Simulated ground motions for seismic risk assessment of structures

Thesis submitted to University College London for the degree of
Doctor of Philosophy

by

Alexandra Tsioulou

Department of Civil, Environmental & Geomatic Engineering

University College London

November 2018

Declaration

I, Alexandra Tsioulou, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed:

_____

Date:

_____

Abstract

The recent advances in computational efficiency and the scarcity/absence of recorded ground motions for specific seismicity scenarios have led to an increasing interest in the use of ground motion simulations for seismic hazard analysis, structural demand assessment through response-history analysis, and ultimately seismic risk assessment. Two categories of ground motion simulations, physics-based and stochastic site-based are considered in this study.

Physics-based ground motion simulations are generated using algorithms that solve the fault rupture and wave propagation problems and can be used for simulating past and future scenarios. Before being used with confidence, they need to be validated against records from past earthquakes. The first part of the study focuses on the development of rating/testing methodologies based on statistical and information theory measures for the validation of ground motion simulations obtained through an online platform for past earthquake events. The testing methodology is applied in a case-study utilising spectral-shape and duration-related intensity measures (IMs) as proxies for the nonlinear peak and cyclic structural response.

Stochastic site-based ground motion simulations model the time-history at a site by fitting a statistical process to ground motion records with known earthquake and site characteristics. To be used in practice, it is important that the output IMs from the developed time-histories are consistent with these prescribed at the site of interest, something that is not necessarily guaranteed by the current models. The second part of the study presents a computationally efficient framework that addresses the modification of stochastic ground motion models for given seismicity scenarios with a dual goal of matching target IMs for specific structures, while preserving desired trends in the physical characteristics of the resultant time-histories. The modification framework is extended to achieve a match to the full probability model of the target IMs. Finally, the proposed modification is validated by comparison to seismic demand of hazard-compatible recorded ground motions.

This study shows that ground motion simulation is a promising tool that can be used for many engineering applications.

Impact Statement

This work contributes to better seismic risk and resilience assessment prediction, that can be used in turn for different purposes. The impact of this study can be summarised as societal, economic and academic.

With respect to the societal impact, results of this study may ultimately be used by public and private organisations to develop emergency response plans and allocate available resources, evaluate cost-effective seismic retrofitting actions and risk mitigation strategies, and evaluate rapid damage-estimation algorithms for effective disaster response. This study is a contribution toward reduction of economic, life and business interruption losses due to earthquakes through advanced, improved, earthquake risk assessment models and tools.

With respect to the economic impact, the outputs of this study may provide the insurance and re-insurance industry, providers of basic services (e.g., civil protection and emergency managers), as well as multidisciplinary consultancy firms with tools to perform an improved characterisation of seismic risk that will enable them to operate more competitively in the global market.

Finally, this study addresses major intellectual challenges by going beyond the state-of-the-art of ground motion simulation and validation, and has resulted in four published journal papers with another one under review.

Acknowledgements

First and foremost, I would like to thank my principal adviser, Professor Carmine Galasso for his tremendous support, invaluable technical guidance and patience. It has been an honour and great pleasure to work with him throughout my PhD studies. He inspired me to become a better researcher and teacher and become inquisitive about my research subject. He encouraged me to expand my research horizons and acquire new skills. I am extremely grateful for his personal support, understanding and for always thinking what is best for me and I consider him a lifelong friend. I would also like to thank my co-adviser, Professor Alexandros Taflanidis for hosting me at Notre Dame, treating me as one of his own PhD students and providing me with the resources, tools and skills necessary to complete this research. He taught me new subjects, provided me with key technical insights and influenced my growth as a researcher. I consider myself very lucky for having the chance to work with such great researchers, teachers and persons like Carmine and Alexandros. Working with them has been a joyful, rewarding and stimulating experience that has transformed me not only as a researcher but as a person.

I would also like to thank my thesis examiners, Professors Agathoklis Giaralis and Fatemeh Jalayer for their insightful comments and feedback that improved the quality of this thesis.

I would like to thank Dr. Gordon Ross who helped me with the statistical part in Chapter 3 and colleagues from SCEC who shared data with me including Dr. Peng Zhong and Dr. Christine Goulet. I also want to thank Professor Tiziana Rossetto for her insightful comments at the beginning of this research.

Contents

List of Figures

List of Tables

Abbreviations

ASCE	**A**merican **S**ociety of **C**ivil **E**ngineers
BBP	**B**road**B**and **P**latform
CDF	**C**ummulative **D**istribution **F**unction
CoV	**C**oefficient **o**f **V**ariation
CRN	**C**ommon **R**andom **N**umbers
CSM	**C**omposite **S**ource **M**odel
DoE	**D**esign **o**f **E**xperiments
EDP	**E**ngineering **D**emand **P**arameter
ESD	**E**uropean **S**trong **M**otion **D**atabase
GMPE	**G**round **M**otion **P**rediction **E**quation
GMSV	**G**round **M**otion **S**imulation **V**alidation
HC	**H**azard **C**ompatibility
IM	**I**ntensity **M**easure
IMC	**I**ntensity **M**easure **C**ompatibility
KDE	**K**ernel **D**ensity **E**stimation
K-T	**K**anai-**T**ajimi
LHS	**L**atin **H**ypercube **S**ampling
MCS	**M**onte **C**arlo **S**imulation
MDoF	**M**ultiple **D**egrees **o**f **F**reedom
MIDR	**M**aximum **I**nterstory **D**rift **R**atio
MMI	**M**odified **M**ercalli **I**ntensity
NGA	**N**ext **G**eneration **A**ttenuation
NLDA	**N**on**L**inear **D**ynamic **A**nalysis

PBEE	**P**erformance-**B**ased **E**arthquake **E**ngineering
PDF	**P**robability **D**ensity **F**unction
PEER	**P**acific **E**arthquake **E**ngineering **R**esearch Centre
PFA	**P**eak **F**loor **A**cceleration
PGA	**P**eak **G**round **A**cceleration
PGD	**P**eak **G**round **D**isplacement
PGV	**P**eak **G**round **V**elocity
PSA	**P**seudo-**S**pectral **A**cceleration
PSHA	**P**robabilistic **S**eismic **H**azard **A**nalysis
SCEC	**S**outhern **C**alifornia **E**arthquake **C**entre
SDoF	**S**ingle **D**egree **o**f **F**reedom
SEISM	**S**oftware **E**nvironment for **I**ntegrated **S**eismic **M**odeling
SIMBAD	**S**elected **I**nput **M**otions for Displacement-**B**ased **A**ssessment and **D**esign
SVM	**S**upport **V**ector **M**achine
SYNTHESIS	**SYNTHE**tic **S**e**IS**mograms Database
TAG	**T**echnical **A**ctivity **G**roup

Chapter 1

Introduction

1.1 Background and motivation

Earthquake ground motions are an important component of seismic risk assessment and have many engineering applications. They are used to characterise the seismic hazard through the development of ground motion prediction equations (GMPEs), as input to response-history analysis to assess the induced damage to structures, and validate catastrophe insurance models. For many of these engineering applications, the complete acceleration time-history is represented by a single parameter called intensity measure (IM); the peak ground acceleration (PGA) or spectral accelerations at different structural periods T, $(S_a(T))$, are some commonly used IMs. Modern building codes prescribe the use of a target response spectrum, which is the output of site-specific seismic hazard assessment. This target response spectrum is then used either for response spectrum analysis of structures or to select a number of suitable ground motions as input to perform nonlinear dynamic analysis (NLDA). For response spectrum analysis, the spectral acceleration estimates are used whereas for applications involving dynamic analysis, the entire ground motion acceleration time-history is needed.

In the past decades, the growing fields of simulation-based seismic risk assessment (Au and Beck, 2003; Gidaris et al., 2016) and performance-based earthquake engineering (PBEE) (Bozorgnia and Bertero, 2004; Goulet et al., 2007), which addresses the entire spectrum of structural response, from linear to highly nonlinear to collapse, have changed the landscape related to ground motion modeling. These approaches require a large number of ground motion time-histories that are compatible with the various hazard levels as input to perform NLDA to determine performance of structures in terms of probability distributions of some engineering demand parameters (EDPs). In recognition of this need, there has been a bulk of research performed in the area of selecting suitable ground motions for performing NLDA of structures (Haselton, 2009; NIST, 2011).

In most advanced codes and state-of-the-art literature, the target response spectrum for structural or geotechnical analysis at a site of interest is identified through probabilistic seismic hazard analysis (PSHA). The outcome of PSHA is then used to develop a uniform hazard spectrum or conditional (mean) spectrum (Baker, 2011) to represent the design response spectrum. The design ground motions to be provided as input to NLDA, should match the target response spectrum, and also be compatible with the earthquake source characteristics relevant for the site of interest (these are also needed in defining a conditional spectrum). The earthquake source features in terms of magnitude (M), source-to-site distance (R) and epsilon (ϵ) that have the largest contribution to the hazard can be identified via disaggregation of PSHA at the structural period of interest (Bazzurro and Cornell, 2007) as presented in Figure 1.1. Therefore, suitable ground motions are those that are recorded during earthquake events of magnitude and source-to-site distance similar to the magnitude and distance of the seismic scenario that drives the hazard at the site where the structure is located. Moreover, other important source and site features, for example the tectonic environment and soil conditions at the site, need to be accounted for when selecting acceleration time-history records. The target response spectrum can alternatively be defined for a given earthquake scenario through the use of GMPEs e.g., scenario-based assessment

prescribed in FEMA P-58 (FEMA, 2012) guidelines for seismic performance assessment of structures. In this case, the earthquake source and site features are these of the scenario earthquake.
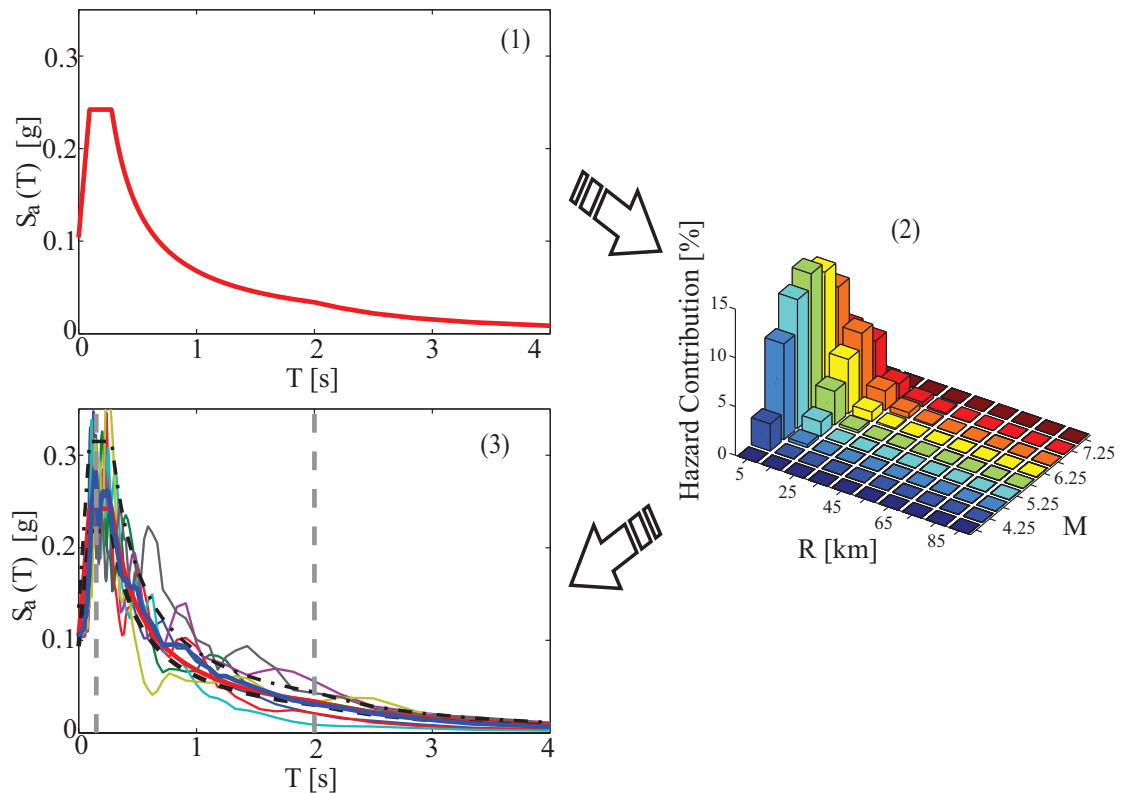
FIGURE 1.1: Steps to define design ground motions according to the hazard at the site from (1) to (3): target spectrum for the site of interest, hazard disaggregation for the spectral ordinate at the fundamental period of the structure, selection of a set of records compatible to disaggregation and matching the target spectrum at that same period (adapted from Galasso, 2010).

Despite the thousands of strong ground motion records available, there is an inherent scarcity or total absence of suitable natural records for some specific earthquake scenarios, for example large magnitude events on nearby faults (see Figure 1.2), as well as records that sample specific combination of source, path and site characteristics. Moreover, a significant number of ground motion records is required for the calibration of GMPEs, which is usually not possible to obtain in regions where earthquakes are infrequent like the stable continental regions of the Central and Eastern United States. For the purposes of PBEE, it is common engineering practice to modify recorded ground motions by scaling or spectrally matching them to a target spectrum. The former approach modifies the amplitude of a "seed"

ground motion record to match spectral acceleration values over a specified range of periods around a target period, while the latter modifies the frequency content of the ground motion record in order to make its spectrum match a target response spectrum.

Synthetic or simulated ground motions is another option for describing seismic excitations that has been recently gaining popularity within the engineering community.
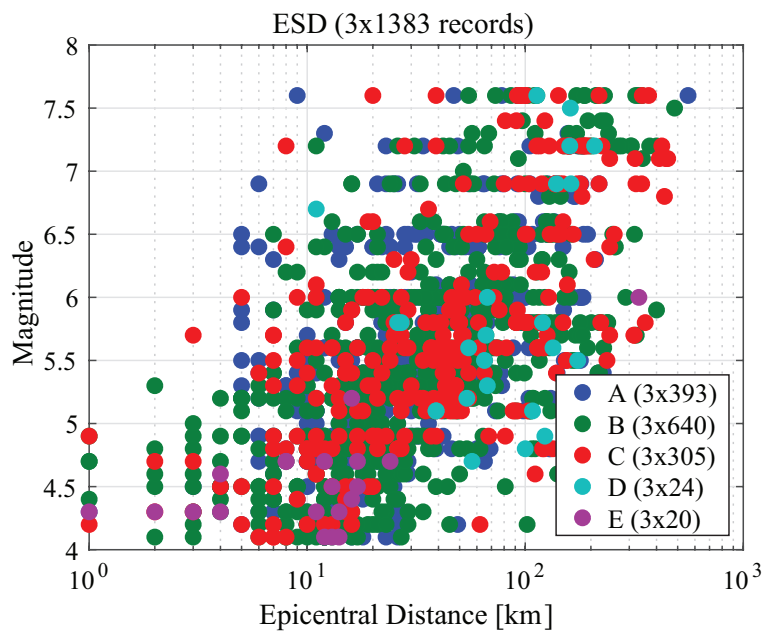
FIGURE 1.2: Lack of large magnitude and short epicentral distance ground motion records in the European strong motion database (ESD). A, B, C, D and E are soil classes according to Eurocode8 (2004).

Simulated ground motions can be used as an alternative to replace or supplement recorded ground motions in cases where these are absent or insufficient. Many ground motion simulation methods have been proposed over the last few decades, but only recently there has been an increasing interest in their use for structural engineering applications. The advances in scientific understanding of the earthquakes' physics and computational power have rendered ground motion simulations easier to compute and use. Two categories of ground motion simulations considered in this study are the physics-based and stochastic site-based.

Physics-based simulations are generated using numerical models that explicitly incorporate the physics of the fault rupture and seismic wave propagation and can thus, be used for simulating future scenarios. Some of their engineering applications may be the design of new or assessment of existing structures in areas where the seismic hazard is dominated by large magnitude events on nearby faults (i.e., in cases that records are scarce or absent), and the calibration of GMPEs in regions where there is not a sufficient number of records to study special topographic or site effects (e.g., study of basin effects). Physics-based ground motions may also represent an attractive option for loss estimation purposes within the catastrophe risk modeling framework, as there is a scarcity of recorded motions for large magnitude and short distance events that can cause nonlinear structural responses. Besides such applications, some researchers have used a rupture-to-rafters simulation approach which simulates the entire phenomenon: from earthquake rupture to nonlinear structural response and damage/loss prediction. Within this approach the seismic hazard and structural analysis are fully coupled. The results of such studies may ultimately be used (1) by public and private organisations to develop emergency response plans, (2) to quantify seismic risk by using a comprehensive collection of earthquake scenarios with associated probabilities, (3) to evaluate cost-effective seismic retrofitting actions and risk mitigation strategies, (4) to evaluate rapid damage-estimation algorithms for effective disaster response, and (5) to use the synthetic data sets to train and test health monitoring algorithms for damage identification (Krishnan et al., 2011).

The stochastic site-based simulation method models the ground motion time-history at a specific site by fitting a statistical process to a suite of recorded ground motions and is thus, relying on ground motion records for predictions. The stochastic ground motions can be used to supplement existing ground motion records in cases where a large number of ground motion time-histories is needed, such as in PBEE or where there is a lack of records with specific combination of source, path and site characteristics. Another very important application of stochastic ground motions models is in simulation-based reliability frameworks

(e.g., for seismic risk assessment) where there is a need for a comprehensive characterisation of acceleration time-histories for a wide range of seismicity scenarios.

Ground motions simulations can be very useful for a wide range of engineering applications as discussed above, provided that they capture the properties of real ground motions as well as their natural variability. However, among stakeholders and engineers the general concern in the context of PBEE, is that simulated ground motions may not be equivalent to real records in estimating seismic structural demand, and hence, the induced damages to structures and losses. Therefore, there is a need for validation of the ground motion simulations so that they can be used with confidence for the various engineering applications. Furthermore, given the significant improvements in computational power, it is evident that simulation-based methods that use stochastic ground motion models will play an important role in the future. This provides another strong motivation for having validated and trust-worthy simulated stochastic ground motion models.

1.2   Scope and aims of research

This study is concerned with the use of simulated ground motions for seismic risk assessment of structures. In particular, two categories of ground motion simulation methods are considered: physics-based and stochastic site-based simulation models. Physics-based models are computationally intensive and can be used to provide realistic synthetic ground motions in areas where ground motion records are scarce or absent; however, they need to be validated against records to be used with confidence in engineering applications. Stochastic site-based models, on the other hand, are fast to compute and have significant potential to be used in practice when a large number of ground motions is needed e.g., in PBEE or simulation-based seismic risk assessment.

The scope of this study is twofold. The first part of the thesis focuses on the introduction of statistics-based metrics for the validation of ground motion simulations

to be used for earthquake engineering applications with a focus on the seismic demand and damage induced in engineering structures. The novelty of this study is the introduction of an information theory-based approach for the engineering validation of ground motion simulations. The approach can be applied to any ground motion simulation method (physics-based or stochastic). Simple proxies for the spectral-shape and duration of the ground motions that are indicative of the peak and cyclic nonlinear structural response, respectively, are used for the validation. These proxies have been proposed in the literature, though, they have not been used in past validation studies.

In the second part of the study, the emphasis is placed on stochastic ground motion simulation models and their use in PBEE and seismic risk assessment applications. The main objective is the development of a computationally efficient framework for modification of existing stochastic ground motion models to achieve compatibility with the conditional seismic hazard at a given site and structure of interest. This conditional hazard is typically characterised through PSHA, for example through disaggregation as discussed above. Essential part of PSHA are the GMPEs. GMPEs provide predictions, as function of seismicity characteristics, for both the median and the dispersion of IMs, determining ultimately the conditional hazard for seismic events corresponding to these characteristics. This facilitates the use of the modified stochastic ground motion models for PBEE or seismic risk assessment applications, where suites of hazard-compatible ground motions are needed. Finally, the proposed modification framework is validated by comparing the seismic demand of recorded ground motions established through NLDA of single degree of freedom (SDoF) case-study systems to the demand of stochastic ground motion models established through the proposed modification.

Figure 1.3 depicts a schematic diagram that explains the most commonly used approach for seismic risk assessment of a specific structure using recorded ground motions (right part of the Figure) as well as the proposed approach (left part of the Figure), where the recorded ground motions have been replaced by ground motion simulations. The current state-of-the-art approach involves the selection
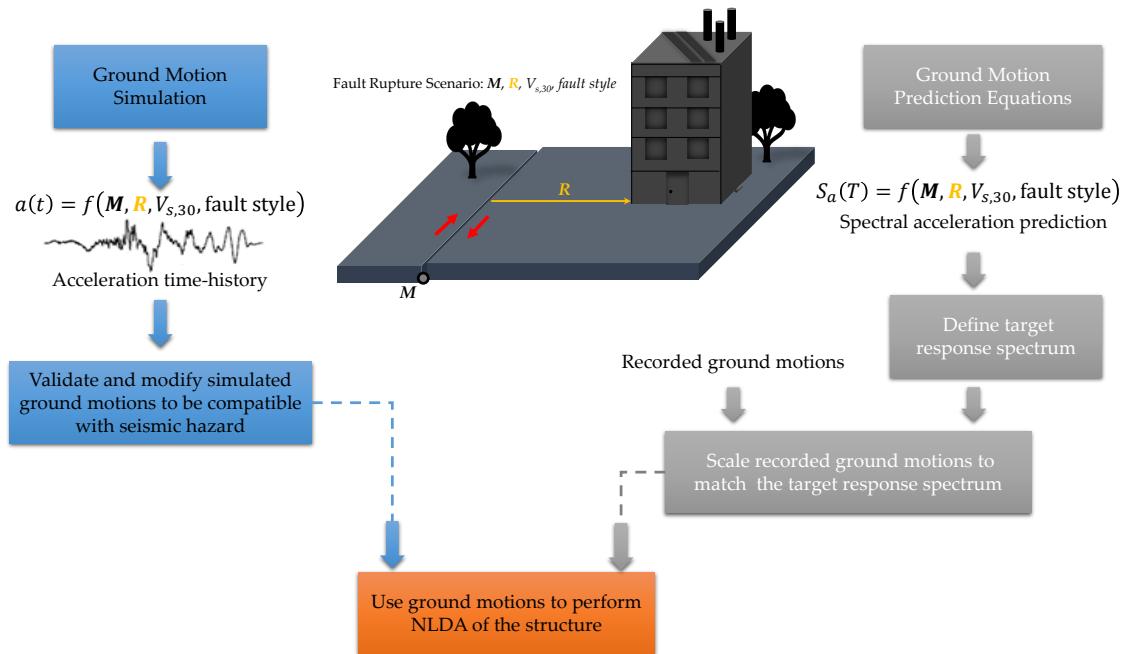
FIGURE 1.3: Schematic diagram of the current and proposed approaches for seismic risk assessment of structures.

of an earthquake scenario (or set of scenarios) which are used to define a target response spectrum. This response spectrum can be derived for a single scenario or based on PSHA incorporating multiple scenarios (resulting in a uniform hazard or conditional (mean) spectrum). The spectral acceleration predictions used in all cases are derived through the use of GMPEs. For the single scenario, these are simply estimated from GMPEs for the given seismicity and site characteristics such as magnitude (M), source-to-site distance (R), shear wave velocity in the upper 30 meters of soil ($V_{s,30}$), and fault style. For deriving a uniform hazard spectrum, the spectral acceleration predictions for a set of earthquake scenarios are aggregated, but still GMPEs are used to estimate spectral accelerations for each scenario. Recorded ground motions are then selected from a database and scaled to match on average the target response spectrum within a period range of interest. In the proposed approach, the recorded ground motions can be replaced by simulated ground motions that are generated based on a given seismicity scenario and site of interest. The scaled recorded or simulated ground motions are finally used to perform NLDA of the structure and calculate probability distributions of EDPs, annual rate or probability of exceedance of given levels of EDPs,

loss or any desired seismic risk metric. The presented approaches can be seamlessly applied to perform seismic risk assessment of a portfolio of structures in a given area. For the simulated ground motions to be used in the proposed seismic risk assessment approach, it is important to validate the stochastic ground motion models to ensure their output IM statistics (mean and dispersion) are compatible to these expected at the site of interest (e.g., as estimated through GMPEs). If that is not the case, modification of the stochastic ground motion model needs to be performed to achieve compatibility.

1.3   Thesis outline

Chapter 2 discusses the literature review on the topics of ground motion simulation and validation. In this chapter, the different ground motion simulation methodologies are presented and common validation approaches are described. Current gaps identified in the literature and the framework adopted in this study are discussed.

Chapter 3 introduces a quantitative approach for the engineering validation of ground motion simulations based on information theory concepts and statistical hypothesis testing. The application of the proposed validation approach is demonstrated to ground motion simulations generated by three simulation methods, including physics-based and stochastic models, for four historical events in California. The validation is performed in terms of spectral-shape and duration-related IMs, acting as proxies for the nonlinear response of more complex engineered systems. The considered IMs are shown to be the optimal IMs in several probabilistic seismic demand models of different structural types, within the framework of PBEE.

Chapter 4 focuses on the use of stochastic site-based ground motion models in PBEE and simulation-based seismic risk assessment applications. This is established by relating the parameters of the stochastic ground motion model to earthquake and site characteristics through predictive relationships. This chapter offers

a computationally efficient framework for the modification of stochastic ground motion models to match target IMs for a specific site and structure of interest. This is set as an optimisation problem with a dual objective. The first objective minimises the discrepancy between the target IMs and the predictions established through the stochastic ground motion model for a chosen earthquake scenario. The second objective constraints the deviation from the model characteristics suggested by existing predictive relationships, guaranteeing that the resultant ground motions not only match the target IMs but are also compatible with regional trends. A framework leveraging kriging surrogate modeling is formulated for performing the resultant multi-objective optimisation.

Chapter 5 discusses the extension of the developed framework in Chapter 4 to perform modification of stochastic ground motion models to establish compatibility with the seismic hazard, described through the mean and dispersion of some structure-specific IM(s), for given seismicity scenarios and structure/site. The modification pertains to the probabilistic predictive models that relate the parameters of the ground motion model to seismicity/site characteristics. These predictive models are defined through a mean prediction and an associated variance and both these properties are modified in the proposed framework. For both the predictive models and the seismic hazard a probabilistic description is considered. The proposed modification is defined as a bi-objective optimisation. The first objective corresponds to comparison for a chosen seismicity scenario between the target hazard and the predictions established through the stochastic ground motion model. The second objective corresponds to comparison of the modified predictive relationships to the pre-existing ones that were developed considering regional data, and guarantees that the resultant ground motions will have features compatible with observed trends. The relative entropy is adopted to quantify both objectives and a computational framework relying on kriging surrogate modeling is established for an efficient optimisation.

Chapter 6 focuses on the validation of the proposed stochastic ground motion

model modification by comparison to seismic demand of recorded ground motions. Suites of hazard-compatible recorded and modified stochastic ground motions whose spectral acceleration statistics match the mean and variance of target spectra within a period range of interest, are used as input to perform response-history analysis of inelastic SDoF case-study systems. EDP distributions are compared to perform the desired validation.

Finally, Chapter 7 presents the main conclusions of this study and recommends future research directions.

The chapters of this thesis are developed to be largely self-contained because they are published as individual journal articles. Because of this, there is some repetition in introductions and background material. In addition, notational conventions were chosen to be simple and clear for the topic of each chapter rather than for the thesis as a whole; because of this, the notational conventions may not be strictly identical for each chapter. Apologies are made for any distraction this causes when reading the thesis as a continuous document.

Chapter 2

Literature review

2.1 Introduction

This chapter provides a review of the current literature regarding ground motion simulation methodologies and validation metrics. The chapter is organised as follows: Section 2.2 presents the most common ground motion simulation methodologies encountered in the literature along with their strengths and limitations. Section 2.3 provides an overview the different approaches and metrics developed over the years for the validation of ground motion simulations. Section 2.4 discusses an open-source platform that enables users to generate ground motion simulations. Finally, Section 2.5 summarises the research gaps that have been identified in the literature and discusses the framework proposed in the thesis.

2.2 Simulation methods

The existing ground motion simulation methods can be grouped in three main categories: (1) Physics-based or deterministic methods (2) Stochastic-process-based or stochastic methods and (3) Hybrid methods. All these methods simulate a ground

motion time-history, whereas parameter prediction methods, the most common of which are GMPEs, typically use a single IM that represents a ground motion time-history (e.g., PGA, $S_a(T)$, etc.). Douglas and Aochi (2008) present a comprehensive review of the existing ground motion simulation methodologies including parameter prediction methods and summarise the history of simulation method development in Figure 2.1 (adapted from Burks, 2014). It is noted that this study focuses on simulation of ground motion time-histories as input to response history dynamic analysis of structures and thus, past studies with this objective will be reviewed. In this chapter the available simulation methods will be presented in detail along with the advantages and disadvantages of each method.
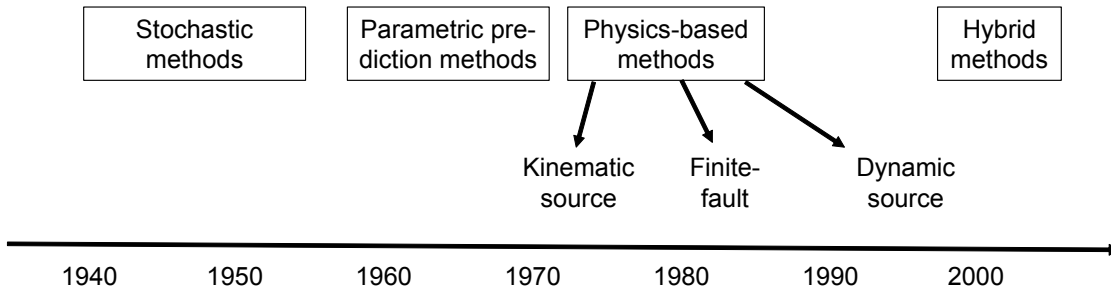
FIGURE 2.1: History of the development of the four main categories of ground motion simulation methods (after Douglas and Aochi (2008) as adapted from Burks (2014)).

2.2.1 Physics-based methods

Physics-based or deterministic methods generate ground motion simulations by using numerical models that explicitly incorporate the physics of the fault rupture and seismic wave propagation. Physics-based simulations were introduced in the 1970s but have been rapidly developed in the last few years due to advances in numerical methods and algorithms, and the growth of computing power and increased availability of parallel computers.

Taborda and Roten (2015) summarise the elements or steps of physics-based simulation methods as described below:

- Selection of a simulation domain;

- Selection of a source and material model;

- Definition of a set of model parameters for the targeted results;

- Implementation of solution methods and operation of a simulation engine, and;

- Execution of simulation

The first three items of the list include the input data for the simulation; that is the selection of the simulation domain, the source model, the material model and the simulation parameters. The source model describes the characteristics of the fault rupture in terms of location, orientation and slip history. The material model describes the properties of the material in the simulation domain. The last input required for simulation are the model parameters which most typically include the maximum targeted frequency in the simulation (f_{max}) and the minimum shear wave velocity ($V_{s,min}$), as well as the number of points per wavelength.

The last two items on the list refer to the solution method and its implementation in a computer code application. The most common numerical methods to estimate the solution to the wave propagation equation are the finite difference method (e.g., Alterman and Karal, 1968; Graves, 1996; Day and Bradley, 2001) and finite element method (e.g., Lysmer and Drake, 1972; Bao et al., 1998). Other numerical methods encountered in the literature include the spectral element method (e.g., Faccioli et al., 1997; Komatitsch and Vilotte, 1998; Komatitsch and Tromp, 1999) and discontinuous Galerkin method (e.g., Dumbser and Käser, 2006).

The physics-based methods can be divided into two main categories: the kinematic methods (e.g., Irikura, 1983; Zeng et al., 1994; Hartzell et al., 1999; Ruiz et al., 2007) and the fully spontaneous or dynamic methods (e.g., Olsen et al., 1997; Hartzell et al., 2005; Pulido and Dalguer, 2009).

The kinematic methods solve the fault rupture and the seismic wave propagation problems separately due to the complexity of the full phenomenon, so that the simulation consists of two steps: (1) simulation of the slip distribution on the fault, and (2) simulation of the resulting wave propagation from the fault to the site of interest. Dynamic rupture simulations, on the other hand, solve the fault rupture evolution combined with the triggered wave propagation problem by prescribing the fault pre-stress, fracture energy and stress drop. Due to their complexity and lack of constraints in the current level of knowledge regarding earthquake rupture dynamics, the dynamic methods are not yet developed to the point that they can be used in engineering applications. Most physics-based simulations use kinematic models to generate simulations and thus, this study will only refer to these models.

As explained above, the input of the physics-based simulations consists of two models used to represent the earthquake source and the propagation media and, that is the source model and the material model, respectively. In the kinematic methods, source models are resolved in the first stage and are then used as basic input in the second stage, which is the ground motion simulation. The source model is most commonly represented by a set of self-balanced forces that are applied to the second stage to trigger the wave propagation.

Taborda and Roten (2015) discusses the two types of source models used by researchers, the point-source and finite-source models. In the case of small magnitude earthquakes ($M < 5$) or sources small enough compared to the wavelength of the radiated energy, the effect of the earthquake rupture and the discontinuity of displacements on the fault can be modelled using a point-source model, that is a single set of equivalent forces acting on a point. In the case of larger magnitude events, the earthquake is modelled as a sum of many smaller earthquakes by discretising the entire rupture area into smaller subfaults, each of which with their own point-source model. The geometry and point-source model of the subfaults are such that when added up, they adjust to the geometry and expected energy release of the entire fault. The simulated ground motions from each subsource are then summed at the site of interest, with appropriate time lags considering the

difference from the triggering time of the subsource and the travel time, to produce the ground motion for the entire fault. These type of source models that consist of a collection of smaller subfaults are called finite-source models. In both point and finite-source models, the point-source model is defined by some geometry and rupture characteristics including the hypocentre location, strike, dip, rake angle, area, and average shear modulus of the subfault as well as the evolution of the slip with time.

The material model is the second input required for the simulation and consists of the material's density (ρ), the P- and S-wave seismic velocities (V_p and V_s) and the material's attenuation properties expressed with the quality factors Q_p and Q_s. The quality factors are defined using attenuation relationships that are empirical functions of V_p and V_s.

The physics-based simulation methods produce realistic synthetic ground motions at low frequencies (less than 1 Hz), but they are computationally intensive and require a large number of seismological information about the rupture area (e.g., crustal structure, rise time, stress drop, cut-off frequency, material properties, basin effects etc.) that is not readily available for many regions. However, as demonstrated by Graves and Aagaard (2011) in the case of a future event, the input parameters for the simulation can either be reliably estimated (e.g., seismic moment and fault dimensions) or parametrically assessed using multiple realisations (e.g., hypocenter location and slip distribution). All other source parameters can be determined using the scaling relations described in Graves and Pitarka (2010).

2.2.2 Stochastic-process-based methods

The stochastic ground motion simulation methods make use of statistical approaches to integrate the physics and characteristics of the earthquake source,

path effects, directivity and site effects into simple functional forms. The parameters of the stochastic ground motion models are represented by probability distributions. They are valid representation of the physics of earthquakes without solving the mathematical problem of fault rupture and wave propagation.

Two approaches (Rezaeian and Sun, 2015) used for the NLDA of structures are: (1) nonlinear response-history dynamic analysis by use of a suite of ground motions and (2) nonlinear stochastic dynamic (i.e., random vibration) analysis by use of a stochastic representation of the ground motion. For the former approach, a suite of recorded or synthetic ground motions or a combination of both can be used and both deterministic and stochastic simulations can be utilised. The stochastic dynamic analysis for yielding/nonlinear structures encompass various statistical linearisation techniques which define, first, an equivalent linear system minimising the error/difference between its response statistics for a given stochastic excitation and the response statistics of the nonlinear system. Such approaches by-pass the need for response-history dynamic analysis as well as selection of recorded ground motions (e.g., Giaralis and Spanos, 2010; Mitseas et al., 2018).

Synthetic ground motions should properly capture the characteristics of the recorded ground motions that control the response of the structures, that is the intensity, duration and frequency content. Many researchers have conducted reviews of stochastic simulations over the last few decades and one important issue raised in all of them is that of temporal and spectral nonstationarities (Liu, 1970; Ahmadi, 1979; Shinozuka and Deodatis, 1988; Kozin, 1988; Conte and Peng, 1997; Rezaeian and Der Kiureghian, 2008). Temporal nonstationarity is the variation of the amplitude of the time-series with time (nonstationarity in time domain), while spectral nonstationarity is the variation of frequency content with time (nonstationarity in the frequency domain). In particular, spectral nonstationarity can affect the nonlinear dynamic structural response (Yeh and Wen, 1990; Conte, 1992a; Wang et al., 2002), because the structure's behaviour becomes nonstationary as it gets in the inelastic response range and is subject to period elongation (Papadimitriou, 1990).

Another important issue for stochastic simulations is the proper representation of the natural variability of recorded ground motions. Some simulation methods underestimate the natural variability of the ground motions because they underestimate the variability of their model parameters. On the other hand, if correlations between the model parameters are not taken into account or if they are underestimated, the variability of the simulations will be overestimated compared to recorded ground motions (Rezaeian and Sun, 2015).

Stochastic models are the oldest ground motion simulation methods. The motivation behind their development was partly the lack of sufficient and dependable recorded ground motions considering that the first strong motion to be recorded by a seismograph next to a fault rupture was not until the 1940 El Centro or Imperial Valley earthquake. Furthermore, stochastic models had the potential to be used in random vibrations as well as nonlinear stochastic dynamics techniques which, whilst approximate, offered meaningful results at times where computers were scarce and nonlinear response-history dynamic analysis was a luxury.

The oldest and most widely used stochastic ground motion model is the Kanai-Tajimi model (K-T) (Kanai, 1957; Tajimi, 1960) expressed through a stationary coloured random process that represents the response of a SDoF oscillator modeling soil deposits to white-noise excitation. In the early 1960's, several enveloping functions in the time-domain have been considered to modulate the K-T yielding uniformly modulated non-stationary stochastic process. In late 1980's, the concept of K-T filters with time-varying properties has been proposed to capture the nonstationary frequency content of recorded accelerograms (Fan and Ahmadi, 1990).

Stochastic ground motion models can be divided in two categories: the source-based and site-based models. Source-based models describe the fault rupture at the source and the propagation of the seismic waves at the site of interest explicitly accounting for the path and site effects, whereas the site-based models describe the ground motion time-history at a specific site by fitting a stochastic process to

recorded ground motions with known earthquake and site characteristics (Reza-eian and Der Kiureghian, 2008). In contrast to source-based models, site-based models account for the source, path and site effects implicitly through empirical calibrations.

Stochastic source-based simulations have been developed by a number of researchers in the past few decades (e.g., Beresnev and Atkinson, 1998; Boore, 2003; Motazedian and Atkinson, 2005; Boore, 2009) based on the work of McGuire and Hanks (1980), that uses the Fourier amplitude spectrum of a ground motion and then combines it with a random phase spectrum, assuming the ground motion to be a band-limited white Gaussian noise with finite duration. Stochastic source-based method was first developed to model far-field ground motions where the earthquake source can be considered as a point (e.g., Boore, 2003), resulting in point-source stochastic models. For simulation of ground motions closer to the earthquake source, the method was improved to consider the rupture progress on a finite fault (Beresnev and Atkinson, 1998), resulting in finite-fault stochastic models. In finite-fault modeling, the fault is discretised into many subfaults, and each subfault is treated as a point-source. The ground motion from each subfault is modeled using the point-source stochastic model with its own amplitude spectrum. The total ground motion at a site is the superposition of the contributions of all subfaults with a proper time lag considering the difference from the triggering time of the subfault and from the travel time between the subfault to the site (Rezaeian and Sun, 2015). One of the disadvantages of the point-source stochastic models is that they don't include variation of the frequency content with time and they can underestimate the natural variability of ground motions as they fix their model parameters. However, the latter can be addressed by assigning probability distributions to the model parameters to introduce parametric uncertainty (Vetter and Taflanidis, 2014). In general, stochastic source-based models require a thorough knowledge of the source, path and site characteristics as well as a good understanding of their underlying principles which makes them not practical to be used by practicing engineers.

Site-based models on the other hand, are easy to compute due to their simple formulations and can be more practical to use in engineering applications, as their parameters can be correlated with basic earthquake (e.g., moment magnitude and rupture distance) and site (e.g., shear wave velocity) characteristics. The assumption behind these models is that the ground motion is a zero-mean Gaussian process. Rezaeian and Der Kiureghian (2008) classifies site-based stochastic ground motion simulations into four categories:

(1) Filtered white-noise processes that are obtained by passing a white-noise signal through a filter with subsequent modulation in time to achieve temporal nonstationarity (e.g., Bolotin, 1960; Shinozuka and Sato, 1967; Amin and Ang, 1968; Iyengar and Iyengar, 1969; Ruiz and Penzien, 1971). A disadvantage of these processes is that they have constant frequency content throughout the time-history. Auto-regressive moving average (ARMA) models (e.g., Jurkevics and Ulrych, 1978; Hoshiya and Hasgur, 1978; Polhemus and Cakmak, 1981; Chang et al., 1982; Kozin, 1988; Conte et al., 1992b; Mobarakeh et al., 2002; Giaralis and Spanos, 2009) also fall under the umbrella of methods that filter white-noise. These models can simulate ground motions with temporal and spectral nonstationarity using time-varying model parameters; however, it is difficult to correlate the model parameters with earthquake and site characteristics and that makes them impractical to use for engineering applications. Rezaeian and Der Kiureghian (2008) developed a fully nonstationary stochastic model with separable temporal and spectral nonstationarities based on previous work by Yeh and Wen (1990) and Papadimitriou (1990). Their model uses a filtered white-noise process in the time domain with the filter having time-varying properties, thus allowing variation of the frequency content with time. The modulated process with spectral nonstationarity is subsequently passed through a time modulating filter to generate a fully nonstationary process as presented in Figure 2.2. The process is finally passed through a high-pass filter to ensure zero residual velocity and displacement and provide reasonable spectral

response estimates at long periods. The model parameters are related to earthquake and site characteristics such as the faulting mechanism, earthquake magnitude, source-to-site distance and shear wave velocity through predictive equations Rezaeian and Der Kiureghian (2010). Vlachos et al. (2016) recently developed a stochastic ground motion model that uses a bimodal, analytical, fully nonstationary spectral version of the K-T model. The model parameters are established in the energy domain and are related to earthquake (e.g., moment magnitude and rupture distance) and site (e.g., shear wave velocity) characteristics to facilitate simulation of stochastic ground motions for specific earthquake scenarios and sites of interest (Vlachos et al., 2018).

(2) Filtered Poisson processes obtained by passing a train of Poisson pulses through a linear filter (e.g., Cornell, 1964; Lin, 1965). These processes can achieve temporal and spectral nonstationarity through modulation in time (Lin, 1986); however matching with natural recordings is difficult.

(3) Various forms of spectral representation (e.g., Saragoni and Hart, 2002; Der Kiureghian and Crempien, 1989; Shinozuka and Deodatis, 1991; Conte and Peng, 1997; Pousse et al., 2006; Yamamoto and Baker, 2013). Contrary to the previous three categories, these models work in the frequency domain and focus on developing a time-varying spectral representation that matches a recorded ground motion. They use a short-time Fourier transform or wavelet transform to develop a time-frequency modulating function.

Emphasis in this study is placed on site-based stochastic ground motion models that use filtered white-noise processes and more specifically, the class of models that generate synthetic ground motions for specific earthquake and site characteristics. The techniques discussed can be extended to any type of stochastic ground motion model as long as a link between the model parameters and earthquake and site characteristics through predictive relationships is provided (Atkinson and Silva, 2000). For illustration of the developed methodologies, the stochastic ground motion model by Rezaeian and Der Kiureghian (2008) is used later in the thesis.
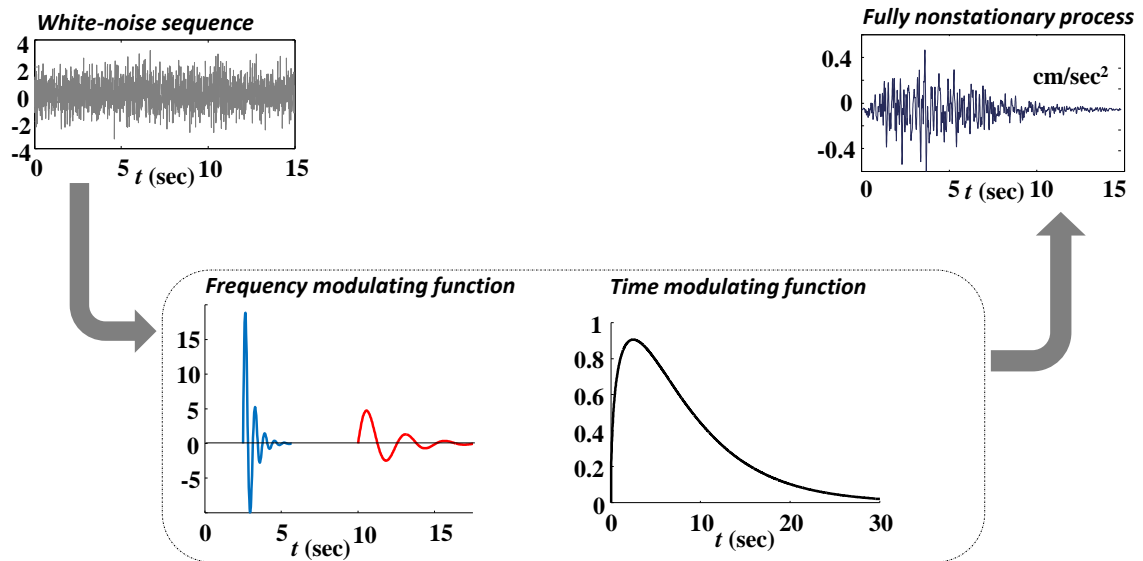
FIGURE 2.2: Development of a fully nonstationary stochastic process according to Rezaeian and Der Kiureghian (2008).

An important concern related to the use of stochastic ground motion models for structural engineering applications is the fact that through current approaches in selecting their predictive relationships, compatibility to the seismic hazard for specific structures and sites is not necessarily obtained. Rezaeian and Der Kiureghian (2010) validated their model by comparing model predictions with spectral acceleration estimates from GMPEs. The comparison is performed for both the median and median $\pm$ one standard deviation of pseudo-acceleration response spectra as predicted by four GMPEs. The results show that a good match to the median and median $\pm$ one standard deviation estimates from GMPEs is not obtained for all considered earthquake scenarios and structural periods. Such a match to some desired GMPE (or target IMs in general) is, though, important for subsequent use of the stochastic ground motion models to describe the seismic hazard.

This observation has motivated researchers to investigate the selection of predictive relationships for stochastic ground motion models so that compatibility with GMPEs is explicitly established. Scherbaum et al. (2006) introduced an explicit optimisation for matching the median predictions of the ground motion model to the spectral acceleration estimates of GMPEs, while maintaining physics-based principles or the matching to trends from real ground motions as an optimisation constraint, in an attempt to preserve desired ground motion characteristics. Vetter et al. (2016) extended the work of Scherbaum et al. (2006) and proposed a versatile and computationally efficient framework, leveraging surrogate modeling for tuning stochastic ground motion models to generate GMPE-compatible stochastic ground motions for a range of seismicity scenarios and structural periods of interest. The parameters of the stochastic ground motion model are tuned to optimise the match of the model predictions to the ones from one or more selected GMPEs in terms of spectral accelerations at different periods ($S_a(T)$). One of the advantages of their model is that unlike its predecessors, it can be directly used within a seismic risk assessment framework on account of its target versatility (the response from different earthquake scenarios and structural periods can be selected as targets) and hazard compatibility. One of the main drawbacks of this tuning approach, though, is that the physical characteristics of the resulting acceleration time-series are incorporated in the optimisation merely as constraints, something that requires significant experience in ground motion characterisation for proper definition of the optimisation problem, and can furthermore lead to synthetic time-histories with unrealistic properties for some seismicity scenarios. Another shortcoming of all the aforementioned studies is that they focused on the mean model characteristics and GMPE estimates. Optimisation utilised only the mean of the predictive relationships of the stochastic ground motion model, whereas, more importantly, match only to the median spectral accelerations from GMPEs was investigated, ignoring any variability in their predictions. That is though, an important constraint since for seismic risk assessment applications hazard compatibility is expressed in terms of both the mean and dispersion of some target IMs.

As a summary, stochastic site-based models are fast to compute and very practical to use for engineering applications when a large number of ground motions is needed e.g., in PBEE. The essential component for such applications is the predictive models relating seismicity/site characteristics to ground motion model parameters. Their parameters are calibrated based on observations from previously recorded ground motions and this limits their use to areas where ground motion data is available. However, the recent advances in physics-based ground motion simulation methodologies can provide valuable data to fill in gaps in recorded ground motion datasets and extend the use of stochastic-process-based models.

2.2.3   Hybrid methods

Hybrid methods compute the low-frequency and high-frequency components separately and then combine them to generate a single ground motion time-history. More specifically, they combine a physics-based simulation method for the low-frequency range with a stochastic simulation for the high-frequency range. The use of different simulation approaches at different frequency bands results from the observation that the effects of fault rupture and wave propagation become stochastic at frequencies of about 1Hz and higher that mainly reflects the lack of knowledge about the details of these phenomena at high frequencies (Graves and Pitarka, 2010). This is also consistent with the observation that physics-based simulations produce realistic synthetic time-histories at low frequencies (typically below 1 Hz). Several hybrid models exist in the literature including the work of Hartzell et al. (1999); Liu et al. (2006); Frankel (2009); Graves and Pitarka (2010); Mai et al. (2010); Graves and Pitarka (2015). These simulations combine the advantages of the physics-based and stochastic simulations and thus, the final ground motion is more realistic across a wider frequency band, which is particularly relevant for tall buildings and multiple degrees of freedom (MDoF) systems with higher frequency modes. This type of simulations requires the same detailed seismological information to describe the source mechanism, wave travel path and

local site effects as physics-based simulations and therefore, they can only be used in regions where seismological data is available.

2.3   Validation of simulated ground motions

Ground motions simulations are an attractive option that can supplement or substitute recorded motions in many engineering applications. In particular, physics-based and hybrid simulations can give insight in cases that scarcity or absence of recorded motions is a problem. Stochastic-based simulations on the other hand, being simpler and faster to compute, may be more practical in PBEE, where a large number of input ground motions is required to assess different damage levels of structures. Before these simulations can be used in engineering applications, it is important to show that they produce ground motions that are equivalent with the real records in estimating seismic demand and induced damage on structures. Several approaches and metrics to validate ground motions exist in the literature.

2.3.1   Validation approaches

Burks (2014) places the existing validation approaches in three categories: (1) the historical approach which compares simulations with past recordings of historical events, (2) the empirical model approach, which compares simulations with predictions from empirical parameter models (e.g., GMPEs), and (3) the similar spectra approach, which compares groups of simulations and recordings with similar elastic response spectra. These approaches can be used to validate simple IMs as well as structural response for a particular event or a response spectrum.

2.3.1.1   Historical approach

The historical approach is used to directly compare the simulations of a historical event with ground motion recordings for the same event and it is the oldest method found in the literature. During these early studies, the researchers that developed the simulation method performed a visual inspection of the acceleration, velocity and displacement waveforms of simulations to test whether they match historical recordings (e.g., Zeng et al., 1994; Olsen et al., 1997; Pitarka et al., 1998; Wang, 1999; Olsen et al., 2003). As the engineering interest in simulations has grown, researchers started using simple validation metrics such as PGA and Modified Mercalli Intensity (MMI) for large sets of simulations and past records, (e.g., Hartzell et al., 1999, 2005; Aagaard et al., 2008).

Eventually, as the understanding of the ground motion properties that are important for the seismic structural performance grew, the simple validation metrics used in the early studies evolved into goodness-of-fit parameters, which can quantify the misfit between simulations and recordings in terms of peak ground values (PGA, peak ground velocity (PGV) and peak ground displacement (PGD)), spectral accelerations at different periods ($S_a(T)$), shaking duration and other engineering-specific metrics, such the ratio of inelastic to elastic displacement (e.g., Anderson, 2004; Kristekova et al., 2006; Olsen and Mayhew, 2010; Dreger et al., 2015). In particular, Olsen and Mayhew (2010) proposed a goodness-of-fit criterion using several IMs and the ratio between inelastic and elastic response spectra. They apply the proposed criterion to the 2008 M 5.4 Chino Hills, California earthquake, concluding that the simulated ground motions yield realistic results for moderate and long structural periods.

Recent studies have moved a step further to perform engineering validation of simulated ground motions for historical events in terms of linear and nonlinear response of SDoF (Bazzurro et al., 2004; Galasso et al., 2012) and MDoF systems (Galasso et al., 2013). More specifically, Bazzurro et al. (2004) using suites of simulated and real records from the 1994 M 6.7 Northridge earthquake conclude

that six out of seven simulation methods considered in their study appear to be biased both in the elastic and post-elastic cases. Similarly, Galasso et al. (2012, 2013) examine engineering validation in terms of elastic and inelastic response of structural systems for four historical earthquakes modeled using the hybrid broadband ground motion simulation methodology by Graves and Pitarka (2010). The validation metrics used include various EDPs, such as the inelastic displacement (Δ_{in}), and equivalent number of cycles (N_e) for SDoF systems and maximum interstory drift ratio (MIDR) and peak floor acceleration (PFA) for MDoF systems. They also compare the intra-event variability of the simulations at different stations with the natural variability observed from the recordings. The statistical significance of the differences observed in the structural response and dispersion from simulations and records was tested using hypothesis testing. The results indicate that there are small differences between median estimates of seismic demand obtained by using real records and simulations especially in the transition area between semistochastic and deterministic simulations (around 1 s). The observed differences can be attributed to the systematic differences in the shape of the elastic and inelastic response spectra. For all the events considered in their study, the intra-event dispersion in the structural response due to simulations is generally lower than that for recordings at short periods. At longer periods, the simulations can contain strong velocity pulses that cause them to overestimate the intra-event dispersion (Galasso et al., 2012).

Rezaeian et al. (2015) propose an engineering validation of ground motion simulations based on their waveform characteristics using three time-dependent validation metrics that capture the nonstationary intensity and frequency features of the ground motions. The proposed metrics are the mean-square intensity of the acceleration time-series which quantifies the evolving intensity of the ground motion, the cumulative number of zero-level up-crossings, which quantifies the evolution of the predominant frequency of the ground motion, and the cumulative number of negative maxima and positive minima, which relates to the evolution of the bandwidth of the ground motion. The proposed validation approach is applied

to recorded and simulated ground motions from the 1994 M 6.7 Northridge event generated by using the simulation method by Graves and Pitarka (2010).

A general concern regarding the historical validation approach is that ground motion simulation developers tend to tune their model parameters to provide the best fit with past records, thus model developers should be cautious in order to avoid the over-fitting of model parameters (Stewart et al., 2001). The historical validation approach cannot be used for future earthquake scenarios where no past recordings are available, so a different validation approach may be applied in this case. Nevertheless, engineering validation of historical events is necessary in order to provide feedback to ground motion developers and help them improve the predictive capabilities of their models.

2.3.1.2   Empirical model approach

The empirical model approach compares simple IMs of simulated ground motions (e.g., PGA, $S_a(T)$) with the output from empirical relations (e.g., GMPEs). There are several GMPEs available in the literature that are used to estimate ground motion IMs for future events, thus they can be used to compare with simulations when past records are not available. There are some examples of simulation validation studies that use predictions from GMPEs as a baseline (Frankel, 2009; Star et al., 2011; Dreger et al., 2013, 2015) and in general show good fit of the IMs used as validation proxies.

In particular, Star et al. (2011) compare simulations of an M 7.8 rupture scenario on San Andreas fault and an M 7.15 Puente Hills blind thrust scenario to median and dispersion predictions from Next Generation Attenuation (NGA) GMPEs in terms of elastic spectral accelerations at different periods. The observed differences can be due to problems with the simulations, GMPEs, or even both. Moving a step further, De Luca et al. (2014) present prediction equations for peak and cyclic inelastic SDoF systems' response, developed based on Italian accelerometric data, and use them as a baseline for the engineering validation of broadband

hybrid ground motion simulations of the 1980 M 6.9 Irpinia earthquake in Italy. Results show that the simulation method may lead to generally acceptable results, although the authors consider this study to be preliminary due to the limited number of simulated records used.

One limitation for the use of the empirical model approach is the fact that the empirical models used are calibrated using observations from past earthquakes and thus, give reliable results for the magnitude and distance range of the underlying data. Since in most of the cases ground motion simulations are generated for rare events for which recordings are scarce or not available, it is difficult to rely on these models for comparison.

2.3.1.3   Similar spectra approach

The third approach used in ground motion simulation validation compares sets of simulations and recordings that have similar elastic response spectra. This approach is important for both performance-based design and code-based applications that require a suite of acceleration time-histories that match a hazard compatible response spectrum as input to nonlinear response-history analysis, and for probabilistic seismic demand analysis that assesses structural response at different intensity levels. Validation studies (e.g., Iervolino, De Luca and Cosenza, 2010; Atkinson and Goda, 2010; Jayaram and Shome, 2012; Burks et al., 2015) show that, in general, the differences in structural performance to synthetics and recordings with similar elastic response spectra are not statistically significant. In particular, Burks et al. (2015) focus on the validation of hybrid broadband ground motion simulations that match the ASCE 7 (ASCE, 2010) building code specified spectrum, for use by structural engineers as input to nonlinear response-history analysis. They conclude that the structural response to simulations and recordings is similar with most of the discrepancies explained by differences in directional characteristics (i.e., orientation of the maximum spectral response).

Burks and Baker (2014) present a simulation validation framework based on the empirical model and similar spectra validation approaches and propose a list of parameters for the response of complex structural systems that can be used as proxies for the validation of ground motion simulations for engineering applications. These proxies include (1) the correlation of spectral acceleration across periods, which is a proxy for the spectral-shape and relevant to the response of structures dominated by higher-mode periods and expected to suffer period elongation, such as tall buildings, (2) the ratio of maximum to median spectral acceleration across all horizontal orientations, which is indicative of the directionality of the ground motion and important for 3-D structures that respond in all orientations, and (3) the ratio of inelastic to elastic displacement that is indicative of nonlinear behaviour and important for structures expected to perform nonlinearly during large earthquake events. All these proxies have reliable empirical models that can be used as baseline for validation of the simulations.

2.3.2   Validation metrics

The credibility of simulated ground motions is assessed based on criteria that are used to quantitatively evaluate the similarity of simulated and recorded acceleration time-histories. One common approach adopted by researches involves the use of goodness-of-fit criteria to compare how well the simulations match the ground motion records. Hypothesis testing has also been employed in recent studies in the literature. The following sections provide an overview of the aforementioned validation approaches.

2.3.2.1   Goodness-of-fit measures

Recognising that an earthquake ground motion is a very complex time series that is very difficult to characterise by means of a single parameter, Anderson (2004) proposed a suite of ten metrics that can be used for validation. The proposed

method first filters the acceleration time-histories in up to ten narrow pass-bands and gives a score to each different metric in the frequency band. The suite of metrics comprise the PGA, PGV, PGD, Arias intensity, the integral of velocity squared, Fourier spectrum and acceleration response spectrum on a frequency-by-frequency basis, the shape of the normalised integrals of acceleration and velocity squared, and the cross correlation. The goodness-of-fit for each metric in the frequency band is expressed by a score between 0 and 10, with 10 indicating perfect match, and is estimated using equation 2.1:

$$S(p_1, p_2) = 10\exp\left\{-\left[\frac{(p_1 - p_2)}{min(p_1, p_2)}\right]^2\right\} \tag{2.1}$$

Where p_1, p_2 are the metrics used for comparison (e.g. PGA).

The scores for all the metrics are averaged over the frequency bands to provide an overall goodness-of-fit measure. In order to examine what the different scores mean in terms of quality of fit, the author performed calibration using two different approaches: (1) by comparison of 1000 pairs of synthetic acceleration time-histories generated by the same simulation method for the same earthquake scenario, and (2) the comparison of two horizontal components of recorded ground motions. Based on the calibration results, the author suggests the following goodness-of-fit classification: A score below 4 is a poor fit, a score of 4 to 6 is a fair fit, a score of 6 to 8 is a good fit, and a score over 8 is an excellent fit.

Olsen and Mayhew (2010) proposed a goodness-of-fit measure that uses similar validation metrics to those considered in Anderson (2004), but they also included structural engineering-specific metrics to account for the nonlinear structural response such as the ratio between the maximum inelastic and elastic displacements. The goodness-of-fit measure is calculated using equation 2.2:

$$GOF = 100 * \text{erfc}\left[\frac{2|x - y|}{x + y}\right] \tag{2.2}$$

Where erfc is the complementary error function of a normalised residual and x and y are two sets of positive scalar metrics yielding a goodness-of-fit value between 0 and 100, with 100 indicating a perfect match. The overall goodness-of-fit value is estimated as the weighted average of the different metrics considered for validation. The authors proposed the following classification of the goodness-of-fit values based on the validation they performed for the 2008 M 5.4 Chino Hills, California earthquake: 80 to 100 is an excellent fit, 65 to 80 is a very good fit, 45 to 65 is a fair fit and 35 to 45 is a poor fit.

Recently, Dreger et al. (2015) suggested an alternative goodness-of-fit metric for validation and implemented it in a large-scale ground motion simulation validation exercise. Details of this metric are discussed in Section 2.4.1 below.

2.4   SCEC Broadband Platform

The Southern California Earthquake Centre (SCEC) released the Broadband Platform (BBP), an open-source software distribution that enables third-party users to compute broadband synthetic ground motions for engineering applications. The BBP contains several physics-based ground motion simulation models, developed by researchers, that consist of computer code incorporated in the platform by the SCEC software development group. The output of the BBP ground motion models provided to the user are acceleration or velocity time series. Time series post-processing is performed within the BBP using common software tools. The BBP also contains software utilities for evaluation and comparison of ground motion simulation results with recordings from past events and GMPEs (Baker et al., 2014). The distribution of the BBP as open-source software provides transparency to the scientific models and computer codes implemented within the platform and consists a reproducible way to simulate ground motion time-series for engineering applications.

2.4.1   SCEC ground motion simulation validation technical activity group

The SCEC BBP can produce large number of simulations for historical and future events and thus, has great potential for validation studies for engineering applications. The SCEC technical activity group (TAG) on ground motion simulation validation (GMSV) was established in 2010, with the objective to develop and implement testing/rating methodologies for simulated ground motions to be used in engineering applications and provide feedback to model developers. So far, a number of validation studies have been funded by the SCEC Software Environment for Integrated Seismic Modeling (SEISM) project and other projects have been funded through the annual SCEC request for proposals. Luco et al. (2013) summarises the group's progress including studies on GMSV for engineering analysis using simple and robust proxies (Rezaeian et al., 2015; Burks and Baker, 2014), GMSV for elastic and inelastic response of SDoF systems (Galasso et al., 2012), GMSV for building code nonlinear response-history analysis (Burks et al., 2015), GMSV for nonlinear response of MDoF structural systems (Galasso et al., 2013; Bijelić et al., 2014), and application of simulated ground motions to duration-sensitive geotechnical systems (Afshari and Stewart, 2013; Rathje and Peterman, 2013; Afshari and Stewart, 2016).

Besides these studies, the SCEC BBP has been recently used in a large-scale GMSV exercise performed by the GMSV TAG of SCEC (Dreger et al., 2015; Maechling et al., 2015). The ground motion simulations are generated using five different broadband finite-source simulation methods implemented in BBP version 14.3 (as of March 2014): a stochastic source-based white-noise method (EXSIM) (Atkinson and Assatourians, 2015), two deterministic approaches including the composite source model (CSM) (Anderson, 2015), and the UCSB method (Crempien and Archuleta, 2015), and two hybrid approaches referred to as Graves and Pitarka (GP) (Graves and Pitarka, 2015), and SDSU (Olsen and Takedatsu, 2015). The ground motion simulations were generated using the BBP version 14.3 for

twelve historical earthquake events (western, central and eastern Unites States and Japan), and four generic strike-slip and reverse scenarios for which GMPEs are considered to be well constrained by data.

According to Goulet et al. (2015), the purpose of this validation exercise is mainly to fill the gap in recorded datasets for pseudo-spectal accelerations (PSA) for two ground motion hazard projects: (1) the southwestern United States utilities project, and (2) the Pacific Earthquake Engineering Research Centre (PEER) NGA project for the central and eastern North America region (NGA-East). Therefore, the ground motion simulation methods are evaluated based on their performance in matching the PSA of recordings for historical earthquakes or predictions from GMPEs for future earthquake scenarios. The selected validation metric is the $RotD_{50}$ 5%-damped PSA, where $RotD_{50}$ is the median value of the resultant of two horizontal PSA components of the ground motion as computed over each degree of rotation from 1° to 180° (Boore, 2010). The comparison of simulations and recordings or predictions from empirical relations is done in terms of PSA using a combined goodness-of-fit parameter, taken as the equally weighted sum of the absolute value of the mean residuals and the mean of the absolute value of the residuals (Dreger et al., 2015). The results from this study indicate that the simulation methods provide reasonable estimates of PSA, however it is suggested that additional research work is necessary to validate ground motion simulations for other applications using different proxies and metrics.

2.5 Research gaps and proposed framework

The scope of this study is twofold. The first part of the thesis focuses on the introduction of information theory and statistics-based metrics for the validation of ground motion simulations used as input to NLDA to assess the seismic demand and damage of engineering structures. The approach can be applied to any ground motion simulation methodology (physics-based, hybrid and stochastic). In the

second part of the thesis, the emphasis is placed on stochastic ground motion simulation models and their potential use in PBEE and seismic risk assessment applications. The literature review results in the identification of several technical gaps related to the scope of this study that are discussed below.

