University of Strathclyde Glasgow

Improving earthquake risk assessment by using richer hazard inputs and better damage outputs

Archie Rudman

Department of Civil and Environmental Engineering

University of Strathclyde

A THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

'DOCTOR OF PHILOSOPHY'

Glasgow, United Kingdom

2026

Declaration of authenticity and author's rights

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed:

Date: 23/01/2026

Statement of co-authorship

The following people and institutions contributed to publications which are part of this thesis:

Candidate	Archie Rudman, University of Strathclyde
Author 1	Enrico Tubaldi, University of Strathclyde
Author 2	John Douglas, Frazer-Nash Consultancy (formerly University of Strathclyde)
Author 3	Fabrizio Scozzese, University of Camerino
Author 4	Roberto Gentile, University College London

Publications in this thesis and contributions of each author:

1. **Paper 1 (Chapter 3 of this thesis)**

   Rudman, A., Douglas, J. & Tubaldi, E. The assessment of probabilistic seismic risk using ground-motion simulations via a Monte Carlo approach. *Nat Hazards* **120**, 6833–6852 (2024). https://doi.org/10.1007/s11069-024-06497-1

   Archie Rudman: Conceptualisation; Formal Analysis; Investigation; Software; Writing – Original Draft Preparation; Writing – Review & Editing.

   John Douglas: Conceptualisation; Supervision; Writing – Review & Editing.

   Enrico Tubaldi: Conceptualisation; Supervision; Writing – Review & Editing.

2. **Paper 2 (Chapter 4 of this thesis)**

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

Archie Rudman: Conceptualisation; Formal Analysis; Investigation; Software; Writing – Original Draft Preparation; Writing – Review & Editing.

Enrico Tubaldi: Conceptualisation; Supervision; Writing – Review & Editing.

John Douglas: Conceptualisation; Supervision; Writing – Review & Editing.

Fabrizio Scozzese: Conceptualisation; Writing – Review & Editing.

3. **Paper 3 (Chapter 5 of this thesis)**

Rudman, A., Tubaldi, E., Gentile, R., Douglas, J. 1 Covariance Structure Modelling of Engineering Demand Parameters in Cloud-Based Seismic Analysis. *Earthquake Engng Struct Dyn*. 2026; 55: 1533–1551. https://doi.org/10.1002/eqe.70151

Archie Rudman: Conceptualisation; Formal Analysis; Investigation; Software; Writing – Original Draft Preparation; Writing – Review & Editing.

Enrico Tubaldi: Conceptualisation; Supervision; Writing – Review & Editing.

Roberto Gentile: Conceptualisation; Writing – Review & Editing.

John Douglas: Writing – Review & Editing.

Acknowledgements

First and foremost, I would like to thank my PhD supervisors, John Douglas and Enrico Tubaldi. Throughout this journey, their mentorship has fostered both my growth and confidence as a researcher. Both John and Enrico have offered me countless opportunities to develop myself, and for that, I will always be grateful.

I wish to express my gratitude to Fabrizio Scozzese and Roberto Gentile, who assisted with Chapters 4 and 5 of this thesis, respectively. The quality of my work would be considerably lower without their insight and guidance. I am also thankful to all the anonymous reviewers who provided constructive feedback on the journal articles that form the main body of this thesis.

I want to thank my family and friends who have always believed in me and been proud of my achievements. In particular, I wish to thank my parents, Helen and Steven, who have always given everything to support me and to foster my curiosity and ambition. I would not be the person I am today without them. I also want to acknowledge my Nan, Jenny, who passed away last year. Nan never stopped learning and had a deep love of science that she helped foster in me. I know she would be very proud of this accomplishment.

I met Hamish Dow during my first week in Glasgow in 2017, and since then we have shared so many ups and downs - though he did manage to finish his PhD well before me! Our lives may now finally take us in different directions, but I am incredibly grateful to have made a lifelong friend.

It would be remiss of me not to mention my 5-a-side football group, ‘Strathletico FC’. Our Tuesday matches have always been a great way to get to know my peers and perhaps vent some stress from the PhD life. Long may it continue! I am also thankful for my friends in Southend who have always been there to offer ‘great schtick’ when it was needed the most.

Last but not least, I would like to thank my partner Rosie - mostly for putting up with me! Rosie is the most loving, supportive partner anyone could wish for. She has been by my side through every moment of stress and celebration, patiently listening to countless hours of my ramblings as I tried to make sense of my thoughts. I could not have done this without her.

Abstract

Probabilistic seismic risk assessments are vital tools in understanding and estimating the losses, to people and infrastructure, that earthquakes can cause. There are two main approaches to earthquake risk assessment: unconditional, where risk is calculated directly from observations of a system, and conditional, where risk is estimated through a series of probabilistic steps. Although unconditional methods are considered more robust, they are computationally expensive and so rarely used. In contrast, there have been considerable developments in conditional risk modelling techniques in recent years, with the growing popularity of machine learning spurring innovation.

This thesis evaluates potential improvements to both hazard and demand modelling techniques (two stages of the conditional approach) and explores how novel methods and machine learning opportunities can enhance the accuracy of risk estimates. Firstly, the ability of different ground-motion models to improve the accuracy of seismic hazard and risk estimates is investigated. Secondly, the combined selection of seismic intensity measure and demand model is evaluated to find the optimal combination for estimating component-level damage assessment. Finally, modelling the covariance of seismic demand estimates is explored, with impacts assessed through loss estimates on a case study structure. The analysis is performed using hypothetical scenarios and stochastic ground motion models, allowing all results to be validated against risk estimates from the unconditional approach.

The findings of this thesis help to demonstrate the potential for machine learning to improve earthquake risk assessment practice, whilst also highlighting that traditional practices can often still out-perform the state-of-the-art. Improving the way in which earthquake risks are modelled helps to better understand, prepare for, and protect against, these risks, ultimately resulting in safer communities and structures.

Table of Contents

List of figures

List of tables

List of abbreviations and acronyms

ANN	Artificial neural network
CCDF	Complimentary cumulative distribution function
CDF	Cumulative distribution function
GMM	Ground motion model
GMPE	Ground motion prediction equation
GPR	Gaussian process regression
IDA	Incremental dynamic analysis
KS test	Kolmogorov-Smirnov test
LB	Lower bound
MAF	Mean annual frequency
MSA	Multiple stripe analysis
PBEE	Performance-based earthquake engineering
PDF	Probability distribution function
PEER Center	Pacific Earthquake Engineering Research Center
PSDM	Probabilistic seismic demand model
PSHA	Probabilistic seismic hazard assessment
RF	Random forest
SDOF system	Single degree of freedom system
SSM	Seismic source model
SVM	Support vector machine
UB	Upper bound

List of symbols

For clarity to the reader, symbols are defined at the start of each chapter, this section summarises some of the main symbols used across this thesis.

a	Rate of earthquakes within the region of interest
b	Relative ratio of small- to large-magnitude earthquakes within the region of interest
DM	Damage measure
$Dmax$	Maximum absolute difference between two cumulative distribution functions
DS	Damage state
DV	Decision variable
EDP	Engineering demand parameter
EMD	Earth mover's distance
$f(X)$	Probability density function of random variable X
$G(X\|Y)$	Complimentary cumulative distribution function of random variable X conditional on random variable Y
$I(X)$	Indicator function equal to 1 if conditions of X are met, and 0 otherwise
IDR	Inter storey drift
IM	Intensity measure
ISW	Importance sampling weight
M	Earthquake magnitude
N_{source}	Number of rupture sources

$P(X)$	Cumulative distribution function of random variable X
$P(X\|Y)$	Cumulative distribution function of random variable X conditional on random variable Y
PFA	Peak floor acceleration
PGA	Peak ground acceleration
R	Source-to-site distance
$S_a(T,\zeta)$	Spectral acceleration at given period, T, and damping, ζ
T	Structural period
$VS30$	Average shear-wave velocity in the top 30m of the ground
B	Lognormal dispersion of a probabilistic seismic demand model
ζ	Damping ratio
Θ	Lognormal mean of a probabilistic seismic demand model
$\lambda(X)$	Frequency of exceedance of random variable X
$\rho(X,Y)$	Correlation of random variables X and Y
σ	Model error parameter, usually evaluated as the standard deviation of residuals
$\Sigma(X,Y)$	Covariance of random variables X and Y
Φ	Cumulative distribution function of the standard normal distribution

1 Introduction

Seismic risk assessments are important tools for earthquake engineering. They express probable losses due to shaking through different metrics (e.g., dollars, downtime, and deaths) (e.g., Musson, 2000) and have uses in a range of sectors (e.g., design, insurance, and policy making) (e.g., Erdik, 2017). As seismicity is a complex and hard-to-predict phenomenon, there is an inherent uncertainty and variability within these risk assessment practices (Moehle & Deierlein, 2004). This makes it favourable to perform seismic risk assessment within a probabilistic framework, allowing for better utilised and more efficient designs and decision making.

Probabilistic seismic risk can be assessed via two distinct approaches (e.g., Scozzese et al., 2020). The first, known as the unconditional approach, directly uses observations of a system's response to ground motions to simply count the rate of exceedance of some loss threshold (e.g., Bradley et al., 2015). Unconditional methods are considered the more robust approach of the two and are noted for their adaptability and conceptual simplicity. However, this approach can be computationally expensive, requiring analysis of a large number of earthquake records, and so is rarely used outside of research.

The second, more popular, approach (known as the conditional approach) was established to solve this problem. Perhaps the most popular conditional method is the

Pacific Earthquake Engineering Research (PEER) Center's framework for performance-based earthquake engineering (PBEE) (e.g., Moehle & Deierlein, 2004), which comprises four intermediate stages, and corresponding parameters, combined into the triple integral of Equation 1-1:

$$\lambda(DV) = \iiint G(DV|DM)|dG(DM|EDP)||dG(EDP|IM)||d\lambda(IM)| \quad (1\text{-}1)$$

where:

- $\lambda(X)$, is the mean annual frequency (MAF) that X exceeds some threshold.
- $G(X|Y)$ is the complementary cumulative distribution function (CCDF) of X conditional on Y.
- $d\lambda(IM)$, known as the hazard assessment, expresses the ground shaking at the site of interest and is represented by an intensity measure (*IM*).
- $G(EDP|IM)$, known as the demand assessment, describes the structural response to ground motions and is represented by an engineering demand parameter (*EDP*).
- $G(DM|EDP)$, known as the damage assessment, gives an indication of the system's damage due to the structural response, and is represented by a damage measure (*DM*).
- Finally, $G(DV|DM)$, known as the loss assessment, estimates the losses in the system due to the estimated damage and is represented by a decision variable (*DV*).

The discrete nature of this approach allows for separate analyses to be performed at each stage. There is an ever-growing array of interchangeable modelling techniques available for use at each stage. In particular, there has been considerable recent innovation in both hazard and demand modelling, with advancements in machine learning techniques fuelling research into these disciplines (e.g., Kong et al., 2018;

Kubo et al., 2024). The introduction of new techniques and the growing number of available hazard and demand models warrant an exploration into their ability to estimate risk.

Moreover, the approach is heavily reliant on the initial selection of the *IM*. This choice should have little impact on risk estimates, provided that it is *sufficient* (i.e., statistically independent of earthquake characteristics, such as magnitude and distance) (Luco & Cornell, 2007), and adequate ground-motion records are used to characterise the system's demand. However, these conditions cannot always be met, and this has led to a proliferation of candidate *IM*s in the literature (e.g., Katsanos et al., 2010). An optimal method for selecting *IM*s in the direct context of PBEE, given the growing variety of hazard and demand models, is lacking.

The conditional approach improves upon the unconditional approach by allowing for a computationally efficient description of seismic risk in a region, based on far fewer observations. However, there is a clear lack in quality and quantity of data used to evaluate earthquake risk in many regions of the world (e.g., Kramer, 1996; Xie et al., 2020), and the improper use of this data could introduce bias and uncertainty into risk models.

Stochastic ground motion models could potentially solve this issue, helping to bolster empirical datasets to produce more confident risk estimates. Alternatively, the output of these simulation models could be applied to the unconditional approach, providing the large number of ground motions needed to provide accurate risk estimates. This creates an opportunity to compare the two approaches directly, allowing validation of conditional techniques through the robust unconditional approach.

1.1 Thesis aim and objectives

The overall aim of this thesis is to investigate potential improvements to conditional earthquake risk assessment processes and explore how novel techniques and machine learning opportunities can enhance the accuracy of risk estimates. In particular, the research contained within this thesis focuses on how hazard and demand modelling techniques can be applied to ensure that risk estimates are as accurate as possible.

To demonstrate the impact of the risk modelling choices explored, it is necessary to assess their impact across the overall conditional risk assessment framework. As such, risk estimates are evaluated in the form of demand hazard, component-level damage, and loss estimates in subsequent chapters of this work. This allows the entire PBEE risk assessment framework to be explored across this thesis.

The project's aim is met through six objectives:

1. Review current conditional risk assessment practice, in particular hazard and demand modelling techniques, and identify methods with potential to improve upon standard practice, e.g., machine learning methods.
2. Evaluate various methods of performing ground motion modelling and hazard assessment to recommend optimal methods for application to risk assessment.
3. Compare and analyse different demand modelling techniques to identify ways in which current practice can be further enhanced with respect to the accuracy of risk estimates.
4. Use stochastic ground motion modelling to benchmark conditional risk assessment approaches against an unconditional approach.

5. To investigate how techniques explored within this thesis can have downstream impacts on risk assessment outputs (e.g., *DM*s and *DV*s).
6. Develop appropriate methods to evaluate the performance of risk modelling techniques in order to aid risk model validation and assist with decision making.

1.2 Outline of thesis

This thesis is organised into six chapters, comprising a background review (Chapter 2), three distinct research articles formatted for peer-reviewed journals (Chapters 3-5), and finally the overall project conclusions (Chapter 6). Figure 1-1 shows a schematic of the PBEE risk assessment framework, and highlights which stages are considered within each chapter of the main research body.

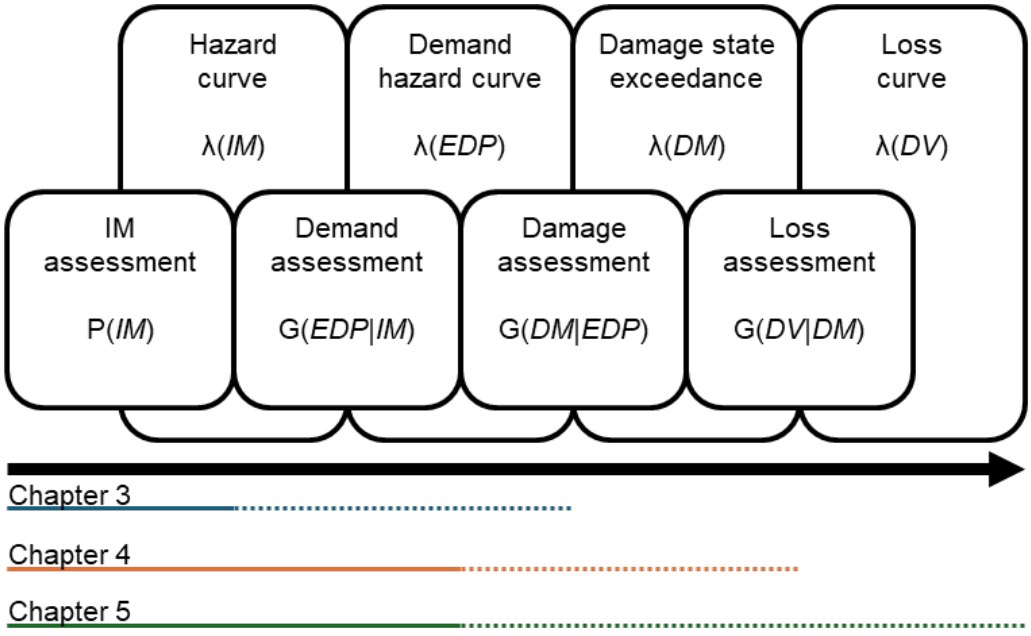


Figure 1-1: Schematic of earthquake risk assessment framework with the topics covered by each chapter of the main research body overlaid. Solid lines indicate stages of the process that are investigated, whilst dashed lines demonstrate areas where risk estimates are evaluated.

A brief outline of the contents of each chapter is provided below:

- **Chapter 2** provides a background review on the relevant work to this thesis, specifically focusing on ground-motion modelling and *IM* selection, probabilistic seismic demand modelling techniques, and machine learning opportunities within these stages of earthquake risk assessment. This chapter identifies and highlights the research opportunities that form the foundation of this thesis, meeting Objective 1 of this thesis.
- **Chapter 3** investigates how the use of different ground-motion modelling techniques can affect seismic hazard and risk estimates. Relaxing the homoscedastic assumption for the variance of the *IM* is explored, while an observational-based hazard assessment that uses recorded strong motions to directly quantify hazard is also proposed. Findings highlight the ability of machine learning models to estimate ground motion intensity. This work contributes to Objectives 2 and 4 to 6 of this thesis.
- **Chapter 4** evaluates how the combined selection of *IM* and demand models can impact seismic risk estimates, features that are often considered separately. This chapter also introduces a novel risk model performance evaluation metric that could be applied at all stages of a conditional risk assessment framework. Findings highlight a need to consider *IM* and demand model selection in tandem to produce accurate risk estimates, whilst also suggesting that traditional *IM* performance metrics may not be sufficient for assessing the accuracy of risk estimates. This work helps to meet Objectives 2 to 6 of this thesis.

- **Chapter 5** proposes a method of incorporating covariance amongst *EDP*s into seismic risk estimates. This study examines various types of Cloud analysis-based demand models and different formulations of the covariance structure through residual modelling of both the variance and correlation of *EDP*s. Loss estimates are produced to fully evaluate the impact that considering covariance has on the risk assessment procedure. The Earth Mover's Distance is applied as a new metric for assessing seismic risk accuracy. Results indicate that heteroscedastic variance modelling improves the accuracy of risk estimates, as does accounting for correlation; however, the method of considering correlation is not necessarily crucial. This study helped to meet Objectives 3 to 6 of this thesis.
- **Chapter 6** contains the overall thesis conclusions and highlights potential scope for future work.

Although each of the three research chapters evaluates a different stage of the conditional risk assessment framework, they share a broadly consistent experimental setup: featuring a stochastic ground motion model, a SSM, and a structural model. However, these components are progressively refined throughout the thesis to better address the specific objectives of each chapter.

Chapter 3 adopts the Atkinson-Silva stochastic model (Atkinson & Silva, 2000), which is based on Californian ground motions. This enables more consistent comparison with the observational hazard model derived from the NGA-West2 database (Ancheta et al., 2014). In contrast, Chapters 4 and 5 utilise the SIGMA stochastic model (Fiorentino et al., 2025; Sabetta & Pugliese, 1996; Sabetta et al., 2021), reflecting a more state-of-the-art approach, with ground motions calibrated and validated against recent Italian earthquakes.

The SSM used in Chapters 4 and 5 is largely consistent with that in Chapter 3, with some alterations. In Chapter 3, multiple stations are included to capture ground motion path effects within the hypothetical region, helping to enhance the realism of the created hazard assessments. However, as Monte Carlo hazard assessment is employed in Chapters 4 and 5, this detail is no longer required; and so, the SSM is simplified to a single central station. Additionally, in Chapter 5, fault distances are reduced and maximum magnitudes increased to generate stronger ground motions, allowing the better investigate of the case study structure's nonlinear response range.

Finally, the structural modelling approach also evolves. Chapter 3 employs a single degree of freedom (SDOF) system, as the primary focus is on hazard characterisation, with risk assessment included as a simple extension. In Chapters 4 and 5, a more advanced structural model is adopted to develop more realistic demand models and improve the applicability of the results.

1.3 List of publications

The following is a comprehensive list of peer-reviewed journal articles and conference papers related to this thesis.

- Relating to Chapter 3 of this thesis:
  - Rudman, A., Douglas, J., & Tubaldi, E. (2023). *Using ground-motion simulations within a Monte Carlo approach to assess probabilistic seismic risk,* SECED 2023 Conference, Cambridge, UK.
  - Rudman, A., Douglas, J., & Tubaldi, E. (2024). The assessment of probabilistic seismic risk using ground-motion simulations via a Monte Carlo approach.

Natural Hazards, *120*(7), 6833-6852. https://doi.org/10.1007/s11069-024-06497-1

- Relating to Chapter 4 of this thesis:
  - Rudman, A., Douglas, J., Tubaldi, E., Scozzese, F., & Turchetti, F. (2024). *Evaluating the impact of intensity measure selection on conditional seismic risk,* 18th World Conference on Earthquake Engineering, Milan, Italy.
  - Rudman, A., Tubaldi, E., Douglas, J., & Scozzese, F. (2024). The impact of the choice of intensity measure and seismic demand model on seismic risk estimates with respect to an unconditional benchmark. *Earthquake Engineering & Structural Dynamics*, *53*(14), 4183-4202. https://doi.org/https://doi.org/10.1002/eqe.4208
- Relating to Chapter 5 of this thesis:
  - Rudman, A., Tubaldi, E., Gentile, R., Douglas, J. (2026). Covariance Structure Modelling of Engineering Demand Parameters in Cloud-Based Seismic Analysis. *Earthquake Engineering & Structural Dynamics*. 2026; 55: 1533-1551. https://doi.org/10.1002/eqe.70151

2 Background review

This Chapter provides a background review of the components of the conditional risk assessment framework explored within this thesis. First, the concept of probabilistic seismic hazard assessment (PSHA) is reviewed (Section 2.1), including the techniques available to model ground motions and the metrics used to select the most appropriate seismic intensity measure (*IM*). Next, in Section 2.2, available techniques for assessing the seismic demand on structures are evaluated. After this, the role of ground motion simulations in seismic risk assessment is examined (Section 2.3). Finally, the use of machine learning techniques in both hazard and demand modelling are summarised (Section 2.4).

2.1 Hazard assessment

PSHA is the foundation of a probabilistic seismic risk assessment, establishing the frequency of earthquake occurrence at the site under investigation over a specified period of time, usually one year (e.g., McGuire, 2008). This method was developed in the 1960s by Cornell (1968) and Esteva (1967), and further enhanced by Cornell (1971) who translated PSHA into a total probability theorem-based framework, as per Equation 2-1:

$$\lambda(IM > im) = \sum_{i=1}^{n_{source}} \lambda_i(M > m_{min}) \int_{m_{min}}^{m_{max}} \int_{0}^{r_{max}} P(IM > im|m,r) f_M(m) f_R(r)\, dr dm \quad (2\text{-}1)$$

where, $\lambda(IM > im)$ is the recurrence rate of earthquakes with intensity, *IM*, greater than some threshold, im, $\lambda_i(M > m_{min})$ is the recurrence rate of earthquakes with magnitude, *M*, greater than a threshold, m, $f_M(m)$ is the probability distribution of earthquake magnitudes at each source, and finally $f_R(r)$ is the distribution of earthquake source-to-site distances for each potential source. This integral is performed and summed over the number of sources present in the region, n_{source}. The workflow of this equation is represented by Figure 2-1, and consists of five steps:

1. Capture all earthquake sources within the region of interest within a seismic source model (SSM).
2. For each source, characterise the distribution of potential earthquake magnitudes.
3. For each source, characterise the distribution of source-to-site distances.
4. For each source, estimate the probabilistic distribution of an *IM* at the site as a function of earthquake magnitude, distance, and other relevant parameters, using a ground motion model (GMM).
5. Combine information from all steps and sources to estimate site hazard.

Essentially, the PSHA equation could be thought of as the combination of two models. The SSM is the first of these models and describes the characteristics of shaking within a region. The second model describes how ground motions are generated and propagate from the source and interact with the site of interest and is known as a GMM.

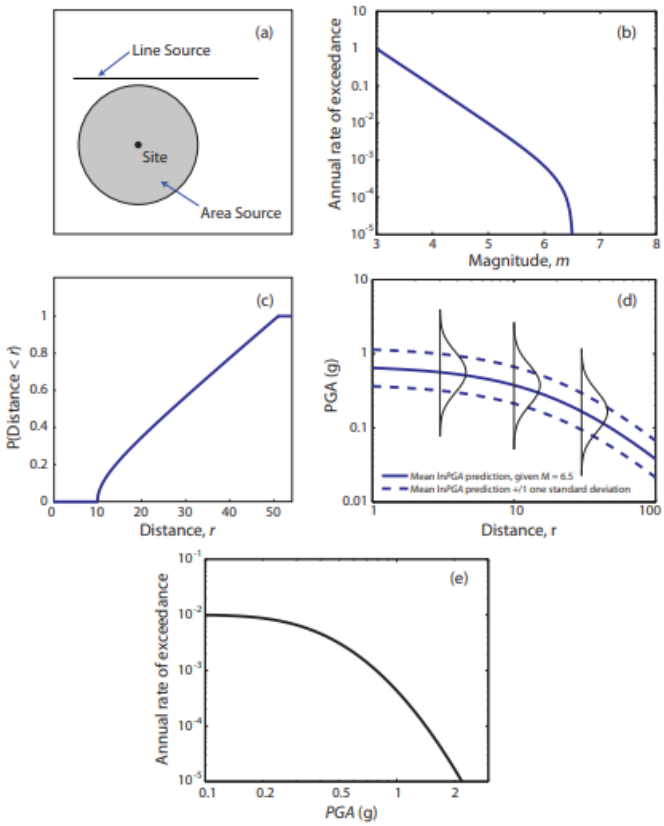


Figure 2-1: probabilistic seismic hazard assessment workflow, taken from Baker et al. (2021) (a) example seismic source model, (b) earthquake magnitude probability distribution, (c) earthquake source-to-site distance probability distribution, (d) example ground motion model for a single magnitude, (e) seismic hazard curve.

An SSM is developed by defining the type of earthquake fault, the location of the potential rupture, and finally its potential size and recurrence (e.g., Baker et al., 2021). Earthquake faults are most commonly represented as points, one-dimensional lines, or as two-dimensional areas, although more complex models may represent three-dimensional zones of faulting (e.g., Kramer, 1996). The potential distribution of earthquake magnitudes is often modelled using a Gutenberg-Richter recurrence relationship (Gutenberg & Richter, 1944), which provides an estimate of how many earthquakes of a certain size occur in a given time period, usually a year. This is represented by Equation 2-2:

$$log_{10}\lambda = a - bm \tag{2-2}$$

where, λ is the number of earthquakes occurring within a given period greater than magnitude, m, and a and b are constants estimated by statistical analysis of past earthquakes within the region. a represents the rate of earthquakes within the region and b represents the relative ratio of small- to large-magnitude earthquakes, usually found (or assumed) to be one. Often, the bounded Gutenberg-Richter relationship is used to describe the probability distribution of earthquake magnitudes at a source (e.g., Baker et al., 2021), where a maximum and minimum magnitude of earthquakes at the source are specified. This probability distribution is calculated as per Equation 2-3:

$$f_M(m) = \frac{b \cdot ln(10) \cdot 10^{-b(m-m_{min})}}{1 - 10^{-b(m_{max}-m_{min})}}, m_{min} < m < m_{max} \tag{2-3}$$

where, m_{min} is the minimum possible magnitude at the source, and m_{max} is the largest possible magnitude at the source. The recurrence of earthquakes at a particular source is often assumed to follow a Poisson distribution, with mean rate determined from the Gutenberg-Richter relationship (e.g., Kramer, 1996).

Earthquake occurrence at a source is often assumed to follow a uniform distribution, i.e., for a line source, an earthquake is equally likely to occur anywhere along that line (e.g., Reiter, 1990). This means that the calculation of earthquake source-to-site distances is relatively straightforward and based purely on the geometry of the SSM. One question, however, within this stage of the PSHA process is deciding which distance is most accurate, with several alternatives offered in the literature, such as epicentral distance, hypocentral distance, or Joyner-Boore distance (e.g., Scherbaum et al., 2004).

The processes used to define seismic source characteristics often rely heavily on expert judgement, and SSMs may be based on conflicting assumptions. For example, modelling earthquake occurrence with a Poisson distribution assumes independence among earthquake events, when realistically (as suggested by elastic rebound theory), events are not independent of one and other (e.g., Kramer, 1996). This can create large uncertainties which need to be properly accounted for within seismic hazard formulations.

The second model used for PSHA is the GMM. GMMs are empirical relationships that describe the relationships between earthquake source, path, and site parameters (e.g. earthquake magnitude and source-to-site distance) to an *IM* describing the level of ground shaking at a site of interest (e.g., Baker et al., 2021). The predictive capability of GMMs is limited by sparse and heterogeneous earthquake datasets in many regions (e.g., Kramer, 1996; Xie et al., 2020). They are discussed further in Section 2.1.1.

By performing PSHA at a range of *IM* thresholds, users can construct a seismic hazard curve. Hazard curves summarise the likely occurrence rates of different *IM*s over a given period, usually one year, thereby representing the entire hazard at the site of interest (e.g., McGuire, 2004). An example hazard curve is presented in Figure 2-1. This output can then be fed into the rest of the probabilistic seismic risk assessment framework to estimate the risk of losses to a structure on the site of interest. Hazard curves are also useful for identifying key hazard levels/scenarios for use in seismic design codes (e.g., Eurocode 8, ASCE 7, NBCC), which prescribe *IM* levels associated with predefined probabilities of exceedance over the design life of a structure. Commonly adopted levels include events with 10% and 2% probabilities of

exceedance in 50 years, corresponding to mean return periods of approximately 475 and 2475 years, respectively.

An alternative approach to hazard assessment is through a Monte Carlo-based procedure, a process that estimates seismic hazard at a site numerically rather than analytically (e.g., Musson, 2000). The first step of Monte Carlo hazard assessment is to draw many samples of earthquake magnitude and distance from their respective distributions, as characterised by the SSM. Sample size is often determined by a specified time horizon and the known earthquake recurrence rate in the region. These magnitude–distance samples are then propagated through a GMM to estimate ground motion intensities at the site of interest.

Finally, a simple counting procedure allows users to assess the hazard curve at the site of interest, as per Equation 2-4:

$$\lambda(IM > im) = \sum_{i=1}^{n_{source}} \lambda_{source_i} \sum_{j=1}^{N_j} \frac{I_{i,j}(IM > im)}{N_j} \quad (2\text{-}4)$$

where, N_j is the number of magnitude-distance samples considered for a given i-th source, and $I_{i,j}(IM > im)$ is an indicator function equal to one if for the i^{th} record any *IM* exceeds the *IM* threshold, im, and zero otherwise. Essentially, the frequency of *IM* exceeding some threshold is simply a count of how many records are greater than or equal to that threshold, divided by the number of records (e.g., Musson, 2000).

2.1.1 Ground motion modelling

GMMs, also known as ground-motion prediction equations or attenuation relationships, were first introduced by Esteva and Rosenblueth (1964). They estimate the chosen *IM* using site, source, and path parameters as predictor variables. They

model both the median ground motion at a site and its variability (e.g., Abrahamson & Silva, 2008). All GMMs estimate *IM* using earthquake magnitude and source-to-site distance (but many include other parameters too), with the simplest models following the functional form represented by Equation 2-5, which is also visually displayed by Figure 2-2:

$$\ln(IM) = C_0 + C_1 M + C_2 \ln(R) + \sigma \quad (2\text{-}5)$$

where, M is earthquake magnitude, R is the source-to-site distance, C are coefficients fit through regression analysis, and σ is the model error term. Other parameters are often included in GMMs. For instance, the well-known Chiou and Youngs (2014) GMM included 15 predictor variables, accounting for earthquake magnitude, several source-to-site distance metrics, and measures of faulting style and site conditions, e.g., the average shear wave velocity of the first 30m of soil (*VS30*). This helps demonstrate how GMMs are becoming more complex as more data become available and modelling techniques enable more detailed analyses.

The example GMMs plotted in Figure 2-2 helps to illustrate one of the biggest flaws of this modelling technique. There is a general lack of ground motion data available in the area of greatest interest to risk modellers (i.e., high-magnitude, short-distance events). This is partly down to the long time required to collect sufficient data for a comprehensive range of earthquake magnitudes and distances, especially in areas of low seismic activity (e.g., Douglas & Edwards, 2016). To avoid this issue, GMMs are often developed using data from similar geological regions worldwide, which is called the ergodic assumption (e.g., Abrahamson & Silva, 2008; Lavrentiadis et al., 2023). For example, the NGA West 2 GMMs (Ancheta et al., 2014) use data from California,

Turkey, and Taiwan, among other countries. The ergodic assumption currently helps improve ground motion prediction practices in areas with insufficient local data and, in turn, risk assessment. However, there is a growing trend towards site-specific ground motion prediction, with advanced modelling techniques and better data making this more possible (e.g., Bradley, 2025; Lavrentiadis et al., 2023).

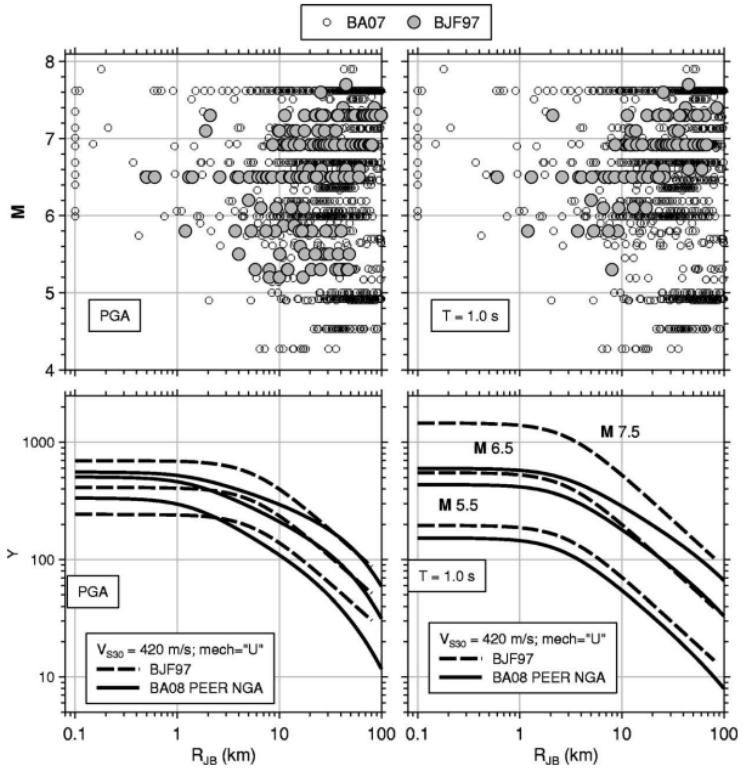


Figure 2-2: Taken from Boore and Atkinson (2008). The top two panels show the observed magnitude-distance-intensity (in the form of peak ground acceleration, and spectral acceleration with a period of 1.0s) relationships for records used to construct the Boore et al. (1997) (BFJ97) and Boore and Atkinson (2008) (BA08) ground motion models. The bottom two panels show the median estimates from both the BFJ97 and BA08 ground motion models for a variety of magnitudes and distances.

Despite there being limited data available for the creation of GMMs, a large number of GMMs are available in the literature, and this is growing at a considerable rate each year (Douglas, 2024). Newer GMMs often focus on capturing model uncertainty and on delineating between epistemic uncertainties and aleatory variabilities, reducing

epistemic uncertainties as far as possible (e.g., Aldama-Bustos et al., 2023; Liou & Abrahamson, 2024). Nonparametric techniques have also become more popular (e.g., Douglas & Edwards, 2016). Both machine learning (Section 2.4) and ground motion simulation (Section 2.3.1) approaches have been used to increase predictive power and reduce uncertainties, within ground motion prediction and ultimately hazard assessment (e.g., Boore, 2003; Khosravikia et al., 2018). The growing number of GMMs using these more advanced techniques warrants a comparison of the methods available to predict ground motion, which is presented in Chapter 3.

2.1.2 Intensity measures

*IM*s are the response variable of a GMM, representing the ground shaking intensity at the site of interest, and are used to evaluate the site hazard, as well as to define the structural response by means of a probabilistic seismic demand model (PSDM, see Section 2.2.1). In general, *IM*s represent one or more of three ground motion characteristics: amplitude, frequency, and duration (Baker et al., 2021).

Historically, the most common *IM*s used were peak ground motion amplitudes (e.g., the largest absolute acceleration); now, spectral ordinates are more common (e.g., the 5% damped spectral acceleration at the structure's fundamental period). Alternatively, it has been suggested that a vector of *IM*s would better describe ground motions (e.g., Baker & Cornell, 2005; Gehl et al., 2013).

Conditional seismic risk assessments are highly reliant on the selection of the *IM*, making their selection important, as each has distinct trade-offs and uses in different contexts. As such, a range of *IM*s has been proposed in the literature, with some of the most popular candidate *IM*s identified and described in Table 2-1.

Table 2-1: A collection of commonly used and investigated seismic intensity measures found in the literature

Intensity Measure	Amplitude (A), Frequency (F), Duration (D)?	Definition	Units
A95	A,D	Acceleration amplitude corresponding to 95% of cumulative Arias Intensity (e.g., Sarma & Yang, 1987).	m/s^2
Absolute/Relative Bracketed Duration	D	Time between first and last exceedance of a specified threshold acceleration (usually 50 cms^{-2}) (e.g., Bommer et al., 2009).	s
Absolute/Relative Significant Duration	D	Time interval between two percentages (often 5%–95%) of Arias Intensity (e.g., Bommer et al., 2009).	s
Absolute/Relative Uniform Duration	D	Total time period within a ground motion where the amplitude has surpassed a given threshold (usually 50 cms^{-2}) (e.g., Bommer et al., 2009).	s
Acceleration/Velocity Spectral Intensity	A,F	Integral of acceleration/velocity response spectrum over a defined period range (e.g., Sarma & Yang, 1987).	$m/s^2 \cdot s$, $m/s \cdot s$, $m \cdot s$
Arias Intensity (IA)	A,D	Integral of squared acceleration over time, representing total energy per unit mass (e.g., Arias, 1969).	m/s
Bandwidth	F	Measure of frequency content spread (e.g., based on spectral moments) (e.g., Kramer, 1996).	-
Central Frequency (fc)	F	Frequency at which spectral energy is centred (ratio of second to zero spectral moment) (e.g., Kramer, 1996).	Hz
Characteristic Intensity (IC)	A	Function of Arias Intensity and duration used to estimate structural damage potential (e.g., Park et al., 1985).	m/s
Cumulative Absolute Velocity (CAV)	A,D	Integral of absolute velocities (e.g., Reed et al., 1988).	m/s

Cyclic Damage Parameters	A,D	Quantify cumulative inelastic energy demand based on cyclic loading (e.g., Park–Ang index) (e.g., Gehl et al., 2013).	-
Housner Intensity (IH)	A,F	Integral of the the pseudo-velocity spectrum usually between a period range of 0.1–2.5s (e.g., Housner, 1952).	m·s
Number of Cycles	D	Effective number of significant loading cycles contributing to damage (e.g., Hancock & Bommer, 2005).	-
PGA, PGV, PGD	A	Peak ground acceleration/velocity/displacement of a ground motion (e.g., Baker et al., 2021).	m/s^2, m/s, m
Predominant Period (Tp)	F	Period corresponding to maximum Fourier amplitude of the motion (e.g., Kramer, 1996).	s
Root Mean Square Acceleration/Velocity/Displacement	A	Root-mean-square acceleration/velocity/displacement (e.g., Kramer, 1996).	m/s^2, m/s, m
Sa, Sv, Sd	A,F	Spectral acceleration/velocity/displacement for a given period and damping ratio (e.g., Baker et al., 2021).	m/s^2, m/s, m
Sa_avg	A,F	Average spectral acceleration over a specified period range (e.g., Vamvatsikos & Cornell, 2005).	m/s^2
Slope (of Husid plot)	A,D	Average slope of the Husid plot (Arias Intensity vs. time) over significant duration, indicating energy rate (e.g., Gehl et al., 2013).	m/s^3
Total Duration	D	Total duration of the recorded motion (e.g., Bommer et al., 2009).	s
vmax/amax	A,F	Ratio of peak velocity to peak acceleration (e.g., Kramer, 1996).	s

2.1.3 Intensity measure selection

Due to the large number of *IM*s in existence, the topic of selecting an optimal *IM* for risk assessment has become a well-covered and wide-ranging debate - see e.g., Katsanos et al. (2010) for a review of *IM* development and selection in the context of ground motion record selection. It has proved challenging to find a comprehensive *IM* that is optimal for all types of earthquake risk assessment. This, in turn, has led to many research efforts devoted to evaluating *IM* selection across a range of contexts. For instance, Mackie and Stojadinovic (2003) evaluated *IM* selection for probabilistic seismic demand models of highway bridges in California. Bray and Travasarou (2007) investigated the impacts of *IM* selection for estimating seismic slope displacements. Kohrangi et al. (2016) compared the impact of selecting eight different structure-specific *IM*s on repair cost estimates of three different 3D building models.

