Received: 15 October 2023 | Revised: 5 March 2024 | Accepted: 8 March 2024

DOI: 10.1002/eqe.4116

RESEARCH ARTICLE

WILEY

Surrogate model-aided global sensitivity analysis framework for seismic consequences estimation in buildings

Jiajun Du[1,2] | **Wei Wang**[1,2]

[1]State Key Laboratory of Disaster Reduction in Civil Engineering & Department of Structural Engineering, Tongji University, Shanghai, China

[2]Shanghai Engineering Research Center for Resilient Cities and Intelligent Disaster Mitigation, Tongji University, Shanghai, China

Correspondence
Wei Wang, Department of Structural Engineering, College of Civil Engineering, Tongji University, 1239 Siping Road, Shanghai 200092, China.
Email: weiwang@tongji.edu.cn

Funding information
Natural Science Foundation of China, Grant/Award Numbers: 52378182, 52078366; Top Discipline Plan of Shanghai Universities-Class I, Grant/Award Number: 2022-3-YB-18; Shanghai 2022 Science and Technology Innovation Action Plan Social Development Science and Technology Research Project, Grant/Award Number: 22dz1201700

Abstract

Seismic consequences estimation for individual buildings is valuable for various stakeholders, including government entities, building owners, and insurers. The robustness of estimation results in the presence of incomplete input information can typically be investigated through sensitivity analysis. However, the estimation process's complexity and the sensitivity analysis's computational burden hinder its practical application, which requires a more efficient procedure to facilitate broader use. This paper proposes a novel framework for sensitivity analysis of seismic consequences estimation to improve the efficiency and reliability of such analysis. The proposed approach encompasses three key components: (1) stochastic ground motion modeling (SGMM)-based seismic consequences estimation to evaluate the economic, environmental, and social consequences given specific buildings by considering different hazard levels, (2) the training of surrogate model (Gaussian process model) for structural analysis to reduce the computational cost of the evaluation process, and (3) variance-based global sensitivity analysis to investigate the importance of parameters of concern in the estimation process. The entire procedure is implemented in Python, adhering to object-oriented programming, and does not rely on external software. Then, the proposed methodology is applied to two distinct three-story steel moment-resistant frames (SMRFs) subjected to four different hazard levels to demonstrate its effectiveness. The SGMM method can generate specific ground motions for each hazard level, mitigating the potential for result bias from using ground motions with unrealistic characteristics. Furthermore, the SGMM method is particularly suitable for automated analysis processes reducing the laborious task of screening ground motions from the database. Comparative analysis with surrogate-free estimation reveals that the surrogate-based analysis delivers reliable results with significantly reduced computational cost. The results of analyzing different structures under varying hazard levels reflect the variability in sensitivity analysis of consequences estimation, highlighting the necessity of the proposed flexible and efficient framework. Furthermore, the proposed framework's advantages, limitations, and future research needs are discussed.

KEYWORDS
Gaussian process model, global sensitivity analysis, seismic consequences estimation, stochastic ground motion modeling, surrogate model

1 | INTRODUCTION

As urbanization continues to gain momentum, there is a notable concentration of population, wealth, and social activities. This phenomenon also implies that the potential threat of earthquake disasters on human communities could be significantly more severe. In recent decades, several medium or high-intensity earthquakes have resulted in unexpected and unacceptable consequences.[1,2] For instance, in the 2011 Christchurch earthquake, most buildings in the city's central business district were forced to rebuild due to severe damage despite meeting available seismic codes and not collapsing.[3] The design philosophy that previously focused on the life safety of the occupants appears to be no longer sufficient to meet the stringent demands of contemporary structural excellence.[4] Instead, it should focus more on specific seismic consequences, such as monetary losses, carbon emissions, and recovery time. The structural engineering community has recognized these issues and is dedicated to developing methodologies to appropriately estimate the potential seismic consequences during the building design phase to guide the real seismic resilient building design.

To date, the second-generation Performance-Based Earthquake Engineering (PBEE-2)[5] concept pioneered by the Pacific Earthquake Engineering Research (PEER) Center stands as the most robust concept for evaluating probabilistic seismic consequences. In contrast to some earlier technical documents, such as the Vision 2000 report,[6] ATC-40,[7] FEMA-273,[8] and FEMA-356,[9] PBEE-2 employs performance measures of direct interest to stakeholders, such as monetary losses, as opposed to solely focusing on structural response or damage states. Subsequently, leveraging PBEE-2 as a pivotal foundation, methodologies like HAZUS,[10] FEMA P-58,[11] and Arup's REDi[12] accompanied by their respective databases were developed and widely adopted. Of these methods, FEMA P-58 has been commonly applied in disaster mitigation,[13–15] structural optimization,[16,17] decision-making,[18,19] and so on. In recent developments, specific approaches have further enhanced FEMA P-58 to facilitate the quantification of post-earthquake functional recovery, such as the F-Rec framework,[20] ATC-138,[21] and TREADS.[22]

In the framework of PBEE-2, the seismic consequence analysis process is modularized into four parts: seismic hazard analysis, structural analysis, damage analysis, and loss analysis.[5] In this manner, the impact of seismic risk on buildings is progressively transformed into probabilistic performance measures of interest. The structural analysis typically requires nonlinear dynamic time-history analysis at different hazard levels to obtain engineering demand parameters (EDPs).[11] Therefore, selecting appropriate motion records that match the intensity measure (IM) values corresponding to different hazard levels is necessary. Due to the limited motion records for design scenarios, records from locations distinct from the project site are usually selected and then modified to different IM levels by scaling or spectrum matching.[23] This approach, though typical, is controversial for its potential to produce unrealistic seismic characteristics.[24] Relevant studies have revealed the potential bias that may rise from the use of scaled motions in structural demand analysis,[24,25] fragility analysis,[26] and so forth, compared to unscaled motions due to such unrealistic representation. Stochastic ground motion modeling (SGMM) is regarded as an alternative solution to circumvent this problem.[27] Such models modulate stochastic sequences by applying functions that address the spectral and temporal characteristics of the recorded ground motion to generate synthetic ground motion samples with realistic characteristics. It is noteworthy that there are ongoing debates concerning the SGMM methods, particularly regarding potential discrepancies in high-energy content and long-period components[27] when compared to recorded ground motions. This necessitates careful validation of various SGMM methods. Despite these concerns, the need to simulate realistic ground motions has led to extensive research into SGMM. In the case of their application to PBEE, Lamprou et al.[28] adopted a point-source SGMM in seismic economic loss estimation. Nevertheless, compared with their wide application in seismic risk and structural analysis,[29–31] application in seismic consequence estimation still needs further exploration.

The seismic consequences estimation is a complex issue that involves numerous sources of uncertainty, including but not limited to ground motions, modeling assumptions, and fragility data.[32] Especially during the design phase, design information may be crude and incomplete. Sensitivity analysis can effectively help evaluate the relative importance of various uncertainty sources by identifying their contributions to the overall uncertainty. Specifically, it may be feasible to reduce computational expense by adopting best estimate values for variables that make minor contributions to overall

uncertainty rather than using statistic distribution. Meanwhile, significantly contributing variables can be prioritized for further study to reduce their uncertainty in the future.[33]

However, compared with substantial research focused on earthquake consequence estimation methods, corresponding sensitivity analysis research is limited. Porter et al.[32] performed a sensitivity analysis for a 1960s high-rise nonductile moment frame building to evaluate the contribution of input variables to seismic economic loss. Dyanati et al.[34] investigated the influence of earthquake events, EDP, and hazard calculation formulation selection on the seismic performance of a six-story steel-braced frame building. Although these studies provided insights into the impact of critical factors in PBEE-2, they focus on using local rather than global sensitivity analysis (GSA) methods. Local analysis only evaluates model sensitivity to individual input changes locally, whereas GSA examines model sensitivity across all potential input values globally.[35] For example, Lamprou et al.[28] presented an advanced stochastic sampling-based GSA to assess the impact of uncertain earthquake characteristics and damage parameters on seismic repair costs for a four-story concrete moment-frame building. Cremen and Baker[36] utilized the variance-based GSA to examine the impact of high-level input parameters (e.g., building age, occupancy type) on the seismic economic loss ratio and re-occupancy time. Nevertheless, there is still a lack of knowledge and application on GSA of seismic consequences estimation methods like FEMA P-58 or other advanced methods.

The broad application of GSA in seismic consequences estimation is impeded by two key factors. First, there is a deficiency in an integrated and adaptable workflow. Seismic consequence analysis is inherently rife with sources of uncertainty, including variations in design information for different structural solutions. Consequently, sensitivity analysis outcomes can vary across diverse scenarios, necessitating specific sensitivity analyses tailored to each situation. In such circumstances, the need for a unified and flexible workflow becomes imperative. This paper addresses this challenge by harnessing the principles of object-oriented programming and employing modular seismic consequences estimation methods like FEMA P-58.

Another factor is the formidable computational burden. When multiple factors are under consideration, the need for numerous evaluations to ensure the convergence of sensitivity analysis outcomes results in a substantial computational load. To address this issue, a promising solution emerging in structural engineering involves the application of surrogate models. These models serve to reduce evaluation costs by substituting the partial or complete evaluation model. For example, Giudice et al.[37] utilized polynomial chaos expansion to replace the 3D printing process and subsequently conducted a sensitivity analysis on some key parameters in additive manufacturing. Tesfamariam et al.[38] developed a predictive model of maximum and residual drifts for timber-steel hybrid buildings based on the Gaussian process (GP) model. Anwar and Dong[39] used the GP model as a substitute for PBEE evaluation, thereby facilitating the multi-objective optimization of the community buildings portfolios. Nevertheless, existing research has yet to explore the utility of surrogate models in the GSA of seismic consequence analysis. Furthermore, whether surrogate models yield reliable sensitivity analysis results warrants further investigation.

Nowadays, there is much valuable literature available on the application of surrogate models to earthquake loss estimation (e.g., structural analysis, damage analysis, loss analysis). For example, in the area of structural response prediction, Solelmani-Babakamali and Esteghamati[40] leverage deep learning techniques to estimate seismic demand models through nonlinear pushover analysis. Luo and Paal[41] integrate ML with hysteresis models for predicting the lateral stiffness of structural components. Gudipati and Cha[42] explore three different surrogate models (support vector machines, kriging, and neural networks) for structural response prediction of a building class. Suarez et al.[43] utilize GP regression to build surrogate models for response prediction to facilitate the direct loss-based seismic design of low-rise base-isolated structures. Fayaz et al.[25] use an interpretable GP regression to assess the impact of ground motion amplitude scaling on the EDPs estimation. In the area of damage analysis, Nguyen et al.[44] utilize various machine learning (ML) techniques to predict the damage states of steel moment frames and achieve commendable prediction accuracy. Li et al.[45] construct surrogate models for efficient regional seismic damage risk assessment using deep generative learning. Gentile and Galasso[46] facilitate the seismic fragility assessment of building portfolios by leveraging GP regression. To better serve PBEE-based design directly, Esteghamati and Flint[47] develop data-driven surrogate models to map design parameters with performance indicators (e.g., seismic-induced consequence) for performance-based assessment of mid-rise RC frame buildings. The primary goal of employing ML to construct surrogate models, as illustrated in these studies, is to enhance efficiency and potential accuracy compared to traditional methods. This paper also aligns with this objective. In such existing research, high-dimensional input features, such as the number of floors, are often required to build the ML models, and the output is mostly the maximum response of the entire structure, which is difficult to refine to the response of each floor. Besides, they pursue strong generalization capabilities in the expectation of obtaining applicability to various types of building arrangements and hazard scenarios, and thus often require a very large amount of model

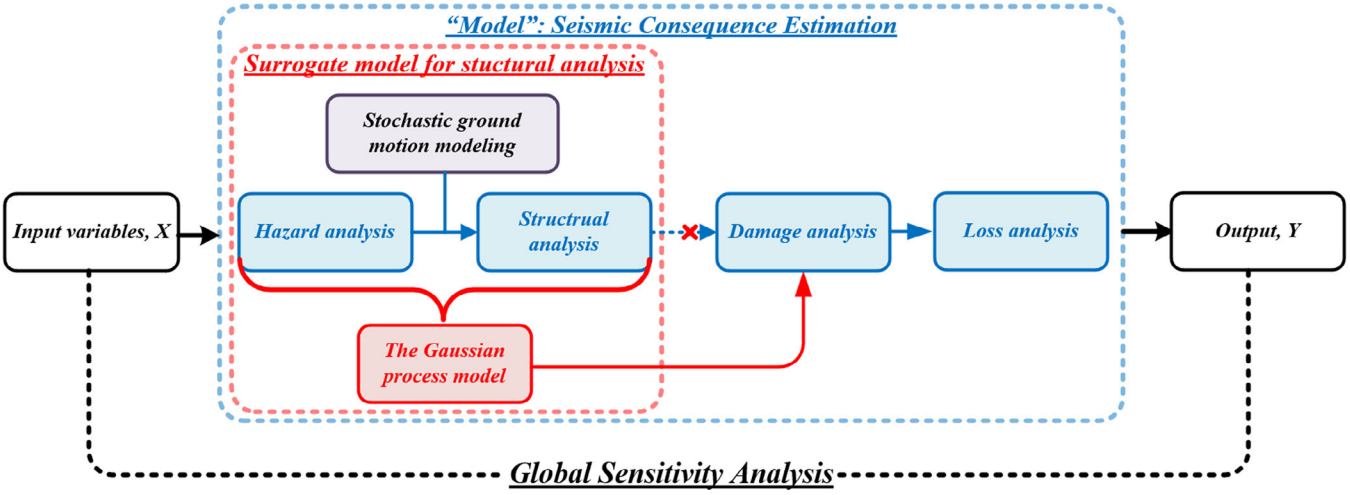


FIGURE 1 The proposed sensitivity analysis framework.

training data. However, despite the generalization capabilities demonstrated by these models in their respective focused issues, they may not necessarily satisfy the need for flexible parameter selection in sensitivity analysis which is the focus of this paper. For instance, the effects of damping ratios, material properties, and so forth, may be of more interest in structure-specific sensitivity analyses than high-dimensional parameters (e.g., number of floors). Furthermore, almost all these models only give the maximum response of the entire structure, rather than detailing the response of each floor required for seismic loss estimation. For surrogate models that can directly output seismic damage risk or loss, although they can serve performance assessment more directly, their modeling is generally based on a set of recommended parameters (e.g., Hazus-MH guidelines[48]), and their input features do not incorporate the key parameters of the loss estimation, and thus cannot be applied to various sensitivity analyses. Therefore, a surrogate model-building strategy that is more suitable for the characteristics of structure-specific sensitivity analysis needs to be further investigated.

Based on the above, this paper presents a surrogate model-aided GSA framework for computationally expensive seismic consequence estimation. First, a GP model is constructed through a series of nonlinear time-history analyses (NTHA) to replace the most computationally expensive part of the seismic consequence analysis, that is, structural analysis. Notably, the surrogate model is built for structure-specific structural analysis, and the input features are picked to serve sensitivity analysis. To facilitate an efficient workflow and reduce bias in results, an SGMM approach is utilized to generate the ground motions required for each design scenario in NTHA. Then, the computational cost-reduced seismic consequence estimation is used to perform a variance-based GSA. The proposed framework is implemented as object-oriented programming in Python and forms an integrated workflow in a modular form, which is flexible and extensible form. The framework is later illustrated on two different three-story steel moment-resisting frames (SMRFs) to demonstrate its feasibility, flexibility, and effectiveness. Intensity-based (i.e., four different hazard levels with 43-, 475-, 975-, and 2475-year return period) seismic consequences estimations are considered. The results of the surrogate model-aided analysis are compared with those of the surrogate-free model to verify the effectiveness of the proposed framework.

2 | PROPOSED SURROGATE MODEL-AIDED SENSITIVITY ANALYSIS FRAMEWORK FOR SEISMIC CONSEQUENCES ESTIMATION

The proposed surrogate model-aided sensitivity analysis framework for seismic consequences estimation is shown in Figure 1. The framework is organized into three main parts: (1) SGMM-based seismic consequence estimation, (2) surrogate model for structural analysis, and (3) GSA.

