Chen, Y., and Baker, J. W. (2019). "Spatial correlations in CyberShake physics-based ground motion simulations." *Bulletin of the Seismological Society of America*, 109(6), 2447-2458.
https://doi.org/10.1785/0120190065

Spatial correlations in CyberShake physics-based ground motion simulations

Yilin Chen Jack W. Baker
Department of Civil and Environmental Engineering, Stanford University

yilinc2@stanford.edu | bakerjw@stanford.edu

Abstract

When studying the performance of distributed infrastructure in earthquakes, spatial variations in strong ground motion have a significant impact. Currently, prediction models for spatial ground motion variations in future earthquakes are calibrated using ground motion observations from densely recorded earthquakes. While useful, that calibration process requires strong assumptions about stationarity and isotropy of correlations. This paper reports results from conducting analogous spatial variation estimation using physics-based simulations from the CyberShake platform. This platform contains simulated ground motions from hundreds of thousands of rupture realizations, at locations throughout Southern California, providing a synthetic ground motion catalog that is much richer than we could ever hope to achieve from recordings. That richness allows significant relaxation of stationarity and isotropy assumptions, and provides new insights regarding the role of source and path heterogeneity on the spatial correlation of ground motion amplitudes. The results suggest that geological conditions, source effects and path effects have significant impacts on spatial correlations. Additionally, this work serves as a new dimension of ground motion simulation validation, as the estimated correlations can be compared to results from past earthquakes.

Introduction

During an earthquake, the shaking intensity varies spatially. Predicting the ground motion intensities is important for evaluating performance of distributed infrastructure in earthquakes (e.g., Chang *et al.*, 2000). In order to predict the ground motion intensities across a region, ground motion prediction equations (GMPEs, e.g., Chiou and Youngs, 2014; Abrahamson *et al.*, 2014; Boore *et al.*, 2014; Campbell and Bozorgnia, 2014; Idriss, 2014) have been developed to predict the median and standard deviation of ground motion intensity measures, such as spectral acceleration at a given period. The difference between an observed amplitude and its median prediction is referred to as a ground motion residual. Modeling these residuals is essential to understand the spatial characteristics of ground motion caused by earthquakes, and is important for obtaining unbiased estimates of lifeline risk (Jayaram and Baker, 2010).

Previous studies have investigated the spatial correlation of the residuals using empirical earthquake records and simulations. Wang and Takada (2005) studied the correlation in residuals from K-NET and KiK-NET data, and pointed out that the correlation coefficients decay exponentially as distance increases. Park *et al.* (2007) discussed potential causes of correlations due to source and path effects. Goda and Hong (2008) investigated the spatial correlation of spectral acceleration

residuals at different periods using California and Chi-Chi earthquake records. Jayaram and Baker (2009) applied geostatistical tools to seven earthquake datasets with magnitude from 5 to 7 and suggested that correlations are affected by variability in near-surface site stiffness. Sokolov *et al.* (2012) studied the correlation in residuals from ground motions in Taiwan, and showed that larger correlation in residuals is related to larger correlation in V_{s30}. Infantino *et al.* (2018) studied the spatial correlation of ground motion using broad-band physics-based numerical simulations, and found that the spatial correlations in simulations are consistent with an empirical model at long and short periods. Although the data used by these studies are different, they result in similar prediction models.

While useful, these spatial correlation models all assumed stationarity and isotropy: the authors all assumed that any pair of sites with equal separation distance within an earthquake has the same correlation (stationarity), and that the correlation is independent of orientation (isotropy). Therefore, most of these models depend only on the separation distance between the sites of interest. These assumptions are used out of necessity due to the lack of repetitive ground motion intensity records at the same site. Therefore, pairs of stations are aggregated based on their separation distance, and other factors that might influence the spatial correlation tend to be hidden. Recently, physics-based simulations of strong ground motions have been extensively studied. Compared with recorded earthquakes, simulations provide repetitive well-structured spatially distributed ground motion data, which conceptually is a preferable data source to quantify spatial correlations.

In this paper, we relax the assumptions of stationarity and isotropy by adopting CyberShake simulations to calculate the correlation coefficients for every pair of stations in a region. CyberShake is a computational study to calculate ground motion hazard in the Los Angeles region (Graves *et al.*, 2011), and has previously been used for ground motion studies and simulation validation (Wang and Jordan, 2014; Villani and Abrahamson, 2015). It includes simulations of over 415,000 rupture realizations at 336 sites, and simulates wave propagation through a three-dimensional velocity model that reflects the impact of sedimentary basins and near-surface materials on ground motion. It provides an opportunity to investigate the impact of other factors, such as site geological condition and rupture property, on spatial correlation.

Modeling spatial correlation of ground motions

Data Sources

In this study, we used 5%-damped response spectra of ground motions simulated at 336 stations using the version 15.12 CyberShake runs (see Data and Resources Section). Figure 1 shows the locations of the stations considered in this study. Two ruptures, Puente Hills (M_w = 6.65) and San Andreas (M_w = 8.05) are studied in depth for comparison of stationary and non-stationary correlations. A number of additional ruptures were studied to explore patterns in observed correlations, and illustrative results from five additional ruptures are provided below. There are 32 and 355 rupture realizations for the Puente Hills and San Andreas ruptures, respectively. Here 'rupture' refers to a specified earthquake magnitude and spatial extent, and 'rupture realization' refers to a particular simulation of a hypocenter and slip distribution. Strain Green tensors are computed for each source-site pair of interest, and then for each rupture realization the relevant strain Green tensors are post-processed using reciprocity to simulate seismograms, which are then used to compute intensity measures such as spectral accelerations. These rupture realizations are used to estimate the residuals of ground motions and are essential to reveal rupture-specific spatial correlation behavior.

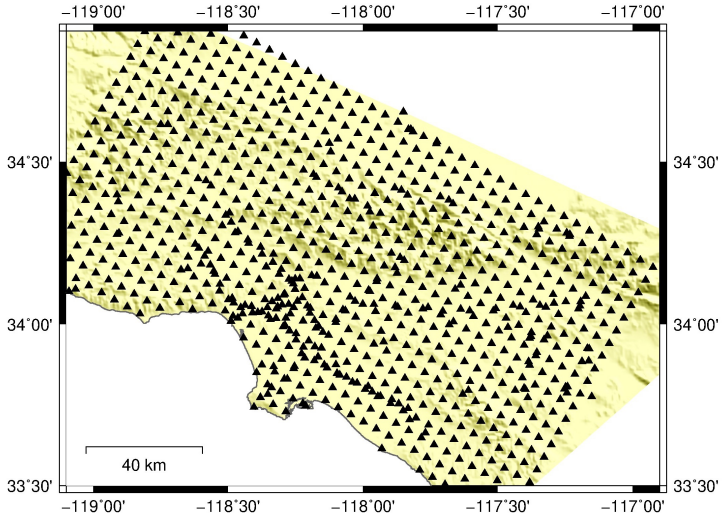

Figure 1: Locations of stations in the CyberShake study.

