Safety, Reliability, Risk, Resilience and Sustainability of Structures and Infrastructure
12th Int. Conf. on Structural Safety and Reliability, Vienna, Austria, 6–10 August 2017
Christian Bucher, Bruce R. Ellingwood, Dan M. Frangopol (Editors)

Effect of ground motion correlation on regional seismic loss estimation: application to Lima, Peru using a cross-correlated principal component analysis model

Maryia Markhvida[a], Luis Ceferino[a], and Jack Baker[a]
[a]Department of Civil and Environmental Engineering, Stanford University

Abstract: This paper investigates the effects of spatially cross-correlated ground motions on regional loss estimation. Unlike risk assessments at a single site, regional seismic risk analyses must consider spatial correlation that is exhibited in ground motions. In addition, when considering a region with heterogeneous building typologies that use fragility functions defined for spectral accelerations at differing periods, cross-correlation must be accounted for. In order to capture spatial cross-correlation in ground motions, a computationally efficient simulation model that uses principal component analysis is introduced. Application of five ground motion correlation models, ranging from no correlation to full spatial cross-correlation, considering a $M_w = 8.8$ earthquake scenario in Lima, Peru, is then illustrated. Loss estimation for all of Lima is performed using different correlation models and a regional exposure model consisting of 36 building classes and associated fragility functions. Results from the five models provide insight into how different ground motion correlation models effect the distribution of aggregate regional losses and losses for building classes of varying vulnerability.

1 Introduction

Regional seismic risk assessments reveal regional vulnerabilities and can help inform policies related to risk reduction and management. One of the challenges in conducting risk assessments is the quantification of large uncertainties associated with earthquake events and their consequences. The uncertainties can exist in every step of regional assessment: hazard analysis, building distribution, structural response and damage of building, and associated losses.

When simulating earthquake events, uncertainties in ground motion associated with different sites and different earthquake events are represented by within-event and between-event residuals, respectively. A general formulation of a ground motion model (GMM) that predicts an intensity measure (IM) from ground shaking is the following:

$$\ln IM_{k,j} = \mu_{lnIM} + \delta B_k + \delta W_{k,j} \tag{1}$$

where $\ln IM_{k,j}$ is the logarithm of the intensity measure of interest, μ_{lnIM} is the predicted mean of the log IM; δB_k is the between-event (inter-event) residual for earthquake k with a mean of 0 and standard deviation denoted by τ; and $\delta W_{k,j}$ is the within-event (intra-event) residual for site j and earthquake k with a mean of 0 and standard deviation denoted by ϕ.

Studies have shown that both between-event and within-event residuals are correlated [3, 10, 14, 8]. Between-event residuals exhibit correlation between intensity measures, also known as cross-correlation [10]. Within-event residuals, in addition to being cross-correlated, are correlated spatially [10]. Several models exist for single intensity measure spatial correlation [e.g., 25, 14, 11], and a few models have been proposed for spatial cross-correlation of within-event residuals for multiple intensity measures [10, 17, 18].

Simultaneous consideration of multiple intensity measures becomes significant on a regional level, when damages to different types of structures are best quantified using spectral acceleration at different periods. Therefore, cross-correlation models should be used for regional risk assessments to better represent the extent of damages and losses. Consideration of ground motion uncertainties has been shown to have significant impact on loss estimation in a portfolio of assets, where underestimating spatial correlation tends to overestimate more frequent losses and underestimate more rare ones [21]. A study of the effect of cross-correlation of within-event residuals on portfolio losses showed that cross-correlation has a great impact on the extreme losses [26].

This paper explores the effect of different ground motion correlation models on a large regional risk assessment. Five models for uncertainty and correlation of within-event residuals are considered, ordered from most approximate to most precise: (1) median spectral acceleration (no uncertainty on the ground motion), (2) no correlation (spatial or cross-*IM*), (3) spatial correlation only (no cross-correlation), (4) Markov-type cross-correlation model and (5) full spatial cross-correlation model. The base model used to simulate the full spatial cross-correlation is a geostatistical model that uses Principal Component Analysis (PCA) for simultaneous simulation of spectral accelerations at different periods [18]. The case study presented is a loss estimation for the city of Lima, Peru, for an earthquake rupture of magnitude $M_w = 8.8$.

