17th World Conference on Earthquake Engineering, 17WCEE

Sendai, Japan - September 13th to 18th 2020

Paper N° XXXX (Abstract ID)

Registration Code: S-XXXXXXXX

PROGRESS IN MEASURING SPATIAL CORRELATIONS IN GROUND MOTION INTENSITY

J.W. Baker[(1)], M. Markhvida[(2)], Y. Chen[(3)]

[(1)] *Professor, Stanford University, bakerjw@stanford.edu*
[(2)] *Ph.D. Graduate, Stanford University, markhvid@stanford.edu*
[(3)] *Graduate Student, Stanford University, yilinc2@stanford.edu*

Abstract

For nearly 20 years, studies of spatial correlations in ground motion intensity have been an area of active research. Models of spatial correlation are integral for assessing seismic risk to distributed infrastructure systems, portfolios of insured properties, or any other systems affected by shaking at more than one location. These spatial correlation models have been calibrated using ground motion data from Japan, USA, Italy, New Zealand, and elsewhere around the world. Models have been developed for a range of ground motion intensity metrics, including spectral acceleration at several periods, peak ground acceleration, peak ground velocity, and Arias Intensity. Moreover, studies have been performed to evaluate whether these correlations depend upon the tectonic region, the underlying geotechnical conditions, or the properties of the causal earthquake rupture.

Despite the importance of these models, their development is significantly constrained by the availability of suitable recorded ground motions at closely spaced stations. With that context, this paper reviews recent developments by the authors in two areas: (1) new statistical techniques for estimating spatial correlations from data, (2) measurement of spatial correlations from physics-based ground motion simulations. These developments will be described, their potential impact on seismic risk assessment will be explained, and prospects for future developments will be discussed.

Keywords: ground motion; spatial correlation; regional risk assessment

1 Introduction

There are a number of application areas where understanding of the spatial distribution of ground shaking intensity is important. When considering probabilistic assessments of risk, ground shaking intensity must be characterized probabilistically. And it is not just the marginal distributions of shaking intensity at individual sites that must be considered. The joint distribution across sites is needed in order to understand risk of large cumulative losses, or the risk of disruption of infrastructure systems that are dependent upon components at a number of locations. As an example, Fig. 1 shows loss results for a distributed transportation network, where earthquake occurrence and resulting ground motion variability are random. The figure shows that the likelihood of extreme losses is larger when spatial correlations in ground motion amplitudes are considered. This is because it is more likely that extended regions of the study area all experience shaking that is stronger-than-expected in a given event. The same phenomenon has been observed in several other studies [1–5].

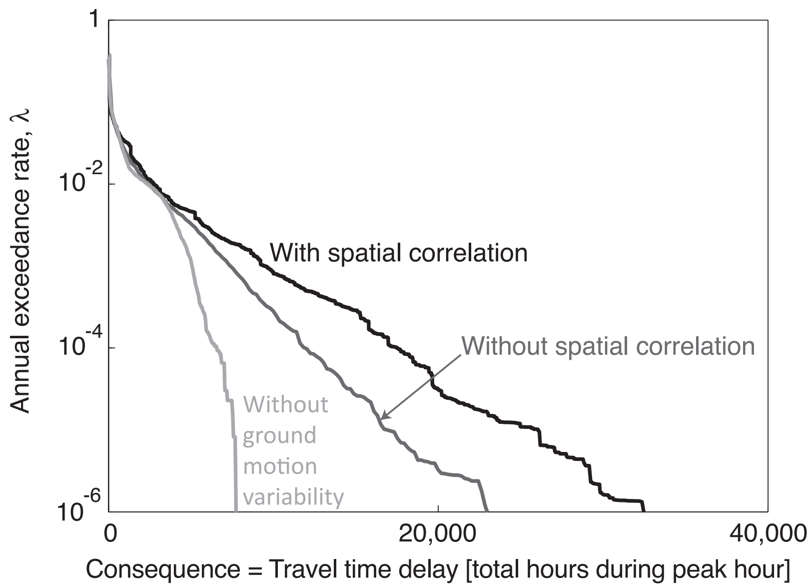

Fig. 1 – Effect of spatial correlation in ground motions on estimated risk of travel time disruption over a transportation network (from [6]).

The acquisition (beginning in the 1990’s) of ground motion recordings on relatively dense scales allowed for the empirical study of such spatial correlations. Boore et al. [7] performed a statistical assessment of ground motion recordings from the 1994 Northridge earthquake. Takada and Wang [8] performed a similar analysis of the 1999 Chi-Chi earthquake, and five Japanese earthquakes that occurred between 2000 and 2003—analyses that were possible due to the recently installed dense seismic networks in those countries. In the subsequent years, a number of studies have performed similar analyses, using larger data sets, data sets from other regions, and refined analysis algorithms. At present, there are a number of empirical models that predict ground motion intensity correlations for regions around the world, and for a variety of ground motion intensity metrics [9–16].

