Received: 17 August 2022 | Revised: 12 December 2022 | Accepted: 18 December 2022

DOI: 10.1002/eqe.3807

RESEARCH ARTICLE

WILEY

Hybrid data-driven hazard-consistent drift models for SMRF

Faridah Zahra[1] | **Jorge Macedo**[2] | **Christian Málaga-Chuquitaype**[1]

[1]Department of Civil and Environmental Engineering, Imperial College London, London, UK

[2]School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA

Correspondence
Christian Málaga-Chuquitaype, Department of Civil and Environmental Engineering, Imperial College London, London, UK.
Email: c.malaga@imperial.ac.uk

Abstract

The seismic design and assessment of steel moment resisting frames (SMRFs) rely heavily on drifts. It is unsurprising, therefore, that several simplified methods have been proposed to predict lateral deformations in SMRFs, ranging from the purely mechanics-based to the wholly data-driven, which aim to alleviate the structural engineer's burden of conducting detailed nonlinear analyses either as part of preliminary design iterations or during regional seismic assessments. While many of these methods have been incorporated in design codes or are commonly used in research, they all suffer from a lack of consideration of the causal link between the seismic hazard level and the ground-motion suite used for their formulation. In this paper, we propose hybrid data-driven models that preserve this critical relationship of hazard-consistency. To this end, we assemble a large database of non-linear response history analyses (NRHA) on 24 SMRFs of different structural characteristics. These structural models are subjected to 816 ground-motion records whose occurrence rates and spectral shapes are selected to ensure the hazard consistency of our outputs. Two sites with different seismic hazards are examined to enable comparisons under different seismic demands. An initial examination of the resulting drift hazard curves allows us to re-visit the influence of salient structural modelling assumptions such as plastic resistance, geometric configurations and joint deterioration modelling. This is followed by a machine learning (ML)-guided feature selection process that considers structural and seismic parameters as well as key static response features, hence the hybrid nature of our models. New models for inter-storey drift and roof displacements are then developed. A comparison with currently available formulations highlights the significant levels of overestimation associated with previously proposed non-hazard consistent models.

KEYWORDS

drift demands, earthquake engineering, hazard-consistent, ML, moment-resisting frames, steel

This is an open access article under the terms of the Creative Commons Attribution 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.

1 | INTRODUCTION

Performance-based earthquake engineering (PBEE) methodologies have enabled the design process to hinge around the attainment of specified performance objectives for pre-defined levels of seismic demand. Famously introduced in the SEAOC[1] guidelines, the PBEE philosophy has been adopted in many seismic design codes, such as Eurocode 8[2] and ASCE.[3] The most widely embraced PBEE framework, known as the Pacific Earthquake Engineering Research or PEER–PBEE framework, conceived by Cornel and Krawinkler[4] and contained in FEMA P-58,[5] enables structural designs to be targeted to different stakeholders' performance objectives. In order to verify the satisfaction of those objectives, consequences are expressed in terms that are easily understood by the stakeholder, such as casualties, repair or replacement cost, downtime and environmental impact.[6]

One way to present the expected structural response at a given site within the PBEE framework is by depicting the probability of exceeding various structural response levels during a specified time (usually a year). These curves are known as EDP hazard curves, where EDP stands for engineering demand parameter (EDP), which is a parameter that describes the structural response (e.g., peak deformations, peak accelerations). In the case of steel moment resisting frames (SMRFs), the EDP *par excellence* is the inter-storey drift, or in its absence, the roof drift. This is because SMRFs are flexible structures whose design is usually controlled by serviceability constraints on their allowable deformation. Unsurprisingly, numerous simplified models have been proposed in the scientific literature for the estimation of drifts in SMRFs.

Predictive models of structural drift can generally be divided into: (i) classical mechanics-based, (ii) hybrid and (iii) purely data-driven models.[7] Mechanics-based models, e.g., Refs.[8–10] are derived solely from classical mechanics theory, which makes them easy to generalise and interpret, although they are also often oversimplified and imprecise. On the opposite side of the spectrum are the purely data-driven models, e.g., Refs.[11–13] that connect the model predictors directly to the prediction model target using an advanced statistical or machine learning (ML) algorithm. These data-driven models are especially practical in the absence of an analytical model.[14] However, they use complex methods that are arduous in their interpretation or effectively work as 'black boxes' that can only be tested from the outside. In between these two types of models are the hybrid models, e.g., Refs.[5,15–17] that aim to offer a balance between interpretability and applicability. Hybrid models are initiated by mechanics-based analysis (e.g., modal or linear static analysis) to obtain certain intermediate results. These intermediate results are then connected with the target by linear, multi-linear regression or ML models. Although dealing with a different structural system than the one studied herein and sometimes focused on residual displacements, which is not the focus of this paper, the following studies[18–20] also represent an example of hybrid models.

Most of the models proposed to date have been constructed on the basis of limited ground-motion records scaled up to mimic increasing seismic intensities, or record sets selected to represent a particular type of ground motion (e.g., pulse-like, long-duration). Notwithstanding the limitations of a scalar intensity measure (IM) to describe the ground shaking,[21] another crucial aspect that has been universally neglected in past modelling formulations is the preservation of the causal connection between the hazard level and the set of records employed. Since the ground-motion characteristics change as a function of the earthquake intensity, simply scaling the spectral ordinates would not account for this. A step forward to the most common approaches used in previous efforts, and the one followed herein, is to enable a hazard-consistent selection of ground motion recordings on the basis of which new predictive relationships can be developed.

In this study, we evaluate the inter-storey drift and roof displacement demands of SMRFs within a hazard-consistent framework and provide predictive models for them. To this end, non-linear response history analyses (NRHA) on 24 SMRFs with different structural characteristics are performed. The frames are subjected to 816 ground motions representative of two different sites whose rate of occurrence is estimated using the conditional scenario spectra (CSS) methodology[22] to ensure the desired hazard consistency. The results from these comprehensive NRHA are expressed in the form of EDP hazard curves, and the effects of the modelling approach and key structural characteristics are first examined. Subsequently, we use a number of ML feature selection techniques to inform the choice of structural and seismic parameters to be incorporated into our predictive models. Then, several regression models are trialled to develop predictive relationships for inter-storey drift and roof displacement. Finally, our models are compared with currently available predictive models, and the significant differences between their estimations are quantified.

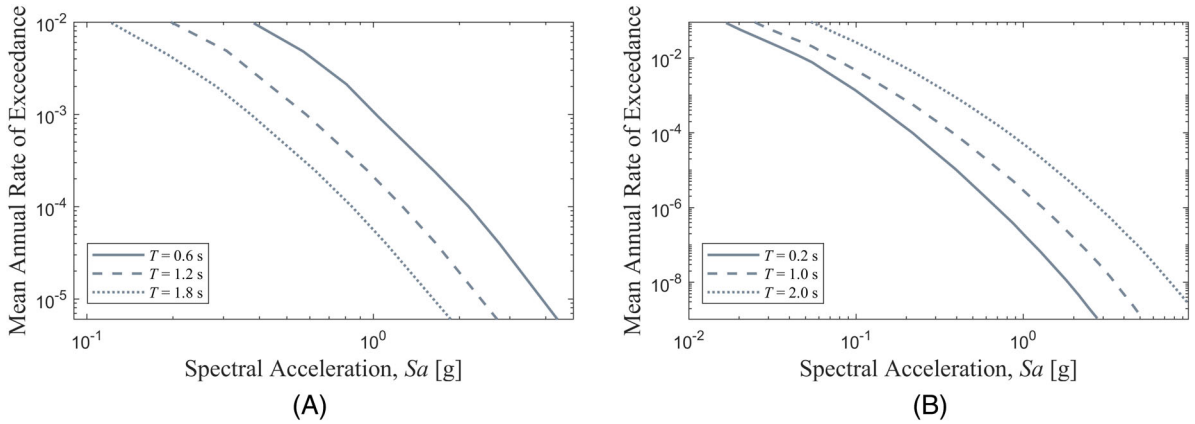


FIGURE 1 Hazard curves of (A) Oakland and (B) Sonora sites at different periods from the PSHA

2 | OVERVIEW OF HAZARD-CONSISTENT SEISMIC DEMAND ANALYSIS

Various frameworks have been proposed for the estimation of the probability of exceedance of EDPs as a function of ground-motion intensity. However, most available studies have achieved this by means of incremental dynamic analyses (IDA),[23] which suffer from the underlying assumption that a single set of ground motions scaled to achieve various IM levels is representative of the change in ground-motion characteristics with earthquake intensity. Consequently, IDA result in rather unrealistic ground motion spectral shapes that may lead to correspondingly unrealistic structural responses that are not hazard-consistent.[22] The effects of this can be quantified by comparing the outputs of previous models generated based on IDAs against the hazard-consistent estimations produced herein, as examined later in this paper.

An alternative method, and the one adopted in this study, is to obtain the EDP using a set of ground motions whose rates of occurrence are determined from CSS as proposed by Abrahamson et al.[22,24,25] The CSS presents a set of realistic earthquake spectra with an assigned rate of occurrences based on their spectral shape and intensity. Hence, the ground motion spectra in the CSS are developed to be spectrally consistent with the hazard at the selected site at all relevant periods.

To develop the CSS, we have selected ground motions from hazard disaggregation at two sites and have estimated the target uniform hazard spectra (UHS) at several hazard levels (i.e., several exceedance rates). For each hazard level, the ground motions are scaled to account for the variability of the peaks and troughs around the conditional mean spectra (CMS). The CMS ensure that each spectrum has the correct shape and provides the geometric mean response spectrum for all periods of interest.[26,27] Each ground motion is then assigned an initial rate of occurrence according to the hazard level of the UHS at the CMS conditioning period. Finally, the rate of occurrence given to each ground motion is numerically optimised such that its calculated hazard matches the target hazard curve.

The CSS method has been previously used to evaluate the drift and base shear demands in RC frames by Arteta and Abrahamson[22,25] and to assess the non-linear truss model capability to simulate the seismic behaviour of RC walls by Arteta et al.[28] The CSS was also employed to perform fragility analysis of concrete shear walls[29] and seismic risk assessment of ground-mounted rigid sliding equipment and contents.[30] However, to date, the CSS has not been applied to steel structures.

In this study, we use CSS to generate a database for the training of our drift predictive models for SMRFs, as will be described later in this paper. To this end, site-specific probabilistic seismic hazard analyses (PSHA) were carried out for Oakland (37.803 °N, 122.287 °W) and Sonora (38.158 °N, 120.000 °W), both in California. The hazard curves obtained are presented in Figure 1. One of the main features of the CSS is that the hazard at a site can be recovered over a whole range of periods by using a suite of records with specific rates of occurrence. The target and reconstructed CSS spectra for different hazard levels are compared in Figure 2, where they show a good match. This figure also shows the Eurocode 8 Type 1 Spectrum for Soil Class B matching well the high-hazard site spectrum for a return period, T_R, of 500 years. This spectra compatibility implies the similarity of the demand at the site and the Eurocode 8[2] provisions, and justifies the selection of these spectra to test the Eurocode 8 design assumptions that follow in this paper. It can be appreciated from these figures that the two selected sites have different seismicity levels. The Oakland site has the highest seismicity of the two with a peak ground acceleration (PGA) for $T_R = 2475$ years at 0.835 g, which is around 3.5 times the corresponding PGA in Sonora.

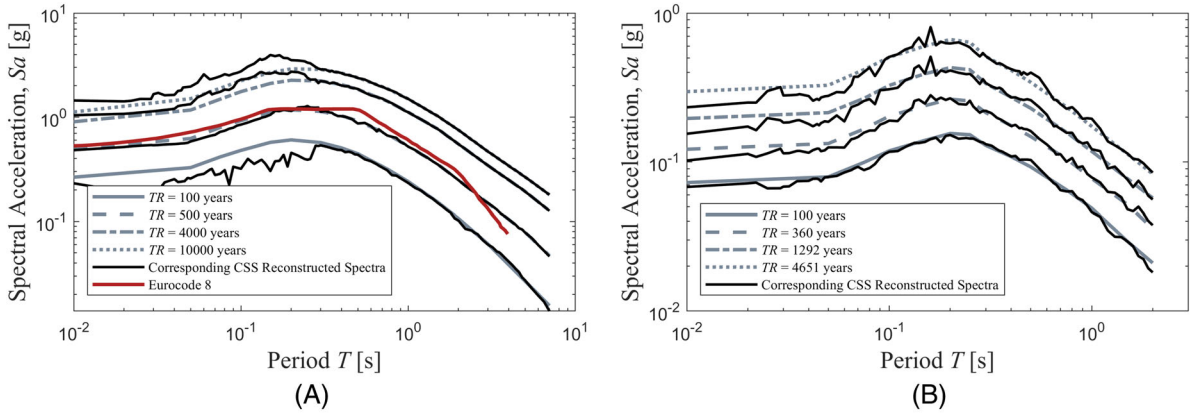


FIGURE 2 Comparison of target and CSS re-constructed spectra of (A) Oakland and (B) Sonora sites

