FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

DEVELOPMENT OF INNOVATIVE METHODOLOGIES FOR THE TREATMENT OF UNCERTAINTIES IN THE EARTHQUAKE LOSS ESTIMATION OF BUILDING PORTFOLIOS

Luis Miguel Costa Sousa

Doctoral Program in Civil Engineering

Supervisor: Mário António Lage Alves Marques

Co-supervisor: Humberto Salazar Amorim Varum

Co-supervisor: Vitor Emanuel Marta da Silva

June, 2017

This page is intentionally left blank

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

DEVELOPMENT OF INNOVATIVE METHODOLOGIES FOR THE TREATMENT OF UNCERTAINTIES IN THE EARTHQUAKE LOSS ESTIMATION OF BUILDING PORTFOLIOS

A thesis submitted to the Faculty of Engineering of the University of Porto in fulfilment of the requirements for the degree of Doctor of Philosophy

Scientific coordination of Dr. Mário António Lage Alves Marques, post-doctoral researcher at the Faculty of Engineering of the University of Porto; with co-supervision of Professor Humberto Salazar Amorim Varum, full professor at the Faculty of Engineering of the University of Porto, and Dr. Vitor Emanuel Marta da Silva, post-doctoral researcher at the Department of Civil Engineering of the University of Aveiro

Orientação científica do Doutor Mário António Lage Alves Marques, investigador post-doc da Faculdade de Engenharia da Universidade do Porto; e coorientação do Professor Doutor Humberto Salazar Amorim Varum, professor catedrático da Faculdade de Engenharia da Universidade do Porto, e Doutor Vitor Emanuel Marta da Silva, investigador post-doc do Departamento de Engenharia Civil da Universidade de Aveiro

This page is intentionally left blank

Abstract

Earthquake losses registered worldwide over the last century have triggered crippling effects on the economic and social systems of wealthy and undeveloped countries alike. In the face of ever increasing impacts, earthquake loss modelling is essential for the prediction, prevention and mitigation of the adverse effect of future seismic events. Given the complex nature of the process, it is utopian to seek for absolute certainty when gathering the required resources. As an inherent property of any analytical process, uncertainty must not be 'removed' from the equation. One must seek to improve our knowledge to the extent imposed by practical limitations. However, a seismic risk assessment can only be meaningful if fully coupled with its accompanying analysis of uncertainty, be that aleatory or knowledge-based.

In the context of seismic risk, several questions remain entirely unanswered or lack a deeper understanding. Therefore, in the present work, important issues related with the treatment of uncertainty in portfolio loss estimation are addressed, focusing on the building exposure and vulnerability counterparts. With regard to building exposure, an innovative algorithm is proposed, providing an automated tool for the development of exposure datasets of industrial buildings in Europe, based on open-access data and Volunteered Geographic Information (VGI). With respect to building fragility and vulnerability, on the other hand, the present work sheds light on several problems and limitations in current practice, focusing on the impact that (commonly assumed) hazard disaggregation approximations have on various risk metrics. Building on the latter findings, the appropriate treatment of structural capacity and seismic demand variability is further addressed. More specifically, questions regarding the treatment of such sources of uncertainty are studied from a statistical significance point of view, providing novel and robust approaches to the problems of hazard-compatible ground motion selection, estimation of building response variability, and representation of uncertainty and (spatial) correlation of damage exceedance probabilities. This contribution is subsequently extended to the definition of building vulnerability, whereby innovative *conditional fragility functions* and the resulting vulnerability model are proposed and included in a novel loss assessment framework. The proposed methodology allows a robust evaluation of the impact of variability and spatial correlation of building vulnerability, highlighting important limitations of state-of-the-art methods.

The combined contribution of the aforementioned efforts results in the final undertaking of this work, which consists of the development of a methodology that is able to adequately propagate the epistemic uncertainty of hazard modelling into the corresponding risk results. The latter ensures a robust and statistically meaningful representation of aleatory (and epistemic) variability in structural response and damage exceedance probability, as well as the explicit modelling of the (hazard-consistent) spatial correlation of building losses.

This page is intentionally left blank

Resumo

Perdas devidas à acção sísmica registadas a nível global durante o último século têm provocado efeitos devastadores no tecido económico e social de diversos países, mais ou menos desenvolvidos. Perante este impacto, a modelação analítica de perdas sísmica torna-se uma ferramenta essencial para a previsão, prevenção e mitigação do efeito adverso de futuros eventos sísmicos. Dada a incerteza associada a este processo, é utópico pretender obter resultados com absoluta certeza. Como parte inerente de qualquer estudo analítico, a incerteza não deve ser 'removida' da equação. O objectivo reside na melhoria do conhecimento associado às diferentes variáveis envolvidas, na medida imposta por limitações de ordem práctica. No entanto, a robustez de um estudo de risco sísmico é garantida apenas quando este é devidamente acompanhado de uma análise das várias incertezas, sejam estas de ordem epistémica ou aleatória.

No âmbito do risco sísmico, várias questões continuam inteiramente sem resposta ou necessitam de um estudo mais aprofundado. Deste modo, vários problemas relacionados com o tratamento da incerteza são abordados neste trabalho, com especial enfase nas components de avaliação do património exposto e sua vulnerabilidade sísmica. No que respeita ao primeiro, é desenvolvido um inovador algoritmo automático para a caracterização da localização espacial de edifícios industriais na Europa, com base em fontes públicos de informação geo-referenciada. No que diz respeito à fragilidade de edifícios, o presente trabalho centra-se em diversas limitações identificadas na literatura, focando-se no impacto que aproximações generalizadamente assumidas durante o precesso de desagregação de perigosidade sísmica têm em diferentes medidas de risco. Com base nestes resultados, o tratamento da variabilidade na resposta estrutural e acção sísmica é subsequentemente abordado. Mais concretamente, as referidas fontes de incerteza são estudadas do ponto de vista estatístico, conduzindo a propostas alternativas e robustas com vista à análise de problemas de: selecção de accelerogramas naturais compatíveis com a perigosidade local, estimativa da variabilidade na resposta estrutural, e representação da incerteza e correlação espacial entre probabilidades de excedencia de dano. Esta contribuíção é ainda extendida à definição de vulnerabilidade, meio pelo qual as propostas *funções de fragilidade condicionais* são integradas num nova metodologia para a estimativa de perdas. Esta metodologia permite efectuar a avaliação probabilística do impacto da variabilidade e correlação espacial da vulnerabilidade de uma forma robusta, evidenciando limitações importantes no estado da arte.

A conjugação dos esforços acima mencionados resulta no estudo final deste trabalho, que consiste no desenvolvimento de uma metodologia que é capaz de, de forma consistente, ter em conta a incerteza epistemica desde a avaliação de perigosidade até ao cálculo de perdas. É assim possível uma representação da incerteza aleatória (e epistemica) na resposta estrutural e probabilidade de dano, bem como a modelação adequada da correlação espacial de dano sísmico.

This page is intentionally left blank

Acknowledgements

I would like to start by expressing my most sincere gratitude to Dr. Mário Marques, my supervisor, for his encouragement, knowledge, and human qualities, constantly present throughout this work. I am deeply grateful for his motivation and wisdom, for his selfless help, and more importantly, his friendship.

To Dr. Vitor Silva, my co-advisor and friend, I would like to express my deepest appreciation for the crucial role in this thesis, particularly for encouraging and enabling me to start this work. His personal, scientific and professional qualities are one of the main drivers of my efforts to achieve increasingly demanding goals.

To Professor Humberto Varum, my co-supervisor, I must extend my admiration and gratitude for the contributions and support that ultimately made this thesis possible.

To Dr. Helen Crowley, Dr. Graeme Weatherill and Dr. Paolo Bazzurro, co-authors in four of the scientific papers developed within the context of this thesis, I would like to express my infinite appreciation for their example, scientific knowledge and invaluable contribution to this work.

My sincere gratitude goes to Dr. Rui Pinho, Professor Gian Michele Calvi and Dr. Ricardo Monteiro, who through their example, energy and leadership have decisively contributed not only for my personal and scientific achievements during the time I spent in Pavia, but also for enabling future prospects.

To Professor Raimundo Delgado, I would like to express my admiration and gratitude for the selfless contribution to this thesis, as well as the decisive influence on making it possible. His example and help throughout this journey are infinitely appreciated.

To my friends in Pavia, namely Romain Sousa, Cecilia Nievas, Jennisie Tipler, James Brown, Vitor Silva, Daniela Rodrigues, Rui Figueiredo and Catarina Costa, I would like express my sincere appreciation for the shared moments of companionship. A special thanks goes to Caterina Manieri, for her kind affection and friendship.

To all my childhood friends, especially José Barbosa, João Azevedo and Hugo Garcês, I am thankful for all the diatribes of nonsensical conversations and laughter that make efforts like this worthwhile.

To my Parents, my amazing brothers, my beautiful sisters, and my joyful nephews and god-daughter, I would like to say a special thank you, for all the love and support.

Finally, I would like to express my deepest admiration and love to Venetia Despotaki. This accomplishment would not exist without the love, patience and warmth of such a beautiful human being.

This page is intentionally left blank

Contents

List of Figures

List of Tables

This page is intentionally left blank

Chapter 1 INTRODUCTION

1.1 Earthquake risk

Seismic action is paramount among natural hazards impacting civil infrastructure and human activity all over the globe (Ellingwood & Kinali, 2009). Major earthquakes have been responsible for a death toll of over 60,000 people per year in the last decades, as well as economic losses that can reach a great fraction of a country's welfare (Silva, 2013). In Europe, countries such as Romania, Greece, Turkey and Italy, in particular, have experienced substantial material destruction and loss of life in the past 50 years, despite significant advances in building construction and design standards.

In the United States, the Northridge earthquake of 1994 is perhaps the most important in a series of events registered in California over the past 30 years. It will long be remembered for the unprecedented losses incurred as a result of a moderate-size event, which amount to as much as 40 billion USD, excluding indirect effects (Eguchi, et al., 1998). This makes the Northridge earthquake identically severe to the Kobe event that occurred exactly one year after, in Japan, adding to the reality that developed countries are equally exposed to high earthquake risks. In this context, the Great East Japan Earthquake and tsunami (2011) sent a clear message to countries and regions believed to be prepared to cope with seismic hazard. In fact, the year of 2011 was the most expensive year ever registered, far exceeding the 2005 economic losses, which held the previous record mainly due to the effects of hurricane Katrina. From the overall cost of 380 billion USD, the earthquake disasters in Japan and New Zealand alone accounted for 60% of this value (Silva, 2013).

1.2 Earthquake loss modelling and uncertainties

Earthquake losses have a crippling effect on the economy of affected countries, either by the legal liability of governments to cover the full costs of rebuilding, and/or by the financial burden imposed upon private companies and individuals. In the face of this problem, earthquake loss modelling serves as the foundation to risk prediction and prevention, as a way to mitigate the adverse impacts of future events. A particularly relevant example is the creation of the Turkish Catastrophe Insurance Pool (TCIP), after the enormous financial burden imposed by the 1999 Kocaeli and Duzce earthquakes, due to the country's statuary obligation in covering the costs of reconstruction. By means of this initiative, the creation of an earthquake loss model allowed a large part of the financial risk to be transferred to the world's reinsurance markets, further enabling the evaluation of catastrophe risk impacts in the Turkish economy (Bommer, et al., 2002).

Earthquake loss modelling also serves as the base to many other seismic risk mitigation actions. These may include prioritization of zones within a country where the structural seismic vulnerability of the building stock should be improved, planning of post-disaster emergency response, or definition of mandatory seismic-proof construction practices (Silva, 2013). However, the required resources, datasets and tools seldom exist in a way that is compatible with a comprehensive assessment of seismic risk. Despite of great advances made in the last decades in the areas of: probabilistic seismic hazard assessment (e.g. Abrahamson (2006), Bommer & Abrahamson (2006)), evaluation of building vulnerability (e.g. Calvi et. al. (2006)), and collection of information regarding elements exposed to hazard (e.g. Dell'Acqua et. al. (2013)), several limitations still exist in the way that uncertainties in each of these aspects can consistently be taken into account. In the presence of uncertainties, risk to civil infrastructure from earthquakes cannot be eliminated, but must be managed in the public interest by the entities involved in its evaluation (Ellingwood & Kinali, 2009). Therefore, any meaningful seismic risk or safety assessment must be fully-coupled with its accompanying analysis of uncertainty, be that aleatory or knowledge-based (i.e. epistemic).

Structural reliability concepts and probabilistic risk analysis tools (e.g. Silva et. al. (2014)) provide an essential framework to model uncertainties associated with earthquake prediction, exposure definition, and infrastructure response. In addition to the continued improvement in the characterization of random and epistemic uncertainties in seismic hazard, recent years have seen a major swing in emphasis towards the explicit inclusion

of uncertainties in the performance assessment of single buildings (Bradley, 2013). These have been mainly related with the treatment of sources of uncertainty such as the (hazard-consistent) record-to-record variability and/or the random nature of geometric and structural parameters (e.g. Jalayer et. al. (2010)), in the evaluation of the seismic response. However, in the context of portfolio risk assessment, several important questions remain entirely unanswered or lack a deepest understanding.

The main focus of this thesis consists of the treatment of random and epistemic uncertainties in the hazard-consistent evaluation of building fragility, vulnerability and risk assessments. Matters of statistical significance and spatial correlation of damage and loss are investigated in the context of the probabilistic loss estimation of building portfolios. As a result, innovative methodologies for the treatment of uncertainties are proposed, demonstrating its increased accuracy and robustness when compared with state-of-the-art risk assessment frameworks.

1.3 Objectives and thesis organization

With the aim of addressing several important issues related with the treatment of uncertainties in the earthquake loss modelling of building portfolios, the present thesis is divided into seven chapters. The first and present one consists of an introductory presentation of the various subjects addressed in this thesis, Chapter 7 discusses the main conclusions and possible future developments, and the remaining five chapters can further be grouped into two main subjects: building exposure and vulnerability. With regard to the former, Chapter 2 deals with current limitations in the development of exposure datasets of buildings of industrial use, at the European scale, whereas Chapter 3 to Chapter 6 are concerned with the treatment of random and epistemic uncertainty in the development of fragility and vulnerability models for loss estimation of building portfolios. More specifically, Chapter 3 presents the study of the impact of simplifications generally accepted as state-of-the-art by researchers and practitioners, in the context of seismic hazard disaggregation and record selection for fragility analysis and earthquake loss estimation, Chapter 4 consists of the evaluation of several sources of uncertainty in the process of analytically deriving fragility functions for the loss estimation of building portfolios, Chapter 5 addresses the modelling of spatial correlation of vulnerability uncertainty in portfolio risk analysis, and Chapter 6 deals with the study of the impact of epistemic uncertainty in the development of hazard-consistent fragility models and

corresponding risk estimates. The contents of each chapter are subsequently described in further detail.

The second chapter presents the recent developments in Volunteered Geographic Information (VGI), such as the OpenStreetMap initiative, highlighting the potential of these datasets as supplementary or alternative sources of spatially-based building information. Its increase in usefulness is particularly evidenced when combined with additional Open-Access data such as the CORINE initiative, which provides the geo-referenced distribution of non-residential areas in Europe. In this context, this chapter presents the development of an algorithm that, based on open-access information, provides an automated tool for the development of exposure datasets of industrial buildings in Europe, at the 30 arc-second resolution.

The third chapter of this thesis sheds light on several problems and limitations in current practice of hazard-consistent ground-motion selection and fragility analysis, focusing on the impact that (commonly assumed) approximations in disaggregation outputs have on the distinct risk metrics, as opposed to an exact solution. These issues are investigated for several building typologies, seismicity models and ground motion prediction equations (GMPE), and appropriate guidelines are provided to researchers and practitioners.

In the fourth chapter, it is demonstrated that several questions exist regarding the appropriate treatment of structural capacity and seismic demand variability, from a statistical significance perspective, in the context of the fragility evaluation of building portfolios. As a result, matters such as: the minimum number of ground-motion records necessary for a statistically meaningful evaluation of structural response, the statistical significance of analytically determined damage exceedance probabilities, and the statistically meaningful representation of uncertainty and correlation in the estimation of intensity-dependent damage exceedance probabilities are addressed.

In the fifth chapter, the concepts developed in Chapter 4 are extended to the definition of building Vulnerability, whereby vulnerability functions are characterized by hazard-consistent distributions of *damage ratio* (i.e. ratio between cost to repair and total value of the building) per level of primary seismic intensity parameter. The latter is further included in a loss assessment framework, in which the impact of variability and spatial

correlation of *damage ratio* in the probabilistic evaluation of seismic loss is accounted for, using several building portfolios as test-bed cases. The proposed methodology is evaluated in comparison with current state-of-the-art methods of vulnerability and loss calculation, highlighting the discrepancies that can arise in loss estimates when the variability and spatial distribution of *damage ratio* are not appropriately taken into account.

The sixth chapter consists of the application of the results and methodologies developed in Chapter 3, Chapter 4 and Chapter 5, with the aim of developing a statistically significant framework that allows the hazard-consistent propagation of uncertainty from fragility to loss estimates. In this study, several independent hazard modelling options are considered, in order to infer the repercussion from using fragility functions that are consistent with each hazard modelling approach, on the appraised risk metrics. In light of the appraised results, a methodology for the fragility assessment of building portfolios is presented, in which the epistemic uncertainty of the hazard model can be adequately propagated into the fragility results.

Finally, Chapter 7 summarizes the main conclusions of the present work, providing a description of future developments that, in light of the presented findings, are envisaged in order to improve the results and methodologies addressed herein.

This page is intentionally left blank

Chapter 2 USING OPEN-ACCESS DATA IN THE DEVELOPMENT OF EXPOSURE DATASETS OF INDUSTRIAL BUILDINGS

This chapter is based on the following reference:

Sousa, L.; Silva, V.; Bazzurro, P (2017) Using open-access data in the development of exposure datasets of industrial buildings for earthquake risk modelling. *Earthquake Spectra*, 33(1): 63-84. doi: 10.1193/020316EQS027M

Summary

Recent developments in Volunteered Geographic Information (VGI), such as the OpenStreetMap (OSM) initiative, highlight the potential of these datasets as supplementary or alternative sources of spatially-based building information. Its increase in usefulness is particularly evident when combined with additional Open-Access data such as the CORINE initiative, which provides the geo-referenced distribution of non-residential areas in Europe. However, the systematic application of VGI in the development of exposure models for catastrophe risk assessment has been the subject of limited research. This chapter describes an algorithm that, based on open-access information, provides an automated tool for the development of exposure datasets of industrial buildings in Europe, at the 30 arc-second resolution. Its practical application shows that results obtained at national and regional scales are in excellent agreement with data collected from cadastral agencies in Denmark, Italy and Portugal, which highlights the potential of the algorithm when real building information is scarce or non-existent.

2.1 Introduction

Earthquake impacts in buildings of industrial use have been particularly relevant in recent events, highlighting the importance of a detailed modelling of direct and indirect losses resulting from this type of structures.

The widespread damage to welded steel moment resisting frame systems that happened in different types of buildings including industrial facilities, was one of the

major overall lessons of the Northridge earthquake (Youssef, Bonowitz, & Gross, 1995) that occurred in 1994 in Southern California. In this event, damage occurred in new as well as old buildings, despite the fact that most damaged structures were constructed according to modern codes and standards of practice. In the case of the Kocaeli earthquake, which hit north-western Turkey in 1999, extensive damage to industrial buildings has been reported in a region where approximately forty percent of heavy industry in Turkey was concentrated (Sezen & Whittaker, 2004). The widespread damage to industrial facilities had a substantial impact on the economy of the region in terms of direct losses resulting from structural and non-structural damage, and from indirect losses stemming from business interruption, loss of utilities and loss of transportation infrastructure. The relevance of these findings is extremely important considering that many of the inspected structures were designed in accordance with U.S. and European standards, providing valuable indications as to the likely performance of industrial facilities in other seismically active regions of the world.

In Europe, evidence of satisfactory earthquake response goes hand in hand with reports of collapse of precast buildings, which is the most commonly structural typology used in industrial buildings (Bournas, Negro, & Taucer, 2014). In Italy, specifically, a seismic sequence struck the region of Emilia Romagna in 2012, with two main events on May 20th and May 29th, with local magnitude of 5.9 and 5.8, respectively. These earthquakes caused approximately three quarters of the precast concrete industrial buildings designed with non-seismic provisions in the affected area to suffer significant damage, with one quarter of the total presenting partial or total collapse of the roof. The severe damage that affected these buildings has been probably the most controversial issue raised by these events, given the high level of exposure in terms of human life, building contents, and the importance of the continuity of production processes in the socio-economic activity of the region.

In the context of risk assessment, characteristics such as occupancy type, geometry and material properties are essential to describe the uniqueness of each individual building or building class. However, the quantification of building stock, and the definition of its spatial distribution and structural characteristics are resource intensive and strenuous problems to tackle (Dell'Acqua, Gamba, & Jaiswal, 2013).

Due to the several practical challenges, the engineering and loss modelling community often rely on aggregated statistical data on building stock as the input to the loss estimation procedure. Of course, the various census datasets, collected by national

statistical offices, are an important source of building inventory, providing rigorous and detailed information unlike any other dataset at a national scale. However, because statistical offices do not usually allow public users to access data at the building level, and independent field surveys are impractical for relatively large regions, the building datasets used for the purposes of exposure modelling are usually coarse in their spatial resolution.

Unfortunately, be it parish, municipality, regional or national level, the use of aggregated portfolios can introduce significant errors in risk assessment procedures due to the unbalanced spatial distribution of the building stock with respect to the considered hazard (Bazzurro & Park (2007) and Silva et. al. (2015b)). Thus, disaggregating an exposure dataset at a finer resolution but still compatible with available computational resources is of utmost importance. A suitable level of disaggregation is not self-evident and is also peril dependent. For example, losses estimated for natural events such as earthquakes are still sensitive to the level of detail of the location of the exposed assets (e.g., due to local site effects) but less so than events with narrower and more irregular footprints, such as hailstorms or floods (Chen, et al., 2004).

In the context of earthquake loss estimation, several sources of proxy data are routinely used in the regional disaggregation of coarse exposure datasets. Population density is arguably the more readily available and easy to apply, as demonstrated by Silva et. al. (2015b) in the disaggregation of the Portuguese building stock at parish level based on the population distribution on a 30 arc sec grid. This is a reasonable and well-established approach in the case of residential and public buildings, for which there is an obvious correlation between population density and built-up area. However, this approach is generally not robust for commercial and industrial buildings, which, because of their use, are usually located outside residential areas.

In the case of industrial buildings, not only census data or other type of cadastral information are extremely difficult to obtain at any level of regional aggregation, but also spatial disaggregation methods typically applied in residential building portfolios are not suitable. As a result, alternative approaches need to be explored, particularly with respect to sources of spatial-based building data. In this chapter, attention is given to open-access datasets and Volunteered Geographic Information, including the information provided by the CORINE and OpenStreetMap initiatives in an automated tool for the development of exposure datasets of industrial buildings.

2.2 Objective and area of interest

In this study we describe a methodology that, based on open-access datasets and Volunteered Geographic Information, allows users to automatically develop geo-referenced exposure datasets of industrial buildings in the region of interest. Given the nature of the OSM information, which provides distinct degrees of completeness (Hecht, Kunze, & Hahmann, 2013) in different locations, particular focus is given to the development of methods that are able to identify and overcome possible data shortage, in a statistically meaningful way.

Following a detailed presentation of the datasets, the methodology proposed herein is validated by comparing exposure models derived for Denmark, Italy and Portugal (at different regional scales) with real data independently collected from mapping and cadastral agencies in those countries. In order to demonstrate the algorithm capabilities, exposure datasets are additionally derived for the 36 European countries: Albania, Austria, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Macedonia, Montenegro, Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey and United Kingdom. Results are computed with a spatial resolution of 30 arc-seconds, and further aggregated at the first administrative level in each country, for visual clarity. These databases were utilized to underpin the earthquake loss estimation models developed for Europe by RED and ERN (www.redpavia.com).

2.3 Input data

2.3.1 CORINE

The IMAGE 2000 & CORINE Land Cover project (EEA-ETC/TE, 2002) was launched in 1994 by the European Environmental Agency (EAA) and the Joint Research Centre (JRC) of the European Commission. Its main objective was to provide an up-to-date land cover database, as well as information regarding general land cover changes in Europe between 1990 and 2000 (Steenmans & Perdigao, 2001).

These datasets were derived from satellite imagery and other ancillary data such as aerial photos, digital elevation datasets and hydrology models. As a result, CORINE classifies the European territory into 44 classes at a spatial detail comparable to that of a

paper map on a scale of 1:100,000. Amongst the total set of identified land cover classes, the industrial and commercial units are of interest for this research, as further presented in detail. According to the CORINE technical guide (Bossard, Feranec, & Otahel, 2000), these units correspond to areas mainly occupied by industrial activities of transformation and manufacturing, trade, financial activities and services. For the sake of synthesis, these mixed activity areas are referred herein as non-residential areas, and their location is shown in Figure 2.1.

2.3.2 OpenStreetMap (OSM)

OpenStreetMap (OSM) is a collaborative project founded in 2004 in the University College London, with the aim of creating a free geographic database of the entire world (Ramm, Topf, & Chilton, 2010). Its launch marked a new approach to gathering geodata, made possible by the increasing proliferation of GPS devices amongst private users and by the availability of web-based mapping services. In the case of exposure datasets for earthquake loss assessment specifically for industrial buildings, several georeferenced features provided by OSM are of interest. Available information regarding building footprints and location, building height and land use areas are of evident and increasing usefulness, especially when combined with additional open-access sources such as the CORINE initiative (Bossard, Feranec, & Otahel, 2000).

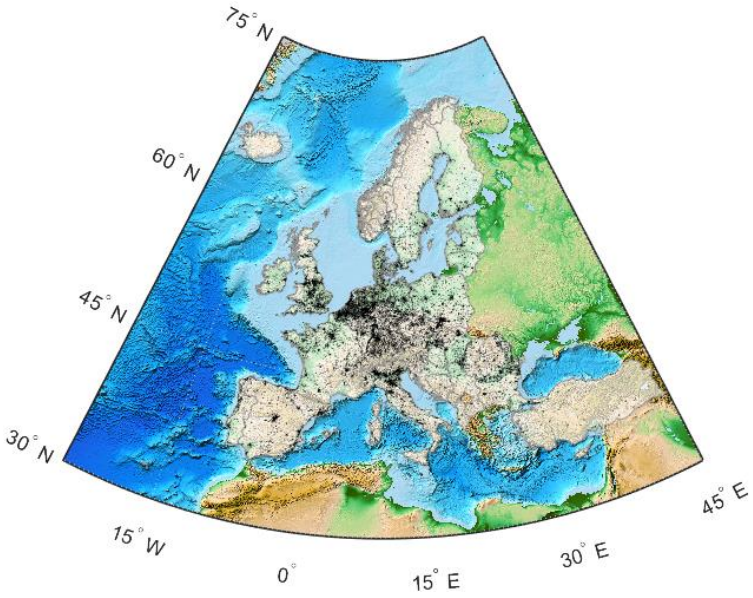


Figure 2.1 – Georeferenced database of non-residential areas (black) provided by the CORINE Land Cover project (CORINE, 2006) for 36 European countries.

Because many sources of geographic data, even those publicly available, are provided with restrictions to their use, OSM's data are distributed under the "Creative

Commons Attribute-ShareAlike 2.0 license", which allows freedom of use by the public (Girres & Touya, 2010).

OSM is probably the most popular and successful VGI initiative, as supported by recent investigations on its completeness and quality. As demonstrated by Haklay (2010) and by Neis *et al.* (2012), urban areas in Central Europe have already been partially mapped with an impressive level of detail, and OSM is well ahead of only mapping the street network. A plethora of spatial data such as roads, buildings, land use areas or points of interest exist in the project's database (OSM Statistics , 2015), emphasizing the potential of its use in the development of catastrophe exposure models in Europe. However, as further highlighted in this chapter, information regarding building height, type of material and age of construction is not provided in a large number of cases. Moreover, each building footprint might enclose several structures, which carries an important limitation when the actual number of assets is of interest.

In this study, data concerning building location and footprint, height, area and land use are used as input for the proposed algorithm, and issues related to the aforementioned data completeness are addressed in their systematic application.

For the sake of illustration, Figure 2.2 presents the industrial land use areas provided by the OSM for the aforementioned 36 European countries, as of October 2015.

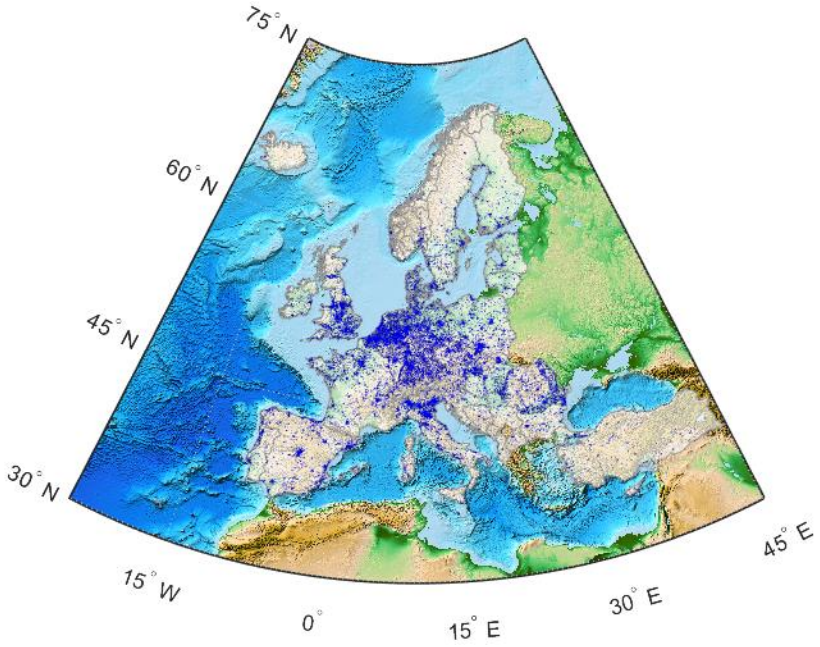


Figure 2.2 – Industrial land use areas provided by the OSM database (blue) for 36 European countries, as of October 2015.

2.4 Methodology and algorithm

The general methodology and the corresponding algorithm developed in this research study are schematically presented in Figure 2.3. Starting from the envelope of non-residential areas containing undifferentiated industrial and commercial units (CORINE), the algorithm identifies areas containing only industrial (not commercial) assets via the *Area Calculator*. The identification is based on the input from OSM, which includes differentiated industrial and commercial areas (see Figure 2.4). Note that in this framework we use OSM information only when complete (i.e., whenever it accurately reflects the reality of the industrial built inventory). Issues concerning data completeness are addressed later in this study.

The various identified polygons containing industrial assets (Figure 2.4b) are further used to determine the footprints of all the industrial buildings inside, as provided by the OSM datasets (Figure 2.5). It shall be mentioned that, in some rare cases, a footprint may include more than one structure, a caveat that should be kept in mind in those rare applications when a precise building count is of interest. However, this is not relevant in the present situation, where only the total footprint area matters.

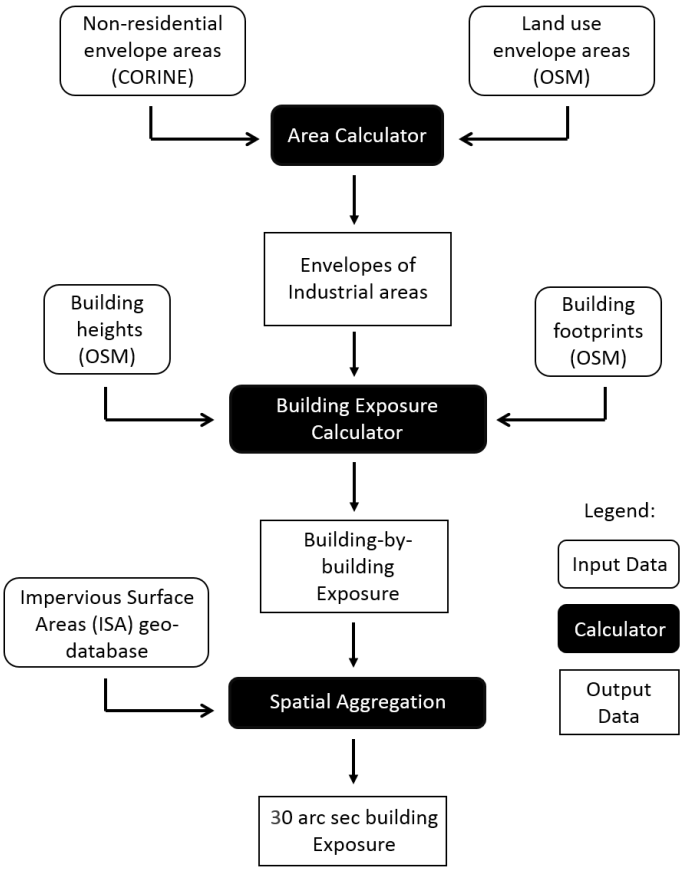


Figure 2.3 – Schematic representation of the main algorithm workflow

The outcome of this process is the total industrial built area in a given region, obtained by inferring the number of floors of each building from the total height associated to the corresponding footprint (extracted from OSM). The final database is assembled by the *Spatial Aggregation Calculator*, in which the total area of all the industrial buildings determined as a function of building's footprint and height are aggregated in a grid of 30 arc sec spacing, such as the one depicted in Figure 2.5b.

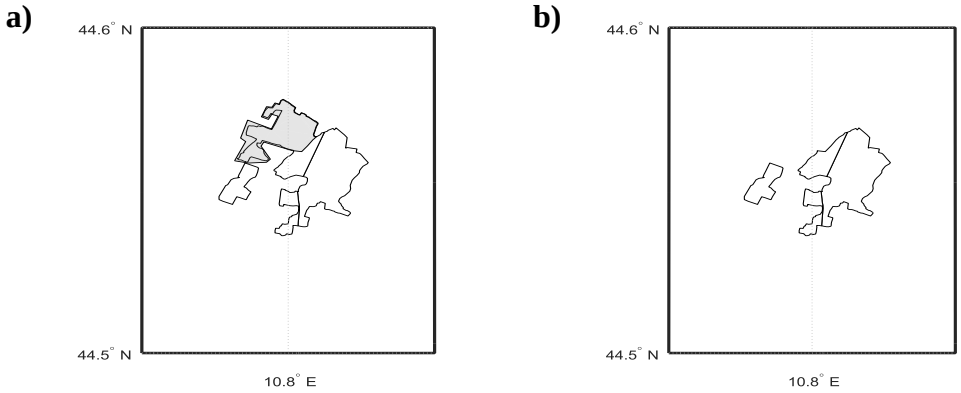


Figure 2.4 – Schematic representation of the operations performed by the *Area Calculator* for an example area: a) Inputs: CORINE non-residential area (white) and OSM commercial area (grey), and b) Industrial area output.

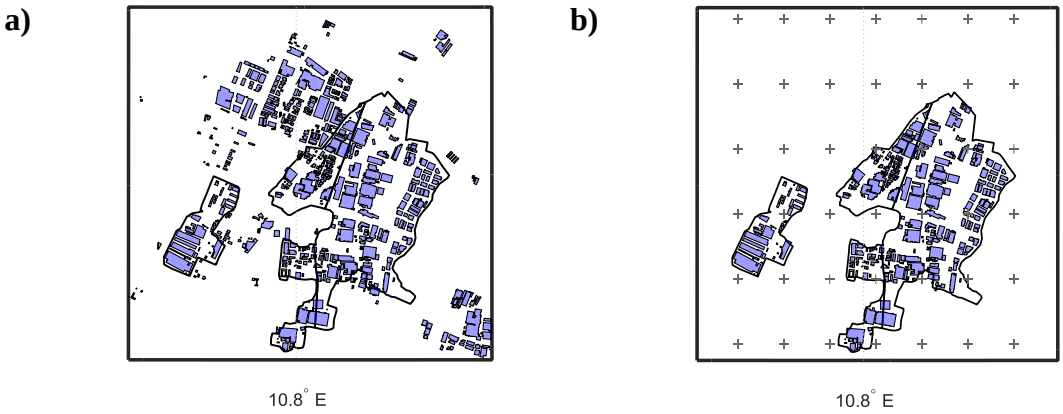


Figure 2.5 – Schematic representation of the operations performed by the *Building Exposure Calculator* for an example area: a) Inputs, and b) Building-by-building output.

2.4.1 Data completeness

In order to overcome potential lack of data in a statistically robust way, particular attention is given to the level of completeness of the datasets. Assuming that a satisfactory level of coverage is provided by CORINE (i.e., all the non-residential areas in Europe are mapped), an "ideal" model would identify both building footprint and height of all the industrial buildings inside the CORINE polygons using OSM as input.

However, such a level of completeness cannot realistically be expected everywhere. At the time of this writing, some areas do have missing data as demonstrated by recent

investigations by Haklay (2010) who reports the level of completeness of the OSM datasets in Europe. Thus, in "data-incomplete" regions for which information provided by OSM is not exhaustive, the "actual" industrial built area in each of the CORINE polygons is estimated using statistical techniques. This is done here according to the optimization algorithm described in Figure 2.6.

Based on the CORINE polygons in which OSM data is complete, *reference* probabilistic joint distributions $f(x, y)$ – are empirically derived for: x = building height, and y = ratio between surface of industrial building footprints and total area of the corresponding CORINE polygon. It is possible that the height, x, of the industrial buildings is statistically dependent on the density of the built area within a polygon, which is expressed by y. It is assumed, however, that x|y is independent on the area of the CORINE polygon. As a result, the *reference* $f(x, y)$ distribution is used to jointly sample the values of these random variables *x* and *y* in the CORINE polygons with incomplete OSM data. This process is shown in Figure 2.6.

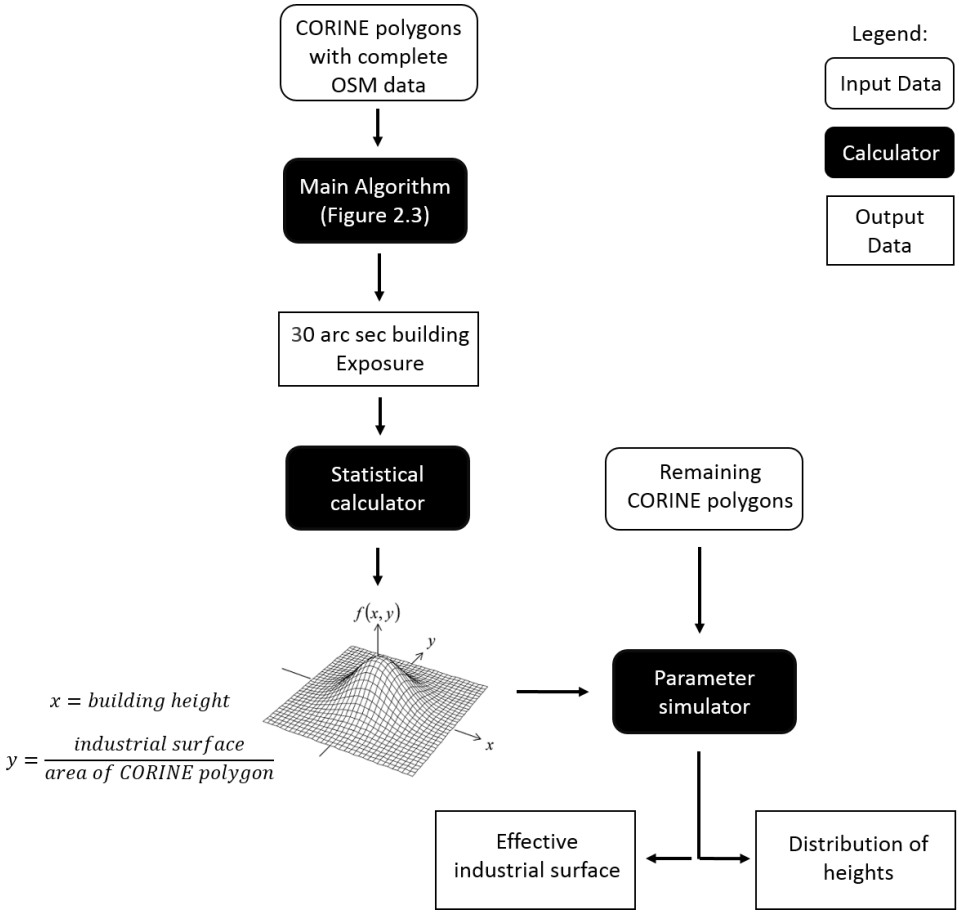


Figure 2.6 – Schematic representation of the *completeness* algorithm workflow

Of course, the robustness of the reference $f(x,y)$ distribution used to simulate y and x|y for the polygons with incomplete data depends on the fraction of polygons within a certain region that have complete data. This fraction in turn depends on the geo-referenced boundaries enclosing regions of OSM data completeness, which are defined by the user as input to the procedure shown in Figure 2.6. In this study, these regions have been manually identified based on the comparison of building footprints and land use with respect to the information retrieved from publicly available sources. In other words, building footprints and corresponding type of use, obtained from OSM, have been compared against cadastral information and satellite imagery (e.g. the Flashearth datasets (2015)), in order to define the limits of the regions in which OSM information can be considered to be in agreement with the real built environment and, therefore, complete.

If the number of CORINE polygons enclosed in the specified boundaries of OSM completeness is not sufficient to ensure that $f(x,y)$ reflects the actual distribution of *x,y* in the country of interest, the final results might not adequately reflect the real building distribution. However, how does the user know if the reference $f(x,y)$ distribution is robust enough? To answer this question an empirical exercise has been devised, to estimate the minimum fraction of polygons with complete data that is necessary to achieve a statistically acceptable estimate of the reference $f(x,y)$ distribution. This exercise is based on the 442 CORINE polygons with complete OSM data located in the region of Emilia Romagna (Italy), which comprises 353 municipalities. This set of 442 polygons, which represent all the industrial areas in the region, can be used to compute the "reference" $f(x,y)$ distribution. Then, it is possible to randomly select many sets, *S*, of increasing number of municipalities (Figure 2.7a) and for each set *S*, only the CORINE polygons inside the selected municipality boundaries are used to derive $f(x,y)_S$. This allows a statistical comparison between $f(x,y)$ and $f(x,y)_S$. However, because $f(x,y)_S$ and $f(x,y)$ computed in this way are not independent - i.e., $f(x,y)_S$ includes observations that exist in $f(x,y)$ as well -, particular care has been employed in the procedure of statistical comparison between samples. As a result, a randomly selected set of 265 out of the 353 boundaries (approximately 75%) is herein considered to be a reasonable approximation to the "reference" $f(x,y)$ distribution, and the CORINE polygons from sets of 1 to the total of remaining municipalities (353-265=88) are used to empirically derive $f(x,y)_S$. This way, samples do not share similar observations, and the

bivariate extension of the non-parametric, two-sample Kolmogorov-Smirnov (*B-KS*) test (Justel, Pena, & Zamar, 1997) has been used for comparing $f(x,y)_S$ and $f(x,y)$.

Loosely speaking, the null hypothesis that $f(x,y)_S$ and $f(x,y)$ are two samples from the same parent distribution is tested by checking that the largest discrepancy between the two cumulative distributions is not too extreme. The so-called *p-values* computed in this B-KS test represent the probabilities that the largest discrepancy could indeed be observed if the two cumulative distributions were sampled from the same parent one. Therefore, if the *p*-value is small, we can conclude that the two distributions do not share the same parent distributions and are indeed different. On the other hand, in practice if the *p*-value is large, we could conclude that $f(x,y)_S$ is a good approximation of $f(x,y)$. Conventionally, if the *p*-values are smaller than a reference value α, say, 10%, then it can be assumed that the largest discrepancy is not generated by the vagaries of the sample but that the two distributions are indeed statistically significantly different.

The *p*-values obtained by simulating 88 sets of 1 to 88 municipalities are plotted as a function of the fraction of polygons with complete data in Figure 2.7b, As illustrated, whenever at least 20% of all the CORINE polygons are enclosed in the boundaries of the selected set of municipalities, the p-values are larger than our reference acceptance value of 10% and in those cases we can consider that the corresponding $f(x,y)_S$ is an acceptable approximation of the "reference" $f(x,y)$.

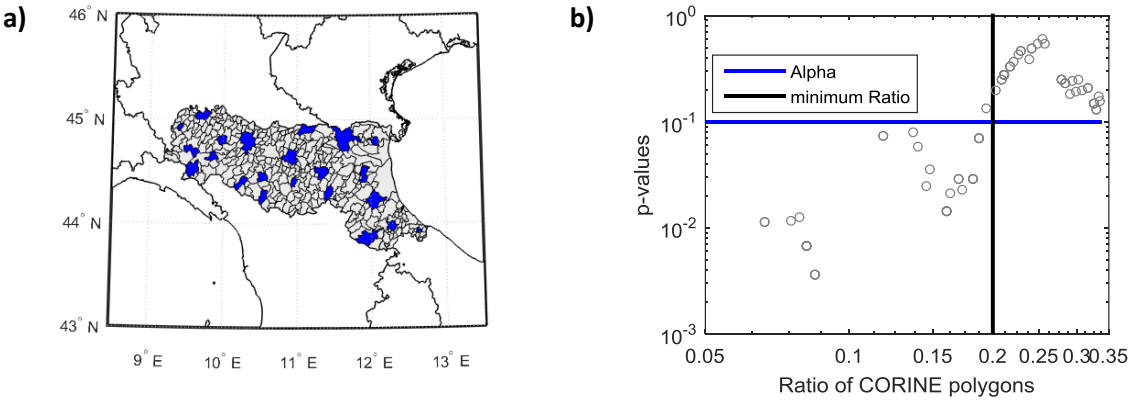


Figure 2.7 – a) Region of Emilia Romagna, Italy (grey), and one simulated set *S* of 22 administrative boundaries, (blue) randomly selected from the total of 353 municipalities, b) *p-values* of the *B-KS* test as a function of the fraction of polygons with complete data

Of course, fully acknowledging the limitations inherent in a single test, it is herein considered as a "rule of thumb" that when at least 20% of the total number of CORINE polygons in any region of Europe are enclosed in the selected "data-complete" boundaries, then the empirically derived $f(x,y)_S$ is acceptable for the purposes of

developing an industrial building inventory. It should be noted, however, that this is only valid if the selection of data-completeness boundaries is done randomly within the assessed country or region (Figure 2.7a), in order to avoid bias in the corresponding statistics.

The final stage of the correction process (Figure 2.6) consists of disaggregating the total built areas computed for each "data-incomplete" CORINE polygon in a grid of 30 arc sec. To do so, the areas of industrial activity in each CORINE polygon are distributed proportionally to the spatial density of constructed impervious surfaces (ISA, 2010). The dataset of ISA values, defined in a grid of 30 arc sec resolution, provides ratios that, for each grid cell, reflect the proportion of the cell area that is occupied by impervious surfaces resulting from man-made constructions such as buildings and roads. This way, within a given CORINE polygon, the computed area of industrial buildings is distributed amongst the centroids of the ISA cells enclosed by the CORINE boundary, proportionally to the corresponding ISA values.

2.5 Validation

For the purpose of validation, the methodology and assumptions presented above are applied to three distinct regions of Europe – Denmark, Portugal and Emilia Romagna (Italy) – for which cadastral information (referred herein as "real" data) is available at different resolutions. In the cases of Denmark and Portugal, total built areas have been collected for 98 municipalities and 18 districts, respectively, as presented in Figure 2.8a and Figure 2.8b. Instead, the data about industrial buildings in Emilia Romagna were available at the building level in a geo-referenced environment and were later aggregated according to the administrative boundaries of the 353 municipalities for the purpose of this test, as shown in Figure 2.8c.

The results of the proposed algorithm have been aggregated in the same spatial resolution as the available information, in order to provide the means for a direct comparison. The industrial building areas for 98 municipalities in Denmark inferred by our methodology (Figure 2.9) show a remarkable agreement with the "real" areas depicted in the same figure. In order to further assess this relationship, particularly the potential bias of predictions with respect to "real" values, the scatter of results is presented in Figure 2.10a, as a function of the corresponding collected data.

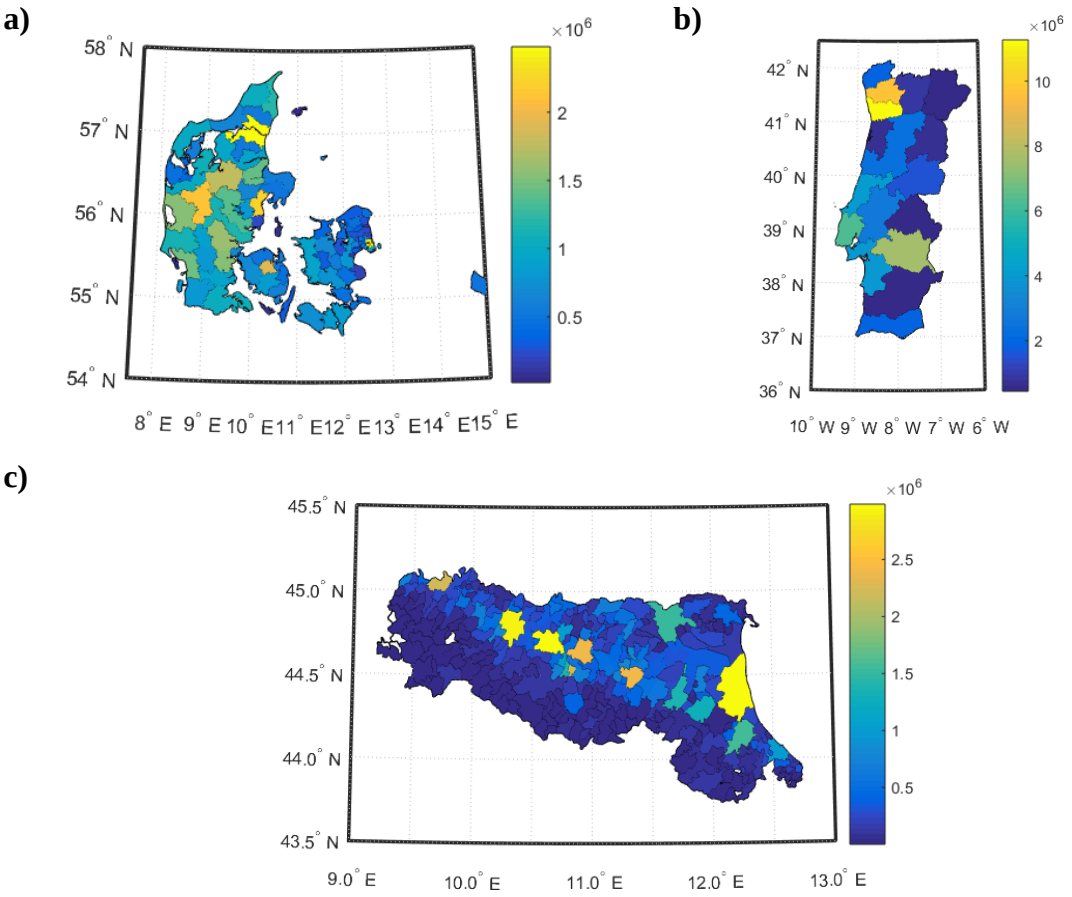


Figure 2.8 – a) Total industrial built area in Denmark (98 municipalities), as obtained from Statistics Denmark (2015), b) Industrial built area of 18 districts in Portugal, according to PRISE (2013-2015) and Araújo *et. al.* (2015), and c) industrial built area in Emilia Romagna, Italy (353 municipalities), as obtained from Geoportale Emilia Romagna (2015).

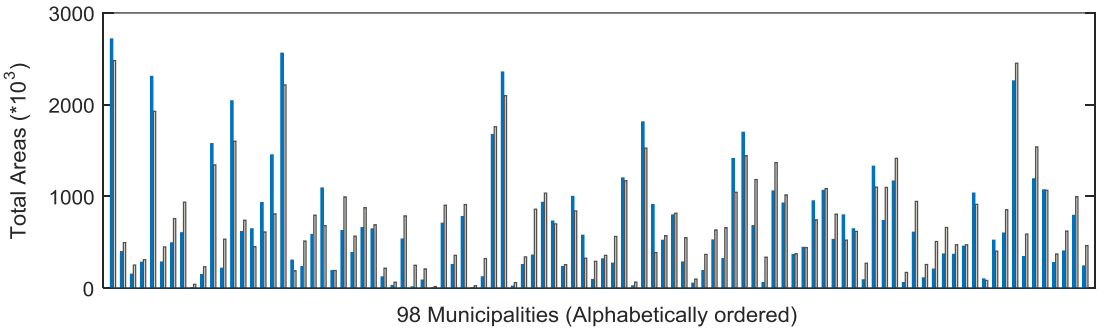


Figure 2.9 – Bar chart comparing "real" (blue) and predicted values (grey) of total built area of industrial buildings, for 98 municipalities in Denmark

Figure 2.10b shows that our methodology, in this specific case, tends to over-predict the total built area when the values are lower than $100*10^3$ m^2. The agreement improves significantly for industrial areas with larger built areas. The reasons for this bias for low-built area cases, which appears is this case only, is not known and is not a defect of the proposed methodology, as will made clear by the following two examples.

