



Spatial correlation of systematic effects of non-ergodic ground motion models in the Ridgecrest area

Chenyang Liu¹ · Jorge Macedo¹ · Nicolas Kuehn²

Received: 7 June 2021 / Accepted: 24 May 2022 / Published online: 2 July 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

The current state of the practice in probabilistic seismic hazard analysis (PSHA) employs ergodic ground motion models (GMMs), which assume that the ground motion variability observed in a global database is the same as the variability in ground motion at a single site-source combination. However, the fast-growing empirical ground motion databases indicate significant regional differences in ground motions due to repeatable and systematic source, path, and site effects. These systematic effects, which are spatially correlated, are not consistent with the ergodic assumption, promoting the transition to non-ergodic GMMs for PSHA. In this study, we use Gaussian processes with different covariance functions to model the spatial correlation structures of systematic source, path, and site effects for the Ridgecrest area. We compare the proposed correlation models for Ridgecrest with those previously developed for the ANZA array in terms of predictive performance. We also evaluate the effects of the cell-specific attenuation approach on the spatial correlation structures of path effects. We find that the spatial correlation of systematic source and path effects is best characterized by anisotropic non-stationary covariance functions in Gaussian processes. We also find that the cell-specific attenuation approach with squared grids has limitations in predicting path effects and does not affect the correlation structures significantly for the Ridgecrest database. Finally, comparisons with the ANZA array suggest that the spatial correlation structures of path effects derived from the ANZA array may be transferable to the Ridgecrest area, potentially due to their similarity in crustal heterogeneity.

Keywords Non-ergodic ground motion model · Spatial correlation · Non-stationary covariancemodel · Gaussian process

1 Introduction

Probabilistic seismic hazard analysis (PSHA) requires ground motion models (GMMs) to describe the probability distribution of ground motion intensity measures (IMs) for given earthquake scenarios and site conditions. The probability distribution of an IM is often

✉ Jorge Macedo
jorge.macedo@ce.gatech.edu

¹ Georgia Institute of Technology, Atlanta, GA, USA

² University of California, Los Angeles, California, USA

characterized by a log-normal distribution with a median and standard deviation which are estimated from GMMs as functions of earthquake characteristics (e.g., earthquake magnitude— M , rupture distance— R_{rup} , etc.) and site conditions (e.g., time-averaged shear wave velocity in the top 30 m of the soil profile— V_{s30}).

Current GMMs are based on the ergodic assumption that the distribution of ground motion IMs over time at a single site is the same as the distribution of ground motion IMs over space (Anderson and Brune 1999). Under this assumption, GMMs are developed using global databases, providing median and standard deviation estimates that are the same for a single source-site combination. However, the standard deviations of ergodic GMMs are affected by location-specific systematic and repeatable effects; hence they represent an inflated aleatory variability, which has become evident as regional differences in ground motions have been observed in recent studies (Walling 2009). As a result, the ergodic assumption leads to biased estimates of IM hazard curves due to an inadequate trade-off between aleatory variability and epistemic uncertainties, where the latter is associated with repeatable and systematic effects (Walling 2009). On the contrary, non-ergodic GMMs model location-specific systematic and repeatable effects as adjustments to the median ground motions and hence yield a better trade-off between aleatory variability and epistemic uncertainties. The reduced standard deviations (associated with the reduced aleatory variability) are compensated by additional epistemic uncertainties introduced during the estimation of systematic effects (Abrahamson et al. 2019; Landwehr et al. 2016; Liu et al. 2022; Lavrentiadis et al. 2021). Many studies show that non-ergodic GMMs have significant impacts on PSHA, as the median adjustments shift the hazard curves horizontally while smaller standard deviations lead to steeper hazard curves (e.g., Rodriguez-Marek et al. (2014); Stewart et al. (2017); Abrahamson et al. (2019)), which is schematically illustrated in Figure 1 (see details in the figure caption).

The fast-growing ground motion databases enable relaxing the ergodic assumption and transitioning from ergodic PSHA to non-ergodic PSHA, which requires estimating location-specific systematic source, path, and site effects. Several efforts with different levels of complexities have been made to develop non-ergodic GMMs that account for systematic and repeatable effects for non-ergodic PSHA. An initial effort is the so-called single-station-sigma GMMs (e.g., Rodriguez-Marek et al. (2013)), which estimate systematic site effects for well-recorded sites. Specifically, if repeated ground motion recordings from multiple earthquakes at a single site are available, then it is possible to estimate and remove

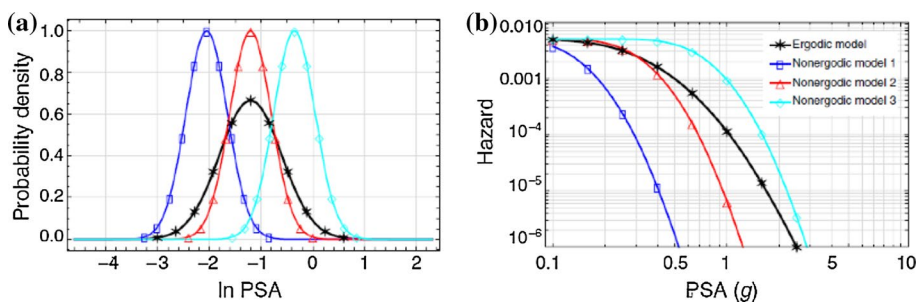


Fig. 1 **a** The probability density of an ergodic model is compared with the densities of three non-ergodic models with smaller standard deviations, but different medians. **b** The non-ergodic models lead to steeper and horizontally shifted hazard curves as compared to the ergodic one (adapted from Abrahamson et al. (2019))

the systematic site effect of that site from the residuals and hence reduce the aleatory standard deviation. Examples of single-station-sigma GMMs include models developed by Rodriguez-Marek et al. (2013), Abrahamson and Hollenback (2012), Rodriguez-Marek et al. (2011), Atkinson (2006), Chen and Tsai (2002), and Lin et al. (2011a). Another approach for developing non-ergodic GMMs is to regionalize the functional forms that represent the scaling of IMs. For example, Kuehn et al. (2020) developed a non-ergodic GMM with several regionalized coefficients (e.g., Japan, Taiwan, Southern America, etc.) to capture regional differences in the scaling of spectral accelerations. Dawood and Rodriguez-Marek (2013) developed a GMM for Japan with different anelastic attenuation scaling for different small regions, which is known as the cell-specific attenuation. In a different effort, Landwehr et al. (2016) developed a varying-coefficient GMM for California with coefficients varying smoothly by geographical coordinates of earthquakes and sites. Lanzano et al. (2021) developed a non-ergodic GMM with spatially varying coefficients for crustal earthquakes in Italy, using a multi-source geographically weighted regression (Caramenti et al. 2022). More recent studies decompose the residuals of ergodic GMMs into systematic and aleatory components for source, path, and site terms. The systematic effects are then modeled as functions of geographical coordinates, while the aleatory residuals represent the inherent randomness of the process. For example, Kuehn and Abrahamson (2020) developed ergodic GMMs for Taiwan and California from which residuals were calculated and used as inputs into Gaussian processes that evaluated the systematic source and path effects in the estimated residuals. The estimated systematic effects were subsequently used in the implementation of non-ergodic GMMs for Taiwan and California. In addition, Sgobba et al. (2021) developed a non-ergodic GMM by modeling the event, source, path, and site effects in the residuals of a region-specific GMM for central Italy, which was subsequently used to generate shaking maps.

Although significant efforts have been made to provide valuable insights into the transition from ergodic to non-ergodic GMMs, there are still several gaps to be bridged. For example, one of the challenges is the estimation of the spatial correlation structures for systematic source, path, and site effects, which are expected to affect PSHA and the risk assessment of engineering systems significantly (Sgobba et al. 2019, 2021; Park et al. 2007; Macedo et al. 2020, 2022; Liu et al. 2021; Patel et al. 2021; Ceferino et al. 2020). Hence, it is important to develop accurate and robust spatial correlation models for systematic effects. However, most studies developed spatial correlation models based on ergodic or partially ergodic assumptions, or have not fully exploited the functional forms of spatial correlation models. For example, Jayaram and Baker (2009) modeled the spatial correlation of within-event residuals of ergodic GMMs using an isotropic semi-variogram that solely depends on the distances between sites. However, the within-event residuals contain both systematic path and site effects that cannot be fully modeled by an isotropic semi-variogram alone (Kuehn and Abrahamson 2020; Kuehn et al. 2019; Villani and Abrahamson 2015; Lin et al. 2011b; Schiappapietra and Smerzini 2021). Foulser-Piggott and Goda (2015) evaluated the spatial correlation for site effects in single-station-sigma GMMs, but they did not investigate the correlation structures of source and path effects. More recently, Kuehn and Abrahamson (2020) used Gaussian processes with several covariance functions to estimate the spatial correlation of systematic source and path effects from the ANZA array (Berger et al. 1984) database. However, only a smaller number of correlation models were tested for path effects, and a single model was used for source effects. Schiappapietra and Douglas (2020) have conducted a comprehensive review of previously developed correlation models, among which most of them are developed based on semivariograms with isotropic and stationary exponential models. They also pointed out that a single rate of

decay of the correlation as a function of the inter-site separation distance may not be sufficient for seismic hazard and risk assessment. Moreover, most of the previous studies have used ground motions databases that are too sparse to estimate repeatable effects effectively (e.g., Jayaram and Baker (2009); Hong et al. (2009)). Lastly, the transferability of spatial correlation structures across different regions is not fully understood. Schiappapietra and Douglas (2020) observed that differences in ground motion database and regional geological conditions could result in a different rate of decay of the correlation with increasing inter-site separation distance, in the context of second-order stationary and isotropic correlation models. Kuehn and Abrahamson (2020) investigated the possibility of applying the correlation structures of non-stationary path-effect models developed using the ANZA array in California to Taiwan. They found this transfer of correlation structures (i.e., from California to Taiwan) promising if the magnitude extrapolation was properly accounted for. Even though the previous referred studies provided interesting insights, more efforts are required to investigate the transferability of correlation structures. Exploring the transferability of correlation structures is beneficial as it could reduce the computational efforts in PSHA assessments with non-ergodic GMMs and also improve the prediction of IMs for areas with scarce data (Abrahamson and Kuehn 2021).

In this study, we develop several spatial correlation models for systematic source, path, and site effects for peak ground acceleration (*PGA*) using the Ridgecrest database (Rekoske et al. 2020). The Ridgecrest database contains a dense set of ground motions in a localized region with a range of magnitudes, which is suitable for estimating systematic effects (Liu et al. 2022). We evaluate the efficacy of different correlation models based on their predictive performance on a hold-out set of ground motions representing future earthquakes. We also investigate the impact of the cell-specific attenuation approach (Dawood and Rodriguez-Marek 2013) on the correlation models for path effects. The performance of the different models developed in this study are compared, and insights from the comparison are shared. Finally, we also compare the performance of our models with the Kuehn and Abrahamson (2020) models, which are developed using a different database, to evaluate the transference of correlation structures across different regions.