Regarding the validation of ground motion simulations (first part of the thesis), typically, in the literature the validation is performed as a paired comparison (i.e., at the same recording locations for historical events) between the recorded and simulated IM or EDP datasets in terms of the first two moments of their empirical distribution (mean and standard deviation). A validation approach assessing the overall similarity of the probability distributions of the studied IMs or EDPs for recorded and simulated ground motions would be useful when validation in terms of the nonlinear structural demands or expected loss for a portfolio of structures (or infrastructure) is of interest. Another useful application would be to measure the similarity of the distributions of seismic response to sets of simulations and recordings matching a target (elastic) response spectrum mean and variance, as in the "similar spectra" validation approach.

Most previous studies focus on the validation of ground motion simulations using metrics such as the peak ground values, spectral acceleration values at different periods $(S_a(T_i))$, significant duration of the ground shaking and various EDPs for SDoF and MDoF elastic and inelastic systems. In the recent years several studies have shown the influence of the spectral-shape in the response of structures dominated by higher-mode effects and structures that are expected to perform non-linearly during strong shaking (Giovenale et al., 2004), however, spectral-shape proxies have not been used in the literature for the validation of ground motion simulations. Burks and Baker (2014) introduce the correlation of spectral acceleration across periods as a good proxy for the spectral-shape, however this proxy is highly dependent on GMPE estimates and it can thus, be used only in areas for which reliable GMPEs are available.

The second part of the thesis investigates the use of stochastic site-based ground

motion simulations for seismic risk assessment. The parameters of stochastic ground motion models are related to seismicity and site characteristics through predictive relationships, something that facilitates a comprehensive description of the seismic hazard. For such applications, it is important that the output IMs from the simulated acceleration time-histories are consistent with these prescribed at the site of interest (e.g., $S_a(T)$ estimates from GMPEs) for specific structures. That is though, not necessarily guaranteed through the current approaches in selecting the predictive relationships of the stochastic ground motion model parameters. Past studies have investigated the selection of predictive relationships for stochastic ground motion models so that compatibility with GMPEs is explicitly established, though a few shortcomings have been identified.

One of the main drawbacks of these past approaches, is that the physical characteristics of the resulting acceleration time-series are incorporated in the optimisation merely as constraints, something that requires significant experience in ground motion characterisation for proper definition of the optimisation problem, and can furthermore lead to synthetic time-histories with unrealistic properties for some seismicity scenarios. Another drawback is that past studies have focused only on the mean model characteristics and GMPE estimates. The approaches in the literature utilised only the mean of the predictive relationships of the stochastic ground motion model, whereas, more importantly, match only to the median spectral accelerations from GMPEs was investigated, ignoring any variability in their predictions. That is though, an important constraint since for seismic risk assessment applications, hazard compatibility is expressed in terms of both the mean and dispersion of some target IMs. The aim of this study is to address both shortcomings of past studies and propose a methodology to modify stochastic ground motion models for hazard compatibility, while preserving the main features of the resultant ground acceleration time-series.

This study aims to address the gaps identified in the GMSV literature described above through the steps summarised below:

1. Propose a novel validation approach that is based on statistics and information theory concepts to assess the overall similarity of the probability distributions of the studied IMs or EDPs for recorded and simulated ground motions.

2. Apply the proposed approach to a GMSV case-study using simple proxies for the spectral-shape and duration of the ground motions that are indicative of the peak and cyclic nonlinear structural response, respectively.

   As a summary, Figure 2.3 presents the GMSV studies performed so far (shown in black colour) in addition to the GMSV exercises carried out in this study (shown in red colour) and potential future studies (shown in green colour). The GMSV studies are placed on a line based on the level of complexity of the ground motion features or structural systems considered in the validation. In this context, moving to the right of the graph indicates a more complex ground motion feature or structural response.

   The engineering validation of simulated ground motions using spectral-shape and duration-related IMs is placed on the axis between validation in terms of conventional IMs (PGA, PGV, PGD, $S_a(T)$) and SDoF system's inelastic response. These spectral-shape and duration-related IMs are referred to as advanced IMs on the graph and throughout this study to distinguish from conventional IMs, as the former contain more information related to the structural response. Since SDoF and MDoF system's inelastic peak and cyclic response is strongly correlated with the shape of the elastic response spectra and the integral ground motion parameters (ground motion duration) respectively (Galasso et al., 2012), the advanced IMs are good proxies for validation of the more complex inelastic SDoF and MDoF system's response. The validation can further be performed in risk-based or performance-based terms, i.e., damage and loss, but this falls outside the scope of this study.

3. Propose a computationally efficient framework to modify stochastic ground motion models for specific seismicity scenarios with a dual goal of (i) matching a target IM for a specific structure (or range of structures) while (ii)

preserving desired trends and correlations in the physical characteristics of the resultant ground acceleration time-series.

4. Extend the proposed modification framework of stochastic ground motion models to (i) match the prescribed conditional hazard (mean and dispersion of IMs) for a specific site and structure (or range of structures) while (ii) preserving desired trends and correlations in the physical characteristics of the resultant ground acceleration time-series, including consideration of the variability of these characteristics.

5. Perform a validation of the proposed modification framework of the two previous steps by comparing the inelastic response of SDoF case-study structural systems to suites of hazard-compatible stochastic and recorded ground motions.
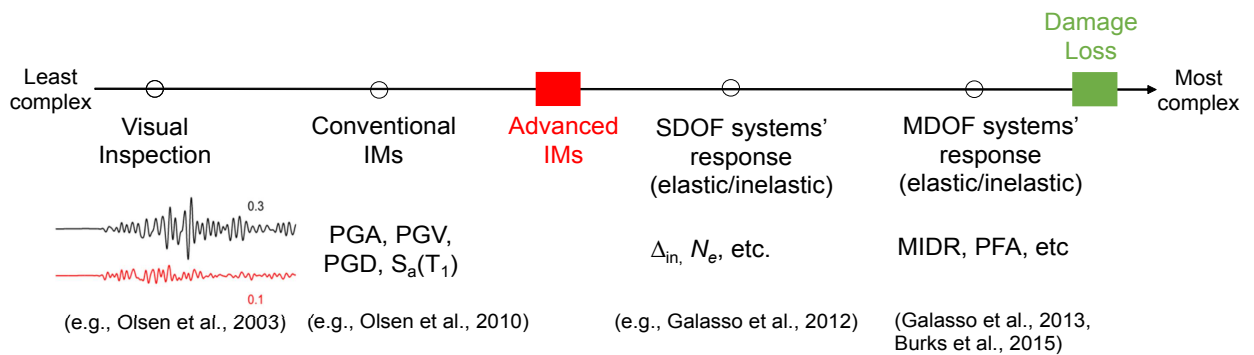
FIGURE 2.3: Summary of validation proxies used in past studies (black colour), in the current study (red colour) and future studies (green colour). The axis indicates increasing levels of complexity in the validation.

Chapter 3

Information theory measures for the engineering validation of ground motion simulations

Adapted from Tsioulou, A. and Galasso, C. (2018). Information theory measures for the engineering validation of ground motion simulations, *Earthquake Engineering & Structural Dynamics* **47**(4): 1095-1104.

3.1 Introduction

Recent advances in high-performance computing and understanding of complex seismic source features, path effects, and site effects, along with the scarcity or total absence of suitable recorded ground motion for specific earthquake scenarios (e.g., large magnitude crustal events recorded at close distance) have led to an increasing interest in ground motion simulation. Simulated ground motions are now considered a valuable supplement to recorded ground motions, fulfilling a variety of engineering needs (e.g., Bradley et al., 2017), such as seismic hazard assessment or assessment of seismic demand on structural and geotechnical systems through

response-history dynamic analysis, within the framework of PBEE. Among engineers, the general concern is that simulated ground motions may not be equivalent to real records in estimating seismic demand, and hence, in estimating the induced damage and loss to structures. Moreover, synthetic ground motions are not yet widely available in the engineering practice, especially in regions where seismogenic faults and characteristics and the regional velocity structure are not well established. On the other hand, in California, the recently released SCEC BBP (Maechling et al., 2015) provides scientists and engineers with a suite of open-source tools to compute and validate broadband synthetic ground motions by using several physics-based ground motion simulation methods. A GMSV TAG has been established by SCEC to develop and implement testing/rating methodologies via collaboration between ground motion modelers and engineering users. A similar effort is also being made in Italy, through a recently released web-repository (SYNTHESIS: SYNTHEtic SeISmograms database) containing synthetic waveforms for Italian scenario earthquakes coming from different simulation techniques (D' Amico et al., 2017).

This chapter proposes the use of information theory concepts and statistical hypothesis testing to quantitatively test a specific simulation method as well as to rate different simulation methods, consistently with the objectives of the SCEC GMSV TAG. We focus on the engineering validation of ground motion simulations in terms of spectral-shape and duration-related IMs. These metrics are common proxies for assessing the similarity of the expected nonlinear structural response and damage potential of simulated and recorded ground motions for many actual structural types. For illustrative purposes, the proposed testing/rating methodology is applied to the considered spectral-shape and duration-related IMs, obtained for different systems (i.e., structural periods) considering three broadband simulation methods: Graves and Pitarka's hybrid broadband method, the composite source model (CSM) deterministic method, and the EXSIM stochastic simulation method. These methods are used to compute simulations for several past Californian earthquakes. In fact, past events provide an important opportunity to test

the ability to use ground motion simulation methods to generate synthetic ground motions consistent (i.e., at the same locations) with those observed. Following a validation exercise, as the one presented in this article, end-users can decide regarding which model to use for their forward simulations of earthquake scenarios for which no observations exist. The confidence in using simulation methods beyond the tested limits must also be assessed considering the science behind each method (Goulet et al., 2015).

The next section briefly reviews some recent approaches and studies aiming at the engineering validation of ground motion simulation. This is followed by an introduction to the proposed validation approach. An illustrative implementation of the proposed approach is then presented and results of the application are finally discussed.

3.2 Engineering validation of ground motion simulations

A significant bulk of research has been developed in recent years to validate ground motion simulation methods for engineering applications, including (1) the comparison of simulations and recordings in terms of waveforms (e.g., by visual inspection), IMs and structural response for historical events; (2) the comparison in terms of IMs of simulations and predictions from empirical models (e.g., GMPEs), for both historical events and future scenarios; and (3) the comparison in terms of structural response of sets of simulations and recordings with similar elastic response spectra, consistently with guidelines for ground motion selection and scaling for building code applications. As a recent example of (1), Galasso et al. (2012) and Galasso et al. (2013) have investigated whether simulated ground motions are comparable to real records in terms of their nonlinear response in the domain of SDoF and MDoF linear and nonlinear building systems. The authors consider four historical earthquakes modeled by using the hybrid broadband ground motion

simulation method by Graves and Pitarka (2010). The validation exercise using various EDPs and formal statistical hypothesis testing indicates that, in most cases, the differences found in seismic demands produced by real and synthetic records are not significant, increasing the trust in the use of simulated motions for engineering applications. Rezaeian et al. (2015) propose a validation framework at the waveform level and considering three time-dependent validation metrics capturing the nonstationary features of intensity and frequency contents of earthquake ground motions. The proposed validation methodology is demonstrated by using examples of recorded and simulated ground motions from the Northridge event computed with the method by Graves and Pitarka (2010). As a recent example of (2), ground motion simulations computed by five different simulation methods implemented on the SCEC BBP v14.3 are compared with records from twelve earthquake events (western, central, and eastern Unites States and Japan), and published GMPEs in the recent studies by Dreger et al. (2015) and Goulet et al. (2015). The validation is performed in these studies with a focus on spectral accelerations. Four generic strike-slip and reverse scenarios for which GMPEs are considered to be well constrained by data are considered to compare spectral accelerations produced by simulation with predictions from the selected GMPEs. The results from this study indicate that the simulation methods provide reasonable estimates of spectral acceleration; however, it is suggested that additional research work is necessary to validate ground motion simulations for other applications by using different proxies and methods. As a recent example of (3), Burks et al. (2015) have investigated the validation of hybrid broadband simulations for use by structural engineers as input to nonlinear response-history analysis following the ASCE Standard (ASCE, 2010). The authors consider a set of "appropriate" hybrid broadband simulations (computed by using different simulation methods) and a comparable set of recordings to analyze a building in Berkeley, CA, and compare the predicted structural performance using the two sets. Results show that the structural behaviour resulting from recordings and simulations is similar, and most discrepancies are explained by differences in directional properties such as orientation of the maximum spectral response. These results suggest that

when simulations meet the criteria outlined for recordings in ASCE/SEI 7-10 and properties such as directionality are realistically represented, simulations provide useful results for structural analysis and design. Finally, Burks and Baker (2014) have developed a simulation validation framework combining the empirical models and similar spectra validation approaches (i.e., 2 and 3), proposing a list of parameters for the response of complex structural systems that can be used as proxies for the validation of ground motion simulations for engineering applications. The primary list of parameters includes correlation of spectral acceleration across periods, ratio of maximum to median spectral acceleration across all horizontal orientations, and the ratio of inelastic to elastic displacement, all of which have reliable empirical models against which simulations can be compared. The authors also describe secondary parameters, such as directivity pulse periods and structural collapse capacity, that do not have robust empirical models (so, the historical validation approach needs to be used) but are important for engineering analysis. The authors demonstrate the application of the proposed framework to example simulations computed by using a variety of simulation methods. Results show that each simulation method matches empirical models for some parameters and not others, indicating that all relevant parameters need to be carefully validated.

3.3   Proposed validation approach

As discussed, the validity of simulated ground motions is typically assessed based on criteria that are used to quantitatively evaluate the similarity of simulated and recorded time series in terms of IMs or structural response (i.e., EDPs). One common approach adopted by researchers involves the use of some goodness-of-fit criteria to compare how well the simulations match the ground motion records (e.g., Anderson, 2004; Olsen and Mayhew, 2010; Dreger et al., 2015). We propose the use of a novel validation approach based on information theory as a possible

testing/rating methodology for simulated ground motions to be used in engineering applications. Information theory concepts can be used to test the similarity of two datasets, which herein refers to the considered IMs (or EDPs) for simulated and recorded ground motions. Specifically, the relative entropy, also called the Kullback-Leibler divergence (Kullback, 1959) or cross entropy, is proposed here to measure the difference between two probability distributions p and q. In our applications, $p(\text{IM})$ represents the "true" distribution of a given IM (or EDP—IM), i.e., the empirical distribution of the IM (or EDP—IM) values derived from the recorded ground motions (for example, for a given past event or for a selected hazard-compatible ground motion set), while $q(\text{IM})$ represents a model or approximation of $p(\text{IM})$, i.e., the empirical distribution of the IM (or EDP—IM) values derived from the simulated ground motions (for the given past event or selected set and by using a given simulation method). The Kullback–Leibler divergence of $q(\text{IM})$ from $p(\text{IM})$, denoted D_{KL}, is a measure of the amount of information lost when $q(\text{IM})$ is used to approximate $p(\text{IM})$ and is defined as

$$D_{KL} = \int_{-\infty}^{+\infty} p(\text{IM}) \log_2 \left(\frac{p(\text{IM})}{q(\text{IM})} \right) \, d\text{IM} \tag{3.1}$$

If the logarithm is calculated in base 2, is expressed in terms of bits of information. D_{KL} has been used in earthquake engineering applications to compare the relative sufficiency of alternative IMs in predicting structural response (Jalayer et al., 2012).

In the context of ground motion simulation validation, the empirical distribution of the observed IMs estimated from the records, $p(\text{IM})$, and the empirical distribution of the IMs calculated from the simulated ground motions, $q(\text{IM})$, are constructed by using kernel density estimation (KDE) based on n available IM samples as

$$p(\text{IM}) \text{ or } q(\text{IM}) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} K \left(\frac{\text{IM} - \text{IM}_i}{h} \right) \tag{3.2}$$

where $K(.)$ is the chosen kernel and h is the kernel bandwidth. in this study the Epanechnikov kernel is employed, which is given by the following expression:

$$K(t) = \begin{cases} \frac{3}{4}\left(1 - t^2\right), & \text{if } -1 \leq t \leq 1 \\ 0, & \text{otherwise} \end{cases} \qquad (3.3)$$

with bandwidth h chosen to minimise the asymptotic mean integrated squared error between the KDE and the target distribution to be approximated (Scott and Sain, 2005). The entropy in Equation 3.1 can be then approximated by using the KDE estimates of p(IM) and q(IM) in Equation 3.2, with the 1-D integral calculated through numerical integration; for example, by using the trapezoidal rule.

Given that the estimated D_{KL} values are not standardised nor do they have an upper bound, it may be challenging assessing how extreme the calculated D_{KL} value is and drawing conclusions regarding the similarity of the two datasets. To overcome this, a procedure using the bootstrapping technique to construct an empirical distribution of D_{KL} is proposed and statistical hypothesis testing is used to assess the similarity of the two datasets from the observed D_{KL} value. This procedure is summarised below, where samples of IMs estimated from real records are called X and samples of IMs from simulations are referred to as Y for simplicity.

In the first step of the proposed procedure, we compute the Kullback-Leibler divergence D_{KL} between X and Y, referred to as $D_{KL,obs}$. In statistical hypothesis testing, the p-value for $D_{KL,obs}$ is the probability that when the null hypothesis is true, D_{KL} would be the same as or more extreme than the actual observed value. In this case, the null hypothesis is that X and Y have the same probability distribution. If this is true, then X and Y can be merged into a single sample and be treated as being one larger draw from the same distribution. This is the second step of the proposed procedure. The bootstrapping technique is then used in the third step of the proposed procedure to get the empirical distribution of

D_{KL} for each considered IM. To achieve this, two new vectors, X_{boot} and Y_{boot}, that have the same length as X and Y are drawn, by sampling observations at random from the combined X and Y data with replacement, so that observations from the original X sample may end up in the bootstrapped Y_{boot} sample and vice versa. For each set of new vectors, X_{boot} and Y_{boot}, the Kullback-Leibler divergence, $D_{KL,boot}$, can be calculated. The third step is performed many times, 1000 in this specific exercise, yielding 1000 samples of $D_{KL,boot}$. Finally, in the fourth step of the proposed procedure, the p-value for the observed $D_{KL,obs}$ is computed by finding the proportion of the 1000 $D_{KL,boot}$ samples that are more extreme (i.e., larger) than the $D_{KL,obs}$ value computed by using the original X and Y vectors. The obtained p-value represents the level of statistical significance in assuming that X and Y have the same probability distribution. Reasonable pass/fail thresholds can be applied to the obtained p-value results, for instance 95%, as in traditional hypothesis testing and in the illustrative application presented below. The 95th percentile of the empirical distribution corresponds to a 5% statistical significance level as only $D_{KL,obs}$ values that lie above the 95th percentile are significant. The hypothesis test is an one-sided test in this case.

The relative entropy can be interpreted as the expected value of the information gain about a certain IM rendered by a specific probability distribution (specific simulation methodology) in comparison to the target distribution (from recorded ground motions). The relative entropy would provide a mutual divergence quantification; and as such, it can be used to compare different models (distributions from different simulation methodologies) to the same target distribution. In this case it does not seem necessary to try to bound or to express a specific judgment on the absolute value of the entropy measured in bits of information. Nevertheless, the bootstrap analysis can still be used for estimating the empirical distribution for the expected value of information gain.

The proposed validation approach distinguishes itself from the past studies and other proposed goodness-of-fit criteria by assessing the overall similarity of the probability distributions of the studied IMs for recorded and simulated ground

motions. Thus, it does not just provide a paired comparison (i.e., at the same recording locations, for historical events) between the recorded and simulated IM datasets in terms of mean and standard deviation of their distributions. This represents a useful tool for the engineering validation of simulated ground motions in terms of the nonlinear structural demands or expected loss for a portfolio of structures (or infrastructure) where an overall as opposed to a paired comparison of the records and simulations is of interest, for example for catastrophe risk modeling purposes (Sørensen and Lang, 2015). The proposed approach can also be used to measure the similarity of the distributions of seismic response to sets of simulations and recordings matching a target (elastic) response spectrum mean and variance, consistently with the current practice in ground motion selection and scaling for building code applications (Jayaram et al., 2011).

3.4   Illustrative application

The illustrative implementation of the proposed validation approach considers ground motion simulations generated by the SCEC BBP v13.5 and 13.6 using three broadband, finite-source simulation methods: the hybrid approach by Graves and Pitarka (2010), referred to as G&P (2010); the deterministic CSM approach (Zeng et al., 1994), herein referred to as CSM; and a band-limited stochastic white-noise method called EXSIM (Motazedian and Atkinson, 2005) based on previous work by Boore (2009). G&P (2010), widely used in past validation studies, is a hybrid broadband (0-10 Hz) ground motion simulation method that combines a physics-based deterministic approach at low frequency ($f \leq 1$ Hz) with a semistochastic approach at high frequency ($f > 1$ Hz). The low- and high-frequency waveforms are computed separately and then combined to produce a single time-history through a matching filter. The use of different simulation approaches for the different frequency bands results from the seismological observation that source radiation and wave propagation effects tend to become stochastic at frequencies of about 1 Hz and higher, primarily reflecting the relative lack of knowledge about

these phenomena's details at higher frequencies. The CSM method uses a kinematic source model for rupture on a finite fault. This source is propagated to the station by using a flat-layered velocity model, scattering, and attenuation that can be measured from independent seismological observations. The objective is to reproduce the wave propagation entirely within the constraints of the measured velocity and Q structure (Anderson, 2015). As described in Atkinson and Assatourians (2015), EXSIM divides the fault plane in an array of subsources, each of which is treated as point source. The ground motion from each subsource is treated as random Gaussian noise of a specified duration. The duration of motion for each subsource comes from the source duration plus the path duration.

The simulations used here are computed by the G&P (2010) and EXSIM methods as implemented on SCEC BBP v13.6 and the CSM method implemented on SCEC BBP v13.5, as the CSM method on BBP v13.6 is only available for validation against GMPEs and not against recorded events (personal communication with C.A. Goulet, 2016). The four historical events considered herein are the 1989 M 6.8 Loma Prieta, 1992 M 7.2 Landers, 1986 M 6.1 North Palm Springs, and 1994 M 6.7 Northridge. For each simulation method and each earthquake event, 50 different simulations were obtained based on the same number of realisations of different kinematic source models (e.g., amount of slip, slip velocity, rise time), yielding a total of 50 realisations of ground motion simulations per station. The validation is performed on the average results from those 50 realisations. Moreover, as explained in Goulet et al. (2015), the simulation methods do not focus much on near-surface effects coming from nonlinear site response. In fact, a single generic site profile with a $V_{s,30}$ value of 863m/s was used for all the simulations. To make the simulations comparable to the as-recorded site conditions, empirical site effect models should be applied increasing the epistemic uncertainty of the problem. Therefore, to reduce the uncertainties arising from applying site amplification factors, this study only includes recordings from sites with $V_{s,30}$ close enough to the $V_{s,30}$ used for the BBP simulations (863m/s). Stations with $V_{s,30}$ values greater than 700 m/s are identified to be of "similar" $V_{s,30}$ to the reference value

used in the simulations (SCEC GMSV TAG). This leads to datasets of relatively small size (less than 10 ground motions) for each considered earthquake event. The estimation of empirical probability distributions from such small datasets can result in unreliable values of D_{KL}. In this case, due to the purely illustrative nature of the application presented here, all the events for each simulation method were combined, focusing on assessing the overall performance of a given simulation method rather than the specific performance for a given earthquake event.

3.4.1   Considered intensity measures

An IM is a single ground motion parameter (scalar IM) or set of ground motion parameters (vector IM), which are representative of the earthquake damage potential with respect to a specific class of engineered systems. Typical engineering applications (e.g., performance-based assessment and design) require the choice of an IM which is suitable to predict the response of the system with the smallest scatter ("efficiency") and providing a significant amount of information, downgrading the effect of other seismological parameters ("sufficiency") to predict the response quantities involved in the performance objectives. In addition, many researchers have investigated other IM selection criteria, related for example to "hazard computability", "proficiency", and "practicality" (Padgett et al., 2008). Conventional IMs, including PGA, PGV, PGD, and (pseudo-) spectral acceleration at the initial fundamental period (for a damping ratio of 5%), $S_a(T_1)$, are the most commonly used IMs. In general, PGA and $S_a(T_1)$ poorly predict the structural response of mid-rise to high-rise moment resisting frames, although the latter IM sufficiently captures the elastic behaviour of first-mode dominated MDoF systems, especially in the case of low to moderate fundamental periods (Shome et al., 1998). However, the behaviour of highly nonlinear structures (sensitive to periods greater than T_1 due to period lengthening) or structures dominated by higher-mode periods (less than T_1) are not very well represented by utilising $S_a(T_1)$, due to the lack of information on the spectral-shape provided by this IM.

Therefore, it has become essential implementing advanced IMs that account for the elongated periods and/or consider nonlinear demand-dependent structural parameters. Kazantzi and Vamvatsikos (2015) and Kohrangi et al. (2017) among several others have investigated the adequacy of numerous advanced scalar IMs that take into consideration the aforementioned parameters.

For the illustrative application presented here, we then use the advanced scalar IM proposed by Bojórquez and Iervolino (2011). This IM, denoted as I_{N_p}, is based on $S_a(T_1)$ and the parameter N_p, and is defined as

$$I_{N_p} = S_a(T_1)N_p^\alpha \tag{3.4}$$

where the parameter α is taken as $\alpha = 0.4$ based on the tests conducted by the authors and N_p is defined as

$$N_p = \frac{S_{a,avg}(T_1, ..., T_N)}{S_a(T_1)} = \frac{\left[\prod_i^N S_a(T_i)\right]^{1/N}}{S_a(T_1)} \tag{3.5}$$

T_N corresponds to the maximum period of interest and lies within a range of 2 and $2.5T_1$, as suggested by the authors. In this study, I_{N_p} is computed for four different fundamental periods T_1: 0.5, 1, 2, and 4 s. For the N_p computation, 3 periods are considered: T_1, $1.5T_1$ and $2T_1$. Figure 3.1 shows an example of scatter plot for the structural demand in terms of inelastic displacement versus I_{N_p} for an inelastic SDoF with $T_1 = 1$s, a strength reduction factor (R_μ) equal to 8 (typical of severely inelastic structures), and a non-degrading elastic-plastic with positive strain-hardening, $\alpha = 3\%$, model (EPH). 121, 2-component, ground motion records from the Northridge earthquake have been used; see Galasso et al. (2012) for details. For the considered case, I_{N_p} outperforms all the conventional and advanced scalar IMs in terms of all the criteria for optimal IMs.

Integral (i.e., duration-related) IMs, such as the Arias intensity or significant ground motion duration, are commonly used, but they are considered to be related
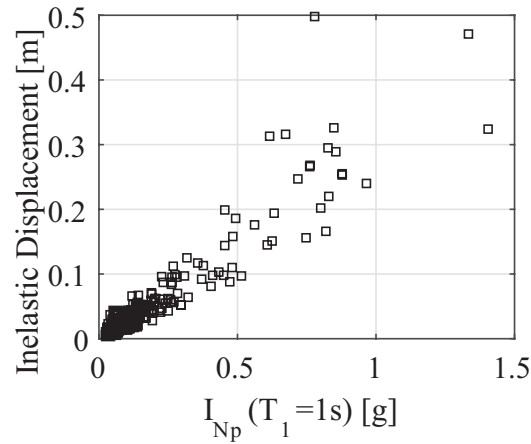
FIGURE 3.1: Example of scatter plot of the inelastic displacement versus I_{N_p} for an inelastic SDoF with $T_1 = 1$s, $R_\mu = 8$, and EPH model with, $\alpha = 3\%$ (Northridge earthquake). See Galasso et al. (2012) for details.

more to the cyclic energy dissipation rather than to the peak structural response. In fact, some studies (e.g., Iervolino et al., 2006) investigated how ground motion duration-related parameters affect nonlinear structural response and particularly structural collapse (e.g., Raghunandan and Liel, 2013; Chandramohan et al., 2016). It is widely acknowledged that, generally, spectral ordinates are sufficient (i.e., duration does not add much information) if one is interested in the ductility demand, while duration-related measures do play a role only if the hysteretic structural response is to be assessed; i.e., in those cases in which cyclic deterioration and cumulative damage potential of the earthquake are of concern. Chandramohan et al. (2016) highlight the need to consider ground motion duration, in addition to intensity and response spectral-shape, in regions where significant hazard due to long duration shaking exists, such as locations susceptible to large magnitude, subduction zone earthquakes. Finally, integral IMs are also important for several other engineering applications, for example, in geotechnical engineering, such as landslide and liquefaction risk assessment. Therefore, the engineering validation of simulated ground motions in terms of duration-related parameters is also of significant importance.

The term duration is typically used to identify only the portion of a record in which the ground motion amplitude can potentially cause damage to engineering

and geotechnical structures. Several definitions are proposed to this aim; the most commonly used one is the significant duration, introduced by Trifunac and Brady (1975), defined as the time interval over which the integral of the square of the ground acceleration (Husid plot, Husid, 1969) is within a given range of its total value. Usually, this range is between 5% and 95% (as in this study), denoted as D_{5-95}, or between 5% and 75%.

Finally, Cosenza and Manfredi (1997) introduced the dimensionless I_D-factor defined as

$$I_D = \frac{\int_0^{t_E} a^2(t)\, dt}{PGAPGV} \tag{3.6}$$

which has proven to be a good proxy for cyclic structural response (Manfredi, 2001). Here, $a(t)$ is the acceleration time-history and t_E is the complete duration of the ground motion (length of the record). Figure 3.2 shows an example of scatter plot for the structural demand in terms of equivalent number of cycles (N_e – i.e., the cumulative hysteretic energy normalised with respect to the largest cycle) versus I_D and D_{5-95} for an inelastic SDoF with $T_1 = 1$s, a strength reduction factor (R_μ) equal to 2 (typical of mildly inelastic structures), and a degrading/evolutionary model (ESD) comprising a negative strain-hardening (i.e., a softening branch), $-\alpha = 10\%$, and a residual strength equal to 10% of the maximum strength. The simple peak-oriented model is considered to account for the cyclic stiffness degradation. Also in this case, 121, 2-component, ground motion records from the Northridge earthquake have been used; see Galasso et al. (2012) for details. For the considered case, D_{5-95} outperforms other integral IMs (including I_D) in terms of all the criteria for optimal IMs. However, the authors found that this result is dependent on the considered level of nonlinearity, with I_D outperforming the other integral IMs in the case of severely inelastic structures (i.e., $R_\mu \geq 4$). Therefore, both metrics are kept in our validation exercise.

It is worth noting that the main objective of the BBP validation exercise presented
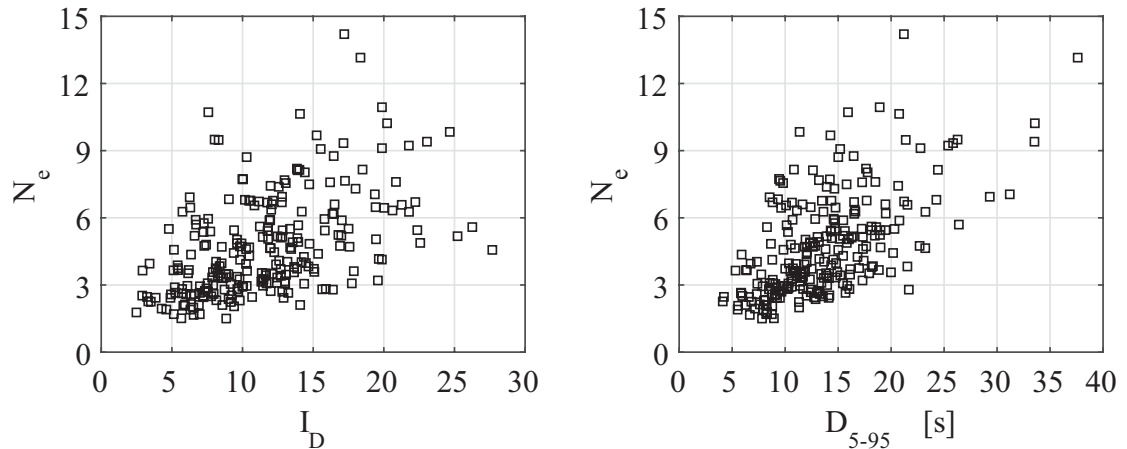
FIGURE 3.2: Example of scatter plot of the equivalent number of cycles versus I_D (left) and D_{5-95} (right) for an inelastic SDoF with $T_1 = 1$s, $R_\mu = 2$, and ESD model with, $\alpha = 10\%$ (Northridge earthquake). See Galasso et al. (2012) for details.

in Dreger et al. (2015) was to validate elastic spectral response by using the BBP v14.3. The parameters proposed in our study - as well as those introduced in Burks and Baker (2014)- are intended as a supplement, not a replacement, to that validation. It is understood that many other metrics would be necessary to fully assess the simulation methods' ability to produce reasonable ground motions as a whole. An important property of the proposed validation parameters is that they are hazard computable, i.e., empirical models or GMPEs exist (e.g., for I_D and D_{5-95} see Iervolino, Giorgio, Galasso and Manfredi, 2010) or may be easily derived (e.g., for I_{N_p} Bojórquez and Iervolino, 2011) combining existing tools and can be used as a baseline comparison against simulations for a very broad range of conditions, including future earthquake scenarios.

3.5   Validation results

All ground motions (recorded and simulated) selected for each simulation method are used as input to compute the selected IMs described above. Only the horizontal components of ground motions (i.e., north-south, and east-west) are used, while the vertical component is neglected, consistently with other studies.

Table 3.1 summarises the D_{KL} values for all considered IM distributions for each of the three simulation methods implemented on BBP v13.5 and 13.6, as discussed above. The mean of the 50 values of IMs obtained from the same number of realizations for the two horizontal components at each station is computed and then combined into an "average" value by using the geometric mean. In this case, as explained above, the number of data per event is limited, and thus, the D_{KL} values are estimated for each simulation method by grouping the simulations from all the earthquake events together and compare them with the records. This allows the comparison of the performance of the three simulation methods in estimating the probability distributions of spectral-shape and duration-related IMs. As discussed above, the estimated D_{KL} value is a measure of the amount of information loss incurred from using the distribution of simulated IMs to approximate the "true" distribution of recorded IMs. Thus, when comparing two or more ground motion simulation methods, the method yielding the smallest D_{KL} value performs best in matching the distribution of recorded IMs; these cases are shown in bold font in Table 3.1.

As explained in a previous section, statistical hypothesis testing can be performed by using the bootstrapping technique to assess how large the observed D_{KL} values are in each case and draw conclusions regarding the similarity of the two datasets, recorded and simulated, for a given simulation method. For the hypothesis tests yielding a p-value less than 0.05 (5%), there is strong evidence to reject the null hypothesis, and thus, the differences in the IM probability distributions from simulations and real records can be considered statistically significant. This means that the observed D_{KL} value lies above the 95th percentile of the empirical cumulative distribution function (CDF) for D_{KL}. These cases are highlighted with the grey color in Table 3.1. For p-values greater than 0.05 (5%), there is not sufficient evidence to reject the null hypothesis, meaning that the differences in the IM probability distributions from simulations and real records are not statistically significant. In this case, the observed D_{KL} values fall below the 95th percentile of the CDF for D_{KL}.

TABLE 3.1: D_{KL} values for spectral-shape and duration-related IMs for each
simulation method

IM	Simulation Method	$D_{KL,obs}$ value			
		T_1=0.5 s	T_1=1 s	T_1=2 s	T_1=4 s
I_{N_p}	CSM	0.33	0.37	0.19	0.27
	EXSIM	**0.23**	**0.18**	0.17	**0.24**
	G&P (2010)	0.31	0.22	**0.06**	0.33
I_D	CSM	0.56			
	EXSIM	0.54			
	G&P (2010)	**0.40**			
D_{5-95}	CSM	1.42			
	EXSIM	3.42			
	G&P (2010)	**0.14**			

The results in Table 3.1 reveal that the performance of the simulation methods
in estimating spectral-shape proxies greatly depends on the advanced IM and
period considered. In particular, CSM method performs worse than the other
two methods in estimating I_{N_p} across all periods. G&P (2010) method performs
best in estimating I_{N_p} only for 2s period. EXSIM method gives the most accurate
predictions for 0.5, 1 and 4 s periods for I_{N_p}. Overall, EXSIM method outperforms
the other two, having the highest number of best performances for the spectral-
shape-related IMs considered. On the other hand, there is a single best performing
simulation method for all the duration-related IMs examined. Based on the results
in Table 3.1, the G&P (2010) method results in the most accurate predictions of
the I_D and D_{5-95} distribution.

With respect to the results of the hypothesis testing, all the observed D_{KL} values
are within the non-rejection region established through bootstrapping, except for
the D_{KL} values for D_{5-95} calculated from CSM and EXSIM methods. To shed
further light on this result, histograms of the D_{5-95} samples from recorded and
simulated ground motions are plotted in Figure 3.3 for the CSM and EXSIM
methods. The white bars correspond to the simulated IMs, whereas the grey bars

refer to the IMs from recorded time-histories. A histogram plot of I_{N_p} samples at 2 s period for G&P (2010) that corresponds to a case of no rejection is presented in Figure 3.4 for comparison. It is evident that large differences exist between the resulting histograms (and derived probability density functions from KDE) for the simulated and recorded D_{5-95} values for the CSM and EXSIM methods as shown in Figure 3.3. By contrast, the histograms of the simulated and recorded I_{N_p} samples at 2 s period for G&P (2010) are very similar, making it the best performing method for this specific validation metric, as shown in Figure 3.4.
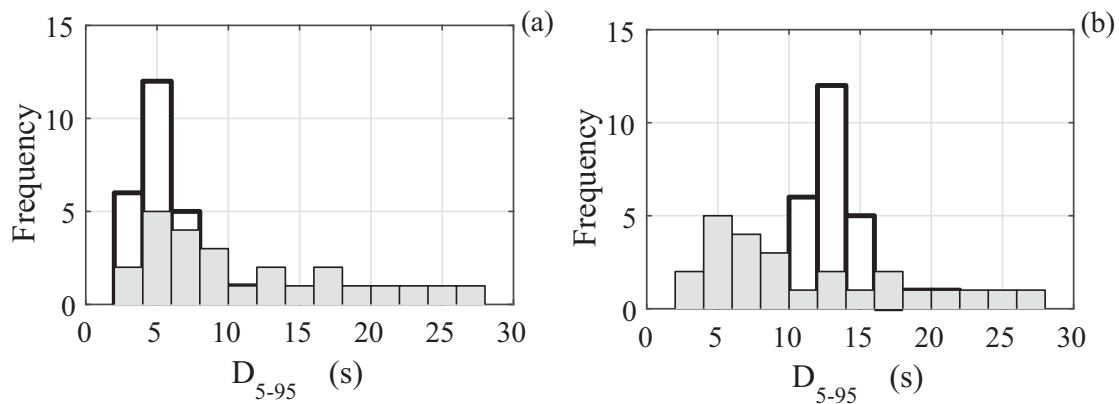
FIGURE 3.3: Histograms of the D_{5-95} samples from recorded (gray bars) and simulated (white bars) ground motions for (a) the CSM, and (b) the EXSIM methods.
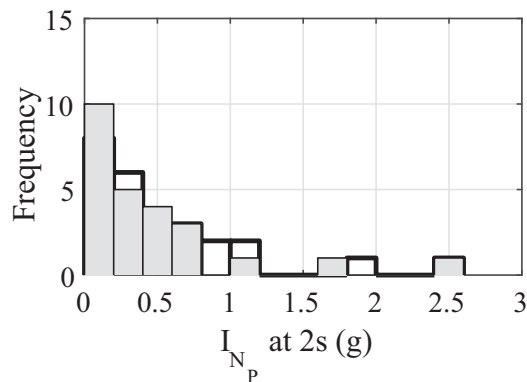
FIGURE 3.4: Histograms of the I_{N_p} samples at 2 s period from recorded (gray bars) and simulated (white bars) ground motions for the G&P (2010) method.

In addition to the visual comparison in Figures 3.3 and 3.4, the proposed validation approach was compared with the more commonly used hypothesis testing method. Standard t-tests for the equality of means of the studied IMs were performed (e.g., Galasso et al., 2012, 2013) and the results were consistent with the results of the

proposed approach based on D_{KL} values. However, since the latter approach evaluates the match between the full distributions (i.e., probability density functions) of the two data sets, it is also necessary to perform hypothesis tests on higher moments of the distribution and not just the mean (first moment) to compare the results. Hence, F-tests for the equality of variances of the two datasets were performed in addition to the t-tests. The results showed that there were several rejections for the F-test that did not seem to be justified by the empirical IM distributions (histogram plots of the IMs for these rejection cases were created to visually assess the equality of variances, similar to Figures 3.3 and 3.4). On the other hand, the method proposed here seems to accurately detect differences in the full distribution of the IMs as shown in Figures 3.3 and 3.4.

3.6 Conclusions

The design of new structures or the assessment of existing ones may be complicated by the inherent rareness or total absence of suitable recorded ground motions for the earthquake scenarios that dominate the seismic hazard at a given site. Therefore, broadband synthetic records may be an attractive option as input to NLDA, if an accurate and transparent engineering validation for the considered simulation method is carried out. To this aim, the focus of this chapter was on the design of such a validation exercise by proposing a novel quantitative approach for testing/rating ground motion simulation methods, based on information theory measures coupled with statistical hypothesis testing. The proposed approach assesses the overall similarity of the probability distributions of the recorded and simulated IMs and uses the relative entropy to quantify their distance. Statistical hypothesis testing relying on the bootstrapping technique is then used to test the significance of the estimated distance. Ultimately, the approach can be used to rank the performance of different ground motion simulation methods and it is part of a larger, longer-term, and broader ongoing plan for the validation of simulated ground motions for engineering applications.

The application of the proposed evaluation criteria was demonstrated by using a group of ground motion simulations computed by Graves and Pitarka's (2010), CSM, and EXSIM simulation methods implemented on v13.5 and 13.6 of the SCEC BBP for four past earthquakes: 1989 M 6.8 Loma Prieta, 1992 M 7.2 Landers, 1986 M 6.1 North Palm Springs, and 1994 M 6.7 Northridge. The illustrative application considers three ground motion IMs: one spectral-shape and two duration-related that have been shown to be optimal proxies for the (nonlinear) seismic response of actual buildings and geotechnical systems. The proposed validation metrics are hazard computable and their empirical models can be used as baseline for comparison for future earthquake scenarios. The list of IMs considered in this study is not exhaustive and can be used to supplement other validation metrics encountered in the literature. Finally, for the specific simulated ground motion data considered in the illustrative example, the EXSIM and Graves and Pitarka's (2010) ground motion simulation methodologies perform best in predicting the probability distributions of the spectral-shape and duration-related IMs, respectively.

It is worth noting that since broadband simulation methods evolve very fast, the intent here is not to provide a definite judgment about the specific simulation methods, but rather to illustrate the proposed validation metrics and approaches and discuss possible outcomes. Indeed, these types of validation exercises can highlight the similarities and differences between simulated and recorded ground motion for a given simulation method. The similarities should provide confidence in using the simulation method for engineering applications, while the discrepancies should help in improving the generation of synthetic records.

The proposed validation approach (Tsioulou and Galasso, 2018) has been applied to perform validation of ground motion simulations in terms of seismic response of skewed bridges by Galasso et al. (2018).

Acknowledgments

Chapter 4

Modification of stochastic ground motion models for matching target intensity measures

Adapted from Tsioulou, A., Taflanidis, A. A. and Galasso, C. (2018a). Modification of stochastic ground motion models for matching target intensity measures, *Earthquake Engineering & Structural Dynamics* **47**(1): 3-24.

4.1 Introduction

The growing interest in PBEE (Bozorgnia and Bertero, 2004; Goulet et al., 2007) and in simulation-based, seismic risk assessment approaches (Au and Beck, 2003; Jensen and Kusanovic, 2014) has increased in the past decades the relevance of ground motion modeling techniques. These techniques describe the entire time-series of seismic excitations, providing a characterisation appropriate for dynamic time-history analysis. Undoubtedly the most popular methodology for performing this task for seismic risk assessment (or seismic design) applications is the selection and modification of real (i.e., recorded from past events) ground motions

based on a target IM level (Lin et al., 2013; Katsanos et al., 2010), e.g., an elastic pseudo-acceleration response spectrum. For seismic risk assessment such modification is performed for specific seismicity scenarios (typically defined through moment magnitude and source-to-site distance) contributing to the seismic hazard for the chosen site, with the target IM commonly (McGuire, 2004) derived through GMPEs (Stewart et al., 2016; Campbell and Bozorgnia, 2008).

An alternative philosophy for describing seismic excitations is to use simulated ground motions (Jalayer and Beck, 2008; Galasso et al., 2013). A specific modeling approach for the latter which has been steadily gaining increasing attention by the structural engineering community (Vetter and Taflanidis, 2014; Broccardo and Der Kiureghian, 2015) is the use of stochastic ground motion models (Rezaeian and Der Kiureghian, 2010; Gavin and Dickinson, 2010; Yamamoto and Baker, 2013; Vlachos et al., 2016; Boore, 2003; Atkinson and Silva, 2000). These models are based on modulation of a stochastic sequence, through functions (filters) that address spectral and temporal characteristics of the excitation. The parameters of these filters are related to seismicity (e.g., moment magnitude and rupture distance) and site characteristics (e.g., shear wave velocity for soil profile) through predictive relationships (Rezaeian and Der Kiureghian, 2010; Boore, 2003). Sample ground motions for any desired seismicity scenario can be generated by determining the parameters of the stochastic ground motion model through these predictive relationships and by using a sample stochastic sequence.

The essential component of stochastic ground motion models is the development of the associated predictive relationships, and various approaches have been established to accomplish this, with main representatives being record-based and physics-based models. Record-based models (also known as site-based) are developed by fitting a preselected "waveform" to a suite of recorded regional ground motions (Rezaeian and Der Kiureghian, 2010; Vlachos et al., 2016; Papadimitriou, 1990). On the other hand, stochastic physics-based models (also known as source-based) rely on physical modeling of the rupture and wave propagation mechanisms (Boore, 2003; Atkinson and Silva, 2000). Emphasis in this study will be on the

former models, though the techniques discussed can be extended to any type of stochastic ground motion model.

An important concern related to the use of stochastic ground motion models for structural engineering applications is the fact that through current approaches in selecting their predictive relationships, compatibility to the seismic hazard for specific structures and sites is not necessarily obtained (Rezaeian and Der Kiureghian, 2010) (this is also shown in the illustrative example considered later). Although validation of these models is frequently performed by comparison of their spectral acceleration outputs to GMPEs (Rezaeian and Der Kiureghian, 2010; Yamamoto and Baker, 2013), the match to GMPEs is not explicitly incorporated in the predictive relationships development. Such a match to some desired GMPE (or target IMs in general) is though important for subsequent use of the stochastic ground motion models to describe the seismic hazard. Take for example the recent FEMA P-58 (FEMA, 2012) guidelines for seismic performance assessment of structures; the scenario-based description of the seismic hazard requires match of the median response to the one described by a GMPE (for the specific seismicity scenario examined). Similarly, the intensity- and time-based descriptions in FEMA P-58 require compatibility with seismic hazard curves which are ultimately defined through use of GMPEs (Petersen et al., 2008).

This realisation has motivated researchers to investigate the selection of predictive relationships for stochastic ground motion models so that compatibility with GMPEs is explicitly established (Scherbaum et al., 2006). The formulation introduces an explicit optimisation for matching the median predictions of the ground motion model to the spectral acceleration estimates of GMPEs, while maintaining physics-based principles or the matching to trends from real ground motions as an optimisation constraint, in an attempt to preserve desired ground motion characteristics. Vetter et al. (2016) recently extended the work of Scherbaum et al. (2006) by providing a versatile and computationally efficient approach, leveraging surrogate modeling principles, for tuning stochastic ground motion models to establish compatibility with the median GMPE predictions for a range of structural

periods and seismicity scenarios of interest. One of the main drawbacks of this tuning approach, though, is that the physical characteristics of the resulting acceleration time-series are incorporated in the optimisation merely as constraints, something that requires significant experience in ground motion characterisation for proper definition of the optimisation problem and can furthermore lead to synthetic time-histories with unrealistic properties for some seismicity scenarios.

The current study addresses this critical shortcoming and looks at the modification of stochastic ground motion models for specific seismicity scenarios with a dual goal of (i) matching a target IM for a specific structure (or range of structures) while (ii) preserving desired trends and correlations in the physical characteristics of the resultant ground acceleration time-series. This is ultimately formulated as a multi-objective optimisation problem. The first objective is to minimise the discrepancy between the median ground motion output and the target IM for a given seismicity scenario. Any desired IM can be used for this purpose with only requirement to have a corresponding seismicity scenario. For instance, if the target IM is derived through PSHA, a corresponding seismicity scenario (or "design earthquake") can be derived through the disaggregation of seismic hazard (McGuire, 2004) for a given hazard level. The second objective is to establish the smallest deviation from the model characteristics suggested by existing predictive relationships. This second objective aims at maintaining regional physical characteristics and parameter correlations with respect to existing predictive relationships. The approach differs significantly from Vetter et al. (2016); rather than tuning the ground motion for hazard-compatibility ignoring any existing predictive relationships, goal here is the minimum modification of the existing relationships that will yield the desired compatibility. This is ultimately posed as a multi-objective problem, to better investigate the compromise between the two different objectives, and for efficiently solving it a surrogate modeling approach is adopted, similar to that of Vetter et al. (2016). A surrogate model (i.e., metamodel) is trained based on an initial database of ground motion simulations, and ultimately provides a highly efficient approximation for the spectral acceleration predictions of the stochastic

ground motion model. The surrogate model is then leveraged to solve the opti-
misation problem. Emphasis is also placed here on the selection of the database
to inform the metamodel development, which constitutes a significant advance-
ment over the approach by Vetter et al. (2016). Blind search and gradient-based
approaches are considered for the multi-objective optimisation and the relative
computational benefits of each are explored.

In the next section, the general problem of developing simulated ground motions
compatible with target IMs is defined and then specific aspects of the framework
are discussed in detail.

4.2   Problem formulation

Consider a stochastic ground motion model that provides acceleration time-histories
$\ddot{\alpha}(t|\boldsymbol{\theta},\boldsymbol{w})$ by modulating a Gaussian white-noise sequence, $\boldsymbol{w}$, through appropri-
ate time/frequency functions that are parameterised through the n_θ-dimensional
model parameter vector $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_{n_\theta}] \in \mathbb{R}^{n_\theta}$. This vector completely defines
the model and is typically composed of various excitation properties such as Arias
intensity, strong ground motion duration or parameters related to frequency char-
acteristics of the ground motion. A specific example for such a model, the one
used in the illustrative example in Section 4.5, is provided in Appendix A. This
particular record-based model efficiently addressed both temporal and spectral
nonstationarities. The former is established through a time-domain modulating
envelope function, whereas the latter is achieved by filtering a white-noise process
by a filter with characteristics that vary in time.

Synthetic time-histories can be created by relating $\boldsymbol{\theta}$ to seismicity and local site
properties through predictive relationships. The vector of these properties, called
seismological parameters, is denoted as $\boldsymbol{z}$. Common characteristics used for $\boldsymbol{z}$
(Rezaeian and Der Kiureghian, 2010; Boore, 2003) include the fault type F, the
moment magnitude, M, the rupture distance, R, and the shear wave velocity in

the upper 30 meters of soil, $V_{s,30}$. For record-based models the standard approach for development of these predictive relationships (Rezaeian and Der Kiureghian, 2010; Medel-Vera and Ji, 2016) relies on first matching the waveform characteristics to recorded ground motions (i.e., identify first $\boldsymbol{\theta}$ for each of the recorded ground motions in a given database) and then performing a regression to relate $\boldsymbol{\theta}$ to $\boldsymbol{z}$. This leads ultimately to a probabilistic regression model for $\boldsymbol{\theta}$ with mean predictive relationship $\boldsymbol{\mu_\theta}(\boldsymbol{z})$ that is dependent on $\boldsymbol{z}$ and some associated uncertainty characterization U, identified from the residuals of the regression, that is independent of $\boldsymbol{z}$. Typically, this is performed by first transforming the problem to the standard Gaussian space through a nonlinear mapping for each component θ_i. The transformed Gaussian vector is denoted $\boldsymbol{v}(\boldsymbol{\theta})$ herein. Approach ultimately leads to a Gaussian probability model $\boldsymbol{v} \sim N(\boldsymbol{\mu}(\boldsymbol{z}), \boldsymbol{\Sigma})$ with mean $\boldsymbol{\mu}(\boldsymbol{z})$ and covariance matrix $\boldsymbol{\Sigma}$. In this case, the uncertainty characterization U corresponds to the covariance matrix $\boldsymbol{\Sigma}$ and to the fact that probability model for $\boldsymbol{z}$ is identified as Gaussian. Appendix A includes more details for a specific ground motion model (Rezaeian and Der Kiureghian, 2010). Note that a similar description can be established for physics-based models. In this case the predictive relationships $\boldsymbol{\mu_\theta}(\boldsymbol{z})$ are obtained through rupture and wave propagation principles (Boore, 2003; Boore and Thompson, 2015), whereas the uncertainty characterization U can be established by assigning probability models for $\boldsymbol{\theta}$ through an epistemic uncertainty treatment (Vetter and Taflanidis, 2014; Atkinson, 2008).

As discussed in Section 4.1, this formulation for the predictive relationships of stochastic ground motion models, prioritising a match to regional trends, provides synthetic ground motions whose output IMs do not necessarily match hazard-compatible IMs (e.g., as derived from GMPEs). For this purpose, a modification of the model parameter vector $\boldsymbol{\theta}$ is proposed for specific seismicity scenarios defined by $\boldsymbol{z}$ with objective to (i) match a target IM vector, while (ii) maintaining similarity to the predictive relationships already established for the model. Equivalently, this can be viewed as identifying the model characteristics $\boldsymbol{\theta}$ that are closest to the established model $\boldsymbol{\mu_\theta}(\boldsymbol{z})$ (considering, when available, any additional information

provided by U) and also match the intended hazard (described through some IM). The IM vector may include different response quantities of interest, for example, (i) direct characteristics of the ground motion, such as PGA, PGV and PGD; (ii) elastic and inelastic spectral responses for different periods of an SDoF oscillator; or (iii) more complex spectral or ground motion related quantities proposed by different researchers (Bojórquez and Iervolino, 2011; Cosenza and Manfredi, 1997; Cordova et al., 2000). The target for most of these IMs can be described through a GMPE (e.g., Campbell and Bozorgnia, 2008). However, this is not necessary; any IM description can be used, with only requirement that a corresponding seismicity scenario is defined. Note that if match to spectral responses (i.e., a spectral plot) is of interest, then a range of structural periods for which the match is established needs to be determined.

To formalise these concepts mathematically, let, $Y_i(\boldsymbol{z}); i = 1, \ldots, n_y$ denote the target response quantities of interest and $Y_i^m(\boldsymbol{\theta})$ the median predictions for the same quantities provided through the stochastic ground motion model. The median predictions are obtained through the following process:

Step 1: Generate n_w sample acceleration time-histories for different white-noise sequences $\ddot{\alpha}^k(t|\boldsymbol{\theta}, \boldsymbol{w}^k); k = 1, \ldots, n_w$.

Step 2: For each sample evaluate the responses of interest. For spectral quantities this will entail numerical simulation of SDoF responses.

Step 3: Estimate the statistics (median) over the established sample-set.

The modification problem is ultimately formulated as bi-objective optimisation problem:

$$\boldsymbol{\theta} = \arg\min \{F_1(\boldsymbol{\theta}|\boldsymbol{z}), F_2(\boldsymbol{\theta}|\boldsymbol{z})\} \tag{4.1}$$

The first objective F_1 corresponds to a measure of the discrepancy from the chosen target. One choice for this measure is the average weighted square error given by

$$F_1\left(\boldsymbol{\theta}|\boldsymbol{z}\right) = \frac{1}{n_y}\sum_{i=1}^{n_y}\gamma_i^2\left(Y_i(\boldsymbol{z}) - Y_i^m(\boldsymbol{\theta})\right)^2 \tag{4.2}$$

with γ_i corresponding to the weights prioritising the match to different IM components. A typical selection for γ_i is $1/Y_i(\boldsymbol{z})$, then the quantity in Equation 4.2 corresponds to the average squared relative error. Alternative formulations for this first objective, facilitating perhaps a better physical intuition, are the average absolute error or the maximum absolute error over the different IMs, given, respectively by,

$$F_{1r}\left(\boldsymbol{\theta}|\boldsymbol{z}\right) = \frac{1}{n_y}\sum_{i=1}^{n_y}\gamma_i\left|Y_i(\boldsymbol{z}) - Y_i^m(\boldsymbol{\theta})\right| \tag{4.3}$$

$$F_{1m}\left(\boldsymbol{\theta}|\boldsymbol{z}\right) = \max_{i=1,\dots,n_y}\gamma_i\left|Y_i(\boldsymbol{z}) - Y_i^m(\boldsymbol{\theta})\right| \tag{4.4}$$

Both these objectives lead, though, to more challenging optimisation problems as they correspond to nondifferentiable functions. Thus, preference will be here for the objective given by Equation 4.2. The alternative measures will be used to evaluate the suitability of different solutions.

The second objective F_2 measures the discrepancy of $\boldsymbol{\theta}$ from the established predictive relationships. One choice for $F_2(\boldsymbol{\theta}|\boldsymbol{z})$ could be $[\boldsymbol{\theta}-\boldsymbol{\mu_\theta}(\boldsymbol{z})]^T[\boldsymbol{\theta}-\boldsymbol{\mu_\theta}(\boldsymbol{z})]$, i.e., the discrepancy from the mean predictive relationships. If there is no available uncertainty characterization U, then this is the only option that can be made. If this characterisation is available, as is the case typically with record-based models (Rezaeian and Der Kiureghian, 2010), correlation and variability information for the model parameters can be additionally incorporated. This is established by considering the maximisation of the likelihood of θ based on the probability

model described through U. For the stochastic ground motion model described in Appendix A, this corresponds to the maximisation of the probability density for $\boldsymbol{v}(\boldsymbol{\theta})$, leading to

$$F_2\left(\boldsymbol{\theta}|\boldsymbol{z}\right) = \left[\boldsymbol{v}(\boldsymbol{\theta}) - \boldsymbol{\mu}(\boldsymbol{z})\right]^T \boldsymbol{\Sigma}^{-1} [\boldsymbol{v}(\boldsymbol{\theta}) - \boldsymbol{\mu}(\boldsymbol{z})] \tag{4.5}$$

The covariance matrix $\boldsymbol{\Sigma}$ incorporates in the formulation the correlation between the model characteristics as well as the fact that variability for each of these characteristics is different.

Objective function F_1 enforces the match to the target IMs. Objective F_2 guarantees compatibility of the physical characteristics of the resultant ground motions with the regional trends observed in recorded ground motions. Solution of the multi-objective optimisation of Equation 4.1 ultimately leads to a Pareto set of dominant solutions $\{\boldsymbol{\theta}_p; p = 1, \ldots, n_p\}$ that express a different compromise between the competing objectives F_1 and F_2. A solution is characterised as dominant (and belongs in the Pareto set) if there is no other solution that simultaneously improves both objectives F_1 and F_2. The representation of the Pareto set in the performance objective $[F_1, F_2]$ space, $\{[F_1(\boldsymbol{\theta}_p|\boldsymbol{z}), F_2(\boldsymbol{\theta}_p|\boldsymbol{z})]; p = 1, \ldots, n_p\}$ is termed as the Pareto front. Illustrations of such Pareto fronts are included in the example discussed later. One extreme point of this front will always correspond to the unmodified model with $\boldsymbol{\theta} = \boldsymbol{\mu_\theta}(\boldsymbol{z})$, representing the minimum of objective $F_2=0$. Unless this point also yields a match to the targeted hazard (i.e., corresponds to $F_1=0$), optimisation of Equation 4.1 will identify points that improve upon $F_1(\boldsymbol{\mu_\theta}(\boldsymbol{z})|\mathbf{z})$ while deviating from the unmodified model ($F_2 > 0$). One can eventually select a model configuration from the identified Pareto set that yields the desired IM-compatibility without deviating significantly from regional ground motion characteristics. This will be further discussed in the illustrative implementation in Section 4.5.

Identifying the Pareto set for this problem is challenging because the computational burden in evaluation of objective F_1 is significant, requiring $n_w \cdot n_y$ time-history analyses for each objective function estimation. To facilitate an efficient optimisation that can be repeated for any desired seismicity scenario $\boldsymbol{z}$, a surrogate modeling approach is adopted here, similar to the one used by Vetter et al. (2016). Specifically, kriging is used as metamodel since it has a proven capability to approximate highly complex functions (Lophaven et al., 2002), while simultaneously providing gradient information that will be directly exploited in the optimisation and allowing to explicitly consider the local metamodel approximation error within the optimisation formulation. These aspects of the optimisation problem will be discussed in Section 4.4. The details of the kriging metamodel development are discussed first in the next section.

4.3   Kriging metamodel development

The kriging metamodel is developed to provide an efficient approximation to the input-output relationship $\boldsymbol{\theta} - Y_i^m(\boldsymbol{\theta})$ considering every potential response quantity of interest that can be eventually used for the definition of objective F_1. A further simplification can be established if the relationship between some components of $\boldsymbol{\theta}$ and the response output $Y_i^m(\boldsymbol{\theta})$ is explicitly known. This is true for stochastic ground motion models that include a parameter, denoted θ_s herein, that directly controls the amplitude of the excitation. This means that $Y_i^m(\boldsymbol{\theta}) = \theta_s \cdot s_i^m(\boldsymbol{x})$ with $\boldsymbol{x}$ corresponding to the remaining model parameters excluding θ_s and $s_i^m(\boldsymbol{x})$ representing the output $Y_i^m(\boldsymbol{\theta})$ for $\theta_s = 1$. For the model described in Appendix A, $\theta_s = \sqrt{I_a}$ and $\boldsymbol{x} = \{D_{5-95}, t_{mid}, \omega_{mid}, \omega', \zeta_f\}$. Without loss of generality, we will adopt here this assumption, i.e., representation $Y_i^m(\boldsymbol{\theta}) = \theta_s \cdot s_i^m(\boldsymbol{x})$. In this case, the metamodel needs to be established to approximate only relationship $\boldsymbol{x} - s_i^m(\boldsymbol{x})$.

For developing the metamodel, a database with n observations is initially obtained that provides information for the $\boldsymbol{x} - s_i^m(\boldsymbol{x})$ pair. For this purpose n samples for

$\{\boldsymbol{x}^j; j = 1, \ldots, n\}$, also known as support points or experiments, are obtained over the domain of interest for $\boldsymbol{x}$. This domain, denoted X, should encompass the anticipated range that the metamodel will be implemented in and its determination is discussed later in this section. The median predictions provided through the ground motion model are then established through the 3-step process discussed in Section 4.2 considering n_w white-noise samples. Using the dataset $\{\boldsymbol{x}^j - s_i^m(\boldsymbol{x}^j); j = 1, \ldots, n\}$, the kriging model can be formulated. Details for this formulation may be found in the studies of Sacks et al. (1989) or Vetter et al. (2016), with the latter reference focusing on a similar application as the one considered here, looking at approximating the predictions of stochastic ground motion models.

This approach ultimately leads to a kriging predictor that has a Gaussian nature with predictive mean $\hat{s}_i^m(\boldsymbol{x})$ and local prediction variance, which is also a function of $\boldsymbol{x}$, $\sigma_i^2(\boldsymbol{x})$ (Sacks et al., 1989). Each response output can be approximated through this predictor leading to

$$Y_i^m(\boldsymbol{\theta}) = \theta_s \hat{s}_i^m(\boldsymbol{x}) + \epsilon_i \theta_s \sigma_i(\boldsymbol{x}) \tag{4.6}$$

where ϵ_i is a standard Gaussian variable. This facilitates a computationally efficient approximation to $Y_i^m(\boldsymbol{\theta})$ for each $\boldsymbol{\theta}$. This efficiency can be improved by ignoring the metamodel prediction error (i.e., setting $\epsilon_i{=}0$) since calculation of predictive variance $\sigma_i^2(\boldsymbol{x})$ entails a significant higher computational cost than estimation of predictive mean. The computationally intensive aspect of the entire formulation is the development of the database $\{\boldsymbol{x}^j - s_i^m(\boldsymbol{x}^j); j = 1, \ldots, n\}$, which requires response-history analysis for a large number of model parameters to populate X and a sufficient number of white-noise samples to address the resultant variability in the response. However, this needs to be performed only once. As soon as the kriging metamodel is established based on this database, it can be then used to efficiently predict the responses for any other $\boldsymbol{\theta}$ desired. Calculation of $\hat{s}_i^m(\boldsymbol{x})$ and $\sigma_i^2(\boldsymbol{x})$ can be also vectorised (Jia and Taflanidis, 2013), something

that will be leveraged in the numerical optimisation discussed in the next section.

