



# Exploring the impact of spatial correlations of earthquake ground motions in the catastrophe modelling process: a case study for Italy

E. Schiappapietra<sup>1</sup> · S. Stripajová<sup>2</sup> · P. Pažák<sup>2</sup> · J. Douglas<sup>1</sup> · G. Trendafiloski<sup>3</sup>

Received: 5 October 2021 / Accepted: 25 April 2022 / Published online: 30 May 2022  
© The Author(s) 2022

## Abstract

Catastrophe models are important tools to provide proper assessment and financial management of earthquake-related emergencies, which still create the largest protection gap across all perils. Earthquake catastrophe models include three main components, namely: (1) the earthquake hazard model, (2) the exposure model and, (3) the vulnerability model. Simulating spatially distributed ground-motion fields within either deterministic or probabilistic seismic hazard assessments poses a major challenge when site-related financial protection products are required. In this framework, we develop ad hoc correlation models for different Italian regions (specifically northern, central and southern Italy) and thereafter we perform both deterministic scenario-based and probabilistic event-based hazard and risk assessments in order to advance the understanding of spatial correlations within the catastrophe modelling process. We employ the OpenQuake engine for our calculations. This is an open-source tool suitable for accounting for the spatial correlation of earthquake ground-motion residuals. Our outcomes, albeit preliminary, demonstrate the importance of considering not only the spatial correlation of ground motions, but also its associated uncertainty in risk analyses. Although loss exceedance probability curves for the return periods of interest for the (re)insurance industry show similar trends, both hazard and risk footprints in terms of average annual losses feature less noisy and more realistic patterns if spatial correlation is taken into account. Such results will have implications for (re)insurance companies evaluating the risk to high-value civil engineering infrastructures.

**Keywords** Spatial correlation · Seismic risk assessment · OpenQuake engine · Catastrophe model · Loss assessment

---

✉ E. Schiappapietra  
erika.schiappapietra@strath.ac.uk

<sup>1</sup> Department of Civil and Environmental Engineering, University of Strathclyde, Glasgow G1 1XJ, UK

<sup>2</sup> Aon Impact Forecasting, Prague 110 00, Czech Republic

<sup>3</sup> Aon Impact Forecasting, London EC3V 4AN, UK

## 1 Introduction

Catastrophe models are important tools to provide proper estimations and financial management of earthquake-related emergencies, which still have the largest protection gap across all perils. The mitigation of socio-economic losses and the development of insurance and reinsurance strategies are, therefore, central targets in the assessment of the seismic risk at urban and regional scales. Earthquake catastrophe models include three main components: hazard, vulnerability and exposure (Grossi and Kunreuther 2005). The earthquake hazard requires a proper description of the earthquake ground motion (intensity measures—IMs) and its spatial variability across the region using deterministic or probabilistic earthquake hazard models. In particular, three main issues need to be identified to achieve reliable estimates of local intensity and therefore assess catastrophe loss: (1) locations of potential future events, (2) their frequency of occurrence, and (3) their severity. The vulnerability component estimates the probability that a structure's damage will exceed various levels as a result of ground motion. The vulnerability component usually defines loss ratios as a function of IM values depending on material of construction, seismic design, building height, occupancy and other modifiers. The exposure component provides geocoding of the analysed assets, i.e. (re)insurance portfolio, including asset's vulnerability modifiers, total insured value and policy conditions.

Simulating spatially-distributed ground-motion fields poses a major challenge when site-related financial protection products are required. Indeed, the loss assessment of spatially-distributed systems requires the consideration of: (1) the cross-correlation among different IMs at the same site, (2) the spatial correlation of the same IM at different sites, and (3) the spatial cross-correlation among different IMs at closely-spaced sites (Weatherill et al. 2015a, b). Several authors have demonstrated the importance of including correlation models in risk analyses. For example, Park et al. (2007) and Sokolov and Wenzel (2011) evidenced that neglecting the spatial correlation may cause a bias in loss estimates, whose magnitude depends on the considered portfolio. Similarly, Weatherill et al. (2015a, b) proved that the effects of including spatial cross-correlations is limited when larger footprints are considered or a building typology is predominant in the portfolio. While investigating the losses in a district of Istanbul as a consequence of a Mw 7.2 earthquake scenario, Wagener et al. (2016) found that the loss distribution changes clearly depend on the correlation model in contrast to the mean loss, which is not affected. Wesson and Perkins (2001) also demonstrated that the spatial correlation is crucial for explaining the variance of losses to a spatially-distributed system, but not the mean loss. Costa et al. (2018) evaluated the consequences of seismic events on transportation networks and observed that the probability of disruption is higher when the spatial correlation is neglected.

It is well-established that IMs at multiple locations during the same earthquake are spatially correlated and their degree of correlation tends to be higher for closely-spaced sites and for low-frequency ground motions. In addition, several authors (Chen and Baker 2019; Schiappapietra and Douglas 2020; Infantino et al. 2021; Schiappapietra and Smerzini 2021) have demonstrated that the spatial correlation of earthquake ground-motion is period-, regionally- and scenario-dependent, so that the implementation of a unique correlation model calibrated on worldwide databases may represent an oversimplification. Besides, spatial correlation models are usually inferred from a set of multiple events due to the limited number of ground-motion recordings from a single earthquake. The event-to-event correlation variability should be therefore included

in regional probabilistic risk analysis to enable more realistic estimations of both the ground motion and corresponding losses (Heresi and Miranda 2019).

In this context, we firstly calibrate different correlation models based on observations of the Mw 6.5 2016 Norcia (central Italy) earthquake. These models are used in deterministic scenario-based calculations for the same event. The primary aim of such analyses is to better investigate to what extent different input models may affect the overall economic losses. Thereafter, we perform event-based probabilistic seismic hazard assessment (PSHA) calculations in order to advance the understanding of spatial correlations within the catastrophe modelling process. Indeed, the probabilistic seismic risk assessment is usually required for decision-making and insurance/reinsurance purposes (Erdik 2017; Kohrangi et al. 2021). To account for the event-to-event correlation variability, we follow the methodology proposed by Heresi and Miranda (2019), based on the median range value and its associated dispersion. We employ the OpenQuake engine (Silva et al. 2012; Pagani et al. 2014) for our calculations, which is an open-source seismic hazard and risk modelling tool that can account for the spatial correlation of earthquake ground-motion residuals. In particular, we develop custom spatial correlation models, and we fine-tune the already implemented approach to simulate spatially-correlated random fields to significantly reduce the computational cost and to enable whole country event-based PSHA calculations.

The goals of this study are: (1) to illustrate the impact of spatially dependent modelling on the resulting earthquake shaking losses of building portfolios or spatially-distributed infrastructures and, (2) to investigate the effects on per-event and in-location loss estimates for underwriting purposes. The results of this project, albeit only illustrative, will have implications for (re)insurance companies evaluating the risk to high-value civil engineering infrastructures.

This article is organized as follows. Section 2 describes the characterization of the spatial correlation of ground motion IMs, with particular emphasis on the custom correlation models derived for the regions of interest. Section 3 details the parameters and models required as input to the OpenQuake engine. We mainly focus on the seismic hazard assessment aspect, and only a brief description of the vulnerability and exposure components is given. Finally, Sect. 4 examines the outcomes of the risk analyses with the aim of illustrating the effects of spatial correlations within the catastrophe modelling process. The article ends with some brief conclusions.

## 2 Modelling spatial correlation of ground motion intensity measures

Empirical ground motion models (GMMs) provide an estimate of the ground shaking at an individual site as a function of explanatory variables describing the source, the propagation path and the site conditions (e.g. Douglas and Edwards 2016):

$$\log \text{IM}_{ei} = \mu_{\log \text{IM}_{ei}} + \delta B_e + \delta W_{ei} \quad (1)$$

where  $\text{IM}_{ei}$  is the intensity measure of interest at site  $i$  due to the event  $e$ , whereas  $\mu_{\log \text{IM}_{ei}}$  is the predicted mean of  $\log \text{IM}_{ei}$ .  $\delta B_e$  and  $\delta W_{ei}$  are the between-event and within-event residuals, respectively, which are assumed to be independent and normally distributed with zero mean and standard deviation  $\tau$  and  $\varphi$ , respectively.  $\delta B_e$  represents the average shift of the observed IM of an individual event  $e$  from the median predicted by the model and it is common for all sites, whereas the various  $\delta W_{ei}$  account for site-to-site differences of observations from the median event-corrected predictions. GMMs express the IMs of interest as

lognormally distributed random variables, providing only their marginal probability distribution at an individual site. However, for regional seismic hazard and risk assessments, one needs to quantify the joint probability of occurrence of IMs at multiple locations over the region of interest. Therefore, it is necessary to model the spatial variability of  $\delta W_{ei}$ , namely to define how the  $\delta W_{ei}$  vary in space.

## 2.1 Characterization of the spatial variability of univariate random functions

In geostatistical analysis, the most common tool adopted to describe the spatial correlation of a random function is the empirical semivariogram, which measures the average dissimilarity of a pair of random variables  $z(x_i)$  and  $z(x_j)$  at locations  $x_i$  and  $x_j$ , separated by an inter-site distance  $h$ :

$$\gamma = \frac{1}{2}E[(z_{x_i} - z_{x_j})^2] \quad (2)$$

where  $E[\ ]$  denotes the expected value. In seismological applications, the  $z(x_i)$  represent the (normalized) within-event residuals:

$$z(x_i) = \frac{\delta W_{ei}}{\sigma_e} = \frac{\log \text{IM}_{ei} - \mu_{\log \text{IM}_{ei}} - \delta B_e}{\sigma_e} \quad (3)$$

in which  $\sigma_e$  is the sample standard deviation of the within-event residuals for earthquake  $e$ .

Due to the lack of repeated observations of ground motions at each site from a given earthquake, further simplifications are needed to estimate  $\gamma$ , and the hypothesis of second-order stationarity and isotropy are commonly assumed. This means that the correlation between any pairs of sites with equal separation distance is the same, independently of the source-to-site distance and orientation. Therefore, under these assumptions, Eq. (2) can be rewritten as:

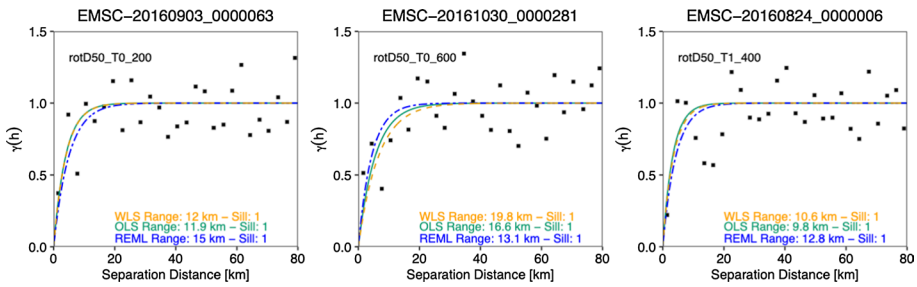
$$\gamma = \frac{1}{2}E[(z_{x_i} - z_{x_i+h})^2] \quad (4)$$

where  $h$  is the inter-site distance.

