Training and Communication for Earthquake Risk Assessment (TREQ) Project
Deliverable 3.1.1a

Probabilistic Seismic Hazard Analysis (PSHA) Training Manual

Version No. 1.0.0

Global Earthquake Model (GEM) Foundation

An example-based guide to building PSHA models using open-source data and tools

www.globalquakemodel.org

Authors

Kendra Johnson [1], Marco Pagani[1]

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

Email address for all current authors:
<firstname.lastname>@globalquakemodel.org

Citation

Please cite this document as:
GEM Foundation (2021). *Probabilistic Seismic Hazard Analysis (PSHA) Training Manual. doi: 10.13117/OQ_HAZARD_TRAINING_MANUAL*

This manual is made possible by the support of the American People through the United States Agency for International Development (USAID.) The contents of this study are the sole responsibility of the GEM Foundation and do not necessarily reflect the views of USAID or the United States Government.

Contents

Preface

This document is an example-based instructional manual that teaches how to construct a simple seismic hazard model. The intended audience is the novice hazard modeler at university level or higher, and thus instructs on the basic state-of-practice. The methodologies presented within are meant to be accessible to a wide range of users, and so an emphasis is given to openly available data sets and tools. The provided examples use public data sets, and toolkits developed by the GEM Foundation.

The users of this manual will benefit from introductory level prerequisite knowledge of the following topics:

- probability theory
- earthquakes and tectonics
- geospatial information system (GIS)
- beginner- to intermediate-level Python

The instruction provided herein is complemented by several available resources, which we aim to build on but not replace. In particular, we refer to earlier articles and manuals that explain in greater detail many of the concepts or algorithms included in this training material. These resources are:

- *Seismic Hazard and Risk Analysis* (McGuire, 2004)
- statistical seismology training prepared by CORSSA: the Community Online Resource for Statistical Seismicity Analysis
- Introduction to Probabilistic Seismic Hazard Analysis, by Jack W. Baker
- OpenQuake Engine Underlying Hazard Science
- the OpenQuake Engine User Instruction Manual
- the Hazard Modeller's - Toolkit User's Guide

The development of this training material was funded by USAID through the Training and Communication for Earthquake Risk Assessment (TREQ) project. The development efforts were led by the Hazard Team of the Global Earthquake Model Foundation, a public-private partnership initiated by the Global Science Forum of the Organisation for Economic Co-operation and Development (OECD)[1].

[1] A short description of the process promoted by OECD is available here: http://www.oecd.org/science/sci-tech/theglobalearthquakemodelgem.htm

Acronyms

r_{rup}
closest distance between the site and rupture.

AFE
annual frequency of exceedence.

GAF-DB
Global Active Faults Database.

GEM
Global Earthquake Model Foundation.

GHEA
Global Historical Earthquake Archive.

GMC
Ground Motion Characterization.

GMM
Ground Motion Model.

GMPE
Ground Motion Prediction Equation.

GMPE-SMTK
GMPE Strong Motion Modeller's Toolkit.

HMTK
Hazard Modeller's Toolkit.

IASPEI
International Association of Seismology and Physics of the Earth's Interior.

ISC
International Seismological Center.

ISF
IASPEI Seismic Format.

MFD
Magnitude-Frequency Distribution.

oq-engine
OpenQuake-engine.

OQ-MBTK
OpenQuake Model Building Toolkit.

PGA

Peak Ground Acceleration.

PGV

Peak Ground Velocity.

PSHA

Classical PSHA.

SA

Spectral Acceleration.

SSC

Seismic Source Characterization.

SSM

Seismic Source Model.

TRT

tectonic region type.

UHS

uniform hazard spectrum.

Part I

Introduction

1 Introduction to seismic hazard analysis

1.1 What is a seismic hazard analysis?

Seismic hazard analysis is a procedure used to determine the level of danger that may occur due to earthquakes within an area of interest over a given time frame. The two major types are determinstic seismic hazard analysis, which is interested in the worst case event that can be expected at a site, and probabilistic seismic hazard analysis, which calculates the rates at which each shaking level will occur or the probability that each level will be exceeded during a fixed time frame.

This instructional manual teaches classical probabilistic seismic hazard analysis (PSHA), and how to construct the underlying input models. Classical PSHA (Cornell, 1968) uses the hazard integral, which considers all possible seismic sources expected to affect a site and their rates (or, alternatively, probabilities) of occurrence, and the expected ground motions that will occur for earthquakes produced by these sources. The hazard integral evaluates the probability that a specified ground motion intensity level will be exceeded over an investigation period or the frequency at which the intensity level is exceeded.

A number of software options are available for developing seismic hazard models and executing Classical PSHA calculations. This manual uses the OpenQuake Engine (Pagani et al., 2014) and its accompanying toolkits.

1.2 Brief introduction to the OpenQuake-Engine

The OpenQuake Engine is a seismic hazard and risk calculation software that's developement and maintainence is led by the Global Earthquake Model Foundation (GEM) (Pagani et al., 2014; Silva et al., 2014). The Engine is written in Python and supported on Linux, MacOS, Windows, and cloud platforms. The software emerged from a 2010 project aiming to develop an open and transparent hazard and risk calculation engine with community-driven development, and includes contributions from many of GEM's collaborators. The Engine's development emphasises reproducibility, and thus is founded on several testing approaches including unit tests, comparisons against other PSHA codes and benchmarks (e.g. PEER, Hale et al., 2018), and comparisons of outputs from other PSHA codes using end-to-end calculations.

The Engine is hosted on GitHub (https://github.com/gem/oq-engine).

Additional reading: Pagani et al. (2014) and Silva et al. (2014), the OpenQuake Engine User Instruction Manual, and the OpenQuake Underlying Hazard Science

1.3 Brief introduction to OpenQuake Hazard Toolkits

The examples included in this training manual use three toolkits: the Hazard Modeller's Toolkit (HMTK), the OpenQuake Model Building Toolkit (OQ-MBTK), and the GMPE Strong Motion Modeller's Toolkit (GMPE-SMTK). Like the OpenQuake Engine, all three toolkits are Python-based. The HMTK is built into the Engine, while the other two toolkits must be installed separately.

1.3.1 HMTK

The HMTK is a Python library that provides tools that assist in developing seismic hazard models, and in particular seismic source models, in the OpenQuake Engine format. The toolkit does not eliminate the requirement of modeller judgement in the hazard modelling process, but provides modellers with implmentations of many state-of-practice algorithms used to develop seismic source models.

The HMTK is the most established of the toolkits used in this training manual. As a component of the OpenQuake Engine, it is thoroughly tested. It is accompanied by a users manual available here.

1.3.2 OQ-MBTK

The OQ-MBTK is a container for a number of separate packages that are useful for building seismic hazard models and analysing their performance. The two OQ-MBTK pacakges most relevant for this training manual are the catalogue toolkit (CAT) that assists modellers in preparing the earthquake catalogue to be used in the model-building process and the model-building toolkit (MBT) that streamlines the use of the HMTK for preparing seismic source models inputs in the OpenQuake Engine format.

The OQ-MBTK can be downloaded at https://github.com/GEMScienceTools/oq-mbtk.

1.3.3 GMPE-SMTK

The GMPE-SMTK is a suite of Python tools that helps hazard modellers to select GMPEs (ground motion prediction equations, also called ground motion models) to be used in a seismic hazard model. The toolkit is based upon the GMPEs included in the OpenQuake Engine.

The GMPE-SMTK can be downloaded at https://github.com/GEMScienceTools/gmpe-smtk.

2 Components of a seismic hazard model

A seismic hazard model consists of two primary components — the source characterization, ground motion characterization — and their corresponding uncertainties (Fig. 2.1). Together, these components capture the expected earthquakes and the predicted shaking that will result from these earthquakes.

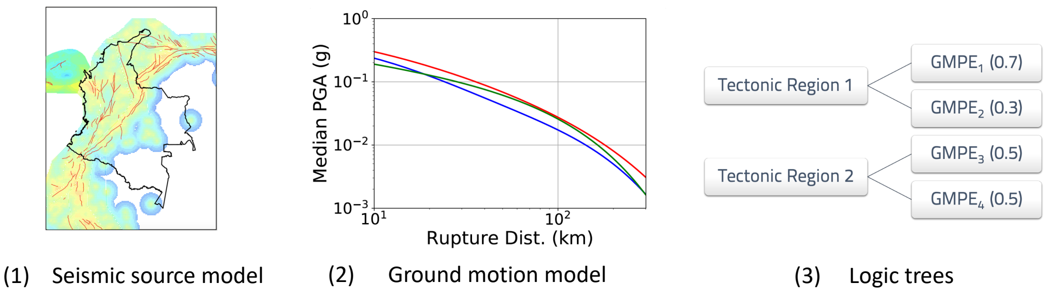


Figure 2.1 – The components of a PSHA input model. (1) An example of a seismic source model: a collection of seismic sources that together generate all the seismicity affecting an area of interest. (2) An example of a ground-motion model, showing that the ground-motion value attenuates with distance. (3) Example of a logic tree structure used to incorporate uncertainty in components (1) and (2).

2.1 Seismic source characterization

Seismic source characterization is the description of earthquake occurence and the associated epistemic uncertainties. Traditionally, a seismic source model SSM defines earthquake occurrence, containing the locations, sizes, geometries, and frequencies of all observed and anticipated earthquakes that will affect the sites of interest. The procedure for building an SSM can be summarized by the following steps:

1. Identify all the seismic sources that may produce damaging ground motions in the region of interest
2. Characterize the spatial distribution of the earthquake sources, assigning geometry and position to each source
3. Characterize the magnitude distribution of each earthquake source, assigning each source a magnitude-dependent rate of occurrence

Figure 2.2 shows an example of a seismic source model.

Epistemic uncertainty in the SSC can be described by using a number of SSMs, each one with an associated weight, or by some initial SSMs and a set of modifications that describe the various epistemic uncerataintes of source parameters.

