

RESEARCH ARTICLE

Modeling spatially correlated spectral accelerations at multiple periods using principal component analysis and geostatistics

Maryia Markhvida  | Luis Ceferino  | Jack W. Baker 

Department of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305-4020, U.S.A.

Correspondence

Maryia Markhvida, Department of Civil and Environmental Engineering, Stanford University, Stanford, CA, 94305-4020, U.S.A.

Email: markhvid@stanford.edu

Funding information

National Science Foundation, Grant/Award Number: CMMI 0952402

Summary

Regional seismic risk assessments and quantification of portfolio losses often require simulation of spatially distributed ground motions at multiple intensity measures. For a given earthquake, distributed ground motions are characterized by spatial correlation and correlation between different intensity measures, known as cross-correlation. This study proposes a new spatial cross-correlation model for within-event spectral acceleration residuals that uses a combination of principal component analysis (PCA) and geostatistics. Records from 45 earthquakes are used to investigate earthquake-to-earthquake trends in application of PCA to spectral acceleration residuals. Based on the findings, PCA is used to determine coefficients that linearly transform cross-correlated residuals to independent principal components. Nested semivariogram models are then fit to empirical semivariograms to quantify the spatial correlation of principal components. The resultant PCA spatial cross-correlation model is shown to be accurate and computationally efficient. A step-by-step procedure and an example are presented to illustrate the use of the predictive model for rapid simulation of spatially cross-correlated spectral accelerations at multiple periods.

KEYWORDS

cross-correlation, principal component analysis, spatial correlation, spectral accelerations, seismic risk

1 | INTRODUCTION

Assessment of earthquake risk at a community level requires proper quantification of spatially distributed ground motions. Empirical ground motion models consider only the marginal probability distribution of ground motion intensity for an earthquake scenario.¹⁻⁵ However, for a risk analysis at a regional level or for a portfolio of assets, one needs to estimate ground motion intensities over the entire area of interest.

In general, ground motion models quantify ground motion at a particular site by predicting the median intensity measure (IM) and adding within- and between-event residuals (see Equation 1 below). Within-event residuals quantify IM variability from location-to-location for one earthquake event, whereas between-event residuals quantify variability from earthquake-to-earthquake that is seen at all locations for a given event. The within-event residuals exhibit spatial correlation and correlation between IMs (ie, cross-correlation), and the between-event residuals exhibit cross-correlation.⁶⁻⁹ A number of studies have proposed models for quantifying the spatial correlation for a single IM.^{7,8,10,11} However, when

considering a regional risk assessment or an assessment of a portfolio comprised of heterogeneous structures, there is a need to use multiple IMs to best quantify earthquake consequences.

A previous study of the effect of spatial cross-correlation on portfolio losses showed that the effect of cross-correlation can be significant, with greater impact on analysis of the extreme losses.¹² In addition, several risk estimation frameworks use multiple IMs to quantify losses: the FEMA P-58 methodology uses spectral acceleration at the fundamental period of the building to predict building response¹³; the Global Earthquake Model has a catalog of fragility curves, comprised of a variety of structural building typologies that use spectral acceleration at differing periods as input¹⁴; and the HAZUS methodology uses peak ground acceleration for buildings and spectral acceleration at period $T = 1.0$ second for bridges to estimate potential damages.¹⁵ Therefore, to conduct a comprehensive regional risk analysis, it is necessary to consider both the spatial dependencies of individual IMs and the spatial dependencies across multiple IMs.

Two approaches have been previously proposed for modeling correlation between multiple sites at different periods: a Markov-type screening hypothesis⁷ and a linear model of coregionalization (LMC).¹⁶ To obtain reliable regional damage or loss estimates, these models need to produce a large number of IM realizations at a large number of locations. Therefore, for regional analyses, these models can be computationally expensive.

This paper proposes a novel methodology for rapid and accurate modeling of spatial cross-correlation of spectral accelerations using principal component analysis (PCA). PCA has two key characteristics. First, it allows one to work in an uncorrelated space by applying an orthogonal transformation to a set of correlated variables. Second, it allows for dimensionality reduction when simulating data that exhibits high correlation among its variables. For these reasons, PCA is a powerful tool for simulating spectral accelerations at different periods, since they exhibit significant correlation across periods. The research described in this paper applies PCA to a large set of recorded ground motions. Based on the results, an accurate and efficient framework for modeling spatially cross-correlated ground motions at multiple sites is proposed.

2 | MODELING SPATIAL VARIABILITY OF CROSS-CORRELATED VARIABLES

This section describes the procedure for calculating ground motion IM residuals, the methodology for characterizing spatial variability of both univariate and multivariate random functions and different techniques to model the random functions.

2.1 | Ground motion residuals

Current ground motion models assess the marginal probability distribution of earthquake IMs at a single site, given an earthquake scenario and site conditions.¹⁻⁵ Equation 1 shows a typical ground motion model format at site j due to an event k :

$$\ln IM_{k,j} = \mu_{\ln IM} + \delta B_k + \delta W_{k,j}, \quad (1)$$

where $\ln IM_{k,j}$ is the logarithm of IM of interest, $\mu_{\ln IM}$ 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 .

In addition to exhibiting spatial cross-correlation, ground motion residuals have several important characteristics. Empirical studies¹⁷ have shown that (1) marginal between- and within-event residuals of spectral accelerations can be represented by a normal distribution, (2) a set of between- and within-event residuals at different periods at one site can be represented as a multivariate normal distribution, and (3) the within-event residuals at different sites can be represented by a bivariate normal distribution. Extending these conclusions, we assume that the within-event residuals of the spectral accelerations across different periods at multiple sites can be approximated by a multivariate normal distribution. In this case, only the covariance matrix is needed to construct the multivariate distribution, since the mean of the within-event residuals is 0.

2.2 | Calculation of within-event residuals

This study used the NGA-West2 empirical ground motion database to select appropriate records.¹⁸ The events were filtered using several criteria. Firstly, only ground motions matching the Chiou and Youngs⁴ selection criteria were chosen, as this

ground motion model was used for the calculation of the residuals. Further, only records having usable residuals for all 19 chosen spectral periods ranging from 0.01 to 5 seconds were considered. Lastly, only earthquakes with more records than the number of periods, were chosen (to be able to perform PCA for each of the earthquake events). In total, the analysis was performed on 45 earthquakes having 4910 associated records. The number of records per earthquake ranged from 19 to 602. The magnitudes ranged from $3.68 \leq \mathbf{M} \leq 7.90$, where 61% of the records came from events of $\mathbf{M} \geq 6.0$.

Within-event residuals were calculated using Chiou and Youngs' ground motion model,⁴ as per Equation 2, where $im_{k,j}$ is the recorded IM from earthquake k at site j , δb_k is the observed between-event residual of earthquake k , and $\delta w_{k,j}$ is the resultant within-event residual. The resultant residuals were further normalized by σ_k , the sample standard deviation of the within-event residuals for earthquake k .

$$z(x) = \frac{\delta w_{k,j}}{\sigma_k} = \frac{\ln(im_{k,j}) - \mu_{\ln IM} - \delta b_k}{\sigma_k} \quad (2)$$

This normalization was performed because the focus of this study is the investigation of spatial cross-correlation of within-event residuals and not the magnitude of variability. For convenience, the normalized residuals for earthquake k and site j will be denoted as $z(x)$, where x is a vector denoting the site's location.

2.3 | Spatial variability of univariate joint distribution

In geostatistics, random variables that are distributed over space and exhibit spatial continuity are represented by a random function, $Z(x)$, where x is the spatial position of the variable. In the case of ground motions, the normalized residuals $z(x)$ for spectral acceleration at a specific natural period of vibration T can be considered samples from a random function $Z(x)$. For a given earthquake, the covariance of a random function $Z(x)$ at 2 different locations x and x' is given by Equation 3:

$$\text{Cov}(Z(x), Z(x')) = \mathbf{E} [(Z(x) - \mathbf{E}[Z(x)]) (Z(x') - \mathbf{E}[Z(x')])], \quad (3)$$

where $\text{Cov}()$ denotes covariance and $\mathbf{E}[]$ denotes expectation. Due to the typical lack of repeated observations of Z at a given location (ie, the lack of multiple ground motions at a given site), the calculation of $\mathbf{E}[Z(x)]$ at different points x in space cannot be done, and therefore, it is common to assume second-order stationarity.¹⁹ Second-order stationarity assumes that the first 2 moments of the joint probability distribution are preserved under spatial translation. Under this assumption the mean $\mathbf{E}[Z(x)]$ and the variance $\text{Var}[Z(x)]$ have constant values μ_Z and σ_Z^2 , respectively, in all the domain, and the covariance will not depend on specific locations x and x' but only on the relative separation distance $h = x - x'$. In the case of ground motions, h is a 2-dimensional vector, since the residuals are observed on a 2-dimensional surface. Past investigations have shown that the covariance between residuals is not dependent on a particular direction, and therefore, the covariance function can be assumed to be isotropic.^{8,20} Consequently, in this study, the covariances are considered independent of the direction, where calculations are based on the magnitude of the separation distance, denoted h .

