Earthquake Spectra

RESEARCH PAPER

Spatial Cross-Correlation Models for Next-Generation Amplitude and Cumulative Intensity Measures

Vitor A. Monteiro | Savvinos Aristeidou | Gerard J. O'Reilly

Centre for Training and Research on Reduction of Seismic Risk (ROSE Centre), Scuola Universitaria Superiore IUSS di Pavia, Pavia, Italy

Correspondence: Gerard J. O'Reilly (gerard.oreilly@iusspavia.it)

Received: 23 October 2025 | **Revised:** 31 March 2026 | **Accepted:** 14 April 2026

Keywords: cross-spatial correlation | intensity measures | regional assessment | seismic risk

ABSTRACT

Spatial correlation of ground motion intensity measures (IMs) plays a critical role in seismic hazard and risk assessment, particularly for understanding how IMs vary across sites during an earthquake event. While conventional IMs such as *PGA*, *PGV* and $Sa(T)$ have been widely studied, there is a growing interest in so-called next-generation IMs, such as $Sa_{avg}(T)$ and *FIV*3, due to their increased accuracy in characterising seismic vulnerability. However, to date, no spatial correlation models exist for these next-generation IMs, nor for their cross-spatial correlations with traditional IMs. This study proposes such a model for within-event residuals among five IM types, including both conventional and next-generation ones. Principal component analysis (PCA) and geo-statistical methods were used to develop the model based on 8484 ground motion records from the NGA-West2 and ESM databases. Two modelling strategies were explored: a global PCA formulation that captures the joint spatial behaviour of all IMs and a pairwise-PCA approach. The pairwise strategy improves the representation of semivariograms and cross-semivariograms for specific IM combinations. However, it is primarily intended for pairwise analysis, while the global model provides a consistent framework for joint simulation of multiple IMs. The proposed model shows good agreement with the existing spatial correlation models for traditional IMs. The study found that clustering datasets by moment magnitude, M_w, did not significantly impact spatial correlations. In contrast, clustering based on site conditions, specifically V_{s30} (i.e., soft soil vs. hard rock), notably influenced the correlation structure and was hence considered. Furthermore, illustrative applications indicate that neglecting cross-IM correlation can lead to underestimation of joint exceedance probabilities. Given the current lack of spatial correlation models for next-generation IMs, this study fills a critical gap by providing tools that support their integration into more accurate regional seismic hazard and risk modelling frameworks.

1 | Introduction

Regional-scale seismic risk assessment has been an active research topic since the early 2000s (e.g., Wang and Takada 2005; Park et al. 2007) and continues to evolve with recent developments (e.g., Bodenmann et al. 2023; Heresi and Miranda 2023; Acevedo et al. 2025; Bantis et al. 2025). A central objective at this scale is to quantify how earthquake-induced losses and risk are distributed across large building portfolios and geographic regions. This has

motivated several research groups to develop spatial correlation models for ground shaking since neglecting spatial correlation has been shown to significantly bias regional loss estimates (e.g., Park et al. 2007).

Spatial correlation of ground motion intensity measures (IMs) has been modelled extensively over the past two decades using various methodologies and datasets. Approaches including analytical formulations (e.g., Goda and Atkinson 2009; Aldea et al. 2022), geo-statistical tools such as semivariograms (e.g., Jayaram and Baker 2009; Esposito and Iervolino 2011) and more advanced methods accounting for path and site effects (e.g., Bodenmann et al. 2023) are all examples of possible models that can be found in the literature (Monteiro and O'Reilly 2026). Spatial cross-correlation models, which describe the joint spatial behaviour of multiple IMs, have also been developed using diverse methods. These include the Markov-type screening hypothesis (e.g., Goda and Hong 2008), linear model of coregionalisation (LMC) (e.g., Loth and Baker 2013), principal component analysis (PCA) (e.g., Markhvida et al. 2018; Du and Ning 2021) and latent dimensions (LD) (e.g., Abbasnejadfard et al. 2020). A recent review by Monteiro and O'Reilly (2026) compiled over 45 studies on spatial and non-spatial correlation modelling, comparing methodologies, IMs and datasets used in regional risk assessments.

While significant advances have been made in developing spatial correlation models that allow the simulation of ground motion fields (GMFs) for a wide range of IMs (e.g., Weatherill et al. 2014; Silva and Horspool 2019), parallel progress in vulnerability modelling has led to increased use of more sophisticated IMs, such as average spectral acceleration, $Sa_{avg}(T)$, and filtered incremental velocity, $FIV3$, among others, which better capture structural demands (e.g., O'Reilly 2021; Eads et al. 2015; Kazantzi and Vamvatsikos 2015; Dávalos and Miranda 2019). However, the development of spatial correlation models that allow GMFs to be developed for regional studies has not kept pace with these trends. In particular, there is a specific lack of established models for the spatial and cross-spatial correlation of these next-generation IMs, or their relationship with traditional IMs like $Sa(T)$, PGA, PGV, etc. This gap limits analysts' ability to perform more advanced and accurate regional seismic risk assessments using the most informative IMs.

This article proposes a spatial cross-correlation model for within-event residuals across several IMs, including both conventional and next-generation types. The model is developed using PCA and geo-statistical tools based on a combined dataset of 8484 ground motions from the next generation attenuation relationships (NGA-W2) for Western US (Ancheta et al. 2014) and European Strong Motion (ESM) databases (Lanzano et al. 2019). Two modelling strategies are explored: a global PCA-based formulation that captures the joint spatial behaviour of all considered IMs and a simplified pairwise fitting approach in which PCA models are calibrated individually for each IM pair. While both approaches are developed and compared in this study, they serve different purposes: the global PCA formulation provides a unified framework for multi-IM simulation, whereas the pairwise approach is introduced to improve the accuracy of individual IM-pair spatial relationships.

In addition to comparing the proposed model with existing cross-spatial correlation models, the study also investigates the influence of clustering the dataset by moment magnitude, M_w, site soil condition, V_{s30}, and the ground motion models (GMMs) used to quantify the residuals and develop correlation models fit for practical implementation. Finally, an illustrative example is presented to demonstrate the generation of cross-GMFs between traditional and next-generation IMs, along with a quantitative analysis of the influence of cross-IM spatial correlation on portfolio-level exceedance metrics.

2 | Ground Motions

2.1 | Available Datasets

A number of extensive ground motion datasets have played a critical role in developing spatial correlation models. Commonly employed datasets include Japan's K-NET and KiK-net networks (Aoi et al. 2004), the RESORCE database (Akkar et al. 2014b), ITACA (Luzi et al. 2008) and the Ridgecrest database (Rekoske et al. 2020). Although each of these could have been suitable for this study, two major databases were selected: the NGA-W2 database compiled by Ancheta et al. (2014) and the ESM database curated by Lanzano et al. (2019). These datasets were used to develop a global cross-IM spatial correlation model for active shallow crustal regions. Here, the term *cross-IM* indicates that the model establishes correlations between different IMs. A total of 59 earthquakes were selected from the combined NGA-W2 and ESM datasets. The records were filtered according to the same criteria adopted in Aristeidou et al. (2024), as summarised in Table 1. In addition to the selection criteria applied to the ground-motion parameters, a minimum usable frequency threshold of 0.17Hz was imposed, which corresponds to a maximum usable period of 6 s for each ground-motion record considered in the database. Furthermore, only earthquakes with at least 40 associated records were retained, meaning that events with fewer records were excluded from the analysis. This procedure resulted in a final dataset comprising 8484 ground-motion

TABLE 1 | Range of input parameters used in the database filtering.

Description	Min	Max
Moment magnitude, M_w	4.5	7.9
Rupture distance, R_{rup} (km)	0.0	300
Joyner–Boore distance, R_{jb} (km)	0.0	300
Distance perpendicular to fault strike, R_x (km)	−297.13	292.39
Hypocentral depth, D_{hyp} (km)	2.3	18.65
Basin depth, $Z_{2.5}$ (m)	0.0	7780
Minimum useable frequency, f (Hz)	0.0	0.17
Depth to top of rupture, Z_{tor} (km)	0.0	16.23
Time-averaged shear-wave velocity, V_{s30} (m/s)	100	1270
Style of faulting	0	4

records, of which 5155 were obtained from NGA-W2 and 3329 from ESM. Figure 1 presents the distribution of moment magnitude, M_w, Joyner–Boore distance, R_{jb}, and site soil condition, V_{s30}, for the final dataset.

2.2 | IMs

The IMs considered in the cross-IM spatial correlation modelling in this study include:

- *PGA*: peak ground acceleration;
- *PGV*: peak ground velocity;
- $Sa(T)$: 5% damped spectral acceleration at a vibration period, T, where $T \in \{0.01, 0.05, 0.075, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, 1.0, 1.5, 2.0, 3.0, 4.0, 5.0\}$ s.
- $Sa_{avg}(T)$: average spectral acceleration, defined as the geometric mean of N linearly spaced spectral acceleration values. Several studies have proven how this IM improved efficiency, sufficiency and predictability in collapse risk and economic loss estimation (e.g., Eads et al. 2015; Kohrangi et al. 2017). The formulation for $Sa_{avg}(T)$ is:

$$\ln\ Sa_{avg}(T) = \frac{1}{N}\sum_{i=1}^{N} \ln\ Sa(c_i T) \tag{1}$$

$$Sa_{avg}(T) = \left(\prod_{i=1}^{N} Sa(c_i T)\right)^{\frac{1}{N}} \tag{2}$$

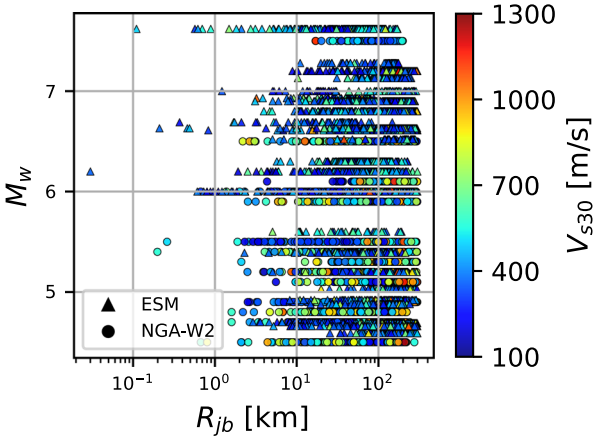

FIGURE 1 | M_w, R_{jb} and V_{s30} distribution of the filtered combined ground motion database.

where $Sa(c_iT)$ corresponds to the 5% damped spectral acceleration value, $N = 10$ typically, c_i is a factor ranging uniformly from 0.2 to 2.0 and 0.2 to 3.0 for what were termed (Shahnazaryan and O'Reilly 2024) $Sa_{avg2}(T)$ and $Sa_{avg3}(T)$, respectively. The periods T selected for Sa_{avg2} are: $T \in \{0.1, 0.15, 0.2, 0.3, 0.4, 0.5, 0.6, 0.75, 0.8, 0.9, 1.0, 1.2, 1.5, 2.0, 2.5, 3.0\}$ s., and for $Sa_{avg3}(T)$ are: $T \in \{0.1, 0.15, 0.2, 0.3, 0.4, 0.5, 0.60, 0.75, 0.8, 0.9, 1.0, 1.2, 1.5, 2.0\}$ s;

- *FIV*3: proposed by Dávalos and Miranda (2019), has shown promising results regarding the efficiency and sufficiency in characterising the collapse performance of the buildings. Basically, it captures the cumulative effect of ground motion pulses by integrating a filtered version of the acceleration signal over a sliding time window. The periods selected are: $T \in \{0.1, 0.15, 0.2, 0.3, 0.4, 0.5, 0.6, 0.75, 0.8, 0.9, 1.0, 1.2, 1.5, 2.0, 2.5, 3.0, 4.0\}$ s. This IM can be briefly explained as

$$FIV3 = max\{V_{s,max1} + V_{s,max2} + V_{s,max3}; |V_{s,min1} + V_{s,min2} + V_{s,min3}|\} \tag{3}$$

$$V_s(t) = \int_t^{t+\alpha T} \ddot{u}_{gf}(t)\, dt, \quad \forall t < t_{end} - \alpha T \tag{4}$$

where $V_s(t)$ represents a sequence of incremental velocity (IV) values computed over moving time windows of length of αT. From this series, the three largest and three smallest local extremes, $V_{s,max1}$, $V_{s,max2}$, $V_{s,max3}$ and $V_{s,min1}$, $V_{s,min2}$, $V_{s,min3}$, respectively, are extracted. The variable T corresponds to the period of interest, while t_end denotes the final time step of the acceleration time history. The acceleration signal $\ddot{u}_{gf}$ is obtained by applying a second-order Butterworth low-pass filter to the original motion, using a cutoff frequency $f_c = \beta f$, where $f = 1/T$ and β is a scaling factor. Following the recommendation of Dávalos and Miranda (2019), values of $\alpha = 0.7$ and $\beta = T$ were adopted in the generalised GMM (GGMM) employed here and consequently in this study.

