

Assessment of ground motion amplitude scaling using interpretable Gaussian process regression: Application to steel moment frames

Jawad Fayaz^{1,2}  | Pablo Torres-Rodas³ | Miguel Medalla⁴ | Farzad Naeim^{5,6}

¹School of Computing, Engineering, and Digital Technologies, Teesside University, Middlesbrough, UK

²Department of Civil, Environmental, and Geomatic Engineering, University College London, London, UK

³Colegio de Ciencias e Ingenierías, Campus Cumbayá, Universidad San Francisco de Quito, Quito, Ecuador

⁴Facultad de Ingeniería y Ciencias Aplicadas, Universidad de los Andes, Santiago, Chile

⁵Farzad Naeim Inc., Irvine, USA

⁶Department of Civil and Environmental Engineering, University of California Irvine, Irvine, USA

Correspondence

Jawad Fayaz, School of Computing, Engineering, and Digital Technologies, Teesside University, Middlesbrough, TS1 3BX, UK.

Email: j.fayaz@tees.ac.uk

Abstract

The process of ground motion selection and scaling is an integral part of hazard- and risk-consistent seismic demand analysis of structures. Due to the lack of ground motion records that naturally possess high amplitude and intensity, the research community generally relies on scaling the records to match a target hazard intensity level. The scaling factors used are frequently as high as 10. Due to the criticism received in previous research studies, the extent of amplitude scaling and its process has become a matter of debate, and various constraints on the scaling factors have been proposed. The primary argument against unrestricted amplitude scaling is the unrealistic nature of the scaled records and the possible biases caused in the engineering demand parameters (EDPs) of structures. This study presents a framework to utilize machine-learning and statistical techniques for the assessment of ground motion amplitude scaling for nonlinear time-history analysis (NTHA) of structures. The framework utilizes Bayesian non-parametric Gaussian process regressions (GPRs) as surrogate models to obtain statistical estimates of EDPs for scaled and unscaled ground motions. The GPR surrogate models are developed based on a large-scale analysis of five steel moment frames (SMFs) using 200 unscaled as-recorded ground motions for ten spectral acceleration levels, ($S_a(T_1)$) (ranging from 0 g to maximum considered earthquake, MCE) and 2500 scaled ground motions representing 50 scale factors (SF), and the 10 $S_a(T_1)$ levels for each SMF. For each building, two types of EDPs are considered: i) peak inter-story drift ratio (PIDR) and ii) peak floor acceleration (PFA). To provide a better interpretation of the GPR surrogate models, the concept of explainable artificial intelligence (i.e., Shapley additive explanation, SHAP) is used to obtain insights into the decision-making process of the GPR models with respect to the SF and $S_a(T_1)$. Then, for the 10 $S_a(T_1)$ levels, the GPR-based EDP estimates under scaled ground motions corresponding to 50 different SFs are compared with the EDP estimates of unscaled ground motions. The

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Earthquake Engineering & Structural Dynamics* published by John Wiley & Sons Ltd.

comparison is conducted using Kolmogorov–Smirnov (KS) statistical hypothesis test. Results indicate that the range of allowable SF s depends on two factors: i) intensity level (characterized by $S_a(T_1)$), and ii) the dynamic properties of the building. In general, it is noticed that allowable SF s range between 0.5 and 3.0 for PIDRs, and from 0.6 to 2.0 for PFAs. Finally, the EDP between the unscaled and scaled ground motions are adhered to various discrepancies observed in different intensity measures representing amplitude-, duration-, energy-, and frequency-content of the two sets of ground motions.

KEYWORDS

amplitude scaling, Bayesian non-parametric, explainability, Gaussian process regression, ground motion scaling, interpretability, machine learning, scale factors, surrogate models, XAI

1 | INTRODUCTION

The advances in computational capacities have driven the use of nonlinear time history analysis (NTHA) to assess the performance of buildings under the framework of performance-based earthquake engineering (PBEE). In this context, an essential step is the selection and scaling process of ground motions which are used for the NTHA of the building structures. The method of ground motion selection and scaling forms a link between the fields of seismology and earthquake engineering. Considering the limited number of strong recorded ground motions available in current ground motion databases and hence the inability to represent the typical hazard levels for performance-based analysis (e.g., design basis earthquake DE, maximum considered earthquake MCE, or collapse), it is a common practice to select moderate-to-low amplitude ground motions and scale them to reach higher spectral amplitudes that represent the required hazard levels. Thus, code provisions such as the American Society of Civil Engineers (ASCE 7–16)¹ and performance-based guidelines such as the Los Angeles Tall Buildings Structural Design Council (LATBSDC),² among others, have included procedures to select and scale recorded ground motions for conducting NTHA.

For decades, several researchers (specially seismologists) have criticized the validity of scaling ground motion records to achieve the desired intensity level.^{3,4} The primary concerns in this regard are the differences in the frequency content and duration of ground motions recorded from different magnitudes and distances. It is well-known that for a given rupture distance (R_{rup}) and a class site, the spectral acceleration, on average, tends to increase with an increase in earthquake moment magnitude (M_w). On the other hand, for a given magnitude and site class, an increase in the source-to-distance (such as R_{rup}) leads to a decrease in spectral acceleration amplitude (attenuation). Thus, amplitude scaling can be interpreted as an artificial increment or decrement of M_w or R_{rup} , and thereby can lead to inconsistencies in the correlations between intensity measures (IMs) and corresponding M_w or R_{rup} of the ground motions.⁴ Due to this, studies such as Bommer and Acevedo⁵ have postulated that M_w strongly influences the frequency content and strong motion duration and must be considered as a parameter in the ground motion selection process. They proposed limiting M_w of the selected ground motion records to be within ± 0.20 units of the disaggregated M_w for the desired hazard level.

As a remedial measure, many studies have proposed to set limits on the scaling factors that can be used in the ground motion selection and scaling process. For example, Krinitzsky and Chang⁶ proposed a scaling factor limit of 4 without exhaustive scientific support for this criterion. Later on, Cornell and his group^{7,8} studied the nonlinear response of a multiple-degree-of-freedom (MDOF) system subjected to ground motion records grouped into four different bins in terms of M_w and R_{rup} to explore alternative criteria for the record selection and scaling. These scenarios included unscaled and scaled ground motion records. The study concluded that the bias from scaled records (in terms of the MDOF's response) is minimized when the records are scaled to the median spectral acceleration at the structural first period ($S_a(T_1)$) of the corresponding M_w and R_{rup} bin. These results were further confirmed by Iervolino and Cornell,⁹ where the study concluded that the nonlinear response of structures is not affected by conducting record selection using M_w and R_{rup} scenarios, by scaling records for sites with no directivity effects and for records without systematic peak-valley effects.

Subsequent studies on the effect of amplitude scaling focus on estimates of the potential bias introduced by scaling ground motion records on the seismic performance of structures. For instance, Luco and Bazzurro¹⁰ quantified the bias introduced in the nonlinear structural drift response of single-degrees-of-freedom (SDOFs) and MDOF systems by scaling records randomly selected from a $M_w - R_{rup}$ range to a target intensity level. The scaled ground motions were then used

NOVELTY

This article is a step towards utilization of artificial intelligence and machine learning concepts to solve earthquake engineering problems such as ground motion selection and scaling. In particular, this article utilizes Gaussian process regression (GPR) to carefully develop a data-driven surrogate model for statistical comparison of structural demands for steel moment frames obtained under scaled and unscaled ground motion records. This is one of the first studies that clearly provides the range of allowable scale factors for ground motion scaling different intensity levels and dynamic properties of the building structures. Furthermore, to provide a better interpretation of the GPR surrogate models, the concept of explainable artificial intelligence (i.e., Shapley additive explanation, SHAP) is used to acquire insights into the decision-making process of the GPR models. Hence, this study also aims to implement solutions to reduce the black-box nature of machine learning models and provide engineering interpretation to the models.

Due to its novelty, its topical subject matter, and its strong potential to have a high impact on the field of earthquake engineering, we believe this work will be of broad and considerable interest to the readers of EESD and its special issue on *AI and Data-driven Methods in Earthquake Engineering*.

for NTHA of the SDOF and MDOF systems, and the obtained responses were compared with responses from unscaled records corresponding to the specific intensity level. Thus, bias was calculated as the ratio between the median structural response of scaled records to the median structural response of unscaled records at the target $S_a(T_1)$. Their analyses indicated that the record scaling introduces a bias in the nonlinear response of the systems and their engineering demand parameters (EDPs) (i.e., lateral drifts). Specifically, this bias was found to depend on the scaling factor, the fundamental period of the structure, the overall overstrength of the structure, higher mode effects, and on the $M_w - R_{rup}$ range of the records scaled. The computed bias ranged from 15% to 60% for SDOF systems (with different periods) for scaling factors from 2 to 10, respectively, while for MDOF systems, the bias spanned from 25% to 80% for scaling factors from 2 to 10, respectively. The study concluded that this bias is due to the differences in the response spectra of the scaled and unscaled records. Similar observations were confirmed by Backer and Cornell,¹¹ who proposed to select ground motions based on the parameter ϵ , which is considered a measure of spectral shape.

More recently, Zacharenaki et al.,¹² Du et al.,¹³ and Davalos and Miranda^{3,4} studied the influence of amplitude scaling of ground motions on the probability of collapse of SDOF and MDOF systems with different fundamental periods. Specifically, Davalos and Miranda³ investigated the impact of using spectral shape for the ground motion selection process on the assessment of bias in the seismic response of several SDOF and MDOFs systems, including 2- and 4-story reinforced concrete moment frames. This study compared the median structural response and collapse probabilities using scaled and unscaled records. It indicated that the bias in drifts (overestimation) is observed even though the ground motions are selected and scaled considering the mean and variance of spectral shape. The study concluded that the bias occurs due to the presence of a large number of pulses with large incremental velocities in the scaled ground motions compared to the unscaled records. Thus, the differences in energy distribution, frequency, and duration of the scaled and unscaled ground motions for the same intensity level can be adhered to explain the statistical mismatch in the corresponding EDPs.

On the other hand, Du et al.¹³ investigated the effect of scaling factor limits on ground motions selected by the conditional spectra criterion¹⁴ for four different hazard scenarios. The authors studied the effect of scaling on the ground motions characteristics, conducted NTHA on SDOF and MDOF systems, and recommended a scaling limit of 3 to 5 for general use when ground motions are selected from the Next Generation Attenuation West 2 (NGA-West 2) database.¹⁵ Modern studies on the topic have included artificial intelligence techniques in their analyses. For instance, Fayaz and Zareian¹⁶ presented an algorithm to simulate site-based ground motions that reach a target IM between a range of periods. This algorithm used predictive relations between the model parameters and the spectral acceleration spectrum of the site-based ground motions and was used to generate ground motions for the NTHA of a bridge structure. The response of this bridge was compared to three sets of ground motions, including unscaled, scaled, and simulated ground motions. Results from this study indicate that the set of scaled ground motions led to significant bias in the EDP estimates compared to the unscaled and simulated sets. Besides these studies, Tsalouchidis and Adam (2022)¹⁷ investigated the impact of amplitude scaling on the structural response of buildings. For this purpose, ten steel moment frames (SMFs) were subjected to NTHA. They studied the potential biases introduced in the peak inter-story drift ratio (PIDR) due to the scaling of ground motions. The study utilized cloud analysis and generalized linear regression models to deduce any

changes in the PIDR trends caused by the scaling of spectrally compatible and scenario-compatible ground motions. The comparisons were made for scaling factors ranging from 2 to 20 and were primarily conducted by checking the ratios between scaled and unscaled EDPs. They concluded that the scaling process doesn't cause a major bias in the EDPs however recommended caution while using large scaling factors. Although this finding is consistent with some earlier studies on the topic (e.g.,^{9,18}), many modern studies (e.g.,^{3,4,10}) have indicated that amplitude scaling can lead to major biases even if ground motions are selected with proper consideration of the spectral shape. One of the major limitations of the Tsalouchidis and Adam (2022)¹⁷ study relies in their ground motion selection criteria, where it is evident that the ground motions with smaller IMs are over-represented in the dataset of cloud analysis which can cause biases in the regression models and hence in the conclusions.

In summary, various research studies (e.g.,^{3,4,10-13}) have indicated that the scaling of the ground motions for NTHA leads to bias in the EDPs of structural systems. Most studies agree that the bias is not statistically significant for smaller scaling factors. However, there is no clear scientific rationale for what scaling factors are acceptable and do not lead to substantial biases in the structural demands. Also, ground motion scaling for hazard levels of engineering interest such as DE or MCE is barely studied. Furthermore, it is evident from the previous studies that allowable scaling factors may also vary based on the target intensity level. Moreover, most of these studies were conducted on non-degrading systems (e.g.,⁷) or did not include structures of different configurations, limiting the generalization of their results. In addition, the selected ground motions for their analyses do not provide a vast range of intensity levels or focus on a specific limit state such as collapse level (e.g.,^{3,4}). No proper alternatives were explored to overcome the limitations of high intensity ground motion records. Furthermore, only traditional statistical tools were used to estimate bias (e.g., comparing the mean drift response from scaled or unscaled records) and no information was provided about their uncertainty distributions.

