Bulletin of Earthquake Engineering
https://doi.org/10.1007/s10518-026-02377-0

REVIEW ARTICLE

# A review of ground motion correlation modelling for regional seismic risk analysis

**Vitor A. Monteiro**[1] · **Gerard J. O'Reilly**[1]

Received: 21 October 2025 / Accepted: 19 January 2026


**Abstract**
Inter and intra-site correlation of ground-motion intensity measures (IMs) plays a critical role in seismic hazard and risk assessment. Ignoring such correlations can lead to significant misrepresentation of losses in regional-scale studies and misrepresentation of ground motion field simulations. Accurate correlation modelling is essential for scenario-based risk assessments, emergency preparedness planning, and understanding systematic infrastructure vulnerabilities in portfolio risk analyses. This study presents a detailed overview of intra-site (non-spatial) and inter-site (spatial) correlation models developed over the past two decades. It reviews over 45 models proposed in the literature, encompassing diverse methodologies applied to different regional databases and a variety of IM. The analyses reveal considerable variability among models, particularly in short-range spatial correlation and in how inter-IM correlations are treated. Despite this diversity, most models rely on simplifying assumptions such as stationarity and isotropy, which may not fully capture the complexities of real-world ground motion patterns. This work provides a valuable resource for researchers and practitioners by summarising the current state of correlation modelling and offering guidance on model selection based on database, regional context, and engineering application. It underscores the importance of informed model choice for improving the accuracy of hazard and risk assessments in spatially distributed systems.

**Keywords** Seismic risk · Correlation · Spatial correlation · Regional assessment · Intensity measures

## 1 Introduction

In seismic risk assessment studies, the correlation of ground shaking plays a significant role, particularly in understanding how different intensity measures (IMs) relate to one another both at the same site location and spatially across several locations. These models have

✉ Gerard J. O'Reilly
gerard.oreilly@iusspavia.it

[1] Centre for Training and Research on Reduction of Seismic Risk (ROSE Centre), Scuola Universitaria Superiore IUSS di Pavia, Piazza della Vittoria 15, 27100 Pavia, Italy

Published online: 04 February 2026

several applications in seismic risk analyses. In ground motion selection procedures, such as the conditional spectrum (CS) method (Lin et al. 2013; Baker and Lee 2018) and the generalised conditional intensity measure (GCIM) method (Bradley 2010), correlations are key parameters for matching the statistical characteristics of target distributions. Additionally, spatial correlation models are critical for generating spatially coherent ground motion fields, which are essential for simulating earthquake shaking over broad geographic areas (e.g., Bradley 2014; Weatherill et al. 2014, 2015) and for performing loss estimation for spatially distributed assets (e.g., Park et al. 2007).

The study of correlations can be broadly classified into two categories: inter-site correlation (herein termed spatial correlation), which examines dependencies between the shaking intensity at different site locations; and intra-site correlation (herein termed non-spatial correlation), which focuses on relationships in ground shaking at a single site location. These spatial and non-spatial correlation models can be further distinguished according to the IMs they address, with models mapping a single IM referred to as *same IM* models herein, and *cross IM* models refer to when different IMs are mapped. Table 1 categorises these and lists the numbers of studies reviewed here based on whether they address intra- or inter-site correlations, as well as whether same or cross IMs.

This review provides a summary of existing research in correlation modelling using a wide range of methodologies and databases over the past two decades. Over 40 different studies were reviewed and are critically discussed herein. A detailed description of the database, methodology, IMs, and correlation types is summarised in a table hosted on GitHub and referenced in Sect. 7. The overall goal is to provide a background and relative comparison of the various models developed, their scope and possible limitations for seismic risk analysts looking to implement them in different contexts.

## 2 Background and structure

Before diving into the different studies that are reviewed in the following sections, several background aspects common to most studies are first presented. These relate to the IMs adopted, the ground motion databases utilised, the methods to model and compute correlation, and the techniques applied to fit and produce usable models for seismic engineering applications.

A general overview is illustrated in Fig. 1. The first distinguishing aspect is between spatial and non-spatial correlation, followed by whether the same or different IMs were investigated. Consequently, the main body is divided into three main sections, beginning with non-spatial correlation with different IMs in Sect. 3. This is followed by Sect. 4 that introduces spatial correlation modelling for the same IMs, followed by Sect. 5 that examines spatial correlation for different IMs. The redundant case of non-spatial correlation for the same IM is not discussed.

**Table 1** Number of intra- and inter-site models for same and cross IMs

| | Same IM | Cross IMs |
|---|---|---|
| Intra-site | - | 13 |
| Inter-site | 25 | 8 |

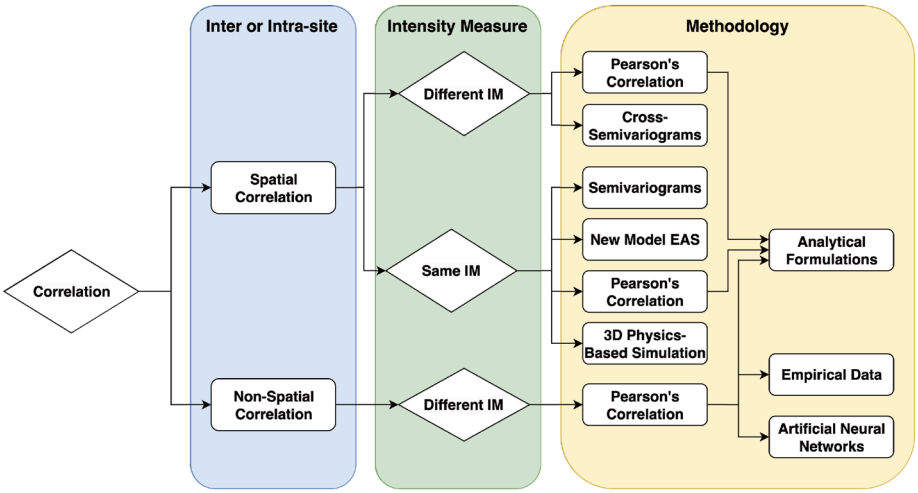


**Fig. 1** Overview of correlation modelling for seismic shaking

## 2.1 Intensity measures

The choice of IM when conducting probabilistic seismic hazard analysis (PSHA) or seismic fragility and vulnerability modelling is a pertinent aspect that depends on several criteria. These often relate to the specific nature of the structural systems analysed, the analysis approach used, and the broader goals of the study. The purpose of this article is not to explore these motivations in detail, but rather to note that a variety of IMs exist, each serving different purposes. These can be broadly categorised as peak amplitude or cumulative IMs. Examples of peak amplitude IMs include peak ground acceleration, $PGA$, peak ground velocity, $PGV$, peak ground displacement, $PGD$, spectral acceleration at different periods, $Sa(T)$, for example. Examples of cumulative IMs include Arias intensity, $Ia$, significant duration, $Ds_{575}$ and $Ds_{595}$, cumulative absolute velocity, $CAV$, acceleration spectrum intensity, $ASI$, spectrum intensity, $SI$, and displacement spectrum intensity, $DSI$. Another group of so-called next-generation IMs have been the focus of much research in recent years, offering several advantages when characterising risk in different settings. Examples include average spectral acceleration, $Sa_{avg}(T)$, and filtered incremental velocity, $FIV3(T)$, which will be discussed further below.

The Husid plot (Husid 1969) represents the cumulative $Ia$ normalised by the total $Ia$ and is defined as:

$$H(t) = \frac{1}{Ia} \int_0^t [a(\tau)]^2 \, d\tau \times 100\% \tag{1}$$

where $a(t)$ is the ground acceleration, and $Ia$ is given by:

$$Ia = \int_0^T [a(t)]^2 dt \tag{2}$$

The significant duration $Ds_{xy}$ is then defined as the interval between instants $t_x$ and $t_y$ at which the Husid plot reaches $x\%$ and $y\%$ of $Ia$, respectively:

$$Ds_{xy} = t_y - t_x \tag{3}$$

Average spectral acceleration, $Sa_{avg}(T)$, is computed as the geometric mean of $N$ spectral acceleration values within a range of periods:

$$\begin{aligned} Sa_{avg}(T) &= \left( \prod_{i=i}^{N} Sa(c_i T) \right)^{\frac{1}{N}} \\ \ln Sa_{avg}(T) &= \frac{1}{N} \sum_{i=1}^{N} \ln Sa(c_i T) \end{aligned} \tag{4}$$

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 filtered incremental velocity metric, $FIV3(T)$, proposed by Dávalos and Miranda (2019), has shown promising results regarding the efficiency and sufficiency in characterising the collapse performance of buildings. It captures the cumulative effect of ground motion pulses by integrating a filtered acceleration signal over a sliding time window. This IM can be briefly explained as:

$$FIV3(T) = max\{V_{s,max1} + V_{s,max2} + V_{s,max3}; |V_{s,min1} + V_{s,min2} + V_{s,min3}|\} \tag{5}$$

$$V_s(t) = \int_t^{t+\alpha T} \ddot{u}_{gf}(t)\, dt, \quad \forall t < t_{end} - \alpha T \tag{6}$$

where $V_s(t)$ represents a sequence of incremental velocity (IV) values computed over moving time windows of length of $\alpha T$. 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, implying it is a period-dependent IM, 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 cut-off frequency $f_c = \beta f$, where $f = 1/T$ and $\beta$ is a scaling factor.

When evaluating correlations between spectral acceleration, $Sa(T)$, an important consideration is the orientation of horizontal ground motion components. A common practice is to take the geometric mean of the two orthogonal station recordings, $Sa^{gm}(T)$. However, this is sensitive to the initial orientation in which the components were recorded. Boore (2010) introduced rotated measures such as $Sa^{RotD50}(T)$ and $Sa^{RotD100}(T)$. The $RotD50$ definition represents the median (50th percentile) across all non-redundant orientations of the horizontal ground motion, effectively capturing a typical directional response. In contrast, the $RotD100$ definition corresponds to the maximum $Sa(T)$ observed across all

orientations, representing a conservative, envelope-type estimate. Where relevant, the orientation of these IMs will be specified herein.

### 2.2 Ground motion databases

Another aspect is the database of ground motions utilised. These can range from the classical approach of using natural ground motions, albeit with some filtering and correction techniques applied, recorded from notable past earthquakes, or using simulated ground motion recordings obtained from physics-based simulations (PBS).

Several correlation models have been developed using different ground motion databases. Knowing which databases were considered can help engineers and researchers apply a model to a specific region. A table summarising the databases used for the correlation models reviewed in this study, along with additional relevant detail, is available on GitHub (see Sect. 7).

In recent years, the increasing availability of computational resources has significantly enhanced the capability of PBS to simulate earthquake ground shaking with a high degree of physical realism. These simulations, grounded in the numerical solution of the elastodynamic equations, offer a valuable alternative to recorded ground motion data, particularly in regions with sparse instrumentation or complex geology (e.g., Chen and Baker 2019; Infantino et al. 2021; Schiappapietra and Smerzini 2021; Lin and Smerzini 2022; Zolfaghari and Forghani 2024). A more detailed discussion is provided in Sect. 4.

Lastly, it is important to note that all correlation models discussed in this work pertain to mainshock events, where the issue of aftershocks has not been included in the review, but several studies have examined these (e.g., Zhu et al. 2017; Papadopoulos et al. 2019; Ming-Yang and Da-Gang 2024).

### 2.3 Ground motion models and residuals

Correlation modelling, both spatial and non-spatial, relies on the use of appropriate ground 140 motion models (GMMs). GMMs are essential to provide estimates of ground shaking intensity as a function of earthquake rupture characteristics, path effects, and local site conditions. They typically follow a lognormal distribution as a function of explanatory variables and a residual term, generally written as:

$$\ln IM_{i,k,m} = \mu_i(X_{k,m}, \theta) + \delta t_{i,k,m} \cdot \sigma_i \tag{7}$$

where, $\ln IM_{i,k,m}$ is the natural logarithm of $i^{th}$ IM for event $k$ and recording site $m$, $\mu_i(X_{k,m}, \theta)$ is the mean value from the GMM based on explanatory variables $X_{k,m}$ (e.g., moment magnitude, $M_w$, Joyner-Boore distance, $R_{jb}$, site conditions, $V_{s30}$, etc.), and model parameters, $\theta$. $\delta t_{i,k,m}$ and $\sigma_i$ are the normalised total residual and the total standard deviation in logarithmic space, respectively. To better separate sources of variability, modern GMMs often use mixed-effects regression models (Abrahamson and Youngs 1992), which decompose the total variability into between- and within-event components using $\delta t_{i,k,m} \cdot \sigma_i = \delta b_{i,k} \cdot \tau_i + \delta w_{i,k,m} \cdot \phi_i$, noting that the $m$ subscript is dropped from $\tau$ since it is invariant for all site locations between events. The form of such a model is written:

$$\begin{aligned} \ln IM_{i,k,m} &= \mu_i(X_{k,m},\theta) + \delta B_{i,k} + \delta W_{i,k,m} \\ &= \mu_i(X_{k,m},\theta) + \delta b_{i,k}\cdot\tau_i + \delta w_{i,k,m}\cdot\phi_i \end{aligned} \tag{8}$$

where $\delta b_{i,k}$ and $\delta w_{i,k,m}$ correspond to the normalised between-event and within-event residuals. $\tau_i$ and $\phi_i$ are the between-event and within-event logarithmic standard deviations, respectively.

Although this representation aligns with standard mixed-effects formulations, it is worth noting that the separation of the residual components becomes more nuanced when the GMM includes non-linear site-response terms. In such cases, the between-event variability interacts with the non-linear site amplification rather than passing through the model unchanged (e.g., Chiou and Youngs 2008, 2014; Atik and Abrahamson 2010). This interaction means that the within-event and between-event terms are not strictly additive, which in turn can influence how correlations should be interpreted or applied. From a practical perspective, users applying existing spatial correlation or non-spatial correlation models to GMMs with non-linear site terms should be aware that most correlation models were derived assuming linear site behaviour. As a result, some deviation in the effective within-event variability is expected, and performing basic sensitivity checks may be advisable when non-linear site effects are significant.

Correlation models are typically developed based on within-event residuals, comparing the observed ground GMM's prediction, rather than the absolute IM value recorded. Specifically, the correlation models analysed in this study are based on two types of residuals: the total residuals, denoted by $\delta T$, and the within-event residuals $\delta W$.

For non-spatial correlation models, the total normalised residuals are typically used, as given in Eq. (9), where the total standard deviation $\sigma_i$ is decomposed into $\tau_i$ and $\phi_i$ components according to: $\sigma_i = \sqrt{\tau_i^2 + \phi_i^2}$:

$$\delta t_{i,k,m} = \frac{\ln IM_{i,k,m} - \mu_i(X_{k,m},\theta)}{\sqrt{\tau_i^2 + \phi_i^2}} \tag{9}$$

In contrast, spatial correlation models generally focus on the spatial pattern of variability *within* the same seismic event. Therefore, they utilise normalised within-event residuals and are computed as:

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

Just to note, spatial correlation models are typically developed by first estimating the residuals separately for each earthquake and then combining the results across all events to obtain a generalised model. For each individual event, the between-event term ($\delta b_{i,k}\cdot\tau_i$) is often assumed constant across sites, which is a reasonable approximation for conventional, homoscedastic GMMs. As a result, existing inter-IM spatial correlation models generally rely on within-event residuals. While this approach works well in most cases, it may not fully capture spatial variability when heteroscedastic effects or, as mentioned above, non-linear site responses are significant. Exploring approaches that incorporate total residuals, accounting for both between-event and within-event variability, could be an avenue for future research, particularly in the context of cross-IM correlations.

### 2.4 Methodologies

The previous section outlined the basis for utilising GMMs and computing residual terms. However, with these datasets of residuals for different IMs and event sets etc., several methodologies can be employed to model both spatial and non-spatial correlations, as illustrated in Fig. 1.

It is important to note that the majority of existing spatial and non-spatial correlation models reviewed in this study, whether developed using analytical formulations, empirical fitting, or artificial neural networks (ANNs), are embedded within multi-stage estimation frameworks. In such approaches GMM parameters are first estimated, followed by separate estimation of correlation structures using the resulting residuals. While widely adopted in practice, multi-stage algorithms may suffer from statistical inefficiency and potential inconsistency, particularly when correlation parameters interact with the GMM structure. Recent studies have shown that one-stage estimation algorithms, which jointly estimate GMM and correlation parameters, can provide improved statistical properties, albeit at the cost of increased numerical complexity. For completeness, the reader is referred to Ming et al. (2019) for a detailed discussion of one-stage estimation frameworks and their implications for correlation modelling.

#### 2.4.1 Pearson's correlation coefficient

Pearson's correlation coefficient (Ang and Tang 2007) is the most widely used formulation for computing correlations. Its basic form is shown in Eq. (11), where $X$ and $Y$ represent the random variables of interest (e.g., two different IMs), while $X_i$ and $Y_i$ are individual observations from datasets of size $n$. The $\mu$ terms denote the sample mean. Once the correlation coefficient is determined, analytical formulations are typically used (e.g., Boore et al. 2003; Baker and Cornell 2006; Park et al. 2007; Baker and Jayaram 2008; Goda and Hong 2008; Goda and Atkinson 2009; Sokolov et al. 2010; Bradley 2011a, 2011b, 2012; Cimellaro 2013; Heresi and Miranda 2019), and empirical studies have also been explored (e.g., Bradley 2011a; Akkar et al. 2014; Baker and Bradley 2017; Heresi and Miranda 2021; Tarbali et al. 2023). More recently, techniques such as artificial neural networks (e.g., Aristeidou et al. 2024) have emerged as an alternative.

$$\rho_{X,Y} = \frac{\sum_{i=1}^{n}(X_i - \mu_X)(Y_i - \mu_Y)}{\sqrt{\sum_{i=1}^{n}(X_i - \mu_X)^2(Y_i - \mu_Y)^2}} \tag{11}$$

#### 2.4.2 Semivariograms

As far as spatial correlation is concerned, semivariograms are commonly used in geostatistics. Here, random variables distributed over space and exhibiting spatial continuity are represented by a random function $Z(x)$, where $x$ denotes a spatial position. For an univariate random field $Z(x)$, the covariance function between two spatial locations $x$ and $x + h$ is defined as:

$$C(h) = \mathrm{Cov}\left(Z(x), Z(x+h)\right) = \mathbb{E}\left[\left(Z(x) - \mathbb{E}\left[Z(x)\right]\right)\left(Z(x+h) - \mathbb{E}\left[Z(x+h)\right]\right)\right] \tag{12}$$

where for a given earthquake $Z(x)$ and $Z(x+h)$ are normalised residuals at sites separated by a distance $h$. Due to sparse spatial coverage, irregular station spacing, and insufficient sample size to reliably estimate location-dependent means and covariances a common assumption in modelling spatial fields is second-order stationarity, which implies a constant mean and covariance that depend only on the separation vector $h$, not on the absolute location $x$ and $x+h$. Furthermore, assuming isotropy means that spatial dependence is a function of absolute distance only (i.e., $h = \|h\|$) and the mean becomes $\mu_Z = \mathbb{E}[Z(x)] = \mathbb{E}[Z(x+h)]$, simplifying the covariance function to:

$$C(h) = \mathrm{Cov}\left(Z(x), Z(x+h)\right) = \mathbb{E}\left[\left(Z(x) - \mu_Z\right)\left(Z(x+h) - \mu_Z\right)\right] \tag{13}$$

Before adopting assumptions such as second-order stationarity or isotropy, several statistical tests are available to evaluate whether these properties are supported by the data. Both parametric and non-parametric approaches exist for this purpose. For example, Weller and Hoeting (2016) provide a comprehensive review of parametric and non-parametric tests for detecting anisotropy in spatial datasets. Additionally, tools implemented in the spTest package in R (Weller 2018) similarly allow testing isotropy and stationarity through semivariogram-based contrasts and resampling procedures. Such diagnostic tests can offer preliminary insight into the spatial structure of the data and help guide the selection of an appropriate correlation model or identify cases where more flexible, non-stationary formulations may be necessary.