TABLE 1 SMRFs considered

Frame ID	n_s	n_b	Sections
1	3	3	240-330(1-3)
2	3	3	260-330(1-3)
3	3	3	280-330(1-3)
4	3	6	240-330(1-3)
5	3	6	260-330(1-3)
6	3	6	280-330(1-3)
7	6	3	280-360(1-4)+260-330(5-6)
8	6	3	300-360(1-4)+280-330(5-6)
9	6	3	320-360(1-4)+300-330(5-6)
10	6	6	280-360(1-4)+260-330(5-6)
11	6	6	300-360(1-4)+280-330(5-6)
12	6	6	320-360(1-4)+300-330(5-6)
13	9	3	340-360(1)+340-400(2-5)+320-360(6-7)+300-330(8-9)
14	9	3	360-360(1)+360-400(2-5)+340-360(6-7)+320-330(8-9)
15	9	3	400-360(1)+400-400(2-5)+360-360(6-7)+340-330(8-9)
16	9	6	340-360(1)+340-400(2-5)+320-360(6-7)+300-330(8-9)
17	9	6	360-360(1)+360-400(2-5)+340-360(6-7)+320-330(8-9)
18	9	6	400-360(1)+400-400(2-5)+360-360(6-7)+340-330(8-9)
19	12	3	400-360(1)+400-400(2-3)+400-450(4-5)+360-400(6-7)+340-400(8-9)+340-360(10)+340-330(11-12)
20	12	3	450-360(1)+450-400(2-3)+450-450(4-5)+400-450(6-7)+360-400(8-9)+360-360(10)+360-330(11-12)
21	12	3	500-360(1)+500-400(2-3)+500-450(4-5)+450-450(6-7)+400-400(8-9)+400-360(10-11)+400-330(12)
22	12	6	400-360(1)+400-400(2-3)+400-450(4-5)+360-400(6-7)+340-400(8-9)+340-360(10)+340-330(11-12)
23	12	6	450-360(1)+450-400(2-3)+450-450(4-5)+400-450(6-7)+360-400(8-9)+360-360(10)+360-330(11-12)
24	12	6	500-360(1)+500-400(2-3)+500-450(4-5)+450-450(6-7)+400-400(8-9)+400-360(10-11)+400-330(12)

3 | DATABASE OF SMRF MODELS AND GROUND-MOTION RECORDS

3.1 | SMRF models

Twenty-four plane SMRFs are considered in this study. The building archetypes used are based on the frame designs by Karavasilis et al.[16] based on Eurocode 3[31] and Eurocode 8[2] provisions. Different sections, number of storeys (n_s), all 3-m height, and number of bays (n_b), all of 5-m width are considered as shown in Table 1. A typical configuration of an SMRF is shown in Figure 3. HEB sections are used for the columns, and IPE sections for the beams. The expressions

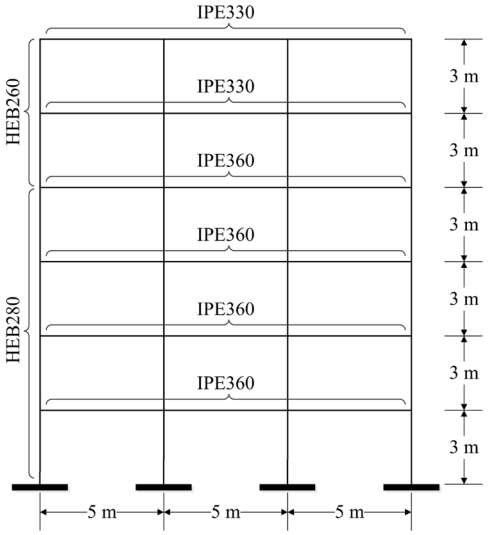


FIGURE 3 Dimension and section of a typical SMRF, Frame 07 (6S-3B) (280-360(1-4)+260-330(5-6)), employed in this study

under the heading 'Sections' in Table 1 follow the nomenclature 'column–beam(storeys)', for example, 280-360(1-4)+260-330(5-6) means that floors 1–4 have HEB280 columns and IPE360 beams and floors 5–6 have HEB260 columns and IPE330 beams. The gravity load on the beams is set to 27.5 kN/m. It should be noted that although the steel frames are designed according to the Eurocodes by Karavasilis et al.,[16] different configurations of steel sections are considered in order to obtain a family of frames that covers the more flexible and more rigid ends of the spectrum. This is the reason why three different designs are used for each unique set of n_s and n_b. The more flexible frames can be regarded as the product of lower seismicity levels or more relaxed values of serviceability drift limits, while stiffer frames can be associated with higher seismic design demands.

The frames were analysed using the open-source software OpenSees.[32] The uniaxial bilinear *Steel01* material model was used. The beams and columns were modelled with *force-based* elements comprising four integration points along their length and sections discretised in around 20–40 fibres along the flange and web, depending on their section size. The shear compliance of the panel zone was modelled using the scissors model.[33] Geometric nonlinearities were incorporated by means of the *corotational transformation*. The non-linear behaviour of the joints was represented by rotational springs added via *zero-length* elements adopting the *modified-Ibarra–Krawinkler deterioration model*[34] with the parameters determined in accordance with the empirical relationships proposed by Lignos and Krawinkler.[35]

Table 2 shows several structural characteristics that describe the dynamic behaviour of the frames obtained from eigenvalue and pushover analyses like the fundamental period (T_1), the stiffness ratio (ρ), plasticity resistance ratio (α), relative storey stiffness ratio (β_3), first mode participation factor (γ) and global ductility parameter (μ).

The stiffness ratio (ρ) controls the frame behaviour in the elastic range,[16] and is calculated as $\rho = \frac{\sum(I/l)_b}{\sum(I/l)_c}$ where I and l are the second moment of inertia and the length of the beam (b) and column (c), respectively. As ρ increases, the behaviour of the structure changes from flexural-dominated to shear-dominated. The influence of ρ diminishes as the structure moves into the inelastic range and approaches the formation of a mechanism.[16] The plasticity resistance ratio (α) is used to describe the characteristic of the structure in the inelastic range, it is defined as $\alpha = \frac{M_{RC,1,av}}{M_{RB,av}}$, where $M_{RC,1,av}$ is the average of the first storey column sections plastic moments and $M_{RB,av}$ is the average of the beam sections plastic moments. For a certain level of ductility and beam cross-section, a higher value of α will delay the mechanism formation.

The relative storey stiffness ratio (β) accounts for the variation in the stiffness or strength in the upper storeys. This parameter may influence the response of the structure since relatively flexible or weak upper storeys may result in early yielding and relatively high roof drift demand due to higher mode effects. β is calculated using the inter-storey drift profile corresponding to the first mode shape of the frame obtained from eigenvalue analysis. The relative storey stiffness ratio used in this study is the β_3, which is the ratio of the maximum drift of the upper half of the frame to the maximum drift of the lower half of the frame.[17]

The first mode participation factor (γ) is also used to account for higher mode effects in the drift response. A higher γ indicates less significant higher mode effects in the structural response. The global ductility parameter (μ) is calculated as

TABLE 2 Structural characteristics of the SMRFs under study

Frame ID	T_1	ρ	α	β_3	γ	Δ_u	Δ_y	μ
1 (3S-3B)	0.678	0.470	1.300	0.649	0.851	0.442	0.116	3.809
2 (3S-3B)	0.635	0.360	1.600	0.675	0.843	0.445	0.128	3.489
3 (3S-3B)	0.601	0.280	1.900	0.702	0.835	0.535	0.129	4.152
4 (3S-6B)	0.695	0.540	1.300	0.637	0.856	0.420	0.104	4.053
5 (3S-6B)	0.650	0.410	1.600	0.660	0.848	0.430	0.117	3.664
6 (3S-6B)	0.614	0.310	1.900	0.684	0.840	0.483	0.129	3.746
7 (6S-3B)	1.124	0.380	1.600	0.793	0.821	0.674	0.177	3.810
8 (6S-3B)	1.075	0.290	1.970	0.840	0.802	0.735	0.214	3.443
9 (6S-3B)	1.038	0.240	2.270	0.938	0.789	0.759	0.224	3.383
10 (6S-6B)	1.135	0.430	1.600	0.890	0.803	0.567	0.186	3.051
11 (6S-6B)	1.082	0.330	1.970	0.917	0.797	0.571	0.195	2.935
12 (6S-6B)	1.048	0.270	2.270	0.935	0.792	0.582	0.200	2.908
13 (9S-3B)	1.414	0.280	2.190	0.828	0.794	0.960	0.248	3.877
14 (9S-3B)	1.379	0.240	2.430	0.844	0.791	0.972	0.256	3.794
15 (9S-3B)	1.324	0.180	2.930	0.895	0.783	0.980	0.270	3.633
16 (9S-6B)	1.423	0.320	2.190	0.794	0.800	0.829	0.235	3.525
17 (9S-6B)	1.381	0.280	2.430	0.812	0.796	0.861	0.242	3.559
18 (9S-6B)	1.322	0.210	2.930	0.865	0.788	1.008	0.248	4.058
19 (12S-3B)	1.710	0.240	2.600	0.858	0.791	1.187	0.395	3.002
20 (12S-3B)	1.599	0.260	3.000	0.726	0.793	1.315	0.409	3.218
21 (12S-3B)	1.527	0.190	3.630	0.738	0.789	1.435	0.412	3.485
22 (12S-6B)	1.690	0.280	2.600	0.811	0.799	1.088	0.359	3.034
23 (12S-6B)	1.574	0.300	3.000	0.691	0.801	1.193	0.380	3.137
24 (12S-6B)	1.508	0.220	3.630	0.693	0.798	1.295	0.390	3.317

the ratio of ultimate displacement (Δ_u) to the yield displacement (Δ_y) at the top of the structure,[36] $\mu = \frac{\Delta_u}{\Delta_y}$. The ultimate displacement (Δ_u) is measured at the instant of the formation of the collapse mechanism, while the yield displacement (Δ_y) indicates the point beyond which the structural response is strongly non-linear. Currently, there is no universally accepted definition of Δ_u and Δ_y, which has led to a variety of interpretations of global ductility assessment results for MRFs. The definitions adopted herein are the ones in Annex B of Eurocode 8,[2] which describes (Δ_u) as the displacement where the plastic mechanism occurs and provides an equation to calculate the (Δ_y) as a function of deformation energy and yield force.

3.2 | Ground-motion records

596 ground motions for the Oakland site and 220 ground motions for the Sonora site were selected from the PEER NGA-West 2 database[37] based on the disaggregation of the PSHA. The difference in the number of ground motions required at each site is a reflection of the lower variability associated with the Sonora site in comparison with Oakland, hence requiring fewer records to obtain comparably stable mean estimates. This is because of the higher seismicity at the Oakland site, then requiring a larger number of ground motions within the CSS framework to reproduce the seismic demand across a range of periods. Figure 4 presents the associated horizontal-component ground motion spectra where the rate of occurrence of each spectrum was calculated using the CSS methodology.

Figure 5 presents a distribution plot matrix of key parameters of the ground-motion database employed, including the *PGA*, peak ground velocity (*PGV*), significant duration (D_{5-95}), ratio between the period of the building and mean period of the ground motion (T_1 /T_m) and the spectral acceleration at the fundamental period ($Sa(T_1)$) for all buildings and ground-motions considered. The kernel density estimate plots in Figure 5 show the probability density function of one

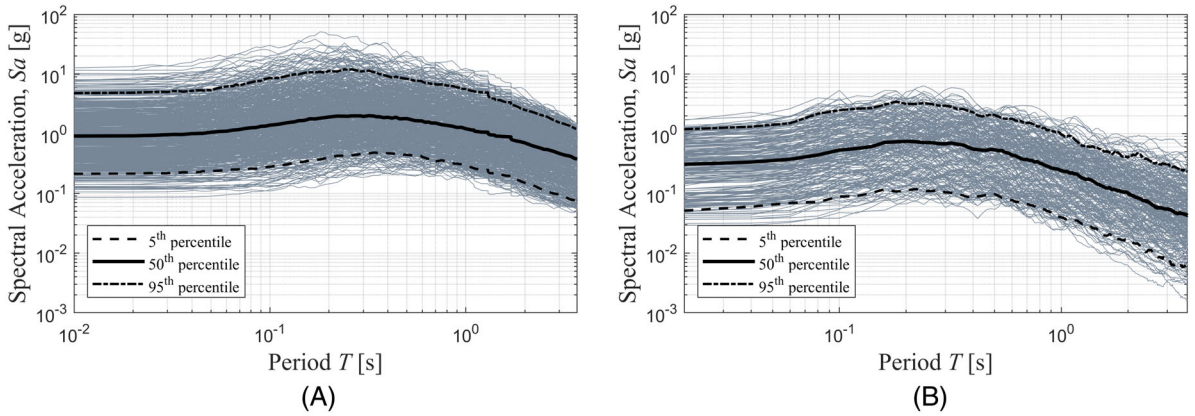


FIGURE 4 Scenario spectra (5% damped) of (A) Oakland and (B) Sonora sites

parameter as a function of another, while the associated histograms are shown in the diagonal. It can be seen from Figure 5 that the dataset employed has a generally normal and uniform distribution of parameters.

4 | PROBABILISTIC SEISMIC DEMAND ANALYSIS

The 24 frames described before were subjected to the 816 ground motions from both sites and analysed using Opensees[32] on the high-performance computing facility at Imperial College London.[38] These 19,584 analyses resulted in an extensive dataset of structural responses that ranged from mildly elastic to highly inelastic. These results are used to develop the EDP hazard curves, as shown in Figure 6. In this figure, the maximum inter-storey drifts of Frame 03 (3S-3B) are plotted against the rate of occurrence of the corresponding ground motions. The EDP annual rate of exceedance, $\nu[EDP > d]$, is calculated using the following expression[22]:

$$\nu[EDP > d] = \sum_{i=1}^{\#records} Rate_{\mathrm{CSS},i} H[EDP - d] \tag{1}$$

where $Rate_{\mathrm{CSS},i}$ is the assigned rate of occurrence of the ground motions obtained from the CSS, and $H[EDP - d]$ is the Heaviside function.