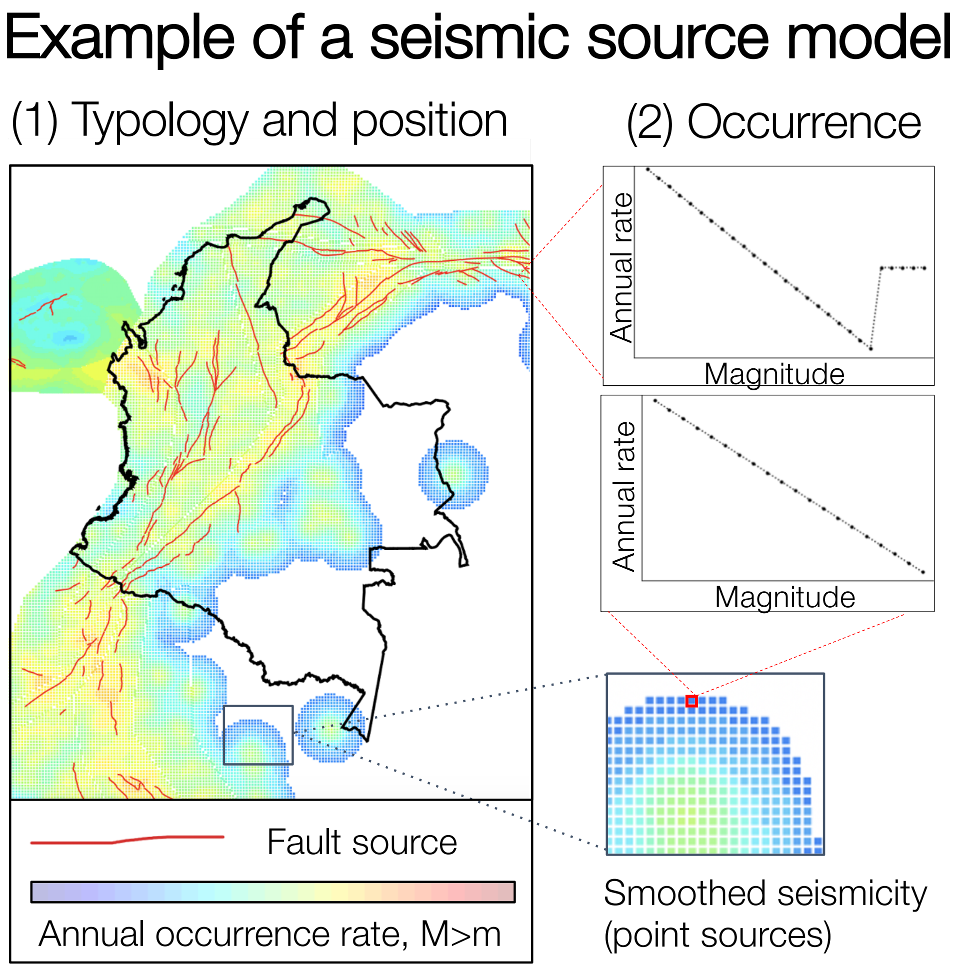


Figure 2.2 – A simplified example of a seismic source model that shows the position, source typology, and magnitude frequency distributions of sources covering a region of interest. The left panel shows two source typologies: faults, with surface traces denoted by red lines, and point sources colored by the cumulative annual occurrence rate of earthquakes with magnitude $M > m$, where m is a minimum magnitude threshold. The top of the right panel shows two examples of magnitude frequency distributions, which indicate the occurrence rates of sources denoted by the dashed lines. (NB: this is a purely hypothetical source model.)

Hazard modellers use the following information and considerations to complete seismic source characterization.

- *Recorded seismicity:* do instrumental and historical earthquake catalogues contain events in this region? How frequent are recorded earthquakes, and what are their sizes?
- *Distribution of observed seismicity*: are patterns observed in the occurring seismicity? Do observed earthquakes tend to concentrate in delineable areas/volumes (in terms

of both lateral and depth distribution), or are hypocenters more sparsely distributed? How are the earthquakes distributed throughout time? What is the ratio of large to small earthquakes?

- *Tectonic setting/geology:* what is known about the tectonic activity in the area covered by the model? For example, is it close to a plate boundary, or in the stable continental interior? If it's close to a plate boundary, what kind of earthquakes does that boundary produce? What are the relative plate motions?
- *Active faults:* other than plate boundary faults, does the region include other known faults, and are they thought to be seismically active? Are they covered by paleoseismic or geodetic data that could reveal their activity rates?

2.2 Ground motion characterization

Ground motion characterization provides everything that is needed to compute the ground motion at a site given an earthquake rupture. In the simplest case, this means the creation or selection of one ground-motion model, i.e. a Ground Motion Prediction Equation (GMPE). These models predict the level of shaking for the range of spectral periods expected to impact the human environment.

Hazard modellers use the following information and considerations during ground motion characterization:

1. *Tectonic region type:* do the seismic sources included in the model occur in cold, dense, stable crust, or are they in more active environments with warmer, less dense underlying crust, where seismic waves attenuate more rapidly?
2. *Ground motion records:* do seismic waveforms, strong motion records, or intensity data exist for past earthquakes in the region (either for earthquakes with local hypocenters, or local stations recording distant earthquakes)
3. *Existing ground motion models:* have any ground motion models been developed specifically for the region of interest? Which models cover the tectonic region types that the model encompasses? Which ground motion models cover the ranges of magnitude and source-to-site distance in the model?

Figure 2.3 shows an example of the components of a ground motion characterization workflow.

2.3 Logic trees

Logic trees are structures used to incorporate epistemic uncertainty into hazard models (e.g. Figure 7.1). Epistemic uncertainty is the type of uncertainty that results from our limited knowledge or data about a phenomenon that we try to model. Logic trees can be applied in both the source and ground motion characterizations in order to include multiple hypotheses that encompass the range of possible values for uncertain parameters or methods (e.g.

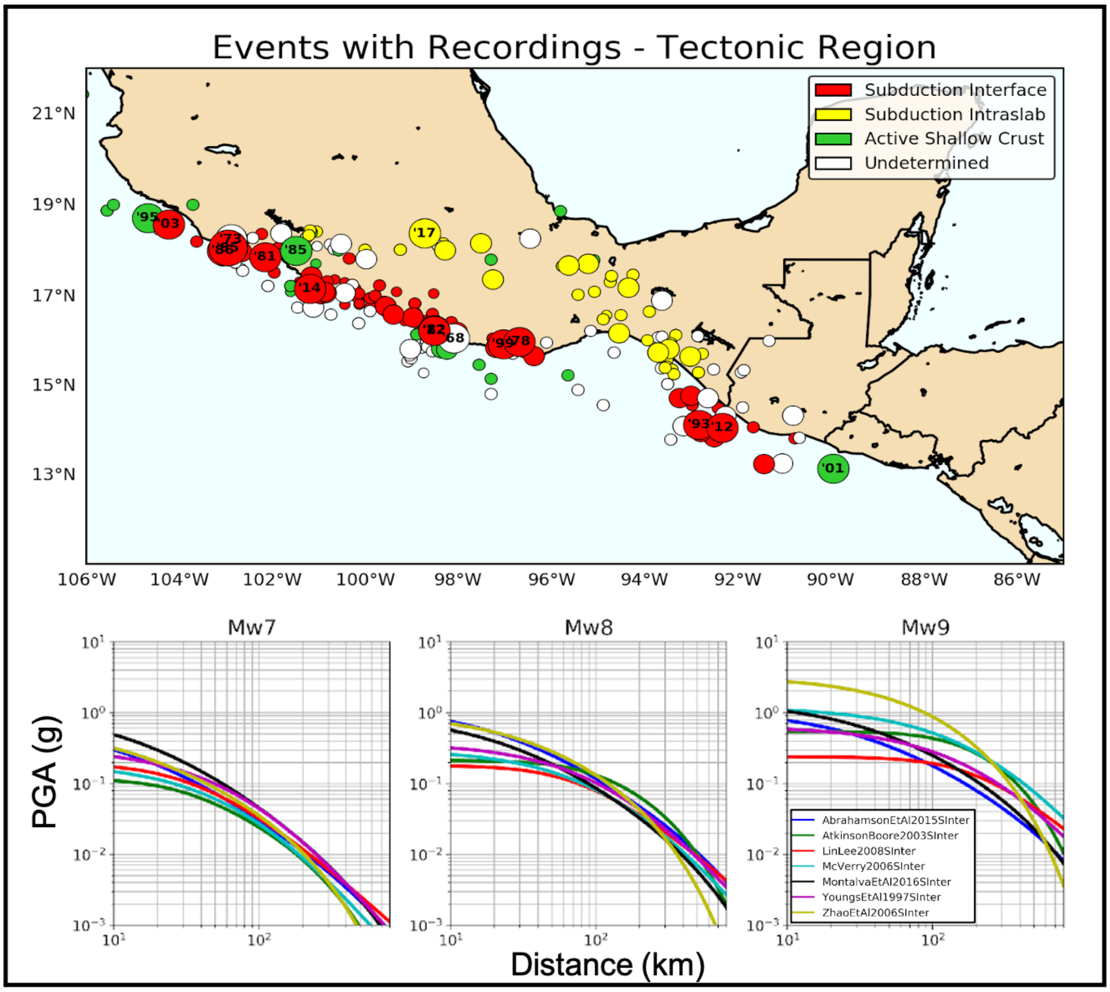


Figure 2.3 – Components of the ground motion characterization workflow. Top: Locations of earthquakes near Mexico for which strong motion data are available, colored by tectonic region type and scaled by magnitude. Bottom: Plots showing how the computed peak ground acceleration (PGA) changes with distance for three different magnitudes using seven existing ground motion models for subduction interfaces.

