14th International Conference on Structural Safety and Reliability - ICOSSAR'25
1-6 June, 2025 - Los Angeles California, USA
R. Ghanem, E. Johnson, S. Masri and A. Olivier (Eds)

STATISTICAL APPLICATION OF DEEP LEARNING-BASED GENERALIZED GROUND MOTION MODELS – FROM PROBABILISTIC SEISMIC HAZARD ANALYSIS TO STRUCTURAL FRAGILITY ANALYSIS

Jawad Fayaz[1], Pablo Torres-Rodas*[2], Miguel Medalla[3] and Yijun Xiang[4]
[1]*University of Exeter*
J.Fayaz@exeter.ac.uk
[2]*Universidad San Francisco de Quito*
patorresr@usfq.edu.ec
[3]*Pontificia Universidad Católica de Chile*
mmedalla@ing.puc.cl
[4]*Georgia Institute of Technology*
yijunx3@uci.edu

Key words: Generalized ground motion models, performance-based earthquake engineering, deep learning.

Abstract

This study presents a comprehensive implementation and evaluation of deep learning-based generalized ground motion models (GGMMs) applied across key stages of performance-based earthquake engineering (PBEE). Traditional ground motion models (GMMs) typically rely on assumed empirical relationships between earthquake characteristics—such as magnitude and rupture distance—and intensity measures (IMs) like peak ground acceleration (PGA), spectral acceleration (S_a), etc. They also treat the prediction of IMs independently, disregarding the internal cross-dependencies among them. These limitations, combined with their generally low predictive accuracies, have prompted the development of various data-driven machine learning-based GMMs. In this study, data-driven GGMM proposed by the authors for crustal ground motions respectively, is employed. The GGMMs are based on long short-term memory (LSTM) recurrent neural networks that capture amplitude-based (e.g., PGA, S_a), frequency-based (e.g., peak ground velocity), energy-based (e.g., Arias intensity, cumulative absolute velocity), and duration-based (e.g., significant duration) IMs as a sequence of cross-dependent parameters. Using the earthquake characteristics and site parameters, such as soil shear-wave velocity, as inputs, the GGMMs output a cross-dependent IM vector. Using the earthquake rupture forecast (ERF) data, the GGMMs are applied to conduct probabilistic seismic hazard analysis (PSHA) for Los Angeles, representing crustal seismicity—yielding hazard curves. The hazard curves guide the selection and scaling of 30 ground motions used in multi-strip analysis for three distinct steel moment frames (SMFs) across 10 hazard levels. Nonlinear time-history analyses are performed to derive fragility curves for each SMF, quantifying their vulnerability under different seismic conditions. This process is replicated with conventional GMMs and the conditional mean spectrum (CMS) approach. The results obtained at each stage, i.e., hazard

curves, hazard spectra, selected motions, and SMF seismic demands from the multi-strip analysis, are statistically compared between the conventional GMM and deep learning-based GGMM approaches. The findings underscore the value of GGMMs in both hazard estimation and structural response assessment and finally provide a complete framework for the deployment of the machine learning-based GMMs for PBEE.

1 INTRODUCTION

The core objective of modern earthquake engineering is to ensure that structures can withstand seismic shaking while meeting predefined performance objectives. This performance-based approach evolved in response to the limitations of traditional, code-based prescriptive design, which primarily focuses on life safety rather than controlling damage or ensuring predictable performance. The shortcomings of prescriptive design became evident after major earthquakes, such as the 1994 Northridge and 1995 Kobe events, where unexpected failures—particularly in steel structures—exposed weaknesses in the conventional approach. These failures catalyzed the development of performance-based earthquake engineering (PBEE), which enables a more comprehensive quantification of structural performance across different hazard levels [1].

PBEE adopts a probabilistic framework that integrates four key components: (a) seismic hazard assessment, (b) structural performance modeling, (c) damage assessment, and (d) loss estimation. The first step, seismic hazard assessment, involves a site-specific estimation of ground motion intensity using probabilistic seismic hazard analysis (PSHA). This process relies on ground motion models (GMMs), which utilize the earthquake source (such as magnitude M) and site parameters (site's distance from the rupture) to estimate intensity measures (IMs) such as peak ground acceleration (PGA), spectral acceleration at the first mode period ($S_a(T_1)$), Arias intensity (I_a), cumulative absolute velocity (CAV), significant duration (D_s), and others [2]. Traditionally, GMMs are parametric functional forms where IMs are statistically modeled based on earthquake source and site parameters such as M, closest rupture distance (R_{rup}), soil's average shear-wave velocity in the topmost 30m (V_{s30}), and depth to the top of the rupture (Z_{TOR}). For instance, Abrahamson et al. [3] (ASK14) developed a GMM for shallow crustal earthquakes in active seismic regions like California, Japan, China, and Taiwan. Similarly, Campbell and Bozorgnia [4] (CB14) refined their earlier 2008 model by incorporating regionally independent geometric attenuation, regionally dependent anelastic attenuation, and magnitude-dependent aleatory variability. Chiou and Youngs [5] (CY14) further improved their GMM by addressing regional differences in far-source distances and site effects in active tectonic regions.

Despite their advancements, traditional GMMs exhibit key limitations. Most notably, low prediction accuracy [6], [7], and missing intrinsic cross-correlations between IMs [8], [9]. With the computational advent of machine learning methods, these limitations have driven the development of data-driven deep learning models. For example, Dhanya and Raghukanth [10] employed artificial neural networks (ANNs) combined with genetic algorithms to predict *PGA*, *PGV*, and ($S_a(T_1)$)across 26 periods. More recently, Fayaz et al. [6] proposed generalized ground motion models (GGMM) that use long short-term memory (LSTM) recurrent neural

networks (RNNs) to predict a vector of amplitude-, duration-, energy-, and frequency-based IMs while explicitly capturing their cross-dependencies. This vector of IMs includes Arias intensity (I_a), cumulative absolute velocity (CAV), significant duration (D_{5-95}), and response spectrum ($S_a(T)$) across multiple periods. This RNN was trained using the database from the next generation attenuation (NGA) west-2 [11] for crustal ground motions. In addition to this GGMM and as an extension of this work, Fayaz et al. [7] trained an RNN with Chilean subduction ground motion using the database from SIBER-RISK [12]. This study included an application of the GGMM in the estimation of structural demands from an archetype frame.