2.3 | GMM and Residuals

The IMs values and corresponding residuals were computed using the GGMM proposed by Aristeidou et al. (2024). Similar to other GMMs developed in the past, this model adopts the conventional lognormal functional form expressed as

$$\log_{10}\ IM_{i,k,m} = \mu_i(X_{k,m}, \theta) + \delta b_{i,k} \cdot \tau_i + \delta w_{i,k,m} \cdot \phi_i \tag{5}$$

where $\log_{10}\ IM_{i_{k,m}}$ is the logarithm of ith IM of interest, $\mu_i(X_{k,m}, \theta)$ is the predicted mean obtained from the GGMM based on explanatory variables $X_{k,m}$ (e.g., moment magnitude, M_w, rupture distance, R_{jb}, site conditions, V_{s30}, etc.), model parameters θ for a given earthquake k at site m, $\delta b_{i,k}$ and $\delta w_{i,k,m}$ are the normalised between- and within-event residuals of IM_i, respectively, and finally τ_i and ϕ_i are the between- and within-event logarithmic standard deviations, respectively. In spatial correlation models, normalised within-event residuals are utilised, so the formulation to calculate them can be deduced from Equation (5), solving for $\delta w_{i,k,m}$ as

$$\delta w_{i,k,m} = \frac{\log_{10}\ IM_{i,k,m} - \mu_i(X_{k,m}, \theta) - \delta b_{i,k} \cdot \tau_i}{\phi_i} \tag{6}$$

Previous studies have shown that normalised residuals are well represented by a normal distribution, with noticeable deviations only beyond approximately four standard deviations Baker et al. (2021). Furthermore, IM residuals at multiple sites are commonly modelled as following a multivariate normal distribution with mean zero and an associated covariance structure Jayaram and Baker (2008).

3 | Spatial Correlation Model Development

In this section, the theoretical basis to develop a spatial correlation model is first presented using normalised within-event residuals for each IM mentioned before. The approach then follows the PCA-based methodology initially employed by Markhvida et al. (2018), combining PCA and geo-statistical methods to develop the proposed spatial correlation model.

3.1 | Theoretical Background

When analysing spatial correlation across multiple ground motion parameters or IMs (PGA, PGV, $Sa(T)$, $Sa_{avg}(T)$, $FIV3$), a multivariate approach is necessary. A multivariate random field extends the univariate case by representing the random function as a vector of variables at each spatial location. To quantify spatial dependence within and between different components of this multivariate field, the cross-semivariogram has been used (e.g., Loth and Baker 2013; Wang and Du 2013; Markhvida et al. 2018; Du and Ning 2021). For two random functions $Z_{IM_i}(x)$ and $Z_{IM_j}(x)$, representing residuals of IMs IM_i and IM_j at spatial locations x and $x+h$, the cross-semivariogram is defined as

$$\gamma_{IM_i,IM_j}(h) = \frac{1}{2}\mathbb{E}[(Z_{IM_i}(x) - Z_{IM_i}(x+h))(Z_{IM_j}(x) - Z_{IM_j}(x+h))] \tag{7}$$

GMMs already account for the systematic scaling of IMs with respect to magnitude, site conditions, distance, etc. After removing this trend, the remaining residuals represent the unexplained variability, within which a component exhibits spatial dependence. Residuals, therefore, provide the appropriate input to the cross-semivariogram, ensuring that the dependence structure is not biased by large-scale deterministic effects. Since the theoretical formulation in Equation (7) cannot be evaluated analytically, it is common to approximate it empirically using observed residuals from recorded ground motions. This leads to the empirical cross-semivariogram, given as

$$\gamma_{IM_i,IM_j}(h) = \frac{1}{2N_{i,j}(h)}\sum_{\alpha=1}^{N_{i,j}(h)}[(z_{IM_i}(x_\alpha) - z_{IM_i}(x_\alpha+h))(z_{IM_j}(x_\alpha) - z_{IM_j}(x_\alpha+h))] \tag{8}$$

where $N_{i,j}(h)$ is the number of observed data pairs separated by an absolute distance h (normally computed using the haversine formula for geographic coordinates). This empirical function summarises how cross-IM similarity decays with separation distance.

Starting from Equation (7), the cross-semivariogram can be expanded as

$$\begin{aligned}\gamma_{IM_i,IM_j}(h) &= \tfrac{1}{2}\mathbb{E}[(Z_{IM_i}(x) - Z_{IM_i}(x+h))(Z_{IM_j}(x) - Z_{IM_j}(x+h))]\\ &= \tfrac{1}{2}\mathbb{E}\big[Z_{IM_i}(x)\cdot Z_{IM_j}(x) - Z_{IM_i}(x)\cdot Z_{IM_j}(x+h) - Z_{IM_j}(x)\cdot Z_{IM_i}(x+h)\\ &\quad + Z_{IM_i}(x+h)\cdot Z_{IM_j}(x+h)\big]\\ &= \tfrac{1}{2}\Big(\mathbb{E}[Z_{IM_i}(x)\cdot Z_{IM_j}(x)] - \mathbb{E}[Z_{IM_i}(x)\cdot Z_{IM_j}(x+h)]\\ &\quad - \mathbb{E}[Z_{IM_j}(x)\cdot Z_{IM_i}(x+h)] + \mathbb{E}[Z_{IM_i}(x+h)\cdot Z_{IM_j}(x+h)]\Big)\end{aligned} \tag{9}$$

Under second-order stationarity, the mean and covariance structure are assumed not to vary with location, which can then be exploited to give

$$\begin{aligned}\mathbb{E}[Z_{IM_i}(x)\cdot Z_{IM_j}(x)] &= C_{IM_i,IM_j}(0) + \mu_{IM_i}\cdot\mu_{IM_j}\\ \mathbb{E}[Z_{IM_i}(x)\cdot Z_{IM_j}(x+h)] &= C_{IM_i,IM_j}(h) + \mu_{IM_i}\cdot\mu_{IM_j}\\ \mathbb{E}[Z_{IM_j}(x)\cdot Z_{IM_i}(x+h)] &= C_{IM_j,IM_i}(h) + \mu_{IM_i}\cdot\mu_{IM_j}\\ \mathbb{E}[Z_{IM_i}(x+h)\cdot Z_{IM_j}(x+h)] &= C_{IM_i,IM_j}(0) + \mu_{IM_i}\cdot\mu_{IM_j}\end{aligned} \tag{10}$$

where $C_{IM_i,IM_j}(0)$ and $C_{IM_i,IM_j}(h)$ are the cross-covariance at zero distance and at distance h, respectively. Substituting into the expression for the cross-semivariogram gives

$$\begin{aligned}&= \tfrac{1}{2}\Big[\Big(C_{IM_i,IM_j}(0) + \mu_{IM_i}\cdot\mu_{IM_j}\Big) - 2\Big(C_{IM_i,IM_j}(h) + \mu_{I+M_i}\cdot\mu_{IM_i}\Big)\\ &\quad + \Big(C_{IM_j,IM_i}(0) + \mu_{IM_i}\cdot\mu_{IM_j}\Big)\Big]\\ &= \tfrac{1}{2}\Big[2C_{IM_i,IM_j}(0) + 2\mu_{IM_i}\cdot\mu_{IM_j} - 2C_{IM_j,IM_i}(h) - 2\mu_{IM_i}\cdot\mu_{IM_j}\Big]\end{aligned} \tag{11}$$

If the cross-covariance function is symmetric (i.e., $C_{IM_i,IM_j}(h) = C_{IM_j,IM_i}(h)$), the expression simplifies to

$$\gamma_{IM_i,IM_j}(h) = C_{IM_i,IM_j}(0) - C_{IM_i,IM_j}(h) \tag{12}$$

and the cross-covariance function can be written in terms of the cross-semivariogram as follows:

$$C_{IM_i,IM_j}(h) = C_{IM_i,IM_j}(0) - \gamma_{IM_i,IM_j}(h) \tag{13}$$

From this, the spatial correlation function can be derived as (Wang and Du 2013)

$$\rho_{IM_i,IM_j}(h) = \frac{C_{IM_i,IM_j}(0)}{\sqrt{C_{IM_i,IM_i}(0) \times C_{IM_j,IM_j}(0)}} - \frac{\gamma_{IM_i,IM_j}(h)}{\sqrt{C_{IM_i,IM_i}(0) \times C_{IM_j,IM_j}(0)}} \tag{14}$$

To represent the spatial relationships among all N number IMs, the cross-semivariogram matrix of terms, $\Gamma(h)$, the corresponding covariance matrix, $C(h)$, and the correlation matrix, $R(h)$, are defined as

$$\Gamma(h) = [\gamma_{IM_i,IM_j}(h)] = \begin{bmatrix} \gamma_{IM_1,IM_1}(h) & \cdots & \gamma_{IM_1,IM_N}(h) \\ \vdots & \ddots & \cdots \\ \gamma_{IM_N,IM_1}(h) & \cdots & \gamma_{IM_N,IM_N}(h) \end{bmatrix} \tag{15}$$

$$C(h) = [C_{IM_i,IM_j}(h)] = \begin{bmatrix} C_{IM_1,IM_1}(h) & \cdots & C_{IM_1,IM_N}(h) \\ \vdots & \ddots & \cdots \\ C_{IM_N,IM_1}(h) & \cdots & C_{IM_N,IM_N}(h) \end{bmatrix} \tag{16}$$

$$R(h) = [\rho_{IM_i,IM_j}(h)] = \begin{bmatrix} \rho_{IM_1,IM_1}(h) & \cdots & \rho_{IM_1,IM_N}(h) \\ \vdots & \ddots & \cdots \\ \rho_{IM_N,IM_1}(h) & \cdots & \rho_{IM_N,IM_N}(h) \end{bmatrix} \tag{17}$$

Since normalised within-event residuals are often modelled as following a multivariate normal distribution, their spatial variability for a specific earthquake event k can be fully described by their mean and covariance. The mean is assumed to be zero, and the covariance matrix, Σ, captures the spatial dependence across M observation sites, For event k, the full covariance matrix is assembled using the $N \times N$ sub-matrices $C(h)$ for each inter-site separation:

$$\Sigma(event \quad k) = \begin{bmatrix} C(h_{11}) & C(h_{12}) & \cdots & C(h_{1M}) \\ C(h_{21}) & C(h_{22}) & \cdots & C(h_{2M}) \\ \vdots & \vdots & \ddots & \vdots \\ C(h_{M1}) & \cdots & \cdots & C(h_{MM}) \end{bmatrix} \tag{18}$$

Maintaining a positive definite covariance matrix is essential for valid spatial modelling and ensures physically realistic simulation when developing GMFs. Several studies (e.g., Loth and Baker 2013; Wang and Du 2013; Markhvida et al. 2018; Abbasnejadfard et al. 2020; Du and Ning 2021) have proposed different methods to fit cross-semivariograms models while preserving this property.

3.2 | PCA

In PCA, normalised within-event residuals are linearly transformed to an orthogonal basis, which produces uncorrelated principal component (PC) projections in a new coordinate system. Equations (19) and (20) define that linear transformation, where P is the orthogonal linear transformation matrix with size $N \times N$, Z is the matrix with normalised IM within-event residuals for each record so the size is $N \times M$ and Y is the final matrix with the transformed uncorrelated variables with the same shape as matrix Z, where N is the total number of IMs and M is the total number of records.

$$PZ = Y \tag{19}$$

$$\begin{bmatrix} p_{1,IM_1} & \cdots & p_{1,IM_N} \\ \vdots & \ddots & \vdots \\ p_{N,IM_1} & \cdots & p_{N,IM_N} \end{bmatrix} \begin{bmatrix} Z_{IM_1}(x_1) & \cdots & Z_{IM_1}(x_M) \\ \vdots & \ddots & \vdots \\ Z_{IM_N}(x_1) & \cdots & Z_{IM_N}(x_M) \end{bmatrix} = \begin{bmatrix} y_1(x_1) & \cdots & y_1(x_M) \\ \vdots & \ddots & \vdots \\ y_N(x_1) & \cdots & y_N(x_M) \end{bmatrix} \tag{20}$$

The original residuals can be recovered from the PCs as

$$Z = P^{-1}Y = P^{T}Y \tag{21}$$

For a better understanding of how PCA works, Figure 2 demonstrates the steps necessary to go from matrix Z to matrix Y. The process begins with the residuals matrix, denoted as Z, which contains the normalised IM within-event residuals, where each row corresponds to a record and each column to an IM.

PCA reduces data complexity by projecting the original dataset into a smaller set of orthogonal components that capture the main spatial variations. These PCs can be thought of as capturing the main modes of spatial variation in the data, which are much easier to analyse compared to the original multi-dimensional data. This methodology was applied to pooled earthquake data, as there was a good agreement in the mode behaviour between each earthquake individually and the pooled data. Using this methodology, 21 models were developed for each IM pair used in this study. Table 2 summarises the cross-IM spatial correlation models developed in this study, denoted as MAO26, and also some other cross-IM spatial correlation models from the literature.

