

“OpenQuake: Calculate, share, explore”

# The OpenQuake-engine Book: Hazard

Copyright © 2014 GEM Foundation

PUBLISHED BY GEM FOUNDATION

GLOBALQUAKEMODEL.ORG/OPENQUAKE

### Citation

Please cite this document as:

Pagani, M., Monelli, D., Weatherill, G. A. and Garcia, J. (2014) The OpenQuake-engine Book: Hazard. *Global Earthquake Model (GEM). Technical Report*

### Disclaimer

The OpenQuake Engine Book Hazard is distributed in the hope that it will be useful, but without any warranty: without even the implied warranty of merchantability or fitness for a particular purpose. While every precaution has been taken in the preparation of this document, in no event shall the authors of the manual and the GEM Foundation be liable to any party for direct, indirect, special, incidental, or consequential damages, including lost profits, arising out of the use of information contained in this document or from the use of programs and source code that may accompany it, even if the authors and GEM Foundation have been advised of the possibility of such damage. The Book provided hereunder is on as "as is" basis, and the authors and GEM Foundation have no obligations to provide maintenance, support, updates, enhancements, or modifications.

This needs to be updated or removed

The current version of the book has been revised only by members of the GEM model facility and it must be considered a draft copy.

### License

This Book is distributed under the Creative Common License Attribution-NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0) (see link below). You can download this Book and share it with others as long as you provide proper credit, but you cannot change it in any way or use it commercially.

*First printing, October 2014*

# Contents

<b>Foreword</b>	5
<b>1 Introduction</b>	7
1.1 Overview of the OpenQuake-engine	7
1.2 Overview of the OpenQuake-engine development process	8
1.3 The basics of the OpenQuake-engine hazard component	9
1.4 Input and output description	12
1.5 Description of book structure	12
<b>2 PSHA with the OpenQuake-engine</b>	13
2.1 Basic concepts	13
2.2 Classical PSHA	14
2.3 Event-based PSHA	16
2.4 Disaggregation	18
<b>3 Seismic Source Models</b>	21
3.1 Basic concepts	21
3.2 The Point and Area sources	23
3.3 The Simple Fault source	25
3.4 The Complex Fault source	26
3.5 The Characteristic Fault source	29

<b>4</b>	<b>Ground Motion Prediction Equations .....</b>	<b>31</b>
4.1	Introduction	31
4.2	Implementation and use of GMPEs in seismic hazard analysis	34
4.3	Future developments	37
<b>5</b>	<b>Logic-trees .....</b>	<b>39</b>
5.1	Introduction	39
5.2	The OpenQuake-engine logic-tree structure	40
5.3	Logic tree processing	43
5.4	Future developments	46
<b>6</b>	<b>Calculation Examples .....</b>	<b>47</b>
6.1	Classical PSHA with an Area Source	47
6.2	Classical PSHA with complex logic tree	50
6.3	Convergence between Classical and Event-based PSHA	51
6.4	Disaggregation analysis	51
<b>7</b>	<b>Bibliography .....</b>	<b>55</b>
7.1	Books	55
7.2	Articles	55
7.3	Other Sources	56
	<b>Index .....</b>	<b>59</b>

## **Foreword**

Here we should have the Foreword



<b>Overview of the OpenQuake-engine</b>
Structure of the OpenQuake-engine
<b>Overview of the OpenQuake-engine development process</b>
Development tools
Programming language
<b>The basics of the OpenQuake-engine hazard component</b>
Calculation workflows
Testing and Quality Assurance
<b>Input and output description</b>
<b>Description of book structure</b>

# 1. Introduction

This chapter provides an overview of the OpenQuake-engine (OQ-engine), its structure and the processes adopted for its development. A particular emphasis is placed on transparency, reproducibility, community-based development and testing (Pagani et al., 2014a), the central tenets of the development process adopted since the early stages of the project.

## 1.1 Overview of the OpenQuake-engine

The OQ-engine is an open-source hazard and risk calculation engine developed by the Global Earthquake Model (GEM) initiative. The OQ-engine is part of OpenQuake, a suite of open-source software packages developed by GEM (Figure 1.1), which comprises the OQ-engine, the OpenQuake-platform and a large set of tools of which the most interesting from a hazard perspective are the hazard modellers' toolkit (see <https://github.com/GEMScienceTools/hmtk>).

The development of the engine - started in 2010 and currently in progress - follows established standards adopted for the development of open-source software such as open access of the source code through an easily accessible, website and transparency of the development process<sup>1</sup>. The engine was designed to operate on computational hardware with different properties ranging from a simple laptop to a heterogeneous cluster of multi-core machines. The operative system currently supported is Ubuntu Linux (additional information on the supported version and on the installation procedure can be found on the GEM area on github, accessible at the following link: <https://github.com/gem/oq-engine>).

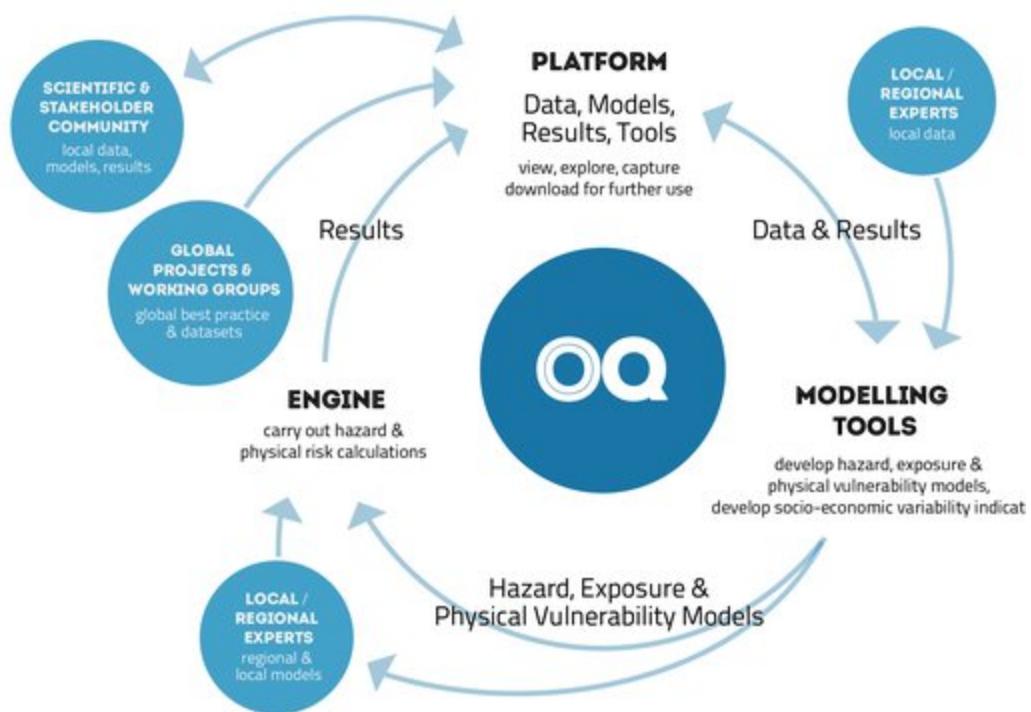
### 1.1.1 Structure of the OpenQuake-engine

The OQ-engine is the combination of different and sometimes self-sufficient libraries. Below we provide a short description for each of them.

**oq-hazardlib** Contains the code used to describe seismic sources, create the Earthquake Rupture Forecast (ERF), calculate hazard curves, create stochastic event sets, compute ground motion fields and calculate seismic hazard disaggregation.

**oq-risklib** Comprises the code used to describe exposure, vulnerability and fragility curves, and for the computation of losses.

<sup>1</sup>See for example the documentation available on the website of the Open-Source Initiative for a more comprehensive description of the development standards commonly adopted within the open-source software community



**Figure 1.1 – A schematic describing the OpenQuake suite.**

**oq-nrmllib** Includes the code relating to the reading, writing and validation of the full suite of OQ-engine input and output files. The majority of these files are formatted according to a dialect of XML called Natural hazard Risk Markup Language (NRML).

**oq-commonlib** Includes common code for OQ-engine applications, such as - for example - the code used to describe logic tree structures.

**oq-engine** It incorporates the core of the OQ-engine; the code in this library acts as the glue that sticks the different libraries together and lets the user easily perform calculations according to an established set of calculation options.

## 1.2 Overview of the OpenQuake-engine development process

The OQ-engine is developed through a close and continuous collaboration between the GEM scientific and IT teams. The development process is operated in the open in order to promote the participation of experts working in the disciplines of earthquake hazard and risk analysis, as well as those specialising in software development.

### 1.2.1 Development tools

The OQ-engine development process is based on a number of open-source tools, which guarantee an open and transparent process. For example, each new feature improvement or bug fix before being implemented is described in a bug tracking system (in our case, Launchpad - see Table 1.1). Bug tracking systems such as Launchpad keep a log of bugs and errors identified by users of the software, in addition to requests for new features.

The tools used to maintain and make publicly available the OQ-engine repository and to manage the continual improvement and enhancement process are git and a git-based repository

**Table 1.1 – Main services and websites related to the OQ-engine**

<b>Service</b>	<b>Link</b>
OQ-engine main website	<a href="http://www.globalquakemodel.org/openquake/start/engine/">http://www.globalquakemodel.org/openquake/start/engine/</a>
OQ-engine bug tracking system	<a href="https://launchpad.net/openquake">https://launchpad.net/openquake</a>
OQ-engine web repository	<a href="http://github.com/gem">http://github.com/gem</a>

hosting service called GitHub (see Table 1.1). This process ensures comprehensive version control, facilitating the tracking of feature implementation and bug fixing. It also ensures that previous versions of the software can be easily retrieved. When a developer commits new code to the main repository the record of the change is kept. If the code is intended to resolve a bug or error identified in the bug-tracking system, or implement a new feature in response to a request, the log of the code contribution should indicate the specific bug, error or feature that the code change is intended to resolve. Thus an exhaustive and audible record is kept of each problem identified and the changes to the code taken to resolve it. Table 1.1 provides a short summary of the main resources related to the OQ-engine.

### 1.2.2 Programming language

The core of the OQ-engine is developed in Python. Python is a high-level and open-source programming language extensively used in the scientific community which can run on almost all the operative systems currently available.

#### Main libraries to which the OQ-engine depends upon

The engine relies in on a number of open-source libraries such as:

**Redis** A key-value store

**RabbitMQ** A messaging system

**Celery** An asynchronous task queue/job queue.

## 1.3 The basics of the OpenQuake-engine hazard component

The hazard component of the OQ-engine has been developed mostly following an object oriented programming paradigm taking, in some cases, concepts introduced in the development of OpenSHA, a seismic hazard analysis library developed within a joint SCEC-USGS collaboration (Field et al., 2003).

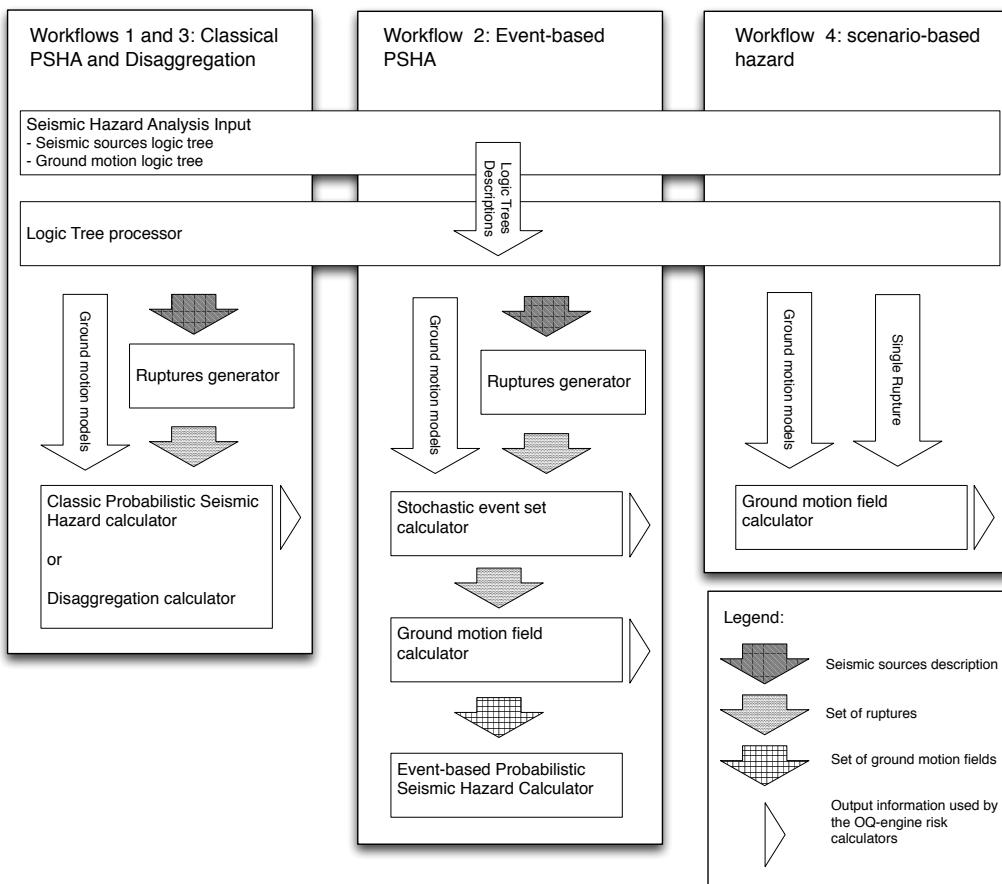
From a conceptual point of view, the main objects adopted in the development of the oq-hazardlib follows quite closely the classical schematic proposed by Reiter (1991) i.e. a seismic source, a ground shaking intensity model and a calculator that using this information computes the hazard at the site.

The OQ-engine builds on top of oq-hazardlib and expands this concept by taking into account not just the essential objects needed to compute the hazard at a site discussed before but also the associated epistemic uncertainties.

### 1.3.1 Calculation workflows

The hazard component of the OQ-engine provides four main calculation workflows (see Figure 1.2):

- Classical Probabilistic Seismic Hazard Analysis (PSHA). Calculates hazard curves, hazard maps, and uniform hazard spectra by solving the PSHA integration procedure, as proposed by Field et al. (2003). This is the usual approach adopted in regional/national-scale hazard



**Figure 1.2 – A schematic describing the main OpenQuake-engine calculation workflows available in the hazard component.**

assessment, as well as in site-specific studies. Using the risk component of the OQ-engine, the computed hazard curves can be combined with a vulnerability and exposure model to derive asset-specific loss exceedance curves and loss maps for various return periods. Such analyses are useful for comparative risk assessment between assets at different locations, or to understand the areas where mitigation actions should be concentrated. Crowley and Bommer (2006) suggest this methodology tends to overestimate losses at high return periods for portfolios of structures and recommend the use of methods capable to account for the spatial correlation of ground motion residuals.

- **Event-based PSHA.** Computes stochastic event sets (i.e., synthetic catalogs of earthquake ruptures) and ground-motion fields for each rupture, possibly taking into account the spatial correlation of within-event residuals. This is essentially a Monte Carlo-based PSHA calculator (Musson, 2000). The computed synthetic catalogs can be used for comparisons against a real catalog, whereas hazard curves and hazard maps can be derived from post-processing the ground-motion fields (Ebel and Kafka, 1999). Ground-motion fields are essential input for loss estimations, whereby loss exceedance curves and loss maps are calculated for a collection of assets by combining a vulnerability and exposure model with these sets of ground-motion fields. Because the spatial correlation of the ground-motion residuals can be taken into account in this calculator, the losses to each asset can be summed per ground-motion field, and a total loss exceedance curve representative of the whole collection of assets can be derived. These results are important for deriving reliable

- estimates of the variance of the total losses.
- Disaggregation. Given a PSHA model, it computes the earthquake scenarios contributing the most to a given hazard level at a specific site (Bazzurro and Cornell, 1999). Currently this is done following the classical PSHA methodology; this functionality will be added to the event-based calculator in subsequent development phases.
  - Scenario-based Seismic Hazard Analysis (SHA). Given an earthquake rupture and a ground-shaking model, a set of ground-motion fields can be computed. This is a typical use case for urban-scale loss analysis. This set of ground-motion fields can be employed with a fragility/vulnerability model to calculate distribution of damage/losses for a collection of assets. Such results are of importance for emergency management planning and for raising societal awareness of risk.

### 1.3.2 Testing and Quality Assurance

Quality Assurance is an aspect carefully and diligently considered in the development of the OQ-engine. There are several different reasons for the adoption of this approach.

The first and most practical one is dictated by the development process which involves experts from different disciplines (e.g. seismic hazard and information technology). In this context the use of a formal testing process is a way through which developers confirm the compliance of the tools developed against the requirements defined by the scientific team and it is also a process through which it can be demonstrated that the entire code fulfills minimum quality criteria (e.g. the code complies with the PEP 8 standard<sup>2</sup>, the code before getting into the master repository is revised by at least one separate developer and is clearly documented).

The second motivation relates to the specific goal of building a dynamic tool (i.e. offering a large flexibility and expandability) while constantly assuring the stability and reliability of the supported calculation workflows.

The implementation of tests is usually done in parallel with code development, but tests are also added for example every time a bug is fixed. This improves the overall robustness and reliability of the code and reduces drastically the possibility of regressions.

The following approaches represent the four-level suite of tests applied to the OQ-engine and therefore provide high quality assurance standards. Further information can be found in the OQ-engine testing and quality assurance report (Pagani et al., 2014b)

**Unit-testing** A testing methodology which checks discrete units of code against associated control data, expected behaviors and operating procedures. A special set of unit-tests are the ones systematically created for every Ground Shaking Intensity Model (GSIM) implemented (additional information about this specific topic is available within Chapter 4)

**Testing against benchmark results** The results provided by the OQ-engine are compared against benchmark results. Several of the tests defined by Thomas et al. (2010) are used to check the reliability and correctness of the results provided.