To assess the spatial correlation structure on the residuals with semivariograms, the main steps are: (1) compute the experimental semivariogram  $\hat{\gamma}(h)$ ; (2) choose a parametric function (e.g. exponential and spherical models); and (3) estimate the correlation parameters, namely the sill (i.e. the overall variance) and the practical range (i.e. the distance beyond which the correlation is negligible) through regression analysis. In this study, the experimental semivariogram is computed based on the robust estimator proposed by Cresie (1985) and the exponential function is used to model the sample semivariogram, as this is the most commonly-adopted function because it often provides the best fit to the data:

$$\gamma(h) = a \left[ 1 - \exp\left(-\frac{3h}{b}\right) \right] \quad (5)$$

where  $a$  represents the sill and  $b$  is the practical range. In the case of the exponential model,  $b$  is the distance at which 95% of the sill is reached. The sill is equal to 1 because of the normalization of the within-event residuals and therefore, the range is the only correlation parameter to be estimated. We opt for a weighted least-squares regression to fit the sample semivariogram, in which the weights depend both on the number of pairs and the separation distance, following Baker and Chen (2020):



**Fig. 1** Experimental semivariograms and corresponding semivariogram models for three different events and IMs. WLS stands for weighted least squares regression, in which the weights are computed by using Eq. 6; OLS stands for ordinary least squares regression; REML stands for the maximum-likelihood approach. WLS is the preferred method used in this study

$$w_i = N(h_i) \exp\left(-\frac{h_i}{c}\right) \tag{6}$$

in which  $N$  is the number of pairs in each bin and  $c$  is a parameter that describes the decay of the weights as the distances increases. We set it equal to 5 km, based on preliminary analysis. We note that correlation parameters were also estimated through maximum-likelihood approaches in preliminary analyses. In general, the results were similar in most of the cases and therefore we favoured the method that provided a better fit and is more common.

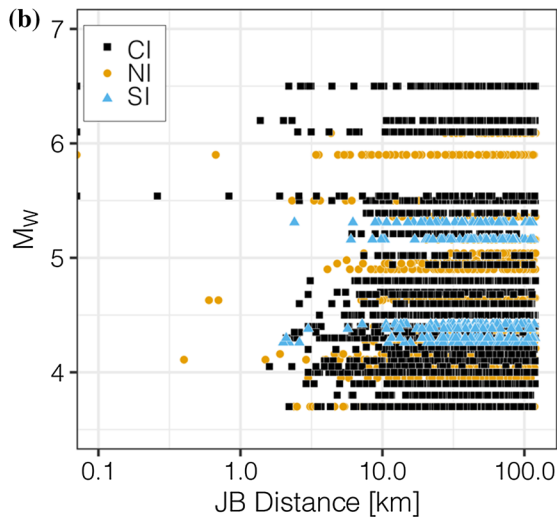
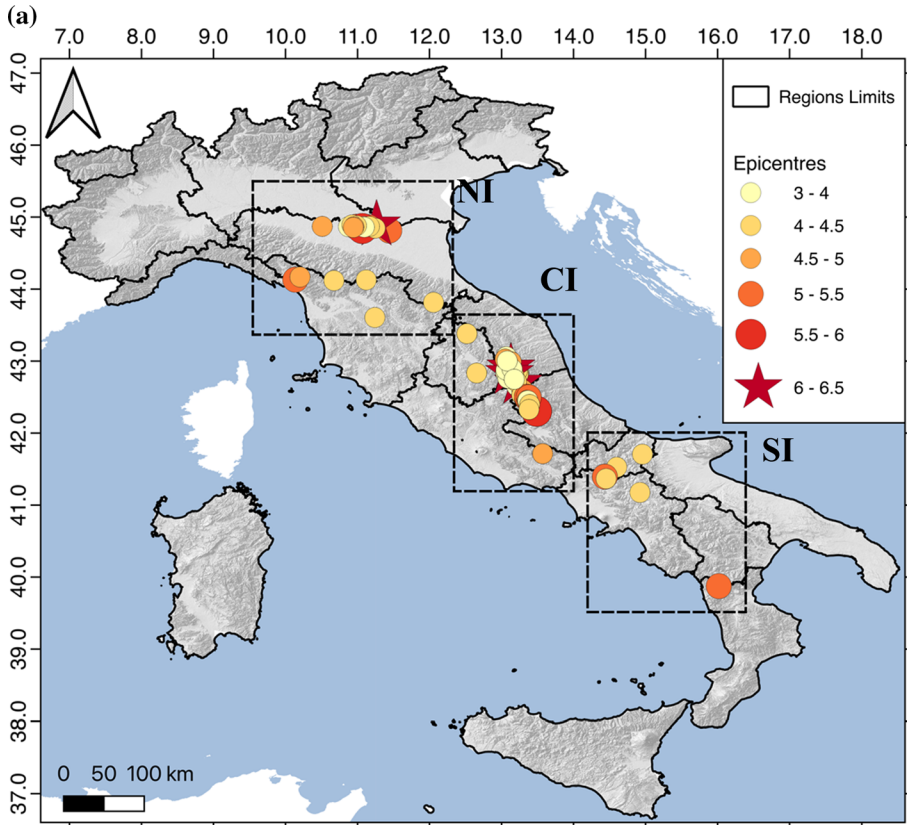
Figure 1 shows examples of experimental semivariograms and the corresponding fitted models obtained by using three different regression approaches. The weights for the weighted least squares regression are computed through Eq. 6. As can be observed, the exponential function generally fits the empirical data well.

The reader should refer to Schiappapietra and Douglas (2020) and Schiappapietra and Douglas (2021) for further details on correlation modelling, including on the choice of the exponential function.

**2.2 Databases**

We employ the European Strong-Motion (ESM) dataset (Lanzano et al. 2019) to develop spatial correlation models for Italy. In particular, we apply the selection criteria proposed by Kotha et al. (2020) to select only shallow crustal earthquakes and records with an usable frequency range. Only those events with  $\geq 40$  records (within a Joyner-Boore distance  $R_{JB} \leq 120$  km) in the dataset are used to guarantee more robust correlation estimates. We further subdivide the dataset into three sub-categories based on macro-regions: (1) Northern, (2) Central, and (3) Southern Italy to calibrate ad hoc correlation models tailored to the specific macro-region. Figure 2a summarises the characteristics of the three sub-datasets, whereas Fig. 2b shows the magnitude-distance scatter plot. We note that the boundaries of these macro-regions are slightly arbitrary and rectangles are used for convenience. The boundaries of the macro-regions are not based on clear topographic, geological or cultural differences. The macro-regions encompass the main concentrations of available data with no overlap.

The Northern Italy dataset includes 23 earthquakes (1408 records) that mainly occurred during the 2012 Emilia seismic sequence (in the Po Plain) with characteristic moment



**Fig. 2** Set of Italian data: **a** Location of the epicentres as a function of  $M_w$ ; and **b** magnitude-distance scatter plot. JB stands for Joyner-Boore distance. The rectangles (in dashed black) indicate the three sub-regions selected in this study: NI: Northern Italy; CI: Central Italy; and SI: Southern Italy

magnitudes of 4.0–5.0. All the events are recorded on average by 61 stations, which mainly belong to the soil class C (according to the EC8 soil classes, CEN (2004)), i.e. soft soil characterized by a  $V_{s30}$  (average shear wave velocity of the top 30 m of the soil column) between 180 and 360 m/s. The prevailing style of faulting is thrust (TF) and only a few events have normal (NF) or strike slip (SS) mechanisms.

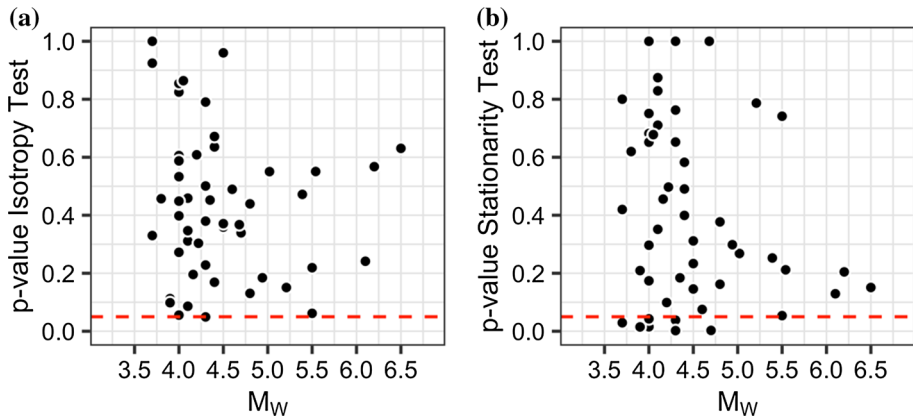
The Central Italy dataset includes 4363 records from 50 events characterized mainly by normal-fault mechanisms (48 NF and 2 SS) in the central Apennines. The majority of data belongs to either the 2009 L'Aquila or the 2016 Central Italy sequences, and are from fault distances smaller than 70 km, as well as moment magnitudes in the range 4.0–5.0. The largest earthquake is the Mw 6.5 30th October 2016 Norcia event. All the earthquakes are recorded on average by nearly 90 stations, which are mainly classified as soil class B, i.e. stiff soil with a  $V_{s30}$  in the range 360–800 m/s.

Finally, the Southern Italy dataset contains only 333 records from six events in the southern Apennines. Four out of the six events have a magnitude of 4.5 and a NF mechanism. Each event is recorded on average by 55 stations, which mainly belong to soil class B. We are aware that the Southern Italy correlation model is not well constrained due to the shortage of data in this area. However, we prefer to distinguish the three macro-regions as correlation models are found to be period- and regionally- dependent (e.g. Schiappapietra and Douglas 2020).

### 2.3 Stationarity and isotropy assumptions

We use the test by Bowman and Crujeiras (2013) to verify the suitability of a stationary and isotropic model to represent the spatial correlation. The validity of the stationary assumption is determined through a test statistic which compares an estimated semivariogram based on separation distance and location with that obtained as a function of only separation distance. Analogously, the test of isotropy compares an estimated semivariogram based on separation distance and azimuth with that obtained as a function of only separation distance. If the  $P$ -value of the statistical tests is greater than 0.05, the evidence of non-stationarity and anisotropy is not statistically significant (at a 5% significance level, a commonly chosen level for hypothesis testing) and therefore the assumption of stationarity and isotropy cannot be rejected.

