Earthquake Spectra

RESEARCH PAPER

Notes on Spatial Correlation for Average Spectral Acceleration: Direct and Indirect Approaches

Vitor A. Monteiro | Gerard J. O'Reilly

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

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

Received: 27 October 2025 | **Revised:** 19 December 2025 | **Accepted:** 12 January 2026

Keywords: average spectral acceleration | intensity measures | seismic risk | spatial correlation

ABSTRACT

When seismic risk assessments are performed on a regional scale, simulating representative ground motion fields (GMFs) at several locations over the area of interest is a key step. This is in terms of the intensity of shaking at each single site location with respect to the considered earthquake rupture scenario and also the coherence of the simulated shaking intensity values between each site location. For this, spatial correlation models are crucial and have received much attention in recent years. They have been developed mainly for well-known intensity measures (IMs) such as peak ground acceleration and spectral acceleration, $\mathrm{Sa}(T)$. For next-generation IMs, such as average spectral acceleration, $\mathrm{Sa}_{\mathrm{avg}}(T)$, there are no spatial correlation models specifically developed for this IM, inhibiting its application in practical regional applications. A recent proposal has developed an indirect formulation for this correlation model, capitalising on the fundamental definition of $\mathrm{Sa}_{\mathrm{avg}}(T)$ and the availability of other models that can be used to infer it. In this study, we compare the indirect approach with a recently developed direct spatial correlation model for $\mathrm{Sa}_{\mathrm{avg}}(T)$. Results show that both methods exhibit very similar trends, with the indirect approach closely matching the direct formulation. Analysis of total and within-event residuals indicates that total-residual correlations approach an asymptotic value of around 0.2, while lack of care when generating GMFs using an indirect method can overestimate the IM by up to two orders of magnitude compared to the observed values. These findings highlight the need for care when modelling spatial correlation of $\mathrm{Sa}_{\mathrm{avg}}(T)$, as well as potential unintended pitfalls in routine applications when applying an indirect modelling approach.

1 | Introduction

In performance-based earthquake engineering (PBEE), regional seismic assessment plays a crucial role in shaping public policy, informing decision makers, and guiding infrastructure retrofitting strategies (Heresi and Miranda 2023). An essential component of regional seismic assessment is evaluating the broader impacts under specific earthquake rupture scenarios to examine their propagation. To do this, spatially consistent ground motion fields (GMFs) are generated and represent the distribution of expected ground shaking intensity over a geographic area during a given earthquake rupture scenario. These GMFs must not only preserve site-specific means and variances of expected intensity measures (IMs), but also accurately capture the spatial correlation of ground motion intensities between different site locations (Park et al. 2007).

Over the past two decades, considerable progress has been made in developing models for the spatial correlation of commonly used IMs, such as peak ground acceleration (PGA) and spectral acceleration at individual periods, $Sa(T)$. For example, some studies [e.g., (Boore 2003; Sokolov et al. 2010; Esposito and Iervolino 2011)] have developed models for PGA at different site locations. Other studies [e.g., (Goda and Hong 2008; Heresi and Miranda 2019; Aldea et al. 2022)] have developed models for the spectral acceleration at the same period of vibration T between sites, and even fewer studies [e.g., (Goda and Atkinson 2009; Loth and Baker 2013; Markhvida et al. 2018)] have developed correlation models for when the period T differs at two site locations. These former examples are generally referred to as intra-IM spatial correlation models because they map the same IM spatially, whereas the latter is referred to herein as an inter-IM spatial correlation model because it maps different IMs spatially. A more detailed critical review of these models and their development is available in Monteiro and O'Reilly (2026).

Notwithstanding this continued development of spatial correlation modelling, emerging research has pointed out the limitations of using legacy IMs such as PGA or other period-specific IMs like $Sa(T)$ when evaluating individual or collective seismic vulnerability. Specifically, studies have shown that these IMs may lead to less efficient or biased predictions of structural response [e.g., (O'Reilly 2021a)]. In this context, average spectral acceleration, denoted as $Sa_{avg}(T)$, has been increasingly advocated as a more robust IM. It is defined as the geometric mean of spectral accelerations over a range of periods and offers improved efficiency, sufficiency and predictability as an IM in collapse risk and economic loss estimation [e.g., (Eads et al. 2015; Kohrangi et al. 2017)]. Its superior performance has been demonstrated across various structural typologies and for different types of ground motions. As a result, $Sa_{avg}(T)$ is now widely employed in advanced seismic hazard and vulnerability analyses [e.g., (Kazantzi and Vamvatsikos 2015; O'Reilly 2021a, 2021b; Nafeh and O'Reilly 2024)]. Recent studies have developed ground motion models (GMMs) to enable its prediction and correlation to other IMs [e.g., (Kohrangi et al. 2018; Dávalos and Miranda 2021; Aristeidou et al. 2024, 2025)] for site-specific analysis. However, with available spatial correlation models currently not so plentiful, its use as an IM is often limited to site-specific analyses and somewhat inhibited in regional applications, as will be discussed herein.

To extend the advantages observed for $Sa_{avg}(T)$ in structure- and site-specific studies to a regional seismic risk context, a spatial correlation model is needed to generate suitable GMFs. Despite the increased use of $Sa_{avg}(T)$, a direct spatial correlation, a model was yet to be developed, representing a key inhibitor for regional applications (Monteiro and O'Reilly 2026). To address this gap, Heresi and Miranda (2021) presented an indirect derivation approach that leverages the mathematical formulation of $Sa_{avg}(T)$ as an average of multiple $Sa(T)$ terms. By combining existing spatial correlation models for individual periods and aggregating them, an indirect correlation model for $Sa_{avg}(T)$ was thus inferred, and will be discussed below.

This document revisits the formulation of $Sa_{avg}(T)$ as an IM and the indirect formulation of a spatial correlation model, as described by Heresi and Miranda (2021). Commentary on the potential and unintended misuse of such an approach is discussed alongside the relative magnitude of the impacts. Other issues, such as treatments of between-event and within-event residuals are further developed and discussed. Lastly, a recently developed direct spatial correlation model proposed by the authors (Monteiro et al. 2026) for $Sa_{avg}(T)$ is presented and critiqued with reference to the currently available indirect formulations.

2 | Indirect Formulation of Spatial Correlation Model

2.1 | Definitions and Notation

First, some basic definitions are recalled, and condensed notation is defined to reduce clutter in lengthy expressions that will be encountered later. Average spectral acceleration is defined as the logarithmic average of N spectral acceleration values:

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

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

Several definitions of these coefficients c_i and N exist in the literature. This has sometimes been treated as an optimisation problem for different structural typologies and analysis objectives. It has contributed to the confusion of an apparent uniqueness in the definition of $Sa_{avg}(T)$, masking its dependence on the period range used and the number of values considered. Hence, while some studies claim to use $\text{Sa}_{\text{avg}}(T)$ as the IM, it often has an ad hoc formulation whose precise definition is overlooked, making its results or proposed values difficult to generalise. To address this, some standard definitions exist, and two notable examples were identified in Shahnazaryan and O'Reilly (2024) are adopted herein, both with slightly different objectives. Eads et al. (2015) defined what will be herein termed $\text{Sa}_{\text{avg3}}(T)$ for collapse prediction, whereas Kohrangi et al. (2017) defined what will be herein termed $\text{Sa}_{\text{avg2}}(T)$ for moderate non-linear behaviour suitable for loss estimation. Both definitions utilise $N = 10$ and c_i is a linearly spaced coefficient ranging from $c_i \in [0.2, 2.0]$ for $\text{Sa}_{\text{avg2}}(T)$ and $c_i \in [0.2, 3.0]$ for $\text{Sa}_{\text{avg3}}(T)$. These will be recalled when needed in what follows, but for now, the generic definition of $\text{Sa}_{\text{avg}}(T)$ will used and can be interpreted to imply either definition, unless otherwise stated.