Definition of residuals

Consider a typical ground motion model (e.g., Atik *et al.*, 2010) for spectral accelerations

$$\ln SA_{i,j} = \mu_{\ln SA}(Rup_i, Site_j) + \delta B_i + \delta W_{i,j} \tag{1}$$

where $SA_{i,j}$ is the spectral acceleration at site j caused by the i^{th} rupture (with a given period and damping); $\mu_{\ln SA}(Rup_i, Site_j)$ is the natural logarithmic mean of spectral acceleration intensity. In this study, we utilized the simulations themselves to decompose the ground motion amplitudes into this form. The $\mu_{\ln SA}(Rup_i, Site_j)$ term is calculated as the sample mean of $\ln SA$ at site j over a series of rupture realizations with the same rupture extent. This was done to avoid any potential inconsistency between the simulation amplitudes and those predicted by an empirical ground motion model. The difference between the mean prediction and the i^{th} SA observation is characterized by δB_i, the between-event residual for the i^{th} rupture, and $\delta W_{i,j}$, the within-event residual for site j from the i^{th} rupture. The standard deviations of δB_i and $\delta W_{i,j}$ are denoted τ and ϕ, respectively. The total residual $\varepsilon^t_{i,j}$ is the sum of the within- and between-event residuals:

$$\varepsilon^t_{i,j} = \delta B_i + \delta W_{i,j} = \ln SA_{i,j} - \mu_{\ln SA}(Rup_i, Site_j) \tag{2}$$

Under the assumptions of the above ground motion model, $\delta W_{i,j}$ is a normal random variable with zero mean. Since there are 336 stations ($n = 336$), a reasonable estimate of δB_i without utilizing mixed effects regression is given by:

$$\delta \hat{B}_i = \frac{1}{n}\sum_{j=1}^{n} \varepsilon^t_{i,j} = \frac{1}{n}\sum_{j=1}^{n} \ln SA_{i,j} - \mu_{\ln SA}(Rup_i, Site_j) \tag{3}$$

where $\hat{}$ denotes an estimate.

Then, the within-event residual can be estimated by:

$$\delta \hat{W}_{i,j} = \ln SA_{i,j} - \mu_{\ln SA}(Rup_i, Site_j) - \delta \hat{B}_i \tag{4}$$

Using Equation 2, the following relationship between the correlation in within-event residuals and correlation in total residuals can be derived (Park *et al.*, 2007):

$$\rho^t(h) = \frac{\tau^2 + \phi^2 \rho(h)}{\tau^2 + \phi^2} \tag{5}$$

where $\rho(h)$ is the correlation coefficient for within-event residuals separated by a distance h, and $\rho^t(h)$ is the correlation coefficient for total residuals. Note that the difference between $\rho(h)$ and $\rho^t(h)$ is caused by the variation of between-event residual. If $\tau = 0$, $\rho(h)$ will be same as $\rho^t(h)$. As $h \to \infty$, $\rho^t(h)$ is the ratio of the between-event residual variance to the total residual variance. We note that, per Equation 1, correlation in spectral acceleration amplitudes can both come from correlations in residuals as discussed above, as well as correlations in mean predictions over random future ruptures. The latter source of correlations is already characterized by standard ground motion models, so it is the former source that merits further study (Baker and Miller, 2011; Giorgio and Iervolino, 2016).

Pearson's correlation coefficient

In this paper, we used two methods, semivariogram and Pearson's correlation coefficient, to evaluate the spatial correlation of residuals in CyberShake simulations. Pearson's correlation coefficient is a measure of the linear correlation between two variables. The correlation coefficient of ground motion residuals between site j and site k can be estimated as:

$$\hat{\rho}(j,k) = \frac{\sum_{i=1}^{n} z_{i,j} z_{i,k}}{\sqrt[2]{\sum_{i=1}^{n} {z_{i,j}}^2} \sqrt[2]{\sum_{i=1}^{n} {z_{i,k}}^2}} \tag{6}$$

where n is the number of rupture realizations and z is the within-event residual or total residual. Note that within-event residuals and total residuals have differing correlation coefficient estimates because the latter includes a common δB in both z values (Equation 2). The mean of residuals of n rupture realizations is zero according to Equation 3 and 4. Pearson's correlation coefficient quantifies the spatial correlation in a non-stationary and anisotropic fashion, since it gives the correlation coefficient between any site j and site k instead of a given separation distance. However, it requires multiple observations of residual z at site j and site k, and thus is not commonly used when studying empirical earthquake data. The richness of CyberShake data allows us to use Pearson's correlation coefficient to study the non-stationary and anisotropic behaviors of spatial correlation in ground motions.

We note here that these correlations are referring to correlation in ground motion amplitude, rather than ground motion coherence (i.e., correlation in the ground motion time series at two sites). To illustrate, Figure 2 shows example data used to compute theses correlations. Figure 2a shows ground motion residuals at two stations with positive correlation, and Figure 2b shows residuals for two stations with negative correlation. In the positive correlation case, these data indicate that if the s389 station has $SA(3s)$ greater than its mean amplitude for this rupture, then the $SA(3s)$ amplitude at s385 is also likely to be larger than its mean amplitude. In the negative correlation case, if the s389 has $SA(3s)$ greater than its mean amplitude, then s393 is likely to have $SA(3s)$ less than its mean amplitude. In the limits, a correlation of 1 indicates that the amplitudes at two sites are perfectly linearly related, and a correlation of -1 indicates that they are perfectly oppositely related.

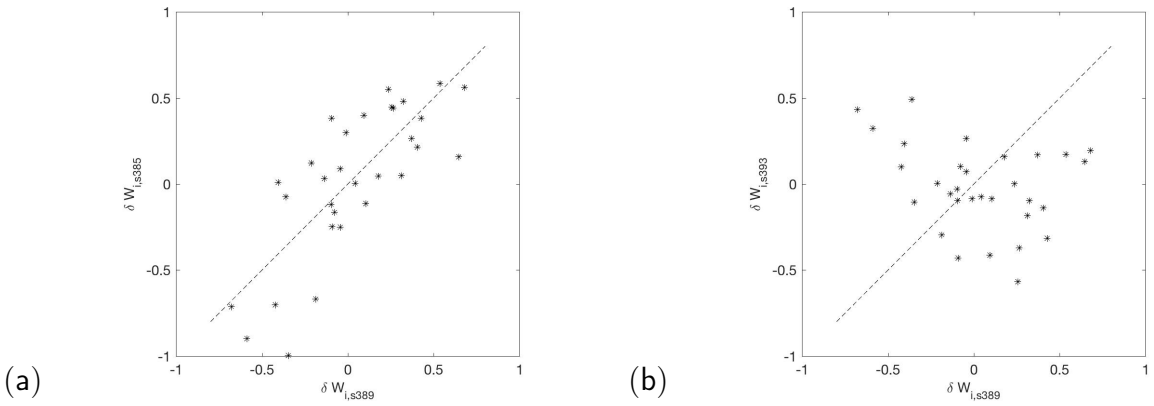

Figure 2: Residuals of spectral acceleration at $T = 3s$ at (a) stations s389 and s385 ($\hat{\rho} = 0.76$) and (b) s389 and s393 ($\hat{\rho}$ = -0.43). Both station pairs have $h = 20km$, and all data come from simulations of the $M_w = 6.65$ Puente Hills rupture.

