"OpenQuake: Shaken not stirred"

The OpenQuake Book

February 4, 2011

Authors:
Helen Crowley[1], Leonardo Garrido[1], Damiano Monelli[3], Marco Pagani[1], Vitor Silva[1], Graeme Weatherill[2]

[1] GEM Foundation
via Ferrata, 1
20133 Pavia
Italy

[2] GEM Model Facility
via Ferrata, 1
20133 Pavia
Italy

[3] GEM Model Facility
SED - ETHZ
Sonneggstrasse, 5
CH-8092 Zurich
Switzerland

Email address (for all the authors):
<name.surname>@globalquakemodel.org

Contents

Symbols and acronyms

Quantity or Description	Symbol or acronym	Unit of measure
Distance	D or d	km
Frequency-Magnitude Distribution	FMD	
Gutenberg-Richter (frequency-magnitude distribution)	G-R	
Ground motion (GM) or intensity measure type (IMT)	U or u	g
Ground Motion Field	GMF	
Ground Motion Prediction Equation	GMPE	
Magnitude	M or m	
Natural hazards' Risk Markup Language	nrML	
Occurrence Rate	λ	events/yr
OpenQuake	OQ	
Rupture	rup	
Stochastic Event Set	SES	

PART I

Introduction

CHAPTER 1

OpenQuake background

An OpenQuake schema evidencing main inputs and calculators is represented in Figure 1.1.

Figure 1.1.: OpenQuake schema. Purple boxes are the calculators included in the the hazard part of OQ; green boxes are the risk calculators. The method of Wesson et al. [2009] is represented in a separate box since it incorporates hazard and risk calculations.

CHAPTER 2

Brief desciption of OpenQuake IT aspects

2.1. Risk-lib

PART II

Hazard

CHAPTER 3

Introduction

Introduction to the hazard component of OpenQuake

CHAPTER 4

OpenQuake input description

An OpenQuake Source model (i.e. a PSHA input model) contains four distinct information blocks: (1) information regarding location, geometry, and seismicity generation properties of seismic sources, (2) information about ground motion prediction equations to be adopted in the calculation, (3) information about the epistemic uncertainties related to the two aforementioned points, and (4) information specifying calculation settings.

The description of epistemic uncertainties is accomplished via a logic-tree structure, described branching level by branching level. In OpenQuake it's possible to define two logic-tree structures: one describing epistemic uncertainties associate with the creation of the ERF and one considering the uncertainties connected with the use of GMPEs. For the sake of homogeneity and simplicity, the OpenQuake PSHA input models are always defined using two logic tree structures. In cases when epistemic uncertainties are non accounted for, the two logic tree structures simply have one branching level with one branch (with weight equal to 1). Further explanations on the way we describe and model logic-tree is provided in section 4.1 at page 20.

The description of OpenQuake seismic source properties is subdivided into source location and geometry and source frequency-magnitude distribution. Section 4.2.1 provides an in deep description of the properties pertinent to the different source types supported, together with examples of each single source type nrML description.

The selection of ground motion prediction equations (GMPE) in OpenQuake is basically accomplished by specifying a label in a particular file. The user can associate a GMPE to each tectonic region considered in the analysis (Active Shallow tectonic, Stable Continental etc.). Examples of GMPEs selection are provided in section 4.3 at page 27.

The last, but not less important, block of information characterizing a PSHA input model contains parameters specifying the way calculations must be carried out. The information to be included in this part of the input is strictly related to the calculation properties of the engine. Section 4.4 provides an outlook of the hazard specific calculation setting supported by OpenQuake.

4.1. Logic-tree description

Logic-trees are a tool designed to consider in a systematic manner the epistemic uncertainties of models and parameters included in a hazard analysis.

In OpenQuake, the description of a logic-tree structure uses as its principal component a branching level where a branching level consists on (1) the definition of the parameter - or model - affected by uncertainty, (2) the specification of the type of uncertainty (3) the listing of the - mutually exclusive and collectively exhaustive [Bommer and Scherbaum, 2008] - alternative hypotheses and, (4) the index of the branches of the previous level to which this branching level applies. Each hypothesis (i.e. branch) included in a branching level has an associated value and a corresponding weight expressing - according to different interpretations available in the literature - "probabilities or simply subjective indications of relative merit" [Bommer and Scherbaum, 2008, pagecon 999].

Figure 4.1 depicts a branching level defining epistemic uncertainties on the dip angle of simple fault sources. In this case the possible values of the dip are specified on each branch composing the branching level (i.e. 30, 45 and 60 degrees). This means that these three values are the only ones admitted for all the sources included in the Source Model considered in this example.

Branching level definition: .... Simple Fault Dip Angle
Branching level uncertainty type: .... Absolute values
Applies to: .... Simple faults
Correlated branches: .... Yes

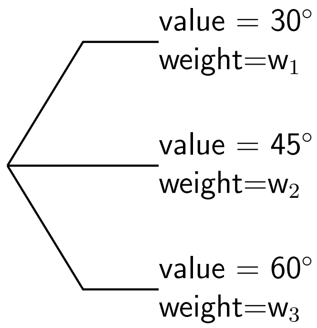

Figure 4.1.: Branching level description example. The upper example shows a branching level describing epistemic uncertainties on faults dip angle

As in the case of Figure 4.1, Figure 4.2 also shows a branching level defining epistemic uncertainties on the dip angle of simple fault sources. However - contrary to the example in Figure 4.1, in this case the values specified for each branch aren't absolute dip angles but instead delta values to be added - or subtracted - to the initial dip value specified for each simple fault source inclued in the initial source model.

Branching level definition: .... Simple Fault Dip Angle
Branching level uncertainty type: .... Relative values
Applies to: .... All previous branches
Correlated branches: .... Yes

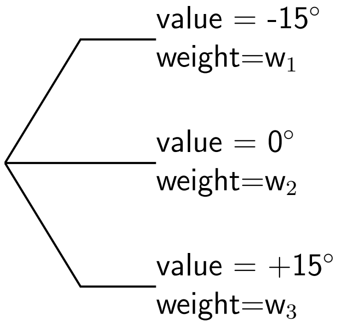

Figure 4.2.: Branching level description example. The upper example shows a branching level describing epistemic uncertainties on faults dip angle

Once a set of branching levels are defined, they can be easily and flexibly combined to create an entire logic-tree structure. Figure 4.3 shows an example of a logic tree structure obtained by combining two distinct branching levels.

marco: we probably need abstract classes for all these components of the LT structure

Figure 4.4 shows an example of a nrML file describing the logic tree structure.

4.2. The OpenQuake source model - Description of input data pertinent to the creation of the ERF

The creation of the Earthquake Rupture Forecast (or seismicity probabilistic occurrence model) is the first step in the calculation of hazard following a probabilistic procedure.

As mentioned in the introductory part of this Chapter, information describing the input data for the creating the ERF in OpenQuake is always organized into a logic tree structure.

