Modelling the Spatial Correlation of Earthquake Ground Motion: Insights from the literature, data from the 2016-2017 Central Italy earthquake sequence and ground-motion simulations

Erika Schiappapietra[a,1], John Douglas[a]

[a] *Department of Civil and Environmental Engineering, University of Strathclyde, Glasgow, G11XJ, UK.*

ABSTRACT

Over the past decades, researchers have given increasing attention to the modelling of the spatial correlation of earthquake ground motion intensity measures (IMs), particularly when the seismic risk of spatially distributed systems is being assessed. The quantification of the seismic performance of these systems requires the estimation of simultaneous IMs at multiple locations during the same earthquake, for which the correlation between pairs of locations needs to be defined. Numerous spatial correlation models of common IMs, such as peak ground acceleration and spectral acceleration, have been published. Although the functional forms of the models are generally similar, significant discrepancies exist in terms of the rate of decay of the correlation with increasing inter-site separation distance. The main reasons for such differences lie with the selected databases, the ground-motion models used to derive the spatial correlation models, estimation approaches and regional geological conditions.

In this study, we aim to provide a comprehensive review of spatial correlation models, analysing factors that most affect the spatial dependency of IMs. We use strong-motion records from the 2016-2017 Central Italy earthquake sequence combined with ground-motion simulations to examine the influence of various factors on spatial correlation models. We investigate the dependency on: (1) the estimation method and model fitting technique; (2) the magnitude; (3) the response-spectral period; and (4) local-soil conditions. Our results suggest that the rate of decay is not only period-dependent, but also regionally-dependent, so that a single universal correlation model based on large datasets is

[1] Corresponding author.
E-mail address: erika.schiappapietra@strath.ac.uk (E. Schiappapietra)

not appropriate when describing the correlation behaviour of small geographical areas. Our outcomes could be used to guide the development of new spatial correlation models.

Keywords: *spatial correlation; ground-motion intensity measures; seismic hazard; strong ground motion; earthquake.*

1. Introduction

Stakeholders, such as government, search-and-rescue organizations and private companies, require a reliable evaluation of the ground-motion field to assess the effects of an earthquake for more informed risk management and decision making designed to reduce economic and human losses (e.g. Park *et al.*, 2007; Weatherill *et al.*, 2015). The probabilistic assessment of ground-motion intensity measures (IMs) at a single site is now a well-established technique, and sophisticated methods have been developed for this aim. Traditional seismic hazard and risk analysis tools usually determine the ground motion caused by an earthquake through ground motion prediction equations (GMPEs). GMPEs provide an estimate of the ground shaking and its associated aleatory variability at a given site, considering IMs at different sites as independent. The seismic risk assessment of spatially-distributed systems, such as long bridges, water and power lifelines and portfolios of buildings, however, require not only the estimation of simultaneous IMs at multiple locations during the same earthquake, but also the quantification of the correlation structure (e.g. Goda and Atkinson, 2009, 2010; Jayaram and Baker, 2009; Esposito and Iervolino, 2011, 2012; Weatherill *et al.*, 2015; Wagener *et al.*, 2016; Heresi and Miranda, 2019). Indeed, understanding the spatial characteristics of the ground motion arising from similarities in the seismic wave paths and local-site effects is needed to provide a more accurate representation of ground-motion fields (Verros *et al.*, 2017). For instance, correlation models can be used to generate spatially-correlated random fields for use in developing either scenarios for future earthquakes or retro-scenarios of past events in terms of ground motion shaking, as well as for loss estimates. Iervolino (2013), among others, discuss the importance of considering regional hazard for calculation of aggregate risk. This author compared the annual rate

of exceedance of a specific IM in at least one site among several with the corresponding site-specific hazard assessment, demonstrating that consideration of the joint probability of occurrence in the hazard computation leads to larger values than those obtained for individual sites. Similar conclusions are found in Sokolov and Ismail-Zadeh (2016) and Sokolov and Wenzel (2019). These studies outline the discrepancies between site-specific PSHA and multiple-site PSHA: smaller within-event correlations combined with larger reference areas make the hazard assessment for individual sites and multiple sites prone to remarkable differences. Besides, Sokolov and Wenzel (2019) demonstrated that the fraction of the reference area in which the design ground motion level will be exceeded depends not only on the considered return period, but also on the correlation model. For instance, for a return period of 475 years, there is a 10% of probability that the design ground motion level will be exceeded at least once in 50 years in 20% and 40% of the reference area when within-event residuals are uncorrelated and perfectly correlated, respectively, with corresponding impacts on loss estimates. The importance of defining spatially-correlated ground-motion fields in seismic risk assessment was also demonstrated by Park *et al.* (2007) and Sokolov and Wenzel (2011). Neglecting the spatial correlation may cause a bias in loss estimates, overestimating the most likely losses and underestimating rare losses. On the contrary, overestimating the correlation may lead to the opposite result. As remarked in Park *et al.* (2007), the effects of including or not the spatial correlation depend also on the considered portfolio. Analogous outcomes are provided in Crowley *et al*. (2008), in which the authors compared the variation of the mean damage ratio of the loss model, obtained considering uncorrelated and correlated ground motion fields as well as correlated ground motion fields constrained to the recordings available at a specified number of sites. Given observations at recording stations and including proper correlation models makes the variability of the ground motions lower, and hence the variability of the losses narrower.

Over the past decade, the number of studies on spatial correlation has increased significantly. Many models have been published (**Table 1**) and common findings suggest that: (1) within-event correlation

decays rapidly with increasing inter-site separation distances, and (2) ground-motion IMs associated with longer response-spectral periods have larger correlation lengths (Goda and Hong, 2008; Jayaram and Baker, 2009; Esposito and Iervolino, 2012; Wagener *et al.*, 2016; Sgobba *et al.*, 2019). Nevertheless, a thorough comparison among the proposed models demonstrates significant inconsistencies, which may make the assessed seismic risk prone to large uncertainties (**Figure 1**). The majority of the studies are for different regions (e.g. California, Japan and Taiwan), suggesting that the underlying database as well as regional and local-site effects are likely to play first-order roles in the observed differences. Besides, the use of either existing GMPEs or *ad hoc* study-specific GMPEs as well as different estimation approaches may contribute to different outcomes. Further analyses are thus required to draw firm conclusions on the spatial correlation structures of different IMs.

References	**Database**	**Considered GMPEs**	**Method**	**Spatial Correlation Model**	**Ground Motion Intensity Measure**
Boore et al. (2003)	Mw 6.7 1994 Northridge earthquake	-	Standard deviation of the difference of the logarithm of the PGA	Exponential	PGA
Wang and Takada (2005)	6 earthquakes recorded in Japan and Taiwan (6.2 ≤ Mw ≤ 8.0)	GMPE by Annaka et al. (1997) and by Midorikawa and Ohtake (2002)	Direct estimation of the sample correlation coefficient	Exponential	PGV
Goda and Hong (2008)	6 Californian earthquakes (5.9 ≤ Mw ≤ 7.1)	*Ad hoc* GMPE calibrated on 592 records from 39 Californian earthquake	Semivariogram and direct estimation of the correlation coefficient	Exponential	PGA and SA for T up to 3 s
	Mw 7.6 Chi Chi Earthquake	Boore and Atkinson (2007)			PGA and SA at T = 0.3; 1; 3 s
Goda and Atkinson (2009)	K-NET and KiK-NET database: 106 earthquakes with Mw >= 5.5 and depth < 200 km	*Ad hoc* GMPE	Semivariogram and direct estimation of the correlation coefficient	Exponential	PGA and SA (the final model is independent of T)
Hong et al. (2009)	39 Californian earthquakes with Mw ≥ 5.0	*Ad hoc* GMPE	Correlation is computed within the	Exponential	PGA and SA for T up to 3 s

			GMPE regression		
Jayaram and Baker (2009)	6 Californian earthquakes ($5.1 \leq Mw \leq 6.7$) Mw 7.6 Chi Chi Earthquake	Boore and Atkinson (2008) and Chiou and Youngs (2008)	Semivariogram (Classic estimator)	Exponential	SA for T up to 10 s
Goda and Atkinson (2010)	SK-NET, K-NET and KiK-NET database: 20 earthquakes with $5.6 \leq Mw \leq 6.8$ and 7 km $\leq$ depth $\leq$ 70 km	GMPE by Goda and Atkinson (2009)	Semivariogram	Exponential	PGA and SA (the final model is independent of T)
Jayaram and Baker (2010)	dataset used in Campbell and Bozorgnia (2008)	GMPE by Campbell and Bozorgnia (2008)	Correlation is computed within the GMPE regression	-	PGA and SA for T up to 10 s
Sokolov et al. (2010)	TSMIP network in Taiwan: 66 earthquakes with $Ml \geq 4.5$ and depth $\leq$ 30 km	*Ad hoc* GMPE	Direct estimation of the sample correlation coefficient	Exponential	PGA
Esposito and Iervolino (2011)	Subset of ESM (Mw 5-7.6 and R_{jb} 0-100km) and ITACA (Mw 4-6.9 and R_{jb} 0-196km)	ESM: Akkar and Bommer (2010) ITACA: Bindi et al. (2010)	Semivariogram (Robust and classic estimator)	Exponential	PGA and PGV
Goda (2011)	PEER-NGA, K-NET, KiK-NET and SK-net database: 41 earthquakes with $Mw \geq 5.5$	GMPE by Boore and Atkinson (2008) and by Goda and Atkinson (2009)	Semivariogram	-	SA for T up to 2 s
Esposito and Iervolino (2012)	Subset of ESM (Mw 5-7.6 and R_{jb} 0-100km) and ITACA (Mw 4-6.9 and R_{jb} 0-196km)	ESM: Akkar and Bommer (2010) ITACA: Bindi et al. (2011)	Semivariogram (Robust and classic estimator)	Exponential	SA: T up to 2 s for ITACA and 2.85 s for ESM
Sokolov et al. (2012)	TSMIP network in Taiwan: 54 earthquakes with $Ml \geq 5.0$ and depth $\leq$ 30 km	GMPE by Sokolov et al. (2010) and by Tsai et al. (2006)	Semivariogram and direct estimation of the correlation coefficient	Exponential	PGA
Du and Wang (2013)	5 Californian earthquakes ($5.1 \leq Mw \leq 6.7$) 3 Japanese earthquakes ($6.6 \leq Mw \leq 6.8$) Mw 7.6 Chi Chi Earthquake	GMPE by Campbell and Bozorgnia (2008, 2010, 2012)	Semivariogram (Robust estimator)	Exponential	SA for T up to 5 s, CAV and Ia
Sokolov and Wenzel (2013)	84 Japanese earthquakes ($4.2 \leq Mw \leq 7.4$)	GMPE by Kanno et al. (2006) and by Goda and Atkinon (2009)	Semivariogram and direct estimation of the correlation coefficient	Exponential	PGA and PGV

Bradley (2014)	2010-2011 Canterbury earthquakes	GMPE by Bradley (2010-2013)	Direct estimation of the sample correlation coefficient	-	SA for T up to 6 s
Foulser-Piggott and Goda (2015)	Subset of 203 Japanese earthquakes with more than 100 records	*ad hoc* GMPE calibrated on 661 Japanese earthquakes with Mw ≥ 5.0, depth ≤ 150 km and R ≤ 300 km.	Semivariogram	-	Ia and CAV
Wagener et al. (2016)	8 earthquakes (3.5 ≤ Mw ≤ 5.1) recorded in the Marmara region	GMPE by Akkar and Bommer (2010)	Semivariogram	Exponential	PGA and SA for T up to 1 s
Garakaninezhad and Bastami (2017)	9 earthquakes (5.2 ≤ Mw ≤ 7.6) occurred in California, Japan and Taiwan	GMPE by Campbell and Bozorgnia (2014)	Semivariogram	-	PGA and SA for T up to 10 s
Infantino et al. (2018)	Data from ground motion simulations carried out using the numerical code SPEED (Mw 6.0 Po Plain earthquake, Mw 6.5 Volvi earthquake, Mw 7.0 Istanbul scenario, Mw 6.5 Beijing scenario	*Ad hoc* GMPE	Semivariogram (Classic estimator)	Exponential	PGA and SA for T up to 3.5 s
Stafford et al. (2018)	24 earthquakes occurred in the Groningen field (2.5 ≤ Ml ≤ 3.6)	GMPE by Bommer et al. (2018)	Semivariogram (Classic estimator)	Exponential	SA for T up to 1 s
Chen and Baker (2019)	Physics-based simulations from the CyberShake platform (scenarios for souther California)	-	Semivariogram and direct estimation of the correlation coefficient	Exponential	PGA and SA for T up to 10 s
Heresi and Miranda (2019)	39 well recorded worldwide earthquakes	GMPE by Boore et al. (2014)	Semivariogram	Exponential	PGA and SA for T up to 10 s
Huang and Galasso (2019)	233 Italian earthquakes (4 ≤ Mw ≤ 6.9)	*Ad hoc* GMPE	Correlation is computed within the GMPE regression	Exponential	PGA and SA for T up to 4 s
Ming et al. (2019)	Simulations based on 62 Italian earthquakes (5 ≤ Mw ≤ 6.9) to validate the scoring algorithm approach	*Ad hoc* GMPE	Correlation is computed within the GMPE regression		PGA

Sgobba et al. (2019)	29 earthquakes of the 2012 Emilia sequence	GMPE by Lanzano et al. (2016)	Semivariogram (Classic estimator)	Exponential	PGA and SA for T up to 4 s

Table 1: Main published correlation models. PGA is peak ground acceleration, PGV is peak ground velocity, SA is response spectral acceleration, Ia is Arias intensity, CAV is cumulative absolute velocity and T is spectral period. Mw is moment magnitude and R_{JB} is Joyner-Boore distance. GMPEs is ground motion prediction equations.

