



Simulation of seismic ground motion fields via object-oriented spatial statistics with an application in Northern Italy

Alessandra Menafoglio¹ · Sara Sgobba² · Giovanni Lanzano² · Francesca Pacor²

Published online: 4 September 2020
© The Author(s) 2020

Abstract

This work offers a novel methodological framework to address the problem of generating data-driven earthquake shaking fields at different vibration periods, which are key to support decision making and civil protection planning. We propose to analyse the entire profiles of spectral accelerations and project their information content to unsampled locations in the system, based on the theory of Object Oriented Spatial Statistics. The proposed methodology combines a non-ergodic ground motion model with a fully functional model for the residual term, the latter consisting of (i) the spatially-varying systematic effects due to source, site and path, and (ii) the remaining aleatory error. The proposed methodology allows to generate multiple shaking scenarios conditioned on the data, jointly and consistently for all the vibration periods, overcoming the intrinsic limitations of existing multivariate approaches to the problem. The approach is tested on a vast dataset of ground motion records collected in the study-area of the Po Plain (Northern Italy), for which a region-specific fully non-ergodic GMM was previously calibrated. Our validation tests demonstrate the potentiality of the approach, which is capable to effectively simulate spectral acceleration profiles, while keeping the ability to capture the main physical features of ground motion patterns in the region.

Keywords Uncertainty quantification · Functional data · Ground motion models · Shaking fields

1 Introduction

Seismic shaking maps are tools to support decision making at a given site and a key-topic for civil protection planning and engineering purposes such as for loss assessment and risk analysis. Currently, empirical approaches adopted to simulate seismic shaking fields are based on the use of ground motion models (GMMs), which allow to estimate seismic intensity measures dependent on the vibration period as a function of several parameters related to the

reference earthquake scenario (magnitude, distance, soil category, etc.).

One of the main hypotheses in the formulation of GMMs is the ergodicity, which states that the variability in ground motion at a single site-source combination is the same as the variability in ground motion observed in a more global dataset (Anderson and Brune 1999; Al-Atik et al. 2010). In recent years, the increasing availability of strong-motion data has allowed to relax this assumption and to build non-ergodic GMMs, see, e.g., Rodriguez-Marek et al. (2013), Villani and Abrahamson (2015), Baltay et al. (2017), Lanzano et al. (2017), and Kuehn et al. (2019). These GMMs, together with the associated variability, enable one to describe the effects of earthquakes on a regional scale, rather than at global level as in ergodic GMMs. The key to removing the ergodic assumption grounds on the decomposition of model residuals into systematic corrective terms, computed from repeated sampling of the site, path, and source effects, gathered from multiple events and multiple stations. Once computed, these repeatable effects are typically considered as adjustment terms of the median GMM prediction, which

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00477-020-01847-4>) contains supplementary material, which is available to authorized users.

✉ Alessandra Menafoglio
alessandra.menafoglio@polimi.it

¹ MOX, Department of Mathematics, Politecnico di Milano, Milan, Italy

² INGV Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Milano, Milan, Italy

allow one to move (and reduce) part of the aleatory variability into epistemic uncertainty (Anderson and Uchiyama 2011; Villani and Abrahamson 2015; Kotha et al. 2016; Lanzano et al. 2017).

Another common finding in the literature is that the residuals of a GMM are characterized by a non-negligible spatial correlation (Jayaram and Baker 2010; Lin et al. 2011; Walling 2009), which can be modelled through geospatial analysis and exploited to predict shaking intensities at unobserved sites with the aim to produce shaking scenarios. This approach has been proposed for instance for well-sampled regions in California (Landwehr et al. 2016; Sahakian et al. 2019; Kuehn et al. 2019) and Italy (Lanzano et al. 2018; Sgobba et al. 2019). In these contexts, the method adopted to model the spatial correlation is of paramount importance, as it is found that different approaches can lead to significantly different results on the ground motion spatial distribution (Sokolov and Wenzel 2011). The majority of the previous studies are based on univariate applications of the traditional spatial interpolation techniques, such as kriging or conditional Gaussian simulations (Jayaram and Baker 2010 and others). More recent studies have focused on multivariate geostatistical approaches (Goda and Hong 2008; Goda and Atkinson 2010; Loth and Baker 2013; Verros and Ganesh 2017; Worden et al. 2018; Huang and Galasso 2019), which allow predicting the intensity measures at multiple periods and at different sites. An example is related to ShakeMap (Worden and Wald 2016; Michelini et al. 2019), an online tool where shaking maps are made available in near-real time for a discrete set of spectral ordinates, thus requiring to predict the unobserved intensities both at unsampled locations and at different spectral periods. These multivariate approaches are all based on the cross-variograms of the spatially distributed measures, thus inevitably suffering from the curse of dimensionality as the number of considered periods increases (Weatherill et al. 2015).

The present work aims to overcome the intrinsic limitations of existing approaches for multivariate scenario-based shaking predictions, by grounding on the viewpoint of functional data analysis (FDA, Ramsay and Silverman 2005). The latter framework allows for the statistical analysis of datasets of curves, by generalizing the methods typical of multivariate statistics to the infinite-dimensional setting. In FDA, the curse of dimensionality is overcome by considering the data as elements of a mathematical space (e.g., the functional space L^2 of square-integrable function), whose geometrical structure should be chosen as to capture the key features of the data (Ferraty and Vieu 2006). This work thus stems from the need to develop advanced probabilistic and statistical tools able to appropriately model and account for the complex sources of

uncertainty associated with seismic hazard—as thoroughly discussed by Porcu et al. (2017) (and references therein). In this vein, the focus of this work shall be on the population of spectral acceleration profiles $\{SA_s(T), T \in \mathcal{T}\}$ — T denoting the vibration periods ranging in an interval \mathcal{T} and s the geographic location—rather than on the vectors of their discrete evaluations $(SA_s(T_1), \dots, SA_s(T_p))$. Grounding on the theory of object-oriented spatial statistics (O2S2, see e.g., Menafoglio and Secchi 2017, 2019), we here extend to the functional setting the modeling approach of Sgobba et al. (2019), in which the period-dependent source, site and path systematic residuals of the reference GMM are assumed to be varying with the geographical location, and characterized by non-negligible spatial correlation. A functional characterization of the dependence among spectral acceleration profiles is here used to provide data-driven simulations of shaking scenarios (Fig. 1), which are key for several engineering applications (e.g., Hacıefendioğlu and Alpaslan 2014; Hacıefendioğlu et al. 2015). In this conceptual framework, the present study takes advantage of the large availability of seismic records in Northern Italy to build a novel model for the spectral profiles in the region, based on a non-ergodic GMM developed by Lanzano et al. (2017) for the same area.

The remaining part of the work is organized as follows. Section 2 recalls the seismological background and the data considered in the present study. Section 3 develops the model used for the spectral acceleration profiles. Section 4 describes the results of the modeling framework when applied to the field case here considered, whereas Sect. 5 reports an analysis for the model validation. Section 6 presents a comparison of the proposed model with state-of-the-art univariate and multivariate approaches. The proposed strategy of analysis is tested on an independent event in Sect. 7. Section 8 finally concludes the paper. The Supplementary Material presents further analyses and validation.

2 Data and model

2.1 Basic definitions

In the field of engineering seismology, the effect of the earthquake shaking is generally described through empirical GMMs, which quantify the conditional distribution of a ground motion intensity measure IM (e.g. peak ground acceleration, PGA or spectral acceleration, SA), given a set of explanatory parameters (magnitude, distance or local shear wave velocity) related to a given earthquake event e and an observing site s in a study area \mathcal{D} . The ground motion parameter is typically assumed to be log-normally

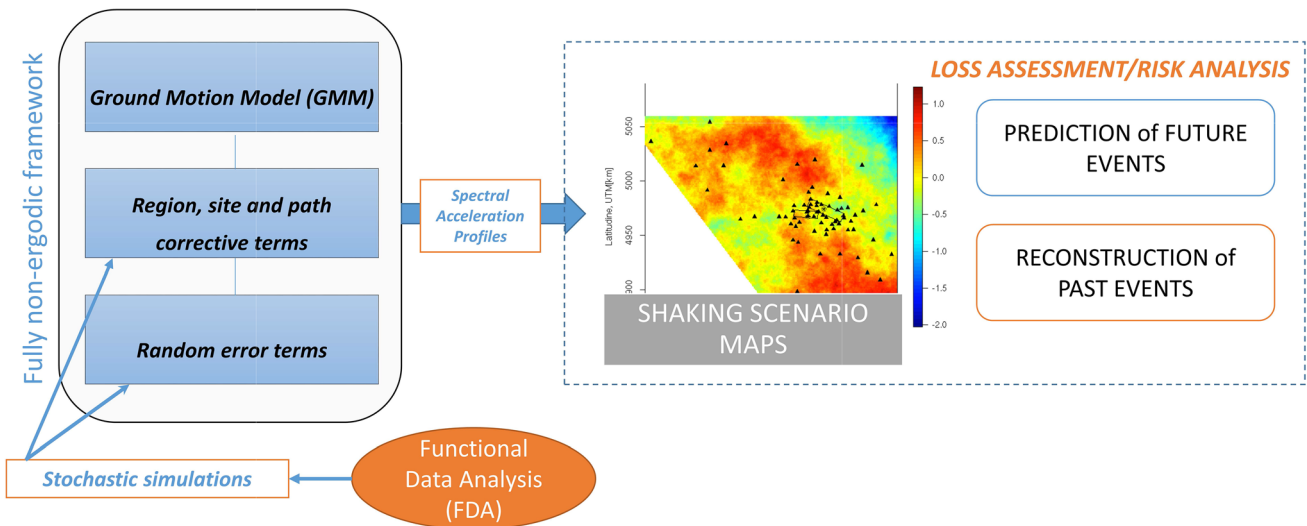


Fig. 1 Schematic representation of the proposed approach

distributed (i.e., $\log_{10}(IM_{es}(T))$ is derived from a normal distribution) and, in a (fully) non-ergodic framework, it is expressed as

$$\log_{10}(IM_{es}(T)) = \mu_{es}(T) + \mathcal{X}_{rs}(T) + \delta B_e(T) + \epsilon_s(T), \tag{1}$$

where $\mu_{es}(T)$ is the (deterministic) mean of the logarithm of the IM_{es} predicted by the GMM at the period T , and the other terms are, following the notation by Al-Atik et al. (2010), (i) a (random) spatially-varying corrective term $\mathcal{X}_{rs}(T)$ describing systematic effects of ground motion related to the seismogenic source region r , (ii) the between-event term $\delta B_e(T)$, which is a (random) spatially constant model bias, and (iii) a (random) remaining term $\epsilon_s(T)$ (i.e. the residuals corrected for event, region, site and path effects). For any given T in \mathcal{T} , the terms $\mathcal{X}_{rs}(T)$, $\delta B_e(T)$, $\epsilon_s(T)$ are assumed to be mutually independent and Gaussian. Note that, in (1), the systematic component of the (log) IM_{es} is described by the deterministic term $\mu_{es}(T)$ and by the random terms $\mathcal{X}_{rs}(T)$, $\delta B_e(T)$ —characterized by *epistemic* uncertainty. Here, the latter is intended as the uncertainty due to incomplete information or incomplete knowledge of the earthquake process. The remaining term $\epsilon_s(T)$ captures a purely *aleatory* (i.e., not-systematic) variability, instead, which is intended as the inherent randomness of ground motion. In this sense, the non-ergodic approach allows one to move part of the aleatory variability to (random) systematic components characterized by epistemic uncertainty only. In this framework, the corrective term \mathcal{X}_{rs} appearing in (1) is obtained upon decoupling the total residual into independent, zero-mean Gaussian terms describing site, path and source effects, as

$$\mathcal{X}_{rs}(T) = \delta S2S_s(T) + \delta P2P_{sr}(T) + \delta L2L_r(T). \tag{2}$$

Here the site-to-site term $\delta S2S_s$ quantifies the average misfit of ground motion at the site, with respect to the event-corrected median value predicted by the GMM. The location-to-location term $\delta L2L_r$ indicates how the ground motion of an event e recorded in a given small seismogenic region r differs from the mean prediction of the events in the entire source region. The path-to-path term $\delta P2P_{sr}$ represents how the specific characteristics of a travel path lead to ground motions that are systematically different from the ground motion predicted by the GMM. It is calculated as the mean of the event- and site-corrected residuals from earthquakes recorded at the station s with reference to the region r . One may also quantify the *aleatory* standard deviation associated with the non-ergodic IM_{es} prediction of Eq.(1), which is that of the remaining term $\epsilon_s(T)$, and is denoted in previous works by σ (Sgobba et al. 2019). We refer the reader to Lanzano et al. (2017) for a full account of the modeling and estimation frameworks used in previous works. In Sect. 3 we shall however discuss the compatibility of our model assumptions with those used in previous works.

2.2 Dataset and assumptions

The present study aims to improve the existing methodologies on the generation of empirical multiple scenario-based realizations, by taking into account the entire functional form of the intensity measures $\log_{10}(IM_{es}(T))$, $T \in \mathcal{T}$ (Fig. 1) defined in (1). For this purpose, we consider a proper region-specific GMM and a corresponding set of corrective terms, which are used to adjust the median predictions of IM_{es} .

