
Efficient Treatment and Quantification of Uncertainty in Probabilistic Seismic Hazard and Risk Analysis

Christoph Scheingraber



München 2018

Efficient Treatment and Quantification of Uncertainty in Probabilistic Seismic Hazard and Risk Analysis

Christoph Scheingraber

Dissertation
an der Fakultät für Geowissenschaften der
Ludwig–Maximilians–Universität
München

vorgelegt von
Christoph Scheingraber

München, den 11.09.2018

Erstgutachter: Prof. Dr. Heiner Igel

Zweitgutachter: PD Dr. habil. Martin Käser

Tag der mündlichen Prüfung: 06.02.2019

Contents

Summary	xix
1 Introduction and Motivation	1
1.1 Introduction	1
1.2 Objectives and Outline	4
2 Probabilistic Seismic Hazard and Risk Analysis	7
2.1 Probabilistic Seismic Hazard Analysis	7
2.1.1 Historical Background	7
2.1.2 Classical Integration Approach	9
2.1.3 Event-Based Simulation Approach	13
2.1.4 Ground Motion Modeling	15
2.2 Probabilistic Seismic Risk Analysis	17
2.2.1 Historical Background	17
2.2.2 Components of Seismic Risk Analysis	18
2.2.3 Exposure Models	19
2.2.4 Vulnerability Models	21
3 A Framework for the Treatment of Ground Motion and Damage Uncertainty in Probabilistic Seismic Hazard and Risk Analysis	25
3.1 Introduction	25
3.2 Monte Carlo Simulation and Random Number Generation	26
3.2.1 Monte Carlo Simulation	26
3.2.2 Uniform Random Number Generators	28
3.2.3 Simulating Nonuniform Random Variables	30
3.3 Framework for Repeatable Stochastic Simulation for Probabilistic Seismic Risk Analysis on Distributed Systems	31
3.3.1 Computational Infrastructure	31
3.3.2 Selection of PRNG Substreams using Hash Functions	32
3.4 Framework for Stochastic Simulation of Ground Motion Uncertainty	34
3.4.1 Uncorrelated Ground Motion Uncertainty	34
3.4.2 Spatially Correlated Ground Motion Uncertainty	36
3.4.3 Spatially Cross-Correlated Ground Motion Uncertainty	39

3.5	Damage Uncertainty Framework	42
3.5.1	Zero-One-Inflated Distributions	42
3.5.2	Kumaraswamy Distribution	43
3.5.3	Zero-One-Inflated Kumaraswamy Distribution	44
3.5.4	Damage Correlation	45
4	The Impact of Portfolio Location Uncertainty in Probabilistic Seismic Risk Analysis	47
4.1	Introduction	48
4.2	Background: Seismic Hazard and Risk Analysis	50
4.3	Proposed Framework for Treatment of Location Uncertainty	51
4.3.1	Problem Statement	51
4.3.2	Sampling Location Uncertainty on a Weighted Irregular Grid	51
4.3.3	Shared Hazard Computation on Unique Risk Locations	51
4.3.4	Empirical Probable Maximum Loss Distributions	53
4.4	Modeling Seismic Hazard in Western Indonesia	53
4.4.1	Study Area	53
4.4.2	Seismo-Tectonic Setting	54
4.4.3	Modeling the Seismic Hazard	54
4.5	Case Study: Impact of Location Uncertainty in Western Indonesia	56
4.5.1	Creation of Synthetic Portfolios	56
4.6	Impact of Location Uncertainty on Probabilistic Seismic Loss	60
4.6.1	Flat Hazard Model: Influence of Spatial Risk Item Clustering and Loss Aggregation	60
4.6.2	Western Indonesia Controlled Example: Upper Bound of the Isolated Influence of Location Uncertainty	62
4.6.3	Western Indonesia Realistic Example: The Influence of Location Uncertainty in Conjunction with Aleatory Ground Motion Uncertainty	67
4.7	Conclusions	70
5	Spatial Seismic Hazard Variation and Adaptive Sampling of Portfolio Location Uncertainty in Probabilistic Seismic Risk Analysis	73
5.1	Introduction	74
5.2	Background	75
5.2.1	Probabilistic Seismic Hazard and Risk Analysis	75
5.2.2	Portfolio Location Uncertainty	76
5.2.3	Evaluation of the Proposed Sampling Scheme	77
5.2.4	Generation of Synthetic Portfolios	79
5.3	Case Study: Spatial Seismic Hazard Variation in Western Indonesia	80
5.3.1	Hazard Model	80
5.3.2	Spatial Seismic Hazard Variation	81
5.3.3	Loss Rate Variation	84
5.4	A Framework for Adaptive Sampling of Portfolio Location Uncertainty	84

5.4.1	Risk Location Index Mapping Table	85
5.4.2	Criterion I: Coefficient of Variation of Loss Rate	85
5.4.3	Criterion II: Number of Risk Items	87
5.4.4	Criterion III: Value Distribution	88
5.4.5	Combination of Criteria	89
5.5	Results	90
5.5.1	Spatial Variation Parameter Study	90
5.5.2	Performance of the Final Implementation	92
5.6	Conclusions	96
6	Uncertainty Quantification for Seismic Risk Assessment using Latin Hypercube Sampling and Quasi Monte Carlo Simulation	97
6.1	Introduction	98
6.2	Methodology	99
6.2.1	Modeling Framework and Uncertainty Treatment	99
6.2.2	Sampling Strategies and Error Estimation	101
6.2.3	Sensitivity Analysis	103
6.2.4	Earthquake Scenarios and Portfolios	105
6.3	Results	108
6.3.1	Error Convergence of Sampling Strategies	108
6.3.2	Results of Global Sensitivity Analysis	111
6.4	Conclusions	113
7	Conclusion and Outlook	117
Appendix A OpenQuake Model Explorer: Visualization and Conversion of Earthquake Models in NRML format		
A.1	Introduction	143
A.2	Graphical User Interface	144
A.2.1	Visualization of Sources and Seismicity	144
A.2.2	Extraction of Geographical Subsets	146
A.2.3	Converting and Saving a Source Model	146
A.3	Usage of OpenQuake Model Explorer from a Python Script	146
Appendix B Source Zone Creator: Graphical Tool for Seismic Hazard Model Development		
B.1	Introduction	149
B.2	Graphical User Interface	149
B.2.1	Source Model and Catalogue Visualization	150
B.2.2	Creating and Modifying the Geometry of Seismic Source Zones	151
B.2.3	Obtaining the Final Model Variables	151
Acknowledgements		153

List of Figures

1.1	Overall and Insured Losses from Natural Catastrophe Events 1980–2017	1
1.2	World Map of Natural Catastrophes in 2011	2
1.3	Distribution of Losses among different Peril Types in 2011	3
1.4	Distribution of Losses among different Peril Types in 2017	3
2.1	Workflow of Classical PSHA	11
2.2	Workflow of Event-Based PSHA	14
2.3	Components of Risk Analysis	19
2.4	Aggregated Exposure Model	20
2.5	Empirical Vulnerability Curves	23
3.1	Mean and Uncorrelated Ground Motion Uncertainty	35
3.2	Footprint with Spatially Correlated Ground Motion Uncertainty	38
3.3	Footprints with Cross-Correlated Ground Motion Uncertainty using the Linear Model of Coregionalization	40
3.4	Footprints with Full Block Cross-Correlated Ground Motion Uncertainty	41
3.5	Different Shapes of the Kumaraswamy Distribution	43
3.6	Probability Density Function of the Zero-One-Inflated Kumaraswamy Distribution	44
4.1	Variable Resolution Grid	52
4.2	Population Density in Indonesia	54
4.3	Tectonic Plates in Indonesia	55
4.4	Seismic Hazard in Western Indonesia	57
4.5	Insured Values of Risk Items for the Mixed Exponential Value Distribution	59
4.6	PML Curve for a Synthetic Portfolio of 100 Risk Items With 80% Unknown Coordinates, using a Flat Value Distribution and a Flat Hazard Model	61
4.7	PML Curve for a Synthetic Portfolio of 300 Risk Items With 20% Unknown Coordinates, using a Flat Value Distribution and the Indonesia Hazard Model	63
4.8	PML Curve for a Synthetic Portfolio of 300 Risk Items With 20% Unknown Coordinates, using a Mixed Exponential Value Distribution and the Indonesia Hazard Model	64

4.9	Comparison of the Reference PML Curve for the Original Portfolio With the Portfolio With 20% Unknown Coordinates	65
4.10	Kernel Density Estimation of PML at a Return Period of 10 Years	66
4.11	Coefficient of Variation of PML at a Return Period of 10 Years for Synthetic Portfolios	68
4.12	Coefficient of Variation of PML at a Return Period of 10 years for Synthetic Portfolios with Aleatory Ground Motion Uncertainty	69
5.1	Weighted Grid Used as an Insurance Density Proxy	76
5.2	Coefficient of Variation of Peak Ground Acceleration with an Exceedance Probability of 10% in 50 Years Inside Provinces in Western Indonesia	80
5.3	Coefficient of Variation of Peak Ground Acceleration with an Exceedance Probability of 10% in 50 Years Inside Regencies and Cities in Western Indonesia	81
5.4	Coefficient of Variation of Ground Motion Predicted to be Exceeded at Various Return Periods for the <i>Jambi</i> Province	82
5.5	Coefficient of Variation of Ground Motion Predicted to be Exceeded at Various Return Periods for the <i>Kalimantan Timur</i> Province	83
5.6	Coefficient of Variation of the Loss Rate Inside Provinces in Western Indonesia	84
5.7	Location Sampling Scheme Criterion I	86
5.8	Location Sampling Scheme Criterion II.a	87
5.9	Location Sampling Scheme Criterion II.b	88
5.10	Location Sampling Scheme Criterion III	89
5.11	Results of the Systematic Parameter Study	91
5.12	Convergence Plots Showing Relative Probable Maximum Loss at a Return Period of 100 Years Against the Number of Used Hazard Locations for Portfolios of 10 Risk Items	92
5.13	Convergence Plots Showing Relative Probable Maximum Loss at a Return Period of 100 Years Against the Number of Used Hazard Locations for Portfolios of 100 Risk Items	93
5.14	Relative Standard Errors of Probable Maximum Loss at a Return Period of 100 Years Against the Total Number of Used Hazard Locations	94
6.1	Weighted Irregular Grid and Sunda Straight	100
6.2	Footprint of Southern Sumatra Subduction Fault Event	106
6.3	Footprint of Northern Sumatra Fault Zone Event	107
6.4	Event Loss Variation for Different Sampling Schemes	109
6.5	Variance Reduction Results for Medan Portfolios	110
6.6	Adaptive Variance Reduction Results for Jakarta Portfolios	111
6.7	Scatter Plot for Sumatra Subduction Zone Event	112
6.8	Sobol Indices for Sumatra Subduction Zone Event	113
6.9	Sobol Indices for Sumatra Fault Zone Event	114

A.1	OpenQuake Model Explorer Main Window	144
A.2	OpenQuake Model Explorer Seismic Source Popup Window	145
A.3	Plot of B-Values in Indonesia	148
B.1	Source Zone Creator Main Window with B-Values	150
B.2	Source Zone Creator Main Window with Smoothed Seismicity from Catalogue	151
B.3	Histogram of Hypocentral Depths	152

List of Tables

2.1	Example of Exposure Model	21
2.2	Typical Natural Resonance Periods of Building Rise Classes	22
3.1	Supported Ground Motion Correlation Models	34
4.1	Risk Location Index Mapping Table	53
4.2	Ground Motion Model Logic Tree	56
5.1	Risk Location Index Mapping Table	85
5.2	Mean Runtime Speedup and Standard Errors	95

List of Abbreviations

ANOVA	Analysis Of Variances
API	Application Programming Interface
CDF	Cumulative Distribution Function
CMRG	Combined Multiple Recursive Generator
CPU	Central Processing Unit
CV	Coefficient of Variation
DPM	Damage Probability Matrix
DR	Damage Ratio
ERF	Earthquake Rupture Forecast
EU	European Union
FBCC	Full Block Cross-Correlation
FEMA	Federal Emergency Management Agency
GEM	Global Earthquake Model
GMPE	Ground Motion Prediction Equation
GNP	Gross National Product

GUI	Graphical User Interface
HRNG	Hardware Random Number Generator
IID	Independent and Identically Distributed
IPC	Inter-Process Communication
LCG	Linear Congruential Generator
LFSR	Linear Feedback Shift Register
LHS	Latin Hypercube Sampling
LMCR	Linear Model of Co-Regionalization
LSS	Location Sampling Scheme
MC	Monte Carlo
MD5	Message-Digest Algorithm 5
MDR	Mean Damage Ratio
MFD	Magnitude-Frequency Distribution
MMI	Maximum Mercalli Intensity
MRG	Multiple Recursive Generator
NatCat	Natural Catastrophe
NGA	Next Generation of Ground-Motion Attenuation Models
NRML	Natural hazards' Risk Markup Language
OECD	Organization for Economic Co-operation and Development

PDF	Probability Density Function
PGA	Peak Ground Acceleration
PGD	Peak Ground Displacement
PGV	Peak Ground Velocity
PML	Probable Maximum Loss
PRNG	Pseudo Random Number Generator
PSA	Peak Spectral Acceleration
PSHA	Probabilistic Seismic Hazard Analysis
PSRA	Probabilistic Seismic Risk Analysis
QMC	Quasi Monte Carlo
RNG	Random Number Generator
RORAC	Return on Risk-Adjusted Capital
RSE	Relative Standard Error
SA	Sensitivity Analysis
SCR	Solvency Capital Requirements
SDOF	Single Degree Of Freedom
SE	Standard Error
SEM	Standard Error of the Mean
SES	Stochastic Event Set

SFZ	Sumatra Fault Zone
SPD	Symmetric Positive Definite
SSHAC	Senior Seismic Hazard Committee
TOM	Temporal Occurrence Model
TSI	Total Sum Insured
UNAM	Universidad Nacional Autonoma de Mexico
US	United States
USGS	United States Geological Service
VI	Value Insured
VR	Variance Reduction

Summary

The main goals of this thesis are the development of a computationally efficient framework for stochastic treatment of various important uncertainties in probabilistic seismic hazard and risk assessment, its application to a newly created seismic risk model of Indonesia, and the analysis and quantification of the impact of these uncertainties on the distribution of estimated seismic losses for a large number of synthetic portfolios modeled after real-world counterparts. The treatment and quantification of uncertainty in probabilistic seismic hazard and risk analysis has already been identified as an area that could benefit from increased research attention. Furthermore, it has become evident that the lack of research considering the development and application of suitable sampling schemes to increase the computational efficiency of the stochastic simulation represents a bottleneck for applications where model runtime is an important factor.

In this research study, the development and state of the art of probabilistic seismic hazard and risk analysis is first reviewed and opportunities for improved treatment of uncertainties are identified. A newly developed framework for the stochastic treatment of portfolio location uncertainty as well as ground motion and damage uncertainty is presented. The framework is then optimized with respect to computational efficiency. Amongst other techniques, a novel variance reduction scheme for portfolio location uncertainty is developed. Furthermore, in this thesis, some well-known variance reduction schemes such as Quasi Monte Carlo, Latin Hypercube Sampling and MISER (locally adaptive recursive stratified sampling) are applied for the first time to seismic hazard and risk assessment. The effectiveness and applicability of all used schemes is analyzed. Several chapters of this monograph describe the theory, implementation and some exemplary applications of the framework. To conduct these exemplary applications, a seismic hazard model for Indonesia was developed and used for the analysis and quantification of loss uncertainty for a large collection of synthetic portfolios.

As part of this work, the new framework was integrated into a probabilistic seismic hazard and risk assessment software suite developed and used by Munich Reinsurance Group. Furthermore, those parts of the framework that deal with location and damage uncertainties are also used by the flood and storm natural catastrophe model development groups at Munich Reinsurance for their risk models.

Chapter 1

Introduction and Motivation

1.1 Introduction

In recent decades, we have experienced a strong increase of the adverse effects of natural catastrophe events. This can also be seen in the upwards tendency of suffered losses (see Figure 1.1). Among other reasons, this can be explained by a growing world population, continuing urbanization, and high vulnerability of present-day technology and societies (Smolka et al., 2004). Almost every part of the world is affected (see Figure 1.2), which has raised global awareness to the urgent need to quantify the associated risk in a probabilistic framework.

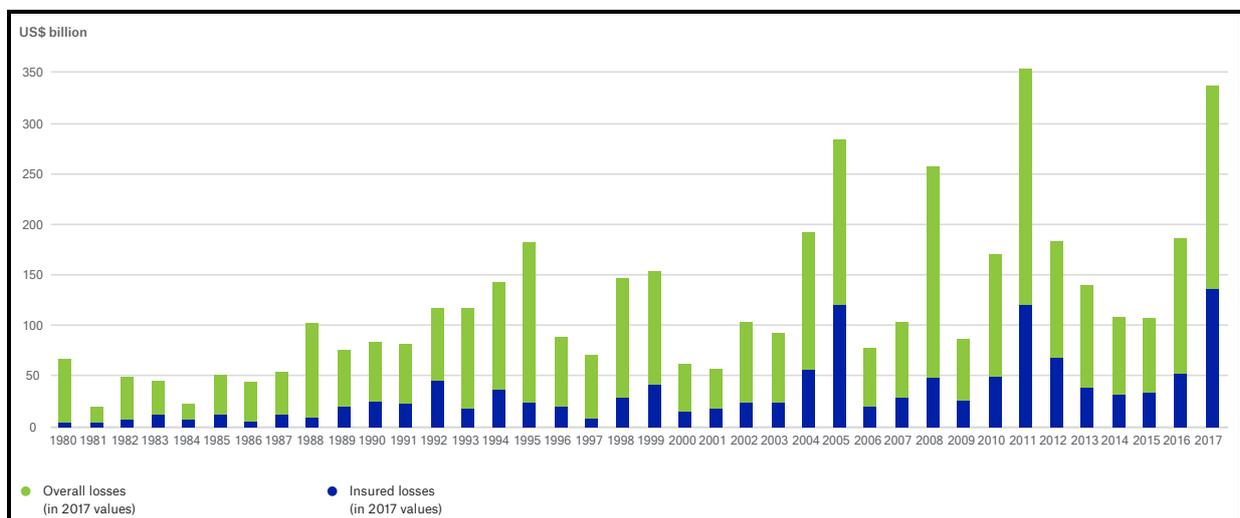


Figure 1.1: Overall and insured losses from relevant natural catastrophe events 1980 to 2017, normalized to 2017 US Dollars by taking consumer price index and exchange rate fluctuations into account. Noteworthy is the strong annual fluctuation of losses and a relatively low insurance penetration. From Munich Re NatCatSERVICE (2018b).

As of this writing, earthquakes and their secondary effects represented 6 of the 10 costliest natural catastrophe events by overall losses (Munich Re NatCatSERVICE, 2018c). This portion is smaller if measured by insured losses, due to the higher insurance penetration for other perils. The distribution of losses among different peril types can fluctuate significantly, due to accumulation of events for a peril. Earthquake losses in particular are often driven by relatively few but costly events. For example, in 2011 geophysical events represented 52.7% of overall losses (see Figure 1.3), but only 3.0% in 2017 (see Figure 1.4).

The 2011 Tōhoku earthquake and tsunami has been the costliest natural catastrophe event in terms of losses to the overall economy (~210 billion US dollars), and second costliest in terms of insured losses (~40 billion US dollars). In less developed countries, the absolute monetary value of losses caused by natural catastrophe events might be lower due to a smaller gross national product (GNP), but the impact on the national economy can be even larger. For example, the three largest losses from 1972 to 1990 caused by earthquakes as a fraction of GNP occurred in the Central American countries of Nicaragua (1972, 40% GNP), Guatemala (1976, 18% GNP) and El Salvador (1986, 31% GNP) (Calvi et al., 2006; Coburn and Spence, 2002).

Munich Re Group is a reinsurance company based in Munich, Germany. As “one of

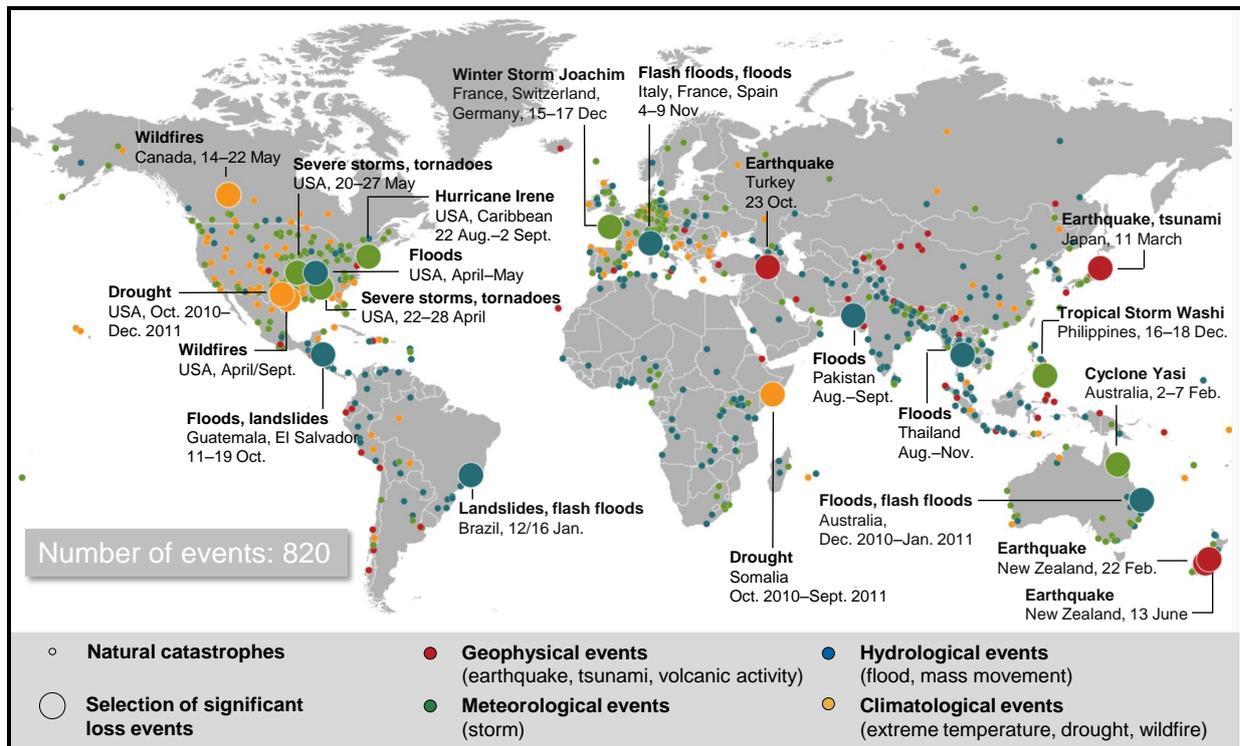


Figure 1.2: World Map of Natural Catastrophes in 2011. From Munich Re NatCatSERVICE (2018a).

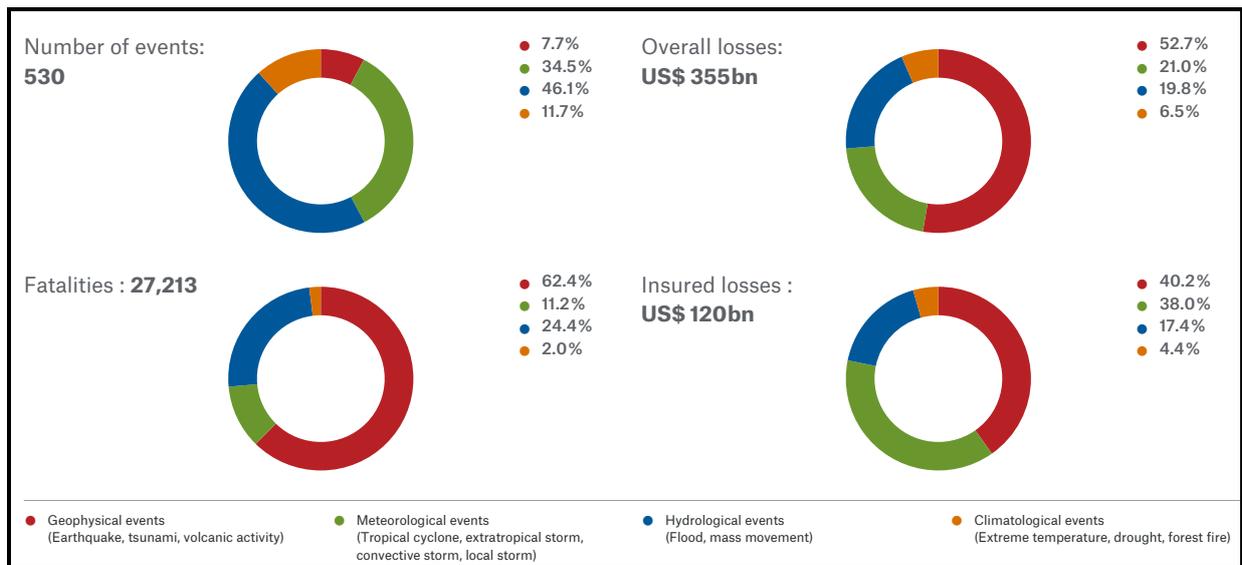


Figure 1.3: Distribution of losses among different peril types in 2011. Geophysical events represent only a small fraction in terms of number of loss events, but account for a large fraction of monetary loss. From Munich Re NatCatSERVICE (2018a).

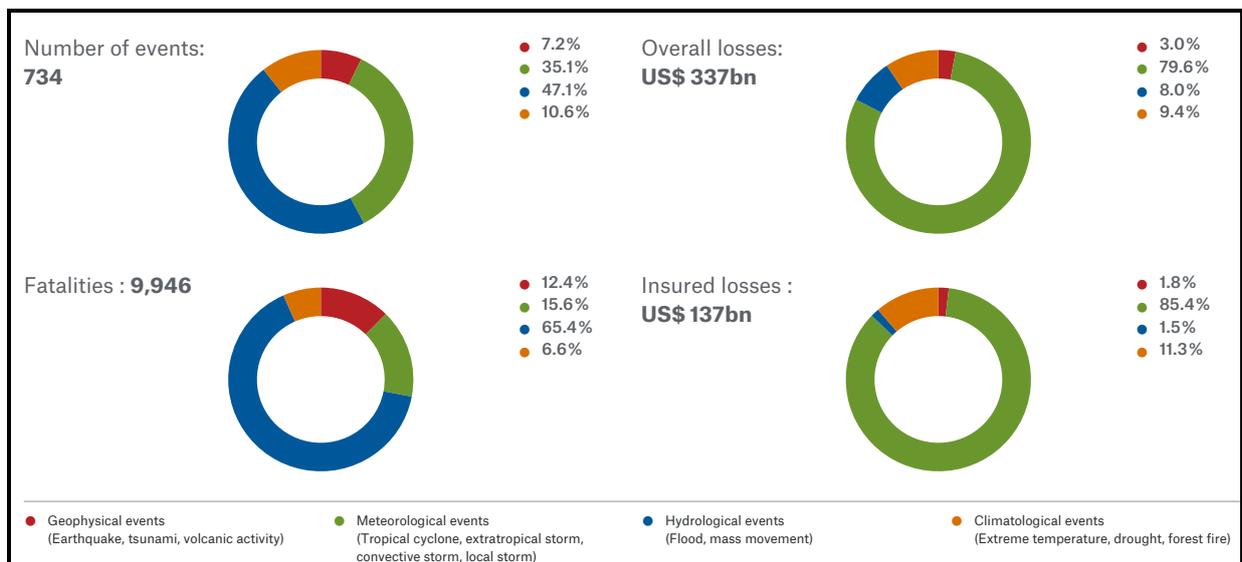


Figure 1.4: Distribution of losses among different peril types in 2017. Like most years, meteorological events account for the majority of losses. From Munich Re NatCatSERVICE (2018a).

the world's leading risk carriers"¹, in 2017 Munich Re generated a revenue of about 50 billion euros (Munich Re, 2018), mostly from insurance premiums. In the reinsurance industry, solvency capital needs to be reserved to reduce the risk of insolvency. *Solvency II* (European Commission, 2009) is a directive in the European Union and its implementation became mandatory for member states on January 1st, 2016. Using an internal risk model, insurance companies model an economic profit and loss distribution over all fields of business. The solvency capital requirements (SCR) require to hold capital accounting for the estimated value at risk over a one-year time horizon of this distribution with a confidence level of 99.5%. In 2017, the SCR for Munich Re Group (i.e., reinsurance and primary insurance divisions) excluding tax and diversification effects was about 29 billion euros. The property-casualty field of business amounted for 6.6 billion euros of the group SCR, and it is noteworthy that natural catastrophe scenarios account for about two thirds of this. To be conservative and get a high grade rating, Munich Re multiplies the required minimum SCR by a factor of 1.75 to obtain the actually maintained risk capital.

Munich Re set itself a long-term target of 15% for the return on risk-adjusted capital (RORAC), which is a mixture of accounting ratios and economic indicators. This target was set in times when premiums had higher margins, and is hard to reach in the current low interest-rate market environment (Munich Re, 2015). In 2016, Munich Re reached a RORAC of 11.5%. Mostly due to an accumulation of unusually costly meteorological events, in 2017 Munich Re remained far from this goal, only achieving a RORAC of 1.5%.

More than ever, continuing pressure on premiums results in the necessity to make careful and educated underwriting decisions, including taking the uncertainty of the associated risk model into account.

1.2 Objectives and Outline

Probabilistic seismic hazard and risk analysis involves large uncertainties of various kinds, the treatment of which can be computationally intensive. While this has already been identified as an area that could benefit from increased research attention, so far the amount of research has been limited. This study is concerned with the development of an efficient framework for the stochastic treatment of portfolio location uncertainty as well as ground motion and damage uncertainty. The efficiency of the framework and the impact of the different uncertainty types on the results of seismic risk analysis are analyzed using a seismic risk model of Indonesia and a large number of synthetic but realistic portfolios.

This monograph aims to be self-supporting, in the sense that it can be read without the need to consult further literature on the side. All necessary theory will be presented and outlined in a way that supports a profound conceptual understanding. However, it would be beyond the scope of this thesis to rigorously show every step of the derivation of every part of the theory. Numerous references are provided as suggestions for further reading.

¹<http://www.munichre.com>

Structure

The central chapters of this thesis can be summarized as follows:

Chapter 2 reviews the historical and theoretical background as well as the current state of the art of probabilistic seismic hazard and risk assessment.

Chapter 3 describes the theory underlying the treatment of ground motion and damage uncertainty in the context of probabilistic seismic hazard and risk analysis and presents a framework for the simulation of these uncertainties developed in the course of this thesis.

Chapter 4 presents a new framework for the treatment of portfolio location uncertainty and analyzes the impact on probabilistic seismic risk analysis.

Chapter 5 introduces and analyzes an efficient sampling scheme specifically designed for the treatment of portfolio location uncertainty as presented in the previous chapter.

Chapter 6 investigates the applicability of several well known variance reduction techniques to seismic hazard and risk analysis and shows the results of a global sensitivity analysis for a seismic risk model.

Chapter 7 presents a conclusion and outlook for future work.

Appendix

A certain part of this thesis is concerned with the development of new auxiliary software, i.e. tools to facilitate seismic hazard model development. While this represents an important building block supporting the preparation of this thesis and should arguably be included here, not all of it aligns with science in the traditional sense and therefore the corresponding chapters have been moved to the appendix. They describe new functionalities and graphical tools that were developed to enable and facilitate the creation of the Indonesian hazard model described in Chapter 4 as well as other models.

Chapter A describes *OpenQuake Model Explorer*, a graphical tool and modular framework for hazard model visualization and conversion.

Chapter B describes *Source Zone Creator*, a graphical tool for seismic area-source zonation based on analysis of an event catalogue in conjunction with a source model converted using the OpenQuake Model Explorer framework.

Chapter 2

Probabilistic Seismic Hazard and Risk Analysis

This chapter reviews the historical and theoretical background as well as the current state of the art of probabilistic seismic hazard and risk assessment. It aims to provide a solid understanding of the most important foundations but can not describe all intricacies. For a deeper understanding, the reader is encouraged to consult comprehensive learning material on the side, such as the classic textbook *Earthquake Hazard Analysis* by Reiter (1990) or the more recent monograph *Seismic Hazard and Risk Analysis* by McGuire (2004).

2.1 Probabilistic Seismic Hazard Analysis

Probabilistic Seismic Hazard Analysis (PSHA) is a widely used framework to quantify earthquake hazard. It is a method to compute the probability of exceeding specific ground motion levels at a site which can be integrated well into a risk analysis framework.

The concept of classical PSHA is introduced in Section 2.1.2. Classical PSHA is a deterministic method and an important foundation for understanding any work in seismic hazard assessment.

Section 2.1.3 introduces event-based PSHA, which is the stochastic approach for which the new uncertainty framework introduced in this thesis has been developed.

2.1.1 Historical Background

The initial development of **classical PSHA** can be traced back to separate efforts by two researchers in the **early 1960s**, which were brought together in 1966 (McGuire, 2008). In his doctoral dissertation at Stanford University, Allin Cornell studied factors influencing earthquake engineering design decisions (Cornell, 1964). He was using the concept of derived distributions in which the probability distribution of a dependent variable is derived using its relationship to predictor variables and their respective distributions.

At the same time, PhD student Luis Esteva and Prof. Emilio Rosenblueth at the Universidad Nacional Autonoma de Mexico (UNAM) studied the dependence of earthquake ground motions on magnitude and distance, the frequency of earthquakes and the frequency of ground motions at a site using Bayesian updating. This led to the publication of the **first seismic zone maps** (Esteva, 1963). They were also relating probability of structural failure to ground motion.

In the same decade, Cornell (1968) published the **mathematical derivation** of classical PSHA in a seminal paper. Several aspects were influenced by his exchange with the UNAM group, such as the use of the Gutenberg-Richter magnitude-frequency distribution (Gutenberg and Richter, 1942, 1944) and the functional form of the chosen ground motion distribution, which had been published previously by Esteva and Rosenblueth (1964).

Esteva focused more on the engineering design process (Esteva, 1967, 1968). Esteva was also the first to use the term *aleatory uncertainty* to describe natural variability of ground motion due to processes about which we can not gain knowledge. This is in contrast to *epistemic uncertainty*, which one could theoretically treat correctly but fails in practice due to insufficient information. While the distinction of epistemic and aleatory uncertainty is somewhat arbitrary (Kiureghian and Ditlevsen, 2009), it can be helpful in practice.

To assess the earthquake related ground motions, most authors at that time used attenuation functions of the same functional form, and just reported different values for coefficients. Some reported a (fixed) standard deviation, while others did not report a standard deviation at all (Douglas, 2018). In early classical PSHA codes, such as a noteworthy implementation called *EQRISK* by McGuire (1976) a fixed form of the ground motion attenuation function (Equation 2.8) and a constant standard deviation were assumed.

With increasing computational resources, studies that employed **event-based PSHA** started to become more common in the **1980s** (e.g. Rosenhauer, 1983; Shapira, 1983). This approach evaluates the tail risk by computing individual event realizations for a finite event set. The event set is usually generated stochastically and the method became known as stochastic modeling. This approach provides the advantage that it is applicable to seismic risk analysis. For each event, a vulnerability function can be evaluated to obtain a damage or loss measure. In classical PSHA, this would have required integration of the vulnerability function, which is a method that was hardly ever used (Musson, 2000). This new approach also allowed a new technique called **deaggregation**¹ to identify the contribution of individual events to the hazard. The study of ground motion variability lead to the development of **correlation models** for the ground motion residual (e.g. Jayaram and Baker, 2009; Goda and Atkinson, 2010), which are of interest for loss analysis of portfolios, in particular when these are spatially clustered. Lately, spectral cross-correlation has been applied to co-simulate ground motion fields at various periods (e.g. Baker and

¹In this work, the term deaggregation is adopted in favor of disaggregation for consistency with other terms commonly used in the field of seismology, such as deconvolution.

Cornell, 2006; Loth and Baker, 2012), which is of interest for heterogeneous portfolios where ground motion influences damage at multiple natural frequencies (Weatherill et al., 2015). The application of these correlation models results in more realistic patterns in the event footprint.

Defining a parametric model for the standard deviation of ground motion residuals did not become customary until the year **2008** with the publication of *Next Generation of Ground-Motion Attenuation Models* (NGA) in a special issue of Earthquake Spectra (Power et al., 2008). In addition to providing parametric models for standard deviation, the authors developed a new set of sophisticated models for median ground motion, including more explanatory variables for source and site effects. For **source effects**, footwall and hanging wall terms, style-of-faulting terms, and rupture area terms were included. Soil and basin amplification terms were developed to treat **site effects**, which has been found to significantly improve PSHA (Field and SCEC Phase III Working Group, 2000). In general, these efforts helped to identify some of the aleatory ground motion uncertainty and move it into the realm of epistemic uncertainty.

2.1.2 Classical Integration Approach

Methodology

Let y be a ground motion intensity scale parameter, a physical ground motion parameter, or a response function of a system exposed to ground motion, such as Maximum Mercalli intensity, peak ground acceleration, or peak spectral acceleration, respectively (see Section 2.1.4).

The idea of PSHA is to derive the probability $P(y_0, \mathbf{r}_0, t_0)$ of exceeding some specific ground motion level y_0 within a time interval t_0 at a site location \mathbf{r}_0 . PSHA presumes that there exists a mean occurrence rate curve $\lambda = \lambda(y_0, \mathbf{r}_0)$ at which ground motion level y_0 is exceeded (Anderson and Brune, 1999).

Assuming that earthquakes are random events occurring independently of time leads to the Poisson temporal occurrence model. The probability of observing exactly k random events, i.e. observing exactly k earthquakes exceeding ground motion level y_0 in time interval t_0 is then given by the probability mass function

$$p(k, t_0) = \frac{(\lambda t_0)^k e^{-\lambda t_0}}{k!}. \quad (2.1)$$

The probability of exceeding ground motion level y_0 at least once within the time interval t_0 is then given by the probability of more than zero events, i.e. by evaluating the complementary cumulative distribution function \bar{P} at $k = 0$:

$$\begin{aligned} P(y_0, t_0, \lambda)[y \geq y_0] &= p(k > 0) = \bar{P}(k = 0) = 1 - p(k = 0) \\ &= 1 - \frac{(\lambda t_0)^0 e^{-\lambda t_0}}{0!} = 1 - e^{-\lambda t_0}. \end{aligned} \quad (2.2)$$

To find the mean seismic occurrence rate $\lambda(y_0, \mathbf{r}_0)$, the total probability theorem is invoked. The probability of exceeding ground motion y_0 at a site \mathbf{r}_0 due to the contribution of the i th source having a rupture of magnitude m at distance \mathbf{r} can then be expressed by a multiplication of the conditional probability $P[y \geq y_0|m, \mathbf{r}, \mathbf{r}_0]$ of exceeding ground motion y_0 at site \mathbf{r}_0 given a rupture of magnitude m at a distance \mathbf{r} by the probability densities of the magnitude and distance to the i th fault, $f_i(m)$ and $f_i(\mathbf{r}|m)$, respectively. To get the contribution for all ruptures of the i th source, a double integration over distances \mathbf{r} and magnitudes m is performed.

In the classical formula, the probability obtained from the aforementioned double integration is multiplied by the mean annual rupture occurrence rate ν_i of the i th source to get the final mean annual exceedance rate (Senior Seismic Hazard Committee (SSHAC), 1997):

$$\lambda(y_0, \mathbf{r}_0)[y \geq y_0] = \sum_{i=1}^I \nu_i \int_{\mathbf{r}} \int_{m_{\min}}^{m_{\max}} P[y \geq y_0|m, \mathbf{r}, \mathbf{r}_0] f_i(m) f_i(\mathbf{r}|m) d\mathbf{r} dm. \quad (2.3)$$

The conditional probability $P[y \geq y_0|m, \mathbf{r}, \mathbf{r}_0]$ can be derived by evaluating the complementary cumulative distribution function (tail distribution) of a ground motion probability model (see Section 2.1.4). For the distribution of magnitudes, a common assumption is to use results of Benno Gutenberg and Charles Richter (1942, 1944):

$$\log_{10} n_m = a - bm. \quad (2.4)$$

Here, m denotes magnitude, n_m is the number of events with magnitudes $\geq m$, and a and b are constants.

For classical PSHA, the integral of Equation 2.3 is solved numerically or analytically. For instance, Cornell (see Section 2.1.1) derived a closed analytical form for line, point and area sources. For intensity measures such as Maximum Mercalli Intensity (MMI), he found Type 1 extreme value distribution (Gumbel, 1935), while for ground motion parameters such as peak ground acceleration, he finds Type 2 (asymptotic extreme value distribution). It is important to note that these results are not based on the asymptotic extreme value argument (Gumbel, 1958), but dependent on the functional form of the assumptions.

The assumptions for magnitude distribution (Equation 2.4) and conditional ground motion probability (Section 2.1.4) can relatively easily be replaced with other functional forms. This merely changes the results of some analytical integrations. For example, in a later work, Equation 2.3 was integrated for a quadratic magnitude-frequency law (Merz and Cornell, 1973). Removing the assumption that earthquake occurrences can be modeled by a Poisson process is less straightforward and can make analytical integration very intricate, however.

See Figure 2.1 for the schematic workflow of classical PSHA. Modern implementations typically compute an earthquake rupture forecast (e.g. Field et al., 2003; Paganì et al., 2014), from which the hazard output is obtained by integration of event probabilities. Note that no realization of the ground motion distribution needs to be explicitly computed.

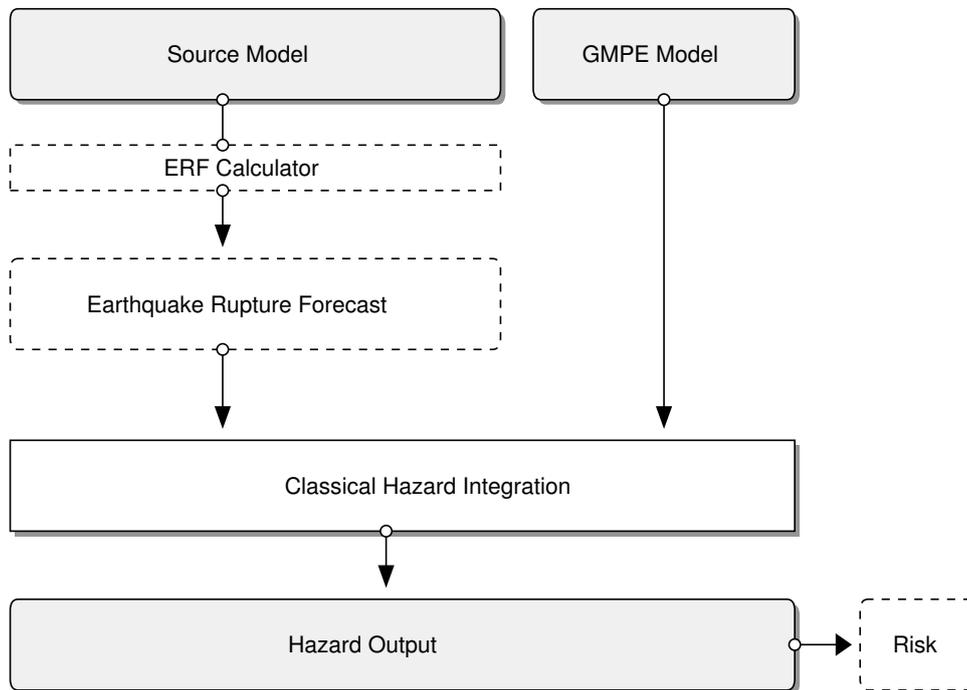


Figure 2.1: Schematic workflow of classical PSHA. Modern implementations typically compute an earthquake rupture forecast (ERF), from which the hazard output is obtained by integration of event probabilities. Illustration inspired by the OpenQuake Book (Crowley et al., 2011).

The probabilities of exceeding a certain threshold are directly obtained by integrating Equation 2.3. Although risk results can be obtained from classical hazard output (Cao et al., 1999), this is usually performed through event-based PSHA (see Section 2.1.3).

Classical Hazard Results

After computing the probability for a range of ground motion levels y_0 for a constant t_0 , the result can be visualized for a single site \mathbf{r}_0 by plotting the obtained probability values versus y . This is called a hazard curve.

Evaluating the probability at a range of locations \mathbf{r} for a specific t_0 enables producing a so-called hazard map. This is usually a color-coded plot of the spatial distribution of ground motion levels y corresponding to some constant probability level p_0 .

Note that although obtaining risk results from classical PSHA output is possible (e.g. Cao et al., 1999), for this purpose event-based PSHA (see Section 2.1.3) is usually preferred. Another procedure to directly compute the probability distribution for losses largely analogous to the classic PSHA procedure also exists (Wesson et al., 2009).

Classical PSHA: Analytical Integration of Uncertainty

In the first publication by Cornell (1968), the distribution of earthquake magnitudes and site-to-source distance was treated by analytical integration. The original formulation did not take any aleatory ground motion variability into account, but just used the median prediction. However, only two years later, more sophisticated treatment of the conditional ground motion probability was added by Esteva (1970). In classical PSHA it is customary to analytically or numerically integrate the ground motion exceedance probability (using Equation 2.3 and assumptions).

In the classical formulation of PSHA, explicit computation of the ground motion for individual event realizations is not necessary. Ground motion exceedance probability is integrated directly from source parameters and the conditional ground motion distribution.

It is also important to note that classical PSHA does not involve any stochastic processes. The ground motion distribution is either integrated analytically, or numerically using deterministic sampling. This is in contrast to Monte Carlo integration, which is usually employed for event-based PSHA (see Section 2.1.3). For a more comprehensive discussion of classical PSHA, its advantages and disadvantages, confer to a report by the Senior Seismic Hazard Committee (SSHAC) (1997).

Alternative Formulation

The traditional formulation first sums mean annual rates and then computes probabilities from them. A more efficient formulation of classical PSHA was introduced later (e.g. Field et al., 2003; Pagani et al., 2014): Assuming that the probability of multiple ruptures of the same source within the investigated time interval t_0 (see Section 2.1.2) can be neglected allows to work with probabilities instead of mean annual occurrences. The probability of exceeding intensity measure value y_0 is then given by

$$P(y \geq y_0) = 1 - \prod_{i=1}^I \left(1 - \sum_{n=1}^{N(i)} P(y \geq y_0 | \mathbf{rup}_{i,n}) P(\mathbf{rup}_{i,n}) \right). \quad (2.5)$$

Here, $N(i)$ is the maximum number of ruptures on source i , $\mathbf{rup}_{i,n}$ is a shorthand notation for the n th rupture of the i th source, and $P(\mathbf{rup}_{i,n})$ denotes the probability of that rupture. $P(y \geq y_0 | \mathbf{rup}_{i,n})$ denotes the conditional probability of exceeding intensity level y_0 given that rupture (Field et al., 2003).

This formulation allows to handle time-dependent probabilities explicitly and does not assume Poissonian statistics. For a derivation of this formula from the classical method under the assumption that multiple ruptures can be neglected, see Appendix A of Field et al. (2003).

2.1.3 Event-Based Simulation Approach

Methodology

The **main loop for classical PSHA is over sites**. For each site, the hazard is integrated taking all relevant sources into account. This is performed evaluating the integral, which was analytically derived beforehand for a specific magnitude-frequency relationship, for a rupture occurrence model (generally assumed to be Poissonian for the classical approach), and for specific source typologies such as point or line sources.

Classical PSHA in this form is the most efficient way of obtaining traditional hazard results. It is recommended to be used if these results are the only desired output, e.g. for an engineering project at a single site (Crowley et al., 2011). It is less well-suited when individual event ground motions are explicitly needed. This involves, for example, modeling the risk of losses to a portfolio in the insurance industry, where per-event losses are required, or advanced techniques such as ground motion correlation or deaggregation¹ (see footnote on page 8).

In that case, an alternative procedure is **event-based PSHA** (see Figure 2.2) where the **main loop is over events**. In this approach a synthetic seismicity catalogue (also known as stochastic event set, SES) is generated from an earthquake rupture forecast (ERF) for a given time interval t_0 using a temporal (rupture) occurrence model (TOM). An ERF is a list of possible ruptures and their associated probabilities obtained from a seismic source model. For each rupture **rup** that can theoretically be generated by a source, this is usually performed by sampling the number of ruptures k within a time interval t_0 from the probability distribution $P_{\text{rup}}(k, t_0)$ corresponding to the temporal occurrence model (e.g. Poissonian, see Equation 2.1 for the probability mass function). In other words, a stochastic event set is a sample of the population of all ruptures contained in the ERF (Crowley et al., 2011).

For each event in the SES, one or more realizations of the ground motion field is then explicitly computed at all necessary sites. This usually involves computing the mean ground motion as well as inter- and intra-event residuals (see Section 3.4). Since both the seismicity catalogue as well as the realizations of event ground motion are usually obtained using Monte Carlo simulation, this method is also known as the stochastic or Monte Carlo method.

Obtaining Classical Hazard Results from Event-Based PSHA

Also when employing event-based PSHA, one might want to obtain classical hazard results, i.e. hazard curves and hazard maps. These can be obtained by post-processing the simulated ground motion fields. Note that this does not imply that the results of all scenarios need to be stored explicitly. Post-processing can occur incrementally, e.g. following immediately after the simulation of one or several events.

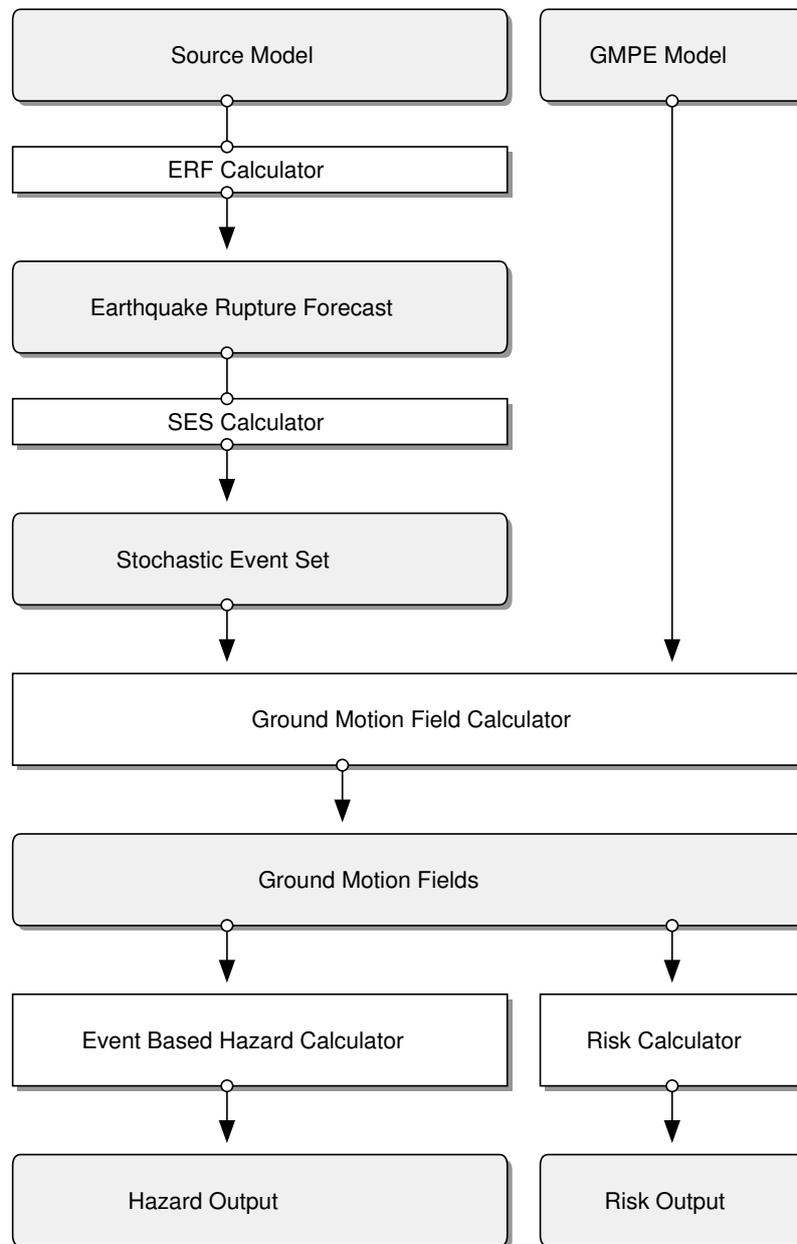


Figure 2.2: Schematic workflow of a typical implementation of event-based PSHA and event-based risk analysis. Note that a stochastic event set (SES) is obtained from the earthquake rupture forecast (ERF), and that the hazard and risk output is obtained from the collection of ground motion fields simulated using Ground Motion Prediction Equations (GMPEs). Illustration inspired by the Openquake Book (Crowley et al., 2011).

The mean rate of exceeding ground motion level y_0 at a site \mathbf{r}_0 can be obtained as

$$\lambda(y_0, \mathbf{r}_0)[y \geq y_0] = \frac{\sum_{k=1}^K \mathcal{H}(y_k - y_0)}{T_0}, \quad (2.6)$$

where y_k is the ground motion produced by the k th rupture, K is the number of simulated ruptures, T_0 is the duration of the stochastic event set simulation, and \mathcal{H} is the Heaviside function (Kafka et al., 1999; Crowley et al., 2011).

Equation 2.6 estimates the mean ground motion exceedance rate from a simulation of duration T_0 . The estimate becomes more accurate with increasing T_0 . It is also possible to join multiple stochastic event sets with shorter duration T_0 . To obtain the probability of exceedance in time interval t_0 , a temporal occurrence model for the ruptures needs to be integrated. If the temporal occurrence model of the sources from which the stochastic event set was generated were all Poissonian, this can be performed by using the result of Equation 2.6 with the cumulative distribution function of the Poissonian model given by Equation 2.2. Other approaches employ time-dependent temporal occurrence models (e.g. Anagnos and Kiremidjian, 1988; Matthews et al., 2002; Fitzenz and Nyst, 2015). However, as with many aspects of PSHA, calibration and verification of these models is difficult due to the limited amount of available data.

2.1.4 Ground Motion Modeling

Ground Motion Prediction Equations

A ground motion prediction equation (GMPE) is a statistical model predicting the ground motion at a site for a given earthquake rupture. A GMPE predicts the ground motion in terms of some ground motion parameter which is a function of the full time history of the ground motion occurring at a site during an earthquake. Commonly used parameters include peak ground acceleration (PGA), peak ground velocity (PGV), peak ground displacement (PGD), and peak spectral acceleration (PSA). While the first three have some direct physical meaning, peak spectral acceleration is a response function of a damped single-degree-of-freedom (SDOF) oscillator subjected to ground motions. Some processing of the seismogram is necessary to obtain these parameters (see e.g. Krinitzsky, 2002; Douglas, 2003; Pulido et al., 2004; Güllü, 2012).

In general, if y is a ground motion parameter of interest, then the ground motion prediction at the i th site for the j th event, denoted by y_{ij} , is given by

$$\ln(y_{ij}) = \ln(\bar{y}_{ij}) + \epsilon_{ij} + \eta_j. \quad (2.7)$$

Here, \bar{y}_{ij} denotes the predicted median ground motion intensity, ϵ_{ij} denotes intra-event residual at the i th site for the j th event, while η_j denotes the inter-event residual for the j th event (Jayaram and Baker, 2009). ϵ_{ij} and η_j are normal random variables with zero mean. The multivariate normality assumption for the joint distribution of intra- and

inter-event residuals has been confirmed by statistical tests (e.g. Jayaram and Baker, 2008; Strasser et al., 2008).