2 Database

This study uses the Ridgecrest ground motion database developed by the United States Geological Survey (Rekoske et al. 2020), which contains 22,375 recordings at 968 stations from 133 earthquakes. In this database, multiple recordings are available from a single event or site, making it suitable for estimating the spatial correlation of systematic effects. The database is further filtered based on two criteria. First, ground motions recorded at distances larger than 200 km are removed as PSHA studies commonly exclude scenarios with rupture distances beyond 200 km. Second, an ergodic GMM (i.e., see the “Functional form of the ground motion model” section) is fitted using the database, and additional records with high leverages (Hastie et al. 2009) or abnormal residuals are removed. The filtering process yields a final subset of 12,612 recordings from 131 earthquakes at 458 stations with rupture distances up to 200 km and magnitudes between 3.6 to 7.1. To evaluate the performance of spatial correlation models, some earthquakes and stations are removed to a hold-out test set while the remaining recordings are used as a training set. The training set contains around 70% (9206 recordings from 87 events at 382 stations) while the test set contains around

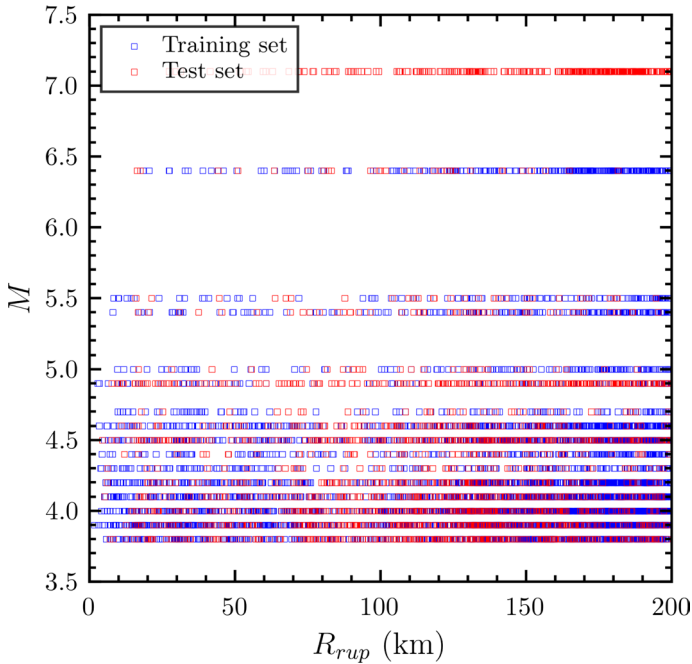


Fig. 2 Magnitude-distance distribution of ground motion recordings in the Ridgecrest database

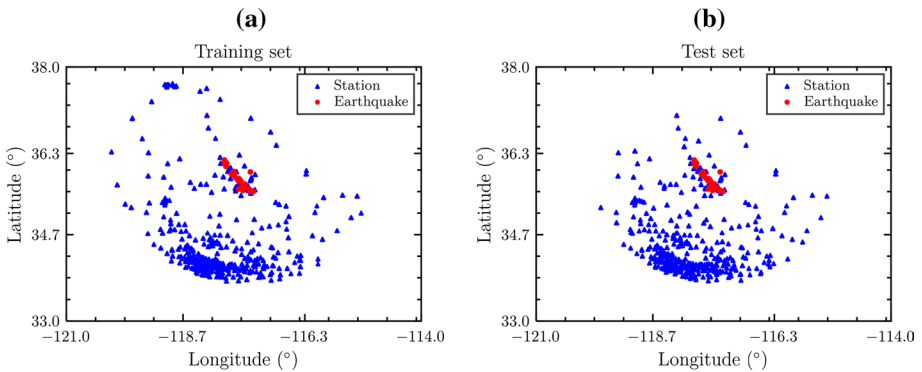


Fig. 3 Spatial distribution of earthquakes and sites in a the training set and b the test set

30% (3405 recordings from 73 events at 347 stations) of the ground motion recordings in the final subset. As elaborated upon later, correlation models are trained using the training set, and their performance is evaluated based on their predictions on the test set. Figures 2 and 3 show the magnitude-distance distribution and the spatial

distribution of sites and earthquake epicenters for ground motion recordings in this database, respectively.

3 Estimation of systematic effects

3.1 Components of ground motion residuals

Systematic effects can be estimated from residuals of an ergodic GMM, as shown in Eq. 1.

$$\ln Y = f_{ergodic}(M, R_{rup}, V_{s30}) + \delta B + \delta W \quad (1)$$

where Y is the IM of interest (i.e., PGA in this study), $f_{ergodic}$ is a function of earthquake characteristics (e.g., M , R_{rup} , etc.) and soil conditions (e.g., V_{s30} , etc.), which provides median estimates for the IM of interest. δB and δW are the between-event and within-event residuals, respectively, which can be estimated by random-effect regressions (e.g., Abrahamson and Youngs (1992)). δB and δW have Gaussian distributions with zero means and standard deviations of τ and ϕ , respectively.

Using a single-station-sigma model, δW can be divided into two components: the between-site residual δS , which contains systematic site effects; and the event-site-corrected residual δWS , which includes the systematic path effects. Hence, δB , δS , and δWS follow Gaussian distributions with zero means and standard deviations of τ , ϕ_S , and ϕ_{SS} , respectively, as shown in Eqs. 2 to 4.

$$\delta B \sim \mathcal{N}(0, \tau^2) \quad (2)$$

$$\delta S \sim \mathcal{N}(0, \phi_S^2) \quad (3)$$

$$\delta WS \sim \mathcal{N}(0, \phi_{SS}^2) \quad (4)$$

Following the notations in Atik et al. (2010), δB , δS , and δWS can be further partitioned into epistemic components accounting for the systematic effects and aleatory components representing the natural randomness of the process:

$$\delta B = \delta L2L + \delta B_0 \quad (5)$$

$$\tau^2 = \tau_{L2L}^2 + \tau_0^2 \quad (6)$$

$$\delta WS = \delta P2P + \delta WS_0 \quad (7)$$

$$\phi_{SS}^2 = \phi_{P2P}^2 + \phi_{0,SS}^2 \quad (8)$$

$$\delta S = \delta S2S + \delta S_0 \quad (9)$$

$$\phi_S^2 = \phi_{S2S}^2 + \phi_{0,S2S}^2 \quad (10)$$

where $\delta L2L$, $\delta P2P$, and $\delta S2S$ represent systematic source, path, and site effects, respectively. δB_0 , δWS_0 , and δS_0 are the remaining aleatory between-event, event-site-corrected, and between-site residuals, respectively. τ_{L2L} , ϕ_{P2P} , and ϕ_{S2S} are (epistemic) standard deviations of systematic source, path, and site effects; τ_0 , $\phi_{0,SS}$, and $\phi_{0,S2S}$ are standard deviations of aleatory between-event, event-site-corrected, and between-site residuals, respectively.

3.2 Modeling spatial correlation of systematic effects

We model systematic source, path, and site effects as functions of their geographic coordinates as defined in Eqs. 11 to 13:

$$\delta L2L = f_1(x_e) \tag{11}$$

$$\delta P2P = f_2(x_e, x_s) \tag{12}$$

$$\delta S2S = f_3(x_s) \tag{13}$$

in which x_e and x_s are column vectors containing geographical coordinates of earthquakes and sites, respectively. Since the functional forms of f_i 's (for $i = 1, 2$, and 3) are unknown, we assume they are drawn from a Gaussian process (GP) prior:

$$f_1(x_e) \sim GP(0, k(x_e, x'_e)) \tag{14}$$

$$f_2(x_e, x_s) \sim GP(0, k([x_e, x_s], [x'_e, x'_s])) \tag{15}$$

$$f_3(x_s) \sim GP(0, k(x_s, x'_s)) \tag{16}$$

where $k(x, x')$ is a covariance function that estimates the correlation between locations x and x' . The variations of f_i 's are controlled by the choices of $k(x, x')$. A covariance function $k(x, x')$ is called stationary if $k(x, x') = g(h)$ with $h = x' - x$ (i.e., the covariance only depends on the vector from x to x'). Further, if $k(x, x') = g(\|h\|)$ (i.e., the covariance only depends on the distance between x and x'), $k(x, x')$ is called isotropic. Hence, isotropy implies stationarity in covariance functions; namely, a non-stationary covariance function is also anisotropic. As detailed in later sections, we build different isotropic stationary and anisotropic non-stationary covariance functions and evaluate their performance on estimating spatial correlation of systematic effects. Modeling spatial correlation of systematic effects as GPs is an alternative to geostatistical methods such as fitting semivariograms, which have been used in previous studies (e.g., Jayaram and Baker (2009), Foulser-Piggott and Goda (2015), etc).

Considering that the aleatory residuals δB_0 , δWS_0 , and δS_0 are each independent and identically distributed Gaussian random variables, δB , δWS , and δS follow multivariate Gaussian distributions:

$$\delta B \sim \mathcal{N}(0, k(x_e, x'_e) + \delta_{ij}\tau_0^2) \tag{17}$$

$$\delta WS \sim \mathcal{N}(0, k([x_e, x_s], [x'_e, x'_s]) + \delta_{ij}\phi_{0,SS}^2) \tag{18}$$

$$\delta S \sim \mathcal{N}(0, k(x_s, x'_s) + \delta_{ij} \phi_{0, S2S}^2) \tag{19}$$

where δ_{ij} is the Kronecker delta with a value of 1 if two earthquakes, paths, or sites are the same in Eqs. 17, 18, and 19, respectively. When predictions of systematic effects at new locations are required, one can calculate the conditional distributions of systematic effects given the observed δB , δWS , and δS in the training set. Details on mathematical derivations for predictions are outlined in Rasmussen (2003) and Kuehn and Abrahamson (2020). In this study, we first compute the residuals of the Ridgecrest database (including both training and test sets) using an ergodic GMM. Then, we estimate correlation structures of systematic effects based on the residuals in the training set and evaluate their predictive performance on the test set.

4 Functional form of the ground motion model

We use residuals of the ergodic GMM defined in Eqs. 20 and 21 to model spatial correlation of systematic effects. This functional form has also been used by GeoPentech (2015) and Abrahamson et al. (2019) in previous studies for other areas in California.

$$f_{ergodic} = \theta_0 + g(M) + (\lambda_4 + \lambda_5(M - 5)) \ln \sqrt{R_{rup}^2 + \lambda_6^2} + \lambda_7 R_{rup} + \theta_8 Z_{hyp} + \lambda_9 F_n + \lambda_{10} F_{rv} + \theta_{11} \ln \frac{V_{s30}}{760} \tag{20}$$

$$g(M) = \begin{cases} -\lambda_1 + \lambda_2(M - 5.5), & \text{if } M < 5.5 \\ \lambda_1(M - 6.5), & \text{if } 5.5 \leq M \leq 6.5 \\ \lambda_3(M - 6.5), & \text{if } M > 6.5 \end{cases} \tag{21}$$

where Z_{hyp} is the hypocentral depth, F_n and F_{rv} are flags with values of 1 for normal and reverse fault mechanisms, respectively. λ_i and θ_i are coefficients that are determined using Bayesian regression (Gelman et al. 2013). Considering that the earthquakes in the Ridgecrest database have mainly strike-slip fault mechanisms and only small fractions of ground motions recorded from large magnitudes at short distances are available, the coefficients representing magnitude scaling, geometric spreading, anelastic attenuation, and fault styles (i.e., λ_i 's) are constrained using the NGA-West2 ground motion database (Ancheta et al. 2014), while the remaining coefficients (i.e., θ_i 's) are fitted using the Ridgecrest training set.

