Supplementary Material: Scalable Scenario-based Earthquake Risk Modeling via Linearized Ground-Motion–Fragility Coupling and Probabilistic PCA

SoungEil Houng and Luis Ceferino

S1 Sensitivity of damage probability to ground-shaking and fragility parameters

We demonstrate how the linearized formulation provides a simpler and more direct perspective of seismic risk, while maintaining PBEE rigor. First, we examine how the mean shaking intensity μ influences the resulting damage. For a fixed structural vulnerability $\ln\theta_k$ (vertical gray dotted line in Figure S1(a)), let us first assume that $\mu < \ln\theta_k$. This scenario corresponds to either a strong structure or weak mean shaking intensity (light or dark red curves in Figure S1(a)). From Figure S1(a), we infer that the probability of failure is less than 0.5, as the majority of the shaking intensity probability mass lies to the left of $\ln\theta_k$, in the region where the failure probability is low. This is reflected in Figure S1(b), showing $P(DS \geq k)$ as a function of μ and σ: regardless of σ, a value of μ less than $\ln\theta_k$ yields $P(DS \geq k) < 0.5$ (red shaded region). This trend holds true even when β changes (Figure S1(c)). Conversely, if $\mu > \ln\theta_k$—corresponding to either a weak structure or strong mean shaking intensity—the $P(DS \geq k)$ exceeds 0.5. In this case, the majority of the shaking intensity probability mass lies to the right of $\ln\theta_k$ (dark blue curve in Figure S1(a)), leading to a $P(DS \geq k)$ greater than 0.5 (blue shaded regions in (b) and (c)).

Our linearized formulation facilitates a straightforward interpretation of this property. When $\mu < \ln\theta_k$, g_k exhibits a Gaussian curve with a positive mean (light or dark red curves in Figure S1(d); see Equation 10). The probability of failure $P(DS \geq k)$ corresponds to the area under the Gaussian curve where g_k is negative (Equation 11; gray shaded area under the reddish curves in Figure S1(d)). Since the mean of g_k is positive, this area is strictly less than 0.5. Similarly, if $\mu > \ln\theta_k$, g_k has a negative mean (blue curves in Figure S1(d)), resulting in a failure probability greater than 0.5.

Secondly, we investigate the impact of shaking intensity variability, σ. The impact of σ on risk depends on the position of μ relative to $\ln\theta_k$. In the traditional framework, when $\mu < \ln\theta_k$, a larger σ (light red curve in Figure S1(a) vs. dark red curve) increases risk. This is because a larger σ broadens the shaking intensity distribution, increasing the probability of extreme tail events. As observed in the red-shaded region of Figure S1(b) (where $\mu < \ln\theta_k$), the failure probability increases as σ increases. On the other hand, when $\mu > \ln\theta_k$, an increase in σ has the opposite effect: it increases the probability of realizing lower shaking intensities, thereby reducing the risk (light blue curve in Figure S1(a) vs. dark blue curve). As observed in the blue-shaded region of Figure S1(b) (where $\mu > \ln\theta_k$), the failure probability decreases as σ increases.

This phenomenon is also readily explained by our linearized formulation. For $\mu < \ln\theta_k$ (light and dark red curves in Figure S1(d)), the shaded failure area is larger for the high-σ case (light red curve), indicating higher risk. Conversely, if $\mu > \ln \theta_k$ (light and dark blue curves), an increase in σ (light blue curve) reduces the shaded area where $g_k < 0$, indicating reduced risk. Although increasing σ can theoretically either reduce or increase the risk depending on μ, low mean intensity events (small magnitude/far distance, where $\mu < \ln \theta_k$) occur more frequently than high mean intensity events. Consequently, increasing σ generally increases the total estimated risk [1].

Thirdly, we investigate the influence of the fragility curve dispersion parameter, β. As β decreases, the fragility curve approaches a step function (black solid curve in Figure S1(a) vs. black dashed curve). In the limit as $\beta \to 0$, it becomes a deterministic damage function: intensities below $\ln \theta_k$ yield 0% failure probability, while those above yield 100%. Conversely, as β increases, the fragility curve flattens (black dashed line in Figure S1(a)). In the limit as $\beta \to \infty$, the probability of failure converges to 0.5 across all ground motion intensities.

For small β of 0.4 (Figure S1(b)), the probability of failure $P(DS \geq k)$ changes abruptly with variations in μ and σ, reflecting a high sensitivity of risk to ground motion parameters. For example, for σ of 0.6, $P(DS \geq k)$ changes from 0.10 to 0.84 when μ varies from -1.6 to 0 (0.2 g to 1 g). However, for large β of 1.0 (Figure S1(c)), $P(DS \geq k)$ changes more gradually. This reduced sensitivity arises because the large uncertainty in the structural capacity washes out the specific details of the ground motion intensity. For example, for σ of 0.6, $P(DS \geq k)$ changes from 0.22 to 0.73 when μ varies from -1.6 to 0 (0.2 g to 1 g).