Against this backdrop, this study proposes a comprehensive framework based on principles of machine learning and statistical surrogacy to assess the impact of ground motion amplitude scaling on NTHA of structural systems. In particular, the paper presents a sophisticated statistical investigation using Gaussian process regression (GPR) to provide insights concerning the bias introduced by scaling ground motion records for NTHA of steel moment frames. Specifically, the objectives of this paper are threefold: a) investigate potential bias in EDPs caused by scaling of ground motions using machine-learning-based surrogate models; b) understand and interpret the decision-making process of developed surrogate-models using the concepts of explainable artificial intelligence (XAI); c) provide recommendations on the use of scale limits based on the target intensity level of ground motions. For this study, a series of NTHA is conducted on five archetype SMFs (i.e., 2-,4-,8-,12-, and 20-story) to obtain two EDPs: i) peak inter-story drift ratio (PIDR); and ii) peak floor acceleration (PFA). A total of 2500 scaled and 200 unscaled ground motion records are selected from the NGAWest2 database¹⁵ to match ten intensity levels ranging from 0 g to MCE level for each SMF. The results of the analyses are then used to develop GPR-based surrogate models for scaled and unscaled demands, which are then interpreted and explained using Shapley additive explanation (SHAP) to provide insights. The surrogate models are then used to sample the EDP posterior distributions for different intensity levels and scaling factors (SF). The distributions are utilized and compared to understand the joint impact of SF and $S_a(T_1)$ on the EDPs; the comparison is conducted through hypothesis testing of the scaled and unscaled EDP distributions to provide recommendations on the record scaling process. Finally, the observed bias in the EDPs under scaled ground motions as compared to the unscaled ground motions is briefly investigated by analyzing the differences in the intensity measures representing amplitude-, duration-, energy-, and frequency- content of the two sets of ground motions.

1.1 | Assessment framework

The framework is based on a selection of ground motion records from an extensive ground motion database (such as the NGAWest2 database¹⁵) and relies on large-scale NTHA of the selected structural application. The general scientific background of this investigation is based on a ten-stepped procedure described below (the text provided in parenthesis for the steps explains the corresponding items used in this study).

1. Select a ground motion IM (e.g., $S_a(T_1)$) and corresponding target IM level (e.g., MCE level) based on the structural application (e.g., steel SMFs) and hazard level of interest (e.g., MCE level).
2. Obtain n stripes of IMs between the minimum IM level of interest and the target IM level by dividing the range into n groups (e.g., $n = 10$ equal stripes between 0 g and MCE level).

3. For each of the n IM stripes, from the ground motion database, obtain i number of unscaled ground motions possessing IM closest to the IM stripe (e.g., $i = 20$ is used to select 20 unscaled ground motions). Hence a total of $n \times i$ unscaled ground motions are selected (e.g., $10 \times 20 = 200$ unscaled ground motions are selected for each SMF).
4. For each of the n IM stripes, compute the average IM of the selected i unscaled ground motions.
5. Select m values of SFs to obtain scaled ground motions for the computed average IM of each IM stripe (e.g., $m = 50$ is used with 25 divisive SFs (< 1) and 25 multiplicative SFs (> 1)).
6. Obtain j scaled ground motions (e.g., $j = 5$ is used to select 5 scaled ground motions) for each of the m values of SFs by selecting the j ground motions that require m^{th} SF to attain the average IM of the i unscaled ground motions. Hence a total of $n \times j \times m$ scaled ground motions are obtained (e.g., $10 \times 5 \times 50 = 2500$ scaled ground motions are selected for each SMF).
7. Utilize the selected scaled and unscaled ground motions to conduct NTHA of the selected structure (e.g., steel SMF) and obtain the EDPs of interest (e.g., PIDRs and PFAs). Hence a total of $n \times j \times m$ scaled + $n \times i$ unscaled EDPs are finally obtained for the selected structure.
8. Due to the limitation of recorded ground motions, the EDP grid constructed using $n \times m$ scaled and unscaled ground motions (unscaled represents $SF = 1$) are likely to be sparse and would not adequately cover all ranges of IMs and SFs . Hence utilize appropriate machine-learning/statistical methods to develop a comprehensive surrogate model which can be used to complete the dataset. The model needs to be trained using the IMs and SFs as inputs and should be able to provide proper statistical estimates of corresponding EDPs (e.g., Bayesian non-parametric GPR models are trained using the scaled and unscaled datasets).
9. For a particular IM level, use the surrogate model to obtain the estimate for the unscaled EDPs for $SF = 1$ and scaled EDP for the m different SFs ; the estimates can be point estimates of a statistical moment or complete statistical distribution (e.g., complete posterior distributions of EDPs are obtained using the GPR models).
10. Use an appropriate hypothesis test to statistically compare the similarity of one estimate of unscaled EDP against m scaled EDP estimates corresponding to m SFs (e.g., KS test is used in this study). Finally obtain the SFs whose corresponding scaled EDP estimates are deemed statistically similar to the unscaled EDP estimate, thereby showing no bias.

This process should be repeated for other IM levels of interest to obtain the corresponding SFs leading to no significant bias in EDPs. The results can then be used to provide recommendations for the allowable SF ranges for different IM levels and EDPs. While the proposed framework is general and can be used for any type of structure and ground motions, in this study, the process is applied to steel SMFs and crustal ground motions, and the results are discussed in the subsequent sections of the paper. The sections: “Structural Models” provide details of the five SMFs utilized in this study; “Ground Motion Selection” discusses the selection and scaling process of ground motions (i.e., steps 1 to 6); “Seismic Demand Surrogate Model using GPR” entails the description of NTHA and development of GPR-based surrogate modelling (i.e., steps 7 and 8); “Interpretability of the developed GPR models” provides supplementary interpretability of the developed GPRs for engineering analysis; “Hypothesis Testing and Results” discusses the results of KS hypothesis test and provides recommendations for allowable scaling limits (i.e., steps 9 and 10); and finally “Conclusions” section provides the final analysis of the framework within the setting of steel SMF structures.

1.2 | Structural models

For this study, five (i.e., 2-, 4-, 8-, 12- and 20-story) different archetype SMFs are selected as the prototype structures for analysis. The five SMFs mathematical models vary only in their height since this parameter is a quite important characteristic of buildings, as pointed out in several parametric studies conducted in the past (e.g.,^{19–22}). Thus, the height of the SMFs varies from 8.50 m (2-story) to 79.85 m (20-story).

The SMFs are designed as per ASCE 7–16.¹ A dead load of 4.78 kN/m² and a live load of 2.38 kN/m² are applied to all floors. In addition, a perimeter load of 1.20 kN/m² is applied to simulate the cladding. Figure 1 illustrates the elevation and the plan view of the selected archetype frames used in this investigation. As per Figure 1, the SMFs share common characteristics, such as the plan view. The SMFs are located at the perimeter of the building plan and consist of three-bay frames with a bay width of 6 m. The first story’s height is 4.5 m, while the typical height of the rest of the stories is 3.90 m. The beam-column connections are detailed as reduced beam section (RBS) connections as per the American Institute of Steel Construction (AISC) AISC-341-16²³ and AISC-358-16.²⁴ Table 1 summarizes the first mode’s participation factors and natural periods of the SMFs. For further details about the archetype frames, the readers are referred to Kircher et al.²⁵

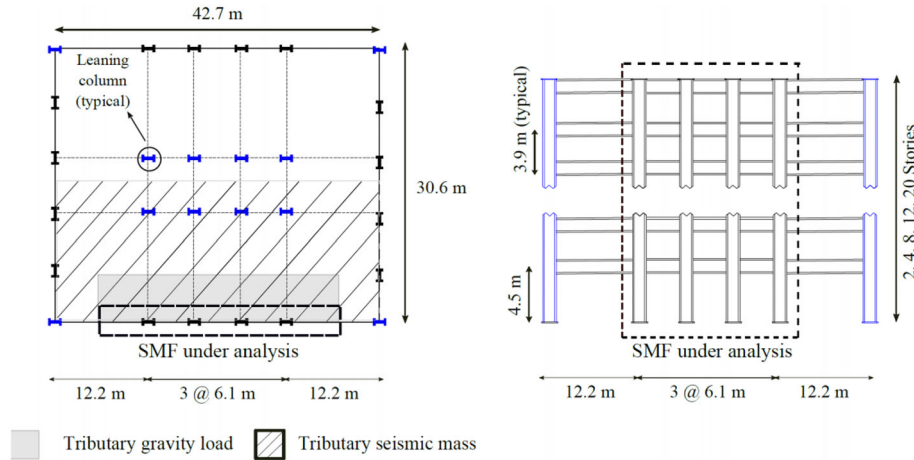


FIGURE 1 Archetype SMFs' outline

TABLE 1 SMFs' modal properties

	2-story	4-story	8-story	12-story	20-story
First period (s)	1.03	1.6	2.21	2.97	4.64
First mode mass participation	95.31%	83.80%	80.41%	78.28%	74.89%

The SMFs are modeled as plane frames (2D) using the software OpenSEES for all the NTHA since this platform has been extensively verified²⁶ for nonlinear analysis. Beam and column elements of the SMFs are modeled as linear elastic elements with rotational springs located at their ends. The nonlinear rotational springs aim to capture the hysteretic behavior of the beam-column connections and are represented by Ibarra-Medina-Krawinkler (IMK) bilinear model.²⁷ This phenomenological model consists of a trilinear backbone curve, rules to predefine the hysteretic behavior, and rules to capture cyclic strength and stiffness deterioration. These springs' properties are computed per the guidelines for nonlinear structural analysis for the design of buildings provided by the National Institute of Standards and Technology (NIST GCR 17-917-46v2).²⁸ However, these phenomenological springs cannot capture the moment-axial load interaction effect in the columns of the SMFs. Thus, this phenomenon is tackled in an approximate manner. First, based on the recommendations given by NIST GCR 17-917-46v2[28], the gravity loads are assumed as an average value of the axial load obtained during the NTHA of the frames. Next, the bending strength of the columns is reduced using the AISC 318-16²⁹ interaction equations for combined forces. Finally, this reduced strength is used in all column springs. This procedure is appropriate since SMFs are typically drift-controlled rather than force-controlled, entailing that the SMF columns are only lightly loaded. Moreover, experimental studies in the past (e.g.,³⁰) on shake tables have shown good agreement with mathematical models calibrated with the preceding approach. Because of these reasons, several studies on SMFs^{20,31,32} conducted in the past have adopted similar modeling decisions.

The panel zones are modeled as a hinge parallelogram assembly by rigid elements with a nonlinear spring in one of the corners to simulate shear distortion. This spring is defined by the yield strength, full plastic strength, and the corresponding stiffness parameters.³³ A leaning column simulates P -delta effects through large displacement geometric nonlinearity. This leaning column is loaded with gravity loads corresponding to half of the plan-building on each floor and is connected to the SMFs through rigid elements. Finally, column base connections are simulated with the mathematical models developed by Torres-Rodas et al.³⁴ and Torres-Rodas et al.³⁵ for exposed and embedded base connections, respectively. Consistent with current engineering practice,³⁶ the exposed base plate detail is used for the low-rise buildings (i.e., 2-, 4-story), while for the taller frames (i.e., 8-, 12- and 20-story), the embedded base connection detail is preferred.