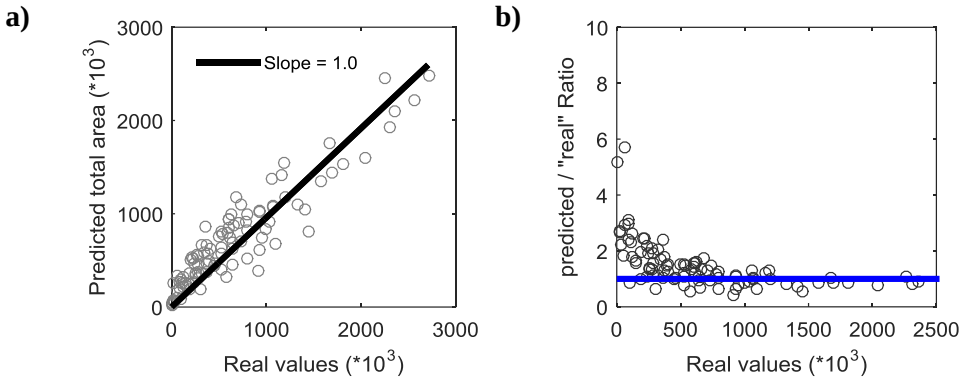


Figure 2.10 – a) Inferred versus "real" values of industrial building built areas for 98 municipalities in Denmark, and b) corresponding ratio between predicted and "real" and values.

In the application to Portugal, results have been aggregated at the district level, as depicted in Figure 2.11. Again, the agreement between inferred and "real" data is very good and, in this case, for all values of built areas, large and small (Figure 2.12a and Figure 2.12b).

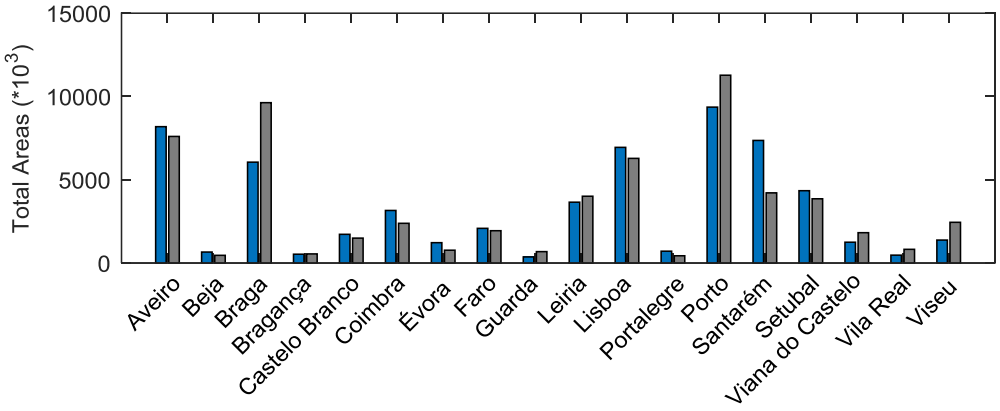


Figure 2.11 – Bar chart comparing real (blue) and predicted values (grey) of total built area of industrial buildings, for 18 districts in Portugal

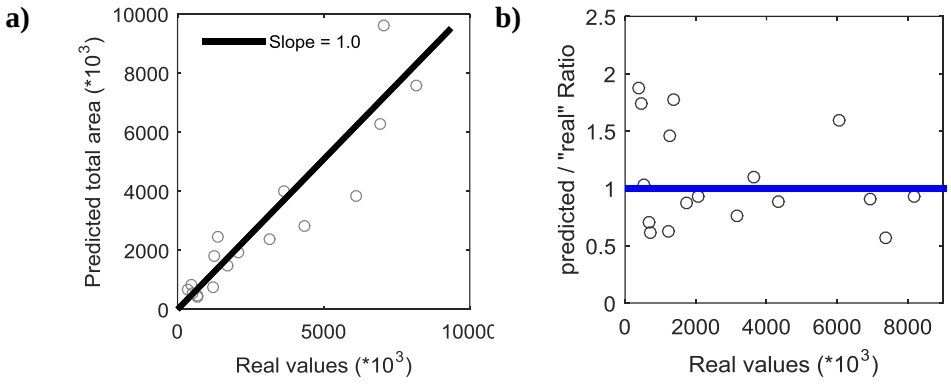


Figure 2.12 – a) Inferred versus "real" values of industrial building built areas for 18 districts in Portugal, and b) corresponding ratio between predicted and "real" and values.

A similar unbiased outcome has been reached for the third and last example that deals with Emilia Romagna. Figure 2.13 presents the aggregated results at the province

level, whereas Figure 2.14 illustrates the relationship between estimates in 353 municipalities. Again, in this case as well there is no obvious indication that the methodology provides biased low estimates for low-value built areas especially when looking at the municipality results shown in Figure 2.13a and Figure 2.13b.

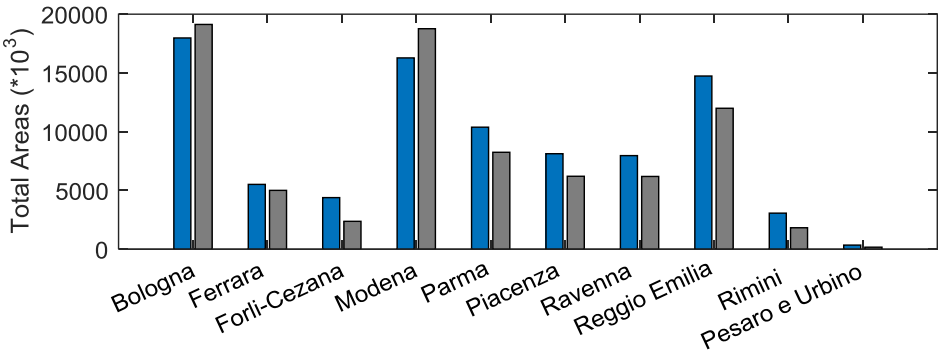


Figure 2.13 – Bar chart comparing "real" (blue) and predicted values (grey) of total built area of industrial buildings, for 10 provinces in the region of Emilia Romagna, Italy

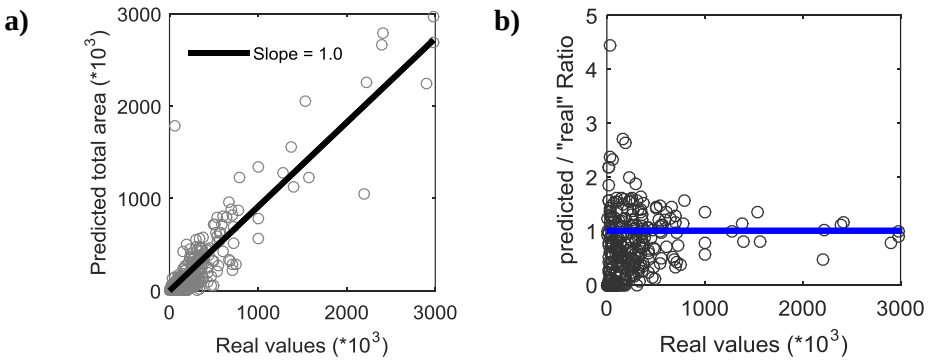


Figure 2.14 – a) Inferred versus "real" values of industrial building built areas, for 353 municipalities in the region of Emilia Romagna, Italy, and b) corresponding ratio between predicted and "real" and values.

2.6 Application and results

An exposure model for 36 countries in Europe has been created, reflecting the total built area of industrial buildings in an evenly spaced grid at 30 arc sec. The average building density in the region has been estimated to range from 500 $^{m^2}/_{km^2}$ or less in countries such as Austria, Croatia, Finland, Sweden and Turkey, to values of 1000 to 1500 $^{m^2}/_{km^2}$ in France, Germany, Denmark, Italy, Czech Republic, Hungary, Poland, Portugal, Romania, Switzerland and the United Kingdom. Overall maximums of approximately 5000 and 7000 $^{m^2}/_{km^2}$ are found in Belgium and the Netherlands, respectively. These indicators represent the ratio between total area of industrial buildings

in a given country and the area of its territory. Figure 2.15 illustrates the aggregation of results at the first administrative level in each country, for visual clarity.

The loss estimation of the building portfolios is beyond the scope of this research, in which we did not tackle the issue of vulnerability to ground shaking of buildings with different characteristics in the various countries. However, we qualitatively evaluated the level of risk in Europe through the assessment of what is herein referred as "exposure at risk". To this aim, the probabilistic seismic hazard results provided by the SHARE project (Giardini, 2013) have been employed in order to determine the proportion of industrial areas subjected to ground motion intensities with different mean return periods of exceedance.

Since structural properties are expected to be significantly heterogeneous and, therefore, enclose a wide range of dynamic behaviour, ground motion intensity is herein characterized in terms of peak ground acceleration (PGA). Figure 2.16, shows the mean hazard map for PGA values an exceedance probability of 10% in 50 years, or, conversely, a 475-year mean return period (MRP). This threshold has been selected because it is relevant to engineering practice, since most standards define such MRP as the basis for selecting design seismic actions.

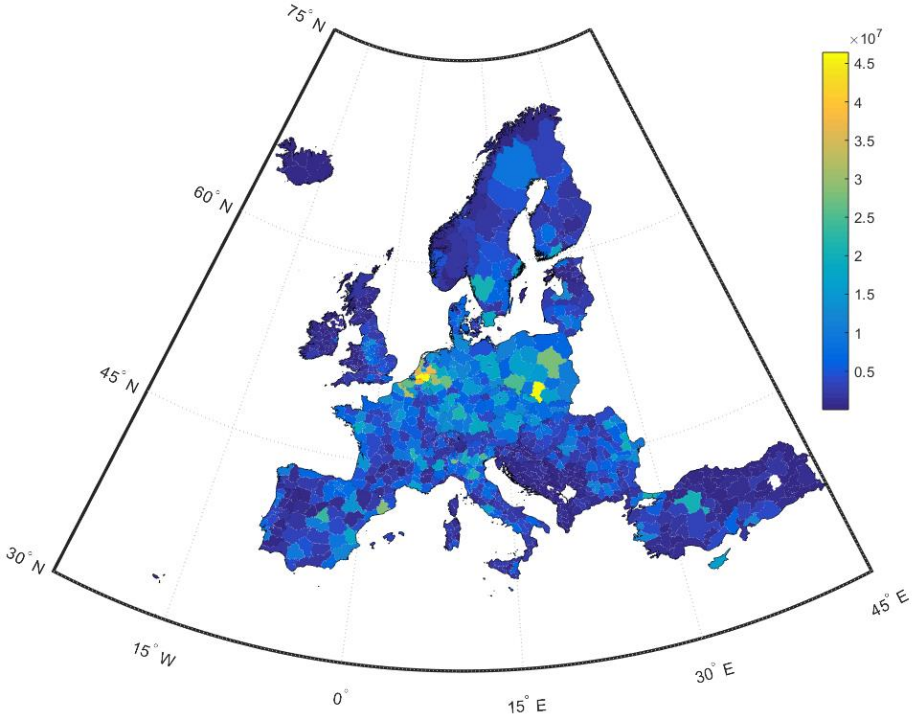


Figure 2.15 – Exposure dataset of industrial building areas (m^2) developed for 36 countries in Europe, aggregated at the first administrative level (m^2).

Figure 2.17 summarizes the hazard shown on Figure 2.16 by differentiating territories subjected to low, medium and high seismicity that are associated here to threshold values of 0.05g, 0.15g, and 0.30g, respectively, for the aforementioned MRP. These contours are plotted on top of the industrial exposure data shown in Figure 2.15.

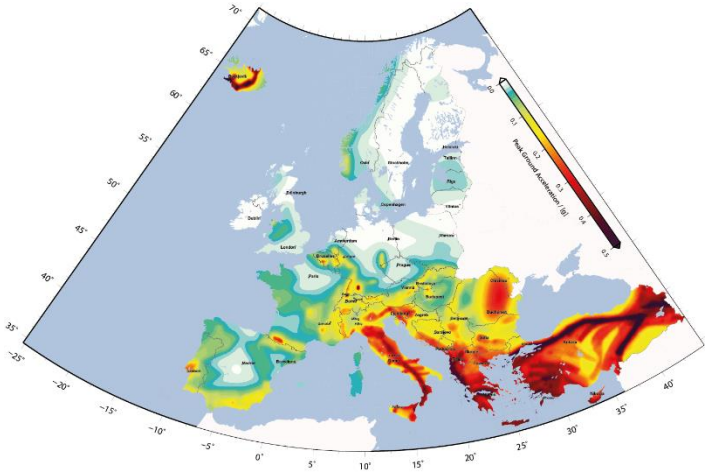


Figure 2.16 – European hazard map of peak ground acceleration (PGA) with 10% exceedance probability in 50 years (in units of g), adapted from SHARE (2015)

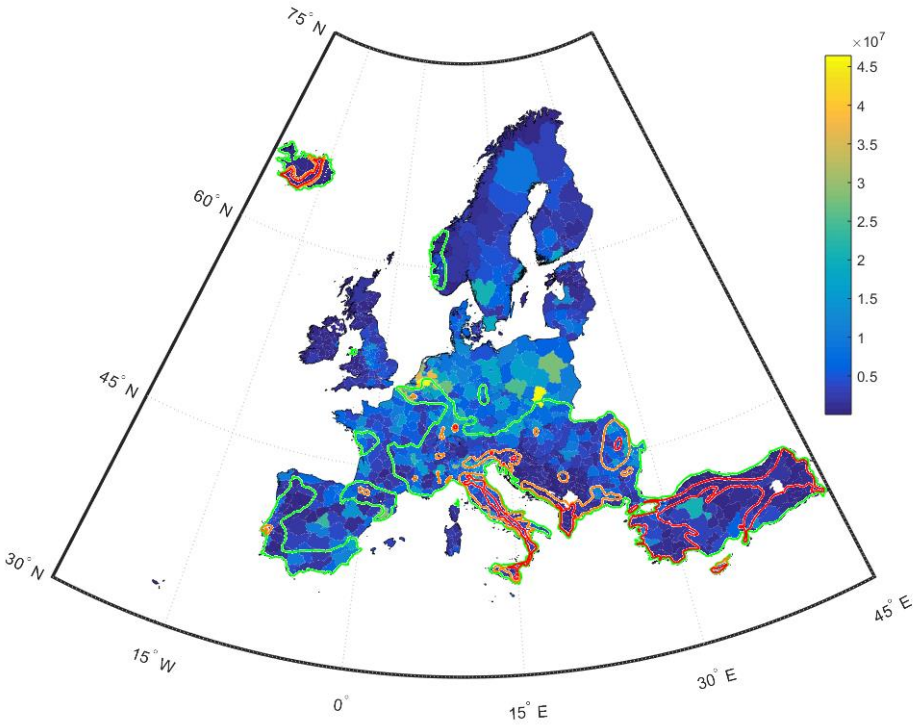


Figure 2.17 – Illustration of industrial "exposure at risk" in Europe. The solid contour lines enclose regions where PGA values of 0.05g (green), 0.15g (orange), and 0.30g (red), respectively, are expected to be exceeded with 10% probability in 50 years.

The majority of highly industrialized regions in Germany, France, Poland, the Netherlands and the United Kingdom is subjected to low seismic hazard characterized by PGA lower than 0.05g at the 475 year MRP level. Italy, on the other hand, which

possesses one of the largest building portfolios, is characterized by medium to high seismic hazard and, therefore, has the highest fraction of industrial "exposure at risk". Additional critical regions are those of Romania and part of Turkey, as determined by the conjunction of high seismic hazard and exposure.

As a summary of "exposure at risk", Table 2.1 shows the aggregated area of industrial assets in the metropolitan boundaries of 36 European capitals along with the corresponding mean level of PGA with 10% probability of exceedance in 50 years.

Table 2.1– Total area of exposed assets in 36 European capitals, and corresponding hazard level defined by the peak ground acceleration (PGA) with 10% exceedance probability in 50 years

Country	**Capital**	**Industrial area (10^3 m^2)**	**PGA (g)**
Albania	Tirana	402.9	0.21
Austria	Vienna	9407.0	0.11
Belgium	Brussels	33491.3	0.06
Bosnia and Herzegovina	Sarajevo	1299.2	0.13
Bulgaria	Sofia	5635.0	0.28
Croatia	Zagreb	2567.0	0.24
Cyprus	Nicosia	1602.3	0.26
Czech Republic	Prague	5111.0	0.01
Denmark	Copenhagen	9344.2	0.02
Estonia	Tallinn	4023.9	0.02
Finland	Helsinki	11654.6	0.01
France	Paris	50762.6	0.01
Germany	Berlin	10606.7	0.01
Hungary	Budapest	10198.6	0.08
Iceland	Reykjavik	1697.9	0.46
Ireland	Dublin	7286.3	0.01
Italy	Rome	6121.8	0.24
Latvia	Riga	6236.5	0.03
Lithuania	Vilnius	3066.9	0.01
Luxembourg	Luxembourg	643.7	0.03
Malta	Valetta	3232.0	0.06
Macedonia	Skopje	1809.4	0.26
Montenegro	Podgorica	402.7	0.01
Netherlands	Amsterdam	7398.7	0.02
Norway	Oslo	5705.2	0.02
Poland	Warsaw	11196.9	0.02
Portugal	Lisbon	4331.3	0.24
Romania	Bucharest	11286.4	0.25
Serbia	Belgrade	3870.5	0.11

Table 2.1 (continued)

Country	Capital	Industrial area (10^3 m^2)	PGA (g)
Slovakia	Bratislava	3166.7	0.01
Slovenia	Ljubljana	2972.3	0.25
Spain	Madrid	10974.6	0.01
Sweden	Stockholm	9828.4	0.01
Switzerland	Bern	1185.4	0.02
Turkey	Istanbul	21147.1	0.42
United Kingdom	London	24999.0	0.01

The results presented in Table 2.1 allow the identification of the European areas in which particular attention should be devoted during vulnerability and loss assessment studies. In addition, Figure 2.18 and Figure 2.19 rank the top 10 medium and high seismicity countries based on the total industrial areas subjected to different ground motion intensities at the 475-year hazard level. The former figure shows the ranking in absolute values of total areas exposed while the latter shows the ranking by area exposed normalized by the total area of industrial buildings in each country.

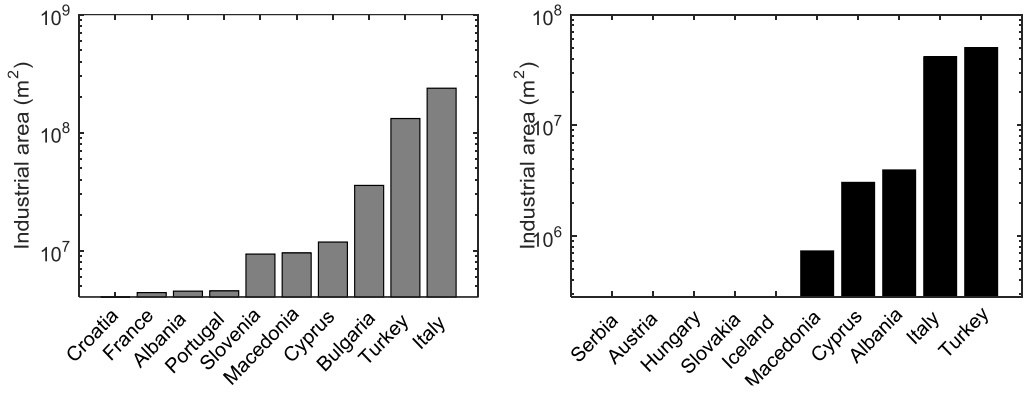


Figure 2.18 – Total area of industrial assets exposed to PGA of 0.15g (grey) and 0.30g (black) or higher with a mean return period of 475 years. Graph include only the top 10 countries

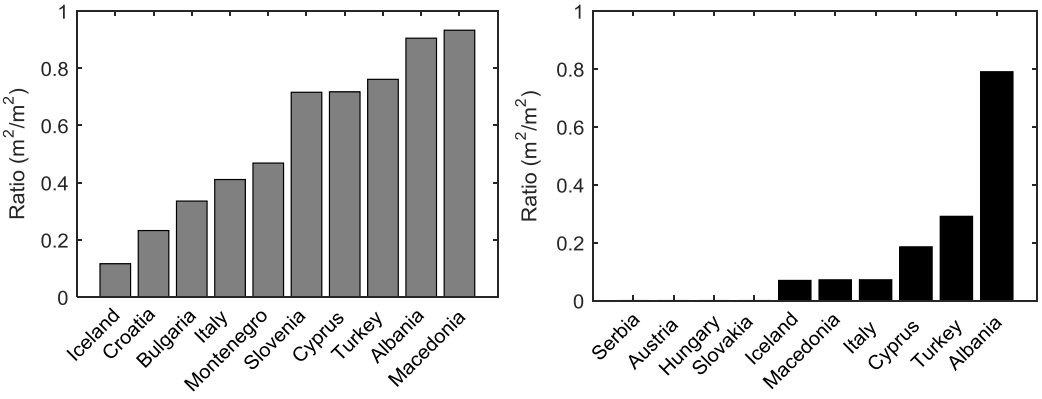


Figure 2.19 – Same as in Figure 2.18 but normalized by the total area of industrial buildings in each country.

2.7 Limitations and caveats

The proposed procedure delivers a dataset of industrial buildings with no differentiation of building typology and associated replacement cost. For loss estimation purposes, however, the final output of this procedure needs to be further refined by differentiating buildings in terms of construction material, lateral load resisting system, and unit replacement cost. This additional step is not tackled here. The focus of this research is to provide users with a methodology that is able to provide accurate estimates of industrial building exposure, as determined by the geo-referenced distribution of construction areas. While not sufficient for the purposes of earthquake loss estimation, the definition of a reliable geo-referenced building exposure is arguably the most time consuming and challenging task in the process of building a full exposure model that also includes the differentiation in building classes and related replacement costs. Building characteristics and constructions costs can be inferred from several open-access sources such as census data and information provided by various statistical offices. Hence, the presented algorithm and methodology are intended to provide the means to overcome significant technical difficulties in a process where lack of reliable data is widely acknowledged by practitioners.

2.8 Final remarks

This chapter presented a methodology for the development of exposure datasets of industrial buildings, using Volunteered Geographic Information (VGI) and Open-Access data as the main sources of information. This framework was implemented in an algorithm that provides an automated tool for the development of such models at the 30 arc sec resolution, providing the means to overcome the potential lack of input data in a statistically meaningful way. In its development, data completeness was assessed through an optimization process that uses spatial statistics in order to complement the necessary inputs, in regions where such data does not adequate reflect the real building environment.

This procedure was validated using exogenously provided data on industrial buildings in three distinct regions in Europe - Denmark, Portugal and Emilia Romagna, Italy. More precisely, the validation entailed a comparison of the exposure estimates obtained with this procedure with cadastral information collected at different resolutions. The findings of this exercise demonstrated the excellent agreement between real and inferred exposed areas.

In order to demonstrate its capabilities, the presented framework was used to develop exposure models of industrial buildings in 36 European countries, at the 30 arc sec resolution. The spatial distribution of industrial building area in Europe was then coupled with the probabilistic seismic hazard results provided by the SHARE initiative (Giardini, 2013) in order to present a qualitative estimate of industrial "exposure at risk". As a measure of this indicator, areas of industrial assets in the metropolitan boundaries of 36 European capitals have been determined, and further compared with the corresponding mean level of PGA for a 475-year mean return period of exceedance. Turkey and Italy are the two countries with the largest number of industrial assets located in medium and high seismicity areas.

This page is intentionally left blank

Chapter 3 HAZARD DISAGGREGATION AND RECORD SELECTION FOR FRAGILITY ANALYSIS AND EARTHQUAKE LOSS ESTIMATION

This chapter is based on the following reference:

Sousa, L.; Marques, M.; Silva, V.; Varum, H. (2017) Hazard disaggregation and record selection for fragility analysis and earthquake loss estimation. *Earthquake Spectra*, 33(2): 529-549. doi: 10.1193/062016EQS101M

Summary

Economic losses and collapse probability are critical measures for evaluating the safety of existing buildings. In this context, the study presented in this chapter sheds light on several problems and limitations in current practice of hazard-consistent ground-motion selection and fragility analysis, focusing on the impact that (commonly assumed) approximations in disaggregation outputs have on the aforementioned risk metrics, as opposed to an exact solution. These issues are investigated for several building classes, seismicity models and ground motion prediction equations (GMPE), for a site in the city of Lisbon (Portugal). It is observed that only an exact (i.e. rupture-by-rupture) disaggregation can lead to satisfactory results in terms of accuracy, when limit state criteria are not structure-specific. On the other hand, an approximate method is proposed, which still leads to statistically valid results regardless of the chosen structural class, seismicity model or GMPE.

3.1 Introduction

Earthquake-induced collapse and economic losses are critical measures for evaluating the safety of existing buildings, as well as assessing the effectiveness of risk reduction schemes (Liel, 2008). As a result, such metrics must be as accurate as possible, appropriately reflecting the various sources of uncertainty associated with the evaluation of seismic risk. Otherwise, biased, unreliable or inadequate results may be achieved, resulting in ill-informed decisions.

With respect to seismic fragility and vulnerability, the largest source of uncertainty lies in the characterization of the input earthquake ground motion (Liel, Haselton, Deierlein, & Baker, 2009). Given a particular level of seismic intensity, these uncertainties are associated with several ground motion attributes that influence the designated 'record-to-record' variability. Therefore, in this study we focus on the appropriate treatment of record-to-record variability in the context of fragility and vulnerability evaluation. In particular, we investigate what is the impact on earthquake loss estimations, when different degrees of accuracy are considered in the hazard-compatible ground motion selection for non-linear response history analysis.

In the context of performance-based engineering, the selection of natural (or synthetic) ground motion records provides the link between the seismology and the earthquake engineering counterparts. In other words, an appropriate record selection shall be consistent with the seismic hazard at the site at which the engineered system of interest is located. A rigorous approach requires both the determination of a 'target' to compare the appropriateness of different ground motions, as well as an objective method for the selection, simulation and/or modification of ground motions to 'match' this 'target' (Bradley, 2010a).

Despite the consensus on the differences between the uniform hazard spectrum (UHS) resulting from probabilistic seismic hazard analysis (PSHA) and individual earthquake scenarios, the UHS is still widely used as the basis for record selection and scaling. However, given the limitations associated with the UHS (e.g. Baker (2011), NIST (2011)), several methodologies have recently been proposed in the literature, providing alternatives that allow the direct link between ground motion properties and PSHA. Amongst many studies, the conditional spectrum (CS), initially proposed by Baker & Cornell (2006b) and further developed by Jayaram, Lin, & Baker (2011), provides the mean and variance of spectral ordinates conditioned on the occurrence of a specific value of a single spectral acceleration, as determined by PSHA. This method has its fundamental basis in the assumption that spectral accelerations follow a multivariate lognormal distribution. As hypothesized by Bradley (2010a), this is not restricted to spectral accelerations, and can be extended to any arbitrary vector of ground motion intensity measures of interest (*IM*). As a result, the proposed general conditional intensity measure approach (GCIM) establishes that, for a given earthquake rupture (or scenario) discretized in the source model used in PSHA, a conditional vector *IM* has also a multivariate lognormal distribution. \

According to the GCIM approach, considered in this research, an exact estimation of the 'target' distributions of several intensity measures used for record selection and scaling shall be obtained as the contribution of all the possible (independent) ruptures influencing the hazard at a given site. As such, the contribution of each rupture conditioned on a particular level (of a given intensity measure) can be defined by seismic hazard disaggregation (Bazzurro & Cornell, 1999). Because this "rupture-by-rupture" disaggregation is computationally demanding, it is not available as a standard output of most PSHA tools. Therefore, the simplification commonly used consists in grouping individual ruptures into 'rupture scenarios' usually defined by a pair of causal magnitude (M) / distance (R). This approach is adopted in virtually all hazard disaggregation platforms. However, the impact of considering the hazard contribution of 'rupture scenarios' (e.g. M/R intervals) in the computation of 'targets' for record selection (e.g. Lin *et. al.* (2013)), as opposite to an exact 'rupture-by-rupture' disaggregation, has been subject of limited scrutiny. More specifically, the level of error introduced by considering this approach instead of a more robust 'rupture-by-rupture' disaggregation is still not clear.

In this research, the fragility, vulnerability and loss assessment of 9 distinct structural typologies is performed, using sets of ground motion records selected according to the GCIM approach. Target distributions are computed based on rupture contributions determined by seismic hazard disaggregation for Lisbon, Portugal, and several degrees of approximation are used, ranging from: a) the exact consideration of all the possible ruptures contributing to hazard, to b) a commonly used approach in which more coarse rupture scenarios are grouped into magnitude / distance bins. With respect to the latter, different M / R intervals are considered, in order to determine the level of error introduced by different approximations, as well as possible suitable levels of approximation recommended to be used by the research and practitioner communities. Furthermore, in this framework, distinct source models and ground motion prediction equations (GMPE) are considered, in order to assess how sensible the probabilistic loss estimation results are to the aforementioned approximations, when different GMPEs and seismicity models are used.

3.2 Numerical models

The numerical models considered herein correspond to the most typical class of buildings constructed in Portugal at different time periods: reinforced concrete (RC) structures with masonry infills. Drawing upon the study by Silva *et al.* (2015c), in which material and geometrical properties of Portuguese building classes were characterized, statistical distributions of such properties have been used to create synthetic sets of 100 structures per building class, from which the numerical model corresponding to the median capacity has been selected, for each class. To account for the evolution of seismic design and its effect on seismic response, three different classes have been defined: *pre-code, mid-code,* and *post-code*. The first refers to buildings constructed before 1958 (i.e. previous to the publication of the first simplified seismic design code), *mid-code* buildings have been designed between 1958 and 1983 with a more rigorous code, and the last category includes structures built after 1983, time at which the modern seismic regulations were enforced in Portugal. In addition, three building heights per code class have been considered (2, 5 and 8 floors), leading to a total of 9 different building classes (Table 3.1).

Table 3.1– Considered building classes and corresponding fundamental periods of vibration

Acronym	Code class	Number of floors	Mean period (sec.)
PC-2	Pre-code	2	0.260
PC-5	Pre-code	5	0.450
PC-8	Pre-code	8	0.700
MC-2	Mid-code	2	0.312
MC-5	Mid-code	5	0.683
MC-8	Mid-code	8	0.705
C-2	Post-code	2	0.316
C-5	Post-code	5	0.521
C-8	Post-code	8	0.615

Dynamic properties are characterized by the mean fundamental periods of vibration obtained from random generation of 100 structures with varying geometrical and material properties. These have been found to range from 0.260 to 0.705 seconds, as presented in Table 3.1. The percentage of reinforcement in the beams and columns is calculated following the applicable code regulations and practices for ultimate and serviceability limit states, in accordance with the sampled geometrical and material characteristics.

To maintain the computational effort at a reasonable level, each structure is modelled as a single infilled moment frame with three bays. As schematically presented in Figure 3.1, for the case of 5 story buildings, each frame was modelled in a 2D environment using the open-source software OpenSees (McKenna, Fenves, Scott, & Jeremic, 2000), with force-based distributed plasticity beam-column elements.

For the sake of synthesis, readers are referred to the aforementioned work by Silva *et al.* (2015c) for details on the numerical considerations adopted with regards to the cross section discretization and integration points of the elements, the material constitutive relationships, P-delta effects, the infill panel modelling approach, and applicable design provisions.

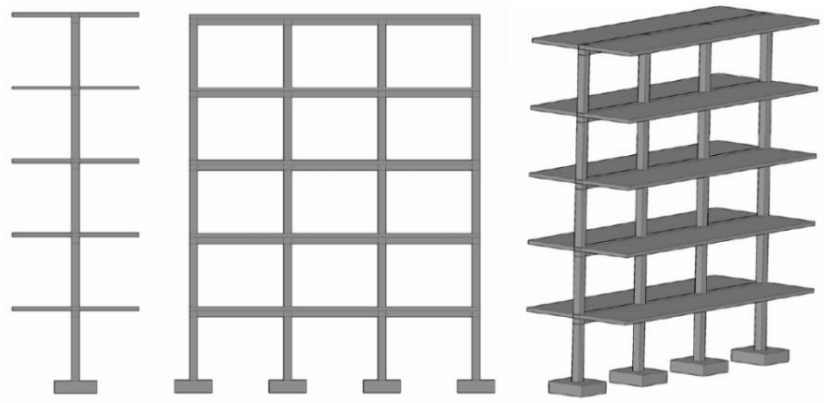


Figure 3.1 – Schematic view of the five-story RC frame model: front (left), side (centre) and isometric view (right) without infills, adapted from Silva et. al. (2014).

3.3 Methodology

The methodology implemented in this research consists of: a) seismic hazard and disaggregation for the site of interest – Lisbon, Portugal, b) record selection following the GCIM framework, compatible with the disaggregation results appraised in a), c) non-linear response history analysis (NLRHA) for the 9 building models (one per building class) subjected to each set of records selected in b), d) fragility assessment for each building class and set of ground motion records resulting from distinct discretization methods used in disaggregation, and e) seismic vulnerability and probabilistic loss estimation for each fragility model defined in d). The final stage of the assessment includes an evaluation of fragility functions obtained for each level of discretization used

in seismic hazard disaggregation, for each structural typology, as well as the probabilistic comparison between the corresponding losses.

3.3.1 Description of the probabilistic seismic hazard analysis models

In order to account for the epistemic uncertainty associated with the definition of the seismicity in the area of interest, PSHA is performed based on four distinct source models. The first (referred herein as VF-model) has been obtained from the study of Vilanova & Fonseca (2007), in which the Portuguese seismic catalogue has been reviewed in order to define a new seismic zonation, whilst the remaining three have been developed within the FP7 SHARE project – a collaborative program supported by the European Commission, in which a community-based seismic hazard model for the Euro-Mediterranean region has been developed (Woessner, et al., 2015). The latter includes an area source model (AS-model) based on the definition of areal sources for which earthquake activity is defined individually, a kernel-smoothed zonation-free stochastic earthquake rate model (Hiemer, et al., 2014) that considers seismicity and accumulated fault moment (SEIFA-model), and a fault source/background seismicity model (FSBG-model), based on the identification of large seismogenic sources using tectonic and geophysical evidence (Haller & Basili, 2011).

For consistency with the aforementioned studies, hazard and disaggregation calculations have been performed using the ground motion prediction equations (GMPE) of Akkar & Bommer (2010) and Chiou & Youngs (2008), separately, for each source model. These equations correspond to GMPEs to which the experts involved in the SHARE initiative attribute higher degree of confidence for application in the two tectonic environments applicable to Portugal: Active Shallow and Stable Continental Crust (Delavaud, et al., 2012).

It is acknowledged that the consideration of only two ground motion prediction equations is not sufficient to capture the effect of epistemic uncertainty in this region. However, the present objectives are: a) to verify the impact of different attenuation relationships and corresponding set of rupture defining parameters in the computation of hazard-consistent fragility, for a given seismicity model, and b) assess the influence of different seismicity models on the hazard-consistent fragility, when considering GMPEs with different degrees of detail in the definition of rupture properties (Table 3.2).

Table 3.2– List of rupture parameters in Akkar & Bommer (2010) and Chiou & Youngs (2008). R_{rup} is the closest distance to the rupture surface, R_x is the shortest horizontal distance to a line defined by extending the fault trace to infinity in both directions, and R_{jb} is the Joyner-Boore distance.

GMPE – acronym	Mw	Rake / fault mechanism	Rupture Depth	Dip angle	R_{rup}	R_x	R_{jb}	Hanging / foot wall
Akkar & Bommer (2010) – AB10	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Chiou & Youngs (2008) – CY08	Yes	Yes	-	-	-	-	Yes	-

3.3.2 Seismic hazard disaggregation

Probabilistic seismic hazard analysis and disaggregation were performed using the OpenQuake-engine (Pagani, et al., 2014), in accordance with the theoretical background established by McGuire (2004) and Bazzurro & Cornell (1999), respectively. As a result, hazard disaggregation has been performed for the following rupture discretization approaches:

1. Rupture-by-rupture, i.e. the contribution of all the independent ruptures generated by the earthquake rupture forecast (ERF) is computed, as illustrated in Figure 3.2;
2. Scenario ruptures, i.e. the ruptures generated by the ERF are classified and grouped according to magnitude (M) / distance bins, assuming Joyner-Boore (R_{jb}) as the distance measure:
    a. Magnitude interval (ΔM_w) = 0.2 / Distance interval (ΔR_{jb}) = 2 km;
    b. ΔM_w = 0.4 / ΔR_{jb} = 4 km;
    c. ΔM_w = 0.6 / ΔR_{jb} = 8 km;
    d. ΔM_w = 0.8 / ΔR_{jb} = 12 km;
    e. ΔM_w = 1.0 / ΔR_{jb} = 20 km.

With respect to methods 2a to 2e, one should note that evaluating the relative influence of increasing the interval of magnitude or distance individually is not the objective of this study. Differently, one is interested in determining the impact of using disaggregation methods with increasing approximation levels. As a result, the considered increase in magnitude intervals is proportional to that of distance.

In addition, the parameters necessary for the application of the GMPEs cannot directly be obtained when methods 2a to 2e are used, because various ruptures are

identified in each M / R_{jb} bin. Therefore, for each bin, the considered M and R_{jb} pair corresponds to the average values in the bin interval, whereas the remaining parameters (see Table 3.2) have been inferred from the rupture with higher disaggregation contribution to hazard. The OpenQuake-engine does not provide the contributions on a rupture-by-rupture basis. Instead, these are classified and grouped, as presented in 2). However, given the open-source nature of this tool, it was possible to produce the necessary intermediate results for the computation of $P_{rup_{i}|_{IM_j=imj}}$ - the probability that the rupture properties are those of a particular rup_i, given that a ground motion (IM_j) has occurred, with an intensity of imj, as defined in Equations 3.1 and 3.2:

$$P_{rup_{i}|_{IM_j \geq imj}} = \frac{ln\left[1 - P_{IM_j \geq imj|_{rup_i}}\right]}{ln\left[\prod_{i=1}^{I}(P_{IM_j \geq imj|_{rup_i}})\right]} \quad (3.1)$$

$$P_{rup_{i}|_{IM_j \geq imj}} \cong \frac{1}{v_{IM_j \geq imj} - v_{IM_j \geq imj + \Delta imj}} *$$

$$*\left[P_{rup_{i}|_{IM_j \geq imj}} . v_{IM_j \geq imj} - P_{rup_{i}|_{IM_j \geq imj + \Delta imj}} . v_{IM_j \geq imj + \Delta imj}\right] \quad (3.2)$$

Where I is the total number of ruptures in the ERF, $v_{IM \geq im}$ is the annual rate of exceedance of a ground motion with an intensity level of imj, and Δimj is a small increment of IM_j, relative to imj. As highlighted by Bradley (2010a), Equation 3.2 becomes exact in the limit as $\Delta imj \to 0$.

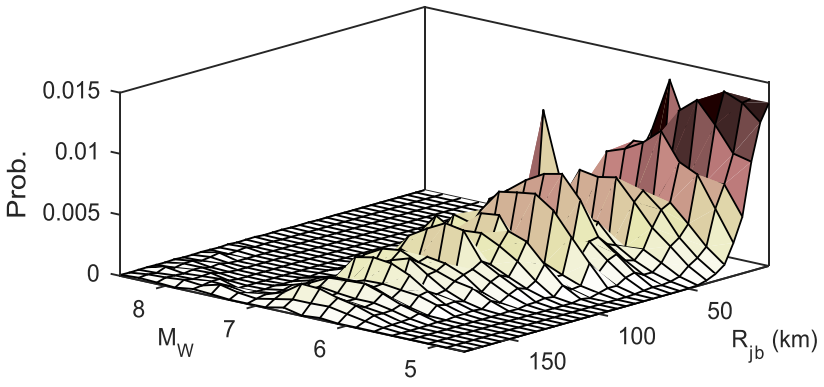


Figure 3.2 – Rupture-by-rupture disaggregation (method 1), using the ERF generated for the FSBG-model, and $Sa(T_1)=0.5g$. Ruptures are grouped with $\Delta M_w = 0.2$ ΔR_{jb} =5 km, for visual clarity.

3.3.3 Hazard consistent record selection

The Generalized Conditional Intensity Measure (GCIM) approach is adopted herein for the purpose of record selection, as it allows the *predictability* (Kramer & Mitchell, 2006) of all the intensity measures verified to influence the seismic response of the assessed structures. Readers are referred to the work of Bradley (2010a) for a detailed description of the theoretical background of the methodology. In brief, the fundamental basis of the GCIM is that any set of ground motion parameters can be assumed to follow a multivariate lognormal distribution, and the conditional distribution given a) a rupture scenario, and b) the occurrence of a specific value of IM_j, has a univariate lognormal distribution.

Upon definition of the ground motion prediction equation of interest and the correlation between the considered intensity measures, the conditional distribution of a certain intensity measure (*IMi*) given a level *imj* of a second intensity parameter *(IMj)* is given by:

$$f_{IMi_{|IMj=imj}} = \sum_{i=1}^{I} f_{IMi_{|rup_i,IMj=imj}} \cdot P_{rup_{i|IMj=imj}} \quad (3.3)$$

Where $f_{IMi_{|IMj=imj}}$ is the probability density function (pdf) of *IMi* given *IMj=imj*, $f_{IMi_{|rup_i,IMj=imj}}$ is the pdf of *IMi* given *IMj=imj* and *Rup=rup*$_i$, and $P_{rup_{i|IMj=imj}}$ is obtained from disaggregation (methods 1 and 2a to 2e).

It shall be noted that, building on the mathematical formulation presented in Appendix E of the study by Baker and Cornell (2005), any of the methods 2a to 2e could potentially be used with little discrepancies in both mean and variance of $f_{IMi_{|IMj=imj}}$ (with respect to the exact method 1), provided that certain corrections are performed. More specifically, Equation 3.3, which accounts for the contribution of all the ruptures in the ERF, is equivalent to Equation 3.4, which simplifies Equation 3.3 by determining $f_{IMi_{|IMj=imj}}$ as the weighted contribution of each disaggregation bin:

$$f_{IMi_{|IMj=imj}} = \sum_{k=1}^{K} f_{IMi_{|bin_k,IMj=imj}} \cdot P_{bin_{k|IMj=imj}} \quad (3.4)$$

Where *K* is the total number of disaggregation bins, and $f_{IMi_{|bin_k,IMj=imj}}$ and $P_{bin_{k|IMj=imj}}$ are determined as follows:

$$f_{IMi_{|bin_k,IMj=imj}} = \sum_{j=1}^{J_k} f_{IMi_{|rup_j,IMj=imj}} . P_{rup_{j_{|IMj=imj}}} \quad (3.5)$$

$$P_{bin_{k_{|IMj=imj}}} = \sum_{j=1}^{J_k} P_{rup_{j_{|IMj=imj}}} \quad (3.6)$$

and J_k is the number of individual ruptures identified and grouped in bin *k*.

Knowing that both $f_{IMi_{|bin_k,IMj=imj}}$ and $f_{IMi_{|rup_j,IMj=imj}}$ are lognormal probabilistic functions (Ang & Tang, 2007), a satisfactory approximation to the conditional mean and variance of $f_{IMi_{|bin_k,IMj=imj}}$ can be obtained by the First-Order, Second Moment (Melchers, 1999) method, as follows:

$$\mu_{f_{IMi_{|bin_k,IMj=imj}}} = \mu_{f_{IMi|\overline{M,R_{jb}}}} + \sigma_{f_{IMi|\overline{M,R_{jb}}}} . \rho_{lnIMi,lnIMj} . \bar{\varepsilon} \quad (3.7)$$

$$\sigma_{f_{IMi_{|bin_k,IMj=imj}}} = \sigma_{f_{IMi|\overline{M,R_{jb}}}} \sqrt{1 - \rho^2_{lnIMi,lnIMj}} \quad (3.8)$$

Where $\overline{M}$,$\overline{R_{jb}}$,$\bar{\varepsilon}$ are the weighted mean of M, R_{jb}, ε (weights are the disaggregation probabilities of the J_k ruptures identified in the bin of interest – $P_{rup_{j_{|IMj=imj}}}$), and $\mu_{f_{IMi|\overline{M,R_{jb}}}}$, $\sigma_{f_{IMi|\overline{M,R_{jb}}}}$ are the mean and variance of f_{IMi}, respectively, evaluated at $\overline{M}$, $\overline{R_{jb}}$.

Despite the possible advantages of this correction to methods 2a to 2e, a "rupture-by-rupture" disaggregation (method 1) would still have to be performed in order to identify each rupture probability ($P_{rup_{j_{|IMj=imj}}}$), therefore defeating the purpose of an approximate solution. For this reason, this approach was not addressed in this work. Differently, the present objective is to determine whether and at which degree of discretization an approximate approach such as 2a to 2e is able to provide accurate results in terms of fragility, vulnerability and loss, when compared with an exact solution (method 1).

3.3.4 'Targets' for record selection

The vector of intensity measures considered in this research (i.e. *IM*) includes intensity parameters (i.e. *IMi*) of peak ground acceleration (PGA), Housner intensity (HI) and spectral ordinates within the range of periods of 0.05 to 3.0 seconds, conditioned on the spectral acceleration (*IMj*) at the mean fundamental period of vibration of each class – $Sa(T_1)$, as further (thoroughly) justified in Chapter 4 (section 4.3.3). In the referred

section, the issue of *predictability* has been foreseen in the definition of *IM*, and matters of *efficiency* (Shome & Cornell, 1999), *sufficiency* (Luco, 2002) and *scaling robustness* (Tothong & Luco, 2007) have been verified when analysing similar structural models. As recommended in Chapter 4 (section 4.4.2.1), a number of 60 ground motion records have been selected and scaled per level of $Sa(T_1)$, with the latter ranging from 0.1g to 1.0g, with intervals of 0.1g.

The probabilistic distribution of the selected *IM* vector conditioned on a given level of $Sa(T_1)$ is designated henceforth as $f_{IM_{|Sa(T_1)=a}}$, being determined according to the hazard-consistent probabilistic distribution of each *IMi* given *Sa(T1)=a,* as established in Equation 3.3, and the correlation models enunciated in section 4.3.3 (Table 4.1). For details regarding the database of natural ground motion records, readers are also referred to Chapter 4, where an explanation of all the issues addressed above is presented in detail.

For the sake of illustration, the 'target' probabilistic distributions of *HI* and spectral ordinates of periods ranging between 0.1 and 3.0 seconds are presented in Figure 3.3. In this figure, 5-story (mid-code) structures are considered, and probabilistic distributions computed according to methods 1 and 2a to 2e are illustrated, in order to highlight the differences between hazard-consistent record-selection 'targets' when different rupture discretization options are considered in hazard disaggregation.

As illustrated in Figure 3.3, in which the FSBG-model and CY08 GMPE have been used, significant differences are verified between the mean of exact and approximate solutions, with discrepancies increasing proportionally to the increase of the discretization interval. In other words, the accuracy of predictions decreases from method 2a to 2e, with respect to method 1. In terms of variance, this discrepancy is reflected in the increase of uncertainty from method 2a to 2e, as verified in Figure 3.3b (i.e. loosely speaking, it is possible to visually confirm that the standard deviation of curve 2a is lower than that of 2e because curve 2a is 'more vertical' than 2a). For the sake of synthesis, only a particular set of results are presented in Figure 3.3. However, similar differences in accuracy have been verified for each source model and GMPE investigated, for all the structures and levels of $Sa(T_1)$.

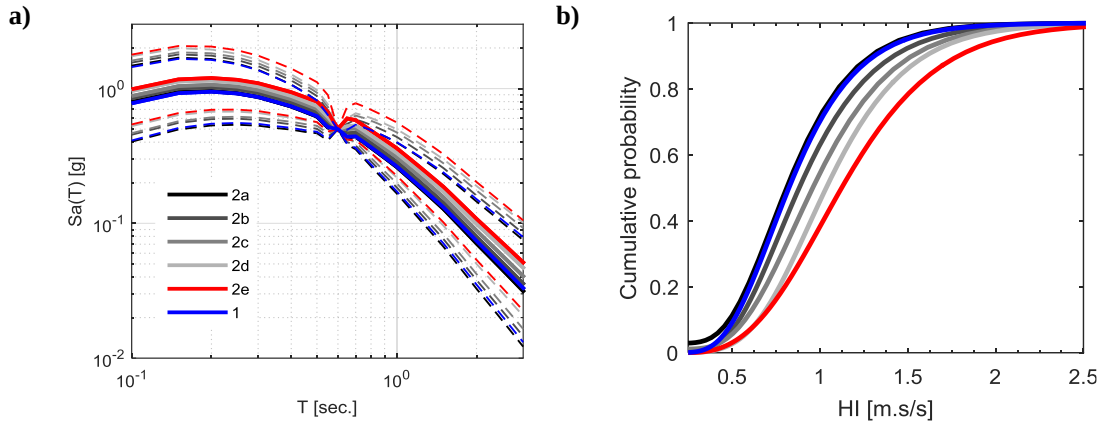


Figure 3.3 – **a)** Target spectral ordinates of periods ranging between 0.1 and 3.0 seconds (solid lines correspond to the mean and dashed lines represent 16 and 84 percentiles, i.e. mean +/- 1 standard deviation), and **b)** target cumulative probabilistic distribution functions of *HI*. In both cases, the results from disaggregation methods 1 and 2a to 2e (FSBG-model, CY08) are illustrated, considering a mid-code structure of 5 floors, for a conditional $Sa(T_1)=0.5g$.

A possible way to assess how differences in 'targets' for ground motion selection propagate to discrepancies in fragility and loss results is to evaluate what is the degree of 'similarity' between distributions of *IMs* in the resulting record sets. 'Similarity' is obviously a qualitative concept; therefore, in order to quantitatively compare conditional 'target' distributions of IMs obtained from method 1 and 2a to 2e, a statistical approach was implemented. In this methodology, record sets selected based on approximate 'target' distributions (from 2a to 2e) were individually compared with those obtained using method 1. For a given level of $Sa(T_1)$ and building class (and a certain combination of source model/GMPE), one is able to evaluate what is the empirical distribution of each *IMi* (see section 3.3.3), for each set of records selected using methods 1 and 2a to 2e. Therefore, the null hypothesis that a given approximate empirical distribution follows the same underlying normal distribution as the corresponding exact one (method 1) can be assessed using the two-sample Kolmogorov-Smirnov (KS) test (Ang & Tang, 2007).

Loosely speaking, the so-called *p-values* computed in this KS test represent the probabilities that the largest discrepancy between two samples could indeed be observed if the two respective cumulative distributions were sampled from the same parent one. In practice, if the *p*-value is large, we cannot reject the null hypothesis, concluding that the evaluated approximate distribution is a good approximation to the exact one (method 1). Conversely, if the *p*-values are smaller than a reference value α, say, 10%, it can be assumed that the two distributions are statistically significantly different. For illustration, the *p-values* resulting from comparing the empirical distributions of $Sa(T=0.05-3.0\ sec.)$ conditioned on $Sa(0.624\ sec.)=0.5g$ (MC-5) are presented in Figure 3.4, using the FSBG

source model. For each *T*, the empirical distribution of *Sa(T)* obtained when records are selected for 'targets' computed using methods 2a to 2e are individually compared with those resulting from ground motions selected with the exact 'targets' (method 1).

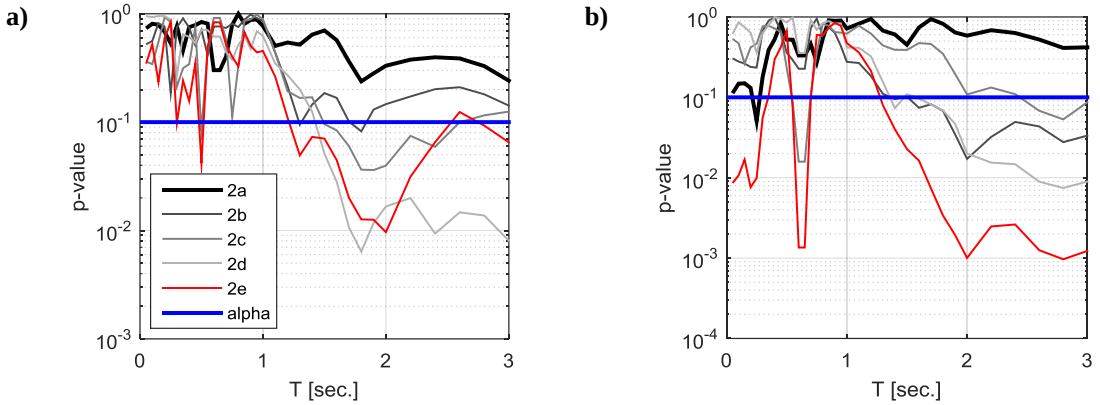


Figure 3.4 – **a)** p-values obtained with the KS test, when comparing empirical distributions of *Sa(T=0.05 to 3.0 sec.)* derived from records selected with approximate target distributions (methods 2a to 2e), as opposed to those obtained with the exact method 1. Conditioning $Sa(T_1)=0.5g$ (MC-5), FSBG source model and AB10 GMPE. **b)** same as **a)** but with CY08 GMPE.

As illustrated, the corresponding *p-values* are presented for each approximate method (from 2a to 2e), for each *T*. Based on these results, one is able to recognize that the *p-values* generally decrease from method 2a to 2e. In fact, these are lower than α for methods 2b, 2c, 2d and 2e, at least at one *T* instance, which implies that only in the case of method 2a the approximation is statistically significant.