In general, an optimal *IM* is selected from one, or a combination of, four key metrics: *efficiency*, *sufficiency*, *practicality*, and *hazard computability*. These are represented by Figure 2-3.

The first, *efficiency*, describes the variability in the system demand, which is characterised by an engineering demand parameter (*EDP*) conditional on the *IM* (Luco & Cornell, 2007). The efficiency of an *IM* can be calculated as the standard deviation of residuals from a linear regression fit between *IM* and *EDP*, with a smaller value indicating a more efficient *IM*.

The *sufficiency* of an *IM* is a measure of its statistical independence from earthquake characteristics, e.g., magnitude and distance (Luco & Cornell, 2007). *Sufficiency* can be calculated by fitting linear regression models to the residuals from the regression of

IM on *EDP*, with either magnitude or distance as predictor; the p-value of the regression line slope would serve as the *sufficiency* measure, with a large p-value indicating that the slope is zero and as such the two variables are independent of each other. Although any *IM* with a *sufficiency* value above the confidence threshold (usually 0.05) can be defined as *sufficient*, it could be considered that the higher the p-value, the lower the evidence of an *insufficient IM*.

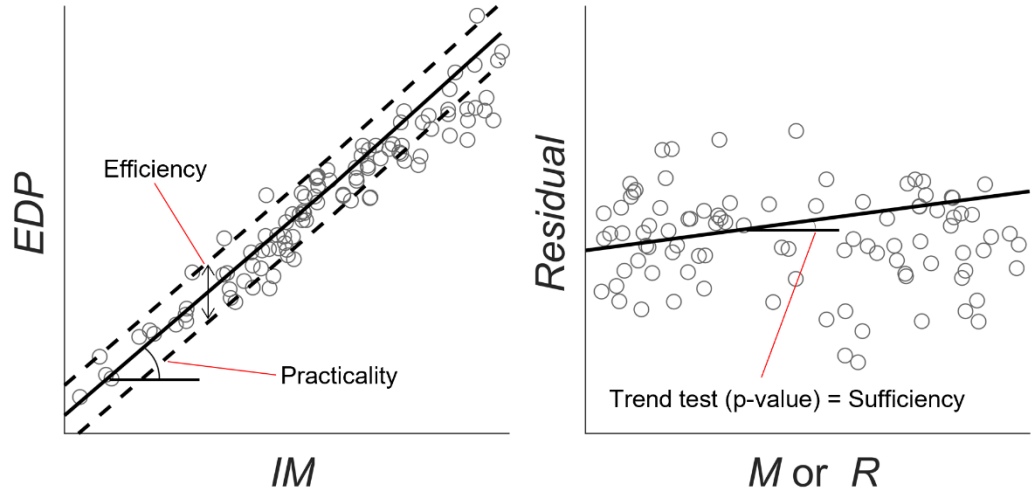


Figure 2-3: Commonly used intensity measure performance metrics efficiency, sufficiency, and practicality. Hazard computability cannot be graphically displayed. Note that sufficiency is represented by the p-value of the slope, determining if a statistically significant relationship is present between magnitude or distance and the model residuals.

The concept of *hazard computability* was introduced by Giovenale et al. (2004) and is defined as the ease with which a hazard curve can be computed for a given *IM*. Essentially, this is determined by whether a hazard curve exists for, or could exist for, the *IM* under investigation. For example, a simple way to evaluate hazard computability is to consult the Douglas (2024) GMM compendium to determine whether a model exists for the specified *IM* (for the region of interest).

Finally, Padgett et al. (2008), among others, introduced *practicality*, which is the correlation between the *IM* and *EDP* within the regression model. This metric is defined by the slope of the linear regression model; a higher slope indicates a stronger correlation and therefore greater *practicality*.

Other metrics have been proposed in the literature, but none as widely used as these four, in particular, *efficiency* and *sufficiency*. Padgett et al. (2008) also introduced the term *proficiency* (*efficiency* divided by *practicality*), and Tothong and Luco (2007) introduced *scaling robustness* (lack of bias in response estimation from scaled records). Du et al. (2020) used a joint-entropy-based approach to determine the optimal *IM* for site-specific seismic risk assessment. Qian and Dong (2022) used multi-criteria decision-making principles to identify the optimal *IM*. Finally, Khosravikia and Clayton (2020) offered updated definitions of efficiency, practicality, and proficiency, suggesting that current *IM* selection metric formulations do not always guarantee the selection of the optimal *IM*.

Given the crucial role the *IM* plays in seismic risk assessment, it is understandable that a wide range of *IM*s have been proposed in the literature and that many methods exist to choose the optimal *IM*. However, there has been little-to-no research developing an *IM* selection methodology based on the unconditional approach to risk estimation (Section 2.3), and research has not been performed to validate existing *IM* selection metrics by comparing them against benchmark, unconditional, risk estimates.

2.2 Demand assessment

The next stage in the PBEE framework is to estimate the frequency of exceedance of demand that the structure undergoes - this term is also often called the system response. The demand assessment does this by combining a model of the system demand conditional on the seismic intensity, known as a PSDM (Section 2.2.1), with the site hazard curve (as per the total probability theorem). This is explained by Equation 2-6 (e.g., Bradley, 2013):

$$\lambda(EDP > edp) = \int_0^\infty P(EDP > edp \mid IM = im_i) \left| \frac{d\lambda(IM > im_i)}{dIM} \right| dIM \quad (2\text{-}6)$$

where, $\lambda(EDP > edp)$ is the frequency of *EDP* exceeding some threshold, edp, $P(EDP > edp \mid IM = im_i)$ is the complementary cumulative distribution function (CCDF) of the structural demand conditional on the *IM* of choice (found via the PSDM), and finally $\left|\frac{\mathrm{d}\lambda(IM > im_i)}{dIM}\right|$ is the absolute derivative of the hazard curve at threshold, im_i, with the absolute value taken to ensure a positive frequency is returned.

As with hazard assessment, the output of this stage can be plotted as a demand curve, which summarises the MAF of exceedance of different *EDP* levels (e.g., Bradley, 2013). The output of the demand assessment could subsequently be used in a damage and loss assessment; however, these stages of the overarching framework are beyond the scope of this review.

The CCDF of the structural demand conditional on the *IM* is often derived analytically from the output of a PSDM, assuming a lognormal distribution (e.g., Baker, 2015; Shome, 1999), as per Equation 2-7:

$$P(EDP > edp \mid IM = im_i) = \Phi\left(\frac{\ln\left(\frac{im_i}{\theta}\right)}{\beta}\right) \quad (2\text{-}7)$$

where Φ is the cumulative distribution function (CDF) of the standard normal distribution, and θ and β are the lognormal mean and standard deviation of the CCDF, estimated through statistical fitting. A variety of fitting approaches for these parameters exists in the literature. The method of moments approach finds θ and β by matching them to the associated mean and standard deviation, as determined from data produced by structural analyses (e.g., Baker, 2015). Alternatively, users could use a maximum likelihood approach, which finds the θ and β values that maximise the likelihood of having generated the data (e.g., Shinozuka, 1983). Finally, the least squares approach estimates θ and β values to minimise the sum of squared errors relative to the data generated by structural analyses (e.g., Shome, 1999).

To accurately fit this CCDF, many *IM-EDP* sample pairs are required, whether derived through observation or estimation. This can be impractical and inefficient to achieve, given the large time and computational power required to perform thousands of structural analyses on even a relatively simple structural model (e.g., Kim & Kim, 2025). Instead, models that surrogate the structural response have been developed, enabling efficient simulation of *EDP* across the entire *IM* domain. The formulation of these PSDMs are discussed further in Section 2.2.1.

As with hazard assessment, a direct demand assessment approach is also available, based on a Monte Carlo procedure. This forgoes the use of intermediate modelling steps and can use the output of structural analyses to estimate structural demand directly (e.g., Bradley, 2013; Scozzese et al., 2020). This is discussed further in Section 2.3.

2.2.1 Probabilistic seismic demand modelling

PSDMs bridge the gap between hazard assessment and demand assessment by deriving a relationship between *IM* and *EDP*. There are three main approaches to developing PSDMs, as illustrated by Figure 2-4. One of these methods is Incremental Dynamic Analysis (IDA) (Vamvatsikos & Cornell, 2002). IDA involves performing structural analyses under a set of ground motions, each scaled to a different *IM* level, until collapse is achieved. These results can then be plotted to illustrate the structure's demand to shaking across its whole response profile, as shown by Figure 2-4a. As the entire response profile can be explored, demand can be estimated by simply summarising the analysis results through a method of moments approach. Alternatively, demand estimates can be derived from an IDA by locating the fractiles of the IDA curves, thus providing insight into the distribution of structural demand (e.g., Baker, 2015).

Another method, Multiple Stripe Analysis (MSA) (e.g., Mackie & Stojadinovic, 2005; Scozzese et al., 2020), follows a similar methodology, except that a set of ground motions is used at discrete *IM* levels (e.g., Jalayer & Cornell, 2009). It is not required to perform MSA up to collapse; moreover, because different ground motions are used at each *IM* level, the response shape may differ from that for IDA. Maximum likelihood estimation or least squares regression is often used to fit demand CCDFs from the estimated output of MSA (e.g., Baker, 2015). Both IDA and MSA often require scaling of input parameters and are heavily reliant on the *efficiency* and *sufficiency* of the *IM* used and the selected ground motions. This increases uncertainty in the risk assessment process, making it important to ensure that demand output is as accurate as possible to obtain the most accurate loss estimates.

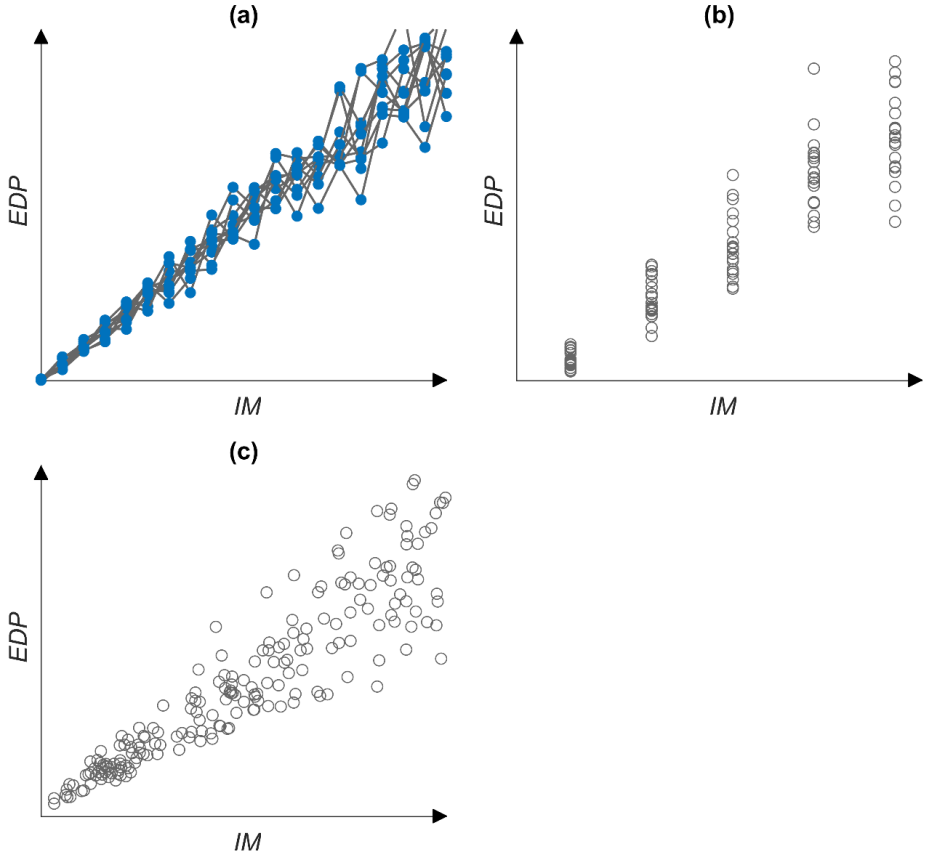


Figure 2-4: Example applications of three major probabilistic seismic demand modelling approaches (a) Incremental Dynamic Analysis (b) Multiple Stripe Analysis (c) Cloud analysis.

Finally, Cloud analysis fits a regression-type model directly between *EDP* and unscaled *IM*s, using all available ground motions regardless of their *IM*, thus creating a "cloud of points" as seen in Figure 2-4c (e.g., Shome, 1999). Typically, Cloud analysis assumes a linear relationship between *IM* and *EDP* (e.g., Baker & Cornell, 2006), where the median prediction from this model represents θ, and the standard deviation of model residuals represents β. The homoscedasticity assumption is often introduced, i.e., β is assumed to be constant across various *IM* levels (e.g., Du & Padgett, 2020). Research has shown that this assumption is not necessarily sound, with several studies investigating the ability of heteroscedastic models to improve upon this

(e.g., Bradley & Lee, 2010; Du & Padgett, 2020; Jayaram et al., 2012). However, little research has compared homoscedastic and heteroscedastic models, and few studies have attempted to validate heteroscedastic approaches. This topic is investigated within Chapter 5 of this thesis.

Recent research has shown that a bilinear fit could improve the accuracy of Cloud analysis PSDMs (e.g., O'Reilly & Monteiro, 2019; Tubaldi et al., 2016). The benefits of the bilinear demand model are evident as the second branch of the model can account for the change in demand once the structure enters the nonlinear demand range, providing a physical meaning behind the mathematical function. Another benefit of bilinear models is that they allow two values of the model dispersion parameter, β, thereby slightly relaxing the homoscedastic assumption. PSDMs can also be developed to account for multiple predictors (e.g., other *IM*s or structural parameters) through multivariate regression models, such as the polynomial response surface model and the multivariate adaptive regression spline technique (Ghosh et al., 2013; Towashiraporn, 2004). Furthermore, there is a growing trend in the use of machine learning models for the development of PSDMs (e.g., Soleimani & Hajializadeh, 2022), as further discussed in Section 2.4. Given the increasing number of techniques available for developing Cloud analysis PSDMs, a comparison of these methods is necessary to identify the best approaches, in terms of both median demand estimation and the capturing of uncertainty. This comparison is provided in Chapters 4 and 5 of this thesis.

2.3 Unconditional approach for seismic risk assessment

As outlined in Chapter 1, one alternative to the PBEE approach is to directly estimate losses through a simulation-based approach, known as the unconditional approach

(e.g., Cornell, 2005; Scozzese et al., 2020). This involves directly estimating structural demand and, ultimately, losses from a large number of ground motion samples (Au & Beck, 2003; Cremen et al., 2022), rather than breaking the procedure into conditional steps. Scozzese et al. (2020) outlines the advantages and disadvantages of this unconditional approach, when compared to conditional approaches; these are summarised in Table 2-2.

Table 2-2: Summary of the arguments between using either the conditional or unconditional approach, taken from Scozzese et al. (2020)

Unconditional approach	Conditional approach
• Research-oriented • No conditioning on *IM* • Large number of simulations (structural analyses) required • Robust and tool for seismic risk estimation • Requires a large number of ground motions to describe the seismic input i.e., from a stochastic ground motion simulation model	• Practice-oriented • Dependent on the initial selection of *IM* • Reduced number of structural analyses required if *IM* Is efficient • Potentially biased if *IM* is not sufficient and ground motion records are not representative of the hazard • Can be applied using recorded ground motions, but in this case, the accuracy cannot be checked

Similar to the Monte Carlo hazard assessment procedure outlined in Section 2.1, this approach first requires the collection of a large number of ground motions at the site consistent with the SSM for the region of interest. In the numerical approach, structural

analysis is then performed on the structure in question using each of these ground motions to obtain a catalogue of structural demands which can be processed into *EDP*s. These can then be counted to estimate the demand on the structure, using Equation 2-8:

$$v(edp) = \sum_{j=1}^{sources} \lambda_{source_j} \sum_{i=1}^{N_j} \frac{I_{i,j}(edp)}{N_j} \tag{2-8}$$

This approach has been used by several researchers in the literature. For instance, Azar and Dabaghi (2021) and Bijelić et al. (2019) examined unconditional hazard and risk assessments, respectively, using the CyberShake software (Graves et al., 2011). Scozzese et al. (2020) used simulated ground motions to compare a conditional MSA-based approach of estimating structural demand with a direct approach based on subset simulation. The results showed that, when the MSA is optimised, the conditional approach can achieve similar results to the directly simulated method but required far less computation time.

Studies that adopt the unconditional approach mainly focus on validating conditional-based models. For instance, Bradley et al. (2015) and Kwong et al. (2015a, 2015b) use this conditional approach as a benchmark to compare the ability of different ground motion selection procedures, with both studies utilising nonlinear single-degree-of-freedom systems. Moreover, Franchin et al. (2012) compared demand estimates of a 15-storey reinforced concrete structure generated from various conditional and unconditional methods to validate the conditional-based methods. However, it is important to note that this study considered only a single *EDP* (maximum interstorey drift) and did not consider frequencies of exceedance below 10^{-3}/yr.

One limitation of the unconditional simulation method is that it requires a large number of realisations to achieve accurate demand estimates. This is due to the very small exceedance probability of high-intensity, damaging events, which are of greatest interest to seismic risk assessment (e.g., McGuire, 2008). To address this inefficiency, a sampling approach could be employed to obtain accurate results with far fewer samples, thus improving computational efficiency. As such, Au and Beck (2003) propose a strategy known as 'subset simulation', a reliability-based approach that helps compute small failure probabilities by combining a series of intermediate conditional failure events. Further, Jayaram and Baker (2010) developed a strategy that used importance sampling to select events of interest, followed by a K-Means Clustering algorithm to further refine the sample size when generating ground-motion intensity maps for lifeline risk assessment.

2.3.1 Ground motion simulations

The large number of samples required for the unconditional approach makes it unlikely that an accurate risk assessment could be conducted with purely empirical ground motions compatible with the site's SSM. To resolve this, ground motion simulation methods can be used. These techniques are often used to bolster existing ground-motion catalogues, helping to cover the whole possible range of earthquakes that could occur in a region, where there may be insufficient data to represent less likely events (e.g., Beauval et al., 2009; Douglas et al., 2024). However, they can also be used to create a comprehensive dataset of simulated ground motions for seismic risk assessment (e.g., Stupazzini et al., 2021).

Different techniques exist to simulate ground motions. Stochastic methods, such as the well-known Boore (2003) model, simulate ground motions as a random process, using

statistically derived parameters to calibrate the realisations to the specific site, path, and source effects of the region. Alternatively, more complex physics-based simulation models have also been developed, like the finite difference method (e.g., Oprsal & Zahradnik, 2002) and the wavelet packet transformation method (Yamamoto & Baker, 2013), which can both create more realistic time-histories, but are more computationally expensive and potentially impractical for detailed risk assessments. For a thorough comparison of the techniques available to simulate ground motions, see Douglas and Aochi (2008).

The method presented by Boore (2003) demonstrates the fundamentals of stochastic ground motion simulation. It is a simple single-source model that generates synthetic ground motions from a small number of inputs. This technique first computes a Fourier spectrum of ground motion using a combination of simple functional forms (i.e., source, path, and site factors), then converts it to the time domain to produce a realistic time series.

As stochastic models are fit to empirical ground motions, they can be calibrated to specific regions in the world. For instance, the Atkinson-Silva stochastic model (Atkinson & Silva, 2000) is derived for earthquakes in California. Whilst the SIGMA model (Fiorentino et al., 2025; Sabetta & Pugliese, 1996; Sabetta et al., 2021) is calibrated to a dataset of recent Italian earthquakes (Lanzano et al., 2019). This means care is required when selecting a stochastic model to ensure the correct ground motions are simulated for the specific risk assessment context.

Although using an entirely simulated dataset of ground motions may not fully capture the complexity of real ground motions, these datasets can help enhance empirical

ground motion datasets by filling gaps where real-world data may be lacking, thus avoiding record scaling, which is not always desirable. This is done by Meirova et al. (2018), Kowsari et al. (2021), and Zolfaghari (2015), who used ground-motion simulations to improve PSHA in Israel, Iceland, and Iran, respectively.

Beauval et al. (2009) compared different hazard analysis techniques, using an empirical Green's function to simulate data for larger events on the island of Guadeloupe (France) from smaller earthquakes. A hybrid PSHA model is built for the island, with hazard curves shown alongside results from the GMM of Ambraseys et al. (2005). Moreover, Stupazzini et al. (2021) used ground-motion simulations to assess hazard. The study also examined the impact of simulations on the conditional risk assessment for a portfolio of buildings in Istanbul. Medel-Vera and Ji (2016) also proposed an approach to use ground-motion simulations in an unconditional risk assessment, although this study was specifically for nuclear facilities in the UK. They found the results comparable to current practice, justifying the method's use for earthquake risk assessments.

2.4 Machine learning opportunities in seismic risk assessment

The use of machine learning in seismic risk modelling is a particularly interesting research area, with growing applications in earthquake engineering over recent years (e.g., Kong et al., 2018). Machine learning is a branch of artificial intelligence that uses algorithms to learn patterns from data and to make inferences from these patterns to predict new data. There are two broad classes of machine learning algorithms, supervised and unsupervised. Supervised learning is the branch that relates most to earthquake risk modelling, using labelled data (i.e., predictor variables that map onto

corresponding response variables) to help solve regression or classification problems (e.g., Jordan & Mitchell, 2015).

Supervised learning techniques should help improve both ground motion prediction and demand estimation, as they are not constrained by the rigid functional forms of traditional GMMs, which may not fully capture the complexity of the underlying physical processes of shaking (e.g., Alidadi & Pezeshk, 2025). Of the many algorithms available to help model GMMs and PSDMs, some of the most popular are artificial neural networks (ANNs), random forest (RF) regression, and Gaussian process regression (GPR).

ANNs are a supervised learning technique inspired by the structure of the human brain, consisting of layers of interconnected nodes that collectively process data. The classic structure of an ANN consists of an input and output layer that map the predictor variables to the response variables, respectively (e.g., Derras et al., 2014). These are connected by a hidden layer that adjusts the input variables based on weightings and transformations optimised by both the machine and the user. Weightings are usually optimised by a backpropagation algorithm, which looks to minimise the error between the input and output (e.g., Pradhan & Lee, 2009).

RF regression is a machine learning method that builds an ensemble of decision trees to produce more accurate and stable predictions than a single tree on its own. Each tree in the forest is trained on a random subset of the data and decision tree hyperparameters before being averaged to produce a final output (e.g., Xie et al., 2020). By averaging the predictions of many diverse trees, the model reduces the risk of 'overfitting' and improves the models' ability to estimate unseen data (e.g., Izquierdo-Horna et al.,

2022). The simplicity of the RF approach makes this type of model easy to optimise as few hyperparameters are needed to be input by the user.

Finally, GPR is a technique that fits a range of possible functions to the data and finds a distribution over all these functions that fit the data. GPRs are characterised by a mean function and a covariance function (or kernel) (e.g., Minas et al., 2018). The kernel captures the similarity between data points, allowing the model to infer smooth patterns and correlations in the input space (e.g., Gentile & Galasso, 2022). This allows GPR models to incorporate both the mean prediction as well as associated variability into a single model (e.g., Gentile & Galasso, 2022).

Each of these models are unconstrained by a proscribed functional form, they can handle multiple inputs and outputs, and they often require minimal user input to be optimised (e.g., Xie et al., 2020). This gives these tools great flexibility and allows them to detect and estimate complicated patterns and uncertainties within data and make predictions based on these (e.g., Alidadi & Pezeshk, 2025). This potentially makes machine learning techniques ideal candidates for ground motion and structural demand modelling, contexts in which processes are driven by complex nonlinear physical processes that traditional, fixed-form regression practices are not necessarily well-adapted.

Supervised learning is already being used within earthquake risk research. Khosravikia and Clayton (2021) compared the performance of three machine learning approaches for ground-motion modelling with that of a traditional linear regression model. The three considered models, an ANN, an RF regression, and a support vector machine (SVM), all performed better than the traditional approach, with the RF model

providing the most accurate estimates. These techniques have also been evaluated for demand modelling, demonstrating great promise (e.g., Li et al., 2022; Mangalathu et al., 2019).

ANN models have received particular attention from ground-motion modellers. One distinct benefit of an ANN is that it can be reconfigured to provide a functional form and thus is not entirely a black-box model, unlike other machine learning techniques (e.g., Khosravikia et al., 2018). Khosravikia et al. (2018) created an ANN to estimate ground motions in the southwestern United States, primarily due to induced seismicity. Derras et al. (2012) discussed the use of ANNs to estimate *IM*s and developed one based on the KiK-net strong-motion database. This approach was then extended to create ANN-based GMMs for a range of candidate *IM*s (Derras et al., 2014; Derras et al., 2016). See recent review articles by Xie et al. (2020), Alidadi and Pezeshk (2025) and Xie (2025) for a thorough review of the use of machine learning in both ground motion prediction and estimation of structural demand.

It is important to note that machine learning models are not perfect predictive tools. They are heavily reliant on the data used to train them, and this can lead to biases and inaccuracies within GMMs (e.g., Kubo et al., 2020). Moreover, they are prone to overfitting, which can make them poor at extrapolation and at predicting events within the range of the training data that have not occurred before (e.g., Monterrubio-Velasco et al., 2024). The black-box nature of many machine learning models creates a lack of transparency in the modelling techniques and can make results harder to interpret or to give physical meaning to the underlying processes (e.g., Khosravikia et al., 2018). Nevertheless, the significant potential of machine learning techniques to enhance the accuracy of risk estimates necessitates evaluating them not only against one another

but also against current risk assessment practice, thereby identifying the most accurate techniques. As such, machine learning techniques are evaluated in Chapters 3-5 of this thesis for both ground motion modelling (Chapter 3) and demand modelling (Chapters 4 and 5).

3 The assessment of probabilistic seismic risk using ground-motion simulations via a Monte Carlo approach

This Chapter has been adapted from:

Rudman, A., Douglas, J. & Tubaldi, E. The assessment of probabilistic seismic risk using ground-motion simulations via a Monte Carlo approach. *Nat Hazards* **120**, 6833–6852 (2024). https://doi.org/10.1007/s11069-024-06497-1

Accurately characterizing ground motions is crucial for estimating probabilistic seismic hazard and risk. The growing number of ground-motion models, and increased use of simulations in hazard and risk assessments, warrants a comparison between the different techniques available to predict ground motions. This chapter aims at investigating how the use of different ground-motion models can affect seismic hazard and risk estimates. For this purpose, a case study is considered with a circular seismic source zone and two line sources. A stochastic ground-motion model is used within a Monte Carlo analysis to create a benchmark hazard output. This approach allows the generation of many records, helping to capture details of the ground-motion median and variability, which a ground motion prediction equation may fail to properly model. A variety of ground-motion models are fitted to the simulated ground motion data, with fixed and magnitude-dependant standard deviations (sigmas) considered. These include classic ground motion prediction equations (with basic and more complex functional forms), and a model using an artificial neural network. Hazard is estimated from these models and then we extend the approach to a risk assessment for an inelastic single-degree-of-freedom-system. Only the artificial neural network produces accurate hazard results below an annual frequency of exceedance of 1x10-3 years-1. This has a direct impact on risk estimates - with ground motions from large, close-to-site events having more influence on results than expected. Finally, an alternative to ground-motion modelling is explored through an observational-based hazard assessment which uses recorded strong-motions to directly quantify hazard.

3.1 Introduction

Seismic risk expresses the expected probable losses due to shaking, measured through different metrics e.g.: economic, social, and environmental (e.g., Musson, 2000). The basis of estimating seismic risk is through a hazard analysis, which establishes the likelihood of earthquake ground motion of a given intensity occurring in the area under investigation. Probabilistic predictions of earthquake hazard and risk can be obtained via two distinct approaches: unconditional and conditional (e.g., Scozzese et al., 2020).

Unconditional methods use direct observations to estimate hazard and risk, an example of this being Monte Carlo hazard assessment (e.g., Musson, 2000) or subset simulation (e.g., Au & Beck, 2003). The unconditional method is noted for its adaptability, flexibility, and conceptual simplicity, and has been used frequently in research to good effect, such as EqHaz (Assatourians & Atkinson, 2013) a program to assess seismic hazard. Unconditional methods are considered more robust approaches to estimate risk; however, they are computationally expensive, meaning they are rarely used in practice.

Conditional methods are a more practice-orientated approach to estimate risk. Such methods include the PEER Center's PBEE approach (Moehle & Deierlein, 2004). Cornell (2005), among others, discusses both the benefits and drawbacks of this type of approach, which is illustrated by Figure 3-1. These techniques require the definition of an intensity measure (*IM*) which describes the ground-motion intensity at the site of interest. Seismic hazard can then be evaluated by characterising the seismic source zones of the site and combining this with a GMM, to fully describe the frequency of exceeding this *IM* during a time period of interest. The next step is to perform structural analyses for a given system to calculate the conditional probability of exceeding a

given engineering demand parameter (*EDP*) at certain *IM* values; this can be done through various approaches, with the most common being IDA (Vamvatsikos & Cornell, 2002) and MSA (e.g., Mackie & Stojadinovic, 2005; Scozzese et al., 2020). Finally, risk is estimated by convolving the results from both these steps. It is worth noting here that this chapter uses two different terms to describe methods of predicting ground motion. Ground motion prediction equations (GMPEs) refer to traditional methods of predicting ground motions, i.e., those relying on regression analysis and a prescribed functional form, whereas GMM is used to refer to all models that predict ground motion.

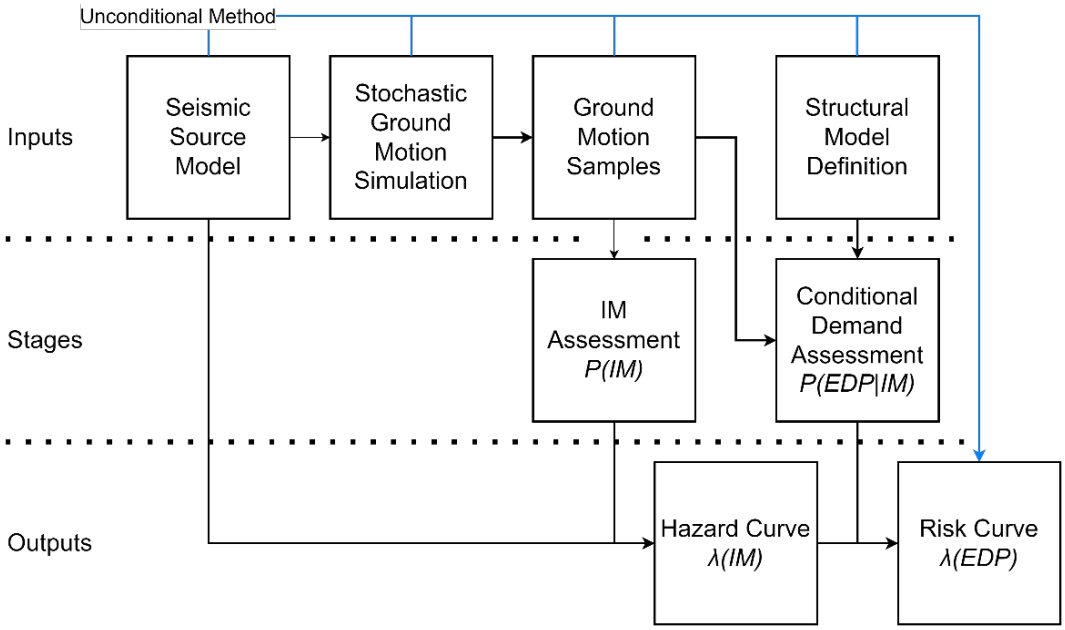


Figure 3-1: Workflow of the conditional approach to risk assessment, the unconditional approach is also presented (blue line). P(IM) represents the complementary cumulative distribution function (CCDF) of the intensity measure (IM) within the region of interest, whilst P(EDP|IM) is the CCDF of the engineering demand parameter (EDP) conditional on the IM.

One disadvantage of the unconditional approach is that it needs a large number of ground motions to be accurate; therefore, a rapid method of simulating ground motions is required. Stochastic ground motion simulations model the randomness of the

earthquake rupture process and seismic wave propagation, which cause ground motions, thereby creating samples from just a few seismological inputs. This research uses the same terminology as Boore (2003), i.e. the general means of simulating ground motions is referred to as the stochastic method, whilst the specific application of this method is called a stochastic model. Stochastic methods can create a large number of records rapidly, helping to remove potential gaps and biases in empirical data. This makes them attractive for conditional risk assessments where stochastic models can be used to bolster empirical datasets and improve PSHA, as is done by Meirova et al. (2018) building upon the SvE approach described in Shapira and van Eck (1993) to produce an updated PSHA for Israel. Kowsari et al. (2021) and Zolfaghari (2015) also used ground motion simulations to help improve PSHA in Iceland and Iran, respectively.

This study makes use of a stochastic model to directly estimate hazard and risk from simulated ground motions through the unconditional approach. Few studies explicitly compare conditional and unconditional methods. Both Jalayer and Beck (2008) and Franchin et al. (2012) explored the effects of using conditional and unconditional approaches on seismic risk estimates for a reinforced-concrete frame structure. Whilst Azar and Dabaghi (2021) and Bijelić et al. (2019) examined unconditional hazard and risk assessments, respectively, using the CyberShake software (Graves et al., 2011). Unconditional techniques also provide a good reference solution to evaluate different methods of carrying out a conditional risk assessment. Bradley et al. (2015) uses this direct method of estimating hazard and risk to compare different ground motion selection strategies when evaluating the peak displacement response of a nonlinear

SDOF system. Scozzese et al. (2020) uses the outputs of Monte Carlo analyses as a reference solution to investigate the accuracy of a conditional method based on MSA.

Moreover, Beauval et al. (2009) used earthquake simulations in the unconditional method to derive a hybrid deterministic-probabilistic hazard assessment, for the island of Guadeloupe (France). Comparisons were made between the built model and the Ambraseys et al. (2005) GMPE, but this is purely for illustrative purposes - as Douglas et al. (2006) showed that GMPE not to be well adapted for predicting earthquakes in the region. The study by Beauval et al. (2009) built on the work by Convertito et al. (2006) and Hutchings et al. (2007) to develop unconditional hazard assessments based on simulations of ground motion.

Stupazzini et al. (2021) also used ground motion simulations to perform hazard assessment using different GMMs and compare these results. The study also looked at their impact on conditional risk assessment procedure, estimating the risk to a portfolio of buildings for a case-study in Istanbul. Medel-Vera and Ji (2016) also proposed an approach to use ground-motion simulations in an unconditional risk assessment, although this study was specifically for nuclear facilities in the UK. They found results to be comparable to current practice, justifying the method's use for earthquake risk assessments.

When estimating hazard, and subsequently conditional risk, conditional methods are far more prevalent in the literature - as seen by the ever-growing number, and variety, of GMMs Douglas (2022). It is difficult to make an effective comparison between different conditional methods for risk assessment, especially with respect to the precision of the GMM. This is because there are a lack of high-quality ground-motion

data in many regions of the world (e.g., Xie et al., 2020) and so it is hard to establish an effective benchmark for models to be compared against.

Moreover, the data are being constantly updated, meaning that hazard models of the same region can provide different estimates. Gkimprixis et al. (2021) showed that using two different hazard models (built in different years) at the same site in Italy, yielded significantly different hazard results, which directly impacted on the estimated risk. Again, this shows a difficulty in comparing hazard models, as the benchmark that does exist is constantly changing, alongside its associated uncertainties.

The need for any comparison between hazard assessments could be made redundant if enough high-quality ground motion data was present in a region of interest, as the need for GMMs could then be removed completely. Sufficient data collected from observations of shaking used in a Monte Carlo hazard assessment could fully capture all forms of variability and uncertainty without the need for modelling, making the most accurate hazard assessment possible. The collation of large strong-motion databases, such as the NGA-West2 database (Ancheta et al., 2014), presents a dataset for this idea to be tested.

This paper aims to compare the impact of using different GMMs on estimates of the seismic hazard, and subsequently risk. For this purpose, a stochastic model is employed within a fictive scenario to simulate ground motions (Section 3.2), allowing the creation of benchmark hazard and risk estimates through the unconditional approach. The following sections describe the creation of three GMMs (Section 3.3), before investigating hazard assessment results from these GMMs against the created benchmark (Section 3.4). These results are extended to a risk assessment of a simple

structural model for both conditional and unconditional approaches (Section 3.5). For the conditional approaches an empirical fragility curve is created from the stochastic model (Section 3.5), allowing a convolution with the hazard estimates from each of the GMMs (Section 3.5). This procedure allows for direct comparison between the unconditional and conditional approaches, allowing judgements to be made on the impact that each of the conditional approaches has on risk. Finally, an observation-based hazard assessment approach is demonstrated (Section 3.6), which directly estimates hazard using a strong-motion database, to investigate how well recorded data can match the benchmark hazard.

3.2 Seismic Scenario and Stochastic Model

For this study, a fictive scenario is established with a circular source zone of radius 100km, and two faults of length 75km and 25km. Ground motions are computed at eleven stations along a line through the centre of the areal source, with hazard and risk calculated for a site at the centre-point of the region. This allows for different distances from the faults to be sampled – and so properly account for ground motion attenuation from the faults. 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 3-2.

The Atkinson and Silva (2000) stochastic model is implemented within this study. The model is altered by the generalised double-corner-frequency source spectral approach of Boore et al. (2014) to allow the stress drop of each source (detailed in Figure 3-2) to be changed, therefore changing the strength of each source, and allowing the ground motions from each source to be accounted for when developing hazard assessments.

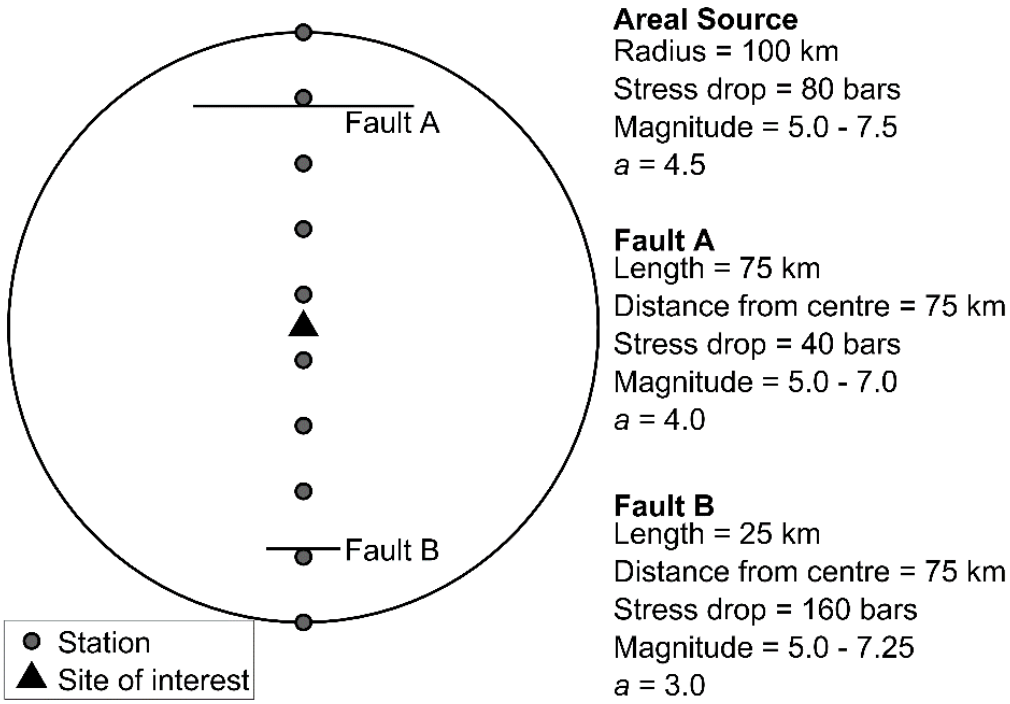


Figure 3-2: The seismic source model used for this study

The stochastic model is used to obtain ground motions for this scenario, with the *IM* assumed as the spectral acceleration (S_a *(T, ζ))* corresponding to a damping ratio ζ=5% and a period T=0.2s. One hundred catalogues are generated for a time period of 25,000 years to form the hazard and risk assessments. The number of events occurring from each source is a function of their yearly activity. Multiplying this by the number of years to produce a total of 10,605 events within the time frame from all sources. Each event is recorded by all 11 stations, yielding a total of 116,655 records per catalogue. All stations are used to derive GMMs; however, hazard and risk is only evaluated at the site of interest, defined in Figure 2. A large number of simulations are used within this research to capture the precision of both the unconditional and conditional approaches, with mean hazard and risk estimates presented alongside their accompanying uncertainties.

