Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Scalable Scenario-based Regional Earthquake Risk Assessment via Gaussianized Ground-motion–damage Modeling and PPCA

Soung Eil Houng | **Luis Ceferino**

[1]Civil and Environmental Engineering, University of California, Berkeley, CA

Correspondence
Corresponding author Soung Eil Houng
Email: shoung@berkeley.edu

Present address
Davis Hall, Berkeley, CA 94706

Abstract

In scenario-based regional risk modeling, the traditional workflow simulates spatially correlated ground motions, and subsequently samples building damage states from lognormal fragility functions. In this procedure, the dimensionality grows with the number of assets (N) and quickly becomes computationally prohibitive for large cities. To overcome this limitation, we introduce a scalable computational framework that (i) recasts the traditional two-step (ground-motion, then damage) simulation as a single, N-dimensional Gaussian sampling problem via an exact change of variables, and (ii) identifies low-dimensional latent variables that make this sampling efficient by employing probabilistic principal component analysis (PPCA). We validate the proposed approach on San Francisco's downtown portfolio of 1,000 buildings, benchmarking against SimCenter R2D's computational testbed. The modal damage states of $> 95\%$ buildings match exactly, with a mean difference below 0.04 (on a 0–4 ordinal scale representing none to complete damage), confirming the framework's accuracy. In tests on downtown San Francisco (15,836 buildings) and the broader Bay Area, a single latent dimension and 20 dimensions, respectively, reproduce the benchmark loss distributions within $<\sim 2.5\%$. The achievable dimensionality reduction depends primarily on the portfolio's spatial extent rather than building density. As a result, the computational complexity drops by one order in N relative to the traditional approach – from $O(N^3)$ to $O(N^2)$ in the pre-processing step and from $O(N^2M)$ to $O(NM)$ in the simulation step, where M is the number of simulations. For 30,000 buildings, the method yields roughly $9\times$ faster pre-processing and $97\times$ faster simulation, with speedups growing linearly with portfolio size, resulting in $63\times$ faster total computation time for $M = 10^5$, compared to the traditional framework. Overall, the framework substantially lowers the computational barrier for regional seismic risk assessment of dense urban building portfolios.

KEYWORDS
regional seismic risk; scenario-based risk modeling, ground-motion–fragility coupling, dimensionality reduction, probabilistic principal component analysis

1 | INTRODUCTION

Scenario-based seismic risk analysis (SRA) estimates losses for a building portfolio subjected to a specified earthquake scenario (Figure 1). By focusing on a single representative event rather than aggregating outcomes across all possible earthquakes, this approach provides an actionable view of risk, supporting risk mitigation decisions on where and how much to invest, which assets to prioritize for retrofitting, and how to plan evacuations [1,2,3,4,5]. The type of loss considered depends on the intended application and may include financial losses, casualties and injuries, or infrastructure downtime [6,7,8,9,10,11]. However, most of these analyses rely on computationally intensive frameworks originally developed for individual structures, limiting their scalability to large regional portfolios.

The performance-based earthquake engineering (PBEE) framework [12] represents one of the most widely adopted methodologies for individual structures. This framework sequentially quantifies loss by first generating ground motion intensities, simulating

structural responses, estimating damage states based on those responses, and finally calculating the resulting losses. Regional-scale risk analysis fundamentally follows this same logic, albeit by utilizing spatially distributed quantities such as ground motion intensity maps, damage maps, and loss maps (Figure 1). At this scale, explicit response analysis is frequently bypassed due to its high computational cost; instead, fragility curves are employed to directly evaluate building damage from ground shaking intensity. The total regional loss is then determined by aggregating the losses of all constituent buildings.

While this workflow is conceptually straightforward, its implementation at scale faces computational hurdles. Specifically, the generation of the ground shaking intensity maps is recognized as one of the most computationally demanding phase of regional SRA (Figure 1 (b) and (c)). Regional shaking is typically modeled using empirical ground-motion models integrated within a Gaussian process framework to account for its spatial correlation (Figure 1 (b))[13,14,15,16]. As a result, this step presents a major bottleneck as it necessitates simulating multivariate normally distributed random variables (one per building) through an expensive Cholesky decomposition of a large, dense correlation matrix and repeated sampling of within-event residuals[17,18]. Consequently, the computational time scales poorly with the building inventory size N and the number of simulations M: $O(N^3)$ for the factorization and $O(N^2M)$ for the sampling, with $O(N^2)$ memory[19].

Once the ground shaking intensity map is established, building damage is modeled using fragility curves, which at a given ground shaking intensity defines the probability that a building will exceed a specific damage state; thus, this process is inherently stochastic (Figure 1 (d) and (e)). In this stage, because damage must be simulated across all portfolio locations, this phase remains high-dimensional, however, generally less computationally burdensome than ground shaking modeling since it is typically modeled as independent between buildings. Following the damage assessment, loss metrics—such as repair costs or recovery times—are modeled based on the predicted damage states (Figure 1 (f) and (g)). The complexity of the loss model significantly influences the overall computational demand of the risk analysis[20]. The simplest approach utilizes look-up tables or deterministic ratios for each damage state[21].

To date, efforts to reduce the computational cost of regional SRA have focused mainly on reducing the number of scenarios—through importance sampling[22,23,24,25] and optimization-based scenario selection[26,27,28,29]. In addition, advanced sampling methods developed for rare-event simulation[30,31,32,33] are available to similarly reduce the number of required simulations. Beyond such scenario-reduction methods, the computational cost can be reduced further by combining them with per-simulation cost-reduction methods. In particular, when a simple Hazus-like loss model[21] and standard ground-motion–damage fragility curves are used (Figure 1), ground-shaking simulation becomes the primary bottleneck. Thus, reducing its computation time is beneficial, especially for densely packed building portfolios[34].

However, comparatively little effort has been directed toward efficiently simulating the spatially correlated ground-shaking maps required in this setting. The most common strategy is to employ coarse "ground-shaking grids" to alleviate this cost, assigning each asset the intensity of the nearest grid point or an interpolated value[17]. Although such techniques substantially reduce the overhead of generating correlated ground motions, they can introduce bias by assigning identical intensities to distinct sites that, in reality, experience different shaking[35,36,34]. Other studies have applied principal component analysis (PCA) to accelerate the simulation of ground shaking intensities[37,38]; however, these approaches primarily benefit the modeling of multiple intensity measures across different periods and do not fundamentally address spatial scalability. Yang et al.[39] accelerated spatially correlated ground shaking simulation by combining exact Cholesky decomposition on coarse-grid locations with kriging interpolation of the remaining sites, which is efficient for smooth correlation kernels. More recently, sparse Cholesky decomposition has been applied to efficiently simulate spatially correlated non-ergodic ground shaking[40].

Generating a spatially correlated ground-shaking map amounts to sampling an N-dimensional Gaussian vector. A broad body of dimensionality-reduction and covariance-approximation techniques can be brought to bear on this problem[41,42,43], and they may help reduce the computational burden of large-scale, per-asset ground-shaking map simulation. In this study, however, rather than applying those techniques directly to ground-shaking map simulation, we first recast the traditional two-step ground-shaking–damage simulation to a single, unified N-dimensional Gaussian sampling problem, i.e., Gaussianization of the ground-motion–damage modeling. This is achieved through a linearizing transformation of the ground-motion–damage coupling, using a mathematically exact change-of-variable rather than an approximation. This transformation is novel and has not been reported previously. This Gaussianization 1) allows simpler damage sampling in which ground shaking is simulated implicitly and 2) transparently shows how the coupled ground-motion and damage parameters affect the final damage field. We then integrate probabilistic principal component analysis (PPCA) into this SRA modeling framework to identify low-dimensional latent variables and the residual independent component of the Gaussianized formulation, thereby enabling high-performance computations for large-scale building portfolios.

The proposed framework is mathematically formulated, interpreted using simple examples, and its performance is evaluated through a case study of the San Francisco Bay Area—extending from the urban downtown to the entire regional scale—under a major earthquake scenario on the San Andreas Fault.

2 | TRADITIONAL REGIONAL SCENARIO-BASED SEISMIC RISK ANALYSIS

Following the principles of PBEE, the resulting loss distribution is generally mathematically intractable in closed form. Thus, Monte Carlo (MC) simulation is utilized to achieve numerical convergence. The required number of MC realizations is typically dictated by the specific exceedance probabilities or tail risks of interest. The following sections detail the traditional implementation of each phase, including the simulation of ground-shaking (Figure 1(b) and (c)), damage and subsequent loss (Figure 1(d)–(g)).

2.1 | Ground Shaking

Given an earthquake scenario, the ground motion $y \in \mathbb{R}_+$ at a single site is lognormally distributed. Thus, we can simulate it using the logarithmic mean of the ground motion (μ) and samples of between-event and within-event residuals, expressed as a linear combination of two standard normal random variables. For a single site:

$$\ln y = \mu + \tau\,\zeta + \phi\,\xi, \tag{1}$$

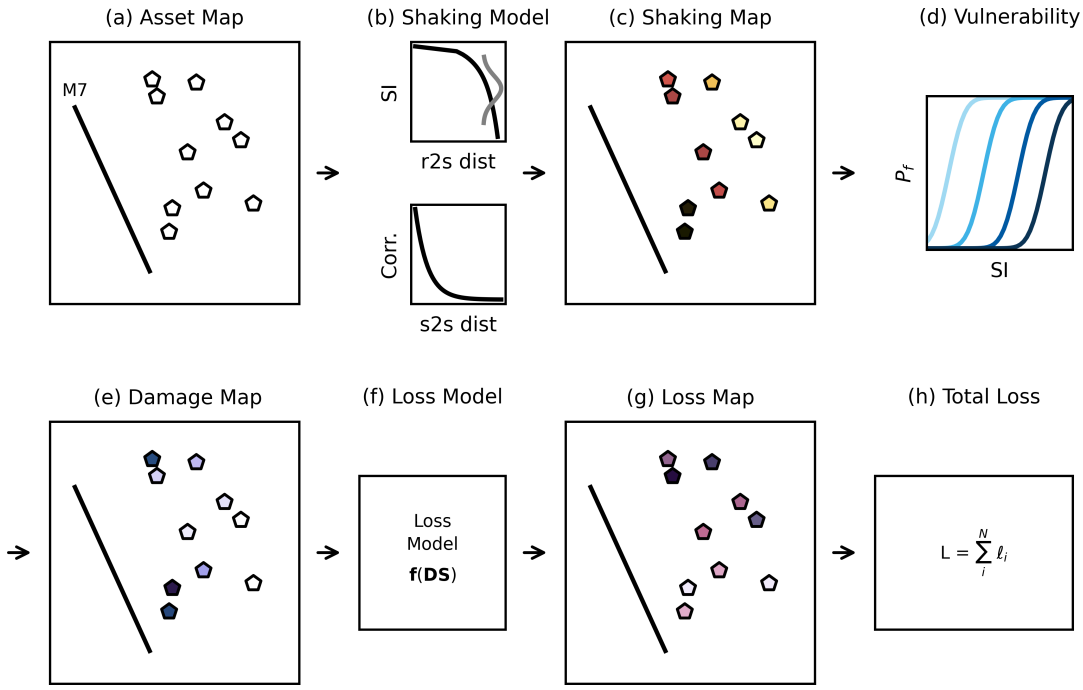

FIGURE 1 Schematic diagram of the traditional scenario-based risk modeling framework. (a) Asset distribution map featuring individual assets (pentagons) and a scenario earthquake rupture of M 7.0 (solid black line). (b) Models for generating shaking maps: (top) Ground Motion Model (GMM) and (bottom) correlation model. r2s dist: rupture-to-site distance; SI: shaking intensity; s2s distance: inter-site distance; and Corr.: within-event spatial correlation. (c) A realization of a shaking intensity map, where increasing color intensity indicates higher ground shaking. (d) Fragility curves representing vulnerability for multiple damage limit states (P_f). (e) A realization of a damage map. (f) Implementation of a user-defined loss model. (g) Resultant loss map illustrating the distribution of consequences. (h) Total loss calculated as the sum of all individual losses.

where τ and ϕ are between- and within-event standard deviations, respectively; and ζ and ξ are independent standard normal random variables, $\mathcal{N}(0,1)$. The terms $\tau\zeta$ and $\phi\xi$ provide stochastic realizations of the between-event and within-event residuals, respectively. The term μ is logarithmic mean ground motion intensity, and τ and ϕ are between-event and within-event standard deviations, respectively.

For a region, we model ground shaking $\mathbf{y}$ at multiple sites simultaneously, explicitly incorporating the spatial correlation of ground motions. The within-event correlation matrix is typically modeled as a function of inter-site distance, with higher correlation assigned to adjacent sites and lower correlation to more distant sites[15,14,13]. The correlation matrix is $\mathbf{C} = [\rho_{ij}] \in \mathbb{R}^{N\times N}$, where $\rho_{ii} = 1$ because each site has a unique ground-shaking intensity, and $\rho_{ij} = \rho_{ji}$ since the correlation between sites i and j depends solely on their inter-site distance, which is symmetric. To model the within-event residuals with correlation matrix $\mathbf{C}$, we seek any matrix $\mathbf{L}$ such that $\mathbf{L}\mathbf{L}^{\mathsf{T}} = \mathbf{C}$. The matrix $\mathbf{L}$ is typically obtained via the Cholesky decomposition of $\mathbf{C}$, for its simplicity and numerical stability[44]. Accordingly, $\mathbf{L}$ is the lower triangular matrix. Therefore, for N buildings, the ground motions $\mathbf{y} \in \mathbb{R}_+^N$ across the sites are given by

$$\ln \mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\tau}\,\zeta + \boldsymbol{\phi} \circ (\mathbf{L}\,\boldsymbol{\xi}), \tag{2}$$

where $\boldsymbol{\mu} \in \mathbb{R}^N$ is the vector of logarithm of median shaking intensities; $\boldsymbol{\tau} \in \mathbb{R}^N$ and $\boldsymbol{\phi} \in \mathbb{R}^N$ are the vectors of between- and within-event standard deviations; $\mathbf{L} \in \mathbb{R}^{N\times N}$ is the lower-triangular Cholesky factor of the within-event spatial correlation matrix $\mathbf{C} \in \mathbb{R}^{N\times N}$; $\zeta \in \mathbb{R} \sim \mathcal{N}(0,1)$ is the scalar standard normal random variable; and $\boldsymbol{\xi} \in \mathbb{R}^N \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is the standard normal random vector. The operator $\circ$ denotes element-wise multiplication. Note that ζ, which generates the between-event residual, is a scalar for a given earthquake scenario and affects all sites the same. By contrast, the within-event residuals are generated by a correlated N-dimensional random vector and therefore require $\boldsymbol{\xi} \in \mathbb{R}^N$. The total covariance matrix of the ground shaking intensity given earthquake scenario ($\ln \mathbf{y}$ in Equation 2) is $\mathbf{Cov}(\ln \mathbf{y}) = \boldsymbol{\tau}\boldsymbol{\tau}^{\mathsf{T}} + \mathbf{FCF}$, where $\mathbf{F} = \mathrm{diag}(\boldsymbol{\phi})$. Also, the total correlation of the ground shaking intensity can be obtained by diagonal normalization of $\mathbf{Cov}(\ln \mathbf{y})$, which is $\mathbf{C}_{\ln \mathbf{y}} = \mathrm{diag}\{\mathbf{Cov}(\ln \mathbf{y})\}^{-1/2}\ \mathbf{Cov}(\ln \mathbf{y})\ \mathrm{diag}\{\mathbf{Cov}(\ln \mathbf{y})\}^{-1/2}$

In most recent ground-motion models, $\boldsymbol{\phi}$ and $\boldsymbol{\tau}$ are site dependent and vary with local soil conditions, particularly when the shaking intensity amplification term is nonlinear with respect to ground-shaking intensity[45,46,47,48]. If nonlinearity is neglected, some models can be modeled as magnitude-dependent only. Thus, for a fixed scenario, we could treat these standard deviations across sites as constant[48]. However, for generality, we retain the vector notation. For peak ground acceleration (PGA), $\boldsymbol{\tau}$ typical ranges from 0.30 to 0.50, and $\boldsymbol{\phi}$ from 0.50 to 0.75, with both tending to decrease for larger magnitudes; exact ranges depend on the model[49].