The accuracy of this metamodel depends on the number of experiments n used as well as the exact selection of these experiments. A large value of n can improve accuracy but at the same time can reduce significantly computational efficiency. Since the latter is important for solving the challenging multi-objective optimisation problem discussed here, the value of n needs to be kept moderately low. Therefore, the improvement of metamodel accuracy is primarily sought after through an adaptive design of experiments (DoE). The adaptive DoE strategy gradually increases the number of support points, leveraging the metamodel developed through the existing support points to guide the selection of the new experiments. This leads to an iterative identification of support points, whereas the specific strategy adopted here for selecting the support points in each of the iterations corresponds to a sample-based DoE (Dubourg et al., 2011; Gidaris et al., 2015). In the first iteration, since no metamodel is yet available, the initial n_1 experiments are obtained using Latin hypercube sampling in X, i.e., a space-filling DoE. Subsequent iterations adopt the adaptive DoE strategy. At the kth iteration, a surrogate model is developed using the available $n_{av}^{(k)}$ support points. The prediction error of this metamodel is then leveraged to identify new experiments in regions with low metamodel accuracy. This is accomplished through a sample-based implementation: a large number of candidate experiments is first sampled within X that are distributed proportional to $\sigma_i^2(\boldsymbol{x})$ (e.g., through rejection sampling (Robert and Casella, 2004)) and these experiments are then clustered (e.g., using k-means clustering (Hartigan and Wong, 1979)) to the desired number n_a of additional experiments which are added to the existing experiments for a total of $n_{av}^{(k+1)} = n_{av}^{(k)} + n_a$ support points. The clustering is important for avoiding close-proximity support points that ultimately provide information over the same domain in X. A new metamodel is then developed using the $n_{av}^{(k+1)}$ experiments, and its accuracy is assessed, e.g., by calculating error statistics through cross-validation (Gidaris et al., 2015). If sufficient accuracy is achieved, then adaptive DoE is terminated and $n = n_{av}^{(k+1)}$, else algorithm proceeds to the $k+1$st iteration.

Another important feature of the metamodel development is the definition of domain X. This domain needs to (i) cover the entire range of values that the metamodel will be eventually used for (to avoid extrapolations that unavoidably have reduced accuracy) while (ii) avoiding unnecessarily broad definitions that lead to computational effort spent in subdomains within X of no practical interest. Here the definition of X is established by ignoring subdomains that correspond to high values for $F_2(\boldsymbol{\theta}|\boldsymbol{z})$ for any potential seismicity scenario $\boldsymbol{z}$, i.e., $\boldsymbol{x}$ values that are away from the current predictive relationships. Solutions within such subdomains will not be selected in the multi-objective optimisation because they correspond to large values for one of the objectives. The X definition is established through these steps:

Step 1: Create a range of seismicity scenarios $\{\boldsymbol{z}_c^l; l = 1, \ldots, n_c\}$ that are representative of the scenarios that could be eventually considered in the ground motion model tuning.

Step 2: Define a box-bounded domain X^d that is expected to be a superset of X, and create a large number of samples for $\{\boldsymbol{x}_d^j; j = 1, \ldots, n_s\}$ uniformly distributed in X^d. Range defined through the scenarios in Step 1, $\{\boldsymbol{\mu}_{\boldsymbol{\theta}}(\boldsymbol{z}_c^l); l = 1, \ldots, n_c\}$, can be also used to guide selection of X^d.

Step 3: For each candidate sample $\boldsymbol{x}_d^j$ calculate value $F_2(\boldsymbol{\theta}_d^j|\boldsymbol{z}_c^l)$ for each seismicity scenario $\boldsymbol{z}_c^l$ where $\boldsymbol{\theta}_d^j$ is the sample corresponding to $\boldsymbol{x}_d^j$ and the mean value for θ_s (for the given $\boldsymbol{z}_c^l$). Evaluate

$$D^j = \min_{l=1,\ldots,n_c} F_2\left(\boldsymbol{\theta}_d^j|\boldsymbol{z}_c^l\right) \tag{4.7}$$

which corresponds to the smallest distance for $\boldsymbol{x}_d^j$ from the current predictive relationships for any potential seismicity scenario.

Step 4: Set a threshold δ_d and then classify each sample $\boldsymbol{x}_d^j$ as belonging in set X if D^j is smaller than δ_d (sample is given membership classification 1) or not otherwise (sample is given membership classification 0). Using this

classification information, domain X can be characterised through support vector machine (SVM) (Schölkopf, Smola et al., 2002).
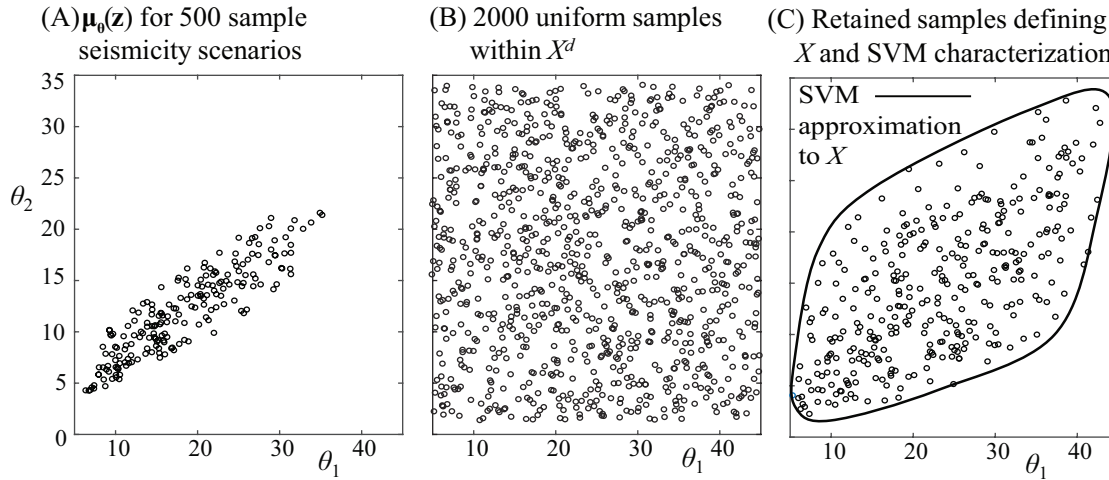
FIGURE 4.1: Steps for SVM classification of X: (A) samples $\boldsymbol{\mu_\theta}(\boldsymbol{z})$ for a range of seismicity scenarios; (B) uniform samples created within X^d; (C) samples as belonging to X along with developed SVM (curve).

This approach ultimately leads to an SVM characterization of X. Figure 4.1 demonstrates some of the steps of this process. Finally, samples within X, as needed for the adaptive DoE, can be generated with negligible computational effort by creating first samples within box-bounded domain X^d and then using the SVM classifier to maintain only the samples belonging in X.

4.4 Multi-objective optimisation to match target IMs supported by Kriging metamodeling

The multi-objective optimisation of Equation 4.1 can be efficiently performed by using the kriging approximation given by Equation 4.6 when evaluating performance objective F_1. Note that calculating objective F_2 is computationally trivial. Additionally, the approximation error of the metamodel can be incorporated in the objective function definition, leading to the following modification (Vetter et al., 2016):

$$F_1\left(\boldsymbol{\theta}|\boldsymbol{z}\right) = \frac{1}{n_y}\sum_{i=1}^{n_y}\gamma_i^2\left[\left(Y_i(\boldsymbol{z}) - \theta_s\hat{s}_i^m(\boldsymbol{x})\right)^2 + \theta_s^2\sigma_i^2(\boldsymbol{x})\right] \qquad (4.8)$$

The motivation for incorporating this error is to improve the robustness of the optimisation and avoid convergence to erroneous solutions due to poor quality of the metamodel. This feature will be further explored in the illustrative implementation in Section 4.5.

For solving the multi-objective optimisation a variety of numerical approaches can be used (Marler and Arora, 2004). Here two are considered, one gradient-free and one gradient-based, offering different advantages. The first approach adopts an exhaustive search (Coello et al., 2007). A very large number of n_{bc} samples for $\boldsymbol{\theta}$ are generated that are close to $\boldsymbol{\mu_\theta}(\boldsymbol{z})$, and objective functions F_1 and F_2 are calculated. Estimation of objective F_1 in this case leverages the computational efficiency of the metamodel in performing vectorised predictions: the calculations are simultaneously performed for all n_{bc} samples, or using subsets with a lower number of samples depending on the available computational resources (memory can be a problem for vectorising operation). This greatly reduces computational time for estimating F_1. The dominant solutions representing the Pareto front can be then readily identified by comparing the values for the two objectives. The challenge in this case is that the value of n_{bc} needs to be large in order to obtain an adequate representation of the Pareto front. The advantage is that vectorised calculations can be used for the metamodel predictions. For the stochastic ground motion model described in Appendix A, the samples for $\boldsymbol{\theta}$ can be generated by obtaining samples for $\boldsymbol{v}$ from Gaussian distribution $N(\boldsymbol{\mu}(\boldsymbol{z}), \boldsymbol{\Sigma})$ and then transforming these to samples for $\boldsymbol{\theta}$ through the inverse of Equation A.6 in Appendix A. This guarantees that samples will correspond to lower values for F_2 (i.e., are close to $\boldsymbol{\mu_\theta}(\boldsymbol{z})$) and can therefore emerge as dominant solutions. Note that since this approach does not use gradients, it can seamlessly accommodate the alternative objective functions for F_1 given by Equations 4.3 and 4.4. The gradient-free optimisation was implemented in MATLAB without any toolboxes used.

The second optimisation approach is a gradient-based one. The epsilon constraint approach (Mavrotas, 2009) is specifically adopted due to its ability to explicitly define the value for one of the objectives. This method converts the multi-objective optimisation problem to a set of single-objective constraint optimisation problems with different constraint bounds ϵ^r. Through systematic variations of this constraint, different Pareto optimums can be obtained. Here objective function F_2 is adopted as objective and F_1 as constraint. This allows identification of the stochastic ground motion model that provides a specific compatibility with the target hazard (the prescribed constraint). The range of values ϵ^r of interest is determined by identifying first the anchor points of the Pareto front, corresponding to the minimum of objective functions $F_1(\boldsymbol{\theta}|\boldsymbol{z})$ and $F_2(\boldsymbol{\theta}|\boldsymbol{z})$ (unconstrained single-objective optimisations). Evidently optimisation for $F_2(\boldsymbol{\theta}|\boldsymbol{z})$ yields solution $\boldsymbol{\mu_\theta}(\boldsymbol{z})$. The range for ϵ^r corresponds then to $[\min F_1 \, F_1(\boldsymbol{\mu_\theta}(\boldsymbol{z}))]$ and a number of different constraint values can be considered, with the exact number depending on the desired resolution of the front. For each such value the single-objective, constrained optimisation is solved

$$\boldsymbol{\theta} = \arg\min F_2(\boldsymbol{\theta}|\boldsymbol{z})$$
$$\text{subject to } F_1(\boldsymbol{\theta}|\boldsymbol{z}) \leq \epsilon^r \tag{4.9}$$

This optimisation problem is not convex, and a gradient-based approach appropriate for constrained global optimisation problems needs to be adopted. This is accomplished through an approximate multistart approach that addressed both the potential existence of multiple local minima as well as challenges associated with identifying feasible starting points for the gradient-based approach. Initially, a large value n_{init} of trial solutions for $\boldsymbol{\theta}$ is examined, then the solutions corresponding to lower values of $F_2(\boldsymbol{\theta}|\boldsymbol{z})$ while satisfying constraint ϵ^r for $F_1(\boldsymbol{\theta}|\boldsymbol{z})$ are taken as initial points for a gradient-based optimisation. The latter is achieved through a SOL solver implemented through the TOMLAB optimisation environment (Holmstrom et al., 2009). The same candidate solutions are used for all

values of ϵ^r (no need to repeat this step). Evaluation of $F_1(\boldsymbol{\theta}|\boldsymbol{z})$ over the n_{init} candidate solutions is vectorised, so has small burden, whereas the efficiency of the gradient-based optimisation is improved by obtaining analytically the gradients for both objectives. For objective F_1, the components of the gradient vector are

$$\frac{\partial F_1}{\partial \theta_i} = \frac{1}{n_y} \sum_{i=1}^{n_y} \gamma_i^2 \left[2\left(Y_i(\boldsymbol{z}) - \theta_s \hat{s}_i^m(\boldsymbol{x})\right) \frac{\partial \theta_s \hat{s}_i^m(\boldsymbol{x})}{\partial \theta_i} + \frac{\partial \theta_s^2 \sigma_i^2(\boldsymbol{x})}{\partial \theta_i} \right] \qquad (4.10)$$

where the partial derivatives inside the brackets can be readily obtained through the metamodel (Vetter et al., 2016). For the specific objective function F_2 that will be used in the illustrative example later, given by Equation 4.5, the components of the gradient vector are calculated by the chain rule as

$$\frac{\partial F_2}{\partial \theta_i} = \frac{\partial F_2}{\partial v_i} \frac{\partial v_i}{\partial \theta_i} = \frac{\partial F_2}{\partial v_i} \frac{F'_{\theta_i}(\theta_i)}{\phi(\Phi^{-1}(F_{\theta_i}(\theta_i)))} \qquad (4.11)$$

where $\phi(.)$ stands for the standard Gaussian probability density function (PDF), the partial derivative $\partial v_i / \partial \theta_i$ was calculated by differentiating Equation A.6 in Appendix A, and the partial derivatives $\partial F_2 / \partial v_i$ correspond to the components of the gradient row vector

$$\nabla_v F_2 = 2\left[\boldsymbol{v}(\boldsymbol{\theta}) - \boldsymbol{\mu}(\boldsymbol{z})\right]^T \Sigma^{-1} \qquad (4.12)$$

The challenge for this optimisation approach is that a gradient-based step needs to be repeated multiple times (for each different value of ϵ^r) and cannot leverage vectorised calculations for the metamodel because the metamodel is separately used for each sequential objective function evaluation. The advantage is that gradient information can improve computational efficiency and that the optimisation can be performed for specific values of ϵ^r. This allows the identification of a specific part for the Pareto-front if desired, e.g., the front that corresponds to specific levels of compatibility to the chosen IMs.

Once the Pareto front has been identified, a dominant solution can be adopted using any desired criterion, e.g., the solution that provides a specific compromise between the two objective functions. This will be further discussed in the illustrative implementation of the next section.

4.5   Illustrative implementation

The illustrative implementation considers the stochastic ground motion model developed by Rezaeian and Der Kiureghian (2010) and reviewed in Appendix A. For the target IMs, GMPEs used in the Western US are considered here (Abrahamson and Silva, 2008; Boore and Atkinson, 2008; Campbell and Bozorgnia, 2008; Chiou and Youngs, 2008), where the suggestions by Kaklamanos et al. (2011) were adopted to estimate unknown inputs for some of the GMPEs. As target, IM predictions from individual GMPEs as well as the average of their predictions will be adopted later. Note that the latter still provides a single target IM for each structural period examined. All computations are performed in a quad-core 3.0 GHz Xeon processor with 16 Gb of RAM and all computational times reported herein are for this processor.

4.5.1   Details for metamodel development

The box-bounded domain X^d was determined based on the ranges reported by Rezaeian and Der Kiureghian (2010) as [5 45] s for D_{5-95}, [0.5 40]s for t_{mid}, [0.1 30] Hz for $\omega_{mid}/2\pi$, [-2, 0.5] Hz/s for $\omega'/2\pi$, and [0.02 0.99] for ζ_f. For the response output, the peak pseudo-acceleration (Y=PSA) for an SDoF system with 5% damping ratio and for 22 different periods, the ones used by the aforementioned GMPEs, is adopted. The white-noise samples are chosen as n_w=200. For the domain X characterisation n_c=1000 seismicity scenarios are considered in range [6 8] for M, [10 100] km for R, [300 1600] m/s for $V_{s,30}$ and discrete 0,1 for F. These

ranges correspond to the ones for which the initial predictive relationships (Reza-eian and Der Kiureghian, 2010) were developed. The samples for the SVM-based characterisation of X are chosen as $n_s=10000$ with $\delta_d=9$ (latter corresponding to radius of 3 standard deviations away from mean). The adaptive DoE discussed in Section 4.3 is implemented with $n_1=600$ and $n_a=300$. Three different accuracy criteria are selected, with associated coefficient of determination (averaged over all outputs) 0.95, 0.97, and 0.985. This leads to number of support points 1500, 3000, and 4500, respectively. The SVM-characterisation and adaptive DoE was implemented in MATLAB, whereas for the tuning of the metamodel parameters the MATLAB toolbox DACE (Lophaven et al., 2002) was utilised.

For generating a total of 900,000 time-histories and performing the required 19,800,000 simulations to develop the database for the metamodel for the elastic responses, close to 600 CPU hours were required. For these computations, a high performance cluster (Persephone) was used. Although this computational burden is significant, it should be stressed that it corresponds to an initial only overhead of the approach. Once the metamodel is developed, it can be then used for any required predictions because the established accuracy is high. This large burden should be also attributed to the large number of white-noise sequences (200), periods (22), and the wide seismicity range examined. The former provides high accuracy for the calculation of the relevant statistics, whereas the latter two support a wide applicability of the developed metamodel, as it can provide accurate predictions for all the responses of interest for the considered stochastic ground motion model (covers its range of applicability) and GMPEs (covers all the periods addressed). The burden for a metamodel that considered smaller number of stochastic sequences or constrained seismicity or period ranges would be drastically decreased.

Estimation of metamodel response for 10000 samples requires 4.9s, 8.8s, and 14.05s for the metamodels with 1500, 3000, and 4500 support points, respectively. When the metamodel prediction error variance is not computed, the corresponding times are 3.6s, 7.5s, and 10.7s, respectively. Note that adoption of larger value of samples prohibits efficient vectorisation of operations for the $n=4500$ points due to

memory restrictions. The calculation of objective function F_1 along with its gradient requires 0.031s, 0.052s, and 0.091s for the metamodels with 1500, 3000, and 4500 support points, respectively. When the metamodel prediction error variance is not included in the calculation, the corresponding times are 0.008s, 0.0107s, and 0.015s, respectively. Comparison of these computational times shows that: (i) increase of the support points has a considerable effect on computational efficiency, (ii) vectorisation of calculations provides significant benefits, and (iii) inclusion of the prediction error variance in the calculations increases the computational burden, especially when gradient information needs to be obtained. All these aspects should be taken into account when choosing computational details for the optimisation problem.

4.5.2   Comparison of optimisation approaches

The focus is first placed on the numerical solution of the optimisation problem. The target used in this subsection, and also in the next one 4.5.3, corresponds to structural periods T_s=[0.4 0.5 0.75 1.0 1.5 2.0] s and IM described by the average of the considered GMPEs (Abrahamson and Silva, 2008; Boore and Atkinson, 2008; Campbell and Bozorgnia, 2008; Chiou and Youngs, 2008). Weights are chosen as γ_i=$1/Y_i(\boldsymbol{z})$, so that objective function is expressed in terms of the relative error. In other words, no specific structural period is prioritized in evaluating the match to the target IMs. Three reference seismicity scenarios are examined in this section, corresponding to M=6-R=20km, M=7.8-R=30km, and M=7-R=40km, for a strike-slip fault (F=0) and $V_{s,30}$=800 m/s. The first two scenarios correspond to cases where the unmodified stochastic ground motion model does not provide an adequate match to the target GMPEs (Rezaeian and Der Kiureghian, 2010) (significantly over predicts for the former, moderately under predicts for the latter) and the last to a case where the unmodified model facilitates a good match. The discussion will focus around specific cases of interest. The metamodel with 4500 support points is used in this section. Comparison across metamodels with

different number of support points (and therefore accuracy) will be discussed in the next section.
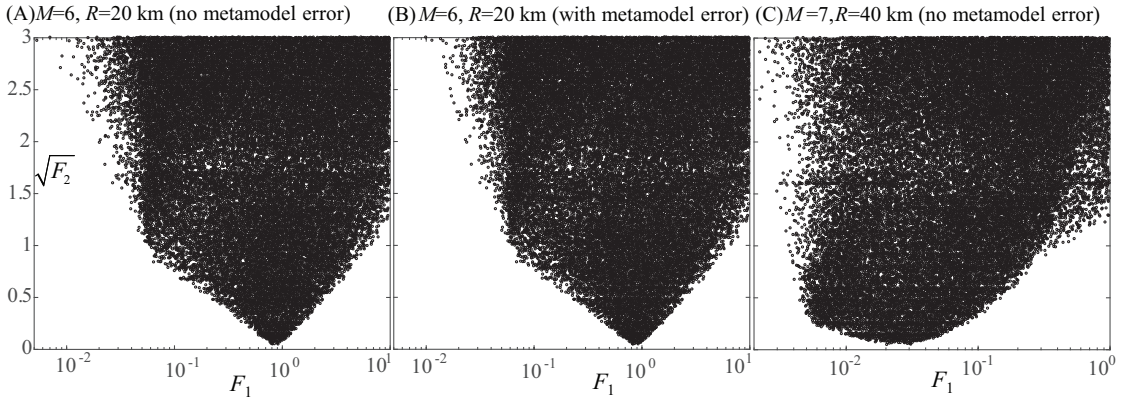
FIGURE 4.2: Projection to the objective space of swarm of 50000 candidate solutions obtained using an exhaustive search for (A,B) $M=6$-$R=20$km and (C) $M=7$-$R=40$km. Objective F_1 is calculated without the metamodel error for (A) and (C), and with the metamodel error for (B).

Figure 4.2 presents illustrative swarms of candidate solutions in the objective space from the exhaustive search using $n_{bc}=50000$ samples for seismicity $M=6$-$R=20$km and $M=7$-$R=40$km, in the former case examining the case with and without metamodel error in the calculation of objective F_1. For the second objective, results are reported with respect to $\sqrt{F_2}$, which corresponds to the distance between the initial and the modified predictive relationships (not the squared distance) and offers a better normalisation for the results in the comparison. The solutions located at the left boundary of the swarms in Figure 4.2 correspond ultimately to Pareto optimal solutions because there is no other solution that can simultaneously improve both performance objectives. It is evident from these swarms that modification of the predictive relationships can indeed facilitate a difference in the IM match (check the range of F_1 values obtained), whereas for smaller F_1 values, the candidate solutions deviate more from the rest of the swarm. This means, ultimately, that there are fewer model configurations that can provide a good match to the target IMs (small F_1 values). When the unmodified model is closer to the target IMs ($M=7$-$R=40$km seismicity scenario), higher compatibility to these IMs can be obtained through modification of the predictive relationships (compare case (C)

to the other two), whereas addition of the metamodel error (compare case (B) to (A)) increases function F_1, especially for smaller F_1 values.
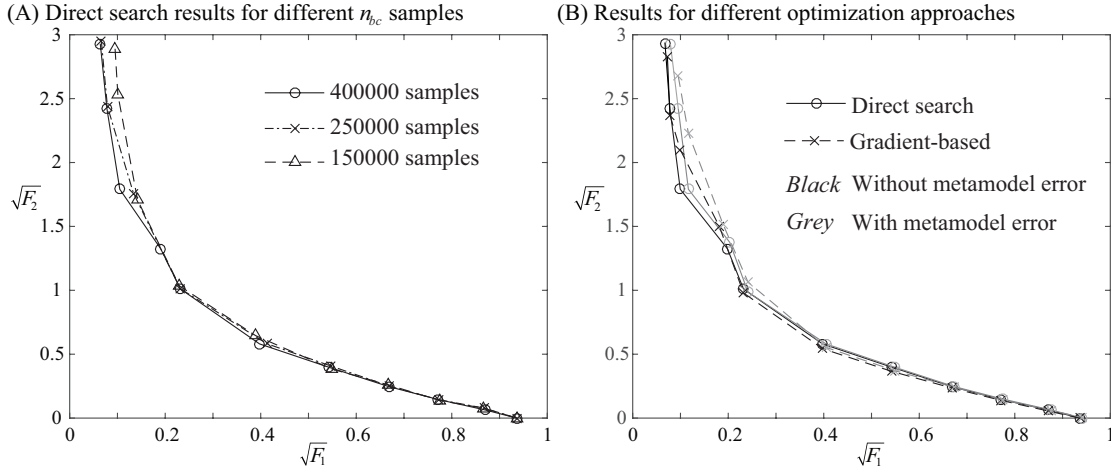
FIGURE 4.3: (A) Pareto fronts identified by exhaustive search for different n_{bc} values. (B) Comparison of Pareto fronts obtained by exhaustive search and epsilon constraint approach, considering or not the metamodel error. Case presented corresponds to seismicity scenario M=6-R=20km.

Figure 4.3 examines different approaches for the solution of the multi-objective problem for the reference seismicity case of M=6-R=20km (similar trends hold for the other cases). Part (A) shows the Pareto fronts identified through the exhaustive search using three different n_{bc} values. Results are reported herein with respect to $\sqrt{F_1}$ and $\sqrt{F_2}$ since this facilitates an easier comparison (differences of extreme values easier to discern). Only ten representative solutions are shown, and not the entire front. It is evident that minor differences only exist between the identified fronts for different n_{bc} values, and such differences occur primarily for small F_1 - large F_2 combinations. This comparison shows that a value of n_{bc} around 200,000 to 400,000 should be considered as sufficient for efficiently identifying the front. For the remaining of the manuscript results for the exhaustive search will be presented for a value of n_{bc} equal to 400,000. Part (B) of Figure 4.3 then compares the Pareto fronts obtained by the exhaustive search and the gradient-based (epsilon constraint) approaches for the case that the metamodel error is considered or not in the objective function F_1. For the epsilon constraint, n_{init} is taken as 10000, whereas the gradient-based optimisation is performed using a couple only different initial points. For small F_1 constraints, some challenges

were encountered in converging to an admissible solution. This should be attributed to the trend identified in Figure 4.2; for such values, a smaller number of candidate solutions exist that can satisfy the constraint, and a potential increase in n_{init} might be needed to identify feasible initial points for the gradient-based optimisation. The comparison in Figure 4.3(B) shows that the two approaches identify similar Pareto fronts, with the gradient-based optimisation converging to suboptimal solutions for lower F_1 values, evidently due to the existence of local minima with greater differences in achieved performance. For such performance ranges (i.e, with an already good match to the target IM, as indicated by the lower F_1 value), these differences might be unimportant. With respect to the computational burden, the exhaustive search requires 15 s per 10000 candidate solutions examined (11 s if metamodel error is disregarded). For the epsilon constraint the computational cost is 15 s for the initial 10000 trials and 20 s (2 s if metamodel error is disregarded) for each different constraint examined. These comparisons show that both optimisation approaches may be considered as adequate and preferred, depending on the application context. Overall, some preference exists for the exhaustive search due to the fact that epsilon constraint method needs identification of an appropriate starting point to avoid convergence to suboptimal local minima. The epsilon constraint approach, though, might be beneficial when a single solution is sought after, the one that satisfies a desired match to the target IM, rather than the entire front. Consideration or not of the metamodel error has no effect on the differences between the optimisation approaches. The computational efficiency for the gradient-based search is reduced, as discussed in the previous section, when this error is included. Table 4.1 provides summary of these results, including some details discussed in the next section.

Finally, Figure 4.4 presents a comparison between the alternative objective function selections for quantifying the discrepancy from the target IM, i.e., comparison between F_1, F_{1r}, and F_{1m}. Part (A) discusses F_{1r} and part (B) F_{1m}. The exact Pareto front, i.e., optimisation using F_{1r} (or F_{1m}) and F_2 as objectives, as well as an approximate front are compared. The approximate front is obtained by using

TABLE 4.1: Summary of pros/cons characteristics for the different implementation cases

	Pros	Cons
Optimisation algorithms		
Gradient-free exhaustive search	Can leverage vectorised metamodel predictions for increased computational efficiency	Large number of evaluations required and will identify always the entire front. Number of samples used in exhaustive search has (small) effect on solutions
Gradient-based	Very efficient when a single only solution is sought, rather than entire front. Allows identification of a specific part of the Pareto front	Requires identification of an appropriate starting point to avoid convergence to local minima. Greater computational burden when metamodel error is considered
Metamodel characteristics		
Larger number of support points	Higher accuracy in predictions established. Reduces necessity to include metamodel error for obtaining high quality solutions	Increased computational burden. Remedied by the fact that metamodel error does not need to be considered
Inclusion of metamodel error	Facilitates greater robustness, avoiding identification of erroneous points	Considerably increased computational burden especially when combined with gradient-based optimisation algorithms

F_1 and F_2 as objectives, identifying the Pareto set for $\boldsymbol{\theta}$ and then evaluating objective F_{1r} (or F_{1m}) over that Pareto set. As discussed earlier, this optimisation is less challenging but evidently identifies a sub-optimal solution. The results in Figure 4.4 show that there is overall very good correlation between the different objective functions and that the approximate solutions have only small deviations from the optimal front and those only for small values of the first objective. Therefore, use of F_1 as objective may be considered as an adequate surrogate even when the interest is in objectives F_{1r} or F_{1m} for quantifying discrepancy from the target IMs.
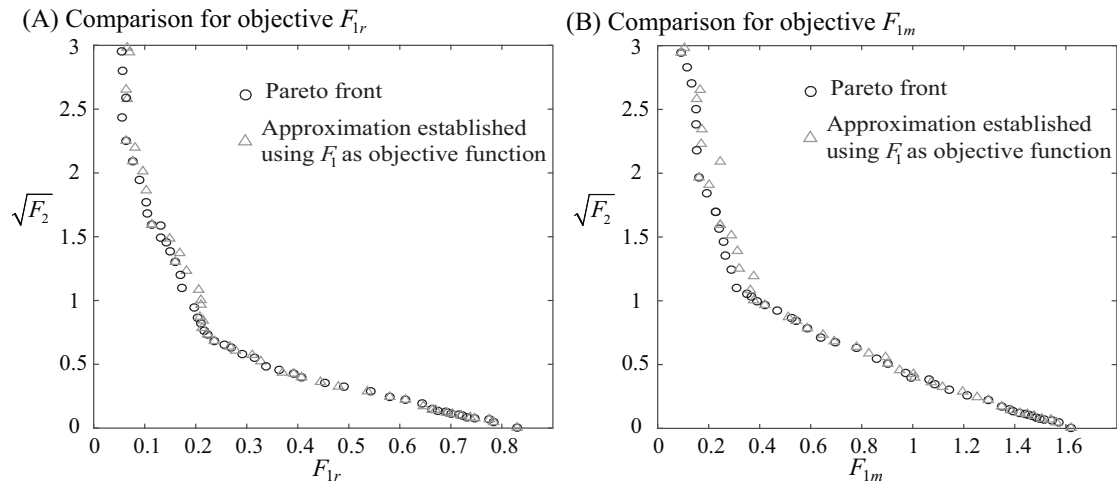
FIGURE 4.4: (A) Comparison between exact and approximate Pareto fronts for objective functions (A) F_{1r} and (B) F_{1m} describing the discrepancy from the target IMs. Case presented corresponds to seismicity scenario M=6-R=20km.

4.5.3   Impact of metamodel accuracy

The discussion moves next to the examination of the impact of the metamodel accuracy. This is established by considering additionally the results obtained by using the exact stochastic ground motion model (i.e., not relying on metamodel predictions), which represents the measure for evaluating the actual hazard compatibility of the identified ground motion model. The details for the study are the ones used in the previous section. Results for seismicity scenario M=6-R=20km are presented in Figures 4.5 and 4.6. Figure 4.5 shows the Pareto fronts identified by using the metamodels with the three different number of support points. Cases with or without the metamodel error in the estimation of objective function F_1 are separately shown. This leads to two different Pareto sets, one without error and one with error, and for each two different fronts are reported, one corresponding to metamodel predictions and one to the use of the exact stochastic ground motion model. Then Figure 4.6 shows the spectral plot comparisons for the solution (among the Pareto set identified in each case) corresponding to the minimum of F_1. The period range used in Figure 4.6 corresponds to the target structural periods T_s. In all cases, the exhaustive search is implemented with the same candidate solutions to facilitate a consistency in the corresponding comparisons. Figure 4.7

then shows Pareto front results for different seismicity scenario, $M=7.8$-$R=30$km. Note that for seismicity scenario $M=7$-$R=40$km (another case discussed in the previous section), results are of limited interest since the unmodified ground motion model provides a good compatibility to target IMs.

(A) Metamodel with 1500 support points   (B) Metamodel with 3000 support points (C) Metamodel with 4500 support points
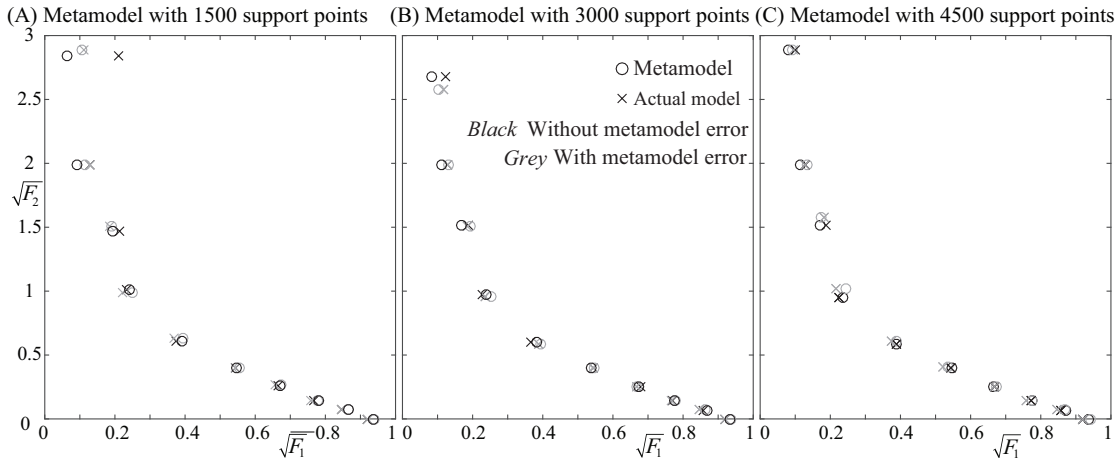
FIGURE 4.5: Pareto fronts identified using metamodels with (A) 1500, (B) 3000, or (C) 4500 support points and comparison to predictions by exact stochastic ground motion model. Case presented corresponds to seismicity scenario $M=6$-$R=20$km.

(A) Metamodel with 1500 support points (B) Metamodel with 3000 support points (C) Metamodel with 4500 support points
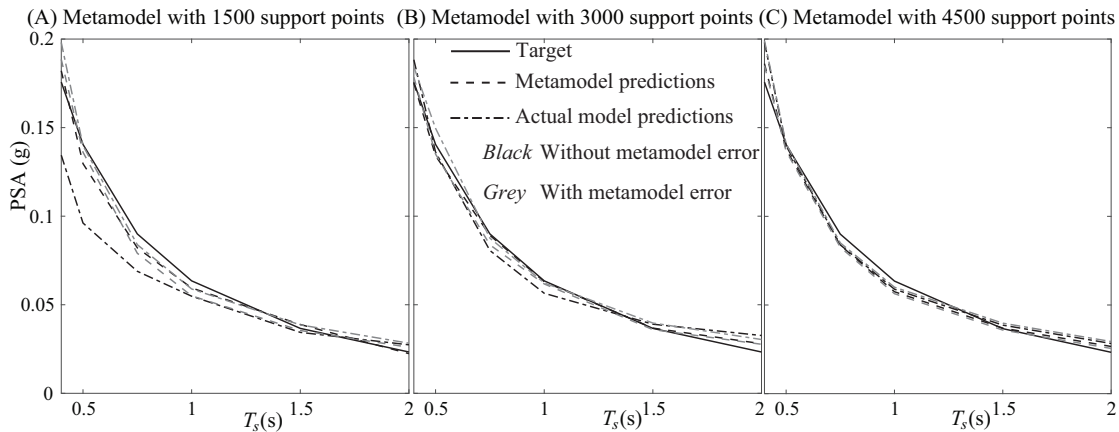
FIGURE 4.6: Spectral plots for the solutions corresponding to minimum of F_1 in the Pareto fronts identified in Figure 4.5.

The results show that for the higher accuracy metamodel (4500 support points) good agreement is established between the metamodel predictions and the actual model predictions along the Pareto front, whereas the inclusion of the metamodel error has only a small effect on the identified Pareto front. This is observed in both Pareto fronts (Figures 4.5 and 4.7) as well as in the corresponding spectral
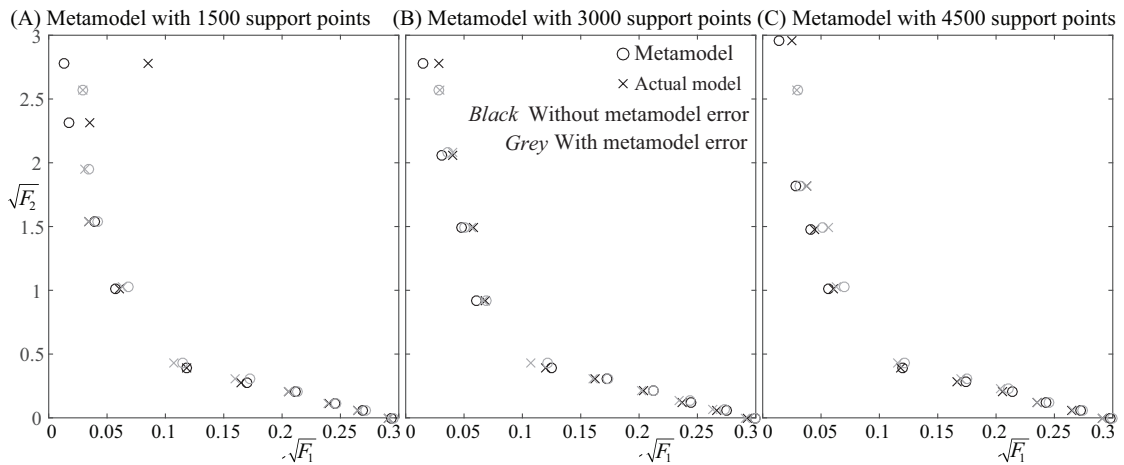
FIGURE 4.7: Pareto fronts identified using metamodels with (A) 1500 (B) 3000 or (C) 4500 support points and comparison to predictions by exact stochastic ground motion model. Case presented corresponds to seismicity scenario M=7.8-R=30km.

plots (Figure 4.6). Different trends are observed, though, for the lower accuracy metamodel (1500 support points). For lower F_1 values, and therefore large F_2 values, the Pareto optimal solutions identified when metamodel error is not included in the problem formulation lead to erroneously modified ground motion models. Based on the metamodel predictions, these models provide a very good match to the target IMs (small predicted F_1 value), but when response is evaluated with the exact model larger differences are observed from the target IMs. This is particularly evident in the spectral plots shown in Figure 4.6. Evidently the respective solutions identified correspond to parameters $\boldsymbol{\theta}$ for which the metamodel accuracy is low. The moderate accuracy metamodel (3000 support points) falls in between the two aforementioned cases, with characteristics that resemble more closely the ones for the high accuracy metamodel.

Note that the seismicity case examined here is ideal for exploring vulnerabilities in the optimisation associated with lower metamodel accuracy, as it corresponds to a case at the boundary of the scenarios used to define domain X. Therefore, for larger values of F_2, the corresponding parameters $\boldsymbol{\theta}$ are expected to be close to the boundary of X, where metamodel accuracy is lower. However, even for this challenging case, metamodels with higher accuracy (3000 or 4500 support points) face small challenges, whereas the inclusion of the metamodel error for the

lower accuracy metamodel (1500 support points) greatly improves the robustness of the optimisation, leading to identified solutions with good agreement between metamodel and actual model. Figure 4.8 sheds further light into this topic. It focuses on the case examined in part (A) of Figure 4.6 but besides the mean metamodel predictions it includes the predictions that are 1.5 standard deviations (σ) from the mean based on the estimated error variance. The solution identified when metamodel error is not included in the evaluation of F_1 is associated with a larger anticipated error (part (A) of Figure 4.8). Note that the actual model is actually even further away than the plotted 1.5σ. When the metamodel error is included in the evaluation of F_1, such $\boldsymbol{\theta}$ values with large associated σ are avoided, since the large σ contributes to larger values for the objective function. This ultimately contributes to identification of solutions with greater robustness, i.e., better agreement between metamodel and actual model (part (B) of Figure 4.8).
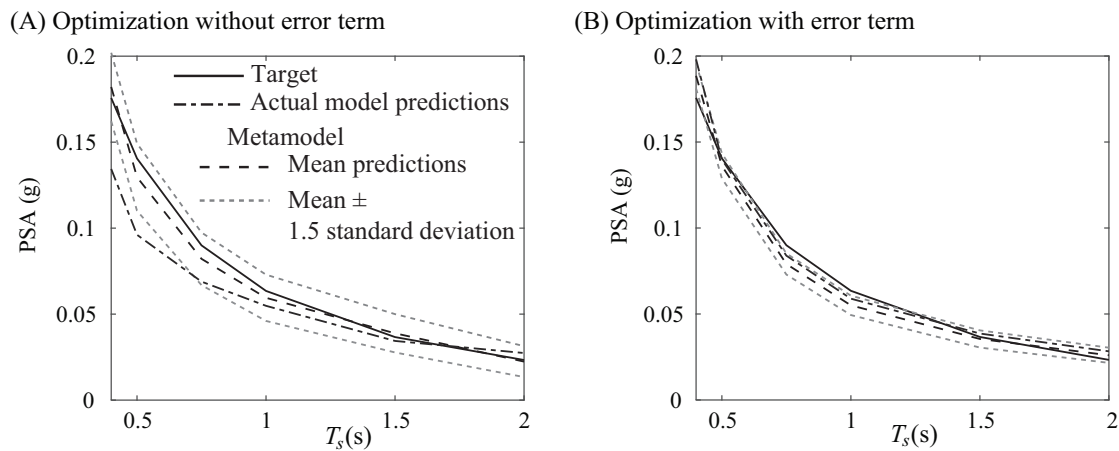
FIGURE 4.8: Spectral plots for the solutions corresponding to minimum of F_1 in the Pareto fronts identified in Figure 4.5 for the metamodel with 1500 support points. For the metamodel predictions the mean predictions and the predictions within 1.5 standard deviations from the mean are shown.

Overall, the above discussion shows that metamodels with higher accuracy (coefficient of determination 98%) can be considered as a good surrogate for the proposed optimisation, whereas the inclusion of the prediction error greatly improves the robustness of this optimisation, avoiding identification of erroneous solutions, even

when metamodels with lower accuracy are adopted. Consideration of this error is not necessary, though, when higher accuracy metamodels are used.

4.5.4   Implementation for different seismicity scenarios

With the computational details ironed out, the discussion moves finally to the IM compatibility established by the proposed modification of the ground motion model. Figure 4.9 shows results for three seismicity scenarios targeting PSA given by the average of the aforementioned GMPEs for two different ranges for T_s: T_s=[0.4 0.5 0.75 1.0 1.5 2.0] s and T_s=[0.4 0.5 0.75] s. These two different cases are referenced herein as long and short, respectively, period ranges. The proposed approach identified in each case a Pareto front that clearly demonstrates the compromise between the two objectives, with different characteristics in each case, depending on how close the unmodified ground motion model was to the target IM. Choosing a shorter period range for this target facilitates an overall better match; this is anticipated because objective F_1 imposes less strict requirements in terms of IM compatibility (fewer number of components to match).

The question finally arises which point should be selected within the identified Pareto set. Various approaches have been proposed in the greater multi-objective optimisation literature for making this choice (Keeney and Raiffa, 1993). Perhaps the most common one is to select the solution that has the smallest normalised distance from the utopia point, defined as the point in the Pareto front (i.e., objective function space) that corresponds to the minimum of the two objectives across the front. This utopia point represents the best but unachievable performance [$\min F_1$ $\min F_2$]. Normalisation is typically established with respect to the maximum of each performance objective across the Pareto front, leading to distance metric

$$D_p(\boldsymbol{\theta}) = \sqrt{\sum_{i=1,2} \left(\frac{F_i(\boldsymbol{\theta}|\boldsymbol{z}) - \min F_i}{\max F_i - \min F_i} \right)^2};$$

$$\max F_i = \max_{\{\boldsymbol{\theta}_p; i=1,...,n_p\}} (F_i), \ \min F_i = \min_{\{\boldsymbol{\theta}_p; i=1,...,n_p\}} (F_i)$$

(4.13)

Instead of F_1 and F_2, this can be implemented for $\sqrt{F_1}$ and $\sqrt{F_2}$ due to the better normalisation properties. This point is identified in all cases in Figure 4.9. Another choice would have been to choose the solution that satisfies a predetermined threshold for the match to the targeted IMs. Perhaps this is better set with respect to objectives F_{1r} or F_{1m} rather than F_1.
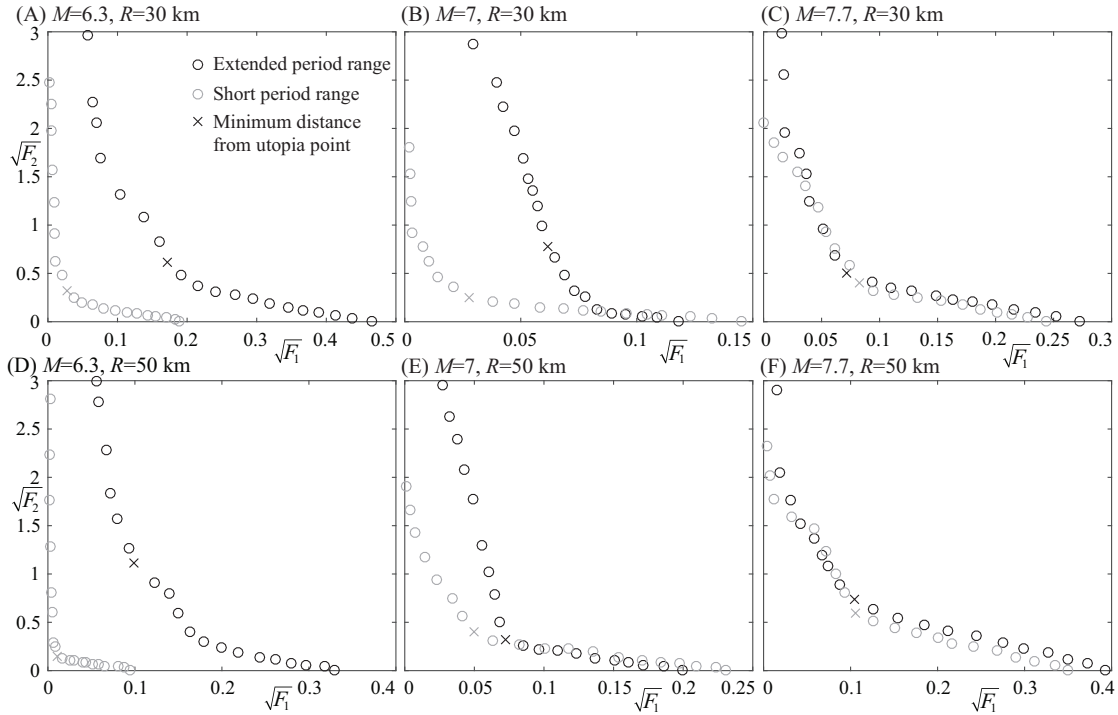
FIGURE 4.9: Pareto fronts for different seismicity cases considering match to long (black) or short (gray) period range IMs. In each plot the Pareto point with minimum distance from utopia point is shown with x.

These selections are finally demonstrated for a wide range of seismicity scenarios (M in range [6 8] and R in range [10 100] km) in Figures 4.10 to 4.12. For each scenario, three different Pareto points are selected, the one with smallest distance $D_p(\boldsymbol{\theta})$ from the utopia point and the ones with average absolute relative error $\sqrt{F_1}$ smaller than 0.15 or 0.05. These three cases are denoted, U_t, C_l, and C_s,

respectively. The thresholds for C_l and C_s modifications were chosen so that to reflect medium and small, respectively, incompatibility to the target hazard. In addition, results for the unmodified model are presented denoted U_n. Figure 4.10 shows plots for (i - first row) $\sqrt{F_1}$ for U_t and U_n ($\sqrt{F_1}$ is constrained for the other two cases) and for (ii – second and third rows) $\sqrt{F_2}$ for U_t, C_l, and C_s ($\sqrt{F_2}$ is zero for U_n). To better demonstrate the differences, results are presented separately for U_t (second row) and for the pair C_l and C_s (third row) in the latter case. The three different columns in the figure correspond to three different implementation cases: target IM given by the average of the aforementioned four GMPEs for both the (A) long and (B) short period ranges as well as (C) target IM given only by GMPE Boore and Atkinson (2008) for the long period range. These scenarios are denoted herein as SC_1, SC_2, and SC_3, respectively. Figure 4.11 shows spectral plots for a smaller selection of seismicity scenarios, defined by combinations of M [6.2, 6.8, 7.4, 8] and R [30, 60, 90] km, for SC_1. For each of the 12 M-R combinations the curves corresponding to the target IM, the unmodified model and the predictions by the three aforementioned model modifications are shown to facilitate comparisons. Finally, Figure 4.12 shows for all examined seismicity scenarios the model parameters $\boldsymbol{\theta}$ for the unmodified ground motion model as well as for the modified model corresponding to the Pareto point with smallest distance from the utopia point (U_t case) for SC_1. Note that some of the curves shown in these figures have nonsmooth characteristics. This should be attributed to multiple facts: (i) a discrete representation of the Pareto front is identified, rather than the actual Pareto front; (ii) problem has multiple local minima as discussed earlier, especially for larger F_2 values; and (iii) algorithms used for the optimisation have characteristics of stochastic search, well known to lead to discontinuous results.

The unmodified model U_n does not provide a good match to the target IMs for the entire seismicity range, with trends observed in first row of Figure 4.10 and in the spectral plots in Figure 4.11 being similar to the ones reported by Rezaeian and Der Kiureghian (2010), i.e., greater challenges in lower moment magnitudes and combination of higher magnitudes and larger distances. Figure 4.10 shows that
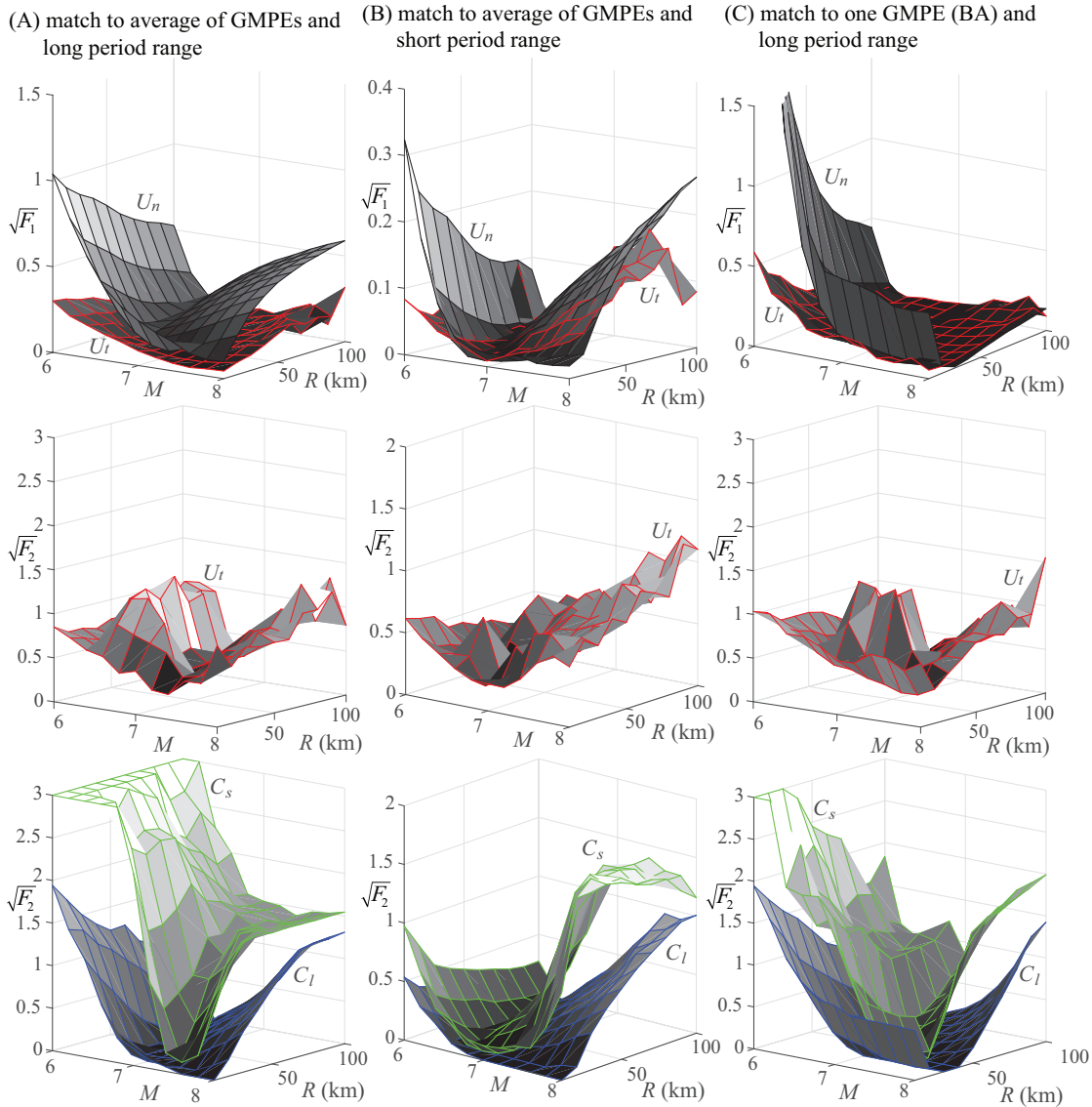
FIGURE 4.10: Results for $\sqrt{F_1}$ and $\sqrt{F_2}$ for unmodified ground motion model (U_n) and modified ground motion model corresponding to three different selection criteria: Minimum distance form utopia point (U_t) and value $\sqrt{F_1}$ smaller than 0.15 (C_l) or 0.05 (C_s). Implementations in the different columns correspond to (A) long and (B) short period ranges for matching to the average considered GMPEs and (C) long period ranges for match to GMPE Boore and Atkinson (2008).

the proposed modification (cases U_t, C_l and C_s) improves this match, establishing a balance between F_1 and F_2, with the characteristics of the balance depending on the criteria for selection of the final model among the Pareto optimal solutions (i.e., which specific case is chosen). When the unmodified model has larger discrepancies from the target IMs, then the modifications lead to larger values for F_2, but still

successfully identify models, independent of the implementation case, that provide an improved match to the IM target. The U_t modification identifies a model with moderate discrepancy from the unmodified one, corresponding to values of F_2 in the range of 0.3-1, whereas the two other modification approaches, C_l and C_s, identify models with greater variability across the different seismicity scenarios. For scenarios in the range of M=7-7.5 the unmodified model provides a good match to the target IMs and therefore modification of it provides limited advantages. This is perhaps better captured by the C_l case. This discussion shows that selection of the Pareto optimal model based on a targeted accuracy to the GMPEs, i.e, value for F_1 below a certain threshold as in the C_s and C_l cases, provides a more rational selection for the final model as it allows a more direct recognition of the seismicity ranges where modification is not truly required. However, when that threshold is selected small (C_s case) and the unmodified model has larger discrepancy from the target IM, then the modification leads to identification of a model with big differences from the original one (large F_2 values). This model will typically be far away from the U_t case and will belong in a steep part of the Pareto front, meaning that small improvements in F_1 come at a large increase of F_2 (check the Pareto fronts shown in previous figures). A multilevel selection criterion seems therefore more appropriate: select the Pareto optimal solution that satisfies a certain accuracy threshold for F_1 unless this solution leads to a greater F_2 value than the Pareto optimal solution with minimum distance from the utopia point. If the latter happens, then select the Pareto optimal solution with minimum distance from the utopia point.

With respect to the different implementation cases shown in Figure 4.10, selection of a shorter period range [compare cases (B) and (A)] promotes, as also identified earlier, an easier match to the target (smaller overall values for F_1 and F_2). This demonstrates the importance of carefully selecting the target IMs, as this selection affects the ability to match this target, and of the availability of a framework, as the one developed here, that allows you to do so. The exact selection of the target [compare cases (A) and (C)] does not impose any additional constraint in the
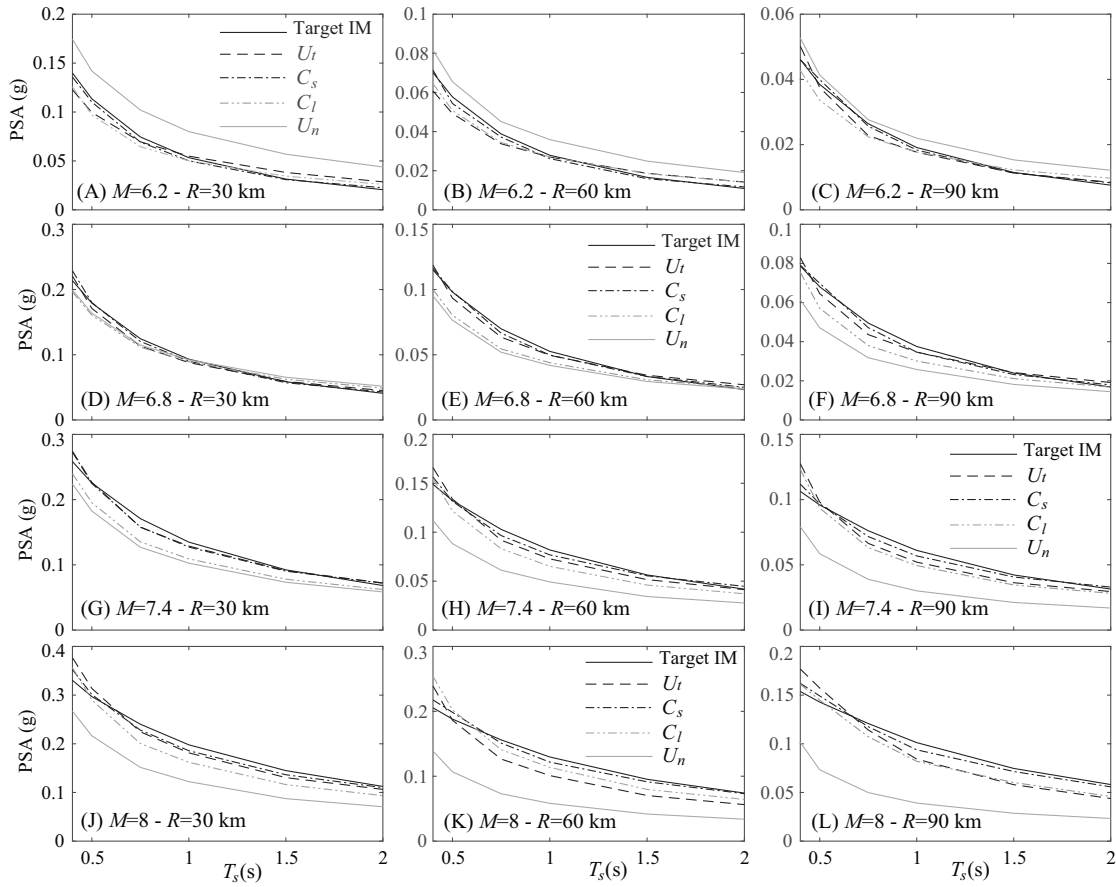
FIGURE 4.11: Spectral plots for seismicity scenarios (different subplots) corresponding to combinations of M [6.2, 6.8, 7.4, 8] and R [30, 60, 90] km, for the target IM (target), the unmodified ground motion model (U_n) and modified ground motion model corresponding to three different selections criteria: minimum distance form utopia point (U_t) and value $\sqrt{F_1}$ smaller than 0.15 (C_l) or 0.05 (C_s). Implementation scenario shown corresponds to matching to the average considered GMPEs and long period range.

optimisation implementation, though the results evidently change.

The spectral plots in Figure 4.11 provide further validation for the observed accuracy of the different examined cases, showing the decomposition of the overall F_1 match to the different structural periods. The ground motion model modification greatly improves the match to the target spectral curves, with C_s modification always providing a better match compared to the C_l one. The U_t modification exhibits a varying behaviour, frequently in between the C_l and C_s cases, in other instances outside their envelope, with characteristics that ultimately depend as discussed above (based on Figure 4.10) on the ability of the unmodified model to

provide an adequate match to the target IMs (i.e., how much is the modification truly needed). Of course the proper evaluation of the proposed modifications comes from examining both F_1 and F_2 values, as detailed above (discussion focusing on Figure 4.10).

The physical ground motion model parameters $\boldsymbol{\theta}$ for the unmodified U_n and modified ground motion models U_t and C_s for the 12 seismicity scenarios of Figure 4.11 are presented in Table C.1 in Appendix C. The spectral acceleration estimates (PSA) shown in Figure 4.11 are presented for the same ground motion model cases along with the target IMs (average of four GMPEs) in Table C.2.
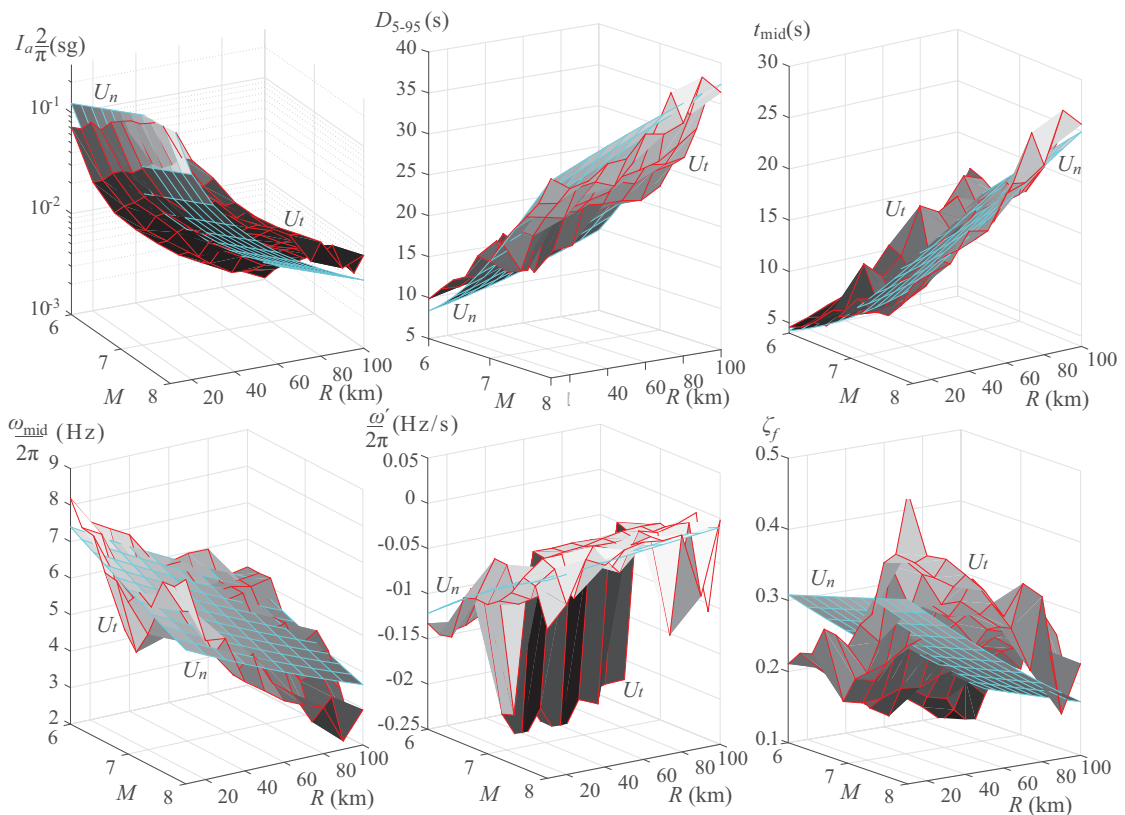
FIGURE 4.12: Physical ground motion model parameters $\boldsymbol{\theta}$ corresponding to unmodified ground motion model (U_n) and modified ground motion model with minimum distance form utopia point (U_t) for implementation scenario of matching to the average considered GMPEs and long period range.

Finally, the results for the ground motion model parameters in Figure 4.12 show that the model modification leads to similar trends as observed for the unmodified model. This is the direct results of incorporating the difference between these

parameters as an objective in the problem formulation (objective F_2) rather than merely as a constraint. Parameters I_a, ζ_f, and ω' show bigger variability compared with their initial values. This should be attributed to a greater sensitivity with respect to them of the resultant ground motions.

The overall discussion shows the importance of the established framework: once the initial metamodel is developed, through the adaptive guidelines established here, it can support the efficient identification of ground motion models that (i) match any desired IMs for any chosen period range while (ii) maintaining a small deviation from the initial predictive relationships. This can be seamlessly repeated for any seismicity scenario. The final ground motion model modification can be chosen based on the criteria discussed earlier.

4.6   Conclusions

The modification of stochastic ground motion models to match target IMs for specific seismicity scenarios was presented in this chapter. This was formulated as a multi-objective optimisation problem with first objective (F_1) quantifying the discrepancy between the ground motion predictions and the target IMs, and the second objective (F_2) corresponding to the discrepancy between the new model characteristics and the model characteristics suggested by existing predictive relationships (i.e., the unmodified model). The second objective explicitly incorporates in the modification process physical characteristics and parameter correlations described in the initial ground motion model. The developed framework facilitates a match to any desired IM or to a collection of them, e.g., spectral accelerations over a period range, for any chosen seismicity scenario. Repeating process for different seismicity scenarios can then facilitate the development of a suite of models that can support comprehensive seismic risk assessment. Computational efficiency was achieved by adopting a metamodel for approximating the median ground motion

model predictions for the targeted IMs. Although the upfront cost for development of this metamodel is significant, once established, it can be used to support a highly efficient multi-objective optimisation. Gradient-based and gradient-free approaches were discussed for the latter whereas to reduce the computational burden for the metamodel development, an adaptive DoE was proposed for selecting the database informing the metamodel.