3.3 | Spatial Correlation Using PCs

Empirical semivariograms and cross-semivariograms were computed for each transformed dataset Y following Equation (8) using a distance bin of 2 km. For this model, the pooled data across all stations were used, resulting in a total of 1,004,030 station pairs. To characterise the spatial dependence structure of the PCs, a nested semivariogram model proposed by Markhvida et al. (2018) was employed. This model, presented in Equation (22), accounts for the initial discontinuity at the origin, commonly referred to as the nugget effect, as well as both short-range and long-range spatial correlation behaviours. In this formulation, $\zeta_{h=0}$ is an indicator function that takes the value of 1 when $h=0$, and 0 otherwise. The parameters c_{0k}, c_{1k}, c_{2k}, a_{1k} and a_{2k} are regression coefficients estimated for each PC k.

$$\gamma(k) = c_{0k}(1-\zeta_{h=0}) + c_{1k}\left(1-\exp\left(\frac{-3h}{a_{1k}}\right)\right) + c_{2k}\left(1-\exp\left(\frac{-3h}{a_{2k}}\right)\right) \tag{22}$$

Figure 3 illustrates the empirical semivariograms and the corresponding fitted nested models for the first 10 PCs. Figure 3a corresponds to the first three PCs, while Figure 3b shows results from the 4th to the 10th PCs. The semivariogram model coefficients, including the nugget, sill and range parameters, were estimated using an ordinary least squares approach, which minimises the sum of squared differences between the empirical semivariance values $\hat{\gamma}(h_i)$ and the values predicted by the nested model $\gamma(h_i;\theta)$:

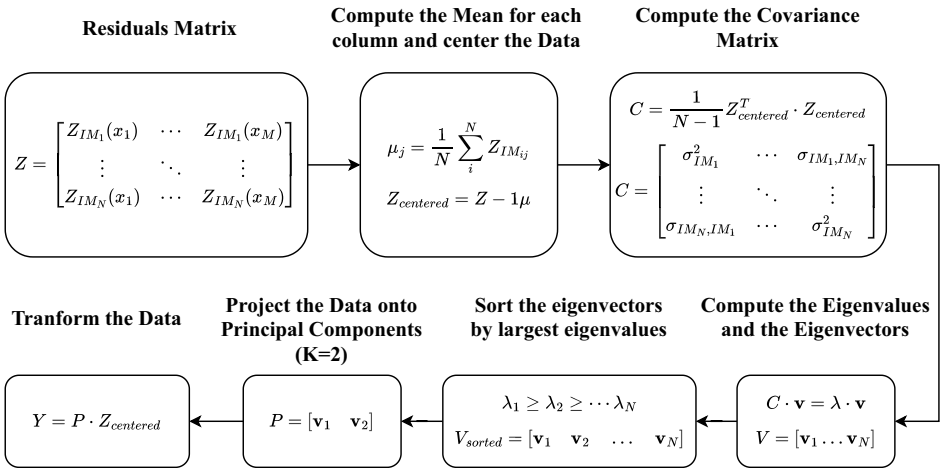

FIGURE 2 | PCA workflow.

TABLE 2 | Cross-IM spatial correlation models proposed here and existing ones from the literature.

IMs	$Sa(T)$	$Sa_{avg2}(T)$	$Sa_{avg3}(T)$	*PGA*	*PGV*	*FIV*3
$Sa(T)$	MAO26, LB13, MCB18, DN21	MAO26	MAO26	MAO26, DN21	MAO26, DN21	MAO26
$Sa_{avg2}(T)$		MAO26	MAO26	MAO26	MAO26	MAO26
$Sa_{avg3}(T)$			MAO26	MAO26	MAO26	MAO26
PGA				MAO26, DN21	MAO26, DN21	MAO26
PGV					MAO26, DN21	MAO26
*FIV*3						MAO26

Note: MAO26: This study; LB13: Loth and Baker (2013); MCB18: Markhvida et al. (2018); DN21: Du and Ning (2021).

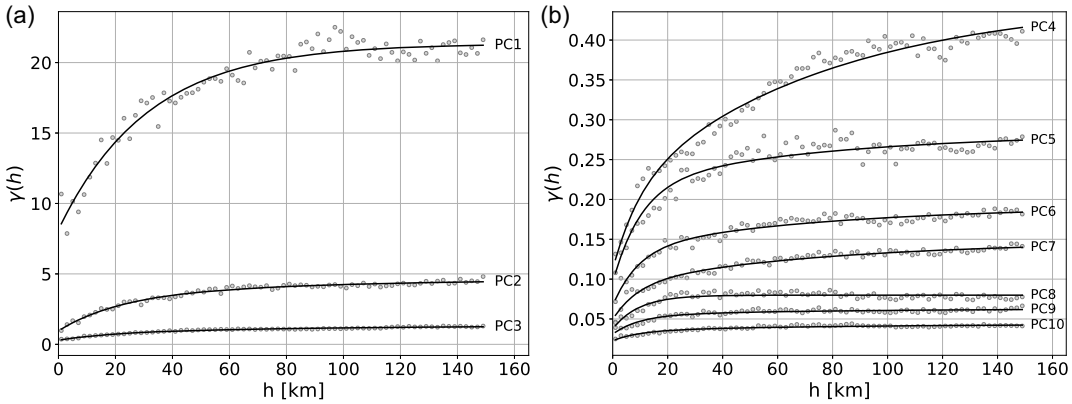

FIGURE 3 | Empirical semivariograms and the fitted models for (a) the first three PCs and (b) for the 4th − 10th PCs.

$$\hat{\theta} = \arg\ \min_{\theta} \sum_{i=1}^{n} (\hat{\gamma}(h_i) - \gamma(h_i;\theta))^2 \tag{23}$$

where $\theta = \{c_0, c_1, a_1, c_2, a_2\}$ are the model parameters (nugget, sills and ranges). This approach provides a standard, conceptually simple fit of the nested semivariograms to the observed data for each PCs. Several other methodologies for semivariogram fitting, including ordinary and weighted least squares (WLS) approaches, are reviewed by Baker and Chen (2020).

As observed in Figure 3, the spatial correlation structure varies considerably with PC order. In this context, it is useful to identify how many PCs of the transformed data are necessary to retain a significant amount of the original variance since the first PC captures most of the variance and so on.

In many cases, particularly for higher order components, the empirical semivariance increases only gradually with distance, which would justify the use of a simplified model that accounts only for the nugget effect.

4 | Results

4.1 | Model Fitting

To validate the proposed model, it is essential to compare the spatial variability derived from empirical ground motion residuals in the original normalised space Z with the modelled cross-semivariograms. This comparison relies on a

multivariate adaptation of the covariance structure to incorporate inter-IM relationships as expressed in Equation (14). Following the formulation by Markhvida et al. (2018), the cross-semivariogram between two IMs, IM_i and IM_j, can be written as

$$\begin{aligned}\gamma_{IM_i,IM_j}(h) &= C_{IM_i,IM_j}(0) - C_{IM_i,IM_j}(h) \\ &= \sum_{k=1}^{N} p_{k,IM_i} p_{k,IM_i} \mathrm{COV}(Y_k(x), Y_k(x)) - \sum_{k=1}^{N} p_{k,IM_j} p_{k,IM_j} \mathrm{COV}(Y_k(x), Y_k(x+h)) \\ &= \sum_{k=1}^{N} p_{k,IM_i} p_{k,IM_j} (C_k(0) - C_k(h)) \\ &= \sum_{k=1}^{N} p_{k,IM_i} p_{k,IM_j} \gamma_k(h)\end{aligned} \tag{24}$$

Here, p_{k,IM_i} and p_{k,IM_j} are the coefficients of the kth PC for the respective IMs, obtained from the PCA transformation matrix P, and N is the total number of PCs used. The term $\gamma_k(h)$ represents the semivariogram of the kth PC evaluated at separation distance h, calculated using Equation (22). This decomposition enables the reconstruction of the cross-semivariogram in the original space as a weighted sum of individual PC semivariograms. Empirical semivariograms and cross-semivariograms are estimated using the residuals of the original IMs, following the procedure described in Equation (8). By comparing these empirical estimates with the model-based reconstructions, it is possible to assess how well the PCA-based approach captures the spatial correlation structure across different IMs.

The semivariogram, variance and covariance functions must be normalised by the explained variance percentage to account for the missing variance. This normalisation was applied by Markhvida et al. (2018) and Du and Ning (2021) as shown in Equations (25), (26) and (27), respectively.

$$\gamma_k(h) = \frac{c_{0k}(1-\zeta_{h=0}) + c_{1k}\left(1 - \exp\left(\frac{-3h}{a_{1k}}\right)\right) + c_{2k}\left(1 - \exp\left(\frac{-3h}{a_{2k}}\right)\right)}{\%\sigma^2_{\text{expl.cum}}} \tag{25}$$

$$C_k(0) = \frac{c_{0k} + c_{1k} + c_{2k}}{\%\sigma^2_{\text{expl.cum}}} \tag{26}$$

$$\begin{aligned}C_k(h) &= C_k(0) - \gamma_k(h) \\ &= \frac{c_{0k}(\zeta_{h=0}) + c_{1k}\exp\left(\frac{-3h}{a_{1k}}\right) + c_{2k}\exp\left(\frac{-3h}{a_{2k}}\right)}{\%\sigma^2_{\text{expl.cum}}}\end{aligned} \tag{27}$$

The cross-variance and cross-covariance for each pair of IMs can be computed with the same methodology as Equation (24), as shown in Equations (28) and (29), respectively. The final correlation, $\rho_{IM_i,IM_j}(h)$, is then obtained by applying Equation (14).

$$C_{IM_i,IM_j}(0) = \sum_{k=1}^{N} p_{k,IM_i} \cdot p_{k,IM_j} \cdot C_k(0) \tag{28}$$

$$C_{IM_i,IM_j}(h) = \sum_{k=1}^{N} p_{k,IM_i} \cdot p_{k,IM_j} \cdot C_k(h) \tag{29}$$

In order to understand how the fitted semivariograms and cross-semivariograms compare to the empirical values, Figure 4 presents examples of these comparisons for different IM pairs. Each plot shows both empirical and modelled semivariograms and cross-semivariograms using all selected PCs in the summation. The results indicate that while the PCA-based model generally preserves spatial correlation patterns across different IM types, including $Sa(T)$, $Sa_{avg}(T)$ and other measures like $FIV3$, some deviations are observed. In particular, pairs like $Sa(T) - FIV3(T)/PGV$ and $PGA - FIV3(T)$ show less accurate fitting, which will be explored further in the next section.

4.2 | Global versus Pairwise Fitting of PCs

The imperfect fitting observed in Figure 4 for some IM pairs was already observed in Markhvida et al. (2018) and Du and Ning (2021), which suggests that the full PCA-based model, which incorporates many IMs simultaneously (Equation (20)),

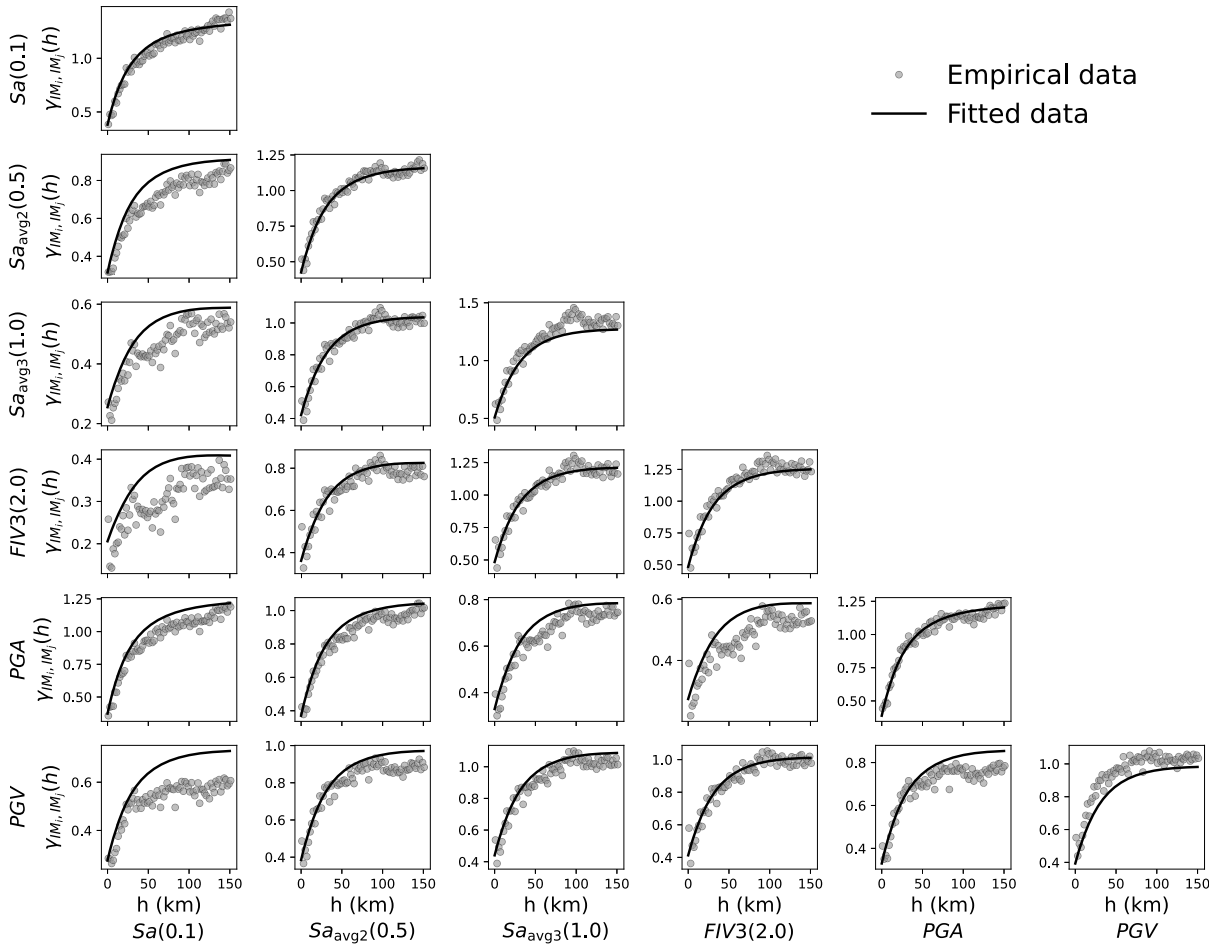

FIGURE 4 | Comparison between empirical semivariogram and model semivariogram using Equation (24) with all PCs for the model used.

sometimes fail to capture pairwise spatial correlations accurately. This is because including multiple IMs compromises the fitting accuracy of specific pairs while trying to maintain overall accuracy. To address this, a simplified modelling approach that focuses on each pair of IMs individually was investigated, using only two PCs relevant to that pair and one PC in the case of using the same IM.

