Modelling anisotropic spatial cross-correlations of multiple ground-motion intensity measures

Ran Yuan[a]

[a] Key Laboratory of Transportation Tunnel Engineering, Ministry of Education, Southwest Jiaotong University, Chengdu 610031, China

Email: yuanran@swjtu.edu.cn

Yi Qiu[b]

[b] School of Civil Engineering, Southwest Jiaotong University, Chengdu 610031, China

Email: 2020200001@my.swjtu.edu.cn

Tong-Tong Liu[c]

[c] Northwest China Electric Power Design Institute Construction Group Co., Ltd., Xian 710065, China

Email: liutongtong@my.swjtu.edu.cn

Yin Cheng[b*]

[b] School of Civil Engineering, Southwest Jiaotong University, Chengdu 610031, China

Email: yin.cheng@swjtu.edu.cn

Chaolie Ning[d]

[d] College of Civil Engineering (State Key Laboratory of Disaster Reduction in Civil Engineering), Tongji University, 1239 Siping Road, Shanghai 200092, China,

Email: ningchaolie@126.com

Dariusz Wanatowski[e]

[e] Faculty of Engineering and Physical Sciences, University of Leeds, Leeds LS2 9JT, UK.

Email: D.Wan@leeds.ac.uk

Yi He[f]

[f] Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 611756, China

Email: heyi@swjtu.edu.cn

Abstract

Although anisotropy in spatial correlations of intensity measures (IMs) has been acknowledged, few models specifically address the anisotropy, and consider cross-correlations of IMs. This study introduces anisotropic spatial cross-correlation models for 16 IMs: PGA, PGV, I_a, CAV, $D_{s5\text{-}75}$, T_m, spectral V_{EIr}, and SA at periods of 0.05s, 0.2, 0.5s, 1s and 2s. The models can predict for four anisotropy directions, which are angled (i.e., 0°, 45°, 90°, and 135°) relative to the fault direction. Finally, the proposed model was utilized for the regional seismic landslide hazard assessment to show a practical application.

Keywords: anisotropy; spatial cross-correlations; linear model of coregionalization; intensity measures

1 Introduction

Assessing ground-motion intensity measures (IMs) across a geographically extensive area is crucial for evaluating the earthquake risk associated with spatially distributed assets of both structures and geotechnics over a spatially distributed region (Garakaninezhad and Bastami, 2017). Numerous ground-motion prediction equations (GMPEs) have been formulated to predict IMs based on factors like earthquake magnitude, source-to-site distance, and site-specific geological conditions. These predictions encompass parameters like peak ground acceleration (PGA), peak ground velocity (PGV), and spectral acceleration (SA). However, these GMPEs often overlook the spatial relationships of ground-motion IMs across different sites within the same earthquake event. IMs at multiple locations during the same earthquake are spatially correlated (Schiappapietra et al., 2022), this relationship is called the spatial correlation of IMs, while the spatial cross-correlation delves into the relationships between different IMs within a given spatial domain. Studies (e.g., Abbasnejadfard et al., 2021b; Garakaninezhad and Bastami, 2019) have indicated that ignoring the spatial correlation of IMs can significantly skew seismic loss estimates of spatially distributed systems or portfolios. Therefore, it is essential to account for these spatial correlations to ensure a more accurate and comprehensive risk assessment.

Modeling the spatial correlation between different sites is essential for a precise evaluation of seismic risk on a regional scale. Over the past few years, a variety of correlation models have been introduced to understand the variability in IMs. These investigations have revealed that certain parameters, which are not randomly distributed, influence the correlation of PGA residuals. Such parameters include the size of the earthquake magnitude and the depth of sediment layers (e.g., Boore et al., 2003; Goda and Hong, 2008; Jayaram and Baker, 2009; Sokolov and Wenzel, 2013; Wang and Takada, 2005).

The configuration of the correlation model has a significant influence on the generation of simulations for areas with the cross-correlation of ground-motion IMs, such as the seismic activity within a particular region. It should be noted that the majority of the spatial correlation models for IMs are normally based on the assumption of isotropy (e.g., Esposito and Iervolino, 2011; Wang and Du, 2013). The assumptions of isotropy and stationarity are considered as the basis for studying spatial correlation with respect to the 1994 Northridge and the 1999 Chi–Chi earthquakes (Jayaram and Baker, 2009). Using the directional semivariograms, it has been concluded that the assumption of isotropy is reasonable for both earthquakes, i.e., Northridge and Chi-Chi earthquakes (Jayaram and Baker, 2009). The spatial correlation of SAs, cumulative absolute velocity (CAV), and Arias intensity (I_a) have been investigated in Du and Wang (2013). However, Garakaninezhad and Bastami (2017) used statistical tests introduced by Bowman and Crujeiras (2013) to examine the isotropic assumptions of PGA and SA residuals within events and concluded that the isotropic assumption is generally invalid. Furthermore, they proposed spatial correlation models for the anisotropy of PGA and SA residuals within events. In addition, Abbasnejadfard et al. (2021a) conducted statistical tests to investigate the isotropic assumptions of PGV and peak ground displacement (PGD) residuals within events and found that the residuals of PGV and PGD should be considered as realizations of an anisotropic random field, which is related to the local site-specific anisotropy.

Nevertheless, there are a few concerns that need to be addressed, especially when considering the spatial cross-correlation of IMs under isotropic and anisotropic conditions. Most current studies focus on the spatial correlation of IMs while neglecting the spatial cross-correlation between IMs. A few notable empirical models have confirmed the importance of considering the anisotropic spatial cross-correlation of multiple earthquake IMs, e.g., the model proposed by Abbasnejadfard et al. (2020); they only included a few IMs that are suitable for structures, i.e., PGA, PGV, PGD and SAs, which are suitable for seismic risk assessment of portfolios of buildings or structures; however, these IMs are insufficient, especially for the hazard assessment of geotechnical problems, e.g., earthquake-induced landslides from a wide range of slope conditions in nature. Although the effects of anisotropic spatial correlations of IMs have been analyzed for seismic hazard assessment of the portfolio of buildings and infrastructure systems (Abbasnejadfard et al., 2021b), the effects of anisotropic cross-correlations of IMs have not been fully addressed on the hazard assessment in geotechnical problems. However, such effects are crucial to evaluating the risk/loss of spatially distributed infrastructure systems induced by seismic slope failures or landslides, since it is commonly acknowledged that soils are non-homogeneous, demonstrating characteristics of anisotropy and space variety. Hence, the seismic responses of soils subjected to seismic loading are complicated and various (Cheng et al., 2020; Du and Wang, 2013).

Hence, this study intends to propose anisotropic spatial cross-correlation models for 16 vector-IMs, including the relative input energy equivalent velocity ($V_{EIr}(T)$) and SA ordinates at periods (T) of 0.05s, 0.2, 0.5s, 1s and 2s, PGA, PGV, I_a, CAV, significant durations ($D_{s5\text{-}75}$), and the mean period (T_m), which are suitable for further seismic risk assessment of spatially distributed assets (both structural and geotechnical conditions), based on the proper ground-motion records that are selected from NGA-West2. Furthermore, as a case study, the influence of the anisotropic spatial cross-correlations of the IMs on the fully probabilistic seismic landslide hazard assessment of slopes is studied under different conditions, i.e., critical yield accelerations (a_c), permanent displacement models, site conditions, and rupture distances. This method provides a theoretical benchmark for regional seismic hazard and risk assessment of the portfolio of buildings/slopes considering the anisotropic cross-correlations of IMs.

2 Modelling anisotropic spatial cross-correlations based on geostatistical tools

2.1 Isotropy in multivariate spatial cross-correlations (Cheng et al., 2020)

Evaluation of the univariate spatial correlation of within-event residuals of IMs needs to consider the recorded residuals as the realization of a function that takes random values in spatial locations **u**. The function of the random variable is known as random fields in statistics, defined as:

$$\{Z(\mathrm{u}):\mathrm{u} \in \mathrm{R}^2\} \quad (1)$$

Where $Z(\mathbf{u})$ is the random variable at position **u**.