1.3 | Ground motion selection

Proper ground motion record selection is essential to understand the joint impact of SF and $S_a(T_1)$ on the response of steel structures. The study uses a large subset from the NGAWest2 database¹⁵ consisting of ~ 7000 mainshock records

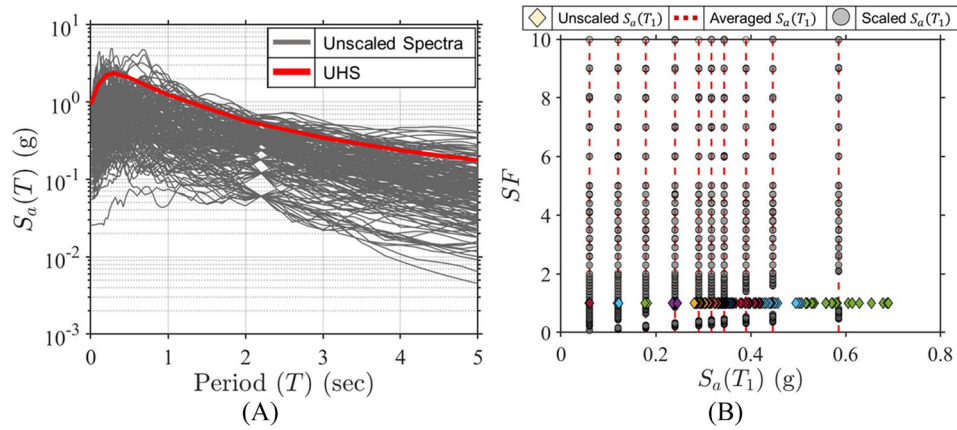


FIGURE 2 (A) Spectra of selected ground motions and MCE UHS of LADT; (B) selected $S_a(T_1)$ and SF grid for eight-story SMF

from ~ 277 earthquakes with moment magnitudes (M_w) ranging from $3 \leq M_w \leq 8$ and rupture distance (R_{rup}) ranging from $0 \leq R_{rup} \leq 100$ km. More details about the subset database can be found in Fayaz et al.³⁷

Since the steel structures are designed for the hazard levels of the Los Angeles Downtown (LADT)²⁵ site (latitude = 34.05204, longitude = -118.25713, $V_{s30} = 360$ m/s), the MCE level $S_a(T_1)$ is obtained from the uniform hazard spectrum (UHS) of the site using the unified hazard tool (UHT)³⁸ (corresponding to step 1). The $S_a(T_1)$ levels are then uniformly divided into 10 equal stripes from 0 g to the obtained MCE $S_a(T_1)$ level (corresponding to step 2). For each stripe of $S_a(T_1)$, 20 unscaled ground motion records that closely match (without any scaling) the respective $S_a(T_1)$ are selected from the ~ 7000 records database. Hence a total of 200 unscaled ground motions are selected for each building for this study (corresponding to step 3). The selected ground motion spectra are shown along with the UHS of LADT in Figure 2A for the eight-story building ($T_1 = 2.21$ s). The $S_a(T_1)$ of the selected unscaled records are also plotted against $SF = 1$ in Figure 2B for the eight-story building ($T_1 = 2.21$ s). The different colors of the diamond-shaped data points represent the ground motions selected for different $S_a(T_1)$ stripe. It can be observed from the data points that as the stripe $S_a(T_1)$ increases, the variance of $S_a(T_1)$ of the selected ground motions also increases. This is due to the lack of enough ground motions in the recorded database for large $S_a(T_1)$ levels. Though the $S_a(T_1)$ distribution and accuracy of the selected 20 ground motions are not the same for higher $S_a(T_1)$ levels as compared to the lower $S_a(T_1)$, the differences in the unscaled $S_a(T_1)$ are not expected to lead to significant differences in the final analyses, as will be described in the following sections of selecting scaled ground motions and developing/utilizing GPR-based surrogate models.

For each group of the selected 20 unscaled ground motions, a corresponding set of 250 scaled ground motions is obtained. As described above, the selected unscaled ground motions do not necessarily have the same $S_a(T_1)$ as the target $S_a(T_1)$ of the different stripes. Hence, to select a statistically compatible set of scaled ground motions, $S_a(T_1)$ of the selected 20 unscaled ground motions for each stripe is averaged (shown in dashed red lines in Figure 2B) (corresponding to step 4). Then the averaged $S_a(T_1)$ is used as a target $S_a(T_1)$ to select five scaled ground motions for 50 different SF s. The 50 SF s are determined by dividing SF s from 0.1 (divided by 10) to 1 in 25 intervals and from 1 to 10 (multiplied by 10) in 25 intervals (corresponding to step 5). Then for each SF , five ground motions are selected for the respective $S_a(T_1)$ from a subgroup of the ground motions database with $M_{w,min} - 0.5 \leq M_w \leq M_{w,max} + 0.5$, where $M_{w,min}$ and $M_{w,max}$ denote the minimum and maximum M_w of the corresponding 20 unscaled records, respectively (corresponding to step 6). This is done to reduce any additional sources of variability between the scaled and unscaled ground motions that may occur due to differences in causal parameters. The selection is made by first obtaining the required SF s for all the ground motions in the subgroup to achieve the target averaged $S_a(T_1)$ and then selecting the five closest to each of the 50 SF levels. Hence, a total of 2500 scaled ground motions are set (5 ground motions \times 50 SF s \times 10 $S_a(T_1)$ stripes) for each SMF (corresponding to step 6). The selected scaled ground motions are shown in grey data points in Figure 2B. It should be noted here that in the complete record selection process (including scaled and unscaled), once a ground motion is selected for any level of scaled or unscaled $S_a(T_1)$ and SF , it is removed from the database to avoid any repetitive selection of the same ground motion. This selection process is repeated for each SMF (five in total) analyzed herein. In this manner, a total of 13,500 ground motions (2500 scaled + 200 unscaled = 2700 ground motions \times 5 SMFs) are selected for NTHA of the steel structures.

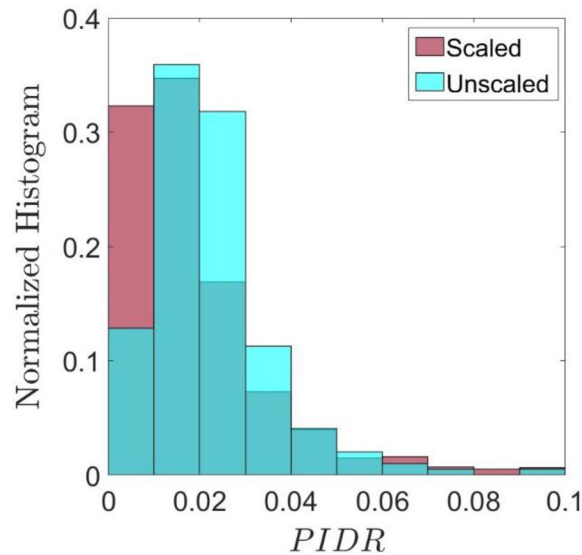


FIGURE 3 Histogram of $PIDR$ for scaled (2500) and unscaled (200) records for eight-story SMF

1.4 | Seismic demand surrogate model using GPR

The selected 2700 ground motions are used to conduct NTHA of the five SMFs to obtain two peak responses: i) peak inter-story drift ratio ($PIDR$) and ii) peak floor acceleration (PFA) (corresponding to step 7). Figure 3 illustrates a normalized histogram of the $PIDR$ distributions for the 8-story frame. As per Figure 3, most of the $PIDR$ s ($\sim 80\%$) are below 0.025 (i.e., are code compliant), implying modest nonlinear behavior. Similar trends are observed in the other SMFs. The aim is to compare the 2500 scaled responses for the selected 50 scaling factors against the 200 unscaled responses. As described in the previous section, due to the lack of unscaled records for large $S_a(T_1)$ and scaled records for the combination of large $S_a(T_1)$ and $SF < 1$, the obtained EDPs data are not viable for the direct comparison. This can be thought of as a “missing data” or “kriging” problem. Hence, in this study, Bayesian non-parametric and machine-learning-based GPR³⁹ is deployed to develop surrogate models, which are used to complete the data for the entire selected $S_a(T_1)$ and SF grid (corresponding to step 8).

The basis of GPRs relies on the Gaussian process (GP), which is defined as a collection of random variables, any finite number of which have a joint (multivariate) Gaussian distribution.⁴⁰ GPR is a powerful and effective tool for nonlinear regression problems since it has a simple structure and, unlike most regression methods, provides non-uniform uncertainty estimations based on the proximity of features.⁴¹ GPR serves as an efficient tool for performing inference both passively (i.e., interpolation, such as completing missing data, or extrapolation, such as forecasting or prediction) as well as actively (i.e., filtering and smoothing).³⁹ It can accurately capture a wide variety of relations between the features and targets by utilizing a theoretically infinite number of parameters and allowing the data to determine the level of complexity through Bayesian inference.⁴² Furthermore, it is observed that many Bayesian regression models based on artificial neural networks converge to the Gaussian process with an infinite number of hidden units.⁴³ Due to the efficacy of GPRs, it has been widely used as surrogate models for various modeling purposes and applications (e.g.,^{44–49}).

A GPR model attempts to describe the data in terms of a signal term ($f(x)$) and noise term (ϵ), as explained in Equation 1. Similar to other regression models, the noise term ϵ reflects the inherent aleatory randomness in the data and is assumed to be normally distributed with zero mean and variance (σ^2). The signal term ($f(x)$) is also assumed to be a random variable following a particular statistical distribution, thereby reflecting the uncertainty of the function. In GPRs, this distribution is assumed to be a Gaussian process with a mean ($m(\mathbf{x})$) and covariance ($k(\mathbf{x}, \mathbf{x}')$) function, as mentioned in Equation 2. The mean function ($m(\mathbf{x})$) describes the expected value ($E[.]$) of the function $f(x)$ at input \mathbf{x} as shown in Equation 3. The covariance function ($k(\mathbf{x}, \mathbf{x}')$) explains the dependence between the values of the function at inputs \mathbf{x} and \mathbf{x}' . The covariance function $k(\mathbf{x}, \mathbf{x}')$ involves the utilization of a kernel that models the correlation between two points based on the distance between the points. This means that closer points are expected to behave more similarly than points that are further away from each other. Many available kernels are used to develop the covariance function, such as radial-basis function (RBF), white, matérn, rational quadratic, pairwise, etc.³⁹

Based on the trade-off between explainability, generalization (prevent overfitting), and accuracy, this study uses the summation of matérn (Equation 5) and white (Equation 6) kernels as described in Equation 7. The class of matérn kernels is a generalization of RBF kernels. It is parameterized by two factors, including i) length scale (l), which is greater than zero, and ii) smoothness parameter (ν). The length scale parameter (l) can either be a scalar (an isotropic variant of the kernel) or a vector with the same number of dimensions as the inputs \mathbf{x} (an anisotropic variant of the kernel). The smaller ν , the less smooth the approximated function is. As $\nu \rightarrow \infty$, the kernel becomes equivalent to the RBF kernel. When $\nu = 1/2$, the matérn kernel becomes identical to the absolute exponential kernel. Other important intermediate values include $\nu = 1.5$ (once differentiable functions) and $\nu = 2.5$ (twice differentiable functions). Furthermore, $d(\cdot)$, $K_\nu(\cdot)$, and $\Gamma(\cdot)$ represent the Euclidean distance, modified Bessel function, and gamma function, respectively. The white kernel is mainly used as a sum kernel and explains the noise (ϵ) of the signal as an independent and identical normal distribution with variance ϕ^2 . Lastly, η represents the scaling factor of the used kernel function $k(\mathbf{x}, \mathbf{x}')$ (Equation 7). Due to the lognormal nature of seismic demands and IMs, the GPR models are trained in the lognormal domain.^{50,51}

$$y = f(x) + \epsilon (\sim N(0, \sigma^2)) \quad (1)$$

$$f(x) \sim GP(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \quad (2)$$

$$m(\mathbf{x}) = E[f(\mathbf{x})] \quad (3)$$

$$k(\mathbf{x}, \mathbf{x}') = E[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))] \quad (4)$$

$$k(\mathbf{x}, \mathbf{x}')_{mat} = \frac{1}{\Gamma(\nu) 2^{\nu-1}} \left(\frac{\sqrt{2\nu}}{l} d(\mathbf{x}, \mathbf{x}') \right)^\nu K_\nu \left(\frac{\sqrt{2\nu}}{l} d(\mathbf{x}, \mathbf{x}') \right) \quad (5)$$

$$k(\mathbf{x}, \mathbf{x}')_{white} = \epsilon (\sim N(0, \phi^2)); \text{ if } \mathbf{x} = \mathbf{x}' \text{ else } 0 \quad (6)$$

$$k(\mathbf{x}, \mathbf{x}') = \eta^2 \times (k(\mathbf{x}, \mathbf{x}')_{mat} + k(\mathbf{x}, \mathbf{x}')_{white}) \quad (7)$$

For each SMF, two types of GPRs are trained as surrogate models using the resulting EDPs of i) 200 unscaled ground motions with $S_a(T_1)$ as the input features, and ii) 2500 scaled ground motions with $S_a(T_1)$ and SF s as the input features (corresponding to step 8). The two EDPs (i.e., PFA and PIDR) are independently used as the target variables for the GPRs. Hence a total of four GPRs are developed for each SMF. Since each combination of $S_a(T_1)$ and SF , five scaled ground motions are selected and used for NTHA, and the GPRs are developed by obtaining the mean EDPs (in the lognormal domain) and SF s. Hence for each five scaled ground motions obtained at the intersection of $S_a(T_1)$ and SF , the respective mean EDPs are obtained to represent the scaled response. This is presented in Figure 4A for eight-story SMF, where the yellow data points show the 200 unscaled records against $SF = 1$, grey points show the 2500 scaled records, and red markers show the 500 mean values (i.e., one mean value for five scaled records) for the scaled records. This is done for both PFAs and PIDRs, and the obtained 500 mean values are utilized to develop GPR surrogate models for scaled responses, while the 200 unscaled values are used to develop GPR surrogate models for unscaled responses (corresponding to step 8).