4.2.1. Seismic source typologies description

OpenQuake, at present time, provides four seismic source typologies, for the most part defined in the course of the GEM1 project [Pagani et al., 2010]:

- Area source - This source type is the one that - at least for the time being - is most frequently adopted in national and regional PSHA models.
- Grid source - Grid sources can be considered a replacement for area sources since they both model distributed seismicity;

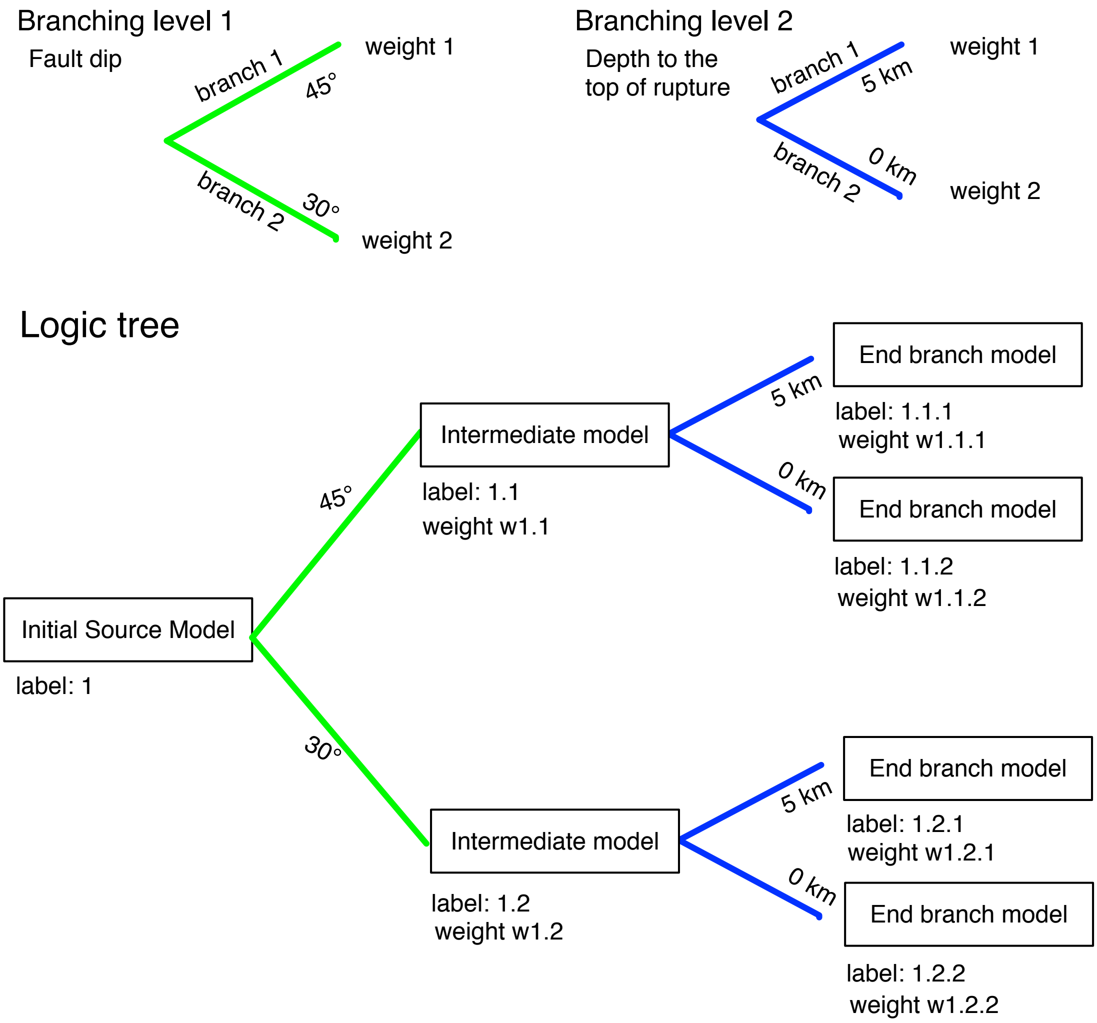

Figure 4.3.: Example of a logic tree structure as defined in OpenQuake. The upper part of the Figure depicts two branching levels.

```
<?xml version="1.0" encoding="UTF-8"?>
<ns2:logicTreeSet
 xmlns:ns1="http://www.opengis.net/gml/profile/sfgml/1.0"
 xmlns:ns2="http://openquake.org/xmlns/nrml/0.1"
 xmlns:ns3="http://www.w3.org/1999/xlink"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xsi:schemaLocation="http://openquake.org/xmlns/nrml/0.1
    file:/Users/damianomonelli/Projects/opengem/docs/schema/nrml_seismic.xsd">
    <ns2:logicTree tectonicRegion="Active Shallow Crust">
        <ns2:logicTreeBranchSet ns2:uncertaintyType="gmpeModel">
            <ns2:logicTreeBranch>
                <ns2:uncertaintyModel>
                    BA_2008_AttenRel
                </ns2:uncertaintyModel>
                <ns2:uncertaintyWeight>0.5</ns2:uncertaintyWeight>
            </ns2:logicTreeBranch>
            <ns2:logicTreeBranch>
                <ns2:uncertaintyModel>
                    CB_2008_AttenRel
                </ns2:uncertaintyModel>
                <ns2:uncertaintyWeight>0.5</ns2:uncertaintyWeight>
            </ns2:logicTreeBranch>
        </ns2:logicTreeBranchSet>
    </ns2:logicTree>
    <ns2:logicTree tectonicRegion="Subduction Interface">
        <ns2:logicTreeBranchSet ns2:uncertaintyType="gmpeModel">
            <ns2:logicTreeBranch>
                <ns2:uncertaintyModel>
                    McVerryetal_2000_AttenRel
                </ns2:uncertaintyModel>
                <ns2:uncertaintyWeight>1.0</ns2:uncertaintyWeight>
            </ns2:logicTreeBranch>
        </ns2:logicTreeBranchSet>
    </ns2:logicTree>
</ns2:logicTreeSet>
```

Figure 4.4.: nrML example

- Simple fault sources - Simple faults are the easiest modality available to specify the parameters need to characterize a fault source. This typology is usually adopted to describe shallow seismogenic fault sources.

- Complex fault sources - Complex faults is usually adopted to model subduction interface sources with a complex geometry.

The basic assumptions adopted in the definition of these source typologies are the following:

- In the case of area and fault sources, the seismicity is homogeneously distributed over the source;
- Seismicity temporal occurrence follows a Poissonian model;
- The frequency-magnitude distribution can be approximated to a evenly discretized distribution.

4.2.1.1. Area sources

Area sources usually model the seismicity occurring over wide areas where fault sources identification or characterization - i.e. the unambiguous definition of seismicity occurrence parameters - is difficult.

The Senior Seismic Hazard Analysis Committee [1997] defined three main types of area seismic sources using as a discriminant their extension:

1. Area sources enclosing concentrated zones of seismicity;
2. Regional area sources;
3. Background area sources.

As a general rule, the criteria - and the related uncertainties - adopted for their definition varies according to each area source type. From a hazard computation standpoint we do not introduce any difference between these three area types.

marco: In the future we may support a specialized background area source type

Parameters

- A polygon that identifies the external border of the area. Eventually, internal borders can be specified so as to create holes inside an area.

marco: I don't think nrML supports area sources with holes.

- One (or many) couples of the following objects:
  - A discrete Frequency-Magnitude Distribution (FMD)
  - Strike, dip, and rake angles characterizing the seismicity specified in the associated FMD and occurring in the area source under consideration. For example, Coppersmith et al. [2009] defines a discrete distribution of strike values (dip is not considered because the source-site metrics they use is the Joyner-Boore distance).

This area source specification permits the accurate characterization of seismicity occurrence within an area by explicitly distributing the seismicity on the existing faulting trends.

- An array to specify the depth to the top of rupture dependency on magnitude. The array contains two columns and one or many <depth, magnitude> tuples. Each tuple specifies the depth to the top of rupture for magnitudes equal or greater than the specific value.

- A value to indicate the hypocentral depth in case of punctual sources. By convention all the events with magnitude lower than the lowest value contained in the array used to specify the depth to the top of rupture are modelled considering a punctual source. On the opposite, ruptures with magnitude equal or greater than the lowest value of magnitude contained in the depth to the top of rupture array are modelled considering their finite dimensions.

4.2.1.2. Grid sources

A grid source is a typology used to model distributed seismicity - usually of low and intermediate magnitude.