Figure 4 shows the residuals of the ergodic GMM against M , Z_{hyp} , R_{rup} , and V_{s30} , where there are no noticeable trends, indicating that the ergodic GMM is adequately constrained. The coefficients and standard deviations of the ergodic GMM are presented in Table 1.

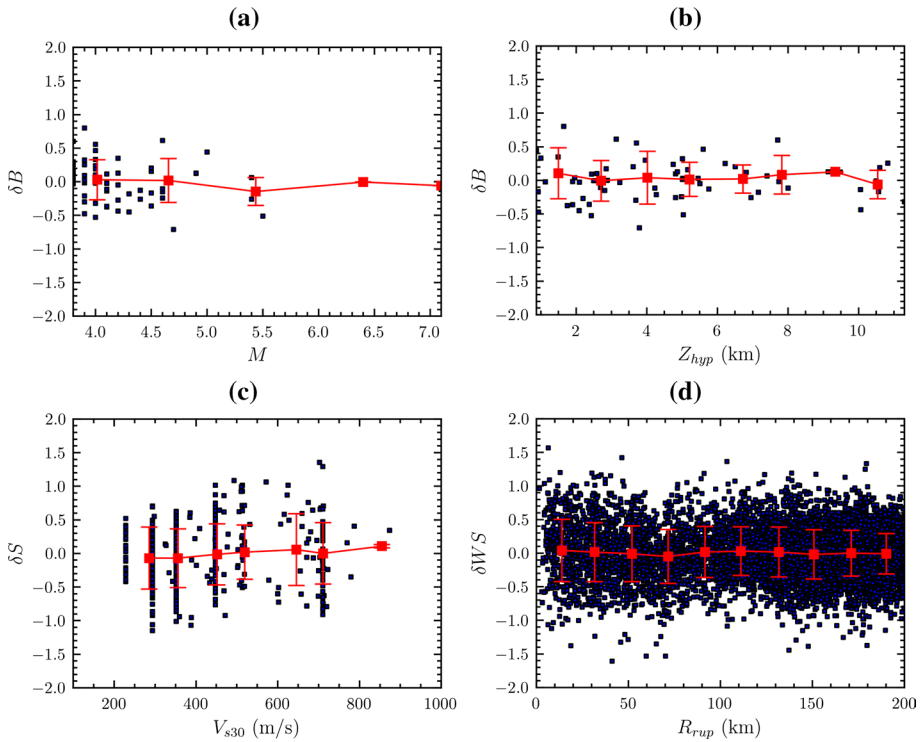


Fig. 4 Residuals of the ergodic GMM for *PGA* against *M*, *Z_{hyp}*, *V_{s30}* and *R_{rup}*

Table 1 Coefficients and standard deviations of the ergodic GMM

θ_0	λ_1	λ_2	λ_3	λ_4	λ_5	θ_6	λ_7
0.178	0.035	1.026	0.064	-1.155	0.196	3.244	-0.007
θ_8	λ_9	λ_{10}	θ_{11}	τ	ϕ_{SS}	ϕ_S	
0.073	0.035	-0.045	-0.538	0.304	0.374	0.465	

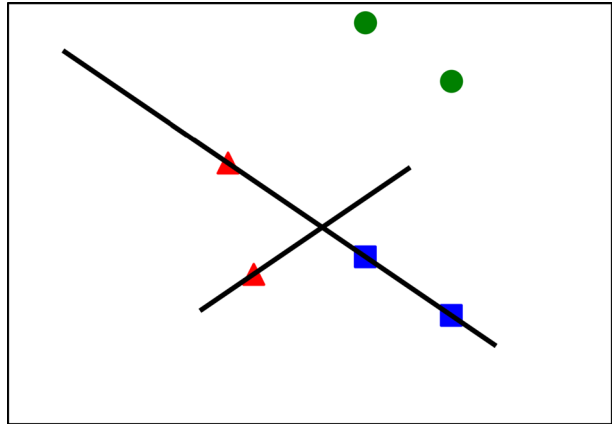
5 Spatial correlation of systematic source effects

Systematic source effects $\delta L2L$ are often modeled using an isotropic covariance function for the GP defined in Eq. 14 (Abrahamson et al. 2019; Liu et al. 2022; Kuehn and Abrahamson 2020), which means that the correlation only depends on the distance between the locations of two earthquakes. A typical isotropic covariance function is defined as:

$$k(x_e, x'_e) = \tau_{L2L}^2 \exp\left(-\frac{\|x_e - x'_e\|}{\ell}\right) \tag{22}$$

where ℓ is the correlation length that controls the decaying rate of correlation with respect to distances between earthquakes. x_e is a 2-dimensional column vector representing the longitude and latitude of an earthquake epicenter (all the coordinates are converted into the

Fig. 5 Three pairs of earthquake epicenters with equal separation distances are marked with circles, triangles, and squares; in addition, two faults are represented by two solid lines. An isotropic stationary covariance function, in this case, would be unable to differentiate the correlation among the three earthquake pairs



Universal Transverse Mercator coordinates in this study). One limitation of Eq. 22 is that it ignores the geometry of a fault and hence could inappropriately extrapolate source effects to distant regions. This is illustrated in Figure 5 and subsequent sections.

In Figure 5, conceivably, we consider 3 pairs of equidistant earthquake epicenters (denoted with circles, squares, and triangles) and 2 fault segments (denoted with black lines), which is representative of the two main faults in the Ridgecrest area. Since all 3 pairs of earthquakes have the same separation distances, Eq. 22 generates the same correlations for the 3 cases. However, the pair of earthquakes located on the same fault (marked as squares) should be more correlated than the pair on different faults (marked as triangles). In addition, the correlation for the pair marked as circles is expected to be the weakest. Considering this observation, we propose an anisotropic non-stationary covariance function by making the correlation length ℓ in Eq. 22 depend on the distance from the earthquake to the fault. Since the covariance function should be positive definite, we use the methodology proposed by Paciorek and Schervish (2006) to construct an anisotropic non-stationary positive definite covariance function as defined in Eqs. 23 and 24.

$$Q(x_e, x'_e) = (x_e - x'_e)^T \left(\frac{\Lambda(x_e) + \Lambda(x'_e)}{2} \right)^{-1} (x_e - x'_e) \tag{23}$$

$$k(x_e, x'_e) = \tau_{L2L}^2 2^{D/2} |\Lambda(x_e)|^{0.25} |\Lambda(x'_e)|^{0.25} |\Lambda(x_e) + \Lambda(x'_e)|^{-0.5} \exp(-\sqrt{Q(x_e, x'_e)}) \tag{24}$$

in which D is the dimension of x_e (i.e., $D = 2$ in this case). Λ is a D by D matrix-valued function. The premultiplication terms in Eq. 24 ensure the positive definiteness of $k(x_e, x'_e)$. $\Lambda(x)$ describes the relationship between correlation length and the earthquake location x and should also be positive definite. As a result, we model the correlation length ℓ as a positive function of x_e and put it on the diagonal entries of Λ :

$$\Lambda(x) = \begin{pmatrix} \ell^2(x) & 0 \\ 0 & \ell^2(x) \end{pmatrix} \tag{25}$$

if $i = j$:

$$\ell^2(x_e) = (a * \exp(-bd(x_e, i)))^2 \tag{26}$$

$$\ell^2(x'_e) = (a * \exp(-bd(x'_e, j)))^2 \tag{27}$$

if $i \neq j$:

$$\ell^2(x_e) = (a * \exp(-bd(x_e, j)))^2 \tag{28}$$

$$\ell^2(x'_e) = (a * \exp(-bd(x'_e, i)))^2 \tag{29}$$

where i and j are the index numbers of the fault segments, to which x_e and x'_e are closest, respectively. $d(x_e, i)$ is the shortest distance from x_e to the fault segment i . a and b are positive coefficients to be estimated. ℓ is an exponentially decaying function of distances from earthquake locations to fault segments. As a result, the correlation between two earthquakes x_e and x'_e decreases as their distances to the fault segments increase.

We train the isotropic stationary (Eq. 22) and anisotropic non-stationary models (Eqs. 23 to 29) using the Ridgecrest training set and evaluate their predictive performance

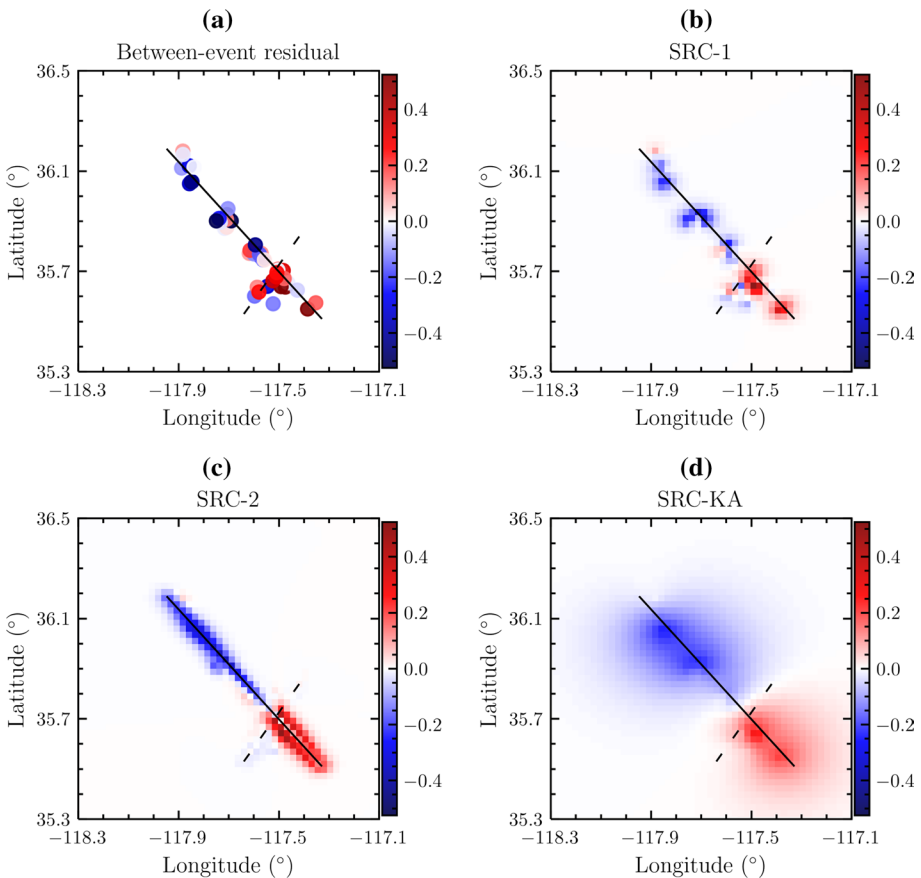


Fig. 6 Comparison of the spatial distribution of **a** between-event residuals in the training set, and predicted source effects, $\delta L2L$, by the **b** SRC-1, **c** SRC-2, and **d** SRC-KA models. The Eastern Little Lake and the Southern Little Lake faults in the Ridgecrest area are represented by the solid and dashed black lines, respectively

on the test set. In addition, we compare our models with the one developed in Kuehn and Abrahamson (2020). The Kuehn and Abrahamson (2020) model shares the same functional form as our isotropic stationary model but has coefficients estimated using the ANZA array data. The coefficients in Eqs. 22 to 29 are determined by maximum a posteriori (MAP) estimation using the program STAN (Carpenter et al. 2017). For simplicity, we consider the two main faults in the Ridgecrest area (i.e., the Eastern Little Lake fault and the Southern Little Lake fault, as described in Plesch et al. (2020)) and parameterize them as two straight line segments, as shown in Fig. 6. It is important to note that the anisotropic non-stationary model also works for faults with nonlinear geometries. For subsequent discussion, we denote the isotropic stationary and anisotropic non-stationary models developed in this study for source effects as SRC-1 and SRC-2, respectively. In addition, we denote the source effects model developed in Kuehn and Abrahamson (2020) as SRC-KA. The estimated coefficients for these models are shown in Table 2.