Building on this foundation, the present study conducts a comprehensive evaluation of traditional and deep-learning-based GMMs. Specifically, the models ASK14, CB14, CY14, and the RNN model proposed by Fayaz [6] et al. (referred to hereafter as GGMM21) are used to conduct PSHA at Los Angeles Downtown (LADT), United States. The resulting hazard curves from the four models are statistically compared. The hazard curves are further used to inform the selection and scaling of 30 ground motion records for multi-stripe nonlinear time-history analysis (NTHA) of three steel moment frames (SMFs) at 10 hazard levels. The SMFs vary in height (3-, 8-, and 12- stories) and are designed according to the American Society of Civil Engineering (ASCE) 7-22 [13] and American Institute of Steel Construction (AISC) 341-22 [14]. The study further analyzes the engineering demand parameters (EDPs), such as the peak inter-story drift ratios (PIDRs) and peak floor accelerations (PFAs), and compares them across the four GGMMs.

2 PERFORMANCE-BASED EARTHQUAKE ENGINEERING FRAMEWORK

The PBEE framework integrates four modules: a) PSHA (through IMs), b) structural demand analysis (through EDPs), c) damage assessment (through damage measure, DM), and d) loss estimation (through decision variable, DV). This is expressed in Equation 1 where G represents the probability of exceedance, λ_{IM} is the rate of exceedance of the IM, and λ_{DV} is the rate of exceedance of the DV. This study is limited to the EDP level, and this section summarises the steps involved in the process.

$$\lambda_{DV} = \sum\sum\sum G(DV|DM)dG(DM|EDP)dG(EDP|IM)d\lambda_{IM} \quad (1)$$

2.1 Probabilistic Seismic Hazard Analysis

PSHA, in general, involves five steps: a) Identification of the earthquake sources, b) determination of the distribution of earthquake magnitudes, c) determination of the distribution of source-to-site-distances, d) prediction of the distribution of ground motion intensity given the casual and site parameters, and e) combinations of uncertainties in earthquake size, location, and intensity resulting in a complete distribution of ground motion intensity at different levels. This is done through Equation 2, where $\lambda(M_i > m_{min})$ is the rate of occurrence of earthquakes greater than m_{min} from the source, $\lambda(IM > x)$ is the rate of IM greater than a set IM value x, $n_{sources}$ is the number of earthquake sources, and M and R represent the magnitude and rupture distance. This process leads to the generation of hazard curves for a given location.

$$\lambda(IM > x) = \sum_{i=1}^{n_{sources}} \lambda(M_i > m_{min}) \sum_{j=1}^{n_M} \sum_{k=1}^{n_R} G(IM|m_j, r_k) P(M_i = m_j) P(R_i = r_k) \quad (2)$$

This study utilizes the uniform California earthquake rupture forecast, version 2 (UCERF2) [15] for obtaining the source and site parameters for the LADT site. These parameters are used regularly by seismic hazard platforms such as the OpenSHA software [16]. The obtained parameters are used as inputs to the four GMMs (i.e., ASK14, CB14, CY14, and GGMM21) to predict IMs and generate hazard curves using Equation 2.

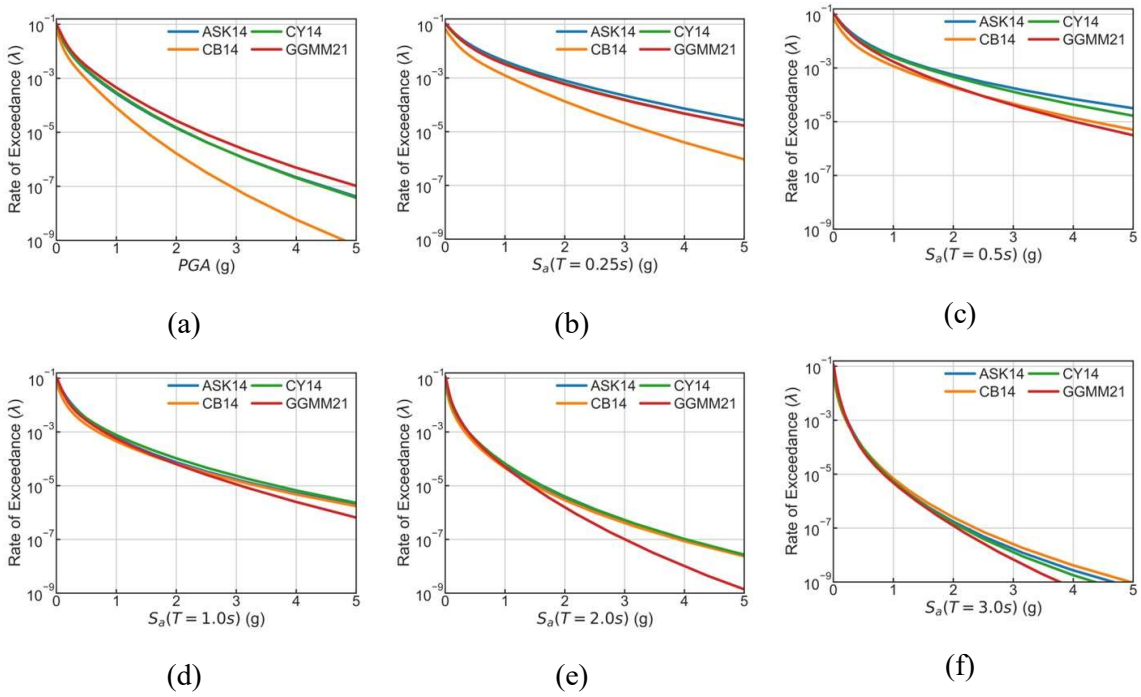


Figure 1: $S_a(T)$ hazard curves computed from the four GMMs for LADT for six periods: (a) PGA ($T = 0$s), (b) $T = 0.25$s, (c) $T = 0.5$s, (d) $T = 1$s, (e) $T = 2$s, and (f) $T = 3$s