Both magnitude and distance samples are generated based on their respective probability distributions to create ground-motion samples using the stochastic model. The distributions of the simulated magnitude, distance, and *Sa* were evaluated to ensure that they conform to expectations.

The assumption is made that the stochastic model used produces accurate and realistic ground motion intensities for the earthquakes modelled in this study. This is justified by various studies that validate the stochastic method (e.g., Silva et al., 1996; Tsioulou et al., 2019). This assumes the unconditional hazard and risk estimates from the stochastic model are *the truth* and so act as a benchmark for the created GMMs to be compared against.

3.3 Ground-Motion Models

Separate to the unconditional analysis, the simulated ground motions are used to create three GMMs. These include a basic and more complex GMPE created through least squares regression analysis, with the functional forms shown in Equation 3-1 and Equation 3-2 respectively:

$$\ln(Sa) = C_0 + C_1 M + C_2 \ln(R + 5) + C_3 Fault\_A + C_4 Fault\_B \quad (3\text{-}1)$$

where, $C_0 = -2.6642$,$C_1 = 1.110$, $C_2 = -1.6812$, $C_3 = -0.4639$, and $C_4 = 0.2926$.

$$\ln(Sa) = C_0 + C_1 M + C_2 \ln(R + 5) + C_3 M^2 + C_4 R + C_5 Fault\_A + C_6 Fault\_B \quad (3\text{-}2)$$

where, $C_0 = -8.0230$, $C_1 = 2.4141$, $C_2 = -1.1646$, $C_3 = -0.1134$, $C_4 = -0.0073$, $C_5 = -0.4154$, and $C_6 = 0.3748$.

In these equations, M represents magnitude, R distance (km) and Sa is 5% damped spectral acceleration (g). The terms $Fault\_A$ and $Fault\_B$ equal 1 to predict ground motions from Fault A or Fault B and 0 otherwise.

A feedforward ANN is also considered, consisting of a single hidden layer of five nodes. The input of the ANN is the same as the GMPEs (i.e., M, R, $Fault\_A$ and $Fault\_B$) and the output remains as $\ln(Sa)$. The ANN uses the Levenberg–Marquardt optimisation technique (e.g., Dhanya & Raghukanth, 2018) and five nodes are used to prevent over-fitting (e.g., Derras et al., 2014).

Machine learning tools have become more common in the field of engineering seismology in recent years (e.g., Kong et al., 2018). There is clear potential for their use in ground-motion prediction, helping to model complex nonlinear behaviours of ground motion that the fixed functional form of a GMPE may fail to capture (e.g., Alavi & Gandomi, 2011), providing there are enough data. An ANN was selected for this study as its use has been well investigated for the purpose of ground-motion prediction; further research could investigate the use of other machine learning tools, as discussed by Khosravikia and Clayton (2021).

Each of these GMMs are used in PSHA to predict ground motion intensity samples based on the simulated magnitude-distance combinations from the stochastic earthquake catalogue. The conditional hazard analyses rely on the standard deviation (sigma) of each GMM to introduce variability in results when estimating *Sa*, whilst the unconditional approach already models the variability in ground motions by using every simulated record in the hazard analysis. On top of this, a model-error parameter (ε_{mod}) is introduced to the stochastic model, as proposed by Jalayer and Beck (2008).

This parameter scales the radiation spectra calculated by the stochastic model in order to account for modelling uncertainty. It is characterised by a normal distribution with mean of zero and standard deviation (in natural logarithm) of 0.5. For this study, only the total sigma is considered within the GMMs created. Future research could consider the effects of inter-event and intra-event variabilities separately.

3.4 Results

In this section, results from the conditional hazard models are presented and compared to each other, and against the results obtained using the unconditional approach. This includes comparing the residuals from each GMM, the returned median spectral acceleration values from the GMMs, a comparison of the hazard results produced by each of the models, and an investigation into the differences in these hazard results.

3.4.1 Residual Analysis

Residual plotting of the three models indicated that magnitude-dependant sigmas could be considered (e.g., Youngs et al., 1997). As such, plotting magnitude (binned at 0.1 intervals) against sigma for each of these intervals showed a relationship between these two variables. An ANN (with a single hidden layer of two neurons) was fitted for each of the models to predict sigma based on magnitude. The ANN was selected to achieve a good fit of these data, ensuring accurate sigma prediction. A summary of the sigma values for each of these six considered models is presented in Table 3-1. The basic GMPE constructed has a higher sigma value than both the complex GMPE and ANN, implying it has a worse fit to the true ground-motion samples than the other two models. The sigmas are slightly smaller than generally observed for GMPEs obtained from actual ground-motion records (e.g., Douglas et al.,

2014). This is because the stochastic model does not include all sources of variability in earthquake ground motions.

Table 3-1: Sigma values from the ground motion models created in this study.

Model	Sigma (natural logarithm)
Basic GMPE	0.5548
Basic GMPE (magnitude dependant sigma)	0.5411 – 0.5557
Complex GMPE	0.5455
Complex GMPE (magnitude dependant sigma)	0.5167 – 0.5498
ANN	0.5437
ANN (magnitude dependant sigma)	0.5135 – 0.5479

3.4.2 Predicting spectral acceleration

Median predictions of *Sa* for the three conditional models are presented on Figure 3-3 for fixed distances of 25km and 75km, and fixed magnitudes of 5.75 and 6.75: with dashed lines on Figure 3-3 representing plus and minus one standard deviation from the median. Magnitude-dependant sigma models are not included in this plot as median *Sa* predictions are not affected by this. As the differences between each of the GMMs was consistent across all three of the seismic sources, only the median Sa predictions from the background source are presented here.

For the fixed distance plots, predictions from the ANN appear almost identical to that of the complex GMPE: whilst all three models appear to vary from each other at extreme distances in the fixed magnitude plot. There are similarities in median *Sa* prediction at points where there are a wealth of magnitude and distance samples, allowing each of the models to be sufficiently trained. The differences in predictions appear where fewer events are expected in the catalogue. This is most noticeable with the basic GMPE, which considerably overpredicts *Sa* values for higher magnitudes

and extreme distances due to the simplicity of its functional form, meaning it cannot fully describe the possible ground motions for all possible earthquakes.

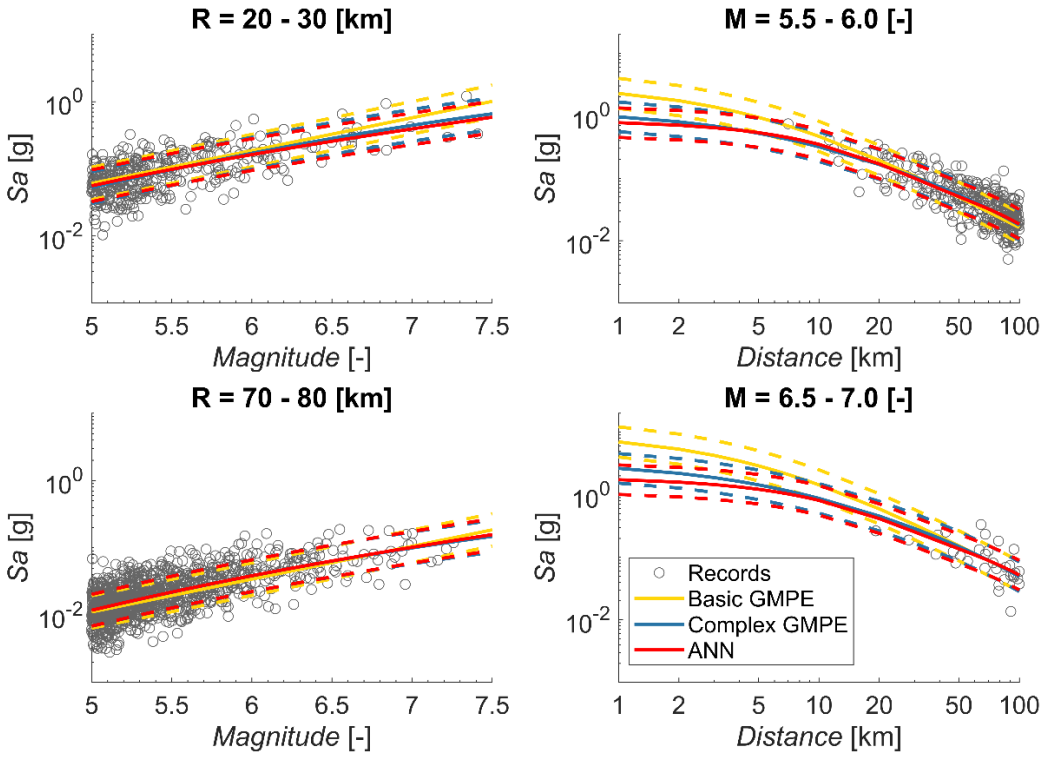


Figure 3-3: Median spectral acceleration predictions from the area source for all three conditional models (solid lines), at fixed distances of 25km and 75km, and fixed magnitude of 5.75 and 6.75, dashed lines represent plus and minus one standard deviation of the median.

3.4.3 Assessing hazard

The hazard, or MAF, is estimated for each source by finding the number of ground motions that exceed an *Sa* threshold and multiplying it by the seismicity rate of the source: this is performed on a series of *Sa* thresholds (finely discretised at 1.0×10^{-3}g spacing between 0g and 2.5g) and summed across all sources at these thresholds, to produce the overall site hazard. The process is repeated for 100 sets of records to obtain the MAF. This procedure is performed on the simulated ground motions in the case of the unconditional Monte-Carlo based hazard assessment, and on the predicted ground

motions from the GMMs in the case of the conditional methods. Figure 3-4 shows the mean hazard curves for all created models: with dashed lines representing the 16^{th} and 84^{th} percentiles of the mean.

The standard deviation of MAF values is also computed, allowing the calculation of the 16^{th} and 84^{th} percentiles - assuming a lognormal distribution. All models appear to have a similar predictive quality at more frequent hazard occurrences, matching until an MAF of exceedance of 0.025 $years^{-1}$, but are quite different for lower MAFs. The hazard curves assuming magnitude-dependant sigma models show similar behaviour to the fixed sigma alternatives and so results from these models are not shown in the following plots. This similarity between magnitude-dependent and magnitude-independent GMMs could be because the magnitude dependency of sigma is quite small in the developed GMMs.

Since the differences between the hazard curves are large when using a wealth of simulated data, they will likely be even more significant when using actual strong-motion records, which are fewer in number and sparser in distribution. This indicates that GMM selection can be important when carrying out a hazard analysis, as these inaccuracies will be propagated through to the risk assessment, leading to poorer loss estimates.

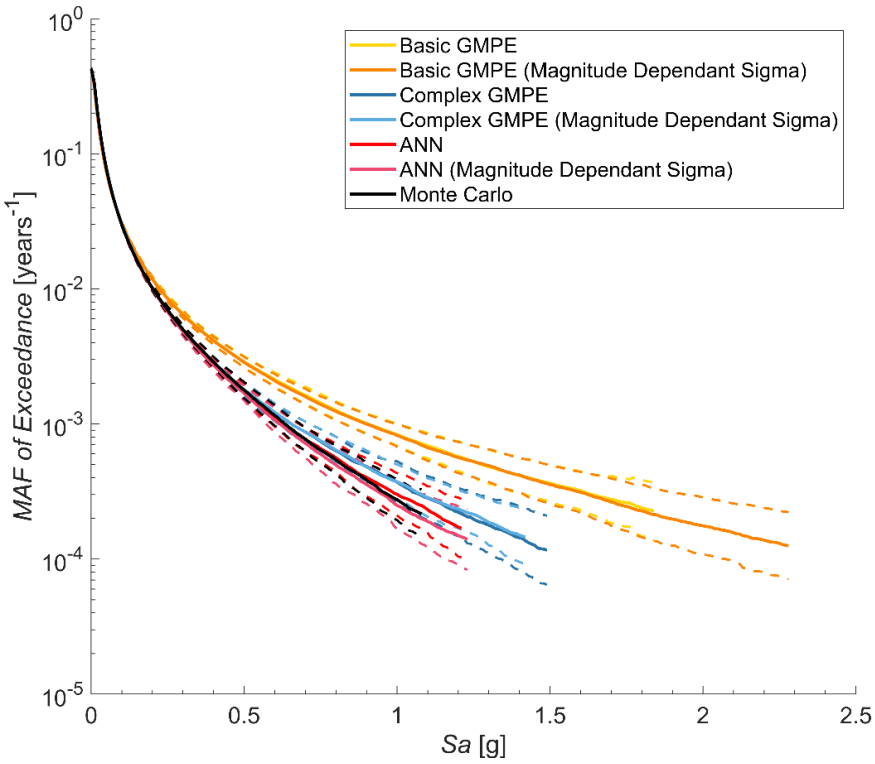


Figure 3-4: Mean hazard curves for all created models, solid lines show the mean hazard results whilst dashed lines show the 16th and 84th percentiles of the mean. Spectral acceleration at a period of 0.2 seconds is plotted on a linear scale to better show the differences between the results.

To check whether the differences in hazard curves were persistent, the same procedure was carried out to predict *Sa* at a period of *T*=1.0s. Figure 3-5 presents the MAF of exceedance hazard curves for three GMMs derived for *T*=1.0s. There is still a difference between the results from each of the GMMs, especially with respect to the basic GMPE; however, the models are more similar than when evaluated at *T*=0.2s. Further research could investigate whether the differences between conditional hazard models is maintained for a range of *IM*s. For this study, the differences in hazard estimates of *Sa* at *T*=1.0s are not further investigated as the differences for this spectral period still indicate the importance of GMM selection for risk assessment.

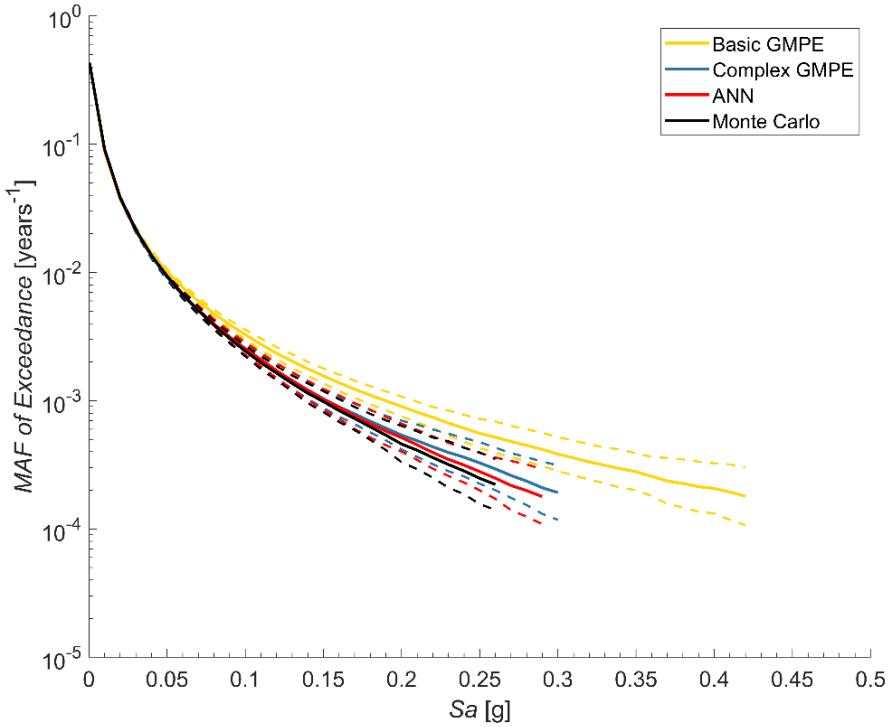


Figure 3-5: Mean hazard curves for all created models using Spectral acceleration at a period of T=1s, solid lines show the mean hazard results whilst dashed lines show the 16th and 84th percentiles.

A different set of mean hazard results can also be obtained by finding the mean ground motion intensity for a fixed set of MAF of exceedances. The suitability of these two approaches has been discussed previously and they were found to yield distinct hazard curves (e.g., Bommer & Scherbaum, 2008). Interestingly, with the large suite of simulated data created by this study, both approaches to calculate the mean hazard yield very similar results, as shown in Figure 3-6.

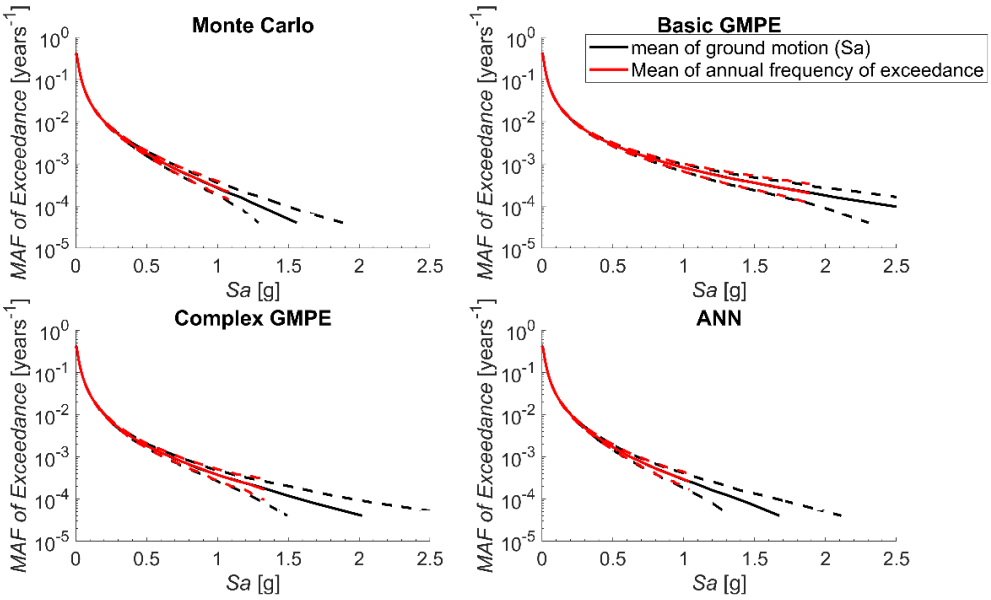


Figure 3-6: Comparison of mean hazard calculations via two approaches; calculating the mean of the annual frequency of exceedance; and calculating the mean spectral acceleration. Solid lines show the mean hazard, with dashed lines showing the 16th and 84th percentiles.

3.4.4 Evaluating differences between hazard results

The Kolmogorov–Smirnov (KS) test (e.g., Stephens, 1974) is used to compare results from the conditional hazard models against the benchmark hazard results. The test is carried out by finding the maximum absolute difference between two CDFs tested. Although different techniques exist to compare hazard results such as Cohen's effective size (e.g., Malhotra, 2014), the KS test was chosen for its ease of application. As a non-parametric test, the KS test does not rely on the assumption and description of a probability distribution. It evaluates differences between the entire range of the two distributions, so it can ascertain differences at the extremes of the distribution, including stronger ground motions, which are more likely to be different in the various hazard results.

The null hypothesis of the KS test is that the two CDFs are drawn from the same population. A CDF can be obtained from a hazard curve by first converting the frequency of exceedance for a range of *Sa* values to a probability. Given that our SSM assumes a Poisson process, Equation 3-3 converts MAF of exceedance to annual probability of exceedance:

$$P = 1 - \mathrm{e}^{-\lambda} \qquad (3\text{-}3)$$

Where, P is the annual probability of exceeding a certain *Sa* value, and λ is the MAF of exceeding the same *Sa* value at the site i.e., the cumulative MAF of exceedance from each of the three sources. One minus the annual probability of exceedance provides the CDF value for the given *Sa*. CDFs for each hazard model are obtained by performing this procedure on hazard results for *Sa* values at 0.01g intervals between 0 and 2.5g.

Each of the CDFs from the conditional approaches can then be tested against the CDF of the reference solution. Failing the test implies that the created GMM did not generate similar hazard estimates to the reference curve. This would imply that it would not be an appropriate model to use for risk assessment.

Table 3-2 presents the returned *p*-value of the KS test, with only the ANN model passing the test at the 5% significance level – agreeing with visual inspection of the hazard curves. Thus, this is the only model that could be considered as coming from the same distribution as the simulated data and so provides the best prediction of the hazard.

Table 3-2: Results from Kolmogorov-Smirnov test on the three conditional hazard models.

Model	Basic GMPE	Complex GMPE	ANN
P-value	<0.001	0.0136	1.0000
Reject null hypothesis?	✓	✓	✗

The KS test is known for its sensitivity when the CDF is described at different sampling points, which could influence the returned result from the test, and this holds true for this scenario. For instance, carrying out the test at intervals of 10^{-4}g spacing between 0 and 2.5g rejected the null hypothesis for all three models; whilst using intervals of 0.1g for the same *Sa* bounds only rejected the null hypothesis for the basic GMPE. Ultimately, the interval spacing of 0.01g used to perform the test was considered acceptable for this scenario as this broader spacing is more likely to be used when comparing conditional hazard models to real world data. Nevertheless, further research should consider using other statistical tests to test the similarity between the hazard results, because of the sensitivity of the KS test.

3.4.5 Hazard disaggregation

To investigate the differences in hazard predictions, hazard disaggregation (e.g., Bazzurro & Cornell, 1999) is performed. This breaks down the ground motions into the factors that contribute towards hazard, i.e., in this case magnitude, distance, and epsilon. Figure 3-7 plots disaggregation results from both the Monte Carlo and conditional hazard analysis approaches for *Sa* values of 0.1g and 1.0g – mean magnitude and distance are also provided. Results are very similar at 0.1g, as expected by the agreement of the hazard curves (Figure 3-4) at this value of *Sa*. However, when disaggregation is performed at 1.0g, results vary between each of the models, as seen by differences in the plots.

The disaggregation results at 1.0g show a change in the dominant earthquake scenario. The mean magnitude-distance combination for the Monte Carlo-based approach at 0.1g is 5.86 and 33.56km, respectively, whilst at 1.0g this changes to 6.51 and 10km. At smaller magnitudes and further distances, the GMPEs accurately predict ground motions but at higher magnitudes and shorter distances, the GMPEs poorly predict the less abundant, stronger ground motions. This creates greater inaccuracies within the hazard assessment, which could lead to poor quality risk assessments, reaffirming the importance of GMM selection.

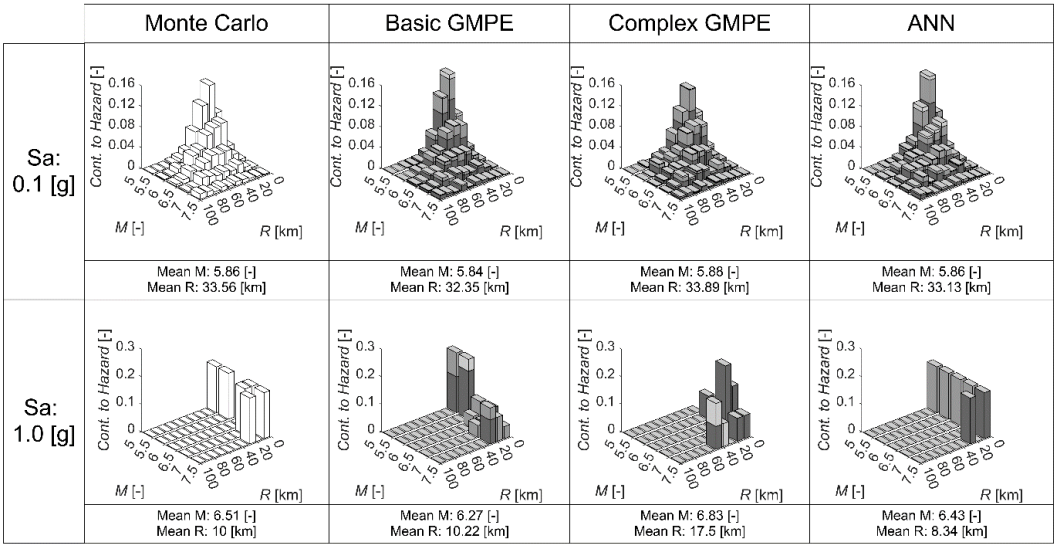


Figure 3-7: Hazard disaggregation results from the Monte Carlo-based approach and three conditional hazard assessment approaches, for spectral acceleration values of 0.1g and 1.0g.

3.4.6 Restricting the magnitude and distance ranges of seismicity

All hazard models appear to perform well for smaller *Sa*. This is well indicated by the hazard curves of Figure 3-3 and the hazard disaggregation of Figure 3-7. To confirm this agreement, hazard assessment is performed again on each of the models, but this time the seismicity is restricted to magnitude and distance ranges of 5.5-6.5 and 10-50km, respectively. In order to create this restricted scenario, both faults are halved in

length and in their distance from the centre of the site, so that all distances simulated from the stochastic model will be between 10 and 50km. The mean hazard results for this restricted scenario are plotted on Figure 3-8.

When comparing the hazard curves created by this new, restricted, scenario, all models match quite well, with the ANN mirroring the true values for the whole range of values under investigation. This shows that the models are a good fit for the simulated data for this range of interest, and that discrepancies in hazard predictions between the models (when considering the whole range of magnitudes and distances), is likely down to the influence of ground motions caused by higher magnitudes and smaller distances, as suggested by the hazard disaggregation.

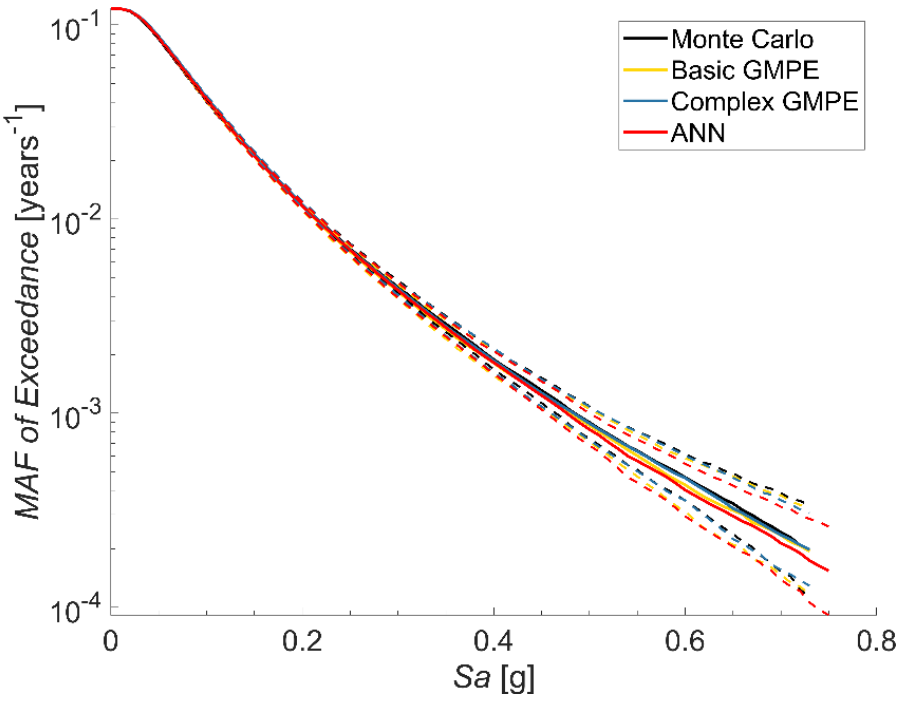


Figure 3-8: Mean hazard curves produced by the Monte Carlo-based hazard approach and three conditional hazard approaches, for restricted distance range of 10 to 50km and magnitude range of 5.5 to 6.5, plotted by solid lines. Dashed lines show 16th and 84th percentiles of mean hazard.

3.5 Extension to risk assessment

To demonstrate the impact of GMM selection on seismic risk assessment, each hazard model was extended to assess the risk of an inelastic SDOF system, with system ductility selected as the *EDP*. The system has elastic-perfectly plastic behaviour designed to withstand a ductility factor of q=4 for a MAF of exceedance of 2.1x10^{-3} years^{-1}; with elastic period T=0.2s, and yield displacement μ_y=0.0013m. The yield displacement was calculated for the ductility factor, based on the assumption that the SDOF system is in the medium ductility class defined in Eurocode 8 (European Committee for Standardization (CEN), 2004). As the structure is in the short period range, the equal energy rule is used (Equation 3-4) as inferred by the N2 method (Fajfar, 1999) and Eurocode 8:

$$q_d = \frac{\mu}{\mu_y} = 1 + (q-1)\frac{T_c}{T} > q \tag{3-4}$$

where q_d is the modified ductility factor, u is the inelastic displacement of the SDOF, μ_y is the yield displacement, q is the ductility factor, T_c is the corner period of the SDOF (T_c=0.5s according to Eurocode 8 (EN-1998-1)) and T is the period of the SDOF. The inelastic displacement of the SDOF, u, is estimated by multiplying the design elastic displacement with the Newmark and Hall inelastic displacement coefficient, C, (Newmark & Hall, 1982), where C was found to equal 2.34.

Nonlinear dynamic structural analyses were carried out for all 100 sets of 10,605 ground-motion samples at the site of interest to calculate the maximum system displacement for each record. Displacements are normalised by the system's yield threshold to produce the system ductility demand.

For the unconditional Monte Carlo-based approach, seismic risk can be assessed by finding the annual frequency of exceeding a ductility demand threshold, as illustrated by Equation 2-8, using ductility demand (μ) as the EDP. The mean risk estimate is the average rate that the ductility demand value is exceeded for all 100 sets of records. Reproducing this assessment for a series of ductility demand thresholds creates a Monte Carlo-based risk curve.

To estimate risk for the conditional approach, a fragility curve must be created and combined with the hazard results. An empirical fragility curve is created for each ductility threshold by finding the conditional probability that the system ductility exceeds the threshold level, given the intensity measure of shaking. These are derived from the simulated hazard and ductility values, with fragility curves for ductility thresholds of 1, 2, 3, 4 and 5 shown in Figure 3-9. To produce a fragility curve, *Sa* is separated into 40 bins: displayed by the dashed lines on Figure 3-9. The probability of exceeding a ductility threshold is then calculated for each bin for all 100 sets of records, with the process repeated for each ductility threshold to create a series of fragility curves.

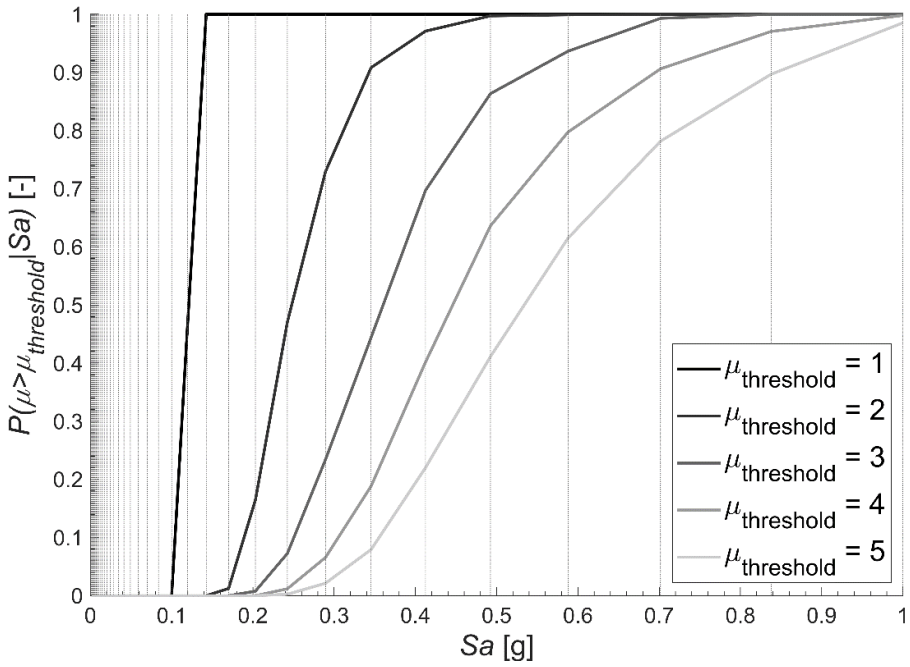


Figure 3-9: Example empirical fragility curves created for conditional risk assessment, at ductility thresholds 1, 2, 3, 4 and 5. Curves are created as a mean from all 100 sets of catalogues.

Conditional risk estimates are made for each ductility threshold by convolving the corresponding fragility curve with the hazard curves for each GMM (e.g., Baker et al., 2021). Risk curves for both the Monte Carlo based approach and conditional approaches are presented in Figure 3-10, with mean risk (solid lines) and the 16th and 84th percentiles of the mean risk (dashed lines) presented.

The basic GMPE considerably overpredicts risk at estimates lower than approximately 0.025 years^{-1}. For instance, for a ductility of 2.0, the basic GMPE overpredicts risk by 24% and at a ductility of 4 the benchmark risk is overpredicted by 35%.

Both the complex GMPE and ANN lead to far better risk estimates, with the ANN marginally better than the complex GMPE, especially at higher ductility values. This extension reinforces the results from the hazard assessment, demonstrating the importance of GMM selection when carrying out a risk assessment.

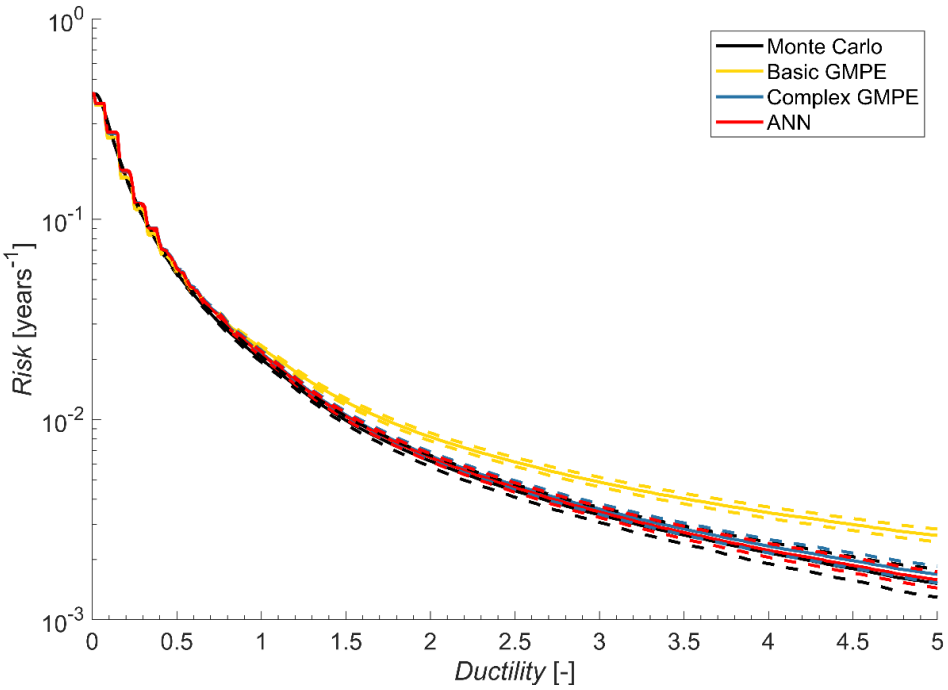


Figure 3-10: Mean risk curves for all created models evaluating a single degree of freedom system with period T=0.2s, solid lines show the mean risk whilst dashed lines show the 16th and 84th percentiles of the mean.

For completeness, risk is also estimated for an SDOF where the period of the system is changed to T=1.0s. As this structure is in the medium-long period range, the equal displacement rule can be used, as per Eurocode 8 (EN-1998-1). This assumes that the peak elastic displacement of the system is equal to the peak inelastic displacement of the system. Therefore, dividing the calculated elastic displacement of the SDOF by the ductility factor q=4, the system is found to have yield displacement u_y=0.0067m. Figure 3-11 presents mean risk curves with both 16th and 84th percentiles of the mean (dashed lines) for the unconditional and conditional approaches investigate. Similar to the hazard assessment for a period of T=1.0s, the conditional risk estimates appear to match the unconditional approach more closely – however there is still a noticeable difference between the two approaches.

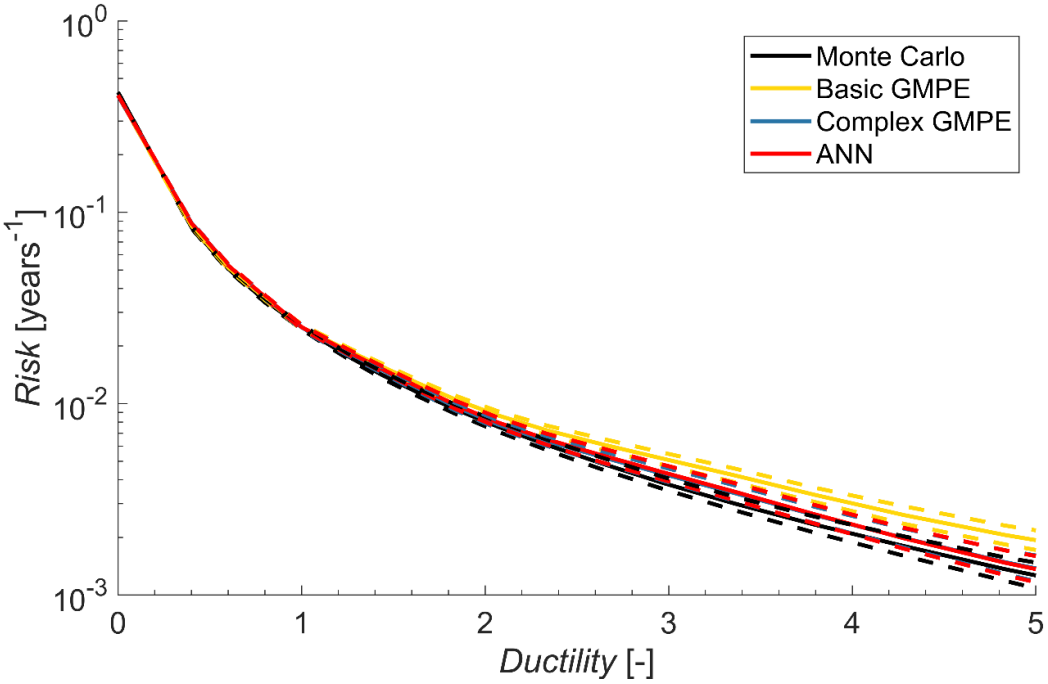


Figure 3-11: Mean risk curves for all created models evaluating a single degree of freedom system with period T=1.0s, solid lines show the mean risk whilst dashed lines show the 16th and 84th percentiles of the mean.

3.6 Observation-based hazard assessment

So far, this study has compared the impact that different GMMs have on risk estimates. However, if sufficient real-world ground motions had been recorded, stochastic models and GMMs would no longer be needed to assess seismic risk. Instead, the empirical data could be implemented directly within a Monte Carlo analysis to estimate the hazard and risk in the area – as has been done with the simulated ground motions in this study. To test this idea, an "observation-based" hazard assessment is presented here, where the hazard estimated using real strong-motion records is compared against the benchmark hazard.