Furthermore, Pearson's correlation coefficient under stationarity and isotropy assumptions can be evaluated by averaging the correlation coefficients of all pairs of sites separated by the same distance. Specifically, the expected value of Pearson's correlation coefficient under stationarity and isotropy assumptions at distance h can be regarded as:

$$\mathbb{E}[\rho(h)] = \mathop{\mathbb{E}}_{d(j,k)=h}[\rho(j,k)] \tag{7}$$

which can be estimated by

$$\hat{\rho}(h) = \frac{1}{N} \sum_{|d(j,k)-h|<\Delta h} \rho(j,k) \tag{8}$$

where $d(j,k)$ is the separation distance between site j and site k, and the summation condition takes the average of the correlation coefficients of all pairs of sites with separation distances between $h-\Delta h$ and $h+\Delta h$, Δh is a distance threshold and N is the number of such pairs.

Semivariogram

The semivariogram is a geostatistical tool to measure the dissimilarity of spatially distributed data (Journel and Huijbregts, 1978; Goovaerts, 1997), and is commonly used in modeling geological data (e.g., Marinoni, 2003; Zhang and Zhu, 2018) and evaluating spatial correlations of ground motions (e.g., Goda and Hong, 2008; Jayaram and Baker, 2009; Goda and Atkinson, 2010; Esposito and Iervolino, 2011; Foulser-Piggott and Stafford, 2012; Loth and Baker, 2013; Infantino *et al.*, 2018). It measures half the mean squared difference between variable z at location j and k:

$$\gamma(j,k) = \frac{1}{2}\mathbb{E}[(z(j) - z(k))^2] \tag{9}$$

which can be empirically estimated by:

$$\hat{\gamma}(j,k) = \frac{1}{2N} \sum_{\alpha=1}^{N} [z(j_\alpha) - z(k_\alpha)]^2 \tag{10}$$

where α represents α^{th} data pair and N is the number of data pairs at location (j, k). Since one often does not possess several observations of ground motion intensity at a given pair of sites, the assumptions of stationarity and isotropy have to be made in order to evaluate Equation 10. Under these assumptions, the semivariogram is specified by the separation distance h instead of location (j, k):

$$\gamma(h) = \frac{1}{2N}\sum[z(u) - z(u+h)]^2 = \frac{1}{2N}\sum_{|d(j,k)-h|<\Delta h}[z(j) - z(k)]^2 \tag{11}$$

Note that, in contrast to Equation 6, this result is the same whether z represents within-event or total residuals, because for total residuals the δB is common to $z(j)$ and $z(k)$ and so will cancel. Furthermore, the correlation coefficient between $z(u)$ and $z(u+h)$ is:

$$\rho(h) = 1 - \frac{\gamma(h)}{\sigma_z^2} \tag{12}$$

where σ_z is the standard deviation of $z(u)$. We fit the semivariogram using the following exponential function, with two parameters referred to as the sill (S) and range (R):

$$\tilde{\gamma}(h) = S\left(1 - \exp\left(\frac{-3h}{R}\right)\right) \tag{13}$$

Figure 3 shows an example of a fitted semivariogram for ground motion residuals caused by the San Andreas rupture, and we can use Equation 12 to convert the semivariogram to equivalent correlation coefficients.

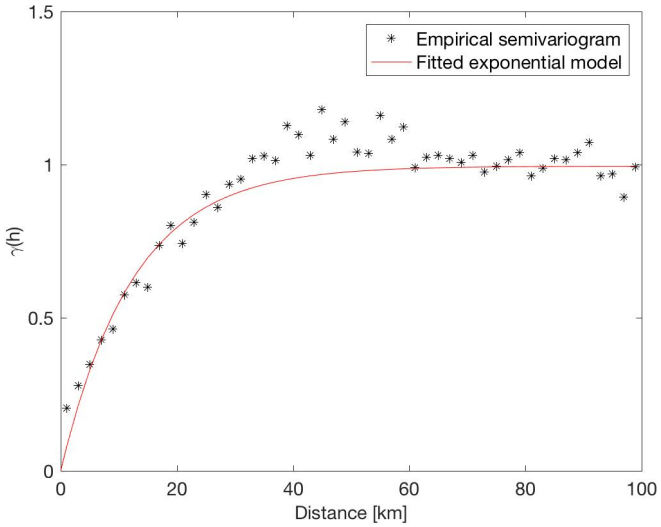

Figure 3: Empirical and fitted exponential semivariogram for ground motion residuals given the $M_w = 8.05$ San Andreas rupture. Individual points on the plot are estimated using Equation 11, and the fitted model is from Equation 13.

Spatial correlations without stationarity and isotropy assumptions

We used Equation 6 to calculate the Pearson's correlation coefficients between every pair of sites across the region, and used the results to examine the relationship between correlation coefficients

and factors other than separation distance.

Dependence of spatial correlations on geological condition

In order to investigate the effect of geological condition on spatial correlations, we select a reference site and calculate the correlation coefficient between it and every other site. Then a heatmap of correlation coefficients is provided to visualize spatial patterns in the results.

Heatmaps of correlation coefficients of within-event residuals for spectral acceleration at $T = 3s$ caused by the San Andreas rupture are shown in Figure 4. Two reference sites on different site conditions, site s345 in the Los Angeles basin and site s383 on rock, are selected for comparison. Correlation coefficients predicted by an empirical model from Jayaram and Baker (2009) are used here as a reference. The empirical correlation coefficients decay uniformly in all orientations away from the reference site, because this model depends on separation distance only. The correlation level also generally decreases with distance from the reference site in CyberShake simulations. However, the basin region shows a higher correlation level when the reference site is also in the basin. Similarly, the rock region shows a higher correlation level when the reference site is on rock. For reference, Figure 5 shows basin depths in the CyberShake model, where $Z_{1.0}$ is the depth below ground surface to a shear-wave velocity of $1km/s$ (with large values indicating sedimentary basins) and surficial site conditions, where V_{s30} is the time-averaged shear-wave velocity to $30m$ depth (with small values indicating surficial soils). This observation is consistent with the idea that earthquakes may generate similar ground motions in regions that have similar geology.

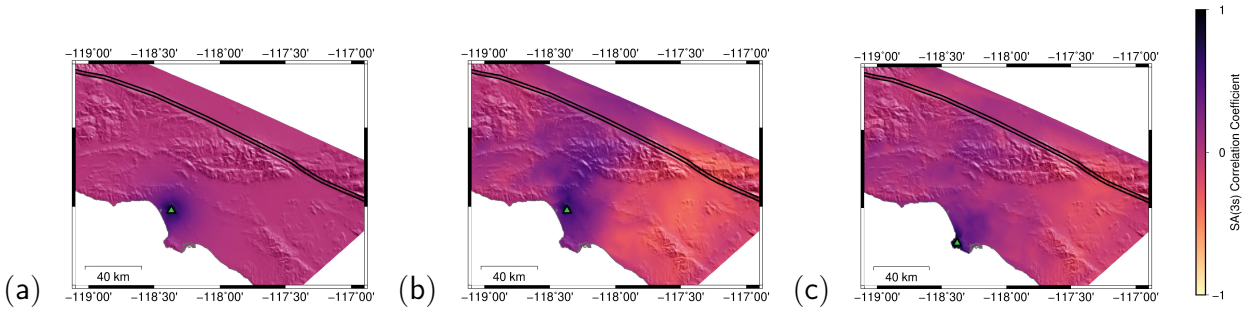

Figure 4: Heatmaps of correlation coefficients of within-event residuals for spectral acceleration at $T = 3s$ caused by the $M_w = 8.05$ San Andreas rupture. (a) Empirical Model, reference site: s345; (b) CyberShake, reference site: s345; (c) CyberShake, reference site: s383. Black lines show the surface projection of the rupture. The triangle shows the reference site. Shading shows the correlation coefficient with the reference site.