Grid sources can be considered as a PSHA source model alternative to area sources since they both try to represent distributed seismicity. Grid sources usually derive from the application of seismicity smoothing algorithms [Frankel, 1995; Woo, 1996].

The use of these algorithms carries some advantages compared to area sources, indeed, (1) they remove most of the unavoidable degree of subjectivity due to the definition of the geometries and (2) they define a seismicity spatial pattern that is, usually, more similar to reality. Nevertheless, some smoothing algorithms require the a-priori definition of some setup parameters that expose the calculation to a certain partiality level.

Grid source models are modelled in OpenQuake simply as set of point sources. The next section describes the parameters required to characterize a point source.

Parameters

For each grid node:

- A location specified in terms of the ¡latitude,longitude¿ tuple;

- Similarly to area sources, one (or many) couples of the following objects:
  - A discrete Frequency-Magnitude Distribution (FMD)
  - Strike, dip, and rake angles characterizing the seismicity specified in the associated FMD.

- An array to specify the dependency on magnitude of the depth to the top of rupture. This array contains two columns and one or many <depth, magnitude> tuples where each tuple specifies the depth to the top of rupture for magnitudes equal or greater than a specific value.
- A value to indicate the hypocentral depth in case of punctual sources. The same convention specified for area sources applies here.

4.2.2. Accounting for rupture finiteness in case of areal and grid sources

In the scientific literature is well known that for magnitudes approximately greater than six the finite dimension of the rupture cannot be neglected in the calculation of the source-site distance.

marco: we need at least one citation

marco: it's not clear if and how much ruptures can extend outside the area border

To correctly calculate the source-site distance in case of area and grid sources two are the approaches available. The first is to multiply the epicentral or hypocentral distance by a correction factor (see for example Harmsen [2008]) the second requires to place on each node a number of ruptures with different orientation.

4.2.2.1. Simple faults

Simple Faults are the most common source type used to model faults; the "simple" adjective here refers to the geometry description of the source which is basically obtained by projecting a trace along a representative dip direction.

Parameters

- A fault trace (usually a multi-segment line)
- A FMD
- A representative value of the dip angle (Aki-Richards convention; see Aki and Richards [2002]),
- Rake angle (Aki-Richards convention; see Aki and Richards [2002])
- Upper and lower values of depth limiting the seismogenic interval
- A boolean flag that specifies if ruptures should follow a magnitude scaling relationship and thus be distributed homogeneously over the fault surface or it is accepted that ruptures of whatever magnitude (of course of the ones admitted by the FMD) will rupture the whole fault surface.

4.2.2.2. Complex faults

Complex faults differ from simple fault just by the way geometry is described and, consequently in the way the fault surface is created. The input parameters used to describe complex faults are, for the most part, the same used to describe the simple fault typology. In particular, in the case of complex faults the dip angle is not requested while the fault trace is substituted by two fault traces used to limit at top and bottom the fault surface.

Representation of complex faults

Usually, we use complex faults to model intraplate megathrust faults such as the big subduction structures active in the Pacific (Sumatra, South America, Japan).

4.3. GMPEs description

4.4. Calculation settings description

CHAPTER 5

Earthquake Rupture Forecast calculator

The Earthquake Rupture Forecast calculator creates a list of all the possible ruptures

5.0.0.3. The Poisson model

The Poisson distribution gives the probability of occurrence of n events in a time interval t, provided a value of the occurrence rate λ:

$$P(N = n|t, \lambda) = \frac{(\lambda t)^n \exp(-\lambda t)}{n!} \tag{5.1}$$

5.0.0.4. Brownian Time Passage (BTP) model

The Brownian Time Passage model was originally proposed by Matthews and Ellsworth [2002]. Not extensively used in PSHA; the Japan J-SHIS is one representative model containing this temporal occurrence model.

5.0.1. Frequency-magnitude distribution models

The frequency-magnitude distribution describes the density of earthquakes of a given magnitude in a given time interval.

5.0.1.1. Gutenberg-Richter distribution

Truncated Gutenberg-Richter distribution.

5.1. ERF creation in case of Area sources

Area sources (see also Section 4.2.1.1 at page 24)

5.2. ERF creation in case of Grid sources

5.3. ERF creation in case of Fault sources

5.3.1. Fault sources with simple geometry

5.3.2. Fault sources with complex geometry

CHAPTER 6

Classical PSHA calculator

OpenQuake computes classical PSHA [Cornell, 1968; McGuire, 2004] following the methodology proposed by Field et al. [2003]. This methodology has the distinctive property of performing the entire calculation using probabilities, as orginally proposed by Chiang et al. [1984], instead of working with occurrence rates like in most of the commonest PSHA codes [see for instance Bender and Perkins, 1987]. This methodology has the clear advantage of decoupling the creation of the probabilistic seismicity occurrence model (in the OpenSHA terminology this is defined as the Earthquake Rupture Forecast) from the assumption of a Poissonian temporal occurrence model. Field et al. [2003] demonstrated the congruence between the original PSHA formulation based on occurrences and the methodology here adopted in the case of a Poissonian temporal model.

From a more strict calculation perspective two, are the main steps composing the procedure:

- Creation of the probabilistic seismicity occurrence model, i.e. a discrete distribution giving the probability of occurrence of each possible rupture occurring on one of the seismic sources defined in the PSHA input model in a given time span. This step of the procedure was already discussed within Chapter 5 at page 29.

- Calculation of hazard at the site by combining the probabilistic occurrence model with a ground motion prediction equation (also Intensity Measure Relationship in the OpenSHA jargon). This second step will the topic of the current chapter.

6.1. Calculation kernel

The classical PSHA calculation kernel takes an Earthquake Rupture Forecast (ERF) as an input; an ERF is simply a list of all the possible ruptures occurring on all the seismic sources included in a Source Model; each rupture Rup is associated with a probability of occurrence in the time span t fixed for the analysis.

6.2. Hazard calculation: traditional formulation in terms of probabilities

Following Field et al. [2003], the traditional formulation in terms of probabilities:

$$P(U \geq u) = 1 - \prod_{i=1}^{l} \left[\sum_{s=0}^{+\infty} \left(P(S=s) \left(1 - \sum_{j=0}^{j(i)} \sum_{s=0}^{K(i,j)} P(m_{i,j}) P(R_{i,j,k}|m_{i,j}) P(U \geq u | m_{i,j}, R_{i,j,k}) \right) \right)^{s} \right. \tag{6.1}$$

where l i the number of sources

If the probability of multiple occurrences is assumed to be negligible the probability to get an exceedance of x in a given time span corresponds to:

$$P(X \geq x) = 1 - \prod_{i=1}^{l} \left(1 - \sum_{n=1}^{N(i)} P(Rup_{i,n}) P(X \geq x | Rup_{i,n}) \right) \tag{6.2}$$

6.2.1. Example

In case of two punctual sources each one generating a single rupture, the probability of exceedance of ground motion u in a given site corresponds to:

$$\begin{aligned} P(U \geq u) = 1 - \Big(& \left[1 - P(Rup_{i,1}) P(U \geq u | Rup_{i,1}) \right] \\ & \left[1 - P(Rup_{i,2}) P(U \geq u | Rup_{i,2}) \right] \Big) \end{aligned} \tag{6.3}$$

CHAPTER 7

Stochastic event set and ground motion field calculators

The calculation of stochastic event sets and the corresponding ground motion fields is a methodology tightly connected with a specific seismic risk analysis of common use within the insurance and re-insurance industry (CITATION). OpenQuake given an Earthquake Rupture Forecast can generate a number of seismicity histories, each one representing a possible realisation of the seismicity that the sources included in a Source Model can generate within a given time span (fixed by the user). The

Each rupture in a seismicity history is successively associated with a ground motion field, an object describing the spatial distribution of a scalar parameter representative of the intensity of shaking (e.g. PGA or Spectral Acceleration). OpenQuake has the capability to generate

7.1. Stochastic Event Set Calculator

The Inverse Trasform CITATION is a methodology widely used to randomly sample a discrete probabilistic distribution

7.2. Ground Motion Fields calculator

We generate ground motion fields using the

7.2.1. Spatially correlated ground motion fields

The genearation of spatially correlated ground motion fields is based on the work of Jayaram and Baker [2009] where they propose a model - based on a semivarigram -

for describing

$$\gamma(h) = a\left[1 - \exp\left(-\frac{3h}{b}\right)\right] \tag{7.1}$$

7.3. Stochastic PSHA calculator

The OpeQuake stochastic PSHA calculator provides a way to check the consistency between the result provided by a classical PSHA calculator and the results that can be obtained by a number of ground motion fields, representative of the

CHAPTER 8

Wesson et al. [2009] risk calculation implementation

In the procedure proposed by Wesson et al. [2009] the creation of the ERF follows the classical approach.