The model for the median prediction \bar{y} and the standard deviation terms are developed via regression. For example, a simple equation for the median ground motion model is

$$\bar{y}(m, \mathbf{r}) = c_1 + c_2 m + c_3 \ln(\mathbf{r} + r_0), \quad (2.8)$$

with the magnitude m and the focal distance \mathbf{r} . Coefficients c_1 , c_2 , c_3 and r_0 describe the attenuation relationship.

The standard deviation σ of the observations can be obtained via

$$\sigma = \sqrt{\frac{1}{n} \sum (y_i - \bar{y}(m, \mathbf{r}))^2}, \quad (2.9)$$

where n is the number of observations used for the regression, and y_i is the i th observation. To obtain the conditional probability of exceeding some ground motion level y_0 , the probability density of Equation 2.7 is integrated:

$$P[y \geq y_0 | m, \mathbf{r}] = \int_{y_0}^{\infty} f_y(y | m, \mathbf{r}) dy, \quad (2.10)$$

$f_y(y | m, \mathbf{r})$ is the probability density of $y(m, \mathbf{r})$. The integration is often truncated at two or three standard deviations above the mean (Grünthal et al., 2013). It is still under debate whether this can be seen in seismic data (e.g. Anderson and Brune, 1999; Jayaram and Baker, 2008), but it is probably safe to assume that most researchers will agree that at some point a truncation can be justified on physical grounds. However, due to the lack of a statistically significant correlation between physical mechanisms and large residuals, the deduction of a general truncation scheme is extremely difficult (Strasser et al., 2008).

For classical PSHA, the integration of Equation 2.10 can be performed by inverting the GMPE and evaluating the complementary cumulative distribution function of the standard normal distribution. For example, when using Equation 2.7 with Equation 2.8 and a constant standard deviation σ_c , this results in:

$$P[y \geq y_0 | m, \mathbf{r}] = \Phi^* \left(\frac{y_0 - c_1 - c_2 m - c_3 \ln(\mathbf{r} + r_0)}{\sigma_c} \right). \quad (2.11)$$

Here, Φ^* denotes the complementary cumulative distribution function of the standard normal distribution, and the argument is the inverted GMPE (McGuire, 1976).

In the development of most GMPEs, seismic data is jointly used from multiple earthquakes and sites. In effect, this means that the problem of limited data for a specific location is overcome by making an ergodic assumption (Anderson and Brune, 1999). An ergodic process is a random process where the probability distribution of a random variable at a fixed time in space is equal to its probability distribution at a fixed point in time.

Modern GMPEs are usually more complex than the form presented in Equation 2.8. Some incorporate the influence of source mechanism, path and site effects, such as amplification by soil and basins, local geology, faulting mechanism and directivity effects. In

addition, instead of reporting some fixed total standard deviation, modern GMPEs provide parametric models for inter- and intra-event standard deviation.

When individual event realizations are needed, numerical integration using event-based PSHA is employed. In this case the integral of Equation 2.10 is evaluated via post-processing the results of event-based ground motion footprints. The ground motion uncertainty treatment framework implemented in the course of this thesis (see Section 3.4) is based on this approach.

Numerical Waveform Modeling

The large amount of earthquakes that usually need to be taken into account for a PSHA study result in the necessity of using a relatively simple ground motion model of limited computational cost. Due to the extreme computational requirements (10000 CPU-hours or more on a modern supercomputer for a single event), integrating numerical simulation of physics-based three-dimensional wave propagation into a large-scale seismic hazard and risk analysis instead of using GMPEs remains the exception. Nevertheless, recent pioneering work (e.g. Graves et al., 2011; Infantino, 2016; Paolucci et al., 2018) is investigating the possibility of incorporating dynamic wave simulations based on numerically solving the wave equation into PSHA.

Field and SCEC Phase III Working Group (2000) found that the aleatory variability of GMPEs represents complex effects caused by the unique wave propagation from the source to a site. They concluded that in order to significantly improve PSHA, numerical waveform simulation needs to be adopted instead of conventional regression-based ground motion modeling.

2.2 Probabilistic Seismic Risk Analysis

2.2.1 Historical Background

Seismic vulnerability assessment started in the **1970s** with the application of **empirical methods** using macroseismic intensity measures. At the time, using physical ground motion parameters was avoided due to large scatter (Calvi et al., 2006). **Damage probability matrices** (DPMs) were first proposed by Whitman et al. (1973). The usage of DPMs is still popular to account for damage state uncertainty (e.g. Dolce et al., 2003; Pasquale et al., 2005). The **vulnerability index method** has been introduced by Benedetti and Petrini in **1984**. It defines a vulnerability index for a building based on a weighted sum. It uses building parameters such as plan and elevation configuration, foundation type, structural and non-structural elements, state of conservation and material type and quality. Since the vulnerability index method is not only based on the building typology as single predictor variable, this so-called indirect method allows to determine the vulnerability characteristics of a specific building stock somewhat more precisely. **Continuous vulnerability functions** were introduced shortly after DPMs. Their derivation was more complicated since seismic intensity is not a continuous variable (Calvi et al., 2006). The

problem was overcome by converting seismic intensity to a physical measure (e.g. Spence et al., 1992; Orsini, 1999).

Enabled by the increased availability of spectral seismic hazard input, **analytical and mechanical methods** started to be used more extensively in the **1990s**. Obtaining analytical vulnerability curves through non-linear analysis is a time-consuming procedure and computationally extremely intensive. Therefore, analytical methods are often used in conjunction with empirical DPMs and vulnerability curves (Calvi et al., 2006). **Hybrid methods** combine empirical data obtained after an earthquake with analytical results obtained from a mathematical model (e.g. Kappos et al., 1995; Barbat et al., 1996). Another analytical vulnerability method is the **capacity spectrum method**. For example, it is used by HAZUS (Hazard United States; Whitman et al., 1997; Federal Emergency Management Agency (FEMA), 2001), a methodology to estimate losses from earthquakes on a regional basis. The capacity spectrum method finds the structural performance of a building type as the intersection point of two curves: the acceleration-displacement spectrum and the capacity spectrum (pushover curve). The former curve represents ground motion, while the latter represents the horizontal displacement of the structure under increasing lateral load (Kircher et al., 1997).

2.2.2 Components of Seismic Risk Analysis

Probabilistic Seismic Risk Analysis (PSRA) builds on top of PSHA and consists of three distinct parts: a hazard model, an exposure model, and a vulnerability model (see Figure 2.3).

The **hazard model** is part of a PSHA modeling framework and — as explained in Chapter 2.1 — estimates an exceedance rate per ground motion intensity. An **exposure model** comprises the locations and characteristics of exposed risk items, such as buildings and infrastructure, or the content of buildings (e.g. industrial machinery). They are grouped into certain classes according to their structural and nonstructural characteristics. A **vulnerability model** provides damage probabilities for different types of exposed risk items depending on the ground motion intensity they have been exposed to.

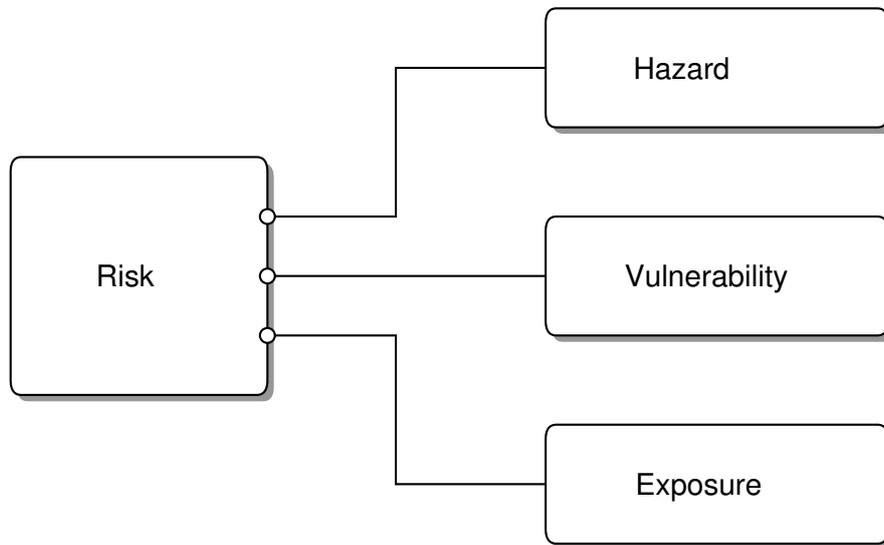


Figure 2.3: $\text{Risk} = \mathcal{F}(\text{Hazard}, \text{Vulnerability}, \text{Exposure})$.

2.2.3 Exposure Models

In contrast to hazard modeling, less studies have focused on the field of exposure modeling for purposes of seismic risk assessment. Depending on the type of data and available computational resources, two distinct types of exposure models are used: **aggregated models** and **single-risk models** (Crowley, 2014). There are some differences in the corresponding workflows when using either of these in probabilistic seismic risk analysis.

Aggregated Models

In aggregated exposure models, individual buildings within a given geographic area such as grid cells or administrative zones are aggregated to a single information record. Administrative zones, e.g. counties or postcodes, can be of varying shape and size (see Figure 2.4). This type of exposure model is frequently developed with very little available data, resulting in the necessity to make numerous assumptions about the true distribution and characteristics of exposure. Aggregated models are still very common, in particular for country-wide or continental scale risk modeling outside of the insurance industry (see e.g. Yepes et al., 2010; Silva et al., 2014). Even in the insurance industry, where smaller portfolios containing only few high value risk items are sometimes relatively well known, country-wide exposure currently almost never is.

In some aggregated exposure models, all risk items are simply located at a single loca-

tion, such as the centroid of their zone. This introduces the implicit assumption of full correlation of ground motion uncertainty for all buildings which are aggregated to a single point.

In practice most aggregated exposure models neglect uncertainty treatment. This is being justified by the law of large numbers, i.e. by assuming that the aggregated data represents a very large number of individual buildings so that it suffices to assume that the arithmetic mean damage ratio of modeled buildings equals the expected value of the vulnerability distribution. As an attempt to improve this, aggregated exposure is sometimes deaggregated onto a weighted grid using population census and density data, as well

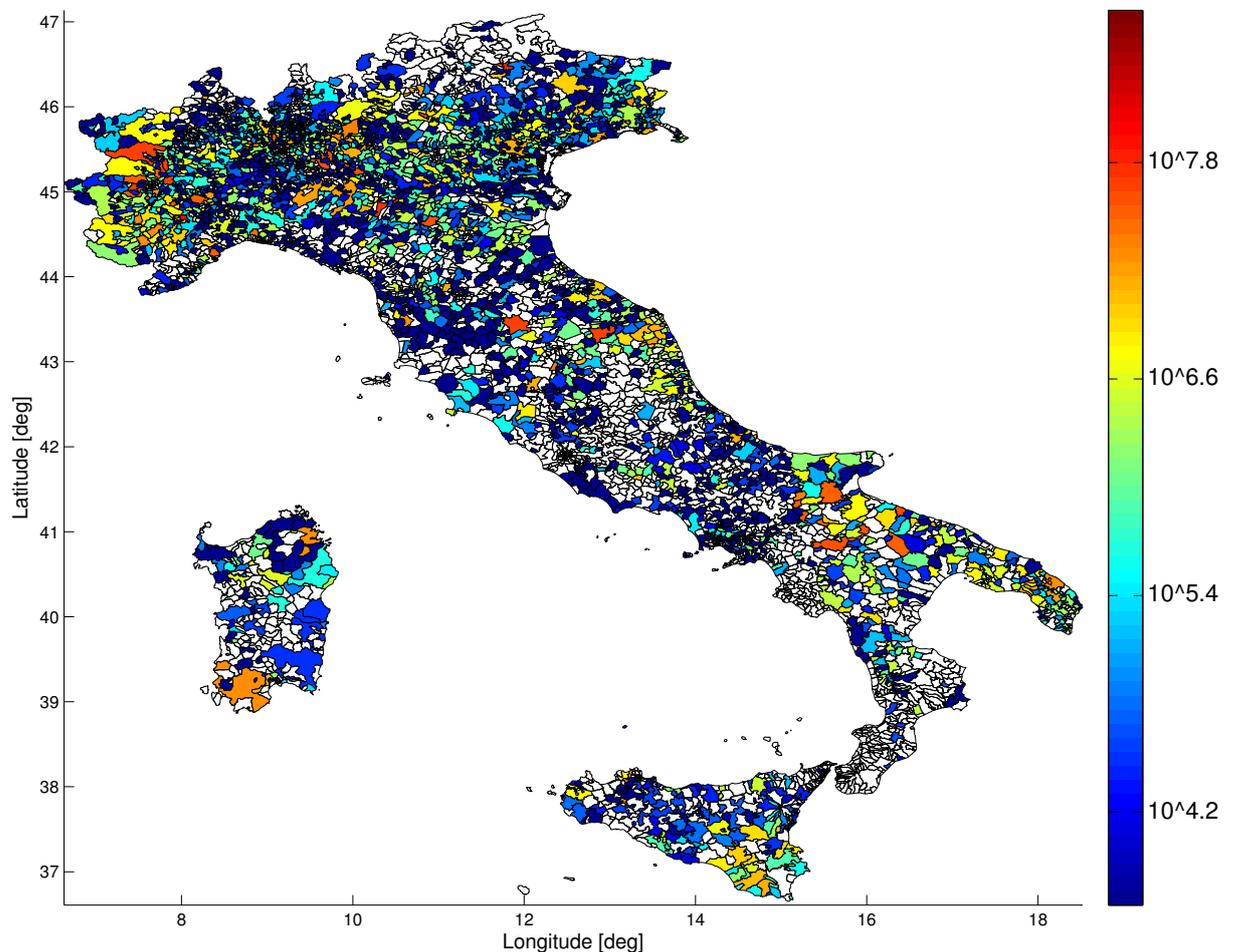


Figure 2.4: Example of the geographical value distribution of an aggregated exposure model provided on postcodes in Italy. Note this does not correspond to a real portfolio, but is provided for illustrative purposes only.

as publicly or commercially available building stock data. However, this merely results in smaller areas of aggregation corresponding to the Voronoi cells of the grid, and the assumptions mentioned above are still employed. Furthermore, aggregated models do not provide detailed information on the individual insurance conditions on site or policy levels, which can have a very significant effect on the finally modeled loss-frequency curve of insured losses.

Single-Risk Models

Conversely, single-risk (or detailed) exposure models contain a single information record per individual risk item. These exposure models comprise all information related to the collection of assets for which risk assessment is performed. This includes data such as location, insured value, construction type and year, building height, and insurance conditions. Table 2.1 shows an example of this kind of model.

More detailed exposure models in turn allow to use a more specific vulnerability model, which might take into account these additional parameters for the prediction of damage. Also for single-risk exposure models, the quality of data can vary. Some information might be missing or not plausible. For example, precise coordinate location information might not be available for all risk items. Different parts of the portfolio can be known on different geographic resolutions, e.g. coordinate level, postcodes, or administrative zones of varying size. Note that this does not necessarily mean that these risk items are aggregated, since the value of individual risk items might still be known while precise coordinates are not.

Geo.-Resolution	Geo.-Zone	Lon.	Lat.	Value	Construction Type	...
Postal Code	10130	<i>unknown</i>	<i>unknown</i>	12340	reinforced concrete	...
City District	DKI Jakarta	<i>unknown</i>	<i>unknown</i>	34343	reinforced concrete	...
Coordinate	Palembang	104.9	-3.0	34343	masonry	...
City	Medan	<i>unknown</i>	<i>unknown</i>	23456	<i>unknown</i>	...
...

Table 2.1: Hypothetic example of a detailed exposure model. Note that a dataset can be quite heterogeneous, e.g. the quality of location information can vary or some records might have unknown fields.

2.2.4 Vulnerability Models

The concept of vulnerability provides a link between hazard and risk. Vulnerability models define the susceptibility of risk items (e.g. buildings, bridges, office equipment, or wine

barrels) to be damaged by a certain level of ground motion.

Some important parameters influencing vulnerability are:

- construction material
- construction type (load bearing structural systems)
- building design
- official building construction code
- building height (number of stories)
- occupancy type (type of usage, e.g. school, office, hospital)

A vulnerability model is a set of vulnerability curves and associated uncertainty distributions for different building classes defined on the basis of these parameters. Vulnerability functions provide a probabilistic model of damage depending on seismic demand. Seismic demand is commonly quantified using a ground motion intensity parameter such as PGA or PSA (see Section 2.1.4). Depending on its height, a building will respond more or less to acceleration at different frequencies (see Table 2.2).

Rise Class	Stories	Typical Natural Period
Low-Rise	< 3	< 0.3 seconds
Mid-Rise	3 – 8	0.3 – 1.0 seconds
High-Rise	9 – 20	1.0 – 2.0 seconds
Skyscraper	> 20	> 2.0 seconds

Table 2.2: Typical natural resonance periods of building rise classes.

In the insurance industry, damage is usually measured as the damage ratio (DR) defined as

$$DR = \frac{\text{cost}_{\text{repair}}}{\text{value}_{\text{insured}}}, \quad (2.12)$$

where $\text{cost}_{\text{repair}}$ is the cost of repairing or rebuilding and $\text{value}_{\text{insured}}$ is the insured value. $\text{cost}_{\text{repair}}$ is also referred to as *ground-up loss*.

As mentioned in Section 2.2.1, a number of different methodologies have been developed. Judgement based methods, as well as analytical and mechanical methods are in widespread use — such as the FEMA loss estimation methodology and HAZUS (Kircher et al., 1997; Whitman et al., 1997; Federal Emergency Management Agency (FEMA), 2001).

In particular in the insurance industry, another promising approach to construct vulnerability models is to analyze ground motion and loss data from past earthquakes. If the quality of a dataset is good, an empirical vulnerability curve and uncertainty distribution can be obtained this way (e.g. Sarabandi et al., 2004; Rossetto et al., 2013; Scheingraber, 2015). Figure 2.5 shows an example of this for three different construction types. However, in many cases this approach is still hindered by the lack of reliable ground-up loss data.

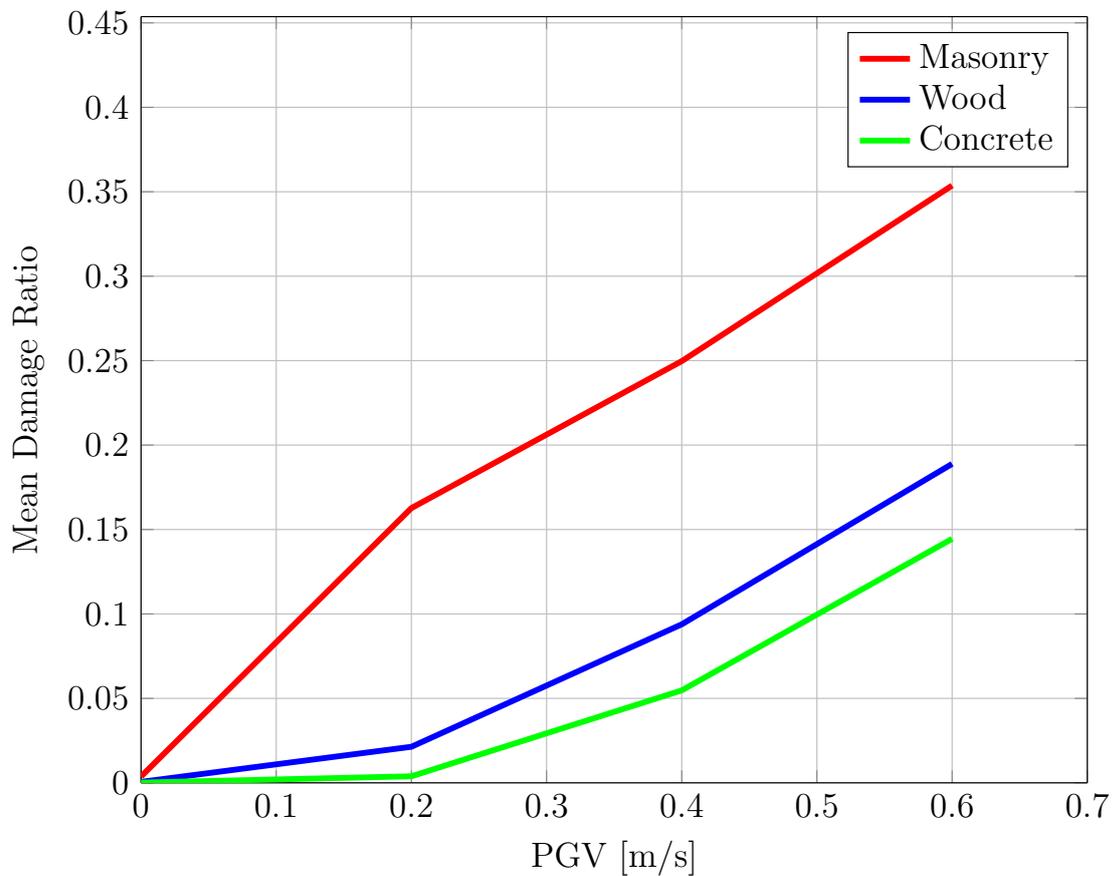


Figure 2.5: Example of empirical vulnerability curves for different construction types. This shows mean damage ratio (see Equation 2.12) against peak ground velocity (PGV, in m s^{-1}). Color denotes different construction types. These vulnerability curves have been obtained from ground motion and loss data of the 2011 Tōhoku-Oki earthquake (Japan) using a sliding window approach with equidistant window width of 0.2 m s^{-1} and step size of 0.2 m s^{-1} (Scheingraber, 2015).

Chapter 3

A Framework for the Treatment of Ground Motion and Damage Uncertainty in Probabilistic Seismic Hazard and Risk Analysis

This chapter describes the theory behind a new framework for the treatment of ground motion and damage uncertainties by repeatable stochastic simulation, as implemented during the course of this thesis for a seismic risk assessment framework used by *Munich Re* (see Chapter 1.1).

3.1 Introduction

The financial impact of an earthquake on insured portfolios is a convoluted process involving a multitude of highly complex physical phenomena, such as the nucleation and propagation of a rupture, the emission and propagation of seismic waves in a heterogeneous medium, the response and eventual failure of building materials, and the evaluation of the value and the repair cost of an affected structure. Remarkable progress has already been made to model many of these processes in a manner that accurately reflects their physical nature, such as numerical modeling of 3D waveform propagation for strong scenarios in selected regions for probabilistic seismic hazard assessment (e.g. Graves et al., 2011; Infantino, 2016; Mert et al., 2016), or numerical simulation of the non-linear response of structures subjected to ground motion in the area of earthquake engineering (e.g. Mazzoni et al., 2005; Deierlein et al., 2010; McKenna, 2011; Ni et al., 2012; Chen et al., 2013). However, fully doing so for the purpose of seismic risk assessment in the insurance industry seems to remain impractical or unfeasible for the foreseeable future. The reasons for this are manifold and include the finiteness of financial and computational resources to physically model the tremendous number of earthquake scenarios which are necessary to be included in the analysis, limited knowledge of the composition of the crust,

and often a surprisingly high degree of uncertainty about the construction materials as well as the condition and location of insured buildings. Therefore, current practice necessarily involves utilization of simpler models, thereby reducing computational demands and time expenditure for model development.

To model ground motion at a site, nowadays the most common approach is to use a ground motion prediction equation (GMPE; see Section 2.1.4) obtained from regression analysis of site ground motion of records of real events — sometimes supplemented with synthetic events. Modern GMPEs depend on local soil, varying distance measures to the earthquake, and a few source parameters such as earthquake magnitude or faulting mechanism. To model the damage to a building due to ground motion, vulnerability functions (see Section 2.2.4) are used. Like GMPEs, these are relatively simple curves and depend on a few factors such as the region, height and age of the building, and the construction type of load-bearing structures.

Even when neglecting the uncertainty in data, these relatively simple models will always be associated with a high degree of aleatory variability, as they are not designed to capture the physical processes which influence the seismic wave along its path, the response of structures, and losses of an insured portfolio. An important requirement for a modern seismic risk assessment framework is therefore to take the full uncertainty associated with ground motion and vulnerability modeling into account.

3.2 Monte Carlo Simulation and Random Number Generation

In order to take uncertainty of a numeric model into account, it is usually necessary to evaluate the model at a large point set. Depending on the problem, there is a multitude of suitable methods to generate this set. The Monte Carlo method uses random numbers in a controlled manner. Monte Carlo is the standard integration scheme for high-dimensional problems such as Probabilistic Seismic Hazard Assessment (PSHA) and Probabilistic Seismic Hazard Assessment (PSRA) — in particular if individual event losses are needed or correlation models are used.

3.2.1 Monte Carlo Simulation

The Monte Carlo method was invented during the Manhattan Project at Los Alamos Laboratory and first published by Metropolis and Ulam (1949). Simple Monte Carlo in its original form is based on uncorrelated sampling using uniformly distributed random numbers, generating a set of n realizations $\{Y_1, Y_2, \dots, Y_n\}$ of a random variable Y . A parameter of interest, such as the expectation or variance of Y , can then be estimated from the sample using a suitable estimator. For example, an unbiased estimator for the expectation $\mu = \mathbb{E}(Y)$ is given by

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n Y_i, \tag{3.1}$$

where $Y_i = f(X_i)$ is the random variable depending on the i th random number. An unbiased estimator for the variance σ^2 is given by

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \hat{\mu})^2. \quad (3.2)$$

In real applications, generating Y can be quite complicated. It might depend on a large number of underlying random variables with potentially discontinuous probability distributions (see e.g. Section 3.5.3), as well as additional underlying random processes involving decisions and jumps (see e.g. Chapter 6). However, simple Monte Carlo can still be used as long as Y can be averaged (Owen, 2013).

Convergence and Error Estimation

The strong law of large numbers ensures convergence of simple Monte Carlo. When the sample size n grows to infinity, the probability \mathbb{P} that the estimator $\hat{\mu}$ is equal to the expectation $\mu = \mathbb{E}(Y)$ is one:

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} |\hat{\mu}_n - \mu| = 0\right) = 1. \quad (3.3)$$

Since Monte Carlo is a stochastic method, there are no exact bounds for the error $E = |\hat{\mu}_n - \mu|$. The error at a given sample size n is instead distributed according to a probability distribution, which is called **sampling distribution**. The standard deviation of the sampling distribution is referred to as **standard error** (E_{SE}). This quantity is often used as an indication of size of the true error.

If the sampling distribution is known to be normal, then for many statistics of interest there are formulas available to directly estimate standard errors using the sample size n and variance σ^2 (see e.g. Harding et al., 2014). For example, under the normal assumption for the sampling distribution, the standard error of the mean $E_{SEM,normal}$ can be estimated as

$$E_{SEM,normal} = \frac{\hat{\sigma}}{n}. \quad (3.4)$$

In practice, Equation 3.4 is often applied without knowledge about the sampling distribution. However, if the sampling distribution is not known a better approach is either using repeated simulation with the same sample size (e.g. Koehler et al., 2009) or estimating the error from the same sample using bootstrapping (Efron, 1979; Efron and Gong, 1983; Efron, 1982). This resampling-based method assumes that the sample already contains all information about the underlying population. A set of bootstrap resamples is obtained by sampling *with replacement* from the original sample. Each member of the set is of the same size n as the original sample. The standard deviation of the estimates obtained from the bootstrap are then used as an estimate for the standard error E_{SE} .

In the studies presented in Chapters 5 and 6 of this thesis, repeated simulation was used to estimate standard errors of different sampling schemes, and bootstrapping was used to estimate corresponding confidence intervals.

3.2.2 Uniform Random Number Generators

As implied in Section 3.2.1, Monte Carlo simulation — and probabilistic algorithms in general — rely on the availability of a supposedly endless stream of independent and identically distributed (IID) random variables following different probability distributions. Generating random variables of a nonuniform distribution in turn requires a theoretically endless stream of IID uniformly distributed random variables (Robert and Casella, 2004), which is provided by a random number generator (RNG). A good RNG is therefore a key ingredient of Monte Carlo simulation.

True and Pseudo Random Number Generators

In computing, two different kinds of RNGs are generally used: hardware random number generators (HRNGs; also called “true” random number generators), and pseudo-random number generators (PRNGs; also called deterministic random bit generators). HRNGs are based on collecting entropy from physical sources, such as thermal and acoustic noise or quantum phenomena. Modern computer architectures like Intel Ivy Bridge (Hofemeier, 2012) and even smartphones provide dedicated HRNG devices. Random numbers based on the collected entropy are made available to programs by the operating system¹. HRNGs are most important for cryptographic applications which not only need *statistical randomness*, but also the stronger requirement of *unpredictability*. While there is ongoing research concerning faster HRNGs (e.g. Reidler et al., 2009), the devices that are nowadays commonly used can not collect entropy at a rate that would suffice to produce random numbers for direct application. Theoretically, this is not necessary anyway, because entropy is not a quantity that can be consumed at any measurable rate. Thus, HRNGs are used to obtain a random *seed* to generate an initial state for a cryptographically secure PRNG, which acts as an *extensor* of randomness and stretches the seed into a sequence of apparently random numbers (L’Ecuyer, 2014).

For stochastic simulation and computational statistics, PRNGs are better suited because they can generate random variables at a fast rate and fulfill another important requirement: *repeatability* of the generated “random” sequences. Of course, there is nothing truly random about a sequence produced by a PRNG — the dilemma was famously boiled down to its essence by Von Neumann (1951): “Any one who considers arithmetical methods of producing random digits is, of course, in a state of sin.” In fact, PRNGs are deterministic algorithms which only appear to behave like a random variable.

Uniform Pseudo Random Number Generators

As mentioned before, (Pseudo-)RNGs usually imitate the behavior of a random variable following a uniform distribution. The probability density $f(x)$ of the continuous uniform

¹On Unix-like systems, the random digits are made available through `/dev/*random` for non-dedicated devices and through `/dev/hrng` for dedicated HRNG devices. On Windows, they are made available through the `CryptGenRandom` and `CNG` APIs.

distribution $U(a, b)$ with $a < b$ is given by:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b], \\ 0 & \text{otherwise.} \end{cases} \quad (3.5)$$

According to L'Ecuyer (1994), a uniform PRNG can be defined as a structure $(\mathcal{S}, \mu, f, \mathcal{U}, g)$, with \mathcal{S} a finite set of *states*, μ a probability distribution on \mathcal{S} used to select the initial state s_0 , $f : \mathcal{S} \rightarrow \mathcal{S}$ the transition function between consecutive states, and $g : \mathcal{S} \rightarrow \mathcal{U}$ the output function to the output set \mathcal{U} . Most PRNGs model a random variable following the standard uniform distribution $U(0, 1)$, which implies that the output set equals the unit interval, $\mathcal{U} = [0, 1]$. Consecutive states s_i with $i \geq 1$ are obtained using a recurrence relationship $s_i = f(s_{i-1})$. Because \mathcal{S} is finite, starting from a state s_i , after a certain number of state transitions j eventually s_i will be reached again: $s_{i+j} = s_i$. The smallest j for which this happens is called the *period length* ρ . A good PRNG carries only a small state \mathcal{S} , but preferably features a large period length at the same time. It is also important that a PRNG performs well on statistical randomness tests². Bad PRNGs generate correlated sequences which can cause undesired and surprising effects in probabilistic algorithms and wrong results in stochastic simulations.

Almost all PRNGs which are nowadays used are based on linear recurrences:

$$X_i = (a_1 X_{i-1} + \dots + a_k X_{i-k} + c) \pmod{m}. \quad (3.6)$$

Prime numbers are good choices for m , because the period length can then be maximized to $\rho = m^k - 1$ with well chosen a_i (Knuth, 1981). The well-known class of **linear congruential generators** (LCGs) use $k = 1$ with different coefficients a_1 and c . LCGs are easy to implement, very fast, and only require to store a small state in memory, which makes them a good choice for e.g. embedded systems. The quality of a LCG strongly depends on the parameter choices, and in a lot of older implementations — which have unfortunately also been used in many scientific studies which are now often being questioned — bad choices were made (e.g. IBM's RANDU). Today, there are better PRNGs for a large scale Monte Carlo simulation where high quality random numbers are important.

Instead of a prime number, **linear feedback shift register** (LFSR) generators use a modulus of $m = 2$, which can be implemented efficiently using simple bitwise operations (**bit-shift** and **xor**). Notable examples are *Mersenne Twister* (Matsumoto and Nishimura, 1998), *xorshift* (Marsaglia, 2003), and *WELL* (Panneton et al., 2006). The name is derived from using a *Mersenne prime* for its huge period length ($\rho = 2^{19937} - 1$). It was the first PRNG purposefully designed to fix problems with statistical randomness that were common in LCGs. It was initially perceived as a very good generator and is the default PRNG in virtually all modern software libraries and computer programs. While Mersenne

²Good random number generator testing suites are e.g. *TestU01* (L'Ecuyer and Simard, 2007) and *dieharder* (<http://webhome.phy.duke.edu/~rgb/General/dieharder.php>).

Twister still has a good reputation, it has numerous disadvantages despite the huge period; for example it fails some newer statistical randomness tests already for sequences of 50000 numbers, and has an unnecessarily large state resulting in slow seeding and high memory requirements (O’Neill, 2014). For parallel simulation, it is necessary to divide the (periodic) PRNG sequence into multiple statistically independent *substreams*. Mersenne Twister was not designed to support substreams; it also has a very slow skip-ahead. Furthermore, two seedings are not guaranteed to be statistically independent, especially if the generated sequence is not very large. This makes Mersenne Twister ill-suited for parallel simulation, in particular if results need to be exactly reproducible for repeated simulations.

Multiple recursive generators (MRGs) use Equation 3.6 with a large m and define the output as $U_i = X_i/m$. **Combined multiple recursive generators** (CMRGs) are constituted of several MRGs:

$$X_i = \left(\sum_{j=1}^k (-1)^{j-1} s_{i,j} \right) \pmod{(m_1 - 1)}, \quad (3.7)$$

where $s_{i,j}$ is the i th input from the j th RNG, m_1 is the modulus of the first RNG, and X_i is the i th random integer generated by the CMRG (L’Ecuyer, 1988; L’Ecuyer, 1996). The random numbers of the uniformly distributed output U are then given by

$$U_i = \begin{cases} X_i/m_1 & \text{for } X_i > 0 \\ (m_1 - 1)/m_1 & \text{for } X_i = 0. \end{cases} \quad (3.8)$$

A good combination of MRGs can result in more complicated composite structure and improved statistical randomness of the produced sequence. At the same time, by keeping the underlying generators simple, CMRGs run fast and only need to carry a relatively small state (Panneton et al., 2006). A good representative of this category is e.g. *MRG32K3a* L’Ecuyer (1999). MRG32K3a has a total period of $\rho \approx 2^{191} \approx 3.1 \cdot 10^{57}$ and supports $1.8 \cdot 10^{19}$ independent substreams which each have a period of $7.6 \cdot 10^{22}$. MRG32K3a has a fast skip-ahead, i.e. individual substreams can be selected efficiently. These properties make it a good candidate for parallelizing large-scale Monte Carlo simulations on distributed systems while retaining repeatability.

In this thesis, Mersenne Twister was used for the studies presented Chapters 4 and 6, and MRG32K3a for the final implementation of the framework described in this chapter and the study presented Chapter 5.

3.2.3 Simulating Nonuniform Random Variables

From a uniformly distributed output set \mathcal{U} as obtained from a PRNG or a low discrepancy sequence, random variables following other distributions can be simulated by applying an appropriate transformation (Hörmann et al., 2013). The most understandable and most

commonly used transformation to obtain a random variable Y with cumulative distribution function $F(y)$ is the *inverse transform method*:

$$Y = F^{-1}(U) := \min\{y|F(y) \geq U\}, \tag{3.9}$$

where F^{-1} is the inverse of F , which is also known as *quantile function*. This method is well suited to sample distributions where an analytical expression of the corresponding quantile function is available, such as the Kumaraswamy distribution which in this framework is used to model damage uncertainty (see Section 3.5.2). When the quantile function is not explicitly known — like for the normal distribution used to model ground motion uncertainty — other methods are better suited; for example the *Box–Muller transform* (Box and Muller, 1958), or rejection sampling methods such as *Ziggurat* (Marsaglia and Tsang, 1984; Marsaglia and Tsang, 2000).

When simple sampling of ground motion uncertainty without multidimensional variance reduction techniques is performed, the framework described in this chapter uses the Ziggurat method to sample from the untruncated normal distribution, and the efficient rejection sampling algorithm proposed by Botev (2017) to sample from the truncated normal distribution. When variance reduction techniques are applied (see e.g. Chapter 6), inverse transform sampling in conjunction with an efficient approximation of the quantile function of the normal distribution is used (Beasley and Springer, 1977; Moro, 1995; Glasserman, 2003).

3.3 Framework for Repeatable Stochastic Simulation for Probabilistic Seismic Risk Analysis on Distributed Systems

An important requirement for PSRA in the insurance industry is the *exact* reproducibility of results obtained through stochastic simulation. Note that since exact reproducibility is required, it is not sufficient to ensure stochastic convergence in the sense that the standard error is below a certain threshold. This requirement is not in contradiction with the usage of stochastic sampling, because PRNGs are, in essence, combined recurrence relationships following a deterministic algorithm to calculate random numbers based on the current state and to move from one state to the next (see Section 3.2.2). It is however not completely straightforward to achieve this requirement on a distributed compute cluster with many individual nodes and processors, and therefore the theory and implementation decisions behind the framework are outlined in the following.

3.3.1 Computational Infrastructure

The uncertainty treatment framework has been integrated into a risk assessment framework developed by the Natural Catastrophe Modeling Group at MunichRe (see Section 1.1).

The framework is designed to run on a modern cluster compute environment, managed running Apache Spark³ on Apache Mesos⁴ with dynamic resource allocation. Individual worker processes run in Docker containers⁵ and the exposure data is piped to the MATLAB process using a custom binary inter-process communication (IPC) layer. Parallelization is performed using tiling of exposure and events.

3.3.2 Selection of PRNG Substreams using Hash Functions

In a cluster environment with dynamic resource allocation, it can not be taken for granted that input data will be partitioned identically when a computation is repeated in the same or a similar way at a later point in time. Exposure and events might be ordered differently than the first time and individual worker processes might receive different tiles of exposure and events. There might even be a different number of total worker nodes assigned to the same job.

For stochastic simulation requiring exact reproducibility, this poses special problems. Starting from a given state, RNGs always generate same sequence which might e.g. be used to stochastically select quantiles of various uncertainty distributions. With variable input data partitioning and ordering, these quantiles would however inevitably be assigned to different event and exposure tiles, which would imply obtaining different results. It is thus necessary to select a state of a RNG depending on the combination of an event and exposure. This is achieved by choosing a substream of MRG32K3a (see Section 3.2.2) based on the exposure properties of an individual risk item. For some sampling procedures such as sampling location uncertainty, it is necessary to obtain a repeatable sequence independent of the current event. In the framework, an individual risk item is uniquely identified by:

- occupancy type (type of usage, e.g. school, office, hospital)
- construction type (load bearing structural systems)
- year of construction
- latitude and longitude
- geographic/administrative zone
- geographic resolution of original data.

When random number sequences need to be obtained based on the combination of an event and exposure, these data are furthermore supplemented by the event group index.

While the year of construction, geographic resolution, occupancy type, construction type are encoded as integers, latitude and longitude are floating point numbers, and the

³Apache Spark is a unified analytics engine for big data. See <https://spark.apache.org>.

⁴Apache Mesos abstracts CPU, memory, storage and other resources. See <http://mesos.apache.org>.

⁵Docker provides operating-system-level virtualization, see <https://www.docker.com>.

geographic zone is a string. It was thus necessary to develop a method to map from these heterogeneous data to the $1.8 \cdot 10^{19}$ statistically independent substreams supported by MRG32K3a. A **hash function** does exactly this: it maps from data of arbitrary size to data of fixed size. The framework supports three suitable hash functions:

- Donald Knuth's hash proposed in Chapter 6.4 of *The Art of Computer Programming Volume 3* (Knuth, 1997)
- *SDBM Hash*⁶
- *Message-Digest Algorithm 5* (MD5)⁷.

Donald Knuth's hash and the SDBM hash are both simple and fast hash functions based on a few bit-wise operations. For example, the SDBM hash works by adding up the original bit-sequence shifted to the left by 6 bits and the original sequence shifted to the right by 16 bits, and subtracting the original sequence:

```
1 unsigned int SDBMHash(char* str, unsigned int length) {
2     unsigned int hash = 0;
3     unsigned int i = 0;
4
5     for (i = 0; i < length; str++, i++)
6     {
7         hash = (*str) + (hash << 6) + (hash << 16) - hash;
8     }
9
10    return hash;
11 }
```

For its relative simplicity, SDBM was found to have a good distribution and few collisions, i.e. it seldom happens that two different input data are mapped to the same output. MD5 is a more expensive algorithm which was initially designed for encryption. While MD5 is no longer considered cryptographically secure, due to relatively few collisions it suffices the requirements for an uncertainty sampling framework.

In the framework introduced in this section, the MD5 hash is always used to select a substream of MRG32K3a for the sampling of location uncertainty. For this uncertainty type it is especially important to avoid collisions, because due to the underlying cluster and datastore architecture it can only be sampled in advance for the entire stochastic event set and not per individual event. For the sampling of ground motion and damage uncertainty, the faster SDBM hash is used by default but all supported hash functions can be selected. In order to support both repeatable as well as stochastically independent results for repeated simulations, a fixed or a random seed can be set in the job parameter file of the framework.

⁶This is the hash function implemented in the public domain database library SDBM. See <http://www.cse.yorku.ca/~oz/sdbm.bun> and <http://www.cse.yorku.ca/~oz/hash.html>.

⁷See RFC 1321: <https://tools.ietf.org/html/rfc1321>.

3.4 Framework for Stochastic Simulation of Ground Motion Uncertainty

The modeling of the likely impact of earthquakes on insured portfolios requires not only computation of the expected ground motion but also involves evaluation and simulation of ground motion uncertainty, which can have a strong impact on the results. By design, GMPEs model the marginal distribution of one ground motion parameter at a single site. To accurately analyze seismic risk for heterogeneous portfolios consisting of different rise types (see Section 2.2.4), it is necessary to use a vector of multiple ground motion parameters corresponding to Peak Spectral Acceleration (PSA) at different response periods with correlated residuals (Bazzurro and Cornell, 2002; Baker and Cornell, 2006). Spatial cross-correlation — i.e. the simultaneous treatment of period-to-period and spatial correlation — is important when assessing the impact of earthquake events on a spatially clustered portfolio with heterogeneous building types, such as a portfolio distributed over a typical city-scale region consisting of risk items with different rise types (Weatherill et al., 2015).

The framework described in this chapter supports evaluation of uncorrelated and spatially correlated ground motion uncertainty, as well as spatially cross-correlated ground motion uncertainty using either the linear model of coregionalization or full block cross-correlation methodologies. Table 3.1 gives an overview of supported ground motion correlation models.

Correlation Model	Type
Jayaram and Baker (2009)	spatial
Baker and Cornell (2006)	period-to-period
Loth and Baker (2012)	LMCR
Jayaram and Baker (2009) with Baker and Cornell (2006)	FBCC

Table 3.1: Ground motion correlation models supported by the framework (LMCR = linear model of coregionalization. FBCC = full block cross correlation.).

3.4.1 Uncorrelated Ground Motion Uncertainty

Recall Equation 2.7, which for the i th site and j th event models the ground motion parameter y_{ij} as

$$\ln(y_{ij}) = \ln(\bar{y}_{ij}) + \epsilon_{ij} + \eta_j, \quad (3.10)$$

with \bar{y}_{ij} the predicted median ground motion intensity, ϵ_{ij} the intra-event residual, and η_j the inter-event residual. ϵ_{ij} and η_j are normal random variables with zero mean and standard deviation σ_{ij} and τ_j , respectively.

The new event-based uncertainty framework numerically simulates Equation 3.10, sampling ground motion uncertainty on all locations and for all GMPEs used by a seismic hazard model. First, the standard normal distribution is sampled to obtain standard normal random variates μ_{ij} and ν_j . Stochastic realizations of ϵ_{ij} and η_j are then obtained by multiplication of the standard normal random variates with the corresponding standard deviation as returned by the GMPE:

$$\epsilon_{ij} = \xi_{ij}\sigma_{ij} \quad (3.11)$$

$$\eta_j = \nu_j\tau_j \quad (3.12)$$

Depending on the settings of the computation, ξ_{ij} and ν_j and therefore also ϵ_{ij} and η_j can be truncated normal random variates, obtained via the algorithm of Botev (2017) or inverse transform sampling in conjunction with an efficient approximation of the quantile function of the normal distribution (Beasley and Springer, 1977; Moro, 1995; Glasserman, 2003; see Section 3.2.3). A truncation of the normal distribution is arguably justified for physical reasons, but based on current data it is not clear which truncation level to adopt (see Section 2.1.4).

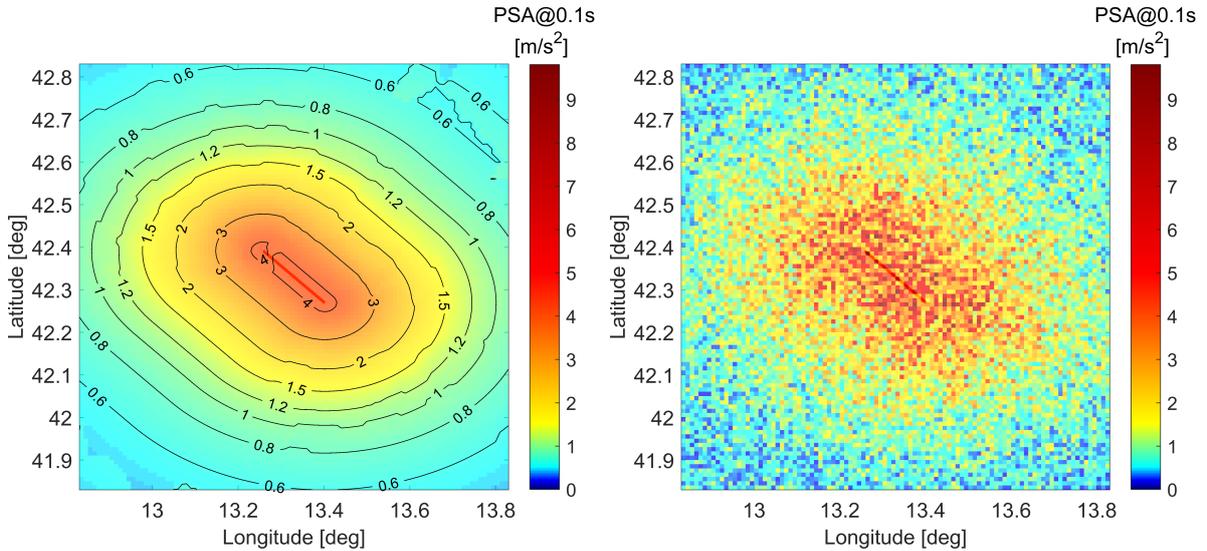


Figure 3.1: Footprints of mean ground motion (left) and ground motion with sampled uncorrelated intra-event uncertainty (right) for the M_w 6.3 2009 L'Aquila event. Color shows Peak Spectral Acceleration (PSA) in m s^{-2} at a response period of $T = 0.1$ s obtained using the ground motion model of Chiou and Youngs (2008). The surface trace of the rupture is superimposed as a red line, black isolines in the left plot indicate equal levels of ground motion.

To provide an illustration, Figure 3.1 shows a footprint of mean ground motion and a footprint of ground motion with sampled uncorrelated intra-event uncertainty for the M_w 6.3 earthquake which occurred near L’Acquila in central Italy on April 6th, 2009. The GMPE by Chiou and Youngs (2008) has been used to obtain PSA at a response period of $T = 0.1$ s.

3.4.2 Spatially Correlated Ground Motion Uncertainty

Many portfolios in the insurance industry contain risk items which are spatially clustered, for example in areas of high population density or large industrial plants. In these cases, the modeling of the spatial correlation of the intra-event ground motion residuals (ϵ_j) at a collection of sites is of high importance to correctly assess the likely impact during an event, since it can increase the likelihood of particularly high — and sometimes low — losses for spatially clustered portfolios (Weatherill et al., 2015). Spatial correlation between sites is caused by several reasons, such as common source effects (e.g. similar stress-drop or slip pattern) or common path effects because seismic waves travel a similar path from the rupture to nearby sites (Jayaram and Baker, 2010).

Geostatistics and Spatial Correlation Models for Ground Motion Uncertainty

Because the intra-event residuals $\epsilon_j = (\epsilon_{1j}, \epsilon_{2j}, \epsilon_{nj})$ have been shown to follow a multivariate normal distribution (Jayaram and Baker, 2008), the distribution of ϵ_j can be fully defined by its mean and standard deviation, and the correlation between all ϵ_{i_1j} and ϵ_{i_2j} pairs. Because the residuals are centered around zero, the mean of ϵ_j is $\mathbf{0}$. While the standard deviation is part of the GMPE, usually defined as a function of the spectral period and distance to the epicenter, the correlations are not.

In the literature several spatial correlation models are proposed, usually derived from empirical data with geostatistical tools. Ground motion is modeled as a stationary random process $Z(\mathbf{x}) = \mu + \epsilon(\mathbf{x})$ with mean μ and residual ϵ (compare Equation 3.10). Using the assumption of intrinsic stationarity

$$\mathbb{E}[Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})] = 0, \tag{3.13}$$

i.e. that the expected differences between two sites \mathbf{x} and $\mathbf{x} + \mathbf{h}$ separated by a distance vector \mathbf{h} are zero, a semi-variogram can be used to model the correlation structure (Oliver and Webster, 2014):

$$\gamma(\mathbf{h}) = \frac{1}{2} \mathbb{E}[\{Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})\}^2], \tag{3.14}$$

where Z is a random variable corresponding to ground motion, and $Z(\mathbf{x})$ and $Z(\mathbf{x} + \mathbf{h})$ its values at two sites.

Using the method of moments, an experimental variogram can be obtained from a finite number of realizations $z(\mathbf{x}_i)$ corresponding to data recorded at sites \mathbf{x}_i as (Matheron, 1965):

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} \{z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})\}^2, \quad (3.15)$$

with $m(\mathbf{h})$ the number of paired comparisons at distance \mathbf{h} . To obtain a continuous model for the correlation as a function of *lag distance* $h = |\mathbf{h}|$, a smooth curve is fit to the experimental variogram $\hat{\gamma}$. For example, a popular model is the isotropic spherical-plus-nugget model:

$$\gamma(h) = \begin{cases} c_0 + c \left\{ \frac{3h}{2r} - \frac{1}{2} \left(\frac{h}{r} \right)^3 \right\}, & \text{for } 0 < h \leq r, \\ c_0 + c, & \text{for } h > r, \\ 0, & \text{for } h = 0, \end{cases} \quad (3.16)$$

with c_0 the nugget variance, c the spatially correlated variance, and r the range.

Most spatial correlation models for intra-event ground motion residuals ϵ_j (e.g. Wang and Takada, 2005; Goda and Atkinson, 2009; Jayaram and Baker, 2009; Garakaninezhad et al., 2017) fit an exponential model of the form

$$\gamma(h, T) = \exp[a(T)h^{b(T)}], \quad (3.17)$$

with h the distance, and a and b coefficients depending on the period T .

The spatial correlation model by Jayaram and Baker (2009) which was implemented for this framework distinguishes between regions where the values of V_{S30} (the shear wave velocity in the upper 30 m) show or do not show clustering. It models the correlation γ between two sites separated by a distance h as

$$\gamma(h) = \exp(-3h/b), \quad (3.18)$$

where

$$b = \begin{cases} 8.5 + 17.2T, & \text{for } T < 1 \text{ s and } V_{S30} \text{ clustering,} \\ 40.7 - 15.0T, & \text{for } T < 1 \text{ s and no } V_{S30} \text{ clustering,} \\ 22.0 + 3.7T, & \text{for } T \geq 1 \text{ s independent of } V_{S30} \text{ clustering,} \end{cases} \quad (3.19)$$

where T is the spectral period.

Using the exponential model for the spatial correlation implicitly assumes that the intra-event ground motion residuals are joint normally distributed and that the field is isotropic and homogeneous (Weatherill et al., 2015). These assumptions have already been shown to be reasonable (Wang and Takada, 2005; Baker and Jayaram, 2008; Jayaram and Baker, 2009).

Simulation of Spatially Correlated Random Fields

The most common methodology to simulate a spatially correlated normal random field is to use the classical decomposition approach as described by Davis (1987):

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{L}\mathbf{Z}, \quad (3.20)$$

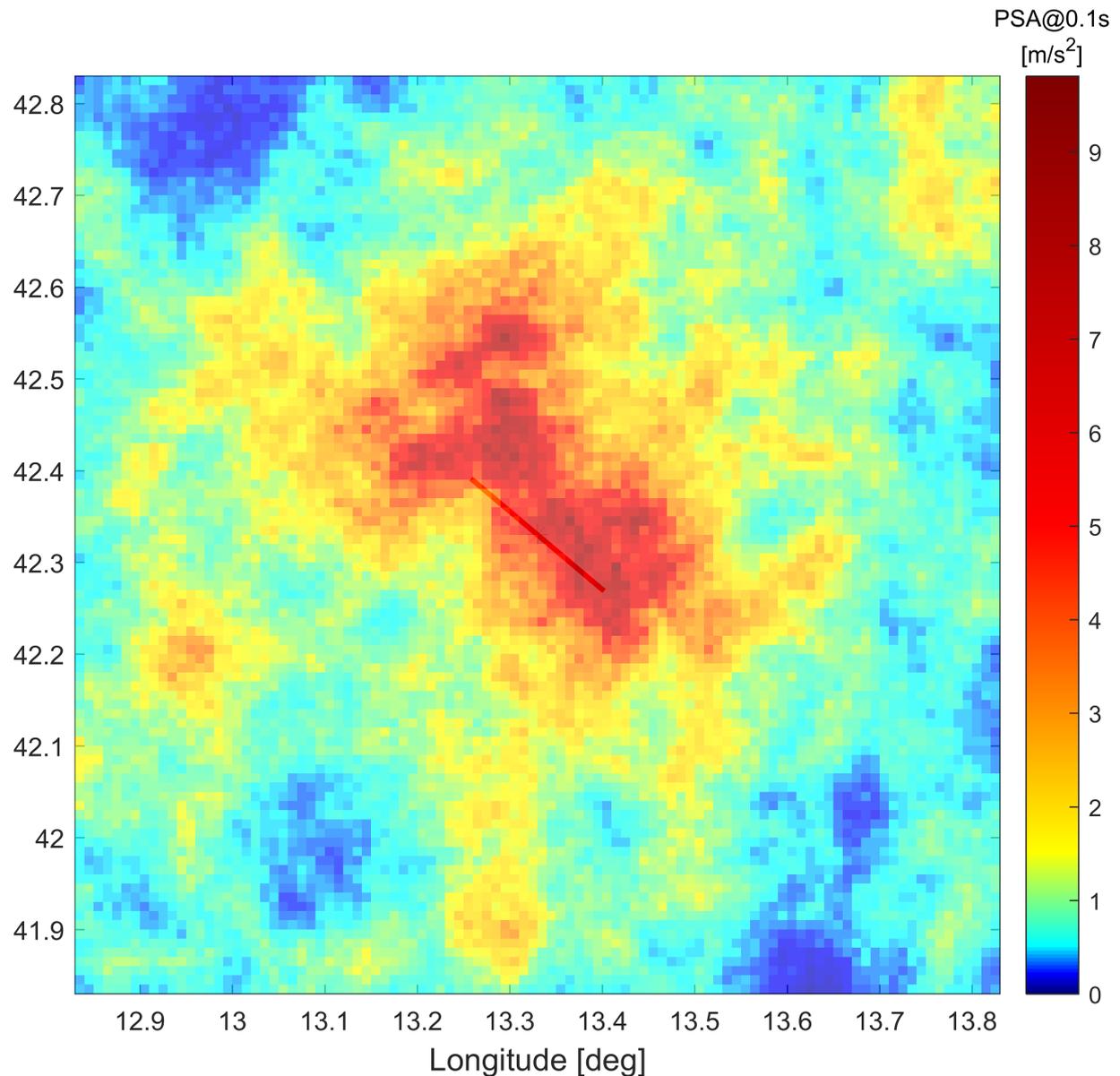


Figure 3.2: Footprint of Aquila event with spatially correlated ground motion uncertainty. This shows Peak Spectral Acceleration (PSA) in m s^{-2} at $T = 0.1\text{s}$ using the ground motion model of Chiou and Youngs (2008) and the spatial correlation model by Jayaram and Baker (2009) with V_{S30} clustering.

where \mathbf{Y} is the resulting vectorized random field with mean μ , \mathbf{Z} is a vector of IID standard normal random variables, and \mathbf{L} is the lower-triangular matrix obtained from the Cholesky decomposition

$$\mathbf{L}\mathbf{L}^T = \mathbf{C}, \quad (3.21)$$

with \mathbf{C} being the symmetric positive definite (SPD) correlation or covariance matrix:

$$\mathbf{C} = \begin{bmatrix} 1 & \gamma(h_{1,2}) & \gamma(h_{1,3}) & \cdots & \gamma(h_{1,n}) \\ & 1 & \gamma(h_{2,3}) & & \gamma(h_{2,n}) \\ & & \ddots & \ddots & \vdots \\ & & & 1 & \gamma(h_{n-1,n}) \\ \text{sym.} & & & & 1 \end{bmatrix}, \quad (3.22)$$

and $\gamma(h_{i,j})$ the correlation or covariance between two sites separated by a distance $h_{i,j}$.