Figure 1 illustrates hazard curves generated from the four GMMs, i.e., ASK14 (blue), CY14 (green), CB14 (orange), and GGMM21 (red) for LADT at six distinct periods ranging from PGA (i.e., $T = 0$s) to 3.0s. In comparing these curves, CB14 lies notably below the other three across the short period range of PGA ($T = 0$s) and $S_a(T = 0.25s)$. The GGMM21 tends to be higher than the other three GMMs for these short-period domains, indicating higher expected hazard demands than the other GMMs. From mid-period ranges $S_a(T = 0.5s)$ and $S_a(T = 1s)$, the GGMM21 tends to be lower than ASK14 and CY14 but coincides mostly with the CB14. As the period increases, the GGMM21 leads to lower demands than the other three GMMs, leading to significantly lower demands. On the other hand, CB14 leads to increased demands with an increase in the period. This is due to the fact the CB14 doesn't capture the decay in the spectral values with an increase in the period. As conventional GMMs like ASK14, CB14, and

CY14 are fitted through statistical regression techniques, they tend to fit average trends that are influenced by extreme values much more than deep learning-based models. This is also the reason that post-processing methods like conditional mean spectrum (CMS) [17] and generalized conditional intensity measure (GCIM) [9] are utilized in conjunction with the GMMs to obtain realistic spectral shapes, including cross-correlations. Due to the recurrent structure of the neural networks in GGMM21, the outputs don't require any post-processing, and they perform well in making sure predictions are accurate and implicitly cross-correlated.

2.2 Ground Motion Selection and Scaling

The generated hazard curves are used for ground motion selection and scaling. This is done by obtaining the uniform hazard spectra (UHS) for 10 return periods using the hazard curves of the four GMMs. The UHS from ASK14, CB14, and CY14 are post-processed into CMS for three periods corresponding to the first model periods of the three SMFs used in this study, i.e., 1.1s, 2.4s, and 3.0s. The CMS incorporates the cross S_a correlations to give the target spectra a realistic shape. As mentioned earlier, this post-processing is not required for the GGMM21 due to the inherent nature of capturing cross-IM dependencies.

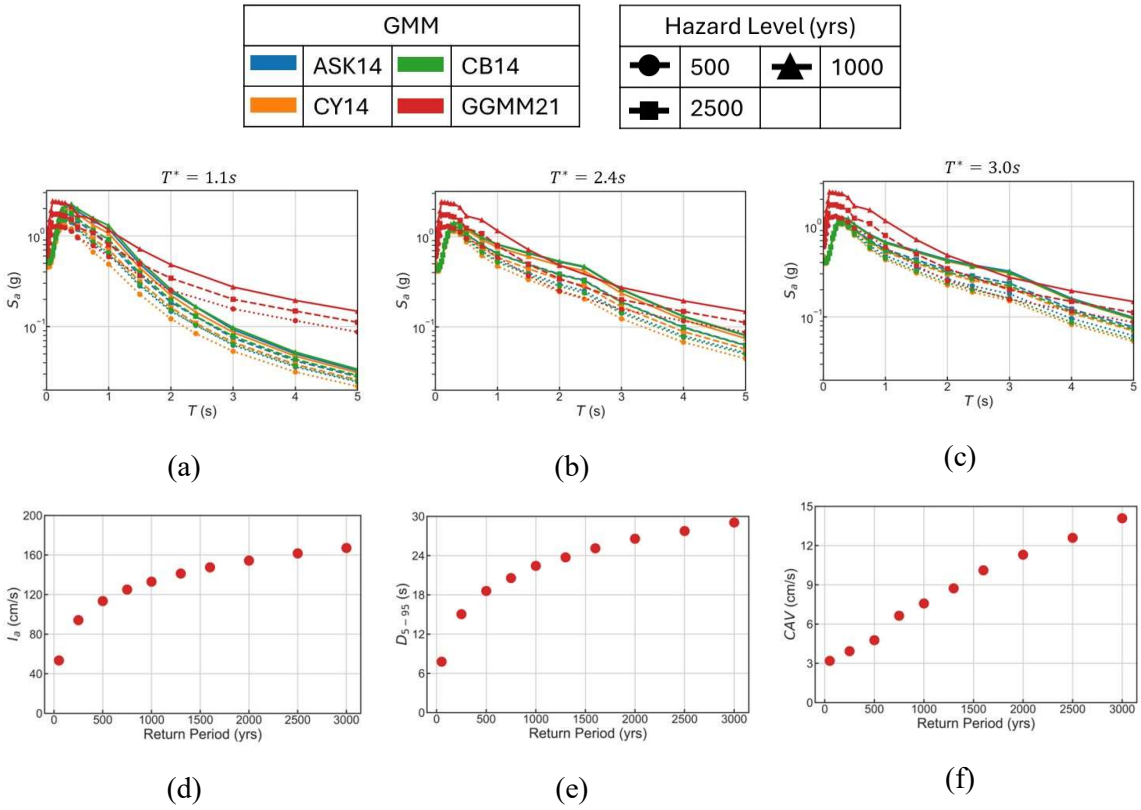


Figure 2: CMSs conditioned on three structural periods: (a) $T = 1.1$s, (b) $T = 2.4$s, and (c) $T = 3.0$s with UHS from GGMM21, and other IMs from GGMM21: (d) I_a, (e) D_{5-95}, and (f) CAV across different return periods

Figure 2 illustrates the CMS conditioned three periods (1.1s, 2.4s, and 3.0s), together with the UHS from GGMM21 (referred to as GMS from here on) and additional IMs from GGMM21 across different return periods. CMS of three GMMs – ASK14 (blue), CY14 (green), and CB14 (orange), and GMS from GGMM21 (red) are plotted, each with three levels of return periods marked by a distinct shape – circles for 500 years, triangles for 1000 years and squares for 2500 years. At $T = 1.1$s, the GGMM21 model closely follows the other three GMMs from periods ranging from 0s to 2s, particularly near the conditional period of 1.1s. However, beyond $T = 2.0$s, its S_a are higher than the other three models, highlighting a notable shift in the predicted hazard curves for longer periods. This is observed for all three hazard levels, with an increase in hazard levels increasing the discrepancies between the CMSs and the GMS. By contrast, for longer periods, i.e., $T = 2.4$s and $T = 3.0$s, the GMS essentially coincides with the other three models throughout the entire range of spectral periods. The four models predict nearly identical spectral accelerations, especially near the conditional periods, and this alignment holds across multiple return periods. The high consistency implies that design engineers can make relatively confident decisions during these intermediate-to-long periods, as the models offer minimal divergence.

