Received: 23 February 2024 | Revised: 24 July 2024 | Accepted: 25 July 2024

DOI: 10.1002/eqe.4208

RESEARCH ARTICLE

WILEY

The impact of the choice of intensity measure and seismic demand model on seismic risk estimates with respect to an unconditional benchmark

Archie Rudman[1] | **Enrico Tubaldi**[1] | **John Douglas**[1] | **Fabrizio Scozzese**[2]

[1]Department of Civil and Environmental Engineering, University of Strathclyde, Glasgow, UK

[2]School of Architecture and Design, University of Camerino, Camerino, Italy

Correspondence
Archie Rudman, James Weir Building Level 5, 75 Montrose Street, University of Strathclyde, Glasgow, G1 1XJ, UK.
Email: archie.rudman@strath.ac.uk

Funding information
University of Strathclyde

Abstract

Many methods for seismic risk assessment rely on the selection of a seismic intensity measure (*IM*) and the development of models of the seismic demand conditional on the *IM*. The *individual* importance of these two features to accurately assess seismic performance is well known. In contrast, this study aims to evaluate the impact that the *combined* selection of *IM* and the demand model has on risk estimates. Using a hypothetical seismic source model and a non-stationary stochastic ground-motion model, we present risk estimates for a mid-rise steel structure for 15 different *IM*s and five demand models derived by cloud analysis (four based on regression and a fifth based on an empirical binning approach). The impact of these choices is investigated through a novel method of model performance evaluation using a benchmark solution obtained via the unconditional approach (i.e., directly estimating demand exceedance frequencies from simulated ground motion time histories). The obtained results are also compared against traditional *IM* performance metrics, for example, efficiency and sufficiency. Finally, we demonstrate how risk estimate inaccuracies are propagated by performing a damage assessment on two example components. The results show that, for the scenario under investigation, Arias intensity combined with the binned demand model provides the best risk estimates, if sufficient samples are available, whilst ground displacement and duration-based *IM*s ranked worst, irrespective of the demand model. The findings highlight the importance and interconnectedness of the selection of the *IM* and the demand model when using cloud analysis and present a clear method of determining the most accurate combination for risk assessments.

KEYWORDS
cloud analysis, intensity measure selection, machine learning, seismic demand modelling, seismic risk assessment, stochastic ground motion models

1 | INTRODUCTION

The seismic risk of structures can be assessed via two distinct approaches.[1,2] The first, known as the unconditional method, directly uses observations of a system's response to ground motions to estimate the rate of exceedance of some loss threshold.[3] This is a conceptually simple, but computationally expensive, approach—and so is rarely used outside of research. The second approach, known in the literature as the conditional method, was established to overcome this problem. Cornell,[4] among others, describes the basis of the conditional method, where a series of intermediate modelling steps are made to assess the risk of a system (Figure 1). This method allows the efficient description of seismic risk in a region with far fewer simulations and with a considerable reduction in computational expense. More detailed descriptions of the conditional method can be found elsewhere.[5,6]

One of the most popular conditional approaches for assessing earthquake risk is the PEER's performance-based earthquake engineering (PBEE)[7,8] framework. The original intention of the framework was to report risk estimates that represent the whole structure. However, performing performance assessments at the component level can provide a more thorough understanding of a building's performance,[9] given the high contribution of non-structural components to overall earthquake-induced losses.[10] Therefore, more recent implementations of the PBEE approach (e.g., the FEMA P-58 method[11]) includes component-level analyses within seismic risk assessments.

Conditional approaches rely on defining a set of conditioning parameters that represent each stage of the assessment process. Decision variables (*DVs*), damage states (*DSs*) and engineering demand parameters (*EDPs*) describe respectively the system's losses (e.g., deaths, dollars and downtime), damage (e.g., cracking of concrete, or buckling of beams and columns) and response to shaking (e.g., exceedance of an inter-storey drift ratio threshold), respectively. These three parameters are reliant on the initial selection of an intensity measure (*IM*), which represents the ground shaking at the site of interest and is used to evaluate the site hazard, as well as describing the structural and non-structural component response.

The importance of *IM* selection is well known, with a range of *IM*s proposed in the literature. These generally consider one or more of three ground-motion characteristics: amplitude, frequency and duration.[12] Historically, the most common *IM* used was the peak value of ground-motion amplitude (e.g., largest absolute acceleration); whereas now it is more common for spectral ordinates to be used (e.g., 5% damped spectral acceleration at the fundamental period of the structure or a combination of spectral accelerations at different periods). Alternatively, it has been suggested that a vector of *IM*s would describe ground motions more accurately.[13–15] In an ideal scenario, *IM* selection should have little impact on risk estimates, provided that the *IM* is *sufficient* (statistically independent from earthquake characteristics, e.g., magnitude and distance),[16] and enough ground-motion records are used to characterise the system's response. However, satisfying these criteria is not always possible, and the need to limit the number of numerical simulations has led to a proliferation of alternative *IM*s and comparisons in terms of *efficiency*.

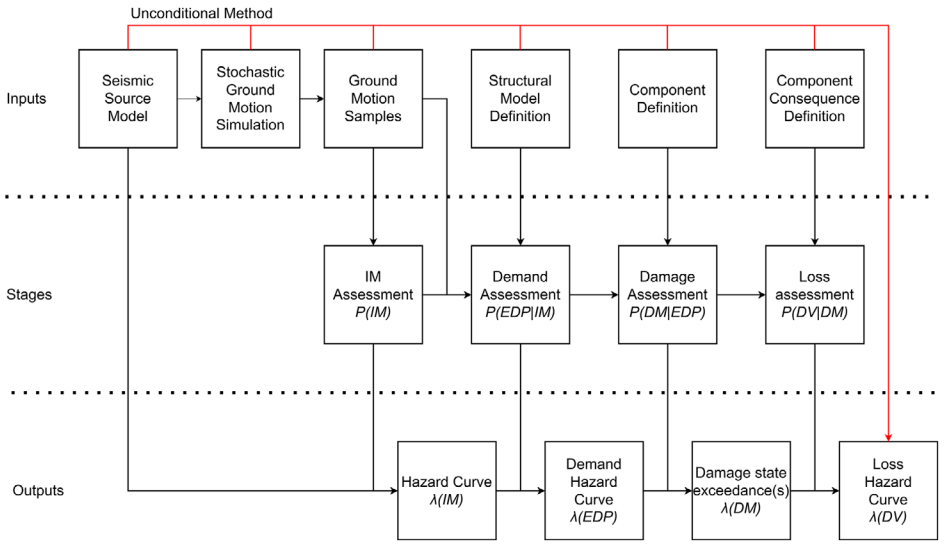

FIGURE 1 Workflow of the conditional approach to risk assessment when ground motion simulations are utilised. The unconditional approach can be described by the red workflow.

Due to the large number of *IM*s in existence, the topic of selecting an optimal *IM* has become a well-covered, and wide-ranging debate (see e.g., Katsanos et al.[17] for a review of *IM* development and selection in the context of ground motion record selection). It has proved difficult to find a comprehensive *IM* that is optimal for all types of earthquake risk assessment. This in turn has led to many research efforts being devoted to evaluating *IM* selection for a range of contexts. For instance, Mackie and Stojadinovic[18] evaluated *IM* selection for probabilistic seismic demand models of highway bridges in California. Bray and Travasarou[19] investigated the impacts of *IM* selection for estimating seismic slope displacements. Kohrangi et al.[20] compared the impact of selecting eight different structure-specific *IM*s on repair cost estimates of three different 3D building models.

Despite this coverage, there is little research on the impact that *IM* selection has on the final output of PBEE assessments. Some notable examples of studies that did consider this final output are Kohrangi et al.,[20] who made loss estimates within their study on 3D building models, O'Reilly[21] who investigated *IM* selection for seismic risk assessment of bridges, and Du et al.[22] who provided insight on the influence of *IM* selection in the context of regional seismic risk assessment. However, most studies that compare *IM*s only go as far as predicting the statistics of the demand conditional on the *IM* or the frequency of exceeding an *EDP* threshold (the demand hazard), instead of also examining the results that would more likely be useful to end users, for example, damage estimates. There has also been little evaluation of *IM* impact through comparison with an unconditional benchmark, although Kwong et al.[23,24] have previously demonstrated how this comparison can be used to evaluate ground motion selection procedures at the demand hazard step of risk assessment.

The *IM* choice plays an important role in characterising the seismic demand, that is, the response of a system due to ground motions, represented by the conditional demand assessment stage in Figure 1. Several different methods have been proposed to describe this relationship, including multiple stripe analysis[1,25] and incremental dynamic analysis.[26] Possibly the simplest of these methods to implement is cloud analysis[25] which fits a regression model with *IM* as the independent variable, and *EDP* as the dependent variable. The most practical, but least sophisticated, regression models are based on a linear fit between *IM* and *EDP*, whereas more complex models use bilinear fits to account for the nonlinearity in the structural response[27–29]; machine learning tools have also been used to improve this characterisation.[30] Nevertheless, there are few studies comparing these different methods of modelling the seismic demand, especially against an unconditional benchmark and considering different *IM*s.

This research aims to investigate the impact that both the *IM* selection and the method of characterising the *IM-EDP* relationship have on component-level earthquake risk and damage estimates. Using a numerical model of a benchmark building widely employed in PBEE-based studies and a non-stationary stochastic ground-motion model for a hypothetical seismic source scenario, a total of 15 different *IM*s and five seismic demand models are considered, with the subsequent impact of each combination on the risk estimates evaluated. Ground-motion simulations are used to generate a large amount of data to properly evaluate uncertainties within the risk assessment procedure and also to allow the estimation of risk via the unconditional approach, which acts as a benchmark to compare against conditional estimates.

The following sections describe: the seismic scenario and method of ground-motion simulation employed (Section 2), the structural model (Section 3) considered for the study, and the *IM*s evaluated (Section 4). Current methods for evaluating the optimal *IM* are also discussed, and a new method to determine the optimal *IM* is presented (Section 4). The conditional risk estimates obtained for each *IM* using all five seismic demand models are presented (Section 5), before the impact of each *IM* and demand model is evaluated (Section 6). Finally, damage estimates are made for two components with the comparison between the benchmark (unconditional) and conditional damage estimates also shown, to demonstrate the impact that *IM* selection and seismic demand modelling can have on FEMA P-58 component-level analysis (Section 6).

2 | SEISMIC SCENARIO

For this study, a fictive scenario is established with a circular source zone of radius 100 km, and two faults of length 75 and 25 km. Simulated ground-motion 'recordings' are made at a single station at the centre of the circular source zone. The location and details of each of these seismic sources are shown in Figure 2. All earthquakes in the scenario follow the Gutenberg-Richter relationship with minimum magnitude 5 and $b = 1.0$; the maximum magnitude and the a-value for each source are provided in Figure 2.