The framework was demonstrated in an illustrative example considering a recently developed record-based stochastic ground motion model and IMs described through ground motion prediction equations. It was shown in the context of this example that the metamodel-aided optimisation can support an accurate identification of the Pareto front of dominant solutions, provided that the metamodel accuracy is significantly high, and that inclusion of the metamodel error in the optimisation formulation greatly improves the robustness of this optimisation, avoiding the identification of erroneous solutions. Comparisons between the two optimisation approaches showed that the gradient-free one demonstrates overall preferable attributes, since the gradient-based one might converge to suboptimal local minima, especially for lower F_1 values. However, the gradient-based approach provides greater relative efficiency when identification of a single solution, rather than of the entire front, is desired. Still, an adequate representation of the overall Pareto front can be obtained in as little as two minutes using the blind search, gradient-free optimisation, which should be considered as an acceptable computational burden. Different approaches can be then used to select the final ground motion model, e.g., the Pareto optimal point that has the minimum distance from the utopia point or the point that satisfies a desired compatibility to the target IMs (i.e., value of F_1 smaller than a threshold). Implementation for a range of seismicity scenarios showcased the advantages of the proposed framework: small modifications of the original ground motion model (small to moderate F_2 values) provided significant improvement for the match to the target IM (F_1 value) for seismicity ranges where the unmodified model faces challenges in matching the target IMs.

With respect to the selection of the final ground motion model the following suggestion is provided, after carefully examining the various trends observed: select the Pareto optimal solution that satisfies a certain accuracy threshold for F_1 unless this solution leads to a greater F_2 value than the Pareto optimal solution with minimum distance from the utopia point. If the latter happens, then select the Pareto optimal solution with minimum distance from the utopia point.

One of the limitations of this framework is that it utilises only the mean of the predictive relationships of the stochastic ground motion model, whereas, more importantly, match only to some target IMs (e.g., median spectral accelerations from GMPEs) was investigated, ignoring any variability in the IM predictions. This is a significant constraint for seismic risk assessment applications, where hazard compatibility is expressed in terms of both mean and dispersion of some target IMs. Chapter 5 addresses this limitation by extending the proposed framework to modify stochastic ground motion models for hazard compatibility.

Chapter 5

Hazard-compatible modification of stochastic ground motion models

Adapted from Tsioulou, A., Taflanidis, A. A. and Galasso, C. (2018b). Hazard-compatible modification of stochastic ground motion models, *Earthquake Engineering & Structural Dynamics* **47**(8): 1774-1798.

5.1   Introduction

As discussed in Chapter 4, stochastic ground motion models (Rezaeian and Der Kiureghian, 2010; Gavin and Dickinson, 2010; Yamamoto and Baker, 2013; Vlachos et al., 2016; Boore, 2003; Atkinson and Silva, 2000) is as an alternative approach for describing seismic excitations used in PBEE (Goulet et al., 2007; Gidaris and Taflanidis, 2015) and simulation-based probabilistic seismic risk assessment (Au and Beck, 2003; Jensen and Kusanovic, 2014; Gidaris et al., 2016). They are based on a parametric description of the spectral and temporal characteristics of

the excitation, with synthetic time-histories obtained by filtering a stochastic sequence through the resultant frequency and time domain modulating functions. The parameters involved in this description are related to seismicity (e.g., moment magnitude and rupture distance) and site characteristics (e.g., shear wave velocity for soil profile) through predictive models/relationships (Rezaeian and Der Kiureghian, 2010; Boore, 2003). Sample ground motions for a specific seismicity scenario and site can be generated by determining the parameters of the stochastic ground motion model through these predictive relationships and by utilising a sample stochastic sequence. This approach may ultimately support a comprehensive description of the seismic hazard (Gidaris and Taflanidis, 2015), and its essential component is the predictive models relating seismicity/site characteristics to ground motion model parameters.

The two main methodologies for establishing such stochastic ground motion models are record-based and physics-based approaches. Record-based models are developed by fitting a preselected "waveform" to a suite of recorded regional ground motions. Regression analysis is used for establishing the predictive models, which leads to a probabilistic characterisation described by mean predictions along with an associated variability (Rezaeian and Der Kiureghian, 2010; Vlachos et al., 2018). On the other hand, stochastic physics-based models rely on physical modeling of the rupture and wave propagation mechanisms (Boore, 2003; Atkinson and Silva, 2000). The predictive relationships in this case are typically described by deterministic models that represent the underlying mean physical properties, though approaches exist for addressing variability in these properties (Vetter and Taflanidis, 2014; Atkinson, 2008). Similar to Chapter 4, emphasis here will be on record-based models, though the techniques discussed can be extended to any type of stochastic ground motion model, with the assumption that the corresponding predictive models are characterised by both a mean prediction and an associated variance.

As discussed in Chapter 4 important concern related to the use of stochastic ground

motion models for structural engineering applications is the fact that through current approaches in selecting their predictive models, compatibility to the seismic hazard for specific structures and sites is not necessarily obtained. This hazard is typically characterised through PSHA (McGuire, 2004), for example through disaggregation that identifies the seismicity scenarios, described through relevant seismicity characteristics, mainly the moment magnitude and rupture distance, that have the largest contribution to the hazard for a specific structure. Essential part of PSHA are the GMPEs. GMPEs provide predictions, as function of seismicity characteristics, for both the median and the dispersion of IMs, determining ultimately the conditional hazard for seismic events corresponding to these characteristics. In other words, for a given seismicity scenario, defined for example by the moment magnitude and source-to-site distance, the conditional hazard is described through the mean and the dispersion of some structure-specific IMs.

Recognising the importance of matching stochastic ground motion models to some target IM, the modification of the stochastic ground motion model for accommodating such a match was examined in Chapter 4. The proposed modification was performed for specific seismicity scenarios and identified the ground motion model that achieves the minimum modification of the existing predictive relationships that will yield the desired compatibility with the target IM.

The study in Chapter 4 and all past studies focused, though, on the mean model characteristics and associated hazard. Optimisation utilised only the mean of the predictive relationships of the stochastic ground motion model, whereas, more importantly, match only to a target IM vector (taken to represent the mean hazard) was investigated, ignoring any variability in the IM predictions. The latter is an important constraint because for seismic risk assessment applications hazard compatibility is expressed in terms of both the mean and dispersion of the target IMs (McGuire, 2004). From a practical standpoint, capturing the actual variability of target IMs is essential to capture extreme structural response values and, therefore, in properly assessing the likelihood of consequences due to such extreme seismic demand values.

The current chapter extends the approach presented in Chapter 4 to (i) match the prescribed conditional hazard (not simply mean IMs) for a specific site and structure (or range of structures) while (ii) preserving desired trends and correlations in the physical characteristics of the resultant ground acceleration time-series, including consideration of the variability of these characteristics. This is again formulated as a bi-objective optimisation problem. The first objective is to minimise the discrepancy between the statistics (mean and dispersion) of the outputs for a suite of acceleration time-histories obtained from the ground motion model and the target IM statistics for a given seismicity scenario. The second objective is to establish the smallest deviation from the model characteristics suggested by existing predictive models, examining both the mean and the variance of the model parameters. Both objectives are expressed as comparison between probability distributions, and the relative entropy is adopted to quantify them. This setting creates a fundamental difference to the previous study in Chapter 4 with respect to both the optimisation characteristics (alter both mean and variance of predictive models) as well as the goal (match to mean and dispersion for conditional hazard). In Chapter 4 the ground motion model was tuned so that outputs from a single parametric description of the ground motion model match a target IM vector for each seismicity scenario. The goal of the current study is to produce an ensemble of ground motion models whose output statistics yield the desired compatibility with the hazard (IM mean and dispersion) for that scenario. For efficiently solving the multi-objective optimisation problem a surrogate modeling approach is adopted (Vetter et al., 2016; Tsioulou et al., 2018a) for approximating the desired IMs for specific values of the ground motion model parameters. Emphasis is placed here on the efficient estimation of the response statistics for the modified ground motion model output, leveraging Monte Carlo techniques. This requires further extension of the surrogate modeling framework, compared with the approach adopted in Chapter 4, for facilitating this estimation. Different assumptions are also examined for the evaluation of the entropy for the first objective. The corresponding bi-objective optimisation is finally solved using a random

search approach. The novelty of the current work stems from both the fundamentally different theoretical framework for formulating the stochastic ground motion modification as well as the computational advances required for efficiently calculating the response statistics for the first objective and performing the associated optimisation.

In the next section, the general problem of developing simulated ground motions compatible with target IM distributions is defined, and then specific aspects of the framework are discussed in detail.

5.2   Problem formulation

5.2.1   Preliminaries and baseline predictive relationships formulation

The foundation of the problem formulation is the same as in Chapter 4. A stochastic ground motion model is considered that provides acceleration time-histories $\ddot{\alpha}(t|\boldsymbol{\theta}, \boldsymbol{w})$ by modulating a discretised Gaussian white-noise sequence, $\boldsymbol{w}$, through appropriate time/frequency functions that are parameterised through the n_θ-dimensional model parameter vector $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_{n_\theta}] \in \mathbb{R}^{n_\theta}$. $\boldsymbol{\theta}$ completely defines the parametric description of the model, i.e., along with $\boldsymbol{w}$ facilitates the generation of a time-history $\ddot{\alpha}(t|\boldsymbol{\theta}, \boldsymbol{w})$, and is typically composed of various excitation properties such as Arias intensity, strong ground motion duration or parameters related to frequency characteristics of the ground motion. It should be noted that $\boldsymbol{\theta}$ corresponds typically to a low-dimensional vector and $\boldsymbol{w}$ to a high-dimensional sequence.

Predictive models/relationships are utilised to relate $\boldsymbol{\theta}$ to seismicity and local site characteristics, such as the fault type F, the moment magnitude, M, the rupture distance, R, and the shear wave velocity in the upper 30 m of soil, $V_{s,30}$ (Rezaeian

and Der Kiureghian, 2010; Boore, 2003). The vector of these characteristics is denoted as $\boldsymbol{z}$. For record-based models the standard approach for development of these predictive relationships (Rezaeian and Der Kiureghian, 2010; Medel-Vera and Ji, 2016) relies on first matching the waveform of recorded ground motions (i.e., identify first $\boldsymbol{\theta}$ for each of the recorded ground motions in a given database) and then carrying out a regression to relate $\boldsymbol{\theta}$ to $\boldsymbol{z}$. Typically, this is performed by first transforming problem to the standard Gaussian space through a nonlinear mapping for each component θ_i (Rezaeian and Der Kiureghian, 2010). The transformed Gaussian vector is denoted $\boldsymbol{v}$. Approach ultimately leads to a Gaussian probability model $\boldsymbol{v} \sim N(\boldsymbol{\mu}_r(\boldsymbol{z}), \boldsymbol{\Sigma}_r)$ with mean $\boldsymbol{\mu}_r(\boldsymbol{z})$ and covariance matrix $\boldsymbol{\Sigma}_r$. Note that the latter is independent of $\boldsymbol{z}$. The notation $N(a, b)$ stands for Normal distribution with mean α and covariance b whereas notation $c \sim d$ stands for random variable c following distribution d. The resultant probability model for $\boldsymbol{\theta}$ is denoted $p(\boldsymbol{\theta}|\boldsymbol{\mu}_r(\boldsymbol{z}), \boldsymbol{\Sigma}_r)$ and determines the predictive model for the ground motion model parameters. Note that a similar description can be readily established for physics-based models (Tsioulou et al., 2018a), with the uncertainty characterisation stemming from an explicit treatment of the epistemic uncertainties associated with the physics-based formulation (Vetter and Taflanidis, 2014; Atkinson, 2008). The predictive model for $\boldsymbol{\theta}$ will be denoted herein as $p(\boldsymbol{\theta}|\boldsymbol{\mu}_r(\boldsymbol{z}), \boldsymbol{\Sigma}_r)$, with the understanding that it is not necessarily constrained to models established through regression analysis, rather simply can be parameterised by quantities $\boldsymbol{\mu}_r(\boldsymbol{z})$ and $\boldsymbol{\Sigma}_r$, with $\boldsymbol{\mu}_r(\boldsymbol{z})$ representing the mean predictions, and $\boldsymbol{\Sigma}_r$ the variability of these predictions.

5.2.2 Modification of predictive models for hazard matching

As discussed in Section 5.1, the formulation of the predictive model for $\boldsymbol{\theta}$ provides synthetic ground motions whose statistics (mean and dispersion) of output IMs do not necessarily match the intended hazard for specific structures and sites. For

accommodating such a match a modification of the existing predictive model for $\boldsymbol{\theta}$ is proposed for specific seismicity scenarios defined through $\boldsymbol{z}$, with objective to get a suite of acceleration time-series for that scenario whose (i) mean and dispersion match a target IM mean and dispersion vectors, while (ii) maintaining similarity to the predictive relationships already established for the model. Equivalently this can be viewed as identifying the updated probabilistic model $p(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ that is closest to the established model $p(\boldsymbol{\theta}|\boldsymbol{\mu}_r(\boldsymbol{z}), \boldsymbol{\Sigma}_r)$ and also matches the intended conditional hazard. $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ represent the updated parametric description for the probability model of $\boldsymbol{\theta}$. In the context of record-based models these correspond, respectively, to the mean vector and covariance matrix for $\boldsymbol{v}$. The IM vector may include different response quantities of interest as discussed in Chapter 4, e.g., direct characteristics of the ground motion, such as PGA or elastic and inelastic spectral responses for different periods of an SDoF oscillator. The conditional hazard for most of these IMs may be described through a GMPE (Bozorgnia et al., 2010; Stewart et al., 2016). Since the proposed modification refers to the conditional hazard, for simplifying terminology for the remainder of the paper, the description as "conditional" will be removed: term hazard corresponds to conditional hazard.

To formalise these concepts mathematically, let, $Y_i(\boldsymbol{z}); i = 1, \ldots, n_y$ denote the IMs of interest. The target hazard for them is quantified through a probabilistic description, denoted by $p_t(\ln(Y_i)|\boldsymbol{z})$. As is common in earthquake engineering and without loss of generality, the statistics are assumed here to be determined for the logarithm of the IM. To better align approach with current GMPE standards a log-normal underlying model is assumed, leading to $\ln(Y_i(\boldsymbol{z}))^t \sim N(\ln(\bar{Y}_i(\boldsymbol{z})), \sigma_i^2(\boldsymbol{z}))$ with $\ln(\bar{Y}_i(\boldsymbol{z}))$ and $\sigma_i^2(\boldsymbol{z})$ corresponding to the mean and variance, respectively, of the logarithmic IM. Note, though, that the computational framework can support any probabilistic IM description $p_t(\ln(Y_i)|\boldsymbol{z})$, not constrained to one provided by GMPEs or defined through a lognormal probabilistic model. Also, the superscripts/subscripts t and g (the latter defined in the next paragraph) are used herein to distinguish between target prediction and prediction facilitated through

the stochastic ground motion model.

To quantify the hazard predictions through the stochastic ground motion model, let $Y_i^g(\boldsymbol{\theta}, \boldsymbol{w})$ denote the estimate for Y_i established through this model for specific values of the model parameter vector $\boldsymbol{\theta}$ and a specific white-noise sequence $\boldsymbol{w}$ (i.e., for a specific ground motion time-history $\ddot{\alpha}(t|\boldsymbol{\theta}, \boldsymbol{w})$). $Y_i^g(\boldsymbol{\theta}, \boldsymbol{w})$ will be referenced herein as response output of the ground motion model. The statistical characterisation for $\ln(Y_i)$ through the stochastic ground motion model for the updated parametric description is denoted by $p_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and using the total probability theorem is equal to:

$$
\begin{aligned}
p_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \iint p(\ln(Y_i^g)|\boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) p(\boldsymbol{\theta}, \boldsymbol{w}, |\boldsymbol{\mu}, \boldsymbol{\Sigma}) \, d\boldsymbol{\theta} \, d\boldsymbol{w} \\
&= \iint p(\ln(Y_i^g)|\boldsymbol{\theta}, \boldsymbol{w}) p(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) p(\boldsymbol{w}) \, d\boldsymbol{\theta} \, d\boldsymbol{w}
\end{aligned}
\tag{5.1}
$$

where $p(\boldsymbol{w})$ is the probability distribution for the stochastic sequence $\boldsymbol{w}$, which by definition is independent of $p(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In deriving the second equality in Equation 5.1 the fact that the stochastic ground motion model is completely defined by pair $\boldsymbol{\theta}, \boldsymbol{w}$ was also used. This simplifies $p(\ln(Y_i^g)|\boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = p(\ln(Y_i^g)|\boldsymbol{\theta}, \boldsymbol{w})$ because knowledge of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is redundant if $\boldsymbol{\theta}$ is also known. Note that Y_i^g is itself a random variable with randomness in its description stemming from both $\boldsymbol{w}$ and $\boldsymbol{\theta}$. In this context, $Y_i^g(\boldsymbol{\theta}, \boldsymbol{w})$ may be viewed as a realisation of the random variable.

Distribution $p_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ can be approximated through KDE as discussed in Appendix B. To further simplify framework, the same hypothesis as for the hazard can be adopted, assuming a lognormal distribution, leading to $\ln(Y_i)^g \sim N(\ln(\bar{Y}_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma})), (\sigma_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma}))^2)$, where the mean and variance for $\ln(Y_i)^g$, considering the variability in both the low-dimensional $\boldsymbol{\theta}$ and high-dimensional $\boldsymbol{w}$, are:

$$
\ln(\bar{Y}_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma})) = \mathrm{E}[\ln(Y_i)^g] = \iint \ln(Y_i^g(\boldsymbol{\theta}, \boldsymbol{w})) p(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) p(\boldsymbol{w}) \, d\boldsymbol{\theta} \, d\boldsymbol{w}
\tag{5.2}
$$

$$\sigma_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma}))^2 = \text{Var}[\ln(Y_i^g)] = \iint \left[\ln(Y_i^g(\boldsymbol{\theta}, \boldsymbol{w})) - \ln(\bar{Y}_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma}))\right]^2 p(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})p(\boldsymbol{w}) \, d\boldsymbol{\theta} \, d\boldsymbol{w}$$

(5.3)

with E[.], Var[.] denoting the expectation and variance operators, respectively. The functional dependence of the statistics of $\ln(Y_i(\boldsymbol{z}))^g$ on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is explicitly noted herein to facilitate an easier understanding of the ground motion model modification framework. The lognormal assumption will be primarily used for $p_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and its validity will be examined in the illustrative example later by comparing to the KDE-based estimation.

The hazard-compatible modeling corresponds to modification of the probability model for $\boldsymbol{\theta}$, ultimately of the parametric description defined through $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, and is formulated as multi-objective optimisation problem with two competing objectives

$$[\boldsymbol{\mu}, \boldsymbol{\Sigma}]^* = \arg\min \{F_{p1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\boldsymbol{z}), F_{p2}(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\boldsymbol{z})\} \tag{5.4}$$

The first objective F_{p1} corresponds to the weighted discrepancy of the target seismic hazard to the hazard predicted through the ground motion model, i.e., to a comparison between $p_g(\ln(Y_i(\boldsymbol{z}))|\boldsymbol{\mu}(\boldsymbol{z}), \boldsymbol{\Sigma})$ and $p_t(\ln(Y_i)|\boldsymbol{z})$. The relative entropy, a popular measure to quantify differences between distributions (Gibbs and Su, 2002), is utilised as metric for this discrepancy, given by:

$$D[p_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma})||p_t(\ln(Y_i)|\boldsymbol{z})] = \int_{\mathbb{R}} p_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \log\left[\frac{p_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma})}{p_t(\ln(Y_i)|\boldsymbol{z})}\right] \, d\ln(Y_i)$$

(5.5)

This leads to definition of F_{p1} as

$$F_{p1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\boldsymbol{z}) = \frac{1}{\sum_{i=1}^{n_y} \gamma_i} \sum_{i=1}^{n_y} \gamma_i D[p_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma})||p_t(\ln(Y_i)|\boldsymbol{z})]$$

$$= \frac{1}{\sum_{i=1}^{n_y} \gamma_i} \sum_{i=1}^{n_y} \gamma_i D[p_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma})||N(\ln(\bar{Y}_i(\boldsymbol{z})), \sigma_i^2(\boldsymbol{z}))] \tag{5.6}$$

with γ_i corresponding to the weights prioritising the match to different IM components (e.g., spectral accelerations at different structural periods). For the assumption of lognormal distribution for Y_i^g, this simplifies to

$$F_{p1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\boldsymbol{z}) = \frac{1}{\sum_{i=1}^{n_y} \gamma_i} \sum_{i=1}^{n_y} \gamma_i D[N(\ln(\bar{Y}_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma})), (\sigma_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma}))^2)||N(\ln(\bar{Y}_i(\boldsymbol{z})), \sigma_i^2(\boldsymbol{z}))]$$

$$\tag{5.7}$$

with closed-form solution (Gibbs and Su, 2002):

$$D[N(\ln(\bar{Y}_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma})), (\sigma_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma}))^2)||N(\ln(\bar{Y}_i(\boldsymbol{z})), \sigma_i^2(\boldsymbol{z}))] =$$

$$\frac{\left(\ln(\bar{Y}_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma})) - \ln(\bar{Y}_i(\boldsymbol{z}))\right)^2}{2\sigma_i^2(\boldsymbol{z})} + \frac{1}{2} \left[\frac{(\sigma_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma}))^2}{\sigma_i^2(\boldsymbol{z})} - 1 - \ln\left(\frac{(\sigma_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma}))^2}{\sigma_i^2(\boldsymbol{z})} \right) \right] \tag{5.8}$$

Objective F_{p2} measures that discrepancy between the initial predictive model for $\boldsymbol{\theta}$, $p(\boldsymbol{\theta}|\boldsymbol{\mu}_r(\boldsymbol{z}), \boldsymbol{\Sigma}_r)$, and the modified one, $p(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The relative entropy is utilised again as measure to quantify differences:

$$F_{p2}(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\boldsymbol{z}) = D[p(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})||p(\boldsymbol{\theta}|\boldsymbol{\mu}_r(\boldsymbol{z}), \boldsymbol{\Sigma}_r)]$$

$$= \int_{\mathbb{R}^{n_\theta}} p(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \log\left[\frac{p(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})}{p(\boldsymbol{\theta}|\boldsymbol{\mu}_r(\boldsymbol{z}), \boldsymbol{\Sigma}_r)} \right] d\theta \tag{5.9}$$

Since the relative entropy is invariant under a coordinate transformation for $\boldsymbol{\theta}$ (Gibbs and Su, 2002), the comparison can be established in the transformed Gaussian space, leading to

$$F_{p2}(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\boldsymbol{z}) = D[N(\boldsymbol{\mu}, \boldsymbol{\Sigma})||N(\boldsymbol{\mu}_r(\boldsymbol{z}), \boldsymbol{\Sigma}_r)] \qquad (5.10)$$

with the latter expression readily evaluated (Gibbs and Su, 2002) as

$$
\begin{aligned}
&D[N(\boldsymbol{\mu}, \boldsymbol{\Sigma})||N(\boldsymbol{\mu}_r(\boldsymbol{z}), \boldsymbol{\Sigma}_r)] = \\
&1/2 \left[tr[\boldsymbol{\Sigma}\boldsymbol{\Sigma}_r^{-1}] + (\boldsymbol{\mu}_r(\boldsymbol{z}) - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_r^{-1}(\boldsymbol{\mu}_r(\boldsymbol{z}) - \boldsymbol{\mu}) - n_\theta - \ln\left(\det[\boldsymbol{\Sigma}\boldsymbol{\Sigma}_r^{-1}]\right) \right]
\end{aligned}
\qquad (5.11)
$$

where $tr[.]$ and $\det[.]$ stand for trace and determinant, respectively.

Objective F_{p1} enforces hazard compatibility (mean and dispersion), whereas objective F_{p2} guarantees compatibility of the physical characteristics of the resultant ground motions with the regional trends observed in recorded ground motions. Solution of the multi-objective optimisation of Equation 5.4 ultimately leads to a Pareto set of dominant solutions $\{(\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p); p = 1, \ldots, n_p\}$ that expresses a different compromise between the competing objectives F_{p1} and F_{p2}. A solution is characterised as dominant (or Pareto optimal) and belongs in the Pareto set if there is no other solution that simultaneously improves both objectives F_{p1} and F_{p2}. The representation of the Pareto set in the performance objective $[F_{p1}, F_{p2}]$ space, $\{[F_{p1}(\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p), F_{p2}(\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p)]; p = 1, \ldots, n_p\}$ is termed as the Pareto front. Illustrations of such Pareto fronts are included in the example discussed later. One can eventually select a model configuration from the identified Pareto set that yields the desired hazard compatibility without deviating significantly from regional ground motion characteristics. This will be further discussed in the illustrative implementation. Note that a Pareto set for optimisation of Equation 5.4 always exists apart from the extreme case that the original stochastic ground motion model provides an exact match to the hazard, i.e., $F_{p1}(\boldsymbol{\mu}_r(\boldsymbol{z}), \boldsymbol{\Sigma}_r|\boldsymbol{z})=0$. In that case, the entire set corresponds to a single point $(\boldsymbol{\mu}_r(\boldsymbol{z}), \boldsymbol{\Sigma}_r)$.

Identifying the Pareto set is, though, challenging because the computational burden in evaluation of objective F_{p1} is significant, requiring calculation of the multidimensional integrals of Equations 5.2 and 5.3 (or samples needed for the KDE approximation discussed in Appendix B) which can be only performed numerically and entails hundreds estimates of the response output of the ground motion model. To facilitate an efficient optimisation that can be repeated for any desired seismicity scenario $\boldsymbol{z}$, a surrogate modeling (i.e., metamodeling) approach is adopted, specifically selecting kriging as metamodel due to its proven capability to approximate well even complex functions (Lophaven et al., 2002). As input for the metamodel, the low-dimensional vector $\boldsymbol{\theta}$ is chosen. This choice corresponds to the smallest possible dimension for the metamodel input, something that can greatly enhance accuracy (Lophaven et al., 2002), a very important consideration since metamodel will be eventually used for optimisation, and larger metamodel errors can lead to identification of suboptimal solutions. Alternative choices were to additionally include $\boldsymbol{w}$ in the metamodel input definition which is completely impractical (due to high-dimensionality of $\boldsymbol{w}$), or to use directly pair $\{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}$ as the metamodel inputs (and statistics of Equations 5.2 and 5.3 as outputs) which significantly increases, however, the input dimension. Under this selection of $\boldsymbol{\theta}$ as metamodel input, the metamodel output corresponds to the response output statistics considering variability with respect to $\boldsymbol{w}$ only (i.e., conditional on $\boldsymbol{\theta}$ statistics). Here statistics refers to all quantities needed to eventually calculate F_{p1}. The variability with respect to $\boldsymbol{\theta}$ in quantifying the hazard (and ultimately objective F_{p1}) will be addressed through the approach discussed in Section 5.4.

This creates a similar setting as in Chapter 4 for the metamodel formulation, with the additional requirement, though, to eventually incorporate variability stemming from $\boldsymbol{\theta}$ in the overall framework, so that variability in both $\boldsymbol{\theta}$ and $\boldsymbol{w}$ is explicitly considered. The kriging metamodel development is briefly reviewed in the next section, and the details of the kriging-aided optimisation problem are discussed in Section 5.4. A schematic of the overall optimisation approach is provided in Figure 5.1.

Common (pre-optimisation) tasks

Select stochastic ground motion model parameterised through vector $\boldsymbol{\theta}$, and initial predictive model for it $p(\boldsymbol{\theta}|\boldsymbol{\mu}_r(\mathbf{z}),\boldsymbol{\Sigma}_r)$ *

Assuming ground motion model includes a global scaling parameter θ_s set metamodel input $\mathbf{x}$ to $\boldsymbol{\theta}$ excluding θ_s and define normalized response $s_i(\mathbf{x},\mathbf{w})$ to $Y_i^g(\boldsymbol{\theta},\mathbf{w})$ for $\theta_s{=}1$

Develop **surrogate model** for approximating statistics (logarithmic mean and variance) of $s_i(\mathbf{x},\mathbf{w})$ with respect to $\mathbf{w}$ $\ln(\bar{s}_i(\mathbf{x}))$ and $(\sigma_i^s(\mathbf{x}))^2$ for all desired IMs {details in Chapter 4}

* Other terminology used
$\mathbf{z}$: vector of seismicity/site characteristics defining predictive model for $\boldsymbol{\theta}$
$\mathbf{w}$: stochastic sequence
$Y_i^g(\boldsymbol{\theta},\mathbf{w})$:IM prediction through ground motion model

Set ground motion modification problem

Select seismicity scenario defined through $\mathbf{z}$ Define IMs of interest Y_i $\{i{=}1,\ldots,n_y\}$, the target conditional hazard for them $N(\ln(\overline{Y}_i(\mathbf{z})),\sigma_i^2(\mathbf{z}))$ and weights γ_i

Random search optimisation (repeat for new seismicity scenarios)

Generate candidate solutions for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ Generate common random number samples to use in Monte Carlo simulation (MCS)

Repeat for each candidate solution $\{\boldsymbol{\mu},\boldsymbol{\Sigma}\}$

Calculate statistics $\ln(\overline{Y}_i^g(\boldsymbol{\mu},\boldsymbol{\Sigma}))$ and $(\sigma_i^g(\boldsymbol{\mu},\boldsymbol{\Sigma}))^2$ through Eqs. (14-15), evaluating analytically statistics involving only θ_s and using MCS for components in Eqs. (16-18), utilising the **surrogate model** predictions for the latter

Calculate $F_{p2}(\boldsymbol{\mu},\boldsymbol{\Sigma}\,|\,\mathbf{z}) = D\big[N(\boldsymbol{\mu},\boldsymbol{\Sigma})\,\|\,N(\boldsymbol{\mu}_r(\mathbf{z}),\boldsymbol{\Sigma}_r)\big]$ using Eq.(11)

Calculate using Eq. (8) [or KDE approach] $F_{p1}(\boldsymbol{\mu},\boldsymbol{\Sigma}\,|\,\mathbf{z}) =$
$$\frac{\sum_{i=1}^{n_y}\gamma_i D\Big[N(\ln(\overline{Y}_i^g(\boldsymbol{\mu},\boldsymbol{\Sigma})),(\sigma_i^g(\boldsymbol{\mu},\boldsymbol{\Sigma}))^2)\,\|\,N(\ln(\overline{Y}_i(\mathbf{z})),\sigma_i^2(\mathbf{z}))\Big]}{\sum_{i=1}^{n_y}\gamma_i}$$

(after evaluation for all candidate solutions)

Identify dominant solutions among candidate solutions → Pareto set   Choose final model using any desired criterion

FIGURE 5.1: Overview schematic of the hazard-compatible stochastic ground motion modeling approach.

5.3 Kriging metamodel development

As discussed in the previous section, $\boldsymbol{\theta}$ is chosen as metamodel input. A modification of this input should be further adopted if the relationship between some components of $\boldsymbol{\theta}$ and the response output $Y_i^g(\boldsymbol{\theta},\boldsymbol{w})$ is explicitly known. This is true for stochastic ground motion models that include a scaling parameter, denoted θ_s herein, that directly controls the amplitude of the excitation as shown in Chapter 4. For example, for the model that will be used in the illustrative example (Rezaeian and Der Kiureghian, 2010) θ_s corresponds to the Arias intensity. This means that $Y_i^g(\boldsymbol{\theta},\boldsymbol{w}) = f(\theta_s)s_i(\boldsymbol{x},\boldsymbol{w})$ with $\boldsymbol{x}$ corresponding to the remaining model parameters excluding θ_s, representing the output $Y_i^g(\boldsymbol{\theta},\boldsymbol{w})$ for $\theta_s{=}1$ and $f(\theta_s)$ being a simple function of θ_s. Without loss of generality we will adopt here this assumption, i.e., representation $Y_i^g(\boldsymbol{\theta},\boldsymbol{w}) = \sqrt{\theta_s}s_i(\boldsymbol{x},\boldsymbol{w})$. The choice of

square root for function $f(.)$ is made here simply to match the model used in the illustrative example. Numerical approach discussed herein can accommodate any other function. This setting leads to modification of the metamodel input to $\boldsymbol{x}$, and respectively of metamodel output to the statistics of $s_i(\boldsymbol{x}, \boldsymbol{w})$ conditional on $\boldsymbol{x}$. Once the latter statistics are known, the statistics for $Y_i^g(\boldsymbol{\theta}, \boldsymbol{w})$ conditional on $\boldsymbol{\theta}$ can be easily obtained as will be shown in Section 5.4, because the relationship to the remaining component of vector $\boldsymbol{\theta}$, θ_s, is explicitly known. This modification ultimately reduces the dimension of the metamodel input, which as discussed earlier facilitates improved accuracy. Note that similar to Y_i^g, s_i is also a random variable with randomness in its description stemming from $\boldsymbol{w}$ and $\boldsymbol{x}$.

The approximated statistics, the ones required for calculating F_{p1}, are the conditional on $\boldsymbol{x}$ mean and variance of $\ln(s_i)$, given by

$$\ln(\bar{s}_i(\boldsymbol{x})) = \mathrm{E}[\ln(s_i)|\boldsymbol{x}] = \int \ln(s_i(\boldsymbol{x}, \boldsymbol{w}))p(\boldsymbol{w})\,d\boldsymbol{w} \tag{5.12}$$

$$(\sigma_i^s(\boldsymbol{x}))^2 = \mathrm{Var}[\ln(s_i)|\boldsymbol{x}] = \int \left[\ln(s_i(\boldsymbol{x}, \boldsymbol{w})) - \ln(\bar{s}_i(\boldsymbol{x}))\right]^2 p(\boldsymbol{w})\,d\boldsymbol{w} \tag{5.13}$$

where independence of $\boldsymbol{w}$ from $\boldsymbol{x}$ was used in deriving the last equalities in both expressions. These statistics define the metamodel output. It should be pointed out that this output is independent of the predictive relationships; rather is simply functions of $\boldsymbol{x}$ (reason for introducing the functional dependence on $\boldsymbol{x}$ notation). This is what contributes to the efficiency of the approach: the surrogate model is established with respect to the low-dimensional vector $\boldsymbol{x}$ and can be then leveraged to evaluate the required statistics for different selections of the predictive models (i.e. different $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$) as detailed in the next section.

For developing the metamodel, a database with n observations is initially obtained that provides information for the $\boldsymbol{x} - \ln(\bar{s}_i(\boldsymbol{x}))$ and $\boldsymbol{x} - \sigma_i^s(\boldsymbol{x})$ input/output pairs. For this purpose, n samples for $\{\boldsymbol{x}^j; j = 1, \ldots, n\}$ also known as support

points or experiments are obtained over the domain of interest for $\boldsymbol{x}$. This domain, denoted X, should encompass the anticipated range that the metamodel will be implemented in to avoid extrapolations, i.e., domain covered by $p(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, approximated in this case as domain for $p(\boldsymbol{\theta}|\boldsymbol{\mu}_r(\boldsymbol{z}), \boldsymbol{\Sigma}_r)$. The definition of X is discussed in detail in Chapter 4. The predictions provided through the ground motion model for each $\boldsymbol{x}^j$ are then established through the following process:

Step 1: Generate n_w sample acceleration time-histories for different white-noise sequences $\{\boldsymbol{w}^k; k = 1, \ldots, n_w\}$ using $\theta_s{=}1$ for all samples;

Step 2: For each sample evaluate the response outputs of interest $\{s_i(\boldsymbol{x}^j, \boldsymbol{w}^k); i = 1, \ldots, n_y\}$ using response-history analysis for spectral IMs;

Step 3: Estimate the statistics (logarithmic mean and variance) over the sample-set to obtain $\ln(\bar{s}_i(\boldsymbol{x}^j))$ and $\sigma_i^s(\boldsymbol{x}^j)$.

Using this database the surrogate model can be formulated following the approach presented in Chapter 4. Only difference is that the output for metamodel also includes $\sigma_i^s(\boldsymbol{x})$. The computationally intensive aspect of the formulation is the development of the database which requires response-history analysis. This needs to be performed, though, only once. As soon as the metamodel is established using this database, it can predict $\ln(\bar{s}_i(\boldsymbol{x}))$ and $\sigma_i^s(\boldsymbol{x})$ for any other $\boldsymbol{x}$ desired. Metamodel predictions can be also vectorised (Jia and Taflanidis, 2013), something that will be leveraged in the numerical optimisation discussed in the next section. The accuracy of the metamodel depends on the number of experiments n as well as the exact selection of these experiments. Chapter 4 presents details for both these tasks, including a sequential, adaptive metamodel formulation that gradually increases n, stopping when pre-specified accuracy criteria are satisfied.

5.4 Multi-objective optimisation to match conditional hazard supported by Kriging meta-modeling

5.4.1 Calculation of statistics of interest

Evaluation of objective F_{p1} boils down to estimation of Equations 5.2 and 5.3. Using representation $Y_i^g(\boldsymbol{\theta}, \boldsymbol{w}) = \sqrt{\theta_s} s_i(\boldsymbol{x}, \boldsymbol{w})$ this simplifies to

$$\ln(\bar{Y}_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma})) = \mathrm{E}[\ln(Y_i^g)] = \mathrm{E}[\ln(\sqrt{\theta_s} s_i)] = \mathrm{E}[\ln(\theta_s)]/2 + \mathrm{E}[\ln(s_i)] \qquad (5.14)$$

$$(\sigma_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma}))^2 = \mathrm{Var}[\ln(\sqrt{\theta_s} s_i)] = \mathrm{Var}[\ln(\theta_s)]/4 + \mathrm{Var}[\ln(s_i)] + \mathrm{Cov}[\ln(\theta_s), \ln(s_i)]$$
$$(5.15)$$

where $\mathrm{Cov}[a, b]$ stands for the covariance between random variables a and b. The statistics with respect to θ_s, that is the mean $\mathrm{E}[\ln(\theta_s)]$ and variance $\mathrm{Var}[\ln(\theta_s)]$ can be readily calculated using the marginal distribution $p(\theta_s | \boldsymbol{\mu}, \boldsymbol{\Sigma})$. The statistics that involve s_i may be calculated using the metamodel to approximate variability with respect to $\boldsymbol{w}$. Using the laws of total expectation and variance we have for these statistics

$$\mathrm{E}[\ln(s_i)] = \mathrm{E}[\mathrm{E}[\ln(s_i)|\boldsymbol{x}]] = \mathrm{E}[\ln(\bar{s}_i(\boldsymbol{x}))] = \int \ln(\bar{s}_i(\boldsymbol{x})) p(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \, d\boldsymbol{x} \qquad (5.16)$$

$$\mathrm{Var}[\ln(s_i)] = \mathrm{E}[\mathrm{Var}[\ln(s_i)|\boldsymbol{x}]] + \mathrm{Var}[\mathrm{E}[\ln(s_i)|\boldsymbol{x}]] = \mathrm{E}[(\sigma_i^s(\boldsymbol{x}))^2] + \mathrm{Var}[\ln(\bar{s}_i(\boldsymbol{x}))]$$

$$= \int (\sigma_i^s(\boldsymbol{x}))^2 p(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \, d\boldsymbol{x} + \int (\ln(\bar{s}_i(\boldsymbol{x})) - \mathrm{E}[\ln(\bar{s}_i(\boldsymbol{x}))])^2 p(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \, d\boldsymbol{x}$$

$$(5.17)$$

$$\begin{aligned}
\mathrm{Cov}[\ln(\theta_s), \ln(s_i)] &= \mathrm{E}[\ln(\theta_s) \ln(s_i)] - \mathrm{E}[\ln(\theta_s)] \mathrm{E}[\ln(s_i)] \\
&= \mathrm{E}[\mathrm{E}[\ln(\theta_s) \ln(s_i)|\boldsymbol{x}]] - \mathrm{E}[\ln(\theta_s)] \mathrm{E}[\ln(s_i)] \\
&= \mathrm{E}[\ln(\theta_s) \mathrm{E}[\ln(s_i)|\boldsymbol{x}]] - \mathrm{E}[\ln(\theta_s)] \mathrm{E}[\ln(s_i)] \\
&= \mathrm{E}[\ln(\theta_s) \ln(\bar{s}_i(\boldsymbol{x}))] - \mathrm{E}[\ln(\theta_s)] \mathrm{E}[\ln(s_i)] \\
&= \int \ln(\theta_s) \ln(\bar{s}_i(\boldsymbol{x})) p(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \, d\boldsymbol{\theta} - \mathrm{E}[\ln(\theta_s)] \mathrm{E}[\ln(s_i)]
\end{aligned}$$

$$(5.18)$$

where the expectation (integrals) with respect to $\boldsymbol{x}$ or $\boldsymbol{\theta}$ have been explicitly expressed in all these equations. These integrals address the variability stemming from the predictive relationships and need to be calculated through Monte Carlo simulation (MCS). Rather than performing two different MCS: one for the integrals involved in the expectation $\mathrm{E}[\ln(s_i)]$ and variance $\mathrm{Var}[\ln(s_i)]$, which require samples from the marginal distribution $p(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and one for the covariance $\mathrm{Cov}[\ln(\theta_s), \ln(s_i)]$, which requires samples from joint distribution $p(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, a single MCS is performed, utilising a common set of samples for all these statistics. This leads to the following approximations for the quantities needed for Equations 5.14 and 5.15

$$\mathrm{E}[\ln(s_i)] \approx \frac{1}{N_s} \sum_{j=1}^{N_s} \ln(\bar{s}_i(\boldsymbol{x}^j)) \tag{5.19}$$

$$\mathrm{Var}[\ln(s_i)] \approx \frac{1}{N_s} \sum_{j=1}^{N_s} (\sigma_i^s(\boldsymbol{x}^j))^2 + \frac{1}{N_s} \sum_{j=1}^{N_s} (\ln(\bar{s}_i(\boldsymbol{x}^j)) - \mathrm{E}[\ln(s_i)])^2 \tag{5.20}$$

$$\text{Cov}[\ln(\theta_s), \ln(s_i)] \approx \frac{1}{N_s} \sum_{j=1}^{N_s} \ln(\theta_s^j) \ln(\bar{s}_i(\boldsymbol{x}^j)) - \text{E}[\ln(\theta_s)]\text{E}[\ln(s_i)] \qquad (5.21)$$

where $[\theta_s^j, \boldsymbol{x}^j]$ correspond to samples from $p(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, N_s is the total number of samples used, and $\ln(\bar{s}_i(\boldsymbol{x}))$ and $\sigma_i^s(\boldsymbol{x})$ are approximated through the kriging meta-model for each one of these samples. Utilising vectorised manipulations for the metamodel predictions both these quantities can be calculated with very small computational effort, meaning that the MCS-based estimation of Equations 5.19 to 5.21 can be performed very efficiently. Further numerical details (computational times) are provided in the illustrative example.

For reducing the relative importance of the MCS estimation error within the multi-objective problem of Equation 5.4, common random numbers (CRN) are utilised (Spall, 2003). This is facilitated by getting first samples in the standard Gaussian space (which is independent of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$) and then transforming them to the desired samples for $\boldsymbol{\theta}$. Approach is equivalent to transforming integrals in Equations 5.16 to 5.18 to standard Gaussian space and for record-based model is easily performed, as these models typically entail (Rezaeian and Der Kiureghian, 2010) a Gaussian predictive model for $\boldsymbol{v}$. The CRN is implemented by using the same sample set for the standard Gaussian samples across the entire optimisation as also shown earlier in Figure 5.1. Adoption of CRN creates a consistent estimation error in the MCS application for the different examined $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ values and therefore improves the optimisation accuracy for identifying the correct Pareto front (Spall, 2003). In other words, it allows use of smaller value for N_s because it reduces the relative importance of the MCS estimation error within the optimisation.

5.4.2   Multi-objective optimisation

Calculation of statistics given by Equations 5.14 and 5.15, utilising MCS estimates of Equations 5.19 to 5.21, facilitates an efficient approximation for performance objective F_{p1} given by Equation 5.8. If the lognormal assumption for the distribution

of Y_i^g is not used, then F_{p1} can be estimated through the approach discussed in Appendix B, leveraging again MCS principles. Note that calculation of objective F_{p2} is computationally trivial. Therefore, through the introduction of the metamodel an efficient estimation of both objectives involved in the optimisation described by Equation 5.4 can be established. It should be noted that in Chapter 4 the explicit incorporation in the optimisation of the metamodel error was also considered, established through appropriate modification of the equivalent objective function F_{p1}. It was shown, though, that this reduces computational efficiency and is not necessary if the underlying surrogate model has high accuracy. As calculation of objective F_{p1} requires N_s evaluations of the metamodel in a MCS setting (only one evaluation was needed in the framework of Chapter 4) the inclusion of the metamodel error is not advocated here as the computational burden is expected to be higher. Additionally, the benefits of explicitly including this error in the current formulation are expected to be smaller because objective F_{p1} represents a statistical quantity over $\boldsymbol{\theta}$. In evaluating the necessary statistic, the response, and therefore associated metamodel errors over different $\boldsymbol{\theta}$ values are averaged. Potential large errors that may exist for specific $\boldsymbol{\theta}$ values end up averaged with smaller errors from other $\boldsymbol{\theta}$'s and therefore do not impact F_{p1} estimates as much as they would if these $\boldsymbol{\theta}$ values were the only contribution to F_{p1} as was the case in Chapter 4.

For solving the resultant multi-objective optimisation, a variety of numerical approaches can be utilised (Marler and Arora, 2004). In Chapter 4, two such approaches were examined, one gradient-free and one gradient-based, and preference was ultimately given to the former since it can support higher robustness and computational efficiency for identifying the entire Pareto front. This recommendation is adopted here and a gradient-free, random search approach is implemented as follows. A large number of n_{bc} candidate solutions for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are generated that are close to $\boldsymbol{\mu}_r(\boldsymbol{z})$ and $\boldsymbol{\Sigma}_r$. This is established by creating uniform random samples centered around $\boldsymbol{\mu}_r(\boldsymbol{z})$ and $\boldsymbol{\Sigma}_r$. The range for these samples is chosen so that the

value of F_{p2} does not become excessively large, because the latter indicates significant departure of the modified model from the initial one, which might produce ground motions with unrealistic characteristics. Since estimation of F_{p2} is simple, a pre-screening can be implemented, rejecting any candidate solution with (large) value of F_{p2} over some desired threshold. Value close to 10 is appropriate for the latter threshold. This choice corresponds to modification of the standard deviation up to 50% and modification of $\boldsymbol{\mu}$ within a hypersphere 5 standard deviations away from $\boldsymbol{\mu}_r(\boldsymbol{z})$. Once all n_{bc} candidate solutions are obtained, objective functions F_{p1} and F_{p2} are calculated for each of them. Estimation of objective F_{p1} in this case leverages the computational efficiency of the metamodel in performing vectorised predictions: the calculations are simultaneously performed for all n_{bc} candidate solutions, or using subsets with a lower number of members depending on the available computational resources (memory can be a problem for vectorised operations depending on n and n_{bc}). The dominant solutions representing the Pareto front can be then readily identified by comparing the values for the two objectives over all candidate solutions. The only challenge is that the value of n_{bc} needs to be large in order to obtain an adequate representation of the Pareto front. Proper selection for n_{bc} is examined in the illustrative example. Once the Pareto front has been identified, a solution (among the Pareto set) can be adopted using any desired criterion. This will be further discussed in the illustrative implementation.

5.5   Illustrative implementation

The illustrative implementation considers the stochastic ground motion model developed by Rezaeian and Der Kiureghian (2010) and presented in Appendix A with model parameter vector including the arias intensity I_a, the significant duration D_{5-95}, the time corresponding to 50% of the intensity t_{mid} and the associated spectral frequency ω_{mid}, the rate of change for that frequency ω', and the damping ratio ζ_f for the excitation spectrum. The arias intensity simply scales the ground motion, so it corresponds to parameter θ_s in representation $Y_i^g(\boldsymbol{\theta}, \boldsymbol{w}) = \sqrt{\theta_s} s_i(\boldsymbol{x}, \boldsymbol{w})$,

with the five remaining parameters corresponding to $\boldsymbol{x}$. For the target hazard, GMPEs used in the Western US are selected here (Abrahamson and Silva, 2008; Boore and Atkinson, 2008; Campbell and Bozorgnia, 2008; Chiou and Youngs, 2008), whereas the suggestions in Kaklamanos et al. (2011) are adopted to estimate unknown inputs for some of the GMPEs. As target, IM predictions of the logarithmic mean and variance from individual GMPEs as well as the average of their predictions is used. Note that the latter provides single target hazard for each structural period examined, simply that hazard is obtained by averaging information from multiple GMPEs. All computations are performed in a quad-core 3.0 GHz Xeon processor with 16 Gb of RAM and all computational times reported herein are for this processor. Fault and site characteristics are taken, respectively, as strike-slip fault and shear wave velocity in upper 30 m of soil $V_{s,30}$=800 m/s. For moment magnitude M and rupture distance R different values will be examined. As IM for the seismic hazard description, the peak pseudo-acceleration PSA (S_a) for an SDoF system with 5% damping ratio is utilised. Different ranges of structural periods will be examined for S_a and, unless otherwise specified, the objective function F_{p1} is estimated as the weighted average of the entropies for each scalar IM as in Equation 5.7, with the weights chosen as γ_i=1, so that no specific structural period is prioritised.

5.5.1   Details for metamodel development

For the characterisation of domain X and the selection of the support points for the metamodel development, the same approach as in Chapter 4 is adopted, the only difference being that a larger domain is considered here with a relative increase of 70% compared to that considered in Chapter 4. The reason for the latter is to support higher accuracy in the MCS implementation: in Chapter 4, evaluation only at the new (modified) predictive relationships was required so proximity to the initial predictive relationships needed to account only for that modification, whereas here the MCS estimation will need to utilise the metamodel predictions

for parameters even further away from the modified (and therefore from the initial) mean predictive relationships.

Logarithmic mean and variance for S_a for different periods, the ones used by the aforementioned GMPEs, is adopted as the response output for the metamodel development whereas the white-noise samples are chosen as $n_w=100$. Three different accuracy criteria with associated coefficient of determination 0.92, 0.94 and 0.96 are selected for the adaptive metamodel development as presented in Chapter 4. This leads to metamodels with 1500, 3000 and 4500 number of support points, respectively. Note that the accuracy of the established metamodels is much higher for prediction of the logarithmic mean rather than the logarithmic variance. For example for the metamodel with 4500 support points the average coefficient of determination is 0.99 for the logarithmic mean and 0.94 for the logarithmic variance. This trend agrees with past studies (Gidaris et al., 2015) that have examined the use of metamodels in approximating seismic hazard when stochastic ground motion models are utilised for the description of the latter. This should be also viewed as a positive attribute of the metamodel approximation for the purposes of this study: the logarithmic mean has a higher importance towards the hazard description for the typical earthquake engineering applications that have logarithmic variance smaller than 1 (the scaling of difference of logarithmic means by $1/\sigma_i^2(\boldsymbol{z}) > 1$ in Equation 5.8 when estimating discrepancy form target also reveals this), which means that a higher degree of confidence exists for properly approximating this hazard than the average coefficient of determination (averaged over both logarithmic mean and variance) indicates, because the accuracy for the more important component (logarithmic mean) is higher.

Estimation of metamodel response for 10000 samples requires 3.6, 7.5, and 10.7 s for the metamodels with 1500, 3000, and 4500 support points, respectively. Note that adoption of larger value of samples prohibits efficient vectorisation of operations for the $n=4500$ points due to memory restrictions. Calculation of objective function F_{p1} for $N_s=100$ MCS samples requires 0.005, 0.10, and 0.16 s for the metamodels with 1500, 3000, and 4500 support points, respectively. It should be

stressed that, as advocated earlier, evaluation of the metamodel across multiple candidate solutions is simultaneously performed to better leverage the computational efficiency allotted by the vectorised metamodel evaluation. For example, for N_s=100, 100 different candidate solutions are examined; this leads to a total of 100·100=10000 samples for the metamodel evaluation, which avoids memory problems even for the metamodel with 4500 support points.

5.5.2 Validation of lognormal distribution assumption
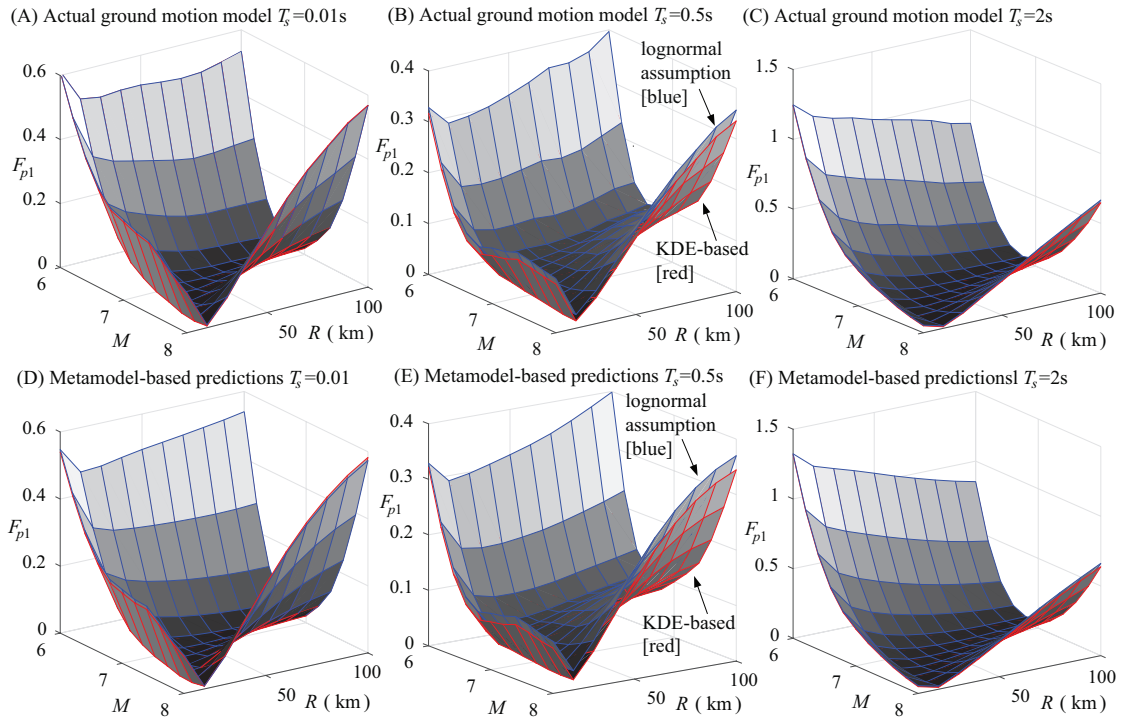
FIGURE 5.2: Comparison for F_{p1} evaluated either through lognormal assumption for Y_i^g or KDE approximation utilising samples for actual ground motion model (top row) and metamodel predictions (bottom row). Columns correspond to three different Y_i, representing, respectively, PSA for 0.01, 0.5 or 2 s.

First the validation of the lognormal assumption for the response output Y_i^g distribution is examined. This is performed separately for the response output from the exact stochastic ground motion model as well as for the metamodel approximation. Distribution $p_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is approximated though KDE (Equation B.1) as discussed in Appendix B using n_s=1000 samples of Y_i^g. Comparison is primarily expressed though objective F_{p1} because this is the critical quantity utilised in

the proposed framework. The target hazard corresponds to the average of the aforementioned four GMPEs. The predictive model proposed in Rezaeian and Der Kiureghian (2010), referenced herein as unmodified model, is used for defining $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, but comparison extends over a wide range of seismicity scenarios, M in range [6 8] and R in range [10 100] km. The latter guarantees that validation extends over the wider range that the optimisation is going to examine with respect to the predictive relationships. Figure 5.2 shows F_{p1} evaluated either through the lognormal assumption or through the KDE approximation for three periods T_s=[0.01 0.5 2] s (columns of the figure). In this instance F_{p1} is evaluated for each period separately to assess impact over specific structural characteristics, rather than averaged over the different IMs. Top row corresponds to comparison with respect to the actual stochastic ground motion model and bottom row to comparison with respect to the metamodel approximation. The comparisons in this figure show exceptionally close agreement for objective F_{p1} between the lognormal assumption and the KDE approximation (compare the two curves in each subplot) over the entire seismicity range and for all structural periods. This agreement holds for both the actual model and the metamodel. The results in the figure allow for some additional observations. First there is a very good agreement between the metamodel and the actual model (compare the results across the rows). This provides a preliminary validation, examined further in the following sections, of the proposed, metamodel-based framework. Secondarily, for certain seismicity ranges, the unmodified model has large discrepancy (large F_{p1} values) from the target hazard. This validates the claim that motivated this study, that existing approaches for selecting predictive relationships do not necessarily provide a close match to the desired hazard for some structures or seismicity scenarios. For assessing what constitutes large value of F_{p1} based on Equation 5.8, values close to 0.005 should be considered as small, values close to 0.02 as moderate and values over 0.05 as large.

The compatibility with respect to the lognormal assumption is further examined in Figure 5.3, which shows the CDF for Y_i^g for a particular seismicity scenario,

corresponding to M=6 and R=20 km. Similarly to Figure 5.2 each column corresponds to a different period T_s. The results in Figure 5.3 show, again, a very good match not just between the CDFs using the KDE approximation or the lognormal assumption but also between the CDFs using the actual model and the metamodel. Overall, the comparisons in Figures 5.2 and 5.3 justify the use of a lognormal assumption for the distribution of Y_i^g. Additionally, they provide a first validation for the accuracy of the metamodel approximation, when compared with the actual ground motion model, with respect to both the entropy (Figure 5.2) but also the exact distribution of Y_i^g (Figure 5.3).
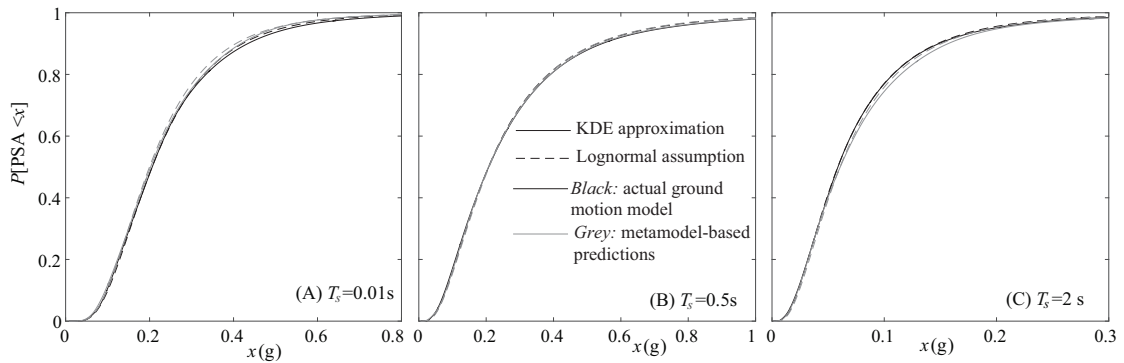
FIGURE 5.3: Comparison for the CDF Y_i^g = PSA ($P[\text{PSA} < x]$) based on lognormal assumption or KDE approximation utilising samples for actual ground motion model and metamodel predictions. Cases correspond to PSA for (A) 0.01, (B) 0.5 or (C) 2 s.

5.5.3 Optimisation details and metamodel accuracy

This section examines details of the numerical solution of the optimisation problem. For the MCS of Equations 5.19 to 5.21, Latin hypercube sampling (LHS) was adopted, and for the sample size N_s different values in range [20 150] were examined. The selection of the exact value of N_s is a compromise between numerical efficiency (it proportionally impacts the computational burden) and robustness of the multi-objective optimisation (identification of correct Pareto front). After sensitivity analysis, a value of N_s=70 was chosen. The coefficient of variation (CoV) for F_{p1} using MCS estimates of Equations 5.19 to 5.21 with N_s=70 was found to be in range 3% to 4% which should be deemed sufficient considering the fact that

CRN are further utilised (Spall, 2003) to reduce the importance of the associated estimation error in the comparisons established across the optimisation.

The discussion moves next to the impact of the number of points used in the random search n_{bc}. The considerations are similar to the selection of N_s (efficiency versus robustness) although in this case the selection is not as straightforward because there are no simple statistics like the CoV to compare. The sensitivity analysis is performed instead by solving the optimisation problem for different values of n_{bc}. The target used in this section corresponds to structural periods T_s=[0.4 0.5 0.75 1.0 1.5 2.0] s and hazard described by the average of the considered GMPEs. Three different seismicity scenarios M=6-R=20km, M=7-R=30km, and M=8-R=50km are considered. The first scenario represents a case that the unmodified model provides a poor match to the target hazard, the second one achieves a good match, and the last case lies in between the other two. Figure 5.4 presents the Pareto fronts identified through random search for different number of points n_{bc}. The metamodel with 4500 support points is used in this case. Only a few representative solutions and not the entire front are shown for clarity. Results are reported in this figure and in the remaining of the manuscript with respect to $\sqrt{F_{p1}}$ and $\sqrt{F_{p2}}$ to facilitate an easier comparison (differences of extreme values easier to discern). Figure 5.4 shows that the differences between the identified fronts for different n_{bc} values are only minor, and they occur primarily for small $\sqrt{F_{p1}}$ - large $\sqrt{F_{p2}}$ combinations. Note that the objective function $\sqrt{F_{p1}}$ values for the M=7-R=30km scenario are very low (due to an already good match of the unmodified model to the target hazard); therefore, any identified differences between the Pareto fronts for different number of points n_{bc} are of smaller relevance. This comparison shows that a value of n_{bc} around 100,000 to 250,000 should be considered as sufficient for identifying an adequate representation of the Pareto front. An n_{bc} value equal to 150,000 is adopted for the random search results presented in the remaining of the manuscript.

The impact of the metamodel accuracy is examined next. This is established by considering additionally the results obtained by using the exact stochastic ground
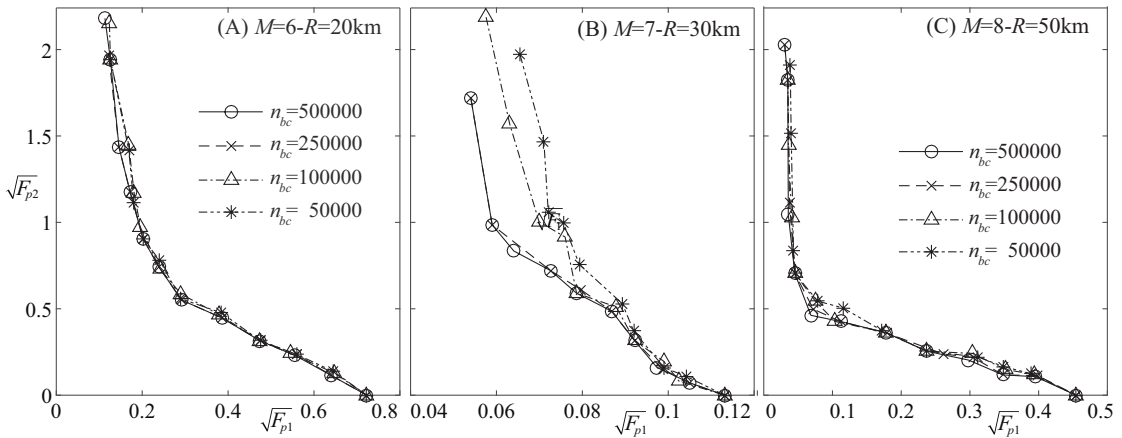
FIGURE 5.4: Pareto fronts identified by random search for different number of candidate solutions n_{bc} for different seismicity scenarios (A) M=6-R=20km, (B) M=7-R=30km, and (C) M=8-R=50km.
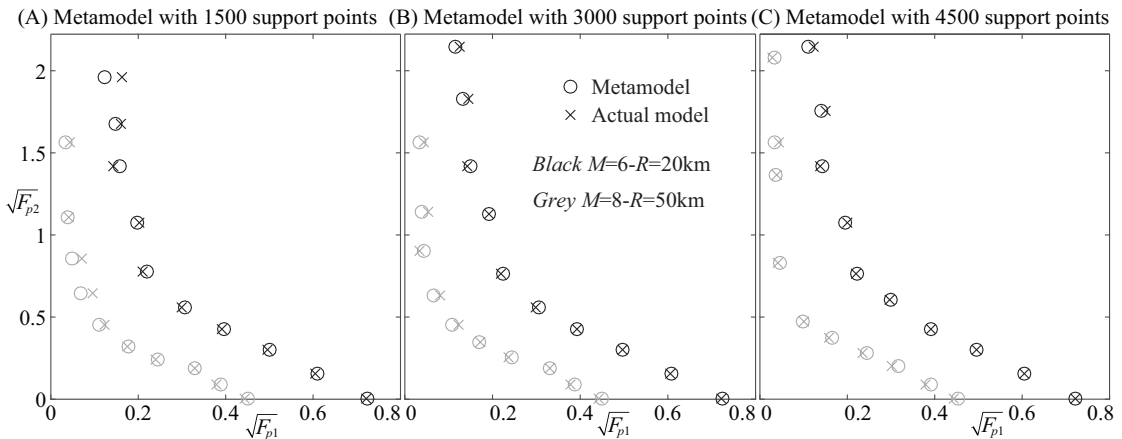
FIGURE 5.5: Pareto fronts identified using metamodels with (A) 1500, (B) 3000, or (C) 4500 support points and comparison to predictions by exact stochastic ground motion model. Two seismicity scenarios examined M=6-R=20km [black] and M=8-R=50km [gray].

motion model, which represents the measure for evaluating the actual hazard compatibility of the identified ground motion model. Comparison is performed across different Pareto optimal solutions. The Pareto fronts identified by using the metamodels with the three different number of support points are presented in Figure 5.5 for seismicity scenario M=6-R=20km and M=8-R=50km. Note that for seismicity scenario M=7-R=30km which was also presented in Figure 5.4, the results are of limited interest since the unmodified ground motion model provides a good compatibility to the target hazard. This is the reason that this seismicity scenario is not presented here. In all cases, the random search is implemented

with the same candidate solutions, to facilitate a consistency in the corresponding comparisons. Then, Figure 5.6 presents spectral plot comparisons for the solution (among the Pareto set identified in each case) corresponding to the minimum of F_{p1} for the seismicity scenario M=6-R=20km. Similar trends hold for the M=8-R=50km seismicity scenario. In all plots the predictions using the metamodel and the actual ground motion model are reported. Figure 5.6 offers comparisons in context of both average response (top row) using curves corresponding to median and 14th to 86th percentiles (denoted as median $\pm\sigma_{log}$ herein), and logarithmic standard deviation (bottom row). The former assesses the hazard compatibility with respect to different IM statistics and the latter explicitly with respect to the IM dispersion.
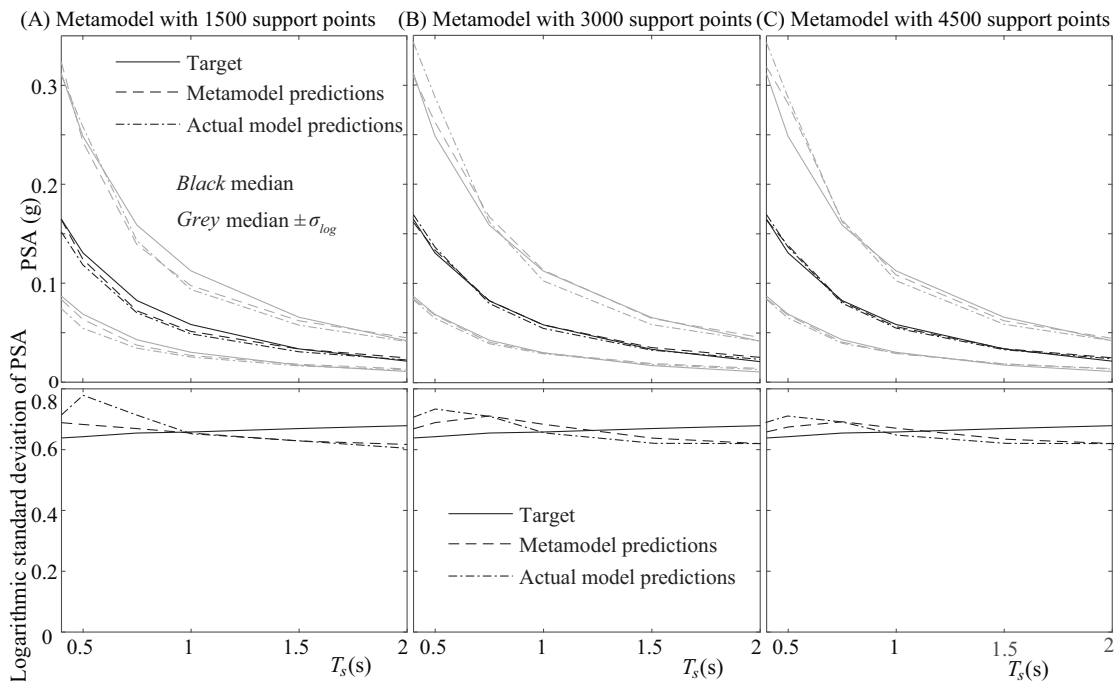
FIGURE 5.6: Spectral plots for the solutions corresponding to minimum of F_{p1} in the Pareto fronts identified in Figure 5.5 for the M=6-R=20km seismicity scenario. Top row shows curves corresponding to median and median $\pm\sigma_{log}$ for the response. Bottom row shows logarithmic standard deviation (σ_{log}).