Figure 6 shows the spatial distribution of the between-event residuals δB in the training set and systematic source effects ($\delta L2L$) predicted for the Ridgecrest area by SRC-1, SRC-2, and SRC-KA models. The lower half of the Eastern Little Lake fault shows on average higher between-event residuals than the upper half of the fault. This spatial pattern is captured by all three models. The earthquake epicenters in the training set exhibit clustered structures, leading to the circular contours around earthquake epicenters in the SRC-1 model. In contrast, the source effects predicted by SRC-2 show no clusters; instead, the contours of the source effects are interpolated and constrained along the two faults. As opposed to SRC-1 and SRC-2, the predicted source effects from SRC-KA extrapolate to distant locations where there are no data to validate the extrapolation. This result is likely influenced by the sparsely distributed seismicity in the ANZA array (based on which SRC-KA is developed) as compared to the clustered epicenters in the Ridgecrest area, which yields a longer correlation length in SRC-KA (i.e., 11.8 km vs. 2 km as shown in Table 2).

It is difficult to compare the distribution of source effects based on simple visualizations; as a result, we quantify the performance of these models by computing the root-mean-square-error (RMSE) and mean negative log likelihood (MNLL) on the test set. The test set consists of earthquakes with different epicenter locations than the training set, which reduces the bias of the performance estimation. The RMSE and MNLL are defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \mu_i)^2}{N}} \tag{30}$$

$$MNLL = \frac{\sum_{i=1}^N \frac{1}{2} \log 2\pi\sigma_i^2 + \frac{(y_i - \mu_i)^2}{2\sigma_i^2}}{N} \tag{31}$$

Table 2 Coefficients of correlation models for path effects

Model	a	b	ℓ	τ_{L2L}	τ_0
SRC-1	–	–	2.013	0.302	0.037
SRC-2	51.086	1.001	–	0.239	0.188
SRC-KA	–	–	11.818	0.202	0.249

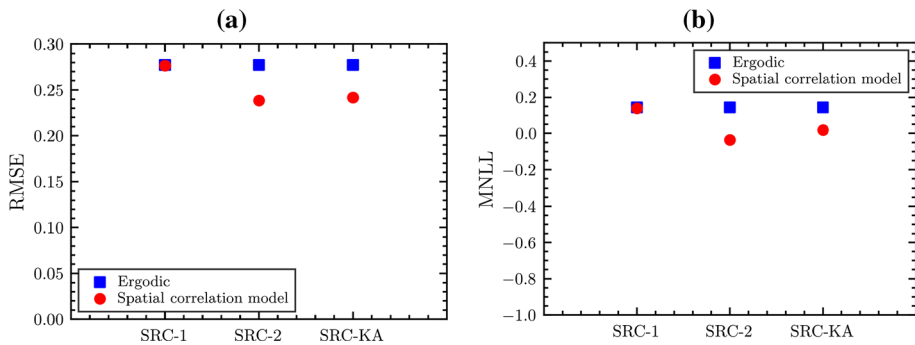


Fig. 7 Performance comparison of correlation models for source effects in terms of (a) RMSE and (b) MNLL on the test set

where N is the total number of data, μ_i is the predicted mean source, path, or site effects, and y_i corresponds to δB , δWS , or δS , respectively (for example, in the case of source terms, N represents the total number of earthquakes, y_i and μ_i are the observed between-event residual (δB) and predicted mean source effect for the i th earthquake). σ_i is the total standard deviation (including both the epistemic and aleatory standard deviations) associated with the i th prediction. RMSE measures the differences between predicted and observed values, whereas the MNLL takes into account the uncertainty associated with the predictions (Bosman and Thierens 2000). Specifically, the MNLL is proportional to Kullback-Leibler divergence (Kullback and Leibler 1951); hence, it measures the difference between distributions predicted by different models and the true distribution of the data. For both metrics, RMSE and MNLL, a lower value indicates a better predictive performance.

The RMSE and MNLL results are presented in Figure 7 for the three source effects models. The anisotropic non-stationary SRC-2 model has better performance (lowest RMSE and MNLL) as compared to the other models. The SRC-1 model has similar RMSE and MNLL to the ergodic model (without any spatial correlation modeling of systematic source effects), indicating that it has a similar prediction error as the ergodic model. This is because the SRC-1 model is limited in extrapolating the source effect to new locations along the faults compared to the SRC-2 model (see Figure 6) and hence over-fits the data (see τ_0 in Table 2). Interestingly, the SRC-KA model has comparable performance to the SRC-2 model. One possible explanation for this may be due to the similarity in earthquake characteristics of the database used in its derivation and the one used in this study. Both the ANZA array and the Ridgecrest databases contain mainly earthquakes in a narrow magnitude range, and the earthquakes in both databases are within the same tectonic region (e.g., both in Southern California). Hence, it is reasonable for the two regions to have similar spatial correlations of source effects. This may also be related to the fact that most of the earthquakes in the Ridgecrest database are located close to the considered faults, and the SRC-2 and SRC-KA models predict similar source effects along the faults (see Figure 6), which may be an artifact of the correlation length in the SRC-KA model. However, the extrapolation of source effects to distant areas in the SRC-KA model is not desired. It is worth highlighting that the performance of the SRC-2 and SRC-KA models should be further evaluated considering earthquakes that are far from the faults, which is not assessed in this study due to the characteristics of the Ridgecrest ground motion database (i.e., most earthquakes occur

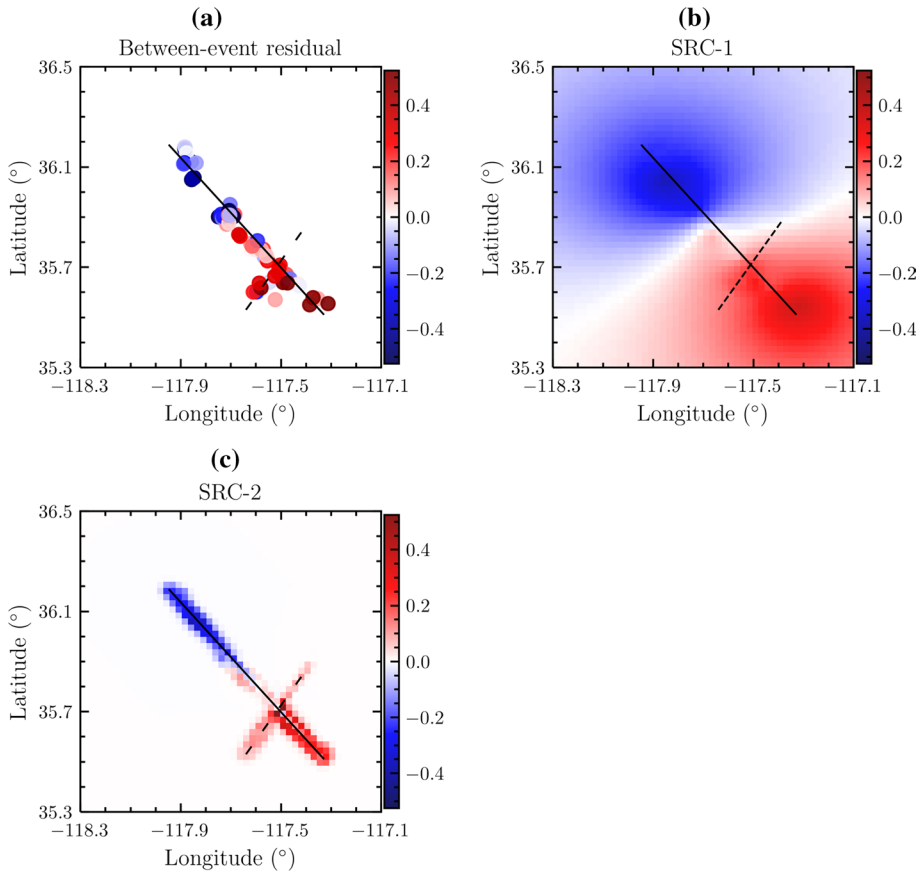
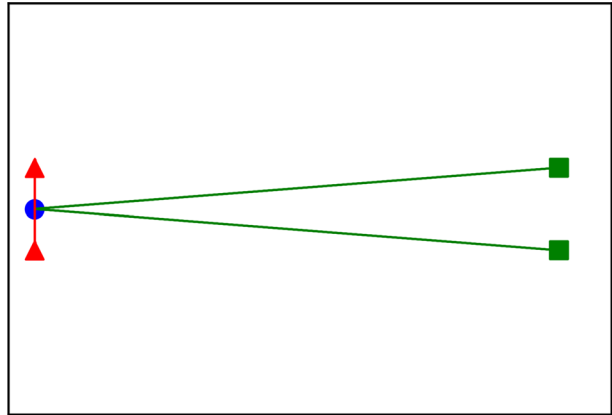


Fig. 8 Comparison of the spatial distribution of (a) between-event residuals, and predicted source effects, $\delta L2L$, by (b) the SRC-1 and (c) SRC-2 models for a different ground motion training subset. The Eastern Little Lake and the Southern Little Lake faults in the Ridgecrest area are represented by the solid and dashed black lines, respectively

near the faults). Moreover, the SRC-2 model takes into account the fault geometry; it is not only controlled by the spatial patterns of the between-event residuals, which is the case for the isotropic stationary model. This advantage is illustrated in Figure 8, where we select a different training set to estimate the coefficients of the SRC-1 and SRC-2 models. In this dataset, the between-event residuals show strong negative and positive biases in the upper and lower portions of the Eastern Little Lake fault, respectively. In contrast with Figure 6, the SRC-1 model in Figure 8 shows a stronger extrapolation of source effects to distant locations due to its large correlation length (30 km—influenced by the different training set), while this extrapolation is constrained by the fault geometries for the SRC-2 model, regardless of the training set. In certain scenarios where fault characteristics are not known or explicitly modeled, the traditional isotropic stationary correlation model for source effects could be used with engineering judgment to assess the model extrapolation carefully.