Using an ERF, for each rupture Rup is possible to calculate the probability that a given ground motion U is in the interval $u_x \pm \Delta u$ given an inter-event variability ϵ_{inter} (this corresponds to equation 3 of Wesson et al. [2009]):

$$\begin{aligned} P(u_x - \Delta u \leq U < u_x + \Delta u | Rup, \epsilon_{inter}) &= \Phi\left(\frac{(\ln(u_x - \Delta u) - \ln(u_0)}{\sigma_{intra}}\right) - \\ &- \Phi\left(\frac{(\ln(u_x + \Delta u) - \ln(u_0)}{\sigma_{intra}}\right) \end{aligned} \tag{8.1}$$

where $\ln(u_0)$ is the mean of the GMPE computed considering a value ϵ_{intra} and a rupture Rup, (generally characterized by a geometry and a magnitude) and Φ is the standard normal CDF.

The next step is to calculate the PMF of losses for a given rupture. Given an asset, the probability of suffering a loss value in the interval $[l - \Delta l, l + \Delta[$ given a ground motion value in the interval $[u_x - \Delta u, u_x + \Delta u[$ (note that in this case the distribution of ground motion u will depend on ϵ_{intra}) corresponds to:

$$\begin{aligned} P(l - \Delta l \leq L < \; & l - \Delta l | Rup, \epsilon_{inter}) = \\ \sum_{x=0}^{\infty} \; & P(l - \Delta l \leq L < l - \Delta l | u_x + \Delta u \leq U < u_x + \Delta u) \\ & P(u_x - \Delta u \leq U < u_x + \Delta u | R, \epsilon_{inter})) \end{aligned} \tag{8.2}$$

If $P^i(L = l|Rup, \epsilon_{inter})$ corresponds to the conditional probability mass function describing the discrete probability of having a loss in the interval $[l - \Delta l, l - \Delta[$ for the asset with index i, the probability of cumulated losses to a portfolio can be computed as:

$$P_{CL}(CL = cl|M, \epsilon_{inter}) = P^1(L = l|Rup, \epsilon_{inter}) * \ldots * P^n(L = l|Rup, \epsilon_{inter}) \tag{8.3}$$

where symbol $*$ stands for convolution.

Finally, the total probability of exceeding a given level of cumulated losses cl computed considering the contributions of all the ruptures occurring on all the seismic sources considered is (note that this expression extends equation A10 of Field et al. [2003]):

$$P(CL \geq cl) = 1 - \prod_{i=1}^{n} \left(1 - \sum_{n=1}^{N(i)} \sum_{k=-3}^{3} P(Rup_{i,n}) P(CL \geq cl | Rup_{i,n}, \epsilon_k) \right) \tag{8.4}$$

where $P(CL \geq cl|Rup_{i,n}, \epsilon_k)$ can be simply derived from the PMF $P_{CL}(CL = cl|M, \epsilon_{inter})$, n is the number of seismic sources.

PART III

Risk

CHAPTER 9

OpenQuake Input Description: Risk

The two main sources of input information required for a risk calculation with OpenQuake are an exposure model and a vulnerability model (in addition to the calculation type, such as those described in the subsequent chapters, and the region of interest). An exposure model for a given asset category describes, at each location of interest within a given region, the value of each asset typology. The vulnerability model describes the vulnerability characteristics of each asset typology.

9.1. Exposure

The OpenQuake engine requires an exposure model that needs to be stored according to the respective NRML schema. This file format can withstand several types of exposure elements such as population count, value of dwellings, building count among others. The following parameters are currently being used to describe each asset of the exposure model:

- Asset ID: A unique key used to identify the asset instance;
- Asset description: Brief description of the asset category;
- Asset value: Numerical value of the quantity of the asset;
- Location: Geographic coordinates of the asset expressed in decimal degrees.

This list of parameters will be further extended in future releases of OpenQuake once more complex data will need to be stored (e.g. value of contents or number of occupants per building).

9.2. Vulnerability

Vulnerability is defined as the probability distribution of loss, given an intensity measure level. Vulnerability functions can be derived directly, usually through empirical methods where the losses from past events at given locations are related to the levels of intensity of ground motion at those locations, or they can be derived by combining fragility functions and consequence functions. Fragility functions describe the probability of exceeding a set of limit states, given an intensity measure level; limit states describe the limits to performance levels, such as damage or injury levels. Fragility functions can be derived empirically (using observed data) or analytically, by explicitly modeling the behavior of a given asset typology when subjected to increasing levels of ground motion. Consequence functions describe the probability distribution of loss, given a performance level and are generally derived empirically. Version 0.2 of OpenQuake only supports vulnerability functions. However, the possibility to describe the vulnerability characteristics of the exposure assets with fragility and consequence functions is planned in OpenQuake such that users can view intermediate results of seismic loss calculations, such as the distribution of damage or injury levels.

9.2.1. Vulnerability Function

9.2.1.1. Discrete Vulnerability Functions

In the current version of OpenQuake (0.2) discrete vulnerability functions are used to directly estimate human and economic losses in OpenQuake. Discrete vulnerability functions are described by a list of intensity measure levels and corresponding mean loss ratios (ratio of loss to exposed value), associated coefficients of variation and probability distribution. The uncertainty on the loss ratio is assumed in OpenQuake v0.2 to follow a lognormal distribution, however different probabilistic distributions for the uncertainty will be developed in future versions such as the normal or beta distribution. Figure 9.1 illustrates a discrete vulnerability function.

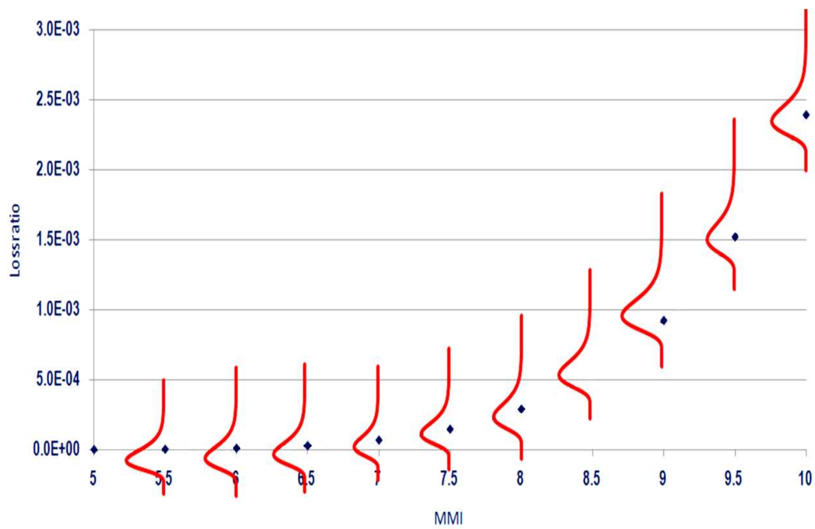

Figure 9.1.: Discrete vulnerability function.

