Received: 2 January 2021 | Revised: 23 March 2021 | Accepted: 2 April 2021

DOI: 10.1002/eqe.3461

WILEY

RESEARCH ARTICLE

Spatial correlation of broadband ground motions from physics-based numerical simulations

Maria Infantino | **Chiara Smerzini** | **Jiayue Lin**

Department of Civil and Environmental Engineering, Politecnico di Milano, Milan, Italy

Correspondence
Chiara Smerzini, Department of Civil and Environmental Engineering, Politecnico di Milano, Milan, Italy.
Email: chiara.smerzini@polimi.it

Funding information
Project URBASIS "New challenges for Urban Engineering Seismology", H2020-MSCA-ITN-2018, Grant Agreement 813137

Abstract

This study assesses the spatial correlation of broadband earthquake ground motions from 3D physics-based numerical simulations in near-source conditions. State-of-the-art models for predicting the spatial correlation are derived from wide datasets including densely recorded earthquakes in different areas worldwide and, therefore, they may be poorly representative of specific regions and near-source effects. A large set of broadband ground motions simulated by the SPEED code, and enriched in the high-frequency range with an Artificial Neural Network technique, is used to investigate the sensitivity of crucial parameters in geostatistical analysis (number of receivers), as well as of source, path, and site effects on spatial correlation, with a level of detail which could not be possible otherwise due to the paucity of recordings. First of all, the comparison of our results with those derived from earthquake recordings validates successfully the numerical approach in predicting the spatial correlation in a broad frequency range. Furthermore, the study points out that spatial correlation of response spectral accelerations is significantly affected by the magnitude, forward directivity effects, ground-motion directionality (fault normal versus fault parallel), and relative position from the causative fault. These features may make critical the use of isotropic and stationary models especially in near-fault conditions.

KEYWORDS
spatial correlation, broadband ground motions, near-source effect, 3D physics-based numerical simulations, seismic risk

1 | INTRODUCTION

One of the key issues in the seismic risk assessment of spatially distributed portfolios or infrastructural systems (eg, lifelines), which are the backbone of modern urban environments, is the modeling of the spatial correlation of ground-motion intensity measures (IMs). Indeed, if during the same earthquake all structures of a portfolio can be affected, it is not correct to treat the IMs at all sites as independent variables, especially if they are located closely in space. To provide some examples, Weatherill et al.[1] and Park et al.[2] considered the impact of spatial correlation on the aggregated loss analysis of synthetic building portfolios and concluded that, on one side, considering spatial correlation produces greater losses at longer return periods, and, on the other side, that such an impact is sensitive to the spatial scale and type of the portfolio. In a similar study, Goda and Atkinson[3] showed that the inclusion of the correlation structure of ground motion may have

a remarkable impact on the estimated seismic losses by a factor as high as 50%. Recently, Garakaninezhad and Bastami[4] investigated the issues related to the correlation of vertical ground motions and found that neglecting spatial correlation may provide unsafe evaluations of seismic performance of an idealized bridge network.

These considerations led to the development of several spatial correlation models for different IMs, mainly in terms of Spectral Accelerations (SA) at different vibration periods (T), using datasets of ground-motion recordings.[5–13] For a thorough overview of the existing empirical models, we refer the reader to the recent review published by Schiappapietra and Douglas.[14]

However, it should be noted that the ground-motion correlation structure depends strongly on local geologic conditions (region-specific) as well on source and path effects (scenario-specific), therefore, a single generalized spatial correlation model may not be appropriate to represent specific regions, as pointed out by Sokolov et al.[9] for the Taiwan case, and by Schiappapietra and Douglas[14] for Central Italy. Thus, the spatial correlation models proposed in literature, based on the analysis of wide strong motion datasets, may be not fully representative of the specific area under study.

On this regard, a powerful tool capable of producing detailed and region-specific estimates of earthquake ground shaking and of its spatial variability is represented by 3D Physics-Based Numerical Simulations (PBS). The numerical simulations are based on physical models of the seismic source, the propagation path from the source to the site and the local geologic structure. In this work, we used the 3D PBS generated by the Discontinuous Galerkin Spectral Element code SPEED[15] (http://speed.mox.polimi.it/) already used in several case studies both for verification within international benchmarks, for validation with recorded near-source ground motion or for generation of realistic earthquake ground-motion scenarios for risk assessments.[16–18] To enrich the frequency content of PBS, limited typically to about 1.5-2 Hz, broadband ground motions have been generated by coupling SPEED results with an Artificial Neural Network (ANN) technique.[19] The advantages of PBS in estimating spatial correlation are essentially two: (1) availability of thousands of receivers "recording" the same (simulated) earthquake; (2) inclusion of the physics of the earthquake, from the source up to the site. It follows that PBSs allow for a systematic investigation of the impact of crucial parameters in geostatistical analysis, such as the number of receivers, and of physical factors, such as source, propagation path and local site features, for a variety of "virtual", albeit realistic, conditions. Such analyses can be hardly obtained from recordings because of their limited number and, hence, their unfavorable spatial sampling.

With this premise, the goal of this work is twofold: first, to verify the accuracy of the broadband simulation approach to reproduce a realistic spatial correlation of ground motion, as inferred from the analysis of strong motion recordings, and, second, to investigate systematically the influence of different physical factors, namely, magnitude, source directivity, ground-motion directionality, and source-to-site path. To this end standard geostatistical tools, based on the computation of both semivariograms and Pearson's correlation coefficient (with and without the hypotheses of stationarity and isotropy, respectively), are used to estimate the correlation of SA(T) separately for different horizontal components.

A similar study based on PBS has been done by Chen and Baker,[20] who developed spatial correlation maps for different earthquake scenarios and reference sites in the Los Angeles region using Cybershake simulations. Compared to Chen and Baker,[20] we introduced some novelties, namely: (i) spatial correlation analyses are obtained from simulated ground motions reliable in a broad frequency range, that is, not only limited to the low-frequency range propagated by the numerical model; (ii) the PBS dataset covers different regions worldwide (Po Plain, L'Aquila, Marsica, Sulmona, Norcia, in Italy; Istanbul, in Turkey; Thessaloniki, in Greece); (iii) the role of directionality of ground shaking and of factors related to the fault rupture realization, such as magnitude and forward directivity effects, is addressed in greater detail.

The article is organized as follows. First the simulation dataset (numerical approach and case studies considered) and the geostatistical methods are described. Then, after some sensitivity analyses on the geostatistical tool (eg, number of receivers; choice of median regression for computing intraevent residuals), an overview of the main results derived from the PBS dataset is presented. Furthermore, the dependence of spatial correlation, that is, correlation length (range) and variance (sill), on magnitude, ground-motion component, source-to-site path is explored and discussed. Finally, the spatial correlation is analyzed relaxing the hypotheses of isotropy and stationarity and maps of the Pearson's correlation coefficient are shown for earthquake scenarios in the Istanbul area.

2 | THE NUMERICAL BROADBAND APPROACH AND DATASET

The numerical procedure adopted in this study to generate the broadband ground motions consists of two main steps: (1) a full 3D wave propagation simulation by SPEED and (2) a technique based on ANN to enrich the high-frequency content of the simulated waveforms.

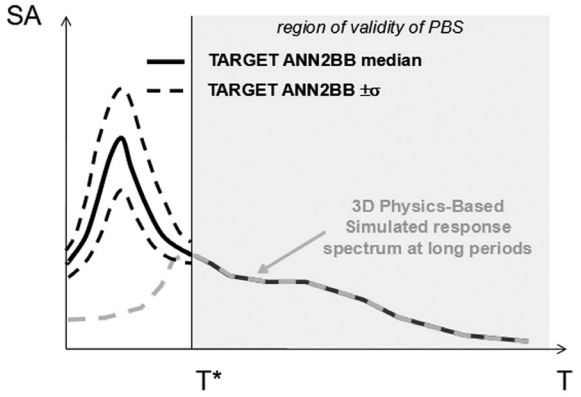

FIGURE 1 Construction of the target broadband spectrum using the ANN2BB approach (after Paolucci et al.[19])

To accomplish the first step, a proper 3D numerical model has to be defined. Once the topographical, geotechnical, and geological information are collected, a 3D soil model can be constructed combining: (a) the digital elevation model; (b) the crustal structure generally described in form of a layered model of S and P wave velocity (V_S and V_P, respectively), and (c) the local shallow geological structure with a model of V_S and V_P variable both in the horizontal and vertical direction, and possibly including the corresponding models for internal soil damping and nonlinear curves (variation of shear modulus and damping ratio as a function of shear strain). Note that the present version of SPEED can account for stochastic fluctuations of seismic wave velocity, as discussed in a recent work on the simulation of induced earthquakes in Groningen,[17] whose impact on spatial correlation could be addressed in future studies. The geological model is then combined with that of the source based on seismotectonic knowledge. For the case studies discussed in the present work, a kinematic rupture generator has been used applying along the fault a heterogeneous slip distribution combined with a slip source function, typically a sigmoid one, with initiation time and length depending on the local rupture velocity and rise time. Note that in SPEED two kinematic rupture generators are available, namely Herrero and Bernard[21] and Schmedes et al.[22], in the sequel referred to as HB94 and SEA12, respectively. The source and velocity models are then condensed into a spectral element numerical model consisting of hexahedral elements with a spectral degree selected in order to propagate a maximum frequency typically of about 1.5-2 Hz. This frequency threshold is mainly related to the level of detail of the information regarding both the seismic source and 3D surface geology and to the computational cost of fine grids.

It follows that, in order to make numerically simulated ground motions usable for earthquake engineering applications, such as risk analyses, a fundamental step of SPEED post-processing concerns the generation of broadband ground motions, where the low-frequency simulated waveforms are enriched in the high-frequency range to produce time histories with a realistically broad frequency content. In this study, broadband ground motions have been obtained through an approach based on ANN, referred to as ANN2BB, presented and validated against real case studies in Paolucci et al.[19] Referring the reader to the relevant publication for a detailed description of the methodology, we limit herein to recall the main features of the approach. In the ANN2BB approach, an ANN is trained based on a dataset of strong-motion records (eg, SIMBAD[23] or NGA-West2[24]) to predict short period spectral ordinates ($T < T^*$) based on long period ones ($T \geq T^*$), being T^* the minimum period of the numerical model. Then, for each receiver, a target ANN2BB response spectrum is computed (see Figure 1), the spectral ordinates of which, for $T \geq T^*$, coincide with the simulated ones, while, for $T < T^*$, they are obtained from the ANN (median value over 20 ANN realizations, separately for horizontal and vertical components). Although the ANN2BB technique appears as a promising approach to produce broadband ground motions generally consistent with recorded waveforms and with realistic features in terms of spatial variability (as further demonstrated in this work), it still does not overcome the main drawbacks related to the lack of a physical modeling of high-frequency ground-motion components. For this reason, the technique suffers from some limitations that are typical of machine learning methods such as the dependence on the training dataset and overfitting issues.