The non-stationary stochastic model of Sabetta et al.[31] is used to generate realistic ground motions. The model simulates ground motions with just a few easily defined inputs: moment magnitude, source-to-site distance, time-averaged

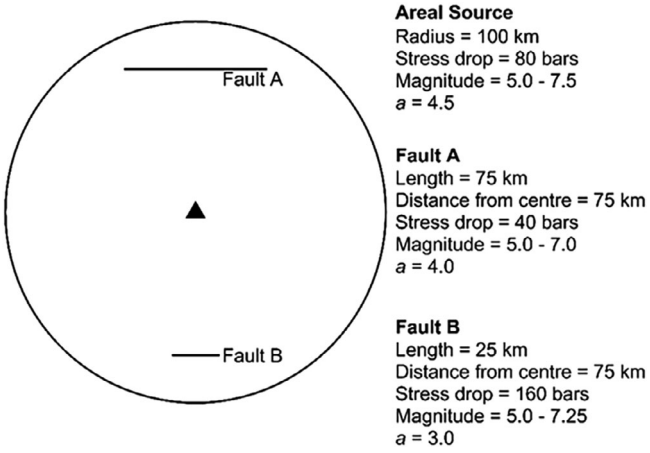

FIGURE 2 Seismic source model for the site of interest.

shear-wave velocity in the upper 30 m of the site (Vs30) and style of faulting. These are input to models estimating Arias intensity, significant duration, central frequency and frequency bandwidth; these in turn are used to simulate ground motions by filtering, scaling and windowing Fourier amplitudes in the frequency and temporal domains.

The coefficients within the stochastic model are calibrated to a recent dataset of Italian earthquakes.[32] Therefore, the seismic scenario developed in this study could be considered suitable for a high seismicity region such as central/southern Italy or California, where the most important earthquake scenarios are moderate to large events ($M > 5.5$) at short to moderate distances ($R < 50$ km). Normal faulting is considered within the stochastic model, and a fixed site with $Vs30 = 255$ m/s being used to replicate a soft-soil site. By fixing the values of the style of faulting and Vs30, only moment magnitude and source-to-site distance are required as inputs to the stochastic model. In an additional step, to allow differences in ground motion characteristics from each source, the code has been altered to allow the stress drop to be changed within the model (Figure 2).

Generating magnitude and distance inputs directly from their probability distributions would require many simulations to capture sufficient extreme events—high magnitudes and short distances. This would be computationally unaffordable and would lead to the generation of many events that would not significantly contribute to the hazard and risk at the site under investigation. To combat this, the importance sampling approach described in Jayaram and Baker[33] is used to simulate magnitudes and distances. This method involves generating a uniformly distributed sample of magnitudes and distances, then attributing an 'importance weight' to these values based on the ratio of their expected (from the original distributions) and their actual probabilities. The 'importance weight' can then be used within the analysis, so that sampling bias is not introduced into the risk assessment process.

In total 100 sets of 1000 magnitude-distance pairs are sampled at the site of interest using the importance sampling procedure, with ground motions simulated from these. This large number of ground motions being simulated allows uncertainty and variability within the risk assessment process to be modelled. We have used a similar approach in a recent study focussed on seismic hazard and risk estimates using single-degree-of-freedom systems.[34]

3 | STRUCTURAL MODEL

The structural model employed is replicated from the SAC phase 2 steel project.[35] This case study was selected as it is a well-studied structure with many previous uses in the literature, thus allowing the developed model to be validated. It is a typical three-storey office building designed to the local code for Los Angeles[36] using post-Northridge connections. The front elevation of the structure is made up of three moment-resisting bays and one simply-supported bay, all of 9.15 m in width, with each storey being 3.96 m in height.

Both plan and front elevations, as well as steel member sizes of the structure are provided in Figure 3. The typical floor dead load is 4.6 kN/m^2, the roof dead load is 4.0 kN/m^2 and the typical live load on all floors is 1.0 kN/m^2. Column steel strength is 397 MPa while the steel used for the beams has a strength of 339 MPa, as reported by Barroso.[37] Further details regarding the structural model definition can be found in many articles.[38,39]

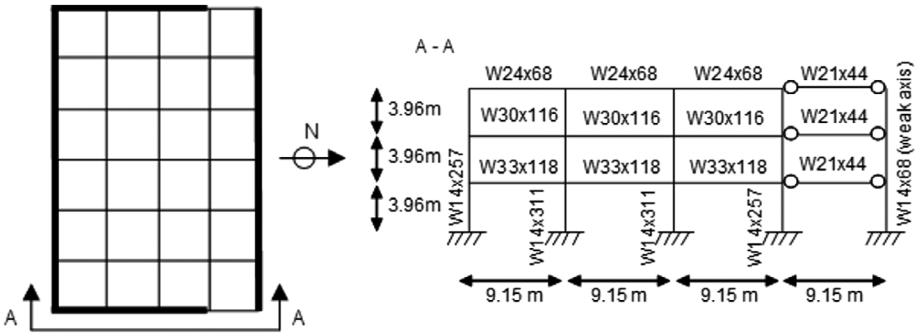

FIGURE 3 Plan and front elevations of the structural model under investigation [taken from Scozzese et al. (2020)[1]]. Thick lines highlight moment-resisting frames.

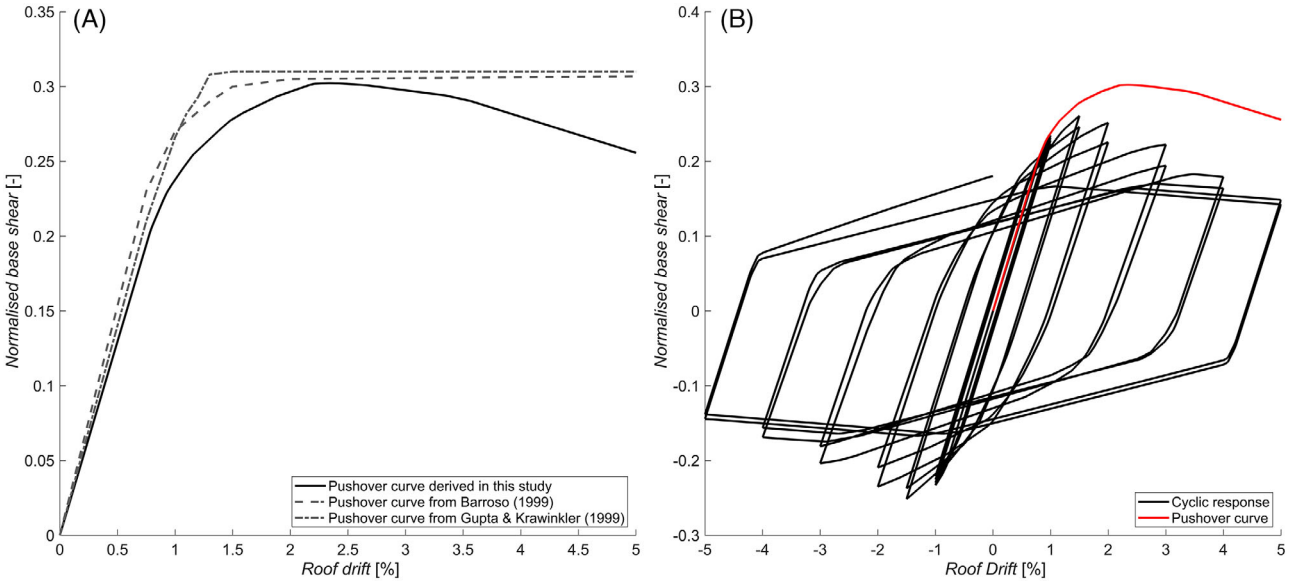

FIGURE 4 (A) Static pushover curve and (B) cyclic response, of the structure under investigation.

The structure is modelled in OpenSees,[40,41] via the open-source Frame-Modeller 2D software,[42] which facilitates modelling and analysis in OpenSees. However, for this study, Frame-Modeller 2D was only used for model development, with the outputted structural model used directly in OpenSees and MATLAB[43] for structural analysis.

The first, second and third structural periods of the model are 1.00, 0.30 and 0.14 s, respectively. The static pushover curve of the structure is presented in Figure 4A, where roof drift (roof displacement normalised by the total height) is plotted against normalised base shear (the base shear divided by the structure's self-weight). The pushover curves derived by Barroso[37] and Gupta and Krawinkler[38] are also plotted in Figure 4A for comparison. The cyclic response of the structure is presented in Figure 4B, which demonstrates the strength degradation of the structure under loading. As vibration periods and static analysis results are in good agreement with the literature,[37–39] it can be assumed that the structural model is sufficiently accurate.

Several modelling assumptions were introduced for this structure, which would likely explain any differences between the studied model and those from the literature: these assumptions also explain the more significant stiffness reduction and softening observed in the pushover curve of the developed model in Figure 4. Columns within the structure are modelled using the Ibarra–Medina–Krawinkler deterioration model[44] and a lumped plasticity approach. Deformation of the structural panel zones is accounted for using the parallelogram model discussed in Gupta and Krawinkler,[38] which allows the model to better account for degradation within the system—an important aspect to consider when evaluating ground motion duration as an *IM*. This is achieved by modelling the corner of the idealised panel zone with a nonlinear rotational spring formed with a degrading hysteretic material model. Finally, a fictitious bay that represents all the pinned/simple/gravity connections in the structure is introduced, as per Gupta and Krawinkler[38] in order to account for P-Delta effects.

TABLE 1 *IM*s under investigation within this study.

Name	Description	Units
PGA, PGV & PGD	Peak ground acceleration/velocity/displacement	g, cm/s & cm
Sa	5% damped spectral acceleration at fundamental period (T = 1.00 s)	g
Sa(avg)	Geometric mean of 5% damped spectral accelerations between periods of 0.2 and 3.0 s[15]	g
ABDur	Absolute bracketed duration using an acceleration threshold of 50 cm/s^2	s
AUDur	Absolute uniform duration using an acceleration threshold of 50 cm/s^2	s
AI	Arias intensity[12]	m/s
CAV	Cumulative absolute velocity: integral of absolute velocities[12]	cm/s
ARMS, VRMS & DRMS	Root-mean-square acceleration/velocity/displacement[46]	g, cm/s & cm
ASI	Acceleration spectrum intensity: integral of *Sa* between periods of 0.1 and 2.5 s[46]	g/s
CI	Characteristic intensity[47]	–
HI	Housner intensity: integral of the 5% damped pseudo-velocity spectrum between 0.1 and 2.5 s[46]	cm

4 | INTENSITY MEASURE SELECTION

In total 15 different *IM*s were considered within this study, with their names, descriptions and units provided in Table 1. Most of these *IM*s are well known (e.g., *PGA*, *PGV* and *PGD*) but some require clarification as to how they were defined. Further description of absolute durations (*ABDur* and *AUDur*) are available in Bommer et al.[45]

4.1 | Current practice to determine an optimal intensity measure

To determine an optimal *IM*, several indicators have been proposed. The two most used indicators were introduced by Luco and Cornell,[16] who described the concept of *efficiency* (variability in the *EDP* conditional on the *IM*) and *sufficiency* (statistical independence from earthquake characteristics, e.g., magnitude and distance). Moreover, Giovenale et al.[48] added the concept of *hazard computability* (ease of computing a hazard curve for the given *IM*), and Padgett et al.,[49] among others, included *practicality* (correlation between *IM* and *EDP*). Padgett et al.[49] also introduced the term *proficiency* (*efficiency* divided by *practicality*), and Tothong and Luco[50] introduced *scaling robustness* (lack of bias in response estimation from scaled records). For this study, *efficiency*, *magnitude-sufficiency*, *distance-sufficiency*, *practicality* and *hazard computability* are used to evaluate the impact that *IM* selection has on risk estimates.