The posterior mean prediction surface and curve using the scaled and unscaled GPR models, respectively, are shown in Figure 4B (other views are given in Figures 4C and D) for the eight-story building. As can be seen from the figure, the GPR models demonstrate a good prediction power for the PIDRs with a coefficient of determination (R^2) of 0.75 and 0.85 for unscaled and scaled responses, respectively. Similarly, R^2 of 0.74 and 0.9 are observed for unscaled and scaled PFAs, respectively. Similar results are observed for other SMFs and EDPs. Hence based on the good prediction power of the developed GPRs for the dataset, they can be used as surrogate models for the data completion. It should be noted that the

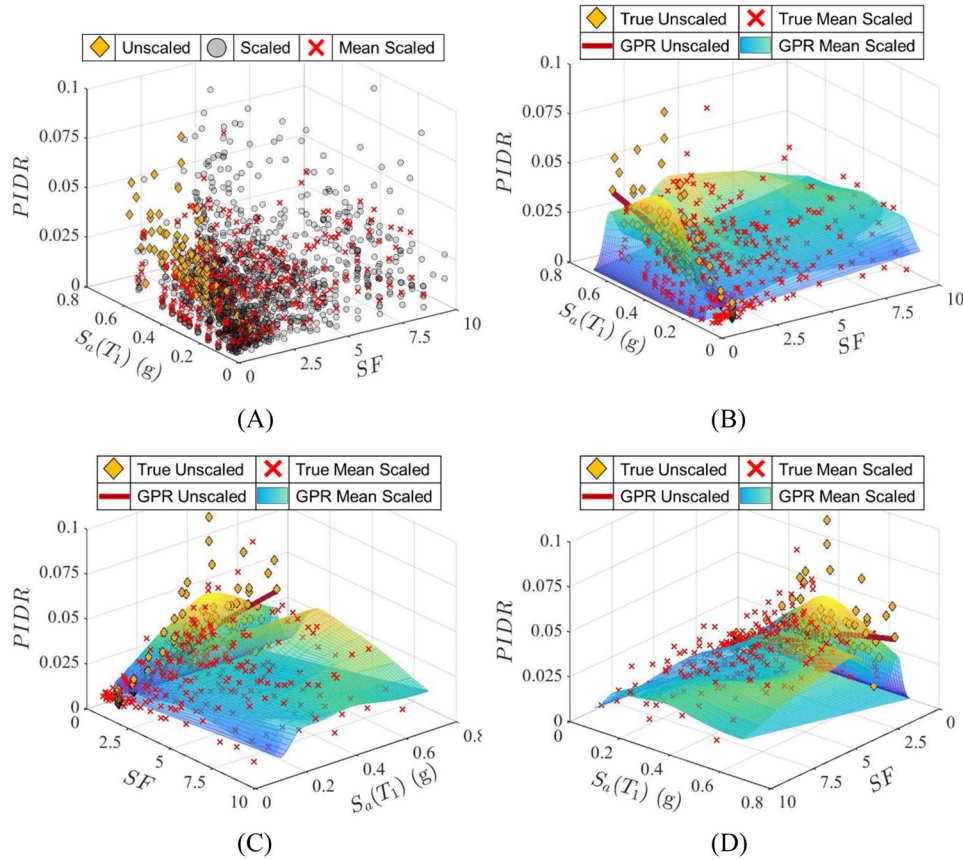


FIGURE 4 (A) True computed, (B) GPR predictions (view 1), (C) GPR predictions (view 2), (D) GPR predictions (view 3), for unscaled and mean scaled PIDR for eight-story SMF

developed GPRs are not proposed as general surrogate models but are only used as kriging models to complete the EDPs dataset in this study. It is important to point out that only non-collapse cases of analysis have been considered in this study. Thus, a PIDR limit of 0.10 has been set as a threshold between collapse or non-collapse, following notable publications on the topic (e.g.,⁵²).

Several previous research studies (such^{16,53–56}) have described the importance of using IMs that explain spectral accelerations at a range of periods rather than $S_a(T_1)$. This is mainly to include the period elongation (due to non-linearity) that can occur during the time history of high-intensity ground motions and to include higher mode effects. However, such conclusions are usually observed in the cases of structures that are expected to undergo huge non-linearity during the ground motion and have a higher number of participating modes. In this study, the modal mass contributions of the first mode for the SMFs are 95.31% for two-story, 83.80 for four-story, 80.41% for eight-story, and 78.28% for 12-story, and 74.89% for 20-story (as mentioned in Table 1). Thus, higher modes' contributions are not deemed significant to the total response of the used SMFs. Furthermore, the recorded EDPs are observed to be low (code compliance), and consequently, the level of inelasticity/non-linearity occurring during the ground motion is also small. Hence, $S_a(T_1)$ is postulated to be a sufficient IM in the study. This hypothesis is tested by comparing the relative feature importance of $S_a(T_1)$ against the average spectral acceleration ($S_{a,avg}$) and $S_{a,ratio}$ using random forests algorithm. $S_{a,avg}$ is defined as the geometric mean of spectral acceleration values between periods $a.T_1$ and $b.T_1$ where a and b are non-negative constants ($a \leq b$).⁵⁷ In this study $a = 0.2$ and $b = 3$ are used. On the other hand, $S_{a,ratio}$ is a metric used to quantify spectral shape, and defined as $S_a(T_1)$ normalized by $S_{a,avg}$ over a period range.⁵⁵ This data-driven method checks the relative importance of the candidate features for explaining a target variable.⁵³ To avoid any bias or collinearity issues, in this case, the algorithm is employed in two settings; $S_a(T_1)$ is compared against $S_{a,avg}$ and $S_{a,ratio}$ for the target PIDR for the 2700 (scaled and unscaled) responses for all five SMFs. It can be observed from Figures 5A and B for the dataset of the eight-story SMF, $S_a(T_1)$ is observed to be more significant in describing PIDR as compared to $S_{a,avg}$ and $S_{a,ratio}$, thereby validating the hypothesis. Also, to make the conclusions simpler and interpretable, it is essential to minimize the number of features

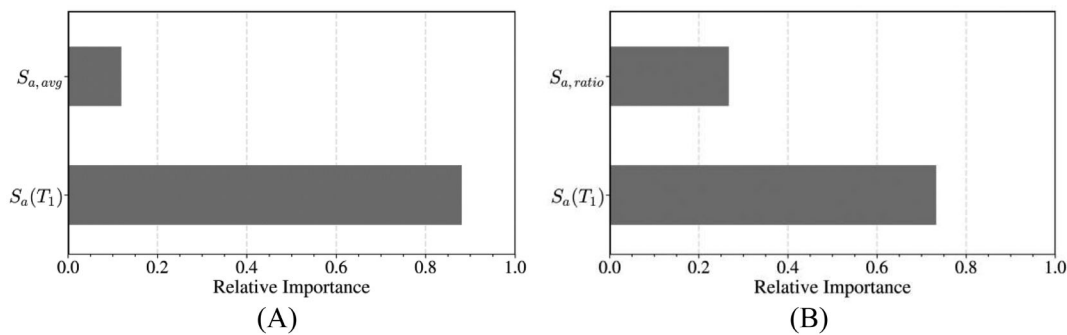


FIGURE 5 Relative feature importance of $S_a(T_1)$ against: (A) $S_{a,avg}$ and (B) $S_{a,ratio}$, for PIDR of eight-story SMF

and use scalar quantities of IMs as the inputs to the GPRs. Hence it can be concluded that for the used dataset, $S_a(T_1)$ is sufficient to explain the demands of the used SMFs. The readers can conduct similar feature analysis for their datasets and utilize the appropriate IMs (corresponding to step 1).

1.5 | Interpretability of the developed GPR models

Due to the versatility of the machine-learning models, they have been widely used in engineering applications. However, due to the “black-box” nature of these models, there is a general reluctance in the research community to propose and utilize such models. Hence it is critical to provide sufficient analytics for model interpretability and its response in terms of predictions based on variability in the input features. With the onset of XAI, various algorithms have been developed that provide different methods that allow interpretability of these “black-box” models and step towards “grey-box” and even “white-box” nature (such as linear regression, decision trees, etc.).

This study uses SHAP to analyze and interpret the nature of the developed GPRs.⁵⁸ SHAP is a model-agnostic procedure that provides insights to explain individual predictions of the model based on the game’s theoretically optimal Shapley values. Shapley values are a widely used approach from cooperative game theory with desirable properties.⁵⁹ The feature values of a data instance act as players in a coalition. The Shapley value is the average marginal contribution of a feature value across all possible coalitions. Shapley values are the only solution that satisfies properties of efficiency, symmetry, dummy, and additivity.⁵⁸

SHAP belongs to the class of models called “additive feature attribution methods,” where the explanation is expressed as a linear function of features, as described by Equation 8. In this equation, $g(z)$ represents a local surrogate model of the original model $f(x)$ (in this case $f(x)$ is the GPR model), M is the number of input features of the model $f(x)$ (i.e. $S_a(T_1)$ and SF), θ_0 denotes the bias term, which represents the base value of the predictions made by the model $f(x)$, θ_i is the contribution of the i^{th} feature towards the final output (i.e., SHAP value), and z_i is a binary variable that takes a value of 1 for the feature corresponding to θ_i contribution and 0 otherwise. θ_i evaluates the difference to the final predictions made by the model $f(x)$ by including the i^{th} feature for all combinations of features other than i . This is expressed in Equation 9, where S represents a subset of features among all N features except the i^{th} feature (denoted as $S \subseteq N \setminus \{i\}$), $[f_x(\{i\}) - f_x(S)]$ is the difference in the outputs made by the i^{th} feature, and $\frac{|S|!(M-|S|-1)!}{M!}$ is the weighing factor counting the number of permutations of the subset S . $f_x(S)$ in the different part of the equation represents the expected output given the subset of features S , which is similar to the marginal average on all other features other than the subset S . Hence, in a nutshell, SHAP values explain the contribution of the features to the respective outputs in a quantitative manner, thereby allowing interpretation. These are analogically similar to the coefficients of a regression model, which provide the impact of the corresponding feature on the target variable. Due to the computational complexity of computing Equation 10, SHAP values are approximated using various types of explainers such as kernel-explainer, tree-explainer, deep-explainer, etc.⁶⁰ In this study, kernel-explainer uses a special weighted linear regression to compute the importance of each feature. The computed importance values are Shapley values from game theory and coefficients from a local linear regression.⁶⁰

$$g(z) = \theta_0 + \sum_{i=1}^M \theta_i z_i \quad (8)$$

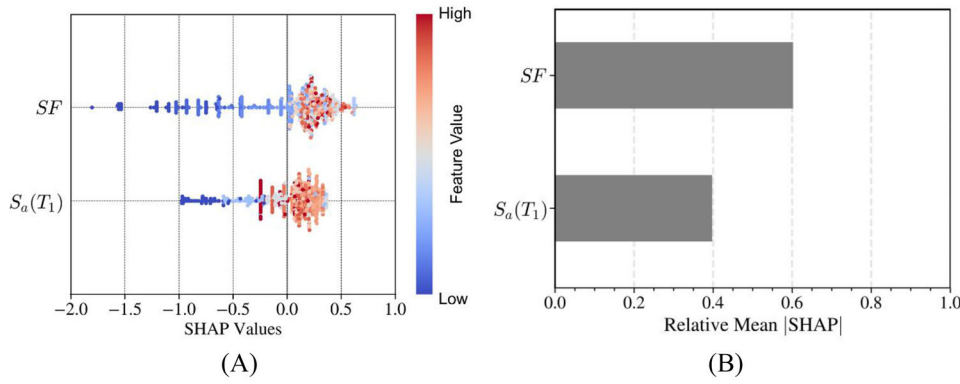


FIGURE 6 (A) SHAP values and (B) relative mean absolute SHAP values for SF and $S_a(T_1)$, for PIDR of eight-story SMF

$$\theta_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(M - |S| - 1)!}{M!} [f_x(S \cup \{i\}) - f_x(S)] \quad (9)$$