Geostatistical analysis suggested that the spatial correlation of a random variable is usually quantified by semivariograms (Journel and Huijbregts, 1976). The semivariogram $\gamma(\mathrm{h})$ is a plot of semivariance versus separation distance, measuring the average dissimilarity between data separated by a vector **h** (Goovaerts, 1997). The semivariogram function can be given as (Goovaerts, 1997):

$$\gamma(\mathrm{h}) = \frac{1}{2}\, Var[Z(\mathrm{u}+\mathrm{h}) - Z(\mathrm{u})] = \frac{1}{2}\, E[(Z(\mathrm{u}+\mathrm{h}) - Z(\mathrm{u}))^2] \quad (2)$$

Where $Var[.]$ denotes variance; **h** is lag distance; $Z(\mathrm{u})$ and $Z(\mathrm{u}+\mathrm{h})$ are variables of *Z* at locations with a lag distance of **h**.

For a multivariate random field, consisting of *k* random variables ($\{Z(\mathrm{u}) = [Z_1(\mathrm{u}), \dots, Z_k(\mathrm{u})]' : \mathrm{u} \in R^2\}$), which is assumed to be isotropic and second-order stationary, the empirical estimator of the semivariogram can be defined as (Goovaerts, 1997):

$$\gamma_{mw}(\mathrm{h}) = \frac{1}{2}\,\frac{1}{|N(h)|}\sum_{i=1}^{|N(h)|}[(Z_m(\mathrm{u}_i) - Z_m(\mathrm{u}_i+\mathrm{h}))(Z_w(\mathrm{u}_i) - Z_w(\mathrm{u}_i+\mathrm{h}))] \quad (3)$$

Where $Z_m(\mathrm{u}_i)$ and $Z_m(\mathrm{u}_i+\mathrm{h})$ denote the i^{th} data pair in the **h**-distance bin for random variable Z_m. $|N(\mathrm{h})|$ is the dimension of the set of all distinct points which fall in the bin of the distance **h**.

In addition, the cross-covariance function $C_{mw}(\mathrm{h})$, measuring the similarity of data of the multivariate random field with a lag distance of **h** is defined as (Cressie, 1993):

$$C_{mw}(\mathrm{h}) = \frac{1}{|N(\mathrm{h})|} \sum_{i=1}^{|N(\mathrm{h})|} [(Z_m(\mathrm{u}_i) - \overline{Z_m})(Z_w(\mathrm{u}_i) - \overline{Z_w})] \quad (4)$$

Where $\overline{Z_k} = \frac{1}{n} \sum_{i=1}^{n} Z_k(\mathrm{u}_i)$ is the sample mean value of the variable *k*; $i(=1, \ldots, n)$ is the position in the lag distance bin.

By combining Equations 3 and 4, $\lim_{\mathrm{h}\to\infty} \gamma_{mw}(\mathrm{h}) = Cov[Z_m(\mathrm{u}_i), Z_w(\mathrm{u}_i)] = C_{mw}(0)$ (Goovaerts, 1997), in which *Cov*[.] denotes covariance.

The relation between cross-covariance and cross-semivariogram can be obtained as (Goovaerts, 1997):

$$C_{mw}(\mathrm{h}) = C_{mw}(0) - \gamma_{mw}(\mathrm{h}) = C_{mw}(0)\left(1 - \frac{\gamma_{mw}(\mathrm{h})}{C_{mw}(0)}\right) \quad (5)$$

As indicated by Equation 5, $\lim_{h\to 0} \gamma_{mw}(\mathrm{h}) = 0$.

The spatial correlation coefficient $\rho_{mw}(\mathrm{h})$ between Z_m and Z_w at lag distance *h* can be defined as (Goovaerts, 1997):

$$\rho_{mw}(\mathrm{h}) = \frac{C_{mw}(\mathrm{h})}{\{C_{mm}(h) \times C_{ww}(h)\}^{1/2}} = \frac{C_{mw}(0)}{\{C_{mm}(0) \times C_{ww}(0)\}^{1/2}} - \frac{\gamma_{mw}(\mathrm{h})}{\{C_{mm}(0) \times C_{ww}(0)\}^{1/2}} \quad (6)$$

It should be noted that if the cross-semivariogram and cross-covariance in Equation 6 are respectively substituted by Equation 4 and Equation 5, then the spatial correlation coefficient is determined.

The first- and second-order moments (expectation and covariance) exist and are invariant under spatial translation if the second-order stationary is assumed. The semivariogram is independent of the position **u**, but it is only in relation to the lag distance **h**. The isotropy is assumed if the semivariogram of a second-order stationary random field is irrespective of directions. Under the assumption of isotropy, the cross-semivariogram and the cross-covariance depend solely on the relative separation distance between positions, i.e., $h = \|\mathrm{h}\|$.

2.2 Parametric functions to fit semivariograms

The empirical semivariogram value is calculated based on the recorded seismic data. By fitting a model to the empirical values, an effective semivariogram function model can be obtained. It is worth noting that if the model is a conditionally non-negative definite function, it is considered as an effective semivariogram function (Cressie, 1993). Semivariogram analysis provides a set of empirical values for discrete lag distances *h*, and requires a continuous function that must satisfy positive definiteness to fit these empirical data for practical use. Notable models for fitting the empirical semivariogram can be classified as: Exponential model, Gauss model and

Spherical model (Wang and Du, 2013), and the comparisons of their fitting performance are shown in Fig.1. It is demonstrated in Fig.1 that the exponential fitting model shows a good performance. Hence, the exponential fitting model, as shown in Equation 7, will be used in this study.

$$\widehat{\gamma}(h) = a[1 - exp(-3h/b)] \tag{7}$$

Where *a* is the sill of the semivariogram, representing the overall variance of the empirical data *Z*; *b* is the practical range of semivariogram, representing the lag distance at which 95% of the sill *a* is reached.

Following Abbasnejadfard et al. (2020), for the specific direction, we used the isotropic variogram (Jäckel., 2002) form to calculate the semivariograms. Anisotropic characteristics are handled separately for each direction (grouped into four directions in this study), utilizing an isotropic variogram format that is based on data specific to that direction. This may have limitations as the directional semivariograms may not be fully considered. A possible solution can be to establish a holistic model encompassing all directions using the latent dimensions method; however, this will complicate the current study and will be considered in future work. Theoretically, the semivariogram value should equal to 0 at zero separation distance, i.e., *h*=0. However, the semivariogram often demonstrates a nugget effect in practice.

Hence, the exponential model for fitting semivariogram considering the nugget effect can be given as (e.g., Cheng et al., 2020; Du and Pan, 2017):

$$\hat{\gamma}(h) = \begin{cases} 0 & h = 0 \\ C_0 + (a - C_0)[1 - exp(-3h/b)] & h > 0 \end{cases} \quad (8)$$

Where C_0 denotes the nugget effect.

2.3 Linear model of coregionalization (LMC)

Since the variance of any linear combination of $Z(\mathrm{u}_i)$ at a random position u_i must be non-negative, the covariance function associated with the semi-variogram must be positive definite. The exponential permissible model is usually adopted in the earthquake engineering community for fitting the empirical covariance values at discrete distance lags. In practice, the positive definiteness can be ensured through the positive definite linear combination. It is worth noting that not all combinations of the permissible basic models introduced earlier bring in a permissible (i.e., positive definite) covariance function. The easiest way to develop a permissible model lies in building a random function first. In recent years, an advanced LMC method, which was first proposed by Goovaerts (1997), is a commonly suggested statistical tool for modelling effective multivariate semivariogram matrix of IMs. The LMC method employs a combination of short-range exponential function and long-range exponential function, together with a term of nugget effect. A nugget effect is incorporated into the LMC to improve the fitting performance at small separation distances. The expression for the LMC semivariogram model can be defined as:

$$\Gamma(h) = \mathrm{B}^0(1 - I_h) + \mathrm{B}^1[1 - \exp(-3h/r_1)] + \mathrm{B}^2[1 - \exp(-3h/r_2)] \quad (9)$$

Where the *h* denotes lag distance with a unit of km; $\Gamma(h)$ is semivariogram matrix for *h*, r_1 and r_2 represent the short-range coefficient and long-range coefficient; and B^0, B^1, B^2 denote the average nugget effect, short-range, and long-range coregionalization matrices for ten earthquake events; I_h is a bivariate indicator (equal to 1 at *h*= 0 or 0 otherwise).