As an example of typical results, the inter-storey drift hazard curve of Frame 03 (3S-3B) is shown in Figure 6B for Oakland. The annual rates of exceedance of 1/475 (life safety) and 1/2475 (collapse prevention) are also shown in this figure to facilitate interpretation. The corresponding maximum inter-storey drift of Frame 03 associated with the life safety and collapse prevention performance levels are 1.38 and 1.98%, respectively. These values are well below the performance level objective for conventional structures based on FEMA 356,[39] which is 2.5% transient drift for life safety and 5% for collapse prevention performance level. By contrast, the current version of Eurocode 8[2] only mentions a drift limit for immediate occupancy performance level ('damage limitation state') at 1% drift.

The process of developing the hazard curve was repeated for every frame at both Oakland and Sonora sites. These hazard-consistent representations of structural response can be used to compare the effects of key structural and non-structural characteristics over the whole range of responses and annual rates. Below, these comparisons are offered for site seismicity, joint deterioration and geometrical structural configurations.

4.1 | Influence of site seismicity

Figure 7 shows that the structural demands of the Oakland site are remarkably higher than those for the Sonora site, as mentioned in Sections 2 and 3. While the same frame will experience a 2.6% drift for 2475 years return period in Oakland, it will experience 0.5% in Sonora, both of which are well below the code requirement. This significant difference in seismic demands between the different sites also emphasises that site-specific performance assessments are needed in order to

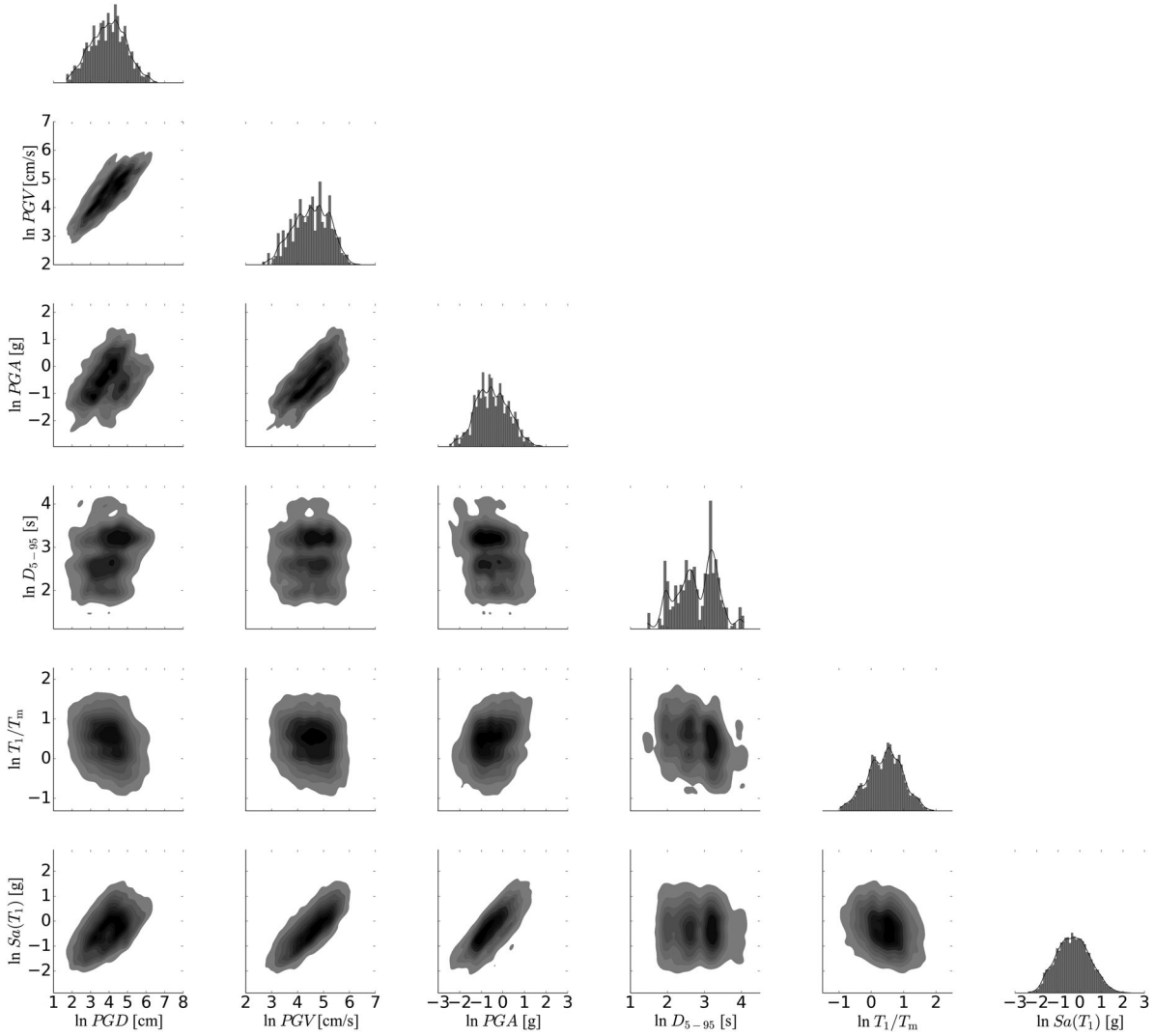


FIGURE 5 Distribution plot matrix of the key record characteristics of the ground-motion dataset employed

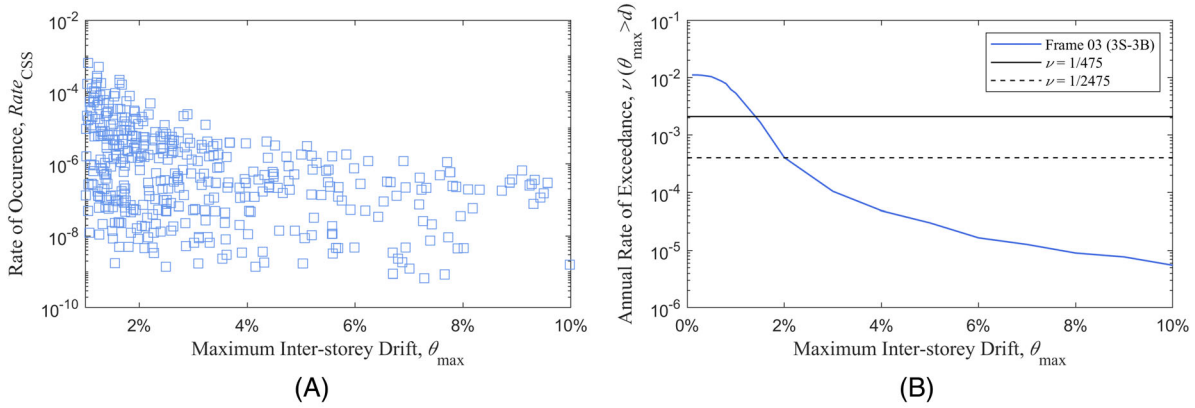


FIGURE 6 (A) Scatter plot of inter-storey drift rate of occurrence based on CSS of Frame 03 (3S-3B) and (B) its hazard curve at Oakland site

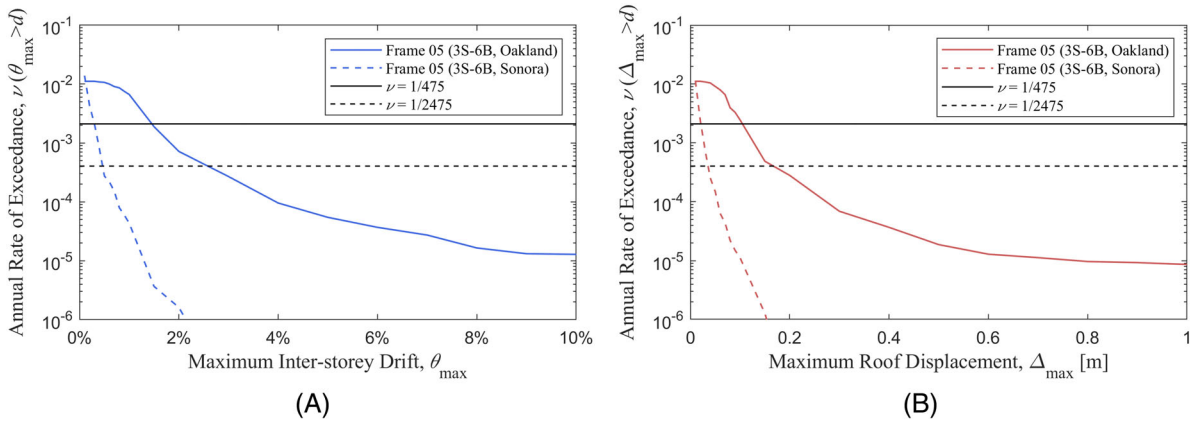


FIGURE 7 Influence of site location on (A) inter-storey drift and (B) roof displacement hazard curve of Frame 05 (3S-6B)

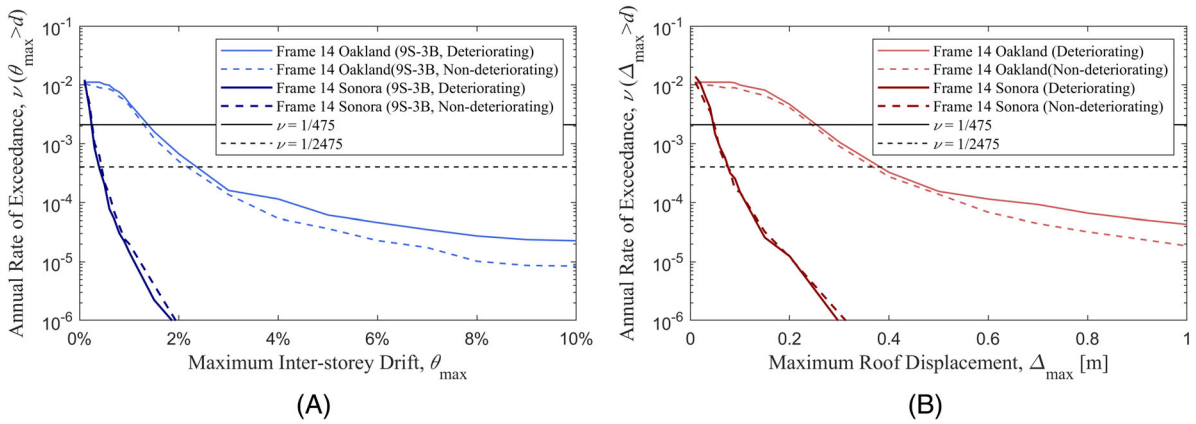


FIGURE 8 Influence of joint deterioration model on (A) inter-storey drift and (B) roof displacement hazard curve of Frame 14 (9S-3B)

ensure economical designs. This aspect will be discussed further during the development of the predictive models later in this paper.

4.2 | Influence of joint deterioration modelling

Although simplified non-deteriorating models have been shown to predict reasonably well mean global structural responses, for example,[40,41] a deterioration model that captures the strength and stiffness degradation of the structural component is generally expected to be more reliable even at lower levels of earthquake demand,[42] and there have been recent calls to use the higher levels of model sophistication even for design level demand assessments.[43]

To examine the accuracy of these claims, alternative models without deterioration were assessed, and their outputs were compared with the deteriorating ones. The difference in their estimations is shown in Figure 8. In general, the deteriorating frames exhibit slightly higher inter-storey drift demand than the non-deteriorating ones, although this difference is generally more noticeable at higher hazard levels ($\nu < 1/2475$). At the lower hazard level, the deteriorating frames appear to predict similar or sometimes even lower demands. Accordingly, the Sonora site, which has significantly lower seismicity than Oakland, shows a very small difference between deteriorating and non-deteriorating models at all relevant hazard levels. It follows from the previous discussion that site seismicity can inform the need to incorporate expensive deterioration effects in beams and columns. This also stresses the possibility of avoiding the complexities of deterioration modelling for assessments that are not intended to evaluate collapse scenarios, a fact that was already highlighted by Krawinkler.[44]

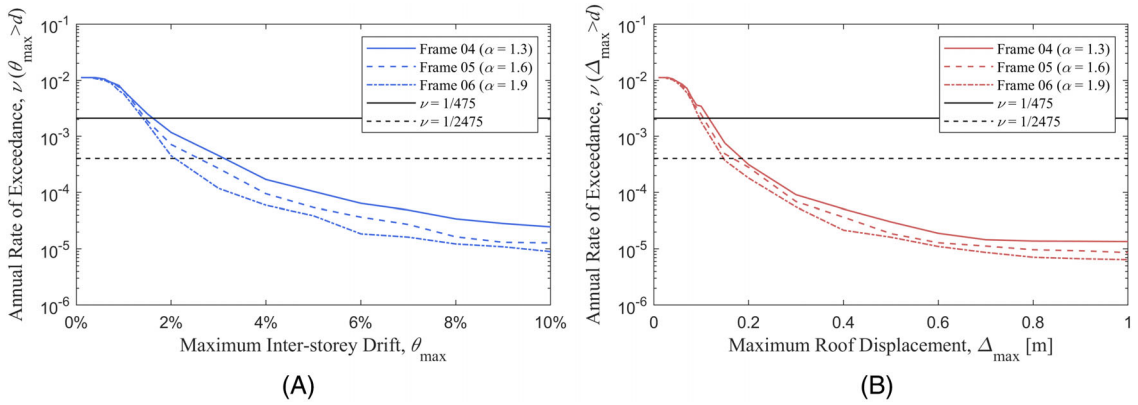


FIGURE 9 Influence of plasticity resistance ratio, α, on (A) inter-storey drift and (B) roof displacement hazard curve