$$f_x(S) = E[f(x) | x_S] \quad (10)$$

The 50 SF s and 10 $S_a(T_1)$ are used as the inputs (i.e., $50 \times 10 = 500$ samples) to the GPR surrogate model for scaled ground motions and the respective SHAP values for SF (i.e., $SHAP_{SF}$) and $S_a(T_1)$ (i.e., $SHAP_{S_a(T_1)}$) are computed. The SHAP values for the 500 samples for PIDR of eight-story SMF are presented in Figure 6a, where the color of the data points represents the magnitude of the feature values, while low refers to values close to 0 for both SF s and $S_a(T_1)$, high refers to values close to 10 and 0.6 for SF s and $S_a(T_1)$, respectively. As described above, the sign of SHAP indicates the direction of contribution and the absolute value of SHAP indicates the magnitude of contribution. Hence negative SHAP values indicate that the corresponding feature lowers the final target prediction while positive SHAP indicates an increase in the final output of the model. It can be observed that, in general, with an increase in the value of both features, their corresponding SHAP values tend to move from a negative sense to a positive sense. The trend is more evident for SF s, which means for central values of the feature (i.e., $SF = 1$), the contribution tends to be close to zero. Hence this means a higher the magnitude of scaling factor (division or multiplication) and $S_a(T_1)$ leads to a higher magnitude of contribution to the output PIDR. Furthermore, it can be observed from the Figure 6A that extreme values of SF lead to a higher magnitude of SHAP and hence lead to more significant changes in the output PIDR as compared to $S_a(T_1)$. Also, Figure 6B presents the relative feature importance of SF and $S_a(T_1)$ on the GPR output, in terms of their mean absolute SHAP values ($|SHAP|$). This is done by computing the mean $|SHAP|$ values for the 500 samples for both SF and $S_a(T_1)$ and then dividing them by the sum of the two mean $|SHAP|$. Since the SHAP values represent the contribution of the features to the model output, computing their relative sum for the two features in an absolute sense signifies the importance of the respective feature in predicting PIDR. It can be observed from the figure that SF tends to have a higher contribution in the GPR model to predict the PIDR as compared to $S_a(T_1)$. The difference in the contribution is close to 60%:40% in favor of SF against $S_a(T_1)$. This means the scaling processes require more careful examination in terms of both SF and $S_a(T_1)$.

Figure 7 presents the dependence contribution plots for $SHAP_{SF}$ and $SHAP_{S_a(T_1)}$ against SF and $S_a(T_1)$ and the colors provide the dimension of other features. Dependence plots allow better understanding and insights into the features' interactions and the respective SHAP values. Figure 7A shows the 500 SF s plotted against the corresponding $SHAP_{SF}$ while Figure 7B shows the 500 $S_a(T_1)$ plotted against the corresponding $SHAP_{S_a(T_1)}$. The plots also provide box plots for the respective SHAP values for showing the statistics. The box plots show that the range of $SHAP_{SF}$ tends to be higher than $SHAP_{S_a(T_1)}$, which means that sensitivity of PIDR is higher towards the change in SF as compared to $S_a(T_1)$. Furthermore, it can be observed from Figure 7A, $SHAP_{SF}$ tends to be close to 0 for $SF \sim 1$ showing no contribution to the GPR model predictions. This is in line with the expectations since, for $SF \sim 1$, $S_a(T_1)$ is the only parameter that impacts the PIDR. It is further observed that as the value of SF decreases less than 1, the magnitude of $SHAP_{SF}$ increases linearly in a negative sense with significantly less variability due to $S_a(T_1)$. On the other hand, an increase in SF greater than 1 leads to a plateau effect in terms of $SHAP_{SF}$, however, the variability in the SHAP values increases significantly due to $S_a(T_1)$. This means the contribution of $SF < 1$ (i.e., divisive scaling factors) to the PIDR doesn't vary significantly with $S_a(T_1)$, while

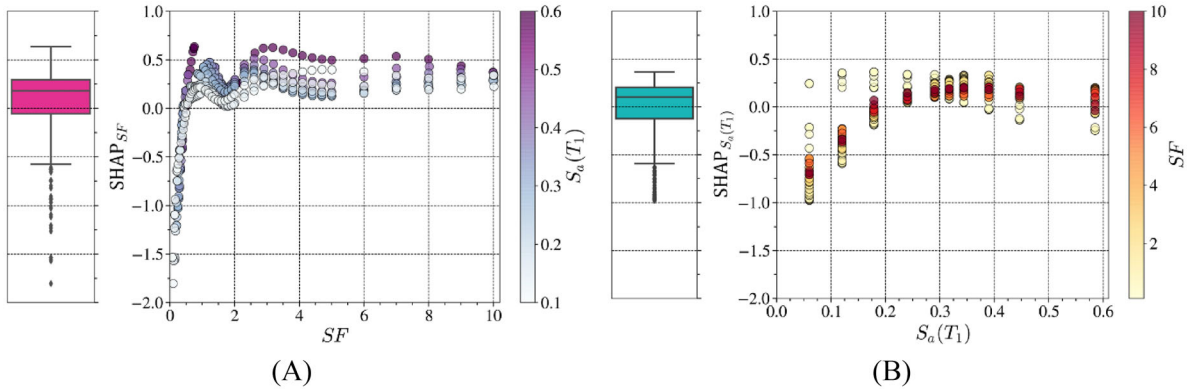


FIGURE 7 Dependence contribution plots: (A) $SHAP_{SF}$ vs. SF , and (B) $SHAP_{S_a(T_1)}$ vs. $S_a(T_1)$, for PIDR of eight-story SMF

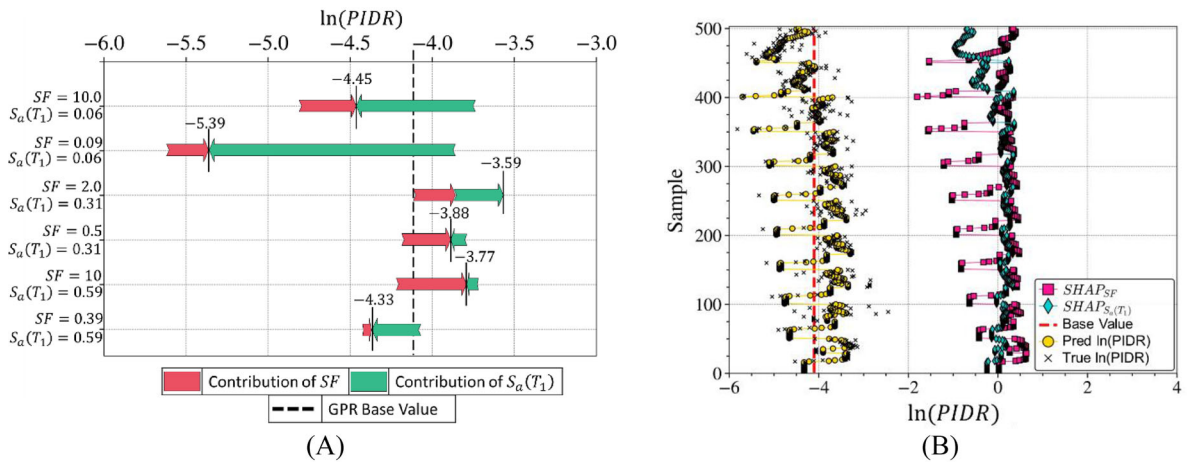


FIGURE 8 SHAP: (A) force plot and (B) decision plot for PIDR of eight-story SMF

for $SF > 1$ (i.e., multiplicative scaling factors), the contribution to the $PIDR$ varies considerably with $S_a(T_1)$. Similarly, in Figure 7B, it can be observed that with an increase in $S_a(T_1)$, $SHAP_{S_a(T_1)}$ increases considerably, and interaction with SF leads to significant variability in the SHAP values. This means the contribution of $S_a(T_1)$ to predict $PIDR$ is significantly affected by the level of scaling involved. Hence it is essential to develop regulations for amplitude scaling in conjunction with the IMs (i.e., $S_a(T_1)$ in this case).

Lastly, to interpret the decision-making process of the GPR model, Figure 8A presents force plots for six grid samples among the 500 samples ($50 SF \times 10 S_a(T_1)$). The selected six samples represent the four corners (maximums and minimums of SF and $S_a(T_1)$) and two central (factor of 2 for SF and median $S_a(T_1)$) examples of the selected SF and $S_a(T_1)$ grid. The dashed line in the figure shows the base value obtained from the GPR model, which represents the average prediction for $PIDR$ and hence is the same across all predictions. The overall prediction from the GPR model is computed as described in Equation 11. The two-colored arrows in the figure show how the SF and $S_a(T_1)$ influence the GPR base value (i.e., SHAP value) and arrive at the final prediction of $PIDR$ based on the input values. It can be observed in general that SF tends to have a decreasing impact on the model outcome while $S_a(T_1)$ tends to raise the values, thereby having an increasing impact. Specifically, it is observed that for smaller values of $S_a(T_1)$ the model outputs are influenced by $S_a(T_1)$ to a more considerable extent as compared to SF , which is consistent with Figure 7b. Figure 8B presents the final decision plots for the 500 samples and shows the contributions of SF and $S_a(T_1)$ on the GPR base value and thereby how the model arrives at the final prediction as per Equation 11. It can be observed that the GPR predictions marked with cross marker are in close vicinity to the true $PIDR$ values thereby showing good surrogacy power of the GPR model. The values of $SHAP_{SF}$ are observed to mainly cause a decrease in the GPR base values and $SHAP_{S_a(T_1)}$ tend to increase the base values. While the results are presented here only for the eight-story SMF, similar analyses are conducted for the other four SMFs and PFA . These results are expected to allow the readers to interpret the GPR surrogate model and gain insights into how

the GPR model makes the final decision and lead to predictions of the EDPs .

$$\text{Pred ln}(PIDR) = \text{GPR Base Value} + \text{SHAP}_{SF} + \text{SHAP}_{S_a(T_1)} \quad (11)$$

1.6 | Hypothesis testing and results

The GPR-based surrogate models are utilized to provide recommendations for careful scaling of ground motions to conduct NTHA of the steel SMFs. This is done by using the scaled GPR model to sample the posterior distributions of EDPs for 50 SF s (25 divisive and multiplicative) \times 50 $S_a(T_1)$ (0 g to MCE in 50 equal intervals) grid; and using the unscaled GPR model to sample the posterior distribution of EDPs for the 50 $S_a(T_1)$ (unscaled EDPs correspond to $SF = 1$). Hence, after the sampling, there are 50 distributions for the scaled EDPs (corresponding to 50 SF s) and one distribution for unscaled EDPs (corresponding to $SF = 1$) against each of 50 $S_a(T_1)$ (corresponding to step 9). The obtained 50 scaled EDP distributions are compared against the one unscaled EDP distribution for each $S_a(T_1)$ to identify the SF s that lead to significant statistical bias in EDPs. Herein, it should be noted that “scaled EDPs” do not mean that the EDPs are scaled themselves, but rather refers to the EDPs obtained for the scaled ground motions (surrogated by the GPR model). The comparison between the scaled and unscaled EDP distributions is conducted using Kolmogorov–Smirnov (KS) hypothesis test for each $S_a(T_1)$ (corresponding to step 10). The null hypothesis proposes that the two distributions are statistically similar, which thereby suggests that the scaled EDPs for the corresponding SF are identical to the unscaled EDPs for the given $S_a(T_1)$. Hence the p -values greater than a significance level (typically 0.05) indicate no significant bias in the EDPs due to scaling of ground motions, while p -values lower than the significance level mean that the corresponding SF introduces a bias in the EDPs.

Figures 9 and 10 summarize the findings by illustrating the results from the KS test using contour plots of the obtained p -values for the grid. In these figures, the horizontal axis represents the scaling factor while the vertical axis is the IM (defined here $S_a(T_1)$). The color bar represents the p -value from the KS test. Thus, the green-colored regions in the figures illustrate the grid cases with p -value \sim 0, thereby rejecting the null hypothesis of statistical similarity between the scaled and unscaled EDPs and indicating bias due to the scaling of ground motions. Similarly, other regions represent grid cases where the test fails to reject the null hypothesis suggesting no significant bias with different probability levels (i.e., p -values). While the consensus in statistical research is to use p -value = 0.05 for hypothesis testing and balance type I and type II errors, in this study, the contours are presented for different p -values to allow stakeholders to choose their appropriate levels of required confidence.

The hypothesis test is conducted for both EDPs: PIDR and PFA (for all the archetype frames). Figures 9 and 10 present the results for PIDR and PFA, respectively, for all five SMFs. In both figures, it is observed that large scaling (both multiplicative and divisive) leads to p -values \sim 0, indicating bias in the scaled PIDRs and PFAs. Also, both figures show that for all SMFs, the bias is dependent on both $S_a(T_1)$ and SF ; and generally allows a higher range of SF for lower levels of $S_a(T_1)$ and leads to a more restricted range of SF for higher levels of $S_a(T_1)$. Furthermore, it is noticed that the results are not symmetric for $SF < 1$ and $SF > 1$ for both PIDR and PFA. The allowable SF s tends to decrease faster with increasing $S_a(T_1)$ for $SF > 1$, and the slope is observed to be lower for $SF < 1$.