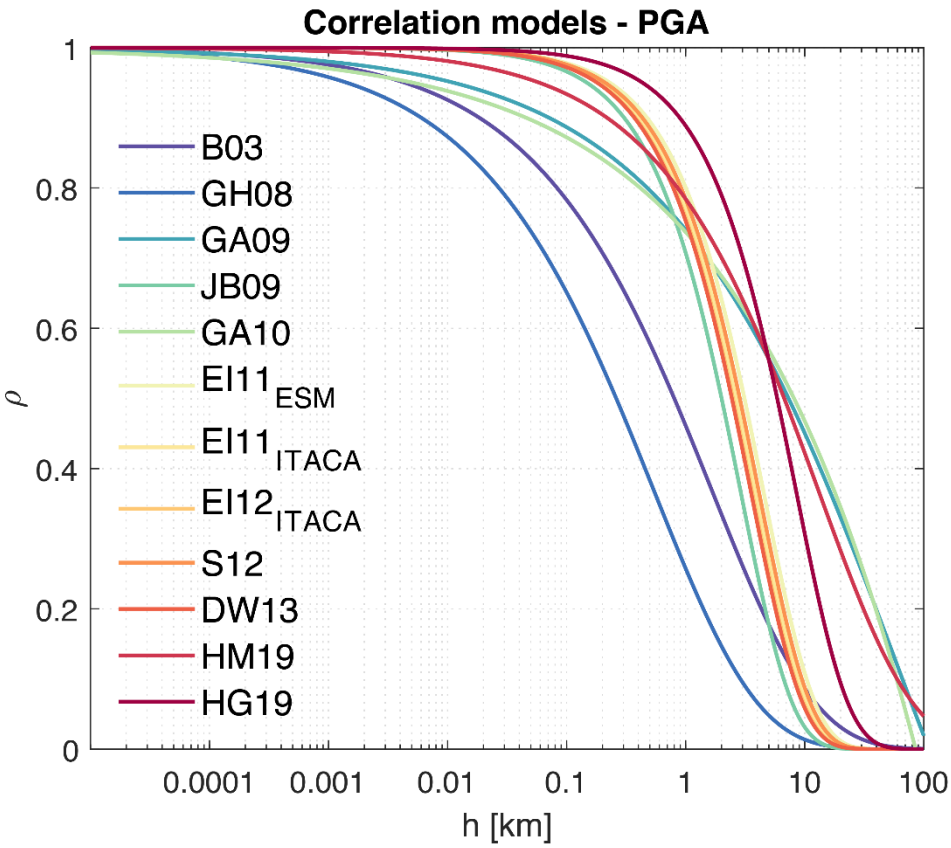

Figure 1: A typical selection of spatial correlation models for PGA: B03: Boore et al. (2003); GH08: Goda and Hong (2008); GA09: Goda and Atkinson (2009); JB09: Jayaram and Baker (2009); GA10: Goda and Atkinson (2010); $EI11_{ESM}$: Esposito and Iervolino (2011) ESM database; $EI11_{ITACA}$: Esposito and Iervolino (2011) ITACA database; $EI12_{ITACA}$: Esposito and Iervolino (2012) ITACA database; S12: Sokolov et al. (2012); DW13: Du and Wang (2013); HM19: Heresi and Miranda (2019); HG19: Huang and Galasso (2019).

To address these issues, we use large databases of recorded strong ground motion from previous earthquakes and ground-motion simulations. In particular, we carry out a geostatistical analysis using the database of the 2016-2017 Central Italy earthquake sequence, which includes nearly 6900 records from 63 Mw ≥ 3.7 events (and nearly 1600 records from nine events with Mw ≥ 5.0) that occurred over a period of five months (August 2016 – January 2017). We choose this dataset because it allows: (1) removal of some of the uncertainties related to the underlying region; and (2) quantification of the variability of spatial correlation among different earthquakes when the same area is considered.

Conversely, simulation of spatially-correlated ground-motion fields provides a controlled environment to test the factors that most influence the correlation structure.

In this article our goal is to provide a well-structured review critically summarising the main findings to advance understanding of the spatial correlation of strong ground motion. We employ the outcomes from the simulations and observations to address the main issues and research gaps.

2. Basic Definitions

The correlation of ground-motion IMs from an earthquake includes three main elements, namely: (1) the spatial correlation among residuals of the same IM for adjacent sites; (2) the correlation among residuals of different IMs at the same location; (3) the spatial cross-correlation among residuals of different IMs for closely spaced sites (Weatherill *et al.*, 2015). We focus on the first of the above-mentioned elements.

In general, the similarity of ground-motion IMs at two different sites depends on: (1) the earthquake source; (2) the propagation path from the source to the sites and local-site effects; (3) the position of closely-spaced sites in near-source conditions with respect to the main fault asperities (Park *et al.*, 2007). The first of these aspects is commonly accounted for by the between-event residual provided by the GMPE. GMPEs relate a ground motion IM (e.g. peak ground acceleration, PGA; peak ground velocity, PGV; peak ground displacement, PGD; or pseudo-spectral acceleration for 5% of critical damping, SA) to a set of explanatory variables describing the source (e.g. magnitude and faulting mechanism), the wave propagation path (e.g. distance metric and regional effects) and the site conditions (e.g. soil classification) (e.g. Douglas and Edwards, 2016). IMs are commonly modelled as lognormally-distributed random variables, through a mixed-effects approach. Therefore, GMPEs take the form:

$$\log_{10} Y_{ij} = \log_{10} \bar{Y}_{ij}(M, R, S, \theta) + \varepsilon_{ij} + \eta_i \quad (1)$$

where Y_{ij} is the IM of interest at the jth site due to the ith event, whereas $\bar{Y}_{ij}$ is the predicted median function of magnitude (*M*), distance from the source (*R*), local-site conditions (*S*) and other explanatory variables (θ). η_i is the between-event residual term, assumed as an independent, identically and normally distributed random variable with zero mean and standard deviation τ. It denotes the systematic deviation of observed IMs associated to an event with respect to the GMPE prediction and does not depend on the site. Conversely, ε_{ij} represents the independent within-event residual term, which is site dependent as it accounts for differences from the average model due to the path and local-site effects. ε_{ij} follows a multivariate Gaussian distribution, completely defined by its mean function $E[\varepsilon_{ij}]$, and covariance function, which reflects the correlation of within-event residuals:

$$COV_j(i,k) = E\left[\varepsilon_{ij} \cdot \varepsilon_{kj}\right] - E\left[\varepsilon_{ij}\right] \cdot E\left[\varepsilon_{kj}\right] \tag{2}$$

The within-event correlation between all pairs of sites during an earthquake due to the similarity of travel path and local-site effects depends on the inter-site separation distance. Typically, it is modelled as an exponential function:

$$\rho_\varepsilon(h) = \exp\left(\alpha \cdot h^\beta\right) \tag{3}$$

where α and β are the model coefficients, usually inferred through a least-squares approach, and h is the separation distance.

As above-mentioned, the between-event correlation, which arises from commonality of the rupture process, is taken into account by the η_i term and is defined as the ratio between the variability components:

$$\rho_\eta = \frac{\tau^2}{\sigma_T^2} \tag{4}$$

in which $\sigma_T = \sqrt{\tau^2 + \phi^2}$ is the total variance and ϕ is the standard deviation of the within-event residuals. Therefore, the total correlation among residuals for closely spaced sites, which considers both the between-event and within-event variability, is given by:

$$\rho_T(h) = \rho_\eta + \rho_\varepsilon(h)\frac{\phi^2}{\sigma_T^2} \tag{5}$$

It is noted that the between-event correlation among different pairs is always positive, as a result of all sites sharing the same source rupture. As a matter of fact, the total correlation will never drop to zero, although $\rho_\varepsilon(h)$ decays to zero with increasing separation distance (Heresi and Miranda, 2019). In this study, we focus only on the within-event correlation because the within-event residual term is the only component of the total variability that varies from site to site and thus affects the spatial dependency of IMs (Stafford, 2012).

Furthermore, only one observation from a given earthquake is available for each station, making it impossible to draw any inferences from it (Webster and Oliver, 2007). Therefore, the hypothesis of second-order stationarity and isotropy are generally assumed, so that the mean function of the random variable (ε_{ij}) is constant for all sites, and the covariance $COV_j(i,k)$, and thus the correlation, depends only on the separation distance (h) between two sites and not on their absolute position and orientation $\left[\rho\left(\varepsilon_{ij}, \varepsilon_{jk}\right) = \rho(h)\right]$.

Finally, one should be aware that the correlation is usually computed on the residual terms rather than directly on the IM. Indeed, the diverse underlying distributions of each IM value, due to different explanatory variables (such as R, S, θ), would make the assessment of the IMs correlation inappropriate (Heresi and Miranda, 2019).

3. Databases

3.1. Central Italy earthquake sequence

Starting from 24th August 2016, one of the most important earthquake sequences ever recorded in Italy struck the Central Apennines between the municipalities of Amatrice and Norcia, causing widespread damage, thousands of homeless and invaluable losses for the historical heritage of the region. The first mainshock (Mw 6.0) struck on 24th August 2016 at 01:36 UTC near Amatrice and it was followed, within less than an hour, by a Mw 5.4 aftershock (Chiaraluce et al., 2017). After two months, two other large earthquakes occurred: a Mw 5.9 on 26th October 2016 at 19:18 UTC, near the village of Ussita and a Mw 6.5 on 30th October 2016 at 06:40 UTC with an epicentre close to Norcia (Luzi et al., 2017). Four other Mw ≥ 5.0 earthquakes occurred on 18th January 2017 near the villages of Campotosto and Montereale (**Figure 2, Table 2**). Event and stations metadata are from Lanzano *et al.* (2018). All the events, generated by normal fault segments, were recorded by permanent and temporary networks, set up to monitor the earthquake sequence at a high resolution and to retrieve more accurate observations of the ground shaking in the near-source region (Luzi et al., 2017).

In this study, we select data from strong-motion stations within an epicentral distance of 200 km. The distributions of the selected data with respect to distance, magnitude and distance, magnitude and number of stations per EC8 soil classes (Eurocode 8, 2004) are summarized in **Figure 3**. The site conditions at each strong-motion station are expressed through the EC8 soil categories, which is based either on the average shear-wave velocity of the upper-most 30 m ($V_{S,30}$) or on the available geological information. Most of the selected stations are classified as site class B/stiff soil.

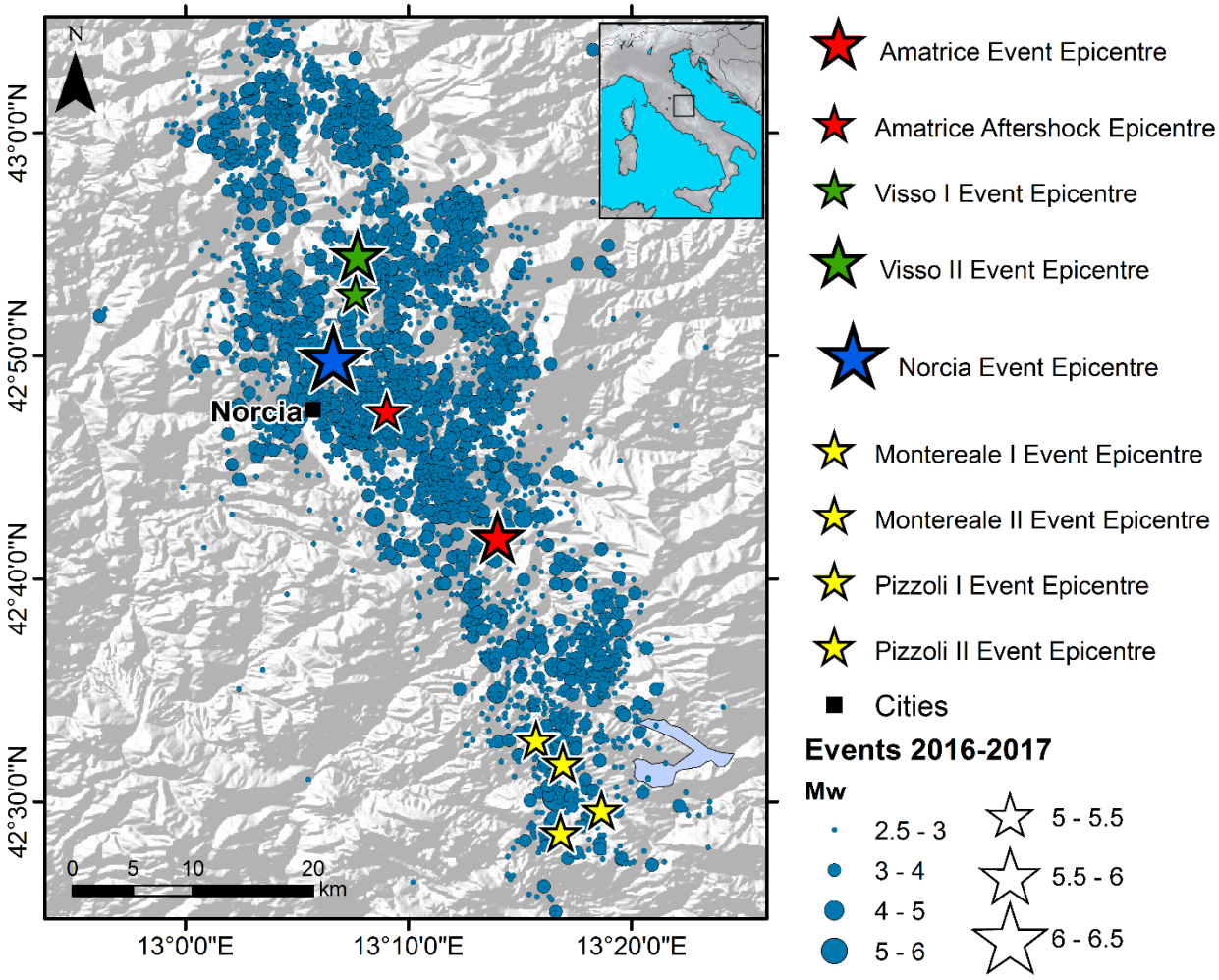

Figure 2: Epicentres (stars) of the nine Mw ≥ 5.0 earthquakes of the sequence. Mw ≥ 2.5 aftershocks for one year after the Amatrice mainshock are also mapped with blue dots. [source: http://cnt.rm.ingv.it/]

Event	Event Time	Latitude [°]	Longitude [°]	Depth [km]	Mw	# Records
Amatrice	24/08/2016 01:36	42.70	13.23	8.1	6.0	158
Amatrice Aftershock	24/08/2016 02:33	42.79	13.15	8.0	5.3	138
Visso I	26/10/2016 17:10	42.88	13.13	8.7	5.4	158
Visso II	26/10/2016 19:18	42.91	13.13	7.5	5.9	166
Norcia	30/10/2016 06:40	42.83	13.11	9.2	6.5	159
Montereale I	18/01/2017 09:25	42.55	13.26	9.2	5.1	127
Montereale II	18/01/2017 10:14	42.53	13.28	9.1	5.5	141
Pizzoli I	18/01/2017 10:25	42.49	13.31	8.9	5.4	129
Pizzoli II	18/01/2017 10:33	42.48	13.28	10.0	5.0	124

Table 2: Main characteristics of the largest events of the Central Italy sequence [source: Lanzano et al. (2018)].