Although an extensive set of intensity measures has been considered for record selection (with that set including the IMs for which higher correlation with response quantities has been observed in Chapter 4 in the analysis of similar structural models), the non-linear response of structures is a complex phenomenon that is influenced by ground motion properties that may not comprehensively be represented by the set of IMs considered herein. Therefore, the results presented in Figure 3.4, despite providing a valuable insight, care for appropriate validation through NLRHA and corresponding fragility and loss estimates, subsequently presented.

As in the case of Figure 3.3, only a particular set of results are presented in Figure 3.4, for the sake of synthesis. Nonetheless, similar results have been attained for each combination of source model/GMPE investigated (as well as all structures and levels of $Sa(T_1)$). More specifically, the assessed null hypothesis is systematically rejected for methods 2b, 2c, 2d and 2e, whereas method 2a tend to lead to 'target' distributions that

can be considered statistically identical to those referring to the exact method 1, at the 10% significance level.

3.4 Fragility and loss assessment

As discussed by Silva et. al. (2015c), the use of local criteria (e.g. local plastic hinge rotation) to define limit states when generating fragility curves for a population of buildings may not be appropriate. Hence, in the study herein, structural response will be evaluated based on the maximum inter-story drift (*ISD*) and global drift (*GD*), considering four damage states: Slight Damage (*SD*), Moderate Damage (*MD*), Extensive Damage (*ED*) and Collapse (*Col*). In this context, *GD* corresponds to the maximum roof drift ratio, computed as the fraction between maximum roof displacement and building height.

Global Drift limits are determined according to the evaluation of capacity of each frame through a displacement-based adaptive pushover (Antoniou & Pinho, 2004). Similarly to what has been considered by other authors (e.g. Erberik (2007)), displacement thresholds at each limit state are defined for each sampled frame without masonry infills (bare frame) according to the following assumptions:

- Slight damage: global drift at 50% of maximum base shear capacity;
- Moderate damage: global drift when 75% of maximum base shear capacity is achieved;
- Extensive damage: global drift at maximum base shear capacity;
- Collapse: global drift when 20% decrease of the base shear capacity is verified, or 75% of the ultimate global drift attained, whichever is achieved first. In this framework, ultimate capacity is considered to be achieved once lack of numerical convergence is verified, provided that maximum shear strength has been achieved prior to numerical instability. With respect to numerical simulation assumptions, readers are referred to the work of Silva et. al. (2015c).

The influence of infill panels, which translates to a significant decrease of displacement capacity, is accounted for by applying the reduction factors proposed by Bal et. al. (2010) for each aforementioned limit state.

For what concerns *ISD*, a fixed set of values per limit state are defined based on the evaluation of global damage with increasing inter-story drift from 25 dynamic tests performed in real reinforced concrete moment resisting frames by Rossetto & Elnashai

(2003). In order to adapt the six damage states proposed by the latter Authors with the one being considered in this study, *light/slight damage* and *partial collapse/collapse* damage states have been merged, as follows:

- Slight damage: 0.08% maximum inter-story drift;
- Moderate damage: 0.30% maximum inter-story drift;
- Extensive damage: 1.15% maximum inter-story drift;
- Collapse: 2.80% or higher maximum inter-story drift.

3.4.1 Fragility comparison

For the purpose of fragility assessment, lognormal cumulative distribution functions have been fitted to the scatter of exceedance probabilities obtained for all the combinations of source model / GMPE. Based on the visual inspection of Figure 3.5 and Figure 3.6, is possible to conclude that, within each combination of source model / GMPE, the differences between approximate methods and method 1 tend to increase proportionally to the decrease of accuracy of the approximation (i.e. method 2a tends to lead to fragility curves that are 'closer' to the those obtained with method 1, whereas method 2e is the one for which the differences with respect to method 1 are higher). However, these differences are not consistent across combinations of source model / GMPE, and are in fact exacerbated in the cases of AS and SEIFA source models.

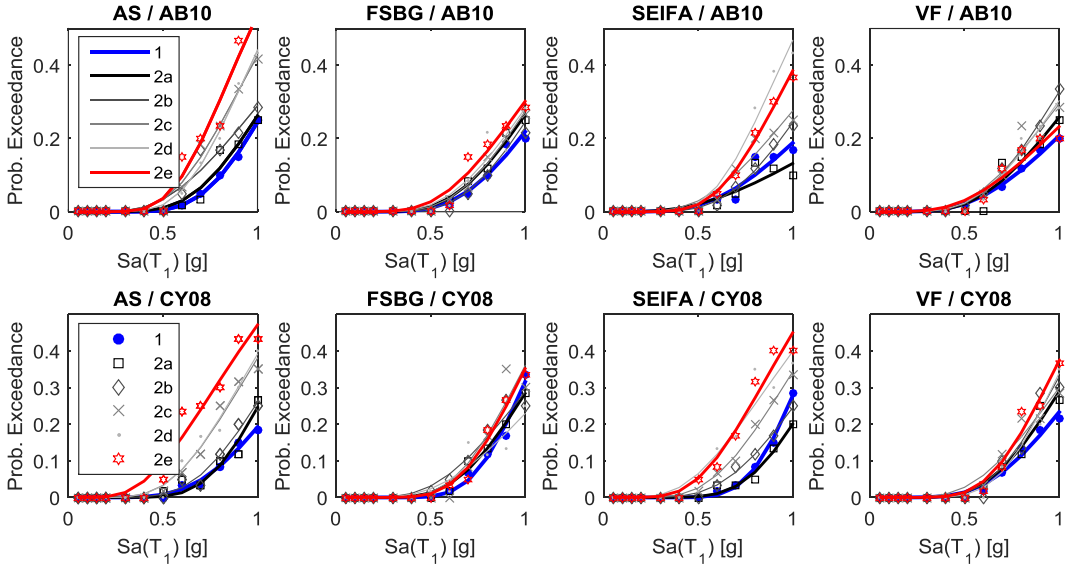


Figure 3.5 – Fragility functions obtained with records selected based on methods 1 and 2a to 2e, for all the combinations of GMPE / source model. Limit state of Collapse using the ISD criteria, C-5 building class.

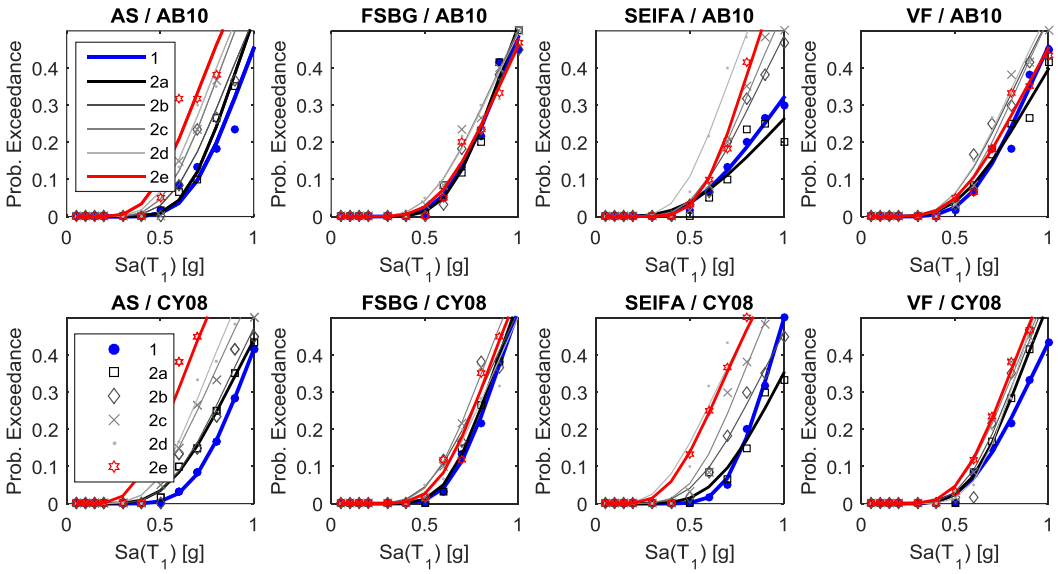


Figure 3.6 – Fragility functions obtained with records selected based on methods 1 and 2a to 2e, for all the combinations of GMPE / source model. Limit state of Collapse (GD criteria), C-5 structural typology.

The degrees of discrepancy between approximate methods and method 1 verified in the case of AS and SEIFA models (versus those of FSBG and VF models) can be explained by the rupture-by-rupture disaggregation results obtained for each source model, as exemplified in Figure 3.7 for the case of methods 1 and 2c. In this figure, the C-5 building and $Sa(T_1)=0.5g$ are selected for the sake of illustration. However, it has been verified that the trends exhibited here are common for all the structural classes and GMPEs used.

More specifically, it has been verified that exact disaggregation probabilities (method 1) obtained with source models AS and SEIFA are generally significantly more clustered in a limited range of small magnitudes and distances, whereas in the case of FSBG and VF, disaggregation contributions tend to be more homogeneous across the entire range of M and R_{jb}. For this reason, aggregating hazard contribution probabilities into more coarse M/R_{jb} intervals (as illustrated in Figure 3.7 for method 2c), does not result in significant changes when FSBG and VF models are considered (i.e. the distribution of probabilities remains fairly homogeneous across the overall range of M and R_{jb}). In the case of AS and SEIFA models, on the other hand, one can verify that the "spike" registered for low magnitudes and distances is much more pronounced in the case of method 2c than in method 1. For this reason, as the discretization intervals increase (from methods 2a to 2e), the differences in disaggregation results increase more significantly when considering AS and SEIFA models. As a result, discrepancies between

approximate and exact fragilities are also more significant in the case of AS and SEIFA models.

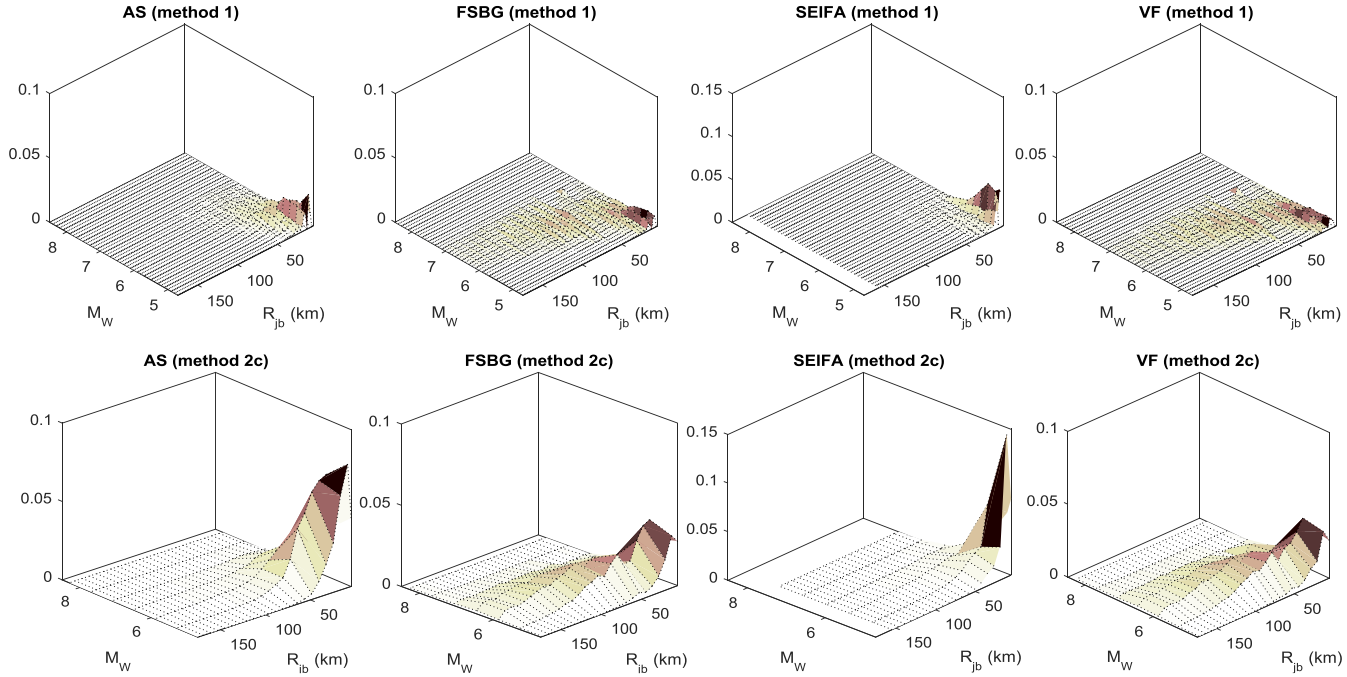


Figure 3.7 – Disaggregation results for methods 1 and 2c, considering AB10 GMPE and the ERF generated for the AS, FSBG, SEIFA and VF models, for *Sa(T_1)=0.5g (C-5)*. Ruptures in the exact method 1 are grouped with ΔM_w = 0.1 / ΔR_{jb} =1 km, for visual clarity, and, in practice, contribution probabilities of method 2c result from aggregation of those of method 1 into coarser M, R_{jb} intervals.

In addition, one might argue that non-negligible differences are also obtained between exact fragilities (method 1) across different combinations of source model / GMPE. This can can be explained by the fact that the used intensity measure is not *sufficient*. In other words, because *Sa(T_1)* cannot comprehensively reflect all the ground motion properties influencing the seismic response of the assessed structures, fragility results are dependent on the ground motion properties of the selected set of records. Therefore, because different combinations of source model / GMPE provide different 'targets' for selection (even for method 1), the exact fragilities are in fact distinct across combinations of source model / GMPE.

Given the wide range of seismic hazard modelling options evaluated, it would not be practical to present the comparison presented in Figure 3.5 and Figure 3.6 for all the investigated building classes (and damage states). For this reason, only the results corresponding to the Collapse limit state of post-code 5-story building (C-5) are illustrated. Nonetheless, it shall be noted that these are in fact representative of the results obtained for all the remaining structural typologies.

In this section, the comparison between fragilities is done visually, since a statistical comparison between different parametric curves would inevitably imply that all the assessed functions are different at any significance level. As a practical example, one might think of the KS test for comparison between two data samples. If one were to randomly simulate an extremely large number of values of $Sa(T_1)$ and obtain the corresponding damage exceedance probabilities from two particular fragility functions, one would be able to test the null hypothesis of the samples arising from the same distribution by comparing the corresponding cumulative distributions (CDF) of exceedance probabilities. However, for an infinitely large number of sampled values, it follows from the theoretical outline of the KS method (Ang & Tang, 2007) that the maximum allowable difference between CDFs in order not to reject the null hypothesis would tend to zero. In this situation, all the differences between CDFs would be higher than the limit (zero), inevitably rejecting the null hypothesis of the samples being drawn from the same distribution, for any two fragility curves.

The example above is outlined in order to demonstrate that such comparison would not provide any meaningful insight on the impact of the differences visually identified in Figure 3.5 and Figure 3.6. For this reason, the evaluation of the influence of those differences is performed herein by investigating what is the impact on the corresponding loss estimates.

3.4.2 Vulnerability and loss estimation

In this section, vulnerability functions are drawn from the fragility curves presented above, using the consequence model adopted by Silva *et. al.* (2015c) in the evaluation of vulnerability of similar structural typologies as those studied herein. As a result, deterministic damage ratios (ratio between cost to repair and total building value) of 0.1, 0.3, 0.6 and 1.0 are adopted for damage states *SD, MD, ED* and *Col.*, respectively. Although not presented for the sake of synthesis, fragility functions for limit states other than Collapse present similar scatter as those presented in Figure 3.5 and Figure 3.6. This is confirmed in Figure 3.8, where a sample of vulnerability results is shown. As illustrated, similar relative differences between curves obtained with methods 1 and 2a to 2e are observed for the selected cases, which are representative of all the results.

a) **b)**

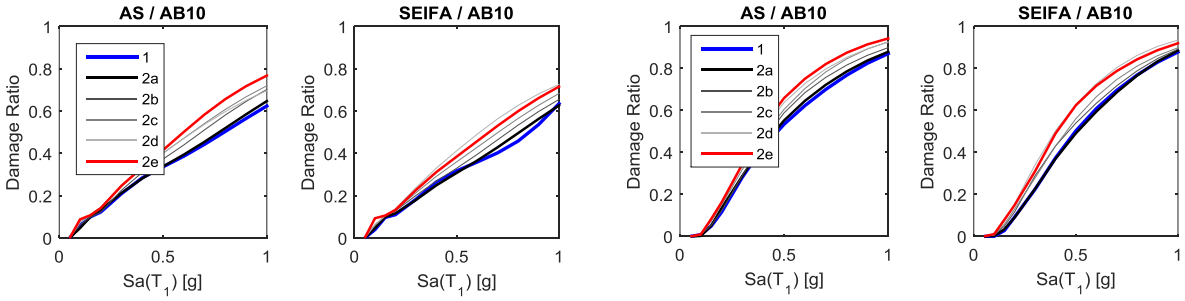


Figure 3.8 – **a)** Vulnerability functions obtained with records selected based on methods 1 and 2a to 2e, for AS/AB10 (left) and SEIFA/AB10 (right), combinations. ISD criteria and C-5 structural typology, **b)** same as **a)**, considering GD criteria.

In order to account for the propagation of uncertainty from fragility to vulnerability, the variability of damage ratios, for a specific level of $Sa(T_1)$, is directly associated with the uncertainty in fragility regression. More specifically, 200 fragility functions are fitted to equal number of synthetic datasets randomly generated by bootstrap sampling with replacement (Wasserman, 2004) from the original sets of intensity-specific probabilities.

Because the bootstrapping is consistent across limit states (i.e. for a given simulation, the indexes of the probabilities sampled for *SD* are the same ones used to build the corresponding samples of *MD*, *ED* and *Col*. probabilities), it is possible to compute a distinct vulnerability function for each bootstrapped fragility (Figure 3.9). As a result, loss exceedance curves are also computed for each of the bootstrapped vulnerability functions, for each method (1 and 2a to 2e), structural model and combination of source model/GMPE.

In this framework, the uncertainty in vulnerability is related with the uncertainty in fragility regression, which is further propagated into the estimation of loss exceedance probabilities. In other words, for a given building class and combination of source model/GMPE, one obtains 200 distinct loss exceedance curves, in accordance with the 200 bootstrapped vulnerability functions.

For the purpose of comparing loss estimates obtained from approximate methods (2a to 2e) with those of method 1, average annual loss (AAL) values are computed for each of the bootstrapped loss exceedance curves. As a result, 200 independent AAL values are obtained for each disaggregation method, making it possible to compute 200 normalized differences between a) AAL obtained with an approximate method (2a to 2e), and b) AAL computed for method 1 (for each structural class and combination of source model/GMPE).

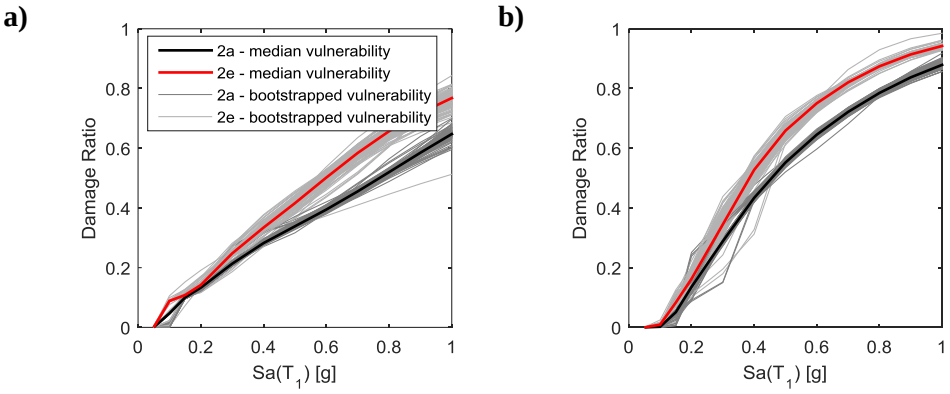


Figure 3.9 – **a)** Uncertainty in the vulnerability model obtained from bootstrapping with replacement from the original sets of damage exceedance probabilities. Structural typology of C-5, methods 2a and 2e, AS/AB10 combination, and ISD criteria, **b)** same as **a)**, considering GD criteria.

Using the sets of 200 AAL values, one could use statistical tests to compare distributions of (200) AAL values obtained with approximate and exact methods. However, because these results are obtained via bootstrap sampling, using statistical tests would not be a sound approach. More specifically, because the rejection of the null hypothesis (of the samples being drawn from the same distribution) is dependent on the sample size, it is possible to manipulate the number of bootstrap samples in order to reject or not the null hypothesis. Therefore, instead of comparing distributions of AAL from a statistical point of view, one is interested in computing the probability that the difference between approximate and exact AAL values is higher than a limit of interest. This is furthermore understood as a more informative exercise, since simply comparing the distributions would not render a qualitative measure of the resulting error.

In the present case, the Author has decided that, based on subjective judgement, an error of 10% is an adequate compromise between computational efficiency and accuracy of results, i.e. one considers an approximate method to be satisfactory if the resulting error is lower than 10%. Therefore, the suitability of each approximate method is herein evaluated as the probability of the attained error being higher than 10%, as shown in Figure 3.10 and Figure 3.11.

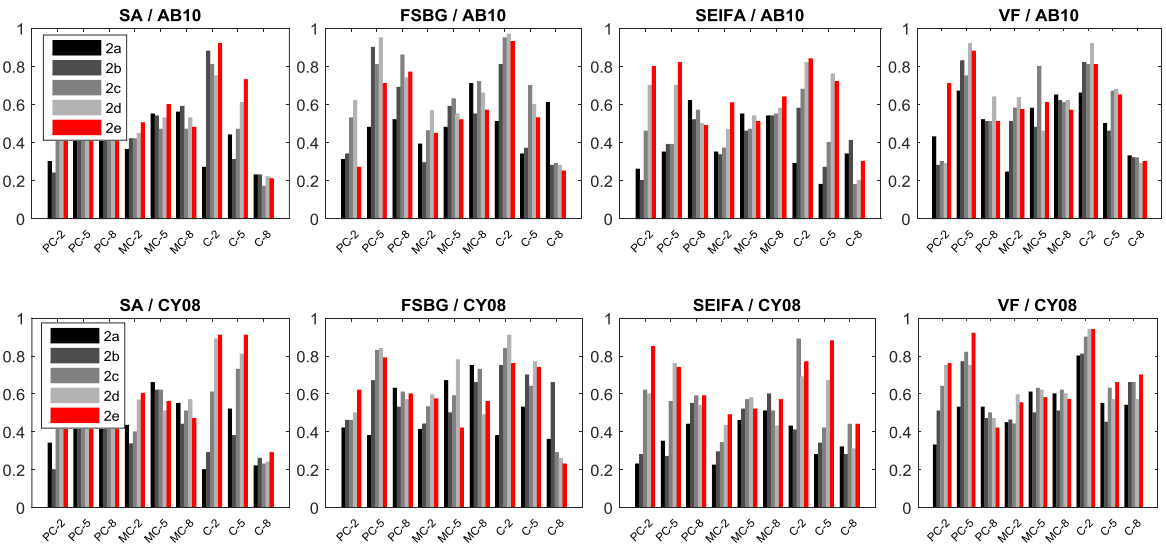


Figure 3.10 – Probability of 'Error' being higher than 10%, where 'Error' is the absolute (normalized) difference between: a) AAL obtained from methods 2a to 2e, and b) AAL computed using method 1. Inter-story Drift Criteria.

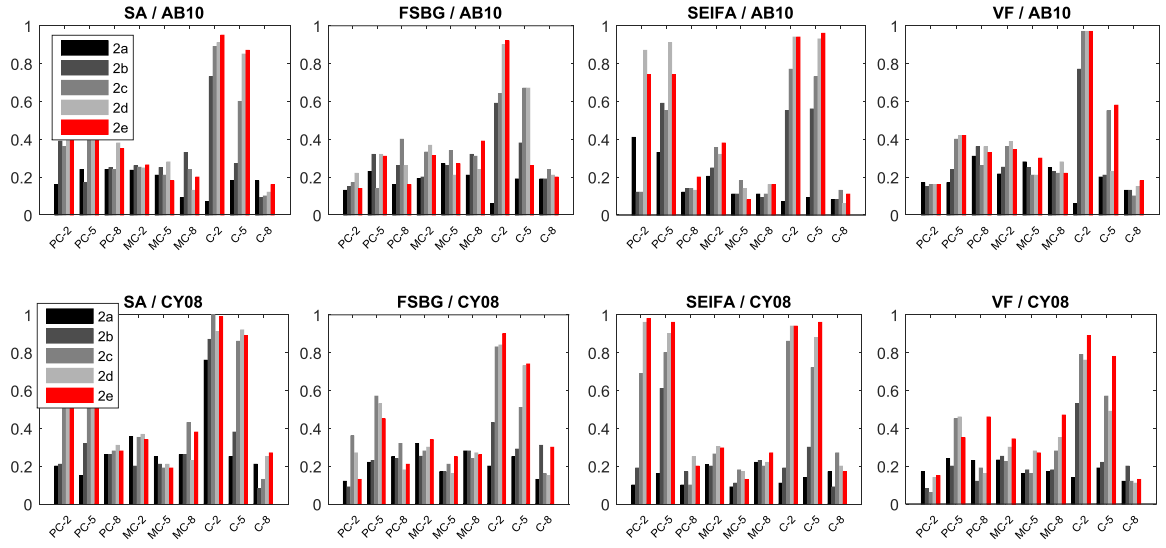


Figure 3.11 – Probability of 'Error' being higher than 10%, where 'Error' is the absolute (normalized) difference between a) AAL obtained from methods 2a to 2e, and b) AAL computed using method 1. Global Drift Criteria.

As illustrated, the probabilities of obtaining errors higher than 10% (herein simply referred as 'error probabilities') generally increase proportionally with the decrease in accuracy of the approximation (i.e., from method 2a to 2e), irrespectively of the combination of source model/GMPE and structural class (more specifically, in approximately 80% of the assessed building class / source model / GMPE combinations). Moreover, it was verified that such probabilities are significantly higher when ISD criteria are used to derive the fragility models (Figure 3.10), in comparison with the cases where GD criteria are considered (Figure 3.11). In the case of GD, limit state thresholds are

dependent on the properties of the sampled building, whereas in the case of ISD the specified limits are common to all the sampled structures in a class. As a result, despite the fact that response parameters are obtained for the same sampled building in each class, there is a higher uncertainty introduced when limits are not building-specific, leading to increased probability of obtaining an error higher than 10%, in the case of ISD criteria.

Considering method 2a as the one for which an approximate approach generally imparts lower errors, Figure 3.12 presents the corresponding 'error probabilities', for all the combinations of source model / GMPE, and structural classes. As in Figure 3.10 and Figure 3.11, the differences between results of ISD and GD criteria are evident. However, it is also possible to conclude that, despite the differences registered in fragility functions obtained for AS and SEIFA models (as opposed to FSBG and VF models – see section 3.4.1), method 2a leads to error probabilities that do not exhibit a particular trend across source models (irrespectively of the GMPE and structural model).

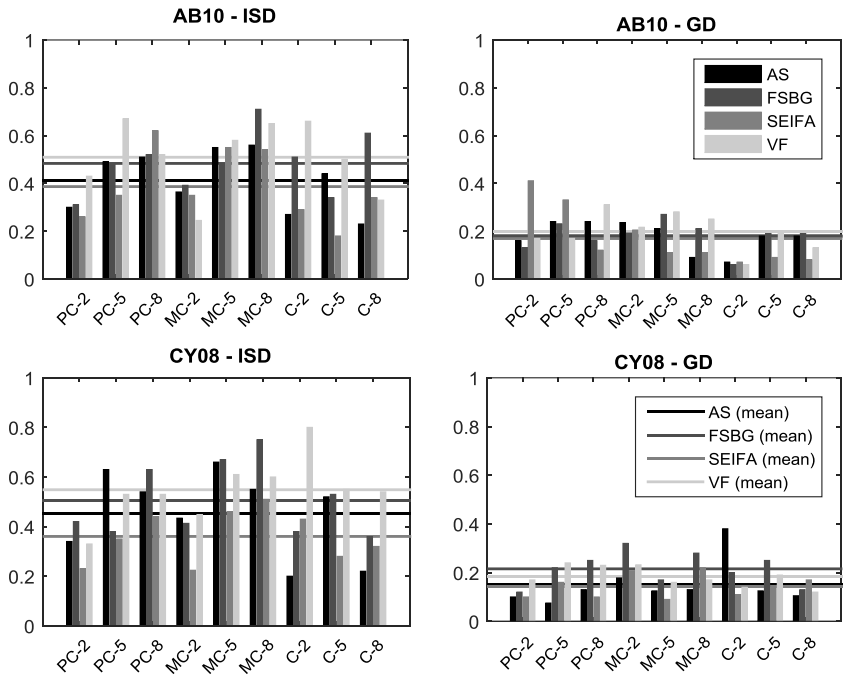


Figure 3.12 – Probability of 'Error' being higher than 10% obtained with method 2a, for all the combinations of source model / GMPE, structural classes, and limit state criteria. For clarity, horizontal lines correspond to the mean errors of all the structural classes, for each of the different source models.

This can be explained by the fact that, despite errors in fragility for a less accurate method (say, method 2e) are significantly higher in the case of AS and SEIFA models, method 2a provides the same level of approximation irrespectively of the source model. More specifically, when GD criteria are used, an average 'error probability' of approximately 20% is obtained, whereas average values of 50% are obtained for ISD.

These results indicate that, for the range of structural classes and source model / GMPE combinations assessed, all the approximate disaggregation methods are inadequate when limit state criteria are not building-specific. On the other hand, when limit state thresholds are building-specific, method 2a seems to constitute a valid approximation, given the low value (20%) of probability of attaining errors higher than 10%.

3.4.3 Collapse risk assessment

In this section, risk estimates are further evaluated in the context of the collapse risk assessment of the buildings under scrutiny. For this purpose, the average annual probability of collapse has been determined and compared with limits prescribed by different authors and seismic design regulations. More specifically, the present objective is to evaluate the differences in 'safety tagging' one would obtain by considering an approximate method, with respect to the exact probability of collapse. In this context, 'safety tagging' is understood as the process of classifying a building's seismic safety, when comparing the appraised annual probability of collapse with the selected acceptable limits.

As presented in the study by Silva et. al. (2015a), in which the collapse probability of several building typologies has been assessed in order to determine the so-called 'risk-targeted' hazard maps for Europe, ASCE (2010) establishes an acceptable risk of 1% in 50 years (approximately 2.0×10^{-4} annually) for the territory of the United States. On the other hand, Douglas *et al*. (2013) has adopted a value of 1.0×10^{-5} annually as a reasonable limit, following the literature review of several studies in which the annual probability of collapse of a number of structures designed according to modern regulations in France was determined. Given the significant difference between the two proposals, both values (2.0×10^{-4} and 1.0×10^{-5}) are considered herein for completeness, as boundary limits. For illustration, Figure 3.13 presents the average annual probability of collapse for the mid-code and post-code buildings (for methods 1 to 2e), as well as the selected limits, for the AS / AB10 combination.

In this figure, it is possible to verify that, for both limits of acceptable risk, method 2a leads to similar "safety tagging" as that obtained with the exact method 1, for both ISD and GD criteria. For the sake of synthesis, the results of pre-code structures are not illustrated, since the corresponding collapse probability values are systematically higher

than 2.0×10^{-4}, for all the methods (1 to 2e). In addition, only the AS / AB10 combination is illustrated; however, similar results (in terms of difference in "safety tagging" between exact and approximate methods) were verified for the remaining source models /GMPE.

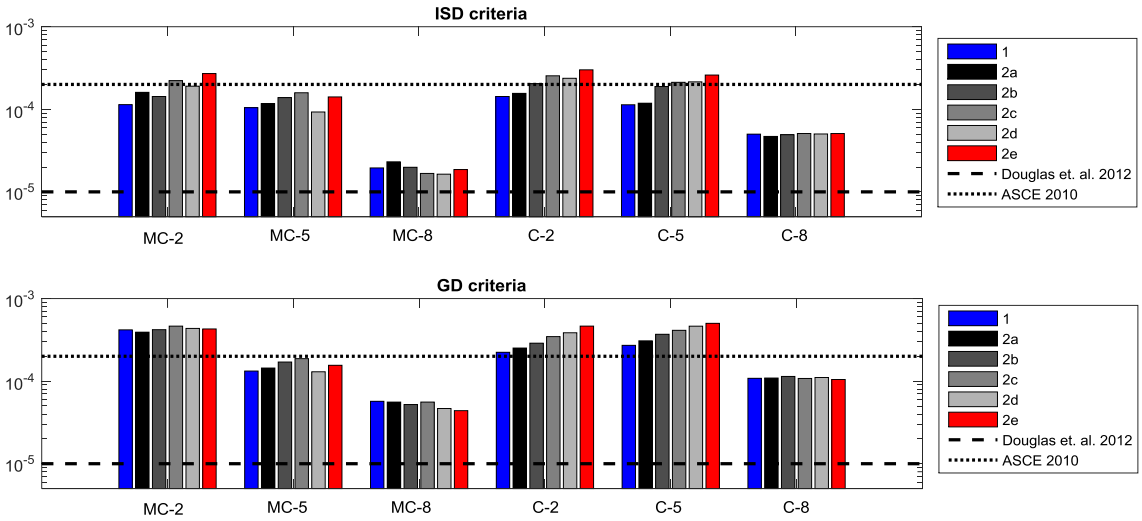


Figure 3.13 – Average annual collapse probability (mid-code and post-code buildings) computed using methods 1 and 2a to 2e, for the AS/AB10 combination.

3.5 Final remarks

In this chapter, hazard-consistent ground motion selection and fragility assessment were performed for 9 building classes representing existing structures in Portugal. Target distributions for record selection were computed based on rupture contributions determined by seismic hazard disaggregation for Lisbon, Portugal, and several degrees of accuracy are used, ranging from: 1) the consideration of all the possible ruptures contributing to hazard, to 2) a more coarse approach in which rupture scenarios are grouped into magnitude (M) / distance (R_{jb}) bins.

In order to account for the epistemic uncertainty associated with the definition of seismicity in the area of interest, PSHA was performed based on four distinct source models and two different GMPEs, with the objective of: a) verifying the impact of different attenuation relationships and corresponding set of rupture defining parameters in the computation of hazard-consistent fragility, for a given seismicity model, and b) assess the influence of different seismicity modelling approaches on the hazard-consistent fragility, when considering GMPEs with different degrees of detail in the definition of rupture properties.

Based on the corresponding risk results evaluated in terms of AAL, it was verified that only an exact disaggregation method (i.e. contribution of all the possible independent

ruptures) guarantees a satisfactory outcome in terms of accuracy, when limit state criteria are not building-specific, irrespectively of the source model/GMPE combination used. On the other hand, when limit state thresholds are building-specific, an approximate method in which magnitude/distance bins are defined with ΔM / ΔR_{jb} equal to 0.2 / 2km (designated here as method 2a) systematically leads to a valid approximation (for the structures, source models, GMPEs, and site investigated herein). However, when the risk metric investigated is the annual probability of collapse, method 2a tends to provide appropriate results, irrespectively of building class, source model, GMPE and limit state definition criteria. In this case, the base of comparison between exact and approximate results was a pre-determined level of acceptable risk, defined as a maximum allowable annual collapse probability.

This study demonstrated the possible advantages and limitations of considering approximate solutions to the problem of hazard-compatible record selection and subsequent analytical fragility and loss assessments. Therefore, within the wide range of structural properties, response parameters, seismological source modelling options, and ground motion prediction equations assessed, the main contribution of this study lies on the robust proposal of suitable levels of approximation recommended to be used by the research and practitioner communities.

This page is intentionally left blank

Chapter 4 ON THE TREATMENT OF UNCERTAINTIES IN THE DEVELOPMENT OF FRAGILITY FUNCTIONS FOR EARTHQUAKE LOSS ESTIMATION OF BUILDING PORTFOLIOS

This chapter is based on the following reference:

Sousa, L.; Silva, V.; Marques, M.; Crowley, H. (2016). On the treatment of uncertainties in the development of fragility functions for earthquake loss estimation of building portfolios. *Earthquake Engineering and Structural Dynamics*, 45: 1955–1976. doi: 10.1002/eqe.2734.

Summary

State-of-the-art methods for the assessment of building fragility consider the structural capacity and seismic demand variability in the estimation of the probability of exceeding different damage states. However, questions remain regarding the appropriate treatment of such sources of uncertainty from a statistical significance perspective. In this study, material, geometrical and mechanical properties of a number of building classes are simulated by means of a Monte Carlo sampling process in which the statistical distribution of the aforementioned parameters is taken into consideration. Building on the findings of Chapter 3, record selection is performed in accordance with (exact) hazard-consistent distributions of a comprehensive set of intensity measures, and issues related with sufficiency, efficiency, predictability and scaling robustness are addressed. Based on the appraised minimum number of ground motion records required to achieve statistically meaningful estimates of response variability conditioned on different levels of seismic intensity, the concept of conditional fragility functions is presented. These functions translate the probability of exceeding a set of damage states as a function of a secondary sufficient intensity measure, when records are selected and scaled for a particular level of primary seismic intensity parameter. It is demonstrated that this process allows a hazard-consistent and statistically meaningful representation of uncertainty and correlation in the estimation of intensity-dependent damage exceedance probabilities.

4.1 Introduction

The various sources of aleatory variability (and the correlation of their residuals) associated with ground-motion and structural response predictions cannot be neglected in loss assessment procedures, as demonstrated by several authors (e.g. Bommer & Crowley (2006), Bazzurro & Luco (2005), Weatherill et. al. (2015)). Hence, the purpose of this study is to investigate the appropriate treatment of material, geometrical and record-to-record variability in the derivation of fragility models for the earthquake loss estimation of building portfolios. In this context, one of the many challenges that arises when using analytical approaches to predict structural response is the choice of seismic intensity measure and ground motion input to apply in numerical simulations. A number of key ground-motion characteristics such as frequency content and spectral shape (e.g. Baker & Cornell (2006b)), peak ground motion (e.g. Bradley et. al. (2009b)), and duration (e.g. Iervolino et. al. (2006)) have been demonstrated to significantly influence predictions of the response of nonlinear systems, which typically renders record-to-record variability the main source of aleatory (i.e. random) variability (Shome & Cornell, 1999).

As stated by Bradley et. al. (2009b), intensity measures (*IMs*) shall ideally embody features of: *efficiency*, which is the ability of the intensity measure to predict the structural response with a small standard deviation, for a given set of records and statistical level of confidence (Shome & Cornell, 1999), *sufficiency*, which guarantees independence of response from parameters other than the intensity measure value of interest (Luco, 2002), *predictability*, which relates to the feasibility to estimate the measure from a ground motion prediction equation (Kramer & Mitchell, 2006), and *scaling robustness* (Tothong & Luco, 2007), which seeks to determine if the distribution of engineering demand parameters (*EDPs*) obtained using scaled ground motions is biased compared with that obtained using unscaled records. However, it is acknowledged in many applications (e.g. Bradley et. al. (2009b), (Luco & Cornell, 2007)) that none of the commonly used intensity measures (*IMs*) are *sufficient* with respect to the distribution of ground motion characteristics – namely, magnitude (*M*), distance (*R*), and *epsilon* (ε) – expected at a given site, as determined by probabilistic seismic hazard analysis *(PSHA*) (Cornell, 1968). Thus, it is clear that the response from nonlinear analysis will be dependent on the suite of selected records, as demonstrated by Haselton et. al. (2011), who assessed the influence of *epsilon* in the collapse fragility of a large number of structures.

As evidenced by Haselton et. al. (2012) a robust mechanism to determine structural response variability for a particular level of seismic action shall be based on a record-selection procedure that incorporates the prediction of both mean and variance of the considered intensity defining parameters. To this end, the Generalized Conditional Intensity Measure (*GCIM*) approach (Bradley, 2010a)] is employed in the selection of natural ground-motion records that are primarily scaled to match increasing levels of spectral ordinates at the mean fundamental period of vibration of the classes of interest - $Sa(T_1)$. According to the latter, conditional distributions of a relevant set of *IMs* are determined by taking into account all the rupture scenarios that influence the seismic hazard at the site of interest – Lisbon, Portugal – by means of the relative contribution of magnitude, distance and ground motion prediction models obtained from disaggregation (Bazzurro & Cornell, 1999), as formulated in Lin et. al. (2013).

In this study, thousands of nonlinear dynamic analyses were performed within a probabilistic methodology, developed by Silva et. al. (2015c), wherein hundreds of reinforced concrete frame models (with distributed plasticity) are simulated in a 2D environment. Through Monte Carlo simulation, the variability in the geometrical and material properties of typical two, five and eight-story pre-code reinforced concrete buildings in mainland Portugal is taken into account. As a result, the minimum number of ground-motion records necessary to achieve robust predictions of response variability is appraised, in order to achieve hazard-consistent and statistically meaningful distributions of structural response, conditioned on different levels of seismic intensity - $Sa(T_1)$. In this framework, matters of *efficiency*, *sufficiency*, *predictability* and *scaling robustness* are taken into account. Nonlinear response analysis of 100 structural models is performed for each selected ground-motion record, and damage exceedance probabilities are determined for each record, at each level of $Sa(T_1)$. The importance of computing "record-specific" probabilities is highlighted in the context of the loss estimation of building portfolios, which is strongly influenced by the spatial correlation of ground motion intensity parameters. To this end, it is demonstrated that the verified variability of "record-specific" probabilities is conditional on each level of $Sa(T_1)$, and dependent on intensity measures other than $Sa(T_1)$. Thus, these probabilities can be expressed as a function of a conditional intensity measure ($IMi_{|Sa(T_1)}$), which establishes the proposed concept of *conditional fragility function*.

4.2 Numerical Models

The approach used herein is similar to that presented in Chapter 3 (section 3.2), with the main difference being the generation of 100 assets per building class, rather than simply considering the structure that corresponds to the median capacity. More specifically, drawing upon the study by Silva et. al. (2015c), the numerical models considered herein represent typical buildings constructed before 1958, the year when the first seismic design provisions were enforced, and are thus defined as *pre-code*.

In this section, dynamic properties are characterized by the mean fundamental periods of vibration of the random generation of 100 assets with varying geometrical and material statistical distributions. These have been found to be 0.26, 0.45 and 0.70 seconds, for the two, five and eight story buildings, respectively. The percentage of reinforcement in the beams and columns is calculated following the pre-code regulations and practices corresponding to the ultimate and serviceability limit states, for each asset, in accordance with the sampled geometrical and material characteristics.

When using a Monte Carlo approach to randomly generate portfolios of buildings, it is important to ensure that convergence in the results is achieved. Accordingly, as demonstrated in a study by Silva et. al. (2014), in which a similar sampling framework was implemented, the use of one hundred assets is necessary to guarantee the statistical significance of the generated distribution of structural capacity. To maintain the computational effort at a reasonable level, each structure is modelled as a single infilled moment frame with three bays. As schematically presented in Figure 4.1 for the case of 5 story buildings, each frame was modelled in a 2D environment using the open-source software OpenSees (McKenna, Fenves, Scott, & Jeremic, 2000), with force-based distributed plasticity beam-column elements.

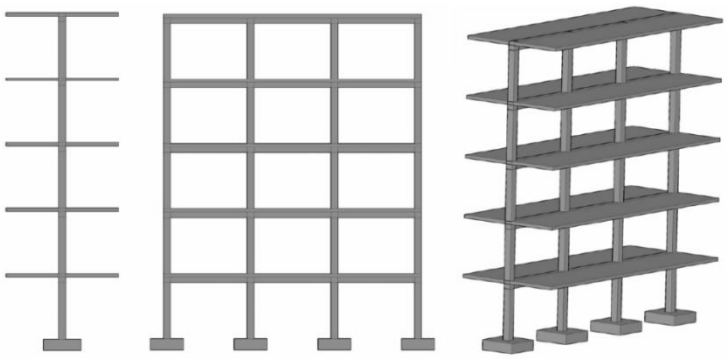


Figure 4.1 – Schematic view of the five-story RC frame model: front (left), side (centre) and isometric view (right) without infills, adapted from Silva et. al. (2014).

For the sake of synthesis herein, readers are referred to the aforementioned work by Silva et. al. (2015c) for details of the numerical considerations adopted with regards to the cross section discretization and integration points of the elements, the material constitutive relationships, P-delta effects, and the infill panel modelling approach.

4.3 Record selection methodology

In the analytical assessment of building fragility, the record-to-record variability should be robustly modelled given its significant influence on the estimated distribution of structural response (Lin, Haselton, & Baker, 2013). Amongst the available ground motion selection procedures, the Conditional Spectrum (CS), initially proposed by Baker (2011) and further developed by Jayaram et. al. (2011), provides a mechanism for estimating both the target mean and variance of spectral ordinates that a set of selected records should match, thus adequately accounting for the record-to-record variability. However, a limitation of the latter approach is that only the characteristics of ground motion represented in terms of spectral ordinates are considered. Thus, the Generalized Conditional Intensity Measure (GCIM) approach proposed by Bradley (2010a) is adopted herein for record selection, as it allows all the intensity measures identified as necessary to ensure that *efficiency*, *sufficiency*, *scaling robustness* and *predictability* are accounted for. A brief summary of the theoretical concepts behind the GCIM is provided below, and its application in the present study is presented further in the following sections.

The fundamental basis of the conditional response spectrum is that spectral accelerations at multiple vibration periods can be assumed to have a multivariate lognormal distribution, and the conditional distribution of spectral acceleration ordinates, for a single earthquake scenario, given the occurrence of a specific value of the spectral acceleration at some period, has a univariate lognormal distribution (Jayaram & Baker, 2008). In the GCIM, this concept is extended to any ground motion parameter of interest. In other words, the distribution of any *IMi* given an earthquake scenario, or rupture *Rup* (*IMi*|*Rup*), conditioned on the occurrence of a particular level of another intensity parameter (*IMj*), $f_{IMi|Rup,IMj}\left(im_i \middle| rup_k, im_j\right)$, can be assumed to have a lognormal distribution.

Upon the definition of appropriate Ground Motion Prediction Equations (*GMPE*) and correlation models between different intensity measures (*IMi),* the conditional

distribution of each *IMi* given *IMj = imj* is obtained via the total probability theorem as follows:

$$f(IM_i|_{IM_j=im_j}) = \sum_{n=1}^{N_{Rup}} f_{IMi|Rup,IMj}(im_i|rup_n, im_j) P_{Rup|IMj}(rup_n|im_j) \quad (4.1)$$

Where $f_{IMi|IMj}(im_i|im_j)$ is the probability density function (*pdf*) of *IMi* given *IMj=imj*, $f_{IMi|Rup,IMj}(im_i|rup_n, im_j)$ is the *pdf* of *IMi* given *IMj=imj* and *Rup=rup*$_n$, and $P_{Rup|IMj}(rup_n|im_j)$ is the contribution weight of *Rup=rup*$_n$, determined from seismic hazard disaggregation. From the assumption that the vector of all the considered *IMi* (herein referred as *IM*) is characterized by a multivariate lognormal distribution, it follows that for each *IMi*, the function $f_{IMi|Rup,IMj}(im_i|rup_n, im_j)$ has a univariate lognormal distribution, which can be defined by its conditional mean and standard deviation parameters:

$$\mu_{lnIM_i|Rup,IM_j}(rup_n, im_j) = \mu_{lnIM_i|Rup(rup_n)} + \\ + \sigma_{lnIM_i|Rup(rup_n)} \rho_{lnIM_i,lnIM_j} \varepsilon_{lnIM_j} \quad (4.2)$$

$$\sigma_{lnIM_i|Rup,IM_j}(rup_n, im_j) = \sigma_{lnIM_i|Rup}(rup_n) \sqrt{1 - \rho^2_{lnIM_i,lnIM_j}} \quad (4.3)$$

Which are determined as a function of ε_{lnIM_j}, the number of standard deviations, $\sigma_{lnIM_j|Rup}(rup_n)$, by which the logarithm of *IMj=im*$_j$ differs from the mean prediction of a particular *GMPE*, $\mu_{lnIM_j|Rup}(rup_n)$, for a given rupture scenario, *Rup=rup*$_n$:

$$\varepsilon_{lnIMj} = \frac{lnIM_j - \mu_{lnIM_j|Rup}(rup_n)}{\sigma_{lnIM_j|Rup}(rup_n)} \quad (4.4)$$

Accordingly, $\rho_{lnIM_i,lnIM_j}$corresponds to the correlation of residuals between different intensity parameters, presented in detail in section 4.3.3.

4.3.1 Probabilistic seismic hazard and disaggregation

A number of seismic hazard models exist for Portugal (e.g. Vilanova & Fonseca (2007), Sousa & Campos Costa (2009)), but only one of these has been selected herein for the purposes of demonstrating the methodology to link nonlinear response analysis with *PSHA*. The seismological source model has been taken from the study by Vilanova and Fonseca (2007), whilst the selection of the *GMPEs* was performed based on the

findings of Vilanova et. al. (2007), in which regional ground motion data from moderate magnitude earthquakes was used to verify the performance of different *GMPEs* in the Iberian region. Subsequently, the models developed by Atkinson & Boore (2006) and Akkar & Bommer (2010) are considered herein, with 0.70 and 0.30 logic tree weights, respectively, as further discussed in Silva et. al. (2015b).

Typically, causal earthquake magnitude, source-to-site distance and fault properties are considered in the definition of scenarios that contribute to the hazard in a given site, and are established by disaggregation of the PSHA (Bazzurro & Cornell, 1999). However, following the developments made by Lin et. al. (2013) to the Conditional Spectrum framework, seismic hazard disaggregation should not be limited to Magnitude (*M*) and Distance (*R*), but also consider the influence of different *GMPEs*, in order to ensure the consistency between the target distributions of all considered intensity measures, *IMi*, and the variability of ground motion properties expected at the site of interest (Lisbon, Lat. = 38.373, Lon. = –9.143). Thus, $f(IM_i|_{IM_j=im_j})$ is estimated for each conditioning intensity level (see section 4.3.3) according to the contribution of all N_{Rup} scenarios and set of *GMPEs* considered, as described below:

$$f\left(IM_i|_{IM_j=im_j}\right) =$$

$$\sum_{m=1}^{N_{GMPE}} \sum_{n=1}^{N_{Rup}} f_{IM_i|Rup,IM_j}\left(im_i \middle| rup_n, im_j, GMPE_m\right) P_{Rup,GMPE_m|IMj}(rup_n, GMPE_m|im_j) \quad (4.5)$$

The OpenQuake-engine (Monelli et. al. (2012), Silva et. al. (2014)) which has been used herein for the probabilistic seismic hazard analysis based on rock site conditions (i.e. shear wave velocity in the top 30 m of the soil of 760 m/s) does not currently address 3D disaggregation on *M*, *R* and *GMPE* (Lin, Harmsen, Baker, & Luco, 2013); however, due to its open-source nature, it was possible to produce the necessary intermediate results for the computation of $P_{Rup,GMPE_m|IM_j}(rup_n, GMPE_m|im_j)$, as demonstrated below:

$$P_{Rup,GMPE_m|IM_j}\left(rup_n, GMPE_m \middle| im_j\right) = \frac{v\left(IM_j, Rup \middle| GMPE_m\right).P(GMPE_m)}{v(IMj)} \quad (4.6)$$

Where $P(GMPE_m)$ stands for the logic-tree weight assigned to $GMPE_m$, $v(IMj, Rup|GMPE_m)$ is the rate corresponding to the conditional probability of *IMj=imj*, using $GMPE_m$, assuming a Poissonian process, and $v(IMj)$ is the rate of occurrence of *IMj=imj*, computed from the corresponding rate of exceedance, as proposed by Bradley (2010a).

4.3.2 Record database

Only three seismic events with significant ground motion were ever recorded in Portugal. For this reason, in order to create a sufficiently large database of candidate records for selection, accelerograms from other regions in the world with similar geological and tectonic characteristics were gathered (e.g. Spain, France, Switzerland, and East United States). The properties of stable continent and active shallow crustal regions, as well as the corresponding faults influencing the seismic hazard were respected to the maximum extent, based on the information provided by Vilanova & Fonseca (2007) and Sousa & Campos Costa (2009).

The horizontal orthogonal components of 911 non-pulse ground motions were selected from PEER (2015) and ESMD (2015) databases, with 815 and 96 records from each, respectively. The large number of selected records from the PEER database could give rise to concern as to the influence they may have on the results presented herein. However, the underlying assumption of the record selection procedure employed herein is that the influence that the selected suites of natural ground motion records have on the nonlinear response of structures is only a function of the expected distribution of seismic intensity parameters to which they are matched (see section 4.3.3).

4.3.3 Selected intensity measures

As demonstrated in a study by Sousa et. al. (2014), in which *efficiency* of an extensive set of *IMs* has been evaluated in the context of fragility estimation, there are a number of intensity measures related to duration and number of cycles that do not provide statistically meaningful correlation with the structural response of the building classes considered herein. On the other hand, the intensity measure types that incorporate velocity and spectral shape characteristics systematically provide increased correlations with damage exceedance probabilities. Theoretically, any intensity parameter can be considered in the *GCIM* selection approach (Bradley, 2012a). However, the latter assumption hinges on a number of constraints that, in practice, currently limit the number of *IMi* that can be considered:

a) *Predictability* must be ensured, based on the availability of *GMPEs* for predicting marginal mean and standard deviation of the logarithm of each *IMi*;
b) It must be possible to determine the *correlation* between each intensity parameter considered.