2 Principal component cross-correlation model

Marginal within-event residuals at a single site can be represented by a normal distribution. Previous studies have also shown that the residuals at different periods and at different sites can be represented by a multivariate normal distribution with a mean vector of zeros [15]. Therefore, the formulation in Equation 2 for simulating normally distributed correlated random variables, $\mathcal{N}(0,\Sigma)$, can be used to simulate cross-correlated residuals at multiple sites and multiple periods. In this equation $\mathbf{x}$ is a q-vector of correlated variables, $\mathbf{b}$ is a q-vector of standard normal deviates, and $\mathbf{T}$ is a $q \times q$ matrix such that $\mathbf{T}^T\mathbf{T} = \Sigma$.

$$\mathbf{x} = \mathbf{T}^T\mathbf{b} \tag{2}$$

In order to build the covariance matrix, Σ, for multiple sites and periods, a spatial cross-correlation model needs to be used. The details on the correlation model used in this study can be found in [18]. This model was built by using within-event residuals from records of 42 earthquakes, based on the NGA-West2 empirical ground-motion database [5]. The residuals at different periods were first linearly transformed to the principal component space using PCA [13], as shown in Equation 3a, where $\mathbf{P}$ is an orthonormal linear transformation matrix; $\mathbf{Z}$ is the matrix of original data where each row represents different observations of variables; and $\mathbf{Y}$ is a matrix of transformed variables whose rows represent uncorrelated principal components.

$$\mathbf{PZ} = \mathbf{Y} \tag{3a}$$

$$\begin{bmatrix} p_{1,T_1} & \cdots & p_{1,T_m} \\ \vdots & \ddots & \vdots \\ p_{m,T_1} & \cdots & p_{m,T_m} \end{bmatrix} \begin{bmatrix} z_{T_1}(x_1) & \cdots & z_{T_1}(x_n) \\ \vdots & \ddots & \vdots \\ z_{T_m}(x_1) & \cdots & z_{T_m}(x_n) \end{bmatrix} = \begin{bmatrix} y_1(x_1) & \cdots & y_1(x_n) \\ \vdots & \ddots & \vdots \\ y_m(x_1) & \cdots & y_m(x_n) \end{bmatrix} \tag{3b}$$

There are two advantages to using PCA when considering cross-correlated residuals at different periods: (1) the transformed residuals, known as principal components, are uncorrelated between components but retain spatial correlation, and (2) since PCA is a dimensionality reduction technique, m periods can be represented by a linear combination of $m' \leq m$ number of variables.

Once the residuals are transformed into the principal component space, semivariograms are used to characterize the spatial variability of each of the principal components. Semivariograms relate the separation distance between two sites (h) with the associated semivariance - a measure of spatial decorrelation or dissimilarity. Under the assumption of second-order stationarity, the semivariance is only a function of the separation distance, which allows for relatively straightforward quantification of the spatial correlation.

The fitted models use nested exponential semivariograms of the form expressed in Equation 4, where $\gamma_i(h)$ is the semivariance of principal component i at separation distance h; $\mathscr{I}_{h=0}$ is the indicator function that evaluates to 1 if $h = 0$, and 0 if $h \neq 0$; and c_{oi}, c_{1i}, c_{2i}, a_{1i}, a_{2i} are fitted model coefficients.

$$\gamma_i(h) = c_{oi}\left(1 - \mathscr{I}_{h=0}\right) + c_{1i}\left(1 - \exp\left(\frac{-3h}{a_{1i}}\right)\right) + c_{2i}\left(1 - \exp\left(\frac{-3h}{a_{2i}}\right)\right) \tag{4}$$

Figure 1 shows the empirical semivariogram and the fitted model for the spatial variability of different principal components.

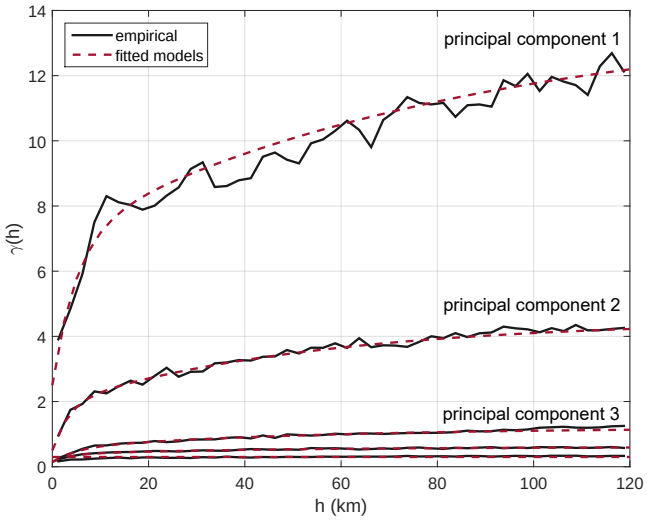


Figure 1: Empirical semivariograms and fitted semivariogram models for the first five principal components (decreasing semivariance with each component). Adapted with permission from [18].