To quantify the dissimilarity between random variables at different locations, the semivariogram $\gamma(h)$ is defined as the expected ($\mathbb{E}$) squared difference between values of the random function at locations separated by a distance $h$ (Gooverts 1997). It is particularly useful in situations with limited repeated observations at a single location, which is common in earthquake ground motion studies. It is written as:

$$\begin{aligned}
\gamma(h) &= \frac{1}{2}\mathbb{E}\left[\left(Z(x) - Z(x+h)\right)^2\right] \\
&= \frac{1}{2}\mathbb{E}\left[Z(x)^2 - 2\cdot Z(x)\cdot Z(x+h) + Z(x+h)^2\right] \\
&= \frac{1}{2}\left(\mathbb{E}\left[Z(x)^2\right] - 2\cdot\mathbb{E}\left[Z(x)\cdot Z(x+h)\right] + \mathbb{E}\left[Z(x+h)^2\right]\right)
\end{aligned} \tag{14}$$

From the second-order stationarity assumption, one obtains:

$$\begin{aligned}
\mathbb{E}\left[Z(x)^2\right] &= \mathrm{Var}\left[Z(x)\right] + \mu^2 = C(0) + \mu^2 \\
\mathbb{E}\left[Z(x+h)^2\right] &= C(0) + \mu^2 \\
\mathbb{E}\left[Z(x)\cdot Z(x+h)\right] &= C(h) + \mu^2
\end{aligned} \tag{15}$$

and substituting into the expression for $\gamma(h)$:

$$\begin{aligned}\gamma(h) &= \frac{1}{2}\left(\left(C(0)+\mu^2\right) - 2\left(C(h)+\mu^2\right) + \left(C(0)+\mu^2\right)\right) \\ &= \frac{1}{2}\left(2C(0) + 2\mu^2 - 2C(h) - 2\mu^2\right) \\ &= \frac{1}{2}\cdot 2\left(C(0) - C(h)\right) \\ &= C(0) - C(h)\end{aligned} \tag{16}$$

Hence, the covariance function $C(h)$ can be expressed in terms of the semivariogram $\gamma(h)$ as:

$$C(h) = C(0) - \gamma(h) \tag{17}$$

where $\mathrm{Cov}(\cdot)$ denotes covariance, $\mathrm{Var}(\cdot)$ denotes variance, and $C(0)$ is the covariance at zero distance and represents the variance of the random field. Still under the second-order stationarity assumption, the correlation coefficient, often referred to as a correlogram denoted by $\rho(h)$, describes how correlation decays with increasing separation. It is defined as:

$$\begin{aligned}\rho(h) =& \frac{\mathrm{Cov}\left(Z(x), Z(x+h)\right)}{\sqrt{\mathrm{Var}\left[Z(x)\right]\cdot \mathrm{Var}\left[Z(x+h)\right]}} \\ =& \frac{C(h)}{C(0)} \\ =& \frac{C(0)-\gamma(h)}{C(0)} \\ =& 1 - \frac{\gamma(h)}{C(0)}\end{aligned} \tag{18}$$

The correlation depends solely on the distance rather than on specific locations. Consequently, two different pairs of stations distributed in three different locations that are equidistant are assumed to exhibit the same correlation value, regardless of their absolute positions. Recent research by Bodenmann et al. (2023) has developed this concept to incorporate further details beyond these simplifying assumptions, which will be discussed in Sect. 4.3.

The stationary semivariogram can be empirically estimated from earthquake ground motion data using the following functional form:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} \left[z(x_\alpha + h) - z(x_\alpha)\right]^2 \tag{19}$$

where $N(h)$ is the number of observation pairs separated by the distance $||h||$ (often computed using the haversine formula for geographic coordinates) and $\{z(x_\alpha), z(x_\alpha + h)\}$ denotes the $\alpha^{th}$ pair. A valid semivariogram must be conditionally negative definite. That is, for any finite set of locations $x_1, \ldots, x_m$, and any set of real weights $w_1, \ldots, w_m$ such that $\sum_{i=1}^{m} w_i = 0$, the following condition must hold (Cressie 1993):

$$\sum_{i=1}^{m}\sum_{j=1}^{m} w_i w_j \gamma(x_i - x_j) \leq 0 \tag{20}$$

This condition ensures that the semivariogram leads to a positive semi-definite covariance matrix, thereby preserving non-negative variances and conditional variances. To satisfy this requirement, semivariogram models are typically constructed as linear combinations of basic functions such as exponential, Gaussian, spherical, and nugget effect models, represented by Eqs. (21), (22), (23), and (24), respectively.

$$\gamma(h) = a\left[1 - \exp\left(\frac{-3h}{b}\right)\right] \tag{21}$$

$$\gamma(h) = a\left[1 - \exp\left(\frac{-3h^2}{b^2}\right)\right] \tag{22}$$

$$\gamma(h) = \begin{cases} a\left[\frac{3}{2}\frac{h}{b} - \frac{1}{2}\left(\frac{h}{b}\right)^3\right], & 0 \leq h \leq b \\ a, & h > b \end{cases} \tag{23}$$

$$\gamma(h) = \begin{cases} 0, & \text{if h} = 0 \\ a, & \text{if h} > 0 \end{cases} \tag{24}$$

In these models, two key parameters define their shape. The sill, $a$, is the asymptotic value the semivariogram approaches as the distance between data points increases. In practical terms, it represents the total variance in the system, where the semivariogram value becomes constant beyond a certain distance. The range, $b$, is the distance beyond which the semivariogram reaches the sill value. It represents the distance at which the spatial correlation between data points becomes almost negligible.

Figure 2 shows the common exponential model fitting (Eq. (21)) of the semivariogram and the visual representation of the sill and range. Semivariograms have been applied to

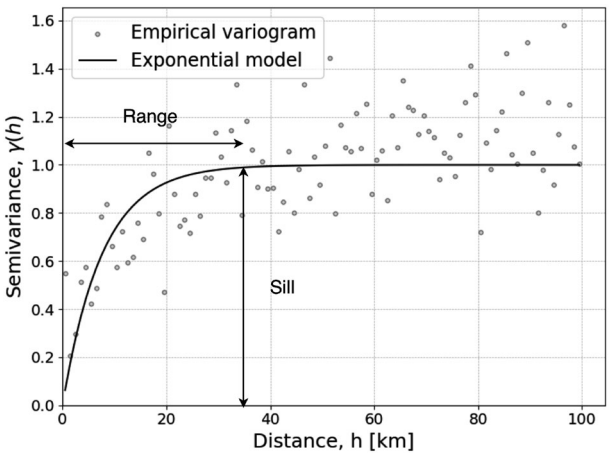


**Fig. 2** Empirical semivariogram and fitted exponential model for a univariate random field

quantifying how the variability of IMs changes with parameters such as distance or even soil type (e.g., Wang and Takada 2005; Jayaram and Baker 2009; Esposito and Iervolino 2011, 2012; Du and Wang 2013; Loth and Baker 2013; Sokolov and Wenzel 2013; Wang and Wenqi 2013; Markhvida et al. 2018; Stafford et al. 2019; Abbasnejadfard et al. 2020, 2021; Wenqi and Ning 2021).

### 2.4.3 Non-ergodic ground motion models

Conventional GMMs are typically developed under the assumption of ergodicity, which treats aleatory variability as spatially independent and estimates it using global datasets. While this assumption simplifies model development, it can obscure important spatial patterns in ground motion variability. As a result, it may introduce bias in seismic hazard assessments, specifically when applied to specific sites or regions.

To overcome this limitation of using the basic GMM functional forms given by Eq. (8), the ergodic assumption is relaxed in what are termed non-ergodic GMMs. These models form a broader class of approaches that introduce spatially varying source, path and site terms directly into the model coefficients. A comprehensive description of this framework is provided by Lavrentiadis et al. (2023), and a full treatment is beyond the scope of this review. However, a subset of recent non-ergodic modelling studies, such as Huang and Galasso (2019), Sgobba et al. (2019), Kuehn and Abrahamson (2020), Liu et al. (2023), incorporate spatial correlation within the residual terms, which is directly relevant to the focus of this manuscript. For example, Huang and Galasso (2019), for example, inferred spatial correlation directly within the GMM and defined a covariance function for within-event residuals, $\delta W_{i,k,m}$, as:

$$Cov(\delta W_{i,k,m}) = \phi_i^2 \cdot \Omega_{i,k,m}(\omega) \tag{25}$$

where $\phi_i$ is the standard deviation of the $\delta W_{i,k,m}$, and $\Omega_{i,k,m}$ represents the correlation matrix for $IM_i$ during earthquake $k$ at recording site $m$, and $\omega$ represents the vector of fitting parameters for the non-ergodic GMM. Assuming stationarity and isotropy, Huang and Galasso (2019) modelled spatial correlation between two locations $x_m$ and $x_n$, separated by distance $h$, corresponding to the upper part of Eq. (28) ($\delta WS$), using the following covariance function:

$$\Omega_{i,k,m}(\omega) = k(x_m, x_n) = \exp\left(\frac{-3h}{b}\right) = \rho(h, Sa(T_i)) \tag{26}$$

where $h$ is the distance and $b$ is the fitted parameter.

Further partitioning the between-event, $\delta B$, and within-event residuals, $\delta W$, helps reduce the aleatory variability and better accounts for non-ergodic effects, and several studies (e.g., Lanzano et al. 2017; Kuehn and Abrahamson 2020; Liu et al. 2023) have shown this. Specifically, they decomposed the between-event residuals to isolate systematic source variability, $\delta L2L$, and the within-event residuals to account for systematic path and site effects, denoted as $\delta P2P$ and $\delta S2S$, respectively, and are expressed as follows:

$$\delta B = \delta L2L + \delta B_0 \tag{27}$$

$$\begin{aligned}\delta W &= \delta WS + \delta S \\ &= \delta P2P + \delta WS_0 + \delta S2S + \delta S_0\end{aligned} \tag{28}$$

where $\delta B_0$, $\delta WS_0$, and $\delta S_0$ represent the remaining aleatory components of the between-event, within-site, and between-site residuals, respectively. Thus, the GMM functional form from Eq. (8) can be rewritten as:

$$\ln IM = \mu(M, R_{rup}, Vs30) + \delta B_0 + \delta L2L + \delta P2P + \delta WS_0 + \delta S2S + \delta S_0 \tag{29}$$

assuming that $\delta B$, $\delta S$, and $\delta WS$ follow Gaussian distributions with zero means and standard deviations of $\tau$, $\phi_S$, and $\phi_{WS}$, respectively. A detailed discussion is provided in Sect. 4.4.

## 3 Non-spatial correlation modelling

### 3.1 Analytical formulations

In the early 2000s, several studies emerged focusing on the development of analytical formulations for different ground motion intensity measures. The term *analytical formulation* herein refers to the application of Pearson's correlation (Sect. 2.4.1) and fitting of a relatively simple parametric functional form. An overview of the key studies examined here is illustrated in Fig. 3, where they are distinguished based on the IM they examine and the ground motion database they utilise.

One of the first notable studies adopting this approach of utilising Pearson's correlation and a fitted analytical functional form using GMM total residuals was by Baker and Cornell (2006) using the pacific earthquake engineering research centers (PEERs) 2000 database for $Sa(T)$ among different period values denoted $T_i$ and $T_j$ in Fig. 3. Concerning the orientation definition of $Sa(T)$, the correlation presented in Baker and Cornell (2006) were based on fixed horizontal orientations and did not involve any rotations to utilise geometric mean or other definitions previously described in Sect. 2.1. Their analysis computed correlations for both horizontal and vertical ground motion components (e.g., $x$-$x$, and $z$-$z$) as well as for different orthogonal orientations (e.g., $x$-$y$, and $x$-$z$), which are given by Eqs. (30), (31), (32), and (33), respectively.

$$\rho^{xx}_{T_i,T_j} = 1 - \cos\left(\frac{\pi}{2} - \left(0.359 + 0.163 I_{(T_{min}<0.189)} \ln\frac{T_{min}}{0.189}\right)\ln\frac{T_{max}}{T_{min}}\right) \tag{30}$$

$$\rho^{zz}_{T_i,T_j} = 1 - 0.77\ln\frac{T_{max}}{T_{min}} + 0.315\left(\ln\frac{T_{max}}{T_{min}}\right)^{1.4} \tag{31}$$

$$\begin{aligned}\rho^{xy}_{T_i,T_j} = &\left(0.79 - 0.023\cdot\ln\sqrt{T_{min}T_{max}}\right)\cdot \\ &\left(1 - \cos\left(\frac{\pi}{2} - \left(0.359 + 0.163 I_{(T_{min}<0.189)} \ln\frac{T_{min}}{0.189}\right)\ln\frac{T_{max}}{T_{min}}\right)\right)\end{aligned} \tag{32}$$

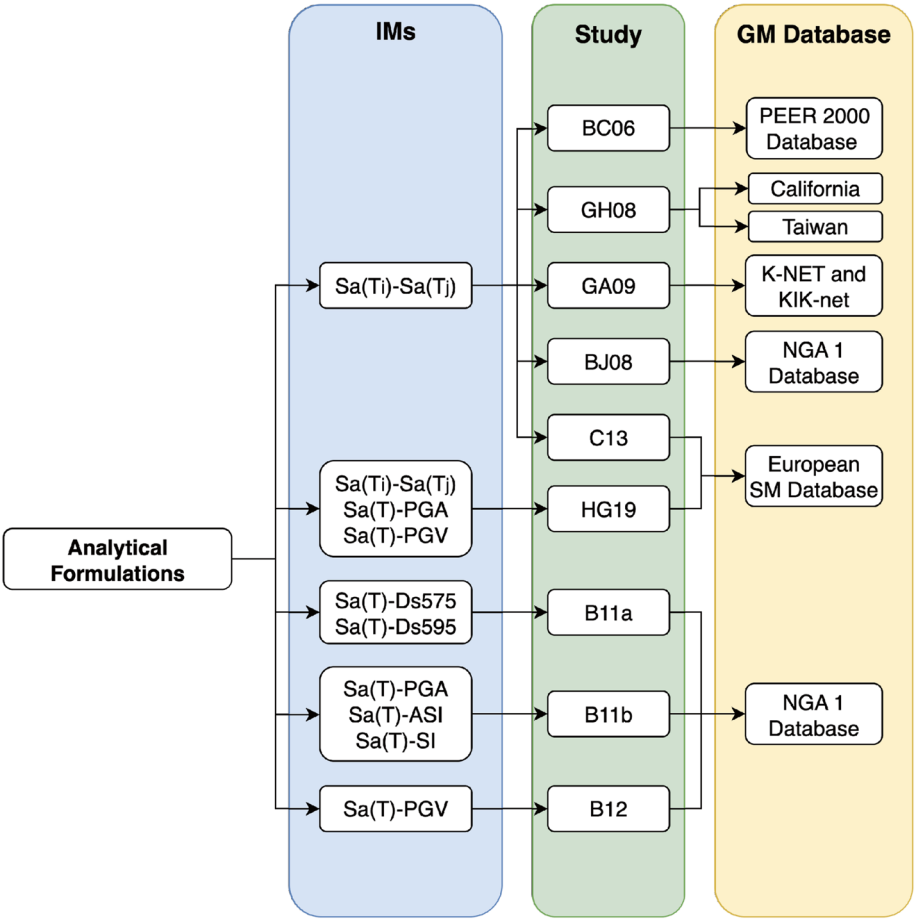


**Fig. 3** Overview of key studies employing analytical formulations to model non-spatial correlation across different IMss. Note: BC06:Baker and Cornell (2006), GH08:Goda and Hong (2008), BJ08:Baker and Jayaram (2008), GA09:Goda and Atkinson (2009), C13:Cimellaro (2013), B11a:Bradley (2011a), B11b:Bradley (2011b), B12:Bradley (2012)

$$\rho^{xz}_{T_i,T_j} = (0.64 - 0.021 \cdot \ln\sqrt{T_{min}T_{max}} \cdot \left(1 - \cos\left(\frac{\pi}{2} - \left(\ln\frac{T_{max}}{T_{min}}\right)\left(0.29 + 0.094 I_{(T_{min<0.189})} \ln\frac{T_{min}}{0.189}\right)\right)\right) \tag{33}$$

where $I_{(T_{min<0.189})}$ equals 1 when $T_{min<0.189}$ and 0 otherwise. Here, $T_{max}$ and $T_{min}$ represent the maximum and the minimum of the two periods of interest, $T_i$ and $T_j$.

Similarly, Baker and Jayaram (2008) computed Pearson's correlation for total residuals of $Sa(T)$ by accounting for rotation of the horizontal components of the acceleration using geometric mean. Baker and Jayaram (2008) utilised the next generation attenuation (NGA)-W1 ground-motion database (Chiou et al. 2008) to develop a fitting model that evaluates a broader range of spectral acceleration periods, described by the equations detailed below:

$$
\begin{aligned}
C_1 &= 1 - \cos\left(\frac{\pi}{2} - 0.366 \ln\left(\frac{T_{\max}}{\max(T_{\min}, 0.109)}\right)\right), \\
C_2 &= \begin{cases} 1 - 0.105\left(1 - \frac{1}{1+e^{100T_{\max}-5}}\right)\left(\frac{T_{\max}-T_{\min}}{T_{\max}} - 0.0099\right) & \text{if } T_{\max} < 0.2, \\ 0 & \text{otherwise.} \end{cases} \\
C_3 &= \begin{cases} C_2 & \text{if } T_{\max} < 0.2 \\ C_1 & \text{otherwise} \end{cases} \\
C_4 &= C_1 + 0.5(\sqrt{C_3} - C3)\left(1 + \cos\left(\frac{\pi T_{min}}{0.109}\right)\right)
\end{aligned}
\tag{34}
$$

$$
\begin{aligned}
\rho_{T_i,T_j} &= C_2, \quad \text{if } T_{\max} < 0.109 \\
\rho_{T_i,T_j} &= C_1, \quad \text{else if } T_{\max} > 0.109, \\
\rho_{T_i,T_j} &= \min(C_2, C_3), \quad \text{else if } T_{\max} < 0.2, \\
\rho_{T_i,T_j} &= C_4, \quad \text{else}
\end{aligned}
\tag{35}
$$

Expanding on this approach, Huang and Galasso (2019) developed a non-ergodic GMM for Italy, incorporating empirical correlations for amplitude-type IMs, including $PGA$ and $Sa(T)$ at 29 periods ranging from 0.01 to 4 seconds. The analysis focused on total residuals and used empirical data to derive analytical correlation models specific to Italy. A key finding was that correlations between $Sa(T_i) - Sa(T_j)$, and $Sa(T) - PGA$ exhibit no significant dependence on magnitude and distance. In contrast, $Sa(T) - PGV$ correlations were shown to be significantly influenced by the large magnitude and short distances. The correlation formulation between $Sa(T)$ was defined as follows:

$$
\begin{aligned}
\rho_{T_i,T_j} &= C_1 = 1 - \cos\left(\frac{\pi}{2} - 0.2351 \ln\left(\frac{T_{\max}}{\max(T_{\min}, 0.1)}\right)\right), \quad \text{if } T_{\max} > 0.1 \\
&= C_2 = 1 - 0.0617\left(1 - \frac{1}{1+e^{100T_{\max}-5}}\right) \times \left(\frac{T_{\max}-T_{\min}}{T_{\max}} - 0.0099\right), \quad \text{else if } T_{\max} \geq 0.1 \\
&= \min(C_2, C_3), \quad \text{else if } T_{\max} \geq 0.2 \\
&= C_3 = C_1 + 0.3131\left(\sqrt{C_1} - C_1\right) \times \left[1 + \cos\left(\frac{\pi T_{\min}}{0.1}\right)\right], \quad \text{else}
\end{aligned}
\tag{36}
$$

For correlations involving $Sa(T) - PGA$ and $Sa(T) - PGV$, the following general model was formulated:

$$
\rho_{Sa(T),IM_i} = \frac{(\phi_1 + \phi_2)}{2} - \frac{(\phi_1 - \phi_2)}{2} \tanh\left[\phi_4 \ln\left(\frac{T}{\phi_3}\right)\right] \tag{37}
$$

where $\phi_1, \phi_2, \phi_3$ and $\phi_4$ are model fitting parameters, whose complete description can be found in Huang and Galasso (2019)).