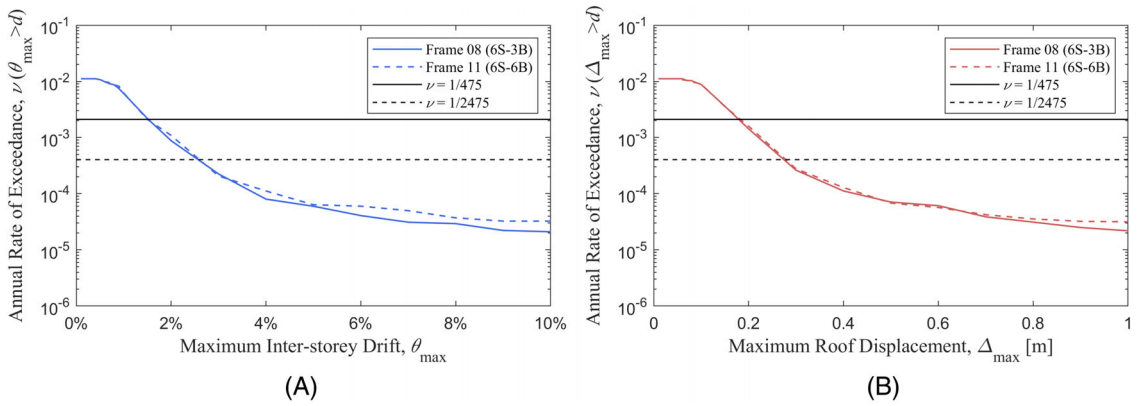


FIGURE 10 Influence of the number of bays on (A) inter-storey drift and (B) roof displacement hazard curve

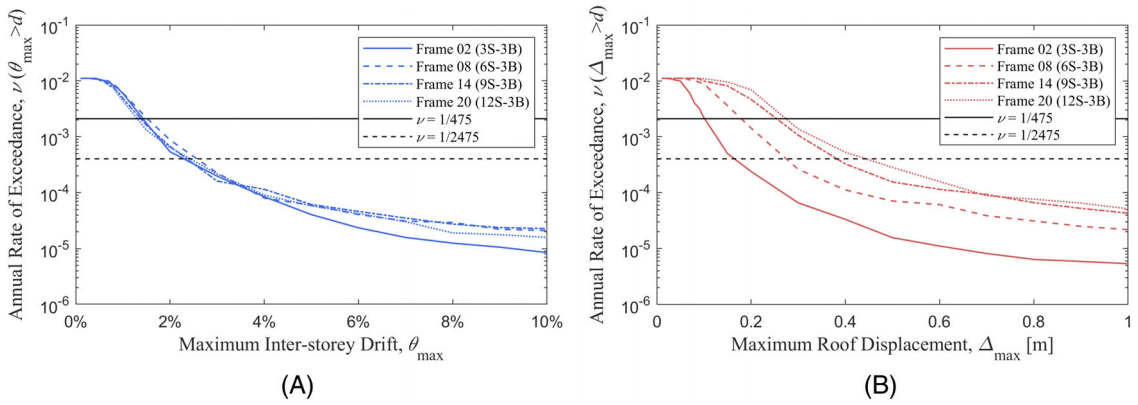


FIGURE 11 Influence of number of storeys on (A) inter-storey drift and (B) roof displacement hazard curve

4.3 | Influence of structural configuration

The influence of different structural configurations on the inter-storey drift and roof displacement hazard curves can be discussed with reference to Figures 9–11, taking the Oakland site as an example. In Figure 9, frames with a higher plasticity resistance ratio, α, have noticeable lower inter-storey drift and roof displacement demand for the same hazard level as the high α delays the formation of plastic hinges. It can be seen that a 0.3% increase in α results in a 0.5% increase in inter-storey drift demand at a 2475 year return period rate. The hazard curves also show the influence of the number of bays in Figure 10A,B, where frames with a different number of bays show similar demands up to 2475 years return period,

although a higher number of bays are associated with slightly higher demand at higher hazard level with 0.4% difference at 10,000 years return period. This finding on the influence of the number of bays defies the commonly accepted perception that a larger number of bays is conducive to a better seismic response of structures.[45,46] Last, Figure 11A shows almost no difference in inter-storey drift demands for different number of storeys at both 475 and 2475 years return period. Although the difference in the demand is more noticeable in higher hazard levels, there is no distinguished pattern of the influence of the number of storeys in the inter-storey drift demand. On the other hand, the influence of the number of storeys becomes significant in the roof displacement demand in Figure 11B as the roof displacement is obviously higher for taller frames. These results show that the plasticity resistance ratio influences the inter-storey drift more than the number of bays and storeys.

All the hazard curves obtained in this study indicate that code-compliant SMRFs are safely within the required performance level objectives of conventional structures based on FEMA 356,[39] which are 2.5% transient drift for life safety and 5% for collapse prevention. This is especially true of the Sonora site, where the low seismicity results in considerably low and uniform structural drifts. Besides, Figures 7–11 show that the inter-storey drift demands of all frames at the Oakland are bounded at around 0.04% for immediate occupancy performance level ($\nu = 1/95$). However, there is a 70 and 53% variation in the inter-storey drift demands for life safety and collapse prevention level, respectively, at the Oakland site. This increased variation at higher hazard levels, where the responses are generally inelastic, indicates that the inelastic drifts could be controlled by specific characteristics of ground motions beyond spectral content that are not well incorporated in the elastic spectral shape used in this study.[21,42] Besides, the drift plots presented above illustrate in a striking manner the inability of Eurocode 8 to ensure a consistent exceedance level for a given EDP, especially at a lower hazard. Arguably, the better the code, the more it will result in well-controlled structural designs, where similar risk levels are expected on all structures depending on their function.

5 | MACHINE-LEARNING-BASED DRIFT MODELS

The first important step towards the formulation of the new predictive models proposed in this study is feature selection, which is conducted herein using several ML algorithms. Feature selection aims to carefully pick the features that would improve the accuracy of the model and reduce its complexity, increase training efficiency and avoid overfitting. Feature selection methods can be classified into filter, wrapper and embedded methods. Filter methods are generally used to pre-process the features by assigning a correlation score between them and the target variable. Wrapper methods train subsets of features by recursively adding or removing them to select the best-performing subset based on feedback from the ML algorithm. Last, embedded methods combine the advantages of filter and wrapper methods by performing feature selection and training of the algorithm in parallel rather than through a discrete process where the features are either kept or discarded. All three methods are used in this study.

A preliminary assessment was carried out by filtering features on the basis of their Pearson's correlation coefficient, which quantifies linear dependencies between each feature and the response variable of interest. Then, three wrapper methods are used: Forward Selection, Backward Elimination and Exhaustive Feature Selection, all carried out by means of the MLXtend Python Library.[47] In the case of Forward Selection, the model is started by fitting one feature at a time and selecting the best-performing one. The remaining features are then added to the model, and the best-performing pair is chosen. This process is repeated until the best-performing feature set with the desired number of features is obtained. Conversely, in Backward Elimination, the model starts with all the features, and the least significant feature is eliminated at each iteration. On the other hand, Exhaustive Feature Selection tries every possible combination of features, which makes it computationally expensive.

In terms of Embedded Methods, Random Forest[48] was used in this study by means of the Scikit-Learn Python Library.[49] Random Forest builds decision trees over a random extraction of the observations from the dataset and a random feature extraction. The random forest algorithm calculates the relative importance of each feature based on the mean decrease in impurity (or gini importance). Thus, the algorithm measures how effective the feature is at reducing uncertainty when creating the decision trees. However, this *default* type of feature importance (as opposed to *permutation* or *drop column* types) has a known limitation in that it tends to escalate the importance of continuous or high-cardinality features, so that it can be unreliable for datasets whose features vary in the scale of measurement.[50,51] To overcome this limitation, permutation importance[52,53] is used here. This method permutes a specific column and measures the decrease in the accuracy of the overall regressor. The random forest with permutation importance was carried out using the Rfpimp Python Library[54] within Scikit-learn. However, Strobl et al.[51] showed that permutation importance overestimates the importance of

TABLE 3 Structural and ground motion parameters considered in this study

No	Features	Definition
1	T_1	Fundamental period of structure
2	γ	First mode mass participation factor
3	q	Behaviour factor of structure
4	β_3	Relative storey stiffness ratio[17]
5	Δ_u	Ultimate displacement of structure
6	$\Delta_{1,roof}$	Roof displacement at the first yield
7	$\theta_{1,max}$	Maximum inter-storey drift at the first yield
8	PGD	Peak ground displacement
9	PGV	Peak ground velocity
10	PGA	Peak ground acceleration
11	CAV	Cumulative absolute velocity.[57] $CAV = \int_0^{t_{tot}} a(t)\,dt$
12	AI	Arias intensity.[58] $AI = (\frac{\pi}{2g}) \int_0^{t_{tot}} a^2(t)\,dt$
13	D_{5-75}	Significant duration for 5%–75% Arias intensity[59]
14	D_{5-95}	Significant duration for 5%–95% Arias intensity[59]
15	T_m	Mean period of ground motion.[60] $T_m = \frac{\sum_i C_i^2 (1/f_i)}{\sum_i C_i^2}$ C_i = Fourier amplitude coefficient corresponding to a frequency (f_i) obtained from a discrete Fast Fourier Transform (FFT)
16	T_1/T_m	Period ratio
17	$Sa(0.2T_1)$	Spectral acceleration at $0.2T_1$ s
18	$Sa(0.5T_1)$	Spectral acceleration at $0.5T_1$ s
19	$Sa(1.5T_1)$	Spectral acceleration at $1.5T_1$ s
20	$Sa(1.8T_1)$	Spectral acceleration at $1.8T_1$ s
21	$Sa(2.0T_1)$	Spectral acceleration at $2.0T_1$ s
22	$Sa(2.5T_1)$	Spectral acceleration at $2.5T_1$ s
23	$Sa(3.0T_1)$	Spectral acceleration at $3.0T_1$ s
24	Sa_{avg}	Average spectral acceleration[61] $Sa_{avg}(c_1T_1, \ldots, c_NT_1) = (\prod_{i=1}^{N} Sa(c_iT_i)^{1/N})$ N = the number of spectral ordinates considered in the averaging process; $c_1 = 0.2$; $c_N = 3$
25	$Sa(T_1)$	Spectral acceleration at T_1 s

correlated predictor variables. Last, the most computationally expensive random forest feature importance is drop-column importance, which defines the importance of a feature as the difference between a model with all features (the baseline) and a model with the feature dropped using a given accuracy metric.[50,55]

As defined by Guan et al.,[7] a hybrid predictive model requires the user to perform some type of mechanics-based analysis to generate intermediate results. Subsequently, those intermediate results are used within a ML framework to develop a function approximator for the final target parameter, in our case, the drift. Hybrid models represent a good balance between the purely data-driven models, whose interpretation is difficult or even wholly concealed, and the solely mechanistic models, whose accuracy can be low. In this context, the features used in the predictive model formulated in this study can be divided into structural (some of which demand some form of structural analysis) and ground motion parameters. The initial set of features explored is listed in Table 3. More specifically, the predictive models developed below are constructed in terms of a maximum drift modification factor, θ_{mod}, and a global drift modification factor, Δ_{mod} that can be written as

$$\theta_{mod} = \frac{\theta_{max}}{q \times \theta_{1,max}} \tag{2}$$

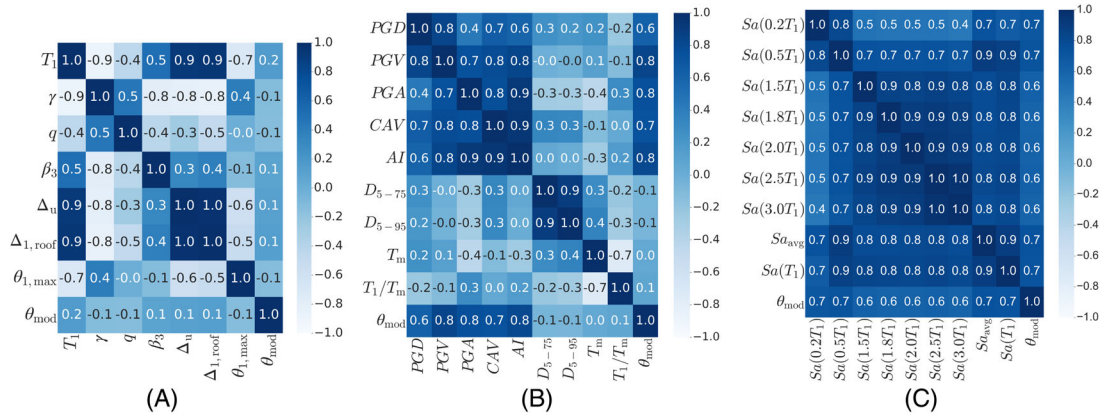


FIGURE 12 Heatmaps of the correlation coefficient between (A) structural parameters, (B) ground motion parameters and (C) ground motion spectral accelerations and θ_{mod} at the Oakland site

and

$$\Delta_{\text{mod}} = \frac{\Delta_{\text{max}}}{q \times \Delta_{1,\text{roof}}} \tag{3}$$

where θ_{max} is the inter-storey drift for a given behaviour factor, q, and $\theta_{1,\text{max}}$ is the maximum inter-storey drift at the first yield obtained from pushover analysis with a first-mode lateral load profile. The behaviour factor, q, was estimated by means of the expressions proposed by Karavasilis et al.[16] Similarly, Δ_{mod} is defined on the basis of the roof displacement, Δ_{max}. Besides, Δ_{max} is the maximum roof displacement, and $\Delta_{1,\text{roof}}$ is the roof displacement at the first yield obtained from pushover analysis with lateral load profile based on the first mode distribution. Since the distributions of feature values and EDPs in this study follow the power law, they were then transformed into lognormal space, as suggested by Cornell et al.[56]

5.1 | Inter-storey drift model

5.1.1 | Feature selection

For brevity, only the results corresponding to the Oakland site are presented in this section. A similar process was followed for the Sonora site. First, Pearson's correlation coefficients were calculated between the features, and between the features and the target variable. These correlation coefficients are presented in the form of heatmaps in Figure 12. Several features have high collinearity between each other, as is the case for Δ_{u} and $\Delta_{1,\text{roof}}$, or D_{5-75} and D_{5-95}, as well as several spectral acceleration values, as expected. It is worth noting that the term collinear is used here to stress that, since Pearson's correlation is used, a low correlation coefficient does not imply that the referred feature is unimportant or unrelated to the other, but just that it is not linearly correlated. Although collinearity between features may pose as a problem,[62] it is not always the case,[63] and it is still interesting to observe which of these collinear features is more influential to the seismic drift response of SMRFs.