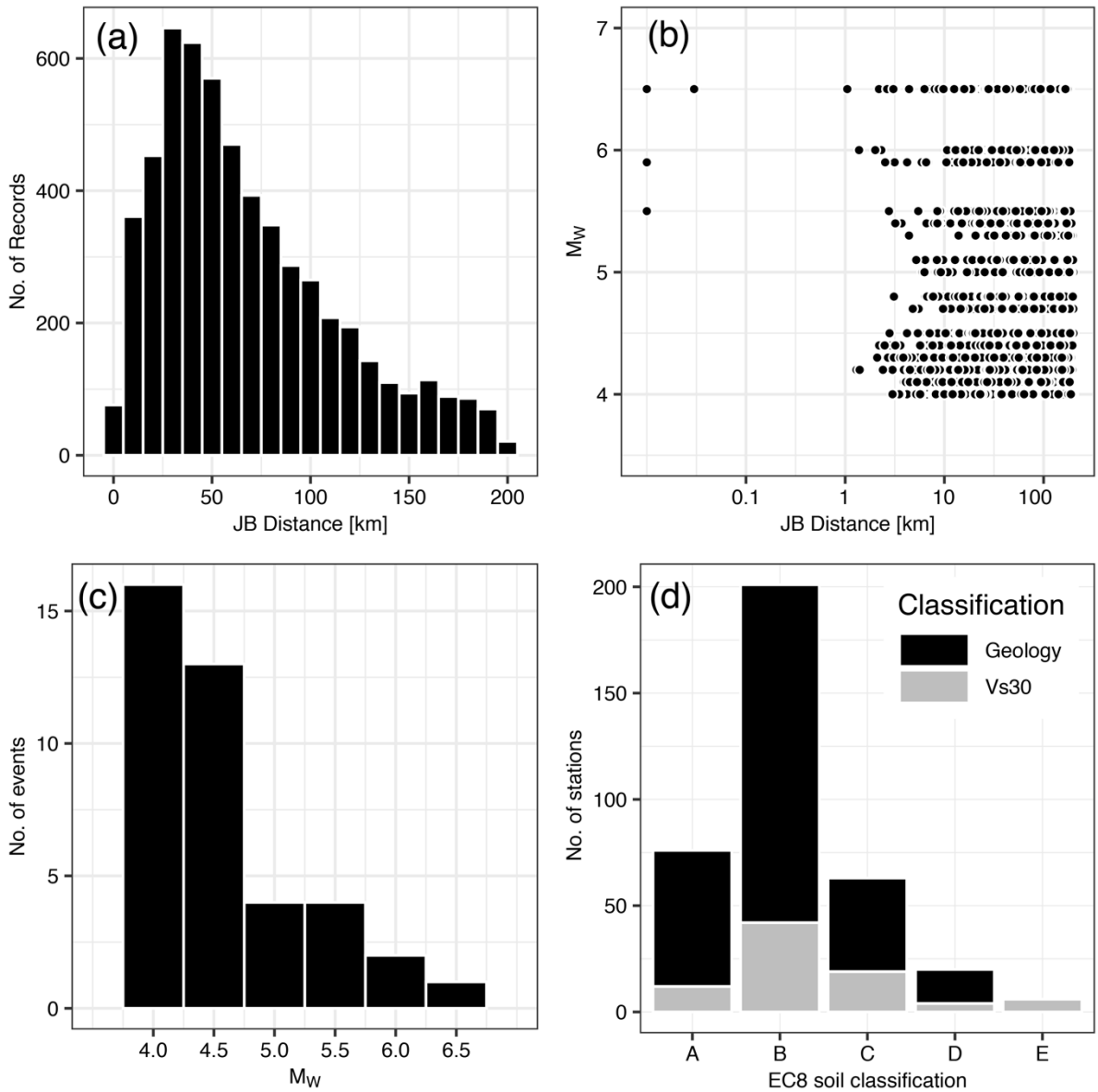

Figure 3: (a) Records-distance distribution; (b) Magnitude-distance distribution; (c) Events-magnitude distribution; (d) Distribution of stations in terms of EC8 site classification. Distances are in terms of Joyner and Boore distance (R_{jb}, i.e. the distance to the surface projection of the rupture).

3.2. Simulated ground-motions fields

To simulate spatially-correlated ground-motion fields, we employ the approach described in Strasser and Bommer (2009), which accounts for both the between- and within-event variabilities. The former is specified through the between-event standard deviation in the IMs computation, whereas the latter is included through the computation of a random field based on a multivariate normal distribution. The procedure includes the following steps:

1. Generation of deterministic fields using GMPEs (GM_{det}). We employ the GMPE by Lanzano *et al.* (2019) for shallow crustal earthquake in Italy, considering homogeneous rock site conditions (**Figure 4a**).
2. Generation of ground motion random fields (**Figure 4c**).

   The spatially correlated random field $P(h_0, \Phi)$ is computed through a multivariate normal distribution characterized by an exponential correlation model with correlation length h_0 and standard deviation Φ (see section 2). We choose different values of h_0 (5, 10 and 30 km) and Φ (0.1, 0.32 and 0.5) to allow a deeper insight into the impact of these parameters on the analysis.
3. Generation of ground-motion stochastic fields (**Figure 4b**).

   This is obtained by combining steps 1 and 2:

$$GM_{sto} = 10^{[GM_{det} + P(h_0, \Phi)]} \tag{6}$$

The ground-motion fields are generated on a 200 km × 200 km grid with a 2 km resolution. Finally, we randomly locate strong-motion recording stations throughout the region and use the simulated IMs at these stations. We repeat the process 1000 times for each $h_0 - \Phi$ pair to obtain stable results.

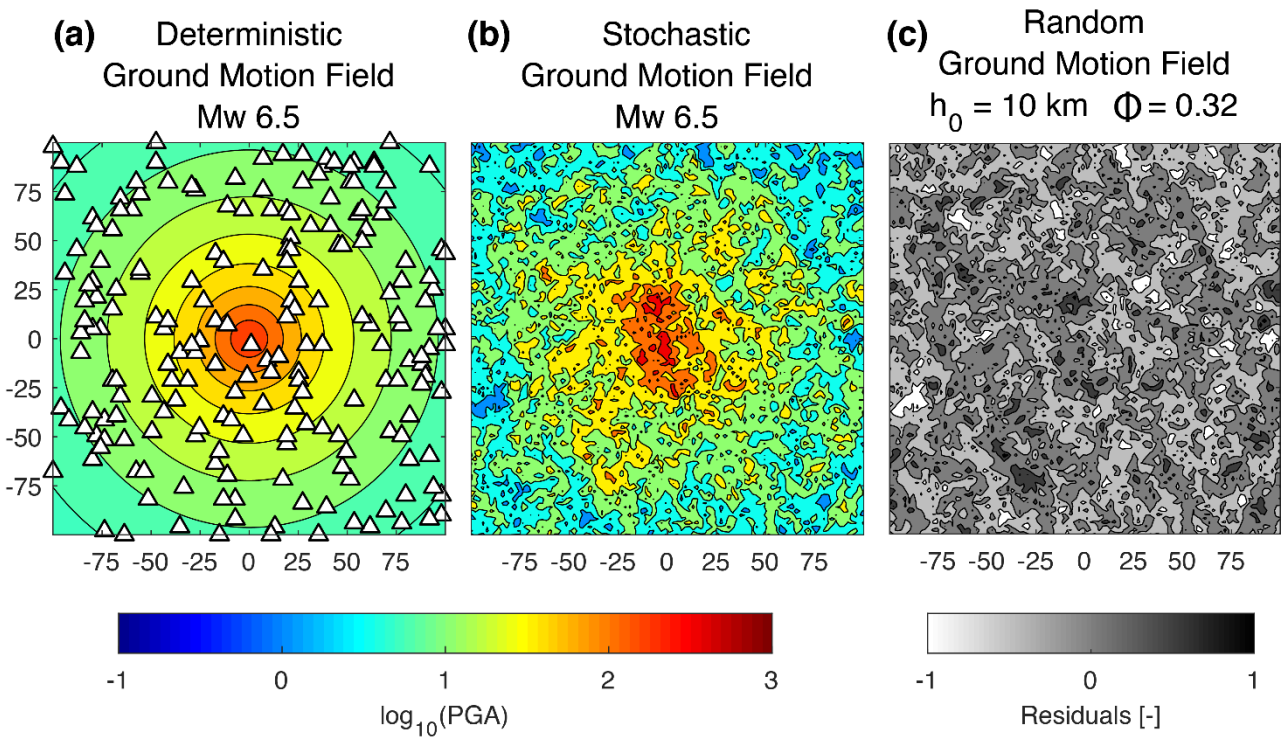

Figure 4: Example of deterministic (a), stochastic (b) and random (c) ground-motion fields. obtained through the above-described approach. PGAs are expressed in cm/s^2. The triangles represent the stations used in the analysis.

4. Modelling of spatial correlation

4.1. Estimation of the within-event spatial correlation

It is common practice to adopt an existing GMPE or an *ad hoc* study-specific GMPE to compute the within-event residuals at each site in order to assess the within-event spatial correlation $\rho_\varepsilon(h)$. This can be estimated with two different approaches, namely: (1) computing directly the covariance and the correlation coefficient (e.g. Wang and Takada, 2005; Sokolov *et al.*, 2010, 2012), or (2) calculating the sample semivariogram, which measures the average dissimilarity between spatially distributed data (e.g. Jayaram and Baker, 2009; Esposito and Iervolino, 2011, 2012; Wagener *et al.*, 2016; Heresi and Miranda, 2019; Sgobba *et al.*, 2019). In the first method, the spatial correlation is estimated as:

$$\rho_\varepsilon(h) = \frac{\mathrm{COV}(\varepsilon_{ij}, \varepsilon_{kj})}{\phi^2} \tag{7}$$

where COV denotes the covariance function, whereas ε_{ij} andε_{kj} are the within-event variabilities for the ith and kth sites during the jth event with zero mean and standard deviation ϕ.

The second method computes the experimental semivariogram to represent the spatial dependency of IMs values with varying separation distance, which is defined as:

$$\hat{\gamma}(h) = \frac{1}{2}\mathrm{Var}\left[\varepsilon_{ij} - \varepsilon_{kj}\right] \tag{8}$$

in which Var indicates the variance. It is common practice to adopt two different estimators to compute the sample semivariogram:

1) Method of moments (Matheron, 1962)

$$\hat{\gamma}(h) = \frac{1}{2N(h)}\sum_{N(h)}\left\{\varepsilon_{ij} - \varepsilon_{kj}\right\}^2 \tag{9}$$

where $\hat{\gamma}(h)$ represents the empirical semivariogram and *N(h)* is the number of pairs separated by *h*.

2) Estimator proposed by Cressie (1985)

$$\hat{\gamma}(h) = \frac{1}{2}\left\{\frac{\left[\frac{1}{|N(h)|}\sum_{N(h)}\left|\varepsilon_{ij} - \varepsilon_{kj}\right|^{0.5}\right]^4}{0.457 + \frac{0.494}{|N(h)|}}\right\} \tag{10}$$

This is found to be a more robust estimator, being less sensitive to outliers (Esposito and Iervolino, 2012; Du and Wang, 2013; Oliver and Webster, 2014).

Both the covariance and the semivariogram are calculated for each pair of stations (x_i, x_k) whose inter-site spacing falls in a distance bin defined as $h - \Delta h/2 \leq \left|x_i - x_j\right| \leq h + \Delta h/2$. Esposito and Iervolino (2012) and Du and Wang (2013) suggest setting the bin size in such a way that there are at least 30 pairs in each bin. Other authors (e.g. Wagener *et al.*, 2016) advise having at least 100 pairs per bin to have more reliable and representative estimations. Furthermore, the definition of ϕ is of primary importance. ϕ can be estimated either from the sample semivariogram at large separation distances, where the within-event residuals are assumed to be uncorrelated, or as the standard

deviation of the within-event residuals for a given event (Goda and Atkinson, 2010). Alternatively, Esposito and Iervolino (2012) employ the standard deviation related to the GMPE as ϕ. However, this approach is strongly discouraged by other studies, since using a constant value of ϕ for different events might lead to a biased within-event correlation (Heresi and Miranda, 2019, Foulser-Piggott and Goda, 2015).

Parametric functions are used to fit the experimental values computed either through the sample semivariogram or covariance approaches. This allows for the spatial variation of ε_{ij} to be retrieved for any separation distance h. Examples of basic second-order stationary and isotropic models are: (1) Matérn; (2) Exponential; and (3) Gaussian (Webster and Oliver, 2007). The exponential model (**Figure 5**), which usually provides the best performance, takes the form:

$$\gamma(h) = a\left[1 - \exp\left(-\frac{ch}{b}\right)\right] \tag{11}$$

where h is the separation distance, a and b are the sill and the range of the semivariogram and c is a given positive constant set to 3. The sill equals the variance of the data, whereas the range represents the distance beyond which the correlation between sites is negligible. The range is the distance at which $\gamma(h)$ equals 0.95 times the sill. It is worth mentioning that in some studies (e.g. Wang and Takada, 2005; Huang and Galasso, 2019) c is set equal to 1. Consequently, the range outlines the distance at which the correlation equals $\exp(-c) = \exp(-1) \cong 0.37$. The implementation of either one or the other value leads to a different meaning of the range, without however affecting the spatial correlation structure of the IM of interest. Moreover, under the hypothesis of second-order stationarity, the semivariogram and the covariance function are equivalent, so that the following relation holds (Oliver and Webster, 2014):

$$\gamma(h) = \Phi^2 - \mathrm{COV}\left(\varepsilon_{ij}, \varepsilon_{kj}\right) = \Phi^2[1 - \rho_\varepsilon(h)] \tag{12}$$

being $\rho_\varepsilon(h) = \mathrm{COV}\left(\varepsilon_{ij}, \varepsilon_{kj}\right) / \Phi^2$.