2.2 | Damage States

We define damage as an ordinal variable with n_{ds} states, indexed by $k = 0, 1, \ldots, n_{\mathrm{ds}} - 1$, where larger k indicates more severe damage. A damage state of $k = 0$ denotes no damage, and $k = n_{\mathrm{ds}} - 1$ denotes the most severe state. In typical applications, we take $n_{\mathrm{ds}} = 5$, corresponding to the following categories: No damage ($k = 0$), Slight ($k = 1$), Moderate ($k = 2$), Extensive ($k = 3$), and Complete ($k = 4 = n_{\mathrm{ds}} - 1$)[21].

The damage probabilities are typically modeled via lognormal fragility functions, which describe the likelihood of exceeding a given damage threshold as a function of ground-motion intensity. For damage state k, the exceedance probability is expressed as

$$P(DS \geq k \mid y) = \Phi\left(\frac{\ln y - \ln \theta_k}{\beta}\right), \tag{3}$$

where $P(DS \geq k \mid y)$ is the probability that the building experiences a damage state k or greater at shaking intensity y (e.g., PGA of 0.3g), $\ln \theta_k$ is the logarithm of the shaking intensity with 50% probability that $DS \geq k$, β is the logarithmic standard deviation, which is often assumed constant across damage states to avoid crossing fragility curves[21], and $\Phi(\cdot)$ is the CDF of $\mathcal{N}(0,1)$. In this study, damage is modeled independently conditioned on the ground shaking intensity, which is a common assumption in regional-scale risk models. However, damage states can be correlated when structure–soil–structure interaction (SSSI) is significant for adjacent large structures, or when one building collapses onto another[50,51]. Nonetheless, regional models have not been developed yet to account for these effects in large building portfolios.

In practice, to simulate damage, we draw $u \sim \mathcal{U}(0,1)$. If $u > P(DS \geq k \mid y)$ or equivalently, $u - P(DS \geq k \mid y) > 0$, then $DS < k$; otherwise, $DS \geq k$. In order to derive our approach, we define the limit-state function G_k:

$$G_k = \begin{cases} -\infty, & \text{if } k = 0, \\ u - \Phi\left(\frac{\ln y - \ln \theta_k}{\beta}\right), & \text{if } k = 1, \ldots, n_{\text{ds}} - 1. \end{cases} \tag{4}$$

If $u > P(DS \geq k \mid y)$, then $G_k > 0$, implying that the damage state is lower than k ($DS < k$). Conversely, if $u < P(DS \geq k \mid y)$, then $G_k < 0$, implying that the damage state is at least k ($DS \geq k$). The simulated damage state is determined by selecting the maximum k such that $G_k < 0$:

$$DS = \max\{k \in \{0, \ldots, n_{\text{ds}} - 1\} \mid G_k < 0\} \tag{5}$$

That is, DS is the largest k for which $G_k < 0$. Note that such a k always exist because we set $G_0 = -\infty$ (Equation 4), ensuring $G_0 < 0$. Thus, the minimum possible value is $DS = 0$ (No damage). Importantly, as long as the sign of G_k is preserved, the simulated damage state does not change. Given DS, the loss associated with DS, l, is computed by multiplying the building's value V (e.g., replacement cost, number of occupants) by the loss ratio function, $r(\cdot)$:

$$l = V \times r(DS). \tag{6}$$

Also, the total loss is then obtained by summing over all buildings, $l_{\text{t}} = \sum_{i=1}^{N} l_i$.

3 | RECASTING OF GROUND-MOTION–DAMAGE COUPLING INTO A GAUSSIAN SAMPLING PROBLEM

As discussed earlier, seismic risk is typically computed in two separate stages: ground shaking and damage simulations. Integrating these stages is challenging because the damage simulation involves a nonlinear fragility curve relating the Gaussian shaking intensity, $\ln y$ (Equation 1) to the uniform damage-state random variable, u (Equation 4). In this section, we reformulate this traditional two-step procedure into a single Gaussian sampling operation. This is achieved through an exact linearizing transformation of ground-motion–damage coupling using a change of variables, without assumptions or approximations such as a Taylor Expansion. This transformation is remarkably simple and intuitive. We first derive this Gaussianization formulation for a single site, give it an intuitive geometric interpretation, and then extend it to the multi-site case.

3.1 | Single Site Gaussianization

We begin by transforming u via change-of-variable (or probit transformation, inverse transform sampling method): $u = \Phi(\varepsilon)$, where ε is a standard normal random variable and $\Phi(\cdot)$ is the standard normal cumulative distribution function (CDF). Then, Equation 4 for $k \geq 1$ becomes:

$$G_k = \Phi(\varepsilon) - \Phi\left(\frac{\ln y - \ln \theta_k}{\beta}\right) \tag{7}$$

Since $\Phi(\cdot)$ is monotonically increasing, it always preserves the order of inputs; thus, G_k has the same sign as the difference of its inputs. Thus, we can model damage exactly via the sign of the following variable, g_k:

$$g_k = \varepsilon + \frac{1}{\beta}(\ln \theta_k - \ln y) \tag{8}$$

This yields the transformed formulation for a single-building damage simulation by substituting Equation 1 for $\ln y$:

$$g_k = -\frac{1}{\beta}(\tau\zeta + \phi\xi - \beta\varepsilon) + \frac{1}{\beta}(\ln \theta_k - \mu) \tag{9}$$

The first term of Equation 9, $-\frac{1}{\beta}(\tau\zeta + \phi\xi - \beta\varepsilon)$, is a linear combination of three independent standard normal variables (ε, ζ, and ξ), which still is a normally distributed and its variance is $1 + (\tau^2 + \phi^2)/\beta^2$ or $1 + \sigma^2/\beta^2$, where $\sigma = \sqrt{\tau^2 + \phi^2}$ is the total

standard deviation of the ground shaking intensity. The second term, $\frac{1}{\beta}(\ln\theta_k - \mu)$, contains no random variables and thus acts as a deterministic bias, i.e., a shift in the mean of the distribution. Therefore, g_k is normally distributed as:

$$g_k \sim \mathcal{N}\left(\frac{\ln\theta_k - \mu}{\beta}, 1 + \frac{\sigma^2}{\beta^2}\right) \tag{10}$$

Damage simulation can be performed similarly to the traditional approach (Equation 5), with the substitution of G_k by g_k:

$$DS = \max\{k \in \{0, \ldots, n_{\text{ds}} - 1\} \mid g_k < 0\} \tag{11}$$

Recall that in the traditional framework, damage simulation involves a two-step procedure: i) simulating ground motion (Equation 1), and ii) evaluating Equations 3 to compute G_k. The procedure incorporates three random variables—ζ, ξ, and u (Equations 1 and 4). In contrast, the Gaussianized damage simulation using g_k based on Equations 10 streamlines the traditional two-step modeling process into a single-step formulation. Furthermore, this approach eliminates redundant random variables, enabling the damage to be modeled using a single Gaussian random variable.

Furthermore, the Gaussianization provides a transparent perspective on the coupling between ground-motion models and fragility curve parameters, elucidating their joint influence on the resulting risk (Equation 10). For instance, an increase in θ_k or a decrease in μ shifts the distribution toward the positive domain, thereby reducing the probability of damage. An increase in σ flattens the distribution, which leads to an increase in risk when $\ln\theta_k > \mu$ and a decrease when $\ln\theta_k < \mu$. Finally, as β increases, the distribution converges toward $\mathcal{N}(0, 1)$; in this limiting case, $P[DS > k]$ approaches 0.5, and the specific values of μ and θ_k become increasingly irrelevant to the risk assessment. Further details regarding these parametric sensitivities are provided in the Section S1 and Figure S1 of Supplementary Material.

3.2 | Geometric Interpretation of the Single-Site Gaussianization

We demonstrate geometrically how the Gaussianization reduces the traditional framework's three-dimensional system (ζ, ξ, and u) to a single dimension. We begin by analyzing the distribution of G_k (Equation 4) in the traditional framework within the space spanned by the ground shaking intensity, $\ln y$, and the uniform damage sample, u. For a given k, we map the regions where G_k is negative or positive within the $(\ln y, u)$ space (Figure 2(a)). As shown in Figure 2(a), the regions corresponding to $G_k > 0$ (or $DS < k$) and $G_k \leq 0$ (or $DS \geq k$) are well separated, with $G_k = 0$ acting as the boundary (red and blue shaded regions separated by a black curve). This boundary naturally traces the normal CDF fragility function, forming a non-linear curve, as the limit state $G_k = 0$ is defined by $u = P(DS \geq k)$ (Equation 4). Due to this non-linearity, the gradient vector of G_k—the direction of maximum change—varies depending on the specific values of $\ln y$ and u. This is evident from the varying orientations of the contour lines (gray dotted lines in Figure 2(a)).

In contrast, the transformed formulation employing g_k (Equation 9) converts this non-linear boundary into a linear one characterized by a unique gradient direction of g_k. The limit-state boundary transforms into a straight line in the $(\ln y, \varepsilon)$ space (Figure 2(b); green and purple shaded regions separated by a black line). Thus, to maintain $g_k = 0$, a change in $\ln y$ requires a constant proportional change in ε. This linearity applies not only to $g_k = 0$ but to any constant value of g_k; thus, the contours (isolines) of g_k are parallel to the limit-state line. Therefore, the variation of g_k occurs exclusively in the direction perpendicular to these contours, corresponding to a slope of $1/\beta$ (Figure 2(b)).

We extend this concept to three-dimensional space by decomposing $\ln y$ into its components, ξ and ζ (Equation 1). In the traditional framework, a curved surface (black surface in Figure 2(c)) separates the failure ($G_k < 0$) and non-failure ($G_k \geq 0$) regions (red and blue shaded regions). As discussed with Figure 2(a), the curvature of this surface prevents the identification of a single, constant direction responsible for the variation in G_k. In contrast, our formulation yields a linear planar boundary in the three-dimensional $(\xi, \zeta, \varepsilon)$ space (Figure 2(d); green and purple shaded regions separated by a plane), analogous to the linear boundary in Figure 2(b). Consequently, the variation in g_k occurs entirely along the unique direction defined by the normal vector of this planar boundary. This normal vector (**n**) is $[\,\tau/\beta,\ \phi/\beta,\ -1\,]$, derived from Equation 9. This property allows us to model g_k effectively as a one-dimensional problem along this direction using a single Gaussian random variable, as any vector component orthogonal to **n** does not cause variation in g_k (Equation 10).

We can extend this visualization to include all possible k values, drawing the limit state boundaries for multiple damage state categories (Figure 3). Figure 3(a) illustrates these boundaries in the traditional framework using uniform damage samples u, while Figure 3(b) shows them in the transformed space with standard normal damage samples ε. In the traditional framework, the limit state boundaries appear as non-linear curved surfaces with slopes that vary by location and k. Conversely, our transformed

formulation exhibits perfectly linear planar boundaries, defined by the normal vector $\mathbf{n}$ (Figure 3(b)). Notably, if the same β is used for all k, these linear boundaries are parallel as seen in Figure 3(b). If different β values are used, they do not share $\mathbf{n}$, thus the planes are not parallel and may intersect.

3.3 | Multiple-Site Gaussianization

For a region, the damage function Gaussianization for ith building can be expressed similar to single building case by incorporating the standard normal damage variable ε_i instead of u_i (Equation 9), but incorporating the within-event spatial correlation of ground shaking (Equation 2):

$$g_{ik} = -\frac{1}{\beta_i}\left(\tau_i\zeta + \phi_i(\mathbf{L}_i \cdot \boldsymbol{\xi})\right) + \varepsilon_i + \frac{1}{\beta_i}(\ln\theta_i^k - \mu_i), \tag{12}$$

where $\mathbf{L}_i$ is the ith row of $\mathbf{L}$ from Equation 2, and $\boldsymbol{\xi} \in \mathbb{R}^N \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. This can be expressed in vector form as:

$$\mathbf{g}_k = \mathbf{A}\mathbf{z} + \mathbf{b}_k \tag{13}$$

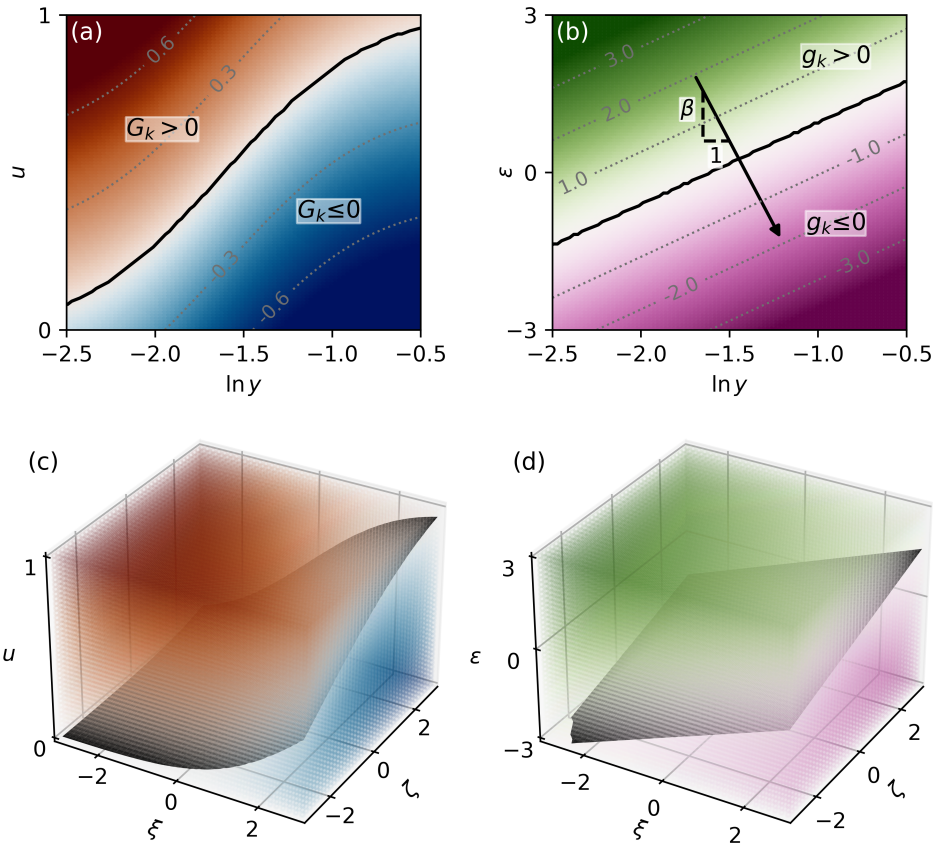

FIGURE 2 Comparison of the damage function behavior in the traditional and transformed space. (a) Variation of G_k in the traditional $(\ln y, u)$ space. (b) Variation of g_k in the transformed $(\ln y, \varepsilon)$ space. Red and green shaded zones indicate regions with positive values ($DS < k$), while blue and purple zones indicate negative values ($DS \geq k$). The thick black line indicates the limit state ($G_k = 0$ or $g_k = 0$), and gray dotted lines represent contours. Darker colors indicate larger absolute values. The black arrow in (b) denotes the direction of variation for g_k. (c) Variation of G_k in the 3D (ξ, ζ, u) space. (d) Variation of g_k in the 3D $(\xi, \zeta, \varepsilon)$ space. The color code matches (a) and (b): (c) uses red/blue and (d) uses green/purple. The black curved surface in (c) and the planar surface in (d) represent the limit states.