The *efficiency* of an *IM* can be described by the standard deviation of a linear regression model fit between *IM* and *EDP*, with a smaller value indicating a more efficient *IM*. *Practicality* can be described by the slope of the same linear regression model—the higher the slope the greater the practicality. *Sufficiency* can be represented by fitting linear regression models between the residuals of the *IM* versus *EDP* model and both magnitude and distance, with the *p*-values of the regression line slopes being considered to demonstrate the *IM*s *sufficiency*. Although any *IM* with a sufficiency value above the confidence threshold (in this case 0.05) can be determined as sufficient, it could be considered that the higher the *p*-value, the lower the evidence of an insufficient *IM*.[49] Finally, *hazard computability* is assessed here by only investigating *IM*s for which a recent ground-motion model exists. These were identified by cross-referencing *IM*s within the literature, for example, using the Douglas[51] GMPE compendium.

4.2 | Proposed method to identify optimal intensity measure

As traditional *IM* performance metrics are only reliant on knowing the *IM-EDP* relationship, the impact of the *IM* choice for damage and loss assessments is not directly considered. This could potentially lead to the use of suboptimal *IM*s and demand models when performing a risk or loss assessment.

The proposed method of identifying an optimal *IM* involves comparing the conditional risk assessment procedure against the unconditional benchmark. To this end, the Kolmogorov–Smirnov (KS) test[52] is considered as an alternative to the traditional *IM* performance metrics. This test finds the maximum absolute difference ($D_{\max}$) between two cumulative

probability distributions (CDFs), as per Equation 1:

$$D_{max} = \max_{x} \left(\left| F_1(x) - F_2(x) \right| \right) \tag{1}$$

where, $F_1(x)$ and $F_2(x)$ are both CDFs as a function of, x. As a non-parametric test, the KS test does not rely on the assumption and description of a probability distribution. The KS test compares two CDFs and evaluates if they are from the same population; however, the demand hazard is in the form of mean annual frequency (MAF) of exceedance. To overcome this, a MAF of exceedance is converted to an annual probability of exceedance using Equation 2, assuming that the exceedance events follow a Poisson process:

$$P = 1 - e^{-\lambda t} \tag{2}$$

where P is the annual probability of exceeding some threshold, t is the time frame in years and λ is the MAF of exceeding the same threshold. The complement of P provides the CDF of the demand hazard. The null hypothesis of the KS test is that the two CDFs are drawn from the same population. This is evaluated by comparing the returned p-value of the test against an assumed significance level (in this case 0.05, i.e., 5%).

The investigation found that comparing the returned p-value for the KS test did not provide enough information to be able to properly evaluate and compare the performance of each *IM*. Instead, we propose to simply use $D_{\max}$ as a method of quantifying the difference between an IM-based demand hazard assessment and the unconditional assessment. As this is a measure of closeness, a smaller value of $D_{\max}$ would indicate a CDF that is more similar to the benchmark assessment, thus implying a better *IM*. These results are investigated/evaluated in Section 6.2.

An alternative approach would have been to use the relative entropy method[53,54] which employs a cumulative measure of the difference between two probability distributions. This study opts for the maximum difference between the two as it is less sensitive to the shape of the underlying distribution of the data and will highlight any large local deviations between the two distributions—which may not be fully represented by a cumulative change. The technique for calculating $D_{\max}$ is also computationally simpler, making the technique more attractive for future use.

As all of the outputs from the conditional risk assessment workflow are in the form of MAF of exceedance (as demonstrated by Figure 1), the outlined procedure can be performed at any stage of a risk assessment. This allows for a more direct comparison between the suitability of *IM*s for the specific use under question.

5 | RISK ASSESSMENT PROCEDURE

First, the evaluation of the demand hazard via the unconditional approach is described in this section. Second, each of the seismic demand models used to estimate structural response, and demand hazard, via the conditional approach are discussed. Finally, the demand hazard estimates from each of the *IM*s and each of the demand models are compared against the benchmark estimates.

5.1 | Benchmark unconditional procedure

The benchmark demand hazard (referred to from now on as the *unconditional* estimate) can be obtained using Equation 3:

$$\upsilon(EDP) = \sum_{j=1}^{sources} \lambda_{source_j} \sum_{i=1}^{N_j} \frac{I_{i,j}(edp)}{N_j} \times \frac{ISW_j}{\sum ISW} \tag{3}$$

where $\upsilon(EDP)$ is the mean annual frequency of *EDP* exceedance, λ_{source_j} describes the activity-rate of each source, N_j is the number of magnitude-distance pairs simulated for a given j-th source, $I_{i,j}(edp)$ is an indicator function equal to one if for the i^{th} record the *EDP* threshold *edp* is exceeded and zero otherwise, and ISW is the importance sampling weight for each event. Performing this sum over a range of *EDP* thresholds, and over all sets of ground-motion samples, allows the formation of a demand hazard curve, which describes the MAF of exceeding a demand hazard level.

5.2 | Conditional risk assessment procedure

Demand hazard is estimated through convolution of the hazard curve, describing the MAF of exceeding an *IM, λ(IM)*, and the complementary cumulative distribution function (CCDF) of an *EDP* conditional on the *IM*, *P(EDP|IM)*. Hazard curves are created for each *IM* through Monte Carlo hazard assessment with importance sampling, with the hazard evaluated at 50 *IM* thresholds logarithmically spaced between the minimum and maximum value of each *IM*, across all sets of samples.

After performing dynamic analyses in OpenSees, top-storey *IDR* (*IDR* between the roof and third floor) was selected as the *EDP*. This was chosen as it highlights the impact of *IM* and demand model selection; however, the following results are not strongly dependent on the choice of *EDP*. The structural response is then analysed using cloud analysis.[25] Within this study, four different regression models have been considered. They are a linear fit between *IM* and *EDP*; a more complex bilinear relationship between the predictor and response that follows the functional form of Tubaldi et al.[27]; and two simple machine learning-based regression models: a random forest (RF) and an artificial neural network (ANN). The functional form of the linear and bilinear fits are provided in Equations 2 and 3, respectively:

$$\ln(EDP) = a_1 + b_1 \ln(IM) + \sigma \tag{4}$$

$$\ln(EDP) = (a_1 + b_1 \ln(IM))H_1 + [a_1 + (b_1 - b_2)\ln IM^* + b_2 \ln IM](1 - H_1) + \sigma \tag{5}$$

where, a_1, b_1, a_2, b_2, H_1 and IM^* are coefficients fit by the regression model, and σ is the model error. a_1, b_1, a_2 and b_2 control the intercept (a) and slope (b) of the first and second segments, respectively, whilst H_1 is a step function to determine which segment *IM* lies in, finally IM^* identifies the breakpoint between the two segments. For the bilinear model σ can take two values depending on which segment the *IM* lies in.

Two machine learning models have been used in this study so that the risk estimates from one can be cross-verified by the estimates of the other. The ANN consists of a single hidden layer of three neurons, using the natural logarithm of *IM* as input to predict the natural logarithm of *EDP*. The ANN uses the Levenberg–Marquardt optimisation technique, as well as a sigmoid activation function, as both are known to be effective at solving nonlinear models. The RF model is an ensemble of decision trees that are developed using Bayesian optimisation to control their hyperparameters, this technique helps to avoid overfitting. The RF model uses the same input and output as the ANN. Relaxing the assumption of homoscedasticity was considered for the machine learning models, by estimating the model dispersion values as a function of *IM*. However, this change proved to have little impact on the accuracy of the demand models, and so heteroscedasticity was not included within either machine learning model. Instead, single (homoscedastic) dispersion σ values were calculated for both machine learning models, which equal the standard deviations of model residuals in the two cases.

ANNs and RFs have been used in various studies to create seismic demand models.[55,56] In this study, both techniques have been implemented to demonstrate the capabilities of machine learning for this application. These models could be trained with a wide range of training data, such as other ground-motion characteristics and structural information. However, in this study, the only input to the machine learning models is *IM*, allowing direct comparisons with the linear and bilinear demand models.

When investigating both duration-based *IM*s (*ABDur* and *AUDur*), it was found that a log-linear regression model better described the relationship between *IM* and *EDP*. The cloud analysis models for these *IM*s are generated by simply replacing the ln(*IM*) terms in Equations 4 and 5 with *IM*, for example, Equation 5 becomes:

$$\ln(EDP) = (a_1 + b_1 IM)H_1 + [a_1 + (b_1 - b_2)IM^* + b_2 IM](1 - H_1) + \sigma \tag{6}$$

Each cloud analysis model is shown in Figure 5 for one set of simulations with *PGA* (Figure 5A) and *PGD* (Figure 5B) as the *IM*, and top-storey *IDR* as the *EDP*. In Figure 5A, the linear cloud analysis model appears to fit the data well until 0.2 g *PGA* at which the system begins to respond nonlinearly. This will likely lead to over-prediction of seismic demand at larger ground motions. The bilinear fit improves upon this and fits the data much better within this nonlinear segment. Both machine learning models also fit the data well, with the RF and ANN providing median predictions somewhere between the linear and bilinear models, before tending toward the bilinear model in the nonlinear range. The RF appears to be simply following the median of the dataset. However, demand model estimates outside of the data range for both machine learning models are poor, as they have not been trained at these values, making extrapolation inaccurate.

The fit for all models appears to be very similar in Figure 5B. As there is no significant non-linearity, the linear model appears to match the trend in the data as successfully as the other models. However, there is a much larger variation

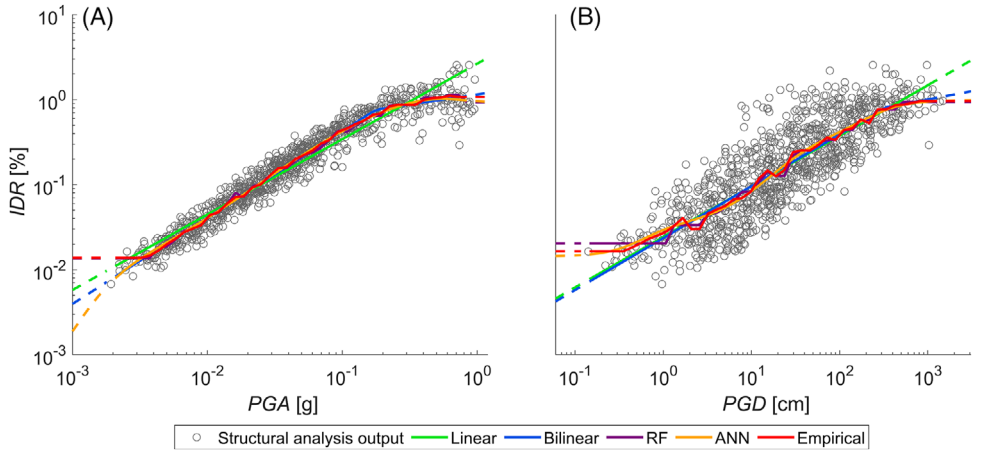

FIGURE 5 Demand models for (A) *PGA* (B) *PGD*, the *EDP* is top-storey *IDR*.

in response than is visible in Figure 5A. This implies that *PGD* is a less efficient *IM*, and that the models fitted are not as suitable as those for *PGA*. Once again, demand model estimates outside of the data range for both machine learning models are poor.