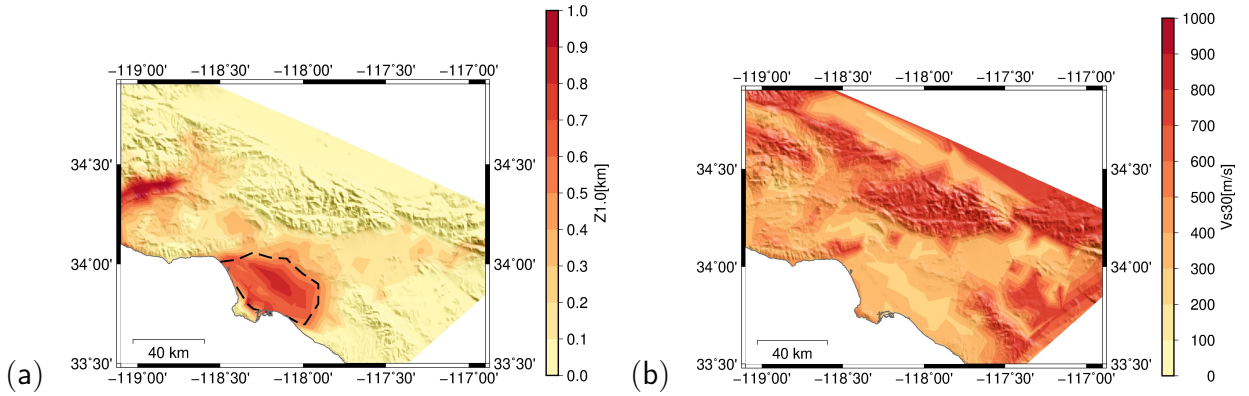

Figure 5: (a) Map of $Z_{1.0}$ values in CyberShake region. Red regions indicate sedimentary basins. Dashed lines show the boundary of the basin region. Data from Small *et al.* (2017). (b) Map of V_{s30} values in CyberShake region.

Dependence of spatial correlations on period

We next vary the SA period of interest to evaluate the role of period in correlation coefficients. In Figure 6, the regional distribution of correlation coefficients of within-event residuals at periods $T = 1, 5, 10s$ are shown. The correlations increase as the period of interest increases for both CyberShake simulations and the empirical model. For sites separated by more than $20km$, increasing the period of interest has less impact on the correlation coefficient. It is noticeable that the decay of correlation with distance in CyberShake simulations is slower than the empirical model, especially for larger periods. Additionally, the direction perpendicular to the rupture shows higher correlation, for a given distance from the reference site.

Dependence of spatial correlations on source

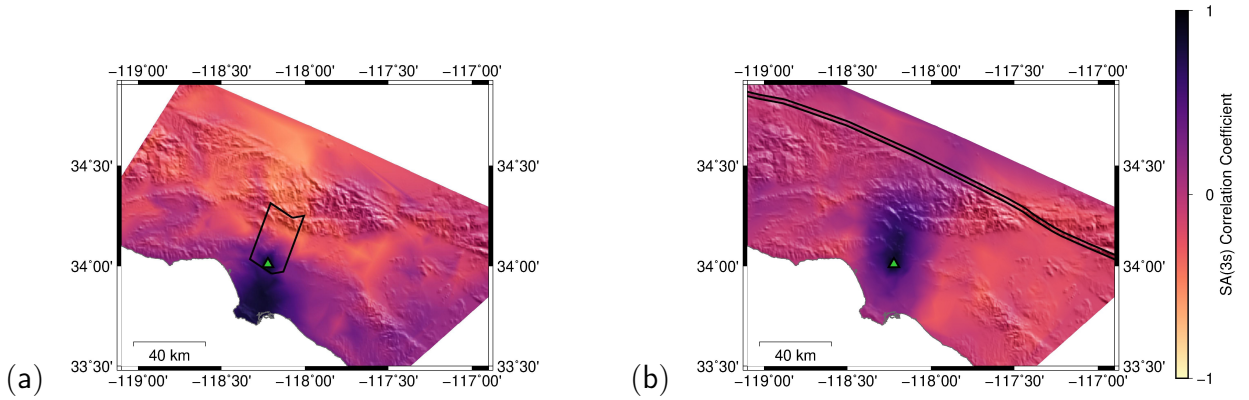

Figure 7: Heatmap of correlation coefficients of within-event residual for spectral acceleration at $T = 3s$, (a) Puente Hills rupture; (b) San Andreas rupture. Reference site: s389. Black lines show the surface projection of the rupture.