The semivariance can be related to correlation $\rho(h)$, by the relationship shown in 5, where $C_i(0)$ is the covariance of principal component i at separation distance 0, which can be determined

from the semivariogram model.

$$\rho_i(h) = \frac{C_i(0) - \gamma_i(h)}{C_i(0)} \tag{5}$$

By combining Equations 4 and 5, covariance matrices, Σ_i, can be constructed for each of the principal components. Equation 2 can then be used to simulate the principal components. Then the transpose of orthonormal matrix P in Equation 3a is used to calculate the within-event residuals. The resulting residuals are correlated both spatially and across periods. Further details on the procedure and the required model coefficients can be found in [18].

The above model for spatial cross-correlation has two advantages when compared to previously proposed models: it does not require simultaneous fitting of cross-semivariograms, and it is computationally more efficient for simulating residuals at a large number of sites, for two or more periods.

3 Regional loss estimation case study: Lima, Peru

The city of Lima, Peru, was chosen as a case study to evaluate the effects of ground motion correlation on regional losses. Peru is one of the most seismically hazardous countries in South America, which saw four great earthquakes ($M_w \geq 8$) over the last century [4]. The capital city, Lima, possesses great seismic risk, as it is located next to a subduction zone in the coastal region of Lima and houses about 10 million people [12]. In addition, a large portion of the city consists of non-formal and incremental construction, which makes it particularly vulnerable to earthquakes. The following section describes the model used in the loss estimation, where the hazard, exposure and vulnerability models are based on a regional Lima post-earthquake health demand study [6].

3.1 Hazard model

Earthquake scenario: the considered earthquake rupture location and dimensions were chosen to match a M_w ~8.8 earthquake that occurred in the region in 1746 [4, 7, 22].

Ground motion model: a subduction slab GMM was used to predict the median *IM*s and standard deviations for between- and within-event residuals, τ and ϕ [27]. PGA and spectral accelerations at 0.3s and 1s were considered as intensity measures. The near-surface shear wave velocity (V_{s30}) for the region was based on the existing microzonation information [1], and a proxy method using topographic slope was used where such information was unavailable [24].

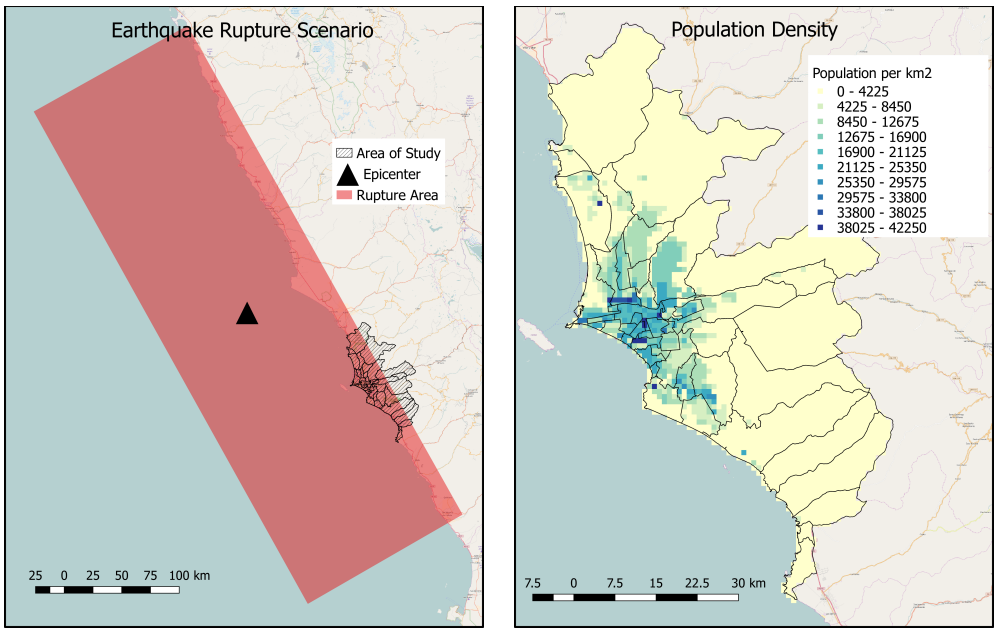
The between- and within-event residuals were simulated 1000 times to account for the uncertainty. The correlation of between-event residuals was modeled as per [10]. Five models for correlation of within-event residuals were considered:

1. Median ground motion for the three intensity measures, with no uncertainty consideration.

2. No correlation. Residuals were simulated as independent random variables for each site and intensity.

3. Spatial correlation based on [18], but no cross-correlation.

4. Simplified cross-correlation model. A Markov screening hypothesis, where the co-located residual for period T_1 screens the effect of other T_1 residuals on the T_2 residual, is used



(a) Earthquake rupture

(b) Spatial distribution of population in Lima (population/km^2)

Figure 2: Earthquake event on the subduction zone off the Coast of Lima and Lima population distribution. Adapted with permission from [6].

[16]. Correlations then have the following formulation [10]:

$$\rho_\varepsilon(h,T_1,T_2) \approx \rho_0(T_1,T_2)\rho_\varepsilon(h,T_{max},T_{max}) \tag{6}$$

where ρ_ε is the spatial cross-correlation of two spectral accelerations for periods T_1 and T_2 separated by a distance h, T_{max} is the larger of the two periods, and ρ_0 represents the overall correlation of the logarithm of the ground motion parameter. ρ_ε is estimated based on [18] and ρ_0 is based on [2].

5. Full spatial cross-correlation model based on [18].

Examples of correlation matrices for within-event residuals (**R**) at two sites and two periods are presented below. The entries are for residuals of PGA_1, PGA_2, $Sa_1(1s)$, $Sa_2(1s)$, where the subscript 1 and 2 denote two sites separated by a distance $h = 53km$, and $Sa(1s)$ denotes spectral acceleration with a period of 1s.

$$\mathbf{R} = \begin{matrix}\text{Model (2)}\\ \begin{bmatrix}1.00 & 0.00 & 0.00 & 0.00\\ 0.00 & 1.00 & 0.00 & 0.00\\ 0.00 & 0.00 & 1.00 & 0.00\\ 0.00 & 0.00 & 0.00 & 1.00\end{bmatrix}\end{matrix} \quad \begin{matrix}\text{Model (3)}\\ \begin{bmatrix}1.00 & 0.25 & 0.00 & 0.00\\ 0.25 & 1.00 & 0.00 & 0.00\\ 0.00 & 0.00 & 1.00 & 0.22\\ 0.00 & 0.00 & 0.22 & 1.00\end{bmatrix}\end{matrix}$$

$$\begin{matrix}\text{Model (4)}\\ \begin{bmatrix}1.00 & 0.25 & 0.52 & 0.11\\ 0.25 & 1.00 & 0.11 & 0.52\\ 0.52 & 0.11 & 1.00 & 0.22\\ 0.11 & 0.52 & 0.22 & 1.00\end{bmatrix}\end{matrix} \quad \begin{matrix}\text{Model (5)}\\ \begin{bmatrix}1.00 & 0.25 & 0.55 & 0.15\\ 0.25 & 1.00 & 0.15 & 0.55\\ 0.55 & 0.15 & 1.00 & 0.22\\ 0.15 & 0.55 & 0.22 & 1.00\end{bmatrix}\end{matrix} \tag{7}$$

3.2 Exposure and vulnerability models

Building classes: the exposure model consisted of 36 building classes based on the Global Earthquake Model's (GEM) South America exposure catalog [9]. The building classes in the

catalog are classified based on lateral load resisting system and its material, the ductility level and the range of number of stories.

Building distribution: the city of Lima is divided into 44 districts. For each district, the number of buildings and the building class distribution were obtained from GEM exposure catalog [9]. LandScan population density [20] was then used to obtain the spatial distribution of the number of buildings in each of the building classes at a 1×1 km^2 resolution.

Fragility functions: GEM fragility functions were used to simulate damage states for each of the building classes [23]. Depending of the building class, PGA and spectral accelerations at 0.3s and 1s periods were used as input intensity measures. Five damage states were considered: no damage, slight damage, moderate damage, extensive damage, and collapse.