An alternative to performing cloud analysis is generating the CCDFs for the *EDP* using an empirical binning procedure. For this purpose, the *IM* is separated into 50 equally sized bins (each containing 2% of the population) between its minimum and maximum value, and the probability of exceeding a given *EDP* threshold, within each bin, calculated via Equation 7:

$$P\left(EDP > edp|IM_j\right) = \sum_{i=1}^{N_j} \frac{I_{i,j}}{N_j} \tag{7}$$

where $P(EDP > edp|IM_j)$ is the CCDF of *EDP*, given that the ground motion falls into bin IM_j, N_j is the number of ground motions that fall into bin IM_j, and $I_{i,j}$ is an indicator function that equals one if the i^{th} record in bin j is greater than the *EDP* threshold and zero otherwise.

Interpolation can then be used to estimate the CCDF for the relevant thresholds. Any CCDF value conditioned on an *IM* outside of the range of data is assigned the next nearest value for which data are available. The median of each of these bins is also plotted in Figure 5 for comparison with the cloud analysis models. The empirical model fluctuates significantly as it is tracking the local trends (i.e., following the median of each bin). This feature is expected to yield accurate estimates of the demand as the model is most responsive to the data that it is built on.

The CCDF for each regression model is found using the median model prediction at the *IM* threshold of interest, and the accompanying model σ—assuming that *P(EDP|IM)* follows a lognormal distribution. In total 50 CCDFs are created for each model, conditioning *EDP* on the same *IM* thresholds as for the hazard curves, thus making the convolution for demand hazard straightforward. Figure 6 plots fragility curves for the same set of samples, with *PGA* (Figure 6A) and *PGD* (Figure 6B) as the *IM*, and top-storey *IDR* as the *EDP*. The CCDFs are plotted for *IDR* thresholds of 0.2%, 0.7% and 1.0%. Example empirical fragility curves for the same *IDR* thresholds are also provided in Figure 6. Most demand models produce similar CCDFs, except the linear model which underpredicts the drift demand at 0.2% and 0.7%, and overpredicts the drift demand at 1.0% (in agreement with Figure 5). The variation in CCDF values between the other models is mostly due to their associated model σ. Both the RF and empirical models produce less smooth CCDFs than the other demand models; this is because their median model estimations are not always increasing (as is clear from Figure 5), and so the CCDF value will not always increase.

5.3 | Comparing mean risk estimates

Demand hazard is estimated through this procedure for all *IM*s, all seismic demand models, and all 100 sets of samples, allowing the MAF of top-storey *IDR* exceedance to be calculated. Figure 7 presents the mean of this demand hazard for

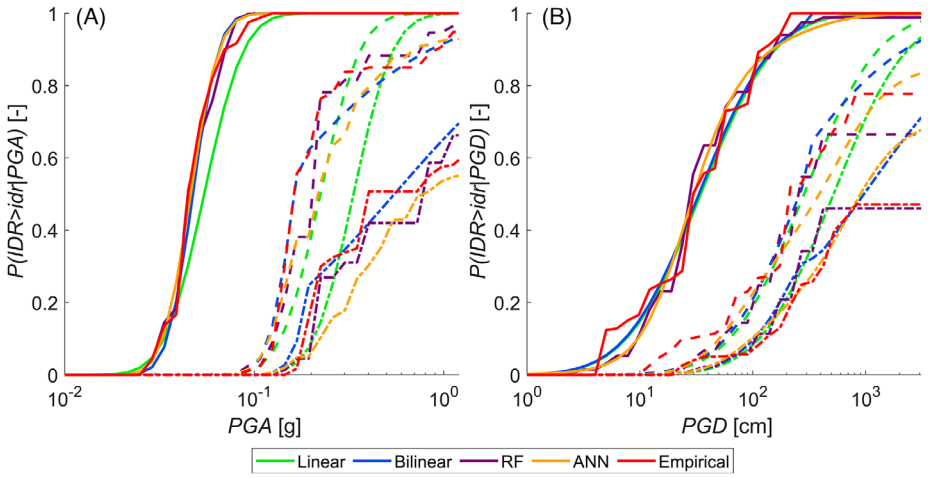

FIGURE 6 *P(EDP|IM)* curves for (A) *PGA* and (B) *PGD*, and each of the seismic demand models, at *IDR* thresholds of 0.2% (solid lines), 0.7% (dashed lines) and 1.0% (dot-dash lines).

each of the *IM*s and demand models under investigation, comparing each against the unconditional estimates. Most *IM*s, and all demand models, appear to fit relatively well until an *IDR* between 1% and 1.5%. This is approximately the range at which the *IM*|*IDR* relationship in the log-log plane becomes nonlinear due to the structural component's yielding (see pushover curve in Figure 4) and so the *IM* choice and demand model becomes more important here.

The linear model overestimates the demand hazard for all *IM*s, with no visually discernible best-fitting *IM*. The bilinear fit significantly improves upon the linear fit as the nonlinear system response is now much better accounted for, with demand hazard estimates for *Sa(avg), AI* and *CAV* in particular appearing to match the unconditional estimates most closely when compared to the other *IM*s. Both machine learning models (the RF and ANN) produce almost identical demand hazard estimates for all *IM*s across all *EDP* thresholds, with these estimates appearing to be superior to that of the bilinear model. These models even appear to improve estimates for both *ABDur* and *AUDur*, both of which could not be well described by the linear and bilinear models. However, care must be taken in the design and implementation of machine learning models, especially at the edges of the model range, where overfitting can lead to poor response estimates. Finally, the empirical model seems to estimate the unconditional demand hazard very closely, with all *IM*s appearing to be good choices using this technique. Nonetheless, this result warrants further exploration to see how well the empirical model performs when estimating the benchmark demand hazard. Moreover, Figure 7 shows it would be difficult to draw a conclusion as to the optimal *IM* for any of the demand models from visual inspection alone, so quantitative *IM* performance metrics need to be used to help decide this.

6 | INVESTIGATING IMPACT OF INTENSITY MEASURE SELECTION

In this section, traditional *IM* performance metrics are evaluated and analysed for each of the seismic demand models. After this, a novel technique to measure *IM* performance is proposed. A comparison between the different techniques is then provided and discussed.

6.1 | Traditional IM performance metrics

The *efficiency*, *magnitude-sufficiency*, *distance-sufficiency* and *practicality* of all *IM*s using the linear demand model are presented in Figure 8. These measures were assessed through the procedures described in Section 4 and repeated for all 100 sets of ground-motion samples — thus allowing the assessment of their distributions (as shown by the boxplots presented, where the boxes represent the median, 25th and 75th percentiles, and the whiskers represent the maximum and minimum values).

Overall, there appears to be no one optimal *IM* for the estimation of demand hazard that outperforms the others in terms of all the above performance metrics, although it is evident that *PGD*, *ABDur*, *AUDur* and *DRMS* rank worst of all

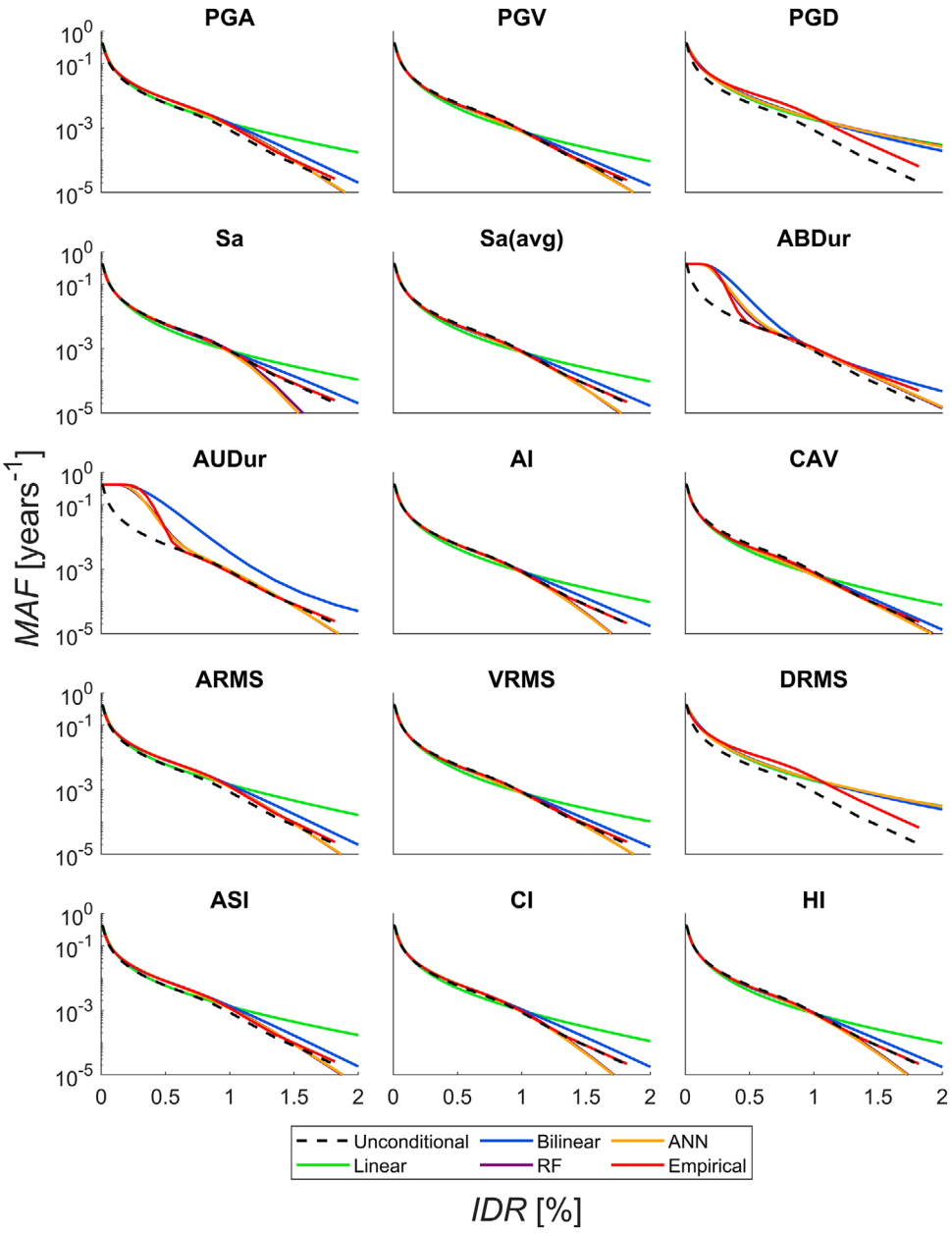

FIGURE 7 Mean demand hazard estimates for all *IM*s and seismic demand models under investigation.