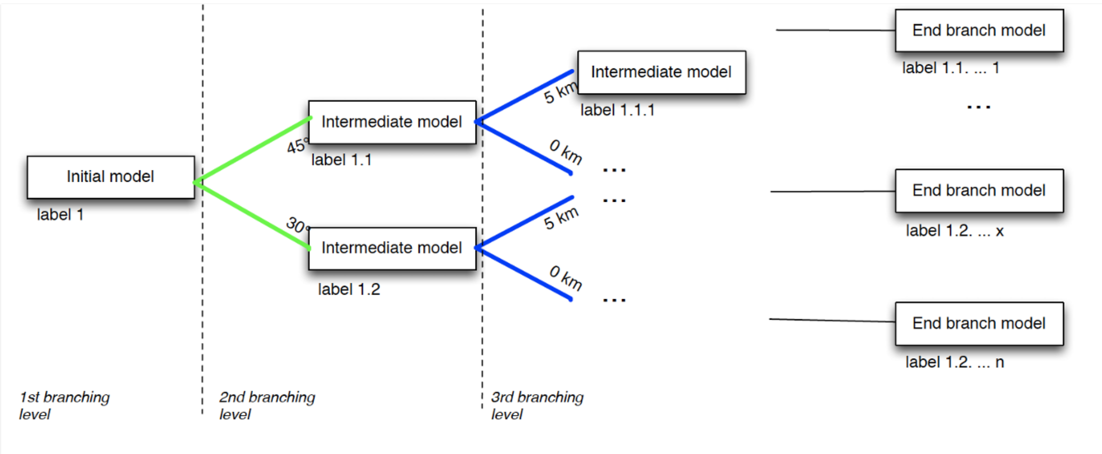


Figure 2.4 – An example of a logic tree, with labels corresponding to the terminology of the OpenQuake Engine (Pagani et al., 2014). Each branching level introduces an additional epistemic uncertainty to the input model.

declustering). In the case of the source characterization, epistemic uncertainties include, for example, the seismic source delineation and geometry, maximum magnitude admitted, the parameters of the magnitude-frequency distribution, and any other parameter that cannot be reduced to a single hypothesis. In the case of the ground motion characterization, the epistemic uncertainties include the regional inter- and intra-event variability of ground motions for a given magnitude-distance pair, which prohibits us from fitting recorded seismic recordings for a tectonic region using a just single ground motion model, and data gaps (or general lack of data), such as for high magnitudes, or at long and short site-source distances.

Part II

Seismic source characterisation

Seismic source characterization is the definition of one or more seismic source models that represent all earthquake sources that could produce seismicity within the model coverage area during an investigation period and the associated epistemic uncertainties. In this chapter, we introduce:

1. types of data that are used in seismic source characterization, including datasets thats that are open-source and commonly used for PSHA
2. the source typologies that are most frequently used to represent seismic sources
3. a seismic source characterisation workflow, from catalogue processing to defining seismic sources

3 Basic data

This chapter introduces three types of data that are commonly used to build seismic hazard models:

- Earthquake catalogues
- Fault data
- Geodetic data or information

3.1 Earthquake catalogues

3.1.1 What are earthquake catalogues?

Earthquake catalogues are compilations of past earthquakes and their parameters. Each entry includes the basic information about an earthquake (often called an event) on record, at minimum including the earthquake time, magnitude, and hypocenter. Earthquake catalogues can be categorized spatially and temporally.

Temporal categorization: There are three categories of earthquake catalogue divided based on time frame, and discussed here from most recent to oldest records. (1) Instrumental catalogues include events for which the event information is obtained by processing data recorded by seismometers. (2) Historical catalogues include events from before the time of seismometers, but that were felt by humans. Event information is mostly retrieved from anecdotal evidence about an earthquake's impacts (e.g., intensity data), but may also include analysis of waveforms recorded from the earliest seismographs. (3) Pre-historic catalogues include events from even earlier, which are revealed through paleoseismic trench or other geological investigations. The data in instrumental catalogues is considered to have more accurate parameters and a more complete record, e.g., including more instances of earthquakes above a lower magnitude threshold, than the historical catalogues, which are more complete than pre-historic catalogues. Thus, instrumental catalogues also include the most earthquakes per unit of time.

The first instantiations of instrumental catalogues are typically produced algorithmically; that is, automated codes derive earthquake parameters from seismometers that delivered data in real-time from network seismic stations to the network data center. Depending on the network, different thresholds and priorities are used to determine which events will also be reviewed by a human analyst.

Spatial categorization: Spatially, earthquake catalogues can be categorized as (1) global, (2) regional, and (3) local; these categories are intuitive as to the spatial ranges they cover. Global catalogues are produced from seismic networks that have long-period seismometers distributed across the globe, positioned densely enough to detect all earthquakes with M_W $\sim$ 5.5 and larger; or are collections of regional and/or local catalogues. Regional catalogues(>100 km) and local catalogues (<100 km) have high station density, but covering only sections of the globe. Thus, they have the advantage of a lower magnitude completeness threshold, but only applying to a small area.

In addition to origin time, magnitude, and location, some catalogues also include focal mechanism parameters, providing either the nodal plane geometries or moment tensors.

3.1.2 Utility in seismic hazard analysis

Earthquake catalogues are useful for identifying and delineating the geometry of seismic sources to be used in seismic hazard analysis. Instrumental catalogues in particular are useful for assigning earthquake rates to sources for seismic hazard analysis using statistical analysis; however, they show only a small snapshot (e.g., 10s of to $\sim$ 100 years) of time rather than averaged long-term rates. Historic and pre-historic catalogues can be helpful for constraining longer-term rates (e.g., 1000s to 10000s of years) of high magnitude seismicity; however, the spatial and temporal completeness of this information can be rather uncertain.

Catalogues that include focal mechanism parameters (e.g., centroid moment tensors or nodal planes) are useful for identifying the primary faulting types within an area of interest, and assigning these to the seismic sources of a hazard model.

Figure 3.1 shows a map of earthquakes in the ISC-GEM extended catalogue (Weatherill et al. (2016); see Section 3) for the Aegean Sea region.

3.2 Fault data

3.2.1 Fault investigations

Targeted investigations of faults can improve our understanding of their geometries and activity histories. This text will not instruct on executing detailed fault investigations, but on the types of investigations used to recover fault characteristics that are needed to construct a seismic source for hazard analysis: (1) the magnitude potential of the fault, (2) evidence of earthquake timing, and (3) indicators of fault dip direction and sense of motion.

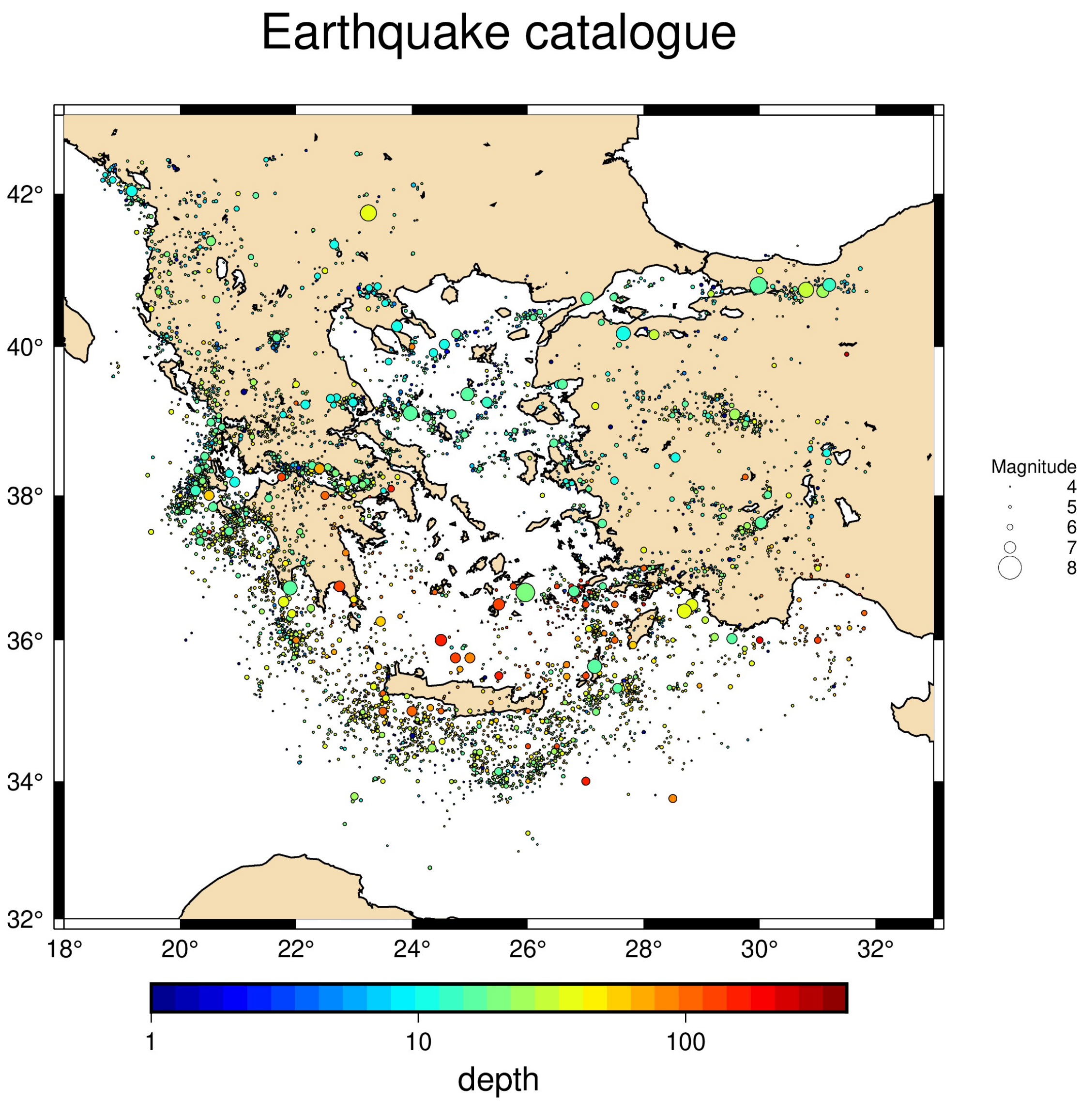


Figure 3.1 – Example of an earthquake catalogue: the ISC-GEM extended catalogue coverage of the Aegean Sea region Weatherill et al., 2016. Hypocenters are colored by depth and size is scaled by magnitude.