Recent further developments in the field relate to studies and algorithms for estimation of cross-correlations of multiple ground motion intensity metrics, and studies of finer-scale variations in ground shaking intensity using physics-based ground motion simulations. These topics will be reviewed in the following sections, followed by conclusions and remarks on prospects for further progress in this field.

2 Estimation of spatial correlations

In this section we briefly introduce the spatial correlation phenomenon of interest here. These correlations are defined with respect to the typical formulation of ground motion amplitudes using empirical Ground Motion Models (GMMs). A GMM predicts a ground motion intensity measure (*IM*) from earthquake rupture *i* at site *j* (here *j*=1 and 2, indicated by subscripts) as a function of rupture and site properties.

$$\begin{aligned} \ln IM_{i,1} &= \mu_{\ln IM}(rup_i, site_1) + \delta B_i + \delta W_{i,1} \\ \ln IM_{i,2} &= \mu_{\ln IM}(rup_i, site_2) + \delta B_i + \delta W_{i,2} \end{aligned} \tag{1}$$

where $\mu_{\ln IM}(\ )$ is the mean predicted ln*IM* value, as a function of rupture (*rup*) parameters such as magnitude and faulting type, and site (*site*) parameters such as source-to-site distance and soil conditions, among others [e.g., 17,18]. The δB_i and $\delta W_{i,j}$ terms are the between- and within-event residuals, representing deviations between observed and mean predicted ln*IM* values. These residuals are normally distributed random variables with means of zero and standard deviations denoted τ_i and ϕ_i (where the *i* index denotes that the standard deviations are in some cases a function of rupture properties). The between-event residual, δB_i, is common for the two sites because it depends upon the rupture and not the specific site.

Standard GMMs provide the functional form of $\mu_{\ln IM}(\ )$, and the standard deviations τ_i and ϕ_i. So, for this spatial correlation application, the only uncharacterized portion of the model is the dependence of $\delta W_{i,1}$ and $\delta W_{i,2}$, which requires an additional model for how within-event residuals vary in space.

For spatial correlation calculations, we invert equation (1) to solve for $\delta W_{i,j}$. Further, it is helpful to normalize by the standard deviation, in order to get a standardized random variable *z*, as follows

$$z_{i,j} = \frac{\delta W_{i,j}}{\phi_i} = \frac{\ln IM_{i,j} - \mu_{\ln IM}(rup_i, site_j) - \delta B_i}{\phi_i} \tag{2}$$

The data required for spatial correlation estimation are the residuals from well-recorded earthquake events, as illustrated in Fig. 2. Fig. 2a shows *SA*(1s) values recorded from the 2000 M_w 6.6 Tottori, Japan earthquake. These are the ln*IM* values with respect to equation (1). The Chiou and Youngs GMM [18] is then used to compute $\mu_{\ln IM}(\ )$ for each recording condition, and the normalized residuals are computed for each station location using equation (2), as shown in Fig. 2b. The areas of stations with similar coloring in Fig. 2b indicate the presence of spatial correlations in these data.

These data can then be used to estimate a semivariance, which is a measure of dissimilarity.

$$\gamma_{j,k} = \frac{1}{2} E\left[\left(Z_j - Z_k\right)^2\right] \tag{3}$$

where *j* and *k* are two locations of interest, γ denotes the semivariogram, and *E*[ ] denotes expectation [19].

If we assume that the semivariance depends only upon the separation distance between the two locations (*h*), it can be estimated from empirical data as

$$\hat{\gamma}(h) = \frac{1}{2n(h)} \sum_{d(j,k)=h} \left(z_{i,j} - z_{i,k}\right)^2 \tag{4}$$

where *d(j,k)* is the distance between sites *j* and *k*, the summation is over all *j* and *k* with separation distance *h* (within a user-specified tolerance), and *n(h)* is the number of observed station pairs with separation distance *h*. Typically, an empirical semivariance is computed from data using equation (4), and then a parametric equation is fit to the resulting data to create a predictive model. An example will be shown later in Fig. 4.