The results show that good agreement is established between the metamodel predictions and the actual model predictions along the Pareto front for all three metamodel cases: 1500, 3000, and 4500 support points. This is evident in both the Pareto fronts (Figure 5.5) and the corresponding spectral plots (Figure 5.6),

the latter indicating a good match in terms of both the mean and variability of the hazard. For the lower accuracy metamodel utilising only 1500 support points greater differences exist, especially for smaller F_{p1} values, but the discrepancies even for it are overall quite small, significantly smaller than the ones reported in Chapter 4 where this metamodel was shown to lead to erroneous results, identifying suboptimal solutions when the metamodel error was not explicitly considered in the optimisation. This should be attributed to the fact, discussed also in Section 5.4.2, that objective F_{p1} represents a statistical quantity over $\boldsymbol{\theta}$, and averaging over different $\boldsymbol{\theta}$ values for evaluating F_{p1} reduces the potential influence of larger errors for specific $\boldsymbol{\theta}$ values. This discussion and the good agreement reported in Figures 5.5 and 5.6 also further validate the choice to avoid the explicit consideration of the metamodel prediction error in the problem formulation (see discussion in Section 5.4.2). This consideration would increase the computational burden while providing negligible benefits in terms of the quality of the identified Pareto fronts. Overall, all examined here metamodels provide adequate accuracy in the identification of the Pareto front, with no need to explicitly consider the metamodel prediction error, with preference towards the metamodel with 3000 or 4500 support points. The latter will be utilised in all remaining comparisons in this manuscript.

Some final remarks are warranted with respect to the overall computational cost. The primary computational burden of optimisation of Equation 5.4 stems from the $N_s \cdot n_{bc}$ evaluations of the metamodel required for the MCS of Equations 5.19 to 5.21 across the search for the Pareto front. For the scenario advocated earlier with N_s=70, n_{bc}=150,000 and use of metamodel with 4500 support points, the time is 180 minutes per seismicity scenario. If the metamodel with 3000 support points is used instead the required time reduces to 120 minutes. Compared with the study in Chapter 4 this represents an important increase of the computational cost: there is ultimately a N_s-fold increase of this cost for same n_{bc} value, stemming from the MCS step.

5.5.4   Implementation for different seismicity scenarios

With the computational details ironed out the discussion moves next to the hazard compatibility established by the proposed modification of the ground motion model. Figure 5.7 shows results for six seismicity scenarios targeting seismic hazard given by the average of the aforementioned GMPEs for two different ranges of T_s: T_s=[0.4 0.5 0.75 1.0 1.5 2.0] s and T_s=[0.4 0.5 0.75] s. These two different cases are referenced herein as long and short, respectively, period ranges. The proposed framework identifies in each case a Pareto front that clearly demonstrates the compromise between the two objectives, with different characteristics in each case, depending on how close the unmodified ground motion model was to the target hazard. Choosing a shorter period range for the target IMs facilitates an overall better match (smaller F_{p1} values); this is anticipated because objective F_{p1} imposes less strict requirements (fewer number of components to match) in this case.
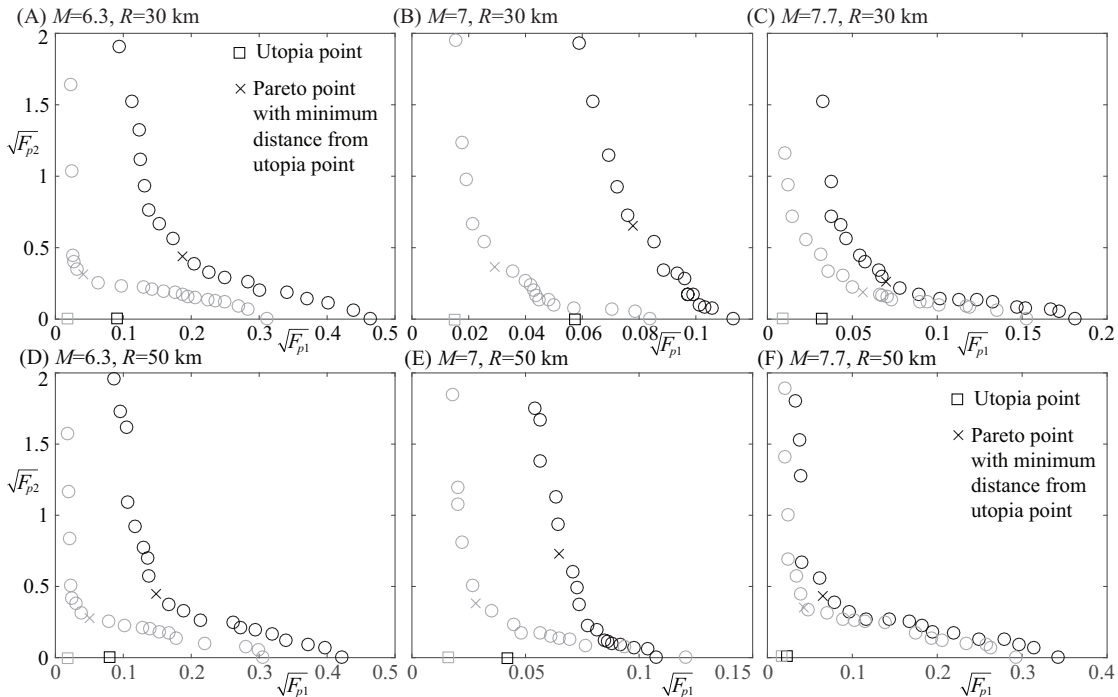
FIGURE 5.7: Pareto fronts for different seismicity scenarios considering match to long (black) or short (gray) period range IMs for defining the seismic hazard. In each plot, the Pareto point with minimum distance from utopia point is shown with x and the utopia point with a square.

To select a solution within the identified Pareto set, one of the most common approaches within the multi-objective optimisation literature (Keeney and Raiffa, 1993) is to choose the one that has the smallest normalised distance from the utopia point, defined as the point in the Pareto front that corresponds to the idealised (unachievable) minimum of the two objectives across the front. Following the guidelines in Chapter 4 for choosing the normalisation, the following distance metric is chosen

$$D_p(\boldsymbol{\theta}) = \sqrt{\sum_{i=1,2} \left(\frac{\sqrt{F_{pi}(\boldsymbol{\theta}|\boldsymbol{z})} - \min\sqrt{F_{pi}}}{\max\sqrt{F_{pi}} - \min\sqrt{F_{pi}}} \right)^2};$$
$$\max\sqrt{F_{pi}} = \max_{\{\boldsymbol{\theta}_p; i=1,\dots,n_p\}} (\sqrt{F_{pi}}), \ \min\sqrt{F_{pi}} = \min_{\{\boldsymbol{\theta}_p; i=1,\dots,n_p\}} (\sqrt{F_{pi}}) \quad (5.22)$$

The corresponding point is identified in all cases in Figure 5.7. Another option would have been to choose the solution that satisfies a pre-determined threshold for the match to the targeted IMs.

These selections are finally demonstrated for a wide range of seismicity scenarios, M in range [6 8] and R in range [10 100] km, in Figures 5.8 to 5.13. For each scenario, three different Pareto points are selected, the one with smallest distance $D_p(\boldsymbol{\theta})$ from the utopia point and the ones with objective $\sqrt{F_{p1}}$ smaller than 0.15 or 0.075. These three cases are denoted, U_t, C_l, and C_s, respectively. The thresholds for C_l and C_s modifications were chosen so that to reflect medium and small, respectively, incompatibility to the target hazard. In addition, results for the unmodified model are presented, denoted U_n. Figure 5.8 shows plots for (i - first row) $\sqrt{F_{p1}}$ for U_t and U_n ($\sqrt{F_{p1}}$ is constrained for the other two cases) and for (ii – second and third rows) $\sqrt{F_{p2}}$ for U_t, C_l and C_s ($\sqrt{F_{p2}}$ is zero for U_n). To better demonstrate the differences, results are presented separately for U_t (second row) and for the pair C_l and C_s (third row) in the latter case. The three different columns in the figure correspond to three different implementation cases: target hazard given by the average of the aforementioned four GMPEs for both the (A) long and (B) short period ranges as well as (C) target hazard given only by GMPE

Boore and Atkinson (2008) for the long period range. These scenarios are denoted herein as SC_1, SC_2 and SC_3 respectively. Figures 5.9 to 5.11 show spectral plots for a selection of seismicity scenarios, defined by combinations of M [6.2, 6.8, 7.4, 8] and R [30, 60, 90] km, for $SC1$. For each of the 12 M-R combinations spectral curves corresponding to the target hazard, the unmodified model and the predictions by the three aforementioned model modifications are shown to facilitate comparisons. More specifically, Figure 5.9 shows curves corresponding to different statistics of the response (median and median $\pm\sigma_{log}$) for the unmodified model and the model corresponding to the Pareto point with smallest distance from the utopia point (U_t case). Figure 5.10 presents the same curves for the models corresponding to the Pareto points with average relative entropy thresholds C_l and C_s. Figure 5.11 presents same comparison directly in terms of logarithmic standard deviation (i.e., IM dispersion). In all these figures the curves corresponding to the target are also shown.

Finally, Figures 5.12 and 5.13 show the predictive relationships for all examined seismicity scenarios for the modified model corresponding to the Pareto point with smallest distance from the utopia point (U_t case) for SC_1. In particular, Figure 5.12 shows the mean model parameters $\boldsymbol{\theta}$ of the modified predictive model (corresponding to $\boldsymbol{\mu}$ for the aforementioned Pareto point). The unmodified ground motion model parameters (corresponding to $\boldsymbol{\mu}_r$) as well as another case that will be discussed in the next section are also included. Then, Figure 5.13 shows the ratio of standard deviation for the modified and unmodified model (comparison of square root of the diagonal points of $\boldsymbol{\Sigma}$ for the aforementioned Pareto point and $\boldsymbol{\Sigma}_r$). Note that some of the curves shown in these figures have non-smooth characteristics. As discussed in Chapter 4 this should be attributed to the fact that a discrete representation of the Pareto front is obtained and to the random search characteristics of the adopted optimisation algorithm for identifying the Pareto front.
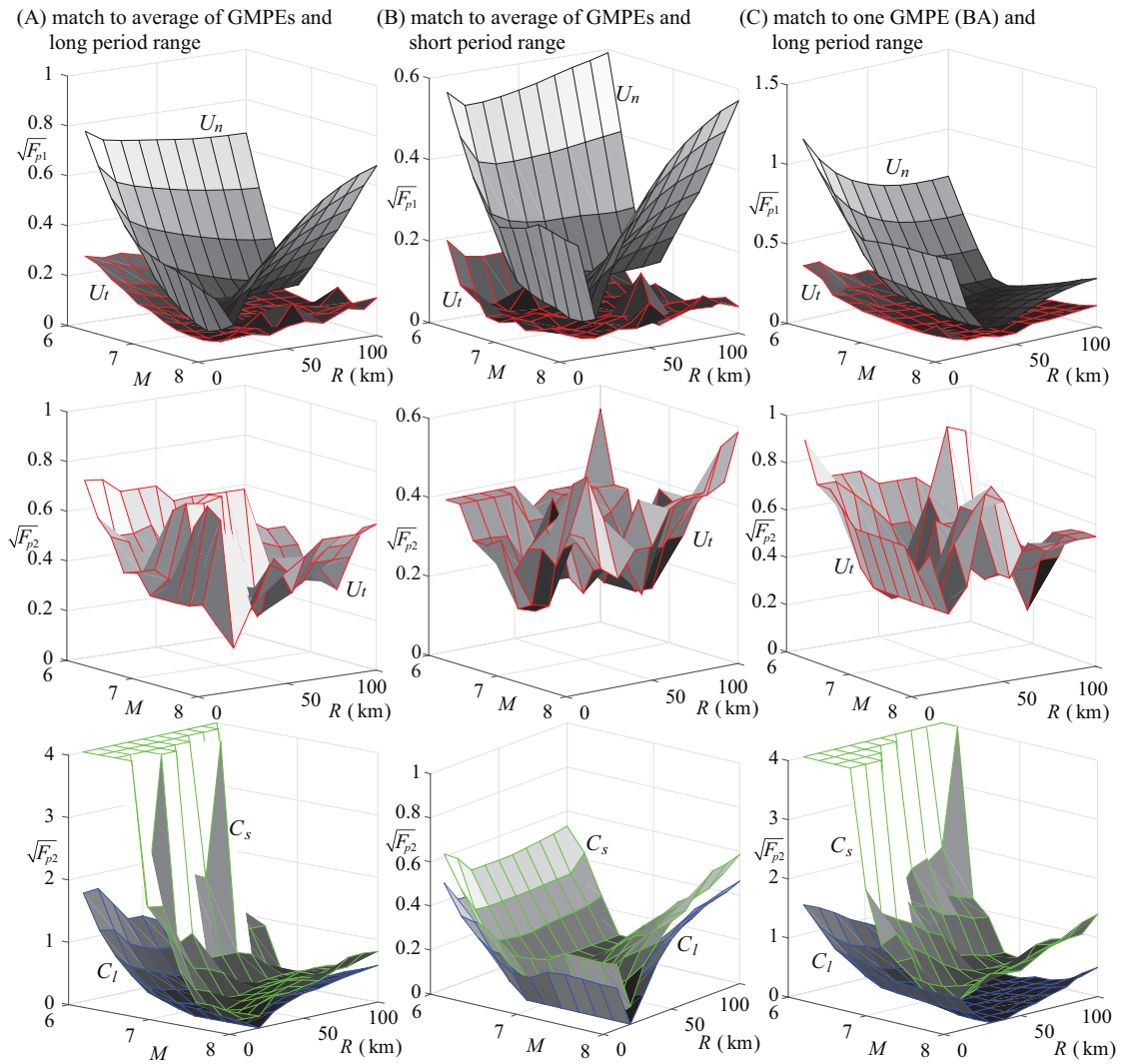
FIGURE 5.8: Results for $\sqrt{F_{p1}}$ and $\sqrt{F_{p2}}$ for unmodified ground motion model (U_n) and modified ground motion model corresponding to three different selection criteria: Minimum distance form utopia point (U_t) and value $\sqrt{F_{p1}}$ smaller than 0.15 (C_l) or 0.075 (C_s). Implementations in the different columns correspond to (A) long and (B) short period ranges for matching to the average considered GMPEs and (C) long period ranges for match to GMPE Boore and Atkinson (2008).

The baseline trends observed in the figures are similar to the ones in Chapter 4. These trends are enhanced in this study with additional considerations with respect to the variability associated with the seismic hazard. The unmodified model does not provide a good match to the target hazard for the entire seismicity range as observed in first row of Figure 5.8 and also in the spectral plots in Figures 5.9 and 5.11. This is particularly true for the logarithmic mean, i.e., median (Figure 5.9), with logarithmic variance, i.e., dispersion (Figure 5.11), showing better
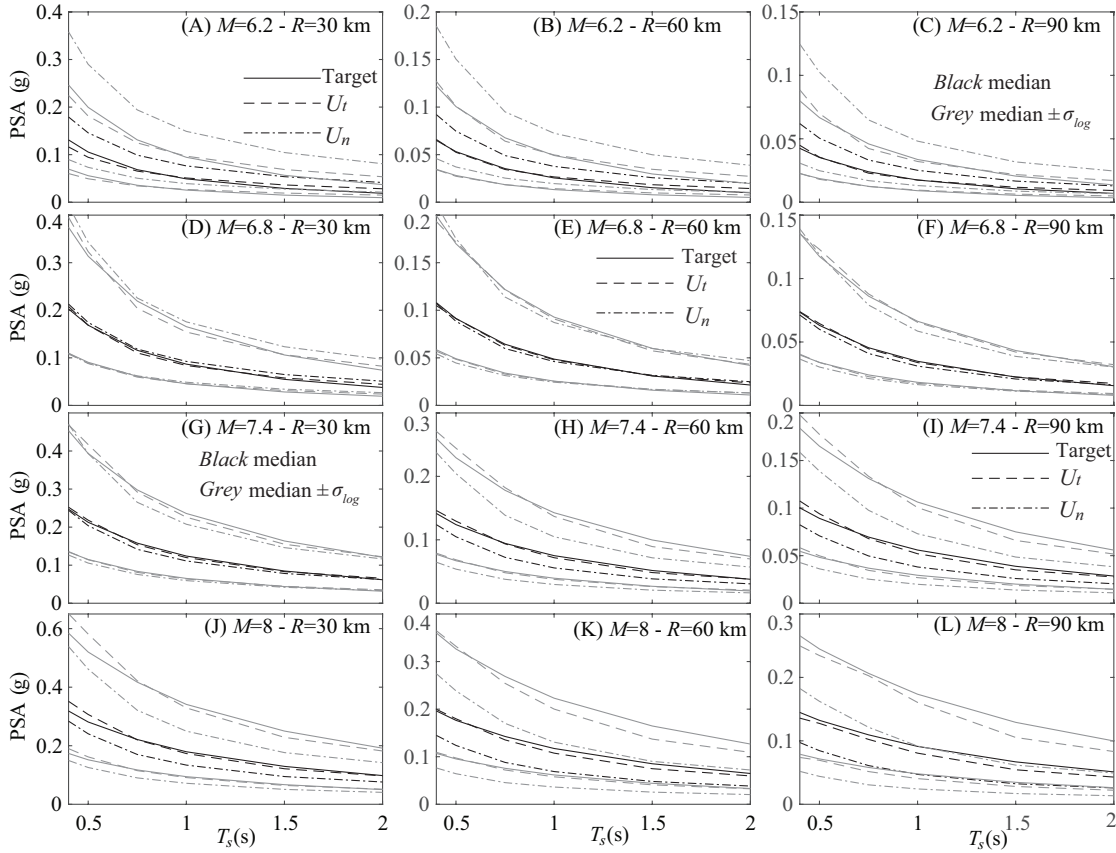
FIGURE 5.9: Spectral plots of average response for seismicity scenarios (different subplots) corresponding to combinations of M [6.2, 6.8, 7.4, 8] and R [30, 60, 90] km, for the target hazard (target), the unmodified ground motion model (U_n) and the modified ground motion model corresponding to the minimum distance from utopia point (U_t). Implementation scenario shown corresponds to matching to the average considered GMPEs and long period range. Curves corresponding to median and median $\pm\sigma_{log}$ shown.

compatibility to start with. The proposed modification (cases U_t, C_l, and C_s) significantly improve the match to the hazard (Figure 5.8), establishing a balance between F_{p1} and F_{p2}, with the characteristics of the balance depending on the criteria for selection of the final model among the Pareto optimal solutions. Similar to Chapter 4 when the unmodified model has larger discrepancies from the target hazard, then the modifications lead to larger values for F_{p2}, but still successfully identify models, independent of the implementation case, that provide an improved match to the IM target mean and variance. This is clearly observed for C_l and C_s cases; C_s imposes a smaller discrepancy between the modified model and the target hazard and the modification leads to identification of a model with bigger

differences from the original one (larger values for F_{p2}). The U_t modification iden-
tifies a model with moderate discrepancy from the unmodified one, corresponding
to values of $\sqrt{F_{p2}}$ in the range of 0.2 to 0.8, whereas the two other modification
approaches, C_l and C_s, identify models with greater variability across the differ-
ent seismicity scenarios. For scenarios in the range of M=7-7.5, the unmodified
model provides a good match to the target hazard and therefore modification of
it provides limited benefits. This is perhaps better captured by the C_l case, which
corresponds to low $\sqrt{F_{p2}}$ values for this seismicity range, significantly lower than
the U_t modification. This indicates that a satisfactory match to the target hazard,
i.e., a match satisfying the predefined compatibility threshold, is established with
no need to greatly modify the initial predictive models. Thus, selection of the
Pareto optimal model based on a targeted accuracy to the GMPEs, i.e., value for
F_{p1} below a certain threshold as in the C_s and C_l cases, provides a more rational
selection for the final model as it allows a more direct recognition of the seismicity
ranges where modification is not truly required. Selection of a small threshold (C_s
case), however, results in identification of a model with large discrepancies from
the unmodified model (large F_{p2} values). This model will typically be far away
from the U_t case and belong in a steep part of the Pareto front (check Figure 5.7
earlier), meaning that small improvements in F_{p1} come at a large increase of F_{p2}.
A multi-level selection criterion similar to the one proposed in Chapter 4 may
be therefore advocated: select the Pareto optimal solution that satisfies a certain
accuracy threshold for F_{p1} unless this solution leads to a greater F_{p2} value than
the Pareto optimal solution with minimum distance from the utopia point. If the
latter happens, then select the Pareto optimal solution with minimum distance
from the utopia point.

From the comparison of the different implementation cases shown in Figure 5.8,
it is evident that selection of a shorter period range (compare cases (B) and (A))
establishes an easier match to the target, as it results in smaller overall values
for F_{p1} and F_{p2}. The selection of a different target (compare cases (A) and (C)),
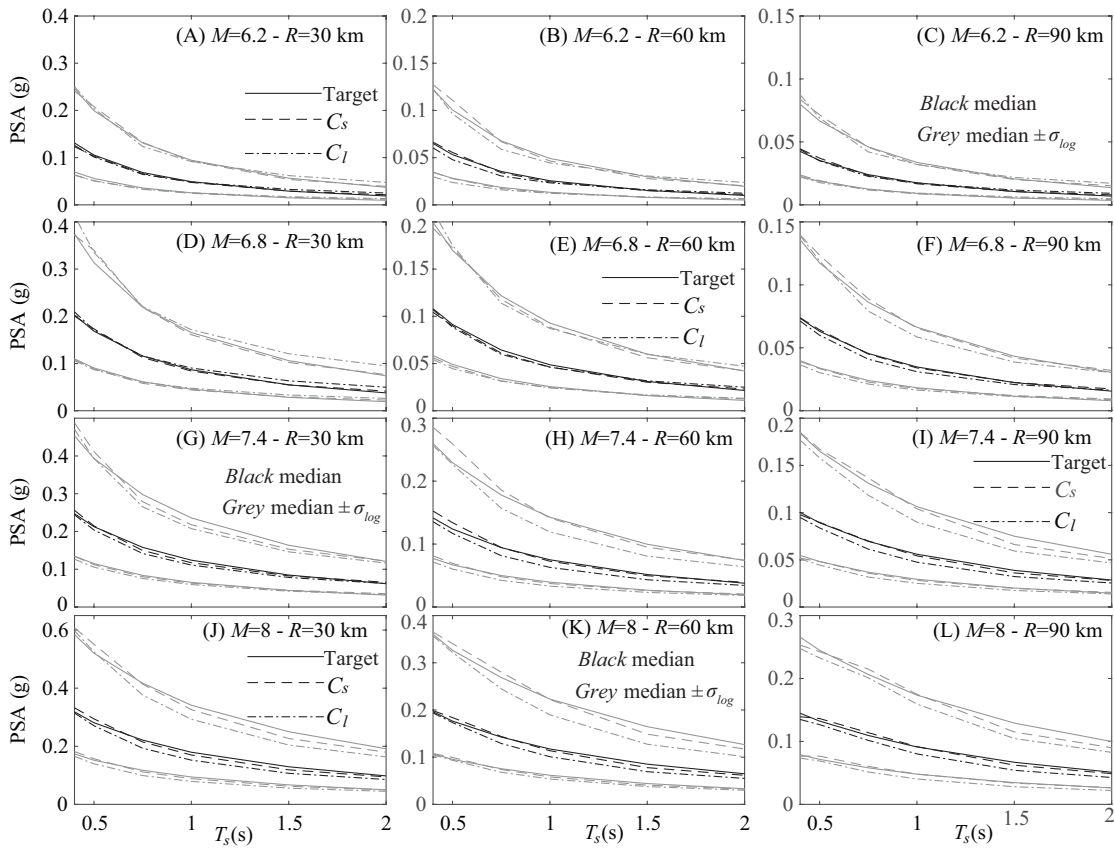
FIGURE 5.10: Spectral plots of average response for same seismicity scenarios as in Figure 5.9 (different subplots) for the target hazard (target), and the modified ground motion model corresponding to value $\sqrt{F_{p1}}$ smaller than 0.15 (C_l) or 0.075 (C_s). Implementation scenario shown corresponds to matching to the average considered GMPEs and long period range. Curves corresponding to median and median $\pm\sigma_{log}$ shown.

although not imposing any additional constraint in the optimisation implementation, leads to different results. Overall, the discrepancy between the unmodified model and the target hazard follows the same trends in all implementation cases. This demonstrates the versatility of the proposed framework as the modification facilitates an enhanced hazard compatibility for all IMs and hazard targets considered.

The spectral plots in Figures 5.9 to 5.11 provide the decomposition of the overall F_{p1} match to the different structural periods and the statistics of the IM distribution (median and dispersion). The comparison between the unmodified and modified cases shows that the match to both the target median and dispersion
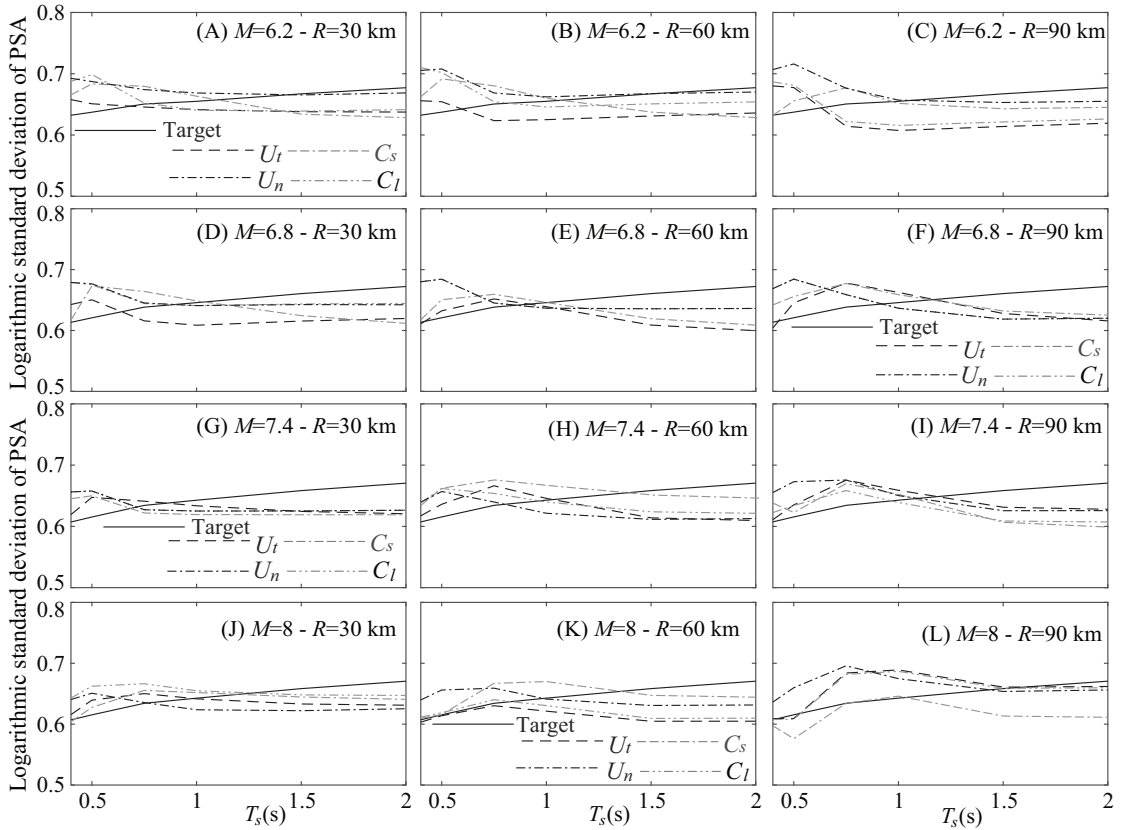
FIGURE 5.11: Spectral plots for logarithmic standard deviation of response for the same cases examined in Figures 5.9 and 5.10.

are separately addressed within the proposed optimisation. This is also supported by the trends in Figures 5.12 and 5.13; not only the mean (Figure 5.12) but also the standard deviation (Figure 5.13) of the predictive model is adjusted. Overall for the ground motion model examined here, the match to the target dispersion offered by the unmodified model, and subsequently by the modified one, is quite good (compare relative discrepancies for U_n in Figures 5.8 and 5.11). Although the proposed modifications also impact this dispersion (Figure 5.11), the impact on the median (Figures 5.9 and 5.10) is more substantial. This should be attributed to ability to influence value of objective F_{p1} more by adjustments in $\ln(\bar{Y}_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma}))$ rather than $\sigma_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (formulation of Equation 5.8 and scaling of differences from the target for the former by $\sigma_i^2(\boldsymbol{z})$ also reveals that), something automatically leveraged by the optimisation. With respect to the different modifications, the C_s always provides a better match compared with C_l one, although the spectral curves of the two modifications are very close to each other especially for lower magnitude

ranges M=6.2-6.8. The U_t modification also provides a good match to the target, frequently very close to the C_l and C_s cases, depending on the original match of the unmodified model (i.e., how much the modification is truly needed). This observation further supports the multi-level selection criterion discussed earlier.

The results in Figures 5.12 and 5.13 show that the model modification leads to similar characteristics as observed for the unmodified model. This guarantees that the proposed ground motion model modification does not deviate significantly from observed regional trends and was achieved by incorporating this deviation as an objective in the problem formulation (objective F_{p2}). Parameters I_a, ζ_f, and ω' show bigger variability compared with their initial mean values (Figure 5.12). This should be attributed to a greater sensitivity with respect to them of the resultant ground motions and agrees with the trends reported in Chapter 4. With respect to the adjustment of the variability of the predictive model, the ratio of standard deviations remains close to 1 (Figure 5.13) with values in the range of 0.8 to 1.05, indicating small (but not negligible) overall adjustment.

The physical ground motion model parameters $\boldsymbol{\theta}$ corresponding to the mean predictive relationships of the unmodified U_n (i.e., $\boldsymbol{\mu}_r$) and modified ground motion models U_t and C_s (i.e., $\boldsymbol{\mu}$) for the 12 seismicity scenarios of Figure 5.9 are presented in Table C.3 in Appendix C. Table C.4 shows the coefficient vector $\boldsymbol{\beta}$ that can be used to derive the covariance matrix of the modified predictive relationships $\boldsymbol{\Sigma}$ as $\boldsymbol{C} \cdot \boldsymbol{B} \cdot \boldsymbol{C}^T$, where $\boldsymbol{C}$ is the Cholesky decomposition of the original covariance matrix $\boldsymbol{\Sigma}_r$ and $\boldsymbol{B}$ is a diagonal matrix whose main diagonal comprises the elements of vector $\boldsymbol{\beta}$. The median and σ_{log} spectral acceleration estimates (PSA) are shown for the same ground motion model cases along with the target estimates (average of four GMPEs) in Tables C.5 and C.6, respectively.

The overall discussion shows the importance of the established framework: once the initial metamodel is developed, it can support the efficient identification of ground motion models that (i) match conditional hazard for any desired IMs and
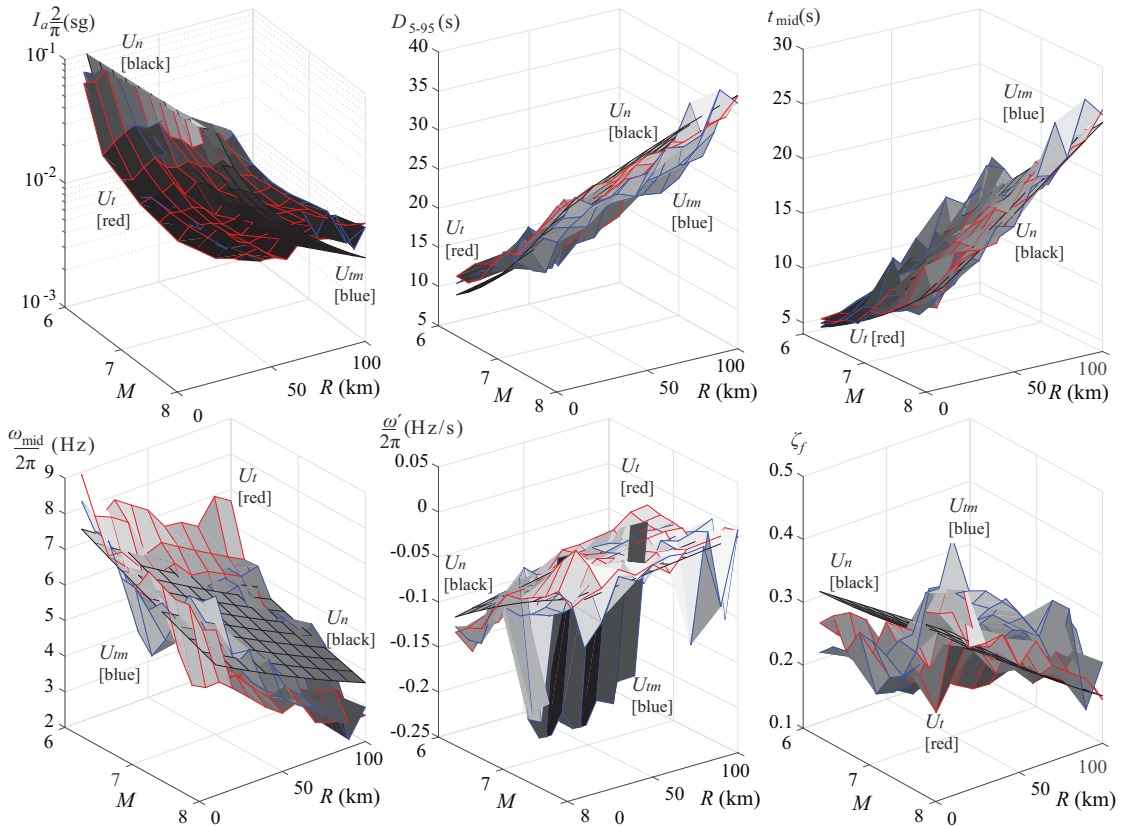
FIGURE 5.12: Mean for physical ground motion model parameters $\boldsymbol{\theta}$ corresponding to unmodified ground motion model (U_n), modified ground motion model with minimum distance from utopia point for matching the complete probabilistic hazard (U_t) or the hazard corresponding to mean predictive relationships (U_{tm}). Implementation scenario corresponds to matching to the average considered GMPEs and long period range.

chosen period range while (ii) maintaining a small deviation from the initial predictive models. This can be seamlessly repeated for any seismicity scenario. The final ground motion model modification can be chosen based on the criteria discussed earlier.

5.5.5 Comparison to modification of mean value characteristics only

As discussed in Section 5.5.3 the computational cost of the proposed modification to match the probabilistic seismic hazard is significantly higher than previous efforts to match only the IM predictions corresponding to the mean predictive
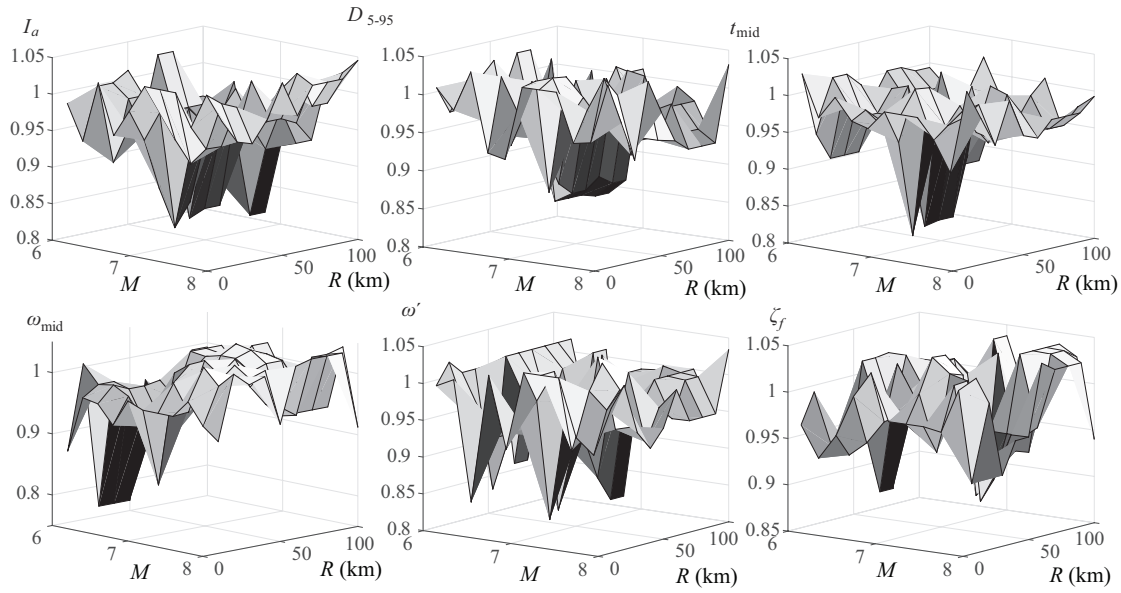
FIGURE 5.13: Ratio of standard deviations for the ground motion model parameters between modified and unmodified ground motion model (U_n and U_t cases in Figure 5.12).

relationships for the ground motion model presented in Chapter 4, i.e., completely ignoring the variability stemming from $\boldsymbol{\Sigma}$ in the predictive model for $\boldsymbol{\theta}$ (using $\boldsymbol{\Sigma}{=}0$). This increase in computational burden stems ultimately from the need to estimate the statistics of the response when the variability of the predictive model $p(\boldsymbol{\theta}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ is considered. It is therefore of interest to examine whether the computationally less demanding problem of matching only the IM predictions corresponding to the mean of the predictive relationships (approach presented in Chapter 4) can be adopted as a surrogate for the problem of interest here. This is equivalent to assuming $\boldsymbol{\Sigma}{=}0$ in the optimisation of Equation 5.4 and leads to modification of the mean only predictive relationships, while greatly reduces computational cost of the numerical optimisation as it entails no MCS step (because $\boldsymbol{\Sigma}{=}0$). This facilitates a N_s-fold reduction of the computational burden as discussed earlier. The resultant modification of the predictive relationships will be denoted $\boldsymbol{\mu}_m$ herein.

The quality of the solution obtained from this approximate problem may be then assessed by evaluating objectives F_{p1} and F_{p2} assuming distribution $p(\boldsymbol{\theta}|\boldsymbol{\mu}_m,\boldsymbol{\Sigma}_r)$, i.e., adopting the initial variability of the predictive models $\boldsymbol{\Sigma}_r$, or $p(\boldsymbol{\theta}|\boldsymbol{\mu}_m,0)$, i.e.,
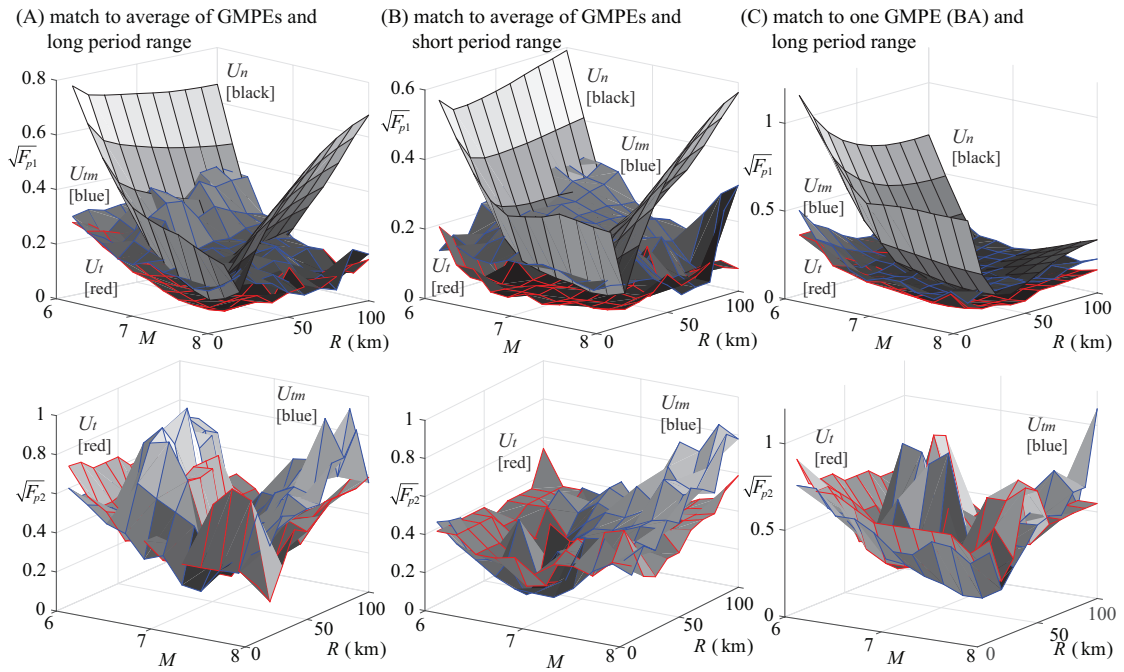
FIGURE 5.14: Results for $\sqrt{F_{p1}}$ and $\sqrt{F_{p2}}$ for unmodified ground motion model (U_n) and modified ground motion model with minimum distance form utopia point for matching the complete probabilistic hazard(U_t) or the hazard corresponding to mean predictive relationships (U_{tm}). For the latter estimation of objectives adopts variability of the initial (U_n) predictive model but with the corresponding updated predictive mean. Implementations in the different columns correspond to (A) long and (B) short period ranges for matching to the average considered GMPEs and (C) long period ranges for match to GMPE Boore and Atkinson (2008).

ignoring any variability in the predictive models and utilising only the variability stemming from the white-noise to calculate response statistics. The comparison can be performed with respect to the entire Pareto front for limited number of seismicity scenarios or with respect to a specific solution along the front over a wider range of scenarios. The latter comparison is reported here, with specific solution chosen as the point with minimum distance from the Utopia point. The solution available from Chapter 4 is directly utilised for $\boldsymbol{\mu}_m$. The case corresponding to $p(\boldsymbol{\theta}|\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_r)$ is denoted as U_{tm} and case corresponding to $p(\boldsymbol{\theta}|\boldsymbol{\mu}_m, 0)$ as U_{tmn}. Results are shown in Figures 5.14 and 5.15, following same guidelines as the study reported in Figures 5.8 and 5.9. Figure 5.14 compares U_n, U_t, and U_{tm} across both objectives over a range of seismicity scenarios (i.e., adds U_{tm} curve in the results reported in Figure 5.8), and Figure 5.15 presents spectral plots for the

target seismic hazard, U_{tm} and U_{tmn}. Latter figure should be compared directly to Figure 5.9 to evaluate the relative advantages of U_t (adding that curve in this plot is avoided to improve clarity of the presentation). Solution for the physical parameters corresponding to $\boldsymbol{\mu}_m$ has been also reported earlier in Figure 5.12. Note that this solution is same for both U_{tm} and U_{tmn}.
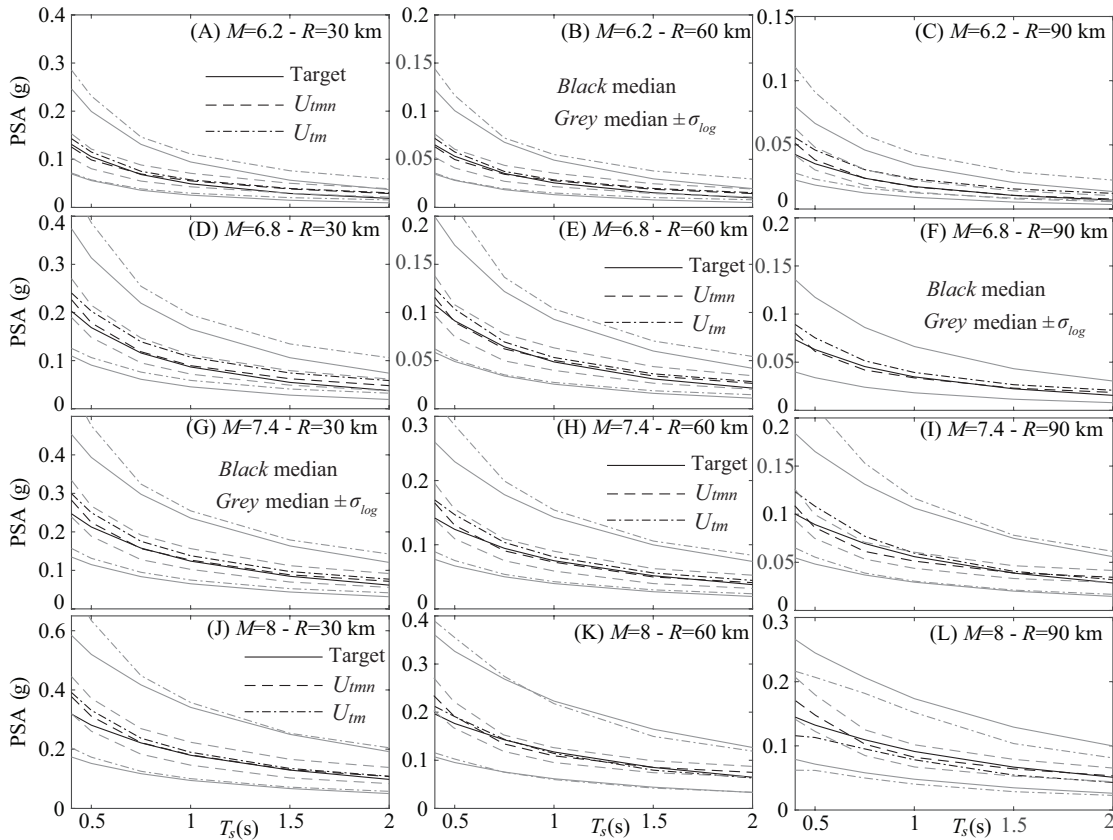
FIGURE 5.15: Spectral plots for seismicity scenarios (different subplots) corresponding to combinations of M [6.2, 6.8, 7.4, 8] and R [30, 60, 90] km, for the target hazard (target), and modified ground motion model with minimum distance from utopia point for matching the hazard corresponding to mean predictive relationships. For the latter the cases adopting variability of the initial predictive model (U_{tm}) or no variability (U_{tmn}) are shown. Implementation scenario shown corresponds to matching to the average considered GMPEs and long period range. Curves corresponding to median and median $\pm\sigma_{log}$ shown.

With respect to U_{tmn} first (Figure 5.15) the results show the importance of considering the variability of the predictive relationships. Reliance only on the variability stemming from the white-noise provides significantly lower variability for the seismic hazard than the variability prescribed by GMPEs (compare median and median $\pm\sigma_{log}$ curves for target and U_{tm} cases). This ultimately leads to large

values for F_{p1} for U_{tmn} which is the reason that results for it are not reported
in Figure 5.14. Comparison now between U_t and U_{tm} in Figure 5.14 shows that
the explicit optimisation for the probabilistic hazard provides for some scenarios
a noticeably better match (smaller F_{p1} values for U_t) for the same level of modi-
fication of the initial probability model (similar F_{p2} values for U_t and U_{tm}). U_{tm}
modifications even underperform the unmodified model U_n for some scenarios,
corresponding to cases for which U_n provides an adequate match to the target
hazard to start with. Same trends are observed in the spectral plots in Figure
5.15. Overall U_{tm} is shown to provide an improvement over U_n when the initial
match to the hazard is not adequate, though it might underperform U_t due to
its inability to explicitly accommodate the hazard variability. It should be also
pointed out that based on the results in Figure 5.12, even the mean value vectors
for the predictive relationships identified by different modification approaches (U_t
and U_{tm}) are different. A final interesting comparison can be established with re-
spect to the mean spectral response for U_{tm} and U_{tmn} (black curves in Figure 5.15).
Consideration of the variability in the predictive model after the optimisation (U_{tm}
case) impacts even the mean of the predicted response, not only the variability
of that response, and overall moves that response further away from the desired
target, which, recall, was the objective in the underlying modification for the mean
predictive relationships $\boldsymbol{\mu}_m$. This discussion shows that even though the surrogate
optimisation for only $\boldsymbol{\mu}_m$ may provide an improvement over the unmodified model
for scenarios of high initial discrepancy from the target probabilistic hazard, the
post-consideration of variability in the predictive models is problematic. Setting
initially $\boldsymbol{\Sigma}=0$ to identify $\boldsymbol{\mu}_m$ and then calculating the hazard for predictive model
($\boldsymbol{\mu}_m$, $\boldsymbol{\Sigma}_r$) provides ultimately a lower quality fit to the target hazard. Despite the
higher computational burden associated with it, the simultaneous modification of
the entire predictive model for $\boldsymbol{\theta}$ (both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$) is therefore advocated.

5.6   Conclusions

The modification of stochastic ground motion models to establish hazard compatibility for specific seismicity scenarios was discussed in this chapter. The hazard for each scenario was described with respect to some IM of interest and a probabilistic description was adopted for it, for example defined through mean and dispersion characteristics. The modification of the ground motion model was defined as an adjustment of the probabilistic predictive models/relationships that relate the parameters of the ground motion model to seismicity characteristics. Both the mean of the predictive model and the associated variance were adjusted. The proposed modification was defined as a bi-objective optimisation with dual objective of minimising the discrepancy between the hazard for a given structure/site and the predictions established through the stochastic ground motion model, while maintaining a small deviation from the original predictive relationships, assumed to facilitate similarity to observed regional trends. This setting extends the work of the previous Chapter 4 that examined modification of only the mean predictive relationships with goal to match the corresponding IM predictions, ignoring any variability in either of these two components. The relative entropy was adopted as metric to quantify the objectives considered in the multi-objective optimisation problem, whereas the same surrogate modeling framework as in Chapter 4 was utilised for an efficient optimisation. Emphasis was placed on the estimation of the various response statistics needed for the entropy calculation, and a MCS approach was advocated for it coupled with assumption for lognormal distribution of the response when considering the variability of the predictive models/relationships. The computational approach explicitly considered the fact that most ground motion models involve a separate parameter that impacts their scaling. This parameter was separately treated with respect to both the surrogate model development and the MCS. Different statistical assumption for the distribution of the ground motion model output was also examined for the evaluation of the entropy for the first objective.

In the illustrative example, the proposed framework was applied using a recently developed record-based stochastic ground motion model. It was shown that log-normal distribution assumption for calculating the first objective provides an adequate approximation for the application at hand, whereas the metamodel-aided optimisation can facilitate an accurate identification of the Pareto front even when lower accuracy metamodels are utilised, a feature that did not hold in the study presented in Chapter 4. The necessity to calculate response statistics through MCS increases, though, the computational burden of the approach. Application to wide range of seismicity scenarios and different approaches for determining the seismic hazard (different IMs or sources for the target values) were examined, illustrating the advantages for the proposed framework: it allows significant improvements to the target hazard match, especially for seismicity scenarios for which the un-modified model provides a poor initial fit, with minor only modifications to the original predictive model, something that can guarantee a good agreement with observed regional trends. With respect to selection of the final model across the identified Pareto front, same recommendation as in Chapter 4 is made: select the Pareto optimal solution that satisfies a certain accuracy threshold for match to the target hazard (F_{p1} constraint) unless this solution leads to a greater modification for the predictive model (F_{p2} value) than the Pareto optimal solution with minimum distance from the Utopia point. Finally, the approach presented in this chapter was compared with the approach of Chapter 4: modification of only the mean predictive relationships to match the corresponding hazard for these mean predictions. It was shown in this case that the latter approach may provide an adequate surrogate for seismicity scenarios with high initial discrepancy to the target hazard, though overall it is better to explicitly consider the impact of the variability in the predictive models (i.e., calculate response statistics) and simultaneously modify the entire predictive model (i.e., not only focus on the mean of the predictive relationships).

The main limitation of the approach is the significant computational burden for performing the multi-objective optimisation to identify the Pareto front, a burden

stemming from the MCS step. Considering the fact that the proposed modification needs to be repeated for each seismicity scenario of interest, further reduction of this burden, which will have to come from a more computationally efficient implementation of the surrogate model predictions, is an important extension of this work, which is not part of the scope of this thesis.

Finally, the validation of the proposed stochastic ground motion modification approaches presented in Chapters 4 and 5 is performed next in Chapter 6.

Chapter 6

Validation of stochastic ground motion model modification by comparison to seismic demand of recorded ground motions

6.1   Introduction

Chapters 4 and 5 presented a computationally efficient framework to modify existing stochastic ground motion models with a dual goal of (i) facilitating compatibility with the target conditional hazard described through any chosen IM while (ii) preserving desired trends and correlations in the physical characteristics of the resultant ground acceleration time-series. For a given seismicity scenario the framework identifies the modified predictive relationships of the stochastic ground

motion model that can sufficiently match the target conditional hazard while maintaining similarity to preexisting predictive relationships, so that observed regional physical characteristics of ground motions are retained. The conditional target hazard is described through the conditional mean and the dispersion of some target IM (Chapter 5) or simply through only the conditional mean (Chapter 4). The modification is posed as a multi-objective optimisation problem, with different criteria established for selecting the final predictive relationships.

This chapter extends this effort through a validation study by comparing the seismic demand of hazard-compatible recorded ground motions to the demands of stochastic ground motion models that are modified to match the same target hazard. Suites of recorded and stochastic ground motions, whose spectral acceleration statistics match the mean and variance of target spectra within a period range of interest, are utilised as input to perform response-history analysis of inelastic SDoF case-study systems. The resultant EDP distributions are then compared to perform the desired validation. Validation extends to different seismicity scenarios and different inelastic and hysteretic characteristics for the SDoF systems.

6.2 Stochastic ground motion model and proposed modification

Similar to the applications in Chapters 4 and 5, the stochastic ground motion model considered (and modified) is the one developed by Rezaeian and Der Kiureghian (2010), which combines a time-domain modulating envelope function with a frequency-spectrum with time varying spectral properties. The model parameter vector, denote as $\boldsymbol{\theta}$ herein, consists of: the parameters of the envelope function corresponding to the Arias intensity I_a, the significant duration D_{5-95}, and the time at the middle of the strong-shaking phase t_{mid}; and the parameters of the frequency-spectrum corresponding to the damping ratio ζ_f, the spectral

frequency ω_{mid} at t_{mid}, and the rate of change for that frequency ω' (linear varia-
tion is assumed for the spectral frequency). These model parameters are related
through predictive relationships to seismicity and local site parameters: the mo-
ment magnitude, M, the rupture distance, R, the fault type, F, and the shear
wave velocity in the upper 30 meters of soil, $V_{s,30}$. The vector of these four param-
eters is denoted as $\boldsymbol{z}$ herein. The predictive relationships developed by Rezaeian
and Der Kiureghian (2010) ultimately define a conditional probability distribution
that relates $\boldsymbol{\theta}$ to $\boldsymbol{z}$, denoted herein as $p(\boldsymbol{\theta}|\boldsymbol{\mu}_r(\boldsymbol{z}), \boldsymbol{\Sigma}_r)$, where $\boldsymbol{\mu}_r(\boldsymbol{z})$ are the mean
predictions and $\boldsymbol{\Sigma}_r$ represents the variability of these predictions. This ground mo-
tion model description ultimately provides a probabilistic prediction for any IM of
interest, with variability in the predictions stemming from both (i) the stochastic
characteristics of the ground motion model (i.e., fact that it entails a white-noise
sequence); and (ii) the probabilistic description of the predictive relationship be-
tween $\boldsymbol{z}$ and $\boldsymbol{\theta}$ (i.e., the fact that $\boldsymbol{\Sigma}_r$ exists). For spectral acceleration at a given
period T_i, which is the IM utilised in this chapter, the probabilistic description
through the ground motion model is denoted as $p_g(\ln(S_a(T_i))|\boldsymbol{\mu}_r(\boldsymbol{z}), \boldsymbol{\Sigma}_r)$ and, as
shown in Chapter 5, can be approximated very well as a lognormal distribution
utilising simply the median and dispersion (under the aforementioned two sources
of variability (i-ii)) of $S_a(T_i)$. A complete mathematical description of all these
statistics is available in Chapter 5.

The modification framework developed and discussed in Chapter 5 adjusts $\boldsymbol{\mu}_r(\boldsymbol{z})$
and $\boldsymbol{\Sigma}_r$ (replaces them with $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively) for each examined $\boldsymbol{z}$ so that
the conditional (to the seismicity scenario defined by $\boldsymbol{z}$) seismic hazard established
through the modified model, $p_g(\ln(S_a(T_i))|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, provides a closer match to the
desired target seismic hazard for the IM, $p_t(\ln(S_a(T_i))|\boldsymbol{z})$. In the context of this
study, the latter is determined through GMPE predictions for the mean and dis-
persion of $S_a(T_i)$ considering a range of periods T_i. This ultimately facilitates a
GMPE-based (or scenario-based) spectra compatibility of the modified stochastic

ground motion model. This modification is expressed as a multi-objective optimisation problem with two competing objectives. The first objective, F_1, is to minimise the discrepancy of the target seismic hazard to the hazard predicted through the ground motion model, i.e., to a comparison between $p_g(\ln(S_a(T_i))|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $p_t(\ln(S_a(T_i))|\boldsymbol{z})$. The second objective, F_2, is to establish the smallest deviation between the updated probability model $p(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and the initial predictive relationships $p(\boldsymbol{\theta}|\boldsymbol{\mu}_r(\boldsymbol{z}), \boldsymbol{\Sigma}_r)$, so that consistency of the physical characteristics of the resultant ground motion simulations with the regional trends observed in recorded ground motions is achieved. The relative entropy is utilised to quantify both these objectives, corresponding ultimately to the difference between probability distributions, and a computational framework relying on surrogate modeling is leveraged to efficiently solve the resultant multi-objective optimisation. A simplified implementation of this framework is presented in Chapters 4, where variability in the predictive relationships is completely ignored, i.e., enforces $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_r = 0$ (variability stemming from stochastic features of ground motion model still considered), and establishes compatibility with respect to the median IM predictions, rather than the complete hazard (median and dispersion of predictions). This simplified version yields significantly higher computational efficiency with the caveat, of course, that the dispersion of the predictions is not explicitly optimised. Objective F_1 is expressed in this case as the average squared relative error for $S_a(T_i)$ between the ground motion predictions and the GMPE-target across the considered periods, whereas objective F_2 as the weighted squared difference between $\boldsymbol{\mu}$ and $\boldsymbol{\mu}_r(\boldsymbol{z})$. The simplified implementation is references herein as IMC (IM compatibility) with the full one referenced as HC (Hazard compatibility). Scaling of objectives F_1 and F_2 by $1/2$ is utilised in this chapter for the IMC case compared to the study in Chapter 4 to facilitate a more direct comparison to the HC case, since in the relative entropy-based quantification (HC) the squared differences appearing in the IMC objectives shown in Chapter 5 are scaled by factor $1/2$.

The solution of this multi-objective optimisation problem for either case leads to a Pareto set of dominant solutions expressing a different compromise between the

two competing objectives. The representation of the Pareto set in the performance objective $[F_1, F_2]$ space, is termed as the Pareto front. For better comparative normalisation of the solutions the front is represented through the square root of the objectives $[\sqrt{F_1}, \sqrt{F_2}]$. Figure 6.1 shows representative Pareto fronts for all seismicity scenarios discussed later in this paper. The front ranges from the unmodified model, denoted U_n, corresponding to $F_2 = 0$ and higher discrepancy from the IM-target (larger F_1 values), to models that establish high compatibility to the IM-target (small F_1 values) at the expense of significant deviation of the model characteristics from the initial predictive relationships (large F_2 values). Further reduction of F_1 (ultimately achieving $F_1 = 0$) is achieved by deviation from the initial predictive relationships that might yield unrealistic characteristics for the resultant ground motions and this part of the front is not identified through the use of appropriate constraints in the multi-objective optimisation as discussed in Chapters 4 and 5. One can eventually select a model configuration from the identified Pareto set that yields the desired hazard compatibility (or strictly IM compatibility for IMC) based on objective F_1 without deviating significantly from regional ground motion characteristics based on objective F_2. Following recommendations in Chapters 4 and 5, three specific points are examined, also shown in Figure 6.1. The first one, denoted U_t, is the point with minimum distance from the Utopia point, corresponding to the minimum of the two objectives across the Pareto front (this performance is unachievable due to the conflicting nature of the objectives). U_t offers a balanced compromise between the competing objectives and, as shown also in Figure 6.1, improvement of one objective is typically established away from that point with greater sacrifices in the other objective (front has steep slope). The other two chosen points are defined as the ones that achieve a predetermined compatibility with respect to target hazard, i.e., a specific threshold value of objective F_1. The first of these points, denoted C_s, corresponds to high compatibility (small F_1 threshold), whereas the second point, denoted C_l, is defined following the multi-level criterion proposed in Chapters 4 and 5: select the point that provides a moderate compatibility (larger threshold for F_1 compared to C_s) unless that point provides a larger compatibility to the target than U_t;

for those instances update $C_l = U_t$. This update avoids defining a C_l point that belongs to the steep part of the Pareto front with respect to objective F_2. Also for the C_s case, if no point in the Pareto front satisfies the desired threshold, the extreme of the front with respect to F_1 is used instead. The thresholds for $\sqrt{F_1}$ defining C_s and C_l points are taken as 0.014 and 0.05 for the IMC and HC cases, respectively, for C_s and 0.07 and 0.15 for the IMC and HC cases, respectively, for C_l. These thresholds are chosen to represent high and moderate compatibility for C_s and C_l respectively based on the features of objective F_1 for each of the cases examined.
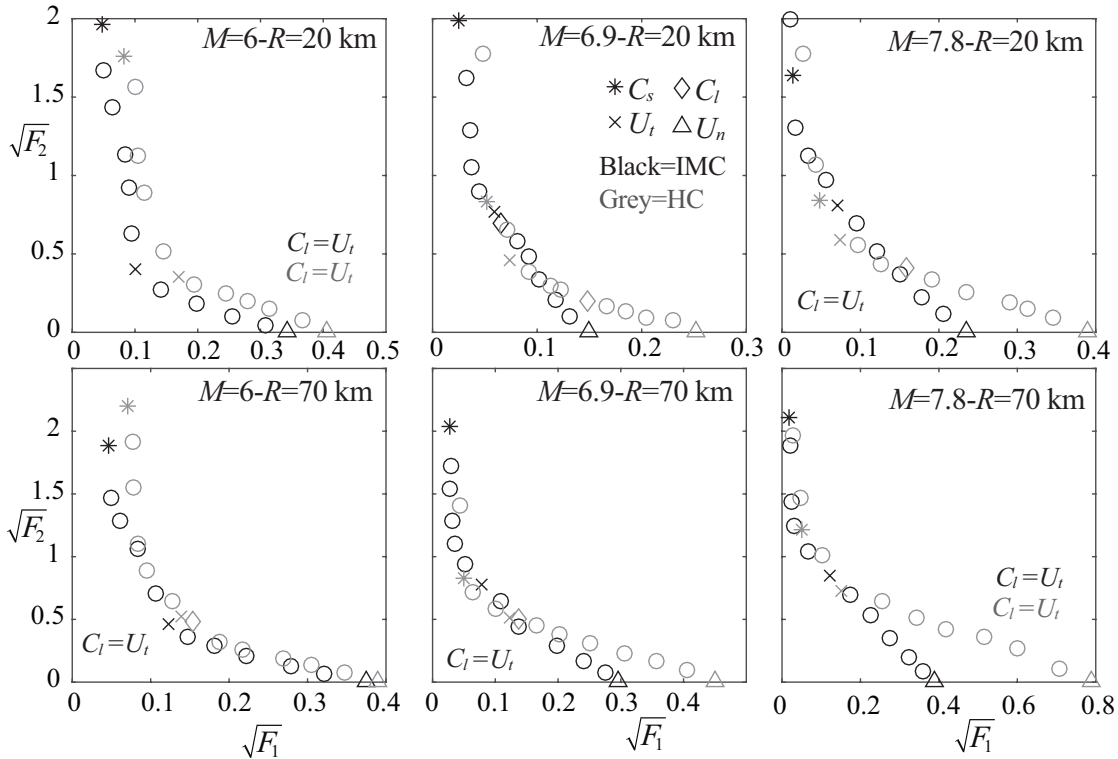
FIGURE 6.1: Pareto fronts for the stochastic ground motion modification for IMC (black) and HC (gray).