In our linearized framework, this behavior is observed as changes in both the mean and standard deviation of g_k (Equation 10). When β becomes large, the mean of g_k approaches zero, and the standard deviation decreases towards 1. This results in g_k distributions that are centered near zero with short tails (Figure S1(e) compared to (d)). Due to this reduced variability in g_k, the probability of failure remains relatively similar across different ground motion parameters, rendering the risk estimate less sensitive to ground motion characteristics. Intuitively, for large β, the fragility curve does not know how strong or weak the structure is, thus, the variation of the parameters does not matter. Intuitively, as $\beta \to \infty$, the failure probability converges toward 0.5 across the entire range of shaking intensities, effectively eliminating any distinction between them; consequently, the influence of the parameters on the overall risk estimate becomes negligible.

Finally, we consider the role of θ_k. This parameter appears only in the mean-shift term of g_k and enters with a sign opposite to that of μ (Equation 10). Therefore, its effect on $P(DS \geq k)$ is inverse to that of μ: if $\ln \theta_k > \mu$, $P(DS \geq k)$ is less than 0.5, whereas if $\ln \theta_k < \mu$, $P(DS \geq k)$ is greater than 0.5.

S2 Comparison Between Standard PCA and Probabilistic PCA: Toy example

This section demonstrates the mechanics of standard Principal Component Analysis (PCA) and probabilistic PCA, illustrating how these techniques facilitate efficient simulations of $\mathbf{g}_k$, using a two-dimensional toy example (two sites). We begin by examining standard PCA with two different examples exhibiting strong and weak spatial dependencies. In the case of strong dependency (Figure S2(a)), the eigenvalues of $\mathbf{Cov}(\mathbf{g}_k)$ are distinct: the primary eigenvalue is 1.9, while the secondary is 0.1. Conversely, in the weak dependency case (Figure S2(b)), this gap narrows significantly, with eigenvalues of 1.2 and 0.8. Figure S2(c) illustrates the sorted eigenvalue spectra for both scenarios. For the strong dependency case (solid line), the eigenvalues decay rapidly; the leading eigenvalue dominates the second, which is nearly zero. In contrast, the eigenvalues for the weak dependency case (dashed line) remain comparable in magnitude and do not converge toward zero.

From the perspective of dimensionality reduction, strong dependency implies that a full two-dimensional simulation is redundant. Simulating a one-dimensional Gaussian along the direction of the first principal component (λ_1) is sufficient, as the variability along the second direction (λ_2) is negligible (Figure S2(a)). In other words, the variance along λ_1 captures the vast majority of the total system variance, allowing the dimensionality to be reduced from two to one. However, in the weak dependency case (Figure S2(b)), PCA-based dimensionality reduction is infeasible. The variance along the λ_2 axis is non-negligible, meaning the second component still accounts for a significant proportion of the total variance and cannot be discarded without substantial information loss.

To formalize this, we define m_0 as the minimum number of eigenpairs required to capture a target proportion of the total variance F_m (Equation 16):

$$m_0 = \min\{m \mid F_m \geq q\} \tag{S1}$$

where q is a threshold typically set close to 100% (e.g., $q = 90\%$). The value of m_0 indicates the efficacy of PCA. If m_0 is significantly smaller than N, PCA is considered effective, as the first few eigenpairs allow the data to be approximated by an m_0-dimensional Gaussian random variable. If m_0 is close to N, PCA is ineffective, as the reduced representation requires a number of axes comparable to the original dimension N.

Figure S2(d) illustrates F_m for the strong and weak dependency cases. In the high-dependency scenario, the first eigenpair alone explains 95% of the total variance, i.e., $m_0 = 1$. In the low-dependency case, it explains only 60%, and two eigenpairs are required to explain the data variance, i.e., $m_0 = 2$. This implies that dimensionality reduction from two to one in the latter case would result in a significant deviation from the true $\vec{g}_k$. This concept extends to higher-dimensional systems, where the utility of PCA depends on whether the eigenvalue spectrum of the covariance matrix $\mathbf{Cov}(\mathbf{g}_k)$ decays quickly to zero enough to be represented by a few dominant components.

For cases characterized by low correlation, Probabilistic PCA offers a more robust framework for dimensionality reduction. In PPCA, the original data is decomposed into a highly correlated component and an independent noise component, allowing the data to be modeled as the sum of these two parts. Figure S3(a) illustrates the original data, while (b) and (c) represent the correlated component and the independent noise component, respectively.

In this example, the correlated portion explains 22% of the total variance $((0.42+0.02)/(0.8+1.2)\times 100)$, while the noise component accounts for 78% $((0.78 + 0.78)/(0.8 + 1.2) \times 100)$. The dominance of the noise component is visually evident in the spatial coverage shown in Figure S3 (b) and (c), where the data occupies a larger area in (c). Notably, within the correlated portion, the first principal component explains 95% of that portion's variance, making standard PCA applicable to this sub-component. By modeling the correlated part using only the first PC and adding the noise term, the model captures 98.4% of the total variance (represented by the first square at λ_2 on the dash-dotted curve in Figure S3(d)). In comparison, standard PCA using only the first PC explains only 60% of the total variance (the first triangle on the dashed curve). This demonstrates the superior efficiency of PPCA for datasets with low-to-moderate correlations.

(a) Eigenvalues and corresponding eigenvectors illustrating a low degree of linear dependency in $\mathbf{g}_k$ between two sites, as in

S3 Implementation of the Proposed Scenario-Based Regional Risk Modeling Framework

This section summarizes the implementation procedure of the proposed framework for the scenario-based regional risk modeling of spatially distributed portfolios. We first present the itemized step-by-step procedures for the pre-processing and simulation steps, and then demonstrate how to numerically calculate this using a simple, exemplary four-site case.