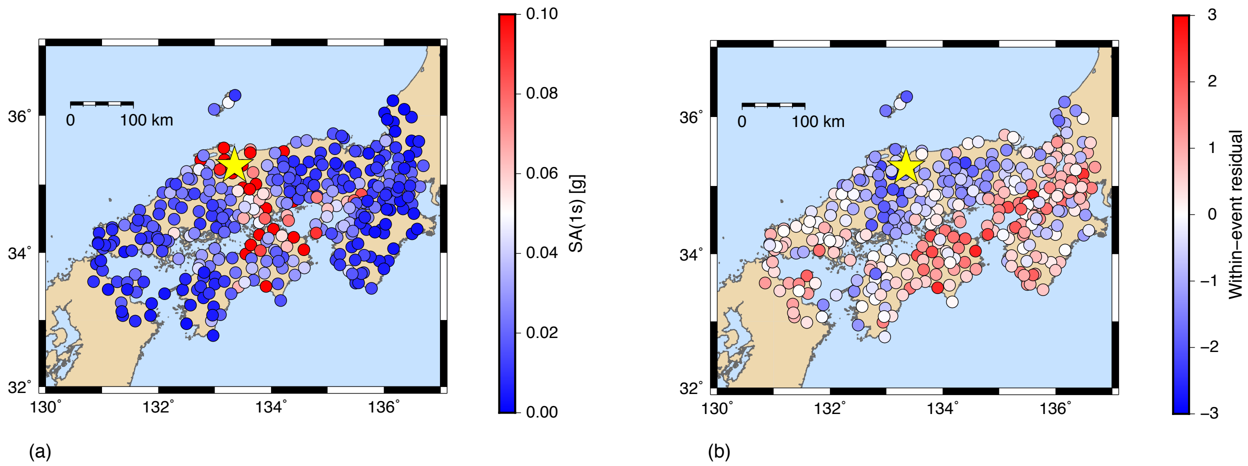

Fig. 2 (a) Observed *SA*(1s) values from the 2000 M_w 6.6 Tottori, Japan earthquake. (b) SA(1s) within-event residuals from the Tottori earthquake, with respect to the Chiou and Youngs GMM [18]. Data are from the NGA-West 2 database [20]. The earthquake epicenter is denoted with the star.

3 Estimation of spatial cross-correlations

In addition to the estimation of correlations of an individual *IM* across space, it is sometimes of interest to know the correlations between multiple *IM*s at multiple locations. Most frequently, the *IM*s would be spectral acceleration values at multiple periods, but they could also be any other type of *IM* metric. Generically, if we denote two types of intensity measures as *IM* and *IM**, then equation (1) can be modified and written as

$$\begin{aligned} \ln IM_{i,1} &= \mu_{\ln IM}(rup_i, site_1) + \delta B_i + \delta W_{i,1} \\ \ln IM^*_{i,2} &= \mu^*_{\ln IM}(rup_i, site_2) + \delta B^*_i + \delta W^*_{i,2} \end{aligned} \quad (5)$$

As before, it is the correlation between $\delta W_{i,j}$ and $\delta W^*_{i,j}$ that needs to be specified, in addition to the information available from GMMs. Additionally, the correlation δB_i and δB^*_i is now needed; that can be obtained from a single-location correlation model [e.g., 21]. Goda and Hong [15] built a correlation model for spectral accelerations at multiple periods using the product of a formula for cross-*IM* correlation at a single location, multiplied by a spatial correlation term for spectral acceleration at the longer of the two periods of interest. Loth and Baker [22] proposed a linear model of coregionalization, which assumed that the cross-covariances of spectral accelerations at multiple periods are a linear combination of several underlying structures, and then performed a fitting exercise to estimate those structures. Both of these models provide probabilistically valid predictions of covariances (i.e., non-negative definite covariance matrices). The Goda and Hong model makes a somewhat strong (though empirically reasonable) assumption about correlation structure, and the Loth and Baker model is somewhat difficult to fit. Perhaps more importantly, both models produce large covariance matrices that are difficult to work with for forward simulations of large numbers of *IM* locations and types.

With that motivation, Markhvida et al. [23] recently proposed a strategy to deal with spatially correlated *IM*s using Principal Component Analysis (PCA) [24] to translate the *IM* vector at a given location into an uncorrelated set of principal components. Let **Z** be an *m* x *n* matrix of within-event residuals where each row represents a distinct *IM* metric, and each column a given recording station (in this case, spectral accelerations at $m = 19$ periods ranging from 0.01 to 5 seconds, and pooled $n = 4910$ records from 45 earthquakes). Let **Y** be an *m* x *n* matrix of *m* uncorrelated principal components. PCA is a linear transformation as follows:

$$\mathbf{PZ} = \mathbf{Y} \tag{6}$$

where **P** is an *m* x *m* orthogonal linear transformation matrix determined through the PCA analysis.

For the ground motion within-event residual data of interest here, Fig. 3 illustrates the terms in **P** relating to various principal components. Each subfigure graphically illustrates the 19 **P** coefficients for one principal component. This provides some intuition as to what these principal components represent. The first component has comparable coefficients at all periods, so it indicates the general amplitude of the response spectrum. The second component has positive coefficients at long periods and negative coefficients at short periods, so it indicates the relative amplitudes of long- and short-period spectral values. Further components reflect more subtle features in the spectra, and their overall contribution to **Z** decreases with increasing component number.