Goda and Hong (2008) and Goda and Atkinson (2009) extended the study of non-spatial correlations for different $Sa(T)$ periods $T_i$ and $T_j$ (i.e., cross-IM correlation). Although based on different databases, both works partitioned the total residuals computed by Baker and Cornell (2006) into within-event and between-event residuals using Eq. (38), and developed correlation functions for each type of residual.

$$\rho_{\delta T_i,\delta T_j} = \frac{\rho_{\delta B_i,\delta B_j}\tau_i\tau_j + \rho_{\delta W_i,\delta W_j}\phi_i\phi_j}{\sigma_i\sigma_j} \tag{38}$$

where $\rho_{\delta T_i,\delta T_j}$, $\rho_{\delta B_i,\delta B_j}$, and $\rho_{\delta W_i,\delta W_j}$ represent the correlations for total, between-event, and within-event residuals for $IM_i$ and $IM_j$, respectively, and the corresponding standard deviations are $\sigma$, $\tau$, and $\phi$. Those studies utilised 592 ground motions from California earthquakes and 8557 ground motions from K-net and KiK-net databases, respectively, to model the correlation of total and within-event residuals across spectral acceleration periods, following the same methodology of Baker and Cornell (2006). Due to data limitations, Goda and Hong (2008) assumed the between-event residual correlation to follow Eq. (30), as proposed by Baker and Cornell (2006). In contrast, Goda and Atkinson (2009) using K-net, and KiK-net databases derived the following correlation function for between-event residuals:

$$\begin{aligned}\rho_{\delta B_i,\delta B_j} =& \frac{1}{3}\left(1-\cos\left\{\frac{\pi}{2}-\left[1.374+5.586I_{T_{min<0.25}}\times\left(\frac{T_{min}}{T_{max}}\right)^{0.728}\right.\right.\right.\\ &\left.\left.\left.\times\log_{10}\left(\frac{T_{max}}{0.25}\right)\right]\log_{10}\left(\frac{T_{max}}{T_{min}}\right)\right\}\right)\\ &+\frac{1}{3}\left\{1+\cos\left[-1.5\log_{10}\left(\frac{T_{max}}{T_{min}}\right)\right]\right\}\end{aligned} \tag{39}$$

Other studies have focused on non-spatial correlation of IMs beyond the spectral acceleration at different periods. Bradley (2011a, 2011b), and Bradley (2012) developed analytical formulations to fit empirical correlation coefficients using ground motion data from the NGA-W1 Database. These models describe the correlation between $Sa(T)$ and other IMs, including significant duration, $Ds_{575}$, $Ds_{595}$, $PGA$, $ASI$, $SI$, and $PGV$. Bradley (2011a) proposed a functional form to represent the correlation coefficient between significant duration (both $Ds_{575}$ and $Ds_{595}$) and $Sa(T)$ over periods ranging 0.01 - 10s:

$$\rho_{\ln Ds_{xy},\ln Sa(T)} = a_{n-1} + \frac{\ln\left(\frac{T}{b_{n-1}}\right)}{\ln\left(\frac{b_n}{b_{n-1}}\right)}[a_n - a_{n-1}] \quad b_{n-1} \le T < b_n \tag{40}$$

where $Ds_{xy}$ represents the two possible significant durations ($Ds_{575}$ and $Ds_{595}$), and $a_n$, $b_n$ are empirical constants defining correlation and specific periods of vibration, respectively, with further details available in Bradley (2011a)). On the other hand, Bradley (2011b) and Bradley (2012) developed a different functional form for the correlation between $Sa(T)$ and remaining IMs mentioned before with the functional form:

$$\rho_{\ln IM_i,\ln S_a(T)} = \frac{a_n+b_n}{2} - \frac{a_n-b_n}{2}\tanh\left[d_n\ln\left(\frac{T}{c_n}\right)\right] \quad e_{n-1} \le T < e_n \tag{41}$$

where $IM_i$ represents the considered intensity measure, $a_n$, $b_n$, $c_n$ and $d_n$ are empirical constants specific to each correlation model and $e_n$ represent period ranges for different segments. Again, for details see Bradley (2011b) and Bradley (2012).

### 3.2 Empirical data

Similar to the works previously described that utilised Pearson's correlation followed by analytical functional forms, further studies, illustrated in Fig. 4, focused on simply providing the raw empirical correlation coefficients. The empirical correlation was tabulated by extracting paired observations across various ground motions and IMs. This was computed at discrete points, with the intention being that where direct data were unavailable for certain intensity measure pairs, linear interpolation is employed to estimate intermediate correlation values.

Studies like Akkar et al. (2014), Baker and Bradley (2017), Tarbali et al. (2023) focused on correlating IMs representing various ground motion characteristics, such as amplitude ($PGA$ and $PGV$), frequency content ($Sa(T)$, $ASI$, $SI$, and $DSI$), significant duration ($Ds_{575}$, and $Ds_{595}$), and cumulative effects ($CAV$, and $Ia$), on the other hand, Bradley (2011a) computed correlations for $Sa(T)$ residuals alone. Bradley (2011a) used seven GMMs and primarily sought to understand the relationship between significant durations and several other IMs. It was found that $Ds_{575}$ and $Ds_{595}$ tend to be negatively correlated with high-frequency amplitude-based IMs, weakly negatively correlated with moderate-frequency amplitude-based IMs, and weakly positively correlated with low-frequency amplitude-based IMs and cumulative absolute velocity. Akkar et al. (2014) developed GMMs to estimate horizontal and vertical damping scaling factors, as well as vertical-to-horizontal $Sa(T)$ ratios, using a subset of the pan-European strong-motion RESORCE database. With these GMMs, empirical correlations between $Sa(T)$ residuals for periods ranging 0.01 and 4 seconds were computed, and are presented in Fig. 5(a).

Baker and Bradley (2017) and Tarbali et al. (2023) investigated IM correlations using a NGA-W2 ground motion database. Baker and Bradley (2017) also found that IM correlations are largely independent of magnitude, distance, and other site parameters. An example of the $Sa(T) - Sa(T)$ correlation developed by Baker and Bradley (2017) is illustrated in Fig. 5(b), showcasing how this correlation behaves across different periods. Building on previous work and using the same database as Baker and Bradley (2017), Tarbali et al. (2023) analysed the correlation of various IMs for ground motions containing near-fault

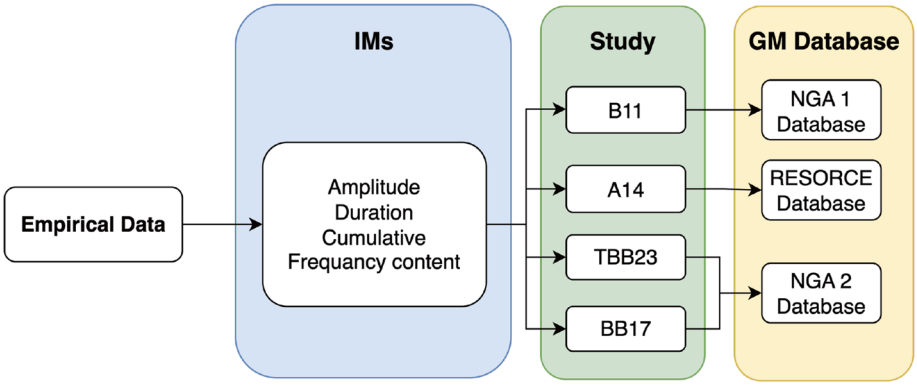


**Fig. 4** Overview of key studies employing empirical data from Pearson's correlation to model non-spatial correlation. B11:Bradley (2011a), A14:Akkar et al. (2014), TBB23:Tarbali et al. (2023), BB17:Baker and Bradley (2017)

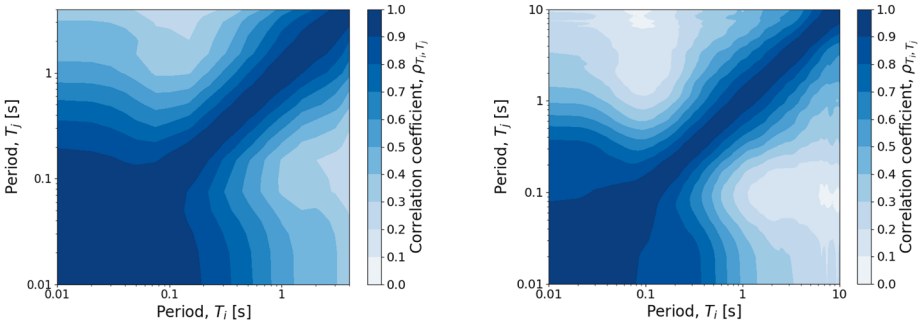


**Fig. 5** Correlation contour of empirical $Sa(T) - Sa(T)$ coefficients developed by (**a**) Akkar et al. (2014) and (**b**) Baker and Bradley (2017)

directivity velocity pulses. Their findings indicate that IM correlations for directivity ground motions are generally consistent with existing models such as Baker and Bradley (2017) for non-pulse-like ground motions. Any observed differences were attributed to variations in the ground-motion distribution or sample size (Tarbali et al. 2023).

### 3.3 Artificial neural networks

While the previous sections have developed models based on either empirical tables of coefficients or by fitting analytical functional forms to such data, machine learning methods have emerged as transformative tools in earthquake engineering and have also been implemented in correlation modelling. In this context, it is important to note that the data modelled still consists of residuals derived from GMMs, and the fundamental assumptions underlying these residuals remain unchanged. What distinguishes the use of ANNs is the replacement of traditional analytical fitting functions, which rely on predefined functional forms and assumptions to model correlation coefficients, with a data-driven approach capable of directly representing complex relationships from the data. Aristeidou et al. (2024) utilised ANN-based regression models to estimate correlations for several well-established and next-generation IMs, including $PGA$, $PGV$, $PGD$, $Sa(T)$, $Ds_{575}$, $Ds_{595}$, $Sa_{avg}(T)$, and $FIV3(T)$. Their study showed that ANN-derived correlation estimates exhibit lower residual errors and better alignment with empirical data.

## 4 Spatial correlation modelling using the same IMs

### 4.1 Analytical formulations

Contrary to the formulations presented in Sect. 3, which focus on non-spatial correlation, several studies have developed models to describe inter-site correlation for the same IM. These models predominantly use exponential decay functions to explain the decreasing correlation with increased distance. Most studies in the literature, particularly those involving typical IMs such as $PGA$, $PGV$, and $Sa(T)$ with the same period $T$, have developed spa-

tial correlation models using different ground motion databases, depending on the specific goals of the study, as illustrated in Fig. 6.

Goda and Hong (2008) investigated spatial correlation relationships for $PGA$ and $Sa(T)$. For the latter IM, they derived equations to fit the correlation coefficients obtained using Eq. (38), adapted for total residuals of the same $T_i$ in $Sa(T)$ as follows:

$$\rho_{\delta T}(h, T_i) = \frac{\rho_{\delta B}(T_i) \cdot [\tau(T_i)]^2 + \rho_{\delta W}(h, T_i) \cdot [\phi(T_i)]^2}{[\sigma(T_i)]^2} \tag{42}$$

Since the analysis considers ground motions from the same seismic event across multiple site locations, the between-event residuals are fully correlated, meaning $\rho_{\delta B}(T_i) = 1$ and the expression simplifies to:

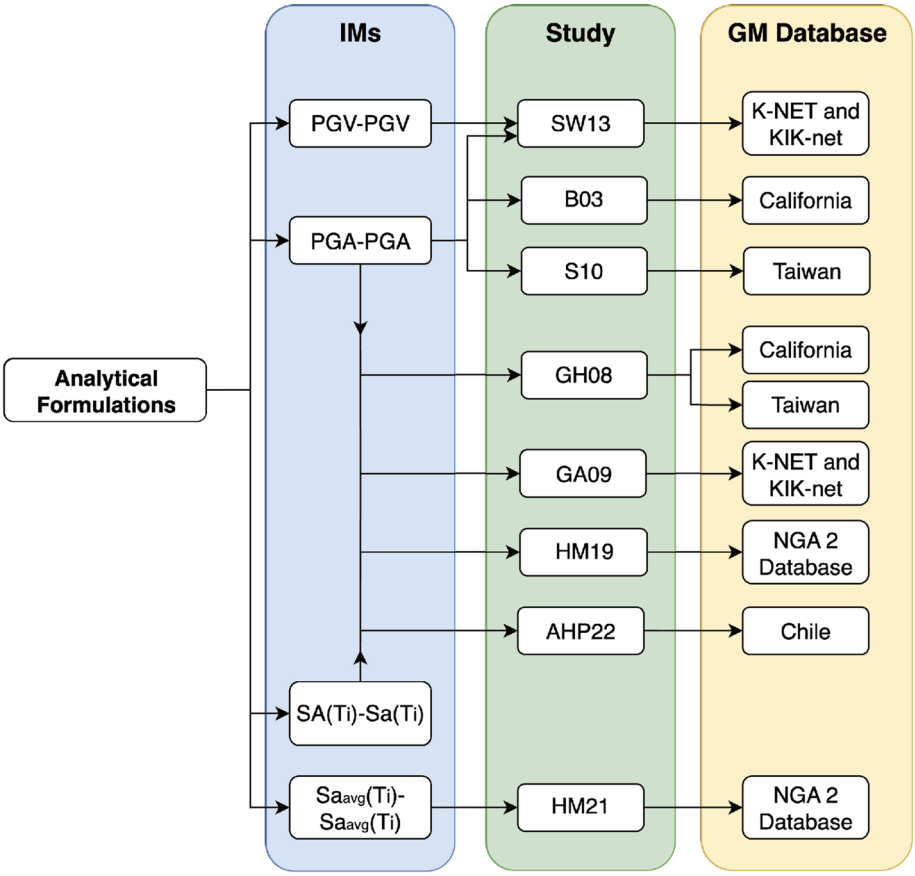


**Fig. 6** Overview of key studies that employed analytical formulations to model spatial correlation for a single IM. B03:Boore et al. (2003), GH08:Goda and Hong (2008), GA09:Goda and Atkinson (2009), S10:Sokolov et al. (2010), SW13:Sokolov and Wenzel (2013), HM19:Heresi and Miranda (2019), HM21:Heresi and Miranda (2021), AHP22:Aldea et al. (2022)

$$\rho_{\delta T}(h,T_i) = \left(\frac{\tau(T_i)}{\sigma(T_i)}\right)^2 + \rho_{\delta W}(h,T_i)\left(\frac{\phi(T_i)}{\sigma(T_i)}\right)^2 \tag{43}$$

Alternatively, Eq. (44) was also proposed by Goda and Hong (2008) to calculate the spatial correlation of total residuals introducing term $[\sigma_d(h,T_i)]^2$, firstly proposed by Boore et al. (2003), representing the variance of the difference of within-event residuals between two locations separated by a distance $h$, (i.e., $\delta W_{n,k}(T_i) - \delta W_{m,k}(T_i)$, where $k$ denotes the earthquake event and $n$ and $m$ represent distinct site locations). This term $[\sigma_d(h,T_i)]^2$, captures the spatial correlation structure of within-event residuals and was initially investigated by Boore et al. (2003) in the context of $PGA$ residuals.

$$\rho_{\delta T}(h,T_i) = 1 - \frac{[\sigma_d(h,T_i)]^2}{2[\sigma(T_i)]^2} \tag{44}$$

Goda and Hong (2008) applied two approaches to model within-event spatial correlation of residuals. The first approach involved the direct evaluation of the sample Pearson's correlation coefficient, as given by Eq. (45). The second approach was based on $[\sigma_d(h,T_i)]^2/2$ and formulated by Eq. (46). Several studies have employed exponential fitting to model the decay of spatial correlation with distance, using two fitted parameters, $\alpha$, and $\beta$ that depend on the IM used, as shown in Eq. (47). For Californian earthquakes, Goda and Hong (2008) explored this exponential fitting. Similarly, Sokolov et al. (2010) applied the same format of exponential decay model to fit within-event residuals from Taiwanese earthquakes, incorporating correction factors into the GMM of Morikawa et al. (2008), who demonstrated that grouping ground motion data at specific stations could reduce model uncertainty. Likewise, Sokolov and Wenzel (2013) investigated the same decay pattern shown in Eq. (47) for Japanese earthquakes, considering different soil type classifications and GMMs in the residual calculations.

$$\rho_{\delta W}(h,T_i) = \frac{\mathrm{COV}[\delta W_{k,n}(T_i), \delta W_{k,m}(T_i)]}{[\phi(T_i)]^2} \tag{45}$$

$$\rho_{\delta W}(h,T_i) = 1 - \frac{[\sigma_d(h,T_i)]^2}{2[\phi(T_i)]^2} \tag{46}$$

$$\rho_{\delta W}(h,T_i) = \exp(\alpha(T_i)\cdot h^{\beta(T_i)}) \tag{47}$$

A similar study by Goda and Atkinson (2009) using 7,780 records from 106 earthquakes from the K-net and KiK-net databases, extended this approach. In addition to the previous methods for calculating $\rho_{\delta W}(h,T_i)$ (Eq. (45) and (46)), a modified fitting method incorporating a third parameter, $\gamma(T_i)$, was introduced:

$$\rho_{\delta W}(h,T_i) = \max\left[\gamma(T_i)\cdot\exp[-\alpha(T_i)\cdot h^{\beta(T_i)}] - \gamma(T_i) + 1; 0\right] \tag{48}$$

Other studies, such as Heresi and Miranda (2019) and Aldea et al. (2022), used a different exponential function to those presented earlier to describe the decay of spatial correlation

with distance. This alternative exponential formulation has also two parameters $\alpha$, and $\beta$, as:

$$\rho_{\delta W}(h, T_i) = \exp\left[-\left(\frac{h}{\beta(T_i)}\right)^{\alpha(T_i)}\right] \tag{49}$$

Both studies developed spatial correlation models using Eq. (49). Heresi and Miranda (2019) analysed 39 worldwide seismic events from active shallow crust regions and found that the parameter $\alpha(T_i)$ remained relatively constant across different IMs and fixed it at 0.55. Similarly, Aldea et al. (2022), based on earthquakes from the Chilean subduction zone, also fixed $\alpha$ at 0.59 to simplify the spatial correlation model. The $\beta$ parameter, however, was fitted individually for each event. Heresi and Miranda (2019) also performed Monte Carlo simulations to incorporate the event-to-event variability of the within-event term in the spatial correlation model, rather than considering one correlation model from a single event or set of events. As a result, they developed spatial correlation models for $PGA$ and $Sa(T)$ with median, $\hat{\beta}$, and standard deviation, $\sigma$, represented as:

$$\hat{\beta} = \begin{cases} 4.231T^2 - 5.180T + 13.392 & \text{if} \quad T < 1.37\text{s} \\ 0.140T^2 - 2.249T + 17.050 & \text{if} \quad T \geq 1.37\text{s} \end{cases} \tag{50}$$

$$\sigma(T) = 4.63 \times 10^{-3}T^2 + 0.028T + 0.713 \tag{51}$$

For the Aldea et al. (2022) model, the median $\hat{\beta}$ is given by:

$$\hat{\beta} = \begin{cases} 14.400 - 17.000T & \text{if} \quad T \leq 0.40\text{s} \\ 14.743 + 7.795\ln(T) & \text{if} \quad 0.40\text{s} < T \leq 0.75\text{s} \\ 12.500 & \text{if} \quad 0.75\text{s} < T \leq 3.00\text{s} \\ 5.063 + 6.769\ln(T) & \text{if} \quad 3.0\text{s} < T \leq 10.00\text{s} \end{cases} \tag{52}$$

For the Heresi and Miranda (2019) model, if no uncertainty is considered, the standard deviation is given by Eq. (51). Otherwise $\beta$ is modelled as a lognormal random variable with parameters $\ln(\hat{\beta})$ and $\sigma(T)$. If additional epistemic uncertainty is considered, $\sigma(T)$ may be scaled by a user-defined uncertainty factor.