Since $\text{Sa}_{\text{avg}}(T)$ is a combination of $\text{Sa}(T)$ values, its mean, μ, and variance, σ^2, can be assembled assuming a normal distribution (i.e., $\mathcal{N}(\mu, \sigma^2)$), for a given rupture scenario rup using existing GMMs as follows [e.g., Bianchini et al. (2009); Bojórquez and Iervolino (2011); Kohrangi et al. (2017)]:

$$\mu_{\ln \text{Sa}_{\text{avg}}(T)|\text{rup}} = \frac{1}{N}\sum_{i=1}^{N} \mu_{\ln \text{Sa}(c_i T)|\text{rup}} \tag{3}$$

$$\sigma^2_{\ln \text{Sa}_{\text{avg}}(T)|\text{rup}} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \rho_{\ln \text{Sa}(c_i T),\, \ln \text{Sa}(c_j T)}\, \sigma_{\ln \text{Sa}(c_i T)|\text{rup}}\, \sigma_{\ln \text{Sa}(c_j T)|\text{rup}} \tag{4}$$

where $\mu_{\ln \text{Sa}(c_i T)|\text{rup}}$ and $\sigma_{\ln \text{Sa}(c_i T)|\text{rup}}$ are the mean and standard deviation obtained from a suitable GMM for $\text{IM} = \text{Sa}(c_i T)$ and $\rho_{\ln \text{Sa}(c_i T),\, \ln \text{Sa}(c_j T)}$ is the correlation between $\text{Sa}(T)$ values at periods $c_i T$ and $c_j T$ which is generally assumed to be independent of the rup parameters (Baker and Bradley (2017)).

In order to simplify the expressions above in the sections below, the following reduced notation will be used. The indices i and j will be used to refer to periods of vibration, and n and m will indicate the site locations. This way, the $\text{Sa}(T)$-based term at a particular site n is defined as:

$$Y(i, n) = \ln \text{Sa}(c_i T)_n \tag{5}$$

Since $\text{Sa}_{\text{avg}}(T)$ at the same site n is simply an aggregation of several $\text{Sa}(T)$ values, the following is also written:

$$X(n) = \ln \text{Sa}_{\text{avg}}(T)_n \tag{6}$$

$$= \frac{1}{N}\sum_{i=1}^{N} \ln \text{Sa}(c_i T)_n \tag{7}$$

$$= \frac{1}{N}\sum_{i=1}^{N} Y(i, n) \tag{8}$$

The above can also be written for other combinations of the indices, giving $Y(i, n)$, $Y(j, n)$, $Y(i, m)$, $Y(j, m)$ alongside $X(n)$ and $X(m)$. It then follows that the mean value of the $\text{Sa}_{\text{avg}}(T)$, which was previously defined in Equation (3), can be written for site n and rupture rup as:

$$\mu_{X(n)} = \mu_{\ln \text{Sa}_{\text{avg}}(T)_n|\text{rup}} \tag{9}$$

$$= \frac{1}{N}\sum_{i=1}^{N} \mu_{\ln \text{Sa}(c_i T)_n|\text{rup}} \tag{10}$$

$$= \frac{1}{N}\sum_{i=1}^{N} \mu_{Y(i,n)} \tag{11}$$

and the variance σ^2 can be written as:

$$\sigma^2_{X(n)} = \sigma^2_{\ln \text{Sa}_{\text{avg}}(T)_n|\text{rup}} \tag{12}$$

$$= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \rho_{\ln \text{Sa}(c_i T)_n, \ln \text{Sa}(c_j T)_n} \sigma_{\ln \text{Sa}(c_i T)_n|\text{rup}} \sigma_{\ln \text{Sa}(c_j T)_n|\text{rup}} \tag{13}$$

$$= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \rho_{Y(i,n),Y(j,n)} \sigma_{Y(i,n)} \sigma_{Y(j,n)} \tag{14}$$

The rup notation has been dropped from the subscript at each site n and m, assuming it is implicit, and the definition of $\text{Sa}_{\text{avg}}(T)$ is kept generic. The purpose of this notation X and Y is to simplify the lengthy expression when reproducing the indirect spatial correlation described by Heresi and Miranda (2021) discussed next, and also when extending to handle between- and within-event variability together.

2.2 | Indirect Formulation of Spatial Correlation

The correlation between two random variables A and B, $\rho_{A,B}$, can be written as:

$$\rho_{A,B} = \frac{E[(A - \mu_A)(B - \mu_B)]}{\sigma_A \sigma_B} \tag{15}$$

Given that in this case, it is the correlation between two random variables that represent the $\text{Sa}_{\text{avg}}(T)$ at sites n and m that is sought, and can be written using the simplified notation defined above as:

$$\rho_{X(n),X(m)} = \frac{E\left[\left(X(n) - \mu_{X(n)}\right)\left(X(m) - \mu_{X(m)}\right)\right]}{\sigma_{X(n)} \sigma_{X(m)}} \tag{16}$$

which is the first step of the derivation presented by Heresi and Miranda (2021) and followed the same steps but with notation listed above. From the definition of X above, these are substituted to get:

$$\rho_{X(n),X(m)} = \frac{E\left[\left(\frac{1}{N}\sum_{i=1}^{N} Y(i,n) - \frac{1}{N}\sum_{i=1}^{N} \mu_{Y(i,n)}\right)\left(\frac{1}{N}\sum_{j=1}^{N} Y(j,m) - \frac{1}{N}\sum_{j=1}^{N} \mu_{Y(j,m)}\right)\right]}{\sqrt{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \rho_{Y(i,n),Y(j,n)} \sigma_{Y(i,n)} \sigma_{Y(j,n)}} \sqrt{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \rho_{Y(i,m),Y(j,m)} \sigma_{Y(i,m)} \sigma_{Y(j,m)}}} \tag{17}$$

$$= \frac{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} E\left[\left(Y(i,n) - \mu_{Y(i,n)}\right)\left(Y(j,m) - \mu_{Y(j,m)}\right)\right]}{\sqrt{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \rho_{Y(i,n),Y(j,n)} \sigma_{Y(i,n)} \sigma_{Y(j,n)}} \sqrt{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \rho_{Y(i,m),Y(j,m)} \sigma_{Y(i,m)} \sigma_{Y(j,m)}}} \tag{18}$$

Recall that for $\text{Sa}(T)$ at two sites n and m, the spatial correlation of that IM at periods $c_i T$ and $c_j T$ would be, using the simplified notation above:

$$\rho_{Y(i,n),Y(j,m)} = \frac{E\left[\left(Y(i,n) - \mu_{Y(i,n)}\right)\left(Y(j,m) - \mu_{Y(j,m)}\right)\right]}{\sigma_{Y(i,n)} \sigma_{Y(j,m)}} \tag{19}$$

which when rearranged, gives:

$$E\left[\left(Y(i,n) - \mu_{Y(i,n)}\right)\left(Y(j,m) - \mu_{Y(j,m)}\right)\right] = \rho_{Y(i,n),Y(j,m)} \sigma_{Y(i,n)} \sigma_{Y(j,m)} \tag{20}$$

and can be substituted in the above to give:

$$\rho_{X(n),X(m)} = \frac{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \rho_{Y(i,n),Y(j,m)} \sigma_{Y(i,n)} \sigma_{Y(j,m)}}{\sqrt{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \rho_{Y(i,n),Y(j,n)} \sigma_{Y(i,n)} \sigma_{Y(j,n)}} \sqrt{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \rho_{Y(i,m),Y(j,m)} \sigma_{Y(i,m)} \sigma_{Y(j,m)}}} \tag{21}$$