For each of the 10 hazard levels, the CMSs and GMSs are utilized to select and scale ground motions from the NGA west-2 database [11]. The ground motions are selected through range matching, where the ground motion spectra and target spectra are matched for S_a between the periods of $0.5T_1$ to $2T_1$ where T_1 represents the first mode period of the SMFs, i.e., 1.1s, 2.4s, and 3.0s. 30 ground motions are selected for each of the four GMMs (three CMSs and one GMS for each hazard level and SMF) while limiting the scaling factors between 0.3 to 3 to minimize biases due to amplitude scaling [18]. Hence, four sets of 30 ground motions are selected for each hazard level and SMF. To further utilize the other IMs from GGMM21 in the ground motion selection and scaling process, fifth set of ground motions is selected using a weighted objective function as expressed in equations 3 to 5 [7]. In this case, the selection is conducted using S_a between the periods of $0.5T_1$ to $2T_1$ and the other three IMs (CAV, I_a, and D_{5-95}). It should be noted here that the scaling process alters all the elements of IM vector (*i.e.,* CAV and $S_a(T)$ are scaled linearly by scaling factor and I_a is scaled by the square of the scaling factor) of the ground motions except for the duration D_{5-95} (which is not scaled). Hence, to select the ground motions that match the IM vectors, firstly, the ground motions that have D_{5-95} between 75% to 150% of the target D_{5-95} for the given hazard level are used as the ground motion selection pool. Subsequently, the other IMs of the pooled ground motions are then scaled from 0.3 to 3.0 and compared against the target IM vector of the given hazard level. Due to the different scales of the IM values, the comparisons are performed using a weighted average accuracy, where accuracies for I_a and CAV are computed using the normalized accuracy metric given in Equation 3, and the accuracy of the $S_a(T)$ spectrum is computed using the index of agreement in Equation 4 [19]. Both equations 3 and 4 are bounded between 0 and 1, where 0 is the worst match, and 1 is the best match. Then, the accuracy measures are combined using a weighted average accuracy (Acc_{avg}), computed as per Equation 5. For illustrative purposes, in this case, the weights are chosen as $w_1 = 0.2$; $w_2 = 0.2$; and $w_3 = 0.6$; however, users can freely choose weights based on their own criteria and/or structural system sensitivity. This leads

to five sets of 30 ground motions for each combination of the 10 hazard levels and the three SMFs.

$$Acc_{IM} = 1 - \text{abs}\left(\frac{IM_{Tar} - IM_{GM}}{IM_{Tar} + IM_{GM}}\right) \tag{3}$$

$$IA_{S_a} = 1 - \frac{\sum_{i=1}^{n}\left[S_{a,Tar}(T_i) - S_{a,GM}(T_i)\right]^2}{\sum_{i=1}^{n}\left[\left|S_{a,Tar}(T_i) - \frac{\sum_{i=1}^{n} S_{a,Tar}(T_i)}{n}\right| + \left|S_{a,GM}(T_i) - \frac{\sum_{i=1}^{n} S_{a,Tar}(T_i)}{n}\right|\right]^2} \tag{4}$$

$$Acc_{avg} = \frac{w_1 Acc_{I_a} + w_2 Acc_{CAV} + w_3 IA_{S_a}}{w_1 + w_2 + w_3} \tag{5}$$

2.3 Structural Nonlinear Time-History Analysis

The ground motions are then applied to conduct multi-strip NTHA of the three archetype SMFs. The SMFs share the same plan as the archetypes developed for the evaluation of the Federal Emergency Management Agency (FEMA) P-695 methodology project [20]. Since these archetypes were designed as per ASCE 7-05 [21], in this study, the cross sections are updated based on the newest building codes (i.e., ASCE 7-22 [22] and AISC 341-22 [23]). Thus, the three SMFs vary in height (3-, 8-, and 12-story) while sharing the same plan view. A dead load of 4.78 kN/m^2 and a live load of 2.38 kN/m^2 are applied to all floors. Additionally, a perimeter load of 1.20 kN/m^2 was applied to simulate cladding. The first story height is 4.5m, while the typical story height is 3.90m. The SMFs are located at the perimeter of the building and consist of three bay frames.

The finite element models are developed in the open system for earthquake engineering simulation (OpenSEES) platform for NTHA [24]. Beams and column members are idealized as linear elastic elements with concentrated rotational hinges at their ends. These nonlinear springs are represented by the well-known IMK bilinear models [25] with a trilinear backbone curve, predefined rules for defining the hysteretic behavior, and rules for capturing cyclic deterioration. The panel zones are modeled as a hinge parallelogram assembly by rigid elements with a nonlinear spring at one of the corners. In addition, a leaning column is used to simulate the P-delta effects through large geometric nonlinearities. The parameters of the IMK model are determined based on the recommendations given at NIST GCR 17-917-46v2 [26].
With this setting, the multi-strip ground motions are used for the NTHA of the three SMFs and the peak inter-story drift ratios ($PIDR$) and peak floor accelerations (PFA) are recorded. The 10 different seismic hazard levels chosen for these analyses include: 50, 250, 500, 750, 1000, 1300, 1600, 2000, 2500, and 3000 years. Hence a total of 5 ground motion sets x 10 stripes x 30 GMs x 3 SMFs = 4500 NTHA are conducted.
Figure 3 illustrates the probability density histograms for two EDPs (PIDR and PFA) for the 8-story SMF under the ground motions selected for the 2500 return period. The figure compares EDPs from the ground motions selected using ASK14 (blue), CY14 (green), CB14 (orange), GGMM21 (red), and GGMM21 with all IMs (GGMM21$_{\text{allIMs}}$; cyan). Both GGMM21 and GGMM21$_{\text{allIMs}}$ exhibit slightly elevated probability densities for PIDR >0.03, indicating a high

level of nonlinearity. This indicates a modest shift towards higher inter-story drifts for the ground motions selected by these two models' criteria. By contrast, ASK14 and CY14 concentrate more on PIDR probability density around 0.01, suggesting a tendency towards low inter-story drifts. Notably, CB14 shows a secondary peak near 0.04. In terms of the variance, $GGMM21_{allIMs}$ leads to the highest variability with multi-modal effects in the PIDR. This is expected since the ground motions are selected in a multi-IM-objective manner, leading to ground motions with different traits being selected. Within the spectral only criteria, GGMM21 leads to higher PIDR as compared to the other four.