Table 3 summarises the number of PCs used for each model in this work, alongside comparisons to previous studies. For more information regarding the PC coefficients used for each combination pair of IM, please refer to the online GitHub repository. If an IM period not explicitly listed is required, the PC coefficients can be linearly interpolated without significantly impacting the model or its parameters, as noted by Markhvida et al. (2018).

Instead of reconstructing, for example, $Sa(0.1) - Sa(0.5)$ cross-correlation from the global model with 22 PCs covering all $Sa(T)$ periods, each pair was modelled via Equation (20) with just two PCs specifically capturing their joint behaviour. Figure 5a illustrates this approach with the example of $Sa_{avg2}(0.1) - Sa_{avg2}(0.5)$, where the full model uses 22 PCs, while the simplified use only 2 PCs. The simplified model demonstrates a much improved fit to the empirical semivariograms, with no real computational penalty, aside from needing to fit several individual models rather than a single one. Applying this simplified modelling approach, where each IM pair is analysed individually, Figure 5b presents the WLS errors for all pairs that the global model was developed. In this context, the WLS quantifies the discrepancy between the empirical and the modelled semivariograms obtained using Equation (24). As expected, the simplified two-IM model often yields lower WLS errors, indicating a better fit to the observed spatial variability.

The pairwise modelling strategy adopted in this study enhances the fitting accuracy of individual semivariograms and cross-semivariograms by constructing PCA transformations independently for each IM pair. Because each pairwise model is formulated in a two-dimensional space, it produces a valid 2×2 covariance structure. However, when multiple independently calibrated pairwise models are assembled into a higher dimensional covariance matrix, positive semi-definiteness is not formally guaranteed for arbitrary combinations of IMs. This limitation reflects a trade-off between

TABLE 3 | Comparison of the number of principal components used in the simplified models between this study and models developed by Markhvida et al. (2018) and Du and Ning (2021).

Models	This study (global model)	This study (pairwise model)	Markhvida et al. (2018)	Du and Ning (2021)
No. of earthquakes	59	59	45	28
No. of ground motions	8484	8484	4910	3797
No. of PCs used in total	22	2/1*	19	23
No. of PCs used in reduced model	3	2/1*	5	7
$\%\sigma^2_{\text{expl.cum}}$	95	100	95	90

Note: *2 PC were used when applying PCA to different IM pairs (e.g., *PGA–PGV*); only 1 PC was used when the same IM was paired with itself (e.g., *PGA–PGA*).

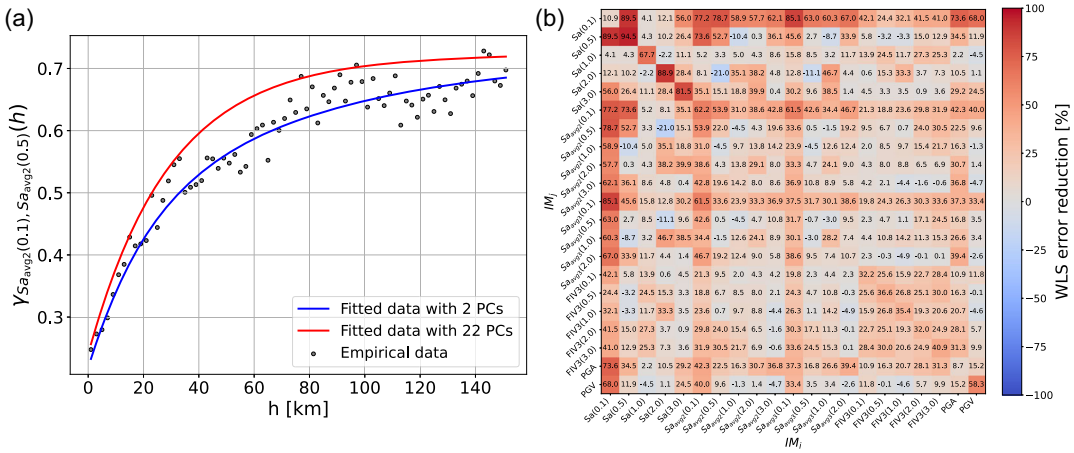

FIGURE 5 | (a) Comparison between PCA-based models with fewer and total amount of PCs for $Sa_{avgsa2}(0.1) - Sa_{avgsa2}(0.5)$ cross-semivariogram model, and (b) error reduction for all semivariograms and cross-semivariograms when used two specific IM pairs and all of the IM pairs simultaneously.

local fitting accuracy and global structural consistency. Joint latent-variable formulations, such as those proposed by Markhvida et al. (2018), or linear models of coregionalisation (e.g., Loth and Baker 2013), enforce positive semi-definiteness through a unified covariance structure. However, when a large number of IMs are included simultaneously, such global approaches may compromise the accuracy of individual pairwise relationships. In contrast, the present methodology prioritises accurate representation of pairwise spatial dependence across a broad set of IMs, without imposing a single global covariance constraint. As a result, the pairwise formulation should be interpreted primarily as a modelling tool for accurately representing pairwise spatial dependence between two IMs. For applications requiring the joint simulation of multiple IMs, the global PCA model should be used, as it provides a unified covariance structure that is internally consistent by construction.

While direct fitting of semivariograms could potentially provide slightly improved agreement with empirical correlations, it does not naturally provide a generative framework for simulating spatially correlated residual fields across multiple IMs. In contrast, the global PCA-based formulation introduces a latent-variable structure in which a reduced number of spatially correlated PCs can be simulated and subsequently transformed back into the original IM space. This enables efficient generation of spatially correlated GMFs while preserving both spatial correlation and cross-IM correlation. In the pairwise formulation, PCA is applied in a two-dimensional setting, and, therefore, no dimensionality reduction is achieved. The computational advantage observed in this case arises from the factorisation of smaller covariance matrices rather than from a reduction in the dimensionality of the system. By contrast, global PCA models involving multiple IMs can achieve true dimensionality reduction and provide a more efficient framework for high-dimensional simulation problems.

Although the pairwise framework can be extended to multi-IM applications, assembling independently calibrated models does not guarantee a positive semi-definite covariance matrix, which may lead to computation inconsistencies. This issue can be addressed through reconstruction of a consistent correlation matrix (see Supplemental Material), for example by projection onto the nearest positive semi-definite matrix (Higham 2002), followed by transformation to covariance form. However such procedures introduce additional processing steps and depart from the unified PCA-based latent-variable structure of the global PCA formulation. A comparative example is provided later in Section 6, indicating that this approach can still achieve reasonable accuracy for the case study considered. Both the pairwise and global models developed in this study are available in the accompanying GitHub repository to enable future study on the feasibility of this alternative.

4.3 | Comparison with Previous Models

The performance of the proposed global and pairwise inter-IM spatial correlation models are compared against previous studies that developed cross-spatial correlation models for several IMs as mentioned in Table 2. Figure 6 presents representative examples of correlograms (plots of correlation coefficients as a function of separation distance) and cross-correlograms for $Sa(T)$ at periods of 0.01, 0.1, 0.5, 1, 2 and 5 s, chosen to illustrate the behaviour across short-, intermediate- and long-period ranges.

Overall, the proposed model exhibits trends broadly consistent with existing formulations. The model of Du and Ning (2021) systematically predicts higher correlation values across periods and separation distances compared to the other models. In contrast, the proposed models shows a somewhat faster decay of spatial correlation with distance than Markhvida et al. (2018), while generally predicting higher correlations than Loth and Baker (2013).

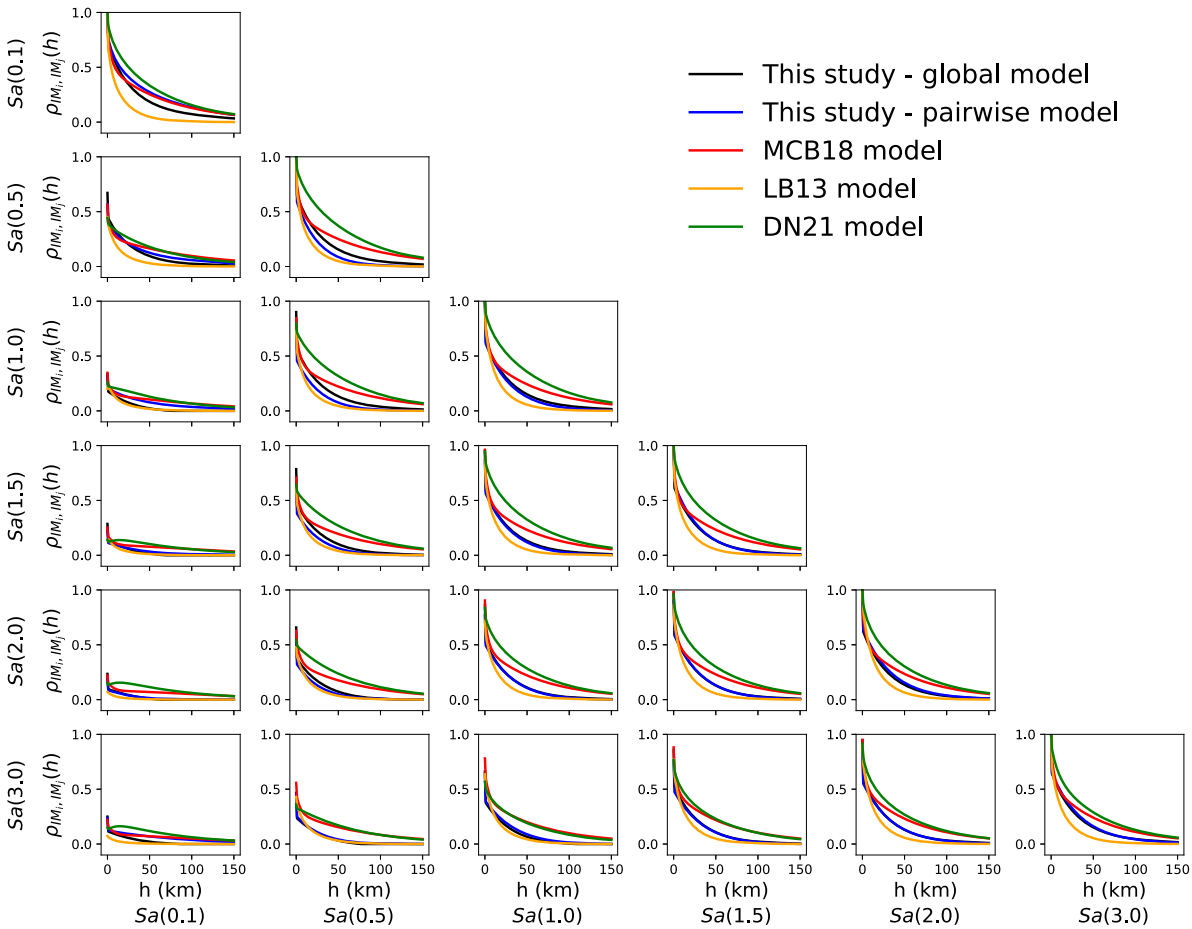

FIGURE 6 | Comparison of the proposed correlograms and cross-correlograms for different $Sa(T)$ between this study and other studies.