where the damage for the i-th building is simulated using the sign of the i-th entry of $\mathbf{g}_k \in \mathbb{R}^N$ (Equation 11). The components are defined as $\mathbf{A} \in \mathbb{R}^{N\times(2N+1)}$, $\mathbf{b}_k \in \mathbb{R}^N$, and $\mathbf{z} \in \mathbb{R}^{2N+1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$:

$$A = \begin{bmatrix} -\mathbf{B}\boldsymbol{\tau} & -\mathbf{BFL} & \mathbf{I} \end{bmatrix}, \qquad \mathbf{b}_k = \mathbf{B}(\ln \boldsymbol{\theta}_k - \boldsymbol{\mu}), \qquad \mathbf{z} = \begin{bmatrix} \zeta \\ \boldsymbol{\xi} \\ \boldsymbol{\varepsilon} \end{bmatrix}. \tag{14}$$

Here, $\mathbf{B} = \text{diag}(1/\beta_1, \ldots, 1/\beta_N)$ and $\mathbf{F} = \text{diag}(\phi_1, \ldots, \phi_N)$. The matrix $\mathbf{A}$ contains the uncertainty-related components: the ground-motion standard deviations ($\boldsymbol{\tau}$ and $\boldsymbol{\phi}$), the Cholesky factor $\mathbf{L}$, and the dispersions of the fragility curves $\boldsymbol{\beta}$. Similar to Equation 9, the vector $\mathbf{b}_k$ contains the deterministic, mean-shifting components: the median ground shaking of the fragility curve $\boldsymbol{\theta}_k$ minus logarithmic median of ground shakings $\boldsymbol{\mu}$, divided by dispersion parameter of fragility curve $\boldsymbol{\beta}$.

Equation 13 defines a transformation of a $(2N + 1)$-dimensional standard normal vector $\mathbf{z}$ into an N-dimensional normally distributed vector, $\mathbf{g}_k$, with mean $\mathbf{b}_k$ and covariance matrix $\mathbf{AA}^\mathsf{T}$:

$$\mathbf{g}_k \sim \mathcal{N}(\mathbf{b}_k, \mathbf{AA}^\mathsf{T}), \tag{15}$$

where the covariance matrix $\mathbf{AA}^\mathsf{T} \in \mathbb{R}^{N\times N}$ is:

$$\mathbf{Cov}(\mathbf{g_k}) = \mathbf{AA}^\mathsf{T} = \mathbf{I} + (\mathbf{B}\boldsymbol{\tau})(\mathbf{B}\boldsymbol{\tau})^\mathsf{T} + \mathbf{BFCFB}. \tag{16}$$

This indicates that the damage to N buildings can be modeled using a set of correlated N-dimensional Gaussian random variables, where the damage of the ith building corresponds to the ith Gaussian random variable, and the categorical damage states are encoded by the signs of these random samples. Negative entries of $\mathbf{g}_k$ correspond to damage states $DS \geq k$, whereas non-negative entries correspond to $DS < k$. Also, note that the damages are simulated using a single realization of $\mathbf{g}_k$ per MC simulation; that is, the zero-mean random samples are generated once for all ks, and mean shifting term ($\mathbf{b}_k$) is applied to decide if $DS \geq k$ or not.

A key advantage of this formulation is its simplicity. It eliminates the need to evaluate damage-state probabilities from the fragility curves as a function of the simulated ground-shaking intensity, since this information is already embedded in the mean and covariance structure of $\mathbf{g}_k$. In addition, similar to the single-building case, damage modeling of N buildings in the traditional framework requires $2N + 1$ random variables: $N + 1$ for ground-motion simulation (Equation 2) and N for damage modeling (Equation 5). By contrast, the proposed formulation enables a one-shot procedure using Equation 15 and reduces the dimensionality by approximately half for large N ($2N + 1$ to N). All of this is achieved without introducing additional mathematical assumptions.

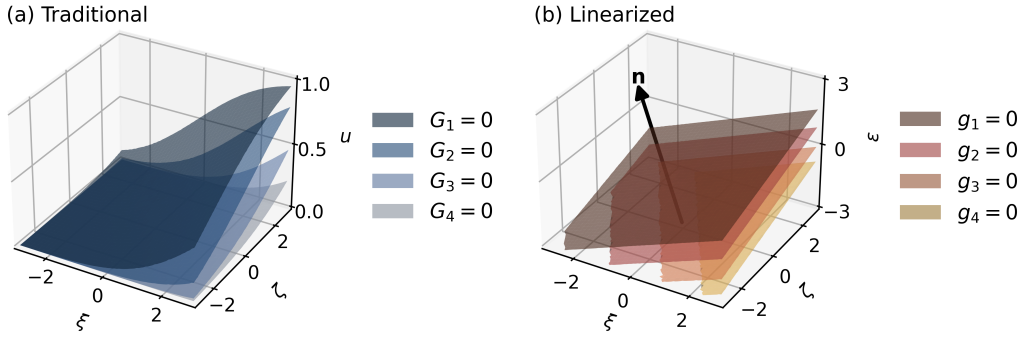

FIGURE 3 Boundaries separating damage state categories ($k = 1$ to 4) in: (a) the (ξ, ζ, u) space using the traditional framework, and (b) the transformed $(\xi, \zeta, \varepsilon)$ space in which the ground-motion–damage coupling is linearized. Darker planes indicate lower damage state thresholds, while lighter planes indicate higher damage state thresholds. The black arrow indicates the normal vector (**n**)–the direction along which g_k varies–which is common to all damage states.

However, this dimensionality reduction does not by itself imply a significant improvement in computational efficiency. Although it simplifies the traditional framework, the simulation still requires an N-dimensional correlated Gaussian sampling. Its computational burden may be comparable to that of the spatially correlated within-event residual computation—the main computational bottleneck of the traditional two-step method. Further investigation is therefore needed to reduce the computational burden of this formulation.

For more detailed and intuitive statistical characteristics of $\mathbf{g}_k$, see Section S2 and Figure S2 of the Supplementary Material.

4 | DIMENSIONALITY REDUCTION IN REGIONAL SEISMIC RISK

We investigate the potential of PPCA for further dimensionality reduction of the Gaussianized formulation (Equations 13 and 15) to reduce its computational burden. Given the structure of its covariance $\mathbf{Cov}(\mathbf{g}_k)$ (Equation 16), PPCA is a promising strategy because it decomposes the field variance into a strongly correlated low-rank component plus an independent noise floor. Other dimensionality reduction methods are available, such as principal component analysis (PCA)[52], the Nyström approximation[53,54], and pivoted Cholesky decomposition[55]. However, in general, these methods are effective only when the variance is heavily concentrated in a low-dimensional subspace spanned by a few low-rank components. If a substantial share of the variance is spread independently across all dimensions, these methods must retain many dimensions to accurately model the total variance.

Beyond dimensionality reduction, other methods also exploit the sparsity of information in the covariance matrix, e.g., sparse-precision methods[56,43,57], covariance tapering[58], and hierarchical matrix methods[59]. However, these methods have the caveat that they do not reduce the fundamental dimensionality of the problem, making it difficult to combine with the efficient sampling methods in regional-scale seismic risk analysis[60]. Furthermore, some of these methods are generally not applicable to the structure of $\mathbf{Cov}(\mathbf{g}_k)$ in our problem. For example, sparse approximations such as the Vecchia approximation or covariance tapering assume that distant sites become approximately conditionally independent given a few neighbors; however, the second term in Equation 16, $(\mathbf{B}\boldsymbol{\tau})(\mathbf{B}\boldsymbol{\tau})^{\mathsf{T}}$, correlates all assets simultaneously. Moreover, for the third term in Equation 16, $\mathbf{BFCFB}$., these methods may be less effective for densely populated building portfolios in a small area, because the model's correlation length is not sufficiently smaller than the spatial extent of the portfolio.

In this section, we first show how PPCA differs from PCA, a widely used dimensionality reduction method, and explain why PPCA is more appropriate (Section 4.1). Then, we further investigate the properties of $\mathbf{Cov}(\mathbf{g}_k)$ by analytically deriving its eigenvalues (Section 4.2) and examining the impact of portfolio distribution on them (Section 4.3. Finally, we formulate a PPCA $\mathbf{g}_k$ damage simulation based on these derivations and experiments (Section 4.4).

4.1 | PCA vs. PPCA

PCA is a widely used dimensionality reduction technique for multivariate Gaussian random variables[52,37]. The method approximates these variables as linear combinations of lower-dimensional principal components by exploiting their underlying linear correlations. These dependencies are encoded within the covariance matrix—specifically $\mathbf{Cov}(\mathbf{g}_k)$ in this study—whose eigenvalues are real and non-negative due to the matrix's positive semi-definite property. The eigenvector corresponding to the largest eigenvalue of $\mathbf{Cov}(\mathbf{g}_k)$ identifies the primary axis of linear dependency, with the eigenvalue quantifying the variance captured along this vector. Subsequent eigenvectors, ordered by their corresponding eigenvalues, identify directions of decreasing variance under the constraint of orthogonality to preceding components. Since the sum of the eigenvalues represents the variance within the space spanned by the corresponding eigenvectors, the number of components m required to capture a significant proportion (e.g., 90%) of the total variance in $\mathbf{g}_k$ is determined by:

$$F_m = \frac{\sum_{i=1}^{m} \lambda_i}{\sum_{j=1}^{N_b} \lambda_j} \times 100, \tag{17}$$

where F_m is the cumulative variance explained, expressed as a percentage. A reduction in the value of m necessary to reach a specific threshold indicates stronger correlations between the g_k values across various sites, thereby enhancing the efficiency of PCA representation.

PPCA extends the standard PCA framework[42]. This approach decomposes the covariance matrix $\mathbf{Cov}(\mathbf{g}_k)$ into a highly correlated component and an independent noise component, such that $\mathbf{Cov}(\mathbf{g}_k) = \mathbf{Cov}(\mathbf{g}_k)^{(cor)} + \mathbf{Cov}(\mathbf{g}_k)^{(ind)}$. Standard PCA is then applied to the correlated portion, $\mathbf{g}_k^{(cor)}$, to obtain the dimensionality reduced approximation of $\tilde{\mathbf{g}}_k^{(cor)}$. Consequently, $\mathbf{g}_k$

is represented by the low-dimensional $\tilde{\mathbf{g}}_k^{(cor)}$ and the full-dimensional $\mathbf{g}_k^{(ind)}$. Since standard PCA is still applied to $\tilde{\mathbf{g}}_k^{(cor)}$, the efficiency of PPCA heavily lies in how to decompose it among infinite possible combinations to maximize the linear correlations of variables in $\mathbf{Cov}(\mathbf{g}_k)^{(cor)}$. The more concentrated $\mathbf{g}_k^{(cor)}$'s variance is in a lower-dimensional subspace, the more efficient PPCA becomes.

$\mathbf{Cov}(\mathbf{g}_k)$ is well suited to PPCA since its structure already admits a natural decomposition of this form. The second term in Equation 16, $(\mathbf{B}\boldsymbol{\tau})(\mathbf{B}\boldsymbol{\tau})^\mathsf{T}$, represents a rank-1, globally and fully correlated field, which can be embedded in $\mathbf{Cov}(\mathbf{g}_k)^{(cor)}$. Also, the first term , $\mathbf{I}$, exhibits isotropic independent noise, which can be cleanly separated into $\mathbf{Cov}(\mathbf{g}_k)^{(ind)}$. In addition, the third term, **BFCFB**, whose correlation structure arises from within-event correlation, has a nugget effect, where the correlation between nearby sites does not converge to 1, which can be included in $\mathbf{Cov}(\mathbf{g}_k)^{(ind)}$. On the other hand, because of the independent components spread across all dimensions in the structure of $\mathbf{Cov}(\mathbf{g}_k)$ – which PCA cannot capture using a low-dimensional subspace – standard PCA is likely to be inefficient. For a more detailed comparison between standard PCA and PPCA using illustrative examples, see Section S3 and Figures S3 and S4 of the Supplementary Material. In the following sections (Sections 4.2 and 4.3), we further investigate this property of $\mathbf{Cov}(\mathbf{g}_k)$ for the formulation of PPCA-based simulation.

4.2 | Analytical Derivation of Eigenvalues of $\mathbf{Cov}(\mathbf{g}_k)$

To determine the optimal decomposition of the covariance matrix $\mathbf{Cov}(\mathbf{g}_k)$ for PPCA, we examine its eigenvalue properties. We focus primarily on the lower-order eigenvalues of $\mathbf{Cov}(\mathbf{g}_k)$, which are instrumental in identifying independent noise components and isolating the highly correlated structure. We first present an analytical derivation of the lower bound eigenvalues for $\mathbf{Cov}(\mathbf{g}_k)$ using an idealized spatial distribution that represents a limiting case in regional risk analysis. Subsequently, we investigate more realistic scenarios through illustrative examples across various spatial configurations. These cases serve to validate the analytical derivation and demonstrate how the eigenvalues evolve in response to diverse building distributions.

The theoretical analysis focuses on a limiting geometric configuration designed to facilitate a formal derivation: t building clusters separated by large spatial distances, with each cluster containing densely packed buildings. A building cluster represents a group of buildings in close spatial proximity. Each cluster i ($i = 1, 2, \ldots, t$) contains n_i buildings, satisfying $\sum_{i=1}^{t} n_i = N$. The correlation between buildings belonging to distinct clusters is assumed to be zero. Within each cluster, all the inter-building within-event ground shaking correlation is assumed to be a constant $\rho_{\max} < 1$. This value represents the maximum achievable inter-site correlation, accounting for the nugget effect[14].

Under these idealized conditions, the lower bound eigenvalues of $\mathbf{Cov}(\mathbf{g}_k)$ are derived as $1 + \min_i \left(\frac{\phi_i^2}{\beta_i^2}\right)(1 - \rho_{\max})$ (see Section S4 of Supplementary Material). In the specific case where the within-event standard deviations (ϕ) and fragility curve dispersion parameter (β) are constant across all sites, there are exactly $(N - t)$ minimum eigenvalues equal to:

$$\lambda_{\min} = 1 + \frac{\phi^2}{\beta^2}(1 - \rho_{\max}). \tag{18}$$

Therefore, there are t eigenvalues that strictly exceed this minimum value, where t is the assumed number of building clusters.