The coregionalization matrices are established by minimizing the weighted sum of squares (WSS) that represents the discrepancy between the predicted and empirical semivariograms (Goulard and Voltz, 1992). WSS is evaluated as in Equation 10.

$$\mathrm{WSS} = \sum_{k=1}^{K} \sum_{m=1}^{n} \sum_{w=1}^{n} w(h_k) \frac{[\tilde{\gamma}_{mw}(h_k) - \gamma_{mw}(h_k)]^2}{\hat{\sigma}_m \hat{\sigma}_w} \quad (10)$$

In this context, $\widetilde{\gamma}_{mw}(h_k)$ represents the empirical semivariogram, while $\gamma_{mw}(h_k)$ indicates the semivariogram predicted by the LMC model. The term $w(h_k)$ refers to the weight applied at the lag distance h_k, and $\widehat{\sigma}_m$ signifies the standard deviation of the random variable Z_w. For further details on the algorithm, please refer to the work of Loth and Baker (2013).

2.4 Anisotropy in the multivariate spatial cross-correlations

As mentioned in Section 2.3, the Linear Model of Coregionalization (LMC) ensures positive-definite semivariogram and covariance matrix functions, a prerequisite for stochastic simulation of ground-motion fields (Goovaerts, 1997). However, LMC is typically applied to isotropic spatial fields. Anisotropy is accounted for when the semivariogram of a second-order stationary random field depends on both the direction and magnitude of the lag vector **h**. For a multivariate random field with anisotropy, the lag vector incorporates both magnitude and direction.

This study addresses anisotropic spatial correlations separately for each direction using an isotropic semivariogram LMC model specific to that direction. To ensure the resulting semivariogram, covariance, and correlation matrices are positive-definite, a detailed approach is proposed in this section. These positive-definite matrices make the method suitable for engineering applications.

2.4.1 Definition of anisotropy directions

A directional semivariogram is derived by calculating the semivariogram for data pairs that lie within specific directional bands and at a set bin distance. The study area is divided into different directional zones, where the azimuth of the direction vector, angular tolerance, and bandwidth are illustrated in Fig.2. These parameters can be employed to compute the directional semivariance, subsequently enabling the calculation of directional covariance functions (Cressie, 1993). For example, with respect to the target region as shown in Fig.3, the semivariogram values corresponding to an anisotropy direction of θ and an angle tolerance of $\Delta/2$ can be computed following the two steps: first, chose a reference site A, and identify all sites that fall within the anisotropy direction range of $[\theta - \frac{\Delta}{2}, \theta + \frac{\Delta}{2}]$. These sites are then organized based on the bin distances relevant to the analysis. This process must be conducted for every site in the dataset. By following the method described in Section 2.1, the empirical semivariograms for the specified anisotropy direction can be computed.

2.4.2 Procedures to consider the anisotropy in semivariograms

To ensure the covariance matrix for sampling spatially correlated IM residuals with anisotropic characteristics is positive-definite, the following procedure is used to construct the final matrix.

1) Examine the empirical semivariogram figures for each anisotropic direction; pre-select the short-range coefficient r_1 and long-range coefficient r_2 for each anisotropy direction θ for Equation 11 according to the guidelines of this study:

- r_1: controls the rapid spatial variation over small distances; it is selected such that the semivariogram reflects 10%–30% of the sill (*a*) in the exponential fitting model (Equation 8) when $r_1 = b = h$.
- r_2: controls the gradual spatial variation over larger distances; it is chosen such that the semivariogram reflects 70%–95% of sill (*a*) in the exponential fitting model (Equation 8) when with $r_2 = b = h$.

2) Using the dataset for each anisotropic direction θ, which includes records from the examined earthquake events, the pre-defined values of r_1 and r_2 are employed in a regression analysis to derive the coregionalization matrices $\mathbf{B}^0$, $\mathbf{B}^1$, and $\mathbf{B}^2$ for each anisotropy direction θ for each earthquake event. This is accomplished by minimizing the weighted sum of squares (WSS) of the difference between the predicted and the empirical semivariograms, as outlined in Equation 10. The average coregionalization matrices $\mathbf{B}^0$, $\mathbf{B}^1$, and $\mathbf{B}^2$ for all earthquakes are calculated and used as those for that specific anisotropy direction.

3) The general WSS for all empirical semivariograms from all earthquakes is

documented, and various combinations of r_1 and r_2 are tested until a relatively low WSS is achieved while adhering to the established criteria for r_1 and r_2, as stated in Step 1). Ultimately, the values of $\mathbf{B}^0$, $\mathbf{B}^1$, $\mathbf{B}^2$, r_1 and r_2 for each anisotropy direction θ are determined.

4) The average values of $\mathbf{B}^0$, $\mathbf{B}^1$, and $\mathbf{B}^2$ across four isotropy directions are obtained, respectively, while r_1 and r_2 values for the four directions remain.

5) The resulting covariance matrix for each anisotropic direction is obtained through Equation 14 based on average $\mathbf{B}^0$, $\mathbf{B}^1$ and $\mathbf{B}^2$ obtained from Step 4), along with the r_1 and r_2 obtained from Step 3) for each anisotropic direction.

6) If the elements in the resulting covariance matrix involve more than one anisotropic direction, the element from the covariance matrix of that direction is abstracted to construct the final covariance matrix.

In the end, the semivariogram matrix function for the anisotropy direction θ at the lag distance h is expressed as:

$$\Gamma^{\theta}(h) = \mathrm{B}^0(1 - I_h) + \mathrm{B}^1[1 - \exp(-3h/r_1^{\theta})] + \mathrm{B}^2[1 - \exp(-3h/r_2^{\theta})] \tag{11}$$

Where r_1^{θ} and r_2^{θ} represent the determined short-range coefficient and long-range coefficient for the anisotropy direction θ; and B^0, B^1, B^2 denote the average nugget effect, short-range, and long-range coregionalization matrices regressed across the different directions; I_h is a bivariate indicator (equal to 1 at h= 0 or 0 otherwise).

It is important to note that if the coregionalization matrices B^l (for l=0, 1, and 2) and are positive semidefinite, then the covariance matrix will also be positive semidefinite throughout the entire random field (Goovaerts, 1997). Although various anisotropic directions share the same B^l but have distinct short- and long-range coefficients, the linear combination of the positive-semidefinite B^l, as indicated in Equation 14, ensures that the resulting combined covariance matrix from the different anisotropic directions remains positive-semidefinite.

2.4.3 The covariance and correlation matrices with a single anisotropy direction

Once the semivariogram matrix function $\Gamma^{\theta}(h)$ is determined as described in the above procedure, the covariance matrix $\mathrm{C}^{\theta}(h)$ for the direction (θ) can be calculated as follows:

$$\mathrm{C}^{\theta}(h) = \mathrm{C}^{\theta}(0) - \Gamma^{\theta}(h) \qquad (12)$$

Where $\mathrm{C}^{\theta}(0)$ can be expressed as:

$$\mathrm{C}^{\theta}(0) = \lim_{h\to\infty} \Gamma(\mathrm{h}) = \mathrm{B}^0 + \mathrm{B}^1 + \mathrm{B}^2 \qquad (13$$

)

The covariance matrix function $C^{\theta}(h)$ can be obtained through Equations 11 and 12, expressed as

$$C^{\theta}(h) = B^{0} I_{h} + B^{1}\ exp\left(-3h/r_{1}^{\theta}\right) + B^{2}\ exp\left(-3h/r_{2}^{\theta}\right) \tag{14}$$

Then combining the covariance matrix and the spatial correlation coefficient as shown in Equation 6, the spatial cross-correlation matrix $\mathbf{R}^{\theta}(h)$ for the anisotropy direction θ can be determined as:

$$R^{\theta}(h) = [\rho_{ij}(h)] = P^{0} I_{h} + P^{1}\ exp\left(-3h/r_{1}^{\theta}\right) + P^{2}\ exp\left(-3h/r_{2}^{\theta}\right) \tag{15}$$

Where h is the separation distance and $P^{l} = [\rho_{ij}^{l}]$ $(l = 0,1, \text{and } 2)$ is the standardized coregionalization matrix:

$$\rho_{ij}^{l} = \frac{b_{ij}^{l}}{(\sqrt{b_{ii}^{0} + b_{ii}^{1} + b_{ii}^{2}}) \times (\sqrt{b_{jj}^{0} + b_{jj}^{1} + b_{jj}^{2}})} \quad l = 0,1, and\ 2 \tag{16}$$

Where b_{ij}^{l} denotes the element in B^{l}.