Due to the current lack of direct formulations for calculating the spatial correlation for $Sa_{avg}(T)$, Heresi and Miranda (2021) investigated an indirect approach based on $Sa(T)$ -based correlation models and the intrinsic relationship between these two IMs. The results was an equivalent correlation model for $Sa_{avg}(T)$ at two distinct locations denoted as sites $m$ and $n$ and is given as:

$$\rho_{\ln Sa_{avg}(T)_m, \ln Sa_{avg}(T)_n} = \frac{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\rho_{\ln Sa(T_i)_m, \ln Sa(T_j)_n} \cdot \sigma_{\ln Sa(T_i)_m} \cdot \sigma_{\ln Sa(T_j)_n}}{\sigma_{\ln Sa_{avg}(T)_m} \cdot \sigma_{\ln Sa_{avg}(T)_n}} \tag{53}$$

where $\rho_{\ln_{Sa(T_i)m},\ln_{Sa(T_j)n}}$ represents the cross-IM spatial correlation coefficient (Sect. 5) between two sites $m$ and $n$ at spectral periods $T_i$ and $T_j$, and $N$ is the total number of periods considered in the range $c$ in Eq. (4). The term $\sigma_{\ln_{Sa(T_i)m}}$ corresponds to the standard deviation obtained from the GMM used to compute the residuals for $Sa(T)$ at site $m$ and period $T_i$. Meanwhile, $\sigma_{\ln_{Sa_{avg}(T),m}}$ denotes the standard deviation of $Sa_{avg}(T)$, which can be derived indirectly using standard deviations of $Sa(T)$ values, following the approach (Kohrangi et al. 2017):

$$\sigma_{\ln_{Sa_{avg}(T)}} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\rho_{\ln_{Sa(T_i)},\ln_{Sa(T_j)}}\cdot\sigma_{\ln_{Sa(T_i)}}\cdot\sigma_{\ln_{Sa(T_j)}} \tag{54}$$

### 4.2 Semivariograms

Several studies, as shown in Fig. 7, have employed semivariograms, introduced in Sect. 2.4.2 to model or quantify spatial correlation using a single IM. In practice, the semivariogram models are often applied to data with specific parameter values, such as the sill and range.

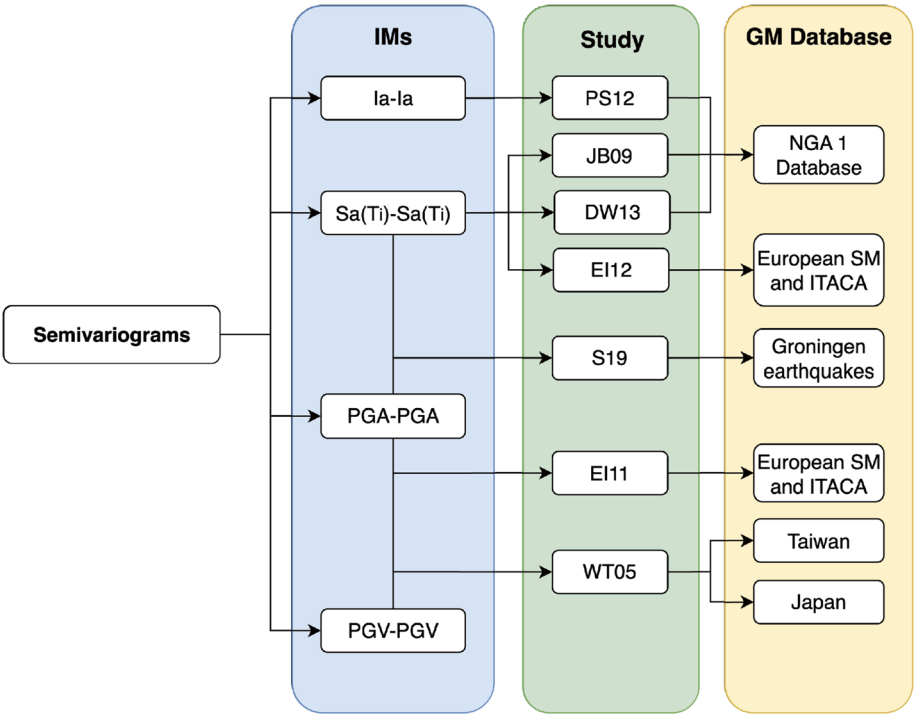


**Fig. 7** Overview of key studies that employed semivariograms to model spatial correlation for a single IM. WT05:Wang and Takada (2005), JB09:Jayaram and Baker (2009), EI11:Esposito and Iervolino (2011), EI12:Esposito and Iervolino (2012), PS12:Foulser-piggott and Stafford and (2012), DW13:Du and Wang (2013), S19:Stafford et al. (2019)

While semivariograms and analytical formulation both aim to characterise spatial dependence in ground motion residuals, they do so from different statistical perspectives. Analytical models typically regress Pearson's correlation coefficients against inter-station distance, providing a direct measure of linear correlation. In contrast, semivariograms quantify how the variance of the residuals increases with separation distance and infer the underlying covariance structure from this relationship. As a result, semivariogram-based approaches offer greater flexibility in representing spatial variability, particularly when the variance does not adhere to the assumptions implicit in Pearson-based regressions. These methodological differences may lead to variations in the inferred correlation lengths and have implications for their practical application in seismic hazard analyses.

One of the early studies to apply semivariogram-based modelling to characterise the spatial correlation of total residuals was Wang and Takada (2005). In their analysis of $PGV$ residuals from five Japanese earthquakes and the 1999 Chi-Chi earthquake in Taiwan, they proposed a macro-spatial correlation model with an exponential decay function of the form:

$$\rho(h) = \exp(-h/b) \tag{55}$$

where $h$ is the separation distance and $b$ is a correlation length parameter estimated from the data (i.e., range). Two different GMMs were considered for the calculation of $PGV$ values (Annaka et al. (1997) and (Midorikawa and Ohtake 2002)) and consequently, using Eq. (9) to calculate $PGV$ residuals, to be applied in the semivariogram.

The classical semivariogram estimator, originating from geostatistics, is thoroughly documented in Cressie (1993). In the context of earthquake engineering, spatial correlation modelling of ground motion within-event residuals was introduced by Jayaram and Baker (2009). Their study developed a global correlation model that accounts for the clustering of $V_{s30}$ values, utilising data from seven historical earthquakes. The spatial correlation was modelled using an exponential function:

$$\rho(h) = \exp(-3h/b) \tag{56}$$

In this formulation, the parameter $b$ is a function of $T$ and the presence of $V_{s30}$ clustering. The expressions for $b$ are as follows:

- No clustering of $V_{s30}$ values (for $T < 1$ second):

$$b = 8.5 + 17.2T \tag{57}$$

- With clustering of $V_{s30}$ values (for $T < 1$ second):

$$b = 40.7 + 15.0T \tag{58}$$

- For $T > 1$ second (regardless of clustering):

$$b = 22.0 + 3.7T \tag{59}$$

However, the classical estimator is known to be sensitive to outliers. To address this issue, robust alternatives, such as Cressie-Hawkins estimator (Cressie and Hawkins 1980), are often preferred to enhance reliability. Several studies have compared classical and robust semivariograms in the development of spatial correlation models across different ground motion databases.

For instance, Schiappapietra et al. (2022) proposed three spatial correlation models tailored for distinct regions of Italy, namely, northern, central and southern Italy. These models underscore the importance of accounting for spatial correlation and associated uncertainties in seismic risk analyses. Each model follows the exponential form of Eq. (56), with the parameter $b$ expressed as a function of the period $T$. The models, valid for $PGA$ and $Sa(T)$ up to 2 seconds, are given as:

- Northern Italy:

$$b(T) = \begin{cases} 27.48 - 52.20 \cdot (T - 0.55), & \text{for } T \leq 0.55 \\ 27.48 + 15.81 \cdot (T - 0.55), & \text{for } T > 0.55 \end{cases} \tag{60}$$

- Central Italy:

$$b(T) = \begin{cases} 17.87 - 8.52 \cdot (T - 1), & T \leq 1 \\ 17.87 + 7.85 \cdot (T - 1), & T > 1 \end{cases} \tag{61}$$

- Southern Italy:

$$b(T) = 23.25 - 5.44 \cdot T \tag{62}$$

In contrast, Esposito and Iervolino (2011) and Esposito and Iervolino (2012) focused on broader applications of robust semivariogram estimators, analysing spatial correlations in ground motion residuals across European and Italian ground motion databases for a variety of IMs, including $PGA$, $PGV$, and $Sa(T)$.

Using the NGA-W1 database (Chiou et al. 2008), studies such as Jayaram and Baker (2009) and Du and Wang (2013) made significant contributions to spatial correlation research involving $Sa(T)$. While both studies focused on spatial correlation, Du and Wang (2013) extended the analysis to include correlations based on $V_{s30}$ values for $CAV$ and $Ia$. In this study, a similar global spatial correlation model for $Sa(T)$ was developed using the same formulation for calculating correlation as detailed in Eq. (56). Additionally, Foulser-Piggott and Stafford and (2012) conducted a focused study on $Ia$ using the same NGA-W1 database. Stafford et al. (2019), on the other hand, developed a spatial correlation model for $Sa(T)$ using semivariograms applied to a dataset of induced earthquakes from the Groningen gas field in the Netherlands. While the methodological approach is similar, their study differs in the type of seismicity and its regional focus, addressing spatial correlation under the specific characteristics of induced seismic events.

### 4.3 Methodology accounting for path and site effects

Recent advancements have emphasised non-stationary and anisotropy patterns of spatial correlations. For example, Bodenmann et al. (2023) introduced an innovative framework for considering these concepts in spatial correlations of ground motion IMs with site-specific effects, addressing localised geological variability and improving the practical application of spatial correlation models. In their methodology, they followed three assumptions, starting with the assumption that the correlation between sites decreases as the Euclidean distance between them increases, and for this assumption, they use an exponential function described by Eq. (63), which has the same formulation presented, for example, in Eq. (49).

$$\rho_E(d_E;\psi_E) = \exp(-(d_E/l_E)^{\gamma_E}) \tag{63}$$

where $d_E$ is the Euclidean distance between two sites (see Fig. 8), $l_E$ is the length scale in kilometres, $\gamma_E \in (0,2)$, and $\psi_E$ is a vector that contains all possible combinations for $l_E$ and $\gamma_E$.

The second assumption is that the correlation between two sites may also depend on their position relative to the earthquake rupture, so the epicentral azimuth, $\theta$, was used to characterise this relative position and assumed that correlation between sites decreases as the difference in their azimuths increases as given by:

$$\rho_A(d_A;l_A) = (1 + d_A/l_A)(1 - d_A/180)^{180/l_A}) \tag{64}$$

where $d_A$ is the angular distance, represented in Fig. 8 and $l_A$ is the length scale in degrees and can vary between $0°$ and $45°$.

The last assumption is that sites with similar soil conditions may have stronger correlations. As shown by Eq. (65), an exponential form was used for the correlation decay using $V_{s30}$ values.

$$\rho_S(d_S;l_S) = \exp(-d_S/l_S) \tag{65}$$

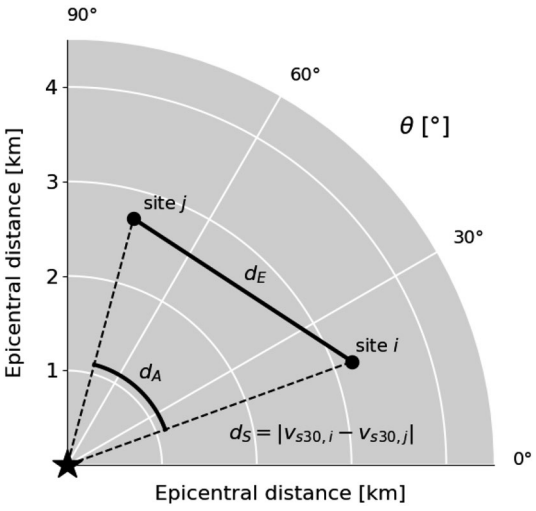


**Fig. 8** Schematic diagram showing the euclidean distance, $d_E$, azimuth distance, $d_A$, and site effect, $d_S$, parameters described in Bodenmann et al. (2023)

where $d_S$ is the absolute difference between two sites $V_{s30}$ values, illustrated in Fig. 8 and $l_S$ is the length scale in $ms^{-1}$ ($\in \mathbb{R}^+$). In the end, to simultaneously account for spatial proximity, path, and site effects they introduced the correlation methodology EAS represented by Eq. (66). Their methodology emphasised enhanced accuracy in regional risk assessments by considering variations in spatial correlation patterns across multiple IMs and leveraging high-resolution site characterisation data.

$$\rho_{EAS}(d_E; d_A; d_S; \psi_{EAS}) = \rho_E(d_E) \cdot (w\rho_A(d_A) + (1-w)\rho_S(d_S)) \tag{66}$$

where $w$ is the weight parameter ($\in (0, 1)$), and $\psi_{EAS}$ is a vector that collects all possible combinations for $\gamma_E, l_E, l_A, l_S$ and $w$.

Beyond its mathematical formulation, this methodology also carries several practical implications. By explicitly incorporating spatial proximity, path geometry, and site-condition similarity, the EAS model can reduce the effective aleatory variability in predicted ground motion, particularly in regions where local geologic variability strongly controls spatial patterns. The model is calibrated using recorded ground motion in combination with high-resolution site-characterisation data, enabling its parameters to reflect regional attenuation behaviour. An advantage of this approach is its ability to represent non-stationary and anisotropic correlation structures that traditional isotropic models discussed in Sec.ts 4.1 and 4.2 cannot capture. However, this flexibility requires more detailed site information and a sufficiently dense seismic network for robust calibration, which may limit its applicability in data-scarce regions. Although the methodology is relatively recent, it has already motivated extensions in related fields; for example, similar principles have been applied to spatial correlation in liquefaction hazard modelling (e.g., Pretell et al. 2024), indicating growing interest in path- and site-aware spatial models.

### 4.4 Non-ergodic ground motion models

Traditional studies have modelled spatial correlation as a function of inter-site distance, assuming stationarity and isotropy, but emerging research has shown that these assumptions may not hold in all cases, particularly in regions with complex geological structures or site-specific effects (Bodenmann et al. 2023). Recent studies have made significant advances in refining spatial correlation models within the context of non-ergodic ground motion models. For example, Sgobba et al. (2019), using the non-ergodic GMM developed by Lanzano et al. (2017), investigated the spatial correlation of ground motion residuals in Italy, highlighting the impact of regional geology. Huang and Galasso (2019) analysed the implications of spatial correlation for infrastructure vulnerability assessments, providing insights into site-to-site variability. Kuehn and Abrahamson (2020) proposed methodologies to incorporate non-ergodic corrections into GMMs, emphasising the need for improved epistemic uncertainty qualification, and Liu et al. (2023) explored regionalised GMMs that account for spatially varying source, path, and site effects using records from the Ridgecrest database (Rekoske et al. 2020).

Kuehn and Abrahamson (2020) and Liu et al. (2023), additionally, developed spatial correlation models specifically for systematic source, path, and site effects, which can be seen in Eq. (29), modelled as functions of geographical coordinates of earthquakes, $x_k$, and sites, $x_s$, assuming Gaussian processess (GPs):

$$\delta L2L = f_1(x_k) \sim GP(0, k(x_k, x_k')) \tag{67}$$

$$\delta P2P = f_2(x_k, x_s) \sim GP(0, k([x_k, x_s], [x_k', x_s'])) \tag{68}$$

$$\delta S2S = f_3(x_s) \sim GP(0, k(x_s, x_s')) \tag{69}$$

where $k(x, x')$, $k([x_k, x_s], [x_k', x_s'])$, and $k(x_s, x_s')$ are covariance functions for source, path and site effects, respectively. Thus, the main residuals ($\delta B, \delta WS$, and $\delta S$) follow a multivariate Gaussian distribution where the variance is represented by the systematic effects explained by covariance functions and the aleatory effects explained by $\delta_{ij}$ multiplied by the aleatory standard deviations as (Liu et al. 2023):

$$\delta B \sim \mathcal{N}(0, k(x_e, x_e') + \delta_{ij}\tau_0^2) \tag{70}$$

$$\delta WS \sim \mathcal{N}(0, k([x_e, x_s], [x_e', x_s']) + \delta_{ij}\phi_{0,WS}^2) \tag{71}$$

$$\delta S \sim \mathcal{N}(0, k(x_s, x_s') + \delta_{ij}\phi_{0,S2S}^2) \tag{72}$$

where $\delta_{ij}$ takes the value 1 if two earthquakes, paths or sites are identical. For more detailed information regarding the covariance functions see, for example, Paciorek and Schervish (2006), Kuehn and Abrahamson (2020), Liu et al. (2023).

To model source-effects, Kuehn and Abrahamson (2020) assumed several stationary covariance functions, such as exponential, squared exponential, spherical and a special case of the Matérn class of covariance functions. Eq. (73) shows the typical isotropic covariance function (exponential form) used by Kuehn and Abrahamson (2020):

$$k(x_e, x_e') = \tau_{L2L}^2 \left( -\frac{||x_e - x_e'||}{l} \right) \tag{73}$$

To demonstrate what a stationary covariance function represents for the source-effect spatial correlation, an exponential covariance function was used for two events in Fig. 9. Since the covariance function only depends on the distance between two events, the shapes of the two correlations for the two events are the same.

To account for fault geometry, using an isotropic covariance function may be inappropriate. For instance, in a fictitious scenario like the one shown in Fig. 10(a), a pair of earthquake epicentres on the same fault (red pair) should exhibit higher correlation than a pair on different faults (blue pair), even if they share the same separation distance. To address this, Liu et al. (2023) applied the following equation, originally proposed by Paciorek and Schervish (2006), to construct an anisotropic and non-stationary positive definite covariance function:

$$k(x_e, x_e') = \tau_{L2L}^2 2^{D/2} |\Lambda(x_e)|^{0.25} |\Lambda(x_e')|^{0.25} |\Lambda(x_e) + \Lambda(x_e')|^{-0.5} \exp(-\sqrt{Q(x_e, x_e')}) \tag{74}$$

Equation (74) follows the form of a non-stationary squared exponential covariance function. However, instead of a spatially constant Kernel matrix, the function incorporated $\Lambda(x_e)$, which gives non-stationary and anisotropic behaviour. The premultiplication in Eq. (74)

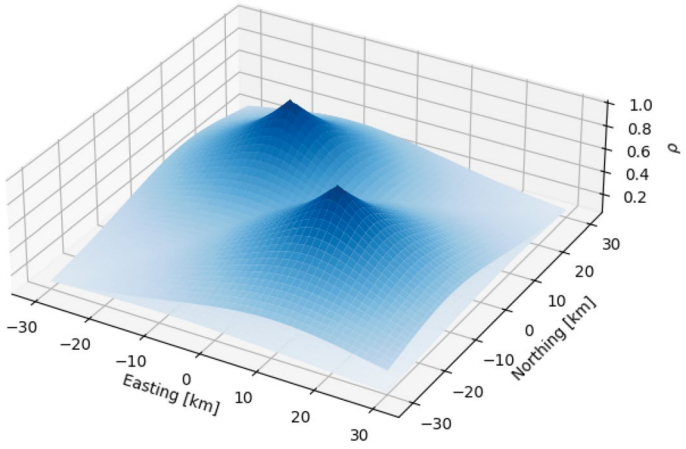


**Fig. 9** Correlation of source effects for two sites using an exponential stationary covariance function

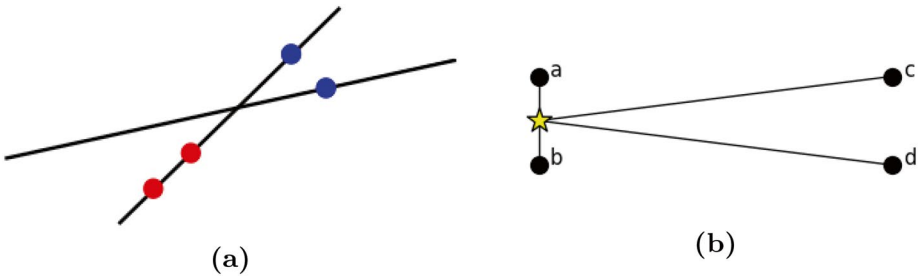


**Fig. 10** (**a**) Illustration of spatial stationarity: two pairs of earthquake epicentres (blue and red circles) have equal separation distances, but are located on different fictitious faults. The red pair lies along the same fault, while the blue pair spans across two distinct faults (black lines), potentially resulting in similar correlation despite differing geological contexts and (**b**) illustration of a stationary covariance function for path effects. Two pairs of stations ($a$–$b$ and $c$–$d$) have equal inter-station distances and therefore exhibit the same correlation, despite pair $a$–$b$ being farther from the source region and expected to show higher correlation due to more similar propagation paths

ensures the positive definiteness of the covariance function (Liu et al. 2023). Here, $D$ represents the dimension of $x_e$, $\Lambda(x_e)$ is a $D$ by $D$ matrix-values function that describes the relationship between correlation length and the earthquake location $x$, and $Q(x_e, x'_e)$ defines a generalised squared distance (i.e., a Mahalanobis-type quadratic form) between points $x_e$ and $x'_e$, according to their respective anisotropy matrices (for more details, see Paciorek and Schervish (2006)).