4.3 | Impact of Building Distribution on the Eigenvalues of $\mathbf{Cov}(\mathbf{g}_k)$

Building upon this analytical derivation, we investigate how the eigenvalues evolve under more realistic spatial configurations. More realistic scenarios are constructed by relaxing the idealized constraints in two primary ways. First, the inter-cluster correlation constraint is loosened; by adjusting the spatial separation between building clusters, we generate geometries with various inter-cluster correlations. (Figure 4 (a)). These configurations are categorized as close, intermediate, or distant, corresponding to within-event ground shaking intensity correlations between cluster centers of 0.8, 0.38, and 0, respectively. Second, the intra-cluster correlation constraint is relaxed; by increasing the spatial extent of a cluster, the within-event correlations between buildings inside that cluster may fall below $\rho_{\max}$ (Figure 4 (b)). These cases are categorized by building density as high, moderate, or low, corresponding to average within-event correlations between buildings within the cluster of 0.8, 0.23, and 0.002, respectively. Together, these modifications represent a broad spectrum of spatial distributions encountered in practice. We demonstrate these effects using a system of 15 buildings organized into three clusters ($t = 3$), subjected to varying inter-cluster distances and within-cluster densities.

Figure 4 (c) shows the eigenvalue spectrum change when the inter-cluster distance is changed. First, the minimum eigenvalues of 1.29, which precisely match with the theoretical prediction of Equation 18, remain invariant with respect to the change in inter-cluster distances. Furthermore, the number of eigenvalues exceeding this minimum is exactly $t = 3$, the number of building clusters. The primary variation occurs in the relative magnitudes; as inter-cluster distances decrease, the second and third eigenvalues converge toward the theoretically derived minimum value. This behavior is consistent with the physical interpretation that as clusters move closer, they begin to behave as a single aggregate cluster rather than three distinct entities, effectively reducing the number of eigenvalues greater than the theoretically derived minimum toward one.

Figure 4 (d) illustrates the change in the eigenvalue spectrum as the building density within each cluster varies. These results demonstrate that within-cluster building density influences the minimum eigenvalues of the covariance matrix. The high-density case shown in Figure 4(b) exhibits distinctive three (t) non-minimum eigenvalues and 12 ($= N - t$) minimum eigenvalues. However, as the spatial extent of each cluster increases—thereby reducing within-event correlations below $\rho_{\max}$—the eigenvalues no longer maintain this constant minimum value. This transition results in a gradual decay as observed in the "moderate" case. It decreases the first three eigenvalues, while the trailing eigenvalues increase. We also found that the minimum eigenvalue in this case closely approximates the value predicted by Equation 18 when $\rho_{\max}$ is replaced by the observed maximum correlation $\hat{\rho}_{\max}$ for the specific building geometry: $\lambda_{\min} \approx 1 + \frac{\phi^2}{\beta^2}(1 - \hat{\rho}_{\max})$.

When the cluster size increases to the point that intra-cluster correlations approach zero (the "low density" case in Figure 4(b) and (d)), the spectrum reveals 14 nearly homogeneous minimum eigenvalues, with only a single eigenvalue exceeding this threshold. This behavior mirrors the "Close cluster" case shown in Figures 4(a) and (c), which also functions as a single cluster. In this state, the system effectively behaves as a single large cluster characterized by negligible spatial correlation across all buildings. Consequently, the number of building clusters t within our Gaussianized formulation is defined in a relative sense, rather than by absolute physical distances. The primary distinction lies in the converged eigenvalue levels and the relative magnitude between the dominant and secondary eigenvalues. The converged eigenvalue in this case also follows Equation 18 with $\hat{\rho}_{\max} = 0$.

Across all examined cases, it is evident that a substantial number of principal components are required to explain a significant fraction of the total variance in $\mathbf{g}_k$. In the scenarios presented in Figure 4 (a), all three configurations require 12 out of 15 components to capture 90% of the variance (Figure 4 (e)). Similarly, in Figure 4 (b), all three cases require 13 out of 15 components to reach the same 90% threshold (Figure 4 (f)). Consequently, standard PCA yields a dimensionality reduction of only 13% to 20%, consistent with previous findings. However, having established the convergence behavior of these eigenvalues and close correlation between the first few dominant eigenvalue to the number of building clusters t, we can now leverage these insights to formulate a PPCA framework for the Gaussianized damage function $\mathbf{g}_k$. This approach allows us to circumvent the computationally expensive full eigendecomposition of $\mathbf{Cov}(\mathbf{g}_k)$.

4.4 | Decomposition of $\mathbf{Cov}(\mathbf{g}_k)$ for PPCA

We adopt PPCA because the structure of $\mathbf{Cov}(\mathbf{g}_k)$ is appropriate for PPCA. This strategy reformulates the generative model of $\mathbf{g}_k$ by extracting the highly correlated component and isolating the independent component. This separation allows us to apply dimensionality reduction to the correlated component efficiently and facilitates fast simulation of the independent component.

4.4.1 | Decomposition of $\mathbf{g}_k$

Instead of Equation 13, which simulates $\mathbf{g}_k$ using $N \times (2N + 1)$ matrix-vector multiplication, $\mathbf{Az}$, we rewrite the generative model of $\mathbf{g}_k$ as:

$$\mathbf{g}_k = \mathbf{W}\mathbf{x} + c\mathbf{z} + \mathbf{b}_k, \tag{19}$$

where $\mathbf{W} \in \mathbb{R}^{N\times N}$, c is a scalar constant, and $\mathbf{x}, \mathbf{z} \in \mathbb{R}^N$ are independent standard normal random vectors. To ensure this formulation reproduces the statistics of the original model (Equation 13), the covariance of the right-hand side of Equation 19 must match $\mathbf{Cov}(\mathbf{g}_k)$ (Equation 16). This condition requires (see Section S5 of the Supplementary Material for derivation):

$$\mathbf{W} = \mathbf{U}\boldsymbol{\Sigma} \tag{20}$$

where $\boldsymbol{\Sigma} = (\boldsymbol{\Lambda} - c^2\mathbf{I})^{1/2}$. Here, $\boldsymbol{\Lambda} \in \mathbb{R}^{N\times N}$ is a diagonal matrix containing the eigenvalues of $\mathbf{Cov}(\mathbf{g}_k)$ ordered in descending magnitude, $\mathbf{I} \in \mathbb{R}^{N\times N}$ is the identity matrix, and $\mathbf{U} \in \mathbb{R}^{N\times N} = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_N]$ is the matrix of eigenvectors corresponding

to $\mathbf{\Lambda}$ (= diag($[\sqrt{s_1}, \sqrt{s_2}, \cdots, \sqrt{s_N}]$)), where $s_i = \lambda_i - c^2$. The λ_i are the eigenvalues of $\mathbf{Cov}(\mathbf{g}_k)$ and $\mathbf{u}_i$ are the corresponding eigenvectors.

4.4.2 | Selection of $\mathbf{\Sigma}$ and c

While infinite pairs of $\mathbf{W}$ and c can satisfy Equation 19, our goal is to minimize computational cost of $\mathbf{g}_k$ while maintaining its accuracy. The primary burden is the calculation of the first term $\boldsymbol{\omega} = \mathbf{Wx}$, which scales with $O(N^2)$, while the remaining terms ($c\mathbf{z}$ and $\mathbf{b}_k$) scale with $O(N)$. Therefore, we aim to select an optimal constant c and corresponding $\mathbf{\Sigma}$ that allows reducing the effective number of columns of matrix $\mathbf{W}$ from N to a smaller dimension.

To do that, we desire the covariance of $\boldsymbol{\omega}$ to have many zero eigenvalues and only a few dominant ones to effectively reduce dimensionality. The covariance of $\boldsymbol{\omega}$ is given by:

$$\mathbf{Cov}(\boldsymbol{\omega}) = \mathbf{W}\mathbf{W}^{\mathsf{T}} = \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^{\mathsf{T}}, \tag{21}$$

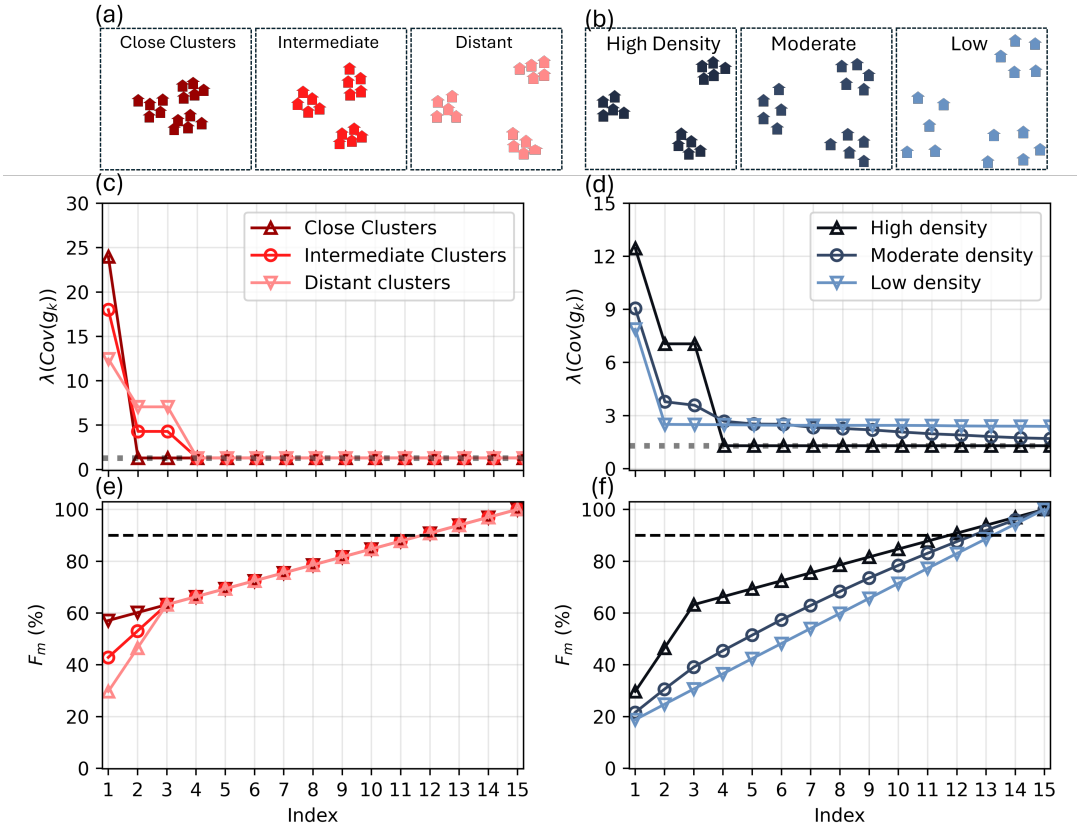

FIGURE 4 Eigenvalue analysis of building configurations across varying spatial distributions. (a) Configurations with varying inter-cluster distances, categorized as Close (dark red), Intermediate (red), and Distant (light red) clusters, corresponding to within-event ground shaking intensity correlations between cluster centers of 0.8, 0.38, and 0, respectively. (b) Configurations with varying cluster sizes, categorized as High (dark blue), Moderate (blue), and Low (light blue) density, corresponding to average within-event correlations between buildings within the cluster of 0.8, 0.23, and 0.002, respectively. (c) and (d): the eigenvalues for the configurations in (a) and (b), respectively; the theoretically driven lower bound is indicated by the gray dotted horizontal lines. (e) and (f): the cumulative variance contribution (F_m) for the configurations in (a) and (b), respectively, with the dashed black line denoting the 90% threshold.

where $\mathbf{U}^{\mathsf{T}}$ is the transpose of $\mathbf{U}$. Note that the constant c is selected as the square root of the lower bound eigenvalue of $\mathbf{Cov}(\mathbf{g}_k)$ (Equation 18 , with $\rho_{\max}$ replaced by the observed maximum correlation coefficient $\hat{\rho}_{\max}$). This choice ensures that s_i values for $i > t$ remain close to zero (See Section S6 of the Supplementary Material for derivation):

$$c^* = \sqrt{1 + (1 - \hat{\rho}_{\max})\frac{\phi^2}{\beta^2}}. \tag{22}$$

We substitute this value into Equation 20. Then, we also define the corresponding operator $\mathbf{W}^*$ as $\mathbf{U}\mathbf{\Sigma}^*$, where $\mathbf{\Sigma}^* = (\mathbf{\Lambda} - {c^*}^2\mathbf{I})^{1/2}$ and $s_i^* = \lambda_i - (c^*)^2$. The Equation 19 can then be re-written as $\mathbf{g}_k = \mathbf{W}^*\mathbf{x} + c^*\mathbf{z} + \mathbf{b}_k$.

The total variance of $\mathbf{g}_k$ explained by the first m largest principal components of $\mathbf{Cov}(\boldsymbol{\omega}^*)$ is defined as:

$$F_m = \frac{Nc^* + \sum_{i=1}^{m} \lambda_i}{Nc^* + \sum_{j=1}^{N_b} \lambda_j} \times 100, \tag{23}$$

where Nc^* represents the variance contribution from the independent noise component. Because Nc^* acts as a non-zero constant contributor to the total variance regardless of the chosen m, and since the high correlation within $\boldsymbol{\omega}^*$ allows for a minimal m to explain the majority of the variance in $\mathbf{g}_k$, this formulation is expected to reach the variance contribution threshold more efficiently than standard PCA.

Figure 5 illustrates the impact of applying the optimal c^* to the eigenvalues of $\mathbf{Cov}(\boldsymbol{\omega})(i.e., \mathbf{Cov}(\boldsymbol{\omega}^*))$ for the building configurations previously analyzed in Figure 4, along with the resulting cumulative variance contribution to $\mathbf{g}_k$ (Equation 23). Figure 5(a) displays the eigenvalue spectra of $\mathbf{Cov}(\boldsymbol{\omega}^*)$ for the building configurations shown in Figure 4(a). In contrast to the original covariance $\mathbf{Cov}(\mathbf{g}_k)$, $\mathbf{Cov}(\boldsymbol{\omega}^*)$ possesses zero eigenvalues, as seen by comparing Figure 5(a) to Figure 4(c). Furthermore, the number of non-zero eigenvalues is identical to the number of building clusters, which is $t = 3$ in this case. If the density of buildings within each cluster decreases (e.g., the moderate density case in Figure 4(b) and Figure 5(b)), several trailing components remain non-zero, although they converge toward zero. In cases where the building density within a cluster is low (Figure 5(b)), the spectra contain many zeros with only one non-zero eigenvalue because the framework treats them as a single cluster with no ground-shaking correlation.

The cumulative variance contribution shown in Figures 5(c) and (d) reaches the threshold far more rapidly than the case where independent noise components are not extracted from $\mathbf{Cov}(\mathbf{g}_k)$. In all instances, the variance contribution exceeds 90% within only three principal components, whereas 12 or 13 components were previously required (Figures 4(e) and (f)). Interestingly, the least favorable scenario for this PPCA framework occurs when the building distribution is neither highly dense nor highly sparse. A "moderate density" distribution results in the slowest convergence of the total variance contribution (intermediate blue curve of Figure 5(d)). This indicates that the framework is highly efficient for dense metropolitan areas. For cases with moderate ground-motion correlations, the approach remains efficient, though the efficiency gains are less pronounced than in dense urban settings. For independently located buildings, the framework is also efficient as the system can be modeled with a single principal component. However, the ground-shaking and damage modeling for independent buildings is already computationally manageable, as the complexity scales with $O(N)$ when correlations are neglected.