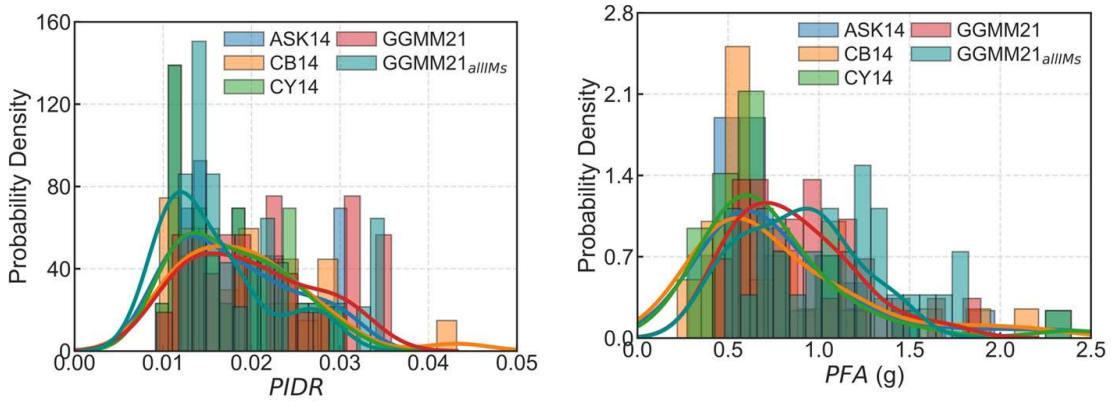


Figure 3: EDPs of 8-story for 2500-year return period: (a) PIDR and (b) PFA

For PFA, $GGMM21_{allIMs}$ similarly show a modest concentration within the 1.0-1.5g range, indicative of slightly higher demand levels. Meanwhile, ASK14, CY14, and CB14 peak near 0.5g which is on a comparatively lower acceleration regime. GGMM21 exhibits no clear spike, implying a more evenly spread probability across various PFAs. Nevertheless, CB14 and CY14 both introduce small spikes above 2.0g and even 2.5g, reflecting a nontrivial probability of very high acceleration. Similar to the PIDR, $GGMM21_{allIMs}$ leads to the highest variability in the PFA. Overall, the EDPs from GGMM21 and $GGMM21_{allIMs}$ shift portions of their probability density towards higher values, though these differences remain subtle compared to the extreme upper bins observed for CB14 (in both PIDR and PFA) and CY14 (in PFA). While these variations may appear minor, even modest differences in the upper tails can significantly impact design decisions.

3 Statistical Comparative Analysis

The results from section 2 are used to compare the hazard curves (for the different periods), target response spectra (for the 10 different hazard levels), and EDPs across the five sets of ground motions and structural responses. The following sections describe the details for each comparison.

3.1 Hazard Curves

The hazard curves are compared through hypothesis testing of surrogate functional forms of the hazard curves for each period. Four-degree polynomials are employed to characterize the hazard curves in log-log space for the four GMMs and all periods. This polynomial form is presented in Equation 6. For each case, the polynomial fits exceptionally well with the coefficient of determination (R^2) greater than 0.99 in most cases.

$$ln(\lambda) = c_4(ln(S_a))^4 + c_3(ln(S_a))^3 + c_2(ln(S_a))^2 + c_1 ln(S_a) + c_0 \quad (6)$$

The coefficients c_0, c_1, c_2, c_3, and c_4 control the shape of the hazard curve. The intercept c_0 shifts the entire curve up or down, indicating baseline IM values when the return period is 1. The linear coefficient c_1 governs the primary slope of the curve, where higher values indicate steeper hazard curves and increased IMs for longer return periods. The quadratic term c_2 influences the curve's curvature; positive values make the curve convex, while negative values make it concave, affecting hazard predictions at intermediate return periods. The cubic coefficient c_3 adjusts inflection points, allowing the curve to capture more complex hazard variations, while the quartic coefficient c_4 fine-tunes the curve at the extreme ends, helping model subtle variations in hazard levels

Figure 4 shows *p*-values resulting from a series of t-tests that compare the coefficients of the fitted fourth-degree polynomial. The hypothesis testing is conducted in a pair-wise manner for each period with the GGMM21 as the base model, and its coefficients are compared against the three GMMs: ASK14, CB14, and CY14. The t-test, performed at a significance level of 0.05, evaluates whether the two hazard curves differ in a statistically significant manner. The figure presents the *p*-values of the hypothesis testing of the coefficients estimated from GGMM21 are compared with those of the other three GMMs. The horizontal axis in the figure represents the period, while the vertical axis shows the examined *p*-values. The five colors – pink, yellow, green, orange, and blue, correspond to the coefficients c_0 through c_4, and the three marker shapes – circles, triangles, and rectangles, represent the comparison GMM: ASK14, CB14 and CY14, respectively. A *p*-value less than 0.05 means that the difference in the polynomial coefficients between the GGMM21 and other GMM is statistically insignificant.