We adopt the non-ergodic GMM developed by Lanzano et al. (2016) (named NI15), which is specifically calibrated

on the area \mathcal{D} covering the main source regions of Northern Italy (Fig. 2): the model predicts as IM_{es} , the geometric mean of the horizontal components of PGAs and the 5%-damping response spectral ordinates (SA) for 25 fundamental natural periods (T) ranging within the interval $\mathcal{T} = [0.04, 4]s$ (log10 units). The use of a spectral acceleration-based intensity measure IM_{es} comes from its relevance in the engineering practice, since it provides a suitable static force for the design of structures under seismic loading. The acceleration response spectrum SA, represents indeed the maximum response of a single-degree-of-freedom system of oscillators, characterised by different periods T and damping values, and subjected to an earthquake ground motion input. In the following, the IMs of Eq. (1) will refer to PGA and SA, and denoted by $SA(T)$, for T in \mathcal{T} .

The NI15 model was calibrated on a highly dense dataset compiled for the Po-Plain area (i.e. more than 2000 records triggered by 71 stations inside an area of 5000 km^2) after the occurrence of the 2012 seismic sequence (M_w 6.1, 20/05/2012 Emilia 1st mainshock and M_w 6.0, 29/05/2012 Emilia 2nd mainshock). This large availability of records allowed the previous authors to develop the fully non-ergodic NI15 model, and to perform residual decomposition in a non-ergodic framework.

The corrective terms $\mathcal{X}_{rs}(T)$ are taken from the work by Sgobba et al. (2019), who used Eq. (1) based on NI15 model, on a regular grid across a small Po-Plain area, to develop scenario-based shaking fields at 25 discrete periods T in $\mathcal{T} = [0.04, 4]s$. These authors considered a univariate geostatistical approach for the simulation of the marginal fields of corrective terms $\mathcal{X}_{rs}(T)$ (see also Sect. 6 for a more detailed comparison). Such a dataset of \mathcal{X}_{rs} includes the $\delta L2L_r$ region terms and the $\delta P2P_{sr}$ path terms related to

a single source zone (ZS912—according to the Italian hazard map MPS04, Meletti et al. (2008)—that is the most contributing inside the area considered for calibrating NI15 and the one generating the earthquakes of the 2012 seismic sequence), thus resembling a single-source single-path set. For this reason, the subscript r appearing in \mathcal{X}_{rs} will be dropped in the following notation. The term δB_e was also estimated by Sgobba et al. (2019), and is here considered as fixed, given that the analysis here presented refers to the same datasets of events of Sgobba et al. (2019); otherwise, analysis of the term δB_e could follow a very analogous pipeline as that detailed in this work.

Finally, the remaining term ϵ_s of Eq. (1) are herein estimated by difference with respect to the real observed values of $\log_{10}(SA_{es}(T))$ recorded by the 71 stations and referred to the calibration events of Sgobba et al. (2019) (Fig. 4a). In the following, the corrective component $\mathcal{X}_s(T)$ and the term $\epsilon_s(T)$ are assumed to be independent as in Sgobba et al. (2019), and thus analysed separately. Moreover, the study domain is restricted to the same geologically homogeneous area $D \subset \mathcal{D}$ of the Po plain considered by Sgobba et al. (2019). The present application appears suitable for comparison purposes with these previous works, being grounded on compatible assumptions and study areas.

For our aims, we shall consider a functional data analysis perspective, allowing to provide random realizations both for the corrective term \mathcal{X}_s and for the residual term ϵ_s . The combination of these realizations with the predictions of the NI15 model will allow to provide multiple realizations of shaking fields jointly and consistently for all the periods $T \in \mathcal{T}$. As discussed in Sect. 1, this purpose would be hardly achievable with multivariate geostatistical approaches based on cross-variography, due to (i) the high dimensionality of the spectral accelerations considered in this study (25 periods, which would require estimating a set of 325 variograms and cross-variograms), and (ii) the infinite dimensionality of the targeted curve IM_{es} . The following Sect. 3 describes the model and methods used for the term \mathcal{X}_s (Sect. 3.1) and for ϵ_s (Sect. 3.2).

3 A functional simulation approach for the prediction of IMs

3.1 Conditional simulation of the corrective term

The focus of this subsection is on the corrective terms \mathcal{X}_s introduced in equations (1) and (2). At a site $s \in D$, the corrective term \mathcal{X}_s is considered as a functional random element $\{\mathcal{X}_s(T), T \in \mathcal{T}\}$, valued in the space $L^2(\mathcal{T})$ (or L^2 for short) of square-integrable functions, defined on the

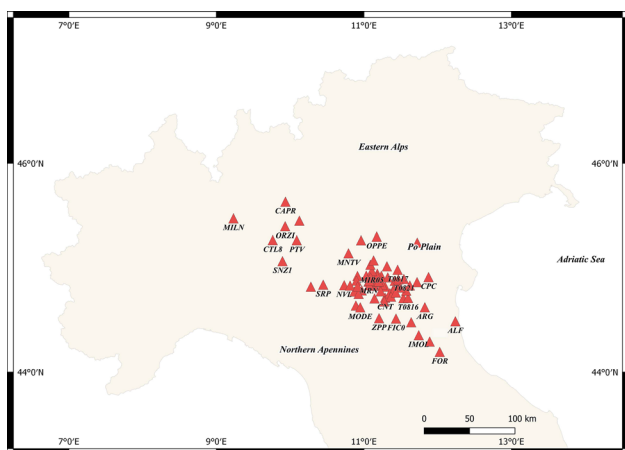


Fig. 2 Geographic location of the study-area in Northern Italy (Po Plain). Red triangles indicate the strong motion recording stations belonging to the networks of the Italian Department of the Civil Protection (DPC) and the Istituto Nazionale di Geofisica e Vulcanologia (INGV)

closed interval \mathcal{T} . The collection of functional corrective terms $\{\mathcal{X}_s, s \in D\}$ is modeled as a functional random field (Delicado et al. 2010; Menafoglio and Secchi 2017) in L^2 . The field $\{\mathcal{X}_s, s \in D\}$ is assumed to be a stationary Gaussian random field, with constant spatial mean m

$$m(T) = \mathbb{E}[\mathcal{X}_s(T)], \quad T \in \mathcal{T}, s \in D,$$

and stationary cross-covariance operator C

$$C(s_1 - s_2)x = \mathbb{E}[\langle \mathcal{X}_{s_1} - m, x \rangle (\mathcal{X}_{s_2} - m)], \quad x \in L^2, s_1, s_2 \in D, \tag{3}$$

$\langle \cdot, \cdot \rangle$ denoting the inner product in L^2 . The cross-covariance operator $C(s_1 - s_2)$ in (3) allows to determine the *covariability* between elements of the field at two locations s_1, s_2 ; it thus represents the generalization of the multivariate covariance function (associating a covariance matrix to any pair of locations) in use in multivariate geostatistics. Moreover, stationarity and Gaussianity of the functional field $\{\mathcal{X}_s, s \in D\}$ entails the stationarity and Gaussianity of the (univariate) marginal fields $\{\mathcal{X}_s(T), s \in D\}$, for any $T \in \mathcal{T}$, as assumed in Sgobba et al. (2019) (see also Sects. 2, 6). Note that, in general, the field $\{\mathcal{X}_s, s \in D\}$ may be nonstationary, particularly if the term $\delta P2P_{sr}$ in (2) plays a relevant role in the determination of the variability of the corrective term. We here detail the methodology under stationarity, as this framework is consistent with the data at hand, as shown in Sect. 4 and previously in Sgobba et al. (2019). Nonetheless, extensions of these methods to nonstationary contexts have been discussed in O2S2, e.g., in Menafoglio et al. (2013), Menafoglio et al. (2016a). These could be applied in this context too, without substantial modifications of the strategy here proposed.

The purpose of the study is to provide random realizations of the field $\{\mathcal{X}_s, s \in D\}$, conditional to the observations available at sampled stations s_1, \dots, s_n in the system. Such random realizations represent a set of scenarios which are consistent with the corrective terms available at the sampling locations. Note that this modeling framework is motivated by the observation that, in the presence of an independent event, the corrective terms at the sampling locations would not be changed, as they represent systematic effect ‘known’ at the sampling sites. The generation of conditional scenarios thus aims to represent the uncertainty on the corrective terms at unobserved locations in the system.

To generate realizations from a functional random field, this work follows the methodological proposal of Menafoglio et al. (2016b), and pursues an approach based on (i) a dimensionality reduction of the functional data by projection over an orthonormal functional basis, (ii) generation of the coefficients of the functional representation, through multivariate (Gaussian) geostatistical simulation, and (iii)

functional kriging of the residuals of the projection, to interpolate the data at the sampling locations. In this setting, each element \mathcal{X}_s of the random field $\{\mathcal{X}_s, s \in D\}$ is represented as

$$\mathcal{X}_s(T) = m(T) + \sum_{k=1}^K x_k(s)\varphi_k(T) + \xi(T), \quad T \in \mathcal{T}, \tag{4}$$

where $\{\varphi_k, 0 \leq k \leq K\}$ is a truncated orthonormal basis of L^2 , $x_k(s)$ is the coefficient of \mathcal{X}_s over the k th element of the basis φ_k , i.e., $x_k(s) = \langle \mathcal{X}_s - m, \varphi_k \rangle$, and ξ is a residual term from the projection, assumed to be uncorrelated of the $x_k(s)$. In principle, any (orthonormal) basis $\{\varphi_k, 0 \leq k \leq K\}$ could be used (e.g., Fourier, polynomial, wavelet bases). However, Menafoglio et al. (2016b) suggest the use of the orthonormal basis generated by the first K functional principal components (FPCs, see, e.g., Ramsay and Silverman 2005), because this provides a best approximation of the field in the mean square sense, uniformly in D . The use of projections over the FPCs was also advocated by Nerini et al. (2010), Menafoglio et al. (2016a) in the context of functional spatial prediction (kriging). Recall that the FPCs are the directions associated with the maximum variability of the dataset, i.e., the k th FPC e_k is found by maximizing, over $\varphi \in L^2$,

$$\sum_{i=1}^n \langle \mathcal{X}_{s_i} - \bar{\mathcal{X}}, \varphi \rangle^2 \quad \text{subject to} \quad \|\varphi\|^2 = 1, \langle \varphi, e_j \rangle = 0, j < k,$$

where the last orthogonality constraint is meaningful only for $k > 1$. The FPCs are found as the ordered eigenfunctions of the sample zero-lag covariance operator $S(\theta)$,

$$S(\theta)x = \frac{1}{n-1} \sum_{i=1}^n \langle \mathcal{X}_{s_i} - \bar{\mathcal{X}}, x \rangle (\mathcal{X}_{s_i} - \bar{\mathcal{X}}), \quad x \in L^2, s \in D, \tag{5}$$

i.e., as solution of the eigenproblem

$$S(\theta)e_k = \lambda_k e_k, \quad k = 1, 2, \dots$$

where $\lambda_1 > \lambda_2 > \dots \geq 0$. Note that, in some cases (e.g., if the GMM was estimated via least-squares and all the estimated corrected terms were considered), the mean m and the sample mean $\bar{\mathcal{X}}$ of the field \mathcal{X}_s may be null. However, we prefer to keep reference to the terms $m, \bar{\mathcal{X}}$ to account for (i) the possible bias induced by alternative estimation methods for the GMM, and (ii) the effect of possible smoothing procedures in the data pre-processing, and (iii) the possible selection of a subset of the estimated corrected terms to perform the scenario-based analysis—as in the present study, where the study domain D is a subset

of the area \mathcal{D} upon which the calibration of the GMM is based.

Estimated the FPCs e_1, \dots, e_K , each observation is represented by (a) a vector of coefficients (a.k.a. scores) $\mathbf{x}(s_i) = (x_1(s_i), \dots, x_K(s_i))^T$, obtained by projecting the centered observations over the basis, i.e., $x_k(s_i) = \langle \mathcal{X}_{s_i} - \bar{\mathcal{X}}, e_k \rangle$, $k = 1, \dots, K$, and (b) a residual term $\xi_s = \mathcal{X}_s - \bar{\mathcal{X}} - \sum_{k=1}^K x_k(s)e_k$. Conditional realizations of the functional field $\{\mathcal{X}_s, s \in D\}$ can be then obtained, based on (4), by generating random realizations of the (multivariate) field of coefficients' vector $\{\mathbf{x}(s), s \in D\}$ conditional on the observations $\mathbf{x}(s_1), \dots, \mathbf{x}(s_n)$ at the sampled sites, and by interpolating the residual term ξ_s via functional kriging (Delicado et al. 2010; Menafoglio and Secchi 2017). Note that the assumptions of stationarity and of Gaussianity for the field $\{\mathcal{X}_s, s \in D\}$ entail the same assumptions for the field $\{\mathbf{x}(s), s \in D\}$ in \mathbb{R}^K . Furthermore, the cross-covariance operator (3) is coherently estimated via the variograms and cross-variograms of $\{x_k(s), s \in D\}$, $k = 1, \dots, K$, as discussed, e.g., in Menafoglio and Petris (2016) and Nerini et al. (2010). At a target location s_0 , the conditional simulation of the coefficient vector $\mathbf{x}(s_0)$ may follow any multivariate method for geostatistical simulation (e.g., Chilès and Delfiner 1999; Mariethoz and Caers 2015). In this work, sequential Gaussian simulation is used, following the approach of Abrahamsen and Benth (2001), as implemented in the package `gstat` (Pebesma 2004) of the software R (R Core Team 2013). Functional kriging of the residual term ξ_s can be performed by modeling its stationary trace-semivariogram

$$2\gamma(s_i - s_j) = \mathbb{E}[\|\xi_{s_i} - \xi_{s_j}\|^2], \quad s_i, s_j \in D, \tag{6}$$

estimated from the data, and solving a linear system analogous to that used in the scalar case. Note that the trace-semivariogram (6) is a real-valued function (thus independent of the period T), which provides a global measure of spatial dependence for the functional field $\{\xi_s, s \in D\}$. The modeling effort required for its estimation is thus virtually the same as in the univariate case (for further details, see, e.g., Menafoglio and Secchi 2017).