*IM*s considered. A visual comparison of these measurements suggests that *Sa* performs the best out of all considered *IM*s. This is unsurprising as the structure's response is dominated by its first mode. Figure 8 shows that *Sa* is the most *efficient IM*, the third most *practical IM*, the most *magnitude-sufficient IM*, and fifth most *distance-sufficient IM*. In this article, 'most *sufficient*' means the one with the highest *p*-value, that is, the one that is furthest from failing the significance test based on a *p*-value threshold.

Figure 8D shows *CAV* to clearly be the most *practical*. Results appear to suggest that *efficiency* and *practicality* are highly correlated. Figure 8B shows that most *IM*s satisfy the *magnitude-sufficiency* threshold of 0.05, however, no *IM* is clearly the most *magnitude-sufficient*. On the other hand, Figure 8C indicates that *distance-sufficiency* is a harder requirement to achieve with 7 of the 15 *IM*s being deemed *insufficient* according to their mean values, and a much larger spread of results clearly visible. *VRMS* appears to be the most *distance-sufficient IM*.

As the assumption of homoscedasticity is relaxed for the bilinear cloud analysis model, two error terms are produced (one for each segment of the model), this makes it harder to compare *IM*s in terms of *efficiency* when the bilinear model is employed. To resolve this, an overall *efficiency* term is calculated for the bilinear model by finding the standard deviation

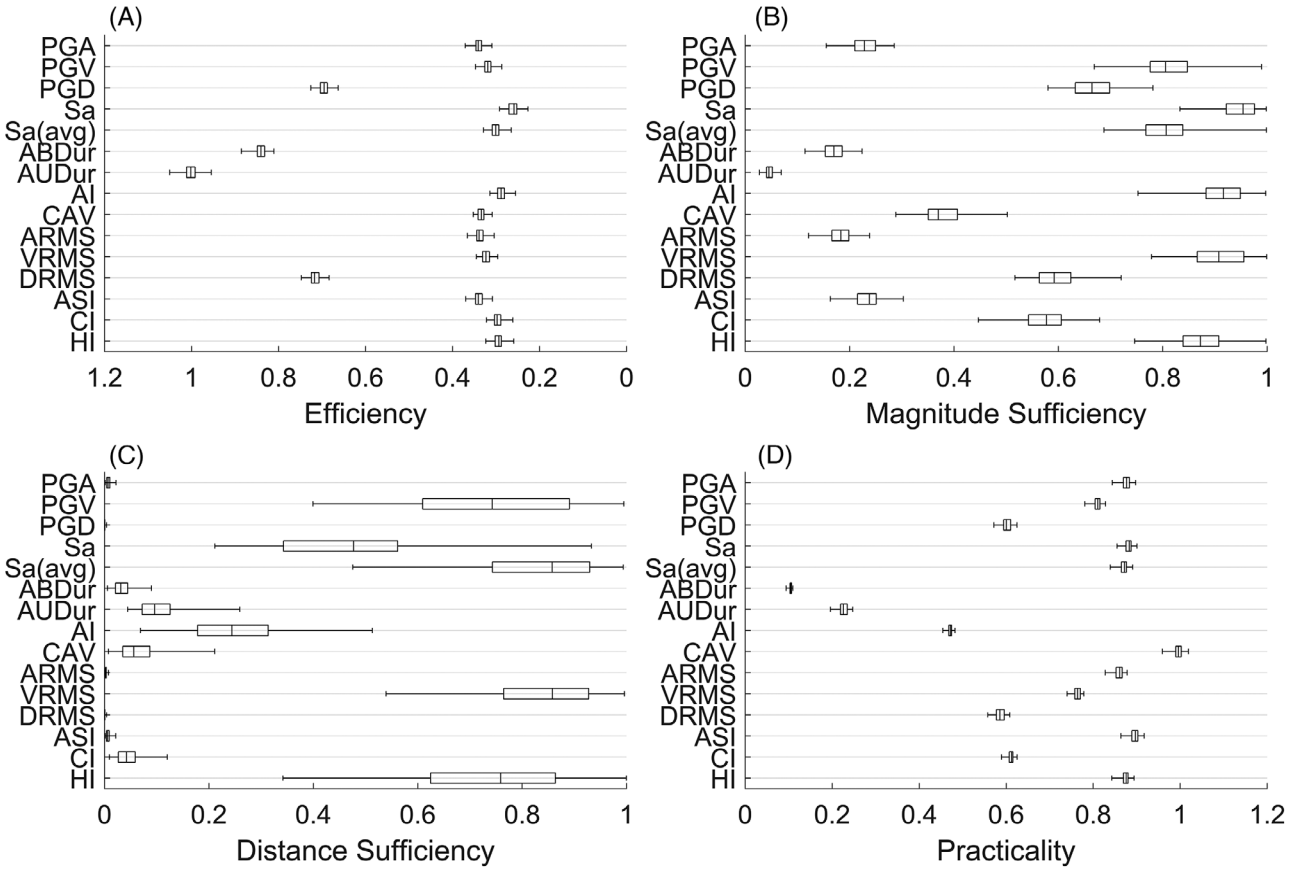

FIGURE 8 Boxplot showing the (A) *efficiency* (B) *magnitude-sufficiency* (C) *distance-sufficiency* (where *sufficiency* is the *p*-value of the slope of the residuals from the *P(EDP|IM)* model against magnitude and distance respectively) and (D) *practicality* of each IM when considering a linear demand model for the top-storey *IDR*. Note the reversed x-axis in (A) as a lower value of *efficiency* is best (smallest standard deviation).

of all residuals. Moreover, the linear model is the only demand model where the issue of *practicality* can be easily resolved as each of the other models produce multiple slopes, making it harder to compare *IM*s for this metric. As such, *practicality* is only evaluated for the linear model and ignored for all others.

Figure 9 shows how the *efficiency*, *magnitude-sufficiency* and *distance-sufficiency* of each *IM* changes for every demand model. *Efficiency* is improved for all *IM*s (except *ABDur* and *AUDur*) when using either the bilinear or machine learning demand models, reducing by 10%−25% from the linear to the bilinear model. The ANN and RF models vary just 5% from the bilinear model so return very similar *efficiency* values. *Sa* improves the most with the *efficiency* value reducing 36% from the linear model. *Magnitude* and *distance-sufficiency* are improved across all *IM*s for the bilinear model, with all *IM*s deemed *sufficient*—again except *ABDur* and *AUDur*. However, there is no consistent improvement of *sufficiency* results between the machine learning models and the linear demand model.

6.2 | Proposed performance metric

Figure 10 presents the boxplot of the $D_{\max}$ values (i.e., the difference to the unconditional results) calculated according to Equation 1 for each *IM* and each seismic demand model, for all 100 sets of simulations. From these boxplots it is clear that in general, the linear cloud analysis model produces the most different demand hazard estimates to the unconditional estimates (i.e., the least accurate). The bilinear, RF and ANN models both provide similarly close demand hazard estimates, with the ANN appearing to produce slightly superior estimates. However, the ranking of these three models for each *IM* vary, which would not have been obvious from visual inspection of the demand hazard curves alone. For all *IM*s the empirical demand model produces the most similar estimates to the unconditional approach.

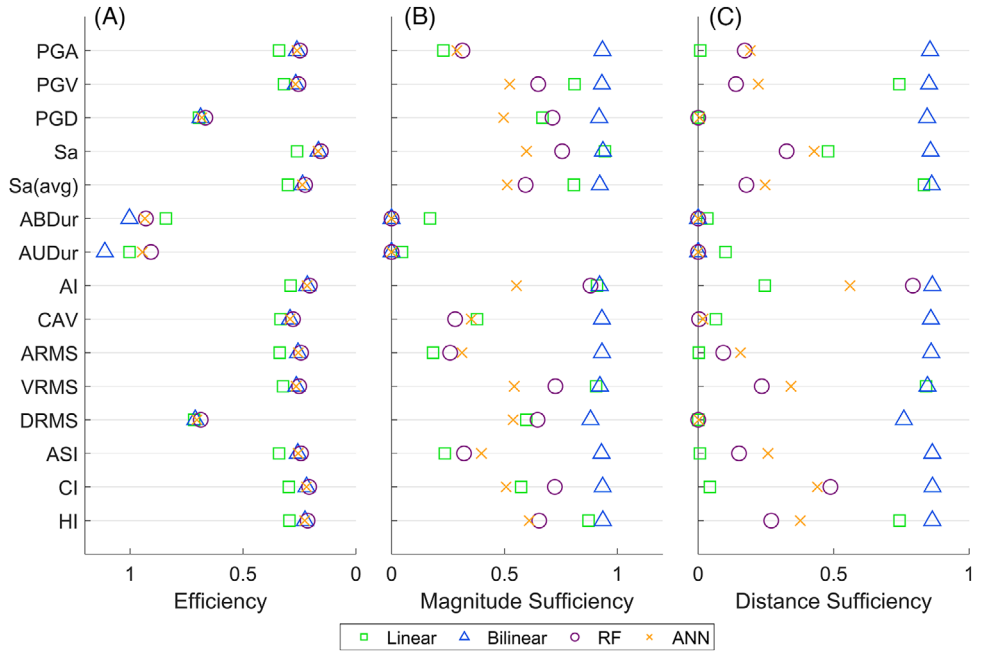

FIGURE 9 Plot showing how the mean of (A) *efficiency* (B) *magnitude-sufficiency* and (C) *distance-sufficiency* changes for each *IM* and seismic demand model combination. Note the reversed x-axis in (A) as a lower value of efficiency is best (smallest standard deviation).

As with the traditional performance metrics, it is hard to identify one optimal *IM* from these results. According to the proposed performance metric, the combination of *AI* as *IM* and the empirical demand model is the optimal one, in the sense that it yields the lowest mean $D_{\max}$ value. As has been seen throughout, both duration and displacement-based *IM*s perform considerably worse than all other *IM*s. Velocity-based measures (e.g., *PGV*, *CAV* and *VRMS*) always outperform their acceleration-based counterparts (e.g., *PGA* and *ARMS*). This is likely because the velocity-based measures capture the longer period effects of ground-motion better, leading to a better description of the demand hazard. Spectral ordinate-based *IM*s, *Sa* and *Sa(avg)*, are generally the most used *IM*s. Both *Sa* and *Sa(avg)* perform similarly well as measured by $D_{\max}$, providing the fourth and fifth overall most accurate demand hazard estimates, respectively. *Sa* provides the most accurate demand hazard estimates when using either the bilinear or ANN models, and *Sa(avg)* provides the second most accurate demand hazard estimates when using the linear model—after *CAV*. This supports their continued use throughout earthquake engineering.

6.3 | Correlation between the new and existing performance metrics

As $D_{\max}$ is a direct evaluation of the similarity of risk estimates with a benchmark truth, comparing $D_{\max}$ values for each *IM* with the traditional *IM* performance metrics can provide information on the suitability of these commonly used indicators. Linear regression analyses were carried out between the mean $D_{\max}$ value for each *IM* (using the linear demand model), and the corresponding mean values of each of the four previously examined performance metrics. R^2 values were then calculated to evaluate the correlation between $D_{\max}$ and each of the four metrics. A strong correlation was found between $D_{\max}$ and both *efficiency* and *practicality*, with R^2 values of 0.787 and 0.729, respectively. In contrast, there is little to no correlation between $D_{\max}$ and both sufficiency metrics, with R^2 values of 0.359 and 0.113 between $D_{\max}$ and *magnitude-sufficiency*, and *distance-sufficiency*, respectively.