**Tests against provided by other PSHA codes: simple cases** The result computed with the OQ-engine for simple models (e.g. one area source) are compared against the results calculated using independent PSHA software.

**Tests against provided by other PSHA codes: national or regional PSHA input models**

The result computed with the OQ-engine using national or regional models are compared against the results calculated using independent PSHA software.

---

<sup>2</sup>As Python is a rapidly advancing language, the Python Enhancement Proposal (PEP) is the mechanism through which new features in the language are proposed, debated and documented. Compliance with approved PEP standards ensures correctness of structure and implementation of code, thus providing clarity and facilitating continual compatibility with changes to the language.

Here we need to provide a description consistent with the structure of the document on testing

## 1.4 Input and output description

An OQ-engine PSHA input model is organised into a number of files each one including a specific information subset. The main file is a configuration file which contains information on the principal calculation settings (e.g. names of the files containing the source model and ground-motion model logic trees, type of analysis, kind of results).

An OQ-engine PSHA Input model also includes a file which describe the source model logic tree, at least one file describing the initial set of seismic sources and their properties and one file with the ground motion model logic tree.

## 1.5 Description of book structure

This part will be completed later.

## Basic concepts

### Classical PSHA

Equivalence with the rate-based equation

### Event-based PSHA

Calculation of hazard curves from ground motion fields

### Disaggregation

Disaggregation histograms

Comparison between OpenQuake-engine disaggregation and *traditional* disaggregation

## 2. PSHA with the OpenQuake-engine

This chapter describes the mathematical framework for PSHA implemented by the OpenQuake-engine. Seismicity is assumed to be described as a time-independent Poissonian process. Seismicity modeling relies therefore on the following three assumptions:

- seismicity in a region is described by a collection of *independent seismic sources* (i.e. the occurrence of an earthquake rupture in a source does not affect the probability of earthquake occurrence in the other sources)
- each source generates *independent earthquake ruptures* (i.e. the occurrence of an earthquake rupture in a source does not affect the probability of occurrence of the other potential earthquake ruptures in the same source)

If we already think to expand the engine in order to accomodate mutually exclusive ruptures, I would move this condition below this list.

- the number of occurrences for each earthquake rupture in a source follows a *Poissonian distribution*

Is this true now that we have also the non parametric source?

### 2.1 Basic concepts

The Classical, Event-Based, and Disaggregation analysis requires the definition of two main components: the *seismic source model*, that is a collection of seismic sources describing the seismic activity in a region of interest, and the *ground motion model*, that is a mathematical relationship defining the probability distribution of a ground motion parameter at a site given the occurrence of an earthquake rupture.

The design of a seismic source model involves the specification of a number of sources whose main parameters are the geometry, constraining the earthquake rupture locations, and the *magnitude-frequency distribution*, defining the average annual occurrence rate over a magnitude range. A seismic source model (*SSM*) can be therefore defined as a set of  $I$  seismic sources ( $Src$ ):

$$SSM = \{Src_1, Src_2, \dots, Src_I\} \quad (2.1)$$

Chapter 3 provides a detailed description of the different source typologies supported by the OQ-engine. However, independently of the typology, in a PSHA each source undergoes a

discretization process which effectively generates a number of distinct earthquake ruptures. A generic  $i$ -th source defines therefore a set of  $J$  earthquake ruptures:

$$Src_i = \{Rup_{i1}, Rup_{i2}, \dots, Rup_{iJ}\} \quad (2.2)$$

By indicating with  $v_{ij}$  the average annual occurrence rate for the  $j$ -th rupture in the  $i$ -th source, the associated probability of occurring  $k$  times in a time span  $T$  can be written, accordingly with the Poissonian assumption, as:

$$P_{rup_{ij}}(k|T) = e^{-v_{ij}T} \frac{(v_{ij}T)^k}{k!} \quad (2.3)$$

## 2.2 Classical PSHA

The classical PSHA analysis allows calculating the probabilities of exceeding, at least once in a given time span, and at a given site, a set of ground motion parameter levels considering all possible earthquake ruptures defined in a seismic source model. Such a list of probability values is usually referred to as *hazard curve*.

We indicate with  $P(X \geq x|T)$  the probability that a ground-motion parameter  $X$  exceeds, at least once in a time span  $T$ , a level  $x$ .  $P(X \geq x|T)$  can be computed as 1 minus the probability that none of the sources is causing a ground motion exceedance. By assuming *independent sources*, the probability that none of the sources is causing an exceedance is equal to the product of the probabilities that each source does not cause an exceedance, that is:

$$\begin{aligned} P(X \geq x|T) &= 1 - P_{src1}(X < x|T) * P_{src2}(X < x|T) * \dots * P_{srcI}(X < x|T) \\ &= 1 - \prod_{i=1}^I P_{src_i}(X < x|T) \end{aligned} \quad (2.4)$$

where  $P_{src_i}(X < x|T)$  is the probability that the  $i$ -th source is not causing an exceedance and  $I$  is the total number of sources in the source model.

By further assuming each source generates *independent earthquake ruptures*, we can compute  $P_{src_i}(X < x|T)$  as the product of the probabilities that each rupture does not cause an exceedance, that is:

$$\begin{aligned} P_{src_i}(X < x|T) &= P_{rup_{i1}}(X < x|T) * P_{rup_{i2}}(X < x|T) * \dots * P_{rup_{iJ_i}}(X < x|T) \\ &= \prod_{j=1}^{J_i} P_{rup_{ij}}(X < x|T) \end{aligned} \quad (2.5)$$

where  $P_{rup_{ij}}(X < x|T)$  is the probability that the  $j$ -th rupture in the  $i$ -th source is not causing an exceedance and  $J_i$  is the total number of ruptures generated by the  $i$ -th source. Relying therefore on the assumptions of independent sources and independent earthquake ruptures generated by each source, we can write:

$$P(X \geq x|T) = 1 - \prod_{i=1}^I \prod_{j=1}^{J_i} P_{rup_{ij}}(X < x|T) \quad (2.6)$$

Intuitively, the fact that a rupture does not cause any exceedance in a given time span  $T$  can be due to the fact that the rupture does not occur at all or that the rupture occurs once but without causing an exceedance, or that the rupture occurs twice but both times without causing

an exceedance, and so on. Given that all these events are mutually exclusive, by using the total probability theorem we can write:

$$\begin{aligned} P_{rup_{ij}}(X < x|T) &= P_{rup_{ij}}(n = 0|T) + P_{rup_{ij}}(n = 1|T) * P(X < x|rup_{ij}) + \\ &\quad P_{rup_{ij}}(n = 2|T) * P(X < x|rup_{ij})^2 + \dots \\ &= \sum_{k=0}^{\infty} P_{rup_{ij}}(k|T) * P(X < x|rup_{ij})^k \end{aligned} \quad (2.7)$$

where  $P_{rup_{ij}}(k|T)$  is the probability that the  $j$ -th rupture in the  $i$ -th source is occurring  $k$  times in time span  $T$  and  $P(X < x|rup_{ij})$  is the conditional probability that parameter  $X$  is not exceeding level  $x$  given an occurrence of  $rup_{ij}$ .

By now assuming that the *number of occurrences of each rupture in a time span T follows a Poissonian distribution* we can place equation 2.3 in 2.7 and thus write:

$$\begin{aligned} P_{rup_{ij}}(X < x|T) &= \sum_{k=0}^{\infty} e^{-v_{ij}T} \frac{(v_{ij}T)^k}{k!} * P(X < x|rup_{ij})^k \\ &= e^{-v_{ij}T} \sum_{k=0}^{\infty} \frac{(v_{ij}T * P(X < x|rup_{ij}))^k}{k!} \end{aligned} \quad (2.8)$$

Making use of the property:

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \quad (2.9)$$

we can rewrite 2.8 as:

$$\begin{aligned} P_{rup_{ij}}(X < x|T) &= e^{-v_{ij}T} e^{v_{ij}T * P(X < x|rup_{ij})} \\ &= e^{-v_{ij}T * (1 - P(X < x|rup_{ij}))} \\ &= e^{-v_{ij}T * P(X \geq x|rup_{ij})} \end{aligned} \quad (2.10)$$

By now recognizing that, according to the Poissonian distribution, the probability of at least one occurrence (that is one or more) in a time span  $T$  of  $rup_{ij}$  is:

$$P_{rup_{ij}}(n \geq 1|T) = 1 - e^{-v_{ij}T} \quad (2.11)$$

we can write equation 2.10 as:

$$P_{rup_{ij}}(X < x|T) = (1 - P_{rup_{ij}}(n \geq 1|T))^{P(X \geq x|rup_{ij})} \quad (2.12)$$

By placing equation 2.12 in 2.6, we can write:

$$P(X \geq x|T) = 1 - \prod_{i=1}^I \prod_{j=1}^{J_i} (1 - P_{rup_{ij}}(n \geq 1|T))^{P(X \geq x|rup_{ij})} \quad (2.13)$$

Equation 2.13 is used by the OQ-engine for the calculation of hazard curves when performing Classical PSHA. To our knowledge, this equation has been first proposed by Field et al. (2003), derived from the traditional rate-based formulation converted in terms of probabilities (their equation A8). Instead, we derive it from the assumptions of a source model consisting of independent sources, independent earthquake ruptures generated by each source, and ruptures obeying to a Poissonian temporal occurrence model.

### 2.2.1 Equivalence with the rate-based equation

It is worth noticing how equation 2.13 is equivalent to the more traditional rate-based hazard equation (McGuire, 1995). Indeed, by assuming ground motion occurrence to follow a Poissonian distribution in time, and indicating with  $v$  the mean annual rate of exceeding a ground motion level  $x$ , we can write:

$$P(X \geq x|T) = 1 - e^{-vT} \quad (2.14)$$

We can also rewrite 2.13 as:

$$\begin{aligned} P(X \geq x|T) &= 1 - \prod_{i=1}^I \prod_{j=1}^{J_i} (1 - P_{rup_{ij}}(n \geq 1|T))^{P(X \geq x|rup_{ij})} \\ &= 1 - \prod_{i=1}^I \prod_{j=1}^{J_i} e^{-v_{ij} T * P(X \geq x|rup_{ij})} \\ &= 1 - e^{-\sum_{i=1}^I \sum_{j=1}^{J_i} v_{ij} T * P(X \geq x|rup_{ij})} \end{aligned} \quad (2.15)$$

The equivalence between equations 2.14 and 2.15 is possible if and only if:

$$v = \sum_{i=1}^I \sum_{j=1}^{J_i} v_{ij} * P(X \geq x|rup_{ij}) \quad (2.16)$$

Assuming now, for the sake of simplicity, that a rupture is completely characterized by magnitude and distance from a site, we can write the rate of occurrence of the  $j$ -th rupture as:

$$v_{ij} = v_i * f_i(m, r) \quad (2.17)$$

where  $v_i$  is the total occurrence rate for the  $i$ -th source, and  $f_i(m, r)$  is the probability, for the  $i$ -th source, of generating a rupture of magnitude  $m$  and distance  $r$ . By placing 2.17 in 2.16 and by replacing the discrete summation over ruptures with a continuous integral over magnitude and distance, we can write:

$$v = \sum_{i=1}^I v_i \iint f_i(m, r) P(X \geq x|m, r) dm dr \quad (2.18)$$

which is the traditional equation for calculating ground motion exceedance rates (McGuire, 1995).

### 2.3 Event-based PSHA

The goal of an Event-based PSHA is to simulate seismicity in a region as described by a source model and to simulate ground shaking on a set of locations accordingly with a ground motion model. In both cases, simulation involves a Monte Carlo (i.e. random) sampling procedure.

Seismicity is simulated by generating a *stochastic event set* (also known as *synthetic catalog*) for a given time span  $T$ . For each rupture generated by a source, the number of occurrences in a time span  $T$  is simulated by sampling the corresponding probability distribution as given in equation 2.3. A stochastic event set is therefore a *sample* of the full population of ruptures as defined by a seismic source model. Each rupture is present zero, one or more times, depending on its probability. Symbolically, we can define a stochastic event set (*SES*) as:

$$SES(T) = \{k \times rup, k \sim P_{rup}(k|T) \ \forall rup \text{ in } Src \ \forall Src \text{ in } SSM\} \quad (2.19)$$

where  $k$ , the number of occurrences, is a random sample of  $P_{rup}(k|T)$ , and  $k \times rup$  means that rupture  $rup$  is repeated  $k$  times in the stochastic event set.

Given an earthquake rupture, the simulation of ground shaking values on a set of locations ( $\mathbf{x} = (x_1, x_2, \dots, x_N)$ ) forms a *ground motion field*. In a Event-based PSHA, for each rupture in a stochastic event set, the ground motion field is obtained by sampling the probability distribution defined by the ground motion model. As described in Chapter 4, the ground motion distribution at a site is assumed to be a Normal distribution. The aleatory variability is described in terms of an *inter-event* (also known as *between-events*) standard deviation ( $\tau$ ) and intra-event (also known as *within-event*) standard deviation ( $\sigma$ ). The simulation of a ground motion field is therefore the result of the summation of three terms, the logarithmic mean of the ground motion distribution:

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_N) \quad (2.20)$$

the inter-event variability:

$$\boldsymbol{\eta} = (\eta, \eta, \dots, \eta), \text{ where } \eta \sim N(0, \tau) \quad (2.21)$$

and the intra-event variability:

$$\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N) \sim N(\mathbf{0}, \boldsymbol{\Sigma}) \quad (2.22)$$

where:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} & \cdots & \sigma_1\sigma_N\rho_{1N} \\ \sigma_2\sigma_1\rho_{21} & \sigma_2^2 & \cdots & \sigma_2\sigma_N\rho_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_N\sigma_1\rho_{N1} & \sigma_N\sigma_2\rho_{N2} & \cdots & \sigma_N^2 \end{bmatrix} \quad (2.23)$$

It is worth noticing how the inter-event variability, uniform for all sites given an earthquake rupture, is drawn from an univariate normal distribution of mean 0 and standard deviation  $\tau$ , while the intra-event variability is a random sample of a multivariate normal distribution of mean  $\mathbf{0}$  and covariance matrix  $\boldsymbol{\Sigma}$ .  $\boldsymbol{\Sigma}$  is a diagonal matrix when considering no correlation in the intra-event variability, while it has non-zero off-diagonal elements when considering a correlation model ( $\rho$ ) for the intra-event aleatory variability.

### 2.3.1 Calculation of hazard curves from ground motion fields

The ground motion fields simulated for each rupture in a stochastic event set can be used to compute hazard curves. Indeed, indicating with  $T_0$  the duration associated with a stochastic event set, and with  $K$  the number of ground motion fields (and associated ruptures) simulated in time  $T_0$ , we can compute the rate of exceedance of a ground motion level  $x$  at a site as (Ebel and Kafka, 1999):

$$v = \frac{\sum_{k=1}^K H(x_k - x)}{T_0} \quad (2.24)$$

where  $H$  is the Heaviside function and  $x_k$  is the ground motion parameter value at the considered site associated with the  $k$ -th ground motion field. The exceedance rate obtained from equation 2.24 can then be used to compute the probability of at least one occurrence in any time span  $T$ , accordingly with the Poissonian distribution, using equation 2.14.

As the stochastic event set duration  $T_0$  increases, equation 2.24 provides an increasingly more accurate estimate of the actual rate of exceedance. A larger  $T_0$  can be achieved not only by simulating a single stochastic event set with longer duration, but also by simulating multiple stochastic event sets. These can then be joined together to form a stochastic event which has a large enough duration to provide a stable estimate of the rates of ground motion exceedance.

## 2.4 Disaggregation

The disaggregation analysis allows investigating how the different earthquake ruptures defined in a source model contribute to the probability of exceeding a certain ground motion level  $x$  at a given site. Given the very large number of earthquake ruptures associated with a source model, contributions cannot be investigated on a rupture by rupture basis but a classification scheme is used instead. Ruptures are classified in terms of the following parameters:

- magnitude ( $M$ )
- distance to rupture surface-projection (Joyner-Boore distance) ( $r_{jb}$ )
- longitude and latitude of rupture surface-projection closest point ( $\lambda, \phi$ )
- tectonic region type ( $TRT$ )

For each earthquake rupture, the associated conditional probability of ground motion exceedance –  $P(X \geq x|T, rup)$  – is computed for different  $\varepsilon$  bins, where  $\varepsilon$  is the difference, in terms of number of total standard deviations, between  $x$  and the mean ground motion  $\mu$  as predicted by the ground motion model, that is:

$$\varepsilon = \frac{x - \mu}{\sigma_{total}} \quad (2.25)$$

The disaggregation in terms of  $\varepsilon$  allows investigating how the different regions of the ground motion distributions contribute to the probability of exceedance.

The rupture parameters ( $M, r_{jb}, \lambda, \phi, TRT$ ) together with the  $\varepsilon$  parameter effectively create a 6-dimensional model space which, discretized into a number of bins, is used to classify the probability of exceedance for different combination of rupture parameters and  $\varepsilon$  values.

For a given model space bin  $\mathbf{m} = (M, r_{jb}, \lambda, \phi, TRT, \varepsilon)$  the probability of exceeding level  $x$  at least once in a time span  $T$  is computed using equation 2.13, that is:

$$P(X > x|T, \mathbf{m}) = 1 - \prod_{i=1}^I \prod_{j=1}^{J_i} \begin{cases} (1 - P_{rup_{ij}}(n \geq 1|T))^{P(X \geq x|rup_{ij}, \varepsilon)} & \text{if } rup_{ij} \in \mathbf{m} \\ 1 & \text{if } rup_{ij} \notin \mathbf{m} \end{cases} \quad (2.26)$$

In other words, if a rupture belongs to the considered bin, then the probability of not causing a ground motion exceedance is computed according to equation 2.12, otherwise the probability is 1 (that is, given that the rupture does not belong to the bin it can never cause a ground motion exceedance).