3.2 Unconditional simulation of the error term

The term left to be considered in model (1) is the remaining term ϵ_s , the term δB_e being fixed and known in our case study. The term ϵ_s describes the aleatory variability of the (log) IMs, which cannot be ascribed to any systematic effect. We here model the error field $\{\epsilon_s, s \in D\}$ as a zero-mean Gaussian functional random field valued in L^2 , uncorrelated in space and independent on the field $\{\mathcal{X}_s, s \in D\}$. Again, these assumptions entail the similar assumptions made in Sgobba et al. (2019) (see also the

discussion in Sect. 6). Such modeling choice is consistent with the assumption that the entire spatial dependence of the residual from the GMM, $R_s = \mathcal{X}_s + \delta B_e + \epsilon_s$, can be associated with the systematic effects already captured in the term \mathcal{X}_s . In this work, we shall also assume that the field $\{\epsilon_s, s \in D\}$ is stationary.

Simulation of the functional error term ϵ_s will follow an unconditional approach, i.e., we will generate realizations of the field $\{\epsilon_s, s \in D\}$ without conditioning to the observations $\epsilon_{s_1}, \dots, \epsilon_{s_n}$. Indeed, for an independent event, the error ϵ_s at the sampling locations s_1, \dots, s_n would not be the same as that already observed, as ϵ_s does not capture any systematic effect. Unconditional simulation of ϵ_s follows a similar scheme as that described in Sect. 3.1. It consists of (i) dimensionality reduction (via FPCA) yielding a decomposition $\epsilon_s(T) = \sum_{l=1}^L v_l(s)\psi_l(T)$, $v_l(s)$ denoting the l th score and $\psi_l(T)$ the l th FPC, $l = 1, \dots, L$, and (ii) unconditional simulation of the scores. Note that the scores along the FPCs can be now simulated as random realizations of a white noise, as the field $\{\epsilon_s, s \in D\}$ is assumed Gaussian and uncorrelated. Given the purely random nature of ϵ_s , the residual terms $\sum_{l=1}^{N-1} v_l(s)\psi_l(T)$ (N denoting the total number of observations of ϵ_s available at the sampling locations) will not be considered in the simulation of ϵ_s .

3.3 Simulation of the spectral acceleration

The spectral acceleration $SA(T)$ is eventually reconstructed, for $T \in \mathcal{T}$, as

$$\begin{aligned} \log(SA_{es_0}^*)(T) &= \mu_{es_0}(T) + \bar{\mathcal{X}}(T) + \sum_{k=1}^K x_k^*(s_0)e_k(T) \\ &\quad + \xi_{s_0}^*(T) + \delta B_e(T) + \sum_{l=1}^L v_l^*(s_0)\psi_l(T), \end{aligned} \tag{7}$$

where $\mu_{es_0}(T)$ denotes the mean acceleration predicted by the GMM, δB_e denotes a fixed and known inter-event term (described in Sect. 2), and the other terms are associated with the random realization of the corrective term at the target location s_0 , with the kriged residual term and with the random realization of the remaining aleatory term ϵ_s , respectively. In the following, we shall indicate by $\mathcal{X}_s^K, \epsilon_s^L$ the projected component, i.e.,

$$\mathcal{X}_s^K = \bar{\mathcal{X}}(T) + \sum_{k=1}^K x_k(s)e_k; \quad \epsilon_s^L = \sum_{l=1}^L v_l(s)\psi_l(T),$$

respectively, and by $\mathcal{X}_s^{*K}, \epsilon_s^{*L}$ the corresponding simulation, i.e.,

$$\mathcal{X}_s^{*K} = \bar{X}(T) + \sum_{k=1}^K x_k^*(s)e_k; \quad \epsilon_s^{*L} = \sum_{l=1}^L v_l^*(s)\psi_l(T).$$

Note that the described simulation procedure reproduces the projected components $\mathcal{X}_s^K, \epsilon_s^L$, whereas it neglects part of the variability associated with the residual term ζ_s , and the entire variability of $\sum_{l=L+1}^{N-1} v_l(s)\psi_l(T)$. The latter sources of uncertainty are however marginal with respect to the variability of the terms $\mathcal{X}_s^K, \epsilon_s^L$, as these are precisely found as to capture the respective main modes of variability. The choice of the truncation orders K, L reflect directly on the amount of variability captured by the representation—hence by the simulation. A trade-off in the choice of K, L is however apparent, as they need to balance the ability to reproduce the system variability and the modeling/computational effort needed for the simulations, particularly for the term \mathcal{X}_s^{*K} , whose scores require spatial modeling (i.e., variography analysis). Further discussion on the point is provided in Sect. 4.

4 Generation of shaking fields in the study area

Data preprocessing To apply the methodology detailed in Sect. 3, raw data (Figs. 3a–4a) were smoothed by using cubic splines. Interpolating cubic splines with 25 equally spaced knots were used to represent the corrective terms, yielding the functional forms reported in Fig. 3b. Analyses were also performed upon considering data smoothed via smoothing B-splines, without substantial differences in the results described hereafter (not shown). Error terms were smoothed by using smoothing cubic splines with 25 equally spaced knots. Smoothing coefficients were determined by minimizing a least squares criterion with a penalization on the curvature of the resulting solution. The weight of the

penalization was selected via generalized cross-validation (Ramsay and Silverman 2005). Figure 4b reports the smoothed data. This set of choices allowed to cope with the fact that some error terms were only partially available along the whole domain T , as the set of accelerometric waveforms originally used to compute the corrective terms were processed assuming variable high- and low-pass filter periods (Lanzano et al. 2018), thus leading to different usable period ranges for each signal.

4.1 Spatial analysis of the corrective term

Dimensionality reduction of \mathcal{X}_s Functional principal component analysis was performed on the smoothed data displayed in Fig. 3b; results are reported in Fig. 5. The first $K = 4$ FPCs together explain about the 98% of the variability of the dataset. The reconstruction of the data when using the first $K = 4$ principal components (i.e., $\mathcal{X}_{s_i}^K, K = 4, i = 1, \dots, 71$) is displayed in Fig. 5h. A comparison with the smoothed data suggests that the FPC representation tends to smooth the small fluctuations presented by the data at long periods. Fluctuations in the data at short periods are better reproduced by the FPC approximation, as these are precisely associated with the second FPC (see Fig. 5d). The variability of the sample for high periods could be better captured if a higher number of components was considered ($K > 10$). Such a high truncation order would lead to a simulation procedure more demanding than that described here, and devoting a significant effort to explore only a small portion of the data variability. Note that, even as one was resorting to $K = 10$ FPCs, a relevant dimensionality reduction would be attained with respect to a standard multivariate setting, yielding a substantially smaller cross-variogram structure (55 models for $K = 10$ against the 325 models needed in the multivariate framework), with the further advantage of obtaining predictions for the full profile instead of a discrete set of periods.

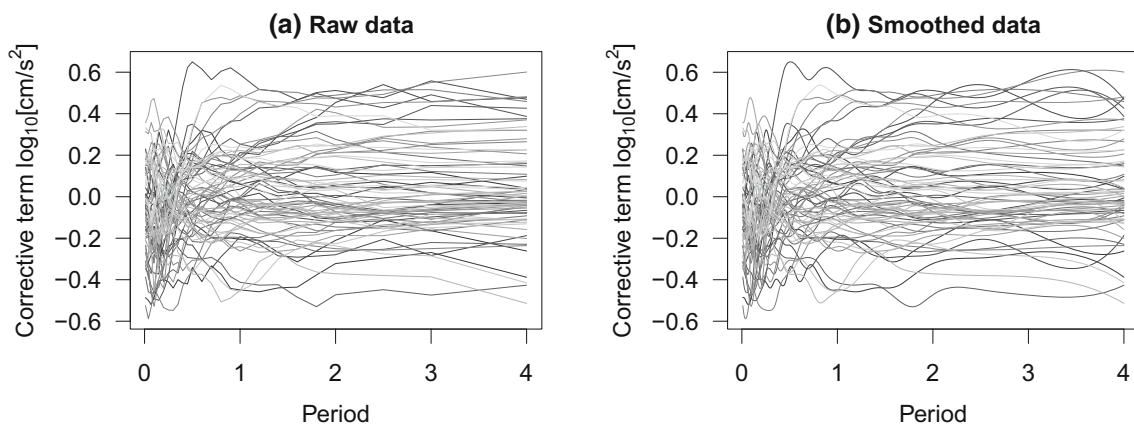


Fig. 3 Corrective terms at the 71 recording stations. a Raw data. b Smoothed data

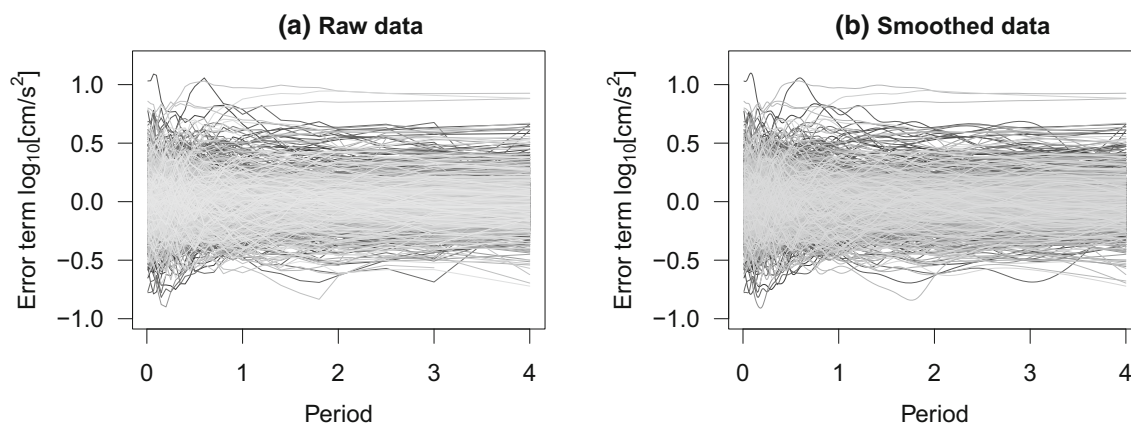


Fig. 4 Error ϵ_s at the 71 recording stations. **a** Raw data. **b** Smoothed data

Further comparison with a multivariate approach is discussed in Sect. 6. In the following, we consider a $K = 4$ -dimensional approximation of the data, and ascribe the residual variability to the term ξ_s in (4).

Geostatistical analysis of the scores x_s The variograms and cross-variograms of the coefficient vectors $\mathbf{x}(s_1), \dots, \mathbf{x}(s_n)$ were modeled by fitting a linear model of coregionalization to the empirical estimates. Figure 6 depicts the estimated linear model of coregionalization, based on an exponential model. Empirical variograms, represented as points in Fig. 6, support the stationarity assumption. Furthermore, one can notice that the random field of coefficient vectors displays a non-negligible spatial correlation, indicating that a systematic component is indeed present in the residuals of the GMM.

The estimated model was used to generate the realizations of the coefficient vectors, over a grid of locations in the spatial domain D having resolution 750 [m] in both directions X and Y . The conditional realizations were obtained by sequential Gaussian simulation (Chilès and Delfiner 1999), setting to 0 the mean of the fields $\{\mathbf{x}_k(s), s \in D\}$, $k = 1, \dots, K$. The realizations of the coefficient vectors were used to determine the associated realizations $\mathcal{X}_{s_0}^{*K}$ of the projected component $\mathcal{X}_{s_0}^K$, as in (7). An example of realization is displayed in Fig. 7a.

Geostatistical analysis of the residual term ξ_s The trace-semivariogram of the residual term ξ_s was modeled by fitting an exponential model with nugget to the empirical estimate (estimated parameters: nugget 0.0014; partial sill 0.0020; practical range 59.34 km). The empirical estimate supports the stationarity assumption (not shown). The estimated trace-semivariogram model was used to perform functional kriging of the residual term. A sample of kriged residuals ξ_{s_0} is reported in Fig. 7b. Visual comparison between Fig. 7a, b confirms that the residual terms indeed represent a marginal variability of the corrective term.

Simulation of the corrective term The realizations of the corrective term $\mathcal{X}_{s_0}^*$ were eventually obtained by adding the kriged residuals to the realization of the projected component. Figure 8 displays a spatial representation of the sample realization $\mathcal{X}_{s_0}^*$ of the functional field, evaluated at the periods $T = 0$ [s] (PGA) and $T = 4$ [s].

The obtained fields are also consistent with the results provided by Sgobba et al. (2019) for the univariate approach (see also Sect. 6 and the Supplementary Material). Hence, the functional fields of the corrective terms, while reducing the high dimensionality of the problem, also allow to capture the physically-related aspects of the shaking pattern. Indeed, the simulated fields reproduce the main spectral features compliant with the geomorphology of the area, such as the systematic enhancement effects observed at long-periods ($T = 4$ [s]) in ground motion amplitudes along the North–South direction (Fig. 8).