9.2.1.2. Continuous Vulnerability Functions

In version 0.3 of OpenQuake, continuous vulnerability functions will be implemented. Continuous vulnerability functions are described by continuous distributions of mean loss ratio and coefficient of variation with ground motion intensity. The following figure illustrates this type of functions.

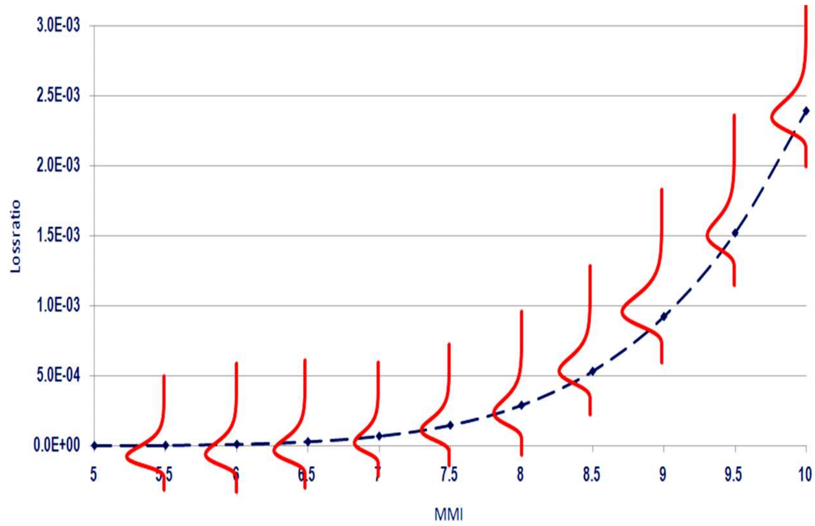

Figure 9.2.: Continuous vulnerability function.

9.2.2. Fragility Functions

Fragility functions describe the probability of exceeding a set of limit states, given an intensity measure level. When the asset category concerns structures (e.g. buildings), the intensity measure can either be structure-independent or structure-dependent. The former can be calculated directly from recorded measurements of ground shaking (e.g. peak ground acceleration, peak ground velocity, spectral acceleration at a given period of vibration, or even macroseismic intensity). The latter requires information on the characteristics of the structures in order to be calculated, for example spectral acceleration at the fundamental period of vibration, or spectral displacement at the limit state period of vibration. The calculation of these characteristics might be through a simple formula (e.g. a yield period-height equation, see e.g. Crowley and Pinho [2004]) or through so-called non-linear static methods, which are needed when the intensity measure is a non-linear response quantity such as spectral displacement at the limit state period of vibration (see e.g. FEMA-440:ATC [2005]). Discrete and continuous fragility functions with structure-independent intensity measures aim to be implemented in version 0.3 of OpenQuake. Fragility functions with structure-dependent intensity measures (and the methods necessary to calculate them) will be planned for the version 0.5 release.

9.2.2.1. Discrete Fragility Functions

Fragility functions can be defined in a discrete way by providing for each limit state a list of intensity measure levels and respective probabilities of exceedance. Figure 9.3 presents a set of discrete fragility functions using a macroseismic intensity.

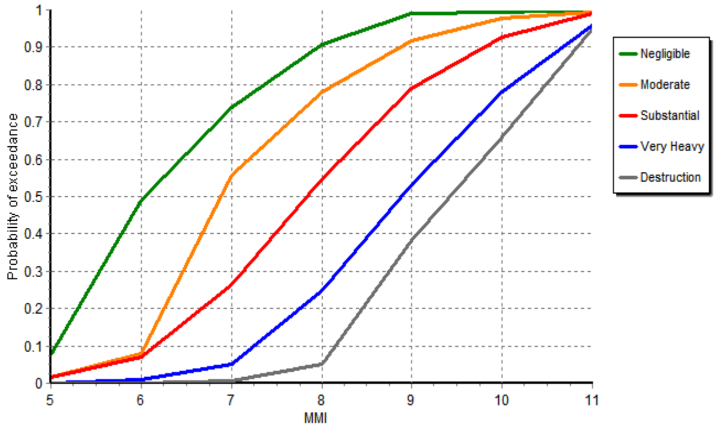

Figure 9.3.: Set of discrete fragility function.

9.2.2.2. Continuous Fragility Functions

Continuous fragility functions are defined by the parameters of a cumulative distribution function. The following figure presents an example of a set of continuous fragility functions with a structure-dependent intensity measure.

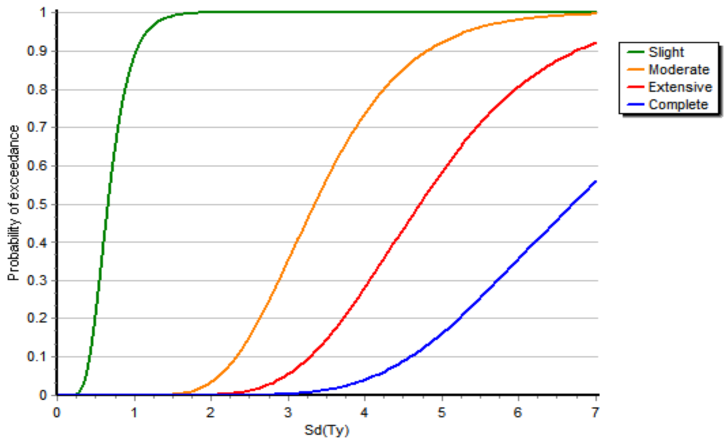

Figure 9.4.: Set of continuous fragility function.

9.2.3. Consequence Functions

Consequence functions describe the probability distribution of loss, given a performance level. For example, if the asset category is buildings and the performance level is significant damage, the consequence function will describe the mean loss ratio, coefficient of variation and probability distribution. The following figure presents the mean damage ratio for a set of performance levels proposed by two different sources:

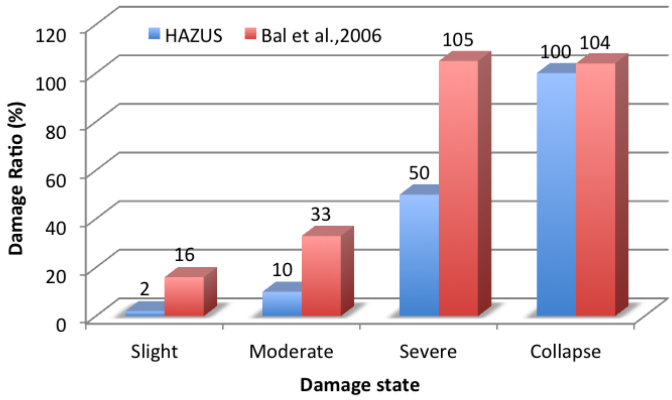

Figure 9.5.: Consequence functions adapted from Bal et al. [2010]

CHAPTER 10

Deterministic event based calculator

The deterministic event-based calculator is capable of computing losses and loss statistics for a single event for a collection of assets. Depending of the type of hazard input that is provided, two separate approaches can be followed. In the first one, the calculator uses two maps, one with the distribution of the mean ground motion and a second one with the associated aleatory variability. In the second approach, the event is repeated many times to model the variation in the inter-event variability and for each event, a ground motion field is generated taking into account the intra-event variability (and possibly the spatial correlation of the latter). The following scheme describes the architecture of this calculator:

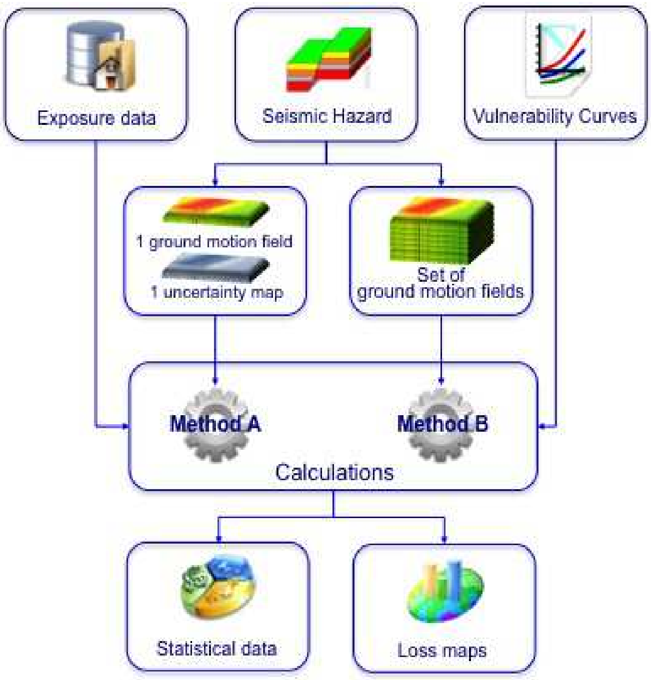

Figure 10.1.: Architecture of the deterministic event-based calculator.

10.1. Method A

10.1.1. Description

In this approach, after providing the two aforementioned maps, the mean ground motion and coefficient of variation at each site are used together with the assigned vulnerability function for each asset to calculate a mean loss ratio. The aleatory variability in the ground motion is combined with the uncertainty in the vulnerability functions through the total probability theorem in order to calculate the standard deviation of the loss ratio for each asset.

10.1.2. Calculations workflow

To compute the mean loss:

1. In order to compute the probability of occurrence of each intensity measure level defined on the vulnerability function, an upper and lower bound needs to be calculated for each value. The limits for an IML_n can be given by the following formulae:

$$Lowerbound = \frac{IML_n + IML_{n-1}}{2} \tag{10.1}$$

$$Upperbound = \frac{IML_{n+1} + IML_n}{2} \tag{10.2}$$

The following figure illustrates the intervals that were computed based on 4 intensity measure levels that compose a given vulnerability function:

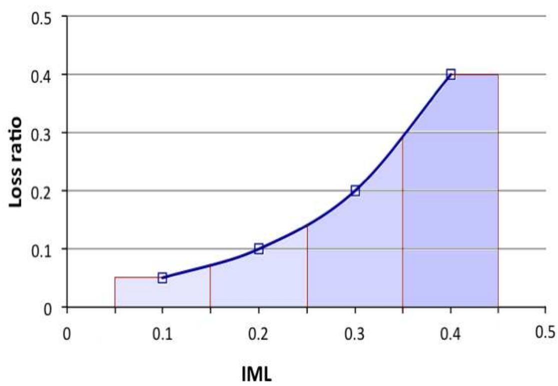

Figure 10.2.: Intervals for each intensity measure level on a discrete vulnerability function.

Note that for the first and last intensity measure level, there are not any value before or after respectively. In these cases, the lower and upper bound need to be compute based on the distance between the respective intensity measure level and the bound that was computed through one of the aforementioned formulae. Thus, the lower bound for the first intensity measure level can be given by:

$$Lowerbound = IML_1 - \frac{IML_2 + IML_1}{2} \tag{10.3}$$

And the upper bound for the last intensity measure level can be computed using:

$$Upperbound[IML_n] = IML_n + \frac{IML_n - IML_{n-1}}{2} \tag{10.4}$$

2. Once the intervals for each intensity measure level are defined, the probability of occurrence can be computed through the expression:

$$PO[IML_n] = F(UB, \mu, \sigma) - F(LB, \mu, \sigma) \tag{10.5}$$

Where F stands for the cumulative distribution function, UB and LB stand for the upper and lower bound respectively of the IML_n and μ and σ stand for the mean ground motion and associated standard deviation of the normal distribution of the logarithm of the ground motion values.

3. Then, the mean loss ratio for each asset can be computed through the formula:

$$LR = \sum_{n=1}^{m} PO[IML_n] \times LR_n \tag{10.6}$$

4. The absolute loss can be computed by multiplying the mean loss ratio by the value of the asset contained on the exposure model file.

To compute the standard deviation of the mean loss:

1. In order to compute this parameter, the total probability theorem needs to be used. The first step is to compute $E[LR_n^2]$, which is given by the following formula:

$$E[LR_n^2] = SD[LR_n]^2 + E[LR_n]^2 \tag{10.7}$$

Where $SD[LR_n]$ stands for the standard deviation of the distribution of loss ratios and $E[LR_n]$ stands for the mean loss ratio.

2. Then, the total $E[LR^2]$ can be derived using the formula:

$$E[LR^2] = \sum_{n=1}^{m} PO[IML_n] \times E[LR_n]^2 \tag{10.8}$$

3. Subsequently the standard deviation of the mean loss ratio can be computed using the expression:

$$SD[LR] = \sqrt{E[LR^2] - E[LR]^2} \tag{10.9}$$

   Where $E[LR]$ stands for the mean loss ratio computed previously.

4. The standard deviation of the mean loss can finally be computed by multiplying the standard deviation of the mean loss ratio by the value of the respective asset.

10.2. Method B

10.2.1. Description

In this approach, for each ground motion field, the intensity measure level at a given site is used to calculate the mean and standard deviation of loss ratio using the vulnerability functions for each asset contained in the exposure file. Using these results, the mean and standard deviation of loss ratio across all events can be calculated. Again, loss ratios are converted into losses by multiplying by the value of the asset given in the exposure file. For this method, it is possible to aggregate the losses throughout the region and to compute the standard deviation of the aggregated loss.

10.2.2. Calculations workflow

To compute the mean loss:

1. For each ground motion field, the intensity measure levels are related with the vulnerability functions to compute the loss ratio for each asset. Since currently the vulnerability functions are being defined in a discrete way, it is quite probable that the intensity measure level provided by the ground motion field is not contained in the vulnerability function. In these cases, linear interpolation methods are being employed to derived .

2. Once the loss ratios for all assets across all ground motion fields were computed, the mean loss ratio for each asset can be calculated through the formula:

$$LR = \frac{\sum_{n=1}^{m} LR_n|IML}{m} \tag{10.10}$$

Where m stands for the number of ground motion fields.

3. The mean loss can then be derived by multiplying the mean loss ratio for the value of the asset contained in the exposure model file.

To compute the standard deviation of the mean loss:

1. Again, the total probability theorem is employed. $E[LR_n^2]$ is computed using the following formula:

$$E[LR_n^2] = SD[LR_n]^2 + E[LR_n]^2 \tag{10.11}$$

Where $SD[LR_n]$ stands for the standard deviation of the distribution of loss ratios and $E[LR_n]$ stands for the mean loss ratio (for each ground motion field).

2. Then, the total $E[LR^2]$ can be derived using the formula:

$$E[LR^2] = \frac{\sum_{n=1}^{m} E[LR_n]^2}{m} \tag{10.12}$$

3. The standard deviation of the mean loss ratio can be computed using the expression:

$$SD[LR] = \sqrt{E[LR^2] - E[LR]^2} \tag{10.13}$$

Where $E[LR]$ stands for the mean loss ratio computed previously.

4. The standard deviation of the mean loss can finally be computed by multiplying the standard deviation of the mean loss ratio by the value of the respective asset.