Interestingly, if the inaccurate *IM*s (i.e., *ABDur*, *AUDur*, *PGD* and *DRMS*) are removed from this correlation calculation, the R^2 values change considerably. R^2 values for *Efficiency* and *practicality* fall to 0.260 and 0.003, respectively, whilst the R^2 values for *magnitude-sufficiency* and *distance-sufficiency* rise to 0.602 and 0.492, respectively. This indicates that when considering only accurate *IM*s, *efficiency* and *practicality* are poor predictors of *IM*-optimality and both *sufficiency* values are acceptable predictors. This suggests that *IM*s that perform well in terms of traditional *IM* performance metrics do not necessarily yield accurate demand hazard estimates. It is noteworthy that in the demand hazard assessment using the conditional demand approach, a very high number of records (i.e., 1000) has been considered for building the demand

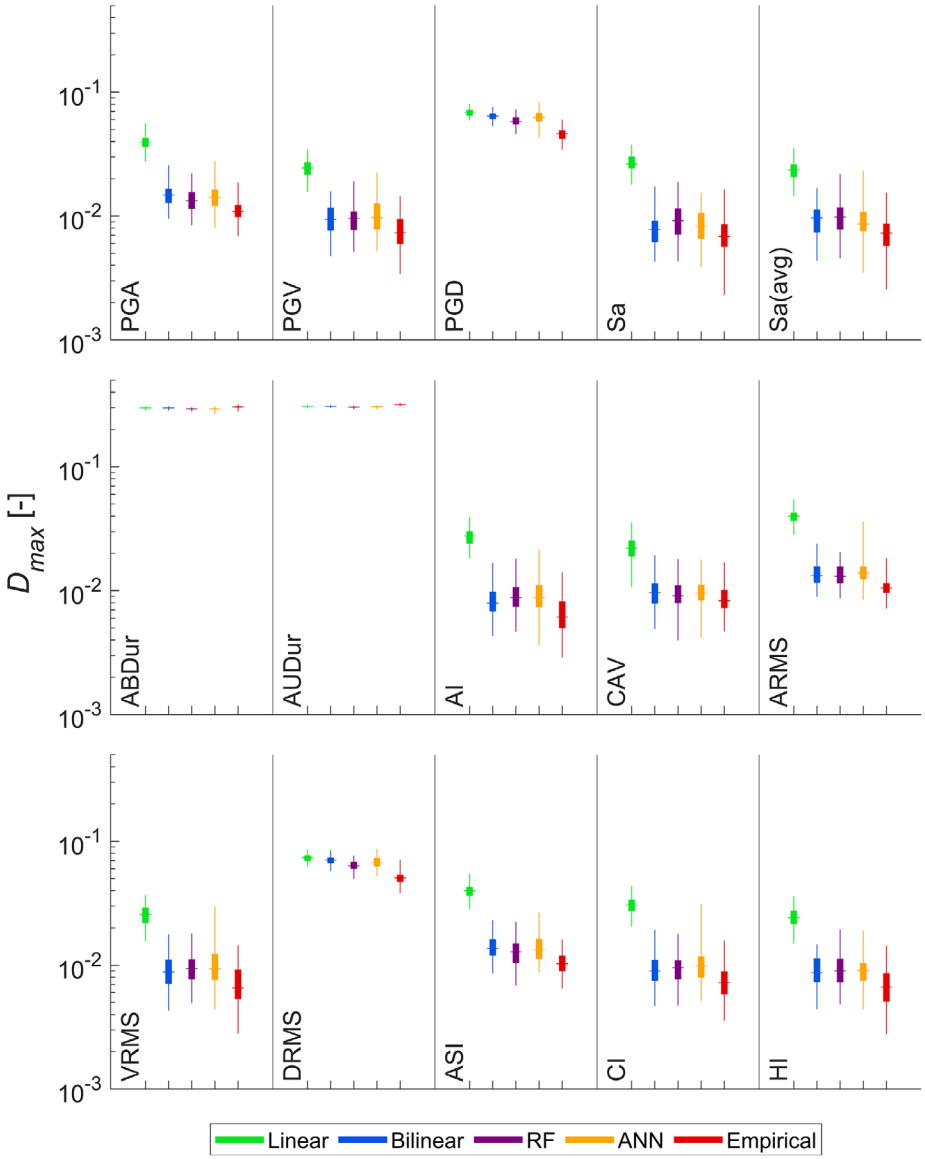

FIGURE 10 Boxplot showing D_{max} for each *IM* and each seismic demand model.

models, which is much higher than the number of records typically used in cloud analysis. Thus, it is necessary to evaluate how the obtained results are affected by the number of records, which is investigated in the next section.

7 | SENSITIVITY ANALYSIS

To further investigate the optimal *IM* and demand model combination, a sensitivity analysis was performed, evaluating how changing the number of ground motion samples affects the reported D_{max} value, for every *IM* and demand model. For this purpose, the mean D_{max} was calculated using the risk estimation procedures previously described, for samples between 50 and 1000 records. Figure 11 shows the sensitivity analysis results for (A) *Sa* and (B) *AI* for all five demand models. The sensitivity of estimates for these *IM*s is considered representative of all other *IM*s, except the displacement (*PGD* and *DRMS*) and duration (*ABDur* and *AUDur)* based *IM*s — where risk estimates remain inaccurate irrespective of the number of samples.

As already demonstrated, the empirical model is the superior demand model when all 1000 samples are used, yielding a mean D_{max} value 8% lower than the next best model for *Sa* and 21% for *AI*. However, for *Sa* the empirical model becomes less accurate than the bilinear model when less than 900 samples are used, and this is also true for *AI* when less than 700

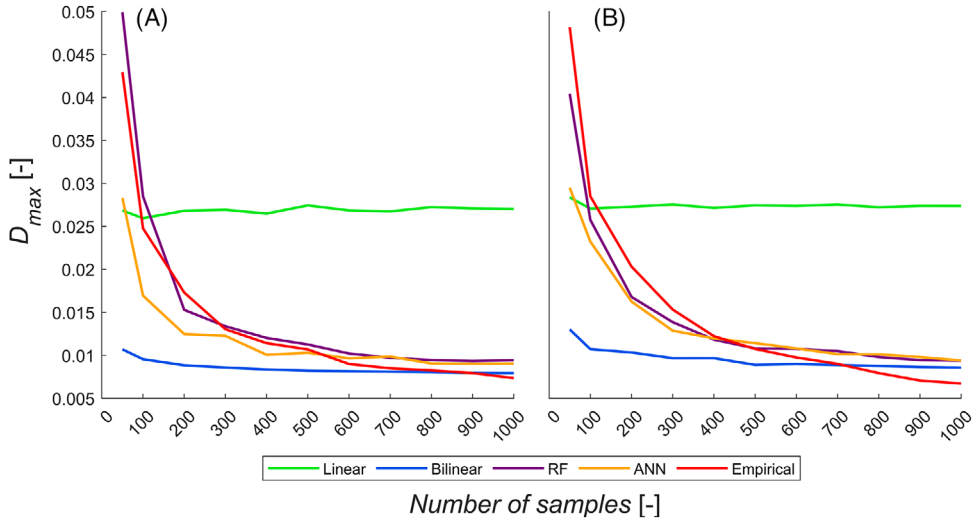

FIGURE 11 Sensitivity of $D_{\max}$ to the number of samples used in the risk assessment procedure. The *IM*s used are (A) *Sa* and (B) *AI*, which are considered representative of all accurate IMs.

samples are used. When just 50 samples are used the empirical model is the second least accurate model for *Sa*, with the RF model being the least accurate. Whilst, for *AI* when 50 samples are used the empirical model is considerably the least accurate demand model, with a $D_{\max}$ value 20% higher than the next worst. This result is expected as a smaller sample size would mean fewer records in each bin, thus limiting the methods ability to account for variability in conditional demand and reducing its accuracy.

The linear model is by far the worst model when all 1000 samples are used; however, this demand model is very insensitive to the record number, and it is the second best one for both *Sa* and *AI* when just 50 samples are used. Both machine learning models and the bilinear models perform similarly with large amounts of samples—in agreement with Figure 10. However, the machine learning models are far more sensitive and have a large drop in accuracy when using less than 200 samples for *Sa*, and 300 samples for *AI*—making these methods unsuitable for small sample sizes. Restructuring, or using more complex architectures for the machine learning models, may improve the demand estimates at smaller sample sizes. From this sensitivity analysis it can be concluded that the empirical demand model should only be used when enough samples are available, otherwise users are recommended to consider a bilinear demand model when assessing risk.

8 | COMPONENT DAMAGE ANALYSIS

To fully demonstrate the impact that both *IM* and demand model selection have on risk assessment, the demand hazard estimates are extended to damage estimates on typical components within the structure under investigation. The FEMA P-58 Normative Quantity Estimation Tool[11] is used to populate the structural model with typical structural and non-structural components, based on the structure's size and occupancy. Of these components, damage estimates from two are presented within this article: C2011.001b, a prefabricated steel staircase, and C1011.001a, a wall partition, both components are affected by *IDR*. These components are located on the top storey of the structure. The FEMA P-58 fragility database provides fragility information for both components—with the fragility curves for each plotted in Figure 12. Each DS is sequential, meaning that each DS can only occur if the preceding one has already occurred.

Fragility curves are convolved with the demand hazard estimates from Section 5.3 to produce estimates of the MAF of exceedance of each DS for each component. This is repeated for all combinations of *IM* and demand model, and all 100 sets of simulated ground motions, allowing the mean result to be plotted along with the corresponding uncertainties. The unconditional damage estimate is also produced through a convolution with the fragility curves.

Figure 13 presents estimates for the MAF of exceedance of each component's DSs using the empirical demand model, as this has been found to be the most accurate when using all 1000 samples. All except the displacement (*PGD* and *DRMS*) and duration (*ABDur* and *AUDur*) based *IM*s have been plotted in Figure 13, with these four *IM*s excluded due to their poor accuracy in terms of demand hazard. The mean damage estimate for each DS is plotted by the black circle, with error bars denoting the 16th and 84th percentiles of the mean. The unconditional damage hazard for each DS is plotted by the

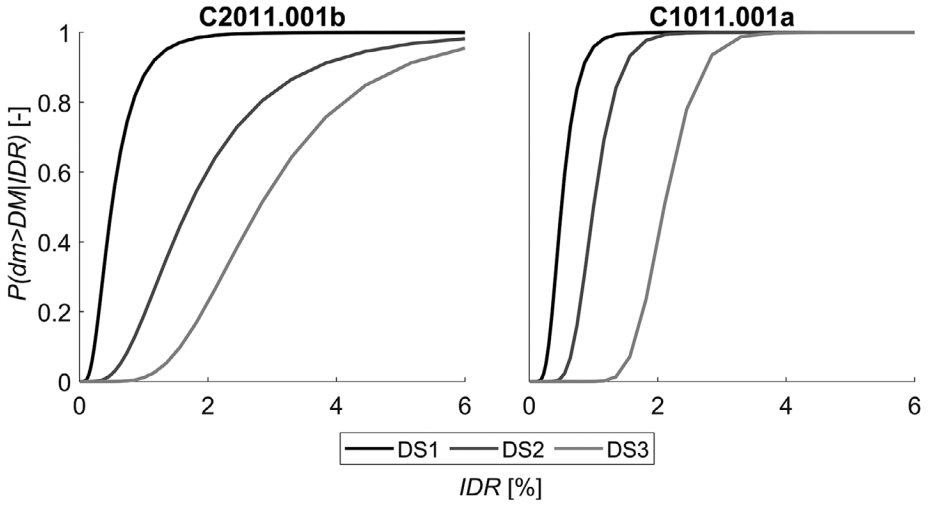

FIGURE 12 Fragility curves for both components under investigation.