The seismic consequence estimation part, the "model" for sensitivity analysis in this paper, can consider the possible consequences under various seismic hazard scenarios in a specific region. This part is generally structured in four steps: (1) hazard analysis; (2) structural analysis; (3) damage analysis; (4) loss analysis. The seismic hazard analysis provides a reasonable evaluation of the possible seismic hazard scenarios that may occur on the site of the building, as well as their corresponding intensity levels and probability of occurrence. The SGMM technique generates ground motion records

consistent with site characteristics and selected intensity levels for further use in subsequent NTHA. Compared with the prevailing ground motion selection method, the SGMM method is more efficient and suitable for automatic workflow. Moreover, synthetic ground motions can be used without scaling or with a relatively small scale factor, thus avoiding the response bias caused by unrealistic ground motions. Then, a finite element model capable of capturing nonlinear behavior is built to obtain the structural response (i.e., EDPs) under seismic excitation. Subsequently, the obtained EDPs are mapped to every considered building component, including structural and nonstructural components. In damage analysis, each component's damage state and corresponding probability at given EDPs can be derived from empirical fragility functions obtained from experimental tests or professional judgment. Finally, repair measures and contractor quotes used to repair damaged components can be used to predict the final consequences, such as monetary loss.

The most computationally expensive part of the "model" is generally the nonlinear time history analysis (NTHA), and to obtain converged sensitivity analysis results, this process often needs to be executed thousands of times. To reduce such computational burden, the NTHA part is surrogated by a surrogate model (GP model). Establishing a surrogate model still requires NTHAs, but the amount is relatively small. The computational cost of sensitivity analysis with such a well-trained surrogate model is almost negligible compared to using the original model directly. It is worth noting that the strategy of this paper is not to establish a surrogate model with strong generalization ability, that is, it does not pursue the ability to predict the response to high-dimensional input features such as different numbers of stories or bays. Instead, a dedicated surrogate model is built for specific individual buildings in specific hazard scenarios. The advantage of this move is that a highly accurate surrogate model can be obtained with a small computational cost. Moreover, this strategy is more consistent with the requirements of sensitivity analysis. The input features of the surrogate model can be adjusted accordingly as the types of structural parameters considered in the sensitivity analysis change.

In the sensitivity analysis, the engineer should first select the parameters to be investigated based on the level of detail available with the design information. Sampling the ranges and estimated distributions of these parameters, inputting them into the model, and then observing the degree of change in the model output to determine the model's sensitivity to these parameters provides a more in-depth view for model inspection. Note that the sensitivity analysis results are relative and will vary with the type and number of parameters investigated. In addition, when the indicators used to characterize the model output are different (such as mean or median value), the results will also differ.

An object-oriented programming structure for GSA of seismic consequence estimation is presented to organize an integrated and flexible workflow, as shown in Figure 2. The overall programming structure is implemented in Python without needing other external software. Each critical part of the proposed sensitivity analysis framework is written as a Python class. The SGMM module generates several scenario-specific ground motions for a given hazard scenario (site and earthquake characteristics). The structural analysis module, implemented in *OpenSeesPy*[49] (the Python version of seismic structural analysis software OpenSees[50]), enables fast and automatic construction of structural models based on the given geometric and loading information. More specifically, for the same structure type, such as SMRF, the construction process for OpenSees models is almost consistent. Variations such as the number of stories, floor height, and component sections, can all serve as input parameters for parametric modeling. Once the parametric modeling process is developed for the interested structure type, the corresponding analytical model for a specific building can be efficiently constructed by inputting specific parameters (e.g., floor height). This facilitates engineers' rapid modeling of archetype structures, thus avoiding repetitive modeling of different design solutions. The benefits and effectiveness of such an automatic model concept have been shown in previous studies.[51,52]

This paper takes SMRFs as the investigated object. The concentrated plasticity model is used to capture the frame's nonlinear behavior, which comprises linear elastic beam and column elements connected with nonlinear elements intended to localize plastic deformations, as illustrated in Figure 3. Nonlinear behavior is captured by the plastic hinges at the ends of beam and column elements and in the joint panel zones. The modified IMK material model[53,54] is used to model the former, and corresponding parameters are derived according to Refs.[55–57] Note that elastic analysis (i.e., the *ElasticAnalysis* class in the structural analysis module) needs to be performed first to obtain the key parameter "axial compression ratio" in constructing the column hinges. In the joint panel zone model, eight elastic rigid elements are employed to form a parallelogram consistent with the dimension of the actual joint. To account for potential shear yielding, as per Krawinkler's approach,[58] a trilinear rotational spring is positioned at the upper right corner, whereas the other three corners are represented as pinned connections. Besides, the P-Δ effect is considered by constructing a leaning column connected by rotational stiffness-negligible springs and connected to the resisting frames through rigid truss elements. The floor mass is uniformly added to each node on the same floor. Gravity loads on the SMRF are applied uniformly to the beam elements, whereas loads on the unmodeled gravity frames are applied as concentrated forces on the leaning column. Notably, the application of this framework to other structures requires the expansion of the structural analysis module, that is, the

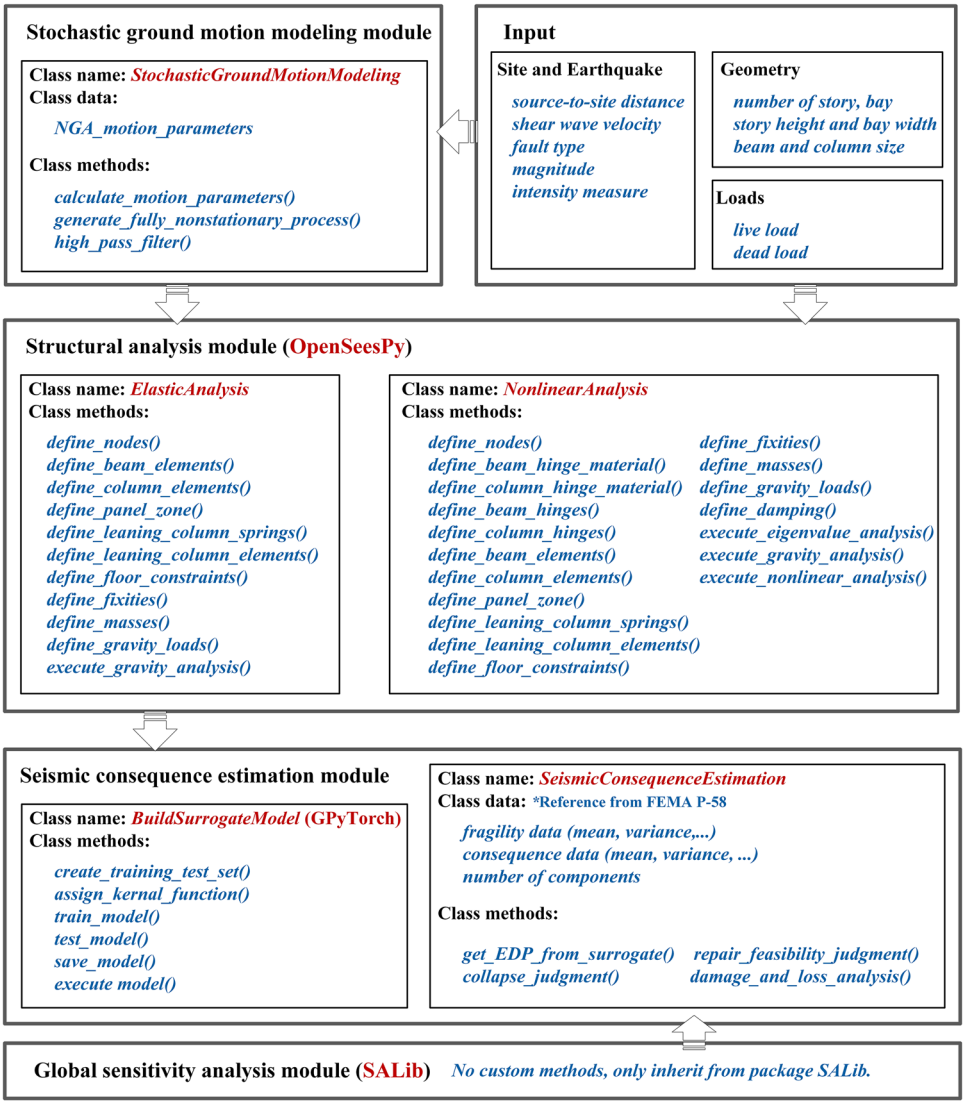


FIGURE 2 Programming structure of the proposed global sensitivity analysis process.

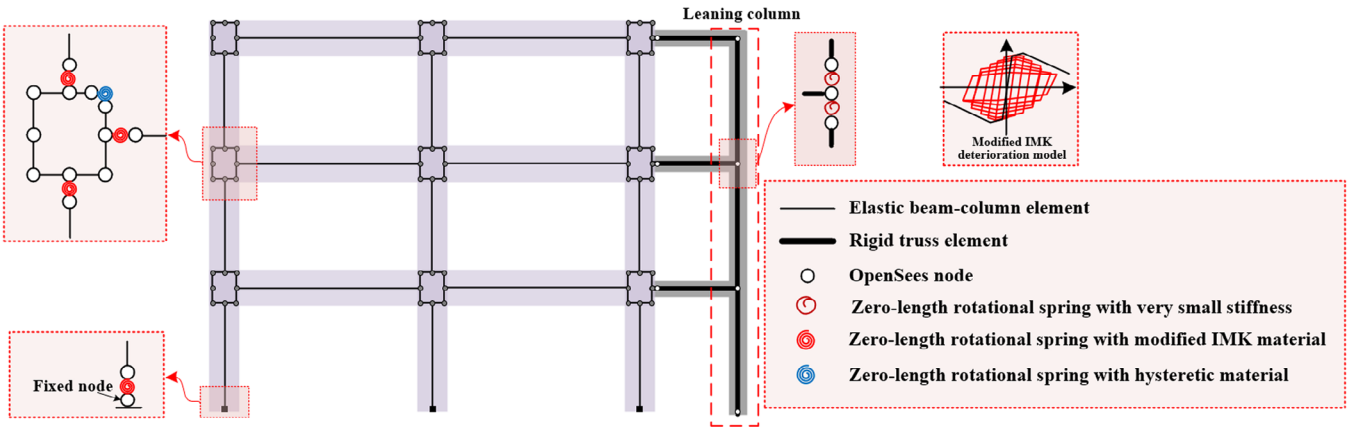


FIGURE 3 Structural model for steel moment resisting frames (SMRFs) in OpenSeesPy.

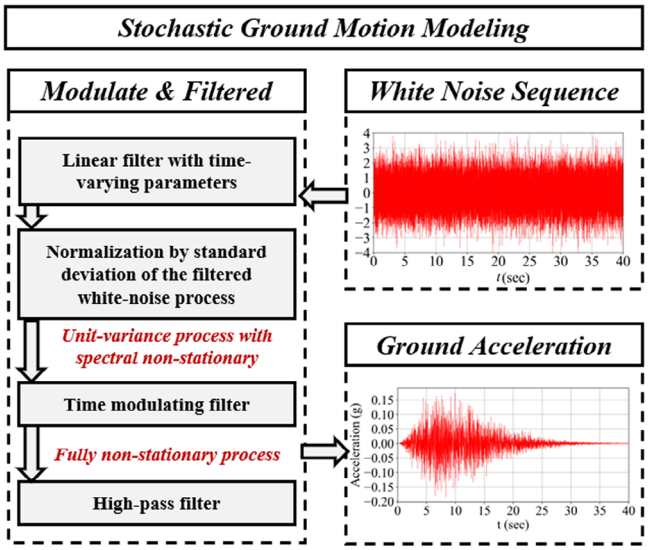


FIGURE 4 The procedure of stochastic ground motion modeling.

development of specific automated structural model construction approaches for different structural types. Corresponding construction methods for some structures (such as RC moment resisting frames,[47] and concrete frames with infills[52]) have been proposed and can be used to extend the framework proposed in this study. More different structural system types (e.g., steel braced frame, concrete frame), other modeling methods (e.g., fiber section model), or 3D models can be considered in the future.

The seismic consequence estimation module comprises the *BuildSurrogateModel* class and the *SeismicConsequenceEstimation* class. The former can call the motion modeling and structural analysis modules to build a dataset and then use *GPyTorch*[59] to perform tasks such as training, testing, saving, and executing models. The latter embeds the fragility data and consequence data provided by the FEMA P-58 document, converting the predicted EDPs obtained by the surrogate model into possible seismic consequences. Finally, the *SALib*[60,61] is used for sensitivity analysis on the constructed seismic consequence estimation module.

The steps involved in the SGMM, surrogate model, and GSA methods parts are further elaborated in the subsequent sections.

3 | STOCHASTIC GROUND MOTION MODELING-BASED SEISMIC CONSEQUENCE ESTIMATION

3.1 | Stochastic ground motion modeling method

Synthetic ground motion models can be broadly categorized into three main types: those employing physics-based seismological models to describe source mechanisms and wave propagation paths, parameterized stochastic models tailored to recorded ground motions, and hybrid models that combine the former two approaches.[62] Of these three types, the seismological models require extensive knowledge and computations, and their applicability is limited in data-scarce regions due to regional parameter variations. Therefore, the model employed in this study belongs to the second type and was developed by Rezaeian and Kiureghian,[27] under the auspices of the PEER Center.[63]

In the selected model, the white-noise process $\omega(\tau)$ passes through linear filtering with time-varying parameters, standard-deviation normalization, and time modulation successively, thereby obtaining a fully nonstationary process $x(t)$, as briefly shown in Figure 4. Its continuous form can be expressed as

$$x(t) = q(t,\alpha)\left\{\frac{1}{\sigma_h(t)}\int_{-\infty}^{t} h[t-\tau,\lambda(\tau)]\omega(\tau)d\tau\right\} \quad (1)$$

where $q(t, \boldsymbol{\alpha})$ denotes a non-negative, deterministic, time-modulating function that controls the temporal nonstationary. The entire term inside the brace is a unit-variance process that controls the spectral nonstationary. $h[t\text{-}\tau, \boldsymbol{\lambda}(\tau)]$ is the impulse-response function of the filter with time-varying parameters $\boldsymbol{\lambda}(\tau)$. The $\sigma_h(t)$ is the variance of the integral term. Through the product of temporal and spectral nonstationary control terms, a fully nonstationary process $x(t)$ is obtained. More specifically, the gamma function is chosen as $q(t, \alpha)$:

$$q(t, \alpha) = \alpha_1 t^{\alpha_2 - 1} \exp(-\alpha_3 t) \tag{2}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$. Both α_1 and α_3 are positive, which, respectively, control the intensity and duration of the process; $\alpha_2 > 1$, controls the shape of $q(t, \boldsymbol{\alpha})$. For the filter impulse-response function, a form related to the pseudo-acceleration response of a single-degree-of-freedom linear oscillator is adopted, as Equation (3).

$$h[t - \tau, \boldsymbol{\lambda}(\tau)] = \begin{cases} \frac{\omega_f(\tau)}{\sqrt{1-\zeta_f^2(t)}} \exp[-\zeta_f(t)\omega_f(\tau)(t-\tau)] \times \sin[\omega_f(\tau)\sqrt{1-\zeta_f^2(t)}(t-\tau)], \tau \le t \\ 0, \tau > t \end{cases} \tag{3}$$

where $\boldsymbol{\lambda}(\tau) = (\omega_f(\tau), \zeta_f(\tau))$ denotes the time-varying parameters. Of these two parameters, $\zeta_f(\tau)$ denotes the damping ratio of the filter, which controls the bandwidth of the process, and $\omega_f(\tau)$ denotes its frequency, which controls the evolutionary predominant frequency.

To facilitate numerical calculation, the discrete form of Equation (1) can be expressed as

$$\hat{x}(t) = q(t, \alpha) \sum_{i=1}^{n} \{s_i(t, \lambda(t_i)) u_i\} \tag{4}$$

$$s_i(t, \lambda(t_i)) = \begin{cases} \frac{h[t-t_i, \lambda(t_i)]}{\sqrt{\sum_{j=1}^{k} h^2[t-t_j, \lambda(t_j)]}}, t_k \le t < t_{k+1} \\ 0, t < t_i \end{cases} \quad \text{when } 0 < i \le k,\ 1 \le k \le n-1 \tag{5}$$

where $t_i, i = 0, 1, \ldots, n$, and $t_0 = 0$, t_n represents the last time point of the acceleration time history. u_i is a standard normally distributed random pulse at a discrete time point t_i. Finally, the nonstationary process $x(t)$ needs to be high-pass filtered. Without this step, generated ground motions may have nonzero residual velocity and displacement. Therefore, a critically damped oscillator is selected, as Equation (6).

$$\ddot{z}(t) + 2\omega_c \dot{z}(t) + \omega_c^2 z(t) = \hat{x}(t) \tag{6}$$

where ω_c denotes the frequency of the high-pass filter. In practice, $\omega_c/2\pi$ can be taken as a constant 0.1 Hz.