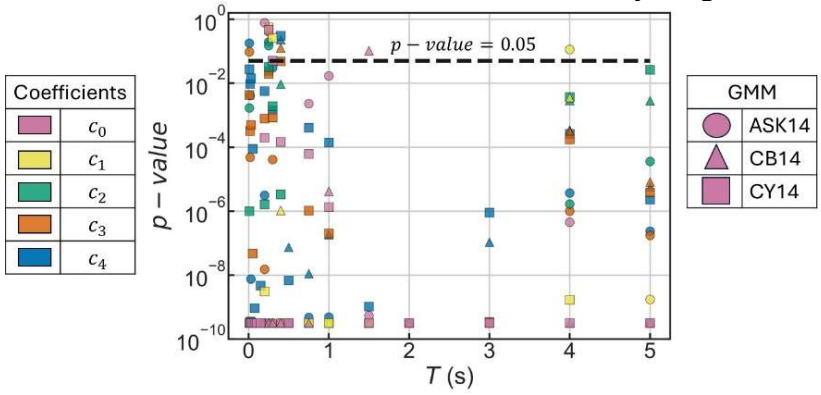


Figure 4: Comparison of hazard curve coefficients through t-test: *p*-values *vs* period

Differences in these coefficients across models (e.g., GGMM21 versus traditional GMMs like ASK14, CB14, and CY14) reflect variations in how each model estimates hazard levels. For example, a higher c_1 in GGMM21 suggests that it predicts faster-increasing hazard levels for longer return periods compared to traditional models. Similarly, significant variations in c_2, c_3, and c_4 indicate differences in how these models represent hazard variability across short to long periods. Examining the *p*-value results, most periods exhibit significant coefficient differences between GGMM21 and the other models, suggesting that GGMM21's polynomial fit diverges systematically in one or more parameters from the other GMMs. However, within the short-period range from PGA to $T = 0.5$s, there is a small subset of coefficients that do not show significance, implying an overlap in the fitted polynomial forms for short periods. These non-significant cases may pertain to any of the five coefficients and any of the three models, indicating no particular bias in which terms align more closely with GGMM21. In practice, this implies that for very short periods, engineers might see less variability among the four models, whereas intermediate-to-long period demands reflect more substantial parameter distinctions.

The systematic differences in hazard curves between GGMM21 and traditional GMMs (ASK14, CB14, and CY14) have direct implications for seismic risk analysis, particularly in quantifying the probability of structural damage or failure under varying seismic scenarios. The polynomial regression analysis revealed that GGMM21 predicts higher S_a for intermediate-to-long periods (e.g., $T =$1s to $T =$3s) at longer return periods (e.g., 1000-year and 3000-year). This divergence suggests that traditional GMMs may underestimate seismic risk for structures with longer natural periods, such as high-rise buildings or long-span bridges, which are more sensitive to these spectral ranges.

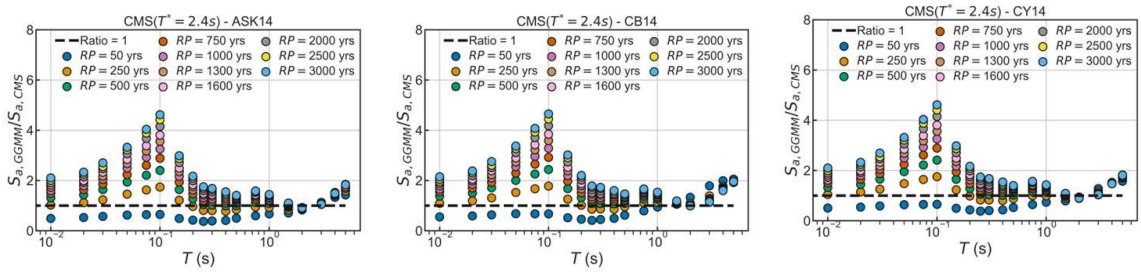
For example, at $T = 2$s, GGMM21 predicts S_a values that are approximately 1.5 to 2 times higher than those predicted by CB14 for return periods exceeding 1000 years. This discrepancy implies that structures designed using traditional GMMs may be under-designed for these periods, leading to higher vulnerability during rare but severe earthquakes. In contrast, for short periods (e.g., $T = 0.1$s to $T = 0.5$s), the hazard curves from GGMM21 and traditional GMMs show closer alignment, suggesting that risk estimates for low-rise structures or rigid systems may be less affected by the choice of GMM.

The higher hazard predictions by GGMM21 for long periods also have implications for risk-based decision-making. For instance, in regions with high seismic activity, such as Los Angeles, the use of GGMM21 could lead to higher insurance premiums or more stringent building code requirements for tall structures. Additionally, the ability of GGMM21 to capture cross-dependencies among IMs (e.g., S_a, I_a, CAV) without post-processing simplifies the risk analysis process, as it eliminates the need for additional steps like the CMS. This could lead to more accurate and efficient risk assessments, particularly for complex structures where multiple IMs influence performance.

3.2 Target Hazard Spectra

The S_a of GMS (from GGMM21) and CMSs from the other three GMMs are compared for the 10-return period across the three conditioned periods. Figure 5 presents the ratios for the

comparison of S_a of GMS (from GGMM21) and CMSs from the other three GMM conditioned at T=2.4s. On the vertical axis, a ratio greater than 1 indicates that GGMM21 yields a higher S_a than the reference model, whereas a ratio less than 1 suggests the opposite case. Different colored markers depict the 10 return periods, from 50 years to 3000 years. Overall, the comparisons reveal major discrepancies across different models, specifically for the short-period domains. The S_a ratios tend to converge to a ratio of 1 at the conditioning period, i.e., 2.4s. The ratios are particularly close to unity for a return period of 500 years at the conditioning period. Across most periods, the ratios remain within a broad band of approximately 0.5 to 1.5. The low hazard levels of 50 and 250 years lead to values close to or lower than 1 indicating lower values from GGMM21 as compared to the other GMMs. As the hazard level increases, the ratio grows higher signifying higher values from GGMM21.