In Figure 9, it is observed that for $S_a(T_1)$ values close to 0.1 g, the allowable range of scaling factors is: between 0.4 and 3.5 for p -value \sim 0.025, between 0.5 and 3 for p -value \sim 0.05, and between 0.6 and 2 for p -value \sim 0.1. In contrast, for higher levels of intensity (e.g., $S_a(T_1) = 0.7$ g for eight-story SMF) the allowable SF range narrows down to: between 0.7 and 2 for p -value \sim 0.025, between 0.75 and 1.75 for p -value \sim 0.05 and between 0.85 and 1.15 for p -value \sim 0.1. This phenomenon can be due to the mismatch in time-histories created by the SF for low amplitude ground motions compared to the true high amplitude unscaled ground motions. For example, it can be postulated that the use of $SF = 2$ for ground motions with $PGA = 0.1$ g and $PGA = 1$ g may not lead to the same changes in their other respective ground motions parameters. Hence the impact of scaling should be considered along with the level of ground motion intensity. Furthermore, a closer inspection of Figure 9 indicates that the bias in EDPs is also affected by the height of the SMFs. For instance, for the 12-story frame, at $S_a(T_1) = 0.5$ g the range of no-bias scaling factors is between 0.70 and 2.00, while for the two-story, at the same intensity level, the range spans from 0.40 to 3.00. This trend is observed across the five frames. This phenomenon may be explained due to the differences in the plastic strain distribution along with the height of the buildings. For instance, taller buildings (e.g., 12- and 20- story) generally lead to concentrated plasticity in a few elements in the lower stories. In contrast, in short buildings (e.g., two- and four-story), the plasticity is more evenly distributed along with the height. Thus, differences in the first period of the structure (associated with building height) entail an influence on the range of allowable SF s. This analysis is consistent with the mass modal participation factors mentioned

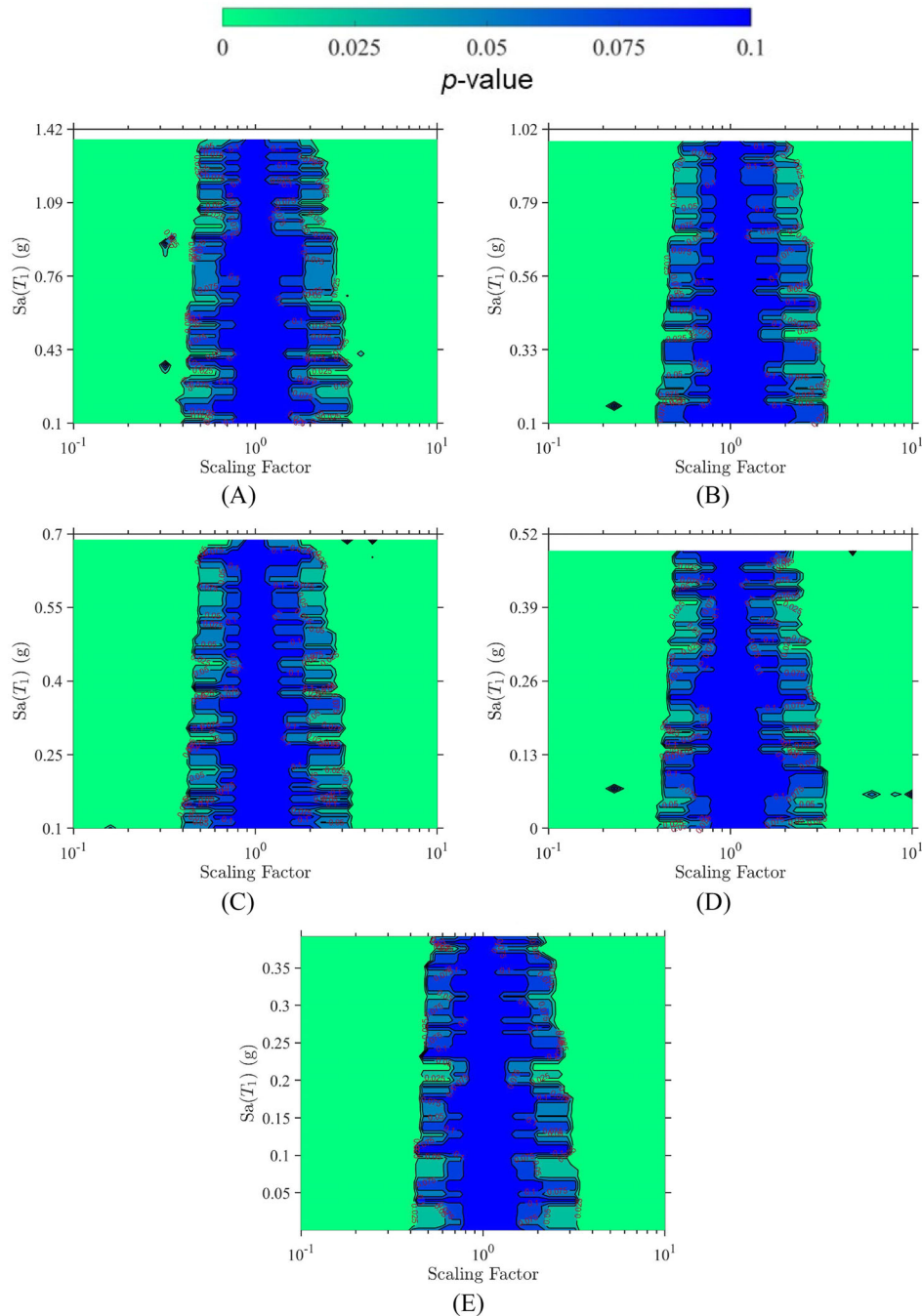


FIGURE 9 p -value contours for PIDR for: (A) 2-story; (B) 4-story; (C) 8-story; (D) 12-story; (E) 20-story

in the earlier sections, where the first mode leads to the participation of 95.31% for the two-story SMF and reduces to 74.89% for the 20-story SMF. Furthermore, as mentioned above, the slope of allowable SF s with respect to $S_a(T_1)$ tends to be steeper for $SF > 1$ as compared to $SF < 1$. This means the contribution of $SF < 1$ (i.e., divisive scaling factors) to the $PIDR$ doesn't vary significantly with $S_a(T_1)$, while for $SF > 1$ (i.e., multiplicative scaling factors), the SF contribution to the $PIDR$ varies considerably with $S_a(T_1)$. This observation is consistent with the results obtained from SHAP analysis.

PFAs play an important role in detailing nonstructural components of buildings.^{61,62} Hence it is also important to consider the impact of ground motion scaling on the structural PFAs. Figure 10 presents the KS test results for the PFAs of the five SMFs. In general, it can be observed that the range of allowable scaling factors is narrower for PFAs as compared to PIDRs. Similar to the cases in Figure 9, the bias due to record scaling on PFAs depends on the intensity level. It can be observed that for $S_a(T_1)$ values close to 0.1 g, the allowable range of scaling factors is between 0.45 and 2.5 for

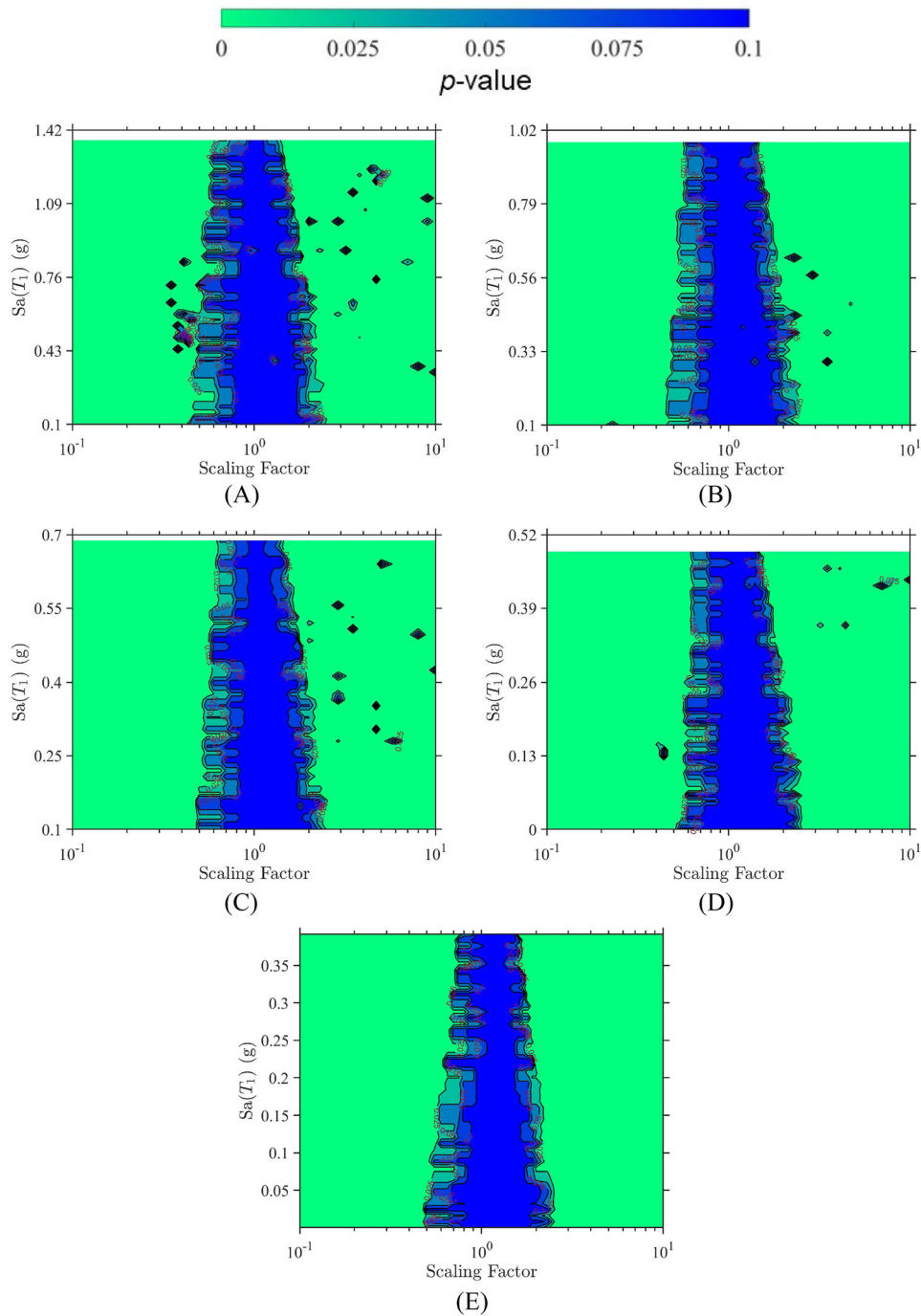


FIGURE 10 p -value contours for PFA for: (A) 2-story; (B) 4-story; (C) 8-story; (D) 12-story; (E) 20-story

p -value ~ 0.025 , 0.6, and 2.1 for p -value ~ 0.05 , and 0.7 and 2 for p -value ~ 0.1 . In contrast, for higher levels of intensity (e.g., $S_a(T_1) = 0.7$ g for eight-story SMF) the range narrows down to: between 0.65 and 1.8 for p -value ~ 0.025 , between 0.8 and 1.5 for p -value ~ 0.05 and between 0.9 and 1.1 for p -value ~ 0.1 . The results indicate that the scaling bias in PFAs has lower variability for the same intensity level than in the case of PIDRs. However, the slope of allowable SF s concerning $S_a(T_1)$ is observed to be more prominent for $SF > 1$ in the case of PFAs as compared to PIDRs. Lastly, in contrast with Figure 9, the scaling bias is less sensitive to the building height, implying lower effects of $S_a(T_1)$ on the ground motion scaling in PFAs. These differences in the results between PFAs and PIDRs can be due to the tendency of PFAs to decrease (as reported by⁶³) nonlinearly with levels of plasticity to a higher degree in contrast to the PIDRs.