The last column and row of the heatmaps in Figure 12 show the collinearity of each feature with the target variable, θ_{mod}. Interestingly, most structural parameters (Figure 12A) and duration-related metrics (Figure 12B) have very low collinearity with θ_{mod}. Although ground motion duration is not expected to be used independently to predict seismic responses, it has been recently argued that duration may play a role in the seismic response of yielding systems, even at low levels of demand.[42] On the other hand, the velocity-, acceleration-, (Figure 12B) and spectral acceleration-related parameters (Figure 12C) have prominent collinearity with θ_{mod}.

With regard to so-called wrapper methods, the optimal number of features to be included in a predictive model can be informed by observing the model's performance with certain features in the forward selection and backward elimination results. No significant change in the performance after five features was found. This number of features is also considered to be practical in the context of reducing the complexity of a predictive model equation. The selected feature subsets from

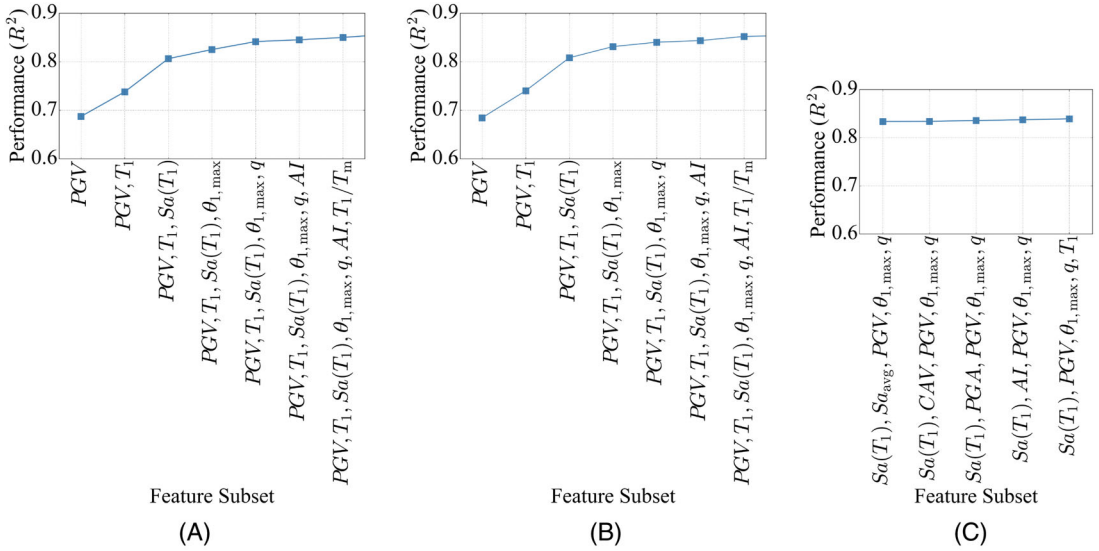


FIGURE 13 Best performing feature subsets from (A) forward selection, (B) backward elimination and (C) exhaustive feature selection for θ_{mod} predictive model at Oakland site

forward selection and backward elimination are presented in Figure 13A,B, respectively. On the other hand, Figure 13C presents the corresponding fiive-feature subsets with the highest performance score obtained from exhaustive feature selection. In this last case, all possible five-feature combinations from all the parameter sets were evaluated (53,130 in total). The results from the wrapper methods consistently select the same features. Some of these selected features are not surprising. For example, $Sa(T_1)$ is widely accepted to govern the seismic response of structures,[64] and *PGV* is recognised as a better indicator of the structural seismic response than other peak values like *PGA*.[65] It is also worth noting that acceleration-related features, *CAV*, *PGA* and *AI*, also appear in the subsets with the highest performance scores identified by the exhaustive feature selection results as they have high collinearity with θ_{mod}.

Finally, the results of the embedded method carried out by means of Random Forest are depicted in Figure 14. The random forest default feature importance and permutation feature importance are shown in this figure. Generally, consistent results are observed among the different methods where *PGV*, $Sa(T_1)$, T_1, q and $\theta_{1,max}$, all appear at the top rows of default and permutation importance methods. However, several unexpected features with very low collinearity with θ_{mod} are also identified (e.g., T_1 /T_m, *PGD*). The appearance of the mean ground-motion period, T_m, is in line with previous studies that highlighted its role in expected drift demands[66–68] and *PGD* may describe well drifts for taller buildings experiencing limited non-linear deformations. Besides, in contrast with the results from the other feature selection methods, the drop column importance shows that the behaviour factor (q), maximum inter-storey drift at the first yield ($\theta_{1,max}$) and several spectral accelerations for periods past the fundamental structural period have negative importance. These unforeseen results from the embedded method can be partially attributed to the collinearity between the features.[55] It should be noted that the feature importance in embedded methods is calculated by considering all features to behave individually and independently of each other and that collinearity acts differently in random forest models. In the random forest with default and permutation feature importance, the importance of collinear features is shared with each other. The amount of this share was found to be a function of how much noise there is between the collinear features by Parr et al.[55] For example, D_{5-75} and D_{5-95} are assigned equal importance as they are highly collinear and have very low noise between them. The effect of collinearity is also present in the drop column importance, as this method calculates the importance from a drop in accuracy. Hence, when one of the highly correlated features is dropped, the accuracy will hardly change, and the importance assigned to this feature will be very small, regardless of its actual importance causing, for example, some spectral accelerations (which are highly collinear between each other) to be assigned very low importance values.[55]

5.1.2 | Model formulation

The differences in feature selection outputs discussed above stress the need to carefully examine the results of ML algorithms as they may behave differently for different datasets. With these differences in mind, several subsets were trialled

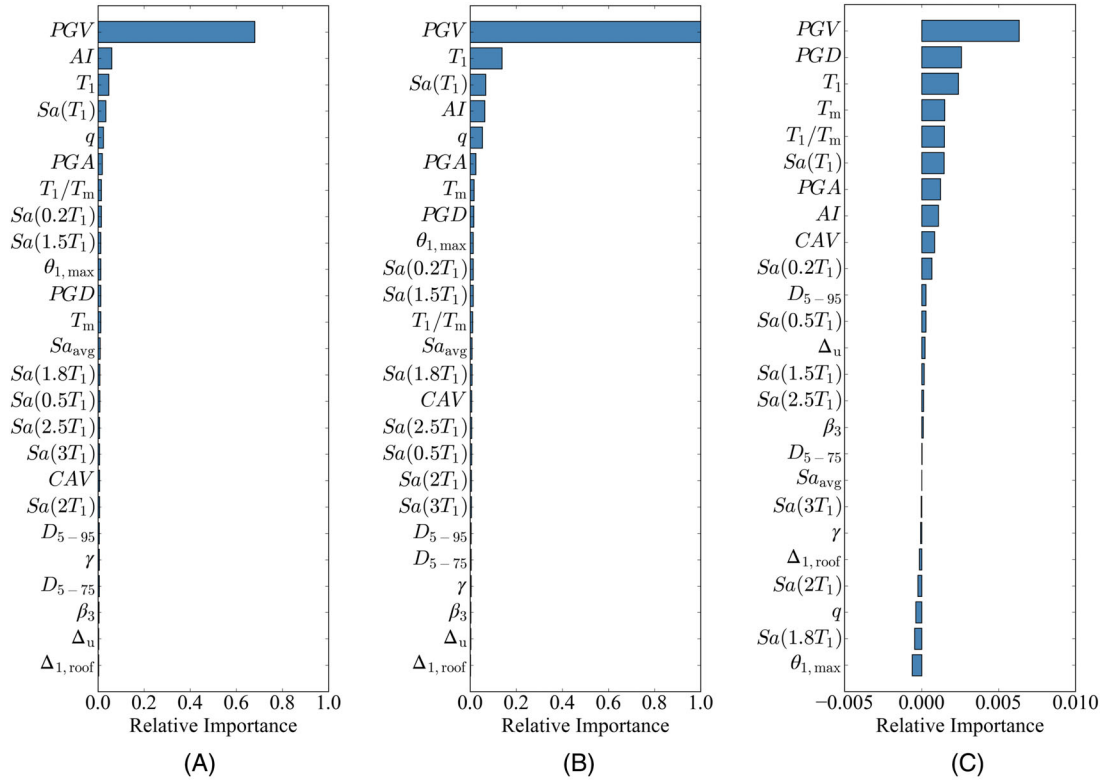


FIGURE 14 Relative importance of the features resulted from random forest (A) default, (B) permutation and (C) drop-column feature importance for θ_{mod} predictive model at Oakland site

TABLE 4 Coefficients of the θ_{mod} predictive model

a_0	a_1	a_2	a_3	a_4	a_5
−5.546	0.265	0.666	0.560	−1.253	−1.278

to arrive at the predictive model with a balance between accuracy and practicality. In the end, the five features used in the predictive model for θ_{mod} at both sites (Oakland and Sonora) are PGV, $Sa(T_1)$, T_1, q and $\theta_{1,max}$. This is an important development facilitated by the observed results since it allows us to formulate models with the same functional form and coefficients that preserve the hazard-consistency attributes of the ground-motion database.

As usual, the database generated as described earlier in this paper was divided into 80% training and 20% validation. Multi-linear regression was performed on the training datasets to derive the EDP predictive model for θ_{mod} as

$$\ln \theta_{mod} = a_0 + a_1 \ln PGV + a_2 \ln Sa(T_1) + a_3 \ln T_1 + a_4 \ln q + a_5 \ln \theta_{1,max} \quad (4)$$

where PGV is expressed in cm/s, $Sa(T_1)$ in g, and the coefficients a_0 to a_5 take the values presented in Table 4.

The model was validated using the testing datasets of each site. Figure 15 shows the model's prediction accuracies in colour shade format in terms of kernel density distribution.[69] It can be seen that most of the data, denoted by darker shade, falls around the reference line, where the predicted target is similar to the observed data. The accuracy of the model can be evaluated by the coefficient of determination, R^2, which measures how much variability in the dependent variable can be explained by the model.[70] While R^2 is a relative measure, the mean square error (MSE) is an absolute measure of the fit, obtained by calculating the sum of the square of prediction errors divided by the number of data points. The root mean square error ($RMSE$) square roots the MSE back to the same level of prediction error and makes it easier to interpret. In addition, to assess the relative difference, the median absolute relative derivation ($MARD$)[71] is also provided to highlight the central tendency of the relative deviation of the prediction from the true value. Besides, the $D_{30\%}$ parameter proposed by Guan et al.[7] that shows the fraction of the dataset whose relative difference does not exceed 30% is also given. The accuracy measures of the model are shown in Figure 15, where they are generally acceptable and comparable to other

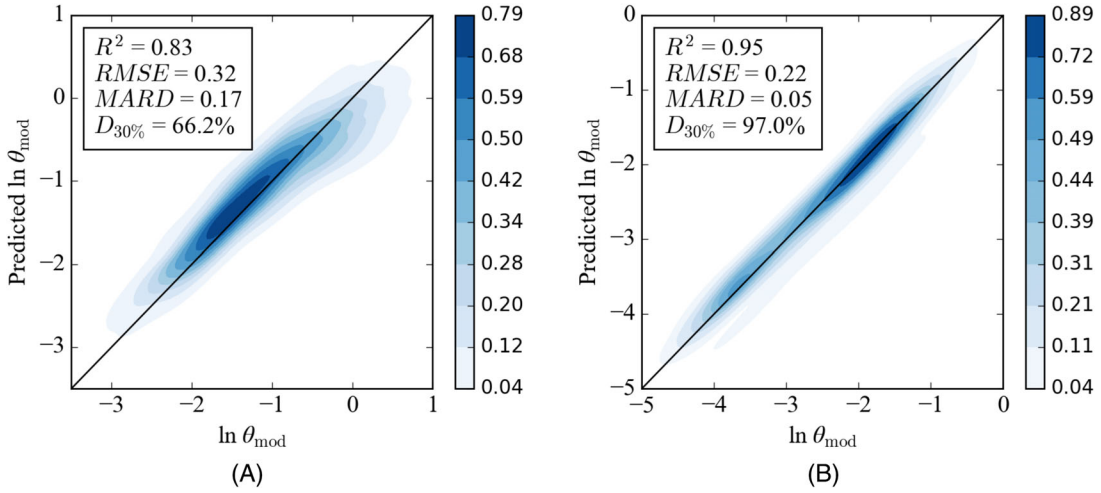


FIGURE 15 Observed θ_{mod} from NRHA versus θ_{mod} from the predictive model at (A) Oakland and (B) Sonora sites