According to the above equation, the key to obtaining $\hat{x}(t)$ is to determine expressions of $q(t, \boldsymbol{\alpha})$ and $h[t - \tau, \boldsymbol{\lambda}(\tau)]$, that is, to determine the parameters $(\alpha_1, \alpha_2, \alpha_3)$ and $(\omega_f(\tau), \zeta_f(\tau))$. The former is related to $(\bar{I}_a, D_{5\text{–}95}, t_{\text{mid}})$. $\bar{I}_a$ denotes the expected Arias intensity of the process $\hat{x}(t)$, as defined in Equation (7). $D_{5\text{–}95}$ denotes the interval between the time points at which 5 and 95% $\bar{I}_a$ are reached, which represents the effect duration of $\hat{x}(t)$. And t_{mid} can be considered as the time point at which 45% $\bar{I}_a$ is reached, representing the middle of the strong-shaking phase.

$$\bar{I}_a = E\left[\frac{\pi}{2g}\int_0^{t_n} \hat{x}^2(t)dt\right] = \frac{\pi}{2g}\int_0^{t_n} q^2(t, \boldsymbol{\alpha})dt \tag{7}$$

Equation (2) states that $q^2(t, \boldsymbol{\alpha})$ is proportional to a gamma probability density function with variables $2\alpha_2 - 1$ and $2\alpha_3$. Then, the parameters α_2 and α_3 can be derived by Equations (8) and (9). The left side of the equation is motion-related parameters, and the right side is the time-modulating function-related parameters. Furthermore, α_1 can be obtained by Equation (10).

$$D_{5\text{–}95} = t_{95} - t_5 \tag{8}$$

$$t_{\text{mid}} = t_{45} \tag{9}$$

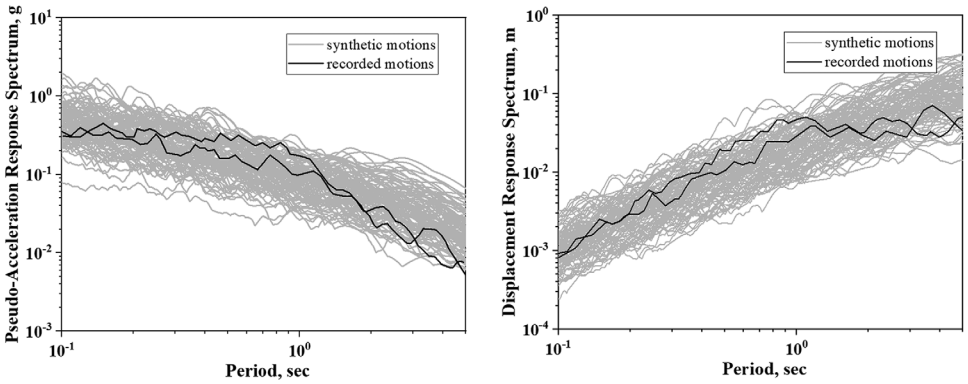


FIGURE 5 Elastic response spectra (5% damping ratio) of recorded and synthetic ground motions.

$$\alpha_1 = \sqrt{\bar{I}_a \frac{(2\alpha_3)^{2\alpha_2-1}}{\Gamma(2\alpha_2-1)}} \tag{10}$$

where t_p represents the gamma inverse cumulative distribution function value at p%. For parameters $\omega_f(\tau)$ and $\zeta_f(\tau)$, the following approximation Equations (11) and (12) can be adopted.

$$\omega_f(\tau) = \omega_{mid} + \omega'(\tau - t_{mid}) \tag{11}$$

$$\zeta_f(\tau) = \zeta_f \tag{12}$$

where ω_{mid} denotes the filter frequency at t_{mid}; ω' denotes the rate of change of the filter frequency with pulse application time. The relationship of the motion parameter set ($\bar{I}_a$, D_{5-95}, t_{mid}, ω_{mid}, ω', ζ_f) to site and earthquake characteristics (i.e., earthquake magnitude M, source-to-site distance R, shear-wave velocity of the site V, the type of faulting F) is developed through a subset of the NGA database. By this way, for given (M, R, F, V), the possible motion parameters are generated. Then, the time-modulating and impulse response functions are determined through Equations (2), (3), (8)–(12). Eventually, the simulated ground acceleration process $\ddot{z}(t)$ can be obtained through Equations (4) and (5).

Taking the 1994 Northridge earthquake recorded at the LA-Wonderland Ave as an example, the elastic response spectrum of the two recorded and 100 synthetic ground motions is shown in Figure 5. All these motions are corresponded to $M = 6.69$, $R = 20.3$ km, $V = 1223$ m/s, $F = 1$ (i.e., reverse fault). It is seen that the recorded motions can be considered as a subset of possible ground motions with similar site and earthquake characteristics.

3.2 | Seismic consequences estimation

FEMA P-58 method is adopted in this paper, one of the seismic consequence estimation methodology paradigms, as shown in Figure 6. Following the core concept of PBEE-2, the FEMA P-58 method converts the seismic risk faced by the building into the specific possible seismic consequence loss through the multiple-integral shown in Equation (13).

$$\lambda(DV) = \iiint G(DV|DM)\mathrm{d}G(DM|EDP)\mathrm{d}G(EDP|IM)\mathrm{d}\lambda(IM) \tag{13}$$

where $\lambda(X)$ denotes the mean annual frequency of X. $G(.|.)$ denotes the conditional probability density. DV represents the decision variable, that is, the "seismic consequence" in this paper, generally indicators such as economic loss, carbon emissions, casualties, and so forth. DM represents the damage measure, usually the damage state a component is in. EDP represents the EDPs, which should be the structural response indicators that best reflect the components' damage state, for example, peak floor acceleration (PFA), peak interstory drift ratio (PIDR), and residual interstory drift ratio (RIDR). IM denotes the intensity measure of earthquake events, such as the spectral acceleration corresponding to the building's fundamental period and 5% damping ratio, $S_a(T_1, 5\%)$.

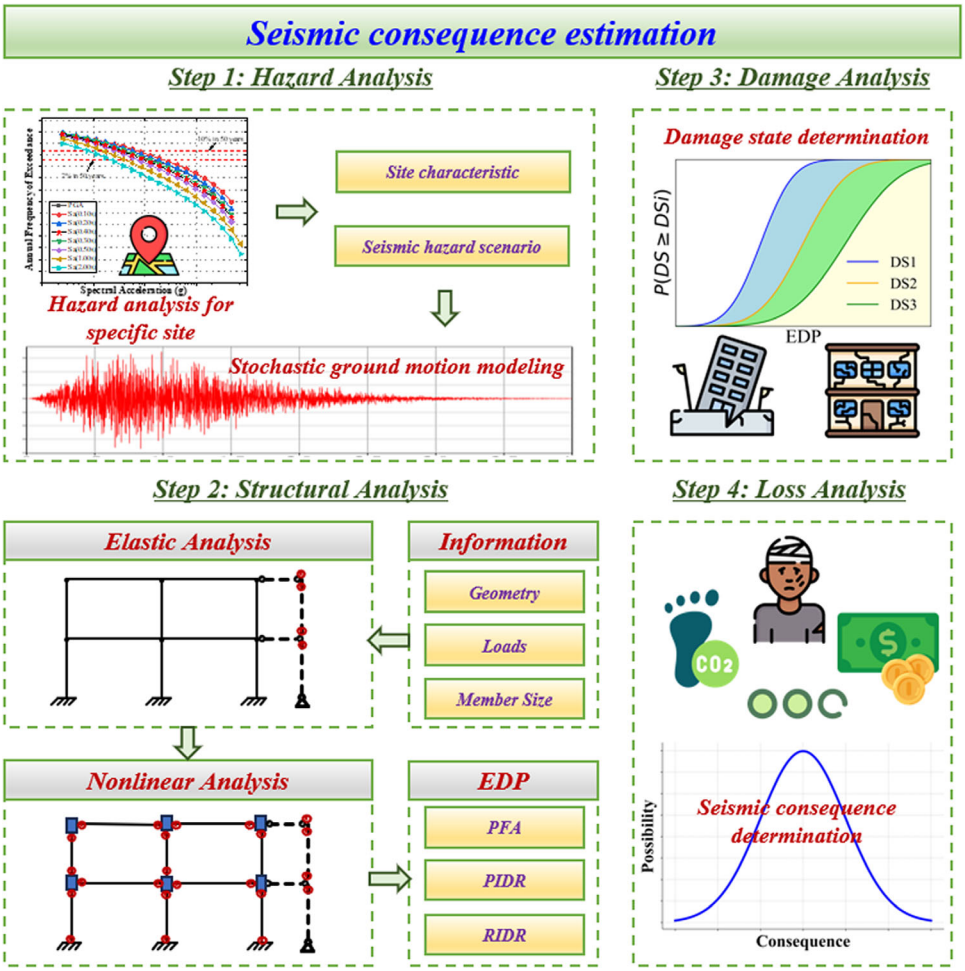


FIGURE 6 The flowchart of seismic consequence estimation.

The relationship between the four steps shown in Figure 6 is related to the multiple integrals shown in Equation (13). More specifically, hazard analysis provides the IM value for a given hazard scenario, and then the appropriate ground motion set under the considered IM value is used to obtain the corresponding structural response distribution in structural analysis. After obtaining the distribution of structural response, the possible damage states and corresponding probabilities of each considered component can be obtained by incorporating the component fragility function provided by FEMA P-58 documents in the damage analysis. In the loss analysis, the identified damage states of each component, combined with the consequence function provided by FEMA P-58 can be further translated into the corresponding repair cost and probability. So, data are transmitted unidirectionally in order between the four modules, and their integration realizes the transformation from site and structure-related seismic risk to structural seismic consequences.

More specially, the fragility specifications for each component's damage state detail repair actions, materials, and quantities needed for cost estimation, which are crucial for creating consequence functions. These costs account for all construction activities required to restore components to their pre-earthquake state, assuming "in-kind" repairs or replacements without updating to newer criteria. Costs are derived from outlined repair measures, encompassing every step a contractor would take. According to FEMA P-58, the consequence functions for repair costs are based on 2011 construction cost estimates from Northern California, excluding uncertainties like contractor pricing or cost escalation. They also factor in economies of scale and operational efficiencies in construction. More detailed information about the method for evaluating repair costs can be found in the FEMA P-58 documents.[11]

Notably, it is impractical to solve the multiple-integral analytically, so the Monte Carlo simulation is typically used to address this problem. The distributions of EDP under N_{GM} ground motions for N_{IM} IM levels are obtained through NTHA. Then, a 3D Monte Carlo sampling table with $N_{GM} \times N_{IM}$ columns can be established to illustrate the process, as shown in Figure 7 visually. The N_S cubes in the (i, j)th column represent the N_S "consequence" samples under ith ground motion

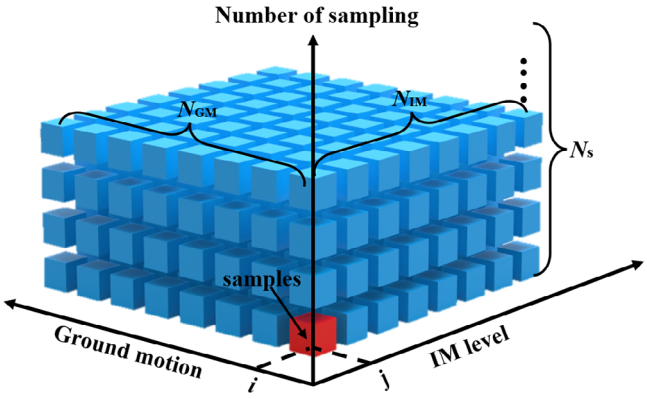


FIGURE 7 Schematic illustration of Monte Carlo simulation.

at *j*th IM level. For each column, if the corresponding PIDR exceeds the collapse limit, such as 10%[64] adopted in this paper, then the column is filled with building replacement consequences. Otherwise, the RIDR is employed to obtain the probability of feasibility of building repair, P_{nr}. The probability of irrepair exceeding a certain RIDR is typically assumed to satisfy the lognormal distribution.[11] Then, the $P_{nr} \times N_S$ cubes can be seen as irreparable and will also be filled with building replacement consequences. For remaining $N_S \times (1 - P_{nr})$ cubes, that is, the building is considered repairable, each building component's possible damage state is first sampled according to the EDP. Subsequently, the possible repair consequence of the damage state is sampled. The sampling results of all components are aggregated to derive the overall sampling results for the building, that is, the "consequence" filled in cubes. Finally, through studying the resulting 3D sampling table, the distribution of seismic consequence under each considered IM can be obtained. The above procedure is also called "intensity-based seismic consequence estimation." Furthermore, suppose the occurrence probability of IM is considered to obtain the value and probability of seismic consequence within a given time interval. In that case, this is called "time-based seismic consequence analysis."[11]

4 | SURROGATE MODEL: GAUSSIAN PROCESS MODEL

The GP model is adopted as the surrogate model in this paper, which originated from the stochastic process in probability and statistics and then evolved into a powerful supervised machine learning method with wide application. Among many machine learning techniques, the GP is highly recommended because it can provide a reasonable mechanism to infer high-dimensional attributes of observed data and a probability distribution of predicted results.

According to the existence theorem of GP, it can be uniquely defined by a mean and covariance function. The mean function and the covariance function of a real process satisfies:

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})] \tag{14}$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))] \tag{15}$$

where $\mathbf{x}$, $\mathbf{x}'$ are arbitrary input vectors, $m(\mathbf{x})$ is the mean function of GP, $k(\mathbf{x}, \mathbf{x}')$ is the covariance function of GP, $f(\mathbf{x})$ is the GP latent function values, $\mathrm{E}[f(\mathbf{x})]$ is the expectation of $f(\mathbf{x})$. And then the function $f(\mathbf{x})$ distributed as a GP can be denoted as

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \tag{16}$$

A general Gaussian regression model with stochastic noise can be expressed in the form of Equation (17). The term $f(\mathbf{x})$ here is a GP with a specific form of mean function and covariance function but parameters to be determined. Such undetermined parameters are also called "hyperparameters." Different forms of functions may have different types of hyperparameters. The rest term ε is the added noise, which satisfies $\varepsilon \sim \mathrm{N}(0, \sigma_n^2)$.

$$y = f(\mathbf{x}) + \varepsilon \tag{17}$$

Assume that the observed data satisfy such model, that is, the prior. Then, the joint distribution of the observed outputs, $\mathbf{y}$, and the predicted outputs $\mathbf{f}_*$ according to the prior is

$$\begin{bmatrix} \mathbf{y} \\ f_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(X,X)+\sigma_n^2 I & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right) \tag{18}$$

where $X = \{x_1, \ldots, x_i, \ldots, x_n\}$ denotes n data points, and each $x_i = [x_{i1}, \ldots, x_{im}]$ denotes m-dimensional input variables. X_* denotes n_* points to predict. $K(X, X)$, $K(X, X_*)$, $K(X_*, X)$ are $n \times n$, $n \times n_*$, $n_* \times n$ matrixes, respectively, and their corresponding (i, j)th elements are $k(x_i, x_j)$, $k(x_i, x_{j*})$, $k(x_{i*}, x_j)$, respectively. The prior mean function is generally set to zero.

Then, according to the conditional distribution properties of Gaussian distribution, the key predictive equations for GP regression can be derived as follows:

$$\mathbf{f}_*|X, \boldsymbol{y}, X_* \sim \mathcal{N}(\bar{\mathbf{f}}_*, cov(\mathbf{f}_*)) \tag{19}$$

$$\bar{\mathbf{f}}_* \triangleq \mathbb{E}[\mathbf{f}_*|X, \boldsymbol{y}, X_*] = K(X_*,X)\left[K(X,X)+\sigma_n^2 I\right]^{-1}\boldsymbol{y} \tag{20}$$

$$\text{cov}(\mathbf{f}_*) = K(X_*,X_*) - K(X_*,X)\left[K(X,X)+\sigma_n^2 I\right]^{-1}K(X,X_*) \tag{21}$$

Then, the predicted distribution of test targets $\mathbf{y}_*$ can be computed by adding $\sigma_n^2 I$ to covariance in Equation (21). While the key prediction equations have been derived, the hyperparameters θ that make the previous assumptions hold still need to be obtained through training. In practice, the type II maximum likelihood estimation is adopted, that is, through the maximum log marginal likelihood function as defined in Equation (22).

$$\mathcal{L}(\theta) = \log p(y|X,\theta) = -\frac{1}{2}\mathbf{y}^T\left(K_\theta + \sigma_n^2 I\right)^{-1}\mathbf{y} - \frac{1}{2}\log\left|K_\theta + \sigma_n^2 I\right| - \frac{n}{2}\log 2\pi \tag{22}$$

$$\theta_{MLE} = \arg\max_\theta \mathcal{L}(\theta) \tag{23}$$

where K_θ denotes $K(X, X)$.