S3.1 Procedure Summary

Given N target buildings ($i = 1, 2, ..., N$), their location and the fragility curve parameters β_i and θ_{ik} are provided, where $k = 0, 1, ..., n_{\text{ds}}$ denotes the categorical damage states. Before begin, make sure that the $\mathbf{B}$ from fragility curve parameters, $\boldsymbol{\tau}$ and $\mathbf{F}$ from ground motion standard deviation model, and $\mathbf{C}$ from within-event ground motion correlation model be ready.

The pre-processing step follows two steps:

p-1) Calculate the constant c^* using Equation 27. [Output: c^*]

p-2) Create $\mathbf{Cov}(\mathbf{g}_k)$ using Equation 15, calculate top t largest eigenvalues and corresponding eigenvectors of $\mathbf{Cov}(\mathbf{g}_k)$, and construct the matrix $\mathbf{W}_t^*$ using Equation 36. [Output: $\mathbf{W}_t^*$]

The simulation step follows four steps below:

s-1) Sample $t + N$ standard normal random variables to create $\mathbf{x}$ (length t) and $\mathbf{z}$ (length N).

s-2) Compute

$$\boldsymbol{\gamma} = \mathbf{W}_t^* \mathbf{x} + c^* \mathbf{z}.$$

s-3) For each damage state $k = 1, 2, ..., n_{\text{ds}}$, compute

$$\mathbf{g}_k = \boldsymbol{\gamma} + \mathbf{b}_k, \quad \text{where} \quad \mathbf{b}_k = \mathbf{B}(\ln \boldsymbol{\theta}_k - \boldsymbol{\mu}),$$

s-4) Determine the damage state vector $\mathbf{DS}$, where each component DS_i is the maximum k such that $g_{ik} < 0$.

Repeat steps s-1 to s-4 to generate a sufficient number of damage maps.

S3.2 Illustrative Example

We assume four buildings with the geometry specified in Figure S4. Each building's fragility curve is given in Table S1. Step numbers correspond to those in the previous section. For readability, all numbers are rounded to three decimal places; thus, some intermediate values may differ slightly from full-precision calculations.

From fragility curve parameter (Table S1), we obtain

$$\mathbf{B} = \begin{bmatrix} 1/0.4 & 0 & 0 & 0 \\ 0 & 1/0.4 & 0 & 0 \\ 0 & 0 & 1/0.4 & 0 \\ 0 & 0 & 0 & 1/0.4 \end{bmatrix} = \begin{bmatrix} 2.5 & 0 & 0 & 0 \\ 0 & 2.5 & 0 & 0 \\ 0 & 0 & 2.5 & 0 \\ 0 & 0 & 0 & 2.5 \end{bmatrix}$$

From ground motion model,

$$\boldsymbol{\tau} = \begin{bmatrix} 0.4 \\ 0.4 \\ 0.4 \\ 0.4 \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} 0.6 & 0 & 0 & 0 \\ 0 & 0.6 & 0 & 0 \\ 0 & 0 & 0.6 & 0 \\ 0 & 0 & 0 & 0.6 \end{bmatrix}$$

From a simple within-event correlation model of $\rho_{ij} = \exp[-3 \times h_{ij}/10]$, where h_{ij} are the inter-site distances in Figure S4:

$$\mathbf{C} = \begin{bmatrix} 1.000 & 0.942 & 0.741 & 0.216 \\ 0.942 & 1.000 & 0.736 & 0.230 \\ 0.741 & 0.736 & 1.000 & 0.223 \\ 0.216 & 0.230 & 0.223 & 1.000 \end{bmatrix}, \quad \rho_{\max} = 0.942$$

Table S1: Assumed fragility curves (PGA–Damage State) for the implementation example

Site id	β $(\ln, g)$	θ_1 (Slight) (g)	θ_2 (Moderate) (g)	θ_3 (Extensive) (g)	θ_4 (Complete) (g)
S_1	0.4	0.10	0.12	0.21	0.36
S_2	0.4	0.12	0.19	0.37	0.60
S_3	0.4	0.10	0.14	0.26	0.47
S_4	0.4	0.19	0.31	0.64	1.49

S3.2.1 Pre-processing

p-1) Calculate c^* using Equation 27:

$$c^* = \sqrt{1 + (1 - \rho_{\max})\frac{\phi^2}{\beta^2}} = \sqrt{1 + (1 - 0.942)\frac{0.6^2}{0.4^2}} = 1.064$$

p-2) First, construct $\mathbf{Cov}(\mathbf{g}_k) = \mathbf{I} + (\mathbf{B}\boldsymbol{\tau})(\mathbf{B}\boldsymbol{\tau})^\mathsf{T} + \mathbf{BFCBF}$:

$$\mathbf{Cov}(\mathbf{g}_k) = \begin{bmatrix} 4.250 & 3.119 & 2.666 & 1.487 \\ 3.119 & 4.250 & 2.657 & 1.516 \\ 2.666 & 2.657 & 4.250 & 1.502 \\ 1.487 & 1.516 & 1.502 & 4.250 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} + \begin{bmatrix} 2.250 & 2.119 & 1.666 & 0.487 \\ 2.119 & 2.250 & 1.657 & 0.516 \\ 1.666 & 1.657 & 2.250 & 0.502 \\ 0.487 & 0.516 & 0.502 & 2.250 \end{bmatrix}$$