### 2.4.1 Disaggregation histograms

Disaggregation values as given by equation 2.26 can be aggregated in order to investigate earthquake rupture contributions over a reduced model space. The following disaggregation histograms are provided by the OpenQuake-engine.

Magnitude disaggregation:

$$P(X > x|T, M) = 1 - \prod_{r_{jb}} \prod_{\lambda} \prod_{\phi} \prod_{TRT} \prod_{\varepsilon} (1 - P(X > x|T, \mathbf{m})) \quad (2.27)$$

Distance disaggregation:

$$P(X > x|T, r_{jb}) = 1 - \prod_M \prod_{\lambda} \prod_{\phi} \prod_{TRT} \prod_{\varepsilon} (1 - P(X > x|T, \mathbf{m})) \quad (2.28)$$

Tectonic region type disaggregation:

$$P(X > x|T, TRT) = 1 - \prod_M \prod_{r_{jb}} \prod_{\lambda} \prod_{\phi} \prod_{\varepsilon} (1 - P(X > x|T, \mathbf{m})) \quad (2.29)$$

Magnitude-Distance disaggregation:

$$P(X > x|T, M, r_{jb}) = 1 - \prod_{\lambda} \prod_{\phi} \prod_{TRT} \prod_{\epsilon} (1 - P(X > x|T, \mathbf{m})) \quad (2.30)$$

Magnitude-Distance-Epsilon disaggregation:

$$P(X > x|T, M, r_{jb}, \epsilon) = 1 - \prod_{\lambda} \prod_{\phi} \prod_{TRT} (1 - P(X > x|T, \mathbf{m})) \quad (2.31)$$

Longitude-Latitude disaggregation:

$$P(X > x|T, \lambda, \phi) = 1 - \prod_M \prod_{r_{jb}} \prod_{TRT} \prod_{\epsilon} (1 - P(X > x|T, \mathbf{m})) \quad (2.32)$$

Longitude-Latitude-Magnitude disaggregation:

$$P(X > x|T, \lambda, \phi, M) = 1 - \prod_{r_{jb}} \prod_{TRT} \prod_{\epsilon} (1 - P(X > x|T, \mathbf{m})) \quad (2.33)$$

Longitude-Latitude-Tectonic Region Type disaggregation:

$$P(X > x|T, \lambda, \phi, TRT) = 1 - \prod_M \prod_{r_{jb}} \prod_{\epsilon} (1 - P(X > x|T, \mathbf{m})) \quad (2.34)$$

All the above equations are based on the assumption that earthquake ruptures in different bins are independent, therefore probabilities can be aggregated by using the multiplication rule for independent events. The probability of a ground motion exceedance over a reduced model space is computed simply as 1 minus the probability of non-exceedance over the remaining model space dimensions.

#### 2.4.2 Comparison between OpenQuake-engine disaggregation and traditional disaggregation

The traditional disaggregation analysis as commonly known in literature (e.g. Bazzurro and Cornell, 1999) differs from the one provided by the OpenQuake-engine. Indeed, a disaggregation analysis typically provides the conditional probability of observing an earthquake scenario of given properties (magnitude, distance, epsilon, ...) given that a ground motion exceedance is occurred, which can be written (following the same notation used in this chapter) as:

$$P(\mathbf{m}|X > x) \quad (2.35)$$

On the contrary, the OpenQuake-engine (as described in equation 2.26) provides the conditional probability of observing at least one ground motion exceedance in a time span  $T$  given the occurrence of earthquake ruptures of given properties  $\mathbf{m}$ , that is:

$$P(X > x|T, \mathbf{m}) \quad (2.36)$$

The probabilities given in equations 2.35 and 2.36 are clearly different. Indeed, for different  $\mathbf{m}$ , values given by equation 2.35 must sum up to 1, while this is not the case for equation 2.36. For the former equation different  $\mathbf{m}$  represent mutually exclusive events, while for the latter they represent independent events.

Using the Poissonian assumption it is possible however to derive equation 2.35 from equation 2.36. Indeed, indicating with  $v_{\mathbf{m}}$  the rate of ground motion exceedance ( $X > x$ ) associated with

earthquake ruptures of properties  $\mathbf{m}$  and with  $v$  the rate of ground motion exceedance associated with all earthquake ruptures, we can write equation 2.35 as:

$$P(\mathbf{m}|X > x) = \frac{v_{\mathbf{m}}}{v} \quad (2.37)$$

By solving the Poissonian equation 2.14 for the rate of exceedance, we can write  $v_{\mathbf{m}}$  as:

$$v_{\mathbf{m}} = -\frac{\ln(1 - P(X > x|T, \mathbf{m}))}{T} \quad (2.38)$$

$v$  can be obtained using the same equation above but considering  $P(X > x|T)$  instead of  $P(X > x|T, \mathbf{m})$ , where  $P(X > x|T)$  is obtained by aggregating, using the multiplicative rule, the probabilities over the different  $\mathbf{m}$ , that is:

$$P(X > x|T) = 1 - \prod_{\mathbf{m}} (1 - P(X > x|T, \mathbf{m})) \quad (2.39)$$

By computing  $v_{\mathbf{m}}$  and  $v$  from  $P(X > x|T, \mathbf{m})$  it is hence possible to obtain the more traditional disaggregation results as given in equation 2.35.

## Basic concepts

### The Point and Area sources

#### The Simple Fault source

The rupture floating algorithm for a Simple Fault source

#### The Complex Fault source

Mesh construction in a Complex Fault source

The rupture floating algorithm for a Complex Fault source

#### The Characteristic Fault source

## 3. Seismic Source Models

This chapter describes the seismic source typologies supported by the OQ-engine: *Point* and *Area* sources for modeling distributed seismicity, *Simple Fault*, *Complex Fault*, and *Characteristic Fault* for modeling fault-based seismicity.

### 3.1 Basic concepts

The OQ-engine provides several seismic source typologies to accommodate different modeling approaches. For modeling distributed seismicity, that is seismic activity occurring over a geographical region and not tied to specific well characterized fault structures, the OQ-engine provides the *Point* and *Area* sources. The former defines seismic activity *nucleating in a single geographical location*, while the second seismicity occurring *uniformly over a geographical region*. Both sources define a seismogenic layer which constrains rupture location and extension along depth. A collection of Point sources can be used to model seismicity with spatially variable parameters (as obtained from a smoothed seismicity approach for instance), while Area sources can be used to model seismicity in geographic zones usually defined by expert judgments taking into account seismological, geological and geodetic information. Both source typologies allow for the modeling of earthquake ruptures as extended surfaces (that is as rectangular planes) with potentially multiple orientations and inclinations (that is multiple strike and dip angles) and also placed at different depth levels. Earthquake ruptures can extend without barriers along strike but cannot cross the seismogenic layer.

For modeling fault-based seismicity, that is seismic activity occurring on a well identified and characterized fault zone, the OQ-engine provides two main options, the *Simple Fault* and the *Complex Fault* sources. Both source typologies distribute seismicity *uniformly over a fault surface*, with the only constraint that an earthquake rupture cannot extend outside of the defined fault surface. In both sources an earthquake of a given magnitude is defined as a portion of the fault surface. To simulate all possible rupture locations, an earthquake rupture is moved, or *floated*, over the entire fault surface. The two source typologies differ instead in terms of the geometrical complexity they can accommodate when modeling a fault surface. In particular, a Simple Fault source can model a fault surface as a plain rectangle in the simplest case, or as a set of connected parallelograms in the most complex case. A Complex Fault source can instead model an arbitrarily complex quadrilateral surface, which can therefore accommodate changes in dip angle along depth or along strike or changes in fault width. The Complex Fault source is

therefore particularly suitable for modeling large subduction interface faults, while Simple Fault sources can be used for modeling crustal faults. The *Characteristic Fault* source can model both simple and complex geometries, but instead of simulating floating ruptures, each earthquake breaks the entire fault surface independently of the associated magnitude. The Characteristic Source typology can be used to model individual faults or fault segments that tend to produce essentially same size earthquakes (Schwartz and Coppersmith, 1984).

Independently of the typology, all sources require the definition of the following main parameters:

- magnitude-frequency distribution
- tectonic region type
- magnitude-area scaling relationship (for all but the Characteristic Fault source)
- rupture-aspect ratio (for all but the Characteristic Fault source)
- upper seismogenic depth
- lower seismogenic depth

The magnitude-frequency distribution defines the total moment rate released by a source. It is therefore a key parameter which determines the influence of a source in a PSHA. The OQ-engine supports the definition of the traditional *double-truncated Gutenberg-Richter* magnitude-frequency distribution which is widely used in PSHA. Required parameters are the cumulative *a-value* (defined as the intercept of the *cumulative* distribution at  $M = 0$  in a  $\log_{10}$  scale), the *b-value*, the minimum and the maximum magnitudes. However, to accommodate also other possible parametric (and non-parametric) forms, the OQ-engine provides a generic discrete *Incremental* magnitude-frequency distribution defined through a list of annual occurrence rates, associated to an equally spaced set of magnitude values.

In a PSHA, the magnitude-frequency distribution is subject to a discretization process which defines a set of equally spaced magnitude bins and the associated annual occurrence rates (for an *Incremental* magnitude-frequency distribution such a discretization is part of its definition). For each source, the annual occurrence rate associated to a magnitude bin is *uniformly* distributed over all the ruptures associated to the same magnitude bin value. In other words, while the particular geometry of a source determines location and number of ruptures of a given magnitude, the occurrence rate (and thus the occurrence probability) is uniform over ruptures with the same magnitude.

The second typology-independent parameter associated to a seismic source is the tectonic region type. This is a source attribute that is used as a key to associate a seismic source to a ground motion model. Given a source model describing seismicity occurring in a region where different tectonic settings overlap or are close to each other, the associated ground motion model may prescribe different equations for the different tectonic settings. The mapping between a ground motion model equation and a seismic source is therefore achieved through the tectonic region type attribute.

All sources model earthquake ruptures as extended surfaces. With the only exception of the Characteristic Fault source, the area of an earthquake rupture surface is magnitude dependent. To constrain the rupture surface area, the Point, Area, Simple and Complex Fault sources require the definition of a *magnitude-area* scaling relationship. Together with a *rupture aspect ratio* (defined as ratio between length and width), the rupture extension and shape (assumed rectangular) is completely defined. Indeed, indicating with  $A$  the rupture area and with  $ar$  the rupture aspect ratio, rupture length ( $L$ ) and width ( $W$ ) can be computed as:

$$\begin{aligned} L &= \sqrt{A * ar} \\ W &= \sqrt{A / ar} \end{aligned} \tag{3.1}$$

In all sources, the rupture aspect ratio is used to constrain the initial rupture shape. However, if

**Table 3.1 – Table summarizing parameters, and their functions, required for the definition of Area and Point sources in the OQ-engine**

Parameter	Purpose
Magnitude frequency distribution	Defines total moment rate
Magnitude area scaling relationship	Define sizes and shapes of rupture planes
Rupture aspect ratio (length / width)	
Nodal plane distribution (each nodal plane being defined by strike, dip, rake)	Defines orientations and faulting styles of ruptures
Hypocentral depth distribution	Defines centroids of rupture planes
Upper seismogenic depth	Constrains rupture planes inside seismogenic layer
Lower seismogenic depth	

this conflicts with other source-dependent geometrical constrains, the rupture is reshaped so as to conserve the area as given by the scaling relationship.

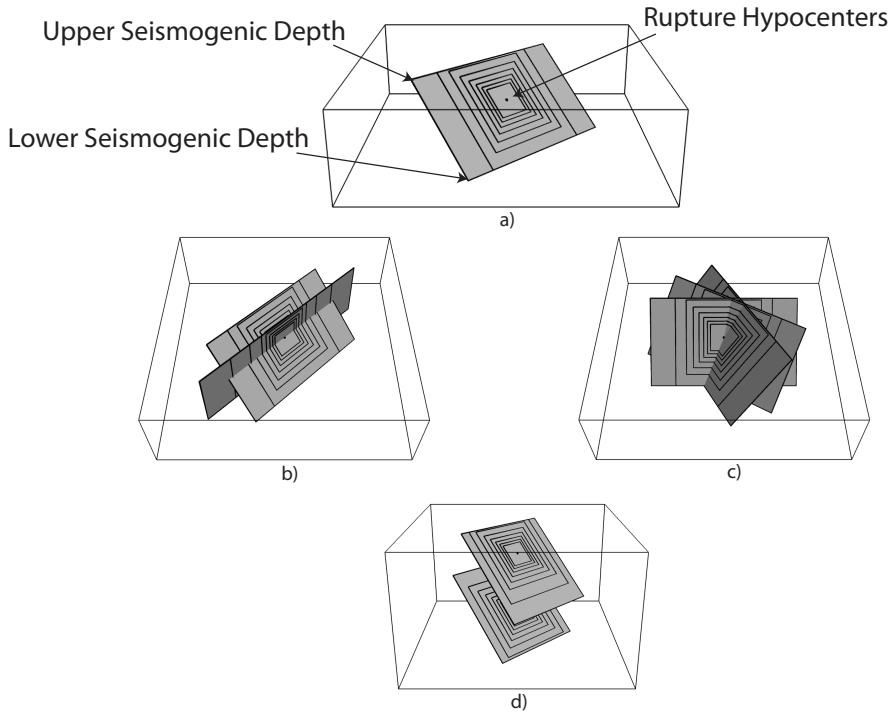
The upper and lower seismogenic depths define the seismogenic layer, that is the depth range over which earthquake ruptures can occur. The definition of a seismogenic layer is required to avoid an uncontrolled extension of the earthquake ruptures along depth which can lead, especially for large magnitude events, to unrealistic scenarios. The definition of a seismogenic layer thickness effectively induces a magnitude-dependent rupture aspect ratio. Indeed, as the rupture size increases with increasing magnitude values, the rupture width reaches the maximum allowed width, and the rupture aspect ratio starts deviating, that is increasing, from the original value.

## 3.2 The Point and Area sources

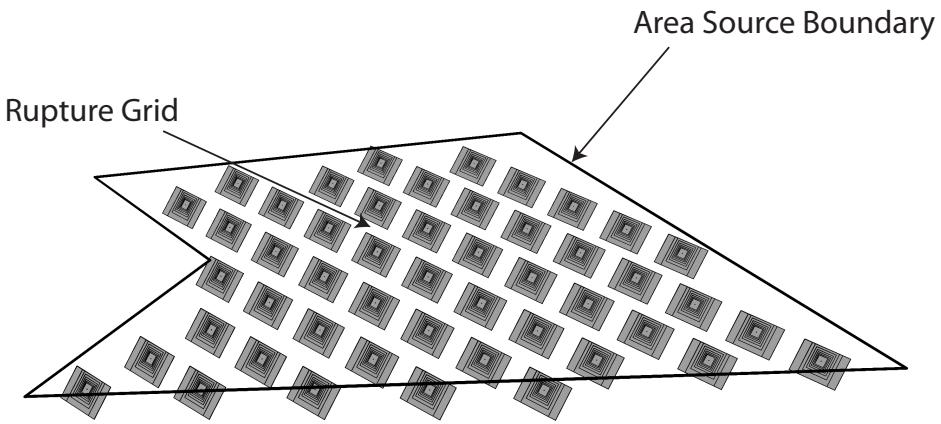
The parameters specific to the definition of Point and Area sources, and their associated function, are listed in Table 3.1. Sources are parameterized so that earthquake ruptures are modeled as rectangular planes. In a point-source representation (Figure 3.1) ruptures are generated underneath a single geographical location, and can be potentially distributed over multiple orientations, faulting styles, and depth levels. Rupture centroids are co-located with the point-source location and are positioned at depths specified by the hypocentral depth distribution. Rupture shapes follow the given aspect ratio. However, if for a given aspect ratio and hypocentral depth the rupture plane crosses either boundary (upper or lower) of the seismogenic layer, the plane is shifted along the dip direction so as to fit within the upper and lower seismogenic depths. As a consequence, the hypocentral location no longer corresponds with the plane centroid. If this adjustment is insufficient to avoid crossing either boundary of the seismogenic layer, the plane is reshaped; the width becomes the maximum allowed by the seismogenic layer thickness, and the length is increased so as to conserve rupture area (at the expense of the aspect ratio).

In an area source (Figure 3.2), earthquake ruptures are distributed over a regular grid (equally-spaced in distance) covering a geographical region as defined by a seismic zone. Generation of ruptures follows the same algorithm as for point sources.

For both sources, the rate associated to each rupture plane is the original rate associated to



**Figure 3.1 – Graphical representation of the earthquake ruptures as generated by a Point Source.** a) Given a geographical location on the Earth surface, ruptures are generated underneath according to a scaling relationship and aspect ratio value and forced to not exceed the upper and lower seismogenic depths. Ruptures can be distributed over multiple dips b), strikes c) and hypocentral depths d).



**Figure 3.2 – Earthquake ruptures generated by an area source in the OQ-engine.** Ruptures are distributed uniformly over a regular grid within the area. In this plot, for better visualization, ruptures are modeled only according to a single nodal plane and hypocentral depth, but actual calculations may involve multiple orientations and hypocentral depths. Ruptures originating from different grid nodes may also overlap and cross each other.

**Table 3.2 – Table summarizing parameters, and their functions, required for the definition of a Simple Fault source in the OQ-engine**

Parameter	Purpose
Magnitude frequency distribution	Defines total moment rate
Magnitude area scaling relationship	Define sizes and shapes of rupture planes
Rupture aspect ratio (length / width)	
Fault trace	Define fault surface
Upper seismogenic depth	
Lower seismogenic depth	
Dip angle	
Rake angle	Defines faulting style

the corresponding magnitude bin, scaled by the location weight (1 for a point source and 1/N for an area source, where N is the total number of grid points in the area), the nodal plane (that is orientation and faulting style) weight, and the hypocentral depth weight.