Since the diagonal entries of $\mathbf{\Sigma}^*$ become zero for all indices greater than t, we can effectively truncate $\mathbf{W}^*$ without loss of information. The reduced matrix $\mathbf{W}_t^*$ is defined as:

$$\mathbf{W}_t^* = \mathbf{U}_t\mathbf{\Sigma}_t, \tag{24}$$

where $\mathbf{W}_t^* (\in \mathbb{R}^{N\times t})$ is $[\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_t]$ and $\mathbf{\Sigma}_t (\in \mathbb{R}^{t\times t})$ is $\mathrm{diag}([\sqrt{s_1^*}, \sqrt{s_2^*}, \cdots, \sqrt{s_t^*}])$. Using this reduced representation, Equation 19 is approximated as:

$$\mathbf{g}_k \approx \mathbf{W}_t^*\mathbf{x}_t + c^*\mathbf{z} + \mathbf{b}_k, \tag{25}$$

where $\mathbf{x}_t \in \mathbb{R}^t$ and $\mathbf{z} \in \mathbb{R}^N$ are independent standard normal random vectors. Equation 25 constitutes our proposed computational framework for regional damage modeling. It integrates the Gaussianized damage formulation and the dimensionality reduction using PPCA, reducing the computational complexity of the matrix-vector product from $O(N^2)$ to $O(tN)$—a significant efficiency gain when $t \ll N$.

5 | PROPOSED COMPUTATIONAL FRAMEWORK

A schematic comparison of our proposed framework in comparison with the traditional framework is presented in Figure 6. The procedures are decomposed into two: 1) pre-processing step and 2) simulation step. Pre-processing is executed once for a given portfolio and does not repeat over MC realizations. In contrast, the simulation step includes operations repeated for each MC realization (M times).

In the pre-processing step, for both traditional and proposed frameworks, we need to specify the parameters of the buildings' fragility curves ($\boldsymbol{\beta}_k$ and $\boldsymbol{\theta}_k$), the ground motion between- and within-event standard deviation models ($\boldsymbol{\tau}$ and $\boldsymbol{\phi}$), and the within-event ground shaking correlation matrix $\mathbf{C}$. Afterward, in the traditional framework, one must perform the Cholesky decomposition of $\mathbf{C}$ to create $\mathbf{L}$, which is used to simulate the spatially correlated ground shaking. However, in the proposed framework, first we construct the covariance matrix $\mathbf{g}_k$, or $\mathbf{A}\mathbf{A}^{\mathsf{T}}$, using Equation 16. Note that the Cholesky decomposition is not required in our framework. Then, $\mathbf{W}_t^*$ and c^* are computed using Equations 22 and 24.

In the simulation step, in both frameworks, mean ground shaking intensities ($\boldsymbol{\mu}$) at the buildings are calculated using empirical ground motion models for the earthquake scenario. This step is performed once for a scenario-based risk model. For an event-based model (random rupture magnitude and location), this process is repeated inside the MC loop (red dotted box in Figure 6). Next, in the traditional framework, the calculated $\boldsymbol{\mu}$ is combined with the standard deviation models and $\mathbf{L}$ to simulate the random ground shaking vector ($\ln \mathbf{y}$). In the proposed framework, however, $\boldsymbol{\mu}$ is not used at this stage. Instead, the matrix $\mathbf{W}_t^*$ and c^* are used to calculate the zero-mean component of $\mathbf{g}_k$ (denoted as $\boldsymbol{\gamma}$ in Figure 6), which is the part of Equation 25 before adding the mean $\mathbf{b}_k$. For the damage simulation, the traditional framework requires N uniform samples ($\mathbf{u}$) per MC realization and the calculation of the standard normal CDF ($\Phi(\cdot)$). The proposed framework, at this point, adds the mean-shifting term $\mathbf{b}_k$ (which depends on $\boldsymbol{\mu}$) to $\boldsymbol{\gamma}$ to get the final $\mathbf{g}_k$. Both frameworks require a repetitive loop over the number of categorical damage states (n_{ds}). Given $\mathbf{G}_k$ or $\mathbf{g}_k$, the damage state is determined using Equation 5 or 11, and the corresponding loss is also calculated using Equation 6.

Table 1 summarizes the computational complexities of both the traditional and proposed frameworks for regional scenario-based seismic risk models. In the traditional framework, the pre-processing bottleneck lies in the Cholesky decomposition of

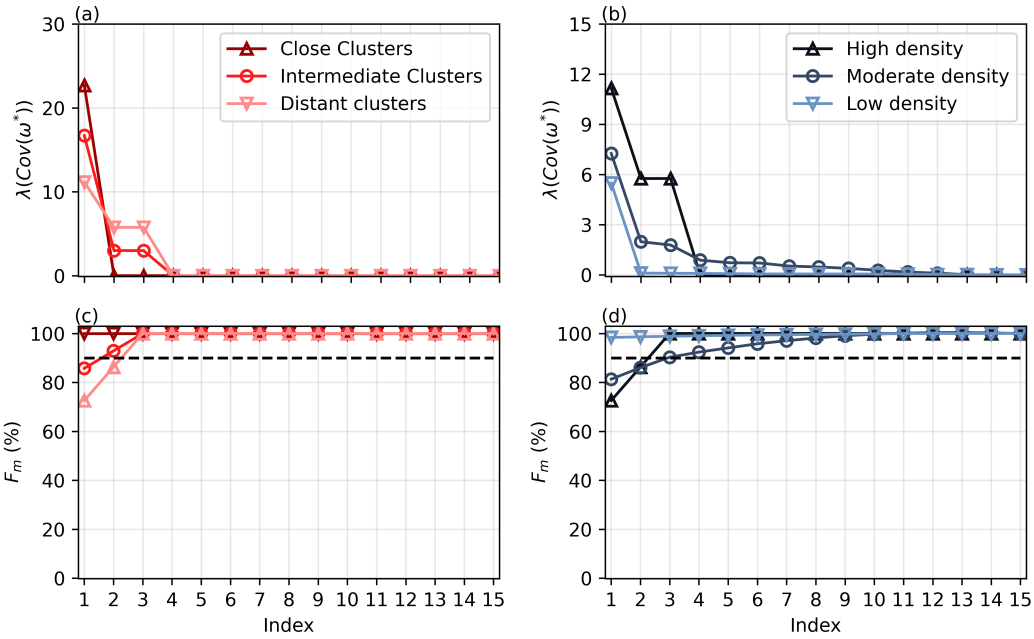

FIGURE 5 Eigenvalue analysis of $\mathbf{Cov}(\boldsymbol{\omega}^*)$ for the spatial configurations depicted in Figure 4(a) and (b): (a) Eigenvalue spectra corresponding to the cases in Figure 4(a); (b) Eigenvalue spectra corresponding to the cases in Figure 4(b); (c) and (d) Cumulative variance explained (F_m) as a function of the number of retained principal components. The horizontal dashed line denotes the 90% threshold.

$\mathbf{C}$ to construct the lower triangular matrix $\mathbf{L}$, which entails a computational complexity of $O(N^3)$. Conversely, the proposed framework bypasses the Cholesky decomposition during pre-processing. Instead, it requires the construction of $\mathbf{W}_t^*$ via the eigen-decomposition of $\mathbf{Cov}(\mathbf{g}_k)$. While a full decomposition scales as $O(N^3)$, computing only the top t eigenpairs reduces the complexity to $O(N^2 t)$, which is effectively $O(N^2)$ when $t \ll N$. Regarding the simulation step, the primary computational bottleneck in the traditional framework is ground-motion generation, which scales with $O(N^2 M)$ for large values of N and M. In contrast, the dominant procedure in the proposed framework is the simulation of $\mathbf{g}_k$, which scales with $O(tNM)$, or effectively $O(NM)$ when $t \ll N$. Overall, the proposed framework reduces the computational complexity by one order of magnitude: from $O(N^3)$ to effectively $O(N^2)$ in the pre-processing stage, and from $O(N^2 M)$ to effectively $O(NM)$ in the simulation stage.

For a detailed implementation of the algorithm and an illustrative case study, see Section S7, Figure S5 and Table S1 of the Supplementary Material. A more comprehensive discussion of the computational complexity is also provided in Section S8 of the Supplementary Material. In the subsequent sections, we provide a numerical demonstration of the proposed framework for a building portfolio in the San Francisco Bay Area.

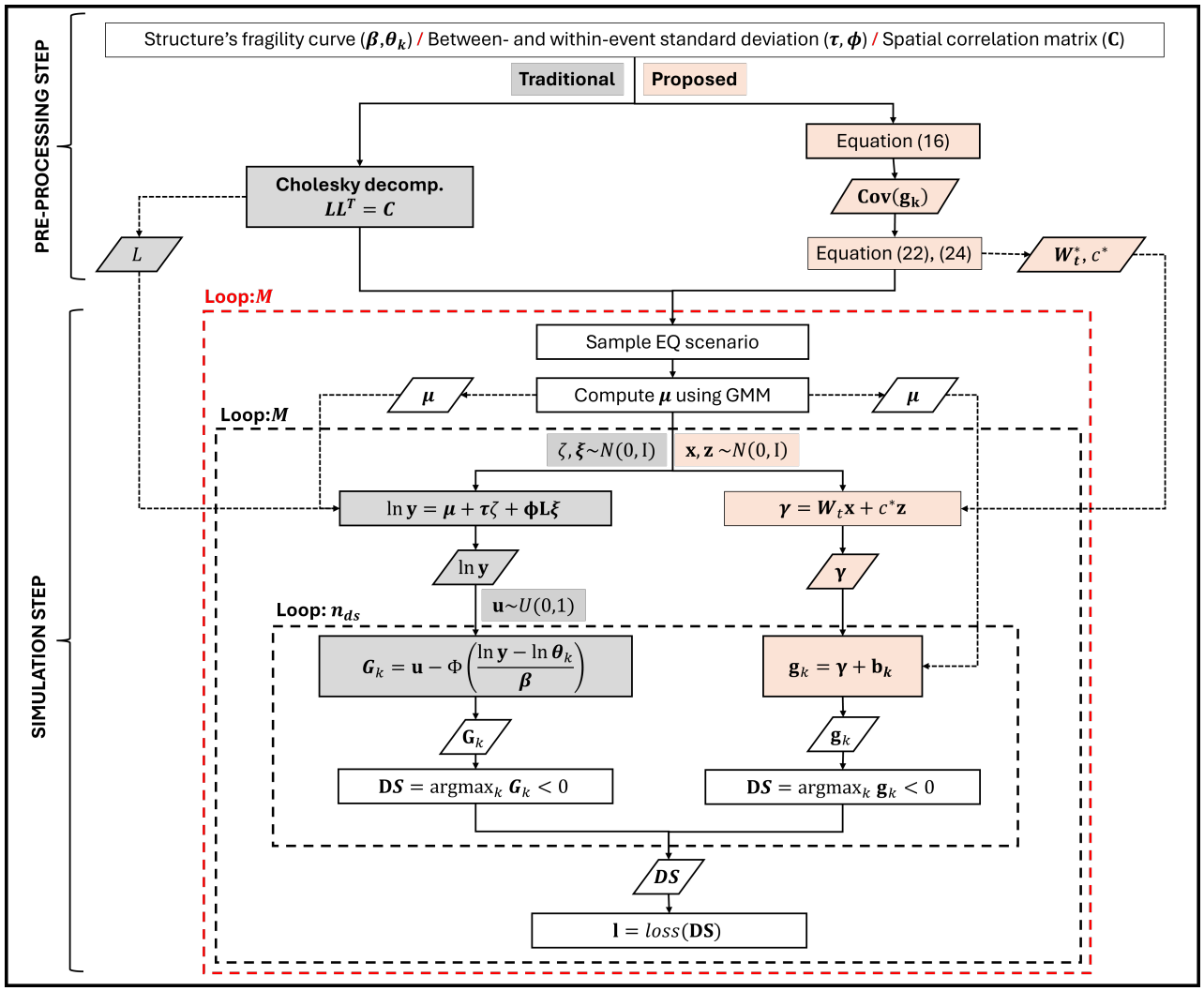

FIGURE 6 Schematic diagram comparing the traditional and the proposed frameworks for regional-scale seismic risk analysis. The steps for both traditional and proposed frameworks are white boxes; unique processes in the traditional framework are filled with light gray, while unique steps in the proposed framework are filled with light orange. The loops are presented as dotted line boxes. The red dotted loop over the M box corresponds to event-based (random rupture magnitude and location) risk modeling, while the black loop corresponds to scenario-based (specified rupture magnitude and location) risk modeling.

TABLE 1 Computational complexity of our proposed framework in comparison with the traditional method for regional scenario-based seismic risk model. N denotes the number of buildings, and M denotes the number of MC simulations

Traditional		**Proposed**	
Pre-processing		**Pre-processing**	
Calc. $\mathbf{C}$	$O(N^2)$	Calc. $\mathbf{C}$	$O(N^2)$
Calc. $\mathbf{L}$	$O(N^3)$	Construct $\mathbf{Cov}(\mathbf{g}_k)$	$O(N^2)$
		Construct $\mathbf{W}_t^*$	$O(tN^2)$
Simulation		**Simulation**	
Calc. $\boldsymbol{\mu}$	$O(N)$	Calc. $\boldsymbol{\mu}$	$O(N)$
Sampling ζ, $\boldsymbol{\xi}$, $\mathbf{u}$	$O(NM)$	Sampling $\mathbf{x}$, $\mathbf{z}$	$O((N+t)M)$
GM simulation (Equation 2)	$O(N^2M)$	$\mathbf{g}_k$ simulation	$O(tNM)$
DS simulation (Equation 4)	$O(NM)$		

6 | TESTBEDS: FROM SAN FRANCISCO DOWNTOWN TO BAY AREA

We conduct empirical analyses on our proposed framework using San Francisco (SF) Downtown and Bay Area building portfolios. In Example 1, we demonstrate the accuracy of the framework using 1,000 buildings in SF downtown. In Example 2, we examine t, the building cluster parameter, to keep high accuracy for different region extents and building portfolio sizes, using cases for SF downtown and the whole Bay Area comprising 15,836 buildings (Figure 8). In Example 3, we study our framework's computational efficiency in the northern SF peninsula region (Figure 10(a)).

The earthquake rupture scenario for all the examples is a magnitude 7.2 event along the northern segment of the vertically dipping San Andreas Fault (Figure 8 (b)). The rupture is assumed to span 50 km in length and 10 km in width, extending from the west coast of San Francisco to the southern Peninsula. The mean, between-event, and within-event standard deviations are computed using the ground-motion model developed for shallow crustal earthquakes in California[61]. The within-event spatial correlation is computed using the model developed based on extensive crustal earthquake records[15]. The building inventory was obtained from the SimCenter Earthquake Testbed[62], and the structural-type–dependent fragility functions were sourced from Hazus 6.1[21]. These are lognormal fragility functions comprising four limit states and five damage states (No Damage, Slight, Moderate, Extensive, and Complete) that take Peak ground acceleration (PGA) as input.

6.1 | Example 1: Accuracy for San Francisco Downtown

We compared the results of our framework against a benchmark generated using the R2D software[63,64]. The benchmark utilizes the building portfolio from R2D Example 1 – Basic HAZUS, which performs analysis via HAZUS loss assessment (the traditional framework). From the dataset containing 15,836 buildings in downtown San Francisco, a subset of 1,000 buildings was selected to avoid excessive computational overhead when running the R2D software on a desktop-level machine (Figure 7(a)). This subset captures the full spectrum of characteristics found in the original dataset, including spatial locations, design levels, and structural types. The selected portfolio consists of 544 Pre-Code, 35 Low-Code, 322 Moderate-Code, and 99 High-Code buildings. In terms of structure, the set includes 267 wood-frame, 217 steel, 321 concrete, 194 masonry, and one unreinforced masonry building. Damage for each building was simulated over 10,000 MC realizations to ensure accuracy.