First, to improve the match between the benchmark hazard and the empirical data, a simpler seismic scenario is considered from the previous study: leaving just the areal source. To provide strong motion records the NGA-West2 database (Ancheta et al.,

2014) was selected. This is one of the largest databases yet compiled so there is a good chance of containing enough records to provide a comparable hazard assessment to the benchmark. Moreover, NGA-West2 includes mainly records from the same geographical region as the Atkinson-Silva stochastic model (i.e., western United States) – so ground motions from this database should be similar to those simulated by the stochastic model: making a comparison of this method against the benchmark hazard suitable.

To estimate hazard from this observational data, both the simulated magnitude-distance pairs from the stochastic model, and magnitude-distance combinations from the NGA-West2 database, are binned. The bins, and the number of magnitude-distance pairs in each bin, are shown in Figure 3-12. Records that fall outside of these bins (i.e., with magnitudes less than 5 and greater than 7.5, and hypocentral distances greater than 100km) are discarded as they fall outside the range of the stochastic model used.

Each simulated magnitude-distance combination is randomly matched to an NGA-West2 record that falls into the same bin – with the corresponding empirical ground motion assigned to this simulated event. For example, if the simulated magnitude-distance pair fell in the range of 5.5-6.0 magnitude and 25.12-39.81km, the ground motion assigned to this simulated event would be one of the 129 records that fell into the same bin. Records are randomly selected from each bin to introduce variability into the hazard model. The process is repeated for all 100 sets of records to produce mean hazard results.

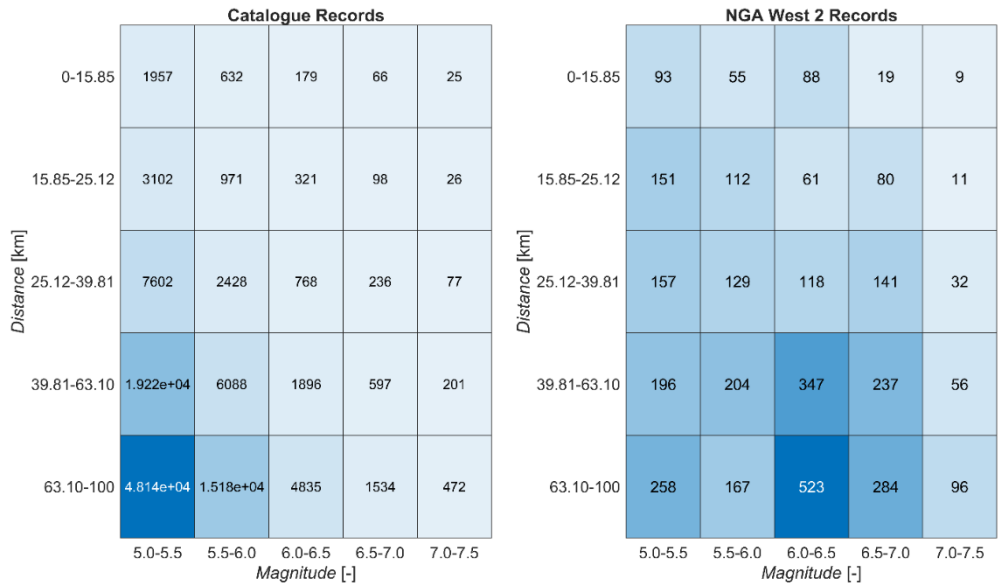


Figure 3-12: Heat-map showing number of records belonging to each magnitude-distance bin from (a) an example simulated catalogue and (b) the NGA-West2 database.

The stochastic model used in this research is calibrated to rock sites; however, there was insufficient data to only include rock sites from the NGA-West2 database in this analysis. To account for this, two different approaches were considered. The first where all sites are used within hazard assessment, irrespective of the average shear-wave velocity in the top 30m of the ground (*Vs30*). The second, where only sites with *Vs30* greater than 450 m/s are considered. Figure 13 shows the mean hazard curves produced from these two approaches, with dashed lines representing the 16th and 84th percentiles.

Figure 3-13 shows that the purely observational method with no site restrictions consistently over-predicts hazard. This is likely down to the NGA-West2 database being dominated by sites with relatively low *Vs30*, meaning site amplification is greater and stronger ground motions are produced than the fixed *Vs30* of 620 m/s in the benchmark stochastic model. The site restricted model only produces an improved

hazard prediction at small *Sa*, suggesting that there are an abundance of sites with a low *Vs30* producing small ground motions.

To try and achieve a better fit to the benchmark hazard assessment, a *Vs30*-based adjustment is introduced. This adjustment scales the observed ground motions by a site amplification factor. The site effects equation from the Chiou and Youngs (2014) GMPE was used to estimate this site amplification factor. This model was calibrated for the NGA-West2 database and hence is appropriate for this adjustment. This modelling was used for both the site restricted and unrestricted approaches to produce two new observation-based hazard curves, also shown in Figure 3-13.

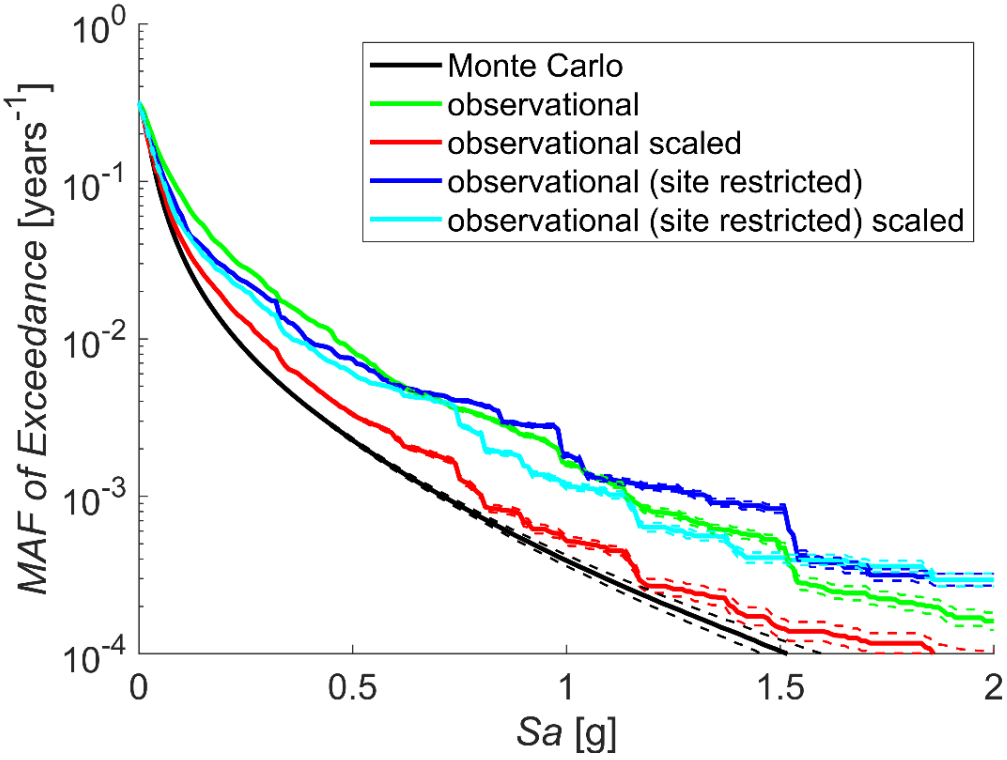


Figure 3-13: Mean hazard curves for the four observation-based hazard assessment models compared to the unconditional benchmark hazard, solid lines show the mean hazard whilst dashed lines show the 16th and 84th percentiles of the mean

Out of these two new hazard curves, only the non-restricted model with site effects scaling (denoted as "observational scaled" in Figure 3-13) makes a great improvement

on results. This model matches the benchmark considerably better than the other curves, across the whole range of *Sa* values, only slightly overpredicting hazard. Although promising, this method does not provide a particularly accurate description of the scenario hazard, when compared against the benchmark. To improve this, further study could consider a larger database with a closer match to the site conditions of the stochastic model in the hopes to obtain a better hazard assessment without the need for the introduction of any ground-motion modelling at all.

It is worth noting that the uncertainties presented by Figure 3-13 appear small. This is likely down to the poor sampling that can be observed for the observation-based models. For example, in the bin for magnitudes 5.5-6.0 and distances 39.81-63.10km there are 204 empirical records and 6088 simulated records – meaning each record would be sampled on average 30 times for this set of data. Using a bigger database may resolve this issue to capture the natural variability of ground motions more accurately.

3.7 Conclusions

In this study, a comparison of different ground motion prediction methods has been carried out, in regard to their impact on hazard and risk estimates. A fictive scenario was established, with a stochastic model employed to simulate ground motions and build these models.

Three ground-motion models were considered: a basic and more complex classical ground motion prediction equation, and an artificial neural network. An empirical fragility curve was also created from simulated ductility data, before being convolved with each of the ground-motion models to produce risk estimates via the conditional

approach. Alongside this, simulations were directly used in a Monte Carlo analysis to directly estimate benchmark hazard and risk results, to which the ground-motion model-based estimates were compared.

Finally, an observation-based hazard assessment technique was outlined that demonstrated potential to estimate hazard when compared against the benchmark results, if a site-effect scaling was applied. This technique indicates that in certain situations it may be possible to estimate hazard and risk purely from empirical data, without needing a ground-motion model.

Conditional results show that careful selection of a ground-motion model is required to obtain the best estimates of seismic risk. Out of the three ground-motion models created, only the artificial neural network appears to produce hazard estimates similar to the benchmark results, and the same outcome is visible from the risk estimates. Although it is important to note that such models are only successful when trained on large, complete, datasets - something that is hard to replicate in the real world.

Hazard disaggregation was performed on each hazard model. With the ground motion prediction equations struggling to predict high magnitude, short distance, events at higher ground-motion levels, causing the over-prediction of hazard, and ultimately risk. It may be possible to partially reduce this problem by using more complex functional forms, but these are difficult to constrain without large datasets. Uncertain risk estimates may result in inaccurate risk assessment and design, and ill-informed decision making. If estimates are poor in a data-rich scenario, they will be worse in the real world where data are less comprehensive.

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

This Chapter has been adapted from:

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

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 chapter aims at evaluating 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 IMs 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, e.g., 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 provide the best risk estimates, if sufficient samples are available, whilst ground displacement and duration-based IMs 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.

4.1 Introduction

The seismic risk of structures can be assessed via two distinct approaches (e.g., Bazzurro et al., 1998; Scozzese et al., 2020). 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 (e.g., Bradley et al., 2015). 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 (2005), 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 4-1). 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 (e.g., Bradley, 2013; Cornell et al., 2002).

One of the most popular conditional approaches for assessing earthquake risk is the PEER's PBEE framework (e.g., Cornell & Krawinkler, 2000; Moehle & Deierlein, 2004). The original intention of the framework was to report risk estimates that represent the whole structure. However, performing performance assessments at component-level can provide a more thorough understanding of a building's performance (e.g., Cremen & Baker, 2019), given the high contribution of non-structural components to overall earthquake-induced losses (e.g., Taghavi, 2003). Therefore, more recent implementations of the PBEE approach (e.g. the FEMA P-58 method, FEMA, 2012) include 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.

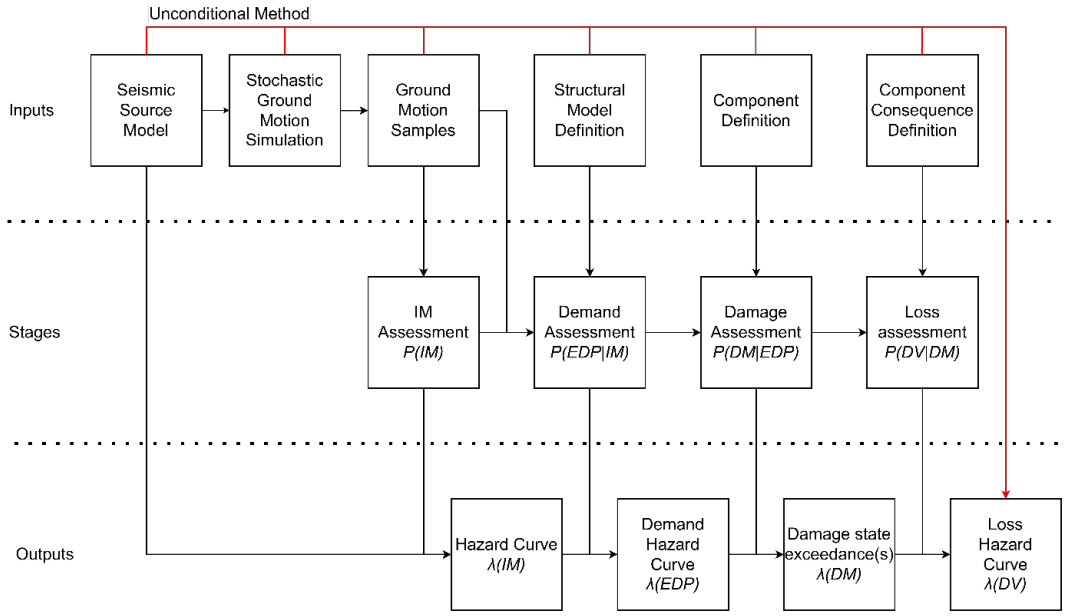


Figure 4-1: Workflow of the conditional approach to risk assessment when ground motion simulations are utilised. The unconditional approach can be described by the red workflow.

The importance of *IM* selection is well known, with a range of *IMs* proposed in the literature. These generally consider one or more of three ground-motion characteristics: amplitude, frequency, and duration (e.g., Baker et al., 2021). 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 *IMs* would describe ground motions more accurately (e.g., Baker & Cornell, 2005; Gehl et al., 2013; Vamvatsikos & Cornell, 2005). 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) (Luco & Cornell, 2007), and enough ground-motion records are used to characterise the system's response. However, satisfying these criteria is not always possible, and the need of limiting the number of numerical simulations has led to a proliferation of alternative *IMs* and comparisons in terms of *efficiency*.

Due to the large number of *IMs* in existence, the topic of selecting an optimal *IM* has become a well-covered, and wide-ranging debate (see e.g. Katsanos et al., 2010, 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 (2003) evaluated *IM* selection for PSDMs of highway bridges in California. Whilst Bray and Travasarou (2007) investigated the impacts of *IM* selection for estimating seismic slope displacements. Kohrangi et al. (2016) compared the impact of selecting eight different structure-specific *IMs* 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. (2016), who made loss estimates within their study on 3D building models, O'Reilly (2021) who investigated *IM* selection for seismic risk assessment of bridges, and Du et al. (2020) who provided insight on the influence of *IM* selection in the context of regional seismic risk assessment. However, most studies that compare *IMs* 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, e.g., damage estimates. There has also been little evaluation of *IM* impact through comparison with an unconditional benchmark, although Kwong et al. (2015a, 2015b) 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, i.e., the response of a system due to ground motions, represented by the conditional demand assessment stage in Figure 4-1. Several different methods have been proposed to describe this relationship, including MSA (e.g., Mackie & Stojadinovic, 2005; Scozzese et al., 2020) and IDA (Vamvatsikos & Cornell, 2002). Possibly the simplest of these methods to implement is Cloud analysis (e.g., Mackie & Stojadinovic, 2005) 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 (e.g., O'Reilly & Monteiro, 2019; Tubaldi et al., 2016); machine learning tools have also been used to improve this

characterisation (e.g., Soleimani & Hajializadeh, 2022). 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 *IMs* 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 4.2), the structural model (Section 4.3) considered for the study, and the *IMs* evaluated (Section 4.4). Current methods for evaluating the optimal *IM* are also discussed, and a new method to determine the optimal *IM* is presented (Section 4.4). The conditional risk estimates obtained for each *IM* using all five seismic demand models are presented (Section 4.5), before the impact of each *IM* and demand model is evaluated (Section 4.6). Finally, damage estimates are made for two components with 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 4.8).

4.2 Seismic scenario

For this study, a fictive scenario is established with a circular source zone of radius 100km, and two faults of length 75km and 25km. 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 4-2. 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 4-2.

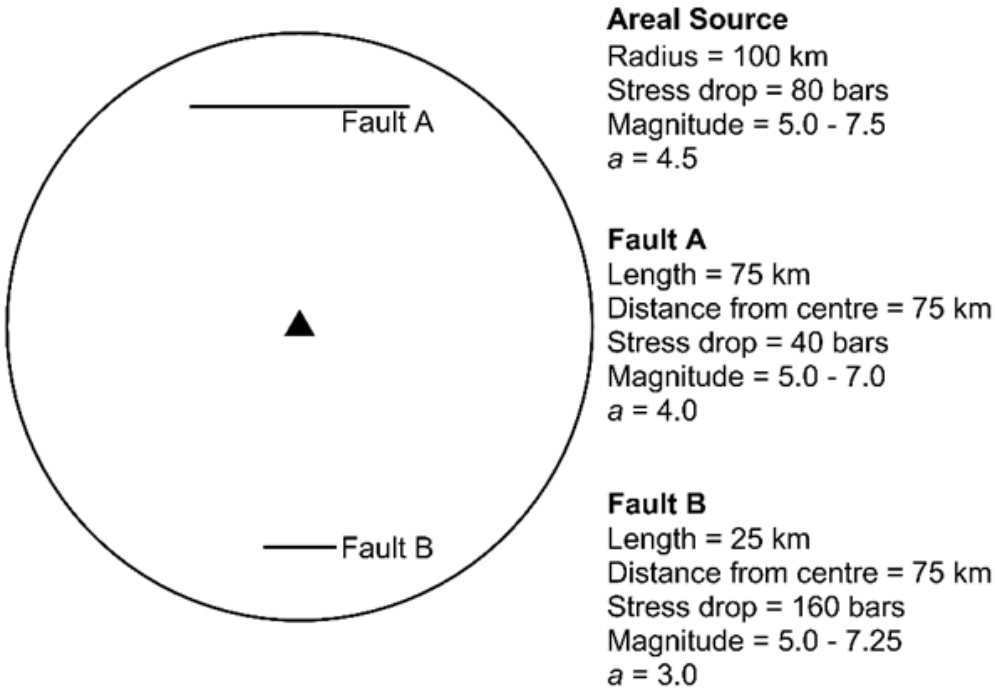


Figure 4-2: Seismic source model for the site of interest.

The non-stationary stochastic model of Sabetta et al. (2021) 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 shear-wave velocity in the upper 30m 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 (Lanzano et al., 2019).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<50km). Normal faulting is considered within the stochastic model, and a fixed site with *Vs30*=255m/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 4-2).

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 actually contribute to the hazard and risk at the site under investigation. To combat this, the importance sampling approach described in Jayaram and Baker (2010) 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 1,000 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 (Rudman, Douglas, et al., 2024).

4.3 Structural model

The structural model employed is replicated from the SAC phase 2 steel project (SAC, 1994). 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 (UBC, 1994) 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 9.15m width, with each storey being 3.96m in height.

Both plan and front elevations, as well as steel member sizes of the structure are provided in Figure 4-3. The typical floor dead load is 4.6kN/m^2, the roof dead load is 4.0kN/m^2, and the typical live load on all floors is 1.0kN/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 and Smith (1999). Further details regarding the structural model definition can be found in many articles (e.g., Barroso & Smith, 1999; Gupta & Krawinkler, 1999; Ohtori et al., 2004).

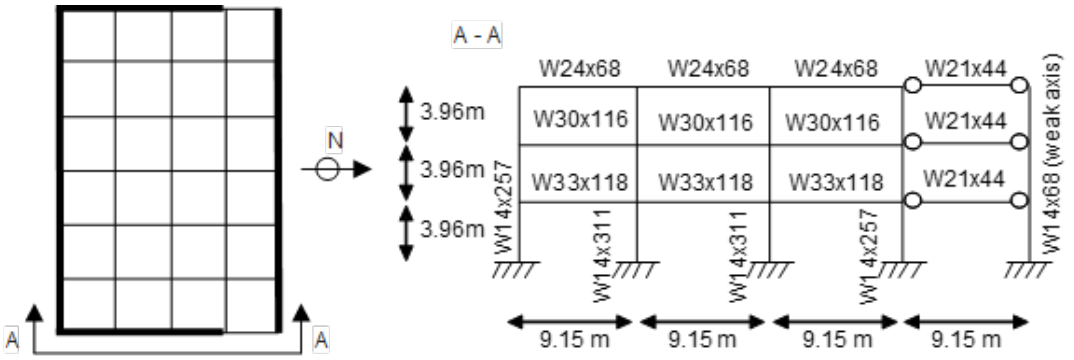


Figure 4-3: Plan and front elevations of the structural model under investigation [taken from Scozzese et al. (2020)] Thick lines highlight moment-resisting frames.

The structure is modelled in OpenSees (e.g., Mazzoni et al., 2006; McKenna, 2011), via the open-source Frame-Modeller 2D software (Elkady, 2022), 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 (The MathWorks Inc., 2023) for structural analysis.

The model's columns are represented by elastic beam-column elements with nonlinear rotational springs at their end (i.e., through a lumped plasticity approach), using a corotational transformation to account for geometric nonlinearity and assist with model convergence. These springs are assigned the modified Ibarra-Median-Krawinkler deterioration model (Ibarra et al., 2005), using the IMKbilin material in OpenSees. The parameters for the material assigned to each spring are defined by Lignos et al. (2019) who defines a set of regression equations that determine both the first-cycle envelope and monotonic backbone properties of the material depending on different steel member sizes. These material models are selected as they help ensure a good match with the benchmark models used for validation whilst capturing more realistic degradation (e.g., Barroso & Smith, 1999; Gupta & Krawinkler, 1999).

Deformation of the structural panel zones is accounted for using the parallelogram model discussed in Gupta and Krawinkler (1999), 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 available in OpenSees. The parameters for this material are defined by Skiadopoulos et al. (2021) who developed a set of panel zone stiffness and shear strength equations (based on structural analyses on different panel zone geometries and compositions) at discrete levels of panel zone shear distortion i.e., at yielding, four times yielding, and six times yielding.

Frame-Modeller 2D is able to directly handle the equations referenced above, meaning that users are only required to input minimal information such as steel member sizes and building geometry, with the material properties calculated from these. Thus, readers are also referred to Elkady (2022) for further information regarding the numerical modelling of the structure. The numerical model used in this work is publicly available via a Zenodo repository (Rudman, 2026).

The first, second and third structural periods of the model are 1.00s, 0.30s and 0.14s respectively. The static pushover curve of the structure is presented in Figure 4-4a, 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 and Smith (1999) and Gupta and Krawinkler (1999) are also plotted in Figure 4-4a for comparison. The cyclic response of the structure is presented in Figure 4-4b, which demonstrates the strength degradation of the structure under loading. As vibration periods and static analysis results are in good

agreement with the literature (e.g., Barroso & Smith, 1999; Gupta & Krawinkler, 1999; Ohtori et al., 2004), it can be assumed that the structural model is sufficiently accurate.

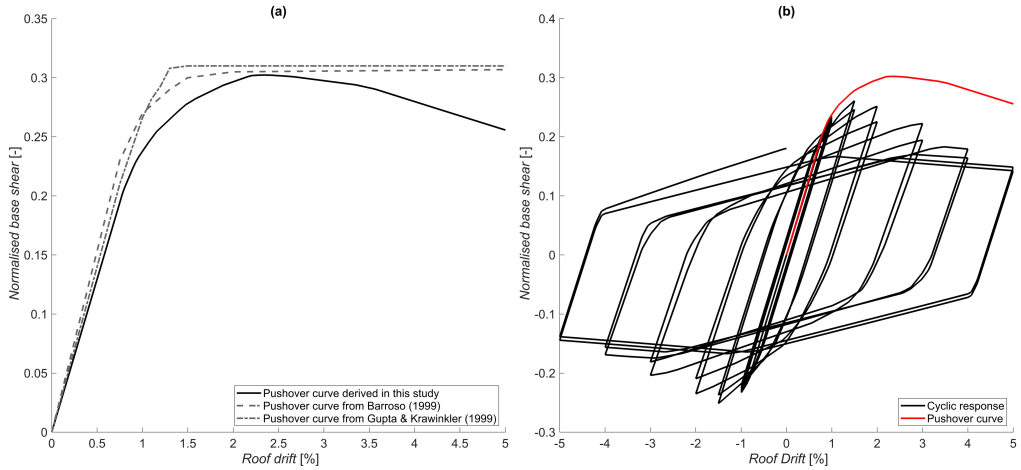


Figure 4-4: (a) Static pushover curve and (b) cyclic response, of the structure under investigation.

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 4-4. Columns within the structure are modelled using the Ibarra-Medina-Krawinkler deterioration model (Ibarra et al., 2005) and a lumped plasticity approach. Deformation of the structural panel zones is accounted for using the parallelogram model discussed in Gupta and Krawinkler (1999). Finally, a fictitious bay that represents all the pinned/simple/gravity connections in the structure is introduced, as per Gupta and Krawinkler (1999) in order to account for P-delta effects.

4.4 Intensity measure selection

In total 15 different *IMs* were considered within this study, with their names, descriptions and units provided in Table 4-1. Most of these *IMs* 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. (2009).

Table 4-1: IMs 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=1.00s)	g
Sa(avg)	Geometric mean of 5% damped spectral accelerations between periods of 0.2s and 3.0s (e.g., Vamvatsikos & Cornell, 2005)	g
ABDur	Absolute bracketed duration using acceleration threshold of 50 cms^{-2}	s
AUDur	Absolute uniform duration using acceleration threshold of 50 cms^{-2}	s
AI	Arias intensity (e.g., Baker et al., 2021)	m/s
CAV	Cumulative absolute velocity: integral of absolute velocities (e.g., Baker et al., 2021)	cm/s
ARMS, VRMS & DRMS	Root-mean-square acceleration/velocity/displacement (e.g., Kramer, 1996)	g, cm/s & cm
ASI	Acceleration spectrum intensity: integral of *Sa* between periods of 0.1s and 2.5s (e.g., Kramer, 1996)	gs
CI	Characteristic intensity (e.g., Park et al., 1985)	-
HI	Housner intensity: integral of the 5% damped pseudo-velocity spectrum between of 0.1s and 2.5s (e.g., Kramer, 1996)	cm

4.4.1 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 (2007), 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. (2004) added the concept of *hazard computability* (ease of computing a hazard curve for the given *IM*), and Padgett et al. (2008), among others, included *practicality* (correlation between *IM* and *EDP*). Padgett et al. (2008) also introduced the term *proficiency* (*efficiency* divided by *practicality*), and Tothong and Luco (2007) 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 *IMs 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* (e.g., Padgett et al., 2008). Finally, *hazard computability* is assessed here by only investigating *IMs* for which a recent ground-motion model

exists. These were identified by cross-referencing *IMs* within the literature, e.g., using the Douglas (2022) GMPE compendium.

4.4.2 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 *IMs* 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 KS test (e.g., Stephens, 1974) 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 4-1:

$$D_{max} = \max_{x}(|P_1(X) - P_2(X)|) \tag{4-1}$$

where, $P_1(X)$ and $P_2(X)$ are both CDFs at a value, 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 MAF of exceedance. To overcome this, a MAF of exceedance is converted to an annual probability of exceedance using Equation 3-3, assuming that the exceedance events are a Poisson process. 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%).

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 4.6.2.

An alternative approach would have been to use the relative entropy method (e.g., Jalayer et al., 2012; Tsioulou & Galasso, 2018) 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 4-1), the outlined procedure can be performed at any stage of a risk assessment. This allows for more direct comparison between the suitability of *IMs* for the specific use under question.

4.5 Risk assessment procedure

Firstly, the evaluation of the demand hazard via the unconditional approach is described in this section. Secondly, 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 *IMs* and each of the demand models are compared against the benchmark estimates.

4.5.1 Benchmark unconditional procedure

The benchmark demand hazard (referred to from now on as the *unconditional* estimate) can be obtained using Equation 4-2:

$$v(edp) = \sum_{j=1}^{sources} \lambda_{source_j} \sum_{i=1}^{N_j} \frac{I_{i,j}(edp)}{N_j} \frac{ISW_j}{\sum ISW} \tag{4-2}$$

where $v(edp)$ is the MAF 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.

4.5.2 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 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*. While the top-storey *IDR* may be

expected to exhibit less nonlinearity than the ground storey, the structure's response profile indicates that it consistently produces the largest drifts within the structure. Therefore, this *EDP* was selected, as this study is primarily focused on peak structural demands and exploring the full range of response. However, it is worth noting that the following results are not strongly dependent on the choice of *EDP*.

The structural response is then analysed using cloud analysis (e.g., Mackie & Stojadinovic, 2005). 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. (2016); and two simple machine learning based regression models: a RF regression and an ANN. The functional form of the linear and bilinear fits are provided in Equations 4-3 and 4-4(4-4 respectively:

$$\ln(EDP) = a_1 + b_1 \ln(IM) + \sigma \tag{4-3}$$

$$\ln(EDP) = (a_1 + b_1 \ln(IM))H_1 + [a_1 + (b_1 - b_2)\ln IM^* + b_2 \ln IM](1 - H_1) + \sigma \tag{4-4}$$

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. It would be expected that considering a different *EDP* such as ground-storey *IDR*, which tends to exhibit a greater nonlinearity in response, may see a greater use for heteroscedastic models, and further research could investigate this. Instead, for this study, 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 (e.g., Li et al., 2022; Soleimani & Liu, 2022). 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 *IMs* (*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 *IMs* are generated by simply replacing the $\ln(IM)$ terms in Equations 4-3 and 4-4 with IM, for example Equation 4-4 becomes:

$$\ln(EDP) = (a_1 + b_1 IM)H_1 + [a_1 + (b_1 - b_2)IM^* + b_2 IM](1 - H_1) + \sigma \quad (4\text{-}5)$$

Each cloud analysis model is shown in Figure 4-5 for one set of simulations with *PGA* (Figure 4-5a) and *PGD* (Figure 4-5b) as the *IM*, and top-storey *IDR* as the *EDP.* In Figure 4-5a*,* the linear cloud analysis model appears to fit the data well until 0.2g *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 towards 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 4-5b. 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 in response than is visible in Figure 4-5a. 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 in 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 4-6:

$$P\left(EDP > edp \middle| IM_j\right) = \sum_{i=1}^{N_j} \frac{I_{i,j}}{N_j} \tag{4-6}$$

where $P\left(EDP > edp \middle| IM_j\right)$ 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.

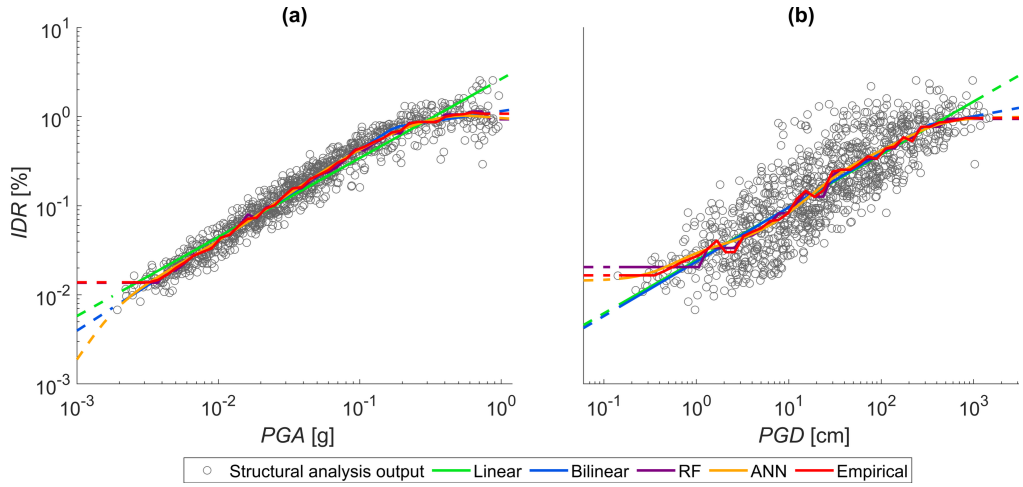


Figure 4-5: Demand models for (a) peak ground acceleration (b) peak ground displacement, the engineering demand parameter is top-storey interstorey drift ratio.

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 on Figure 4-5 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 4-6 plots fragility curves for the same set of samples, with *PGA* (Figure 4-6a) and *PGD* (Figure 4-6b) 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 4-6. 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 4-5). 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 4-5), and so the CCDF value will not always increase.

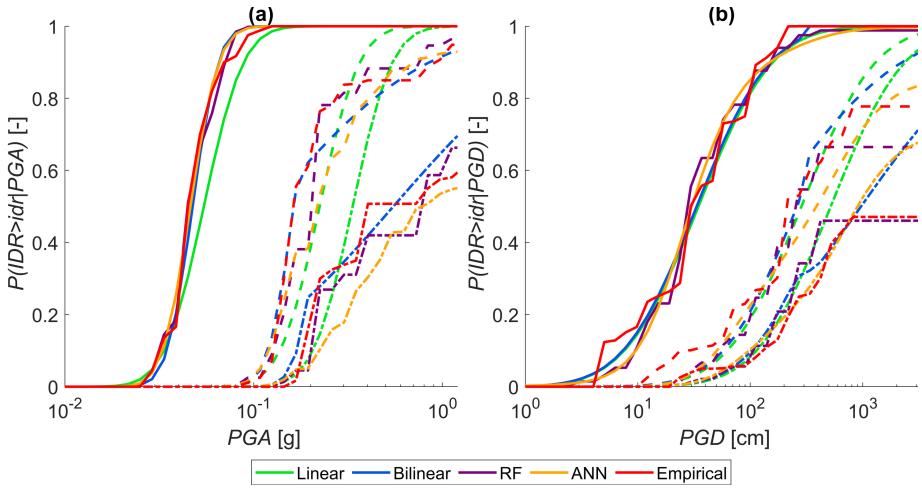


Figure 4-6: P(EDP|IM) curves for (a) peak ground acceleration and (b) peak ground displacement, and each of the seismic demand models, at interstorey drift ratio thresholds of 0.2% (solid lines), 0.7% (dashed lines) and 1.0% (dot-dash lines).

4.5.3 Comparing mean risk estimates

Demand hazard is estimated through this procedure for all *IMs*, all seismic demand models, and all 100 sets of samples, allowing the MAF of top-storey *IDR* exceedance to be calculated. Figure 4-7 presents the mean of this demand hazard for each of the *IMs* and demand models under investigation, comparing each against the unconditional estimates. Most *IMs*, 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 structural component's yielding (see pushover curve in Figure 4-4) and so the *IM* choice and demand model becomes more important here.

The linear model overestimates the demand hazard for all *IMs,* 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 *IMs*. Both machine learning

models (the RF and ANN) produce almost identical demand hazard estimates for all *IMs* 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 *IMs* 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 4-7 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.

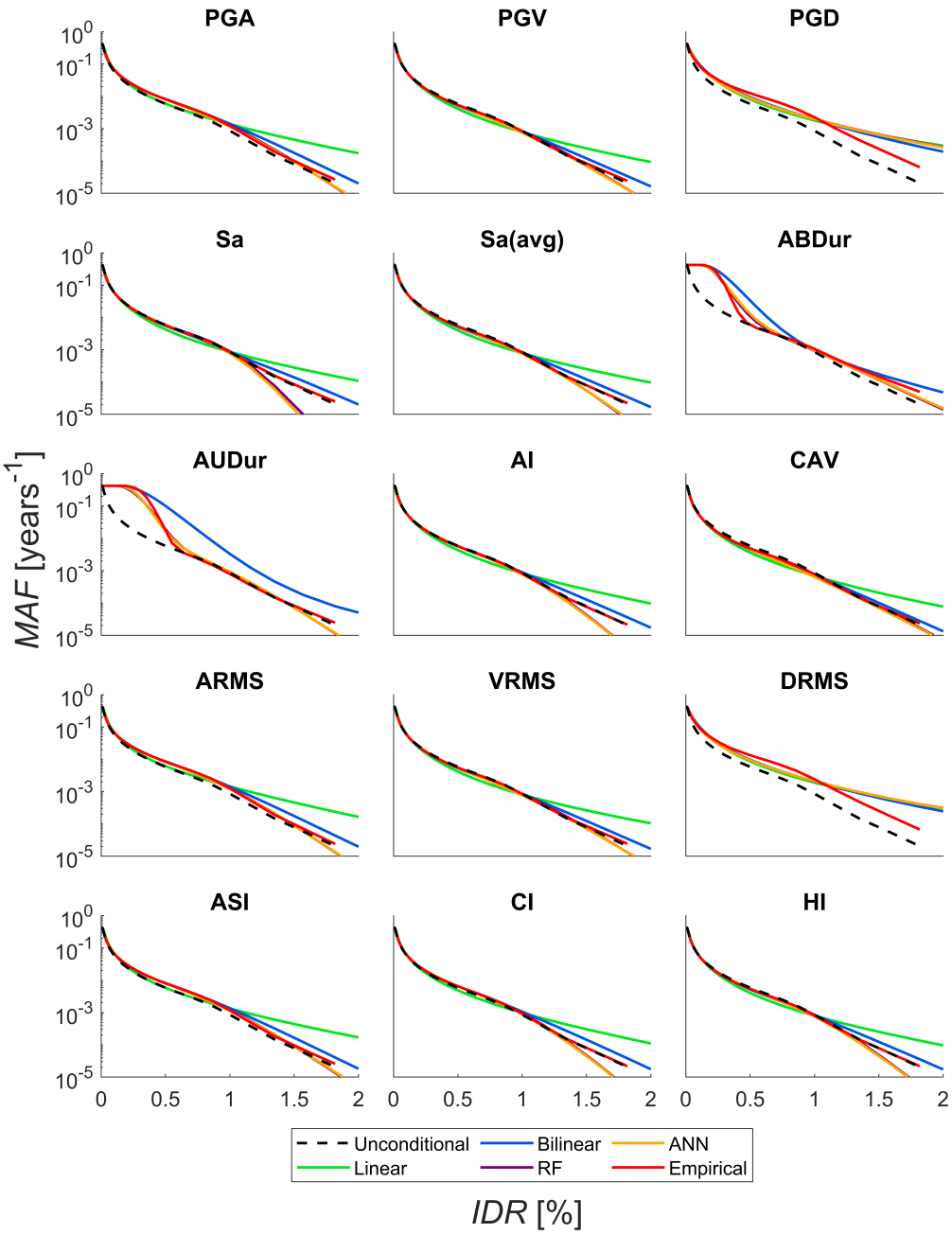


Figure 4-7: Mean demand hazard estimates for all intensity measures and seismic demand models under investigation

4.6 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.

4.6.1 Traditional IM performance metrics

The *efficiency*, *magnitude-sufficiency*, *distance-sufficiency*, and *practicality* of all *IMs* using the linear demand model are presented in Figure 4-8. These measures were assessed through the procedures described in Section 4.4 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 *IMs* considered. A visual comparison of these measurements suggests that *Sa* performs the best out of all considered *IMs*. This is unsurprising as the structure's response is dominated by its first mode. Figure 4-8 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 thesis, "most sufficient" means the one with the highest p-value, i.e. the one that is furthest from failing the significance test based on a p-value threshold.