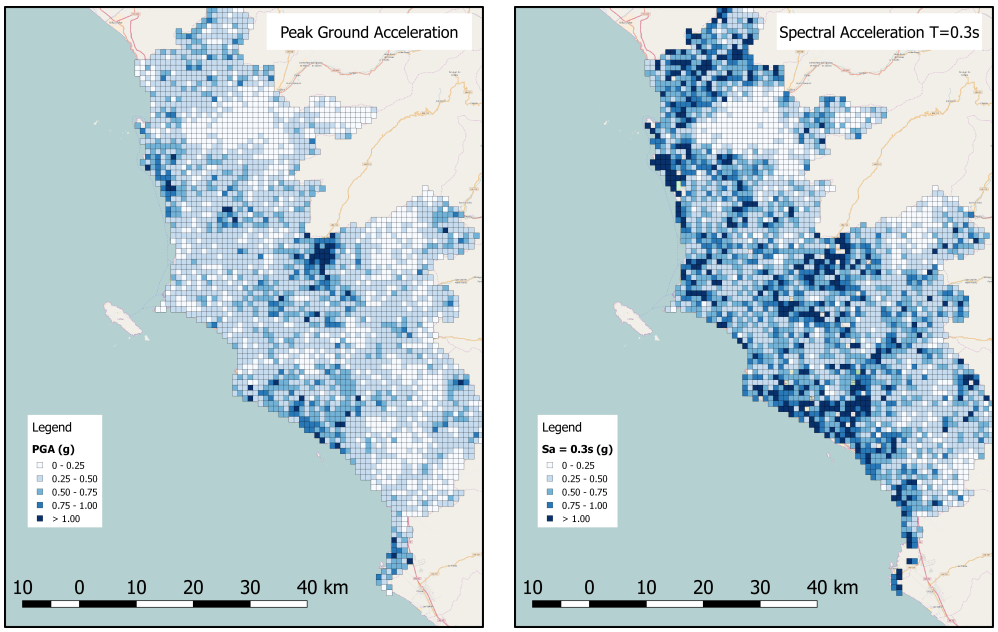
3.3 Loss estimation

Given a damage state, damage-to-loss ratios for each of the damage states were drawn from a truncated normal distribution, with mean values adapted from GEM's South America Risk Assessment project (Vitor Silva, personal communication, 2016) and standard deviations based on a damage-to-loss model for reinforced concrete moment-frames [19]. The replacement cost per building class was calculated based on the average floor area per dwelling and average replacement cost per area [9]. The replacement costs include the cost of the lateral load resisting system and non-structural components, and are expressed in USD.

4 Results

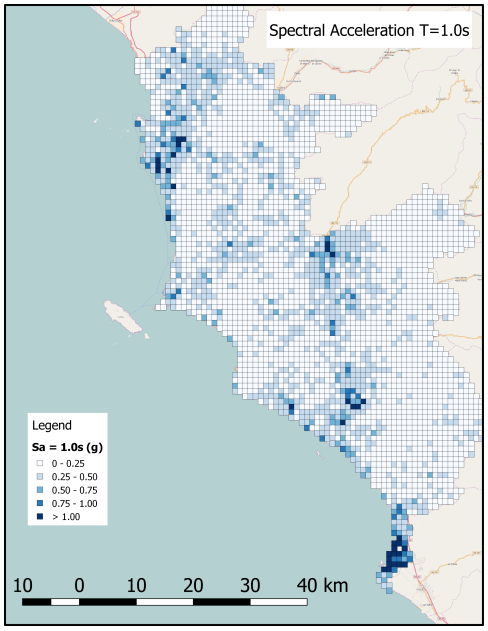
For each of the models the loss results were aggregated in two ways: regionally for the whole city of Lima, and marginally for different building classes. Figure 3 shows sample results for one of the simulation using the full spatial cross-correlation model. Figures 3a, 3b, 3c demonstrate that while accelerations for different periods vary, the spatial correlation and correlation between different intensity measures can clearly be seen, where the patches of similar color (similar acceleration magnitudes) are seen across different periods. The losses for the simulation (Figure 3d) are a result of the simulated ground motions, building distribution, buildings' damage state realization, and the associated losses. It can be seen that the high losses are concentrated in the central part of Lima, where population density is higher.