For an area source, the boundary is assumed ‘leaky’, that is earthquake ruptures can extend out of it. Because of rupture area conservation, earthquake surfaces associated to large magnitudes can extend well beyond the source boundaries. If the rupture orientation is considered random then this behavior can potentially lead to unrealistic scenarios, that is earthquake ruptures that are not consistent with the area geometry and the tectonic feature it is meant to represent. The design of an area source requires therefore a careful estimation not only of the associated activity rates but also of the predominant faulting orientations.

The OpenQuake-engine does not currently provide the possibility to define non-leaky boundaries. The main difficulty in the implementation of such a feature is the definition of a clear algorithm specifying how hard boundaries would influence the generation of earthquake ruptures within the area source. Several options are available. The easiest approach would be to remove, from the set of generated ruptures, the ones that extend outside of the boundary. This approach requires however a careful calculation of the occurrence rates to be assigned to the earthquake ruptures. These cannot be calculated anymore *a priori* (that is from the number of grid points in the area source), but only after all the ruptures have been generated and the ones crossing the boundary excluded. Additionally, the removal of ruptures may also introduce a non-uniform hazard pattern within the area source. An alternative approach would be to truncate earthquake rupture surfaces that extend outside of the area boundary. However, without a careful analysis of the consistency between the main rupture orientations and source geometry, this approach may potentially lead to large magnitude events developing over very small rupture surfaces. A third approach would be to adjust the earthquake orientation/location so that the rupture surface does not extend beyond the area boundary. Depending on the source geometry, such an adjustment may not be always possible (that is, there may no be an orientation/location which allows a rupture to fully lie within the area source). This last strategy can be seen as a way to minimize the rupture extension outside of the area source.

### 3.3 The Simple Fault source

Parameters required for the definition of a Simple Fault source are given in Table 3.2. The fault surface is constructed by translating the fault trace (defined as the intersection between the fault

surface and the Earth's surface) from the upper to the lower seismogenic depth along a direction perpendicular to the fault trace (measured as the difference between the azimuths of the first and last coordinates of the trace) and with an inclination equal to the dip angle. The surface so defined is effectively modeled as a regular (i.e. equally spaced in distance) mesh (Figure 3.3a). For each magnitude bin defined in the magnitude-frequency distribution, an earthquake rupture is modeled as a portion of the fault surface, accordingly with the magnitude scaling relationship and the rupture aspect ratio (Figure 3.3b). To simulate all possible rupture locations, each earthquake rupture is *floated*, that is moved, along both the strike and dip directions (Figure 3.3c). The floating step is assumed equal to the mesh discretization step. The occurrence rate associated to a given magnitude bin is distributed uniformly over all the ruptures associated with the same magnitude value.

### 3.3.1 The rupture floating algorithm for a Simple Fault source

We describe here in more detail the algorithm adopted for modeling floating ruptures in a Simple Fault source. We indicate with  $\Delta$  the mesh spacing, and with  $n_{strike}^{fault}$  and  $n_{dip}^{fault}$  the number of nodes along the strike and dip directions in the mesh representing the entire fault surface. By indicating with  $L(M)$  and  $W(M)$  the length and width of a rupture of magnitude  $M$  (obtained from equation 3.1), the equivalent number of nodes (along strike and dip) representing a rupture on the mesh can be computed as:

$$\begin{aligned} n_{strike}^{rup}(M) &= L(M)/\Delta + 1 \\ n_{dip}^{rup}(M) &= W(M)/\Delta + 1 \end{aligned} \quad (3.2)$$

By further assuming that a rupture is floated along strike and dip with a step equal to the mesh spacing ( $\Delta$ ), we can compute the total number of ruptures along strike and dip as:

$$\begin{aligned} N_{rup}^{strike}(M) &= n_{strike}^{fault} - (n_{strike}^{rup}(M) - 1) \\ N_{rup}^{dip}(M) &= n_{dip}^{fault} - (n_{dip}^{rup}(M) - 1) \end{aligned} \quad (3.3)$$

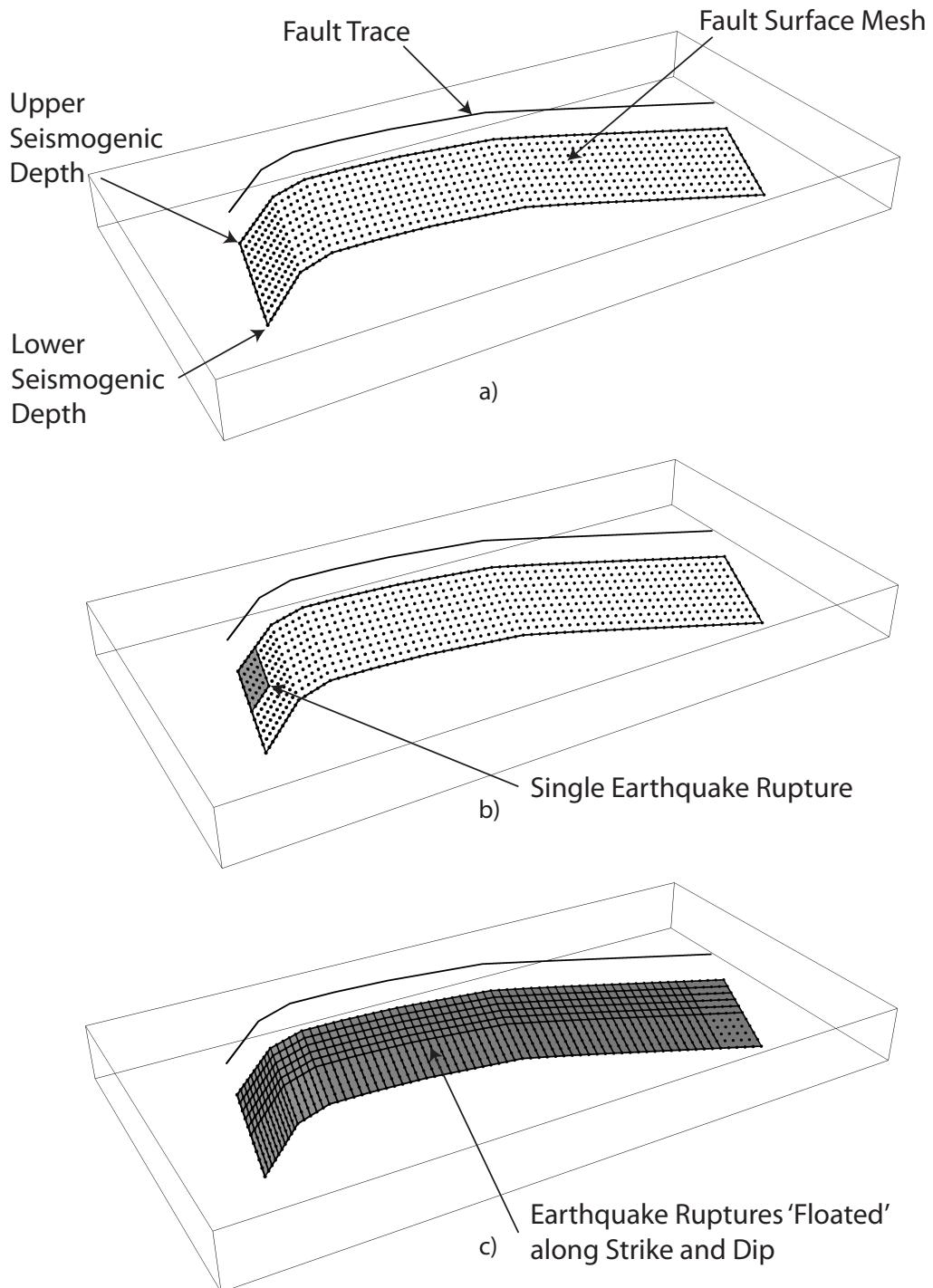
That is a rupture can propagate until its boundary reaches the fault boundary, but not beyond. Therefore the total number of possible rupture locations along a certain dimension is equal to the total number of nodes minus the number of nodes required by the rupture reduced by 1, which represents the number of positions that a rupture cannot occupy because it would otherwise extend, at least by one mesh spacing, outside of the fault boundary. Indicating with  $v(M)$  the annual occurrence rate as given by the magnitude-frequency distribution, the annual occurrence rate associated to each rupture of magnitude  $M$  is given by:

$$v_{rup}(M) = \frac{v(M)}{N_{rup}^{strike}(M) * N_{rup}^{dip}(M)} \quad (3.4)$$

The occurrence rate scaling factor is therefore magnitude dependent (in contrast with the Area Source where the scaling is constant for all magnitudes).

### 3.4 The Complex Fault source

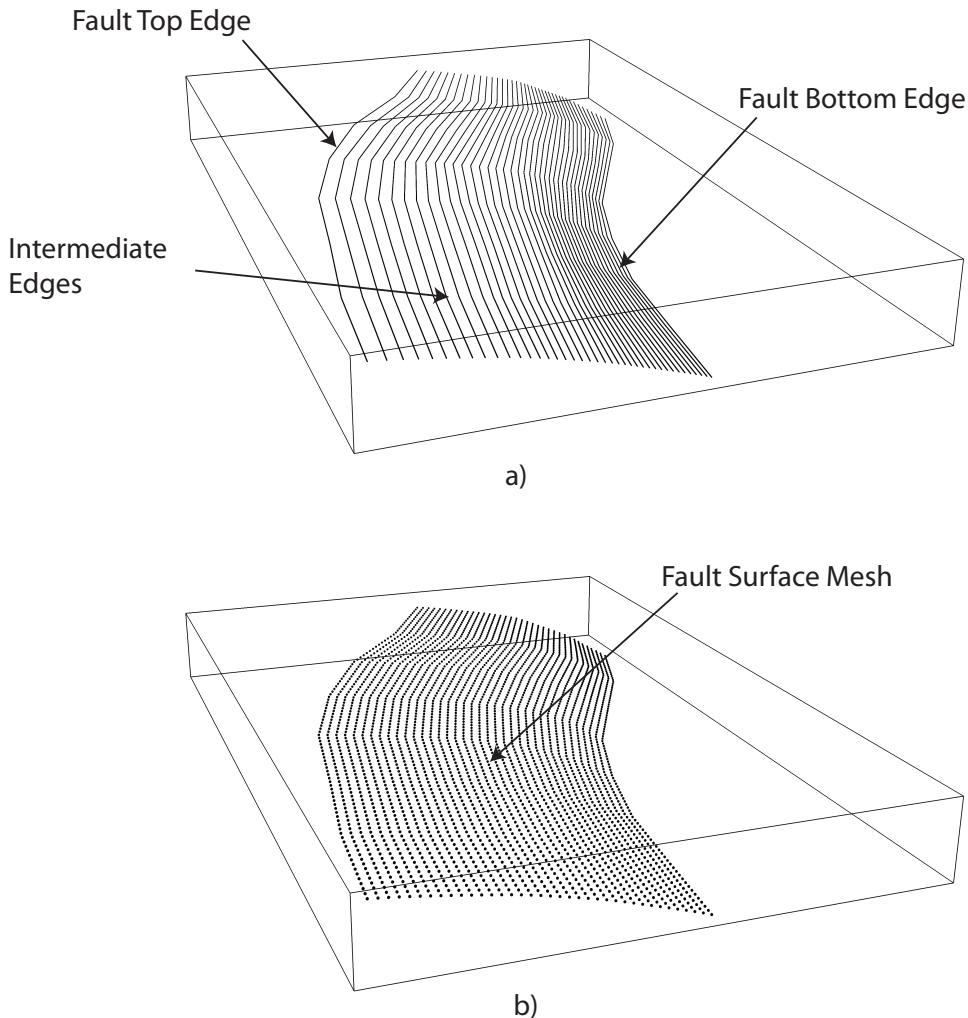
Parameters required for the definition of a Complex Fault source are given in Table 3.3. To accommodate the definition of irregular geometries, the Complex Fault source requires the specification of the coordinates of, at least, the top and bottom edges of the fault surface, and optionally, of one or more intermediate edges (Figure 3.4a). Edges can have different and variable directions and a single edge can develop over different depth levels. This gives a very large flexibility in the definition of the fault surface, allowing for changes in width and inclination both along strike and along dip.



**Figure 3.3 – Simple Fault source in the OQ-engine.** The fault surface is obtained by translating the fault trace from the Earth's surface to the lower seismogenic depth with an inclination equal to the dip angle. The upper seismogenic depth delimits the fault top edge. A mesh representation of the fault surface is then constructed a). An earthquake rupture is defined as a portion of the fault surface b), and all possible rupture locations are simulated by floating the rupture surface both along strike and along dip c)

**Table 3.3 – Table summarizing parameters, and their functions, required for the definition of a Complex Fault source in the OQ-engine**

Parameter	Purpose
Magnitude frequency distribution	Defines total moment rate
Magnitude area scaling relationship	Define sizes and shapes of rupture planes
Rupture aspect ratio (length / width)	
Top edge	Define fault surface
Intermediate edges (optional)	
Bottom edge	
Rake angle	Defines faulting style



**Figure 3.4 – Complex Fault source**

### 3.4.1 Mesh construction in a Complex Fault source

The fault edges are used to construct a mesh representing the fault surface (Figure 3.4b). The mesh is, in general, not uniform; that is, the mesh spacing may be spatially variable to accommodate the irregular geometry of the surface. The construction of the mesh relies on the following algorithm. By indicating with  $\bar{L}_{edge}$  the average edge length, and with  $\Delta$  the desired mesh spacing, the number of mesh points along the strike direction is computed as:

$$n_{strike} = \bar{L}_{edge}/\Delta + 1 \quad (3.5)$$

Each edge is then resampled into  $n_{strike}$  points. By connecting the different edges along nodes that are on the same positions, dipping lines can be constructed. By indicating with  $\bar{W}_{fault}$  the average fault width computed from the set of dipping lines, the number of mesh points along dip is computed as:

$$n_{dip} = \bar{W}_{fault}/\Delta + 1 \quad (3.6)$$

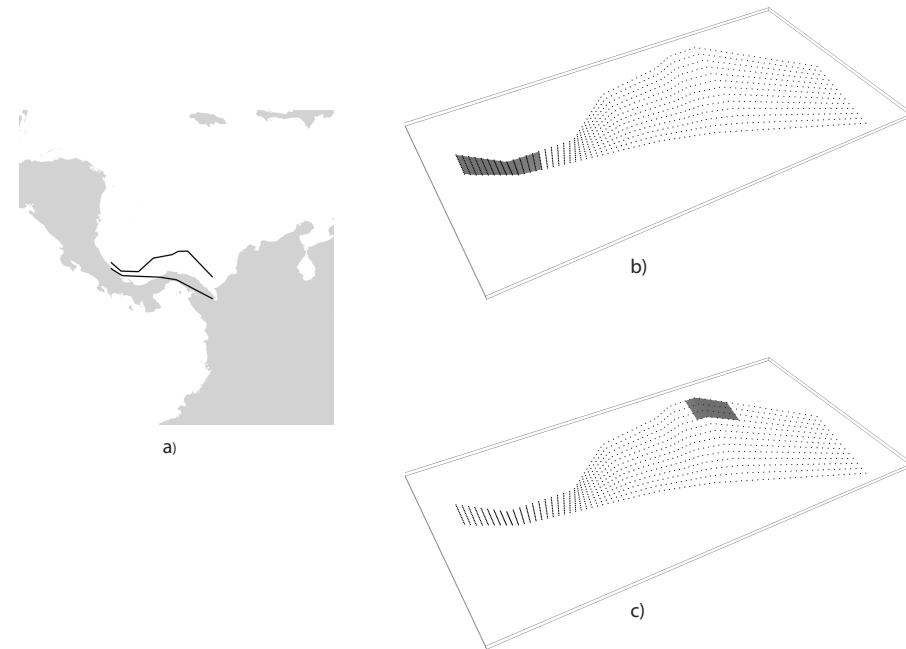
Each dipping line is then resampled into  $n_{dip}$  points. This completes the construction of the mesh which is therefore represented by  $n_{strike} \times n_{dip}$  nodes. It is worth noticing how the resampling of the edges as well as of the dipping lines allows the construction of a rectangular mesh which is however non-uniform. The actual mesh spacing varies from values smaller than  $\Delta$ , in regions where the fault length or width is smaller than the corresponding average, to values larger than  $\Delta$  where the opposite occurs.  $\Delta$ , which is basically used to compute the number of nodes along strike and dip, should therefore be considered as an *average* mesh spacing.