CHAPTER 11

Classical PSHA based calculator

11.1. Description

The Classical PSHA-based risk calculator determines the loss exceedance curves for a collection of assets. This calculator takes the configuration information, one or several vulnerability models, information describing the assets and using hazard curves either contained in a file or computed using a PSHA input model, computes a loss exceedance curve for each asset. The following scheme illustrates the architecture of this calculator:

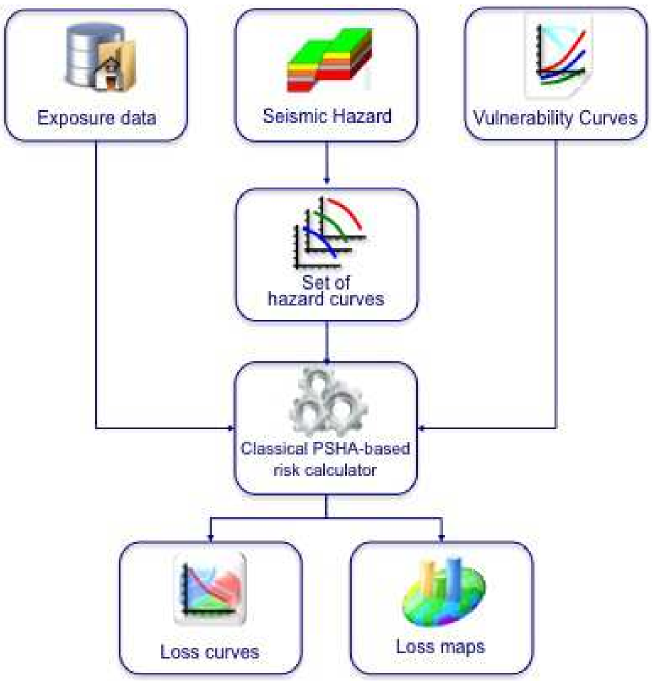

Figure 11.1.: Architecture of the Classical PSHA-based risk calculator.

11.2. Calculations workflow

1. By default, the hazard component of the OpenQuake engine computes the hazard curves for a set of intensity measure levels that are pre-defined in the configuration file. With the integration of the hazard and risk components in the engine, a feature was implemented with the purpose of verifying that this set of values covers the range of intensity measure levels defined in the vulnerability functions. If not, the set of values in which the hazard curves are going to be computed is extended based on the minimum and maximum values of the vulnerability functions.

2. To use this calculator, the hazard curves need first to be converted into probability mass functions (e.g. probability of occurrence of a discrete set of intensity measure levels). To do so, the engine starts by reading the intensity measure levels from the vulnerability function, and computing the middle value between consecutive levels. Each consecutive pair of values define an interval for each intensity measure level and relating the limits of each interval with the hazard curve, the engine computes the corresponding probabilities of exceedance. Figure 11.2 contains a vulnerability function (bottom chart) and a hazard curve (top chart) in which the definition of the interval for a given intensity measure level and associated estimation of the probabilities of exceedance of each limit are highlighted.

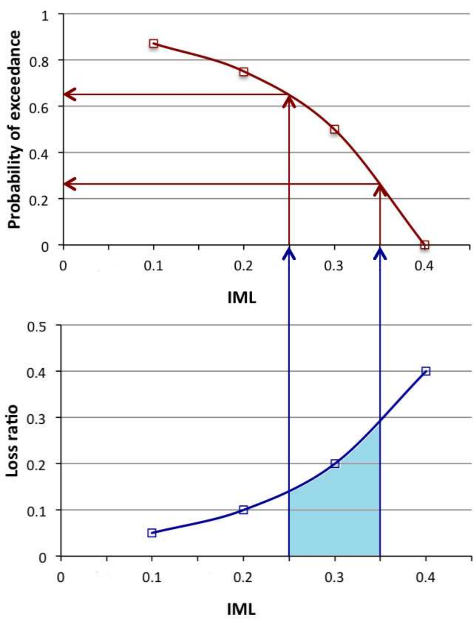

Figure 11.2.: Workflow to estimate the probability of exceedance of each interval bound.

3. The probability of occurrence of any level within each interval can be derived by subtracting the probabilities of exceedance of the lower and upper limits, just like described by he following formula:

$$PO = PE[lowerbound] - PE[upperbound] \tag{11.1}$$

4. The vulnerability functions for each asset are converted into loss ratio exceedance matrices (e.g. matrices which describe the probability of exceedance of each loss ratio for a discrete set of intensity measure levels). These matrices have a number of columns equal to the amount of intensity measure levels defined on the vulnerability function and a number of rows that can go from the amount of loss ratios, up to any multiple of this number. In order to properly incorporate the probabilistic distribution of loss ratios per intensity measure level, the probabilities of exceedance should be computed not just for the loss ratios defined on the vulnerability function, but also for many intermediate values between consecutive loss ratios. This is the reason why the number of rows could go from the amount of loss ratios defined on the vulnerability function, up to any value. Currently, the OpenQuake engine considers 5 intermediate values between consecutive loss ratios however, this is a parameter that can be easily adjusted. The following figure contains an example of a discrete vulnerability function and the respective loss ratio exceedance matrix (in light gray):

IML	LR	COV
0.1	0.05	0.5
0.2	0.08	0.3
0.4	0.20	0.2
0.6	0.40	0.1

	0.1	0.2	0.4	0.6
0.00	1.00	1.00	1.00	1.00
0.03	0.89	1.00	1.00	1.00
0.05	0.41	0.93	1.00	1.00
0.07	0.21	0.71	1.00	1.00
0.08	0.11	0.44	1.00	1.00
0.14	0.01	0.02	0.96	1.00
0.20	0.00	0.00	0.46	1.00
0.30	0.00	0.00	0.02	1.00
0.40	0.00	0.00	0.00	0.48
0.70	0.00	0.00	0.00	0.00
1.00	0.00	0.00	0.00	0.00

Figure 11.3.: Example of a discrete vulnerability function and respective loss ratio exceedance matrix.

Note that for this example only one intermediate value was considered between consecutive loss ratios and in order to consider the whole distribution of the loss ratios, the matrix was computed considering a minimum and maximum loss ratio of 0 and 1 respectively.

5. Finally, each column of the aforementioned matrix is multiplied by the probability of occurrence of the respective intensity measure level (extracted from the hazard curves) to produce a conditional loss ratio exceedance matrix. Then, for each loss ratio the probabilities of exceedance are summed, leading to a loss ratio exceedance curve, whose set of loss ratios can be multiplied by the value of the asset given by the exposure file to obtain a loss curve.

CHAPTER 12

Probabilistic event based calculator

12.1. Description

This method uses stochastic event sets and associated ground motion fields to compute loss curves for each asset contained in an exposure file, as illustrated in the following scheme:

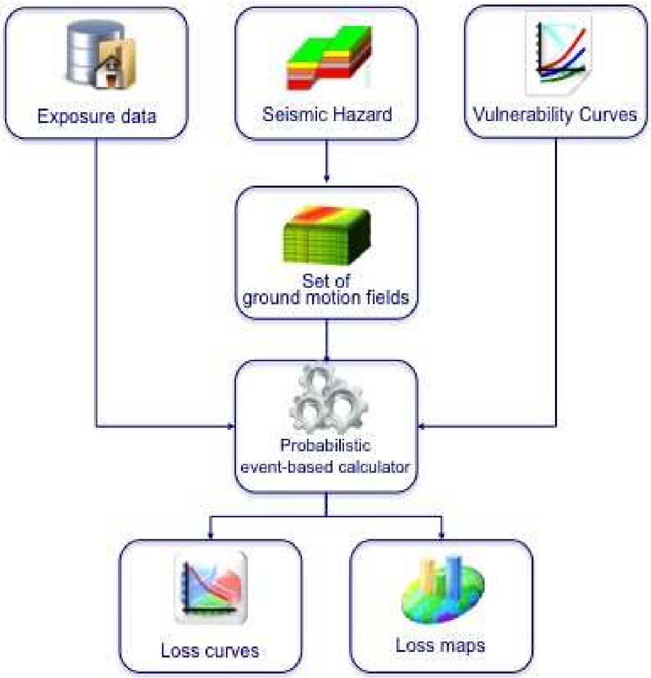

Figure 12.1.: Architecture of the probabilistic event-based calculator.