Then, obtain the top t eigenvalues and eigenvectors of $\mathbf{Cov}(\mathbf{g}_k)$. In this example, we set $t = 2$:

$$\mathbf{u}_1 = \begin{bmatrix} 0.546 \\ 0.547 \\ 0.520 \\ 0.364 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -0.238 \\ -0.224 \\ -0.167 \\ 0.930 \end{bmatrix} \qquad \lambda_1 = 10.900, \quad \lambda_2 = 3.236$$

Finally, $\mathbf{W}_t^* = \mathbf{U}_t\mathbf{\Sigma}_t$:

$$\mathbf{U}_t = \begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2\end{bmatrix} = \begin{bmatrix}0.546 & -0.238\\ 0.547 & -0.224\\ 0.520 & -0.167\\ 0.364 & 0.930\end{bmatrix}, \qquad \mathbf{\Sigma}_t = \begin{bmatrix}\sqrt{\lambda_1 - (c^*)^2} & 0\\ 0 & \sqrt{\lambda_2 - (c^*)^2}\end{bmatrix} = \begin{bmatrix}3.125 & 0\\ 0 & 1.450\end{bmatrix}$$

$$\mathbf{W}_t^* = \begin{bmatrix}0.546 & -0.238\\ 0.547 & -0.224\\ 0.520 & -0.167\\ 0.364 & 0.930\end{bmatrix}\begin{bmatrix}3.125 & 0\\ 0 & 1.450\end{bmatrix} = \begin{bmatrix}1.707 & -0.345\\ 1.709 & -0.325\\ 1.624 & -0.242\\ 1.138 & 1.350\end{bmatrix}$$

S3.2.2 Simulation

s-1) Sample $t + N$ (=2+4=6) standard normal variables to create $\mathbf{x}$ and $\mathbf{z}$ (these vary each simulation):

$$\mathbf{x} = \begin{bmatrix}-0.298\\ 1.564\end{bmatrix} \qquad \mathbf{z} = \begin{bmatrix}1.463\\ 1.320\\ -0.885\\ 0.661\end{bmatrix}$$

s-2) Calculate $\boldsymbol{\gamma} = \mathbf{W}_t^*\mathbf{x} + c\mathbf{z}$:

$$\boldsymbol{\gamma} = \begin{bmatrix}1.707 & -0.345\\ 1.709 & -0.325\\ 1.624 & -0.242\\ 1.138 & 1.350\end{bmatrix}\begin{bmatrix}-0.298\\ 1.564\end{bmatrix} + 1.064 \times \begin{bmatrix}1.463\\ 1.320\\ -0.885\\ 0.661\end{bmatrix} = \begin{bmatrix}0.508\\ 0.387\\ -1.804\\ 2.475\end{bmatrix}$$

s-3) For each k, calculate $\mathbf{g}_k = \boldsymbol{\gamma} + \mathbf{b}_k$, where $\mathbf{b}_k = \mathbf{B}(\ln\boldsymbol{\theta}_k - \boldsymbol{\mu})$. The median ground motion $\boldsymbol{\mu}$ is 0.545 g, 0.533 g, 0.545 g, and 0.353 g for sites 1–4, respectively ($\ln Y = -2.35 + 0.57M - \ln R_{rup} - 0.0059R_{rup}$). For example, the four buildings' limit-state vector for damage state 1, $\mathbf{g}_1$, is

$$\mathbf{g}_1 = \begin{bmatrix}0.508\\ 0.387\\ -1.804\\ 2.475\end{bmatrix} + \begin{bmatrix}2.5 & 0 & 0 & 0\\ 0 & 2.5 & 0 & 0\\ 0 & 0 & 2.5 & 0\\ 0 & 0 & 0 & 2.5\end{bmatrix}\left(\begin{bmatrix}\ln(0.10)\\ \ln(0.12)\\ \ln(0.10)\\ \ln(0.19)\end{bmatrix} - \begin{bmatrix}\ln(0.545)\\ \ln(0.533)\\ \ln(0.545)\\ \ln(0.353)\end{bmatrix}\right) = \begin{bmatrix}-3.729\\ -3.342\\ -6.041\\ 0.930\end{bmatrix},$$

The remaining $\mathbf{g}_k$ vectors are

$$\mathbf{g}_2 = \begin{bmatrix}-3.274\\ -2.193\\ -5.200\\ 2.154\end{bmatrix}, \quad \mathbf{g}_3 = \begin{bmatrix}-1.874\\ -0.527\\ -3.653\\ 3.966\end{bmatrix}, \quad \mathbf{g}_4 = \begin{bmatrix}-0.527\\ 0.681\\ -2.172\\ 6.078\end{bmatrix}.$$

s-4) Form $\mathbf{g} = [\mathbf{g}_1,\ \mathbf{g}_2,\ \mathbf{g}_3,\ \mathbf{g}_4]$ and obtain the damage-state vector $\vec{d}$ by taking, for each row i, the maximum column index k such that $g_{ik} < 0$. (Rows are buildings; columns are damage states.)