5 | VARIANCE-BASED GLOBAL SENSITIVITY ANALYSIS METHOD

The variance-based sensitivity analysis is utilized as the GSA method in this study. For a given model of the form $Y = f(\mathbf{X})$, the variance-based method is a probabilistic sensitivity analysis that quantifies the sensitivity of Y to $\mathbf{X}$ in terms of the reduction in the variance of Y. The function $f(.)$ in this study is the seismic loss estimation method. Here, model inputs ($\mathbf{X}$) involve some building's properties (e.g., building mass, damping ratio, component fragility, and consequence parameters) and the interested earthquake intensity measure.

In the variance-based method, the total variance of the "model" output Y over the entire input space X can be decomposed as

$$V(Y) = \sum_{i=1}^{p} V_i + \sum_{1\le i<j<p} V_{ij} + \cdots + V_{1\ldots p} \tag{24}$$

where V_i is the variance of individual input variables X_i, V_{ij} is the variance of the interaction between X_i and X_j, and $V_{1\ldots p}$ is the variance of the interaction between all parameters.

As defined in Equation (24), V_i can represent the main effect of X_i on Y. Then the first-order sensitivity index is calculated by normalizing the main effect by the total variance as Equation (25). The index S_i can be interpreted as the average degree of reduction of the total variance when X_i is fixed.

$$S_i = \frac{V_i}{V(Y)} \tag{25}$$

Let $V_{\sim i}$ denote the variance contribution of all parameters except X_i. Then, another critical index-total sensitivity index S_{Ti} can be derived by Equation (26). The index S_{Ti} is the sum of the X_i-related effects on Y, including the main and interaction effects. The difference between S_i *and* S_{Ti} can be used to measure the degree of the parameter interaction.

$$S_{Ti} = 1 - \frac{V_{\sim i}}{V(Y)} \tag{26}$$

In practice, the interested model is generally complex and has no explicit expression, so it is necessary to use numerical calculations to obtain an approximate result close to the analytical one. First, two independent parameter sample matrices are generated, that is, M_1 and M_2:

$$M_1 = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \dots & & & \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}, M_2 = \begin{pmatrix} x'_{11} & x'_{12} & \dots & x'_{1p} \\ x'_{21} & x'_{22} & \dots & x'_{2p} \\ \dots & & & \\ x'_{n1} & x'_{n2} & \dots & x'_{np} \end{pmatrix} \tag{27}$$

where p is the number of parameters and n is the sample size. Then, the estimated total variance $\hat{V}(Y)$ is derived using M_1 and M_2, as

$$\hat{V}(Y) = \frac{1}{2n-1} \sum_{i=1}^{n} \left\{ f^2(x_{i1}, x_{i2}, \dots, x_{ip}) + f^2\left(x'_{i1}, x'_{i2}, \dots, x'_{ip}\right) \right\} - \hat{f}_0^2 \tag{28}$$

where $f(.)$ denotes the model output value for given input parameters, $\hat{f}_0$ denotes the expectation of the model estimated by

$$\hat{f}_0^2 = \frac{1}{n} \sum_{i=1}^{n} f(x_{i1}, x_{i2}, \dots, x_{ip}) \times f\left(x'_{i1}, x'_{i2}, \dots, x'_{ip}\right) \tag{29}$$

The estimated variance of individual parameter X_i, that is, $\hat{V}_i$, can be derived by

$$\hat{V}_i = \frac{1}{n-1} \sum_{i=1}^{n} \left\{ f\left(x_{i1}, x_{i2}, \dots, x_{ip}\right) \times f\left(x'_{i1}, x'_{i2}, \dots, x'_{i(j-1)}, x_{ij}, x'_{i(j+1)}, x'_{ip}\right) \right\} - \hat{f}_0^2 \tag{30}$$

The estimated total variance of all parameters except X_i, that is, $\hat{V}_{\sim i}$, can be derived by

$$\hat{V}_{\sim i} = \frac{1}{n-1} \sum_{i=1}^{n} \left\{ f\left(x'_{i1}, x'_{i2}, \dots, x'_{ip}\right) \times f\left(x'_{i1}, x'_{i2}, \dots, x'_{i(j-1)}, x_{ij}, x'_{i(j+1)}, x'_{ip}\right) \right\} - \hat{f}_0^2 \tag{31}$$

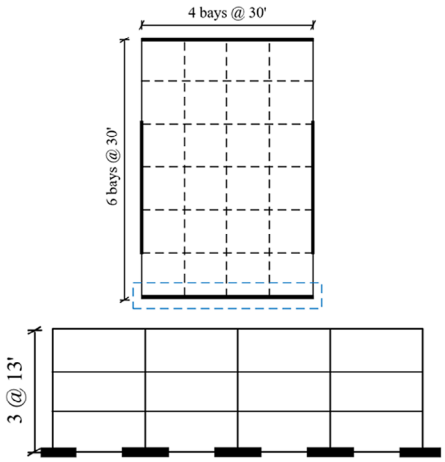
The total computational cost for obtaining S_i and S_{Ti} is $n(p+2)$, where $2n$ times model executions are used to estimate the $\hat{V}(Y)$ as Equation (28), $p \times n$ times are used to estimate S_i. According to Equation (31), no more model evaluations need to be executed to calculate the S_{Ti}.

Besides, the bootstrap technique is used to compute 95% confidence intervals of the sensitivity indices to assess the robustness of sensitivity analysis. This technique resamples the output samples multiple times and calculates the corresponding sensitivity indices for each bootstrap resample to obtain the distribution interval. In this paper, the number of resamples is set to 1000, as recommended in Ref.[65] More detailed information about the variance-based method can be found in Refs.[35,66,67]

6 | ILLUSTRATIVE EXAMPLES

6.1 | Basic information

Two three-story SMRF prototype buildings in FEMA P-2012[68] are used to demonstrate the proposed framework. The two buildings have the same plan and elevation dimensions, one is designed for the high-seismicity zone (e.g., Los Angeles),



Design for high-seismicity zone:

Story	Column		Girder
	Exterior	Interior	
Roof	W14×211	W14×311	W27×94
3rd	W14×311	W14×370	W33×130
2nd	W14×311	W14×370	W33×130

Design for low-seismicity zone:

Story	Column		Girder
	Exterior	Interior	
Roof	W14×53	W14×99	W21×44
3rd	W14×53	W14×99	W24×62
2nd	W14×53	W14×99	W24×62

FIGURE 8 The elevation, plan view, geometry dimensions, and member size of the frame.

and the other for the low-seismicity zone (e.g., Boston). As shown in Figure 8, each building has two SMRFs in each orthogonal direction, and one SMRF in the north-south direction (i.e., marked by the blue dotted line) is used in this paper to characterize the earthquake resistance in both directions uniformly. The elevation, plan view, geometry dimensions, and member size are shown in Figure 8. The design dead loads on the floor and roof are typically 106 and 83 psf, respectively, while the design live loads are 50 and 20 psf, respectively. The above geometric and load information is written in a Txt file and then read by the structural analysis module to construct the OpenSeesPy model automatically. The fundamental periods of the two models are 0.8 and 1.93 s, respectively.

Table 1 presents information on the structural and nonstructural components considered in the three-story building. When building type (office building) and floor area are given, the quantities of nonstructural components can be estimated by *Normative Quantity Estimation Tool*[69] provided by FEMA P-58. This tool offers the mean value and key parameter (beta) of the lognormal distribution satisfied by the quantities of nonstructural components. The damage states of structural components are typically determined by PIDR. Nonstructural components are divided into acceleration-sensitive and displacement-sensitive types. Facades and ceilings are typically displacement-sensitive in nonstructural components, while mechanical, electrical, and plumbing (MEP) systems are typically acceleration-sensitive. Note that the content, such as bookshelves, computers, and so forth, are not considered in this paper, as there is a lack of relevant fragility data.

6.2 | Interested input variables (*X*)

Multiple parameters, including structure, fragility, and consequence parameters, are considered in this paper, as presented in Table 2. The building mass and damp ratio are selected for structure parameters, which are also the main sources of uncertainty in the structural model considered in FEMA P-58. The building mass or dead loads will change due to construction quality, errors in the mass of the preset components during design, and so forth. As Ellingwood et al.[70] summarize, the Gaussian distribution with a mean value equal to nominal dead load and a coefficient of variation (CV) of 0.10 is an approximation model for the distribution of dead load. Therefore, the 10th and 90th percentiles of such Gaussian distribution can be taken as the lower and upper limits of mass/dead load to set a truncated Gaussian distribution. For the damping ratio, 2% is typically used for steel buildings in practice, while this value is varied in previous studies. Bernal et al.[71] investigated 81 responses from steel buildings and found that damping ratios were typically larger than 2%. In the latest Indian seismic code IS 1893 (Part-1): 2016, a damping ratio of 5% is fixed for all types of buildings.[72] Meanwhile, in the Chinese seismic code GB50011-2010,[73] the damping ratio ranges from 0.02~0.05 depending on the structure height, analysis type, and earthquake intensity. To study the uncertainty contribution of the damping ratio, it is assumed here that the damping ratio is uniformly distributed between 2 and 5%.

The mean values of the fragility functions of some major structural and nonstructural components are considered for fragility parameters. Uncertainties about these parameters are lacking investigation. Therefore, these mean values' distributions are assumed to satisfy the normal distribution with the corresponding mean values provided in FEMA P-58 and a uniform CV of 0.3. Like building mass, the 10th and 90th percentiles are taken as the lower and upper limits. Besides, the

TABLE 1 The information on structural and nonstructural components considered in this paper.

Category	Component	FEMA P58 ID	Quantity per story	Fragility unit of measure	Quantity beta[a]	EDP	Story
Nonstructural components	Curtain walls	B2022.001	216	30 SF	0.6	PIDR	1–3
	Wall partition	C1011.001a	21.6	100 LF	0.2		1–3
	Suspending ceiling	C3032.001a	86.4	250 SF	0.0		1–3
	Raised access floor	C3027.001	162	100 SF	0.2	PFA	1–3
	Cold or hot potable- pipe	D2021.011a	0.91	1000 LF	0.7		1–3
	HVAC ducting, A ≥ 6 sq.ft	D3041.012a	0.43	1000 LF	0.2		1–3
	HVAC ducting, A ≤ 6 sq.ft	D3041.011a	1.62	1000 LF	0.2		1–3
	HVAC drops/diffusers	D3041.031a	19.44	10 EA	0.5		1–3
	VAV box	D3041.041a	10.8	10 EA	0.2		1–3
	Independent pendant lighting	C3034.002	648	1 EA	0.3		1–3
	Fire sprinkler drop standard threaded steel	D4011.031a	1.94	100 EA	0.2		1–3
	Fire sprinkler water piping	D4011.021a	4.32	1000 LF	0.1		1–3
	Prefabricated steel stair	C2011.001b	2.16	1 EA	0.2		1–3
	Motor control center	D5012.013a	2.59	1 EA	0.5		1
	Air handling unit	D3052.011a	3.78	4000 CF	0.2		1–3
	Cooling tower	D3031.021a	2.46	75 TN	0.1		Roof
	Chiller	D3031.011a	2.46	75 TN	0.1		1
	Traction elevator	D1014.011	1.81	1 EA	0.7		1
	Low voltage switchgear	D5012.021a	0.03	225 AP	0.4		1–3
Structural components (high-seismicity zone)	Steel column base plates	B1031.011c	20	1 EA	—	PIDR	1
	RBS connection at the exterior columns, $D_b \geq$ W30	B1035.002	8				1, 2
	RBS connection at the interior columns, $D_b \geq$ W30	B1035.012	12				1, 2
	RBS connection at the exterior columns, $D_b \leq$ W27	B1035.001	8				3
	RBS connection at the interior columns, $D_b \leq$ W27	B1035.011	12				3
Structural components (low-seismicity zone)	Steel column base plates	B1031.011a	20	1 EA	—	PIDR	1
	RBS connection at the exterior columns, $D_b \leq$ W27	B1035.001	8				1,2,3
	RBS connection at the interior columns, $D_b \leq$ W27	B1035.011	12				1,2,3

Note: D_b denotes the beam depth.

Abbreviations: AP, Amp; CF, cubic foot; EA, each; LF, linear foot; SF, square foot; TN, ton.

[a]Lognormal distribution.

TABLE 2 The interested input variables (X) in this paper.

Source of uncertainties	Symbol	Explanation	Distribution	Range
Structure parameters	m_{b}	Building mass	Normal: Mean = D_n, CV = 0.1	$[0.872D_n, 1.128D_n]$
	ζ	Damp ratio	Uniform	[2%, 5%]
Fragility parameters	M_{bcj}	Mean of the beam–column joint's fragility	Normal: Mean = M_{com}, CV = 0.3 (M_{com} varies with the component type)	$[0.616M_{\mathrm{com}}, 1.384M_{\mathrm{com}}]$
	M_{gcw}	Mean of the curtain wall's fragility		
	M_{wp}	Mean of the wall partition's fragility		
	M_{sc}	Mean of the suspended ceiling's fragility		
	M_{ele}	Mean of the elevator's fragility		
	M_{hvac}	Mean of the HVAC's fragility		
	M_{rf}	Mean of the repair fragility	Uniform	[0.5%, 1.5%]
	S_{rf}	Standard deviation of the repair fragility	Uniform	[0.1, 0.5]
Consequence parameters	P_{nsq}	Nonstructural component quantity percentile	Uniform	[50, 100]
	C_{rep}	Replacement cost	Uniform	$[C_{rmax}, C_{rmax}/0.3]$

Note: CV refers to the coefficient of variation; D_{n} refers to nominal dead load; M_{com} refers to the median of the corresponding component's fragility specified in the FEMA P58 document; C_{rmax} refers to the maximum repair cost potential of the components (i.e., if every single component were damaged in its most expensive damage state).

mean and standard deviation of the repair fragility (i.e., the relationship between the RIDR and the feasibility of repair) are also considered. Some current studies take the mean as 0.5%,[13,74] while some suggest higher values,[75] such as 1%.[11] Most studies take the standard deviation to be 0.3. In this paper, the mean and standard deviation of the repair fragility are assumed to be uniformly distributed in [0.5%, 1.5%] and [0.1, 0.5], respectively.

Nonstructural components are the primary source of earthquake-induced losses in most buildings. Since the FEMA P-58 database provides the possible distribution of the quantities of nonstructural components, the quantity percentile is also selected as the interested variable. Note that it is assumed that the change in percentiles of all components is the same, that is, the quantities of all components are at the same level. Then, for a given component, its corresponding amount can be calculated using the lognormal distribution expression as follows:

$$\ln(q) = G^{-1}\left(\frac{P_{nsq}}{100}\right) = \ln(\hat{q}) + \beta_q \Phi^{-1}\left(\frac{P_{nsq}}{100}\right) \tag{32}$$

where P_{nsq} is the quantity percentile and follows a uniform distribution in [50, 100], $G^{-1}(.)$ is the inverse cumulative density function of the lognormal distribution, $\hat{q}$ is the 50th percentile quantity of the component, β_q is the logarithmic standard deviation of the quantity, that is, the "Quantity beta" presented in Table 1. According to the seismic risk assessment software SP3's website,[76] the ratio of maximum component repair potential (i.e., if every single component were damaged in its most expensive damage state) can reasonably range anywhere from 30 to 150%. For steel frame buildings, the maximum repair potential (C_{rmax}) is typically lower than the building value, because most of the building's value may lie in "rugged" components not covered by the FEMA P-58 model. Therefore, the building replacement cost is considered to satisfy a uniform distribution in $[C_{rmax}, C_{rmax}/0.3]$. Notably, in the future, detailed investigations into the replacement costs of structures of different structural types would be beneficial in reducing the uncertainty in this parameter.