Fig. 9 Illustration of the spatial correlation of path effects. The earthquake epicenter is marked as a circle and sites are denoted as triangles and squares. Path effects for sites in squares are likely more similar than the sites marked as triangles



6 Spatial correlation of systematic path effects

Previous research efforts have found it difficult to model the spatial correlation of path effects using an isotropic stationary covariance model that is a function of the distances between two sites (i.e., $|x_s - x'_s|$) (Loth and Baker 2013; Foulser-Piggott and Stafford 2012; Kuehn and Abrahamson 2020). This difficulty results from the fact that the correlation between two paths depends not only on the distances between sites but also on the source-to-site distances. For example, the two pairs of sites shown in Figure 9 have the same between-site distances but the path effects for the pair far from the earthquake epicenter (represented by squares) should be more correlated because the propagation paths of earthquake waves to the further sites are likely more similar than those to the closer sites.

To capture this phenomenon, we first apply the anisotropic non-stationary covariance function described in Kuehn and Abrahamson (2020), which is similar to the SRC-2 model, to estimate the correlation of path effects, as shown in Eqs. 32 to 38.

If paths $[x_s, x_e]$ and $[x'_s, x'_e]$ come from the same earthquake or travel to the same site:

$$k([x_s, x_e], [x'_s, x'_e]) = \tau_{L2L}^2 2^{D/2} |\Lambda(x_s, x_e)|^{0.25} |\Lambda(x'_s, x'_e)|^{0.25} * |\Lambda(x_s, x_e) + \Lambda(x'_s, x'_e)|^{-0.5} \exp(-\sqrt{Q([x_s, x_e], [x'_s, x'_e])}) \tag{32}$$

Otherwise:

$$k([x_s, x_e], [x'_s, x'_e]) = 0 \tag{33}$$

Specifically, $Q([x_s, x_e], [x'_s, x'_e])$ is modeled as a function of between-site or between-earthquake distances for paths from the same earthquakes or to the same sites, respectively:

$$Q([x_s, x_e], [x'_s, x'_e]) = \begin{cases} (x_s - x'_s)^T \left(\frac{\Lambda(x_s, x_e) + \Lambda(x'_s, x'_e)}{2} \right)^{-1} (x_s - x'_s) & \text{same earthquake} \\ (x_e - x'_e)^T \left(\frac{\Lambda(x_s, x_e) + \Lambda(x'_s, x'_e)}{2} \right)^{-1} (x_e - x'_e) & \text{same site} \end{cases} \tag{34}$$

$$\Lambda(x_s, x_e) = \begin{pmatrix} \ell^2(x_s, x_e) & 0 \\ 0 & \ell^2(x_s, x_e) \end{pmatrix} \tag{35}$$

In the above equations, $D = 2$ since the covariance function depends on either the earthquake or site locations. We have conducted an initial investigation of the spatial correlation of δWS using semivariograms and observed a generally positive relationship between the correlation length of path effects and the earthquake magnitude. This observation is consistent with that in Heresi and Miranda (2019) that spatial correlation parameters for within-event residuals are magnitude dependent. Consistently, Kuehn and Abrahamson (2020) also found that the performance of correlation models for path effects was improved by making the correlation length magnitude-dependent. For simplicity, in this study, we model the correlation length $\ell(x_s, x_e)$ to be dependent on the earthquake magnitude in addition to the source-to-site distance (we use R_{rup} in this study). Regarding the functional forms of $\ell(x)$, we first consider a linear dependence on R_{rup} and M :

$$\ell^2(x_s, x_e) = \ell^2(M, R_{rup}) = (p + qMR_{rup})^2 \tag{36}$$

where p and q are positive coefficients to be estimated, M is the magnitude of the earthquake with epicenter x_e , and R_{rup} is the rupture distance from the site x_s to the earthquake x_e . This functional form is selected to be consistent with Kuehn and Abrahamson (2020) to facilitate comparisons, as discussed later. However, Eq. 36 implies that ℓ is unbounded, and can unrealistically increase to infinity when R_{rup} approaches infinity. To address this issue, we evaluate additional functional forms with a slower increasing rate for ℓ and an upper bound, as shown in Eqs. 37 and 38.

$$\ell^2(x_s, x_e) = \ell^2(M, R_{rup}) = (p + q \ln(MR_{rup} + 1))^2 \tag{37}$$

$$\ell^2(x_s, x_e) = \ell^2(M, R_{rup}) = \left(p + q \frac{1 - \exp(-MR_{rup})}{1 + \exp(-MR_{rup})} \right)^2 \tag{38}$$

In Eq. 37, a natural logarithm is introduced to MR_{rup} to reduce the scaling of ℓ when R_{rup} gets extremely large. Equation 38 is modified based on the sigmoid function (Han and Moraga 1995), where ℓ has a lower bound of p and an upper bound of $p + q$ when MR_{rup} approaches 0 and infinity, respectively. By making ℓ a constant to be estimated from the data, we also develop an isotropic stationary model to benchmark the performances of anisotropic non-stationary models. We denote the isotropic stationary model as PATH-1 and

Table 3 Characteristics of different models for path effects

Model	ℓ	Stationarity	Paths considered as correlated
PATH-1	Constant	Isotropic stationary	From the same station or earthquake
PATH-1a	$p + qMR_{rup}$	Anisotropic non-stationary	From the same station or earthquake
PATH-1b	$p + q \ln(MR_{rup} + 1)$	Anisotropic non-stationary	From the same station or earthquake
PATH-1c	$p + q \frac{1 - \exp(-MR_{rup})}{1 + \exp(-MR_{rup})}$	Anisotropic non-stationary	From the same station or earthquake
PATH-2	constant	Isotropic stationary	All paths
PATH-2a	$p + qMR_{rup}$	Anisotropic non-stationary	All paths
PATH-2b	$p + q \ln(MR_{rup} + 1)$	Anisotropic non-stationary	All paths
PATH-2c	$p + q \frac{1 - \exp(-MR_{rup})}{1 + \exp(-MR_{rup})}$	Anisotropic non-stationary	All paths
PATH-KA	$p + qMR_{rup}$	Anisotropic non-stationary	From the same station or earthquake

the anisotropic non-stationary models with the length functions in Eqs. 36, 37, and 38 as PATH-1a, PATH-1b and PATH-1c, respectively. Similar to SRC-KA, the path-effect model in Kuehn and Abrahamson (2020) is denoted as PATH-KA, which has the same functional form as PATH-1a with different coefficients estimated from the ANZA array. These model definitions are summarized in Table 3. Lastly, it is important to highlight that, also as outlined in Schiappapietra and Smerzini (2021) and Stafford et al. (2019), other factors such as rupture processes, fault extent, etc., could also contribute to the estimation of correlation lengths, especially for near-field ground motions. Hence, more advanced correlation models that account for the different characteristics of far-field and near-field path effects could also be explored with different databases in the future.

It is important to note that the above models ignore the correlation between paths that travel from two different earthquakes to two different sites. This might be a proper assumption for a database with sparsely distributed seismicity and sites. However, for the Ridgecrest database used in this study where earthquakes and sites are closely located, it is reasonable to consider the correlation of paths from different earthquakes to different sites, which results in the following models:

$$k([x_s, x_e], [x'_s, x'_e]) = \tau_{L2L}^2 2^{D/2} |\Lambda(x_s, x_e)|^{0.25} |\Lambda(x'_s, x'_e)|^{0.25} * |\Lambda(x_s, x_e) + \Lambda(x'_s, x'_e)|^{-0.5} \exp(-\sqrt{Q([x_s, x_e], [x'_s, x'_e])}) \tag{39}$$

$$Q([x_s, x_e], [x'_s, x'_e]) = (\hat{x} - \hat{x}')^T \left(\frac{\Lambda(x_s, x_e) + \Lambda(x'_s, x'_e)}{2} \right)^{-1} (\hat{x} - \hat{x}') \tag{40}$$

$$\Lambda(x_s, x_e) = \begin{pmatrix} \ell^2(x_s, x_e) & 0 & 0 & 0 \\ 0 & \ell^2(x_s, x_e) & 0 & 0 \\ 0 & 0 & \ell^2(x_s, x_e) & 0 \\ 0 & 0 & 0 & \ell^2(x_s, x_e) \end{pmatrix} \tag{41}$$

where $\hat{x} = [x_e^T, x_s^T]^T$ is a 4 by 1 column vector, then D in Eq. 39 becomes 4. The functional forms of ℓ are the same as Eqs. 36, 37, and 38. These models are similar to the previous ones (i.e., PATH-1a, 1b, and 1c) except that now the correlation for any two paths depends

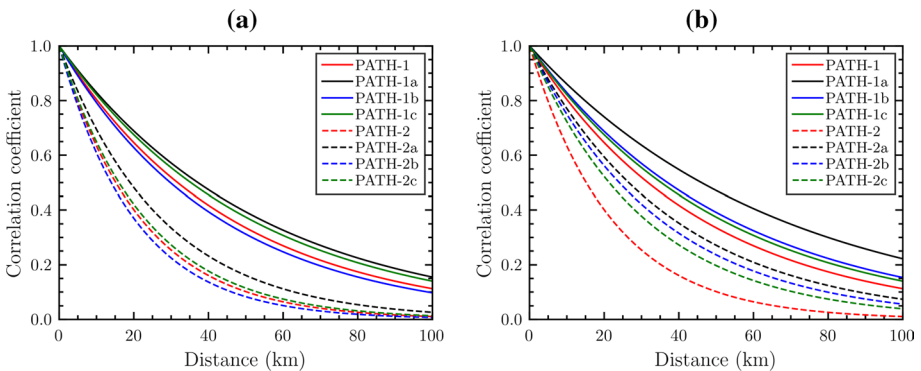


Fig. 10 Correlation of path effects as a function of separation distances between stations and/or earthquakes for different models considering (a) $R_{rup} = 90$ km, $M = 4.3$ and (b) $R_{rup} = 150$ km, $M = 7.5$

on both the distances between two earthquakes and between two sites. Accordingly, we denote these models with length functions in Eq. 36, 37, and 38 as PATH-2a, PATH-2b, and PATH-2c, respectively, and the corresponding isotropic stationary model is denoted as PATH-2. The details of the models are presented in Table 3.

Figure 10a shows the path-effect correlation as a function of separation distance given $R_{rup} = 90$ km and $M = 4.3$ for all correlation models developed in this study. The anisotropic non-stationary models have similar trends compared to the isotropic stationary models. This may be because the considered R_{rup} and M values are close to the average values in the Ridgecrest database. The PATH-2a,b,c models show a lower correlation compared to the PATH-1a,b,c models; this is likely because they consider the correlation of paths from different earthquakes to different sites. Figure 10b shows the correlation considering a different magnitude and rupture distance (i.e., $R_{rup} = 150$ km and $M = 7.5$). While the trends of the isotropic stationary models remain unchanged, the anisotropic non-stationary models exhibit increased correlations due to their magnitude- and distance-dependent length functions. The models with a linearly-dependent length function (e.g., PATH-1a and PATH-2a) show a higher correlation compared to models with other length functions.