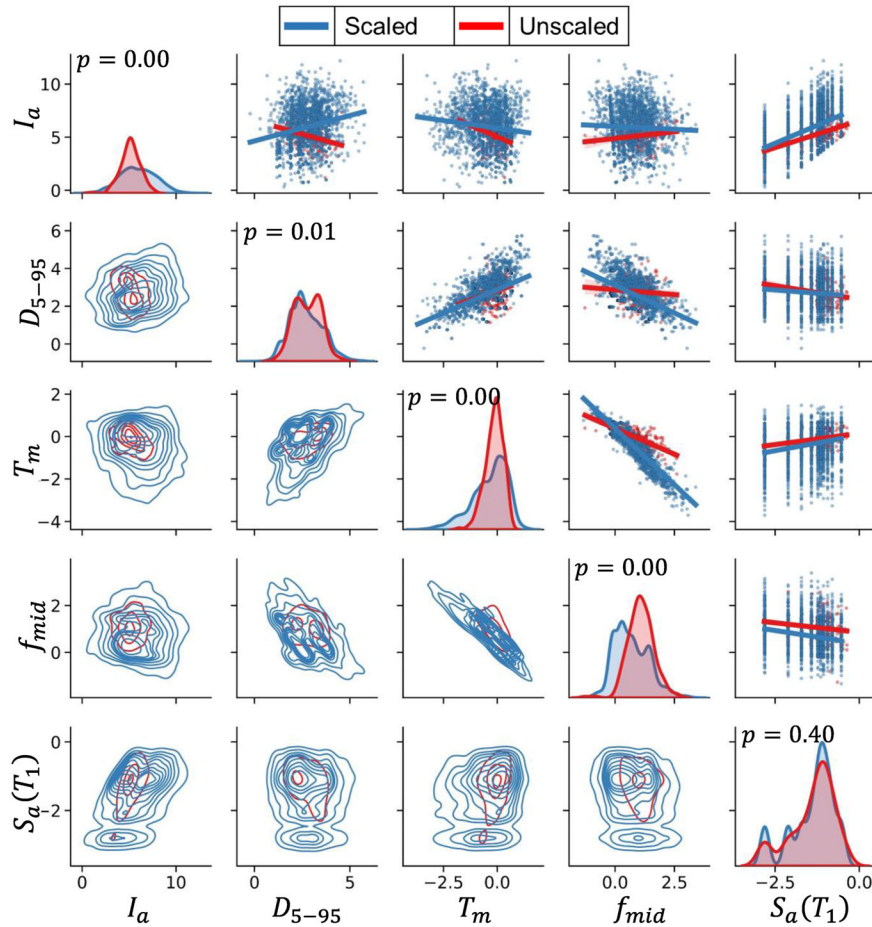


FIGURE 11 Pair plot between various IMs of selected scaled and unscaled ground motions for the eight-story SMF (in the natural logarithmic domain)

The findings presented in Figures 9 and 10 are consistent with the previous studies conducted on the topic (e.g.^{3,4,10,64}), implying that record scaling leads to bias in EDPs. The reasons behind this bias, as pointed out by other authors (e.g.^{3,4,10,64}), can be due to: i) strong ground motion duration (which is related with M_w) is not affected by amplitude scaling and has been reported to influence the behavior of buildings⁶⁵; ii) ground motions with benign amplitude are typically recorded for small M_w or large R_{rup} , and consequently require large scaling for utilization at engineering hazard levels. Such scaled records lead to bias in EDPs due to the fact that the scaled $S_a(T_1)$ is not aligned with the corresponding M_w or R_{rup} and hence other related IMs are not consistently altered^{4,57,66}; iii) as reported by Davalos and Miranda³, the scaled ground motion records lead to larger input energy due to acceleration pulses with larger amplitudes compared to the input energy of unscaled records for similar intensity levels. In a nutshell, the EDP biases can adhere to the fact that the amplitude scaling in terms of $S_a(T_1)$ or any other scalar IM, do not appropriately scale the other correlated IMs of the ground motions (such as energy-content, frequency, duration, etc.). Hence the scaled ground motions do not inherit the internal correlations among the various IMs and event parameters (such as M_w , R_{rup} , etc.). Additionally, as observed from the figures, the range of SFs with possible bias increases with an increase in the ground motion intensity (i.e., $S_a(T_1)$). This is directly related to the level of inelasticity developed in the buildings due to high intensity ground motions. Note that the analyses in this study are conducted to a $S_a(T_1)$ equal to the MCE level. Hence it can be concluded that the collapse-oriented analyses are strongly affected by scaling the low-amplitude records with a factor of ~ 2 and above. Accordingly, it is observed that the levels associated with serviceability, design, and MCE requirements have different allowable scaling ranges that should be considered for ground motion selection and scaling procedures.

To briefly explore the reasoning behind the observed biases and provide new insights, Figure 11 compares different ground motion IMs of the unscaled and scaled records (in the natural logarithmic domain). In this figure, the red and blue colors represent the unscaled and scaled records, respectively. The diagonal insets of the figure present the normalized

kernel density estimations (KDE) of Arias Intensity (I_a), significant duration (D_{5-95}), mean period (T_m), mid-frequency (f_{mid}), and $S_a(T_1)$ for 2500 scaled (blue color) and 200 unscaled (red color) records selected for the eight-story SMF. These IMs are used as proxies for various engineering characteristics of the ground motions: I_a represents the energy content, D_{5-95} is the significant duration of shaking, T_m signifies the dominant frequency content,⁶⁷ f_{mid} represents the frequency at the middle of the shaking,⁶⁸ and $S_a(T_1)$ represents the amplitude of the ground motions. The off-diagonal subplots in Figure 11 show relational plots (scaled and unscaled records) among the five IMs presented for two measures in each subplot. Specifically, the upper off-diagonal insets present the scatter plots (along with mean regressions) between each pair of IMs. In contrast, the lower off-diagonal insets show the density contour of the scatter plots.

The KDEs for the five IMs are compared between the scaled and unscaled ground motions using the KS hypothesis test with the null hypothesis that the two KDEs are statistically similar. It can be observed that except $S_a(T_1)$ all the other comparisons lead to p -values < 0.05 , indicating rejection of the null hypothesis. The KDE match for $S_a(T_1)$ is obvious here since the 2500 scaled ground motions are selected with the criteria to match the $S_a(T_1)$ of the 200 unscaled ground motions. This comparison of IMs signifies that even though the scaled ground motions are selected to make sure they have consistent amplitude-based $S_a(T_1)$ distributions with unscaled ground motions do not necessarily lead to statistical match in other types of IMs. Notably, it can be observed that the two sets of ground motions are significantly different in terms of their energy- and frequency- content (I_a , T_m , and f_{mid}). These observations are consistent with the conclusions made by Davalos and Miranda³ with respect to the difference in energy between scaled and unscaled ground motions.

Furthermore, it can be observed from the upper off-diagonal insets in Figure 11 that the scaled and unscaled ground motion IMs differ considerably in terms of their cross-relations. Notably, the relationships between the other IMs and $S_a(T_1)$ are observed to be different in the two sets except possibly for the case of D_{5-95} . Since the ground motions used in this study only correspond to the crustal seismic sources, D_{5-95} is known to not vary significantly (due to the inclusion of pulse- characteristics).⁶⁹ Hence in Figure 11, it can be observed that, generally, the trends with D_{5-95} are not extremely sensitive to ground motion selection using $S_a(T_1)$. In addition, the trends of other IMs concerning f_{mid} are observed to be considerably different between the scaled and unscaled ground motions. This means that the selection and scaling of ground motions using only amplitude-based $S_a(T_1)$ is not sufficient in eliminating the bias that the mismatches can cause in the other types of IMs. This observation is consistent with the results of Fayaz et al.,⁶⁸ who tested the efficiency and sufficiency of different IMs to explain the EDPs of bridge structures in the context of ground motion simulation validation. Similar observations were made by Li et al.⁷⁰ who pointed out the importance of preserving I_a in addition to $S_a(T_1)$ in liquefaction analysis. While such detailed analysis is out of the scope of this paper, it is worth mention that Bradley⁷¹ proposed a general procedure to built vector-values IMs to develop hazard consistent analysis when more than one IMs are required. The lower off-diagonal elements of Figure 11 compare the density contours between the selected scaled and unscaled ground motions. In general, it can be observed that the scaled ground motions have higher variability as compared to the unscaled ground motions. One of the reasons for this can be due to the sampling bias where only 200 unscaled ground motions are compared against 2500 scaled ground motions. However, it is also viable that natural ground motions corresponding to a $S_a(T_1)$ level have lower variability in terms of other IMs, while scaling of ground motions to match a $S_a(T_1)$ do not appropriately influence the other IMs, thereby leading to higher variability. Also, it is observed from the contours that, in many cases, the mean of the unscaled ground motions contour is not similar to the density contour of scaled ground motions, indicating bias due to scaling. Hence, all these mismatches in the IMs between the scaled and unscaled sets can be the basis of the mismatches observed in the EDPs of SMFs. Although the reasons presented here complement the findings of previous studies, the goal of this study is not to conduct an in-depth investigation of the underlying causes of the EDP bias but to present an overview of the possible explanations and provide recommendations in terms of $S_a(T_1)$ and SF . The underlying reasoning behind the EDP mismatch requires further scrutiny and is currently beyond the scope of this study. Finally, the results of this study suggest that the scaling of ground motions requires careful examination in the context of seismic demand analysis.

2 | CONCLUSION

The lack of high amplitude ground motion records has made ground motion selection and scaling process an essential step for the seismic demand analysis in structural and earthquake engineering. Over the years, various measures have been suggested to minimize the impact of extremeground motion scaling on structural demands. This study is a step to provide statistical and machine-learning-based recommendations for scaling ground motions in terms of IM and SF . The study proposes and implements a 10-stepped procedure to assess the viability of scaling ground motions to match average

IM levels and its impact on the structural responses in conjunction with the ground motion intensity level. The procedure involves careful selection of scaled and unscaled ground motions corresponding to different IM levels and using them for NTHA of structural models. The results are then used to develop machine-learning-based surrogate models and obtain statistical predictions for scaled and unscaled responses for various IM and SF levels. The scaled and unscaled response predictions are then compared through hypotheses testing to determine the SF ranges of no-significant bias and provide recommendations for ground motion scaling process.

As an exercise of the proposed procedure, the study investigates the impact of ground motion amplitude scaling on two seismic EDPs (i.e., PFA and PIDR) of five steel SMFs varying in their heights (2-,4-,8-,12-, and 20-story) using Bayesian non-parametric GPR-based surrogate models. For this purpose, a total of 200 unscaled ground motions and corresponding 2500 scaled ground motions are selected to represent 10 $S_a(T_1)$ levels from 0 g to MCE level (2% in 50 years hazard level). Hence a total of 2700 ground motions are applied to each SMF to conduct NTHA, and the resulting EDPs are utilized to develop GPR-based surrogate models for unscaled and scaled responses. To have better insights into the GPR models, their prediction- and decision-making- processes are explained using the concepts of XAI and SHAP. The scaled GPR surrogate models are then utilized to sample the EDPs for a 50×50 grid of $S_a(T_1)$ and SF levels. The obtained posterior distribution of scaled EDPs for each grid point is then compared against the posterior distribution of unscaled EDPs sampled from the unscaled GPR model corresponding to the same $S_a(T_1)$. The comparison is conducted using the KS test, where the two distributions are analyzed for statistical similarity and hence checked for possible bias in EDPs due to scaling.

Results of this study indicate that the record scaling (both multiplicative and divisive) introduces bias in the responses of SMFs for both EDPs (i.e., PFA and PIDR). It is concluded that the EDP bias caused by scaling depends on three factors: i) scaling level (i.e., SF), ii) target amplitude (i.e., $S_a(T_1)$), and iii) fundamental period (i.e., mass and stiffness distribution). In general, it is noticed that the allowable SF s range between 0.5 and 3.0 for PIDRs, and from 0.6 to 2.0 for PFAs (with p -value = 0.05). It is further seen that the allowable scaling range decreases with an increase in the target $S_a(T_1)$ leading to an allowable SF range between 0.7 and 1.8 for PIDRs, and from 0.8 to 1.6 for PFAs (with p -value = 0.05) for MCE levels. These analyses are essential from both design and analysis points of view since drift limits typically control the design of SMFs, while PFAs define the anchor forces for detailing nonstructural components. Finally, the EDP differences between scaled and unscaled ground motions are briefly explained by analyzing the differences in five different IMs of the ground motions characterizing the energy-, duration-, and frequency content of the ground motions. The analysis shows that a sheer scaling of ground motions using amplitude-based IMs such as $S_a(T_1)$ do not appropriately alter the other IMs, and a significant inconsistency is noticed between the other IM relations for the scaled and unscaled ground motions. These IM inconsistencies are postulated to be the reasons behind the mismatch in the scaled and unscaled EDPs. Based on the results of this study, it is clear that further scrutiny and studies are required to understand the underlying phenomenon of scaling ground motions. Furthermore, it is vital to understand the impact of other IMs in the record selection and scaling process, and also, with the influx of further recorded/simulated data, the analysis can be conducted by taking the variability of ground motions and its consequences into account. This study can provide a resource to the community for the utilization of sophisticated machine-learning tools to obtain insights about ground motion selection and scaling problems.