The procedure described above has been adopted to obtain the ground-motion dataset considered in this study, an overview of which is given in Table 1. The dataset includes seven different regions worldwide (namely, L'Aquila, Sulmona, Po Plain, Marsica and Norcia, in Italy; Thessaloniki in Greece; Istanbul in Turkey) and rupture scenarios with moment magnitude in the range $M_W =$ 6-7.4. The geological conditions are also rather variable, passing from regions with very soft alluvial basins (eg, Marsica) to regions with stiffer sediments (eg, Norcia). It is worth

TABLE 1 List of PBSs computed by SPEED and considered in this work

Region	Fault	Style of faulting	Model size (km^3)	V_s ^{surface} (m/s)	Moment magnitude M_W	No. Scenarios	References
L'Aquila, Central Italy	Paganica	Normal	58 × 58 × 20	300-1700	6.2	1	Evangelista et al.[25]
Sulmona, Central Italy	Mt. Morrone	Normal	49 × 42 × 13	500-1200	6.0	5	Villani et al.[26]
					6.5	3	
Po Plain, Northern Italy	Mirandola	Reverse	74 × 51 × 20	300	6.0	1	Paolucci et al.[27]
Marsica, Central Italy	Fucino	Normal	56 × 46 × 20	100-1000	6.7	1	Paolucci et al.[28]
Norcia, Central Italy	Mt. Vettore–Mt. Bove	Normal	50 × 40 × 21	280-1700	6.5	1	Özcebe et al.[29]
Istanbul, Turkey	North Anatolian Fault (Marmara Sea)	Strike-Slip	165 × 100 × 30	250-1350	7.0	4 + 11*	Infantino et al.[18]
					7.4	1 + 20*	
Thessaloniki, Northern Greece	Gerakarou	Normal	82 × 64 × 31	300-2000	6.5	1	Smerzini et al.[30]
	Anthemountas	Normal			7.0	1	Smerzini et al.[31]

*Scenarios used to compute the Pearson's coefficient.

recalling that the PBSs included in this dataset were already verified and validated in previous projects (see references in Table 1).

A total of 19 scenarios were used to compute the semivariogram, while, referring to the Istanbul case study, further 31 scenarios were considered to calculate the Pearson's correlation coefficient.

3 | GEOSTATISTICAL TOOLS TO MODEL SPATIAL CORRELATION OF GROUND MOTIONS

This section briefly introduces the geostatistical tools commonly used in seismology and in other fields to describe the spatial correlation of a random function.[32]

3.1 | Semivariogram

The semivariogram is the geostatistical tool commonly used in seismology to model the spatial correlation of earthquake ground motion[7,6,8,10,33] as well as the geological data.[34,35] The semivariogram γ quantifies the average dissimilarity between variable Z at locations x_i and x_j:

$$\gamma = \frac{1}{2} Var\left(Z_{x_i} - Z_{x_j}\right) = \frac{1}{2} E\left[\left(Z_{x_i} - Z_{x_j}\right)^2\right] \quad (1)$$

where Var and E denote the variance and expected value, respectively. However, because for each site multiple realizations of a given earthquake are generally not available, the assumptions of second-order (or weaker) stationary and isotropy have to be made to estimate γ in Equation (1).

Under these assumptions, the semivariogram depends only on the separation vector h and covariance and semivariogram are equivalent:

$$\gamma(h) = \frac{1}{2} Var\left(Z_{x_j} - Z_{x_{j+h}}\right) = \frac{1}{2} E\left[\left\{Z_{x_j} - Z_{x_{j+h}}\right\}^2\right] \quad (2)$$

$$\gamma(h) = C(0) - C(h) = \sigma^2 - C(h) \quad (3)$$

where $C(\cdot)$ is the covariance operator and $C\ (0) = \sigma^2$ is the variance of Z. Moreover, we can introduce the spatial correlation coefficient $\rho(h)$ defined as:

$$\rho\ (h) = \frac{C\,(h)}{\sigma^2}\ =\ 1 - \frac{\gamma\,(h)}{\sigma^2} \tag{4}$$

Modeling the spatial correlation with semivariograms is generally carried out by the following steps: (1) calculation of the sample semivariogram $\hat{\gamma}(h)$ from the data set; (2) selection of the proper theoretical semivariogram model fitting the data; (3) estimation of the parameters for the chosen model.

Referring to the first point, the sample semivariogram is estimated through the method of moments[36]:

$$\hat{\gamma}\ (h) = \frac{1}{2N\,(h)} \sum_{j\,=\,1}^{N(h)} \left\{ z_{x_j} - z_{x_j+h} \right\}^2 \tag{5}$$

where $z_{x_j} - z_{x_j+h}$ is the difference between the data at sites separated by h and $N(h)$ denotes the total numbers of pairs at lag h. It is worth mentioning that the sample semivariogram can be computed using the more robust estimator proposed by Cressie and Hawkins,[37] less sensitive to the outliers. Once that sample semivariogram is computed, a functional form (negative definite) fitting the data has to be selected. The basic models generally used to this end are the Exponential, Gaussian, or Spherical models defined, for an isotropic case, respectively by Equations (6a), (6b), and (6c):

$$\gamma\ (h) =\ a\left[1 - exp\left(\frac{-3h}{b}\right)\right] \tag{6a}$$

$$\gamma\ (h) =\ a\left[1 - exp\left(\frac{-3h^2}{b^2}\right)\right] \tag{6b}$$

$$\begin{cases} \gamma\ (h) =\ a\left[\frac{3}{2}\frac{h}{b} - \frac{1}{2}\left(\frac{h}{b}\right)^3\right]\ if\ h \le b \\ a\ otherwise \end{cases} \tag{6c}$$

The parameters a and b of Equation (6) are called *sill* and *range*, and represent the variance of the random process and the separation distance at which the data can be considered fully uncorrelated, respectively.

In this work, we focus on the spatial correlation among residuals of the same IM for sites spatially distributed. Ground-motion models predict intensities at a site s due to an earthquake e and take the general form:

$$log_{10}\ Y_{se} = log_{10}\ \bar{Y}_{se}\,(M_W, R, S, \delta) + \delta B_e + \delta W_{se} \tag{7}$$

where Y_{se} denotes the IM of interest, that is, SA(T); $\bar{Y}_{se}$ is the predicted median value as a function of moment magnitude (M_W), source-to-site distance (R), local-site conditions (S), and others parameters (δ); δB_e is the interevent residual, which is a random variable with zero mean constant during the same earthquake across all the sites, and δW_{se} represents the intraevent residuals, a random variable with zero mean variable site-to-site.

Thus, the spatial correlation is herein computed on the intraevent residuals δW_{se} as the logarithmic (log_{10}) misfit between the simulated *SA(T)*, Y_{se}, at the site x with respect to a median trend $\bar{Y}_{se}$ given by the following relationships:

$$log_{10}\ \bar{Y}_{se} = c_1\ - c_2 log_{10}\,(R_{line} + c_3) + F_{site} \text{ with } F_{site} = \begin{Bmatrix} c_4 log_{10}\left(\frac{V_{S30}}{800}\right)\ (a) \\ c_4 H_{bed}\ (b) \end{Bmatrix} \tag{8}$$

where c_i, with $i = 1,\ldots 4$, are the model coefficients calibrated by Least-Squares (LS) regression, R_{line} is the closest distance from the surface fault projection of the segment at the top edge of the rupture plane (see definition in Paolucci et al.), F_{site} is the site term, defined as a function of either (a) V_{S30} (time-averaged shear wave velocity in the top 30 m) or (b) H_{sed} (depth of the alluvial-bedrock interface). The latter is adopted only for the Po Plain case study, since the corresponding numerical model is characterized by an irregular bedrock interface with a spatially homogeneous V_{S30} distribution. As regards the distance metric, R_{line} is used instead of more standard metrics, such as R_{JB} and R_{rup}, because studies underway suggest

that it is more efficient (i.e. lower dispersion) in near-source conditions. Note that, at variance with most works where the spatial correlation is estimated on the intraevent residuals with respect to the median predictions of existing empirical ground-motion models, in this work, the regression of Equation (8) is calibrated on the specific dataset associated with each PBS.

From an operational point of view, the procedure to compute semivariograms from broadband simulated motions belongs to the multistage class of geostatistical approaches[6] and can be summarized as follows: (1) compute the intraevent residuals on *SA* at periods $T = 0$, 0.2, 0.5, 1, 2 and 3 s, $\delta W_{se} = logSA_{se} - \overline{logSA}_{se}$, separately for the horizontal components, that is, fault normal (FN), fault parallel (FP), and horizontal geometric mean (GMH); (2) divide the separation distance, h, in bins of constant width δ; (3) compute the sample semivariogram by the method of moments; (4) fit the sample semivariogram by weighted LS regression using the exponential theoretical model and, hence, estimate the range and sill parameters.

In the procedure above, the exponential model is used for most of the PBS dataset because it is the model providing the best performance in terms of mean square error, with the exception of Istanbul, for which a spherical model is preferred because of the larger fit with the sample semivariogram. As regards the LS fitting, either a classical or weighted regression, with higher weights for short separation distances, was selected based on the minimum MSE. A bin width $\delta = 2$ km was chosen so that there is a stable trend of correlation with at least 30 pairs per bin.[38] Further details on the sensitivity analyses performed on the geostatistical analysis (median regression, type of LS fitting) are given in the following section.

As opposed to the multistage approach, more recently Ming et al.[39], proposed a novel approach called one-stage scoring algorithm, where the ground-motion model parameters and the spatial correlation parameters are both estimated in a single stage without introducing the semivariogram. The estimators produced by the one-stage scoring approach have good statistical properties (consistency, statistical efficiency, and asymptotic normality) and produce smaller prediction errors than the multistage algorithm. Furthermore, multistage approaches are only capable of estimations of isotropic and stationary correlation structures.

3.2 | Pearson's correlation coefficient

The Pearson's correlation coefficient ρ is a statistical measure of the linear correlation between two variables X and Y. It ranges from −1 to +1: a positive (or negative) ρ implies that the two variables are positive (or negative) correlated, while if $\rho = 0$ the variables are fully uncorrelated. When computed from a sample, the Pearson's correlation coefficient can be estimated as follows:

$$\hat{\rho}\ (x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{\sum_{i=1}^{n} z_{i,x} \cdot z_{i,y}}{\sqrt{\sum_{i=1}^{n} {z_{i,x}}^2}\sqrt{\sum_{i=1}^{n} {z_{i,y}}^2}} \tag{9}$$

where n is the sample size, x_i and y_i are the individual events in sample space, $\bar{x}$ and $\bar{y}$ are the sample means and $z_{i,x}$ and $z_{i,y}$ are the residuals of x_i and y_i with respect to the corresponding sample means (i.e. $z_{i,x} = x_i - \bar{x}$ and $z_{i,y} = y_i - \bar{y}$).

Based on Equation (9), the correlation coefficient of ground-motion residuals between sites j and k can be computed as follows:

$$\hat{\rho}\ (j,k) = \frac{\sum_{i=1}^{n} z_{i,j} \cdot z_{i,k}}{\sqrt{\sum_{i=1}^{n} {z_{i,j}}^2}\sqrt{\sum_{i=1}^{n} {z_{i,k}}^2}} \tag{10}$$

where n is the number of rupture realizations of a given earthquake scenario (an earthquake scenario refers to a specified magnitude M_W and fault rupture location; instead, a rupture realization refers to a particular rupture scenario with prescribed hypocenter and slip distribution) and z is the intraevent residual estimated as:

$$z_{i,j} = \widehat{\delta W}_{i,j} = log_{10}\left(Y_{i,j}\right) - \bar{Y}_{log_{10}(IM_{i,j})} - \widehat{\delta B}_i \tag{11}$$

where $\bar{Y}_{log_{10}(IM_{i,j})}$ is the logarithmic mean of a given IM, (that is, *SA(T)*) and $\widehat{\delta B}_i$ is the between-event residual (that is average misfit of ground motions of a particular event with respect to median prediction).