### 3.4.2 The rupture floating algorithm for a Complex Fault source

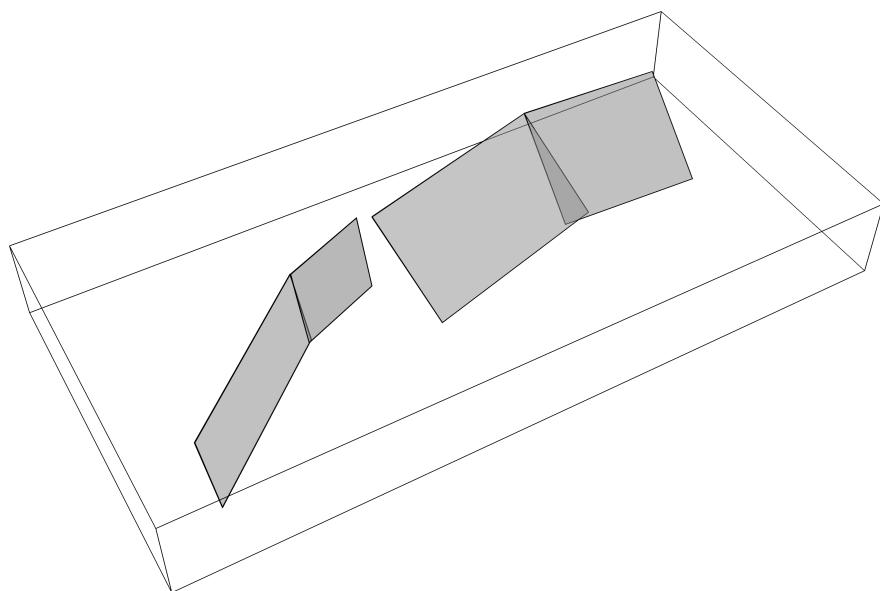
The non-uniformity of the mesh representing the fault surface makes the rupture floating algorithm for a Complex Fault source more problematic than for a Simple Fault source. Indeed, because of the variation of the actual mesh spacing, it is not possible to rely on the mesh spacing  $\Delta$  to compute the number of nodes required by a rupture of a given magnitude (that is equation 3.2). For each possible rupture location, an optimization procedure is used instead which identifies the number of nodes along strike and dip which gives a mesh surface with an area that is the closest to the one predicted by the magnitude area scaling relationship. In other words, the rupture mesh is represented by a number of nodes which is not constant but that may vary along the fault surface. In this context, the rupture aspect ratio is used to define the length of the rupture top edge, while the rupture width results from the optimization procedure. Such optimization procedure is well exemplified in Figure 3.5. The fault surface represents a southward dipping subduction fault located north of Panama defined in the model for South America developed by (Petersen et al., 2010). The figures shows how a  $M = 7.7$  event is modeled in two different parts of the fault surface. Where the fault is narrow (the western part) the rupture mesh contains a large number of nodes, separated by a small distance. Where the fault is large (the eastern part), the rupture mesh contains fewer nodes but separated by a larger distance.

## 3.5 The Characteristic Fault source

The *Characteristic Fault* source is meant to represent individual faults or fault segments that tend to produce earthquakes (Schwartz and Coppersmith, 1984) of essentially the same size. To offer the greatest flexibility in the definition of the associated geometry, a Characteristic Fault can be defined in terms of a simple fault geometry or as a complex fault geometry. A third option is available, that is a collection of planar ruptures, which can be used to model multi-segment ruptures for instance (Figure 3.6). In a Characteristic fault, earthquake ruptures always break the entire fault surface, therefore the rupture floating mechanism is not needed.



**Figure 3.5 – a)** Example of Complex Fault source representing subduction interface fault in North of Panama (Petersen et al., 2010). The mesh modeling a  $M = 7.7$  event is depicted in the eastern part b) and in the western part c)



**Figure 3.6 –** Example of Characteristic Fault sources defined through a collection of planar surfaces modeling a multi-segment rupture

## Introduction

Tectonic regionalisation

Main predictor variables

Supported intensity measure types

## Implementation and use of GMPEs in seismic hazard analysis

Testing

Selection criteria

Spatial correlation of ground motion

## Future developments

Adjusted equations

Non-ergodic sigma

Sigma adjustment

Vs-Kappa correction

Host-to-target adjustment

Spatial cross correlation

Near source directivity effects

# 4. Ground Motion Prediction Equations

This chapter provides an overview of ground shaking intensity modeling based on empirical equations and describes the way GSIMs — more commonly known as ground motion models or Ground Motion Prediction Equations (GMPEs) — are implemented in the OQ-engine.

## 4.1 Introduction

Ground shaking intensity models are empirical equations that - given a set of parameters - compute a value representative of the shaking intensity together with an associated variability. GSIM have a fundamental importance in the overall PSHA architecture.

A ground shaking intensity equation can be schematised as follows (Al Atik et al., 2010):

$$Y = f(X_{es}, \theta) + \Delta \quad (4.1)$$

where  $Y$  is the natural logarithm of a ground shaking intensity measure,  $X_{es}$  is the vector of explanatory (or independent) variables,  $\theta$  is the vector of model coefficients and  $\Delta$  is a random variable describing the overall variability of the ground shaking intensity at the site.

The selection of independent variables and the definition of the structure of the equation is usually done on the basis of physical principles and basic descriptions of the earthquake process, the latter intended as the combination of a rupture occurrence, the synchronous radiation of seismic waves and their propagation to the site.

### 4.1.1 Tectonic regionalisation

The different properties of the ground-motion generated by earthquakes of comparable size but occurring in dissimilar tectonic regions (e.g. stable continental regions, subduction interface) is well recognized in the scientific literature (Abrahamson and Shedlock, 1997).

The OQ-engine computes hazard using Seismic Source Models (SSMs) which may contain sources belonging to different tectonic regions. The assignment of each source to a specific tectonic region is habitually completed using a zonation map called tectonic regionalisation (see for example Delavaud et al., 2012).

Each seismic source contains a label specifying the tectonic region to which it belongs. The OQ-engine automatically selects from the Ground Motion Model (GMM) the associated GSIM.

#### 4.1.2 Main predictor variables

In the current section we give a brief overview of the most important predictor variables supported by the OQ-engine (for general a summary, see Akkar et al., 2013); currently they are organised into three main categories: variables describing the rupture properties, variables describing the rupture-site path and variables used to characterise the site conditions.

Table 4.1 contains a summary of the variables assigned to the three groups.

**Table 4.1 – Principal predictor variables supported by the OQ-engine and corresponding groups.**

Group	Variables
Rupture parameters	- Magnitude - Dip - $Z_{tor}$ - Rake
Rupture-site distances	- See Table 4.2
Site parameters	- $V_{S,30}$ - Depth to the 1 km/s interface interface - Depth to the 2.5 km/s interface

#### Magnitude

Moment magnitude (Hanks and Kanamori, 1979) is the magnitude typology preponderantly used within the most recent GSIMs and - as a consequence - within seismic hazard analysis in general. The OQ-engine, however, does not contain a predefined magnitude typology. It is up to the user to ensure that the magnitude used to define earthquake occurrence in the seismic source model is consistent with the one used in the selected ground shaking intensity models.

#### Distance

The OpenQuake-engine supports almost all the rupture-site distance metrics used by the most recent and complex ground shaking intensity models published in the scientific literature (see Table 4.2 for a comprehensive list). The calculation of distances within the hazard component of the OQ-engine is performed by assuming a spherical Earth with a radius of 6371.0 km.

Since earthquakes are always modelled in the OQ-engine as finite ruptures all rupture-site metrics are always computed instantaneously. For this reason the engine does not contain conversion equations between different metrics (see for example Scherbaum et al., 2004)

#### Rupture mechanism

Many GSIMs compute ground-motion usually using as a categorical variable describing the rupture mechanisms (e.g. normal, strike-slip or reverse).

In the OQ-engine the rupture mechanism of a seismic source is specified in terms of the rake angle (defined according to Aki and Richards, 2002). Since the rake is not used directly as a predictor variable, most of the GSIMs implemented in the OQ-engine contains a mapping between the rake angle and the rupture mechanism classes supported by each specific model (for a review see page 24 of Akkar et al., 2013).

#### Time averaged shear-wave velocity in the uppermost 30m ( $V_{S,30}$ )

Local site conditions and their effects on the ground-motion are routinely incorporated into ground shaking intensity models by means of a scalar quantity corresponding to the time-averaged shear wave velocity measured in the uppermost 30m of the soil column ( $V_{S,30}$ ). Local site conditions in the OQ-engine are specified by means of this parameter.

**Table 4.2 – Rupture-site distances supported by the OQ-engine.**

<b>Distance definition</b>	<b>Symbol</b>	<b>Description</b>
Epicentral	$R_{epi}$	Distance between the epicenter and the site. Note that currently in the OQ-engine the hypocenter is assumed to be at the center of a rupture.
Hypocentral	$R_{hypo}$	Distance between the hypocenter and the site. Note that currently in the OQ-engine the hypocenter is assumed to be at the center of a rupture.
Joyner and Boore distance	$R_{jb}$	Closest distance between the site and the surface projection of the rupture
Closest distance to the rupture	$R_{rup}$	Closest distance between the site and the rupture surface
Horizontal top-edge distance	$R_x$	Horizontal distance between the site and the top edge of the rupture
Top-of-Rupture depth	$Z_{tor}$	Depth to the top edge of the rupture

In case of ground shaking intensity models which support the definition of local soil conditions through soil classes (e.g. hard rock, soft soil) their implementation is done in a way that given a value of  $V_{S,30}$  the corresponding soil class is used to compute the value of ground motion (provided that a mapping between soil classes and  $V_{S,30}$  is defined by the authors).

Additional parameters used to quantitatively describe local geology are the depths to the 1 km/s and 2.5 km/s shear-wave velocity interfaces. These are parameters used in some GSIMs (e.g. Chiou and Youngs, 2008) to capture the influence of the deeper stratigraphy.

#### **Depth to the top-of-rupture ( $Z_{tor}$ )**

The depth to the top of rupture is a parameter introduced in some of the NGA West 1 GSIM such as Chiou and Youngs (2008) and Abrahamson and Silva (2008) following a supposed dependence of ground-motion to the depth of the source, as suggested by Somerville and Pitarka (2006).

#### **4.1.3 Supported intensity measure types**

Each GSIM implemented in the OQ-engine provides a list of the supported Intensity Measure Types (IMTs). Table 4.3 contains a comprehensive list of possible options. The definitions of the

**Table 4.3 – Principal intensity measure types supported.**

<b>Acronym</b>	<b>Description</b>	<b>Unit of measure</b>
PGA	Peak Ground Acceleration	g
PGV	Peak Ground Velocity	cm/s
PGD	Peak Ground Displacement	
SA	Spectral Acceleration	g
IA	Arias intensity	
CAV	Cumulative Absolute Velocity	
RSD	Relative Significative Duration (Trifunac and Brady, 1975)	
MMI	Modified Mercalli Intensity	

ground-motion component supported are instead listed in Table 4.4.

Component	Description
HORIZONTAL	General horizontal component
GEOMETRIC_MEAN	
GMRotI50	Median value of the (period independent) geometric mean (Boore et al., 2006)
RotD50	Median value of the (period dependent) geometric mean (Boore, 2010)
RANDOM_HORIZONTAL	Random horizontal component
VERTICAL	Vertical component of ground-motion

**Table 4.4 – Principal ground-motion components supported (THE CONTENT OF THIS TABLE MUST BE COMPLETELY REVISED)**

## 4.2 Implementation and use of GMPEs in seismic hazard analysis

The OQ-engine contains a large set of GSIMs developed for different tectonic regions. Currently the engine includes GSIMs for shallow earthquakes in active tectonic regions, earthquakes in stable continental regions, subduction regions and geothermal areas.

GSIMs are implemented following a template model (in the Python jargon a base class) which defines the basic behaviour and describes the principles to be followed for their implementation. Each GSIM contains a definition of the independent parameters used to describe the rupture, the site conditions, the rupture-site distance metrics, the intensity measure types supported, the type of standard deviation provided, the tectonic region where the use of the GSIM is recommended.

The main advantage of this approach is that GSIMs, no matter which are their specific properties or features, behave following a common standard. For example, this feature allowed the creation on top of the GSIM library of a universal testing procedure, a standard applied to all the models implemented in the OQ-engine which guarantees a baseline uniform quality assurance level.

A second advantage of the developed library relates to its flexibility and modularity. Once the properties of main objects are defined, GSIMs can be used interchangeably. For example, the OpenQuake Ground Motion Toolkit (Weatherill, 2014) builds on top of this library and includes tools for computing residuals given a dataset of recordings, or for the calculation of trellis plots that compare the scaling of multiple GMPEs side by side in terms of magnitude, distance, site-condition, spectra etc.. Figure 4.1 shows the scaling of ground-motion versus magnitude for some of the GSIMs implemented in the OQ-engine. The ground motion is computed for a site at a  $R_{jb}$  distance of about 33 km with  $V_{S,30}$  equal to 760 m/s from a rupture with a strike of dip of 45 degrees toward south for two different values of rake (i.e. rupture mechanism). The upper panel of Figure 4.1 shows the position of the site and the rupture.

### 4.2.1 Testing

The progressively increasing complexity of ground shaking intensity models is giving more and more emphasis and relevance to the validation of results provided by the GSIM implemented within PSHA codes and the results of original GSIM implementations as described in the scientific literature (or directly provided by the authors).

The standard process adopted for the implementation in the OQ-engine of a ground shaking intensity model requires a set of verification tables each one containing values of ground-motion (or standard deviation) computed using a large number of combinations of the predictor variables. Table 4.5 shows a simplified example of a GSIM verification table; it consists of: a header line with (standard) names for each of the column, a number of lines each one containing values of the predictor variables plus the computed values of ground-motion intensity or standard deviation.

Examples of verification tables are available on the OpenQuake-hazardlib Github repository<sup>1</sup>.

**Table 4.5 – Schematic of a GSIM verification table used in the OQ-engine.**

M	R	V <sub>S,30</sub>	IMT <sub>1</sub>	IMT <sub>2</sub>	...
val <sub>1,1</sub>	val <sub>1,2</sub>	val <sub>1,3</sub>	val <sub>1,4</sub>	val <sub>1,5</sub>	
val <sub>2,1</sub>	val <sub>2,2</sub>	val <sub>2,3</sub>	val <sub>2,4</sub>	val <sub>2,5</sub>	
...					

Using these tables and an automated verification procedure implemented in the OQ-engine, it is possible to verify the consistency between the original results and the corresponding values computed with the version of the GSIM implemented. On average we accept OQ-engine

GEM recommends contextually to the publication of GSIM distributing as an electronic attachment the table of coefficients as well as of a set of verification tables (or a software which allows the generation of these tables). This can certainly improve the reproducibility of the new models proposed and most of all would improve the quality and robustness of the computed hazard.

#### 4.2.2 Selection criteria

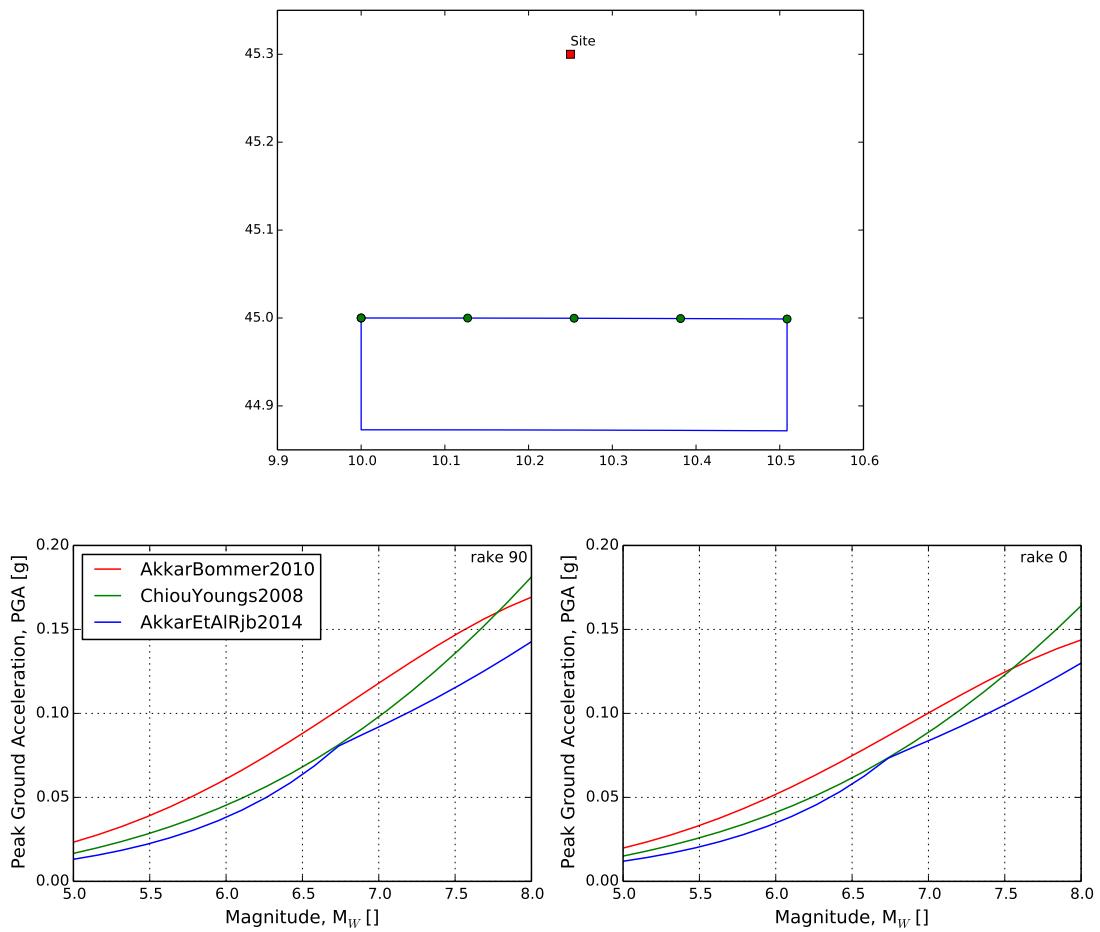
This section will be completed later.

#### 4.2.3 Spatial correlation of ground motion

This section will be completed later.

---

<sup>1</sup> <https://github.com/gem/oq-hazardlib/tree/master/openquake/hazardlib/tests/gsim/data>



**Figure 4.1** – (upper panel) Simple schematic with the surface projection of the rupture and the site (red square) used in this example. The green dots show the position of the top of rupture. (lower panels) Scaling of Peak Ground Acceleration as a function of magnitude obtained by some of the GSIM implemented in the OQ-engine.

### **4.3 Future developments**

This part will completed later.