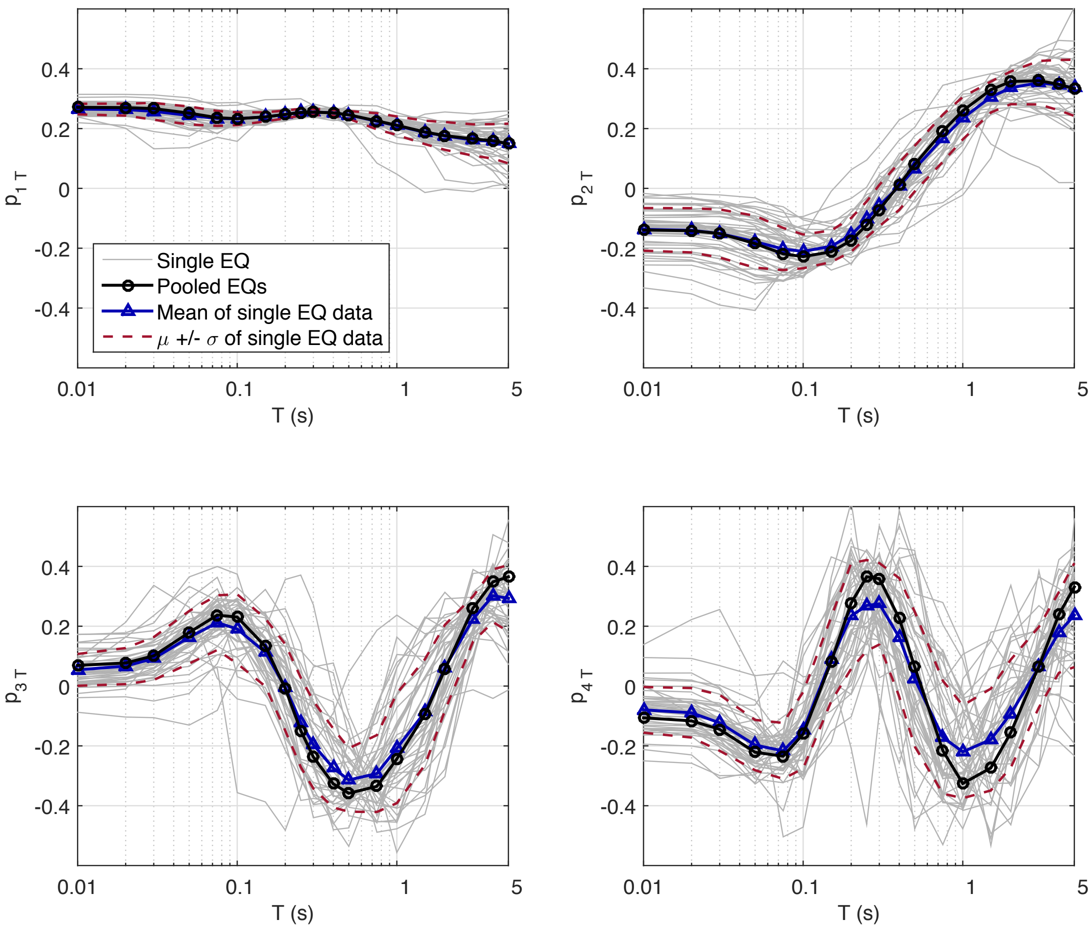

Fig. 3 – Coefficients corresponding to spectral acceleration periods, *T*, for the first four principal components of normalized ground motion residuals ($P_{i,T}$), as per equation (6). From [23]

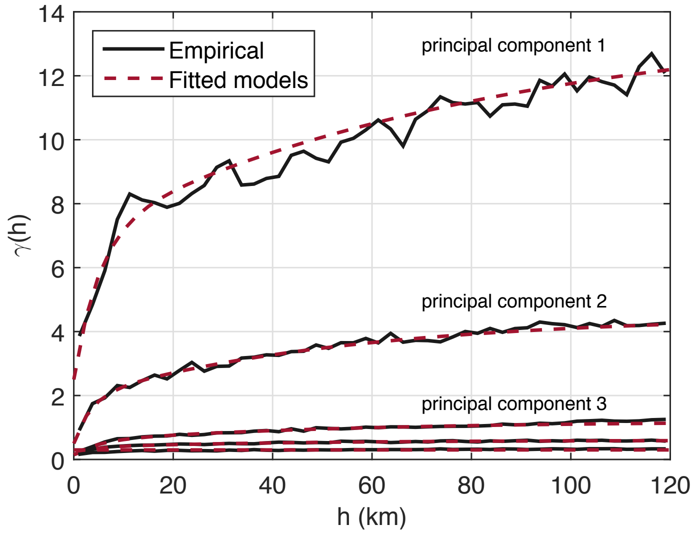

Fig. 4 – Empirical semivariograms and fitted semivariogram models for the first five spectral accelerations principal components (from [23]).