To model spatial correlation for path effects, Liu et al. (2023) and Kuehn and Abrahamson (2020) employed a similar covariance function as in Eq. (74). However, in this case, the quadratic function, $Q$, is expressed in terms of between-site or between-earthquake distance, considering paths from the same earthquakes or to the same sites. This formulation accounts for path effects, as demonstrated in Fig. 10(b), where two site pairs ($a$-$b$ and $c$-$d$) share the same between-site distance. The pair farther from the epicentre ($c - d$) is expected to exhibit higher correlation due to more similar propagation paths (Liu et al. 2023). If path-

effects originate from the same earthquake or travel to the same site, the covariance function is given by:

$$k([x_s, x_e][x'_s, x'_e]) = \phi^2_{P2P} 2^{D/2} |\Lambda(x_s, x_e)|^{0.25} |\Lambda(x'_s, x'_e)|^{0.25} \cdot |\Lambda(x_s, x_e) + \Lambda(x'_s, x'_e)|^{-0.5} \exp\left(-\sqrt{Q([x_s, x_e][x'_s, x'_e])}\right). \quad (75)$$

Otherwise:

$$k([x_s, x_e][x'_s, x'_e]) = 0 \quad (76)$$

In terms of modelling spatial correlation of site effects, Kuehn and Abrahamson (2020) did not account for it; such correlations can be incorporated using nonstationary correlation functions (Chen et al. 2021) or through isotropic and stationary covariance functions based on between-site distances, as described by Liu et al. (2023):

$$k(x_s, x'_s) = \phi^2_{S2S} \exp\left(-\frac{||x_s - x'_s||}{l}\right) \quad (77)$$

Using non-ergodic GMMs enables the partitioning of IM residuals, which allows the spatial correlation structure of systematic effects to be estimated. Studies such as Kuehn and Abrahamson (2020), Liu et al. (2023), among others, developed spatial correlation models based on this approach. They proposed isotropic and stationary models for source and site effects, as well as anisotropic and non-stationary models that account for fault geometry and path effects. Liu et al. (2023) found that anisotropic and non-stationary models more accurately captured fault geometries and extrapolated more reliably to data-sparse regions compared to isotropic and stationary models, particularly for source and path effects. In contrast, the isotropic and stationary model performed well in capturing the spatial distribution of site effects. Liu et al. (2023) compared their results with those from Kuehn and Abrahamson (2020) and found consistent spatial correlation structures for path effects across different regions, suggesting that the models may be transferable.

### 4.5 Physics-based simulations

Physics-based simulation represent a key development in the availability of ground motion signals. In terms of ground motion correlations, a key contribution was made by Chen and Baker (2019) who used PBS to investigate intra-event spatial correlation in California. It highlighted the importance of considering regional geological conditions and source characteristics in shaping spatial correlation patterns. Importantly, they demonstrated that PBSs can replicate key trends observed in empirical models, while also offering the flexibility to explore spatially beyond the limitations of observed data. Their findings suggest that non-stationary and anisotropic spatial correlations depend on source effects, path effects, and the relative location to the rupture, for example at small distances, sites located near the same portion of the fault rupture tend to exhibit stronger correlation due to shared source effects, at moderate to large distances, correlation is more influenced by shared wave propagation paths, and sites on opposite sides of a rupture can even display negative correlation, reflect-

ing complex rupture-relative spatial dependencies. Infantino et al. (2021) applied PBS to simulate the 1978 Volvi earthquake in Northern Greece, using the spectral elements in elastodynamics with discontinuous Galerkin (SPEED) code (see https://speed.mox.polimi.it/). Their study examined the spatial correlation of synthetic ground motion IMs, providing evidence for anisotropic correlation patterns. The results showed a notable difference between fault-normal and fault-parallel components, with correlation ranges increasing with period and reaching values up to $\approx 64km$ for long-period motions. These findings challenged the conventional assumptions of isotropic and stationary correlation structures often adopted in ground motion models and suggested that fault orientation and directivity play an important role in spatial correlation characteristics. Expanding on this, Schiappapietra and Smerzini (2021) conducted a detailed scenario-based simulation for the 2016 Norcia earthquake in Central Italy, again using the SPEED code. They generated over 4,000 synthetic ground motion records across an $80km \times 80km$ domain, capturing various fault-relative directions. Their study systematically evaluated the correlation of ground motion IMs, confirming the non-stationarity and anisotropy of the correlation structure. They demonstrated that spatial correlation varies depending on the relative orientation to the fault, particularly for near-source stations. Furthermore, they provided a comparative assessment between simulated and empirical correlation models (e.g., Jayaram and Baker (2009)), identifying significant deviations and reinforcing the need for simulation-informed models in complex seismo-tectonic regions.

Lin and Smerzini (2022) extended the work by Infantino et al. (2021) for constructing an updated 3D numerical model for the Thessaloniki region, incorporating both Thessaloniki and Mygdonia basins and adjusting the crustal velocity model. They simulated ground motions for earthquake scenarios with magnitudes between 6.5 and 7. They observed a maximum correlation range of $\approx 64km$ for long periods and $\approx 20km$ for shorter periods. Zolfaghari and Forghani (2024) explored spatial correlation variability using a large database of broadband ground motions simulated for Istanbul. Their simulations included 65 earthquake scenarios and over 7,343 virtual stations, allowing for unprecedented resolution in examining spatial correlation across varying source, path and site conditions. They employed both traditional geo-statistical methods and a non-linear, non-stationary approach, calculating correlation coefficients for every station pair. Their results highlighted that spatial correlation is highly sensitive to source rupture complexity, path characteristics, and local geology. They also observed that station pairs with similar soil conditions and those orientated perpendicular to the rupture tend to exhibit higher correlations, consistent with the findings of Monteiro et al. (2026) showcasing the anisotropic and site-dependent nature of spatial correlation.

## 5 Spatial correlation modelling using different IMs

### 5.1 Analytical formulations

As with same IM spatial correlation, several studies have focused on modelling cross IMs spatial correlation through analytical formulations. These formulations typically use decay functions to describe how the correlation between IMs varies with distance. These analytical approaches are particularly useful, as they provide a simplified and computationally efficient

means of estimating spatial dependencies in the available data. Notable studies adopting this approach include Goda and Hong (2008) and Goda and Atkinson (2009). As shown in Fig. 11, the first study investigated the spatial correlation of $Sa(T)$ responses using earthquake records from California and the Chi-Chi in Taiwan. In contrast, employing similar methods (Goda and Atkinson 2009), focused on data from the K-NET and KiK-net networks in Japan. These studies adapted Pearson's correlation formulation, originally described for the same spectral acceleration periods (Eq. (42)), to focus instead on the spatial correlation of residuals between different periods. The adapted formulation is shown in Eq. (78). Specifically, Goda and Hong (2008) and Goda and Atkinson (2009) proposed different analytical formulations to describe spatial correlations between $Sa(T)$ at different periods. The following equations illustrate their methodologies for modelling these inter-period spatial correlations for total and within-event residuals. The total spatial correlation is given by:

$$\rho_{\delta T}(h, T_i, T_j) = \frac{\rho_{\delta B}(T_i, T_j) \cdot \tau(T_i) \cdot \tau(T_j) + \rho_{\delta W}(h, T_i, T_j) \cdot \phi(T_i) \cdot \phi(T_j)}{\sigma(T_i)\sigma(T_j)} \tag{78}$$

where $\rho_{\delta T}(h, T_i, T_j)$, $\rho_{\delta B}(T_i, T_j)$, and $\rho_{\delta W}(h, T_i, T_j)$ represent the spatial correlation of total, between-event, and within-event residuals, respectively. The corresponding standard deviations are denoted as $\sigma(T_i)$ and $\sigma(T_j)$ for total residuals, $\tau(T_i)$ and $\tau(T_j)$ for between-event residuals, and $\phi(T_i)$ and $\phi(T_j)$ for within-event residuals. Goda and Hong (2008), proposed an alternative expression for $\rho_{\delta T}(h, T_i, T_j)$ that is based on the $[\sigma_d(h, T_i, T_j)]^2$, proposed firstly by Boore et al. (2003), which represents the variance of $[\delta B_k(T_i) + \delta W_{k,m}(T_i)] - [\delta B_k(T_j) + \delta W_{k,n}(T_j)]$ for event $k$ and sites $m$ and $n$. This term was previously described in Sect. (4.1) for $Sa(T)$ at a single $T_i$, but it is now extended to different periods $T_i$ and $T_j$. This alternative expression allows for the calculation of the spatial correlation of total residuals, not depending on between-event or within-event models. The quantity $[\sigma_d(h, T_i, T_j)]^2$ can be written as:

$$\begin{aligned}[\sigma_d(h, T_i, T_j)]^2 &= \mathrm{Var}\left((\delta B_k(T_i) + \delta W_{k,m}(T_i)) - (\delta B_k(T_j) + \delta W_{k,n}(T_j))\right) \\ &= \mathrm{Var}\left(\delta B_k(T_i) - \delta B_k(T_j) + \delta W_{k,m}(T_i) - \delta W_{k,n}(T_j)\right)\end{aligned} \tag{79}$$

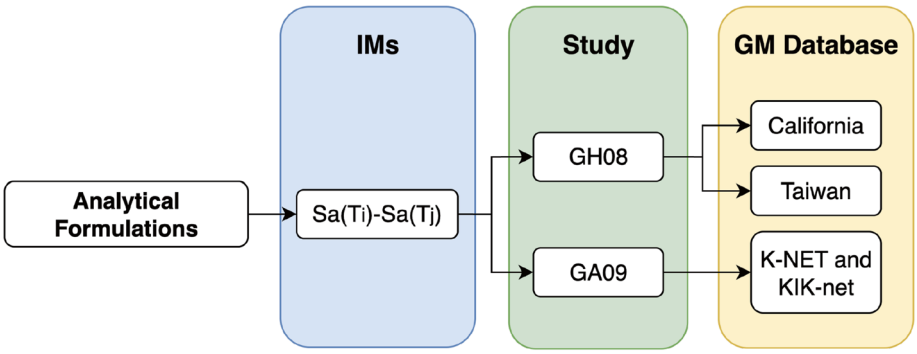


**Fig. 11** Overview of key studies that employed analytical formulations to model spatial correlation for the same IMs. GH08:Goda and Hong (2008), GA09:Goda and Atkinson (2009)

where $h$ represents the distance between sites $m$ and $n$. Because $\delta B$ and $\delta W$ are considered independent, Eq. (79) can be expanded as:

$$\begin{aligned}[\sigma_d(h,T_i,T_j)]^2 &= \mathrm{Var}\left(\delta B_k(T_i) - \delta B_k(T_j)\right) + \mathrm{Var}\left(\delta W_{k,m}(T_i) - \delta W_{k,n}(T_j)\right) \\ &= \tau^2(T_i) + \tau^2(T_j) - 2\cdot\rho_{\delta B}(T_i,T_j)\cdot\tau(T_i)\cdot\tau(T_j) \\ &\quad + \phi^2(T_i) + \phi^2(T_j) - 2\cdot\rho_{\delta W}(h,T_i,T_j)\cdot\phi(T_i)\cdot\phi(T_j) \\ &= \sigma^2(T_i) + \sigma^2(T_j) - 2\cdot\rho_{\delta T}(h,T_i,T_j)\sigma(T_i)\cdot\sigma(T_j)\end{aligned} \tag{80}$$

Therefore, $\rho_{\delta T}(h,T_i,T_j)$ can be rewritten as:

$$\rho_{\delta T}(h,T_i,T_j) = \frac{\sigma^2(T_i) + \sigma^2(T_j) - [\sigma_d(h,T_i,T_j)]^2}{2\cdot\sigma(T_i)\cdot\sigma(T_j)} \tag{81}$$

For between-event correlation, Goda and Atkinson (2009) used the formula presented by Eq. (39) and Goda and Hong (2008), and as mentioned before in Sect. (3.1), due to the lack of data, it was approximated as:

$$\rho_{\delta B}(T_i,T_j) = \rho_{\delta T}(0,T_i,T_j) = \rho_{\delta W}(0,T_i,T_j) \tag{82}$$

which was represented as $\rho^{xy}_{T_i,T_j}$ suggested by Baker and Cornell (2006) in Eq. (32).

For different periods $T_i$ and $T_j$, both studies mentioned before, modelled within-event spatial correlation using the Markov-type screening hypothesis (Journel 1999):

$$\rho_{\delta W}(h,T_i,T_j) \approx \rho^{xy}_{T_i,T_j}\cdot\rho_{\delta W}(h,T_{max}) \tag{83}$$

where $T_{max}$ represents the maximum period between $T_i$ and $T_j$, $\rho^{xy}_{T_i,T_j}$ denotes any non-spatial correlation model for $Sa(T)$ (see Sect. 3) and $\rho_{\delta W}(h,T_{max})$ can be calculated using any spatial correlation model for the same spatial correlation period, $T_{max}$ (see Sect. 4).

With Eq. (82) describing the correlation of between-event residuals and Eq. (83) describing the spatial correlation of within-event residuals, Goda and Hong (2008) and Goda and Atkinson (2009) reformulated Eq. (81), which represents the spatial correlation of total residuals as follows:

$$\rho_{\delta T}(h,T_i,T_j) = \frac{\rho^{xy}_{T_i,T_j}\cdot[\tau(T_i)\cdot\tau(T_j) + \rho_{\delta W}(h,T_{\max})\cdot\phi(T_i)\cdot\phi(T_j)]}{\sigma(T_i)\cdot\sigma(T_j)} \tag{84}$$

### 5.2 Cross-semivariograms

When analysing spatial correlation across multiple ground motion parameters, such as spectral accelerations at different periods or various IMs (e.g., $PGA$, $PGV$, etc.), a multivariate approach is required. The multivariate random field extends the univariate case by representing the random function as a vector, with each component corresponding to a different variable at a given spatial location. In the univariate case, spatial dependence is captured using the semivariogram (see Eq. (14)). In the multivariate setting, the spatial dependence

is of interest not only within each variable but also between different variables at the same location. This is measured using the cross-semivariogram. For instance, let $Z_{IM_i}(x)$ and $Z_{IM_j}(x)$ denote the residuals of the random functions corresponding to two different intensity measures, $IM_i$ and $IM_j$. The cross-semivariogram between these two components is defined as:

$$\gamma_{IM_i,IM_j}(h) = \frac{1}{2}\mathbb{E}[(Z_{IM_i}(x) - Z_{IM_i}(x+h))(Z_{IM_j}(x) - Z_{IM_j}(x+h))] \tag{85}$$

This function captures both spatial variability and the relationship between these two different variables. The empirical cross-semivariogram is then calculated, similar to Eq. (19), as:

$$\gamma_{IM_i,IM_j}(h) = \frac{1}{2N_{i,j}(h)} \sum_{\alpha=1}^{N_{i,j}(h)} [(z_{IM_i}(x_\alpha) - z_{IM_i}(x_\alpha+h))(z_{IM_j}(x_\alpha) - z_{IM_j}(x_\alpha+h))] \tag{86}$$

where $N_{i,j}(h)$ is the number of observation pairs for the two variables at a separation distance $h$ and $\{z_{IM_i}(x_\alpha), z_{IM_i}(x_\alpha+h)\}$ represent the $\alpha^{th}$ data pair for the bin for $i^{th}$ component of the vector IM. Compared to the univariate approach (see Eq. (13)), here the covariance function $C_{IM_i,IM_j}(h)$ is calculated as:

$$\begin{aligned} C_{IM_i,IM_j}(h) =& \text{Cov}\left(Z_{IM_i}(x), Z_{IM_j}(x+h)\right) \\ =& \mathbb{E}\left[(Z_{IM_i}(x) - \mathbb{E}[Z_{IM_i}(x)])(Z_{IM_j}(x+h) - \mathbb{E}[Z_{IM_j}(x)])\right] \end{aligned} \tag{87}$$

where under the stationary assumption, the mean of $Z_{IM}(x)$ is equal the mean of $Z_{IM}(x+h)$. Similar to the relationship mentioned before in Eq. (17), for isotropic and stationary fields, it can be written (Gooverts 1997):

$$C_{IM_i,IM_j}(h) = C_{IM_i,IM_j}(0) - \gamma_{IM_i,IM_j}(h) \tag{88}$$

So the spatial correlation, similar to Eq. (18), but now between two different intensity measures, $IM_i$ and $IM_j$, can be calculated as (Wang and Wenqi 2013):

$$\begin{aligned} \rho_{IM_i,IM_j}(h) &= \frac{C_{IM_i,IM_j}(h)}{\sqrt{C_{IM_i,IM_i}(0) \times C_{IM_j,IM_j}(0)}} \\ &= \frac{C_{IM_i,IM_j}(0)}{\sqrt{C_{IM_i,IM_i}(0) \times C_{IM_j,IM_j}(0)}} - \frac{\gamma_{IM_i,IM_j}(h)}{\sqrt{C_{IM_i,IM_i}(0) \times C_{IM_j,IM_j}(0)}} \end{aligned} \tag{89}$$

The cross-semivariogram matrix, $\Gamma(h)$, Eq. (90), aggregates the semivariograms and cross-semivariograms for all pairs of IMs into a single matrix representation. The covariance matrix $C(h)$, Eq. (91), for the multivariate random field can be derived using the semivariogram matrix, ensuring it remains positive definite, and the correlation matrix $R(h)$, Eq. (92), has the direct correlation coefficients on its main diagonal (when $i = j$), and the off-diagonal elements (when $i \neq j$) represent the cross-correlation coefficients. Note

that $i, j \in \{1, 2, \ldots, N-1, N\}$ represents the different IMs analysed and $\gamma_{IM_i,IM_i}(h)$, $C_{IM_i,IM_i}(h)$, and $\rho_{IM_i,IM_j}(h)$ can be obtained using Eqs. (86), (88), and (89), respectively.

$$\Gamma(h) = [\gamma_{IM_i,IM_j}(h)] = \begin{bmatrix} \gamma_{IM_1,IM_1}(h) & \gamma_{IM_1,IM_2}(h) & \cdots & \gamma_{IM_1,IM_N}(h) \\ & \gamma_{IM_2,IM_2}(h) & \cdots & \gamma_{IM_2,IM_N}(h) \\ & \text{sym} & \ddots & \vdots \\ & & & \gamma_{IM_N,IM_N}(h) \end{bmatrix} \tag{90}$$

$$C(h) = [C_{IM_i,IM_j}(h)] = \begin{bmatrix} C_{IM_1,IM_1}(h) & C_{IM_1,IM_2}(h) & \cdots & C_{IM_1,IM_N}(h) \\ & C_{IM_2,IM_2}(h) & \cdots & C_{IM_2,IM_N}(h) \\ & \text{sym} & \ddots & \vdots \\ & & & C_{IM_N,IM_N}(h) \end{bmatrix} \tag{91}$$

$$R(h) = [\rho_{IM_i,IM_j}(h)] = \begin{bmatrix} \rho_{IM_1,IM_1}(h) & \rho_{IM_1,IM_2}(h) & \cdots & \rho_{IM_1,IM_N}(h) \\ & \rho_{IM_2,IM_2}(h) & \cdots & \rho_{IM_2,IM_N}(h) \\ & \text{sym} & \ddots & \vdots \\ & & & \rho_{IM_N,IM_N}(h) \end{bmatrix} \tag{92}$$

Given that normalised within-event residuals can be reasonably modelled as following a multivariate normal distribution, their spatial variability for a specific earthquake event $k$ is entirely described by their mean and covariance structure. In this context, the mean is assumed to be a vector of zeros, while the covariance matrix, $\Sigma$, captures the spatial dependencies among the $M$ observation sites. For event $k$, the full covariance matrix $\Sigma$ is constructed by assembling the $N \times N$ sub-matrices $C(h)$ (as defined in Eq. (91)) corresponding to all inter-site distances:

$$\Sigma(event \quad k) = \begin{bmatrix} C(h_{11}) & C(h_{12}) & \cdots & C(h_{1M}) \\ C(h_{21}) & C(h_{22}) & \cdots & C(h_{2M}) \\ \vdots & \vdots & \ddots & \vdots \\ C(h_{M1}) & \cdots & \cdots & C(h_{MM}) \end{bmatrix} \tag{93}$$

Maintaining a positive definite covariance matrix is essential for valid spatial modelling, as it ensures the physical plausibility of simulated random fields. Fig. 12 illustrates several studies (e.g., Loth and Baker 2013; Wang and Wenqi 2013; Markhvida et al. 2018; Abbasnejadfard et al. 2020; Wenqi and Ning 2021; Monteiro et al. 2026) that have employed different fitting techniques to fit the cross-semivariogram models and guarantee the positive definiteness of the covariance matrix.