**4.3.1 Adjusted equations**

**4.3.2 Non-ergodic sigma**

**4.3.3 Sigma adjustment**

**4.3.4 Vs-Kappa correction**

**4.3.5 Host-to-target adjustment**

**4.3.6 Spatial cross correlation**

**4.3.7 Near source directivity effects**



<b>Introduction</b>
<b>The OpenQuake-engine logic-tree structure</b>
The seismic source model logic tree
The ground-motion model logic tree
<b>Logic tree processing</b>
Full-path enumeration
Monte Carlo sampling
Calculation of mean and percentiles/quantiles
<b>Future developments</b>
Logic tree pruning/collapsing

## 5. Logic-trees

The logic-tree is an integral component of a PSHA input model for the OpenQuake-engine. An OQ-engine input model always contains a logic tree structure describing the epistemic uncertainties associated with the construction of the seismic source model and a logic-tree used to formally specify epistemic uncertainties related to the GSIM models used in different tectonic regions for the calculation of hazard.

The definition of logic trees in the OpenQuake-engine is based on the combination of predefined modules, each one modelling a specific epistemic uncertainty. There are the main advantages of this approach. The first and most obvious is that the user is not forced to use a predefined logic tree structure and - instead - can create a tailored logic tree which integrally reflects the specific uncertainties he or she wants to model. The second is that the logic tree structure becomes an integral part of a PSHA input model definition. The hazard calculations based on this approach are testable, fully reproducible and do not require pre- or post-processing steps since these are features included in the engine.

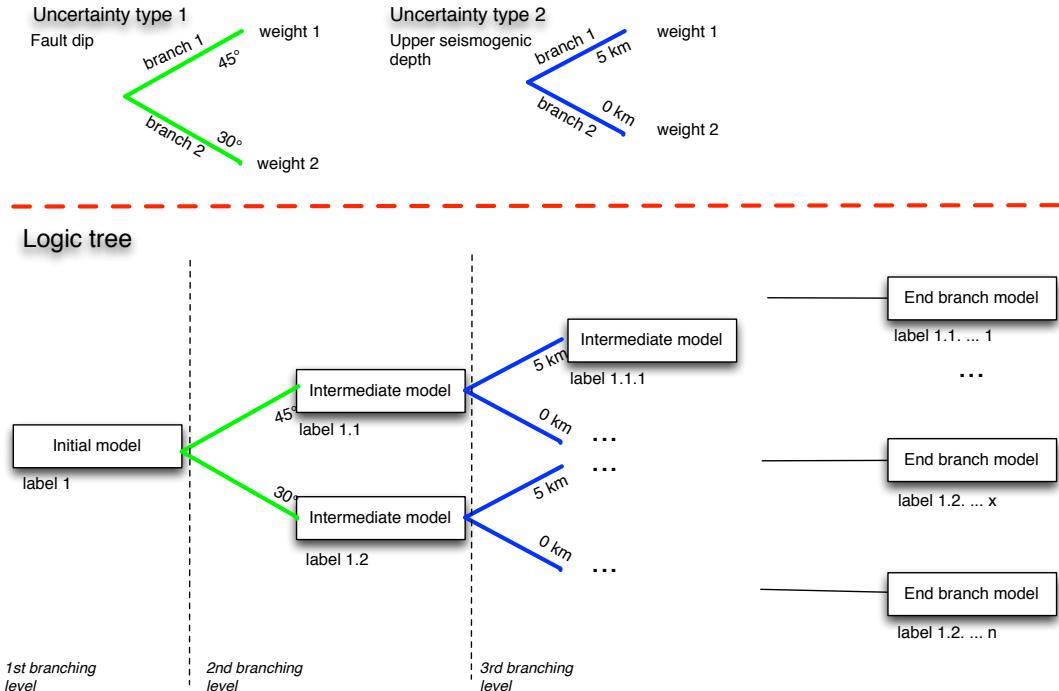
This Chapter is dedicated to the description of the basic theory behind logic-trees and to the delineation of how logic-trees are implemented into the OpenQuake-engine.

### 5.1 Introduction

The use of logic-trees to account for epistemic uncertainties in probabilistic seismic hazard analysis was originally proposed by Kulkarni et al. (1984). Nowadays logic-trees are an essential component of a PSHA input model and represent the formal methodology through which it is possible to synthesize the outcomes of the elicitation process on epistemic uncertainties requested in site-specific seismic hazard analyses (Budnitz et al., 1997) as well as for the creation of state-of-the-art national and regional PSHA input models.

The interpretation of the branches in a logic tree structure, of their corresponding weights and of the following results is still the subject of an intense scientific debate going on in the literature since 2005 (Abrahamson and Bommer, 2005; McGuire et al., 2005; Musson, 2012; Scherbaum and Kuehn, 2011). There is however a general agreement on the fact that the branches used to describe alternative interpretations (or values of a parameter affected by epistemic uncertainty) must be mutually exclusive and collectively exhaustive (Bommer and Scherbaum, 2008). This means that while processing the logic tree, once you choose one option in the implementation of a model you automatically exclude the other possible interpretations (mutual exclusivity) and

that the set of options described by the different branches represents the entire group of options admitted (collective exhaustiveness). While the first assumption is relatively easy to accept - although it presumes the lack of correlation between the uncertainties in the different branches - the second one certainly has implications that are more difficult and delicate to go along with since it presumes a comprehensive knowledge of a specific model uncertainty, knowledge that cannot be assumed *a priori*.



**Figure 5.1** – An example of modular logic tree structure supported by the OQ-engine. The upper part of the figure contains two modules, one modelling the epistemic uncertainty on the dip angle of a fault, the other modelling the uncertainty on faults upper seismogenic depth.

## 5.2 The OpenQuake-engine logic-tree structure

The OQ-engine offers a flexible and modular methodology to create customised logic-tree structures. The main components of this structure are (see for example Figure 5.1):

- **branch**  
It is the elemental component of a logic-tree. A branch represents one possible interpretation of a model or parameter affected by epistemic uncertainty. It is uniquely defined by a tuple consisting of a value and a real number in  $[0, 1]$  that can be either be considered a degree of confidence or a probability.
- **branch set**  
A branch set is a group of branches which collectively describes the (epistemic) uncertainty associated with a parameter or a model; the sum of weights for the branches within a branch set must be equal to one.
- **branching level**  
A branching level defines the position of a branch set within the logic tree structure. The lower is the value of the branching level the closer is the branch set to the roots of the tree. A branch set, as well as a branch, is defined with a unique identifier.

The logic-tree is the combination of a set of linked modules which starting from the roots specify the structure until the uppermost branches. A branch set can applied to all the sources included in the initial seismic source model, to a subset of sources, to a branch included in a branch set occurring before in the logic tree structure or even just to a single source.

Currently, the rules controlling the application of a branch set incorporated into the OQ-engine are the following:

- *applyToBranches*

The current branch set is applied to one or more branches of the previous branching level designated through their unique ID;

- *applyToSources*

The current branch set is applied to one or more sources included in the initial seismic source model designated through their unique ID;

- *applyToSourceType*

The current branch set is applied to all the sources of a specific type (e.g. simple fault sources) included in the initial seismic source model;

- *applyToTectonicRegionType*

The current branch set is applied to all the sources belonging to a selected tectonic region type (e.g. stable continental).

The schematic represented in Figure 5.1 shows an example of the conceptual model adopted to describe a logic tree structure.

### 5.2.1 The seismic source model logic tree

The seismic source model logic tree handles the epistemic uncertainties related to the definition of geometry, position and seismicity occurrence properties of seismic sources capable of generating ground-motion of engineering relevance at the investigated site.

By default, the first branching level of a seismic source model logic tree contains one (or several) initial seismic source model. Currently is not possible to create a logic tree model by incrementally adding different sources as - for example - in the case of some of the logic tree structures included in the recently presented CEUS-SSC model (Technical Report, 2012). This functionality will be added into future versions of the software.

#### Supported epistemic uncertainties

At the present time the OQ-engine provides a limited set of modules describing a specific epistemic uncertainty related to the creation of the seismic source model. A short description of each module is provided below. Note that the rules defined by each branch set are applied to the sources in the input model matching one of the filters specified within section 5.2. If a branch set has not a filter, then the associated epistemic uncertainty will be applied to all the sources included in the seismic source model.

- *Seismic source model*

This module allows the user to load one or several initial seismic source models. Using this module it is possible to use models with different source geometries and properties based on distinct assumptions or interpretations.

- *Relative uncertainty on the b-value of the double truncated Gutenberg-Richter relationship*  
This branch set adds (or subtracts) a delta to the b-value of the double truncated Gutenberg-Richter relationship.

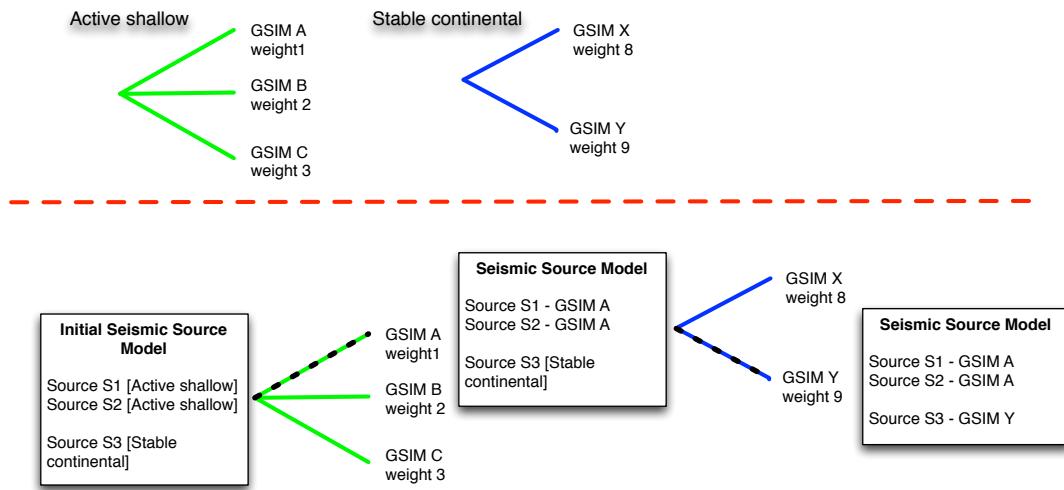
- *Uncertainty on the a-value of the double truncated Gutenberg-Richter relationship*  
This branch set assigns a value to the a-value of the double truncated Gutenberg-Richter relationship.

- *Uncertainty on the maximum magnitude of a double truncated Gutenberg-Richter distribution*

This branch set considers the epistemic uncertainty on the maximum value of magnitude used to define a double truncated Gutenberg-Richter distribution. The application of this branch set adds (or subtracts) a delta value to the maximum magnitude.

- *Uncertainty on the maximum magnitude of a double truncated Gutenberg-Richter distribution*

This branch set considers the epistemic uncertainty on the maximum value of magnitude used to define a double truncated Gutenberg-Richter distribution. The application of this branch set assigns a value to the maximum magnitude of a double truncated Gutenberg-Richter.



**Figure 5.2 – (upper panel)** Example of branch sets belonging to the ground-motion logic tree. **(lower panel)** Example of ground-motion logic tree processing. The initial seismic source model, on the left, is propagated through a simple logic tree structure following the path indicated by the black dashed line. Model information is added incrementally as the input models propagate through the tree structure.

### 5.2.2 The ground-motion model logic tree

The current structure of the ground-motion model logic tree is simple and designed to support just the use of alternative GSIMs models for a single tectonic region.

#### Supported epistemic uncertainties

The epistemic uncertainty allowed for the GSIM logic-tree are the following:

- *Ground shaking intensity models*

This module assigns to each tectonic region one or many GSIMs. This branch set implicitly contains a filter since it is applied only to the seismic sources belonging to the corresponding tectonic region. The example within Figure 5.2 illustrates the common processing of the ground-motion logic tree operated by theOQ-engine. In this example the source model contains seismic sources included in two tectonic domains: active tectonics and stable continental. The branch set defined for 'active shallow' is then applied just to sources 'S1 and 'S3' while the branch set for sources in stable continental regions is utilised only for source 'S3'.

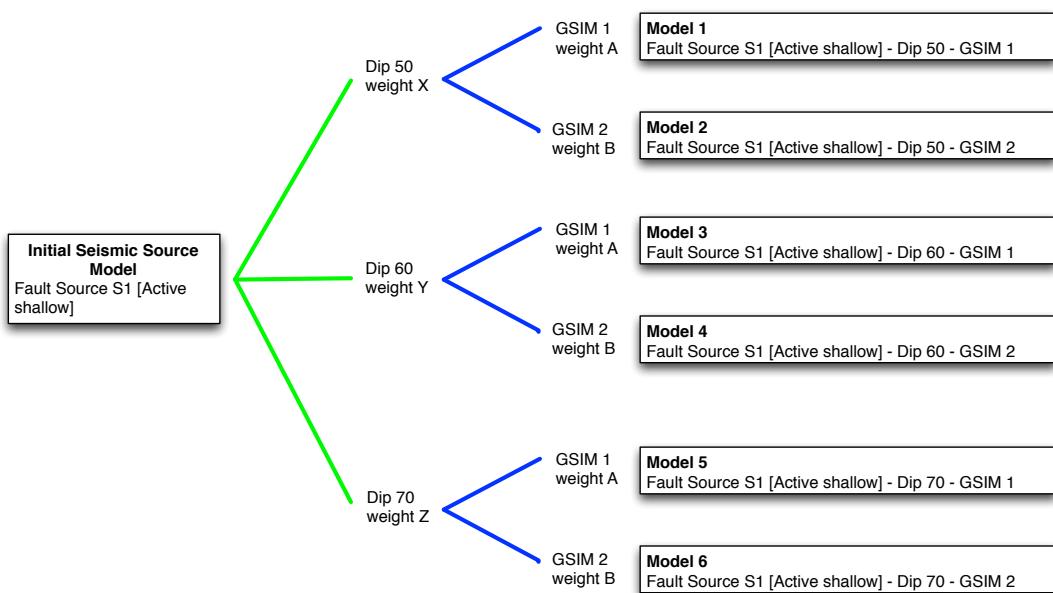
### 5.3 Logic tree processing

The OQ-engine currently provides two distinct ways to process logic-trees: full-path enumeration and Monte Carlo sampling.

Full path enumeration is a methodology which generates all the models admitted by a logic tree structure. For this reason, the use of this methodology is feasible only when the logic tree structure is relatively simple, that is when the number of end branches is at maximum in the order of a few tens.

Monte Carlo sampling is instead a methodology which makes an extensive use of random number generation in order to select a subset of models capable to reliably define the overall uncertainty on the final results produced by the epistemic uncertainties used in the construction of the logic tree structure.

In the following sections we provide a short description of the these two methodologies as implemented in the OQ-engine.



**Figure 5.3 – Logic tree full path enumeration processing.** Note that the first branching level, the one dealing with the definition of the initial seismic source model is neglected since we assume there is no epistemic uncertainty associated with its definition. The final PSHA input model contains the initial sources each one with an associated GSIM to be used in the calculation of hazard for this specific logic tree path.

#### 5.3.1 Full-path enumeration

Full-path enumeration is the simplest methodology implemented in the OQ-engine for logic-tree processing. As previously anticipated, it consists of computing hazard for the entire set of investigated sites using all the possible paths admitted by the specific logic tree structure defined.

Let's consider the example described in Figure 5.3. to illustrate how this method operates. The logic structure depicted in this Figure contains two branching levels each one including a single branch set. Note that for the sake of simplicity and clarity we assume that the first branching level (i.e. the one used to define the initial seismic source model) is not affected by epistemic uncertainty. Note also that the initial seismic source model contains only one fault source. The branch set in the first branching level describes the epistemic uncertainty on the dip

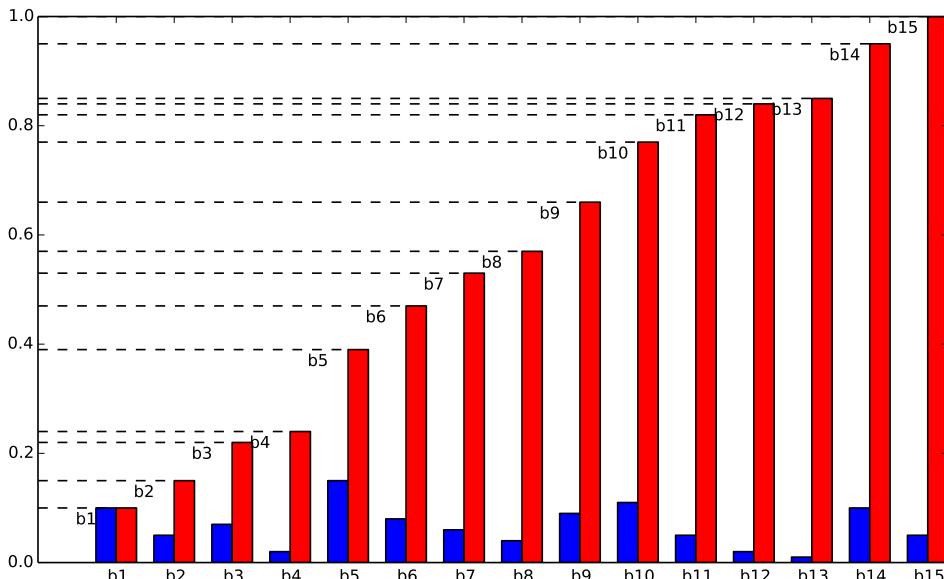
angle; three values, each one with an associated probability, are considered plausible. The second branch set describes the epistemic uncertainties associated with the modelling of ground-motion; two GSIMs are admitted in this case. On the right side of the figure the entire set of models originated by the logic tree structure are briefly described in terms of their distinctive parameters.

### 5.3.2 Monte Carlo sampling

The Monte Carlo sampling of the logic tree is implemented in a simple and straightforward way.