Fig. 3 also shows the transformation coefficients corresponding to different periods for 45 earthquakes with a sufficient number of ground motions to be considered for spatial analysis. Each of the lines in the figure shows the principal component coefficients if each earthquake was considered separately for a PCA transformation. The heavy line shows the coefficients if all data are pooled. The stability of these PCA coefficients indicates that a single transformation can be performed for the pooled data set.

Since the rows of **Y** are uncorrelated, the spatial variability characterization of each row (i.e., each principal component) can be performed independently of other rows. By changing the basis in this way, PCA can provide us with the ability to reduce the dimensionality of the problem to a smaller number of independent dimensions. For spatial correlation estimation, each principal component can have its own variogram fit independently of the others (Fig. 4), which is a more straightforward exercise than the joint fitting of, e.g., [22]. And in forward simulation, the principal components can be (independently) simulated for each location of interest, and then back-transformed to the spectral acceleration residuals via the inverse of equation (3). This decoupled forward simulation greatly reduces computational effort relative to prior models that required a coupled simulation of all *IM*s jointly.

Regarding the forward-simulation of residuals for risk analysis, this approach has great computational advantages. By decoupling the correlations amongst principal components, there is a dramatic improvement in computational time for simulating residuals. This is because there is no longer a need to invert a large covariance matrix representing covariances amongst all locations and all *IM*s—such matrices have thousands of elements for regional risk assessment problems where many locations and *IM*s are of interest. Instead, several smaller matrices are considered. For problems with several thousand locations and several IMs of interest (a typical situation for regional risk analysis), this approach can take less than 1/100th of the time to simulate residuals compared to an approach where all locations and *IM*s are considered in a single covariance matrix [23].

The drawbacks of this method are that the *IM*s of interest must be pre-specified during the fitting stage, and all *IM*s must be computable for each ground motion in the **Z** matrix of equation (6). This first drawback is also relevant to the Loth and Baker [22] model, though not the Goda and Hong [15] model. The second drawback can be slightly limiting for, e.g., recordings from older analog instruments, where long-period spectral accelerations are not resolvable due to signal-to-noise issues. This is a relatively minor issue, however, as older recordings are generally not associated with densely-recorded earthquakes. Overall, these limitations should be minor relative to the advantages for most practical problems.

Example Matlab source code to perform simulations using the approach described in this section is available at https://github.com/bakerjw/Spatial_PCA.

4 Spatial correlations in physics-based ground motion simulations

A fundamental limitation in the estimation of spatial correlations is data availability. It is challenging to obtain recordings of ground motions with sufficient density to estimate spatial correlations—only a few regions in the world have strong ground motion seismic networks sufficient for this task. Further, because earthquakes are rare, and each is unique in size, location and other factors, it is difficult to obtain data to quantify how earthquake properties and regional factors influence spatial correlations.

While we wait for more empirical data to provide insights about spatial correlations, numerical simulation of ground motion offers an intriguing opportunity to perform numerical experiments about this issue [25]. The concept is to use a "physics-based" ground motion simulation that incorporates rupture process and wave propagation in the simulation of earthquake ground shaking. Shaking simulations can be performed for locations of arbitrary configuration, and the rupture processes can be varied systematically, allowing for controlled experiments. The resulting ground motions can then be used to compute *IM*s, and the spatial correlation can be studied. If the physics of the real earthquake process are sufficiently captured by the simulation algorithm (an assumption requiring validation), the resulting spatial correlation results can provide insights about factors causing variations in spatial correlations amongst earthquakes or regions of the world.

Chen and Baker [26] recently used numerical ground motion simulations to study spatial correlations in this way. They considered the CyberShake ground motion simulations [27]—a large set of simulations performed on a dense grid of locations in the Los Angeles region (Fig. 5). Simulated ground motions, and resulting *IM*s, were computed for hundreds of earthquake simulations from the data set.

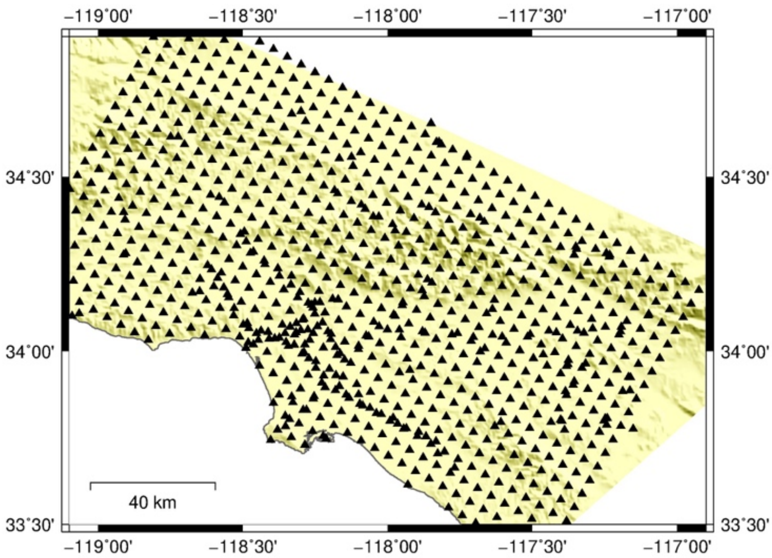

Fig. 5 – Locations of CyberShake ground motion stations considered for spatial correlation estimation. From [26].