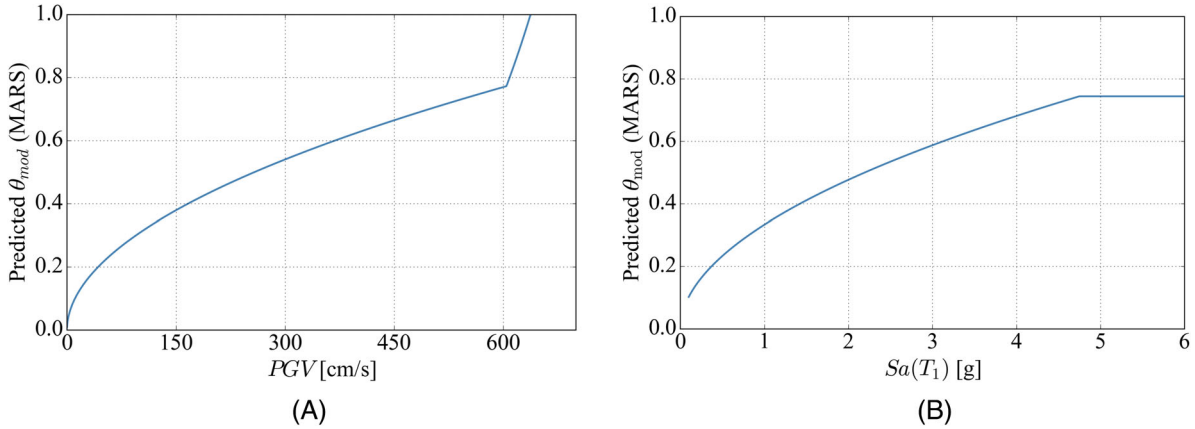


FIGURE 16 Illustration of the (A) PGV and (B) $Sa(T_1)$ knots in MARS-based θ_{mod} predictive model at Oakland site

studies on similar structures, for example, Refs.[7,72,73] or even broader earthquake engineering settings.[74] The Sonora site has a noticeably higher model accuracy ($R^2 = 0.95$; $MARD = 0.05$) compared to Oakland ($R^2 = 0.83$; $MARD = 0.17$), which is attributed to the lower seismic demand and associated dispersion at Sonora, as discussed previously in this paper.

In a second stage, we also explored whether improved models could be obtained by means of multivariate adaptive regression splines (MARS). MARS, first introduced by Friedman,[75] constructs an ensemble of linear functions joined together by one or more hinge functions. The py-earth package[76] was used to conduct the MARS analyses in this study. The predictive model obtained with MARS has the form

$$\begin{aligned}\ln\theta_{\text{mod}} = {} & c_0 + c_1 \ln PGV + c_{2a}\max(0, \ln PGV - d_1) + c_{2b}\max(0, d_1 - \ln PGV) \\ & + c_{3a}\max(0, \ln Sa(T_1) - d_2) + c_{3b}\max(0, d_2 - \ln Sa(T_1)) + c_4 \ln T_1 + c_5 \ln q + c_6 \ln\theta_{1,\max}\end{aligned} \tag{5}$$

where c_i are regression coefficients and d_i are the knots where the behaviour of the model changes. To illustrate these knots, Figure 16 shows the knots for PGV and $Sa(T_1)$ for Frame 24, as an example. The MARS-based predicted θ_{mod} presented in these figures was calculated at fixed values of $Sa(T_1)$ (for Figure 16A) and PGV (for Figure 16B). To this end, the $Sa(T_1)$ for Frame 24 at a 2475-year return period was obtained from the UHS at 0.406 g, while PGV was calculated from the model developed by Huang and Whittaker[77] at a 2475-year return period. It can be seen from Figure 16 that the PGV presents a knot, d_1, at $PGV = 604$ cm/s, whereas the knot for $Sa(T_1)$, d_2, forms at $Sa(T_1) = 4.7$ g. In both cases, the knots are located at ground-motion parameter values associated with a high earthquake intensity.

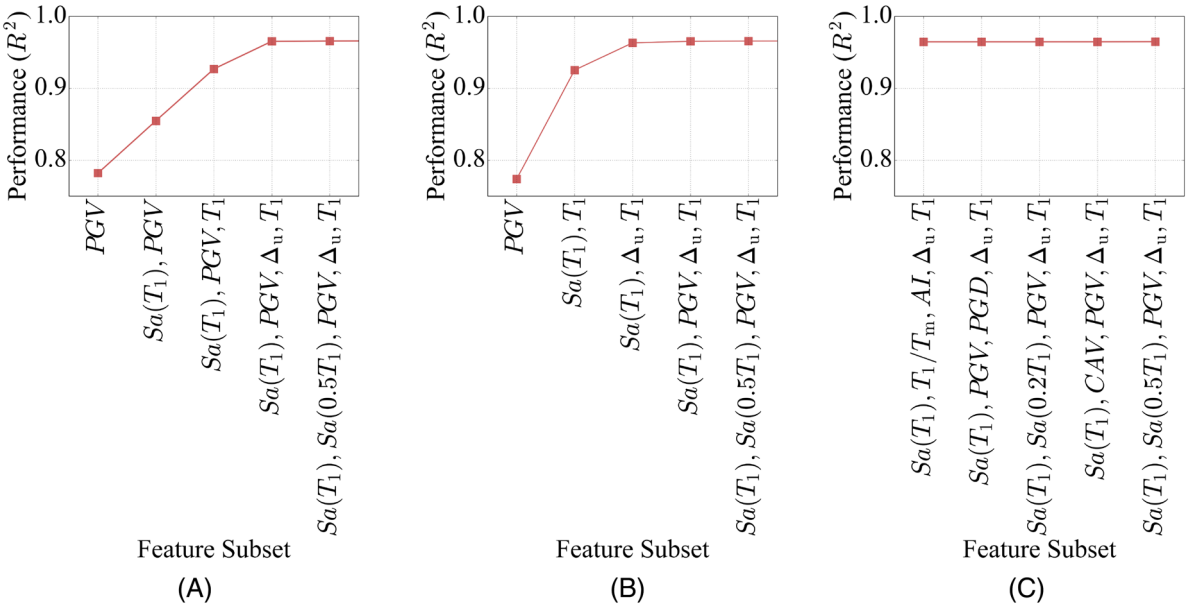


FIGURE 17 Best performing feature subsets from (A) forward selection, (B) backward elimination and (C) exhaustive feature selection for Δ_{mod} predictive model at Sonora site

Despite its sophistication, the MARS-based model does not yield a considerable increase in accuracy in comparison with the multi-linear regression model presented before. A net increase of 0.009 is observed for the R^2 value between our MARS and multi-linear regression models, while the *RMSE* is reduced by only 0.001. Likewise, the *MARD* is only reduced by 0.001, and there is no increase in the $D_{30\%}$ value. This can be due to the knots forming at high earthquake intensities where the available data are limited, for example, although the *PGV* knot at 604 cm/s in Figure 16A changes the tendency dramatically, there are only a few data points available for $PGV > 604$ cm/s. In considering the complexity of the resulting expressions and the overall predictive performance, we have opted to favour the multi-linear regression approach (Equation 4) and hence have decided not to present the MARS coefficients.

5.2 | Global drift model

In addition to inter-storey drifts, it is sometimes important to get an independent estimation of roof drifts. The same steps described above for the formulation of the inter-storey drift models were followed to construct a global drift model. These steps are summarized in this section, although for reasons of brevity, the discussion is succinct.

5.2.1 | Feature selection

A feature selection process in the same line as that explained above for θ_{mod} was also carried out for Δ_{mod} as the target variable at both sites. As before, an optimal number of features of 5 was identified for building the global drift predictive model. The selected feature subsets from forward selection, backward elimination and exhaustive feature selection are presented in Figure 17. Again, the results from the default and permutation importance from the random forest method are slightly different from the results obtained from the wrapper methods, which is attributable to the presence of the collinear features as explained in the case of θ_{mod} in the previous section.

5.2.2 | Model formulation

Several feature subset based on the feature selection results were tried to obtain a predictive model for Δ_{mod} with a balance between accuracy and practicality. *PGV*, $Sa(T_1)$, T_1, q and $\Delta_{1,roof}$ were finally selected, and the following Δ_{mod} predictive

TABLE 5 Coefficients of the Δ_{mod} predictive model at Oakland and Sonora site

b_0	b_1	b_2	b_3	b_4	b_5
−3.097	0.420	0.444	1.431	−0.936	−0.898

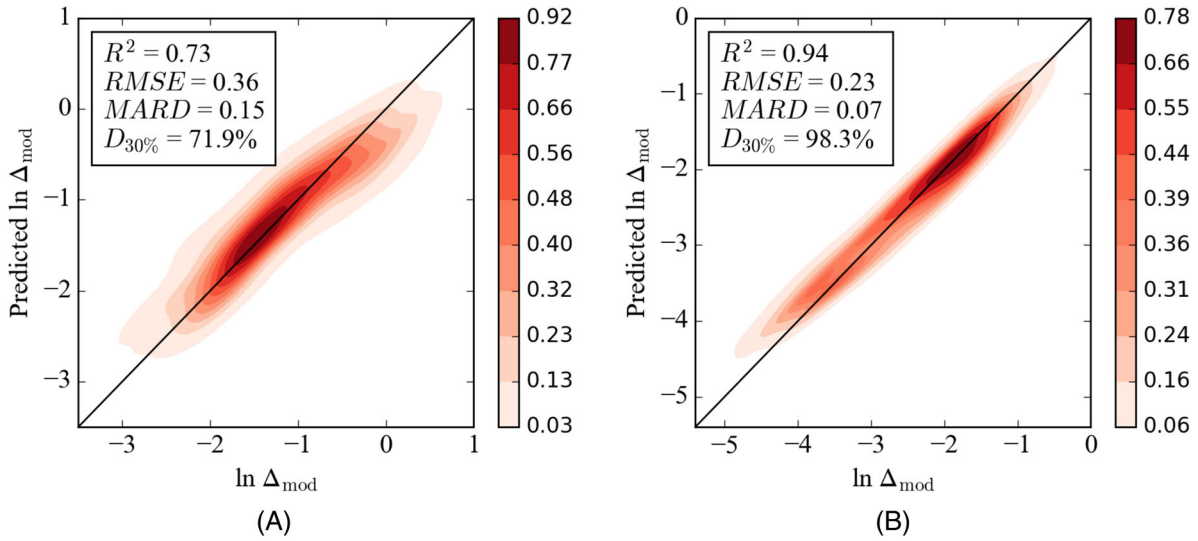


FIGURE 18 Observed Δ_{mod} from NRHA versus Δ_{mod} from the predictive model at (A) Oakland and (B) Sonora sites

model was obtained as

$$\ln \Delta_{mod} = b_0 + b_1 \ln PGV + b_2 \ln Sa(T_1) + b_3 \ln T_1 + b_4 \ln q + b_5 \ln \Delta_{1,roof} \quad (6)$$

whose coefficients b_0 to b_5 are presented in Table 5.

The model was validated using the testing datasets of each site. Figure 18 shows the prediction accuracy together with the R^2 and *RMSE* values. The Sonora site, once again, has noticeably higher model accuracy ($R^2 = 0.94$) compared to the Oakland site ($R^2 = 0.73$). Interestingly, the model for Δ_{mod} was found to have a higher bias in Oakland than its θ_{mod} counterpart. Since the global drift is influenced by the number of storeys and the higher modes, its dispersion is larger than that associated with inter-storey drifts and exacerbated by the higher earthquake intensities expected in Oakland in comparison with Sonora.

As before, the MARS methodology was also tried, but it was associated with a rather small increase in accuracy for a significant increase in the complexity of its functional form. Thus, the multi-linear regression approach was selected.

6 | COMPARISON WITH AVAILABLE MODELS

As mentioned before, several researchers have proposed a number of simplified models for the estimation of seismic drifts in SMRFs. Strictly speaking, such models are not directly comparable with our predictions since the parameter space explored is not identical, and hazard consistency has been overlooked in past investigations. However, some overlap exists, and a comparison is still possible in order to highlight the importance of incorporating hazard considerations into the prediction of local and global drift demands on SMRFs.

Therefore, in this section, we compare our response predictions which can be classed as semi-probabilistic,[78] with the deterministic models proposed by Karavasilis et al.,[16] Kumar et al.[17] and Bravo-Haro et al.[43] It is recognised that, unlike probabilistic approaches like the one followed in this study, there is a number of uncertainties a deterministic procedure will not be able to reflect. It should be noted, however, that all the deterministic models considered in our comparisons used the same seismic design spectrum as the one employed in this paper (Figure 2) and that the building configurations employed are largely similar.

First, the model proposed by Karavasilis et al.[16] for the estimation of Δ_u as a function of the ratio η is expressed as

$$\eta = \frac{\Delta_u}{H\theta_{max}} = 1.0 - 0.193(n_s - 1.0)^{0.54}\rho^{0.144}\alpha^{-0.19} \tag{7}$$

Second, the model proposed by Kumar et al.[17] for the estimation of maximum inter-storey and global drift modification factors that has the following form:

$$\ln\theta_{mod} = e_0 + e_1 q + (e_2 + e_3 q)\ln\left[\min\left(\frac{T_1}{T_m}, 1\right)\right] + e_4 \ln\left[\max\left(\frac{T_1}{1.7T_m}, 1\right)\right] + e_5\gamma + e_6\beta_3 \tag{8}$$

$$\ln\Delta_{mod} = f_0 + f_1 q + (f_2 + f_3 q)\ln\left[\min\left(\frac{T_1}{T_m}, 1\right)\right] + f_4 \ln\left[\max\left(\frac{T_1}{2.782T_m}, 1\right)\right] \tag{9}$$

where e_i and f_i are regression coefficients.