ACKNOWLEDGMENTS

None.

CONFLICT OF INTEREST

The authors confirm that there is no potential conflict of interest with anyone for this study

DATA AVAILABILITY STATEMENT

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

ORCID

Jawad Fayaz  <https://orcid.org/0000-0003-4706-9761>

REFERENCES

1. American Society of Civil Engineers. *Minimum design loads for buildings and other structures*. American Society of Civil Engineers (ASCE). 2016.
2. Cochran BL, Hart GC. Los Angeles Tall Buildings Structural Design Council. *Struct Des Tall Spec Build*. 2010;19(1-2):1-3.

3. Dávalos H, Miranda E. Evaluation of bias on the probability of collapse from amplitude scaling using spectral-shape-matched records. *Earthq Eng Struct Dyn*. 2019;48(8):970-986.
4. Dávalos H, Miranda E. Evaluation of the scaling factor bias influence on the probability of collapse using Sa (T1) as the intensity measure. *Earthq Spectra*. 2019;35(2):679-702.
5. Bommer JJ, Acevedo AB. The use of real earthquake accelerograms as input to dynamic analysis. *J Earthquake Eng*. 2004;8(spec01):43-91.
6. Krinitzky EL, Chang FK. *State-of-the-art for Assessing Earthquake Hazards in the United States: Report 4: Earthquake Intensity and the Selection of Ground Motions for Seismic Design*. US Army Engineer Waterways Experiment Station; 1975.
7. Shome N, Cornell CA, Bazzurro P, Carballo JE. Earthquakes, records, and nonlinear responses. *Earthq Spectra*. 1998;14(3):469-500.
8. Shome N, Cornell CA. Probabilistic seismic demand analysis of nonlinear structures. *Civ Environ Eng Rep*. 1999. Reliability of Marine Structures Program Report No. RMS-35.
9. Iervolino I, Cornell CA. Record selection for nonlinear seismic analysis of structures. *Earthq Spectra*. 2005;21(3):685-713.
10. 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.
11. Baker JW, Allin Cornell C. Spectral shape, epsilon and record selection. *Earthq Eng Struct Dyn*. 2006;35(9):1077-1095.
12. Zacharenaki A, Fragiadakis M, Assimaki D, Papadrakakis M. Bias assessment in incremental dynamic analysis due to record scaling. *Soil Dyn Earthquake Eng*. 2014;67:158-168.
13. Du W, Ning C, Wang G. The effect of amplitude scaling limits on conditional spectrum-based ground motion selection. *Earthq Eng Struct Dyn*. 2019;48(9):1030-1044.
14. Lin T, Haselton CB, Baker JW. Conditional spectrum-based ground motion selection. Part I: hazard consistency for risk-based assessments. *Earthq Eng Struct Dyn*. 2013;42(12):1847-1865.
15. Ancheta TD, Darragh RB, Stewart JP, et al. NGA-West2 database. *Earthq Spectra*. 2014;30(3):989-1005.
16. Fayaz J, Zareian F. An efficient algorithm to simulate site-based ground motions that match a target spectrum. *Earthq Eng Struct Dyn*. 2021;50(13):3532-3549.
17. Konstantinos Theodoros Tsalouchidis; Christoph Adam. Amplitude scaling of ground motions as a potential source of bias: large-scale investigations on structural drifts. *Earthq Eng Struct Dyn*. 2022;51:2904-2924.
18. Baker JW. Measuring bias in structural response caused by ground motion scaling. *Pacific Conference on Earthq Eng*. 2007(056):1-6. doi:10.1002/eqe
19. Zareian F, Kanvinde A. Effect of column-base flexibility on the seismic response and safety of steel moment-resisting frames. *Earthq Spectra*. 2013;29(4):1537-1559. doi:10.1193/030512EQS062M
20. Torres-Rodas P, Flores F, Pozo S, Astudillo BX. Seismic performance of steel moment frames considering the effects of column-base hysteretic behavior and gravity framing system. *Soil Dyn Earthquake Eng*. 2021;144(November 2020):106654. doi:10.1016/j.soildyn.2021.106654
21. Torres-Rodas P, Flores F, Zareian F. Seismic response of steel moment frame considering gravity system and column base flexibility. *NCEE 2018: Integrating Sci Eng and Policy*. 2018;12:7383-7393. *11th National Conference on Earthquake Engineering 2018*.
22. Miranda E, Eads L, Dávalos H, Lignos D. Assessment of the probability of collapse of structures during earthquakes. *Proceedings of the 16th World Conference on Earthq Eng*. 2017;2441(Abstract ID):1-12.
23. American Institute of Steel Construction. *Seismic Provisions for Structural Steel Buildings*. ANSI/AISC. 2016:341-316.
24. American Institute of Steel Construction. *Prequalified Connections for Special and Intermediate Steel Moment Frames for Seismic Applications*. ANSI/AISC. 358-316. 2016.
25. Kircher C, Deierlein G, Hooper J, et al. Evaluation of the FEMA P-695 methodology for quantification of building seismic performance factors 2010.
26. McKenna F, Fenves GL, Scott MH. *OpenSees: Open system for earthquake engineering simulation*. Pacific Earthquake Engineering Research Center, University of California; 2006. <http://OpenseesBerkeleyEdu>
27. Ibarra LF, Medina RA, Krawinkler H. Hysteretic models that incorporate strength and stiffness deterioration. *Earthquake Eng Struct Dyn*. 2005;34(12):1489-1511. doi:10.1002/eqe.495
28. NIST GCR 17-917-46v2. Guidelines for Nonlinear Structural Analysis for Design of Buildings Part IIa – Steel Moment Frames 2017:145.
29. American Institute of Steel Construction. *Specification for Structural Steel Buildings*. ANSI/AISC; 2016:360-316.
30. Kanvinde AM. *Micromechanical simulation of earthquake-induced fracture in steel structures*. Stanford University; 2004.
31. Zareian F. *Simplified performance-based earthquake engineering*. Stanford University; 2006.
32. Zareian F, Krawinkler H. Assessment of probability of collapse and design for collapse safety. *Earthq Eng Struct Dyn*. 2007;36(13):1901-1914.
33. Gupta A, Krawinkler H. *Prediction of Seismic Demands for SMRF's with Ductile Connections and Elements*. SAC Joint Venture; 2002.
34. Rodas PT, Zareian F, Kanvinde A. Hysteretic model for exposed column–base connections. *J Struct Eng*. 2016;142(12):04016137. doi:10.1061/(asce)st.1943-541x.0001602
35. Torres-Rodas P, Zareian F, Kanvinde A. A hysteretic model for the rotational response of embedded column base connections. *Soil Dyn Earthquake Eng*. 2018;115:55-65. doi:10.1016/j.soildyn.2018.08.015
36. Torres-Rodas P, Fayaz J, Zareian F. Strength resistance factors for seismic design of exposed based plate connections in special steel moment resisting frames. *Earthq Spectra*. 2020;36(2):537-553. doi:10.1177/8755293019891714
37. Fayaz J, Xiang Y, Zareian F. Generalized ground motion prediction model using hybrid recurrent neural network. *Earthquake Eng Struct Dyn*. 2020;50(November):1539-1561. doi:10.1002/eqe.3410

38. U.S. Geological Survey 2020. *Earthquake Lists, Maps, and Statistics*.
39. Williams CK, Rasmussen CE. *Gaussian processes for machine learning*. MIT press; 2006.
40. MacKay DJC, Mac Kay DJC. *Information theory, inference and learning algorithms*. Cambridge University Press; 2003.
41. Williams C, Rasmussen C. Gaussian processes for regression. *Adv Neural Inf Process Syst*. 1995;8.
42. Williams CKI. *Prediction with Gaussian processes: From linear regression to linear prediction and beyond*. Learning in graphical models, Springer; 1998.
43. Neal RM. *Bayesian learning for neural networks*. Springer Science & Business Media; 2012.
44. Hoffer JG, Geiger BC, Kern R. Gaussian process surrogates for modeling uncertainties in a use case of forging superalloys. *Appl Sci*. 2022;12(3):1089.
45. Rajaram D, Puranik TG, Renganathan A, et al. *AIAA Scitech 2020 Forum*. 2020. https://smartech.gatech.edu/bitstream/handle/1853/62530/DR_SciTech_2020_Deep_Gaussian_Process_Enabled_Surrogate_Models_for_Aerodynamic_Flows.pdf
46. Gentile R, Galasso C. Gaussian process regression for seismic fragility assessment of building portfolios. *Struct Saf*. 2020;87:101980.
47. Park C, Rao R, Nikolaev P, Maruyama B. Gaussian process surrogate modeling under control uncertainties for yield prediction of carbon nanotube production processes. *J Manuf Sci Eng*. 2022;144(3).
48. Bajer L, Pitra Z, Repický J, Holeňa M. Gaussian process surrogate models for the CMA evolution strategy. *Evol Comput*. 2019;27(4):665-697.
49. Yang Y, Yang Y, Xiao M, Wan M, Zhang W. A Gaussian process regression-based surrogate model of the varying workpiece dynamics for chatter prediction in milling of thin-walled structures. *Int J Mech Syst Dyn*. 2022;2:117-130.
50. Kramer SL. *Geotechnical earthquake engineering*. Pearson Education India; 1996.
51. Baker J, Bradley B, Stafford P. *Seismic hazard and risk analysis*. Cambridge University Press; 2021.
52. Applied Technology Council. *Quantification of building seismic performance factors*. US Department of Homeland Security, FEMA P695; 2009.
53. Fayaz J, Azar S, Dabaghi M, Zareian F. An efficient algorithm to simulate hazard-targeted site-based synthetic ground motions. *Earthq Spectra*. 2021;37(2):876-902.
54. Norouzi A, Poursha M. Effect of the spectral shape of ground motion records on the collapse fragility assessment of degrading SDOF systems. *Earthq Eng Eng Vib*. 2021;20(4):925-941.
55. Eads L, Miranda E, Lignos D. Spectral shape metrics and structural collapse potential. *Earthq Eng Struct Dyn*. 2016;45(10):1643-1659.
56. Bojórquez E, Iervolino I. Spectral shape proxies and nonlinear structural response. *Soil Dyn Earthquake Eng*. 2011;31(7):996-1008.
57. Eads L, Miranda E, Lignos DG. Average spectral acceleration as an intensity measure for collapse risk assessment. *Earthq Eng Struct Dyn*. 2015;44(12):2057-2073.
58. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst*. 2017;30.
59. Roth AE. *The Shapley value: essays in honor of Lloyd S. Shapley*. Cambridge University Press; 1988.
60. Molnar C. Interpretable machine learning. Lulu. com; 2020.
61. Lin J, Mahin SA. Seismic response of light subsystems on inelastic structures. *J Struct Eng*. 1985;111(2):400-417.
62. Taghavi S, Miranda MM. *Response assessment of nonstructural building elements*. Pacific Earthquake Engineering Research Center; 2003.
63. Flores FX, Lopez-Garcia D, Charney FA. Assessment of floor accelerations in special steel moment frames. *J Constr Steel Res*. 2015;106:154-165.
64. Baker JW. *Measuring bias in structural response caused by ground motion scaling*. 2007.
65. Chandramohan R, Baker JW, Deierlein GG. Quantifying the influence of ground motion duration on structural collapse capacity using spectrally equivalent records. *Earthq Spectra*. 2016;32(2):927-950.
66. Luco N, Cornell CA. Structure-specific scalar intensity measures for near-source and ordinary earthquake ground motions. *Earthq Spectra*. 2007;23(2):357-392.
67. Fayaz J, Galasso C. A deep neural network framework for real-time on-site estimation of acceleration response spectra of seismic ground motions. *Comput-Aided Civ Infrastruct Eng*. 2022.
68. Fayaz J, Azar S, Dabaghi M, Zareian F. Methodology for validation of simulated ground motions for seismic response assessment: application to CyberShake source-based ground motions. *Bull Seismol Soc Am*. 2021;111(1):226-241.
69. Fayaz J, Medalla M, Zareian F. Sensitivity of the response of Box-Girder Seat-type bridges to the duration of ground motions arising from crustal and subduction earthquakes. *Eng Struct*. 2020;219:110845.
70. Li L, Huang D, Jin F, Du S. Evaluation of synthetic and recorded ground motions for soil liquefaction analyses. *Soil Dyn Earthquake Eng*. 2021;149:106815.
71. Bradley BA. A generalized conditional intensity measure approach and holistic ground-motion selection. *Earthq Eng Struct Dyn*. 2010;39(12):1321-1342.

How to cite this article: 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. *Earthquake Engng Struct Dyn*. 2023;1-21. <https://doi.org/10.1002/eqe.3810>