We executed both the traditional and the proposed approaches, demonstrating that $t = 1$ is sufficient to identify the damage-state distribution with high accuracy. Figure 7(b) illustrates a representative damage-state histogram for Building X (identified in Figure 7(a)). While the R2D simulation identifies the modal damage as "Extensive" and the proposed method yields "Complete," the overall distribution are nearly equal. The modal damage states for all 1,000 buildings were compared with the R2D simulation results in Figure 7(c) and (e). The two methods show strong agreement across whole buildings: 952 buildings match exactly, while the remaining 48 buildings exhibit a difference of only one damage-state level. For the buildings exhibiting a one-level difference, their distributions still closely resemble each other, as shown in Figure 7(b). The difference in the mean damage state across all 1,000 buildings—which reflects the overall damage distribution more accurately than the mode—ranges from -0.036 to 0.039 (Figure 7(d)). This further confirms that the proposed framework closely approximates the traditional distribution and the discrepancies between the two methods may predominantly originate from the inherent stochasticity of MC simulations rather than inaccuracies in our framework.

6.2 | Example 2: Loss Estimation by Spatial Extent

The proposed framework is applied to loss estimation for two building portfolios with distinct spatial extents to investigate how the building cluster parameter, t, scales with geographic spread: 1) San Francisco Downtown, comprising 15,836 buildings (representing a dense urban area); and 2) the entire Bay Area, comprising 15,836 buildings randomly sampled from a total inventory of approximately 1.3 million structures (representing a metropolitan-scale simulation; Figure 8). By limiting the sample size, the computation of an accurate benchmark loss curve via the traditional framework—utilizing 15,836 distinct building-specific ground motions—remains feasible on a standard workstation without compromising accuracy, e.g. though coarse grid[34,17,65].

The loss curves for the SF Downtown and Bay Area cases are presented in Figures 9(a) and (b), respectively. For both cases, the traditional and proposed framework curves are generated using 100,000 MC realizations. This sample size results in a coefficient of variation (COV) for the MC loss estimation of approximately 1%. Consequently, if the loss curve produced by the proposed framework remains within $\sim 2.5\%$ error margin—corresponding to an approximate 99% confidence interval—the framework is considered to represent the benchmark accurately, as the deviation falls within the range of MC estimation error. The error is calculated as $e_t(p) = \frac{l_t(p) - l(p)}{l(p)} \times 100$, where $l(p)$ and $l_t(p)$ are the losses corresponding to exceedance probability p from the traditional and proposed frameworks, respectively.

In the San Francisco Downtown case (Figure 9(a)), the proposed framework yields a close approximation of the benchmark when $t = 1$. Within the exceedance probability range down to 10^{-3}, the error fluctuates between -0.0% and 1.4% (blue solid

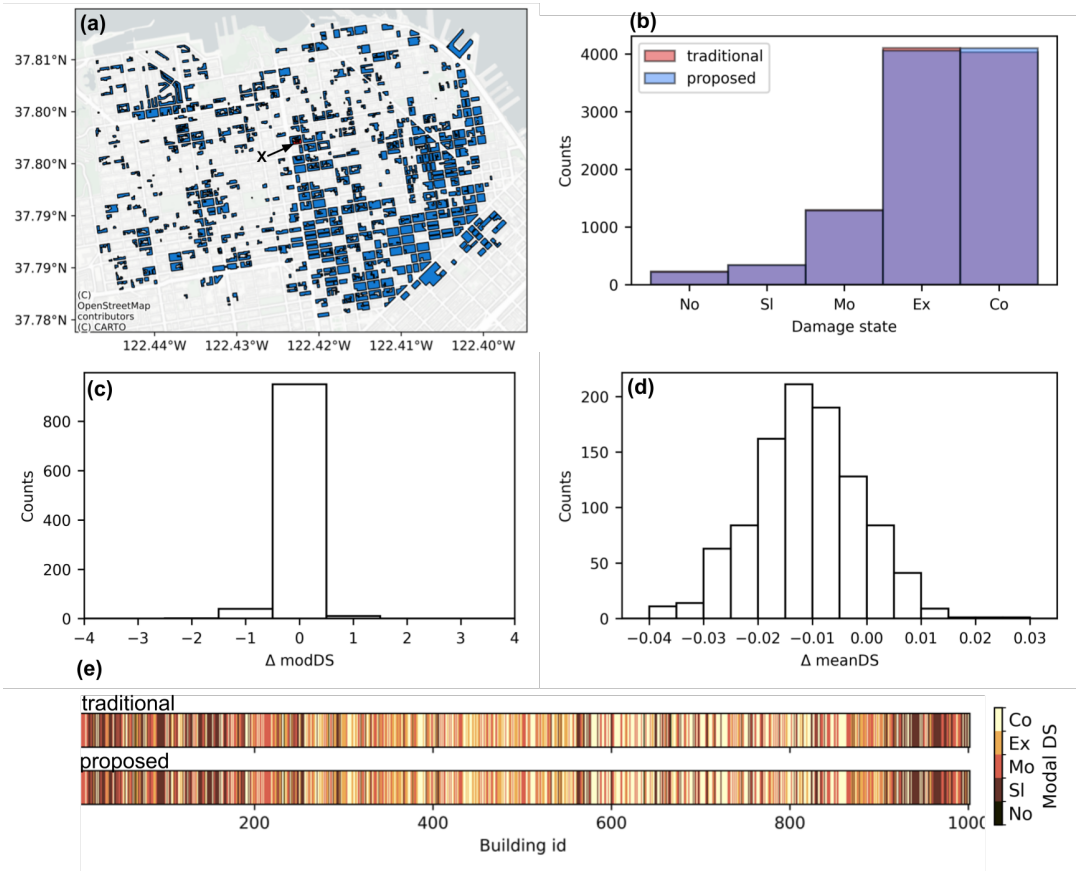

FIGURE 7 (a) Spatial distribution of the building portfolio for Example 1. (b) Comparative damage-state histogram for Building X from panel (a), illustrating a case with a one-step discrepancy in the modal damage state between the two methods. Histogram of (c) modal and (d) mean damage state differences between the R2D baseline and the proposed framework. (e) Comparison of modal damage states simulated using the R2D software (traditional) and the proposed framework ($t = 1$) based on 10,000 MC simulations.

curve, Figure 9(c)). Notably, increasing t beyond 1 does not yield a significant improvement in model fit, as evidenced by the blue solid and dashed curves in Figure 9(c).

Conversely, the Bay Area case exhibits larger discrepancies at $t = 1$, with errors ranging from 2.3% to 9.9% (lightest red dotted curve of Figure 9 (d)). The error decreases as t increases; specifically, at $t = 20$, the error falls within the target 2.5% threshold in general, ranging from 1.2% to 2.9% (red solid curve). Further increases in t lead to asymptotic convergence toward zero, for example for $t = 200$, the error ranges from 0.1% to 1.5%. However, as t increases, the rate of improvement diminishes (Figure 9(d)).

While numerical simulations demonstrate that the proposed framework significantly reduces dimensionality, selecting an optimal value of t is key. We found that the optimal t is invariant regardless of the total building count, provided the spatial extent of the portfolio remains consistent (Section S9 of the Supplementary Material). Thus, we suggest determining t from small, randomly selected subset of buildings that spans the entire region. This is computationally inexpensive due to the small N: 1) Conduct a traditional regional risk analysis using a small, randomly selected subset of buildings that spans the entire region, in order to obtain the benchmark loss curve for the subset. In our experiments, a subset of 100 buildings works well for both larger (the entire SF Bay Area) and smaller (downtown SF) metropolitan areas. However, note that this number may be the minimum requirement for the entire Bay Area, whereas the minimum requirement for the downtown SF case is much smaller. Figure S6 in Section S9 of the Supplementary Material shows how consistent the variance contribution of t is across subsets of various sizes. For applying the method to the user's region of interest, we recommend a subset size of 100 for regions smaller than the entire Bay Area. If the user's region is larger than the extent of the entire Bay Area, users can produce a plot similar to Figure S6(c) with a gradually increasing t, until the variance-contribution curve converges; 2) Perform the risk analysis with the proposed framework on the same subset for various values of t; 3) Determine the optimal t by choosing the smallest t that shows sufficiently good agreement with the benchmark loss curve from step 1; 4) Apply the determined t to the full-scale analysis of the entire building portfolio.

Alternatively, users may select the value of t for which the variance contribution (Equation 23) exceeds a chosen threshold (e.g., 95%). However, for loss calculation, it is not known a priori what level of variance contribution is sufficient, although a larger t from the bigger threshold —and hence a higher variance contribution— improves accuracy. Therefore, for optimal selection of t, we recommend the subset method; that said, if the t required to reach a sufficiently large threshold (e.g., 95%) is not excessively large, this threshold-based criterion can also be used as a simpler alternative.

In addition, we observed that the eigenvectors of $\mathbf{Cov}(\boldsymbol{\omega}^*)$ emerge as spatial harmonics (Figure S7). As the eigenvalues decrease, the spatial frequency of these modes increases, with the order of their magnitudes being directly related to the spatial extent of the building distribution, higher order in the direction. The dominant component corresponds to the direction along

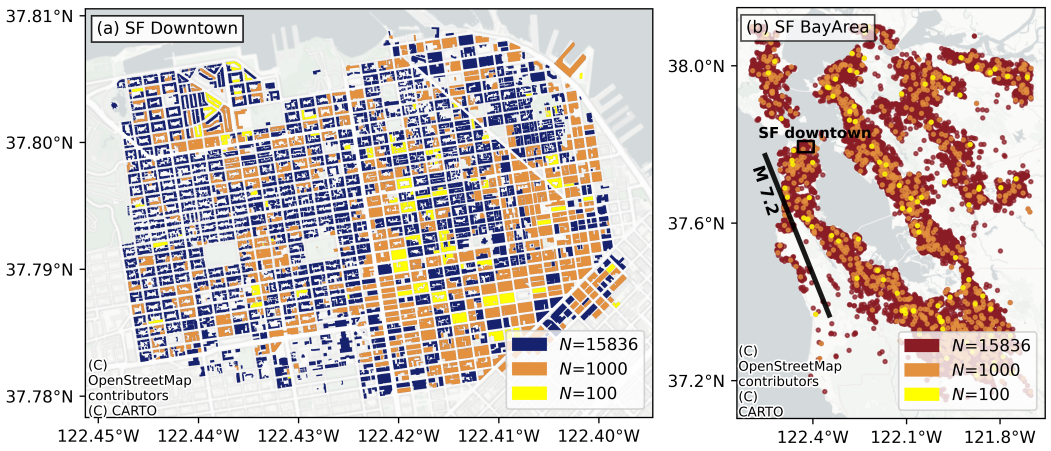

FIGURE 8 Spatial distribution of building portfolios for Example 2: (a) Downtown San Francisco and (b) the broader Bay Area. An assumed M 7.2 rupture is indicated by a thick solid black line, and SF downtown area in panel (a) is denoted by a small black square. The combined yellow, orange, and blue (or red) markers represent the full portfolio of 15,836 buildings; the orange markers denote a selected subset of 1,000 buildings; and the yellow markers specifically highlight a further subset of 100 selected buildings.

which the building distribution is most elongated. Furthermore, the product of each eigenvalue and its corresponding eigenvector illustrates how the magnitude of each eigenvector component diminishes as one moves toward the minor components. For a more detailed analysis of these spatial modes, please refer to Section S10 of the Supplementary Material.

6.3 | Example 3: Gains in Computational Efficiency

To demonstrate the computational efficiency of the proposed framework, we consider the Northern San Francisco Peninsula, which contains more buildings than the downtown area (Figure 10(a)). The first baseline is the traditional two-step approach, in which a ground-shaking map is generated and damage is then modeled from fragility curves. To this baseline, we also applied two computational techniques that reduce the cost of within-event residual simulation—the main bottleneck of the traditional framework: PCA, a widely used dimensionality-reduction method; and PPCA, the same method proposed in this study for $\mathbf{g}_k$ (see Section S11 of the Supplementary Material for the detailed formulation). In total, the comparison comprises four cases: (1) the traditional two-step approach (Trad); and the same approach with within-event residual simulation accelerated by (2) PCA (Trad GM-PCA), (3) PPCA (Trad GM-PPCA), and (4) the proposed method (Proposed).

We measured runtime as a function of N, from 2,000 to 30,000, and of M, at 10^3, 10^4, and 10^5—which yield a coefficient of variation of the MC estimates of 10% down to failure probability levels of $\sim 10^{-1}$, 10^{-2}, and 10^{-3}, respectively. We report the pre-processing and simulation runtimes separately, as well as the total computation time (= pre-processing + simulation), because the simulation scales with both the number of buildings (N) and the number of MC realizations (M), whereas pre-processing scales only with N (Table 1). The total computation time therefore approaches the pre-processing runtime as M becomes small

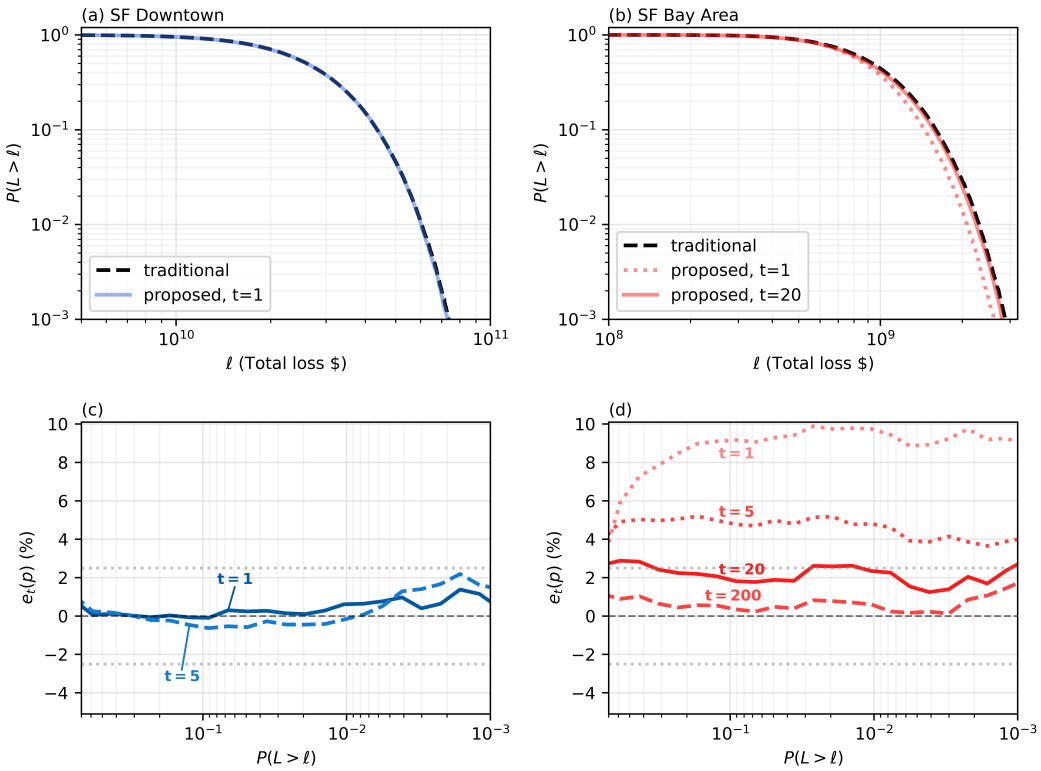

FIGURE 9 (a) Exceedance probability of total loss for the SF Downtown portfolio; the black dashed line represents the traditional framework benchmark, while the blue solid line represents the proposed framework with $t = 1$. (b) Equivalent plot for the Bay Area; the benchmark (black dashed) is compared against the proposed framework with $t = 20$ (red solid line) and $t = 1$ (red dotted line). (c) Relative error of the proposed framework across various exceedance probabilities for SF Downtown ($t = 1$ and $t = 5$, indicated by blue solid and dashed lines, respectively) and (d) the Bay Area ($t = 1$ and 5 indicated by red dotted lines, $t = 20$ by red solid line, and $t = 200$ by red dashed line).

and N becomes large, and approaches the simulation runtime as M becomes large and N becomes small. For Trad GM-PCA, Trad GM-PPCA, and Proposed, runtimes were compared using the value of t at which the top-t eigenpairs capture 95% of the total variance (Equations 17 and 23). One could instead select t to best match the true loss, but the true loss is unknown in an actual simulation, so such a criterion is impractical. For the same reason, and to ensure a fair comparison, we did not use the small-subset method to select t in the proposed framework, relying solely on the variance-contribution criterion.