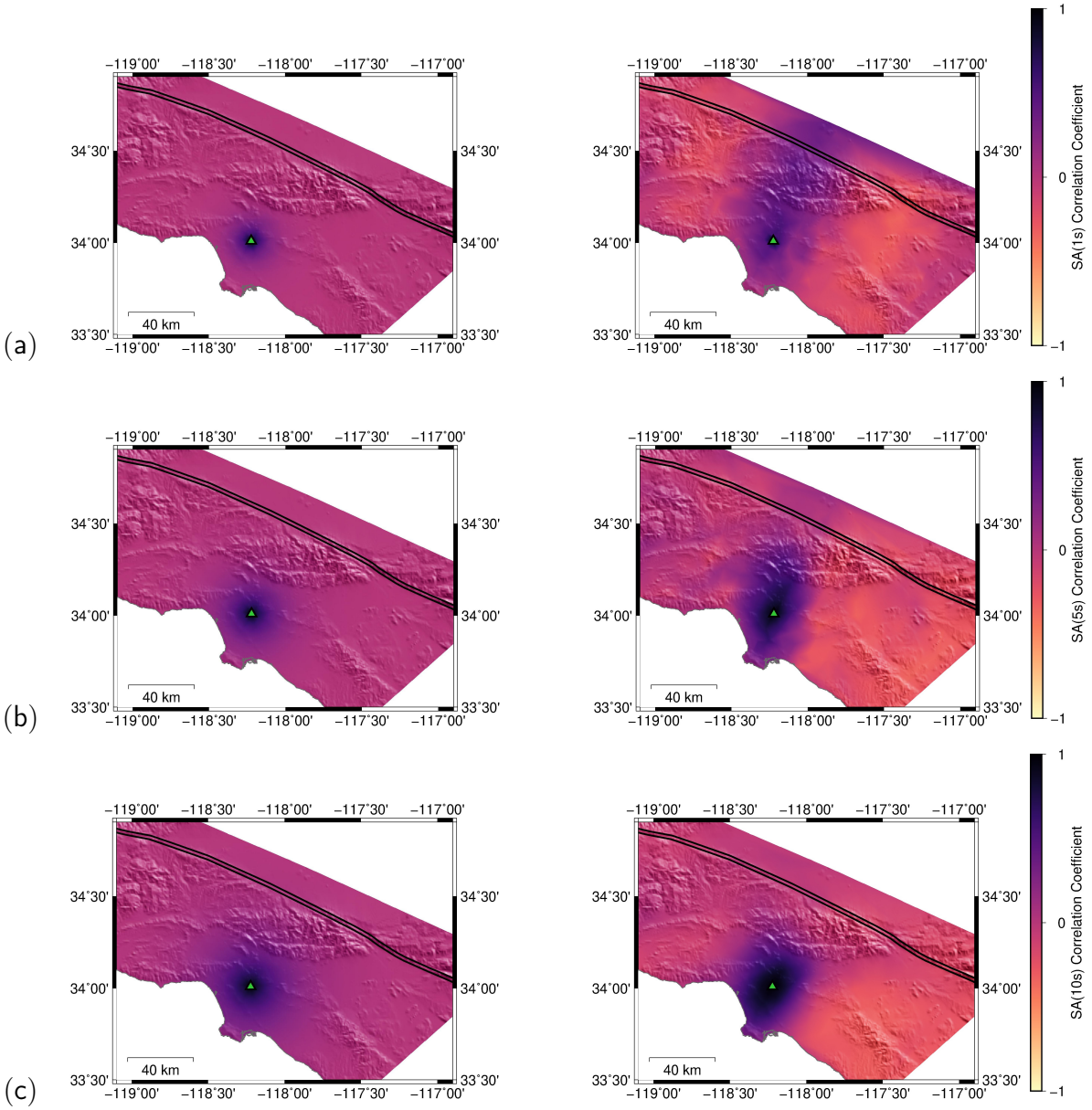

Figure 6: Heatmaps of correlation coefficients of within-event residuals for spectral acceleration at $(a)\ T = 1s, (b)\ T = 5s, (c)\ T = 10s$. Empirical Model (left), CyberShake (right). Reference site: s389.

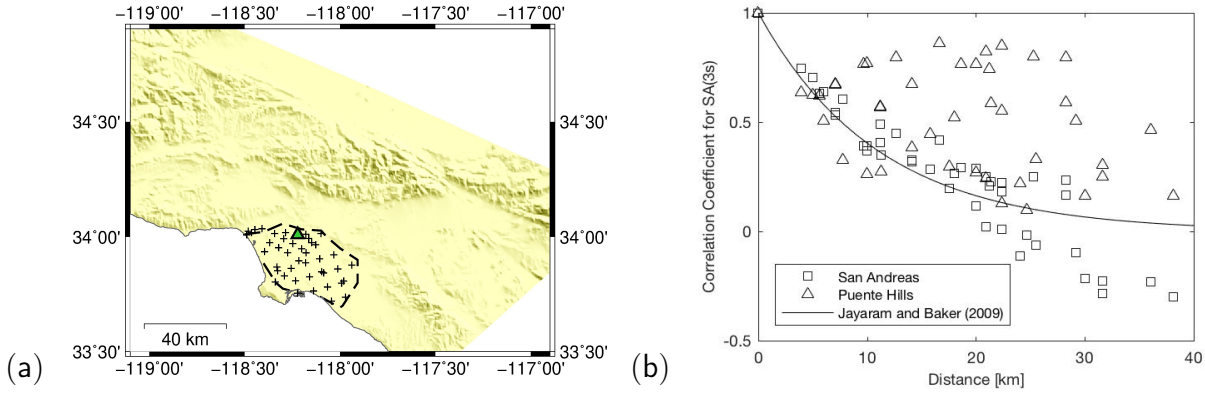

Figure 8: (a) Stations in the basin region considered. (b) Correlation coefficients of within-event residual for sites in the basin region at $T = 3s$. Each of the data point is the correlation coefficient between a site in the basin and the reference site: s389.

Figure 7 shows heatmaps of correlation coefficients of within-event residuals from the Puente Hills and San Andreas ruptures. The spatial patterns of correlation coefficients are very different. In the San Andreas case, the correlation coefficient decays moving away from the reference site, and it becomes almost zero when the separation distance is greater than $40km$. The Puente Hills rupture, however, shows two discrete correlation states in the region. For the area southwest of the rupture (i.e., Los Angeles basin), the correlation level is significantly higher than from the San Andreas rupture; For the area northwest of the rupture, the correlation level is lower and even becomes negative.

To further illustrate the difference of correlation caused by these two ruptures in the basin region, Figure 8 shows the value of correlation coefficients for the sites in the basin region. In the San Andreas case, the decay rate of correlation with respect to distance corresponds well to the empirical model, but the decay of correlation is very slow in the Puente Hills case and the correlation is still significant when separation distance is above $20km$.

The results shown in Figure 7 and 8 can be explained by source effects. In the Puente Hills case, strong and concentrated waveforms propagate southward into the Los Angeles basin (Graves and Somerville, 2006), which results in systematic regional deviation of predictions in the basin region and thus causes high correlation in residuals at large separation distances.

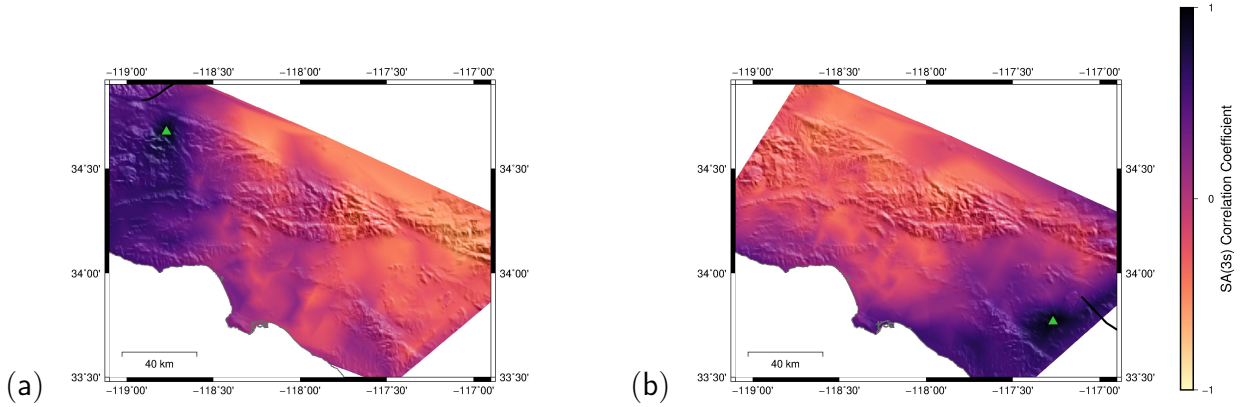

Figure 9: Heatmap of correlation coefficients of within-event residuals for spectral acceleration at $T = 3s$ caused by the (a) $M_w = 6.95$ Garlock rupture, reference site: s078; (b) $M_w = 6.95$ San Jacinto rupture, reference site: s768.