2.4.4 Construction of the covariance and correlation matrices with multiple anisotropy directions

For a given earthquake event *e*, the total covariance matrix C_e for a set of variables $\mathrm{Z} = \{Z_1, \ldots, Z_k\}^T$ is located at *n* spatially distributed positions denoted as $\mathrm{E} = \{\mathrm{Z}(\mathrm{u}_1)^T, \ldots, \mathrm{Z}(\mathrm{u}_n)^T\}^T$. Hence, based on Equation 14, the total covariance matrix C_e can be given as shown in Equation 17 (Genton and Kleiber, 2015; Wang and Du, 2013):

$$\mathrm{C}_e = \begin{bmatrix} \mathrm{C}^{\theta_{11}}(\mathrm{u}_1,\mathrm{u}_1) & \cdots & \mathrm{C}^{\theta_{1n}}(\mathrm{u}_1,\mathrm{u}_n) \\ \vdots & \ddots & \vdots \\ \mathrm{C}^{\theta_{n1}}(\mathrm{u}_n,\mathrm{u}_1) & \cdots & \mathrm{C}^{\theta_{nn}}(\mathrm{u}_n,\mathrm{u}_n) \end{bmatrix} \quad (17)$$

The covariance matrix C_e has the dimensions of $(k \times n, k \times n)$, making the simulation of the *k* correlated multivariate random variables at *n* spatial locations possible. θ_{ij} (i =1…n; j = 1…n) is the anisotropy direction for the path from the location $\mathbf{u}_i$ to $\mathbf{u}_j$. The element of $\mathrm{C}^{\theta_{ij}}(\mathrm{u}_i,\mathrm{u}_j)$, involving the *k* variables, is defined by $\mathrm{C}^{\theta_{ij}}(h)$ in Equation 14, where *h* represents the separation distance between the locations u_i and u_j. Similarly, the total correlation matrix R_e, which has a dimension of $(k \times n, k \times n)$ is expressed as:

$$\mathrm{R}_e = \begin{bmatrix} \mathrm{R}^{\theta_{11}}(\mathrm{u}_1,\mathrm{u}_1) & \cdots & \mathrm{R}^{\theta_{1n}}(\mathrm{u}_1,\mathrm{u}_n) \\ \vdots & \ddots & \vdots \\ \mathrm{R}^{\theta_{n1}}(\mathrm{u}_n,\mathrm{u}_1) & \cdots & \mathrm{R}^{\theta_{nn}}(\mathrm{u}_n,\mathrm{u}_n) \end{bmatrix} \quad (18)$$

As noted by Genton and Kleiber (2015), the covariance matrix should be non-negative definite matrix given arbitrary spatial positions $\mathrm{u}_1, \ldots \mathrm{u}_i$ and arbitrary vector *a*, i.e., $\mathrm{a}^T \mathrm{C}_e \mathrm{a} \geq 0$. This section presents the proposed methodology, which uses LMC for each anisotropy direction while maintaining $\mathbf{B}^l$. By linearly combining semivariograms and covariances for multiple anisotropy directions, it ensures that C_e and R_e remain non-negative definite under various conditions (Goovaerts, 1997).

3 Anisotropic spatial cross-correlation models for the investigated IMs

3.1 Data preparation

Following Wang and Du (2013), three selection criteria are adopted to select appropriate ground-motion records:1) The seismic record data were dense, with a sufficient number of records (more than 30) in each separate distance bin to obtain statistically reliable sample sizes; 2) Necessary seismic and geological information were provided for the selected earthquake events, including source parameters, distance from the site-to-source, and site conditions (Vs_{30} value), so that the median IM and its residuals could be estimated from GMPEs, which stand for Ground-Motion Prediction Equations, also known as ground-motion models (GMMs); 3) The moment magnitude of the earthquake events was greater than 5 and had an adequate number of recorded stations in terms of quantity and distance, making them suitable for spatial correlation analysis. Then a total of 10 earthquake events comprising 2942 high-quality and reliable seismic records were extracted, excluding those with low quality, unreliability, incompleteness, and even aftershock records, from the NGA-West2 database (Pacific Earthquake Engineering Research Center, PEER). List of the earthquake events can be found in Table 1. The distribution of moment magnitude (M_w) and rupture distance (R_{rup}) of the data, as well as the distribution of recorded PGA values and the recorded station's R_{rup} and Vs_{30} values, are shown in Fig.4.

The spatial cross-correlation models for 16 vector-IMs, including $V_{Elr}(T)$ (the spectral ordinates at 5 periods, i.e., 0.05s, 0.2, 0.5s, 1s and 2s), $SA(T)$ (the spectral ordinates at 5 periods, i.e., 0.05s, 0.2, 0.5s, 1s and 2s), PGA, PGV, I_a, CAV, $D_{s5\text{-}75}$, and T_m, are developed based on the aforementioned selected seismic ground-motion data from various regions (i.e., California, Japan and Taiwan). It should be noted that since the seismic data used are based on reverse faults and strike-slip faults, the present anisotropic model is applicable to earthquakes generated by reverse faults and strike-slip faults. More attention is needed if the present model is to be applied to normal faults. In future work, the impact of faulting mechanism will be further analyzed and studied. Additionally, since only seismic data with a moment magnitude greater than 5 and a rupture distance less than 200 km were selected, when applying the newly proposed model, it is important to consider whether the fault type, moment magnitude, and fault distance fall within the applicability range of the model.

As suggested by Abrahamson and Youngs (1992), IMs can be modelled by a lognormal distribution when developing the GMPEs. Hence, for a given earthquake event *i*, the IMs at seismic site *j* can be given as:

$$lnIM_{ij} = \overline{lnIM_{ij}(M,R_{rup},\theta)} + \varepsilon_T\sigma_T = \overline{lnIM_{ij}(M,R_{rup},\theta)} + \eta_i + \varepsilon_{ij} \qquad (19)$$

Where ln denotes the natural logarithm of recorded IM given an earthquake event *i* at seismic site *j*; $\ln IM_{ij}(M,\mathrm{R_{rup}},\theta)$ denotes the logarithm of the predicted median from GMPEs based on magnitude (*M*), R_{rup} and other relative parameters (e.g., location), given an earthquake event *i* at seismic site *j*; σ_T denotes the total residual standard deviation and $\sigma_T = \sqrt{\sigma_{ij}^2 + \tau_i^2}$; ε_T denotes the normalized residual standard deviation; η_i and ε_{ij} represent the between-event and within-event residuals, respectively.

The adopted GMPEs assume that the term of residual standard follows a normal distribution, whilst τ_i and σ_{ij} respectively denote zero mean values and standard derivations (Abrahamson et al., 1991; Jayaram and Baker, 2008; Joyner and Boore, 1993). Assuming that the effects of between-event residuals on the within-event spatial correlation are negligible (Du and Wang, 2013; Jayaram and Baker, 2009), the within-event residual can be calculated as:

$$\varepsilon_{ij} = lnIM_{ij} - \overline{lnIM_{ij}(M,R_{rup},\theta)}$$

(20)

A series of GMPEs developed based on the NGA-West2 database were chosen to calculate the within-event residuals by using Equation 19. The definitions of IMs, together with calculations of GMPEs are illustrated in Table 2.