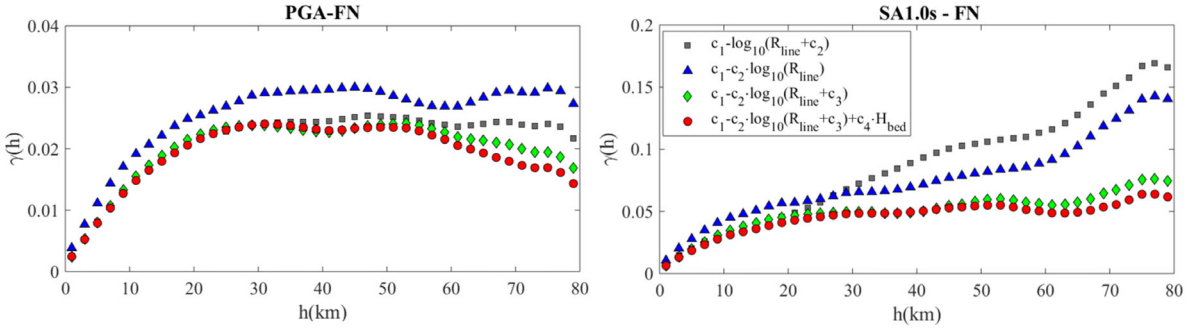

FIGURE 2 Effect of functional form for median trend in estimating the sample semivariogram: Po Plain, PGA-FN (left), and SA1.0s-FN (right)

By setting $IM = SA(T)$, $\bar{Y}_{log_{10}(IM_{i,j})} = \bar{Y}_{log_{10}(SA_{i,j})}$ is computed as sample mean of $log_{10}(SA(T))$ at site j over a suite of rupture realizations $i = 1,\ldots n$. Considering the huge number m of stations available (of the order of thousands) for SPEED simulations, the between-event residual can be estimated as:

$$\hat{\delta}\, B_i = \frac{1}{m} \sum_{j=1}^{m} \left(log_{10}\left(SA_{i,j}\right) - \bar{Y}_{log_{10}(SA_{i,j})} \right) \tag{12}$$

Differently from the semivariogram, which quantifies the correlation as a function of the interstation separation distance, the Pearson's coefficient allows describing the correlation between any site pair, j and k, in a nonstationary (dependence on the absolute position of the sites) and anisotropic (dependence on orientation) fashion.

However, as this approach requires observations of residuals $z_{i,j}$ and $z_{i,k}$ from multiple earthquake realizations, its applications has been limited due to the scarcity of recorded data. On the other hand, the numerousness of simulated ground motions allows using Pearson coefficient as a tool to investigate the ground-motion spatial correlation relaxing the hypotheses of stationarity and isotropy. Analysis of the Pearson's correlation coefficient for the CyberShake numerical simulations of the Los Angeles region has been recently addressed by Chen and Baker.[20]

4 | SENSITIVITY ANALYSES

In this section, some preliminary analyses are presented to discuss the sensitivity of geostatistical analysis on crucial steps such as the residual computation and the receiver selection.

First, the effect of the functional form adopted for the estimation of the median trend, that is, $log_{10}\overline{SA}_{se}$, Equation (8), is investigated to check the robustness of correlation results obtained by means of the semivariogram. Specifically, four different functional forms were considered: (1) $c_1 - log_{10}(R_{line} + c_2)$; (2) $c_1 - c_2 log_{10}(R_{line})$; (3) $c_1 - c_2 log_{10}(R_{line} + c_3)$; (4) $c_1 - c_2 log_{10}(R_{line} + c_3) + c_4 H_{bed}$. Note that functions from (1) to (3) differ only in the geometric attenuation term, while (4) adds the site effect term. It is found that the calibration of a proper median function has a strong impact on the semivariogram trend and, hence, on the robustness of model parameters (range and sill). As a matter of fact, the differences of the square and diamond symbols in Figure 2 indicate that the introduction of a two-coefficient geometric attenuation term allows one to get semivariogram with stable sill especially at long periods, without the need of applying detrending techniques.[32] On the other hand, the introduction of the site term has a more limited effect and mostly at longer periods, as expected for the deep sites in the Po Plain. These verifications, extended also to other case studies (not shown here for sake of brevity), confirmed the accuracy of the selected functional form (4), see Equation (8). Note also that for distances larger than around 60 km, a less representative sample of receivers is used in each bin (lower number and located at model boundaries) and, therefore, the semivariogram shows a decreasing trend with less robust estimates. The use of linear mixed-effect regression models which take into account also the variability of the explanatory variables in the prediction, could be explored in future analyses to extend these sensitivity analyses.

Furthermore, the influence of the number of receivers in computing the semivariogram has been tested. It is well known that the most important factor determining the accuracy of the sample semivariogram is the number of data within each distance bin. In general, the accuracy increases with the number of data.[40] With PBSs, it is possible to check how the

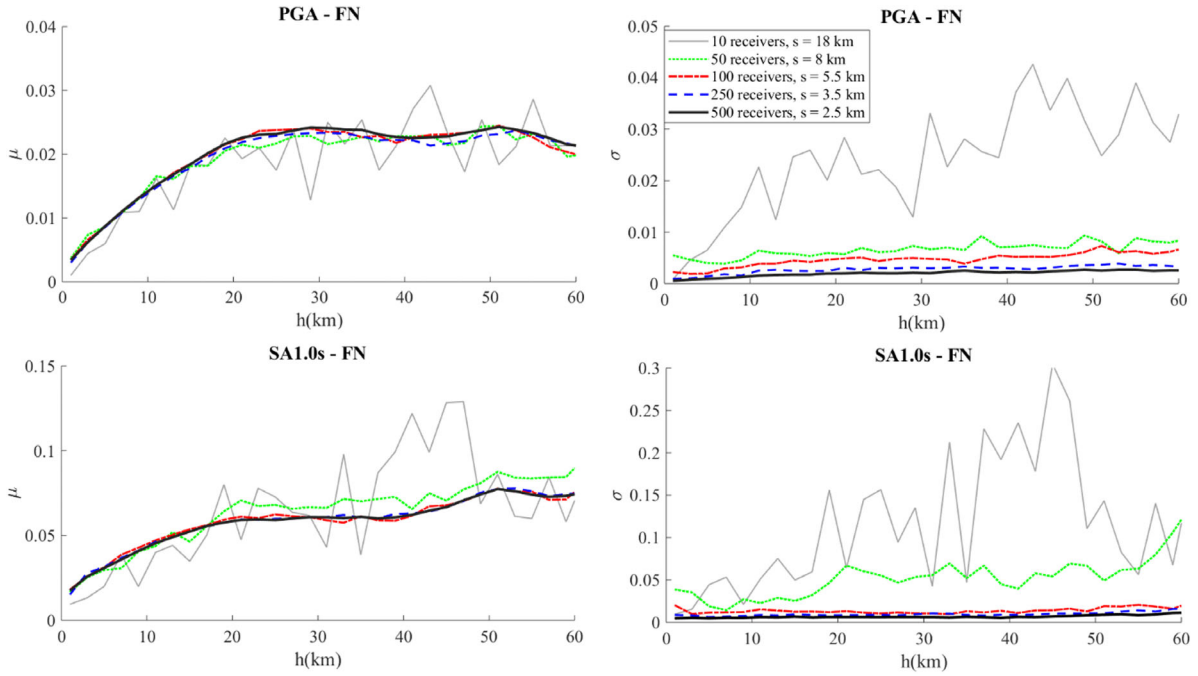

FIGURE 3 Influence of the number of receivers: Po Plain, PGA-FN (top), and SA1.0s-FN (bottom). Mean, μ (left), and standard deviation, σ (right), as a function of h, for 30 random samples of $N = 10$, 50, 100, 250, 500, and 1000 receivers with average spacing $s = 18$, 8, 5.5, 3.5, 2.5, and 1.8 km

accuracy of the computed semivariogram varies with an increasing number of receivers. To this end, for the Po Plain case study and for PGA-FN and SA1.0s-FN, the semivariogram was computed as the average of 30 random realizations of sample of N receivers, with $N = 10$, 50, 100, 250, 500, and 1000, corresponding to an average spacing $s = \sqrt{Area_{model}/N} = 18$, 8, 5.5, 3.5, 2.5 and 1.8 km, respectively. Figure 3 depicts the average (μ) and standard deviation (σ) of the 30 samples.

These analyses point out that an average receiver spacing lower than 5.5 km approximately ($N = 100$) is needed to have accurate estimates with a stable mean and variance. It is worth noting as this result is in agreement with the research study of Baker and Chen[41] who, analyzing the NGA West-2 dataset,[24] concluded that at least 100 stations are needed to limit estimation uncertainty and that for earthquakes with fewer stations, earthquake-specific estimation of spatial correlations should be performed with caution.

After these tests, we show in Figure 4 an illustrative example of the multistage approach adopted in this study for the Po Plain application and SA1.0s, FN component, in terms of: (a) map of $log_{10}SA(1.0s)_{FN}$; (b) map of $log_{10}\overline{SA(1.0s)}_{FN}$, given by Equation (8b); (c) map of logarithmic residuals $\delta W_{se} = log_{10}\, SA(1.0s)_{FN} - log_{10}\overline{SA(1.0s)}_{FN}$; and (d) sample semivariogram (dots) and best-fitting exponential model.

5 | OVERVIEW OF RESULTS AND DEPENDENCE ON PERIOD

To provide an overview of the spatial correlation models calibrated on each PBS, Figure 5 shows the range as a function of period, respectively, for eight representative PBSs. Specifically, for each case study and each causative fault (see Table 1), a single scenario has been considered (namely: M_W6.0 Po Plain–29.05.2012, M_W6.2 L'Aquila–06.04.2009, M_W6.7 Marsica–13.01.1915, M_W6.5 Norcia–30.10.2016, M_W6.0 Sulmona–S002, M_W6.5 and M_W7.0 Thessaloniki –20.06.1978 and S001, and M_W7.0Istanbul-S001) and the model parameters (range and sill) have been computed for six spectral ordinates: PGA, 0.2, 0.5, 1.0, 2.0, and 3.0 s.

The results are shown separately for the FN (filled symbols, continuous lines) and FP components (open symbols, dashed lines). When FN and FP components are not available (that is for segmented faults), the EW and NS components are considered. For the M_W6.0 Po Plain case study (see top left panel of Figure 5), our results are compared with the findings by Sgobba et al.,[42] who estimated the range of the source-, site-, and path-corrective terms of the ground-motion model specific for the Po Plain region using a nonergodic approach. Referring to the Po Plain case, a good agreement is

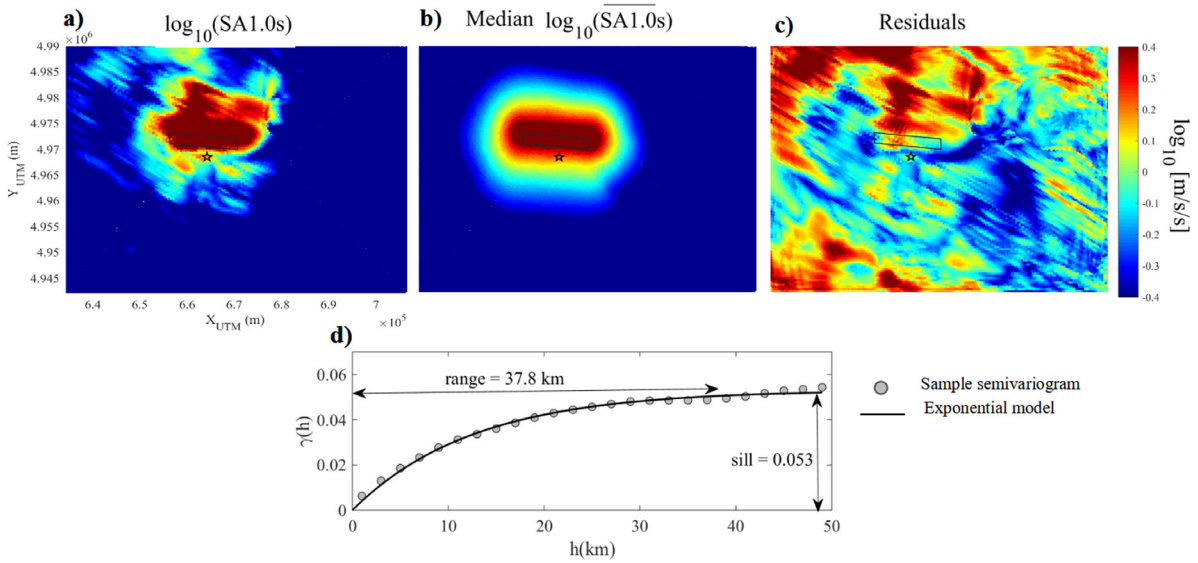

FIGURE 4 Illustrative example of the multistage approach adopted to compute the semivariogram: Po Plain, SA1.0s-FN. (A) Map of $log_{10}SA(1.0s)_{FN}$; (b) map of median $log_{10}\overline{SA(1.0s)}_{FN}$; (c) map of residuals $\delta W_{se} = log_{10}\ SA(1.0s)_{FN} - log_{10}\overline{SA(1.0s)}_{FN}$; (d) sample semivariogram (dots) and best-fitting exponential model obtained by LS regression