Figure 9 provides similar evidence of source-related non-stationarity for Garlock and San Jacinto ruptures. In Figure 9a, the Garlock rupture is located at the northwest of the study region, and the reference site is $15km$ southeast of the rupture. The sites in the northwest region are positively correlated with the reference site, while the eastern region is negatively correlated with the reference site. In Figure 9b, the San Jacinto rupture is in the east with the reference site $15km$ away from the rupture. It shows that the reference site is positively correlated with sites in the southeast region but negatively correlated with the northwest region. It suggests that at small distances sites close to the same part of the rupture have a strong correlation ("source effect").

Dependence of spatial correlation on path

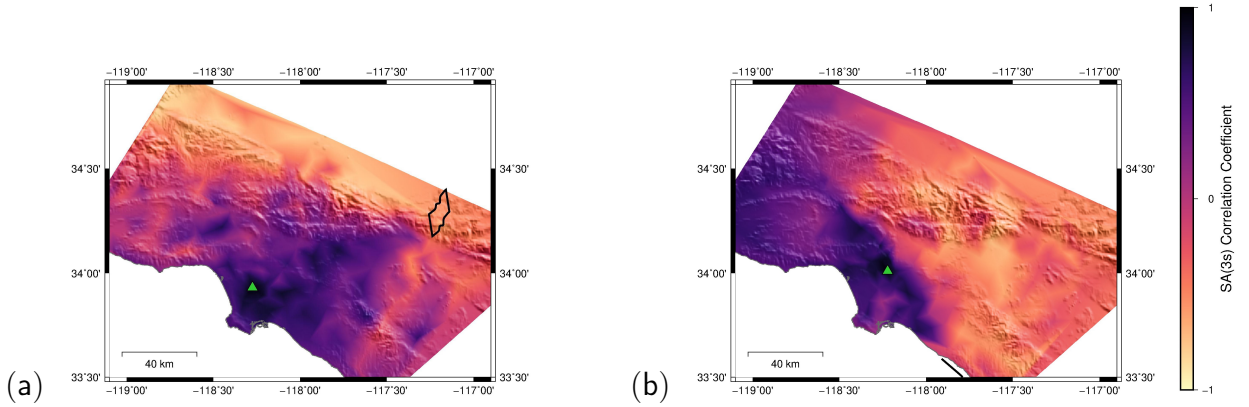

Figure 10: Heatmap of correlation coefficients of within-event residuals for spectral acceleration at $T = 3s$ caused by the (a) $M_w = 6.55$ North Frontal rupture, reference site: s387; (b) $M_w = 6.55$ Newport rupture, reference site: s389.

Figure 10a shows correlations for a reference site approximately $80km$ away from the North Frontal rupture. In this case, there is a band of positively correlated sites along the wave propagation direction between the source and the reference site. Similarly, in Figure 10b, there is a narrow strongly correlated region with the reference site at the end of the Newport rupture. It suggests

that at moderate to large distances (30 to 100km), strong correlations of sites could be caused by common wave propagation paths ("path effect").

Dependence of spatial correlation on relative location to rupture

In Figure 11a, the reference site s558 is about 30km away from the Puente Hills rupture. The region correlated with site s558 is much wider than for the reference site s389 and than predicted by empirical models, and it is negatively correlated with the opposite region of the rupture. Figure 11b shows a similar result from the San Gabriel rupture, where the west and east region are also negatively correlated. It suggests that sites on opposite ends of a rupture are sometimes negatively correlated.

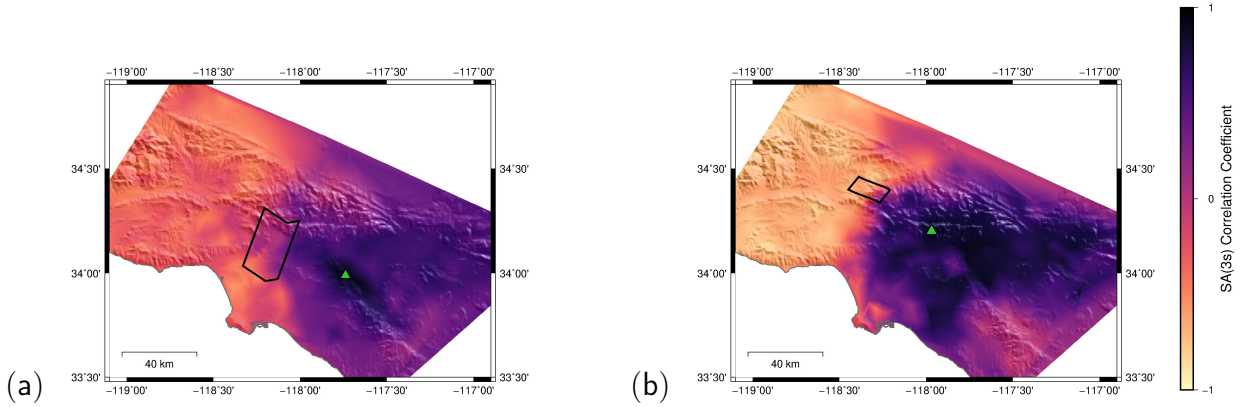

Figure 11: Heatmap of correlation coefficients of within-event residuals for spectral acceleration at $T = 3s$ caused by the (a) $M_w = 6.65$ Puente Hills rupture, reference site: s558; (b) $M_w = 6.55$ San Gabriel rupture, reference site: s439.

Spatial correlations under stationarity and isotropy assumptions

Semivariogram fitting results

We use Equation 4 to calculate the residuals for single rupture realizations, and use Equation 11 to calculate the empirical semivariogram for residuals, with $\Delta h = 1km$. Then the semivariogram is modeled using the exponential function of Equation 13. We estimate the sill and range using the mode of the histogram of the semivariogram values scaled by a Gaussian kernel function (Loth and Baker, 2013). Finally, we use Equation 12 to calculate the correlation coefficients at various separation distances.

In Figure 12, the correlation coefficients are calculated for a Puente Hills and a San Andreas rupture realization. Four previous studies used as references (Goda and Hong, 2008; Jayaram and Baker, 2009; Loth and Baker, 2013; Heresi and Miranda, 2019) are also provided. The CyberShake semivariograms have a similar rate of decay (i.e., range), and are close to the empirical models. Although Figure 8 showed that the Puente Hills rupture had correlations that decayed more slowly with distance in the basin region, the range is similar to San Andreas when fitting the semivariogram. This is due to the averaging of results over many station pairs with comparable distances.

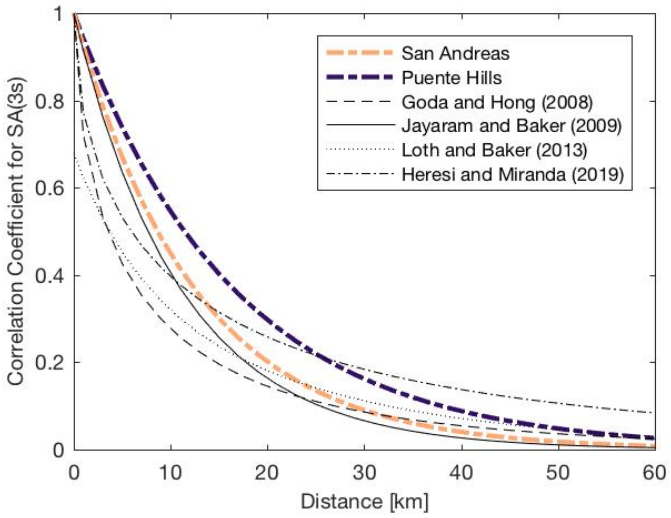

Figure 12: Correlation coefficients for spectral acceleration at $T = 3s$ based on semivariogram fitting.