Converting this back to regular notation, this becomes:

$$\rho_{\ln \text{Sa}_{\text{avg}}(T)_n, \ln \text{Sa}_{\text{avg}}(T)_m} = \frac{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\rho_{\ln \text{Sa}(c_iT)_n, \ln \text{Sa}\left(c_jT\right)_m}\sigma_{\ln \text{Sa}(c_iT)_n}\sigma_{\ln \text{Sa}\left(c_jT\right)_m}}{\sigma_{\ln \text{Sa}_{\text{avg}}(T)_n}\sigma_{\ln \text{Sa}_{\text{avg}}(T)_m}} \tag{22}$$

where the $\sigma_{\ln \text{Sa}_{\text{avg}}(T)}$ terms are given by Equation (12) and not fully expanded here due to the length of the entire expression. This is the generic form of the indirect derivation presented in Heresi and Miranda (2021).

2.3 | Handling Between- and Within-Event Residual Terms

In the previous derivation of the indirect formulation for the spatial correlation of the $\text{Sa}_{\text{avg}}(T)$, the total standard deviation ($\sigma_{\ln \text{Sa}_{\text{avg}}(T)}$) was employed. However, in ground motion simulations, the total variability is typically decomposed into between-event, δB and within-event, δW, components. This distinction is essential because GMFs are generated by first sampling a single between-event term for each rupture event and then producing spatially correlated within-event residuals across sites. This means that while total variability is used, the spatial correlation model employed is a within-event model, herein denoted as ρ^W. To represent this process mathematically, the logarithm of the intensity measure $X(n)$ for a given rupture event k, is assumed to follow a multivariate normal distribution:

$$X(n)_k \sim \mathcal{N}(\mu, \Sigma) \tag{23}$$

where $\sim\mathcal{N}()$ denotes the multivariate normal distribution parameterised by the mean vector μ and covariance matrix Σ defined for M sites:

$$\mu = \begin{bmatrix} X(n)_k \\ X(m)_k \\ \vdots \\ X(M)_k \end{bmatrix} \tag{24}$$

$$\Sigma = \tau^2\mathbf{1} + \phi^2\mathbf{R} \tag{25}$$

where τ and ϕ represent the between-event and within-event standard deviations, respectively (i.e., $\delta B \sim \mathcal{N}(0, \tau^2)$ and $\delta W \sim \mathcal{N}(0, \phi^2)$). $\mathbf{1}$ is a matrix of ones describing the perfect correlation among between-event terms, while $\mathbf{R}$ is the matrix of the within-event spatial correlation coefficients, with diagonal elements equal to unity and off-diagonal elements representing the spatial correlation between sites:

$$R = \begin{bmatrix} 1 & \cdots & \rho^W_{X(n),X(M)} \\ \vdots & \ddots & \vdots \\ \rho^W_{X(M),X(n)} & \cdots & 1 \end{bmatrix} \tag{26}$$

In the previous derivation of the indirect formulation for the spatial correlation of average spectral acceleration, the standard deviation term, $\sigma_{\ln \text{Sa}_{\text{avg}}(T)}$, was expressed in generic form (i.e., it was not specified if it referred to between, within or total residuals). However, for the present case, where $\rho^W_{X(n),X(M)}$ needs to be determined, Equation (22) should be adapted accordingly, such that:

$$\rho_{\ln \text{Sa}(c_iT)_n, \ln \text{Sa}\left(c_jT\right)_m} \rightarrow \rho^W_{\ln \text{Sa}(c_iT)_n, \ln \text{Sa}\left(c_jT\right)_m} \tag{27}$$

$$\sigma_{\ln \text{Sa}(c_iT)_n} \rightarrow \phi_{\ln \text{Sa}(c_iT)_n} \tag{28}$$

$$\sigma_{\ln \text{Sa}_{\text{avg}}(T)_n} \rightarrow \phi_{\ln \text{Sa}_{\text{avg}}(T)_n} \tag{29}$$

To construct a spatial correlation model for the total residuals by combining both the between-event and within-event components, the overall correlation term $\rho_{\ln \text{Sa}(c_iT)_n, \ln \text{Sa}\left(c_jT\right)_m}$ from Equation (22) should be applied in a manner that accounts for both contributions, as follows:

$$\rho_{Y(i,n),Y(j,m)} = \frac{\rho^B_{Y(i,n),Y(j,m)}\tau_{Y(i,n)}\tau_{Y(j,m)} + \rho^W_{Y(i,n),Y(j,m)}\phi_{Y(i,n)}\phi_{Y(j,m)}}{\sigma_{T_{Y(i,n)}}\sigma_{T_{Y(j,m)}}} \tag{30}$$

A similar derivation was used by Goda and Hong (2008) but it was assumed $\rho^B_{Y(i,n),Y(j,m)} \approx \rho^T_{Y(i),Y(j)}$, and a Markov-type model was used (Journel 1999) for the within-event spatial correlation model. This states that for two random variables, $Y(i,n)$ and $Y(j,m)$, the dependence of $Y(j,m)$ on the primary variable $Y(i,n)$ is limited to the colocated value of $Y(i,n)$. Leading to:

$$\rho^W_{Y(i,n),Y(j,m)} \approx \rho^T_{Y(i),Y(j)} \cdot \rho^W_{Y(\max_{[i,j]},n),\,Y(\max_{[i,j]},m)} \tag{31}$$

which when substituting these two simplifications above, gives:

$$\rho_{Y(i,n),Y(j,m)} \approx \frac{\rho^T_{Y(i),Y(j)}\left[\tau_{Y(i,n)}\tau_{Y(j,m)} + \rho^W_{Y(\max_{[i,j]},n),\,Y(\max_{[i,j]},m)}\phi_{Y(i,n)}\phi_{Y(j,m)}\right]}{\sigma_{Y(i,n)}\sigma_{Y(j,m)}} \tag{32}$$

where $\rho^T_{Y(i),Y(j)}$ is an inter-IM total correlation model and is independent of the site location [e.g., Baker and Jayaram (2008); Aristeidou et al. (2025)].