In this study, we test the validity of such assumptions by using the *sm* package (Bowman and Azzalini 2018) in the R software (R Core Team 2019). We perform the test for each event in the database and IM considered (Peak Ground Acceleration, PGA; Spectral Acceleration, SA, at 15 periods between 0.1 and 2 s). Figure 3 shows the  $P$ -values of the statistical tests as a function of Mw considering the events within the Central Italy database and for SA for a period of 1 s. The majority of events have a  $P$ -value larger than 0.05, suggesting that the hypotheses of isotropy and stationarity are satisfied at a 5% significance level. It is noted that the  $P$ -values for the stationarity test for the three largest events are between 0.1 and 0.2, which provides some evidence for non-stationarity. This could result from non-stationarity behaviour in the spatial correlation of path effects. Similar results are obtained for the other IMs and events. These figures are not shown here for the sake of brevity. In the following, we, therefore, describe the spatial correlation for the different regions through stationary and isotropic models. We note that models cannot be extrapolated to other IMs not considered in this study as we cannot guarantee the validity of the isotropy and stationarity assumptions.



**Fig. 3** *P*-value as a function of magnitude  $M_w$ : **a** isotropy test; and **b** stationarity test. Red dashed line indicates the 5% significance level. The results are for the Central Italy dataset and for IM=SA ( $T = 1$  s)

## 2.4 Correlation models for the $M_w$ 6.5 Norcia (Central Italy) event

Here, we propose three different spatial correlation models for the  $M_w$  6.5 Norcia event obtained through the methodology explained in Sect. 2.1. The models differ only by the reference GMM used to compute the within-event residuals. This makes it possible to investigate the effect of the GMM on the spatial correlation and hence on the resulting earthquake shaking losses of building portfolios. We choose the Lanzano et al. (2019) (hereinafter ITA18) and the Kotha et al. (2020) (hereinafter K20) models, which mainly differ in terms of their underlying datasets. The former is calibrated on Italian data whereas the latter includes also European events. We also develop an ad hoc GMM specifically calibrated on the observations of the  $M_w$  6.5 Norcia event (136 records within 120 km and 29 near-source data within 30 km). The latter takes the following form:

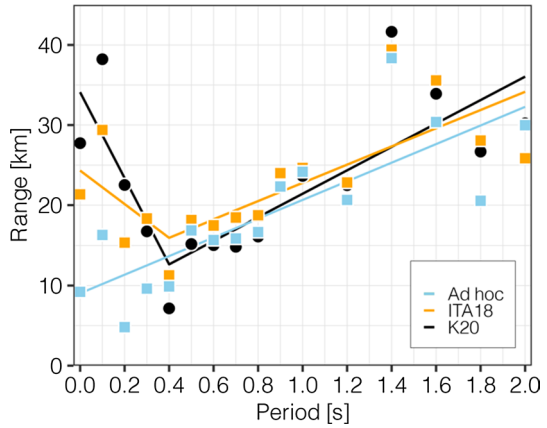
$$\log_{10} Y_{ij} = b_1 + b_2 \log_{10} \sqrt{R^2 + b_3^2} + b_4 \sqrt{R^2 + b_3^2} + b_5 \log_{10} \left( \frac{V_{s30}}{800} \right) \quad (7)$$

where  $R$  is the  $R_{JB}$  distance and  $b_1, \dots, b_5$  are the model coefficients inferred through a one-stage ordinary regression, which is justified as here we are only using data from a single event.

Figure 4 presents the ranges obtained for PGA ( $T = 0$  s) and SA at 15 periods between 0.1 and 2 s and the corresponding fitted models as a function of period. Generally, the three different models have a range that is directly proportional to the period, as previously observed in the literature (e.g. Schiappapietra and Douglas 2020). At intermediate and long periods, the models converge towards very similar values, suggesting that the underlying GMM has a negligible effect on the final outcomes. By contrast, at short periods ( $T \leq 0.4$  s), the three models show different trends. An explanation may lie with the different sensitivity of high-frequency and low-frequency ground motions to the anelastic attenuation. Indeed, Kotha et al. (2020) found larger regional differences of anelastic attenuation at short periods than at longer periods. We believe that the larger ranges of the correlation model based on K20 are due to the faster attenuation of the Central Italy region with respect to the pan-European average (Kotha et al. 2020). Such strong attenuation is not modelled in the ergodic K20 GMM and therefore it manifests



**Fig. 4** Spatial correlation models obtained for the Mw 6.5 Norcia (Central Italy) event. Dots and squares represent the ranges as a function of period, whereas the solid lines represent the fitted models. Ad hoc refers to the GMM calibrated on the Mw 6.5 Norcia observations only; ITA18 refers to the Lanzano et al. (2019) GMM; K20 refers to the Kotha et al. (2020) GMM



**Table 1** Coefficients of the equations fitted to the ranges for the three GMMs (Eq. 8)

	$a_0$	$a_1$	$a_2$	$t$
ad hoc	9.03	11.62		
ITA18	15.96	-20.89	11.37	0.40
K20	12.65	-53.62	14.62	0.40

as apparent spatial correlation. Conversely, the ITA18 and the ad hoc GMMs better capture the attenuation of high-frequency IMs (e.g. term  $b_4\sqrt{R^2 + b_3^2}$  in Eq. 7), resulting in a lower spatial correlation at short periods.

The observed ranges are fitted either with a bilinear (ITA18 and K20) or linear (ad hoc GMM) expression to facilitate its implementation in the OpenQuake software, as follow:

$$b(T) = \begin{cases} a_0 + a_1(T - t), & T \leq t \\ a_0 + a_2(T - t), & T > t \end{cases}, \text{ or} \quad (8)$$

$$a_0 + a_1 T$$

where  $T$  is the period of interest and  $b$  is the range in km. The coefficients  $a_0, a_1, a_2$  and  $t$  obtained through a one-stage regression are reported in Table 1 for the three different cases. We recall that the model cannot be extrapolated to other periods as we cannot guarantee that the hypothesis of isotropy and stationarity holds.

### 2.5 Correlation models for the three macro-regions in Italy

It is recognized that the spatial correlation not only varies with geological context (e.g. Sokolov and Wenzel 2013; Schiappapietra and Douglas 2020), but also from event to event (Goda 2011; Heresi and Miranda 2019). In this study, we propose different correlation models depending on the considered region (Northern, Central and Southern Italy) and we provide an estimate of the event-to-event variability that should be accounted for to obtain more informed regional risk assessments. We follow Heresi and

Miranda (2019), who proposed a simple methodology to include the event-to-event variability of spatial correlation by providing the median range and its associated dispersion. In particular, the main steps are:

- (1) compute the spatial correlation for each event in the database and for each considered IM, following the methodology described in Sect. 2.1. We use the ITA18 GMM to compute the within-event-residuals;
- (2) compute the central tendency of the range for each IM as the weighted geometric mean. Weights are proportional to the square of the number of stations that recorded the earthquake;
- (3) compute the dispersion of the ranges for each IM as the weighted standard deviation of the natural logarithm of  $b$ ;
- (4) compute the empirical cumulative probability distribution of  $b$  for each IM and verify that ranges are lognormally distributed through the Kolmogorov–Smirnov statistical test; and finally,
- (5) fit the computed weighted mean and standard deviation with simple models as a function of the period using Eq. 8 and the following:

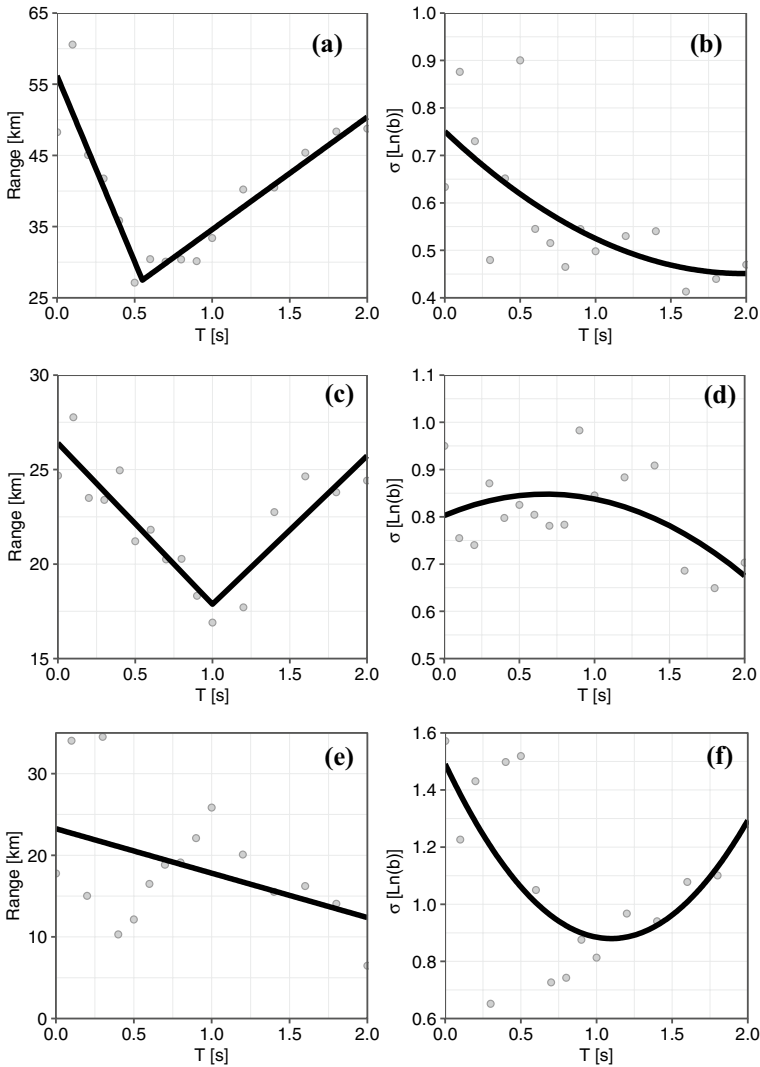
$$\sigma_{\text{Ln}(b)} = a_0 + a_1 T + a_2 T^2 \quad (9)$$

The fitted models for the median and variability are then used in conjunction with Eq. (5) to simulate spatially distributed ground motion fields. For each earthquake scenario in an event-based PSHA,  $b$  is sampled from a lognormal distribution with median and dispersion provided by the fitted model. By so doing, the correlation structure of the different ground motion field varies each time as would be expected in nature, rather than being fixed.

The resulting ranges and their associated variability for the three regions are presented in Fig. 5 along with the fitted models, whereas the model coefficients are reported in Table 2. We note that the appropriateness of these models is measured based on standard criteria, such as the BIC (Bayesian information criteria) and AIC (Akaike's information criterion).

The correlation parameters for the Northern and Central Italy regions are modelled through bilinear (range–Eq. 8a) and quadratic (dispersion–Eq. 9) expressions, whereas the Southern Italy model for the range value is expressed through a simpler linear equation (Eq. 8b). We note that the Southern Italy model features a higher dispersion compared to the other two regions. This is due to both the lower number of events available for this region and the smaller number of stations that on average recorded those events. The Northern Italy region has a lower variability with respect to the Central Italy model. At the same time, the former features larger ranges compared to the latter over all the considered periods, suggesting that the ground motion is on average correlated over longer distances. Such a trend is most likely due to peculiarities of the local site effects and propagation path, which may strongly affect the spatial correlation of a region, as demonstrated for instance by Jayaram and Baker (2009), Sokolov and Wenzel (2013) and Schiappapietra and Douglas (2020). The Central Italy region is indeed characterized by a high degree of heterogeneities in terms of local site effects compared to the Po plain (Northern Italy) region, where soil conditions are more homogeneous in terms of  $V_{s30}$ .