6.3 | Hazard analysis and stochastic ground motion modeling

Assume that the prototype buildings are in Los Angeles, the latitude and longitude coordinates are (34.11197, −118.38544), and the shear wave velocity is assumed as 760 m/s. Then, the corresponding seismic hazard curves are obtained by the Unified Hazard Tool provided by the U.S. Geological Survey website,[77] as shown in Figure 9. The IMs (i.e., the 5% damping ratio spectral acceleration at fundamental period $S_a(5\%, T_1)$) corresponding to four hazard levels (43-year,

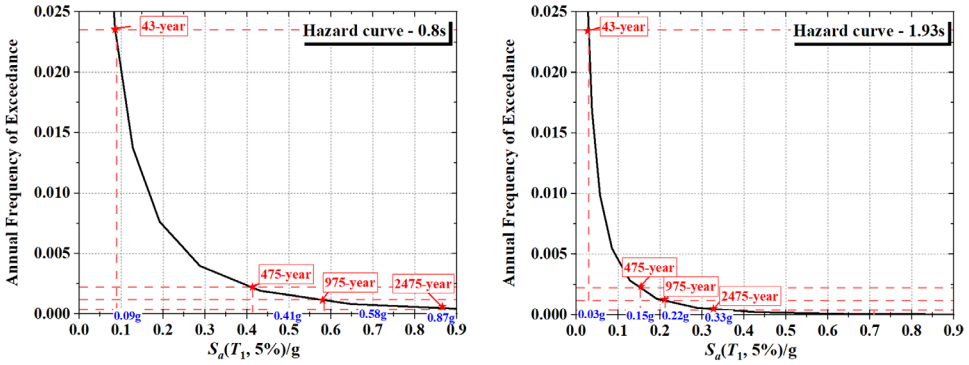


FIGURE 9 Hazard curve.

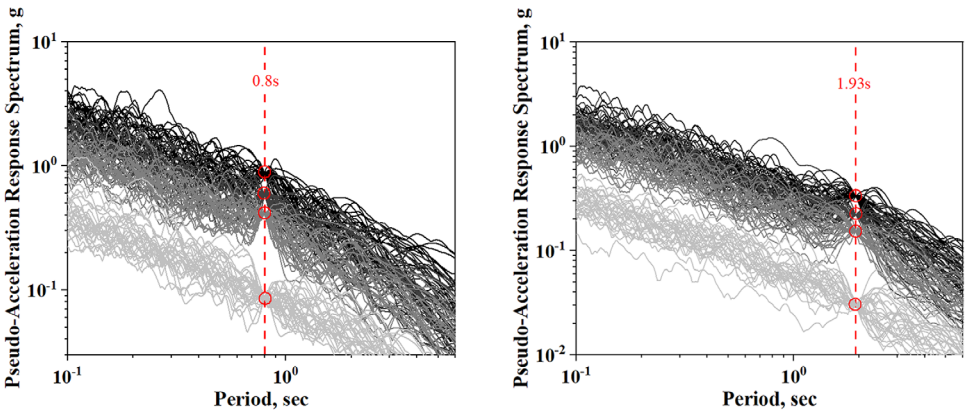


FIGURE 10 The pseudo-acceleration response spectrum of the generated ground motion.

475-year, 975-year, 2475-year) are selected. For the SMRF designed for the high-seismicity zone, the IM values are 0.09, 0.41, 0.58, and 0.87 g, respectively, while for the SMRF designed for the low-seismicity zone, the IM values are 0.03, 0.15, 0.22, and 0.33 g respectively.

Then, for each IM level, 30 specific ground motions are generated, as shown in Figure 10. Note that if the target is only a single IM value when generating, it may require a lot of calculation time. Therefore, a compromise solution is adopted: set a bias interval of 10% for each IM. For instance, if the desired IM ($S_a(5\%, T_1)$) value is 0.3 g, it is advisable to define a target range for generating the ground motion as [0.27, 0.30 g]. Subsequently, a little amplitude modulation can be applied to the generated ground motion to scale them to the desired IM value. Furthermore, unexpected situations may arise, such as when dealing with the 2475-year hazard level when the IM value tends to be relatively high. Given that the ground motion modeling technique relies on a genuine earthquake database, it is possible that such a large IM level earthquake event may not exist within the dataset. Consequently, in such circumstances, a larger scale factor becomes unavoidable. Nevertheless, compared with other methods, the SGMM technique can generate ground motions as close as possible to the desired IM value. In cases where the IM corresponding to the 2475-year hazard level with a fundamental period of 0.80 s exceeds the available limits for naturally occurring ground motions using the modeling technique, amplitude modulation may be performed using a scale factor smaller than 2−4. Such an adjustment within this range is considered reasonable and has been endorsed in previous studies.[23]

6.4 | Build a surrogate model for structural analysis

The synthetic ground motions at different IM levels are input into the structural analysis module for NTHA. For building the surrogate model, 1000 analyses are performed to construct the training set. In addition, 200 analyses are performed to build the testing set to verify the accuracy of the surrogate model at each considered IM level. Eight parameters, including building mass m_b, damping ratio ζ, peak ground acceleration *PGA*, peak ground velocity *PGV*, peak ground displacement *PGD*, pseudo-acceleration response spectrum value S_a, pseudo-velocity response value S_v, displacement

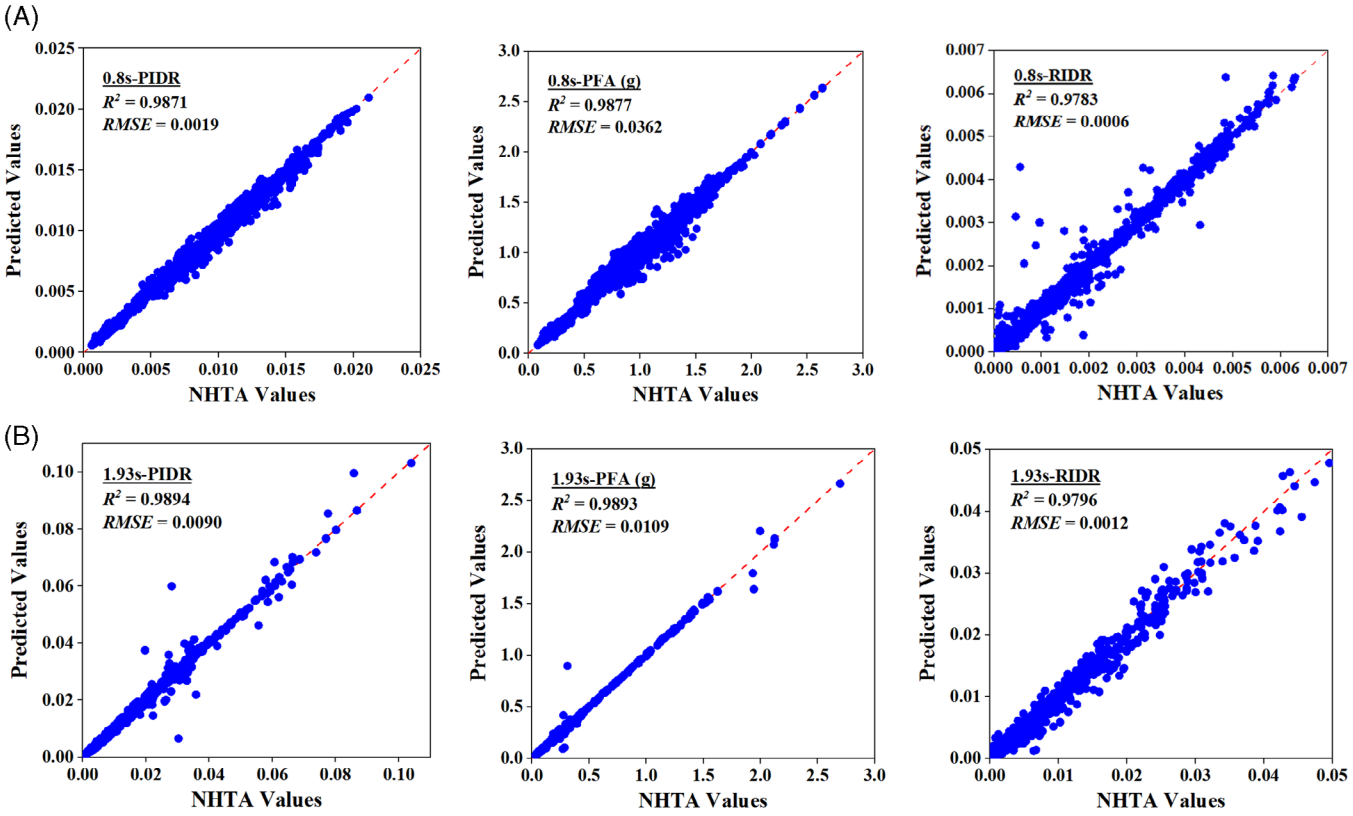


FIGURE 11 Values predicted by the Gaussian process model compared with values obtained by NTHA. (800 test data for each SMRF, and each test data contains the corresponding response values of all stories). (A) SMRF with fundamental period of 0.80s; (B) SMRF with fundamental period of 1.93s.

response spectrum value S_d, are selected as the input features of surrogate model for predicting the PIDR and PFA. These features are generally practical for predicting peak displacement and acceleration responses, but more input features are needed for residual deformation. Therefore, the effective duration of the ground motion (i.e., the $D_{5\text{–}95}$ mentioned in Equation 8) is also selected. Furthermore, the predicted values of the displacement response are also used as input features for the surrogate model to predict the RIDR to improve accuracy. The automatic relevance determination (ARD) technique[78] is used to assign different hyperparameter values for different input features. The kernel function of the GP model is set as the radial basis function (RBF).[78]

It can be noted that to enhance the prediction capabilities of the model, the surrogate models have many additional parameters compared to the parameters to be investigated in the sensitivity analysis, but the impact of these additional parameters on the estimated loss is not investigated in this paper. As stated in Section 5, the sensitivity analysis requires fixing one parameter and varying others. For the interested parameters mentioned in Section 6.2, this operation is reasonable and feasible, whereas for parameters such as PGV, PGA, duration, and so forth, it is not possible to realize a reasonable fixing of PGV and changing only PGA for ground motions, especially when Sa is given, because of the interrelationships between these parameters. Although this can be achieved in the GP model (the surrogate model always outputs results based on the input information), the results obtained are spurious and not only do not represent the realistic ground motion inputs, but also lead to errors in the calculation of the sensitivity of the other parameters. Besides, in the selected loss estimation methodology (i.e., FEMA P-58), one seismic loss estimation (i.e., obtaining one median or mean value) uses the 30 generated ground motions (with different PGAs, PGVs, duration, etc.), so trying to use the same PGA/PGV/duration, and so forth in one estimation is not possible, nor is its sensitivity analysis. Although this paper does not focus on these parameters, it would be interesting to further investigate the impact of these ground motion-related indicators on the seismic loss in the future.

Through verification on the testing set, the established surrogate model has good prediction accuracy for both SMRFs. As shown in Figure 11, the coefficient of determination R^2 of all EDPs is above 0.95, indicating the reliability of the surrogate model.

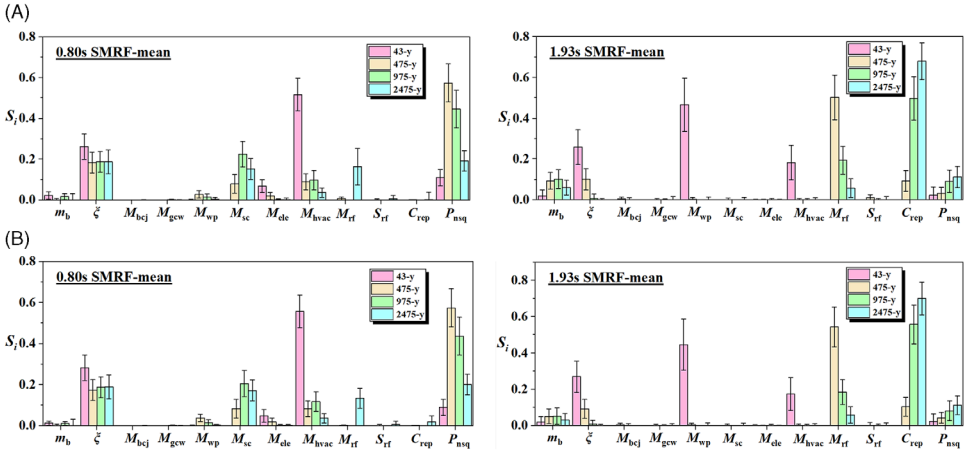


FIGURE 12 Sensitivity analysis results when the target is the mean value of economic loss. (A) Surrogate model-aided sensitivity anlaysis; (B) Surrogate-free sensitivity analysis.

6.5 | Global sensitivity analysis

In this section, leveraging the constructed surrogate model, the sensitivity of seismic economic loss to the considered parameters is examined. First, the mean value is considered as the interested performance indicator. Both surrogate model-aided and surrogate-free sensitivity analysis are performed using an Intel(R) Core(TM) 13980HX processor with 16 threads enabled. The sample size of the analysis is 512, so according to Section 5, 7168 seismic consequence assessment needs to be performed for each hazard level. For surrogate model-aided analysis, the average time per hazard level, which includes the model training time, is approximately 1 h. In contrast, for the analysis conducted without a surrogate model, it took approximately 18 h per hazard level. This substantial reduction in computational time underscores the significant efficiency gains achieved with the surrogate model. Meanwhile, based on the comparison depicted in Figure 12, the results of the sensitivity analysis aided by the surrogate model appear to be reasonable, with no significant deviations.

It is evident from the figure that the sensitivity analysis results for two different structural designs, each with distinct earthquake resistance capabilities, exhibit considerable differences. This difference is caused by the specific site and the two design alternatives: one design is stricter than the site's requirements, while the other has "poorer seismic performance." The "poorer seismic performance" here is a relative concept. As mentioned in Sections 6.1 and 6.3, the two structures are designed for high and low seismic zones, respectively, and are assumed to be in a high seismic zone (i.e., Los Angeles) in this illustrative example. Thus, the frame designed for a low seismic zone would naturally exhibit "poorer seismic performance" when subject to the seismic demands of a high seismic zone. This relative difference in resilience between the two frames forms the basis of our comparison and subsequent analysis. More specially, the former design is better suited for characterizing parameter sensitivities when smaller structures respond, whereas the latter design is more appropriate for capturing sensitivities associated with larger responses. In the case of the more resilient building, the low sensitivity to earthquake replacement costs can be attributed to its small PIDR and RIDR. The earthquake replacement cost is primarily attributed to two factors: collapse or lack of feasibility of repair. The former factor is related to the exceeded PIDR, and the latter factor is related to the exceeded RIDR. Relative smaller PIDR and RIDR indicate a higher repair feasibility and a lower probability of structural collapse, thus resulting in lower sensitivity to replacement costs. Conversely, in the case of the design with poorer seismic performance, replacement costs exhibit a certain level of sensitivity. A similar trend is also observed in the mean value of repair fragility.

The variability exhibited by the sensitivity analysis of different structures in different hazard scenarios is determined by a combination of many factors. First, the correlation between the structural response level and the component fragility characteristics. Since the fragility curves provided by FEMA P-58 tend to satisfy the cumulative distribution function form of normal or lognormal distribution, when the response level of the structure is close to the mean value of the component fragility, the change in the mean value will have a greater impact on the assessment of the damage results, while if the structural response is farther away from the mean (i.e., smaller or larger), the change in the mean value will have less

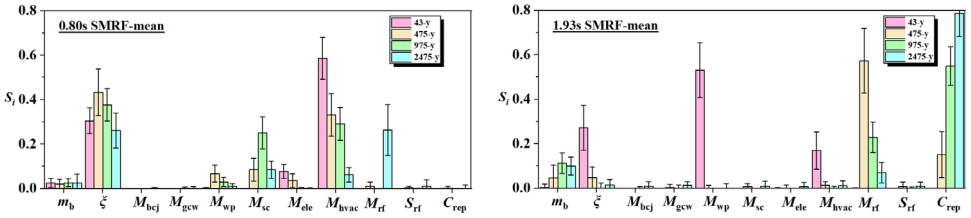


FIGURE 13 Sensitivity analysis results when the target is the mean value of economic loss (with no P_{nsq}).