Coming back to the previous point, some GMMs assume homoscedastic aleatory variability [e.g., Aristeidou et al. (2024)], meaning they treat the standard deviations, σ, τ, and ϕ, as functions of spectral period T only (and do not of magnitude, distance or site). For those models, the notation need not include a site index (n or m), meaning the notation can be simplified to:

$$\begin{aligned}
\rho^B_{Y(i,n),Y(j,m)} = \rho^B_{i,j} \quad &\& \quad \rho^W_{Y(i,n),Y(j,m)} = \rho^W_{i,j,n,m}\\
\tau_{Y(i,n)} = \tau_i \quad &\& \quad \tau_{Y(j,m)} = \tau_j\\
\phi_{Y(i,n)} = \phi_i \quad &\& \quad \phi_{Y(j,m)} = \phi_j\\
\sigma_{Y(i,n)} = \sigma_i \quad &\& \quad \sigma_{Y(j,m)} = \sigma_j\\
\rho^W_{i,j,n,n} = \rho^W_{i,j,0} \quad &\& \quad \rho^W_{i,j,m,m} = \rho^W_{i,j,0}
\end{aligned} \tag{33}$$

Now, Equation (22) can be expanded for the total variability by replacing the correlation coefficient with the total-residual spatial correlation from Equation (30), and with the notation from Equation (33) to give:

$$\begin{aligned}
\rho_{\ln \mathrm{Sa_{avg}}(T)_n,\,\ln \mathrm{Sa_{avg}}(T)_m} &= \frac{\dfrac{1}{N^2}\displaystyle\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\rho^B_{i,j}\tau_i\tau_j + \rho^W_{i,j,n,m}\phi_i\phi_j}{\sigma_i\sigma_j}\sigma_i\sigma_j}{\sqrt{\dfrac{1}{N^2}\displaystyle\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\rho^B_{i,j}\tau_i\tau_j + \rho^W_{i,j,n,n}\phi_i\phi_j}{\sigma_i\sigma_j}\sigma_i\sigma_j}\sqrt{\dfrac{1}{N^2}\displaystyle\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\rho^B_{i,j}\tau_i\tau_j + \rho^W_{i,j,m,m}\phi_i\phi_j}{\sigma_i\sigma_j}\sigma_i\sigma_j}}\\
&= \frac{\displaystyle\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\rho^B_{i,j}\tau_i\tau_j + \rho^W_{i,j,n,m}\phi_i\phi_j\right]}{\displaystyle\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\rho^B_{i,j}\tau_i\tau_j + \rho^W_{i,j,0}\phi_i\phi_j\right]}\\
&= \frac{\displaystyle\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\rho^B_{\ln \mathrm{Sa}(c_iT),\,\ln \mathrm{Sa}(c_jT)}\tau_{\ln \mathrm{Sa}(c_iT)}\tau_{\ln \mathrm{Sa}(c_jT)} + \rho^W_{\ln \mathrm{Sa}(c_iT)_n,\,\ln \mathrm{Sa}(c_jT)_m}\phi_{\ln \mathrm{Sa}(c_iT)}\phi_{\ln \mathrm{Sa}(c_jT)}\right]}{\displaystyle\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\rho^B_{\ln \mathrm{Sa}(c_iT),\,\ln \mathrm{Sa}(c_jT)}\tau_{\ln \mathrm{Sa}(c_iT)}\tau_{\ln \mathrm{Sa}(c_jT)} + \rho^W_{\ln \mathrm{Sa}(c_iT)_n,\,\ln \mathrm{Sa}(c_iT)_n}\phi_{\ln \mathrm{Sa}(c_iT)}\phi_{\ln \mathrm{Sa}(c_jT)}\right]}
\end{aligned} \tag{34}$$

Figure 1 illustrates a comparison of applying Equation (22) to within-event residuals and Equation (34) to total residuals. In these curves, obtained using the two previously mentioned equations, the parameters $\tau_{\ln_{\mathrm{Sa}(c_iT)}}$, $\phi_{\ln_{\mathrm{Sa}(c_iT)}}$ and $\sigma_{\ln_{\mathrm{Sa}(c_iT)}}$ were obtained from the Aristeidou et al. (2024) GMM. For the within-event spatial correlation of spectral acceleration periods, $\rho^W_{\ln \mathrm{Sa}(c_iT)_n,\,\ln \mathrm{Sa}(c_jT)_m}$, the model of Loth and Baker (2013), developed using linear model of coregionalisation (LMC), was employed, as well as the models proposed by Markhvida et al. (2018); Du and Ning (2021); Monteiro et al. (2026) which

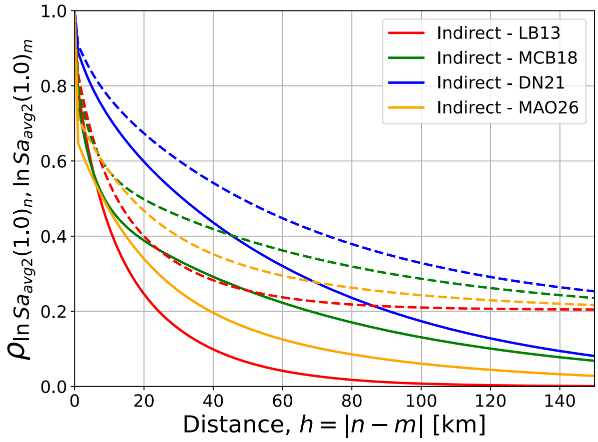

FIGURE 1 | Comparison of indirect approach for spatial correlation for $\mathrm{Sa}_{\mathrm{avg2}}(1.0)$ using within-event (solid lines) and total-event (dashed lines) using different inter-IM spatial correlation models. LB13:Loth and Baker (2013), MCB18:Markhvida et al. (2018), DN21: Du and Ning (2021), MAO26:Monteiro et al. (2026).

were derived using principal component analysis (PCA). For between-event correlation model, $\rho^{B}_{\ln \mathrm{Sa}(c_i T), \ln \mathrm{Sa}(c_j T)}$, the formulation of Goda and Atkinson (2009) was adopted.

It can be observed that when spatial correlation is computed from total residuals, the correlogram generally exhibits a non-zero asymptote at large intersite distances, consistent with observations by Heresi and Miranda (2019), who investigated the effect of between-event variability on spatial correlation models through Monte Carlo simulations. This behaviour arises because total residuals include a between-event component shared by all sites for a given earthquake, which therefore does not decay to zero with increasing distance.

3 | Discussion

While the indirect formulation for estimating the spatial correlation of $\mathrm{Sa}_{\mathrm{avg}}(T)$ is attractive due to its conceptual simplicity and utility of existing $\mathrm{Sa}(T)$-based GMMs, there are some unintended pitfalls that users may not be aware of when utilising such an approach, and are discussed below.

3.1 | Dependence on Inter-IM Spatial Correlation Models

Although the formulation presented in Equation (22) represents an intra-IM spatial correlation model (i.e., it describes how the same $\mathrm{IM} = \mathrm{Sa}_{\mathrm{avg}}(T)$ correlates spatially between locations n and m), it inherently depends on the availability of inter-IM spatial correlation models between individual spectral accelerations, $\mathrm{Sa}(c_i T)$ and $\mathrm{Sa}(c_j T)$, at different sites n and m via the presence of the $\rho^{W}_{\ln \mathrm{Sa}(c_i T)_n, \ln \mathrm{Sa}(c_j T)_m}$ term. However, most spatial correlation models in the literature, such those proposed by Jayaram and Baker (2009); Heresi and Miranda (2019); Aldea et al. (2022), are intra-IM spatial models, meaning they describe spatial correlation for the same period across different locations (i.e., $\rho^{W}_{\ln \mathrm{Sa}(c_i T)_n, \ln \mathrm{Sa}(c_i T)_m}$ and not the required $\rho^{W}_{\ln \mathrm{Sa}(c_i T)_n, \ln \mathrm{Sa}(c_j T)_m}$). Erroneously utilising such intra-IM spatial models in place of the required inter-IM correlations introduces a conceptual mismatch.