Third, the model proposed by Bravo-Haro et al.[43] for the estimation of inter-storey drifts and roof displacement in SMRF is

$$\ln\theta_{mod} = g_0 + g_1 q + (g_2 + g_3 q)\ln\left[\min\left(\frac{T_1}{T_m}, 1\right)\right] + g_4 \ln\left[\max\left(\frac{T_1}{2.818T_m}, 1\right)\right] \tag{10}$$

$$\ln\Delta_{mod} = h_0 + h_1 q + (h_2 + h_3 q)\ln\left[\min\left(\frac{T_1}{T_m}, 1\right)\right] + h_4 \ln\left[\max\left(\frac{T_1}{2.818T_m}, 1\right)\right] \tag{11}$$

where g_i and h_i are regression coefficients.

Karavasilis et al.[16] observed that the equal displacement rule adopted in some seismic codes (e.g., Eurocode 8[2] and NTC 2018[79]), whereby $\mu = q$, leads to inaccurate estimates of q and proposed explicit expressions for a more accurate determination of the behaviour factor. These expressions have been employed herein in conjunction with Equation (7) for consistency.

Also importantly, the expressions by Kumar et al. and Bravo-Haro et al. require the evaluation of the ground-motion mean period, T_m. Since T_m is scenario-dependent, these procedures can be catalogued as 'semi-probabilistic' since they allow for the incorporation of a probabilistic component within a deterministic framework.[80] However, despite being used in several applications in structural and geotechnical engineering,[60,67,81,82] only a limited number of ground motion prediction equations (GMPEs) are available for T_m. In this paper, the GMPE proposed by Du,[83] constructed on the basis of 8491 recordings from the NGA-West 2 database[37] and based on a physical interpretation of T_m was used. In terms of soil conditions, a $V_{s,30}$ of 760 m/s was considered, and the depth of the top of the fault rupture was set equal to 5 km, while directivity was disregarded. Furthermore, since the depth to the 1000 m/s shear wave isosurface, Z_1, was not available, the recommended default value of $Z_1 = 0$ was employed.[83] The same earthquake scenarios used for estimating T_m were also considered for the estimation of *PGV* required for the hybrid ML-based models proposed in this study. To this end, the conditional ground motion model proposed by Huang and Whittaker,[77] as advised by FEMA P-58,[5] was adopted.

Typical results predicted by the models proposed by Karavasilis et al.,[16] Kumar et al.[17] and Bravo-Haro et al.[43] for Frame 03 (3S-3B), Frame 09 (6S-3B) and Frame 15 (9S-3B) are presented in Figure 19 for $\nu = 1/475$ and $\nu = 1/2475$. Since the database employed by Kumar et al.[17] considers SMRFs of up to seven storeys, Kumar et al.'s estimates for nine-storey buildings are not presented in this figure. It is evident from Figure 19 that all available models severely overestimate both the peak inter-story drifts and roof displacements. In fact, all these models predict all frames to have collapsed based on the FEMA 356[39] criteria. The largest overestimations come from Bravo-Haro et al.'s model with a 500% difference for θ_{max} in comparison with the more exhaustive CSS calculations. On the other hand, Kumar et al. and Karavasilis et al. show a four- and three-fold over-prediction, respectively.

Notably, the previously proposed approaches were also found unable to discern between different hazard levels. Karavasilis et al. model provides a single drift and displacement value regardless of the return period being assessed. In comparison, Kumar et al. and Bravo-Haro et al. scaling are determined by T_m alone, whose values for $\nu = 1/475$ and $\nu = 1/2475$ are very close to each other, leading to virtually the same estimates for different earthquake scenarios. On the other hand, the hazard-consistent predictive models developed in this study, also presented in Figure 19, are significantly more consistent with the hazard curves. Although providing a lower bound estimate of responses, the hazard-consistent estimations found with Equations (4) and (6) are able to replicate the hazard level scaling and provide correspondingly higher drift estimates for lower annual exceedance rates.

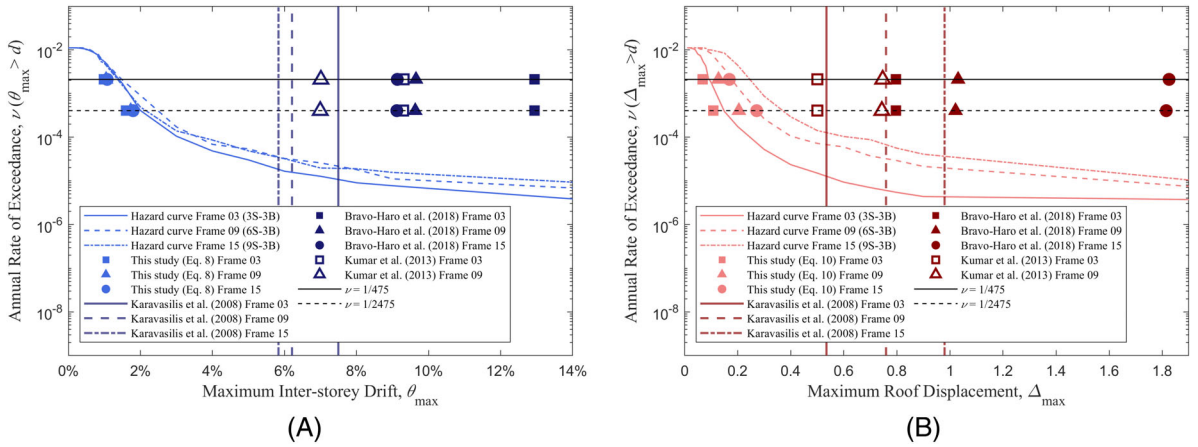


FIGURE 19 Comparison of probabilistic and deterministic estimations of (A) maximum inter-storey drift and (B) maximum roof displacement of Frame 03, 09 and 15

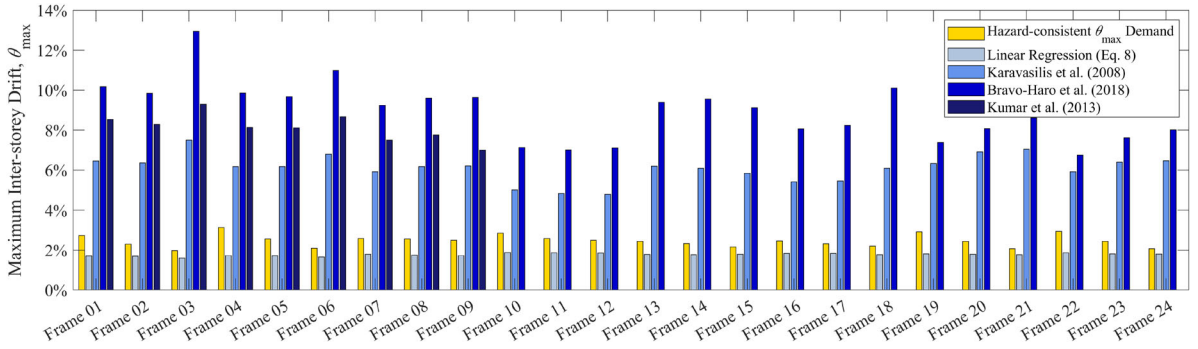


FIGURE 20 Comparison of probabilistic and deterministic estimations of maximum inter-storey drift at Oakland site for $T_R = 2475$ years

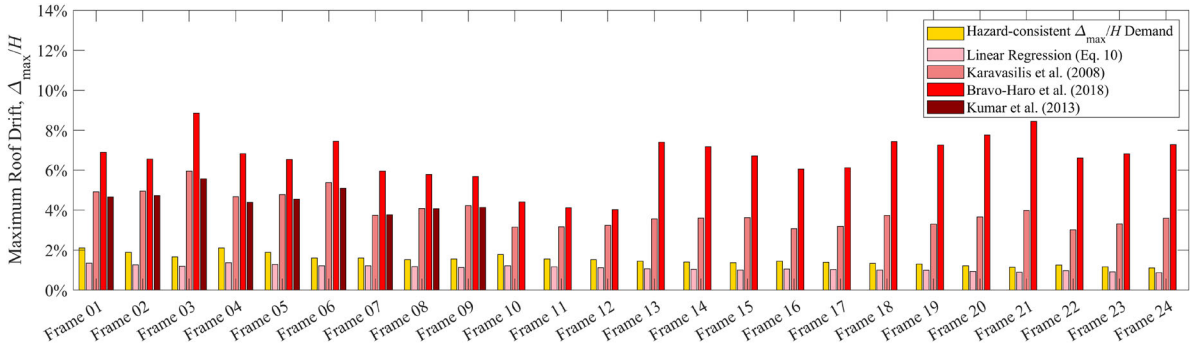


FIGURE 21 Comparison of probabilistic and deterministic estimations of maximum roof drift at Oakland site for $T_R = 2475$ years

A complete comparison between the CSS inter-storey demands for all frames and the estimations from the proposed and available models is presented in Figure 20 for inter-storey drifts demands and in Figure 21 for roof displacements. It can be clearly seen that all previous models severely overestimate the actual demand, sometimes excessively, while the proposed models can predict the demand quite well. It should also be noted that some of the available models, notably,[43] which produce the highest overestimations, were developed using IDA analyses rather than hazard-consistent PSDA, which stresses the importance of a hazard-consistent analysis. Although the bias introduced by deterministic predictive models is now generally recognised, the results presented above stress the sheer level of their overestimations and call for the use of hazard-informed predictions in seismic assessments.

7 | CONCLUSIONS

This study has evaluated the inter-storey drift and roof displacement demands on SMRFs within a hazard-consistent PBEE framework. To this end, NRHA of 24 SMRFs were conducted using a database of 816 ground motions representative of two sites with different seismicity. We have used the CSS methodology to estimate the rate of occurrence of each ground motion, hence ensuring the hazard consistency of our outputs. Several observations can be made from the comparisons of engineering demand parameter curves (EDP-hazard curves): (i) site-specific performance assessments are needed in order to ensure economical designs although EC8 appears inherently unable to offer a consistent level or risk for a given EDP; (ii) contrary to recent suggestions based on non-hazard consistent results that call for widespread use of deteriorating models even at lower demand levels; when looked at from a hazard-consistent perspective, the inherent variability in the response overtakes any deterioration effect making it sensible to avoid the complexity of using advanced deteriorating models for assessments that are not intended to reach collapse scenarios; (iii) the plasticity resistance ratio influences the inter-storey drift demand more than the number of bays or storeys, while the latter has a greatest effect on the roof displacement demands. The insignificant influence of the number of bays at lower hazard levels, and its immateriality at higher hazard levels contradict the commonly held belief that a larger number of bays is conducive to a better seismic response.

The extensive database of drifts constructed in this study was used for the development of EDP predictive models. This was preceded by a careful feature selection process using several ML algorithms. PGV, $Sa(T_1)$, T_1, q and $\theta_{1,\max}$ were selected as the best predictors for inter-storey drift, while PGV, $Sa(T_1)$, T_1, q and $\Delta_{1,\text{roof}}$ were identified for roof displacement. Although these structural and earthquake features have been recognised separately in previous studies as potentially good predictors; their relative importance, as obtained from the ML algorithms used here, can differ. The differences in the results of each ML algorithm described and discussed in this study stress the need to interpret the results of ML tools carefully.

The developed hazard-consistent predictive models were tested, and their good prediction accuracies were confirmed. The Sonora site has a higher model accuracy compared to Oakland, which is attributed to the lower seismic demand and the associated lower response dispersion at Sonora. More sophisticated MARS models were also developed, but they produced only marginal improvements. Although a more advanced ML algorithm may increase the accuracy of the models, this is outside the scope of this paper, which is to assess the influence of hazard-consistency on these models, and to strike the right balance between accuracy, interpretability and, most importantly, practicality (e.g., for hazard integration). Last, our new models were compared with available deterministic and hybrid approaches from the literature. We found that all currently available models can lead to severe overestimations of both peak inter-story drifts and roof displacements. In contrast, the hazard-consistent predictive models developed in this study were able to replicate well the structural drift response and its scaling with seismic hazard level (i.e., provide correspondingly higher drift estimates for lower exceedance rates). The results and findings of this study stress the importance of hazard consistency in PBEE design and assessment and provide the framework to implement it directly in a simple yet accurate way.

ACKNOWLEDGEMENT

The first author would like to thank the support of Indonesia Endowment Funds for Education (LPDP).

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

DATA AVAILABILITY STATEMENT

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

REFERENCES

1. SEAOC. *Performance-Based Seismic Engineering*. Structural Engineers Association of California; 1995.
2. CEN. *EN 1998-1, Eurocode 8: Design of Structures for Earthquake Resistance. Part I: General Rules, Seismic Actions and Rules for Buildings*. European Committee for Standardization; 2005.
3. ASCE. *ASCE 7-16: Minimum Design Loads for Buildings and other Structures*. American Society of Civil Engineers; 2016.
4. Cornell A, Krawinkler H. Progress and challenges in seismic performance assessment. *PEER Newsletter*. 2000.
5. FEMA. *FEMA P-58-1: Seismic Performance Assessment of Buildings (Volume 1 – Methodology)*. Federal Emergency Management Agency; 2018.