All earthquakes occur on faults, discontinuities within the Earth's crust where rock has been crushed or fractured, separating the surrounding earth into separate blocks. Faults range in area from very small (mm^2-scale) to 100s or 1000s of km^2. Some faults are known to scientists, while others are too small to detect; are buried deep underground without surface expressions; or have not ruptured historically. Faults that are thought to be active should be included in seismic hazard models. While this is mostly straightforward in the case of plate boundaries, crustal faults can be more ambiguous, often due to a shortage of data. For this reason, faults are not always included in seismic hazard assessments, even if we know that they are active. This section introduces the fault characteristics that are critical to constructing fault sources, and the ways of collecting these data.

Geological surveys from around the world typically depict the state of known faulting on geologic maps for their regions of oversight, sometimes collecting the faults into a separate database that includes the information available for each. In these databases, each fault includes at minimum a fault trace, which is the curve representing the intersection of the fault surface with the topographic surface. In order to convert a fault into a seismic source for PSHA, we must also assign it a geometry and activity rate. These parameters can be assigned by studying historical earthquakes, and by conducting remote sensing and field investigations.

Field studies of faults known to host past earthquakes

Historical earthquakes that are large enough to be linked to a specific fault (or faults), and in particular those that produce surface ruptures (e.g., offset at the ground surface), indicate currently active faults. Sometimes these earthquakes occur on faults that are known to be active, but other times are the first indicator of ongoing fault activity, or the first evidence that the fault exists. Recent earthquakes that appear in instrumental catalogues are typically studied with a variety of methods – field mapping, seismology, remote sensing, GPS – that provide a 3D image (the geometry) of the causative faults, as well as a robust value of moment magnitude; many of these topics are discussed in the next sections. Studies of earthquakes that occurred earlier in history, before the instrumental period but still observed by humans (those appearing in historical catalogues) and that can be linked to a known fault, can still yield magnitude estimates, both using the surface rupture qualities and by inverting the observed damage field.

Independently, the magnitudes of past earthquakes can only be used to constrain the maximum magnitude of a source, serving as a known *minimum* value of the largest earthquake that a fault can accommodate. The timing of the event can be used to make very coarse estimates of the activity rate, e.g., using the time that has elapsed since the earthquake, and the time that passed between the arrival of the first humans and the earthquake. Typically, these time scales can only indicate the minimum recurrence interval, the period between consecutive events of the given magnitude; the inverse of the recurrence interval is the activity rate. These parameters can be computed more robustly by using data collected from targeted fault investigations using approaches of geology, paleoseismology, and geodesy.

Maximum magnitude

The maximum magnitude potential of a fault can be estimated using scaling relationships that convert from fault length, surface area, or slip to magnitude (discussed more in Chapter 4.4.3).

Fault length is the total length of its surface trace. Importantly, surface rupturing earthquakes do not always rupture the full length of the fault. Fault surface traces can be mapped in the field, or from aerial and satellite data (topography as well as various image types). The applicability of aerial and satellite data to a fault, and the ease of mapping fault surface traces in general, depends on the qualities of the fault site. The easiest faults to map are those that occur in arid, unvegetated landscapes, where fault morphology is preserved and distinguishable even in optical imagery. On the other hand, forested landscapes cover the surface expressions of faults, and in these cases, methods that "see through" vegetation are required to identify the full length of the fault.

Fault surface area is the contact area between the two blocks that comprise a fault. This parameter can be coarsely estimated using basic geometric relations: fault length, dip, and an assumption of the seismogenic zone thickness, the part of the brittle crust that can fail elastically in an earthquake. Dip can be measured from paleoseismic trenches or geologic outcrops that are cut by the fault, especially those that reveal more than a few meters of stratigraphy; or geophysical data such as potential fields methods (gravimetry and magnetometry) and active seismic.

Fault slip is the distance that the fault blocks move relative to each other during an earthquake. In magnitude scaling relationships, slip from a single earthquake event is used with a length-to-slip ratio assumption, and so is most useful as an indicator of magnitude potential for assessing pre-historic fault behavior. One way to measure slip is to identify landform features (e.g., alluvial fans, stream channels, ridges) that span both blocks of a fault, and quantify how they have deformed by earthquakes. For example, a channel crossing a strike-slip fault may be offset during an earthquake such that the upstream part of the channel has moved laterally compared to the downstream channel. Importantly, the slip preserved in a landform does not necessarily indicate the slip produced by a single event.

Slip measurements are most straightforward when the rupture reaches the surface. Slip on ruptures that are preserved in geologic outcrops can be measured from the separation of stratigraphic layers that have been faulted.

One challenge in collecting fault slip data is identifying the number of earthquakes that have occurred since the measured feature was formed. This is addressed with *earthquake timing*.

Earthquake timing

Knowledge of earthquake timing is important for understanding the rate at which faults produce their maximum magnitude events. The most effective way to determine the earthquake ages is to acquire ages of the stratigraphic layers that were deposited before and after faulting events. Usually, this is done by digging a paleoseismic trench, an elongate hole excavated across a fault, and studying the rock and sediment layers that the trench exposes to determine which layers were present when each preserved earthquake occurred. Then, samples of material that can be dated are collected from the layers most closely bracketing earthquakes, (e.g., organic material or minerals using radiometric, cosmogenic, or luminescence dating). These paleoseismic approaches can also be applied to study faulting that does not occur in the shallow crust (e.g., interface) by instead dating formations that preserve the *effects* of earthquakes, such as coastal uplift (e.g. terraces or coral reefs), and turbidites.

When landforms that preserve and bracket earthquake timing are not available, more approximative methods are used. For example, the age of a scarp – a step in topography created by an earthquake – preserved in sediment can be approximated by the comparing the shape of the scarp to local rate of erosion, considering both sediment composition and climate (e.g. rainfall).

One goal of collecting earthquake timing and magnitude (or slip) data is to calculate slip rates – the slip per unit time – for active faults. Slip rates are indicators of how fast the two blocks of a fault are moving relative to each other, and indicate the potential seismic moment release of that fault, that is, the amount of energy stored in a seismogenic fault that may be released by earthquake. In lieu of – or to complement – paleoseismic and geologic slip rate data, GPS stations placed on opposite sides of faults record the velocity of relative block movements; this is discussed more with *Geodesy*.

Fault sense of motion

Fault sense of motion – the way that the fault blocks are moving relative to each other – is important to constructing seismic sources for hazard modelling because the type of fault slip affects seismic wave amplitude and radiation pattern. Seismic sources take into account the fault sense of motion with the rake angle. One way to identify the fault sense of motion is by imposing the regional deformation field (resolved through GPS velocities; see *Geodesy*) onto the geometry of the fault. For example, if the region undergoes shortening along vectors perpendicular to the fault, reverse faulting is expected.

In the absence of deformation data, the fault sense of motion has often been etched into the topography. Each faulting type (normal, strike-slip, and reverse) typically preserves geomorphic features that are uniquely distinct. The kinematics can be assessed by interpretting which sense of motion would lead to the current state of a formerly continuous features, e.g., lateral versus vertical dislocation.

In many cases, the data on characteristics of a given fault are insufficient, however, some parameters can be reasonably estimated. This is discussed more in Chapter 4.3.2 on building fault sources.

3.2.2 Recommended reading

GEM Hazard Team blog post: mapping active faults for fault databases and seismic hazard analysis

3.3 Geodesy

3.3.1 GPS

The type of geodetic data with the highest immediate utility to seismic hazard analysis is GPS velocities. From these data, we can infer how fast the Earth's crust is deforming, and therefore indicate the distribution of earthquake potential across the globe, or a region.

3.4 Open datasets

In this training manual, we focus on the utility of datasets that are openly available for use in seismic hazard modeling. The datasets we present here are composite products that resulted from years of efforts by numerous scientists and projects using the approaches described above, and are now available as compilations that are readily useful.

3.4.1 ISC-GEM catalogue

There are numerous openly available earthquake catalogues available from seismological agencies around the world. The International Seismological Center (ISC) compiles data from 130 of these agencies, reviewing the events from each and compiling them into a single bulletin of IASPEI Seismic Format (ISF), the official bulletin format adopted by the International Association of Seismology and Physics of the Earth's Interior (IASPEI). The ISF format organizes the available data for events into "blocks" that include the parametric data for the event. Read more about the ISF format here:
http://www.isc.ac.uk/standards/isf/

An event catalogue in .csv format can also be downloaded from the ISC, which includes only a single line per event using the prime (preferred) hypocenter. A particular example is the ISC-GEM catalogue (Weatherill et al., 2016).

The ISC-GEM catalogue is a global instrumental catalogue that was originally compiled in a partnership between the ISC and GEM, with Version 1 completed in January 2013; the ISC has since made regular updates and released more recent versions. This catalogue was originally produced as one of GEM's Global Hazard Components, with seismic hazard modelling as one of its primary anticipated utilities. Thus, the catalogue is available in a pre-processed format that expedites its immediate utility for seismic source modelling. In particular, the catalogue reports a single instance of each event with homogenous magnitudes. On the contrary, the catalogue does not include small magnitude events, so has limitations where seismicity rates are low.

The current version (6.0, as of January 2020) of the ISC-GEM catalogue dates back to 1904 and includes events M_w>5.5, with additional M_w>5.0 earthquakes on the continents.

Read more about the ISC-GEM catalogue and download it here:
http://www.isc.ac.uk/iscgem/.