Rather than use a variogram, which pools residuals from a range of conditions, in this study the correlation coefficients were computed for each station pair directly. That is,

$$\hat{\rho}(j,k) = \frac{\sum_{i=1}^{n} z_{i,j} z_{i,k}}{\sqrt[2]{\sum_{i=1}^{n} {z_{i,j}}^2}\sqrt[2]{\sum_{i=1}^{n} {z_{i,k}}^2}} \tag{7}$$

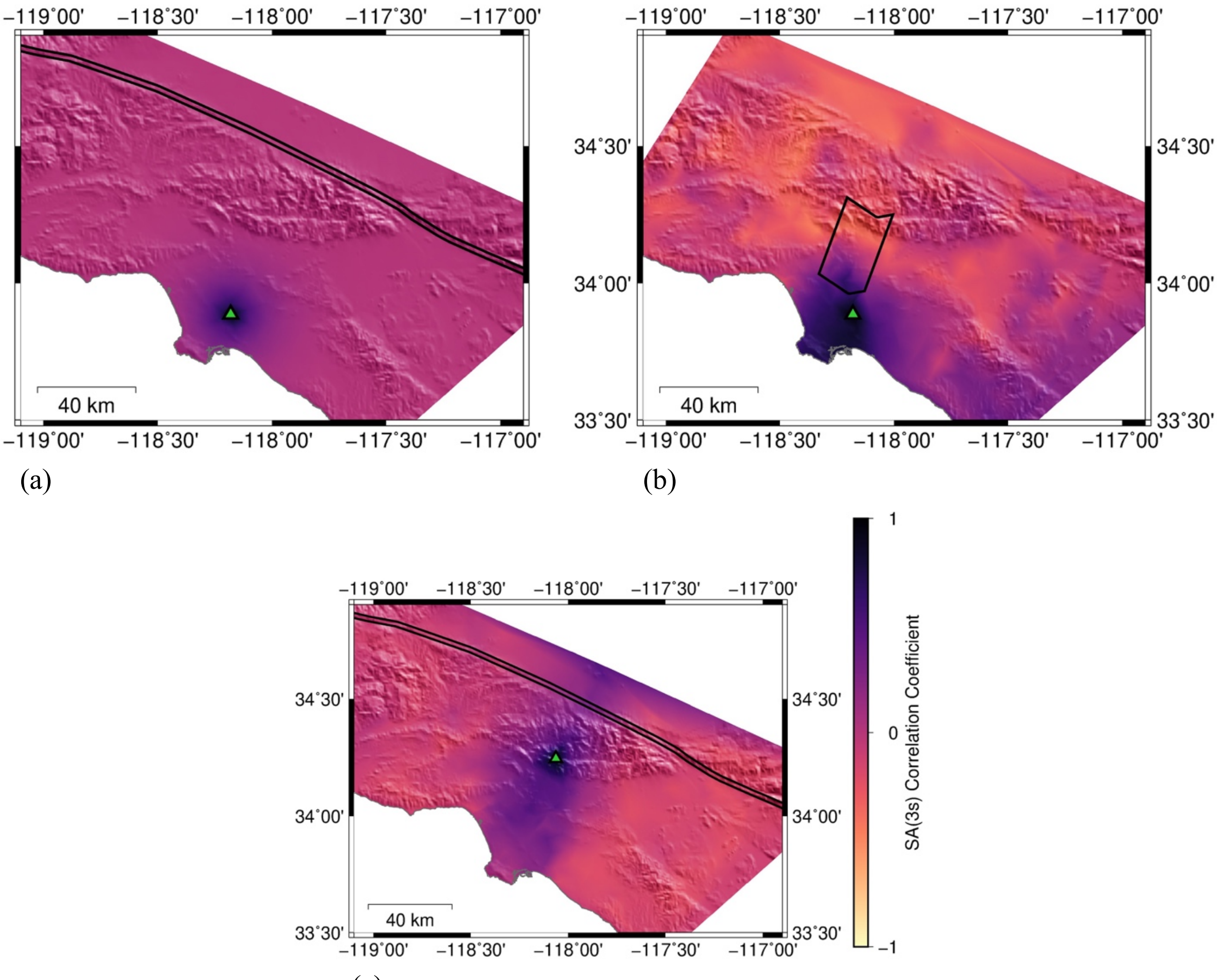

Fig. 6 – Correlations of *SA*(3s) within-event residuals with the *SA*(3s) within-event residual at a reference site (shown with a triangle). (a) Reference model from recorded data [9]. (b) Correlations from CyberShake simulations of a Puente Hills rupture. (c) Correlations from CyberShake simulations of a San Andreas rupture. The surface projection of the reference rupture is shown in black.