Finally, Fig. 6 compares published correlation models with the models proposed in this study. Some observations are worthy of remark. Firstly, although Sgobba et al. (2019)

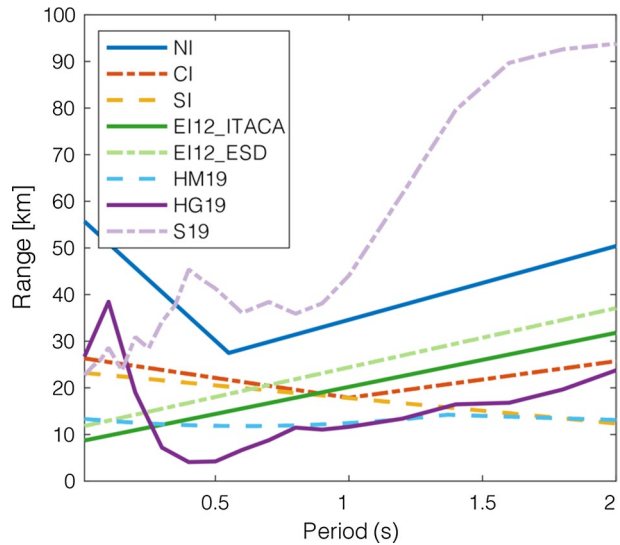


**Fig. 5** (Left hand side) Computed weighted geometric mean and fitted model, and (right hand side) computed weighted standard deviation and fitted model. (Top) Northern Italy, (Middle) Central Italy, and (Bottom) Southern Italy

**Table 2** Coefficients of the equations fitted to the ranges (Eqs. 8 and 9) for the three different regions

Region		$a_0$	$a_1$	$a_2$	$t$
Northern Italy	b [km]	27.48	-52.20	15.81	0.55
	$\sigma$	0.75	-0.30	0.08	-
Central Italy	b [km]	17.87	-8.52	7.85	1
	$\sigma$	0.8	0.13	-0.1	-
Southern Italy	b [km]	23.25	-5.44	-	-
	$\sigma$	1.49	-1.11	0.51	-

**Fig. 6** Comparison among different correlation models: NI-Northern Italy; CI-Central Italy; SI-Southern Italy; EI12\_ITACA-Esposito and Iervolino (2012); EI12\_ESD-Esposito and Iervolino (2012); HM19-Heresi and Miranda (2019); HG19-Huang and Galasso (2019); S19-Sgobba et al. (2019)



provided correlation parameters of the corrective term (e.g. sum of the repeatable terms of variability due to source, path and site effects) and not of the within-event residuals, our Northern Italy model is in agreement with their results, suggesting that this region features on average ground motions correlated over longer distances. The Huang and Galasso (2019) and Esposito and Iervolino (2012)-ITACA models were developed based on events from the entire Italian territory and therefore they provide an average spatial correlation structure. While these models have different trends at shorter periods compared to those proposed in this study, differences with the Central Italy model are not significant for periods longer than  $\sim 0.8$  s. This may likely be due to the datasets on which the correlation parameters were computed, which are dominated by normal faulting events that occurred along the Apennine chain. It is noted that Esposito and Iervolino (2012) proposed two different models based on the ITACA ([http://itaca.mi.ingv.it/ItacaNet\\_31/#/home](http://itaca.mi.ingv.it/ItacaNet_31/#/home)) and ESD (<http://www.isesd.hi.is>) databases, respectively. For comparison, we also show the ESD model, which is based on pan-European (and not just Italian) data. For the sake of completeness, we also show the correlation model proposed by Heresi and Miranda (2019) which is based on 39 well-recorded events from the NGA-West2 database (Ancheta et al. 2014). Generally, this study has lower ranges with respect to the models developed for Italy, suggesting that regional differences in spatial correlation cannot be neglected [although it should be noted that the models of Heresi and Miranda (2019) are based on more records per event than those derived here].

### 3 Openquake-engine input models

#### 3.1 Hazard component

The OpenQuake engine is used here to perform both deterministic and probabilistic seismic hazard assessments for the region of interest. The following subsections describe the

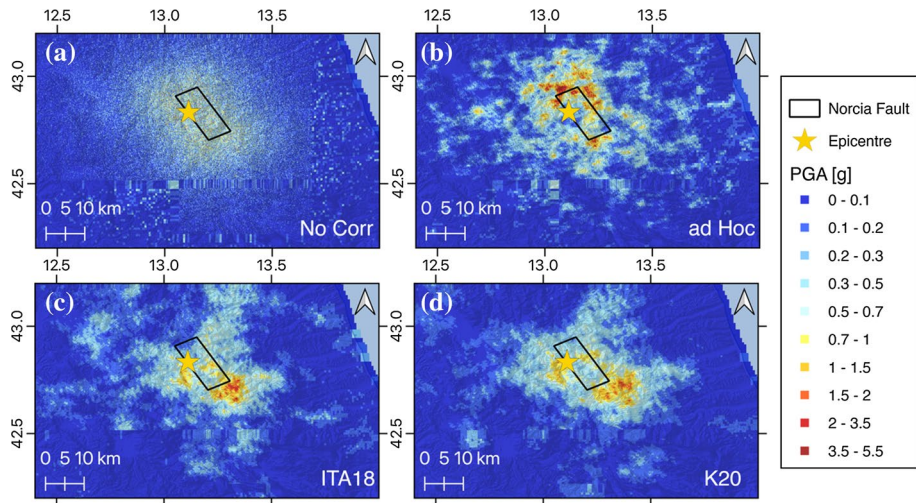
input parameters and models required for both the scenario-based PSHA and event-based PSHA calculations (Pagani et al. 2014).

### 3.1.1 Scenario-based seismic hazard assessment

The computation of ground-motion fields for a specific earthquake scenario requires three main inputs: (1) a fault rupture model that defines the location and geometry of the source, (2) a GMM, and (3) a model of the local site conditions. We consider a recent earthquake, namely the Mw 6.5 30th October 2016 Norcia (Central Italy) event, as the reference to define the rupture. The following parameters ([https://esm-db.eu/#/event/EMSC-20161030\\_0000029](https://esm-db.eu/#/event/EMSC-20161030_0000029)), along with the fault geometry, are used as input to OpenQuake: epicentre latitude 42.82°; epicentre longitude 13.16°; focal depth 6.8 km; Mw 6.5; rake=95°; strike 151°; and dip 47°. We select the ITA10 GMM (Bindi et al. 2011), which is one of the best performing models for shallow active crustal regions in Italy according to Lanzano et al. (2020). We choose an independent GMM compared to the ones used in the correlation modelling, so that the correlation model is the only varying input. We calculate ground-motion fields at multiple resolution grids with ~1 km and ~250 m grid spacing. The finer grid is used for densely populated area with potential of high IMs values to properly account for local site effects. Finally, the local site conditions are taken into account following Mori et al. (2020), who derived a detailed  $V_{s30}$  map for Italy (spatial resolution of 50 × 50 m), which also includes data from Italian seismic microzonations.

The OpenQuake engine allows the simulation of multiple spatially-correlated random fields to account for the aleatory variability in the ground motion by sampling the between- and within-event variability components from the GMM and by considering the spatial correlation of the within-event variability. To ensure the convergence of the mean and associated standard deviations of the results, we generate 1,000 ground motion fields based on the works of Silva (2016) and Costa et al. (2018). The spatial correlation is modelled as mentioned in Sect. 2.4. To investigate to what extent this feature impacts the overall economic losses, we consider four different cases. First, we generate ground-motion fields without considering the spatial correlation. Thereafter, we employ the three models described in Sect. 3. It is noted that the algorithm to simulate spatially-correlated random fields has been fine-tuned to reduce the computational cost and to enable a larger number of sites to be tested. Instead of the original covariance formulation, which is already implemented in OpenQuake, we use the randomization method of the Python package *gstools* (Müller and Schüller 2021). This method represents the spatial random field through the Fourier integral (Wu and Baker 2014) and it evaluates its discretised modes at random frequencies.

We generate spatially-correlated ground-motion fields for PGA and SA at  $T=0.3$ , 0.6 and 1 s. These IMs are used in the vulnerability component to calculate the building damage. A building can be sensitive to vibrations of different periods depending on its height and material of construction. Figure 7 provides a realization of the spatially-correlated ground-motion field for PGA obtained by using the different correlation models. The hazard footprint for the case without spatial correlation appears very noisy. Indeed, the within-event residuals are randomly generated so that the IMs of interest at each site are considered as independent random variables. By contrast, the distribution of PGA values in the other cases show smoother patterns as a result of the spatial correlation, which makes closely-spaced sites likely to experience similar ground-motion levels. The comparison between the ad hoc and K20 models is also noteworthy. The latter has a greater range



**Fig. 7** Spatially-correlated ground-motion maps obtained by using different correlation models: **a** No within-event spatial correlation; **b** Ad hoc spatial correlation model; **c** ITA18 spatial correlation model; and **d** K20 spatial correlation model. The yellow star represents the Mw 6.5 Norcia epicentre, whereas the black rectangle defines the surface fault projection

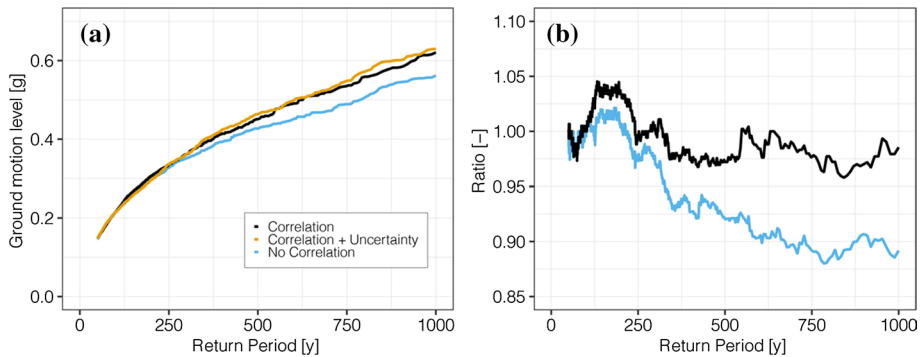
(correlation length) which makes the ground motion correlated over longer distances. Conversely, the shorter range in the ad hoc model results in a patchier distribution of the PGA.

The ground-motion fields are then combined with the vulnerability and exposure components to calculate the losses of the portfolio. The outcomes of the risk analysis are shown in Sect. 4.