To evaluate the effect of different ground motion correlation models, the distribution of losses for the city of Lima was assessed based on the 1000 simulations. A summary of the means and standard deviations for different models is presented in Table 1 and the cumulative distribution functions are shown in Figure 4. Adding uncertainty to the ground motions (models 2-5) significantly impacts the loss distribution, with the means increasing by $\sim$40% and the coefficient of variation increasing from negligible to 0.67 (the small variation in the 'No uncertainty' case arises from variability in individual building damage states and repair costs for the specified ground motion). Since positive correlation increases uncertainty of the sum of random variables, the models with increasing correlation consideration showed significantly larger standard deviations for the total loss. Figure 4 also shows that Markov-type approximation works well for regional loss estimation, where the standard deviation of the model was only 10% below the full spatial cross-correlation PCA model.

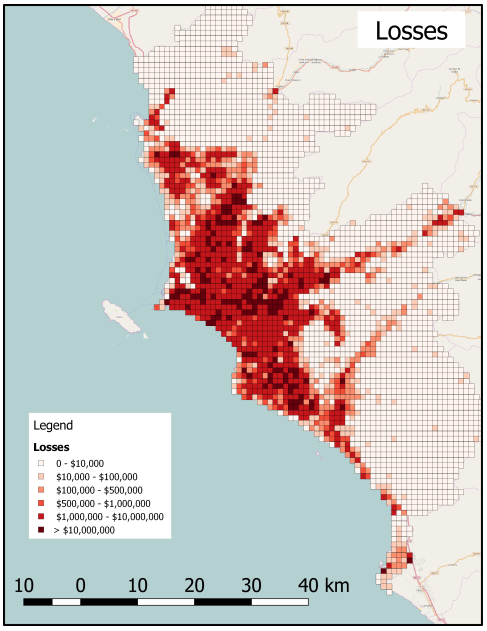


(a) Peak ground acceleration

(b) Spectral acceleration for T = 0.3s



(c) Spectral acceleration for T = 1.0s



(d) Spatial distribution of losses (USD/km^2)

Figure 3: Visualization of results for one of the full cross-correlation model simulations.

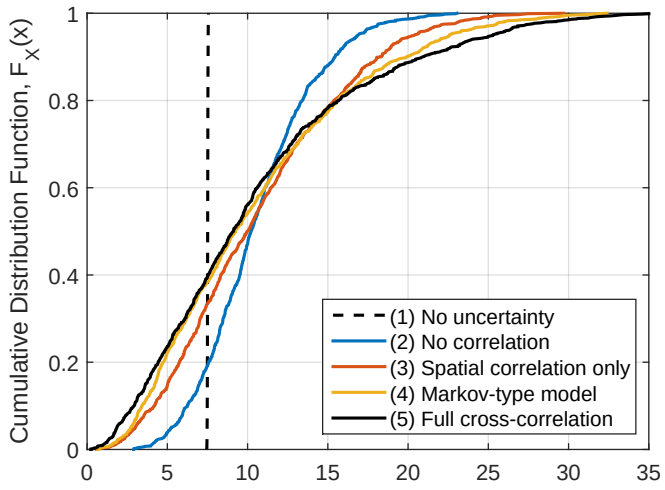
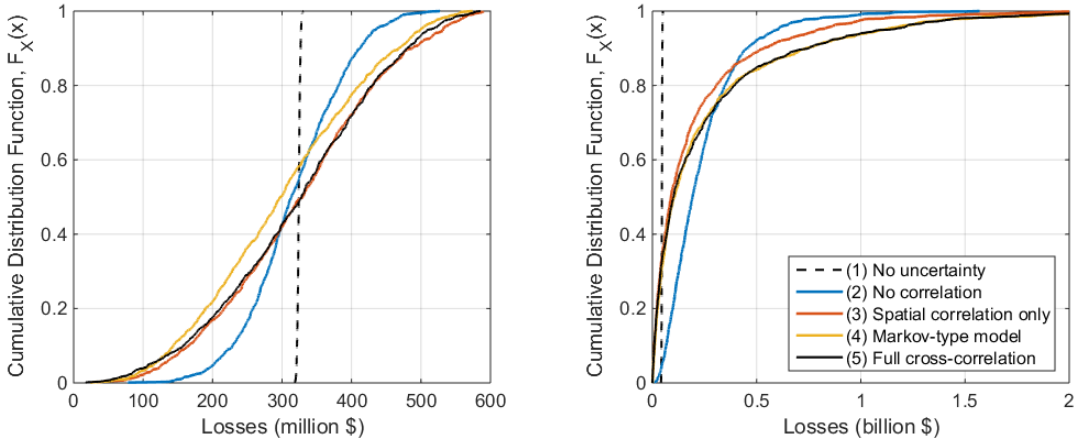


Figure 4: Cumulative distribution function for total Lima losses using the five correlation models.