From these data, the spatial correlations of spectral accelerations in the region were found to be generally consistent with correlations seen from empirical data—a helpful indication that the data are relevant for this exercise. Further, the spatial correlations show patterns related to regional geology and fault geometry. Fig. 6 illustrates the type of data available from this study. These maps plot correlations of within-event residuals, with respect to the residual at some reference site. Fig. 6a shows correlations predicted by an empirical model [9]. Because the predicted correlation is only a function of distance, the correlations decay uniformly away from the reference site. Fig. 6b and 6c show correlations estimated from CyberShake simulations. These correlations have more complex patterns, with apparent effects that relate to the joint orientation of station pairs relative to the rupture location, and some possible effect of a deep sedimentary basin located at the bottom-center of the study area.

Further work by the authors is ongoing at present to more systematically identify and predict features that cause variations in correlations, and to detect whether the same variations appear to be present in recorded ground motion data.

5 Conclusions

Spatial correlations in ground motion intensity are a topic studied in several dozen publications over the past 20 years. This paper has discussed the relevance of these correlations for risk assessment of distributed systems, and described some of the basics of correlation model fitting. The use of these models in conjunction with empirical Ground Motion Models was also discussed. Despite great advances in our understanding of this problem, opportunities for further refinements remain. This manuscript discussed two such opportunities.

The first topic of discussion was the prediction of multiple *IM*s with spatial as well as cross-*IM* correlation. At present, three algorithms have been proposed for estimating these correlations and then performing forward predictions. The recently proposed approach using Principal Component Analysis was described here. This approach works by first transforming the correlated IMs into principal components, whose spatial correlations can then be modeled independently. This has both model-simplicity and computational benefits relative to alternative approaches.

The second topic of discussion was the use of numerical ground motion simulations to study spatial correlations. This is appealing because it allows the controlled study of factors that may influence spatial correlation—a similar opportunity to how simulations have been used for a number of years to study factors influencing the features of shaking at a single location. This approach depends upon the features causing ground motion spatial correlation to be appropriately characterized in the simulation algorithm. However, if that is the case then the simulations can allow much more refined study than empirical ground motion data allow.

As we enter an era of dense seismic instrumentation, and an era with significant effort being devoted to refined ground motion simulation, the opportunities for refined understanding of spatial correlation appear to be substantial. The coming years offer good prospects for refining spatial correlation models to better reflect causal factors, and for incorporating such models into assessments of regional seismic risk.

6 Acknowledgements

The research described in this paper was supported in part by the Southern California Earthquake Center (SCEC; Contribution Number 10023). The SCEC is funded by National Science Foundation (NSF) Cooperative Agreement EAR-1033462 and U.S. Geological Survey (USGS) Cooperative Agreement G12AC20038. The work was also supported in part by the U.S. Geological Survey (USGS) via External Research Program awards G10AP00046 and G20AP00019. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the sponsors.

7 References

[1] Park J, Bazzurro P, Baker JW. Modeling spatial correlation of ground motion intensity measures for regional seismic hazard and portfolio loss estimation. 10th Int. Conf. Appl. Stat. Probab. Civ. Eng. ICASP10, Tokyo, Japan: 2007, p. 8.

[2] Wesson RL, Perkins DM. Spatial Correlation of Probabilistic Earthquake Ground Motion and Loss. Bull Seismol Soc Am 2001;91:1498–515.

[3] Lee R, Kiremidjian AS. Uncertainty and correlation for loss assessment of spatially distributed systems. Earthq Spectra 2007;23:753–770.

[4] Adachi T, Ellingwood B. Impact of infrastructure interdependency and spatial correlation of seismic intensities on performance assessment of a water distribution system. Proc. 10th Int. Conf. Appl. Stat. Probab. Civ. Eng., 2007.

[5] Shiraki N, Shinozuka M, Moore JE, Chang SE, Kameda H, Tanaka S. System risk curves: Probabilistic performance scenarios for highway networks subject to earthquake damage. J Infrastruct Syst 2007;13:43–54.

[6] Jayaram N, Baker JW. Efficient sampling and data reduction techniques for probabilistic seismic lifeline risk assessment. Earthq Eng Struct Dyn 2010;39:1109–31. https://doi.org/10.1002/eqe.988.