We also examine the impact of the cell-specific attenuation approach (Dawood and Rodriguez-Marek 2013) on the spatial correlation of path effects. In this case, we divide the Ridgecrest area into multiple 20-by-20-km cells and compute attenuation coefficients for each cell. Then the anelastic attenuation term $\lambda_7 R_{rup}$ in the ergodic GMM (Eq. 20) is replaced by the cell-specific attenuation term:

$$\sum_{i=1}^{N_{cell}} c_i R_i \tag{42}$$

where N_{cell} is the total number of cells in the Ridgecrest area, c_i is the attenuation coefficient for the i th cell, and R_i is the fraction of a path within the i th cell. The coefficients of the correlation models are re-estimated using the event-site-corrected residuals (δWS) adjusted by the cell-specific attenuation.

Figure 11 shows the RMSE and MNLL of the path effects prediction on the test set using the correlation models without cell-specific attenuation. The isotropic stationary

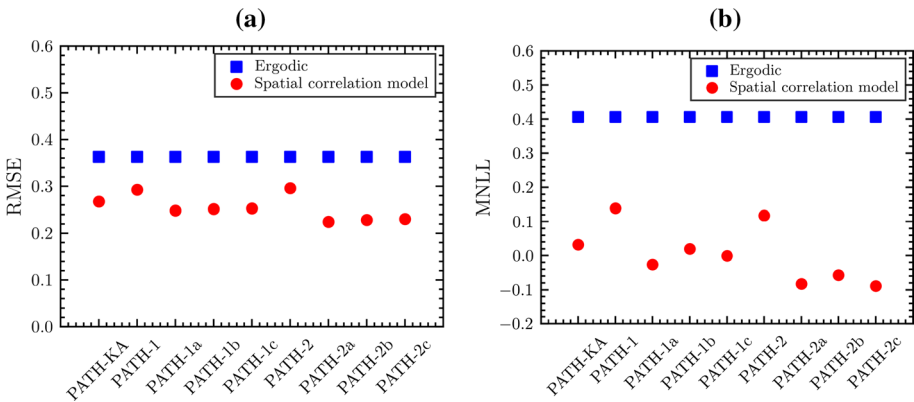


Fig. 11 Comparing the predictive performance of different correlation models for path effects in terms of **a** RMSE and **b** MNLL

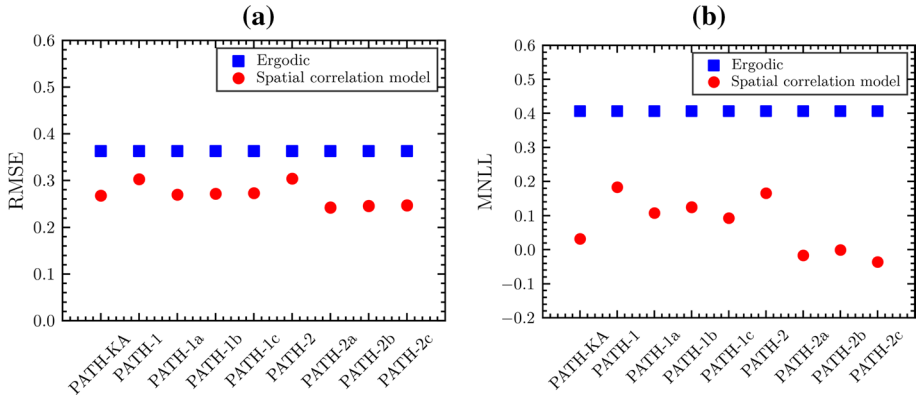


Fig. 12 Comparing the predictive performance of different correlation models with cell-specific attenuation for path effects in terms of **a** RMSE and **b** MNLL

models have a higher RMSE and MNLL than the anisotropic non-stationary models, which show a similar reduction in RMSE and MNLL with respect to the ergodic one. The PATH-2a, 2b, and 2c models have slightly lower RMSE and MNLL than the other models. This may be associated with the Ridgecrest earthquakes being located in a relatively narrow region; hence, there are potential correlations between paths that come from two nearby earthquake sources to two different stations, which cannot be captured by the PATH-1-type models. The influences of the three correlation length functions on the RMSE are

Table 4 Coefficients of correlation models for path effects

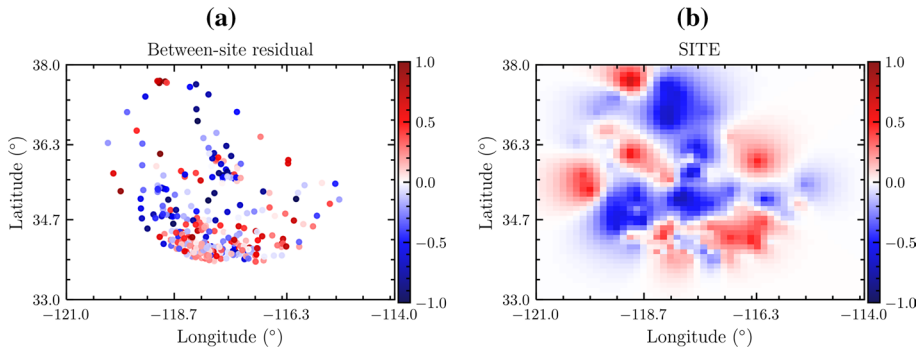
Model	p	q	ℓ	ϕ_{P2P}	$\phi_{0,SS}$
PATH-1	–	–	45.3	0.246	0.288
PATH-1a	2.267	0.092	–	0.252	0.278
PATH-1b	4e-6	7.289	–	0.248	0.280
PATH-1c	45.64	0.112	–	0.244	0.284
PATH-2	–	–	40.3	0.233	0.291
PATH-2a	1.24e-6	0.027	–	0.277	0.269
PATH-2b	6.84e-5	2.907	–	0.249	0.278
PATH-2c	17.667	0.012	–	0.246	0.281

Table 5 Coefficients of correlation models for path effects with cell-specific attenuation adjustments

Model	p	q	ℓ	ϕ_{P2P}	$\phi_{0,SS}$
PATH-1	–	–	44.3	0.245	0.288
PATH-1a	2.296	0.089	–	0.231	0.263
PATH-1b	3e-6	7.074	–	0.227	0.268
PATH-1c	44.09	0.125	–	0.232	0.281
PATH-2	–	–	39.1	0.234	0.293
PATH-2a	9e-7	0.023	–	0.265	0.277
PATH-2b	7.14e-5	2.531	–	0.235	0.281
PATH-2c	16.891	0.012	–	0.221	0.293

Table 6 Coefficient of the correlation model for site effects

Model	ℓ	ϕ_{S2S}	$\phi_{0,S2S}$
SITE	34.798	0.384	0.262

**Fig. 13** Spatial distribution of **a** between-site residuals are compared with **b** predicted site effects by the correlation model

unnoticeable, but the linear functional form (e.g., $\ell^2(M, R_{rup}) = (p + qMR_{rup})^2$, corresponding to PATH-1a and PATH-2a models) shows slightly better MNLL than others. It is also interesting to observe that the PATH-KA model has a comparable performance (in terms of both RMSE and MNLL) to the models developed in this study, which possibly results from the similarity in the scales of crustal heterogeneity between Ridgecrest and ANZA array (Kuehn and Abrahamson 2020).

Similar comparisons for correlation models with cell-specific attenuation are shown in Figure 12. The prediction performance of these models is similar to those shown in Figure 11, indicating that the cell-specific attenuation approach has a minimal impact on the spatial correlation of path effects, which is also reflected by the estimated model coefficients in Tables 4 and 5, where only small changes in coefficients can be observed for models with and without cell-specific attenuation. This observation could be attributed to the spatial distribution of earthquakes and sites in the Ridgecrest database. The clustered earthquake epicenters and stations lead to high correlation and similarity in the propagation paths of seismic waves (e.g., Kotha et al. (2017)). As a result, it is likely that multiple paths with similar lengths pass through the same cells, concealing the potential variations of path effects among these paths. More sophisticated cell grid designs (e.g., grids with heterogeneous cell shapes or sizes and different orientations) could be considered beneficial in future studies to improve the cell-specific attenuation approach.

7 Spatial correlation of systematic site effects

The risk assessment of spatially distributed infrastructures (such as portfolios of buildings, pipeline networks, etc.) requires the estimation of the spatial correlation of systematic site effects for regional PSHA (Giorgio and Iervolino 2016; Chioccarelli et al. 2019). Following the common methodologies in previous studies (Foulser-Piggott and Goda 2015;

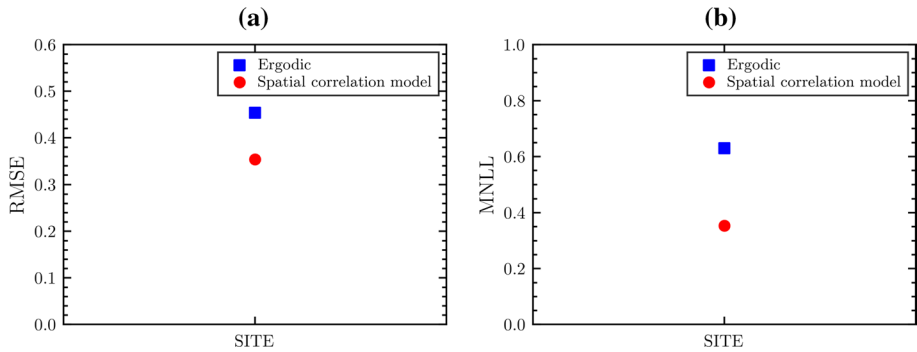


Fig. 14 Comparing the predictive performance of the correlation model for site effects with the ergodic model in terms of **a** RMSE and **b** MNLL

Abrahamson et al. 2019; Jayaram and Baker 2009; Chao et al. 2021; Rahpeyma et al. 2018), we model the spatial correlation of site effects using an isotropic stationary covariance function of between-site distances:

$$k(x_s, x'_s) = \phi_{S2S}^2 \exp\left(-\frac{\|x_s - x'_s\|}{\ell}\right) \tag{43}$$

The estimated coefficients in Eq. 43 are shown in Table 6. The site effects predicted using Eq. 43 are compared with the between-site residuals (δS) in Ridgecrest as shown in Figure 13. The site effects show consistent spatial variation with δS and decrease to zero when extrapolating to distant regions with no data. This spatial variation of site effects represents potential spatial deviations from the average correlation between V_{s30} and deep shear wave velocity profiles as well as potential differences from average topographies effects described by the ergodic GMM. Figure 14 shows the predictive performance of the correlation model on the test set, where a significant reduction in RMSE and MNLL can be observed as compared to the ergodic model. We denote this correlation model as SITE. One could further investigate whether the spatial correlation of site effects could be better modeled with a non-stationary correlation function that takes into account for example differences between stations inside and outside basins (Chen et al. 2021).