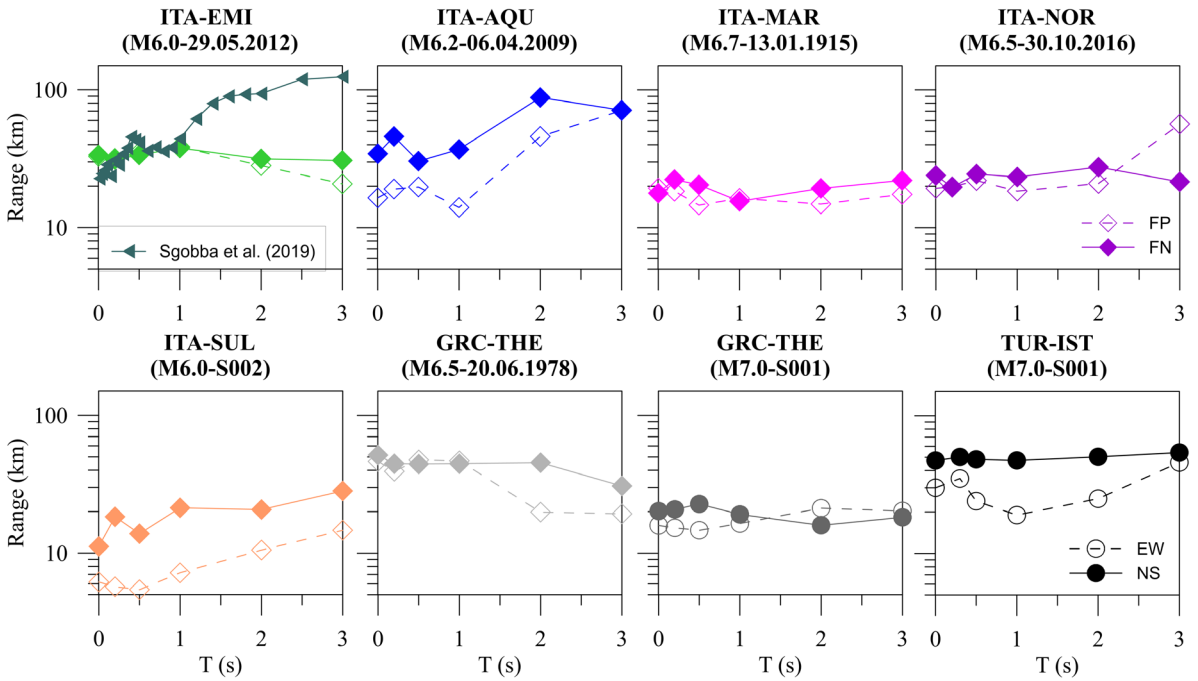

FIGURE 5 Range as a function of period for the different PBSs under consideration

found with Sgobba et al.[42] up to 1 s. The results by Sgobba et al.[42] shows an abrupt increase of range at periods longer than 1 s up to values of about 130 km at $T = 4$ s, while PBS-based estimates show that the range stabilizes between 30 and 40 km at long periods. Such discrepancies may be due to the different processing (residuals and number of data to compute the semivariograms) and to some unmodeled systematic site phenomena which are not fully captured by the residual decomposition in Sgobba et al.[42]

The ranges estimated in this work generally increase with period, similarly to other literature studies. This is consistent with past findings on spatial coherency of ground motion.[43] Being a measure of the linear correlation between two waveforms at a given distance, spatial coherency tends to decrease at increasing separation distance and at higher frequencies.

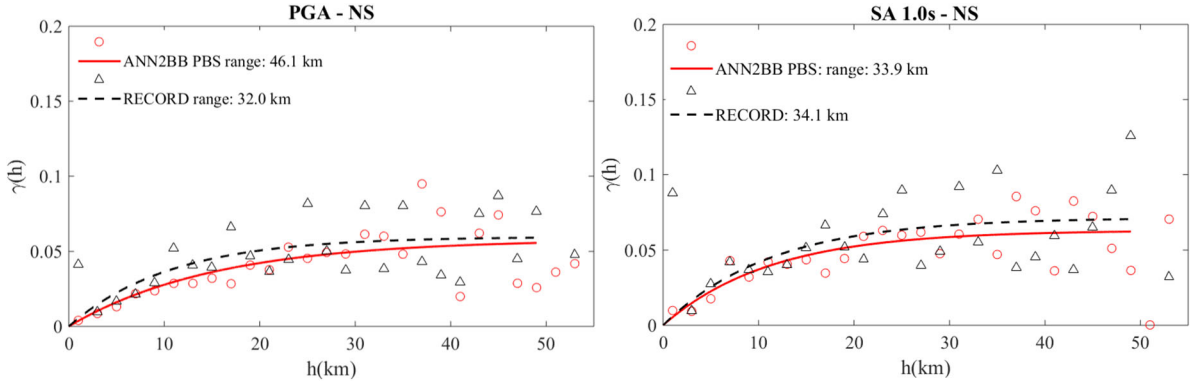

FIGURE 6 Comparison between the semivariograms computed from the ANN2BB simulated motions (dots and solid lines) and recordings (triangles and dashed lines) for the 2012 May 29 Po Plain earthquake: PGA (left) and SA 1.0s (right), NS component

Therefore, high-frequency components are more affected by small-scale heterogeneities and, therefore, they turn out to be less correlated, compared to low-frequency waves.[44] Nevertheless, for some cases (eg, M_W6.0 Po Plain, M_W6.5 Thessaloniki), it is found that the range is nearly constant or slightly decreasing for periods larger than around 1 s. A similar trend was also found in literature studies based on earthquake recordings, as further discussed in the next section.

5.1 | Comparison with recordings and other studies

Comparing the spatial correlation structure of the broadband motions with the recorded one is crucial to test the ability of the ANN2BB procedure to reproduce the actual correlation of ground motion at short periods. To this purpose, the M_W6 earthquake which occurred in the Po Plain (Northern Italy) on 29 May 2012 is an excellent case owing to the availability of 33 near-source recordings at epicentral distances less than 30 km. As an example, Figure 6 shows the semivariograms for PGA and SA1.0s, NS component, computed at the available accelerometric stations from both recordings (in triangles and dashed lines) and broadband simulated ground motions (BB, in dots and solid lines). Note that the simulated SA for $T \geq 0.75$ s are derived entirely from the PBSs, while for $T < 0.75$ s they are the result of the ANN2BB procedure. The NS component is selected because it is the component where a closer match is found between PBS and recorded waveforms (see Paolucci et al.[27]) and the scope herein is not to validate the PBS waveforms in strict sense. A very good agreement is found between the sample semivariograms from records and simulations across all periods. Some differences found in the range estimates for PGA are also related to the lower robustness of the LS regression, given the limited number of station pairs.

As a further validation test, we attempted to make a rather comprehensive comparison between the ranges obtained in this study with the empirical ranges available from other published works based on earthquake recordings. Specifically, in Figure 7 the range (GMH component) as a function of period from PBSs (for the same eight scenarios as considered previously, see Figure 5), is compared with the ranges computed directly on recordings (left panel, a) and with the ranges provided by empirical correlation models (right panel, b).

To this end, the following literature works are considered: Esposito and Iervolino[11] based on the Italian Accelerometric Archive (ITACA) and the European Strong-Motion Database (ESD); Hong et al.[45] on California data; Jayaram and Baker[6] using records from the Anza, Alumn Rock, Parkfield, Chi-Chi, Northridge, Big Bear City, and Chino Hills earthquakes; Sgobba et al.[42] on Po Plain records; Huang and Galasso[13] on the Italian records included in the ESD.

From Figure 7, the following comments can be made:

- In general, the results of the present study are in reasonable agreement with those from literature, both in terms of variability and its trend with period;
- At selected periods, range estimates are characterized by a high variability across different regions which is consistent with the one found in literature works (for example, the strong variability of range for PGA coming from Jayaram and Baker[6]).

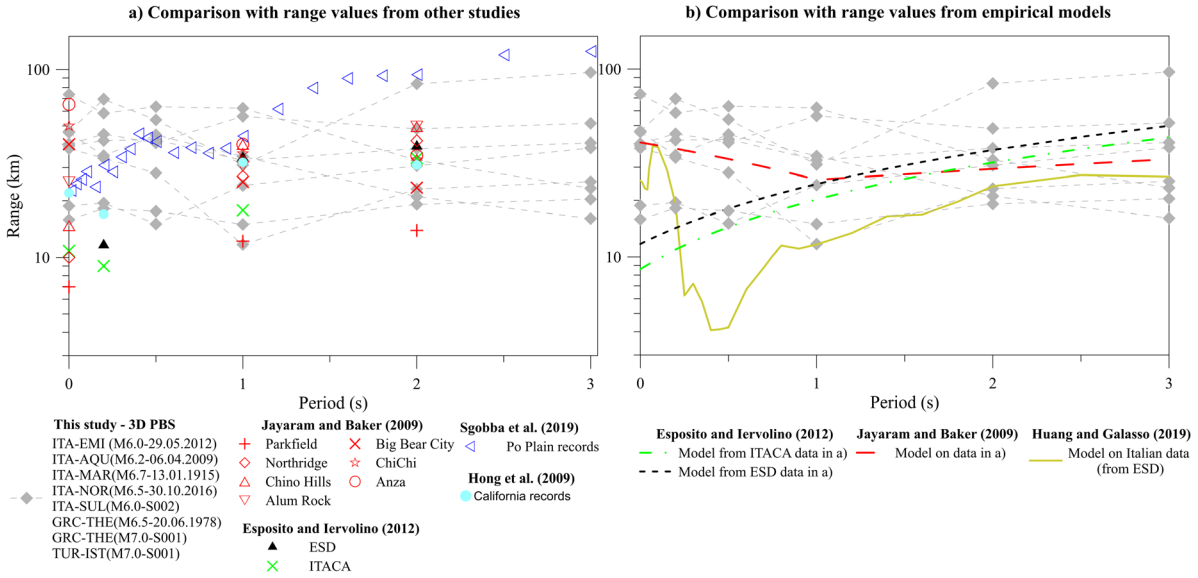

FIGURE 7 Comparison between the ranges (GMH component) estimated in this study (3D PBS, filled diamonds) and those from literature studies as: (a) estimates on specific recorded datasets, (b) predictions by empirical correlation models

- Such differences may be related to different factors affecting the earthquake ground motion, ranging from the frequency content of ground motion, source, path effects and local geology.[14] The dependence of the correlation structure on these factors will be discussed in detail in the following section.