To overcome this limitation, intra-IM spatial models are often combined with inter-IM models [e.g., Baker and Jayaram (2008); Aristeidou et al. (2024)]. This can be achieved through a Markov-type model, as shown in Equation (31). Loth and Baker (2013) compared this methodology with the LMC method and found that the Markov-type model yielded good agreement for shorter periods, but its accuracy diminished as the periods became more widely separated.

Nevertheless, caution is necessary when combining models derived from distinct datasets and methodologies, as this can lead to misleading interpretations of spatial correlation. The fidelity of such indirect approaches is fundamentally constrained by the accuracy of the underlying $\mathrm{Sa}(T)$ spatial models, which may not fully capture the joint dependence structure of spectral accelerations across both periods and locations. In contrast, a direct empirical modelling approach, which quantifies the spatial correlation of $\mathrm{Sa}_{\mathrm{avg}}(T)$ itself, avoids these intermediate assumptions and may offer greater accuracy.

TABLE 1 | Period range limits for $\mathrm{Sa_{avg}}(T)$ when using different inter-IM models.

Inter-IM spatial model	Period range for $\mathrm{Sa}(c_i T)$	T_{max} for $\mathrm{Sa_{avg2}}(T)$	T_{max} for $\mathrm{Sa_{avg3}}(T)$	T_{min} for $\mathrm{Sa_{avg2}}(T)$	T_{min} for $\mathrm{Sa_{avg3}}(T)$
Markhvida et al. (2018)	[0.01–5.00]s	2.50 s	1.67 s	0.05 s	0.05 s
Monteiro et al. (2026)	[0.01–5.00]s	2.50 s	1.67 s	0.05 s	0.05 s
Loth and Baker (2013)	[0.01–10.00]s	5.00 s	3.33 s	0.05 s	0.05 s
Wang and Du (2013)	[0.01–10.00]s	5.00 s	3.33 s	0.05 s	0.05 s

Recognising this need, Monteiro et al. (2026) developed a direct spatial correlation model for $\mathrm{Sa_{avg}}(T)$, which will be discussed below.

3.2 | Limitations on Applicable Period Range

A further constraint concerns the range of periods supported by the inter-IM spatial correlation models. The studies by Loth and Baker (2013) and Wang and Du (2013) model spatial correlations for nine spectral periods ranging from 0.01 to 10 s using LMC. Similarly, Markhvida et al. (2018) developed an inter-IM spatial correlation model for 19 periods from 0.01 to 5 s using PCA. These limits are typically governed by the period range of the GMMs used to develop the correlations, which are in turn limited by the lowest usable frequencies of the supporting ground motion database.

This restriction directly limits the applicability of Equation (34) for computing spatial correlation of $\mathrm{Sa_{avg}}(T)$, especially when using definitions like $\mathrm{Sa_{avg2}}(T)$ and $\mathrm{Sa_{avg3}}(T)$ as they require information at periods two to three times the conditioning period T (i.e., $c_i T$ with $c_i = 2$ and $c_i = 3$, respectively). When using existing inter-$\mathrm{Sa}(T)$ models, the choice of the conditioning period T for $\mathrm{Sa_{avg}}(T)$ becomes explicitly bounded by the upper and lower bounds of the available period range. Table 1 summarises the maximum and minimum values of T, denoted T_{max} and T_{min}, that can be used in $\mathrm{Sa_{avg}}(T)$ computation based on the inter-$\mathrm{Sa}(c_i T)$ spatial correlation models and the indirect formulation mentioned above.

4 | Application Example

4.1 | Comparing Indirect and Direct Models

A direct model for spatial correlation of $\mathrm{Sa_{avg}}(T)$ was recently developed by Monteiro et al. (2026). This model is referred to here as the direct approach because it is derived directly from residuals of $\mathrm{Sa_{avg}}(T)$ computed using a GMM that explicitly predicts the mean values of $\mathrm{Sa_{avg}}(T)$. This contrasts with the indirect approach described in Section 2, which infers the spatial correlation of $\mathrm{Sa_{avg}}(T)$ from the spatial correlation of individual $\mathrm{Sa}(T)$ values.

The direct model is based on a multivariate framework that combines PCA with geostatistical tools, such as cross-semivariograms, to characterise both intra- and inter-IM spatial correlation structures. The use of PCA, originally introduced for spatial correlation modelling by Markhvida et al. (2018), allows the dominant spatial variability patterns across multiple IM to be captured efficiently while reducing the dimensionality of the problem. This is particularly relevant given the size and scope of the dataset used to develop the model, which includes 81 earthquakes and 11,499 ground motion records from a combined NGA-W2 and ESM database. The model was developed to estimate spatial correlations not only for $\mathrm{Sa_{avg}}(T)$ but also for multiple IM combinations, including $\mathrm{Sa_{avg2}}(T)$, $\mathrm{Sa_{avg3}}(T)$, FIV3 and other commonly used IMs, within a unified framework.

Figure 2 compares the within-event spatial correlation of $\mathrm{Sa_{avg2}}(T)$ obtained using the direct (i.e., the model of Monteiro et al. (2026)) and the indirect approach derived from Equation (22). The implementation of the indirect formulation used in this comparison is available on GitHub (https://github.com/vitorazevedomonteiro/notes-spatial-correlation-sa-avg.git).

Figure 2a,b illustrate the spatial correlation values for $\mathrm{Sa_{avg2}}(T)$ obtained through the indirect approach (Equation (22)) for two oscillator periods, 0.1 and 1.0 s, respectively. In both cases, the light-coloured lines surrounding each prediction represent the individual $\mathrm{Sa}(T)$ correlations that were combined to compute the corresponding $\mathrm{Sa_{avg2}}(T)$ correlations. The indirect models for $\mathrm{Sa_{avg2}}(T)$ were developed using inter-IM correlation models ($\rho^{W}_{\ln \mathrm{Sa}(c_i T)_n, \ln \mathrm{Sa}(c_j T)_m}$) derived from the LMC method (Loth and Baker 2013) and from PCA-based approach (Markhvida et al. 2018; Du and Ning 2021;

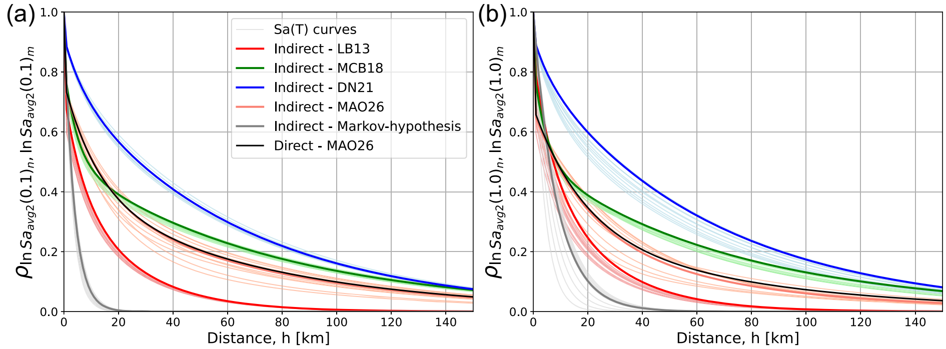

FIGURE 2 | Comparison of using Indirect and direct approach to model spatial correlation for (a) $\mathrm{Sa}_{\mathrm{avg2}}(0.1)$ or (b) $\mathrm{Sa}_{\mathrm{avg2}}(1.0)$. LB13:Loth and Baker (2013); MCB18:Markhvida et al. (2018); DN21:Du and Ning (2021); MAO26:Monteiro et al. (2026).