Using the assumption of stationarity and isotropy, Equation 3 can be rewritten as

$$\text{Cov}(Z(x), Z(x')) = C(h) = \mathbf{E} [(Z(x) - \mu_Z) (Z(x+h) - \mu_Z)]. \quad (4)$$

In geostatistics field, the semivariogram function is used to describe spatial dependence.^{19,21,22} Semivariogram measures the spatial decorrelation or dissimilarity, and it is mathematically defined as in Equation 5. The relationship between the semivariogram and the covariance functions in the case of stationarity is shown in Equation 6.

$$\gamma(h) = \frac{1}{2} \mathbf{E} [(Z(x) - Z(x+h))^2] \quad (5)$$

$$C(h) = C(0) - \gamma(h) \quad (6)$$

The empirical semivariogram can be estimated for a given h from observations using the following expression:

$$\gamma(h) = \frac{1}{2N_h} \sum_{\alpha=1}^{N_h} [z(x_\alpha + h) - z(x_\alpha)]^2, \quad (7)$$

where N_h is the number of observations separated by a distance h . A sample illustration of an empirical semivariogram and a fitted model is shown in Figure 1A. In practice, when the semivariogram is constructed, the set of observations for a given h is chosen with a certain level of tolerance Δ , where the separation distance falls in the interval $[h - \Delta, h + \Delta]$.

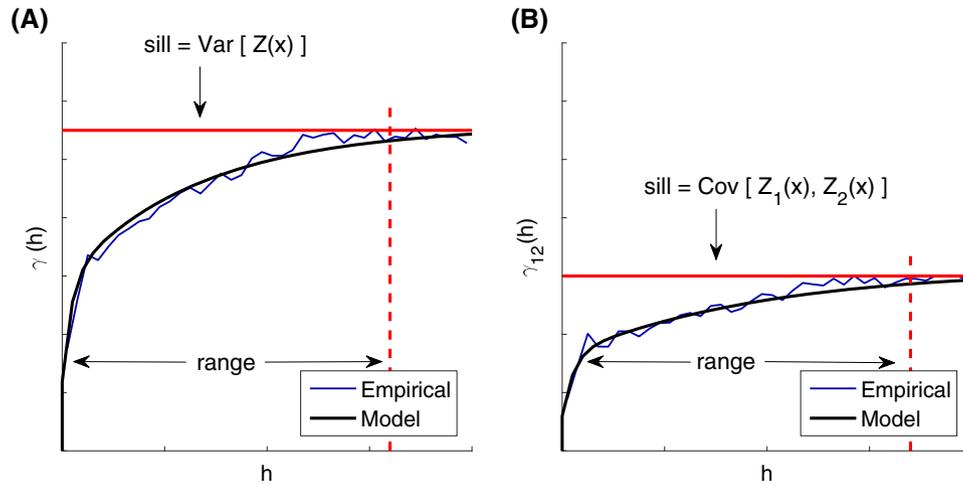


FIGURE 1 Illustration of empirical and fitted models of spatial variabilities for (A) univariate random field represented by semivariogram, and (B) multivariate random field represented by cross-semivariogram

Another tool often used to quantify spatial variability is a correlogram, $\rho(h)$. The correlogram quantifies spatial correlation as a function of separation distance, h , and can be obtained by scaling $C(h)$ in the following manner:

$$\rho(h) = \frac{C(h)}{C(0)}. \quad (8)$$

2.4 | Spatial variability of a multivariate distribution

To model the distribution of several spectral accelerations at different points in space, a vector random function is needed. To obtain this, an extension of the semivariogram for multiple variables, called a cross-semivariogram, is used.

For spectral acceleration at 2 different periods, T_1 and T_2 , the normalized within-event residuals $z_{T_1}(x)$ and $z_{T_2}(x)$ are represented by 2 random functions $Z_{T_1}(x)$ and $Z_{T_2}(x)$, respectively. The cross-semivariogram function is then defined as

$$\gamma_{T_1, T_2}(h) = \frac{1}{2} \mathbf{E} [(Z_{T_1}(x) - Z_{T_1}(x+h)) (Z_{T_2}(x) - Z_{T_2}(x+h))]. \quad (9)$$

The empirical cross-semivariogram can be calculated as

$$\gamma_{T_1, T_2}(h) = \frac{1}{2N_{12,h}} \sum_{\alpha=1}^{N_{12,h}} [(z_{T_1}(x_\alpha) - z_{T_1}(x_\alpha+h)) (z_{T_2}(x_\alpha) - z_{T_2}(x_\alpha+h))], \quad (10)$$

where $N_{12,h}$ is the number of observations of the spectral acceleration at 2 periods T_1 and T_2 separated by a distance h . A sample illustration of an empirical cross-semivariogram and a fitted model is shown in Figure 1B. The collection of cross-semivariograms for different pairs of variables defines the semivariogram matrix as

$$\Gamma(h) = \begin{bmatrix} \gamma_{T_1, T_1}(h) & \dots & \gamma_{T_1, T_m}(h) \\ \vdots & \ddots & \vdots \\ \gamma_{T_m, T_1}(h) & \dots & \gamma_{T_m, T_m}(h) \end{bmatrix}. \quad (11)$$

The semivariogram matrix can be used to construct the covariance matrix; however, to ensure that the covariance matrix is positive definite the cross-semivariograms cannot be fit independently. Several fitting techniques can be used to arrive at simultaneously admissible models,^{23,24} some of which are briefly described in Section 2.6.

2.5 | Univariate spatial variability models

To construct a positive-definite covariance matrix, the empirical semivariograms should be modeled using admissible semivariogram models, which represent different structures of spatial variability. Some admissible models include spherical, cubic, exponential, Gaussian, and nugget.²⁵ Previous studies have shown that ground motion spatial variability is well represented by an exponential semivariogram model with a general formulation^{7,8,10}:

$$\gamma(h) = S \left[1 - \exp \left(\frac{-3h}{R} \right) \right], \quad (12)$$

where the sill (S) represents the semivariance limit or the overall variance, $S = \text{Var}[Z(x)] = \sigma_Z^2$, of the variable. In the case of the exponential semivariogram, this limit is reached asymptotically. For an exponential model, the practical range (R) is the separation distance at which 95% of the sill value is reached.

A second useful semivariogram model is the nugget model (analogous to white noise), which represents discontinuity at short distances and introduces initial variability. The nugget model has the following formulation:

$$\gamma(h) = \begin{cases} 0 & \text{if } h = 0, \\ S & \text{if } h > 0. \end{cases} \quad (13)$$

To represent spatial variability in which multiple structures are present, a linear combination of different semivariogram models can be used.¹⁹ These types of models are referred to as nested semivariogram models.

2.6 | Previously proposed multivariate spatial variability models

This section describes two existing methods that model cross-correlation among spatially distributed spectral accelerations. The PCA methodology for modeling multivariate spatial variability used in this study is described in Section 3.

2.6.1 | Markov-type screening hypothesis model

One way to model spatially cross-correlated spectral accelerations at different periods is by using a Markov-type screening hypothesis, where colocated data $z_1(x)$ at location x is assumed to screen the influence of other data points $z_1(x+h)$ on random secondary variable $Z_2(x)$ ²⁴:

$$\mathbf{E} [Z_2(x)|Z_1(x); Z_1(x+h)] = \mathbf{E} [Z_2(x)|Z_1(x)]. \quad (14)$$

This hypothesis was used in a model proposed by Goda and Hong to model the correlation in spectral accelerations at different periods, denoted by ρ_ϵ , at a distance h .⁷ This model has the following formulation:

$$\rho_\epsilon(h, T_{n1}, T_{n2}) \approx \rho_0(T_{n1}, T_{n2})\rho_\epsilon(h, T_{max}, T_{max}) \quad (15)$$

where T_{max} is the larger of T_{n1} and T_{n2} , ρ_0 represents the correlation coefficient between the spectral accelerations at a given location, and $\rho_\epsilon(h, T_{max}, T_{max})$ is a univariate correlogram for spectral acceleration at T_{max} evaluated at a distance h .