These differences observed in Figure 6 likely stem from several methodological factors, including differences in ground motion datasets, modelling assumptions and functional factors, including differences in ground motion datasets, modelling assumptions and functional forms. In particular, the treatment of dimensionality reduction through PCA can have a significant influence on the resulting correlation structure. The approach adopted by Du and Ning (2021), which involves pooling PCs across earthquake events, may lead to smoother and more uniform correlation estimates depending on how event-specific components are combined. For example, variations in the orientation (sign) of PCs across events can affect the aggregated representation if not explicitly harmonised prior to pooling. Such effects could contribute to the comparatively higher and less rapidly decaying correlations observed in their model when compared to others, such as Markhvida et al. (2018) which used similar data.

By contrast, the present study and Markhvida et al. (2018) retain stronger event-specific structure in the dimensionality reduction stage, which may preserve a greater degree of variability and lead to comparatively lower correlation estimates. These observations highlight the sensitivity of multivariate spatial correlation models to implementation details in PCA-based frameworks, even when broadly similar datasets and modelling philosophies are employed.

These differences, observed in the correlograms and cross-correlograms in Figure 6, arise from several methodological choices. Beyond the selection of ground motion databases and functional forms, the treatment of dimensionality reduction via PCA is particularly influential. For instance, the higher correlations in the Du and Ning (2021) model may be related to the way PCs are pooled across different earthquake events, if the signs of these components are not consistently aligned before averaging, the resulting pooled model can undergo a degree of over-averaging. This process can inadvertently smooth out event-specific variability, leading to the higher and more uniform correlation estimates seen in their results. Conversely, the present study and Markhvida et al. (2018) appear to preserve more of this inherent dispersion, resulting in the comparatively lower correlation values that better reflect the diversity of the underlying seismic records.

Further representative comparisons for other IMs, including $Sa_{avg2}(T)$, $Sa_{avg3}(T)$ and $FIV3$, are shown in Figure 7. To the authors' knowledge, no direct inter-IM spatial correlation models exist for $Sa_{avg}(T)$ or $FIV3$. In this study, an indirect estimation of $Sa_{avg2}(T)$ correlations was derived following the methodology proposed by Heresi and Miranda (2021) and compared with the direct model proposed herein, showing good overall agreement between both approaches. The presented model offers the first insight into the spatial correlation structure of these composite IMs. With respect to $FIV3$, Aristeidou et al. (2025) observed strong internal correlations of the IM with itself across different periods. In particular, the spatial correlation analysis revealed that the cross-correlograms both across different $FIV3$ periods and between $FIV3$ and PGV exhibit highly consistent behaviour.

5 | Sensitivity Analysis

To explore potential dependencies in spatial correlation models, this section investigates their behaviour when applied to clustered datasets. Previous studies have examined the role of magnitude (e.g., Azarbakht et al. 2014; Baker and Bradley 2017), with mixed conclusions: some reporting no significant influence, while others found noticeable effects. Site conditions have also been shown to matter. For example, Jayaram and Baker (2009); Wang and Du (2013) demonstrated that clustering databases by V_{s30} values can alter the observed correlations, as the range differs when earthquakes are analysed in relatively non-homogeneous soil conditions. In the following sensitivity analysis, the simplified pairwise model is adopted for all computations and applied consistently across the different clustering datasets.

5.1 | Impact of Earthquake Magnitude

The influence of earthquake magnitude on the spatial correlation modelling was investigated. The same filtering criteria were maintained, but with a stratification of events based on moment magnitude, M_w, to create two subsets: 1) a low-magnitude ($M_w < 6$) events: 35 earthquakes yielding 3312 ground motion records (894 from NGA-W2 and 2418 from ESM); and 2) a high-magnitude ($M_w \geq 6$) events: 24 earthquakes yielding 5171 ground motion records (4261 from NGA-W2 and 910 from ESM). The correlograms and cross-correlograms derived from these two magnitude-based datasets were compared with those from the full (non-clustered) dataset, as shown in Figure 8 bottom. The comparison indicated minimal differences in spatial correlation patterns between the low- and high-magnitude subsets and the full dataset. Similar behaviour was observed across various IMs, including $Sa(T)$, $Sa_{avg2}(T)$, $Sa_{avg3}(T)$ and $FIV3$ at multiple

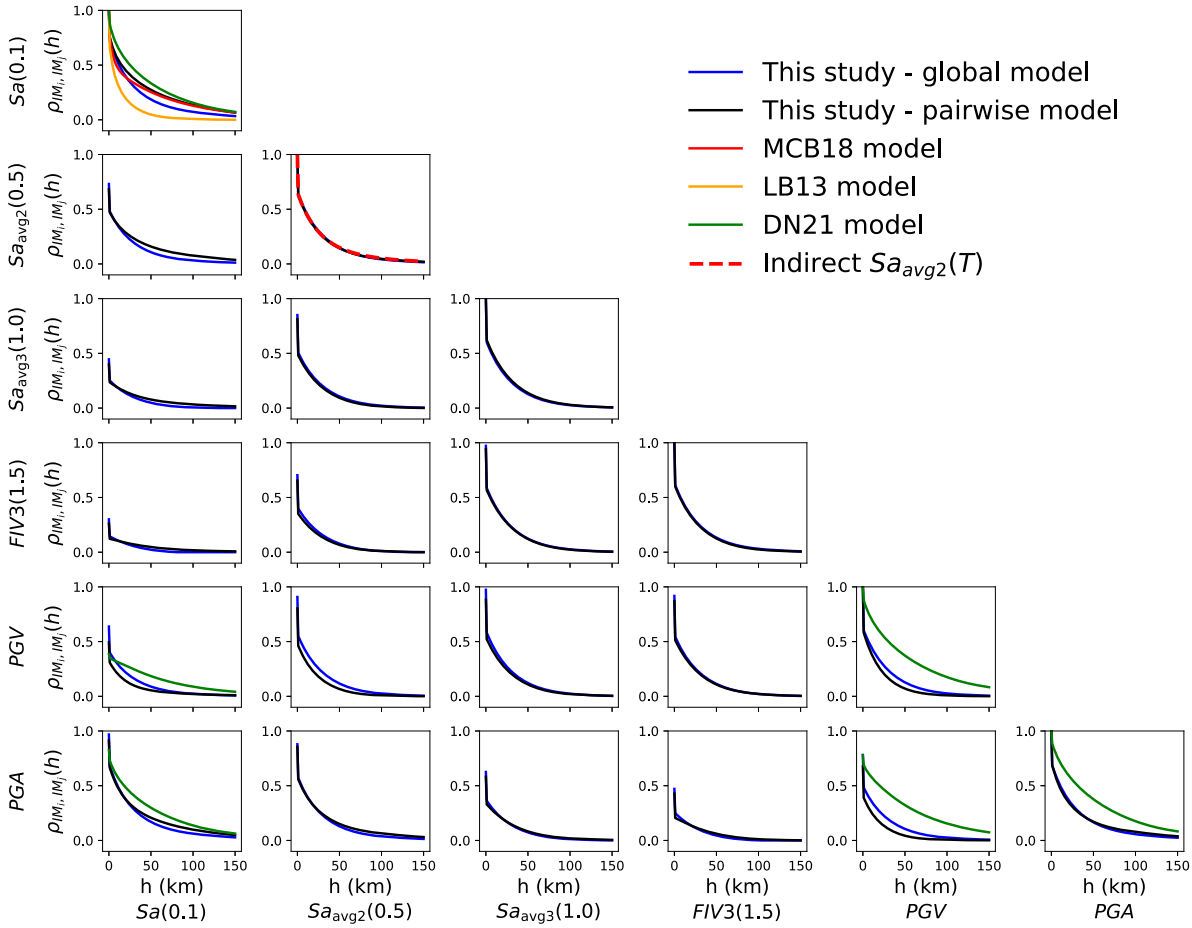

FIGURE 7 | Comparison of predicted correlations and cross-correlations for $Sa(T)$, $Sa_{avg2}(T)$, $FIV3$, PGV and PGA between this study and other studies.

periods. These findings align with previous observations that spatial correlations of IMs are relatively insensitive to earthquake magnitude. However, it is noted that both the low- and high-magnitude subsets exhibit slightly higher $\rho(h)$ values compared to the combined (non-clustered) dataset. A plausible explanation for this behaviour lies in the statistical effects of heterogeneity mixing. When events spanning a broad magnitude range are pooled together, differences in source characteristics and associated between-event variability may introduce additional dispersion in the residuals. This increased variability can effectively dilute the apparent spatial coherence, leading to a modest reduction in the estimated correlation coefficient. In contrast, clustering the dataset by magnitude reduced this inter-event heterogeneity, resulting in more homogeneous residual fields and consequently slightly higher estimated spatial correlations.

5.2 | Impact of Site Soil Conditions

The influence of site soil conditions on the spatial correlation modelling has been previously investigated in the literature (e.g., Jayaram and Baker 2009; Wang and Du 2013). Here, the combined ground motion database was further analysed by stratifying it based on V_{s30} values into two main categories: $180\,\mathrm{m/s} < V_{s30} < 360\,\mathrm{m/s}$ group comprises 16 earthquakes and 1834 recordings (1650 from NGA-W2 and 184 from ESM), while the high-V_{s30} group contains 48 earthquakes and 5334 recordings (2837 from NGA-W2 and 2497 from ESM). Figure 9 top shows the distribution of the number of records by M_w and the corresponding V_{s30} values in each magnitude bin.

Figure 9 bottom presents a comparison of correlograms and cross-correlograms for several IMs, based on the two V_{s30}-clustered datasets and the full (non-clustered) dataset. For most IM pairs, the influence of V_{s30} clustering appears minimal, suggesting that the use of a non-clustered database (black line) is generally reasonable. However, differences emerge with increasing periods and softer soils. As the period of $Sa(T)$ or $Sa_{avg}(T)$ reaches 1.0–1.5 s, the correlation coefficients

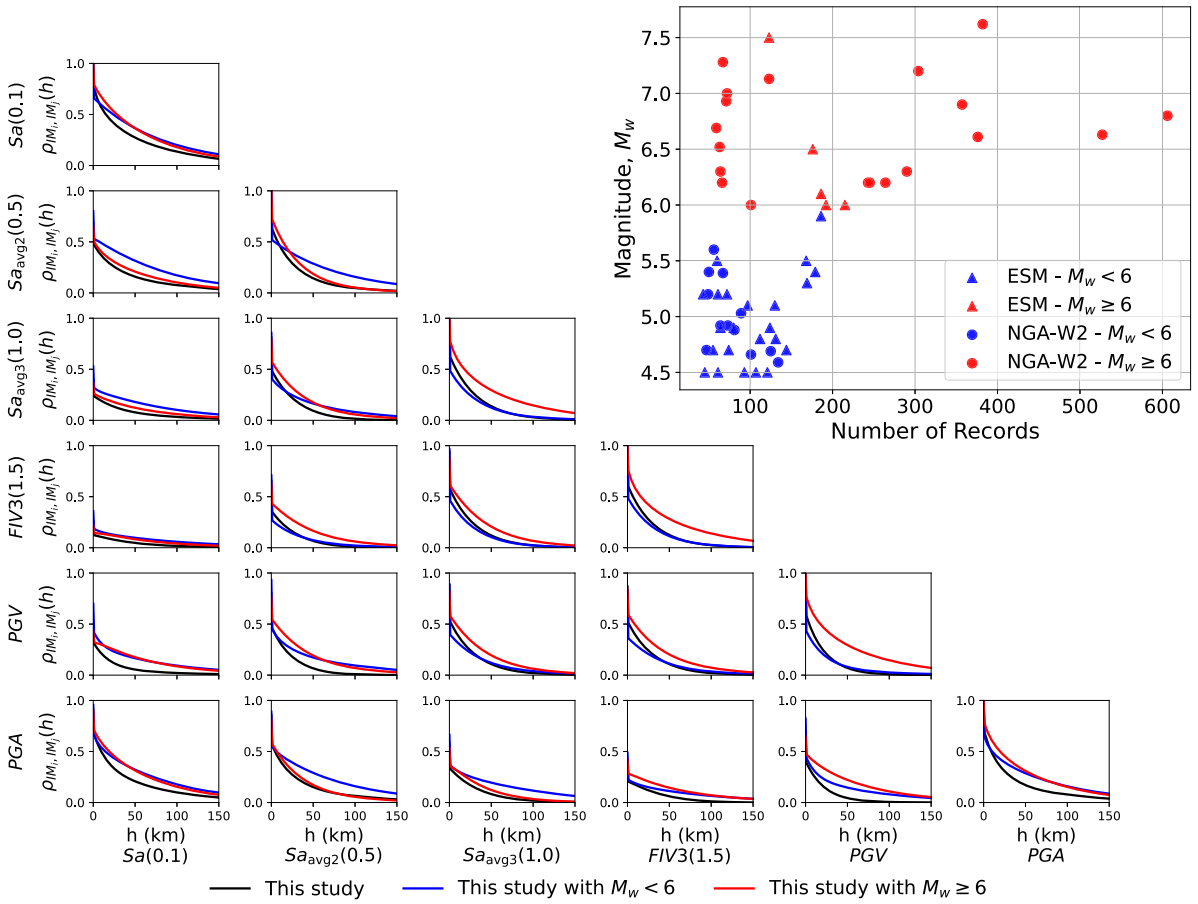

FIGURE 8 | Comparison of predicted correlograms and cross-correlograms for different IMs varying the level of M_w (bottom), and distribution of M_w for the two clustered datasets (top).

of the low and high-V_{s30} groups tend to converge towards the non-clustered dataset. Notably, for shorter periods, the low-V_{s30} group exhibits lower correlation, while at longer periods, it often shows higher correlation compared to the high-V_{s30} group.