(a) (b) (c)

Figure 5: Ratio between S_a from GGMM (i.e., GMS) and CMS

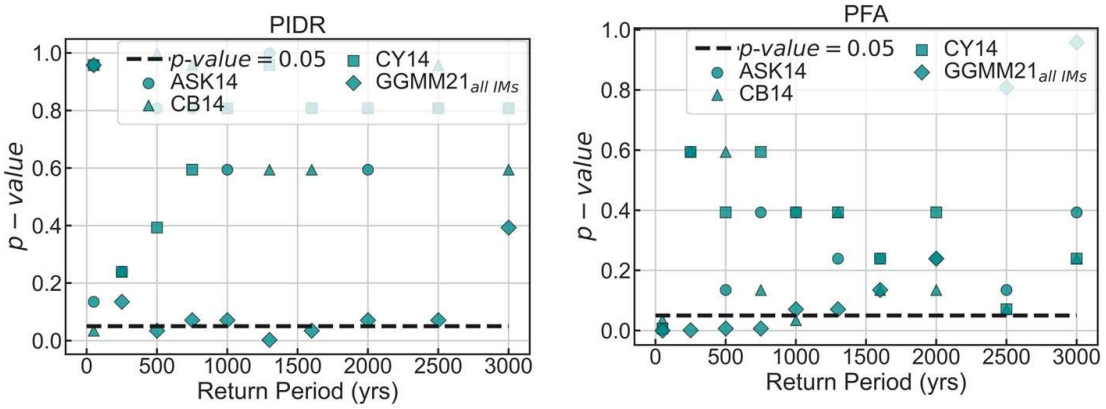
In particular, in the shorter-period range, with a period smaller than 1s, only the 50-year return period curve dips below a ratio of 1, implying GGMM21 is estimating lower S_a, whereas the other return period curves are above a ratio of 1, with the highest ratios observed at a period of 0.1s, which can be due to modeling inaccuracies caused by the high noise in the high-frequency data. Taking 500-year return period for example, its ratio starts near 1.0 at a period of 0.01s and gradually rises to slightly above 2.0 by period of 1.0s. Notably, the ratios peak at around 1sec at a ratio greater than 4 for 3000-year return period, suggesting that away from the conditioning period, GGMM21 can produce substantially higher hazard estimates for certain short-to-intermediate periods. Past the period of 1s, the ratio of 500-year return period tapers from about 1.5 down to 1.0 as the period approaches the conditional period of 2.4s. Beyond the period of 2.4s, all ratio curves again settle near 1.0, indicating lower differences in the high period domain. Due to the lower S_a levels, the conventional CMSs can lead to lower structural designs thereby increasing the structural risk. Similar, results are observed for other cases. Based on the significant accuracy of GGMM21 (as described in the previous studies), it is postulated that the GGMM21 captures off-period hazard more accurately, reflecting limitations of CMSs based on the other GMMs especially when predicting S_a away from their conditional periods.

This suggests that structures designed using traditional GMMs may be under-designed for these critical periods, potentially leading to higher vulnerability during severe earthquakes. If GGMM21 is adopted, designers may need to increase the seismic design forces for these structures, leading to heavier and more robust structural systems. For example, the design base

shear for a building with a fundamental period of 1.2s could increase by 20-30% if GGMM21 is used instead of CB14, depending on the hazard level. This would have cascading effects on material costs, construction practices, and overall project budgets.

3.3 Structural Responses

The EDPs from the five sets of ground motions are compared against each other for each return period. Similar to the previous section, the EDPs from ground motions selected using the GGMM with only S_a as the target, are used as the baseline to make pairwise comparisons against the CMSs of the other three GMMs and the GGMM with all IMs as the target (multi-objective). The comparisons are made through the two-sample Kolmogorov–Smirnov (KS) test, which quantifies the distance between the empirical distribution functions of two EDP sets. The null hypothesis of the test basis that the samples are drawn from the different distributions.



(a) (b)

Figure 6: Comparison of 8-story EDPs under the ground motions selected from GGMM (only using spectra range) against CMSs of other GMMs and the GGMM using all IMs

Figure 6 depicts the *p*-values from t-tests comparing the EDP (PIDR and PFA) distributions obtained from GGMM21 against our other models: ASK14 (circles), CB14 (triangles), CY14 (rectangles), and $GGMM21_{allIMs}$ (diamonds), across the range of the 10 return periods from 50 to 3000 years. For PIDR, the only statistically insignificant differences (p-value < 0.05) arise with CB14 at a return period of 50 years and $GGMM21_{allIMs}$ at a return period of 500, 1300, and 1600 years. In other cases, strong evidence of divergence is detected, indicating significant differences in the EDPs under the ground motions selected from GGMM21 and other GMMs. For PFA, all models exhibit insignificant differences at a return period of 50 years, and $GGMM21_{allIMs}$ also show insignificant differences at the return periods of 250, 500, and 750 years, and CB14 does so at a return period of 1000 years. Outside these highlighted cases, *p*-values remain above 0.05, indicating that GGMM21's PFA predictions are largely different than the other models' distributions for most other return periods. In general, this pattern suggests that the EDP estimates from the five sets of ground motions are distinguished and significantly different, especially for high hazard levels. The higher variability in EDPs observed with

GGMM21, particularly when using all IMs ($GGMM21_{allIMs}$), indicates that this model captures a broader range of ground motion characteristics, leading to more diverse structural responses. This variability is essential for hazard analysis, as it provides a more comprehensive understanding of the potential impacts of different ground motion scenarios.

4 Conclusions

This research presents a comprehensive statistical evaluation of different ground motion models, including data-driven deep-learning models applied to key PBEE steps. The GMMs adopted in this paper include the ASK14, CB14, CY14, and GGMM21. The first three GMMs rely on assumed empirical relationships between earthquake characteristics and $S_a(T)$, treating the prediction of IMs independently and disregarding the internal cross-dependencies among them. To remedy this, the GGMM21 uses LSTM-based RNNs to predict a vector of amplitude, duration-, energy-, and frequency-based IMs while explicitly capturing their cross-dependencies.

The GGMM21 and the empirical GMMs are statistically compared across the hazard curves, target response spectra, and EDPs across the five sets of ground motions and structural responses. The polynomial regression analysis of hazard curves demonstrated a systematic difference between GGMM21 and the traditional GMMs, particularly for intermediate-to-long spectral periods. Hypothesis testing revealed statistically significant coefficient differences in the polynomial fits for most spectral periods, indicating that GGMM21's hazard predictions diverge meaningfully from those of ASK14, CB14, and CY14. However, the differences were less pronounced within the short-period range (PGA to T = 0.5s), with some overlapping coefficients.

On the other hand, the comparison of spectral accelerations between GGMM21-selected ground motion sets and CMS from traditional GMMs revealed substantial variations, especially in the short-period range. The spectral acceleration ratios tended to converge at the conditioning period (e.g., T = 2.4s), with minimal discrepancies observed for return periods of 500 years. However, for lower hazard levels (e.g., 50-year and 250-year return periods), GGMM21 generally predicted lower values than the traditional GMMs, while higher hazard levels (e.g., 1000-year and 3000-year return periods) resulted in larger estimates from GGMM21. These variations highlight the potential influence of GGMMs on design spectra, particularly under high-hazard conditions.