While this model ensures that the resulting covariance matrix is positive definite as long as ρ_0 and ρ_ϵ are admissible, it does not consider the cross-correlation of variables that are not colocated. It has been shown that while this hypothesis is a good approximation for periods that are close to one another, it becomes less accurate when the periods differ substantially.¹⁶

2.6.2 | Linear model of coregionalization

The LMC^{19,23} assumes that the cross-correlated variables are linear combinations of the same underlying structures, so the semivariogram matrix can be formulated as¹⁶

$$\Gamma(h) = \sum_{l=1}^L \mathbf{B}^l g^l(h), \quad (16)$$

where the sum is over L cross-semivariograms, \mathbf{B}^l are the positive definite coregionalization matrices, and $g^l(h)$ are admissible semivariogram functions. Loth and Baker¹⁶ used LMC to propose a model for simulating spectral accelerations at 6 periods: 0.01, 0.1, 0.2, 1, 2, and 5 seconds.

Using the LMC methodology to characterize spatially cross-correlated spectral accelerations at different periods presents some challenges and complexities, especially as one increases the number of periods to be simulated. One of these challenges is to obtain a version of Equation 16 that can accurately represent the large number of cross-semivariograms while keeping the \mathbf{B} matrices positive definite. For m different periods, the set of parameters that define Equation 16 require the fitting of $\frac{m \times (m+1)}{2}$ cross-semivariograms.

The covariance matrix grows quickly with the addition of extra periods for spectral accelerations. If there are n locations and m periods to be simulated, the covariance matrix dimensions become $(nm) \times (nm)$, which can impose computational restrictions for a large number of n or m . This is true for both Markov-type screening and LMC models.

3 | PROPOSED PRINCIPAL COMPONENT MODEL FOR GROUND MOTIONS

This section describes how PCA and semivariograms were used to develop a model for spatial variability of cross-correlated spectral accelerations.

3.1 | Principal component analysis

Principal component analysis performs a linear transformation of the variables of interest to an orthogonal basis, where the resulting projections onto the new basis are uncorrelated and are called principal components. The first principal component has the largest variance, the second principal component has the second largest variance constrained on orthogonality with the previous component, and so on. Principal component analysis yields as many principal components as there are original variables.

There are techniques that can quickly perform PCA and calculate the transformation matrix along with the principal components.²⁶ Equations 17 or 18 define the linear transformation that allows one to pass from the original space (in our case, normalized spectral acceleration residuals at different periods) to a principal component space, where \mathbf{P} is an orthogonal 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 where rows represent different uncorrelated principal components.

$$\mathbf{PZ} = \mathbf{Y} \quad (17)$$

$$\begin{bmatrix} p_{1,T_1} & \dots & p_{1,T_m} \\ \vdots & \ddots & \vdots \\ p_{m,T_1} & \dots & p_{m,T_m} \end{bmatrix} \begin{bmatrix} z_{T_1}(x_1) & \dots & z_{T_1}(x_n) \\ \vdots & \ddots & \vdots \\ z_{T_m}(x_1) & \dots & z_{T_m}(x_n) \end{bmatrix} = \begin{bmatrix} y_1(x_1) & \dots & y_1(x_n) \\ \vdots & \ddots & \vdots \\ y_m(x_1) & \dots & y_m(x_n) \end{bmatrix} \quad (18)$$

Analogously the principal components can be transformed back to the original space as follows:

$$\mathbf{Z} = \mathbf{P}^{-1}\mathbf{Y}. \quad (19)$$

Since \mathbf{P} is an orthogonal matrix, the following equivalence holds:

$$\mathbf{Z} = \mathbf{P}^T\mathbf{Y}. \quad (20)$$

Equation 21 represents the elements of the principal component i . Since principal components are uncorrelated, the spatial variability characterization of Y_i can be done independently of any other component $Y_{i'}$. Only the spatial correlation among the elements of Y_i in Equation 21 should be considered.

$$Y_i = [y_i(x_1) \ y_i(x_2) \ \dots \ y_i(x_n)] \quad (21)$$

The fact that the principal components are uncorrelated eliminates the need for cross-semivariograms and allows one to calculate a semivariogram with Equation 7 for each of the components independently.

By changing the basis, PCA can provide us with the ability to reduce the dimensionality of the problem to a smaller number of independent dimensions. In the case of ground motions, it would mean that the desired number of spectral accelerations at different periods can be simulated with a smaller number of variables. To know how many components we need to simulate to represent most of the variability in the data set, the variance $\text{Var}[Y_i]$ of each component can be calculated. The sum of the variances for all the components adds up to the total variance in the dataset, as shown in Equation 22. In practice, it is common to use enough components such that the variance adds up to anywhere from 70% to 95% of the total variance,^{27,28} as shown in Equation 23 where the 95% threshold was chosen.

$$\text{Var}[\text{Total}] = \sum_{i=1}^m \text{Var}[z_{T_i}(x)] = \sum_{i=1}^m \text{Var}[Y_i] \quad (22)$$

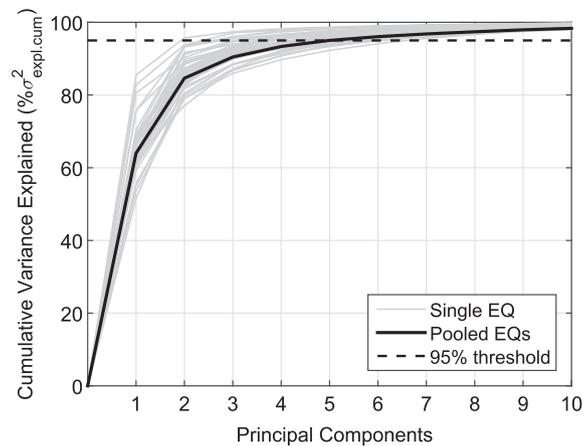


FIGURE 2 Cumulative contribution of each principal component to the explained variance of the variables for individual earthquake events (EQ) and pooled records

$$\% \sigma_{expl.cum}^2 = \frac{\sum_{i=1}^{m_{95}} \text{Var}[Y_i]}{\text{Var}[\text{Total}]} \geq 0.95 \quad (23)$$

3.2 | Principal component coefficients

Principal component analysis was performed on each of the 45 considered earthquakes separately using the approach described in Section 3.1, to obtain principal component coefficient vectors. In addition, PCA was performed on the normalized residuals when the data from the 45 earthquakes was pooled together. Figure 2 shows the cumulative variance explained ($\% \sigma_{expl.cum}^2$) by each of the principal components for single earthquake events, as well as when the records are pooled and analyzed together. When the records are pooled, 5 components explain 95% of the variance. The percent of explained variance ($\% \sigma_{expl}^2$) and cumulative percent of explained variance for each of the principal components is summarized in Table 1. The evident reduction of observed variables to a smaller set of transformed variables that explain most of the variance suggests that different spectral accelerations are measuring the same underlying structures. This implies that spectral accelerations for considered periods can be modeled as a linear combination of a reduced number of variables.

Figure 3 shows the transformation coefficients corresponding to different periods for the 45 single earthquake events. Performing PCA on the 45 events reveals that for each of the principal components, the transformation coefficients follow a similar trend across the periods. This is especially evident in principal components 1 and 2, which capture most of the variables' variance. In addition, the mean of the 45 coefficients for single earthquakes closely resembles the transformation coefficients obtained when all the records are pooled and analyzed together. To create a single model for spatial variability and to ensure that the components remain independent, the coefficients used in the model were obtained by pooling records from the 45 earthquakes.

Table 1 summarizes the transformation coefficients obtained for each of the principal components. In the case that a period not explicitly considered in the model is required, the coefficient can be linearly interpolated between periods for each of the principal components. This methodology was verified by adding an extra period, rerunning PCA and modeling semivariograms in both the principal component and the original spaces. The results of the interpolation and the analysis with an extra period were practically the same and did not have an effect of the modeling parameters.