Consequently, combining Eq. (11) and Eq. (12), we have that:

$$\rho_{\varepsilon}(h) = 1 - \frac{\gamma(h)}{\Phi^2} = \exp\left(-\frac{ch}{b}\right) \tag{13}$$

which is equivalent to Eq. (3).

Furthermore, the hypothesis of second-order stationarity may sometimes not hold. The expected value of the random variable $E[\varepsilon_{ij}]$ may not be constant across all sites, but indeed varying depending on the location. In such cases, the semivariogram $\hat{\gamma}(h)$ increases with separation distance, without reaching a stable sill (Diggle and Ribeiro, 2007). These long-range spatial trends should be removed so that small-range correlation structures can be detected. It is common practice to model spatial trends through trend surface models, namely the mean function is described by either first- or second-order polynomial functions of the geographic coordinates (Diggle and Ribeiro, 2007; Oliver and Webster, 2014). Alternatively, the trend and the semivariogram of the residuals can be estimated all at once through the residual maximum likelihood approach, as suggested by Oliver and Webster (2014).

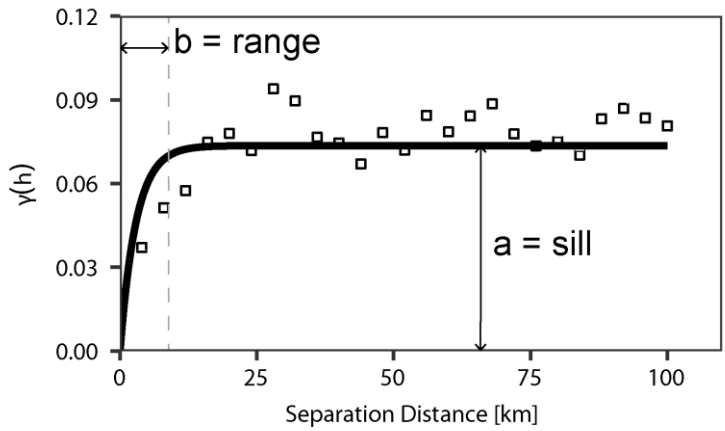

Figure 5: Exponential model. The black solid line is the theoretical model, whereas black squares represent the experimental semivariogram. The sill and range parameters are also highlighted.

Different approaches have been used in previous studies to fit the experimental data, such as the weighted least-squares and manual-fitting techniques. Jayaram and Baker (2009), among others, suggest inferring the model parameters manually so that the experimental data are better fitted at

shorter separation distances. This technique should, however, be discouraged as it involves a certain degree of subjectivity that can lead to prediction errors. Regarding the least-squares approach, weights can depend on either the separation distances (e.g. increasing the weights of data at shorter distances) or the number of pairs in each bin.

4.2. Modelling caveats

The previously described approach suffers from some shortcomings. GMPEs (eq. 1) are usually developed through regression analysis using either the two-stage algorithm proposed by Joyner and Boore (1993) or the one-stage mixed-effects approach by Abrahamson and Youngs (1992), considering independent within-event residuals. The similarity of the earthquake source, path and local-site effects leads to spatially correlated within-event residuals. This issue was firstly investigated by Hong *et al*. (2009), who demonstrated that the inclusion of spatial correlation in GMPE development does not affect the estimated ground motion model coefficients but it does affect the variability of the between- and within-event residuals. In particular, the between-event variance decreases, whereas the within-event component increases, with consequences for the assessment of seismic risk of spatially-distributed systems. Similarly, Jayaram and Baker (2010) adopted a different approach and drew analogous conclusions. While spatial correlation model parameters can be inferred from statistical analysis of residuals independently of GMPE regressions, caution must be applied in the estimation of Φ and τ. Ming *et al.* (2019) reviewed the studies by Hong *et al*. (2009) and Jayaram and Baker (2010), and pinpointed their main limitations. They proposed a new algorithm (the scoring estimation approach) that, while achieving the same results in terms of Φ and τ, provides more statistically robust ground-motion models along with spatial correlation parameters to improve seismic hazard and risk assessments.

In addition, seismic risk analyses usually employ an average estimate of the correlation structure, derived considering heterogeneous databases, without accounting for its event-to-event variability. Goda (2011) investigated the event-to-event variability of within-event spatial correlation, studying

41 different earthquakes individually. He found that the model by Goda and Atkinson (2010) reasonably fits the overall median tendency. However, ± 0.1 units should be added to the same spatial correlation model to take into account the event-to-event uncertainty. Likewise, Heresi and Miranda (2019) quantified the event-to-event variability, comparing the within-event residuals correlation of 39 well-recorded earthquakes. They provided general equations to calculate the median and dispersion of the correlation length (range), so that the uncertainty can be easily accounted for in regional seismic risk computations.

In **Figure 6**, we compare the outcomes obtained analysing the data from the 2016-2017 Central Italy sequence with the models proposed by Goda (2011) and Heresi and Miranda (2019) along with their event-to-event uncertainty. We follow the approach described in Heresi and Miranda (2019) to compute the central tendency and the standard deviation, based on the number of stations that recorded each event. The Central Italy database features a smaller event-to-event variability because all the earthquakes nucleated within the same region. Conversely, the two other models are calibrated on worldwide datasets, leading to a larger variability in terms of correlation length. According to our analysis, a region-specific variability should be considered when performing regional seismic hazard and risk assessment to obtain more accurate results.

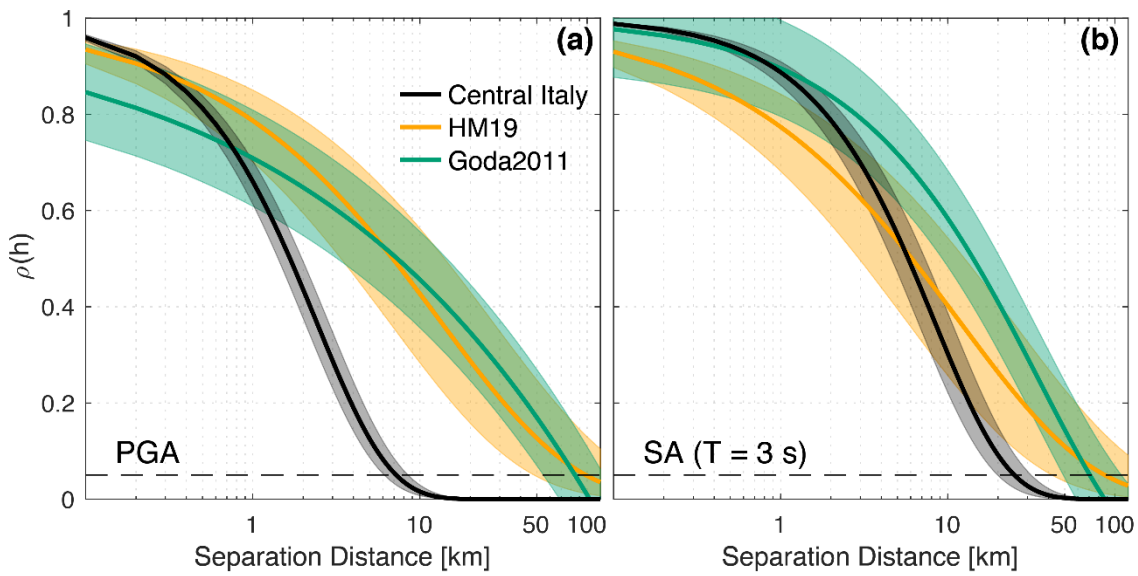

Figure 6: Correlation models proposed by Goda (2011) and Heresi and Miranda (2019) [HM19] and obtained considering data from the Central Italy sequence for PGA (a) and SA(3 s) (b). Solid lines represent the median values, whereas shadow areas indicate the event-to-event uncertainty. Black dashed line helps indicate the distances at which the correlation equals 0.05.

5. Major factors influencing spatial correlation

In this section we present a detailed literature review and the results from our analyses, with the aim of understanding the main factors that affect the spatial correlation of ground-motion IMs.

5.1. Dependence on the estimation approach

In section 4.1 we described two different approaches to estimate the within-event spatial correlation, namely: (1) the covariance and correlation coefficient and (2) the semivariogram. The implementation of either one or the other technique should provide similar results as the semivariogram and the covariance are equivalent for second-order stationary random fields. Goda and Hong (2008) compared the results obtained for the M_w 7.6 Chi Chi earthquake using both the approaches and found that the computation of the covariance provides slightly different estimates. Similarly, Goda and Atkinson (2009), while characterizing the spatial correlation of Japanese data, drew the same conclusions. The correlation decays faster when the correlation coefficient is directly computed, and some inconsistent trends may occur, especially if the number of available data is inadequate. Nevertheless, they provided a general correlation model by averaging the results obtained using both approaches, as subsequently suggested by Sokolov *et al.* (2012, 2013).

In order to address this issue, we carry out a thorough geostatistical analysis of the within-event correlation structure using simulated spatially-correlated ground-motion fields. We compute the range implementing both the techniques and both the sample semivariogram estimators by varying the number of sites. As found by Goda and Hong (2008) and Goda and Atkinson (2009), the direct estimation of the correlation coefficient not only features slightly different results, but also a larger dispersion compared to the semivariogram approach in our simulations (**Figure 7**). Furthermore, the graph well illustrates how the availability of stations plays a crucial role in obtaining more robust

outcomes: even though the mean of h_0 assumes to some extent a constant value, the uncertainty halves as the number of stations increases.

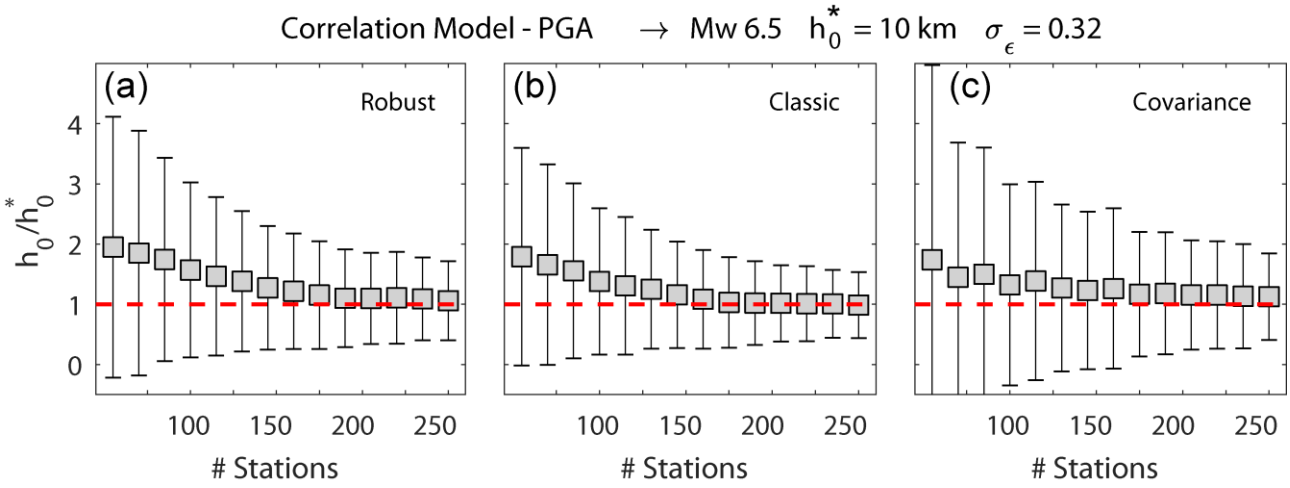

Figure 7: Estimated range h_0 normalized by the true value of the range h^*_0 as a function of the number of sites (stations) using: (a) the Cressie and Hawkins (robust) estimator; the Matheron (classic) estimator; and (c) the covariance approach. Grey squares represent the mean value, whereas the solid vertical lines indicate the variability of the estimates. The red dashed line signifies a ratio of 1.

Most of the proposed models use the classic estimator of Mathéron (1962) to compute the sample semivariogram. The estimates obtained are variances and as such they are sensitive to outliers, which might lead to less reliable semivariograms. A statistically more robust estimator was proposed by Cressie (1985) to down-weight the effects of atypical observations (Oliver and Webster, 2014). Esposito and Iervolino (2011, 2012) implemented both the estimators, concluding that there are negligible differences. Du and Wang (2013) suggested using the robust approach directly to obtain consistent estimations. Clearly, the two different estimators provide almost the same correlation distance in terms of average value, as observed in **Figure 7** and according to Esposito and Iervolino (2011, 2012). However, a more in-depth analysis suggests that: (1) the classic estimator is likely to converge faster; (2) the two different approaches do not provide always the same range, as shown in **Figure 8**. In particular, the smaller the number of considered stations, the more significant are the differences. Although our simulations do not show which is the best semivariogram estimator, Oliver and Webster (2014) demonstrated that if the data features outliers, the classic estimator diverges from the input function, in contrast to the robust estimator.