Figure 4-8d shows *CAV* to clearly be the most *practical*. Results appear to suggest that *efficiency* and *practicality* are highly correlated. Figure 4-8b shows that most *IMs* satisfy the *magnitude sufficiency* threshold of 0.05, however no *IM* is clearly the most *magnitude sufficient*. On the other hand, Figure 4-8c indicates that *distance sufficiency* is a harder requirement to achieve with 7 of the 15 *IMs* 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.*

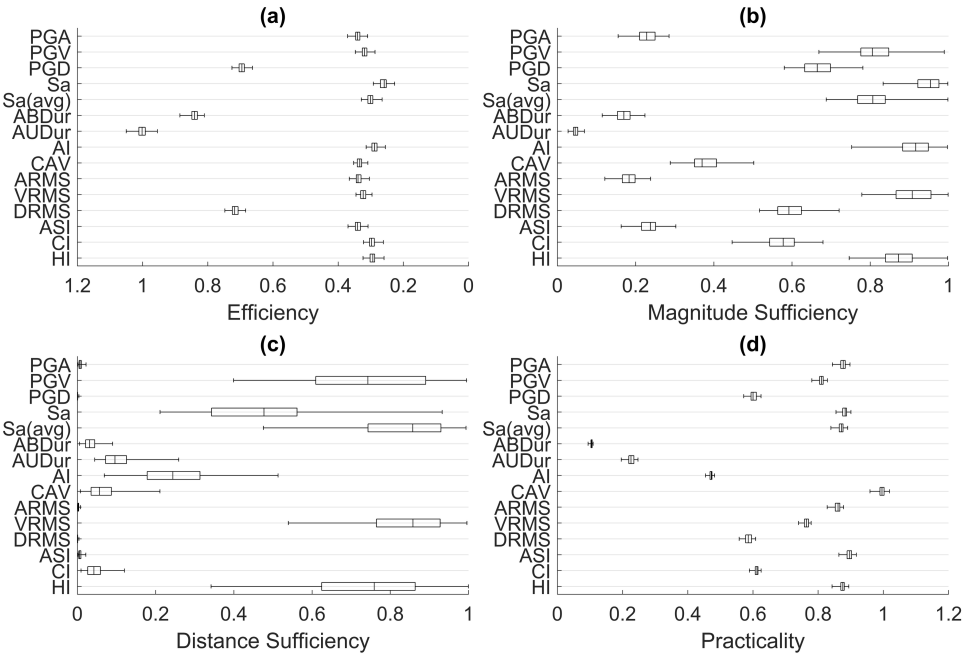


Figure 4-8: Boxplot showing the (a) efficiency (b) magnitude sufficiency (c) distance sufficiency (where sufficiency is the p-value of the slope of the residuals from the P(EDP|IM) model against magnitude and distance respectively) and (d) practicality of each intensity measure when considering a linear demand model for the top-storey interstorey drift ratio. Note the reversed x axis in (a) as a lower value of efficiency is best (smallest standard deviation).

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 *IMs* 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 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 *IMs* for this metric. As such, *practicality* is only evaluated for the linear model and ignored for all others.

Figure 4-9 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 IMs 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.

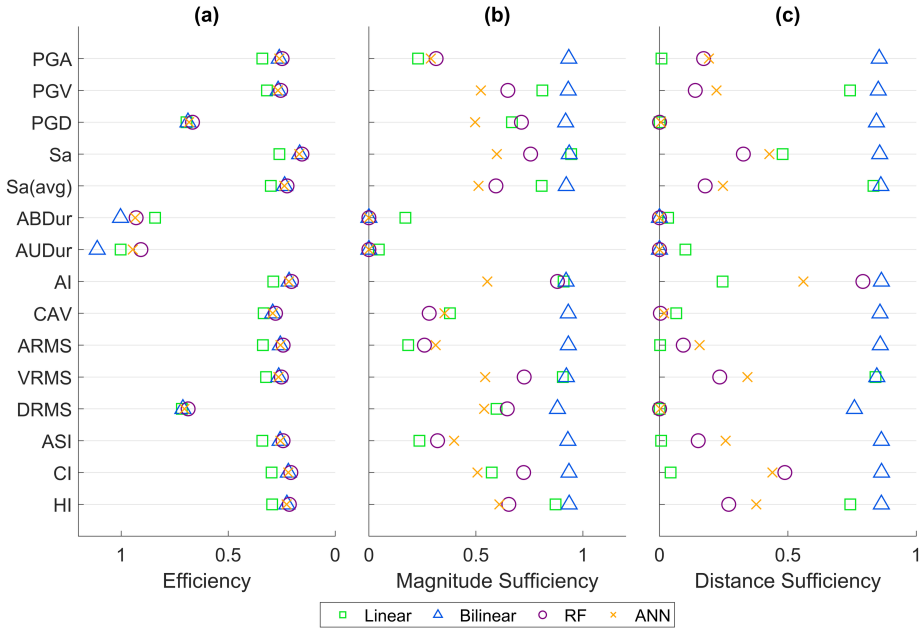


Figure 4-9: Plot showing how the mean of (a) efficiency (b) magnitude sufficiency and (c) distance sufficiency changes for each intensity measure and seismic demand model combination. Note the reversed x axis in (a) as a lower value of efficiency is best (smallest standard deviation).

4.6.2 Proposed performance metric

Figure 4-10 presents the boxplot of the D_{max} values (i.e. the difference to the unconditional results) calculated according to Equation 4-1 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 *IMs* the empirical demand model produces the most similar estimates to the unconditional approach.

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 *IMs* perform considerably worse than all other *IMs*. 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 *IMs*, *Sa* and *Sa(avg),* are generally the most used *IMs*. 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.

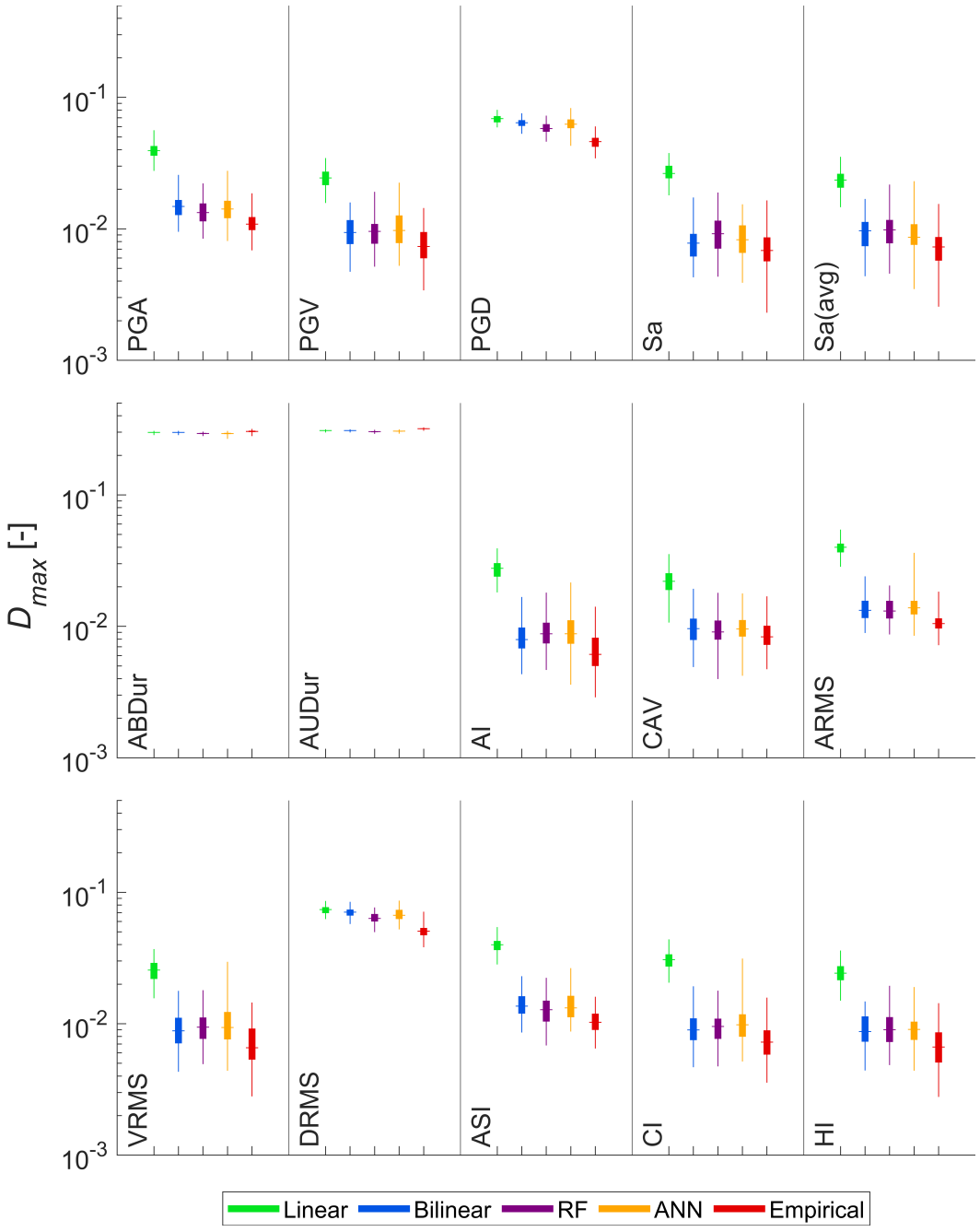


Figure 4-10: Boxplot showing D_{max} for each intensity measure and each seismic demand model.

4.6.3 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 *IMs* (i.e., *ABDur*, *AUDur*, *PGD*, and *DRMS*) are removed from this correlation calculation, the R^2 values change considerably. *Efficiency* and *practicality* fall to 0.260 and 0.003 respectively, whilst the *magnitude-sufficiency* and *distance-sufficiency* rise to 0.602 and 0.492 respectively. This indicates that when considering only accurate *IMs*, *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 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.

4.7 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 4-11 shows the sensitivity analysis results for (a) *Sa* and (b) *AI* for all five demand models. The sensitivity of estimates for these *IMs* is considered representative of all other *IMs*, except the displacement (*PGD* and *DRMS*) and duration (*ABDur* and *AUDur)* based *IMs* - where risk estimates remain inaccurate irrespective of the number of samples.

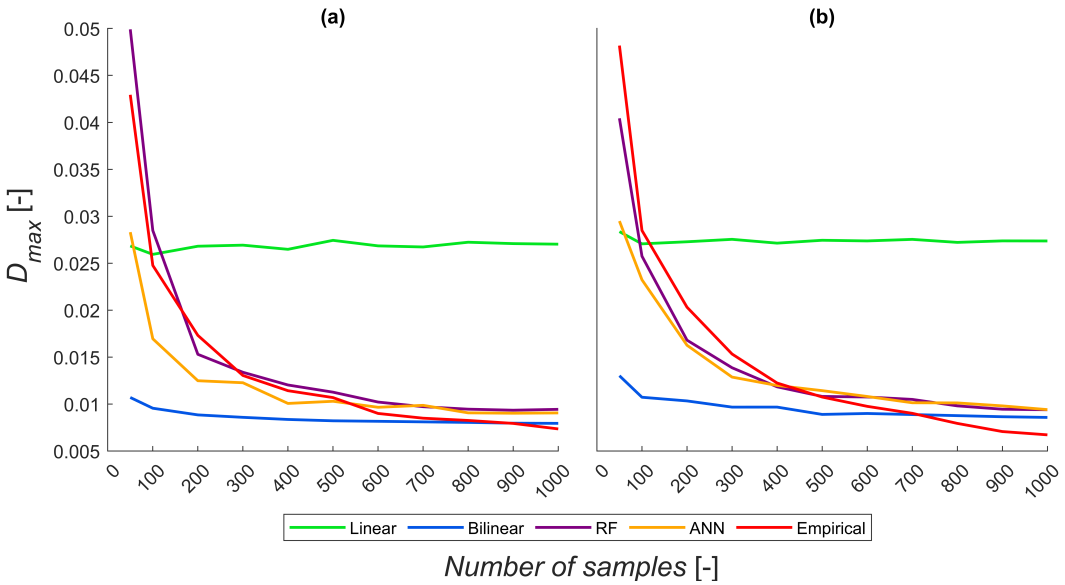


Figure 4-11: Sensitivity of D_{max} to the number of samples used in the risk assessment procedure. The intensity measures (IMs) used are (a)Sa and (b)AI, which are considered representative of all accurate IMs.

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 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 4-10. 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.

4.8 Component damage analysis

To fully demonstrate the impact that both *IM* and demand model selection has 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 (FEMA, 2012) 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 chapter: 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 4-12. Each *DS* is sequential, meaning that each *DS* can only occur if the preceding one has already occurred.

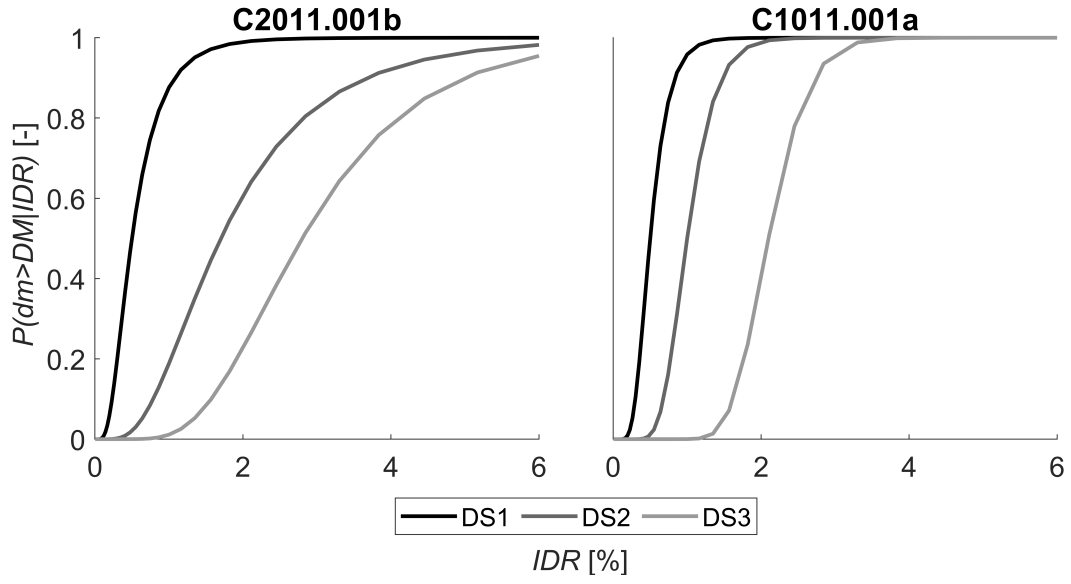


Figure 4-12: Fragility curves for both components under investigation

Fragility curves are convolved with the demand hazard estimates from Section 4.5.3 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 4-13 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 *IMs* have been plotted on Figure 4-13, with these four *IMs* 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 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 *IMs* (*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.

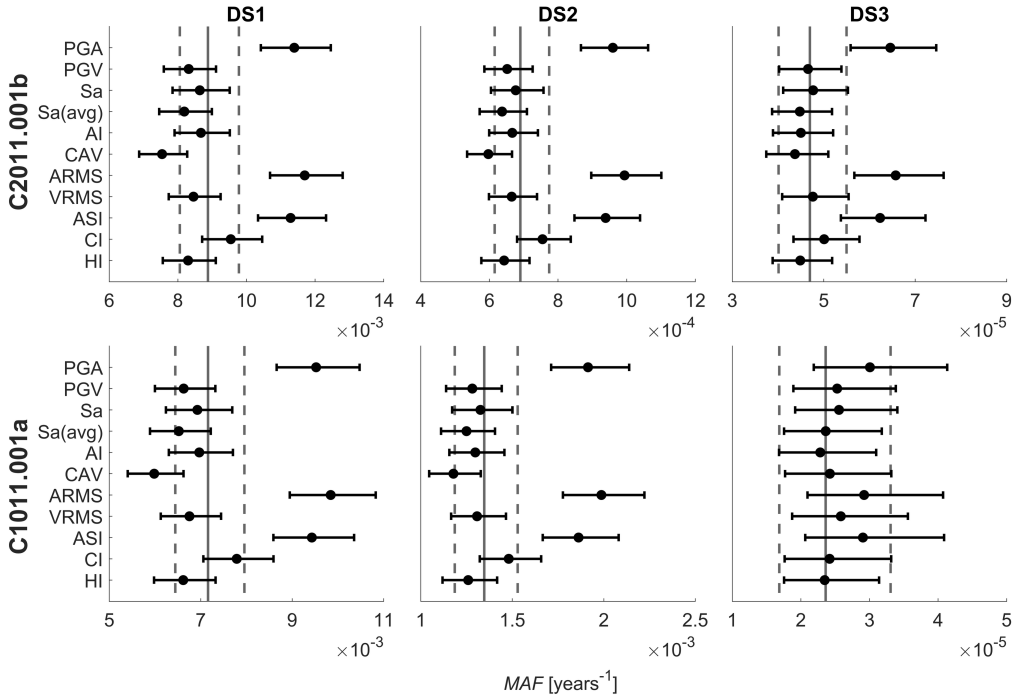


Figure 4-13: Mean annual frequency of exceedance of each damage state for both components using the empirical demand model, with 16th and 84th percentiles plotted as error bars. The unconditional estimate is marked by the solid (mean) and dashed (16th and 84th percentile) lines.

4.9 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 hazard estimates.
- 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 i.e., 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 study 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.

It is noteworthy that the bilinear model is a rational choice for representing the variation of the seismic demand with the intensity measure when considering structures whose response is dominated by the first mode of vibration and exhibits a clear transition between a linear and nonlinear phase (as could be inferred from the capacity curve plot of Figure 4-4). In these cases, the intensity measure value representing this transition is the intensity that corresponds to the yielding of the system (e.g., Freddi et al., 2017; Tubaldi et al., 2016).

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 study provides an objective way of choosing the most appropriate combination of intensity measure and seismic demand model for any seismic risk and loss assessment.

5 Covariance structure modelling of engineering demand parameters in Cloud-based seismic analysis

This Chapter has been adapted from:

Rudman, A., Tubaldi, E., Gentile, R., Douglas, J. 1 Covariance Structure Modelling of Engineering Demand Parameters in Cloud-Based Seismic Analysis. *Earthquake Engng Struct Dyn*. [Submitted]

Probabilistic seismic demand modelling aims to estimate structural demand as a function of ground motion intensity- a critical stage in seismic risk assessment. Although many models exist to describe the structural demand, few consider the covariance among engineering demand parameters (EDPs), potentially overlooking a key factor in improving the accuracy of these models. This chapter aims to investigate the impact of heteroscedastic covariance models on seismic demand hazard and loss estimates, using an illustrative example of a mid-rise steel structure, a hypothetical seismic source model, and a non-stationary stochastic ground-motion model. Cloud analysis is performed to establish median demand estimates using a linear model, a bilinear model, and a Gaussian process regression (GPR) model. Various seismic demand and loss estimates are produced from combinations of four variance and four correlation models, which consider both homoscedasticity and heteroscedasticity. Heteroscedastic models include a step-model, a linear regression-based model, and a model using another GPR. Earth mover's distance is used as a metric to assess the accuracy of each estimate against a benchmark solution obtained via an unconditional approach (i.e., empirically estimating demand exceedance frequencies from simulated ground motion time-histories). The study shows that using more complex, heteroscedastic variance models improves risk and loss estimate accuracy, with the GPR-based variance model proving the most accurate. Although the method used to account for correlation has a smaller impact on model accuracy, considering correlation is also important for covariance formulation, as models that ignore correlation yield the least accurate seismic demand and loss estimates.

5.1 Introduction

The creation of an accurate PSDM is a key step in seismic risk assessment (e.g., Gentile & Galasso, 2020; Scozzese et al., 2020; Tubaldi et al., 2016). PSDMs establish a statistical relationship between one or more ground motion intensity measures (*IM*s) and engineering demand parameters (*EDP*s), such as drift, acceleration, and displacement, within the framework of PBEE. Several methods have been established to create PSDMs, most notably IDA, MSA, and Cloud analysis (e.g., Bradley, 2013). IDA involves scaling a set of ground motions to varying *IM* levels and plotting the structural demand until collapse (Vamvatsikos & Cornell, 2002). MSA differs from IDA in that multiple sets of ground motions are used for each *IM* level to reduce the use of scaling (e.g., Scozzese et al., 2020). Finally, Cloud analysis utilises a large set of natural, unscaled ground motions to fit a regression between *IM* and *EDP*, thus eliminating the need for record scaling (e.g., Mackie & Stojadinovic, 2005).

The simplicity of the Cloud-based approach makes it attractive in risk assessment practice, as it is easy to establish a demand model without having to scale ground motions. This is preferable as scaling ground motions can distort the characteristics of natural ground motions and introduce bias into risk estimates, if an inappropriate ground motion selection method is chosen (e.g., Whittaker et al., 2011). Typically, Cloud analysis models assume a linear relationship between the logarithms of *IM* and *EDP* (e.g., Cornell et al., 2002). However, recent research has shown that a bilinear fit could improve the accuracy of Cloud analysis PSDMs (e.g., Freddi et al., 2017; O'Reilly & Monteiro, 2019; Rudman, Tubaldi, et al., 2024; Tubaldi et al., 2016). The benefits of the bilinear demand model are evident as the second branch of the model can account for the change in demand once the structure enters the nonlinear demand

range, providing a physical meaning behind the mathematical function (e.g., Tubaldi et al., 2016). There is also evidence that more complex fitting techniques, such as GPR, could further enhance Cloud analysis models (Rudman, Tubaldi, et al., 2024). One downside of the Cloud-based approach is that it requires many ground motions to be accurate (e.g., Gehl et al., 2015; Jiang et al., 2023). However, the use of simulated ground motions can overcome this issue by providing an efficient way to establish sufficient samples to explore the full capabilities of a Cloud analysis model (e.g., Rudman, Douglas, et al., 2024; Rudman, Tubaldi, et al., 2024).

In an ideal world, sufficient ground motions and computational power would exist that a PSDM would not be needed at all when assessing seismic risk. Instead, a structure's demand could be fully characterised using an unconditional method. This is where direct observations of a system's demand to ground motions are used to estimate the rate of exceedance of some threshold – again meaning that record scaling is avoided. One drawback of the unconditional approach is that it requires a sufficiently large database of appropriate ground motions to be considered accurate. Nevertheless, Rudman, Douglas, et al. (2024) have demonstrated that unconditional methods can already provide accurate estimates using existing ground motion databases through their observation-based hazard assessment approach. Moreover, simulated ground motions can compensate for a lack of available data, facilitating the unconditional method and creating a benchmark to validate probabilistic "conditional" approaches (i.e., estimating risk through a series of intermediate conditional probabilities, as is done in PBEE-based risk assessments) (e.g., Rudman, Douglas, et al., 2024; Rudman, Tubaldi, et al., 2024; Scozzese et al., 2020).

Whilst few studies have considered different methods of modelling the variance within a PSDM (e.g., Bradley & Lee, 2010; Rossetto et al., 2016), to improve accuracy, the additional consideration of covariance among *EDP*s is a less studied topic, likely due to the perceived increase in complexity and computational demand of correlation modelling (e.g., Bradley & Lee, 2010). Covariance amongst *EDP*s can be accounted for by combining the variance of the individual *EDP*s with the correlation structure of the *EDP*s to produce the covariance structure, which can then be used within a joint probability distribution (e.g., Bradley & Lee, 2010). The interactions among *EDP*s can vary considerably at different seismic intensities and could have a significant impact on the accuracy of PSDMs, and consequently, the accuracy of seismic demand and loss estimates (e.g., Bradley & Lee, 2010; Du & Padgett, 2020). Thus, better accounting for the covariance structure of *EDP*s is important, as it enables more accurate quantification of the probability of damage and associated losses (e.g., Cremen & Baker, 2018).

Table 5-1 summarises previous research efforts that have explicitly accounted for the correlation amongst *EDP*s as part of the risk assessment process. Out of the 17 studies considered, seven use a heteroscedastic covariance term (i.e., the covariance changes in response to an explanatory variable, usually the *IM*). Interestingly, Kang et al. (2023) considered a heteroscedastic covariance term using structural characteristics, such as the first mode period and damping ratio, as predictors instead of the *IM*. Nine studies incorporated a homoscedastic correlation term into their demand model – the covariance is constant. Jayaram et al. (2012) reported that the correlation between *EDP*s with respect to *IM* level was not significant enough to justify using a heteroscedastic correlation matrix term for the six tall buildings considered within their

study (four steel moment resisting frames and two concrete frames). However, the authors state that this could be due to the small number of ground motions used within the study. Only one study by Du and Padgett (2020) compared the impact of considering both homoscedastic and heteroscedastic covariance models on a 3-span concrete girder bridge. The study concluded that their homoscedastic covariance model was optimal for the site under investigation, as fragility estimates from the model were comparable to those from a heteroscedastic model and were far less computationally expensive. The choice between homoscedastic and heteroscedastic residual analysis is especially relevant for bilinear Cloud analysis models, as combining residuals from both segments into one could be considered an oversimplification. This idea is explored by Tubaldi et al. (2016) who considered a bilinear variance term alongside their bilinear Cloud analysis model.

Seven of the 16 studies created their PSDM through IDA or MSA. Many of these studies are reliant on record scaling to derive demand models and covariance matrices. A Cloud analysis-based approach to demand modelling would avoid the issue of record scaling. However, PSDMs based on IDA and MSA are not necessarily applicable to a Cloud analysis approach. There is limited research on the development of correlation models for direct use within Cloud analysis-based demand models. Only the studies by Minas et al. (2018) and Du and Padgett (2020) develop a heteroscedastic correlation model using a Cloud analysis-based approach. The latter opt to recommend a homoscedastic model over their heteroscedastic one, whereas the former recommends their heteroscedastic approach.

Three studies employed copulas to construct the joint probability distribution of the correlated *EDP*s (Goda & Tesfamariam, 2015; Qian & Dong, 2022; Zhou & Li, 2019)

at various fixed intensities. Copulas are not adopted in this study, since modelling a multivariate normal distribution represents a direct extension of current risk assessment practice, and because the assumption of a lognormal relationship between seismic intensity and structural demand is already well established in PBEE (e.g., Baker, 2015; Ge et al., 2021). However, it is worth highlighting that there has been little to no comparison on the performance of copula-based approaches to demand modelling against alternative methods (e.g., multi-output GPR).

Finally, only two studies attempted to validate their PSDM that incorporates correlation. Lupoi et al. (2006) validated their *EDP*-correlated fragility functions by comparing the ultimate limit state fragility function against empirical fragilities at four *IM* values, which are calculated by performing a Monte Carlo procedure on several nonlinear time history analyses at each *IM* level. Yi and Taflanidis (2025) validated their stochastic emulation model by comparing response distribution estimates against a reference sample that follows a lognormal distribution. However, to the authors' knowledge, no other study has attempted to compare their correlation model against an unconditional (i.e., Monte Carlo-based) risk benchmark.

Table 5-1: Summary of studies incorporating EDP correlations into their risk estimates. L = loss estimates, F = fragility curves, R = structural demand estimation, Dam = damage estimates, Dem = demand hazard estimates

Study	Heteroscedastic correlation?	Structural model	Analysis output	Suitable for Cloud analysis?
Aslani and Miranda (2005)	✓	7-storey RC frame (hotel)	L	✗
Nielson (2005)	✗	Portfolio of multi-span bridges	F	✓
Luco and Cornell (2007)	✗	3, 9 & 20-storey steel MRF	R	✓
Lupoi et al. (2006)	✓	11-span continuous beam bridge & 3-storey RC frame	F	✗
Bradley and Lee (2010)	✓	10-storey RC frame (office)	L	✗
Jayaram et al. (2012)	✗	6 x tall buildings (20, 40, and 42 storey)	L	✓
Ghosh et al. (2014)	✗	City-scale bridge/transportation network	F	✓
Goda and Tesfamariam (2015)	✗	4-storey non-ductile RC frame	L	✗
Wang et al. (2018)	✗	Multi-span continuous RC bridge	F	✓

Minas et al. (2018)	✓	2 x 4-storey RC MRFs	F	✓
Zhou and Li (2019)	✗	two-span RC continuous girder-box bridge	F	✓
Du and Padgett (2020)	✓	3-span simply supported concrete girder bridge	F	✓
Bandini et al. (2022)	✓	Continuous concrete girder bridge	Dam	✗
Qian and Dong (2022)	✗	Portfolio of multi-span RC bridges	L	✓
Kang et al. (2023)	✓	Regional assessment of a virtual city	L	✗
Zhong et al. (2023)	✗	3 x MDOFs	Dem	✗
Yi and Taflanidis (2025)	✓	3-storey RC MRF	L	✗

Therefore, this study seeks to fill these knowledge gaps by pursuing the following three aims:

1. To evaluate the impact of modelling covariance among EDPs within Cloud analysis-based PSDMs on seismic demand and loss estimates.
2. To compare the influence of homoscedastic and heteroscedastic assumptions on Cloud analysis-based PSDMs and their subsequent effect on seismic risk estimates.
3. To develop and demonstrate an approach for validating conditional risk models that account for covariance among EDPs, using an unconditional benchmark.

To meet these objectives, both homoscedastic and heteroscedastic approaches are considered, using the interstorey drift ratio (*IDR*) and peak floor acceleration (*PFA*) as the *EDP*s, which are calculated from structural analyses on a benchmark building widely employed in PBEE studies. This research is done in the context of a hypothetical site that utilises simulated ground motions from a non-stationary stochastic ground motion model. This enables the establishment of a benchmark risk estimates created via the unconditional approach, allowing validation of the developed conditional risk models.

Within this chapter, Section 5.2 describes the methodology of developing covariance and risk models that consider the relationship among *EDP*s. The hypothetical SSM and the structural model employed as a case study are presented in Section 5.3, with results presented and analysed in Section 5.4. Finally, a sensitivity analysis on the number of samples used to create different covariance models is carried out in Section 5.5 to explore how these models may perform in the face of the limited availability of

ground motion records for Cloud analysis. The study findings highlight both the individual importance of accurately modelling median demand, variance, and correlation, as well as the interplay between each model in creating the most precise risk estimates possible.

5.2 Methodology

Figure 5-1 visualises the workflow to obtain demand hazard estimates from both unconditional and conditional approaches, as well as demonstrating how the demand hazard estimates can be extended to evaluate the structural losses on a multi-storey structure, which is the focus of this study.

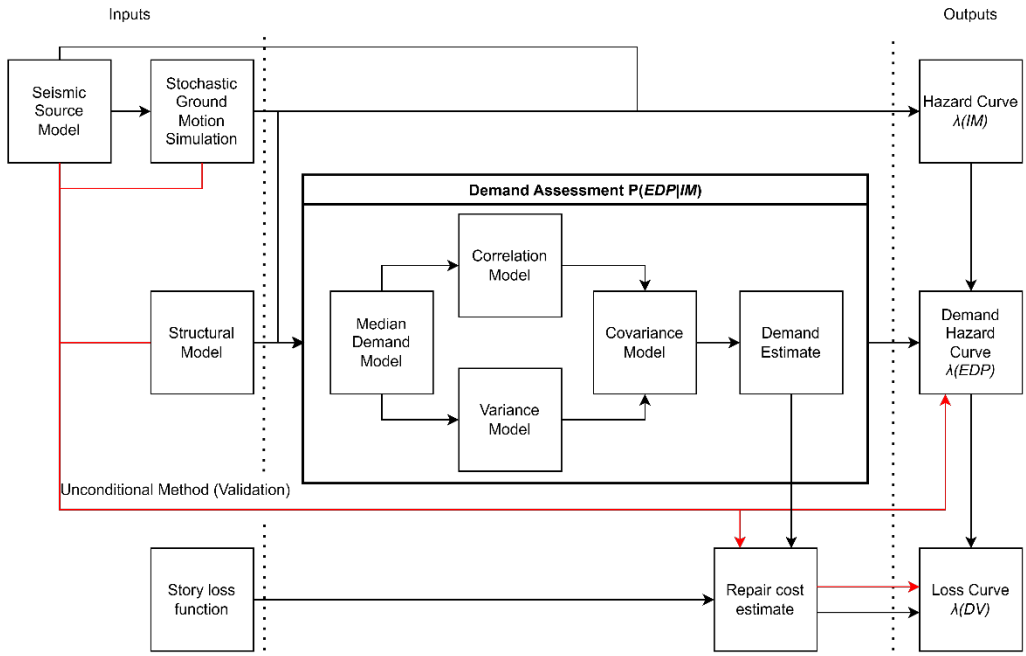


Figure 5-1: Methodology workflow to estimate demand hazard and losses via unconditional (for validation, red) and conditional (black) approaches.

The unconditional approach to estimating demand hazard, demonstrated by the red workflow in Figure 5-1, can be mathematically represented, alongside an importance sampling procedure, by Equation 4-2 (e.g., Bradley et al., 2015; Jayaram & Baker,

2010). As indicated in Figure 5-1, in this study, the unconditional method can only be used to calculate the demand hazard on the structure; as such, it is treated in the same manner as the conditional approach when extended to estimate losses to the structure. Although it is an alternative to the conditional approach, in this study the unconditional approach is used to create a benchmark risk estimate through direct observation of the case study structure's demand, which can act to validate the calibrated conditional risk models.

Conditional approaches to estimating demand hazard and losses first require the development of seismic hazard curves, which describe the MAF of exceeding an *IM*. These are created in this study through Monte Carlo hazard assessment (e.g., Musson, 2000) with importance sampling. Next, a demand assessment is performed using a Cloud analysis approach (e.g., Bradley, 2013), where in this study, a variety of PSDMs are created that consider different formulations of the covariance structure. This demand assessment is then used to estimate demand and losses to the structure. In this study, losses are estimated in the form of direct repair costs, using storey loss functions (e.g., Papadopoulos et al., 2019; Ruiz-García & Miranda, 2010). This forms a simple extension of the demand hazard estimation workflow, allowing the full impact of considering different covariance models for demand estimation to be seen.

The following sections outline the methodology of creating these covariance models (Section 5.2.1), after which the process of estimating demand hazard and loss estimates is presented in Section 5.2.2. Finally, a procedure to compare conditional risk estimates against unconditional benchmark estimates is presented in Section 5.2.3.

5.2.1 Demand assessment

This section presents the process of creating demand models within a Cloud analysis framework that accounts for the covariance amongst *EDP*s typically used to compute losses for a building structure. First, linear, bilinear, and GPR models are created to estimate the median demand of each *EDP*. Next, the variance and correlation models for each of the *EDP*s are established and then combined to estimate covariance.

5.2.1.1 Median demand model

Linear, bilinear, and GPR median demand models are created for the *IDR* and *PFA* at each level of the structure using Cloud analysis, from the nonlinear time history analyses output. These *EDP*s were selected as they can be considered sufficient for representing the damage to structural and non-structural components within the building, respectively (e.g., Miranda & Aslani, 2003; Taghavi, 2003). It is important to note that this process can be easily expanded to include other *EDP*s (e.g., peak floor velocity and residual drift) to shed light on other forms of damage and losses. The functional form of the linear model is presented in Equation 5-1. The bilinear models follow the functional form of Tubaldi et al. (2016), and are provided in Equation 5-2:

$$ln(EDP) = a_1 + b_1\, ln(IM) \tag{5-1}$$

$$ln(EDP) = (a_1 + b_1\, ln(IM))H_1 + [a_1 + (b_1 - b_2)\, ln\, IM^* + b_2\, ln\, IM](1 - H_1) \tag{5-2}$$

where a_1, b_1, a_2, b_2 and H_1 are coefficients fit by the regression model that control the intercept (a) and slope (b) of the first and second segments, respectively, whilst H_1 is a step function to determine which branch of the bilinear model *IM* lies in. IM^* identifies the breakpoint between the two branches, corresponding to the yielding of the structure and evaluated as the yield strength of the structure normalised by the

structure's weight and assessed to be 0.3g from pushover analysis (e.g., O'Reilly & Monteiro, 2019), or fitted through regression analysis (e.g., Tubaldi et al., 2016). As these equations are for the median of the demand, model error is not considered here.

The GPR modelling of the median demand uses a squared exponential kernel function; all other hyperparameters within the GPR are optimised using a Bayesian optimisation algorithm (e.g., Snoek et al., 2012). The squared exponential kernel function - a popular choice for GPR (e.g., Gentile & Galasso, 2022) - is selected since it reflects the "stability" of the involved physical quantities (i.e., a small perturbation of the input parameters produces a slight change in the output parameters). For simplicity, the kernel length scale is also optimised, meaning the model for each *EDP* has a separate length scale value.

While other machine learning methods (e.g., artificial neural networks and random forest regression) could be investigated within the context of this study, GPR was selected due to its recognised stronger performance on small datasets, reduced reliance on heuristic hyperparameter tuning, and smoother response output (Rasmussen & Williams, 2005). Moreover, Rudman, Tubaldi, et al. (2024) showed ANNs and random forests (based on a linear PSDM approach) to be less effective than a bilinear approach to modelling demand. As such other machine learning techniques are not considered within this study.

5.2.1.2 Covariance model

Covariance models are created through residual analysis of each demand model, combining separate models for the variance and correlation of residuals into one structure. This study considers variance and correlation modelling separately, rather

than directly estimating the covariance, as it provides greater flexibility to model the variance and correlation components separately. In fact, previous research has already demonstrated the benefits of a heteroscedastic variance term (e.g., Du & Padgett, 2020); however, a heteroscedastic correlation term may not be preferable (given the expected low impact of the correlation model on results), so this approach allows every combination of variance and correlation model to be tested together to find the most accurate covariance model.

Variance models are first created through residual analysis of each median demand model. The *homoscedastic variance* models are made by finding the total standard deviation of residuals from each demand model. A *step-model* of the variance, consisting of a two-segment homoscedastic model where the segment break is determined by IM^* is also developed by calculating the standard deviation of residuals for each segment of the bilinear demand model.

Heteroscedastic variance models are created by dividing all residuals into 25 equally sized bins according to their *IM* and calculating the standard deviation of residuals in each of these bins. The number of bins is selected as it finds a good balance between practicability and accuracy when creating the risk assessment. Two heteroscedastic models for describing the variance are created through regression, with *IM* as the predictor and the standard deviation of residuals as the response variable. The first model is a *linear regression*, with functional form and fitted coefficients presented by Equation 5-3:

$$\sigma = c_1 + c_2\, ln(IM) \tag{5-3}$$

where σ is the standard deviation of residuals and c_1 and c_2 are coefficients fitted through the regression. The other heteroscedastic model used is a GPR, established in the same fashion as for the median demand model for consistency.

Next, correlation models are established in a similar manner to the variance models. *Homoscedastic correlations* are simply the correlation among all residuals, whereas step-model correlations are the correlations among residuals on either side of *IM**. The data used for heteroscedastic modelling of correlations are derived using the same binning procedure as described for the variance models, but the correlation of residuals in each bin is calculated instead of the standard deviation. Both a *linear regression* and GPR (with the same setup as for the median and variance models) are fit for every element of the correlation structure, using *IM* as the predictor. This creates 36 models, one for every element within the correlation structure (formed of the correlation among the six considered *EDP*s), but only 15 of these models are unique due to the symmetry of the correlation structure and perfect correlations between the same *EDP*s. Additionally, this study considers correlation structures that account for full correlation (referred to as LB – lower bound) and no correlation (referred to as UB – upper bound). These can effectively act as upper and lower bounds of the fragility estimates.

Finally, covariance models are established by combining the variance models for each *EDP* and the correlation structure among *EDP*s as per Equation 5-4:

$$\Sigma(IM) = \sigma(ln(IM))\rho(ln(IM))\sigma(ln(IM)) \tag{5-4}$$

where $\Sigma(IM) \in \mathbb{R}^{d\times d}$ is the covariance matrix among *EDP*s at *IM*, $\sigma(ln(IM)) \in \mathbb{R}^{d\times d}$ is a diagonal matrix containing the standard deviation of each *EDP* at *IM* on the

diagonal, $\rho(ln(IM)) \in \mathbb{R}^{d\times d}$ is the correlation matrix among *EDP*s at *IM*, and d is the number of *EDP*s under consideration. The flexibility of this modelling approach means that combinations of each variance and correlation model can be tested together. Covariance models that exclusively use the homoscedastic terms, or the step-model terms, are the equivalent of estimating the covariance of the residual terms directly, either for all the residuals in the case of the purely homoscedastic term, or for each segment of the bilinear model, when using the step-model terms.

5.2.2 Demand and loss hazards

With median and covariance models established, it is now possible to estimate the demand hazard curves, expressing the MAF of exceedance of various *EDP* levels, for all considered *EDP*s. Since multiple *EDP*s are considered, demand estimates at a given intensity level are made by combining median demand estimates with each considered covariance model to define a multivariate lognormal distribution with probability density function (PDF) as per Equation 5-5 (with the dependency on *IM* level not made explicit):

$$f(x|\mu,\Sigma) = \frac{1}{|\Sigma|^{\frac{1}{2}}(2\pi)^{\frac{d}{2}}}\, e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)} \tag{5-5}$$

where, $x \in \mathbb{R}^d$ is the vector of *EDP* thresholds under consideration, $\mu \in \mathbb{R}^d$ is the vector of median *EDP* values returned from the Cloud analysis model, $\Sigma \in \mathbb{R}^{d\times d}$is the covariance matrix as described in Section 5.2.1.2, and d is the number of *EDP*s under consideration. Demand hazard is then estimated through the convolution of the hazard curves and calculated fragility curves (e.g., Baker et al., 2021), which describe the joint CCDF of *EDPs* conditional on the *IM*.