3.3 | Semivariogram models for principal components

The choice of the semivariogram models is dependent on the shape of empirical semivariograms and previous knowledge about the random function being modeled. Manual visual fits of the empirical semivariograms were performed, following common practice, since the precise fit of the points is less important than the behavior at short separation distances and consideration of what happens at large distances.^{21,29}

The empirical semivariograms for each of the principal components were calculated using lag tolerance, $\Delta = 2.5$ km. The semivariograms were constructed using all of the pooled data, where for each earthquake all combinations of

TABLE 1 Principal components coefficients

Period, s	Y ₁	Y ₂	Y ₃	Y ₄	Y ₅	Y ₆	Y ₇	Y ₈	Y ₉	Y ₁₀	Y ₁₁	Y ₁₂	Y ₁₃	Y ₁₄	Y ₁₅	Y ₁₆	Y ₁₇	Y ₁₈	Y ₁₉
0.01	0.27	-0.14	0.07	-0.11	-0.09	-0.11	-0.19	0.15	-0.16	-0.05	0.11	0.05	-0.08	0.00	0.23	-0.04	-0.30	-0.53	-0.58
0.02	0.27	-0.14	0.08	-0.12	-0.10	-0.12	-0.20	0.16	-0.16	-0.05	0.10	0.05	-0.08	0.01	0.22	-0.04	-0.26	-0.15	0.78
0.03	0.27	-0.15	0.10	-0.14	-0.13	-0.15	-0.22	0.15	-0.14	-0.05	0.09	0.04	-0.06	0.01	0.15	-0.02	-0.03	0.81	-0.23
0.05	0.25	-0.18	0.18	-0.22	-0.18	-0.18	-0.19	0.04	-0.05	-0.03	-0.03	-0.06	0.09	0.02	-0.30	0.06	0.75	-0.21	0.02
0.075	0.24	-0.22	0.24	-0.23	-0.13	-0.04	0.12	-0.27	0.24	0.10	-0.26	-0.12	0.20	0.01	-0.49	0.12	-0.48	0.04	-0.01
0.1	0.23	-0.23	0.23	-0.16	0.04	0.18	0.43	-0.32	0.26	0.14	-0.08	0.05	-0.15	-0.08	0.53	-0.18	0.21	-0.00	0.00
0.15	0.24	-0.21	0.13	0.08	0.33	0.39	0.33	0.16	-0.18	-0.14	0.47	0.18	-0.11	0.09	-0.29	0.26	-0.00	0.02	0.00
0.2	0.25	-0.17	-0.01	0.28	0.40	0.22	-0.08	0.22	-0.17	-0.03	-0.38	-0.24	0.36	-0.09	-0.01	-0.44	0.02	0.01	0.00
0.25	0.25	-0.12	-0.15	0.37	0.25	-0.06	-0.28	-0.08	0.21	0.14	-0.28	-0.04	-0.20	0.02	0.16	0.63	0.05	0.00	0.00
0.3	0.25	-0.07	-0.24	0.36	0.04	-0.25	-0.14	-0.29	0.30	0.06	0.33	0.21	-0.19	0.03	-0.26	-0.48	0.00	0.01	0.00
0.4	0.25	0.01	-0.33	0.23	-0.26	-0.22	0.34	-0.12	-0.06	-0.22	0.21	-0.13	0.58	-0.06	0.20	0.21	0.02	0.00	0.00
0.5	0.25	0.08	-0.36	0.06	-0.34	0.02	0.39	0.18	-0.26	-0.01	-0.38	-0.08	-0.50	0.02	-0.18	-0.07	0.02	0.01	0.00
0.75	0.23	0.19	-0.34	-0.22	-0.17	0.42	-0.14	0.19	0.15	0.53	0.04	0.33	0.27	0.06	0.00	0.01	0.02	0.00	0.00
1	0.21	0.26	-0.24	-0.33	0.08	0.33	-0.22	-0.12	0.27	-0.44	0.15	-0.48	-0.14	-0.04	0.01	-0.02	-0.01	0.00	-0.00
1.5	0.19	0.33	-0.09	-0.27	0.36	-0.15	-0.00	-0.33	-0.27	-0.28	-0.26	0.53	0.07	-0.08	-0.03	0.03	0.01	0.00	-0.00
2	0.18	0.36	0.06	-0.16	0.35	-0.34	0.16	-0.03	-0.21	0.51	0.21	-0.41	-0.04	0.17	-0.00	-0.01	-0.00	0.00	0.00
3	0.17	0.36	0.26	0.07	0.06	-0.22	0.18	0.52	0.46	-0.10	-0.02	0.12	-0.00	-0.42	-0.04	0.02	-0.01	-0.01	0.00
4	0.16	0.35	0.35	0.24	-0.16	0.09	-0.01	0.02	0.11	-0.18	-0.12	0.07	0.06	0.75	0.08	-0.05	0.01	-0.00	-0.00
5	0.15	0.33	0.37	0.33	-0.28	0.28	-0.18	-0.33	-0.31	0.13	0.08	-0.07	-0.05	-0.44	-0.04	0.03	0.00	0.00	0.00
% σ^2_{expl}	0.64	0.21	0.06	0.03	0.02	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
% $\sigma^2_{expl,cum}$	0.64	0.85	0.90	0.93	0.95	0.96	0.97	0.97	0.98	0.98	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00

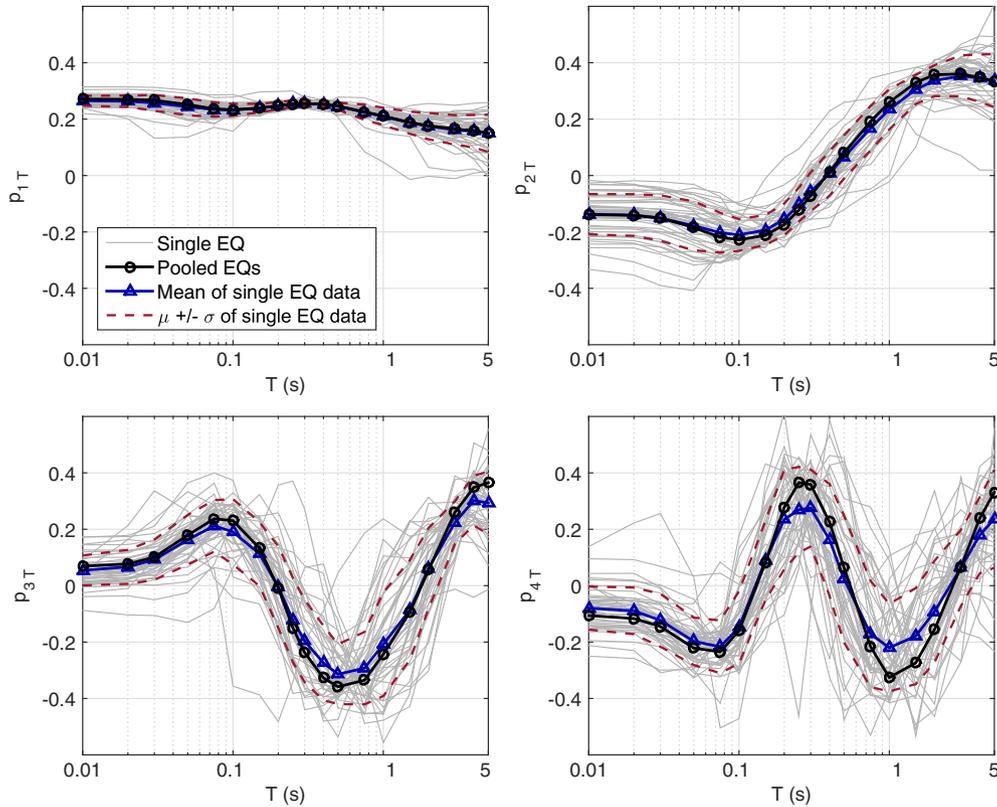


FIGURE 3 Coefficients corresponding to different periods, T , for the first 4 principal components of normalized ground motion residuals, as per Equation 18

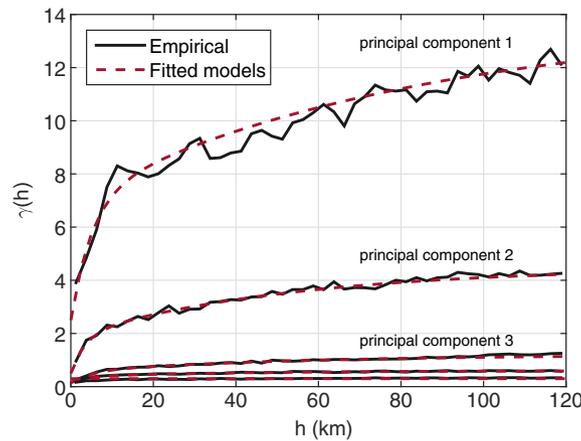


FIGURE 4 Empirical semivariograms and fitted semivariogram models for the first 5 principal components (PCs) (decreasing with each component)

observation pairs were considered. In total, 616219 pairs of observations were used. The shape of the resulting empirical semivariograms (Figure 4) shows that only the first 4 components exhibit decaying spatial correlation, where the semivariance increases with separation distance h . The rest of the components have a more constant semivariance at all separation distances, which implies no spatial correlation, and therefore, the use of a nugget model is appropriate.