### 3.1.2 Event-based probabilistic seismic hazard assessment

We use the event-based calculator from the OpenQuake engine to compute PSHA for Italy through a Monte Carlo (MC) simulation-based approach (Pagani et al. 2014). The first step generates synthetic earthquake catalogues, also called a stochastic event set, by randomly sampling all possible ruptures from the input source model (Musson 2000; Atkinson and Goda 2013). The events from the catalogue are then used to estimate the ground-motion IMs of interest at each site by using a list of suitable GMMs in conjunction with the model of the local site conditions. Likewise to the scenario-case, we adopt the  $V_{s30}$  map proposed by Mori et al. (2020). A logic tree of GMMs is usually employed to describe the ground-motion field with the aim of capturing the epistemic uncertainty. Finally, the IMs estimates obtained at each site for each event are rearranged so that the seismic hazard curves (annual maximum IMs as a function of the probability of exceedance) can be inferred. Musson (2000) demonstrated that for a sufficiently large number of simulations the results of the event-based PSHA are close to the outcomes of the classical PSHA, which uses the numerical integration of the total probability integral. The reader should refer to Atkinson and Goda (2013) for further details and advantages of this approach over the classical one.

In our analyses, we implement the SHARE (Woessner et al. 2015) source and ground-motion models for Italy. In order to account for spatial correlation models for three main macro-regions of Italy (see Sect. 2.5) we assume that only the source area can contribute to the seismic hazard and risk of the region under study. With regard to the GMMs, we



**Fig. 8** **a** Hazard curves showing the return period of different ground motion levels in terms of PGA for three different cases: ground-motion fields are generated without considering correlation; ground-motion fields are generated considering correlation; ground-motion fields are generated considering correlation and its associated uncertainty. **b** Ratio of ground motion levels as a function of the return period

slightly modify the GMM logic tree to enable the modelling of the spatial correlation of the within-event residuals of the ground motion. Indeed, not all the GMMs included in the logic tree (Cauzzi and Faccioli 2008; Toro 2002; Campbell and Bozorgnia 2003) decompose the total aleatory variability into between- and within-event components. For such GMMs, we set the within-event standard deviation ( $\varphi$ ) approximately up to 90% of the total standard deviation ( $\sigma_{TOT}$ ) based on preliminary analyses of the ratio  $\varphi/\sigma_{TOT}$  of the other GMMs. We implement in the OpenQuake engine the spatial correlation models developed in Sect. 2.5 to evaluate the impact of such features on the risk outcomes. We perform three different tests: (1) not taking into account spatial correlation; (2) including the event-to-event spatial correlation variability as described in Sect. 2.5; and (3) considering only the median range to characterise the spatial dependency of the ground motion. Figure 8 shows an example of hazard curves for the Norcia city obtained for the three different case-studies. The curves represent the annual probability of exceedance of various PGA levels. While hazard curves have the same trend at high probabilities of exceedance independently of the correlation structure, the results tend to diverge slightly at lower probabilities of exceedance (return periods of  $\sim 200$  year).

### 3.2 Vulnerability component

The vulnerability component defines loss ratios as a function of IM values depending on the building structure class. In our study, we use the vulnerability functions that have been developed in the framework of the 2020 SERA (Seismology and Earthquake Engineering Research Infrastructure Alliance for Europe, <http://www.sera-eu.org/en/home/>) project. Crowley et al. (2020), Martins and Silva (2020) and Silva et al. (2020) developed about 500 functions to cover the building classes of the European exposure database (Crowley et al. 2020). The building classes for Italy are shown in Sect. 3.3.

The adopted vulnerability functions employ PGA or SA at 0.3, 0.6 and 1 s as ground motion IMs depending on the fundamental period of the different building classes. In general, high-rise structures are associated with long-period IMs, whereas low-rise buildings are associated with short-period IMs, such as the PGA or SA at 0.3 s. It is noted that, for the sake of simplicity, and to highlight the effects of different correlation models, in our

analysis, we do not consider the uncertainties associated with the loss ratio. The reader should refer to Crowley et al. (2021) and Martins and Silva (2020) for further details on the calibration of the vulnerability component.

### 3.3 Exposure component

In our analysis, we implement the European exposure database (Crowley et al. 2020) for Italy, which has been developed within the SERA project. The database reports the distribution of the main residential, industrial and commercial buildings classes along with their replacement costs and numbers of occupants. Buildings are classified according to: (1) the main construction material (e.g. unreinforced masonry–MUR, confined masonry–MCF, reinforced concrete–CR, steel–S); (2) lateral load resisting system; (3) number of stories; (4) seismic design code level (e.g. pre-code–CDN, low code–CDL, moderate code–CDM, high code, CDH); and (5) lateral force coefficient used in the seismic design. The replacement cost is the value of replacing a building based on the latest building codes of the country and it is given by the sum of structural and non-structural costs as well as the cost of contents. As an example, Fig. 9a shows the buildings per each building class according to the European exposure database, and Fig. 9c provides the percentages of total replacement costs for the structural and non-structural components and the contents. These data refer to the area of interest of both the scenario-case and event-based calculations, as presented in Fig. 9b. The reader should refer to Crowley et al. (2020) for further details on the derivation of the exposure database.

## 4 Risk analysis and discussion

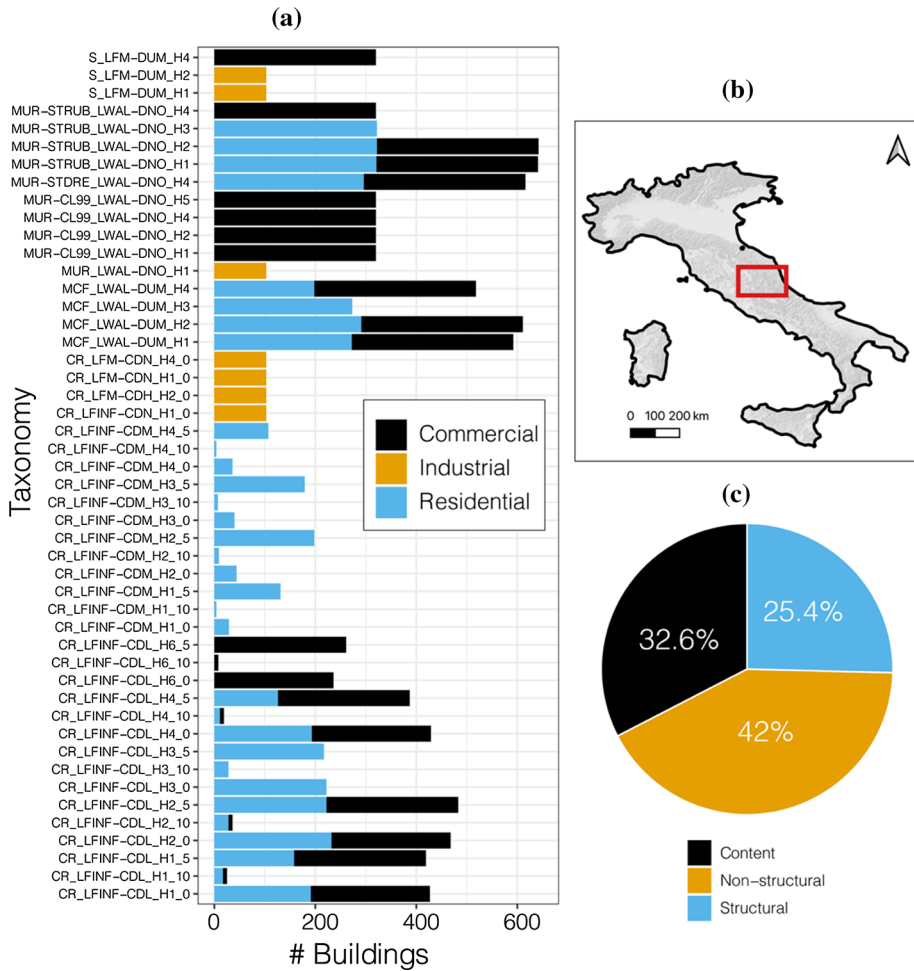
### 4.1 Scenario-base loss assessment

In this section, we present the potential economic losses obtained by combining the vulnerability and exposure components with the hazard model presented in Sect. 3.1.1. Figure 10 shows the distribution of losses with respect to the total portfolio value (%TV). We derive a loss map for each of the correlation models derived for the Norcia event to highlight the potential effects of considering spatially-correlated ground-motion fields. The economic portfolio is disaggregated at the grid level, based on the population density, in contrast to the original exposure model, which assumes that all the buildings are located at the centroid of each municipality. Indeed, such configurations would mask the actual effects of using various correlation models, especially if the inter-site distance among the municipality centroids is larger than the correlation ranges. In particular, we obtain population density data for each grid cell starting from the available raster information (<https://www.eea.europa.eu/>) and we compute weighting factors as population density in a given cell divided by population density in the corresponding municipality. Eventually, the total economic value in each grid cell is calculated by multiplying the total economic value of the municipality by the corresponding weighting factor.

Although differences in the spatial distribution of losses are discernible among the case studies, risk calculations do not seem to be strongly affected by the specific correlation model, as also highlighted in Table 3.

We believe that the exposure model, in terms of heterogeneity, plays a critical role in determining such results. Indeed, as shown in Fig. 4, the correlation models tend to

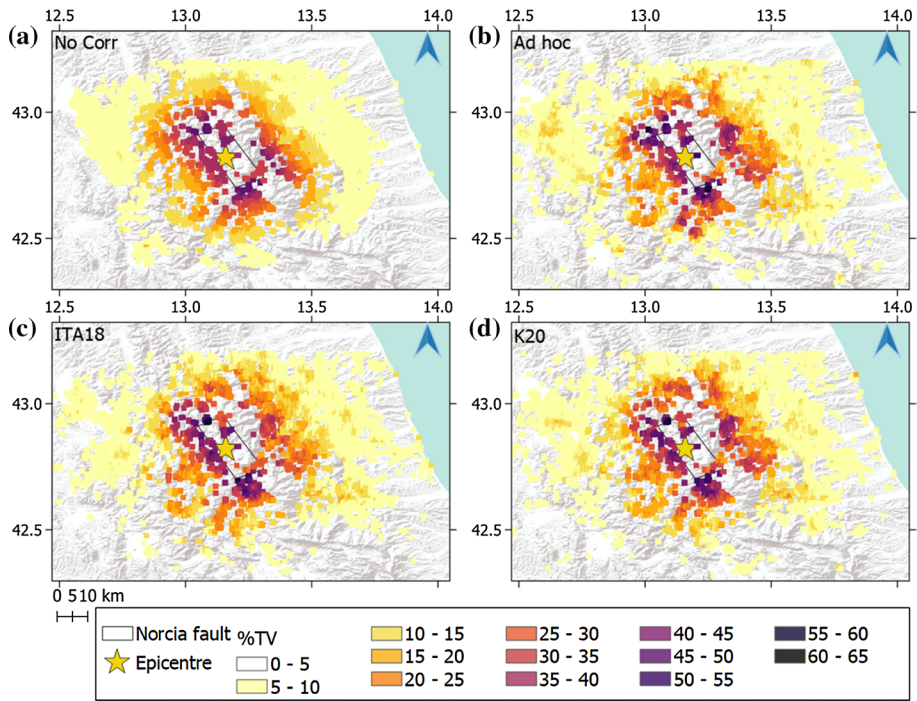




**Fig. 9** **a** Classification of the building classes according to the European exposure model. For further details on the labels, the reader should refer to Crowley et al. (2020); **b** Area of interest: residential, industrial and commercial buildings of graphs (a) and c are within the red rectangle; **c** Percentage of replacement costs, divided into contents and structural and non-structural components

converge toward analogous values as the period increases. Despite almost 90% of the buildings in the portfolio consisting of low-rise structures, only 16% of those buildings are predominantly sensitive to PGA, whereas the remaining 74% are sensitive to longer-period IMs (e.g. SA at 0.3 and 0.6 s). This means that although low-rise structures are predominant within the portfolio, the corresponding losses are similar as a result of ground motions being correlated over larger distances.