To assess whether the differences observed between the two V_{s30}-clustered datasets could be attributed to the disparity in sample sizes, an additional analysis was conducted. Random subsets of approximately 1500 recordings were extracted from the high-V_{s30} dataset (originally containing 5334 recordings), ensuring each subset maintained a minimum of 40 records per earthquake. The results from these subsets closely matched those obtained from the full high-V_{s30} clustered dataset, indicating that the spatial correlation model was not overly dependent on sample size. This finding reinforces the conclusion that the differences between the low and high-V_{s30} groups for $FIV3$ are indeed driven by the underlying site conditions, rather than unequal sample sizes, as previously noted in studies for $Sa(T)$ such as Jayaram and Baker (2009), for example.

5.3 | Impact of GMMs

This study also investigated the impact of different GMMs on spatial and cross-spatial correlation modelling. Figure 10 presents a comparison of correlograms and cross-correlograms obtained in this study with those derived using alternative GMMs, namely, Akkar et al. (2014a); Campbell and Bozorgnia (2014) and Chiou and Youngs (2014), referred to as $ASB14$, $CB14$ and $CY14$, respectively. A simplified version of the Aristeidou et al. (2024) GMM was also included in the comparisons, which can be found on that study's GitHub repository. This version requires fewer input parameters, but comes at the expense of a slightly increased dispersion compared to the original model. As illustrated in Figure 10, the correlation coefficients obtained using $ASO24$ GMM are generally lower, yet they remain in strong overall agreement with those derived from the other GMMs. It is also evident that the simplified version of the Aristeidou et al. (2024) GMM exhibits behaviour similar to $ASB14$, $CB14$ and $CY14$, which can be attributed to differences in residuals computed from each underlying GMM. For the $ASO24$ GMM, residual differences are notably larger at smaller distances compared to the other

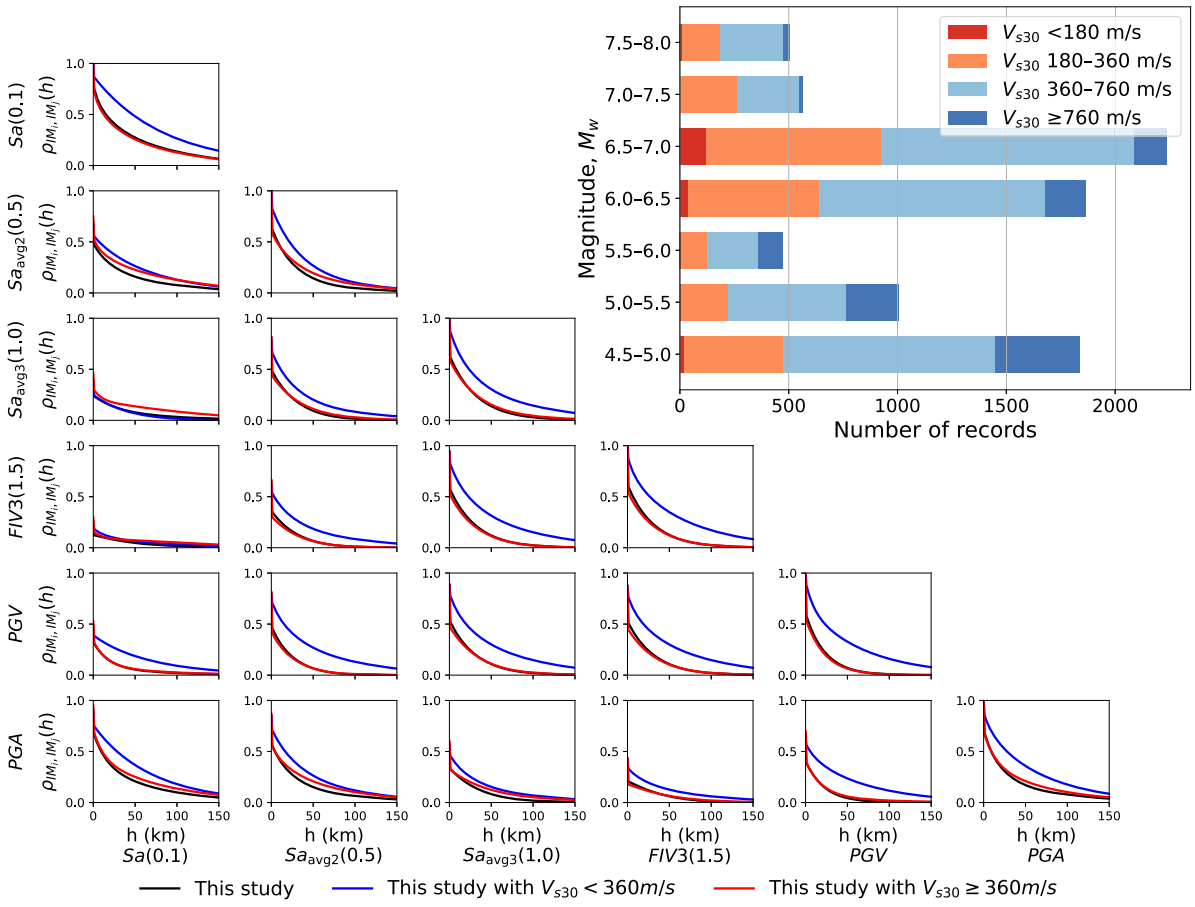

FIGURE 9 | Comparison of predicted correlations and cross-correlations for different IMs varying the level of V_{s30} (bottom), and the distribution of the number of records taking into consideration the M_w and different bins of V_{s30} (top).

models, which alters the shape of the empirical semivariogram, leading to faster decays from small distances. Consequently, when the correlation coefficients are calculated using Equation (14), the correlogram yields slightly different values. In short, although applying different GMMs can introduce some variation in correlation levels, the broader spatial correlation trends remain consistent.

6 | Application Example

To explicitly demonstrate the cross-IM capability of the model, multi-IM simulations were subsequently performed in which spatial and cross-IM correlations were considered simultaneously. For these simulations, the global PCA model was adopted to ensure a consistent latent-variable structure across multiple IMs. While the pairwise PCA formulation provides improved accuracy for individual IM pairs, it does not directly define a consistent latent-variable structure for more than two IMs, consequently when multiple pairwise models are assembled, the resulting covariance matrix is not guaranteed to remain positive semi-definite. Therefore, for applications requiring simultaneous simulation of multiple IMs, the global model is used to ensure internal consistency of the simulated fields. Because between-event correlation models are currently available only for $Sa(T)$, the between-event cross-IM correlation was assumed equal to the corresponding within-event correlation at zero separation distance ($h = 0$), consistent with the common assumptions in prior literature (e.g., Goda and Hong 2008). This enables a coherent joint simulation of all four IMs while remaining consistent with existing empirical knowledge. The resulting cross-correlation fields are shown in Figure 11. In contrast to the independently simulated fields, these joint simulations preserve the spatial co-variability between IMs, such that regions of elevated motion in one IM are systematically associated with elevated levels in the others according to the prescribed cross-correlation structure. This coherence is particularly relevant for applications involving vector-valued IMs or loss estimation frameworks where multiple IMs jointly influence structural response.

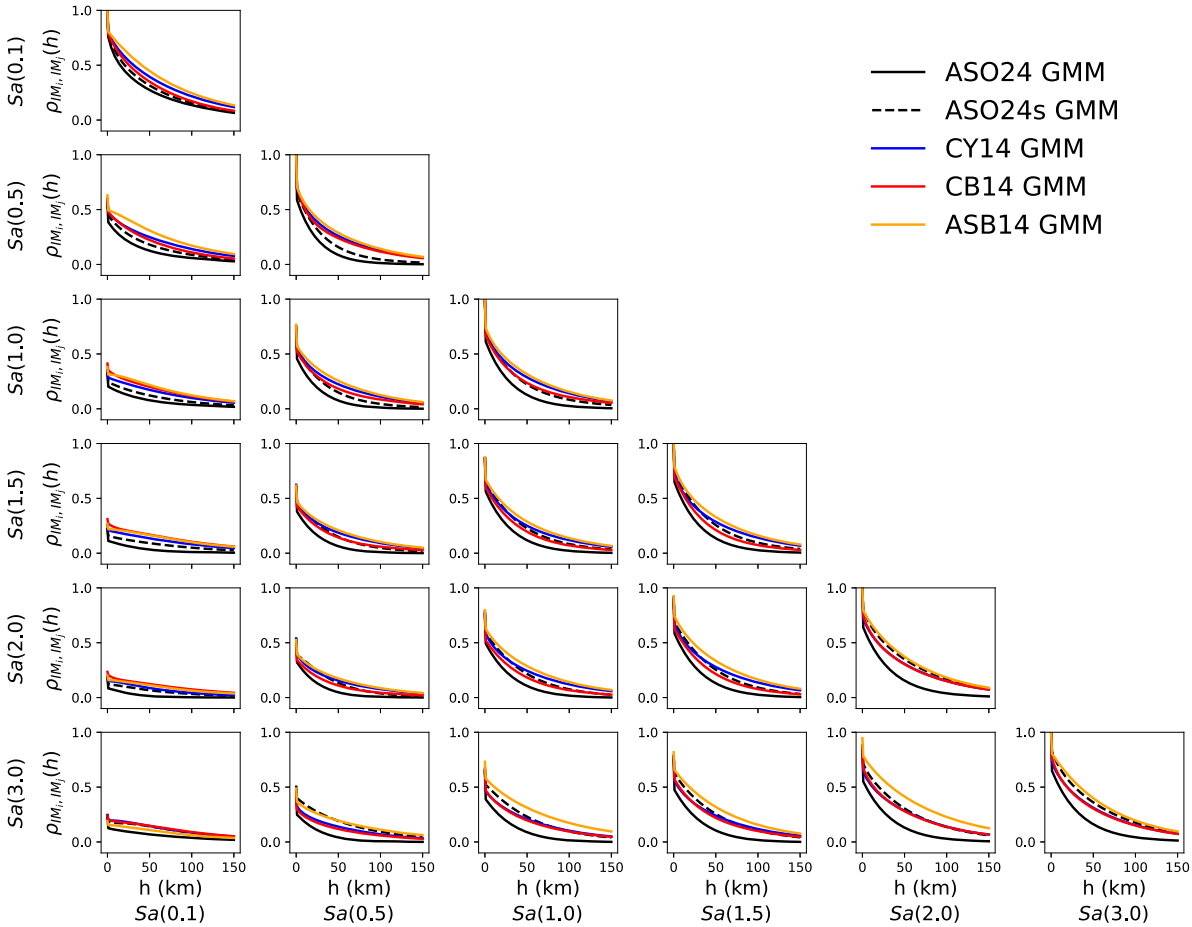

FIGURE 10 | Comparison of predicted correlations and cross-correlations for different IMs and GMMs. ASO24: Aristeidou et al. (2024), ASO24s: simplified version of Aristeidou et al. (2024), CY14: Chiou and Youngs (2014), CB14: Campbell and Bozorgnia (2014), and ASB14: Akkar et al. (2014a).

This application illustrates the practical consequences of incorporating cross-IM spatial correlation in multi-IM simulations. When cross-IM dependence is neglected, spatial patterns of different IMs are implicitly treated as independent, which alters their joint variability structure and may bias joint exceedance probabilities in regional risk assessments. By explicitly modelling both spatial and cross-IM correlation, the proposed framework enables coherent multi-IM simulation within a single probabilistic model.

To provide a quantitative illustration of the implications of cross-IM correlation, a simplified portfolio-level exceedance analysis was performed based on joint IM threshold exceedance. In this example, each spatial location in the GMF was interpreted as an asset and classified as *affected* when all IM thresholds were jointly exceeded. The IMs chosen here were $Sa(1.0)$, PGV and $FIV3(0.5)$, and the thresholds utilised were 0.1g, 10 cm/s and 15 cm/s, respectively. The selected thresholds were deterministic and illustrative, and do not correspond to a specific damage state. Instead, they provide a convenient proxy to examine how joint IM variability propagates to portfolio-scale metrics and could be investigated in more specific detail in future studies for different regions and structural typologies.