It was determined that the nested model of Equation 24 best represents the empirical semivariograms. This form captures three types of behavior: the nugget which models the initial discontinuity, short-range behavior, and long-range behavior.

$$\gamma_i(h) = c_{oi} (1 - \mathcal{I}_{h=0}) + c_{1i} \left(1 - \exp \left(\frac{-3h}{a_{1i}} \right) \right) + c_{2i} \left(1 - \exp \left(\frac{-3h}{a_{2i}} \right) \right) \quad (24)$$

TABLE 2 Coefficients for Equation 24, for each principal component

Principal Component	c_{oi}	c_{1i}	a_{1i}	c_{2i}	a_{2i}
Y_1	2.50	4.52	15	6.78	250
Y_2	0.50	1.40	10	2.60	160
Y_3	0.15	0.42	15	0.63	160
Y_4	0.15	0.23	10	0.23	120
Y_5	0.31	0	...	0	...
Y_6	0.19	0	...	0	...
Y_7	0.14	0	...	0	...
Y_8	0.11	0	...	0	...
Y_9	0.10	0	...	0	...
Y_{10}	0.07	0	...	0	...
Y_{11}	0.06	0	...	0	...
Y_{12}	0.05	0	...	0	...
Y_{13}	0.05	0	...	0	...
Y_{14}	0.04	0	...	0	...
Y_{15}	0.04	0	...	0	...
Y_{16}	0.03	0	...	0	...
Y_{17}	0.02	0	...	0	...
Y_{18}	0	0	...	0	...
Y_{19}	0	0	...	0	...

In Equation 24, $\mathcal{I}_{h=0}$ is the indicator function that evaluates to 1 if $h = 0$, and 0 if $h \neq 0$. Table 2 provides the coefficient values for Equation 24 for each principal component; the resulting semivariogram models are shown in Figure 4. While the first 4 components are fit using the full nested semivariogram model, components 5 to 19 are modeled as pure nugget, where the exponential sills c_1 and c_2 become 0, and range parameters a_1 and a_2 are irrelevant. It can be seen from Tables 1 and 2 that principal components 10 to 19 have negligible contribution to the variance (ie, < 1%) and therefore do not have to be considered in the analysis and modeling of the variables.

3.4 | Covariance models for principal components

It is of interest to find the expression for covariance as a function of separation distance h for each of the components, as it will be required for simulation of ground motion residuals. Recalling Equation 6, to obtain $C(h)$, one must find $C(0)$, which can be found by

$$C(0) = \lim_{h \rightarrow \infty} \gamma(h) = c_o + c_1 + c_2. \quad (25)$$

The resulting expression for covariance, $C_i(h)$, at a separation distance h is the following:

$$C_i(h) = c_{io} \mathcal{I}_{h=0} + c_{i1} \exp\left(\frac{-3h}{a_{1i}}\right) + c_{i2} \exp\left(\frac{-3h}{a_{2i}}\right), \quad (26)$$

where all coefficients are available from Table 2.

4 | RESULTING SPATIAL VARIABILITY AND CROSS-CORRELATION OF THE MODEL

To obtain the resulting semivariogram models in the original space and to verify that the proposed model is accurate, one must use a combination of the PCA coefficients and principal component models. The following set of expressions originating from Equations 6 and 19 are used to derive semivariograms and cross-semivariograms for the original variables:

$$\gamma_{T_1, T_2}(h) = C(0) - C(h) = \text{Cov}\left(\sum_{i=1}^m p_{i, T_1} Y_i(u), \sum_{j=1}^m p_{j, T_2} Y_j(u)\right) - \text{Cov}\left(\sum_{i=1}^m p_{i, T_1} Y_i(u), \sum_{j=1}^m p_{j, T_2} Y_j(u+h)\right). \quad (27)$$

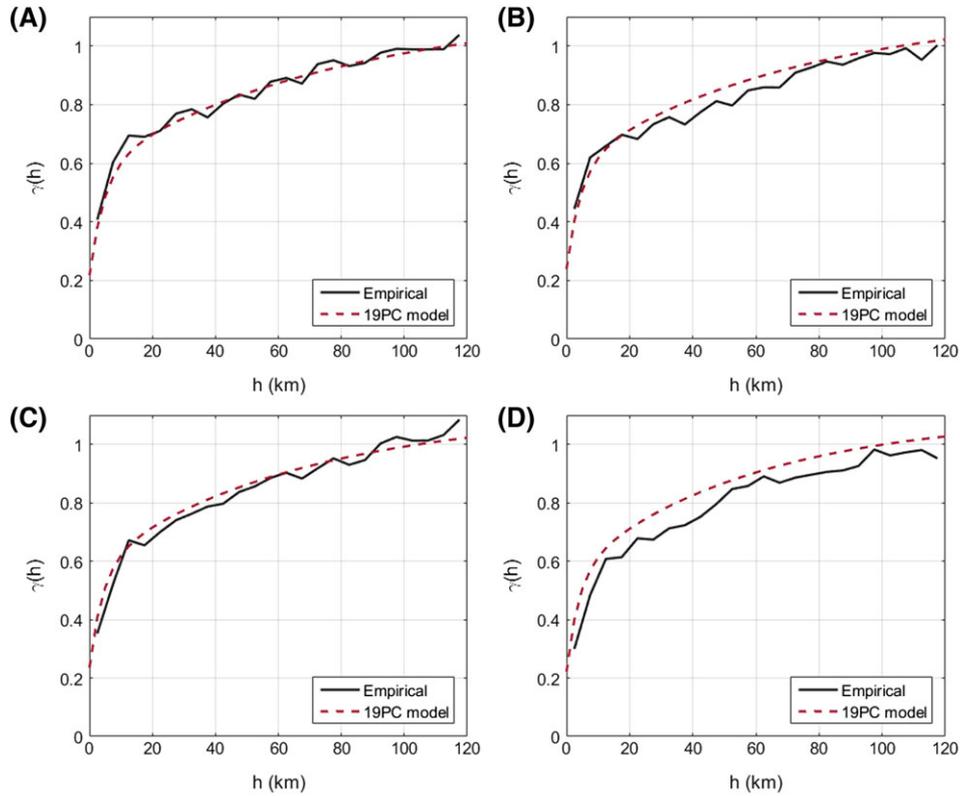


FIGURE 5 Comparison of empirical semivariogram and model semivariogram with 19 principal components (19PC model) in the original space using Equation 28 for (A), $T = 0.01$ s, (B), $T = 0.1$ s, (C), $T = 1$ s, (D), $T = 5$ s

Expanding the terms in Equation 27, and noting that the covariances between different principal components are 0, the resulting expression becomes

$$\begin{aligned}
 \gamma_{T_1, T_2}(h) &= \sum_{i=1}^m p_{i, T_1} p_{i, T_2} \text{Cov}(Y_i(u), Y_i(u)) - \sum_{i=1}^m p_{i, T_1} p_{i, T_2} \text{Cov}(Y_i(u), Y_i(u+h)) \\
 &= \sum_{i=1}^m p_{i, T_1} p_{i, T_2} (C_i(0) - C_i(h)) \\
 &= \sum_{i=1}^m p_{i, T_1} p_{i, T_2} \gamma_i(h).
 \end{aligned} \tag{28}$$

Figure 5 provides comparisons of empirical and model semivariograms for periods $T = 0.01, 0.1, 1,$ and 5 s. It can be seen that the original spatial variability characteristics are preserved for different periods when modeling in the PCA space. Figure 6 shows similar results for cross-semivariograms for differing pairs of periods, signifying that in addition to the spatial correlation, correlation between periods is preserved when using the PCA model.

4.1 | Reduced number of principal components

It is of interest to investigate the minimum number of principal components needed to capture sufficient variance. Section 3.3 shows that while the first few principal components have a large contribution to variance, the later components are modeled as noise and contribute a relatively small amount to the total variance. Figure 7A shows an example of a resulting semivariogram for $T = 0.1$ s when considering only 1, 3, and 5 principal components in the model.

While 5 principal components provide sufficient variance for some periods not shown in the figure, for $T = 0.1$ s, it can be seen that some variance is not captured. One way to mitigate this is to scale the model semivariograms provided in Equation 24 by the cumulative percent of explained variance, $\% \sigma_{expl.cum}^2$ (Table 1).