6 | IMPACT OF SOURCE AND PATH EFFECTS ON SPATIAL CORRELATION

The aim of this Section is to investigate in a systematic way the impact of various factors on spatial correlation, that is, ground-motion directionality, source-related effects, such as magnitude and source directivity, and path effects.

6.1 | Ground-motion directionality: fault normal versus fault parallel components

Typically, empirical spatial correlation models are provided for the GMH component, without any distinction between the horizontal components. The dependence of the spatial correlation structure on the horizontal component of earthquake ground motion is addressed here with reference to the components parallel and normal to the causative fault (FP and FN, respectively). This aspect is relevant because the considered ground shaking scenarios are in the near-source region, where the polarization of ground motion may be significant with large FN/FP ratios.[46,47]

Figure 8 shows the ratio of the range (left, a) and sill (right, b) for the FN over that for the FP for six representative scenarios. It is clear that the FN/FP range and sill ratios tend to be systematically larger than 1 across all periods. This means that the FN motions tend to be correlated over larger distances, because, in near-fault conditions, they are more affected by source directivity effects, with longer period content and larger peak amplitudes (with distinctive pulse-like features), leading to a more coherent ground-motion distribution. A similar trend is found for the sill, with higher values for FN components, consistently with the range FN/FP behavior: this means that an intrinsic higher variance is associated with FN components owing to the stronger influence of the slip asperities on the fault plane.

6.2 | Earthquake magnitude

The availability, for the same region, of different scenarios from the same fault but with different magnitude allows for the study of the possible relation between M_W and spatial correlation (range). As an explanatory case, Figure 9 illustrates,

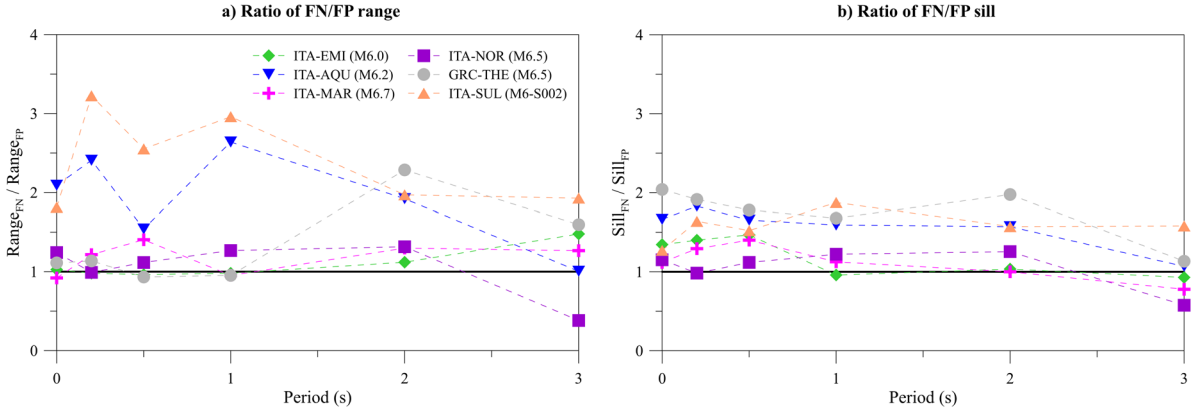

FIGURE 8 Ratio of FN/FP range (a) and sill (b) for the PBSs

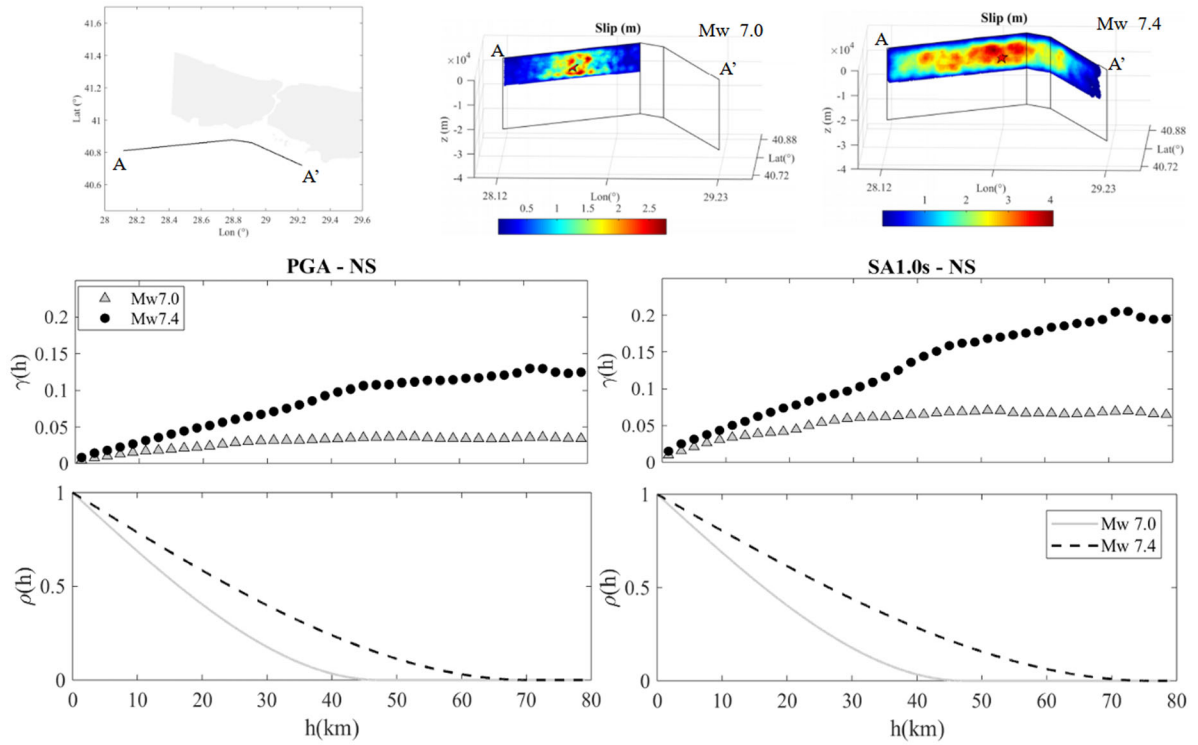

FIGURE 9 Istanbul case study. Top: location of the urban area with respect to the Marmara Sea segments of the NAF (left), fault rupture scenarios with M_W7.0 (center) and M_W7.4 (right). Bottom: comparison of the semivariogram and correlation coefficient for the M_W7.0 (triangles, solid lines) and M_W7.4 (dots, dashed lines) scenarios for PGA-NS (left) and SA1.0s-NS (right)

for the Istanbul region, the semivariogram and the correlation coefficient (ρ, see Equation (4)) of PGA-NS and SA1.0s-NS, evaluated for two rupture scenarios along the Marmara Sea segments of the North Anatolian Fault (NAF) system with $M_W = 7.0$ (top center) and $M_W = 7.4$ (top right).

A pronounced dependence on M_W is found, with larger range and sill for increasing magnitude, owing to the stronger energy content at low frequencies for larger events, which, in turn, yields to larger correlation. A positive correlation with M_W was also found by Sokolov et al.,[48] Sokolov and Wenzel[49] and Foulser-Piggott and Goda,[50] consistently with this study. However, other studies[14,6] did not find any clear evidence of correlation between magnitude and range, most likely because of the simultaneous presence in the recorded dataset of different factors, besides magnitude, affecting the spatial correlation.

In left panel of Figure 10, the range estimated for different M_W for the Sulmona (M_W = 6.0-6.5, left) and Istanbul (M_W = 7.0-7.4, right) regions are shown for a selected horizontal component (FN and NS for Sulmona and Istanbul,

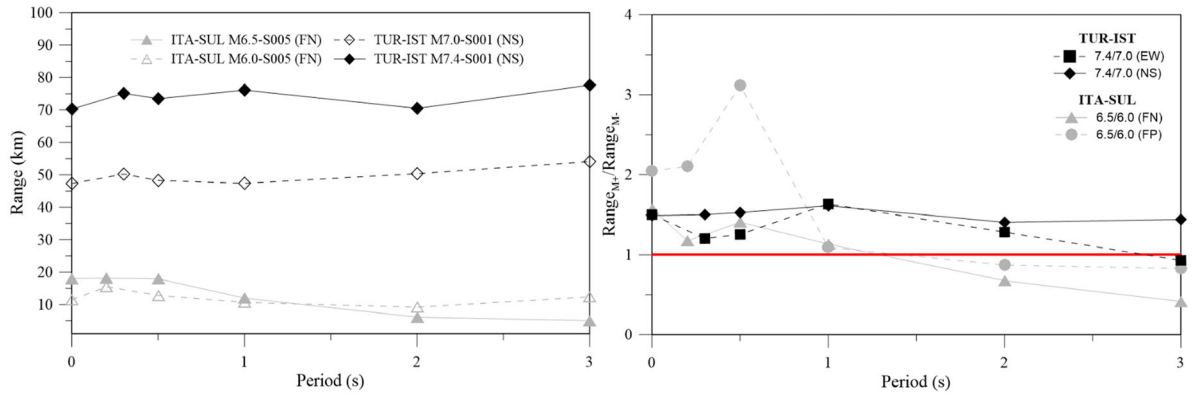

FIGURE 10 Left: range as a function of period for Sulmona FN component (triangles), M_W6.5 versus M_W6.0, and for Istanbul NS component (diamonds), M_W7.4 versus M_W7.0. For each case study, filled and open symbols are for larger and lower M_W respectively. Right: range ratios of larger over lower magnitude earthquakes (M_{W+}/M_{W-}) for both horizontal components

respectively). Note that the Thessaloniki scenarios are excluded from this comparison because they do not share the same causative fault. As previously noted, a positive correlation between M_W and range is found for both regions and for both horizontal components. This is further highlighted by the right panel of Figure 10 depicting the range ratios of larger over lower magnitude earthquakes (M_{W+}/M_{W-}) for both horizontal components. It is noted that for increasing periods the ratio tends to decrease (especially for the Sulmona case), most likely because of the coupling with other effects such as path or basin effects.

6.3 | Source directivity

We examine herein the case of three ground shaking scenarios in Istanbul area which are affected by different directivity features, namely, Forward Directivity (FD), Backward Directivity (BD), and Neutral Directivity (ND), depending on the relative position between the hypocenter and fault asperities with respect to the city of Istanbul. Note that the largest ground-motion amplitudes and directivity pulses are identified especially in the NS component, which approaches the FN component given the fault strike (see Figure 9), for the FD scenario, where the largest asperities are located along the pathway from the hypocenter to Istanbul (for a detailed discussion, see Infantino et al.[18]).

Figure 11 shows the map of SA3.0s-NS (top panel) for three selected scenarios exhibiting FD (left), BD (center), and ND (right) and the corresponding sample semivariograms (center panels) and correlation coefficients (bottom panels), for SA1.0s-NS, SA2.0s-NS, and SA3.0s-NS. Results indicate that the FD scenario is characterized by higher ranges and sills. The reason for these effects can be understood looking at the spatial distribution of SA3.0s: in the FD scenario, larger areas are illuminated by the rupture and are characterized by uniformly high ground-motion peak amplitudes. A clear trend cannot be found in the BD and ND scenarios, as in some cases the ND has larger ranges than BD.

6.4 | Path effects

The role of the propagation path from the causative fault is addressed here. To this end, the Thessaloniki model is considered (see Figure 12) for which two different scenarios from two faults, namely, the Gerakarou Fault ($M_W = 6.5$, left) and the Anthemountas Fault ($M_W = 7.0$, right) are available.