Table 1: Total Lima losses (billion USD) for different within-event residuals cross-correlation models.

Correlation Model	Mean Loss	Standard Deviation	Coef. of Variation
(1) No uncertainty	7.51	0.02	0.003
(2) No correlation	10.60	3.53	0.33
(3) Spatial correlation only	10.62	5.39	0.51
(4) Markov-type model	10.52	6.31	0.60
(5) Full cross-correlation	10.54	7.02	0.67

In addition, the effect of spatial correlation on individual building classes was assessed. Figure 5a shows losses for a relatively vulnerable building class (three story non-ductile confined masonry), and indicates that uncertainty in the ground motion does not significantly affect the mean losses but does affect the standard deviations. On the other hand, for a less vulnerable building class (one story ductile confined masonry, Figure 5b), spatial correlation increases both the mean losses and standard deviations, and causes a much heavier tail in the distribution of losses. Note that the presence or absence of cross-*IM* correlations does not affect these results for either building class, as single-class loss predictions utilize only a single *IM*.



(a) Losses for a vulnerable building class: three floor non-ductile confined masonry

(b) Losses for a less vulnerable building class: one floor ductile confined masonry

Figure 5: Cumulative distribution function for total losses from different building classes.

5 Conclusions

In this study the effect of differing ground motion correlation models on regional earthquake losses was investigated. A $M_w = 8.8$ subduction zone rupture scenario off of the coast of Lima, Peru, was used as a case study. Lima's building inventory was classified into 36 building classes, whose fragility functions used three different intensity measures (PGA and spectral accelerations at 0.3s and 1s periods) to quantify damage. Loss distributions were calculated using 1000 simulations of median ground motion values and ground motions using four different correlation models with increasing correlation complexity.

It was shown that both spatial correlation and cross-correlation of within-event residuals play significant roles in quantification of regional losses. The mean losses increased on the order of ~40%, when any type of uncertainty in the ground motion was introduced. The uncertainty of the losses also increased as models approached the complexity of a full spatial cross-correlation model. Using the full spatial cross-correlation model yields a 0.67 coefficient of variation on

the total losses in Lima. When conducting a regional seismic risk analysis, Markov-type cross-correlation model can be used to approximate the spatial cross-correlation of the within-event residuals. The model yields similar means, and slightly underestimates the standard deviation. If the tail end of the loss distribution is of particular importance, the PCA spatial cross-correlation model should be used. The effect of spatial correlation on the aggregate losses of different building classes varies depending on the vulnerability of the buildings. It was shown that for vulnerable infrastructure, ground motion uncertainty has a large effect on the standard deviation but not the mean of losses. For less vulnerable infrastructure, correlation causes a substantial increase in both the mean and standard deviation of losses and yields loss distributions with heavier right tails.

Acknowledgment

The authors would like to thank Vitor Silva and Catalina Yepes for their advice during the course of this research project.

References

[1] Z. Aguilar et al. "Actualización de la Microzonificación Sísmica de la cuidad de Lima". In: *The International Symposium for CISMID 25th Anniversary*. CISMID, 2013, pp. 1–5.

[2] J. W. Baker and B. A. Bradley. "Intensity measure correlations observed in the NGA-West2 database, and dependence of correlations on rupture and site parameters". In: *Earthquake Spectra* (in press 2017).

[3] J. W. Baker and C. A. Cornell. "Correlation of Response Spectral Values for Multicomponent Ground Motions". en. In: *Bulletin of the Seismological Society of America* 96.1 (Feb. 2006), pp. 215–227.

[4] S. L. Beck and S. P. Nishenko. "Variations in the mode of great earthquake rupture along the Central Peru Subduction Zone". In: *Geophysical Research Letters* 17.11 (1990), pp. 1969–1972.

[5] K. W. Campbell and Y. Bozorgnia. "NGA-West2 Ground Motion Model for the Average Horizontal Components of PGA, PGV, and 5% Damped Linear Acceleration Response Spectra". In: *Earthquake Spectra* 30.3 (Feb. 2014), pp. 1087–1115.