The applicability of the selected *GMPEs* to the specific case of mainland Portugal renders spectral acceleration at a range of periods an obvious initial choice for the target *IMi*. Thus, in order to ensure that target distributions computed for *IMi* other than spectral ordinates are consistent with the ground motion properties to be expected at the site of interest, preference is given to *IMi* for which marginal median and logarithmic standard deviation can directly be determined or indirectly be inferred from the same *GMPEs*. Therefore, the vector of intensity measures considered (i.e. *IM*) includes intensity parameters (i.e. *IMi*) of peak ground acceleration (PGA), peak ground velocity (PGV), acceleration spectrum intensity (ASI) (Von Thun, Roehm, Scott, & Wilson, 1988), Housner intensity (HI) (Housner, 1952) and spectral ordinates within the range of 0.05 to 3.0 seconds, conditioned on the spectral acceleration (*IMj*) at the mean fundamental period of vibration of each class ($Sa(T_1)$). Thus, the probabilistic distribution of the selected *IM* vector conditioned on a given level of $Sa(T_1)$ is designated henceforth as $f(IM|_{Sa(T_1)=a})$, being determined according to the hazard-consistent probabilistic distribution of each *IMi*, given *Sa(T1)=a* - $f(IM_i|_{Sa(T_1)=a})$ - as established in Equation 4.5, and the correlation models summarized in Table 4.1.

For each conditional *M*, *R* and *GMPE*, median predictions of *PGA* and *PGV*, along with the associated logarithmic standard deviations, can directly be obtained from the aforementioned GMPEs. As for the case of *ASI* and *HI,* since both intensity measures result from integrating spectral quantities in the period domain (acceleration and pseudo-velocity, respectively), target distributions can be computed based on the statistical parameters provided for the distribution of spectral ordinates, as demonstrated by Bradley (2010c) and Bradley et. al. (2009a).

Table 4.1– Correlation models considered for application of the GCIM methodology

	$SA(T_i)$	PGA	PGV	ASI	HI
$SA(T_i)$	Baker & Jayaram (2008)	Baker (2007a)	Bradley (2012b)	Bradley (2011)	Bradley (2011)
PGA	-	-	Bradley (2012b)	Bradley (2011)	Bradley (2011)
PGV	-	-	-	Bradley (2012b)	Bradley (2012b)
ASI	-	-	-	-	Bradley (2011)

4.3.4 Record selection for 2D analysis

When analysing a 3D structure, as recommended in the literature (e.g. Haselton et. al. (2009), Jayaram & Baker (2010)), records shall be selected and scaled based on the geometric mean of both orthogonal components for a particular level of conditioning *IMj*, in order to ensure the consistency with the definition of *IMj* used for fragility assessment: the geometric mean of the interested intensity measure (IM_{GM}) in the present study (as in virtually all the cases in the literature). In the case of 3D models, if near-source ground motions are not of interest, a given pair of horizontal components can be randomly applied with respect to the building's orthogonal directions, as demonstrated by Huang et. al. (2009). The latter leads to the inherent conclusion that, when IM_{GM} is considered for selection and scaling, a particular direction of the idealized 3D system will be actually subjected to arbitrary component ground motion: IM_{ARB}. Consequently, a 2D model of a 3D structure shall be analysed according to record sets that reflect the variability of records in a given direction - IM_{ARB} - for each intensity level, rather than that of IM_{GM}, which is smaller.

In the present case, in order to enhance the number of candidates available for record selection, advantage has been taken from the fact that the structures are modelled as 2D frames. Within this framework, each orthogonal component from a given record is considered individually, in practice duplicating the size of the database. However, in order to achieve consistency between the definition of the conditioning *IMj* (i.e. $Sa(T_1)$) in the various stages of record selection, nonlinear analysis and fragility assessment, the target distributions of each *IMi* are computed for the arbitrary component ground motion (IM_{ARB}) rather than the geometric mean of the orthogonal components (IM_{GM}). In practice, this is achieved by correcting the marginal standard deviations provided by the selected *GMPEs*, $\sigma_{lnIM_i|Rup}(rup_n)_{GM}$, which are for the geometric mean, according to the proposal of Baker & Cornell (2006c). Arbitrary ground motion variability of a given *IMi* is thus estimated based on the knowledge of the correlation between residuals (*epsilon*) in each orthogonal direction, $\rho_{lnIM_{iX},lnIM_{iY}}$, as follows:

$$\sigma^2{}_{lnIM_i|Rup}(rup_n)_{ARB} = 2\frac{\sigma^2{}_{lnIM_i|Rup}(rup_k)_{GM}}{1+\rho_{lnIM_{iX},lnIM_{iY}}} \qquad (4.7)$$

Empirically derived equations for $\rho_{lnIM_{iX},lnIM_{iY}}$ are available for spectral ordinates (e.g. Baker & Cornell (2006a), which has been adopted in this study); however, that is

not the case for PGA and PGV, to the Author's knowledge. In that case, by examining the ratio between standard deviations provided for single component and geometric mean given by a GMPE that provides both (Campbell & Bozorgnia (2008) has been selected herein), it is possible to back-calculate the implied arbitrary component correlation coefficient, according to the previously presented equation.

Three key assumptions of the methodology presented in this section are:

a) The correlation models, $\rho_{lnIM_{i_X},lnIM_{i_Y}}$ and $\rho_{lnIM_i,lnIM_j}$, are applicable to the correlation of the residuals from the adopted *GMPEs*;
b) The relationship between arbitrary and geometric mean variability for PGA and PGV (Campbell & Bozorgnia, 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 10s, 2008) is consistent with the seismicity of the site of interest;
c) The adaptation of the process to derive target distributions of each *IMi* for arbitrary ground motion is valid, which assumes that the correlation between residuals of two distinct *IMs*, $\rho_{lnIM_i,lnIM_j}$, is equally applicable to arbitrary components with the same orientation.

As demonstrated by Baker & Jayaram (2008), empirical evidence suggests that the hypothesis outlined in c) can be applied for the case of spectral ordinates. However, further research is necessary to identify whether it is valid for *IMs* other than spectral acceleration. In addition, an appropriate validation exercise will also be required in the future for what concerns points a) and b).

4.4 Fragility assessment framework

As discussed in Chapter 3 (section 3.4), the use of local criteria to define limit states when generating fragility curves for a population of buildings may not be appropriate. Hence, in the study herein, structural response will similarly be evaluated based on the maximum inter-story drift (*ISD*) and global drift (*GD*), considering four damage states: Slight Damage (*SD*), Moderate Damage (*MD*), Extensive Damage (*ED*) and Collapse (*Col*). For the sake of consistency, the limit state criteria established in section 3.4 are considered herein, as further presented, for completeness.

4.4.1 Limit state criteria

GD limits are determined according to the evaluation of capacity of each frame through a displacement-based adaptive pushover (Antoniou & Pinho, 2004). Similarly to what has been considered by other authors (e.g. Erberik (2007)), displacement thresholds at each limit state are defined for each sampled frame without masonry infills (bare frame) according to the following assumptions:

- Slight damage: global drift at 50% of maximum base shear capacity;
- Moderate damage: global drift when 75% of maximum base shear capacity is achieved;
- Extensive damage: global drift at maximum base shear capacity;
- Collapse: global drift when 20% decrease of the base shear capacity is verified, or 75% of the ultimate global drift attained, whichever is achieved first. In this framework, ultimate capacity is considered to be achieved once lack of numerical convergence is verified, provided that maximum shear strength has been achieved prior to numerical instability. With respect to numerical simulation assumptions, readers are referred to the work of Silva et. al. (2015c).

The influence of infill panels, which translates to a significant decrease of displacement capacity, is accounted for by applying the reduction factors proposed by Bal et. al. (2010) for each aforementioned limit state.

For what concerns *ISD*, a fixed set of values per limit state are defined based on the evaluation of global damage with increasing inter-story drift from 25 dynamic tests performed in real reinforced concrete moment resisting frames by Rossetto & Elnashai (2003). In order to adapt the six damage states proposed by the latter Authors with the one being considered in this study, *light/slight damage* and *partial collapse/collapse* damage states have been merged, as follows:

- Slight damage: 0.08% maximum inter-story drift;
- Moderate damage: 0.30% maximum inter-story drift;
- Extensive damage: 1.15% maximum inter-story drift;
- Collapse: 2.80% or higher maximum inter-story drift.

4.4.2 Uncertainty in structural response

The treatment of uncertainty is one of the key aspects of the proposed framework. More specifically, an attempt is made to determine the influence of all the addressed

sources of variability – namely, material, geometrical and seismic input – on the resulting distribution of damage exceedance probabilities. It must be possible to achieve reasonable confidence in the estimated response variability, in a statistically meaningful manner, in order to appropriately determine the probability of exceeding each damage state, for each level of seismic intensity.

As schematically summarized in Figure 4.2, the distribution of *ISD* and *GD* is determined for different levels of $Sa(T_1)$ through nonlinear response history analysis (*NLRHA*) of a number of frames (N_F) subjected to a set of N_R ground-motion records.

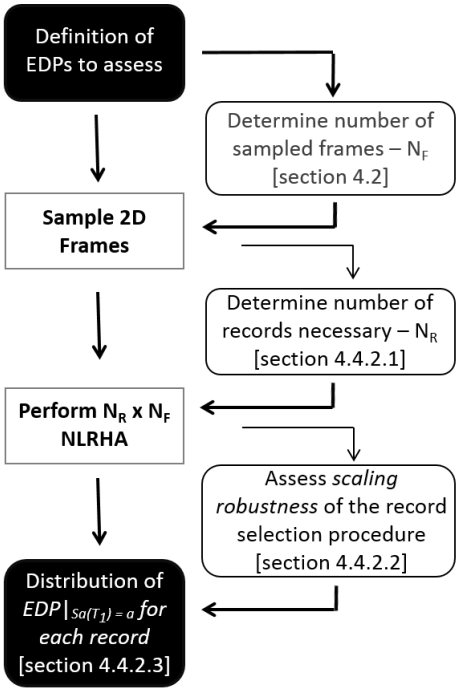


Figure 4.2 – Evaluation of building response distribution – methodology flowchart.

In this framework, variability in structural capacity, taken into account through the sampling of $N_F = 100$ frames, has been addressed through a probabilistic approach towards the modelling of material, geometrical and mechanical properties, as described in section 4.2. However, since records are selected and scaled based on target distributions of a set of *IMi* (section 4.3.3) that have distinct impacts on the spatial distribution of seismic demand (e.g. Bradley et. al. (2009b)), the number of ground motions required to achieve reasonable confidence in the estimated response variability is not known *a priori* (NIST, 2011). It is recognized in the literature that a large number (greater than thirty) is necessary for the aforementioned purposes (e.g. Haselton et. al. (2012), Lin et. al. (2013)); nonetheless, an accurate estimate is highly dependent on the parameters used to characterize response, as well as the structural properties itself. This matter is addressed

in the following section 4.4.2.1, as presented in Figure 4.2 whereas the bias potentially induced by record scaling in the probabilistic distribution of response parameters is verified in section 4.4.2.2.

4.4.2.1 Response variability and record selection – minimum number of records

A total of 150 records, selected according to the *GCIM* methodology to match target distributions of $IM - f(IM|_{Sa(T_1)=a})$ - is hereby assumed as a sufficiently large sample to provide an accurate evaluation of inter-story drift (*ISD*) and global drift (*GD*) distributions at each level of $Sa(T_1)$ in each sampled frame.

The minimum number of records necessary to achieve identical distributions within a given statistical significance level can thus be determined by comparing the latter with responses resulting from record sets of increasing size, selected to match the same target *IM*. Accordingly, the following methodology is devised, for the purposes of determining a minimum number of records necessary for nonlinear response analysis of 100 synthetically generated structures:

1. Distributions of *ISD* and *GD* resulting from nonlinear dynamic analysis of each of the 100 simulated frames (for 2, 5 and 8 story classes), are determined, using a set of 150 records for each level of $Sa(T_1)$;
2. A similar exercise is repeated for samples of 10 to 140 records (with steps of 10 records). These sets are selected to match the empirical distribution of *IM* derived from the *reference* set of 150 records, which ensures statistical consistency between distributions of *IM* amongst record samples, as determined by Kolmogorov-Smirnov (KS) goodness-of-fit tests within a 10 % significance level;
3. Empirical probabilistic distributions of *ISD* and *GD* obtained in step 2 are individually compared with the *reference* computed in step 1, for conditioning levels of $Sa(T_1)$ ranging from 0.1g to 1.0g (with 0.1g intervals).

The assumption that the logarithm of the response variables follows a normal distribution is further evaluated in section 4.4.2.3. Moreover, it is assumed that convergence on the mean prediction is achieved prior to convergence on the variance. Therefore, step 3 is performed using the Brown–Forsythe (*BF*) test (Morton & Forsythe, 1974), according to which the hypothesis that two sets of data have equal variance is assessed at the 5% significance level. Figure 4.3 illustrates the *BF* test statistic (*p-value*)

when comparing variances appraised in step 2 against the *reference* distributions computed in step 1.

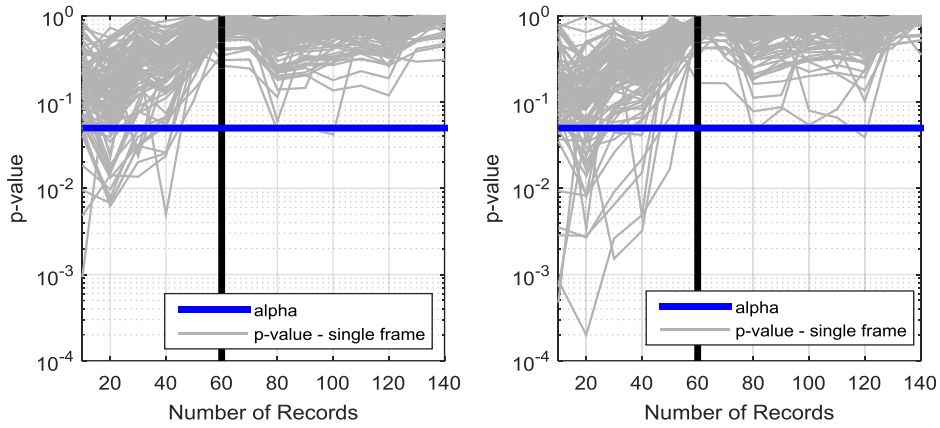


Figure 4.3 – BF test statistic (*p-value*) for 100 synthetic 5 story frames, records selected and scaled to a level of $Sa(T_1)$=1.0g. *P-values* higher than 0.05 indicate that the null hypothesis of equal variance cannot be rejected at 5 % significance, for *GD* (left) and *ISD* (right).

This test is preferred over other inference tools such as the F-test for equality of variances (Ang & Tang, 2007), which is highly sensitive to departures from normality. As illustrated in Figure 4.3, a value of 60 records is considered to provide an adequate compromise between computational effort and statistical significance of results in terms of variance in distribution of Global Drift and Inter-story Drift. Although only the results pertaining to 5 story frames and $Sa(T_1)$ equal to 1.0g are presented, the same conclusion is attained for samples of two and eight story frames, at all considered levels of conditioning seismic intensity parameter, $Sa(T_1)$.

4.4.2.2 Response variability and record selection – scaling robustness

According to the results of the previous section, sets of 60 records initially scaled in order to match the considered level of $Sa(T_1)$ are selected for the purposes of nonlinear response analysis of 2, 5 and 8 story frames. Conditional target distributions of *IM* (defined in section 4.3.3) given $Sa(T_1)=a$ (in which a ranges from 0.1g to 1.0g, with intervals of 0.1g) are computed according to the *GCIM* methodology introduced in section 4.3, according to which limits on record scaling are not imposed. In order to do so, the algorithm provided by Bradley (2015) has been modified in order to include the contribution of all the ruptures scenarios influencing the seismic hazard at the interested site, determined by 3D disaggregation on *M*, *R*, *GMPE* (illustrated in Figure 4.4, with respect to conditional $Sa(T_1)=0.5g$, for 2 and 5 story frames).

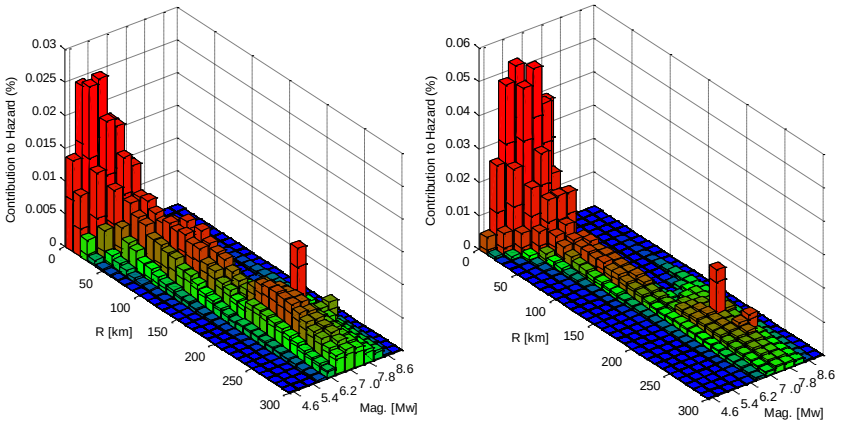


Figure 4.4 – Contribution to hazard determined by disaggregation on *M, R, GMPE* . For each *M, R* pair, lower and upper "bars" illustrate the contribution of Atkinson & Boore (2006) and Akkar & Bommer (2010) *GMPEs*, respectively. Conditional $Sa(T_1)=0.5g$, for 2 story (left) and 5 story frames (right).

For the sake of synthesis, Figure 4.5 provides examples of target and empirical distributions of PGA and spectral ordinates (0.05 to 3.0 sec.) resulting from selection, conditioned on $Sa(T_1)=0.5g$, for the case of 5 story frames (T_1=0.45 sec.).

One of the main advantages of the applied record selection procedure is the possibility to inherently account for the influence of *M*, *R* and *epsilon* on the ground motion properties expected at the site. Thus, matters of *sufficiency* with respect to explanatory variables other than the considered *IMi* are envisaged through evaluation of structural response dependence on scaling factor (i.e. *scaling robustness*). The bias potentially induced by record scaling in the distribution of response parameters conditioned on a particular level of seismic intensity can be examined when linear least-square regression is applied to the logarithm of the variables (Baker, 2007b), i.e. *GD* and *ISD* vs. scaling factor *(SF),* for each sampled frame.

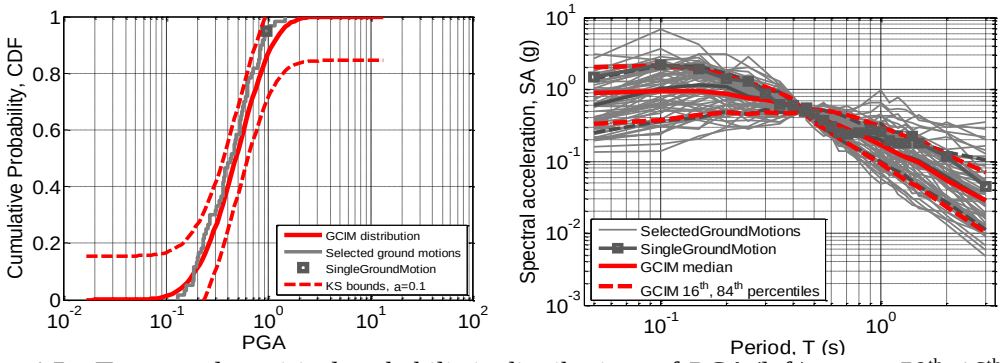


Figure 4.5 – Target and empirical probabilistic distributions of *PGA* (left), target 50th, 16th and 84th percentile spectral ordinates between 0.05 sec and 3.0 sec, and selected ground motions (right). Conditioning $Sa(T_1)=0.5g$, for 5 story frames (provided by selection algorithm, adapted from Bradley (2015)).

Assuming that the logarithms of *GD* and *ISD* are normally distributed, as later verified, the F-test (Ang & Tang, 2007) is used to determine the statistical significance of the "slope" obtained from regression analysis of the response of each sampled frame to the total set of 60 records. The null hypothesis that the predicted regression slope is zero is thus tested versus the alternative of being at least as large as the one obtained from regression. As a result, *p-values* (inferred from the F-test) lower than 1% are hereby considered to indicate statistical significance of the observed slopes, i.e. dependence of structural response on scaling factors. As depicted in Figure 4.6, where *p-values* corresponding to the regressions performed for each of the simulated frames are illustrated (as well as the 16% percentiles out of the total sample of 100 assets), although limits on *SF* have not been imposed, the distributions of response parameters do not show statistically significant dependence on the latter.

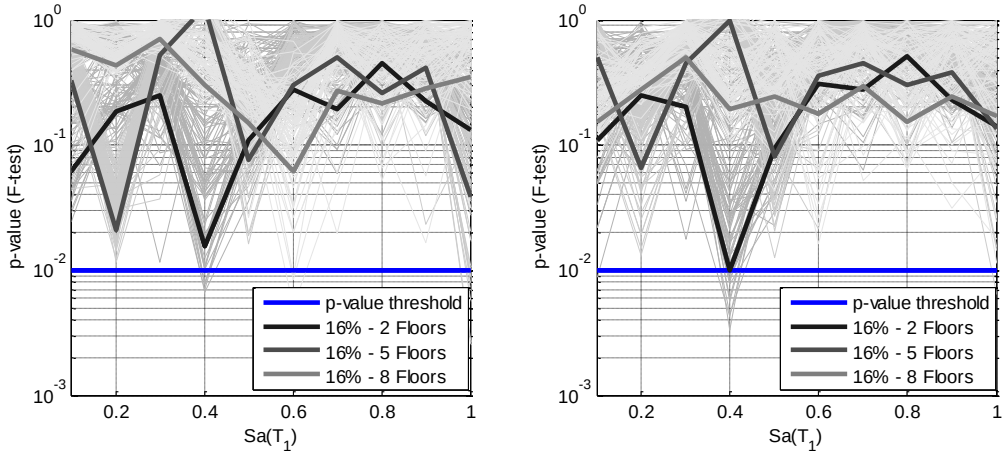


Figure 4.6 – Assessment of *scaling robustness* of the employed record selection methodology for demand parameters of Global Drift (left) and Inter-story Drift (right).

4.4.2.3 Response variability and record selection – distribution of EDPs given intensity

Existing studies have demonstrated that distributions of several demand parameters conditioned on different intensity levels can be assumed to follow a lognormal distribution (e.g. Shome & Cornell (1999)); however, a similar exercise is performed herein, in order to validate the assumption made in sections 4.4.2.1 and 4.4.2.2. The maximum likelihood method is applied to fit lognormal probability density functions to distributions of *ISD* and *GD* conditioned on all levels of $Sa(T_1)$, for 2, 5 and 8 story frames. Based on the latter, KS goodness-of-fit tests are performed in order to assess whether the null hypothesis that the logarithm of the response variables follow a normal

distribution is rejected at a 5% significance level. Since nonlinear response analyses were performed using 60 ground motion records applied to the 100 simulated frames, it is possible to determine "record-specific" distributions of *ISD* and *GD*.

As depicted in Figure 4.7, in which the statistical results for the three building classes are presented, only a residual sample of the results - below the 5% percentile - fails to confirm the assumption of normal distribution of response parameters, at the statistical significance level of interest. However, the latter can be explained by the finding that as structures with more significant ductility capacity approach severe stages of damage, small increments in ground motion intensity produce very large deformations. In extreme cases, the latter are considered to be *outliers* to the sample's underlying probability function, and a procedure developed by Grubbs (1969) has been used to identify observations whose deviation can be considered as undesired mechanisms or errors.

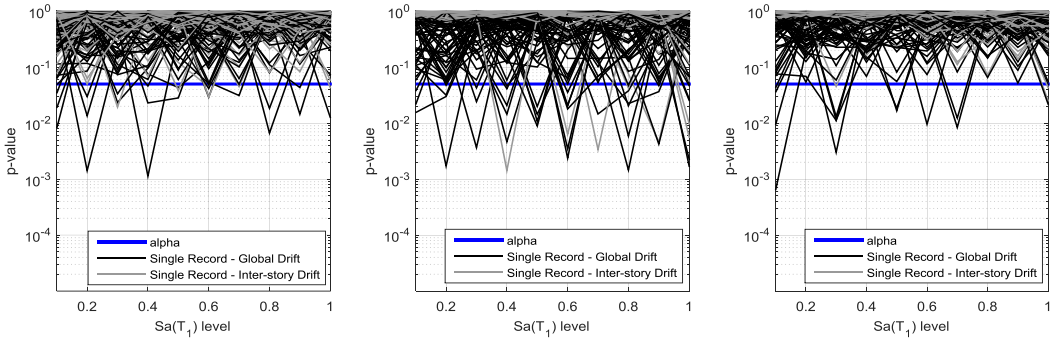


Figure 4.7 – P-value of the KS test obtained for each ground motion record, 2 story (left), 5 story (middle) and 8 story (right) classes. P-value higher than 0.05 indicate that the null hypothesis that the sample follows a normal distribution cannot be rejected at a 5 % significance level, for *GD* and *ISD*.

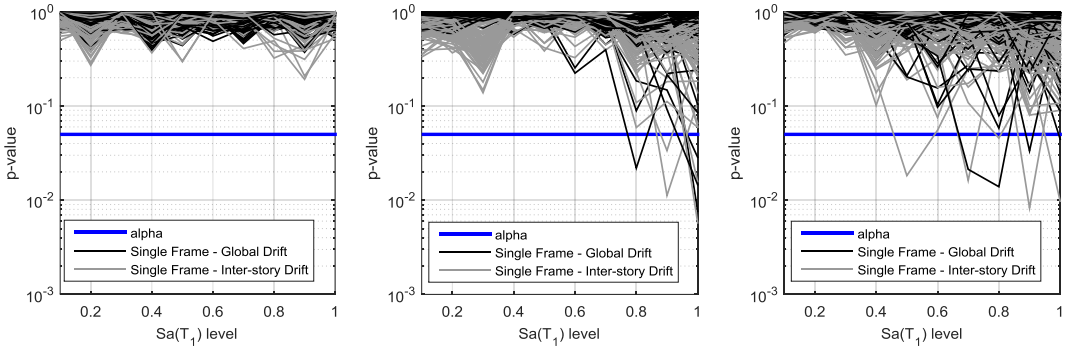


Figure 4.8 – P-value of the KS test obtained for each synthetically generated frame, 2 story (left), 5 story (middle) and 8 story (right) classes. P-value higher than 0.05 indicate that the null hypothesis that the sample follows a normal distribution cannot be rejected at a 5 % significance level, for *GD* and *ISD*.

Nonetheless, cases where such observations cannot be rejected, within a given statistical significance level, lead to departures from the assumed distribution, as

illustrated in Figure 4.8 by the decrease in *p-values*, for higher levels of $Sa(T_1)$, where "frame-specific" distributions of *ISD* and *GD* are evaluated based on the response attained in the analysis of each record.

4.4.3 Uncertainty in damage exceedance probability

Fragility Functions typically describe the uncertainty in the capacity of a structural system when subjected to sets of ground motions representing increasing values of seismic demand. In the present framework, as schematically presented in Figure 4.9, building fragility is characterised through probabilistic distributions of damage exceedance probabilities referring to damage states of *SD*, *MD*, *ED* and *Collapse (Col)*, for each level of $Sa(T_1)$. The latter distributions, presented in section 4.4.4, are thus evaluated by means of the computation of record-specific exceedance probabilities based on *ISD* and *GD* criteria, according to the methodology presented in sections 4.4.3.1 and 4.4.3.2.

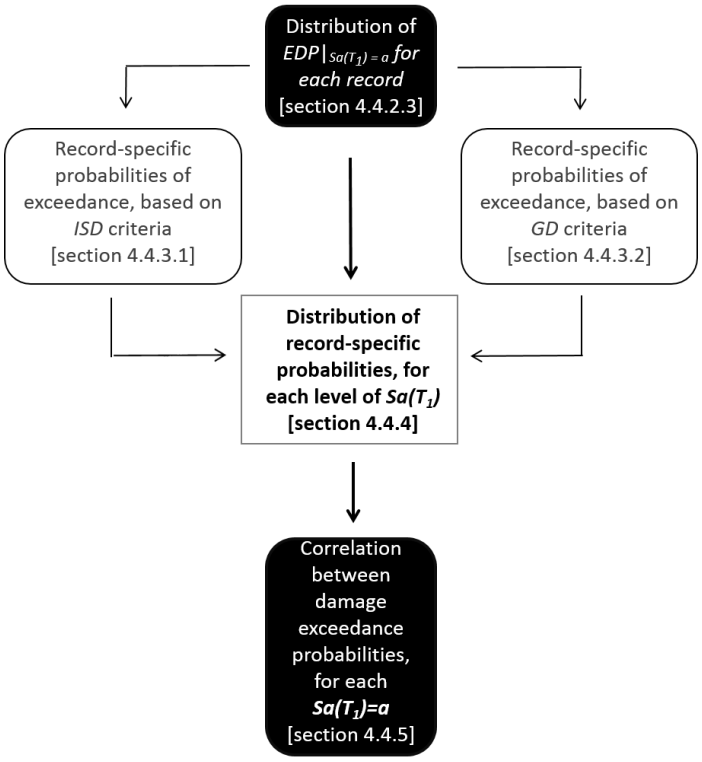


Figure 4.9 – Evaluation of uncertainty and correlation in damage exceedance probabilities – methodology flowchart.

Matters of correlation between the aforementioned probabilistic distributions are furthermore addressed in the present framework, whereby the concept of *conditional fragility function* is introduced. The latter reflects the probability of exceedance of

different damage states as a function of a conditional intensity measure, *IMi*, when records are selected and scaled for a particular level of $Sa(T_1)$. It allows a hazard-consistent treatment of record-to-record variability in the evaluation of fragility, while, as presented in 4.4.5, establishing the means by which different aspects of damage correlation in spatially distributed building portfolios are evaluated.

4.4.3.1 Record specific probabilities of exceedance – ISD criteria

As depicted in Figure 4.10, record-specific probabilities of exceedance based on *ISD* criteria are evaluated upon the verified assumption that distributions of structural response follow a lognormal probabilistic function, as follows:

$$Pls_{i_{|rec_j,Sa(T_1)=a}} = 1 - F_{rec_j,Sa(T_1)=a}(Ls_i) \qquad (4.8)$$

Where $Pls_{i_{|rec_j,Sa(T_1)=a}}$ is the probability of exceeding the limit state ls_i for record rec_j and $Sa(T_1)=a$, and $F_{rec_j,Sa(T_1)=a}(Ls_i)$ is the probability that *ISD* is equal or lower than the response limit ls_i (given rec_j and $Sa(T_1)=a$).

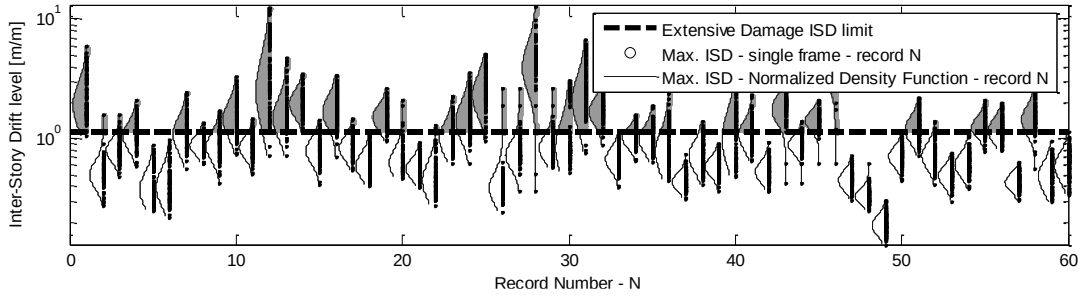


Figure 4.10 – Record-specific distributions of EDP and corresponding probabilities of exceedance of *ED*, determined according to ISD criteria for 8-story frames. Records selected and scaled for $Sa(T_1)=1.0g$.

Equation 4.8 reflects a view according to which the probability of exceedance of a given limit state is determined based on the underlying probabilistic distribution of *EDP* (Figure 4.10). An alternative approach is based on a *frequentist* interpretation of probability (Ross, 2009), i.e. the probability of exceeding a certain limit state equals the ratio between number of exceedances and the total number of performed analysis. Despite being used in several studies (e.g. Dumova-Jovanoska (2000), Rossetto & Elnashai (2005)), a *frequentist* interpretation imparts a significant shortcoming, since the fractions of exceedance are dependent on the total number of performed analyses. On the other hand, when representing reality via statistically significant idealizations of (lognormal)

random variables, it is possible to overcome sample size limitations and reach results that fully reflect the underlying probabilistic distribution of the assessed structural response parameter (Equation 4.8). In other words, the minimum number of analysis (100 per record) ensures that the null hypothesis that a "record-specific" distribution of response follows a given lognormal probabilistic distribution cannot be rejected (see Figure 4.7). Therefore, if one would have analyzed 101 frames per record (or any number higher than 100, for that matter), that would have changed the fraction of exceedances, but would not (in theory) affect the exceedance probabilities appraised based on the underlying theoretical probabilistic distribution.

Figure 4.11 illustrates the absolute error attained when evaluating record-specific probabilities of *SD, MD, ED* and *Collapse* computed as the ratio between number of exceedances and total number of analysed frames (one hundred) for each ground motion input, for all the conditional levels of $Sa(T_1)$. For comparison purposes, the latter results are plotted as a function of the result obtained through Equation 4.8.

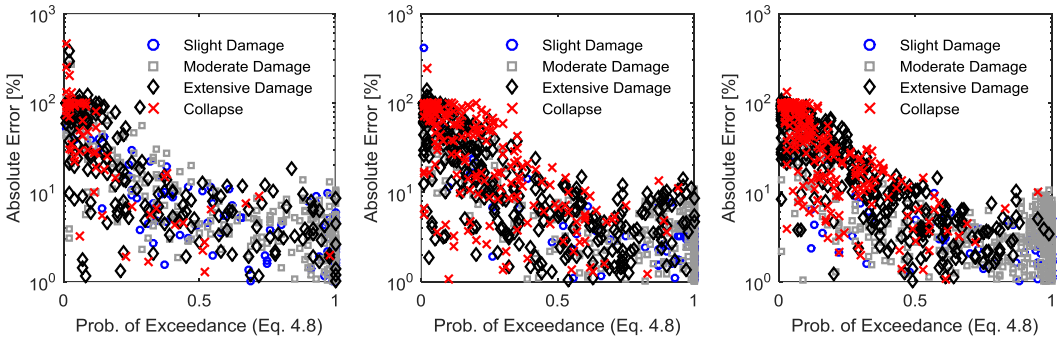


Figure 4.11 – Absolute Error attained when evaluating 600 record-specific probabilities of exceedance (60 ground motions times 10 levels of $Sa(T_1)$) computed as the ratio between number of exceedances and total number of analysis, as a function of the result obtained through Equation 4.8. Response criteria of *ISD*, for 2 (left), 5 (middle) and 8 story buildings (right).

As evidently demonstrated, significant errors (as substantial as 1000%) can arise from a *frequentist* interpretation of probabilities, with a near exponential trend that increases as the specified limit state approaches the right hand tale of the *ISD* distribution, i.e. higher errors are verified for smaller levels of probability.

4.4.3.2 Record specific probabilities of exceedance – GD criteria

The case of *GD* criteria offers a more challenging exercise. Unlike the *ISD* criteria, where damage state thresholds are similar for all the structures, *GD* criteria are specific to each sampled frame (see section 4.4.1). Therefore, even though the distribution of *GD* can be approximated by a lognormal distribution (section 4.4.2.3), Equation 4.8 cannot

be applied herein. In this case, probabilities of exceedance can be perceived as the number of successes in a sequence of independent experiments that yield success with identical probability. The latter is the theoretical background of the discrete binomial distribution, according to which the probability of the number of exceedances (*X*) in a total number of analyses (*N)* being equal to a given value *k* (i.e. *P(X=k)*) is given by the following equation:

$$P(X = k) = \frac{N!}{k!(N-k)!} (Pls_{i_{|rec_j, Sa(T_1)=a}})^k \, (1 - Pls_{i_{|rec_j, Sa(T_1)=a}})^{N-k} \qquad (4.9)$$

Where $Pls_{i_{|rec_j, Sa(T_1)=a}}$, further referred in this section as p_i for simplicity, is the probability that the record rec_j with $Sa(T_1)=a$ will cause the exceedance of limit state ls_i.

Assuming that the value of p_i is exactly equal to the fraction of exceedances in a sample of 100 frames would be a misleading *frequentist* interpretation of the phenomenon, according to which additional observations do not influence the appraised statistics. In fact, unless an extremely large number of analyses are performed in order to ensure the statistical significance of p_i, then the latter is simply one estimate of its "true" value.

A way to overcome the aforementioned limitation when a fixed number of frames (i.e. *N=100*) are analysed for each record is described subsequently. Based on the assumption that *P(X=k)* follows a discrete binomial distribution with parameters p_i and *N=100*, one can determine an empirical discrete distribution of *P(X=k)*. To do so, 10000 sets of 100 *GD* values are generated by means of a bootstrap simulation (Wasserman, 2004) where each dataset is obtained by sampling with replacement from the original group of 100 frame-specific *GD* values obtained from *NLRHA*. As a result, for each record and level of $Sa(T_1)$, *P(X=k)* for different *k* values is equal to the number of bootstrapped sets in which *X* is equal to *k*, divided by 10000. Therefore, *P(X=k)* can be approximated by a parametric binomial model with *N=100*, in which the fitted p_i is a statistically significant estimate of $Pls_{i_{|rec_j, Sa(T_1)=a}}$.

Similarly to what has been demonstrated in section 4.4.3.1, the absolute error attained when evaluating 600 record-specific probabilities of exceedance computed as the ratio between the number of exceedances and total number of analyses is illustrated in Figure 4.12, as a function of the value of $Pls_{i_{|rec_j, Sa(T_1)=a}}$ estimated through the methodology presented in this section.

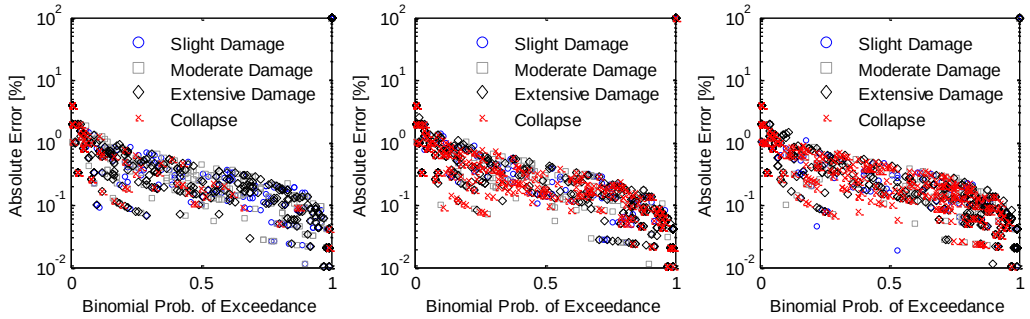


Figure 4.12 – Absolute Error attained when evaluating 600 record-specific probabilities of exceedance computed as the ratio between number of exceedances and total number of analysis, as a function of the result obtained through the methodology presented in this section. Response criteria of *GD*, for 2 (left), 5 (middle) and 8 story buildings (right).

4.4.3.3 Summary and conclusions from 4.4.3.1 and 4.4.3.2.

Although a similar trend is verified in Figure 4.11 and Figure 4.12, the maximum error registered in the case of *GD* criteria – 10% - is significantly lower when compared with that resulting from *ISD* criteria, which can be explained by the nature of the limit state definition criteria. In the case of *GD*, limit states are frame-specific and therefore dependent on the structural capacity of each sampled building, whereas in the case of *ISD* the specified limits are similar for all the frames and not directly related with its structural capacity. Thus, despite the fact that *ISD* and *GD* distributions are estimated for the same sets of frames, there is a higher standard error associated with each record-specific probability of exceedance computed based on *ISD* criteria, when the latter is evaluated as the fraction of exceedances in 100 frames.

The previous sections demonstrate that even in the case where the number of analyses is sufficient to guarantee a statistically significant distribution of response (section 4.2), non-negligible errors are attained if the estimation of probabilities is performed as a fraction of exceedances over the total number of sampled frames. Moreover, given the defined number of sampled buildings, these errors are significantly higher when constant (*ISD*) limits are used, as opposite to the case where the damage thresholds are derived for each frame.

Although an appropriate study of the impact of such discrepancies on loss estimates is submitted to further research, its possible implications cannot be ignored. As widely acknowledged (e.g. Baker (2015)), it is more important to adequately quantify the lower range of probabilities of exceedance, because ground motions with low values of intensity occur much more frequently and thus influence loss estimates more significantly than the

higher probabilities of exceedance, usually associated with less frequent high levels of intensity.

4.4.4 Uncertainty in record-specific probabilities of exceedance

As briefly presented in 4.4.3, the present framework foresees the characterization of building fragility through probabilistic distributions of damage exceedance probability, denoted as $f[Pls_{i_{|Sa(T_1)=a}}]$. In order to do so, 60 record-specific probabilities of exceedance of *SD, MD, ED* and *Col.* are estimated according to *ISD* and *GD* criteria for *Sa(T₁)=0.1g* to *1.0g*, as illustrated in Figure 4.13 for 5-story buildings.

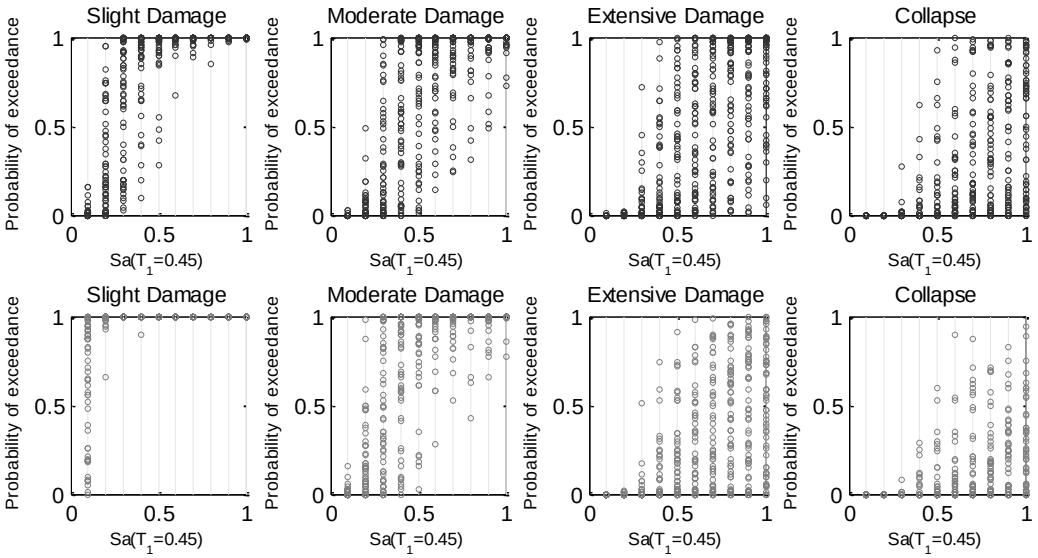


Figure 4.13 – Record-specific probabilities of exceedance of *SD, MD, ED* and *Col*, as a function of *GD* (upper) and *ISD* criteria (lower), for 5-story buildings.

As presented in Figure 4.14, the aforementioned probabilities are considered as realizations of random variables, based on which it is possible to determine the associated empirical density function, for each level of *Sa(T₁)* and damage state. Accordingly, the approximation of a parametric function is evaluated through Kolmogorov-Smirnov (KS) goodness-of-fit tests (Ang & Tang, 2007) tests used to assess the null hypothesis that the underlying distributions follow a *Beta* probabilistic model (Ross, 2009).

A visual inspection of the fit between theoretical and empirical distributions illustrated in Figure 4.14 highlights the capability of the considered model to take into account variations of probability density in the interval]0.0, 1.0[across different levels of *Sa(T₁)*. The aforementioned null hypothesis is thus verified according to KS tests performed on empirical and theoretical cumulative distribution functions, and cannot be rejected at a 10% significance level.

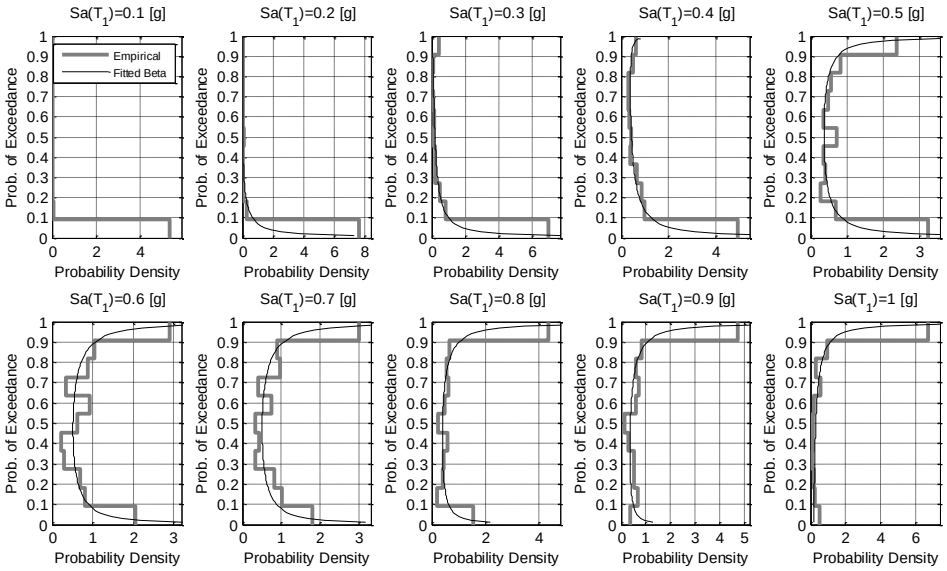


Figure 4.14 – Empirical probability density - $f[Pls_{i_{|Sa(T_1)=a}}]$ - of damage exceedance probability of *Extensive Damage* and corresponding fitted *Beta* models, damage criteria of *ISD* for 5-story frames.

However, it should be highlighted that whenever the *Beta* model is used for the purposes of earthquake loss estimation, appropriate attention should be given to the fact that exceedance probability values of 0.0 and 1.0 cannot be sampled from the latter. For the sake of synthesis, only the results pertaining to 5-story frames and damage state of *Extensive Damage* evaluated in terms of *GD* criteria are illustrated in Figure 4.14. Nonetheless, similar findings regarding the applicability of the selected theoretical model were attained for all structural classes and damage states.

4.4.4.1 Why determine record-specific probabilities of exceedance?

In the context of performance-based engineering, it is widely accepted that, in order to appropriately provide a link between seismic hazard and structural response, an "optimal" intensity measure - *Sa(T1)* in the present case - must embody features of *efficiency* (Shome & Cornell, 1999), *sufficiency* (2002), *predictability* (Kramer & Mitchell, 2006) and *scaling robustness* (Tothong & Luco, 2007). Moreover, provided that *sufficiency*, *predictability* and *scaling robustness* requirements are met, as demonstrated in sections 4.3 and 4.4, *efficiency* matters are related with the number of analyses necessary for the estimation of a satisfactory approximation to the "true" value of exceedance probability within a specified standard error limit, designated herein as $\tilde{P}ls_{i_{|Sa(T_1)=a}}$.

In the present case, the minimum number of sampled frames and ground motion records required for a statistically significant characterization of structural response have been determined. However, distributions of *EDP* and corresponding damage exceedance probabilities are estimated for each ground motion record in each level of $Sa(T_1)$. One might argue that this is an unnecessary step, because $\tilde{P}ls_{i|Sa(T_1)=a}$ can simply be obtained from the distribution of 6000 *EDP* values (60 ground motion records *x* 100 frames) for each level of $Sa(T_1)$, as illustrated in Figure 4.15, in which $\tilde{P}ls_{i|Sa(T_1)=a}$ is plotted against the results previously presented in Figure 4.13.

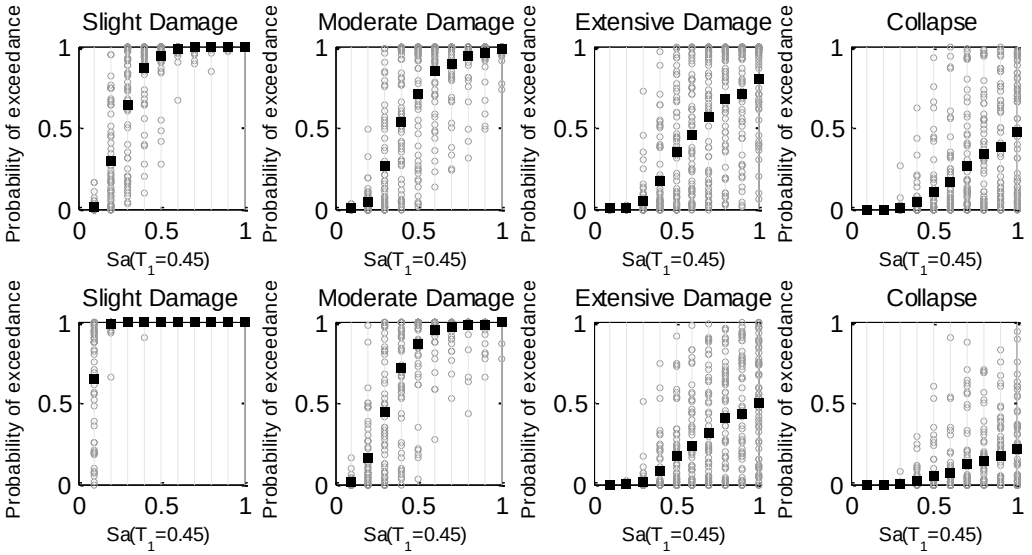


Figure 4.15 – Record-specific probabilities of exceedance of *SD*, *MD*, *ED* and *Col*, as a function of *GD* (upper) and *ISD* criteria (lower), and corresponding $\tilde{P}ls_{i|Sa(T_1)=a}$ (illustrated by the black squares), for 5-story buildings.

However, it is argued by the Author that considering $\tilde{P}ls_{i|Sa(T_1)=a}$ rather than $f[Pls_{i|Sa(T_1)=a}]$ leads to a misrepresentation of the impact of record-to-record variability in the appraised damage exceedance probabilities and, consequently, in the results of seismic loss estimation.

In order to demonstrate the aforementioned statement, a simple example is presented: the building damage of 2, 5 and 8-story buildings is associated simply with the *Collapse* damage state, with a corresponding *damage ratio, DR,* (ratio between the attained loss and the total replacement value of the asset) of 1.0. Strictly speaking, for the purpose of this exercise, the distribution of probabilities of collapse conditioned on $Sa(T_1)=0.5g$ ($f[PCol._{|Sa(T_1)=0.5g}]$) is equal to the distribution of damage ratios conditional on $Sa(T_1)=0.5g$. Thus, when considering a hypothetical portfolio of 100 buildings of the same structural class located at 100 different sites subjected to the same

value of $Sa(T_1)$ *(0.5g* for the purposes of this exercise), the mean (μ_{DR}) and variance (σ^2_{DR}) of the final distribution of aggregated *DR* can be computed according to Equation 4.10 and Equation 4.11, respectively:

$$\mu_{DR} = \sum_{k=1}^{100} \mu_{f[PCol._{|Sa(T_1)=0.5g}]} \quad (4.10)$$

$$\sigma^2{}_{DR} = \sum_{m=1}^{100} \sum_{n=1}^{100} \rho_{m,n} . \sigma^2{}_{f[PCol._{|Sa(T_1)=0.5g}]} \quad (4.11)$$

Where $\mu_{f[PCol._{|Sa(T_1)=0.5g}]}$ and $\sigma^2{}_{f[PCol._{|Sa(T_1)=0.5g}]}$ are respectively the mean and variance of $f[PCol._{|Sa(T_1)=0.5g}]$, which is considered similar in all the *k* locations, and $\rho_{m,n}$ is the spatial correlation coefficient between "record-specific" probabilities at two *m,n* locations.

The characterization of $\rho_{m,n}$ is further addressed in section 4.4.5. Nevertheless, it is clear from Equation 4.11 that it plays a very significant role in the loss estimation of spatially distributed portfolios, as evidenced in Figure 4.16, where the empirical distribution of *aggregated loss* computed through numerical simulation of the *Beta* approximation to $f[PCol._{|Sa(T_1)=0.5g}]$ at each site is illustrated for two extreme cases of zero and full spatial correlation.