Table 7 Different combinations of correlation models for systematic effects to generate *PGA* prediction on the test set

Model	Combination
A	SITE+SRC-1+PATH-1a
B	SITE+SRC-1+PATH-2a
C	SITE+SRC-2+PATH-1a
D	SITE+SRC-2+PATH-2a
KA	SITE+SRC-KA+PATH-KA

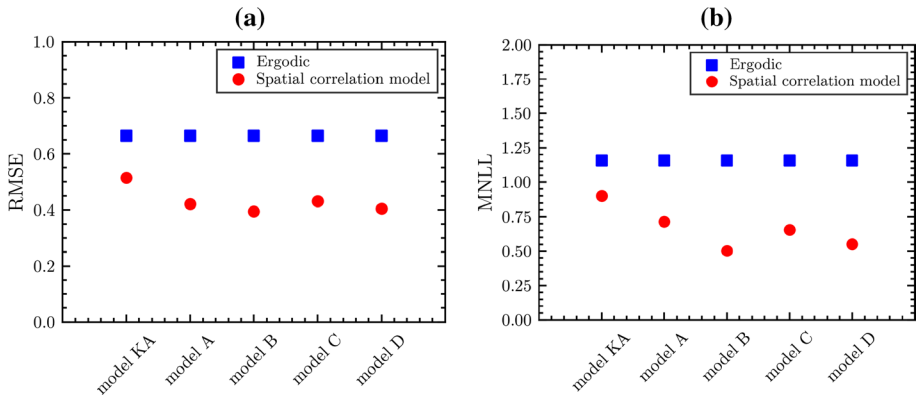


Fig. 15 Comparing the predictive performance of different combinations of correlation models for systematic effects in terms of **a** RMSE and **b** MNLL

8 Evaluation of overall performance

In this section, we compare the performance of different combinations of source-, path-, and site-effect models on the *PGA* prediction for the test set. The comparisons also include the ergodic GMM to provide additional insights. For simplicity, we consider only the PATH-1a and PATH-2a models without cell-specific attenuation for path effects since all path-effect models have similar performance. Hence, five models are compared, as summarized in Table 7.

Figure 15 shows the RMSE and MNLL for the five models. Significant reductions in RMSE and MNLL are observed for all models. The similar performance between Models A and C, as well as between Models B and D, is associated with the similar performance of the SRC-1 and SRC-2 models, which in turn only have a minimal improvement compared to the ergodic source-effect model (see Figure 7). By considering the correlation between paths from different earthquakes to different sites, Models B and D show slightly lower RMSE and MNLL than Models A and C. This is associated with the slightly better performance of the PATH-2a model over the PATH-1a model. Interestingly, despite being developed using a different GMM database, Model KA exhibits significant improvement with respect to the ergodic GMM. The models developed in this study show a slightly better performance than Model KA. Considering the prediction performance and computational efficiency, we recommend the use of Model C. Model B could be preferred for its highest prediction accuracy and precision if the computational cost is not a major concern.

9 Discussion

Compared to the correlation models for systematic path and site effects, the models for systematic source effects do not show a significant improvement in predictive performance with respect to their ergodic counterpart. This observation could be attributed to the database used in this study. The Ridgecrest database mainly consists of aftershocks with similar magnitudes ranging from 3 to 5 and with epicenters densely located in a small region.

Hence, it is challenging to model the spatial correlation of source effects at large scales and assess the potential magnitude dependence of correlation structures. Similarly, the cell-specific attenuation approach shows a minimal impact on the prediction of path effects and the coefficients of correlation models for the reasons discussed in the previous section. Additional correlation models for source effects and alternatives to the current formulation of the cell-specific attenuation approach should be further investigated when more comprehensive ground motion databases are available.

The PATH-2a to 2c models have slightly better predictive performance than the PATH-1a to 1c models. However, it is worth pointing out that the PATH-1a, 1b, and 1c models produce sparse covariance matrices (Eq. 33) as compared to the PATH-2a, 2b, and 2c models that generate large dense matrices; hence, the PATH-1a, 1b, and 1c models can have a higher computational efficiency by using specialized matrix computation algorithms (Lawrence et al. 2003; Melkumyan and Ramos 2009). For example, in this study, the coefficients in the PATH-1a, 1b, and 1c models are estimated using STAN on a local CPU, while the coefficients in the PATH-2a, 2b, and 2c models have to be solved using TensorFlow Probability (Dillon et al. 2017) with TPU (Tensor Processing Unit, Jouppi et al. (2017)) acceleration on the cloud (Bisong 2019). Specifically, the PATH-2a, 2b, and 2c models would take 30 times more time than the PATH-1a, 1b, and 1c models if they are run in a local CPU. More sophisticated correlation models for path effects should be further explored in future studies to account for both computational efficiency and predictive performance.

It is also interesting to highlight that the correlation models for systematic path effects developed in Kuehn and Abrahamson (2020) for the ANZA array show a performance that is comparable to the models developed in this study. Incidentally, Kuehn and Abrahamson (2020) also found that the correlation structures of path effects for the ANZA array were similar to those for a ground motion database in Taiwan (NCREE 2015). These observations potentially suggest that the spatial correlation structures for path effects in different regions might be potentially similar. This makes promising the transference of the spatial correlation structures from one region to another by assuming similar scales of crustal heterogeneity across regions, which is consistent with the observations in Kuehn and Abrahamson (2020). Lastly, the correlation models investigated in this study are restricted to *PGA*; correlation models for spectral accelerations at other periods or other IMs should be evaluated in future studies.

10 Conclusions

In this study, we have developed spatial correlation models for systematic source, path, and site effects for the Ridgecrest area and compared their performance against existing correlation models. Specifically, we first develop an ergodic GMM and then partition its residuals to estimate the spatial correlation structures of systematic effects. We propose an anisotropic non-stationary correlation model that considers fault geometries and an isotropic stationary model for source effects, several isotropic stationary or anisotropic non-stationary models for path effects, and an isotropic stationary model for site effects.

We find that the anisotropic non-stationary correlation model for source effects captures the fault geometries and extrapolates more reasonably to regions with no data than isotropic stationary correlation models. In terms of path effects, the proposed anisotropic non-stationary correlation models show similar overall performance, which is better than

that of the isotropic stationary models. In addition, we find that an isotropic stationary model performs well in capturing the spatial distribution of site effects, which is consistent with previous studies. We also observe that the performance gained from correlation models for source effects is lower compared to the gain in performance by including path and site effects, which may be associated with the spatial distribution of seismicity in the Ridgecrest area. Among the models considered in this study, we recommend the use of Model C (SITE+SRC-2+PATH-1a) as it provides a good balance in terms of prediction performance and computational efficiency. Model B (SITE+SRC-1+PATH-2a) has the highest prediction accuracy and could be preferred if the computational cost is not a major concern. The different correlation models developed in this study can also be incorporated in a logic tree for non-ergodic PSHA if additional epistemic uncertainties associated with the spatial correlation structures of systematic effects need to be considered.

The correlation models developed in this study are also compared with those developed by Kuehn and Abrahamson (2020) using different databases. We find that models for path effects in Kuehn and Abrahamson (2020) have comparable performance to those developed in this study, which may suggest that the spatial correlation structures of path effects are similar for different regions, making them potentially transferable. Interestingly, these models are developed based on very different databases (e.g., number of recordings, locations, spatial density, etc.); hence, this observation should be future investigated in future studies, emphasizing potential physical constraints. Lastly, this study investigated correlation structures for *PGA*; future studies should consider spectral accelerations at other periods and other IMs.

11 Data and resources

The United States Geological Survey Ridgecrest (Rekoske et al. 2020) and the Pacific Earthquake Engineering Research (PEER) NGA-West2 (Ancheta et al. 2014) ground motion databases are used in this study. The programs STAN (Carpenter et al. 2017) and TensorFlow Probability (Dillon et al. 2017) are used to infer the parameters of correlation models.

Acknowledgements We acknowledge the financial support provided by the Civil and Environmental Engineering department of the Georgia Institute of Technology and the Pacific Gas and Electric company. We also thank the U.S. non-ergodic hazard working group for the valuable discussion on spatial correlation models and for the comments provided on our study. In particular, we thank Dr. Yousef Bozorgnia, Dr. Grigoris Lavrentiadis, Dr. Christine Goulet, and Dr. Albert Kottke for providing valuable insights. Finally, we thank the anonymous reviewers and the editor who provided constructive comments that helped to improve the quality of our manuscript.

Funding This work was partially supported by the Georgia Institute of Technology (Award DE00000701), and Pacific Gas and Electric.

Availability of data and material The Ridgecrest ground motion database are available in the USGS website (https://www.strongmotioncenter.org/specialstudies/rekoske_2019ridgecrest/).

Code availability The code for correlation models can be acquired by email request to the first author.

Declarations

Conflicts of interest We do not have any conflict of interest.