3.4.2 Global Historical Earthquake Archive

The Global Historical Earthquake Archive (GHEA) was also compiled as a component for modelling seismic hazard globally, and was coordinated by the Instituto Nazionale di Geofisica e Vulcanologia (INGV), the British Geological Survey (BGS), and GEM. The archive includes the available literature and other information on earthquakes M_W >7.0 that occurred during the time range 1000-1903. The data are available via a web GUI, and through downloadable ASCII and KMZ files that include a single entry per earthquake, using the parameters deemed most accurate by the committee. Read more about the archive here:
https://storage.globalquakemodel.org/what/seismic-hazard/historical-catalogue/
and download it here: https://emidius.eu/GEH/

3.4.3 Global CMT focal mechanisms

The Global Centroid-Moment-Tensor (CMT) project computes centroid moment tensors information for M_W>5 earthquakes and compiles the event information into a catalogue. The catalogue stays up to date to within a few months, with M_W>5.5 earthquakes computed more rapidly. Global CMT uses .ndk format, an ASCII format in which each earthquake uses five lines that indicate the event timing and location, the data and assumptions used to resolve the CMT, and the resolved moment tensor and nodal planes. Read more about the .ndk format here:
https://www.ldeo.columbia.edu/∼ gcmt/projects/CMT/catalog/allorder.ndk_explained

Read more about the Global CMT catalogue and download it here:
https://www.globalcmt.org.

3.4.4 Global Active Faults Database

Typically, national and regional geological agencies produce fault databases for their given domains. Across these databases, the formatting, and included information, and completeness is inconsistent, given the range of availability of resources and the duration of active fault work for each respective area.

Since 2017, the GEM Global Active Faults Database (GAF-DB) has aimed to merge these databases into a single open-source database which includes the basic level of information for each fault that is needed to produce seismic sources to be used in hazard analyses. At minimum, this includes:

- fault surface trace coordinates,
- dip direction and sense of motion, and
- slip rate

As of January 2020, GAF-DB has near-global coverage, and can be downloaded in GeoJSON, GeoPackage, ESRI Shapefile, and KML formats; the OpenQuake tools mostly make use of the the GeoJSON format. The database is kept up-to-date using Git versioning, and so users can easily update as the database evolves to include more faults, and can more easily contribute to the database.

Read more about GAF-DB and download the data here:
https://github.com/GEMScienceTools/gem-global-active-faults.

3.4.5 Open geophysical models

In addition to datasets, a number of openly available geophysical models are useful for seismic hazard analysis. In particular, these include models of Earth structure and deformation properties.

Slab 1.0 and 2.0

The Slab 1.0 and 2.0 models (Hayes et al., 2012; Hayes et al., 2018) were developed by the USGS in order to provide subduction zone slab geometries for the purpose of seismic hazard modeling, subduction earthquake finite fault inversion and scenario analyses, and in general as an input to subduction zone research (segmentation, coupling, etc.). The models are based on earthquake hypocenters, seismic tomography, receiver functions, and active source seismology, providing subduction zone geometries in terms of depth to slab, slab thickness, strike, and dip for all known slabs globally. The geometries are available in a number of text and grid file formats, as well as a global KMZ. Read more about the Slab models here:
https://www.sciencebase.gov/catalog/folder/5aa1b00ee4b0b1c392e86467.

Geodetic strain rate model

The geodetic strain rate model was developed as one of the GEM global components for utility in seismic hazard analysis. The model used thousands of GPS velocities to calculate deformation rates of the Earth's crust, discretized into cells that are mostly within wide bands surrounding plate boundaries. These deformation rates are a complement dataset to instrumental seismicity catalogues and active fault data for calculating expected seismicity rates.

The effort was led by the University of Nevada Reno, University of California Los Angeles, Chinese Earthquake Administration, and UNAVCO. Read more about the model here: https://gsrm2.unavco.org/intro/intro.html.

4 Seismic source model

In this chapter, we will learn how to build a seismic source model using the open-source datasets that were introduced in Chapter 1. In the applications here, we will create time-independent source models, which mean that the source model considers only average rates of earthquake occurrence, and thus produces the same hazard for any given time. We present the workflow in the following generalized steps:

1. Preparing the seismic catalogue
2. Dividing the region of interest into tectonic domains
3. Choosing and defining the geometries and typologies of seismic sources
4. Assigning earthquake activity rates to the seismic sources

The approach we choose to teach in this manual works well with open-source data and tools. However, we emphasize that agencies around the world use a variety of legitimate approaches to seismic source modeling.

4.1 Catalogue processing

In most cases, probabilistic seismic hazard models are founded on seismic catalogues encompassing the region of interest. In this section, we learn the steps to build a single catalogue to be the basis of a seismic hazard model.

For a detailed example of preparing a catalogue using open toolkits (the OQ-CATK), read Weatherill et al. (2016), which presents an extended version of the ISC-GEM catalogue.

4.1.1 Merging catalogues

While global catalogues such as the ISC-GEM are a good starting point for seismic hazard analysis, it is useful to incorporate more localized bulletins that include earthquakes of smaller magnitudes, especially in areas that are relatively stable. Earthquakes from any number of sources can be combined to form a single catalogue used for hazard analysis. However, the catalogues cannot simply be concatenated, because some earthquakes are likely to be found in more than one source. Thus, the duplicate earthquakes must first be identified, and only a single one selected to appear in the merged catalogue.

Duplicate entries are found by comparing the time stamps and locations of all events in all catalogues under consideration. For two events to be considered duplicates, they must match to within defined time and space “windows” . These windows should be chosen according to the precision of the catalogues in question. For example, older entries may only provide time stamps to the day or hour, whereas newer entries are reported to sub-second. The spatial accuracy may depend on the number of and proximity to recording stations of the earthquakes, and the accuracy of the velocity model used by the agency to record to locate the event.

Once the duplicate events are identified, criteria must be applied to choose which event to include in the final catalogue. For example, the criteria might include considerations on the reliability of one network compared to the others, or the primary area covered by a monitoring network. In some cases, the origin may be taken from one catalogue, while the magnitude is taken from another.

4.1.2 Magnitude homogenization

We use earthquake catalogues to understand the rates and distribution of earthquakes in space and time, but also—on a variety of spatial scales—the relative proportion of large to small earthquakes. In order to compare earthquake sizes, all magnitudes must be reported using the same scale. This is completed using magnitude homogenization, which uses regression analysis between pairs of magnitude scales to convert magnitudes into a single scale.

Magnitude homogenization must be applied to when different magnitude scales are used (e.g., M_W, M_S, M_L, m_b), but also when magnitudes are assigned by different agencies; for example, the National Earthquake Information Center (NEIC) and ISC may report different values of M_W for the same earthquake.

4.1.3 Tectonic regionalization

Seismic sources are often defined on the basis of their tectonic region type, that is, on the tectonic environment in which they occur (e.g., active shallow crust, subduction interface, subduction intraslab. . .). For models that include sources from many tectonic region types, the catalogue must be first divided, or classified, into subcatalogues in which each includes only the earthquakes thought to have occurred in a certain tectonic region; these subcatalogues can then be used to characterize the seismic sources.

There are many approaches to classifying seismicity, most using references that are treated as seismogenic boundaries. These can be surfaces that are specifically and carefully defined to delimit a tectonic unit, including Earth structure models, or more arbitrary assignments based on simplified assumptions about where seismicity takes place.

In a simplified case, polygons and depth limits can be used. For example, all the earthquakes within a given polygon can be considered “subduction earthquakes” , where those with depths shallower than 50 km are classified as interface, and those deeper than 50 km are classified as intraslab. More rigorous methodologies additionally use the proximity of each earthquake to a 3D surface, such as Slab2.0 or Crust1.0.

In complex tectonic environments, “hierarchies” must be used in addition to spatial constraints; this is because some earthquakes will classify into multiple categories, and so the more likely tectonic region type must be chosen.

Additional reading: Garcia et al. (2012), Pagani et al. (2020), and Zhao et al. (2015)

Open-source classification tools: USGS STREC, OpenQuake Model Building Toolkit

Basic levels of tectonic regionalisation

Shallow Crustal Seismicity

Seismicity occurring in the brittle crust can be segregated using depth filters that take into account the properties of both the lithosphere and the catalogues to be classified. Younger and more buoyant active crust is likely to have a thinner seismogenic zone, while older and denser stable crust may have a thicker seismogenic zone. Regarding catalogues, some may include a large number of events that have been assigned a fixed depth; this is especially true for global catalogues. For example, in the ISC-GEM catalogue, depths of 30, 33, and 35 km are very common, including in regions of thinner oceanic crust.

The correct depth filters to be used should be carefully selected by the modeler, however, existing global models of the Earth’s lithosphere can help in the decision making. Often, a single depth value cannot be applied to the whole region; instead, a surface such as that from an existing model (Crust1.0, Lithos1.0), or polygons with independently assigned depths, can be used. Similarly, not all crustal seismicity belongs in a single classification category; it can be divided into – at least – active and stable crust, and further subdivided into “zones”, discussed more in Section 4.2.

Subduction sources

Subduction zones produce the largest earthquakes globally, and account for the vast majority of seismic moment release. Therefore, it is very important to correctly classify the seismicity that is produced by these sources, and significant scientific effort has focused on understanding subduction zone structures and productivity.

At a basic level, subduction sources are divisible into two tectonic region types: interface and intraslab. Interface earthquakes occur on the contact surface between the two tectonic plates that are moving toward each other; the earthquakes accommodate the plate convergence. Intraslab earthquakes occur within the plate that is being subducted, e.g., within the slab. The majority of intraslab earthquakes occur at greater depths than the subduction interface, which permits the simplest level of classifying subduction earthquakes: using a depth cutoff that considers shallower earthquakes to be interface, and deeper earthquakes to be intraslab.

There are many resources available to modellers that help to choose a cutoff depth between interface and intraslab events. The works of Heuret et al. (2011), Christopherson et al. (2015), and Hayes et al. (2012, 2018) include compilations of subduction zone parameters, including the depth ranges of locked interfaces for individual subduction zones. Among other data, these models assign depths by inspecting the focal mechanisms, assuming that most reverse faulting occurs on the interface, while most normal faulting occurs within the slab.

Subduction earthquake classification by depth is easy to implement, but causes many crustal earthquakes and those intraslab earthquakes that occur directly beneath the interface close to the trench to be classified as subduction interface. A more advanced approach is to define a volume within which most interface earthquakes occur, establishing a “buffer zone” around a surface that represents the interface. In the past, this was often done using a planar surface with a dip defined by the observed focal mechanisms, but recent work has produced high resolution models of subduction interfaces that consider many types of data. This approach is used in the methodologies of Garcia et al. (2012), Zhao et al. (2015), and Pagani et al. (2020).