Such conclusions can also be drawn from Fig. 11, which presents the histograms of economic losses computed based on the different hazard models. We do not observe significant differences in the loss distributions, in contrast to Wagener et al. (2016), who found an increasing coefficient of variation (CV), i.e. the ratio between the standard deviation and the mean, with increasing correlation length. The CV is indeed almost the same (0.55) for all cases. In addition to the above-mentioned explanations, these



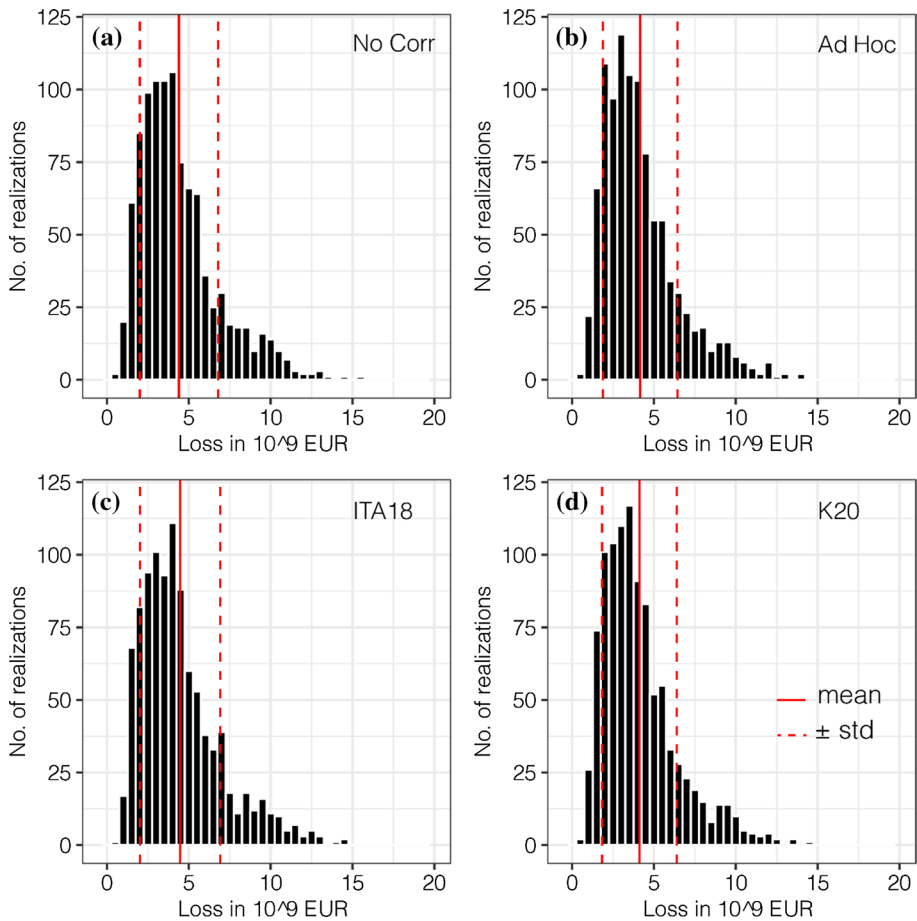
**Fig. 10** Losses (%TV computed based on different hazard models expressed in terms of percentage of total value of the portfolio: **a** Spatial correlation is not considered; **b** ad hoc correlation model; **c** ITA18 correlation model; and **d** K20 correlation model. The yellow star represents the Mw 6.5 Norcia epicentre, whereas the black rectangle defines the surface fault projection

**Table 3** Overall loss in %TV (percentage of total value of the portfolio) for the four different case studies

Correlation model	Overall loss in %TV
No Correlation	4.56
Ad hoc	4.30
ITA18	4.63
K20	4.26

diverging outcomes could be associated with the portfolio dimension. Park et al. (2007) and Weatherill et al. (2015a, b) demonstrated that the effects of including spatial correlations are greater when smaller portfolios (with a footprint within the correlation length) are considered. In such cases, it is much more likely that ground motions are substantially higher/lower than the median values at all sites within the exposure database, compared to larger portfolio.

To further highlight the impact of spatial correlation models within the risk assessment, we therefore consider a homogeneous spatially-distributed flat portfolio for each IM (e.g. PGA and SA at 1 s). In this case, only one building type is considered at all grid points with same building cost.

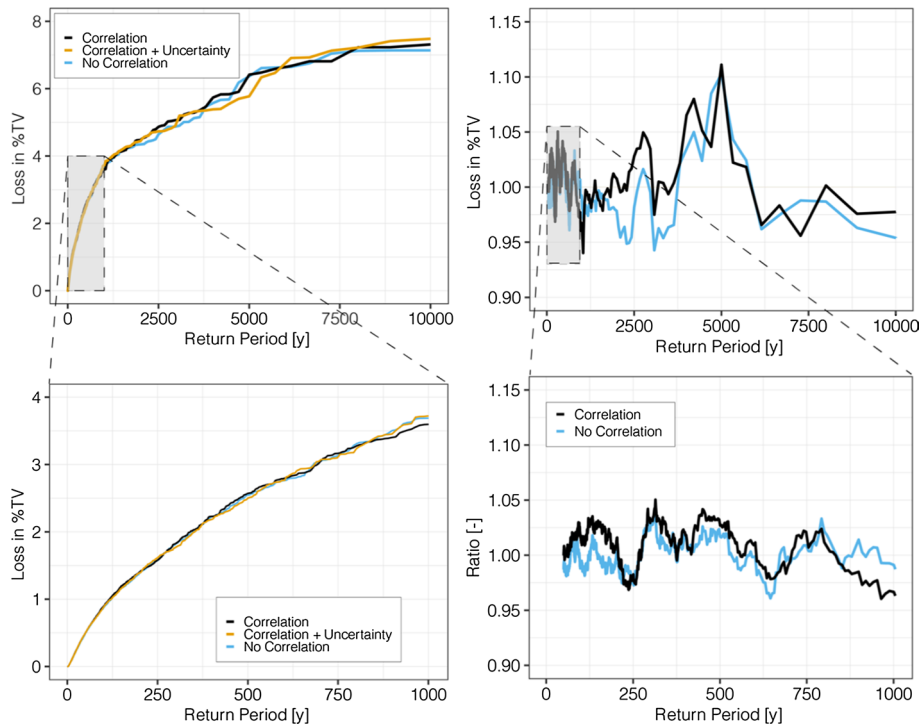


**Fig. 11** Histograms of economic losses computed based on different hazard models: **a** Spatial correlation is not considered; **b** ad hoc correlation model; **c** ITA18 correlation model; and **d** K20 correlation model. Red lines indicate the mean value, whereas the red dashed lines indicate the mean  $\pm$  std

**Table 4** Mean and Coefficient of variation of the losses obtained from the different correlation models for two different IMs

Correlation model	Mean ( $10^9$ )-PGA	CV-PGA	Mean ( $10^9$ )-SA 1 s	CV-SA 1 s
No correlation	6.4	0.79	29.2	0.64
Ad hoc	6.1	0.80	29.6	0.65
ITA18	6.4	0.88	30	0.65
K20	6.7	0.91	29.7	0.65

We compute loss distribution parameters, such as mean, standard deviation and CV (Table 4), for each case to identify trends. In agreement with Wagener et al. (2016), we now note that while the mean is almost constant, CV tends to increase as the correlation length increases for short-periods IMs. On the other hand, the CVs remain stable for long-period



**Fig. 12** Left: EP curves in terms of loss in %TV for three different cases: ground-motion fields generated without considering correlation; ground-motion fields generated considering correlation; and ground-motion fields generated considering correlation and its associated uncertainty. TV stands for total value of the economic portfolio (2020 SERA exposure model). In the inset, we show EP curves up to a return period of 1000 years, which is usually the range of interest for (re)insurance industry. Right: Loss ratio as a function of the return period for the cases in which ground-motion fields are generated without considering correlation and considering correlation. The case in which ground-motion fields are generated considering correlation and its associated uncertainty is taken as a reference. In the inset, we show loss ratio curves up to a return period of 1000 years

SAs. Indeed, the correlation models proposed in Sect. 2.4 converge towards comparable correlation lengths, thus resulting in similar loss distributions.

## 4.2 Event-based probabilistic loss assessment

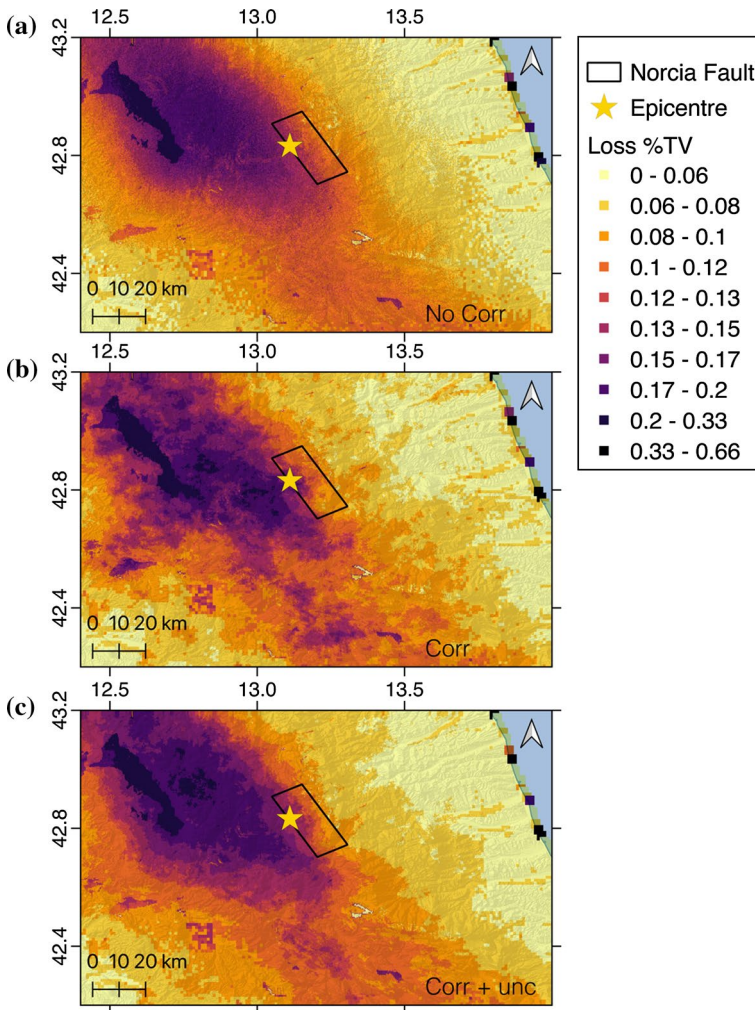
In this section, we present the potential economic losses obtained by combining the vulnerability and exposure components with the hazard model presented in Sect. 3.1.2. We focus only on the same geographical area (Central Italy) to better understand the impact of including the spatial correlation in our analysis. Risk assessment for the whole Italian mainland will be the object of future developments. Figure 12 shows the exceedance probability (EP) curves in terms of losses for three different case-studies: (1) ground-motion spatial correlation is neglected; (2) ground-motion fields are generated by considering a median correlation model as presented in Sect. 2.5; and (3) event-to-event spatial correlation uncertainty is taken into account as explained in Sect. 2.5. In general, we observe