Monteiro et al. 2026). Additionally, the indirect approach is based on the Markov-type hypothesis described in Equation (31). For all the cases presented here, as similar to Figure 1, $\phi_{\ln_{\mathrm{Sa}(c_i T)}}$ is computed using Aristeidou et al. (2024).

Across both periods, a clear trend can be observed: the Markov-type model consistently exhibits the lowest spatial correlation coefficients with increasing distance, whereas the models developed using PCA generally yield higher correlation values, a general observation consistent with Monteiro and O'Reilly (2026). Notably, when the same inter-IM spatial correlation model is employed in both the indirect estimation using $\mathrm{Sa}(c_i T)$ and the direct formulation of $\mathrm{Sa}_{\mathrm{avg}}(T)$, the resulting correlations are remarkably similar. This consistency demonstrates that the indirect method is valid when benchmarked against the direct one, provided the same base models are used. Nevertheless, the direct approach provides a much more streamlined and computationally efficient means of estimating the spatial correlation of $\mathrm{Sa}_{\mathrm{avg}}(T)$, which is discussed next.

4.2 | Ground Motion Fields for Average Spectral Acceleration

To illustrate the impact of direct spatial correlation models on $\mathrm{Sa}_{\mathrm{avg}}(T)$ compared to an indirect way hereinafter referred to as the aggregate $\mathrm{Sa}(T)$-based GMF approach, GMFs conditioned on observed data were generated for the 1994 Northridge earthquake (Mw = 6.7), considering $\mathrm{Sa}_{\mathrm{avg2}}(1.0)$ and using the OpenQuake engine (Pagani et al. 2014). The aggregate $\mathrm{Sa}(T)$-based GMF procedure was first adopted to generate spatially consistent GMFs for $\mathrm{Sa}_{\mathrm{avg}}(T)$ using individual GMF simulations of $\mathrm{Sa}(T)$. In this approach, $\mathrm{Sa}(T)$-based GMFs were independently simulated for each value of $c_i T$ required from $\mathrm{Sa}_{\mathrm{avg}}(T)$'s definition. The mean $\ln\,\mathrm{Sa}_{\mathrm{avg}}(T)$ GMF was computed from each of the individual $\ln\,\mathrm{Sa}(T)$ simulations at each site, as per Equation (1). Under this assumption, spatial correlation and coherency can be thought to have been automatically enforced, and therefore no explicit correlation model for $\mathrm{Sa}_{\mathrm{avg}}(T)$ was required (Figure 3a). However, this approach is not only cumbersome, since it involves the independent sampling of N number of GMFs, but it is also conceptually incorrect. The variance and covariance structure of $\ln\,\mathrm{Sa}_{\mathrm{avg}}(T)$ cannot be recovered from independent per-period GMF simulations unless the cross-period correlation of spectral ordinates is explicitly accounted for. If GMFs are generated independently at each period $c_i T$, the interperiod covariance terms are dropped, which leads to an incorrect estimation of both the variance and the spatial coherence of $\ln\,\mathrm{Sa}_{\mathrm{avg}}(T)$.

Following the notation defined in Section 2.1, the variance of $X(n) = \ln\,\mathrm{Sa}_{\mathrm{avg}}(T)$ can be written as:

$$\begin{aligned}\mathrm{Var}[X(n)] &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{Cov}[Y(i,n),Y(j,n)] \\ &= \frac{1}{N^2}\left(\sum_{i=1}^{N}\mathrm{Var}[Y(i,n)] + \sum_{i\neq j}^{N}\mathrm{Cov}[Y(i,n),Y(j,n)]\right)\end{aligned} \tag{35}$$

If per-period GMFs are simulated independently, as was described above, this implies that $\mathrm{Cov}[Y(i,n),Y(j,n)] = 0$ for $i \neq j$, which gives:

$$\mathrm{Var}[X(n)]_{\mathrm{indep}} = \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{Var}[Y(i,n)] \tag{36}$$

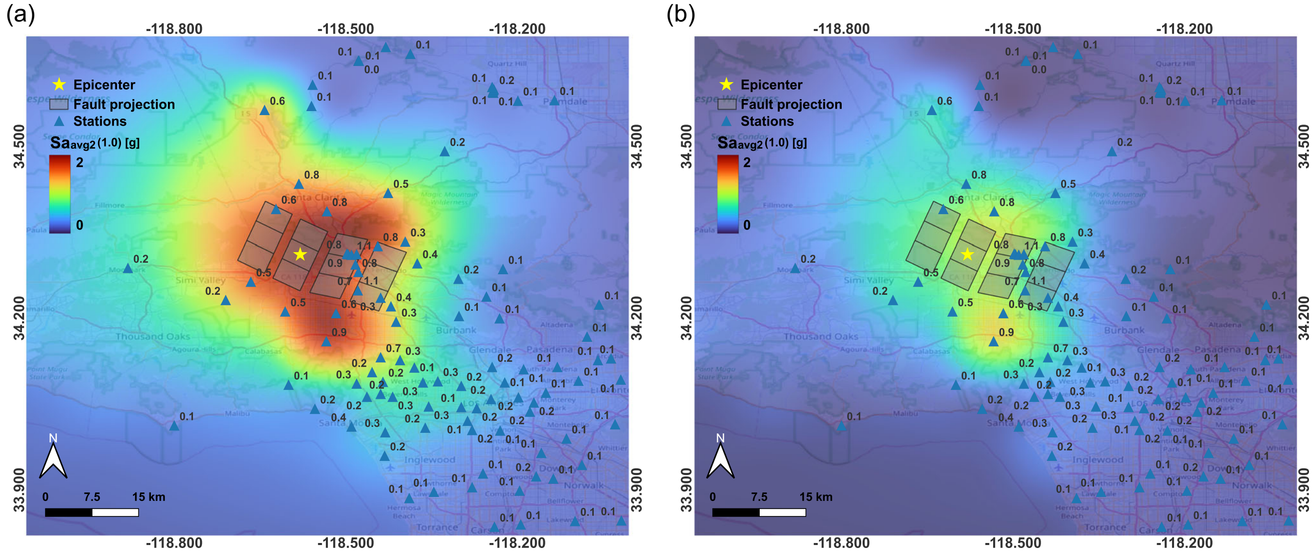

FIGURE 3 | Comparison of GMF for $\mathrm{Sa}_{\mathrm{avg2}}(1.0)$ using an (a) aggregate $\mathrm{Sa}(T)$-based approach and a (b) direct approach. The numbers shown above the stations correspond to the observed $\mathrm{Sa}_{\mathrm{avg2}}(1.0)$ values.

This formulation omits all cross-period covariance terms. Since residuals at nearby periods are empirically known to be positively correlated, neglecting these underestimates the true covariance of $\ln \mathrm{Sa}_{\mathrm{avg}}(T)$.