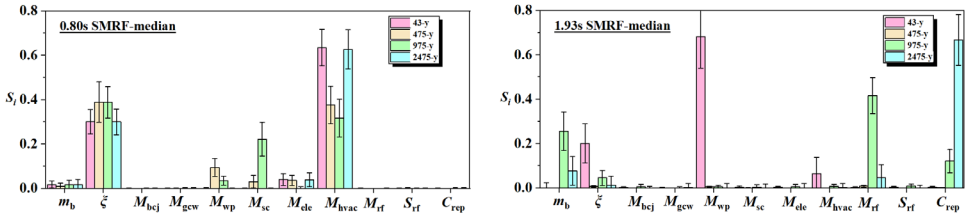


FIGURE 14 Sensitivity analysis results when the target is the median value of economic loss (with no P_{nsq}).

impact on the analysis results. Besides, different components are damaged with different loss values, and the sensitivity increases further as the structural response approaches the mean value of the fragility of the component with larger losses. Moreover, as the structural nonlinearity develops further and the probability of collapse or irreparable damage increases, the sensitivities of the parameters associated with collapse and irreparable damage will dominate, as shown by the results of the SMRF with poorer seismic performance. The multiple factors mentioned above are intertwined together to form the results of the variable sensitivity analysis. Because of this, it is not very recommended to draw conclusions with generalizability for reference, such as which component has the high sensitivity in a given hazard scenario, as this is related to the level of structural design, structural configurations, number of floors, number of bays, considered components quantities and types, structural form, and many other factors. Therefore, this paper focuses more on the proposed framework itself, which makes feasible and efficient the sensitivity analysis of a specific design of a structure at a specific hazard level.

In practical applications, when the analysis result reveals that a particular parameter exerts a significant impact on the estimation outcomes, it may warrant further investigation on similar buildings to mitigate such uncertainty. Assume that the designer conducts an exhaustive investigation of the number of nonstructural components (set as the 50th percentile), then a sensitivity analysis is conducted once more with P_{nsq} fixed at 50%, as illustrated in Figure 13. During this analysis, it was determined that the parameters ξ and M_{hvac} in the design scheme for high-seismicity zones and the parameters M_p and C_{rep} in the design scheme for low-seismicity zones warrant further investigation. This approach facilitates a more exhaustive examination of various design scenarios, ultimately leading to the generation of more robust and reliable analysis results.

The selection of interested performance indicators also affects the results. For instance, when the median value rather than the mean value is chosen as the indicator, it is found that the sensitivity to residual deformation and replacement cost is weakened, as shown in Figure 14. In addition, the sensitivity of other parameters also changed, especially M_{hvac}, from a low-sensitivity parameter to an extremely high-sensitivity parameter at 2475 year for the 0.8 s SMRF. This phenomenon arises because the median focuses exclusively on the values at the 50th percentile, unlike the mean, and an increase in extreme values, which is limited to no more than half of the total values, has no impact on it. Furthermore, taking the SMRF with $T_1 = 0.8$ s as an example, it can be seen from the comparison between Figure 13 (left) and Figure 14 (left) that if the structural response is small under the hazard scenario considered, selecting two indicators will have little impact on the results, but if the structural response is large, that is, the structure has a certain probability of collapse or irreparability, then selecting the mean value rather than median value can effectively reflect the sensitivity of parameters related to structure replacement. Based on the findings of the illustrative example, the mean value is more suitable as a performance indicator for sensitivity analysis. However, it is worth noting that it would be better if it was possible to obtain results for multiple indicators, which would help to compare and verify different results to enhance the robustness of sensitivity analysis. In addition, the selection of indicators should also consider the preferences of stakeholders (such as insurance companies).

TABLE 3 The investigation on the impact of data diversity on model prediction capabilities (* average R^2), taking the SMRF with 0.8-s fundamental period as an example.

Train data	Test data	
X	**No *X* year-related data**	**Only *X* year-related data**
$X = 43$	0.9821	0.8602
$X = 475$	0.9703	0.9037
$X = 975$	0.9601	0.8912
$X = 2475$	0.9884	0.8241
	All four scenarios	
All four scenarios	0.9852	

*Average R^2 is the average of the R^2 for various predicted responses (i.e., PIDR, PFA, RIDR).

The above illustrative example investigates some parameters that may have an impact on the results of seismic consequence estimation yet are mostly set to be deterministic in existing structural resilience assessment. After the sensitivity analysis, parameters that are considered insensitive can still be considered deterministic. In contrast, parameters that are considered to have a greater impact on the results should be carefully investigated to reduce their uncertainties, and if such investigation is not feasible, then their uncertainties are considered to be incorporated into the loss estimation. The inclusion of additional uncertainty is convenient for the FEMA P58 approach. As mentioned in Section 3.2, the Monte Carlo approach can be used to manage these uncertainties. More specifically, the seismic loss estimation is a highly repetitive process with a large number of realizations, each of which represents one possible loss outcome for the building, considering possible combinations of uncertain inputs. For example, if the uncertainty in replacement cost is expected to be incorporated, then by sampling its likely distribution at each loss estimation, its uncertainty will naturally be passed on to the results.

6.6 | The impact of data quality on the predictive capabilities of the surrogate model

In this paper, there are two main factors affecting data quality: one is the diversity of training data; and the other is the amount of training data. Here, the diversity of the training data mainly refers to whether it contains the various hazard scenarios considered in the sensitivity analysis. Then, the impact of data diversity is investigated in this section to understand the effect of missing data from a hazard scenario on the model's predictive capabilities, as shown in Table 3. In the investigation, the amount of training data is almost consistent with the aforementioned example, that is, 999, which means that the training set contains 333 data for each scenario (originally 250 for each of the four scenarios in the aforementioned example). There are also 200 independent data (50 for each of the four scenarios) as the testing set. It can be seen from the results that as long as the hazard scenarios are included in the training data, their corresponding prediction capabilities can be guaranteed. And if the corresponding hazard scenarios are not included in the training data, the prediction capabilities will be weakened. For the SMRF with 0.8-s fundamental period, the S_a corresponding to the four hazard scenarios are 0.09, 0.41, 0.58, and 0.87 g, respectively. For the case where the response data at 0.09 or 0.87 g is missing, the prediction capability in the corresponding scenario decreases more. This is because the structural nonlinear behavior under these two values is significantly different from the structural nonlinear behavior under other values. For the amount of training data, the requirement to give specific values may not be of practical significance. Compared to the large number of analyses required for sensitivity analysis, the data required for training data are relatively small. When the diversity of training data is satisfied, try to increase the amount of data so that its R^2 is above 0.90 to achieve better prediction performance.

6.7 | The impact of surrogate model predictive capabilities on the precision of the sensitivity analysis results

To investigate the impact of model predictive capabilities on the analysis results, the amount of training data is adjusted to 700 (corresponding to an average $R^2 = 0.8986$) and 500 (corresponding to an average $R^2 = 0.7585$) to obtain surrogate models with different predictive capabilities. And then the trained prediction model is used to facilitate sensitivity analysis.

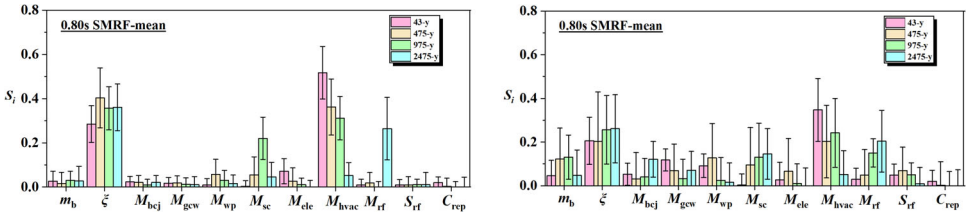


FIGURE 15 Sensitivity analysis results facilitated by models with different predictive capabilities (left: average $R^2 = 0.8986$; right: average $R^2 = 0.7585$).

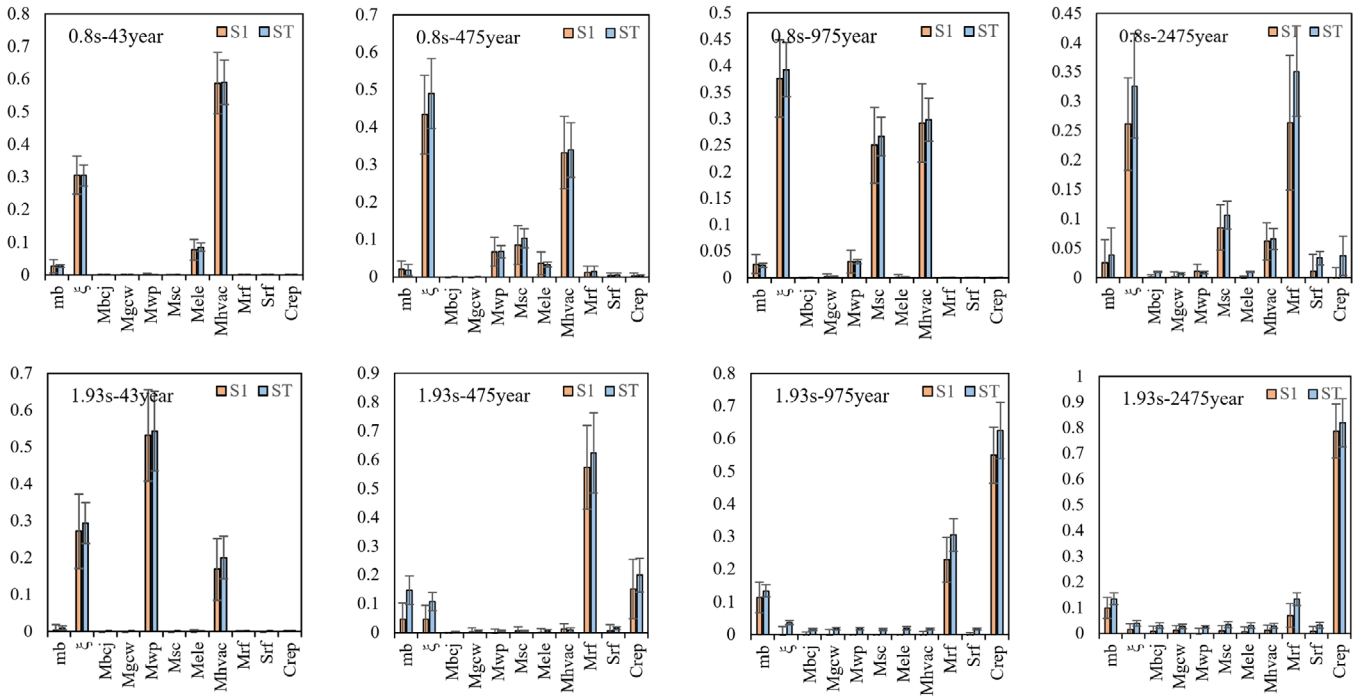


FIGURE 16 Comparison of first and total sensitivity index.

As shown in Figure 15, for the model with average $R^2 = 0.8986$, reasonable sensitivity analysis results can be obtained, but the error bars are larger than using the model with $R^2 = 0.9852$, which shows that the reduction in prediction capabilities brings disturbance to the convergence of the sensitivity analysis. For the model with $R^2 = 0.7585$, the prediction capability is greatly reduced, resulting in the inability to obtain reasonable and convergent sensitivity results. Therefore, in practice, controlling data quality as much as possible (increasing the diversity and quantity of training models) to obtain better prediction capabilities is conducive to obtaining stable sensitivity analysis results. It is worth noting that the computational cost required to build the surrogate model is relatively small compared to the number of model executions required for sensitivity analysis (i.e., 7168). Moreover, as more parameters are investigated, the computational cost of sensitivity analysis will be higher, and the advantages of using surrogate models will be greater.

6.8 | The degree of interaction effect between parameters

The difference between S_i and S_{Ti} can be used to measure the degree of the parameter interaction, as we mentioned in Section 5. As shown in Figure 16, the results of S_{Ti} are slightly higher than those of S_i, but the difference is not significant, and combined with the error bars given by the resampling technique, the synthesis suggests that the interaction effect between the investigated parameters is not strong enough to produce conclusions that are different from the first sensitivity index. The relative importance of each parameter can also be observed from the relative value of the sensitivity index.

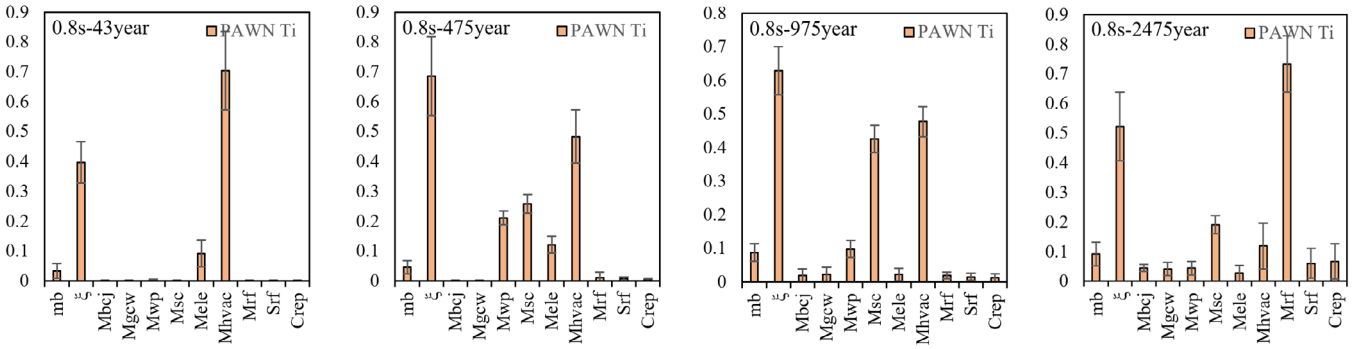


FIGURE 17 PAWN sensitivity index of four scenarios.

6.9 | Comparison with density-based sensitivity method (PAWN method)

Another sensitivity analysis method, called the PAWN method,[79] is also used to conduct the sensitivity analysis for comparison with the variance-based method. Different from the variance-based method, the PAWN method is a density-based method, where the entire model output distribution, rather than only its variance, is used to quantify the relative impact of the influence of the parameters on the model output. The PAWN method measures the sensitivity indices of parameter X_i by the distance between the unconditional CDF of Y, which is obtained by varying all parameters simultaneously, and the conditional CDFs of Y, which are obtained by varying all parameters but X_i.[80] More specifically, the distance between the conditional and unconditional CDFs is measured by the Kolmogorov–Smirnov statistic (KS),[81] as shown in Equation (33).

$$KS(X_i) = \max_Y \left| F_Y(Y) - F_{Y|X_i}(Y) \right| \tag{33}$$

where F_Y(Y) is the unconditional CDF of the output Y and $F_{Y|Xi}(Y)$ is the conditional CDF when X_i is fixed. If the $KS(X_i)$ is equal to zero, it means that the uncertainty of X_i does not affect the output. A large $KS(X_i)$, instead, indicates the uncertainty of X_i has a large effect on the output. Since KS depends on the conditioning values of X_i, the PAWN sensitivity index T_i considers a statistic over all possible values of X_i, as shown in Equation (34). The PAWN sensitivity index of a certain parameter varies between 0 and 1, the higher the value, the more influential of the parameter. Interested readers are referred to literature (e.g., Refs.[79,80,82]) for a more comprehensive description of the PAWN method.

$$T_i = statistic_{X_i}\left[KS(X_i)\right] \tag{34}$$

Taking the steel frame with 0.8-s fundamental period as an example, the relative ordering of the investigated parameters obtained using the PAWN method is the same as that of the variance-based method, as shown in Figure 17. Considering that the focus of this paper is not to compare the differences between different sensitivity methods, the investigated parameters or performance indicators are not intentionally selected to obtain potentially different sensitivity analysis results. However, according to the existing literature,[80] if the analysis is concerned with the first sensitivity index and one wants to know the degree of interaction between different parameters, the variance-based method is more suitable, because the PAWN method cannot investigate the first-order effect. And if one is only concerned with the sensitivity ranking after considering the interactions, the PAWN method is theoretically more advantageous because compared to variance-based methods, it still has good applicability when faced with performance indicators that exhibit highly skewed or multimodel distribution.[80] Notably, if possible, it is best to apply both methods for comparison, which can be achieved with the Python package *SALib*[60] employed in this paper.

7 | FURTHER DISCUSSION

So far, following the guidance of the proposed surrogate-aided sensitivity analysis framework, the entire procedure has been demonstrated with illustrative case studies to illustrate its effectiveness. However, considering the complexity of the problem intended to be solved, there is still room for continued study and improvement in the future.

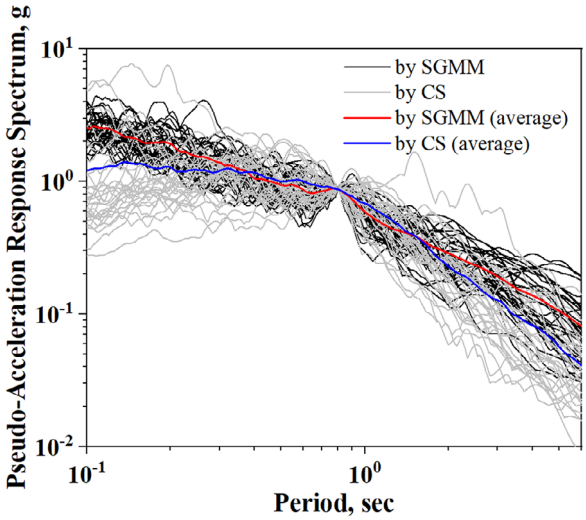


FIGURE 18 Comparison of acceleration response spectra of ground motions obtained by CS and SGMM.