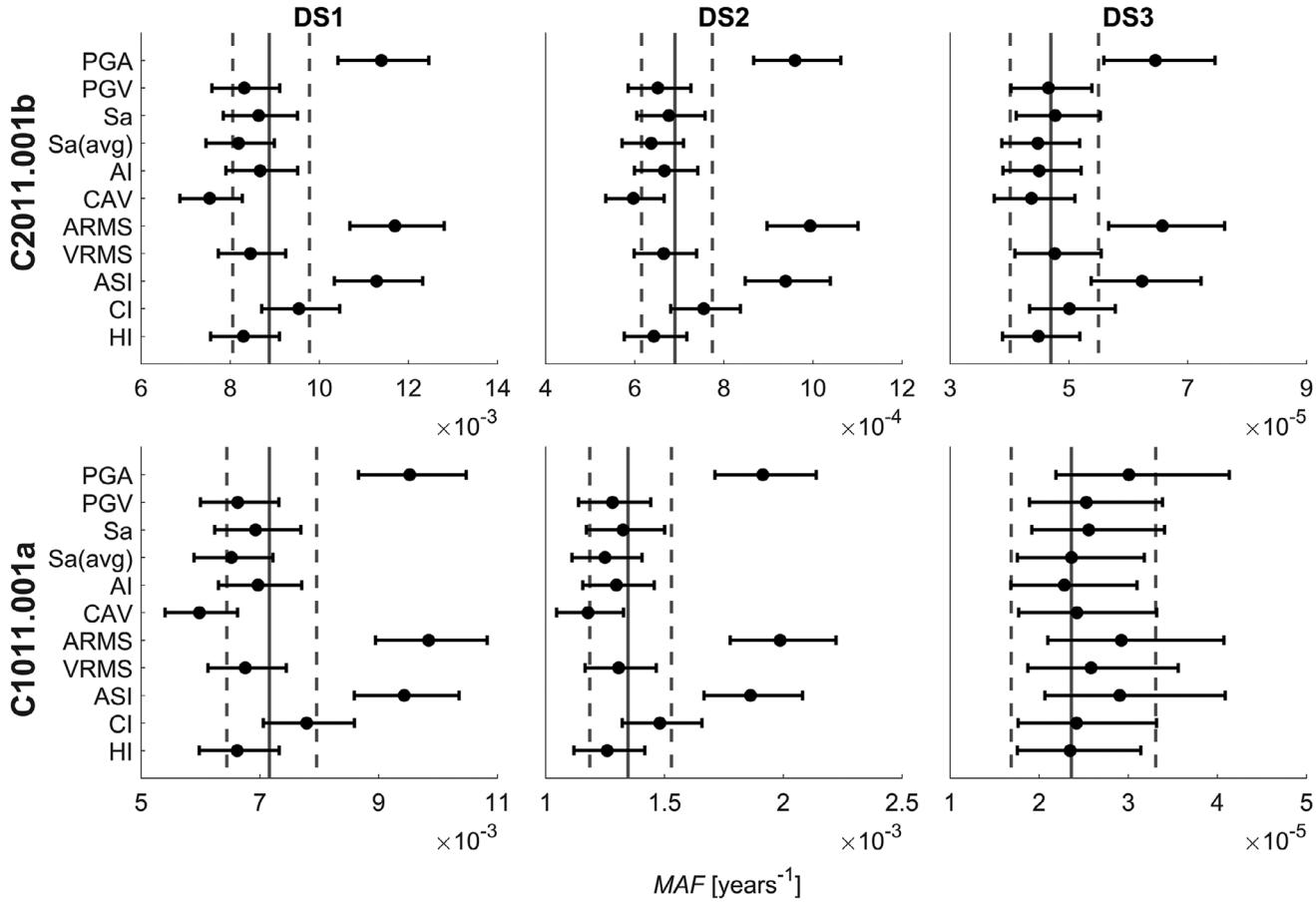

FIGURE 13 MAF of exceedance of each DS for both components using the empirical demand model, with 16th and 84th percentiles plotted as error bars. The unconditional estimate is marked by the solid (mean) and dashed (16th and 84th percentile) lines.

solid line (mean) and dashed lines (16th and 84th percentiles) for comparison against the estimates from the conditional demand models.

These results show how an improper choice of *IM* can lead to poor estimates of risk, with the inaccuracies of the demand hazard stage propagated through to the damage estimates. For instance, acceleration-based *IM*s (*PGA*, *ASI* and *ARMS*) overpredict all DSs for C2011.001b, and DS1 and DS2 of C1011.001a, a result that would not have been clearly indicated

from the demand hazard results and traditional *IM* performance metrics. For both components, *Sa*, *AI* and *VRMS* appear to give the overall most accurate estimates of damage hazard for each DS—although *CAV*, *CI*, and *HI* give considerably better damage hazard estimates for DS3 than for DS1 or DS2. This demonstrates how it is difficult to determine a single comprehensive *IM* with which to perform a risk assessment and reaffirms that it is not necessarily best practice to determine *IM* optimality at the demand hazard level.

9 | CONCLUSIONS

The impact of the combined selection of intensity measure and seismic demand model on demand hazard estimates has been evaluated for a mid-rise steel structure. Using a fictive scenario and stochastic ground motion simulations, we first estimated a benchmark demand hazard for the system via the unconditional approach, before making conditional demand hazard estimates based on 15 different intensity measures, and five conditional seismic demand models using cloud analysis—four developed through regression and a fifth through an empirical binning technique. Visual comparisons were made for all 75 demand hazard estimates against the unconditional approach, with further investigation performed by evaluating traditional performance metrics: *efficiency*, *sufficiency* and *practicality*.

After this, we presented a novel performance metric to assess the accuracy of intensity measures, and demand modelling choices, based on the maximum distance between the demand hazard estimates obtained using the conditional approach and the unconditional benchmark. This metric can be applied to any level of a conditional risk assessment framework, providing the estimates are in the form of frequencies of exceedance, and could also be used to evaluate a wider range of risk assessment modelling choices. A sensitivity analysis was performed, using the new metric, to evaluate how the number of records used within the demand model impacted the quality of demand hazard estimates. Finally, damage estimates were made for two typical components to evaluate how the choice of intensity measure and demand model can impact the latter stages of a loss assessment.

From this study, the following main conclusions can be drawn for the scenario under investigation:

- Intensity measures based on ground displacements and duration performed consistently poorly at estimating risk.
- Our proposed metric demonstrated that a combination of Arias intensity with an empirical demand model provided the most accurate demand hazard estimates, but a more traditional intensity measure such as the spectral acceleration at the fundamental vibration period of the structure also performed quite well.
- Comparing our new measure against the traditional performance metrics indicated that optimal intensity measures in terms of the traditional performance metrics did not necessarily guarantee accurate demand hazard estimates when using a large set of records for developing the demand model.
- The sensitivity analysis revealed that when using more than 700 samples, Arias intensity with the empirical demand model still provided the most accurate combination of intensity measure and demand model. Yet, when fewer than 700 records were used the most accurate combination was the spectral acceleration at the fundamental vibration period of the structure alongside a bilinear seismic demand model.
- For all intensity measures, the empirical demand model was superior when sample sizes were large. However, if very few records were used to develop the demand model, the bilinear model provided the best demand hazardestimates. The bilinear model is a rational choice for representing the variation of the seismic demand with the intensity measure when considering structures like the one analysed whose response is dominated by the first mode of vibration and exhibits a clear transition between a linear and nonlinear phase.
- Ground acceleration-based intensity measures were found to be poor estimators of the damage state, whilst spectral acceleration and Arias intensity continued to perform well.

It is important to caveat the findings of this study due to its relatively narrow scope. Only two components are considered that are both affected by the same single demand parameter, in one structure, and results from other types of demand models are not evaluated, that is, through incremental dynamic analysis or multiple stripe analysis. A different demand model and intensity measure combination may be optimal when considering a different risk assessment context; likewise, the intensity measures shown to be inaccurate within this scenario may perform better in another. Further research would be needed to make more generalised conclusions on the ability of different combinations of intensity measures and demand models to estimate seismic risk. This article has, however, demonstrated the importance of selecting both the

intensity measure and the demand model in a joint fashion, and has presented a clear method of doing this that can easily be applied to different contexts.

Future studies could expand on this work by using the same methodology on a variety of different structures and consider a range of different demand parameters. Further work could also extend damage estimates to evaluate the impact that intensity measure and seismic demand model selection has on seismic loss estimates. It would also be useful for future research to explore whether using different ground-motion models affects the conclusions of this study.

In conclusion, it is essential that the optimal combination of conditioning intensity measure and seismic demand model are selected to ensure earthquake losses are accurately described. The approach presented in this article provides an objective way of choosing the most appropriate combination of intensity measure and seismic demand model for any seismic risk and loss assessment.

ACKNOWLEDGEMENTS

This research was funded by a University of Strathclyde studentship to the first author. An earlier version of this study was presented at the 18th World Conference on Earthquake Engineering (WCEE) 2024.[57] We thank two anonymous reviewers for their detailed comments on a previous version of this study.