The comparison of the spatial correlation coefficients, for both PGA-GMH and SA1.0s-GMH, indicates larger correlations for the Gerakarou scenario, even though the latter is characterized by lower magnitude. In this case, we believe that the position of the fault coupled with the soil conditions plays a predominant role. As a matter of fact, the Gerakarou fault is located at around 20 km NE of Thessaloniki on homogeneous hard rock ($V_S = 2000$ m/s), while the Anthemountas fault is within the soft soils of the Anthemountas plain. Therefore, it is expected that coupling of fault rupture with soft soil conditions leads to more variable ground motions. This is further highlighted on the bottom right panel of Figure 12 where the range obtained for the Gerakarou and Anthemountas scenarios for GMH component is compared.

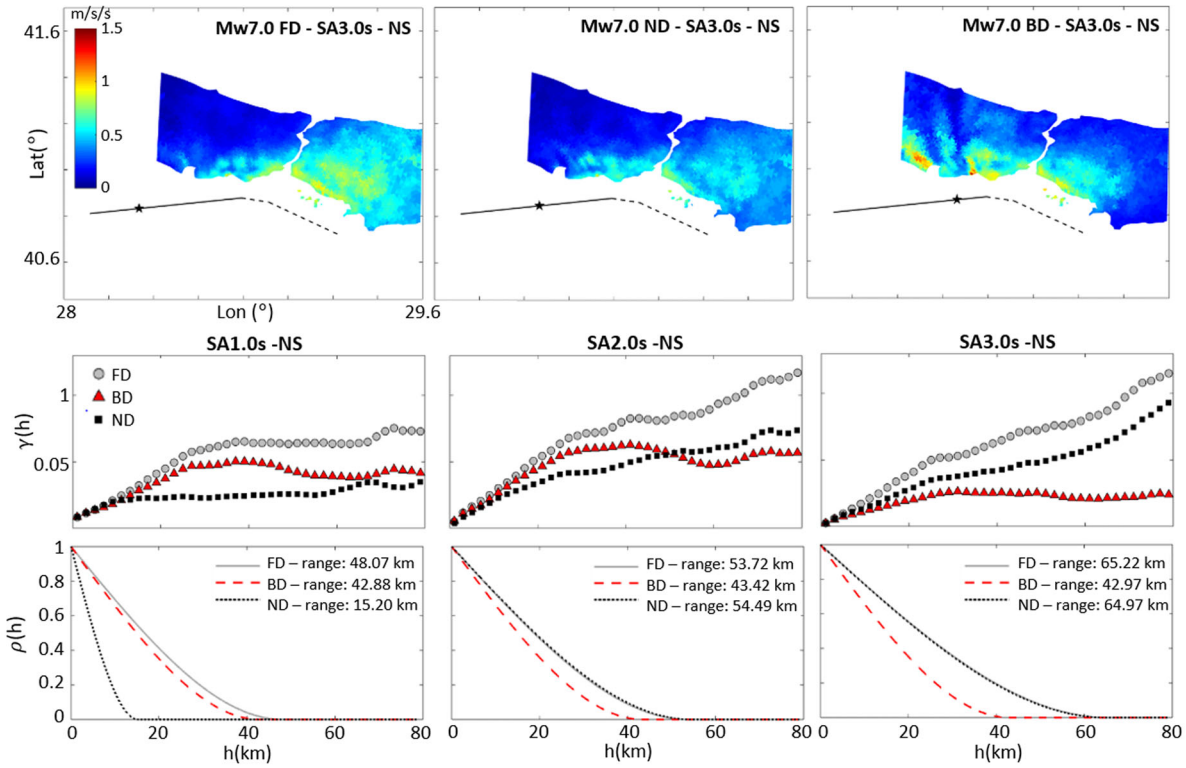

FIGURE 11 Comparison of M_W7.0 Istanbul ground shaking scenarios with Forward Directivity (FD), Neutral Directivity (ND), and Backward Directivity (BD): maps of SA3.0s-NS (top); sample semivariograms (center) and correlation coefficients (bottom) for SA1.0s-NS (left), SA2.0s-NS (center), and SA3.0s-NS (right)

7 | MAPS OF PEARSON'S CORRELATION COEFFICIENT

In this Section, the spatial correlation of PBS ground-motion residuals is estimated by means of the Pearson's correlation coefficient. Since this approach requires multiple rupture realizations of a given earthquake scenario, the Istanbul case study has been selected, for which a representative number of earthquake realizations with given M_W and fault rupture extent are available. More specifically, the considered earthquake scenarios (see distribution of hypocenters in Figure 13) are 20 scenarios of M_W7.4 and 11 of M_W7.0. The M_W7.4 ruptures involve the entire length of the multisegment portion of the NAF included in the numerical model (Marmara Sea segments), while the M_W7.0 ones, herein considered, only the west segment.

Thus, the Pearson's correlation coefficients between each pair of sites within the urban area of Istanbul can be computed by means of Equation (10). By selecting a reference site, the correlation coefficients between the chosen reference site and all the other sites have been estimated and displayed as Pearson's correlation coefficient map (Figures 14–16).

First of all, the effect of the geological conditions as well as of the propagation path is investigated by considering reference sites characterized by different soil properties and location with respect to the fault. Such a comparison is shown in Figure 14 which illustrates the Pearson's correlation maps for SA3.0s-GMH obtained by using the entire set of 20 M_W7.4 rupture realizations for three reference sites (from left to right, $V_{S30} = 400, 1020, 1430$ m/s). Moreover, below each Pearson's map the sites characterized by V_{S30} values similar to that of the reference site (within the range ±100 m/s) are highlighted. Comparing the three maps it is evident as the spatial pattern changes significantly as a function of the reference site. As expected, for all cases, the correlation decreases with the distance from the reference site but such a decay is not homogeneous in all the directions away from the reference site, as would expected in an isotropic case. Indeed, sites affected by similar source directivity phenomena and characterized by similar soil properties show higher correlation.

As a second analysis, Figure 15 shows the comparison of SA3.0s-GMH Pearson's maps obtained using the set of M_W7.4 (left) scenarios and the set of M_W7.0 (center) for the same reference site with $V_{S30} = 400$ m/s. The active fault for the two cases is indicated with a solid line (three segments for M_W7.4 and one segment for M_W7.0). The different spatial correlation distributions can be explained by the rupture propagation effect as a consequence of the fault extent and fault

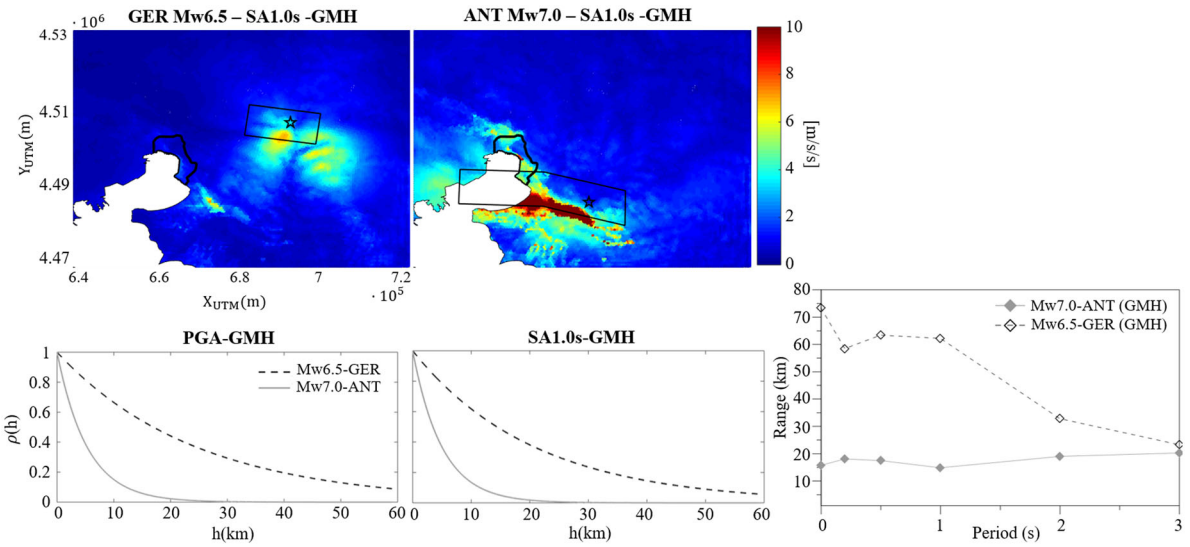

FIGURE 12 Top: map of SA1.0s-GMH for two rupture scenarios along two causative faults, Gerakarou (M_W6.5–GER, left) and Anthemountas (M_W7.0–ANT, right). The borders of Thessaloniki urban area are highlighted with a solid line. Bottom: correlation coefficients, PGA-GMH (left) and SA(1.0s)-GMH (center), and range as a function of period (right) for both scenarios

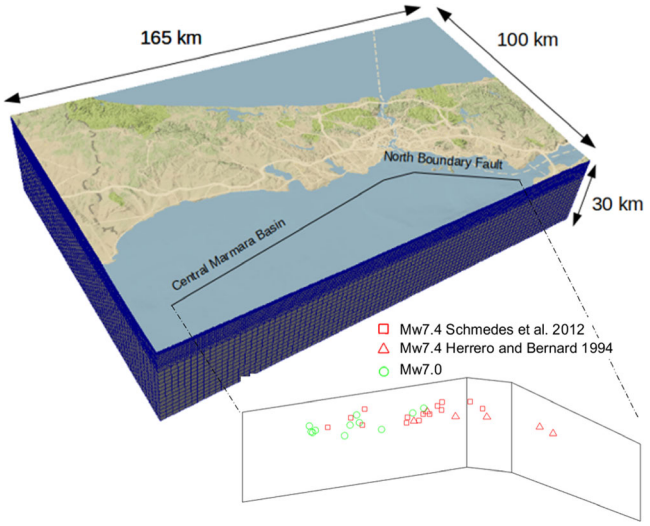

FIGURE 13 Three-dimensional numerical model of Istanbul and distribution of the hypocenters of the earthquake ruptures along the Marmara Sea segments of the NAF: circles M_W7.0 (11 ruptures), squares M_W7.4 generated according to the HB94 source model (6 ruptures), triangles M_W7.4 generated according to the SEA12 source model (14 ruptures)

geometry or orientation. Indeed, in the case of M_W7.0, for which only the west segment breaks, strong and concentrated waveforms propagate mainly along the west coast of the Bosphorus strait because of the fault orientation, leading to positive correlation in that region. On the other hand, for M_W7.4 the concave shape of the three-segmented active fault produces higher correlation with the reference site over the west region.

To further highlight the anisotropic features of the spatial correlation, on the right side of Figure 15 it is shown, for the same reference site, the correlation map obtained by using the semivariogram-based correlation model calibrated on ESD data by Esposito and Iervolino[11] based on the stationarity and isotropic assumptions. It is worth noting as such a model is magnitude independent and underestimate the spatial correlation extent for both magnitude levels.