The analysis was performed in Python 3.11 using NumPy vectorization and Numba[66,67], and all simulations were executed on an Intel Core i7-13700 (2.1 GHz) with 64 GB of RAM, without parallelization.

Figure 10 shows the computation time as a function of N at $M = 10^5$. The proposed method achieved the lowest pre-processing and simulation runtimes for every N, and therefore the lowest total runtime as well. Relative to Trad, at N = 2,000 the pre-processing and simulation runtimes of the proposed framework are 3.2 and 13.9× faster, respectively, giving a 4.5, 10.6, and 13.3× faster total computation time at $M = 10^3$, 10^4, and 10^5, respectively. This gap widens at larger N: at N = 30,000, the pre-processing and simulation runtimes are 9.2 and 97.3× faster, giving a 10.7, 21.1, and 62.7× faster total computation time at $M = 10^3$, 10^4, and 10^5, respectively. This speed-up—and its widening with more buildings—arises from the order-reduced computational complexity in N: $O(N^3)$ to $O(N^2)$ in the pre-processing step, and $O(N^2M)$ to $O(NM)$ in the simulation step (solid black and dark-red curves in Figure 10(c) and (d); Table 1).

For Trad GM-PCA, the pre-processing is even more expensive than Trad, while the gain in simulation is very small (dashed light-blue curve in Figure 10(c) and (d)). This is because the number of principal components required to reach 95% of the total variance is large—roughly 80% of the number of buildings—and this fraction does not change with N (Table S2). The pre-processing complexity of $O(tN^2)$ therefore becomes effectively $O(N^3)$, the same as Trad; and Trad GM-PCA's pre-processing is in fact more expensive, because the eigendecomposition carries a larger constant than the Cholesky factorization (Figure 10(c)). On the simulation side, $O(tNM)$ likewise becomes effectively $O(N^2M)$, so PCA yields only a slight gain over Trad (Figure 10(d)) from the smaller number of columns in the matrix multiplication ($\sim 0.8N$). This gain is negligible compared with the increase in pre-processing, yielding the largest total computational cost (Figure 10(b)).

Trad GM-PPCA, in contrast, successfully reduces the computation time relative to Trad (dotted dark-blue curve in Figure 10(b)–(d)). At N = 2,000, both the pre-processing and simulation runtimes are $\sim$2.7× faster than Trad, giving a $\sim$2.7× faster total computation time at any M. At N = 30,000, the pre-processing and simulation runtimes are 6.9 and 19× faster than Trad, respectively, giving a 7.6, 11.4, and 17× faster total computation time at $M = 10^3$, 10^4, and 10^5, respectively. Under Trad GM-PPCA, the number of principal components (t) required for within-event residual simulation is a constant 27 across a wide range of N—the same behavior as the proposed framework, and in sharp contrast to GM-PCA (Table S2). This value (t = 27) also well captures the within-event residual correlation structure (Figure S8(a)). Trad GM-PPCA thus effectively shares the computational complexity of the proposed framework: $O(N^2)$ for pre-processing and $O(NM)$ for simulation (Table S3).

GM-PPCA is thus an attractive option for users who wish to retain the traditional framework and existing software. To maximize efficiency, however, the proposed framework remains preferable, because the number of components it requires (5) is about five times smaller than that of GM-PPCA (27) (Table S2). Relative to Trad GM-PPCA, for any N the pre-processing runtime of the proposed method is 1.3× faster and the simulation runtime is 5.1× faster, giving a 1.4, 1.8, and 3.6× faster total computation time at $M = 10^3$, 10^4, and 10^5, respectively. The simulation time scales exactly with the ratio of principal components (27/5), whereas the pre-processing time blurs this effect, because it also includes construction of the correlation matrix.

The difference in t between Trad GM-PPCA and the proposed framework can be explained as follows. The larger (but still small) t = 27 is needed for an accurate within-event residual simulation, which in turn yields accurate damage modeling; however, it is unnecessarily large for that purpose. The proposed framework instead targets the variance of the final damage variable $\mathbf{g}_k$, which blends the within-event residual (with β normalization) with the between-event term and the identity matrix (Equation 16). This blending renders the effective within-event field only as accurate as damage modeling requires—rather than unnecessarily accurate—yielding a more efficient simulation.

Finally, we found that sparse-approximation methods for within-event residual simulation are inappropriate for the ground-shaking modeling of extremely dense urban building portfolios. As a representative example, we tested the Vecchia approximation[43,57], which Zhao et al.[40] employed, from the same family of methods, for non-ergodic ground-motion modeling. It assumes conditional independence in the spatial field given the K nearest neighbors of each site, with a computational complexity of $O(NK^3)$ for pre-processing and $O(NK\,M)$ for sampling. We found, however, that such methods are not efficient for this very dense urban region: accurately reproducing the within-event residual requires $K \approx 30\%$ of N (Figure S8(b)), and this fraction does not decrease with N. The complexity therefore converges effectively to $O(N^4)$ for pre-processing and $O(N^2M)$ for

sampling. This is intuitive: for a fixed spatial extent, adding more sites places more buildings within the correlation length of each site, so more conditioning neighbors are needed for accurate simulation.

7 | SUMMARY

We developed a scalable computational framework for regional scenario-based seismic risk modeling, built on an exact linearizing transformation of ground-motion–damage coupling, together with dimensionality reduction via PPCA. First, we demonstrate that canonical coupling between ground motions and lognormal fragility functions can be linearized exactly through a change of variables. This transformation reveals that the traditional two-step process to simulate ground-motion–fragility coupling of N assets can be simplified into a one-shot sampling of an N-dimensional correlated Gaussian vector.

Then, we reconstructed this Gaussianized formulation using PPCA-based dimensionality reduction. This is efficient given the structure of the covariance matrix of this formulation, because the isotropic, independent noise component can be cleanly extracted, leaving the highly correlated portion to be approximated efficiently with standard PCA. We identified the magnitude of the noise component by analytically deriving a lower bound on the eigenvalues of the covariance matrix. Additionally, through numerical experiments, we found that the number of eigenvalues greater than this lower bound corresponds to the number of spatial building clusters (t) in the portfolio, with this same number of principal components being retained for the highly correlated portion.

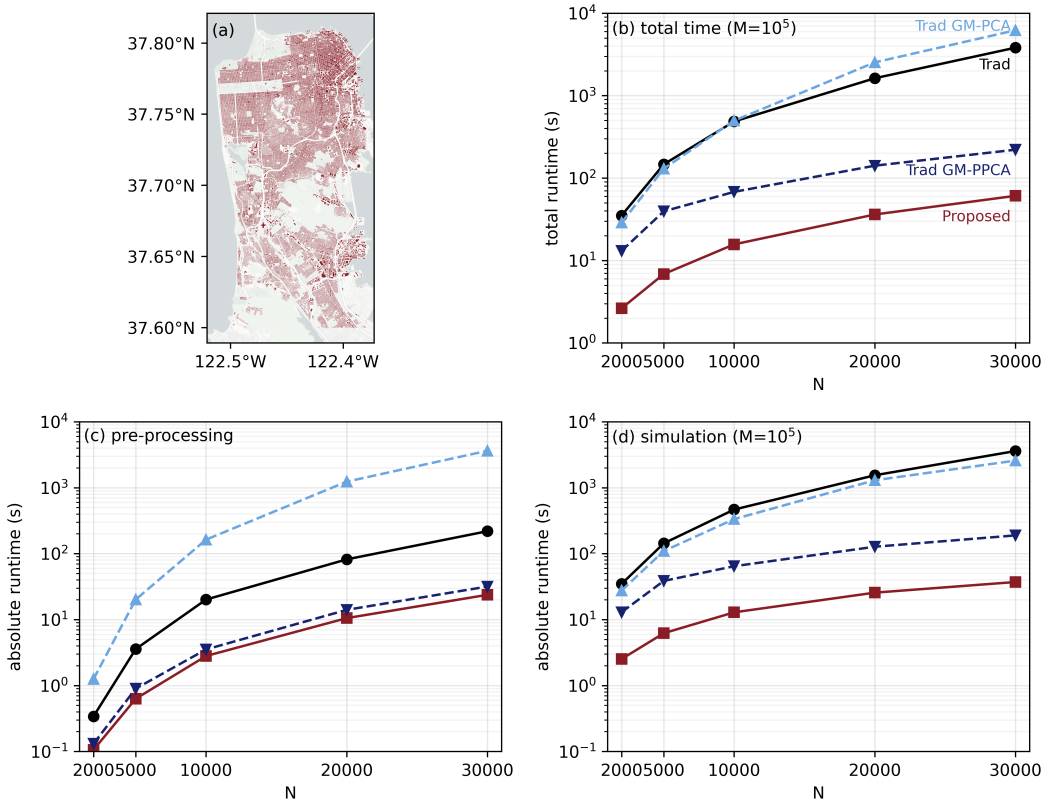

FIGURE 10 (a) The San Francisco Northern Peninsula building portfolio (N = 30,000) used to measure the computation time of the proposed framework. (b) Total runtime as a function of the number of buildings N for $M = 10^5$: proposed framework (Proposed; dark-red solid line with squares), Traditional (Trad; black solid line with circles), Traditional with PCA on the ground-shaking map (Trad GM-PCA; light-blue dashed line with triangles), and Traditional with PPCA on the ground-shaking map (Trad GM-PPCA; dark-blue dashed line with inverted triangles). (c) Pre-processing time and (d) simulation time as a function of the number of buildings N, for $M = 10^5$. For each N, the sum of (c) and (d) gives the total time shown in (b).

Verification using the R2D example—an established software for regional seismic risk modeling—demonstrates that the resulting damage-state distributions are nearly identical to those obtained from the traditional framework. The mean damage state differs by at most ~ 0.04. We also found that for the San Francisco downtown portfolio containing 15,836 buildings, a value of $t = 1$ suffices for accurate modeling. In contrast, for the entire Bay Area with the same number of buildings, $t = 20$ is required, yielding a loss estimation error within approximately 2.5% of the benchmark. Importantly, in this realistic portfolio, the building cluster parameter t depends primarily on the spatial extent of the portfolio rather than the building density. This suggests that the building cluster parameter t can be determined through computationally efficient small-sample simulations, e.g., using 100 subsamples for a full inventory of 15,836 buildings. Additionally, we found that the eigenvectors exhibit spatially varying 2-d sinusoidal harmonics, where the order of these harmonics is closely related to the spatial elongation of the portfolio.

The computational gain of the proposed framework was also theoretically and experimentally verified. The proposed framework reduces the computational complexity of risk modeling from cubic complexity (e.g., $O(N^3)$ or $O(N^2M)$) effectively to quadratic complexity ($O(N^2)$ or $O(NM)$), where M represents the number of MC simulations. For a 30,000-building portfolio of the San Francisco Northern Peninsula, this translates to 9× faster pre-processing and 97× faster simulation. This advantage is expected to grow with the portfolio size N. Alternatively, applying PPCA to the within-event residual simulation of ground-shaking intensity within the traditional framework also improved computation time (6.9× faster pre-processing and 19× faster simulation); this lets users adopt PPCA within existing code and workflows, although its simulation runtime remains about fivefold slower than that of the proposed framework.

Overall, the proposed framework enables large-scale regional seismic risk analysis that remains computationally tractable without compromising accuracy. It substantially reduces computation time while mitigating the uncertainties associated with ground-shaking gridding and exposure resolution in the analysis. This improvement is particularly significant for metropolitan areas characterized by densely distributed building portfolios.

ACKNOWLEDGEMENTS

The authors acknowledge financial support from the Department of Civil and Environmental Engineering at the University of California, Berkeley, and the National Science Foundation under Grant No. CMMI-2541220. The authors also thank the editor, Dr. Iunio Iervolino, and two anonymous reviewers for their constructive feedback. This work was previously made available as a preprint on Engineering Archive (https://doi.org/10.31224/6578)

DATA AVAILABILITY STATEMENT

The source code for computing regional scenario-based damage modeling using the proposed framework explained in this paper is available at DesignSafe (https://doi.org/10.17603/ds2-9hbb-c610) or GitHub (https://github.com/sehoung/Linearization_PPCA).

References

1. Ceferino L, Merino Y, Pizarro S, Moya L, Ozturk B. Placing engineering in the earthquake response and the survival chain. *Nature Communications.* 2024;15(1):4298.

2. Ceferino L, Kukunoor C, Zhao J, et al. Accessing acute care hospitals in the San Francisco Bay Area after a major hayward earthquake. *Nature Communications.* 2025;16(1):9328.

3. Kiremidjian A, Moore J, Fan YY, Yazlali O, Basoz N, Williams M. Seismic risk assessment of transportation network systems. *Journal of Earthquake Engineering.* 2007;11(3):371–382.

4. Silva-Lopez R, Baker JW. Optimal Bridge retrofitting selection for seismic risk management using genetic algorithms and neural Network–Based surrogate models. *Journal of Infrastructure Systems.* 2023;29(4):04023030.

5. Chen R, Jaiswal KS, Bausch D, Seligson H, Wills C. Annualized earthquake loss estimates for California and their sensitivity to site amplification. *Seismological Research Letters.* 2016;87(6):1363–1372.

6. Ceferino L, Kiremidjian A, Deierlein G. Regional multiseverity casualty estimation due to building damage following a Mw 8.8 earthquake scenario in Lima, Peru. *Earthquake Spectra.* 2018;34(4):1739–1761.

7. Ceferino L, Mitrani-Reiser J, Kiremidjian A, Deierlein G, Bambarén C. Effective plans for hospital system response to earthquake emergencies. *Nature communications.* 2020;11(1):4325.

8. Liu Y, Wotherspoon L, Nair NKC, Blake D. Quantifying the seismic risk for electric power distribution systems. *Structure and Infrastructure Engineering.* 2021;17(2):217–232.

9. Yoshikawa H, Goda K. Financial seismic risk analysis of building portfolios. *Natural Hazards Review.* 2014;15(2):112–120.

10. Goda K, Hong H. Estimation of seismic loss for spatially distributed buildings. *Earthquake Spectra.* 2008;24(4):889–910.

11. Crowley H, Bommer JJ. Modelling seismic hazard in earthquake loss models with spatially distributed exposure. *Bulletin of Earthquake Engineering.* 2006;4:249–273.