Pearson's correlation coefficient results

In order to evaluate the Pearson's correlation coefficient results in CyberShake simulations under stationarity and isotropy assumptions, we average the correlation coefficients of pairs of sites separated by the same distance. In this case, the non-stationary and anisotropic effects are averaged out. We use Equation 8 to calculate the averaged correlation coefficients with the reference site s389 in CyberShake simulations, and those results are compared to the empirical models in Figure 13. In Figure 13, each data point is an average of the correlation coefficients with $\Delta h = 2.5km$. The correlation coefficients observed in CyberShake simulations show the same pattern of decay with distance as the empirical models. The total residual shows slightly higher correlation. The standard deviation of between-event residuals in the San Andreas case ($\hat{\tau} = 0.12$) is higher than for Puente Hills ($\hat{\tau} = 0.08$), which causes a larger difference in the two correlation coefficients. In both cases, the correlation coefficients of within-event residuals correspond well to the empirical models. However, the patterns of correlation coefficients without stationarity and isotropy assumptions in these two cases (shown in Figure 7) are quite different. Figure 14 shows the averaged correlation coefficients of all pairs of sites in the CyberShake region. A 90% confidence interval at distance h is estimated by the 5% and 95% empirical quantile of the data with $h \in [h-2.5km, h+2.5km]$. Compared with the results in Figure 13, the averaged coefficients of all pairs of sites are slightly higher than the empirical models, which is caused by the higher correlated pairs (e.g., s558 in Figure 11a). In addition, the confidence interval of the correlation coefficient is wider in Puente Hills case, which is related to the strong non-stationary effect as shown in Figure 7. While the estimated correlation coefficients for Puente Hills rupture have slightly higher finite-sample variability due to the smaller number of rupture realizations, this variability is relatively small compared with the width of the confidence interval in Figure 14, and so does not have a material effect on these conclusions.

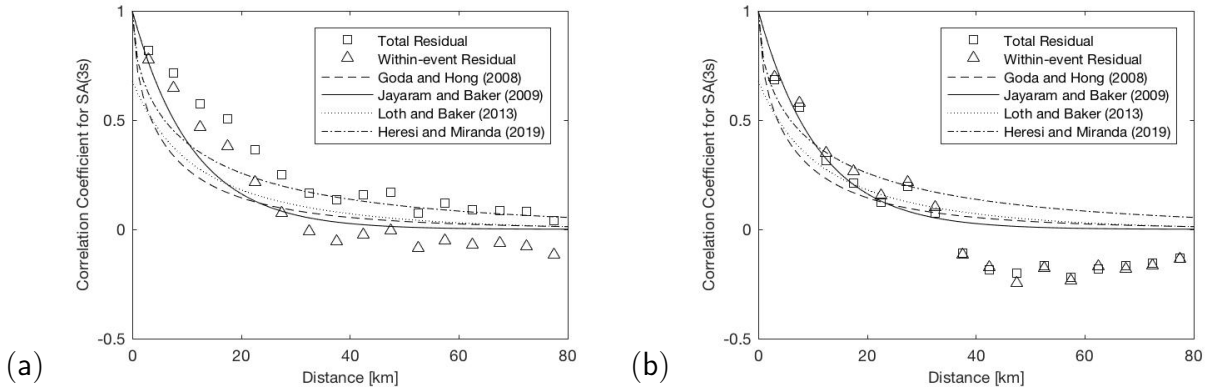

Figure 13: Averaged correlation coefficients of within-event residuals and total residuals for spectral acceleration at $T = 3s$ varied with distance. (a) San Andreas rupture, (b) Puente Hills rupture. Reference site: s389.

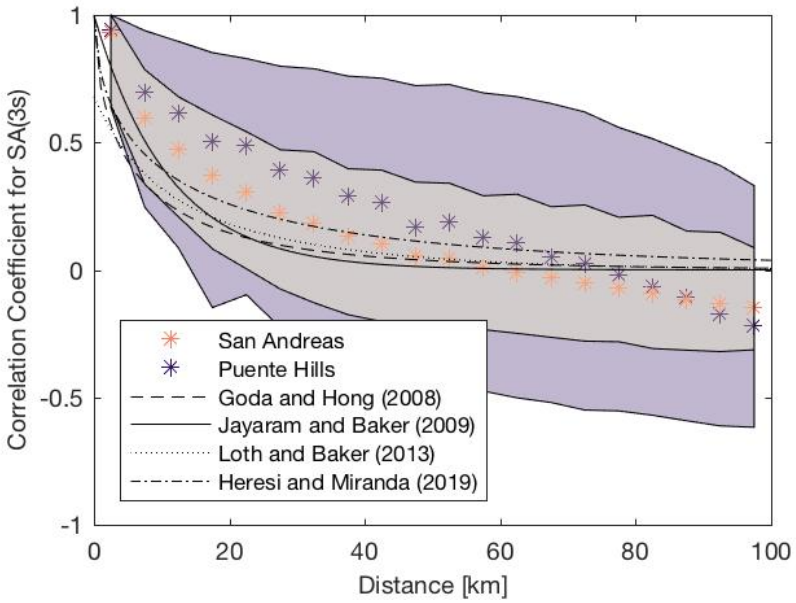

Figure 14: Averaged correlation coefficients of within-event residuals for all pairs of sites at $T = 3s$. Shading shows the 90% confidence interval.

Conclusions

We evaluated the spatial correlation in spectral acceleration ground motion residuals using CyberShake ground motion simulations. Semivariogram and Pearson's correlation coefficient calculations were applied to quantify the correlation in ground motion residuals. The method of Pearson's correlation coefficient is able to show the non-stationary and anisotropic behavior of spatial correlation. The results suggest that the correlation in within-event residuals will be higher for pairs of sites with similar geological conditions. The correlation level tends to increase as the period of interest increases. It is observed that there is anisotropy in spatial correlations for long periods: the

correlation coefficients are higher perpendicular to rupture than parallel to rupture direction for the San Andreas rupture data. In addition, the rupture property also has significant influence on spatial correlation. The Puente Hills fault generates strong correlation along the wave propagation direction, while the correlation coefficients can even be negative when the stations are in opposite directions. It suggests that non-stationary and anisotropic spatial correlations depend on the source effect, path effect and relative location to rupture. At small distances, sites close to the same part of the rupture tend to have stronger correlation ("source effect"); at moderate to large distances, sites with common wave propagation paths tend to have stronger correlation ("path effect"); sites on opposite ends of a rupture are sometimes negatively correlated ("relative location to rupture").