In some cases, a modeler may also want to also divide a subduction zone along strike, or in other words, to impose segmentation. In this case, each interface or intraslab segment would be treated as a separate source, and so the seismicity should also be separated into sub-classes that correspond to the segments.

4.1.4 Declustering

Declustering is the procedure by which the mainshock events in an earthquake catalogue are separated from aftershock events. PSHA usually uses declustered earthquake catalogues in order to evaluate only the seismic hazard that is due to independent earthquakes, as opposed to those that are “triggered" by other earthquakes (aftershocks). The resulting mainshock catalogue is then treated as Poissonian or "memory-less", and thus events of a given magnitude occur at rates which are homogenous in time.

The most common algorithmic procedure is to compare earthquakes within magnitude-dependent space-time “windows" (e.g. Gardner and Knopoff, 1974), treating the largest earthquake within a window as the mainshock for the encompassed events; this methodology is used in the examples that accompany this training material. Before proceeding to the examples, please read HMTK User Guide Section 2.2.1, which explains three windowing options and an additional declustering algorithm included in the HMTK.

Read more about declustering methodologies in Stiphout et al. (2010).

4.1.5 Completeness analysis

Earthquake catalogues do not include all the seismicity that has occured within the space-time volume covered by the catalogue; instead, they are subject to a completeness magnitude, M_C, above which 100% of earthquakes are thought to be included. This occurs for several reasons, for example, some earthquake do not produce strong enough shaking to be detected by the present seismic network (either because the earthquakes have such low magnitudes, or because the network itself is sparse). Small earthquakes can also get “lost" in the signals of larger earthquakes, or the agencies compiling the catalogues may not be interested in the smallest events.

Knowledge of the catalogue completeness is important in PSHA (and, more generally, in applications of statistical seismology) because the occurrence rates assigned to sources are often constrained by earthquake observations. The occurrence rates of sources that are modelled using a Gutenberg-Richter magnitude-frequency distribution (MFD; introduced in Section 4.4) depend on the relative proportion of large to small earthquakes; if too low of a magnitude is used as M_C, the resolved proportion is distorted.

Global catalogues, such as the ISC and ISC-GEM catalogues, typically have a magnitude completeness that varies in both time and space, since global network coverage is not spatially homogenous and station locations change with time. Therefore, in order for these catalogues to be used in PSHA, the completeness must be analysed for the proximity of the model coverage. Use of the most conservative magnitude threshold would eliminate a large number of valuable observations. Instead, a time-dependent M_C can be used, allowing for more robust occurrence rates of higher magnitudes that are based on longer observations periods.

There are many analytical ways to find M_C for a catalogue, including both deterministic and probabilistic methods, and those that are catalogue- or network-based. Read about these methodologies, and more about M_C in this review by CORSSA (Mignan et al., 2010). The included methods are developed to find a single value of M_C. In cases where a time-dependent completeness magnitude is desired, the analytical methods in Mignan et al. (2010) can be used with windowing.

The HMTK provides two means of defining time-dependent M_C. The first is the Stepp (1971) algorithm, which is described in the HMTK Users Manual. The second is to manually assign a completeness table by observing time-magnitude density plots, such as the example in Figure 4.1.

4.2 Source typologies: "state-of-practice" and utility

During seismic source characterization, a set of sources is constructed to represent the anticipated seismicity that pose hazard to the region. The current state-of-practice in hazard analysis includes a few source typologies, that is, geometric representations of modelled seismicity, which vary depending on the level of knowledge regarding these sources. Usually, a model contains many source types. Here, we introduce the most common, discussing them in two groups: faults and distributed seismicity (e.g. Figure 4.2. In this section, we focus on the characteristics of the source types, but do not cover the specific attributes that must be assigned to these sources until Sections 4.3 and 4.4.

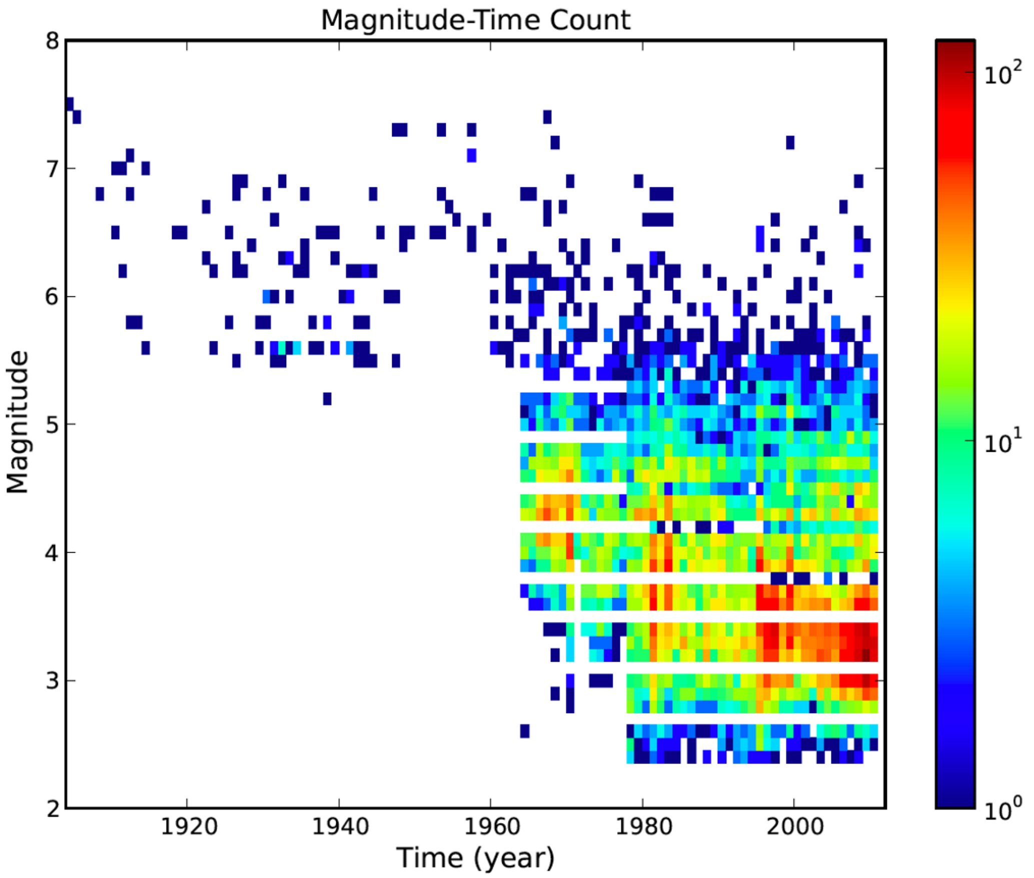


Figure 4.1 – Example of a time-magnitude density plot.

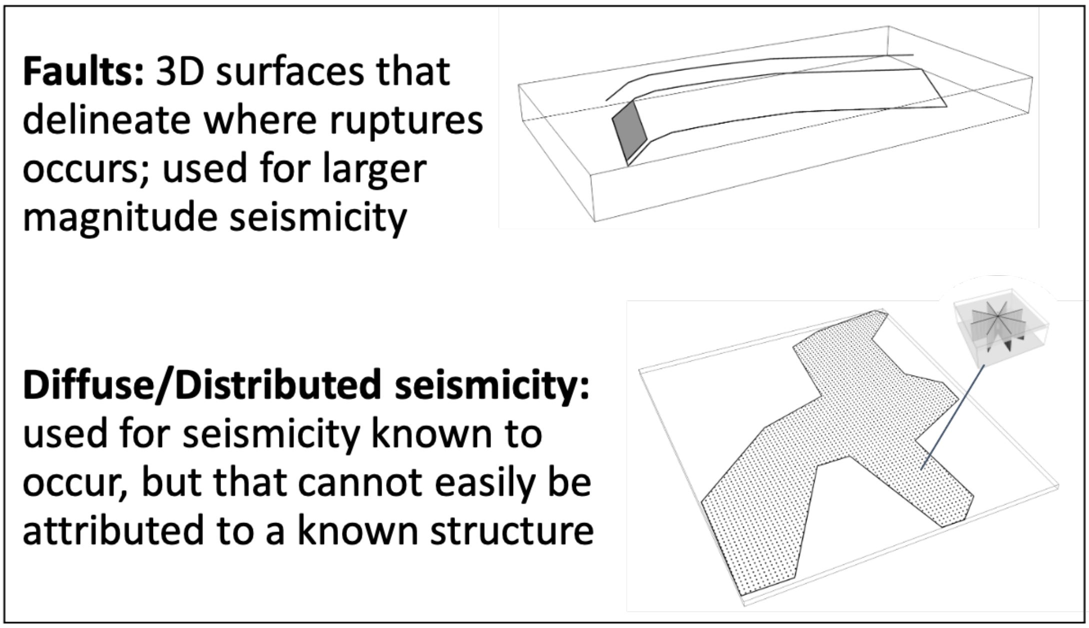
The OpenQuake Engine and Hazard Science Manuals describe how each of the various source typologies are implemented in and treated by the OpenQuake Engine.

4.2.1 Faults

Geometrically, fault sources (sometimes called line sources) are modelled as finite planes or surfaces. Typically, these source types are comprised of discrete faults with a given sense of motion, and during PSHA calculation, ruptures are generated that correspond to individual earthquakes that could occur on the fault surface. Each fault source is prescribed its own magnitude-frequency distribution, which is represented in the OpenQuake Engine as magnitude and occurrence rate pairs.