$$\mathbf{g} = \begin{bmatrix} -3.729 & -3.274 & -1.874 & \mathbf{-0.527} \\ -3.342 & -2.193 & \mathbf{-0.527} & 0.681 \\ -6.041 & -5.200 & -3.653 & \mathbf{-2.172} \\ 0.930 & 2.154 & 3.966 & 6.078 \end{bmatrix}, \quad \vec{d} = \begin{bmatrix} 4 \text{ (Complete)} \\ 3 \text{ (Extensive)} \\ 4 \text{ (Complete)} \\ 0 \text{ (No damage)} \end{bmatrix}$$

S4 Invariant optimal t given spatial extent

We demonstrate that t remains invariant regardless of the total building count, provided the spatial extent of the portfolio remains consistent. Figures S5(a) and (b) present loss curves generated from randomly sampled subsets of 100 and 1,000 buildings for both SF downtown and Bay Area. The results—using $t = 1$ for SF Downtown and $t = 20$ for the Bay Area—consistently align with the benchmarks similar to the full-scale 15,386-building cases. This confirms that t is primarily a function of the number of building clusters, which is intrinsically linked to the spatial extent of the region.

This is well-reflected in the cumulative variance contribution plot (Figure S5 (c)). For SF downtown case, the variance from the first principal component of $\mathbf{W}^*$ (or $\mathbf{W}_1$) alone explains 96.98, 96.68, and 96.65% of the total variance of $\mathbf{g}_k$ for N of 100, 1,000, and 15,836, almost invariant (bluish curves of Figure S5 (c)). For the Bay Area case, the first 20 components, the requirement to achieve the error range less than $\sim 2.5\%$, explains 93.6, 92.0, and 91.8% of the total variance for N of 100, 1,000, and 15,836, still greater than 90%, explains most of the variances and almost invariant as well.

This N-invariant t is further supported by the exponential decay of eigenvalues of $\mathbf{Cov}(\boldsymbol{\omega})$ (Figure S5(d)), where the decay rate (α) remains constant of -1.5 in log-log domain across different values of N for a fixed spatial extent. Since α less than -1 guarantees the convergence of the sum of these eigenvalues as $N \to \infty$, and a first few eigenvalues dominate the total sum, the number of principal component required to achieve a certain variance contribution threshold remain invariant across N.

S5 Spatial Modes of Portfolio Damage

We now examine the eigenvectors ($\mathbf{u}_i$) corresponding to the largest eigenvalues (s_i^*) of $\mathbf{Cov}(\boldsymbol{\omega}^*)$ to elucidate how they characterize the variability of the damage function in a latent space and how the proposed framework captures the total variance using a limited number of components. Figure S6 presents the dominant eigenvectors for both cases. It should be noted that $\mathbf{Cov}(\boldsymbol{\omega}^*)$ and $\mathbf{Cov}(\mathbf{g}_k)$ share identical eigenvectors (Equation 25); thus, the eigenvectors visualized here represent the variation of the damage function $\mathbf{g}_k$ as well as $\boldsymbol{\omega}^*$.

In the SF Downtown case (Figure S6(a)), the eigenvectors emerge as spatial harmonics; as the eigenvalues decrease, the spatial frequency of the modes increases. The first component, $\mathbf{u}_1$, consists of nearly constant entries, indicating that the direction of maximum variance is the vector sum of all axes. This represents the collective damage variation across the entire portfolio. It also confirms that the dominant component of $\mathbf{g}_k$ is positively correlated, which is physically consistent with the ground motion correlations that drive the variance. The second and third components ($\mathbf{u}_2$ and $\mathbf{u}_3$) represent macro-scale variations in the east-west (E-W) and north-south (N-S) directions, respectively. For SF Downtown, the E-W component precedes the N-S component, suggesting that variance is more pronounced along the E-W axis, which aligns with the portfolio's elongated spatial distribution (Figure 9(a)). Higher-order modes ($\mathbf{u}_4$ and higher) continue to capture increasingly complex spatial variations.

To quantify the contribution of each component to the total variance of $\boldsymbol{\omega}$, the eigenvector coordinates are weighted by their corresponding eigenvalues, s_i^*. As shown in Figure S6(b), the weighted values for $i \geq 2$ are significantly smaller than $s_1^*\mathbf{u}_1$, and variations are not visible. This dominance of the first principal component explains why $t = 1$ is sufficient for the SF Downtown portfolio. The variations from $i \geq 2$ are negligible to the values compared to the largest component.

In the Bay Area case, the first component ($\mathbf{u}_1$) is likewise nearly constant (Figure S6(c)). However, unlike the SF Downtown case, the fundamental ($\mathbf{u}_2$) and higher-order ($\mathbf{u}_3$) variations along the NW-SE axis take precedence over the NE-SW direction. This reflects the significantly larger spatial extent along the NW-SE axis than NW-SE direction (Figure 9 (b)). Higher-order modes continue to capture more intricate spatial patterns. Regarding variance contribution, the lower-order components for the Bay Area exhibit non-negligible variation through $i = 8$ (Figure S6(d)). These higher-order modes contribute meaningfully to the total variability, necessitating a larger parameter ($t = 20$) for the metropolitan region.

References

[1] Nirmal Jayaram and Jack W Baker. Efficient sampling and data reduction techniques for probabilistic seismic lifeline risk assessment. *Earthquake Engineering & Structural Dynamics*, 39(10):1109–1131, 2010.