Similarly, the covariance between two different sites n and m is:

$$\mathrm{Cov}[X(n), X(m)] = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathrm{Cov}[Y(i,n), Y(j,m)] \tag{37}$$

and calculating the GMFs for each individual $Y(i) = \ln \mathrm{Sa}(c_i T)$, that is, admitting separability of the covariance structure, it can be calculated as follows:

$$\mathrm{Cov}[Y(i,n), Y(j,m)] = C_{\mathrm{period}}[i.j] \cdot C_{\mathrm{space}}[n,m] \tag{38}$$

which gives:

$$\mathrm{Cov}[X(n), X(m)] = \frac{1}{N^2} \left(\sum_{i=1}^{N} \sum_{j=1}^{N} C_{\mathrm{period}}[i,j] \right) C_{\mathrm{space}}[n,m] \tag{39}$$

where C_{period} denotes the covariance matrix across periods, constructed form interperiod correlations and standard deviations, while C_{space} represents the spatial covariance across sites. The complete covariance matrix for all periods and sites is obtained as the Kronecker product (Tismenetsky 1983):

$$C_{\mathrm{full}} = C_{\mathrm{period}} \otimes C_{\mathrm{space}} \tag{40}$$

with dimension $(N \cdot M) \times (N \cdot M)$. For example, in the case shown in Figure 3, where $M = 1,107$ locations and $N = 10$ periods in the range $c_i \in [0.2, 1.5]$ were used, the full covariance matrix has size $11{,}070 \times 11{,}070$. This highlights that, although this aggregate $\mathrm{Sa}(T)$-based GMF approach of developing a GMF initially appears attractive, it is also computationally very demanding and effectively discards the interperiod correlation structure, which is essential for preserving the correct variance and spatial coherence of $\ln \mathrm{Sa}_{\mathrm{avg}}(T)$. Hence, not only would a GMF for a given rupture scenario be incorrect, it would also be very slow to execute computationally.

To address this limitation, a direct calculation of GMF for $\mathrm{Sa}_{\mathrm{avg2}}(T)$ was performed using the spatial correlation for $\mathrm{Sa}_{\mathrm{avg2}}(T)$ developed by Monteiro et al. (2026). The expected GMF, obtained as the geometric mean of 10,000 simulations, is shown in Figure 3b, which exhibits noticeably lower $\mathrm{Sa}_{\mathrm{avg2}}(1.0)$ values compared to those obtained with the aggregated $\mathrm{Sa}(T)$-based GMF approach presented in Figure 3a. This confirms that the latter procedure leads to an overestimation of the expected GMF.

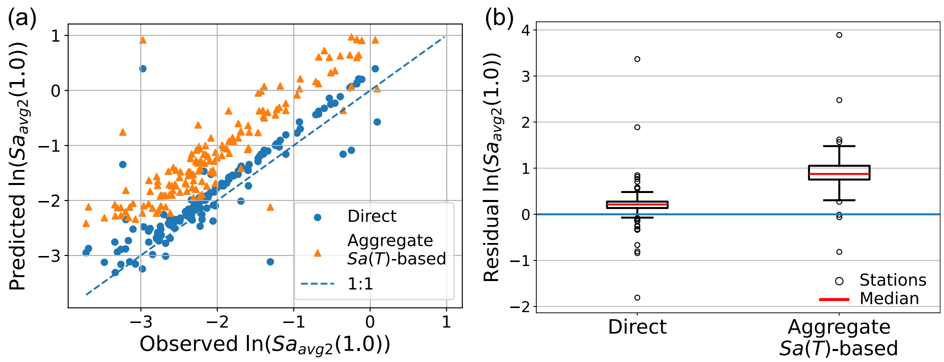

FIGURE 4 | Comparison between the direct and aggregate $\mathrm{Sa}(T)$-based approaches for $ln(\mathrm{Sa}_{\mathrm{avg2}}(1.0))$: (a) Predicted versus observed values at the stations, with the 1:1 line indicating perfect agreement; (b) residual distributions for the two approaches.

Neglecting interperiod interaction leads to an underestimation of the covariance of $\ln \mathrm{Sa}_{\mathrm{avg2}}(1.0)$, as the positive correlations among the contribution spectral ordinates are not accounted for, as evident from Equation (37). While this reduced covariance might appear to only affect the dispersion, it also impacts the conditional mean in the context of GMFs. Specifically, when conditioning on observed data, the underestimated covariance limits the model's ability to distribute variability consistently across correlated periods. As a result, the conditional simulation compensated by increasing the predicted mean of $\mathrm{Sa}_{\mathrm{avg2}}(1.0)$, leads to systematic overestimation. This effect is particularly pronounced in GMFs conditioned on stations, explaining the large overestimations observed in Figure 3.

An alternative could be applying Equation (23) to represent the spatial correlation for the considered IM. However, as observed in Figure 2, both models yield very similar correlation structures, and consequently, the resulting ground-motion fields are nearly indistinguishable with this indirect formulation just requiring a few more ingredients in terms of input models to build upon, which users need to exhibit care when choosing.

To further illustrate the differences between the aggregate $\mathrm{Sa}(T)$-based GMF and direct approaches, Figure 4 shows a station-by-station comparison of the predicted $\mathrm{Sa}_{\mathrm{avg2}}(1.0)$ values against the true observations. Figure 4a highlights that the aggregate $\mathrm{Sa}(T)$-based approach systematically overestimates the GMF, while the direct approach provides predictions closer to the observed values. The slight positive offset observed between the geometric mean of the 10,000 simulated $\mathrm{Sa}_{\mathrm{avg2}}(1.0)$ values (direct approach) and the observed values at the stations, also illustrated in Figure 4b, arises naturally from the characteristics of the conditional GMF simulations. Each simulation represents a random sample from the conditional distribution of $\ln\left(\mathrm{Sa}_{\mathrm{avg2}}\right)$ given the station observations, which has mean μ_{cond} and finite variance. Although the observed value at a station is used to condition the distribution, the conditional mean μ_{cond} provides a smoothed estimate reflecting the spatial correlation structure and the covariance model. Consequently, the geometric mean over the 10,000 simulations, which estimates $\exp\left(\mu_{\mathrm{cond}}\right)$, does not necessarily equal the single observed value at the station. This small offset (more notable in Figure 4b) does not indicate a bias in the GMF or the simulations. It reflects the natural variability of a single conditional realisation around the conditional mean, even when using the direct approach. Overall, this comparison quantitatively validates that the direct approach reduces the overestimation observed in the aggregate $\mathrm{Sa}(T)$-based GMF approach, while the minor offsets at the stations are consistent with expected conditional variability.

5 | Summary and Conclusions

This paper presented a comparative study of two methodologies for quantifying the spatial correlation of average spectral acceleration, $\mathrm{Sa}_{\mathrm{avg}}(T)$, which has received increasing interest in seismic vulnerability modelling in recent years. The first method is the so-called indirect approach, initially presented in Heresi and Miranda (2021), which derives the spatial correlation of $\mathrm{Sa}_{\mathrm{avg}}(T)$ from the combination of individual $\mathrm{Sa}(c_iT)$ values. These can be taken as $c_i \in [0.2T, 2T]$ or $c_i \in [0.2T, 3T]$, corresponding to $\mathrm{Sa}_{\mathrm{avg2}}(T)$ and $\mathrm{Sa}_{\mathrm{avg3}}(T)$, respectively. The second method is a direct formulation of the spatial correlation of $\mathrm{Sa}_{\mathrm{avg}}(T)$ developed by Monteiro et al. (2026), which employs PCA and geostatistical tools. Both approaches were described and compared in detail to highlight their similarities, differences and practical implications. Final remarks on their limitations and potential applications were also discussed in the context of potential pitfalls and issues to avoid in future implementation. Based on the findings presented herein, the following conclusions can be drawn:

- The indirect approach depends on the adoption of inter-IM spatial correlation models or, alternatively, on a Markov-type model combining intra-IM spatial correlation with inter-IM correlation, which is an assumption that is not always reliable, especially when periods are widely spaced, which is the case of $Sa_{avg}(T)$;
- The period range applicability of the indirect approach is constrained by the inter-IM spatial correlation models used in the analysis;
- Despite these limitations, the indirect approach yields results consistent with the direct formulation, demonstrating coherence in both methodology and applicability. However, its practical implementation is outperformed by the direct method, since for a 10-period range of $Sa_{avg2}(T)$ or $Sa_{avg3}(T)$, it requires 100 inter- and intraperiod correlation combinations;
- In the context of GMFs, one possible workaround would be to simulate $Sa(T)$-based GMFs individually and then combine together afterwards as per $Sa_{avg}(T)$'s definition. However, this study shows that this strategy neglects inter-period interactions, leading to erroneous estimates of spatial coherence. In the analysed case, this omission led to an overestimation of the hazard. To mitigate this issue, GMFs can now be simulated using a spatial correlation model that correlates $Sa_{avg}(T)$, which is faster, less computationally demanding and arguably easier to implement.