References

- Abrahamson NA, Youngs R (1992) A stable algorithm for regression analyses using the random effects model. *Bull Seismol Soc Am* 82(1):505–510
- Abrahamson NA, Kuehn NM, Walling M, Landwehr N (2019) Probabilistic seismic hazard analysis in california using nonergodic ground-motion models. *Probabilistic seismic hazard analysis in california using nonergodic ground-motion models*. *Bull Seismol Soc Am* 109(4):1235–1249
- Abrahamson N, Hollenback J (2012) Application of single-station sigma ground motion prediction equations in practice. In: *Proceedings of the 15th world conference on earthquake engineering, Lisbon, Portugal, paper*, 2536
- Abrahamson N, Kuehn N (2021) Non-ergodic ground-motion models for crustal earthquakes in georgia. In: *Building Knowledge for Geohazard Assessment and Management in the Caucasus and other Orogenic Regions*. Springer, pp. 169–184
- Ancheta TD, Darragh RB, Stewart JP, Seyhan E, Silva WJ, Chiou BSJ, Wooddell KE, Graves RW, Kottke AR, Boore DM et al (2014) Nga-west2 database. *Earthq Spectra* 30(3):989–1005
- Anderson JG, Brune JN (1999) Probabilistic seismic hazard analysis without the ergodic assumption. *Seismol Res Lett* 70(1):19–28
- Atik LA, Abrahamson N, Bommer JJ, Scherbaum F, Cotton F, Kuehn N (2010) The variability of ground-motion prediction models and its components. *Seismol Res Lett* 81(5):794–801
- Atkinson GM (2006) Single-station sigma. *Bull Seismol Soc Am* 96(2):446–455
- Berger J, Baker L, Brune J, Fletcher J, Hanks T, Vernon F III (1984) The anza array: a high-dynamic-range, broadband, digitally radiotelemetered seismic array. *Bull Seismol Soc Am* 74(4):1469–1481
- Bisong E (2019) Google colab. *Building machine learning and deep learning models on google cloud platform*. Springer, USA, pp 59–64
- Bosman PA, Thierens D (2000) Negative log-likelihood and statistical hypothesis testing as the basis of model selection in ideas. In: *Proceedings of the Tenth Dutch–Netherlands Conference on Machine Learning, Tilburg University*
- Caramenti L, Menafoglio A, Sgobba S, Lanzano G (2022) Multi-source geographically weighted regression for regionalized ground-motion models. *Sp Stat* 47:100610
- Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A (2017) Stan: A probabilistic programming language. *Journal of statistical software* 76(1)
- Ceferino L, Liu C, Alisjahbana I, Patel S, Sun T, Kiremidjian A, Rajagopal R (2020) Earthquake resilience of distributed energy resources. In: *17th World Conference on Earthquake Engineering, Tokyo, Japan*
- Chao SH, Lin CM, Kuo CH, Huang JY, Wen KL, Chen YH (2021) Implementing horizontal-to-vertical fourier spectral ratios and spatial correlation in a ground-motion prediction equation to predict site effects. *Earthq Spectra* 37(2):827–856
- Chen YH, Tsai CCP (2002) A new method for estimation of the attenuation relationship with variance components. *Bull Seismol Soc Am* 92(5):1984–1991
- Chen Y, Bradley BA, Baker JW (2021) Nonstationary spatial correlation in new Zealand strong ground-motion data. *Earthquake Eng Struct Dyn* 50(13):3421–3440
- Chioccarelli E, Cito P, Iervolino I, Giorgio M (2019) Reassess v2.0: software for single-and multi-site probabilistic seismic hazard analysis. *Bull Earthquake Eng* 17(4):1769–1793
- Dawood HM, Rodriguez-Marek A (2013) A method for including path effects in ground-motion prediction equations: An example using the mw 9.0 tohoku earthquake aftershocks. *Method for including path effects in gmpe using mw 9.0 tohoku earthquake aftershocks*. *Bulletin of the Seismological Society of America* 103(2B):1360–1372
- Dillon JV, Langmore I, Tran D, Brevdo E, Vasudevan S, Moore D, Patton B, Alemi A, Hoffman M, Saurous RA (2017) Tensorflow distributions. *arXiv preprint arXiv: 1711.10604*
- Foulser-Piggott R, Goda K (2015) Ground-motion prediction models for arias intensity and cumulative absolute velocity for Japanese earthquakes considering single-station sigma and within-event spatial correlation. *Bull Seismol Soc Am* 105(4):1903–1918
- Foulser-Piggott R, Stafford PJ (2012) A predictive model for arias intensity at multiple sites and consideration of spatial correlations. *Earthquake Eng Struct Dyn* 41(3):431–451
- Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) *Bayesian data analysis*. CRC Press, Boca Raton
- GeoPentech (2015) Southwestern united states ground motion characterization sshac level 3. *Technical Report, rev. 2*
- Giorgio M, Iervolino I (2016) On multisite probabilistic seismic hazard analysis. *Bull Seismol Soc Am* 106(3):1223–1234

- Han J, Moraga C (1995) The influence of the sigmoid function parameters on the speed of backpropagation learning. In: *International Workshop on Artificial Neural Networks*. Springer, pp. 195–201
- Hastie T, Tibshirani R, Friedman JH, Friedman JH (2009) The elements of statistical learning: data mining, inference, and prediction, vol 2. Springer
- Heresi P, Miranda E (2019) Uncertainty in intraevent spatial correlation of elastic pseudo-acceleration spectral ordinates. *Bull Earthquake Eng* 17(3):1099–1115
- Hong H, Zhang Y, Goda K (2009) Effect of spatial correlation on estimated ground-motion prediction equations. *Bull Seismol Soc Am* 99(2A):928–934
- Jayaram N, Baker JW (2009) Correlation model for spatially distributed ground-motion intensities. *Earthquake Eng Struct Dyn* 38(15):1687–1708
- Jouppi NP, Young C, Patil N, Patterson D, Agrawal G, Bajwa R, Bates S, Bhatia S, Boden N, Borchers A et al. (2017) In-datacenter performance analysis of a tensor processing unit. In: *Proceedings of the 44th annual international symposium on computer architecture*. pp. 1–12
- Kotha SR, Bindi D, Cotton F (2017) Site-corrected magnitude- and region-dependent correlations of horizontal peak spectral amplitudes. *Earthquake spectra* 33(4):1415–1432
- Kuehn NM, Abrahamson NA (2020) Spatial correlations of ground motion for non-ergodic seismic hazard analysis. *Earthquake Eng Struct Dyn* 49(1):4–23
- Kuehn NM, Abrahamson NA, Walling MA (2019) Incorporating nonergodic path effects into the ngawest2 ground-motion prediction equations. *Bull Seismol Soc Am* 109(2):575–585
- Kuehn N, Bozorgnia Y, Campbell K, Gregor N (2020) Partially non-ergodic ground-motion model for subduction regions using nga-subduction database. *PEER Rept. No 2020/04*
- Kullback S, Leibler RA (1951) On information and sufficiency. *annal math stat* 22(1):79–86
- Landwehr N, Kuehn NM, Scheffer T, Abrahamson N (2016) A nonergodic ground-motion model for California with spatially varying coefficients. *Bull Seismol Soc Am* 106(6):2574–2583
- Lanzano G, Sgobba S, Caramenti L, Menafoglio A (2021) Ground-motion model for crustal events in Italy by applying the multisource geographically weighted regression (ms-gwr) method. *Bull Seismol Soc Am* 111(6):3297–3313
- Lavrentiadis G, Abrahamson NA, Nicolas KM, Bozorgnia Y, Goulet CA, Babič A, Macedo J, Dolšek M, Gregor N, Kottke AR et al. (2021) Overview and introduction to development of non-ergodic earthquake ground-motion models. *arXiv preprint arXiv: 2111.07921*
- Lawrence N, Seeger M and Herbrich R (2003) Fast sparse gaussian process methods: The informative vector machine. In: *Proceedings of the 16th annual conference on neural information processing systems*, CONF. pp. 609–616
- Lin PS, Chiou B, Abrahamson N, Walling M, Lee CT, Cheng CT (2011) Repeatable source, site, and path effects on the standard deviation for empirical ground-motion prediction models. *Bull Seismol Soc Am* 101(5):2281–2295
- Lin PS, Chiou B, Abrahamson N, Walling M, Lee CT, Cheng CT (2011) Repeatable source, site, and path effects on the standard deviation for empirical ground-motion prediction models. *Bull Seismol Soc Am* 101(5):2281–2295
- Liu C, Macedo J, Candia G (2021) Performance-based probabilistic assessment of liquefaction-induced building settlements. *Soil Dyn Earthquake Eng* 151:106955
- Liu C, Macedo J and Kottke AR (2022) Evaluating the performance of nonergodic ground motion models in the ridgecrest area. *Bulletin of Earthquake Engineering* : 1–27
- Loth C, Baker JW (2013) A spatial cross-correlation model of spectral accelerations at multiple periods. *Earthquake Eng Struct Dyn* 42(3):397–417
- Macedo J, Candia G, Lacour M, Liu C (2020) New developments for the performance-based assessment of seismically-induced slope displacements. *Eng Geol* 277:105786
- Macedo J, Ramesh V, Liu C, Kottke A (2022) Evaluating different approaches for the hazard-consistent assessment of the seismic performance of dams. *Bull Seismol Soc Am* 112(3):1710–1726
- Melkumyan A, Ramos F (2009) A sparse covariance function for exact gaussian process inference in large datasets. *IJCAI* 9:1936–1942
- NCREE (2015) Web page for reevaluation of probabilistic seismic hazard of nuclear facilities in taiwan using sshac level 3 methodology project. available at <http://sshac.ncree.org.tw>
- Paciorek CJ, Schervish MJ (2006) Spatial modelling using a new class of nonstationary covariance functions. *Environ Off J Int Environ Soc* 17(5):483–506
- Park J, Bazzurro P, Baker JW et al. (2007) Modeling spatial correlation of ground motion intensity measures for regional seismic hazard and portfolio loss estimation. *Applications of statistics and probability in civil engineering* : 1–8
- Patel S, Ceferino L, Liu C, Kiremidjian A, Rajagopal R (2021) The disaster resilience value of shared rooftop solar systems in residential communities. *Earthquake Spectra* 37(4):2638–2661

- Plesch A, Shaw JH, Ross ZE, Hauksson E (2020) Detailed 3d fault representations for the 2019 ridgecrest, California, earthquake sequence. *Bull Seismol Soc Am* 110(4):1818–1831
- Rahpeyma S, Halldorsson B, Hrafinkelsson B, Jónsson S (2018) Bayesian hierarchical model for variations in earthquake peak ground acceleration within small-aperture arrays. *Environmetrics* 29(3):e2497
- Rasmussen CE (2003) Gaussian processes in machine learning. In: *Summer school on machine learning*. Springer, pp. 63–71
- Rekoske JM, Thompson EM, Moschetti MP, Hearne MG, Aagaard BT, Parker GA (2020) The 2019 ridgecrest, California, earthquake sequence ground motions: Processed records and derived intensity metrics. *Seismol Res Lett* 91(4):2010–2023
- Rodriguez-Marek A, Montalva GA, Cotton F, Bonilla F (2011) Analysis of single-station standard deviation using the kik-net data. *Bull Seismol Soc Am* 101(3):1242–1258
- Rodriguez-Marek A, Cotton F, Abrahamson NA, Akkar S, Al Atik L, Edwards B, Montalva GA, Dawood HM (2013) A model for single-station standard deviation using data from various tectonic regions. *Bull Seismol Soc Am* 103(6):3149–3163
- Rodriguez-Marek A, Rathje E, Bommer JJ, Scherbaum F, Stafford P (2014) Application of single-station sigma and site-response characterization in a probabilistic seismic-hazard analysis for a new nuclear site. *Bull Seismol Soc Am* 104(4):1601–1619
- Schiappapietra E, Douglas J (2020) Modelling the spatial correlation of earthquake ground motion: Insights from the literature, data from the 2016–2017 central italy earthquake sequence and ground-motion simulations. *Earth-sci reviews* 203:103139
- Schiappapietra E, Smerzini C (2021) Spatial correlation of broadband earthquake ground motion in Norcia (central Italy) from physics-based simulations. *Bull Earthquake Eng* 19(12):4693–4717
- Sgobba S, Lanzano G, Pacor F, Puglia R, D'Amico M, Felicetta C, Luzi L (2019) Spatial correlation model of systematic site and path effects for ground-motion fields in northern Italy. *Bull Seismol Soci Am* 109(4):1419–1434
- Sgobba S, Lanzano G, Pacor F (2021) Empirical nonergodic shaking scenarios based on spatial correlation models: An application to central Italy. *Earthquake Eng Struct Dyn* 50(1):60–80
- Stafford PJ, Zurek BD, Ntinalexis M, Bommer JJ (2019) Extensions to the groningen ground-motion model for seismic risk calculations: component-to-component variability and spatial correlation. *Bull Earthquake Eng* 17(8):4417–4439
- Stewart JP, Afshari K, Goulet CA (2017) Non-ergodic site response in seismic hazard analysis. *Earthquake Spectra* 33(4):1385–1414
- Villani M, Abrahamson NA (2015) Repeatable site and path effects on the ground-motion sigma based on empirical data from southern California and simulated waveforms from the cybershake platform. *Bull Seismol Am* 105(5):2681–2695
- Walling MA (2009) Non-ergodic probabilistic seismic hazard analysis and spatial simulation of variation in ground motion. University of California, Berkeley

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