Under stationarity and isotropy assumptions, both the Pearson's correlation coefficient results and semivariogram results for the reference site s389 show similar correlation in within-event residuals for the Puente Hills and San Andreas ruptures, and they correspond well to empirical models. However, the correlation coefficient of total residual depends on the variation of between-event residual, which can vary with rupture. The spatial correlation considering all pairs of sites in CyberShake simulations is slightly higher than the empirical models. The strong non-stationary spatial correlations present in the Puente Hills simulations appears to cause the wider confidence interval of correlation coefficients for that rupture.

This study identified several factors which influence spatial correlations. It suggests that physics-based simulations could be a useful data source when assessing risks for distributed infrastructure systems, as they could capture non-stationarities in spatial correlations of ground motions. Further, it points towards a path for incorporating such effects in empirical correlation predictive models. However, confirming these findings with real earthquake data is challenging due to insufficient data, especially for longer periods. Future physics-based simulations that produce shorter-period response spectra, as well as future earthquakes recorded by modern instruments, will both facilitate future comparisons. Future studies of infrastructure or other regional systems using these simulations will also provide insight as to whether these non-stationarities cause significant impacts on risk.

While the observations in this study are specific to this CyberShake data set, the analysis approach proposed here illustrates how ground motion simulations can provide insight into the role of rupture and geological conditions on the spatial patterns of ground motions from future earthquakes. And in turn, comparison of simulation results to empirical models allows for validation of simulations, and possibly for calibration of simulation parameters related to spatial heterogeneity that may be difficult to constrain by other means.

Data and Resources

Ground motion data used in this paper are from the version 15.12 CyberShake runs, which can be obtained at `https://scec.usc.edu/scecpedia/CyberShake_Study_15.12` (last accessed March 2019). Some plots were made using the Generic Mapping Tools version 5.4.4 (Wessel and Smith, 1998).

Acknowledgements

We thank Scott Callaghan for help accessing the CyberShake database, and Brendon Bradley and Ethan Thomson for help with the Figure 5a basin map. We thank Yixiao Sheng for help interpreting the results. This research was supported by the Southern California Earthquake Center (Contribution No. 8225). SCEC is funded by NSF Cooperative Agreement EAR-1033462 & USGS Cooperative Agreement G12AC20038.

References

Abrahamson, N. A., W. J. Silva, and R. Kamai (2014). Summary of the ASK14 ground motion relation for active crustal regions, *Earthquake Spectra* **30**(3), 1025–1055.

Atik, L. A., N. Abrahamson, J. J. Bommer, F. Scherbaum, F. Cotton, and N. Kuehn (2010). The variability of ground-motion prediction models and its components, *Seismological Research Letters* **81**(5), 794–801.

Baker, J. W., and M. K. Miller (2011). Effects of earthquake source geometry and site conditions on spatial correlation of earthquake ground motion hazard, in *Plenary lecture at 4th IASPEI/IAEE International Symposium on Effects of Surface Geology on Seismic Motion*, p. 12.

Boore, D. M., J. P. Stewart, E. Seyhan, and G. M. Atkinson (2014). NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes, *Earthquake Spectra* **30**(3), 1057–1085.

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.

Chang, S. E., M. Shinozuka, and J. E. Moore (2000). Probabilistic earthquake scenarios: extending risk analysis methodologies to spatially distributed systems, *Earthquake Spectra* **16**(3), 557–572.

Chiou, B. S.-J., and R. R. Youngs (2014). Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra, *Earthquake Spectra* **30**(3), 1117–1153.

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.

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.

Giorgio, M., and I. Iervolino (2016). On multisite probabilistic seismic hazard analysis, *Bulletin of the Seismological Society of America* **106**(3), 1223–1234.

Goda, K., and G. M. Atkinson (2010). Intraevent spatial correlation of ground-motion parameters using SK-net data, *Bulletin of the Seismological Society of America* **100**(6), 3055–3067.

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.

Goovaerts, P. (1997). *Geostatics for natural resources evaluation*, Oxford University Press.

Graves, R., T. H. Jordan, S. Callaghan, E. Deelman, E. Field, G. Juve, C. Kesselman, P. Maechling, G. Mehta, K. Milner, *et al.* (2011). CyberShake: A physics-based seismic hazard model for southern California, *Pure and Applied Geophysics* **168**(3-4), 367–381.

Graves, R., and P. Somerville (2006). Broadband ground motion simulations for scenario ruptures of the Puente Hills fault, in *Proc. 8th National Conference on Earthquake Engineering*, pp. 18–21.

Heresi, P., and E. Miranda (2019). Uncertainty in intraevent spatial correlation of elastic pseudo-acceleration spectral ordinates, *Bulletin of Earthquake Engineering* **17**(3), 1099–1115.

Idriss, I. (2014). An NGA-West2 empirical model for estimating the horizontal spectral values generated by shallow crustal earthquakes, *Earthquake Spectra* **30**(3), 1155–1177.

Infantino, M., R. Paolucci, C. Smerzini, and M. Stupazzini (2018). Study of the spatial correlation of earthquake ground motion by means of physics-based numerical scenarios, *16th European Conference on Earthquake Engineering* .

Jayaram, N., and J. W. Baker (2009). Correlation model for spatially distributed ground-motion intensities, *Earthquake Engineering & Structural Dynamics* **38**(15), 1687–1708.

Jayaram, N., and J. W. Baker (2010). Efficient sampling and data reduction techniques for probabilistic seismic lifeline risk assessment, *Earthquake Engineering & Structural Dynamics* **39**(10), 1109–1131.

Journel, A. G., and C. J. Huijbregts (1978). *Mining geostatistics*, Academic press London.

Loth, C., and J. W. Baker (2013). A spatial cross-correlation model of spectral accelerations at multiple periods, *Earthquake Engineering & Structural Dynamics* **42**(3), 397–417.

Marinoni, O. (2003). Improving geological models using a combined ordinary–indicator kriging approach, *Engineering Geology* **69**(1-2), 37–45.

Park, J., P. Bazzurro, J. W. Baker, *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* pp. 1–8.

Small, P., D. Gill, P. J. Maechling, R. Taborda, S. Callaghan, T. H. Jordan, K. B. Olsen, G. P. Ely, and C. Goulet (2017). The SCEC unified community velocity model software framework, *Seismological Research Letters* **88**(6), 1539–1552.

Sokolov, V., F. Wenzel, K. L. Wen, and W. Y. Jean (2012). On the influence of site conditions and earthquake magnitude on ground-motion within-earthquake correlation: Analysis of PGA data from TSMIP (Taiwan) network, *Bulletin of Earthquake Engineering* **10**(5), 1401–1429.

Villani, M., and N. A. Abrahamson (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, *Bulletin of the Seismological Society of America* **105**(5), 2681–2695.

Wang, F., and T. H. Jordan (2014). Comparison of probabilistic seismic-hazard models using averaging-based factorization, *Bulletin of the Seismological Society of America* **104**(3), 1230–1257.

Wang, M., and T. Takada (2005). Macrospatial correlation model of seismic ground motions, *Earthquake Spectra* **21**(4), 1137–1156.

Wessel, P., and W. H. Smith (1998). New, improved version of Generic Mapping Tools released, *Eos, Transactions American Geophysical Union* **79**(47), 579–579.

Zhang, Q., and H. Zhu (2018). Collaborative 3d geological modeling analysis based on multi-source data standard, *Engineering Geology* **246**, 233–244.