Overall, this discussion between the indirect and direct formulations of spatial correlation modelling for $Sa_{avg}(T)$ demonstrates that while the indirect approach does function well, it entails additional computational effort that could be prone to misinterpretation and pitfalls due to differences between within- and total-residual formulations and, consequently, in the choice of spatial correlation model type, and can lead to hazard overestimation when applied incorrectly. These findings highlight the potential benefits of adopting a direct formulation for reliable and efficient regional seismic risk assessments.

Acknowledgements

The work presented in this paper has been developed within the framework of the project 'Dipartimenti di Eccellenza 2023–2027,' funded by the Italian Ministry of Education, University and Research at IUSS Pavia. The authors are thankful for the comments and discussions with Peter Stafford, Pablo Heresi and Helen Crowley that helped improve this work.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Some files and functions that implement the comparisons and expansion of average spectral acceleration's spatial correlation to consider both within and between event correlation are available on GitHub at: https://github.com/vitorazevedomonteiro/notes-spatial-correlation-sa-avg.git.

References

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

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

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

Baker, J. W., and B. A. Bradley. 2017. "Intensity Measure Correlations Observed in the NGA-West2 Database, and Dependence of Correlations on Rupture and Site Parameters." *Earthquake Spectra* 33, no. 1: 145–156. https://doi.org/10.1193/060716eqs095m.

Baker, J. W., and N. Jayaram. 2008. "Correlation of Spectral Acceleration Values from NGA Ground Motion Models." *Earthquake Spectra* 24: 299–317. https://doi.org/10.1193/1.2857544.

Bianchini, M., P. P. Diotallevi, and J. W. Baker. 2009. "Prediction of Inelastic Structural Response Using an Average of Spectral Accelerations." In *Proceedings of the 10th International Conference on Structural Safety and Reliability (ICOSSAR09).*

Bojórquez, E., and I. Iervolino. 2011. "Spectral Shape Proxies and Nonlinear Structural Response." *Soil Dynamics and Earthquake Engineering* 31: 996–1008. https://doi.org/10.1016/j.soildyn.2011.03.006.

Boore, D. M. 2003. "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." *Bulletin of the Seismological Society of America* 93: 2737–2751. https://doi.org/10.1785/0120020197.

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

Dávalos, H., and E. Miranda. 2021. "A Ground Motion Prediction Model for Average Spectral Acceleration." *Journal of Earthquake Engineering* 25: 319–342. https://doi.org/10.1080/13632469.2018.1518278.

Eads, L., E. Miranda, and D. G. Lignos. 2015. "Average Spectral Acceleration as an Intensity Measure for Collapse Risk Assessment." *Earthquake Engineering & Structural Dynamics* 44, no. 12: 2057–2073. https://doi.org/http://doi.wiley.com/10.1002/eqe.2575.

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

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

Goda, K., and H. P. Hong. 2008. "Spatial Correlation of Peak Ground Motions and Response Spectra." *Bulletin of the Seismological Society of America* 98, no. 1: 354–365. https://doi.org/10.1785/0120070078.

Heresi, P., and E. Miranda. 2019. "Uncertainty in Intraevent Spatial Correlation of Elastic Pseudo-Acceleration Spectral Ordinates." *Bulletin of Earthquake Engineering* 17, no. 3: 1099–1115. https://doi.org/10.1007/s10518-018-0506-6.

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

Heresi, P., and E. Miranda. 2023. "RPBEE: Performance-Based Earthquake Engineering on a Regional Scale." *Earthquake Spectra* 39, no. 3: 1328–1351. https://doi.org/10.1177/87552930231179491.

Jayaram, N., and J. W. Baker. 2009. "Correlation Model for Spatially Distributed Ground-Motion Intensities." *Earthquake Engineering & Structural Dynamics* 38, no. 15: 1687–1708. https://doi.org/10.1002/eqe.922.

Journel, A. G. 1999. "Markov Models for Cross-Covariances." *Mathematical Geology* 31: 1019–1021. https://doi.org/10.1023/A:1007565316113.

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

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

Kohrangi, M., S. R. Kotha, and P. Bazzurro. 2018. "Ground-Motion Models for Average Spectral Acceleration in a Period Range: Direct and Indirect Methods." *Bulletin of Earthquake Engineering* 16, no. 1: 45–65. https://doi.org/10.1007/s10518-017-0216-5.

Loth, C., and J. W. Baker. 2013. "A Spatial Cross-Correlation Model of Spectral Accelerations at Multiple Periods." *Earthquake Engineering & Structural Dynamics* 42, no. 3: 397–417. https://doi.org/10.1002/eqe.2212.

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

Monteiro, V. A., S. Aristeidou, and G. J. O'Reilly. 2026. "Spatial Cross-Correlation Models for Next-Generation Amplitude and Cumulative Intensity Measures." *Earthquake Spectra*.

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

Nafeh, A. M. B., and G. J. O'Reilly. 2024. "Fragility Functions for Non-Ductile Infilled Reinforced Concrete Buildings Using Next-Generation Intensity Measures Based on Analytical Models and Empirical Data from past Earthquakes." *Bulletin of Earthquake Engineering* 22, no. 10: 4983–5021. https://doi.org/10.1007/s10518-024-01955-4.

O'Reilly, G. J. 2021a. "Limitations of Sa(T1) as an Intensity Measure when Assessing Non-Ductile Infilled Rc Frame Structures." *Bulletin of Earthquake Engineering* 19: 2389–2417. https://doi.org/10.1007/s10518-021-01071-7.

O'Reilly, G. J. 2021b. "Seismic Intensity Measures for Risk Assessment of Bridges." *Bulletin of Earthquake Engineering* 19: 3671–3699. https://doi.org/10.1007/s10518-021-01114-z.

Pagani, M., D. Monelli, G. Weatherill, et al. 2014. "Openquake Engine: An Open Hazard (and Risk) Software for the Global Earthquake Model." *Seismological Research Letters* 85: 692–702. https://doi.org/10.1785/0220130087.

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

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

Sokolov, V., F. Wenzel, W.-Y. Jean, and K.-L. Wen. 2010. "Uncertainty and Spatial Correlation of Earthquake Ground Motion in Taiwan." *Terrestrial, Atmospheric and Oceanic Sciences* 21: 905. https://doi.org/10.3319/TAO.2010.05.03.01(T).

Tismenetsky, M. 1983. "Kronecker Products and Matrix Calculus (Alexander Graham)." *SIAM Review* 25: 420–421. https://doi.org/10.1137/1025098.

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