The calculated within-event residuals can exhibit discrepancies with respect to R_{rup} and Vs_{30}, since the GMPEs are not corrected by using the seismic ground-motion data selected in this study. The discrepancies may inevitably result in inaccuracies in estimating spatial correlations (Foulser-Piggott and Stafford, 2012; Sokolov et al., 2010). Therefore, Du and Wang (2013) proposed a correction of the calculated residuals to eliminate the bias, which has been applied in this study:

$$\varepsilon_{ij}^{corr} = \varepsilon_{ij} - [\varphi_1 + \varphi_2 \ ln(R_{rup}) + \varphi_3 \ ln(V_{S30})] \quad (21)$$

Where φ_1, φ_2 and φ_3 are coefficients for each earthquake event which are obtained through linear regression. The deviations of the corrected V_{Elr} (0.05s) within-event residuals with respect to R_{rup} and Vs_{30} for ten earthquake events are illustrated in Fig.5.

As shown in Fig.5, the corrected residuals do not bias towards R_{rup} and V_{s30} . The corrected residuals will be adopted in the analysis of semi-variogram. The normalized corrected intra-event residuals can be calculated as follows:

$$\varepsilon'_{ij} = \frac{\varepsilon_{ij}^{corr}}{\sigma_{ij}} \approx \frac{lnIM_{ij} - \overline{lnIM_{ij}(M,R_{rup},\theta)} - [\varphi_1 + \varphi_2 \;\; ln(R_{rup}) + \varphi_3 \;\; ln(V_{S30})]}{\sigma_{ij}} \quad (22)$$

Where ε'_{ij} is the normalized within-event residuals; σ_{ij} represents the within-event standard deviation that can be obtained through the sample data or a specific GMPE. In this study, σ_{ij} adopts the variance of within-event residuals of IMs for each earthquake, since the values provided by GMPEs may not be applicable to all considered earthquake events. Equation 22 neglects the interevent residual term because it is constant for each site during an earthquake and therefore does not contribute to the within-event spatial correlation (Wang and Du, 2013). The inconsistencies in GMPEs are expected to have minimal impact on the estimated spatial correlation Du and Wang (2013).

It is worth mentioning that the multi-stage method used by Jayaram and Baker (2009), as well as by Wang and Du (2013), is applied to adjust the residuals in order to remove bias arising from the inconsistency between the GMPE and the empirical data utilized in their analysis. The multi-stage method is widely applied. However, it is important to highlight that Ming et al. (2019)

have recently introduced a one-stage method that has been shown to more accurately estimate all parameters in GMPEs with spatial correlation simultaneously. Utilizing this method may enhance the unbiased residuals for correlation modeling in future work.

3.2 Anisotropic spatial cross-correlation models for IMs

In this section, the anisotropic spatial cross-correlation empirical models for within-event residuals of the investigated IMs are developed, following the theoretical experience described in Section 2.1-2.3 and procedures detailed in Section 2.4.

3.2.1 Anisotropy considering four anisotropy directions

In this section, four anisotropy directions (i.e., θ = 0°, 45° ,$90^{\circ}$and 135°), an angle tolerance (Δ) of 45°, and bandwidth of 10 km to identify all eligible station locations within the specified range, the empirical semi-variogram values are calculated with Equation 3. For each earthquake event, different lag distances are adopted to ensure a sufficient number of station-point pairs within each lag distance (15 data points within the varying lags). Categorizing based on the considered lag distance and repeating this process for all stations can result in the empirical data of the semi-variogram fitting maps for 10 earthquakes selected in Table 1.

3.2.2 LMC Fitting to empirical for different anisotropy directions

The empirical semivariogram maps are first plotted and observed for 16 IMs ($V_{EIr}(T)$ (the spectral ordinates at 5 periods, i.e., 0.05s, 0.2, 0.5s, 1s and 2s), *SA(T)* (the spectral ordinates at 5 periods, i.e., 0.05s, 0.2, 0.5s, 1s and 2s), I_a, PGV, PGA, CAV, $D_{s5\text{-}75}$ and T_m) of 10 investigated earthquakes for four anisotropy directions (θ) of 0°, 45° ,$90^{\circ}$and 135°. Fig. 6 presents an example of the LMC model fitted to empirical semivariogram data at an anisotropy direction of $0°$ (the rest figures are given in the Supplemental Material, Figs. A1-A3). While the LMC curve approximates most of the empirical semivariogram values, certain values are not captured by the fitted model, likely reflecting the complexity introduced by the ten distinct seismic events considered in the analysis. It is worth noting that the empirical semi-variograms for $D_{s5\text{-}75}$ and T_m in relation to some spectral IMs exhibit negative values. Likewise, other research (e.g., Bradley, 2011; Huang et al., 2020; Baker and Bradley, 2017) has also observed negative correlations between these IMs. Given the constraints of space but the significance of these matters regarding negative correlations and semi-variograms, a detailed discussion is provided in Appendix A.

Following the procedure described in Section 2.4.3: the initial r_1^{θ} and r_2^{θ} within the LMC model for each anisotropy direction θ, are established; then, the empirical semivirogram data for each anisotropy direction are fitted by LMC model as given in Equation 14; the seimirogram, variance, and correlation matrix functions are determined for each anisotropy direction by testing different r_1^θ and r_2^θ; in the end, the average of the matrix functions are determined.

Following the procedure described in Section 2.4.3, the short-range coefficient r_1^θ is determined to be 5 km, 15 km, 10 km, and 10 km for the 0°, 45°, 90°, and 135° directions, respectively. The long-range coefficient r_2^θ is determined to be 70 km, 60 km, 65 km, and 80 km, respectively. The coregionalization matrices $\mathbf{B}^l$ within the LMC for the anisotropy directions are regressed, and their average matrices are obtained. As the result, the nugget coregionalization matrix P^0, the short-range coregionalization matrix $\mathbf{P}^1$, the long-range coregionalization matrix P^2 are provided in Tables 3-5, respectively. These coregionalization matrices can then be used to construct the resulting cross-covariance matrix for sampling intra-event residuals that exhibit anisotropic spatial correlation.

3.2.3 Applicability of the proposed anisotropic spatial cross-correlation models

The proposed spatial cross-correlation models are user-friendly for engineering applications. Below are the steps for estimating the cross-correlation of IMs across multiple locations: 1) calculate the separation distance h and determine the anisotropy direction θ (0°, 45°, 90°, or 135°) between every pair of two sites; 2) obtain the corresponding coregionalization matrix coefficients from Tables 3-5 for IMs; 3) compute the correlation coefficient through Equation 14 for each pair of two sites for all the IMs; 4) construct the final correlation matrix through Equation 16 using the elements obtained from Step 3).

A simple case has been given in this section to demonstrate the applicability and validity of the proposed anisotropic spatial cross-correlation models. A regional area of 30 km×30 km is assumed and is divided into 900 units of 1 km^2. Assuming an earthquake with M_w=6.5 occurs along a strike-slip fault, and its epicenter is located at coordinates (0, 0). The direction of the fault follows the north-south orientation, whilst homogeneous ground conditions are assumed with a specified Vs_{30} value of 500 m/s.

First, the earthquake catalog obtained through the Monte Carlo method includes seismic information. Based on the parameters in the simulated earthquake catalog, the mean and variance of three IMs, including the PGA, PGV, and I_a, are calculated using the GMPEs proposed by Campbell and Bozorgnia (2014) (CB14). Then, considering the variance of IMs and the correlation coefficients with spatial cross-correlation computed by those empirical models considering isotropy and anisotropy of the IMs, the residuals of IMs with spatial correlation features are simulated according to the multivariate normal distribution. These residuals, along with the mean values of IMs calculated by using GMPEs, are used to derive simulated values of IMs with spatial correlations. The prediction results are shown in Figs. 7-8.

As shown in Figs. 7-8, compared to the isotropic model, the anisotropic model demonstrates a stronger spatial correlation of the simulated IMs at zero distance. And the anisotropic model exhibits a notably higher spatial correlation intensity at 45 degrees than in other directions.