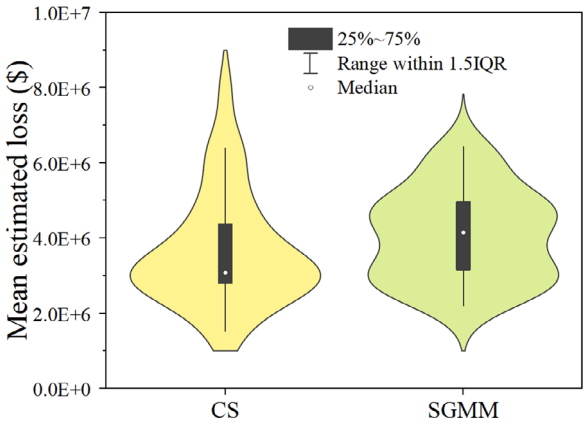


FIGURE 19 Comparison of mean estimated loss using ground motions obtained by CS and SGMM.

7.1 | Further improvement in the hazard compatibility of SGMM

As mentioned in Section 3, the SGMM is established by relating the parameters of stochastic functions to earthquake and site characteristics through predictive relationships. Since these predictive relationships do not explicitly consider seismic risk, the hazard compatibility of SGMM cannot be guaranteed. In this section, the hazard compatibility of selected SGMM is initially explored. A hazard-consisted ground motion selection method, conditional spectrum (CS) method, is used to select real ground motions matching the $S_a(0.8\ s)$ at the 2475-year hazard level. The implementation of the CS method refers to the research of Baker and Lee.[83] Then, the acceleration response spectra of the selected motions are compared with those of the simulated ground motions generated in Section 6, as shown in Figure 18. There exist differences between the distribution of acceleration response spectra of ground motions obtained by SGMM and CS. The average spectral acceleration curves from the two sets reveal that the motions obtained by SGMM pose a higher hazard. This difference results in a larger estimated loss concerning SGMM compared to CS, as shown in Figure 19. Similar findings can also be found in the study by Fayaz et al.,[84] which showed that even with the data-driven model established by a big earthquake database, the hazard compatibility of the ground motion models still cannot be guaranteed. The importance of hazard compatibility of selected ground motions to PBEE has recently been emphasized,[85,86] and therefore it is necessary to further improve the hazard compatibility of SGMM.

Existing research provides two approaches to address such issues. One is to use the CS method to select from simulated ground motions,[83,85] which can ensure that the selected simulated ground motions have the same mean and variance as the selected real ground motions. The other method is to modify the SGMM to establish compatibility with the seismic

hazard for specific seismicity scenarios and a given structure/site, which has been proven feasible on the ground motion model used in this study.[87,88] As for the potential differences between simulated and real ground motions, some validation methods[89] are also applicable to the selected stochastic ground motion model. Thus, the suitability of a stochastic GM model as part of proposed framework can be further improved by incorporating these methods in the future study.

7.2 | Further improvement in the usable scope of SGMM

The utilization of SGMM methods in PBEE serves to generate motion records tailored to local conditions, mitigating result bias that may arise from the use of unrealistic motions. Additionally, it streamlines the process by obviating the need for manual screening of motion records from databases, thereby reducing the reliance on specialized expertise. Nonetheless, it is essential to acknowledge the method's limitations regarding its useable scope. In the context of the method employed in this paper, the input parameters need to fall within the following specified range: $M \geq 6.0$, 10 km $\leq R \leq$ 100 km, $V_{s30} \geq 600$ m/s. If a smaller V_{s30} is expected, one needs to generate appropriate ground motions at the firm soil layer and then propagate through the soil softer soil using the soil dynamics method. This increases the complexity of the work and requires more professional knowledge. The limitation of useable scope is a common problem of existing ground motion modeling methods, which originates from data selection when constructing the motion model.

More specifically, the usable scope of the SGMM is mainly limited by the range of site and earthquake characteristics of the real ground motions used in its development. Therefore, incorporating more real ground motion records to construct the relationship between the motion model parameters and a larger range of site and earthquake characteristics, will effectively enhance the applicability of SGMM. However, after incorporating more ground motion data, restrictions of the originally adopted linear regression models may limit their capabilities in extracting complex nonlinear behaviors in the data. Thus, the use of machine learning as a statistical method is beneficial for constructing complex relationships between model parameters and site/earthquake characteristics.[84,90–92] Besides, the deep learning-based ground motion models also show a more powerful predictive capacity,[93] especially for big data without the assumption of predefined functional forms. These data-driven methods can extract critical information from massive earthquake records and find high-dimensional features to achieve better results. This broadens the applicable scope and simplifies implementation.

7.3 | Further improvement in the surrogate model

It should be noted that the modeling strategy used in this paper has only been tested for some SMRFs and some commonly used hazard scenarios, and its applicability to various other building structures (steel braced frame structures, steel frame-shear wall structures, or concrete structures, etc.) still requires further in-depth studies to modify the current GPR setting to cope with different cases. Considering the strong ability that the GP model has demonstrated in the existing literature,[25,43,46] its wide application to various types of structures is promising. Besides, the Python package for building GPR (GPyTorch[59]) selected in this paper allows for customization, which can be adjusted accordingly for specific training results. For example, the key kernel function can be changed from the "RBF function" used in this paper to the "square exponential (SE) function"[46] or "matérn and white combining function"[25] used in the literature, to adjust the model's suitability in different cases. Moreover, if there are more suitable ML techniques or existing models that can meet the sensitivity analysis needs, they can also be directly applied to the framework of this paper.

7.4 | Possible future extensions of the proposed framework

The seismic consequence estimation method emphasized in this paper is FEMA P-58, primarily due to its widespread use in research. Indeed, the current seismic consequence estimation methods are shifting towards recovery-based design paradigms,[94] representing a sustained effort to progress towards the overarching objective of establishing resilient cities. In the future, exploring these novel recovery-based methods would be worthwhile. Additionally, there are some existing open-source Python packages (e.g., TREADS developed by Hutt et al.[95]) whose open-source nature allows us to modify their internal parameters to study the sensitivity of some inputs or built-in databases.

Expanding surrogate model-aided sensitivity analysis to the regional level would also be a valuable area of research. A community or city is a complex system with intricate interconnections among different functional entities. Recently,

the resilience assessment of these systems has garnered significant attention. Assessing the resilience of such complex systems is undoubtedly time-consuming and fraught with uncertainties, presenting greater challenges compared to the sensitivity analysis of individual buildings. The sensitivity analysis of community/city-level assessment is essential.[96] It aids in identifying and understanding the various factors that can significantly impact the resilience of a community or city. Such analysis is beneficial for policymakers in formulating more effective and targeted strategies to combat the challenges posed by disasters.

8 | SUMMARY AND CONCLUSIONS

This paper proposed an efficient surrogate model-aided sensitivity analysis framework for seismic consequences estimation in buildings. The FEMA P-58 method is selected as the studied seismic consequence estimation method, and the SGMM method is utilized to generate ground motions for each hazard level of concern. The proposed framework reduces the computational cost of the estimation process by substituting the structural analysis part with the GP model, ultimately providing efficient and reliable results. Later, the variance-based GSA is performed to investigate the critical parameters of the estimation process. The whole procedure is implemented in Python following the object-oriented programming concept to construct an automatic workflow. Finally, the proposed framework is illustrated on two different three-story SMRFs to demonstrate its effectiveness and reliability. The main findings and conclusions can be drawn as follows:

- Using the SGMM method not only shortens the time of selecting motion records but can also be easily integrated into the proposed workflow. Meanwhile, suitable, realistic ground motions can be generated for different IM levels, reducing the estimation errors that may result from using unrealistic ground motions from other sites or large amplitude scaling.
- The proposed strategy for building a surrogate model, that is, building a specific surrogate model for a specific structure rather than a model with excess generalization capabilities, yields a model with good training results at a reduced computational cost. Moreover, this strategy aligns with the requirements of sensitivity analysis, wherein it is essential to select different surrogate model input features for different parameters that want to be investigated.
- The proposed framework replaces the expensive structural analysis part of seismic consequence estimation with a computationally cheap GP model, improving the efficiency of sensitivity analysis without losing much accuracy. This enables a higher number of model executions in the sensitivity analysis, thus ensuring the convergence of the sensitivity indices. In the illustrative example, only in terms of the computational cost required for sensitivity analysis, the use of the surrogate model increased the efficiency by 18 times.
- The sensitivity analysis results for earthquake consequence estimation vary depending on the structural system studied, the design scenario, and even the interested indicator (e.g., mean value or median value). Therefore, it is recommended to perform a dedicated sensitivity analysis on specific design information to obtain robust estimation results.

In summary, the proposed framework provides an efficient sensitivity analysis tool to evaluate the importance of interested parameters in the seismic consequences estimation process. This framework can be applied to assess the robustness of existing building resilience assessment and decision-making. Furthermore, owing to the adoption of the object-oriented programming concept, the framework can ultimately be integrated into computational tools with a graphical user interface. This integration simplifies the process for engineers, facilitating the practical application of sensitivity analysis of seismic consequence estimation methods. This, in turn, contributes to more informed and efficacious evaluation and decision-making in PBEE. This paper also acknowledges the current limitations of the proposed framework and outlines directions for future improvements. Specifically, it suggests the adoption of hazard-compatible ground motion selection methods or advanced modeling techniques to enhance the hazard compatibility of SGMM. Additionally, expanding the ground motion models' usable scope through the inclusion of more motion data and data-driven modeling is proposed. Moreover, beyond SMRF, the exploration of more structural types is necessary to validate the framework's applicability. Importantly, while the building-specific strategy employed for constructing surrogate models effectively supports the achievement of the research objective of this paper, it renders the strategy unsuitable for community-level or other nonbuilding-specific scenarios.

ACKNOWLEDGMENTS

The financial support from the Natural Science Foundation of China (NSFC) with Grant Nos. 52378182 and 52078366 are gratefully acknowledged. This study is also supported by the Top Discipline Plan of Shanghai Universities-Class I with

Grant No. 2022-3-YB-18 and Shanghai 2022 Science and Technology Innovation Action Plan Social Development Science and Technology Research Project with Grant No. 22dz1201700.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Jiajun Du https://orcid.org/0000-0001-5702-604X
Wei Wang https://orcid.org/0000-0003-1241-465X

REFERENCES

1. Baird A, Palermo A, Pampanin S. Façade damage assessment of concrete buildings in the 2011 Christchurch earthquake. *Struct Concr*. 2012;13(1):3-13. doi:10.1002/suco.201100040
2. Pan C, Wang H, Huang S, Zhang H. The great East Japan earthquake and tsunami aftermath: preliminary assessment of carbon footprint of housing reconstruction. In: Kontar YA, Santiago-Fandiño V, Takahashi T, eds. *Tsunami Events and Lessons Learned: Environmental and Societal Significance. Advances in Natural and Technological Hazards Research*. Springer; 2014:435-450. doi:10.1007/978-94-007-7269-4_25
3. Bruneau M, MacRae G. *Reconstructing Christchurch: A Seismic Shift in Building Structural Systems*. The Quake Centre, University of Canterbury; 2017.
4. Günay S, Mosalam KM. PEER performance-based earthquake engineering methodology, revisited. *J Earthq Eng*. 2013;17(6):829-858. doi:10.1080/13632469.2013.787377
5. Krawinkler H, Miranda E. Performance-based earthquake engineering. *Earthq Eng Eng Seismol Perform-Based Eng*. 2004;9:1-9.
6. Poland CD, Hill J, Sharpe RL, Soulages J. *Vision 2000: Performance Based Seismic Engineering of Buildings*. Structural Engineers Association of California; 1995.
7. ATC-40, *Seismic Evaluation and Retrofit of Concrete Buildings. Applied Technonlogy Council*; 1996.
8. FEMA 273. *NEHRP Guidelines for the Seismic Rehabilitation of Buildings*. Federal Emergency Management Agency; 1997.
9. FEMA 356. *Prestandart and Commentary for the Seismic Rehabilitation of Buildings*. Federal Emergency Management Agency; 2000.
10. Kircher CA, Whitman RV, Holmes WT. HAZUS earthquake loss estimation methods. *Nat Hazards Rev*. 2006;7(2):45-59. doi:10.1061/(ASCE)1527-6988(2006)7:2(45)
11. FEMA P-58. *Seismic Performance Assessment of Buildings*. 2nd ed. FEMA; 2018.
12. Almufti I, Willford M. REDi™ Rating System: Resilience-based Earthquake Design Initiative for the Next Generation of Buildings. Arup Co.; Published online2013.
13. Zhang R, Wang W, Alam MS. Seismic evaluation of friction spring-based self-centering braced frames based on life-cycle cost. *Earthq Eng Struct Dyn*. 2022;51:3393-3415. doi:10.1002/eqe.3728. n/a(n/a).
14. Hu S, Wang W, Qu B. Seismic economic losses in mid-rise steel buildings with conventional and emerging lateral force resisting systems. *Eng Struct*. 2020;204:110021. doi:10.1016/j.engstruct.2019.110021
15. Fang C, Ping Y, Zheng Y, Chen Y. Probabilistic economic seismic loss estimation of steel braced frames incorporating emerging self-centering technologies. *Eng Struct*. 2021;241:112486. doi:10.1016/j.engstruct.2021.112486
16. Ghasemof A, Mirtaheri M, Karami Mohammadi R. Multi-objective optimization for probabilistic performance-based design of buildings using FEMA P-58 methodology. *Eng Struct*. 2022;254:113856. doi:10.1016/j.engstruct.2022.113856
17. Burton HV, Lee JY, Moradi S, Dastmalchi S. Multi-objective performance-based design optimization of a controlled rocking steel braced frame system. In: Noroozinejad Farsangi E, Takewaki I, Yang TY, Astaneh-Asl A, Gardoni P, eds. *Resilient Structures and Infrastructure*. Springer; 2019:243-268. doi:10.1007/978-981-13-7446-3_10
18. Du J, Wang W, Lou T, Zhou H. Resilience and sustainability-informed probabilistic multi-criteria decision-making framework for design solutions selection. *J Build Eng*. 2023;71:106421. doi:10.1016/j.jobe.2023.106421
19. Anwar GA, Dong Y, Li Y. Performance-based decision-making of buildings under seismic hazard considering long-term loss, sustainability, and resilience. *Struct Infrastruct Eng*. 2021;17(4):454-470. doi:10.1080/15732479.2020.1845751
20. Terzic V, Villanueva PK, Saldana D, Yoo DY. Framework for modelling post-earthquake functional recovery of buildings. *Eng Struct*. 2021;246:113074. doi:10.1016/j.engstruct.2021.113074
21. ATC-138. Seismic performance assessment of buildings volume 8 – methodology for assessment of functional recovery time. Applied Technonlogy Council; 2021.
22. Molina Hutt C, Vahanvaty T, Kourehpaz P. An analytical framework to assess earthquake-induced downtime and model recovery of buildings. *Earthq Spectra*. 2022;38(2):1283-1320. doi:10.1177/87552930211060856
23. Bommer JJ, Acevedo AB. The use of real earthquake accelerograms as input to dynamic analysis. *J Earthq Eng*. 2004;8(sup001):43-91. doi:10.1080/13632460409350521
24. Luco N, Bazzurro P. Does amplitude scaling of ground motion records result in biased nonlinear structural drift responses? *Earthq Eng Struct Dyn*. 2007;36(13):1813-1835. doi:10.1002/eqe.695
25. Fayaz J, Torres-Rodas P, Medalla M, Naeim F. Assessment of ground motion amplitude scaling using interpretable Gaussian process regression: application to steel moment frames. *Earthq Eng Struct Dyn*. 2023;52(8):2339-2359. doi:10.1002/eqe.3810