Finally, Figure 16 illustrates, for the same reference site, the correlation maps obtained with the M_W7.4 scenarios generated by two different source model, HB94 (left) and SEA12 (right). Although in both cases, the reference site is positive correlated with the west region, it is found that for HB94 such correlation is almost perfect (= 1) while for SEA12, although positive, it is lower than 1 and more scattered (between 0.7 and 0.9). Moreover, for SEA12, the correlation decays slower

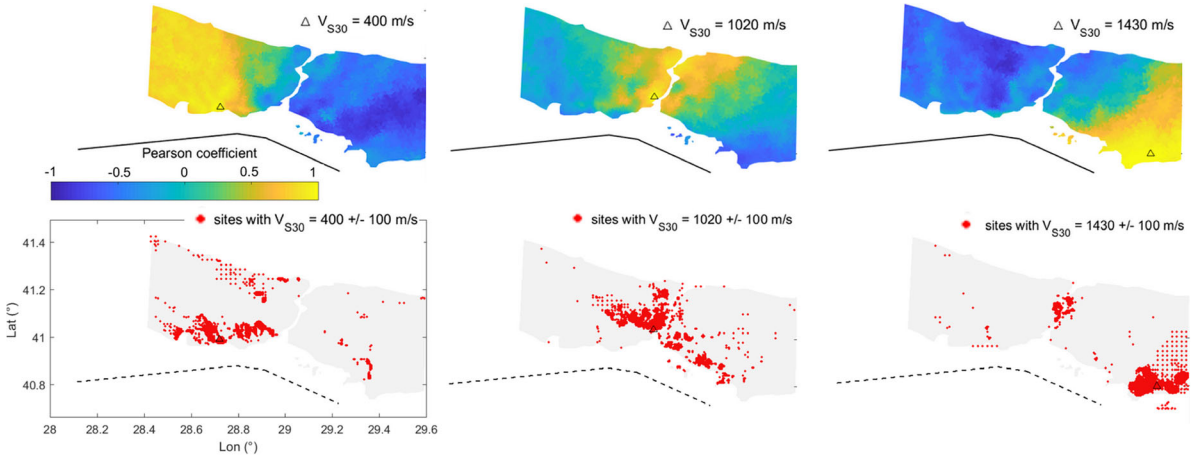

FIGURE 14 Top: Pearson's correlation coefficient maps for SA3.0s-GMH of 20 M_W7.4 earthquake ruptures along the NAF segments (solid line). The maps are shown for three reference receivers (black triangles) with different V_{S30} values: 400, 1020, and 1430 m/s (from the left to the right). Bottom: scattered plot of sites with values of V_{S30} similar ($\pm$100 m/s) to that of the reference site

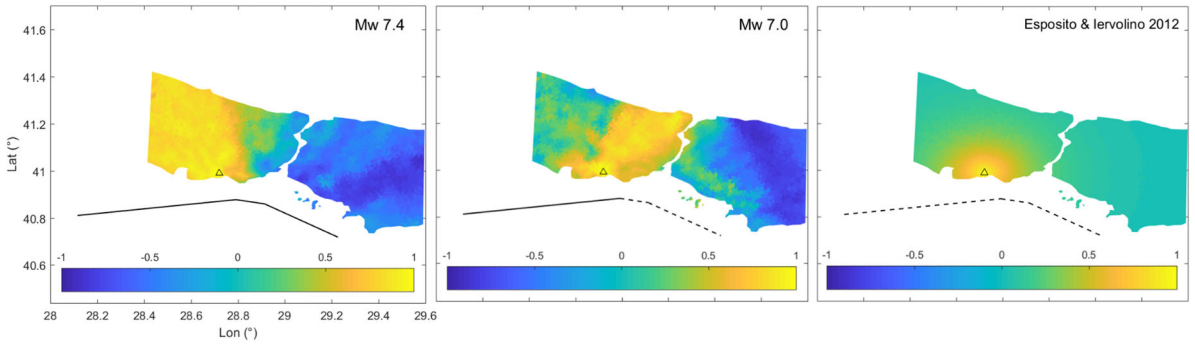

FIGURE 15 Pearson's correlation coefficient maps obtained with PBSs for SA3.0s-GMH of 20 M_W7.4 (left) and 11 M_W7.0 (center). The fault rupture extent considered for each case is indicated with a solid line. Right: correlation map obtained using the empirical correlation by Esposito and Iervolino[11]

with distance with respect to HB94. Such differences are likely due to the slip distribution generated by the two source models (see explanatory slip distribution plots in the bottom panel of Figure 16). The HB94 ruptures are characterized by a more concentrated asperity along the fault leading to larger correlation over a more limited area, while SEA12 ruptures produce "scattered" asperities implying a lower maximum degree of correlation along with a slower decay with the distance. As final remark, it can be observed that correlation coefficients are higher in the FN direction than the FP direction while sites on opposite ends of the ruptures are negatively correlated, as also found by Chen and Baker[20] on CyberShake simulations.

8 | CONCLUSIONS

In this work, we investigated the spatial correlation of broadband response spectral accelerations from 3D PBS for a wide set of rupture scenarios in seven regions worldwide, namely, Po Plain (Northern Italy), L'Aquila, Marsica, Sulmona, Norcia (Central Italy), Thessaloniki (Northern Greece), and Istanbul (Turkey). Geostatistical tools based on the calculation of both semivariogram and Pearson's correlation coefficient were applied to quantify the correlation in ground-motion intraevent residuals. Aspects related to the estimation of spatial coherency of ground motions, which have major implications in earthquake engineering applications related to spatially extended structures, are beyond the scope of this article but will be addressed in a future study, based on ongoing research activities.[51]

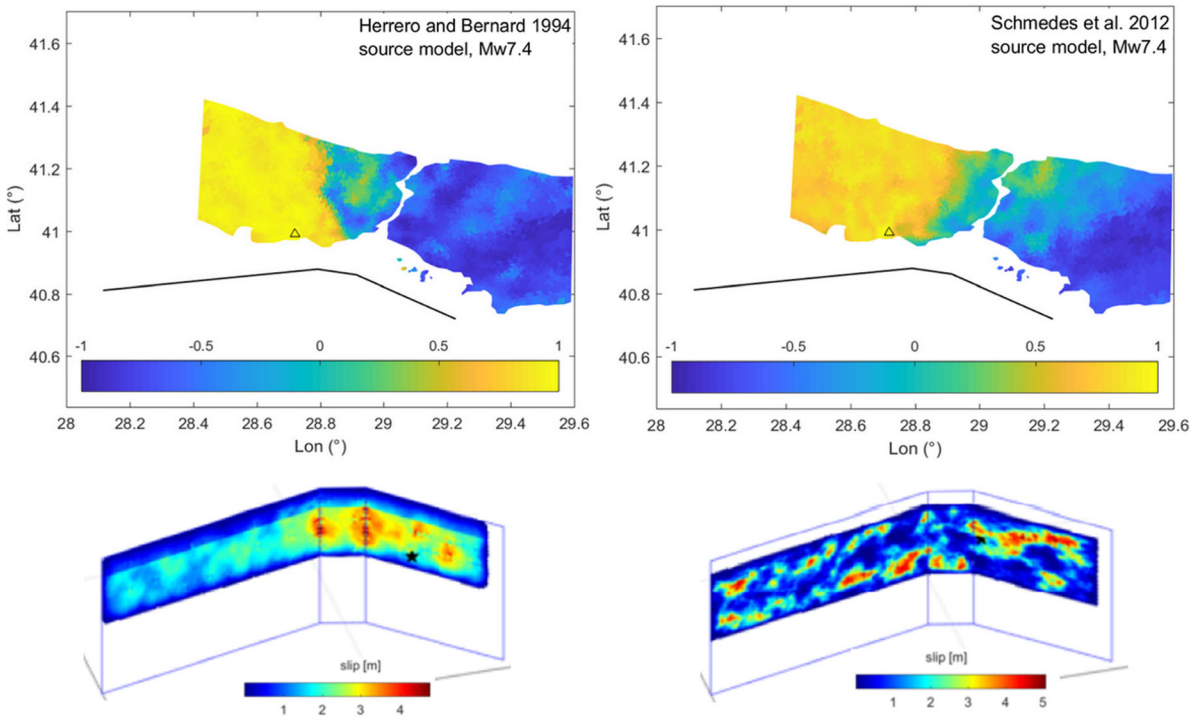

FIGURE 16 Top: Pearson's correlation coefficient maps for SA3.0s-GMH of M_W7.4 ruptures scenarios according to HB94 (left) and SEA12 (right) source model. Bottom: example of slip distributions for two scenarios according to HB94 (left) and SEA12 (right) source model

Two main achievements were reached with respect to, on one hand, the validation (in broad sense) of the broadband numerical approach in predicting spatial correlation by comparison with both recordings for specific events and independent empirical studies, and, on the other one, a systematic analysis of the impact of some relevant physical factors, such as magnitude, source directivity, propagation path, and ground-motion directionality, on the spatial correlation at different periods in near-source conditions. Such effects can be, in fact, hardly investigated using recordings due to their limited spatial sampling. The main conclusions can be summarized as follows.

Comparison with recordings and other empirical studies

- The comparison of the semivariograms computed from simulations and recordings for 33 stations in the Po Plain area during the May 29, 2012 earthquake indicate a very good agreement at both short and long periods, serving as a further validation of the ANN2BB approach to compute broadband ground motions at spatial scale.
- The range estimates from PBSs for the seven regions under consideration turn out to be comparable with the empirical models available in the literature, both in terms of trend with period and overall variability.
- At selected periods, range values are characterized by a high variability across different regions, pointing out that spatial variability of ground motion is region-specific.

Dependence on source, propagation path and geologic conditions

- *Ground-motion directionality*. FN components turn out to be characterized by higher ranges and higher sills at both short and long periods, indicating that FN motions tend to be correlated over larger distances and have an intrinsic larger variance, owing to the influence of source directivity effects.
- *Magnitude*. A clear positive correlation between M_W and range is found with larger ranges for increasing magnitude, owing to the stronger energy content at low frequencies, in agreement with other published works.[48–50]
- *Source directivity*. For the Istanbul scenarios (strike-slip fault), FD scenarios turn out to be characterized by the largest ranges especially at periods longer than 1 s and on fault-normal components. This is consistent with the observation that in FD conditions larger areas are illuminated by the rupture with more coherent "footprints" characterized by lower frequency content.[46]

- *Propagation path and geological conditions*. For the scenarios in Thessaloniki broader area, the causative fault combined with the local geologic conditions were found to play a major role, with larger range for rupture scenarios occurring in homogeneous, rock conditions.

Anisotropy and nonstationarity in spatial correlation

Although limited to the Istanbul case study and SA3.0s-GMH, the Pearson's correlation maps indicate anisotropy and nonstationarity in spatial correlation of SA, as also found by Chen and Baker[20]:

- Correlation is higher for sites with similar geologic conditions;
- sites at opposite ends of the rupture tend to be negatively correlated while sites along the rupture propagation path have stronger correlation.

Overall, this study points out that a single ergodic, stationarity, and isotropic spatial correlation model, calibrated on a large dataset including different regions and events, may not be suitable especially in near-source conditions. In this regard, PBSs can be used as a useful tool, complementary to empirical ground-motion models, for seismic hazard and risk assessment of large urban areas or infrastructural systems. More specifically, in our opinion, two main strategies can be adopted to this end. A first approach consists in the use of PBSs to calibrate region-, scenario-, and ground-motion component-specific spatial correlation models to be used in conjunction with current empirical ground-motion models. For instance, for a specific region, one could take advantage of a suite of PBSs to establish enhanced correlation models which may depend also on selected explanatory variables, such as magnitude, and may provide different estimates for the different ground-motion components. In this approach, the spatial correlation of ground motion is typically incorporated in seismic risk computations by generating spatially correlated random fields,[52] which, however, as a main limitation, rely on the hypothesis of isotropy and stationarity, properties which are hardly supported by data as shown in this work and previous studies.[53] A second approach, that would allow overcoming the aforementioned limitations, consists in using directly PBSs of a suitable set of rupture realizations from earthquake scenarios in an event-based approach to probabilistic seismic hazard and risk assessment relying on Monte Carlo simulations. The rationale behind this approach, referred to as *footprint-based* by Stupazzini et al.,[54] is to introduce the physics-based realizations of each scenario earthquake as multiple equally weighted branches within a logic tree framework. In such a way, the spatial correlation structure is naturally accounted for without the need of defining any spatial correlation model, which inevitably leads to a loss of information specific of each earthquake rupture. Future studies on this subject will help to quantify to which extent the specificity, anisotropy, and nonstationarity of the physics-based ground shaking scenarios may influence risk estimates with respect to standard approaches based on empirical ground-motion models combined with a spatial correlation model for intraevent residuals.