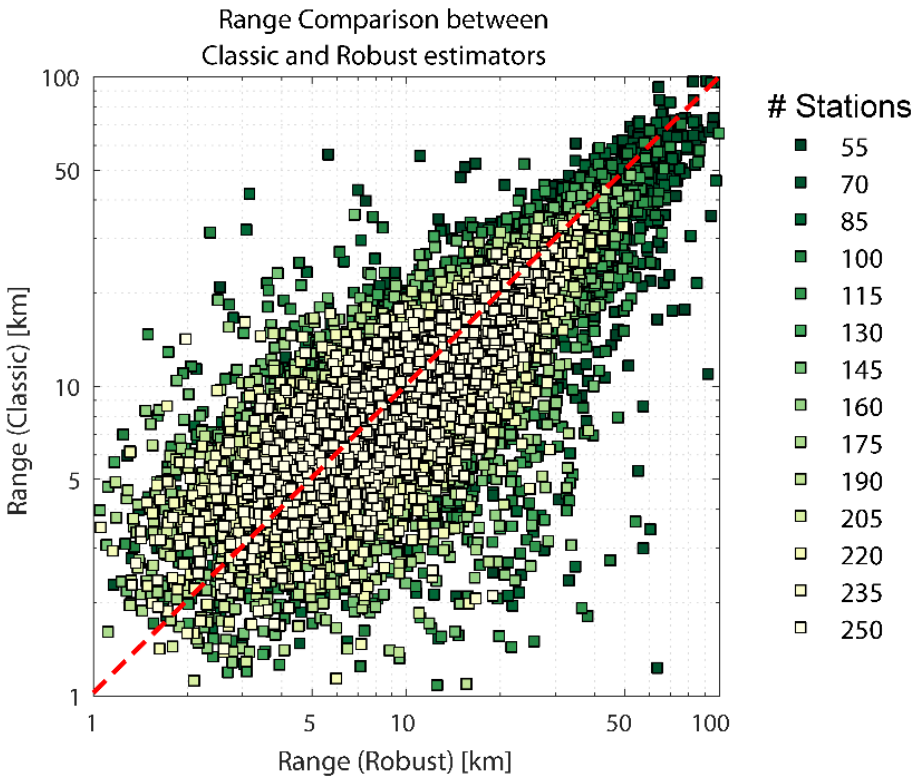

Figure 8: Comparison between the ranges obtained through both the classic and robust estimators, considering different numbers of stations.

5.2. Dependence on the fitting method

A number of methods have been proposed to fit experimental values by means of parametric models. The trial-and-error (manual fitting) and least-squares regression, among others, are the most common techniques. Jayaram and Baker (2009) recommend manually fitting the experimental semivariogram so that shorter separation distances, which are more important for engineering purposes, are weighted higher. Despite its high degree of subjectivity, this approach was chosen also by Du and Wang (2013) for its versatility in fitting the data. Esposito and Iervolino (2011) and Infantino *et al.* (2018) visually fitted experimental semivariograms, assuming the least-squares regression as a basis. Conversely, other studies, such as Esposito and Iervolino (2012), Sokolov *et al.* (2012), Wagener *et al.* (2016), Heresi and Miranda (2019) and Sgobba *et al.* (2019), adopt the weighted least-squares regression approach, using different weights. Some of the authors prioritise either the short separation distances or the number of pairs in each bin; others apply the same weight to all the experimental values, so that the model is fit equally over the full range of data. In our analysis, we opt for the weighted least

squares regression, in which weights are computed based both on the number of pairs and on the separation distance, so that the impact of longer distances is minimized as they are associated with low correlations (Jayaram and Baker, 2009). However, we carry out a further analysis in which all the data are equally weighted to understand the influence of the fitting method on the correlation structure. **Figure 9** illustrates how the different weightings lead to dissimilar results. A similar comparison can be found in Jayaram and Baker (2009), where the authors criticised the study by Wang and Takada (2005). Stafford *et al.* (2018) achieved similar outcomes, arguing that there is variability in the ranges due to the adopted technique. Therefore, according to Jayaram and Baker (2009), among others, we suggest prioritising models that fit the experimental data well at short separation distances, which have a significant impact on seismic risk estimates.

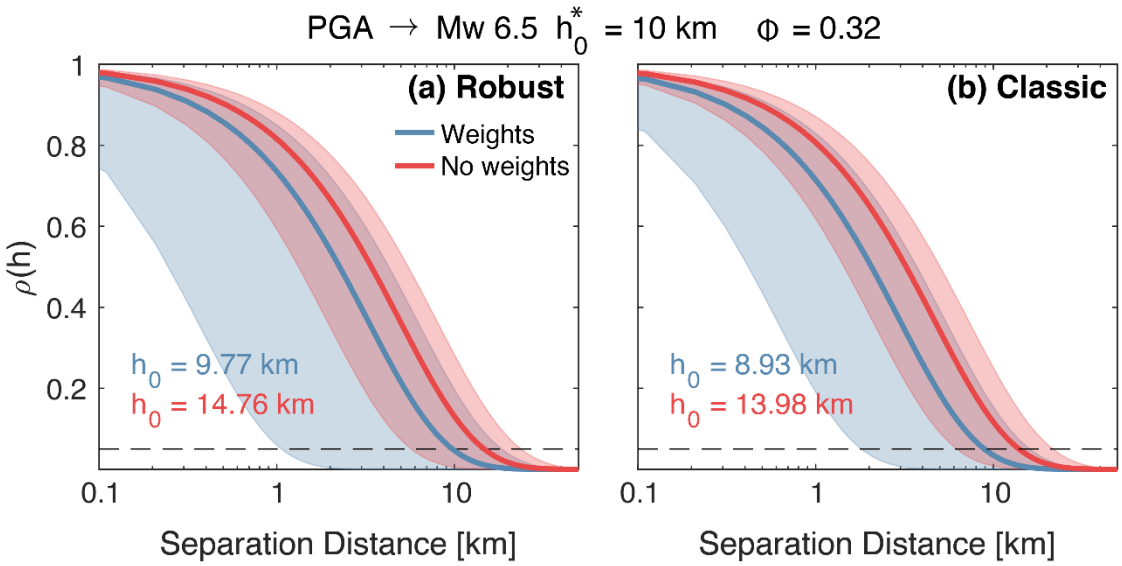

Figure 9: Correlation models obtained through the robust (a) and classic (b) estimators. Blue curves indicate the results in which the weights used to fit the experimental data are computed based on the number of pairs and separation distance; red curves represent the results in which equal weights are applied to all the data. Shadow area represent the standard deviation of the results computed on 1000 simulations. The standard deviation of the model with weights (8.7 and 7.12 km for the robust and classic estimators respectively) is smaller than the standard deviation of the model without weights (9.02 and 7.73 km).

5.3. Dependence on magnitude

The relationship between magnitude and correlation length is one of the most debated topics in studies of spatial correlation. Some studies (e.g. Sokolov *et al.*, 2012, 2013; Foulser-Piggott and Goda, 2015) argue that the range tends to increase with increasing magnitude because moderate-to-large

earthquakes feature lower frequency content and hence an additional non-random component. Conversely, other authors (e.g. Jayaram and Baker, 2009) did not find any correlation between these two parameters. Heresi and Miranda (2019) performed a statistical analysis and found that only a small ratio of the variability in terms of correlation length can be explained by considering the magnitude, in particular for SA at longer periods. Garakaninezhad and Bastami (2017) showed that there is a positive correlation between magnitude and anisotropy ratio (ratio between the ranges of the direction with the largest and smallest spatial correlations, respectively) of residuals for the nine earthquakes analysed. This is valid only for PGA, however, and it does not apply to spectral accelerations.

In **Figure 10** we compare the ranges obtained for each $M_w \geq 4.0$ earthquake belonging to the Central Italy seismic sequence as a function of magnitude. The results do not suggest any clear relationship between range and magnitude, at least for this M_w interval, in agreement with the findings of Jayaram and Baker (2009). It is worth noting that both Sokolov *et al.* (2012, 2013) and Foulser-Piggott and Goda (2015) analysed a wider range of events in terms of magnitude, grouping as moderate-to-large earthquakes all events with $M_w \geq 6.0$. We believe that our analysed database represents too narrow a magnitude interval to draw sound conclusions on the relationship between magnitude and range. Besides, it is biased towards small-magnitude earthquakes, which could influence the result. Other factors should be considered to explain the variability in terms of correlation length, especially when the same seismic region is considered. Stafford *et al.* (2018) demonstrated that the rupture process of events of equal magnitude has a significant contribution on the variability in the range. Likewise, while studying correlation spatial patterns using physics-based ground-motion simulations, Chen and Baker (2019) found a dependence of the correlation on source effects and on the relative position of a site with respect to the fault asperities.

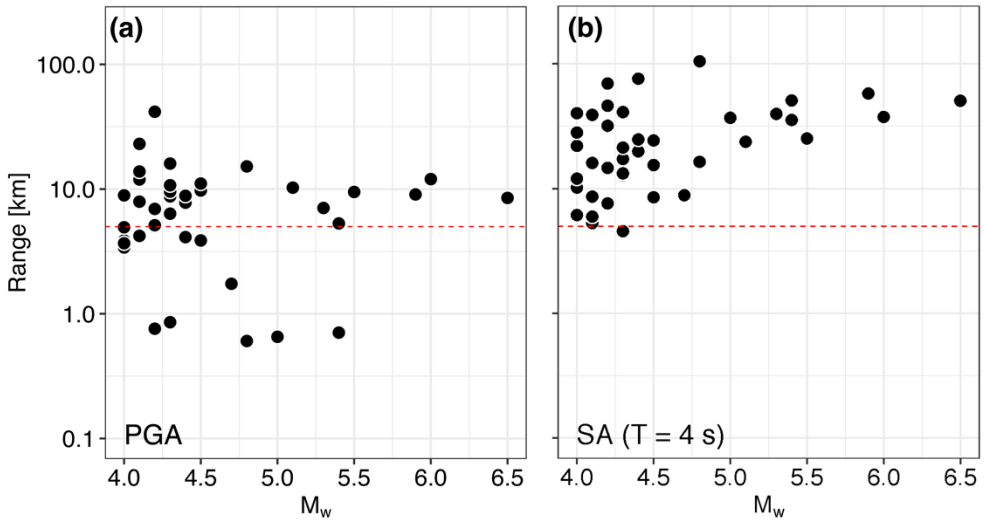

Figure 10: Range as a function of magnitude: (a) PGA; (b) SA(4s). Ranges are obtained by means of the robust semivariogram. We also apply a detrending processing using second-order surface models to remove long-range spatial trends.

5.4. Dependence on period

Previous research (e.g. Goda and Hong, 2008; Jayaram and Baker, 2009; Esposito and Iervolino, 2012; Wagener *et al.*, 2016; Infantino *et al.*, 2018; Sgobba *et al.*, 2019; Chen and Baker, 2019) has demonstrated that the spatial correlation structure is affected by the response-spectral period considered. Range and period are found to be directly proportional. This is consistent with studies on ground-motion coherency, a measure of similarity between waveforms and phases of two ground-motion time histories recorded at two different sites. Hough and Field (1996) demonstrated that ground motions at frequencies up to 2-3 Hz and distances up to 3-4 km are characterized by a high level of waveform coherence, which by contrast drops for shorter period ground motions. Zerva and Zervas (2002), while investigating spatial coherency, proved that small-scale heterogeneities in the travelling path strongly affect the short-period waves (short wavelengths), which therefore tend to be less coherent, in contrast to long-period waves. **Figure 11** provides an overview of the correlation lengths as a function of period of the above-mentioned studies. All the range values in **Figure 11** are computed as the distance at which the correlation equals 0.05 in order to compare the results. Esposito and Iervolino (2012) and Wagener *et al.* (2016) showed that at short periods (up to 0.5 s) it is not

possible to clearly delineate a trend between range and period. Goda and Atkinson (2010) found that the correlation is clearly affected by the vibration period especially at short separation distances (h < 10 km), in contrast to greater distances (h > 50 km), where the models tend to converge towards similar results. Consequently, they proposed an average constant model (**Figure 11)**, calibrated on the results obtained at nine different periods. To further test the hypothesis of a dependence of the correlation on period, here we present a model for the Central Italy sequence computed by pooling all the data from the nine $M_w \geq 5.0$ events.

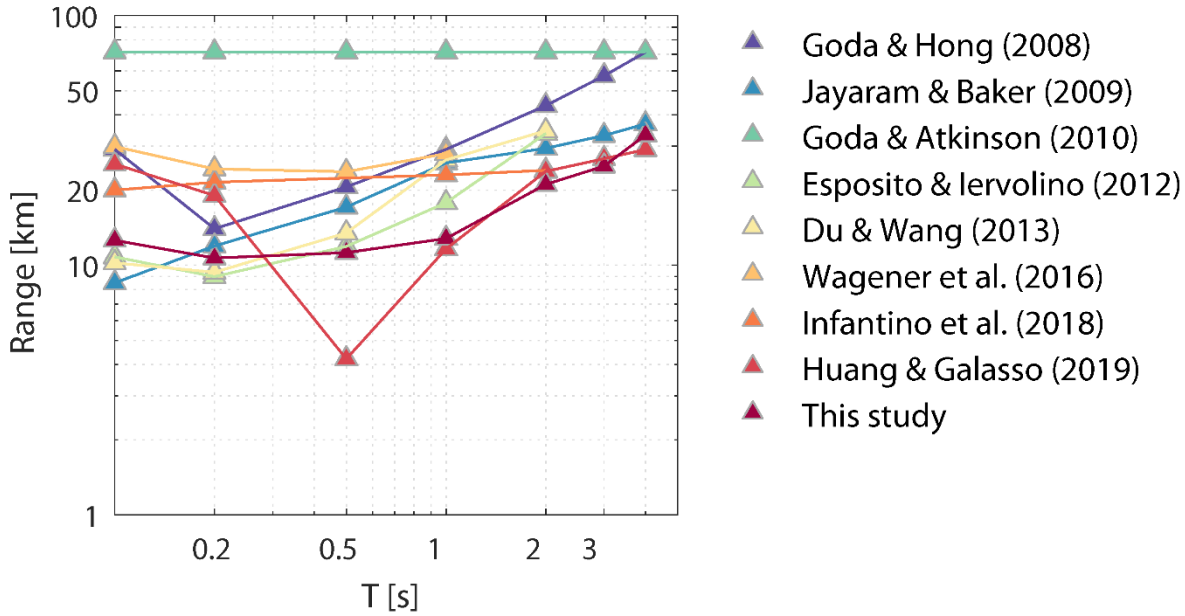

Figure 11: Range as a function of response-spectral period, T. The main studies are reported along with a model for the Central Italy seismic sequence, denoted as "This study". This model is calibrated by pooling all the data from the nine Mw ≥ 5.0 events. The model of Jayaram and Baker (2009) refers to their model for heterogeneous soil condition, whereas the model of Du and Wang (2013) is computed considering a V_{s30} correlation of 4.5 km.

As we summarise in **Figure 12**, we observe the increasing trend with period for all $M_w \geq 5.0$ events of the sequence taken individually, as well.