6. Freddi F, Galasso C, Cremen G, et al. Innovations in earthquake risk reduction for resilience: recent advances and challenges. *Int J Disaster Risk Reduct*. 2021;60:102267.
7. Guan X, Burton H, Shokrabadi M, Yi Z. Seismic drift demand estimation for steel moment frame buildings: from mechanics-based to data-driven models. *J Struct Eng*. 2021;147(6):04021058.
8. Miranda E. Approximate seismic lateral deformation demands in multistory buildings. *J Struct Eng*. 1999;125(4):417-425.
9. Miranda E, Reyes CJ. Approximate lateral drift demands in multistory buildings with nonuniform stiffness. *J Struct Eng*. 2002;128(7):840-849.
10. Lin YY, Miranda E. Estimation of maximum roof displacement demands in regular multistory buildings. *J Eng Mech*. 2010;136(1):1-11.
11. Morfidis K, Kostinakis K. Seismic parameters' combinations for the optimum prediction of the damage state of R/C buildings using neural networks. *Adv Eng Softw*. 2017;106:1-16.
12. Cook D, Wade K, Haselton C, Baker J, DeBock D. A structural response prediction engine to support advanced seismic risk assessment. In Eleventh US National Conference on Earthquake Engineering, Los Angeles, CA. 2018.
13. Zhang R, Chen Z, Chen S, Zheng J, Büyüköztürk O, Sun H. Deep long short-term memory networks for nonlinear structural seismic response prediction. *Comput Struct*. 2019;220:55-68.
14. Málaga-Chuquitaype C. Machine learning in structural design: an opinionated review. *Front Built Environ*. 2022;8:815717.
15. Gupta A, Krawinkler H. Behavior of ductile SMRFs at various seismic hazard levels. *J Struct Eng*. 2000;126(1):98-107.
16. Karavasilis TL, Bazeos N, Beskos DE. Drift and ductility estimates in regular steel MRF subjected to ordinary ground motions: a design-oriented approach. *Earthq Spectra*. 2008;24(2):431-451.
17. Kumar M, Stafford PJ, Elghazouli A. Influence of ground motion characteristics on drift demands in steel moment frames designed to eurocode 8. *Eng Struct*. 2013;52:502-517.
18. Hu S, Wang W, Alam MS. Probabilistic nonlinear displacement ratio prediction of self-centering energy-absorbing dual rocking core system under near-fault ground motions using machine learning. *J Earthq Eng*. 2021;31:1-32.
19. Hu S, Zhu S, Alam MS, Wang W. Machine learning-aided peak and residual displacement-based design method for enhancing seismic performance of steel moment-resisting frames by installing self-centering braces. *Eng Struct*. 2022;271:114935.
20. Hu S, Zhu S, Wang W. Machine learning-driven probabilistic residual displacement-based design method for improving post-earthquake repairability of steel moment-resisting frames using self-centering braces. *J Building Eng*. 2022;61:105225.
21. Málaga-Chuquitaype C, Bougatsas K. Vector-im-based assessment of alternative framing systems under bi-directional ground-motion. *Eng Struct*. 2017;132:188-204.
22. Arteta C, Abrahamson N. Conditional scenario spectra (CSS) for hazard-consistent analysis of engineering systems. *Earthq Spectra*. 2019;35(2):737-757.
23. Vamvatsikos D, Cornell CA. Incremental dynamic analysis. *Earthq Eng Struct Dyn*. 2002;31(3):491-514.
24. Abrahamson N, Al Atik L. Scenario spectra for design ground motions and risk calculation. In 9th US National and 10th Canadian Conference on Earthquake Engineering. 2010:12.
25. Arteta C, Abrahamson N. Methodology based on conditional scenario spectra to estimate engineering demand parameter risk. In Proceedings of the 16th World Conference on Earthquake Engineering, Santiago, Chile. 2017.
26. Abrahamson N, Yunatci AA. Ground motion occurrence rates for scenario spectra. In Fifth International Conference on Recent Advances in Geotechnical Earthquake Engineering and Soil Dynamics. 2010.
27. Baker JW. Conditional mean spectrum: tool for ground-motion selection. *J Struct Eng*. 2011;137(3):322-331.
28. Arteta C, Araújo GA, Torregroza AM, Martínez AF, Lu Y. Hybrid approach for simulating shear-flexure interaction in rc walls with nonlinear truss and fiber models. *Bullet Earthq Eng*. 2019;17(12):6437-6462.
29. Buniya M, Simpson B, Macedo J, Vergaray L, Barbosa A. Collapse fragility function development using conditional scenario spectra: application to a multi-story reinforced concrete shear wall. In 17th World Conference on Earthquake Engineering, 17WCEE. Sendai. 2020.
30. Jaimes MA, Candia G. Seismic risk of sliding ground-mounted rigid equipment. *Eng Struct*. 2020;204:110066.
31. CEN. *EN 1993-1-1, Eurocode 3: Design of Steel Structures. Part 1.1: General Rules for Buildings*. European Committee for Standardization. 2005.
32. McKenna F, Fenves GL, Scott MH. *Open System for Earthquake Engineering Simulation*. University of California; 2000.
33. Charney FA, Marshall J. A comparison of the Krawinkler and scissors models for including beam-column joint deformations in the analysis of moment-resisting steel frames. *Eng J-Am Inst Steel Constr INC*. 2006;43(1):31-48.
34. Ibarra LF, Medina RA, Krawinkler H. Hysteretic models that incorporate strength and stiffness deterioration. *Earthq Eng Struct Dyn*. 2005;34(12):1489-1511.
35. Lignos DG, Krawinkler H. Deterioration modeling of steel components in support of collapse prediction of steel moment frames under earthquake loading. *J Struct Eng*. 2011;137(11):1291-1302.
36. Newmark NM, Hall WJ. *Earthquake Spectra and Design. Engineering Monographs on Earthquake Criteria*. Earthquake Engineering Research Institute; 1982.
37. Bozorgnia Y, Abrahamson N, Atik LA, et al. NGA-West2 research project. *Earthq Spectra*. 2014;30(3):973-987.
38. HPC-Imperial. Imperial college research computing service. 2019.
39. FEMA. *FEMA 356: Prestandard and Commentary for the Seismic Rehabilitation of Buildings*. Federal Emergency Management Agency; 2000.

40. Málaga-Chuquitaype C, Elghazouli A, Bento R. Rigid-plastic models for the seismic design and assessment of steel framed structures. *Earthq Eng Struct Dyn*. 2009;38(14):1609-1630.
41. Paglietti A, Porcu MC. Rigid-plastic approximation to predict plastic motion under strong earthquakes. *Earthq Eng Struct Dyn*. 2001;30(1):115-126.
42. Málaga-Chuquitaype C. Strong-motion duration and response scaling of yielding and degrading eccentric structures. *Earthq Eng Struct Dyn*. 2021;50(2):635-654.
43. Bravo-Haro M, Tsitos A, Elghazouli A. Drift and rotation demands in steel frames incorporating degradation effects. *Bull Earthq Eng*. 2018;16(10):4919-4950.
44. Krawinkler H. Importance of good nonlinear analysis. *Struct Design Tall Special Build*. 2006;15(5):515-531.
45. Wen Y, Song SH. Structural reliability/redundancy under earthquakes. *J Struct Eng*. 2003;129(1):56-67.
46. Arturo TC, Cortés-Benítez JA. Assessment of redundancy in the seismic design of moment resisting reinforced concrete frames. In 11th Canadian Conference on Earthquake Engineering. 2015.
47. Raschka S. MLxtend: providing machine learning and data science utilities and extensions to Python's scientific computing stack. *J Open Source Softw*. 2018;3(24):638.
48. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5-32.
49. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825-2830.
50. Parr T, Wilson JD, Hamrick J. Nonparametric feature impact and importance. *arXiv preprint arXiv:2006.04750*. 2020.
51. Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A. Conditional variable importance for random forests. *BMC Bioinform*. 2008;9(1):1-11.
52. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinform*. 2007;8(1):1-21.
53. Breiman L, Cutler A. Random forest. 2004. Accessed December 12, 2021. https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm
54. Parr T, Turgutlu K. Rfpimp 1.3.7. 2021. Accessed December 12, 2021. https://pypi.org/project/rfpimp/
55. Parr T, Turgutlu K, Csiszar C, Howard J. Beware default random forest importances. 2018. Accessed December 12, 2021. http://parrt.cs.usfca.edu/doc/rf-importance/index.html
56. Cornell CA, Jalayer F, Hamburger RO, Foutch DA. Probabilistic basis for 2000 sac federal emergency management agency steel moment frame guidelines. *J Struct Eng*. 2002;128(4):526-533.
57. Reed JW, Kassawara RP. A criterion for determining exceedance of the operating basis earthquake. *Nucl Eng Des*. 1990;123(2-3):387-396.
58. Arias A. A measure of earthquake intensity. In *Seismic Design for Nuclear Power Plants*. Massachusetts Institute of Technology; 1970.
59. Bommer JJ, Martinez-Pereira A. The effective duration of earthquake strong motion. *J Earthq Eng*. 1999;3(02):127-172.
60. Rathje EM, Abrahamson N, Bray JD. Simplified frequency content estimates of earthquake ground motions. *J Geotech Geoenviron Eng*. 1998;124(2):150-159.
61. Dávalos H, Miranda E. A ground motion prediction model for average spectral acceleration. *J Earthq Eng*. 2021;25(2):319-342.
62. Allen MP. *Understanding Regression Analysis*. Springer Science & Business Media; 2004.
63. Morrissey MB, Ruxton GD. Multiple regression is not multiple regressions: the meaning of multiple regression and the non-problem of collinearity. *Philos, Theory, Pract Biol*. 2018;10(3).
64. Shome N, Cornell CA, Bazzurro P, Carballo JE. Earthquakes, records, and nonlinear responses. *Earthq Spectra*. 1998;14(3):469-500.
65. Bommer JJ. Predicting and monitoring ground motions induced by hydraulic fracturing. 2017. Accessed December 12, 2021. https://www.ogauthority.co.uk/media/3693/pgv-thresholds-for-control-of-induced-seismic-hazard_v5.pdf
66. Málaga-Chuquitaype C, Elghazouli A. Inelastic displacement demands in steel structures and their relationship with earthquake frequency content parameters. *Earthq Eng Struct Dyn*. 2012;41(5):831-852.
67. Málaga-Chuquitaype C. Estimation of peak displacements in steel structures through dimensional analysis and the efficiency of alternative ground-motion time and length scales. *Eng Struct*. 2015;101:264-278.
68. Hickey J, Broderick B. Influence of mean period of ground motion on inelastic drift demands in CBFs designed to Eurocode 8. *Eng Struct*. 2019;182:172-184.
69. Sammut C, Webb GI. *Encyclopedia of Machine Learning and Data Mining*. Springer Publishing Company; 2017.
70. Mendenhall WM, Sincich TL. *Statistics for Engineering and the Sciences*. CRC Press LLC; 2015.
71. Sun H, Burton H, Zhang Y, Wallace J. Interbuilding interpolation of peak seismic response using spatially correlated demand parameters. *Earthq Eng Struct Dyn*. 2018;47(5):1148-1168.
72. Di Sarno L, Wu JR. Fragility assessment of existing low-rise steel moment-resisting frames with masonry infills under mainshock-aftershock earthquake sequences. *Bull Earthq Eng*. 2021;19(6):2483-2504.
73. Basim MC, Pourreza F, Mousazadeh M, Hamed AA. The effects of modeling uncertainties on the residual drift of steel structures under mainshock-aftershock sequences. *Structures*. 2022;36:912-926.
74. Macedo J, Liu C, Soleimani F. Machine-learning-based predictive models for estimating seismically-induced slope displacements. *Int J Soil Dyn Earthq Eng*. 2021;148:106795.
75. Friedman JH. Multivariate adaptive regression splines. *Ann Stat*. 1991;19(1):1-67.
76. Rudy J. Py-earth. 2013. Accessed November 1, 2021. https://contrib.scikit-learn.org/py-earth/content.html
77. Huang YN, Whittaker AS. *Calculation of Peak Ground Velocity from 1-second Spectral Velocity. Background Document FEMA P-58/BD-3.7.5*. Federal Emergency Management Agency; 2012.

78. Macedo J, Ramesh V, Liu C, Kottke A. Evaluating different approaches for the hazard-consistent assessment of the seismic performance of dams. *Bull Seismol Soc Am*. 2022;112:1710-1726.
79. NTC18. *Circolare 21 gennaio 2019, n. 7 istruzioni per l'applicazione dell'aggiornamento delle norme tecniche per le costruzioni di cui al D.M. 17 gennaio 2018*. Consiglio Superiore dei Lavori Pubblici, Gazzetta Ufficiale della Repubblica Italiana; 2019.
80. Macedo J, Bray J, Abrahamson N, Travasarou T. Performance-based probabilistic seismic slope displacement procedure. *Earthq Spectra*. 2018;34(2):673-695.
81. Demirci C, Málaga-Chuquitaype C, Macorini L. Seismic shear and acceleration demands in multi-storey cross-laminated timber buildings. *Eng Struct*. 2019;198:109467.
82. Saygili G, Rathje EM. Empirical predictive models for earthquake-induced sliding displacements of slopes. *J Geotech Geoenviron Eng*. 2008;134(6):790-803.
83. Du W. An empirical model for the mean period (Tm) of ground motions using the NGA-West2 database. *Bull Earthq Eng*. 2017;15(7):2673-2693.

How to cite this article: Zahra F, Macedo J, Málaga-Chuquitaype C. Hybrid data-driven hazard-consistent drift models for SMRF. *Earthquake Engng Struct Dyn*. 2023;1-24. https://doi.org/10.1002/eqe.3807