[6] L. Ceferino, A. Kiremidjian, and G Deierlein. "Regional multi-severity casualty estimation due to building damage following a Mw 8.8 earthquake in Lima, Peru". In: *Earthquake Engineering & Structural Dynamics* In Preparation (2017).

[7] L. Dorbath, A. Cisternas, and C. Dorbath. "Assessment of the size of large and great historical earthquakes in Peru". In: *Bulletin of the Seismological Society of America* 80.3 (1990), pp. 551–576.

[8] S. Esposito and I. Iervolino. "PGA and PGV Spatial Correlation Models Based on European Multievent Datasets". en. In: *Bulletin of the Seismological Society of America* 101.5 (Oct. 2011), pp. 2532–2541.

[9] Global Earthquake Model. *South America Risk Assessment*. 2013.

[10] K. Goda and H. P. Hong. "Spatial correlation of peak ground motions and response spectra". In: *Bulletin of the Seismological Society of America* 98.1 (2008), pp. 354–365.

[11] K. Goda and G. M. Atkinson. "Intraevent Spatial Correlation of Ground-Motion Parameters Using SK-net Data". en. In: *Bulletin of the Seismological Society of America* 100.6 (Dec. 2010), pp. 3055–3067.

[12] Instituto Nacional de Estadistica e Informatica. “Cerca de 10 millones de personas viven en Lima Metropolitana”. In: (Jan. 2016).

[13] J. E. Jackson. *A user’s guide to principal components*. Vol. 587. John Wiley & Sons, 2005.

[14] N. Jayaram and J. W. Baker. “Correlation model for spatially distributed ground-motion intensities”. en. In: *Earthquake Engineering & Structural Dynamics* 38.15 (Dec. 2009), pp. 1687–1708.

[15] N. Jayaram and J. W. Baker. “Statistical Tests of the Joint Distribution of Spectral Acceleration Values”. en. In: *Bulletin of the Seismological Society of America* 98.5 (Oct. 2008), pp. 2231–2243.

[16] A. G. Journel. “Markov Models for Cross-Covariances”. In: *Mathematical Geology* 31.8 (1999), pp. 955–964.

[17] C. Loth and J. W. Baker. “A spatial cross-correlation model of spectral accelerations at multiple periods”. en. In: *Earthquake Engineering & Structural Dynamics* 42.3 (Mar. 2013), pp. 397–417.

[18] M. Markhvida, L. Ceferino, and J. Baker. “Modeling spatially correlated spectral accelerations at multiple periods using principal component analysis and geostatistics”. In: *Earthquake Engineering & Structural Dynamics* In Review (2017).

[19] L. Martins et al. “Development and assessment of damage-to-loss models for moment-frame reinforced concrete buildings”. In: *Earthquake Engineering & Structural Dynamics* (2015).

[20] Oak Ridge National Laboratory. *LandScan global population database 2013*. Minneapolis, 2013.

[21] J. Park, P. Bazzurro, J. Baker, et al. “Modeling spatial correlation of ground motion intensity measures for regional seismic hazard and portfolio loss estimation”. In: *Applications of statistics and probability in civil engineering. Taylor & Francis Group, London* (2007), pp. 1–8.

[22] F. O. Strasser, M. Arango, and J. J. Bommer. “Scaling of the Source Dimensions of Interface and Intraslab Subduction-zone Earthquakes with Moment Magnitude”. In: *Seismological Research Letters* 81.6 (2010), pp. 951–954.

[23] M. Villar-Vega et al. “Development of a fragility model for the residential building stock in South America”. In: *Earthquake Spectra* 33.2 (2017), pp. 581–604.

[24] D. J. Wald and T. I. Allen. “Topographic slope as a proxy for seismic site conditions and amplification”. In: *Bulletin of the Seismological Society of America* 97.5 (2007), pp. 1379–1395.

[25] M. Wang and T. Takada. “Macrospatial Correlation Model of Seismic Ground Motions”. In: *Earthquake Spectra* 21.4 (Nov. 2005), pp. 1137–1156.

[26] G. A. Weatherill et al. “Exploring the impact of spatial correlations and uncertainties for portfolio analysis in probabilistic seismic loss estimation”. en. In: *Bulletin of Earthquake Engineering* 13.4 (Jan. 2015), pp. 957–981.

[27] J. X. Zhao et al. “Attenuation relations of strong ground motion in Japan using site classification based on predominant period”. In: *Bulletin of the Seismological Society of America* 96.3 (2006), pp. 898–913.