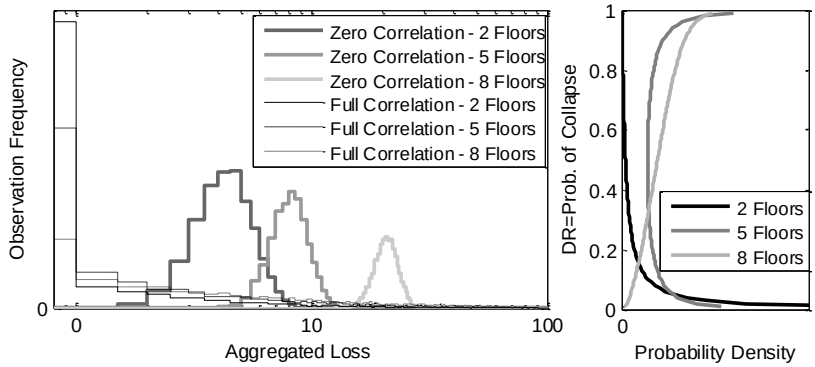
4.4.5 Correlation between damage exceedance probability

The previous sections highlight the importance of characterizing building fragility through probabilistic distributions of damage exceedance probability per level of $Sa(T_1)$. However, two important questions shall be addressed:

a) Is there a physical meaning underlying the assumption of spatial correlation between $f[Pls_{i_{|Sa(T_1)=a}}]$ (i.e. between "record-specific" damage exceedance probabilities) at different sites?
b) How can such correlation be adequately taken into account, in a hazard-consistent manner?

Regarding a) and b) above, two important aspects shall be evidenced. Firstly, the damage exceedance probability distributions presented in this framework arise from the computation of "record-specific" probabilities of exceedance for each level of $Sa(T_1)$. In this context, it is verified that the scatter depicted in Figure 4.13 for each level of $Sa(T_1)$ is the result of record-to-record variability. In other words, for a given level of $Sa(T_1)$, the variability in "record-specific" probabilities relates to the variation of a secondary

(conditional) intensity measure – $IMi|_{Sa(T_1)=a}$, which can be different for different levels of $Sa(T_1)$.



a)

Figure 4.16 – Empirical probabilistic distributions of *aggregated loss* computed for a hypothetical portfolio of 100 spatially distributed buildings subjected to $Sa(T_1)=0.5g$, with zero and full correlation between $f[PCol._{|Sa(T_1)=0.5g}]$ at each of the 100 sites. Limit state criteria of *GD* (left). Beta approximation to $f[PCol._{|Sa(T_1)=0.5g}]$ (right). Although not presented for the sake of visual clarity, the means of distributions with full and zero correlation are equal, as determined by Equation 4.10, whereas the variability changes proportionally to $\rho_{m,n}$ (Equation 4.11).

When assigning the record-specific values of each $IMi|_{Sa(T_1)=a}$ to the corresponding record-specific probabilities, the more *efficient IMi* can be selected as the one for which the correlation with damage exceedance probabilities of *SD*, *MD*, *ED* and *Col* is higher. Furthermore, it is demonstrated in Figure 4.17 that, for such $IMi|_{Sa(T_1)=a}$, regression analysis can be performed in order to fit a cumulative lognormal function to the scatter of *IMi*-dependent damage exceedance probabilities.

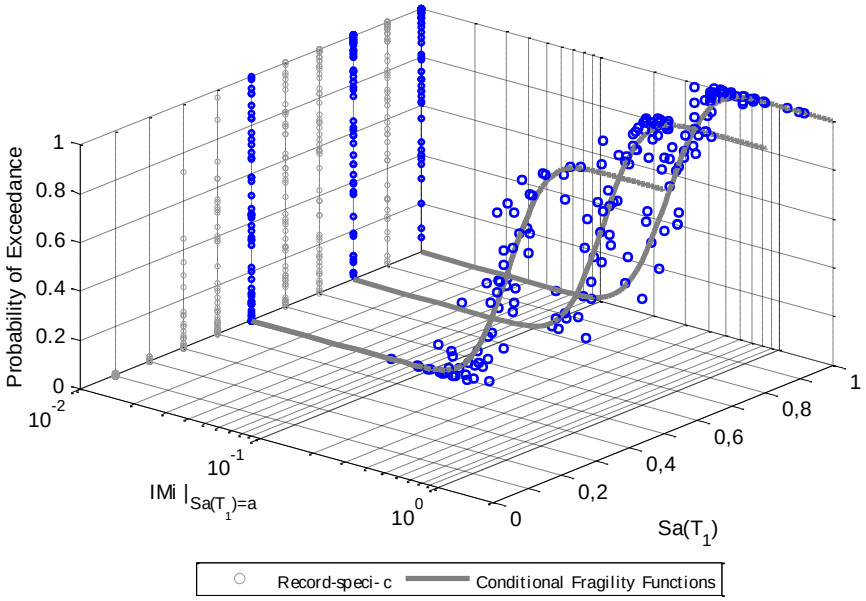


Figure 4.17 – Record-specific probabilities of exceedance of *ED*, as a function of *GD* criteria, and corresponding *conditional fragility functions* for the cases of $Sa(T_1)$=0.5g, 0.8g and 1.0g, for 5-story buildings

Given its conditional nature, such curves are hereby designated as *Conditional Fragility Functions*, providing a parametric relationship between $IMi|_{Sa(T_1)=a}$ and damage exceedance probabilities when records are selected and scaled for $Sa(T_1)=a$.

For the sake of illustration, Figure 4.18 presents the set of conditional *IMi* for which the correlation with damage exceedance probabilities is higher, for each level of $Sa(T_1)$. As demonstrated, different conditional *IMi* (and, therefore, different *conditional fragility functions*) are selected for each level of $Sa(T_1)$, which highlights the conditional nature of the proposed fragility model.

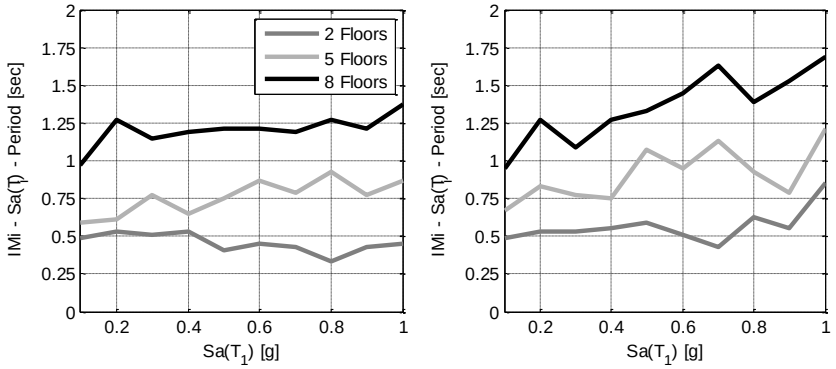


Figure 4.18 – Most *efficient* conditional *IMi* for each structural class and level of $Sa(T_1)$. For each level of $Sa(T_1)$, the correspondent *IMi* (which is also a spectral ordinate) is represented by the corresponding period of vibration. Fragility assessment in terms of *GD* (left) and ISD criteria (right).

It has been established that $f[Pls_{i|Sa(T_1)=a}]$ can be depicted as a function of a secondary intensity measure conditioned on $Sa(T_1)=a$, i.e. $IMi|_{Sa(T_1)=a}$ (Figure 4.17). Thus, the second important aspect to be highlighted is the fact that, if the spatial correlation between $IMi|_{Sa(T_1)=a}$ at different sites subjected to identical $Sa(T_1)=a$ can be determined, then $\rho_{m,n}$ (the spatial correlation between "record-specific" probabilities) has in fact a physical meaning. Since, as established by the *conditional fragility functions*, $f[Pls_{i|Sa(T_1)=a}]$ is a function of $IMi|_{Sa(T_1)=a}$, then "record-specific" damage exceedance probabilities can be assumed as random variables whose uncertainty relates to the record-to-record variability expressed by $IMi|_{Sa(T_1)=a}$. As a result, the spatial correlation between damage exceedance probabilities at different sites subjected to $Sa(T_1)=a$ is a function of the correlation between the values of $IMi|_{Sa(T_1)=a}$ at those same sites. In this context, the appropriate definition of the correlation between values of $IMi|_{Sa(T_1)=a}$ at different sites subjected to $Sa(T_1)=a$ and the impact of its consideration in a loss estimation procedure is addressed in Chapter 5.

4.5 Final remarks

This chapter presents a framework according to which multiple ground motion intensity measures are included in the characterization of building fragility through probabilistic distributions of damage exceedance probability for each level of $Sa(T_1)$. Variability of structural capacity and seismic demand have been considered in an analytical exercise where statistically significant distributions of response have been determined. Moreover, it is demonstrated that even in the case where the number of performed analyses are sufficient to ensure statistically significant distributions of structural response, non-negligible errors can be attained in estimation of damage exceedance probabilities if such computation is not performed in a statistically consistent manner. These errors have furthermore been verified to be dependent on the way the definition of response limit states is performed.

The relevance of the presented novel approach has been demonstrated within the context of loss estimation of building portfolios, where the spatial correlation of ground motion residuals plays a significant role. To this end, the importance of the introduced *conditional fragility functions* is illustrated by demonstrating its capability of consistently take into account record-to-record variability in the evaluation of fragility, while establishing the means by which spatial correlation between damage exceedance probability distributions can be taken into account.

Chapter 5 MODELLING SPATIAL CORRELATION OF DAMAGE RATIO RESIDUALS IN PORTFOLIO RISK ASSESSMENT

This chapter is based on the following reference:

Sousa, L.; Silva, V.; Marques, M.; Crowley, H. (2017). On the treatment of uncertainty in seismic vulnerability and portfolio risk assessment. *Earthquake Engineering and Structural Dynamics* (in press)

Summary

In the previous chapter, building fragility is represented in terms of intensity-specific distributions of damage exceedance probability of various damage states. The contribution of the latter has been demonstrated in the context of loss estimation of building portfolios, where it is shown that the proposed concept of conditional fragility functions provides the link between seismic intensity and the uncertainty in damage exceedance probabilities. In the present study, this methodology is extended to the definition of building vulnerability, whereby vulnerability functions are characterized by hazard-consistent distributions of damage ratio per level of primary seismic intensity parameter – $Sa(T_1)$. The latter is further included in a loss assessment framework, in which the impact of variability and spatial correlation of damage ratio in the probabilistic evaluation of seismic loss is accounted for, using test-bed portfolios of two, five and eight-story pre-code reinforced concrete buildings located in the district of Lisbon, Portugal. This methodology is evaluated in comparison with current state-of-the-art methods of vulnerability and loss calculation, highlighting the discrepancies that can arise in loss estimates when the variability and spatial distribution of damage ratio, influenced by ground motion properties other than the considered primary intensity measure, are not taken into account.

5.1 Introduction

In the context of earthquake risk modelling, state-of-the-art open-source software such as the OpenQuake-engine (Silva, Crowley, Pagani, Monelli, & Pinho, 2014)

provides a set of calculators capable of computing economic loss for a spatially distributed collection of assets by considering the probability of all possible events that might occur within a region over a certain time span. In the latter framework, vulnerability functions are described by a discrete list of intensity measure levels and corresponding mean loss ratio, as well as uncertainty, which is typically modelled using lognormal or beta distributions (Silva, Crowley, Pagani, Monelli, & Pinho, 2014).

In the process of risk computations, several studies have demonstrated the importance of accounting for spatial cross-correlation of ground motion residuals in the evaluation of portfolio losses (e.g. Park et. al. (2007), Weatherill et. al. (2015), Silva (2016)). However, when the correlation of uncertainty in vulnerability is incorporated in loss estimation procedures (e.g. Silva et. al. (2014)), it is done such that when sampling the uncertainty in the vulnerability of two assets with the same building class, the residuals are assumed to be either uncorrelated of perfectly correlated. Furthermore, as demonstrated by Bradley (2010b) and Silva et. al. (2013) in the context of component and building fragility, respectively, the propagation of uncertainty from fragility to vulnerability is related to the scatter of results to which a parametric (usually lognormal) fragility curve is fitted. However, although sound from a statistical point of view, the assumption of lognormal or beta distributions to model vulnerability uncertainty has in fact no evident physical meaning when analytical methodologies are used to derive fragility and vulnerability functions.

As highlighted by Taylor (2015), intensity and associated *damage ratio* (i.e. ratio between attained loss and replacement value of an asset) must in fact be jointly sampled in a way such that the underlying correlation reflects the physical phenomenon influencing the spatial variability (and correlation) of building response. Moreover, the importance of considering loss distributions with complete statistics as opposed to convenient but unrealistic simplifications such as the aforementioned beta or lognormal models is stressed by the author (Taylor, 2015). In this context, the research presented herein builds on the results presented in Chapter 4, extending that methodology to the derivation of vulnerability functions that reflect a non-parametric (site-specific) histogram of *damage ratios* per level of primary intensity measure - $Sa(T_1)$. This framework provides the link between vulnerability uncertainty and seismic hazard, such that when sampling the uncertainty in the vulnerability of two assets with the same building class and level of $Sa(T_1)$, the correlation of its residuals is physically explained by the spatial distribution (and correlation) of ground motion properties other than $Sa(T_1)$.

The aforementioned methodology is based on the concept of *conditional fragility functions*, presented in Chapter 4, which is validated in herein. Based on the latter, matters of uncertainty and correlation between damage state probabilities (i.e. probability of being exactly in a given damage state) are foreseen in the computation of vulnerability functions. As a result, a novel probabilistic loss assessment framework is presented, featuring the simulation of spatially correlated random fields of a set of intensity measures (*IMi*) conditioned on various levels of $Sa(T_1)$. The latter enables the explicit consideration of spatial correlation of residuals when sampling the uncertainty in the vulnerability of different building portfolios, demonstrated herein in the loss assessment of the building classes assessed in Chapter 4, i.e. two, five and eight-story pre-code reinforced concrete buildings located in the district of Lisbon, Portugal.

5.2 Conditional fragility functions: validation

As presented in the previous chapter, *conditional fragility functions* provide a (cumulative lognormal) parametric relationship between: a) record-specific probabilities of exceeding a certain limit state, and b) the corresponding record-specific values of a *sufficient IMi*, when records are selected and scaled for a particular level of $Sa(T_1)$. For a given structural class and level of $Sa(T_1)$, 60 ground motion records are selected and nonlinear response history analyses (NLRHA) are performed for a set of 100 numerical models that represent the variability in structural capacity (i.e. 60*100 analyses are carried out per level of $Sa(T_1)$). As a result, damage exceedance probabilities are evaluated for each record (denoted as 'record-specific' probabilities), based on the distribution of 100 values of maximum global drift (*GD*) and inter-story drift (ISD).

The Generalized Conditional Intensity Measure (*GCIM*) approach (Bradley, 2010a) is used in the selection of sets of 60 natural ground-motion records, whereby the resulting variability is consistent with the conditional probabilistic distribution of various intensity measures (*IMi*), conditioned on $Sa(T_1)$ being equal to a given value *a*. In this case, the considered vector of distinct *IMi* is referred as *IM*, and readers are referred to the aforementioned Chapter 4 for details regarding: (i) considered set of *IMi*, (ii) definition of the number of structural models and ground-motion records used per level of $Sa(T_1)$, (iii) computation of record-specific exceedance probabilities, and (iv) adopted limit states and associated definition criteria.

5.2.1 Conditional fragility functions: sufficiency of $IMi_{|Sa(T_1)=a}$

According to Chapter 4, one can assume that, in some form, $f[Pls_{i_{|Sa(T_1)=a}}]$ (i.e. the probabilistic distribution of record-specific probabilities of exceeding limit state ls_i given $Sa(T_1)=a$) is a function of $IM|_{Sa(T_1)=a}$, the vector of ground motion intensity measures considered in the process of record selection for $Sa(T_1)=a$, as mathematically formulated in the following Equation:

$$f[Pls_{i_{|Sa(T_1)=a}}] = \int_{IM} F\left[ls_{i_{|IM,Sa(T_1)=a}}\right].f\left(IM_{|Sa(T_1)=a}\right).dIM \qquad (5.1)$$

Where $F\left[ls_{i_{|IM,Sa(T_1)=a}}\right]$ represents the structural capacity, as the probability of exceeding limit state ls_i given IM and $Sa(T_1)=a$, $f\left(IM_{|Sa(T_1)=a}\right)$ is the probability density function of IM given $Sa(T_1)=a$, and IM is an intermediate variable that allows the problem of explicitly determining $f[Pls_{i_{|Sa(T_1)=a}}]$ to be tackled through the separate evaluation of $F\left[ls_{i_{|IM,Sa(T_1)=a}}\right]$ and $f\left(IM_{|Sa(T_1)=a}\right)$.

Determining $F\left[ls_{i_{|IM,Sa(T_1)=a}}\right]$ is clearly a very challenging task, due to the significant number of ground motion properties considered in the vector of intensity measures given $Sa(T_1)=a$, i.e. $IM_{|Sa(T_1)=a}$. Thus, the simplification introduced in the previous chapter is the assumption that a single *IMi* – the more *efficient* (Shome & Cornell, 1999) - is *sufficient* to account for all the explanatory variables in $IM|_{Sa(T_1)=a}$. In other words, it is proposed that $f[Pls_{i_{|Sa(T_1)=a}}]$ can be assumed as a function of $IMi|_{Sa(T_1)=a}$, rather than $IM|_{Sa(T_1)=a}$:

$$f[Pls_{i_{|Sa(T_1)=a}}] = \int_{IM_i} F\left[ls_{i_{|IM_i,Sa(T_1)=a}}\right].f\left(IM_{i_{|Sa(T_1)=a}}\right).dIM_i \qquad (5.2)$$

In which $F\left[ls_{i_{|IM_i,Sa(T_1)=a}}\right]$ is the *conditional fragility function* that translates the probability of exceeding a limit state ls_i, as a function of *IMi*, when records are selected and scaled for $Sa(T_1)=a$.

In the commonly accepted formulation, *sufficiency* is considered as the independence of structural response from parameters other than the intensity measure of interest (Luco, 2002), namely, magnitude (M), distance (R) and epsilon (ε). However, the influence of M, R and ε on $IM_{|Sa(T_1)=a}$ (and, therefore, on $IM_{i_{|Sa(T_1)=a}}$) has been taken into account in the process of record selection used in 4.3. Therefore, *sufficiency* is hereby

considered as the suitability of *IMi* to, as an intermediate variable, ensure the compatibility between the results appraised through Equation 5.1 and those obtained via Equation 5.2, as described below.

According to 4.4.4, Equation 5.1 is numerically solved by determining 60 record-specific damage exceedance probabilities for a given limit state, ls_i. and level of $Sa(T_1)$. In other words, one starts by selecting 60 ground motion records that comply with the theoretical distribution of $IM_{|Sa(T_1)=a}$, i.e. the obtained record set represents 60 realizations of $f\left(IM_{|Sa(T_1)=a}\right)$. As a result, NLRHA is performed for each realization (i.e. each record), and $f[Pls_{i_{|Sa(T_1)=a}}]$ is defined by the corresponding set of 60 record-specific damage exceedance probabilities. In the case of Equation 5.2, since both $F\left[ls_{i_{|IM_i,Sa(T_1)=a}}\right]$ and $f\left(IM_{i_{|Sa(T_1)=a}}\right)$ are known, a numerical method can be applied to find a numerical solution. If one samples a significantly large number of *IMi* values from $f\left(IM_{i_{|Sa(T_1)=a}}\right)$, it is possible to obtain $f\left[Pls_{i_{|Sa(T_1)=a}}\right]$ from the set of damage exceedance probabilities evaluated on $F\left[ls_{i_{|IM_i,Sa(T_1)=a}}\right]$ for each sampled value of *IMi*. Furthermore, as exemplified in Figure 5.1, $F\left[ls_{i_{|IM_i,Sa(T_1)=a}}\right]$ is a random variable whose regression uncertainty is herein determined by means of a bootstrap sampling method with replacement (Wasserman, 2004), using 200 synthetic datasets randomly generated from the original sets of 60 record-specific damage exceedance probabilities.

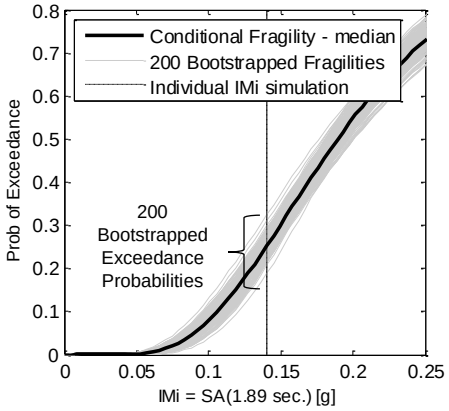


Figure 5.1 – Regression uncertainty of $F\left[Collapse_{|IM_i,Sa(T_1)=0.5g}\right]$, the *Conditional Fragility Function* of *Collapse* for 5-story frames, given $Sa(T_1)=0.5g$ and damage criteria of *GD*. For the considered level of $Sa(T_1)$, the most efficient *IMi* is the spectral acceleration at a period of vibration of 1.89 sec.

More specifically, since a distinct *conditional fragility function* is fitted to each synthetically generated set, 200 damage exceedance probabilities are determined for each simulated value of *IMi*. Therefore, for a sample of 2000 values of *IMi* obtained through

Monte Carlo sampling based on $f\left(IM_{i_{|Sa(T_1)=a}}\right)$, $f\left[Pls_{i_{|Sa(T_1)=a}}\right]$ determined through Equation 2 corresponds to the histogram of 2000 *x* 200 damage exceedance probabilities of the limit state and seismic intensity value of interest.

The Kolmogorov-Smirnov (KS) goodness-of-fit test (Ang & Tang, 2007) performed when comparing: a) distributions of damage exccedance probability attained with the methodology associated with Equation 5.1 - hereby referred as "exact", and b) corresponding distributions obtained with Equation 5.2 - further mentioned as "simulated", are illustrated in Figure 5.2.

Only the results pertaining to the limit state of *Collapse* and of $Sa(T_1)$ ranging from 0.7g to 1.0g are illustrated in Figure 5.2 (for 5 story frames), for simplicity. However, as presented in Figure 5.3, the null hypothesis that distributions of damage exceedance probability obtained from Equations 5.1 and 5.2 arise from the same underlying probabilistic distribution cannot be rejected at a 10% significance level, for all limit states and levels of $Sa(T_1)$ considered: *0.1g* to *1.0g* (with *0.1g* intervals).

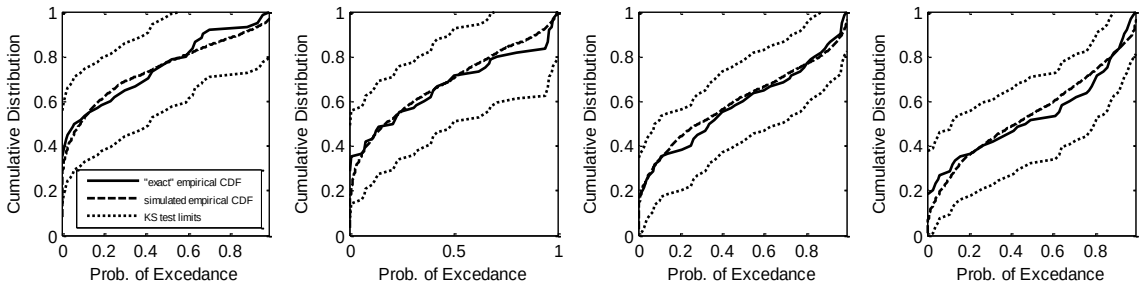


Figure 5.2 – Graphical illustration of the KS test performed when comparing distributions of collapse probability attained through Equation 5.1 ("exact") and Equation 5.2 (simulated) according to *Global Drift* criteria for 5 story frames and levels of $Sa(T_1)$ ranging from 0.7g to 1.0g.

Within the adopted statistical significance level, the various *IMi* selected for each level of $Sa(T_1)$ (illustrated from 0.1 to 1.0g in Figure 4.18) are *sufficient* to account for all the *IMi* influencing the nonlinear response and inherent damage exceedance probabilities of the assessed structures, conditioned on the presented range of $Sa(T_1)$. Consequently, it is herein assumed that Equation 5.2 provides a statistically significant approximation to Equation 5.1, demonstrating the validity of the proposed *conditional fragility curves*, which are a function of a single *IMi*.

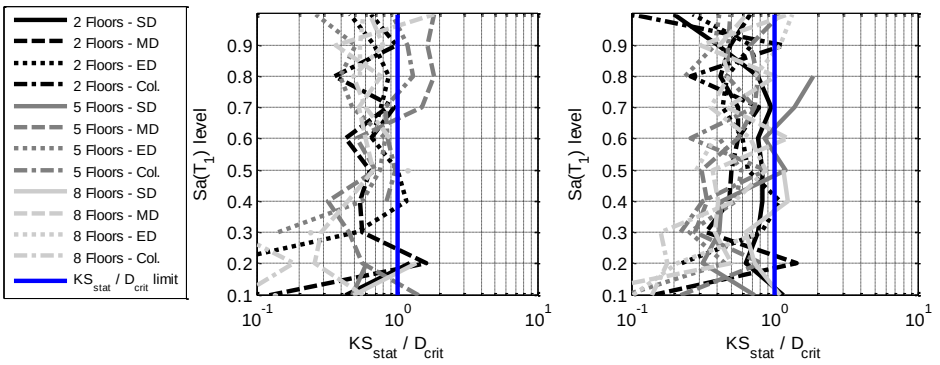


Figure 5.3 – Ratio between KS test statistics and critical value (*Dcrit*) when comparing distributions of exceedance probability determined based on Equation 1 and Equation 2. Ratios inferior to 1.0 indicate that the null hypothesis that the samples follow identical distributions cannot be rejected at a 10 % significance level, in terms of *ISD* (left) and *GD* (right).

5.3 From fragility to vulnerability: conditional fragility functions and loss estimation

In this section, it is demonstrated how the *conditional fragility functions* validated in the previous section are incorporated in a loss estimation framework. As further presented, this methodology allows for the representation of vulnerability functions through non-parametric (site-specific) probabilistic distributions of *damage ratio* per level of primary intensity measure - $Sa(T_1)$, as well as incorporating the spatial correlation of damage ratio residuals in the loss estimation of building portfolios. In this context, the damage states of *Slight Damage (SD), Moderate Damage (MD), Extensive Damage (ED)* and *Collapse (Col)* defined in terms of *ISD and GD* criteria are herein considered, as defined in section 4.4.1.

5.3.1 Probabilistic loss assessment methodology

The risk to earthquake action of a given portfolio of buildings is commonly described through a loss exceedance curve that specifies the frequency, usually expressed annually, with which specific values of loss will be exceeded. When using a single Ground Motion Prediction Equation (*GMPE*), this annual frequency, or rate, can be computed based on the application of the total probability theorem, as follows:

$$\gamma(L > l)_{GMPE_m} = \sum_{n=1}^{N} P\left(L > l|_{Rup_n, GMPE_m}\right).\gamma_n \qquad (5.3)$$

Where $\gamma(L > l)_{GMPE_m}$ is the annual rate of exceedance of loss l determined when using $GMPE_m$, $P\left(L > l|_{Rup_n, GMPE_m}\right)$ is the probability that loss L exceeds a specific value

l given the seismic event, or rupture, Rup_n , using $GMPE_m$, γ_n is the annual rate of occurence of Rup_n, and N is the number of different (assumed independent) possible earthquake ruptures determined by an earthquake rupture forecast (*ERF*) (Pagani, et al., 2014). In the present case, the seismological model developed by Vilanova & Fonseca (2007) has been used for the purposes of building the *ERF*, as described in Chapter 4.

The probability of exceedance of loss l given the occurrence of a Rup_n is an uncertain variable, expressed by the following equation:

$$P\left(L > l|_{Rup_n,GMPE_m}\right) = \int_I P(L > l|_I).f\left(I|_{Rup_n,GMPE_m}\right).dI \qquad (5.4)$$

In which I represents the spatial distribution of seismic intensity – $Sa(T_1)$ in the present case - across the portfolio of interest, $f\left(I|_{Rup_n,GMPE_m}\right)$ is the probability density function of I determined by $GMPE_m$, conditioned on Rup_n, and $P(L > l|_I)$ is the probability of exceedance of loss l given the spatial distribution of seismic intensity I.

Equation 5.4 considers the fact that, given a Rup_n, the distribution of intensity across a spatially distributed portfolio of assets is uncertain. Generally, it is not practical to solve the aforementioned equation in its closed form. Thus, given an earthquake rupture, the characterization of $f\left(I|_{Rup_n,GMPE_m}\right)$ in the dI domain is herein performed through the generation of a number of ground motion fields (Pagani, et al., 2014) of $Sa(T_1)$ that incorporates different simulations of spatially correlated ground shaking values at the location of the collection of assets.

Several studies have addressed the issue of generating spatially correlated ground motion fields of spectral ordinates (e.g. Weatherill et. al. (2015), Silva (2016)) and such matter will not be herein addressed in detail. However, provided that a number of random fields J is large enough to adequately reflect $f\left(I|_{Rup_n,GMPE_m}\right)$, then Equation 5.4 is numerically solved as presented in Equation 5.5:

$$P\left(L > l|_{Rup_n,GMPE_m}\right) = \sum_{j=1}^{J} P\left(L > l|_{Rup_n,GMF_j,GMPE_m}\right).\frac{1}{J} \qquad (5.5)$$

Where $P\left(L > l|_{Rup_n,GMF,GMPE_m}\right)$ is the deterministic value of probability of exceedance of loss l given the ground motion field j generated for the Rup_n and $1/J$ is one equiprobable realization of $f\left(I|_{Rup_n,GMPE_m}\right)dI$.

Equation 5.5 embodies what is commonly referred as a model mixture (Surajit & Lindsay, 2005) (i.e. sum of probability densities), according to which, if one considers

$P\left(L > l|_{Rup_n,GMF_j,GMPE_m}\right)$ as the mean of a normal random variable with zero variance, the probability density function of $P\left(L > l|_{Rup_n,GMPE_m}\right)$ is completely defined by a Gaussian mixture (*GM*) model parameterized by:

$$\mu^{GM}{}_{P\left(L>l|_{Rup_n,GMPE_m}\right)} = \left\{P\left(L > l|_{Rup_n,GMF_j,GMPE_m}\right)\right\};\ j = 1 \dots J \qquad (5.6)$$

$$\omega^{GM}{}_{P\left(L>l|_{Rup_n,GMPE_m}\right)} = \left\{\frac{1}{J}\right\};\ j = 1 \dots J \qquad (5.7)$$

In which $\mu^{GM}{}_{P\left(L>l|_{Rup_n,GMPE_m}\right)}$ and $\omega^{GM}{}_{P\left(L>l|_{Rup_n,GMPE_m}\right)}$ are respectively the mixture mean and weight vectors from all component densities.

Equation 5.3 is thus a weighted sum of (assumed) independent random variables, based on which its mean and variance are established by the following Equations 5.8 and 5.9.

$$\mu_{\gamma(L>l)_{GMPE_m}} = \sum_{n=1}^{N} \gamma_n \cdot \mu_{P\left(L>l|_{Rup_n,GMPE_m}\right)} \qquad (5.8)$$

$$\sigma^2{}_{\gamma(L>l)_{GMPE_m}} = \sum_{n=1}^{N} \gamma_n{}^2 \cdot \sigma^2{}_{P\left(L>l|_{Rup_n,GMPE_m}\right)} \qquad (5.9)$$

Where $\mu_{P\left(L>l|_{Rup_n,GMPE_m}\right)}$ and $\sigma^2{}_{P\left(L>l|_{Rup_n,GMPE_m}\right)}$ are the mean and variance of $P\left(L > l|_{Rup_n,GMPE_m}\right)$, respectively.

Despite the fact that $\mu_{P\left(L>l|_{Rup_n,GMPE_m}\right)}$ and $\sigma^2{}_{P\left(L>l|_{Rup_n,GMPE_m}\right)}$ can explicitly be determined, the body of the probabilistic distribution of $\gamma(L > l)_{GMPE_m}$ is not known *a priori*. Therefore, its probability density is herein empirically established through the numerical simulation of a sufficiently large number, *R*, of realizations (denoted as $\gamma^r(L > l)_{GMPE_m}$), with $r = 1 \dots R = 2000$), as follows:

$$\gamma^r(L > l)_{GMPE_m} = \sum_{n=1}^{N} P^r\left(L > l|_{Rup_n,GMPE_m}\right) \cdot \gamma_n \qquad (5.10)$$

In which $P^r\left(L > l|_{Rup_n,GMPE_m}\right)$ are (uncorrelated) random realizations of $P\left(L > l|_{Rup_n,GMPE_m}\right)$ from each of the *N* corresponding distributions.

5.3.1.1 Epistemic uncertainty and loss estimation results

A logic tree approach is used in this study in order to consider the epistemic uncertainty associated with the choice of *GMPEs* to use in conjunction with the

aforementioned seismological model, and, as presented in Chapter 4, the models developed by Atkinson & Boore (2006) and Akkar & Bommer (2010) are considered, with 0.70 and 0.30 logic tree weights, respectively.

A fundamental issue in deciding how to treat the epistemic uncertainty in hazard analysis is the interpretation of what the weights on the logic tree branches represent (Abrahamson & Bommer, 2005). As stated by Vick (2002), one interpretation favours the assumption that the weights are frequency-based probabilities of the alternative models being correct, whereas the alternative view is one according to which logic tree branches represent our relative confidence in the alternative models. As established in Chapter 4, the distributions of a given *IMi* conditioned on *Sa(T1)=a* used as target for record selection take into account the contribution of all the scenarios defined by 3D disaggregation on Magnitude, Distance and *GMPE,* which implies a frequency-based interpretation of logic tree weights assigned to different *GMPEs.* Therefore, the probabilistic distribution of loss exceedance rate determined by the contribution of all the considered *GMPEs* is herein determined by the following equation:

$$\gamma(L > l) = \sum_{m=1}^{N_{GMPE}} \gamma(L > l)_{GMPE_m} . P(GMPE_m) \qquad (5.11)$$

Where $P(GMPE_m)$ is the logic tree weight assigned to $GMPE_m$, and N_{GMPE} is the total number of *GMPEs* considered.

Similarly to Equation 5.5, Equation 5.11 represents the weighted sum of probability densities. However, since the distribution of $\gamma(L > l)_{GMPE_m}$ is defined numerically, the probabilistic density function of $\mu_{\gamma(L>l)}$ is herein computed through the simulation of N_{GMPE} sets of Q^m realizations of $\gamma^r(L > l)_{GMPE_m}$, in which $Q^m = \frac{P(GMPE_m)}{min\{P(GMPE_m)\}} \times R$ and $min\{P(GMPE_m)\}$ is the minimum of the logic tree weights assigned to each of the considered $GMPE_m$.

As highlighted by Abrahamson & Bommer (2005), if one considers that logic tree weights are, on the other hand, measures of the relative merit of each *GMPE*, then the mean value of $\gamma(L > l)$ as defined in Equation 5.11 does not correspond to the expected value in its strict statistical sense. The Author acknowledges the importance of this matter; however, such discussion is considered beyond the scope of this study.

5.3.2 Conditional fragility functions and computation of $P\left(L > l|_{Rup_n,GMF_j,GMPE_m}\right)$

According to Equations 5.6 and 5.7, $\gamma(L > l)_{GMPE_m}$ is defined by the contribution of all *J* ground motion fields of *Sa(T₁)* generated for each of the *N* ruptures in the *ERF*. Therefore, its computation is based on $P\left(L > l|_{Rup_n,GMF_j,GMPE_m}\right)$ determined for each of the *NxJ* simulations, as illustrated in Figure 5.4.

In this exercise, the OpenQuake-engine's hazard algorithms (Pagani, et al., 2014), openly available in its online repository (OpenQuake, 2016), are used for the purposes of implementing the aforementioned seismological model, the generation of the correspondent *ERF* and computation of ground motion fields of *Sa(T₁)*, using the spatial correlation model developed by Jayaram & Baker (2009). Furthermore, $P\left(L > l|_{Rup_n,GMF_j,GMPE_m}\right)$ is computed based on the generation of conditional ground motion fields of $IMi_{|Sa(T_1)=a}$, as subsequently presented.

5.3.2.1 Generation of ground-motion fields of $IMi_{|Sa(T_1)=A}$

Given GMF_j generated for Rup_n with $GMPE_m$, one can perceive *A* as the vector of simulated values of *Sa(T₁)* across a number of locations (L_p) of a portfolio (Figure 5.4).

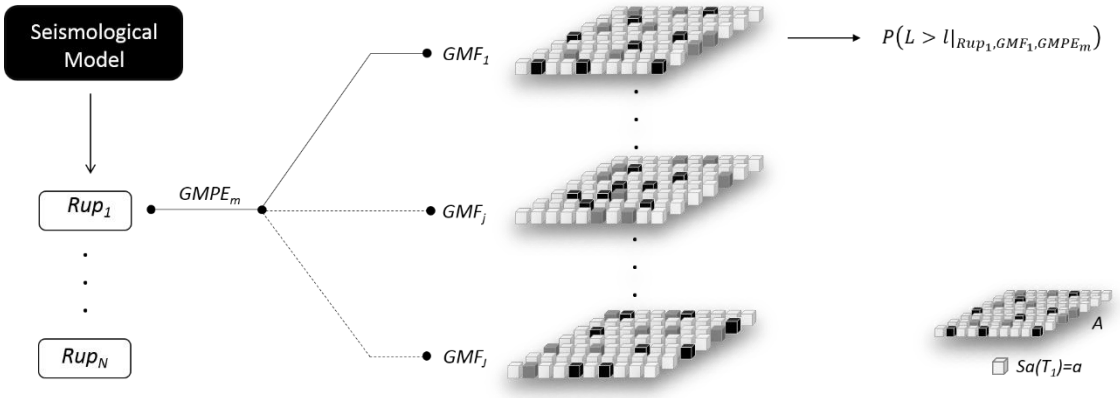


Figure 5.4 – Ratio between Schematic representation of the event-based simulation of ground motion fields of *Sa(T₁)*.

Thus, according to the mathematical formulation presented in Appendix 5.1 (based on the linear model of coregionalization (*LMCR*) proposed by Loth & Baker (2013)), it is shown that, for a given *A*, it is possible to obtain the mean and spatial covariance matrix of $IMi_{|Sa(T_1)=A}$. It shall be highlighted that, as illustrated in Figure 4.18, different *IMi* are selected for different values of *Sa(T₁)*. However, for simplicity, a given *IMi* selected for

a particular level of $Sa(T_1)$ is further generically referred as $IM_{i_{|Sa(T_1)=a}}$ (with i ranging from 1 to N^*), and IM^* is the vector that incorporates the set of N^* *efficient* and *sufficient* *IMi* selected for the N^*levels of $Sa(T_1)$ of interest. In this context, the mean and spatial covariance matrix of $IM^*_{|Sa(T_1)=A}$ are further denoted as $u_{IM^*|_{Sa(T_1)=A}}$ and $\Sigma_{IM^*|_{Sa(T_1)=A}}$, respectively.

Based on $u_{IM^*|_{Sa(T_1)=A}}$ and $\Sigma_{IM^*|_{Sa(T_1)=A}}$ (see Equations A.5.6 and A.5.7 of Appendix 5.1), it is possible to generate S sets of N^* spatially cross-correlated Gaussian random fields of IM^* conditioned on $Sa(T_1)=A$, according to Equation 5.12:

$$\begin{bmatrix} Y_{IM_1|_{Sa(T_1)=A}} \\ \vdots \\ Y_{IM_i|_{Sa(T_1)=A}} \\ \vdots \\ Y_{IM_{N^*}|_{Sa(T_1)=A}} \end{bmatrix} \text{with size } \begin{bmatrix} L_p x\ 1 \\ \vdots \\ L_p x\ 1 \\ \vdots \\ L_p x\ 1 \end{bmatrix} = u_{IM^*|_{Sa(T_1)=A}} + L.\begin{bmatrix} Z_1 \\ \vdots \\ Z_i \\ \vdots \\ Z_{N^*} \end{bmatrix} \qquad (5.12)$$

Where L is the lower triangular matrix obtained from Cholesky factorization such that $L.L^T = \Sigma_{IM^*|_{Sa(T_1)=A}}$, as demonstrated by Oliver (2003), Z_1 to Z_{N^*} are vectors of independent standard Gaussian distributed random values $(z_1, z_2 \dots z_{L_p})$, and $Y_{IM_i|_{Sa(T_1)=A}}$ is the spatial cross-correlated random field of IM_i conditioned on $Sa(T_1)=A$.

As schematically illustrated in Figure 5.5, it is assumed, for simplicity, that the values of $Sa(T_1)$ in A assume only two possible values x and y, for which IM_x and IM_y, respectively, are the corresponding conditional intensity measures. As a result, $Y^1_{IM_x|_{Sa(T_1)=x}}$ defines the values of $IM_x|_{Sa(T_1)=A}$ when $Sa(T_1)=x$, and $Y^1_{IM_y|_{Sa(T_1)=y}}$ includes the values of $IM_y|_{Sa(T_1)=A}$ when $Sa(T_1)=y$, for simulation S_1.

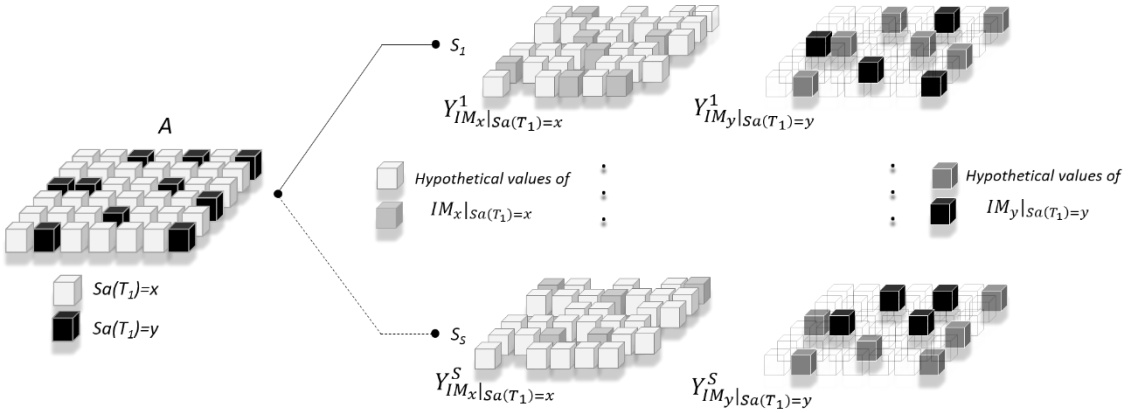


Figure 5.5 – Schematic illustration of the generation of S conditional spatially cross-correlated ground motion fields of $IM_{i_{|Sa(T_1)=A}}$ for GMF_j and Rup_n.

From a generic point of view, any $Y_{IM_i|_{Sa(T_1)=a}}$ (of length $L_L < L_p$) can be organized in the form of a matrix of *S* values of IM_i (referred as *imi*) per location, as illustrated in Figure 5.6.

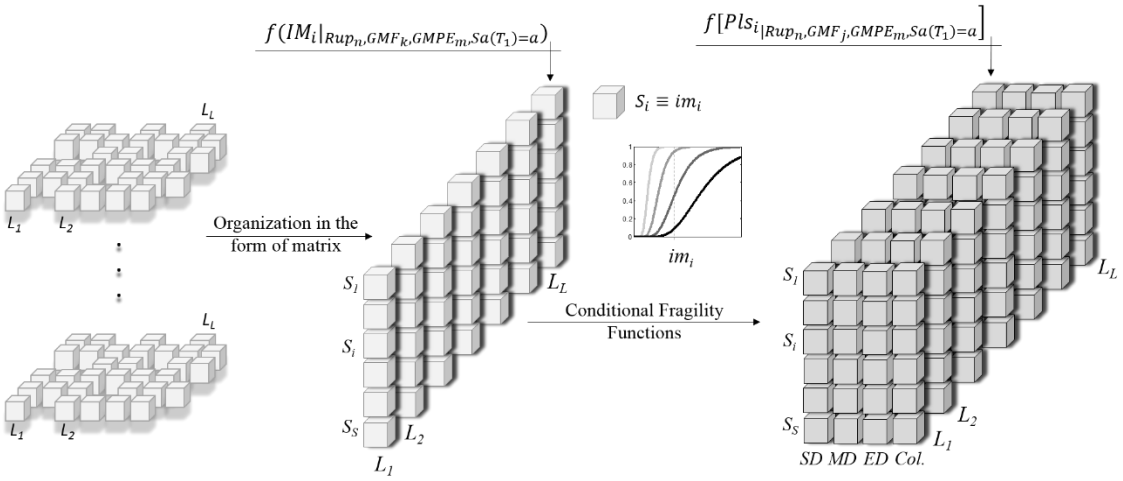


Figure 5.6 – Schematic representation of simulation of spatially correlated values of $IMi|_{Sa(T_1)=a}$ for GMF_j and Rup_n, and corresponding damage exceedance matrix.

This matrix establishes $f\left(IM_{i_{|Rup_n,GMF_j,GMPE_m,Sa(T_1)=a}}\right)$ at each site where $Sa(T_1)=a$ and, as a result, it follows from Equation 5.2 that $f\left[Pls_{i_{|Rup_n,GMF_j,GMPE_m,Sa(T_1)=a}}\right]$ at each site reflects the numerical application of the following Equation 5.13.

$$f\left[Pls_{i_{|Rup_n,GMF_j,GMPE_m,Sa(T_1)=a}}\right] =$$

$$= \int_{IM_i} F\left[ls_{i_{|IM_i,Sa(T_1)=a}}\right] . f\left(IM_{i_{|Rup_n,GMF_j,GMPE_m,Sa(T_1)=a}}\right) dIM_i \qquad (5.13)$$

5.3.2.2 Damage state probabilities for Rup_n and GMF_k

Because the bootstrapped $F\left[ls_{i_{|IM_i,Sa(T_1)=a}}\right]$ is an uncertain variable (see section 5.2.1), the probability of being exactly in a damage state ls_i given a certain value of *IMi* and $Sa(T1)=a$, herein referred as $P\left[ds_{i_{|IM_i=im_i,Sa(T_1)=a}}\right]$, is also a random variable. More specifically, Equations 5.14 and 5.15 show that, for a given level of *IMi* and $Sa(T1)=a$, $P\left[ds_{i_{|IM_i=im_i,Sa(T_1)=a}}\right]$ is evaluated based on the damage exceedance probabilities of different damage states defined by the *corresponding conditional fragility functions*:

$$P\left[ds_{i_{|IM_i=im_i,Sa(T_1)=a}}\right] = F\left[ls_{i_{|IM_i=im_i,Sa(T_1)=a}}\right] - F\left[ls_{i+1_{|IM_i=im_i,Sa(T_1)=a}}\right] \qquad (5.14)$$

$$\text{Or } P\left[ds_{i_{|IM_i=im_i,Sa(T_1)=a}}\right] = F\left[ls_{i_{|IM_i=im_i,Sa(T_1)=a}}\right], \text{ if } ds_i \text{ is } Collapse \quad (5.15)$$

Therefore, the mean and variance of $P\left[ds_{i_{|IM_i=im_i,Sa(T_1)=a}}\right]$ is determined as according to Equation 5.16 and Equation 5.17.

$$\mu_{P\left[ds_{i_{|IMi=im_i,Sa(T_1)=a}}\right]} = \mu_{F\left[ls_{i_{|IMi=im_i,Sa(T_1)=a}}\right]} - \mu_{F\left[ls_{i+1_{|IMi=im_i,Sa(T_1)=a}}\right]} \quad (5.16)$$

$$\sigma^2_{P\left[ls_{i_{|IMi=im_i,Sa(T_1)=a}}\right]} = \sigma^2_{F\left[ls_{i_{|IMi=im_i,Sa(T_1)=a}}\right]} + \sigma^2_{F\left[ls_{i+1_{|IMi=im_i,Sa(T_1)=a}}\right]} - 2 \times \rho_{ls_i,ls_{i+1}|IMi=im_i,Sa(T_1)=a} \times \sigma_{F\left[ls_{i_{|IMi=im_i,Sa(T_1)=a}}\right]} \times \sigma_{F\left[ls_{i+1_{|IMi=im_i,Sa(T_1)=a}}\right]} \quad (5.17)$$

Where $\mu_{F\left[ds_{i_{|IMi=im_i,Sa(T_1)=a}}\right]}$ and $\sigma^2_{F\left[ds_{i_{|IMi=im_i,Sa(T_1)=a}}\right]}$ correspond to the mean and variance of $F\left[ds_{i_{|IM_i=im_i,Sa(T_1)=a}}\right]$, and $\rho_{ls_i,ls_{i+1}|IMi=im_i,Sa(T_1)=a} \times \sigma_{F\left[ls_{i_{|IMi=im_i,Sa(T_1)=a}}\right]} \times \sigma_{F\left[ls_{i+1_{|IMi=im_i,Sa(T_1)=a}}\right]}$ is the covariance between distributions of damage exceedance probability for damage states ls_i and ls_{i+1} conditioned on $IMi=im_i$ and $Sa(T_1)=a$ (computed as a function of $\rho_{ls_i,ls_{i+1}|IMi=im_i,Sa(T_1)=a}$: the correlation between regression uncertainty of $F\left[ls_{i_{|IM_i=im_i,Sa(T_1)=a}}\right]$ and $F\left[ls_{i+1_{|IM_i=im_i,Sa(T_1)=a}}\right]$).

As illustrated in Figure 5.7, the uncertainty in fragility regression for $IMi=im_i$ and $Sa(T_1)=a$ is determined through a bootstrap method that generates 200 *conditional fragility curves* for each limit state. Thus, as theoretically demonstrated by Bradley (2010b) and Silva *et al*. (2013) in similar fragility simulation exercises, $F\left[ls_{i_{|IM_i=im_i,Sa(T_1)=a}}\right]$ can be approximated by a normal distribution with parameters $\mu_{F\left[ls_{i_{|IMi=im_i,Sa(T_1)=a}}\right]}$ and $\sigma_{F\left[ls_{i_{|IMi=im_i,Sa(T_1)=a}}\right]}$, as illustrated in Figure 5.7. Moreover, each bootstrap simulation is consistent across all damage states, i.e. the indices of the values that are drawn from the original dataset of *SD* exceedance probabilities are considered in order to constitute the corresponding bootstrap samples of *ED*, *MD* and *Collapse* probabilities for a given $Sa(T_1)=a$. Therefore, $\rho_{ls_i,ls_{i+1}|IMi=im_i,Sa(T_1)=a}$ is determined as the sample correlation between $F\left[ls_{i_{|IM_i=im_i,Sa(T_1)=a}}\right]$ and $F\left[ls_{i+1_{|IM_i=im_i,Sa(T_1)=a}}\right]$ obtained from the aforementioned consistently bootstrapped *conditional fragility functions*, as follows:

$$\rho_{ls_i, ls_{i+1}|_{IMi=im_i, Sa(T_1)=a}} = \frac{\sum_{k=1}^{200}\left(F_k\left[ls_{i|IM_i=im_i,Sa(T_1)=a}\right]-\mu_{F\left[ls_{i|IMi=im_i,Sa(T_1)=a}\right]}\right).\left(F_k\left[ls_{i+1|IM_i=im_i,Sa(T_1)=a}\right]-\mu_{F\left[ls_{i+1|IMi=im_i,Sa(T_1)=a}\right]}\right)}{\sqrt{\left(F_k\left[ls_{i|IM_i=im_i,Sa(T_1)=a}\right]-\mu_{F\left[ls_{i|IMi=im_i,Sa(T_1)=a}\right]}\right)^2}.\sqrt{\left(F_k\left[ls_{i+1|IM_i=im_i,Sa(T_1)=a}\right]-\mu_{F\left[ls_{i+1|IMi=im_i,Sa(T_1)=a}\right]}\right)^2}}$$

(5.18)

In which $F_k\left[ls_{i|IM_i=im_i,Sa(T_1)=a}\right]$ is the *k* bootstrapped *conditional fragility function* evaluated at $IMi=im_i$, and $\mu_{F\left[ls_{i|IMi=im_i,Sa(T_1)=a}\right]}$ is the mean of the 200 values of $F_k\left[ls_{i|IM_i=im_i,Sa(T_1)=a}\right]$ evaluated at $IMi=im_i$ for $Sa(T_1)=a$, represented in Figure 5.7.

Because $F\left[ls_{i|IM_i=im_i,Sa(T_1)=a}\right]$ is a normal random variable, it follows from Equations 5.14 and 5.15 that $P\left[ds_{i|IM_i=im_i,Sa(T_1)=a}\right]$ is also a normal random variable with mean and variance determined according to Equations 5.16 and 5.17. Consequently, the probability of being exactly in damage state ds_i for a given level of $Sa(T_1)$ (for GMF_j of Rup_n), designated $f\left[Pds_{i|Rup_n,GMF_j,GMPE_m,Sa(T_1)=a}\right]$, is an uncertain variable determined by the following mathematical formulation:

$$f\left[Pds_{i|Rup_n,GMF_j,GMPE_m,Sa(T_1)=a}\right] =$$

$$= \int_{IMi} P\left[ds_{i|IM_i=im_i,Sa(T_1)=a}\right].f\left(IM_{i|Rup_n,GMF_j,GMPE_m,Sa(T_1)=a}\right).dIM_i \qquad (5.19)$$

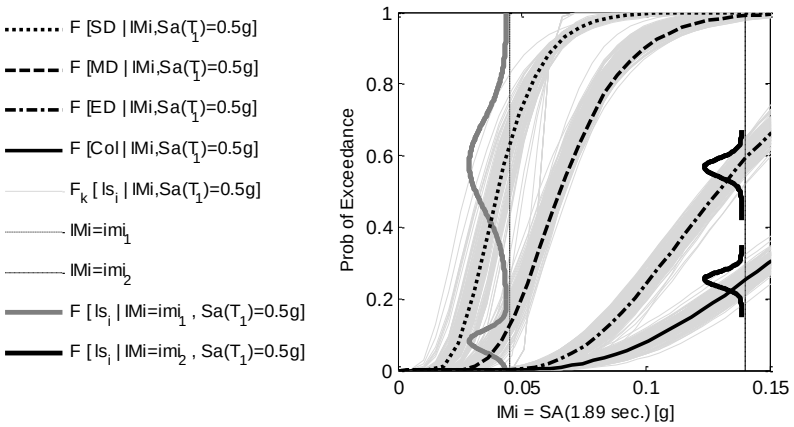


Figure 5.7 – Illustration of probabilistic normal distribution of $F\left[ls_{i|IM_i=im_i,Sa(T_1)=0.5g}\right]$ for two distinct levels of *IMi: im_{i1} and im_{i2}. Conditional Fragility Function*s of 5-story frames given $Sa(T_1)=0.5g$ and damage criteria of *GD.*

Since, as illustrated in Figure 5.7, each simulation S_i defines a given $IMi=im_i$ for each location, Equations 5.14 to 5.19 are herein solved through a numerical exercise schematically illustrated in Figure 5.8.