4 Case study: regional seismic landslide hazard analysis considering the anisotropic spatial cross-correlations of IMs

Seismic hazard analysis has always been a hot topic in geotechnical engineering research (Cheng et al., 2023; Yuan et al., 2024). This section will conduct seismic landslide hazard analysis based on the consideration of spatial cross-correlation of IMs. Taking the seismic area, Wenchuan city, located in Sichuan Province, Southwest of China, as an example, the maximum seismic magnitude is assumed as Ms=8.0, which is acceptable since the notable most heavy earthquake in history is Wenchuan Earthquake happened in 2008 with Ms equal to 8.0. Guo et al. (2013) proposed a relation between the earthquake magnitude and rupture distance to calculate the rupture distance:

$$M_s = 3.3 + 2.1 \ lgL \tag{23}$$

Where M_s is the surface wave magnitude; L denotes the rupture distance.

Based on Equation 23, the rupture distance is calculated as 173 km given that M_s=8.0. This study then references to this value and assumes the rupture distance to be 170km. With respect to the fault site, it is assumed that there exists a strike-slip fault with a fault angle of 90°, located in Southwest China. Following Wang (2022), three different locations of slope sites are taken as case study to perform the regional seismic hazard assessment of slopes considering the anisotropic cross-correlations of IMs. Fig.9 illustrates the fault site and three slope sites, where the vertical distances between the fault site and slope sites are 10km (Site 1), 30km (Site 2) and 60km (Site 3), evaluating effects of different rupture distance to the seismic slope hazard analysis.

According to National Earthquake Hazards Reduction Program (NEHRP) and actual Vs_{30} (average shear wave velocity of soil layers within 30m below the ground) recorded in NGA-West2, Seyhan and Stewart (2014) suggested 5 types of sites, which is shown in Table 6. This study adopted three representative sites to investigate the effects of sites, i.e., Vs_{30}=913 m/s (Site Class B), Vs_{30}=489 m/s (Site Class C) and Vs_{30}=266 m/s (Site Class D).

Seismic catalog is essential data for analyzing regional seismic hazards

(Shao, 2018). In this study, the Monte Carlo method is chosen to simulate and generate a stochastic earthquake catalog. In order to study the impact of major earthquakes on seismic slopes, the truncated Rutenberg-Richter model is employed as the seismic recurrence model, as shown in the following equation:

$$lgN(M_w) = 3.86 - 0.83M_w (4.0 \leq M_w \leq 8.0) \tag{24}$$

A total of 1,732,809 earthquakes with a magnitude $M_w \geq 4.0$ were generated through simulations of earthquake records with a recurrence period of 100 years and 50000 occurrences by using the Monte Carlo method by introducing either isotropy or anisotropic models of IMs proposed in this study.

4.1 XGB-permanent slope displacement prediction model

Following the selection criteria of critical acceleration employed by Liu et al. (2017) and Li et al. (2019), two critical yield accelerations (a_c), i.e., 0.02g and 0.1g, are adopted in the present study. These values reflect the decreasing susceptibility of seismic slopes.

The machine learning framework based on multiple ground motion parameters for predicting permanent landslide displacement (referred to as the XGB model), which was proposed by Wang et al. (2020) is adopted in this study. This model demonstrates strong generalization abilities in displacement prediction and, by optimizing hyperparameters, mitigates the risk of overfitting. The data-driven Newmark displacement prediction model developed in this approach better adheres to sufficiency and efficiency standards. In comparison to traditional empirical models, the generated standard deviation is significantly smaller. Additionally, the application of the model in probabilistic seismic slope displacement hazard analysis is also demonstrated. In order to better apply the XGB model to practical use, the fitting of the residuals is shown as:

$$\sigma_{lnD} = \begin{cases} 0.798 & D_{pred} \leq 0.003cm \\ -0.147log_{10}(D_{pred}) + 0.427 & 0.003 \leq D_{pred} \leq 20cm \\ 0.235 & D_{pred} > 20cm \end{cases} \quad (25)$$

Where σ_{lnD} is the standard deviation; D_{pred} is the predicted displacement.

In order to evaluate the effects of anisotropic spatial cross-correlations of vector-IMs on the regional hazard assessment of seismic slopes, four conditions have been analyzed, and those are the isotropic spatial cross-correlation model (Cheng et al., 2020), the anisotropic spatial cross-correlation model with four anisotropy directions, and without considering spatial cross-correlations of IMs (Not relevant).

4.2 Full probabilistic seismic landslide hazard analysis

The XGB permanent displacement prediction model, which is a vector model, require input parameters consist of PGA, PGV and I_a. It is necessary to consider the joint density function of scalar IMs, and the probability of exceedance can be given as:

$$\lambda_D = \int P(D \geq x | IM_1 = y,\ IM_2 = z,\ IM_3 = m)\, f_{IM_1, IM_2, IM_3}(y, z, m)\, dm\, dz\, dy \tag{26}$$

Where λ_D represents a mean annual rate of sliding displacement *D* exceeding that specified value; using a series of specified value (*x*), e.g., from 0 to 100cm at interval of 1cm, seismic hazard curves of permanent sliding displacements can be obtained. $P(D \geq x | IM_1 = y,\ IM_2 = z,\ IM_3 = m)$ indicates the probability of sliding displacement larger than a given value *x* when IM_1=*y*, $IM_2 = z$ and IM_3=*m*; $f_{IM_1, IM_2, IM_3}(y, z, m)$ is the probabilistic density function of IM_1=*y*, $IM_2 = z$ and IM_3=*m*.

In addition, the annual exceedance rate calculation for the joint seismic slope-site permanent displacement is (Akkar and Cheng, 2016):

$$\lambda_D = \frac{all\ sites\ D > x\ events}{total\ number\ of\ events\ that\ simulate\ all\ sites \times statistical\ period} \tag{27}$$

Different annual exceedance rates correspond to different values of slope permanent displacement, representing various levels of hazard. Therefore, it is possible to assess the hazard based on the calculated permanent displacement values according to the specific slope.

The Monte Carlo simulation for the Equation 26 is employed to study the regional seismic hazard assessment of slopes that subjected to different critical yield acceleration (a_c), site classification, and rupture distance.

4.3 Results and discussions

Assessing seismic risk and hazard at a regional scale differs from evaluating an individual site (e.g., Jeon and O'Rourke, 2005; Wang and Takada, 2005). The joint hazard is of importance to be addressed, whilst the single hazard for an individual site is negligible (Du and Ning, 2020). Hence, the joint seismic hazard curves for Site 1, Site 2 and Site 3 are evaluated through the annual exceedance rate calculation given in Equation 27 for different site classes. The results are plotted in Fig.10, illustrating the mean annual exceedance rate of the slope permanent displacements for two critical accelerations (a_c=0.02g and 0.1g), considering the isotropy and anisotropy of the cross-correlations of IMs. The hazard curves for isotropy are computed using spatial cross-correlation model proposed by Cheng et al. (2020), and those for anisotropy are computed based on the model proposed in this study.

As shown in Fig. 10, the solid line indicates the results of isotropy, the dashed line represents the results of anisotropy, and the dot-dashed line represents the condition without considering spatial cross-correlations of IMs (Not relevant). It is evident that ignoring the spatial cross-correlations of IMs would underestimate the hazard of the slope permanent displacement. As soil conditions soften, as seen in Site Class D, the divergence between hazard displacement estimates from isotropic and anisotropic spatial correlations becomes more pronounced, especially at lower mean annual exceedance rates and a_c=0.1. These results highlight the importance of accounting for the anisotropic characteristics of intensity measures when analyzing soft soils.

5 Conclusion

Although anisotropy of spatial cross-correlations of IMs has long been identified, only a few IMs correlation models considering anisotropy have been developed. This study proposed the anisotropic spatial cross-correlation models for 16 common IMs, including $V_{EIr}(T)$ (the spectral ordinates at 0.05s, 0.2, 0.5s, 1s and 2s), SA(T) (the spectral ordinates at 0.05s, 0.2, 0.5s, 1s and 2s), PGA, PGV, I_a, CAV, D_{s5-75}, and T_m. The model for four anisotropy directions (0°, 45°,90°and 135°) inclined to the fault was developed. In total, 2942 ground-motion records selected from 10 earthquake events were used to compute the empirical semivariograms for the within-event residuals of the selected 16 IMs. The commonly used LMC function was employed to construct permissible spatial correlation models. It was observed that for 16 IMs, the developed anisotropic spatial cross-correlation models would fit the empirical semivariograms and cross-semivariograms reasonably well.