To illustrate the effect of spatial correlation, Figure 3.2 shows a footprint for the same April 6th 2009 L'Acquila M_w 6.3 earthquake as in Figure 3.1. The GMPE by Chiou and Youngs (2008) has been used to obtain PSA at a response period of $T = 0.1$ s in combination with the spatial correlation model by Jayaram and Baker (2009).

3.4.3 Spatially Cross-Correlated Ground Motion Uncertainty

For the analysis of portfolios which contain risk items with heterogeneous rise types (see Section 2.2.4), considering only a scalar ground motion parameter such as PSA at a single period is usually not sufficient to accurately model the likelihood of damage to the building stock, since the typical natural period of a building strongly depends on its rise types (see Table 2.2). In these cases, it is advantageous to employ vector-valued PSHA (Bazzurro and Cornell, 2002) using PSA at different response periods. Spectral GMPEs provide a marginal probability distribution for PSA at a period, but do not provide the joint probability distribution for a vector of PSA at multiple periods for a single site.

Baker and Cornell (2006) investigate observed spectral acceleration values for correlation and provide an analytical model for period-to-period correlation. For the analysis of portfolios with heterogeneous rise types which are also spatially clustered, it is furthermore important to simultaneously take spatial correlation into account.

In the framework described in this chapter, two methodologies are available to achieve this cross-correlation in a stochastic simulation: linear model of coregionalization (LMCR) and full block cross correlation (FBCC).

Linear Model of Coregionalization

Loth and Baker (2012) fit a LMCR (Matheron, 1982) to a set of experimental variograms and cross-variograms of total residuals $\boldsymbol{\tau} = \boldsymbol{\epsilon}_j + \boldsymbol{\eta}_j$. Using the algorithm proposed by Goulard and Voltz (1992), they fit the following functional form for the SPD

cross-correlation matrix \mathbf{C} :

$$\mathbf{C} = \mathbf{B}^1 \exp\left(\frac{-3h}{20}\right) + \mathbf{B}^2 \exp\left(\frac{-3h}{70}\right) + \mathbf{B}^3 \mathbb{1}_{h=0}, \quad (3.23)$$

where h is the distance, $\mathbb{1}$ and indicator function equal to 1 for $h = 0$, \mathbf{B}^1 , \mathbf{B}^2 , and \mathbf{B}^3 are standardized coregionalization matrices describing the correlation for short, long and zero distances, respectively. After vectorizing and appropriately concatenating $\mathbf{Y}_{i=1\dots N}$ and $\mathbf{Z}_{i=1\dots N}$ for all N fields corresponding to N spectral response periods, stochastic simulation can be performed according to Equation 3.20.

Because Loth and Baker (2012) fitted the LMCR to the total residual $\boldsymbol{\tau}$, the model should therefore also be applied to $\boldsymbol{\tau}$ when used for a simulation. The LMCR methodology does not separate spatial and period-to-period correlation, but instead provides a single model incorporating both correlation types. For other regions or events it might therefore be necessary to fit the coregionalization matrices again if differences in the associated correlation structure exist (Weatherill et al., 2015).

To illustrate the effect of spatial correlation using the LMCR, Figure 3.3 shows again footprints for the April 6th 2009 L’Aquila M_w 6.3 earthquake. The GMPE by Chiou and Youngs (2008) has been used to obtain PSA at response periods of $T = 0.1$ s and $T = 0.3$ s in combination with the cross-correlation model by Loth and Baker (2012). The

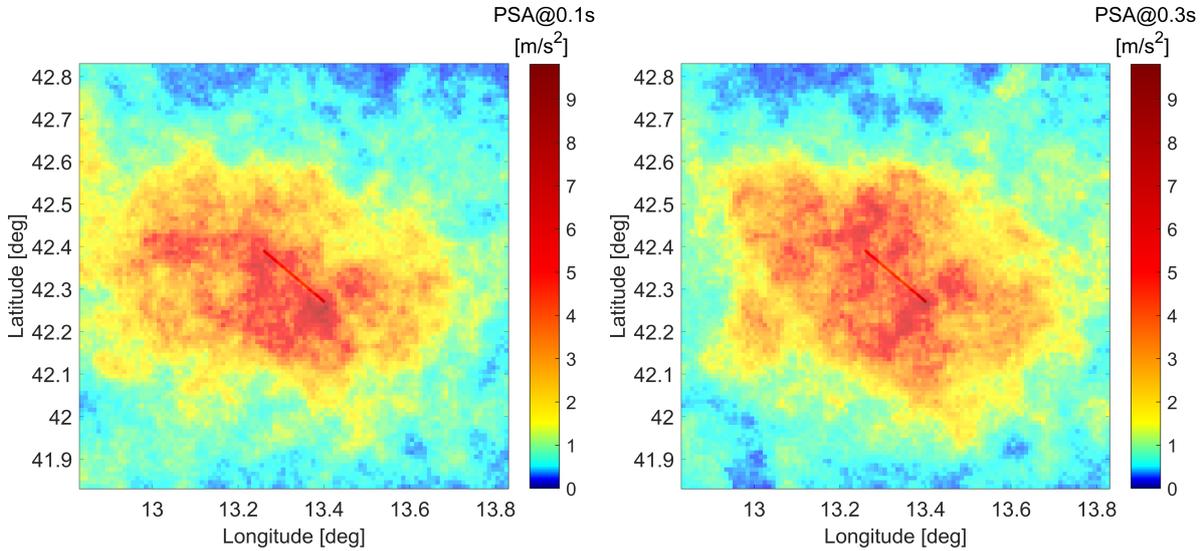


Figure 3.3: Footprints of the L’Aquila event with cross-correlated ground motion uncertainty via the linear model of coregionalization showing Peak Spectral Acceleration (PSA) in m s^{-2} at $T = 0.1$ s (left) and $T = 0.3$ s (right). The ground motion model of Chiou and Youngs (2008) with the cross-correlation model of Loth and Baker (2012) is used.

footprints corresponding to the two periods are cross-correlated, i.e. regions with large or small residuals at $T = 0.1$ s are more likely to have large or small residuals at $T = 0.3$ s, respectively.

Full Block Cross Correlation

The classical decomposition approach (see Equation 3.20) can be extended to be used for stochastic co-simulation of N vectorized random fields $\mathbf{Y}_{i=1\dots N}$ corresponding to N spectral response periods in the following manner (Oliver, 2003; Weatherill et al., 2015):

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_N \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_N \end{bmatrix} + \begin{bmatrix} \mathbf{L}_1 \mathbf{L}_1^T & \rho_{1,2} \mathbf{L}_1 \mathbf{L}_2^T & \cdots & \rho_{1,N} \mathbf{L}_1 \mathbf{L}_N^T \\ \rho_{1,2} \mathbf{L}_2 \mathbf{L}_1^T & \mathbf{L}_2 \mathbf{L}_2^T & \cdots & \rho_{2,N} \mathbf{L}_2 \mathbf{L}_N^T \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1,N} \mathbf{L}_N \mathbf{L}_1^T & \rho_{2,N} \mathbf{L}_N \mathbf{L}_2^T & \cdots & \mathbf{L}_N \mathbf{L}_N^T \end{bmatrix} \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \vdots \\ \mathbf{Z}_N \end{bmatrix}, \quad (3.24)$$

where $\mathbf{Z}_{i=1\dots N}$ are normal random variables, $\rho_{i,j}$ is the period-to-period correlation between the i th and j th response period, and \mathbf{L}_i is the lower-triangular matrix obtained from the Cholesky decomposition

$$\mathbf{L}_i \mathbf{L}_i^T = \mathbf{C}_i, \quad (3.25)$$

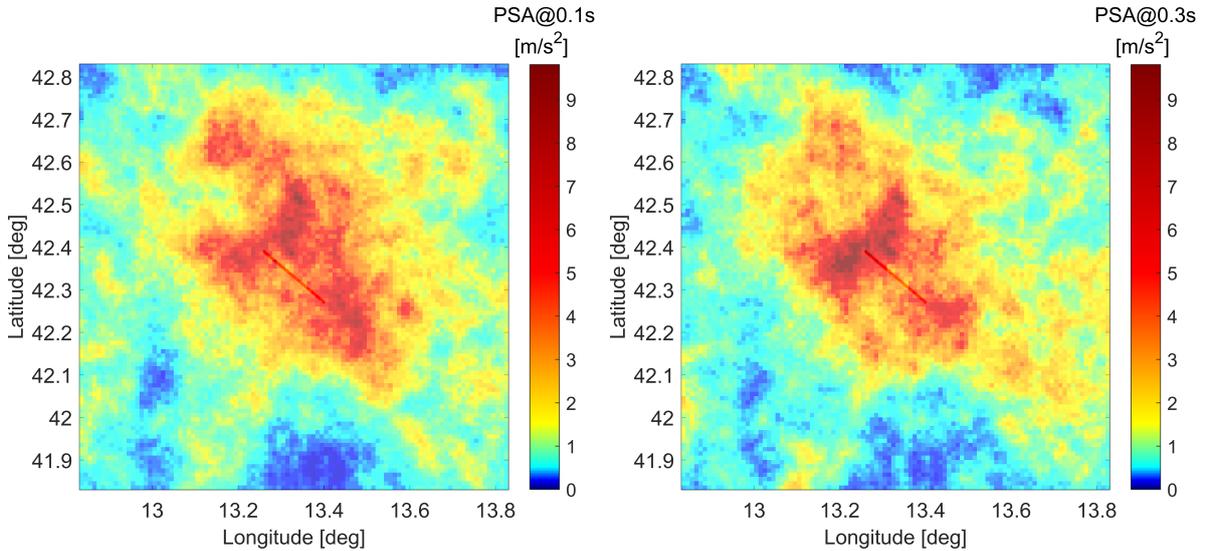


Figure 3.4: Footprints of the L’Aquila event with full block cross-correlated ground motion uncertainty showing Peak Spectral Acceleration (PSA) in m s^{-2} at $T = 0.1$ s (left) and $T = 0.3$ s (right). The ground motion model of Chiou and Youngs (2008) is used, with full block cross-correlation using the period-to-period correlation model of Baker and Cornell (2006) in conjunction with the spatial correlation model of Jayaram and Baker (2009) with V_{S30} clustering.

where \mathbf{C}_i is the spatial covariance or correlation matrix corresponding to the i th response period.

FBCC provides a general methodology to combine any spatial correlation model with any period-to-period correlation model. The spatial correlation remains unaltered for two equal periods i and j , but is reduced for $i \neq j$. In the framework described in this chapter, any available spatial correlation model can be combined with any period-to-period correlation model. In case the resulting cross-correlation matrix is not SPD, the Cholesky decomposition of the nearest SPD matrix in the Frobenius norm (Higham, 1988) is automatically used.

To illustrate the effect of spatial correlation using FBCC, Figure 3.4 shows footprints for the same April 6th 2009 L’Acquila M_w 6.3 earthquake as before. The GMPE by Chiou and Youngs (2008) has been used to obtain PSA at response periods of $T = 0.1$ s and $T = 0.3$ s using FBCC with the spatial correlation model of Jayaram and Baker (2009) to compute the spatial correlation matrices \mathbf{C}_i and the period-to-period correlation model of Baker and Cornell (2006) for the coefficients $\rho_{i,j}$. The footprints corresponding to the two periods are cross-correlated, i.e. regions with large or small residuals at one period are more likely to have large or small residuals at the other period, respectively.

3.5 Damage Uncertainty Framework

The framework implemented during the course of this thesis samples damage uncertainty given a particular mean damage ratio (MDR) and distributional parameters using the inverse transform method (see Section 3.2.3). The MDR of a risk item is given by evaluating the corresponding vulnerability function and depends on the local ground motion. The parameters of the damage distribution are defined as part of the vulnerability model (see Section 2.2.4).

3.5.1 Zero-One-Inflated Distributions

Real loss or damage ratio data from earthquake or other NatCat events usually includes many buildings that are not damaged at all or completely destroyed, corresponding to a damage ratio (DR) of 0 or 1, respectively. This can be modeled using a zero-one-inflated distribution (Ospina and Ferrari, 2010, 2012), which is a class of mixed continuous-discrete distributions. Zero-one inflated distributions consist of a continuous distribution on the open interval $(0, 1)$ and a Bernoulli distribution on $\{0; 1\}$:

$$f^*(x; \delta_0; \delta_1; \alpha; \beta) = \begin{cases} \delta_0, & \text{if } x = 0, \\ \delta_1, & \text{if } x = 1, \\ (1 - \delta_0 - \delta_1)f(x; \alpha; \beta), & \text{if } x \in (0, 1), \end{cases} \quad (3.26)$$

where $f(x; \alpha; \beta)$ is a continuous probability density function (PDF) on $(0, 1)$ with shape parameters α and β . The discrete probability mass for 0 and 1 is given by δ_0 and δ_1 , respectively. In this parameterization, the restriction $0 < \delta_0 + \delta_1 < 1$ applies.

3.5.2 Kumaraswamy Distribution

The Kumaraswamy distribution (Kumaraswamy, 1980; Mitnik and Baek, 2013) is a double bounded continuous probability distribution defined on the interval $[0, 1]$. It has a simple closed form for the PDF as well as the cumulative distribution function (CDF). The PDF is given by

$$f(x; \alpha; \beta) = \alpha\beta x^{\alpha-1}(1-x^\alpha)^{\beta-1}, \quad (3.27)$$

and the CDF by

$$F(x; \alpha; \beta) = 1 - (1-x^\alpha)^\beta. \quad (3.28)$$

The Kumaraswamy distribution can assume a wide amount of different shapes, which is illustrated in Figure 3.5.

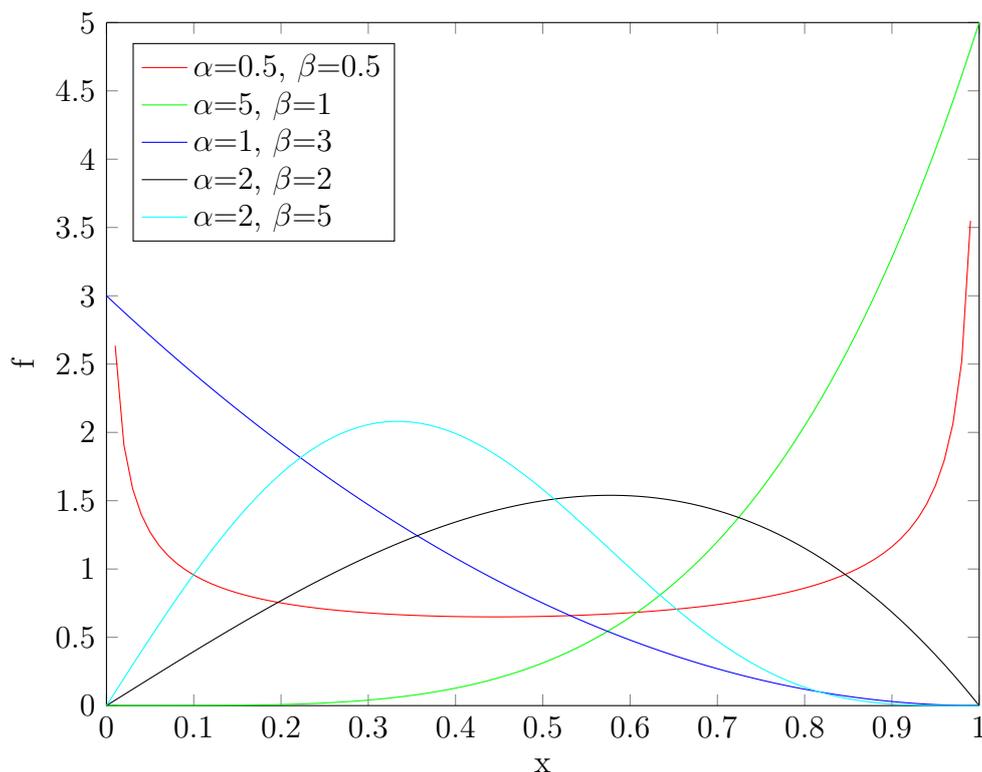


Figure 3.5: Probability density functions for different parameters of the Kumaraswamy distribution. This two-parameter distribution is very flexible and can assume a wide variety of different shapes. Illustration from Scheingraber (2015).

3.5.3 Zero-One-Inflated Kumaraswamy Distribution

For the damage uncertainty framework described in this chapter, a zero-one-inflated Kumaraswamy distribution according to Equation 3.26 in conjunction with Equation 3.27 has been used.

A plot of a possible PDF is shown in Figure 3.6. Vertical lines at $x = 0$ and $x = 1$ illustrate the discrete probability mass for no and full damage, respectively.

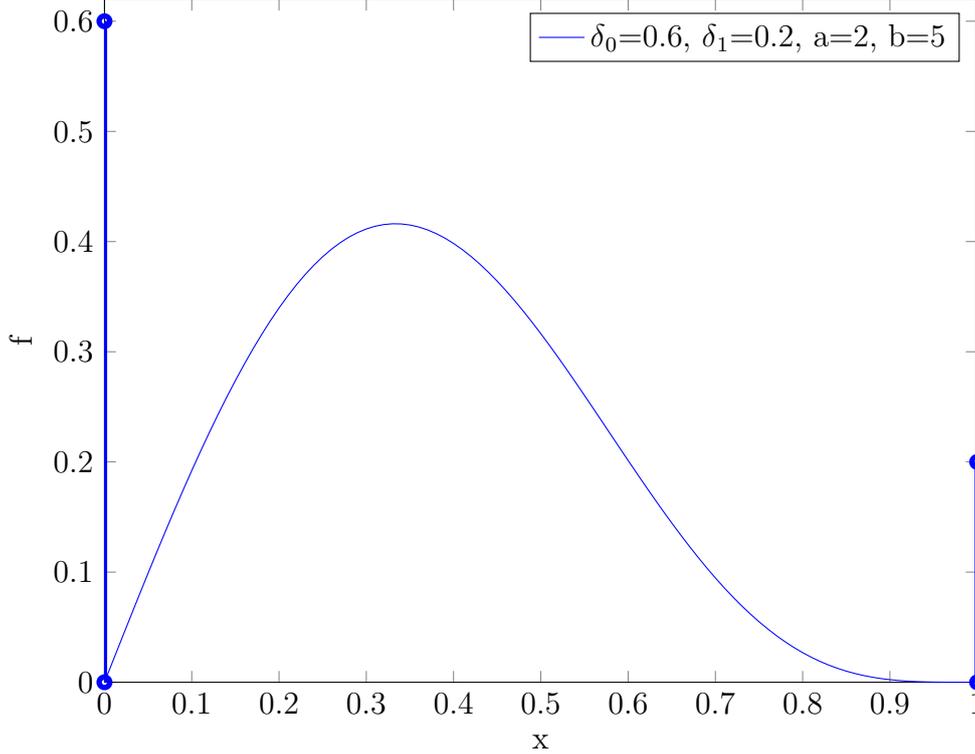


Figure 3.6: Probability density function of the zero-one-inflated (4-parameter) Kumaraswamy distribution. Note the probability density at $y = 0$ and $y = 1$ equals the probability mass given by δ_0 and δ_1 . Illustration from Scheingraber (2015).

The choice of the zero-one-inflated Kumaraswamy distribution to simulate damage uncertainty on the closed interval $[0, 1]$ with inflated probabilities at 0 and 1 allows for an efficient implementation using the simple closed form of the inverse CDF (quantile function) for inverse transform sampling:

$$f^*(x; \delta_0; \delta_1; \alpha; \beta) = \begin{cases} 0, & \text{if } x \leq \delta_0, \\ \left[1 - \left(1 - \frac{x - \delta_0}{1 - \delta_0 - \delta_1}\right)^{1/\beta}\right]^{1/\alpha}, & \text{if } x \in (0, 1), \\ 1, & \text{if } x \geq 1 - \delta_1, \end{cases} \quad (3.29)$$

where α and β are the shape parameters of the Kumaraswamy distribution, and δ_0 and δ_1 the inflated probabilities for no and full damage, respectively.

3.5.4 Damage Correlation

To support stochastic simulation of correlated damage, a multiple factor correlation model developed during a Master Thesis at Ludwig Maximilian University of Munich in collaboration with Munich Re (Reusch, 2017) has been integrated into the damage uncertainty framework.

The factor model generates correlated quantiles as

$$X_i = \sqrt{\rho_g}F_g + \sqrt{1 - \rho_g}F_{l_i}\epsilon_i, \quad (3.30)$$

where F_g is a per-event sample from the uniform distribution corresponding to the event residual, and F_{l_i} is a sample from the uniform distribution corresponding to the location residual, and ρ_g is the *global correlation parameter* controlling the strength of global correlation. For $\rho_g = 0$, the correlation model has no effect. If $\rho_g = 1$, the same quantile is used for all risk items.

The correlated quantiles X are used for inverse transform sampling instead of uniformly distributed quantiles (see Equation 3.9). Different coverages of a single risk item, e.g. a building and its content, always receive the same quantile. Note that if two risk items or coverages share the same quantile, this does not imply that they receive the same damage ratio, because the quantile is used with their individual damage distributions. The distributional parameters are likely different due to different mean damage ratios and vulnerability models (see Section 2.2.4). For further details and a study of the impact of the damage correlation model, please refer to Reusch (2017).

Chapter 4

The Impact of Portfolio Location Uncertainty in Probabilistic Seismic Risk Analysis

Leveraging the work of the previous chapter, we now add treatment of portfolio location uncertainty and analyze its impact in a realistic study of seismic risk in Indonesia.

This chapter has been published in slightly modified form as a peer-reviewed article in:

Scheingraber, C. and Käser, M. (2018).

The Impact of Portfolio Location Uncertainty on Probabilistic Seismic Risk Analysis.

Risk Analysis. *Advance online publication.*

<https://doi.org/10.1111/risa.13176>

Abstract

Probabilistic Seismic Risk Analysis is a well-established method in the insurance industry for modeling portfolio losses from earthquake events. In this context, precise exposure locations are often unknown. However, so far, location uncertainty has not been in the focus of a large amount of research.

In this chapter, we propose a novel framework for treatment of location uncertainty. As a case study, a large number of synthetic portfolios resembling typical real-world cases were created. We investigate the effect of portfolio characteristics such as value distribution, portfolio size, or proportion of risk items with unknown coordinates on the variability of loss frequency estimations.

The results indicate that due to loss aggregation effects and spatial hazard variability, location uncertainty in isolation and in conjunction with ground motion uncertainty can

induce significant variability to probabilistic loss results, especially for portfolios with a small number of risks. After quantifying its effect we conclude that location uncertainty should not be neglected when assessing probabilistic seismic risk, but should be treated stochastically and the resulting variability should be visualized and interpreted carefully.

4.1 Introduction

In the insurance industry Probabilistic Seismic Risk Analysis (PSRA) is widely used to model portfolios losses, usually expressed as an exceedance probability or loss-frequency curve. PSRA builds upon Probabilistic Seismic Hazard Analysis (PSHA, see Section 4.2). PSHA involves a large range of uncertainties, commonly categorized as being either aleatory or epistemic. The *aleatory* category is used for seemingly irreducible uncertainty associated with inherent variability of natural processes like ground motion propagation, modeled in the form of a probability distribution (Atik et al., 2010). The *epistemic* category refers to uncertainty caused by limited knowledge or data, usually captured using a logic tree (Senior Seismic Hazard Committee (SSHAC), 1997). Logic tree weights express the degree of belief in the correctness of a logic tree branch. The weights are determined using different strategies, often a combination of expert judgement and data-based approaches. For example, the choice of ground motion model has a strong effect on probabilistic earthquake loss, but significant reduction of this uncertainty is not foreseeable in the near future (Crowley et al., 2005). To overcome this, various ground motion models might be selected and weighted for the construction of a hazard model (Delavaud et al., 2012). It can also be advantageous to combine alternative source models applicable to a region (Weatherill and Burton, 2010). Arguably, the distinction of epistemic and aleatory uncertainty is somewhat arbitrary (Kiureghian and Ditlevsen, 2009), but it is considered helpful to identify reducible uncertainty.

Empirical ground motion models (also known as ground motion prediction equations; GMPEs) capture aleatory uncertainty using a log-normal probability distribution and often separate the total variability (σ_T) into inter-event variability (σ_E) and intra-event variability (σ_A) (Atkinson and Boore, 2003; Chiou and Youngs, 2008; Bindi et al., 2017). Modern GMPEs consider additional factors to account for more complex source characteristics and site effects (Bommer et al., 2003; Power et al., 2008), which has moved some aleatory uncertainty to the epistemic category. This has been found to significantly improve PSHA (Field and SCEC Phase III Working Group, 2000).

Notwithstanding the potential for further progress, the treatment and impact of uncertainty in the area of PSHA has been extensively researched. Conversely, uncertainty in PSRA has been studied less. A recent study has explored the effect of uncertainties in seismic hazard, site conditions, ground motion intensity conversion, and vulnerability relationship on the annual collapse probabilities of single buildings; uncertainty was propagated through the model to provide a range of probabilities instead of a single point estimate (Foulser-Piggott et al., 2017). In the context of PSRA of a spatially distributed portfolio, exposure uncertainty has already been identified as an area that would particu-

larly benefit from increased research attention (Crowley, 2014). In the present study, we are concerned with a specific type of epistemic exposure uncertainty highly relevant to the insurance industry: namely, that major parts of portfolios lack precise location data. For the natural catastrophe risk modeler, obtaining this information is often not possible. For many parts of the world, geocoding engines are not used systematically or can not reliably derive geographical coordinates from policy address information. In some cases, it might be too much effort for the primary insurer or the involved insurance broker — or sometimes there even seems to be little interest — to provide risk locations of large portfolios on coordinate level. In fact, the quality of this data also depends on the capabilities of the insurers' data management systems. This problem mainly exists for mass business with thousands to millions of risks, but also appears for smaller portfolios (i.e. less than several hundred risks) containing only few high-value exposures. However, administrative zones, such as postal codes, can usually be extracted automatically from the policy. In those cases, the associated polygon is the only constraint on the locus of risk items. Nonetheless, until recently, portfolio location uncertainty has apparently not been viewed as an important topic of research. It has sometimes been treated approximately, e.g. by modification of the aleatory ground motion variability (Stafford, 2012). A recent study has investigated the effect of geo-resolution on event loss (Bal et al., 2010). In that experiment, four different levels of geo-resolution were used to aggregate exposure data; ground motion was always modeled at the same resolution. This was performed for an idealized city as well as five scenario earthquakes on the same vertical strike-slip fault in the Marmara region south of Istanbul, Turkey. For these scenarios, the study found that a very low geo-resolution level does not imply a significantly biased loss result, but the results were associated with a higher standard deviation.

In the present chapter we look at a slightly different problem and take an alternative approach. We propose a framework for stochastic treatment of portfolio location uncertainty in detailed (“single-risk”) PSRA. We assume that the value and number of risk items is known, but precise coordinate-based location information is not known for the whole portfolio. In the insurance industry, this situation is encountered when insufficient location information is provided for an otherwise detailed portfolio, or when previous (e.g. third-party) deaggregation of a portfolio has occurred. Using a full event set for the region of western Indonesia, we investigate the effect of location uncertainty and loss aggregation and the extent of potential misjudgment if it is neglected.

Robust decision making and adaptive risk management under deep uncertainties requires effective communication and visualization of modeling results including the associated uncertainty (Tesfamariam et al., 2010; Cox, 2012). To address this, in this work we utilize novel visualization techniques to effectively communicate loss-frequency results and the associated uncertainty to decision makers.

4.2 Background: Seismic Hazard and Risk Analysis

As stated before, PSRA builds upon PSHA (Senior Seismic Hazard Committee (SSHAC), 1997; Cornell, 1968; McGuire, 2004). The central idea of PSHA is to express the exceedance rate of ground motion level y_0 at a site \mathbf{r}_0 by the hazard integral:

$$\lambda(y_0, \mathbf{r}_0)[y \geq y_0] = \int_V \int_{m_{min}}^{m_{max}} P[y \geq y_0 | m, \mathbf{r}, \mathbf{r}_0] \cdot \nu(m, \mathbf{r}) dmd\mathbf{r}. \quad (4.1)$$

Here, $\nu(m, \mathbf{r})dmd\mathbf{r}$ is the seismic rate density. $P[y \geq y_0 | m, \mathbf{r}, \mathbf{r}_0]$ is the conditional probability of exceeding ground motion y_0 at site \mathbf{r}_0 , given a rupture of magnitude m at source location \mathbf{r} . It is given by the complementary distribution function of a GMPE. The spatial integration volume V needs to contain all sources which can cause relevant ground motion at the site.

Assuming that the occurrence of earthquake events is a temporal Poisson process, the probability of at least one exceedance within time interval t_0 is given by

$$P(y_0, t_0, \lambda)[y \geq y_0] = 1 - e^{-\lambda t_0}, \quad (4.2)$$

where λ is the mean annual recurrence rate. In the insurance industry, common practice is to employ Monte-Carlo integration to solve Equation 4.1 (Pagani et al., 2014). In this approach, a stochastic event set (SES) is generated, which represents a possible seismicity history during time interval t_0 . For each event in the SES, a realization of the probabilistic ground motion model is explicitly computed, resulting in a set of ground motion fields $\hat{\mathbf{Y}}$. For PSRA, the Monte-Carlo simulation can then be taken one step further to obtain the probability that loss ι_0 is exceeded:

$$P(\hat{\mathbf{Y}}, \Theta)[\iota \geq \iota_0] = \sum_{i=1}^n \int_{\iota_0}^{\infty} f_{\iota}(\iota | \hat{\mathbf{Y}}_i, \Theta) d\iota. \quad (4.3)$$

Here, $f_{\iota}(\iota | \hat{\mathbf{Y}}_i, \Theta)$ is the loss probability density function for a portfolio Θ given ground motion field $\hat{\mathbf{Y}}_i$ of the i th event in the SES. The total exceedance probability is obtained by summing up the contribution of all n events.

$f_{\iota}(\iota | \hat{\mathbf{Y}}_i, \Theta)$ is a complicated function and is usually not expressed analytically. It depends on the chosen vulnerability models as well as on properties of the portfolio, such as construction types, spatial scale and spatial clustering of risk item locations, value distribution of risk items, and associated uncertainties. Equation 4.3 is used to obtain a loss exceedance probability curve. Using the assumption that temporal occurrence of earthquake events follows a Poisson process, a Probable Maximum Loss (PML) curve, showing loss against mean return period, can be obtained from the loss exceedance probability curve. Solving a first-order Taylor approximation of Equation 4.2 for the return period $T = 1/\lambda$ under the assumption $t_0 \ll T$ yields

$$T = \frac{t_0}{P(y_0, t_0, \lambda)[y \geq y_0]}, \quad (4.4)$$

where t_0 is the period of interest (time interval), which is 1 year for most reinsurance contracts.

4.3 Proposed Framework for Treatment of Location Uncertainty

4.3.1 Problem Statement

The non-linearity of the loss density function (Equation 4.3) renders analytical derivation of the influence of uncertainties such as portfolio location and spatial clustering unfeasible. While Monte-Carlo simulation in itself already is a computationally intensive method, the treatment of exposure uncertainty poses further challenges. Propagation of location uncertainty needs to be handled in an efficient way; in particular avoiding unnecessary, repetitive computation of hazard. In addition, it is advantageous to exploit known information, such as insured exposure density, which can be approximated by different means (e.g. population density).

4.3.2 Sampling Location Uncertainty on a Weighted Irregular Grid

The simple approach of aggregating risk items at the centroid of the polygon corresponding to the administrative zone introduces several problems, such as incorrect distance calculation and full correlation of the ground motion residual (Crowley, 2014). It also artificially introduces loss aggregation effects, because risk items are jointly affected by fewer events, instead of being affected separately by a larger number of events (see Section 4.6.1).

Presumably, it is more likely that a risk item with an unknown location is located in an area of high rather than low insured exposure density. We distribute risk items with unknown locations inside their respective administrative zones by randomly selecting indices of points on a weighted grid. The probability that a grid point is selected is proportional to its weight. Sampling occurs with replacement, using the *Mersenne Twister* pseudo-random number generator (Matsumoto and Nishimura, 1998). For industrial and commercial exposure, grid point weights are based on *GEG-2013*, a global exposure database produced for the Global Risk Assessment Report 2013 (GAR 2013), which provides a spatial resolution of 5 km. For the residential exposure, grid point weights are based on the global population database *LandScan* (Dobson et al., 2000), which has a grid resolution of 30 arcseconds (corresponding to approx. 1 km at the equator). These datasets combine census data with other higher resolution data, such as digital elevation models and land cover data. Figure 4.1 shows a map of population density around the Sunda straight, as an exemplary proxy for insured exposure density and the weighted grid. In the urban areas of Jakarta and Bandung, the grid point spacing decreases proportional to population density.

4.3.3 Shared Hazard Computation on Unique Risk Locations

For a PSRA, the computation of hazard usually dominates overall runtime. When using a large number of exposure location sets, i.e. a large sample size for location uncertainty,

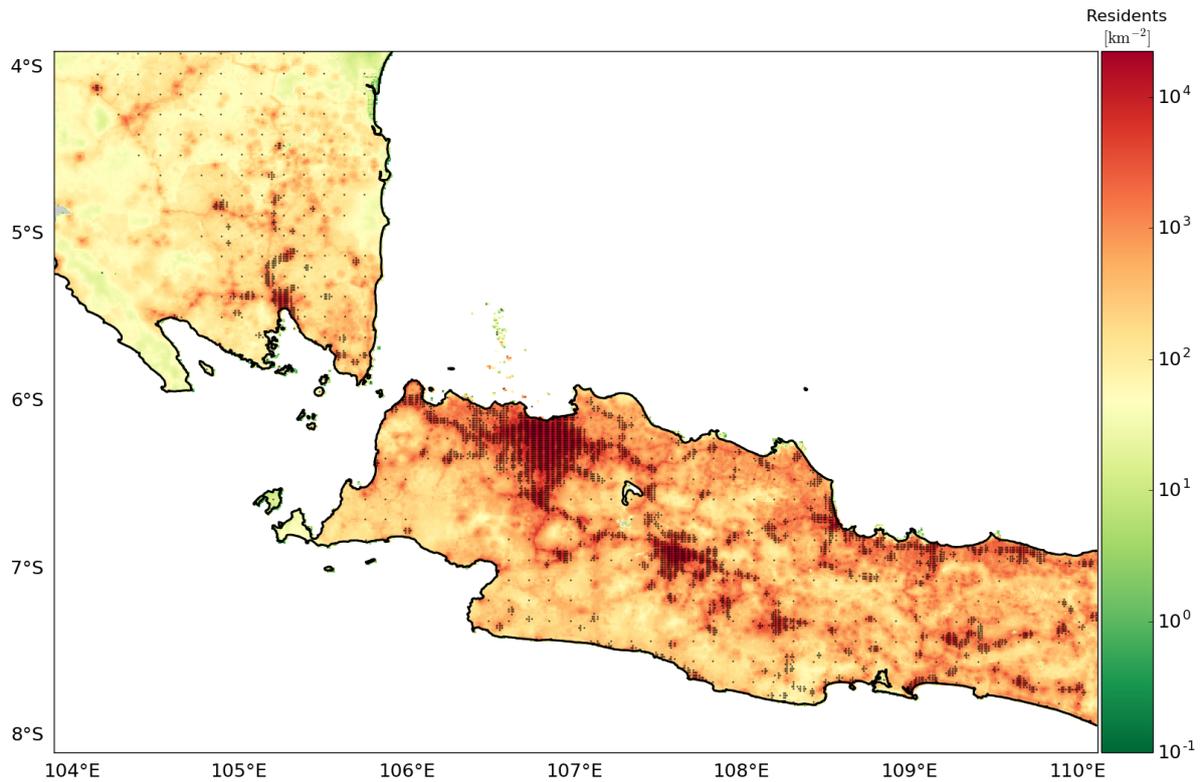


Figure 4.1: Detailed view of Sunda straight between the islands of Sumatra and Java. Color indicates population density (residents per km^2) as a proxy for insured exposure density. Due to licensing restrictions, the population data used for this plot is not the same as used for the study presented in Section 4.4, but a free dataset by Gaughan et al. (2015). Black markers depict grid points of the weighted grid.

it is likely that different exposure location sets share a significant fraction of common grid points. The straightforward approach — computing probabilistic loss for different location sets sequentially — would involve simulating hazard on shared grid points subsets numerous times. In order to increase computational efficiency, in our framework hazard is only computed once for all unique grid points used by all exposure location sets. This is achieved by maintaining an array of unique risk locations, and a mapping of location sets to unique risk location indices. Table 4.1 illustrates the concept. Probabilistic loss for all location sets can then be calculated based on the shared hazard computation. Note that this does not require to explicitly store the SES in memory (see Section 4.2), but Equation 4.3 can be evaluated incrementally following the simulation of each event.

Risk No.	Sample 1	Sample 2	Sample 3	Sample 4
1	43	13	31	51
2	23	28	98	18
3	98	98	98	98

Table 4.1: Risk Location Index Mapping Table. Rows correspond to individual risk items, showing sampled grid point indices. Each column represents a possible spatial distribution of the portfolio. Here, the last risk item has known coordinates and therefore is always located at the same grid point. Accidentally, the location set of the second risk item also shares this point.

4.3.4 Empirical Probable Maximum Loss Distributions

Monte Carlo methods are often employed to estimate expectations of functions $\phi(\mathbf{x})$ of a target density $P(\mathbf{x})$ in a parameter space \mathbf{x} . In this study, \mathbf{x} is the integration domain of the hazard and loss integrals, including the space of all possible risk item location sets for a portfolio Θ (see Section 4.2). $\phi(\mathbf{x})$ could be the mean, variance, or a quantile of PML at a return period as given by Equations 4.3 and 4.4. It is approximated from a sample of $P(\mathbf{x})$ using an estimator $\hat{\phi}(\mathbf{x})$. For example, we use order statistics with linear interpolation (Langford, 2006) to estimate empirical quantile PML curves.

4.4 Modeling Seismic Hazard in Western Indonesia

4.4.1 Study Area

We concentrate on western Indonesia, defined as the area enclosed by the 10° S and 6° N parallels and 95° E and 120° E meridians. This area includes the islands of Sumatra and Java, as well as Kalimantan, the Indonesian sector on the island of Borneo. Indonesia is divided into 34 provinces (administrative zones). Since exposure data is often provided on the basis of these zones, they are used for the analysis conducted in this chapter. According to a census commissioned by the Indonesian government, in 2010 the country had a population of 237.6 Million. The tremendous potential socio-economic impact of natural catastrophe events stems from a high concentration of the population in areas of significant seismic hazard. Figure 4.2 shows population density per province. By choosing the western part of the country, we study the areas of the highest population density.

Indonesia is a seismically highly active region (see Figure 4.3). Throughout recorded history, numerous earthquakes caused severe damage and considerable loss of life, such as the 1797, 1833 and 1861 Sumatra earthquakes, or the 1899 Seram earthquake. Recent events have been particularly consequential, such as the 2004 Sumatra earthquake and tsunami, causing over 230,000 fatalities; or the 2006 Yogyakarta earthquake, with over 5,700 fatalities.

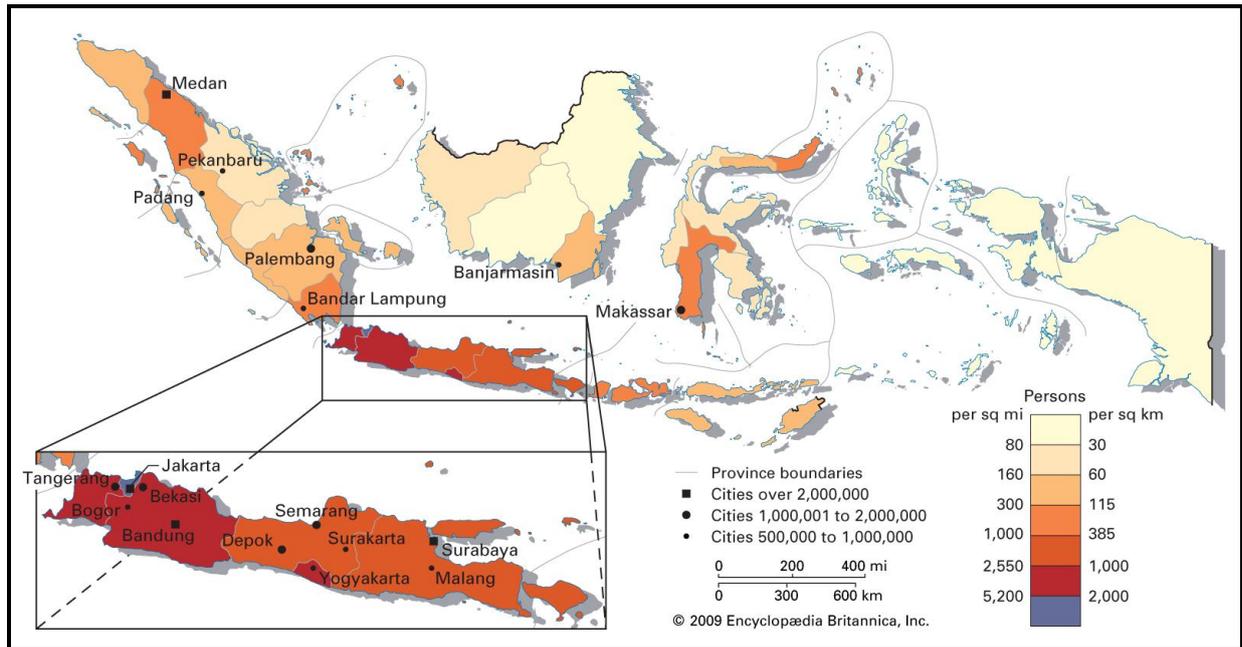


Figure 4.2: Population density in Indonesia. The population density is higher in the western part of the country, with the highest concentration on the island of Java. Map by Encyclopaedia Britannica (2009).

4.4.2 Seismo-Tectonic Setting

Located offshore from Sumatra, the Sunda arc marks the subduction of the Indo-Australian plate under the continental Sunda plate. This classic subduction system consists of the subducting slab along the Sumatra-Java trench, an accretionary wedge where the outer arc ridge forms the backstop, the Bengkulu-Mentawai forearc basins off Sumatra, and the Java forearc basin (Schlüter et al., 2002; Samuel et al., 1997). The largest fault of the country is known as Great Sumatran Fault. Almost 2000 kilometers long, this dextral strike-slip fault is part of the Sumatra fault zone.

4.4.3 Modeling the Seismic Hazard

For this study, a seismic hazard model of Indonesia was created. It is based on the South-East Asia hazard model by Mark Petersen et al. (2007) of the United States Geological Service (USGS), but with some modifications. Site conditions are, with refinements, based on topographic slope (Wald and Allen, 2007). The geometry of the Sumatra subduction has not been taken from the original USGS hazard model, but is a complex fault representation based on *Slab 1.0* (Hayes et al., 2012), which provides a three-dimensional representation of the subduction. For events on the complex fault the model uses a rupture floating mechanism similar to the implementation of OpenQuake (Pagani et al., 2014), a free and open-source seismic hazard and risk software developed as part of the Global Earthquake

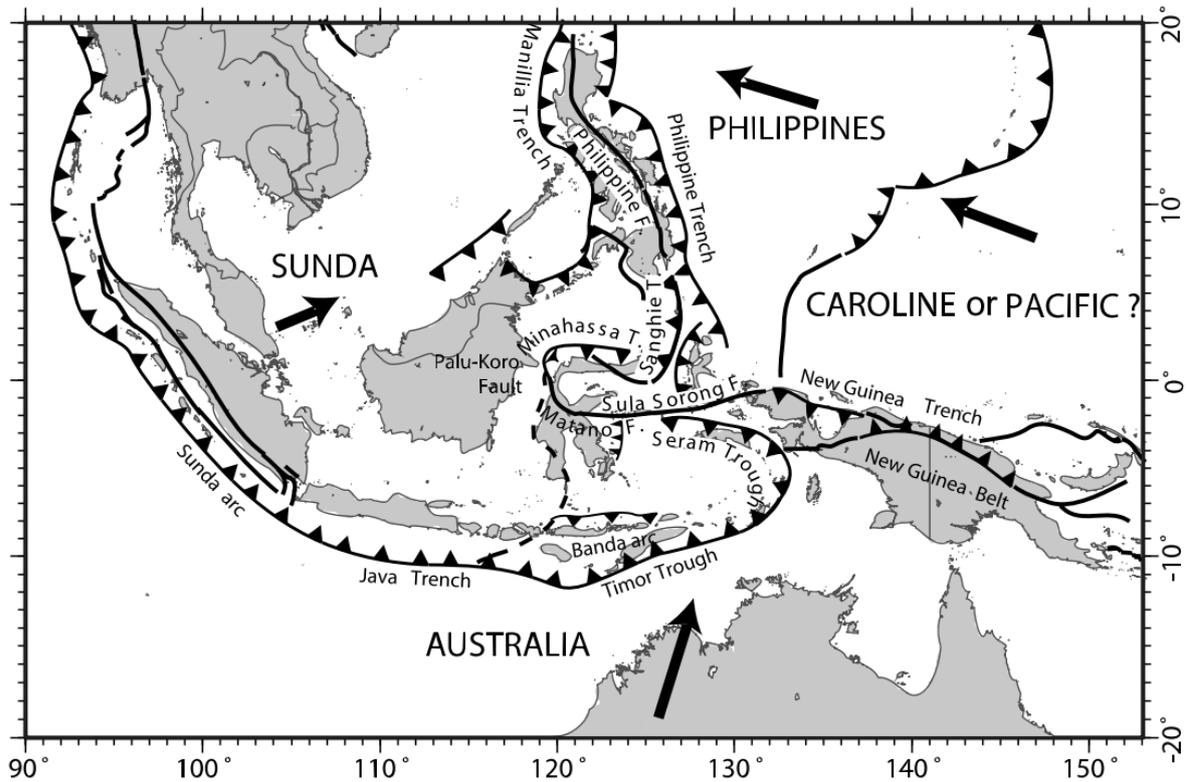


Figure 4.3: Tectonic plates in Indonesia. Arrows show far-field plate velocities with respect to Eurasia. Unmodified from Socquet et al. (2006).

Model (GEM) initiative (Crowley et al., 2013). To reduce computational demands, we have simplified the branch sets of all tectonic region types in the ground motion model logic tree. This allowed to conduct a larger case study of the effect of location uncertainty while retaining overall accordance with the hazard of the original model. Table 4.2 gives an overview of the selected ground motion models and their weights.

Recent work has shown that neglecting epistemic hazard uncertainty can lead to underestimation of hazard levels (Weatherill and Burton, 2010), potentially resulting in biased loss estimates and underestimation of central 70% inter-percentile ranges of important loss metrics (Kotha et al., 2018). In this study, we treat epistemic uncertainty by full logic tree enumeration of the source and ground motion models. On top of this, we investigate the isolated effect of location uncertainty in two controlled settings (Sections 4.6.1 and 4.6.2), as well as — in a more realistic example — in conjunction with aleatory ground motion uncertainty (Section 4.6.3).

Tectonic Region Type	Ground Motion Model	Weight
<i>Active Shallow Crust</i>	Boore and Atkinson (2008)	1/3
	Campbell and Bozorgnia (2008)	1/3
	Chiou and Youngs (2008)	1/3
<i>Stable Continental Crust</i>	Toro et al. (1997)	1
<i>Subduction Interface</i>	Zhao et al. (2006)	2/3
	Youngs et al. (1997)	1/3
<i>Subduction Intraslab</i>	Atkinson and Boore (2003)	1

Table 4.2: Ground motion models for different tectonic region types used by the hazard model.

Hazard Results

Figure 4.4 shows seismic hazard predicted by the model with an exceedance probability of 10% in 50 years, corresponding to an average return period of 475 years. This is the probability level most commonly shown in seismic hazard maps; it is also used by the engineering community for the design of building codes. The hazard map is in general agreement with the original USGS South-East Asia model (Mark Petersen et al., 2007) as well as results obtained using third-party implementations such as by GEM evaluated using the OpenQuake engine¹. On Sumatra, seismic hazard is high and dominated by the Sumatra fault zone (see Section 4.4.2). On Java, the highest hazard levels are obtained in the western part of the island, including the area around Jakarta. Kalimantan is characterized by low hazard levels. Differences to the aforementioned results arise from the inclusion of site conditions, simplified ground motion model logic tree, different source parameterization, improved subduction geometry representation, and slightly modified seismicity rates using the latest ISC-GEM Global Instrumental Earthquake Catalogue (Storchak et al., 2013) and Global Historical Earthquake Catalogue².

4.5 Case Study: Impact of Location Uncertainty in Western Indonesia

4.5.1 Creation of Synthetic Portfolios

For this study, we use synthetic portfolios representing real-world counterparts in terms of spatial distribution of risk items and value distribution among risk items.

¹See https://hazardwiki.openquake.org/sea2007_intro for results obtained using OpenQuake.

²See <http://www.emidius.eu/GEH/>.

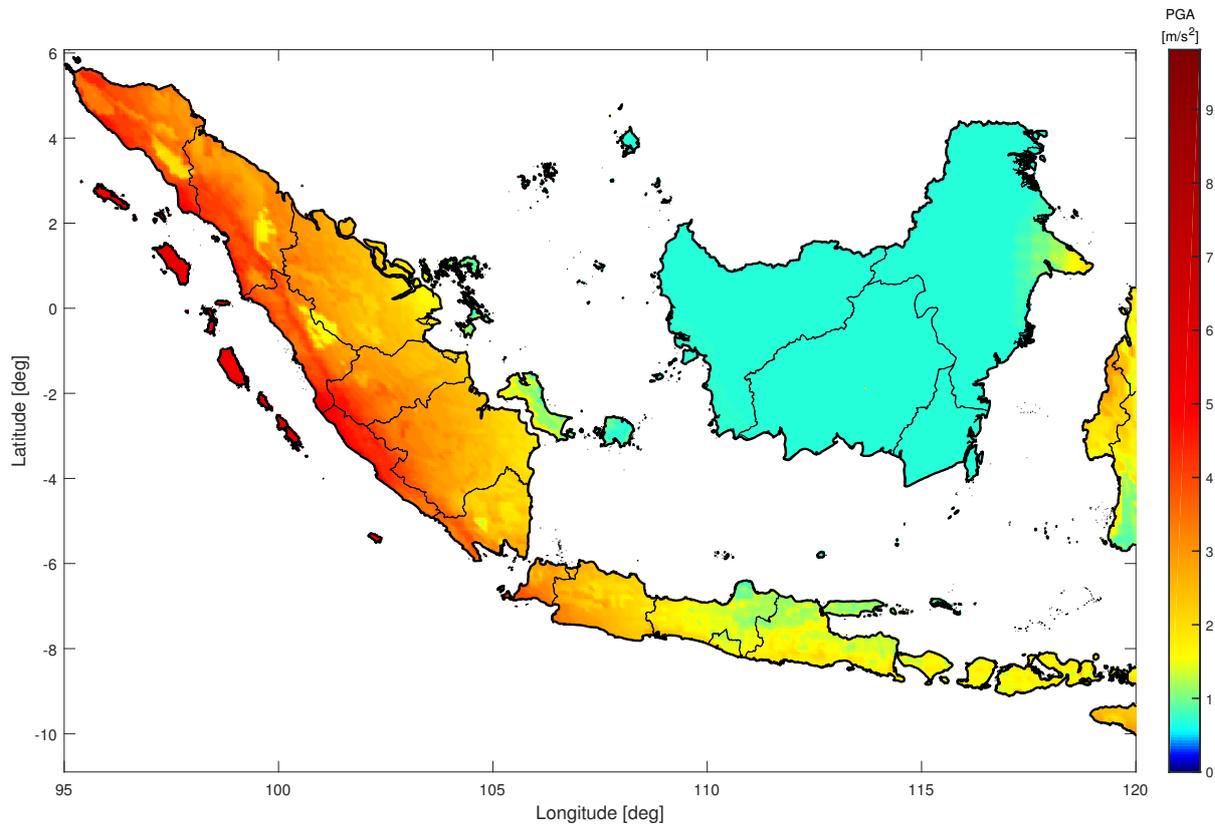


Figure 4.4: Seismic hazard in western Indonesia. Site conditions are based on topographic slope (Wald and Allen, 2007). The area encompasses the islands of Sumatra and Java as well as Kalimantan, the Indonesian sector on the island of Borneo. Color indicates the level of Peak Ground Acceleration (PGA, in m s^{-2}) that is predicted to be exceeded at an average return period of 475 years.

Value Distribution

For all portfolios, the total sum insured (TSI) is kept constant:

$$\text{TSI} = \text{const.} = 1 \cdot 10^6. \quad (4.5)$$

However, the TSI is distributed among a varying number of risk items (portfolio size), ranging from 1 to 10000. Two fundamentally different value distributions are used: *flat* and *mixed exponential*.

Flat Value Distribution In many real residential portfolios the TSI is divided almost equally among risk items. For simplicity, we model this as a completely flat value distribution:

$$\text{VI}_{\text{flat}} = \frac{\text{TSI}}{n}. \quad (4.6)$$

Here, n is the number of risk items and VI_{flat} (“value insured”) is the value assigned to each risk item.