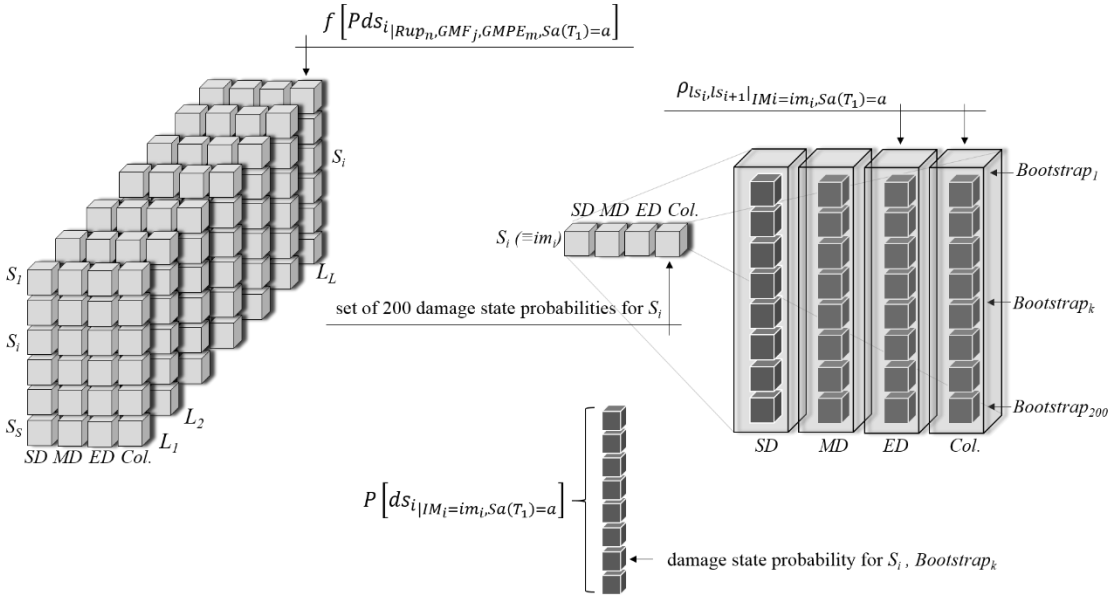


Figure 5.8 – Schematic representation of the damage exceedance matrix resulting from the evaluation of 200 bootstrapped *conditional fragility functions* at each simulated value of *IMi* (designated as im_i)

As presented in Figure 5.8, the bootstrap method presented in the previous section allows one to compute 200 sets of four damage state probabilities that reflect $\rho_{ls_i,ls_{i+1}|IMi=im_i,Sa(T_1)=a}$ and characterize $P\left[ds_{i|IM_i=im_i,Sa(T_1)=a}\right]$ for each S_i, in each location. Therefore, the aforementioned procedure provides a *3D* matrix of [*S* simulations of *IMi* x *200* bootstrap samples] x [4 damage states] x [L_L locations], which defines the distribution of $f\left[Pds_{i|Rup_n,GMF_j,GMPE_m,Sa(T_1)=a}\right]$ (for damage states of *SD*, *MD*, *ED* and *Col.*) at all the locations in which $Sa(T_1)=a$.

5.3.2.3 Intensity-specific distributions of damage ratio and its spatial correlation

Having defined $f\left[Pds_{i|Rup_n,GMF_j,GMPE_m,Sa(T_1)=a}\right]$ for each damage state and each location (Figure 5.8), an appropriate *consequence (or damage-to-loss) function* can be used in order to translate damage state probabilities into a correspondent *damage ratio*. As illustrated in Figure 5.9, each set of 4 damage state probabilities (damage states of *SD*,

MD, *ED* and *Col*) is referred as $S_{i,k}$, which stands for the *bootstrap k* performed for the simulation S_i.

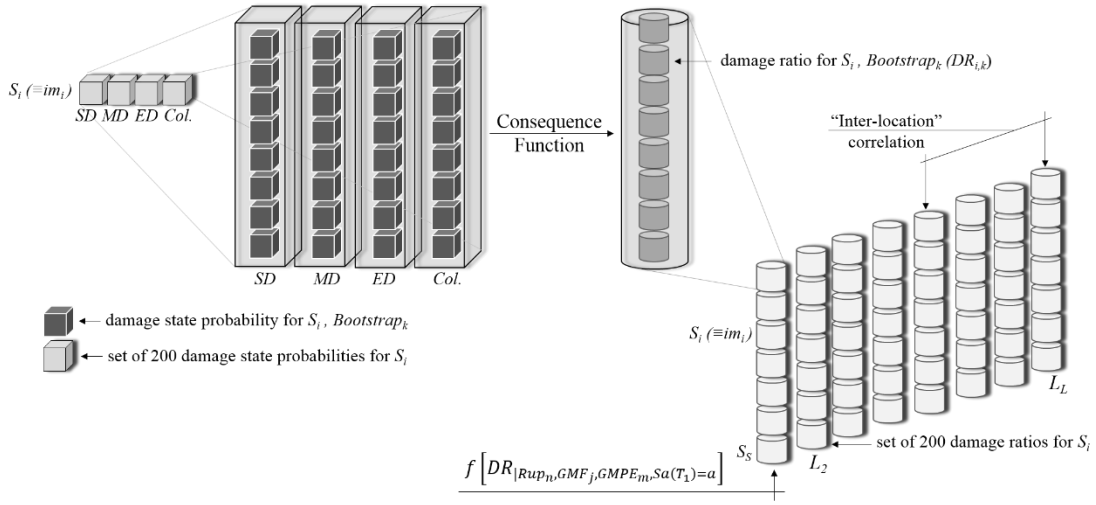


Figure 5.9 – Schematic Schematic representation of the damage matrix resulting from the application of a consequence model to the damage exceedance matrix schematically presented in Figure 5.8.

The resulting *damage ratio* is herein referred as $DR_{i,k}$, and, for each location, the set of *S* x *200 damage ratios* defines $f\left[DR_{|Rup_n,GMF_j,GMPE_m,Sa(T_1)=a}\right]$, i.e. the distribution of damage ratio given $Sa(T_1)=a$, GMF_j, and Rup_n. As a result, the distributions $f\left[DR_{|Rup_n,GMF_j,GMPE_m,Sa(T_1)=a}\right]$ are fully consistent with $f\left(IM_{i_{|Rup_n,GMF_j,GMPE_m,Sa(T_1)=a}}\right)$ at each location (Equation 5.19), and the spatial correlation between damage ratios at different sites (referred as "inter-location" correlation in Figure 5.9) is also taken into account, as the sample correlation between sets of $DR_{i,k}$ values at any two sites of interest.

In this exercise, the consequence model considered by Silva et al. (2015c) in the derivation of a vulnerability model for reinforced concrete buildings in Portugal has been considered. This model provides median (damage ratio) values of 0.10, 0.30, 0.60 and 1.0 for limit states of *SD, MD, ED* and *Col.*, respectively, as well as associated uncertainty. However, for the purpose of this exercise, only median values have been adopted.

5.3.2.4 Distribution of damage ratios and probability of exceedance of loss values

The damage matrix presented in Figure 5.9 has dimensions of [*S* simulations of *IMi* x 200 bootstrap samples] x [L_L locations], and can further be transformed into a vector of [*S* simulations of *IMi* x 200 bootstrap samples], as shown in Figure 5.10.

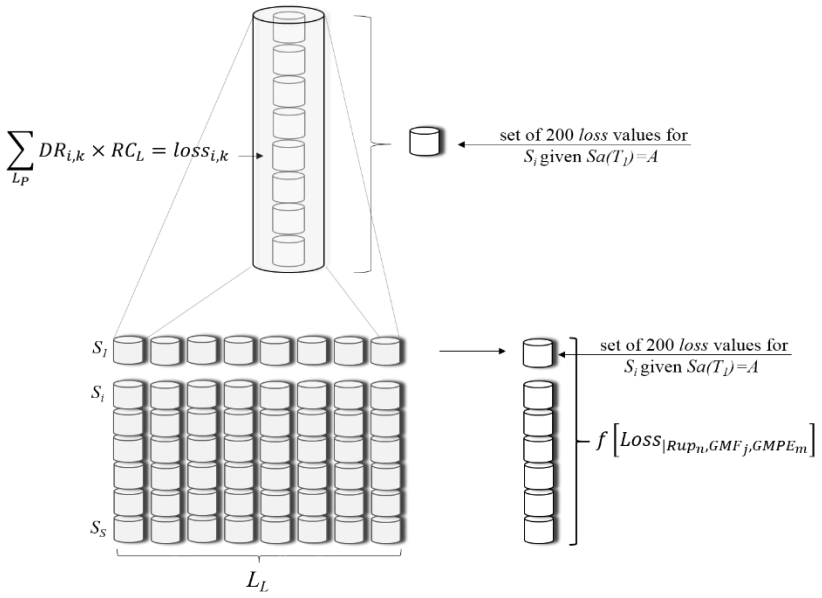


Figure 5.10 – Schematic representation of computation of $f\left[Loss_{|Rup_n,GMF_j,GMPE_m}\right]$ for a portfolio of buildings of a given *Class*, distributed across L_L sites where $Sa(T_1)=a$.

Each element of the resulting vector reflects the sum of $DR_{i,k} * RC_L$ across all the L_L locations $\left(\sum_{L_L} DR_{i,k} \times RC_L\right)$, where $DR_{i,k}$ is *damage ratio* obtained for S_i,$bootstrap_k$, and RC_L is the total replacement value (or cost) of the assets of the construction *Class* of interest (RC), at each location *L*.

This exercise results in [*S* x 200] $loss_{i,k}$ values that represent the probabilistic distribution of *loss* given GMF_j and Rup_n, using $GMPE_m$ (i.e. $f\left[Loss_{|Rup_n,GMF_j,GMPE_m}\right]$). Thus, $P\left(L > l|_{Rup_n,GMF_j,GMPE_m}\right)$ is obtained from $f\left[Loss_{|Rup_n,GMF_j,GMPE_m}\right]$, as the number of occurrences in which the observed loss, *L*, exceeds a certain value, *l*, divided by the length of the aforementioned vector (i.e. *S* x 200).

For the sake of clarity, it shall be noted that, for a given GMF_j, $f\left[Loss_{|Rup_n,GMF_j,GMPE_m}\right]$ results from the aggregation of loss values from all the assets in all the locations of the spatial distribution of $Sa(T_1)$ (denoted as *A*, with length L_p). In fact, *A* can take any possible value of $Sa(T_1)$. However, for simplicity, the computation of $f\left[Loss_{|Rup_n,GMF_j,GMPE_m}\right]$ presented in this section and schematically exemplified in Figure 5.10 takes into account only the hypothetical L_L sites where $Sa(T_1)$ is equal to a given value *a*.

5.3.3 Resulting vulnerability model

As demonstrated in section 5.3.2.3 , $f\left[DR_{|Rup_n,GMF_j,GMPE_m,Sa(T_1)=a}\right]$ is computed for each of the *N x J* simulations (for each site). Therefore, any site-specific $f\left[DR_{|GMPE_m,Sa(T_1)=a}\right]$ at a given location can be obtained according to the following Equation:

$$f\left[DR_{|GMPE_m,Sa(T_1)=a}\right] =$$

$$= \sum_{n=1}^{N}\sum_{j=1}^{J} f\left[DR_{|Rup_n,GMF_j,GMPE_m,Sa(T_1)=a}\right].P\left(Rup_n, GMF_j|_{GMPE_m,Sa(T_1)=a}\right) \quad (5.20)$$

In which $P\left(Rup_n, GMF_j|_{GMPE_m,Sa(T_1)=a}\right)$ is given by the following Equation 5.21:

$$P\left(Rup_n, GMF_j|_{GMPE_m,Sa(T_1)=a}\right) = \frac{\gamma_n}{\gamma_{Sa(T_1)=a}}\frac{1}{J} \quad (5.21)$$

For illustration purposes, Figure 5.11 presents the Vulnerability model of two, five and eight-story buildings, for the site of Lisbon, Portugal (Latitude = 38.373, Longitude = –9.143), as defined by the numerical solution of Equation 5.20. In this case, *N* is defined in the aforementioned earthquake rupture forecast (*ERF)*, *J*=*S*=2000, and GMPE is that of Atkinson & Boore (2006).

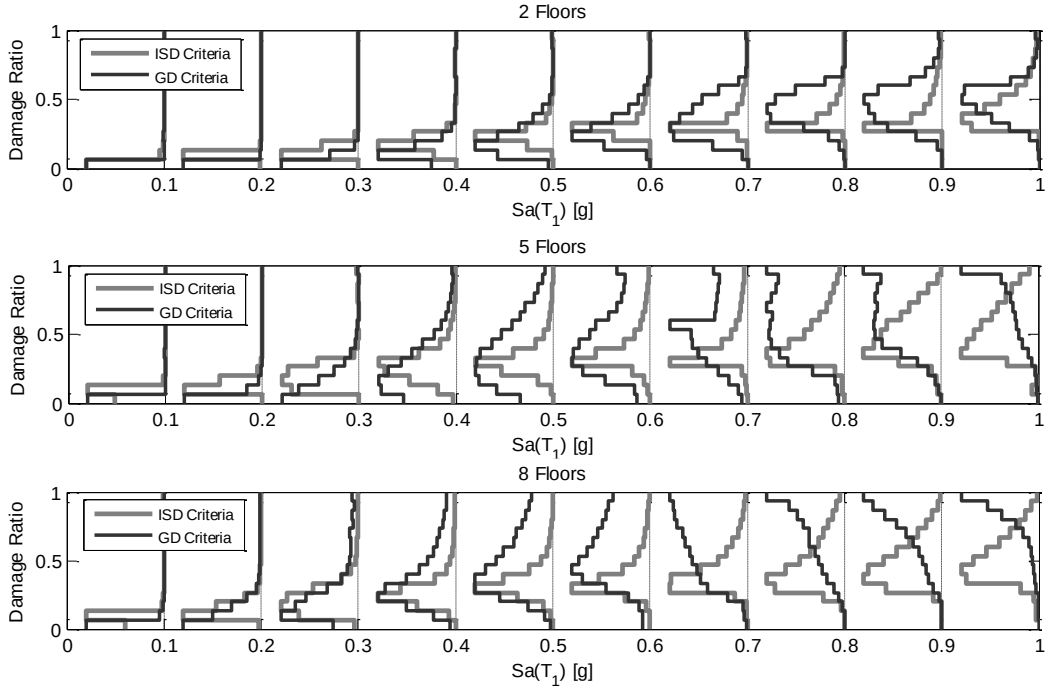


Figure 5.11 – Schematic Vulnerability Model of 2, 5 and 8-story buildings characterized by intensity-specific distributions of damage ratio (for the site of Lisbon, Portugal), using the *GMPE* of Atkinson & Boore (2006).

It is evident in Figure 5.11 that the appraised distributions of damage exceedance probability differ significantly depending on the parameter used to characterize building response and corresponding limit states. As $Sa(T_1)$ increases, distributions obtained using *GD* criteria tend to depart from the ones corresponding to *ISD*, leading to empirical distributions that are skewed towards higher damage ratios. In other words, because limit states are defined differently in the case of *GD* and *ISD* criteria, it is verified that, for a given frame subjected to a certain ground motion record, *GD* tend to lead to overestimated damage exceedance probabilities, with respect to *ISD*. This trend is more evident for higher values of $Sa(T_1)$, and also more pronounced as the level of corresponding conditional *IMi* increases.

5.4 Loss estimation exercise

In this section, the loss estimation framework outlined in chapter 5.3 is separately applied to three different building portfolios, as described below.

5.4.1 Test-bed building portfolios

Using the data from the Portuguese Building Census survey of 2011 (INE, 2015), three building portfolios referring to two, five, and eight-story reinforced concrete pre-code buildings in the district of Lisbon have been considered. The referred survey provides detailed estimation of number of buildings of each class on a parish-level resolution. However, the aggregation of the elements at a single location per parish can introduce a significant error for the larger regions with a very unbalanced spatial distribution of the building stock and seismic hazard, as the ground motion at the area centroid might be significantly different from that at the actual location of the assets (Silva, Crowley, Pinho, & Varum, 2015b). Thus, the GEOSTAT (2011) population distribution dataset, which provides the population count in a grid of 1 km^2 resolution, has been used to distribute the number of buildings in each parish proportionally to the amount of population estimated at each grid cell, as illustrated in Figure 5.12.

Similarly to what has been performed by Silva et al. (2015b), the information provided by the latter has been used to determine the economic value of each building class as the product of: a) the respective average number of dwellings per class (which is a function of the number of floors), b) the average area per dwelling determined for the interested region, in m^2, and c) the average unit cost of replacement in *EUR*/m^2.

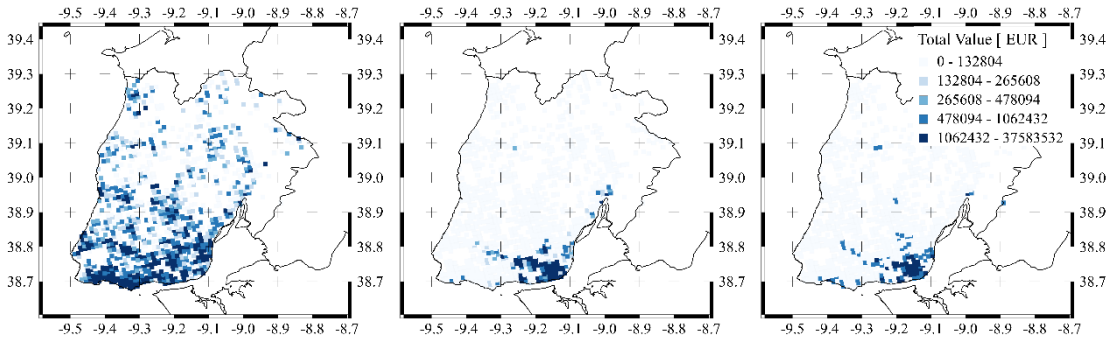


Figure 5.12 – Total replacement value (EUR) of two (upper left), five (upper right) and eight-story (bottom) reinforced concrete pre-code buildings located in the district of Lisbon, Portugal. Spatial resolution of 1km^2.

5.4.2 Fragility models and loss estimation assumptions

State-of-the-art loss estimation frameworks allow the user to account for spatial correlation of ground motion residuals (e.g. Silva et al. (2014)) when simulating random fields of ground motion. However, even in such cases, several limitations are recognized with respect to the way uncertainty in the vulnerability is addressed when using analytically derived models:

1. The correlation of uncertainty is usually incorporated such that when sampling the uncertainty in the vulnerability of two assets with the same building class, the correlation of its residuals does not reflect the physical phenomenon influencing the spatial variability (and correlation) of building response. Usually, boundary conditions inherent to zero and full spatial correlation are applied;
2. The propagation of uncertainty from fragility to vulnerability is commonly related to the scatter to which a parametric (usually cumulative lognormal) curve is fitted to the intensity dependent damage exceedance probabilities (e.g. Bradley (2010b) and Silva et al. (2013)). Therefore, as highlighted by Taylor (2015) the resulting vulnerability models commonly feature convenient but unrealistic simplifications such as beta or lognormal probabilistic distributions to model uncertainty in intensity-dependent *damage ratios*, which in fact has no evident (or demonstrated to the Author's knowledge) physical connection with reality.

In order to overcome the aforementioned limitations, the present loss estimation exercise is performed based on the methodology presented in sections 5.3.1 and 5.3.2, according to which *conditional fragility functions* and conditional random fields of

$IMi|_{Sa(T_1)=a}$ are used when modelling uncertainty and spatial correlation of damage state probabilities and corresponding damage ratios (further designated as fragility model a)).

A second fragility model (further referred as model b) is also considered. In this case, instead of representing fragility for a given level of *Sa(T₁)* as a distribution of damage exceedance probabilities, a single value (denoted as $\tilde{P}ls_{i|Sa(T_1)=a}$) is computed, as commonly considered in the literature. As a result, $\tilde{P}ls_{i|Sa(T_1)=a}$ is obtained from the distribution of 6000 *EDP* values (60 ground motion records *x* 100 frames) for each level of *Sa(T₁),* as presented in Chapter 4. The latter consists of a state-of-the art methodology (e.g. Bradley (2010b) , Silva et al. (2013)) in which the propagation of uncertainty from fragility to vulnerability for a specific *Sa(T₁)=a* is performed through a bootstrap method where 200 (cumulative lognormal) fragility functions are fitted to equal number of synthetic datasets obtained by random sampling with replacement from the original set of $\tilde{P}ls_{i|Sa(T_1)=a}$, as illustrated in Figure 5.13 (dashed black lines).

In the case of model b, the simulation of conditional random fields of $IMi|_{Sa(T_1)=a}$ is not necessary, since $\tilde{P}ls_{i|Sa(T_1)=a}$ is dependent only on values of *Sa(T₁)* obtained in each GMF_j. However, when sampling the exceedance probabilities of two assets with the same building class and level of *Sa(T₁)* from such fragility model, there is no evident physical meaning in the correlation of its residuals (as opposed to the case of fragility model a). Therefore, the aforementioned commonly used boundary conditions are considered in the simulation of the uncertainty of damage exceedance probabilities given *Sa(T₁)=a*, which, as demonstrated by Bradley (2010b) and Silva et al. (2013), can be assumed to follow a normal distribution:

b.1 - Full spatial correlation of residuals;

b.2 - Zero spatial correlation of residuals.

For the purpose of loss estimation, the fragility models a and b have been extended to levels of *Sa(T₁)* corresponding to an *epsilon* value (ε_{max}) selected in order to neglect ground motion values that correspond to (seismic intensity) probabilities of exceedance, *POE* (conditional on a given rupture), that are thought to be too low to matter in the computation of hazard. Thus, following the proposals of Strasser et al. (2008), a conditional *POE* of 10^{-5} and corresponding $\varepsilon_{max} \simeq 4.0$ are herein considered.

Distributions of damage exceedance probability, corresponding *conditional fragility functions* and $\tilde{P}ls_{i|Sa(T_1)=a}$ values are thus additionally determined with intervals

of *0.2g* for levels of $Sa(T_1)$ higher than *1.0g*, for each structural class, according to the methodology presented in Chapter 4. For the sake of illustration, Figure 5.13 presents the record-specific probabilities of exceedance of *Extensive Damage* for $Sa(T_1)=0.1g$ to *5.0g* (5-story buildings) and corresponding *conditional fragility curves* for the specific cases of $Sa(T_1)$=0.5g, 1.2g and 2.0g (i.e. fragility model a). For comparison purposes, the values of $\tilde{P}ls_{i|Sa(T_1)=a}$ and associated uncertainty determined by 200 bootstrapped lognormal fragility curves (i.e. fragility model b) are also illustrated.

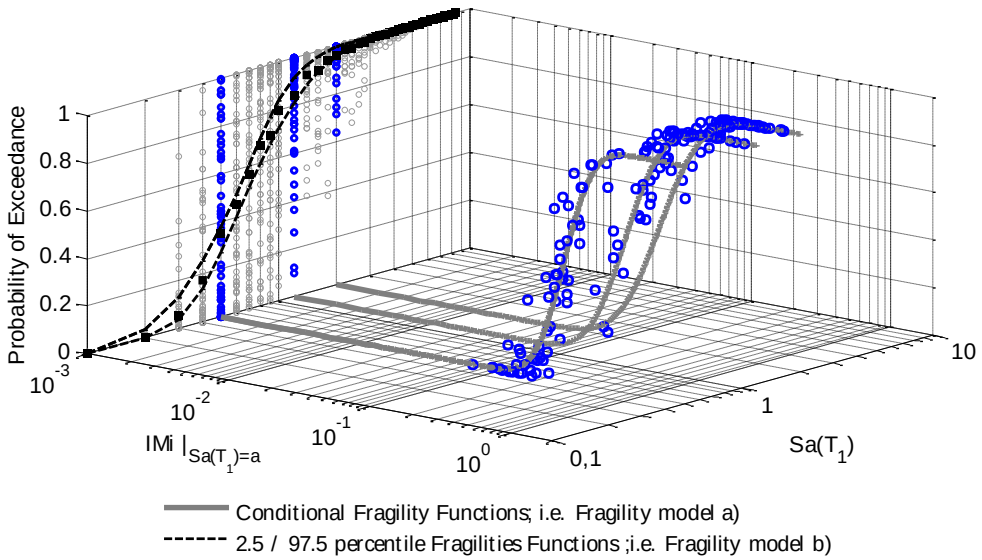


Figure 5.13 – Record-specific probabilities of exceedance of *ED*, as a function of *GD* criteria, and corresponding *conditional fragility functions* for the cases of $Sa(T_1)$=0.5g, 1.2g and 2.0g, as well as the values of $\tilde{P}ls_{i|Sa(T_1)=a}$ (black squares) and associated uncertainty determined by 200 bootstrapped lognormal fragility curves, for 5-story buildings.

In addition, it shall be highlighted that *J=S=2000* (see sections 5.3.1 and 5.3.2) have been verified to be a number of simulations sufficient for obtaining numerically accurate results.

5.4.3 Discussion of results

This section presents the results of the aforementioned loss estimation exercise, whereby annual loss exceedance curves are computed for the portfolios presented in section 5.4.1, using fragility models a and b, derived in terms of *GD* and *ISD* criteria.

As illustrated in Figure 5.14, the loss estimation methodology introduced in this chapter– based on model a - systematically provides lower estimates of annual rate of exceedance for lower (i.e. more frequent) loss values, whereas the opposite trend is verified for higher aggregated losses, in comparison with model b. Despite the fact that

model b represents a state-of-the-art methodology of fragility and vulnerability assessment, it is the Author's opinion that the disparities between the model a and b highlights the strengths (and further justifies the use) of the methodology presented in this chapter.

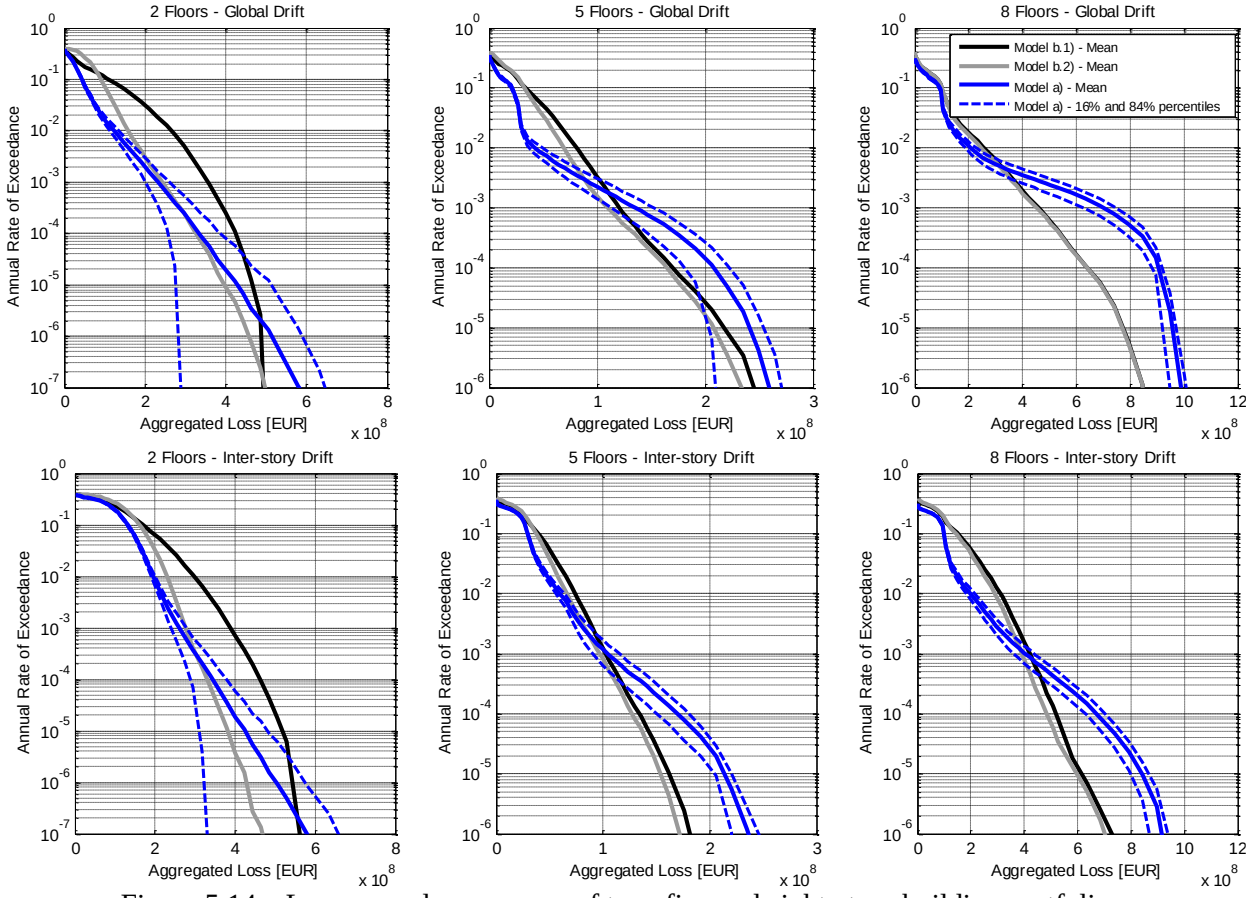


Figure 5.14 – Loss exceedance curves of two, five and eight-story building portfolios, determined using fragility models a and b, derived based GD (upper) and *ISD* criteria (lower).

As illustrated in Figure 5.15, $\int_{\tilde{P}ls_{i|Sa(T_1)=a}}^{1.0} f\left[Pls_{i|Sa(T_1)=a}\right] dp$ (where p ranges from 0 to 1.0) is lower than 0.5 for lower levels of $Sa(T_1)$, being higher than 0.5 for higher levels of $Sa(T_1)$. In other words, for lower levels of $Sa(T_1)$, which are associated with more frequent aggregated loss values, $\tilde{P}ls_{i|Sa(T_1)=a}$ is higher than more than 50% of the conditional damage exceedance probabilities defined by $f\left[Pls_{i|Sa(T_1)=a}\right]$, whereas the opposite trend is verified for higher levels of $Sa(T_1)$ (i.e $\tilde{P}ls_{i|Sa(T_1)=a}$ is lower than more than 50% of conditional damage exceedance probabilities defined by $f\left[Pls_{i|Sa(T_1)=a}\right]$. As a result, $\tilde{P}ls_{i|Sa(T_1)=a}$ systematically overestimates the conditional damage probabilities for lower and more frequent levels of $Sa(T_1)$, and vice-versa for higher

values of *Sa(T1)* (and loss). This is consistent with the results illustrated in Figure 5.14 and highlights the misrepresentation of record-to-record variability associated with model b, and the strength of the proposed framework in that respect.

In addition, it is verified that the variability of loss exceedance rates (depicted in Figure 5.14 by the 16% and 84% percentiles of $\gamma(L > l)$, only for the case of model a for the sake of visual clarity), increases with the size and area of spatial distribution of the building portfolio (see Figure 5.12). The latter is expected and consistent with the theoretical formulation of the variability of the sum of correlated random variables.

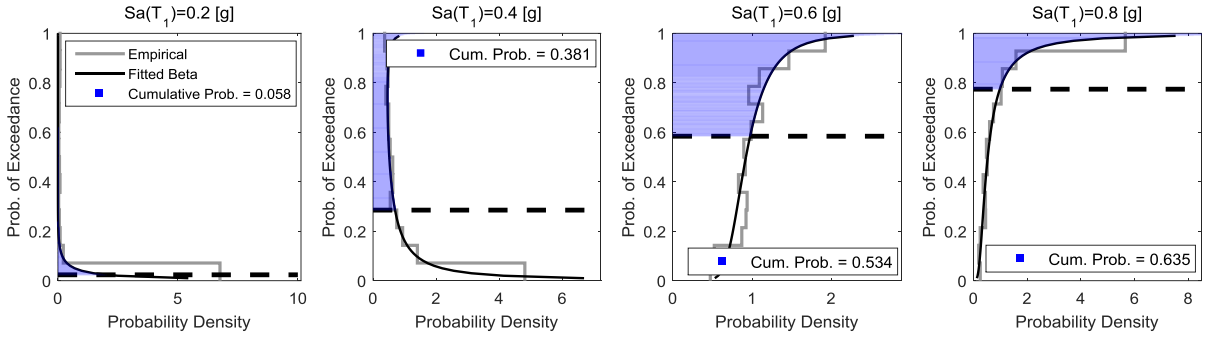


Figure 5.15 – Representation of $\int_{\tilde{P}ls_{i|Sa(T_1)=a}}^{1.0} f\left[Pls_{i|Sa(T_1)=a}\right] dp$ (area in blue, referred as 'cumulative probability', for simplicity), and corresponding values of $\tilde{P}ls_{i|Sa(T_1)=a}$ (black dashed lines). Damage state of *ED* and *GD* criteria, for 5-story frames. "Empirical" and "Fitted Beta" refer, respectively, to the empirical distribution $f[Pls_{i|Sa(T_1)=a}]$ and the corresponding fitted Beta model, as proposed in section 4.4.4.

5.4.4 Final remarks

The research efforts presented in this chapter are built upon the findings described in Chapter 4, extending the methodology to the derivation of vulnerability functions that reflect a non-parametric (site-specific) histogram of *damage ratio* per level of primary intensity measure - *Sa(T1).* This framework provides the link between vulnerability uncertainty and seismic hazard, such that when sampling the uncertainty in the vulnerability of two assets with the same building class and level of *Sa(T1),* the correlation of its residuals is physically explained by the spatial distribution (and correlation) of a conditional intensity measure, designated as *IMi*.

Based on the aforementioned methodology, the simulation of conditional spatial cross-correlated random fields of $IMi_{|Sa(T_1)=a}$ has been included in a novel loss estimation methodology. The latter allows the explicit modelling of spatial correlation of the aforementioned conditional non-parametric histograms of *damage ratio,* as determined by the spatial distribution of $IMi_{|Sa(T_1)=a}$, providing a framework whereby

vulnerability uncertainty can appropriately be taken into account in the context of portfolio loss estimation. Furthermore, this framework uses *conditional fragility functions* introduced in Chapter 4, based on the verified *sufficiency* of $IMi_{|Sa(T_1)=a}$, to provide vulnerability and loss estimations that are fully consistent with the hazard properties at each point of the interested portfolio.

In addition, it has been demonstrated that the proposed loss assessment framework leads to results that differ with respect to state-of-the-art vulnerability assessment methods in a consistent manner across different building classes and damage state definition criteria. More specifically, it has been verified that state-of-the-art methodologies tend to overestimate the annual rate of exceedance of lower (i.e. more frequent) loss values, with respect to the framework presented in this work, whereas the opposite trend is verified for higher aggregated losses. The latter is shown to be related with the presented methodology's more robust representation of the impact of record-to-record variability in loss estimations, highlighting its strengths and contribution to the improvement of fragility, vulnerability and loss assessment of building portfolios.

Appendix 5.1 – Derivation of $u_{IM^*|_{Sa(T_1)=A}}$ and $\Sigma_{IM^*|_{Sa(T_1)=A}}$

As presented in 5.3.2.1, a given *IMi* selected for a particular level of *Sa(T1)* is generically referred as $IM_{i|Sa(T_1)=a}$ (with *i* ranging from 1 to N^*), and IM^* is the vector that incorporates the set of N^* *efficient* and *sufficient IMi* selected for the N^*levels of *Sa(T1)* of interest. In this context, the mean and spatial covariance matrix of $IM^*_{|Sa(T_1)=A}$ are designated $u_{IM^*|_{Sa(T_1)=A}}$ and $\Sigma_{IM^*|_{Sa(T_1)=A}}$, respectively.

If one considers *Sa(T1)* and *IMi* as the random vectors of corresponding intensity measures at a number of locations, L_p (in which L_p is the length of *A*), it follows that:

$$u = \begin{bmatrix} u_{Sa(T_1)} \\ u_{IM_1} \\ \vdots \\ u_{IM_i} \\ \vdots \\ u_{IM_{N^*}} \end{bmatrix} \text{ with size } \begin{bmatrix} L_p x\ 1 \\ L_p x\ 1 \\ \vdots \\ L_p x\ 1 \\ \vdots \\ L_p x\ 1 \end{bmatrix} \tag{A. 5.1}$$

$$\Sigma = \begin{bmatrix} \Sigma_{Sa(T_1),Sa(T_1)} & \Sigma_{Sa(T_1),IM_1} \dots & \Sigma_{Sa(T_1),IM_i} \dots & \Sigma_{Sa(T_1),IM_{N^*}} \\ \Sigma_{IM_1 Sa(T_1)} & \Sigma_{IM_1,IM_1} \ \dots & \Sigma_{IM_1,IM_i} \ \dots & \Sigma_{IM_1,IM_{N^*}} \\ \vdots & \vdots \ \vdots & \vdots \ \vdots & \vdots \\ \Sigma_{IM_i Sa(T_1)} & \Sigma_{IM_i,IM_1} \ \cdots & \Sigma_{IM_i,IM_i} \ \cdots & \Sigma_{IM_i IM_{N^*}} \\ \vdots & \vdots \ \vdots & \vdots \ \vdots & \vdots \\ \Sigma_{IM_{N^*},Sa(T_1)} & \Sigma_{IM_{N^*},IM_1} \ \cdots & \Sigma_{IM_{N^*},IM_i} \ \cdots & \Sigma_{IM_{N^*},IM_{N^*}} \end{bmatrix} \text{ with}$$

$$\text{size } \begin{bmatrix} \mathrm{L_p x\ L_p} & \cdots & \mathrm{L_p x\ L_p} \\ \vdots & \ddots & \vdots \\ \mathrm{L_p x\ L_p} & \dots & \mathrm{L_p x\ L_p} \end{bmatrix} \tag{A. 5.2}$$

In which $u_{Sa(T_1)}$ and u_{IM_i} are the vectors of mean predictions of *Sa(T1)* and *IMi* obtained by the selected Ground Motion Prediction Equation (*GMPEm*) for each of the L_p locations, for Rup_n, and $\Sigma_{Sa(T_1),Sa(T_1)}$, $\Sigma_{Sa(T_1),IM_i}$and $\Sigma_{IM_i,IM_i}$are respectively: the spatial auto-covariance matrix of *Sa(T1)*, the spatial cross-covariance matrix of *IMi* and *Sa(T1)*, and the spatial auto-covariance matrix of *IMi*.

For the purpose of building the aforementioned spatial auto and cross-covariance matrixes, the spatial cross-correlation between fields of ground motion at different spectral periods is obtained via the linear model of coregionalization (*LMCR*) (see

5.3.2.1). More specifically, each element of the matrix Σ presented in Equation A.5.2, denoted herein as $COV_{IM_1{}^{l_a},IM_2{}^{l_b}}$, is determined as follows:

$$COV_{IM_1{}^{l_a},IM_2{}^{l_b}} = \rho_{IM_1{}^{l_a},IM_2{}^{l_b}}.\sigma_{IM_1{}^{l_a}}.\sigma_{IM_2{}^{l_b}} \quad \text{(A. 5.3)}$$

Where $\rho_{IM_1{}^{l_a},IM_2{}^{l_b}}$ is the spatial cross-correlation coefficient between IM_1 at location l_a and IM_2 at location l_b, defined by the *LMCR*, $\sigma_{IM_1{}^{l_a}}$ and $\sigma_{IM_2{}^{l_b}}$ are the predictions of logarithmic standard deviation of IM_1 at location l_a and IM_2 in location l_b, respectively, as determined by the selected $GMPE_m$, for Rup_n. In this context, IM_1 and IM_2 represent any two intensity measures of interest.

It is possible to further partition u and Σ, as follows:

$$u = \begin{bmatrix} u_{Sa(T_1)} \\ u_{IM^*} \end{bmatrix} \text{ with size } \begin{bmatrix} L_p x\ 1 \\ (N^* x L_p) x\ 1 \end{bmatrix} \quad \text{(A. 5.4)}$$

$$\Sigma = \begin{bmatrix} \Sigma_{Sa(T_1),Sa(T_1)} & \Sigma_{Sa(T_1),IM^*} \\ \Sigma_{IM^*,Sa(T_1)} & \Sigma_{IM^*,IM^*} \end{bmatrix} \text{ with}$$

$$\text{size } \begin{bmatrix} L_p x\ L_p & L_p x (N^* x L_p) \\ (N^* x L_p) x\ L_p & (N^* x L_p) x (N^* x L_p) \end{bmatrix} \quad \text{(A. 5.5)}$$

Where $u_{Sa(T_1)}$ and u_{IM^*} are, respectively, the mean vectors of $Sa(T_1)$ and the N^* conditional intensity measures considered at all the L_p locations, and $\Sigma_{Sa(T_1),Sa(T_1)}$, $\Sigma_{IM^*,Sa(T_1)}$ and Σ_{IM^*,IM^*} are the partitions of the Σ matrix defined in accordance with the partition of the mean vector. Consequently, the mean vector and spatial cross-covariance matrix of IM^* conditioned on *A*, in which *A* is the vector of simulated values of $Sa(T_1)$ in each of the L_p locations, for GMF_j generated for Rup_n, are determined as follows:

$$u_{IM^*|_{Sa(T_1)=A}} = u_{IM^*} + \Sigma_{IM^*,Sa(T_1)}.\Sigma_{Sa(T_1),Sa(T_1)}{}^{-1}\left(A - u_{Sa(T_1)}\right) \quad \text{(A. 5.6)}$$

$$\Sigma_{IM^*|_{Sa(T_1)=A}} = \Sigma_{IM^*,IM^*} - \Sigma_{IM^*,Sa(T_1)}.\Sigma_{Sa(T_1),Sa(T_1)}{}^{-1}.\Sigma_{Sa(T_1),IM^*} \quad \text{(A. 5.7)}$$

Chapter 6 SEISMIC HAZARD CONSISTENCY AND EPISTEMIC UNCERTAINTY IN FRAGILITY MODELLING AND PORTFOLIO LOSS ESTIMATION

This chapter is based on the following reference:

Sousa, L.; Marques, M.; Silva, V.; Weatherill, G. (2017). Seismic hazard consistency and epistemic uncertainty in fragility modelling and portfolio loss estimation. *Earthquake Engineering and Structural Dynamics* (under review)

Summary

State-of-the-art methodologies for the seismic risk assessment of building portfolios have benefitted from the continued improvement in the characterization of seismic hazard, especially with respect to the modelling of the associated epistemic uncertainties. Logic-trees have become a standard feature of probabilistic seismic hazard and risk assessments, reflecting distinct hazard modelling options. However, the development of fragility and loss models that are able to adequately reflect these uncertainties in a consistent way has been subject of limited scrutiny. In this research, the subject of 'hazard-dependency' of fragility is addressed, through a methodology that, for each logic-tree branch, incorporates: probabilistic seismic hazard analysis, hazard-compatible record-selection, nonlinear response history analysis, and fragility assessment of three different building classes located at the sites of Lisbon and Faro (Portugal). Three distinct seismological models and twenty combinations of ground motion prediction equations are used, in order to evaluate if fragility functions are in fact dependent on the hazard properties assumed for each branch, when analytical methodologies are used to characterize both hazard and fragility components. Furthermore, the impact of considering a single fragility model common to all the branches, as opposite to a distinct 'hazard-specific' fragility per branch, is investigated in the context of a probabilistic loss estimation exercise. This leads to the proposal of an innovative fragility assessment / loss estimation framework which, based on the concept of conditional fragility functions presented in Chapter 4 and further developed in Chapter 5, is able to ensure the hazard-consistency of the fragility results of each logic-tree branch, while avoiding a time-consuming analytical fragility assessment for the entire logic-tree.

6.1 Introduction

A meaningful evaluation of the seismic risk of either single structures or building portfolios requires the consideration of various sources of uncertainty. The so-called aleatory uncertainty is due to randomness, while epistemic uncertainty is related to lack of knowledge of the process being observed (Bradley, 2009). In addition to the continual improvement in the characterization of seismic hazard (and its link with structural response), recent years have seen a major swing in emphasis towards the explicit inclusion of uncertainties in the performance assessment of structural systems (Bradley, 2013). However, these have been mostly (if not entirely) related with the treatment of record-to-record variability, and/or the random nature of geometric and structural parameters (e.g. Jalayer et al. (2010), Liel et al. (2009)), in the evaluation of the seismic response of buildings. Epistemic uncertainty, on the other hand, has become a standard feature of probabilistic seismic hazard analysis (PSHA), where the goal is to determine the probability of exceeding a given level of a seismic intensity measure of interest (Bommer & Scherbaum, 2008). More specifically, logic-trees (Pagani, et al., 2014) are commonly used to represent distinct hazard modelling possibilities, leading to a set of alternative views of a process that has only one true but unknown result (Der Kiureghian & Ditlevsen, 2008).

In recent studies, record-to-record variability has been addressed in a way that is consistent with the hazard expected at the site of interest. Methods such as the Conditional Spectrum (Baker (2011), Jayaram et al. (2011)) (CS) and the General Conditional Intensity Measure (GCIM) (Bradley, 2010a) allow record selection to reflect ground motion properties determined by the local hazard, providing a direct link between seismic hazard and building response. As a result, a seismic risk assessment framework should, in theory, include a fragility model that reflects these uncertainties in a coherent way. In other words, since structural response is dependent on the set of ground motions to which the building models are subjected (Bradley, Dhakal, Cubrinovski, & MacRae, 2009b) and, in turn, ground motion selection is dependent on hazard, then an analytical fragility model shall not only be structure-specific but also "hazard-specific". In practice, this imparts that, in theory, when the epistemic uncertainty of the hazard model is quantified through a logic-tree approach, distinct record sets shall be selected for each branch (Pagani, et al., 2014) of the logic tree, leading to a distinct fragility function per branch.

The issues above are of theoretical nature and have been subjected to limited scrutiny by the scientific community. Therefore, in this research, the subject of "hazard-specific" fragility is addressed in detail, through a methodology that incorporates: a) probabilistic seismic hazard analysis (PSHA) for the sites of Lisbon and Faro (Portugal), using a logic-tree approach that reflects distinct seismological and ground motion prediction modelling options, b) hazard-compatible record selection and nonlinear response history analysis (NLRHA) for each logic-tree branch and each site, and c) fragility analysis and comparison between fragility functions obtained for each assessed hazard branch, at each site. In this framework, the present study has the aim to assess if the epistemic uncertainty of the hazard model shall appropriately be propagated into the fragility analysis, when analytical methodologies are used to characterize both hazard and fragility components. More specifically, fragility functions obtained for each branch are compared using a statistical approach, in order to investigate whether the appraised differences corroborate the assumed "hazard-dependency" of fragility.

In order to obtain a more meaningful comparison in the context of seismic risk, probabilistic loss estimation is further performed for 6 different building portfolios located in Lisbon and Faro. In this exercise, risk metrics are evaluated for each group of assets and assessed hazard branch, using two distinct fragility assessment approaches: a) a distinct model for each logic-tree branch, consistent with each distinct hazard model, and b) a state-of-the-art method in which only a single fragility model, common to all branches, is used. As further described in this chapter, the objective of this exercise is to verify the hypothesis of 'hazard-dependency' of fragility results, as well as evaluating what are the repercussions of using fragility functions that are consistent with the hazard modelling options of each branch, as opposite to a single fragility model. Finally, in light of the appraised results, a methodology for the fragility assessment of building portfolios is presented, in which the epistemic uncertainty of the hazard model can be adequately propagated into the fragility, vulnerability and loss results. This framework resorts to the concept of conditional fragility functions previously presented in Chapter 4, providing a tool that is able to provide "hazard-consistent" fragility and loss estimates, without the need for the time consuming and computationally demanding record selection / NLRHA for all the branches of the considered hazard logic-tree.

6.2 Fragility assessment methodology

The analytical methodology implemented in this research, presented further in detail, consists of: a) PSHA for Lisbon and Faro, Portugal, using a logic-tree approach in which distinct combinations of seismological and ground motion prediction models are considered in each of the branches, b) record selection and NLRHA for each branch defined in a), for each site and structural class, and c) analysis and comparison between fragility functions derived using the NLRHA results obtained in b).

6.2.1 Probabilistic seismic hazard modelling

In order to perform the probabilistic seismic hazard assessment for the sites of Lisbon and Faro (Portugal), the recent developments of the SHARE initiative (Woessner, et al., 2015) have been considered. As a result, the epistemic uncertainty associated with the definition of seismicity at the sites of interest is defined by three distinct seismological models. The first, designated herein as AS-model, is based on the definition of areal sources for which earthquake activity is evaluated individually, the second relates to a kernel-smoothed zonation-free stochastic earthquake rate model (Hiemer, et al., 2014) that considers seismicity and accumulated fault moment (SEIFA-model), and the last results from the identification of large seismogenic sources using tectonic and geophysical evidence, incorporating a fault source / background seismicity model (FSBG-model) (Haller & Basili, 2011). For the sake of synthesis, further details of these models are not presented herein. For a more comprehensive definition of its properties, readers are referred to the work of Woessner et. al. (2015) and references therein.

For consistency with the above, epistemic uncertainty in ground motion prediction has also been foreseen via the implementation of a logic-tree approach. This way, according to the SHARE methodology, sets of possible GMPEs are defined for Active Shallow Crust (ASC), Stable Continental Crust (SCC), Shield (SH), Subduction (SUB) and Volcanic (VOL) tectonic region types (TRTs). In order to define the applicable sets of GMPEs, the TRT associated with each seismogenic source in each source model has been identified, within a maximum distance of 250 km. As illustrated in Figure 6.1 for the cases of AS and FSGB models, tectonic environments of Stable Continental Crust (Lisbon), and Stable Continental and Active Shallow Crust (Faro) are applicable.

The final logic-tree structure is presented in Figure 6.2, in which all the possible combinations of source model / ground motion prediction equations are schematically

illustrated. In the case of Lisbon, a total of 15 branches (i.e. 3 source models * 5 SCC GMPEs) is applicable, whereas 60 combinations (i.e. 3 source models * 5 SCC GMPEs * 4 ASC GMPEs) exist for Faro, where both ASC and SCC seismic sources are relevant.

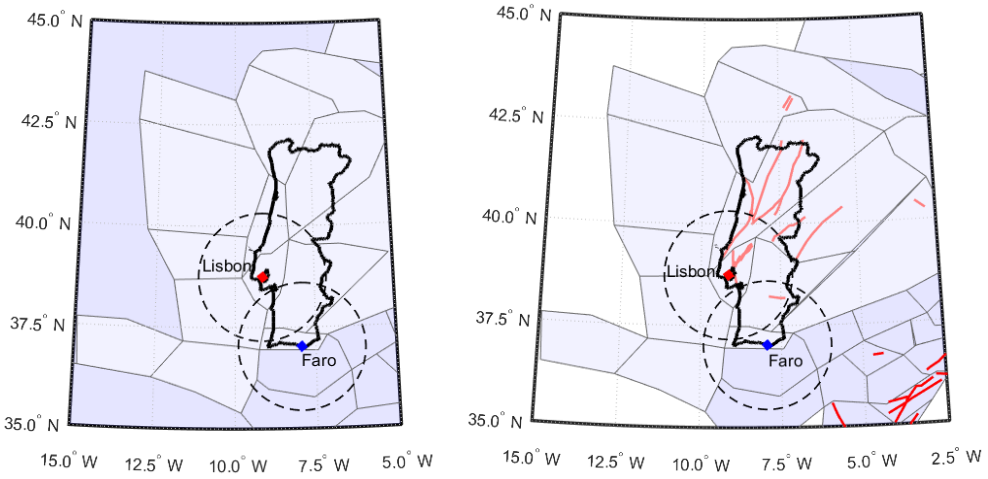


Figure 6.1 – Tectonic region environment (TRT) associated with each of the sources of the AS-model (left) and the FSBG-model (right). Stable Continental and Active Shallow sources are illustrated as: light and darker blue, respectively, for area sources, and light and darker red, respectively, for fault traces. Dashed circles represent the maximum distance of 250 km.

For clarity herein, a given branch is further denoted by an acronym that includes references to the corresponding source model, SCC GMPE and ASC GMPE, by this order. As an example, "AS-AB10-CF08" corresponds to a combination in which the AS-model is used, while the Akkar & Bommer (2010) and Cauzzi & Faccioli (2008) ground motion prediction equations are selected for SCC and ASC tectonic region environments, respectively. In the case of Lisbon, where only SCC applies, acronyms include a single GMPE instance, as in the following example: "AS-AB10".