Lastly, based on EDPs, the structural response comparisons further demonstrated significant differences between GGMM21 and traditional GMMs. KS test results showed that, in most cases, GGMM21 produced statistically distinct EDP distributions compared to ASK14, CB14, and CY14, particularly for high-hazard levels. While some statistically insignificant differences were observed at lower hazard levels (e.g., 50-year return period), the overall trend suggests that GGMM21-based ground motions lead to distinct structural responses compared to those selected using conventional GMMs.

The preliminary findings of this study underscore the potential of deep learning-based GGMMs in advancing seismic hazard assessment and structural performance evaluation. More specifically, GGMMs offer greater flexibility in capturing cross-dependent ground motion characteristics and enhancing hazard-consistent ground motion selection. However, given the observed differences in hazard and response predictions, further validation using larger datasets

and additional structural systems is recommended before widespread adoption in engineering practice.

REFERENCES

[1] S. L. Kramer, "Performance-based earthquake engineering: opportunities and implications for geotechnical engineering practice," *Geotechnical earthquake engineering and soil dynamics IV*, pp. 1–32, 2008.
[2] J. Moehle and G. G. Deierlein, "A framework methodology for performance-based earthquake engineering," in *13th world conference on earthquake engineering*, WCEE Vancouver, 2004, p. 12.
[3] N. A. Abrahamson, W. J. Silva, and R. Kamai, "Summary of the ASK14 ground motion relation for active crustal regions," *Earthquake Spectra*, vol. 30, no. 3, pp. 1025–1055, 2014.
[4] K. W. Campbell and Y. Bozorgnia, "NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra," *Earthquake Spectra*, vol. 30, no. 3, pp. 1087–1115, 2014.
[5] B. S.-J. Chiou and R. R. Youngs, "Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra," *Earthquake Spectra*, vol. 30, no. 3, pp. 1117–1153, 2014.
[6] J. Fayaz, Y. Xiang, and F. Zareian, "Generalized ground motion prediction model using hybrid recurrent neural network," *Earthq Eng Struct Dyn*, vol. 50, no. 6, pp. 1539–1561, 2021.
[7] J. Fayaz, M. Medalla, P. Torres-Rodas, and C. Galasso, "A recurrent-neural-network-based generalized ground-motion model for the Chilean subduction seismic environment," *Structural Safety*, vol. 100, p. 102282, 2023.
[8] J. W. Baker and C. Allin Cornell, "Spectral shape, epsilon and record selection," *Earthq Eng Struct Dyn*, vol. 35, no. 9, pp. 1077–1095, 2006.
[9] B. A. Bradley, "A generalized conditional intensity measure approach and holistic ground-motion selection," *Earthq Eng Struct Dyn*, vol. 39, no. 12, pp. 1321–1342, 2010.
[10] J. Dhanya and S. T. G. Raghukanth, "Ground motion prediction model using artificial neural network," *Pure Appl Geophys*, vol. 175, no. 3, pp. 1035–1064, 2018.
[11] T. D. Ancheta *et al.*, "NGA-West2 database," *Earthquake Spectra*, vol. 30, no. 3, pp. 989–1005, 2014.
[12] J. C. Castro, S., Benavente, R., Crempien, J., Candia, G., de la Llera, "A Consistently Processed Strong Motion Database for Chilean Earthquakes.," 2021.
[13] American Society of Civil Engineers, "Minimum design loads for buildings and other structures, American Society of Civil Engineers (ASCE). Reston, VA.," 2016.
[14] American Institute of Steel Construction., "Seismic Provisions for Structural Steel Buildings. ANSI/AISC 341-16.," 2016.
[15] E. H. Field *et al.*, "Uniform California earthquake rupture forecast, version 2 (UCERF 2)," *Bulletin of the Seismological Society of America*, vol. 99, no. 4, pp. 2053–2107, 2009.

[16] N. SimCenter, “An application framework for regional earthquake simulations,” 2019, *NHERI*.
[17] J. W. Baker, “Conditional mean spectrum: Tool for ground-motion selection,” *Journal of Structural Engineering*, vol. 137, no. 3, pp. 322–331, 2011.
[18] J. Fayaz, P. Torres-Rodas, M. Medalla, and F. Naeim, “Assessment of ground motion amplitude scaling using interpretable Gaussian process regression: Application to steel moment frames,” *Earthq Eng Struct Dyn*, 2023.
[19] C. J. Willmott *et al.*, “Statistics for the evaluation and comparison of models,” *J Geophys Res Oceans*, vol. 90, no. C5, pp. 8995–9005, 1985.
[20] C. Kircher *et al.*, “Evaluation of the FEMA P-695 methodology for quantification of building seismic performance factors,” 2010.
[21] American Society of Civil Engineers, “Minimum design loads for buildings and other structures (ASCE). Reston, VA. 2005,” American Society of Civil Engineers, 2005.
[22] American Society of Civil Engineers, *Minimum Design Loads and Associated Criteria for Buildings and Other Structures (ASCE 7-22)*, 22nd ed. Reston, VA, USA: ASCE, 2022.
[23] AISC, *Seismic Provisions for Structural Steel Buildings (ANSI/AISC 341-22)*. Chicago, IL: American Institute of Steel Construction, 2022.
[24] F. McKenna, G. L. Fenves, and M. H. Scott, “OpenSees: Open system for earthquake engineering simulation,” *Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA., http://opensees. berkeley. edu*, 2006.
[25] L. F. Ibarra, R. A. Medina, and H. Krawinkler, “Hysteretic models that incorporate strength and stiffness deterioration,” *Earthq Eng Struct Dyn*, vol. 34, no. 12, pp. 1489–1511, 2005, doi: 10.1002/eqe.495.
[26] NIST and NEHRP, “Guidelines for Nonlinear Structural Analysis for Design of Buildings,” *NIST GCR 17-917-46v2*, p. 145, 2014.