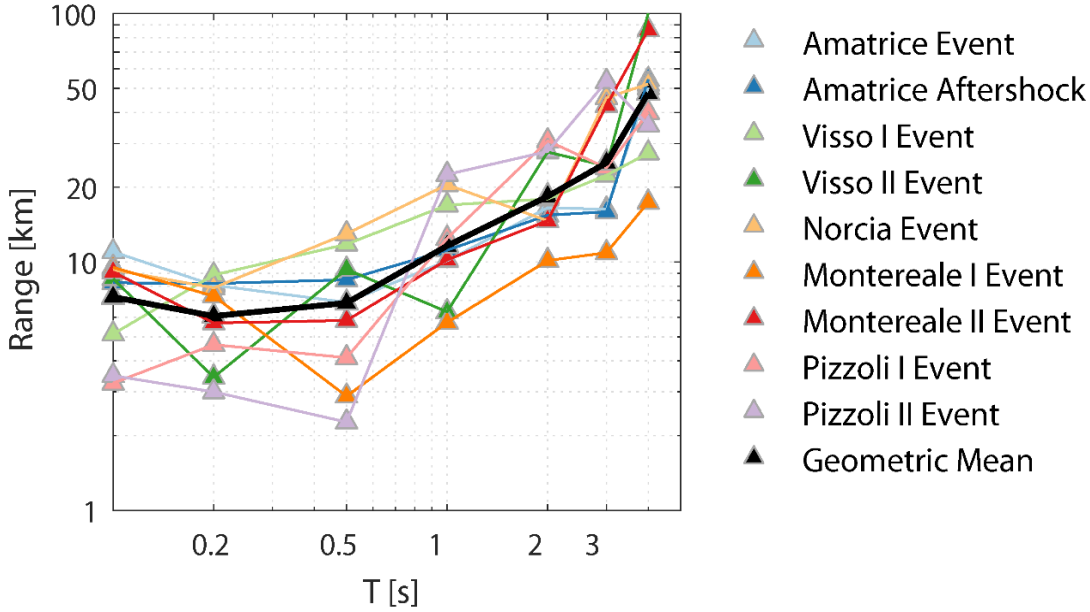

Figure 12: Range as a function of response-spectral period, T, computed using the robust estimator for the nine $M_w \geq 5.0$ events of the sequence. In black we reported the weighted (geometric) average model, in which the weights are given based on the number of stations that recorded each event.

5.5. Dependence on local site effects and geological structure

Several studies (e.g. Jayaram and Baker, 2009; Sokolov *et al.*, 2010, 2012; Du and Wang, 2013; Infantino *et al.*, 2018; Chen and Baker, 2019) indicated that the level of correlation is strongly influenced by regional geologic and site conditions, especially when short-period IMs are of interest. Conversely, Heresi and Miranda (2019) demonstrated that the clustering of V_{s30} values does not explain the high variability in the correlation length. To address this issue, we develop an *ad hoc* GMPE for the Central Italy sequence without accounting for different site conditions and then we introduce a site-specific correction factor to remove any dependency on local-site effects, following Sokolov *et al.* (2010, 2012, 2013). The within-event spatial correlation is understood as a non-random component in the residuals, which arises from neglecting a number of factors in the GMPE, such as site, path and azimuth as well as hanging-wall and footwall effects (Sokolov *et al.*, 2012, 2013). Due to the shortage of data, we could not derive a correlation model for each site class, similarly to Sokolov *et al.* (2012, 2013). The following functional form is assumed:

$$\log_{10} \bar{Y}_{ij} = b_1 + b_2 M_w + (b_3 + b_4 M_w) \log_{10} \sqrt{R_{JB}^2 + b_5^2} \tag{14}$$

where $\bar{Y}_{ij}$ is the PGA or SA(T) for T equal to 0.2, 0.5, 1, 2, 3 and 4 s and R_{JB} is the Joyner-Boore distance (or the epicentral distance for those events where the fault geometry is not defined and the point-source approximation holds). b_1, b_2, b_3, b_4 and b_5 are the model coefficients inferred through the maximum-likelihood regression method by Joyner and Boore (1993). It is noted that our main goal is to estimate the residuals for correlation purposes, and not to develop a new GMPE; therefore, we keep the functional form as simple as possible. The between-event and within-event residual components are evaluated using the mixed-effects approach. In the same way, we compute the site-specific correction term $\delta S2S$, which represents the site-specific deviations from the average model calibrated considering all sites in the dataset, as explained in Kotha *et al*. (2018). We evaluate the sample semivariogram through equation (eq. 10) and we fit the values using the weighted least-squares approach (section 5.2) considering an exponential model (eq. 11). We follow the approach of Esposito and Iervolino (2011, 2012) to pool data from different events. We repeat these steps including the site-specific correction term, so that the corrected IM is:

$$\log_{10} \bar{Y}_{ij,\mathrm{corr}} = \log_{10} \bar{Y}_{ij} + \delta S2S_i \tag{15}$$

Figure 13 provides an overview of the within-event correlation models obtained considering or not the site-specific correction term. Similarly to Sokolov *et al.* (2010, 2012, 2013), the range decreases if a correction factor is included in the analysis, suggesting that neglecting site conditions might add an artificial correlation due to systematic biases in the residuals (Stafford *et al.*, 2018). Besides, the lower level of correlation found at longer periods (grey curve) requires further analysis, since the local-site effects influence is expected to decrease with increasing period, as demonstrated by Jayaram and Baker (2009) and Infantino *et al.* (2018), among others.

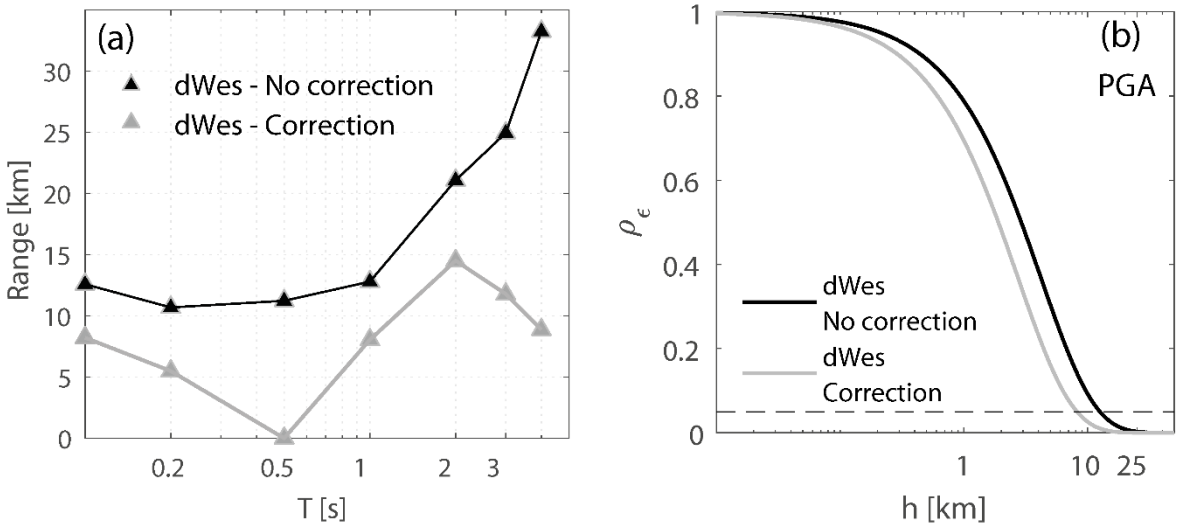

Figure 13: (a) Range as a function of the response-spectral period, T, for within-event residuals ε_{ij} (dWes) with and without the site correction factor; (b) Correlation function in terms of PGA for within-event residuals ε_{ij} with and without the site correction factor. The black dashed line indicates a correlation of 0.05.

Furthermore, the level of correlation of V_{s30} values can be used as a proxy to represent either homogeneous or heterogeneous local-soil conditions in terms of local-site effects, since a smaller correlation indicates a more varied geological setting. Jayaram and Baker (2009) identified a positive relationship between clustering of V_{s30} values and correlation of spectral accelerations, especially at shorter periods. They proposed a model to predict the range as a function of both the period and the V_{s30} range. Similarly, Du and Wang (2013) drew analogous conclusions, illustrating that the influence of local-site effects on the correlation of spectral accelerations drops with increasing period. Sokolov *et al.* (2012) provided a relationship between the PGA range and the V_{s30} correlation, after finding a dependency of these parameters for different regions in Taiwan. Garakaninezhad and Bastami (2017) found that the anisotropy ratio of residuals tends to increase as the anisotropy ratio of V_{s30} increases. However, this positive correlation is significant only for PGA, similarly to the relationship with magnitude. To further emphasis this aspect, **Figure 14** compares the above-mentioned studies along with a model calibrated on the $M_w \geq 5.0$ events of the Central Italy sequence. We develop the latter using the code by Ming *et al.* (2019), which was also used in Huang and Galasso (2019) to study the correlation observed in Italian data. Two main observations are worthy of remark. Firstly, the effects of local-soil conditions on spatial correlation models are much greater at shorter periods, compared

to longer periods where the different models practically overlap. Secondly, the comparison between the Huang and Galasso (2019) and Central Italy models suggests that the correlation strongly depends on the geological characteristics of the considered area, so that a single universal correlation model based on large datasets is not suitable to represent small regions. Indeed, the model of Huang and Galasso (2019) is obtained by including events that occurred in all of the Italian mainland, thus leading to an average spatial correlation model calibrated on diverse geological contexts, with variable spatial dependency of IMs. The same was also previously demonstrated by Sokolov *et al.* (2010, 2012, 2013) for Taiwan. Besides, this behaviour is particularly evident at shorter periods, in agreement with different studies, such as Jayaram and Baker (2009) and Du and Wang (2013), in which these authors demonstrated the impact of local-site effects as a function of period. **Figure 11** illustrates this aspect: the Central Italy and the Huang and Galasso (2019) models nearly coincide for T>1 s.

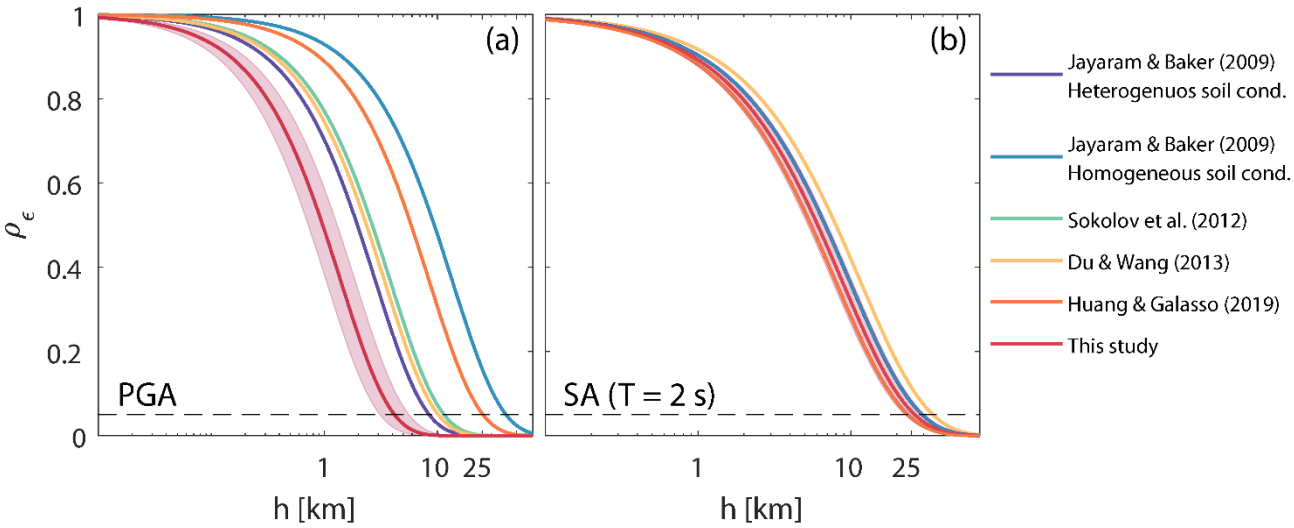

Figure 14: Correlation function in terms of PGA (a) and SA at T = 2 s (b). The model referred as "This study" is computed by means of the code developed by Ming et al. (2019) using data from the nine $M_w \geq 5.0$ central Italy earthquakes. The red shadow area represents the 95% confidence intervals of the central Italy model. The black dashed line indicates a correlation of 0.05.

5.6. Dependence on GMPEs

To develop a correlation model either an existing GMPE or an *ad hoc* relation is used to assess the within-event residuals. Authors, such as Wang and Takada (2005), Jayaram and Baker (2009),

Esposito and Iervolino (2011, 2012) and Wagener *et al.* (2016), employed and existing ground motion model in their analyses. Conversely, other authors such as Goda and Hong (2008), Goda and Atkinson (2009, 2010), Sokolov *et al.* (2010) and Infantino *et al.* (2018) developed a GMPE calibrated on their specific dataset. Du and Wang (2013) asserted that the chosen GMPE does not affect the correlation results. Analogously, Jayaram and Baker (2009) and Goda (2011) demonstrated that similar outcomes are achieved if different GMPEs are used. In contrast, Sokolov *et al.* (2010) maintained that the decomposition of the total aleatory variability into between- and within-event components is influenced by the particular GMPE and hence the spatial correlation is affected in turn, as it is calibrated on the residuals. Infantino *et al.* (2018) built a ground-motion model based on their databases to avoid any dependency on the GMPE. Likewise, we opt for the *ad hoc* GMPE approach and we develop a study-specific ground-motion model, as described in section 5.5. We also calibrate *ad hoc* models for each $M_w \geq 5.0$ event of the Central Italy sequence to obtain a specific correlation model for each earthquake. We assume the following functional form:

$$\log_{10} \bar{Y}_{ij} = b_1 + b_2 \log_{10} \sqrt{R_{JB}^2 + b_4^2} + b_3 \sqrt{R_{JB}^2 + b_4^2} \tag{16}$$

where b_1, b_2, b_3 and b_4 are the model coefficients inferred through a one-stage ordinary regression, which is justified as we are only using data from a single event each time. We show some of the results in **Figure 10** and **Figure 12**.