The probabilistic distribution of the seismic losses can be estimated using storey loss functions. In this work, recourse is made to the loss functions developed by Papadopoulos et al. (2019), which are generalised for use within a PBEE formulation and are suitable for typical steel moment resisting frame office structures, such as the case examined here. In total, six storey loss functions are used, which describe groups of *IDR*-sensitive structural components, *IDR*-sensitive non-structural components, *PFA*-sensitive non-structural components at the ground, roof, and intermediate storeys, respectively, and finally the contents of the structure. The functional form of these storey-loss functions is described by Equation 5-6:

$$C_{repair}(EDP) = MaxCost\left[\varepsilon\frac{EDP^{\alpha}}{\beta^{\alpha}+EDP^{\alpha}} + (1-\varepsilon)\frac{EDP^{\gamma}}{\delta^{\gamma}+EDP^{\gamma}}\right] \quad (5\text{-}6)$$

where, C_{repair} is the estimated repair cost of the storey, $MaxCost$ is the total replacement cost associated with the component group, and ε, α, β, γ, and δ are coefficients fit through regression. Papadopoulos et al. (2019) report coefficient values for each of these parameters, including $MaxCost,$ which can additionally be customised by the user to refine repair cost estimates further if more specific data were available. Repair cost estimates returned from the storey loss functions are per 100m^2 of the storey; as such, the estimated repair costs are scaled to represent the full size of the structure under investigation. The total building repair cost is simply the sum of all six storey loss functions. Finally, the seismic loss hazard curve (i.e., the MAF of exceeding various loss levels) is made through the convolution of the hazard curve with the vulnerability curves, which describe the CCDF of repair costs conditional on the *IM*.

5.2.3 Evaluating the accuracy of demand and loss estimates

Aside from visual comparison, it is also possible to quantify the accuracy of demand and loss hazards against the unconditional estimate. For this purpose, this study proposes the Earth Mover's Distance (*EMD*) metric, also known as the Wasserstein distance (e.g., Rubner et al., 1998). *EMD* is essentially the minimum amount of work required to change one CDF into another, benchmark CDF, and is calculated via Equation 5-7:

$$EMD = \int \left| F_{ref}(X) - F_{est}(X) \right| dx \tag{5-7}$$

where $F_{ref}(X)$is the reference CDF, and $F_{est}(X)$ is the estimated CDF. As losses and demand hazard are in the form of MAF of exceedance, an additional step is needed to reach the CDF form required for evaluating *EMD*. To overcome this, a MAF of exceedance is converted to an annual probability of exceedance using Equation 3-3, assuming that the exceedance events follow a Poisson process.

EMD has previously been used within the earthquake research (e.g., Li et al., 2023; Weatherill & Lilienkamp, 2023), although not for this specific application. Other commonly used metrics for assessing the accuracy of seismic risk estimates are the KS test (e.g., Stephens, 1974) and relative entropy (e.g., Tsioulou & Galasso, 2018) – also known as the Kullback-Leibler divergence. This study prefers *EMD* because it is more sensitive to the tails of the distribution than the KS test, which makes it more effective at capturing extreme values (e.g., low-probability events that cause high *IDR* values) that the KS test might overlook. This means that *EMD* is a better test to perform when the distributions being assessed are very similar to the benchmark distribution, as is the case in this study. Additionally, *EMD* is preferred over relative entropy because it

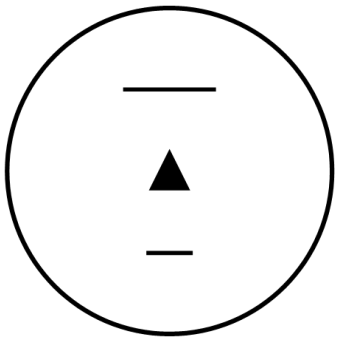
is conceptually and computationally easier to estimate an empirical CDF from risk estimates rather than a PDF, which is required for the relative entropy test. Finally, EMD outputs a metric that allows evaluation of deviation from the benchmark and the magnitude of this deviation. Whilst the KS and relative entropy tests are hypothesis tests which can only determine deviation from the benchmark. This makes EMD a better performance metric as it allows easier comparison of multiple models to the benchmark distribution.

5.3 Seismic source model and structural model

To evaluate the demand of the structure, using the approach described above, a hypothetical SSM similar to that of Rudman, Tubaldi, et al. (2024) is established as a case study. This model consists of a circular source zone of radius 100km, and two linear faults of length 40km and 20km, with the non-stationary stochastic model of Fiorentino et al. (2025) used to generate realistic ground motions. Simulated ground-motion "recordings" are generated for 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 5-2.

All earthquakes in the SSM follow the Gutenberg-Richter relationship with a minimum magnitude of 5 and b=1.0; the maximum magnitude and the a-value for each source are provided in Figure 5-2. Compared to Rudman, Tubaldi, et al. (2024), larger maximum magnitudes are selected to create a greater structural demand within the study structure, allowing for a more thorough investigation into the modelling of the nonlinear demand range. These large magnitudes align with the stochastic model selected (Fiorentino et al., 2025). In total, 100 sets of 1,000 magnitude-distance pairs are sampled at the site of interest using an importance sampling procedure (e.g.,

Jayaram & Baker, 2010), with ground motions simulated from these. The large number of ground motions being simulated enables a fully detailed residual analysis to be carried out, allowing for a proper examination of the correlation among *EDP*s within the structure.



Areal source	**Fault A**	**Fault B**
Radius = 100km	Length = 40km	Length = 20km
Stress drop = 80bars	Distance from centre = 20km	Distance from centre = 20km
Magnitude = 5.0 - 7.8	Stress drop = 40bars	Stress drop = 160bars
a = 4.5	Magnitude = 5.0 - 7.25	Magnitude = 5.0 - 7.25
	a = 4.0	a = 3.0

Figure 5-2: Seismic source model for the site of interest.

The structural model employed is replicated from the SAC phase 2 steel project (SAC, 1994). This is a well-studied structure with numerous previous applications in the literature, making the validation of the created model easier. It is a typical three-storey office building designed to the local code for Los Angeles (UBC, 1994) using post-Northridge connections. The front elevation of the structure consists of three moment-resisting bays and one simply supported bay, all of which are 9.15m in width, with each storey measuring 3.96m in height. Both the plan and front elevations, as well as the steel member sizes of the structure, are provided in Figure 4-3. For more information on the structure, see Barroso and Smith (1999), Gupta and Krawinkler (1999), Ohtori et al. (2004), and Rudman, Tubaldi, et al. (2024).

The structure is modelled in OpenSees (Mazzoni et al., 2006; McKenna, 2011) via the open-source Frame-Modeller 2D software (Elkady, 2022), which facilitates modelling and analysis in OpenSees. However, for this study, Frame-Modeller 2D was only used

for model development, with the output structural model used directly in OpenSees and MATLAB (The MathWorks Inc., 2023) for structural analysis. Modelling details and assumptions are provided in Rudman, Tubaldi, et al. (2024), as well as Section 4.3: with numerical models publicly available via a Zenodo repository (Rudman, 2026).

The first, second, and third structural periods of the model are 1.00s, 0.30s, and 0.14s, respectively, and the associated mass participation percentages are 82.15%, 14.03%, and 3.75%, respectively. The static pushover curve of the structure is presented in Figure 5-3a, 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 and Smith (1999) and Gupta and Krawinkler (1999) are also plotted in Figure 5-3a for comparison. As vibration periods, participation mass factors, and static analysis results are in good agreement with the literature (e.g., Barroso & Smith, 1999; Gupta & Krawinkler, 1999), it can be assumed that the structural model is sufficiently accurate.

For compatibility with the study of Barroso and Smith (1999), 2% damped spectral acceleration (S_a) at the fundamental period of the structure was selected as the *IM* within this study. S_a was also identified as an accurate *IM* to use alongside a bilinear demand model when assessing seismic risk (Rudman, Tubaldi, et al., 2024). Further validation of the results was completed by visually comparing *IDR* values generated through OpenSees analyses with those presented in Barroso and Smith (1999). The close agreement of *IDR* values at similar *IM* levels between each model (as presented in Figure 5-3b) suggests that the model created within this study is sufficiently accurate.

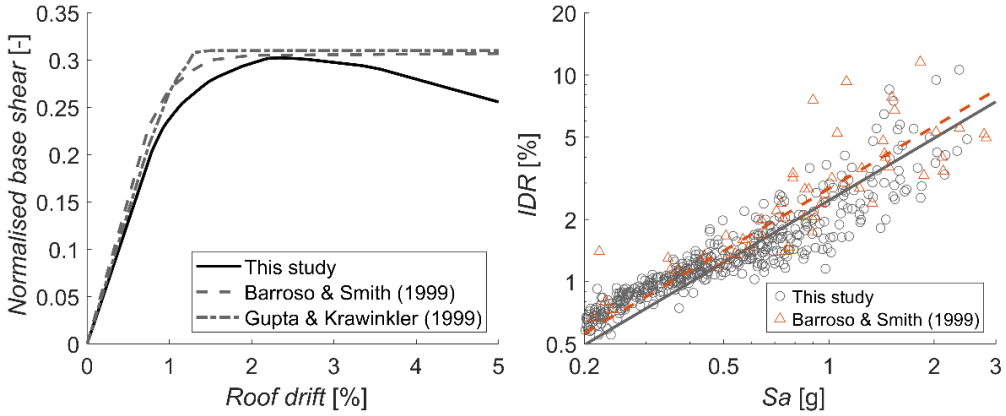


Figure 5-3: (a) Static pushover curve of the case study structure; (b) a Cloud of points returned from structural analyses compared against those from Barroso and Smith (1999), best fit lines are also presented.

5.4 Results

First, the developed demand models that account for covariance amongst *EDP*s are presented; this includes the median demand estimates (Section 5.4.1), variance estimates (Section 5.4.2), and correlation estimates (Section 5.4.3). For brevity, only the variance and correlation models that correspond to the bilinear median demand estimates are presented. Next, demand hazard estimates from all developed risk models are presented and compared (Section 5.4.4). Finally, loss estimates are evaluated in Section 5.4.6.

5.4.1 Median demand models

Figure 5-4 presents the *IDR* and *PFA* values at all three levels of the structure, returned from nonlinear time history analysis on the structural model using one set of records. The figure also presents the three developed models of the median demand: the linear regression model, the bilinear regression model, and the GPR model.

The clear benefit of the bilinear model over its linear counterpart can be seen in Figure 5-4, as it accounts for the obvious change in slope that occurs when the structural

demand enters the nonlinear phase, visually appearing to fit the data at all storeys for *IDR*. The GPR median demand model also fits the data well, with a smoother function accounting for the change in slope as the structure yields. The linear model appears to be an adequate fit for *IDR* at all S_a values; however, for *PFA*, it overpredicts the system response at higher Sa values, as the acceleration response flattens out after yielding.

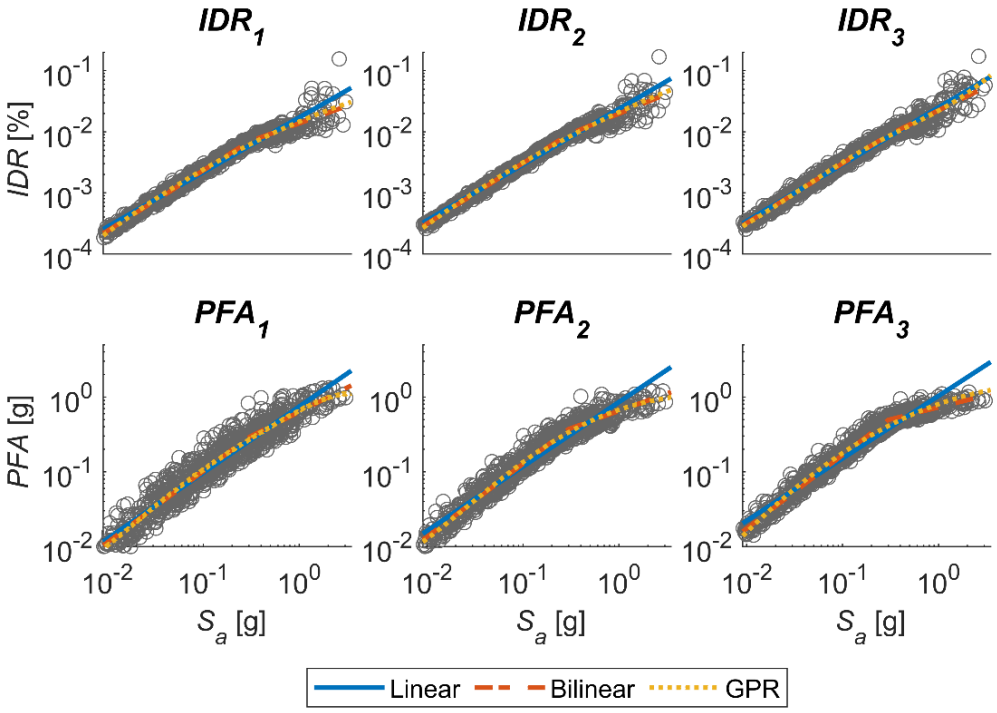


Figure 5-4: Median demand models for all engineering demand parameters (EDPs) considered within this study. Note, the subscript attached to the EDP name refers to the level within the structure.

5.4.2 Variance models

Figure 5-5 shows the standard deviation of residuals evaluated for each *IM* bin, computed for the bilinear median demand model. The homoscedastic model, step-model, and two heteroscedastic models (linear and GPR) are also plotted on this figure, fitted from the residuals of the bilinear median demand model. The homoscedastic variance model is not suitable for capturing the changing variance of *IDR* residuals as

IM increases, as it overpredicts the residual standard deviation for all *IM* values below *IM** and underpredicts the standard deviation of residuals above *IM**. This model provides a much better fit for *PFA* at all three levels of the structure.

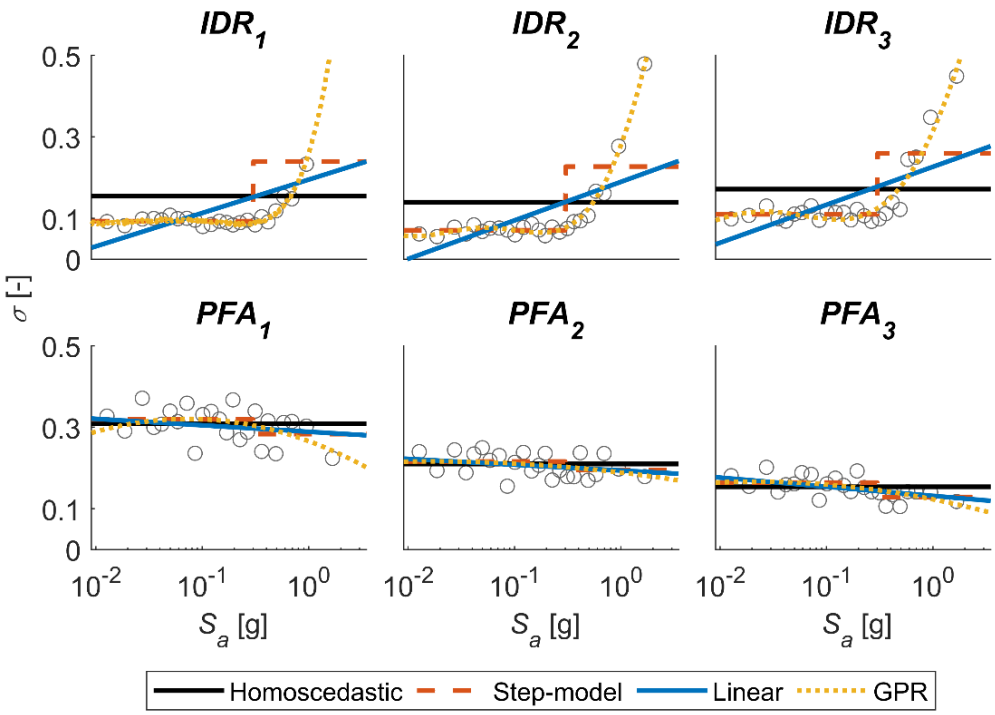


Figure 5-5: Variance models for all engineering demand parameters (EDPs) considered, using the bilinear median demand model. A logarithmic scale on both x and y axes to prevent negative standard deviation predictions. Note, the subscript attached to the EDP name refers to the level within the structure.

Both heteroscedastic models appear to better fit the standard deviation of residuals, with the GPR fully capturing the changing variance due to *IM* for all considered *EDPs*. The ability of the step-model to have two different variance values, depending on the segment of the median model, means that the residual standard deviation yields better variance estimates than the purely homoscedastic model, although not as accurately as either of the heteroscedastic variance models for the *IM* values considered. A benefit of the step-model is that it continues to generate realistic values of standard deviation

even when extrapolation is required, unlike both heteroscedastic models (the GPR model in particular), which are unbounded for *IM* values outside the considered range.

5.4.3 Correlation models

Figure 5-6 presents the correlation between each *EDP* considered as *IM* increases, as well as the homoscedastic, step-model, and both heteroscedastic models fit to these correlation values, for the bilinear median demand model. There is an apparent increase in correlation amongst *IDR* values between each floor of the structure as S_a increases. This is because the increasing damage in the structure is spread across each level, which in turn reduces the importance of higher modes, allowing the monotonically curved "inelastic first mode shape" to dominate the behaviour. In contrast, there appears to be little change in the correlation amongst *PFA* values as S_a increases since the shape of the maximum acceleration profile remains constant. Cross-correlation between *IDR* and *PFA* appears to decrease with seismic intensity for *IDR* at the first and third storeys. This makes sense as the variance of *IDR* dramatically increases after *IM**, whilst there is little change in the *PFA*. However, there is little change in correlation with *IM* between *PFA* and *IDR* on the second floor of the structure*;* this is likely because the second-storey *PFA* is more affected by the higher structural modes than the other two storeys.

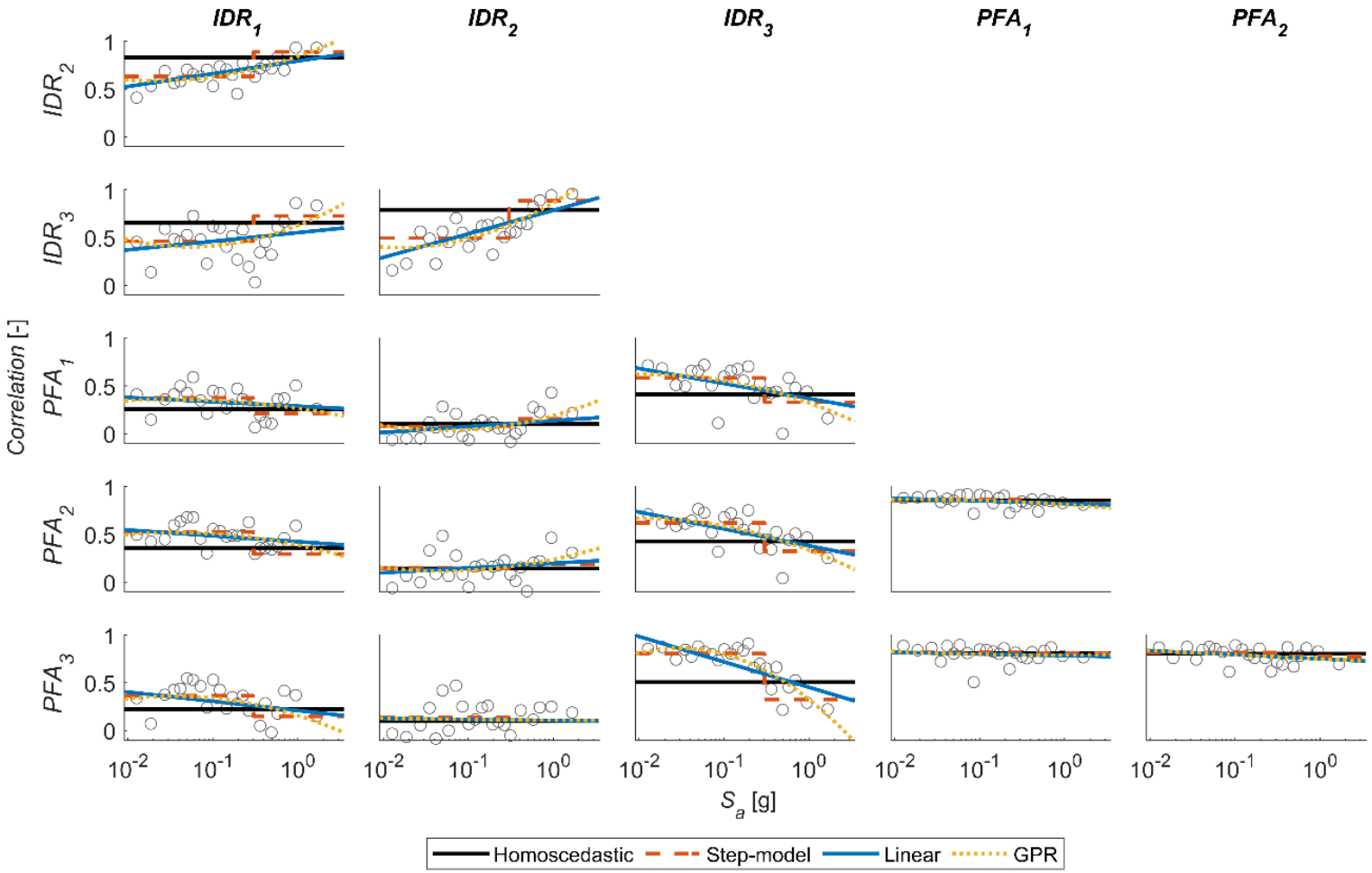


Figure 5-6: Correlation models under investigation. Only the lower half of the correlation structure is shown due to symmetry. Note, the subscript attached to the engineering demand parameter name refers to the level within the structure

Due to the limited variation in inter-*PFA* correlation, the homoscedastic model fits well. However, it does not suitably represent the rest of the correlation structure as S_a increases. This suggests that taking a purely homoscedastic approach to correlation modelling may not be appropriate for modelling the dependence between *EDP*s and thus cause inaccuracies in seismic risk estimates. All other models are a good match to the changing correlation structure along with S_a. The step-model fits the changing correlation structure well, especially for the cross-correlation between *IDR* and *PFA,* where a decrease in correlation is observed at *IM**. Both heteroscedastic models fit the correlation data well, with the GPR visually outperforming the linear regression model as it is flexible enough to cope with the differing relationship between each element of the correlation structure as S_a increases. This is well demonstrated by the flatter shape of the model describing the correlation between the second and third floor *PFA* compared to the curved model representing the relationship between the second and third floor *IDR*.

5.4.4 Demand hazard

The demand hazard curves for the various *EDP*s of interest are estimated by considering all combinations of median demand models and covariance models, and averaging results obtained for 100 sets of ground motions. Figure 5-7 presents the mean demand hazard for each of the covariance models under investigation for the bilinear median demand model, with demand hazard results for the linear and GPR median demand models presented in Appendix A. All results are compared against the corresponding unconditional benchmark hazard estimates. Visual comparisons of these mean risk estimates suggest that, in general, the bilinear median demand model provides the most accurate risk estimates, as it appears to follow the contours of the

unconditional estimate the most successfully, with results from the GPR median demand model performing similarly well. The demand hazard estimates from the linear median demand model are consistently poor.

This result aligns with the study of Rudman, Tubaldi, et al. (2024), which found that using other machine learning-based methods of median demand modelling, namely a RF regression and ANN model (with a homoscedastic variance model), did not improve the accuracy of response estimates with respect to the bilinear median demand model – albeit using the step-model for variance. This is likely due to the bilinear model's ability to produce reasonable, physically based estimates of the case study structure's response across the entire domain of seismic intensity. Whereas machine learning models are liable to poorly extrapolate outside the range of their training dataset. With respect to this study, future work could investigate the development of a more suitable kernel for use within the GPR demand model, as this could improve the model's robustness to extrapolation and its ability to estimate seismic demand.

Increasing the complexity of the variance model used can clearly improve the accuracy of risk estimates. Although the choice of correlation model appears to have little impact on the overall risk estimates presented in Figure 5-7, ignoring correlation (by using the UB model) overestimates the demand hazard for all variance models. This suggests that considering correlation does improve risk estimates, but that the method of accounting for correlation may not be crucial. It is worth noting that no conditional demand model provides a perfect match to the unconditional hazard estimates: increasing the complexity of the median demand model used may be able to improve upon this.

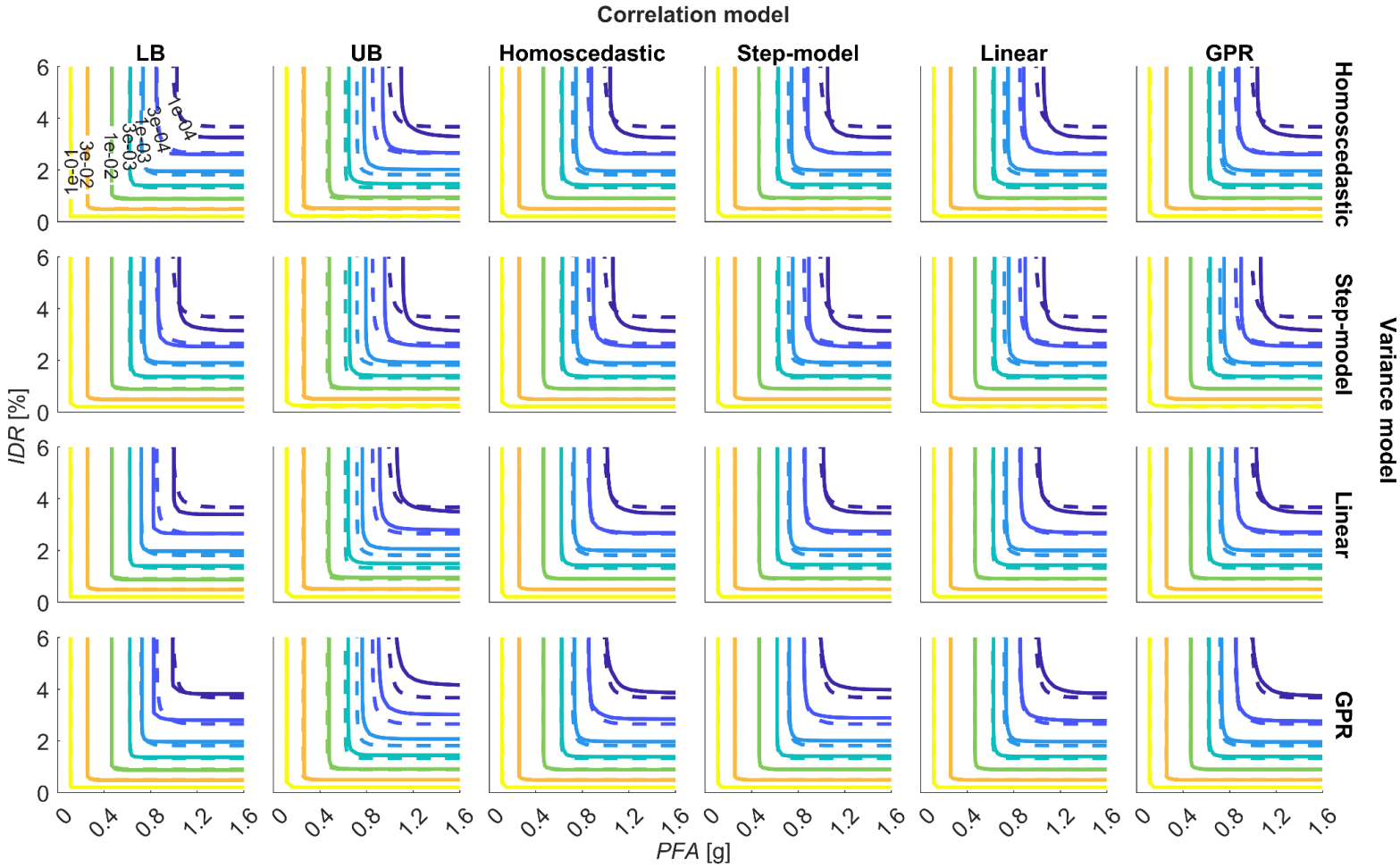


Figure 5-7: Mean demand hazard estimates made with the bilinear median demand model and all associated covariance models via the conditional method (solid lines) compared against the unconditional benchmark (dashed lines).

5.4.5 Comparing demand hazard results

Figure 5-8 presents the *EMD* values for risk estimates from all considered median demand and covariance models across all 100 sample sets. In general, this test reveals that the bilinear median demand model yields the most accurate risk estimates, while the linear median demand model provides the least accurate estimates.

The benefit of accounting for correlation amongst *EDP*s can be clearly seen from Figure 5-8, with the UB model being the least accurate for each of the variance models considered, for both the bilinear and GPR median demand models. Moreover, the least accurate demand hazard estimate from these two median demand models is composed of the GPR median demand model, a bilinear variance model, and the UB correlation model, with a median *EMD* score of 2.2×10^{-3}. This result agrees with Figure 5-7 and Appendix A, demonstrating clearly that incorporating correlation into demand hazard estimates improves their accuracy. Interestingly, however, this is not the case for the linear median demand model, where the UB correlation model provides the most accurate risk estimates across all variance models, except for the step model. This highlights the need for balance in complexity at each stage of the risk assessment process, as the accuracy of the simplistic linear median demand model is further reduced by incorporating more complex modelling techniques (i.e., variance and correlation modelling). Instead, for the more complex bilinear and GPR median demand models, the addition of complex variance and correlation models further increases their accuracy.

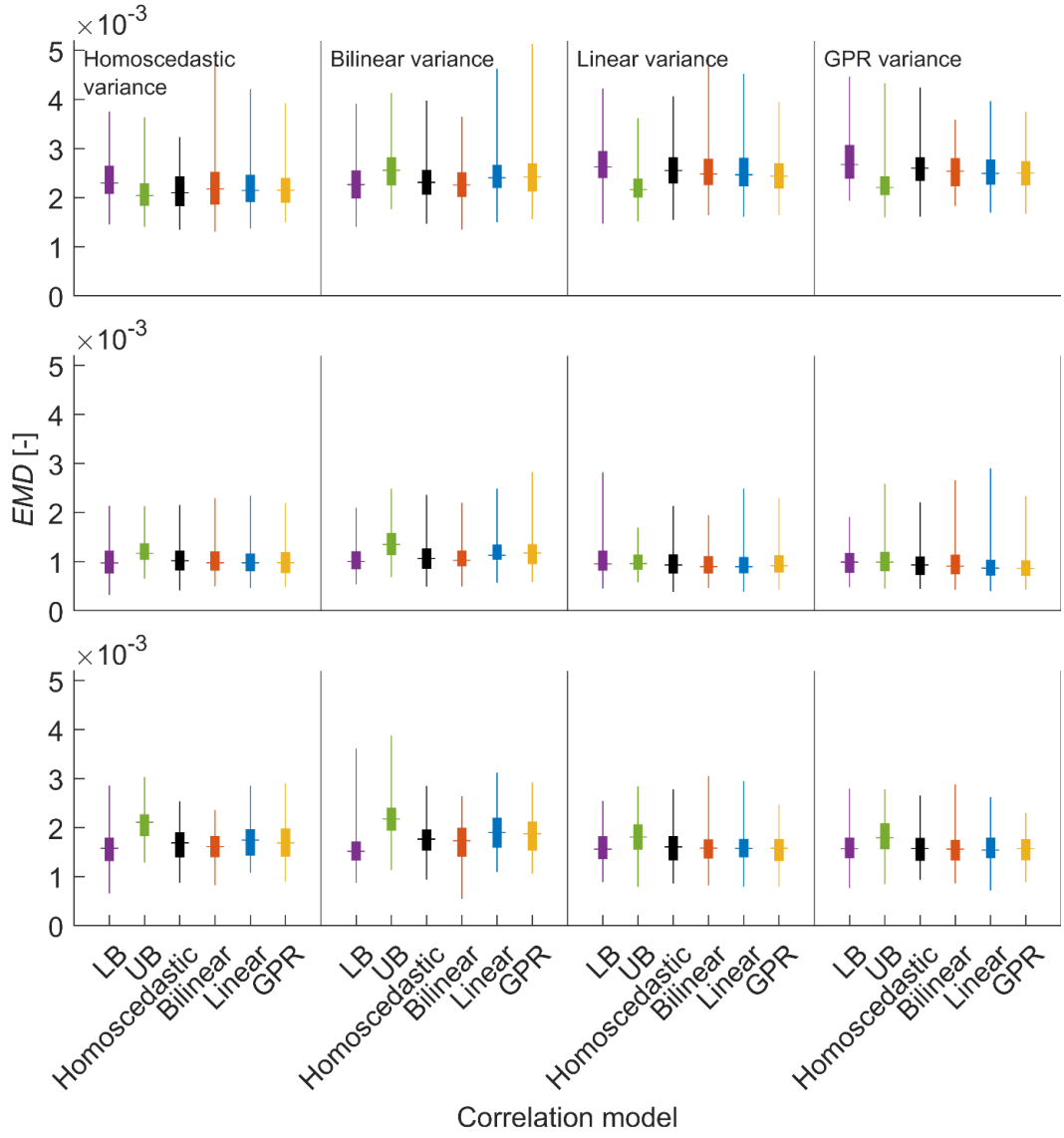


Figure 5-8: Boxplot of EMD values for all demand hazard estimates from the (top) linear, (middle) bilinear, and (bottom) Gaussian process regression median demand models across all 100 sets of records.

The overall most accurate demand hazard estimate is made by the model that uses a bilinear median demand model, a GPR-based variance term, and the GPR correlation model, with a median *EMD* value of 0.85×10^{-4}. However, the accuracy of this estimate is virtually indistinguishable from the linear regression-based variance and correlation models. The risk model that uses the bilinear median demand model, homoscedastic variance model, and LB correlation model, produces an *EMD* value of 0.97×10^{-4}. This

demonstrates that even without increasing the complexity of the variance model, assuming full correlation between *EDP*s can yield accurate demand hazard estimates for this structure. In general, the GPR-based variance term appears to give the most accurate demand hazard estimates for all correlation models. The linear regression-based and step-model for variance also perform well, with the increase in variance model complexity appearing to increase the accuracy of risk estimates.

5.4.6 Loss hazard

Simulated *EDP*s are used to compute repair costs using the storey loss functions described in Section 5.2.2. The mean MAF of losses across all 100 sets of records, for each considered risk model, are presented in Figure 5-9. As with the demand hazard estimates, the linear median demand model performs well in estimating losses at larger MAF values of exceedance but struggles to accurately reflect the unconditional losses at small MAF values, where it fails to capture the drop in MAF of losses at approximately $\$1.5\times10^5$. The drop occurs between $\$1.5\times10^5$ and $\$4.0\times10^5$ and likely exists as this is where most repair cost realisations are concentrated (i.e., this is a range of values that corresponds to the most common intensity-damage values), thus causing a steep decrease in the frequency of repair costs.

Both the bilinear and GPR median demand models appear to fit the unconditional loss estimates more accurately than the linear model, with the risk models derived from the bilinear median demand model providing a slightly superior fit. Models using the step-model for variance and the GPR variance model seem to fit the unconditional estimate particularly well.

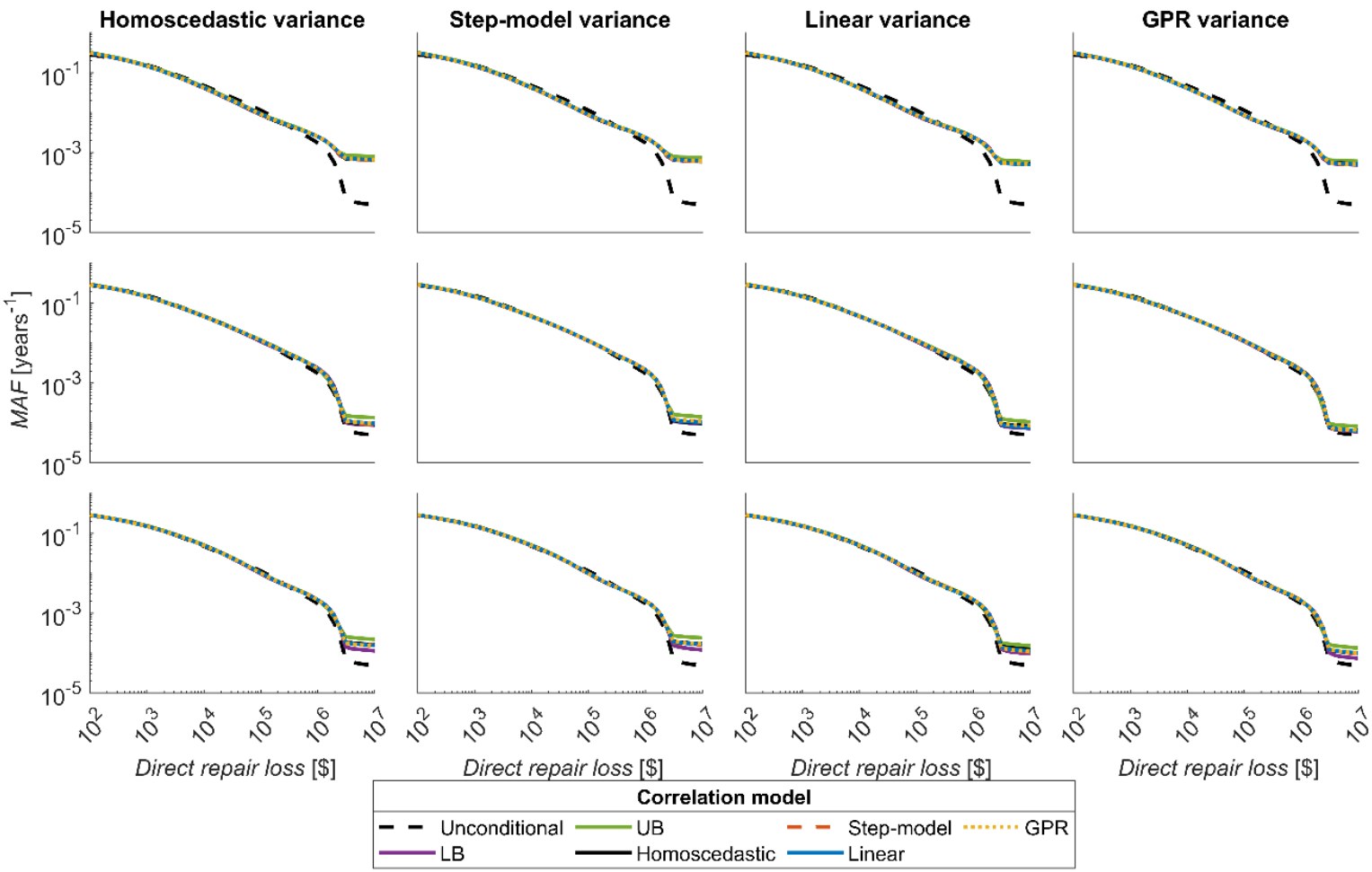


Figure 5-9: Mean loss estimates made from the (top) linear, (middle) bilinear, and (bottom) Gaussian process regression median demand models, and associated covariance models, *across all 100 sets of records.*

Changing the correlation model has little impact on the linear median demand models; however, the effect of each correlation model appears to grow as the complexity of the median demand model increases, with a greater difference in model estimates seen from Figure 5-9. Moreover, the selection of the correlation model appears to have a larger effect for events with a smaller estimated MAF of loss exceedance. The UB model always produces the most conservative loss estimates, again demonstrating the importance of accounting for correlation amongst *EDP*s when assessing losses.

5.4.6.1 Comparing loss results

Figure 5-10 presents the *EMD* values of loss estimates made for all median demand and covariance models across all 100 sets of samples. Note that a logarithmic y-axis is used to allow for proper visualisation of results due to their large spread. These results confirm the visual comparison, showing that loss estimates derived from the linear median demand model are considerably less accurate, with all variance and correlation model variants producing similarly poor results.