Within PSHA, there are sub-typologies of fault sources which differ based on the complexity of the source geometry and restrictions to the rupture geometries or magnitudes that the fault sources can produce. Faults that have been studied more thoroughly can sometimes be modelled with more detailed geometry. For example, it is common practice to use faults with 3D geometries, termed “complex faults" in the OpenQuake Engine terminology, to model subduction interfaces. On the other hand, faults for which only the surface trace

Figure 4.2 – The two main categories of source typologies used in PSHA. Top: Fault sources are 3D surfaces that delineate where ruptures can occur, and correspond to sources for which the geometry is well-defined. The dark gray rectangle indicates a single rupture, while the white surface indicates the fault. Bottom: Distributed seismicity is used where earthquakes are known to occur, but cannot be easily attributed to known structures. Instead, hypothetical ruptures are modelled at many positions and with many geometries and orientations. The figure shows an area source, for which a grid of points inside a perimeter is used to define rupture positions, and other source attributes (described in the text) are used to create ruptures around the grid points.



is known are usually represented by a simpler geometry (“simple faults" in OpenQuake terminology).

All earthquakes occur on faults, but not all faults are known to the scientific community. Fault source typologies are reserved for faults that we *do* know about – cases where we feel confident about the position and orientation of ruptures. However, knowing the position and shape of a fault is not enough to model it using a fault source, we must also have enough information to constrain the occurrence rate of earthquakes on the fault. The methodologies for constraining occurrence rates are covered in Section 4.4.

4.2.2 Distributed seismicity

In some cases, seismicity is not easily attributable to distinct faults. In these cases, seismicity is modeled as distributed; that is, a collection of possible sources with less definitive geometry is defined. Two common distributed seismicity source typologies are area sources and smoothed gridded seismicity.

Area sources are polygons within which there is a uniform probability of occurrence of an earthquake. During hazard calculation in the OpenQuake Engine, area sources are discretized into point sources, each with an equal proportion of the seismic moment rate and identical source attributes. Thus, area sources are appropriate for representing seismicity within a boundary that is thought to have consistent (or similar) characteristics, including occurrence rate. For example, an area source could be used to model seismicity in the outer rise of a subduction zone, or a stable crustal area with no known faults but occasional earthquakes that do not conform to a spatial pattern.

In some cases, area sources are difficult to use effectively. For example, where earthquake rates are spatially variable, use of many area sources would lead to low numbers of earthquakes inside each, which will cause difficulty in constraining occurrences. An alternative approach used for modelling distributed seismicity is called smoothed seismicity, which removes the subjective component associated with the need to define the geometry of an area source polygon. Smoothed seismicity is similar to area sources in that, at first, regions with internally consistent attributes (here called "source zones") are defined and treated as a single source in order to determine occurrence parameters; the source zones are then discretized into a grid of point sources. However, instead of imposing uniform moment rates on the point sources as is done for area sources, smoothed seismicity sources allow for variable occurrence rates across the grid. A fraction of the source zone rate is assigned to each point source by applying a smoothing kernel to the catalogue occurrences. Often, a Gaussian smoothing filter is used.

In both examples of distributed seismicity, seismicity is ultimately modeled as point sources. Despite their name, many calculation engines (including the OpenQuake Engine) do not treat the sources as points during hazard calculations; instead, they use the point source attributes (e.g., hypocenter depth, nodal plane geometry) as a basis for generating ruptures, and the ruptures are used during calculation. This is important because of the impact of rupture finiteness on distance metrics used by ground motion models (GMMs, covered in more detail in Chapter 6.1). For example, many GMMs measure distance to the *rupture*, and not distance to the *hypocenter*. The dimensions of ruptures produced at each point source are based on magnitude-scaling relationships (discussed more somewhere else), and so point sources can still be used for distributed seismicity with larger magnitudes.

4.3 Defining source geometry and postion

One of the fundamental qualities of a seismic source is its geometry, including its position in space (its coordinates), the shape of the source, and the anticipated rupture orientations and positions. The attributes used to define source geometry differ depending on the type of seismic source. During hazard calculation, source geometry is significant because it contributes the source-to-site distance measure used by ground motion models; ground motion models also sometimes use additional source attributes such as rupture sense of motion.

In this chapter, we introduce these attributes on the basis of source typology, and some of the ways that hazard modelers assign these attributes. In addition to the text in this section, please read the OpenQuake Engine User Instruction Manual, which explains the specific attributes required for OpenQuake source typologies.

4.3.1 Distributed seismicity

Point sources

Although termed “point source” , many PSHA softwares actually use a rupture that is built based on the point source and its attributes. Intuitively point sources must correspond to a hypocenter: a latitude, longitude, and depth. The point is converted into a rupture based on a given nodal plane (the rupture’s strike, dip, and rake); a magnitude-scaling relationship, which converts between the magnitude and area of a rupture; and rupture aspect ratio, which controls the ratio between the down-dip width and along-strike length of the rupture. Typically, seismogenic depth limits are also assigned to ensure that ruptures remain within the brittle crust.

Area sources

Because an area source is modelled as a collection of uniform point sources, most of the geometric attributes of the point source persist for area sources. However, instead of defining single latitude-longitude pairs that position the hypocenter, a polygon of coordinates defines the perimeter of the area source. Recall that smoothed seismicity is also initialized with a source zone perimeter.

In order to constrain the source geometry attributes for distributed seismicity, we return to the considerations initially posed in Chapter 2.1. Is there recorded seismicity? How is it distributed? What is the tectonic setting? And are there known active faults?

It is logical to first constrain the larger scale geometric attributes: the perimeters of area sources or source zones. As a pre-modeling step, the basic datasets described in Chapter 3 should be interpreted together for the region of intended coverage. Importantly, the sources that contribute to hazard extend *beyond* the region of interest (ROI), and on this premise we define the first step in seismic source characterization:

1. Define the lateral extent of the seismic source model/s: the region of interest, plus an additional buffer zone around the perimeter.

The width of the buffer zone depends on the region, and should be based on the following consideration:

- **How far away are the seismic sources that contribute to seismic hazard in the ROI?** Larger magnitude earthquakes are felt over longer distances. Earthquakes in older stable crust are felt over longer distances than those in younger active crust.

These latter points also serve as guidelines for selecting an **integration distance** during hazard calculations.

Once the perimeter of the source model is defined, it can be subdivided into sources.

2. Define the perimeters of area sources/source zones that will model distributed seismicity.

Note that while we call this a perimeter, the sources will actually occupy a volume. A source model can consist of many depth layers. Recall that area sources and source zones used for smoothed gridded seismicity should encompass space with internally consistent seismic properties. These properties include:

- Source of deformation
- Faulting style
- Deformation rate
- (for area sources only) spatial distribution of seismicity

There are a multitude of tools that can be used to delineate source perimeters, and the best tool will depend upon the modeler. However, we suggest using QGIS (http://www.qgis.org), since this is an open-source geographic information system (GIS) program that can be used to display numerous datasets and models together and interactively, and to initiate shapefiles (or other vector dataset) that store the source parameters.

Likewise, there are no firm rules that a modeler can use to define the source perimeters, but rather the guidelines above.Seismic source characterization is typically an iterative process, and the source perimeters defined in the first undertaking of Step 2 are liable to change.

Once the source perimeters are defined, the seismicity and tectonic processes occurring inside each can be analyzed in order to assign the other parameters. The OpenQuake Hazard Modeler's Toolkit (HMTK) includes tools to derive these parameters from source perimeters (shapefile format) and catalogues. Before continuing, please learn about the HMTK catalogue format in Section 2.1.1 of the HMTK User Guide!

Many of the next steps to source characterization are completed using subcatalogues that are "clipped" to the source zone perimeter, as well as a depth range if appropriate, as is done in tectonic regionalization. The examples included in this manual use the HMTK to filter events into sub-catalogues, but again, many tools and strategies are available.

Additional reading: QGIS tutorial for making vector datasets, HMTK User Guide

3. Assign hypocentral depth distribution

The hypocentral depths assigned to a source should correspond to the depths of events in the subcatalogue for the area source or source zone. In most seismic hazard softwares (including the OQ Engine) it is possible to assign a hypocentral distribution (e.g., pair of depths and weights summing to 1.0), rather than a single value, thus capturing the aleatory uncertainty in hypocentral depth. Aleatory uncertainty is used to model the range of possibilities that are expected given an infinite observation period, in contrast to epistemic uncertainty, which reflects our incomplete knowledge.

Visual approaches can be helpful for identifying depth concentrations within a subcatalogue, including histograms that count events within a depth range, and time-depth density plots. In addition to revealing where events are occurring, these plots can also reveal characteristics of the earthquake catalogue. For example, hypocentral depths of moderate and small earthquakes may be assigned a fixed depth value (e.g., 10 km, 33 km), and so these precise values, in addition to their encompassing depth ranges, may have excessively high counts. We caution modelers to make themselves aware of these catalogue characteristics.

Additionally, we note that earthquake depths are usually uncertain on the scale of – at least – kilometers. Thus, it is important not to over-interpret the depths of a subcatalogue by assigning a hypocentral distribution that is too finely discretized. Instead, for example, the distribution can use a handful of depths that each represent a range (e.g., a depth of 7.5 km representing hypocenters that span 5-10 kms depth. There are two major reasons for this. First, using a large number of precise hypocenters (e.g., 1.0, 2.0, 3.0, ... n km) does not necessarily improve the model, even if all depths are known with small errors. One reason for this is that ground motion models (GMMs) compute intensity measures for a mean value with a standard deviation that typically allows for larger variation than would be caused by a small shift in hypocenter depth; actually, some GMMs do not consider hypocenetral depth. A second reason is that each additional hypocentral value increases the computational load of the model. These points should be considered while assigning the hypocentral depth.

In some cases, little or no seismicity is available to constrain hypocentral depth. In these cases, a common solution is to apply "rules of thumb" , or generalizations based on the tectonic region type. One simple approach is to assume that all seismicity occurs at the halfway point of the seismogenic thickness.

4. Assign focal mechanisms/nodal planes

Focal mechanisms — which are comprised of two nodal planes, each with a strike, dip, and rake –are used in two ways during hazard calculation. First, for each source included in a seismic source model, earthquake ruptures are defined to use as the basis for GMMs to calculate the intensity measure levels produced by earthquakes at each site. The rupture is oriented according to the source nodal plane strike and dip. Some GMMs also consider the focal mechanism, either applying a scaling term based on the simplified classification (e.g., normal, reverse, or strike-slip), or by the rupture's rake.