The proposed anisotropic spatial cross-correlation model of IMs was applied to regional seismic landslide hazard assessment. Results show that at lower hazard levels, the annual exceedance rate curves for isotropic and anisotropic models align closely. However, anisotropic effects become more significant for soft soils as the annual exceedance rate decreases, while they are negligible for slopes with harder site classes and higher critical accelerations. Therefore, the anisotropy of IMs cannot be ignored in regional seismic landslide hazard analysis, particularly for soft soils and lower critical accelerations.

Declaration of competing interest

No potential conflict of interest was reported by the authors.

Data availability

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

Acknowledgements

The authors sincerely thank the anonymous reviewers and the editor for their valuable comments and suggestions, which have greatly improved the quality of this work. We acknowledge the financial support from the National Natural Science Foundation of China Grant (Nos. 52278413, 42011530170, 52278522; 52361135804), China Postdoctoral Science Foundation (Grant No. 2021M702718).

Reference

Abrahamson N., J. F. Schneider and J. C. Stepp. 1991. "Spatial coherency of shear waves from the Lotung, taiwan large-scale seismic test" *Structural Safety* 10 (1): 145-162. https://doi.org/10.1016/0167-4730(91)90011-W.

Abrahamson N. A. and R. R. Youngs. 1992. "A stable algorithm for regression analyses using the random effects model" *Bulletin of the seismological society of America* 82 (1): 505-510. https://doi.org/10.1785/BSSA0820010505.

Akkar S. and Y. Cheng. 2016. "Application of a Monte-Carlo simulation approach for the probabilistic assessment of seismic hazard for geographically distributed portfolio" *Earthquake Engineering&Structural Dynamics* 45 (4): 525-541. https://doi.org/10.1002/eqe.2667.

Abbasnejadfard M., M. Bastami and A. Fallah. 2020. "Investigation of anisotropic spatial correlations of intra-event residuals of multiple earthquake intensity measures using latent dimensions method" *Geophysical Journal International* 222 (1): 1449-1469. https://doi.org/10.1093/gji/ggaa255.

Abbasnejadfard M., M. Bastami and A. Fallah. 2021a. "Investigating the spatial correlations in univariate random fields of peak ground velocity and peak ground displacement considering anisotropy" *Geoenvironmental Disasters* 8 (1): 24. https://doi.org/10.1186/s40677-021-00196-w.

Abbasnejadfard M., M. Bastami, A. Fallah and A. Garakaninezhad. 2021b. "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" *Bulletin of Earthquake Engineering* 19 (14): 5791-5817. https://doi.org/10.1007/s10518-021-01203-z.

Boore D. M., J. F. Gibbs, W. B. Joyner, J. C. Tinsley and D. J. Ponti. 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 (6): 2737-2751. https://doi.org/10.1785/0120020197.

Bradley, B. A. 2011. Correlation of significant duration with amplitude and cumulative intensity measures and its use in ground motion selection. *Journal of Earthquake Engineering* 15(6): 809-832. https://doi.org/10.1080/13632469.2011.557140.

Bowman A. W. and R. M. Crujeiras. 2013. "Inference for variograms" *Computational Statistics&Data Analysis* 66 19-31. https://doi.org/10.1016/j.csda.2013.02.027.

Cressie N. 1993. "Statistics for spatial data: Wiley series in probability and statistics.".

Campbell K. W. and Y. Bozorgnia. 2014. "NGA-West2 Ground Motion Model for the Average Horizontal Components of PGA, PGV, and 5% Damped Linear Acceleration Response Spectra" *Earthquake Spectra* 30 (3): 1087-1115. https://doi.org/10.1193/062913EQS175M.

Cheng Y., C. L. Ning and W. Du. 2020. "Spatial Cross-Correlation Models for Absolute and Relative Spectral Input Energy Parameters Based on Geostatistical Tools" *Bulletin of the seismological society of America* 110 (6): 2728-2742. https://doi.org/10.1785/0120200142.

Cheng Y., J. Wang and Y. He. 2023. "Prediction Models of Newmark Sliding Displacement of Slopes Using Deep Neural Network and Mixed-effect Regression" *Computers and geotechnics* 156 105264. https://doi.org/https://doi.org/10.1016/j.compgeo.2023.105264.

Du W. and G. Wang. 2013. "Intra-Event Spatial Correlations for Cumulative Absolute Velocity, Arias Intensity, and Spectral Accelerations Based on Regional Site Conditions" *Bulletin of the seismological society of America* 103 (2A): 1117-1129. https://doi.org/10.1785/0120120185.

Du W. and C.-L. Ning. 2020. "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 (1): 486-504. https://doi.org/10.1177/8755293020952442.

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 (5): 2532-2541. https://doi.org/10.1785/0120110117.

Foulser-Piggott R. and P. J. Stafford. 2012. "A predictive model for Arias intensity at multiple sites and consideration of spatial correlations" *Earthquake Engineering&Structural Dynamics* 41 (3): 431-451. https://doi.org/10.1002/eqe.1137.

Goulard, M. and M. Voltz. 1992. Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. *Mathematical Geology*, 24: 269-286. https://doi.org/10.1007/BF00893750.

Goovaerts P. 1997. *Geostatistics for Natural Resources Evaluation*: Oxford University Press . 10.1093/oso/9780195115383.001.0001.

Goda K. and H. P. Hong. 2008. "Spatial Correlation of Peak Ground Motions and Response Spectra" *Bulletin of the seismological society of America* 98 (1): 354-365. https://doi.org/10.1785/0120070078.

Guo A.-N., Z.-J. Guo, J. Z, C.-C. Z and Y.-C. Z. 2013. "Research on the effect of seismogenic structure with complex spatial relationship on earthquake mechanism with the Wenchuan earthquake as an example" *China Earthquake Engineering Journal* 35 (01): 196-202.

https://doi.org/10.3969/j.issn.1000-0844.2013.01.0196. (In Chinese).

Genton M. G. and W. Kleiber. 2015. "Cross-Covariance Functions for Multivariate Geostatistics" *Statistical Science* 30 (2): 147-163. https://doi.org/10.1214/14-STS487.

Garakaninezhad A. and M. Bastami. 2017. "A Novel Spatial Correlation Model Based on Anisotropy of Earthquake Ground-Motion Intensity" *Bulletin of the seismological society of America* 107 (6): 2809-2820. https://doi.org/10.1785/0120160367.

Garakaninezhad A. and M. Bastami. 2019. "Intra-event spatial correlation model for the vertical component of response spectral accelerations" *Journal of Seismology* 23 (4): 853-867. https://doi.org/10.1007/s10950-019-09838-x.

Journel A. G. and C. J. Huijbregts. 1976. *Mining geostatistics*. United Kingdom.

Joyner W. B. and D. M. Boore. 1993. "Methods for regression analysis of strong-motion data" *Bulletin of the seismological society of America* 83 (2): 469-487. https://doi.org/10.1785/BSSA0830020469.

Jeon S.-S. and T. D. O'Rourke. 2005. "Northridge Earthquake Effects on Pipelines and Residential Buildings" *Bulletin of the seismological society of America* 95 (1): 294-318. https://doi.org/10.1785/0120040020.

Jayaram N. and J. W. Baker. 2008. "Statistical Tests of the Joint Distribution of Spectral Acceleration Values" *Bulletin of the seismological society of America* 98 (5): 2231-2243. https://doi.org/10.1785/0120070208.

Jayaram N. and J. W. Baker. 2009. "Correlation model for spatially distributed ground-motion intensities" *Earthquake Engineering&Structural Dynamics* 38 (15): 1687-1708. https://doi.org/10.1002/eqe.922.

Liu J.-m., T. Wang, J.-s. Shi and Z.-t. Li. 2017. "Emergency rapid assessment of landslides induced by the JIUZHAIGOU MS 7.0 earthquake, Sichuan, China" *Journal of Geomechanics* 23 (5): 639-645.(In Chinese).