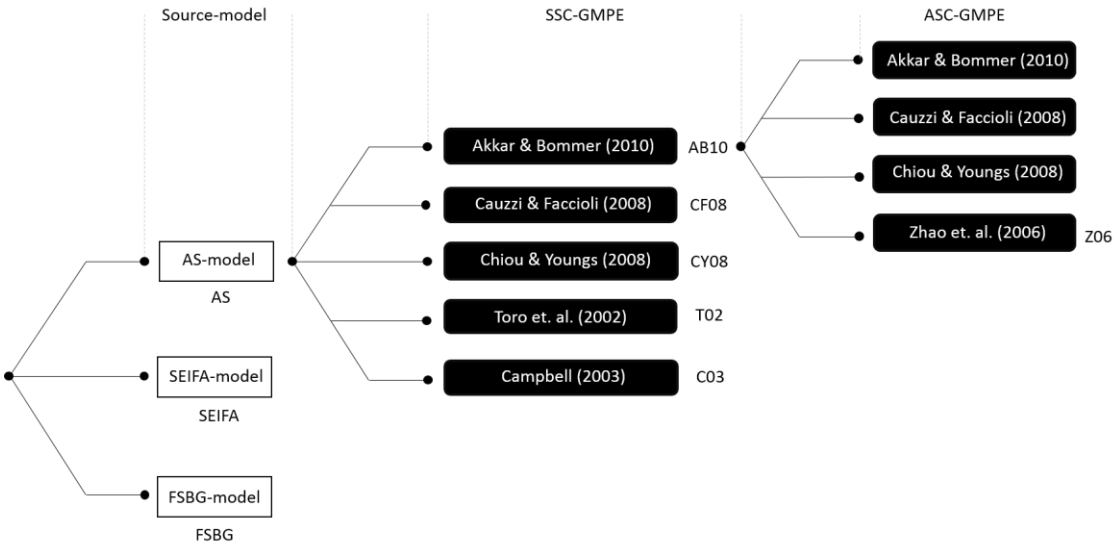


Figure 6.2 – Illustration of the source model / GMPE logic tree adopted for the sites of Faro and Lisbon. Acronyms adopted for each source model / GMPE are presented adjacently.

It shall be noted that the objective of the Author is not to specifically address the evaluation of seismic hazard in Portugal, for which several notable studies (e.g. Vilanova & Fonseca (2007), Vilanova et al. (2007)) exist. The present objective is to adopt a model that, having been scrutinized by the scientific community, provides an adequate platform for the investigation of the subject of "hazard-consistent" fragility. With this respect, the SHARE proposal has been selected amongst all the alternative hazard models proposed for Portugal and/or Europe, as it is the one for which a wider and more comprehensive range of epistemic uncertainties is foreseen.

6.2.2 Hazard-consistent record selection and disaggregation

The Generalized Conditional Intensity Measure (GCIM) approach is adopted herein for the purpose of record selection, as it allows the *predictability* of all the intensity measures verified to influence the seismic response of the assessed structures. Readers are referred to the work of Bradley (2010a) for a detailed description of the theoretical background of the methodology. However, in brief, the fundamental basis of the GCIM is that any set of ground motion parameters can be assumed to follow a multivariate lognormal distribution, and the conditional distribution given a) a rupture scenario, and b) the occurrence of a specific value of an intensity measure parameter (*IMj)*, has a univariate lognormal distribution.

Upon definition of the ground motion prediction equation of interest and the correlation between the considered intensity measures, the conditional distribution of *IMi* given *IMj=imj* is obtained via the total probability theorem as:

$$f_{IMi_{|IMj=imj}} = \sum_{i=1}^{I} f_{IMi_{|rup_i,IMj=imj}} \cdot P_{rup_{i_{|IMj=imj}}} \qquad (6.1)$$

Where *I* is the total number of ruptures in the earthquake rupture forecast (ERF) *(Pagani, et al., 2014)* $f_{IMi_{|IMj=imj}}$ is the probability density function (pdf) of *IMi* given *IMj=imj*, $f_{IMi_{|rup_i,IMj=imj}}$ is the pdf of *IMi* given *IMj=imj* and *Rup=rup*$_i$, and $P_{rup_{i_{|IMj=imj}}}$ is the contribution of *rup*$_i$ to the assessed level of seismic intensity (*im*$_j$).

Probabilistic seismic hazard analysis and disaggregation were developed in the OpenQuake engine (Pagani, et al., 2014), in accordance with the theoretical background established by McGuire (2004) and Bazzurro & Cornell (1999), respectively. In this framework, disaggregation analysis allows an investigation into the contribution of different earthquake ruptures to the probability of a certain ground motion level at the site

of interest. Given the large number of earthquake ruptures generated by the ERF for each source model, OpenQuake does not provide the contributions on a rupture-by-rupture basis. Instead, these are classified and grouped into magnitude (M) / distance (R) bins. However, given the open-source nature of the platform, it was possible to produce the necessary intermediate results for the computation of the contribution of a particular rupture (rup_i) to the occurrence of a given ground motion intensity (IM_j=im_j). The latter is herein denoted as $P_{rup_{i}|_{IM_j=im_j}}$, and defined according to Equations 6.2 and 6.3:

$$P_{rup_{i}|_{IM_j \geq im_j}} = \frac{ln\left[1-P_{IM_j \geq im_j|_{rup_i}}\right]}{ln\left[\prod_{i=1}^{I}(P_{IM_j \geq im_j|_{rup_i}})\right]} \quad (6.2)$$

$$P_{rup_{i}|_{IM_j=im_j}} \cong \frac{1}{v_{IM_j \geq imj} - v_{IM_j \geq imj+\Delta imj}} * \\ * \left[P_{rup_{i}|_{IM_j \geq im_j}} . v_{IM_j \geq imj} - P_{rup_{i}|_{IM_j \geq im_j+\Delta imj}} . v_{IM_j \geq im_j+\Delta imj}\right] \quad (6.3)$$

Where $v_{IM_j \geq imj}$ is the annual rate of exceedance of a ground motion with an intensity level of *imj*, and Δimj is a small increment of *IMj*, relative to *imj*. As highlighted by Bradley (2010a), Equation 6.3 becomes exact in the limit as $\Delta imj \to 0$. For illustration, Figure 6.3 presents the rupture-by-rupture disaggregation results for the ERF of the AS, SEIFA and FSBG source models, considering peak ground acceleration (PGA) as the *IMj*, and the site of Lisbon.

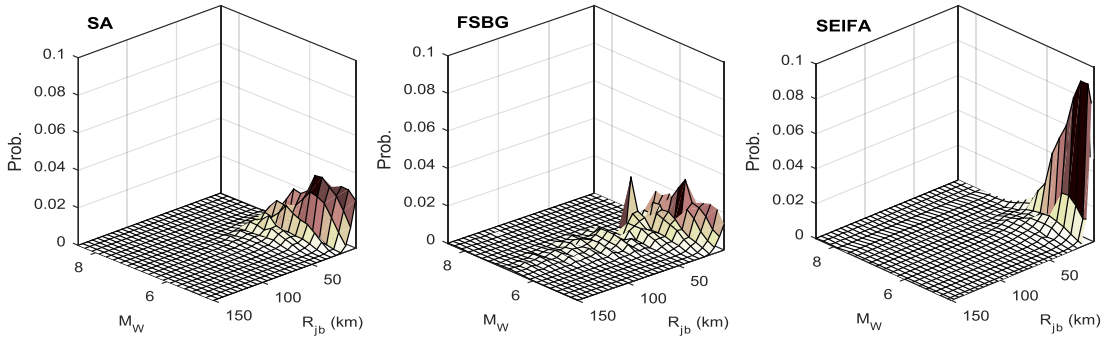


Figure 6.3 – Rupture-by-rupture disaggregation considering the ERF generated for the AS-model (left), FSBG -model (middle) and SEIFA-model (right), for *PGA=0.5g* and site of Lisbon (SCC GMPE of AB10). For visual clarity, ruptures are grouped into M / R bins of 0.2 / 5 km intervals.

6.2.3 Numerical models and record selection

Reinforced concrete construction accounts for approximately 50% of the Portuguese building stock and hosts 60% of the national population. Within this building

class, at the time of the 2011 Census Survey, 49% of the buildings had not been designed to the most recent seismic code (Silva, Crowley, Pinho, Varum, & Sousa, 2015c). Therefore, numerical models considered in this study represent typical reinforced concrete (RC) buildings with masonry infills, constructed in Portugal before 1958. According to the regulations applicable during this construction epoch, structures are designed without any seismic design provisions, and are thus defined herein as pre-code.

In agreement with the work of presented in Chapter 3, Chapter 4 and Chapter 5, statistical distributions of material and geometrical properties have been used to create synthetic portfolios of structures for different building typologies. According to Silva et al. (2015c), a number of 100 assets is necessary to guarantee the statistical significance of the generated distribution of structural capacity. For this reason, sets of 100 structures have been generated for classes of two, five and eight story buildings. In this case, dynamic properties are characterized by the mean fundamental periods of vibration extracted from the total sets of 100 assets. These have been found to be 0.26, 0.45 and 0.70 seconds, for the two, five and eight story buildings, respectively.

The percentage of reinforcement in the beams and columns is calculated following the pre-code regulations and practices corresponding to the ultimate and serviceability limit states, in accordance with the sampled geometrical and material characteristics. Moreover, to maintain the computational effort at a reasonable level, each structure is modelled as a single infilled moment frame with three bays. As schematically presented in Figure 6.4 for the case of 5 story buildings, each frame was modelled in a 2D environment using the open-source software OpenSees (McKenna, Fenves, Scott, & Jeremic, 2000), with force-based distributed plasticity beam-column elements.

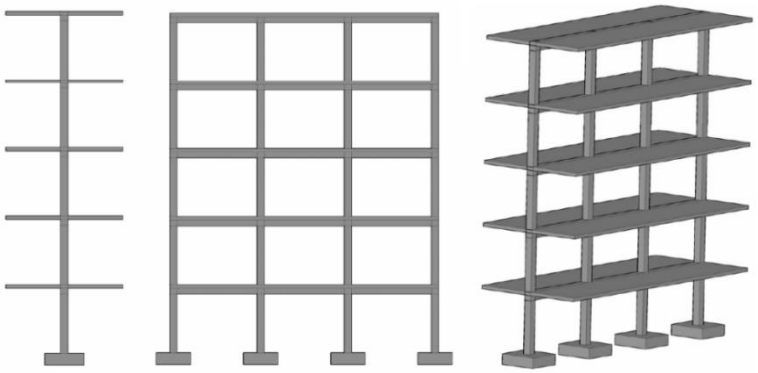


Figure 6.4 – Schematic representation of the five-story RC frame model: front (left), side (centre) and isometric view (right) with infills, adapted from Silva et al. (2015c)

For the sake of synthesis herein, readers are referred to the aforementioned work by Silva et al. (2015c) for details of the numerical considerations adopted with regards to the

cross section discretization and integration points of the elements, material constitutive relationships, P-delta effects, and the infill panel modelling approach.

6.2.3.1 'Targets' for record selection

The vector of intensity measures considered in the definition of 'targets' for ground motion selection (i.e. *IM*) includes intensity parameters (i.e. *IMi*) of peak ground acceleration (PGA), Housner intensity (HI) (Housner, 1952) and spectral ordinates within the range of 0.05 to 3.0 seconds, conditioned on *IMj* being the spectral acceleration at the mean fundamental period of vibration of each class - $Sa(T_1)$. This approach builds on the work presented in Chapter 4, where the issue of *predictability* has been foreseen in the definition of *IM*, and matters of *efficiency*, *sufficiency* and *scaling robustness* have been verified when analysing similar structural models as those used herein. As a result, a number of 60 ground motion records have been selected and scaled per level of $Sa(T_1)$ (for each site, logic tree branch and structural typology), with the latter ranging from 0.1g to 1.0g with intervals of 0.1g.

The probabilistic distribution of the selected *IM* vector conditioned on a given level of $Sa(T_1)$ is designated henceforth as $f_{IM_{|Sa(T_1)=a}}$, being determined according to the hazard-consistent probabilistic distribution of each *IMi* given *Sa(T1)=a,* as established in Equation 6.1, and the correlation models presented in Table 4.1. For details regarding the database of natural ground motion records used, readers are referred to section 4.3.2.

In order to illustrate the differences between 'target' distributions computed for different branches of the logic-tree, Figure 6.5 presents the comparison between probabilistic distributions of spectral ordinates between periods of 0.05 and 3.0 seconds. In this figure, distinction is made only between targets computed for different source models, despite the fact that each individual source model / GMPE combination is shown. The objective is to highlight the significant differences in the mean and variance of 'target' distributions computed for different logic-tree branches, rather than evaluating each of those, individually.

A possible way to assess how the illustrated differences in 'targets' for ground motion selection propagate to discrepancies in fragility results is to evaluate what is the degree of 'similarity' between distributions of *IMs* in the resulting record sets. Therefore, in order to quantitatively compare conditional 'target' distributions of IMs, a statistical approach was implemented. For a given level of $Sa(T_1)$ and structural typology (and a

certain logic-tree branch) one is able to evaluate what is the empirical distribution of each *IMi*. Therefore, the null hypothesis that the empirical distribution of *IMi* obtained for a branch *i* follows the same underlying normal distribution as that of branch *j* can be assessed using the two-sample Kolmogorov-Smirnov (KS) test (Ang & Tang, 2007). In practice, if the so-called p-value computed in the KS test is large, one cannot reject the null hypothesis of the two assessed samples being drawn from the same parent distribution. Conversely, if the p-values are smaller than the statistical significance level of interest (10% in the present case), the null hypothesis is rejected and the two empirical distributions are considered different at the statistical significance level adopted.

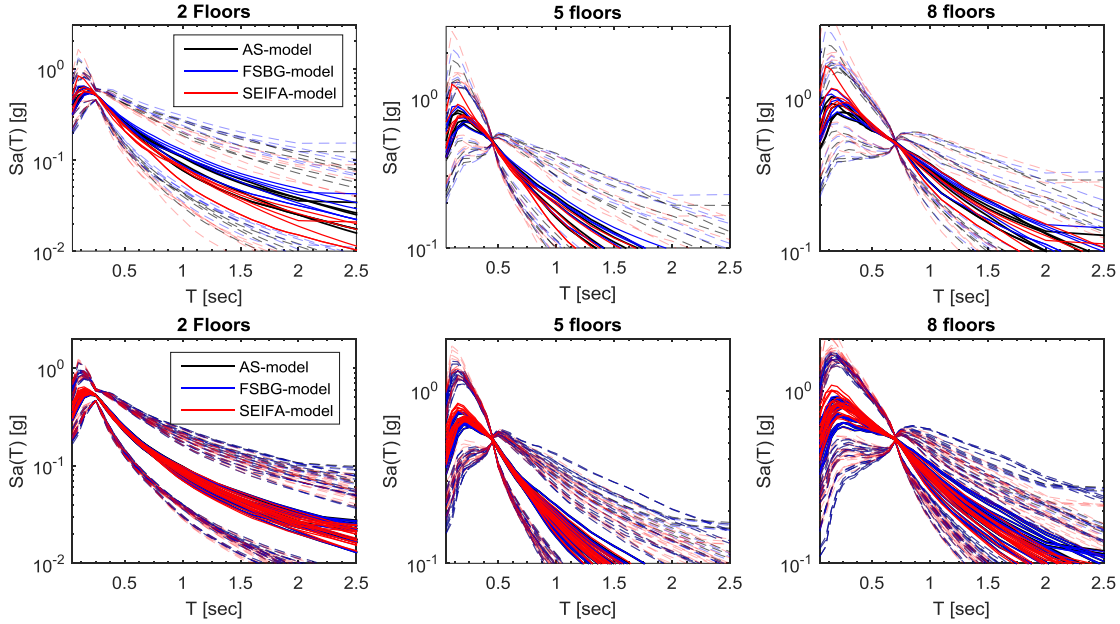


Figure 6.5 – Target spectral ordinates of periods ranging between 0.05 and 3.0 seconds (solid lines correspond to the mean and dashed lines represent 16 and 84 percentiles, i.e., mean +/- 1 standard deviation). Sites of Lisbon (upper) and Faro (lower), considering 2, 5 and 8-floor structures and a conditional $Sa(T_1)=0.5g$.

This exercise is illustrated in Figure 6.6, where the p-values obtained when comparing empirical distributions of HI computed for each of the logic-tree branches applicable to Lisbon are presented. Given the wide range of intensity measures considered in the process of record-selection, it would not be practical to demonstrate the results pertaining to the comparison of each individual *IMi*. For this reason, HI has been selected, as it is able to translate the differences in spectral ordinates at periods between 0.1 and 2.5 seconds (HI is the integral of the pseudo-velocity over the period range of 0.1 to 2.5 sec.).

According to Figure 6.6, p-values lower than 10% are identified in approximately 25%, 35% and 50% (respectively for 2, 5 and 8 floors classes) of the instances where

target distributions of HI are compared between branches *i* and *j* (with *i,j* ranging from 1 to 15, i.e. from 1 to the total number of applicable source model /GMPE combinations, in Lisbon). In other words, 'target' distributions of HI are considered statistically different (at the 10% significance level), in approximately 25%, 35% and 50% of the different combinations of branch pairs, for structural classes of 2, 5 and 8 floors, respectively. Although not presented for the sake of synthesis, similar results have been obtained for the remaining *IMi* and conditional levels of $Sa(T_1)$, when considering the logic-tree branches applicable to both Lisbon and Faro.

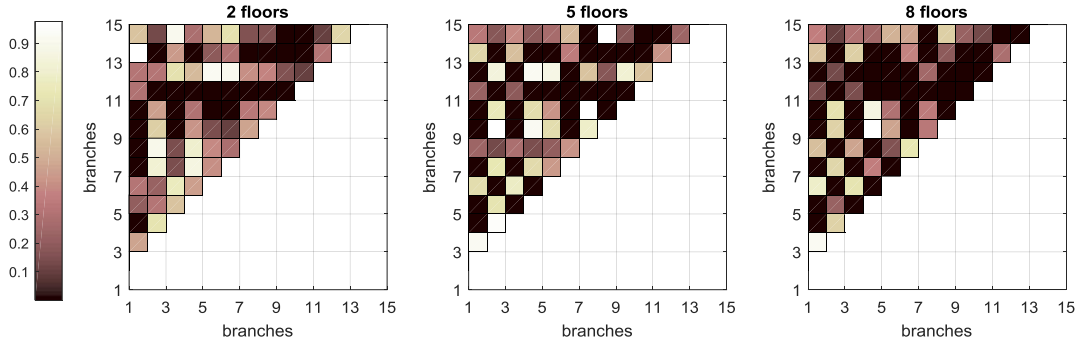


Figure 6.6 – p-values obtained with the KS test when comparing empirical distributions of HI computed for each of the logic-tree branches applicable to the site of Lisbon. Structural typologies of 2, 5 and 8-floors and conditional $Sa(T_1)=0.5g$. p-values lower than 0.1 are plotted in black.

Despite the fact that an extensive set of intensity measures has been considered for record selection (with that set including the IMs for which higher correlation with response quantities has been observed by Sousa et. al. (2014) in the analysis of similar structural models), the non-linear response of structures is a complex phenomenon that is influenced by ground motion properties that may not comprehensively be represented by the set of IMs considered herein. Therefore, despite the valuable insight provided by the results presented in Figure 6.6, a robust exercise requires further NLRHA and corresponding fragility and loss estimation, subsequently presented.

6.3 Hazard-consistency of fragility functions

The use of local criteria to define limit states in the context of building fragility analysis may not be appropriate, as highlighted in section 4.4. Therefore, in accordance with Chapter 3, Chapter 4 and Chapter 5, structural response is evaluated based on the maximum inter-story drift (ISD) and global drift (GD), considering four damage states: Slight Damage (SD), Moderate Damage (MD), Extensive Damage (ED) and Collapse

(Col). For the sake of synthesis, readers are referred to section 4.4.1 for details on the definition of damage state limits, as a function of *ISD* and *GD* response parameters.

6.3.1 Fragility assessment

For the purpose of fragility assessment, the methodology presented in Chapter 4 (where similar structural models are analysed) is adopted. According to the latter, for a given limit state and structural class, instead of computing a single damage exceedance probability for each level of $Sa(T_1)$, one is interested in determining 60 'record-specific' damage exceedance probabilities, conditioned on that same level of $Sa(T_1)$. Because 100 numerical models are analysed for each of the 60 ground-motion records (per level of intensity), this is accomplished by deriving the probabilistic distribution of 100 EDPs for each of the records, as thoroughly explained in sections 4.4.3 and 4.4.4 and illustrated in Figure 6.7.

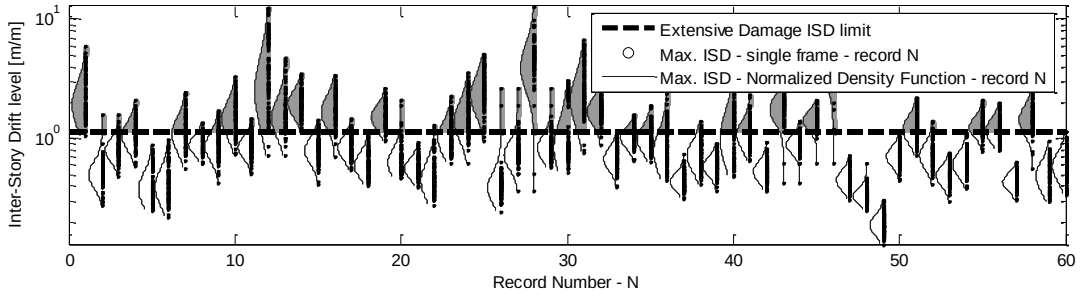


Figure 6.7 – Record-specific distributions of EDP and corresponding probabilities of exceedance of *Extensive Damage*, determined according to *ISD* limit state criteria for 8 story frames. Records selected and scaled for $Sa(T_1)=1.0g$ (previously presented as Figure 4.10)

In brief, when the EDP of interest is *ISD*, any 'record-specific' probability of exceedance of a given limit state corresponds to the area shown in dark grey in Figure 6.7. In the case of *GD* criteria, on the other hand, a distinct approach is adopted. Unlike the *ISD* criteria, where damage state thresholds are similar for all the structures (black dashed line in Figure 6.7), *GD* criteria are specific to each sampled frame (see section 4.4.1). Therefore, probabilities of exceeding a given limit state are perceived as the number of 'successes' (i.e. number of exceedances) in a sequence of 100 independent experiments that result in 'success' with identical probability (section 4.4.3.2).

For illustration, Figure 6.8 shows sets of 60 record-specific probabilities of exceedance of SD, MD, *ED* and *Col.* conditioned on $Sa(T_1)=0.1g$ to $1.0g$, as a function of *GD* and *ISD*. In this case, the site of Lisbon is selected and 5-story frames are considered.

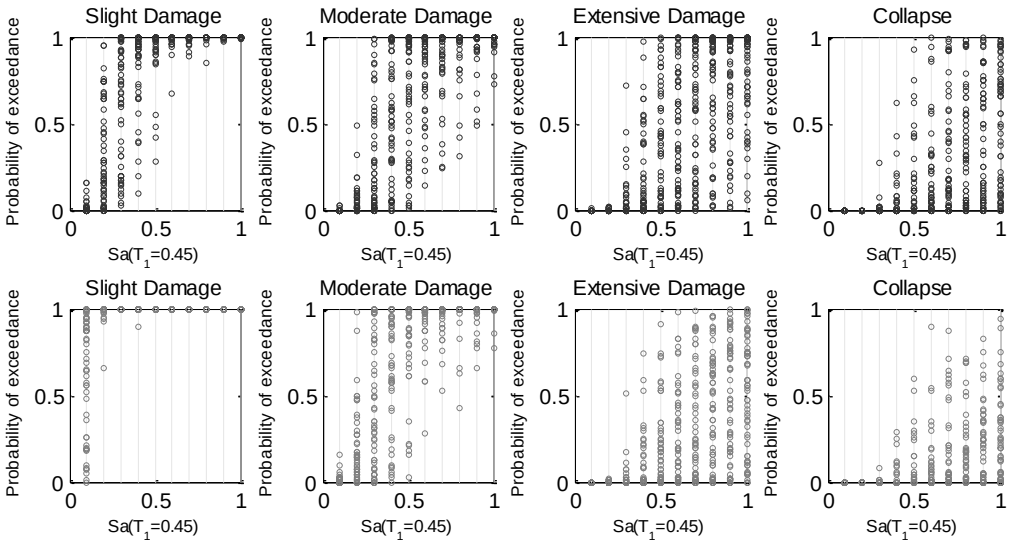


Figure 6.8 – Record-specific probabilities of exceedance of *SD*, *MD*, *ED* and *Col*, as a function of *GD* (upper) and *ISD* criteria (lower). Site of Lisbon and 5-story buildings (previously presented as Figure 4.13).

6.3.2 Fragility comparison

As in Chapter 4, building fragility is herein characterized by intensity-specific distributions of (record-specific) damage exceedance probability. In other words, for a given level of $Sa(T_1)$ and limit state, the 60 corresponding 'record-specific' probabilities follow a certain probabilistic distribution that can be determined empirically. As a result, one can easily recognize that, in order to compare the fragility results obtained for two different logic-tree branches, one can assess the null hypothesis that intensity-specific distributions arising from branch *i* are identical to those corresponding to branch *j*, at a given significance level.

For this purpose, the two-sample Kolmogorov-Smirnov (KS) methodology is used, since it allows also a helpful visual representation of the performed test. Loosely speaking, the null hypothesis that two data samples are drawn from the same parent distribution is tested by checking the largest discrepancy between the two empirical cumulative distributions (CDFs). Based on the significance level of interest (10% in the present case), a maximum allowed difference (D_m) is computed. If the largest discrepancy between CDFs is lower than D_m, one cannot reject the null hypothesis that the two samples share the same parent distribution. On the other hand, one considered the samples to be statistically different if the largest discrepancy is higher than D_m.

In order to illustrate the aforementioned exercise, Figure 6.9 shows the visual representation of the KS test performed for 2, 5 and 8-story frames (GD criteria), when

comparing distributions of *Collapse* probability conditioned on $Sa(T_1)=0.5g$, at the site of Lisbon.

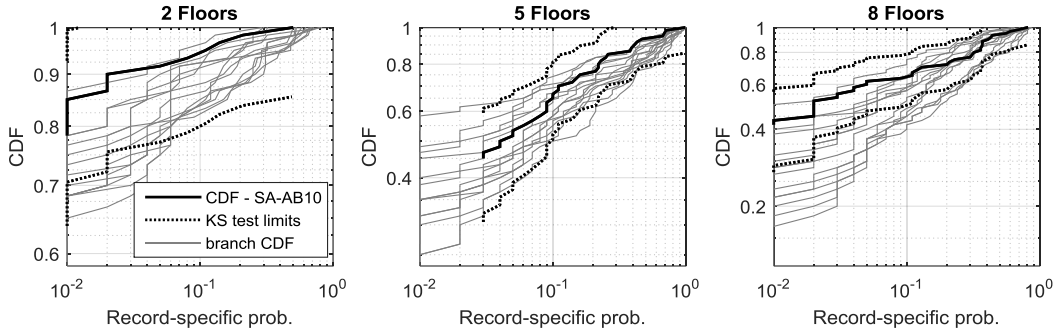


Figure 6.9 – KS test performed when comparing distributions of *Col.* probability conditioned on $Sa(T_1)=0.5g$, at the site of Lisbon. Damage criteria of GD, and structural classes of 2, 5 and 8-story buildings.

In this case, the distributions corresponding to SA-AB10 are compared with the results from the remaining 14 branches. The illustrated differences between CDFs are in the order of magnitude of the allowable limit (D_m), leading to the rejection of the null hypothesis for several branches. Since Figure 6.9 reflects only a fraction of results, Figure 6.10 to Figure 6.12 present the p-values obtained when using the KS test to statistically compare all the branch-specific distributions of damage exceedance probability, for both the sites of Lisbon and Faro. Here, limit states of *SD, MD, ED* and *Col.* are considered, and damage criteria of *ISD* and *GD* are foreseen, at a conditional $Sa(T_1)=0.5g$.

According to Figure 6.10 and Figure 6.11, where the black squares correspond to the rejection of the hypothesis of two samples of 60 damage exceedance probabilities sharing the same parent distribution, discrepancies amongst branches tend to increase from *SD* to *Collapse* (for 5 and 8-story buildings). Evidently, this trend is related with the increase of non-linear excursions as the limit state severity increases. For more severe limit states, a higher degree of uncertainty in structural response and consequent disparities between any two particular branches are verified.

In the case of 2-story frames, however, a distinct trend is exhibited. In this case, the number of discrepancies amongst branches is similar for limit states of *SD, MD* and *ED*, due to the higher influence of infill panel behaviour (in comparison with 5 and 8-story buildings). In the case of collapse, differences are much smaller than for prior limit states (for 2-story frames), because, unlike 5 and 8-story frames, the infill panels are still effective at high levels of deformation, resulting in values of collapse probability close to zero for virtually all the 60 records, irrespectively of the hazard branch.

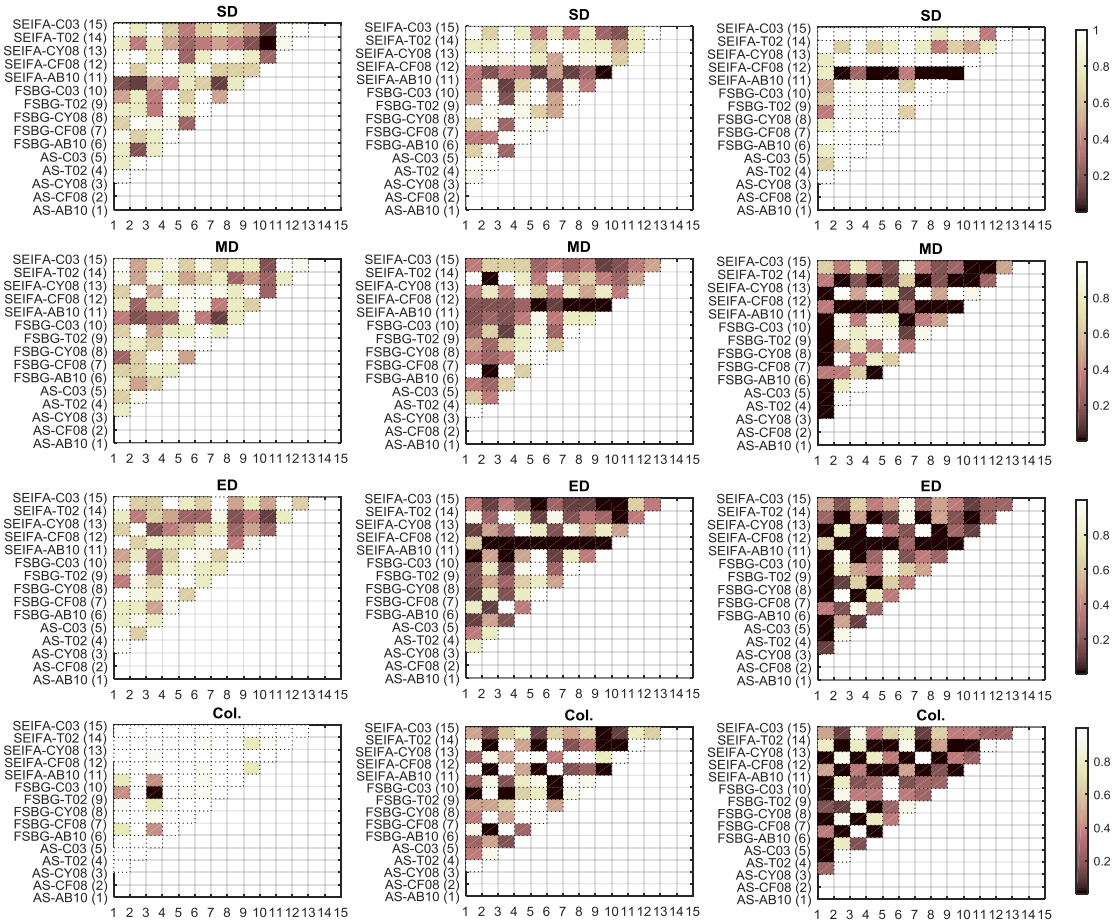


Figure 6.10 – p-values obtained with the KS test when comparing empirical distributions of damage exceedance probability computed for each of the logic-tree branches applicable to the site of Lisbon. Structural typologies of 2 (left), 5 (middle) and 8-floors (right), conditional $Sa(T_1)=0.5g$, and *GD* criteria. p-values lower than 0.1 are plotted in black.

Moreover, it is also possible to conclude that (irrespectively of the limit state) differences in 'branch-specific' samples of damage exceedance probability increase with the number of floors of the assessed structural class (i.e. for a given limit state, not only p-values decrease in general with the increase of the number of floors, as the number of instances in which p-value<alpha increases as well). This finding is consistent with the results presented in Figure 6.6, where (similar) discrepancies between record selection 'targets' from different branches are scrutinized within a similar statistical approach.

It is interesting to note that the discrepancies described above seem to be influenced equally by the differences between seismic source and ground motion prediction models. In other words, there is no evidence that discrepancies amongst branches are more pronounced when changing the GMPE (for a given source model), when compared with the differences registered between different source models, for a given GMPE.

For the sake of synthesis, only the results corresponding to GD criteria are presented in the case of Faro (in Figure 6.12). For the same reason, only $Sa(T_1)=0.5g$ has been

addressed in Figure 6.10 to Figure 6.12. However, results similar to those presented in this section have been verified for all the assessed levels of $Sa(T_1)$ and damage criteria (for locations of Lisbon and Faro).

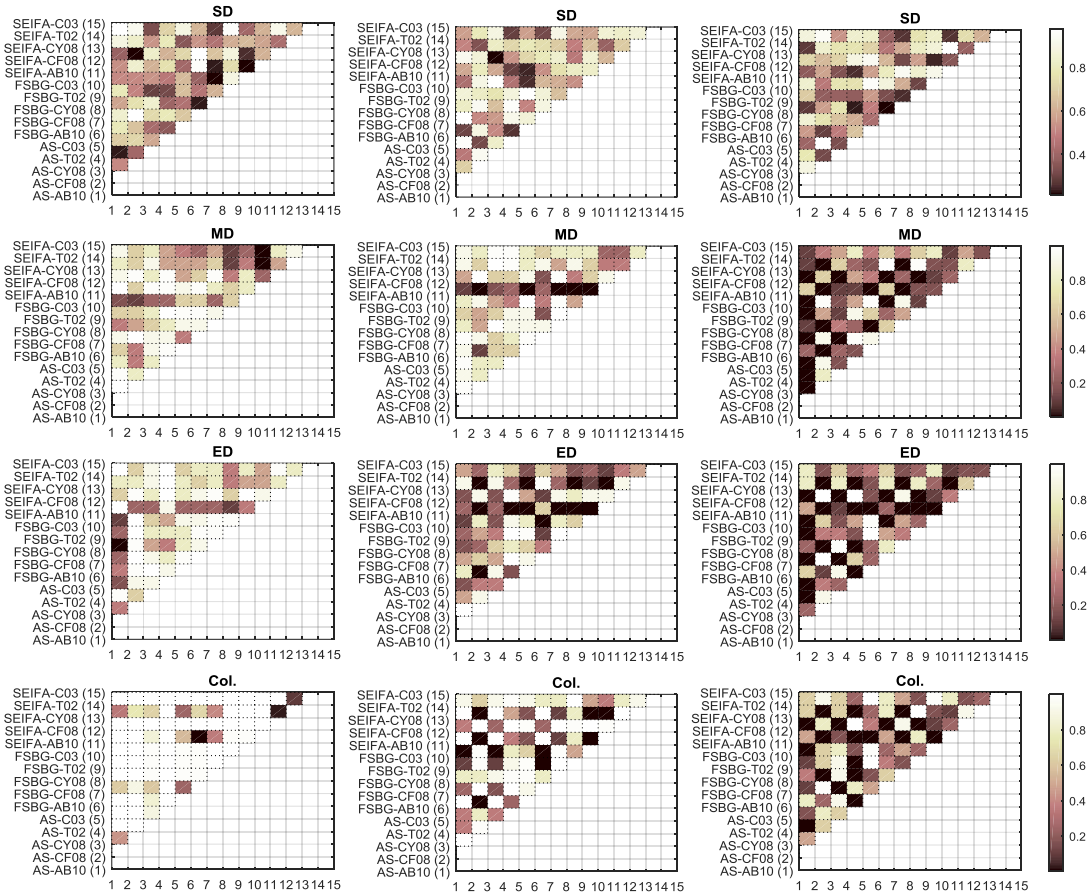


Figure 6.11 – p-values obtained with the KS test when comparing empirical distributions of damage exceedance probability computed for each of the logic-tree branches applicable to the site of Lisbon. Structural typologies of 2 (upper), 5 (middle) and 8-floors (lower), conditional $Sa(T_1)=0.5g$, and *ISD* criteria. p-values lower than 0.1 are plotted in black.

The aforementioned results corroborate the assumption made by the Author in the "Introduction" of this chapter, i.e. when the epistemic uncertainty of the hazard model is quantified through a logic-tree approach, fragility results are in fact "hazard-specific" (or "branch-specific", for simplicity). One might argue that this is the direct result of the selected intensity measure - $Sa(T_1)$ – not being *sufficient* (Luco, 2002). In other words, if one were to use a *sufficient* intensity measure, in which *sufficiency* is understood as the independency of response from parameters other than $Sa(T_1)$, then the differences in record selection 'targets' among different hazard branches (see Figure 6.5) would not be propagated into structural response (i.e. only $Sa(T_1)$ would influence the response). On the other hand, because such a sufficient intensity measure arguably does not exist, *IMs* other than $Sa(T_1)$ do in fact influence structural response, and different probabilistic

distribution of these *IMs* are reflected into distinct ('branch-specific") and fragility results.

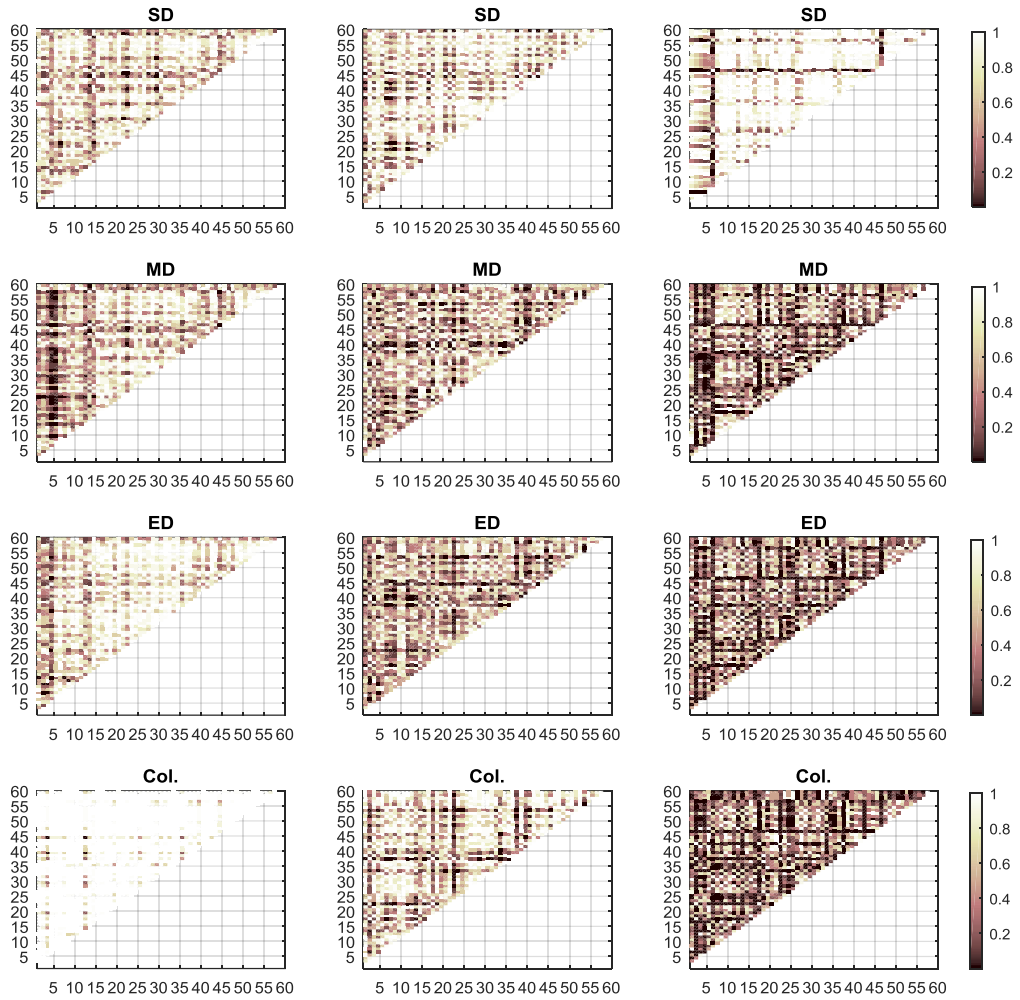


Figure 6.12 – p-values obtained with the KS test when comparing empirical distributions of damage exceedance probability computed for each of the logic-tree branches applicable to the site of Faro. Structural typologies of 2, 5 and 8-floors, conditional $Sa(T_1)=0.5g$, and *GD* criteria. p-values lower than 0.1 are plotted in black, and branches are numbered from 1 to 60.

6.4 Epistemic uncertainty and probabilistic loss estimation

In this section, the epistemic uncertainty of the hazard and fragility models (see previous section) is accounted for in a probabilistic loss estimation exercise in which 6 distinct building portfolios are considered.

6.4.1 Test-bed building portfolios

Using the data from the Portuguese Building Census survey of 2011 (INE, 2015), three building portfolios referring to two, five, and eight-story reinforced concrete pre-code buildings in the districts of Lisbon and Faro have been considered (resulting in a

total of 6 portfolios). Since the referred survey provides building counts at the parish-level resolution, the aggregation of assets at a single location (e.g. the parish centroid) can introduce significant errors. This is particularly relevant in larger regions, where the spatial distributions of seismic hazard and building exposure tend to be significantly unbalanced, i.e. the ground motion at the point where buildings are 'lumped' is virtually always different from that at the actual location of the assets. As a result, the GEOSTAT 2011 population distribution dataset (GEOSTAT, 2011), which provides the population count in a grid of 1 km^2 resolution, has been used to distribute the number of buildings in each parish proportionally to the amount of population estimated at each grid cell, as illustrated in Figure 6.13 and Figure 6.14.

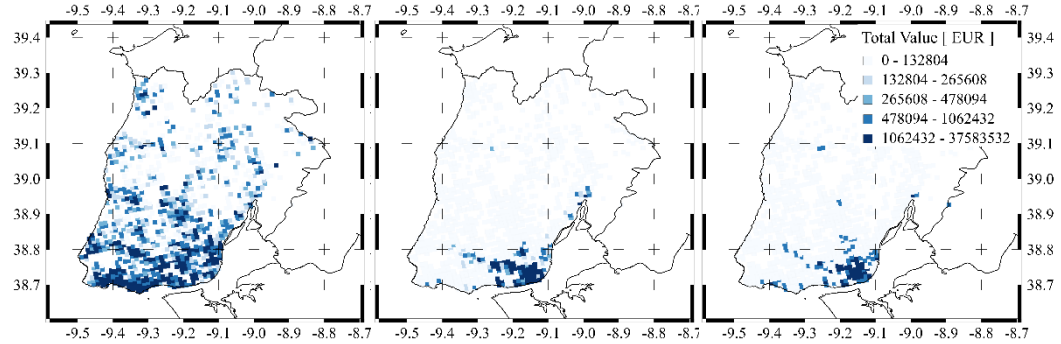


Figure 6.13 – Total replacement value (EUR) of two (left), five (middle) and eight-story (right) reinforced concrete pre-code buildings located in the district of Lisbon, Portugal. Spatial resolution of 1km^2

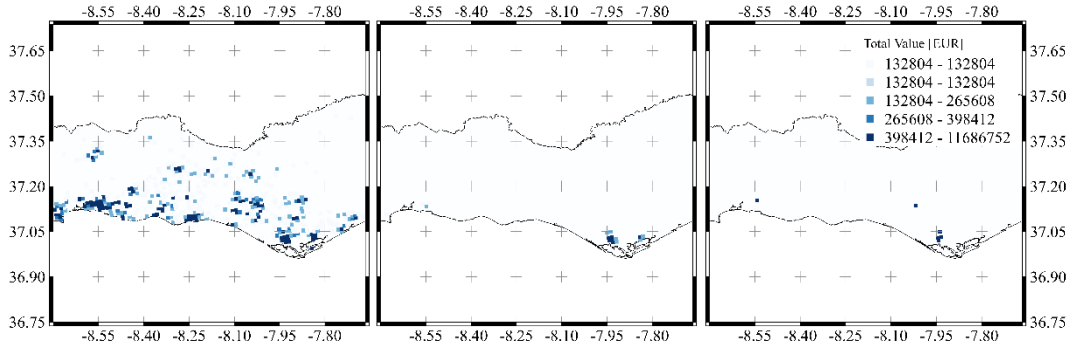


Figure 6.14 – Total replacement value (EUR) of two (left), five (middle) and eight-story (right) reinforced concrete pre-code buildings located in the district of Faro, Portugal. Spatial resolution of 1km^2. For the sake of visual clarity, only a part of the district is shown.

Similarly to what has been performed in section 5.4.1, the information provided by Silva et al. (2015b) has been used to determine the economic value of each building class as the product of: a) the respective average number of dwellings per class (which is a function of the number of floors), b) the average area per dwelling determined for the interested region, in m^2, and c) the average unit cost of replacement in *EUR*/m^2.

6.4.2 Fragility and vulnerability models

In order to assess the impact of propagating the uncertainty of the hazard model into the corresponding "branch-specific" fragilities, one is interested in determining what are the differences in risk estimates when considering:

a) A distinct 'branch-specific' fragility model for each branch (Figure 6.15);

b) A single model for all branches, as in state-of-the-art practice (e.g. (Pagani, et al., 2014)).

For the sake of consistency, it is important that approaches a) and b) are common in terms of the methodology used in their analytical derivation. Therefore, model b) is herein selected as the 'branch-specific' model that corresponds to the median hazard branch. It shall be highlighted that, in this context, the median branch is that associated with the median hazard curve. The 'median' is preferred over the 'mean' definition of hazard, since the computation of 'mean hazard' imparts several considerations regarding what the weights in the logic tree represent, in a probabilistic sense (Abrahamson & Bommer, 2005), which is not the focus of this study.

In Figure 6.15, an example of a 'branch-specific' fragility model is presented, along with the schematic illustration of how it is used to compute the corresponding vulnerability functions. More specifically, fragility results are characterized by what is commonly understood as the "true" value of exceedance probability of a limit state (ls_i), previously designated as $\tilde{P}ls_{i_{|Sa(T_1)=a}}$ (section 4.4.4.1). According to state-of-the-art practice, rather than computing 'record-specific' probabilities of exceedance (as presented in the previous section), one can obtain $\tilde{P}ls_{i_{|Sa(T_1)=a}}$ from the overall distribution of 6000 *EDP* values (60 ground motion records *x* 100 frames) for each level of $Sa(T_1)$. This is illustrated in Figure 6.15, where 'record-specific' exceedance probabilities of *SD, MD, ED* and *Col.* are plotted against the respective values of $\tilde{P}ls_{i_{|Sa(T_1)=a}}$.

As a result, a cumulative lognormal density function can be fitted to the intensity-dependent $\tilde{P}ls_{i_{|Sa(T_1)=a}}$ values, where the regression uncertainty is obtained by means of a bootstrap sampling method with replacement (Wasserman, 2004). In this framework, 200 synthetic datasets are randomly generated from the original values of $\tilde{P}ls_{i_{|Sa(T_1)=a}}$, resulting in 200 corresponding bootstrapped fragility functions, as illustrated in Figure 6.15. In this context, the appropriate correlation between probabilities of

exceedance of different damage states is guaranteed by the consistency between bootstrap samples across different damage states, i.e. the indices of the values that are drawn from the original dataset of *SD* exceedance probabilities are considered in order to constitute the corresponding bootstrap samples of *ED*, *MD* and *Collapse* probabilities for a given $Sa(T_1)=a$.

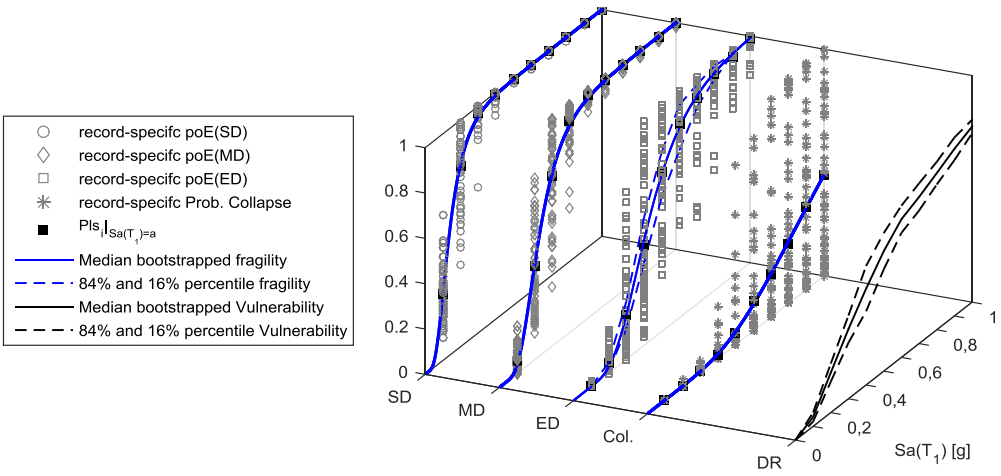


Figure 6.15 – Example of a 'branch-specific' fragility model and corresponding vulnerability, using the damage-to-loss relationship proposed by Silva et al. (2015c) and the 2-story building class. In this case, branch AS-AB10 and the site of Lisbon are selected, and GD criteria are considered. For simplicity, 'poE' stands for 'probability of exceedance'.

The damage-to-loss model used for combining fragility results into vulnerability functions is adapted from that previously in Chapter 5. Deterministic damage ratios (i.e. ratio between attained loss and replacement value of an asset) of 0.10, 0.30, 0.60 and 1.0 are assumed for limit states of *SD*, *MD*, *ED* and *Col.*, respectively. As a result, 200 damage ratios can be computed for each level of $Sa(T_1)$ (one for each bootstrapped fragility), resulting in a distribution of DR that can be approximated by a Beta probability density function (Ross, 2009).

6.4.3 Loss estimation methodology

In the process of risk computations, several studies have demonstrated the importance of accounting for spatial cross-correlation of ground motion residuals in the evaluation of portfolio losses (e.g. Weatherill et al. (2015), Silva (2016)). Therefore, the so-called probabilistic event-based risk tool of the OpenQuake-engine has been used in this study. This calculator uses stochastic event sets and associated spatially correlated ground-motion fields (Pagani, et al., 2014) to compute loss exceedance curves for each asset contained in the exposure model, using the correlation model proposed by Jayaram and Baker (2009).

When the correlation of uncertainty of the vulnerability model is incorporated in loss estimation procedures, it is typically done such that when sampling the vulnerability of two assets with the same building class, the residuals are assumed to be either uncorrelated of perfectly correlated (Taylor, 2015). This is evidently done in order to provide two boundary condition to the problem. However, in cases such as the present one, where the variability of damage ratios is related to the uncertainty of the fragility regression, there is no evident physical meaning behind the assumption of any degree of spatial correlation between damage ratio residuals. For this reason, loss calculations are herein performed assuming that damage ratio residuals are perfectly uncorrelated.

6.4.4 Loss assessment results

The risk to earthquake action of a given building portfolio is commonly described through a loss exceedance curve that specifies the annual frequency of exceedance of a range of possible loss values. Therefore, these curves are used in order to compare loss results obtained when considering methods a) and b) for each branch of the logic-tree applicable to the building portfolios located in Lisbon and Faro (see section 6.4.1).

Unquestionably, loss exceedance curves (and associated uncertainty) are extremely relevant to the comparison presented in this section, as illustrated in Figure 6.16, where the results of models a) and b) are shown for branch AS-AB10 of the logic tree applicable to the 5-story building portfolios located in Lisbon and Faro. However, comparing these curves may not be a straightforward matter, as doing so would require evaluating the differences in exceedance rates obtained for each of the possible loss values. Average Annual Loss values (Pagani, et al., 2014), on the other hand, while still encapsulating the information of the entire loss curve, are more concise and easily understandable outputs.