Mixed Exponential Value Distribution To approximate the value distribution we observe in many country-wide facultative risk portfolios, we developed a “mixed exponential” distribution. Here, the TSI is split into two parts:

$$TSI_{\text{flat}} = \mu \cdot TSI \quad (4.7)$$

$$TSI_{\text{exp}} = (1 - \mu) \cdot TSI. \quad (4.8)$$

The weighting factor μ has been set to 0.3, which ensures that the value assigned to individual risk items does not converge towards zero and assume unrealistically small values for large i , with $i = 1, \dots, n$. TSI_{flat} is divided equally among all risk items according to Equation 4.6. TSI_{exp} is given by the probability density function of an exponential distribution spread across the total number of risk items n :

$$VI_{\text{exp},i}^* = \lambda \cdot e^{-\lambda i} \cdot X_i \cdot \frac{TSI_{\text{exp}}}{n}, \quad (4.9)$$

$$VI_{\text{exp},i} = \frac{TSI_{\text{exp}}}{\sum_{i=1}^n VI_{\text{exp},i}^*} \cdot VI_{\text{exp},i}^*. \quad (4.10)$$

The value distribution we observe in many real portfolios of this type is well represented by choosing a shape factor $\lambda = 1/3$. X_i is a uniform random number in the interval $[1 - p_{\text{exp}}, 1 + p_{\text{exp}}]$. Here, p_{exp} is a perturbation factor set to 0.1, which is consistent with the characteristics of many real portfolios. Equation 4.10 normalizes the n randomly perturbed insured values to ensure $\sum_{i=1}^n VI_{\text{exp},i} = TSI_{\text{exp}}$. Finally, the insured value of the i th risk item for the mixed exponential value distribution is given by

$$VI_{\text{mix},i} = VI_{\text{flat}} + VI_{\text{exp},i}. \quad (4.11)$$

See Figure 4.5 for a plot of insured values generated with this process.

Exposure Locations

For each combination of value distribution (flat and mixed exponential) and portfolio size (ranging from 1 to 10000 risk items), a set of 11 portfolios with an increasing fraction of unknown coordinates is created. For the first portfolio of a set, the coordinates of all risk items are known, randomly distributed over all administrative zones onto the weighted grid (see Section 4.3.2) according to the grid point weights. Coordinate-based location information is then successively removed and replaced by administrative zone information in steps of 10%. For the controlled “worst-case” example of the isolated effect of location uncertainty described in Section 4.6.2, the most valuable risk item is selected first for the removal of coordinate-based location information and less valuable risk items follow subsequently until all coordinates are unknown. In contrast, for the more realistic example including ground motion uncertainty presented in Section 4.6.3, risk items are selected at random for the removal of location information.

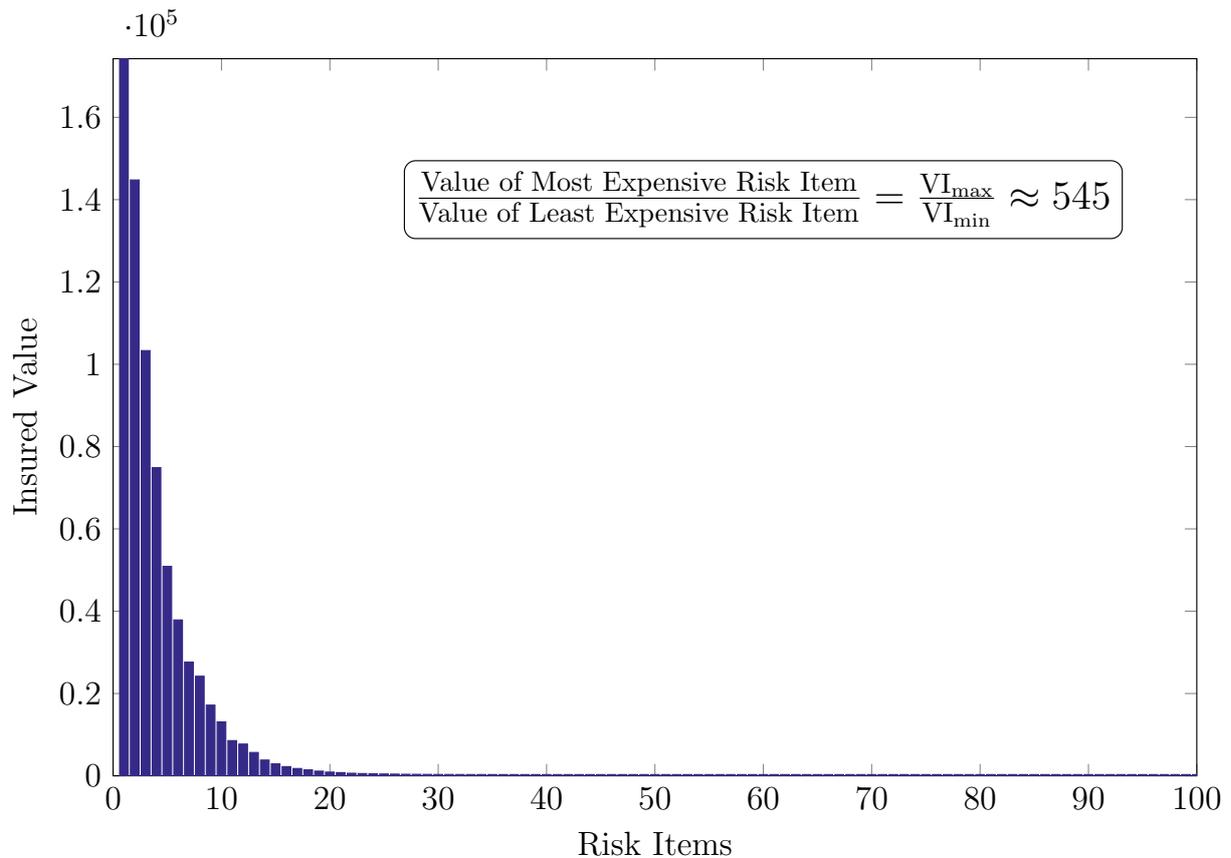


Figure 4.5: Insured values of risk items sorted in descending order for the mixed exponential value distribution. Each bar represents one risk item. Many real-world facultative risk portfolios feature this distribution type, which is characterized by an extremely uneven distribution of values: this plot only shows the 100 most valuable risk items of a portfolio of 1000 risk items. For the portfolio shown in this plot, the ratio of the insured value of the most expensive to the least expensive risk item is $\frac{VI_{\max}}{VI_{\min}} \approx 545$.

Vulnerability Model

As mentioned before, in this study we investigate the influence of location uncertainty in isolation as well as in conjunction with ground motion uncertainty. To eliminate unwanted effects due to vulnerability modeling decisions, we use the same vulnerability function for all risk items.

4.6 Impact of Location Uncertainty on Probabilistic Seismic Loss

For the portfolios described in Section 4.5.1, a series of Monte-Carlo simulations were performed using a sample size of 128. We first investigate the isolated effect of location uncertainty on loss with a flat hazard model in Section 4.6.1. In Section 4.6.2, we add spatial hazard variation by using the model described in Section 4.4.3, but still isolate location uncertainty from aleatory ground motion uncertainty. Finally in Section 4.6.3, we add aleatory ground motion uncertainty sampling to investigate the effect of location uncertainty in a realistic seismic risk analysis.

To present the result for individual portfolios, we plot PML and the associated uncertainty against mean return period obtained as described in Section 4.3.4. The plots show mean PML, lower and upper quartile PML, as well as the full PML range (minimum and maximum PML). These curves connect the respective estimated values between adjacent return periods. Note that the quantile curves do not necessarily correspond to one uncertainty realization, because PML curves of different portfolios can intersect. This is because a given portfolio might be particularly affected by seismic sources active at a specific range of return periods, while another portfolio might be more affected at another range of return periods.

To show a smooth plot of the PML distribution at a return period, we use kernel density estimation with a normal kernel function. To visualize the influence of portfolio properties, we plot the coefficient of variation (CV) of PML at a fixed return period against portfolio size or fraction of unknown coordinates. The CV is defined as

$$CV = \frac{\sigma}{\mu}, \quad (4.12)$$

where σ is the standard deviation and μ the mean.

4.6.1 Flat Hazard Model: Influence of Spatial Risk Item Clustering and Loss Aggregation

In this section, we explore the impact of location uncertainty on PML for portfolios with a flat value distribution using a hypothetical flat hazard model. The flat hazard model is characterized by regularly spaced gridded seismicity with a spatially constant magnitude frequency distribution over the whole area. For this example we perform no sampling of aleatory ground motion uncertainty, which allows to investigate the effect of the spatial distribution of risk items in an isolated manner.

We found that even in this case, location uncertainty can have a significant effect on portfolios consisting of more than one risk item. This can be explained by loss aggregation effects. When risk items of a sampled location set are spatially clustered, they are affected conjunctly by a small number of nearby events, resulting in higher event losses. Conversely, when risk items are spread over a wide area, they are affected by a larger number of events.

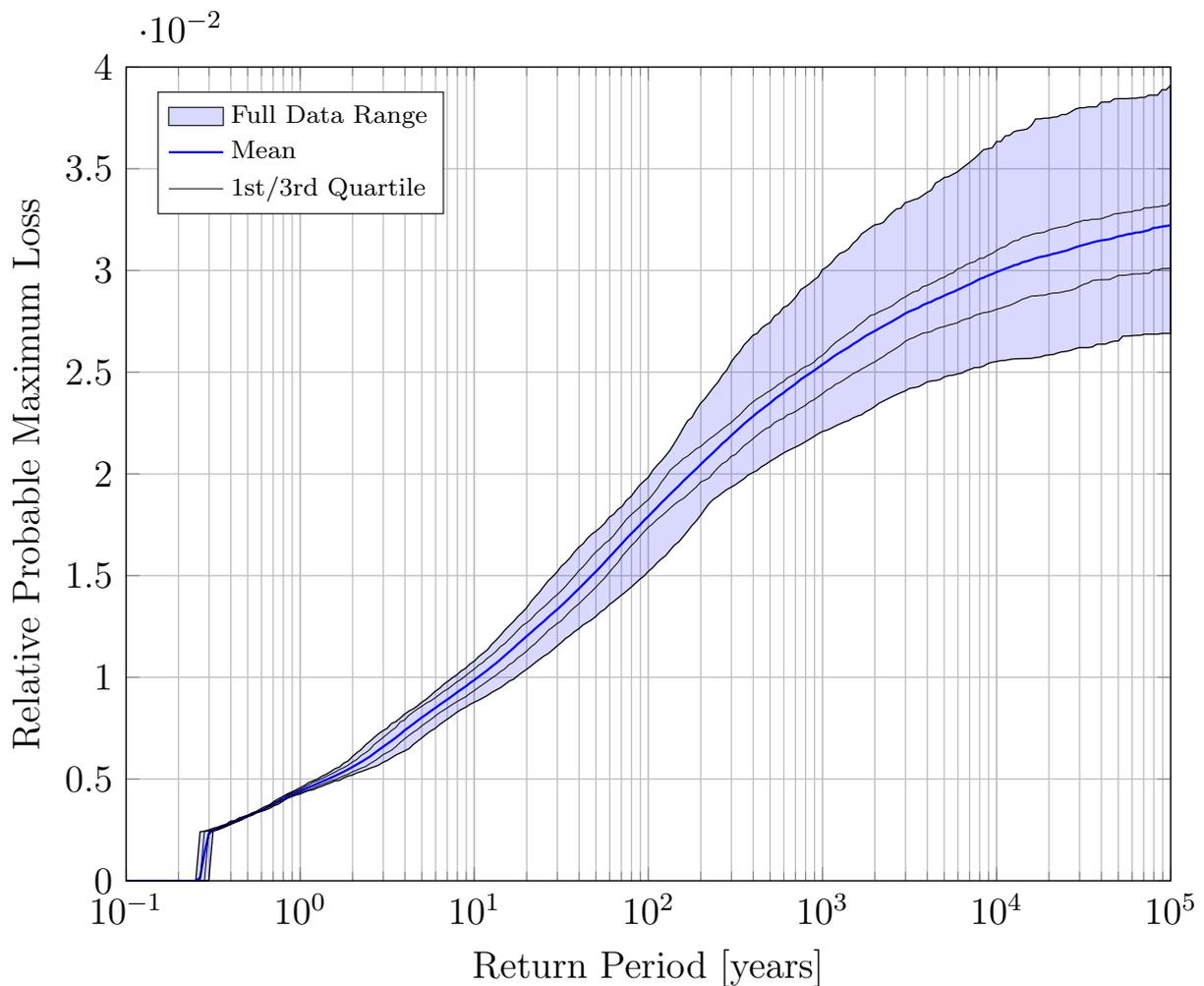


Figure 4.6: Probable Maximum Loss (PML) Curve and associated uncertainty distribution for a synthetic portfolio of 100 risk items with 80% unknown coordinates, using a flat value distribution as described in Section 4.5.1 and a flat hazard model without aleatory ground motion uncertainty sampling. The risk items with unknown locations represent 80% of the portfolio value. Location uncertainty was treated with the framework described in Section 4.3 using a sample size of 128. This shows the mean PML (blue line) as well as lower and upper quartile PML (black lines). PML values are relative to the total sum insured of the portfolio.

Individual events cause smaller losses, since they affect a smaller portion of the portfolio TSI.

For larger portfolios, the effect of location uncertainty was found to be smaller. To consider an extreme case, imagine a portfolio where the number of risk items is much larger than the number of grid points of the weighted grid (see Section 4.3.2). Almost all sampled locations sets for this portfolio will be fairly well distributed according to grid point

weights, therefore most location sets will experience similar loss. On the other hand, for very small portfolios it is more likely that the coordinates are known. As a representative example of loss aggregation, Figure 4.6 shows mean PML and the uncertainty distribution for a flat portfolio of 100 risk items with 80% unknown coordinates, which represents a quite common situation for large reinsurance treaty portfolios, where the majority of risk locations is only known by their administrative zone.

In general, PML uncertainty of a given portfolio relative to the mean PML is smaller for the flat hazard model than for the Indonesia model (see Section 4.6.2). Still, the results show that considerable loss uncertainty results purely from spatial risk item clustering. This fact is often overlooked, but should be kept in mind for any study of seismic risk to spatially distributed exposure.

4.6.2 Western Indonesia Controlled Example: Upper Bound of the Isolated Influence of Location Uncertainty

We now explore the impact of location uncertainty in the presence of spatial hazard variation, but still isolate it by not sampling ground motion residuals. To obtain an upper bound for the western Indonesia model (see Section 4.4.3), we remove the coordinates of risk items in decreasing order of value.

High Potential Impact for the Mixed Exponential Distribution

The value distribution of the portfolio has a strong effect. The potential impact of location uncertainty is higher for the mixed exponential than for the flat value distribution. As explained in Section 4.6.1, for very large portfolios location uncertainty tends to balance out, while for very small portfolios it is likely that the coordinates of all risk items are known. To select meaningful examples in this light, we show PML results for portfolios of 300 risk items with 20% unknown coordinates. For the flat value distribution, PML uncertainty is limited in this case (see Figure 4.7), since the risk items with unknown coordinates only account for 20% of the total portfolio value. The CV of PML is 1.7% at a return period of 10 years, 1.1% at 100 years, and 0.9% at 1000 years. By contrast, 20% of the risk items can represent up to 76% of the portfolio value for the mixed exponential distribution. Figure 4.8 shows this “worst-case” scenario. The CV of PML is 4.2% at a return period of 10 years, 4.9% at 100 years, and 4.8% at 1000 years. The range of PML values is significantly larger than for the flat value distribution. In general, also the mean PML at a given return period is higher. This can be explained by the concentration of value to a few risk items, which increases the likelihood that a high portion of the portfolio TSI is affected jointly by an event. Abrupt, steep jumps in the PML curves at higher return periods can be attributed to low-probability but high impact events.

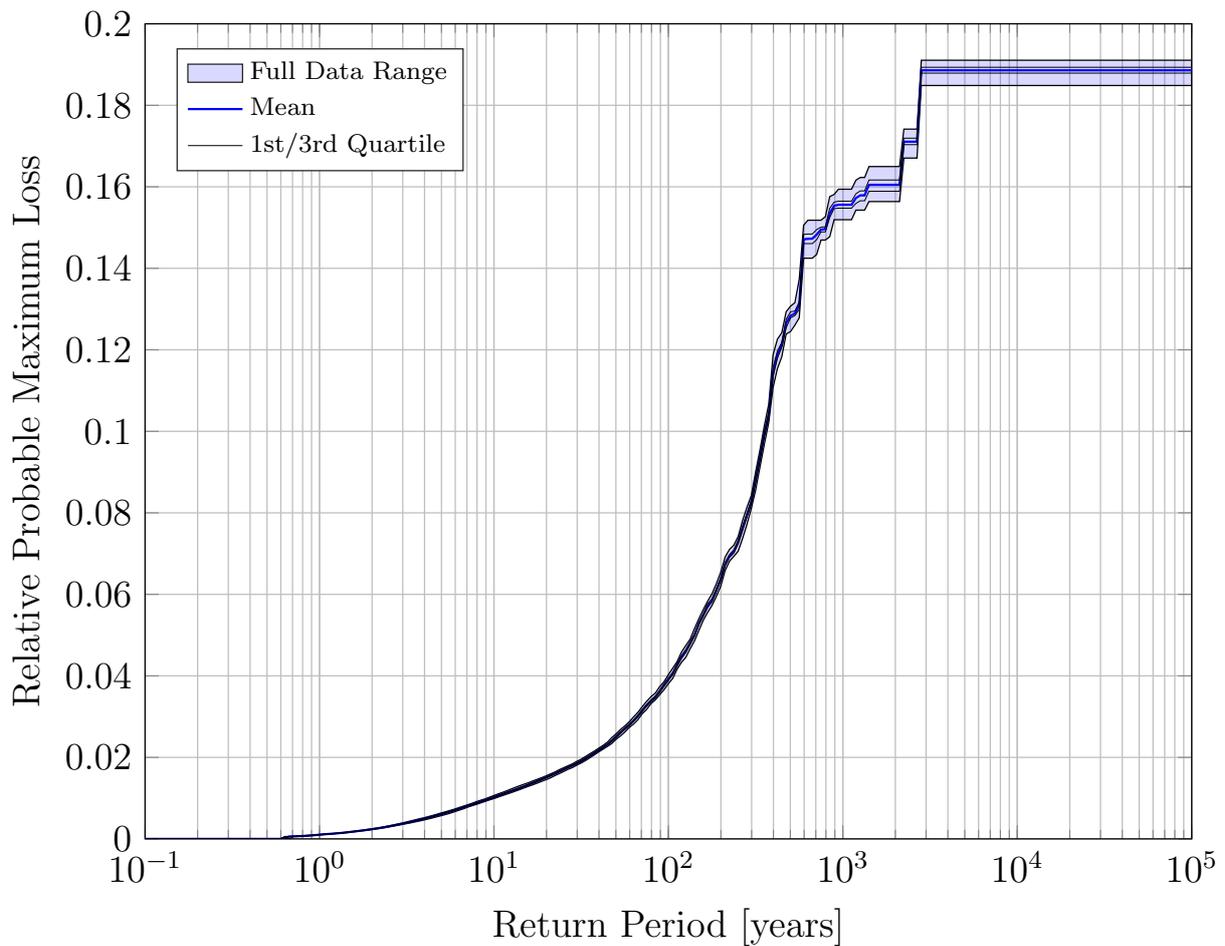


Figure 4.7: Probable Maximum Loss Curve (PML) and associated uncertainty for a synthetic portfolio of 300 risk items with 20% unknown coordinates, using a flat value distribution as described in Section 4.5.1 and the western Indonesia hazard model without aleatory ground motion uncertainty sampling. The risk items with unknown locations represent 20% of the portfolio value. Location uncertainty was treated with the framework described in Section 4.3 using a sample size of 128. The mean PML (blue line) as well as lower and upper quartile PML (black lines) are close. PML values are relative to the total sum insured of the portfolio.

Comparison to Reference Portfolio

In addition to the uncertainty distribution resulting from sampling location uncertainty, Figure 4.8 shows a reference PML curve in red. The curve represents the PML of the original portfolio with known coordinates from which the portfolio with unknown coordinates was derived. The mean of the portfolio with unknown coordinates deviates significantly from the reference solution. For some return periods, the reference curve is even outside the interquartile interval (middle 50%); this is because the spatial distribution of the orig-

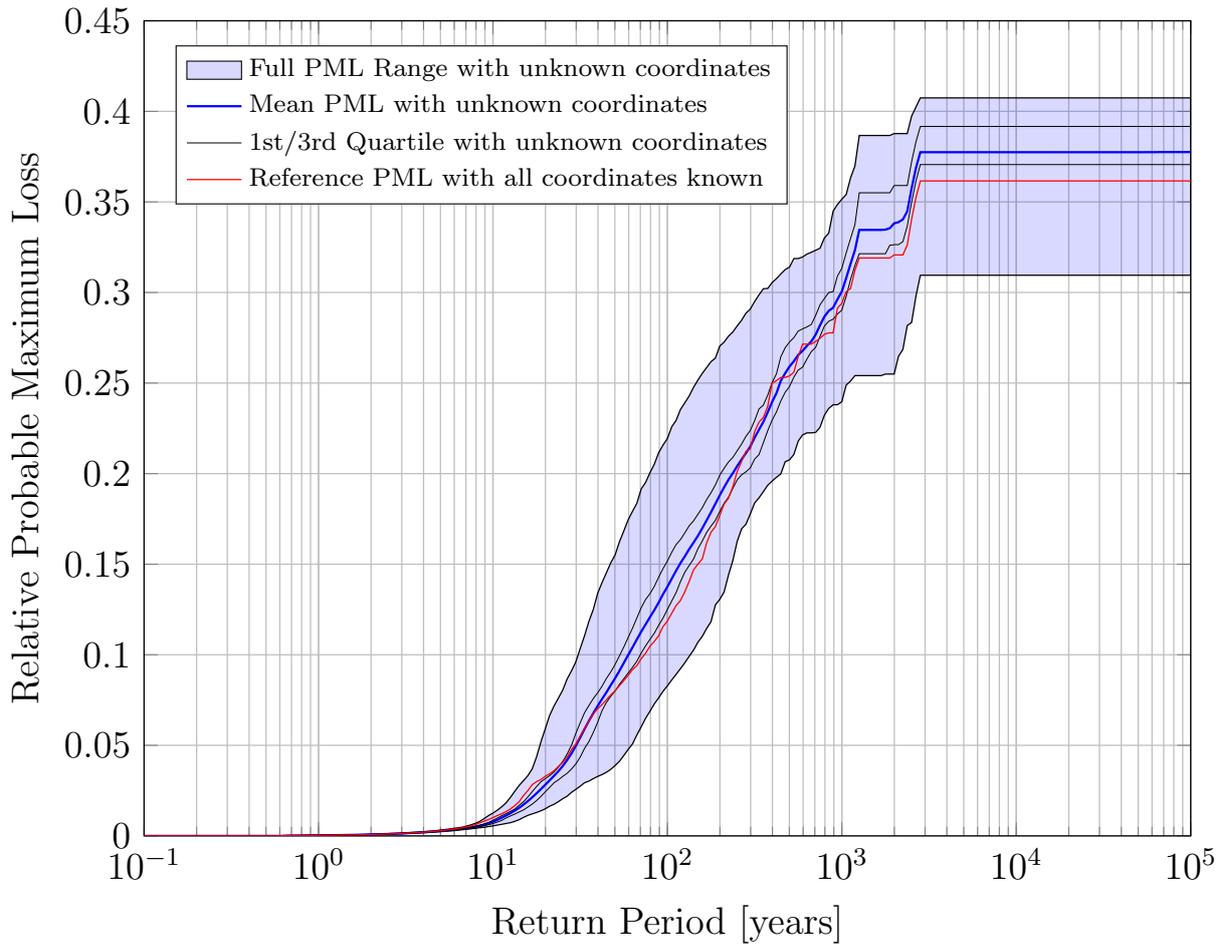


Figure 4.8: Probable Maximum Loss (PML) Curve and associated uncertainty for a synthetic portfolio of 300 risk items with 20% unknown coordinates, using a mixed exponential value distribution as described in Section 4.5.1 and the western Indonesia hazard model without aleatory ground motion uncertainty sampling. This shows the “worst-case” uncertainty for this portfolio, in which the risk items with unknown coordinates account for 76% of the total value. Location uncertainty was treated with the framework described in Section 4.3 using a sample size of 128. The lower and upper quartiles (black lines) and the full PML range (shaded in blue) spread significantly from the mean PML (blue line). The red line shows the PML of the reference portfolio for which all coordinates are known. PML values are relative to the total sum insured of the portfolio.

inal portfolio represents only one uncertainty realization of the portfolio with unknown coordinates. The PML of this realization can behave quite different from the mean PML. Figure 4.9 shows the ratio of the mean PML of the portfolio with unknown coordinates to the reference solution. The reference PML is underestimated for return periods where the blue curve is under the horizontal red line. For this portfolio, this is the case for re-

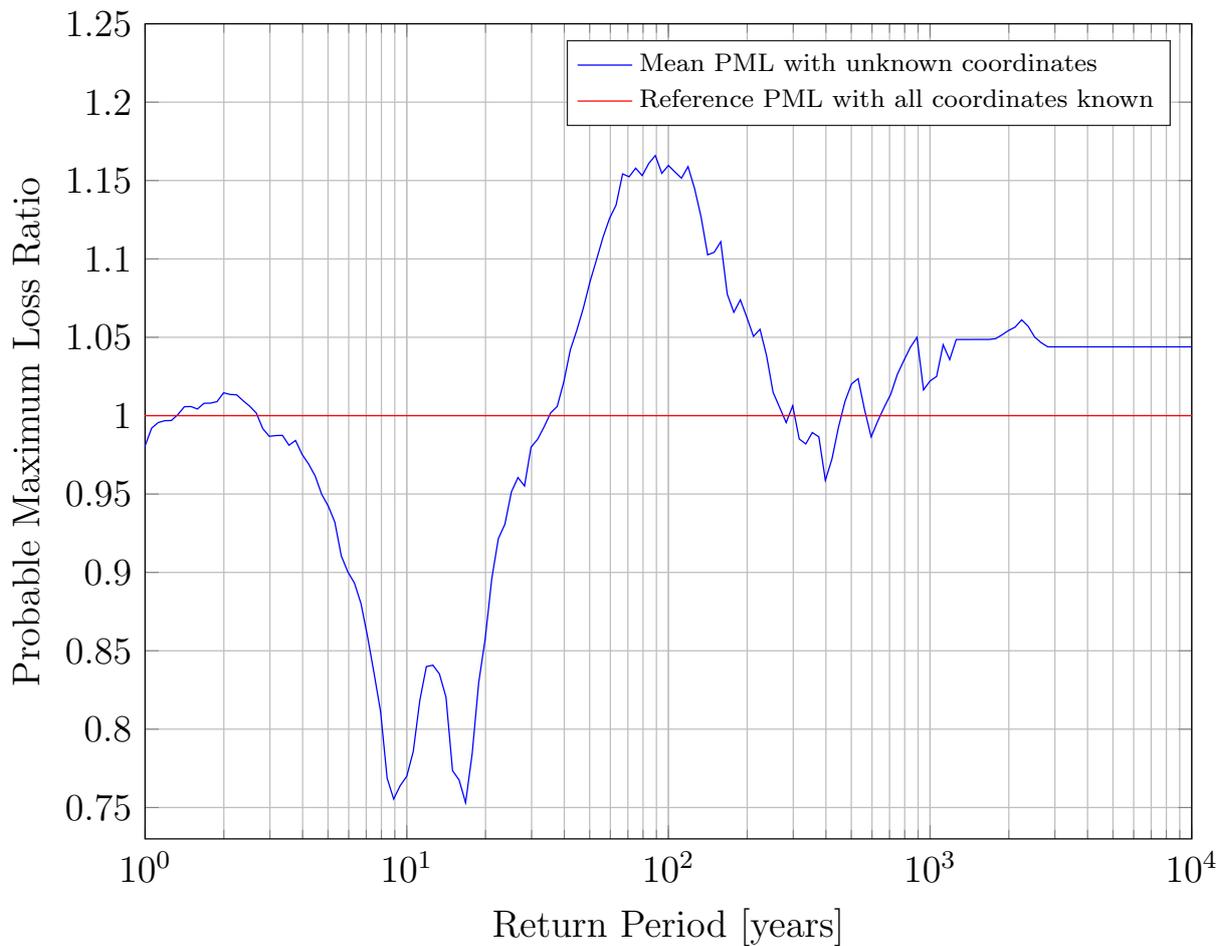


Figure 4.9: Comparison of the reference Probable Maximum Loss (PML) curve for the original portfolio with only known coordinates with the mean PML of the portfolio with 20% unknown coordinates, both using a mixed exponential value distribution as described in Section 4.5.1. Here, risk items with unknown coordinates account for 76% of the total portfolio value. Location uncertainty was treated with the framework described in Section 4.3 using a sample size of 128. Note that for return periods below 40 years, the mean PML of the portfolio with the unknown coordinates underestimates the probabilistic loss of the reference PML, while it tends to overestimate loss for higher return periods.

turn periods below 35 years. The deviation varies for different portfolios; for some, losses at shorter return periods are overestimated and losses at longer return periods underestimated; for others, losses at all return periods are systematically under- or overestimated; occasionally the mean PML is close to the reference solution.

Shape of the PML distribution For the isolated treatment of the location uncertainty performed in this section, we observe a wide range of shapes of the probability density of

PML at different return periods. Sometimes, the shape is symmetric, but often it is left- or right-skewed (e.g. right-skewed for this portfolio at 10 years, see Figure 4.10). In some cases, the distribution is even multi-modal.

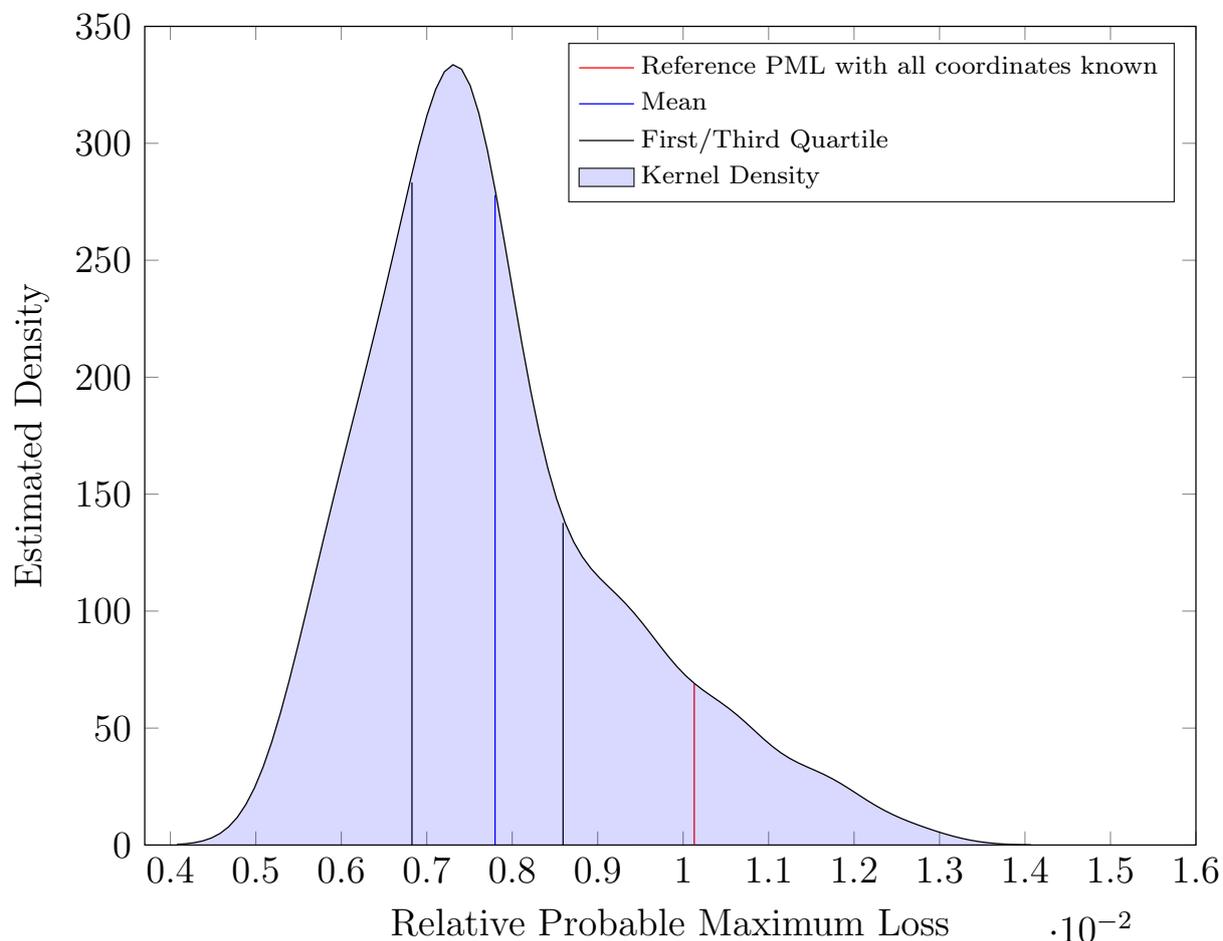


Figure 4.10: “Vertical cut” of the Probable Maximum Loss (PML) curve shown in Figure 4.8 at a return period of 10 years. This shows probability density obtained via kernel density estimation with a normal kernel function. The mean PML of the portfolio with unknown coordinates is shown as a vertical blue line, while lower and upper quartiles are shown in black. The estimated kernel density is shaded in blue. The reference PML of the portfolio with all locations known is shown as a vertical red line; for this return period, it is near the upper tail of the distribution.

Influence of Portfolio Size and Fraction of Unknown Coordinates

For a given value distribution, portfolio size and the fraction of unknown coordinates typically have an important effect. Figure 4.11 shows the CV of PML at 10 years return period against the fraction of unknown coordinates for portfolios of 10, 100, 1000 and

10000 risk items. Solid and dashed lines correspond to the flat and mixed exponential value distribution, respectively. Each curve depicts portfolios of the same size (number of risk items), and each point along the curves represents the CV of PML of one portfolio with the respective fraction of unknown coordinates, obtained from 128 sampled location sets. In particular for smaller portfolios, location uncertainty has a noticeable effect and the CV increases with the fraction of unknown coordinates. For larger portfolios, scenarios with higher losses seem to balance out scenarios with lower losses and therefore the CV increases slower and reaches lower maximum values. We observed a similar behavior at larger return periods than 10 years.

For portfolios with a mixed exponential value distribution, even small percentages of unknown coordinates have a strong effect on loss variability, because for this assessment of the “worst-case” impact, we preferentially remove coordinate information from the most valuable risk items. Since most of the portfolio TSI is assigned to a few risk items, the upward trend when removing coordinate-based location information of further risk items is small.

4.6.3 Western Indonesia Realistic Example: The Influence of Location Uncertainty in Conjunction with Aleatory Ground Motion Uncertainty

In the remainder of this chapter, we analyze the effect of location uncertainty in conjunction with aleatory ground motion uncertainty. To this end, we use the same hazard model and synthetic portfolios as in Sections 4.6.1 and 4.6.2, but also treat aleatory uncertainty for all GMPEs contained in the logic tree (Table 4.2). For GMPEs that provide separate terms for σ_E and σ_A (see Section 4.1), we sample the inter- and intra-event residuals independently. For GMPEs that only provide a term for σ_T , the total residual is directly sampled. For the analysis presented in this section, risk items are selected randomly for the removal of coordinate location information.

The ground motion distribution is known to have a strong influence on the results of PSHA and PSRA, particularly at high return periods corresponding to low annual frequencies of exceedance. While statistical tests and quantile-quantile plots support the assumption that spatially distributed logarithmic ground motion values follow a multivariate normal distribution, it is not clear over which range of quantiles this holds (Jayaram and Baker, 2008; Strasser et al., 2008). Most studies truncate the distribution at a fixed number ε_{\max} of standard deviations σ to exclude unphysical values, but no consensus has been reached on this value. Ideally, ε_{\max} should correspond to the ground motion that would result from the “worst-case” realization of the underlying physical processes, however it has proven to be difficult to establish a robust relationship to predict large residuals. The matter is further complicated by the fact that σ is related to the ground motion model and theoretically ε_{\max} should be too. We follow the recommendation to base this — necessarily subjective — decision on probabilistic considerations instead of defining an integer for ε_{\max} (Strasser et al., 2008). In view of the fact that in this section we investigate

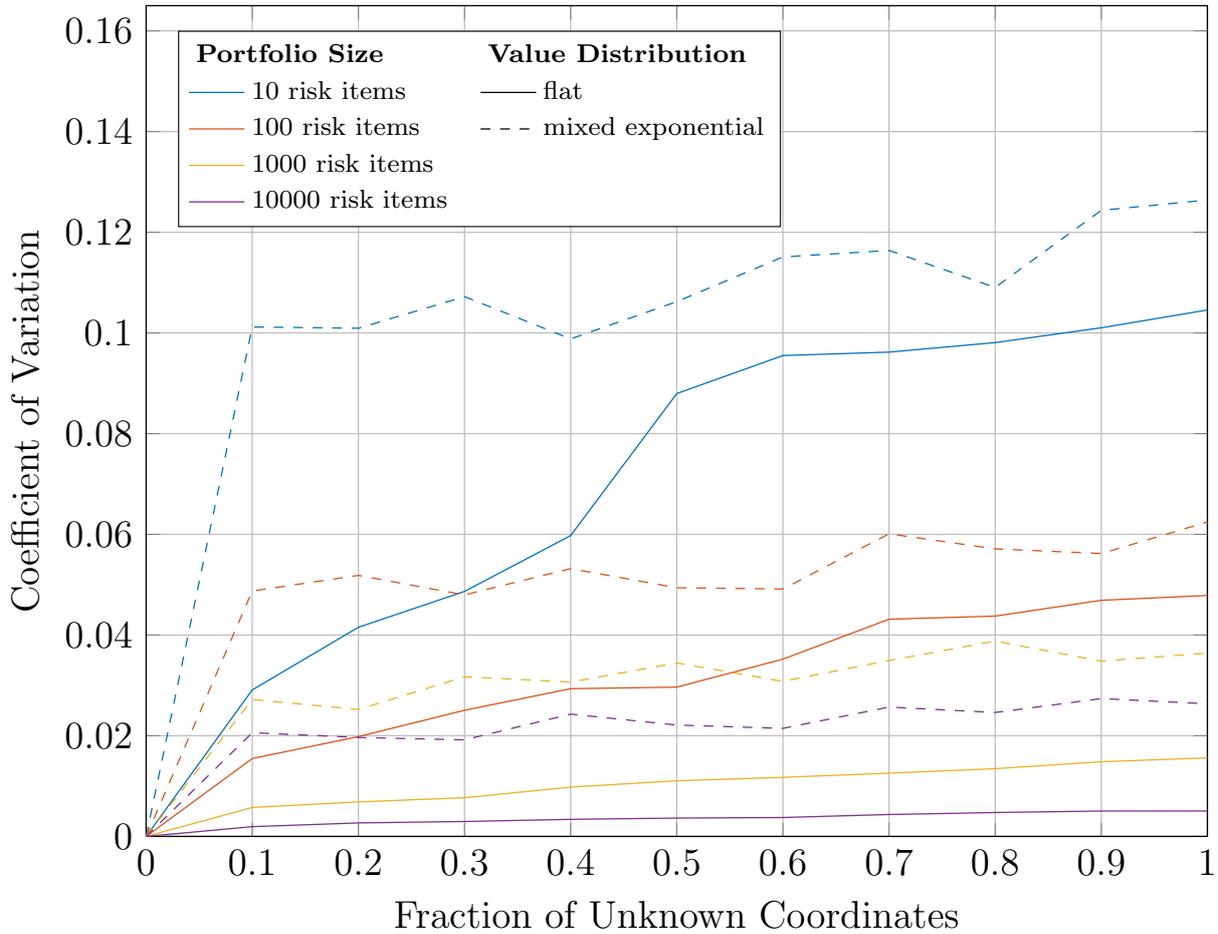


Figure 4.11: Coefficient of Variation of Probable Maximum Loss at a return period of 10 years for synthetic portfolios of 10, 100, 1000 and 10000 risk items with an increasing fraction of unknown coordinates. Solid lines correspond to the flat value distribution, dashed lines to the mixed exponential value distribution (see Section 4.5.1). This shows the “worst-case” uncertainty, i.e. coordinates of high value risk items are preferentially removed. Location uncertainty was treated with the framework described in Section 4.3 using a sample size of 128.

the variability of PML at a relatively small return period for which the truncation has a less pronounced influence, we choose a threshold of 10^{-3} on the conditional probability of exceedance. This corresponds to $\varepsilon_{\max} = 3.09$.

The results show that even for portfolios for which the coordinates of all risk items are known, considerable PML variability can already be caused by aleatory ground motion uncertainty. The degree of the increase in variability due to successively replacing coordinates of randomly selected risk items with administrative zone information depends on the value distribution and portfolio size. In line with the analysis presented in Section 4.6.2, in Figure 4.12 we plot the CV of PML at a return period of 10 years against the fraction

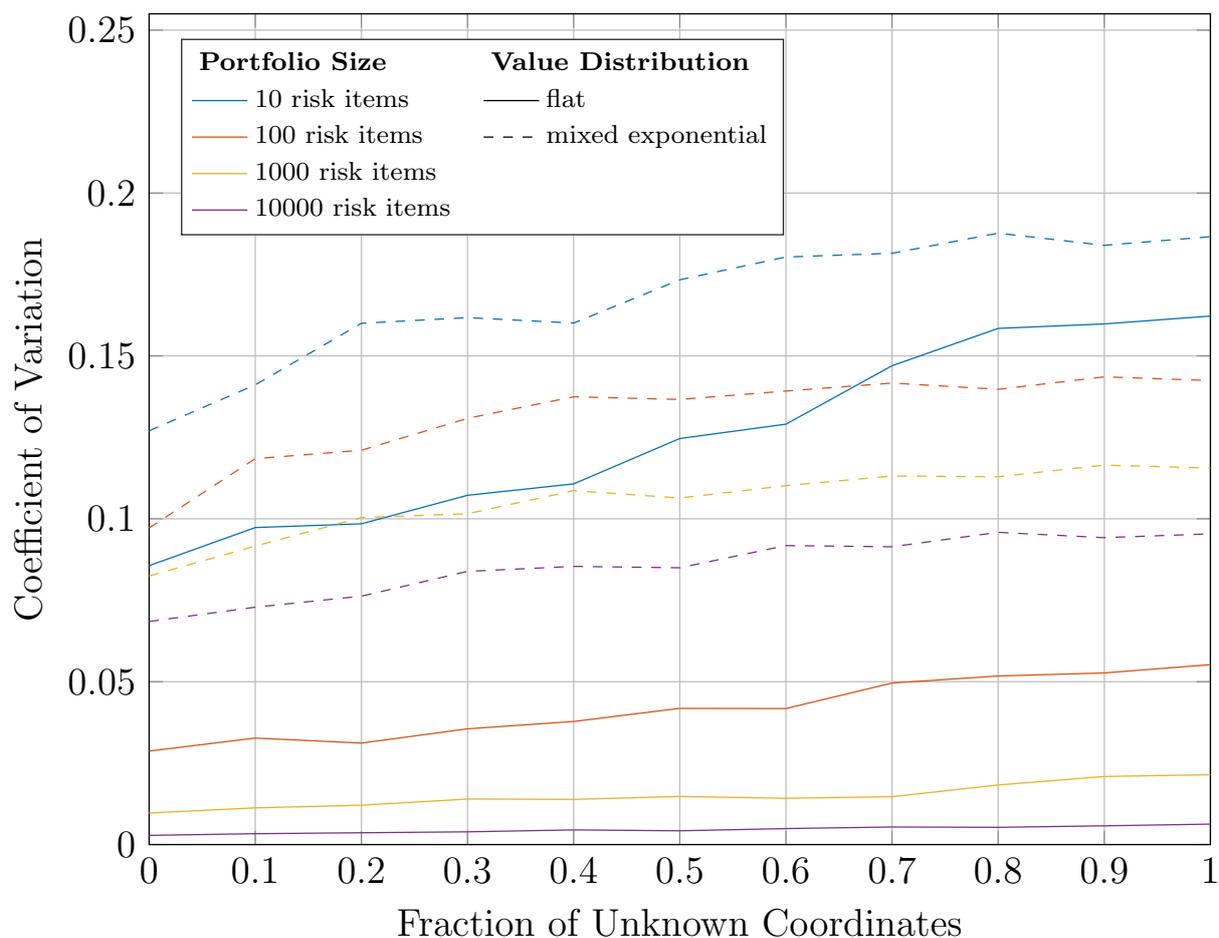


Figure 4.12: Coefficient of Variation of Probable Maximum Loss at a return period of 10 years for synthetic portfolios of 10, 100, 1000 and 10000 risk items with an increasing fraction of unknown coordinates. Solid lines correspond to the flat value distribution, dashed lines to the mixed exponential value distribution (see Section 4.5.1). Risk items are selected randomly for the removal of coordinate location information. In addition to location uncertainty, aleatory ground motion uncertainty was treated using a sample size of 128.

of unknown coordinates for portfolios with the flat (solid lines) and mixed exponential (dashed lines) value distributions. Individual curves correspond to portfolios of the same size, ranging from 10 to 10000 risk items; each point along the curves represents the result for one portfolio with the respective fraction of unknown coordinates. For 10 risk items with a flat value distribution and no unknown coordinates, the CV assumes a value of 8.6% and rises to 16.2% when all coordinates are unknown. While for larger portfolios the absolute increase in variability is smaller, the relative increase from 0% to 100% unknown coordinates remains roughly at a factor of 2.

Like in Section 4.6.2, PML variation is higher for the mixed exponential than for the

flat value distribution, which can once more be attributed to the concentration of value to a few risk items. For a portfolio of 10 risk items, the CV of PML at a return period of 10 years rises from 12.7% to 18.6%. For larger portfolios again the absolute increase from 0% to 100% unknown coordinates is smaller, but the relative factor remains roughly at 1.4.

4.7 Conclusions

Building on a recent study of the influence of exposure geo-resolution on the variability of scenario loss (Bal et al., 2010), in the present chapter we additionally incorporated stochastic sampling of exposure locations and explored the effect on PSRA results. To this end the Monte-Carlo methodology has been successfully applied to the modeling of seismic losses of synthetic but realistic portfolios in western Indonesia; key results have been presented in a novel graphical way that allows to effectively communicate loss estimations with their associated uncertainty.

Even though the results likely do not apply to other studies in terms of exact numbers, the analysis presented here provides valuable insight about the characteristics of location uncertainty. While the conclusions presumably hold for similar cases both in industry and academia, associated differences should be considered carefully, in particular for other regions and portfolios.

This study has shown that it is not advisable to neglect location uncertainty in PSRA. This holds especially for small to moderately sized portfolios of less than a few hundred risks, where loss aggregation — due to the risks' locations relative to each other within an earthquake footprint of damaging ground motion — is important. For portfolios consisting of several hundred risks with a relatively flat value distribution (e.g. residential portfolios), location uncertainty seems to play a less important role. However, even for large portfolios, location uncertainty can have a sizable effect if the value distribution is similar to the mixed exponential type (e.g. larger facultative portfolios). In these cases, it is necessary to investigate how many risks have unknown coordinates, which fraction of the portfolio TSI they represent, and how insured values are distributed. Low quality location information of a few important risk items can already have a significant effect.

The proposed approach to sample location uncertainty using insured exposure density undoubtedly provides more realistic results than previous approximate treatments. However, this study has also shown that it is important to be aware that the sampled mean PML curve does not necessarily converge to the result that would be obtained if the true spatial distribution were known. This is because the true spatial distribution only corresponds to one realization of location uncertainty, which is not related to the average loss. Another interesting observation is that location uncertainty can result in skewed as well as multi-modal loss distributions. This suggests that it is not ideal to treat location uncertainty as part of the ground motion or damage uncertainty sampling, since these uncertainties are usually modeled with a unimodal distribution.

It is appropriate to view the results in the context of overall uncertainties in PSRA. Even for a relatively rough geographical resolution like Indonesian provinces, the influence

of location uncertainty in isolation is not overwhelmingly large in comparison to the effect of ground motion and damage uncertainty (Bommer and Abrahamson, 2006). For most portfolios, the standard deviation of the loss distribution is below 10% of the mean, a finding in accordance with the aforementioned study on scenario loss (Bal et al., 2010). Additionally, it has turned out to be extremely challenging to validate hazard and loss models (Crowley et al., 2008). As a final word of caution, we remark that the range of PML uncertainty can be quite large and that the true spatial distribution of risk items might correspond to an extreme case at the tail of the distribution. Therefore, we propose the usage of appropriate uncertainty visualization techniques, as presented in this chapter, for risk underwriting and risk management in the insurance industry. Careful thought should be given to the graphical representation, which can have a strong effect on decision making (Stone et al., 2017). While this work represents a progressive step towards integrating stochastic treatment of exposure uncertainty into PSRA, opportunities for improvement remain, e.g. taking uncertainty in the insured value of risk items into account. The ultimate goal should be the proper treatment and communication of all involved uncertainties and their dependencies (Bier and Lin, 2013).

Acknowledgements

Financial support for this work was provided by Munich Re through a scholarship granted to the first author. We thank two anonymous reviewers, whose constructive feedback considerably improved the original article.

Chapter 5

Spatial Seismic Hazard Variation and Adaptive Sampling of Portfolio Location Uncertainty in Probabilistic Seismic Risk Analysis

Building upon the previous chapter, we now analyze spatial variation of seismic hazard in Indonesia. Based on this, we develop a novel sampling scheme specifically designed for portfolio location uncertainty in seismic risk analysis.

This chapter has been submitted for publication as a peer-reviewed article in:

Scheingraber, C. and Käser, M. (2018).

Spatial Seismic Hazard Variation and Adaptive Sampling of Portfolio Location Uncertainty in Probabilistic Seismic Risk Analysis.

Risk Analysis. Manuscript submitted for publication.

Abstract

Probabilistic Seismic Risk Analysis is widely used in the insurance industry to model the likelihood and severity of losses to insured portfolios by earthquake events. Due to geocoding issues of address information, risk items are often only known to be located within an administrative geographical zone, but precise coordinates remain unknown to the modeler.

In the first part of this chapter, we analyze spatial seismic hazard and loss rate variation inside administrative geographical zones in western Indonesia. We find that the variation of hazard can vary strongly not only between different zones, but also between different

return periods for a fixed zone. However, the spatial variation of loss rate displays a similar pattern as the variation of hazard, without depending on the return period.

We build upon these results in the second part of this chapter. In a recent work, we introduced a framework for stochastic treatment of portfolio location uncertainty. This results in the necessity to simulate ground motion on a high number of sampled geographical coordinates, which typically dominates the computational effort in Probabilistic Seismic Risk Analysis. We therefore propose a novel sampling scheme to improve the efficiency of stochastic portfolio location uncertainty treatment. Depending on risk item properties and measures of spatial loss rate variation, the scheme dynamically adapts the location sample size individually for insured risk items. We analyze the convergence and variance reduction of the scheme empirically. The results show that the scheme can improve the efficiency of the estimation of loss frequency curves.

5.1 Introduction

Probabilistic Seismic Risk Analysis (PSRA) is widely used in the insurance industry to model the likelihood and severity of losses to insured portfolios due to earthquake events. In this context, precise exposure locations are often unknown, which can have a significant impact on scenario loss as well as on loss frequency curves (Bal et al., 2010, Scheingraber and Käser, 2018, Chapter 4).

The treatment of uncertainty in PSRA of spatially distributed portfolios is usually performed by means of Monte Carlo (MC) simulation (Pagani et al., 2014). This is a computationally intensive process, because the error convergence of MC is relatively slow and a high-dimensional loss integral needs to be evaluated with a sufficient sample size. The hazard component typically dominates the overall model runtime of PSRA. As a result, stochastic treatment of portfolio location uncertainty can be particularly challenging — ground motion needs to be simulated on a large number of sampled risk locations. On the other hand, a fast model runtime is a key requirement for underwriting purposes in the insurance industry. Methods or sampling schemes to improve the error convergence of MC simulation are known as variance reduction techniques. MC simulation is ubiquitous in many areas of science and engineering and a wide variety of sampling schemes exists. Some well-known ideas are common random numbers and control variates (Yang and Nelson, 1991), importance-, stratified- and hypercube sampling, Quasi Monte Carlo Simulation using low-discrepancy sequences, as well as adaptive sampling. The error convergence of different sampling schemes has been investigated for many different types of integrals and application areas (Hess et al., 2006; dos Santos and Beck, 2015). Some work has already been performed on variance reduction for Probabilistic Seismic Hazard Assessment in the form of importance sampling, e.g. preferentially sampling the tails of the magnitude and site ground motion probability distributions (Jayaram and Baker, 2010; Eads et al., 2013). However, to our knowledge so far no study has specifically investigated variance reduction for location uncertainty in PSRA in a modern risk assessment framework. Building on a framework proposed in a recent study (Scheingraber and Käser, 2018, Chapter 4), in

the present chapter we describe a novel variance reduction scheme specifically designed to increase the computational efficiency of stochastic treatment of portfolio location uncertainty in PSRA.

The remainder of this chapter is structured as follows. We outline the most important theoretical background in Section 5.2. Using a seismic risk model of western Indonesia, in Section 5.3 we explore spatial hazard and loss rate variation inside administrative zones. Based on this, in Section 5.4 we propose an adaptive location uncertainty sampling scheme and investigate its performance using several test cases in Section 5.5. In Section 5.6, we give some recommendations on how to apply the results in practice and conclude with possible future improvements.

5.2 Background

5.2.1 Probabilistic Seismic Hazard and Risk Analysis

PSRA is based on Probabilistic Seismic Hazard Analysis (PSHA; Senior Seismic Hazard Committee (SSHAC), 1997; Cornell, 1968; McGuire, 2004), where the exceedance rate λ of ground motion level y_0 at a site \mathbf{r}_0 is expressed by the hazard integral

$$\lambda(y_0, \mathbf{r}_0)[y \geq y_0] = \int_V \int_{m_{\min}}^{m_{\max}} P[y \geq y_0 | m, \mathbf{r}, \mathbf{r}_0] \cdot \nu(m, \mathbf{r}) dm d\mathbf{r}, \quad (5.1)$$

with $\nu(m, \mathbf{r}) dm d\mathbf{r}$ the seismic rate density which describes the spatio-temporal distribution of seismic activity, $P[y \geq y_0 | m, \mathbf{r}, \mathbf{r}_0]$ the conditional probability of exceeding ground motion y_0 at site \mathbf{r}_0 given a rupture of magnitude m at source location \mathbf{r} , and V the spatial integration volume containing all sources which can cause relevant ground motion at \mathbf{r}_0 . Assuming that the occurrence of earthquake events is a temporal Poisson process, the probability of at least one exceedance of y_0 within time interval t_0 is given by

$$P(y_0, t_0, \bar{\lambda})[y \geq y_0] = 1 - e^{-\bar{\lambda}t_0}, \quad (5.2)$$

where $\bar{\lambda}$ is the mean annual recurrence rate.