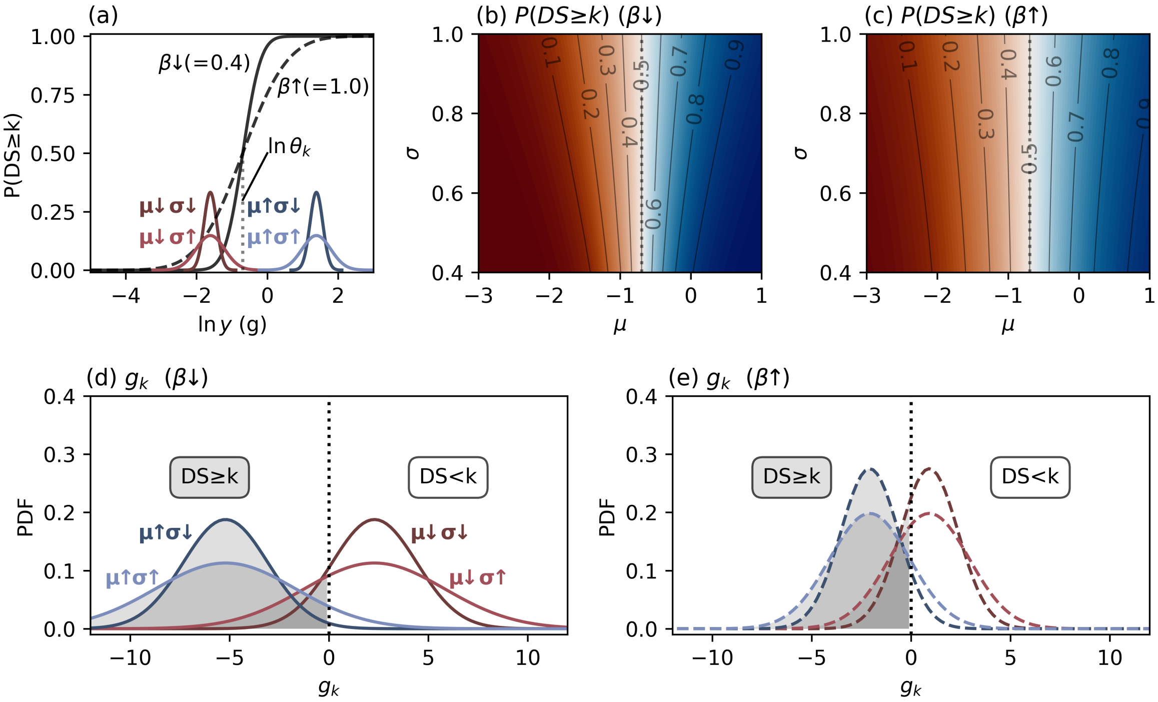

Figure S1: (a) Fragility curves corresponding to small (black solid line) and large (black dotted line) values of β, shown with four ground-motion intensity distributions: low μ with small (dark red) and large (light red) σ, and high μ with small (dark blue) and large (light blue) σ. (b) Variation of $P(DS \geq k)$ with respect to μ and σ for small β. The dotted vertical line indicates $\ln \theta_k$; therefore, the red shaded area corresponds to $\mu < \ln \theta_k$, and the blue shaded area corresponds to $\mu > \ln \theta_k$. (c) Same as (b) but for large β. (d) Distribution of g_k for the small-β case. The vertical dotted line indicates the limit state separating $DS \geq k$ from $DS < k$, and the shaded gray area represents $P(DS \geq k)$. The colors of the curves correspond to those defined in (a). (e) Same as (d), but for the large-β case.