Interestingly, the UB correlation model is the least accurate for the linear median model and all corresponding variance models. This contrasts with the demand hazard estimates (Figure 5-8), where the UB model gave the most accurate results for that median demand model. As the LB correlation model is also the least accurate predictor of losses for all variance models and all other median demand models, it is clear that, when estimating losses, it is never preferable to assume no correlation amongst EDPs for the structure being investigated.

Models derived from the bilinear and GPR median demand models appear to perform more similarly at the loss estimation stage than the demand hazard stage, with the bilinear-related models only just edging out those derived from the GPR median

demand model. The overall most accurate model uses the bilinear median demand model, a GPR variance model, and a GPR correlation model, producing an *EMD* score of 1.0×10^{-3}. The most accurate loss estimation using the GPR median demand model uses a GPR variance model and a step-model for correlation and has an *EMD* score of 1.3×10^{-3}.

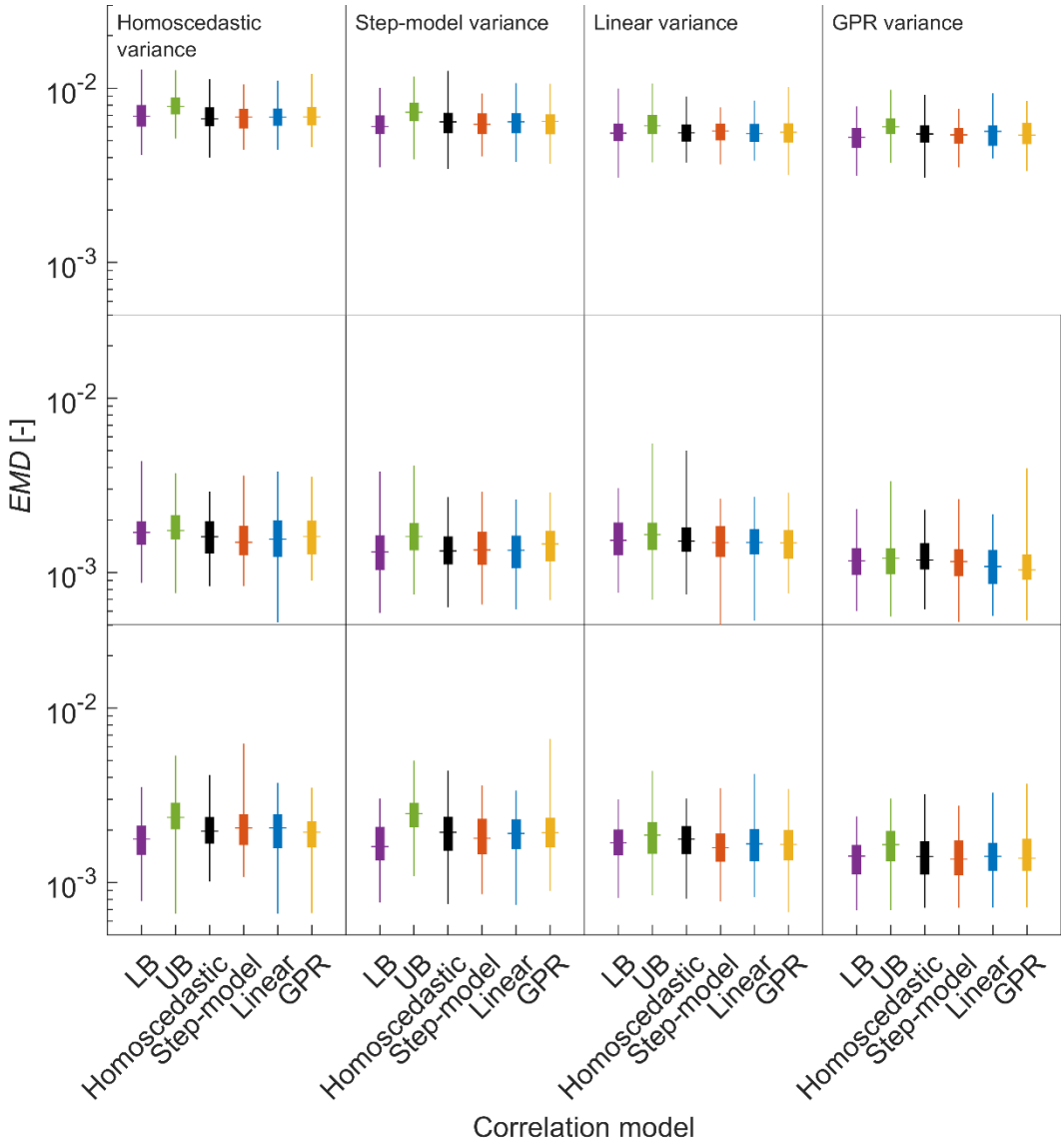


Figure 5-10: Boxplot of EMD values for all loss estimates from the (top) linear, (middle) bilinear, and (bottom) Gaussian process regression median demand models across all 100 sets of records. A logarithmic y-axis is used to display all results on one scale.

It is clear from these results that increasing the complexity of the variance model improves the accuracy of loss estimates. Furthermore, although using a more complex correlation model does not always yield more accurate loss estimates, there is clear evidence that considering correlation (using any method) certainly does improve the accuracy of loss assessment. In fact, all models that assume full correlation amongst *EDP*s perform similarly well to those where the correlation is modelled, so even accounting for this simple assumption may lead to increased accuracy of loss estimates.

Both the KS and relative entropy tests were also performed on loss estimates to help evaluate the effectiveness of *EMD* as a risk model performance metric. Each of the 72 total loss models (3 median models, 4 variance models, and 6 correlation models) were ranked according to their mean *EMD*, KS, and relative entropy scores, respectively. Pairwise rank correlations were then computed among these three metrics. The correlation between *EMD* and KS scores was 0.03, and between *EMD* and relative entropy scores was –0.31, indicating that *EMD* results are largely uncorrelated with these established performance metrics. In contrast, the KS and relative entropy scores showed a correlation of 0.63, suggesting a relatively strong relationship between them. These findings demonstrate that the *EMD* metric provides a distinct evaluation of model performance compared to the KS and relative entropy tests.

5.5 Sensitivity of loss estimates to the number of ground motions

To further investigate the accuracy of each considered covariance model, a sensitivity analysis was performed on the number of samples used to develop the models and create the loss assessment. To that end, variance and correlation models were redeveloped for the bilinear median demand model, using the same procedure described in Section 5.5, for sample sizes used in Cloud analysis between 50 and 1000 records. Figure 5-11 presents the sensitivity results for each considered covariance model, in the form of mean *EMD* values for loss estimates.

For all four variance models, Figure 5-11 shows that the assumption of independence is seldom suitable for seismic risk assessment when using a homoscedastic or step-model for variance. The performance of the UB (independent) model for the Linear and GPR variance models is slightly more comparable with the other correlation models; however, it never provides the most accurate loss estimates, even when using a more realistic sample size, such as 200 samples. The LB model (which assumes full correlation amongst *EDP*s) is almost always more accurate than the UB model across all sample sizes. The performance of both assumptions (no correlation and full correlation) across all sample sizes helps solidify the conclusion that correlations among *EDP*s should always be accounted for in seismic risk and loss assessment.

The GPR variance model appears to provide the least sensitive results to reducing the sample size used to generate the models. Moreover, when only 200 samples are used (an order of magnitude more commonly considered in similar studies), the model that uses the GPR variance model and the homoscedastic correlation model produces the most accurate results, with an *EMD* value of 3.5×10^{-3}. However, assuming full

correlation (LB model) yields an *EMD* value of 3.7×10^{-3}, which is not much higher. The sensitivity of each correlation model appears to be relatively similar, relative to the variance model used. All considered covariance models become considerably less accurate when less than 200 records are used, this is likely due to the low number of records in the bins used to create the models, leading to poor characterisation of the benchmark losses. This is why when just 50 samples are used, the homoscedastic model for correlation becomes more accurate for three of the four variance models.

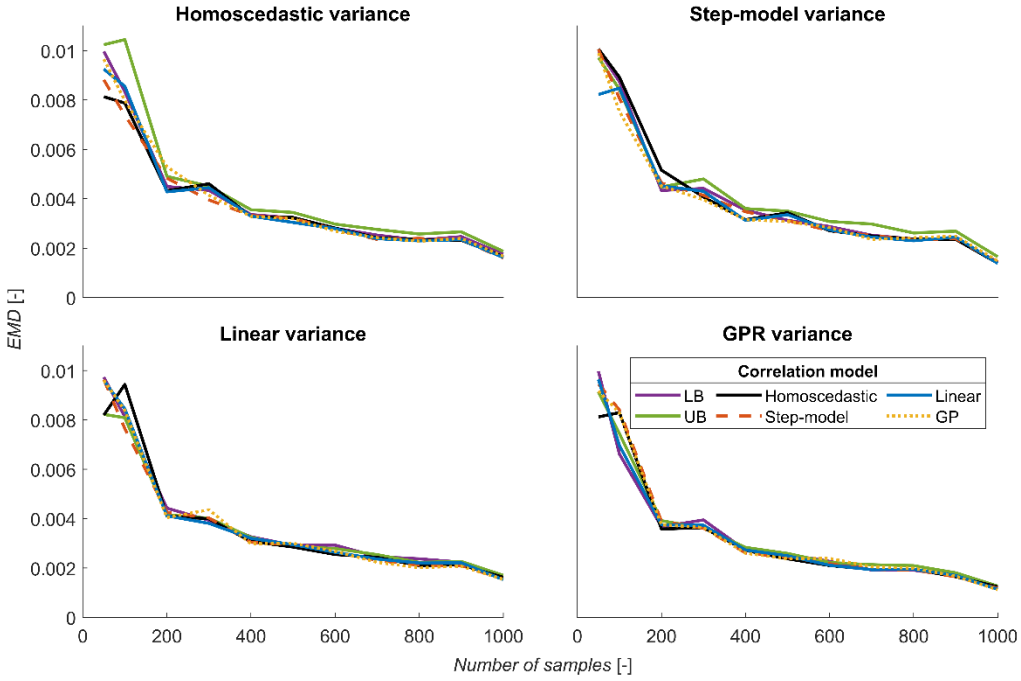


Figure 5-11: Mean EMD value of the loss estimate for each model derived from the bilinear median demand model against the number of samples used to develop each model.

Additional sensitivity analysis was also performed on the most accurate risk model (which consists of a bilinear median demand model, a GPR variance model, and a GPR correlation model) to evaluate the effectiveness of model set up. Figure 5-12.a investigates the sensitivity of *EMD* to the number of bins used to build the variance model, confirming that using 25 bins to develop these models yields the most accurate

loss estimates. Considering 250 bins, each consisting of four records, also appears to yield accurate loss estimates, with an *EMD* value of 1.2×10^{-3}; however, calculating standard deviation and correlation with just four records is likely to be unreliable in other contexts. Figure 5-12.b presents the sensitivity of the chosen risk model to the kernel used to develop the GPR variance and correlation models. This investigation shows the squared exponential kernel leads to the most accurate loss estimates, hence confirming its choice within this study. It is worth noting that all other kernel choices produce an *EMD* value of approximately 1.5×10^{-3}, so these models are relatively insensitive to the GPR kernel choice.

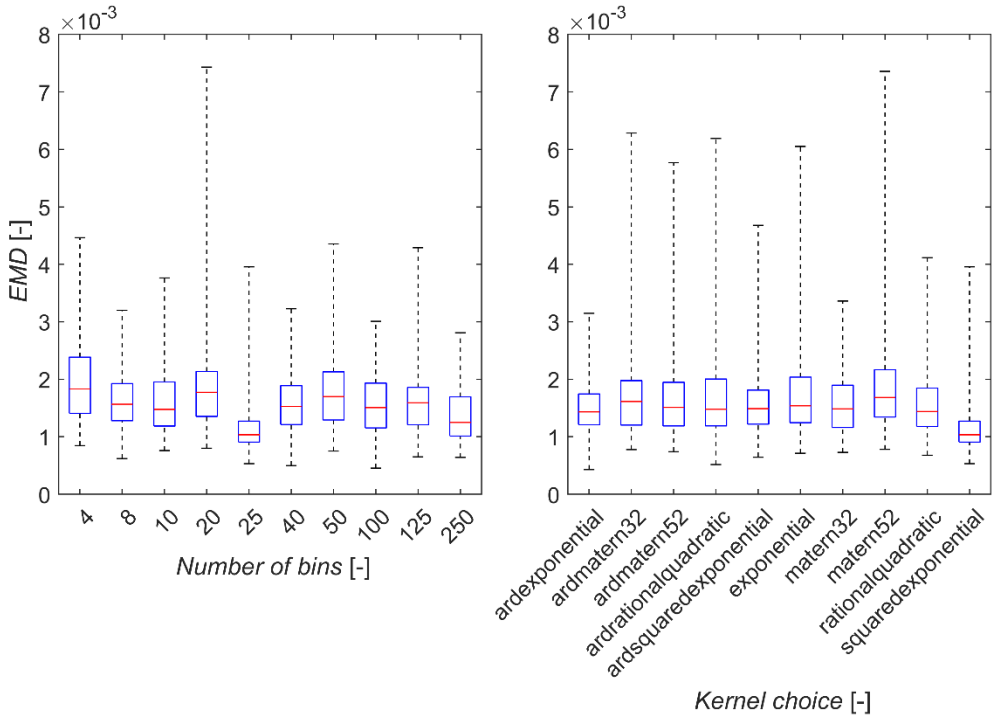


Figure 5-12: Sensitivity analysis results considering the bilinear median demand model, Gaussian process regression (GPR) variance model, and GPR correlation model. (a) presents the sensitivity of this model to the number of bins used to create the variance and correlation models, (b) presents the sensitivity of this model to the kernel choice used within the GPR models.

5.6 Conclusions

This study has investigated the impact of modelling the covariance among *EDP*s on seismic risk assessment of a steel frame using Cloud analysis. Using a hypothetical SSM and a stochastic ground motion model, the seismic response of a three-story steel moment-resisting frame has been evaluated through nonlinear time-history analyses, with interstorey drift and peak floor acceleration at each level of the structure selected as the *EDPs*. The median demand has been estimated using three different Cloud analysis-based models: a linear model, a bilinear model, and a GPR-based model. Next, four distinct correlation and variance models have been developed, with each combination brought together to create a variety of covariance models, each accounting for either homoscedasticity, heteroscedasticity, or a combination of the two assumptions.

Using each median demand model and covariance model, estimates of the demand hazard and losses (in terms of repair costs) have been obtained for the considered structures. Both qualitative and quantitative comparisons have been made at the demand hazard and loss levels, comparing the estimates obtained for each covariance model against those obtained via the unconditional approach (i.e., directly estimating demand exceedance frequencies from simulated ground motion time histories). Earth mover's distance has been used, for the first time, as a metric to quantify the accuracy of these estimates. Finally, a sensitivity analysis has been performed to evaluate how the accuracy of risk estimates using each covariance model changes with the number of samples employed, the number of bins used to create the models, and the GPR kernel used. As this is essentially a structure with a bilinear response dominated by the

first mode, it is reasonable to assume that the obtained results generalise to structures with a similar behaviour.

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

- The bilinear median demand model provides the most accurate risk and loss estimates. The physical meaning behind this model ensures that estimates are stable and bounded across all intensity values, resulting in a slightly better estimator of seismic response than the more complex GPR-based model. Developing a bespoke kernel for use within the GPR could make its application more suitable for this context. The linear demand model lacked the complexity to account for the change in system response post-yielding.
- Using heteroscedastic variance models improves the accuracy of risk estimates for all three median demand models considered. The GPR-based variance model is the most accurate at both the demand hazard and loss estimation stages.
- Accounting for correlation amongst *EDP*s improves the accuracy of risk estimates. Although there is little difference between results that account for demand correlation (whether homoscedastic or heteroscedastic), these are considerably more accurate than those obtained with models that assume independence amongst *EDP*s at both the demand hazard and loss estimation levels.
- The sensitivity analysis has revealed that, at almost all sample sizes evaluated (between 200 and 1000), it is more accurate to assume full correlation amongst *EDP*s instead of simply ignoring correlation.

- When considering a sample size of 200 records, a more practically feasible number of records for risk assessment studies, a bilinear median demand model is recommended, with a covariance model composed of a GPR-based variance model and a homoscedastic correlation model. This finds a good balance between the accuracy of results and the simplicity of the implementation.

Each of these conclusions has broad implications for both seismic risk assessment and engineering design practice for the specific use case explored within this study. The standard approach described by FEMA P-58 is essentially the same as selecting both the homoscedastic variance and correlation terms in this approach, albeit characterised through a multiple stripe analysis-based approach. In essence, this shows that the approach described in this article can make significant enhancement to current risk assessment practice. By choosing the correct median demand model and accounting for variability and correlations, engineers can more accurately predict structural responses to ground shaking, enabling both more reliable loss estimation and more efficient, performance-based structural design.

The method presented in this study could help enhance future seismic risk studies by incorporating an often-overlooked feature of probabilistic seismic demand modelling. There is great potential for incorporating advanced covariance among *EDP*s to improve the accuracy of seismic risk, damage, and loss estimates.

6 Conclusions and recommendations

This thesis explored potential improvements to models used for conditional earthquake risk assessment, and examined how novel methods and machine learning opportunities could enhance the accuracy of risk estimates in the context of performance based earthquake engineering (PBEE). A particular focus was on the hazard and demand modelling stages of the conditional risk assessment framework, due to the wide variety of potential modelling techniques available to the user.

The three chapters within the main research body of this thesis used stochastic ground motion models and hypothetical seismic source models to create realistic case studies. Employing hypothetical case studies facilitated clearer comparisons between the different modelling techniques investigated, while also enabling a comparison against an unconditional "true" risk benchmark (i.e., directly estimating demand exceedance frequencies from simulated ground motion time histories). Evaluating these comparisons provides greater insight into the effectiveness of each modelling technique examined and helps confirm their validity. This would not necessarily be possible using only ground motion recordings, which are insufficient in many regions of the world.

While all chapters were based on a similar experimental setup, each has examined different aspects of the PBEE framework. In Chapter 3, three different ground motion

modelling techniques were compared against a benchmark solution from a hypothetical case study: two ground-motion models with simple and more complex functional forms, respectively, and an artificial neural network. Hazard estimates were extended to estimate the demand hazard of a single degree of freedom system, with ductility demand as the engineering demand parameter (*EDP*). Moreover, the feasibility of an observational-based hazard assessment model was explored using the NGA-West 2 ground motion database, with promising results obtained. From this study, the following conclusions were drawn:

- Out of the three considered models, only the artificial neural network produced statistically similar results to the unconditional benchmark, according to the Kolmogorov-Smirnov test.
- Hazard disaggregation revealed that the ground-motion models with fixed functional forms struggled to predict high ground-motion levels caused by large magnitude, short distance events, causing the over-prediction of hazard, and ultimately risk.
- The performance of the observational-based hazard model indicated that, in certain situations, it may be possible to estimate hazard and risk purely from empirical data, without needing a ground-motion model or conditioning the risk estimate on an intensity measure (*IM*).

In Chapter 4, the well-studied topic of *IM* selection was extended by considering how the joint selection of *IM* and probabilistic seismic demand model impacts seismic risk estimates. Again, using a hypothetical case study, 15 different *IM*s were investigated, along with five demand models: a linear model, a bilinear model, two machine learning

models, and an empirical binning model, using interstorey drift (*IDR*) as the *EDP*. After estimating the demand hazard on the structure, the results were extended to estimate the damage to both structural and non-structural components, helping to understand how these choices impact the risk assessment process downstream. *IM*-demand model performance was assessed via a novel performance metric that measured the distance between the probability distributions of the conditional and unconditional solutions. This was also compared against more traditional *IM* performance metrics, i.e., *efficiency*, *sufficiency*, and *practicality*. This study revealed that:

- The most accurate *IM*-demand model combination was Arias intensity with the empirical demand model. However, more traditionally used *IM*s, such as spectral acceleration at the fundamental vibration period of the structure, also performed well.
- Sensitivity analysis revealed that when carrying out a risk assessment with 200 records, a more usual number for risk assessment, spectral acceleration with a bilinear demand model, was the most accurate combination.
- The developed performance metric showed that traditional *IM* performance measures do not necessarily guarantee the most accurate demand hazard estimates when a large set of records is used to develop demand models.

Finally, Chapter 5 investigated how the covariance amongst *EDP*s could be incorporated into a seismic risk assessment to enhance the accuracy of risk estimates. Three Cloud analysis models were derived: a linear model, a bilinear model, and a Gaussian process regression (GPR) model. These models were based on a hypothetical

case study, using *IDR* and peak floor acceleration as the *EDP*s. Further analyses were conducted to develop various models for the variance and correlation of the residuals of each model, which were then combined to create models of the residual covariance structure. Each covariance model was employed to help estimate both the demand hazard and loss estimates for the considered structure. These estimates were thoroughly compared using, for the first time, Earth mover's distance as a seismic risk performance metric, which compares conditional risk estimates against the unconditional solution. The main findings from this study were:

- The most accurate risk model for estimating losses used the bilinear median demand model, a GPR-based variance model, and a GPR-based correlation model.
- In general, increasing the variance model's complexity improved the accuracy of risk estimates. While this was not necessarily the same for the choice of correlation model, it was clear that considering correlation yielded more accurate risk estimates than ignoring it, which is the more conventional practice.
- Sensitivity analysis showed that when performing a risk assessment with only 200 records, a bilinear median demand model combined with a GPR-based variance model and a homoscedastic correlation model is recommended to achieve the most accurate and easy-to-implement results.

Since each chapter has examined a different aspect of the PBEE risk assessment framework, their conclusions can be combined to recommend a risk assessment approach that provides accurate risk estimates for a mid-rise steel moment-resisting frame structure. Such an idealised hazard-demand model would include:

1. A ground-motion model derived using an artificial neural network,

2. Arias intensity, or spectral acceleration, as the *IM*, depending on whether more or fewer than 700 records are used to build the demand model,
3. A demand model using a bilinear functional form,
4. A model that considers covariance amongst *EDP*s using a GPR-based variance model and a homoscedastic correlation term.

The generalisability of the thesis' findings is constrained by their dependence on the structural model and stochastic ground motion models used, particularly in identifying individual models or modelling approaches that yield the most accurate risk estimates. As such, the quantitative results from this thesis are expected to vary for different structural typologies and seismic environments, and may not be generalisable. However, these findings are expected to be broadly applicable to structural systems exhibiting similar nonlinear behaviour (e.g. bilinear response) and subjected to comparable ground motion characteristics.

Moreover, several of the key conclusions arise from the structure of the PBEE framework itself. These include the dependence of optimal intensity measure selection on the demand model, and the sensitivity of risk estimates to modelling assumptions. As these are inherent features of the PBEE formulation, they are expected to hold across a wide range of earthquake engineering applications.

The findings of this thesis help to offer new perspectives on earthquake risk assessment by re-examining traditional modelling approaches through a different lens and demonstrating how innovative techniques can enhance existing practices. Using better risk models can have an impact across society, resulting in improved decision-making,

safer structures, more accurate insurance rates, and ultimately helping protect people and property.

6.1 Scope for future research

This thesis has identified and evaluated methods for improving earthquake risk assessment practices. While the proposed methods advance current approaches, they do not represent an exhaustive set of solutions. Moreover, the effectiveness of some methods may be influenced by the underlying assumptions and limitations of the models used, as well as by the hazard scenarios and case studies considered. The following is a list of recommendations for future study based on the work contained within this thesis:

- **Exploration of alternative stochastic ground motion models**

  Many stochastic ground motion models exist beyond the two investigated in this thesis. The choice of simulation model will likely influence seismic risk estimates, as models are derived from different ground-motion datasets and employ different derivation techniques. Future studies could compare these models to identify the optimal balance between realism and computational efficiency for studies similar to those within this thesis.

- **Evaluation of magnitude and distance sampling techniques**

  Similarly, a range of techniques for estimating earthquake magnitude and distance are available in the literature. Each of these techniques could present distinct trade-offs in terms of bias, representativeness, efficiency and other such parameters. Future research could assess these techniques to determine whether they lead to more accurate and realistic suites of simulated ground motions. Further, exploring

potential interactions between the sampling method and the stochastic ground motion model would be valuable for determining whether they can be jointly selected.

- **Validation of risk models against empirical ground motion datasets**

  The results in Chapters 3 to 5 depend on how realistic the ground motion simulation method is. Although the simulated datasets are validated, they cannot fully replace real ground motion recordings. It would be helpful for future research to test the modelling approaches recommended in this thesis against real ground motions. This investigation could help clarify how results differ from traditional risk assessment practice and reinforce the relevance of the thesis findings.

- **Integration of physical principles into machine learning models**

  Machine learning techniques often perform poorly in extrapolation because their functional forms may behave unpredictably outside the training domain. Future research could explore integrating physical meaning into machine-learning-derived earthquake risk models. For example, developing a GPR demand model with a physically informed kernel and mean model could improve predictive performance.

- **Refinement of the observation-based hazard assessment approach**

  Section 3.6 introduced an observation-based hazard assessment method using the unconditional approach and data from the NGA West 2 database. This method shows strong potential and should be further developed. This could include testing the approach on alternative ground-motion databases or using the method to estimate damage and losses on a case-study structure, which could then be compared with real-world observations. Developing this observational model

would increase the relevance of this thesis by providing a real-world benchmark for conditional risk assessments, thereby reducing the need for simulated ground motions.

- **Application of research findings to a broader range of structure types**

  The modelling approaches recommended in this thesis could be further tested on a wider variety of structures, including high-rise buildings, bridges, and other complex systems. Such studies would help evaluate the generality and robustness of the proposed methods.

- **Risk assessment across different resolutions**

  Future work could also examine how the proposed techniques perform at varying levels of resolution, from site-specific (non-ergodic) analyses to regional or portfolio-scale assessments. This would help identify modelling techniques that maintain accuracy and stability across different seismic risk contexts.

- **Development of a single PBEE model**

  Loss estimates in PBEE are achieved through a set of discrete, interchangeable stages, offering flexibility but also the potential for significant uncertainties and biases when non-optimal methods are selected. A potential extension of this thesis could be to develop a single machine learning-informed model that maps site and structural parameters directly to losses, thereby optimising the PBEE procedure and reducing epistemic uncertainties attributed to user error.

References

Abrahamson, N., & Silva, W. (2008). Summary of the Abrahamson & Silva NGA Ground-Motion Relations. *Earthquake Spectra*, *24*(1), 67–97. https://doi.org/10.1193/1.2924360

Alavi, A. H., & Gandomi, A. H. (2011). Prediction of principal ground-motion parameters using a hybrid method coupling artificial neural networks and simulated annealing. *Computers & Structures*, *89*(23), 2176–2194. https://doi.org/https://doi.org/10.1016/j.compstruc.2011.08.019

Aldama-Bustos, G., Douglas, J., Strasser, F. O., Daví, M., & MacGregor, A. (2023). Methods for assessing the epistemic uncertainty captured in ground-motion models. *Bulletin of Earthquake Engineering*, *21*(1), 1–26. https://doi.org/10.1007/s10518-022-01515-8

Alidadi, N., & Pezeshk, S. (2025). State of the art: Application of machine learning in ground motion modeling. *Engineering Applications of Artificial Intelligence*, *149*, 110534. https://doi.org/https://doi.org/10.1016/j.engappai.2025.110534

Ambraseys, N. N., Douglas, J., Sarma, S. K., & Smit, P. M. (2005). Equations for the estimation of strong ground motions from shallow crustal earthquakes using data from Europe and the Middle East: Horizontal peak ground acceleration and spectral acceleration. *Bulletin of Earthquake Engineering*, *3*(1), 1–53. https://doi.org/10.1007/s10518-005-0183-0

Ancheta, T. D., Darragh, R. B., Stewart, J. P., Seyhan, E., Silva, W. J., Chiou, B. S.-J., Wooddell, K. E., Graves, R. W., Kottke, A. R., Boore, D. M., Kishida, T., & Donahue, J. L. (2014). NGA-West2 Database. *Earthquake Spectra*, *30*(3), 989–1005. https://doi.org/10.1193/070913eqs197m

Arias, A. (1969). Measure of earthquake intensity. *pp 438-83 of Seismic Design for Nuclear Power Plants. /Hansen, Robert J. (ed.). Cambridge, Mass. Massachusetts Inst. of Tech. Press (1970)*.(None).

Aslani, H., & Miranda, E. (2005). *Probabilistic earthquake loss estimation and loss disaggregation in buildings* (Technical Report 157). http://purl.stanford.edu/kw177ns9250

Assatourians, K., & Atkinson, G. M. (2013). EqHaz; An Open-Source Probabilistic Seismic-Hazard Code Based on the Monte Carlo Simulation Approach. *Seismological Research Letters*, *84*(3), 516–524. https://doi.org/10.1785/0220120102

Atkinson, G. M., & Silva, W. (2000). Stochastic modeling of California ground motions. *Bulletin of the Seismological Society of America*, *90*(2), 255–274. https://doi.org/Doi 10.1785/0119990064

Au, S.-K., & Beck, J. L. (2003). Subset simulation and its application to seismic risk based on dynamic analysis. *Journal of Engineering Mechanics*, *129*(8), 901–917. https://doi.org/https://doi.org/10.1061/(ASCE)0733-9399(2003)129:8(901)

Azar, S., & Dabaghi, M. (2021). Simulation-Based Seismic Hazard Assessment Using Monte-Carlo Earthquake Catalogs: Application to CyberShake. *Bulletin of the Seismological Society of America*, *111*(3), 1481–1493. https://doi.org/10.1785/0120200375

Baker, J., Bradley, B., & Stafford, P. (2021). *Seismic Hazard and Risk Analysis*. Cambridge University Press. https://doi.org/DOI: 10.1017/9781108425056

Baker, J. W. (2015). Efficient Analytical Fragility Function Fitting Using Dynamic Structural Analysis. *Earthquake Spectra*, *31*(1), 579–599. https://doi.org/10.1193/021113eqs025m

Baker, J. W., & Cornell, C. A. (2005). A vector-valued ground motion intensity measure consisting of spectral acceleration and epsilon. *Earthquake Engineering & Structural Dynamics*, *34*(10), 1193–1217. https://doi.org/https://doi.org/10.1002/eqe.474

Baker, J. W., & Cornell, C. A. (2006). Spectral shape, epsilon and record selection. *Earthquake Engineering & Structural Dynamics*, *35*(9), 1077–1095. https://doi.org/https://doi.org/10.1002/eqe.571

Bandini, P. A. C., Siqueira, G. H., Padgett, J. E., & Paultre, P. (2022). Seismic Performance Assessment of a Retrofitted Bridge with Natural Rubber Isolators in Cold Weather Environments Using Fragility Surfaces. *Journal of Bridge Engineering*, *27*(6), 04022040. https://doi.org/doi:10.1061/(ASCE)BE.1943-5592.0001873

Barroso, L., & Smith, H. (1999). *Performance evaluation of vibration controlled steel structures under seismic loading*. John A Blume Earthquake Engineering Center Technical Report 133.

Bazzurro, P., & Cornell, C. A. (1999). Disaggregation of seismic hazard. *Bulletin of the Seismological Society of America*, *89*(2), 501–520. https://doi.org/10.1785/BSSA0890020501

Bazzurro, P., Cornell, C. A., Shome, N., & Carballo, J. E. (1998). Three Proposals for Characterizing MDOF Nonlinear Seismic Response. *Journal of Structural Engineering*, *124*(11), 1281–1289. https://doi.org/doi:10.1061/(ASCE)0733-9445(1998)124:11(1281)

Beauval, C., Honoré, L., & Courboulex, F. (2009). Ground-Motion Variability and Implementation of a Probabilistic–Deterministic Hazard MethodGround-Motion Variability and Implementation of a Probabilistic–Deterministic Hazard Method. *Bulletin of the Seismological Society of America*, *99*(5), 2992–3002. https://doi.org/10.1785/0120080183

Bijelić, N., Lin, T., & Deierlein, G. G. (2019). Quantification of the Influence of Deep Basin Effects on Structural Collapse Using SCEC CyberShake Earthquake Ground Motion Simulations. *Earthquake Spectra*, *35*(4), 1845–1864. https://doi.org/10.1193/080418eqs197m

Bommer, J. J., & Scherbaum, F. (2008). The Use and Misuse of Logic Trees in Probabilistic Seismic Hazard Analysis. *Earthquake Spectra*, *24*(4), 997–1009. https://doi.org/10.1193/1.2977755

Bommer, J. J., Stafford, P. J., & Alarcón, J. E. (2009). Empirical Equations for the Prediction of the Significant, Bracketed, and Uniform Duration of Earthquake Ground Motion. *Bulletin of the Seismological Society of America*, *99*(6), 3217–3233. https://doi.org/10.1785/0120080298

Boore, D. M. (2003). Simulation of Ground Motion Using the Stochastic Method. *pure and applied geophysics*, *160*(3), 635–676. https://doi.org/10.1007/PL00012553

Boore, D. M., & Atkinson, G. M. (2008). Ground-Motion Prediction Equations for the Average Horizontal Component of PGA, PGV, and 5%-Damped PSA at Spectral Periods between 0.01 s and 10.0 s. *Earthquake Spectra*, *24*(1), 99–138. https://doi.org/10.1193/1.2830434

Boore, D. M., Di Alessandro, C., & Abrahamson, N. A. (2014). A Generalization of the Double-Corner-Frequency Source Spectral Model and Its Use in the SCEC BBP Validation Exercise. *Bulletin of the Seismological Society of America*, *104*(5), 2387–2398. https://doi.org/10.1785/0120140138

Boore, D. M., Joyner, W. B., & Fumal, T. E. (1997). Equations for Estimating Horizontal Response Spectra and Peak Acceleration from Western North American Earthquakes: A Summary of Recent Work. *Seismological Research Letters*, *68*(1), 128–153. https://doi.org/10.1785/gssrl.68.1.128

Bradley, B. A. (2013). Practice-oriented estimation of the seismic demand hazard using ground motions at few intensity levels. *Earthquake Engineering & Structural Dynamics*, *42*(14), 2167–2185. https://doi.org/https://doi.org/10.1002/eqe.2319

Bradley, B. A. (2025). Benefits of site-specific hazard analysis for seismic design in New Zealand: Revisited. *Bulletin of the New Zealand Society for Earthquake Engineering*, *58*(4), 231–240. https://doi.org/10.5459/bnzsee.1705

Bradley, B. A., Burks, L. S., & Baker, J. W. (2015). Ground motion selection for simulation-based seismic hazard and structural reliability assessment. *Earthquake Engineering & Structural Dynamics*, *44*(13), 2321–2340. https://doi.org/https://doi.org/10.1002/eqe.2588

Bradley, B. A., & Lee, D. S. (2010). Component correlations in structure-specific seismic loss estimation. *Earthquake Engineering & Structural Dynamics*, *39*(3), 237–258. https://doi.org/https://doi.org/10.1002/eqe.937

Bray, J. D., & Travasarou, T. (2007). Simplified Procedure for Estimating Earthquake-Induced Deviatoric Slope Displacements. *Journal of Geotechnical and Geoenvironmental Engineering*, *133*(4), 381–392. https://doi.org/doi:10.1061/(ASCE)1090-0241(2007)133:4(381)

Chiou, B. S. J., & Youngs, R. R. (2014). Update of the Chiou and Youngs NGA Model for the Average Horizontal Component of Peak Ground Motion and Response Spectra. *Earthquake Spectra*, *30*(3), 1117–1153. https://doi.org/10.1193/072813eqs219m

Convertito, V., Emolo, A., & Zollo, A. (2006). Seismic-Hazard Assessment for a Characteristic Earthquake Scenario: An Integrated Probabilistic–Deterministic Method. *Bulletin of the Seismological Society of America*, *96*(2), 377–391. https://doi.org/10.1785/0120050024

Cornell, C. A. (1968). Engineering seismic risk analysis. *Bulletin of the Seismological Society of America*, *58*(5), 1583–1606. https://doi.org/10.1785/bssa0580051583

Cornell, C. A. (1971). Probabilistic Analysis of Damage to Structures under Seismic Loads.

Cornell, C. A. (2005). *On earthquake record selection for nonlinear dynamic analysis.* Luis Esteva Symposium, Mexico.

Cornell, C. A., Jalayer, F., Hamburger, R. O., & Foutch, D. A. (2002). Probabilistic Basis for 2000 SAC Federal Emergency Management Agency Steel Moment Frame Guidelines. *Journal of Structural Engineering-asce*, *128*, 526–533.

Cornell, C. A., & Krawinkler, H. (2000). Progress and challenges in seismic performance assessment. *PEER Center News*, *3(2)*.

Cremen, G., & Baker, J. W. (2018). Quantifying the benefits of building instruments to FEMA P-58 rapid post-earthquake damage and loss predictions. *Engineering Structures*, *176*, 243–253. https://doi.org/https://doi.org/10.1016/j.engstruct.2018.08.017

Cremen, G., & Baker, J. W. (2019). A Methodology for Evaluating Component-Level Loss Predictions of the FEMA P-58 Seismic Performance Assessment Procedure. *Earthquake Spectra*, *35*(1), 193–210. https://doi.org/10.1193/031618eqs061m

Cremen, G., Galasso, C., & McCloskey, J. (2022). A Simulation-Based Framework for Earthquake Risk-Informed and People-Centered Decision Making on Future Urban Planning. *Earth's Future*, *10*(1), e2021EF002388. https://doi.org/10.1029/2021EF002388

Derras, B., Bard, P. Y., & Cotton, F. (2014). Towards fully data driven ground-motion prediction models for Europe. *Bulletin of Earthquake Engineering*, *12*(1), 495–516. https://doi.org/10.1007/s10518-013-9481-0

Dhanya, J., & Raghukanth, S. T. G. (2018). Ground Motion Prediction Model Using Artificial Neural Network. *Pure and Applied Geophysics*, *175*(3), 1035–1064. https://doi.org/10.1007/s00024-017-1751-3

Douglas, J. (2022). *Ground motion prediction equations 1964-2021*. http://www.gmpe.org.uk.

Douglas, J. (2024). *Ground motion prediction equations 1964-2023 (incomplete)*. figshare. https://figshare.com/articles/online_resource/Ground_motion_prediction_equations_1964-2023_incomplete_/26827750

Douglas, J., Akkar, S., Ameri, G., Bard, P.-Y., Bindi, D., Bommer, J. J., Bora, S. S., Cotton, F., Derras, B., Hermkes, M., Kuehn, N. M., Luzi, L., Massa, M., Pacor, F., Riggelsen, C., Sandıkkaya, M. A., Scherbaum, F., Stafford, P. J., & Traversa, P. (2014). Comparisons among the five ground-motion models developed using RESORCE for the prediction of response spectral accelerations due to earthquakes in Europe and the Middle East. *Bulletin of Earthquake Engineering*, *12*(1), 341–358. https://doi.org/10.1007/s10518-013-9522-8

Douglas, J., Aldama-Bustos, G., Tallett-Williams, S., Daví, M., & Tromans, I. J. (2024). Ground-motion models for earthquakes occurring in the United Kingdom. *Bulletin of Earthquake Engineering*, *22*(9), 4265–4302. https://doi.org/10.1007/s10518-024-01943-8

Douglas, J., & Aochi, H. (2008). A Survey of Techniques for Predicting Earthquake Ground Motions for Engineering Purposes. *Surveys in Geophysics*, *29*(3), 187. https://doi.org/10.1007/s10712-008-9046-y

Douglas, J., Bertil, D., Roulle, A., Dominique, P., & Jousset, P. (2006). A preliminary investigation of strong-motion data from the French Antilles. *Journal of Seismology*, *10*(3), 271–299. https://doi.org/10.1007/s10950-006-9016-0

Douglas, J., & Edwards, B. (2016). Recent and future developments in earthquake ground motion estimation. *Earth-Science Reviews*, *160*, 203–219. https://doi.org/10.1016/j.earscirev.2016.07.005

Du, A., & Padgett, J. E. (2020). Investigation of multivariate seismic surrogate demand modeling for multi-response structural systems. *Engineering Structures*, *207*, 110210. https://doi.org/https://doi.org/10.1016/j.engstruct.2020.110210

Du, A., Padgett, J. E., & Shafieezadeh, A. (2020). Influence of intensity measure selection on simulation-based regional seismic risk assessment. *Earthquake Spectra*, *36*(2), 647–672. https://doi.org/10.1177/8755293019891717

Elkady, A. (2022). FM-2D - open-source platform for the 2-dimensional numerical modeling and seismic analysis of buildings. *SoftwareX*, *17*, 100927. https://doi.org/https://doi.org/10.1016/j.softx.2021.100927

Erdik, M. (2017). Earthquake risk assessment. *Bulletin of Earthquake Engineering*, *15*. https://doi.org/10.1007/s10518-017-0235-2

Esteva, L. (1967). Criterios para la construcción de espectros para diseño sísmico. 3er Simposio Panamericano de Estructuras, Caracas, Venezuela,

Esteva, L., & Rosenblueth, E. (1964). Spectra of earthquakes at moderate and large distances. *Soc. Mex. de Ing. Sismica, Mexico*, *2*(1).