### 5.2.1 Linear model of co-regionalisation

From Fig. 12, it can be seen that Loth and Baker (2013) and Wang and Wenqi (2013) employed a methodology known as the linear model of co-regionalisation (LMC) (Goulard and Voltz 1992). This approach accounts for the effect of multiple spatial scales and assumes that all variables are linear combinations of shared underlying spatial structures (Loth and Baker 2013). The general idea of LMC is to model multivariate spatial variation

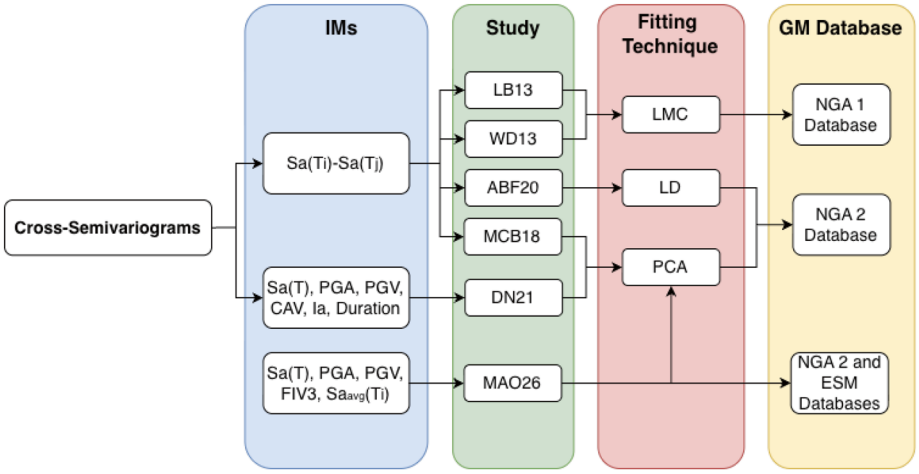


**Fig. 12** Overview of key studies that employed cross-semivariograms to model spatial correlation for different IMs. LB13:Loth and Baker (2013), WD13:Du and Wang (2013), ABF20:Abbasnejadfard et al. (2020), MCB18:Markhvida et al. (2018), DN21:Wenqi and Ning (2021), MAO26:Monteiro et al. (2026), LMC (linear model of co-regionalisation), LD (latent dimensions), PCA (principal component analysis)

by expressing each spatial variable as a linear combination of a common set of orthogonal, second-order stationary random fields. Formally, this starts with a set of $n$ second-order stationary and mutually orthogonal latent random fields $\{Y_r(x), r = 1, \ldots, n\}$, each with its own direct covariance function $C_r(h)$, where $h$ is the spatial lag (see Eq. (17)). Orthogonality implies that the cross-covariances between distinct fields are zero:

$$C_{rr'}(h) = 0 \quad for \quad r \neq r' \tag{94}$$

Let $Z_i(x)$, with $i = 1, \ldots, N$, denote the observed spatial random field associated with the $i^{th}$ IM. Each $Z_i(x)$ is modelled as a linear combination of the shared latent fields:

$$Z_i(x) = \sum_{r=1}^{n} a_{i,r} Y_r(x) \tag{95}$$

where $a_{i,r}$ are scalar coefficients representing the contribution of each latent field $Y_r(x)$ to the observed field $Z_i(x)$. As a result of this construction, the observed fields, $Z_i(x)$, are generally correlated, with their cross-covariances given by:

$$C_{i,j}(h) = \sum_{r=1}^{n} a_{i,r} a_{j,r} C_r(h) \tag{96}$$

This equation shows that the covariance between any two observed variables is a weighted sum of the covariances of the shared latent fields. This enables shared spatial structures to be embedded across multiple observed variables, facilitating multivariate modelling.

To incorporate multiple spatial scales, the standard LMC formulation is extended by associating each latent field with a specific scale $l$. This leads to a multi-scale version of the LMC, where each observed variable $\{Z_i(x)\}$ is modelled as a linear combination of latent fields $Y_r^l(x)$, each indexed by its scale $l$ and component $r$ (Loth and Baker 2013):

$$Z_i(x) = \sum_{l=0}^{L} \sum_{r=1}^{n_l} a_{i,r}^l Y_r^l(x), \quad \forall i = 1, \ldots, N \tag{97}$$

where each $Y_r^l(x)$ is a latent random field at spatial scale $l$ and component $r$, $a_{i,r}^l$ is the coefficient connecting the latent process to the observed variable $Z_i$, $n_l$ is the number of latent components at scale $l$, and $L$ is the total number of scales. Each latent field is assumed to be second-order stationary with zero mean and covariance that satisfies:

$$\begin{aligned} \mathbb{E}[Y_r^l(x)] &= 0, \\ \mathrm{Cov}(Y_r^l(x), Y_{r'}^{l'}(x+h)) &= \begin{cases} \gamma^l(h), & \text{if} \quad r = r' \text{and } l = l' \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{98}$$

This ensures that latent fields are uncorrelated across both component and spatial scales. Analogous to Eq. (96), the multi-scale formulation leads to a decomposition of the semivariogram matrix $\Gamma(h)$, capturing all semivariograms and cross-semivariograms between the observed variables:

$$\Gamma(h) = \sum_{l=0}^{L} B^l \gamma^l(h) \tag{99}$$

where each $\gamma^l(h)$ is an admissible semivariogram function corresponding to the spatial structure $l$, and can be modelled using parametric forms such as exponential, Gaussian, spherical, or nugget models (see Eqs. (21)-(24)). Each term $B^l$ is a positive definite coregionalisation matrix, constructed from the coefficients $a_{i,r}^l$, quantifying how much each structure contributes to the cross-variability of the observed variables. In the case of two IMs, the $B^l$ matrix becomes:

$$B^l = \begin{bmatrix} \sum_{r=1}^{n_l} (a_{1,r}^l)^2 & \sum_{r=1}^{n_l} a_{1,r}^l a_{2,r}^l \\ \sum_{r=1}^{n_l} a_{2,r}^l a_{1,r}^l & \sum_{r=1}^{n_l} (a_{2,r}^l)^2 \end{bmatrix} \tag{100}$$

Thus, the multi-scale LMC provides a flexible framework for modelling multivariate spatial data through a shared set of basic structures, each representing a spatial process operating at a specific scale (Journel and Huijbregts 1979). Loth and Baker (2013) and Wang and Wenqi (2013), used two exponential functions as $\gamma^l(h)$, with two different ranges $b_1$ and $b_2$ as basic models, therefore the cross-semivariogram matrix from Eq. (99) was described as:

$$\Gamma(h) = B^1 \left[1 - \exp\left(\frac{-3h}{b_1}\right)\right] + B^2 \left[1 - \exp\left(\frac{-3h}{b_2}\right)\right] \tag{101}$$

and then the covariance matrix becomes:

$$\begin{aligned} C(h) &= \left[\lim_{h\to\infty} \Gamma(h)\right] - \Gamma(h) \\ &= B^1 \exp\left(-\frac{3h}{b_1}\right) + B^2 \exp\left(-\frac{3h}{b_2}\right) \end{aligned} \tag{102}$$

The total covariance matrix, $\Sigma$, in Eq. (93) is guaranteed to be positive semi-definite regardless of the number of sites considered, if the co-regionalisation matrices $B^l$ are all positive semi-definite (Gooverts 1997).

Loth and Baker (2013) implemented the LMC to simulate cross-spatial correlation in spectral acceleration, $Sa(T)$, at nine different periods, ranging from 0.01 to 10 seconds, considering three co-regionalisation matrices, $B^1$, $B^2$, and $B^3$, obtained using Goulard and Voltz (1992) algorithm, accounting for three components: a short-range structure (20 km) and long-range structure (70 km) and the nugget effect, respectively. The model can be expressed as:

$$R(h) = B^1 \exp\left(\frac{-3h}{20}\right) + B^2 \exp\left(\frac{-3h}{70}\right) + B^3 \zeta_{h=0} \tag{103}$$

where $\zeta_{h=0}$ is 1 when $h = 0$, and 0 otherwise.

Another study to apply the LMC methodology was Wang and Wenqi (2013) to several IMs, such as $PGA$, $PGV$, $Ia$, and $Sa(T)$. Their model for $Sa(T)$ differs significantly from the previous model because it uses data from eleven past earthquakes and not seven, incorporates the impact of site conditions and does not use the same correlation ranges of 0 km, 20 km, and 70 km for the nugget effect, short and long-range, respectively. Instead, it uses a short range of 10 km and omits the nugget effect. Additionally, this study employs the GMM by Campbell and Bozorgnia (2008), in contrast to the one by Boore and Atkinson (2008) used by Loth and Baker (2013). The spatial correlation model was developed for the same nine periods as in the previous study and is given by:

$$R(h) = P^1 \exp\left(\frac{-3h}{10}\right) + P^2 \exp\left(\frac{-3h}{70}\right) \tag{104}$$

For a regionally dependent model, incorporating site conditions, the equation becomes:

$$\begin{aligned} R(h, R_{V_{s30}}) =& P_{SA}^{01} \exp\left(\frac{-3h}{10}\right) + P_{SA}^{02} \exp\left(\frac{-3h}{70}\right) \\ &+ K_{SA}\left(\frac{R_{V_{s30}}}{10}\right)\left[\exp\left(\frac{-3h}{70}\right) - \exp\left(\frac{-3h}{10}\right)\right] \end{aligned} \tag{105}$$

Here, $P^1$ and $P^2$ are co-regionalisation matrices derived from standardising the matrix $B^l$, while $P_{SA}^{01}$, $P_{SA}^{02}$, and $K_{SA}$ are co-regionalisation matrices that account for regional site conditions, representing short-range, long-range, and the influence rate of regional site conditions, respectively.

### 5.2.2 Principal component analysis

The LMC method made it possible to build accurate multivariate semivariogram matrices for IMs, enabling the simultaneous calculation of correlation coefficients for $Sa(T)$ across various periods and locations, as well as the fitting of cross-covariances. However, the challenge of maintaining positive definiteness in matrices as the number of IMs and the number of $n$ locations increase becomes computationally demanding. Addressing this limitation, Markhvida et al. (2018) proposed a framework using principal component analysis (PCA) (Jackson 1991) with geostatistical methods (i.e., semivariograms), providing an efficient solution for co-simulating spatially correlated spectral accelerations across multiple periods in regional seismic analyses. The variables of interest (i.e., the IM residuals computed in Sect. 2.3) are essentially linearly transformed to an orthogonal basis via PCA, which produces principal components-uncorrelated projections onto the new basis. Each of these principal components 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.

With respect to orthogonality with the preceding component, the variance of the first principal component is the largest, followed by the second principal component, which has the second largest variance, and so on. Eqs. (106), and (107) define that linear transformation, where $P$ is the orthogonal linear transformation matrix with the principal component coefficients (i.e., eigenvectors of the covariance matrix), $Z$ is the matrix with the original data (i.e., normalised residuals from several IMs $(Z_{IM_i}, i = 1, \ldots, N)$ and from several locations ($x_1$ to $x_M$)), and $Y$ is the final matrix with the transformed uncorrelated variables (i.e., residuals in the principal component space).

$$PZ = Y \tag{106}$$

$$\begin{bmatrix} p_{1,IM_1} & \cdots & p_{1,IM_N} \\ \vdots & \ddots & \vdots \\ p_{N,IM_1} & \cdots & p_{N,IM_N} \end{bmatrix} \begin{bmatrix} Z_{IM_1}(x_1) & \cdots & Z_{IM_1}(x_M) \\ \vdots & \ddots & \vdots \\ Z_{IM_N}(x_1) & \cdots & Z_{IM_N}(x_M) \end{bmatrix} = \begin{bmatrix} y_1(x_1) & \cdots & y_1(x_M) \\ \vdots & \ddots & \vdots \\ y_N(x_1) & \cdots & y_N(x_M) \end{bmatrix} \tag{107}$$

On the other hand, knowing that $P$ is an orthogonal matrix, $Z$ can be derived from $Y$ in the following way:

$$Z = P^{-1}Y = P^T Y \tag{108}$$

Because principal components are uncorrelated, the simple Eq. (19) can be used to calculate the semivariogram for each component independently, avoiding the need for cross-semivariograms. Markhvida et al. (2018) proposed a simplified model in which only five principal components were used, since they were sufficient to explain at least $95\%$ of the total variance. In this study, Eq. (109) was presented as the best representation of empirical semivariograms using PCA, capturing three types of behaviours:

$$\gamma_k(h) = c_{0k}(I_{nuggect}) + c_{1k}\left(1 - \exp\left(\frac{-3h}{a_{1k}}\right)\right) + c_{2k}\left(1 - \exp\left(\frac{-3h}{a_{2k}}\right)\right) \tag{109}$$

where the first term denotes the nugget effect, where $I_{nuggect} = 0$ if $h = 0$ and $I_{nuggect} = 0$ otherwise. The other two terms represent the short-range and the long-range effects, respectively, using two exponential functions; $k$ represents the $k^{th}$ principal component; and $c_{0k}, c_{1k}, c_{2k}, a_{1k}$, and $a_{2k}$ are regression coefficients. Based on Eqs. (109), and (106), the cross-semivariogram previously described by Eq. (85) can be expressed as (Markhvida et al. 2018):

$$\gamma_{IM_i,IM_j}(h) = C_{IM_i,IM_j}(0) - C_{IM_i,IM_j} = \sum_{k=1}^{N} p_{k,IM_i} p_{k,IM_j} \gamma_k(h) \tag{110}$$

Several factors in the PCA-based analysis can influence the resulting spatial correlations. Wenqi and Ning (2021) investigated this by varying the number of earthquakes, the number of ground motions, the set of IMs considered, and the number of principal components retained. Their results show that the spatial correlation of spectral acceleration at different periods, as well as $CAV$, decreases with increasing period. A similar decreasing trend was observed for $PGA$ and $Ia$. In contrast, significant duration ($D_{s575}$ and $D_{s595}$) exhibited very low and generally negative spatial correlation with most IMs. Furthermore, recognising the gap in models for next-generation IMs, Monteiro et al. (2026) have also developed a cross-spatial model using PCA.

### 5.2.3 Latent dimension method

Although previous studies have examined the multivariate correlation of within-event residuals in earthquake IMs, they have all assumed isotropy in these residuals - an assumption that does not always hold (Garakaninezhad et al. 2017). To address this limitation, Abbasnejadfard et al. (2020) presented a new method that considers anisotropy in spatial correlation of multivariate random fields using the latent dimension latent dimensions (LD) method initially proposed by Genton and Kleiber (2015). This methodology shows how important it is to address anisotropic and non-stationary correlations since conventional isotropic models tend to unrealistic loss estimations and inaccurate resilience assessment (Abbasnejadfard et al. 2021). LD is an efficient approach to construct valid non-separable cross-covariance functions and can be applied to anisotropic multivariate random fields. The key idea is to consider $k$ latent dimensions to represent an $n$-dimensional multivariate random field as a univariate random field in a space $n + k$ dimensions Genton and Kleiber (2015).

## 6 Critical discussion

Three main categories of ground motion correlation modelling were examined in this review, encompassing non-spatial, spatial (same-IM), and spatial (cross-IM) correlations. Comparative analyses were conducted to highlight methodological differences, underlying assumptions, and their implications. The following subsections reflect on these with some critical discussion on some of the benefits, drawbacks and current limitations. A recurring limitation across published correlation models concerns the treatment of uncertainty in the estimated correlation range parameters. Substantial variability in reported range values is

evident across datasets, regions, and modelling approaches, as illustrated by the comparison plots (Figs. 13–14). This variability may be interpreted as reflecting aleatory uncertainty associated with the spatial variability of ground motion residuals, yet most studies rely on single point estimates without explicitly characterising this uncertainty, although studies like Heresi and Miranda (2019) have addressed it. In contrast, differences among published models are more appropriately viewed as epistemic uncertainty arising from data selection, modelling assumptions, and estimation strategies. While point estimates are often adequate for immediate engineering applications, neglecting these distinct sources of uncertainty may lead to overconfident inference when propagated into hazard and risk analyses. Incorporating uncertainty quantification strategies, such as event-wise bootstrap resampling or unified likelihood-based estimation approaches (Ming et al. 2019), represents an important direction for future developments and would enhance the robustness and transparency of ground motion correlation modelling.

### 6.1 Non-spatial correlation modelling

Section (3) reviewed different correlation models developed in the last two decades and Fig. 13 provides an illustrative comparison. Fig. 13(a) compares the correlation between $Sa(T)$ and significant duration ($Ds_{575}$ and $Ds_{595}$) for three models. Although based on

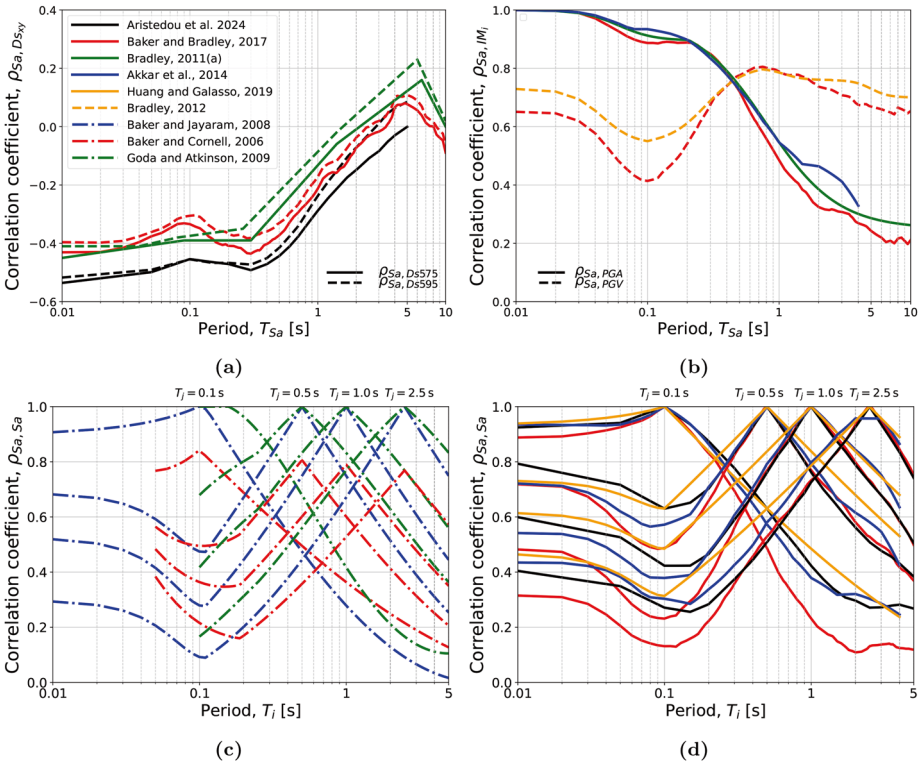


**Fig. 13** Non-spatial correlation and cross-correlation model comparison between **(a)** $Sa(T) - Ds_{575}$ and $Sa(T) - Ds_{595}$, **(b)** $Sa(T) - PGA$ and $Sa(T) - PGV$, **(c)** $Sa(T) - Sa(T)$ using analytical formulations and **(d)** $Sa(T) - Sa(T)$ using various methodologies