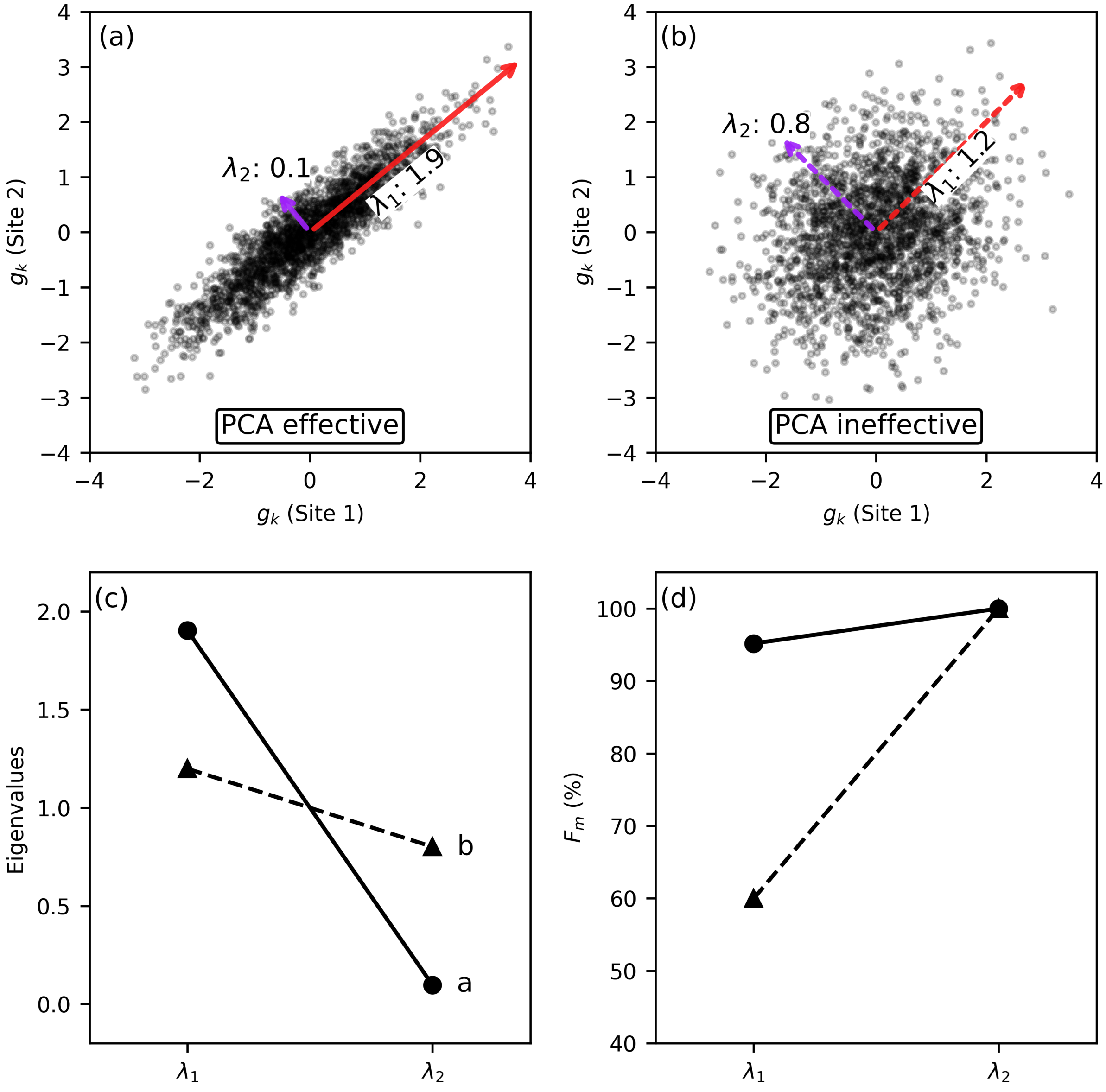

Figure S2: Eigenvalues and corresponding eigenvectors illustrating varying degrees of linear dependency of $\mathbf{g}_k$ between two sites: (a) high dependency, where PCA is effective; and (b) low dependency, where PCA is ineffective. Black dots represent random samples for each case. The red arrow illustrates the eigenvector corresponding to the larger eigenvalue (λ_1), while the purple arrow indicates the eigenvector corresponding to the smaller eigenvalue (λ_2). Note that the lengths of the arrows are not to scale with the magnitude of the eigenvalues. (c) Eigenvalue spectrum comparing the magnitudes of λ_1 and λ_2 for the two cases. Circles and triangles correspond to the cases in panels (a) and (b), respectively. (d) The cumulative variance contribution (F_m) for the two cases.

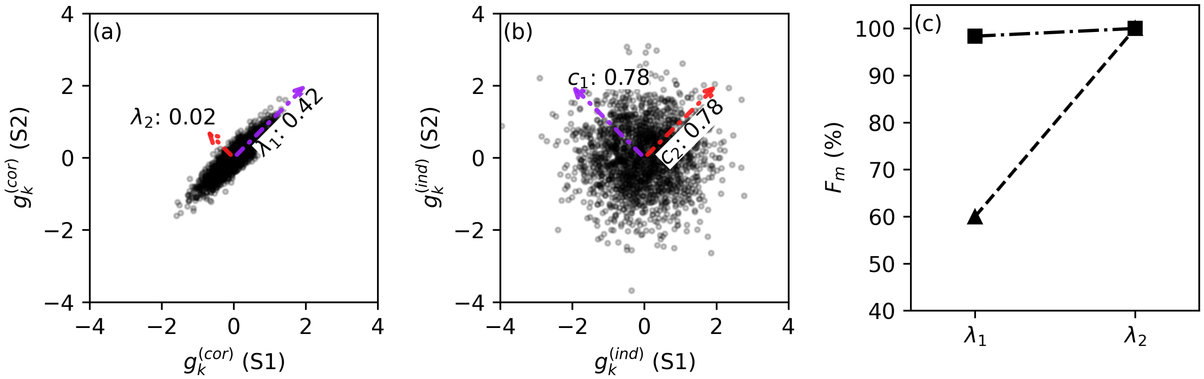

Figure S3: Probabilistic PCA decomposition of data shown in Figure S2 (b) into (a) highly correlated and (b) independent noise components. Black dots represent random samples; the red arrow indicates the eigenvector corresponding to the dominant eigenvalue (λ_1), while the purple arrow indicates the eigenvector corresponding to the secondary eigenvalue (λ_2). Note that the arrow lengths are not scaled to the eigenvalue magnitudes. (c) Cumulative variance contribution (F_m) for the two cases. The triangles and squares correspond to the scenarios presented in Figure S2 (b) and (a)+(b) (Probabilistic PCA), respectively.