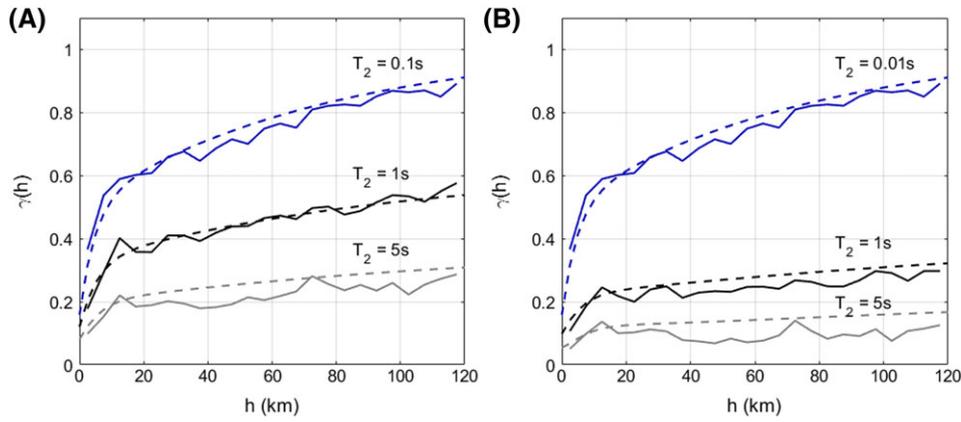


FIGURE 6 Comparison of empirical (solid lines) and model (dashed lines) cross-semivariograms in original space with T_2 periods for (A), $T_1 = 0.01$ s and (B), $T_1 = 0.1$ s

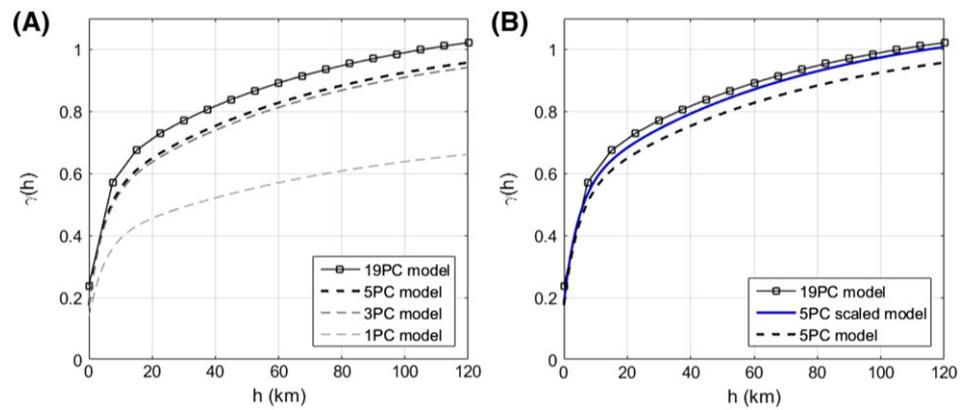


FIGURE 7 Comparison of the full 19 principal component (PC) model with contribution from (A), differing numbers of unscaled components and (B), unscaled and scaled 5 component models for $T = 0.1$ s

The results for the scaled 5 principal component model for $T = 0.1$ s are shown in Figure 7B, and it can be seen that scaling is an effective method to correct for the missing variance. While in some cases, scaling slightly overestimates cross-semivariances, on average, the method works well across all periods.

In an effort to reduce the computational time by not simulating all 19 principal components, it is recommended that the scaled model with 5 principal components be used. The resulting covariance expression for the scaled model for a given principal component is the following:

$$C_i(h) = \frac{c_{i0} \mathcal{I}_{h=0} + c_{i1} \exp\left(\frac{-3h}{a_{i1}}\right) + c_{i2} \exp\left(\frac{-3h}{a_{i2}}\right)}{\% \sigma_{expl.cum}^2}, \quad (29)$$

where $\% \sigma_{expl.cum}^2$ should be corresponding to the number of principal components chosen.

5 | COMPARISON WITH PREVIOUS RESEARCH

5.1 | Spatial variability

The proposed spatial variability model provides an alternative to the existing models discussed in Section 2.6. The LMC model¹⁶ is based on 8 larger earthquakes ($M > 6.0$) whose records were taken from the PEER Center's NGA database.³⁰ The records used in the LMC model constitute 42% of the records used in this study. Figure 8 shows the correlograms and cross-correlograms, calculated using Equation 8, for both models and empirical data. It can be seen that for a given period, the correlations decay faster in the LMC model than they do in the PCA model.

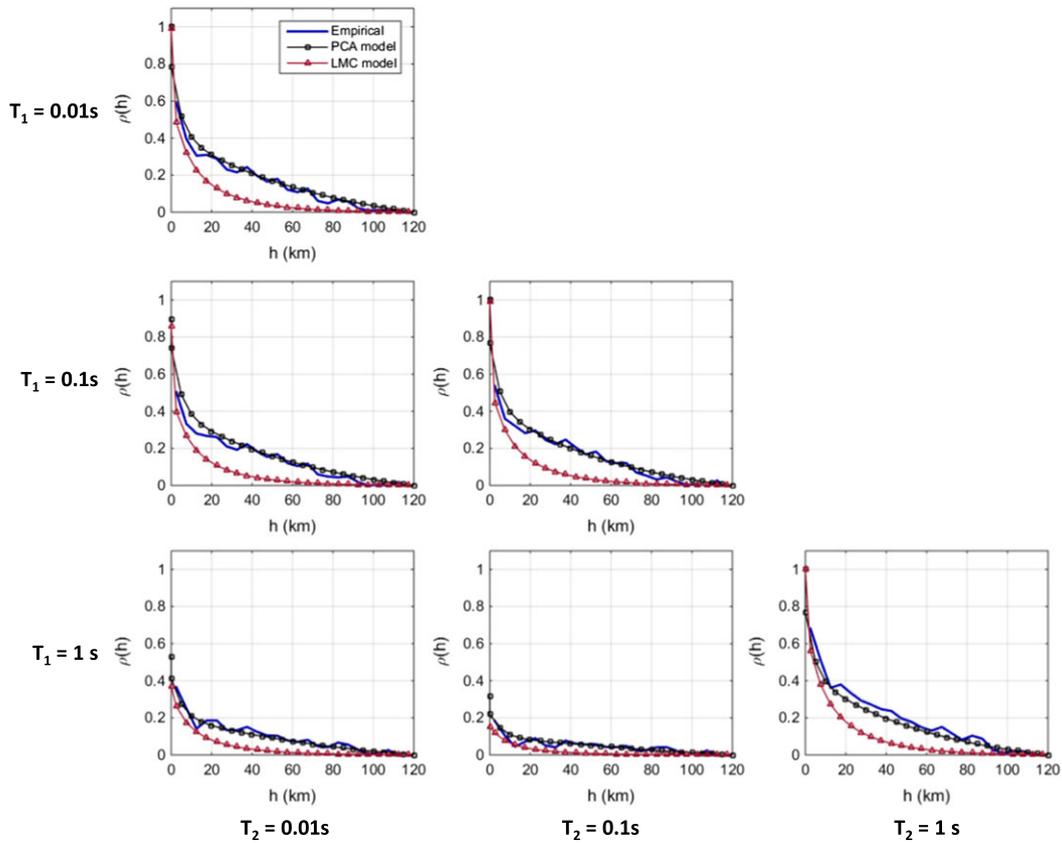


FIGURE 8 Comparison of principal component analysis (PCA) model and linear model of coregionalization (LMC) correlograms and cross-correlograms for different periods

5.2 | Computation time

The PCA model saves substantial computational time when simulating spectral accelerations for multiple periods, because each of the principal components can be simulated independently, thereby reducing the covariance matrix size involved in the simulation of correlated Gaussian random variables.

Equation 30 represents a direct way to simulate multivariate normal random variables with a distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where \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 $\mathbf{T}^T \mathbf{T} = \boldsymbol{\Sigma}$.³¹

$$\mathbf{x} = \mathbf{T}^T \mathbf{b} + \boldsymbol{\mu} \quad (30)$$

The matrix \mathbf{T} is often obtained by Cholesky factorization, which is the most computationally expensive step in the simulation and therefore is the bottleneck in the simulation process. The algorithm's time complexity is $O(q^3)$, which makes factorization of large matrices expensive. For n locations, the LMC and Markov-type screening hypothesis models for ground motion simulation at m periods require the factorization of a $(mn) \times (mn)$ covariance matrix, while the 5 principal component PCA model requires 5 independent factorizations of an $n \times n$ matrix. The theoretical cost of Cholesky factorization, which requires $\frac{1}{3}n^3$ floating-point operations (FLOPS)³² for PCA model and $\frac{1}{3}(mn)^3$ FLOPS for LMC (or Markov-type screening hypothesis) model, is shown in Figure 9B. The computation time for one simulation using both models for various numbers of locations and periods is also presented in Figure 9A, where more than 90% of computation time for generating correlated random variables is taken up by the Cholesky decomposition of the covariance matrix.