26. Grigoriu M. To scale or not to scale seismic ground-acceleration records. *J Eng Mech*. 2011;137(4):284-293. doi:10.1061/(ASCE)EM.1943-7889.0000226
27. Rezaeian S, Der Kiureghian A. Simulation of synthetic ground motions for specified earthquake and site characteristics. *Earthq Eng Struct Dyn*. 2010;39(10):1155-1180. doi:10.1002/eqe.997
28. Lamprou A, Jia G, Taflanidis AA. Life-cycle seismic loss estimation and global sensitivity analysis based on stochastic ground motion modeling. *Eng Struct*. 2013;54:192-206. doi:10.1016/j.engstruct.2013.04.001
29. Cao XY, Feng DC, Li Y. Assessment of various seismic fragility analysis approaches for structures excited by non-stationary stochastic ground motions. *Mech Syst Signal Process*. 2023;186:109838. doi:10.1016/j.ymssp.2022.109838
30. Pang R, Chen K, Fan Q, Xu B. Stochastic ground motion simulation and seismic damage performance assessment of a 3-D subway station structure based on stochastic dynamic and probabilistic analysis. *Tunn Undergr Space Technol*. 2022;126:104568. doi:10.1016/j.tust.2022.104568
31. Vetter C, Taflanidis AA. Global sensitivity analysis for stochastic ground motion modeling in seismic-risk assessment. *Soil Dyn Earthq Eng*. 2012;38:128-143. doi:10.1016/j.soildyn.2012.01.004
32. Porter KA, Beck JL, Shaikhutdinov RV. Sensitivity of building loss estimates to major uncertain variables. *Earthq Spectra*. 2002;18(4):719-743. doi:10.1193/1.1516201
33. Baker JW, Cornell CA. *Uncertainty Specification and Propagation for Loss Estimation Using FOSM Method*. Pacific Earthquake Engineering Research Center, College of Engineering; 2003.
34. Dyanati M, Huang Q, Roke D. Sensitivity analysis of seismic performance assessment and consequent impacts on loss analysis. *Bull Earthq Eng*. 2017;15(11):4751-4790. doi:10.1007/s10518-017-0150-6
35. Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Comput Phys Commun*. 2010;181(2):259-270. doi:10.1016/j.cpc.2009.09.018
36. Cremen G, Baker JW. Variance-based sensitivity analyses and uncertainty quantification for FEMA P-58 consequence predictions. *Earthq Eng Struct Dyn*. 2021;50(3):811-830. doi:10.1002/eqe.3370
37. Del Giudice L, Marelli S, Sudret B, Vassiliou MF. Global sensitivity analysis of 3D printed material with binder jet technology by using surrogate modeling and polynomial chaos expansion. *Prog Addit Manuf*. 2023. doi:10.1007/s40964-023-00459-y. Published online May 30.
38. Tesfamariam S, Loeppky JL, Bezabeh MA. Gaussian process model for maximum and residual drifts of timber-steel hybrid building. *Struct Infrastruct Eng*. 2017;13(5):554-566. doi:10.1080/15732479.2016.1175482
39. Ali Anwar G, Dong Y. Surrogate-based decision-making of community building portfolios under uncertain consequences and risk attitudes. *Eng Struct*. 2022;268:114749. doi:10.1016/j.engstruct.2022.114749
40. Soleimani-Babakamali MH, Zaker Esteghamati M. Estimating seismic demand models of a building inventory from nonlinear static analysis using deep learning methods. *Eng Struct*. 2022;266:114576. doi:10.1016/j.engstruct.2022.114576
41. Luo H, Paal SG. Data-driven seismic response prediction of structural components. *Earthq Spectra*. 2022;38(2):1382-1416. doi:10.1177/87552930211053345
42. Gudipati VK, Cha EJ. Surrogate modeling for structural response prediction of a building class. *Struct Saf*. 2021;89:102041. doi:10.1016/j.strusafe.2020.102041
43. Suarez D, Rubini G, Gentile R, Galasso C. Gaussian process regression-based surrogate modelling for direct loss-based seismic design of low-rise base-isolated structures. *Procedia Struct Integr*. 2023;44:1728-1735. doi:10.1016/j.prostr.2023.01.221
44. Nguyen HD, LaFave JM, Lee YJ, Shin M. Rapid seismic damage-state assessment of steel moment frames using machine learning. *Eng Struct*. 2022;252:113737. doi:10.1016/j.engstruct.2021.113737
45. Li S, Farrar C, Yang Y. Efficient regional seismic risk assessment via deep generative learning of surrogate models. *Earthq Eng Struct Dyn*. 2023;52(11):3435-3454. doi:10.1002/eqe.3849
46. Gentile R, Galasso C. Gaussian process regression for seismic fragility assessment of building portfolios. *Struct Saf*. 2020;87:101980. doi:10.1016/j.strusafe.2020.101980
47. Zaker Esteghamati M, Flint MM. Developing data-driven surrogate models for holistic performance-based assessment of mid-rise RC frame buildings at early design. *Eng Struct*. 2021;245:112971. doi:10.1016/j.engstruct.2021.112971
48. FEMA. Hazus earthquake model technical manual—Hazus 5.1. Published online 2022. Accessed May 9, 2023. https://www.fema.gov/sites/default/files/documents/fema_hazus-earthquake-model-technical-manual-5-1.pdf
49. Zhu M, McKenna F, Scott MH. OpenSeesPy: Python library for the OpenSees finite element framework. *SoftwareX*. 2018;7:6-11. doi:10.1016/j.softx.2017.10.009
50. McKenna F, Scott MH, Fenves GL. Nonlinear finite-element analysis software architecture using object composition. *J Comput Civ Eng*. 2010;24(1):95-107. doi:10.1061/(ASCE)CP.1943-5487.0000002
51. Guan X, Burton H, Sabol T. Python-based computational platform to automate seismic design, nonlinear structural model construction and analysis of steel moment resisting frames. *Eng Struct*. 2020;224:111199. doi:10.1016/j.engstruct.2020.111199
52. Burton H. SPA_INFILL. SPA_INFILL. Accessed October 5, 2023. https://www.henryburtonjr.com/spa-infill
53. Ibarra LF, Medina RA, Krawinkler H. Hysteretic models that incorporate strength and stiffness deterioration. *Earthq Eng Struct Dyn*. 2005;34(12):1489-1511. doi:10.1002/eqe.495
54. Lignos D. Sidesway Collapse of Deteriorating Structural Systems under Seismic Excitations. Stanford University; 2008. Accessed September 24, 2023. https://search.proquest.com/openview/2e867a762ed146fdd5a0dc116a7b8e5a/1?pq-origsite=gscholar&cbl=18750

55. Lignos DG, Krawinkler H. Deterioration modeling of steel components in support of collapse prediction of steel moment frames under earthquake loading. *J Struct Eng-Rest*. 2011;137(11):1291.
56. Applied Technology Council. *Guidelines for Nonlinear Structural Analysis and Design of Buildings. Part IIa—Steel Moment Frames*. National Institute of Standards and Technology; 2017:NIST GCR 17-917-46v2. doi:10.6028/NIST.GCR.17-917-46v2
57. Lignos DG, Hartloper AR, Elkady A, Deierlein GG, Hamburger R. Proposed updates to the ASCE 41 nonlinear modeling parameters for wide-flange steel columns in support of performance-based seismic engineering. *J Struct Eng*. 2019;145(9):04019083. doi:10.1061/(ASCE)ST.1943-541X.0002353
58. Krawinkler H. Shear in beam-column joints in seismic design of steel frames. *Eng J*. 1978;15(3). Accessed September 24, 2023. https://trid.trb.org/view/81133
59. Gardner J, Pleiss G, Weinberger KQ, Bindel D, Wilson AG. GPyTorch: blackbox matrix-matrix Gaussian process inference with GPU acceleration. *Adv Neural Inf Process Syst*. 2018;31. Accessed September 25, 2023. https://proceedings.neurips.cc/paper/2018/hash/27e8e17134dd7083b050476733207ea1-Abstract.html
60. Herman J, Usher W. SALib: an open-source Python library for sensitivity analysis. *J Open Source Softw*. 2017;2(9):97. doi:10.21105/joss.00097
61. Iwanaga T, Usher W, Herman J. Toward SALib 2.0: advancing the accessibility and interpretability of global sensitivity analyses. *Socio-Environ Syst Model*. 2022;4:18155-18155. doi:10.18174/sesmo.18155
62. Douglas J, Aochi H. A survey of techniques for predicting earthquake ground motions for engineering purposes. *Surv Geophys*. 2008;29(3):187-220. doi:10.1007/s10712-008-9046-y
63. Rezaeian S. *Stochastic Modeling and Simulation of Ground Motions for Performance-Based Earthquake Engineering*. Pacific Earthquake Engineering Research Center; 2010.
64. Vamvatsikos D, Allin Cornell C. Incremental dynamic analysis. *Earthq Eng Struct Dyn*. 2002;31(3):491-514. doi:10.1002/eqe.141
65. Archer GEB, Saltelli A, Sobol IM. Sensitivity measures, anova-like techniques and the use of bootstrap. *J Stat Comput Simul*. 1997;58(2):99-120. doi:10.1080/00949659708811825
66. Saltelli A, Chan K, Scott EM. *Sensitivity Analysis*. Wiley; 2000.
67. Sobol′ IM. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math Comput Simul*. 2001;55(1):271-280. doi:10.1016/S0378-4754(00)00270-6
68. FEMA. FEMA P-2012: assessing seismic performance of buildings with configuration irregularities: calibrating current standards and practices. Published online 2018.
69. FEMA P-58. Supporting materials. Accessed September 25, 2023. https://femap58.atcouncil.org/supporting-materials
70. Ellingwood B, Galambos TV, MacGregor JG, Cornell CA. *Development of a Probability Based Load Criterion for American National Standard A58: Building Code Requirements for Minimum Design Loads in Buildings and Other Structures*. Department of Commerce, National Bureau of Standards; 1980. Accessed September 26, 2023. https://books.google.com/books?hl=zh-CN&lr=&id=fT6YDP3R-HAC&oi=fnd&pg=PA11&dq=Development+of+a+Probability-Based+Load+Criterion+for+American+National+Standar&ots=zETzUGso1I&sig=yYtBUyVC2qjZCzgynqi36T1bB7
71. Bernal D, Döhler M, Kojidi SM, Kwan K, Liu Y. First mode damping ratios for buildings. *Earthq Spectra*. 2015;31(1):367-381. doi:10.1193/101812EQS311M
72. Debnath R, Halder L. A comparative study of the seismic provisions of indian seismic code IS 1893–2002 and draft Indian Code IS 1893:2016. In: Rao ARM, Ramanjaneyulu K, eds. *Recent Advances in Structural Engineering. Lecture Notes in Civil Engineering*. Springer; 2019:151-160. doi:10.1007/978-981-13-0365-4_13. Lecture Notes in Civil Engineering.
73. GB 50011–2010. Code for seismic design of buildings. Ministry of Housing and Urban-Rural Development of the People's Republic of China; 2016 (in Chinese).
74. McCormick J, Aburano H, Ikenaga M, Nakashima M, Permissible residual deformation levels for building structures considering both safety and human elements. 2008. Accessed May 11, 2023. https://www.semanticscholar.org/paper/PERMISSIBLE-RESIDUAL-DEFORMATION-LEVELS-FOR-BOTH-McCormick-Aburano/efb47ee40e9205b5f916b5920d849a2b8cbe6a73
75. Dimopoulos AI, Tzimas AS, Karavasilis TL, Vamvatsikos D. Probabilistic economic seismic loss estimation in steel buildings using post-tensioned moment-resisting frames and viscous dampers: seismic Losses in Post-Tensioned Steel Buildings with Viscous Dampers. *Earthq Eng Struct Dyn*. 2016;45(11):1725-1741. doi:10.1002/eqe.2722
76. SP3 Risk. Input page: building layout—knowledge base—home. Accessed September 26, 2023. https://support.sp3risk.com/space/KnowledgeBase/1777633262/Input+Page
77. U.S. Geological Survey. Unified hazard tool. Accessed September 26, 2023. https://earthquake.usgs.gov/hazards/interactive/
78. Rasmussen CE, Williams CKI. *Gaussian Processes for Machine Learning*. MIT Press; 2006.
79. Pianosi F, Wagener T. A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environ Model Softw*. 2015;67:1-11. doi:10.1016/j.envsoft.2015.01.004
80. Khorashadi Zadeh F, Nossent J, Sarrazin F, et al. Comparison of variance-based and moment-independent global sensitivity analysis approaches by application to the SWAT model. *Environ Model Softw*. 2017;91:210-222. doi:10.1016/j.envsoft.2017.02.001
81. Kolmogorov-Smirnov AN, Kolmogorov A, Kolmogorov M, Sulla determinazione empírica di uma legge di distribuzione. 1933. Accessed January 17, 2024. https://www.semanticscholar.org/paper/Sulla-determinazione-emp%C3%ADrica-di-uma-legge-di-Kolmogorov-Smirnov-Kolmogorov/c859d88a2b4386e0be5d4b5c91c8dc8996ef6d64
82. Pianosi F, Wagener T. Distribution-based sensitivity analysis from a generic input-output sample. *Environ Model Softw*. 2018;108:197-207. doi:10.1016/j.envsoft.2018.07.019

83. Baker JW, Lee C. An improved algorithm for selecting ground motions to match a conditional spectrum. *J Earthq Eng*. 2018;22(4):708-723. doi:10.1080/13632469.2016.1264334
84. Fayaz J, Medalla M, Torres-Rodas P, Galasso C. A recurrent-neural-network-based generalized ground-motion model for the Chilean subduction seismic environment. *Struct Saf*. 2023;100:102282. doi:10.1016/j.strusafe.2022.102282
85. Alvarez-Sanchez LG, Iñárritu PG de Q, Šipčić N, Kohrangi M, Bazzurro P. Hazard-consistent simulated earthquake ground motions for PBEE applications on stiff soil and rock sites. *Earthq Eng Struct Dyn*. 2023;52(15):4900-4918. doi:10.1002/eqe.3987
86. Zaker Esteghamati M. A holistic review of GM/IM selection methods from a structural performance-based perspective. *Sustainability*. 2022;14(20):12994. doi:10.3390/su142012994
87. Tsioulou A, Taflanidis AA, Galasso C. Hazard-compatible modification of stochastic ground motion models. *Earthq Eng Struct Dyn*. 2018;47(8):1774-1798. doi:10.1002/eqe.3044
88. Tsioulou A, Taflanidis AA, Galasso C. Modification of stochastic ground motion models for matching target intensity measures. *Earthq Eng Struct Dyn*. 2018;47(1):3-24. doi:10.1002/eqe.2933
89. Tsioulou A, Taflanidis AA, Galasso C. Validation of stochastic ground motion model modification by comparison to seismic demand of recorded ground motions. *Bull Earthq Eng*. 2019;17(6):2871-2898. doi:10.1007/s10518-019-00571-x
90. Karimzadeh S, Mohammadi A, Hussaini SMS, Caicedo D, Askan A, Lourenço PB. ANN-based ground motion model for Turkey using stochastic simulation of earthquakes. *Geophys J Int*. 2024;236(1):413-429. doi:10.1093/gji/ggad432
91. Khosravikia F, Clayton P. Machine learning in ground motion prediction. *Comput Geosci*. 2021;148:104700. doi:10.1016/j.cageo.2021.104700
92. Sreejaya KP, Basu J, Raghukanth STG, Srinagesh D. Prediction of ground motion intensity measures using an artificial neural network. *Pure Appl Geophys*. 2021;178(6):2025-2058. doi:10.1007/s00024-021-02752-9
93. Li C, Ji D, Zhai C, Ma Y, Xie L. Vertical ground motion model for the NGA-West2 database using deep learning method. *Soil Dyn Earthq Eng*. 2023;165:107713. doi:10.1016/j.soildyn.2022.107713
94. Li L, Chang-Richards A, Boston M, Elwood K, Molina Hutt C. Post-disaster functional recovery of the built environment: a systematic review and directions for future research. *Int J Disaster Risk Reduct*. 2023;95:103899. doi:10.1016/j.ijdrr.2023.103899
95. Molina Hutt C. TREADS: tool for recovery estimation and downtime simulation of buildings. Published online May 20, 2023. Accessed September 30, 2023. https://github.com/carlosmolinahutt/treads
96. Blagojević N, Didier M, Stojadinović B. Quantifying component importance for disaster resilience of communities with interdependent civil infrastructure systems. *Reliab Eng Syst Saf*. 2022;228:108747. doi:10.1016/j.ress.2022.108747

How to cite this article: Du J, Wang W. Surrogate model-aided global sensitivity analysis framework for seismic consequences estimation in buildings. *Earthquake Engng Struct Dyn*. 2024;1-30. https://doi.org/10.1002/eqe.4116