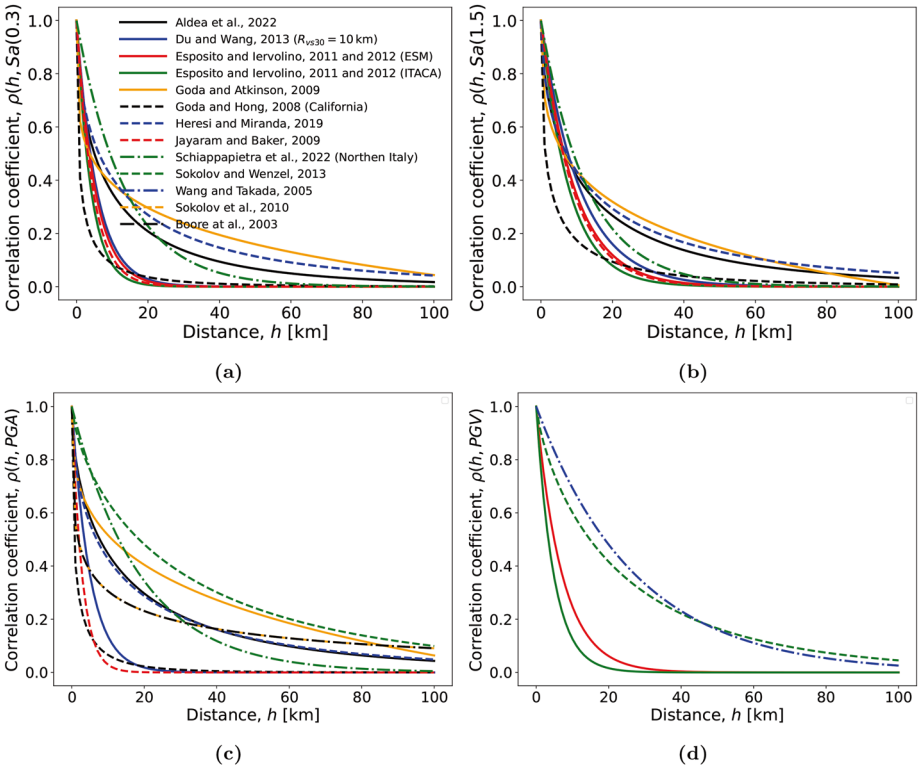


**Fig. 14** Spatial correlation models comparison for (**a**) $Sa(0.3)$, (**b**) $Sa(1.5)$, (**c**) $PGA$, and (**d**) $PGV$

distinct methodologies, all yield broadly consistent trends, with negative correlations at short periods and increasing positive correlations at longer periods. Aristeidou et al. (2024) was observed to be more negatively correlated than the other two available models Bradley (2011a), Baker and Bradley (2017), which was attributed to differences in filtering criteria, with these latter studies including earthquakes of much lower magnitude than the Aristeidou et al. (2024) study. Figure 13(b) compares correlations between $Sa(T)$ and $PGA$ or $PGV$. Despite differences in regional datasets (e.g., NGA-W1, NGA-W2 and European/Middle Eastern), the models show consistent decay of $\rho_{Sa,PGA}$ with period and stable $\rho_{Sa(T),PGV}$ around 0.3 s-1.0s. Given its cross-regional consistency and more recent calibration, the (Baker and Bradley 2017) formulation remains the most broadly applicable. Figures 13(c) and 13(d) show intra-period correlation of spectral acceleration at two periods $T_i$ and $T_j$.

Although all analytical formulations capture the general decay of correlation with increasing separation of periods, they differ in their mathematical structure and treatment of horizontal components. Baker and Cornell (2006) explicitly separate horizontal and vertical components and use trigonometric-logarithmic forms, which can produce sharper changes in correlation for small differences in periods, reflecting a more abrupt decay (Eq. (30)-(33)). In contrast, Baker and Jayaram (2008) (Eq. (34) and (35)) and Huang and Galasso (2019) (Eq. (36)) adopt piecewise or smoothed logarithmic forms, which ensure continuity and smoother variation across periods, particularly when geometric-mean rotations are

used. Mathematically, these differences influence the stability and realism of multivariate spectral acceleration simulations. Ultimately, the choice between these formulations represents two mathematically valid options, each with trade-offs between fidelity to component-specific residuals and smoothness across periods.

The overall behaviour is similar across all studies, although Baker and Bradley (2017) systematically predicts lower correlations for widely separated periods. Subsequent investigations Aristeidou et al. (2024) attributed this to a broader magnitude range in the calibration dataset, implying that such differences stem more from database filtering than fundamental modelling discrepancies (e.g., GMM used). Thus, for applications requiring internally consistent correlations, models constrained to a narrower magnitude range typically utilised in strong shaking, such as Aristeidou et al. (2024), or regionally-specific models such as Akkar et al. (2014), may be more suitable.

### 6.2 Intra-IM spatial correlation modelling

Intra-IM spatial correlation modelling was addressed in Section (4). Over the past two decades, spatial correlation modelling has evolved from analytical and geostatistical formulations toward non-ergodic GMMs, physics-based simulations and Bayesian inference models that better capture spatial variability.

A brief note on the assumptions underlying the earlier analytical and geostatistical approaches is helpful for interpreting the models compared below. Analytical formulations typically rely on simplified structures, often implicitly assuming linear dependence, homogeneity, and stationarity of the underlying spatial field, which make them efficient and easy to apply but limit their ability to represent more complex or scale-dependent behaviour. Geostatistical tools such as semivariograms offer greater flexibility; however, they depend on well-distributed data and on the suitability of the adopted variogram model, irregular sampling, outliers, and non-stationarity trends can all bias the inferred correlation structures. These considerations are particularly relevant when contrasting newer models with analytical or empirical formulations, as differences in underlying assumptions can translate into noticeably different spatial decay patterns.

Figure 15 compares several representative models for $Sa(T)$, $PGA$, and $PGV$. For $Sa(T)$ (Figs. 15(a)-(b)), some models show a noticeably slower decay with distance, maintaining higher correlations (0.20–0.35 at 20 km for $Sa(0.3s)$) compared to older formulations that approach zero. These results imply a broader spatial footprint of ground motion correlation, which may be more realistic in regions with dense seismic instrumentation for crustal earthquakes. However, they also risk overestimating spatial coherence if applied in regions with stronger path variability. For $PGA$ (Fig. 15(c)), the diversity among models is more pronounced. Some approaches retain moderate correlation beyond 40 km, whereas others predict a much faster decay. This variability mainly reflects the large differences among existing $PGA$ spatial correlation models, which results in lower confidence when predicting the behaviour of this IM, as well as differences in data filtering and the definition of inter-site distance, underscoring the sensitivity of $PGA$-based models and the need for regional calibration rather than unquestioned adoption. Finally, the $PGV$ comparison (Fig. 15(d)) reveals that models derived from different databases (e.g., Japanese and European earthquakes) yield completely different correlation coefficients, highlighting the sensitivity of $PGV$-based models and the importance of wisely selecting for the region under study.

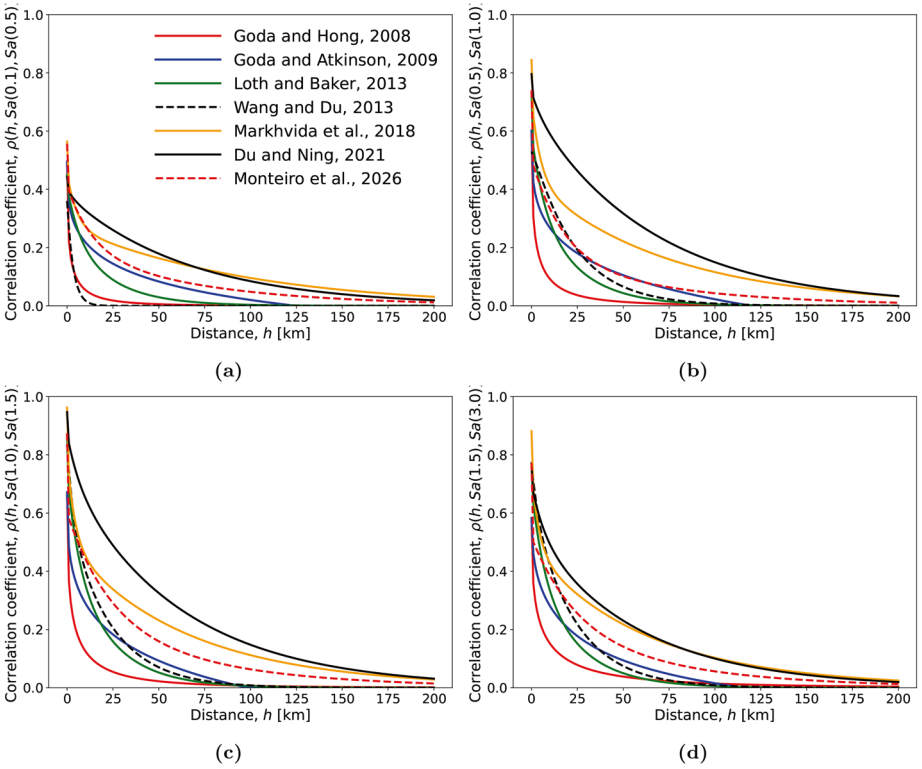


**Fig. 15** Cross-spatial correlation comparison for $Sa(T)$. (**a**) Comparison between $Sa(0.1) - Sa(0.5)$; (**b**) comparison between $Sa(0.5) - Sa(1.0)$; (**c**) comparison between $Sa(1.0) - Sa(1.5)$; (**d**) comparison between $Sa(1.5) - Sa(3.0)$

This comparison between the different spatial correlation models highlights how methodological choices and input databases can significantly influence spatial correlation predictions for various IMs between themselves. Given the variations observed in Fig. 15, careful consideration must be given not only to the IM to adopt, but also to the selection of an appropriate spatial correlation model, particularly in the context of seismic risk assessments.

Section 4.4 discussed the possibility of adopting non-ergodic GMMs for representing spatial correlation. While there are clear advantages to giving further consideration to specific issues like path and site effects, non-ergodic GMMs are not without challenges. One of the main limitations is their dependence on large, spatially dense datasets to reliably constrain regionalised terms and spatial correlation structures, which may not be available. However, the high-resolution datasets becoming increasingly available from PBS (Sect. 4.5) could potentially remedy this current obstacle. The increased complexity of non-ergodic GMMs, particularly when modelling anisotropic and non-stationary behaviours, can also make parameter estimation more difficult and computationally demanding. Moreover, incorporating these models into PSHA frameworks remains non-trivial and is typically more resource-intensive than using conventional ergodic models, compounded by the lack of widely-used software platforms to implement them. As a result, their practical application are currently limited.

Similarly, in the case of PBS, Sect. 4.5 reviewed five different studies that each conveyed the potential of PBS in characterising spatial correlations, which could also, in turn, be useful for non-ergodic GMMs described above. Nevertheless, several challenges remain, such as the high computational demand often required. In addition, the reliability of simulation outputs is closely tied to the quality of input data, such as rupture models, subsurface velocity structures, and site conditions, which are not always well defined. A further limitation is that the underlying models and numerical assumptions can lead to spatial coherence that is stronger and smoother than typically observed in recorded earthquakes, for example, due to simplified velocity structures, limited resolution, or uniform shallow-site representations. Another limitation lies in the validation of simulated spatial correlation patterns, as dense arrays of ground motion records are often scarce, particularly for large events. Thus, while PBS offers a robust framework for exploring complex spatial patterns, the resulting correlation models are currently more suitable for assessing and validating the simulations themselves, rather than for direct use in seismic risk assessments, and should be combined with observational data when possible.

### 6.3 Inter-IM spatial correlation modelling

Section (5) addressed inter-IM spatial correlation modelling, which is probably the least explored topic in correlation modelling, but in practical applications, probably the most needed.

Understanding the assumptions that underpin existing inter-IM correlation models helps contextualise the comparison presented below. Early Pearson- and Markov-type formulations assume linear dependence, stationarity, and a homogeneous distance-decay structure, which makes them straightforward to apply but limits their ability to represent period-dependent or nonlinear interactions between IMs. More flexible approaches, such as cross-semivariograms fitted through the LMC, assume that all inter-IM relationships can be represented as linear combinations of shared basic variogram structures; this requires the spatial field to be jointly second-order stationary on how cross-dependence can vary with scale. PCA-based models rely on decomposing inter-IM covariance into orthogonal modes, meaning that the main directions of variability are assumed to be global and spatially invariant, which can be sensitive to the number of retained components and to the conditioning of the underlying covariance matrix. Latent-dimension formulations embed IMs in a lower-dimension space defined by hidden variables, which provides greater flexibility but introduces identifiability challenges and depends on the choice of regularisation and dimensionality. These differing assumptions across modelling frameworks naturally lead to variations in predicted cross-correlation levels, particularly at short distances and across spectral periods.

Among the models reviewed, it was seen that there are significant variations in how spatial and period-dependent correlations are captured, as shown in Fig. 14. All models predict a monotonic decrease in correlation with distance, but the decay rate and near-field behaviour (within 50 km) depend strongly on the empirical dataset, regional tectonics, and modelling framework. Models with faster decay may underestimate spatial coherence, leading to an underestimation of correlated losses in regions with closely spaced assets. Conversely, models predicting higher short-distance correlations can overestimate spatial dependence, inflating risk estimates if applied outside their calibration range. This near-field sensitivity is particularly important for portfolio-based loss assessments. Additionally, more recent

spatial correlation models developed using PCA (Markhvida et al. 2018; Wenqi and Ning 2021; Monteiro et al. 2026) tend to have higher correlation coefficients, in contrast to the Markov-type models (Goda and Hong 2008; Goda and Atkinson 2009) initially developed.

The spatial co-variation of different IMs reflects underlying physical processes that simultaneously influence multiple IMs. Rupture characteristics such as fault geometry, slip heterogeneity, and directivity generate coherent patterns across a range of frequencies, so IMs of different periods often exhibit correlated variations along preferred directions. Path effects, including attenuation, scattering, and regional crustal structure, impose systematic spatial modifications that affect multiple IMs, though with period-dependent sensitivities. Local site conditions, such as basin-effects, impedance contrasts, and soil nonlinearity, further reinforce short-distance cross-IM correlations. These factors highlight that, beyond inter-station distance, azimuth orientation, regional geology and shared source-path-site effects can strongly influence cross-IM coherence and should be considered when developing or selecting correlation models.

In summary, higher or lower spatial correlation coefficients do not necessarily imply better physical representation but may instead reflect artefacts of underlying GMM fitting, variability, filtering criteria, and other aspects. While there is general agreement on the qualitative nature of spatial correlation decay, significant discrepancies exist among the specific models. Analysts must pay attention to these, possibly referring to Fig. 12, when selecting the most suitable model. Additionally, there remain opportunities to move beyond the current predominant reliance on site-to-site distance as the only predictor of inter-site correlation. Several studies have suggested incorporating additional explanatory parameters, such as site condition contrasts, shared source effects, and regional path characteristics, into correlation models, which could enhance their realism and predictive capacity (e.g., Bodenmann et al. 2023). Likewise, integrating site, path and source effects in a more consistent and scalable manner within global models represents a promising direction for future research.

## 7 Summary

This paper has reviewed over 45 distinct models available in the literature for characterising the correlation between different ground motion intensity measure (IM) types. These encompass both intra-site (non-spatial) and inter-site (spatial) domains, in addition to same IM and cross IM models. The approaches fit Pearson correlation coefficients to observed data using available ground motion databases and characterise the correlation models via analytical functions or other geostatistical techniques, such as semivariograms, depending on the quantity. Fitting methods included the linear model of coregionalisation, and dimensionality reduction approaches like principal component analysis. Other strategies to quantify ground motion correlations both spatially and non-spatially include physics-based simulations and non-ergodic ground motion models. This wide array of modelling techniques, along with the different ground motion datasets used for calibration and validation, highlights the plethora of options available to analysts when performing seismic hazard and risk assessments.

The review provided here aimed to give an overview of the state-of-the-art, provide a critical comparison between models and serve as a valuable resource for future research-

ers and practitioners aiming to select appropriate models for both spatial and non-spatial correlation of IMs. It emphasised the importance of aligning model selection with regional characteristics, the nature of the IMs under consideration, the available ground motion database, and the specific engineering or risk assessment applications. Furthermore, an online repository of these models has been created and shared on GitHub, where analysts can browse and evaluate them, also contributing to their expansion with future implementations.

Special emphasis was made to scrutinise correlation models available for so-called next generation IMs, whether they were direct models, indirect calculations, or yet to be established. This was because improvements have been made in seismic vulnerability modelling with increasing emphasis on more sophisticated IMs, such as $FIV3$ and $Sa_{avg}(T)$, that better characterise structural response of the built environment, helping the development of more accurate vulnerability and fragility models. However, correlation modelling on the seismic hazard side has not kept pace with this evolution. In particular, it was found that there is a general lack of well-established models to quantify spatial and cross-spatial correlation involving these newer IMs. Recent work by Monteiro et al. (2026) has begun to address this, presenting a cross-spatial model for several next-generation IMs and interactions with traditional ones. This gap is particularly relevant in the context of regional risk assessment (e.g., Heresi and Miranda 2023), where different IMs may be used for different structure types or asset classes across a distributed portfolio. By addressing these gaps, the correlation modelling of IMs can develop into more accurate, flexible, and operationally useful models for both non-spatial and spatial correlation of ground motions, enabling more accurate regional seismic assessments.

**Acknowledgments** The authors are thankful for the comments and discussions with Graeme Weatherill, Sinan Akkar and another anonymous review that greatly helped improve this work.

**Author contributions** All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Vitor A. Monteiro. The first draft of the manuscript was written by Vitor A. Monteiro and Gerard J. O'Reilly commented and edited on subsequent versions of the manuscript. All authors read and approved the final manuscript.

**Funding** 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.

**Data availability** The several studies reviewed and comparatively discussed here have been made available via the following GitHub repository: https://github.com/vitorazevedomonteiro/correlation-models.git.

**Code availability** The correlation models reviewed and comparatively discussed here have been made available via the following GitHub repository: https://github.com/vitorazevedomonteiro/correlation-models.git

## Declarations

**Conflicts of interest** The authors have no conflicts of interest to declare that are relevant to the content of this article.