To investigate the dependence of spatial correlation on the chosen ground-motion model, we repeat all the analyses using the GMPE provided by Bindi et al. (2011), which is also calibrated on Italian data. This model was found by Michelini *et al.* (2019) to be the most suitable to predict ground motions in shallow active crustal regions in Italy. **Figure 15** presents a comparison between the correlation models obtained implementing both the *ad hoc* and Bindi et al. (2011) GMPEs. We plot the results illustrated in section 5.5, in which we consider both uncorrected and corrected within-event residuals. Interestingly, the Bindi *et al.* (2011) model provides nearly the same range as the

model obtained including a site correction term for short-period IMs. At longer periods, the Bindi *et al.* (2011) model approaches the range value inferred without incorporating any site correction factor. We interpret these outcomes by also considering the results that we present in section 5.5. The Bindi *et al.* (2011) GMPE takes into account the local-site effects through the definition of coefficients based on the EC8 soil categories. Local-soil conditions most influence the spatial correlation of short-period IMs and thus, it is not just chance that the correlation model inferred through the Bindi *et al.* (2011) GMPE tends to the range obtained implementing a site-specific correction term. Conversely, the influence of site conditions at longer periods is weaker, and the Bindi *et al.* (2011) and the uncorrected correlation models nearly coincide.

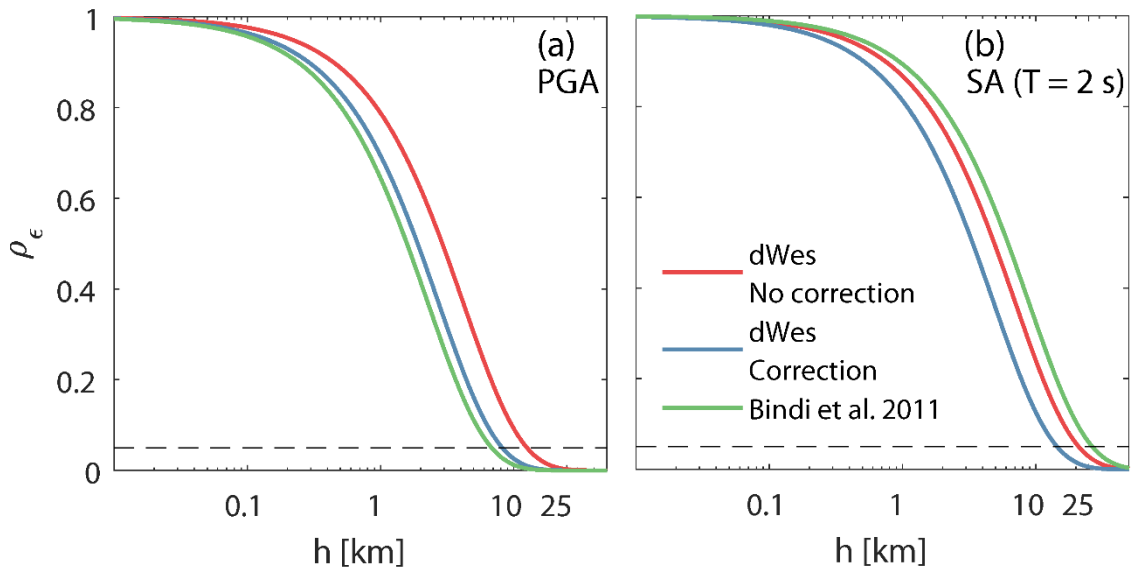

Figure 15: Within-event ε_{ij} (dWes) correlation models for uncorrected and corrected residuals and computed considering the GMPE by Bindi *et al.* (2011): (a) PGA; (b) spectral acceleration for T equals 2 s.

Overall, we believe that any GMPE can be used, provided that the selected ground-motion model fits the data well, so that the residuals do not exhibit biases. Systematic deviations from the predicted median might be mapped as an additional artificial correlation, which inevitably would affect the seismic hazard and risk assessments of spatial distributed systems. This behaviour is shown in **Figure 16**, in which we compare the semivariograms obtained considering both the GMPE by Bindi *et al.* (2011) and Boore and Atkinson (2008). Clearly, the Boore and Atkinson (2008) semivariogram diverges from the Bindi *et al.* (2011) with increasing separation distance. This might be due to the

trend of the within-event residuals with distance found for the Boore and Atkinson (2008) model, which reflects the inability of the GMPE to properly capture the observed IMs.

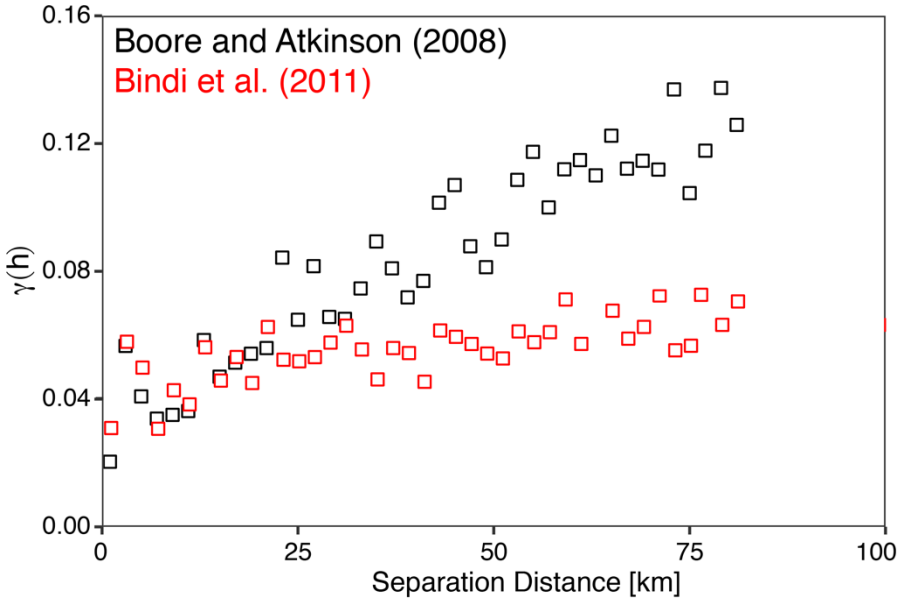

Figure 16: Sample semivariogram obtained using the GMPEs by Boore and Atkinson (2008) and Bindi *et al.* (2011).

6. Conclusions

In this article, we aimed to provide insights into the spatial correlations of earthquake ground-motion intensity measures. A number of studies, such as Park *et al.* (2007) Goda and Atkinson (2009) Esposito and Iervolino (2011) and Verros *et al.* (2017), highlighted the importance of considering such spatial correlations in seismic hazard and risk analysis, especially when the seismic risk of spatially distributed systems has to be evaluated. Nevertheless, the correlation models proposed over the past two decades feature significant discrepancies, which may lead to underestimation or overestimation of the assessed seismic risk. Therefore, in this article, we critically summarised the main findings of previous studies and we attempt to address the primary questions about ground-motion spatial correlation.

We showed the two main approaches to estimate the spatial correlation of ground-motion intensity measures. In principle, the implementation of either one or the other technique is expected to provide similar results, as the semivariogram and the covariance are equivalent for second-order stationary

random fields. Nevertheless, in agreement with Goda and Hong (2008) and Goda and Atkinson (2009), we find that the results slightly diverge. Concerning the computation of the semivariogram, the robust estimator (Cressie) is understood to be less sensitive to atypical observations, as demonstrated by Oliver and Webster (2014). Therefore, we preferred using this estimator in contrast to the classic one throughout the analyses, although the results from our simulations did not find any significant difference between the two estimators. To fit experimental data, we advise prioritising those techniques that weigh the shorter separation distances more highly, since these distances have a strong impact on seismic hazard and risk assessments of spatially-distributed systems. At the same time, we discourage adopting a manual fitting due to the high level of subjectivity involved.

Similarly to Jayaram and Baker (2009), we do not find any relationship between magnitude and correlation distance. We are aware that the analysed dataset is limited in terms of magnitude range (M_w 4.0 to 6.5), so that firm conclusions cannot be drawn on this aspect. However, we believe that magnitude itself cannot explain the large variability of correlation distance values, especially when the same region is considered. Thus, other source effects, such as directivity, azimuth and hanging wall and footwall effects, should be accounted for, as outlined in Stafford *et al.* (2018) and Chen and Baker (2019).

We found a positive correlation between the range and response-spectral period, as expected from the literature. Small-scale heterogeneities in the travel path tend to affect less the long-period waves, which turn out to be more correlated, compared to high-frequency waves.

We analysed the dependency of the spatial correlation on local-soil conditions, illustrating that the influence of local-site effects and regional geology is period-dependent, as demonstrated by several authors, such as Jayaram and Baker (2009), Du and Wang (2013) and Infantino *et al.* (2018). The comparison between our analysis of data from Central Italy and the work proposed by Huang and Galasso (2019) is significant. Firstly, it confirms the hypothesis that the dependency of spatial correlation on site conditions in terms of local-site effects decreases as the period increases. Secondly,

it suggests that a single rate of decay of the correlation as a function of the inter-site separation distance is not suitable for seismic hazard and risk assessment, as the range appears to be regionally-dependent, even though all the data came from the same country (Italy in this case). We believe that region-specific spatial correlation models should be derived.

Finally, our results suggest that any suitable ground-motion model can be applied in correlation studies, provided that the residuals do not exhibit significant trends or biases, which might result in an apparent larger correlation. Therefore, a *single* correlation model could be used in seismic hazard analysis even if *many* different GMPEs are employed to estimate ground-motion parameters. However, as suggested by Sokolov *et al*. (2012), implementing a logic tree for correlation models as already incorporated in probabilistic seismic hazard assessments, should be encouraged to account for the high uncertainty in spatial correlation models.

We note that generally the models are poorly constrained at short separation distances (less than 2 km), due to a shortage of observations (Goda and Atkinson, 2010; Wagener *et al.*, 2016). Results of ground motion simulations could provide better constraints to further advance insights into the spatial correlation of ground motions.

Data and resources

The full dataset of strong-motion waveforms of the nine $M_w \geq 4.0$ earthquakes of the 2016-2017 Central Italy sequences is available at the Engineering Strong Motion database (ESM, http://esm.mi.ingv.it, last accessed September 2019) and the Italian ACcelerometric Archive (ITACA, http://itaca.mi.ingv.it, last accessed September 2019). The full dataset of recordings of the $M_w \geq 2.5$ events of the 2016-2017 Central Italy sequences is available at http://cnt.rm.ingv.it/ (last accessed September 2019).

Acknowledgements

We thank the research group of Dr Francesca Pacor at INGV (Istituto di Geofisica e Vulcanologia) for providing the dataset related to the $M_w \geq 4.0$ events that occurred during the Central Italy earthquake sequence. We also thank Dr Sreeram Reddy Kotha for his helpful suggestions about

implementing mixed-effects regressions in R. We thank Dr Deyu Ming for help with the MATLAB program to implement the approach of Ming et al. (2019). We wish to thank Professor Jack Baker's research group at Stanford University for providing valuable feedback on our analyses. We are grateful to Impact Forecasting for fruitful discussions on hazard footprints. We thank Graeme Weatherill and an anonymous reviewer for their constructive advice on improving the manuscript. Finally, we thank the University of Strathclyde for funding the PhD during which the work was undertaken.

References

Abrahamson, N. A., and Youngs, R. R. (1992). A stable algorithm for regression analyses using the random effects model. *Bulletin of the Seismological Society of America*, 82(1), 505-510.

Akkar, S., and Bommer, J. J. (2010). Empirical equations for the prediction of PGA, PGV, and spectral accelerations in Europe, the Mediterranean region, and the Middle East. *Seismological Research Letters*, 81(2), 195-206.

Annaka, T., Yamazaki, F., and Katahira, F., 1997. Proposal of peak ground velocity and re- sponse spectra based on JMA 87 type accelerometer records, Proceedings, 24th JSCE *Earthquake Engineering Symposium*. Vol. 1, pp. 161–164. Cited in Wang and Takada (2005).

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.

Bindi, D. Pacor, F., Luzi, L., Puglia, R., Massa, M., Ameri, G., Paolucci, R. (2011) 'Ground motion prediction equations derived from the Italian strong motion database', *Bulletin of Earthquake Engineering*, 9(6), pp. 1899–1920. doi: 10.1007/s10518-011-9313-z.

Bommer JJ, Edwards B, Kruiver PP, Rodriguez-Marek A, Stafford PJ, Dost B, Ntinalexis M, Ruigrok E, Spetzler J (2018). V5 ground-motion model for Groningen field. Revision. Cited in Stafford et al. (2018).

Boore, D. M., Gibbs, J. F., Joyner, W. B., Tinsley, J. C., & Ponti, D. J. (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. *Bulletin of the Seismological Society of America*, 93(6), 2737-2751.

Boore, D. M., and G. M. Atkinson (2007). Ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%- damped SA at spectral periods between 0.01 s and 10.0 s, a report of the Next Generation Attenuation Models to the Pacific Earthquake Engineering Research Center, 123 pp.

Boore, D. M., and Atkinson, G. M. (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 10.0 s. *Earthquake Spectra*, *24*(1), 99-138.

Boore, D. M., Stewart, J. P., Seyhan, E., and Atkinson, G. M. (2014). NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes. *Earthquake Spectra*, 30(3), 1057-1085.

Bradley BA. (2010). NZ-specific pseudo-spectral acceleration ground motion prediction equations based on foreign models, report no. 2010-03, Department of Civil and Natural Resources Engineering, University of Canterbury: Christchurch, New Zealand; 324 p

Bradley, B. A. (2013). A New Zealand-specific pseudospectral acceleration ground-motion prediction equation for active shallow crustal earthquakes based on foreign models. *Bulletin of the Seismological Society of America*, 103(3), 1801-1822.

Bradley, B. A. (2014). Site-specific and spatially-distributed ground-motion intensity estimation in the 2010–2011 Canterbury earthquakes. *Soil Dynamics and Earthquake Engineering*, 61, 83-91.

Campbell, K. W., and Bozorgnia, Y. (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 0.01 to 10 s. *Earthquake Spectra*, 24(1), 139-171.