Given a branch set, following the same order used to add the branches we create a cumulative distribution function like the one represented by the red bars in Figure 5.4. A sample model is then obtained from this distribution simply via the generation of a random number (i.e. a real number in the interval [0.0, 1.0]) and the identification of the interval in the cumulative distribution which includes it. In Figure 5.4 the endpoints of the intervals are represented with horizontal dashed segments. Let's assume for example that the random number generator gives a value equal to 0.6. As clearly visible on the y-axis, this value falls within the interval relative to branch 'b9'. Following samples will be generated by repeating the same procedure as many times as needed. Clearly the higher is the weight associated with a branch the higher will be its probability of being sampled. In the example figure the branch with the higher weight is 'b5'.

A full path over the logic tree structure is built starting from the initial seismic source model and repeating this sampling procedure at each branching level.



**Figure 5.4** – On the x-axis an hypothetical list of branches included in a branch set. The height of the blue bar is proportional to the corresponding weight. The red bars show the cumulative distribution function.

### 5.3.3 Calculation of mean and percentiles/quantiles

The calculation of statistical parameters on the computed hazard results is done using the following approach.

From the set of hazard curves computed at a specific site we select the probabilities of exceedance for a given intensity measure level  $P = \{poe_1, poe_2, \dots, poe_n\}$  where  $n$  is the number of realisations i.e. hazard results obtained by processing the logic tree.

In case of a full path enumeration processing methodology, we arrange  $P$  in ascending order and we change the order of the corresponding weights accordingly. Using these weights we compute the cumulative distribution function (CDF). We interpolate the curve defined by the CDF and the probabilities of exceedance (poes) to obtain the values of poes corresponding to the quantiles defined by the user. Using the computed values of the probability of exceedance we find the corresponding curves.

In case of a logic tree processing based on a Monte Carlo sampling the quantiles are computed from this set of probabilities using standard methodologies.

## 5.4 Future developments

This part will completed later.

### 5.4.1 Logic tree pruning/collapsing

## **Classical PSHA with an Area Source**

The area source discretization step  
The effect of dip and rake angles  
The effect of the hypocentral depth distribution

## **Classical PSHA with complex logic tree**

Convergence between Classical and Event-based PSHA

Disaggregation analysis

# **6. Calculation Examples**

This chapter presents practical examples of PSHA done with the OQ-engine. By considering simple synthetic test cases, we illustrate the behaviour of a number of algorithms underlying the seismic source modelling and the logic tree processing. We also present examples of PSHA performed using the national seismic hazard model for U.S. (Petersen et al., 2008) to illustrate the capabilities of the OQ-engine when dealing with complex models.

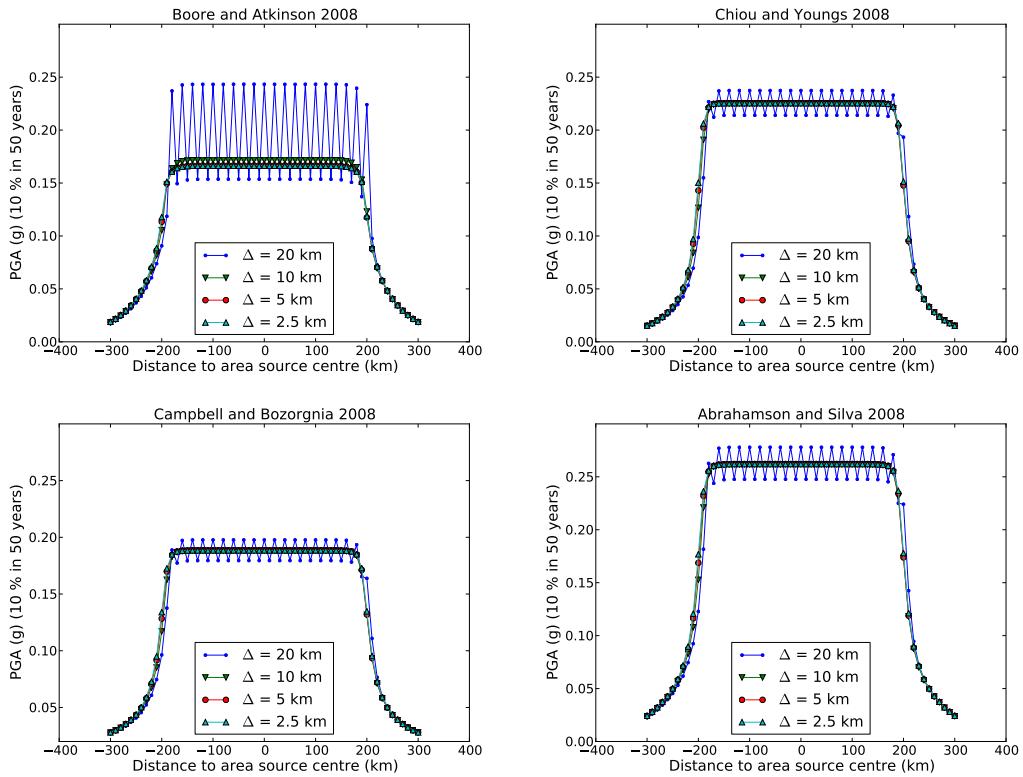
## **6.1 Classical PSHA with an Area Source**

We consider here the case of a single area source. The source has a circular shape of 200 km radius. The activity is described by a truncated Gutenberg-Richter magnitude frequency distribution, with minimum magnitude equal to 5 and maximum magnitude equal to 6.5. The *a-value* is 5 (i.e. the source generates one event per year with  $M \geq 5$ ) and *b-value* is 1. The seismogenic layer extends from 0 to 20 km. Ruptures are associated to a single hypocentral depth (10 km) and nodal plane (strike 0, dip 90, rake 0).

### **6.1.1 The area source discretization step**

We study here the effect of the area source discretization step ( $\Delta$ ) in the calculation of hazard map values. That is, we investigate the effect that the spacing used to discretize the region delimited by the area source boundary has on hazard levels corresponding to a given probability of exceedance. We thus compute hazard curves (for PGA) on a set of locations equally spaced by 10 km defining a profile crossing the centre of the area source, from east to west.

We compute hazard curves using different GMPEs (Boore and Atkinson, 2008, Chiou and Youngs, 2008, Campbell and Bozorgnia, 2008 and Abrahamson and Silva, 2008) to investigate the potential dependence of the results accuracy on the ground motion model. From the hazard curves we extract PGA values corresponding to 10 % in 50 years. Results for four discretization levels (20, 10, 5, and 2.5 km) are shown in Figure 6.1. When using  $\Delta = 20$  km, the hazard map values show strong fluctuations (where the highest are for the Boore and Atkinson, 2008 model) within the area region (that is in the distance range [-200, 200] km). For discretization steps equal to or smaller than 10 km, the different solutions converge instead to the same values.



**Figure 6.1 – The effect of the area source discretization step ( $\Delta$ ) on hazard map calculation**

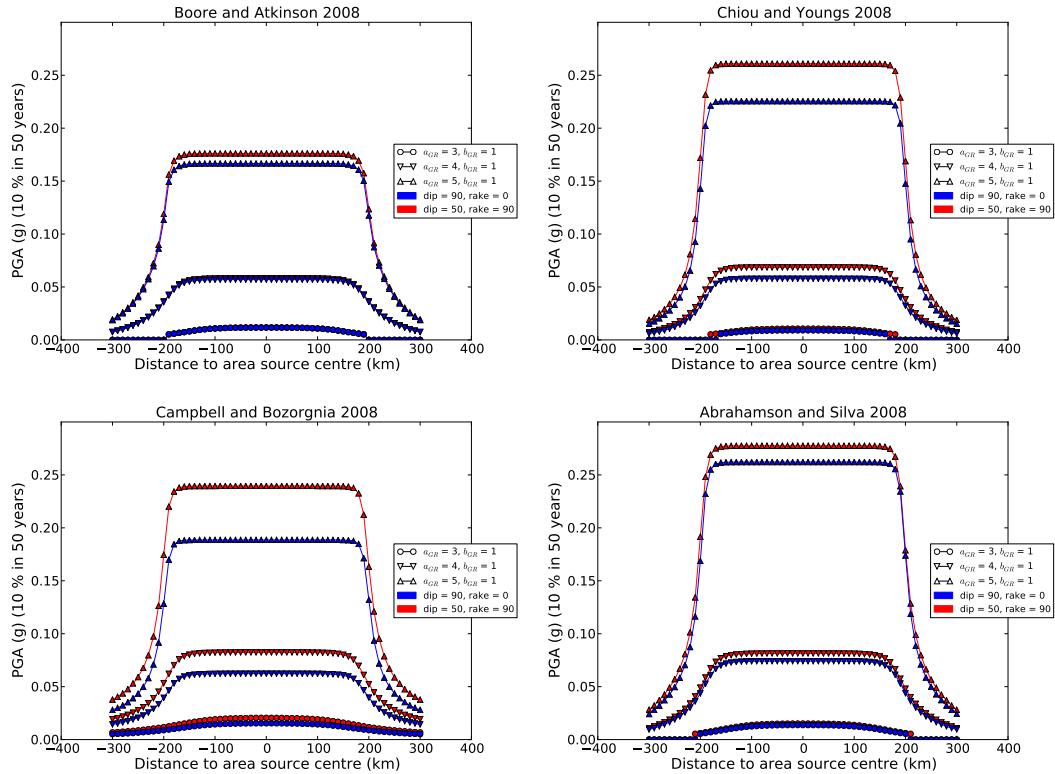
### 6.1.2 The effect of dip and rake angles

To investigate the effect of modelling earthquake ruptures with different inclination (that is dip angle) and faulting style (rake angle), we compare here hazard map values for an area source generating only vertical, strike-slip ruptures and an area source generating dipping (dip=50°), reverse (rake=90°) ruptures.

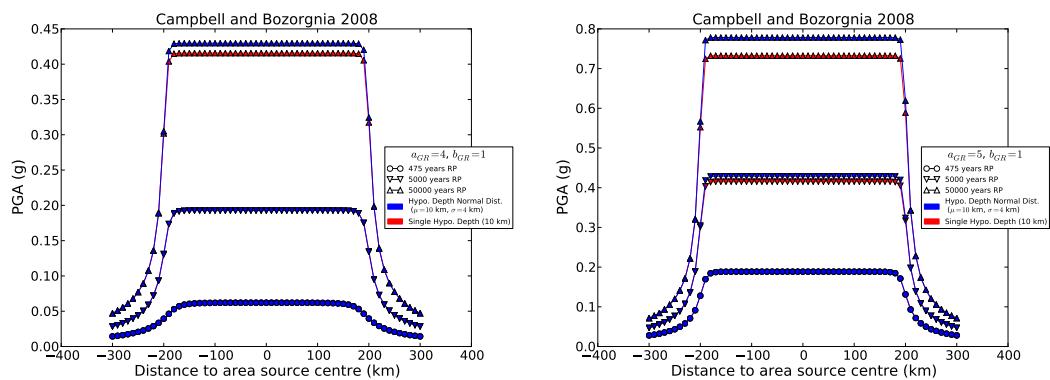
To investigate the potential dependence on the source seismic activity level, we compute hazard maps for area sources having different Gutenberg-Richter  $a$  values ( $a_{GR}$ ) equal to 3, 4 and 5, corresponding to annual occurrence rates above  $M = 5$  of 0.01, 0.1 and 1, respectively. Results are shown in Figure 6.2. Sensitivity of rupture dip and faulting style clearly depends on the source activity level and on the GMPE model. Independently of the GMPE, the highest absolute difference in PGA is for the highest  $a_{GR}$ . Among the different GMPE models, Campbell and Bozorgnia (2008) shows the highest sensitivity (about 20 increase in PGA at  $a_{GR} = 5$ ), while Boore and Atkinson (2008) shows the lowest sensitivity.

### 6.1.3 The effect of the hypocentral depth distribution

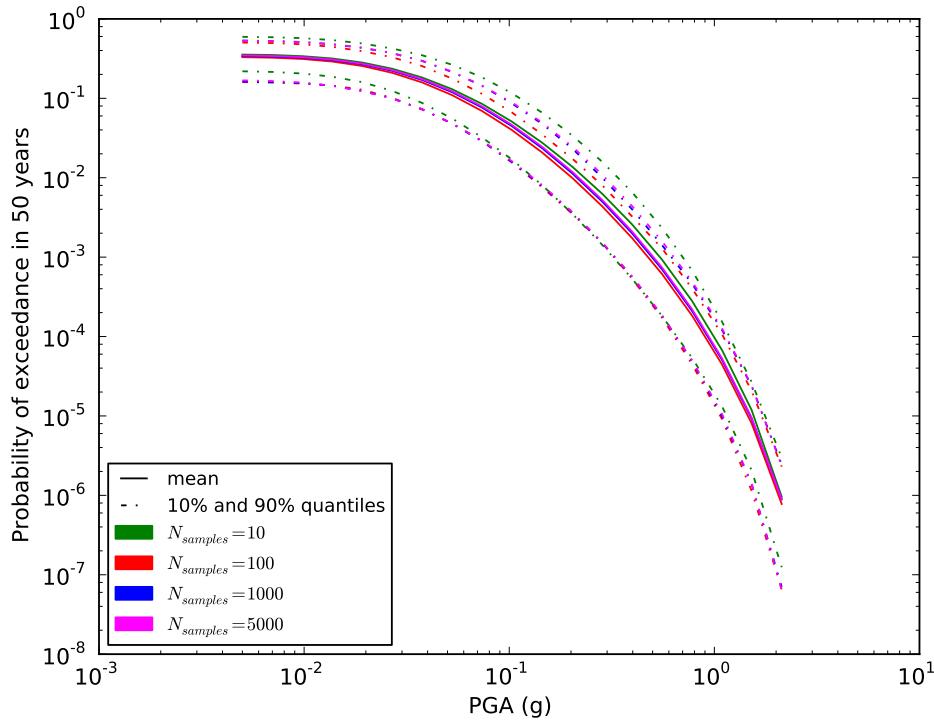
Another modeling parameter which can influence hazard estimates from an area source is the hypocentral depth distribution. We show here the effect of considering a single hypocentral depth value (10 km) versus considering a set of normally distributed values with mean  $\mu = 10$  km and standard deviation  $\sigma = 4$  km. By considering the same source-sites configuration as in the previous analysis, and vertical strike-slip ruptures with single strike (0°), we compute hazard maps considering two  $a_{GR}$  values (4 and 5). We use the GMPE model of Campbell and Bozorgnia (2008). Figure 6.3 shows hazard map values along the site profile for different return periods (RP) and  $a_{GR}$  values. The effect of the distribution of hypocentral values becomes visible when considering long return periods (50000 years) and increases with increasing  $a_{GR}$ .



**Figure 6.2 – The effect of dip and rake angles on hazard map calculation.**



**Figure 6.3 – The effect of hypocentral depth on hazard map calculation.**



**Figure 6.4 – Mean and quantile hazard curves obtained from Monte Carlo sampling of logic tree.**

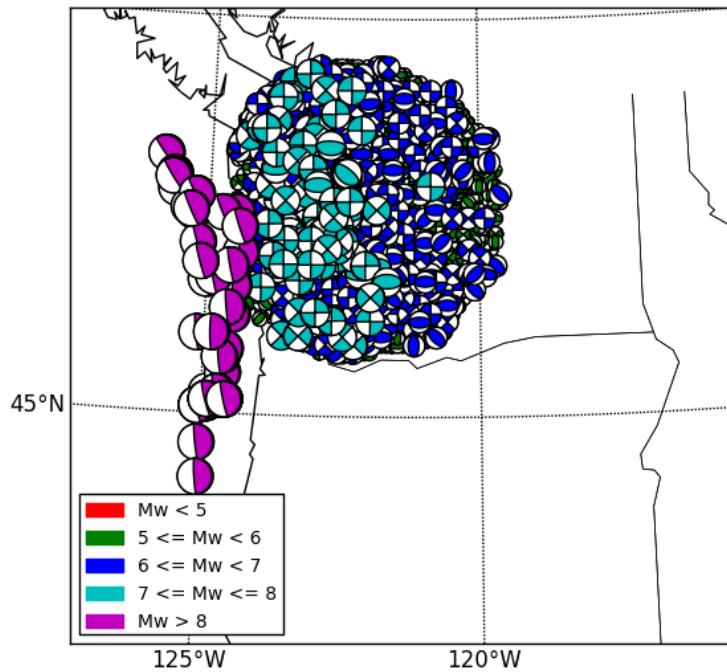
## 6.2 Classical PSHA with complex logic tree

We consider here a synthetic case of a classical PSHA calculation based on a simple source model consisting of 5 identical area sources. Each source has a square shape of 0.5° side. Four sources are arranged to form a regular mesh. The fifth source is placed in the middle of the mesh and overlaps with the other four sources. Each source can have 5 possible ( $a_{GR}$ ,  $b_{GR}$ ) pairs and 3 possible maximum magnitudes. Assuming the uncertainties to be uncorrelated among sources, the total number of possible source parameters combinations can be written as:

$$N = N_{GR}^{N_S} \times N_{MaxMag}^{N_S} \quad (6.1)$$

where  $N_{GR}$  is the number of Gutenberg-Richter parameters (i.e.  $a_{GR}$  -  $b_{GR}$  pairs) for each source,  $N_{MaxMag}$  is the number of maximum magnitudes for each source, and  $N_S$  is the number of sources. In the present case  $N_{GR} = 5$ ,  $N_{MaxMag} = 3$ ,  $N_S = 5$ , and thus  $N = 5^5 \times 3^5 = 759375$ .  $N$  represents the total number of paths in the source model logic tree.

The OQ-engine allows random sampling the logic tree to avoid calculating hazard results for all possible logic tree paths. Figure 6.4 presents mean and quantile hazard curves as obtained from different numbers of samples (10, 100, 1000, 5000). It can be seen how, by increasing the number of samples, results tend to converge to similar values. Indeed, curves obtained from 1000 and 5000 samples are almost indistinguishable. The Monte Carlo sampling offers therefore an effective way to reduce the computational burden associated with a large logic tree and to still obtain reliable results. The results reliability can be controlled by performing a convergence analysis; that is by identifying the number of samples which are required to obtain values that are stable within a certain tolerance level.