Each of the models corresponding to these three modifications (U_t, C_s and C_l) or the unmodified model (U_n) represents a different stochastic ground motion model, always for the specific scenario defined by the vector $\boldsymbol{z}$. For the IMC case, each model corresponds to a single set of ground motion model parameters since variability in the predictive model for $\boldsymbol{\theta}$ is ignored. Synthetic acceleration time-histories are obtained by utilising different white-noise sequence samples each

time, using always the same model parameter vector. For the HC case, a complete probabilistic description is established for $\boldsymbol{\theta}$. Synthetic acceleration time-histories are obtained by utilising different white-noise sequence samples along with different model parameter samples drawn from this probabilistic description.

6.3 Characteristics for validation study

6.3.1 Seismicity scenarios and target IM description

As in similar past studies (e.g., Galasso et al., 2012; Iervolino, De Luca and Cosenza, 2010; Seifried and Baker, 2016), the validation of the stochastic ground motion modification approach discussed in the previous section is performed for specific seismicity scenarios. Six different scenarios are selected, corresponding to combination of moment magnitude values $M = [6, 6.9, 7.8]$ and rupture distance values $R = [20, 70]$ km for a strike-slip fault, with shear wave velocity $V_{s,30} = 600$ m/s. Note that these are the four characteristics needed to define the stochastic ground motion model input (vector $\boldsymbol{z}$). For the remainder of this paper seismicity scenarios with $M=[6, 6.9, 7.8]$ and $R=20$ km are referred to as Scenarios 1, 2 and 3, respectively, and scenarios with $M=[6, 6.9, 7.8]$ and $R = 70$ km as Scenarios 4, 5 and 6, respectively. As target IMs, $S_a(T_i)$ in the period range $0.2T_1$-$1.5T_1$ are utilised, where T_1 is the fundamental period of the structure. An elastic period of $T_1=1$ s is selected, which is typically used as representative fundamental period of mid-rise buildings. Note that the aforementioned period range is chosen based on ASCE 7 (ASCE, 2010) provisions. The median and dispersion for the target IMs are given for each T_i as the average of four GMPEs used in the Western US (Abrahamson and Silva, 2008; Boore and Atkinson, 2008; Campbell and Bozorgnia, 2008; Chiou and Youngs, 2008). Note that this is simply chosen for consistency with Chapters 4 and 5; any other GMPEs, or any other approach that would define an IM description to match (Lin et al., 2013; Bradley, 2010), could have been used instead. The suggestions by Kaklamanos et al. (2011) are adopted

to estimate unknown inputs for some of the GMPEs that need info beyond M, R, $V_{s,30}$ and fault type. For each of the seismicity scenarios the stochastic ground motion model modification is implemented as outlined in the previous section, resulting in the Pareto fronts presented in Figure 6.1. For the U_n, U_t, C_s and C_l models, 200 synthetic acceleration time-histories are then obtained for the IMC and HC cases to be used as input for NLDA. It should be pointed out that the initial (unmodified) model established by Rezaeian and Der Kiureghian (2010) was developed strictly from the perspective of a HC implementation, with predictive models for $\boldsymbol{\theta}$ established explicitly considering the associated variability. Still U_n is examined in both IMC and HC cases here, to demonstrate the benefits of the stochastic ground motion model modification.

6.3.2   Recorded ground motions

The ground motion record set that are utilised as reference in the study, denoted as SR herein, are selected using REXEL (Iervolino, Galasso and Cosenza, 2010), a software that is freely available at `http://www.reluis.it/` and allows users to select records from the European strong motion database (or ESD, `http://www.isesd.hi.is/`), the Italian Accelerometric Archive (`http://itaca.mi.ingv.it/ItacaNet/`), and the Selected Input Motions for Displacement-Based Assessment and Design database (or SIMBAD database, Smerzini et al., 2014), which on average match a code-based or user-defined elastic spectrum in a desired period range and with specified upper and lower bound tolerances. REXEL is able to identify ground motions with desired seismicity and site characteristics (in terms of magnitude, source-to-site distance, and soil profile), which is the reason preferred for this study, as the identified reference ground motions need to have physical properties consistent with the seismicity scenario examined. For each of these six scenarios, a reference set of 30 ground motion records from the SIMBAD database was selected matching the median GMPE predictions discussed in the previous section in period range $0.2T_1$-$1.5T_1$ with a deviation from the target of

± 20%. Note that most of the chosen ground motions also belong to the NGA database (Chiou et al., 2008); this is important since the stochastic ground motion model considered here was calibrated against that database (Rezaeian and Der Kiureghian, 2010). Therefore, a consistent comparison is established between the recorded and stochastic ground motions utilised. The average values of magnitude and distance of the chosen records were [6, 21 km] for Scenario 1, [6.9, 22 km] for Scenario 2, [7.1, 24 km] for Scenario 3, [5.9, 66 km] for Scenario 4, [6.9, 70 km] for Scenario 5, and [7.5, 79 km] for Scenario 6. For the high-magnitude cases the constraint on M was relaxed, as it was not possible to identify the desired large number of ground motion records with the specific seismicity characteristics (lack of large magnitude records in the database). A uniform scaling was applied to all the records so that they match exactly the IM target for the fundamental period T_1=1 s. This was done so that for the elastic SDoF response examined later the reference case is identical to the set target, since SR is taken as the benchmark reference for the response-history analysis (so reasonable to expect match to the target IM for elastic response). Given the relatively small deviation of the selected records from the target, this uniform scaling was in most cases small.

6.3.3   SDoF system characteristics and demand measures

The validation study is performed for a number of inelastic SDoF systems with peak-oriented hysteretic behaviour, strain hardening, and (potentially) degrading characteristics as shown in Figure 6.2. The initial (elastic) SDoF stiffness k_{el} is determined based on the fundamental period T_1 of 1 s whereas a constant mass-proportional viscous damping coefficient corresponding to a 5% critical damping ratio (based on elastic stiffness characteristics) is used. With respect to the characteristics of the inelastic behaviour, the following variations are considered to establish a comprehensive validation setting:

- Strength reduction factors (R_μ). The yield strength, F_y, is chosen based on the elastic demand of the SDoF system through R_μ, defined as the ratio

of elastic base shear demand (peak elastic restoring force) to F_y. Different values of R_μ are considered to describe structural behaviour ranging from mildly inelastic (R_μ=2 and 4) to severely inelastic structures (R_μ=6 and 8). The linear behaviour ($F_y = \infty$) is also considered in this study and, for unification of presentation, it will be frequently referenced as R_μ=1.

- Hysteretic behaviour. Two different systems are examined, a non-degrading one (Figure 6.2a), and a degrading one (Figure 6.2b), referenced herein as EPH and ESD, respectively. Both of them have a strain hardening branch post-yield defined through the ratio α, and a peak-oriented hysteretic behaviour. The ESD system has an additional softening branch after the displacement Δ_u defined through the ratio β, and a residual strength of γF_y. For the EPH system, two different values of α will be examined: 3% and 10%. For the ESD system, α and β are taken to be 3% and 5%, respectively, with value of γ taken as 10% (all correspond to common values appearing in literature). The displacement for the onset of deterioration, Δ_u, is chosen to be proportional to the yield displacement, Δ_y, and the strength reduction factor R_μ, i.e., $\Delta_u = R_\mu \cdot \Delta_y$. This leads to higher ductility to systems with higher R_μ value, an assumption aligned with current design codes.
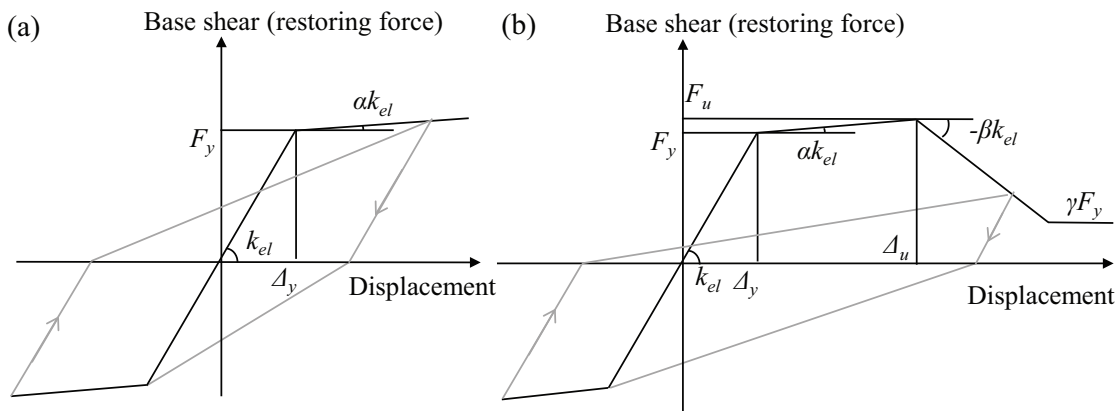
FIGURE 6.2: Hysteretic behavior model for (a) EPH system and (b) ESD system

For determining the elastic base shear demand, and therefore the values for R_μ and F_y, two different approaches are adopted: (a) achieve the same value of R_μ for each record examined; or (b) achieve the same value of F_y for each seismicity

scenario so that desired value of R_μ is obtained in an average sense for the records in the corresponding dataset. Approach (a) will be denoted herein as "constant-R_μ" and approach (b) as "constant-strength". For the "constant-R_μ" approach, the yielding strength of the structure F_y varies, ultimately, from record to record. For each record, the peak elastic base shear, F_{el}, is first calculated assuming linear behaviour (equivalent to $R_\mu=1$) and then, for each R_μ value examined, the yield force is set to $F_y = F_{el}/R_\mu$. This "constant-R_μ" approach guarantees a similar degree of nonlinearity per examined record (same, R_μ value), directly addressing the variability between ground motions by appropriately scaling SDoF strength for each of them. The "constant-strength" approach on the other hand, examines the behaviour of SDoF structures with the same characteristics (same strength) across all examined records and therefore corresponds to the implementation that better represents practical applications (e.g., based on current codes and standards). Strength F_y for each seismicity scenario is chosen based on the median IM target for that scenario, $S_{at}(T_1)$, as $F_y = mS_{at}(T_1)/R$, where m corresponds to the SDoF mass. This implementation ultimately takes the median IM target to represent the design earthquake for that Scenario and designs the SDoF structure according to that earthquake. Rather than the same R_μ for each ground motion record, this approach enforces the target R_μ on average across each examined dataset while adjusting for deviations from design demand $S_{at}(T_1)$; ultimately the average strength reduction factor achieved is $R_\mu S_{am}(T_1)/S_{at}(T_1)$ where $S_{am}(T_1)$ is the average $S_a(T_1)$ for the dataset. As discussed above, this implementation better represents practical applications: same SDoF structure with characteristics designed for each Scenario based on the design event for that Scenario.

Two different EDPs are used as representations of SDoF response: peak inelastic displacement, Δin, and the hysteretic energy, E_H, given by the work of the SDoF restoring force, or equivalently by the area under the restoring force/displacement curve (Figure 6.2). These parameters are considered in order to investigate, respectively, the peak displacement demand and the cyclic behaviour as also done in past SDoF studies (Galasso et al., 2012; Iervolino, De Luca and Cosenza, 2010).

E_H will be presented normalised by SDoF mass m.

6.4   Comparison of synthetic and recorded ground motions to target spectra

Before discussing the validation study in terms of inelastic structural response, the elastic spectra for the synthetic and recorded ground motions are first presented and compared in this section. Figure 6.3 shows the average spectral estimates from the suite of recorded SR and stochastic ground motions corresponding to models U_n, U_t, C_s and C_l for the IMC case and for all six Scenarios. The target spectra are also shown. For SR statistics are shown for the motions obtained directly from REXEL, without the scaling that was utilised to create the reference ground motion set. Figure 6.4 presents the results for the HC case. Figure 6.5, finally, presents the dispersion of the spectral estimates for all implementations examined, covering both the IMC and HC cases for the stochastic ground motion models.

The recorded SR ground motions have high compatibility with the target as shown in Figures 6.3 and 6.4, something attributed to the use of small tolerance (20%) and the fact that the constraints on seismicity characteristics were relaxed for some of the Scenarios in order to satisfy this tolerance. The REXEL optimisation routine exploited this relaxation to provide suites of ground motions with small discrepancy from the target. The dispersion (Figure 6.5) of the stochastic ground motions for the IMC case is much lower than the target as expected because only the white-noise is contributing to the observed dispersion in this case. On the other hand, the HC stochastic ground motions have a higher dispersion that is comparable to the target. The stochastic ground motion model modification achieves a better match to the target dispersion compared to the unmodified model case. Finally, the dispersion of the recorded ground motions (SR) is higher than that of IMC modifications and, for some scenarios, reaches or exceeds the target. Any discrepancies from the target are justifiable since the REXEL optimization only

tries to match the target spectrum through the selection of ground motions and not, additionally, the dispersion of these ground motions (which is only indirectly controlled through the chosen tolerances).
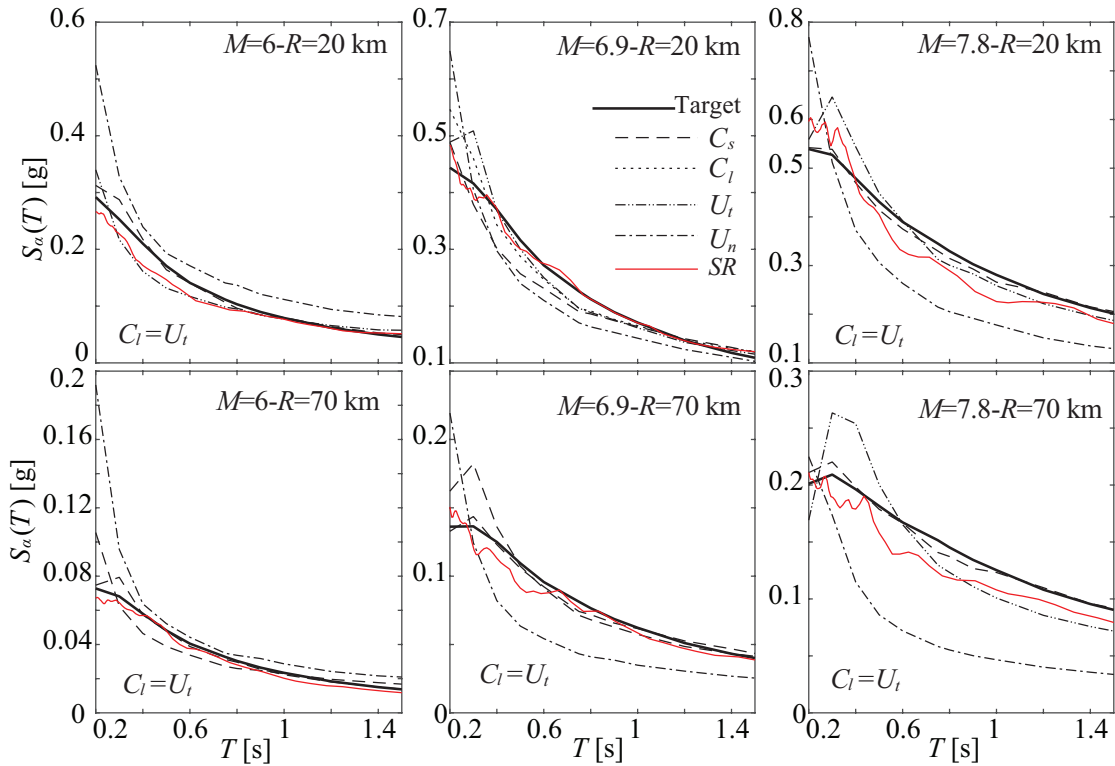
FIGURE 6.3: Spectral plot comparison of target spectra and average predictions of recorded (SR) and stochastic ground motions for IMC.

For the stochastic ground motion suites, the results agree with the ones presented earlier in Figure 6.1 with respect to the discrepancy from the target (F_1 values). The unmodified model, U_n, does not provide a good match to the desired target for some seismicity Scenarios, overpredicting the resultant spectral acceleration values for small M values and underpredicting them for large M values. Note that this trend agrees with the results reported by Rezaeian and Der Kiureghian (2010) when comparing their model to some of the GMPEs utilised here. The HC case (Figure 6.4) provides better match than the IMC case (Figure 6.3) for U_n, which should be expected since, as also commented earlier, the predictive models were developed by Rezaeian and Der Kiureghian (2010) assuming a HC implementation. The proposed modification (U_t, C_s and C_l models) now, facilitates in all instances an improved match for both the IMC (Figure 6.3) and the HC cases (Figure
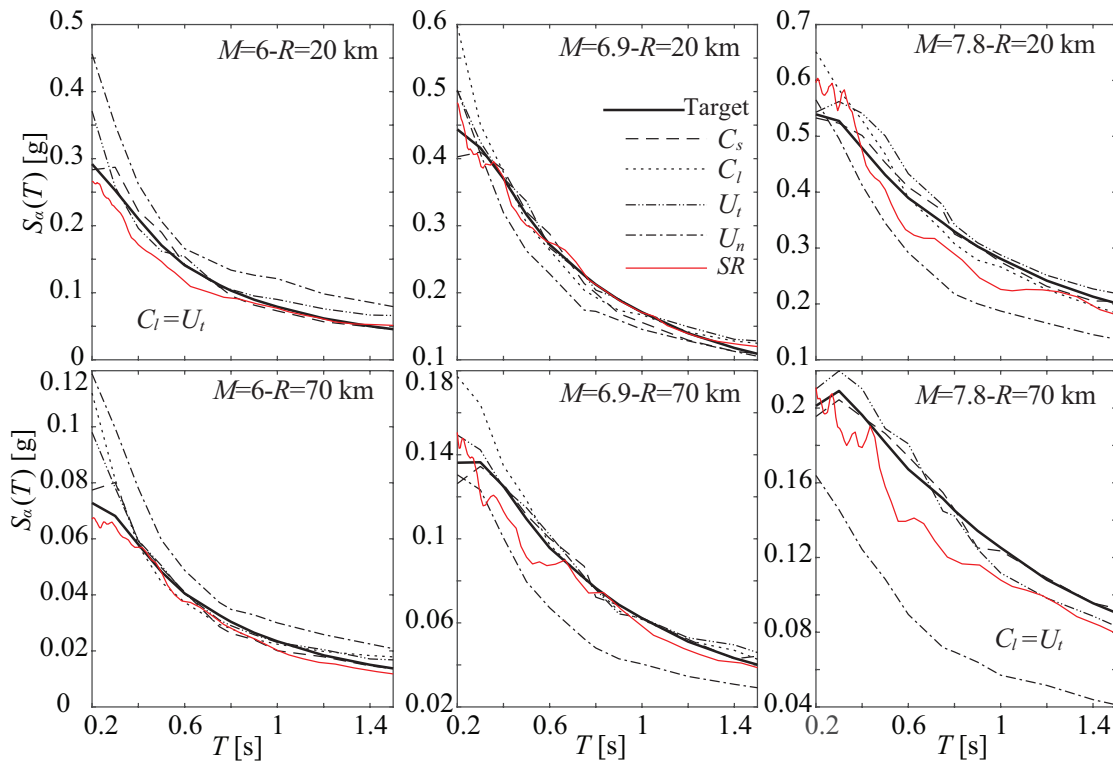
FIGURE 6.4: Spectral plot comparison of target spectra and average predictions of recorded (SR) and stochastic ground motions for HC.

6.4) with C_s achieving in all instances very high compatibility, even better than SR. This further validates the ability of the modification framework proposed in Chapters 4 and 5 in facilitating an improved match to a target IM. With respect to the dispersion (Figure 6.5), all IMC models significantly underestimate the target variability. This is attributed to the fact that the only source of variability for the response stems from the stochastic nature (white-noise sequence) of the models since the variability in the predictive relationship for $\boldsymbol{\theta}$ is ignored. This leads to smaller response dispersion compared to the one observed in the recorded ground motions, and corresponds to an important shortcoming of the IMC modification approach when that dispersion is also of importance (e.g., when assessing collapse risk due to more extreme ground motion records). The HC case, on the other hand, can explicitly control this dispersion through adjustment of $\boldsymbol{\Sigma}$ in the predictive relationships, and as evident from Figure 6.5, high compatibility is achieved for the modified models U_t, C_s and C_l improving upon the unmodified one U_n.
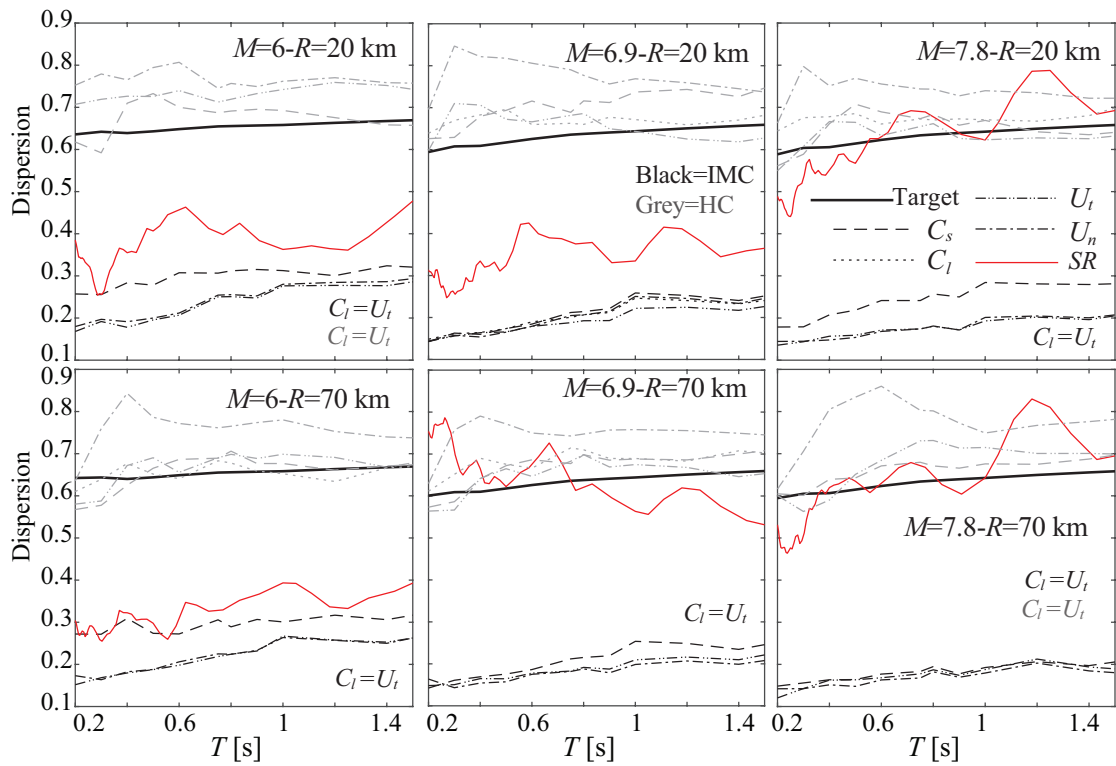
FIGURE 6.5: Comparison of dispersion of target, recorded (SR) and stochastic ground motions (U_n, U_t, C_s) for IMC and HC.

6.5   Comparison of inelastic demand for "constant-R_μ" approach

Focus is shifted next to the validation study in terms of structural response. For each seismicity Scenario, the suites of recorded and simulated ground motions are used as input to the different SDoF systems to perform nonlinear response-history analysis. For each considered system and EDP, Δ_{in} and E_H, the statistics, namely median and CoV, are estimated across each suite. Results are reported for different values of R_μ and further distinction is established between EPH and ESD SDoFs. For the synthetic ground motions, results are reported for the IMC and HC cases together in each figure using color pattern black and gray, respectively. To more clearly depict differences with respect to the reference (benchmark) SR results, the relative error between the response output for any stochastic ground motion modification case and SR is introduced, calculated for the output statistic s as

$$E(s) = \frac{s(SM) - s(SR)}{s(SR)} \tag{6.1}$$

where $s(SM)$ is the output statistic (median or CoV) from the simulated ground motions and $s(SR)$ is the same output statistic from the recorded ground motions. Results for the "constant-R_μ" approach are first discussed in this section.

Figures 6.6 to 6.10 present results for the EPH system with α=3%. Figure 6.6 presents the normalised results for the median Δ_{in} estimates, $\bar{\Delta}_{in}$, while Figure 6.7 the relative error $E(\bar{\Delta}_{in})$. Normalisation in Figure 6.6 is established with respect to the elastic spectral displacement S_d. Figure 6.8 presents results for the median hysteretic energy burned E_H, $\bar{E}_H$. The relative error for the hysteretic energy $E(\bar{E}_H)$ is presented in Appendix D. Figures 6.9 and 6.10 show dispersion results (expressed through the CoV) for Δ_{in} and E_H, respectively. For these, and all remaining figures, the cases where $C_l=U_t$ is not explicitly denoted (as was done in previous figures); simply C_l is not reported for these instances, corresponding to Scenarios 1 and 3 to 6 for IMC and Scenarios 1 and 6 for HC.
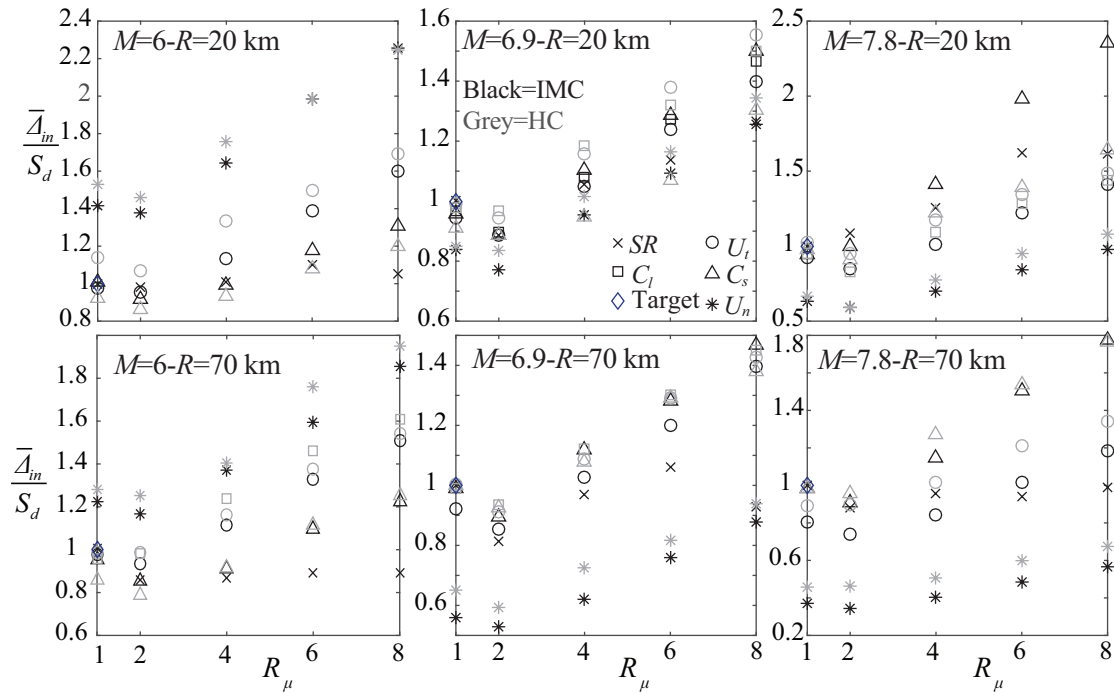
FIGURE 6.6: Normalised median peak inelastic displacements for EPH system with α=3% for "constant-R_μ" approach.
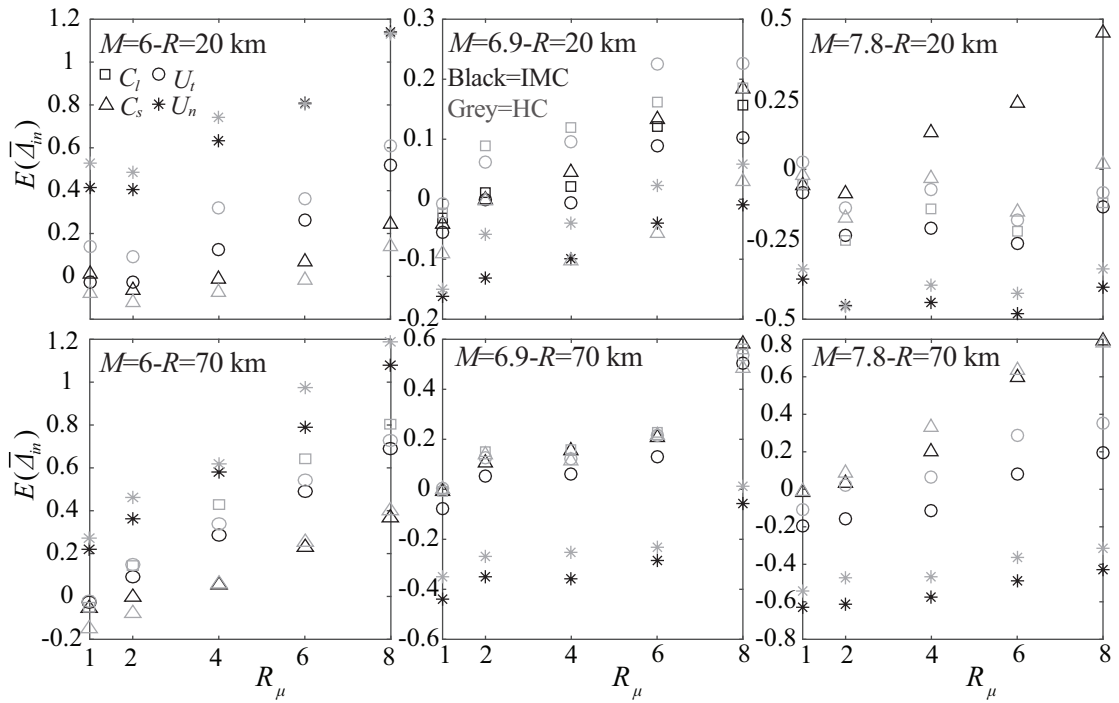
FIGURE 6.7: Relative error compared to reference SR response for the peak inelastic displacements for EPH system with $\alpha=3\%$ for "constant-R_μ" approach.

Results show that the proposed modification facilitates overall a better match to the reference results of the recorded ground motions in terms of median response statistics. Exceptions to this general trend exist only for significant degree of inelastic behavior (R_μ value equal to 8) and for scenarios for which the unmodified model provided a good match to the (elastic) target hazard to start with (Scenarios 2 and 5). In those instances, the unmodified ground motion model has a better match to the SR statistics. Note, though, that the error of the proposed modification in these instances is still small. Overall, the absolute error of all the modified ground motion models stays consistently below 40% to 50% (and in most instances in range of 20% to 30%), with exception of large values of R_μ for Scenario 6. This is not true for the unmodified model which has errors exceeding 100% in some instances. The modification also contributes to smaller sensitivity of the behaviour across the different examined scenarios; even though great variability is observed for the unmodified model U_n across the different scenarios, this variability is reduced for the results of the modified ground motion models. This variability is small for $R_\mu=1$ as expected (since modification matches the target for

elastic behaviour) and increases as degree of inelastic behaviour increases (larger values of R_μ). For small values of R_μ, there is a strong correlation of the results to the $R_\mu=1$ case for Δ_{in} and therefore to the results reported in Figures 6.3 and 6.4 or the reported F_1 values in Figure 6.1. Note that for large values of R_μ, the nonlinear structural response is sensitive to spectral ordinates at periods much larger than the fundamental one (e.g., due to period elongation stemming from the strong nonlinear behaviour); the chosen period range for spectral compatibility (i.e., $0.2T_1$-$1.5T_1$) may not be conservative in those cases (Katsanos and Sextos, 2015), yielding the observed large variability for larger values of R_μ.
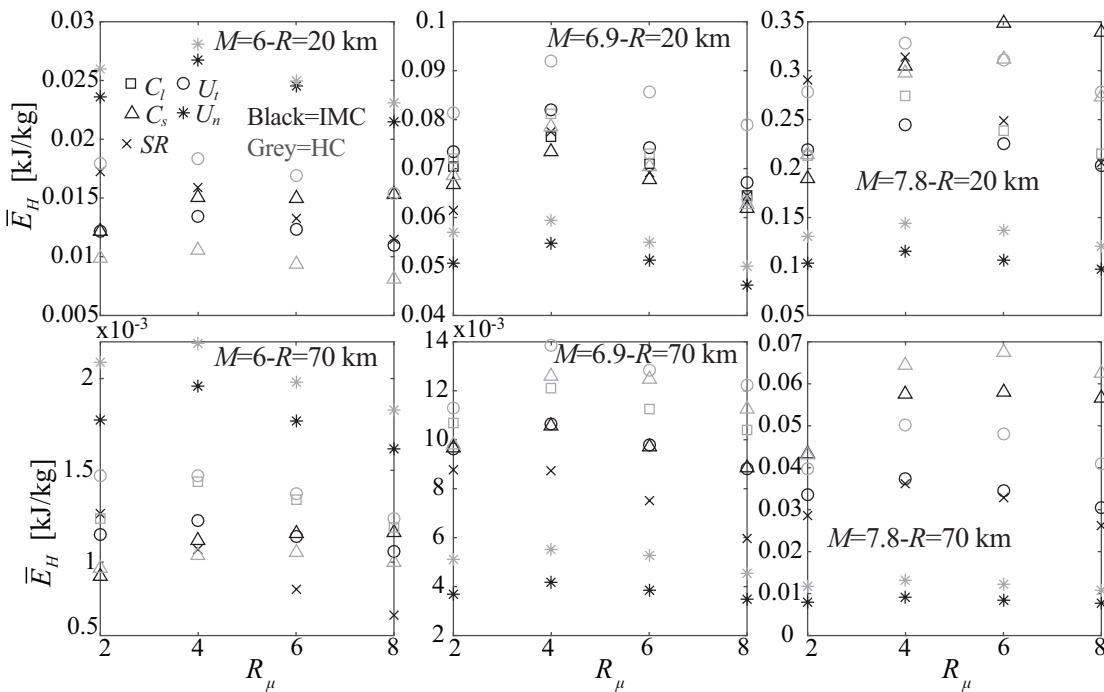
FIGURE 6.8:   Median hysteretic energy for EPH system with $\alpha=3\%$ for "constant-R_μ" approach.

In general, the IMC and HC modification cases yield very similar trends for the median response and similar results, except for some large R_μ value instances for Scenario 3. Comparing the different modification implementations, C_s provides overall the smallest errors, except for large values of R_μ for Scenarios 3 and 6. This might lead someone to conclude that the significant alteration of ground motion physical characteristics, established in the C_s case, might have an impact when looking at high levels of inelastic behaviour. The discrepancies observed could be

attributed, though, to the fact that SR implementation for these two scenarios led to ground motions with different seismicity characteristics than the targeted M, R values, and therefore, possibly, to different physical characteristics for the ground motions than expected for these scenarios. As such, any discrepancies for large degree of inelastic behaviour, that is for the instances these physical characteristics are influential, might not be surprising. Still even for these two scenarios the recommended modification, corresponding to C_l (which recall is equal to U_t in some instances), yields small errors. All these trends hold for both the peak displacement (Figures 6.6 and 6.7) as well as for the hysteretic energy (Figure 6.8). This is an important feature as both these EDPs are commonly used to describe performance in earthquake engineering applications (Ruiz-García and Miranda, 2003; Deniz et al., 2017).

The information theory-based approach proposed in Chapter 3 for the validation of ground motion simulations was applied to the peak inelastic displacements for the EPH system with α=3%. More specifically, the Δ_{in} distributions for the modified U_t, C_s and unmodified U_n stochastic ground motion models were compared with the "benchmark" distribution for the recorded ground motions. The observed D_{KL} values and the corresponding p-values for HC case and Scenario 1 are presented in Tables 6.1 and 6.2, respectively. The cases of rejection characterised by p-values smaller than 0.05 are highlighted gray in the Tables. The case with the smallest D_{KL} in each column is the one that matches best the "benchmark" distribution and is shown in bold font in Table 6.1. The results show that U_n model provides a poor fit to the Δ_{in} distributions from recorded ground motions for all R_μ values considered, except for the elastic case R_μ=1. This can also be observed from the fact that the modified stochastic ground motion models provide the best match to the recorded distribution for all R_μ cases. The U_t case has large discrepancies for large R_μ values (6 and 8), while the C_s case shows rejection for R_μ=4. In general, the higher D_{KL} values in Table 6.1 are associated with smaller p-values in Table 6.2. The results are in agreement with Figures 6.6 or 6.7. For cases where the median peak inelastic displacements have large discrepancies from the

"benchmark" recorded case, the distance between the two distributions is large leading to small p-values. The approach was applied to other EDP cases, SDoF systems and Scenarios and similar trends were observed.

TABLE 6.1: D_{KL} values for Δ_{in} for HC case and Scenario 1

D_{KL}	$R_\mu=1$	$R_\mu=2$	$R_\mu=4$	$R_\mu=6$	$R_\mu=8$
U_t	0.385	0.335	**0.339**	**0.502**	0.605
C_s	**0.216**	**0.196**	0.767	0.542	**0.283**
U_n	0.416	1.173	0.588	1.237	0.928

TABLE 6.2: p-values for Δ_{in} for HC case and Scenario 1

p-value	$R_\mu=1$	$R_\mu=2$	$R_\mu=4$	$R_\mu=6$	$R_\mu=8$
U_t	0.084	0.236	0.254	0.025	0.008
C_s	0.275	0.368	0.027	0.07	0.349
U_n	0.146	0.001	0.07	0	0
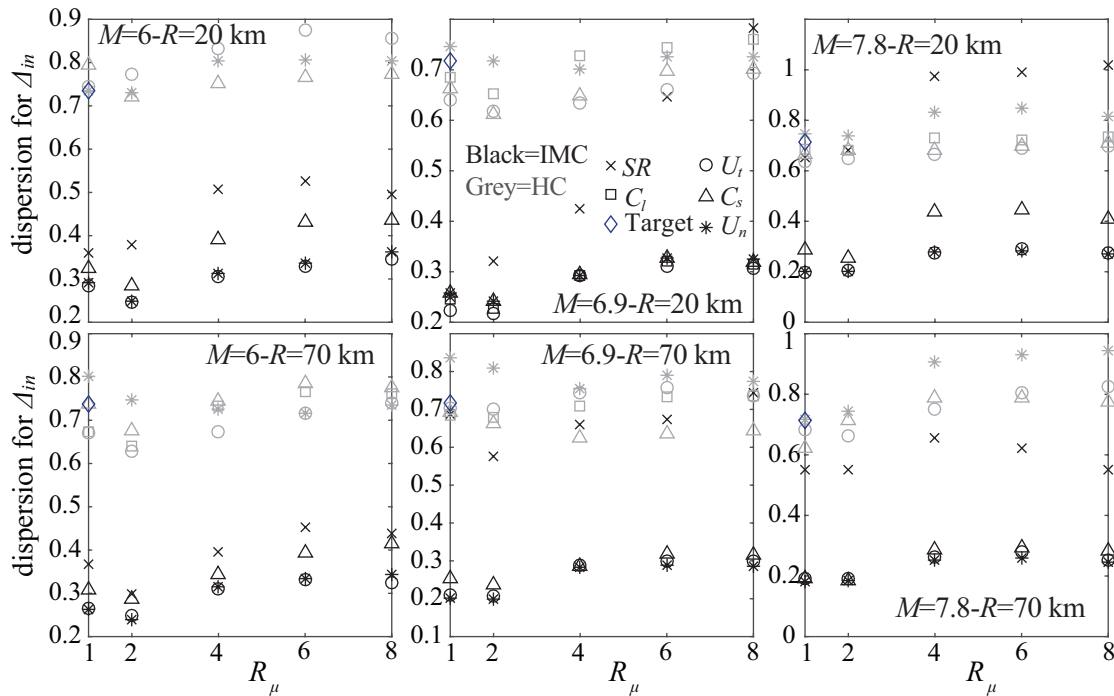
FIGURE 6.9: Dispersion (expressed through CoV) of peak inelastic displacement for EPH system with $\alpha=3\%$ for "constant-R_μ" approach.

For the dispersion characteristics (Figures 6.9 and 6.10), the variability trends reported in Figure 6.5 extend to the inelastic behaviour. Significant differences
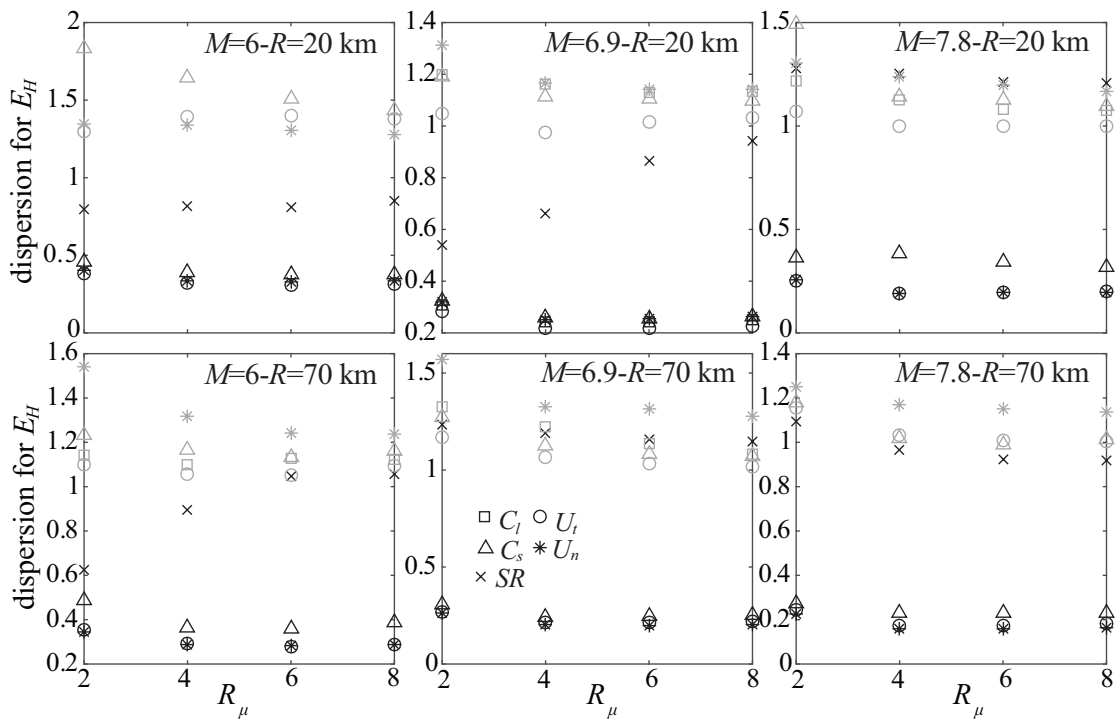
FIGURE 6.10: Dispersion (expressed through CoV) of hysteretic energy for EPH system with α=3% for "constant-R_μ" approach.

exist for this statistic between the IMC and HC applications, as expected, with HC providing enhanced compatibility to the target or reference/benchmark values. This, once more, demonstrates the importance of facilitating hazard compatibility, rather than simply IM compatibility (Tsioulou et al., 2018b; Lin et al., 2013). In general, results for most modification implementations are very similar. This should be attributed to the fact that the unmodified model is close to the target dispersion (so small modifications are only required) and the fact that as explained in Chapter 5 matching of the median statistics is typically more important than matching dispersion statistics for facilitating hazard compatibility, characteristic that leads to smaller modification of the dispersion statistics. Variation of R_μ in general, does not significantly affect the observed dispersion patterns. For the seismicity scenarios for which the spectral dispersion from records (SR) is close to the target, the former is also close to the HC modification. For other scenarios the differences between SR and the HC modifications remain similar to the differences between SR and the target dispersion, apart from Scenario 2 for which SR itself demonstrates a bit of irregular trend, with significant variation of

dispersion across different R_μ values. Overall, trends are again consistent for both the peak displacement and the hysteretic energy.

Figure 6.11 repeats results of Figure 6.6 but for the EPH system with α=10%. Results show exceptionally similar patterns indicating little sensitivity to value of α (as long as latter is in reasonable range). Similar pattern holds for the other statistics, included in Appendix D, thus focus is placed herein on the α=3% case.
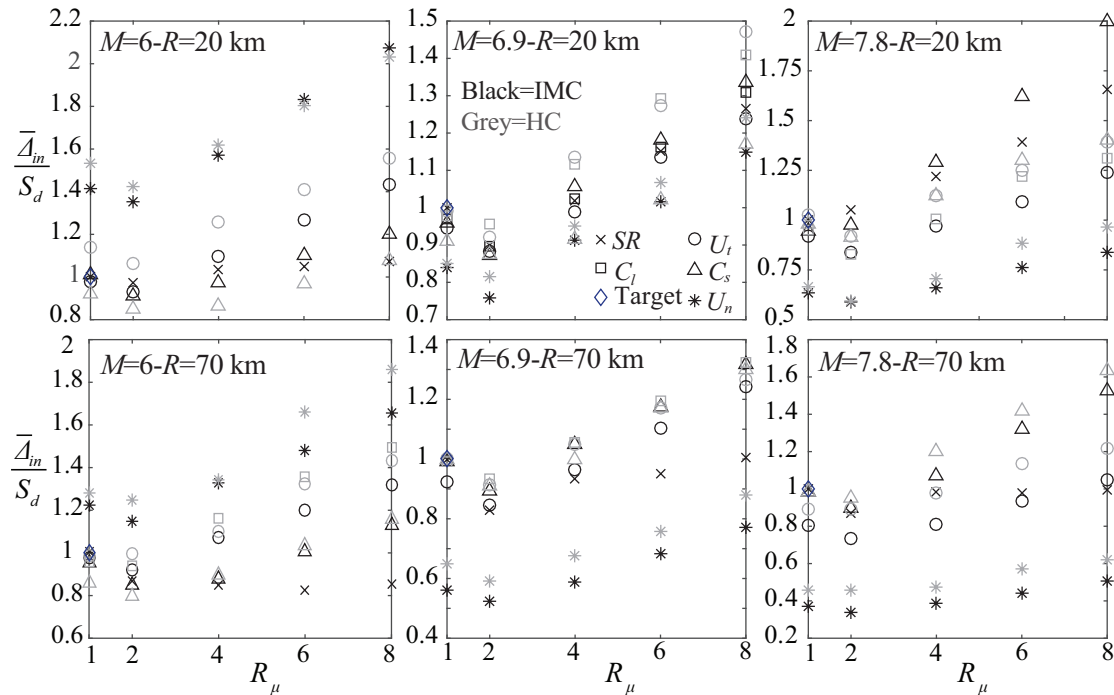
FIGURE 6.11: Normalised median peak inelastic displacements for EPH system with α=10% for "constant-R_μ" approach.

Figures 6.12 to 6.14 present results for the ESD system for Δ_{in}. Results for E_H are included in Appendix D, since they are practically identical to the trends observed for EPH with some reduction in the E_H values for large R_μ values, due to the reduced energy dissipation capabilities when structure enters softening branch of backbone curve.

Differences to the EPH case appear only for the large nonlinearity cases (R_μ=6 or 8), since for lower R_μ values system does not move significantly into the softening branch. Discrepancies are amplified for the ESD model with respect to both the median response (Figures 6.12 and 6.13) as well as dispersion (Figure 6.14).
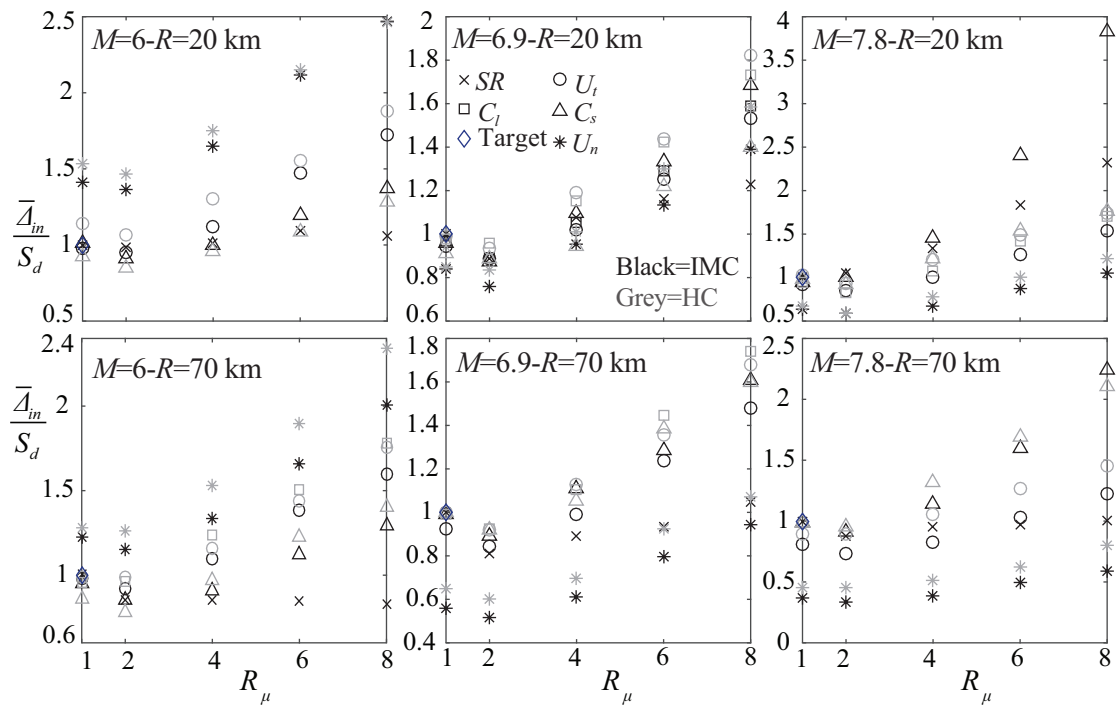
FIGURE 6.12: Normalised median peak inelastic displacements for ESD system
for "constant-R_μ" approach.
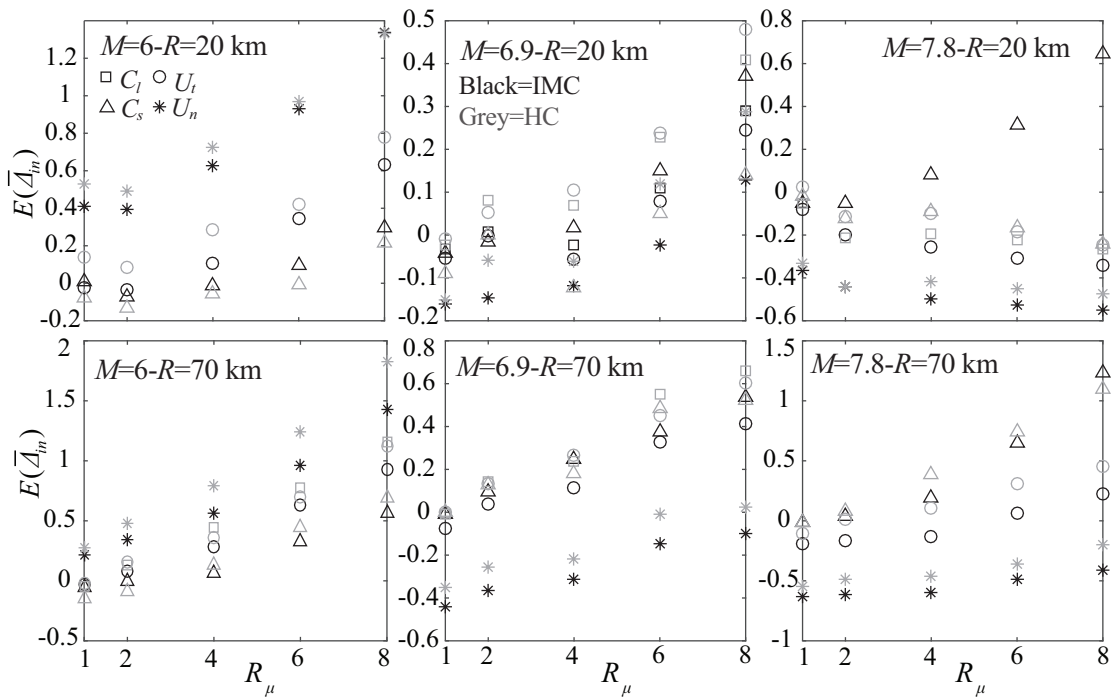
FIGURE 6.13: Relative error compared to reference SR response for the peak
inelastic displacements for ESD system for "constant-R_μ" approach.

For the R_μ=8 case and for a considerable portion of the simulations (over 10%
for most Scenarios), structure reached residual strength, so high variability and

larger discrepancies (Figure 6.14) can be also attributed to that; once the residual strength is reached, the response output is very sensitive to small changes in the ground motion features. For the median statistics, errors are especially amplified for the modification cases that had large errors for the EPH model (e.g., C_s model for Scenarios 3 and 6). For the dispersion statistics the agreement to the reference/benchmark results deteriorates overall.
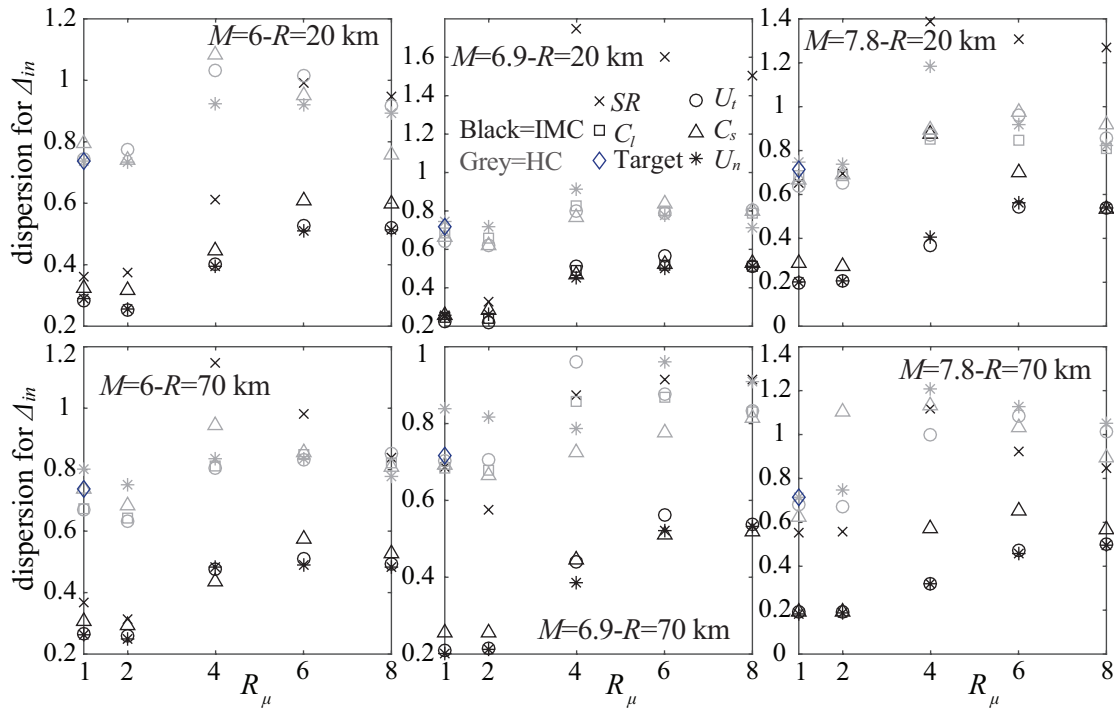
FIGURE 6.14: Dispersion (expressed through CoV) of peak inelastic displacement for ESD system for "constant-R_μ" approach.

6.6 Comparison of inelastic demand for "constant-strength" approach

This section extends comparison to "constant-strength" approach. Figures 6.15 and 6.16 show results for the EPH system with α=3%: the relative error $E(\bar{\Delta}_{in})$ and coefficient of variation for Δ_{in}, respectively. Figures 6.17 and 6.18 present results for the relative error $E(\bar{\Delta}_{in})$ and coefficient of variation for Δ_{in}, respectively

for the ESD system. The results for $\bar{\Delta}_{in}$ are included in Appendix D and the trends are same as these discussed next for $E(\bar{\Delta}_{in})$.
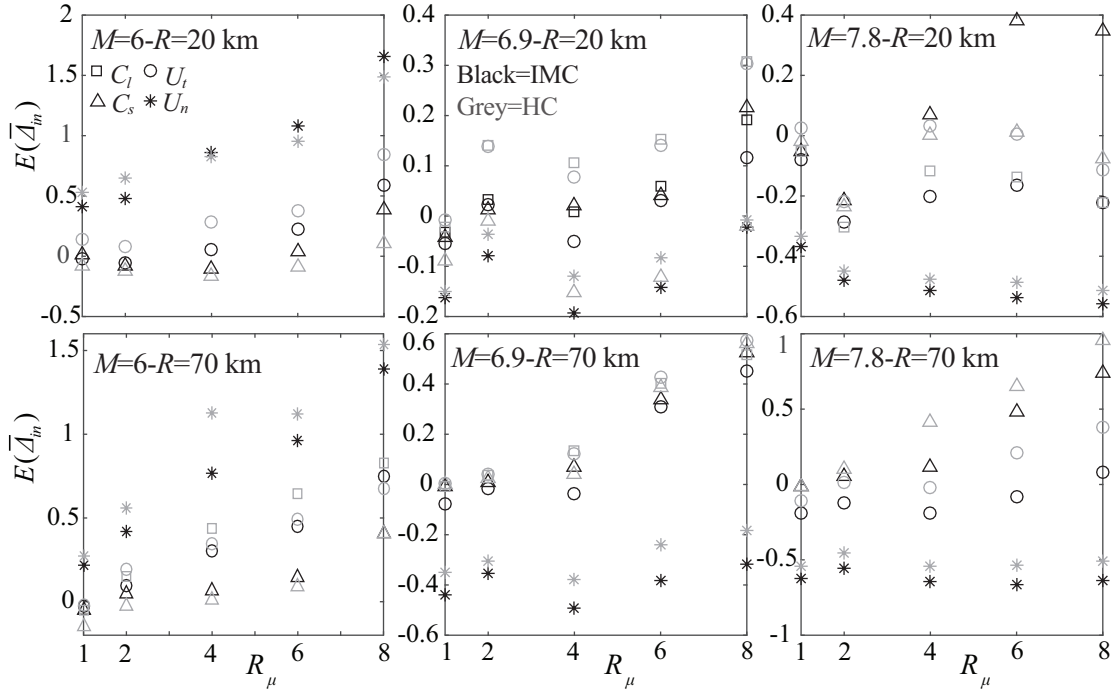
FIGURE 6.15: Relative error compared to reference SR response for the peak inelastic displacements for EPH system with α=3% for "constant-strength" approach.

Results for the median response (Figures 6.15 and 6.17) have similar trends, with notable higher error with respect to reference SR for the U_n compared to the respective plots for "constant-R_μ" (Figures 6.7 and 6.13). Specifically, they show even higher error estimates for Scenarios 1 and 4 and even lower estimates for Scenarios 2, 3, 5 and 6. This difference stems from the significant discrepancies for the elastic case R_μ=1 for U_n which contributes to higher or lower strength compared to the benchmark case for the two groups of Scenarios, respectively. For the modified stochastic ground motion model, on the other hand, only small differences exist with respect to the "constant-R_μ" approach, keeping the errors moderately low. This is expected; since the modification leads to a good match to the target (R_μ=1 case), not significant differences are expected between the "constant-R_μ" and "constant-strength" implementations. This further stresses the importance of the proposed modification, as it is able to facilitate good match to the benchmark results compared to the unmodified ground motion model for

the "constant-strength" implementation which corresponds, as mentioned earlier, to the implementation that is more comparable to practical applications.
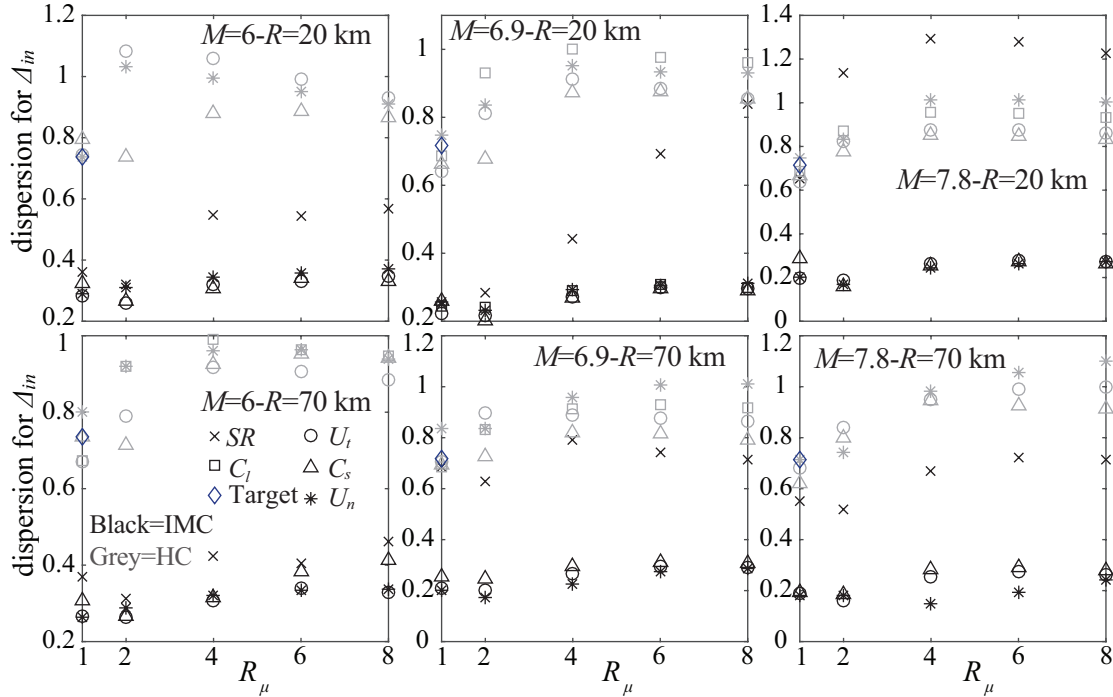
FIGURE 6.16: Dispersion (expressed through CoV) of peak inelastic displacement for EPH system with $\alpha=3\%$ for "constant-strength" approach.

Finally, results in Figures 6.16 and 6.18 show that the "constant-strength" approach leads to higher dispersion estimates for the peak inelastic displacement ratios for the HC modification and SR cases. A possible reason for this is that the higher variability of the ground motion response in these sets is affected more by the adoption of a constant yield strength that results in higher dispersion estimates for this EDP. Still though, there is a similar level of agreement between HC and SR as in the "constant-R_μ" approach. Results for E_H are included in Appendix D and the trends are identical for EPH and ESD systems. It is worth noting that for some Scenarios $\bar{E}_H$ is zero or very small for the unmodified U_n model and case of $R_\mu=2$. This implies that the SDoF systems do not reach the inelastic region and leads to inaccurate dispersion estimates for E_H (unreasonably high CoV estimates) in these cases. These plots are included in the thesis to showcase potential challenges U_n model faces when match to the target is not achieved.