Campbell, K. W., and Bozorgnia, Y. (2010). A ground motion prediction equation for the horizontal component of cumulative absolute velocity (CAV) based on the PEER-NGA strong motion database. *Earthquake Spectra*, 26(3), 635-650.

Campbell, K. W., and Bozorgnia, Y. (2012). A comparison of ground motion prediction equations for Arias intensity and cumulative absolute velocity developed using a consistent database and functional form. *Earthquake Spectra*, 28(3), 931-941.

Campbell, K. W., & Bozorgnia, Y. (2014). NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra. *Earthquake Spectra*, 30(3), 1087-1115.

CEN (2004) EN 1998-1:2004 Eurocode 8: design of structures for earthquake resistance—part 1: general rules, seismic actions and rules for buildings. European Committee for Standardization, Brussels.

Chen, Y. and Baker, J. (2019) 'Spatial Correlations in Cybershake Physics-Based Ground Motion Simulations', *Bulletin of the Seismological Society of America*, In press, pp. 1–17. doi: 10.1785/0120190065.

Chiaraluce, L., Di Stefano, R., Tinti, E., Scognamiglio, L., Michele, M., Casarotti, E., Cattaneo, M., De Gori, P., Chiarabba, C., Monachesi, G., Lombardi, A., Valoroso, L., Latorre, D., Marzorati, S. (2017) 'The 2016 Central Italy Seismic Sequence: A First Look at the Mainshocks, Aftershocks, and Source Models', *Seismological Research Letters*, 88(3), pp. 757–771. doi: 10.1785/0220160221.

Chiou, B. J., & Youngs, R. R. (2008). An NGA model for the average horizontal component of peak ground motion and response spectra. *Earthquake spectra*, 24(1), 173-215.

Cressie, N. (1985), Fitting variogram models by weighted least squares. *Journal of the international Association for mathematical Geology*, 17(5), 563-586.

Crowley, H., Stafford, P. and Bommer, J. J. (2008) "Can earthquake loss models be validated using field observations". *Journal of Earthquake Engineering*, 12(7): 1078-1104.

Diggle, P., and Ribeiro, P. (2007). Model-based Geostatistics [internet resource] (1st ed. 2007, ed., *Springer series in statistics*).

Douglas, J. and Edwards, B. (2016) 'Recent and future developments in earthquake ground motion estimation', *Earth-Science Reviews*. Elsevier B.V., 160, pp. 203–219. doi: 10.1016/j.earscirev.2016.07.005.

Du, W. and Wang, G. (2013) 'Intra-event spatial correlations for cumulative absolute velocity, arias intensity, and spectral accelerations based on regional site conditions', *Bulletin of the Seismological Society of America*, 103(2 A), pp. 1117–1129. doi: 10.1785/0120120185.

Esposito, S. and Iervolino, I. (2011) 'PGA and PGV spatial correlation models based on European multievent datasets', *Bulletin of the Seismological Society of America*, 101(5), pp. 2532–2541. doi: 10.1785/0120110117.

Esposito, S. and Iervolino, I. (2012) 'Spatial correlation of spectral acceleration in European data', *Bulletin of the Seismological Society of America*, 102(6), pp. 2781–2788. doi: 10.1785/0120120068.

Foulser-Piggott, R. and Goda, K. (2015) 'Ground-motion prediction models for arias intensity and cumulative absolute velocity for Japanese earthquakes considering single- station sigma and within-event spatial correlation', *Bulletin of the Seismological Society of America*, 105(4), pp. 1903–1918. doi: 10.1785/0120140316.

Garakaninezhad, A., and Bastami, M. (2017). A Novel Spatial Correlation Model Based on Anisotropy of Earthquake Ground-Motion Intensity. *Bulletin of the Seismological Society of America*, 107(6), 2809-2820.

Goda, K. and Hong, H. P. (2008) Spatial correlation of peak ground motions and response spectra, *Bulletin of the Seismological Society of America*, 98(1), pp. 354–365. doi: 10.1785/0120070078.

Goda, K. and Atkinson, G. M. (2009) 'Probabilistic characterization of spatially correlated response spectra for earthquakes in Japan', *Bulletin of the Seismological Society of America*, 99(5), pp. 3003–3020. doi: 10.1785/0120090007.

Goda, K. and Atkinson, G. M. (2010) 'Intraevent spatial correlation of ground-motion parameters using SK-net data', *Bulletin of the Seismological Society of America*, 100(6), pp. 3055–3067. doi: 10.1785/0120100031.

Goda, K. (2011) 'Interevent variability of spatial correlation of peak ground motions and response spectra', *Bulletin of the Seismological Society of America*, 101(5), pp. 2522–2531. doi: 10.1785/0120110092.

Heresi, P. and Miranda, E. (2019) 'Uncertainty in intraevent spatial correlation of elastic pseudo-acceleration spectral ordinates', *Bulletin of Earthquake Engineering*. Springer Netherlands, 17(3), pp. 1099–1115. doi: 10.1007/s10518-018-0506-6.

Hong, H. P., Zhang, Y. and Goda, K. (2009) 'Effect of spatial correlation on estimated ground-motion prediction equations', *Bulletin of the Seismological Society of America*, 99(2 A), pp. 928–934. doi: 10.1785/0120080172.

Hough, S. E. and Field, E. H. (1996) 'On the coherence of ground motion in the San Fernando valley', *Bulletin of the Seismological Society of America*, 86(6), pp. 1724–1732.

Huang, C. and Galasso, C. (2019) 'Ground-motion intensity measure correlations observed in Italian data', *13th International Conference on Applications of Statistical and Probability in Civil Engineering*.

Iervolino I (2013). Probabilities and fallacies: why hazard maps cannot be validated by individual earthquakes. Earthquake Spectra; 29(3): 1125-1136, https://doi.org/10.1193/1.4000152.

Infantino, M., Paolucci, R., Smerzini, C., and Stupazzini, M. (2018). Study of the spatial correlation of earthquake ground motion by means of physics-based numerical scenarios. In *16th European Conference on Earthquake Engineering*.

Jayaram, N. and Baker, J. W. (2009) 'Correlation model for spatially distributed ground-motion intensities', *Earthquake Engineering and Structural Dynamics*, 38. doi: 10.1002/eqe.

Jayaram, N. and Baker, J. W. (2010) 'Considering spatial correlation in mixed-effects regression and the impact on ground-motion models', *Bulletin of the Seismological Society of America*, 100(6), pp. 3295–3303. doi: 10.1785/0120090366.

Joyner, W. B., and Boore, D. M. (1993). Methods for regression analysis of strong-motion data. *Bulletin of the Seismological Society of America*, *83*(2), 469-487.

Kanno, T., Narita, A., Morikawa, N., Fujiwara, H., and Fukushima, Y. (2006). A new attenuation relation for strong ground motion in Japan based on recorded data. *Bulletin of the Seismological Society of America*, 96(3), 879-897.

Kotha, S. R., Cotton, F. and Bindi, D. (2018) 'A new approach to site classification: Mixed-effects Ground Motion Prediction Equation with spectral clustering of site amplification functions', *Soil Dynamics and Earthquake Engineering*, 110(July 2017), pp. 318–329. doi: 10.1016/j.soildyn.2018.01.051.

Lanzano, G., D'Amico, M., Felicetta, C., Puglia, R., Luzi, L., Pacor, F., & Bindi, D. (2016). Ground-motion prediction equations for region-specific probabilistic seismic-hazard analysis. Bulletin of the *Seismological Society of America*, 106(1), 73-92.

Lanzano G., Puglia R., Russo E., Luzi L., Bindi D., Cotton F., D'Amico M., Felicetta C., Pacor F. and ORFEUS WG5 (2018). *ESM strong-motion flat-file 2018*. Istituto Nazionale di Geofisica e Vulcanologia (INGV), Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum (GFZ), Observatories & Research Facilities for European Seismology (ORFEUS). PID: 11099/ESM_flatfile_2018

Lanzano, G., Luzi, L., Pacor, F., Felicetta, C., Puglia, R., Sgobba, S., and D'Amico, M. (2019). A Revised Ground-Motion Prediction Model for Shallow Crustal Earthquakes in Italy. *Bulletin of the Seismological Society of America*, *109*(2), 525-540.

Luzi, L., Pacor, F., Puglia, R., Lanzano, G., Felicetta, C., D'Amico, M., Michelini, A., Faenza, L., Lauciani, V., Iervolino, I., Baltzopoulos, G., Chioccarelli, E. (2017). The central Italy seismic sequence between August and December 2016: Analysis of strong-motion observations. *Seismological Research Letters*, *88*(5), 1219-1231. doi: 10.1785/0220170037.

Matheron, G., (1962). Traité de géostatistique appliquée. 1 (1962). Editions Technip. Cited in Esposito and Iervolino (2011).

Michelini, A., Faenza, L., Lanzano, G., Lauciani, V., Jozinović, D., Puglia, R., and Luzi, L. (2019). The New ShakeMap in Italy: Progress and Advances in the Last 10 Yr. *Seismological Research Letters*. doi: 10.1785/0220190130.

Ming, D., Huang, C., Peters, G. W., and Galasso, C. (2019). An advanced estimation algorithm for ground-motion models with spatial correlation. *Bulletin of the Seismological Society of America*, 109(2), 541-566. doi: 10.1785/0120180215.

Midorikawa, S., and Ohtake, Y., 2002. Attenuation relationships of peak ground acceleration and velocity considering attenuation characteristics for shallow and deeper earthquakes, *Proceedings, 11th Japan Earthquake Engineering Symposium*, Vol. 1, pp. 609–614. Cited in Wang and Takada (2005).

Oliver, M. A. and Webster, R. (2014) 'A tutorial guide to geostatistics: Computing and modelling variograms and kriging', *Catena*. Elsevier B.V., 113, pp. 56–69. doi: 10.1016/j.catena.2013.09.006.

Park, J., Bazzurro, P. and Baker, J. W. (2007) 'Modeling spatial correlation of ground motion Intensity Measures for regional seismic hazard and portfolio loss estimation', *Applications of Statistics and Probability in Civil Engineering*, pp. 1–8.

Sgobba, S., Lanzano, G., Pacor, F., Puglia, R., D'Amico, M., Felicetta, C., & Luzi, L. (2019). Spatial Correlation Model of Systematic Site and Path Effects for Ground-Motion Fields in Northern Italy. *Bulletin of the Seismological Society of America*. doi: 10.1785/0120180209.

Sokolov, V. and Wenzel, F. (2011) 'Influence of spatial correlation of strong ground motion on uncertainty in earthquake loss estimation', *Earthquake Engineering & Structural Dynamics*, 40, pp. 993–1009. doi: 10.1002/eqe.1074.

Sokolov, V., Wenzel, F., & Kuo-Liang, W. (2010). Uncertainty and spatial correlation of earthquake ground motion in Taiwan. *TAO: Terrestrial, Atmospheric and Oceanic Sciences*, *21*(6), 9. doi: 10.3319/TAO.2010.05.03.01(T).

Sokolov, V., Wenzel, F., Wen, K. L., & Jean, W. Y. (2012). On the influence of site conditions and earthquake magnitude on ground-motion within-earthquake correlation: analysis of PGA data from TSMIP (Taiwan) network. *Bulletin of Earthquake Engineering*, *10*(5), 1401-1429. doi: DOI 10.1007/s10518-012-9368-5.

Sokolov, V. and Wenzel, F. (2013) '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', *Bulletin of Earthquake Engineering*, 11(6), pp. 1909–1926. doi: 10.1007/s10518-013-9493-9.

Sokolov, V. and Ismail-Zadeh, A. (2016). On the use of multiple-site estimations in probabilistic seismic-hazard assessment. Bull. Seismol. Soc. Am. 106(5): 2233–2243, doi: 10.1785/0120150306

Sokolov, V. and Wenzel, F. (2019) Areal exceedance of ground motion as a characteristic of multiple-site seismic hazard: Sensitivity analysis. Soil Dynamics and Earthquake Engineering, 126, 105770, https://doi.org/10.1016/j.soildyn.2019.105770.

Stafford, P. J. (2012). Evaluation of structural performance in the immediate aftermath of an earthquake: a case study of the 2011 Christchurch earthquake. *International Journal of Forensic Engineering*, 1(1), 58-77.

Stafford, P. J. *et al.* (2018) 'Extensions to the Groningen ground-motion model for seismic risk calculations: component-to-component variability and spatial correlation', *Bulletin of Earthquake Engineering*. Springer Netherlands, 17(8), pp. 4417–4439. doi: 10.1007/s10518-018-0425-6.

Strasser, F. O. and Bommer, J. J. (2009) 'Strong ground motions-have we seen the worst?', *Bulletin of the Seismological Society of America*, pp. 2613–2637. doi: 10.1785/0120080300.

Tsai, C. C. P., Chen, Y. H., and Liu, C. H. (2006). The path effect in ground-motion variability: an application of the variance-components technique. *Bulletin of the Seismological Society of America*, 96(3), 1170-1176.

Verros, S. A., Wald, D. J., Worden, C. B., Hearne, M., & Ganesh, M. (2017). Computing spatial correlation of ground motion intensities for ShakeMap. *Computers & geosciences*, *99*, 145-154. doi: 10.1016/j.cageo.2016.11.004.

Wagener, T., Goda, K., Erdik, M., Daniell, J., & Wenzel, F. (2016). A spatial correlation model of peak ground acceleration and response spectra based on data of the Istanbul earthquake rapid response and early warning system. *Soil Dynamics and Earthquake Engineering*, *85*, 166-178. doi: 10.1016/j.soildyn.2016.03.016.

Wang, M. and Takada, T. (2005) 'Macrospatial correlation model of seismic ground motions', *Earthquake Spectra*, 21(4), pp. 1137–1156. doi: 10.1193/1.2083887.

Weatherill, G. A. *et al.* (2015) 'Exploring the impact of spatial correlations and uncertainties for portfolio analysis in probabilistic seismic loss estimation', *Bulletin of Earthquake Engineering*, 13(4), pp. 957–981. doi: 10.1007/s10518-015-9730-5.

Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists*. John Wiley & Sons.

Zerva, A. and Zervas, V. (2002) 'Spatial variation of seismic ground motions: An overview', *Applied Mechanics Reviews*, 55(3), p. 271. doi: 10.1115/1.1458013.