For PSRA in the insurance industry, MC simulation is commonly used to obtain a set of stochastic ground motion fields $\hat{\mathbf{Y}}$ and to then compute the probability that a loss level ι_0 is exceeded as

$$P(\hat{\mathbf{Y}}, \Theta)[\iota \geq \iota_0] = \sum_{i=1}^{n_e} \int_{\iota_0}^{\infty} f_{\iota}(\iota | \hat{\mathbf{Y}}_i, \Theta) d\iota, \quad (5.3)$$

where $f_{\iota}(\iota | \hat{\mathbf{Y}}_i, \Theta)$ is the loss probability density function for a portfolio Θ given the i th ground motion field $\hat{\mathbf{Y}}_i$. Summing up the contribution of all n_e events yields the total loss exceedance probability. A Probable Maximum Loss (PML) curve, showing loss against mean return period T (with $T = 1/\bar{\lambda}$), can be obtained from the loss exceedance probability

curve (Equation 5.3) using a first order Taylor approximation of Equation 5.2:

$$T = \frac{t_0}{P(y_0, t_0, \lambda)[y \geq y_0]}. \quad (5.4)$$

Here, t_0 is the period of interest (time interval), which is 1 year for most reinsurance contracts.

5.2.2 Portfolio Location Uncertainty

Perhaps surprisingly, in the insurance industry, portfolios frequently lack precise coordinate-based location information. Obtaining this information is often not possible, e.g. because

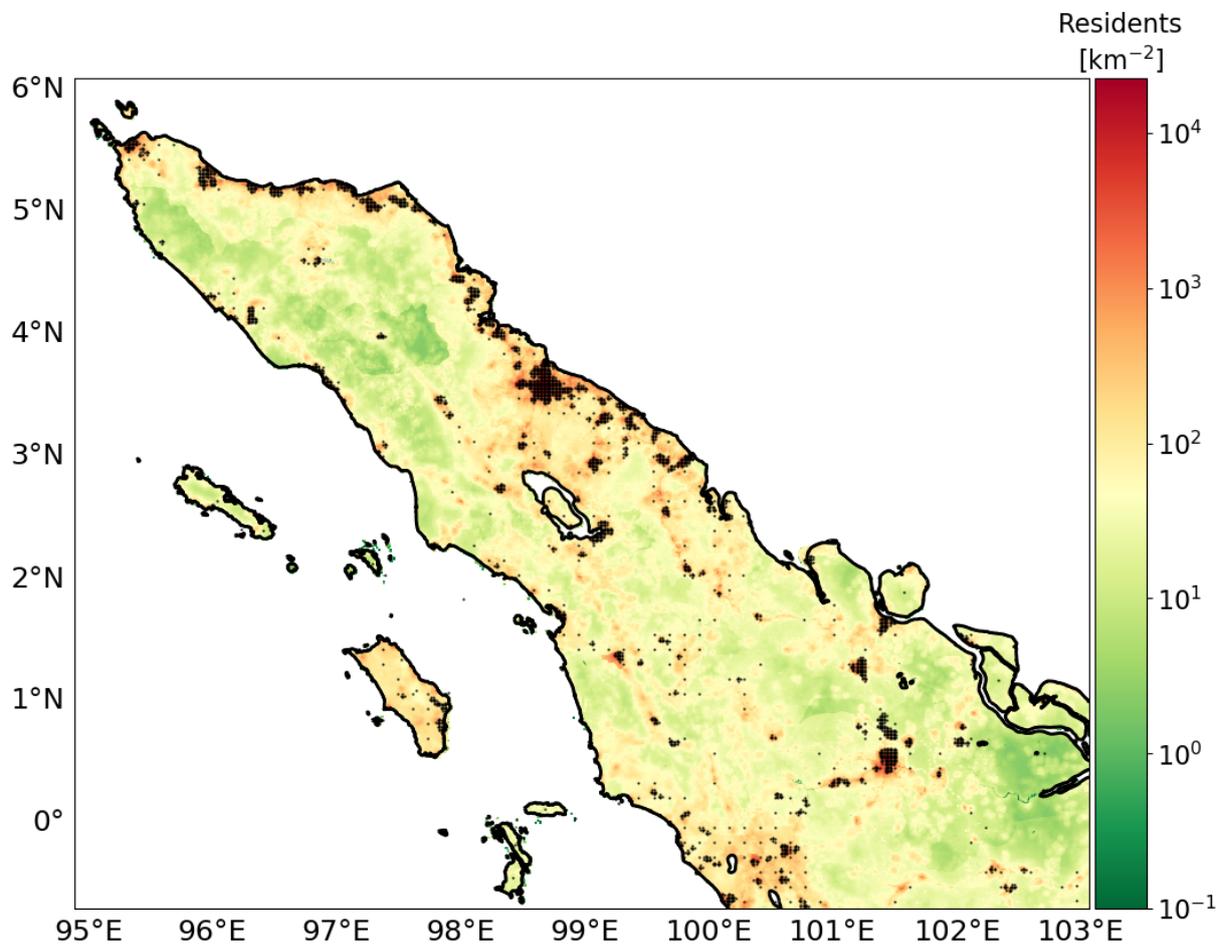


Figure 5.1: An example of a weighted grid used as an insurance density proxy for the location uncertainty framework. This shows northern Sumatra. Color indicates population density (residents per km²) as a proxy for insured exposure density. Black markers depict grid points of the weighted grid. The population data in this plot is based on a free dataset (Gaughan et al., 2015).

geocoding engines are not used systematically or can not reliably obtain coordinates from the policy address of the insured risk. Especially for large treaty portfolios with thousands or millions of risks, it apparently is simply too much effort for the primary insurer or the insurance broker to obtain and provide this information. Unfortunately, this is also not uncommon for smaller portfolios consisting only of a few hundred high-value risks. However, administrative zones, such as postal codes, can easily be obtained from the insurance policy.

Exposure uncertainty has previously been identified as an important area of research (Crowley, 2014), and we already introduced a framework for stochastic treatment of location uncertainty in a recent paper (Scheingraber and Käser, 2018, Chapter 4). In our framework, locations of risk items without precise coordinate location information are sampled with replacement from a weighted irregular grid inside their corresponding administrative zone. The grid weights are used to preferentially sample locations in areas of assumed high insurance density, e.g. based on population density or on commercial and industrial inventory data depending on the type of risk (Dobson et al., 2000). An example of such a weighted grid is shown in Figure 5.1.

In MC simulation, the choice of a pseudo-random number generator is of particular importance. In this study we use *MRG32K3a*, a combined multiple recursive generator which efficiently generates random number sequences with low memory requirements and excellent statistical properties (L'Ecuyer, 1999). *MRG32K3a* supports up to $1.8 \cdot 10^{19}$ statistically independent *streams*. Each stream has a *period*¹ of $7.6 \cdot 10^{22}$. These properties make *MRG32K3a* well suited for a large scale parallel MC simulation of seismic risk.

5.2.3 Evaluation of the Proposed Sampling Scheme

Standard Error

Because MC simulation is a stochastic method, there are no strict error bounds for statistics of interest obtained from a sample of finite size n . The error is therefore usually estimated using the standard deviation of the sampling distribution of the respective statistic, which is referred to as its standard error (E_{SE}). If the sampling distribution is known (e.g. normal), standard errors can often be obtained using a simple closed-form expression (Harding et al., 2014). For the statistics estimated in this study, e.g. PML at a specific return period, we can however not make a valid distribution assumption when taking location uncertainty into account. We therefore use repeated simulation to evaluate the performance of the proposed sampling scheme. The standard error can then be estimated as

$$E_{SE}(\hat{\Phi}_R) = \sqrt{\text{Var}(\hat{\Phi}_R)}, \quad (5.5)$$

¹The period of a pseudo-random number generator refers to the minimum length of a generated sequence before the same random numbers are repeated cyclically.

where $\hat{\Phi}_R$ denotes a set of estimations of a statistic obtained from R repeated simulations and $\text{Var}(\cdot)$ the variance operator. The corresponding relative standard error E_{RSE} can be obtained by dividing E_{SE} by the estimated statistic. To estimate confidence intervals of standard errors, we use bootstrapping with the bias-corrected accelerated percentile method (Efron, 1979; Efron and Tibshirani, 1986).

Bias and Convergence Plots

The bias of an estimator $\hat{\theta}$ is defined as

$$\text{Bias}(\hat{\theta}_n) = \mathbb{E}_\theta(\hat{\theta}_n) - \theta, \quad (5.6)$$

where $\hat{\theta}_n = f(x_1, x_2, \dots, x_n)$ is the estimator depending on the n members of the sample and \mathbb{E}_θ its expected value. Because deriving the bias analytically is infeasible for a complex numerical simulation such as performed by our framework, we use simple MC² with a large sample size as empirical reference and approximation for θ . In addition we use convergence plots, which are a simple yet powerful method to monitor and verify the results (Robert and Casella, 2004). The values estimated using simple MC and the adaptive variance reduction scheme are plotted against increasing sample size n .

Variance Reduction, Convergence Order and Speedup

To quantify the performance of the proposed scheme at a particular sample size n , we use the following well-known definition of variance reduction VR:

$$\text{VR} = \frac{\sigma_{\text{MC}}^2}{\sigma_{\text{LSS}}^2}, \quad (5.7)$$

where σ_{MC}^2 is the variance using simple MC and σ_{LSS}^2 the variance using the proposed location sampling scheme (MacKay, 2005; Juneja and Kalra, 2009).

To describe asymptotic error behavior for growing n , we use the big O notation (\mathcal{O} ; Landau, 1909; Knuth, 1976). For example, the error convergence order of simple MC is always $\mathcal{O}(n^{-0.5})$ independent of the dimensionality of the integrand (Papageorgiou, 2003).

To compare the real runtime required by simple MC and the proposed scheme to reach a specific relative standard error level ε_{RSE} , we use the speedup S defined as

$$S = \frac{t_{\text{MC}}}{t_{\text{LSS}}}, \quad (5.8)$$

where t_{MC} the runtime required by simple MC and t_{LSS} the runtime required by the proposed location sampling scheme.

²For simple MC, the strong law of large numbers guarantees an *almost certain* convergence for $n \rightarrow \infty$.

5.2.4 Generation of Synthetic Portfolios

In this work, we use synthetic portfolios in western Indonesia modeled after real-world counterparts in terms of spatial distribution of risk items as well as value distribution among risk items.

Value Distribution

The total sum insured (TSI) is kept constant for all portfolios:

$$\text{TSI} = \text{const.} = 1 \cdot 10^6. \quad (5.9)$$

However, the TSI is distributed among a varying number of risk items (portfolio size). For this study, we use portfolio sizes n_r of 1, 10, 20, 50, 100, 1000 and 10000 risk items.

The value distribution observed in many real residential portfolios can be approximated well by a randomly perturbed flat value distribution:

$$\text{VI}_{\text{flat},i}^* = \frac{\text{TSI}}{n_r} \cdot X_i, \quad (5.10)$$

$$\text{VI}_{\text{flat},i} = \frac{\text{TSI}}{\sum_{i=1}^{n_r} \text{VI}_{\text{flat},i}^*} \cdot \text{VI}_{\text{flat},i}^*. \quad (5.11)$$

$\text{VI}_{\text{flat},i}$ (“value insured”) is the value assigned to the i th risk item and n_r denotes the number of risk items. X_i is a uniform random number in the interval $[1 - p, 1 + p]$, where p is a perturbation factor set to 0.2, which is consistent with the characteristics of many real portfolios. Equation 5.11 normalizes the n_r randomly perturbed insured values to ensure $\sum_{i=1}^{n_r} \text{VI}_{\text{flat},i} = \text{TSI}$.

Geographical Distribution

For each portfolio size, we created a set of 6 portfolios with an increasing fraction of unknown coordinates: 0%, 20%, 40%, 60%, 80%, and 100% of the risk items have unknown coordinates and are only known on the basis of their administrative zone (Indonesian provinces, or regencies and cities, see Section 5.3).

The geographical distribution of the exposure locations follows the weighted irregular grid described in Section 5.2.2. For each portfolio size, a portfolio with 0% unknown coordinates is initially created by choosing exposure locations from the irregular grid according to the grid point weights. For the other portfolios with the same number of risk items but a higher fraction of unknown coordinates, coordinate-based location information is then removed stepwise from the initial portfolio. In each step, 20% of the risk items are randomly selected for the removal of coordinates until all risk items have unknown coordinates.

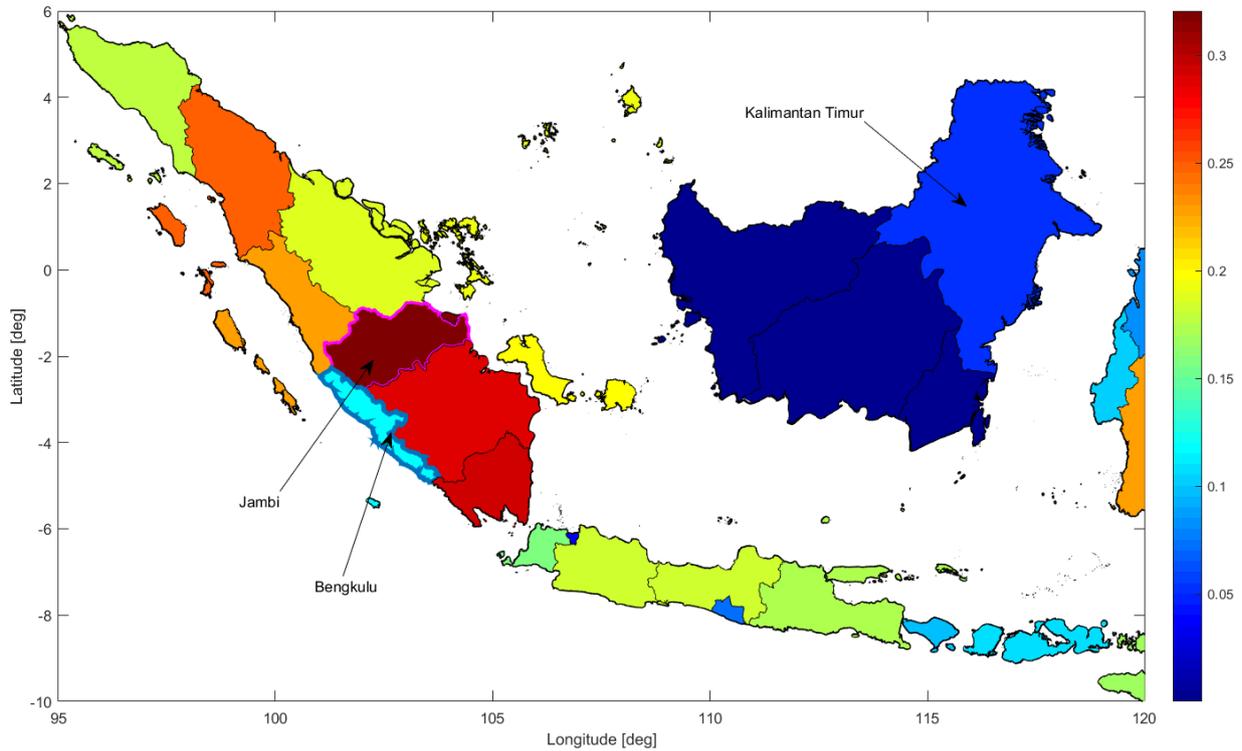


Figure 5.2: Coefficient of Variation (CV) of Peak Ground Acceleration (PGA) with an exceedance probability of 10% in 50 years inside provinces in western Indonesia. Color denotes the CV. Note how the CV is higher in provinces that have a large extent perpendicular to the Sumatra Fault Zone, such as *Jambi* (outlined in pink color), than in provinces with a small extent in that direction, such as *Bengkulu* (outlined in blue).

5.3 Case Study: Spatial Seismic Hazard Variation in Western Indonesia

5.3.1 Hazard Model

We use a proprietary seismic risk model based on the South-East Asia hazard model of the United States Geological Service (USGS) by Mark Petersen et al. (2007). Site conditions are based on topographic slope (Wald and Allen, 2007). The geometry of the Sumatra subduction zone is a complex fault representation based on the three-dimensional *Slab 1.0* model (Hayes et al., 2012). For events on the complex fault, we use a rupture floating mechanism similar to the implementation of OpenQuake (Pagani et al., 2014), a free and open-source seismic hazard and risk software developed as part of the Global Earthquake Model initiative (Crowley et al., 2013). The model is described in greater detail in a recent paper (Scheingraber and Käser, 2018, Chapter 4).

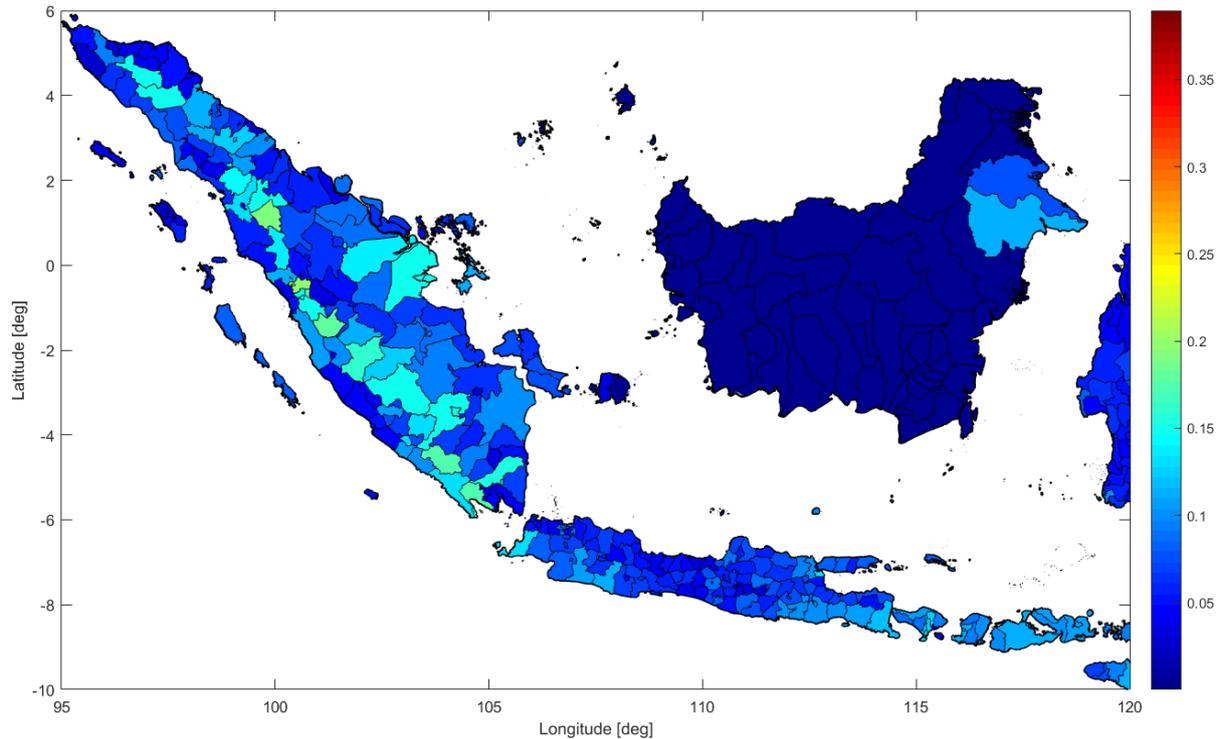


Figure 5.3: Coefficient of Variation (CV) of Peak Ground Acceleration (PGA) with an exceedance probability of 10% in 50 years inside regencies and cities in western Indonesia. Color denotes the CV. At this geographical resolution the CV is lower than for provinces (see Figure 5.2), and the influence of individual seismo-tectonic features, such as the Sumatra fault zone, becomes apparent.

5.3.2 Spatial Seismic Hazard Variation

For this analysis, we compute seismic hazard on a regular grid using a resolution of 0.3° . We investigate the coefficient of variation (CV) of hazard inside administrative geographical zones for different levels of resolution, corresponding to provinces and regencies or cities in Indonesia. The CV is defined as

$$CV = \frac{\sigma}{\mu}, \quad (5.12)$$

where σ is the standard deviation and μ the mean.

Dependence on Resolution Level of Geographical Zones

Figure 5.2 shows the CV of peak ground acceleration with an exceedance probability of 10% in 50 years per province in Indonesia. There is a noticeable decrease of the CV from west to east. The subduction modeled by the complex fault and the Sumatra Fault Zone (SFZ) result in the highest CV on Sumatra (most values 0.2 – 0.3). The CV is also relatively high on Java (around 0.15). The CV is the lowest in Kalimantan (< 0.1) due to the

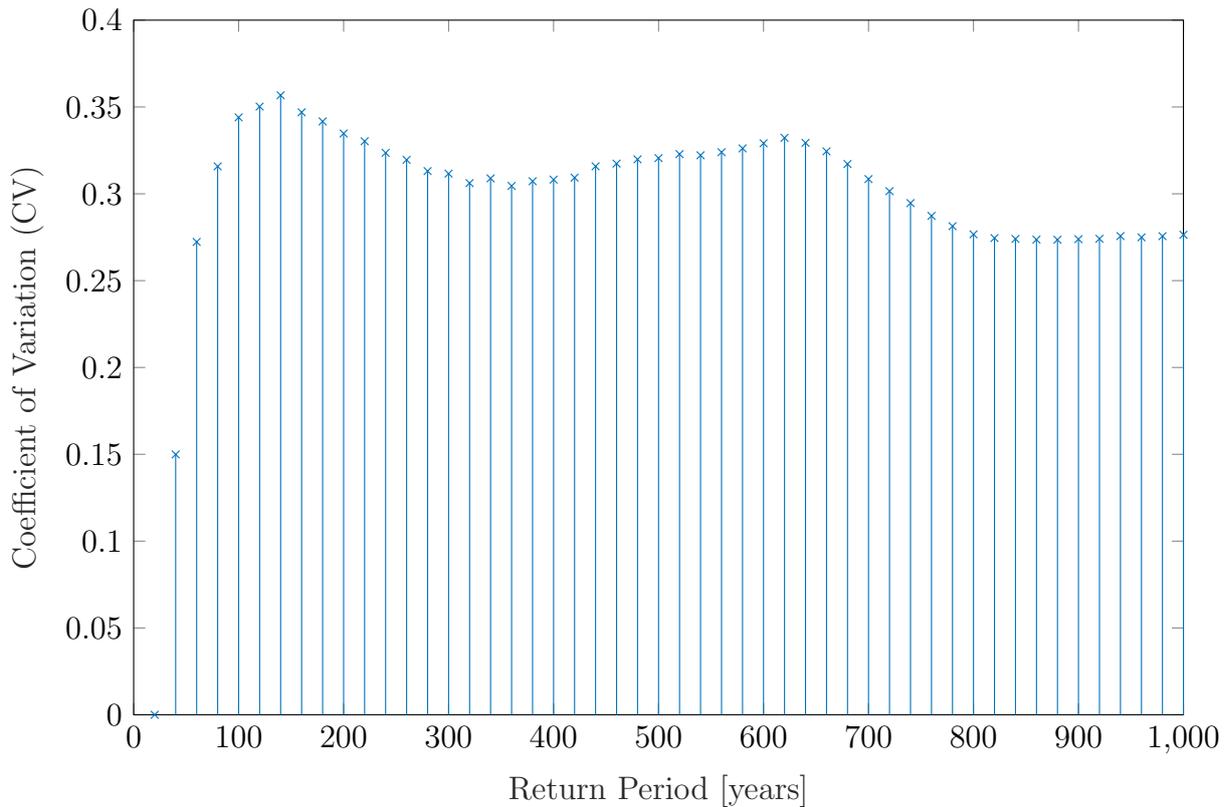


Figure 5.4: Coefficient of Variation (CV) of ground motion predicted to be exceeded at various return periods for the *Jambi* province (see Figure 5.2). The CV remains quite stable over a large range of return periods.

absence of any known or modeled crustal faults. As only gridded seismicity is used in this area, the hazard variation is very small. Furthermore, zones with a large extent perpendicular to the SFZ show a larger CV than zones with a smaller extent along the direction of the steepest hazard gradient. An example of this are the provinces of *Jambi* and *Bengkulu* in Figure 5.2. Arguably, location uncertainty is more important in *Jambi* than in *Bengkulu*.

Figure 5.3 shows the CV per regency or city for the same exceedance probability. Due to the smaller spatial extent of the administrative zones, the CV is in general lower at this more granular resolution of administrative geographical zones. Another observation is that the influence of individual seismo-tectonic features emerges; the CV is higher in the vicinity of modeled faults. While the Sumatra subduction only has a weak influence, the SFZ has a pronounced effect. Near the SFZ, the CV has values of about 0.1 – 0.2. Perpendicular to the SFZ, the CV quickly drops below 0.1.

In general, the CV is highest in zones close to modeled faults of shallow depths, as they result in a higher spatial hazard gradient than compared to areas where hazard is

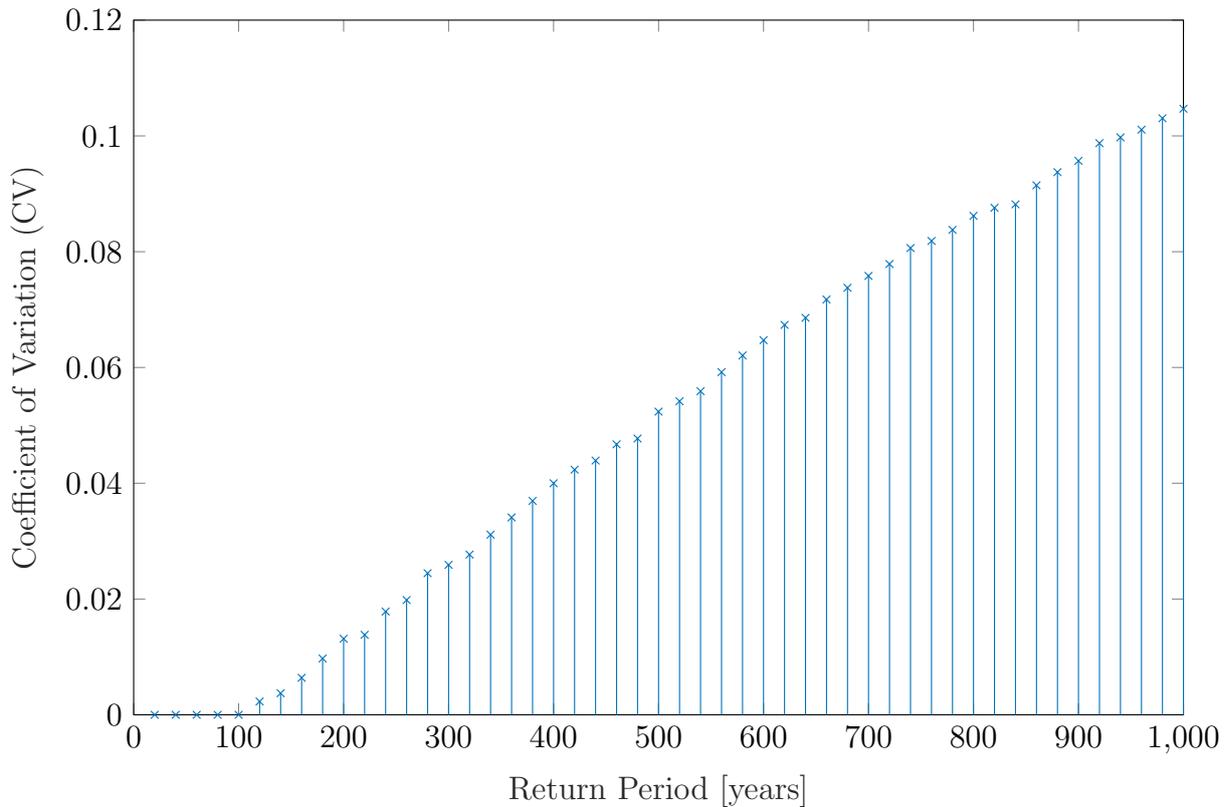


Figure 5.5: Coefficient of Variation (CV) of ground motion predicted to be exceeded at various return periods for the *Kalimantan Timur* province (see Figure 5.2). In this case, it is not possible to determine a range of return periods for which the CV remains in a stable range.

dominated by rather regularly distributed gridded seismicity. A reasonable assumption is that location uncertainty can be particularly high in such zones.

Dependence on Return Period

Analysis of the CV across different return periods for individual zones revealed a similar pattern for most administrative zones. The CV is small for short return periods, and reaches a relatively stable level above a certain return period. An example of this behavior is shown in Figure 5.4 for the province of *Jambi*. However, the CV does not show this pattern in all administrative zones. For some zones, especially at the level of regencies and cities, we could not determine a range of return periods for which the CV is roughly constant, as for example in the province of *Kalimantan Timur* shown in Figure 5.5.

5.3.3 Loss Rate Variation

The variability of the CV over return periods for certain zones makes it difficult to choose a general return period suitable for assessing the spatial variation of hazard inside a zone. To avoid the subjectivity introduced by a manual decision process for a suitable return period, we use the CV of the loss rate per zone, as it considers all return periods. Figure 5.6 shows the CV of the loss rate for Indonesian provinces. The overall pattern agrees with the pattern of the spatial hazard variation in Figure 5.2, but the range of values is much higher, from about 0.1 to 0.9.

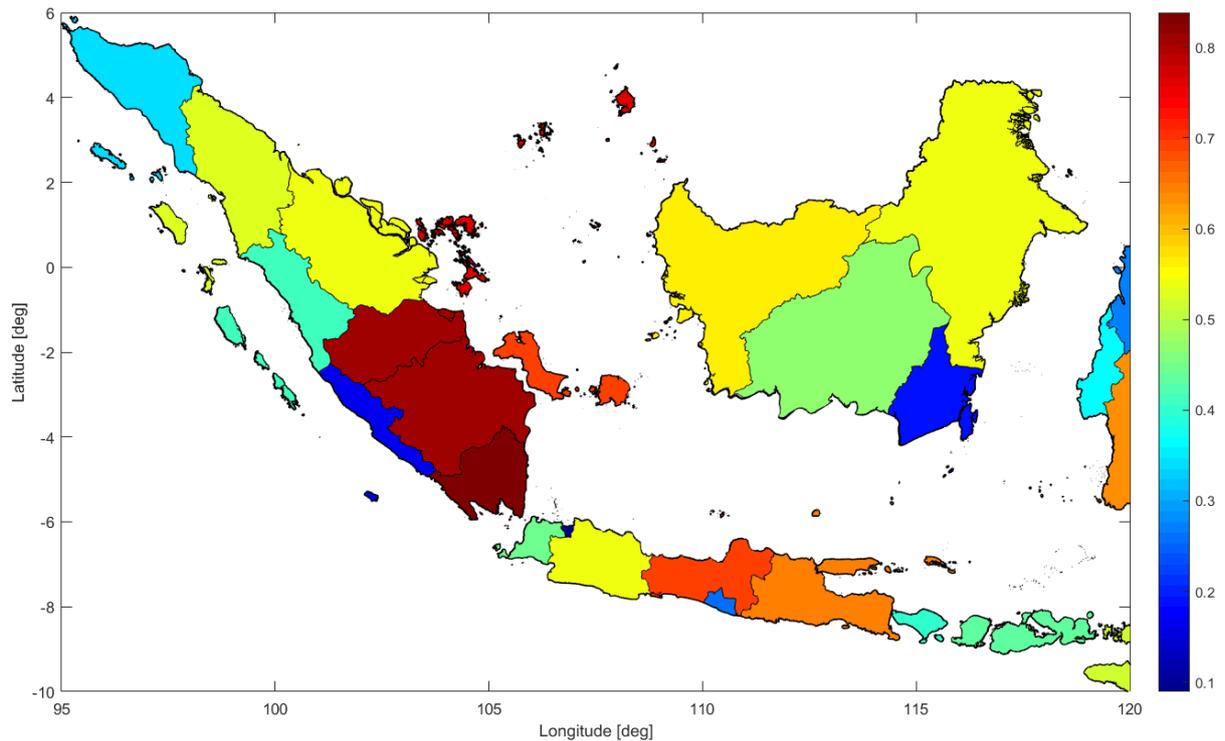


Figure 5.6: Coefficient of Variation (CV) of the loss rate inside provinces in western Indonesia. Color denotes the CV.

5.4 A Framework for Adaptive Sampling of Portfolio Location Uncertainty

To increase efficiency, in our framework ground motion is jointly simulated on all unique locations of all sampled location sets. Since the computation of hazard dominates the overall runtime of PSRA, it is worthwhile to explore possibilities to distribute the number of locations on which hazard is computed in a smart way among risk items. To this end, we introduce three sampling criteria to determine the location sample size individually per

Table 5.1: Risk Location Index Mapping Table. Rows correspond to individual risk items, showing sampled grid point indices. Each column represents a possible spatial distribution of the portfolio. Risk item 1 has the maximum location sample size of $n_{\max} = 4$, but risk items 2 and 3 only have a sample size of 2 and 1, respectively.

Risk Item Index	Sample Size	Sample 1	Sample 2	Sample 3	Sample 4
1	<i>4</i>	43	13	31	51
2	<i>2</i>	23	18	23	18
3	<i>1</i>	51	51	51	51

risk item. A large location sample size is used for risk items for which at all three criteria indicate that location uncertainty has a strong influence. If any of the three criteria predicts that location uncertainty has a lesser effect, a smaller sample size is used. In this way, more computational effort is invested where it is important and a better estimation of the PML curve associated with a lower variance is obtained for a given number of used hazard locations.

To not add noticeable overhead to the calculation, a key requirement is that all criteria can be evaluated very efficiently. To keep the computational overhead small, another design goal is that the framework is adaptive in a sense that it depends directly on properties of the portfolio and a precalculated hazard variability (see Section 5.3), but does not require on-the-fly integral presampling such as used by some general purpose adaptive variance reduction schemes (Press and Farrar, 1990; Jadach, 2003).

5.4.1 Risk Location Index Mapping Table

We store an array containing all unique geographical locations on which ground motion is simulated, and another array storing the sampled location indices per risk item. Table 5.1 illustrates the concept. Each column of the table corresponds to a location set representing a valid realization of location uncertainty for the entire portfolio. To combine unequal sample sizes for risk items without introducing bias due to overemphasis of a subset of a sample, we restrict the sample size to powers of two. The full sample can then be repeated in the mapping table.

5.4.2 Criterion I: Coefficient of Variation of Loss Rate

The first criterion is based on the CV of loss rate within a zone (see Section 5.3), hereafter denoted by CV_z . The values of CV_z can be precomputed for all administrative geographical zones, and therefore the evaluation of this criterion can be implemented in a very efficient manner.

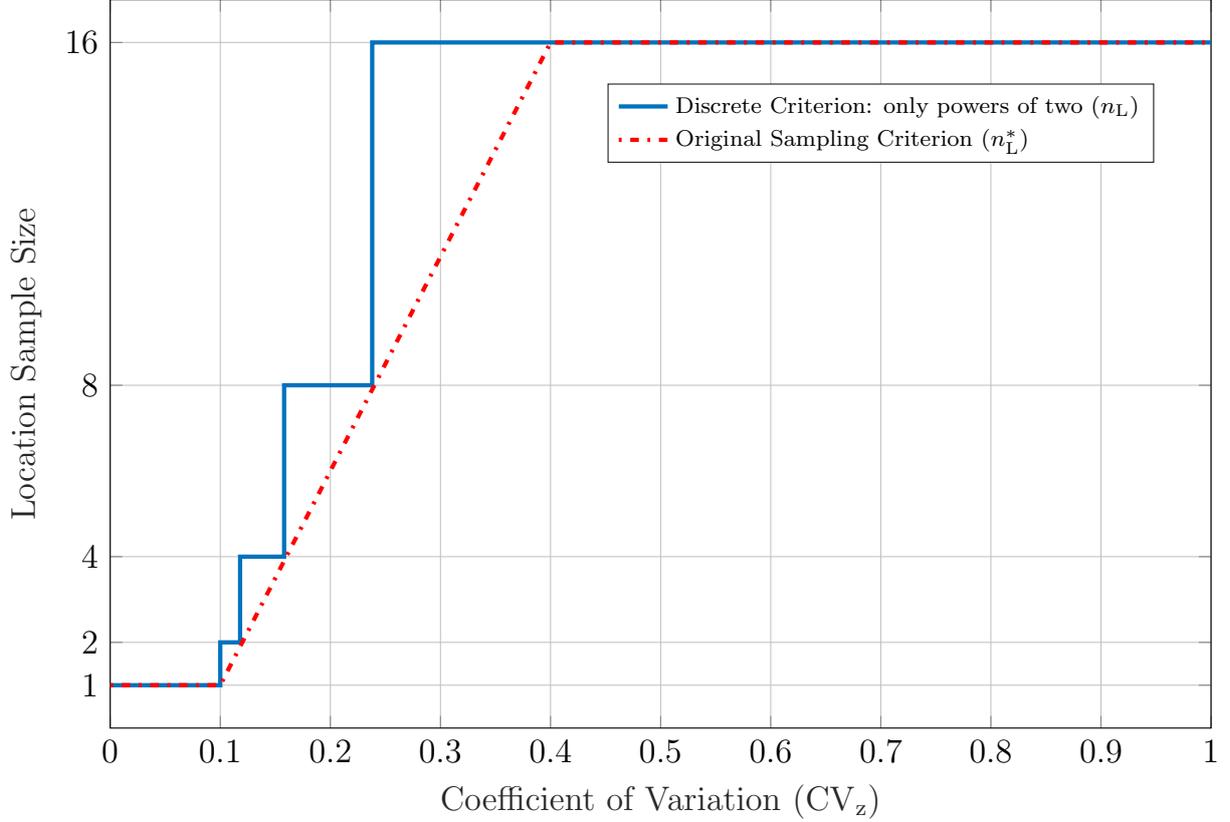


Figure 5.7: Criterion I: Number of samples per zone depending on Coefficient of Variation (CV). The discrete realization of the criterion (limited to powers of two) is shown in blue, while the red line represents the theoretical linear behavior.

The number of samples n_L due to criterion I is defined piecewise:

$$n_L^* = \begin{cases} 1, & \text{if } CV_z \leq t_l, \\ \frac{n_{\max}-1}{t_u-t_l} \cdot CV_z + 1, & \text{if } CV_z \in (t_l, t_u), \\ n_{\max}, & \text{if } CV_z \geq t_u. \end{cases} \quad (5.13)$$

Here, t_l and t_u are lower and upper threshold values. n_{\max} represents the maximum used sample size. We round n_L^* up to the next higher power of two to obtain the final n_L . The criterion is shown in Figure 5.7 for the example $t_l = 0.1$, $t_u = 0.4$ and $n_{\max} = 16$. In our final implementation, t_l and t_u are chosen adaptively as empirical quantiles of the CV distribution ($CV_{0.4}$ for t_l and $CV_{0.6}$ for t_u , i.e. the 40% and 60% percentiles) of the loss rate of all administrative zones of a model (see Section 5.3), which was found to be a reasonable choice for our test cases with the aid of an extensive parameter study (see Section 5.5.1).

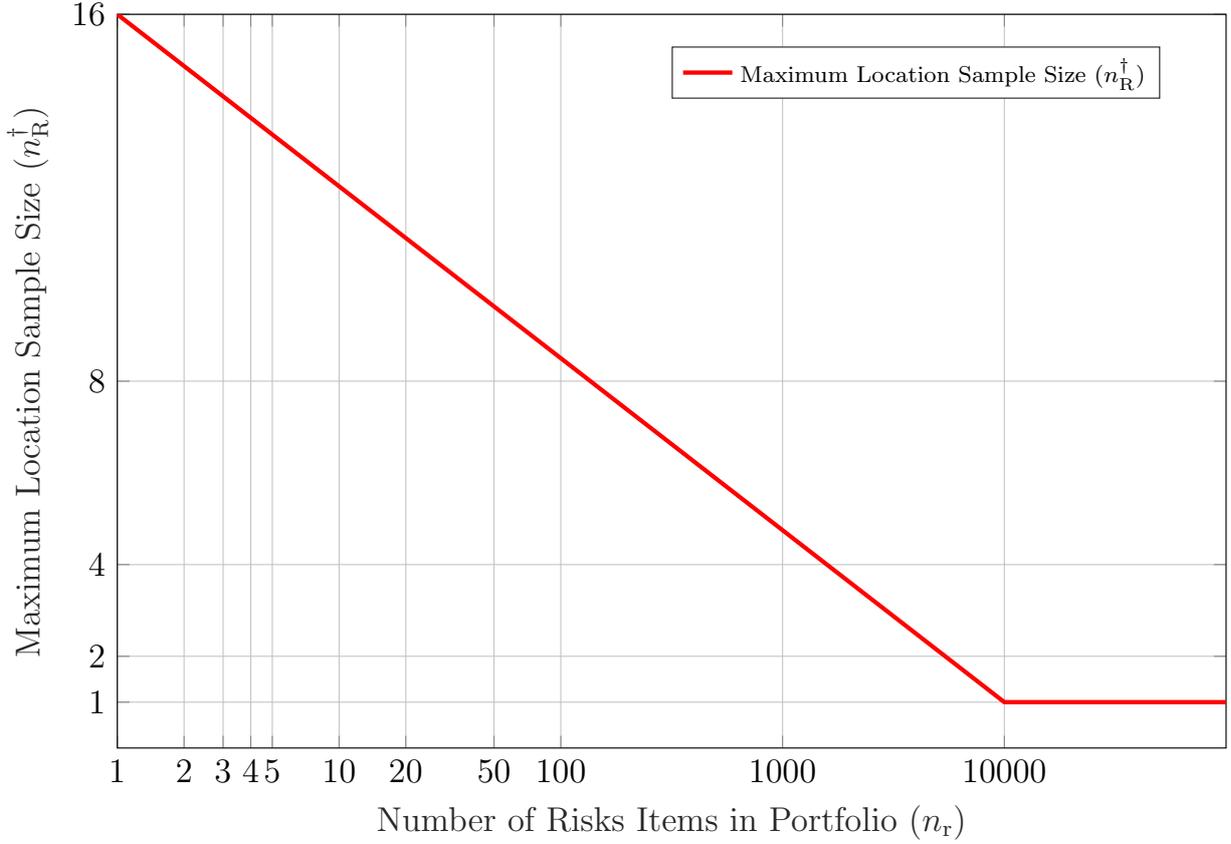


Figure 5.8: Criterion II.a: Number of samples per zone depending on the number of risk items in the portfolio. Note that we do not round up to the next higher power of two, since this plot illustrates Equation 5.14, which is an intermediate step.

5.4.3 Criterion II: Number of Risk Items

The second criterion involves two steps. The first step defines a maximum sample size for the entire portfolio depending on the total number of risk items n_r in the portfolio and a threshold t_p as

$$n_R^\dagger = \begin{cases} -\frac{n_{\max}-1}{\log(t_p-1)} \cdot \log(n_r - 1) + n_{\max}, & \text{if } n_r < t_p, \\ 1, & \text{if } n_r \geq t_p, \end{cases} \quad (5.14)$$

which is then used to obtain a maximum sample size per zone, depending on the number of risk items in a zone n_z and a threshold t_z :

$$n_R^* = \begin{cases} -\frac{n_R^\dagger-1}{t_z-1} \cdot (n_z - 1) + n_R^\dagger, & \text{if } n_z < t_z, \\ 1, & \text{if } n_z \geq t_z. \end{cases} \quad (5.15)$$

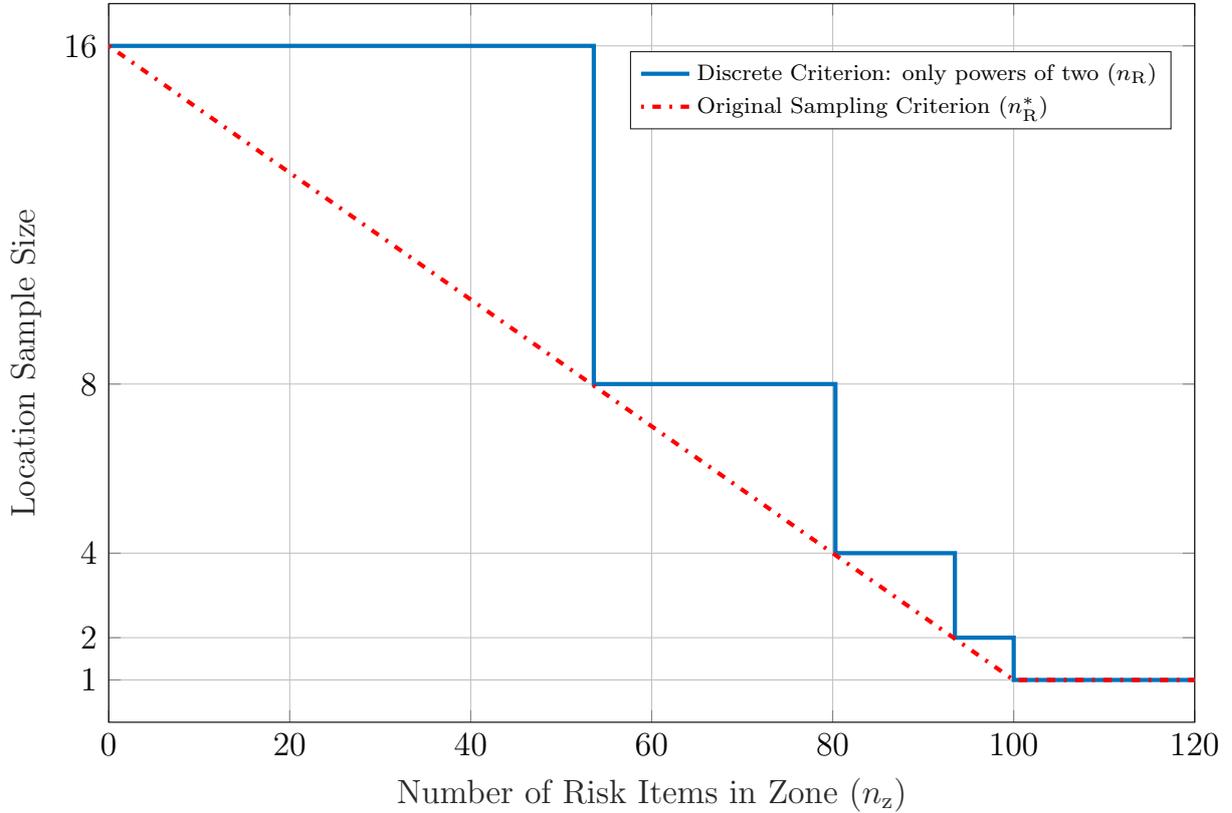


Figure 5.9: Criterion II.b: Maximum number of samples per zone depending on the number of risk items in an administrative zone. The discrete realization of the criterion (limited to powers of two) is shown in blue, while the red line represents the theoretical linear behavior.

We round n_R^* up to the next higher power of two to obtain the final n_R . Figures 5.8 and 5.9 illustrate this criterion for $t_p = 10000$, $t_z = 100$ and $n_{\max} = 16$. In this study, t_p is chosen to be 10000 and t_z is set adaptively to equal the number of grid points of the weighted location uncertainty sampling grid (see Section 5.2.2) inside each administrative zone. The design of this criterion is based on the results of a previous study, in which we systematically investigated the effect of location uncertainty and loss aggregation due to spatial clustering of risk items for a large range of different portfolios. It was found that location uncertainty typically has a neglectable effect for very large portfolios and a roughly flat value distribution (Scheingraber and Käser, 2018, Chapter 4).

5.4.4 Criterion III: Value Distribution

The third criterion depends on the relative insured values of risk items (“sum insured”, SI). Risk items are sorted with respect to their SI, and the index of their sorted order I_r is

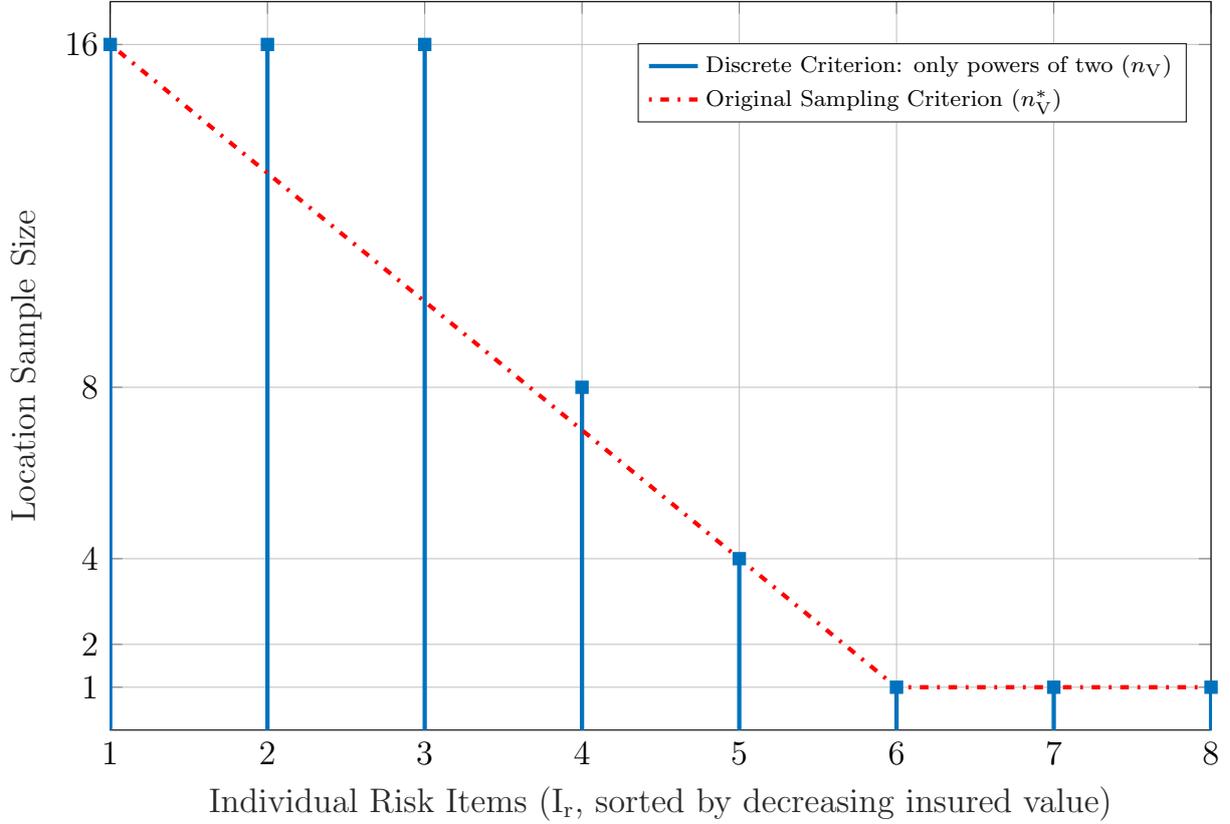


Figure 5.10: Criterion III: Number of samples per zone depending on the insured values of the risk items. The discrete realization of the criterion (limited to powers of two) is shown in blue, while the red line represents the theoretical linear behavior.

used along with a threshold index t_i to determine the maximum sample size per risk item:

$$n_V^* = \begin{cases} -\frac{n_{\max}-1}{t_i-1} \cdot (I_r - 1) + n_{\max}, & \text{if } I_r < t_i, \\ 1, & \text{if } I_r \geq t_i. \end{cases} \quad (5.16)$$

We round n_V^* up to the next higher power of two to obtain the final n_V . Figure 5.10 illustrates this criterion for $t_i = 6$ and $n_{\max} = 16$. In this study, for t_i we adaptively set the index of the first risk item which has a SI higher than the mean of all risk items.

5.4.5 Combination of Criteria

The final sample size for a specific risk item is then given by the minimum of the three criteria:

$$n = \min\{n_L, n_R, n_V\}. \quad (5.17)$$

The rationale behind this decision is that any of the criteria can separately predict that a particular risk item has a low impact on loss uncertainty. For example, if a risk item with

an unknown coordinate has a low insured value, it has a relatively low impact on loss uncertainty even if the variation of hazard or loss rate within the corresponding administrative zone is high, and thus a small location uncertainty sample size can be used.

Vice versa, the impact of location uncertainty is limited if a risk item with an unknown coordinate has a high insured value but the hazard within the corresponding administrative zone is relatively flat. Furthermore, loss uncertainty is also limited if a portfolio contains a very high number of total risk items or the number of risk items belonging to an administrative zone is high compared to the number of grid points within this zone.

5.5 Results

In this section, the variance reduction and speedup obtained with the proposed adaptive location uncertainty sampling scheme is analyzed using the western Indonesia hazard model described in Section 5.3.1 in conjunction with a vulnerability model for regional building stock composition. To this end, loss frequency curves are computed for the synthetic portfolios described in Section 5.2.4 with simple MC as well as the adaptive scheme. The convergence and relative standard errors are evaluated against the number of unique hazard locations used for the loss calculation by either approach and the associated required runtime is compared.

5.5.1 Spatial Variation Parameter Study

We first analyze the performance of the adaptive sampling scheme for different values of the lower (t_l) and upper (t_u) threshold parameters for the spatial variation of loss rate in an administrative zone in comparison to simple sampling. In simple MC, all risk items get the same location uncertainty sample size n_{\max} and there is not restriction to powers of two. For this parameter study, we use values of $n_{\max} = 32, 64, 96, 128, 160, 192, 224, 256$ in order to obtain a smooth curve with a high number of support points.

For the adaptive variance reduction scheme, the sample size is restricted to powers of two and is determined for each risk item individually — potentially smaller than the maximum allowed location uncertainty sample size n_{\max} (see Section 5.4 and Table 5.1). Since the sample size varies between risk items, for a meaningful comparison with simple MC it is necessary to use a measure of the total effort spend for the treatment of location uncertainty of all risk items. We use the total number of unique hazard locations (n_{hazard}) and the runtime spent for the computation of hazard (t_{hazard}). While for simple MC all risk items get the maximum sample size n_{\max} , the adaptive location sampling scheme reduces the sample size for risk items for which location uncertainty likely has a smaller influence. This means that the adaptive location sampling scheme results in a smaller n_{hazard} than simple MC for the same portfolio and n_{\max} . Therefore, in order to obtain a comparable values for n_{hazard} , a larger maximum sample size n_{\max} has to be employed for the adaptive scheme than for simple MC. Here, we use $n_{\max} = 2^i$ with $i = 5, 6, \dots, 8$.