greater losses at higher return period (lower annual probabilities of exceedance) when spatial correlation and its associated uncertainty are included with respect to the case in which the spatial correlation is neglected. When only a median spatial correlation is used, rare losses are slightly overestimated compared to when correlation is neglected. Such results are in agreement, for example, with Park et al. (2007) and Weatherill et al. (2015a, b), even though our study does not show marked differences. An explanation may lie with the heterogeneity of the portfolio along with its dimension, which would also require the consideration of cross-correlation among different IMs to obtain more robust loss estimates. On the other hand, for return period of interest for the (re)insurance industry (up to 1000 years), EP curves show a similar behaviour, suggesting that overall losses are not strongly affected by the different correlation structures. This is in contrast with Park et al. (2007), who demonstrated that for similar return periods losses can be underestimated by 19% when spatial correlation is neglected.

We are aware that our outcomes are preliminary because an accurate estimation of the likelihood of observing rare and frequent losses would require the definition of the cross-correlation, especially in case of a heterogeneous portfolio. Nonetheless, such results provide useful advances to better understand the impact of including not only spatial correlation but also its associated variability. In particular, the consideration of the event-to-event spatial correlation variability leads to an overestimation of the rare losses with respect to the simple spatial correlation case, whereas a trend cannot be identified for frequent losses. We recall that when only a median correlation model is considered, the range is constant for each ground-motion field generated. Conversely, the correlation structure of the different ground-motion fields varies each time when the dispersion of the range is considered in the analysis.

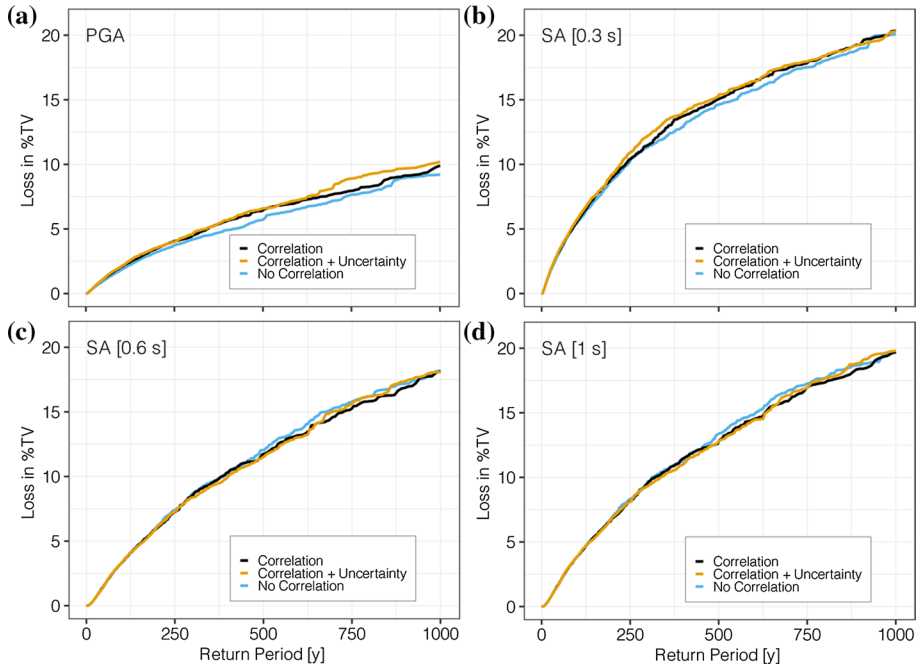
To further highlight the impact of spatial correlation and its associated uncertainty within the probabilistic risk assessment, we therefore account for a homogeneous spatially disaggregated flat portfolio for different IMs, such as PGA and SA at 0.3, 0.6 and 1 s. In this case, only one building type is considered in each grid cell. Figure 13 compares the average annual losses (in %TV) obtained for the same three cases as considered for the SERA economic portfolio (Fig. 12). The inclusion of spatial correlation (with and without the event-to-event variability) makes the risk footprints less noisy and more realistic with respect to the case in which spatial correlation is neglected (Fig. 13a). Indeed, the distribution of losses tends to be very noisy when neglecting spatial correlation as a result of the ground motion being independent at neighbouring sites. Therefore, loss estimates are also independent owing to the homogeneity of the flat portfolio. In contrast, when the correlation is considered, closely spaced sites are likely to experience similar ground-motion levels, so that the distribution of losses shows smoother patterns. Differences are also evident when comparing Fig. 13 b and c, which is when the associated spatial correlation uncertainty is accounted for. Such outcomes demonstrate the importance of considering not only spatial correlation, but also the event-to-event spatial correlation variability in seismic risk assessments. We note that these results refer to PGA; however, similar outcomes are obtained for the other IMs.

For sake of completeness, we show in Fig. 14 the loss EP curves obtained for the different IMs. Rare losses are generally underestimated if spatial correlation is not accounted for. This is more evident for short period IMs, such as PGA and SA at  $T=0.3$  s, compared to long period IMs (SA at  $T=0.6$  and 1 s). Indeed, the spatial correlation model for Central Italy presented in Sect. 2.5 features ranges that decreases as the period increases up to 1 s. Therefore, the effects of considering spatially-correlated ground-motion fields diminish with increasing period. Besides, we note that both



**Fig. 13** Average annual losses (AAL) for the flat homogeneous portfolio (PGA) obtained for three different cases: **a** ground-motion fields are generated without considering correlation; **b** ground-motion fields are generated considering correlation; and **c** ground-motion fields are generated considering correlation and its associated uncertainty. The yellow star represents the Mw 6.5 Norcia epicentre, whereas the black rectangle defines the surface fault projection

frequent and rare losses are generally higher if the correlation uncertainty is accounted for with respect to the simple spatial correlation case for short period IMs. In contrast, sound conclusions are difficult to draw at longer IMs, as the EP curves show negligible differences. Further analyses are still required to advance the understanding of including the event-to-event spatial correlation variability in loss estimates; however, we believe that these findings have significant implications for providing more informed seismic risk assessments.



**Fig. 14** EP curves in terms of loss in %TV for three different cases: ground-motion fields are generated without considering correlation; ground-motion fields are generated considering correlation; and ground-motion fields are generated considering correlation and its associated uncertainty. Losses are computed for flat homogeneous portfolios: **a** PGA; **b** SA at  $T=0.3$  s; **c** SA at  $T=0.6$  s; **d** SA at  $T=1$  s

## 5 Conclusions

In this work, we perform both scenario and event-based seismic hazard and risk calculations to advance the understanding of spatial correlations in the catastrophe modelling process. We first derive custom spatial correlation models based on the Mw 6.5 Norcia earthquake to better investigate to what extent different input models may affect the overall economic losses. Thereafter, we extend our analyses to probabilistic calculations, which are often important for decision-making and (re)insurance and underwriting purposes. In this context, we develop not only median spatial correlation models, but we also provide an estimate of the event-to-event associated dispersion to enable more realistic estimations of both the ground motion and corresponding losses.

Deterministic scenario calculations for the Norcia event suggest that the economic portfolio in terms of both heterogeneity and footprint dimension affect the impact of considering spatially-correlated ground-motion fields in risk analyses. Indeed, risk calculations for the 2020 SERA economic portfolio do not seem to be strongly affected by the specific correlation model, even though differences in the spatial distribution of losses may be spotted. To further highlight the role of different spatial correlation models, we thus assumed a flat portfolio, in which only a single building type is considered at each grid point. In agreement with Wesson and Perkins (2001) and Wagener et al. (2016), we find that the variance of losses varies according to the specific spatial correlation structure used in the generation of ground-motion fields. We note that Wagener et al. (2016) found a coefficient of variation

that is 3.6 times higher when spatial cross-correlation is included compared to the case in which correlation is neglected, whereas the CV in our analyses (only considering spatial correlation) increased by only 1.15 times. Albeit not significantly evident, our outcomes suggest that the hypotheses on which spatial correlation models are grounded (e.g. ground motion models to compute the within-event residuals) play a crucial role in the estimation of the overall potential losses for the area of interest. As a consequence, the implementation of a unique correlation model calibrated on worldwide databases may represent an oversimplification.

Event-based calculations shed light on the importance of considering not only the ground-motion spatial correlation, but also its associated uncertainty in risk analyses. Loss exceedance probability curves for both the SERA economic and flat portfolios indicate that neglecting spatial correlation leads to biases in loss estimates. In particular, smaller losses are expected at lower annual probabilities of exceedance. For return periods of interest for the (re)insurance industry (up to 1000 years), we note that EP curves show a similar behaviour (maximum differences are only about 10% in contrast to Park et al. (2007), who found differences up to 20%), suggesting that overall losses are not strongly affected by the different correlation structures. Note that results strongly depend on the considered portfolio and its relative dimension with respect to the correlation length.

Nonetheless, both hazard and risk footprints in terms of average annual losses feature less noisy and more realistic patterns if spatial correlation is taken into account. Therefore, this study strengthens the idea that the inclusion of spatial correlation and of its associated variability is crucial for enabling more informed risk assessments.

Although our outcomes demonstrate that spatial correlations in probabilistic seismic hazard and risk assessments improves per-event and in-location loss estimates for underwriting purposes, this study presents some caveats. When heterogeneous portfolios are of concern, not only it is important to consider the spatial correlation at closely-spaced sites, but the cross-correlation among different IMs should also be accounted for at once. Neglecting the cross-correlation would lead to independent ground-motion fields for different IMs and thus to inaccurate estimation of the likelihood of observing rare losses. The development of a novel approach to simulate spatially cross-correlated ground-motion fields indicates features for further developments of this preliminary study.