It is worth noting that several authors (see Schiappapietra and Douglas 2020 for a review) demonstrated that the systematic terms constituting $\mathcal{X}_s(T)$ (see also Eq. (2)) are characterised by different scales of variability that are also period-dependent. Indeed, the site terms are generally found to be stationary in space, whereas the path terms often reveal a non-stationarity, being dependent on the absolute position of source and site, particularly at long-periods; finally, the source term is typically spatially uncorrelated. However, the peculiar features of the considered dataset led Sgobba et al. (2019) to model the spatial correlation of the corrective terms as a univariate and stationary Gaussian field. The functional analysis here performed on the same dataset also do not find evidences of non-stationarities. This provides an indication that the dominant features in the dataset variability may be represented by the site terms and the short periods (the latter aspect being also evidenced by FPCA).

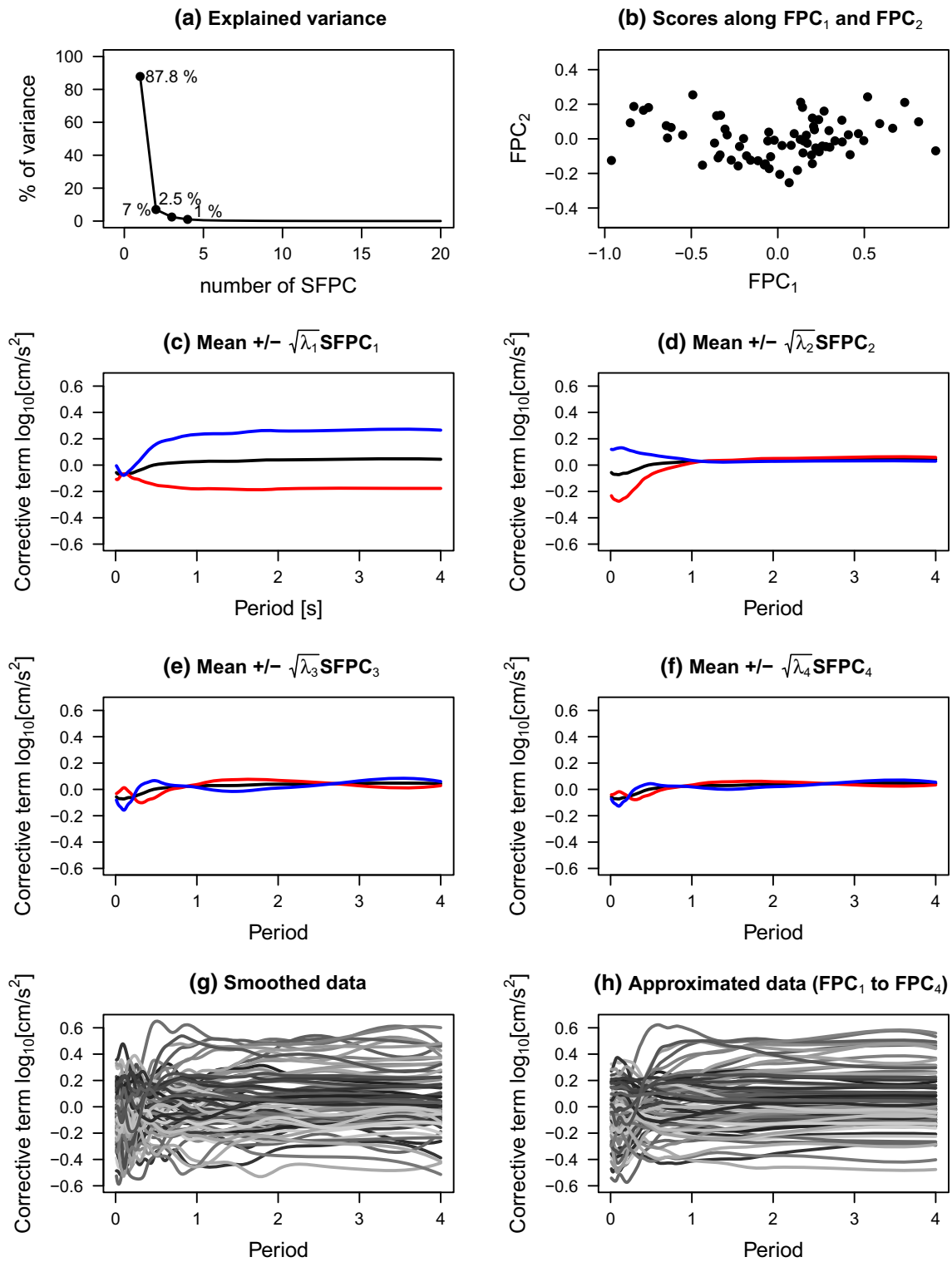


Fig. 5 Results of the FPCA on the dataset of corrective terms $\mathcal{X}_{s_1}, \dots, \mathcal{X}_{s_n}$ ($n = 71$)

4.2 Analysis of the aleatory term ϵ_s

The viability of the assumption of spatial uncorrelation was assessed by inspection of the estimated trace-variogram of

the field (Fig. 9a), which is well represented by a pure-nugget model. To assess the spatial homoscedasticity of the error terms, for each site s_i , $i = 1, \dots, n$, an estimate of the

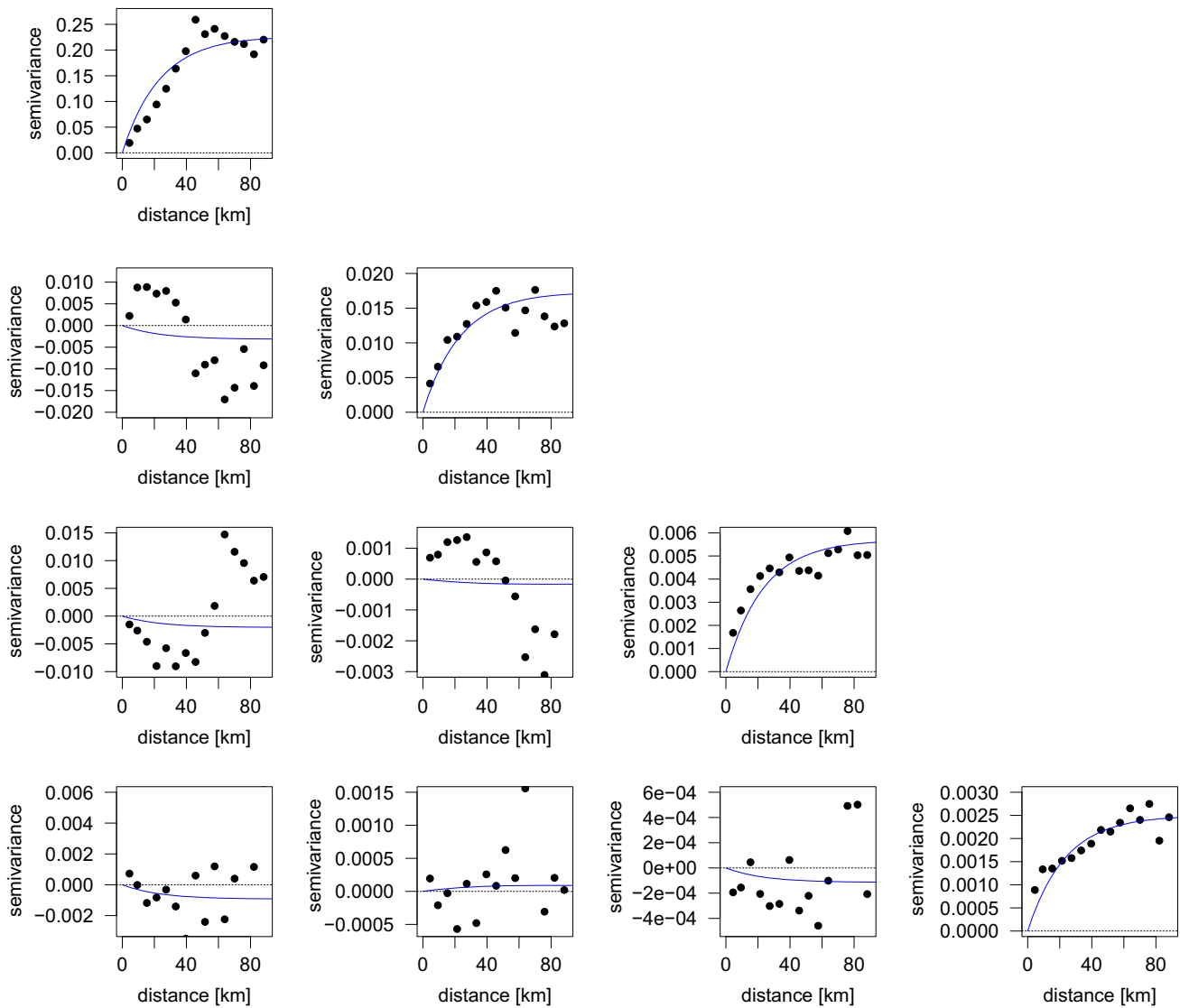


Fig. 6 Semivariograms and cross-semivariograms estimated from the coefficient vectors $\mathbf{x}(s_1), \dots, \mathbf{x}(s_n)$. Empirical semivariograms are indicated with symbols; the estimated linear model of coregionalization is indicated with solid curves

global variance of ϵ_{s_i} was computed from the available repetitions $\epsilon_{s_i,1}, \dots, \epsilon_{s_i,n_i}$, i.e.,

$$\hat{\sigma}_{s_i}^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} \|\epsilon_{s_i,j}\|^2.$$

As shown in Fig. 9b, the estimated variances $\hat{\sigma}_{s_1}^2, \dots, \hat{\sigma}_{s_n}^2$ ($n = 71$) do not display any relevant spatial pattern.

Stochastic simulation of the term ϵ_s thus followed two steps, namely (i) dimensionality reduction via FPCA, obtaining a set of scores and a set of eigenfunctions, and (ii) simulation of the scores from a Gaussian random field, uncorrelated in space. Dimensionality reduction was performed by considering $L = 10$ FPCs, that together explain almost 100% of the variability. A summary of the results of FPCA on ϵ_s is presented in the Supplementary Material.

Note that, when considering the term ϵ_s , the dimensionality of the FPCA representation has a relatively low impact on the modeling and computational effort required for the simulation, as the spatial autocorrelation (i.e., variogram) does not need to be estimated for the scores. Eventually, for $l = 1, \dots, L$ and a target location s_0 in D , the scores $v_l(s_0)$ corresponding to the l th FPC were simulated from a zero-mean Gaussian random noise, uncorrelated in space, with variance equal to the variance of ϵ_s along the l th FPC.

4.3 Shaking fields

The shaking fields were built by combining the prediction of the GMM, the realizations of the corrective term and the realization of the error term, as in (7). Figure 10 displays the simulation of the shaking field for the first mainshock

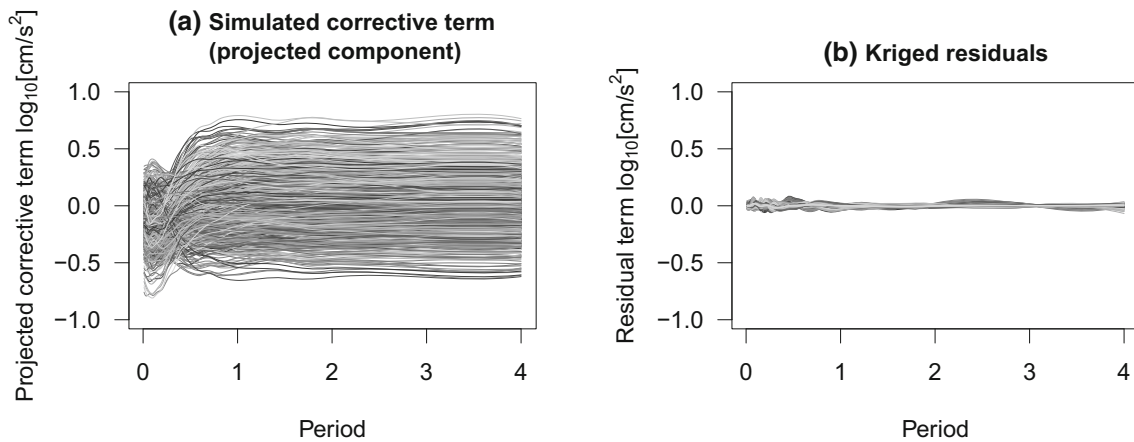


Fig. 7 Simulation of the corrective term \mathcal{X}_s . **a** Sample realization of the projected part $\mathcal{X}_{s_0}^K$. **b** Kriged residual term ξ_s . Grey colors are given consistently in panels **a** and **b**. The realization \mathcal{X}_s^* of the

corrective term \mathcal{X}_s is obtained by adding the kriged residual term to the simulated projected part