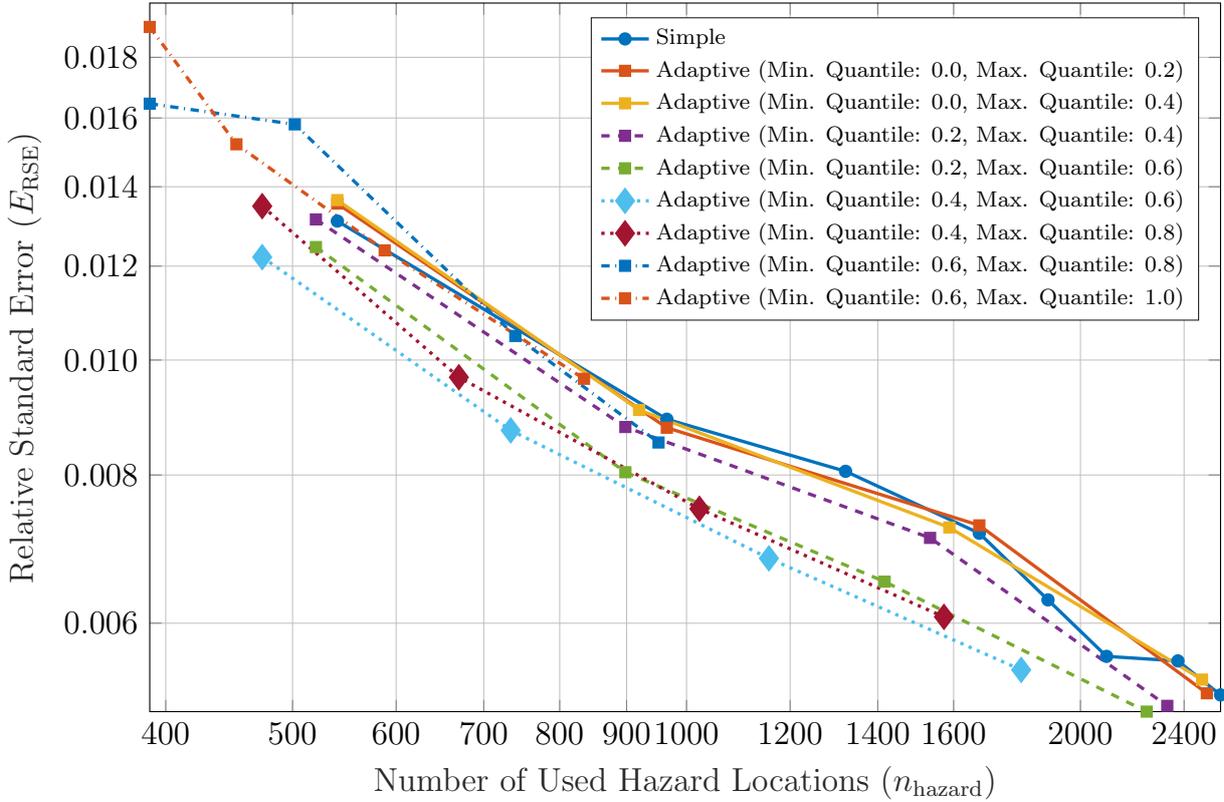


Figure 5.11: Results of a systematic parameter study with the goal of finding good values for the lower threshold (t_l) and upper threshold (t_u) parameters of Criterion I of the adaptive location uncertainty sampling scheme (see Section 5.4), based on the distribution of the Coefficient of Variation (CV) of loss rate in administrative zones. This shows a logarithmic plot of relative standard error (E_{RSE}) of Probable Maximum Loss (PML) at a return period of 100 years against the number of used hazard locations (n_{hazard}) for a portfolio of 20 risk items with 100% unknown coordinates. Color indicates different combinations for the threshold parameters t_l and t_u . Quantiles of the CV distribution around $t_l \in [\text{CV}_{0.2}; \text{CV}_{0.4}]$ in combination with $t_u \in [\text{CV}_{0.6}; \text{CV}_{0.8}]$ work best.

For each sample size, the spatial variation threshold parameters are varied over the distribution of CV values, picking quantiles in constant steps of 0.2. The lower threshold t_l is varied from $\text{CV}_{0.0}$ to $\text{CV}_{0.8}$, and the upper threshold t_u from $\text{CV}_{0.2}$ to $\text{CV}_{1.0}$. For each combination of t_l and t_u , $R = 20$ repeated simulations were performed for each sample size to estimate the respective relative standard error E_{RSE} .

In general, for our test cases the scheme works well around $t_l \in [\text{CV}_{0.2}; \text{CV}_{0.4}]$ in combination with $t_u \in [\text{CV}_{0.6}; \text{CV}_{0.8}]$. For example, for a portfolio of 20 risk items and 100% unknown coordinates, Figure 5.11 shows a logarithmic plot of the relative standard error E_{RSE} of PML at a return period of 100 years against the number of used hazard locations

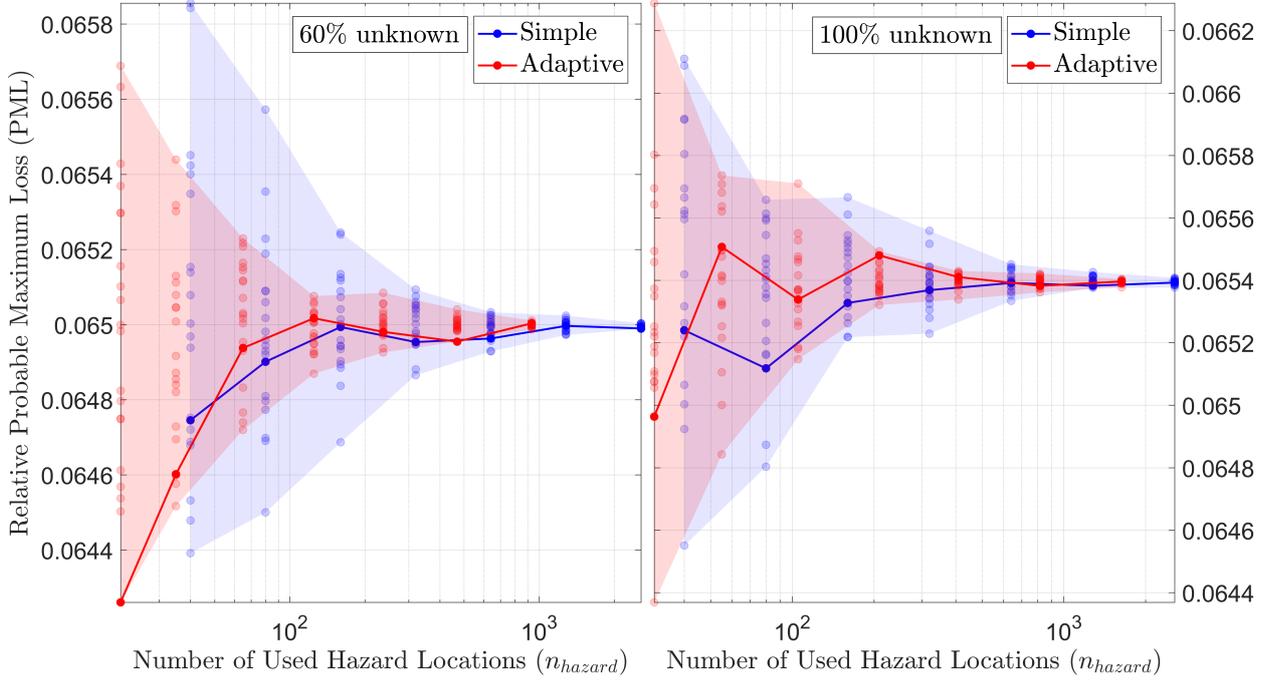


Figure 5.12: Convergence plots showing relative Probable Maximum Loss (PML) at a return period of 100 years against the number of used hazard locations n_{hazard} for portfolios of $n_r = 10$ risk items with 60% (left plot) and 100% (right plot) unknown coordinates using simple MC (shown in blue) as well as the adaptive scheme (shown in red). Semi-transparent circles depict $R = 20$ repeated simulations for each sample size, solid lines highlight one repetition. The transparently shaded background shows the entire range for each sampling scheme. The plots show that the adaptive scheme scatters less and converges faster to the same result as simple sampling.

n_{hazard} for some combinations of t_l and t_u . The error curves for all combinations of t_l and t_u have the same slope as the curve for simple MC and thus the same convergence order of $\mathcal{O}(n^{-0.5})$. For certain combinations, the error curve is below the curve for simple MC, meaning that in these cases the scheme successfully reduces the variance of the estimation and therefore the associated standard error.

For the final implementation, we used $t_l = CV_{0.4}$ and $t_u = CV_{0.6}$, which performed best in this parameter study.

5.5.2 Performance of the Final Implementation

We now evaluate the performance of the final implementation of the adaptive scheme, checking if it results in any unwanted systematic bias and investigating variance reduction and speedup for the calculation of PML for different portfolios.

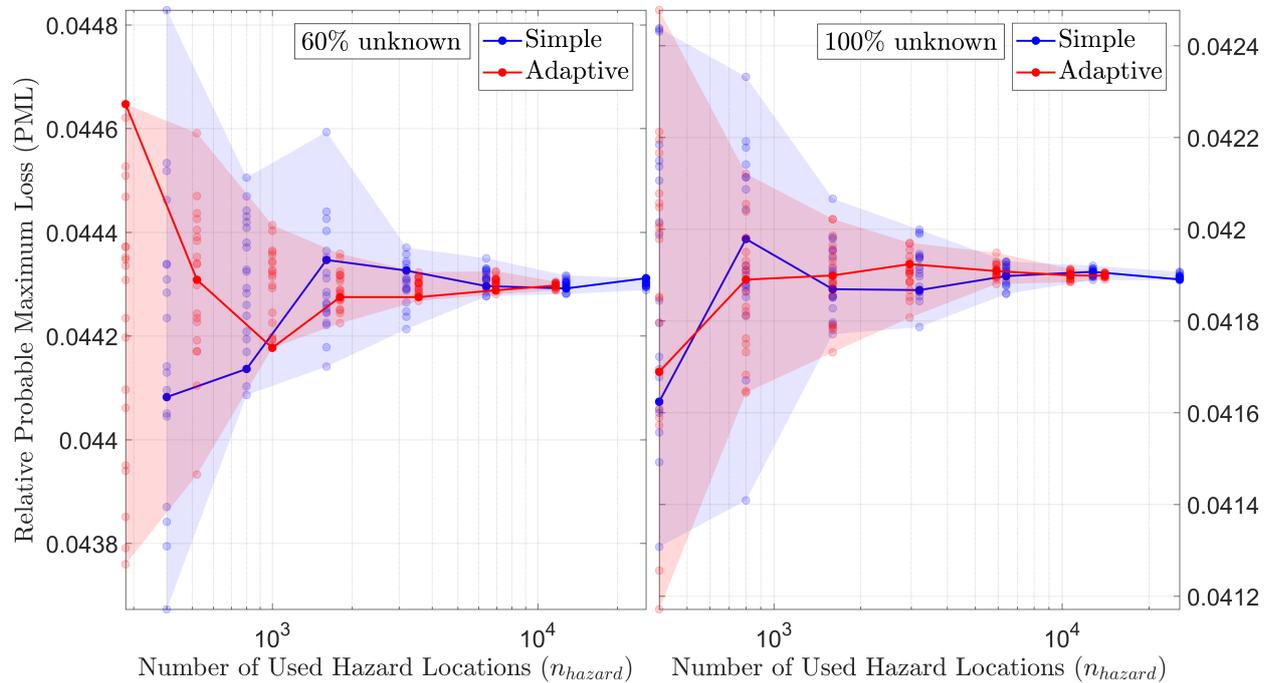


Figure 5.13: Convergence plots showing relative Probable Maximum Loss (PML) at a return period of 100 years against the number of used hazard locations n_{hazard} for portfolios of $n_r = 100$ risk items with 60% (left plot) and 100% (right plot) unknown coordinates using simple MC (shown in blue) as well as the adaptive scheme (shown in red). Semi-transparent circles depict $R = 20$ repeated simulations for each sample size, solid lines highlight one repetition. The transparently shaded background shows the entire range for each sampling scheme. The plots show that the adaptive scheme scatters less and converges faster to the same result as simple sampling.

Convergence and Bias

Figures 5.12 and 5.13 show convergence plots of PML at 100 years return period against the number of used hazard locations n_{hazard} for portfolios with $n_r = 10$ and $n_r = 100$ risk items, respectively. The left plots depict the results for portfolios with 60% unknown coordinates, the right plots the results for portfolios with 100% unknown coordinates. Simple sampling is shown in blue, the adaptive scheme in red. For all portfolios, the sample size n was varied as $n = 2^i$ with $i = 3, 4, \dots, 9$. For each sample size and both sampling schemes $R = 20$ repeated simulations are shown as semi-transparent circles, with solid lines highlighting one individual repetition.

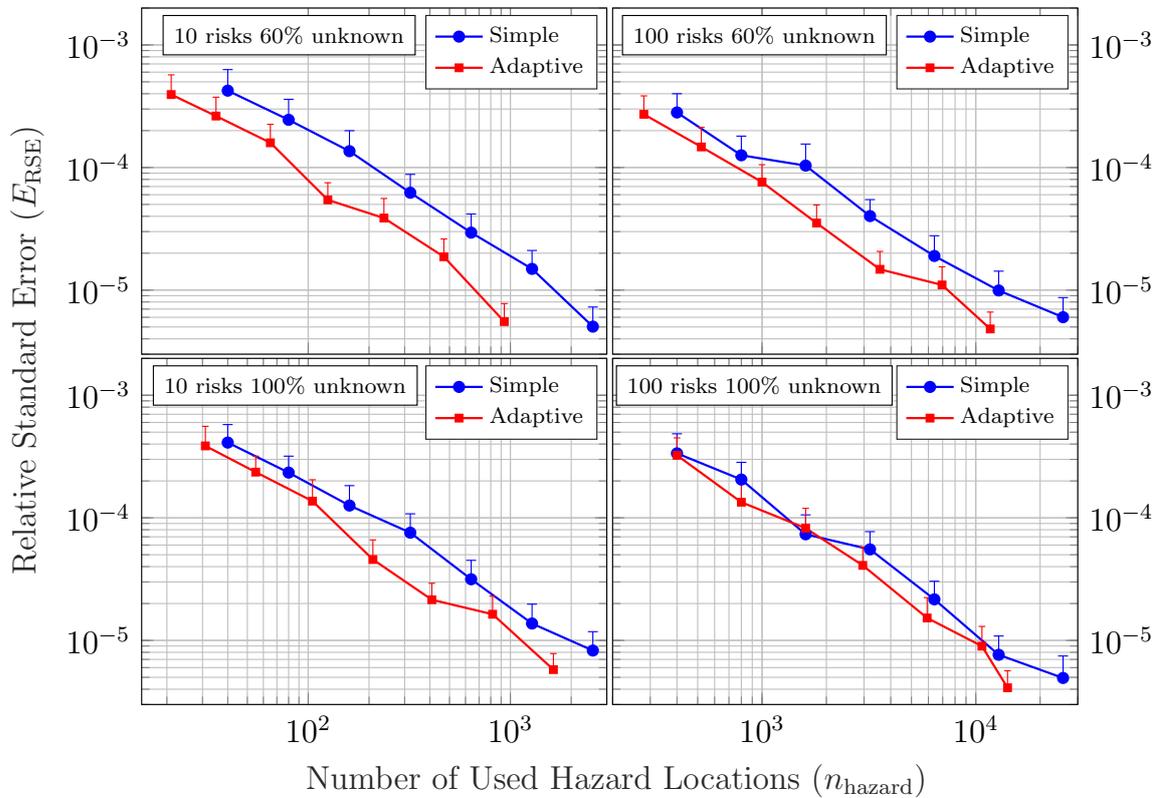


Figure 5.14: Logarithmic plot of relative standard errors E_{RSE} of Probable Maximum Loss (PML) at a return period of 100 years against the total number of used hazard locations n_{hazard} for different portfolios with $n_r = 10$ (left plots) and $n_r = 100$ (right plots) risk items and 60% (upper plots) and 100% (lower plots) unknown coordinates. Simple MC is shown in blue, the adaptive variance reduction scheme in red. All E_{RSE} have been obtained from $R = 20$ repeated simulations, vertical error bars depict upper 95% confidence intervals estimated using bootstrapping with 1000 resamples.

The results show that empirically the adaptive scheme converges to the same result as simple MC for our test cases, meaning that the scheme does not result in any systematic bias. It is also apparent that for a given number of used hazard locations n_{hazard} , the relative PML values obtained with the adaptive scheme scatter less than those estimated with simple MC.

Table 5.2: Mean runtime speedup and standard errors ($S \pm E_{SE}$) of the hazard computation achieved by the adaptive location uncertainty sampling scheme in comparison to simple sampling to obtain relative standard error levels of $\varepsilon_{RSE} = 10^{-4}$ and $\varepsilon_{RSE} = 10^{-5}$, estimated from $R = 20$ repeated simulations. Depending on the portfolio and ε_{RSE} , the mean speedup ranges from 6% to 37%.

Portfolio	Speedup (S)	
	$\varepsilon_{RSE} = 10^{-4}$	$\varepsilon_{RSE} = 10^{-5}$
10 risk items, 60% unknown coordinates	1.24 ± 0.09	1.14 ± 0.04
10 risk items, 100% unknown coordinates	1.35 ± 0.06	1.37 ± 0.09
100 risk items, 60% unknown coordinates	1.08 ± 0.04	1.06 ± 0.03
100 risk items, 100% unknown coordinates	1.09 ± 0.03	1.08 ± 0.02

Variance Reduction and Speedup

For the same portfolios as analyzed in the previous section, Figure 5.14 shows logarithmic plots of the relative standard error E_{RSE} obtained from $R = 20$ repeated simulations against the number of used hazard locations n_{hazard} . Vertical bars depict upper 95% confidence intervals estimated using bootstrapping with 1000 resamples. Simple MC is again shown in blue, the variance reduction sampling scheme in red. While the observed error convergence order of the adaptive scheme remains the same as for simple MC (i.e. $\mathcal{O}(n^{-0.5})$, compare Section 5.5.1), the error curves are below those for simple MC for all portfolios.

The variance reduction quotient (VR, the ratio of the variances of the estimations obtained using simple MC and the adaptive scheme, see Equation 5.7) varies between portfolios with different number of risk items and fractions of unknown coordinates, but generally increases with growing n_{hazard} . For example, for the portfolio with 10 risk items and 60% unknown coordinates, VR is about 6.2 at $n_{\text{hazard}} = 10^2$ and increases to 13.2 at $n_{\text{hazard}} = 10^3$. For the portfolio with 10 risk items and 100% unknown coordinates, $VR \approx 1.8$ at $n_{\text{hazard}} = 10^2$ and 2.2 at $n_{\text{hazard}} = 10^3$. For the portfolios with 100 risk items, the situation is similar. For 60% unknown coordinates, $VR \approx 2.4$ at $n_{\text{hazard}} = 10^3$ and 3.7 at $n_{\text{hazard}} = 10^4$. For 100% unknown coordinates, $VR \approx 1.7$ at $n_{\text{hazard}} = 10^3$ and 3.0 at $n_{\text{hazard}} = 10^4$.

The obtained variance reduction partially leads to a speedup of the computational runtime to reach a specific relative standard error level ε_{RSE} . Table 5.2 shows the speedup S of the scheme to reach relative standard error levels of $\varepsilon_{RSE} = 10^{-4}$ and $\varepsilon_{RSE} = 10^{-5}$ for the same portfolios. Depending on the portfolio, the scheme achieves a speedup between 8% and 35% to reach $\varepsilon_{RSE} = 10^{-4}$, and between between 6% and 37% to reach $\varepsilon_{RSE} = 10^{-5}$. Note that we obtained these speedup values using a highly optimized seismic hazard and risk analysis framework. We suspect that the scheme can result in a significantly higher speedup for less optimized code, especially if the hazard simulation is not vectorized but contains a loop over locations.

5.6 Conclusions

In seismic risk assessment the exact location of risks is often unknown due to geocoding issues of address information. Therefore, in this chapter we propose a novel adaptive sampling strategy to efficiently treat this location uncertainty using a seismic hazard and risk model for western Indonesia. The adaptive scheme considers three criteria to decide how often an unknown risk coordinate has to be sampled within a known administrative zone: (1) the loss rate variation within the zone, (2) the number of risks within the zone, and (3) the individual value of the risk. As the variation of hazard can vary quite strong not only between different administrative geographical zones, but also between different return periods, we use the spatial variation of loss rate which displays a similar pattern as the variation of hazard, but is independent of the return period. Furthermore, the total number of risks in the corresponding administrative zone, as well as the value (importance) of the risk with respect to the entire portfolio are considered by the adaptive scheme.

We investigated the performance of the scheme for a large range of sample sizes using different synthetic portfolios of different levels of unknown risk locations. We have found that the scheme successfully reduces the expected error, i.e. it reaches the same error levels as simple Monte Carlo with less samples of potential risk locations. This results in lower memory requirements and a moderate but appreciable runtime speedup to reach a desired level of reliability when computing loss frequency curves — a critical measure of risk in the insurance industry. The scheme could also be applied to other natural perils, such as probabilistic wind and flood models.

While the proposed scheme already successfully reduces the variance of loss frequency curve estimations, future improvements in the treatment of uncertainty in PSRA are conceivable. The computation might become yet more efficient by the application of variance reduction techniques to other uncertainties, for example in the ground motion and vulnerability models. Moreover, it would be essential to investigate the relative importance of location uncertainty in comparison to these other uncertainty types.

Acknowledgments

This work was supported by a scholarship provided to the first author by Munich Re.

Chapter 6

Uncertainty Quantification for Seismic Risk Assessment using Latin Hypercube Sampling and Quasi Monte Carlo Simulation

After having addressed the efficiency of the Monte Carlo simulation for portfolio location uncertainty in the previous chapter, we now turn our attention to the other uncertainty types described in Chapter 3. We analyze the performance of several well known sampling strategies for the simulation of seismic losses and quantify the relative impact of various uncertainty types using variance-based global sensitivity analysis.

Most of this chapter has been published as a peer-reviewed article in:

Scheingraber, C. and Käser, M. (2018).

Uncertainty Quantification for Seismic Risk Assessment using Latin Hypercube Sampling and Quasi Monte Carlo Simulation.

Paper No. 11221, Proceedings of the 16th European Conference on Earthquake Engineering, Thessaloniki, Greece, 2018.

Abstract

In the insurance industry Seismic Risk Assessment is commonly used for modeling loss to a spatially distributed portfolio. Best practice not only involves the computation of expected loss, but also requires treatment of the uncertainty of all components of the modeling chain. Because the dimensionality is high, this is typically performed with a Monte Carlo simulation of a large number of scenario realizations.

In this study, we first compare the computational efficiency of uncorrelated pseudo-random sampling to variance reduction techniques for scenario loss uncertainty treatment. We observe that Latin Hypercube sampling as well as Quasi Monte Carlo simulation using low-discrepancy sequences can improve the error convergence from $O(n^{-0.5})$ to $O(n^{-1})$ in many cases. The adaptive MISER algorithm, based on recursive stratified sampling, improves the error convergence to $O(n^{-1})$ only for very small portfolios and does not provide any convincing advantage over Latin Hypercube Sampling or Quasi Monte Carlo simulation for our purposes. We then perform a global sensitivity analysis to quantify the contribution of different modeling parameters and their uncertainties to the overall loss uncertainty. To this end we use three scenarios in Indonesia and explore uncertainty in the geographical distribution of portfolio items, structural properties such as building height and quality, as well as ground motion and damage models. We find that depending on the portfolio a significant fraction of the output variance can be attributed to uncertain factors in the exposure and vulnerability models, revealing the importance of their thorough treatment in seismic risk analysis.

6.1 Introduction

Probabilistic Seismic Risk Assessment (PSRA), building upon Probabilistic Seismic Hazard Assessment (PSHA; Cornell, 1968; Senior Seismic Hazard Committee (SSHAC), 1997), is widely used in the insurance industry to model the frequency and severity of losses to a geographically distributed portfolio from the occurrence of earthquake events. In this context it is not sufficient to only compute expected loss, but the treatment of uncertainty in all parts of the modeling chain is of immense importance (Crowley, 2014). For practical purposes, model uncertainty can be categorized into being either epistemic or aleatory. The term epistemic describes uncertainty due to limited knowledge or data and is commonly treated with a logic tree combining multiple alternative models. Weights of the logic tree represent the degree of belief in the correctness of a branch. Aleatory refers to variability inherent to natural processes which is assumed to be irreducible and usually captured with a probability distribution.

Once uncertainty treatment is integrated, result uncertainty can be visualized and communicated to decision makers. Moreover, the contribution of individual factors to the output variation can be quantified using sensitivity analysis (SA). This allows to e.g. identify areas where additional research or effort to reduce the associated uncertainty might be worthwhile. In contrast to local methods which investigate the impact of incremental input perturbations at a base case, global SA aims at exploring the entire space of uncertain input factors and thereby allows to take factor interactions into account. Global SA is a computationally demanding technique, because in general a high-dimensional input space needs to be sampled. This motivates our investigation of the efficiency of various sampling schemes in the first part of this chapter before we perform a global SA for seismic risk analysis using three loss scenarios in Indonesia in the second part.

6.2 Methodology

6.2.1 Modeling Framework and Uncertainty Treatment

Ground Motion Model

Ground Motion Prediction Equations (GMPEs) are used to model the distribution of ground motion at a site \mathbf{r} given the occurrence of an earthquake event of magnitude m . Ground motion levels are expressed by intensity measures such as Peak Ground Acceleration, Peak Ground Velocity, or Pseudo-Spectral Acceleration at a given frequency.

A simplified form for expected ground motion is generally given by

$$y(m, \mathbf{r}) = c_1 + c_2 m + c_3 \ln(\mathbf{r} + r_0), \quad (6.1)$$

with coefficients c_i and r_0 .

Most GMPEs capture the associated aleatory variability using a log-normal probability distribution. The standard deviation of this distribution is part of the GMPE. In this study, we use five different GMPEs (see Section 6.2.4) and sample the total ground motion residual.

Portfolio Location Uncertainty

In the context of Seismic Risk Assessment for insurance purposes, a wide range of portfolio uncertainty exists. For many portfolios, risk items are only known to be located within some administrative zone but exact coordinates are unknown. This can be caused by inaccurate geocoding as well as reduced information accuracy between brokers and reinsurance companies or risk modelers. We treat portfolio location uncertainty in a stochastic manner by sampling risk locations within their respective administrative zone on a weighted irregular grid, which acts as a proxy for insured exposure density so that e.g. residential buildings are preferentially located in areas of high population (see Figure 6.1).

Vulnerability Function and Damage Uncertainty

In addition, it is common that building properties such as the number of stories as well as building age and the related construction quality – due to updated building codes – remain unknown to the modeler. For this study, we assume that building height and construction quality are unknown for all buildings. The associated epistemic uncertainty is treated stochastically with a logic tree approach. To model damage given a ground motion level at a site we use a zero-one-inflated Kumaraswamy distribution. This is a mixture distribution of a Bernoulli distribution on $\{0; 1\}$ and a Kumaraswamy distribution in the open interval $]0, 1[$, which allows the use of discrete probability masses f_0 and f_1 to denote the likelihood that a building suffers no damage or is completely destroyed, respectively. The remaining probability mass $1 - f_0 - f_1$ is then used to scale the Kumaraswamy distribution to model partial damage.

We employ the inverse transform method to sample building damage. Conveniently, the inverse distribution function of the zero-one-inflated Kumaraswamy distribution has a closed-form expression.

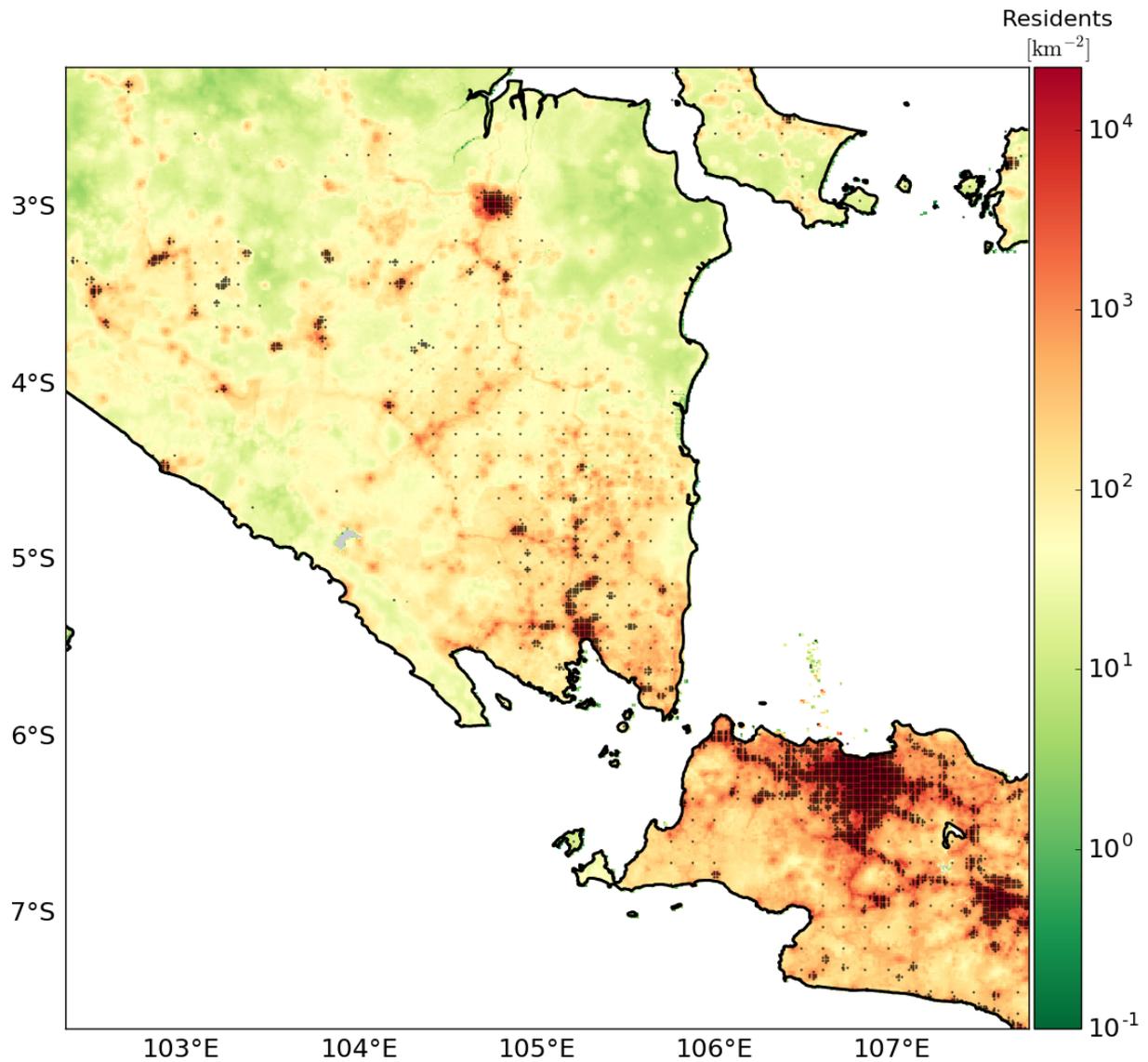


Figure 6.1: Sunda straight with the urban areas of Palembang and Jakarta. Transparent black markers depict the vertices of the weighted irregular grid employed in this study. Color represents population density (residents per km²; Gaughan et al., 2015), which we use as insurance density proxy.

6.2.2 Sampling Strategies and Error Estimation

Because a large number of model evaluations is necessary for sensitivity analysis, it is worth to explore the efficiency of different sampling strategies for our study.

Simple Monte Carlo

Stochastic sampling was first used on an electronic computer (ENIAC) during the Manhattan Project at Los Alamos Laboratory, where the still widely used name Monte Carlo (MC) was also coined. The theory was further developed and first published by Metropolis and Ulam (1949). Simple MC is based on uniform random sampling of the domain and can be used for numerical integration instead of deterministic quadrature rules. We create uncorrelated pseudo-random numbers using the Mersenne Twister pseudo-random number generator (Matsumoto and Nishimura, 1998). Given a sample x of size n , the expectation of some function $\phi(x)$, for example mean scenario loss or values along a probabilistic loss curve, is given by the unbiased estimator

$$\hat{\phi} = \frac{1}{n} \sum_{i=1}^n \phi(x_i). \quad (6.2)$$

Simple MC has a slow error convergence of $O(n^{-0.5})$. However, in contrast to deterministic quadrature schemes such as the trapezoidal rule, the convergence order of simple MC is independent of the number of parameters and only depends on the variance of the estimate $\hat{\phi}$. For this reason, the method is well suited for high-dimensional integrals such as seismic risk assessment with a large portfolio size.

Latin Hypercube Sampling

Uncorrelated random sampling points tend to cluster, which is not ideal because there is little information gain when sampling a point which is close to previous ones. Latin Hypercube Sampling (LHS; McKay et al., 1979) is a variance reduction technique which aims to improve this by stratifying the domain along each dimension. With a sample size of n , for each dimension n strata are created such that every projection to one of the dimensions is itself a stratified sample with n strata. This can result in better asymptotic error convergence, particularly when the function is additive or dominated by one parameter (Owen, 1994). While $O(n^{-1})$ convergence is often observed in practice, theoretically this order could so far only be shown for special cases.

A problem with LHS is that while the domain is stratified in each separate dimension, multidimensional combinations are not. This can be enhanced to some extent by additional design criteria. We use the maximin distance criterion (Johnson et al., 1990), which maximizes the minimum distance between two points. While LHS with sample size n never performs worse than simple MC with sample size $n - 1$ (Owen, 1997), the advantage of LHS can decrease for high-dimensional problems. Thus the performance of LHS for our study needs to be evaluated.

Quasi Monte Carlo

With Quasi Monte Carlo (QMC) methods, deterministic low-discrepancy sequences are used instead of quasi-random numbers to generate sampling points. Low-discrepancy sequences are designed to avoid previous points and fill space evenly. In this study, we use the sequence introduced by (Sobol', 1967). The Koksma-Hlawka inequality states that the QMC integration error E_n of a function $\phi(x)$ in the d -dimensional unit cube is bounded by

$$E_n(\phi(x)) \leq D_n V(\phi(x)), \quad (6.3)$$

where $V(\cdot)$ is the variation in the Hardy-Krause sense which is finite if the integrand is smooth (Moskowitz and Caffisch, 1996). D_n is the *discrepancy* of the sequence which is $O(n^{-1} \log^d n)$ for large n , although it can be worse for intermediate n (Morokoff and Caffisch, 1994). This can potentially be improved with randomized QMC, where a deterministic sequence is scrambled randomly. We use scrambling as described in Owen (1997) and Matoušek (1998).

In practice, many MC simulations involve decisions or functions that are not smooth, like the epistemic uncertainties and the zero-one-inflated loss distribution in this study. We therefore investigate the error convergence of QMC for our model empirically.

MISER

MISER is an adaptive MC scheme based on recursive stratified sampling (Press and Farrar, 1990; Press et al., 2007). It has been extended to be used in conjunction with QMC in Schürer (2004). The basic idea of MISER can be described as follows. First, 10% of the n available sampling points are used to presample the entire d -dimensional integration domain of a function $\phi(x)$ using simple MC or QMC. The standard deviation of $\phi(x)$ in each of the $2d$ possible hyperrectangular halves of the domain is estimated. MISER is then applied recursively to the two disjoint subdomains S_α and S_β (i.e. halves) for which the sum of the standard deviations assumes its minimum, with the remaining 90% of the sampling points split according to

$$\frac{n_\alpha}{n_\beta} = \frac{\hat{\sigma}_\alpha}{\hat{\sigma}_\beta}, \quad (6.4)$$

where n_α is the number of sampling points assigned to the subdomain S_α , n_β the number of sampling points assigned to the subdomain S_β , $\hat{\sigma}_\alpha$ the estimated standard deviation of $\phi(x)$ in S_α , and $\hat{\sigma}_\beta$ the estimated standard deviation of $\phi(x)$ in S_β .

The recursion bottoms out when the number of assigned sampling points falls below a certain threshold, and simple MC or QMC is used to return an estimation of the function $\phi(x)$ in the corresponding subdomain¹. Finally, the estimations $\hat{\phi}_\alpha$ and $\hat{\phi}_\beta$ from two halves

¹An animation of inverse transform sampling of the zero-one-inflated Kumaraswamy distribution using the MISER algorithm is available at <http://archive.scheingraber.net/animation/miser.gif>.

S_α and S_β are combined as

$$\hat{\phi}_{\alpha+\beta} = \frac{\hat{\phi}_\alpha + \hat{\phi}_\beta}{2}, \quad (6.5)$$

where $\hat{\phi}_{\alpha+\beta}$ is the estimation of $\phi(x)$ in the combined region $S_\alpha \cup S_\beta$.

Estimation of Standard Error and Confidence Intervals

We use repeated simulation for MC standard error estimation of the different sampling schemes. Denoting a set of estimations of a statistic obtained from R repeated simulations by $\hat{\phi}_R$ and the variance by $V(\cdot)$, then the standard error E_{SE} is given by

$$E_{SE}(\hat{\phi}_R) = \sqrt{V(\hat{\phi}_R)}. \quad (6.6)$$

In this study, we normalize the standard error of the mean by the sample mean to obtain relative standard error E_{RSE} . For the estimation of confidence intervals of E_{RSE} , we use bootstrapping. This method assumes that the original sample holds all information about the underlying population, and can be used to estimate the sampling distribution of $\hat{\phi}$ by resampling with replacement. Specifically, we employ the bias-corrected accelerated percentile method (Efron and Tibshirani, 1986).

6.2.3 Sensitivity Analysis

Seismic Risk Assessment relies on many uncertain parameters. Awareness of model uncertainties and knowledge of the extent to which certain factors drive output uncertainty under specific circumstances is important for risk modelers as well as end users. With sensitivity analysis (SA), it is possible to quantify the influence of uncertain model input factors. Regulatory documents and official guidelines of the European Commission and the United States Environmental Protection Agency recommend the use of SA and stress the importance to consider factor interactions (Saltelli et al., 2010).

Local SA methods use first-order partial derivatives $\partial Y_i / \partial X_i$ to evaluate the sensitivity of model output Y_i against input X_i at a predefined base case of the input space. If the modeling code itself does not return derivatives, they can e.g. be estimated with finite differences. Another very powerful approach is algorithmic differentiation, which is the automated differentiation of an entire model source code via application of the chain rule. While local methods are relatively cheap, they give only limited insight into the sensitivity of a model with respect to a variable, because they only provide valid information close to the base case where non-linear response can be neglected. They provide no information about other regions of the input space. By contrast, global SA methods explore the entire input space. This allows to quantify the overall sensitivity of the model output with respect to input factors as well as interactions between factors.

For the global SA in this study, we use the variance decomposition introduced by Sobol (2001) based on his *Analysis of Variances* (ANOVA) decomposition. The total variance $V(Y)$ of a scalar model output Y dependent on a model input vector \mathbf{X} can be decomposed into $2^d - 1$ components

$$V(Y) = \sum_{1 \leq i \leq d} V_i + \sum_{1 \leq i < j \leq d} V_{ij} + \cdots + V_{12\dots d}, \quad (6.7)$$

where d denotes the number of input factors, V_i a first-order variance term dependent only on the i th input factor and V_{ij} a second-order term dependent on the i th and j th input. $V_{12\dots d}$ is the highest-order variance term dependent on all d input factors. Higher-order terms represent variance that cannot be explained by lower-order terms, but is caused by some interaction of the involved factors. For example, the variance in V_{ij} cannot be expressed by $V_i + V_j$.

Sobol sensitivity indices express the fraction of the total variance due to a subset of the variance components. In this study, we use the first order (or *main effect*) indices which quantify the fraction of the variance caused by an input factor i to the total output variance without interactions over the whole input space:

$$S_i = \frac{V_{X_i}(E_{\mathbf{X}_{\sim i}}(Y|X_i))}{V(Y)}. \quad (6.8)$$

X_i denotes the i th input factor and V_{X_i} is the partial variance taken over factor X_i . $\mathbf{X}_{\sim i}$ denotes all input factors but the i th, which can be thought of as the “non- X_i direction” (Saltelli et al., 2010). $E_{\mathbf{X}_{\sim i}}$ is the conditional expectation taken over all factors but X_i , which means that variations in other parameters are “averaged” and the variance is taken over these averages.

In addition, we use the *total effect* indices introduced by Homma and Saltelli (1996) to quantify the contribution of the first order effect of the i th input factor together with all higher-order interactions with other factors to the total output variance:

$$S_{T_i} = \frac{E_{\mathbf{X}_{\sim i}}(V_{X_i}(Y|\mathbf{X}_{\sim i}))}{V(Y)}. \quad (6.9)$$

S_{T_i} is called the total effect index of the i th input factor. Note that because individual interaction components of the variance decomposition are reused for the computation of several total effect indices (of all input factors involved in this interaction), the sum of all total effect indices exceeds 1 unless the model itself is purely additive. Because they are normalized by their variances, the exact values of S_i and S_{T_i} are in the interval $[0; 1]$.

Estimating S_i and S_{T_i} is usually performed via MC simulation, which can be computationally demanding because convergence of the indices is often only achieved after a very large number of model evaluations (Sarrazin et al., 2016). Numerous studies are devoted to deriving efficient sampling designs for the joint estimation of main and total effects. In this study, the design proposed by Jansen (1999) is employed.

6.2.4 Earthquake Scenarios and Portfolios

We study scenario loss variations based on two hypothetical earthquake events and several synthetic portfolios in Indonesia. For this purpose we use a proprietary seismic risk assessment framework developed by MunichRe running on MATLAB.

We use three “portfolio sets”, each is only known to be distributed in a different administrative zone corresponding to an Indonesian province. Each portfolio set consists of several portfolios with a varying number of buildings (1, 5, 100, 5000 and 10000). For simplicity we use a total sum insured (TSI) of and a flat value distribution for all portfolios, i.e. losses can be interpreted directly as percentage of the TSI and the TSI is distributed equally among all risk items. For each model evaluation, locations, heights and qualities are sampled independently for all buildings, but we assume a fixed construction type (reinforced concrete with unreinforced masonry infill). To compute mean damage ratios, our model then uses customized vulnerability functions reflecting different structural performance due to varying building heights and construction qualities.

On top of this, we sample either the ground motion residuals or the damage residuals. Note that we do not sample ground motion and damage residuals jointly, because in the reinsurance industry damage distributions are fitted to include variation due to aleatory ground motion uncertainty. It might be possible to correct for this effect in the future when more detailed loss and ground motion data become available, but currently this approach is common practice. The aleatory ground motion uncertainty model is part of each GMPE, while the parameters of the zero-one-inflated Kumaraswamy distribution used to treat aleatory damage uncertainty (see Section 6.2.1) are part of the aforementioned risk assessment framework and depend on mean damage ratios as well as building heights and qualities.

For the SA, we use the SAFE toolbox (Pianosi et al., 2015), which provides workflows for several SA methods. For a variance-based global SA as performed in this study, SAFE provides functions to approximate and visualize main and total effect indices.

Southern Sumatra Subduction Fault Event

The first scenario is a hypothetical M_w 9.2 event on the Sumatra subduction fault near the urban areas of Palembang and Jakarta on the islands of Sumatra and Java, respectively. For this event, we use a three-dimensional representation of the subduction zone based on the Slab 1.0 model (Hayes et al., 2012) and sample from two GMPEs with equal weights: Zhao et al. (2006), and Youngs et al. (1997). We use two sets of portfolios: the first, hereafter labeled “Palembang portfolio set”, is only known to be distributed within the province *Sumatera Selatan* on Sumatra; the second, hereafter labeled “Jakarta portfolio set”, in the province *Daerah Khusus Ibukota Jakarta* on Java. Portfolio locations are sampled onto the weighted irregular grid inside their respective zones (see Section 6.2.1 and Figure 6.1). Figure 6.2 shows a footprint of the event obtained using the GMPE by Zhao et al. (2006) and the outline of the two administrative zones.

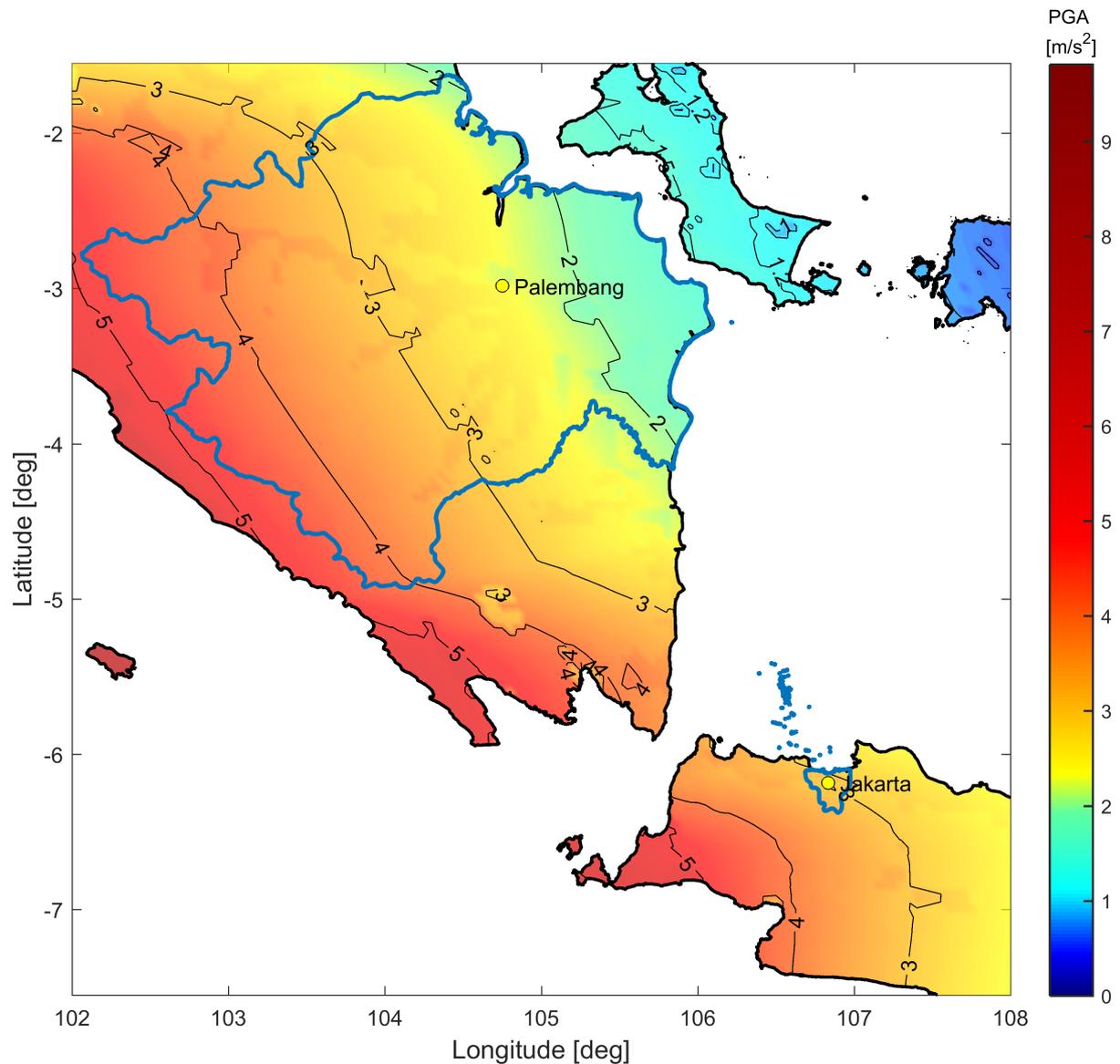


Figure 6.2: Footprint of expected ground motion of the hypothetical M_w 9.2 event on the Sumatra subduction fault near the urban areas of Palembang and Jakarta using the Ground Motion Prediction Equation (GMPE) by Zhao et al. (2006). Color and isolines denote Peak Ground Acceleration (PGA) in m/s^2 . The provinces of *Sumatera Selatan* on the island of Sumatra and *Daerah Khusus Ibukota Jakarta* on Java are outlined by their boundaries in blue color.

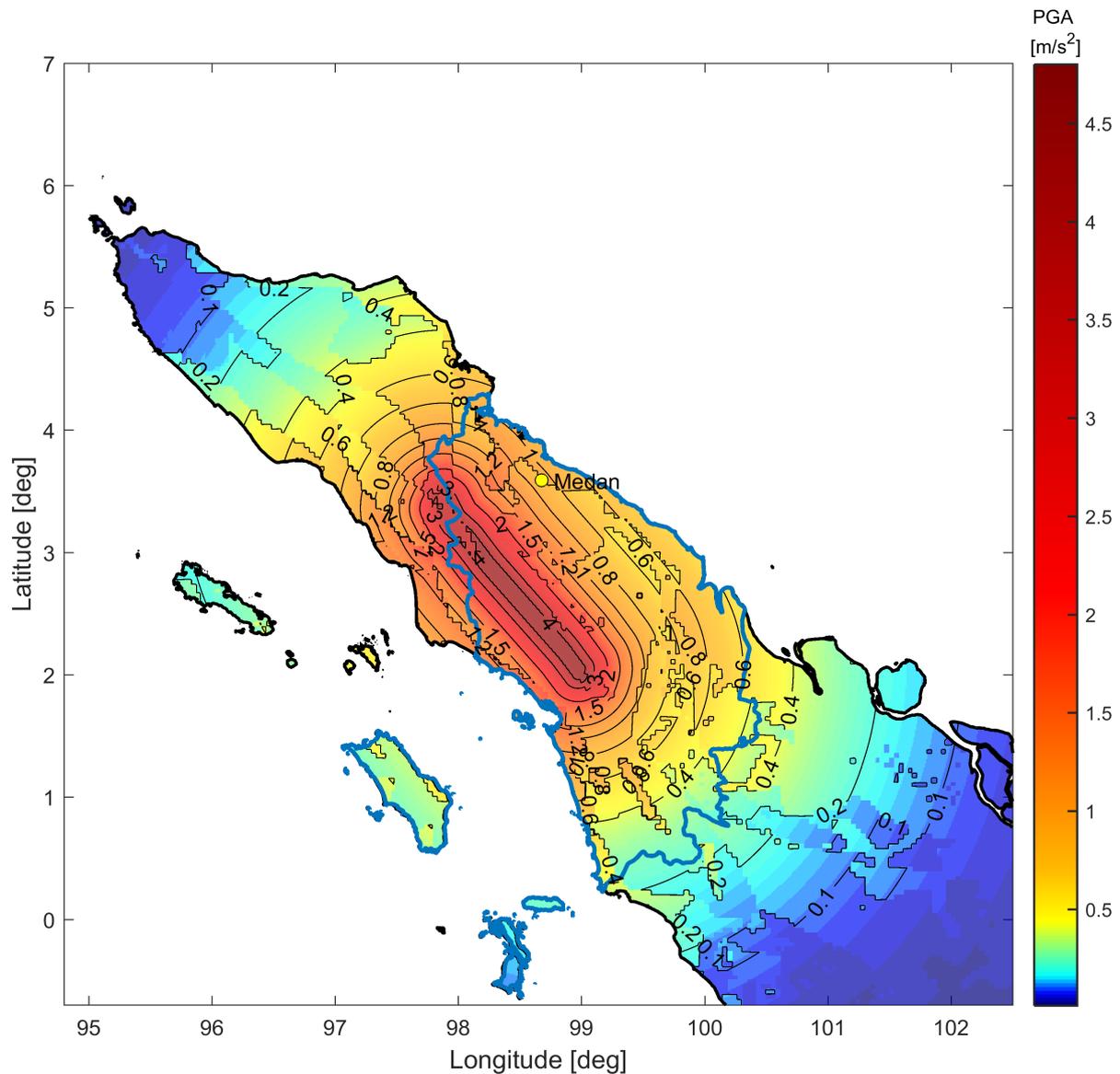


Figure 6.3: Footprint of expected ground motion of the hypothetical M_w 7.8 event on a northern segment of the Sumatra Fault Zone near the city of Medan using the Ground Motion Prediction Equation (GMPE) by Chiou and Youngs (2008). Color and isolines denote Peak Ground Acceleration (PGA) in m/s^2 . The province of *Sumatera Utara* on the island of Sumatra in which we sample location uncertainty is outlined by its boundary in blue color.

Northern Sumatra Fault Zone Event

The second scenario is a hypothetical M_w 7.8 event on a northern segment of the Sumatra fault zone near the urban area of Medan. For this event, we sample from three different GMPEs with equal weights: Boore and Atkinson (2008), Campbell and Bozorgnia (2008), and Chiou and Youngs (2008). We use one portfolio set for this event, hereafter referred to as “Medan portfolio set”, for which risk items are only known to be located somewhere in the administrative zone *Sumatera Utara*. For each model run, risk item locations are sampled onto the weighted irregular grid inside this area (see Section 6.2.1). Figure 6.3 shows a footprint of this event obtained using the GMPE by Chiou and Youngs (2008) and the outline of the administrative zone.

6.3 Results

6.3.1 Error Convergence of Sampling Strategies

In this section, we analyze the performance of different sampling schemes described in Section 6.2.2 for the scenarios and portfolios described in Section 6.2.4. The discontinuous zero-one-inflated damage distribution (see Section 6.2.1) as well as the high dimensionality of large portfolios provide an interesting and challenging test case.

Figure 6.4 shows estimated event losses $\hat{\phi}$ for the Palembang portfolio set with 1 and 5 risk items against ten sample sizes $n = 2^i$ with $i = 2, 3, \dots, 11$ for 50 repeated simulations for each sample size. All sampling schemes converge to the same solution, but the estimations obtained with Latin Hypercube Sampling (LHS) with the maximin design criterion and Quasi Monte Carlo with the scrambled Sobol sequence (SSobol) converge faster with less variable loss estimates. The portfolio with 5 risk items has less variation than the portfolio with 1 risk item due to the diversification induced by uncorrelated sampling of individual building losses. A correlation model such as a spatial ground motion correlation model (e.g. Jayaram and Baker, 2009) with a spatially clustered portfolio or any type of damage correlation model would act to lessen this effect. Other portfolio value distributions than *flat* (see Section 6.2.4) would also show relatively higher variability.

To analyze the convergence order of the different global sampling schemes, Figure 6.5 shows logarithmic plots of the relative standard error E_{RSE} of $\hat{\phi}$ against ten sample sizes $n = 2^i$ with $i = 2, 3, \dots, 11$ for the Medan portfolio set, obtained from r repeated simulations ($r = 200$ for the portfolio with 1 risk item, $r = 50$ for 100 risk items, and $r = 25$ for 5000 and 10000 risk items). The thin blue and red lines indicate theoretical $O(n^{-0.5})$ and $O(n^{-1})$ convergence given the initial E_{RSE} at $n = 4$. As expected, simple MC converges slowest with $O(n^{-0.5})$ for all portfolios independently of the dimensionality. For the small portfolios with 1 and 100 risk items, SSobol and LHS perform about equally well and achieve linear convergence. For the larger portfolio sizes (5000 and 10000 risk items), LHS does not achieve $O(n^{-1})$ convergence but retains some advantage over simple MC. We do

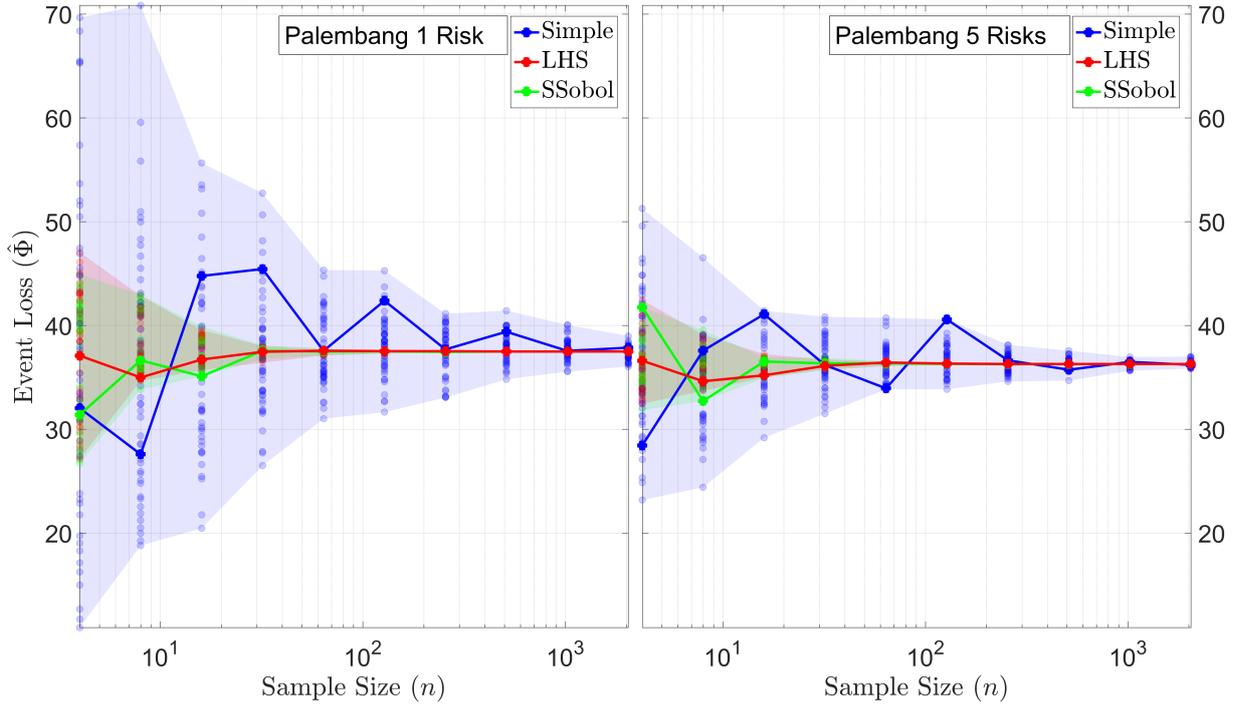


Figure 6.4: Event loss $\hat{\phi}$ versus sample size n for the M_w 9.2 Sumatra subduction zone event and the Palembang portfolio set with 1 risk item (left) and 5 risk items (right). Semi-transparent circles depict 50 repeated simulations for each sample size and sampling scheme, with solid lines highlighting one individual repetition. The transparently shaded background indicates the entire range for each sampling scheme. Estimations obtained using Latin Hypercube Sampling (LHS; red) and Quasi Monte Carlo using a scrambled Sobol sequence (SSobol; green) scatter less than those obtained with simple MC (blue).

not use the Sobol sequence for very large portfolios, because the employed algorithm only supports up to 1111 dimensions (Joe and Kuo, 2003).