Here, the number of affected assets n jointly exceeding these thresholds was computed and expressed as a fraction of the total number of assets considered, which was $N = 1107$. These joint IM exceedances were computed for 1000 GMF, from which the empirical exceedance probability was derived, i.e., $P(IM_1 > im_1, IM_2 > im_2, IM_3 > im_3)$. To simulate the GMF, three modelling strategies were considered. First, each IM was simulated independently, neglecting cross-IM correlation. Second, simulations were performed using the global PCA model. Third, the pairwise PCA model was also extended to multi-IM case. In this latter approach, a full correlation matrix was assembled from pairwise models and subsequently projected onto the nearest positive semi-definite matrix following Higham (2002), as detailed in a Supplemental Material. Figure 12 presents the results for the three-IM combination $Sa(1.0) - PGV - FIV3(0.5)$.

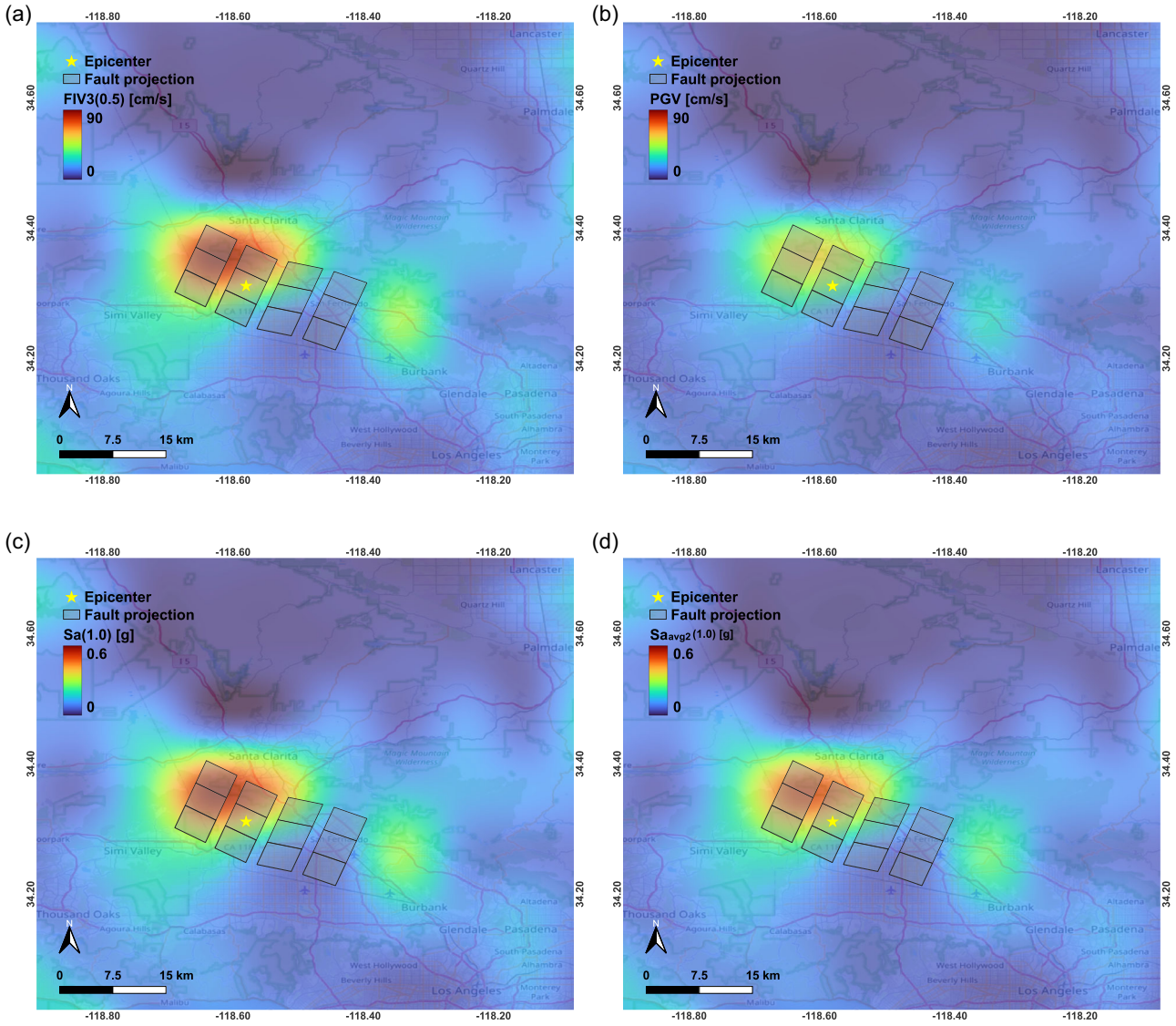

FIGURE 11 | Unconditional cross-IM GMFs for the 1994 Northridge earthquake using the spatial correlation model developed in this study for (a) $FIV3(0.5)$, (b) PGV, (c) $Sa(1.0)$, and (d) $Sa_{avg2}(1.0)$.

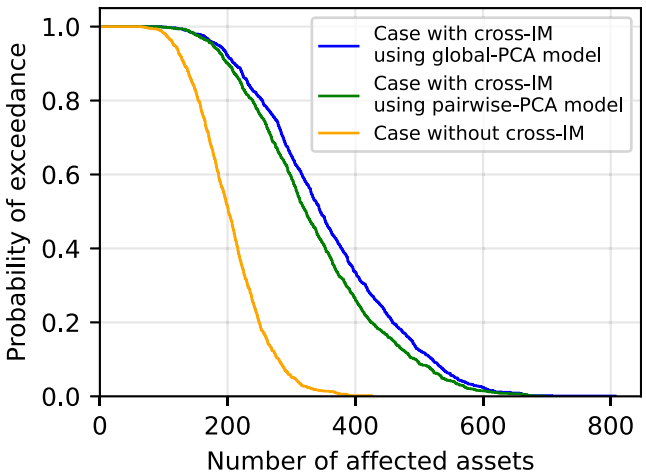

FIGURE 12 | Empirical exceedance probability of the number of affected assets under a joint IM exceedance criterion, comparing simulations with and without cross-IM spatial correlation for $Sa(1.0) - PGV - FIV3(0.5)$.

The results indicate that neglecting cross-IM correlation leads to a systematic underestimation of the number of affected assets. For the combination of IM presented in Figure 12, accounting for cross-IM has a substantial impact on exceedance probabilities. For example, the probability of exceeding 300 affected assets increases from less than 10% to approximately 60%. Additionally, the pairwise and global PCA models yield broadly consistent results, with only moderate differences that are likely attributable to differences in the construction of the correlation matrix in each approach. This figure is purely illustrative and would need more specific case studies and details to give more accurate and representative impact, but the goal here was to show that these differences are not negligible. Importantly, these effects arise despite identical marginal IM distributions, demonstrating that cross-IM correlation primarily governs joint IM behaviour. Overall, these results highlight that cross-IM spatial correlation not only ensures physically consistent multi-IM GMFs but also influences estimates of concurrent exceedance across a portfolio of regionally distributed assets. Also notable with the advent of such cross-IM GMFs is the ability to utilise enable vector IM-based fragility functions (e.g., Ebrahimian et al. 2015; Kohrangi et al. 2016; Du and Padgett 2021) to be utilised, which have shown much promise in past research but seen limited implementation due to practical limitations.

7 | Summary and Conclusions

This article presents a new spatial cross-correlation model for several IMs, developed using PCA and geostatistic tools, tailored for active shallow crustal tectonic regions. The model is based on over 8000 ground motion recordings from two combined databases, NGA-West2 and ESM. It provides spatial coefficients for traditional IMs such as $Sa(T)$, PGA and PGV, as well as next-generation IMs like $Sa_{avg}(T)$ and $FIV3(T)$.

The framework follows the methodology proposed by Markhvida et al. (2018), using PCA to transform spatially correlated normalised within-event residuals into uncorrelated PCs, as described in Equation (20). A key distinction of this study lies in the formulation of pairwise PCA models, in which each IM pair is modelled independently using a low-dimensional representation tailored to that specific combination. This approach leads to a considerable reduction in fitting error and improves the agreement with empirical semivariograms and cross-semivariograms for individual IM pairs. It is important to note, however, that in the pairwise formulation the number of PCs equals the number of IMs (i.e., two), and therefore no dimensionality reduction is achieved. The computational advantage observed for two IMs arises from the factorisation of smaller covariance matrices rather than from a reduction is system dimensionality. By contrast, global PCA formulations involving multiple IMs can achieve true dimensionality reduction and provide a unified latent-variable structure for joint simulation.

While each pairwise formulation produces a valid covariance structure in two dimensions, the pairwise models are calibrated independently and do not define a unique joint latent-variable representation for more than two IMs. As a result, assembling multiple pairwise models into a higher dimensional covariance matrix does not formally guarantee positive semi-definiteness and may lead to inconsistencies in multi-IM simulations. Therefore, the pairwise approach should be implemented primarily as a modelling tool for accurately representing pairwise spatial dependence. For applications requiring the joint simulation of multiple IMs, the global PCA model is recommended, as it ensures internal consistency of the covariance structure by construction. If the pairwise models are extended to multi-IM applications via correlation reconstruction, additional steps such as projection onto the nearest positive semi-definite matrix (Higham 2002) are required (see Supplemental Material), and the resulting framework departs from a unified PCA-based latent-variable formulation.

To validate the proposed model, comparisons were made with existing spatial cross-correlation models, showing strong agreement for common IMs and filling a gap for next-generation IMs such as $Sa_{avg}(T)$ and $FIV3(T)$, for which no prior models currently exist. A sensitive analysis was conducted using clustered datasets based on M_w, V_{s30} and different GMMs. The results indicated that the spatial correlation model is relatively insensitive to M_w clustering and different GMM usage, which suggests that the proposed model can be considered broadly independent of magnitude and GMM, making it suitable for global application in active shallow crustal settings. However, clustering by V_{s30} revealed significant variation and, therefore, should be considered carefully in applications. Specifically, for $Sa(T)$ and $Sa_{avg}(T)$, lower correlation values were observed for the low-V_{s30} subset at short periods, while the opposite trend occurred at longer periods. For $FIV3(T)$, higher spatial correlation values were consistently found for the low-V_{s30} dataset, regardless of period. Finally, an illustrative example demonstrated the application of the proposed model to generate jointly simulated GMFs for $Sa(1.0)$, $Sa_{avg2}(1.0)$, PGV and $FIV3(0.5)$. In this context, the global PCA model was used to ensure a consistent multi-IM simulation framework. The example highlighted the importance of incorporating cross-IM spatial correlation to preserve the joint variability structure of multiple IMs. In particular, neglecting cross-IM correlation was shown to substantially underestimate joint exceedance probabilities, demonstrating that cross-IM correlation can have a direct influence on portfolio-level exceedance metrics, even when marginal IM distributions remain the same. Overall, the findings

of this study indicate that the pairwise PCA formulation improves the modelling accuracy of individual IM pairs, whereas the global PCA formulation should be preferred when consistent multi-IM simulation is required.

While the model assumed stationarity and isotropy, it remains applicable for regional-scale assessments and offers a simpler alternative to more complex frameworks that explicitly incorporate path and site effects (e.g., Bodenmann et al. 2023). By addressing a critical gap in current practice (i.e., the lack of reliable spatial correlation structures for next-generation IMs), the proposed model enables more accurate assessments of seismic risk at regional scales. This advancement facilitates the adoption of modern ground-motion metrics into large-scale hazard and loss frameworks, which was the primary motivation of this research.

Acknowledgments

The work presented in this paper has been developed within the framework of the project "Dipartimenti di Eccellenza 2023–2027," funded by the Italian Ministry of Education, University and Research at IUSS Pavia. The authors are thankful for the comments and suggestions from Jack Baker and one other anonymous reviewer that helped improve this work.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The relevant files and functions that can be used to get the spatial and cross-spatial correlation coefficients presented in this study are available on GitHub at: https://github.com/vitorazevedomonteiro/cross-spatial-correlation-model.git.

References

Abbasnejadfard, M., M. Bastami, and A. Fallah. 2020. "Investigation of Anisotropic Spatial Correlations of Intra-Event Residuals of Multiple Earthquake Intensity Measures Using Latent Dimensions Method." *Geophysical Journal International* 222: 1449–1469. https://doi.org/10.1093/gji/ggaa255. SSN 0956-540X.

Acevedo, A. B., S. Carrascal, J. F. Betancur, and D. González. 2025. "Influence of the Spatial Distribution of Hazard and Exposure Models on Urban Seismic Risk Assessment." *Earthquake Spectra* 41: 3074–3093. https://doi.org/10.1177/87552930251341362. ISSN 8755-2930.

Akkar, S., M. A. Sandıkkaya, and J. J. Bommer. 2014a. "Empirical Ground-Motion Models for Point- and Extended-Source Crustal Earthquake Scenarios in Europe and the middle east." *Bulletin of Earthquake Engineering* 12: 359–387. https://doi.org/10.1007/s10518-013-9461-4. ISSN 1570-761X.