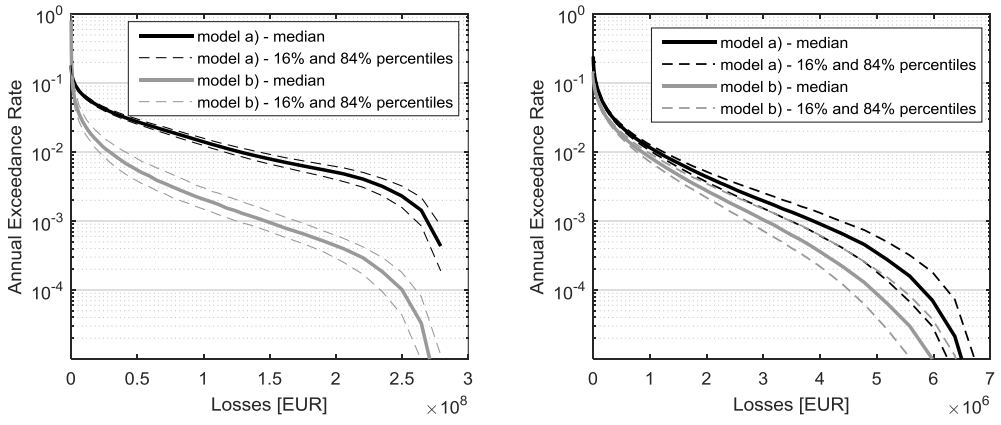


Figure 6.16 – Loss exceedance curves obtained with models a) and b) for branch AS-AB10. 5-story building portfolio located in Lisbon (left) and Faro (right), and GD criteria.

As a result, as shown in Figure 6.17 and Figure 6.18, loss exceedance curves (and associated uncertainty) are used to compute the corresponding AAL values, allowing one to determine the median, 16% and 84% percentile (absolute) differences between the results of approaches a) and b), for each logic-tree branch.

According to Figure 6.17 and Figure 6.18, it is evident that approach b) (i.e. considering a single fragility model for all the logic-tree branches) is not capable of reflecting the 'hazard-dependency' of building fragility and associated epistemic uncertainty. As illustrated, absolute errors with respect to the 'branch-specific' model a) can range up to 1000% for portfolios of 5 and 8-story buildings located in the districts of Lisbon and Faro, irrespectively of damage criteria. In this context, it shall be noted that, for the sake of synthesis, only the results of the 5-story portfolio are presented in the case of Faro. However, the errors obtained for 8-story structures are similar to those illustrated for the 5-story buildings in this location.

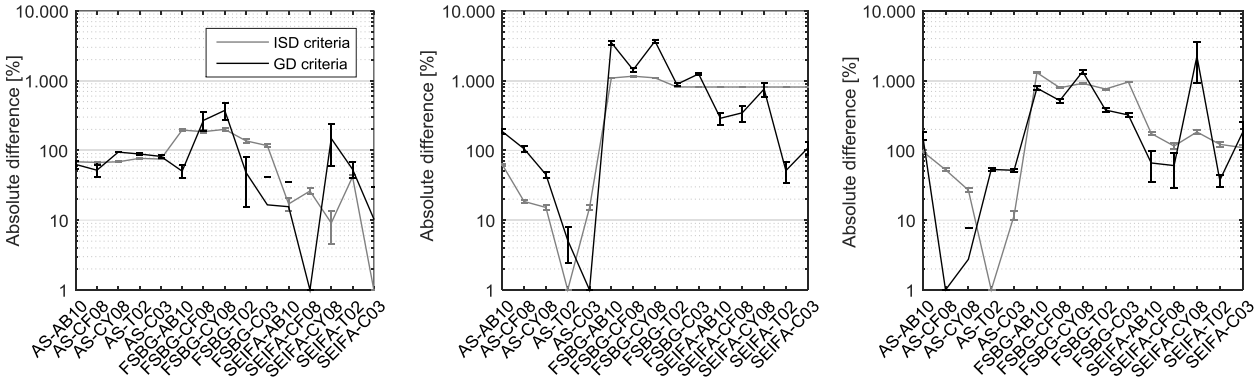


Figure 6.17 – Median, 84% and 16% percentile absolute differences between EAL computed using models a) and b), for all the branches. 2 (left), 5 (middle) and 8-story (right) building portfolios located in Lisbon.

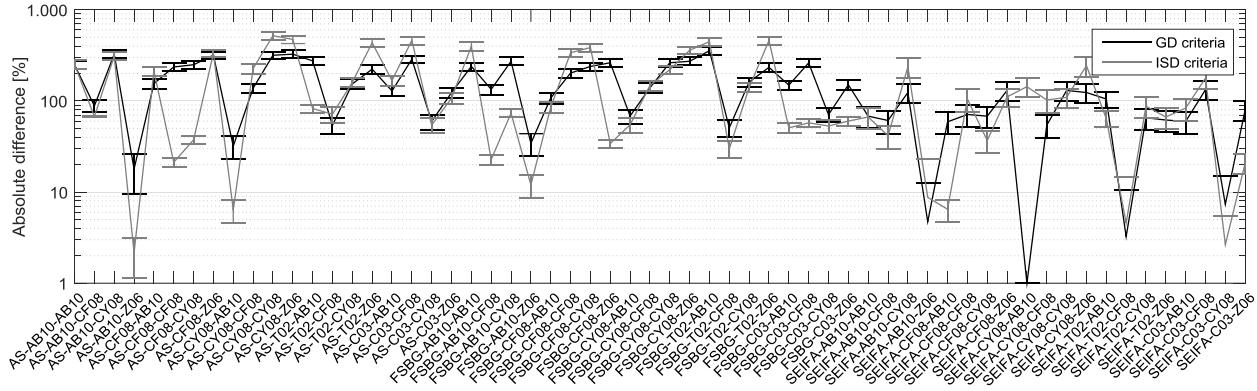


Figure 6.18 – Median, 84% and 16% percentile absolute differences between EAL computed using models a) and b), for all the branches. 5-story building portfolio located in Faro.

In the case of 2-story portfolios, on the other hand, errors tend to be one order of magnitude lower than those of 5 and 8-story buildings (Figure 6.17). This is consistent with the smaller discrepancies between 'branch-specific' fragilities exhibited in the case of the 2-story building class (see section 6.3.2 and Figure 6.10 to Figure 6.12). In other words, because differences between 'branch-specific' fragilities are generally lower in the case of 2-story buildings (especially for damage state of *Col.*), the discrepancies between fragility models a) and b) are also smaller than in the cases of 5 and 8-story buildings. Nonetheless, the errors attained are still in excess of 100% for most branches (Figure 6.17), which is clearly not satisfactory.

6.5 Conditional fragility functions and hazard-consistent fragility

It is clear that the only suitable way to appropriately propagate the epistemic uncertainty of the hazard model into the corresponding loss estimation results is by considering a fragility model that is consistent with each of the considered hazard-modelling approaches. Unless a sufficient intensity measure is used, fragility results are dependent on the properties of the ground-motion records selected for each logic-tree branch, which, in turn, are distinct for different source model / GMPE combinations.

In this framework, it is necessary to develop one fragility model per hazard logic-tree branch, which imparts a very significant computational effort. Therefore, in this section, a methodology based on the concept of conditional fragility functions previously proposed (Chapter 4 and Chapter 5) is presented, allowing one to avoid the effort of performing record selection and NLRHA for each logic-tree branch, while ensuring that the fragility model considered is nonetheless 'branch-specific'.

6.5.1 Conditional fragility functions and its use in loss estimation

As presented in Figure 6.8, building fragility is herein characterized by intensity-specific distributions of (record-specific) damage exceedance probability. Therefore, for a given level of $Sa(T_1)$, if one assigns the record-specific values of each *IMi* determined in section 6.2.3.1 (denoted as $IMi|_{Sa(T_1)=a}$) to the corresponding record-specific probabilities, a cumulative lognormal curve can be fitted to the resulting scatter. According to section 4.4.5, once the most efficient *IMi* is selected (i.e. the one for which the correlation with damage exceedance probabilities of *SD*, *MD*, *ED* and *Col* is higher),

the resulting *conditional fragility functions* provide a parametric relationship between $IMi|_{Sa(T_1)=a}$ and damage exceedance probabilities when records are selected and scaled for $Sa(T_1)=a$ (Figure 6.19).

In Chapter 5, a loss estimation framework that features the simulation of conditional spatially correlated random fields of $IMi|_{Sa(T_1)=a}$ is presented, in order to demonstrate the advantages of using *conditional fragility functions* within the context of loss estimation of building portfolios. More specifically, it is shown that conditional fragility functions are particularly relevant for the 'hazard-consistent' modelling of the uncertainty in fragility and its spatial correlation, which greatly influences the loss estimation of spatially distributed portfolios. For illustration, Figure 6.19 shows one example of a conditional fragility model, as well as the corresponding state-of-the-art approach of considering only the values of $\tilde{P}ls_{i|Sa(T_1)=a}$.

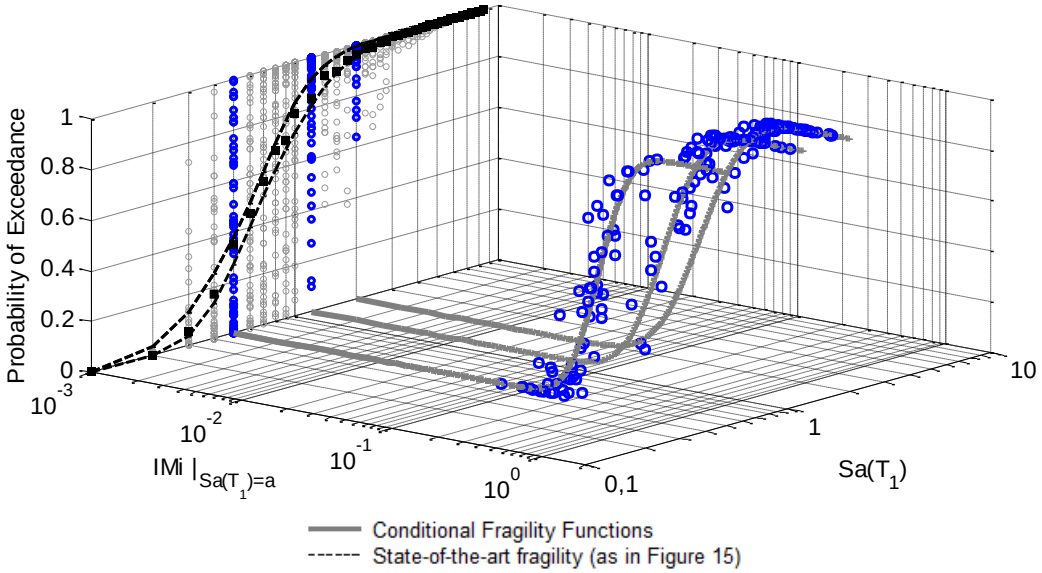


Figure 6.19 – Record-specific probabilities of exceedance of *ED*, as a function of *GD* criteria, and corresponding *conditional fragility functions* for the cases of $Sa(T_1)$=0.5g, 1.2g and 2.0g, as well as the values of $\tilde{P}ls_{i|Sa(T_1)=a}$ (black squares) and associated uncertainty determined by 200 bootstrapped lognormal fragility curves, for 5-story buildings located in Lisbon (previously presented as Figure 5.13).

6.5.2 Fragility models and loss estimation methodology

Here, the matters of spatial correlation of fragility and vulnerability residuals studied in Chapter 5 will not be addressed in detail. Differently, the objective of the present section is to demonstrate how, through the use of *conditional fragility functions*, it is possible to ensure the 'hazard-consistency' of the fragility model, while avoiding the cumbersome NLRHA for each branch of the logic-tree.

In order to do so, it is important to highlight one of the most important characteristics of *conditional fragility functions*. As demonstrated in section 5.2, a single *IMi* used as abscissa of these functions is *sufficient* to account for all the explanatory variables (i.e. all the *IMi*) influencing the seismic response of the assessed structures, at a given level of $Sa(T_1)$. As a result, it follows that although the distributions of 'record-specific' exceedance probability are clearly dependent on the hazard branch of interest (as demonstrated in section 6.3.2), the 'shape' of the *conditional fragility functions* fitted to these probabilities is, in theory, independent of the hazard branch. In other words, because $IMi|_{Sa(T_1)=a}$ is *sufficient*, the 'shape' (i.e. the parameters) of the *conditional fragility functions* is independent of the source model / GMPE combination.

With the aim of better understanding the statement above, one might think of the following analogy. Since the verified *sufficiency* implies that considering $IM|_{Sa(T_1)=a}$ does not provide any additional information with respect to using simply $IMi|_{Sa(T_1)=a}$, the values of exceedance probability for a given record depend only on the corresponding value of $IMi|_{Sa(T_1)=a}$. Therefore, the *trend* according to which the increase of $IMi|_{Sa(T_1)=a}$ will increase the resulting damage exceedance probability depends only on the structural properties of the assessed buildings. This *trend* is thus established by the fitted *conditional fragility function* and, since it depends only on the properties of the assessed buildings, it is independent of the source model / GMPE combination considered.

In this context, possible differences between *conditional fragility functions* (*CFFs*) obtained for different branches are (only) due to the uncertainty that is associated with the regression analysis used to fit the curves. Therefore, in order to verify the assumed independence of *CFFs* with respect to the source model / GMPE, as well evaluating the impact of possible differences between *CFFs* across branches (due to regression uncertainty), a loss estimation study is carried out. Here, similarly to the exercise presented in section 6.4, two distinct fragility models are considered:

a) A distinct 'branch-specific' fragility model for each branch;

b) A single model for all branches, as in state-of-the-art practice.

In accordance with section 6.4.2, model b) is selected as the 'branch-specific' model that corresponds to the median hazard branch. However, the fragility models considered herein consist of the aforementioned *conditional fragility functions*, rather than 'common' fragility curves regressed over the $\tilde{P}ls_{i_{|Sa(T_1)=a}}$ values (Figure 6.19). For the sake

synthesis, readers are referred to Chapter 5 regarding the theoretical details of the framework that resorts to *conditional fragility functions* for the probabilistic computation of losses, which is used herein.

6.5.3 Loss estimation results and comparison

As in Figure 6.17 and Figure 6.18 (section 6.4.4), Figure 6.20 and Figure 6.21 show the median, 16% and 84% percentile (absolute) differences between the AAL values obtained when using approaches a) and b), for each logic-tree branch and building portfolio. As illustrated, the errors obtained using the methodology proposed in the previous sections 6.5.1 and 6.5.2 present maximum values of approximately 10%, as opposed to 100% and 1000% verified in the case of 'state-of-the-art' fragility and loss estimation approach (section 6.4.4).

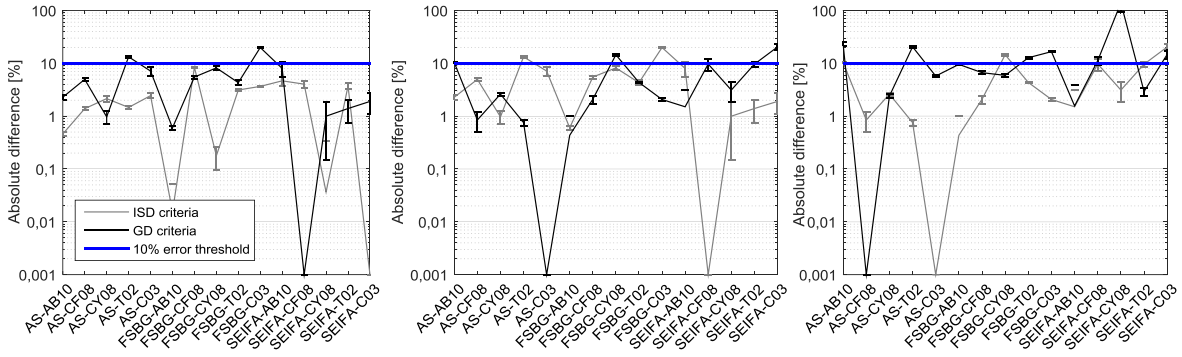


Figure 6.20 – Median, 84% and 16% percentile absolute differences between EAL computed using models a) and b), for all the branches. 2 (left), 5 (middle) and 8-story (right) building portfolios located in Lisbon.

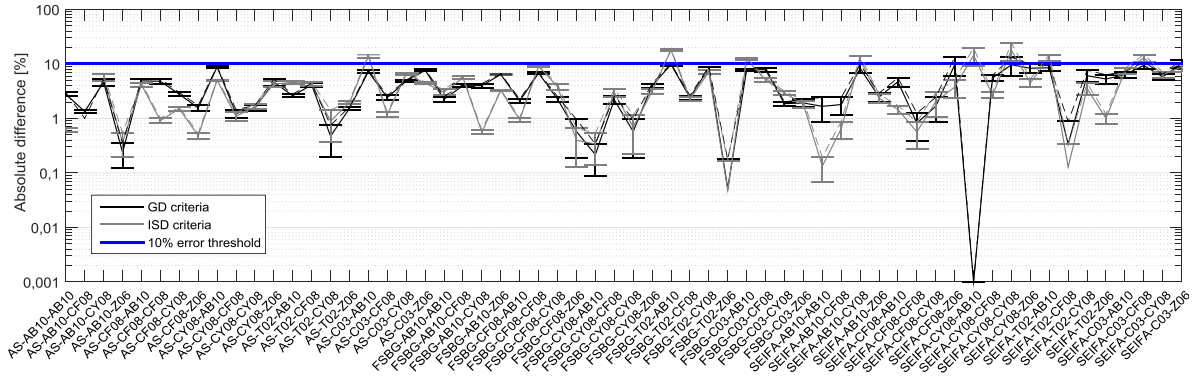


Figure 6.21 – Median, 84% and 16% percentile absolute differences between EAL computed using models a) and b), for all the branches. 5-story building portfolio located in Faro.

The above findings are consistent with the verified *sufficiency* of the *conditional fragility functions*, which validates the proposed methodology as an alternative to the time-consuming evaluation of fragility for each logic-tree branch. More specifically, it is

demonstrated that, when using *conditional fragility functions* and the associated loss estimation methodology (Chapter 5), it is possible to use a single fragility model common to all the hazard branches (e.g. that of the 'median' hazard branch), and still ensure the 'hazard-consistency' of results within a maximum error of 10%.

6.6 Final remarks

In this research, the subject of 'hazard-dependency' of building fragility has been addressed, through a methodology that incorporates: a) probabilistic seismic hazard analysis (PSHA) for the sites of Lisbon and Faro (Portugal), using a logic-tree approach that reflects 3 distinct seismological models and 20 combinations of ground motion prediction equations, b) hazard-compatible record selection and nonlinear response history analysis (NLRHA) for each logic-tree branch and each site, and c) fragility analysis and comparison between fragility functions obtained for each assessed hazard branch, at each site. According to the appraised results, it has been demonstrated that, when analytical methodologies are used to characterize both hazard and fragility components, the epistemic uncertainty of the hazard model shall be propagated into the fragility analysis. More specifically, it has been verified that fragility results are statistically different amongst logic-tree branches, which corroborates the proposed assumption that fragility functions are 'hazard-specific' (or 'branch-specific', for simplicity).

In order to further evaluate the impact of the above findings in the context of seismic risk, the probabilistic loss estimation of 6 different building portfolios located in the districts of Lisbon and Faro has been performed. In this exercise, loss exceedance curves and corresponding Average Annual Losses were computed for each hazard branch (and each portfolio), using two distinct fragility assessment approaches: a) a distinct model for each logic-tree branch, consistent with each distinct hazard model, and b) a state-of-the-art method in which only a single fragility model, common to all branches, is used. As a result, it was demonstrated that the errors associated with using a single fragility model (approach b)), can be as high as 1000%, when compared with the hazard-consistent approach a).

In light of the significance of the above results, it is clear that, in order to appropriately propagate the epistemic uncertainty of the hazard model into the corresponding loss estimation results, one shall define fragility models that are consistent

with each of the considered hazard-modelling approaches (or branches). The more evident way to accomplish this is by performing record-selection and nonlinear response history analysis (NLRHA) for each of the logic-tree branches, which is clearly time-consuming and undesirable. Therefore, an alternative and innovative fragility assessment / loss estimation methodology has been proposed. This framework ensures the hazard-consistency of the fragility model, while avoiding the necessity to perform NLRHA for each branch (i.e. only a single fragility model is used). More specifically, this methodology is based on the concept of *conditional fragility functions* presented in Chapter 4, allowing the use of a single fragility model common to all the hazard branches (e.g. that of the 'median' hazard branch), while ensuring that the loss estimates computed for each branch are consistent with the corresponding hazard model, within a maximum error of 10%.

Chapter 7 CONCLUSIONS AND FUTURE DEVELOPMENTS

7.1 Conclusions

The subjects addressed in this work can be divided into two main areas of research: the improvement of methods and tools for the development of exposure datasets of buildings of industrial use, at the European scale, and the treatment of aleatory and epistemic uncertainty in the stages of fragility, vulnerability and loss estimation of building portfolios.

The proposed exposure modelling algorithm was implemented in an automated tool that provides the means to overcome the potential lack of input data in a statistically meaningful way, using Volunteered Geographic Information (VGI) and Open-Access data as the main sources of information. This procedure was validated using exogenously provided data on industrial buildings in three distinct regions in Europe, demonstrating the excellent agreement between real and inferred exposed areas. Following the application of this framework to 36 European countries, the spatial distribution of industrial buildings in Europe was coupled with probabilistic seismic hazard results, showing that Turkey and Italy are the two countries with the largest number of industrial assets located in medium and high seismicity areas.

As an effort towards the study of uncertainty in seismic risk, possible advantages and limitations of considering approximate solutions to the problem of hazard-compatible record selection and subsequent analytical fragility and loss assessments were investigated. It was demonstrated that, within the wide range of structural properties, response parameters, seismic source modelling options, and ground motion prediction equations assessed, only an exact disaggregation method guarantees a satisfactory

outcome in terms of accuracy, when limit state criteria are not building-specific. In the case where limit state thresholds are building-specific, on the other hand, the main contribution of this study lies on the robust proposal of suitable levels of approximation recommended to be used by the research and practitioner communities.

Regarding the study of aleatory and epistemic uncertainty of structural capacity and seismic demand, as well as its impact on the evaluation of fragility of building portfolios, the importance of the statistical significance of results was demonstrated with respect to: minimum number of selected ground motion records, and estimation of damage exceedance probabilities. The relevance of the resulting fragility model approach was further demonstrated within the context of loss estimation of building portfolios. The importance of the introduced concept of *conditional fragility functions* was verified in the context of its capability to consistently take into account record-to-record variability in the evaluation of fragility, while establishing the means by which spatial correlation between damage exceedance probabilities can be taken into account.

Building upon the above findings, the aforementioned *conditional* fragility assessment methodology was extended to the derivation of vulnerability functions that reflect a non-parametric (site-specific) histogram of *damage ratio* per level of primary intensity measure. As one of the main scientific contributions of this thesis, this framework provides the link between vulnerability uncertainty and seismic hazard, such that when sampling the uncertainty in the vulnerability of spatially distributed assets, the correlation of its residuals is taken into account through an approach that is physically connected with reality. In addition, it was demonstrated that the proposed loss assessment framework leads to results that differ with respect to state-of-the-art methods in a consistent manner across different building classes and damage state definition criteria. More specifically, it was verified that state-of-the-art methodologies tend to overestimate the annual rate of exceedance of lower (i.e. more frequent) loss values, whereas the opposite trend is verified for higher aggregated losses. The latter was shown to be related with the more robust representation of the impact of record-to-record variability ensured by the proposed methodology, highlighting its strengths and contribution to the improvement of fragility, vulnerability and loss assessment of building portfolios.

Finally, the subject of ‘hazard-dependency’ of building fragility was addressed, demonstrating that, when analytical methodologies are used to characterize both hazard

and fragility components, the epistemic uncertainty of the hazard model shall appropriately be propagated into the fragility analysis. More specifically, it was verified that fragility results vary with the selected hazard modelling options, which corroborates the hypothesis outlined in this thesis, i.e. that fragility functions are 'hazard-specific'. As a result of this finding, it was demonstrated that the errors associated with neglecting the hazard-dependency of fragility models (which is common practice in the literature), can lead to errors as high as 1000% in terms of risk estimates, when compared with the proposed hazard-consistent approach. In light of the significance of these results, the previous findings and proposals of this thesis (with regard to the treatment of uncertainties) were combined into the development of an alternative and innovative fragility assessment / loss estimation methodology. The presented framework was proved to guarantee the hazard-consistency of fragility based on the concept of *conditional fragility functions*, while allowing different loss estimates to be in agreement with the corresponding hazard model.

7.2 Future developments

All the tools, methods and theoretical assumptions proposed in this thesis were presented with the level of detail necessary to be used, replicated and disseminated by the research and practitioner communities. However, in order to enhance its reach, applicability, and ease of use, one of the immediate future developments consists of the implementation of the proposed frameworks (specifically those of Chapter 2, Chapter 4, Chapter 5 and Chapter 6) into public and transparent open-source tools. Moreover, in order to engage researchers into the study and further development of the subjects addressed in Chapter 4, Chapter 5 and Chapter 6, the second further objective foresees the development of alternative solutions that, without hampering the robustness and accuracy of the proposed fragility, vulnerability and loss estimation methodologies, guarantees an increased efficiency in terms of computational demand.

With respect to the exposure model algorithm presented in Chapter 2, in particular, the proposed methodology will be improved in order to foresee not only the evaluation of the spatial distribution of industrial areas in Europe, but also the characterization of building vulnerability. More specifically, methods that allow the characterization of the identified building portfolios in terms of construction type and associated vulnerability

will be implemented. This is a crucial step towards the applicability of the proposed framework, as it will allow the direct use of its results in probabilistic loss estimation studies. Furthermore, efforts will be carried out towards the extension of the methodology to other regions in the world. For this purpose, the availability of similar datasets to those used in the European case will be assessed, keeping in mind the intended open-access nature of the tool.

In the study presented in Chapter 3, current limitations may be identified with respect to the range of structural classes considered and structural response uncertainty. Therefore, the variability of structural capacity of the models used in this work will further be addressed through a simulation procedure similar to that of Chapter 4, where 100 building models per construction class are studied. Moreover, additional building classes and locations will be addressed, so as to investigate the validity of the presented assumptions in the context of more comprehensive ranges of structural response and site hazard conditions.

As one of the limitations previously identified in Chapter 3, the level of accuracy of the different approximate solutions is determined by the increase of magnitude and distance bins, in equal proportion. As such, an ongoing development consists of adopting additional approximate solutions in which magnitude and distance bins are increased independently (i.e. increase of distance intervals for a given magnitude value, and vice-versa). This study will be performed in order to study the relative importance of the accuracy of each parameter when compared with the exact solution.

Following the findings presented in Chapter 3, the exact formulation derived for the computation of hazard disaggregation is currently being implemented in a web tool that will be capable of performing ground motion selection fully compatible with the hazard at a given site of interest. This tool will be able to, upon user input, provide ready-to-use selection of natural ground motion records for nonlinear response history analysis of buildings located in any site within Europe, using the SHARE results as the default hazard model. In addition, it will be possible to perform identical task for any additional site outside Europe, upon introduction of an additional hazard model by the user (in the format used by the OpenQuake engine). When epistemic uncertainty is taken into account (such as the case of the built-in SHARE hazard model), it will furthermore be possible to select which hazard output is of interest to the user, for ground motion selection (i.e. mean, median, or any fractile result). The most appropriate ground motion prediction model and

applicable soil type will automatically be assigned by the tool, in accordance with the selected site. However, all the aforementioned options will be available for manual configuration by advanced users.

With respect to the fragility, vulnerability and loss estimation methodologies proposed in Chapter 4, Chapter 5 and Chapter 6, the extension of the loss estimation procedure presented in Chapter 6 will be performed in the immediate future, in order for it to be applicable to portfolios containing multiple building classes. Based on the theoretical formulation presented in the aforementioned chapter, the necessary computation of conditional cross-correlated ground motions fields is currently only possible for a single intensity measure at a time (which imparts the inability to consider portfolios where distinct intensity measures are used for different building classes). In order to make it possible to consider multi-class portfolios, an innovative model for the analytical characterization of the conditional spatial cross-correlation between distinct intensity measures will be developed. For this purpose, cooperation with engineering seismology experts is currently being outlined, in order to develop a methodology that (continues to) ensure a robust link between seismic hazard and building response.

Once the aforementioned objectives are achieved, the subsequent task is the development of a comprehensive conditional fragility and vulnerability model for the entire Portuguese building stock. This will further allow the calibration and enhancement of the aforementioned open-source tools, so as to encourage researchers to apply this methodology in other regions in the world. As a result of these efforts, it will be possible to apply the updated vulnerability model and conditional loss estimation framework in the computation of probabilistic seismic losses and evaluation of meaningful event risk scenarios in Portugal (and other parts of the world, through cooperation with engaged researchers and practitioners).

This page is intentionally left blank

REFERENCES

Abrahamson, A., & Bommer, J. (2005). Probability and Uncertainty in Seismic Hazard Analysis. *Earthquake Spectra*. doi:http://dx.doi.org/10.1193/1.1899158

Abrahamson, N. (2006). Seismic hazard: problems with current practice and future developments. *Proceedings of the 1st European Conference of Earthquake Engineering and Seismology.* Geneva, Switzerland.

Akkar, S., & Bommer, J. (2010). Empirical equations for the prediction of PGA, PGV and spectral accelerations in Europe, the Mediterranean region and the Middle East. *Seismological Research Letters, 81*(2), 195–206.

Ang, A., & Tang, W. (2007). *Probability Concepts in Engineering: Emphasis on Applications in Civil and Environmental Engineering.* Hoboken, New Jersey: John Wiley & Sons,.

Antoniou, S., & Pinho, R. (2004). Development and verification of a displacement-based adaptive pushover procedure. *Journal of Earthquake Engineering, 8*(5), 643–661.

Araújo, M., Castro, M., & Marques, M. (2015). Industrial Seismic Risk Assessment for Mainland Portugal. *Natural Hazards, in review*.

ASCE. (2010). *Minimum Design Loads for Buildings and Other Structures (ASCE 7-10).* Reston, VA: American Society of Civil Engineers.

Atkinson, G., & Boore, D. (2006). Earthquake ground-motion prediction equations for eastern North America. *Bulletin of the Seismological Society of America, 96*(6), 2181–2205.

Baker, J. (2007a). Correlation of ground motion intensity parameters used for predicting structural and geotechnical response. *ICASP10 : Applications of Statistics and Probability in Civil Engineering: Proceedings of the 10th International Conference.* Tokyo, Japan.

Baker, J. (2007b). Measuring bias in structural response caused by ground motion scaling. *Proceedings of the 8th Pacific Conference on Earthquake Engineering.*

Baker, J. (2011). Conditional mean spectrum: tool for ground-motion selection. *Journal of Structural Engineering, SPECIAL ISSUE: Earthquake Ground Motion Selection and Modification for Nonlinear Dynamic Analysis of structures*, 322-331.

Baker, J. (2015). Efficient analytical fragility function fitting using dynamic structural analysis. *Earthquake Spectra, 31*(1), 579-599.

Baker, J., & Cornell, A. (2005). *Vector-valued ground motion intensity measures for probabilistic seismic demand analysis.* Stanford, CA: The John A. Blume Earthquake Engineering Center.

Baker, J., & Cornell, A. (2006a). Correlation of response values for multicomponent ground motions. *Bulletin of the Seismological Society of America*. doi:10.1785/0120050060

Baker, J., & Cornell, A. (2006b). Spectral shape, epsilon and record selection. *Earthquake Engr. & Structural Dynamics, 34*(10), 1193-121.

Baker, J., & Cornell, A. (2006c). Which spectral acceleration are you using? *Earthquake Spectra, 22*(2), 293–312.

Baker, J., & Jayaram, N. (2008). Correlation of spectral acceleration values from NGA ground motion models. *Earthquake Spectra, 24*(1), 299–317.

Bal, I., Crowley, H., & Pinho, R. (2010). *Displacement-based earthquake loss assessment: method development and application to Turkish building stock. ROSE Research Report 2010/02.* Pavia, Italy: IUSS Press.

Bazzurro, P., & Cornell, A. (1999). Disaggregation of seismic hazard. *Bulletin of the Seismological Society of America, 89*(2), 501–520.

Bazzurro, P., & Luco, N. (2005). Accounting for uncertainty and correlation in earthquake loss estimation. *9th International Conference on Structural Safety and Reliability (ICOSSAR'05).* Rome, Italy.

Bazzurro, P., & Park, J. (2007). The effects of portfolio manipulation on earthquake portfolio loss estimates. *10th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP10).* Tokyo.

Bommer, J., & Abrahamson, N. (2006). Why do modern probabilitic seismic hazard analyses often lead to increased hazard estimates? *Bulletin of the Seismological Society of America, 96*, 1967-1977.

Bommer, J., & Crowley, H. (2006). The Influence of Ground-Motion Variability in Earthquake Loss Modelling. *Bulletin of Earthquake Engineering, 4*, 231-248. doi:DOI 10.1007/s10518-006-9008-z

Bommer, J., & Scherbaum, F. (2008). The use and misuse of logic-trees in probabilistic seismic hazard analysis. *Earthquake Spectra, 24*(4), 997-1009.

Bommer, J., Spence, R., Erdik, M., Tabuchi, S., Aydinoglu, N., Booth, E., . . . Peterken, O. (2002). Development of an earthquake loss model for Turkish catastrophe insurance. *Journal of Seismology, 6*, 431–446.

Bossard, M., Feranec, J., & Otahel, J. (2000). *CORINE land cover technical guide – Addendum 2000.* Copenhagen: European Environment Agency.

Bournas, A., Negro, P., & Taucer, F. (2014). Performance of industrial buildings during the Emilia earthquakes in Northern Italy and recommendations for their strengthening. *Bull Earthquake Eng*. doi:10.1007/s10518-013-9466-z

Bradley, B. (2009). Seismic Hazard Epistemic Uncertainty in the San Francisco Bay Area and its role in Performance-based Assessment. *Earthquake Spectra, 25*(4), 733-753.

Bradley, B. (2010a). A generalized conditional intensity measure approach and holistic ground-motion selection. *Earthquake Engineering and Structural Dynamics*. doi:10.1002/eqe.995

Bradley, B. (2010b). Epistemic Uncertainties in Component Fragility Functions. *Earthquake Spectra, 26*(1), 41-62.

Bradley, B. (2010c). Site-specific and spatially distributed ground-motion prediction of acceleration spectrum intensity. *Bulletin of the Seismological Society of America, 100*(2), 792–801.

Bradley, B. (2011). Empirical correlation of PGA, spectral accelerations and spectrum intensities from active shallow crustal regions. *Earthquake Engineering Structural Dynamics, 40*, 1707–1721.

Bradley, B. (2012a). A ground motion selection algorithm based on the generalized conditional intensity measure approach. *Soil Dynamics and Earthquake Engineering*. doi:10.1016/j.soildyn.2012.04.007

Bradley, B. (2012b). Empirical correlations between peak ground velocity and spectrum-based intensity measures. *Earthquake Spectra, 28*(1), 17–35.

Bradley, B. (2013). A critical examination of seismic response uncertainty analysis in earthquake engineering. *Earthquake Engineering and Structural Dynamics, 42*, 1717-1729.

Bradley, B. (2015, January 1). *Personal website—ground motion selection software*. Retrieved from https://sites.google.com/site/brendonabradley/soft-https://sites.google.com/site/brendonabradley/software/

Bradley, B., Dhakal, R., Cubrinovski, M., & MacRae, G. (2009a). Ground-motion prediction equation for SI based on spectral acceleration equations. *Bulletin of the Seismological Society of America, 85*, 277–285.

Bradley, B., Dhakal, R., Cubrinovski, M., & MacRae, G. (2009b). Prediction of spatially distributed seismic demands in structures: ground motion and structural response. *Earthquake Engineering and Structural Dynamics*. doi:10.1002/eqe.954

Calvi, G. M., Pinho, R., Magenes, G., Bommer, J., Restrepo-Veléz, L., & Crowley, H. (2006). Development of seismic vulnerability assessment methodologies over the past 30 years. *ISET Journal of Earthquake Technology, 43*(3), 75-104.

Campbell, K., & 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 10s. *Earthquake Spectra, 24*(1), 139–171.

Cauzzi, C., & Faccioli, E. (2008). Broadband (0.05 to 20 s) prediction of displacement response spectra based on worldwide digital records. *J Seismol*. doi:10.1007/s10950-008-9098-y

Chen, K., McAneney, J., Blong, R., Leigh, R., Hunter, L., & Magill, C. (2004). Defining area at risk and its effect in catastrophe loss estimation: a dasymetric mapping approach. *Appl. Geogr*. doi:10.1016/j.apgeog.2004.03.005

Chiou, B., & Youngs, R. (2008). An NGA Model for the Average Horizontal Component of Peak Ground Motion and Response Spectra. *Earthq. Spectra, 24*, 173-215.

CORINE. (2006). *Corine Land Cover 2006 seamless vector data.* Retrieved from http://www.eea.europa.eu/data-and-maps/data/clc-2006-vector-data-version-3

Cornell, A. (1968). Engineering seismic risk analysis. *Bulletin of the Seismological Society of America, 58*(5), 1583–1606.

Delavaud, E., Cotton, F., Akkar, S., Scherbaum, F., Danciu, L., Beauval, C., . . . Theodoulidis, N. (2012). Towards a ground-motion logic tree for probabilistic seismic hazard assessment in Europe. *J. of Seismology*. doi:10.1007/s10950-012-9281-z

Dell'Acqua, F., Gamba, P., & Jaiswal, K. (2013). Spatial aspects of building and population exposure data. *Nat Hazards*. doi:DOI 10.1007/s11069-012-0241-2

Denmark, S. (2015, October 1). Retrieved from https://www.dst.dk/en

Der Kiureghian, A., & Ditlevsen, O. (2008). Aleatory or epistemic? Does it matter? *Structural Safety, 31*, 105-112.

Douglas, J., Ulrich, T., & Negulescu, C. (2013). Risk-targeted seismic design maps for mainland France. *Natural Hazards*. doi:10.1007/s11069-012-0460-6.

Dumova-Jovanoska, E. (2000). Fragility curves for reinforced concrete structures in Skopje (Macedonia) region. *Soil Dynamics and Earthquake Engineering, 19*, 455-466.

EEA-ETC/TE. (2002). *CORINE land cover update. I&CLC2000 project. Technical Guidelines.* Retrieved from http://terrestrial.eionet.eu.int

Eguchi, R., Goltz, J., Taylor, C., Chang, S., Flores, P., Johnson, L., . . . Blais, N. (1998). Direct economic losses in the Northridge earthquake: a three-year post-event perspective. *Earthquake Spectra, 14*(2), 245-264.

Ellingwood, B., & Kinali, K. (2009). Quantifying and communicating uncertainty in seismic risk assessment. *Structural Safety*. doi:10.1016/j.strusafe.2008.06.001

Erberik, M. (2007). A fragility-based assessment of typical mid-rise and low-rise RC buildings in Turkey. *Engineering Structures, 30*(5), 1360–1374.

ESMD. (2015, January 1). Retrieved from European Strong Motion Database: http://www.isesd.hi.is/esd_local/Database/Database.htm

Flashearth. (2015, October 1). *Satelite imagery*. Retrieved from http://www.flashearth.com/

Geoportale. (2015, October 1). Retrieved from Emilia Romagna: http://geoportale.regione.emilia-romagna.it/en

GEOSTAT. (2011). *Population distribution dataset*. Retrieved from http://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/population-distribution-demography

Giardini, e. a. (2013). Seismic Hazard Harmonization in Europe (SHARE). *Online Data Resource*. doi:10.12686/SED- 00000001-SHARE

Girres, J.-F., & Touya, G. (2010). Quality Assessment of the French OpenStreetMap Dataset. *Transactions in GIS, 14*(4), 435–459.

Grubs, E. (1969). Procedures for detecting outlying observations in samples. *Technometrics, 11*(1), 1–21.

Haklay, M. (2010). How good is volunteered geographical information? A comparative study of openstreetmap and ordnance survey datasets. *Environment and Planning B: Planning and Design, 37*(4), 682-703. doi:10.1068/b35097

Haller, K., & Basili, R. (2011). Developing seismogenic source models based on geologic fault data. *Seis. Res. Let.* doi:10.1785/gssrl.82.4.519

Haselton, C., Baker, J., Bozorgnia, Y., Goulet, C., Kalkan, E., Luco, N., . . . Zareian, F. (2009). *Evaluation of Ground Motion Selection and Modification Methods: Predicting Median Interstory Drift Response of Buildings. PEER Report 2009-01*. Berkeley, CA: Pacific Earthquake Engineering Research Centre, University of California.

Haselton, C., Baker, J., Liel, A., Kircher, C., & Deierlein, G. (2011). Accounting for ground motion spectral shape characteristics in structural collapse assessment through an adjustment for epsilon. *Journal of Structural Engineering. SPECIAL ISSUE: Earthquake Ground Motion Selection and Modification for Nonlinear Dynamic Analysis of structures*, 322–344.

Haselton, C., Whittaker, A., Hortacsu, A., Baker, J., Gray, J., & Grant, D. (2012). Selecting and scaling earthquake ground motions for performing response-history analysis. *Proceeding of the 15th World Conference in Earthquake Engineering.* Lisbon, Portugal.

Hecht, R., Kunze, C., & Hahmann, S. (2013). Measuring Completeness of Building Footprints in OpenStreetMap over Space and Time. *ISPRS Int. J. Geo-Inf, 2*, 1066-1091. doi:10.3390/ijgi2041066

Hiemer, S., Woessner, J., Basili, R., Danciu, L., Giardini, D., & Wiemer, S. (2014). A smoothed stochastic earthquake rate model considering seismicity and fault moment release for Europe. *Geophys. J. Int.* doi:10.1093/gji/ggu186

Housner, G. (1952). Spectrum intensities of strong-motion earthquakes. *Symposium on Earthquakes and Blast Effects on Structures.* Los Angeles, CA: Earthquake Engineering Research Institute.

Huang, Y.-N., Whittaker, A., & Luco, N. (2009). Orientation of maximum spectral demand in the near-fault region. *Earthquake Spectra, 25*(3), 707–717.

Iervolino, I., Manfredi, G., & Cosenza, E. (2006). Ground Motion duration on nonlinear seismic response. *Earthquake Engineering and Structural Dynamics*. doi:10.1002/eqe.529

INE. (2015, October 1). *Portuguse Census 2011*. Retrieved from http://censos.ine.pt/

Jalayer, F., Iervolino, I., & Manfredi, G. (2010). Structural modelling uncertainties and their influence on the assessment of existing RC structures. *Structural Safety, 32*, 220-228.

Jayaram, N., & Baker, J. (2008). Statistical tests of the joint distribution of spectral acceleration values. *Bulletin of the Seismological Society of America, 98*, 2231–2243.

Jayaram, N., & Baker, J. (2009). Correlation model of spatially distributed ground motion intensities. *Earthquake Engineering and Structural Dynamics, 38*, 1687 – 1708.

Jayaram, N., & Baker, J. (2010). Ground-motion selection for PEER transportation systems research program. *Proceedings of the 7th CUEE and 5th ICEE Joint Conference.* Tokyo, Japan.

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

Justel, A., Pena, D., & Zamar, R. (1997). A multivariate kolmogorov-Smirnov test of goodness of fit. *Statistics and Probability Letters, 35*, 251-259.

Kramer, S., & Mitchell, R. (2006). Ground motion intensity measures for liquefaction hazard evaluation. *Earthquake Spectra, 22*(2), 413–438.

Liel, A. (2008). Assessing the collapse risk of California's existing reinforced concrete frame structures: Metrics for seismic safety decisions. Stanford, CA: PhD Thesis, Stanford University.

Liel, A., Haselton, C., Deierlein, G., & Baker, J. (2009). Incorporating modeling uncertainties in the assessment of seismic collapse risk of buildings. *Structural Safety, 31*, 197–211.

Lin, T., Harmsen, S., Baker, J., & Luco, N. (2013). Conditional spectrum computation incorporating multiple causal earthquakes and ground-motion prediction models. *Bulletin of the Seismological Society of America, 103*, 1103–1116.

Lin, T., Haselton, C., & Baker, J. (2013). Conditional-spectrum-based ground motion selection. Part II: intensity-based assessment and evaluation of alternative target

spectra. *Earthquake Engineering and Structural Dynamics*. doi:10.1002/eqe.2337.

Loth, C., & Baker, J. (2013). A spatial cross-correlation model of spectral accelerations at multiple periods. *Earthquake Engineering and Structural Dynamics, 42*(3), 397-417.

Luco, N. (2002). Probabilistic seismic demand analysis, connection fractures, and near-source effects. PhD Thesis. Stanford, CA: Department of Civil and Environmental Engineering, Stanford University.

Luco, N., & Cornell, A. (2007). Structure-specific scalar intensity measures for near-source and ordinary earthquake ground motions. *Earthquake Spectra, 23*(2), 357–392.

McGuire, R. K. (2004). *Seismic Hazard and Risk Analysis.* Earthquake Engineering Research Institute.

McKenna, F., Fenves, G., Scott, M., & Jeremic, B. (2000). *Open System for Earthquake Engineering Simulation (OpenSees).* Berkeley, CA: Pacific Earthquake Engineering Research Center, University of California.

Melchers, R. (1999). *Structural reliability analysis and prediction.* New York: John Wiley.

Monelli, D., Pagani, M., Weatherill, G., Silva, V., & Crowley, H. (2012). The hazard component of OpenQuake: the calculation engine of the Global Earthquake Model. *Proceedings of the 15th World Conference on Earthquake Engineering.* Lisbon, Portugal.

Morton, B., & Forsythe, A. (1974). Robust tests for equality of variances. *Journal of the American Statistical Association*. doi:10.1080/01621459.1974.10482955

Neis, P., Zielstra, D., & Zipf, A. (2012). The street network evolution of crowdsourced maps: Openstreetmap in germany 2007-2011. *Future Internet 2012, 4*, 1-21. doi:10.3390/fi4010001

NIST. (2011). *Selecting and scaling earthquake ground motion for performing response-history analyses. NIST / GCR 11-917-15.* Gaithersburd, Maryland: prepared for NEHRP Consultants Joint Venture for the National Institute of Standards and Technology.

Oliver, D. (2003). Gaussian Cosimulation: Modelling of the cross-covariance. *Mathematical Geology, 365*, 681-698.

OpenQuake. (2016, January 1). *Hazard tools*. Retrieved from https://github.com/gem/oq-hazardlib

OSM Statistics . (2015, October 17). Retrieved from http://www.openstreetmap.org/stats/data_stats.html

Pagani, M., Monelli, D., Weatherill, G., Danciu, L., Crowley, H., Silva, V., . . . Vigano, D. (2014). Openquake Engine: An Open Hazard (and Risk) Software for the

Global Earthquake Model. *Seimological Research Letters*. doi:10.1785/0220130087

Park, J., Bazzurro, P., & Baker, J. (2007). Modeling spatial correlation of ground motion intensity measures for regional seismic hazard and portfolio loss estimation. *10th International Conference on Application of Statistics and Probability in Civil Engineering (ICASP10).* Tokyo, Japan.

PEER. (2015, January 1). Retrieved from strong motion database: http://peer.berkeley.edu/smcat/

PRISE. (2013-2015). Seismic Risk Assessment for Mainland Portugal. *A research project from the University of Porto, Portugal; the University of Aveiro, Portugal; LNEC, Portugal; and EUCENTRE, Italy*.

Ramm, F., Topf, J., & Chilton, S. (2010). *OpenStreetMap: Using and Enhancing the Free Map of the World, 1st ed.* Cambridge, UK: UIT.

Ross, S. (2009). *Introduction to Probability and Statistics for Engineers and Scientists.* Elesevier Academic Press: Elesevier Academic Press.

Rossetto, T., & Elnashai, A. (2003). Derivation of vulnerability functions for European-type RC structures based on observational data. *Engineering Structures, 25*(10), 1241–1263.

Rossetto, T., & Elnashai, A. (2005). A new analytical procedure for the derivation of displacement-based vulnerability curves for populations of RC structures. *Engineering Structures, 27*, 397–409.

Sezen, H., & Whittaker, A. (2004). Performance of Industrial facilities during the 1999 Kocalei, Turkey earthquake. *13th World Conference on Earthquake Engineering* . Vancouver, Canada.

SHARE. (2015, October 1). Retrieved from seismic hazard map: http://www.share-eu.org/node/90

Shome, N., & Cornell, A. (1999). *Probabilistic seismic demand analysis of nonlinear structures. Technical Report RMS-35.* Stanford, CA: RMS Program.

Silva, V. (2013). *Development of open models and tools for seismic risk assessment: Application to Portugal. PhD Thesis.* Aveiro, Portugal: Depertment of Civil Engineering of the University of Aveiro.

Silva, V. (2016). Critical Issues in Earthquake Scenario Loss Modelling. *J. Earthquake Eng.* doi:10.1080/13632469.2016.1138172

Silva, V., Crowley, H., & Bazzurro, P. (2015a). Exploring Risk-Targeted Hazard Maps for Europe. *Earthquake Spectra*. doi:http://dx.doi.org/10.1193/112514EQS198M

Silva, V., Crowley, H., Pagani, M., Monelli, D., & Pinho, R. (2014). Development of the OpenQuake engine, the Global Earthquake Model's open-source software for seismic risk assessment. *Natural Hazards*. doi:10.1007/s11069-013-0618-x

Silva, V., Crowley, H., Pinho, R., & Varum, H. (2013). Extending displacement-based earthquake loss assessment (DBELA) for the computation of fragility curves. *Engineering Structures, 56*, 343-356.

Silva, V., Crowley, H., Pinho, R., & Varum, H. (2015b). Seismic risk assessment for mainland Portugal. *Bull. Earthquake Eng, 13*, 429–457.

Silva, V., Crowley, H., Pinho, R., Varum, H., & Sousa, L. (2015c). Investigation of the characteristics of Portuguese regular moment resisting frame RC buildings and development of a vulnerability model. *Bulletin of Earthquake Engineering, 13*, 1455-149.

Silva, V., Crowley, H., Pinho, R., Varum, H., & Sousa, R. (2014). Evaluation of analytical methodologies used to derive vulnerability functions. *Earthquake Engineering and Structural Dynamics, 43*, 181–204.

Sousa, L., & Campos Costa, A. (2009). Ground motion scenarios consistent with probabilistic seismic hazard disaggregation analysis. Application to Mainland Portugal. *Bulletin of Earthquake Engineering, 7*, 127–147.

Sousa, L., Silva, V., Marques, M., Crowley, H., & Pinho, R. (2014). Including multiple IMTs in the development of fragility functions for earthquake loss estimation. *Proceedings of the 2nd International Conference on Vulnerability and Risk Analysis and Management.* Liverpool, UK.

Steenmans, C., & Perdigao, V. (2001). Update of the CORINE land cover database. In G. Groom, & T. Reed (Eds.), *Strategic Landscape Monitoring for the Nordic Countries* (pp. 101-107). Copenhagen: Nordic Council of Ministers.

Strasser, F., Bommer, J., & Abrahamson, N. (2008). Truncation of the distribution of ground motion residuals. *Journal of Seismology*. doi:10.1007/s10950-007-9073-z

Surajit, R., & Lindsay, B. (2005). The topography of multivariate normal mixtures. *The Annals of Statistics*. doi:10.1214/009053605000000417

Taylor, P. (2015). Calculating Financial Loss from Catastrophes. *Proceedings of the SECED Conference.* Cambridge, UK.

Toro, G. (2003). *Modification of the Toro et al. (1997) Attenuation Equations for Large Magnitudes and Short Distances, Technical Report.* Louisville, CO: Risk Engineering, Inc.

Tothong, P., & Luco, N. (2007). Probabilistic seismic demand analysis using advanced ground motion intensity measures. *Earthquake Engineering and Structural Dynamics, 36*(13), 1837–1860.

Vick, M. (2002). *Degrees of Belief: Subjective Probability and Engineering Judgement.* Reston, VA: ASCE Press.

Vilanova, S., & Fonseca, J. (2007). Probabilistic seismic-hazard assessment for Portugal. *Bulletin of the Seismological Society of America, 97*, 1702–1717.

Vilanova, S., Fonseca, J., & Oliveira, C. (2007). Ground-motion models for seismic-hazard assessment in western Iberia: constraints from instrumental data and intensity observations. *Bulletin of the Seismological Society of America*, 1702–1717. doi:10.1785/012005019

Von Thun, J., Roehm, L., Scott, G., & Wilson, J. (1988). Earthquake ground motions for design and analysis of dams. *Earthquake Engineering and Soil Dynamics II—Recent Advances in Ground-Motion Evaluation, Geotechnical Special Publication, 20,* 463–481.

Wasserman, L. (2004). *All of Statistics: A Concise Course on Statistical Inference.* New York: Springer.

Weatherill, G., Silva, V., Crowley, H., & Bazzurro, P. (2015). Exploring the impact of spatial correlations and uncertainties for portfolio analysis in probabilistic seismic loss estimation. *Bul. of Earthquake Engineering, 3*(4), 957-981.

Woessner, J., Danciu, L., Giardini, D., Crowley, H., Cotton, F., Grünthal, G., . . . Stucchi, M. (2015). The 2013 European Seismic Hazard Model: key components and results. *Bull. Earthq. Eng.* doi:10.1007/s10518-015-9795-1.

Youssef, N., Bonowitz, D., & Gross, J. (1995). *A Survey of Steel Moment Resisting Frame Buildings Affected by the 1994 Northridge Earthquake.* Gaithersburg, MD: NISTR-5625, NIST.

Zhao, J., Zhang, J., Asano, A., Ohno, Y., Oouchi, T., Takahashi, T., . . . Fukushima, Y. (2006). Attenuation Relations of Strong Ground Motion in Japan Using Site Classification Based on Predominant Period. *Bulletin of the Seismological Society of America, 96*(3), 898-913.