Li X., M. Gao and W. Xu. 2019. "Probabilistic seismic slope displacement hazard analysis based on Newmark displacement model：Take the area of Tianshui，Gansu Province，China as an example" *Acta Seismologica Sinica* 41 (6): 795-807. https://doi.org/10.11939/jass.20180075. (In Chinese).

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).

Sokolov V. and F. Wenzel. 2013. "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" *Bulletin of Earthquake Engineering* 11 (6): 1909-1926. https://doi.org/10.1007/s10518-013-9493-9.

Seyhan E. and J. P. Stewart. 2014. "Semi-Empirical Nonlinear Site Amplification from NGA-West2 Data and Simulations" *Earthquake Spectra* 30 (3): 1241-1256. https://doi.org/10.1193/063013EQS181M.

Shao X.-Y. 2018 "Study on Monte Carlo Simulation Based Probabilistic Seismic Hazard Analysis" Master's thesis, Institute of Earthquake Science China Earthquake

Administration. https://kns.cnki.net/kcms2/article/abstract?v=IILC1c-FiAFqudLeCTL9xL5cdv-G9N6wxRLe4gUsulGP253yG0l5uHMfOTD2ro4fzhYGqj4OEihUa_acEqPuElSS6bDIZVbiPTNpgvFgAJDE3GOnh3UKwjwAA2eDjMMW2XFejbL2SyylbyNuwTEY0w==&uniplatform=NZKPT&language=CHS.

Schiappapietra E., S. Stripajová, P. Pažák, J. Douglas and G. Trendafiloski. 2022. "Exploring the impact of spatial correlations of earthquake ground motions in the catastrophe modelling process: a case study for Italy" *Bulletin of Earthquake Engineering* 20 (11): 5747-5773.

Wang M. and T. Takada. 2005. "Macrospatial Correlation Model of Seismic Ground Motions" *Earthquake Spectra* 21 (4): 1137-1156. https://doi.org/10.1193/1.2083887.

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 (6): 3189-3204. https://doi.org/10.1785/0120130061.

Wang J.-F. 2022 "Probabilistic seismic landslide hazard analysis based on deep neural network and mixed-effects prediction model" Master's thesis, Southwest Jiaotong University. https://link.cnki.net/doi/10.27414/d.cnki.gxnju.2022.001799.

Wang M.-X., D. Huang, G. Wang and D.-Q. Li. 2020. "SS-XGBoost: A Machine Learning Framework for Predicting Newmark Sliding Displacements of Slopes" *Journal of Geotechnical and Geoenvironmental Engineering* 146 (9): 04020074. https://doi.org/10.1061/(ASCE)GT.1943-5606.0002297.

Yuan R., Y. Qiu and Y. Cheng. 2024. "A Framework of Probabilistic Seismic Risk Analysis of Urban Shallow-Buried Tunnels Subjected to Active Fault Dislocations" *Journal of Earthquake Engineering* 28 (5): 1312-1330. https://doi.org/10.1080/13632469.2023.2240431.

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(1): 145-156. https://doi.org/10.1193/060716eqs095m.

Huang C., K. Tarbali and C. Galasso. 2020. "Correlation properties of integral ground-motion intensity measures from Italian strong-motion records" *Earthquake Engineering&Structural Dynamics* 49(15): 1581-1598. https://doi.org/10.1002/eqe.3318.

Appendix A

In the study, the negative empirical values were observed for semi-variograms and correlations between non-spectral Intensity Measures (IMs), such as $D_{s5\text{-}75}$

and T_m, and some spectral IMs, such as spectral acceleration (SA) and spectral input energy (V_{EIr}). These issues are examined and analyzed in this appendix.

For a single variable *m*, the empirical semivariogram γ_{mm} is calculated as follows:

$$\gamma_{mm}(\mathrm{h}) = \frac{1}{2} \frac{1}{|N(\mathrm{h})|} \sum_{i=1}^{|N(\mathrm{h})|} [(Z_m(\mathrm{u}_i) - Z_m(\mathrm{u}_i + \mathrm{h}))^2] \quad \text{(A1)}$$

As shown in the Equation (A1), the γ_{mm} cannot be negative. However, for the multiple variables, the empirical cross-seimivariogram is calculated as follows:

$$\gamma_{mw}(\mathrm{h}) = \frac{1}{2} \frac{1}{|N(\mathrm{h})|} \sum_{i=1}^{|N(\mathrm{h})|} [(Z_m(\mathrm{u}_i) - Z_m(\mathrm{u}_i + \mathrm{h}))(Z_w(\mathrm{u}_i) - Z_w(\mathrm{u}_i + \mathrm{h}))] \quad \text{(A2)}$$

As shown in the Equation (A2), the γ_{mw} could be negative. If the differences between the two variables exhibit opposite trends in space (i.e., when one variable increases, the other decreases), the product can be negative, leading to a negative $\gamma_{mw}(\mathbf{h})$. In this study, this situation is likely to occur when a negative empirical cross-semivariogram was observed between some IMs, such as D_{s5-75} and SA (0.2s). It is different from the ordinary semivariogram, which is based on the squared differences of a single variable and is always non-negative.

This study employs a spatial cross-correlation model. The spatial cross-correlation coefficient is calculated through the following general equations:

$$C_{mw}(\mathrm{h}) = C_{mw}(0) - \gamma_{mw}(\mathrm{h}) \tag{A3}$$

$$\lim_{h\to\infty} \gamma_{mw}(\mathrm{h}) = C_{mw}(0) \tag{A4}$$

$$\rho_{mw}(\mathrm{h}) = \frac{C_{mw}(0)}{\{C_{mm}(0) \times C_{ww}(0)\}^{1/2}} - \frac{\gamma_{mw}(\mathrm{h})}{\{\mathrm{C}_{mm}(0) \times \mathrm{C}_{ww}(0)\}^{1/2}} \tag{A5}$$

Equations (A3) and (A4) indicate that if $\gamma_{mw}(\mathrm{h})$ is positive, $C_{mw}(0)$ must also be positive, with $C_{mw}(0) \geq \gamma_{mw}(\mathrm{h})$. Under these conditions, a negative correlation coefficient $\rho_{mw}(\mathrm{h})$, as calculated using Equation (A5), would be impossible. However, if $\gamma_{mw}(\mathrm{h})$ is negative, $C_{mw}(0)$ would also be negative. In this case, both terms in the right side of Equation (A5) are negative, and the absolute value of $C_{mw}(0)$ exceeds that of $\gamma_{mw}(\mathrm{h})$, resulting in a negative value for $\rho_{mw}(\mathrm{h})$.

Additionally, to examine the above argument, we take the Chi-Chi earthquake as an example where the parameters D_{s575} and PGA are selected at a 90° anisotropic direction.

$$C_{mw}(0) = -0.43,\ C_{mm}(0) = 0.72,\ C_{ww}(0) = 0.72$$

h=5km: $\gamma_{mw}(5) = -0.06$

$$C_{mw}(5) = C_{mw}(0) - \gamma_{mw}(5) = -0.37$$

$$\rho_{mw}(5) = \frac{C_{mw}(0)}{\{C_{mm}(0) \times C_{ww}(0)\}^{1/2}} - \frac{\gamma_{mw}(5)}{\{\mathrm{C}_{mm}(0) \times \mathrm{C}_{ww}(0)\}^{1/2}} = -0.51$$

As stated above, the signs (positive or negative) of the semivariograms and correlations are interconnected. The negative semivariograms can lead to negative correlations between IMs. However, there are distinctions between theoretical and empirical semivariogram data. Theoretically, negative semivariograms are not acceptable. However, the empirical semivariograms may yield negative values, possibly due to data randomness and other physical processes that are not fully captured.

The negative correlation between non-spectral IMs and spectral IMs was also observed in previous studies (Bradley 2011; Baker and Bradley, 2017; Huang et al., 2020). However, the exact physical and engineering mechanisms behind the negative correlation are still not fully understood. Certain explanations are available. For instance, Huang et al. (2020) attribute the negative correlation between D_{S5-95} and the short-period spectral accelerations to the fact that ground motions with longer-than-expected durations often result in seismic energy being distributed over an extended timeframe. As a result, the likelihood of generating significant peak responses reduces in a damped oscillator.