Both theoretical and actual computation costs show that PCA is a more efficient method for simulating spectral accelerations for 2 or more periods. This becomes increasingly significant when considering large number of locations at multiple periods, where for 6 periods at 5000 locations the PCA model is over 100 times faster than LMC.

In addition, since all of the operations in the PCA model are performed on smaller or equal size covariance matrices as compared to LMC, substantial memory savings can be achieved.

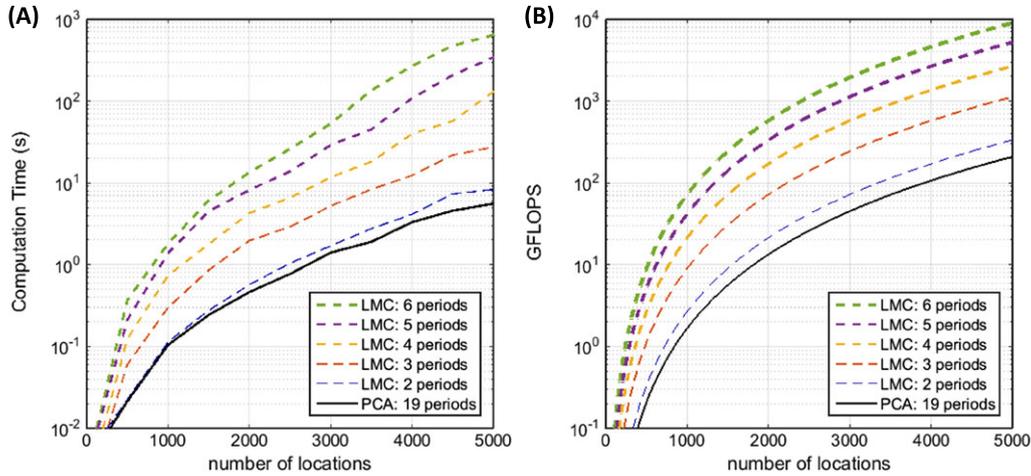


FIGURE 9 Time complexity comparison of principal component analysis (PCA) and linear model of coregionalization (LMC) (or Markov-type model) for simulation of different number of periods: (A), actual run times and (B), theoretical gigaFLOPS (GFLOPS) for Cholesky factorization

6 | IMPLEMENTATION OF THE PREDICTIVE MODEL

The proposed PCA model can be used for simulating ground motion residuals at periods ranging from 0.01 to 5 s. The following is a step-by-step procedure for simulating spatially cross-correlated spectral accelerations. The procedure recommends the use of 5 principal components for simulation, as was shown in Section 4.1; however, the user could include additional components to represent the variance more accurately.

6.1 | Simulation procedure

1. Select n locations and m periods to be simulated.
2. Construct $5 \times n \times n$ covariance matrices $C_i(\mathbf{H})$, for each of the principal components i using Equation 29 with coefficients provided in Table 2 and $\% \sigma_{expl.cum}^2 = 0.95$ (or taken from Table 1 if a different number of components is chosen). \mathbf{H} is an $n \times n$ matrix of separation distances between the considered locations.
3. Simulate spatially correlated variables for each of the components as multivariate normal random variables, $\sim \mathcal{N}(\mathbf{0}, C_i(\mathbf{H}))$.
4. Take a linear combination of the simulated principal components from step 3 and coefficients from Table 1, as per Equation 20, to compute the normalized within-event residuals for each period and location of interest.
5. Multiply the simulated normalized within-event residuals by the standard deviation φ specified in the ground motion model of choice and add predicated median spectral acceleration and between-event residuals as per Equation 1.

The resulting spectral acceleration values for any chosen periods will be correlated both spatially and across periods.

6.2 | Example

A simple example for the simulation of normalized within-event spectral acceleration residuals at 2 periods and 3 locations is presented below. This example considers the use of the first 5 principal components for simulation:

1. Consider 3 locations, located at the coordinates expressed in kilometers: $x_1(0, 0)$, $x_2(7, 12)$, and $x_3(60, 30)$, and simulation of spectral accelerations for periods $T_1 = 0.3$ s and $T_2 = 1$ s.
2. The relative distance between the 3 locations is expressed as the distance matrix, \mathbf{H} :

$$\mathbf{H} = \begin{bmatrix} 0 & 13.89 & 67.08 \\ 13.89 & 0 & 55.97 \\ 67.08 & 55.97 & 0 \end{bmatrix}. \quad (31)$$

The covariance matrices are constructed for each of the principal components using Equation 29 and the corresponding coefficients from Tables 1 and 2. For instance, the covariance between x_1 and x_2 for principal component 1 is

$$C_1(13.89) = \frac{2.50 \times 0 + 4.52 \times \exp\left(\frac{-3 \times 13.89}{15}\right) + 6.78 \times \exp\left(\frac{-3 \times 13.89}{250}\right)}{0.95} = 6.34. \quad (32)$$

Similarly, the covariances are calculated for each of the relative distances for the first 5 principal components and are summarized as $C_i(\mathbf{H})$ below:

$$\begin{aligned} & \begin{matrix} C_1(\mathbf{H}) \\ \begin{bmatrix} 14.53 & 6.34 & 3.19 \\ 6.34 & 14.53 & 3.65 \\ 3.19 & 3.65 & 14.53 \end{bmatrix} \end{matrix} & \begin{matrix} C_2(\mathbf{H}) \\ \begin{bmatrix} 4.74 & 2.13 & 0.78 \\ 2.13 & 4.74 & 0.96 \\ 0.78 & 0.96 & 4.74 \end{bmatrix} \end{matrix} & \begin{matrix} C_3(\mathbf{H}) \\ \begin{bmatrix} 1.26 & 0.54 & 0.19 \\ 0.54 & 1.26 & 0.23 \\ 0.19 & 0.23 & 1.26 \end{bmatrix} \end{matrix} \\ & & \begin{matrix} C_4(\mathbf{H}) \\ \begin{bmatrix} 0.63 & 0.17 & 0.04 \\ 0.17 & 0.63 & 0.06 \\ 0.04 & 0.06 & 0.63 \end{bmatrix} \end{matrix} & \begin{matrix} C_5(\mathbf{H}) \\ \begin{bmatrix} 0.33 & 0.00 & 0.00 \\ 0.00 & 0.33 & 0.00 \\ 0.00 & 0.00 & 0.33 \end{bmatrix} \end{matrix}. \end{aligned} \quad (33)$$

3. Random normally distributed correlated variables are generated using the covariance matrices in Equation 33 for each of the principal components such that $\mathbf{Y}_i \sim \mathcal{N}(\mathbf{0}, C_i(\mathbf{H}))$. Examples are

$$\mathbf{Y}^T = \begin{bmatrix} Y_1 & Y_2 & Y_3 & Y_4 & Y_5 \\ -0.47 & -0.43 & -1.27 & -0.66 & -0.65 \\ -8.92 & -0.57 & -0.34 & 0.23 & 0.03 \\ -0.76 & -0.78 & -0.83 & 0.07 & -0.71 \end{bmatrix} \begin{matrix} x_1 \\ x_2 \\ x_3 \end{matrix}. \quad (34)$$

4. The coefficients for the desired periods for the first 5 principal components are chosen from Table 1.

$$\mathbf{P}^T = \begin{bmatrix} Y_1 & Y_2 & Y_3 & Y_4 & Y_5 \\ 0.25 & 0.07 & -0.24 & 0.36 & 0.04 \\ 0.21 & 0.26 & -0.24 & -0.33 & -0.08 \end{bmatrix} \begin{matrix} T_1 \\ T_2 \end{matrix} \quad (35)$$

To obtain the final normalized residuals for the 2 periods, a linear combination of coefficients and simulated principal components is taken, as per Equation 20:

$$\mathbf{Z} = \mathbf{P}^T \mathbf{Y} = \begin{bmatrix} z_{T1}(x_1) & z_{T1}(x_2) & z_{T1}(x_3) \\ z_{T2}(x_1) & z_{T2}(x_2) & z_{T2}(x_3) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \\ -0.05 & -2.07 & 0.06 \\ 0.26 & -2.02 & -0.24 \end{bmatrix} \begin{matrix} T_1 \\ T_2 \end{matrix}. \quad (36)$$

Code illustrating this type of calculation is also provided as an electronic supplement.

7 | SUMMARY AND CONCLUSIONS

This paper proposes a novel framework for co-simulating spatially correlated spectral accelerations at different periods using a combination of PCA and geostatistical analysis. Data from 45 earthquakes in the NGA-West2 database were used to assess the accuracy and applicability of PCA and to subsequently calibrate a spatial variability model in the principal component space. This model was then compared to other existing models to assess computational efficiency.