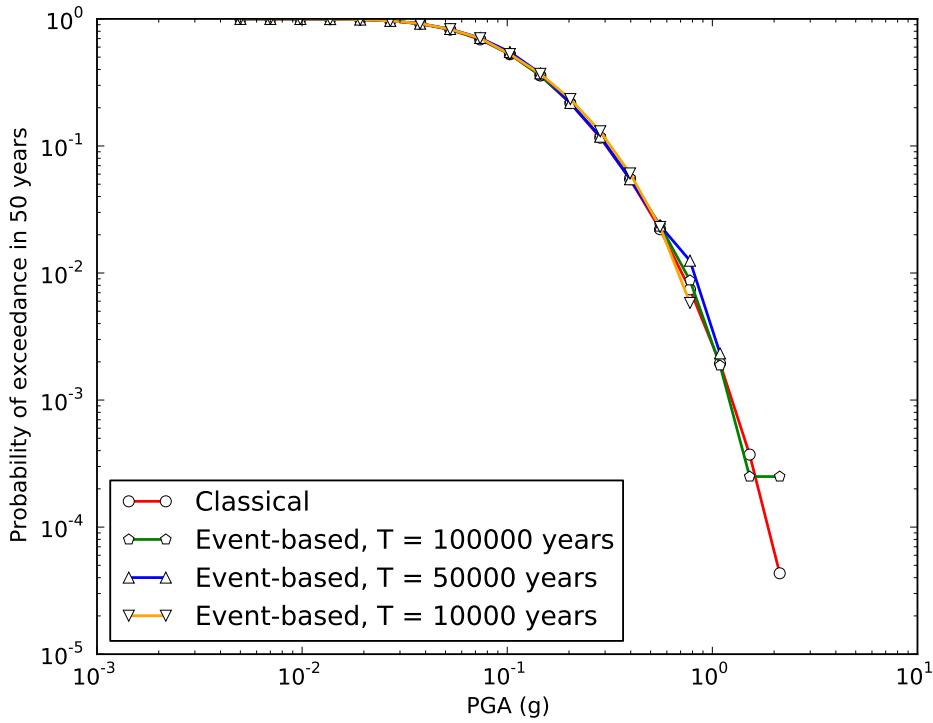
**Figure 6.5 – Stochastic event set for a region surrounding Seattle (U.S.) for a duration of 10000 years.**

### 6.3 Convergence between Classical and Event-based PSHA

The event based approach allows generating stochastic event sets and ground motion fields which can then be used to reproduce the classical results. We present here an event-based calculation for a location corresponding to the city of Seattle. The calculation is done using the 2008 seismic hazard model for conterminous U.S. (Petersen et al., 2008). A stochastic event set corresponding to a duration of 10000 years is generated (Figure 6.5). The event set contains earthquake ruptures within a radius of 200 km from the city of Seattle (longitude = 122.3W, latitude = 47.6N). The event set includes large subduction interface earthquakes generated in the Cascadia region, as well as deep intraslab and shallow active crust earthquakes. From each event, ground shaking values are simulated in the city of Seattle (considering the full set of GMPEs prescribed by the model). From ground motion values, the mean hazard curve (probability of exceedance in 50 years) for PGA is computed, and compared against the one obtained using the classical approach (Figure 6.6). The curve obtained can reliably reproduce the probabilities of exceedance down to  $10^{-2}$ . For lower probabilities a stochastic event set with longer duration is required.

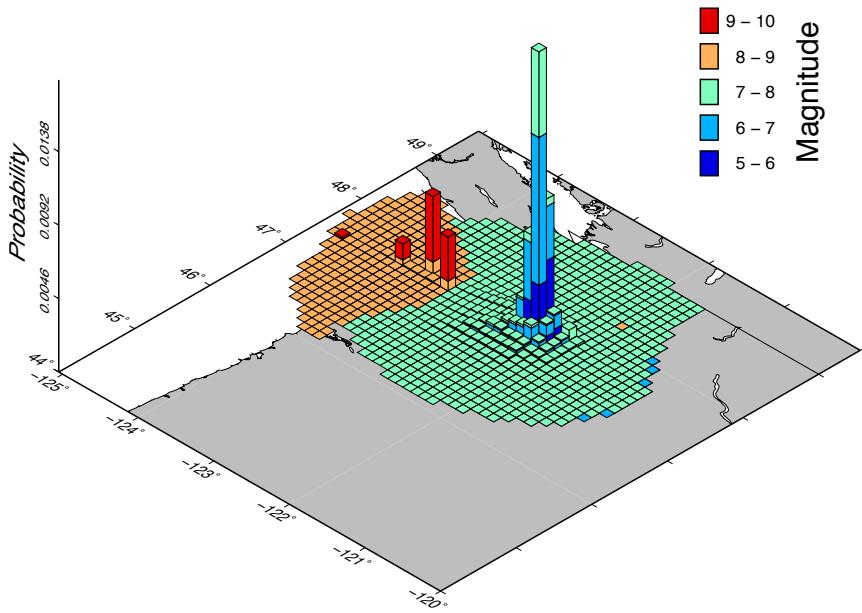
### 6.4 Disaggregation analysis

We present here an example of disaggregation analysis for the city of Seattle, always considering the 2008 national seismic hazard model for U.S. developed by Petersen et al. (2008). In particular, we show the geographic-magnitude (Figure 6.7) and geographic-tectonic region type (Figure 6.8) disaggregation histograms for PGA corresponding to 10% probability of exceedance in 50 years. The geographic disaggregation allows investigating the spatial distribution of the seismic sources contributing to a given level of hazard. By including magnitude and tectonic region type, we can

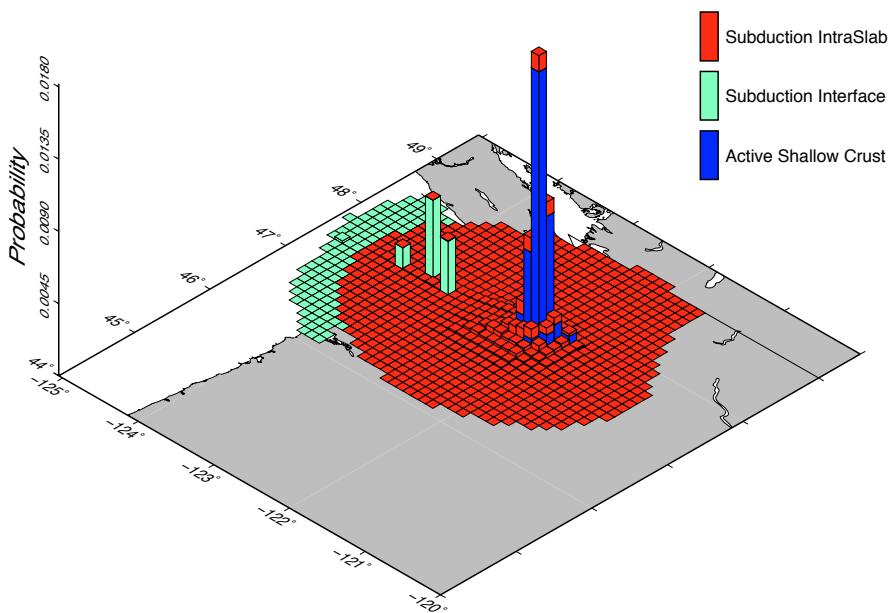


**Figure 6.6 – Hazard curves for Seattle using the Classical and Event-based approaches.**

understand the influence of the different tectonic regions, and also the magnitude ranges involved. Indeed, the disaggregation analysis for the city of Seattle shows that, for a return period of 475 years, the highest probabilities of ground motion exceedance are associated with active shallow crust events with magnitudes in the range 6 to 7. The second highest contributions are from subduction interface events with magnitudes above 9. Subduction intraslab events are instead associated to the lowest contributions.



**Figure 6.7 – Longitude, latitude and magnitude disaggregation for PGA corresponding to 10% probability of exceedance in 50 years.**



**Figure 6.8 – Longitude, latitude and tectonic region type disaggregation for PGA corresponding to 10% probability of exceedance in 50 years.**



## 7. Bibliography

### 7.1 Books

- Aki, K. and P. G. Richards (2002). *Quantitative Seismology*. Sausalito, California: University Science Books (cited on page 32).
- Reiter, L. (1991). *Earthquake hazard analysis*. Columbia University Press (cited on page 9).

### 7.2 Articles

- Abrahamson, N. A. and J. J. Bommer (2005). “Probability and Uncertainty in Seismic Hazard Analysis”. In: *Earthquake Spectra* 21.2, pages 603–607 (cited on page 39).
- Abrahamson, N. A. and K. Shedlock (1997). “Over”. In: *Seismological Research Letters* 68.1, pages 9–23 (cited on page 31).
- Abrahamson, N. A. and W. Silva (Feb. 2008). “Summary of the Abrahamson & Silva NGA Ground-Motion Relations”. In: *Earthquake Spectra* 24.1, pages 67–97 (cited on pages 33, 47).
- Al Atik, L., N. A. Abrahamson, J. J. Bommer, F. Scherbaum, F. Cotton, and N. M. Kuehn (2010). “The Variability of Ground-Motion Prediction Models and Its Components”. In: *Seismological Research Letters* 81.5, pages 794–801. DOI: 10.1785/gssrl.81.5.794 (cited on page 31).
- Bazzurro, P. and C. A. Cornell (1999). “Disaggregation of Seismic Hazard”. In: *Bulletin of the Seismological Society of America* 89.2, pages 501–520 (cited on pages 11, 19).
- Bommer, J. J. and F. Scherbaum (2008). “The Use and Misuse of Logic Trees in Probabilistic Seismic Hazard Analysis”. In: *Earthquake Spectra* 24.4, pages 997–1009 (cited on page 39).
- Boore, D. M. (2010). “Orientation-Independent, Nongeometric-Mean Measures of Seismic Intensity from Two Horizontal Components of Motion”. In: *Bulletin of the Seismological Society of America* 100.4, pages 1830–1835. DOI: 10.1785/0120090400 (cited on page 34).
- Boore, D. M. and G. M. Atkinson (Feb. 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”. In: *Earthquake Spectra* 24.1, pages 99–138 (cited on pages 47, 48).

- Boore, D. M., J. Watson-Lamprey, and N. A. Abrahamson (2006). “Orientation-independent measures of ground motion”. In: *Bulletin of the Seismological Society of America* 96, pages 1502–1511 (cited on page 34).
- Campbell, K. W. and Y. Bozorgnia (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”. In: *Earthquake Spectra* 24.1, pages 139–171 (cited on pages 47, 48).
- Chiou, B. S.-J. and R. R. Youngs (2008). “An NGA Model for the Average Horizontal Component of Peak Ground Motion and Response Spectra”. In: *Earthquake Spectra* 24, pages 173–215 (cited on pages 33, 47).
- Delavaud, E., F. Cotton, S. Akkar, F. Scherbaum, L. Danciu, C. Beauval, S. Drouet, J. Douglas, R. Basili, M. A. Sandikkaya, M. Segou, E. Faccioli, and N. Theodulidis (2012). “Toward a ground-motion logic tree for probabilistic seismic hazard assessment in Europe”. In: *Journal of Seismology* 16, pages 451–473. DOI: 10.1007/s10950-012-9281-z (cited on page 31).
- Ebel, J. E. and A. L. Kafka (1999). “A Monte Carlo Approach to Seismic Hazard Analysis”. In: *Bull. Seism. Soc. Am.* 89.4, pages 854–866 (cited on pages 10, 17).
- Field, E. H., T. H. Jordan, and C. A. Cornell (2003). “OpenSHA - A developing Community-Modeling Environment for Seismic Hazard Analysis”. In: *Seism. Res. Lett.* 74, pages 406–419 (cited on pages 9, 15).
- Hanks, T. C. and H. Kanamori (May 1979). “A moment magnitude scale”. In: *Journal of Geophysical Research* 84.B5, pages 2348–2350 (cited on page 32).
- McGuire, K. K., C. A. Cornell, and G. R. Toro (Aug. 2005). “The Case for Using Mean Seismic Hazard”. In: *Earthquake Spectra* 21.3, pages 879–886 (cited on page 39).
- McGuire, R. K. (1995). “Probabilistic Seismic Hazard Analysis and Design Earthquake: Closing the Loop”. In: *Bulletin of the Seismological Society of America* 85.5, pages 1275–1284 (cited on page 16).
- Musson (2012). “On the Nature of Logic Trees in Probabilistic Seismic Hazard Assessment”. In: *Earthquake Spectra* 28.3, pages 1291–1296. DOI: 10.1193/1.4000062 (cited on page 39).
- Musson, R. M. W. (2000). “The use of Monte Carlo simulations for seismic hazard assessment in the U.K.” In: *Annals of Geophys.* 43.1, pages 1–9 (cited on page 10).
- Pagani, M., D. Monelli, G. Weatherill, L. Danciu, H. Crowley, V. Silva, P. Henshaw, L. Butler, M. Nastasi, L. Panzeri, M. Simionato, and D. Vigano (May 2014a). “OpenQuake Engine: An Open Hazard (and Risk) Software for the Global Earthquake Model”. In: *Seism. Res. Lett.* 85.3, pages 692–702. DOI: 10.1785/0220130087 (cited on page 7).
- Scherbaum, F. and N. M. Kuehn (2011). “Logic tree branch weights and probabilities: Summing up to one is not enough”. In: *Earthquake Spectra* in press (cited on page 39).
- Scherbaum, F., J. Schmedes, and F. Cotton (2004). “On the Conversion of Source-to-Site Distance Measures for Extended Earthquake Source Models”. In: *Bulletin of the Seismological Society of America* 94.3, pages 1053–1069. DOI: 10.1785/0120030055 (cited on page 32).
- Schwartz, D. P. and K. J. Coppersmith (1984). “Fault behavior and characteristic earthquakes: Examples from the Wasatch and San Andreas Fault Zones”. In: *Journal of Geophysical Research: Solid Earth* 89.B7, pages 5681–5698 (cited on pages 22, 29).
- Trifunac, M. D. and A. G. Brady (1975). “A study on the duration of strong earthquake ground-motion”. In: *Bulletin of the Seismological Society of America* (cited on page 33).

### 7.3 Other Sources

- Akkar, S., J. Douglas, C. Di Alessandro, K. W. Campbell, P. G. Somerville, F. Cotton, W. A. Silva, and J. W. Baker (2013). *Define a Consistent Strategy to Model Ground Motion: Consistency in Model Parameters*. Technical report. 71 pages. GEM Foundation (cited on page 32).

- Budnitz, R. J., G. Apostolakis, D. M. Boore, S. L. Cluff, K. J. Coppersmith, C. A. Cornell, and P. A. (1997) Morris (1997). *Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and the Use of Experts*. Technical report NUREG/CR-6372. Two volumes. Washington D.C., United States: U.S. Nuclear Regulatory Commission (cited on page 39).
- Kulkarni, R. B., R. R. Youngs, and K. J. Coppersmith (1984). “Assessment of confidence intervals for results of seismic hazard analysis”. In: *Proceedings, Eighth World Conference on Earthquake Engineering* (cited on page 39).
- Pagani, M., D. Monelli, G. Weatherill, and J. Garcia (2014b). *Testing procedures adopted in the development of the hazard Component of the OpenQuake-engine*. Technical report. GEM Technical Report (cited on page 11).
- Petersen, M. D., A. D. Frankel, S. C. Harmsen, C. S. Mueller, K. M. Haller, R. L. Wheeler, R. L. Wesson, Y. Yzeng, O. S. Boys, D. M. Perkins, N. Luco, E. H. Field, C. J. Wills, and K. S. Rukstales (2008). *Documentation for the 2008 Update of the United States National Seismic Hazard Maps*. Open File Report 2008-1128. U.S. Department of the Interior, U.S. Geological Survey (cited on pages 47, 51).
- Petersen, M., S. Harmsen, K. Haller, C. Mueller, N. Luco, G. Hayes, J. Dewey, and K. Rukstales (2010). “Preliminary Seismic Hazard Model for South America”. In: *LA SISMOLOGÍA EN SUDAMÉRICA Y LOS MECANISMOS DE PREVENCIÓN Y MITIGACIÓN DEL PELIGRO Y RIESGO SÍSMICO*. Edited by Daniel Huaco. Lima, Peru (cited on pages 29, 30).
- Somerville, P. G. and A. Pitarka (2006). “Differences in earthquake source and ground motion characteristics between surface and buried earthquakes”. In: *Proc. Eighth National Conf. Earthquake Engineering*. Paper no. 977 (cited on page 33).
- Technical Report (2012). *Central and Eastern United States Seismic Source Characterization for Nuclear Facilities*. Technical report. EPRI, Palo Alto, CA, U.S. DOE, and U.S. NRC (cited on page 41).
- Thomas, P., I. Wong, and N. A. Abrahamson (May 2010). *Verification of Probabilistic Seismic Hazard Analysis Computer Programs*. PEER Report 2010/106. College of Engineering, University of California, Berkeley: Pacific Earthquake Engineering Centre (cited on page 11).
- Weatherill, G. (2014). *Ground Motion Prediction Equation and Strong Motion Modeller’s Toolkit*. Technical report. GEM Foundation (cited on page 34).



# Index

## G

Ground-motion prediction equation	
Predictor variables .....	32
predictor variables	
Magnitude, 32	
Source-site distance, 32	
Rupture mechanism.....	32
Time averaged shear-wave velocity in the uppermost 30m, $V_{S,30}$ , 32	
Depth to the top of rupture, 33	
Logic-tree .....	39