For each ground motion field, the intensity measure level at a given site is used to calculate the mean loss ratio using the vulnerability functions for each asset defined in the exposure file. The occurrence distribution of mean loss for a given asset is calculated using all of the ground motion fields, leading to a histogram of loss ratios which is then converted into a cumulative histogram, by calculating the number of cumulative occurrences for each interval of loss ratio. The rate of exceedance of each loss ratio is calculated by dividing the number of cumulative occurrences by the number of stochastic event sets multiplied by the length of each event set. By assuming a Poissionian distribution of the occurrence model, the probability of exceedance of each loss ratio is calculated.

If an aggregated loss curve for a portfolio of assets is required, a secondary module is required in order to aggregate the losses from all the assets in the exposure file, per event, before calculating the occurrence distribution of mean loss. If the assets are close enough, it is necessary to generate the ground motion fields taking into account the spatial correlation of ground motion residuals.

12.2. Calculations workflow

1. The engine starts by using the set of ground motion fields to extract the intensity measure levels for the location of each asset.

2. Then the engine takes the vulnerability function assigned to each asset and checks if the uncertainty was specified. If not, the loss ratios are derived based on the mean loss ratio for each intensity measure level. However, if the uncertainty is defined, the loss ratios are sampled following the probabilistic distribution, mean loss ratio and associated coefficient of variation of the respective function, as described bellow:

$$\log LR_n = \mu + \epsilon\sigma \tag{12.1}$$

   Where μ and σ stand for the mean and standard deviation of the logarithm of the loss ratios respectively and ϵ is a term that has a standard normal distribution with a zero mean and a standard deviation of one.

   The method used to sample this parameter can follow two approaches in order to consider the correlation between the vulnerability of similar building typologies:

   - Perfectly correlated: the term ϵ is randomly sampled once for the first asset and this result is used to derive the loss ratio for all the assets.
   - Uncorrelated: the term ϵ is always randomly sampled for each asset and therefore the correlation between the vulnerability of the assets is ignored.

3. In this method an histogram of the loss ratios per asset is required. Before the histogram can be built, it is necessary to define the number and width of

the bins. The former might vary significantly since it might depend of several factors (e.g. number of ground motion fields, range of ground motion covered by the vulnerability model) while the later is related with the minimum and maximum values of loss ratio previously computed and with the number of bins.

4. The histograms for each asset need to be converted into a cumulative histogram. The number of occurrences for each bin can be derived using the following formula:

$$NCO_m = \sum_{n=m} NO_n \tag{12.2}$$

where NCO_m stands for the number of cumulative occurrences of the m^{th} bin of the cumulative histogram and NO_n stands for the number o occurrences of the n^{th} bin of the histogram of the loss ratios.

5. Thereafter, the rate of exceedance of a set of loss ratios needs to be computed for each asset. This set of loss ratios is composed by the middle values of each bin of the cumulative histogram. The following formula is employed to compute this rate:

$$\lambda(LR_n) = \frac{NCO_n}{TSES} \tag{12.3}$$

Where λ stands for the rate of exceedance of the respective loss ratio and $TSES$ stands for the time representative of the stochastic event set which means, the number of stochastic event sets multiplied by the time span of each one.

6. Assuming a poissonion distribution of the occurrence model, the probability of exceedance of the set of loss ratios can be derived using the following formula:

$$PE(LR_n) = 1 - \exp -\lambda_n \times t \tag{12.4}$$

Where t stands for the time span used to produce the stochastic event set.

PART IV
Socio-Economic Impact Assessment

Introduction to the Socio-Economic Impact Assessment

PART V
Modeller's Toolkit

CHAPTER 13
Introduction

Introduction to the hazard input Modellers' Toolkit

CHAPTER 14

Input visualization and preparation

14.1. Hazard

14.2. Risk

14.3. Socio-Economic Impact

PART VI
Appendixes

APPENDIX A
Example of OpenQuake hazard calculation configuration file

This is a test

APPENDIX B

Example of OpenQuake risk calculation configuration file

Bibliography

Aki, K. and Richards, P. G. (2002). *Quantitative Seismology*. University Science Books, Sausalito, California.

Bal, I., Crowley, H., and Pinho, R. (2010). *Displacement-Based Earthquake Loss Assessment: Method Development and Application to Turkish Building Stock*. PhD thesis, Centre for Post-Graduate Training and Research in Earthquake Engineering and Engineering Seismology.

Bender, B. and Perkins, D. M. (1987). Seisrisk III: A computer program for seismic hazard estimation. Bulletin 1772, United States Geological Survey.

Bommer, J. J. and Scherbaum, F. (2008). The Use and Misuse of Logic Trees in Probabilistic Seismic Hazard Analysis. *Earthquake Spectra*, 24(4):9971009.

Chiang, W. L., Guidi, G. A., Scoof, C. G., and Shah, H. C. (1984). Computer Programs for Seismic Hazard Analysis - A User Manual (STASHA). Report 62, The J. A. Blume Earthquake Engineering Center.

Coppersmith, K. J., Youngs, R. R., and Sprecher, C. (2009). Methodology and main results of seismic source characterization for PEGASOS Project, Switzerland. *Swiss J. Geosc.*, 102:91–105.

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

Crowley, H. and Pinho, R. (2004). Period-height relationship for existing european reinforced concrete buildings. *Journal of Earthquake Engineering*, 8:93–120.

FEMA-440:ATC (2005). Improvement of nonlinear static seismic analysis procedures,. Technical report, California, USA.

Field, E. H., Jordan, T. H., and Cornell, C. A. (2003). OpenSHA - A developing Community-Modeling Environment for Seismic Hazard Analysis. *Seism. Res. Lett.*, 74:406–419.

Frankel, A. (1995). Mapping Seismic Hazard in the Central and Eastern United States. *Seismological Research Letters*, 66(4):8–21.

Harmsen, S. (2008). Appendix C. Distance to a Fault with Random Strike. In 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.

Jayaram, N. and Baker, J. W. (2009). Correlation model for spatially distributed ground-motion intensities. *Earthquake Engineering and Structural Dynamics*.

Matthews, M. V. and Ellsworth, W. L. Reasenberg, P. A. (2002). A brownian model for recurrent earthquakes. *Bulletin of the Seismological Society of America*, 92(6):2233–2250.

McGuire, R. K. (2004). *Seismic Hazard and Risk Analysis*. EERI.

Pagani, M., Monelli, D., Crowley, H., Danciu, L., Field, E. H., Wiemer, S., and Giardini, D. (2010). GEM1 Hazard: Description of Input Models, Calculation Engine and Main Results. GEM Technical Report 2010-3, GEM Foundation, Pavia, Italy.

Senior Seismic Hazard Analysis Committee (1997). Recommendations for probabilistic seismic hazard analysis: Guidance on uncertainty and use of experts. Technical Report NUREG/CR-6372, UCRL-ID-122160, Vol. 1, Lawrence Livermore National Laboratory.

Wesson, R. L., Perkins, D. M., Luco, N., and Karaca, E. (2009). Direct calculation of the probability distribution for earthquake losses to a portfolio. *Earthquake Spectra*, 25(3):687–706.

Woo, G. (1996). Kernel estimation methods for seismic hazard area source modeling. *Bulletin of the Seismological Society of America*, 86(2):353–362.