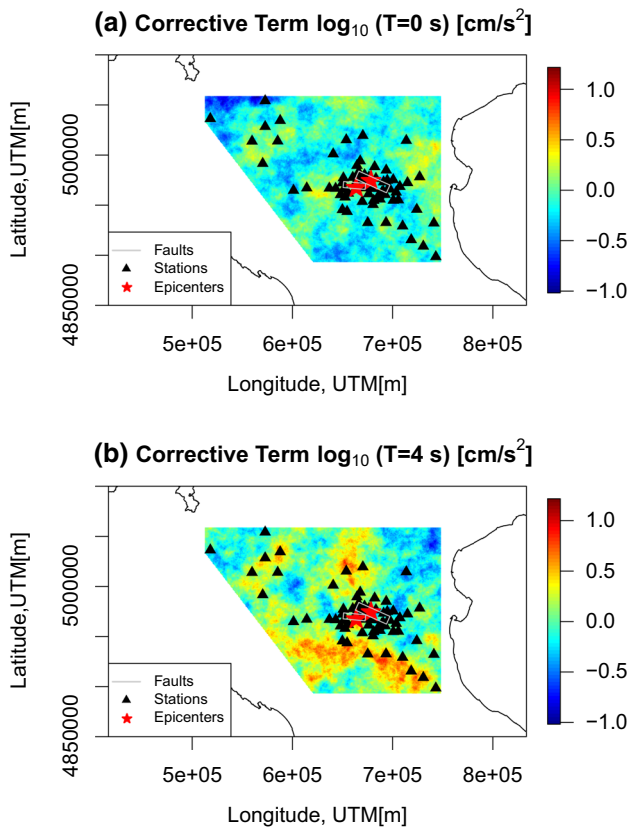


Fig. 8 Realization of the corrective term at periods **a** $T = 0$ [s] and **b** $T = 4$ [s]. Stations are denoted by black triangles. Red stars indicate the epicenters of the two mainshock of the 2012 Emilia seismic sequence

(Mw 6.1, 20/05/2012) of the Emilia sequence (data from the Italian ACcelerometric Archive database—ITACA, <http://itaca.mi.ingv.it/>; Pacor et al. 2011). This was computed as the sum of the GMM and the simulated terms. For comparison, the left panels of Fig. 10 report the prediction

of the GMM. Simulations show that a non-negligible variability can be ascribed to the corrective and error terms. Near the epicenter of the event, the simulated shaking fields present higher values of spectral acceleration than those predicted by the GMM only, which cannot capture realistic near-source pattern as it would assign constant values to the entire area within the surface projection of the fault. At long-periods ($T = 4$ [s]), the simulated ground motion pattern reflects the enhancement observed on the fields of the corrective terms, which is reported as the consequence of focalization effects of the seismic waves at the regions bordering the Po basin (Paolucci et al. 2015; Sgobba et al. 2019).

5 Model validation

Validation of the model was performed by cross-validation. Cross-validation analysis focused on the model of the corrective terms, whose spatial component is non-negligible. The error term was not included in the cross-validation analysis, as it is not associated with systematic spatial patterns. The effect of the error term ϵ_s is further discussed in the Sect. 7, which illustrates the application of the calibrated model to an independent event. We remark that validation of the GMM was already presented by Lanzano et al. (2016), and it is thus outside the scope of the present work.

A nine-fold cross-validation analysis was performed on the corrective terms, to assess the capability of the model to correctly capture the systematic component observed at the sites, and not described by the GMM. The observations $\mathcal{X}_{s_1}, \dots, \mathcal{X}_{s_n}$, $n = 71$, where randomly divided in $\kappa = 9$ batches $\{\mathcal{X}^{(k)}\}$, $k = 1, \dots, 9$, of which the first eight were

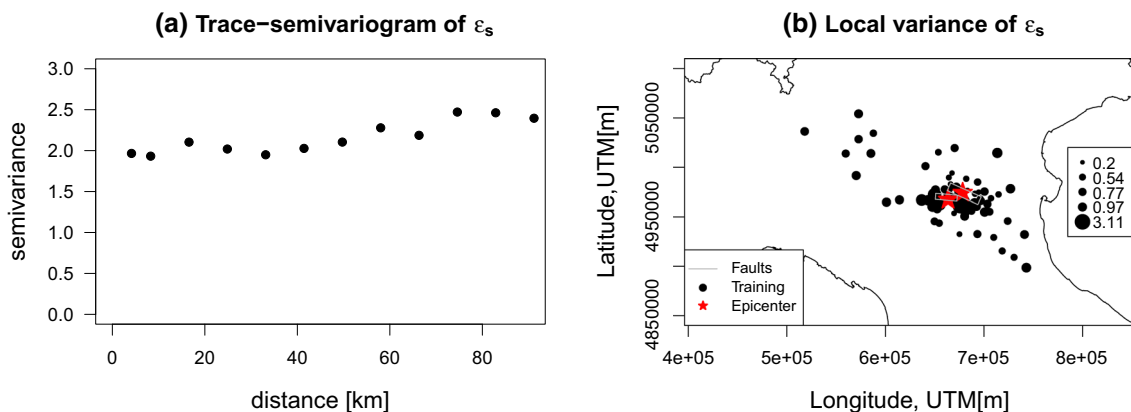


Fig. 9 Analysis of the error term ϵ_s . a Empirical trace-semivariogram. b Average global variance $\hat{\sigma}_{s_i}^2$ at each training site $s_i, i = 1, \dots, n (n=71)$

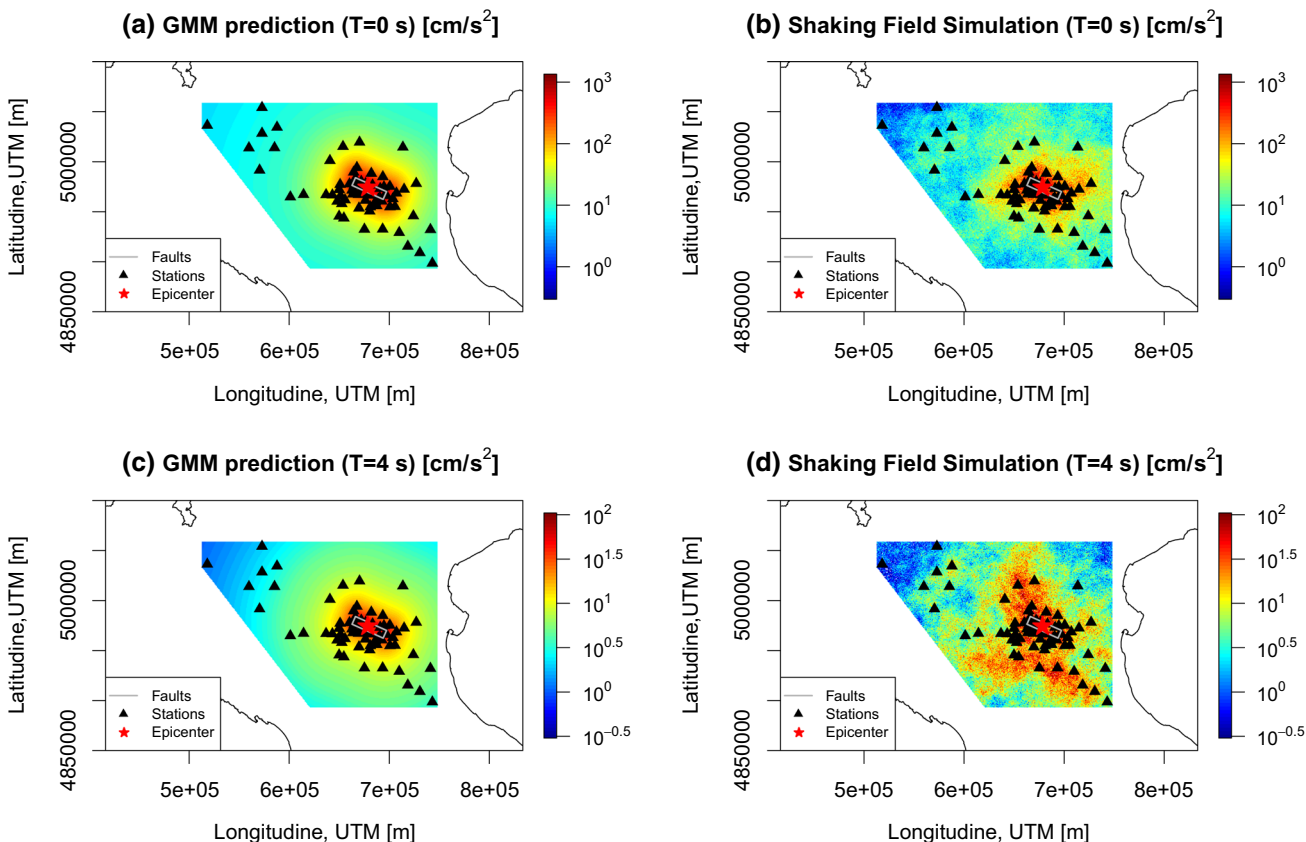


Fig. 10 Comparison between the prediction from the GMM and a realization of the shaking field for the Mw6.1 first-Emilia mainshock. Maps in a and c refer to the GMM prediction for $T = 0$ [s] (PGA) and $T = 4$ [s], respectively. Maps in b and d refer to the simulated

shaking fields for $T = 0$ [s] (PGA) and $T = 4$ [s], respectively. The star indicates the epicenter of the Mw6.1 first-Emilia mainshock. The corresponding fault plane is shown as a closed rectangle. Colors are given on a log-scale

made of $n_k = 8$ observations, and the last of $n_k = 7$ observations.

For a batch $\{\mathcal{X}^{(k)}\}, k = 1, \dots, 9$, the dataset was split in training and test set, the test set consisting of the batch $\{\mathcal{X}^{(k)}\}$, the training set of the remaining $n - n_k$ data. The training set was used to fit the model, by following the methodology described in Sect. 3. A set of $B = 1000$

conditional realizations of the field was generated at each test site, and compared with the corresponding data point in the test set. Figure 11 displays with colors the sites associated with the $\kappa = 9$ batches (Fig. 11a), and the results obtained for two locations in the first test set $\{\mathcal{X}^{(1)}\}$ (MIR05 and MNS, Fig. 11b–c). The latter locations were selected as representative of best/worst scenarios in the

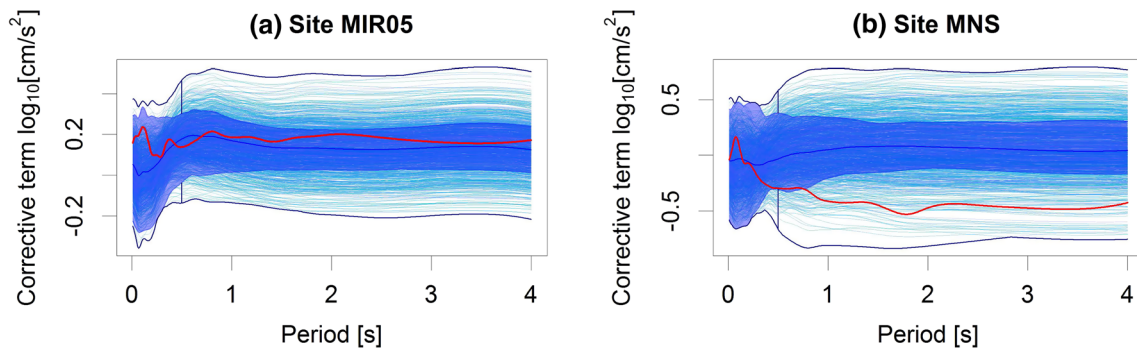


Fig. 11 Cross-validation results. **a** Functional boxplot for the site *MIR05* ($s = (666134, 4982978)$). **b** Functional boxplot for the site *MNS* ($s = (713603, 5014520)$). Site coordinates are in Lat/Lon UTM

model performances. To enhance the interpretation of the plots, functional boxplots based on modified band depths (MBD, Sun and Genton 2009; Ieva and Paganoni 2013) were used to represent the simulated curves. Here, simulated curves are represented as colored light blue lines; the shaded bands represent the envelop of the 50% most central curves in the simulated sample, according to the ordering induced by the MBD measure. The outermost thick blue lines represent the functional analogue of the whiskers in a classical boxplot. The target corrective terms (i.e., the corrective term at each test location) are represented as thick red lines. At the station *MIR05*, Fig. 11b the target corrective term is very central with respect to the simulated sample, showing that the model correctly captures the systematic residual of the GMM at that location. On the contrary, the dynamic observed at site *MNS*, Figure 11b displays a more extreme behavior with respect to the simulated sample, particularly for periods $T > 1$ [s]. Note that errors in cross-validation may also be indicative of informative data in the training set, particularly if the test sites are located at the boundary of the domain (as in the case of the site *MNS*, see Sect. 7).

A summary of the results obtained for all the $n = 71$ stations is reported in Fig. 12. Figure 12b displays, through colors, a table \mathbb{T} whose element $\mathbb{T}_{i,j}$, $i = 1, \dots, \kappa$, $j = 1, \dots, n_i$ ($\kappa = 9$), represents the proportion of simulations *more extreme* than the j th target curve of the i th test set, $\mathcal{X}_j^{(i)}$. More precisely, $\mathbb{T}_{i,j}$ was built by evaluating the depth of $\mathcal{X}_j^{(i)}$ (in terms of MBD) with respect to the sample of simulations, and by computing the proportion of simulations less deep (i.e., more extreme) than the target corrective term. A low (resp. high) value of $\mathbb{T}_{i,j}$ represents a target curve which is very central (resp. very extreme) with respect to the simulated curves. The same color scheme is used for the representation of sites in the map of Fig. 12a. The three stations where the corrective terms are more extreme than 95% of the simulated samples

are *MODE*, *FICO* and *T0828*. These stations are those where the method shows the worst cross-validation performances. We note that, in the functional context, the definition of quantiles is highly non-trivial (see, e.g., Chakraborty and Chaudhuri 2014) and building prediction bands of given coverage is still a field of open research. Although functional boxplots only provide empirical coverage, cross-validation results show that only 4% of the stations are not covered by the deepest 95% of the simulations, confirming a good overall performance of the method.