Previous approaches to characterizing spatial variability of correlated spectral accelerations involve simultaneously fitting a number of cross-semivariogram models and ensuring that the resulting covariance matrix is positive definite, which is a challenging task. The proposed PCA model enables simulation of uncorrelated principal components, thereby eliminating the need to build cross-semivariogram models. In addition, it was shown that only 5 principal components need to be simulated to capture 95% of variability, further reducing the number of required semivariogram models.

A key result was that the coefficients for principal components exhibit a consistent earthquake-to-earthquake trend. The same trend is also seen when all records are analyzed together. This finding allowed for the development of a model generally applicable to spatial variability of spectral accelerations from shallow crustal earthquakes.

Principal component analysis proved to be an accurate and efficient method for co-simulating spectral accelerations at different periods, which is an important factor when conducting regional analyses. Since simulation of uncorrelated variables requires multiple small covariance matrices, as opposed to one large covariance matrix for cross-correlated variables, it allows for substantial computational savings when co-simulating multiple periods. Comparison of time complexity for

the PCA model and alternative methods shows that PCA is a more efficient method for simulating ground motions for multiple periods at multiple locations.

The implementation of the model is invariant to the number of periods that the user is interested in simulating, since the resulting ground motions for any set of periods will be a linear combination of coefficients in Table 1 and a fixed number of simulated principal components. It is recommended that at least 5 principal components be used in simulation to sufficiently capture the variability. The spatially correlated principal components are simulated using covariance matrices that can be built using Equation 29, which is only dependent on the separation distance between locations of interest. The computational efficiency gained by using this method makes it desirable for modeling regional correlated ground motions at multiple periods.

Finally, the proposed model was shown to be accurate by comparing the resultant semivariogram and cross-semivariogram models with the empirical ones. Although the model was built in the principal component basis, it is closely representative of the empirical spatial variability in the original space.

8 | ELECTRONIC SUPPLEMENT

The empirical model coefficients presented in Tables 1 and 2 are provided as comma-separated value files in Supporting Information. In addition, the MATLAB code for implementation of the predictive model for simulating normalized within-event spectral acceleration residuals at multiple sites and periods is available at https://github.com/bakerjw/Spatial_PCA.

ACKNOWLEDGEMENTS

The authors would like to thank Anne Kiremidjian, Jef Caers, Lewis Li, Alexandre Boucher, and Anna Michalak for their insightful comments and advice during the course of this research project. This work was supported by the National Science Foundation under NSF grant number CMMI 0952402. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

ORCID

Maryia Markhvida  <http://orcid.org/0000-0001-5398-0960>

Luis Ceferino  <http://orcid.org/0000-0003-0322-7510>

Jack W. Baker  <http://orcid.org/0000-0003-2744-9599>

REFERENCES

1. Abrahamson NA, Silva WJ, Kamai R. Summary of the ASK14 ground motion relation for active crustal regions. *Earthquake Spectra*. 2014;30(3):1025-1055.
2. Boore DM, Stewart JP, Seyhan E, Atkinson GM. NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes. *Earthquake Spectra*. 2013;30(3):1057-1085.
3. Campbell KW, Bozorgnia Y. NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra. *Earthquake Spectra*. 2014;30(3):1087-1115.
4. Chiou BSJ, Youngs RR. Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra. *Earthquake Spectra*. 2014;30(3):1117-1153.
5. Idriss IM. An NGA-West2 empirical model for estimating the horizontal spectral values generated by shallow crustal earthquakes. *Earthquake Spectra*. 2014;30(3):1155-1177.
6. Baker JW, Cornell CA. Correlation of response spectral values for multicomponent ground motions. *Bull Seismol Soc Am*. 2006;96(1):215-227.
7. Goda K, Hong HP. Spatial correlation of peak ground motions and response spectra. *Bull Seismol Soc Am*. 2008;98(1):354-365.
8. Jayaram N, Baker JW. Correlation model for spatially distributed ground-motion intensities. *Earthquake Eng Struct Dyn*. 2009;38(15):1687-1708en.
9. Esposito S, Iervolino I. PGA and PGV spatial correlation models based on European multievent datasets. *Bull Seismol Soc Am*. 2011;101(5):2532-2541.
10. Wang M, Takada T. Macrospatial correlation model of seismic ground motions. *Earthquake Spectra*. 2005;21(4):1137-1156.

11. Goda K, Atkinson GM. Intraevent spatial correlation of ground-motion parameters using SK-net data. *Bull Seismol Soc Am*. 2010;100(6):3055-3067.
12. Weatherill GA, Silva V, Crowley H, Bazzurro P. Exploring the impact of spatial correlations and uncertainties for portfolio analysis in probabilistic seismic loss estimation. *Bull Earthquake Eng*. 2015;13(4):957-981.
13. FEMA. FEMA P-58-1: Seismic performance assessment of buildings. Volume 1-methodology. Federal Emergency Management Agency Washington, DC, https://www.fema.gov/media-library-data/1396495019848-0c9252aac91dd1854dc378feb9e69216/FEMAP-58_Volume1_508.pdf; 2012.
14. Yepes C, Silva V, Rossetto T, et al. The global earthquake model physical vulnerability database. *Earthquake Spectra*. 2016;32(4):2567-2585.
15. (FEMA) FEMA. *HAZUS MH-2.1 Earthquake Model Technical Manual*. Washington, D.C.: FEMA; 2015.
16. Loth C, Baker JW. A spatial cross-correlation model of spectral accelerations at multiple periods. *Earthquake Eng Struct Dyn*. 2013;42(3):397-417.
17. Jayaram N, Baker JW. Statistical tests of the joint distribution of spectral acceleration values. *Bull Seismol Soc Am*. 2008;98(5):2231-2243.
18. Ancheta TD, Darragh RB, Stewart JP, et al. NGA-West2 database. *Earthquake Spectra*. 2014;30(3):989-1005.
19. Journel AG, Huijbregts CJ. *Mining Geostatistics*. San Diego, CA: Academic press; 1978.
20. Baker JW, Jayaram N. Effects of spatial correlation of ground motion parameters for multi-site seismic risk assessment: collaborative research with Stanford University and AIR. Technical Report, Final Technical Report. Report for US Geological Survey National Earthquake Hazards Reduction Program (NEHRP) External Research Program Award 07HQGR0031; 2008. [https://web.stanford.edu/~bakerjw/Publications/Baker_Jayaram_\(2008\)_spatial_correlation,_USGS_rpt.pdf](https://web.stanford.edu/~bakerjw/Publications/Baker_Jayaram_(2008)_spatial_correlation,_USGS_rpt.pdf)
21. Isaaks EH, Srivastava R. *An Introduction to Applied Geostatistics*. New York, NY: 551.72 I86, Oxford University Press; 1989.
22. Goovaerts P. *Geostatistics for Natural Resources Evaluation*. New York: Oxford University Press; 1997.
23. Goulard M, Voltz M. Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. *Math Geol*. 1992;24(3):269-286.
24. Journel AG. Markov models for cross-covariances. *Math Geol*. 1999;31(8):955-964.
25. Chiles JP, Delfiner P. *Geostatistics: Modeling Spatial Uncertainty*, Vol. 497. Hoboken, NJ: John Wiley & Sons; 2009.
26. Jackson JE. *A User's Guide to Principal Components*, Vol. 587. Hoboken, NJ: John Wiley & Sons; 2005.
27. Cangelosi R, Goriely A. Component retention in principal component analysis with application to cDNA microarray data. *Biol Direct*. 2007;2(1):2. <https://doi.org/10.1186/1745-6150-2-2>
28. Jolliffe I. *Principal Component Analysis*. New York, NY: Wiley Online Library; 2002.
29. Wackernagel H. *Multivariate Geostatistics: An Introduction with Applications*. Berlin, Germany: Springer Science & Business Media; 2003.
30. Chiou B, Darragh R, Gregor N, Silva W. NGA project strong-motion database. *Earthquake Spectra*. 2008;24(1):23-44.
31. Gentle JE. *Random Number Generation and Monte Carlo Methods*. New York, NY: Springer Science & Business Media; 2003.
32. Trefethen LN, Bau DIII. *Numerical Linear Algebra*, Vol. 50. Philadelphia, PA: SIAM; 1997.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Markhvida M, Ceferino L, Baker JW. Modeling spatially correlated spectral accelerations at multiple periods using principal component analysis and geostatistics. *Earthquake Engng Struct Dyn*. 2018;47:1107–1123. <https://doi.org/10.1002/eqe.3007>