12. Moehle J, Deierlein GG. A framework methodology for performance-based earthquake engineering. In: . 679. WCEE Vancouver. 2004:12.

13. Goda K, Hong HP. Spatial correlation of peak ground motions and response spectra. *Bulletin of the Seismological Society of America.* 2008;98(1):354–365.

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

15. Loth C, Baker JW. A spatial cross-correlation model of spectral accelerations at multiple periods. *Earthquake Engineering & Structural Dynamics.* 2013;42(3):397–417. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/eqe.2212doi: 10.1002/eqe.2212

16. Weatherill G, Silva V, Crowley H, Bazzurro P. Exploring the impact of spatial correlations and uncertainties for portfolio analysis in probabilistic seismic loss estimation. *Bulletin of Earthquake Engineering.* 2015;13(4):957–981.

17. Dabbeek J, Crowley H, Silva V, Weatherill G, Paul N, Nievas CI. Impact of exposure spatial resolution on seismic loss estimates in regional portfolios. *Bulletin of Earthquake Engineering.* 2021;19(14):5819–5841.

18. Bhattacharya A, Chakraborty A, Mallick BK. Fast sampling with Gaussian scale mixture priors in high-dimensional regression. *Biometrika.* 2016:asw042.

19. Trefethen LN, Bau D. *Numerical linear algebra*. SIAM, 2022.

20. Li P, Wang Z, Zhao B, Becker T, Soga K. Surrogate modeling for identifying critical bridges in traffic networks under earthquake conditions. *Transportation Research Part D: Transport and Environment.* 2025;138:104512.

21. Federal Emergency Management Agency (FEMA) . *HAZUS Earthquake Model Technical Manual: HAZUS 6.1*. FEMA; : 2024. Version 6.1.

22. Kiremidjian AS, Stergiou E, Lee R. Issues in seismic risk assessment of transportation networks. *Geotechnical, Geological and Earthquake Engineering.* 2007;6:461–480. doi: 10.1007/978-1-4020-5893-6_19

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

24. Houng SE, Ceferino L. Fast probabilistic seismic hazard analysis through adaptive importance sampling. *Bulletin of the Seismological Society of America.* 2025;115(2):646–663.

25. Houng SE, Ceferino L, Abrahamson N. Fast Propagation of Epistemic Uncertainty in Seismic Hazard via Adaptive Importance Sampling. *Bulletin of the Seismological Society of America.* 2026.

26. Han Y, Davidson RA. Probabilistic seismic hazard analysis for spatially distributed infrastructure. *Earthquake Engineering & Structural Dynamics.* 2012;41(15):2141–2158.

27. Miller M, Baker J. Ground-motion intensity and damage map selection for probabilistic infrastructure network risk assessment using optimization. *Earthquake Engineering & Structural Dynamics.* 2015;44(7):1139–1156.

28. Manzour H, Davidson RA, Horspool N, Nozick LK. Seismic hazard and loss analysis for spatially distributed infrastructure in Christchurch, New Zealand. *Earthquake Spectra.* 2016;32(2):697–712.

29. Christou V, Bocchini P, Miranda MJ, Karamlou A. Effective sampling of spatially correlated intensity maps using hazard quantization: Application to seismic events. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering.* 2018;4(1):04017035.

30. Au SK, Beck JL. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic engineering mechanics.* 2001;16(4):263–277.

31. Del Moral P, Doucet A, Jasra A. Sequential monte carlo samplers. *Journal of the Royal Statistical Society Series B: Statistical Methodology.* 2006;68(3):411–436.

32. Wang Z, Song J. Cross-entropy-based adaptive importance sampling using von Mises-Fisher mixture for high dimensional reliability analysis. *Structural Safety.* 2016;59:42–52.

33. Wang Z, Broccardo M, Song J. Hamiltonian Monte Carlo methods for subset simulation in reliability analysis. *Structural Safety.* 2019;76:51–67.

34. Bal IE, Bommer JJ, Stafford PJ, Crowley H, Pinho R. The influence of geographical resolution of urban exposure data in an earthquake loss model for Istanbul. *Earthquake Spectra.* 2010;26(3):619–634.

35. Bazzurro P, Park J. The effects of portfolio manipulation on earthquake portfolio loss estimates. In: . 31. 2007:8.

36. Scheingraber C, Käser MA. The impact of portfolio location uncertainty on probabilistic seismic risk analysis. *Risk Analysis.* 2019;39(3):695–712.

37. Markhvida M, Ceferino L, Baker JW. Modeling spatially correlated spectral accelerations at multiple periods using principal component analysis and geostatistics. *Earthquake Engineering & Structural Dynamics.* 2018;47(5):1107–1123.

38. Du W, Ning CL. Modeling spatial cross-correlation of multiple ground motion intensity measures (SAs, PGA, PGV, Ia, CAV, and significant durations) based on principal component and geostatistical analyses. *Earthquake Spectra.* 2021;37(1):486–504.

39. Yang Y, Wang P, Brandenberg SJ. An algorithm for generating spatially correlated random fields using Cholesky decomposition and ordinary kriging. *Computers and Geotechnics.* 2022;147:104783.

40. Zhao J, Lavrentiadis G, Asimaki D. Scalable Gaussian Process for Learning Non-Ergodic Ground Motion Model from Physics-Based Simulations with Application to Power Infrastructure Assessment. *arXiv preprint arXiv:2605.24284.* 2026.

41. Ghanem RG, Spanos PD. *Stochastic Finite Elements: A Spectral Approach*. New York: Springer-Verlag, 1991

42. Tipping ME, Bishop CM. Probabilistic principal component analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology.* 1999;61(3):611–622.

43. Vecchia AV. Estimation and Model Identification for Continuous Spatial Processes. *Journal of the Royal Statistical Society: Series B (Methodological).* 1988;50(2):297–312. doi: 10.1111/j.2517-6161.1988.tb01729.x

44. Golub GH, Van Loan CF. *Matrix computations*. JHU press, 2013.

45. 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.* 2014;30(3):1057–1085.

46. 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.

47. 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.

48. Abrahamson NA, Silva WJ, Kamai R. Summary of the ASK14 ground motion relation for active crustal regions. *Earthquake Spectra.* 2014;30(3):1025–1055.

49. Bozorgnia Y, Abrahamson NA, Atik LA, et al. NGA-West2 research project. *Earthquake Spectra.* 2014;30(3):973–987.

50. Alexander NA, Ibraim E, Aldaikh H. A simple discrete model for interaction of adjacent buildings during earthquakes. *Computers & Structures.* 2013;124:1–10.

51. Luco JE, Contesse L. Dynamic structure-soil-structure interaction. *Bulletin of the Seismological Society of America.* 1973;63(4):1289–1303.

52. Abdi H, Williams LJ. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics.* 2010;2(4):433–459.

53. Williams C, Seeger M. Using the Nyström method to speed up kernel machines. *Advances in neural information processing systems.* 2000;13.

54. Drineas P, Mahoney MW. On the Nyström Method for Approximating a Gram Matrix for Improved Kernel-Based Learning. *Journal of Machine Learning Research.* 2005;6:2153–2175.

55. Harbrecht H, Peters M, Schneider R. On the Low-Rank Approximation by the Pivoted Cholesky Decomposition. *Applied Numerical Mathematics.* 2012;62(4):428–440. doi: 10.1016/j.apnum.2011.10.001

56. Rue H, Held L. *Gaussian Markov Random Fields: Theory and Applications*. 104 of *Monographs on Statistics and Applied Probability*. Boca Raton, FL: Chapman & Hall/CRC, 2005

57. Katzfuss M, Guinness J. A General Framework for Vecchia Approximations of Gaussian Processes. *Statistical Science.* 2021;36(1):124–141. doi: 10.1214/19-STS755

58. Furrer R, Genton MG, Nychka D. Covariance Tapering for Interpolation of Large Spatial Datasets. *Journal of Computational and Graphical Statistics.* 2006;15(3):502–523. doi: 10.1198/106186006X132178

59. Hackbusch W, others . *Hierarchical matrices: algorithms and analysis*. 49. Springer, 2015.

60. Katafygiotis LS, Zuev KM. Geometric insight into the challenges of solving high-dimensional reliability problems. *Probabilistic Engineering Mechanics.* 2008;23(2-3):208–218.

61. Sadigh K, Chang CY, Egan J, Makdisi F, Youngs RR. Attenuation relationships for shallow crustal earthquakes based on California strong motion data. *Seismological research letters.* 1997;68(1):180–189.

62. Zsarnóczay A, Elhaddad W, Cetiner B, Zhong K, McKenna F, Deierlein G. SimCenter Earthquake Testbed: San Francisco, CA.; 2023

63. Deierlein GG, McKenna F, Zsarnóczay A, et al. A Cloud-Enabled Application Framework for Simulating Regional-Scale Impacts of Natural Hazards on the Built Environment. *Frontiers in Built Environment.* 2020;6. Publisher: Frontiersdoi: 10.3389/fbuil.2020.558706

64. McKenna F, Gavrilovic S, Zhao J, et al. NHERI-SimCenter/R2DTool: Version 5.2.0. https://zenodo.org/records/14219019; 2024

65. Goda K, Zhang L, Tesfamariam S. Portfolio Seismic Loss Estimation and Risk-based Critical Scenarios for Residential Wooden Houses in Victoria, British Columbia, and Canada. *Risk analysis.* 2021;41(6):1019–1037.

66. Lam SK, Pitrou A, Seibert S. Numba: a LLVM-based Python JIT compiler. In: LLVM '15. Association for Computing Machinery 2015; New York, NY, USA:1–6

67. Harris CR, Millman KJ, Van Der Walt SJ, et al. Array programming with NumPy. *nature.* 2020;585(7825):357–362.

68. Kurtz N, Song J. Cross-entropy-based adaptive importance sampling using Gaussian mixture. *Structural Safety.* 2013;42:35–44.

69. Bishop CM. *Pattern Recognition and Machine Learning*. Information Science and StatisticsNew York, NY: Springer, 2006.

70. Guinness J. Permutation and Grouping Methods for Sharpening Gaussian Process Approximations. *Technometrics.* 2018;60(4):415–429. doi: 10.1080/00401706.2018.1437476

71. Watanabe S. Information Theoretical Analysis of Multivariate Correlation. *IBM Journal of Research and Development.* 1960;4(1):66–82. doi: 10.1147/rd.41.0066

72. McFadden D. Conditional Logit Analysis of Qualitative Choice Behavior. In: Zarembka P., ed. *Frontiers in Econometrics*, , New York: Academic Press, 1974:105–142.

73. Ambikasaran S, Foreman-Mackey D, Greengard L, Hogg DW, O'Neil M. Fast Direct Methods for Gaussian Processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2016;38(2):252–265. doi: 10.1109/TPAMI.2015.2448083

74. Solin A, Särkkä S. Hilbert Space Methods for Reduced-Rank Gaussian Process Regression. *Statistics and Computing.* 2020;30(2):419–446. doi: 10.1007/s11222-019-09886-w

75. Rahimi A, Recht B. Random features for large-scale kernel machines. *Advances in neural information processing systems.* 2007;20.

76. Lindgren F, Rue H, Lindström J. An Explicit Link between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology).* 2011;73(4):423–498. doi: 10.1111/j.1467-9868.2011.00777.x

77. Dietrich CR, Newsam GN. Fast and Exact Simulation of Stationary Gaussian Processes through Circulant Embedding of the Covariance Matrix. *SIAM Journal on Scientific Computing.* 1997;18(4):1088–1107. doi: 10.1137/S1064827592240555

78. Wood ATA, Chan G. Simulation of Stationary Gaussian Processes in $[0, 1]^d$. *Journal of Computational and Graphical Statistics.* 1994;3(4):409–432. doi: 10.2307/1390903

79. Datta A, Banerjee S, Finley AO, Gelfand AE. Hierarchical Nearest-Neighbor Gaussian Process Models for Large Geostatistical Datasets. *Journal of the American Statistical Association.* 2016;111(514):800–812. doi: 10.1080/01621459.2015.1044091

80. Huang SP, Quek ST, Phoon KK. Convergence Study of the Truncated Karhunen–Loève Expansion for Simulation of Stochastic Processes. *International Journal for Numerical Methods in Engineering.* 2001;52(9):1029–1043. doi: 10.1002/nme.255

81. Dongarra JJ, Moler CB, Bunch JR, Stewart GW. *LINPACK users' guide*. SIAM, 1979.

82. Marsaglia G, Tsang WW. The ziggurat method for generating random variables. *Journal of statistical software.* 2000;5:1–7.

83. Joyner WB, Boore DM. Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 Imperial Valley, California, earthquake. *Bulletin of the seismological Society of America.* 1981;71(6):2011–2038.

84. Horn RA, Johnson CR. *Matrix Analysis*. Cambridge: Cambridge University Press. 2nd ed., 2013.

85. Sadigh K, Chang CY, Egan J, Makdisi F, Youngs R. Attenuation Relationships for Shallow Crustal Earthquakes Based on California Strong Motion Data. *Seismological Research Letters.* 1997;68(1):180–189. doi: 10.1785/gssrl.68.1.180

86. Park J, Bazzurro P, Baker JW, others . Modeling spatial correlation of ground motion intensity measures for regional seismic hazard and portfolio loss estimation. *Applications of statistics and probability in civil engineering.* 2007;2:1–8.

87. DeBock DJ, Liel AB. A comparative evaluation of probabilistic regional seismic loss assessment methods using scenario case studies. *Journal of Earthquake Engineering.* 2015;19(6):905–937.

88. Antonietti PF, Mazzieri I, Melas L, et al. Three-dimensional physics-based earthquake ground motion simulations for seismic risk assessment in densely populated urban areas. *Mathematics in Engineering.* 2021;3(2):1–31.

89. Graves R, Jordan TH, Callaghan S, et al. CyberShake: A physics-based seismic hazard model for southern California. *Pure and Applied Geophysics.* 2011;168(3):367–381.

90. Smerzini C, Pitilakis K. Seismic risk assessment at urban scale from 3D physics-based numerical modeling: the case of Thessaloniki. *Bulletin of Earthquake Engineering.* 2018;16(7):2609–2631.

91. Boore DM. Simulation of ground motion using the stochastic method. *Pure and applied geophysics.* 2003;160(3):635–676.

92. Irikura K. Prediction of strong acceleration motion using empirical Green's function. In: . 151. 1986:151–156.

93. Lee R, Kiremidjian AS. Uncertainty and correlation for loss assessment of spatially distributed systems. *Earthquake Spectra.* 2007;23(4):753–770.

94. Sokolov V, Wenzel F. Influence of spatial correlation of strong ground motion on uncertainty in earthquake loss estimation. *Earthquake Engineering & Structural Dynamics.* 2011;40(9):993–1009.

95. Heresi P, Miranda E. Structure-to-structure damage correlation for scenario-based regional seismic risk assessment. *Structural Safety.* 2022;95:102155.

96. Mueller CS. Source pulse enhancement by deconvolution of an empirical Green's function. *Geophysical Research Letters.* 1985;12(1):33–36.

97. Field EH, Biasi GP, Bird P, et al. Long-term time-dependent probabilities for the third Uniform California Earthquake Rupture Forecast (UCERF3). *Bulletin of the Seismological Society of America.* 2015;105(2A):511–543.

98. Handy SL, Niemeier DA. Measuring accessibility: an exploration of issues and alternatives. *Environment and planning A.* 1997;29(7):1175–1194.

99. Horn RA, Johnson CR. *Matrix analysis*. Cambridge university press, 2012.