6 Comparison with finite-dimensional approaches

To further explore the relation between our approach and competitor finite-dimensional approaches, we here compare our results with those obtained in the following settings

- *Case UVT* Univariate analysis of the corrective terms $\mathcal{X}_{s_i}(T)$, $i = 1, \dots, n$ at the single periods $T = 0.01$ s and $T = 4$ s, as in the previous work Sgobba et al. (2019);
- *Case MVT-a* Multivariate analysis of the corrective terms $(\mathcal{X}_{s_i}(T_1), \dots, \mathcal{X}_{s_i}(T_K))$ at $K = 4$ selected periods, namely $T_k \in \{0.1, 0.3, 1, 3\}$ representative of low (PGA), medium (0.3–1 s) and long periods (3 s);
- *Case MVT-b* Multivariate analysis of the discrete corrective terms $(\mathcal{X}_{s_i}(T_1), \dots, \mathcal{X}_{s_i}(T_p))$, $i = 1, \dots, n$ at the $p = 25$ periods available in the dataset, by using a dimensionality reduction based on the first $K = 4$ multivariate principal components of the data.

The first case is considered to check the consistency of the obtained results with the state-of-the-art approach to the same field case. Cases MTV-a and MVT-b were selected as competitor multivariate methods.

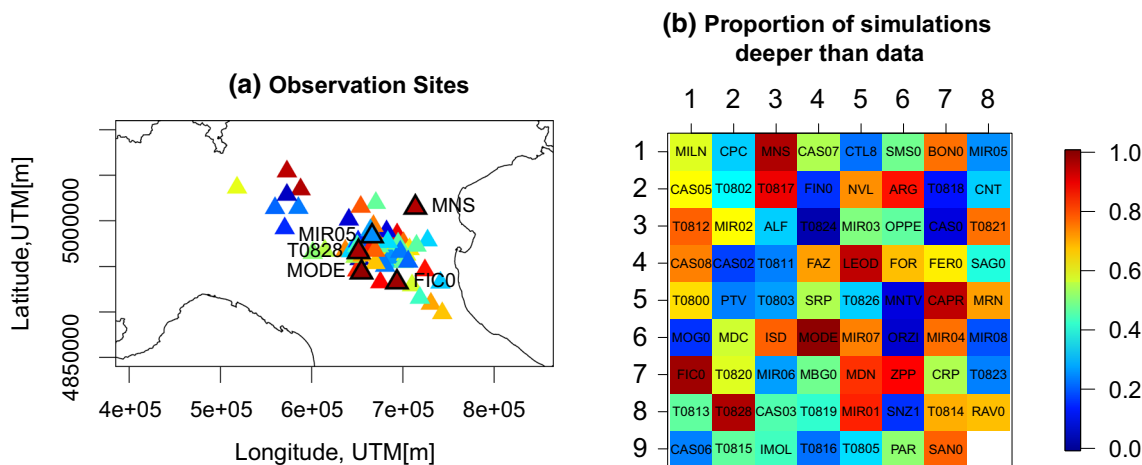


Fig. 12 Cross-validation results. **a** Sampled locations. **b** Representation of the depth (MBD) of the test observations with respect to the simulated realizations; the element i, j represents the proportion of simulated realizations whose MBD is higher than that of the j th

observation in the i th test set. Colors in panels **a** and **b** are given consistently. Codes in the cells in panel **b** correspond to the identification codes of the observation sites

From the methodological viewpoint, the first key advantage of the functional approach over the competitor approaches is that it is the only framework allowing for a joint analysis and uncertainty quantification (UQ) of the entire profile of the corrective term (corrective profile for short). Indeed, univariate models focus on a marginal UQ for fixed periods, whereas multivariate models allow for a joint representation of discretized profiles. This appears to be a crucial point, since spectral ordinates left out in the analysis may contribute to a full assessment of seismic structural response. Beside this, hereafter in this section we quantitatively compare the results of the analyses separately in the univariate and multivariate cases.

6.1 Case UVT

We here perform a comparison of our results with those of Sgobba et al. (2019). These authors analysed the same dataset of corrective terms as that here considered, but with a univariate approach applied separately to the periods T (thus neglecting cross-correlations among different periods), with particular reference to $T = 0.01$ s and $T = 4$ s. For a given period T , the analysis of the corrective terms made by Sgobba et al. (2019) was based on the following key steps: (i) geostatistical analysis of the corrective terms $\mathcal{X}_{s_1}(T), \dots, \mathcal{X}_{s_n}(T)$, and (ii) stochastic simulation of $\mathcal{X}_{s_0}(T)$ and $\epsilon_{s_0}(T)$ for the sites s_0 in a computational grid \mathcal{G} covering the target area. For step (i), the authors estimated a variogram model as in univariate geostatistics (Cressie 1993), and accordingly performed kriging at target location(s) s_0 . For step (ii), they considered for $\mathcal{X}_{s_0}(T)$ a Gaussian random noise uncorrelated in space, centered in the kriging prediction and having variance equal to the kriging variance at s_0 in \mathcal{G} ; for the term $\epsilon_{s_0}(T)$, they

considered a zero-mean Gaussian random noise uncorrelated in space, having variance equal to the estimated aleatory variance (denoted by σ in Sect. 2). The approach proposed in this work differs from that of Sgobba et al. (2019) for two main aspects: (a) a functional approach is considered, and (b) conditional stochastic simulation is used for the corrective terms instead of using an uncorrelated random error. This set of differences leads to markedly different methodological strategies. In particular, the methodology proposed in this work is precisely built to account for the cross-correlations among different periods, and to reflect the spatial covariance structure estimated for the (functional) field of corrective terms (particularly its spatial continuity). We here aim to show that, as far as the mean marginal patterns of the corrective terms are concerned, both the approaches provide consistent results. In terms of UQ, the inference attained by the two models is intrinsically different and not directly comparable. A comparison between UQ properties of functional and finite-dimensional approaches is further discussed in Sect. 6.2.

To allow for a comparison of the results of our model with the mean values of the corrective terms presented by Sgobba et al. (2019), we used the calibrated model as described in Sect. 4, but considered as spatial grid that of Sgobba et al. (2019)—who refer to a smaller target area than that considered in Sect. 4. We thus obtained (i) a set of $B = 100$ conditional realizations of the corrective profiles $\mathcal{X}_s^{*,b}$, and (ii) a set of $B = 100$ unconditional realizations of the error terms $\epsilon_s^{*,b}$. We selected the same periods analyzed in details by Sgobba et al. (2019), namely $T = 0.01$, and $T = 4$ s, and built a set of realizations of the (log-) spectral acceleration for the 20th May 2012 Mw 6.1 Emilia first mainshock at these selected periods (following Sect. 3.3

and Eq. (7)). To compare the results of our model with those of Sgobba et al. (2019), we computed the average of our marginal simulations of $\log_{10}(SA_{es}^*(T))$ at $T = 0.01$, and $T = 4$ s, and compared them with the maps presented in Sgobba et al. (2019). The latter were obtained by adding the prediction $\mu_{es_0}(T)$ of the GMM, the term $\delta B_e(T)$ and the kriging prediction $\mathcal{X}_{s_0}^*(T)$. We show these maps in Fig. 13, and report the summary statistics of their differences in Table 1 (the sign referring to the difference $Fn-UVT$). One can clearly see that the results obtained with the two methods are fully consistent. This confirms that, as far as mean values are concerned, the marginal results of the proposed approach are generally in agreement with the results of univariate approaches.

6.2 Cases MVT-a and MVT-b

Cases MTVa and MVTb are chosen as representative of alternative dimensionality reductions of the dataset leading to substantially equivalent burden for the analysis as the functional approach. Indeed, in all these cases, the

bottleneck stands in the conditional stochastic simulation of the K -dimensional vectors summarizing the profiles, which directly depends on the dimension K .

We run the analyses for both cases MVT-a and MVT-b, and compared the performances of all the methods in terms of (1) results at fixed period T , and (2) cross-validation results. For the sake of brevity, a detailed account of point (1) is provided in the Supplementary Material. We limit to mention that, as far as prediction and UQ at fixed period T is concerned, multivariate and functional approaches provide substantially consistent result for $T = 0.1, 0.3, 1$ s. The major difference between the multivariate approach MVT-a and the others is observed for long periods ($T = 3$ s), where the standard deviations of simulated fields is significantly lower for case MVT-a than for the others. This shows a limitation in approach MVT-a in appropriately characterizing the marginal and joint uncertainty of the corrective profiles, as further demonstrated by the cross-validation analyses detailed hereafter.

To quantitatively compare the method’s performances, we run a 9-fold cross-validation analogous to that

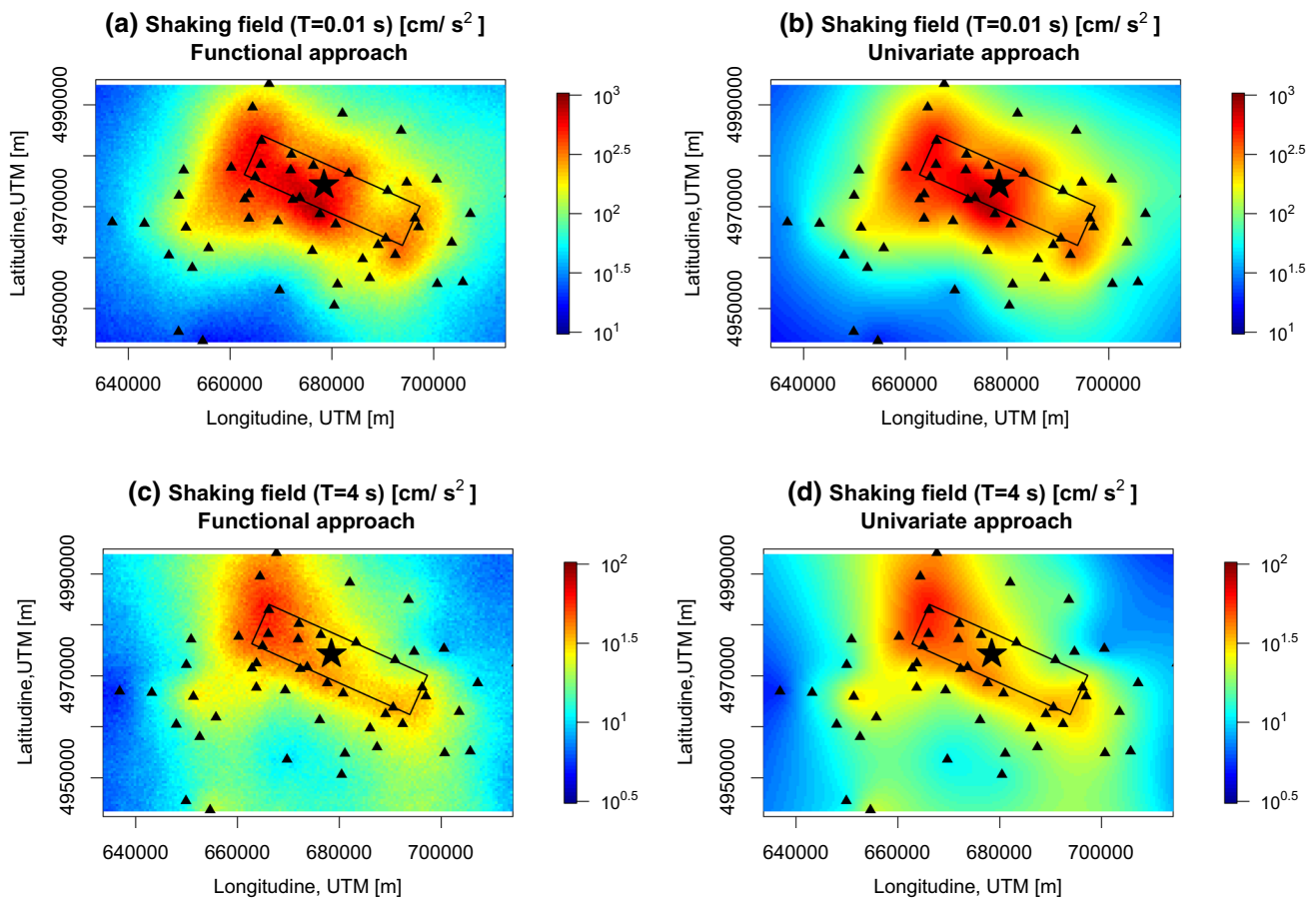


Fig. 13 Comparison between marginal results of the functional approach and the results of Sgobba et al. (2019), for $T = 0.01$ s (a–b) and $T = 4$ (c–d). The star indicates the epicenter of the Mw6.1 first-

Emilia mainshock. The corresponding fault plane is shown as a closed rectangle. The calibration sites are depicted as black triangles

Table 1 Summary statistics of the differences between the fields represented in Fig. 13a–b (first line) and b–c (second line)

	Min	Q1	Median	Mean	Q3	Max
T = 0.01 s	− 0.1441	− 0.0211	0.0017	0.0011	0.0240	0.1419
T = 4 s	− 0.0995	− 0.0222	− 0.0037	0.0024	0.0158	0.1322

considered in Sect. 5. Here, for each training set, the four-dimensional vector data of both cases MVT-a and MVT-b were modelled as multivariate random fields; the four-dimensional cross-covariance structures were estimated by using a linear model of coregionalization; the estimated models were then used to perform kriging prediction and stochastic simulation in the respective cases. To allow for a comparison with the functional setting, we considered a linear interpolation of the discrete predictions of spectral accelerations in both cases MVT-a and MVT-b. Figure 14 reports an example of simulated corrective terms at the same location MIR05, as mentioned in Fig. 11. Evaluation of Fig. 14a clearly shows that a relevant part of the information contained in the corrective profiles is lost when selecting a subset of 4 spectral ordinates (case MVT-a). In particular, the shape of the curve is hard to be reconstructed from the discrete representation. Note that, in case MVT-a, only the spectral ordinates in the range [0.1, 3] can be reconstructed by linear interpolation. Figure 14b shows that a multivariate approach based on the first $K = 4$ multivariate principal components allows one to fairly represent the shape of the profile. Indeed, it can be shown that case MVT-b represents a numerical approximation of the functional setting proposed in the manuscript. The approximation is however suboptimal, as it does not account for the uneven distribution of the 25 periods over the domain [0.01, 4]. This brings a slight discrepancy between the results of multivariate PCA on the vector data and the FPCA on the functional data. For instance, the first $K = 4$ multivariate PCs allows explaining 97.4% of the data variability, whereas the first $K = 4$ FPCs capture

98.3% of the data variability. This turns in an over-smoothing of the right part of the domain of spectral acceleration (see the Supplementary Material for a graphical comparison of FPC results). In terms of coverage properties of the simulated sample of corrective terms, we refer to Fig. 15, which reports tables analogous to that reported in Fig. 12. We again recognize a high degree of similarity between the case MVT-b and the functional one. Concerning the case MVT-a, one clearly observes a detrimental effect of this type of dimensionality reduction on the capability to make inference on the corrective profiles.