It is worth noting that both the simulated and recorded ground motions follow the equal displacement rule for both SDoF systems considered herein.
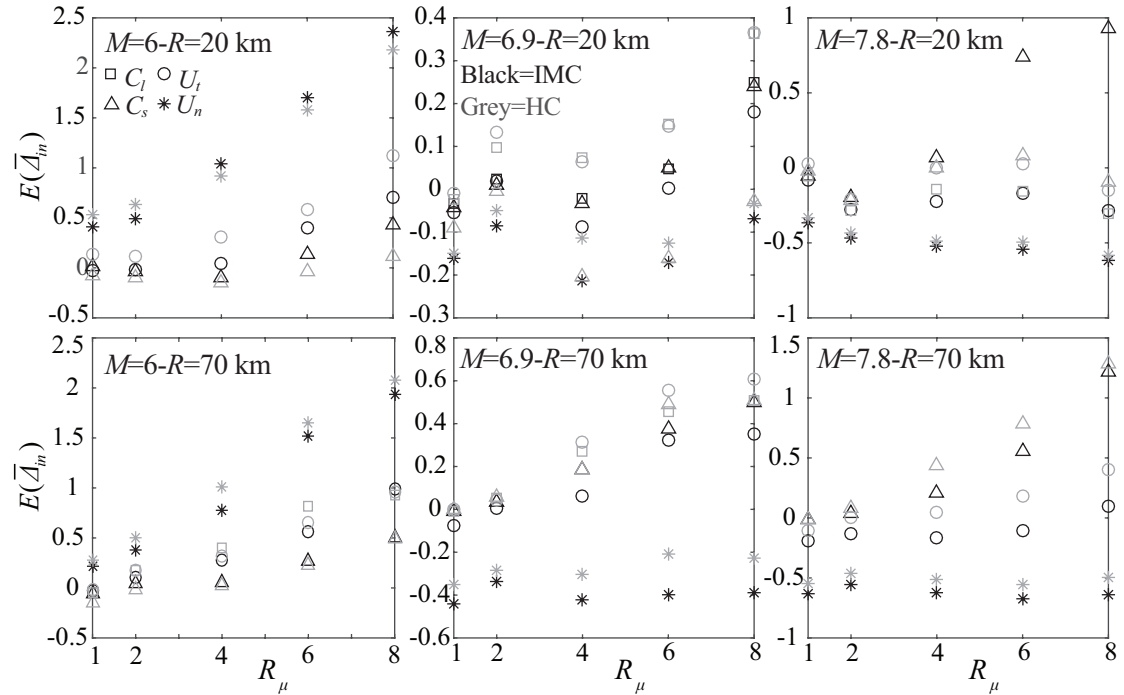
FIGURE 6.17: Relative error compared to reference SR response for the peak inelastic displacements for ESD system for "constant-strength" approach.
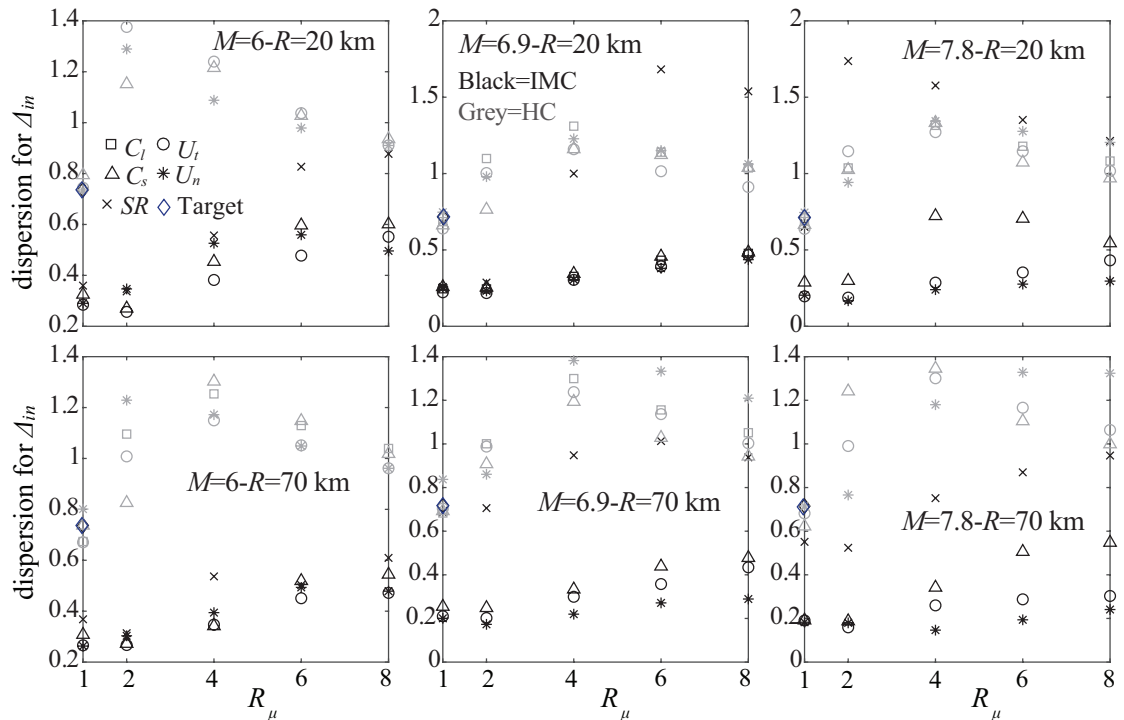
FIGURE 6.18: Dispersion (expressed through CoV) of peak inelastic displacement for ESD system for "constant-strength" approach.

6.7   Conclusions

A validation study for the stochastic ground motion model modification proposed in Chapters 4 and 5 was performed in this chapter by comparing the seismic demand for inelastic SDoF systems of hazard-compatible recorded ground motions to the demand of stochastic ground motion models that are modified to match the same target hazard. Comparison was performed for two different EDPs, peak displacement and hysteretic energy, and for SDoF systems with different degrees of inelastic behaviour, corresponding to different values of strength reduction factor R_μ, and different nonlinear and hysteretic characteristics. For the latter a peak-oriented hysteresis model was chosen with different values for post-yield stiffness, considering non-degrading (EPH system) or degrading (ESD system) strength characteristics. Six different seismicity scenarios were examined, corresponding to different values of moment magnitude and rupture distance for seismic events. The recorded ground motions were obtained through the REXEL software while for the modified stochastic ground motion model, different modification degrees were examined.

Results show that the proposed modification improves significantly the match to the reference (benchmark) results corresponding to the recorded ground motion model. As the degree of inelastic behaviour increases, that is for larger value or R_μ or for ESD system (compared to EPH system), the differences to the reference results increase. Also, for large degrees of modification, larger errors may exist for such instances of significant inelastic behaviour. The moderate modification approach proposed in Chapters 4 and 5 appears to consistently yield good results across all seismicity scenarios and types of inelastic behaviour. Trends were similar for both considered EDPs. With respect to the two types of modifications examined, IM compatibility (IMC) and hazard compatibility (HC), while both match the median statistics similarly well, HC was shown to provide an enhanced match to the target dispersion, with IMC constrained to small dispersion values. It is worth noting that the intent of this study was not to provide a definite judgment

about the specific stochastic ground motion simulation method, but rather to illustrate and validate the proposed modification and discuss possible outcomes. The identified similarities to recorded ground motions should provide confidence in using the modification method for engineering applications, while the discrepancies observed for some seismicity scenarios for highly inelastic response cases, emphasises domains of potential improvement for future stochastic ground motion simulation methods or their potential modification.

Chapter 7

Conclusions

7.1 Summary and conclusions

Ground motion simulations are a valuable alternative or supplement to recorded ground motions that have gained increasing interest within the earthquake engineering community in the past decades. They can be grouped in three categories: physics-based, stochastic-process-based (or stochastic) and hybrid, that are a combination of the other two methods. Physics-based and hybrid models have developed rapidly in the past few years because of advances in computing power and the better understanding of source characteristics and wave propagation. They are based on numerical models that simulate the physics of the rupture and wave propagation phenomena and can thus, be used for future predictions. Stochastic methods can be further grouped into source-based and site-based models. Stochastic source-based models describe the fault rupture at the source and the propagation of the seismic waves at the site of interest and they are similar to physics-based models in that sense. Site-based models on the other hand, describe the ground motion time-history at a specific site by fitting a statistical process to recorded ground motions with known earthquake and site characteristics. They are fast to compute and do not require thorough knowledge of various seismological

parameters, therefore they have potential to be used by practicing engineers.

The contributions, detailed next, established in this thesis related to the field of ground motion simulation are summarised as:

- A novel approach for validation of ground motion simulations

- A stochastic ground motion model modification framework

- Validation of the proposed stochastic ground motion model modification framework

7.1.1 Proposed validation approach

An important consideration regarding ground motion simulations that rely on physics-based modeling and seismological information (physics-based, stochastic source-based and hybrid methods) is that they need to be properly validated to be used with confidence for future predictions. There are three main validation methods: historical method where simulations are compared to records of a historical earthquake, empirical model method that compares simulations to empirical models (i.e., GMPEs), and similar spectra method where simulations and recordings with similar elastic spectra are compared. Several goodness-of-fit measures have been proposed as validation metrics in addition to statistical approaches (i.e., hypothesis testing) to perform a paired comparison between the recorded and simulated IM or EDP datasets in terms of their mean and dispersion. This study proposed a validation approach based on the relative entropy D_{KL} coupled with statistical hypothesis testing to assess the overall similarity of the probability distributions of the studied IMs or EDPs for recorded and simulated ground motions. The approach was demonstrated by using ground motion simulations generated by all three ground motion simulation methods: physics-based, stochastic and hybrid. The validation approach also allows the user to rate the simulation methods, not just provide a judgment about the fit for each method. Simple spectral-shape

and duration-related proxies that are easy to compute and correlate well with the seismic response of more complicated structural systems were introduced as validation metrics. The ground motion simulation methods considered, overall performed well in matching the recorded ground motions. However, since simulation methods evolve very fast, this study was not intended as an evaluation of specific methods, but as a development of a validation approach that can be used in the future.

7.1.2 Proposed stochastic ground motion model modification framework

This study also focused on a class of stochastic site-based ground motion simulations that are based on modulation of a stochastic sequence, through functions (filters) that address spectral and temporal characteristics of the excitation. The parameters of these filters are related to seismicity (e.g., moment magnitude and rupture distance) and site characteristics (e.g., shear wave velocity for soil profile) through predictive relationships. For these models to be used in engineering applications, it is important that the output IMs from the simulated acceleration time-histories are consistent with these prescribed at the site of interest (e.g., $S_a(T)$ estimates from GMPEs) for specific structures. That is though, not necessarily guaranteed through the current approaches in selecting the predictive relationships of the model parameters.

To address this issue, this study has proposed a computationally efficient framework (IMC) to modify stochastic ground motion models for specific seismicity scenarios with a dual goal of (i) matching a target IM for a specific structure while (ii) preserving desired trends and correlations in the physical characteristics of the resultant ground acceleration time-series. The latter goal was identified as a shortcoming of past studies and for that reason was explicitly incorporated as an objective rather than a constraint within the predictive relationship tuning. This was set as an optimisation problem with a dual objective and computational

efficiency in its solution was achieved by adopting a metamodel for approximating the median ground motion model predictions for the targeted IMs. Although the upfront cost for development of this metamodel was significant, this only needs to be performed once and it can subsequently be used to support a highly efficient multi-objective optimisation. To reduce the computational burden for the metamodel development and prediction, an adaptive design of experiments was proposed for selecting the database informing the metamodel. Gradient-based and gradient-free approaches were discussed to solve the resultant optimisation problem, whereas the inclusion or not of the metamodel error was also considered.

The framework was demonstrated in an illustrative example considering an existing record-based stochastic ground motion model and IMs described through GMPEs. It was shown that the metamodel-aided optimisation can support an accurate identification of the Pareto front of dominant solutions, provided that the metamodel accuracy is significant high, and that inclusion of the metamodel error in the optimisation formulation greatly improves the robustness of this optimisation. Comparisons between the two optimisation approaches showed that the gradient-free one demonstrates overall preferable attributes, since the gradient-based one might converge to suboptimal local minima. An adequate representation of the overall Pareto front can be obtained in as little as two minutes using the blind search, gradient-free optimisation, which should be considered as an acceptable computational burden. The gradient-based approach provides greater relative efficiency when identification of a single solution, rather than of the entire front, is desired. It was shown that the proposed framework can lead to significant improvement for the match to the target IM.

The developed framework was extended (HC) to perform modification of stochastic ground motion models to (i) match the prescribed conditional hazard (mean and dispersion of IMs) for a specific site and structure while (ii) preserving desired

trends and correlations in the physical characteristics of the resultant ground acceleration time-series, including consideration of the variability of these characteristics. The proposed modification was defined again as a bi-objective metamodel-optimisation and the relative entropy was adopted as metric to quantify the two objectives. A MCS approach was utilised for the estimation of the various response statistics needed for the entropy calculation, where the metamodel was used to approximate median and dispersion of ground motion model IM predictions. The metamodel-aided optimisation can facilitate an accurate identification of the Pareto front even when lower accuracy metamodels are utilised. The necessity to calculate response statistics through MCS increases, though, the computational burden of the extended approach.

With respect to selection of the final model across the identified Pareto front, same recommendation is made for both approaches: select the Pareto optimal solution that satisfies a certain accuracy threshold for match to the target hazard (or target IM) unless this solution leads to a greater modification for the predictive model than the Pareto optimal solution with minimum distance from the Utopia point, in which case select the latter solution.

The HC approach was also compared with the IMC approach, where modification of only the mean predictive relationships to match the corresponding hazard for these mean predictions is performed. It was shown that the latter may provide an adequate surrogate for seismicity scenarios with high initial discrepancy to the target hazard, though overall, it is better to explicitly consider the impact of the variability in the predictive models and simultaneously modify the entire predictive model.

7.1.3 Validation of the proposed stochastic ground motion model modification framework

A validation study of the proposed modification framework (IMC or HC implementation) was performed by comparing the seismic demand of recorded ground motions to the demand of stochastic ground motion models established through the proposed modification. Suites of recorded and stochastic ground motions, whose spectral acceleration statistics match the mean and variance of target spectra within a period range of interest, were utilised as input to perform response-history analysis of inelastic SDoF systems with different degrees of inelastic behaviour, corresponding to different values of strength reduction factor R_μ, and different nonlinear and hysteretic characteristics. A peak-oriented hysteresis model was chosen with different values for post-yield stiffness, considering non-degrading (EPH system) or degrading (ESD system) strength characteristics. For determining the elastic base shear demand two approaches were adopted: a "constant-R_μ" approach, where the yielding strength of the structure varies from record to record, and a "constant-strength" approach, where the structural yielding strength is constant. Comparison was performed for two different EDPs, peak displacement and hysteretic energy that have been used in past validation studies. Different seismicity scenarios were examined, corresponding to different values of moment magnitude and rupture distance for seismic events. The recorded ground motions were obtained through the REXEL software while for the modified stochastic ground motion model, different modification degrees were examined.

Results showed that the proposed modification improves significantly the match to the reference results corresponding to the recorded ground motion model, especially for larger value or R_μ or for ESD system (compared to EPH system). For large degrees of modification of the stochastic ground motion model, larger errors may exist for such instances of significant inelastic behaviour. The moderate modification approach proposed in this study (discussed in the last paragraph of the previous section) appears to consistently yield good results across all seismicity

scenarios and types of inelastic behaviour. Trends were similar for both considered EDPs. With respect to the two types of modifications examined, IMC and HC, while both match the median statistics similarly well, HC was shown to provide an enhanced match to the target dispersion, with IMC yielding small dispersion values.

7.2 Impact

This work contributes to better seismic risk and resilience assessment prediction, that can be used in turn for different purposes. The impact of this study can be seen in societal, economic and academic terms.

With respect to the societal impact, results of this study may ultimately be used by public and private organisations to develop emergency response plans, evaluate cost-effective seismic retrofitting actions and risk mitigation strategies, and evaluate rapid damage-estimation algorithms for effective disaster response. This study is a contribution toward reduction of economic, life and business interruption losses due to earthquakes through advanced, improved, earthquake risk assessment models and tools.

With respect to the economic impact, the outputs of this study may provide the insurance and re-insurance industry, providers of basic services (e.g., civil protection and emergency managers), as well as multidisciplinary consultancy firms with an improved characterisation of seismic risk to enable them to operate more competitively in the global market.

Finally, this study addresses major intellectual challenges by going beyond the state-of-the-art of ground motion simulation and validation, and the academic impact is summarised below for each contribution.

7.2.1 Proposed validation approach

In the first part of this study, a novel ground motion simulation validation approach has been proposed based on the relative entropy D_{KL} coupled with statistical hypothesis testing. Although the proposed approach was applied to ground motion simulations of historical earthquakes (following the historical method), it is generic enough to be applied to cases where the empirical or similar spectra methods are utilised. In fact, the proposed spectral-shape and duration-related validation metrics are hazard computable and their empirical models can be used as baseline for comparison for future earthquake scenarios as in the empirical method. The proposed approach can readily be used to measure the similarity of the distributions of seismic response to sets of simulations and recordings matching a target (elastic) response spectrum mean and variance (similar spectra approach). It is worth noting that the validation metrics used in this study are intended as a supplement, not a replacement to other proposed validation metrics. Aligned with the broader objectives of the SCEC GMSV TAG, the proposed validation approach expands the efforts of the group (e.g., Galasso et al., 2012, 2013; Burks and Baker, 2014; Burks et al., 2015; Rezaeian et al., 2015) toward the engineering validation of ground motion simulations obtained through BBP. It is thus, expected to be of particular interest to SCEC GMSV TAG or any group that works on GMSV, ground motion simulation developers and seismologists.

7.2.2 Proposed stochastic ground motion model modification framework

The proposed stochastic ground motion model modification framework (extended and simplified implementation) is highly versatile as it facilitates a match to any desired IM (defined through any GMPEs, or any other approach that would define a target IM description) or to a collection of them, e.g., spectral accelerations over

a period range, for any chosen seismicity scenario. It can also achieve compatibility to any target spectrum for ground motion selection purposes, e.g., uniform hazard spectrum derived through PSHA or conditional (mean) spectrum since these are associated with a single seismicity scenario. Repeating process for different seismicity scenarios can then facilitate the development of a suite of models that can support comprehensive seismic risk assessment. Although the framework was demonstrated using a specific record-based stochastic ground motion model, the approach can be applied to any stochastic ground motion model as long as a link between the probability distributions of ground motion model parameters and earthquake and site characteristics through predictive relationships is established. The proposed framework (extended or simplified) can provide significant improvement for the match to the target hazard or IM, respectively for seismicity ranges where the unmodified model faces challenges in matching the target hazard/IMs, with minor only modifications to the original predictive model, something that can guarantee a good agreement with observed regional trends. Therefore, the developed framework can provide a useful tool for engineering users to generate hazard or spectrum-compatible simulated ground motions for seismic risk and loss assessment applications following the approach shown in Figure 1.3. More generally, the developed framework can provide the basis for addressing open problems in risk assessment for different natural hazards.

7.2.3 Validation of the proposed stochastic ground motion model modification framework

The validation study highlighted the impact of the proposed modification framework on the SDoF system response. The proposed modification provides an enhanced match to the EDP statistics estimated through the benchmark recorded ground motion case compared to the unmodified stochastic ground motion that can exceed relative errors of 100% for some scenarios and R_μ factors. These errors

are even higher and can reach values of 200% when the "constant-strength" modeling reference is utilised for the SDoF system, that is the most common modeling option for practical applications. When this modeling reference is adopted, the unmodified model fails to cause the SDoF system to move into the inelastic region for some scenarios and slightly inelastic behaviour (R_μ equal to 2). The modified model also leads to some discrepancies for large degrees of modification factor and significant inelastic behaviour, however, the adoption of a moderate modification approach has good behaviour overall and is the proposed approach.

This study showed that stochastic ground motion models are a promising alternative to ground motion records that can be used with confidence by practicing engineers, if appropriate modifications to match target hazard are performed. In accordance with the objectives of this research, the modified stochastic ground motion models can be used with confidence to replace the most commonly used approach of scaling ground motion records for seismic risk and loss assessment of a single structure or portfolio of structures as shown in Figure 1.3.

7.3 Limitations and future work

7.3.1 Proposed validation approach

In the proposed validation approach we used parameters that are used as proxies for peak inelastic and cyclic response of more complicated structural systems. These parameters have robust empirical models and can be used in an empirical validation approach. This list of proposed validation proxies is though, not extensive and it would be necessary to include other metrics to fully assess the simulation methods' ability to produce reasonable ground motions as a whole. For example, significant duration is important for many geotechnical engineering applications (Afshari and Stewart, 2016) and spatial correlation of IMs (Jayaram, Park, Bazzurro and Tothong, 2010) is important for distributed systems or building

portfolios. Furthermore, validation using metrics that relate to more complicated structural behaviour such as response amplification due to basin effects, landslide displacement and financial losses is necessary.

The ground motion simulations obtained through BBP v13.5 and 13.6 did not include site effects and that created challenges for the validation, since the records from the historical events contain local site effects. To overcome this problem, a large number of ground motions that were recorded on soft soil sites and their corresponding simulations were eliminated from the validation. This significantly decreased that sample size and could influence the outcome of the validation. Site response is currently an active topic of research and the advances in this area are expected to improve the simulation methods implemented on the BBP.

Finally, the aim of this study was not to provide a definite judgment about the predictive capabilities of the considered ground motion simulation methods, but to develop a validation framework that could be used in future validation exercises to promote their use by the engineering community. The proposed validation framework is recommended to be used as a supplement to other existing validation procedures.

7.3.2 Proposed stochastic ground motion model modification framework

The proposed stochastic ground motion model modification framework is very efficient when modification of the mean predictive relationships to match target IMs ignoring variability is considered. An adequate representation of the overall Pareto front can be obtained in as little as two minutes, whereas this can be more efficient when only a single solution from the Pareto front is needed. However, the extended implementation of the framework where variability is also considered, comes with significant computational burden for performing the multi-objective optimisation to identify the Pareto front, a burden stemming from the MCS step.

Considering the fact that the proposed modification needs to be repeated for each seismicity scenario of interest, further reduction of this burden is an important extension of this work. This reduction will have to come from a more computationally efficient implementation of the surrogate model predictions. Once an improved efficiency implementation is established, the proposed framework can be applied in an automated standalone tool that will enable non-technical end-users to generate hazard-compatible stochastic ground motions for use in PBEE or probabilistic seismic risk assessment.

Another limitation of the proposed framework is that it provides a modified ground motion model for a single scenario at a time. This is because a single Pareto front is estimated for each specific seismicity scenario. Therefore, for multiple scenarios (as in PSHA) the modification needs to be performed for each different scenario. Given the significant computational cost for estimating the Pareto fronts for multiple scenarios when variability is also considered, it would be very useful to investigate possible adjustment to the developed framework to facilitate a match to a selection of scenarios at the same time.

7.3.3 Validation of the proposed stochastic ground motion model modification framework

The validation of the proposed modification was performed in terms of inelastic response of SDoF systems. Although different ranges of inelastic behaviour and different nonlinear and hysteretic characteristics were examined, the structural systems, EDPs and fundamental periods considered in this study correspond to first-mode-dominated and moderate height buildings. Therefore, a study considering different period ranges and MDoF systems is an extension of this work that will highlight the impact of the modification on other structural systems.

In this study the match to the target spectrum was established for a period range $0.2T_1$-$1.5T_1$ in accordance with ASCE 7 (ASCE, 2010) provisions. For large values

of R_μ, though, the nonlinear structural response is sensitive to spectral ordinates at periods much larger than the fundamental one and the chosen period range for spectral compatibility may not be conservative. These were the cases where the largest discrepancies for the stochastic ground motion model were observed. Further study extending the spectral compatibility to larger periods would verify this claim.

The validation was performed with respect to two EDPs: the peak inelastic displacement and hysteretic energy that have been used in past validation studies. An important extension of this work is the validation of the proposed stochastic ground motion model modification with respect to history-dependent EDPs, such as the residual deformation, that are important for cumulative damage assessment for a seismic sequence (i.e., aftershocks).

With respect to the ground motion records, this study used 30 ground motion records that was considered a good compromise between accuracy and efficiency. Sensitivity analysis using different number of ground motions per suite would be useful to show the influence of the selected record sample size in the final results. Finally, the dispersion of the recorded ground motions was not explicitly matched to the dispersion of the target spectra, since this is not the current practice, and was in some cases significantly higher or lower than the target. This could have an effect on the discrepancies observed between the benchmark recorded and the stochastic model cases. A validation of the approach where an explicit match to the target dispersion is established for the records may shed light on this matter.

Appendix A

Details for stochastic ground motion model considered in this thesis

The specific ground motion model used in this study is the record-based model proposed by Rezaeian and Der Kiureghian (2010) that efficiently addresses both temporal and spectral nonstationarities. The unfiltered discretised time history of the ground motion according to this model is expressed as

$$\ddot{\underline{q}}(t|\boldsymbol{\theta}, \boldsymbol{w}) = q(t, \boldsymbol{\theta}) \sum_{i=1}^{k} \frac{h[t - t_i, \boldsymbol{\theta}(t_i)]}{\sqrt{\sum_{j=1}^{k} h[t - t_j, \boldsymbol{\theta}(t_j)]^2}} w(i\Delta t); \ k\Delta t < t < (k+1)\Delta t$$

(A.1)

where $\boldsymbol{w} = [w(i\Delta t) : i = 1, 2, \ldots, N_T]$ is a Gaussian white-noise sequence, Δt is the chosen discretisation interval (assumed constant and equal to 0.005s in this study), $q(t, \boldsymbol{\theta})$ is the time-modulating function, and $h[t - \tau, \boldsymbol{\theta}(\tau)]$ an impulse response function corresponding to the pseudo-acceleration response of an SDoF linear oscillator with time varying frequency $\omega_f(\tau)$ and damping ratio $\zeta_f(\tau)$, in

which τ denotes the time of the pulse

$$h[t - \tau, \boldsymbol{\theta}(\tau)] = \frac{\omega_f(\tau)}{\sqrt{1 - \zeta_f^2(\tau)}} \exp\left[-\omega_f(\tau)\zeta_f(\tau)(t - \tau)\right] \sin\left[\omega_f(\tau)\sqrt{1 - \zeta_f^2(\tau)}(t - \tau)\right] ; \ \tau \leq t$$

$$= 0; \ \text{otherwise}$$

$$\text{(A.2)}$$

For the time varying characteristics a linear function has been proposed for the frequency and a constant for the damping

$$\omega_f(\tau) = \omega_{mid} + \omega'(\tau - t_{mid}) \qquad \zeta_f(\tau) = \zeta_f \qquad \text{(A.3)}$$

with ω_{mid} (central frequency), ω' (frequency variation), and ζ_f ultimately corresponding to model parameters for the filter and t_{mid} corresponding to the mid-time of the strong motion duration (defined next). The time envelope $q(t, \boldsymbol{\theta})$ is given by Rezaeian and Der Kiureghian (2010)

$$q(t, I_a, \alpha_2, \alpha_3) = \sqrt{I_a} \left[\sqrt{\frac{(2\alpha_3)^{2\alpha_2-1}}{\Gamma(2\alpha_2 - 1)}}\right] t^{\alpha_2-1} \exp(-\alpha_3 t) \qquad \text{(A.4)}$$

where $\Gamma(.)$ is the gamma function, I_a is the Arias intensity expressed in terms of $g \cdot \pi/2$ (i.e., scaled by that term), and $\{\alpha_2, \alpha_3\}$ are additional parameters controlling the shape and total duration of the envelope that can be related to physical parameters. As advocated by Rezaeian and Der Kiureghian (2010), the strong motion duration D_{5-95} (defined as the duration for the Arias intensity to increase from 5% to 95% of its final value) and t_{mid} corresponding to the time Arias intensity achieves 45% of its final value are used. The pair $\{\alpha_2, \alpha_3\}$ can be then easily determined based on the values of $\{D_{5-95}, t_{mid}\}$ (Rezaeian and Der Kiureghian, 2010). To assure zero residual velocity, the simulated process is eventually high-pass filtered (Rezaeian and Der Kiureghian, 2008). This filter corresponds to a

critically damped oscillator, and the corrected acceleration record is obtained as the solution of the differential equation

$$\ddot{\alpha}(t|\boldsymbol{\theta},\boldsymbol{w}) + 2\omega_c\dot{\alpha}(t|\boldsymbol{\theta},\boldsymbol{w}) + \omega_c^2\alpha(t|\boldsymbol{\theta},\boldsymbol{w}) = \ddot{\underset{\sim}{\alpha}}(t|\boldsymbol{\theta},\boldsymbol{w}); \; \omega_c = 0.5\pi \qquad (A.5)$$

This filter has minimal effect on the response beyond the corner frequency, ω_c (Rezaeian and Der Kiureghian, 2008).

The ground motion model has model parameters $\boldsymbol{\theta} = \{I_a, D_{5-95}, t_{mid}, \omega_{mid}, \omega', \zeta_f\}$, with the first one directly affecting (scaling) the output (thus input-output relationship is known) and the remaining five having a complex nonlinear relationship to that output. Predictive relationships have been established for $\boldsymbol{\theta}$ by fitting the stochastic model to a subset of the NGA relationships strong motion database (Rezaeian and Der Kiureghian, 2010). These predictive relationships relate $\boldsymbol{\theta}$ to the following earthquake and site characteristics, defining seismicity vector $\boldsymbol{z}$: the moment magnitude, M, the rupture distance, R, the type of fault F (F=0 denoting strike slip and F=1 reverse fault) and the shear wave velocity of the top 30m of the site soil, $V_{s,30}$. The predictive relationships were established by first transforming the initial parameters in the standard Gaussian space leading to vector $\boldsymbol{v}$ with components $v_i, i = 1, \ldots, 6$, each of which is related to a component θ_i of vector $\boldsymbol{\theta}$ through

$$v_i(\theta_i) = \Phi^{-1}[F_{\theta_i}(\theta_i)]; \; i = 1, \ldots, 6 \qquad (A.6)$$

where Φ corresponds to the standard Gaussian CDF and F_{θ_i} corresponds to the CDF for the fitted probability distribution to the ith component of vector $\boldsymbol{\theta}$ (Rezaeian and Der Kiureghian, 2008), given in Table A.1 where the two-sided exponential referenced in this table is

$$p_{\omega'/2\pi}(\omega'/2\pi) = \begin{cases} 4.85 \exp(6.77\omega'/2\pi) & -2 < \omega'/2\pi < 0 \\ 4.85 \exp(-17.10\omega'/2\pi) & 0 < \omega'/2\pi < 0.5 \\ 0 & \text{otherwise} \end{cases} \qquad (A.7)$$

TABLE A.1: Fitted probability distributions

Parameter	Fitted distribution	Distribution bounds	Mean	Standard deviation
I_a (sg)	Lognormal	$(0, \infty)$	0.0468	0.164
D_{5-95} (s)	Beta	$[5, 45]$	17.3	9.31
t_{mid} (s)	Beta	$[0.5, 40]$	12.5	7.44
$\omega_{mid}/2\pi$ (Hz)	Gamma	$(0, \infty)$	5.87	3.11
$\omega'/2\pi$ (Hz/s)	Two-sided exponential	$[-2, 0.5]$	-0.0892	0.185
ζ_f	Beta	$[0.02, 1]$	0.213	0.143

The predictive relationships for $\boldsymbol{v}$ are then obtained through regression analysis, providing a probabilistic characterisation with mean vector $\boldsymbol{\mu}(\boldsymbol{z})$ having components (Rezaeian and Der Kiureghian, 2010)

$$\mu_1(\boldsymbol{z}) = c_{1,0} + c_{1,1}F + c_{1,2}\left(\frac{M}{7}\right) + c_{1,3}\ln\left(\frac{R}{25km}\right) + c_{1,4}\ln\left(\frac{V_s}{750m/s}\right)$$
$$\mu_i(\boldsymbol{z}) = c_{i,0} + c_{i,1}F + c_{i,2}\left(\frac{M}{7}\right) + c_{i,3}\left(\frac{R}{25km}\right) + c_{i,4}\left(\frac{V_s}{750m/s}\right) ; \ i = 2,\ldots,6$$
$$(A.8)$$

where $c_{i,j}$ are the regression coefficients provided in Table A.2, and covariance matrix $\boldsymbol{\Sigma}$ corresponding to standard deviation σ_{vi} and correlation coefficient ρ_{ij} for each component that are also shown in Table A.2.

TABLE A.2: Regression coefficients for mean predictive relationships and total
standard deviation and correlation coefficients for covariance matrix

i	$c_{i,0}$	$c_{i,1}$	$c_{i,2}$	$c_{i,3}$	$c_{i,4}$	σ_{vi}	ρ_{i1}	ρ_{i2}	ρ_{i3}	ρ_{i4}	ρ_{i5}	ρ_{i6}
1	-1.844	-0.071	2.944	-1.356	-0.265	0.654	1	-0.36	0.01	-0.15	0.13	-0.01
2	-6.195	-0.703	6.792	0.219	-0.523	0.730		1	0.67	-0.13	-0.16	-0.20
3	-5.011	-0.345	4.638	0.348	-0.185	0.658			1	-0.28	-0.20	-0.22
4	2.253	-0.081	-1.810	-0.211	0.012	1.001		Symmetric		1	-0.20	0.28
5	-2.489	0.044	2.408	0.065	-0.081	0.962					1	-0.01
6	-0.258	-0.477	0.905	-0.289	0.316	1.021						1

Appendix B

Kernel-based approximation of the probability distribution of the ln(IM) for the ground motion model and entropy estimation

This Appendix considers the estimation of $p_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and of objective F_{p1} when the lognormal assumption is not invoked for the distribution of Y_i^g. Approach relies on obtaining samples from $p_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, which when using the actual ground motion model is established through the following process. First generate n_d samples for $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\{\boldsymbol{\theta}^d; d = 1, \ldots, n_d\}$, and n_d sample white-noise sequences, $\{\boldsymbol{w}^d; d = 1, \ldots, n_d\}$, and obtain the corresponding acceleration time-histories $\ddot{\alpha}(t|\boldsymbol{\theta}^d, \boldsymbol{w}^d); d = 1, \ldots, n_d$. For each sample, the response output of interest is estimated providing samples of $\{\ln(Y_i^{g,d}); d = 1, \ldots, n_d\}$ from $p_g(\ln(Y_i)|\boldsymbol{z})$. The latter can be approximated using KDE based on these samples as

$$\tilde{p}_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{n_d} \sum_{d=1}^{n_d} \frac{1}{h} K\left(\frac{\ln(Y_i) - \ln(Y_i^{g,d})}{h}\right);$$

$$K(t) = \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}}$$

(B.1)

where $K(.)$ is the chosen kernel and h is the Kernel bandwidth. In this study, the widely used Gaussian kernel is employed, shown also in Equation B.1 with bandwidth chosen as (Scott, 2015) $h = 1.06 \cdot n_d^{-1/5} \sigma_d$ where σ_d is the standard deviation of the samples $\{\ln(Y_i^{g,d}); d = 1, \ldots, n_d\}$. The entropy in Equation 5.6 can be then approximated using the KDE estimate in Equation B.1, with the scalar integral calculated through numerical integration (trapezoidal rule).

When the IM is approximated through use of the metamodel $Y_i^g(\boldsymbol{\theta}, \boldsymbol{w}) = \sqrt{\theta_s} s_i(\boldsymbol{x}, \boldsymbol{w})$, then the approach for obtaining the samples $\{\ln(Y_i^{g,d}); d = 1, \ldots, n_d\}$ changes and requires, additionally, another statistical assumption for the distribution of $Y_i^g(\boldsymbol{\theta}, \boldsymbol{w})$ under the influence of $\boldsymbol{w}$. Note that this is a different setting than invoking a specific distribution for Y_i^g, since in this case the assumption pertains only to the influence of the white-noise. The standard statistical approximation, with proven accuracy in a number of studies (Gidaris et al., 2015), is lognormal assumption. In this case, the desired samples are obtained through the following process. First generate n_d samples for $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\{\boldsymbol{\theta}^d; d = 1, \ldots, n_d\}$, and estimate through the metamodel outputs $\ln(\bar{s}_i(\boldsymbol{x}^d))$ and $\sigma_i^s(\boldsymbol{x}^d)$ for $d = 1, \ldots, n_d$. Generate n_d samples $\{e^d; d = 1, \ldots, n_d\}$ from a standard Gaussian distribution and obtain each of the desired samples as

$$\ln(Y_i^{g,d}) = \ln(\bar{s}_i(\boldsymbol{x}^d)) + \sigma_i^s(\boldsymbol{x}^d)e^d + \frac{1}{2}\ln\theta_s^d$$

(B.2)

Once these samples are obtained, then the KDE estimation of $p_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and the calculation of F_{p1} follow same approach as in the case that actual ground motion model was utilised.

Appendix C

Additional stochastic ground motion model data for Chapters 4 and 5

TABLE C.1: Physical ground motion model parameters $\boldsymbol{\theta}$ for the unmodified U_n and modified ground motion models U_t and C_s

Scenario	PF point	$I_a \frac{2}{\pi}$ (sg)	D_{5-95} (s)	t_{mid} (s)	$\frac{\omega_{mid}}{2\pi}$ (Hz)	$\frac{\omega'}{2\pi}$ (Hz/s)	ζ_f
M=6.2	U_t	0.0116	12.493	6.996	6.080	-0.1071	0.1815
–	U_n	0.0165	10.876	6.375	6.627	-0.1026	0.2696
R=30km	C_s	0.0043	20.122	10.791	3.560	-0.2695	0.0338
M=6.2	U_t	0.0031	14.243	8.927	6.023	-0.0915	0.1742
–	U_n	0.0043	13.209	9.188	5.803	-0.0909	0.2102
R=60km	C_s	0.0047	15.977	10.355	6.590	-0.4212	0.0541
M=6.2	U_t	0.0016	16.735	11.341	5.171	-0.2238	0.1166
–	U_n	0.0020	15.971	12.568	5.051	-0.0798	0.1591
R=90km	C_s	0.0003	23.828	18.896	2.423	-0.1374	0.1000

M=6.8	U_t	0.0244	16.930	8.520	5.157	-0.2211	0.2018
–	U_n	0.0237	16.615	9.039	6.114	-0.0730	0.2838
R=30km	C_s	0.0347	23.818	16.814	5.415	-0.2608	0.0590
M=6.8	U_t	0.0070	16.359	9.968	5.019	-0.0433	0.2347
–	U_n	0.0062	19.781	12.395	5.334	-0.0629	0.2228
R=60km	C_s	0.0029	32.373	21.587	2.981	-0.1607	0.0571
M=6.8	U_t	0.0040	21.217	15.349	4.625	-0.0437	0.2052
–	U_n	0.0028	23.138	16.164	4.624	-0.0534	0.1697
R=90km	C_s	0.0014	32.370	19.862	2.517	-0.1010	0.1703
M=7.4	U_t	0.0458	22.394	12.179	5.548	-0.0683	0.3200
–	U_n	0.0341	23.871	12.223	5.628	-0.0476	0.2984
R=30km	C_s	0.0378	21.423	10.914	5.524	-0.0353	0.3781
M=7.4	U_t	0.0140	24.339	15.319	4.639	-0.0330	0.2993
–	U_n	0.0089	27.259	15.976	4.891	-0.0390	0.2358
R=60km	C_s	0.0074	27.790	17.408	3.607	-0.2170	0.2355
M=7.4	U_t	0.0076	26.603	18.022	4.110	-0.0215	0.2488
–	U_n	0.0040	30.511	19.935	4.223	-0.0310	0.1809
R=60km	C_s	0.0037	32.826	21.189	3.028	-0.1566	0.2585
M=8	U_t	0.0905	28.314	14.871	5.747	-0.0164	0.4070
–	U_n	0.0489	31.180	15.789	5.168	-0.0261	0.3133
R=30km	C_s	0.0463	33.614	20.359	3.742	-0.1561	0.4215
M=8	U_t	0.0225	32.870	20.521	3.683	-0.0097	0.3130
–	U_n	0.0127	34.100	19.744	4.474	-0.0189	0.2492
R=60km	C_s	0.0151	37.748	24.451	2.766	-0.1522	0.2754
M=8	U_t	0.0068	38.688	27.440	2.221	-0.0821	0.1623
–	U_n	0.0058	36.654	23.672	3.846	-0.0123	0.1924
R=90km	C_s	0.0078	38.687	26.624	2.241	-0.1096	0.2583

TABLE C.2: Spectral acceleration values (PSA) for the unmodified U_n, modified ground motion models U_t and C_s and target IMs

Scenario	Case	PSA (g)					
		T=0.4 s	T=0.5 s	T=0.75 s	T=1 s	T=1.5 s	T=2 s
M=6.2- R=30km	U_t	0.1227	0.0996	0.0695	0.0549	0.0382	0.0286
	U_n	0.1724	0.1400	0.1001	0.0780	0.0549	0.0419
	C_s	0.1357	0.1102	0.0693	0.0501	0.0309	0.0226
	Target	0.1399	0.1135	0.0745	0.0533	0.0315	0.0206
M=6.2- R=60km	U_t	0.0608	0.0492	0.0338	0.0270	0.0186	0.0142
	U_n	0.0817	0.0655	0.0451	0.0359	0.0249	0.0191
	C_s	0.0714	0.0543	0.0372	0.0262	0.0159	0.0117
	Target	0.0701	0.0577	0.0388	0.0279	0.0166	0.0109
M=6.2- R=90km	U_t	0.0502	0.0375	0.0228	0.0177	0.0113	0.0085
	U_n	0.0527	0.0414	0.0276	0.0220	0.0153	0.0121
	C_s	0.0460	0.0400	0.0257	0.0184	0.0113	0.0083
	Target	0.0461	0.0385	0.0264	0.0191	0.0115	0.0076
M=6.8- R=30km	U_t	0.2207	0.1710	0.1132	0.0879	0.0591	0.0451
	U_n	0.1963	0.1607	0.1118	0.0893	0.0625	0.0489
	C_s	0.2288	0.1800	0.1204	0.0901	0.0571	0.0429
	Target	0.2136	0.1791	0.1246	0.0932	0.0586	0.0409
M=6.8- R=60km	U_t	0.1186	0.0934	0.0635	0.0496	0.0343	0.0270
	U_n	0.0951	0.0764	0.0519	0.0418	0.0295	0.0236
	C_s	0.1163	0.0985	0.0666	0.0498	0.0333	0.0250
	Target	0.1150	0.0980	0.0699	0.0526	0.0333	0.0233
M=6.8- R=90km	U_t	0.0830	0.0646	0.0436	0.0347	0.0242	0.0191
	U_n	0.0615	0.0471	0.0318	0.0258	0.0182	0.0144
	C_s	0.0791	0.0699	0.0472	0.0347	0.0232	0.0179
	Target	0.0787	0.0681	0.0495	0.0376	0.0240	0.0168

	U_t	0.2742	0.2265	0.1576	0.1288	0.0911	0.0723
M=7.4-	U_n	0.2241	0.1826	0.1270	0.1026	0.0733	0.0585
R=30km	C_s	0.2727	0.2247	0.1581	0.1272	0.0903	0.0718
	Target	0.2584	0.2255	0.1709	0.1349	0.0923	0.0683
	U_t	0.1661	0.1339	0.0919	0.0728	0.0516	0.0414
M=7.4-	U_n	0.1119	0.0883	0.0612	0.0492	0.0343	0.0277
R=60km	C_s	0.1563	0.1316	0.0966	0.0769	0.0555	0.0447
	Target	0.1490	0.1323	0.1029	0.0819	0.0564	0.0418
	U_t	0.1276	0.0982	0.0665	0.0521	0.0363	0.0295
M=7.4-	U_n	0.0792	0.0585	0.0389	0.0301	0.0212	0.0169
R=60km	C_s	0.1119	0.0968	0.0717	0.0568	0.0407	0.0329
	Target	0.1063	0.0959	0.0762	0.0611	0.0423	0.0314
	U_t	0.3768	0.3138	0.2247	0.1814	0.1305	0.1063
M=8-	U_n	0.2670	0.2164	0.1516	0.1222	0.0873	0.0707
R=30km	C_s	0.3523	0.3012	0.2281	0.1857	0.1364	0.1100
	Target	0.3304	0.2970	0.2402	0.1978	0.1449	0.1124
	U_t	0.2390	0.1863	0.1265	0.1011	0.0703	0.0563
M=8-	U_n	0.1387	0.1068	0.0730	0.0582	0.0418	0.0338
R=60km	C_s	0.2176	0.1985	0.1511	0.1218	0.0914	0.0731
	Target	0.2054	0.1881	0.1560	0.1297	0.0953	0.0741
	U_t	0.1768	0.1573	0.1142	0.0843	0.0580	0.0437
M=8-	U_n	0.0996	0.0713	0.0479	0.0371	0.0267	0.0213
R=90km	C_s	0.1620	0.1492	0.1176	0.0938	0.0715	0.0558
	Target	0.1537	0.1429	0.1208	0.1012	0.0747	0.0581

TABLE C.3: Mean ground motion model parameters $\boldsymbol{\theta}$ for the unmodified U_n (corresponding to $\boldsymbol{\mu}_r$) and modified predictive model U_t and C_s (corresponding to $\boldsymbol{\mu}$ for each Pareto point)

Scenario	PF point	$\boldsymbol{\theta}$					
		$I_a \frac{2}{\pi}$ (sg)	D_{5-95} (s)	t_{mid} (s)	$\frac{\omega_{mid}}{2\pi}$ (Hz)	$\frac{\omega'}{2\pi}$ (Hz/s)	ζ_f
M=6.2	U_t	0.0094	13.489	6.410	7.465	-0.1137	0.2560
–	U_n	0.0165	10.876	6.375	6.627	-0.1026	0.2696
R=30km	C_s	0.0059	8.405	3.798	3.803	-0.1424	0.0624
M=6.2	U_t	0.0030	15.323	9.377	6.281	-0.0929	0.1747
–	U_n	0.0043	13.209	9.188	5.803	-0.0909	0.2102
R=60km	C_s	0.0015	10.632	8.465	3.555	-0.2035	0.0507
M=6.2	U_t	0.0014	18.362	12.787	5.487	-0.0817	0.1297
–	U_n	0.0020	15.971	12.568	5.051	-0.0798	0.1591
R=90km	C_s	0.0004	12.206	12.895	2.999	-0.0461	0.1760
M=6.8	U_t	0.0190	13.823	7.745	5.191	-0.0648	0.2151
–	U_n	0.0237	16.615	9.039	6.114	-0.0730	0.2838
R=30km	C_s	0.0192	17.638	10.306	3.898	-0.0365	0.1342
M=6.8	U_t	0.0050	18.068	11.847	3.928	-0.0531	0.1595
–	U_n	0.0062	19.781	12.395	5.334	-0.0629	0.2228
R=60km	C_s	0.0052	17.894	11.513	4.115	-0.0410	0.1604
M=6.8	U_t	0.0029	21.312	15.032	4.089	-0.0312	0.1767
–	U_n	0.0028	23.138	16.164	4.624	-0.0534	0.1697
R=90km	C_s	0.0026	21.502	16.437	3.901	-0.0439	0.1650
M=7.4	U_t	0.0315	23.987	12.909	4.734	-0.0360	0.3002
–	U_n	0.0341	23.871	12.223	5.628	-0.0476	0.2984
R=30km	C_s	0.0357	23.346	11.980	5.474	-0.0440	0.3080
M=7.4	U_t	0.0107	27.604	16.894	3.831	-0.0351	0.2126
–	U_n	0.0089	27.259	15.976	4.891	-0.0390	0.2358
R=60km	C_s	0.0118	26.338	16.042	4.167	-0.0191	0.2452

M=7.4	U_t	0.0058	29.301	19.199	3.677	-0.0163	0.2014
–	U_n	0.0040	30.511	19.935	4.223	-0.0310	0.1809
R=60km	C_s	0.0047	30.085	20.270	3.181	-0.0549	0.1822
M=8	U_t	0.0652	30.325	15.855	4.416	-0.0084	0.3239
–	U_n	0.0489	31.180	15.789	5.168	-0.0261	0.3133
R=30km	C_s	0.0556	30.221	14.423	4.333	0.0011	0.3448
M=8	U_t	0.0208	32.591	19.806	3.819	-0.0287	0.2993
–	U_n	0.0127	34.100	19.744	4.474	-0.0189	0.2492
R=60km	C_s	0.0208	32.110	19.659	3.207	-0.0098	0.2716
M=8	U_t	0.0095	34.925	23.590	2.712	-0.0039	0.2119
–	U_n	0.0058	36.654	23.672	3.846	-0.0123	0.1924
R=90km	C_s	0.0097	36.969	27.713	2.277	-0.0234	0.1729

TABLE C.4: Coefficient $\boldsymbol{\beta}$ used to derive the covariance matrix of the modified predictive relationship $\boldsymbol{\Sigma}$

Scenario	PF point	$\boldsymbol{\beta}$					
		$I_a\frac{2}{\pi}$ (sg)	D_{5-95} (s)	t_{mid} (s)	$\frac{\omega_{mid}}{2\pi}$ (Hz)	$\frac{\omega'}{2\pi}$ (Hz/s)	ζ_f
M=6.2	U_t	1.0133	1.0293	1.0505	0.9431	0.8963	0.8649
–	U_n	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
R=30km	C_s	0.6987	0.7845	0.4190	1.0239	1.0389	0.6829
M=6.2	U_t	0.9056	0.8564	1.0428	0.8443	1.0528	1.0164
–	U_n	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
R=60km	C_s	0.6019	0.7165	0.5397	0.7593	0.4386	0.8946
M=6.2	U_t	0.9056	0.8564	1.0428	0.8443	1.0528	1.0164
–	U_n	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
R=90km	C_s	0.9914	0.3970	0.8664	0.2094	0.9196	0.9050
M=6.8	U_t	0.7645	1.0768	0.6610	0.7785	0.8907	1.0814
–	U_n	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
R=30km	C_s	0.6115	0.5926	0.6371	0.9442	0.4136	1.0116
M=6.8	U_t	0.6961	0.7438	0.5929	1.0750	0.8104	1.0603
–	U_n	1.000	1.000	1.000	1.000	1.000	1.000
R=60km	C_s	0.6811	0.7792	0.5410	1.0255	0.6465	1.0080
M=6.8	U_t	0.6593	0.7522	1.0674	1.0620	0.6202	0.9653
–	U_n	1.000	1.000	1.000	1.000	1.000	1.000
R=90km	C_s	0.8812	1.0776	0.9140	0.9734	0.8768	1.0211
M=7.4	U_t	0.8865	1.0318	0.8451	1.0616	0.8611	1.0597
–	U_n	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
R=30km	C_s	0.9489	1.0023	1.0523	1.0469	1.0471	1.0435
M=7.4	U_t	0.9329	0.9352	0.8115	1.0272	0.9102	0.9058
–	U_n	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
R=60km	C_s	0.9241	0.8629	0.9744	1.0443	0.8988	0.8458

$M{=}7.4$	U_t	0.9046	0.8715	1.0853	0.9483	1.0187	0.8398
–	U_n	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$R{=}60$km	C_s	0.9900	0.8263	0.8145	1.0272	0.9964	1.0856
$M{=}8$	U_t	0.9241	0.8629	0.9744	1.0443	0.8988	0.8458
–	U_n	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$R{=}30$km	C_s	0.8531	0.8763	1.0636	1.0565	1.0829	0.6929
$M{=}8$	U_t	0.9736	0.8853	0.9369	1.0793	0.9953	1.0858
–	U_n	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$R{=}60$km	C_s	1.0487	0.8416	1.0682	1.0900	0.8855	1.0623
$M{=}8$	U_t	1.0487	0.8416	1.0682	1.0900	0.8855	1.0623
–	U_n	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$R{=}90$km	C_s	0.9003	0.8545	0.9478	1.0128	1.0325	0.6029

TABLE C.5: Median spectral acceleration values (PSA) for the unmodified U_n, modified ground motion models U_t and C_s and target IMs

Scenario	Case	PSA (g)					
		$T=0.4$ s	$T=0.5$ s	$T=0.75$ s	$T=1$ s	$T=1.5$ s	$T=2$ s
$M=6.2$-$R=30$km	U_t	0.1162	0.0945	0.0654	0.0510	0.0363	0.0283
	U_n	0.1791	0.1458	0.0989	0.0763	0.0537	0.0415
	C_s	0.1233	0.1032	0.0660	0.0475	0.0291	0.0213
	Target	0.1310	0.1054	0.0684	0.0491	0.0290	0.0188
$M=6.2$-$R=60$km	U_t	0.0659	0.0525	0.0344	0.0266	0.0184	0.0144
	U_n	0.0921	0.0747	0.0490	0.0377	0.0258	0.0200
	C_s	0.0663	0.0555	0.0340	0.0243	0.0150	0.0110
	Target	0.0650	0.0531	0.0353	0.0255	0.0152	0.0099
$M=6.2$-$R=90$km	U_t	0.0449	0.0355	0.0229	0.0175	0.0118	0.0092
	U_n	0.0619	0.0505	0.0332	0.0251	0.0167	0.0129
	C_s	0.0445	0.0371	0.0232	0.0167	0.0106	0.0079
	Target	0.0425	0.0353	0.0240	0.0175	0.0106	0.0069
$M=6.8$-$R=30$km	U_t	0.2079	0.1687	0.1103	0.0839	0.0576	0.0444
	U_n	0.2130	0.1743	0.1184	0.0925	0.0649	0.0509
	C_s	0.2006	0.1734	0.1127	0.0841	0.0547	0.0415
	Target	0.2027	0.1686	0.1158	0.0868	0.0548	0.0377
$M=6.8$-$R=60$km	U_t	0.1049	0.0913	0.0632	0.0480	0.0312	0.0237
	U_n	0.1072	0.0884	0.0598	0.0461	0.0317	0.0249
	C_s	0.1043	0.0899	0.0611	0.0463	0.0302	0.0230
	Target	0.1080	0.0913	0.0643	0.0487	0.0310	0.0214
$M=6.8$-$R=90$km	U_t	0.0739	0.0645	0.0446	0.0338	0.0224	0.0174
	U_n	0.0713	0.0600	0.0409	0.0310	0.0208	0.0163
	C_s	0.0739	0.0643	0.0450	0.0340	0.0224	0.0173
	Target	0.0735	0.0631	0.0455	0.0348	0.0223	0.0156

	U_t	0.2523	0.2182	0.1539	0.1199	0.0830	0.0656
M=7.4-	U_n	0.2441	0.2038	0.1417	0.1111	0.0782	0.0622
R=30km	C_s	0.2559	0.2148	0.1497	0.1173	0.0824	0.0655
	Target	0.2471	0.2126	0.1579	0.1240	0.0847	0.0617
	U_t	0.1462	0.1294	0.0936	0.0718	0.0484	0.0380
M=7.4-	U_n	0.1239	0.1048	0.0724	0.0557	0.0384	0.0306
R=60km	C_s	0.1524	0.1340	0.0946	0.0727	0.0497	0.0392
	Target	0.1415	0.1239	0.0944	0.0750	0.0516	0.0378
	U_t	0.1074	0.0940	0.0679	0.0520	0.0351	0.0276
M=7.4-	U_n	0.0824	0.0708	0.0497	0.0381	0.0260	0.0205
R=60km	C_s	0.0977	0.0897	0.0700	0.0542	0.0361	0.0282
	Target	0.1002	0.0892	0.0695	0.0559	0.0388	0.0286
	U_t	0.3517	0.3067	0.2201	0.1733	0.1214	0.0968
M=8-	U_n	0.2839	0.2408	0.1702	0.1339	0.0948	0.0761
R=30km	C_s	0.3325	0.2942	0.2149	0.1699	0.1190	0.0947
	Target	0.3186	0.2811	0.2212	0.1794	0.1295	0.0980
	U_t	0.1996	0.1797	0.1354	0.1075	0.0748	0.0596
M=8-	U_n	0.1448	0.1237	0.0880	0.0688	0.0480	0.0383
R=60km	C_s	0.1984	0.1839	0.1438	0.1137	0.0778	0.0616
	Target	0.1962	0.1764	0.1426	0.1173	0.0852	0.0648
	U_t	0.1358	0.1275	0.1023	0.0806	0.0544	0.0429
M=8-	U_n	0.0973	0.0845	0.0610	0.0471	0.0323	0.0257
R=90km	C_s	0.1393	0.1364	0.1139	0.0917	0.0623	0.0489
	Target	0.1448	0.1323	0.1095	0.0913	0.0669	0.0511

TABLE C.6: Logarithmic standard deviation (σ_{log}) of the spectral acceleration (PSA) for the unmodified U_n, modified ground motion models U_t and C_s and target IMs

Scenario	Case	σ_{log}					
		T=0.4 s	T=0.5 s	T=0.75 s	T=1 s	T=1.5 s	T=2 s
M=6.2- R=30km	U_t	0.6577	0.6509	0.06457	0.6414	0.6380	0.6374
	U_n	0.6925	0.6873	0.6741	0.6686	0.6654	0.6687
	C_s	0.6709	0.7020	0.6952	0.6673	0.6165	0.6064
	Target	0.6322	0.6372	0.6506	0.6549	0.6672	0.6772
M=6.2- R=60km	U_t	0.6560	0.6543	0.6234	0.6250	0.6310	0.6359
	U_n	0.6980	0.7004	0.6611	0.6548	0.6601	0.6627
	C_s	0.6514	0.6917	0.6757	0.6477	0.6165	0.6042
	Target	0.6323	0.6373	0.6506	0.6550	0.6672	0.6772
M=6.2- R=90km	U_t	0.6801	0.6774	0.6143	0.6074	0.6139	0.6192
	U_n	0.6994	0.7086	0.6689	0.6495	0.6457	0.6476
	C_s	0.6322	0.6556	0.6771	0.6516	0.6429	0.6452
	Target	0.6324	0.6373	0.6506	0.6550	0.6672	0.6773
M=6.8- R=30km	U_t	0.6427	0.6504	0.6158	0.6087	0.6155	0.6197
	U_n	0.6789	0.6767	0.6452	0.6409	0.6425	0.6423
	C_s	0.6173	0.6731	0.6642	0.6486	0.6244	0.6118
	Target	0.6134	0.6208	0.6383	0.6458	0.6606	0.6722
M=6.8- R=60km	U_t	0.6110	0.6323	0.6518	0.6391	0.6089	0.5997
	U_n	0.6800	0.6843	0.6452	0.6366	0.6357	0.6362
	C_s	0.6177	0.6506	0.6594	0.6460	0.6194	0.6089
	Target	0.6135	0.6209	0.6384	0.6458	0.6606	0.6723
M=6.8- R=90km	U_t	0.6042	0.6454	0.6779	0.6626	0.6279	0.6158
	U_n	0.6688	0.6844	0.6589	0.6366	0.6187	0.6202
	C_s	0.6363	0.6538	0.6809	0.6575	0.6236	0.6149
	Target	0.6136	0.6209	0.6384	0.6458	0.6606	0.6723

	U_t	0.6193	0.6474	0.6411	0.6336	0.6244	0.6206
M=7.4-	U_n	0.6562	0.6577	0.6272	0.6251	0.6252	0.6264
R=30km	C_s	0.6452	0.6499	0.6218	0.6193	0.6190	0.6190
	Target	0.6071	0.6153	0.6343	0.6427	0.6584	0.6706
	U_t	0.6173	0.6361	0.6662	0.6457	0.6139	0.6096
M=7.4-	U_n	0.6506	0.6680	0.6509	0.6325	0.6225	0.6239
R=60km	C_s	0.6272	0.6602	0.6758	0.6654	0.6470	0.6413
	Target	0.6073	0.6154	0.6343	0.6427	0.6584	0.6706
	U_t	0.6109	0.6350	0.6755	0.6589	0.6312	0.6279
M=7.4-	U_n	0.6552	0.6729	0.6756	0.6502	0.6256	0.6257
R=60km	C_s	0.6381	0.6226	0.6700	0.6522	0.6067	0.5989
	Target	0.6073	0.6155	0.6343	0.6427	0.6584	0.6707
	U_t	0.6157	0.6394	0.6502	0.6416	0.6331	0.6311
M=8-	U_n	0.6405	0.6508	0.6354	0.6236	0.6221	0.6253
R=30km	C_s	0.6037	0.6276	0.6554	0.6519	0.6446	0.6407
	Target	0.6071	0.6153	0.6342	0.6427	0.6584	0.6705
	U_t	0.6033	0.6138	0.6303	0.6213	0.6048	0.6048
M=8-	U_n	0.6398	0.6560	0.6591	0.6412	0.6307	0.6315
R=60km	C_s	0.6096	0.6159	0.6670	0.6699	0.6473	0.6443
	Target	0.6072	0.6153	0.6343	0.6427	0.6584	0.6706
	U_t	0.6090	0.6093	0.6840	0.6891	0.6608	0.6618
M=8-	U_n	0.6288	0.6524	0.6884	0.6673	0.6462	0.6487
R=90km	C_s	0.5973	0.5761	0.6343	0.6459	0.6133	0.6112
	Target	0.6072	0.6154	0.6343	0.6427	0.6584	0.6706

Appendix D

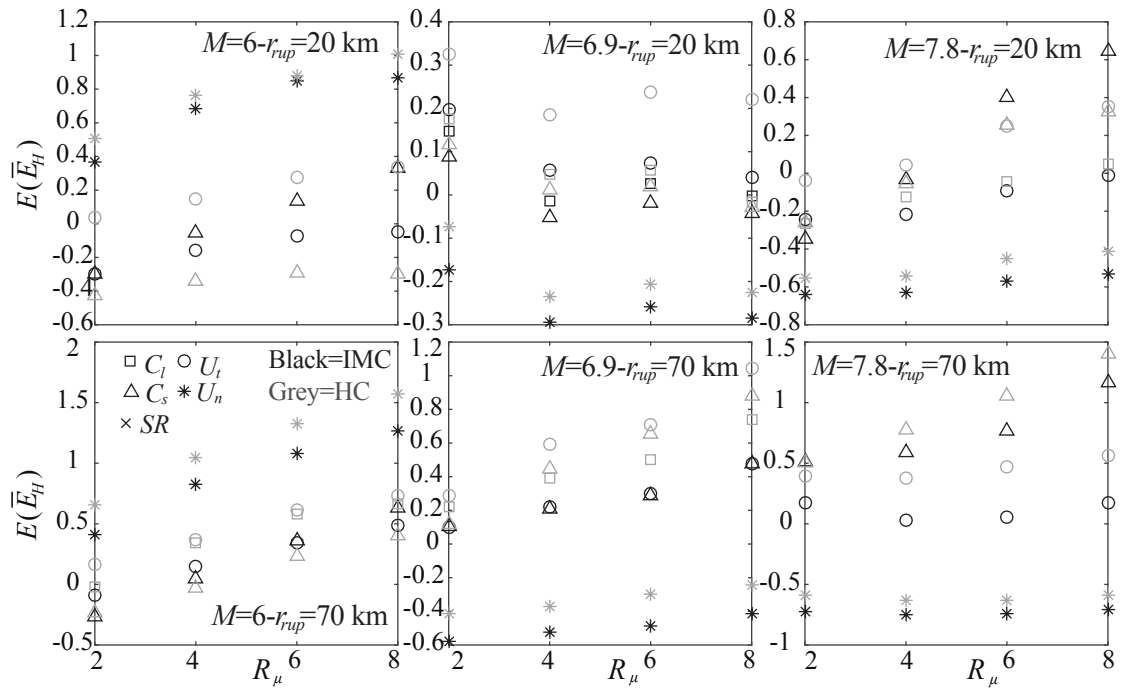# Additional figures for Chapter 6

FIGURE D.1: Relative error compared to reference SR response for the hysteretic energy for EPH system with $\alpha=3\%$ for "constant-R_μ" approach.
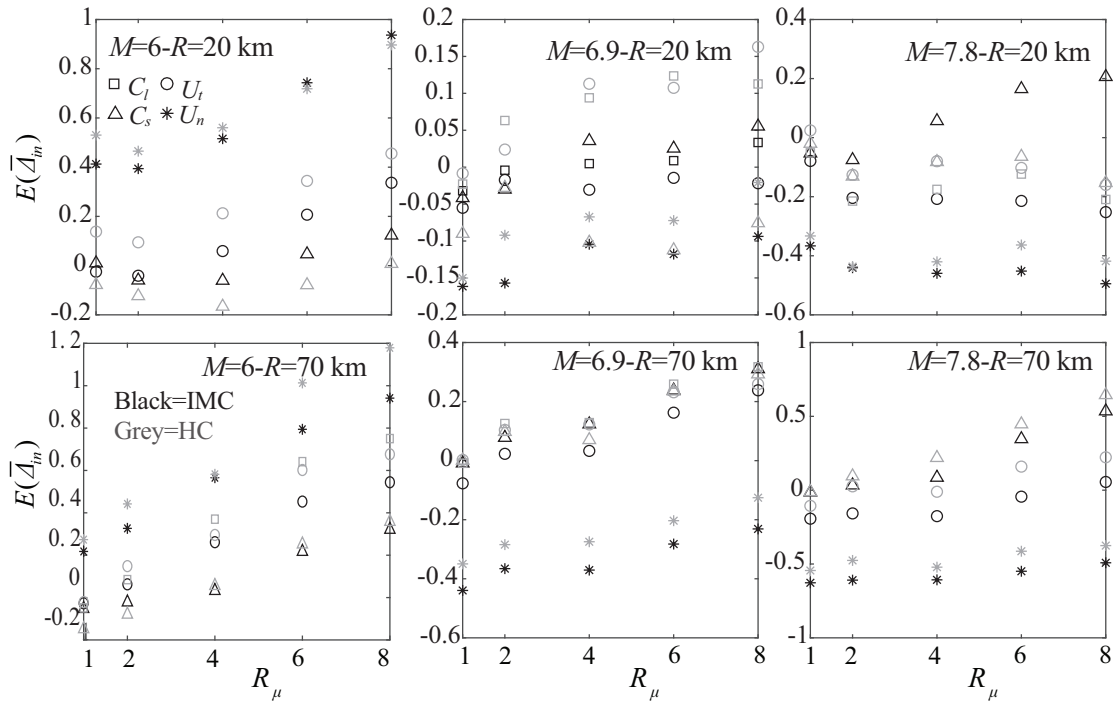
FIGURE D.2: Relative error compared to reference SR response for the peak inelastic displacements for EPH system with α=10% for "constant-R_μ" approach.
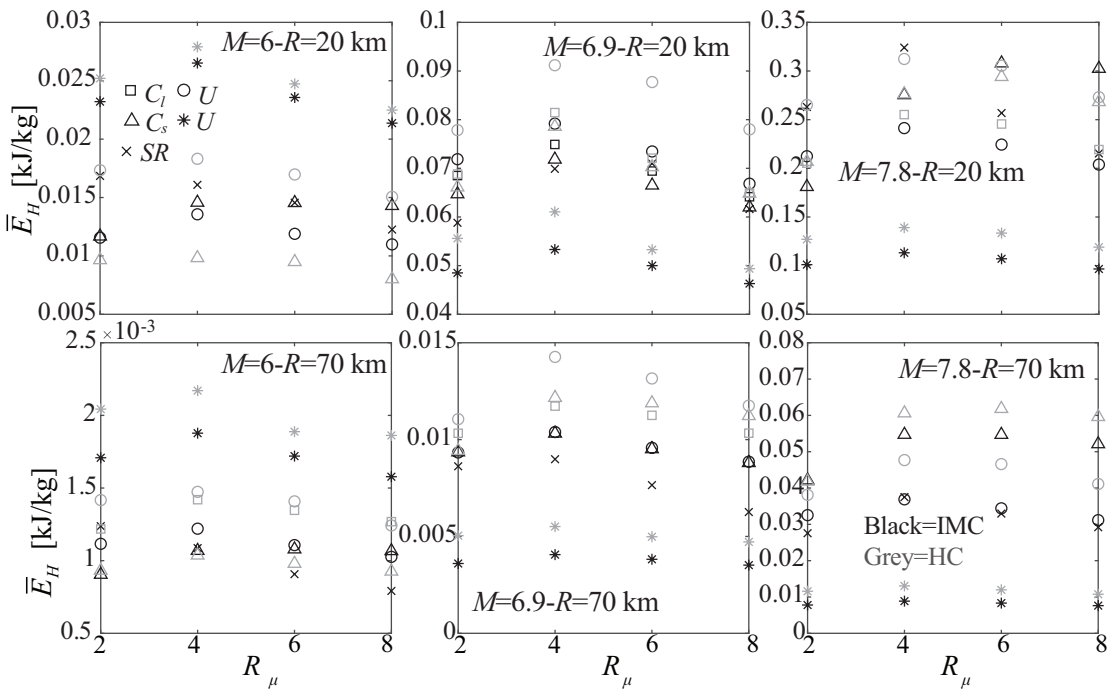