[7] Boore DM, Gibbs JF, Joyner WB, Tinsley JC, Ponti DJ. Estimated Ground Motion From the 1994 Northridge, California, Earthquake at the Site of the Interstate 10 and La Cienega Boulevard Bridge Collapse, West Los Angeles, California. Bull Seismol Soc Am 2003;93:2737–51.

[8] Wang M, Takada T. Macrospatial Correlation Model of Seismic Ground Motions. Earthq Spectra 2005;21:1137–56.

[9] Jayaram N, Baker JW. Correlation model for spatially distributed ground-motion intensities. Earthq Eng Struct Dyn 2009;38:1687–708. https://doi.org/10.1002/eqe.922.

[10] Foulser-Piggott R, Stafford PJ. A predictive model for Arias intensity at multiple sites and consideration of spatial correlations. Earthq Eng Struct Dyn 2012;41:431–51. https://doi.org/10.1002/eqe.1137.

[11] Goda K. Interevent Variability of Spatial Correlation of Peak Ground Motions and Response Spectra. Bull Seismol Soc Am 2011;101:2522–31. https://doi.org/10.1785/0120110092.

[12] Goda K, Atkinson GM. Intraevent Spatial Correlation of Ground-Motion Parameters Using SK-net Data. Bull Seismol Soc Am 2010;100:3055–67. https://doi.org/10.1785/0120100031.

[13] Esposito S, Iervolino I. PGA and PGV Spatial Correlation Models Based on European Multievent Datasets. Bull Seismol Soc Am 2011;101:2532–41. https://doi.org/10.1785/0120110117.

[14] Sokolov V, Wenzel F. Further analysis of the influence of site conditions and earthquake magnitude on ground-motion within-earthquake correlation: analysis of PGA and PGV data from the K-NET and the KiK-net (Japan) networks. Bull Earthq Eng n.d.:1–18. https://doi.org/10.1007/s10518-013-9493-9.

[15] Goda K, Hong HP. Spatial Correlation of Peak Ground Motions and Response Spectra. Bull Seismol Soc Am 2008;98:354–65. https://doi.org/10.1785/0120070078.

[16] Heresi P, Miranda E. Uncertainty in intraevent spatial correlation of elastic pseudo-acceleration spectral ordinates. Bull Earthq Eng 2019;17:1099–115. https://doi.org/10.1007/s10518-018-0506-6.

[17] Gregor N, Abrahamson NA, Atkinson GM, Boore DM, Bozorgnia Y, Campbell KW, et al. Comparison of NGA-West2 GMPEs. Earthq Spectra 2014;30:1179–97. https://doi.org/10.1193/070113EQS186M.

[18] Chiou BS-J, Youngs RR. Update of the Chiou and Youngs NGA Model for the Average Horizontal Component of Peak Ground Motion and Response Spectra. Earthq Spectra 2014;30:1117–53. https://doi.org/10.1193/072813EQS219M.

[19] Journel AG, Huijbregts CJ. Mining geostatistics. Academic press; 1978.

[20] Ancheta TD, Darragh RB, Stewart JP, Seyhan E, Silva WJ, Chiou BS-J, et al. NGA-West2 Database. Earthq Spectra 2014;30:989–1005. https://doi.org/10.1193/070913EQS197M.

[21] Baker JW, Jayaram N. Correlation of spectral acceleration values from NGA ground motion models. Earthq Spectra 2008;24:299–317. https://doi.org/10.1193/1.2857544.

[22] Loth C, Baker JW. A spatial cross-correlation model for ground motion spectral accelerations at multiple periods. Earthq Eng Struct Dyn 2013;42:397–417. https://doi.org/DOI: 10.1002/eqe.2212.

[23] Markhvida M, Ceferino L, Baker JW. Modeling spatially correlated spectral accelerations at multiple periods using principal component analysis and geostatistics. Earthq Eng Struct Dyn 2018;47:1107–23. https://doi.org/10.1002/eqe.3007.

[24] Jackson JE. A user's guide to principal components. vol. 587. John Wiley & Sons; 2005.

[25] Infantino M, Paolucci R, Smerzini C, Stupazzini M. Study of the spatial correlation of earthquake ground motion by means of physics-based numerical scenarios. 16th Eur. Conf. Earthq. Eng., 2018, p. 1–12.

[26] Chen Y, Baker JW. Spatial Correlations in CyberShake Physics-Based Ground-Motion Simulations. Bull Seismol Soc Am 2019;109:2447–58. https://doi.org/10.1785/0120190065.

[27] Graves R, Jordan TH, Callaghan S, Deelman E, Field E, Juve G, et al. CyberShake: A physics-based seismic hazard model for southern California. Pure Appl Geophys 2011;168:367–381.