These analyses justify the use of a functional approach for prediction and UQ on both point-wise evaluations (i.e., at fixed period T in the context of a marginal inference) and on the complete profile (in the context of a joint inference). Overall, our approach seems to perform significantly better than the multivariate approach MVT-a in terms of marginal/joint inference on the corrective profiles. The case study at hand does not evidence dramatic differences between the functional approach and the multivariate approach MVT-b, although the latter just represents a (sub-optimal) discretized implementation of the proposed model.

7 Application to an independent event

The model for the spectral acceleration calibrated in Sect. 4 was used to generate shaking scenarios for an independent event registered in Northern Italy (MI 4.2, 21/08/2018).

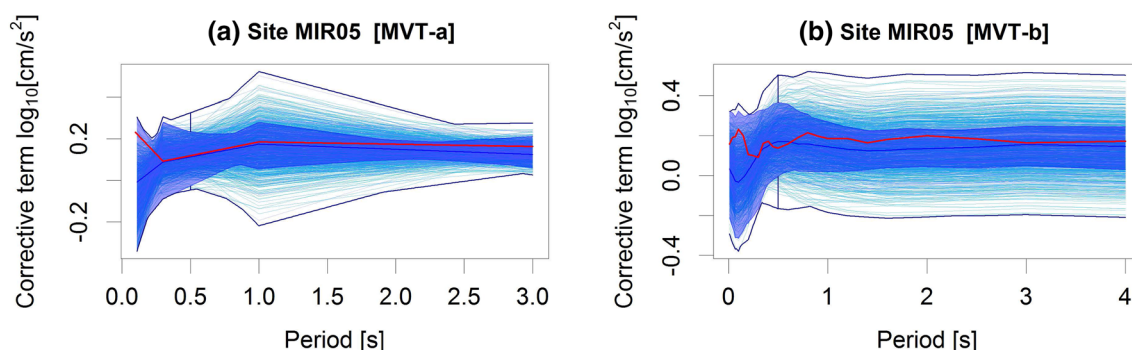


Fig. 14 Comparison of simulated corrective terms at location MIR05, for the cases MVT-a and MVT-b. Note that, in panel a, the x-axis ranges in [0.1, 3], while in panel b it ranges in [0.01, 4]

Fig. 15 Comparison of coverage properties of the simulated sample of corrective terms for the multivariate cases MVT-a (left) and MVT-b (right). Colors represent the proportion of simulations more deep than the data (red colors being associated with test data that are *extreme* with respect to the simulated sample)

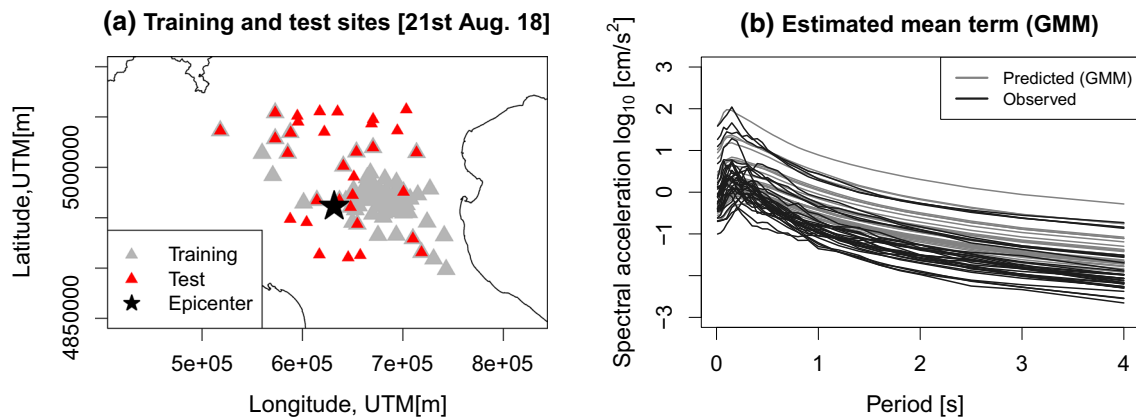
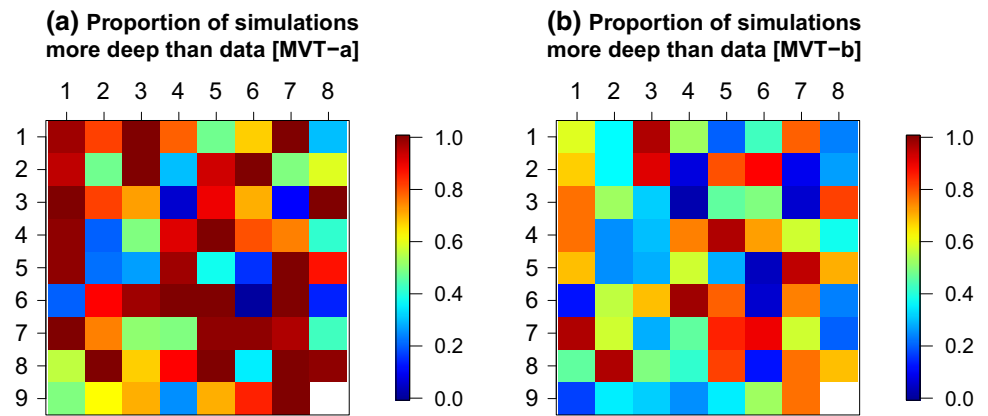


Fig. 16 Application to the independent event MI 4.2 registered in Northern Italy on Aug. 21st, 2018. **a** Locations in the training set and recording stations for the independent event. **b** Estimation of the expected $\log_{10}(SA)$ when using the GMM only, compared with the observations

Fig. 16a reports the positions of the training measurement stations (grey symbols) and those where measurements for the independent event were available (red symbols). Figure 16b displays the estimation of the expected value of $\log_{10}(SA)$ at the test stations as predicted through the NI15 model; here, the values obtained through the NI15 model for a set of 25 periods were interpolated via cubic splines yielding the functional representation of the estimates (light grey curves). Figure 16b also depicts the measured spectral acceleration at the same stations (dark grey curves). A general agreement in the shape of the predicted $\log_{10}(SA)$ can be noted.

The functional geostatistical model estimated in Sect. 4 was used to generate $B = 1000$ realizations of the corrective term (conditional simulation), and of the error term (unconditional simulation). These were added to the estimated GMM as in (7), to obtain $B = 1000$ realizations of the field at the test locations. Figure 17b-c report a visualization of the shaking scenarios simulated at two test stations (*MODE* in Fig. 17b, *MNS* in Fig. 17c); the results at the remaining 30 test stations are reported in the

Supplementary Material. In Fig. 17b and c, similarly as in Sect. (5), functional boxplots are displayed to enhance visualization of the generated curves; thick red lines represent the actual measures, whereas dashed lines represent the estimation of the GMM. At the test station *MODE*, results on the test set show an overall slight overestimation of the shaking field for $T > 1$ [s]. Comparison of the observed measure with the prediction of the GMM suggests that the bias may be due to the corrective terms \mathcal{X}_s . Note that this result is in agreement with the analysis shown in Sect. 5, as the station *MODE* was associated with an extreme behavior in terms of the corrective component. On the contrary, beside the relatively high CV-error at station *MNS* (see Sect. 5), the corrective term appears to balance the error due to the GMM, leading to a set of simulated scenarios perfectly resembling the actual measures.

The map in Fig. 17a displays with colors the centrality of the observed measures with respect to the corresponding set of simulations. Here, the color assigned at a given location is representative of the proportion of simulated responses less deep (in the sense of MBD) than the actual

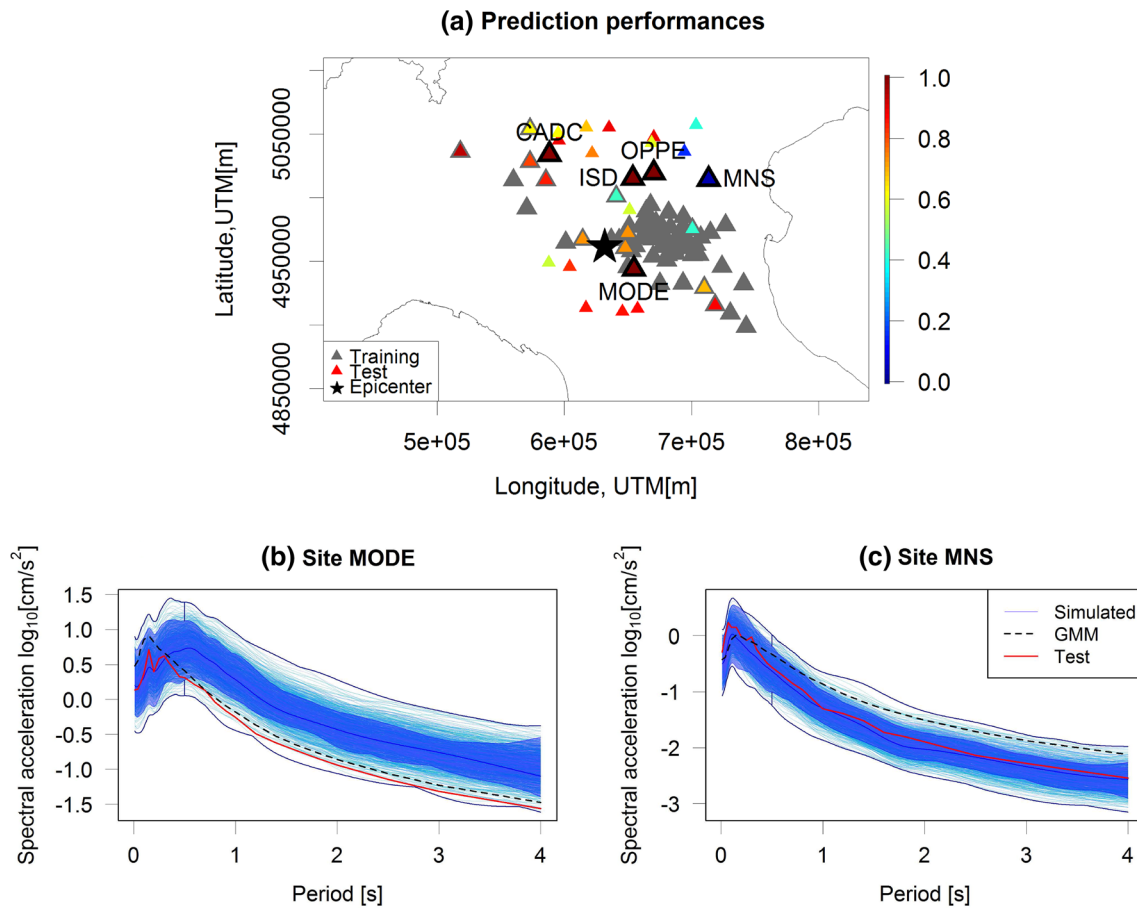


Fig. 17 Shaking scenarios for the independent event M14.2 registered on Aug. 21st, 2018. **a** Locations in the training and test set. **b** Shaking scenarios simulated for the test site *MODE* ($s=(654613, 4943664)$). **c** Shaking scenarios simulated for the test site *MNS* ($s = (713603, 5014520)$). Site coordinates are in Lat/Lon UTM Zone

32N. Colors in panel **a** are given according to the proportion of simulated realization that are less deep (in terms of MBD) than the actual measure in the test set. In panels **b** and **c**, simulated profiles are represented as blue curves, profiles predicted via GMM as dashed black curves, observed profiles as red curves

observations (i.e., red colors are given if the observation is extreme with respect to the simulated sample, blue colors otherwise). Results suggest that a number of stations (particularly *OPPE*, *ISD*, *CADC*) would deserve further attention, as they appear to be associated with ineffective predictions of the corrective components, possibly due to non-systematic effects related either with source or site and only partially captured by the model. Nonetheless, overall, the simulated scenarios appear in agreement with the actual independent measures.