Akkar, S., M. A. Sandıkkaya, M. Şenyurt, et al. 2014b. "Reference Database for Seismic Ground-Motion in Europe (RESORCE)." *Bulletin of Earthquake Engineering* 12: 311–339. https://doi.org/10.1007/s10518-013-9506-8. ISSN 1570-761X.

Aldea, S., P. Heresi, and C. Pastén. 2022. "Within-Event Spatial Correlation of Peak Ground Acceleration and Spectral Pseudo-Acceleration Ordinates in the chilean Subduction Zone." *Earthquake Engineering & Structural Dynamics* 51: 2575–2590. https://doi.org/10.1002/eqe.3674. ISSN 0098-8847.

Ancheta, T. D., R. B. Darragh, J. P. Stewart, et al. 2014. "Nga-west2 Database." *Earthquake Spectra* 30: 989–1005. https://doi.org/10.1193/070913EQS197M. ISSN 8755-2930.

Aoi, S., T. Kunugi, and H. Fujiwara. 2004. "Strong-Motion Seismograph Network Operated by Nied: K-Net and Kik-Net." *Journal of JAEE* 4: 65–74. https://doi.org/10.5610/jaee.4.3_65. ISSN 1884-6246.

Aristeidou, S., D. Shahnazaryan, and G. J. O'Reilly. 2024. "Artificial Neural Network-Based Ground Motion Model for Next-Generation Seismic Intensity Measures." *Soil Dynamics and Earthquake Engineering* 184: 108851. https://doi.org/10.1016/j.soildyn.2024.108851. ISSN 02677261.

Aristeidou, S., D. Shahnazaryan, and G. J. O'Reilly. 2025. "Correlation Models for Next-Generation Amplitude and Cumulative Intensity Measures Using Artificial Neural Networks." *Earthquake Spectra* 41: 851–875. https://doi.org/10.1177/87552930241270563. ISSN 8755-2930.

Azarbakht, A., M. Mousavi, M. Nourizadeh, and M. Shahri. 2014. "Dependence of Correlations between Spectral Accelerations at Multiple Periods on Magnitude and Distance." *Earthquake Engineering & Structural Dynamics* 43: 1193–1204. https://doi.org/10.1002/eqe.2393. ISSN 0098-8847.

Baker, J. W., and B. A. Bradley. 2017. "Intensity Measure Correlations Observed in the Nga-west2 Database, and Dependence of Correlations on Rupture and Site Parameters." *Earthquake Spectra* 33: 145–156. https://doi.org/10.1193/060716eqs095m. ISSN 8755-2930.

Baker, J. W., and Y. Chen. 2020. "Ground Motion Spatial Correlation Fitting Methods and Estimation Uncertainty." *Earthquake Engineering & Structural Dynamics* 49: 1662–1681. https://doi.org/10.1002/eqe.3322. ISSN 0098-8847.

Baker, J., B. Bradley, and P. Stafford. 2021. *Seismic Hazard and Risk Analysis*. Cambridge University Press. https://doi.org/10.1017/9781108425056. ISBN 9781108348157.

Bantis, J., P. Heresi, A. Poulos, and E. Miranda. 2025. "Framework for Regional Seismic Risk Assessments of Groups of Tall Buildings." *Earthquake Engineering & Structural Dynamics* 54: 833–850. https://doi.org/10.1002/eqe.4283. ISSN 0098-8847.

Bodenmann, L., J. W. Baker, and B. Stojadinović. 2023. "Accounting for Path and Site Effects in Spatial Ground-Motion Correlation Models Using Bayesian Inference." *Natural Hazards and Earth System Sciences* 23: 2387–2402. https://doi.org/10.5194/nhess-23-2387-2023. ISSN 1684-9981.

Campbell, K. W., and Y. Bozorgnia. 2014. "Nga-west2 Ground Motion Model for the Average Horizontal Components of Pga, Pgv, and 5." *Earthquake Spectra* 30: 1087–1115. https://doi.org/10.1193/062913EQS175M. ISSN 8755-2930.

Chiou, B. S.-J., and R. R. Youngs. 2014. "Update of the Chiou and Youngs Nga Model for the Average Horizontal Component of Peak Ground Motion and Response *Spectra*." *Earthquake Spectra* 30: 1117–1153. https://doi.org/10.1193/072813EQS219M. ISSN 8755-2930.

Dávalos, H., and E. Miranda. 2019. "Filtered Incremental Velocity: A Novel Approach in Intensity Measures for Seismic Collapse Estimation." *Earthquake Engineering & Structural Dynamics* 48: 1384–1405. https://doi.org/10.1002/eqe.3205. ISSN 0098-8847.

Du, A., and J. E. Padgett. 2021. "Multivariate Return Period-Based Ground Motion Selection for Improved Hazard Consistency over a Vector of Intensity Measures." *Earthquake Engineering & Structural Dynamics* 50: 415–435. https://doi.org/10.1002/eqe.3338. ISSN 0098-8847.

Du, W., and C.-L. Ning. 2021. "Modeling Spatial Cross-Correlation of Multiple Ground Motion Intensity Measures (sas, Pga, Pgv, Ia, Cav, and Significant Durations) Based on Principal Component and Geostatistical Analyses." *Earthquake Spectra* 37: 486–504. https://doi.org/10.1177/8755293020952442. ISSN 8755-2930.

Eads, L., E. Miranda, and D. G. Lignos. 2015. "Average Spectral Acceleration as an Intensity Measure for Collapse Risk Assessment." *Earthquake Engineering & Structural Dynamics* 44: 2057–2073. https://doi.org/10.1002/eqe.2575. ISSN 0098-8847.

Ebrahimian, H., F. Jalayer, A. Lucchini, F. Mollaioli, and G. Manfredi. 2015. "Preliminary Ranking of Alternative Scalar and Vector Intensity Measures of Ground Shaking." *Bulletin of Earthquake Engineering* 13: 2805–2840. https://doi.org/10.1007/s10518-015-9755-9. ISSN 1570-761X.

Esposito, S., and I. Iervolino. 2011. "Pga and Pgv Spatial Correlation Models Based on European Multievent Datasets." *Bulletin of the Seismological Society of America* 101: 2532–2541. https://doi.org/10.1785/0120110117. ISSN 0037-1106.

Goda, K., and G. M. Atkinson. 2009. "Probabilistic Characterization of Spatially Correlated Response Spectra for Earthquakes in japan." *Bulletin of the Seismological Society of America* 99: 3003–3020. https://doi.org/10.1785/0120090007. ISSN 0037-1106.

Goda, K., and H. P. Hong. 2008. "Spatial Correlation of Peak Ground Motions and Response Spectra." *Bulletin of the Seismological Society of America* 98: 354–365. https://doi.org/10.1785/0120070078. ISSN 0037-1106.

Heresi, P., and E. Miranda. 2021. "Intensity Measures for Regional Seismic Risk Assessment of Low-Rise Wood-Frame Residential Construction." *Journal of Structural Engineering* 147, no. 1: 04020287. https://doi.org/10.1061/(ASCE)ST.1943-541X.0002859. ISSN 0733-9445.

Heresi, P., and E. Miranda. 2023. "Rpbee: Performance-Based Earthquake Engineering on a Regional Scale." *Earthquake Spectra* 39: 1328–1351. https://doi.org/10.1177/87552930231179491. ISSN 8755-2930.

Higham, N. J. 2002. "Computing the Nearest Correlation Matrix–a Problem from Finance." *IMA Journal of Numerical Analysis* 22: 329–343. https://doi.org/10.1093/imanum/22.3.329. ISSN 0272-4979.

Jayaram, N., and J. W. Baker. 2008. "Statistical Tests of the Joint Distribution of Spectral Acceleration Values." *Bulletin of the Seismological Society of America* 98: 2231–2243. https://doi.org/10.1785/0120070208. ISSN 00371106.

Jayaram, N., and J. W. Baker. 2009. "Correlation Model for Spatially Distributed Ground-Motion Intensities." *Earthquake Engineering & Structural Dynamics* 38: 1687–1708. https://doi.org/10.1002/eqe.922. ISSN 0098-8847.

Kazantzi, A. K., and D. Vamvatsikos. 2015. "Intensity Measure Selection for Vulnerability Studies of Building Classes." *Earthquake Engineering & Structural Dynamics* 44: 2677–2694. https://doi.org/10.1002/eqe.2603. ISSN 0098-8847.

Kohrangi, M., P. Bazzurro, and D. Vamvatsikos. 2016. "Vector and Scalar Ims in Structural Response Estimation, Part Ii: Building Demand Assessment." *Earthquake Spectra* 32: 1525–1543. https://doi.org/10.1193/053115EQS081M. ISSN 8755-2930.

Kohrangi, M., P. Bazzurro, D. Vamvatsikos, and A. Spillatura. 2017. "Conditional Spectrum-Based Ground Motion Record Selection Using Average Spectral Acceleration." *Earthquake Engineering & Structural Dynamics* 46, no. 10: 1667–1685. https://doi.org/10.1002/eqe.2876. ISSN 00988847.

Lanzano, G., S. Sgobba, L. Luzi, et al. 2019. "The Pan-European Engineering Strong Motion (esm) Flatfile: Compilation Criteria and Data Statistics." *Bulletin of Earthquake Engineering* 17: 561–582. https://doi.org/10.1007/s10518-018-0480-z. ISSN 1570-761X.

Loth, C., and J. W. Baker. 2013. "A Spatial Cross-Correlation Model of Spectral Accelerations at Multiple Periods." *Earthquake Engineering & Structural Dynamics* 42: 397–417. https://doi.org/10.1002/eqe.2212. ISSN 0098-8847.

Luzi, L., S. Hailemikael, D. Bindi, F. Pacor, F. Mele, and F. Sabetta. 2008. "Itaca (italian Accelerometric Archive): A Web Portal for the Dissemination of Italian Strong-Motion Data." *Seismological Research Letters* 79: 716–072. https://doi.org/10.1785/gssrl.79.5.716. ISSN 0895-0695.

Markhvida, M., L. Ceferino, and J. W. Baker. 2018. "Modeling Spatially Correlated Spectral Accelerations at Multiple Periods Using Principal Component Analysis and Geostatistics." *Earthquake Engineering & Structural Dynamics* 47: 1107–1123. https://doi.org/10.1002/eqe.3007. ISSN 0098-8847.

Monteiro, V. A., and G. J. O'Reilly. 2026. "A Review of Ground Motion Correlation Modelling for Regional Seismic Risk Analysis." *Bulletin of Earthquake Engineering*. https://doi.org/10.1007/s10518-026-02377-0.

O'Reilly, G. J. 2021. "Limitations of Sa(t1) as an Intensity Measure when Assessing Non-Ductile Infilled Rc Frame Structures." *Bulletin of Earthquake Engineering* 19: 2389–2417. https://doi.org/10.1007/s10518-021-01071-7. ISSN 1570-761X.

Park, J., P. Bazzurro, and J. W. Baker. 2007. "Modeling Spatial Correlation of Ground Motion Intensity Measures for Regional Seismic Hazard and Portfolio Loss Estimation." In *Applications of 838 Statistics and Probability in Civil Engineering*. 1–8.

Rekoske, J. M., E. M. Thompson, M. P. Moschetti, M. G. Hearne, B. T. Aagaard, and G. A. Parker. 2020. "The 2019 ridgecrest, california, Earthquake Sequence Ground Motions: Processed Records and Derived Intensity Metrics." *Seismological Research Letters* 91: 2010–2023. https://doi.org/10.1785/0220190292. ISSN 0895-0695.

Shahnazaryan, D., and G. J. O'Reilly. 2024. "Next-Generation Non-Linear and Collapse Prediction Models for Short- to Long-Period Systems via Machine Learning Methods." *Engineering Structures* 306: 117801. https://doi.org/10.1016/j.engstruct.2024.117801. ISSN 01410296.

Silva, V., and N. Horspool. 2019. "Combining Usgs Shakemaps and the Openquake-Engine for Damage and Loss Assessment." *Earthquake Engineering and Structural Dynamics* 48: 634–652. https://doi.org/10.1002/eqe.3154. ISSN 10969845.

Wang, G., and W. Du. 2013. "Spatial Cross-Correlation Models for Vector Intensity Measures (pga, Ia, Pgv, and Sas) considering Regional Site Conditions." *Bulletin of the Seismological Society of America* 103: 3189–3204. https://doi.org/10.1785/0120130061. ISSN 00371106.

Wang, M., and T. Takada. 2005. "Macrospatial Correlation Model of Seismic Ground Motions," *Earthquake Spectra* 21: 1137–1156. https://doi.org/10.1193/1.2083887.

Weatherill, G., S. Esposito, I. Iervolino, P. Franchin, and F. Cavalieri. 2014. *Framework for Seismic Hazard Analysis of Spatially Distributed Systems*. Springer Netherlands. https://doi.org/10.1007/978-94-017-8835-9_3. ISBN 978-94-017-8835-9.

Supporting Information

Additional supporting information can be found online in the Supporting Information Section.