To evaluate the performance of the adaptive MISER algorithm in comparison to simple MC, Figure 6.6 shows logarithmic plots of the relative standard error E_{RSE} of $\hat{\phi}$ against ten sample sizes $n = 2^i$ with $i = 2, 3, \dots, 11$ for the Jakarta portfolios with 1 and 5 risk items, obtained from 50 repeated simulations. The thin blue and red lines again indicate theoretical $O(n^{-0.5})$ and $O(n^{-1})$ convergence given the initial E_{RSE} at $n = 4$. MISER improves the convergence rate to $O(n^{-1})$ for the portfolio of 1 risk item. However, for portfolios of 5 risk items and more, the convergence rate is not improved over simple MC. Thus MISER provides no convincing advantage over LHC and QMC for our purposes, even though we use a zero-one-inflated probability distribution to model damage uncertainty.

The variant of MISER using randomized QMC for presampling and at the bottom of the recursion performs slightly better than the original version, but likewise does not improve the convergence rate for larger portfolios.

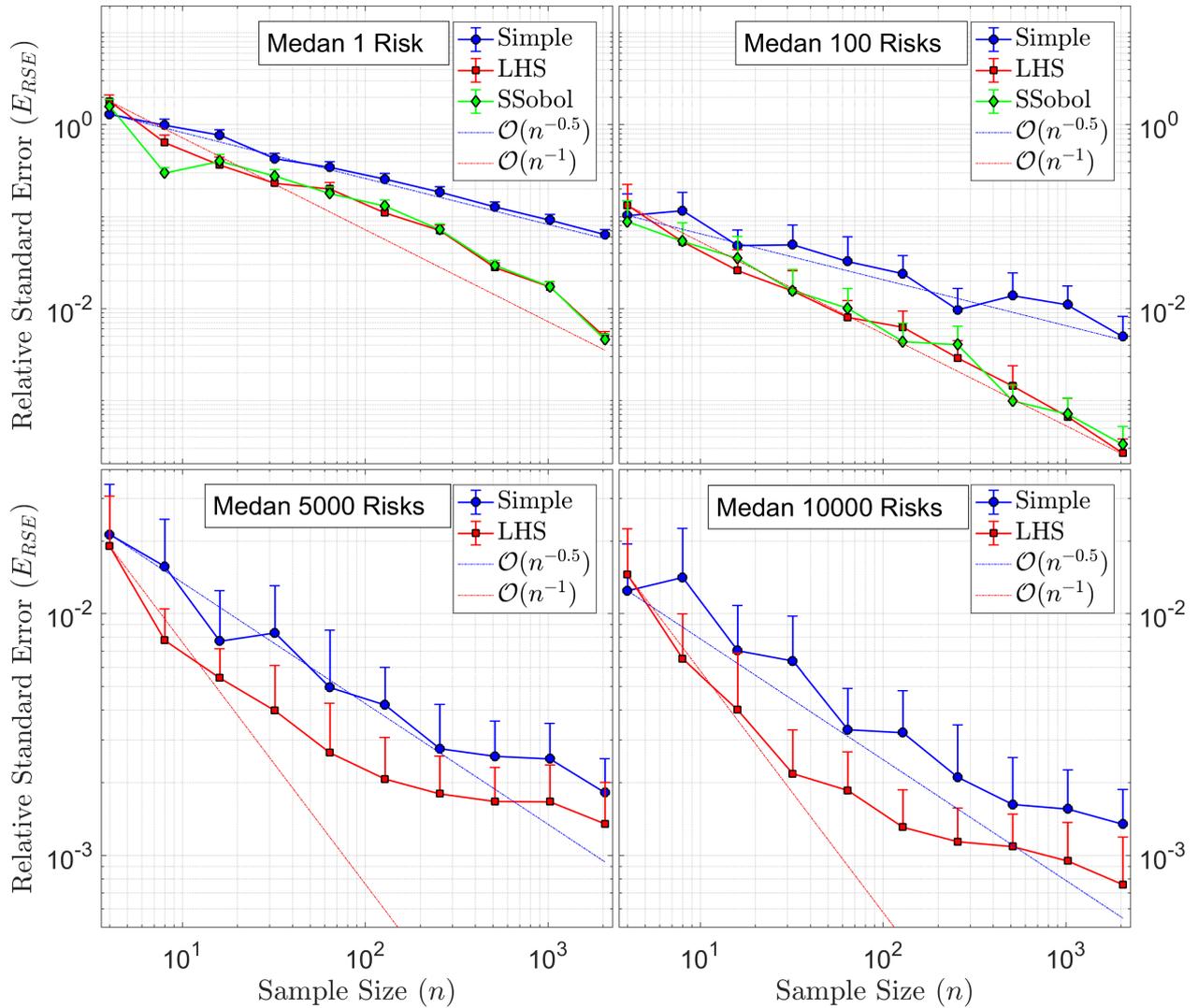


Figure 6.5: Logarithmic plot of relative standard errors E_{RSE} versus sample size n obtained from repeated simulations and bootstrapped upper 95% confidence intervals of event loss for the Medan portfolio set with 1, 100, 5000 and 10000 risk items and the M_w 7.8 Sumatra fault zone scenario. Latin Hypercube Sampling (LHS; red) and Quasi Monte Carlo using a scrambled Sobol sequence (SSobol; green) achieve $O(n^{-1})$ convergence for the small portfolios. While $O(n^{-1})$ is not achieved for larger portfolios, LHS still retains some advantage over simple sampling (blue).

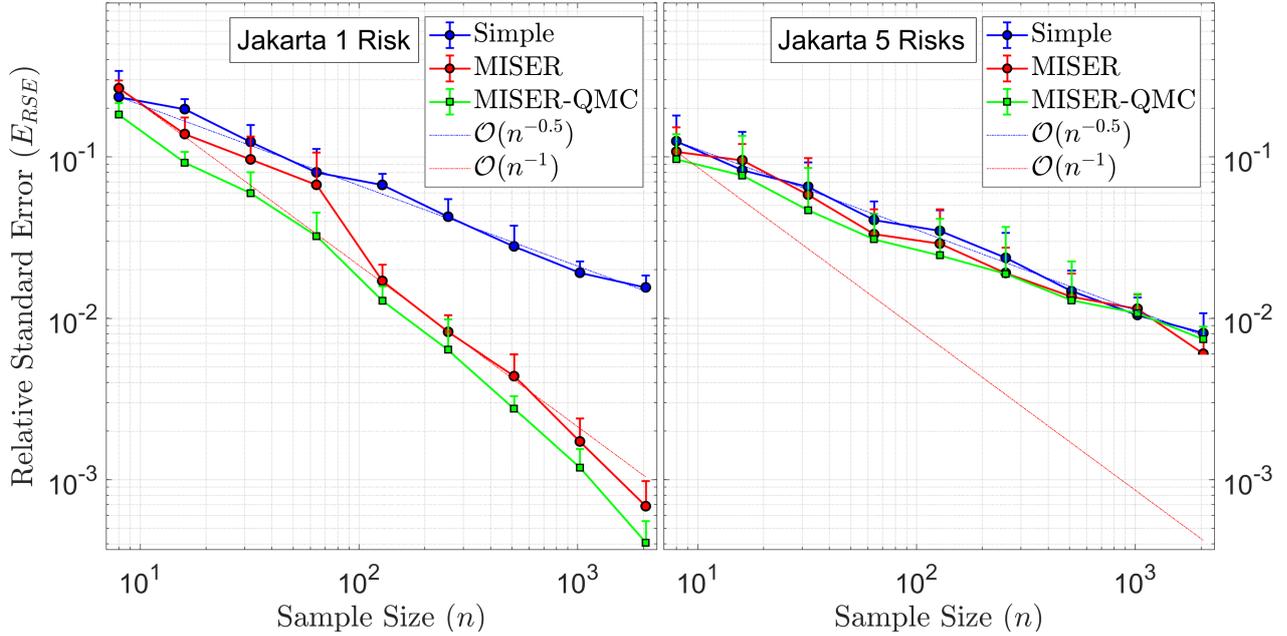


Figure 6.6: Logarithmic plot of relative standard errors E_{RSE} versus sample size n obtained from 50 repeated simulations and bootstrapped upper 95% confidence intervals of event loss for the Jakarta portfolio set with 1 risk item (left) and 5 risk items (right) for the M_w 9.2 Sumatra subduction zone event. The adaptive MISER algorithm (red) and its variant using Quasi Monte Carlo (MISER-QMC; green) improve the convergence order to $\mathcal{O}(n^{-1})$ for the portfolio with 1 risk item. However, for the portfolio with 5 risk items, both MISER and MISER-QMC do not improve the convergence order over simple sampling (blue).

6.3.2 Results of Global Sensitivity Analysis

Using the same hypothetical scenarios and synthetic portfolios in Indonesia, we performed a global SA to investigate the effect of uncertain input factors on event loss estimation.

To obtain a first impression of sensitivities, scatter plots are a simple and powerful tool. This graphical global SA technique allows to quickly assess the first-order effect of varying each factor over its entire range while also taking the global input space of other factors into account. Figure 6.7 shows scatter plots of event losses $\hat{\phi}$ versus five different input factors for the M_w 9.2 Sumatra subduction zone event and a portfolio in *DKI Jakarta* with 1 building, obtained using LHS with size $n = 2^{13} = 8192$. Each plot is a one-dimensional projection of the entire sample, in which one factor (X_i) is varied systematically while all other factors ($\mathbf{X}_{\sim i}$) are taken unconstrained over their full range. The red lines approximate $E_{\mathbf{X}_{\sim i}}(Y|X_i)$ by computing mean values of $\hat{\phi}$ inside a sliding window with length 1/10 of the range of the respective input factor. The steep slope and large range (i.e., large variance) of the red lines of the building quality and damage residual suggest that these uncertainty types have a strong effect for this scenario and portfolio.

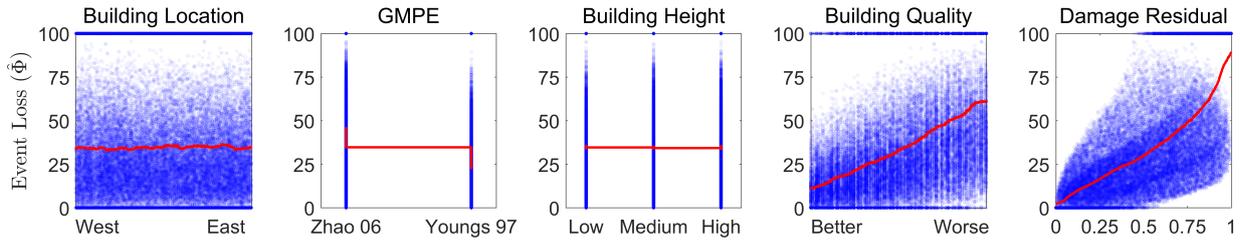


Figure 6.7: Scatter plot of event loss $\hat{\phi}$ versus individual input factors for the M_w 9.2 Sumatra subduction zone event and a portfolio in *DKI Jakarta* with 1 building obtained using Latin Hypercube Sampling. In each plot all other input factors are sampled over their entire range, which corresponds to a projection to one dimension. Semi-transparent blue markers depict individual event loss realizations. The red curves correspond to the conditional mean obtained using a sliding window with a length of 1/10 of the total input interval, thereby approximating $E_{\mathbf{X}_{\setminus i}}(Y|X_i)$.

Figure 6.8 shows the results of a variance-based global SA (see Section 6.2.3) for the same portfolio. To ensure convergence of the sensitivity indices, for this portfolio we computed $N = 229376$ model evaluations using simple MC, corresponding to a base sample size $n = 2^{15} = 32768$ (see Equation 13 of Sarrazin et al., 2016). Main effects (first order Sobol sensitivity indices S_i) are depicted as orange boxes and total effects (total sensitivity indices S_{T_i}) as blue boxes for each input factor. The MC estimation of each effect is indicated by a thin black line inside the corresponding box, while the extent of the boxes depicts 95% confidence intervals obtained using bootstrapping. Confirming the impression obtained from the scatter plot, the GMPE, the building quality and the damage distribution all have an important first order effect. All three are subject to significant higher order interactions with other factors, resulting in larger total sensitivity indices. This underlines the notion that uncertainty quantification for seismic risk analysis should not merely consider first-order effects or local sensitivities, but consider factor interactions and explore the global uncertainty space. For this scenario location uncertainty has little effect, which can be explained by the relatively small extent of the administrative zone *DKI Jakarta* (see Figure 6.2).

Figure 6.9 shows the equivalent plot for the M_w 7.8 Sumatra fault zone event and the Medan portfolio with 100 risk items. The main and total effects for the same uncertainty types as before are investigated. However, in this case the 500 individual input factors are organized into five uncertainty groups (one per uncertainty type), each containing 100 independently sampled factors corresponding to the 100 risk items. Due to the larger portfolio size, $N = 573440$ model evaluations were necessary to achieve convergence of the indices, corresponding to a base sample size $n = 10 \cdot 2^{13} = 81920$. For this scenario, location uncertainty has a strong effect because the province of *Sumatera Utara* has a large spatial extent (see Figure 6.3) even though it is on the same administrative level as *DKI Jakarta*.

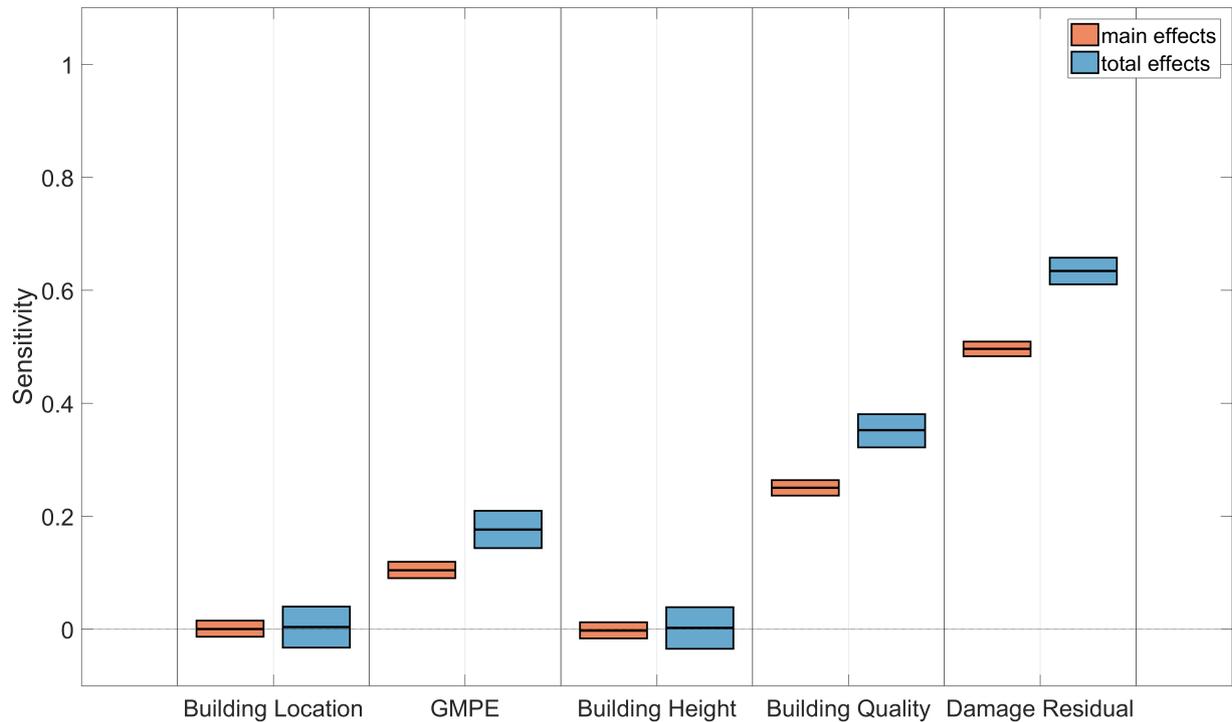


Figure 6.8: Main effects S_i (orange) and total effects S_{T_i} (blue) for the M_W 9.2 Sumatra subduction zone event and a portfolio in *DKI Jakarta* with 1 building. While in this case the building location and height account for a negligible fraction of event loss variance, the Ground Motion Prediction Equation (GMPE), the building quality and the damage distribution all have a sizable effect, in particular in interaction with other factors (S_{T_i}). The vertical extent of the boxes corresponds to bootstrapped 95% confidence intervals.

6.4 Conclusions

The results shown in this study indicate that LHS as well as QMC have the potential to increase the computational efficiency of seismic risk analysis. We observe that error convergence is improved from $O(n^{-0.5})$ to $O(n^{-1})$ for many loss scenarios. While $O(n^{-1})$ convergence is not fully achieved for very large portfolios, it still remains advantageous to use these sampling strategies over simple MC. For our model the adaptive MISER algorithm in its original form as well as in conjunction with QMC does not provide tangible benefits over LHS and QMC.

We have furthermore investigated the impact of uncertainties in the ground motion model as well as in the exposure and vulnerability models. Like many other uncertainty types in the exposure model, uncertainties in building location and building properties are so far often neglected. This study has shown that — depending on the loss scenario — a large fraction of the output variance can be attributed to these factors. Although

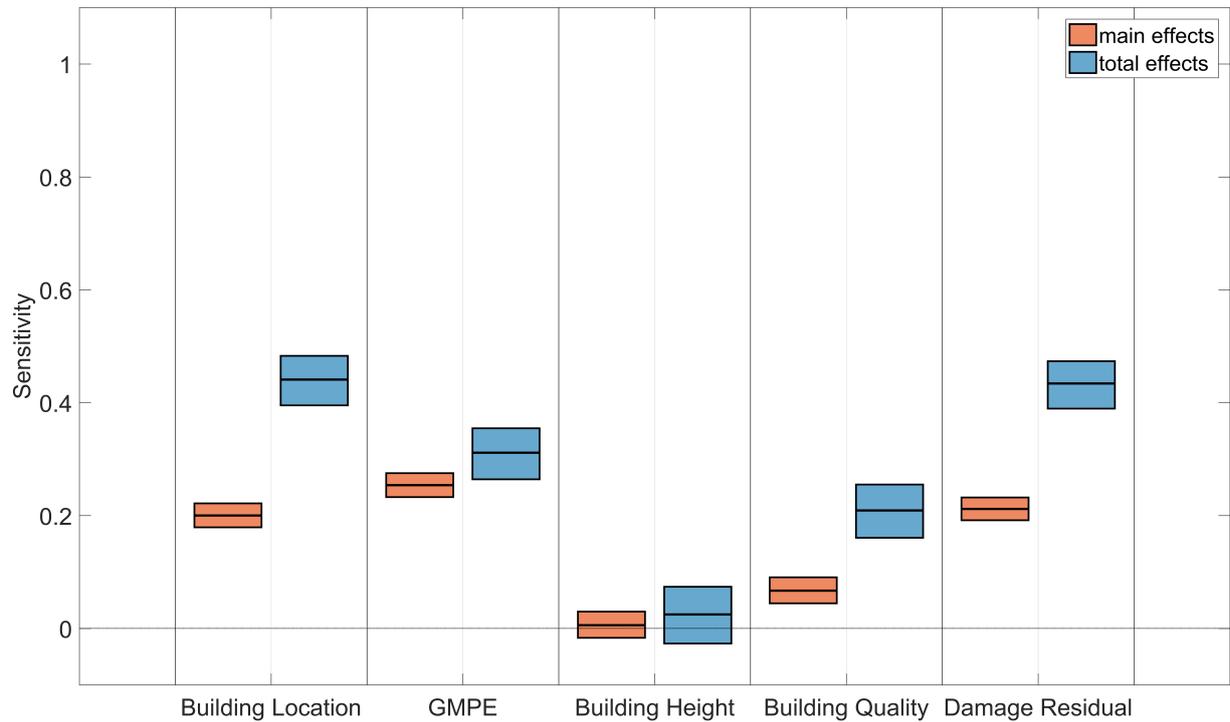


Figure 6.9: Main effects S_i (orange) and total effects S_{T_i} (blue) for the M_w 7.8 Sumatra fault zone event and a portfolio in Medan consisting of 100 buildings. While building heights accommodate a small fraction of the output variance, building locations and qualities, the Ground Motion Prediction Equation (GMPE) as well as the damage distributions have substantial influence. The vertical extent of the boxes corresponds to bootstrapped 95% confidence intervals.

due care must be exercised when transferring the results to other models, they highlight the importance of investigating the uncertainty associated with different factors. Decision makers may then incorporate this knowledge into e.g. regulation, disaster management and response plans, as well as risk mitigation measures and insurance pricing policies.

This work represents a progressive step towards a more comprehensive understanding of uncertainty in seismic risk analysis. Nevertheless, the integration of more factors remains an important task. The results of this study could also be tested using other methods than a variance-based SA, such as the elementary effects test (Morris, 1991) or density based methods (e.g. Pianosi and Wagener, 2015). Another powerful alternative is derivative based global SA (Sobol' and Kucherenko, 2009). This approach is particularly efficient in combination with algorithmic differentiation, which has already been successfully performed for PSHA (Molkenthin et al., 2017).

Acknowledgements

Financial support for this work was provided by Munich Re through a scholarship granted to the first author. We thank two anonymous reviewers, whose constructive feedback helped to improve the original article.

Chapter 7

Conclusion and Outlook

In this work we motivated, developed and analyzed a computationally efficient framework for the treatment of several important uncertainties in probabilistic seismic hazard assessment (PSHA) and probabilistic seismic risk assessment (PSRA). We analyzed the efficiency of different sampling schemes and the impact of various uncertainties on estimated seismic losses using a newly created seismic risk model of Indonesia in combination with synthetic portfolios modeled after real-world counterparts.

In Chapter 2, we reviewed the historical and theoretical background of PSHA and PSRA. We gave an overview of the current state of the art and identified several areas where the treatment and quantification of uncertainty could possibly be improved.

In Chapter 3, we described a framework for the stochastic simulation of ground motion and damage uncertainty, which has been developed during the course of this thesis and not only represents an important building block for all subsequent chapters, but also already facilitates other studies performed by internal and external scientists at Munich Reinsurance Group. We discussed the underlying theory and the most important implementation aspects of the framework, including the generation of uniform pseudo-random numbers, the transformation to non-uniform probability distributions using different algorithms, as well as the theory of Monte Carlo (MC) simulation including convergence and error estimation.

Special attention was also given to the treatment of correlation, which can have a strong effect on the results of PSRA. In the area of ground motion uncertainty, a range of different spatial correlation models as well as period-to-period cross-correlation models were implemented. Extending the framework with other models is straightforward, since the basic methodology for stochastic simulation of spatially cross-correlated random fields is independent of the correlation model. In the area of damage uncertainty, a multiple factor correlation model implemented during a Master thesis at Munich Re was integrated into the framework. The model consists of a global correlation residual sampled per event and a local residual sampled per risk item.

Furthermore, we introduced a new framework to obtain exact reproducibility of stochas-

tic simulation results on distributed systems, which is an important requirement for PSRA in the insurance industry due to stability requirements of model results used for insurance pricing. We described the usage of different hash functions to efficiently select substreams of a suitable pseudo-random number generator (e.g. MRG32K3a) to perform repeatable random number generation. Using a relatively simple and fast hash function such as the SDBM hash seems to be sufficient for the repeatable stochastic simulation of ground motion and damage uncertainty. However, for the repeatable simulation of portfolio location uncertainty, we found that a more sophisticated hash function such as MD5 is a better choice — at least for our framework, in which location uncertainty needs to be sampled beforehand for the entire event set and not per individual event, because of the underlying cluster architecture. Using the same substream repeatedly — due to collisions of the hash function — would carry more weight and should therefore be avoided.

In Chapter 4, we turned our attention to the previously little researched issue that for real portfolios in the insurance industry, exposure locations are often not known on coordinate-level, but instead only on the basis of administrative geographical zones. We proposed a novel framework for the stochastic treatment of portfolio location uncertainty, and investigated its impact on PSRA using a seismic hazard and risk model of western Indonesia. We analyzed the effect of portfolio characteristics such as value distribution, portfolio size, or proportion of risk items with unknown coordinates on the variability of loss frequency curves.

We first obtained an upper bound of the isolated impact of portfolio location uncertainty without aleatory ground motion variability for a flat hazard model and for the western Indonesia hazard model. We found that in both cases, depending on the portfolio properties, the worst-case impact can be quite high, which could be explained by risk item clustering as well as spatial hazard variability.

We proceeded to analyze the typical influence of location uncertainty in a realistic study of seismic risk in western Indonesia. We found that even in conjunction with aleatory ground motion uncertainty, location uncertainty can cause noticeable additional variability in loss frequency curves, in particular for portfolios with a small number of risks or when the insured value is distributed roughly exponentially among risk items.

The results led to the conclusion that location uncertainty in PSRA should not be neglected but treated stochastically, and that the resulting variability of probabilistic loss results should be visualized and interpreted carefully by decision makers. More precise location information might then be requested if the variability due to location uncertainty turns out to be prohibitively large for a particular portfolio.

In Chapter 5, we analyzed the spatial variation of seismic hazard and loss rate. For the example of western Indonesia, we found that the variation of hazard can vary strongly between different administrative geographical zones as well as between return periods. We found that the spatial variation of the loss rate was similar to that of the hazard without depending on the return period. This allowed to quantify spatial variation while avoiding the subjective selection of a suitable return period.

Building on these results, we then proposed a novel sampling scheme to increase the efficiency of portfolio location uncertainty treatment for the estimation of loss frequency curves. The scheme dynamically adapts the location sample size individually for insured risk items depending on (1) the loss rate variation within an administrative geographical zone, (2) the number of risks within the zone, and (3) the value of the risk.

We analyzed the performance of the scheme empirically. While the scheme did not improve the error convergence order, it reached the same error levels as simple MC with less samples of potential risk locations, resulting in lower memory requirements and a moderate runtime speedup. The scheme could also be applied to other natural perils, such as probabilistic wind and flood models.

To improve the efficiency of the MC simulation in the area of ground motion and damage uncertainty, in Chapter 6 we investigated the applicability of some well-known variance reduction schemes for the computation of seismic losses using three scenarios in Indonesia. Latin Hypercube Sampling (LHS) as well as Quasi MC (QMC) simulation using low-discrepancy sequences could improve the error convergence order from $O(n^{-0.5})$ to $O(n^{-1})$ in many cases. The adaptive MISER algorithm, based on recursive stratified sampling, did not provide any convincing advantage over LHS and QMC for our purposes — while the convergence rate could be improved to $O(n^{-1})$ for very small portfolios, it remained at $O(n^{-0.5})$ for portfolios consisting of 5 risk items and more. A variant of MISER combined with QMC slightly improved the performance of the original MISER routine, but likewise did not achieve a better convergence rate.

Using variance-based global sensitivity analysis, we then quantified the relative impact of different uncertainties on seismic loss for the same three scenarios in Indonesia. Depending on the portfolio, we found that a significant fraction of the output variance could be attributed to uncertainties in the exposure and vulnerability models, highlighting the importance of their rigorous treatment in seismic risk analysis.

This work made numerous contributions to a more comprehensive and efficient treatment of uncertainty and thereby represents an important step towards a more holistic view on seismic hazard and risk. Nevertheless, we are certainly far from having integrated all relevant uncertainties and this work hopefully only started the quest to develop and apply more computationally efficient methodologies to their treatment.

Future work could focus on a wide range of possible issues. An important task will be the treatment of further uncertainties, in particular in the exposure model. For example, it is known that the insured values of risk items are often biased or estimated inaccurately, which likely has a strong effect on loss frequency curves. The efficiency of the loss computation can possibly be further optimized, for example with the aid of machine learning techniques to create an adaptive sampling scheme or a surrogate model.

Bibliography

- Anagnos, T. and Kiremidjian, A. S.** (1988). A review of earthquake occurrence models for seismic hazard analysis. *Probabilistic Engineering Mechanics*, **3**(1):3–11.
Referenced on page 15.
- Anderson, J. G. and Brune, J. N.** (1999). Probabilistic seismic hazard analysis without the ergodic assumption. *Seismological Research Letters*, **70**(1):19–28. doi:10.1785/gssrl.70.1.19.
Referenced on pages 9 and 16.
- Atik, L. A., Abrahamson, N., Bommer, J. J., Scherbaum, F., Cotton, F., and Kuehn, N.** (2010). The variability of ground-motion prediction models and its components. *Seismological Research Letters*, **81**(5):659–801. doi:10.1785/gssrl.81.5.794.
Referenced on page 48.
- Atkinson, G. M. and Boore, D. M.** (2003). Empirical ground-motion relations for subduction-zone earthquakes and their application to Cascadia and other regions. *Bulletin of the Seismological Society of America*, **93**(4):1703–1729. doi:10.1785/0120080108.
Referenced on pages 48 and 56.
- Baker, J. W. and Cornell, C. A.** (2006). Correlation of response spectral values for multicomponent ground motions. *Bulletin of the Seismological Society of America*, **96**(1):215–227. doi:10.1785/0120050060.
Referenced on pages 8, 34, 39, 41, and 42.
- Baker, J. W. and Jayaram, N.** (2008). Correlation of spectral acceleration values from NGA ground motion models. *Earthquake Spectra*, **24**(1):299–317. doi:10.1193/1.2857544.
Referenced on page 37.
- Bal, I. E., Bommer, J. J., Stafford, P. J., Crowley, H., and Pinho, R.** (2010). The influence of geographical resolution of urban exposure data in an earthquake loss model for Istanbul. *Earthquake Spectra*, **26**(3):619–634. doi:10.1193/1.3459127.
Referenced on pages 49, 70, 71, and 74.
- Barbat, A. H., Moya, F. Y., and Canas, J.** (1996). Damage scenarios simulation for seismic risk assessment in urban zones. *Earthquake Spectra*, **12**(3):371–394. doi:

10.1193/1.1585889.

Referenced on page 18.

Bazzurro, P. and Cornell, C. A. (2002). Vector-valued probabilistic seismic hazard analysis (VPSHA). *Proceedings of the 7th US National Conference on Earthquake Engineering*, pages 21–25.

Referenced on pages 34 and 39.

Beasley, J. and Springer, S. (1977). Algorithm AS 111: The percentage points of the normal distribution. *Applied Statistics*, **26**:118–121.

Referenced on pages 31 and 35.

Benedetti, D. and Petrini, V. (1984). On seismic vulnerability of masonry buildings: proposal of an evaluation procedure. *L'industria delle Costruzioni*, **18**(2):66–74.

Referenced on page 17.

Bier, V. M. and Lin, S. W. (2013). On the treatment of uncertainty and variability in making decisions about risk. *Risk Analysis*, **33**(10):1899–1907. doi:10.1111/risa.12071.

Referenced on page 71.

Bindi, D., Cotton, F., Kotha, S. R., Bosse, C., Stromeyer, D., and Grünthal, G. (2017). Application-driven ground motion prediction equation for seismic hazard assessments in non-cratonic moderate-seismicity areas. *Journal of Seismology*, **21**(5):1201–1218. doi:10.1007/s10950-017-9661-5.

Referenced on page 48.

Bommer, J. J. and Abrahamson, N. A. (2006). Why do modern probabilistic seismic-hazard analyses often lead to increased hazard estimates? *Bulletin of the Seismological Society of America*, **96**(6):1967–1977. doi:10.1785/0120060043.

Referenced on page 71.

Bommer, J. J., Douglas, J., and Strasser, F. O. (2003). Style-of-faulting in ground-motion prediction equations. *Bulletin of Earthquake Engineering*, **1**(2):171–203. doi:10.1023/A:1026323123154.

Referenced on page 48.

Boore, D. M. and Atkinson, G. M. (2008). Ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods between 0.01 s and 10.0 s. *Earthquake Spectra*, **24**(1):99–138. doi:10.1193/1.2830434.

Referenced on pages 56 and 108.

Botev, Z. I. (2017). The normal law under linear restrictions: Simulation and estimation via minimax tilting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **79**(1):125–148. doi:10.1111/rssb.12162.

Referenced on pages 31 and 35.

- Box, G. E. P. and Muller, M. E.** (1958). A note on the generation of random normal deviates. *The Annals of Mathematical Statistics*, **29**(2):610–611. doi:10.1214/aoms/1177706645.
Referenced on page 31.
- Calvi, G. M., Pinho, R., Magenes, G., Bommer, J. J., Restrepo-Velez, L., and Crowley, H.** (2006). Development of seismic vulnerability assessment methodologies over the past 30 years. *ISET Journal of Earthquake Technology*, **43**(3):75–104.
Referenced on pages 2, 17, and 18.
- Campbell, K. W. and Bozorgnia, Y.** (2008). NGA ground motion model for the geometric mean horizontal component of PGA, PGV, PGD and 5% damped linear elastic response spectra for periods ranging from 0.01 to 10 s. *Earthquake Spectra*, **24**(1):139–171. doi:10.1193/1.2857546.
Referenced on pages 56 and 108.
- Cao, T., Petersen, M. D., Cramer, C. H., Topozada, T. R., Reichle, M. S., and Davis, J. F.** (1999). The calculation of expected loss using probabilistic seismic hazard. *Bulletin of the Seismological Society of America*, **89**(4):867–876.
Referenced on page 11.
- Chen, Z., Chui, Y. H., Ni, C., and Xu, J.** (2013). Seismic response of midrise light wood-frame buildings with portal frames. *Journal of Structural Engineering*, **140**(8):A4013003. doi:10.1061/(ASCE)ST.1943-541X.0000882.
Referenced on page 25.
- Chiou, B. S.-J. and Youngs, R. R.** (2008). An NGA model for the average horizontal component of peak ground motion and response spectra. *Earthquake Spectra*, **24**(1):173–215. doi:10.1193/1.2894832.
Referenced on pages 35, 36, 38, 39, 40, 41, 42, 48, 56, 107, and 108.
- Coburn, A. and Spence, R.** (2002). *Earthquake protection*. John Wiley, 2nd edition.
Referenced on page 2.
- Cornell, C. A.** (1964). *Stochastic processes in civil engineering*. Ph.D. thesis, Department of Civil Engineering, Stanford University.
Referenced on page 7.
- Cornell, C. A.** (1968). Engineering seismic risk analysis. *Bulletin of the Seismological Society of America*, **58**(5):1583–1606.
Referenced on pages 8, 12, 50, 75, and 98.
- Cox, L. A.** (2012). Confronting deep uncertainties in risk analysis. *Risk Analysis*, **32**(10):1607–1629. doi:10.1111/j.1539-6924.2012.01792.x.
Referenced on page 49.

- Crowley, H.** (2014). Earthquake risk assessment: Present shortcomings and future directions. *Geotechnical, Geological and Earthquake Engineering*, volume 34, pages 515–532. doi:10.1007/978-3-319-07118-3_16.
Referenced on pages 19, 49, 51, 77, and 98.
- Crowley, H., Bommer, J. J., Pinho, R., and Bird, J.** (2005). The impact of epistemic uncertainty on an earthquake loss model. *Earthquake Engineering and Structural Dynamics*, **34**(14):1653–1685. doi:10.1002/eqe.498.
Referenced on page 48.
- Crowley, H., D. Monelli, M. Pagani, V. Silva, and G. Weatherill** (2011). *The Openquake book*. The GEM Foundation, Pavia, Italy.
Referenced on pages 11, 13, 14, and 15.
- Crowley, H., Pinho, R., Pagani, M., and Keller, N.** (2013). Assessing global earthquake risks: the Global Earthquake Model (GEM) initiative. *Handbook of Seismic Risk Analysis and Management of Civil Infrastructure Systems*, pages 815–838. Woodhead Publishing Limited. doi:10.1533/9780857098986.5.815.
Referenced on pages 55 and 80.
- Crowley, H., Stafford, P. J., and Bommer, J. J.** (2008). Can earthquake loss models be validated using field observations? *Journal of Earthquake Engineering*, **12**(7). doi:10.1080/13632460802212923.
Referenced on page 71.
- Davis, M. W.** (1987). Production of conditional simulations via the LU triangular decomposition of the covariance matrix. *Mathematical Geology*, **19**(2):91–98. doi:10.1007/BF00898189.
Referenced on page 38.
- Deierlein, G. G., Reinhorn, A. M., and Willford, M. R.** (2010). Nonlinear structural analysis for seismic design. *NEHRP seismic design technical brief*, **4**(4):1–36.
Referenced on page 25.
- Delavaud, E., Cotton, F., Akkar, S., Scherbaum, F., Danciu, L., Beauval, C., Drouet, S., Douglas, J., Basili, R., Sandikkaya, M. A., Segou, M., Faccioli, E., and Theodoulidis, N.** (2012). Toward a ground-motion logic tree for probabilistic seismic hazard assessment in Europe. *Journal of Seismology*, **16**(3):451–473. doi:10.1007/s10950-012-9281-z.
Referenced on page 48.
- Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C., and Worley, B. A.** (2000). LandScan: a global population database for estimating populations at risk. *Photogrammetric engineering and remote sensing*, **66**(7):849–857.
Referenced on pages 51 and 77.

- Dolce, M., Masi, A., Marino, M., and Vona, M.** (2003). Earthquake damage scenarios of the building stock of Potenza (southern Italy) including site effects. *Bulletin of Earthquake Engineering*, **1**(1):115–140. doi:10.1023/A:1024809511362.
Referenced on page 17.
- dos Santos, K. and Beck, A.** (2015). A benchmark study on intelligent sampling techniques in Monte Carlo simulation. *Latin American Journal of Solids and Structures*, **12**(4):624–648. doi:10.1590/1679-78251245.
Referenced on page 74.
- Douglas, J.** (2003). Earthquake ground motion estimation using strong-motion records: a review of equations for the estimation of peak ground acceleration and response spectral ordinates. *Earth-Science Reviews*, **61**:43–104.
Referenced on page 15.
- Douglas, J.** (2018). Ground motion prediction equations 1964–2016, <http://www.gmpe.org.uk>, last accessed on May 16th, 2018. URL <http://www.gmpe.org.uk>.
Referenced on page 8.
- Eads, L., Miranda, E., Krawinkler, H., and Lignos, D. G.** (2013). An efficient method for estimating the collapse risk of structures in seismic regions. *Earthquake Engineering & Structural Dynamics*, **42**(1):25–41. doi:10.1002/eqe.2191.
Referenced on page 74.
- Efron, B.** (1979). Bootstrap methods: another look at the Jackknife. *The Annals of Statistics*, **7**(1):1–26. doi:10.1214/aos/1176344552.
Referenced on pages 27 and 78.
- Efron, B.** (1982). *The jackknife, the bootstrap and other resampling plans*. SIAM.
Referenced on page 27.
- Efron, B. and Gong, G.** (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *American Statistician*, **37**(1):36–48. doi:10.1080/00031305.1983.10483087.
Referenced on page 27.
- Efron, B. and Tibshirani, R.** (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, **1**(1):54–75.
Referenced on pages 78 and 103.
- Encyclopaedia Britannica** (2009). Indonesia: population density. Map. [Internet]. Encyclopaedia Britannica Online; c2009 [cited 2016 Jan 13]. Available from: <https://www.britannica.com/place/Indonesia/images-videos/Population-density-of-Indonesia/138387>.
Referenced on page 54.

- Esteva, L.** (1963). Regionalizacion sismica de la Republica Mexicana. *Revista Sociedad Mexicana do Ingenieria Sismica*, **1**(1):31–35.
Referenced on page 8.
- Esteva, L.** (1967). Criterios para la construccion de espectros para diseno sismico. *Boletin del Instituto de Materiales y Modelos Estructurales, Universidad Central de Venezuela*, **19**(9).
Referenced on page 8.
- Esteva, L.** (1968). *Bases para la formulacion de decisiones de diseno sismico*. Ph.D. thesis, Universidad Autonoma Nacional de Mexico.
Referenced on page 8.
- Esteva, L.** (1970). Seismic risk and seismic design decisions. *Seismic Design for Nuclear Power Plants*, **1**(1):142–182.
Referenced on page 12.
- Esteva, L. and Rosenblueth, E.** (1964). Espectros de temblores a distancias moderadas y grandes. *Boletin Sociedad Mexicana de Ingenieria Sismica*.
Referenced on page 8.
- European Commission** (2009). Directive 2009/138/EC of the European Parliament and of the council of 25 November 2009 on the taking-up and pursuit of the business of insurance and reinsurance (Solvency II).
Referenced on page 4.
- Federal Emergency Management Agency (FEMA)** (2001). HAZUS technical manual. **18**(1986):85–105.
Referenced on pages 18 and 22.
- Field, E. H., Jordan, T. H., and Cornell, C. A.** (2003). OpenSHA: a developing community-modeling environment for seismic hazard analysis. *Seismological Research Letters*, **74**(4):406–419. doi:10.1785/gssrl.74.4.406.
Referenced on pages 10 and 12.
- Field, E. H. and SCEC Phase III Working Group** (2000). Accounting for site effects in probabilistic seismic hazard analysis: overview of the SCEC phase III report. *Bulletin of the Seismological Society of America*, **90**(6B):1–31. doi:10.1785/0120000512.
Referenced on pages 9, 17, and 48.
- Fitzenz, D. D. and Nyst, M.** (2015). Building time-dependent earthquake recurrence models for probabilistic risk computations. *Bulletin of the Seismological Society of America*, **105**(1):120–133. doi:10.1785/0120140055.
Referenced on page 15.

- Foulser-Piggott, R., Bowman, G., and Hughes, M.** (2017). A framework for understanding uncertainty in seismic risk assessment. *Risk Analysis*. doi:10.1111/risa.12919.
Referenced on page 48.
- Garakaninezhad, A., Bastami, M., and Soghrat, M. R.** (2017). Spatial correlation for horizontal and vertical components of acceleration from northern Iran seismic events. *Journal of Seismology*, **21**(6):1505–1516. doi:10.1007/s10950-017-9679-8.
Referenced on page 37.
- Gaughan, A. E., Stevens, F. R., Linard, C., Jia, P., and Tatem, A. J.** (2015). High resolution population distribution maps for Southeast Asia in 2010 and 2015. *PLoS ONE*, **8**(2):e55882. doi:10.1371/journal.pone.0055882.
Referenced on pages 52, 76, and 100.
- Glasserman, P.** (2003). *Monte Carlo methods in financial engineering*, volume 53. Springer Science & Business Media. doi:10.1007/978-0-387-21617-1.
Referenced on pages 31 and 35.
- Goda, K. and Atkinson, G. M.** (2009). Probabilistic characterization of spatially correlated response spectra for earthquakes in Japan. *Bulletin of the Seismological Society of America*, **99**(5):3003–3020. doi:10.1785/0120090007.
Referenced on page 37.
- Goda, K. and Atkinson, G. M.** (2010). Intraevent spatial correlation of ground-motion parameters using SK-net data. *Bulletin of the Seismological Society of America*, **100**(6):3055–3067. doi:10.1785/0120100031.
Referenced on page 8.
- Goulard, M. and Voltz, M.** (1992). Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. *Mathematical Geology*, **24**(3):269–286. doi:10.1007/BF00893750.
Referenced on page 39.
- Graves, R., Jordan, T. H., Callaghan, S., Deelman, E., Field, E., Juve, G., Kesselman, C., Maechling, P., Mehta, G., Milner, K., Okaya, D., Small, P., and Vahi, K.** (2011). CyberShake: a physics-based seismic hazard model for Southern California. *Pure and Applied Geophysics*, **168**(3-4):367–381. doi:10.1007/s00024-010-0161-6.
Referenced on pages 17 and 25.
- Grünthal, G., Stromeyer, D., and Heidbach, O.** (2013). Probabilistische seismische Gefährdungsabschätzungen. *System Erde*, **3**(1):54–61.
Referenced on page 16.
- Güllü, H.** (2012). Prediction of peak ground acceleration by genetic expression programming and regression: a comparison using likelihood-based measure. *Engineering Geology*,

141-142:92–113. doi:10.1016/j.enggeo.2012.05.010.

Referenced on page 15.

Gumbel, E. (1935). Les valeurs extrêmes des distributions statistiques. *Annales de l'institut Henri Poincaré*, **2**:115–158.

Referenced on page 10.

Gumbel, E. (1958). Statistics of extremes. *Columbia Univ. press, New York*.

Referenced on page 10.

Gutenberg, B. and Richter, C. F. (1942). Earthquake magnitude, intensity, energy and acceleration. *Bulletin of the Seismological Society of America*, **32**(3):163–191. doi:10.1785/0120160029.

Referenced on pages 8 and 10.

Gutenberg, B. and Richter, C. F. (1944). Frequency of earthquakes in California. *Bulletin of the Seismological Society of America*, **34**(4):185.

Referenced on pages 8 and 10.

Harding, B., Tremblay, C., and Cousineau, D. (2014). Standard errors: A review and evaluation of standard error estimators using Monte Carlo simulations. *The Quantitative Methods for Psychology*, **10**(2):107–123. doi:10.20982/tqmp.10.2.p107.

Referenced on pages 27 and 77.

Hayes, G. P., Wald, D. J., and Johnson, R. L. (2012). Slab1.0: A three-dimensional model of global subduction zone geometries. *Journal of Geophysical Research*, **117**(1):1–15. doi:10.1029/2011JB008524.

Referenced on pages 54, 80, and 105.

Hess, S., Train, K. E., and Polak, J. W. (2006). On the use of a Modified Latin Hypercube Sampling (MLHS) method in the estimation of a mixed logit model for vehicle choice. *Transportation Research Part B: Methodological*, **40**(2):147–163. doi:10.1016/j.trb.2004.10.005.

Referenced on page 74.

Higham, N. J. (1988). Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra and Its Applications*, **103**(C):103–118. doi:10.1016/0024-3795(88)90223-6.

Referenced on page 42.

Hofemeier, G. (2012). Intel Digital Random Number Generator (DRNG) software implementation guide. *Technical report*.

Referenced on page 28.

Homma, T. and Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, **52**(1):1–17.

Referenced on page 104.

- Hörmann, W., Leydold, J., and Derflinger, G.** (2013). *Automatic nonuniform random variate generation*. Springer Science & Business Media.
Referenced on page 30.
- Infantino, M.** (2016). *From 3D physics-based scenarios to advanced methods for seismic hazard assessment: the case of Istanbul*. Master Thesis, Politecnico di Milano.
Referenced on pages 17 and 25.
- Jadach, S.** (2003). Foam: A general-purpose cellular Monte Carlo event generator. *Computer Physics Communications*, **152**(1):55–100. doi:10.1016/S0010-4655(02)00755-5.
Referenced on page 85.
- Jansen, M. J. W.** (1999). Analysis of variance designs for model output. *Computer Physics Communications*, **117**(1):35–43. doi:10.1016/S0010-4655(98)00154-4.
Referenced on page 104.
- Jayaram, N. and Baker, J. W.** (2008). Statistical tests of the joint distribution of spectral acceleration values. *Bulletin of the Seismological Society of America*, **98**(5):2231–2243. doi:10.1785/0120070208.
Referenced on pages 16, 36, and 67.
- Jayaram, N. and Baker, J. W.** (2009). Correlation model for spatially distributed ground-motion intensities. *Earthquake Engineering & Structural Dynamics*, **41**(11):1549–1568. doi:10.1002/eqe.922.
Referenced on pages 8, 15, 34, 37, 38, 39, 41, 42, and 108.
- Jayaram, N. and Baker, J. W.** (2010). Efficient sampling and data reduction techniques for probabilistic seismic lifeline risk assessment. *Earthquake Engineering and Structural Dynamics*, **39**(10):1109–1131. doi:10.1002/eqe.988.
Referenced on pages 36 and 74.
- Joe, S. and Kuo, F. Y.** (2003). Remark on algorithm 659. *ACM Transactions on Mathematical Software*, **29**(1):49–57. doi:10.1145/641876.641879.
Referenced on page 109.
- Johnson, M. E., Moore, L. M., and Ylvisaker, D.** (1990). Minimax and maximin distance designs. *Journal of Statistical Planning and Inference*, **26**(2):131–148.
Referenced on page 101.
- Juneja, S. and Kalra, H.** (2009). Variance reduction techniques for pricing American options using function approximations. *Journal of Computational Finance*, **12**(3):79–101. doi:10.1.1.509.8171.
Referenced on page 78.
- Kafka, A. L., Ebel, J. E., and Kafka, A. L.** (1999). A Monte Carlo approach to seismic hazard analysis. *Bulletin of the Seismological Society of America*, **89**(4):854–866.
Referenced on page 15.

- Kappos, A., Pitilakis, K., Stylianidis, K., Morfidis, K., and Asimakopoulos, D.** (1995). Cost-benefit analysis for the seismic rehabilitation of buildings in Thessaloniki, based on a hybrid method of vulnerability assessment. *Proceedings of the 5th International Conference on Seismic Zonation*, volume 1, pages 406–413.
Referenced on page 18.
- Kircher, C. A., Nassar, A. A., Kustu, O., and Holmes, W. T.** (1997). Development of building damage functions for earthquake loss estimation. *Earthquake Spectra*, **13**(4):663–682. doi:10.1193/1.1585974.
Referenced on pages 18 and 22.
- Kiureghian, A. D. and Ditlevsen, O.** (2009). Aleatory or epistemic? Does it matter? *Structural Safety*, **31**(2):105–112. doi:10.1016/j.strusafe.2008.06.020.
Referenced on pages 8 and 48.
- Knuth, D. E.** (1976). Big Omicron and big Omega and big Theta. *ACM SIGACT News*, **8**(2):18–24. doi:10.1145/1008328.1008329.
Referenced on page 78.
- Knuth, D. E.** (1981). *The Art of Computer Programming; Volume 2: Seminumerical Algorithms*.
Referenced on page 29.
- Knuth, D. E.** (1997). *The Art of Computer Programming; Volume 3: Sorting and Searching*. Pearson Education.
Referenced on page 33.
- Koehler, E., Brown, E., and Haneuse, S. J.-P. A.** (2009). On the assessment of Monte Carlo error in simulation-based statistical analyses. *The American Statistician*, **63**(2):155–162. doi:10.1198/tast.2009.0030.
Referenced on page 27.
- Kotha, S. R., Bazzurro, P., and Pagani, M.** (2018). Effects of epistemic uncertainty in seismic hazard estimates on building portfolio losses. *Earthquake Spectra*, **34**(1):217–236. doi:10.1193/020515EQS020M.
Referenced on page 55.
- Krinitzsky, E. L.** (2002). How to obtain earthquake ground motions for engineering design. *Engineering Geology*, **65**:1–16.
Referenced on page 15.
- Kumaraswamy, P.** (1980). A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, **46**(1-2):79–88. doi:10.1016/0022-1694(80)90036-0.
Referenced on page 43.

- Landau, E.** (1909). *Handbuch der Lehre von der Verteilung der Primzahlen*, volume 1. B. G. Teubner, Leipzig.
Referenced on page 78.
- Langford, E.** (2006). Quartiles in elementary statistics. *Journal of Statistics Education*, **14**(3). doi:10.1080/10691898.2006.11910589.
Referenced on page 53.
- L'Ecuyer, P.** (1988). Efficient and portable combined random number generators. *Communications of the ACM*, **31**(6):742–751. doi:10.1145/62959.62969.
Referenced on page 30.
- L'Ecuyer, P.** (1994). Uniform random number generation. *Annals of Operations Research*, **53**(1):77–120. doi:10.1007/BF02136827.
Referenced on page 29.
- L'Ecuyer, P.** (1996). Combined multiple recursive random number generators. *Operations Research*, **44**(5):816–822. doi:10.1287/opre.44.5.816.
Referenced on page 30.
- L'Ecuyer, P.** (1999). Good Parameters and Implementations for Combined Multiple Recursive Random Number Generators. *Operations Research*, **47**(1):159–164. doi:10.1287/opre.47.1.159.
Referenced on pages 30 and 77.
- L'Ecuyer, P.** (2014). Random number generation and Quasi-Monte Carlo. *Wiley StatsRef: Statistics Reference Online*, pages 1–12. doi:10.1002/9781118445112.stat04386.pub2.
Referenced on page 28.
- L'Ecuyer, P. and Simard, R.** (2007). TestU01. *ACM Transactions on Mathematical Software*, **33**(4):22–es. doi:10.1145/1268776.1268777.
Referenced on page 29.
- Loth, C. and Baker, J. W.** (2012). A spatial cross-correlation model of spectral accelerations at multiple periods. *Earthquake Engineering & Structural Dynamics*, **41**(11):1549–1568. doi:10.1002/eqe.2212.
Referenced on pages 9, 34, 39, and 40.
- MacKay, D. J. C.** (2005). *Information Theory, Inference, and Learning Algorithms*, volume 100. Cambridge university press. doi:10.1198/jasa.2005.s54.
Referenced on page 78.
- Mark Petersen, Stephen Harmsen, Charles Mueller, Kathleen Haller, James Dewey, Nicolas Luco, Anthony Crone, David Lidke, and Kenneth Rukstales** (2007). Documentation for the Southeast Asia seismic hazard maps. *USGS Administrative Report*.
Referenced on pages 54, 56, 80, 146, 148, and 150.