DATA AVAILABILITY STATEMENT

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

ORCID

Archie Rudman https://orcid.org/0009-0003-7119-4498
Enrico Tubaldi https://orcid.org/0000-0001-8565-8917
John Douglas https://orcid.org/0000-0003-3822-0060
Fabrizio Scozzese https://orcid.org/0000-0002-7378-1763

REFERENCES

1. Scozzese F, Tubaldi E, Dall'Asta A. Assessment of the effectiveness of multiple-stripe analysis by using a stochastic earthquake input model. *Bull Earthquake Eng*. 2020;18(7):3167-3203. doi:10.1007/s10518-020-00815-1
2. Bazzurro P, Cornell CA, Shome N, Carballo JE. Three proposals for characterizing MDOF nonlinear seismic response. *J Struct Eng*. 1998;124(11):1281-1289. doi:10.1061/(ASCE)0733-9445(1998)124:11(1281)
3. Bradley BA, Burks LS, Baker JW. Ground motion selection for simulation-based seismic hazard and structural reliability assessment. *Earthq Eng Struct Dyn*. 2015;44(13):2321-2340. doi:10.1002/eqe.2588
4. Cornell CA. On earthquake record selection for nonlinear dynamic analysis. In: The esteva symposium, Mexico, August 2005.
5. 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;128:526-533.
6. Bradley BA. Practice-oriented estimation of the seismic demand hazard using ground motions at few intensity levels. *Earthq Eng Struct Dyn*. 2013;42(14):2167-2185. doi:10.1002/eqe.2319
7. Moehle JP, Deierlein GG. A framework methodology for performance-based earthquake engineering. Paper presented at: 13th World Conference on Earthquake Engineering; 2004; Vancouver, B.C., Canada.
8. Cornell CA, Krawinkler H. Progress and challenges in seismic performance assessment. *PEER Center News*. 2000;3:1-13.
9. Cremen G, Baker JW. A methodology for evaluating component-level loss predictions of the FEMA P-58 seismic performance assessment procedure. *Earthq Spectra*. 2019;35(1):193-210. doi:10.1193/031618eqs061m
10. Taghavi S, Miranda E. *Response Assessment of Nonstructural Building Elements. PEER Report 2003–05*; Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA. 2003.
11. (FEMA) FEMA. *Seismic Performance Assessment of Buildings, Volume 1, Methodology*. Federal Emergency Management Agency Washington, DC. 2012. FEMA P-58-1.
12. Baker J, Bradley B, Stafford P. *Seismic Hazard and Risk Analysis*. Cambridge University Press; 2021.
13. Baker JW, Allin Cornell C. A vector-valued ground motion intensity measure consisting of spectral acceleration and epsilon. *Earthq Eng Struct Dyn*. 2005;34(10):1193-1217. doi:10.1002/eqe.474
14. Gehl P, Seyedi DM, Douglas J. Vector-valued fragility functions for seismic risk evaluation. *Bull Earthquake Eng*. 2013;11(2):365-384. doi:10.1007/s10518-012-9402-7
15. Vamvatsikos D, Cornell CA. Developing efficient scalar and vector intensity measures for IDA capacity estimation by incorporating elastic spectral shape information. *Earthq Eng Struct Dyn*. 2005;34(13):1573-1600. doi:10.1002/eqe.496
16. Luco N, Cornell CA. Structure-specific scalar intensity measures for near-source and ordinary earthquake ground motions. *Earthq Spectra*. 2007;23(2):357-392. doi:10.1193/1.2723158

17. Katsanos EI, Sextos AG, Manolis GD. Selection of earthquake ground motion records: a state-of-the-art review from a structural engineering perspective. *Soil Dyn Earthquake Eng*. 2010;30(4):157-169. doi:10.1016/j.soildyn.2009.10.005
18. Mackie K, Stojadinovic B. *Seismic Demands for Performance-Based Design of Bridges, PEER Report 2003–16*; Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA. 2003.
19. Bray JD, Travasarou T. Simplified procedure for estimating earthquake-induced deviatoric slope displacements. *J Geotech Geoenviron Eng*. 2007;133(4):381-392. doi:10.1061/(ASCE)1090-0241(2007)133:4(381)
20. Kohrangi M, Vamvatsikos D, Bazzurro P. Implications of intensity measure selection for seismic loss assessment of 3-D buildings. *Earthq Spectra*. 2016;32(4):2167-2189. doi:10.1193/112215eqs177m
21. O'Reilly GJ. Seismic intensity measures for risk assessment of bridges. *Bull Earthquake Eng*. 2021;19(9):3671-3699. doi:10.1007/s10518-021-01114-z
22. Du A, Padgett JE, Shafieezadeh A. Influence of intensity measure selection on simulation-based regional seismic risk assessment. *Earthq Spectra*. 2020;36(2):647-672. doi:10.1177/8755293019891717
23. Kwong NS, Chopra AK, McGuire RK. Evaluation of ground motion selection and modification procedures using synthetic ground motions. *Earthq Eng Struct Dyn*. 2015;44(11):1841-1861. doi:10.1002/eqe.2558
24. Kwong NS, Chopra AK, McGuire RK. A framework for the evaluation of ground motion selection and modification procedures. *Earthq Eng Struct Dyn*. 2015;44(5):795-815. doi:10.1002/eqe.2502
25. Mackie K, Stojadinovic B. Comparison of incremental dynamic, cloud, and stripe methods for computing probabilistic seismic demand models. In Structures Congress 2005; Metropolis and Beyond. 2005:1-11.
26. Vamvatsikos D, Cornell CA. Incremental dynamic analysis. *Earthq Eng Struct Dyn*. 2002;31(3):491-514. doi:10.1002/eqe.141
27. Tubaldi E, Freddi F, Barbato M. Probabilistic seismic demand model for pounding risk assessment. *Earthq Eng Struct Dyn*. 2016;45(11):1743-1758. doi:10.1002/eqe.2725
28. O'Reilly GJ, Monteiro R. Probabilistic models for structures with bilinear demand-intensity relationships. *Earthq Eng Struct Dyn*. 2019;48(2):253-268. doi:10.1002/eqe.3135
29. Freddi F, Padgett JE, Dall'Asta A. Probabilistic seismic demand modeling of local level response parameters of an RC frame. *Bull Earthquake Eng*. 2017;15(1):1-23. doi:10.1007/s10518-016-9948-x
30. Soleimani F, Hajializadeh D. State-of-the-art review on probabilistic seismic demand models of bridges: machine-learning application. *Infrastructures*. 2022;7(5):64. doi:10.3390/infrastructures7050064
31. Sabetta F, Pugliese A, Fiorentino G, Lanzano G, Luzi L. Simulation of non-stationary stochastic ground motions based on recent Italian earthquakes. *Bull Earthquake Eng*. 2021;19(9):3287-3315. doi:10.1007/s10518-021-01077-1
32. Lanzano G, Luzi L, Pacor F, et al. A revised ground-motion prediction model for shallow crustal earthquakes in Italy. *Bull Seismol Soc Am*. 2019;109(2):525-540. doi:10.1785/0120180210
33. Jayaram N, Baker JW. Efficient sampling and data reduction techniques for probabilistic seismic lifeline risk assessment. *Earthq Eng Struct Dyn*. 2010;39(10):1109-1131. doi:10.1002/eqe.988
34. Rudman A, Douglas J, Tubaldi E. The assessment of probabilistic seismic risk using ground-motion simulations via a Monte Carlo approach. *Nat Hazards*. 2024;120:6833-6852.
35. SAC. *SAC Steel Project*; Technical Office, Richmond, Calif. 1994; 94804-4698. http://www.sacsteel.org/
36. UBC. *Structural Engineering Design Provisions*. Vol. 2; 1994. *Uniform Building Code*. International Conference of Building Officials. Whittier, California.
37. Barroso L, Smith HA. *Performance evaluation of vibration controlled steel structures under seismic loading*. John A Blume Earthquake Engineering Center Technical Report, Stanford Digital Repository, 133; 1999. http://purl.stanford.edu/fb322mt0265
38. Gupta A, Krawinkler H. Seismic demands for performance evaluation steel moment resisting frame structures. John A Blume Earthquake Engineering Center Technical Report, Stanford Digital Repository, 132; 1999. http://purl.stanford.edu/fm826wn5553
39. Ohtori Y, Christenson RE, Spencer BF, Dyke SJ. Benchmark control problems for seismically excited nonlinear buildings. *J Eng Mech*. 2004;130(4):366-385. doi:10.1061/(ASCE)0733-9399(2004)130:4(366)
40. Mazzoni S, McKenna F, Scott MH, Fenves GL. OpenSees command language manual. *Pacific earthquake engineering research (PEER) center*. 2006;264(1):137-158.
41. McKenna F. OpenSees: a framework for earthquake engineering simulation. *Comput Sci Eng*. 2011;13(4):58-66. doi:10.1109/MCSE.2011.66
42. Elkady A. FM-2D—open-source platform for the 2-dimensional numerical modeling and seismic analysis of buildings. *SoftwareX*. 2022;17:100927. doi:10.1016/j.softx.2021.100927
43. MATLAB. Version 9.14.0.2254940 (R2023a) Update 2. The MAthWorks Inc; Natick, Massachusetts: The MathWorks Inc. 2023. https://www.mathworks.com
44. Ibarra LF, Medina RA, Krawinkler H. Hysteretic models that incorporate strength and stiffness deterioration. *Earthq Eng Struct Dyn*. 2005;34(12):1489-1511. doi:10.1002/eqe.495
45. Bommer JJ, Stafford PJ, Alarcón JE. Empirical equations for the prediction of the significant, bracketed, and uniform duration of earthquake ground motion. *Bull Seismol Soc Am*. 2009;99(6):3217-3233. doi:10.1785/0120080298
46. Kramer SL. *Geotechnical Earthquake Engineering*. Prentice-Hall; 1996.
47. Park YJ, Ang AHS, Wen YK. Seismic damage analysis of reinforced concrete buildings. *J Struct Eng*. 1985;111(4):740-757. doi:10.1061/(ASCE)0733-9445(1985)111:4(740)

48. Giovenale P, Cornell CA, Esteva L. Comparing the adequacy of alternative ground motion intensity measures for the estimation of structural responses. *Earthq Eng Struct Dyn*. 2004;33(8):951-979. doi:10.1002/eqe.386
49. Padgett JE, Nielson BG, DesRoches R. Selection of optimal intensity measures in probabilistic seismic demand models of highway bridge portfolios. *Earthq Eng Struct Dyn*. 2008;37(5):711-725. doi:10.1002/eqe.782
50. Tothong P, Luco N. Probabilistic seismic demand analysis using advanced ground motion intensity measures. *Earthq Eng Struct Dyn*. 2007;36(13):1837-1860. doi:10.1002/eqe.696
51. Douglas J. Ground motion prediction equations 1964–2021. http://www.gmpe.org.uk
52. Stephens MA. EDF statistics for goodness of fit and some comparisons. *J Am Statist Assoc*. 1974;69(347):730-737. doi:10.2307/2286009
53. Jalayer F, Beck JL, Zareian F. Analyzing the sufficiency of alternative scalar and vector intensity measures of ground shaking based on information theory. *J Eng Mech*. 2012;138(3):307-316. doi:10.1061/(ASCE)EM.1943-7889.0000327
54. Tsioulou A, Galasso C. Information theory measures for the engineering validation of ground-motion simulations. *Earthq Eng Struct Dyn*. 2018;47(4):1095-1104. doi:10.1002/eqe.3015
55. Soleimani F, Liu X. Artificial neural network application in predicting probabilistic seismic demands of bridge components. *Earthq Eng Struct Dyn*. 2022;51(3):612-629. doi:10.1002/eqe.3582
56. Li W, Huang Y, Xie Z. Machine learning-based probabilistic seismic demand model of continuous girder bridges. *Adv Civ Eng*. 2022;2022:3867782. doi:10.1155/2022/3867782
57. Rudman A, Douglas J, Tubaldi E, Scozzese F, Turchetti F, Evaluating the impact of intensity measure selection on conditional seismic risk. Paper presented at: 18th World Conference on Earthquake Engineering, Milan, Italy; 2024.

How to cite this article: Rudman A, Tubaldi E, Douglas J, Scozzese F. The impact of the choice of intensity measure and seismic demand model on seismic risk estimates with respect to an unconditional benchmark. *Earthquake Engng Struct Dyn*. 2024;53:4183–4202. https://doi.org/10.1002/eqe.4208