As a concluding remark, we recall that, for a complete description of ground-motion correlation suitable for risk analyses on heterogeneous portfolios, the spatial cross-correlation among residuals of different IMs on spatially distributed sites should be also addressed.[55,13]

ACKNOWLEDGMENTS

This work has been funded by the URBASIS Project "New challenges for Urban Engineering Seismology," H2020-MSCA-ITN-2018, Grant Agreement No. 813137. The authors are grateful to Roberto Paolucci for his fruitful suggestions during the preparation of the work. The authors wish to thank Sinan Akkar and an anonymous reviewer for their constructive remarks that helped to improve the quality of the article.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

1. Weatherill G, Silva V, Crowley H, Bazzurro P. Exploring the impact of spatial correlations and uncertainties for portfolio analysis in probabilistic seismic loss estimation. *Bull Earthquake Eng*. 2015;13(4):957-981.
2. Park J, Bazzurro P, Baker JW. Modeling spatial correlation of ground motion intensity measures for regional seismic hazard and portfolio loss estimation. *Applications of Statistics and Probability in Civil Engineering*. London: Taylor Francis; 2007:1-8.

3. Goda K, Atkinson GM. Probabilistic characterization of spatially correlated response spectra for earthquakes in Japan. *Bull Seismol Soc Am*. 2009;99:3003-3020.
4. Garakaninezhad A, Bastami M. Intra-event spatial correlation model for the vertical component of response spectral accelerations. *J Seismol*. 2019;23:853-867.
5. Wang M, Takada T. Macrospatial correlation model of seismic ground motions. *Earthquake Spectra*. 2005;21:1137-1156.
6. Jayaram N, Baker JW. Correlation model for spatially distributed ground motion intensities. *Earthquake Eng Struct Dyn*. 2009;38:1687-1708.
7. Goda K, Hong HP. Spatial correlation of peak ground motions and response spectra. *Bull Seismol Soc Am*. 2008;98:354-365.
8. Goda K, Atkinson GM. Intra-event spatial correlation of ground-motion parameters using SK-net data. *Bull Seismol Soc Am*. 2010;100:3055-3067.
9. Sokolov V, Wenzel F, Kuo-Liang W. Uncertainty and spatial correlation of earthquake ground motion in Taiwan. *Terr Atmos Ocean Sci*. 2010;21:905-921.
10. Esposito S, Iervolino I. PGA and PGV spatial correlation models based on European multievent datasets. *Bull Seismol Soc Am*. 2011;101:2532-2541.
11. Esposito S, Iervolino I. Spatial correlation of spectral acceleration in European data. *Bull Seismol Soc Am*. 2012;102:2781-2788.
12. Heresi P, Miranda E. Uncertainty in intraevent spatial correlation of elastic pseudo-acceleration spectral ordinates. *Bull Earthquake Eng*. 2019;17:1099-1115.
13. Huang C, Galasso C. Ground-motion intensity measure correlations observed in Italian strong-motion records. *Earthquake Eng Struct Dyn*. 2019;48:1634-1660.
14. Schiappapietra E, Douglas J. 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*. 2020;203:103-109.
15. Mazzieri I, Stupazzini M, Guidotti R, Smerzini C. SPEED-spectral elements in elastodynamics with discontinuous galerkin: a non-conforming approach for 3D multi-scale problems. *Int J Numer Methods Eng*. 2013;95:991-1010.
16. Smerzini C, Pitilakis K. Seismic risk assessment at urban scale from 3D physics-based numerical modeling: the case of Thessaloniki. *Bull Earthquake Eng*. 2018;16:2609-2631.
17. Paolucci R, Mazzieri I, Piunno G, Smerzini C, Vanini M, Özcebe AG. Earthquake ground motion modeling of induced seismicity in the Groningen gas field. *Earthquake Eng Struct Dyn*. 2021;50:135-154.
18. Infantino M, Mazzieri I, Özcebe AG, Paolucci R, Stupazzini M. 3D physics-based numerical simulations of ground motion in Istanbul from earthquakes along the Marmara segment of the North Anatolian Fault. *Bull Seismol Soc Am*. 2020;110:2559-2576.
19. Paolucci R, Gatti F, Infantino M, Smerzini C, Özcebe AG, Stupazzini M. Broadband ground motions from 3D physics-based numerical simulations using artificial neural networks. *Bull Seismol Soc Am*. 2018;108:1272-1286.
20. Chen Y, Baker JW. Spatial correlations in CyberShake physics-based ground-motion simulations. *Bull Seismol Soc Am*. 2019;109(6):2447-2458.
21. Herrero A, Bernard P. A kinematic self-similar rupture process for earthquakes. *Bull Seismol Soc Am*. 1994;84:1216-1229.
22. Schmedes J, Archuleta RJ, Lavallée D. A kinematic rupture model generator incorporating spatial interdependency of earthquake source parameters. *Geophys J Int*. 2012;192:1116-1131.
23. Smerzini C, Galasso C, Iervolino I, Paolucci R. Ground motion record selection based on broadband spectral compatibility. *Earthquake Spectra*. 2014;30:1427-1448.
24. Ancheta TD, Darragh RB, Stewart JP, et al. NGA-West2 database. *Earthquake Spectra*. 2014;30:989-1005.
25. Evangelista L, Del Gaudio S, Smerzini C, et al. Physics-based seismic input for engineering applications: a case study in the Aterno River valley, Central Italy. *Bullet Earthquake Eng*. 2017;15:2645-2671.
26. Villani M, Faccioli E, Ordaz M, Stupazzini M. High-Resolution seismic hazard analysis in a complex geological configuration: the case of the Sulmona basin in central Italy. *Earthquake Spectra*. 2014;30(4):1801-1824.
27. Paolucci R, Mazzieri I, Smerzini C. Anatomy of strong ground motion: near-source records and 3D physics-based numerical simulations of the Mw 6.0 2012 May 29 Po Plain earthquake, Italy. *Geophys J Int*. 2015;203:2001-2020.
28. Paolucci R, Evangelista L, Mazzieri I, Schiappapietra E. The 3D numerical simulation of near-source ground motion during the Marsica earthquake, central Italy, 100 years later. *Soil Dyn Earthquake Eng*. 2016;91:39-52.
29. Özcebe AG, Smerzini C, Paolucci R, et al. On the comparison of 3D, 2D, and 1D numerical approaches to predict seismic site amplification: the case of Norcia basin during the M6.5 2016 October 30 earthquake. Proceedings of the 7th International Conference on Earthquake Geotechnical Engineering, Rome, Italy, 17-20 June 2019.
30. Smerzini C, Pitilakis K, Hashemi K. Evaluation of earthquake ground motion and site effects in the Thessaloniki urban area by 3D finite-fault numerical simulations. *Bull Earthquake Eng*. 2017;15:787-812.
31. Smerzini C, Cavalieri F, Argyroudis S, Pitilakis K, 3D physics-based numerical modeling as tool for seismic risk assessment of urban infrastructural systems: the case of Thessaloniki, Greece. 16th European Conference on Earthquake Engineering, Thessaloniki, Greece, 18–21 June 2018.
32. Zimmerman DL, Stein M. *Classical Geostatistical Methods. Handbook of Spatial Statistics*. Boca Raton, FL: CRC Press; 2010.
33. Loth C, Baker JW. A spatial cross-correlation model for ground motion spectral accelerations at multiple periods. *Earthquake Eng Struct Dyn*. 2013;42:397-417.
34. Marinoni O. Improving geological models using a combined ordinary-indicator kriging approach. *Eng Geol*. 2003;69:37-45.
35. Zhang Q, Zhu H. Collaborative 3D geological modeling analysis based on multi-source data standard. *Eng Geol*. 2018;246:233-244.

36. Matheron G. *Les Variables Régionalisées et Leur Estimation: Une Application de la Théorie Des Fonctions Aléatoires Aux Sciences de la Nature*. Paris: Masson; 1965:305.
37. Cressie N, Hawkins DM. Robust estimation of the variogram: I. *J Int Assoc Mathematicl Geol*. 1980;12:115-125.
38. Journel AG, Huijbregts CJ. *Mining Geostatistics*. London: Academic Press; 1978:600.
39. Ming D, Huang C, Peters GW, Galasso C. An advanced estimation algorithm for ground-motion models with spatial correlation. *Bull Seismol Soc Am*. 2019;109:541-566.
40. Oliver MA, Webster R. A tutorial guide to geostatistics: computing and modelling variograms and kriging. *Catena*. 2014;113:56-69.
41. Baker JW, Chen Y. Ground motion spatial correlation fitting methods and estimation uncertainty. *Earthquake Eng Struct Dyn*. 2020;49:1662-1681.
42. Sgobba S, Lanzano G, Pacor F, et al. Spatial correlation model of systematic site and path effect for ground-motion fields in northern Italy. *Bull Seismol Soc Am*. 2019;109:1419-1434.
43. Zerva A, Zervas V. Spatial variation of seismic ground motions: an overview. *Appl Mech Rev*. 2002;55:271-297.
44. Der Kiureghian A. A coherency model for spatially varying ground motions. *Earthquake Eng Struct Dyn*. 1996;25:99-111.
45. Hong HP, Zhang Y, Goda K. Effect of spatial correlation on estimated ground-motion prediction equations. *Bull Seismol Soc Am*. 2009;99:928-934.
46. Sommerville P, Smith NF, Graves RW, Abrahamson NA. Modification of empirical strong ground motion attenuation relations to include the amplitude and duration effects of rupture directivity. *Seismol Res Lett*. 1997;68:199-222.
47. Pacor F, Felicetta C, Lanzano G, et al. NESS1: a worldwide collection of strong-motion data to investigate near-source effects. *Seismol Res Lett*. 2018;89:2299-2313.
48. Sokolov V, Wenzel F, Wen KL, Jean WY. On the influence of site conditions and earthquake magnitude on ground-motion within-earthquake correlation: analysis of PGA data from TSMIP (Taiwan) network. *Bull Earthquake Eng*. 2012;10:1401-1429.
49. Sokolov V, Wenzel F. 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 Earthquake Eng*. 2013;11:1909-1926.
50. Foulser-Piggott R, Goda K. Ground-motion prediction models for arias intensity and cumulative absolute velocity for Japanese earthquakes considering single-station sigma and within-event spatial correlation. *Bull Seismol Soc Am*. 2015;105:1903-1918.
51. Smerzini C, Spatial variability of earthquake ground motion from 3D physics-based numerical simulations. Proceedings of the 16th European Conference on Earthquake Engineering, Thessaloniki, Greece, 2018, pp. 12.
52. Crowley H. Earthquake risk assessment: present shortcomings and future directions. In: Ansal A, ed. *Perspectives on European Earthquake Engineering and Seismology. Geotechnical, Geological and Earthquake Engineering*. Switzerland: Springer; 2014.
53. Garakaninezhad A, Bastami M. A novel spatial correlation model based on anisotropy of earthquake ground-motion intensity. *Bull Seismol Soc America*. 2017;107:2809-2820.
54. Stupazzini M, Infantino M, Allmann A, Paolucci R. Physics-based probabilistic seismic hazard and loss assessment in large urban areas: a simplified application to Istanbul. *Earthquake Eng Struct Dyn*. 2021;50:99-115.
55. Baker JW, Jayaram N. Correlation of spectral acceleration values from NGA ground motion models. *Earthquake Spectra*. 2008;24(1):299-317.

How to cite this article: Infantino M, Smerzini C, Lin J. Spatial correlation of broadband ground motions from physics-based numerical simulations. *Earthquake Engng Struct Dyn*. 2021;50:2575–2594.
https://doi.org/10.1002/eqe.3461