European Committee for Standardization (CEN). (2004). Eurocode 8: Design of structures for earthquake resistance – Part 1: General rules, seismic actions and rules for buildings. In (Vol. BS EN 1998-1). Brussels, Belgium: European Committee for Standardization (CEN).

Fajfar, P. (1999). Capacity spectrum method based on inelastic demand spectra. *Earthquake Engineering & Structural Dynamics*, *28*(9), 979–993. https://doi.org/https://doi.org/10.1002/(SICI)1096-9845(199909)28:9<979::AID-EQE850>3.0.CO;2-1

FEMA, F. E. M. A. (2012). *Seismic Performance Assessment of Buildings, Volume 1, Methodology* (FEMA P-58-1).

Fiorentino, G., Sabetta, F., & De Risi, R. (2025). SIGMA: a new tool for the simulation of spectrum-compatible earthquake ground motions. *Natural Hazards*. https://doi.org/10.1007/s11069-025-07387-w

Franchin, P., Cavalieri, F., & Pinto, P. (2012). *Validating IM-based methods for probabilistic seismic performance assessment with higher-level non-conditional simulation* Proceedings of the Fifteenth World Conference on Earthquake Engineering,

Freddi, F., Padgett, J. E., & Dall'Asta, A. (2017). Probabilistic seismic demand modeling of local level response parameters of an RC frame. *Bulletin of Earthquake Engineering*, *15*(1), 1–23. https://doi.org/10.1007/s10518-016-9948-x

Ge, F.-W., Tong, M.-N., & Zhao, Y.-G. (2021). A structural demand model for seismic fragility analysis based on three-parameter lognormal distribution. *Soil Dynamics and Earthquake Engineering*, *147*, 106770. https://doi.org/https://doi.org/10.1016/j.soildyn.2021.106770

Gehl, P., Douglas, J., & Seyedi, D. M. (2015). Influence of the number of dynamic analyses on the accuracy of structural response estimates. *Earthquake Spectra*, *31*(1), 97–113. https://doi.org/10.1193/102912EQS320M

Gehl, P., Seyedi, D. M., & Douglas, J. (2013). Vector-valued fragility functions for seismic risk evaluation. *Bulletin of Earthquake Engineering*, *11*(2), 365–384. https://doi.org/10.1007/s10518-012-9402-7

Gentile, R., & Galasso, C. (2020). Gaussian process regression for seismic fragility assessment of building portfolios. *Structural Safety*, *87*, 101980. https://doi.org/https://doi.org/10.1016/j.strusafe.2020.101980

Gentile, R., & Galasso, C. (2022). Surrogate probabilistic seismic demand modelling of inelastic single-degree-of-freedom systems for efficient earthquake risk applications. *Earthquake Engineering & Structural Dynamics*, *51*(2), 492–511. https://doi.org/https://doi.org/10.1002/eqe.3576

Ghosh, J., Padgett, J. E., & Dueñas-Osorio, L. (2013). Surrogate modeling and failure surface visualization for efficient seismic vulnerability assessment of highway bridges. *Probabilistic Engineering Mechanics*, *34*, 189–199. https://doi.org/https://doi.org/10.1016/j.probengmech.2013.09.003

Ghosh, J., Rokneddin, K., Padgett, J. E., & Dueñas-Osorio, L. (2014). Seismic Reliability Assessment of Aging Highway Bridge Networks with Field Instrumentation Data and Correlated Failures, I: Methodology. *Earthquake Spectra*, *30*(2), 795–817. https://doi.org/10.1193/040512eqs155m

Giovenale, P., Cornell, C. A., & Esteva, L. (2004). Comparing the adequacy of alternative ground motion intensity measures for the estimation of structural responses. *Earthquake Engineering & Structural Dynamics*, *33*(8), 951–979. https://doi.org/https://doi.org/10.1002/eqe.386

Gkimprixis, A., Douglas, J., & Tubaldi, E. (2021). Seismic risk management through insurance and its sensitivity to uncertainty in the hazard model. *Natural Hazards*, *108*(2), 1629–1657. https://doi.org/10.1007/s11069-021-04748-z

Goda, K., & Tesfamariam, S. (2015). Multi-variate seismic demand modelling using copulas: Application to non-ductile reinforced concrete frame in Victoria, Canada. *Structural Safety*, *56*, 39–51. https://doi.org/https://doi.org/10.1016/j.strusafe.2015.05.004

Graves, R., Jordan, T. H., Callaghan, S., Deelman, E., Field, E., Juve, G., Kesselman, C., Maechling, P., Mehta, G., Milner, K., Okaya, D., Small, P., & Vahi, K. (2011). CyberShake: A Physics-Based Seismic Hazard Model for Southern California. *Pure and Applied Geophysics*, *168*(3), 367–381. https://doi.org/10.1007/s00024-010-0161-6

Gupta, A., & Krawinkler, H. (1999). *Seismic demands for performance evaluation steel moment resisting frame structures*. John A Blume Earthquake Engineering Center Technical Report 132.

Gutenberg, B., & Richter, C. F. (1944). Frequency of earthquakes in California*. *Bulletin of the Seismological Society of America*, *34*(4), 185–188. https://doi.org/10.1785/bssa0340040185

Hancock, J., & Bommer, J. J. (2005). The effective number of cycles of earthquake ground motion. *Earthquake Engineering & Structural Dynamics*, *34*(6), 637–664. https://doi.org/https://doi.org/10.1002/eqe.437

Housner, G. W. (1952). Intensity of Ground Motion during Strong Earthquakes.

Hutchings, L., Ioannidou, E., Foxall, W., Voulgaris, N., Savy, J., Kalogeras, I., Scognamiglio, L., & Stavrakakis, G. (2007). A physically based strong ground-motion prediction methodology; application to PSHA and the 1999 Mw = 6.0 Athens earthquake. *Geophysical Journal International*, *168*(2), 659–680. https://doi.org/10.1111/j.1365-246X.2006.03178.x

Ibarra, L. F., Medina, R. A., & Krawinkler, H. (2005). Hysteretic models that incorporate strength and stiffness deterioration. *Earthquake Engineering & Structural Dynamics*, *34*(12), 1489–1511. https://doi.org/https://doi.org/10.1002/eqe.495

Izquierdo-Horna, L., Zevallos, J., & Yepez, Y. (2022). An integrated approach to seismic risk assessment using random forest and hierarchical analysis: Pisco, Peru. *Heliyon*, *8*(10), e10926. https://doi.org/https://doi.org/10.1016/j.heliyon.2022.e10926

Jalayer, F., & Beck, J. L. (2008). Effects of two alternative representations of ground-motion uncertainty on probabilistic seismic demand assessment of structures. *Earthquake Engineering & Structural Dynamics*, *37*(1), 61–79. https://doi.org/https://doi.org/10.1002/eqe.745

Jalayer, F., Beck, J. L., & Zareian, F. (2012). Analyzing the Sufficiency of Alternative Scalar and Vector Intensity Measures of Ground Shaking Based on Information Theory. *Journal of Engineering Mechanics*, *138*(3), 307–316. https://doi.org/doi:10.1061/(ASCE)EM.1943-7889.0000327

Jalayer, F., & Cornell, C. A. (2009). Alternative non-linear demand estimation methods for probability-based seismic assessments. *Earthquake Engineering & Structural Dynamics*, *38*(8), 951–972. https://doi.org/https://doi.org/10.1002/eqe.876

Jayaram, N., & Baker, J. W. (2010). Efficient sampling and data reduction techniques for probabilistic seismic lifeline risk assessment. *Earthquake Engineering & Structural Dynamics*, *39*(10), 1109–1131. https://doi.org/https://doi.org/10.1002/eqe.988

Jayaram, N., Shome, N., & Rahnama, M. (2012). Development of earthquake vulnerability functions for tall buildings. *Earthquake Engineering & Structural Dynamics*, *41*(11), 1495–1514. https://doi.org/https://doi.org/10.1002/eqe.2231

Jiang, J., El Naggar, H., Xu, C., Chen, G., & Du, X. (2023). Seismic fragility analysis for subway station considering varying ground motion ensembles. *Soil Dynamics and Earthquake Engineering*, *165*, 107705. https://doi.org/https://doi.org/10.1016/j.soildyn.2022.107705

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/doi:10.1126/science.aaa8415

Kang, C., Kim, T., Kwon, O.-S., & Song, J. (2023). Deep neural network-based regional seismic loss assessment considering correlation between EDP residuals of building structures. *Earthquake Engineering & Structural Dynamics*, *52*(11), 3414–3434. https://doi.org/https://doi.org/10.1002/eqe.3775

Katsanos, E. I., Sextos, A. G., & Manolis, G. D. (2010). Selection of earthquake ground motion records: A state-of-the-art review from a structural engineering

perspective. *Soil Dynamics and Earthquake Engineering*, *30*(4), 157–169. https://doi.org/https://doi.org/10.1016/j.soildyn.2009.10.005

Khosravikia, F., & Clayton, P. (2020). Updated evaluation metrics for optimal intensity measure selection in probabilistic seismic demand models. *Engineering Structures*, *202*, 109899. https://doi.org/https://doi.org/10.1016/j.engstruct.2019.109899

Khosravikia, F., & Clayton, P. (2021). Machine learning in ground motion prediction. *Computers & Geosciences*, *148*, 104700. https://doi.org/https://doi.org/10.1016/j.cageo.2021.104700

Khosravikia, F., Clayton, P., & Nagy, Z. (2018). Artificial neural network-based framework for developing ground-motion models for natural and induced earthquakes in Oklahoma, Kansas, and Texas. *Seismological Research Letters*, *90*(2A), 604–613. https://doi.org/10.1785/0220180218

Kim, J., & Kim, T. (2025). Efficient seismic fragility analysis considering uncertainties in structural systems and ground motions. *Earthquake Engineering & Structural Dynamics*, *54*(1), 206–226. https://doi.org/https://doi.org/10.1002/eqe.4254

Kohrangi, M., Vamvatsikos, D., & Bazzurro, P. (2016). Implications of Intensity Measure Selection for Seismic Loss Assessment of 3-D Buildings. *Earthquake Spectra*, *32*(4), 2167–2189. https://doi.org/10.1193/112215eqs177m

Kong, Q., Trugman, D. T., Ross, Z. E., Bianco, M. J., Meade, B. J., & Gerstoft, P. (2018). Machine Learning in Seismology: Turning Data into Insights. *Seismological Research Letters*, *90*(1), 3–14. https://doi.org/10.1785/0220180259

Kowsari, M., Halldorsson, B., Snaebjornsson, J. T., & Jonsson, S. (2021). Effects of different empirical ground motion models on seismic hazard maps for North Iceland. *Soil Dynamics and Earthquake Engineering*, *148*, 106513. https://doi.org/ARTN 106513 10.1016/j.soildyn.2020.106513

Kramer, S. L. (1996). *Geotechnical Earthquake Engineering*. Prentice-Hall.

Kubo, H., Kunugi, T., Suzuki, W., Suzuki, S., & Aoi, S. (2020). Hybrid predictor for ground-motion intensity with machine learning and conventional ground motion prediction equation. *Scientific Reports*, *10*(1), 11871. https://doi.org/10.1038/s41598-020-68630-x

Kubo, H., Naoi, M., & Kano, M. (2024). Recent advances in earthquake seismology using machine learning. *Earth, Planets and Space*, *76*(1), 36. https://doi.org/10.1186/s40623-024-01982-0

Kwong, N. S., Chopra, A. K., & McGuire, R. K. (2015a). Evaluation of ground motion selection and modification procedures using synthetic ground motions. *Earthquake Engineering & Structural Dynamics*, *44*(11), 1841–1861. https://doi.org/https://doi.org/10.1002/eqe.2558

Kwong, N. S., Chopra, A. K., & McGuire, R. K. (2015b). A framework for the evaluation of ground motion selection and modification procedures. *Earthquake Engineering & Structural Dynamics*, *44*(5), 795–815. https://doi.org/https://doi.org/10.1002/eqe.2502

Lanzano, G., Luzi, L., Pacor, F., Felicetta, C., Puglia, R., Sgobba, S., & D'Amico, M. (2019). A Revised Ground-Motion Prediction Model for Shallow Crustal Earthquakes in Italy. *Bulletin of the Seismological Society of America*, *109*(2), 525–540. https://doi.org/10.1785/0120180210

Lavrentiadis, G., Abrahamson, N. A., Nicolas, K. M., Bozorgnia, Y., Goulet, C. A., Babič, A., Macedo, J., Dolšek, M., Gregor, N., Kottke, A. R., Lacour, M., Liu, C., Meng, X., Phung, V.-B., Sung, C.-H., & Walling, M. (2023). Overview and introduction to development of non-ergodic earthquake ground-motion models. *Bulletin of Earthquake Engineering*, *21*(11), 5121–5150. https://doi.org/10.1007/s10518-022-01485-x

Li, S., Farrar, C., & Yang, Y. (2023). Efficient regional seismic risk assessment via deep generative learning of surrogate models. *Earthquake Engineering & Structural Dynamics*, *52*(11), 3435–3454. https://doi.org/https://doi.org/10.1002/eqe.3849

Li, W., Huang, Y., & Xie, Z. (2022). Machine Learning-Based Probabilistic Seismic Demand Model of Continuous Girder Bridges. *Advances in Civil Engineering*, *2022*, 3867782. https://doi.org/10.1155/2022/3867782

Lignos, D. G., Hartloper, A. R., Elkady, A., Deierlein, G. G., & Hamburger, R. (2019). Proposed Updates to the ASCE 41 Nonlinear Modeling Parameters for Wide-Flange Steel Columns in Support of Performance-Based Seismic Engineering. *Journal of Structural Engineering*, *145*(9), 04019083. https://doi.org/doi:10.1061/(ASCE)ST.1943-541X.0002353

Liou, I., & Abrahamson, N. (2024). Framework for Aleatory Variability and Epistemic Uncertainty for the Ground-Motion Characterization Based on the Level of Simplification. *Bulletin of the Seismological Society of America*, *115*(1), 296–314. https://doi.org/10.1785/0120240141

Luco, N., & Cornell, C. A. (2007). Structure-Specific Scalar Intensity Measures for Near-Source and Ordinary Earthquake Ground Motions. *Earthquake Spectra*, *23*(2), 357–392. https://doi.org/10.1193/1.2723158

Lupoi, G., Franchin, P., Lupoi, A., & Pinto, P. E. (2006). Seismic Fragility Analysis of Structural Systems. *Journal of Engineering Mechanics*, *132*(4), 385–395. https://doi.org/doi:10.1061/(ASCE)0733-9399(2006)132:4(385)

Mackie, K., & Stojadinovic, B. (2003). *Seismic Demands for Performance-Based Design of Bridges, PEER Report 2003-16*.

Mackie, K., & Stojadinovic, B. (2005). Comparison of Incremental Dynamic, Cloud, and Stripe Methods for Computing Probabilistic Seismic Demand Models. In *Structures Congress 2005* (pp. 1–11). https://doi.org/doi:10.1061/40753(171)184

Malhotra, P. K. (2014). *Cost of Uncertainty in Seismic Hazard*.

Mangalathu, S., Hwang, S.-H., Choi, E., & Jeon, J.-S. (2019). Rapid seismic damage evaluation of bridge portfolios using machine learning techniques. *Engineering Structures*, *201*, 109785. https://doi.org/https://doi.org/10.1016/j.engstruct.2019.109785

Mazzoni, S., McKenna, F., Scott, M. H., & Fenves, G. L. (2006). OpenSees command language manual. *Pacific earthquake engineering research (PEER) center*, *264*(1), 137–158.

McGuire, R. K. (2004). *Seismic hazard and risk analysis*. Earthquake Engineering Research Institute.

McGuire, R. K. (2008). Probabilistic seismic hazard analysis: Early history. *Earthquake Engineering & Structural Dynamics*, *37*(3), 329–338. https://doi.org/https://doi.org/10.1002/eqe.765

McKenna, F. (2011). OpenSees: A Framework for Earthquake Engineering Simulation. *Computing in Science & Engineering*, *13*(4), 58–66. https://doi.org/10.1109/MCSE.2011.66

Medel-Vera, C., & Ji, T. (2016). Seismic probabilistic risk analysis based on stochastic simulation of accelerograms for nuclear power plants in the UK. *Progress in Nuclear Energy*, *91*, 373–388. https://doi.org/https://doi.org/10.1016/j.pnucene.2016.06.005

Meirova, T., Shapira, A., & Eppelbaum, L. (2018). PSHA in Israel by using the synthetic ground motions from simulated seismicity: the modified SvE procedure. *Journal of Seismology*, *22*(5), 1095–1111. https://doi.org/10.1007/s10950-018-9752-y

Minas, S., Chandler, R. E., & Rossetto, T. (2018). BEA: An efficient Bayesian emulation-based approach for probabilistic seismic response. *Structural Safety*, *74*, 32–48. https://doi.org/https://doi.org/10.1016/j.strusafe.2018.04.002

Miranda, E., & Aslani, H. (2003). *Probabilistic response assessment for building-specific loss estimation* (PEER Report 2003-03).

Moehle, J. P., & Deierlein, G. G. (2004). *A framework methodology for performance-based earthquake engineering* 13th World Conference on Earthquake Engineering, Vancouver, B.C., Canada.

Monterrubio-Velasco, M., Callaghan, S., Modesto, D., Carrasco, J. C., Badia, R. M., Pallares, P., Vázquez-Novoa, F., Quintana-Ortí, E. S., Pienkowska, M., & de la Puente, J. (2024). A machine learning estimator trained on synthetic data for real-time earthquake ground-shaking predictions in Southern California. *Communications Earth & Environment*, *5*(1), 258. https://doi.org/10.1038/s43247-024-01436-1

Musson, R. M. W. (2000). The use of Monte Carlo simulations for seismic hazard assessment in the UK. *Annals of Geophysics*, *43*(1), 1–9. https://doi.org/https://doi.org/10.4401/ag-3617

Newmark, N. M., & Hall, W. J. (1982). Earthquake Spectra and Design. *Earthquake Engineering Research Institute Monograph Series*.

Nielson, B. G. (2005). *Analytical fragility curves for highway bridges in moderate seismic zones*. Georgia Institute of Technology.

O'Reilly, G. J. (2021). Seismic intensity measures for risk assessment of bridges. *Bulletin of Earthquake Engineering*, *19*(9), 3671–3699. https://doi.org/10.1007/s10518-021-01114-z

O'Reilly, G. J., & Monteiro, R. (2019). Probabilistic models for structures with bilinear demand-intensity relationships. *Earthquake Engineering & Structural Dynamics*, *48*(2), 253–268. https://doi.org/https://doi.org/10.1002/eqe.3135

Ohtori, Y., Christenson, R. E., Spencer, B. F., & Dyke, S. J. (2004). Benchmark Control Problems for Seismically Excited Nonlinear Buildings. *Journal of Engineering Mechanics*, *130*(4), 366–385. https://doi.org/doi:10.1061/(ASCE)0733-9399(2004)130:4(366)

Oprsal, I., & Zahradnik, J. (2002). Three-dimensional finite difference method and hybrid modeling of earthquake ground motion. *Journal of Geophysical Research: Solid Earth*, *107*(B8), ESE 2–1–ESE 2–16. https://doi.org/https://doi.org/10.1029/2000JB000082

Padgett, J. E., Nielson, B. G., & DesRoches, R. (2008). Selection of optimal intensity measures in probabilistic seismic demand models of highway bridge portfolios.

Earthquake Engineering & Structural Dynamics, *37*(5), 711–725. https://doi.org/https://doi.org/10.1002/eqe.782

Papadopoulos, A. N., Vamvatsikos, D., & Kazantzi, A. K. (2019). Development and Application of FEMA P-58 Compatible Story Loss Functions. *Earthquake Spectra*, *35*(1), 95–112. https://doi.org/10.1193/102417eqs222m

Park, Y. J., Ang, A. H. S., & Wen, Y. K. (1985). Seismic Damage Analysis of Reinforced Concrete Buildings. *Journal of Structural Engineering*, *111*(4), 740–757. https://doi.org/doi:10.1061/(ASCE)0733-9445(1985)111:4(740)

Pradhan, B., & Lee, S. (2009). Landslide risk analysis using artificial neural network model focussing on different training sites. *International Journal of Physical Sciences*, *4*, 1–15.

Qian, J., & Dong, Y. (2022). Surrogate-assisted seismic performance assessment incorporating vine copula captured dependence. *Engineering Structures*, *257*, 114073. https://doi.org/https://doi.org/10.1016/j.engstruct.2022.114073

Rasmussen, C. E., & Williams, C. K. I. (2005). *Gaussian Processes for Machine Learning*. The MIT Press. https://doi.org/10.7551/mitpress/3206.001.0001

Reed, J. W., Anderson, N., Chokshi, N. C., Kennedy, R. P., Metevia, W. J., Ostrom, D. K., & Stevenson, J. D. (1988). *A criterion for determining exceedance of the operating basis earthquake.* (EPRI Report NP-5930).

Reiter, L. (1990). *Earthquake Hazard Analysis: Issues and Insights*. Columbia University Press.

Rossetto, T., Gehl, P., Minas, S., Galasso, C., Duffour, P., Douglas, J., & Cook, O. (2016). FRACAS: A capacity spectrum approach for seismic fragility assessment including record-to-record variability. *Engineering Structures*, *125*, 337–348. https://doi.org/https://doi.org/10.1016/j.engstruct.2016.06.043

Rubner, Y., Tomasi, C., & Guibas, L. J. (1998, 7–7 Jan. 1998). *A metric for distributions with applications to image databases* Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271),

Rudman, A. (2026). *FM2D_SAC_LA3_SMRF* (Zenodo. https://doi.org/10.5281/zenodo.19683348

Rudman, A., Douglas, J., & Tubaldi, E. (2024). The assessment of probabilistic seismic risk using ground-motion simulations via a Monte Carlo approach. *Natural Hazards*, *120*(7), 6833–6852. https://doi.org/10.1007/s11069-024-06497-1

Rudman, A., Tubaldi, E., Douglas, J., & Scozzese, F. (2024). The impact of the choice of intensity measure and seismic demand model on seismic risk estimates with respect to an unconditional benchmark. *Earthquake Engineering & Structural Dynamics*, *53*(14), 4183–4202. https://doi.org/https://doi.org/10.1002/eqe.4208

Ruiz-García, J., & Miranda, E. (2010). Probabilistic estimation of residual drift demands for seismic assessment of multi-story framed buildings. *Engineering Structures*, *32*(1), 11–20. https://doi.org/https://doi.org/10.1016/j.engstruct.2009.08.010

Sabetta, F., & Pugliese, A. (1996). Estimation of response spectra and simulation of nonstationary ground motions. *Bulletin of the Seismological Society of America*, *86*. https://doi.org/10.1785/BSSA0860020337

Sabetta, F., Pugliese, A., Fiorentino, G., Lanzano, G., & Luzi, L. (2021). Simulation of non-stationary stochastic ground motions based on recent Italian

earthquakes. *Bulletin of Earthquake Engineering*, *19*(9), 3287–3315. https://doi.org/10.1007/s10518-021-01077-1

SAC. (1994). *SAC Steel Project*. http://www.sacsteel.org/

Sarma, S. K., & Yang, K. S. (1987). An evaluation of strong motion records and a new parameter A95. *Earthquake Engineering & Structural Dynamics*, *15*(1), 119–132. https://doi.org/https://doi.org/10.1002/eqe.4290150109

Scherbaum, F., Schmedes, J., & Cotton, F. (2004). On the Conversion of Source-to-Site Distance Measures for Extended Earthquake Source Models. *Bulletin of the Seismological Society of America*, *94*(3), 1053–1069. https://doi.org/10.1785/0120030055

Scozzese, F., Tubaldi, E., & Dall'Asta, A. (2020). Assessment of the effectiveness of Multiple-Stripe Analysis by using a stochastic earthquake input model. *Bulletin of Earthquake Engineering*, *18*(7), 3167–3203. https://doi.org/10.1007/s10518-020-00815-1

Shapira, A., & van Eck, T. (1993). Synthetic uniform-hazard site specific response spectrum. *Natural Hazards*, *8*(3), 201–215. https://doi.org/10.1007/BF00690908

Shinozuka, M. (1983). Basic Analysis of Structural Safety. *Journal of Structural Engineering*, *109*(3), 721–740. https://doi.org/doi:10.1061/(ASCE)0733-9445(1983)109:3(721)

Shome, N. (1999). *Probabilistic seismic demand analysis of nonlinear structures*. Stanford University.

Silva, W. J., Abrahamson, N., Toro, G., & Constantino, C. (1996). *Description and validation of the stochastic ground motion model*.

Skiadopoulos, A., Elkady, A., & Lignos, D. G. (2021). Proposed Panel Zone Model for Seismic Design of Steel Moment-Resisting Frames. *Journal of Structural Engineering*, *147*(4), 04021006. https://doi.org/doi:10.1061/(ASCE)ST.1943-541X.0002935

Snoek, J., Larochelle, H., & Adams, R. (2012). Practical Bayesian Optimization of Machine Learning Algorithms. *Advances in Neural Information Processing Systems*, *4*.

Soleimani, F., & Hajializadeh, D. (2022). State-of-the-Art Review on Probabilistic Seismic Demand Models of Bridges: Machine-Learning Application. *Infrastructures*, *7*(5), 64. https://doi.org/doi:10.3390/infrastructures7050064

Soleimani, F., & Liu, X. (2022). Artificial neural network application in predicting probabilistic seismic demands of bridge components. *Earthquake Engineering & Structural Dynamics*, *51*(3), 612–629. https://doi.org/https://doi.org/10.1002/eqe.3582

Stephens, M. A. (1974). EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association*, *69*(347), 730–737. https://doi.org/10.2307/2286009

Stupazzini, M., Infantino, M., Allmann, A., & Paolucci, R. (2021). Physics-based probabilistic seismic hazard and loss assessment in large urban areas: A simplified application to Istanbul. *Earthquake Engineering & Structural Dynamics*, *50*(1), 99–115. https://doi.org/https://doi.org/10.1002/eqe.3365

Taghavi, S. M., E. (2003). *Response Assessment of Nonstructural Building Elements, PEER Report 2003-05*. B. University of California, CA.

The MathWorks Inc. (2023). *MATLAB*. In (Version 9.14.0.2254940 (R2023a) Update 2) The MathWorks Inc.,. https://www.mathworks.com

Tothong, P., & Luco, N. (2007). Probabilistic seismic demand analysis using advanced ground motion intensity measures. *Earthquake Engineering & Structural Dynamics*, *36*(13), 1837–1860. https://doi.org/https://doi.org/10.1002/eqe.696

Towashiraporn, P. (2004). *Building seismic fragilities using response surface metamodels* (Publication Number 3155352) [Ph.D., Georgia Institute of Technology]. ProQuest Dissertations & Theses Global. United States -- Georgia. https://www.proquest.com/dissertations-theses/building-seismic-fragilities-using-response/docview/305189251/se-2?accountid=10673

Tsioulou, A., & Galasso, C. (2018). Information theory measures for the engineering validation of ground-motion simulations. *Earthquake Engineering & Structural Dynamics*, *47*(4), 1095–1104. https://doi.org/https://doi.org/10.1002/eqe.3015

Tsioulou, A., Taflanidis, A. A., & Galasso, C. (2019). Validation of stochastic ground motion model modification by comparison to seismic demand of recorded ground motions. *Bulletin of Earthquake Engineering*, *17*(6), 2871–2898. https://doi.org/10.1007/s10518-019-00571-x

Tubaldi, E., Freddi, F., & Barbato, M. (2016). Probabilistic seismic demand model for pounding risk assessment. *Earthquake Engineering & Structural Dynamics*, *45*(11), 1743–1758. https://doi.org/https://doi.org/10.1002/eqe.2725

UBC. (1994). *Structural Engineering Design Provisions* (Uniform Building Code, Issue. I. C. o. B. Officials.

Vamvatsikos, D., & Cornell, C. A. (2002). Incremental dynamic analysis. *Earthquake Engineering & Structural Dynamics*, *31*(3), 491–514. https://doi.org/https://doi.org/10.1002/eqe.141

Vamvatsikos, D., & Cornell, C. A. (2005). Developing efficient scalar and vector intensity measures for IDA capacity estimation by incorporating elastic spectral shape information. *Earthquake Engineering & Structural Dynamics*, *34*(13), 1573–1600. https://doi.org/https://doi.org/10.1002/eqe.496

Wang, Q. a., Wu, Z., & Liu, S. (2018). Multivariate Probabilistic Seismic Demand Model for the Bridge Multidimensional Fragility Analysis. *KSCE Journal of Civil Engineering*, *22*(9), 3443–3451. https://doi.org/10.1007/s12205-018-0414-y

Weatherill, G., & Lilienkamp, H. (2023). Capturing Directivity in Probabilistic Seismic Hazard Analysis for New Zealand: Challenges, Implications, and a Machine Learning Approach for Implementation. *Bulletin of the Seismological Society of America*, *114*(1), 373–398. https://doi.org/10.1785/0120230161

Whittaker, A., Atkinson, G., Baker, J., Bray, J., Grant, D., Hamburger, R., Haselton, C., & Somerville, P. (2011). Selecting and scaling earthquake ground motions for performing response-history analyses.

Xie, Y. (2025). Deep Learning in Earthquake Engineering: A Comprehensive Review. *ASCE OPEN: Multidisciplinary Journal of Civil Engineering*, *3*(1), 03125001. https://doi.org/doi:10.1061/AOMJAH.AOENG-0080

Xie, Y. Z., Sichani, M. E., Padgett, J. E., & DesRoches, R. (2020). The promise of implementing machine learning in earthquake engineering: A state-of-the-art review. *Earthquake Spectra*, *36*(4), 1769–1801. https://doi.org/10.1177/8755293020919419

Yamamoto, Y., & Baker, J. W. (2013). Stochastic Model for Earthquake Ground Motion Using Wavelet Packets. *Bulletin of the Seismological Society of America*, *103*(6), 3044–3056. https://doi.org/10.1785/0120120312

Yi, S.-r., & Taflanidis, A. A. (2025). Multi-output stochastic emulation with applications to seismic response correlation estimation. *Structural Safety*, *115*, 102578. https://doi.org/https://doi.org/10.1016/j.strusafe.2025.102578

Youngs, R. R., Chiou, S. J., Silva, W. J., & Humphrey, J. R. (1997). Strong Ground Motion Attenuation Relationships for Subduction Zone Earthquakes. *Seismological Research Letters*, *68*(1), 58–73. https://doi.org/10.1785/gssrl.68.1.58

Zhong, K., Navarro, J. G., Govindjee, S., & Deierlein, G. G. (2023). Surrogate modeling of structural seismic response using probabilistic learning on manifolds. *Earthquake Engineering & Structural Dynamics*, *52*(8), 2407–2428. https://doi.org/https://doi.org/10.1002/eqe.3839

Zhou, T., & Li, A.-Q. (2019). Seismic fragility assessment of highway bridges using D-vine copulas. *Bulletin of Earthquake Engineering*, *17*(2), 927–955. https://doi.org/10.1007/s10518-018-0474-x

Zolfaghari, M. R. (2015). Development of a synthetically generated earthquake catalogue towards assessment of probabilistic seismic hazard for Tehran. *Natural Hazards*, *76*(1), 497–514. https://doi.org/10.1007/s11069-014-1500-1

Appendix A

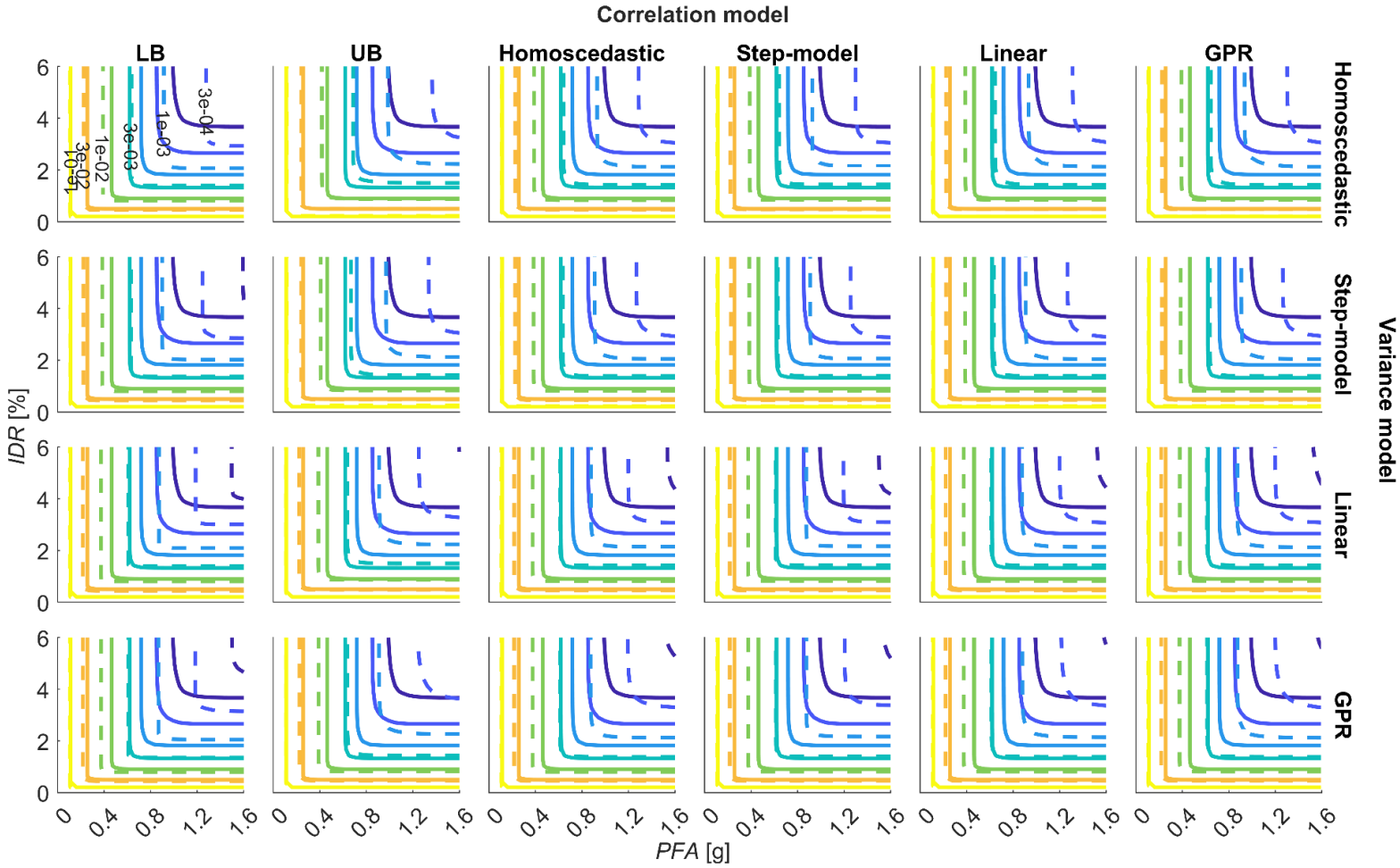


Figure A-1: Mean demand hazard estimates made with the linear median demand model and all associated covariance models via the conditional method (solid lines) compared against the unconditional benchmark (dashed lines).

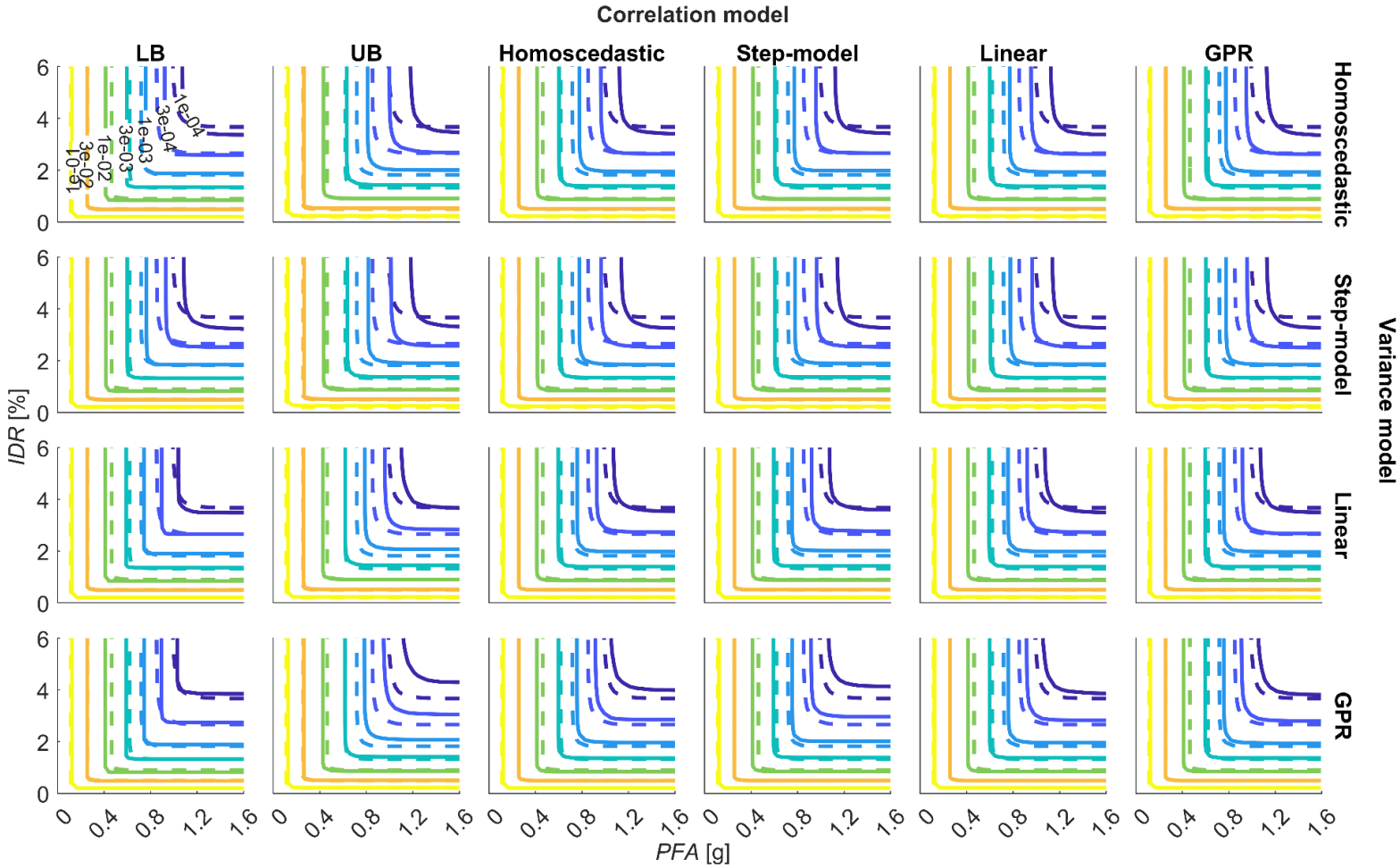


Figure A-2: Mean demand hazard estimates made with the Gaussian process regression median demand model and all associated covariance models via the conditional method (solid lines) compared against the unconditional benchmark (dashed lines).