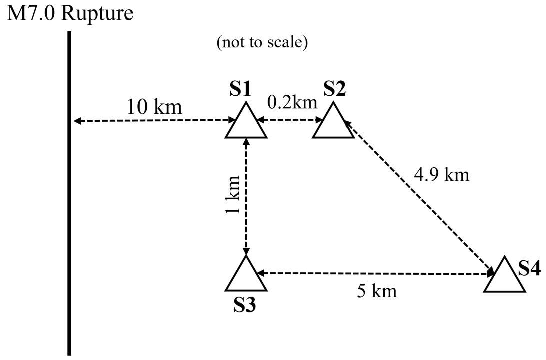

Figure S4: Assumed source and site geometry for demonstrating the implementation of the proposed framework.

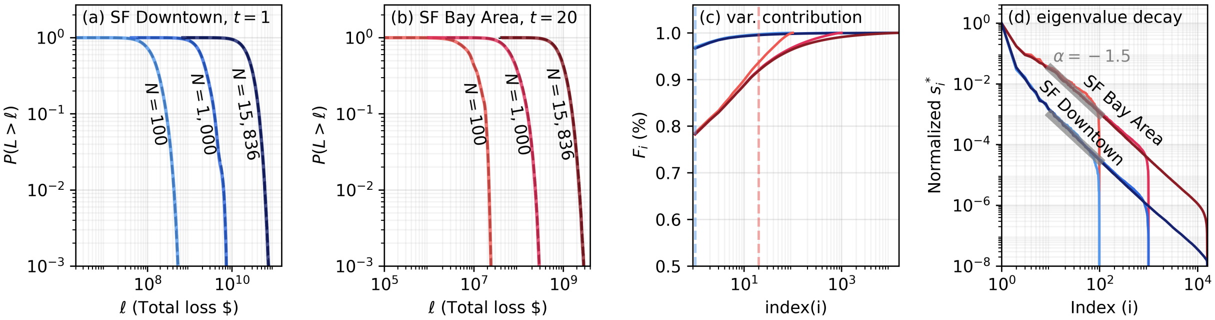

Figure S5: Loss curves with different numbers of buildings in (a) SF Downtown ($t = 1$) and (b) SF Bay Area ($t = 20$). The curves using the traditional framework are black dotted lines; The curves using the proposed framework with N of 100, 1,000, and 15,836 are blue (Downtown) and red (Bay Area), with lighter shades for smaller samples. The proposed curves closely match benchmarks regardless of the number of buildings. (c) The cumulative variance contribution as a function of the number of principal components. (d) Normalized eigenvalues for each case (blue: Downtown; red: Bay Area). The eigenvalue decay rate is consistent within a given region, independent of building density.

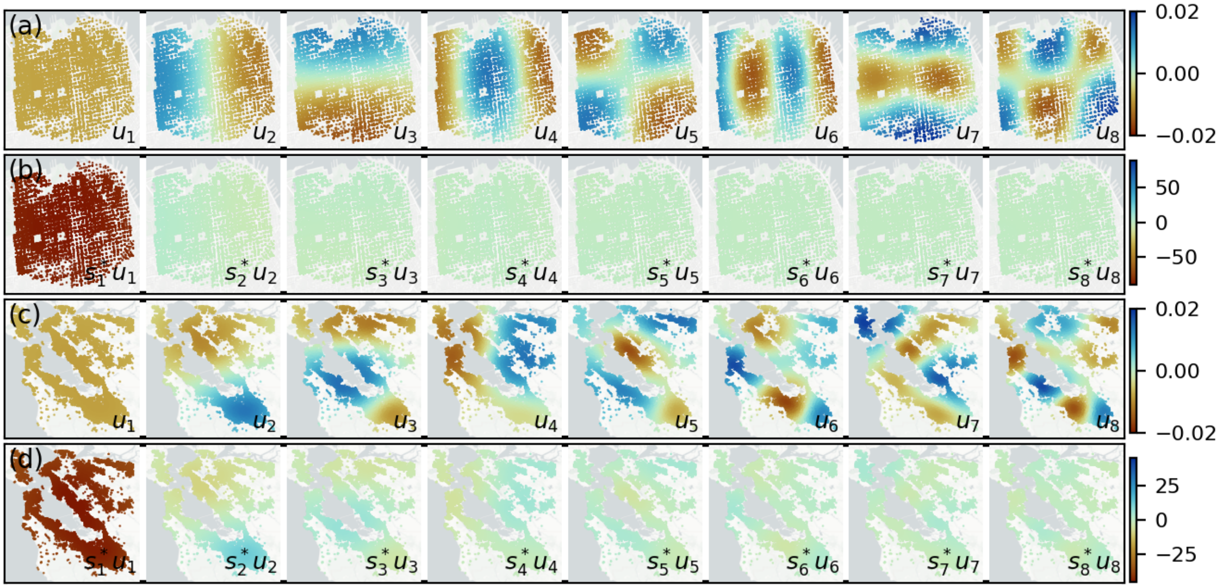

Figure S6: (a, c) Eigenvector coordinates for the eight largest eigenvalues (s_i^*; in descending order) for SF Downtown and Bay Area, respectively. (b, d) Eigenvectors weighted by their corresponding eigenvalues ($s_i^* \mathbf{u}_i$). The colorbars of (b) and (d) are set to be the maximum absolute value for $i = 1$ case. The map is not drawn to scale in latitude and lognitude