FIGURE D.3: Median hysteretic energy for EPH system with α=10% for "constant-R_μ" approach.
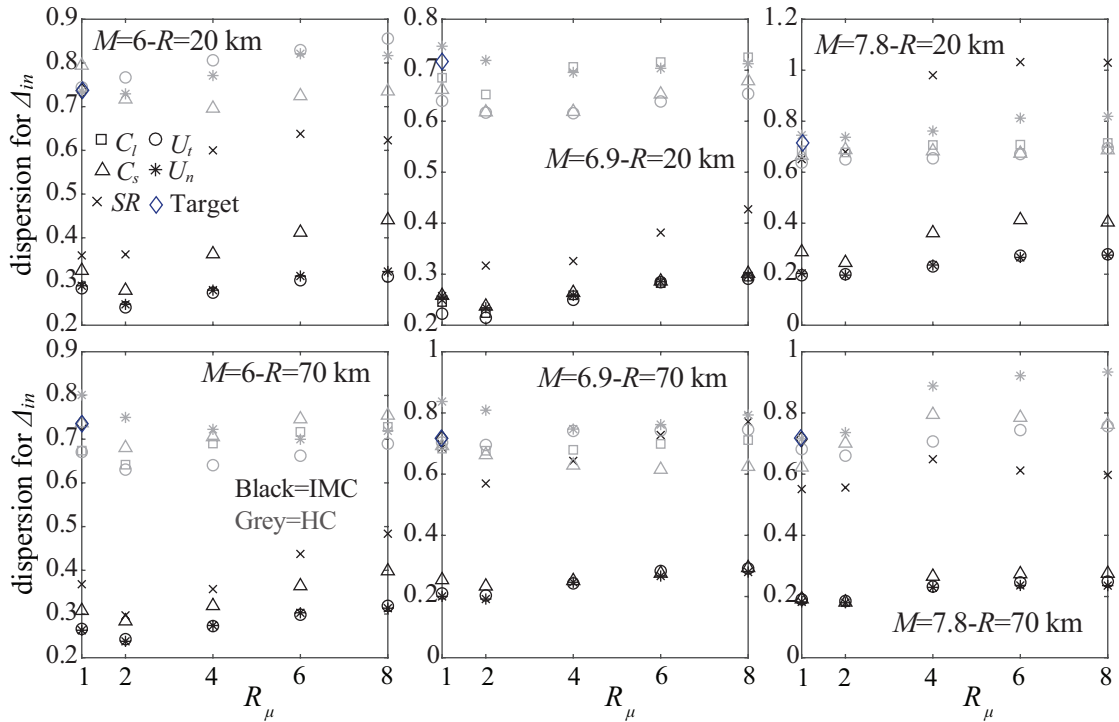
FIGURE D.4: Dispersion (expressed through CoV) of peak inelastic displacement for EPH system with α=10% for "constant-R_μ" approach.
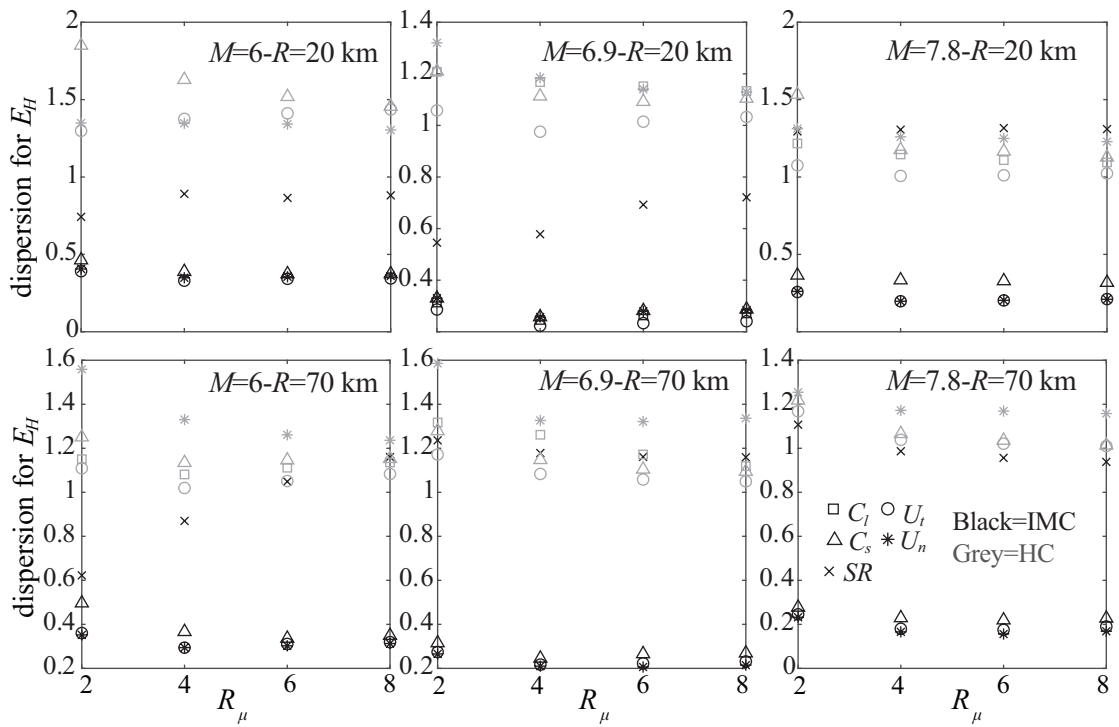

FIGURE D.5: Dispersion (expressed through CoV) of hysteretic energy for EPH system with α=10% for "constant-R_μ" approach.
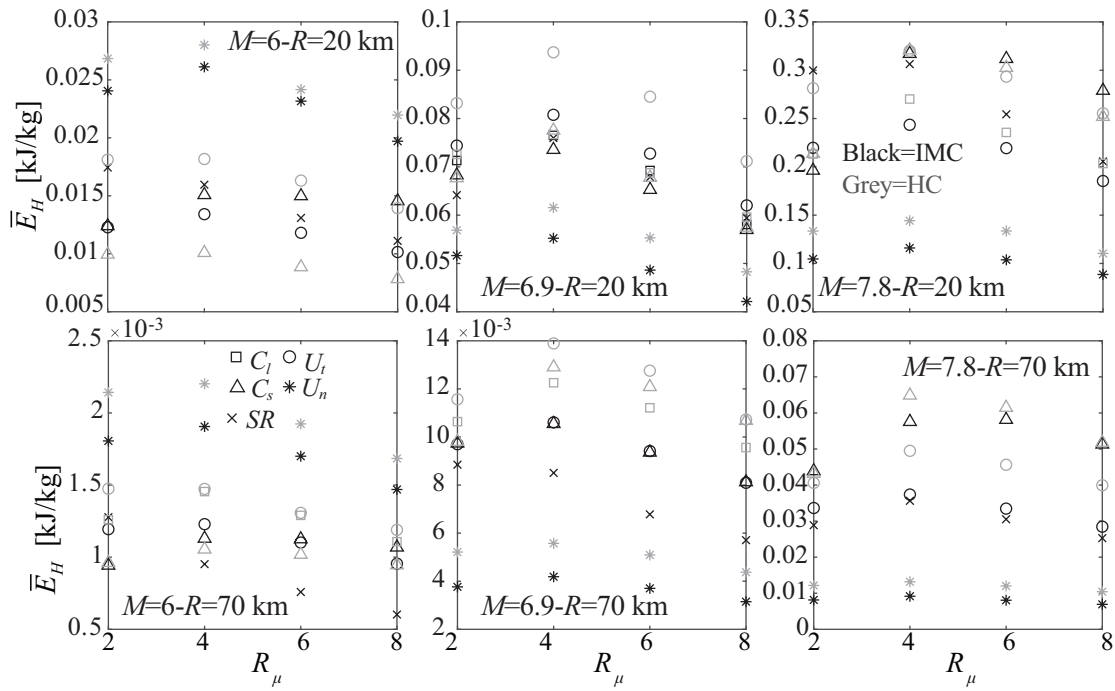
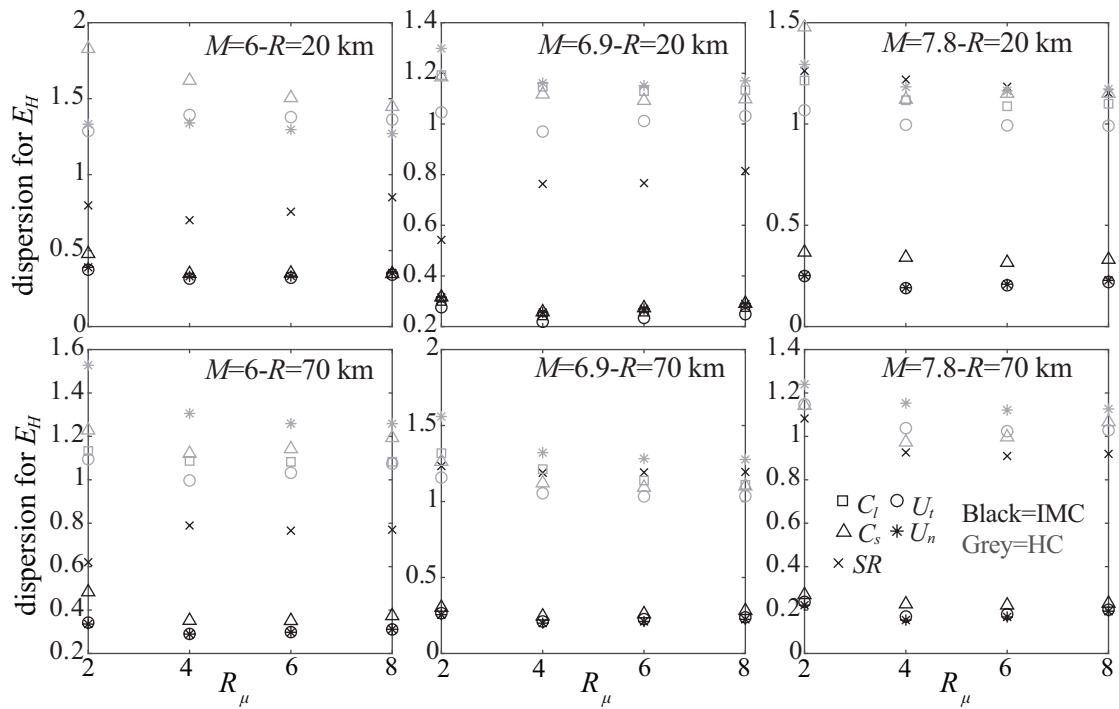FIGURE D.6: Median hysteretic energy for ESD system for "constant-R_μ" approach.

FIGURE D.7: Dispersion (expressed through CoV) of hysteretic energy for ESD system for "constant-R_μ" approach.

FIGURE D.8: Normalised median peak inelastic displacements for EPH system with $\alpha=3\%$ for "constant-strength" approach.

FIGURE D.9: Median hysteretic energy for EPH system with $\alpha=3\%$ for "constant-strength" approach.
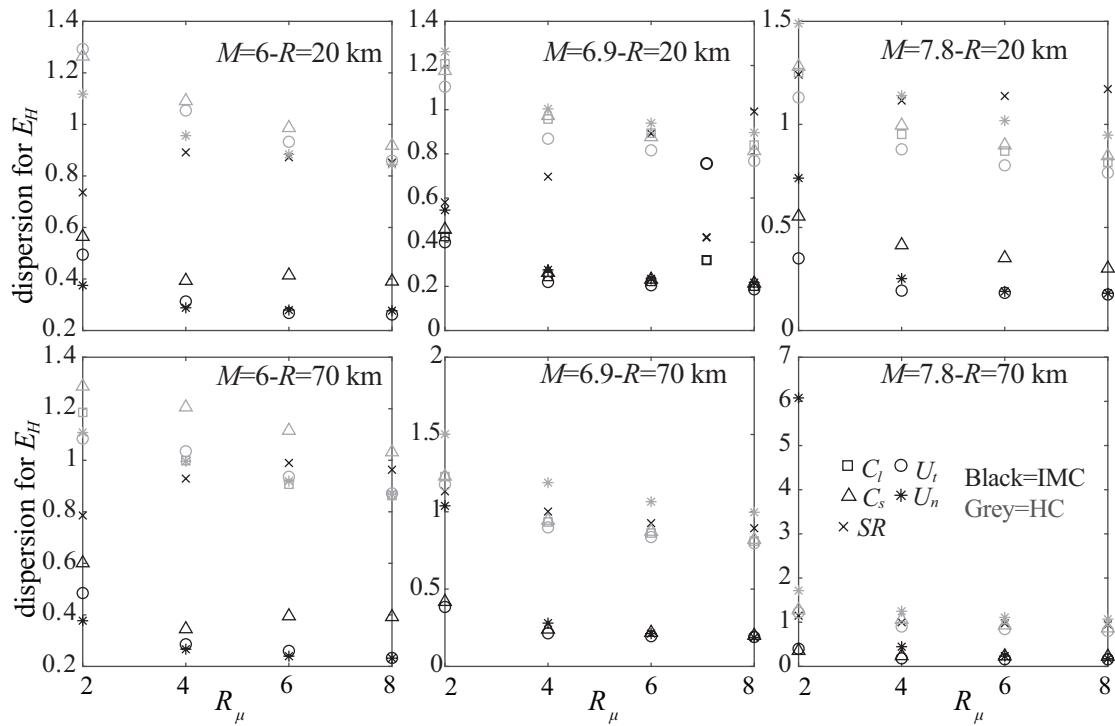
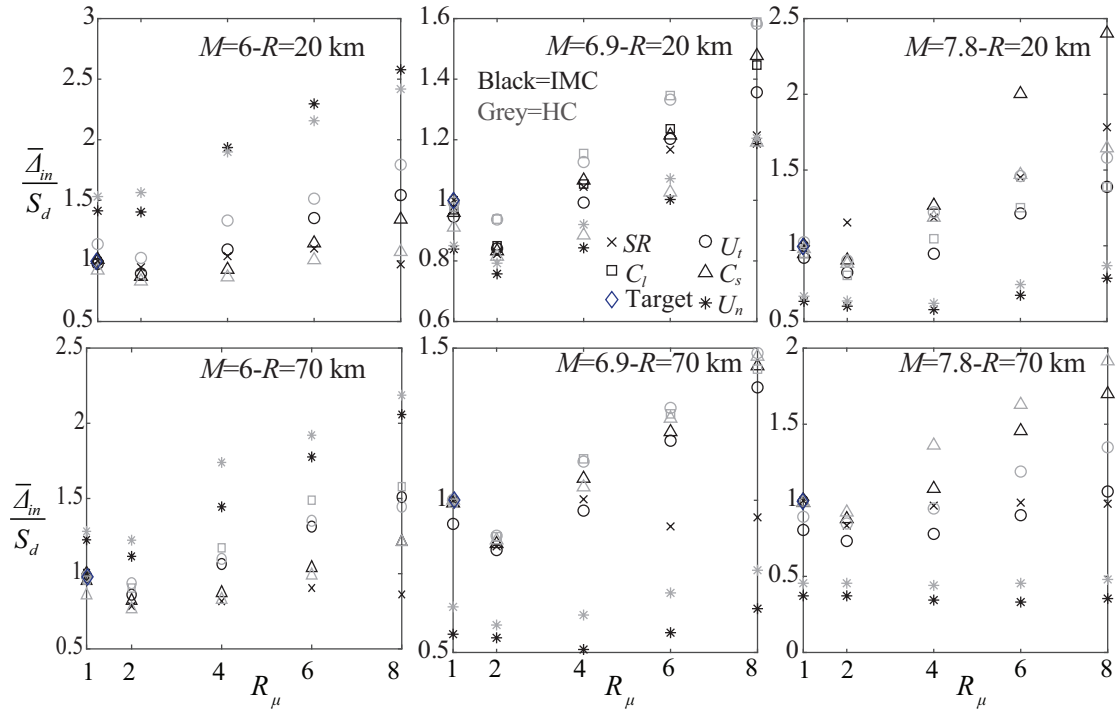FIGURE D.10: Dispersion (expressed through CoV) of hysteretic energy for EPH system with $\alpha=3\%$ for "constant-strength" approach.

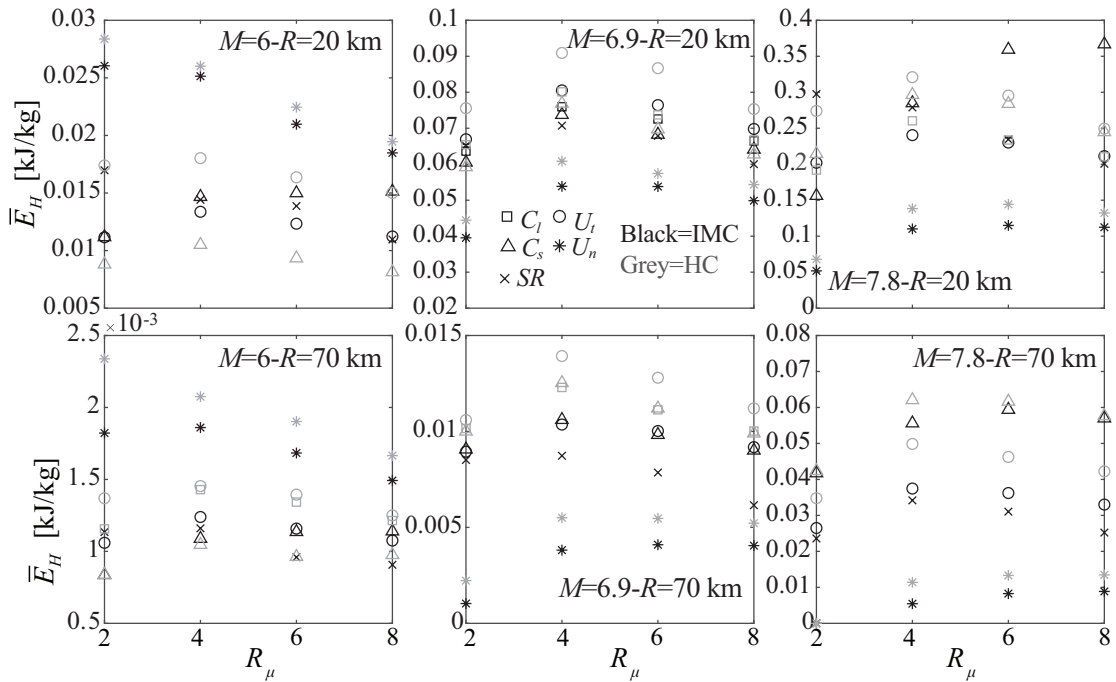FIGURE D.11: Normalised median peak inelastic displacements for ESD system for "constant-strength" approach.

FIGURE D.12:   Median hysteretic energy for ESD system for "constant-strength" approach.
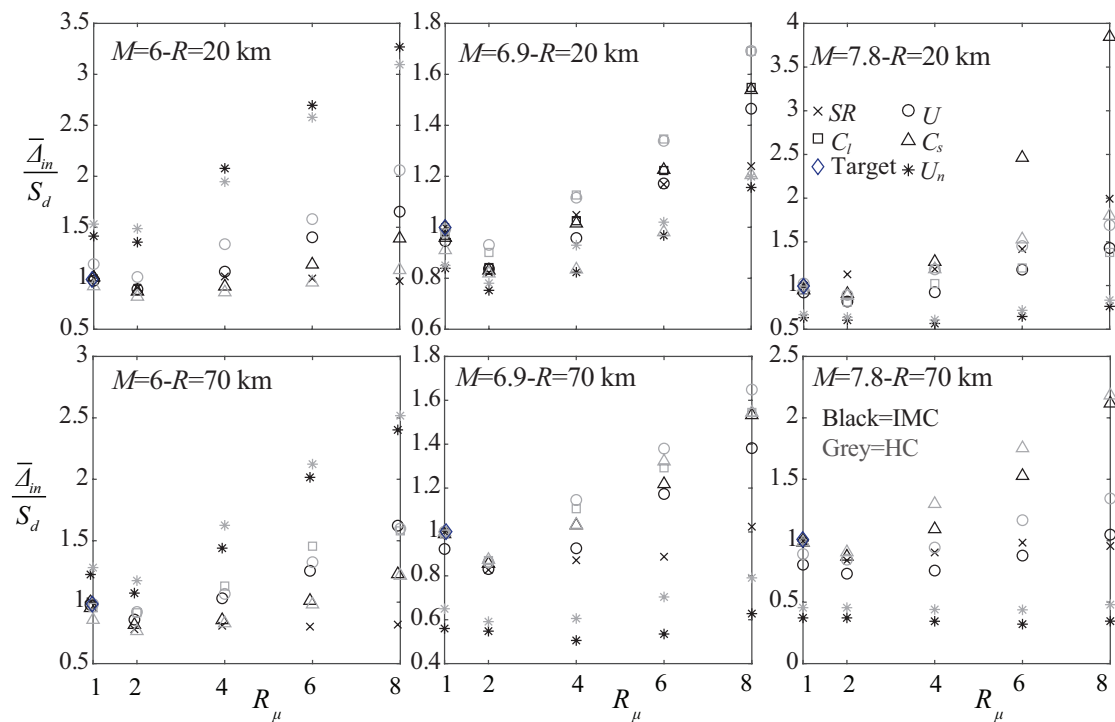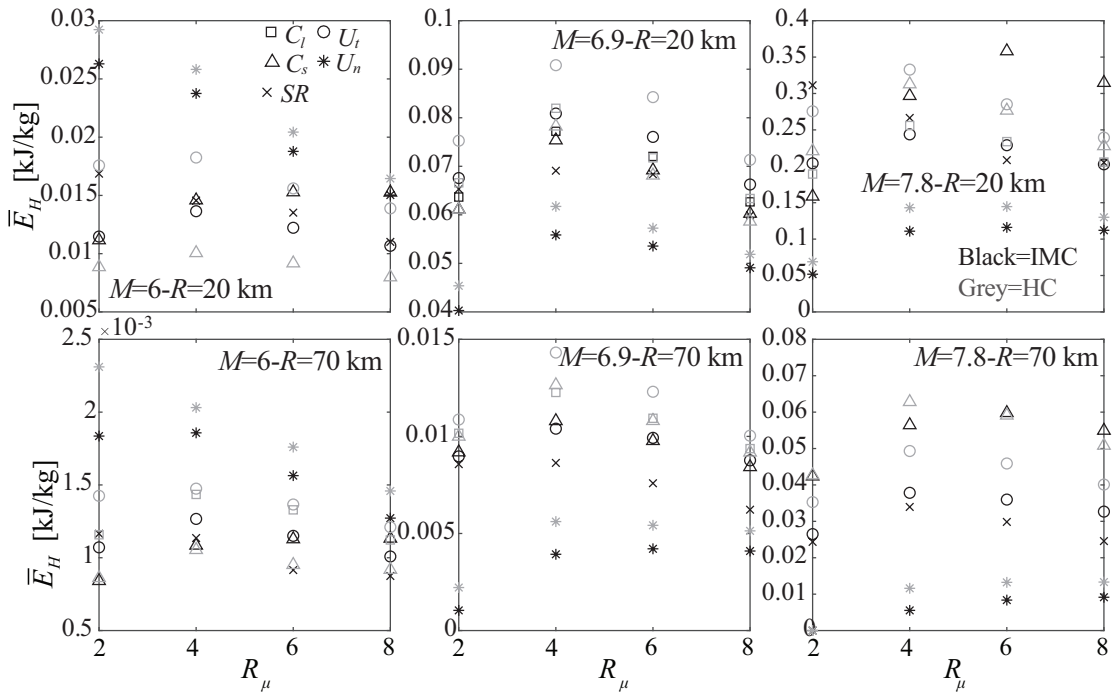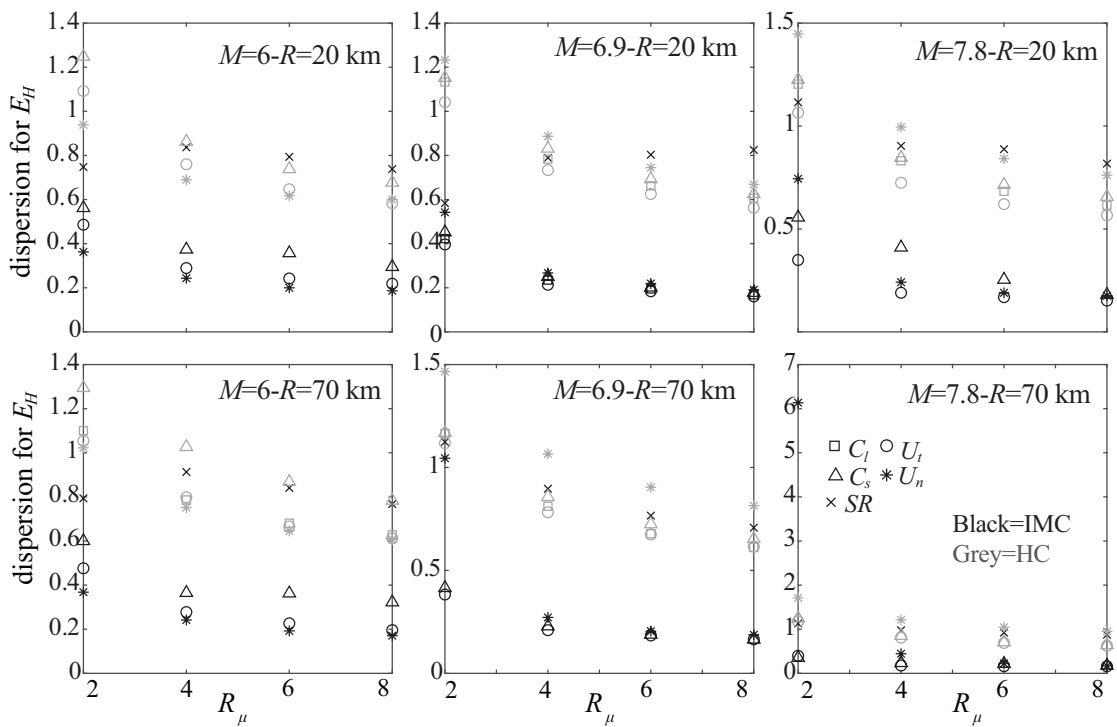

FIGURE D.13:   Dispersion (expressed through CoV) of hysteretic energy for ESD system for "constant-strength" approach.

References

Aagaard, B. T., Brocher, T. M., Dolenc, D., Dreger, D., Graves, R. W., Harmsen, S. C., Hartzell, S., Larsen, S., McCandless, K., Nilsson, S., Petersson, N. A., Rodgers, A., Sjogreen, B. and Zoback, M. L. (2008). Ground-motion modeling of the 1906 San Francisco earthquake, part II: ground-motion estimates for the 1906 earthquake and scenario events, *Bulletin of the Seismological Society of America* **98**(2): 989–1011.

Abrahamson, N. and Silva, W. (2008). Summary of the Abrahamson & Silva NGA ground-motion relations, *Earthquake spectra* **24**(1): 67–97.

Afshari, K. and Stewart, J. P. (2013). Validation approach for application of simulated ground motions to duration-sensitive systems. In GMSV-SEISM TAG Monthly Web Conference, Online, 13, `http://collaborate.scec.org/gmsv_wiki/images/AfshariStewart20131218.pdf`. Accessed: 04-09-2015.

Afshari, K. and Stewart, J. P. (2016). Validation of duration parameters from SCEC broadband platform simulated ground motions, *Seismological Research Letters* **87**(6): 1355–1362.

Ahmadi, G. (1979). Generation of artificial time-histories compatible with given response spectra - a review, *Solid Mechanics Archives* **4**(7): 207–239.

Alterman, Z. and Karal, F. (1968). Propagation of elastic waves in layered media by finite difference methods, *Bulletin of the Seismological Society of America* **58**(1): 367–398.

Amin, M. and Ang, A. (1968). Nonstationary stochastic model of earthquake motions, *Engineering Mechanics Division (ASCE)* **94**(EM2): 556–583.

Anderson, J. (2015). The composite source model for broadband simulations of strong ground motions, *Seismological Research Letters* **86**(1): 68–74.

Anderson, J. G. (2004). Quantitative measure of the goodness-of-fit of synthetic seismogram., *In 13th World Conference on Earthquake Engineering, Earthquake Engineering Research Institute*, Vancouver, B.C., Canada.

ASCE (2010). *Minimum design loads for buildings and other structures*, ASCE 7-10, American Society of Civil Engineers/ Structural Engineering Institute, Reston, VA.

Atkinson, G. M. (2008). Ground-motion prediction equations for eastern North America from a referenced empirical approach: implications for epistemic uncertainty, *Bulletin of the Seismological Society of America* **98**(3): 1304–1318.

Atkinson, G. M. and Assatourians, K. (2015). Implementation and validation of EXSIM (a stochastic finite-fault ground-motion simulation algorithm) on the SCEC Broadband Platform, *Seismological Research Letters* **86**(1): 48–60.

Atkinson, G. M. and Goda, K. (2010). Inelastic seismic demand of real versus simulated ground-motion records for cascadia subduction earthquakes, *Bulletin of the Seismological Society of America* **100**(1): 102–115.

Atkinson, G. M. and Silva, W. (2000). Stochastic modeling of California ground motions, *Bulletin of the Seismological Society of America* **90**(2): 255–274.

Au, S. K. and Beck, J. L. (2003). Subset simulation and its applications to seismic risk based on dynamic analysis, *Journal of Engineering Mechanics, ASCE* **129**(8): 901–917.

Baker, J. (2011). Conditional mean spectrum: Tool for ground motion selection, *Journal of Structural Engineering* **137**(3): 322–331.

Baker, J. W., Luco, N., Abrahamson, N., Graves, R. W., Maechling, P. J. and Olsen, K. B. (2014). Engineering uses of physics-based ground motion simulations, *In 10th National Conference in Earthquake Engineering, Earthquake Engineering Research Institute*, Anchorage, Alaska.

Bao, H., Bielak, J., Ghattas, O., Kallivokas, L., O'Hallaron, D., Shewchuk, J. and Xu, J. (1998). Large-scale simulation of elastic wave propagation in heterogeneous media on parallel computers, *Computer Methods in Applied Mechanics and Engineering* **152**(12): 85–102.

Bazzurro, P. and Cornell, C. (2007). Disaggregation of seismic hazard, *Bulletin of the Seismological Society of America* **89**(2): 501–520.

Bazzurro, P., Sjoberg, B. and Luco, N. (2004). *Post-elastic response of structures to synthetic ground motions. Technical Report PEER 1G00 Addendum 2*, Pacific Earthquake Engineering Research Center, Berkeley, CA, USA.

Beresnev, I. A. and Atkinson, G. M. (1998). Stochastic finite-fault modeling of ground motions from the 1994 Northridge, California, earthquake (I): Validation on rock sites, *Bulletin of the Seismological Society of America* **88**(6): 1392–1401.

Bijelić, N., Lin, T. and Deierlein, G. (2014). Seismic response of a tall building to recorded and simulated ground motions, *In 10th National Conference in Earthquake Engineering, Earthquake Engineering Research Institute*, Anchorage, Alaska.

Bojórquez, E. and Iervolino, I. (2011). Spectral shape proxies and nonlinear structural response, *Soil Dynamics and Earthquake Engineering* **31**(7): 996–1008.

Bolotin, V. V. (1960). Statistical theory of the aseismic design of structures, *In 2nd World Conference on Earthquake Engineering*, Tokyo and Kyoto, Japan.

Boore, D. M. (2003). Simulation of ground motion using the stochastic method, *Pure and Applied Geophysics* **160**(3): 635–676.

Boore, D. M. (2009). Comparing stochastic point-source and finite-source ground-motion simulations: SMSIM and EXSIM., *Bulletin of the Seismological Society of America* **99**(6): 3202–3216.

Boore, D. M. (2010). Orientation-independent, non geometric-mean measures of seismic intensity from two horizontal components of motion, *Bulletin of the Seismological Society of America* **100**(4): 1830–1835.

Boore, D. M. and 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.

Boore, D. M. and Thompson, E. M. (2015). Revisions to some parameters used in stochastic-method simulations of ground motion, *Bulletin of the Seismological Society of America* **105**(2A): 1029–1041.

Bozorgnia, Y. and Bertero, V. V. (2004). *Earthquake engineering: from engineering seismology to performance-based engineering*, CRC press.

Bozorgnia, Y., Hachem, M. M. and Campbell, K. W. (2010). Ground motion prediction equation ("attenuation relationship") for inelastic response spectra, *Earthquake Spectra* **26**(1): 1–23.

Bradley, B. A. (2010). A generalized conditional intensity measure approach and holistic ground-motion selection, *Earthquake Engineering & Structural Dynamics* **39**(12): 1321–1342.

Bradley, B. A., Pettinga, D., Baker, J. W. and Fraser, J. (2017). Guidance on the utilization of earthquake-induced ground motion simulations in engineering practice, *Earthquake Spectra* **33**(3): 809–835.

Broccardo, M. and Der Kiureghian, A. (2015). Multicomponent nonlinear stochastic dynamic analysis by tail-equivalent linearization, *Journal of Engineering Mechanics* **142**(3): 04015100.

Burks, L. S. (2014). *Ground motion simulations: Validation and Application for Civil Engineering Problems*, PhD thesis, Department of Civil and Environmental Engineering, Stanford University, CA, USA.

Burks, L. S. and Baker, J. W. (2014). Validation of ground motion simulations through simple proxies for the response of engineered systems, *Bulletin of the Seismological Society of America* **104**(4): 1930–1946.

Burks, L. S., Zimmerman, R. B. and Baker, J. W. (2015). Evaluation of hybrid broadband ground motion simulations for response history analysis and design, *Earthquake Spectra* **31**(3): 1691–1710.

Campbell, K. W. and Bozorgnia, Y. (2008). NGA ground motion model for the geometric mean horizontal component of PGA, PGV, PGD and 5% damped linear elastic response spectra for periods ranging from 0.01 to 10 s, *Earthquake Spectra* **24**(1): 139–171.

Chandramohan, R., Baker, J. W. and Deierlein, G. G. (2016). Quantifying the influence of ground motion duration on structural collapse capacity using spectrally equivalent records, *Earthquake Spectra* **32**(2): 927–950.

Chang, M., Kwiatkowski, J., Nau, R., Oliver, R. and Pister, K. (1982). ARMA models for earthquake ground motions, *Earthquake Engineering & Structural Dynamics* **10**: 651–662.

Chiou, B., Darragh, R., Gregor, N. and Silva, W. (2008). NGA project strong-motion database, *Earthquake Spectra* **24**(1): 23–44.

Chiou, B.-J. and Youngs, R. R. (2008). An NGA model for the average horizontal component of peak ground motion and response spectra, *Earthquake Spectra* **24**(1): 173–215.

Coello, C. A. C., Lamont, G. B., Van Veldhuizen, D. A. et al. (2007). *Evolutionary algorithms for solving multi-objective problems*, Vol. 5, Springer.

Conte, J. P. (1992a). Effect of earthquake frequency nonstationarity on inelastic structural response, *In 10th World Conference on Earthquake Engineering*, Madrid, Spain.

Conte, J. P. and Peng, B. F. (1997). Fully nonstationarity analytical earthquake ground-motion model, *Journal of Engineering Mechanics* **123**(1): 15–24.

Conte, J. P., Pister, K. S. and Mahin, S. A. (1992b). Nonstationary ARMA modeling of seismic motions, *Soil Dynamics and Earthquake Engineering* **11**(7): 411–426.

Cordova, P. P., Deierlein, G. G., Mehanny, S. S. and Cornell, C. A. (2000). Development of a two-parameter seismic intensity measure and probabilistic assessment procedure, *In 2nd US–Japan Workshop on Performance-Based Earthquake Engineering for Reinforced Concrete Building Structures*, Sapporo, Japan.

Cornell, C. (1964). *Stochastic process models in structural engineering*, PhD thesis, Department of Civil and Environmental Engineering, Stanford University, CA, USA.

Cosenza, E. and Manfredi, G. (1997). The improvement of the seismic-resistant design for existing and new structures using damage criteria, *in* P. Fajfar and H. Krawinkler (eds), *Seismic Design Methodologies for the Next Generation of Codes*, Balkema, Rotterdam: 119-130.

Crempien, J. and Archuleta, R. (2015). UCSB method for broadband ground motion from kinematic simulations of earthquakes, *Seismological Research Letters* **86**(1): 61–67.

Day, S. and Bradley, C. (2001). Memory-efficient simulation of anelastic wave propagation, *Bulletin of the Seismological Society of America* **91**(3): 520–531.

De Luca, F., Ameri, G., Iervolino, I., Pacor, F. and Bindi, D. (2014). Towards validation of simulated accelerograms via prediction equations for nonlinear SDOF response, *Bollettino di geofisica teorica ed applicata* **55**(1): 85–101.

Deniz, D., Song, J. and Hajjar, J. F. (2017). Energy-based seismic collapse criterion for ductile planar structural frames, *Engineering Structures* **141**: 1–13.

Der Kiureghian, A. and Crempien, J. (1989). An evolutionary model for earthquake ground motion, *Structural Safety* **6**: 235–246.

Douglas, J. and Aochi, H. (2008). A survey of techniques for predicting earthquake ground motions for engineering purposes, *Surveys in geophysics* **29**(3): 187.

Dreger, D. S., Beroza, G. C., Day, S. M., Goulet, C. A., Jordan, T. H., Spudich, P. A. and Stewart, J. P. (2013). *Evaluation of SCEC Broadband Platform phase 1 ground motion simulation results*, Southern California Earthquake Center, Los Angeles, CA, USA.

Dreger, D. S., Beroza, G. C., Day, S. M., Goulet, C. A., Jordan, T. H., Spudich, P. A. and Stewart, J. P. (2015). Validation of the SCEC broadband platform V14.3 simulation methods using pseudospectral acceleration data, *Seismological Research Letters* **86**(1): 39–47.

D' Amico, M., Puglia, R., Russo, E., Maini, C., Pacor, F. and Luzi, L. (2017). SYNTHESIS: a web repository of synthetic waveforms, *Bulletin of Earthquake Engineering* **15**(6): 2483–2496.

Dubourg, V., Sudret, B. and Bourinet, J.-M. (2011). Reliability-based design optimization using kriging surrogates and subset simulation, *Structural and Multidisciplinary Optimization* **44**(5): 673–690.

Dumbser, M. and Käser, M. (2006). An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes - ii. the three-dimensional isotropin case, *Geophysical Journal International* **167**(1): 319–336.

Eurocode8 (2004). *Eurocode 8 - Design of structures for earthquake resistance, Part 1-6*, Comité Européen de Normalisation (CEN), EN 1998-1:2004, Brussels, Belgium.

Faccioli, E., Maggio, F., Paolucci, R. and Quarteroni, A. (1997). 2D and 3D elastic wave propagation by a pseudo-spectral domain decomposition method, *Journal of Seismology* **1**(3): 237–251.

Fan, F. and Ahmadi, G. (1990). Nonstationary Kanai-Tajimi models for El Centro 1940 and Mexico City 1985 earthquakes, *Probabilistic Engineering Mechanics* **5**(4): 171–181.

FEMA (2012). FEMA P-58. Seismic performance assessment of buildings, *Technical report*, American Technology Council, Redwood City, CA.

Frankel, A. (2009). A constant stress-drop model for producing broadband synthetic seismograms: Comparison with the Next Generation Attenuation relations, *Bulletin of the Seismological Society of America* **99**(2A): 664–680.

Galasso, C. (2010). *Consolidating record selection for earthquake resistant structural design*, PhD thesis, University of Naples Federico II: Italy.

Galasso, C., Kaviani, P., Tsioulou, A. and Zareian, F. (2018). Validation of ground motion simulations for historical events using skewed bridges, *Journal of Earthquake Engineering* **0**(0): 1–23.

Galasso, C., Zareian, F., Iervolino, I. and Graves, R. W. (2012). Validation of ground-motion simulations for historical events using SDoF systems, *Bulletin of the Seismological Society of America* **102**(6): 2727–2740.

Galasso, C., Zhong, P., Zareian, F., Iervolino, I. and Graves, R. W. (2013). Validation of ground-motion simulations for historical events using MDoF systems, *Earthquake Engineering & Structural Dynamics* **42**(9): 1395–1412.

Gavin, H. P. and Dickinson, B. W. (2010). Generation of uniform-hazard earthquake ground motions, *Journal of Structural Engineering* **137**(3): 423–432.

Giaralis, A. and Spanos, P. (2009). Wavelet-based response spectrum compatible synthesis of accelerograms—Eurocode application (EC8), *Soil Dynamics and Earthquake Engineering* **29**(1): 219–235.

Giaralis, A. and Spanos, P. D. (2010). Effective linear damping and stiffness coefficients of nonlinear systems for design spectrum based analysis, *Soil Dynamics and Earthquake Engineering* **30**(9): 798–810.

Gibbs, A. L. and Su, F. E. (2002). On choosing and bounding probability metrics, *International statistical review* **70**(3): 419–435.

Gidaris, I. and Taflanidis, A. A. (2015). Performance assessment and optimization of fluid viscous dampers through life-cycle cost criteria and comparison to alternative design approaches, *Bulletin of Earthquake Engineering* **13**(4): 1003–1028.

Gidaris, I., Taflanidis, A. A., Lopez-Garcia, D. and Mavroeidis, G. P. (2016). Multi-objective risk-informed design of floor isolation systems, *Earthquake Engineering & Structural Dynamics* **45**(8): 1293–1313.

Gidaris, I., Taflanidis, A. A. and Mavroeidis, G. P. (2015). Kriging metamodeling in seismic risk assessment based on stochastic ground motion models, *Earthquake Engineering & Structural Dynamics* **44**(14): 2377–2399.

Giovenale, P., Cornell, C. A. and 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.

Goulet, C. A., Abrahamson, N. A., Somerville, P. G. and Wooddell, K. E. (2015). The SCEC Broadband Platform validation exercise for pseudo-spectral acceleration: Methodology for code validation in the context of seismic hazard analyses, *Seismological Research Letters* **86**(1): 17–26.

Goulet, C. A., Haselton, C. B., Mitrani-Reiser, J., Beck, J. L., Deierlein, G. G., Porter, K. A. and Stewart, J. P. (2007). Evaluation of the seismic performance of a code-conforming reinforced-concrete frame building—from seismic hazard to collapse safety and economic losses, *Earthquake Engineering & Structural Dynamics* **36**(13): 1973–1997.

Graves, R. (1996). Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences, *Bulletin of the Seismological Society of America* **100**(5A): 2095–2123.

Graves, R. W. and Aagaard, B. T. (2011). Testing long-period ground-motion simulations of scenario earthquakes using the M_w 7.2 El Mayor-Cucapah mainshock; evaluation of finite-fault rupture characterization and 3D seismic velocity models, *Bulletin of the Seismological Society of America* **101**(2): 895–907.

Graves, R. W. and Pitarka, A. (2010). Broadband ground-motion simulation using a hybrid approach, *Bulletin of the Seismological Society of America* **100**(5A): 2095–2123.

Graves, R. W. and Pitarka, A. (2015). Refinements to the Graves and Pitarka (2010) broadband ground-motion simulation method, *Seismological Research Letters* **86**(1): 75–80.

Hartigan, J. A. and Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm, *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **28**(1): 100–108.

Hartzell, S., Harmsen, S., Frankel, A. and Larsen, S. (1999). Calculation of broadband time histories of ground motion: Comparison of methods and validation using strong-ground motion from the 1994 Northridge earthquake, *Bulletin of the Seismological Society of America* **89**(6): 1484–1504.

Hartzell, S., Mariagiovanna, G., Mai, P. M., Liu, P.-C. and Fisk, M. (2005). Calculation of broadband time histories of ground motion, part II: Kinematic and dynamic modeling using theoretical Green's functions and comparison with the 1994 Northridge earthquake, *Bulletin of the Seismological Society of America* **95**(2): 614–645.

Haselton, C. (2009). *Evaluation of Ground Motion Selection and Modification Methods: Predicting Median Interstory Drift Response of Buildings. PEER Report 2009/01*, Pacific Earthquake Engineering Research Center, College of Engineering, University of California, Berkeley, USA.

Holmstrom, K., Goran, A. and Edvall, M. (2009). User's guide for TOMLAB 7, San Diego, CA, Tomlab Optimization Inc, `www.tomopt.com/tomlab/`.

Hoshiya, M. and Hasgur, Z. (1978). AR and MA models of nonstationary ground motion, *Bulletin of the International Institute of Seismology and Earthquake Engineering* **16**: 55–68.

Husid, L. R. (1969). Características de terremotos, análisis general, *Revista del IDIEM 8*, Santiago de Chile: 21-42.

Iervolino, I., De Luca, F. and Cosenza, E. (2010). Spectral shape-based assessment of SDOF nonlinear response to real, adjusted and artificial accelerograms, *Engineering Structures* **32**(9): 2776–2792.

Iervolino, I., Galasso, C. and Cosenza, E. (2010). REXEL: computer aided record selection for code-based seismic structural analysis, *Bulletin of Earthquake Engineering* **8**(2): 339–362.

Iervolino, I., Giorgio, M., Galasso, C. and Manfredi, G. (2010). Conditional hazard maps for secondary intensity measures, *Bulletin of the Seismological Society of America* **100**(6): 3312–3319.

Iervolino, I., Manfredi, G. and Cosenza, E. (2006). Ground-motion duration effects on nonlinear seismic response, *Earthquake Engineering & Structural Dynamics* **35**(1): 21–38.

Irikura, K. (1983). Semi-empirical estimation of strong ground motions during large earthquakes, *Bulletin of Disaster Prevention Research Institute, Kyoto University* **33**(298): 63–104.

Iyengar, R. N. and Iyengar, K. T. S. R. I. (1969). A nonstationary random process model for earthquake acceleration, *Bulletin of the Seismological Society of America* **59**(3): 1163–1188.

Jalayer, F. and Beck, J. (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.

Jalayer, F., Beck, J. and Zareian, F. (2012). Analyzing the sufficiency of alternative scalar and vector intensity measures of ground shaking based on information theory, *ASCE Journal of Engineering Mechanics* **138**(3): 307–316.

Jayaram, N., Lin, T. and Baker, J. W. (2011). A computationally efficient ground-motion selection algorithm for matching a target response spectrum mean and variance, *Earthquake Spectra* **27**(3): 797–815.

Jayaram, N., Park, J., Bazzurro, P. and Tothong, P. (2010). Spatial correlation between spectral accelerations using simulated ground-motion time histories, *In 9th US National and 10th Canadian Conference on Earthquake Engineering*, pp. 1–10.

Jayaram, N. and Shome, N. (2012). A statistical analysis of the response of tall buildings to recorded and simulated ground motions, *In 15th World Conference on Earthquake Engineering*, Lisbon, Portugal.

Jensen, H. and Kusanovic, D. (2014). On the effect of near-field excitations on the reliability-based performance and design of base-isolated structures, *Probabilistic Engineering Mechanics* **36**: 28–44.

Jia, G. and Taflanidis, A. A. (2013). Kriging metamodeling for approximation of high-dimensional wave and surge responses in real-time storm/hurricane risk assessment, *Computer Methods in Applied Mechanics and Engineering* **261**: 24–38.

Jurkevics, A. and Ulrych, T. J. (1978). Representing and simulating strong ground motion, *Bulletin of the Seismological Society of America* **68**(3): 781–801.

Kaklamanos, J., Baise, L. G. and Boore, D. M. (2011). Estimating unknown input parameters when implementing the NGA ground-motion prediction equations in engineering practice, *Earthquake Spectra* **27**(4): 1219–1235.

Kanai, K. (1957). Semi-empirical formula for the seismic characteristics of the ground, *Bulletin of the Earthquake Research Institute* **35**: 309–325.

Katsanos, E. I., Sextos, A. G. and 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.

Katsanos, E. and Sextos, A. (2015). Inelastic spectra to predict period elongation of structures under earthquake loading, *Earthquake Engineering & Structural Dynamics* **44**(11): 1765–1782.

Kazantzi, A. K. and Vamvatsikos, D. (2015). Intensity measure selection for vulnerability studies of building classes, *Earthquake Engineering & Structural Dynamics* **44**(15): 2677–2694.

Keeney, R. L. and Raiffa, H. (1993). *Decisions with multiple objectives: preferences and value trade-offs*, Cambridge university press.

Kohrangi, M., Bazzurro, P., Vamvatsikos, D. and Spillatura, A. (2017). Conditional spectrum-based ground motion record selection using average spectral acceleration, *Earthquake Engineering & Structural Dynamics* **46**(10): 1667–1685.

Komatitsch, D. and Tromp, J. (1999). Introduction to the spectral element method for three-dimensional seismic wave propagation, *Geophysical Journal International* **139**(3): 806–822.

Komatitsch, D. and Vilotte, J.-P. (1998). The spectral element method: An efficient tool to simulate the seismic response of 2D and 3D geological structures, *Bulletin of the Seismological Society of America* **88**(2): 368–392.

Kozin, F. (1988). Autoregressive moving average models of earthquake record, *Probabilistic Engineering Mechanics* **3**(2): 58–63.

Krishnan, S., Muto, M., Mourhatch, R., Bjornsson, A. B. and Siriki, H. (2011). Rupture-to-rafters simulations: Unifying science and engineering for earthquake hazard mitigation, *Computing in Science & Engineering* **13**(4): 28–43.

Kristekova, M., Kristek, J., Moczo, P. and Day, S. M. (2006). Misfit criteria for quantitative comparison of seismograms, *Bulletin of the Seismological Society of America* **96**(5): 1836–1850.

Kullback, S. (1959). *Information and Statistics*, Wiley, New York, NY, USA.

Lin, T., Haselton, C. B. and Baker, J. W. (2013). Conditional spectrum-based ground motion selection. Part I: hazard consistency for risk-based assessments, *Earthquake Engineering & Structural Dynamics* **42**(12): 1847–1865.

Lin, Y. K. (1965). Nonstationary excitation and response in linear systems treated as sequences of random pulses, *Journal of the Acoustical Society of America* **38**: 453–460.

Lin, Y. K. (1986). On random pulse train and its evolutionary spectral representation, *Probabilistic Engineering Mechanics* **1**(4): 219–223.

Liu, P., Archuleta, R. and Hartzell, S. H. (2006). Prediction of broadband ground motion time histories: Frequency method with correlation random source parameters, *Bulletin of the Seismological Society of America* **96**(6): 2118–2130.

Liu, S. C. (1970). Evolutionary power spectral density of strong-motion earthquakes, *Bulletin of the Seismological Society of America* **60**(3): 891–900.

Lophaven, S. N., Nielsen, H. B. and Søndergaard, J. (2002). *DACE - A Matlab Kriging Toolbox, Version 2.0*, Citeseer.

Luco, N., Jordan, T. H. and Rezaeian, S. (2013). Progress of the Southern California Earthquake Center Technical Activity Group on Ground Motion Simulation Validation, *Seismological Research Letters* **84**(2): 336.

Lysmer, J. and Drake, L. A. (1972). A finite element method for seismology, *in* B. Alder, S. Fernbach and B. Bolt (eds), *Methods in Computational Physics (Vol.11, Ch.6)*, Academic Press, New York, USA.

Maechling, P. J., Silva, F., Callaghan, S. and Jordan, T. H. (2015). SCEC broadband platform: System architecture and software implementation, *Seismological Research Letters* **86**(1): 27–38.

Mai, P. M., Imperatori, W. and Olsen, K. B. (2010). Hybrid broadband ground-motion simulations: Combining long-period deterministic synthetics with high-frequency multiple S-to-S backscattering, *Bulletin of the Seismological Society of America* **100**(5A): 2124–2142.

Manfredi, G. (2001). Evaluation of seismic energy demand, *Earthquake Engineering & Structural Dynamics* **30**(4): 485–499.

Marler, R. T. and Arora, J. S. (2004). Survey of multi-objective optimization methods for engineering, *Structural and multidisciplinary optimization* **26**(6): 369–395.

Mavrotas, G. (2009). Effective implementation of the ε-constraint method in multi-objective mathematical programming problems, *Applied mathematics and computation* **213**(2): 455–465.

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

McGuire, R. K. and Hanks, T. C. (1980). RMS accelerations and spectral amplitude of strong ground motion during the San Fernando, California, earthquake, *Bulletin of the Seismological Society of America* **70**(5): 1907–1919.

Medel-Vera, C. and Ji, T. (2016). A stochastic ground motion accelerogram model for Northwest Europe, *Soil Dynamics and Earthquake Engineering* **82**: 170–195.

Mitseas, I. P., Kougioumtzoglou, I. A., Giaralis, A. and Beer, M. (2018). A novel stochastic linearization framework for seismic demand estimation of hysteretic MDOF systems subject to linear response spectra, *Structural Safety* **72**: 84–98.

Mobarakeh, A., Rofooei, F. and Ahmadi, G. (2002). Simulation of earthquake records using time-varying Arma (2,1) model, *Probabilistic Engineering Mechanics* **17**(1): 15–34.

Motazedian, D. and Atkinson, G. M. (2005). Stochastic finite-fault modeling based on a dynamic corner frequency, *Bulletin of the Seismological Society of America* **95**(3): 995–1010.

NIST (2011). *Selecting and Scaling Earthquake Ground Motions for Performing Response-History Analyses. NIST GCR 11-917-15*, National Institute of Standards and Technology, U.S. Department of Commerce, Gaithersburg, MD, USA.

Olsen, K. B., Day, S. M. and R., B. C. (2003). Estimation of Q for long-period (> 2 s) waves in the Los Angeles basin, *Bulletin of the Seismological Society of America* **93**(2): 627–638.

Olsen, K. B., Madariaga, R. and Archuleta, R. J. (1997). Three-dimensional dynamic simulation of the 1992 Landers earthquake, *Science* **278**(5339): 834–838.

Olsen, K. B. and Mayhew, J. E. (2010). Goodness-of-fit criteria for broadband synthetic seismograms, with application to the 2008 M5.4 Chino Hills, California, earthquake, *Seismological Research Letters* **81**(5): 715–723.

Olsen, K. and Takedatsu, R. (2015). The SDSU broadband ground motion generation module BBtoolbox Version 1.5, *Seismological Research Letters* **86**(1): 81–88.

Padgett, J. E., Nielson, B. G. and 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.

Papadimitriou, K. (1990). *Stochastic characterization of strong ground motion and applications to structural response. Report No. EERL 90-03, Earthquake Engineering Research Laboratory*, California Institute of Technology, Pasadena, CA, USA.

Petersen, M. D., Frankel, A. D., Harmsen, S. C., Mueller, C. S., Haller, K. M., Wheeler, R. L., Wesson, R. L., Zeng, Y., Boyd, O. S., Perkins, D. M. et al. (2008). Documentation for the 2008 update of the United States national seismic hazard maps, *Technical report*, Geological Survey (US).

Pitarka, A.and Irikura, K., Iwata, T. and Sekiguchi, H. (1998). Three-dimensional simulation of the near-fault ground motion for the 1995 Hyogo-Ken Nanbu (Kobe), Japan, earthquake, *Bulletin of the Seismological Society of America* **88**(2): 428–440.

Polhemus, N. and Cakmak, A. (1981). Simulation of earthquake ground motions using autoregressive moving averate (ARMA) models, *Earthquake Engineering & Structural Dynamics* **9**: 343–354.

Pousse, G., Bonilla, L. F., Cotton, F. and Margerin, L. (2006). Nonstationary stochastic simulation of strong ground motion time histories including natural variability: Application to the K-net Japanese database, *Bulletin of the Seismological Society of America* **96**(6): 2103–2117.

Pulido, N. and Dalguer, L. A. (2009). Estimation of the high-frequency radiation of the 2000 Tottori (Japan) earthquake based on a dynamic model of fault rupture: Application to the strong ground motion simulation, *Bulletin of the Seismological Society of America* **99**(4): 2305–2322.

Raghunandan, M. and Liel, A. B. (2013). Effect of ground motion duration on earthquake-induced structural collapse, *Structural Safety* **41**: 119–133.

Rathje, E. M. and Peterman, B. (2013). Validation of ground motion simulations for seismic slope stability. In GMSV-SEISM TAG Monthly Web Conference,

Online, 12, `http://www.scec.org/workshops/2013/gmsv/1545_Rathje.pdf`. Accessed: 04-09-2015.

Rezaeian, S. and Der Kiureghian, A. (2008). A stochastic ground motion model with separable temporal and spectral nonstationarities, *Earthquake Engineering & Structural Dynamics* **37**(13): 1565–1584.

Rezaeian, S. and Der Kiureghian, A. (2010). Simulation of synthetic ground motions for specified earthquake and site characteristics, *Earthquake Engineering & Structural Dynamics* **39**(10): 1155–1180.

Rezaeian, S. and Sun, X. (2015). Stochastic ground motion simulation, *in* M. Beer, I. Kougioumtzoglou, E. Patelli and S. Au (eds), *Encyclopedia of Earthquake Engineering*, Springer Berlin Heidelberg, Berlin, Germany.

Rezaeian, S., Zhong, P., Hartzell, S. and Zareian, F. (2015). Validation of simulated earthquake ground motions based on evolution of intensity and frequency content, *Bulletin of the Seismological Society of America* **105**(6): 3036–3049.

Robert, C. and Casella, G. (2004). *Monte Carlo statistical methods*, Springer, New York, NY.

Ruiz-García, J. and Miranda, E. (2003). Inelastic displacement ratios for evaluation of existing structures, *Earthquake Engineering & Structural Dynamics* **32**(8): 1237–1258.

Ruiz, J., Baumont, D., Bernard, P. and Berge-Thierry, C. (2007). New approach in the kinematic k^{-2} source model for generating physical slip velocity functions, *Geophysical Journal International* **171**(2): 739–754.

Ruiz, P. and Penzien, J. (1971). Stochastic seismic response of structures, *Journal of Engineering Mechanics ASCE* **97**(EM2): 441–456.

Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P. (1989). Design and analysis of computer experiments, *Statistical science* pp. 409–423.

Saragoni, G. R. and Hart, G. C. (2002). Simulation of artificial earthquakes, *Earthquake Engineering & Structural Dynamics* **2**: 249–267.

Scherbaum, F., Cotton, F. and Staedtke, H. (2006). The estimation of minimum-misfit stochastic models from empirical ground-motion prediction equations, *Bulletin of the Seismological Society of America* **96**(2): 427–445.

Schölkopf, B., Smola, A. J. et al. (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond*, MIT press, Cambridge, Massaschusetts.

Scott, D. W. (2015). *Multivariate density estimation: theory, practice, and visualisation*, John Wiley & Sons.

Scott, D. W. and Sain, S. R. (2005). Multidimensional density estimation, *Handbook of statistics* **24**: 229–261.

Seifried, A. and Baker, J. (2016). Spectral variability and its relationship to structural response estimated from scaled and spectrum-matched ground motions, *Earthquake Spectra* **32**(4): 2191–2205.

Shinozuka, M. and Deodatis, G. (1988). Stochastic process models for earthquake ground motion, *Probabilistic Engineering Mechanics* **3**(3): 114–123.

Shinozuka, M. and Deodatis, G. (1991). Simulation of stochastic processes by spectral representation, *Applied Mechanics Reviews* **44**(4): 191–204.

Shinozuka, M. and Sato, Y. (1967). Simulation of nonstationary random process, *Journal of Engineering Mechanics ASCE* **93**(EM1): 11–40.

Shome, N., Cornell, C. A., Bazzurro, P. and Carballo, J. E. (1998). Earthquakes, records, and nonlinear responses, *Earthquake Spectra* **14**(3): 469–500.

Smerzini, C., Galasso, C., Iervolino, I. and Paolucci, R. (2014). Ground motion record selection based on broadband spectral compatibility, *Earthquake Spectra* **30**(4): 1427–1448.

Sørensen, M. and Lang, D. (2015). Incorporating simulated ground motion in seismic risk assessment: Application to the lower Indian Himalayas, *Earthquake Spectra* **31**(1): 71–95.

Spall, J. (2003). Introduction to stochastic search and optimization: Estimation, simulation and control, *Wiley* **34**: 54–58.

Star, L. M., Stewart, J. P. and Graves, R. W. (2011). Comparison of ground motions from hybrid simulations to NGA prediction equations, *Earthquake Spectra* **27**(2): 331–350.

Stewart, J. P., Boore, D. M., Seyhan, E. and Atkinson, G. M. (2016). NGA-West2 equations for predicting vertical-component PGA, PGV, and 5%-damped PSA from shallow crustal earthquakes, *Earthquake Spectra* **32**(2): 1005–1031.

Stewart, J. P., Chiou, S., Bray, J. D., Graves, R. W., Somerville, P. G. and Abrahamson, N. A. (2001). *Ground motion evaluation procedures for performance-based design. Technical Report PEER 2001/09*, Pacific Earthquake Engineering Research Center, Berkeley, CA, USA.

Taborda, R. and Roten, D. (2015). Physics-based ground motion simulation, *in* M. Beer, I. Kougioumtzoglou, E. Patelli and S. Au (eds), *Encyclopedia of Earthquake Engineering*, Springer Berlin Heidelberg, Berlin, Germany.

Tajimi, H. (1960). A statistical method of determing the maximum response of a building structure during an earthquake., *Proc. 2nd World Conf. Earthq. Eng.*, pp. 781–797.

Trifunac, M. D. and Brady, A. G. (1975). A study on the duration of strong earthquake ground motion, *Bulletin of the Seismological Society of America* **65**(3): 581–626.

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

Tsioulou, A., Taflanidis, A. A. and Galasso, C. (2018a). Modification of stochastic ground motion models for matching target intensity measures, *Earthquake Engineering & Structural Dynamics* **47**(1): 3–24.

Tsioulou, A., Taflanidis, A. A. and Galasso, C. (2018b). Hazard-compatible modification of stochastic ground motion models, *Earthquake Engineering & Structural Dynamics* **47**(8): 1774–1798.

Vetter, C. R., Taflanidis, A. A. and Mavroeidis, G. P. (2016). Tuning of stochastic ground motion models for compatibility with ground motion prediction equations, *Earthquake Engineering & Structural Dynamics* **45**(6): 893–912.

Vetter, C. and Taflanidis, A. A. (2014). Comparison of alternative stochastic ground motion models for seismic risk characterization, *Soil Dynamics and Earthquake Engineering* **58**: 48–65.

Vlachos, C., Papakonstantinou, K. G. and Deodatis, G. (2016). A multi-modal analytical non-stationary spectral model for characterization and stochastic simulation of earthquake ground motions, *Soil Dynamics and Earthquake Engineering* **80**: 177–191.

Vlachos, C., Papakonstantinou, K. G. and Deodatis, G. (2018). Predictive model for site specific simulation of ground motions based on earthquake scenarios, *Earthquake Engineering & Structural Dynamics* **47**(1): 195–218.

Wang, J., Fan, L., Qian, S. and Zhou, J. (2002). Simulations of non-stationary frequency content and its importance to seismic assessment of structures, *Earthquake Engineering & Structural Dynamics* **31**(4): 993–1005.

Wang, R. (1999). A simple orthonormalization method for stable and efficient computation of Green's functions, *Bulletin of the Seismological Society of America* **89**(3): 733–741.

Yamamoto, Y. and Baker, J. W. (2013). Stochastic model for earthquake groun motion using wavelet packets, *Bulletin of the Seismological Society of America* **103**(6): 3044–3056.

Yeh, C. and Wen, Y. (1990). Modeling of nonstationary ground motion and analysis of inelastic structural response, *Structural Safety* **8**(1-4): 281–298.

Zeng, Y., Anderson, J. G. and Yu, G. (1994). A composite source model for computing realistic synthetic strong ground motions, *Geophysical Research Letters* **21**(8): 725–728.