## References


Abbasnejadfard M, Bastami M, Fallah A. Investigation of anisotropic spatial correlations of intra-event residuals of multiple earthquake intensity measures using latent dimensions method. Geophys J International, 222: 1449–1469, 8 2020. ISSN 0956-540X. https://doi.org/10.1093/gji/ggaa255

Abbasnejadfard M, Bastami M, Fallah A, Garakaninezhad A. Analyzing the effect of anisotropic spatial correlations of earthquake intensity measures on the result of seismic risk and resilience assessment of the portfolio of buildings and infrastructure systems. Bull Earthq Eng, 19: 5791–5817, 11 2021. ISSN 1570-761X. https://doi.org/10.1007/s10518-021-01203-z

Abrahamson NA, Youngs RR. A stable algorithm for regression analyses using the random effects model. Bull Seismolog Soc Ame, 82:505–510, 2 1992. ISSN 1943-3573. https://doi.org/10.1785/BSSA0820010505

Akkar S, Sandıkkaya MA, Ay BÖ. Compatible ground-motion prediction equations for damping scaling factors and vertical-to-horizontal spectral amplitude ratios for the broader europe region. Bull Earthq Eng, 12:517–547, 2 2014. ISSN 1570-761X. https://doi.org/10.1007/s10518-013-9537-1

Akkar S, Sandıkkaya MA, Şenyurt M, Sisi AA, Ay BÖ, Traversa P, Douglas J, Cotton F, Luzi L, Hernandez B, Godey S. Reference database for seismic ground-motion in europe (resorce). Bull Earthq Eng, 12:311–339, 2 2014. ISSN 1570-761X. https://doi.org/10.1007/s10518-013-9506-8

Aldea S, Heresi P, Pastén C. Within-event spatial correlation of peak ground acceleration and spectral pseudo-acceleration ordinates in the Chilean subduction zone. Earthq Eng Struc Dyn, 51:2575–2590, 9 2022. ISSN 0098-8847. https://doi.org/10.1002/eqe.3674

Ang AHS, Tang WH. Probability concepts in engineering: emphasis on applications in civil and environmental engineering. John Wiley & Sons, 2007

Annaka T, Yamazaki F, Katahira F. Proposal of peak ground velocity and response spectra based on jma 87 type accelerometer records. 24th JSCE Earthquake Engineering Symposium, Vol. 1:161–164, 1997

Aristeidou S, Shahnazaryan D, O'Reilly GJ. Correlation models for next-generation amplitude and cumulative intensity measures using artificial neural networks. Earthq Spectra, 10 2024. ISSN 8755–2930. https://doi.org/10.1177/87552930241270563

Atik LA, Abrahamson N. Nonlinear site response effects on the standard deviations of predicted ground motions. Bull Seismolog Soc Ame, 100:1288–1292, 6 2010. ISSN 0037-1106. https://doi.org/10.1785/0120090154

Baker JW, Bradley BA. Intensity measure correlations observed in the nga-west2 database, and dependence of correlations on rupture and site parameters. Earthq Spectra, 33:145–156, 2 2017. ISSN 8755-2930. https://doi.org/10.1193/060716eqs095m

Baker JW, Jayaram N. Correlation of spectral acceleration values from nga ground motion models. Earthq Spectra, 24:299–317, 2008. ISSN 87552930. https://doi.org/10.1193/1.2857544

Baker JW, Lee C. An improved algorithm for selecting ground motions to match a conditional spectrum. J Earthq Eng, 22:708–723, 4 2018. ISSN 1363-2469. https://doi.org/10.1080/13632469.2016.1264334

Baker JW, Cornell CA. Correlation of response spectral values for multicomponent ground motions. Bull Seismolog Soc Ame, 96:215–227, 2 2006. ISSN 0037-1106. https://doi.org/10.1785/0120050060

Bodenmann L, Baker JW, Stojadinović B. Accounting for path and site effects in spatial ground-motion correlation models using bayesian inference. Nat Hazards Earth Sys Sci, 23:2387–2402, 7 2023. ISSN 1684-9981. https://doi.org/10.5194/nhess-23-2387-2023

Boore DM, Atkinson GM. Ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods between 0.01 s and 10.0 s. Earthq Spectra, 24:99–138, 2008. ISSN 87552930. https://doi.org/10.1193/1.2830434

Boore DM, Gibbs JF, Joyner WB, Tinsley JC, Ponti DJ. Estimated ground motion from the 1994 northridge, california, earthquake at the site of the interstate 10 and la cienega boulevard bridge collapse, west los angeles, california. Bull Seismolog Soc Ame, 93:2737–2751, 12 2003. ISSN 0037-1106. https://doi.org/10.1785/0120020197

Boore DM. Orientation-independent, nongeometric-mean measures of seismic intensity from two horizontal components of motion. Bull Seismolog Soc Ame, 100:1830–1835, 8 2010. ISSN 0037-1106. https://doi.org/10.1785/0120090400

Bradley BA. A generalized conditional intensity measure approach and holistic ground-motion selection. Earthq Eng Struc Dyn, 39:1321–1342, 10 2010. ISSN 0098-8847. https://doi.org/10.1002/eqe.995

Bradley BA. Correlation of significant duration with amplitude and cumulative intensity measures and its use in ground motion selection. J Earthq Eng, 15:809–832, 7 2011a. ISSN 13632469. https://doi.org/10.1080/13632469.2011.557140

Bradley BA. Empirical correlation of pga, spectral accelerations and spectrum intensities from active shallow crustal earthquakes. Earthq Eng Struc Dyn, 40:1707–1721, 12 2011b. ISSN 0098–8847. https://doi.org/10.1002/eqe.1110

Bradley BA. Empirical correlations between peak ground velocity and spectrum-based intensity measures. Earthq Spectra, 28:17–35, 2 2012. ISSN 8755-2930. https://doi.org/10.1193/1.3675582
Bradley BA. Site-specific and spatially-distributed ground-motion intensity estimation in the 2010–2011 canterbury earthquakes. Soil Dyn Earthq Eng, 61-62:83–91, 6 2014. ISSN 02677261. https://doi.org/10.1016/j.soildyn.2014.01.025
Campbell KW, Bozorgnia Y. Nga ground motion model for the geometric mean horizontal component of pga, pgv, pgd and 5% damped linear elastic response spectra for periods ranging from 0.01 to 10 s. Earthq Spectra, 24:139–171, 2 2008. ISSN 8755-2930. https://doi.org/10.1193/1.2857546
Chen Y, Baker JW. Spatial correlations in cybershake physics-based ground-motion simulations. Bull Seismolog Soc Ame, 109:2447–2458, 12 2019. ISSN 0037-1106. https://doi.org/10.1785/0120190065
Chen Y, Bradley BA, Baker JW. Nonstationary spatial correlation in New Zealand strong ground-motion data. Earthq EngStruc Dyn, 50:3421–3440, 10 2021. ISSN 0098-8847. https://doi.org/10.1002/eqe.3516
Chiou B, Darragh R, Gregor N, Silva W. Nga project strong-motion database. Earthq Spectra, 24:23–44, 2 2008. ISSN 8755-2930. https://doi.org/10.1193/1.2894831
Chiou BS-J, Youngs RR. Update of the chiou and youngs nga model for the average horizontal component of peak ground motion and response spectra. Earthq Spectra, 30:1117–1153, 8 2014. ISSN 8755-2930. https://doi.org/10.1193/072813EQS219M
Chiou B-J, Youngs RR. An nga model for the average horizontal component of peak ground motion and response spectra. Earthq Spectra, 24:173–215, 2 2008. ISSN 8755-2930. https://doi.org/10.1193/1.2894832
Cimellaro GP. Correlation in spectral accelerations for earthquakes in europe. Earthq EngStruc Dyn, 42:623–633, 4 2013. ISSN 0098-8847. https://doi.org/10.1002/eqe.2248
Cressie N, Hawkins DM. Robust estimation of the variogram: I. J Inter Associat Math Geology, 12:115–125, 4 1980. ISSN 0020-5958. https://doi.org/10.1007/BF01035243
Cressie NAC. Statistics for spatial data. Wiley, 9 1993. ISBN 9780471002550. https://doi.org/10.1002/9781119115151
Dávalos H, Miranda E. Filtered incremental velocity: a novel approach in intensity measures for seismic collapse estimation. Earthq EngStruc Dyn, 48:1384–1405, 10 2019. ISSN 0098-8847. https://doi.org/10.1002/eqe.3205
Du W, Wang G. Intra-event spatial correlations for cumulative absolute velocity, arias intensity, and spectral accelerations based on regional site conditions. Bull Seismolog Soc Ame, 103:1117–1129, 4 2013. ISSN 0037-1106. https://doi.org/10.1785/0120120185
Esposito S, Iervolino I. PGA and PGV spatial correlation models based on European multievent datasets. Bull Seismolog Soc Ame, 101:2532–2541, 10 2011. ISSN 0037-1106. https://doi.org/10.1785/0120110117
Esposito S, Iervolino I. Spatial correlation of spectral acceleration in European data. Bull Seismolog Soc Ame, 102:2781–2788, 12 2012. ISSN 0037-1106. https://doi.org/10.1785/0120120068
Foulser-Piggott, Stafford PJ. A predictive model for arias intensity at multiple sites and consideration of spatial correlations. Earthq Eng Struc Dyn, 41:431–451, 3 2012. ISSN 0098-8847. https://doi.org/10.1002/eqe.1137
Garakaninezhad A, Bastami M, Soghrat MR. Spatial correlation for horizontal and vertical components of acceleration from northern Iran seismic events. J Seismology, 21:1505–1516, 11 2017. ISSN 1573157X. https://doi.org/10.1007/s10950-017-9679-8
Genton MG, Kleiber W. Cross-covariance functions for multivariate geostatistics. Statis Sci, 30, 5 2015. ISSN 0883–4237. https://doi.org/10.1214/14-STS487
Goda K, Atkinson GM. Probabilistic characterization of spatially correlated response spectra for earthquakes in Japan. Bull Seismolog Soc Ame, 99:3003–3020, 10 2009. ISSN 0037-1106. https://doi.org/10.1785/0120090007
Goda K, Hong HP. Spatial correlation of peak ground motions and response spectra. Bull Seismolog Soc Ame, 98:354–365, 2 2008. ISSN 0037-1106. https://doi.org/10.1785/0120070078
Gooverts P. Geostatistics for natural resources evaluation. Oxford University Press, 1997
Goulard M, Voltz M. Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. Math Geology, 24:269–286, 4 1992. ISSN 0882-8121. https://doi.org/10.1007/BF00893750
Heresi P, Miranda E. Uncertainty in intraevent spatial correlation of elastic pseudo-acceleration spectral ordinates. Bull Earthq Eng, 17:1099–1115, 3 2019. ISSN 1570-761X. https://doi.org/10.1007/s10518-018-0506-6
Heresi P, Miranda E. Intensity measures for regional seismic risk assessment of low-rise wood-frame residential construction. J Struct Eng, 147, 1 2021. ISSN 0733–9445. https://doi.org/10.1061/(ASCE)ST.1943-541X.0002859
Heresi P, Miranda E. Rpbee: performance-based earthquake engineering on a regional scale. Earthq Spectra, 39:1328–1351, 8 2023. ISSN 8755-2930. https://doi.org/10.1177/87552930231179491

Huang C, Galasso C. Ground-motion intensity measure correlations observed in Italian strong-motion records. Earthq Eng Struc Dyn, 48:1634–1660, 12 2019. ISSN 0098-8847. https://doi.org/10.1002/eqe.3216

Husid LR. Caracterısticas de terremotos, análisis general. Revista del IDIEM, 8:21–42, 1969

Infantino M, Smerzini C, Lin J. Spatial correlation of broadband ground motions from physics-based numerical simulations. Earthq Eng Struc Dyn, 50:2575–2594, 8 2021. ISSN 0098-8847. https://doi.org/10.1002/eqe.3461

Jackson JE. A User's Guide to principal components. Wiley, 3 1991. ISBN 9780471622673. https://doi.org/10.1002/0471725331

Jayaram N, Baker JW. Correlation model for spatially distributed ground-motion intensities. Earthq Eng Struc Dyn, 38:1687–1708, 12 2009. ISSN 0098-8847. https://doi.org/10.1002/eqe.922

Journel AG. Markov models for cross-covariances. Math Geology, 31:1019–1021, 11 1999. ISSN 08828121. https://doi.org/10.1023/A:1007565316113

Journel AG, Huijbregts CJ. Mining geostatistics. academic press: London; new york, 1978. Mineralog Magazine, 43:563–564, 12 1979. ISSN 0026-461X. https://doi.org/10.1180/minmag.1979.043.328.34

Kohrangi M, Bazzurro P, Vamvatsikos D, Spillatura A. Conditional spectrum-based ground motion record selection using average spectral acceleration. Earthq Eng Struc Dyn, 46:1667–1685, 8 2017. ISSN 0098-8847. https://doi.org/10.1002/eqe.2876

Kuehn NM, Abrahamson NA. Spatial correlations of ground motion for non-ergodic seismic hazard analysis. Earthq Eng Struc Dyn, 49:4–23, 1 2020. ISSN 0098-8847. https://doi.org/10.1002/eqe.3221

Lanzano G, Pacor F, Luzi L, D'Amico M, Puglia R, Felicetta C. Systematic source, path and site effects on ground motion variability: the case study of northern italy. Bull Earthq Eng, 15:4563–4583, 11 2017. ISSN 1570-761X. https://doi.org/10.1007/s10518-017-0170-2

Lavrentiadis G, Abrahamson NA, Nicolas KM, Bozorgnia Y, Goulet CA, Babič A, Macedo J, Dolšek M, Gregor N, Kottke AR, Lacour M, Liu C, Meng X, Phung V-B, Sung C-H, Walling M. Overview and introduction to development of non-ergodic earthquake ground-motion models. Bull Earthq Eng, 21:5121–5150, 9 2023. ISSN 1570-761X. https://doi.org/10.1007/s10518-022-01485-x

Lin J, Smerzini C. Variability of physics-based simulated ground motions in thessaloniki urban area and its implications for seismic risk assessment. Frontiers Earth Sci, 10, 8 2022. ISSN 2296–6463. https://doi.org/10.3389/feart.2022.951781

Lin T, Haselton CB, Baker JW. Conditional spectrum-based ground motion selection. part i: hazard consistency for risk-based assessments. Earthq Eng Struc Dyn, 42:1847–1865, 10 2013. ISSN 0098-8847. https://doi.org/10.1002/eqe.2301

Liu C, Macedo J, Kuehn N. Spatial correlation of systematic effects of non-ergodic ground motion models in the ridgecrest area. Bull Earthq Eng, 21:5319–5345, 9 2023. ISSN 1570-761X. https://doi.org/10.1007/s10518-022-01441-9

Loth C, Baker JW. A spatial cross-correlation model of spectral accelerations at multiple periods. Earthq Eng Struc Dyn, 42:397–417, 3 2013. ISSN 0098-8847. https://doi.org/10.1002/eqe.2212

Markhvida M, Ceferino L, Baker JW. Modeling spatially correlated spectral accelerations at multiple periods using principal component analysis and geostatistics. Earthq Eng Struc Dyn, 47:1107–1123, 4 2018. ISSN 0098-8847. https://doi.org/10.1002/eqe.3007

Midorikawa S, Ohtake Y. Attenuation relationships of peak ground acceleration and velocity considering attenuation characteristics for shallow and deeper earthquakes. *9th Japan Earthquake Engineering Symposium*, Vol. 3:E085–E090, 2002

Ming D, Huang C, Peters GW, Galasso C. An advanced estimation algorithm for ground-motion models with spatial correlation. Bull Seismolog Soc Ame, 109:541–566, 4 2019. ISSN 0037-1106. https://doi.org/10.1785/0120180215

Ming-Yang X, Da-Gang L An extended generalized conditional intensity measure method for aftershock ground motion selection. Earthq Eng Struc Dyn, 53:1509–1536, 4 2024. ISSN 0098-8847. https://doi.org/10.1002/eqe.4082

Monteiro VA, Aristeidou S, O'Reilly GJ. Spatial cross-correlation models for next-generation amplitude and cumulative intensity measures (under review). 2026

Morikawa N, Kanno T, Narita A, Fujiwara H, Okumura T, Fukushima Y, Guerpinar A. Strong motion uncertainty determined from observed records by dense network in Japan. J Seismology, 12:529–546, 10 2008. ISSN 1383-4649. https://doi.org/10.1007/s10950-008-9106-2

Paciorek CJ, Schervish MJ. Spatial modelling using a new class of nonstationary covariance functions. Environmetrics, 17:483–506, 8 2006. ISSN 1180-4009. https://doi.org/10.1002/env.785

Papadopoulos AN, Kohrangi M, Bazzurro P. Correlation of spectral acceleration values of mainshock-aftershock ground motion pairs. Earthq Spectra, 35:39–60, 2 2019. ISSN 8755-2930. https://doi.org/10.1193/020518EQS033M

Park J, Bazzurro P, Baker JW. Modeling spatial correlation of ground motion intensity measures for regional seismic hazard and portfolio loss estimation. Appl Statist Probability Civil Eng, pages 1–8, 2007

Pretell R, Brandenberg SJ, Stewart JP. Ground motion intensity measures at liquefaction field case history sites. Technical report, B. John Garrick Institute for the Risk Sciences, UCLA, October 2024. URL https://doi.org/10.34948/N35K59. GIRS-2024-02 report

Rekoske JM, Thompson EM, Moschetti MP, Hearne MG, Aagaard BT, Parker GA. The 2019 ridgecrest, california, earthquake sequence ground motions: processed records and derived intensity metrics. Seismolog Res Lett, 91:2010–2023, 7 2020. ISSN 0895–0695. https://doi.org/10.1785/0220190292

Schiappapietra E, Stripajová S, Pažák P, Douglas J, Trendafiloski G. Exploring the impact of spatial correlations of earthquake ground motions in the catastrophe modelling process: a case study for italy. Bull Earthq Eng, 20:5747–5773, 9 2022. ISSN 1570-761X. https://doi.org/10.1007/s10518-022-01413-z

Schiappapietra E, Smerzini C. Spatial correlation of broadband earthquake ground motion in norcia (central Italy) from physics-based simulations. Bull Earthq Eng, 19:4693–4717, 9 2021. ISSN 1570-761X. https://doi.org/10.1007/s10518-021-01160-7

Sgobba S, Lanzano G, Pacor F, Puglia R, D'Amico M, Felicetta C, Luzi L. Spatial correlation model of systematic site and path effects for ground-motion fields in northern italy. Bull Seismolog Soc Ame, 109:1419–1434, 8 2019. ISSN 0037-1106. https://doi.org/10.1785/0120180209

Shahnazaryan D, O'Reilly GJ. Next-generation non-linear and collapse prediction models for short- to long-period systems via machine learning methods. Eng Struc, 306:117801, 5 2024. ISSN 01410296. https://doi.org/10.1016/j.engstruct.2024.117801

Sokolov V, Wenzel F. Further analysis of the influence of site conditions and earthquake magnitude on ground-motion within-earthquake correlation: analysis of PGA and pgv data from the k-net and the kik-net (japan) networks. Bull Earthq Eng, 11:1909–1926, 12 2013. ISSN 1570-761X. https://doi.org/10.1007/s10518-013-9493-9

Sokolov V, Wenzel F, Jean W-Y, Wen K-L. Uncertainty and spatial correlation of earthquake ground motion in Taiwan. Terrestrial, Atmospheric Oceanic Sci, 21:905, 2010. ISSN 1017-0839. https://doi.org/10.3319/TAO.2010.05.03.01(T)

Stafford PJ, Zurek BD, Ntinalexis M, Bommer JJ. Extensions to the groningen ground-motion model for seismic risk calculations: component-to-component variability and spatial correlation. Bull Earthq Eng, 17:4417–4439, 8 2019. ISSN 1570-761X. https://doi.org/10.1007/s10518-018-0425-6

Tarbali K, Bradley BA, Baker JW. Effect of near-fault directivity pulses on ground-motion intensity measure correlations from the nga-west2 data set. Earthq Spectra, 39:2263–2280, 11 2023. ISSN 19448201. https://doi.org/10.1177/87552930231199059

Wang G, Wenqi D Spatial cross-correlation models for vector intensity measures (PGA, ia, PGV, and SAs) considering regional site conditions. Bull Seismolog Soc Ame, 103:3189–3204, 12 2013. ISSN 00371106. https://doi.org/10.1785/0120130061

Wang M, Takada T. Macrospatial correlation model of seismic ground motions. Earthq Spectra, 21:1137–1156, 11 2005. ISSN 8755-2930. https://doi.org/10.1193/1.2083887

Weatherill GA, Silva V, Crowley H, Bazzurro P. Exploring the impact of spatial correlations and uncertainties for portfolio analysis in probabilistic seismic loss estimation. Bull Earthq Eng, 13:957–981, 4 2015. ISSN 1570-761X. https://doi.org/10.1007/s10518-015-9730-5

Weatherill G, Esposito S, Iervolino I, Franchin P, Cavalieri F. Framework for seismic hazard analysis of spatially distributed systems, pp 57–88. Springer Netherlands, Dordrecht, 2014. ISBN 978-94-017-8835-9. 10.1007/978-94-017-8835-9_3. URL https://doi.org/10.1007/978-94-017-8835-9_3

Weller ZD. Sptest: an r package implementing nonparametric tests of isotropy. J Statis Soft, 83, 2018. ISSN 1548-7660. https://doi.org/10.18637/jss.v083.i04

Weller ZD, Hoeting JA. A review of nonparametric hypothesis tests of isotropy properties in spatial data. Statis Sci, 31, 8 2016. ISSN 0883-4237. https://doi.org/10.1214/16-STS547

Wenqi D, Ning C-L. 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. Earthq Spectra, 37:486–504, 2 2021. ISSN 8755-2930. https://doi.org/10.1177/8755293020952442

Zhu R-G, Da-Gang L, Xiao-Hui Y, Wang G-Y. Conditional mean spectrum of aftershocks. Bull Seismolog Soc Ame, 7 2017. ISSN 0037-1106. https://doi.org/10.1785/0120160254

Zolfaghari MR, Forghani M. Spatial correlation assessment of multiple earthquake intensity measures using physics-based simulated ground motions. Sci Rep, 14:21235, 9 2024. ISSN 2045-2322. https://doi.org/10.1038/s41598-024-72241-1