Finally, the consideration of isotropic spatial correlation models may represent a strong simplification. Although the non-parametric tests presented in Sect. 2.3 suggest that the hypothesis of isotropy is satisfied at a 5% significance level, several authors (e.g. Chen and Baker 2019; Abbasnejadfar et al. 2020; Infantino et al. 2021; Schiappapietra and Smerzini 2021) demonstrated that anisotropic models better capture the complex ground-motion patterns. Further modelling work should be undertaken to establish the effectiveness of including anisotropic spatial correlation structure in risk analyses. It should be noted that many of results obtained in this study are based on data from a single Mw 6.5 event. Therefore, it is recommended that data from other well-recorded events (e.g. the 26th October Mw 6.0 earthquake) in the same area be considered to verify the generality of the results obtained.

**Acknowledgements** This work has been supported by the University of Strathclyde (Ph.D. of first author). We thank two anonymous reviewers for their detailed and constructive comments on a previous version.

**Authors' contribution** E.S.: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review and editing, Visualization. S.S.: Conceptualization, Software, Formal analysis, Writing – review and editing. P.P.: Software. J. Douglas: Supervision, Writing – review and editing. G.T.: Writing – review and editing.



**Funding** University of Strathclyde (Ph.D. of Erika Schiappapietra).

**Data availability** The ESM (Lanzano et al. 2019) strong-motion flat-file is available at <https://esm.mi.ingv.it/flatfile-2018/>.

**Code availability** The OpenQuake engine is available at <https://www.globalquakemodel.org/oq-get-started>.

## Declarations

**Conflict of interest** The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abbasnejadfar M, Bastami M, Fallah A (2020) Investigation of anisotropic spatial correlations of intra-event residuals of multiple earthquake intensity measures using latent dimensions method. *Geophys J Int* 222(2):1449–1469. <https://doi.org/10.1093/gji/ggaa255>
- Ancheta TD et al (2014) NGA-West2 database. *Earthq Spectra* 30(3):989–1005. <https://doi.org/10.1193/070913EQS197M>
- Atkinson GM, Goda K (2013) Probabilistic seismic hazard analysis of civil infrastructure, handbook of seismic risk analysis and management of civil infrastructure systems. Woodhead Publishing Limited, Woodhead. <https://doi.org/10.1533/9780857098986.1.3>
- Baker JW, Chen Y (2020) Ground motion spatial correlation fitting methods and estimation uncertainty. *Earthq Eng Struct Dyn*. <https://doi.org/10.1002/eqe.3322>
- Bindi D et al (2011) Ground motion prediction equations derived from the Italian strong motion database. *Bull Earthq Eng* 9(6):1899–1920. <https://doi.org/10.1007/s10518-011-9313-z>
- Bowman AW, Crujeiras RM (2013) Inference for variograms. *Comput Stat Data Anal* 66:19–31. <https://doi.org/10.1016/j.csda.2013.02.027>
- Bowman AW, Azzalini A (2018) R package ‘sm’: nonparametric smoothing methods (version 2.2-5.6). <http://www.stats.gla.ac.uk/~adrian/sm>
- Campbell KW, Bozorgnia Y (2003) Updated near-source ground-motion (attenuation) relations for the horizontal and vertical components of peak ground acceleration and acceleration response spectra. *Bull Seismol Soc Am* 93(1):314–331
- Cauzzi C, Faccioli E (2008) Broadband (0.05 to 20 s) prediction of displacement response spectra based on worldwide digital records. *J Seismolog* 12(4):453–475. <https://doi.org/10.1007/s10950-008-9098-y>
- Chen Y, Baker JW (2019) Spatial correlations in cybershake physics-based ground-motion simulations. *Science* 2:10. <https://doi.org/10.1785/0120190065>
- Costa C, Silva V, Bazzurro P (2018) Assessing the impact of earthquake scenarios in transportation networks: the Portuguese mining factory case study. *Bull Earthq Eng* 16(3):1137–1163. <https://doi.org/10.1007/s10518-017-0243-2>
- Cressie N (1985) Fitting variogram models by weighted least squares. *J Int Assoc Math Geol* 17(5):563–586. <https://doi.org/10.1007/BF01032109>
- Crowley H et al (2020) Exposure model for European seismic risk assessment. *Earthq Spectra* 36(1\_suppl):252–273. <https://doi.org/10.1177/8755293020919429>
- Douglas J, Edwards B (2016) Recent and future developments in earthquake ground motion estimation. *Earth-Sci Rev* 160:203–219. <https://doi.org/10.1016/j.earscirev.2016.07.005>

- Erdik M (2017) Earthquake risk assessment, bulletin of earthquake engineering. Springer, Netherlands. <https://doi.org/10.1007/s10518-017-0235-2>
- Esposito S, Iervolino I (2012) Spatial correlation of spectral acceleration in European data. *Bull Seismol Soc Am* 102(6):2781–2788. <https://doi.org/10.1785/0120120068>
- Goda K (2011) Interevent variability of spatial correlation of peak ground motions and response spectra. *Bull Seismol Soc Am* 101(5):2522–2531. <https://doi.org/10.1785/0120110092>
- Grossi P, Kunreuther H (2005) Catastrophe modeling: a new approach to managing risk. Springer, Berlin
- Heresi P, Miranda E (2019) Uncertainty in intraevent spatial correlation of elastic pseudo-acceleration spectral ordinates. *Bull Earthq Eng* 17(3):1099–1115. <https://doi.org/10.1007/s10518-018-0506-6>
- Infantino M, Smerzini C, Lin J (2021) Spatial correlation of broadband ground motions from physics-based numerical simulations. *Earthq Eng Struct Dynam* 50(10):2575–2594. <https://doi.org/10.1002/eqe.3461>
- Jayaram N, Baker JW (2009) Correlation model for spatially distributed ground-motion intensities. *Earthq Eng Struct Dynam* 38:2258
- Kohrangi M, Bazzurro P, Vamvatsikos D (2021) ‘Seismic risk and loss estimation for the building stock in Isfahan: part II—hazard analysis and risk assessment. *Bull Earthq Eng* 19(4):1739–1763. <https://doi.org/10.1007/s10518-020-01037-1>
- Kotha SR et al (2020) A regionally-adaptable ground-motion model for shallow crustal earthquakes in Europe, bulletin of earthquake engineering. Springer, Netherlands
- Lanzano G et al (2019) A revised ground-motion prediction model for shallow crustal earthquakes in Italy. *Bull Seismol Soc Am* 109(2):525–540. <https://doi.org/10.1785/0120180210>
- Lanzano G et al (2020) ‘Ground motion models for the new seismic hazard model of Italy (MPS19): selection for active shallow crustal regions and subduction zones. *Bull Earthq Eng* 18(8):3487–3516. <https://doi.org/10.1007/s10518-020-00850-y>
- Martins L, Silva V (2020) Development of a fragility and vulnerability model for global seismic risk analyses. *Bull Earthq Eng*. <https://doi.org/10.1007/s10518-020-00885-1>
- Mori F et al (2020) A new Vs30 map for Italy based on the seismic microzonation dataset. *Eng Geol* 275:105745. <https://doi.org/10.1016/j.enggeo.2020.105745>
- Müller S, Schüller L (2021) GeoStat-Framework/GSTools. Zenodo. <https://doi.org/10.5281/zenodo.1313628>
- Musson RMW (2000) The use of Monte Carlo simulations for seismic hazard assessment in the U.K. *Annali di Geofisica* 2:1–9
- Pagani M et al (2014) Openquake engine: an open hazard (and risk) software for the global earthquake model. *Seismol Res Lett* 85(3):692–702. <https://doi.org/10.1785/0220130087>
- Park J, Bazzurro P, Baker JW (2007) Modeling spatial correlation of ground motion Intensity Measures for regional seismic hazard and portfolio loss estimation. *Appl Stat Probab Civil Eng* 2:1–8
- Schiappapietra E, Douglas J (2020) 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. *Earth Sci Rev*. <https://doi.org/10.1016/j.earscirev.2020.103139>
- Schiappapietra E, Douglas J (2021) Assessment of the uncertainty in spatial-correlation models for earthquake ground motion due to station layout and derivation method. *Bull Earthq Eng*. <https://doi.org/10.1007/s10518-021-01179-w>
- Schiappapietra E, Smerzini C (2021) Spatial correlation of broadband earthquake ground motion in Norcia (Central Italy) from physics-based simulations. *Bull Earthq Eng* 19(12):4693–4717. <https://doi.org/10.1007/s10518-021-01160-7>
- Sgobba S et al (2019) Spatial correlation model of systematic site and path effects for ground-motion fields in Northern Italy. *Bull Seismol Soc Am*. <https://doi.org/10.1785/0120180209>
- Silva V (2016) Critical issues in earthquake scenario loss modeling. *J Earthq Eng* 20(8):1322–1341. <https://doi.org/10.1080/13632469.2016.1138172>
- Silva V et al (2020) Development of a global seismic risk model. *Earthq Spectra* 36(1\_suppl):372–394. <https://doi.org/10.1177/8755293019899953>
- Sokolov V, Wenzel F (2011) ‘Influence of spatial correlation of strong ground motion on uncertainty in earthquake loss estimation. *Earthq Eng Struct Dyn* 40(9):993–1009. <https://doi.org/10.1002/eqe.1074>
- Sokolov V, 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. *Bull Earthq Eng* 11(6):1909–1926. <https://doi.org/10.1007/s10518-013-9493-9>
- Wagener T et al (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 Dyn Earthq Eng*. 2:166–178. <https://doi.org/10.1016/j.soildyn.2016.03.016>

- Wagener T et al (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 Dyn Earthq Eng* 85:166–178. <https://doi.org/10.1016/j.soildyn.2016.03.016>
- Weatherill GA et al (2015a) Exploring the impact of spatial correlations and uncertainties for portfolio analysis in probabilistic seismic loss estimation. *Bull Earthq Eng* 13(4):957–981. <https://doi.org/10.1007/s10518-015-9730-5>
- Weatherill GA et al (2015b) Exploring the impact of spatial correlations and uncertainties for portfolio analysis in probabilistic seismic loss estimation. *Bull Earthq Eng* 13(4):957–981. <https://doi.org/10.1007/s10518-015-9730-5>
- Woessner J et al (2015) The 2013 European seismic hazard model: key components and results. *Bull Earthq Eng* 13(12):3553–3596. <https://doi.org/10.1007/s10518-015-9795-1>
- Silva V et al. (2012) Development of an open-source platform for calculating losses from earthquakes, pp 1–14. <http://ria.ua.pt/handle/10773/9637>
- Toro GR et al. (2002) Modification of the attenuation equations for large magnitudes and short distances, Risk Engineering Technical Report
- Wesson RL, Perkins DM (2001) Spatial correlation of probabilistic earthquake ground motion and loss, *bulletin of the seismological Society of America*. <https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/91/6/1498/2713201/1498.pdf>
- Wu J, Baker JW (2014) 'round motion modeling for risk and reliability assessment of San Francisco infrastructure systems. In: NCEE 2014-10th U.S. National conference on earthquake engineering: frontiers of earthquake engineering. <https://doi.org/10.4231/D3959C806>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.