8 Conclusions and discussion

In this work, a novel approach to the analysis of the corrective terms of a ground motion model (GMM) has been proposed, leading to the following key conclusions.

1. Spectral accelerations $\{SA_s(T), T \in \mathcal{T}\}$ in a GMM have been here interpreted as functional data, distributed in space. In this vein, we used the theory of Object Oriented Spatial Statistics (O2S2, Menafoglio and Secchi 2017) to provide a novel modeling framework where stochastically generated scenario-based spectral acceleration profiles can be built. A key advantage of our approach over previous methods relies in the possibility of obtaining shaking fields jointly and consistently for all the spectral periods.
2. The modeling effort of this work has been focused on the residual terms of the region-specific GMM model by Lanzano et al. (2016) and the corrective terms calibrated for the Po-Plain area by Sgobba et al. (2019). A functional approach to the estimate of the GMM model (i.e., model for the term μ_{es} in (1)) is currently under investigation by the authors. The combination of the latter model, with that provided in

this work will eventually allow for a fully functional approach to the analysis of spectral acceleration profiles.

3. The method was validated through 10-fold cross-validation. Data left in the test set were compared with the simulated scenarios by using functional boxplots based on Modified Band Depth (MBD) measures. Analyses suggest very good performances of the method. Indeed, test observations were mostly well-captured by the simulated scenarios, in terms of their MBD with respect to the simulated curves.
4. The application of the calibrated model to an independent event M14.2, recorded on Aug. 21st, 2018, demonstrates the ability of the method to well-reproduce shaking scenarios in a predictive framework. Some inconsistencies detected on specific sites may be dependent on non-systematic effects related either with source or site. For instance, the observed shaking pattern of this event provided by Shakemap (<http://shakemap.rm.ingv.it/>) appears to be mainly elongated along the NW–SE direction, thus reflecting some azimuthal effect that cannot be captured by the systematic terms.

In conclusion, the proposed approach, based on the adoption of a fully non-ergodic model in combination with a functional spectral analysis, results in a novel data-driven methodology for shaking simulations that could be potentially implemented as a new tool to reproduce shaking maps. Compared to the existing strategies, it allows to predict the whole spectrum, jointly and consistently on a continuous period range with reasonable computational burden, as opposed to the standard approaches, which typically provide spectral patterns at a discrete set of vibration periods. The possibility to reproduce more spectral parameters could be important to represent the structural response to shaking more thoroughly, for emergency planning. The method is also suitable to reproduce hypothetical scenarios of potential future events, to be used for region-specific risk analysis and loss models. Finally, the application-oriented potential of the method enables for an effective reproduction of the local patterns, which could be designed for near-real time applications, as well as for post-processing of historical earthquake shaking wherever poorly-constrained by a limited number of sampled records.

Acknowledgements The authors are grateful to Filippo Lentoni for the development of part of the codes used in the first steps of the analysis. They would also like to thank the two anonymous reviewers for their constructive comments and efforts towards improving our manuscript.

Funding Open access funding provided by Politecnico di Milano within the CRUI-CARE Agreement.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

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

- Abrahamsen P, Benth F (2001) Kriging with inequality constraints. *Math Geol* 33(6):719–744. <https://doi.org/10.1023/A:1011078716252>
- Al-Atik L, Abrahamson N, Bommer J, Scherbaum F, Cotton F, Kuehn N (2010) The variability of ground-motion prediction models and its components. *Seismol Res Lett* 81(5):794–801
- Anderson J, Brune J (1999) Probabilistic seismic hazard analysis without the ergodic assumption. *Seismol Res Lett* 70(1):19–28. <https://doi.org/10.1785/gssrl.70.1.19>
- Anderson J, Uchiyama Y (2011) A methodology to improve ground-motion prediction equations by including path corrections. *Bull Seismol Soc Am* 101(4):1822–1846. <https://doi.org/10.1785/0120090359>
- Baltay A, Hanks T, Abrahamson N (2017) Uncertainty, variability, and earthquake physics in ground-motion prediction equations. *Bull Seismol Soc Am* 107(4):1754–1772
- Chakraborty A, Chaudhuri P (2014) The spatial distribution in infinite dimensional spaces and related quantiles and depths. *Ann Stat* 42(3):1203–1231
- Chilès JP, Delfiner P (1999) *Geostatistics: modeling spatial uncertainty*. Wiley, New York
- Cressie N (1993) *Statistics for spatial data*. Wiley, New York
- Delicado P, Giraldo R, Comas C, Mateu J (2010) Statistics for spatial functional data. *Environmetrics* 21(3–4):224–239
- Ferraty F, Vieu P (2006) *Nonparametric functional data analysis: theory and practice*. Springer, New York
- Goda K, Atkinson G (2010) Intraevent spatial correlation of ground-motion parameters using SK-net data. *Bull Seismol Soc Am* 100(6):3055–3067. <https://doi.org/10.1785/0120100031>
- Goda K, Hong H (2008) Spatial correlation of peak ground motions and response spectra. *Bull Seismol Soc Am* 98(1):354–365. <https://doi.org/10.1785/0120070078>
- Haciefendioğlu K, Alpaslan E (2014) Stochastically simulated blast-induced ground motion effects on nonlinear response of an industrial masonry chimney. *Stoch Environ Res Risk Assess* 28(2):415–427
- Haciefendioğlu K, Banerjee S, Soyuluk K, Alpaslan E (2015) Stochastic dynamic analysis of a historical masonry bridge under surface blast-induced multi-point ground motion. *Stoch Environ Res Risk Assess* 29(5):1275–1286

- Huang C, Galasso C (2019) Ground-motion intensity measure correlations observed in Italian strong-motion records. *Earthq Eng Struct Dyn* 48(15):1634–1660
- Ieva F, Paganoni A (2013) Depth measures for multivariate functional data. *Commun Stat Theory Methods* 42(7):1265–1276
- Jayaram N, Baker J (2010) Considering spatial correlation in mixed-effects regression and the impact on ground-motion models. *Bull Seismol Soc Am* 100(6):3295–3303. <https://doi.org/10.1785/0120090366>
- Kotha S, Bindi D, Cotton F (2016) Partially non-ergodic region specific GMPE for Europe and middle-east. *Bull Earthq Eng* 14(4):1245–1263. <https://doi.org/10.1007/s10518-016-9875-x>
- Kuehn N, Abrahamson N, Walling M (2019) Incorporating nonergodic path effects into the NGA-West2 ground-motion prediction equations. *Bull Seismol Soc Am* 109(2):575–585. <https://doi.org/10.1785/0120180260>
- Landwehr N, Kuehn N, Scheffer T, Abrahamson N (2016) A nonergodic ground-motion model for California with spatially varying coefficients. *Bull Seismol Soc Am* 106(6):2574–2583. <https://doi.org/10.1785/0120160118>
- Lanzano G, D'Amico M, Felicetta C, Puglia R, Luzi L, Pacor F, Bindi D (2016) Ground-motion prediction equations for region-specific probabilistic seismic-hazard analysis. *Bull Seismol Soc Am* 106(1):73–92. <https://doi.org/10.1785/0120150096>
- Lanzano G, Pacor F, Luzi L, D'Amico M, Puglia R, Felicetta C (2017) Systematic source, path and site effects on ground motion variability: the case study of Northern Italy. *Bull Earthq Eng* 15(11):4563–4583
- Lanzano G, Sgobba S, Pacor F, Luzi L, Puglia R, D'Amico M, Felicetta C (2018) Spatial correlation of the systematic site- and path-specific corrections of a GMPE calibrated in Northern Italy. In: *Proceedings of the 16th European conference earthquake engineering*, pp 1–12
- Lin P, Chiou B, Abrahamson N, Walling M, Lee C, Cheng C (2011) Repeatable source, site, and path effects on the standard deviation for empirical ground-motion prediction models. *Bull Seismol Soc Am* 101(5):2281–2295. <https://doi.org/10.1785/0120090312>
- Loth C, Baker J (2013) A spatial cross-correlation model of spectral accelerations at multiple periods. *Earthq Eng Struct Dyn* 42(3):397–417. <https://doi.org/10.1002/eqe.2212>
- Mariethoz G, Caers J (2015) Multiple-point geostatistics: stochastic modeling with training images. Wiley, Hoboken
- Meletti C, Galadini F, Valensise G, Stucchi M, Basili R, Barba S, Vannucci G, Boschi E (2008) A seismic source zone model for the seismic hazard assessment of the Italian territory. *Tectonophysics* 450(1):85–108
- Menafoglio A, Petris G (2016) Kriging for Hilbert-space valued random fields: the operatorial point of view. *J Multivariate Anal* 146:84–94
- Menafoglio A, Secchi P (2017) Statistical analysis of complex and spatially dependent data: a review of object oriented spatial statistics. *Eur J Oper Res* 258(2):401–410
- Menafoglio A, Secchi P (2019) O2S2: a new venue for computational geostatistics. *Appl Comput Geosci* 2:100007
- Menafoglio A, Secchi P, Dalla Rosa M (2013) A universal kriging predictor for spatially dependent functional data of a Hilbert space. *Electron J Stat* 7:2209–2240
- Menafoglio A, Grujic O, Caers J (2016a) Universal kriging of functional data: trace-variography vs cross-variography? Application to forecasting in unconventional shales. *Spatial Stat* 15:39–55
- Menafoglio A, Guadagnini A, Secchi P (2016b) Stochastic simulation of soil particle-size curves in heterogeneous aquifer systems through a Bayes space approach. *Water Resour Res* 52:5708–5726
- Michellini A, Faenza L, Lanzano G, Lauciani V, Jozinović D, Puglia R, Luzi L (2019) The new ShakeMap in Italy: progress and advances in the last 10 yr. *Seismol Res Lett*. <https://doi.org/10.1785/0220190130>
- Nerini D, Monestiez P, Manté C (2010) Cokriging for spatial functional data. *J Multivariate Anal* 101(2):409–418
- Pacor F, Paolucci R, Luzi L, Sabetta F, Spinelli A, Gorini A, Nicoletti M, Marcucci S, Filippi L, Dolce M (2011) Overview of the Italian strong motion database ITACA 1.0. *Bull Earthq Eng* 9(6):1723–1739
- Paolucci R, Mazzieri I, Smerzini C (2015) Anatomy of strong groundmotion: near-source records and three-dimensional physics-based numerical simulations of the Mw 6.0 (2012 May 29) Po Plain earthquake, Italy. *Geophys J Int* 203(3):2001–2020. <https://doi.org/10.1093/gji/ggv405>
- Pebesma EJ (2004) Multivariable geostatistics in S: the GSTAT package. *Comput Geosci* 30:683–691
- Porcu E, Fassò A, Barrientos S, Catalán PA (2017) *Seismomatics*. *Stoch Environ Res Risk Assess* 31(7):1577–1582
- R Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>
- Ramsay J, Silverman B (2005) *Functional data analysis*, 2nd edn. Springer, New York
- Rodriguez-Marek A, Cotton F, Abrahamson N, Akkar S, Al Atik L, Edwards B, Montalva G, Dawood H (2013) A model for single-station standard deviation using data from various tectonic regions. *Bull Seismol Soc Am* 103(6):3149–3163. <https://doi.org/10.1785/0120130030>
- Sahakian V, Baltay A, Hanks T, Buehler J, Vernon F, Kilb D, Abrahamson N (2019) Ground-motion residuals, path effects, and crustal properties: a pilot study in southern California. *J Geophys Res Solid Earth*. <https://doi.org/10.1029/2018JB016796>
- 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* 203:103139
- Sgobba S, Lanzano G, Pacor F, Puglia R, D'Amico M, Felicetta C, Luzi L (2019) Spatial correlation model of systematic site and path effects for ground-motion fields in northern Italy. *Bull Seismol Soc Am* 109(4):1419–1434
- 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>
- Sun Y, Genton M (2009) Functional boxplots. *J Comput Graph Stat* 20(2):316–334
- Verros DJCBMHSA, Ganesh M (2017) Computing spatial correlation of ground motion intensities for ShakeMap. *Comput Geosci* 99:145–154
- Villani M, Abrahamson N (2015) Repeatable site and path effects on the ground-motion sigma based on empirical data from southern California and simulated waveforms from the CyberShake platform. *Bull Seismol Soc Am* 105(5):2681–2695. <https://doi.org/10.1785/0120140359>
- Walling M (2009) Non-ergodic probabilistic seismic hazard analysis and spatial simulation of variation in ground motion. PhD thesis, University of California, Berkeley
- Weatherill G, Silva V, Crowley H, Bazzurro P (2015) 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>
- Worden C, Thompson E, Baker J, Bradley B, Luco N, Wald D (2018) Spatial and spectral interpolation of ground-motion intensity measure observations. *Bull Seismol Soc Am* 108(2):866–875. <https://doi.org/10.1785/0120170201>
- Worden C, Wald D (2016) Shakemap manual online: technical manual, user's guide, and software guide U.S. Geol. Surv. <https://doi.org/10.5066/F7D21VPQ>

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