- Marsaglia, G.** (2003). Xorshift RNGs. *Journal Of Statistical Software*, **8**(14):1–6. doi:10.18637/jss.v008.i14.
Referenced on page 29.
- Marsaglia, G. and Tsang, W. W.** (1984). A fast, easily implemented method for sampling from decreasing or symmetric unimodal density functions. *SIAM Journal on scientific and statistical computing*, **5**(2):349–359.
Referenced on page 31.
- Marsaglia, G. and Tsang, W. W.** (2000). The Ziggurat method for generating random variables. *Journal of Statistical Software*, **5**(8):1–7. doi:10.1145/355744.355749.
Referenced on page 31.
- Matheron, G.** (1965). Les variables régionalisées et leur estimation: une application de la théorie des fonctions aléatoires aux sciences de la nature. *Masson et CIE*.
Referenced on page 37.
- Matheron, G.** (1982). Pour une analyse krigéante des données régionalisées. *Centre de Géostatistique, Report N-732, Fontainebleau*.
Referenced on page 39.
- Matoušek, J.** (1998). On the L2-discrepancy for anchored boxes. *Journal of Complexity*, **14**(4):527–556.
Referenced on page 102.
- Matsumoto, M. and Nishimura, T.** (1998). Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, **8**(1):3–30. doi:10.1145/272991.272995.
Referenced on pages 29, 51, and 101.
- Matthews, M. V., Ellsworth, W. L., and Reasenber, P. a.** (2002). A Brownian model for recurrent earthquakes. *Bulletin of the Seismological Society of America*, **92**(6):2233–2250. doi:10.1785/0120010267.
Referenced on page 15.
- Mazzoni, S., McKenna, F., and Fenves, G. L.** (2005). OpenSees command language manual. *Pacific Earthquake Engineering Research (PEER) Center*, **264**.
Referenced on page 25.
- McGuire, R. K.** (1976). USGS open-file report 76-67: Fortran computer program for seismic risk analysis.
Referenced on pages 8 and 16.
- McGuire, R. K.** (2004). *Seismic hazard and risk analysis*. Earthquake Engineering Research Institute, 1st edition.
Referenced on pages 7, 50, and 75.

- McGuire, R. K.** (2008). Probabilistic seismic hazard analysis: early history. *Earthquake Engineering & Structural Dynamics*, **37**(3):329–338. doi:10.1002/eqe.
Referenced on page 7.
- McKay, M. D., Beckman, R. J., and Conover, W. J.** (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, **21**(2):239–245. doi:10.1080/00401706.1979.10489755.
Referenced on page 101.
- McKenna, F.** (2011). OpenSees: a framework for earthquake engineering simulation. *Computing in Science & Engineering*, **13**(4):58–66. doi:10.1109/MCSE.2011.66.
Referenced on page 25.
- Mert, A., Fahjan, Y. M., Hutchings, L. J., and Pinar, A.** (2016). Physically based probabilistic seismic hazard analysis using broadband ground motion simulation: a case study for the Prince Islands Fault, Marmara Sea. *Earth, Planets and Space*, **68**(1):146. doi:10.1186/s40623-016-0520-3.
Referenced on page 25.
- Merz, H. A. and Cornell, C. A.** (1973). Seismic Risk Analysis based on a quadratic magnitude-frequency law. *Bulletin of the Seismological Society of America*, **63**(6):1999–2006.
Referenced on page 10.
- Metropolis, N. and Ulam, S.** (1949). The Monte Carlo Method. *Journal of the American Statistical Association*, **44**(247):335–341. doi:10.1080/01621459.1949.10483310.
Referenced on pages 26 and 101.
- Mitnik, P. a. and Baek, S.** (2013). The Kumaraswamy distribution: median-dispersion re-parameterizations for regression modeling and simulation-based estimation. *Statistical Papers*, **54**(1):177–192. doi:10.1007/s00362-011-0417-y.
Referenced on page 43.
- Molkenthin, C., Scherbaum, F., Griewank, A., Leovey, H., Kucherenko, S., and Cotton, F.** (2017). Derivative-based global sensitivity analysis: Upper bounding of sensitivities in seismic-hazard assessment using automatic differentiation. *Bulletin of the Seismological Society of America*, **107**(2):984–1004. doi:10.1785/0120160185.
Referenced on page 114.
- Moro, B.** (1995). The Full Monte. *Risk Magazine*, **8**(2):57–58.
Referenced on pages 31 and 35.
- Morokoff, W. J. and Caflich, R. E.** (1994). Quasi-random sequences and their discrepancies. *SIAM Journal on Scientific Computing*, **15**(6):1251–1279. doi:10.1137/0915077.
Referenced on page 102.

- Morris, M. D.** (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, **33**(2):161–174.
Referenced on page 114.
- Moskowitz, B. and Caflisch, R.** (1996). Smoothness and dimension reduction in Quasi-Monte Carlo methods. *Mathematical and Computer Modelling*, **23**(8-9):37–54. doi:10.1016/0895-7177(96)00038-6.
Referenced on page 102.
- Munich Re** (2015). Knowledge series: Reinsurance in a changing world.
Referenced on page 4.
- Munich Re** (2018). Group Annual Report 2017.
Referenced on page 4.
- Munich Re NatCatSERVICE** (2018a). Annual Statistics 2017. <https://www.munichre.com/touch/naturalthazards/en/natcatservice/annual-statistics/index.html>, last accessed on May 23rd, 2018.
Referenced on pages 2 and 3.
- Munich Re NatCatSERVICE** (2018b). Natural Loss Events Worldwide 1980 - 2017, <https://www.munichre.com/touch/naturalthazards/en/natcatservice/focus-analyses/index.html>, last accessed on May 23rd, 2018.
Referenced on page 1.
- Munich Re NatCatSERVICE** (2018c). Significant Natural Catastrophes. <https://www.munichre.com/touch/naturalthazards/en/natcatservice/significant-natural-catastrophes/index.html>, last accessed on May 23rd, 2018.
Referenced on page 2.
- Musson, R. M. W.** (2000). The use of Monte Carlo simulations for seismic hazard assessment in the U.K. *Annals of Geophysics*, **43**(1):1–9. doi:10.4401/ag-3617.
Referenced on page 8.
- Ni, C., Pei, S., Lindt, J. W. V. D., Kuan, S., van de Lindt, J. W., Kuan, S., and Popovski, M.** (2012). Nonlinear time-history analysis of a six-story wood platform frame buildings in Vancouver, British Columbia. *Earthquake Spectra*, **28**(2):621–637. doi:10.1193/1.4000015.
Referenced on page 25.
- Oliver, D. S.** (2003). Gaussian cosimulation: Modelling of the cross-covariance. *Mathematical Geology*, **35**(6):681–698. doi:10.1023/B:MATG.0000002984.56637.ef.
Referenced on page 41.
- Oliver, M. and Webster, R.** (2014). A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *Catena*, **113**:56–69. doi:10.1016/j.catena.2013.09.006.
Referenced on page 36.

- O’Neill, M. E.** (2014). PCG: A family of simple fast space-efficient statistically good algorithms for random number generation. *Technical Report HMC-CS-2014-0905*, Harvey Mudd College, Claremont, CA.
Referenced on page 30.
- Orsini, G.** (1999). A model for buildings’ vulnerability assessment using the parameterless scale of seismic intensity (PSI). *Earthquake Spectra*, **15**(3):463–483. doi:10.1193/1.1586053.
Referenced on page 18.
- Ospina, R. and Ferrari, S. L.** (2012). A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis*, **56**(6):1609–1623. doi:10.1016/j.csda.2011.10.005.
Referenced on page 42.
- Ospina, R. and Ferrari, S. L. P.** (2010). Inflated beta distributions. *Statistical Papers*, **51**(1):111. doi:10.1007/s00362-008-0125-4.
Referenced on page 42.
- Owen, A. B.** (1994). Controlling correlations in Latin Hypercube samples. *Journal of the American Statistical Association*, **89**(428):1517–1522.
Referenced on page 101.
- Owen, A. B.** (1997). Monte Carlo variance of scrambled net quadrature. *SIAM Journal on Numerical Analysis*, **34**(5):1884–1910.
Referenced on pages 101 and 102.
- Owen, A. B.** (2013). Monte Carlo theory, methods and examples. *Monte Carlo theory, methods and examples*, pages 1–26.
Referenced on page 27.
- Pagani, M., Monelli, D., Weatherill, G., Danciu, L., Crowley, H., Silva, V., Henshaw, P., Butler, L., Nastasi, M., Panzeri, L., Simionato, M., and Vigano, D.** (2014). OpenQuake engine: An open hazard (and risk) software for the Global Earthquake Model. *Seismological Research Letters*, **85**(3):692–702. doi:10.1785/0220130087.
Referenced on pages 10, 12, 50, 54, 74, and 80.
- Panneton, F., L’Ecuyer, P., and Matsumoto, M.** (2006). Improved long-period generators based on linear recurrences modulo 2. *ACM Transactions on Mathematical Software*, **32**(1):1–16. doi:10.1145/1132973.1132974.
Referenced on pages 29 and 30.
- Paolucci, R., Infantino, M., Mazzieri, I., Özcebe, A. G., Smerzini, C., and Stupazzini, M.** (2018). 3D physics-based numerical simulations: Advantages and current limitations of a new frontier to earthquake ground motion prediction. The Istanbul case

- study. *Recent Advances in Earthquake Engineering in Europe: 16th European Conference on Earthquake Engineering-Thessaloniki 2018*, pages 203–223. Springer.
Referenced on page 17.
- Papageorgiou, A.** (2003). Sufficient conditions for fast quasi-Monte Carlo convergence. *Journal of Complexity*, **19**(3):332–351. doi:10.1016/S0885-064X(02)00004-3.
Referenced on page 78.
- Pasquale, G. D., Orsini, G., and Romeo, R. W.** (2005). New developments in seismic risk assessment in Italy. *Bulletin of Earthquake Engineering*, **3**(1):101–128. doi:10.1007/s10518-005-0202-1.
Referenced on page 17.
- Pianosi, F., Sarrazin, F., and Wagener, T.** (2015). A Matlab toolbox for global sensitivity analysis. *Environmental Modelling and Software*, **70**(February 2016):80–85. doi:10.1016/j.envsoft.2015.04.009.
Referenced on page 105.
- Pianosi, F. and Wagener, T.** (2015). A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environmental Modelling and Software*, **67**:1–11. doi:10.1016/j.envsoft.2015.01.004.
Referenced on page 114.
- Power, M., Chiou, B., Abrahamson, N., Bozorgnia, Y., Shantz, T., and Roblee, C.** (2008). An overview of the NGA project. *Earthquake Spectra*, **24**(1):3–21. doi:10.1193/1.2894833.
Referenced on pages 9 and 48.
- Press, W. and Farrar, G.** (1990). Recursive stratified sampling for multidimensional Monte Carlo integration. *Computers in Physics*, **4**(2):190–195.
Referenced on pages 85 and 102.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P.** (2007). *Numerical Recipes*. Cambridge University Press, 3rd edition.
Referenced on page 102.
- Pulido, N., Ojeda, A., Atakan, K., and Kubo, T.** (2004). Strong ground motion estimation in the Sea of Marmara region (Turkey) based on a scenario earthquake. *Tectonophysics*, **391**:357–374. doi:10.1016/j.tecto.2004.07.023.
Referenced on page 15.
- Reidler, I., Aviad, Y., Rosenbluh, M., and Kanter, I.** (2009). Ultrahigh-speed random number generation based on a chaotic semiconductor laser. **103**(2). doi:10.1103/PhysRevLett.103.024102.
Referenced on page 28.

- Reiter, L.** (1990). *Earthquake Hazard Analysis*. Columbia University Press, New York, 1st edition.
Referenced on page 7.
- Reusch, F. A.** (2017). *Development of damage distribution and whose application in NatCat loss models taken into account different kind of correlations*. Master Thesis, Department of Mathematics, Ludwig-Maximilians-Universität München.
Referenced on page 45.
- Robert, C. P. and Casella, G.** (2004). *Monte Carlo statistical methods*. Springer Texts in Statistics. Springer New York. doi:10.1007/978-1-4757-4145-2.
Referenced on pages 28 and 78.
- Rosenhauer, W.** (1983). Methodological aspects encountered in the Lower Rhine Area seismic hazard analysis. *Seismicity and Seismic Risk in the Offshore North Sea Area*, pages 385–396. Springer Netherlands, Dordrecht. doi:10.1007/978-94-009-7046-5_33.
Referenced on page 8.
- Rossetto, T., Ioannou, I., and Grant, D.** (2013). Existing empirical fragility and vulnerability relationships: Compendium and guide for selection. *Technical report*. doi:10.13117/GEM.VULN-MOD.TR2015.01.
Referenced on page 23.
- Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S.** (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, **181**(2):259–270. doi:10.1016/j.cpc.2009.09.018.
Referenced on pages 103 and 104.
- Samuel, M., Harbury, N., Bakri, A., Banner, F., and Hartono, L.** (1997). A new stratigraphy for the islands of the Sumatran Forearc, Indonesia. *Journal of Asian Earth Sciences*, **15**(4):339–380. doi:10.1016/S0743-9547(97)87720-3.
Referenced on page 54.
- Sarabandi, P., Pachakis, D. D., King, S. A., and Kiremidjian, A. S.** (2004). Empirical fragility functions from recent earthquakes. *13th World Conference on Earthquake Engineering*, page Paper No. 1211.
Referenced on page 23.
- Sarrazin, F., Pianosi, F., and Wagener, T.** (2016). Global Sensitivity Analysis of environmental models: Convergence and validation. *Environmental Modelling and Software*, **79**(1):135–152. doi:10.1016/j.envsoft.2016.02.005.
Referenced on pages 104 and 112.
- Scheingraber, C.** (2015). *Uncertainty analysis for the modeling of earthquake risk under consideration of a single risk functionality*. Master Thesis, Department of Informatics,

Technical University of Munich.

Referenced on pages 23, 43, and 44.

Scheingraber, C. and Käser, M. A. (2018). The impact of portfolio location uncertainty on probabilistic seismic risk analysis. *Risk Analysis. Advance online publication*. doi:10.1111/risa.13176.

Referenced on pages 74, 77, 80, and 88.

Schlüter, H. U., Gaedicke, C., Roeser, H., Schreckenberger, B., Meyer, H., and Reichert, C. (2002). Tectonic features of the southern Sumatra-western Java forearc of Indonesia. *Tectonics*, **21**(5). doi:10.1029/2001TC901048.

Referenced on page 54.

Schürer, R. (2004). Adaptive Quasi-Monte Carlo Integration Based on MISER and VEGAS. H. Niederreiter, editor, *Monte Carlo and Quasi-Monte Carlo Methods 2002*, pages 393–406. Springer Verlag.

Referenced on page 102.

Senior Seismic Hazard Committee (SSHAC) (1997). Recommendations for probabilistic seismic hazard analysis: guidance on uncertainty and use of experts. *NUREG/CR-6372*. doi:NUREG/CR-6372Vol.1.

Referenced on pages 10, 12, 48, 50, 75, and 98.

Shapira, A. (1983). Potential earthquake risk estimations by application of a simulation process. *Tectonophysics*, **95**(1):75–89. doi:10.1016/0040-1951(83)90260-3.

Referenced on page 8.

Silva, V., Crowley, H., Varum, H., and Pinho, R. (2014). Seismic risk assessment for mainland Portugal. *Bulletin of Earthquake Engineering*, (2):429–457. doi:10.1007/s10518-014-9630-0.

Referenced on page 19.

Smolka, A., Allmann, A., Hollnack, D., and Thrainsson, H. (2004). The principle of risk partnership and the role of insurance in risk mitigation. *Proceedings of the 13th World Conference on Earthquake Engineering*. Vancouver, Canada.

Referenced on page 1.

Sobol', I. M. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, **7**(4):784–802.

Referenced on page 102.

Sobol, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, **55**(1):271–280. doi:10.1002/wilm.42820050114.

Referenced on page 104.

- Sobol', I. M. and Kucherenko, S.** (2009). Derivative based global sensitivity measures and their link with global sensitivity indices. *Mathematics and Computers in Simulation*, **79**(10):3009–3017. doi:10.1016/j.matcom.2009.01.023.
Referenced on page 114.
- Socquet, A., Simons, W., Vigny, C., McCaffrey, R., Subarya, C., Sarsito, D., Ambrosius, B., and Spakman, W.** (2006). Microblock rotations and fault coupling in SE Asia triple junction (Sulawesi, Indonesia) from GPS and earthquake slip vector data. *Journal of Geophysical Research: Solid Earth*, **111**(B08409). doi:10.1029/2005JB003963.
Referenced on page 55.
- Spence, R. J. S., Coburn, A. W., Pomonis, A., and Sakai, S.** (1992). Correlation of ground motion with building damage: The definition of a new damage-based seismic intensity scale. *Proceedings of the Tenth World Conference on Earthquake Engineering, Madrid, Spain*, volume 1, pages 551–556.
Referenced on page 18.
- Stafford, P. J.** (2012). Evaluation of structural performance in the immediate aftermath of an earthquake: a case study of the 2011 Christchurch earthquake. *International Journal of Forensic Engineering*, **1**(1):58–77. doi:10.1504/IJFE.2012.047447.
Referenced on page 49.
- Stone, E. R., Bruine de Bruin, W., Wilkins, A. M., Boker, E. M., and MacDonald Gibson, J.** (2017). Designing graphs to communicate risks: Understanding how the choice of graphical format influences decision making. *Risk Analysis*, **37**(4):612–628. doi:10.1111/risa.12660.
Referenced on page 71.
- Storchak, D. A., Di Giacomo, D., Bondár, I., Engdahl, E. R., Harris, J., Lee, W. H. K., Villaseñor, A., and Bormann, P.** (2013). Public Release of the ISC–GEM Global Instrumental Earthquake Catalogue (1900–2009). *Seismological Research Letters*, **84**(5):810–815. doi:10.1785/0220130034.
Referenced on pages 56, 149, and 151.
- Strasser, F. O., Bommer, J. J., and Abrahamson, N. A.** (2008). Truncation of the distribution of ground-motion residuals. *Journal of Seismology*, **12**(1):79–105. doi:10.1007/s10950-007-9073-z.
Referenced on pages 16 and 67.
- Tesfamariam, S., Sadiq, R., and Najjaran, H.** (2010). Decision making under uncertainty - An example for seismic risk management. *Risk Analysis*, **30**(1):78–94. doi:10.1111/j.1539-6924.2009.01331.x.
Referenced on page 49.

- Toro, G. R., Abrahamson, N. a., and Schneider, J. F.** (1997). Model of strong ground motions from earthquakes in Central and Eastern North America: best estimates and uncertainties. *Seismological Research Letters*, **68**(1):41–57. doi:10.1785/gssrl.68.1.41.
Referenced on page 56.
- Von Neumann, J.** (1951). Various techniques used in connection with random digits. *Appl. Math Ser*, **12**(36-38):3.
Referenced on page 28.
- Wald, D. J. and Allen, T. I.** (2007). Topographic slope as a proxy for seismic site conditions and amplification. *Bulletin of the Seismological Society of America*, **97**(5):1379–1395. doi:10.1785/0120060267.
Referenced on pages 54, 57, and 80.
- Wang, M. and Takada, T.** (2005). Macrospatial correlation model of seismic ground motions. *Earthquake Spectra*, **21**(4):1137–1156.
Referenced on page 37.
- Weatherill, G. and Burton, P. W.** (2010). An alternative approach to probabilistic seismic hazard analysis in the Aegean region using Monte Carlo simulation. *Tectonophysics*, **492**(1-4):253–278. doi:10.1016/j.tecto.2010.06.022.
Referenced on pages 48 and 55.
- Weatherill, G., Silva, V., Crowley, H., and Bazzurro, P.** (2015). Exploring the impact of spatial correlations and uncertainties for portfolio analysis in probabilistic seismic loss estimation. *Bulletin of Earthquake Engineering*, **13**:957–981. doi:10.1007/s10518-015-9730-5.
Referenced on pages 9, 34, 36, 37, 40, and 41.
- Wesson, R. L., Perkins, D. M., Luco, N., and Karaca, E.** (2009). Direct calculation of the probability distribution for earthquake losses to a portfolio. *Earthquake Spectra*, **25**(3):687–706. doi:10.1193/1.3159475.
Referenced on page 11.
- Whitman, R., Reed, J., and Hong, S.** (1973). Earthquake damage probability matrices. *Proceedings of Fifth World Conference on Earthquake Engineering*, pages 2531–2540. Rome, Italy.
Referenced on page 17.
- Whitman, R. V., Anagnos, T., Kircher, C. A., Lagorio, H. J., Lawson, R. S., and Schneider, P.** (1997). Development of a national earthquake loss estimation methodology. *Earthquake Spectra*, **13**(4):643–661. doi:10.1193/1.1585973.
Referenced on pages 18 and 22.

- Yang, W.-N. and Nelson, B. L.** (1991). Using common random numbers and control variates in multiple-comparison procedures. *Operations Research*, **39**(4):583–591.
Referenced on page 74.
- Yepes, C., Silva, V., Valcárcel, J., Acevedo, A., Tarque, N., Hube, M., Coronel, D., and Santa María, H.** (2010). Modelling the residential building inventory in South America for seismic risk assessment. *Earthquake Spectra*, **33**(1). doi:10.1193/101915EQS155DP.
Referenced on page 19.
- Youngs, R. R., Chiou, S.-J., Silva, W. J., and Humphrey, J. R.** (1997). Strong ground motion attenuation relationships for subduction zone earthquakes. *Seismological Research Letters*, **68**(1):58–73. doi:10.1785/gssrl.68.1.58.
Referenced on pages 56 and 105.
- Zhao, J. X., Zhang, J., Asano, A., Ohno, Y., Oouchi, T., Takahashi, T., Ogawa, H., Irikura, K., Thio, H. K., Somerville, P. G., and Others** (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. doi:10.1785/0120050122.
Referenced on pages 56, 105, and 106.

Appendix A

OpenQuake Model Explorer: Visualization and Conversion of Earthquake Models in NRML format

A.1 Introduction

This chapter describes how to use OpenQuake Model Explorer, a graphical tool which has been developed during the course of this thesis in order to visualize, analyze and process seismic hazard models.

Global Earthquake Model The Global Earthquake Model¹ (GEM) is a public-private partnership initiated by the Organization for Economic Co-operation and Development² (OECD), an internationally sanctioned program which aims to build an independent and open standard for modeling and communicating earthquake risk.

OpenQuake and NRML OpenQuake³ is a free and open-source seismic hazard and risk assessment framework developed as part of the GEM initiative. The OpenQuake engine can be used to perform calculations using hazard and risk models provided by GEM. GEM and OpenQuake use the Natural hazards' Risk Markup Language (NRML), an open file-format for natural catastrophe hazard and risk models.

OpenQuake Model Explorer OpenQuake Model Explorer allows to visualize and analyze earthquake source models in NRML format. Moreover, the tool helps to convert NRML files into a binary format used for model development at Munich Reinsurance (see Section 1.1). OpenQuake Model Explorer is implemented in the Python Programming

¹See <https://www.globalquakemodel.org/>.

²See <https://www.oecd.org/>.

³See <https://www.globalquakemodel.org/openquake/about/> and <https://github.com/gem/oq-engine>.

Language⁴. A visualization-only version has been made available publicly⁵. This version lacks all of the described conversion functionality.

A.2 Graphical User Interface

OpenQuake Model Explorer provides a graphical user interface (GUI), which can be used to visualize, analyze, crop and convert NRML source models. Figure A.1 shows the main window of the interface.

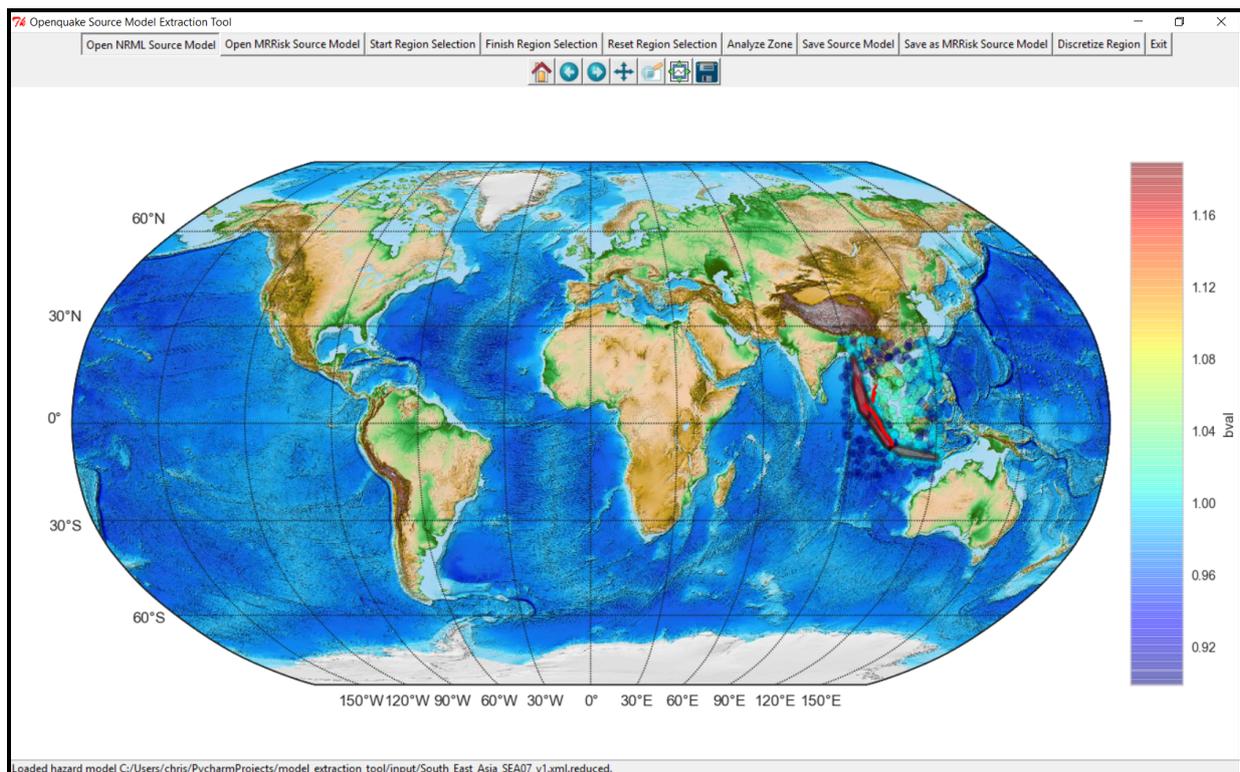


Figure A.1: Main window of OpenQuake Model Explorer after loading a source model. The upper panel contains buttons to use all described functionality, the status bar on the bottom gives information on loaded model files and the current progress of tasks.

A.2.1 Visualization of Sources and Seismicity

Once a NRML model has been loaded using the leftmost button on the top panel, a graphical representation of all seismic sources contained in the model will be plotted into the world map. Point sources are shown as color-coded circles. The color-code denotes the b -value (slope) of a fit of the truncated Gutenberg-Richter magnitude-frequency distribution

⁴See <https://www.python.org/>

⁵See https://github.com/scheingraber/oq_model_explorer.

(MFD) to the *incremental MFD* of the original NRML source, which is a discretization of an arbitrary MFD. Simple faults are shown as red lines, and big transparent red patches show the surface projection of complex faults.

To inspect the model, the world map can be zoomed to any area of interest using the zoom button on the top panel and the right mouse button. Double-clicking on a seismic source opens a popup window (see Figure A.2) with two plots. The left plot shows the annual occurrence rate, the right plot shows the annual exceedance rate of the MFD corresponding to the seismic source. The plots show the original incremental MFD (solid blue) as well as the fitted truncated Gutenberg-Richter MFD (dashed red) used to visualize the b-value.

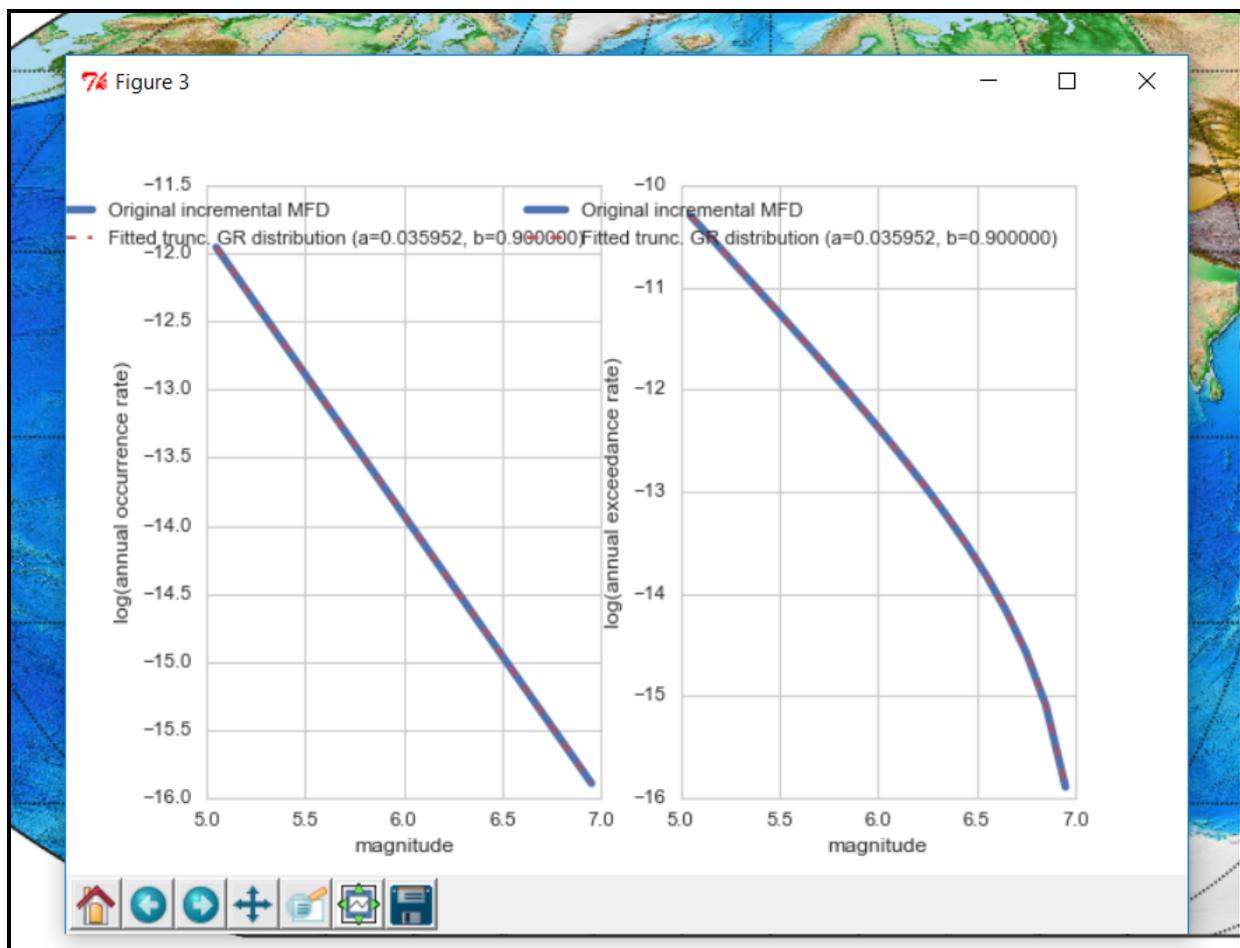


Figure A.2: Popup window after double-clicking on a seismic source showing the discretized magnitude-frequency-distribution of the respective seismic source, as well as a fitted truncated Gutenberg-Richter (GR) distribution. In this case, the discretized magnitude-frequency-distribution was also modeled using a truncated GR distribution, and could therefore be fitted exactly.

A.2.2 Extraction of Geographical Subsets

Using the GUI, a smaller geographical area can be cropped from the original NRML source model in an interactive manner. Clicking the button *Start Region Selection* on the top panel starts the region-selection mode. A left mouse button click adds a new vertex to the polygon connecting to the last vertex via the corresponding great-circle path, and a right mouse button click removes the last vertex. The polygon is closed automatically when the region selection is finished using the *Finish Region Selection* button on the top panel. After a region has been selected, a dialog window opens, allowing to extract the selected region.

For large NRML input models, subset extraction can be a time-consuming operation. The progress is indicated using the status bar on the bottom.

A.2.3 Converting and Saving a Source Model

To support model development at Munich Reinsurance, a model can be converted to a binary MATLAB format from the GUI of OpenQuake Model Explorer. The model is converted automatically when clicking the button *Save MRRisk Source Model* on the top panel, which opens a dialog window asking where to save the file.

A.3 Usage of OpenQuake Model Explorer from a Python Script

Due to the modular, object-oriented software architecture of OpenQuake Model Explorer, the conversion functionality can also be used from a Python script instead of the GUI. This allows for efficient batch processing of a large number of model input files. The application programming interface (API) can be obtained from the reference (docstring) contained in the source code of each module.

For the South-East Asia hazard model by Mark Petersen et al. (2007) of the United States Geological Service (USGS), the following example creates two model containers from NRML input files, separating seismic sources which are shallower than 35 kilometers from seismic sources which are deeper than 35 kilometers. For each model container, model converters are then created, which provide methods to create analytic plots and save the source model to the binary MATLAB output format. In this example, this is performed for both containers:

```
1 #!/usr/bin/env python
2
3 """
4 Read GEM SE-Asia .nrml model, create plots and save information to .mat.
5 """
```

```
6
7 from mr.source.sourceconverter import SourceModelConverter
8 from mr.source.sourcemodelcontainer import NrmlModelContainer,
  MrModelContainer
9
10 # filenames
11 nrml_model_filename = 'South_East_Asia_SEA07_v1.xml'
12 mr_model_filename = 'SouthEastAsia_v1.mat'
13
14 # instantiate source model containers
15 nrml_model_shallow = NrmlModelContainer(nrml_model_filename, min_depth=0,
  max_depth=35)
16 nrml_model_deep = NrmlModelContainer(nrml_model_filename, min_depth=35,
  max_depth=60)
17
18 mr_model = MrModelContainer(mr_model_filename)
19
20 # source model converters hold both containers
21 smc_shallow = SourceModelConverter(nrml_model_shallow, mr_model)
22 smc_deep = SourceModelConverter(nrml_model_deep, mr_model)
23
24 # plot model
25 smc_shallow.plot(filename='se_asia_bval_shallow.png', topo_map=False,
  density_hack=False, dpi=900, point_src_color='bval', point_src_size=5,
  max_point_sources=120000)
26 smc_shallow.plot(filename='se_asia_momentrate_shallow.png', topo_map=False,
  density_hack=False, dpi=900, point_src_color='moment_rate',
  point_src_size=5, max_point_sources=120000)
27
28 smc_deep.plot(filename='se_asia_bval_deep.png', topo_map=False, density_hack
  =False, dpi=900, point_src_color='bval', point_src_size=5,
  max_point_sources=120000)
29 smc_deep.plot(filename='se_asia_momentrate_deep.png', topo_map=False,
  density_hack=False, dpi=900, point_src_color='moment_rate',
  point_src_size=5, max_point_sources=120000)
30
31 # save to .mat format
32 smc_shallow.save_sorted_mat('se_asia_full_sorted')
33 smc_deep.save_sorted_mat('se_asia_full_sorted')
```

The example above generates and saves analytic plots of the b-value of the Gutenberg-Richter distribution as well as the seismic moment rate. The plot of the b-values of shallow sources is shown in Figure A.3.

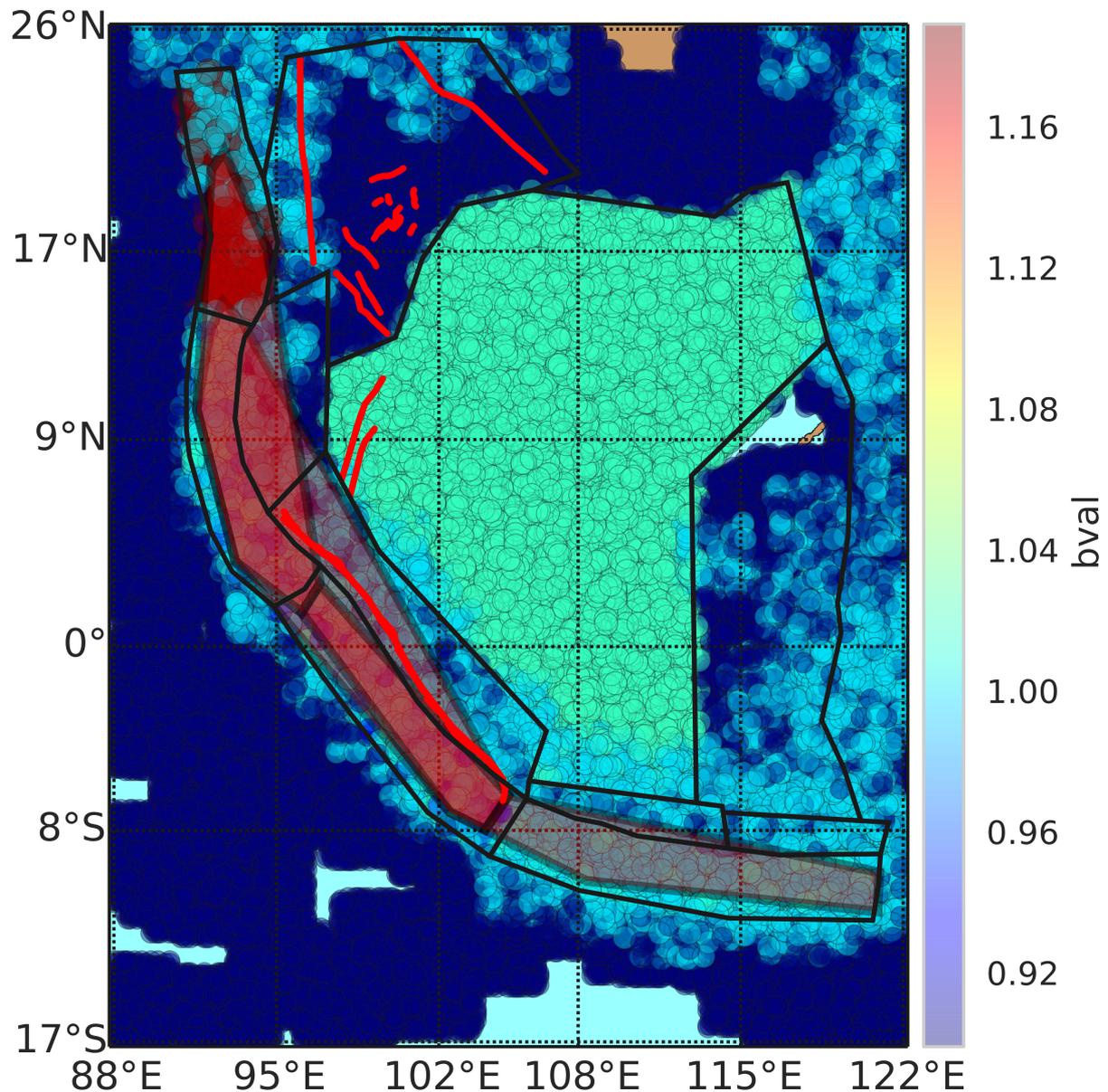


Figure A.3: Example plot of b-values of the Gutenberg-Richter magnitude frequency distribution produced by OpenQuake Model Explorer. This shows b-values for shallow sources in Indonesia contained in the larger South-East Asia hazard model by Mark Petersen et al. (2007) of the United States Geological Service (USGS). Point sources are shown as color-coded circles, with color denoting the b-value of the Gutenberg-Richter magnitude-frequency distribution (MFD) of the seismic source. Simple faults are shown as red lines, and big transparent red patches show the surface projection of complex faults. Area sources are outlined in black.

Appendix B

Source Zone Creator: Graphical Tool for Seismic Hazard Model Development

This chapter describes how to use Source Zone Creator, a graphical tool which has been developed during the course of this thesis to support the creation of seismic source zones in hazard model development.

B.1 Introduction

A common task during seismic hazard model development is the creation of seismic source zones. Seismic source zones share a common magnitude frequency distribution (MFD), which is spatially distributed within the zone using variable weights.

Source Zone Creator is a graphical tool implemented in MATLAB to support the interactive creation of seismic source zones, including their geometry and MFD. Source zones can be created using a catalogue of historical and instrumental seismicity (e.g. Storchak et al., 2013), or distributed point sources from hazard models published by the Global Earthquake Model (GEM, see Section A.1). GEM hazard models need to be converted to a binary MATLAB format using OpenQuake Model Explorer (see Chapter A) before they can be used with Source Zone Creator.

B.2 Graphical User Interface

Source Zone Creator uses a Graphical User Interface (GUI, see Figures B.1 and B.2). On the left hand side, the source zone geometry of the currently loaded model are shown in red, and a country coastline in black.

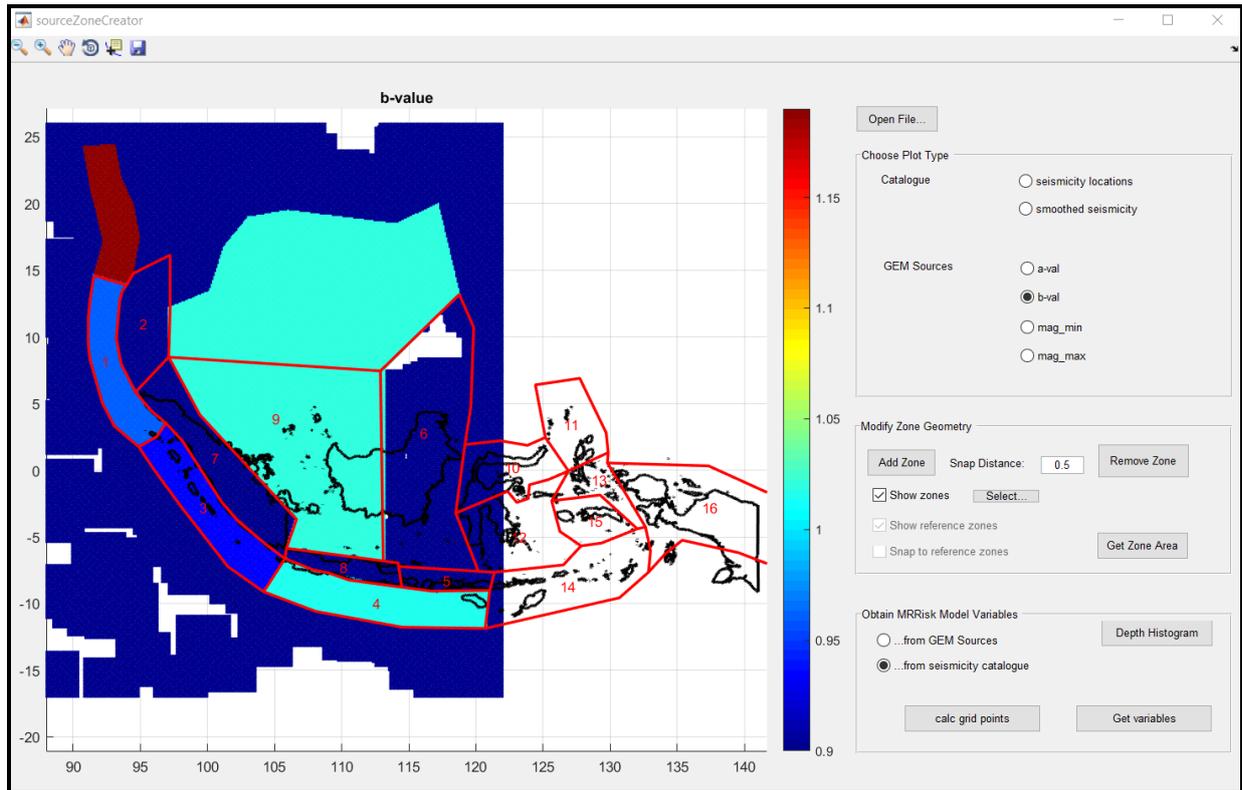


Figure B.1: Main Window of Source Zone Creator showing zones of the model and b-values for shallow point sources in Indonesia contained in the larger South-East Asia hazard model by Mark Petersen et al. (2007) of the United States Geological Service (USGS).

B.2.1 Source Model and Catalogue Visualization

Loading Input Files

Input files such as Munich Re Risk models (see Section 1.1), historical seismicity catalogues or GEM hazard models can be loaded using the topmost button of the panel on the right hand side. In the beginning, most panel elements on the right are greyed out. Once specific input files have been loaded, the corresponding panel elements on the right become active.

Choosing a Plot Type

The radio button under *Choose Plot Type* can be used to plot properties of the MFD of seismic point sources from a loaded GEM source model (e.g. minimum magnitude, maximum magnitude, or a- and b-values, see Figure B.1), or seismicity locations and smoothed seismicity from a loaded catalogue (e.g. Figure B.2). The model view on the left and the corresponding colorbar adapts automatically to the chosen plot type.

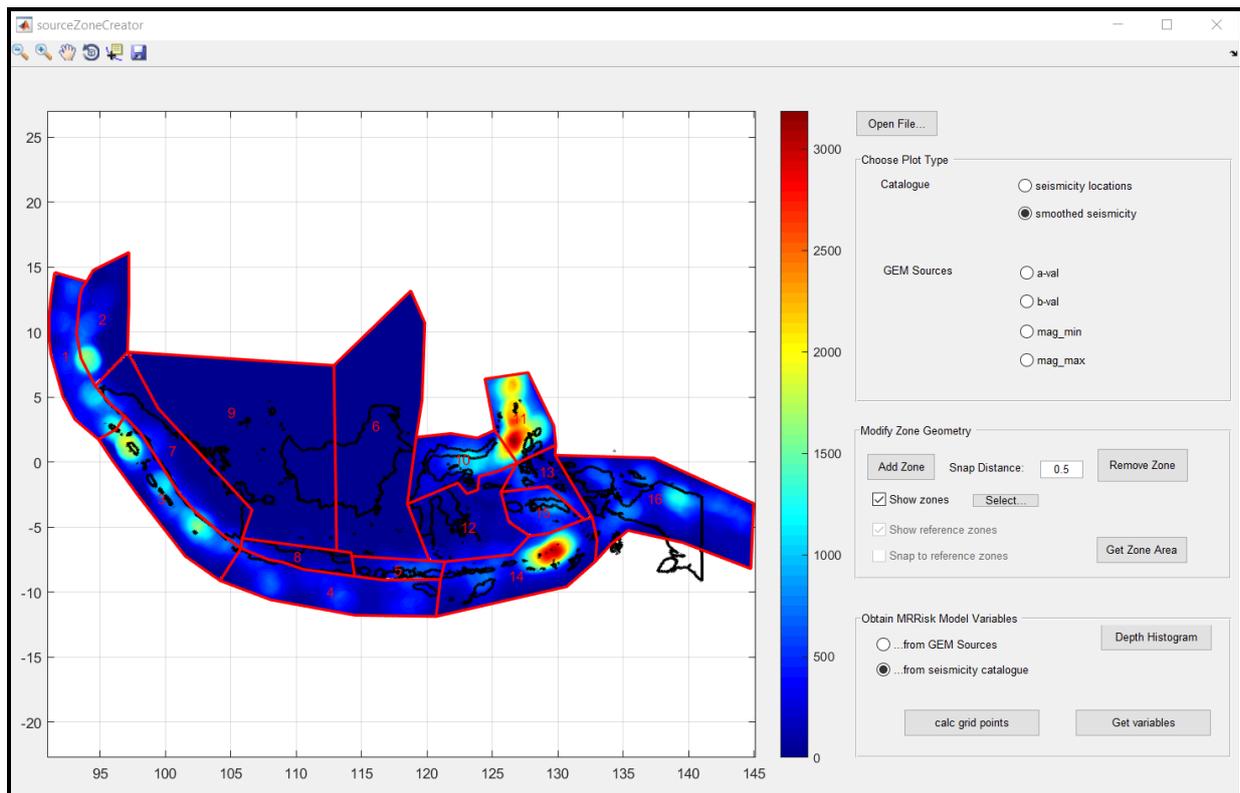


Figure B.2: Main Window of Source Zone Creator showing zones of the model and smoothed seismicity from an historical and instrumental seismic catalogue (Storchak et al., 2013).

B.2.2 Creating and Modifying the Geometry of Seismic Source Zones

Below the *Modify Zone Geometry* header are several buttons to add and remove seismic source zones from the model, and to select which zones to show in the view on the left. When adding zones, a left click adds a new vertex to the zone, a middle click removes the last vertex, and a right click finishes the creation of the zone¹. New vertices can be *snapped* to previous vertices, i.e. a new vertex gets the same coordinates as a previous one when clicking very close.

B.2.3 Obtaining the Final Model Variables

The final model variables for a Munich Re Risk model can be obtained in the lowermost section of the panel on the right hand side, labeled *Obtain MRRisk Model Variables*. Upon clicking the *Get variables* button, source zone geometries, source zones MFDs and grid

¹An animation of the functionality is available at <http://archive.scheingraber.net/animation/sourceZoneCreator.gif>.

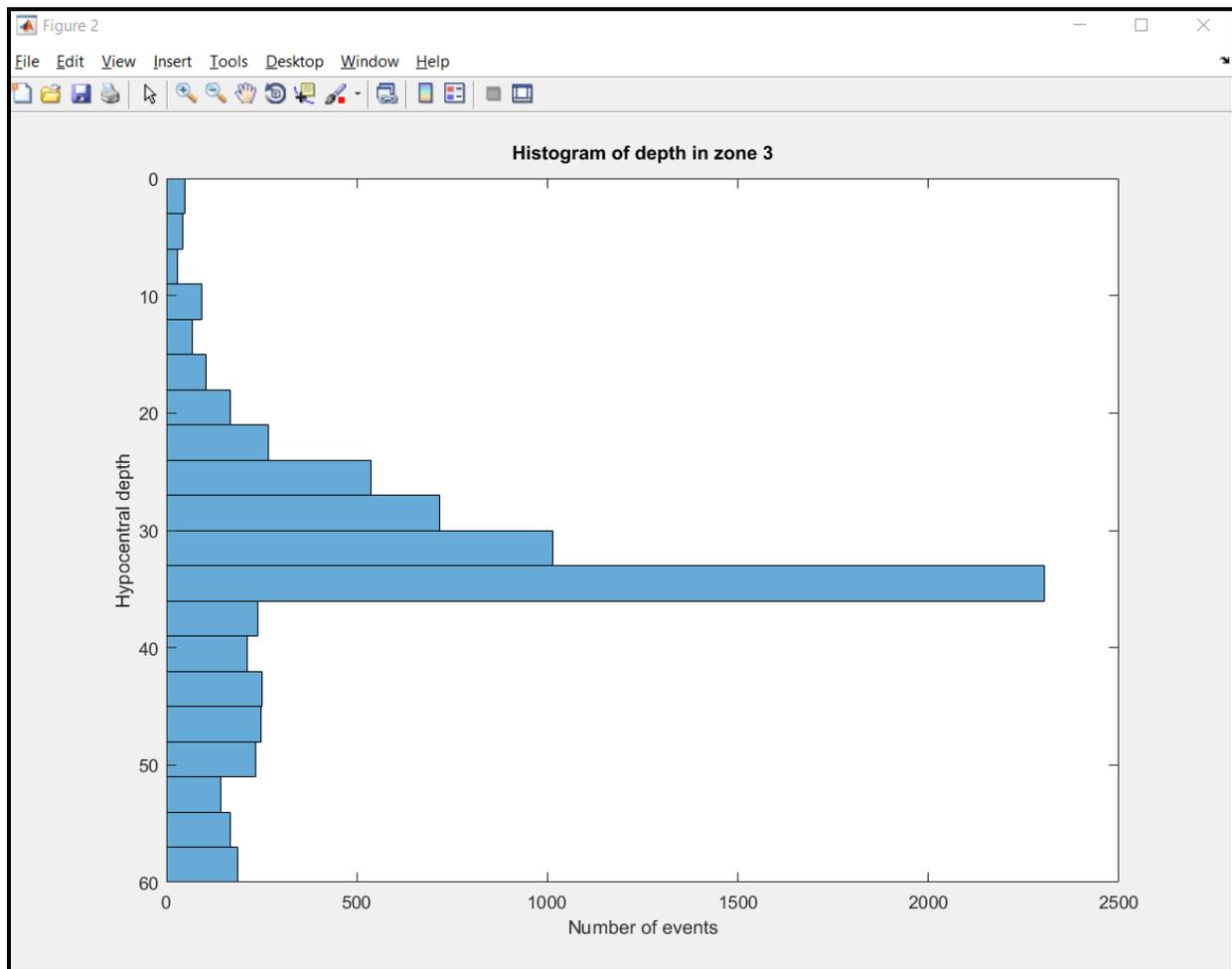


Figure B.3: Histogram of hypocentral depths within a seismic source zone obtained when clicking the *Depth Histogram* button.

point weights are calculated for all selected zones. A radio button is used to select whether this should be performed using the seismicity catalogue or using GEM point sources. Since each source zone can only be assigned 3 discrete depths, the button *Depth histogram* can be used to open a popup window showing a histogram of hypocentral depths of all events within a zone, again showing the seismic catalogue or GEM sources (see Figure B.3).

Acknowledgements

Many thanks to Christoph Bach for improving several aspects of this tool.

Acknowledgements

This work would not have been possible without the direct or indirect support from many people that I would like to thank:

Martin; for being an unmatched teacher, supervisor and mentor throughout these enjoyable years, for the motivating and fruitful discussions, for the extensive, thoughtful and quick support during the creation of every part of this work, for complete backing and always being open and approachable to discuss new ideas, for trusting my motivation and giving me freedom. And for striking the right balance between concentrated work and fun breaks, which enabled a highly motivated and productive atmosphere in the entire team.

Heiner; for reliable and pragmatic backing over many years, for the interesting seminars, for enabling me to travel to many conferences and providing a workplace at LMU. And for sparking my interest in geophysics a long time ago.

Alex; for valuable suggestions which improved many parts of this work, for providing the scholarship as well as the chance to live in Pavia for a month, for the fun team events (CU boulder challenge!) and generally setting an always positive environment, thereby enabling the upbeat and productive atmosphere of the Earthquake and NatCat modeling teams at Munich Re.

Marco, Wilhelm, Christoph, John, Doris, Sophie, Morten, Jan, Anja, Christian, Gabi and many others at Munich Re; for being great officemates and colleagues, for supporting me in many parts of this work, and for the interesting and stimulating discussions about a wide variety of topics during coffee and lunch. I think we were always on the same page and shared many epic stories! John for proofreading Chapter 4 and Chapter 6. Gabi for helping me with many administrative tasks, always in a friendly manner.

Steffi, Sia, Thomas, Tobi, Lion, Kasra, Martin M., and many others at LMU; for the enlightening discussions and fun times. A special thanks to Steffi for proofreading and providing valuable comments that considerably improved Chapter 2.

Jens Ö.; for flawless IT support at LMU and for providing a machine to run OpenQuake.

The administrative team at LMU, especially Elena; for a good deal of paper work and helping me with travel reimbursements, always in a likeable and friendly manner that actually made filling forms fun.

Marco P., Graeme, Vitor, Daniele, and in general the entire hazard, risk and IT teams that were in Pavia when I visited GEM; for the stimulating discussions and for making me feel welcome in Pavia. Daniele for helping with the OpenQuake cluster setup, Marco for helping with numerous questions regarding OpenQuake jobs, Graeme for getting me started and clarifying open questions regarding OpenQuake and stochastic simulation.

Carlos; not only for providing me a place to stay during EGU, but most importantly for being a great friend.

All my other friends; for their friendship, for the fulfilling and fun times that provided the necessary recreation. So stoked on living life with you!

My family; for their unconditional love and support, for the fulfilling and fun times — but I won't really get started on reasons since it doesn't fit in here anyway!

Opa H.; who passed away shortly before I started this work, but actually contributed to it by giving me a first idea and fascination for Monte Carlo simulation (estimating the value of π in BASIC on a 1985 Toshiba T1100) and sparking my fascination for computers and programming pretty much as soon as I could read and type English source code.