Like hypocentral depths, nodal planes can usually be assigned singularly, or as a weighted distribution. Also like hypocentral depths, a distribution of nodal planes should include up to a few possibilities that represent the faulting styles and geometries in the source zone, aiming to find a balance between what is observed in earthquakes and computational expense.

A focal mechanism distribution can be assigned based on the focal mechanisms of earthquakes that have occurred within the source perimeter, and assigned based on a subcatalogue of the GCMT database, or any other catalogue of focal mechanisms. Often, as in the GCMT project, focal mechanisms correspond to an earthquake centroid, not to the hypocenter; this is important to keep in mind while preparing a subcatalogue of focal mechanisms.

Using only focal mechanism catalogues, assigning nodal plane distributions to source zones is not as straightforward as is assigning hypocentral depth distributions. Raw nodal planes cannot be plotted as histograms, since they include three parameters (strike, dip, and rake); however, they can be converted to simplified classifications (e.g. strike-slip, normal, reverse, and the oblique senses of motion in between), and binned by these classifications. Map-view plots of focal mechanism "beachballs" and strike-vs-dip plots – along with the judgement of the modeler – can be used to look more closely at the nodal plane orientation.

Focal mechanisms include two nodal planes: one on which the earthquake occurred, and an auxiliary plane which would result in the same patterns of energy radiation and first arrival polarities. From the focal mechanism alone, it is impossible to know which of the nodal planes hosted the earthquake. This is important for modelers to consider when using strike-vs-dip plots to assign nodal planes to sources. If both nodal planes are plotted, clusters of data will form at the true nodal planes as well as the auxiliary planes. Sometimes, both nodal planes are likely candidates for seismicity within a source zone, such as for normal faulting mechanisms in basin and range-style extensional tectonic settings, whereas for vertically dipping, purely strike-slip mechanisms, one nodal plane may be more likely.

Alongside focal mechanisms, some other data types are useful for assigning nodal planes to distributed sources. GPS and strain rate data will reveal the expected faulting types, and active faults encompassed by the source zone can reveal the orientation of the seismogenic structures. In lieu of other data, Andersonian mechanics can be used to assign dip values to nodal planes.

The final nodal plane distribution for a source or source zone should include few enough nodal plane-weight pairings that their coexistence within a single source can be justified. If there are too many, then the modeler should consider whether the source should be further divided. Typically, this constraint is considered before the source perimeters are delineated. However, observing seismicity characteristics on a smaller scale often reveals patterns that are harder to see in large-scale datasets, and thus source delineation often includes multiple iterations.

5. Assign rupture-scaling attributes

Most hazard calculation engines use magnitude-scaling relationships and rupture aspect ratios to build a set of ruptures from the hypocenters and nodal planes, which should be based on the judgement of the modeler.

The most important consideration regarding magnitude-scaling relationships(MSRs) is that they must be compatible with the tectonic region type (TRT); while the earliest MSRs did not have enough data to divide the available data into TRTs or by faulting style, the more recent MSRs tend to include multiple relationships that each correspond to a category of earthquakes. MSR selection for PSHA is discussed in more detail by Stirling et al. (2013).

Although the depths of hypocenters are assigned to each source, the rupture scaling attributes may create ruptures that extend unrealistically deep, or that extend beyond the Earth's free surface. To prevent this, many calculation engines also require that the lower and upper seismogenic depths are defined, and adjust the rupture dimensions to comply with the intended seismogenic layer (each engine uses its own approach to do this).

4.3.2 Fault sources

Fault sources share some geometric attributes with distributed seismicity sources, while also simplifying some aspects of source definition. Like distributed seismicity, construction of fault sources begins by delineating the bounds of the source:

1. Define the fault surface.

The way that fault source surface geometry is defined can be summarised as by the following approaches:

- Define the 3D coordinates of the fault surface, covering its full extent
- Define the surface trace of the fault, as well as a fault dip and seismogenic depth limits; together, these attributes can be used to compute the 3D coordinates of the fault surface

Fault databases such as the GEM Global Active Faults Database expedite the latter approach, providing both surface traces and fault dips. The former approach is most commonly used if a 3D model of a fault already exists or when a wealth of geophysical data spans the fault (e.g. subduction interfaces, for which Hayes et al. (2018) have released Slab 2.0, a model with global coverage).

When constructing fault sources of any surface complexity, the modeller faces decisions about the extents of faults that should be modelled as a single source. In other words, modellers must decide whether the limits of a fault strand imposed in a fault database correspond to barriers along a fault or fault system that earthquakes will not cross; the fault segmentation used to construct a fault database may not correspond to the supposed boundaries of earthquake rupture. In a first attempt, a modeller may choose to divide a fault trace into several continuous or semi-continuous segments to model as individual fault sources, or alternatively to link multiple fault traces together to form a single fault source. More advanced methodologies may also allow ruptures that incorporate multiple fault sources (e.g. the UCERF model for California by Field et al. (2014)).

The basis for fault segmentation is complicated and not unanimously agreed upon by seismic source modellers, and—given the low occurrence rates of large earthquakes—most valid segmentation hypotheses are extremely difficult to reject. Additionally, recent large multi-fault earthquakes (e.g., El Mayor-Cucapah, Kaikora) have further challenged hazard modellers. This training does not aim to define a procedure for segmenting or joining faults; however, we indicate some data types that are commonly used to make fault segmentation choices:

- Fault kinematics (slip rate and sense of motion)
- Fault coupling, such as in subduction zones
- Topographic of bathymetric features
- Variability in seismicity rates
- Historical surface ruptures or finite fault models

This final data type can most effectively reject a segmentation choice.

Because fault source geometry is defined by a surface, there is no need to define the strike and dip of rupture nodal planes; these parameters can be computed for each rupture that is used during hazard calculation. Only a rake angle is requried, which is usually derived from the fault sense of motion prescribed in the fault database used to construct fault sources. Likewise, hypocenter depth can be calculated from ruptures. In the case of the OpenQuake Engine, the hypocenter is assumed to be centered within the rupture. Thus, instead of requiring nodal plane and hypocentral depth distributions, most calculation engines require parameters that indicate how ruptures should be constructed using the fault surface, as in the following two steps:

2. Define the rupture mesh spacing.

The **rupture mesh spacing**, as it is named in the OpenQuake Engine, indicates the resolution at which the calculation engine will discretize the source and float ruptures across the fault surface.

3.Define rupture-scaling attributes.

This step is consistent with that for distributed seismicity, and assigns rupture aspect ratios, seismogenic limits, and magnitude-scaling relationships that will determine the shape and size of ruptures.

Please see the OpenQuake Engine User Instruction Manual for more explanation on rupture floating parameters. We emphasize that the methods used by different calculation engines to produce ruptures from fault sources and to assign nodal planes and hypocenters are variable, and the modeller must consider the specific software when defining source geometry and rupture-scaling attributes.

4.4 Defining source occurrence

The sources are also assigned earthquake occurence information, e.g., magnitude-frequency distributions (MFDs). MFDs indicate the average frequency at which a source will generate earthquakes of each considered magnitude, usually using units of number of earthquakes per year, the annual rate of occurrence. MFDs are calibrated using either observed seismicity over instrumental and historic time scales, or by slip or strain rates across the seismic sources. In this chapter, we introduce the types of MFDs that are most commonly used in modern PSHA, the methodologies that are used to calibrate the source MFDs, and some applications to certain source typologies.

4.4.1 Types of MFDs

In principle, MFDs are usually built upon one of the following principles:

- a source produces a range of earthquake magnitudes, with occurrence rates that decrease exponentially with increasing magnitude (a Gutenberg-Richter (GR) MFD; Gutenberg and Richter (1944))
- a source produces a “characteristic earthquake" of a single magnitude (Schwartz and Coppersmith, 1984)
- a hybrid between these approaches, (e.g. Youngs and Coppersmith, 1985)

Gutenberg-Richter MFDs

Gutenberg and Richter (1944) were the first to resolve a relationship between magnitude and occurrence rate for earthquakes in a region of interest:

$$logN = a - bm \tag{4.1}$$

where N is the annual rate of earthquakes with magnitude $M \geq m$, a is the rate of all earthquakes, and b is the relative distribution of earthquakes among magnitudes. Equation 4.1, called the *Gutenberg-Richter occurrence relationship*, has a negative exponential functional form, in which occurrence rates decrease with increasing magnitude. Higher b-values release a larger proportion of the seismic moment in small earthquakes, while higher a-values indicate an overall higher rate of seismicity.

While the Gutenberg-Richter occurrence relationship allows computation of the occurrence rate of earthquakes with any magnitude, some sources may only be capable of producing earthquakes up to some maximum magnitude, M_{max} (discussed in Section 4.4.3). To account for these constraints, a truncated GR MFD is used to cut the MFD at M_{max} magnitude.

For calculation purposes, MFDs are also cut at a minimum magnitude (“double-truncated") that signifies a threshold below which earthquakes are not expected to cause damage.

Figure 4.3 shows an example of a truncated GR MFD where a=4.0 and b=1.0, plotting both an incrememntal (or discrete) and a cumulative MFD, and a truncated instance of the cumulative MFD for which M_{max} has been assigned. The cumulative MFD shows the MFD computed from the Gutenberg-Richter recurrence relationship, which indicates the annual occurrence rate of earthquakes with a magnitude greater than or equal to the x-axis magnitude, while the incremental MFD includes only the occurrences within the respective bin, here using a bin width of 0.1. The discrepancy between these two representations of the MFD is significant in terms of hazard calculation, and the modeller must check which MFD is assumed by the utilized calculation engine. Please read the OpenQuake Engine User's Manual to see how to assign MFDs in the case of the OpenQuake Engine.

Catalogue observations classified to a seismic source or source zone (review Section 4.1.3 about catalogue classification) are the most widely used data for constraining GR MFDs. There are several algorithms used to resolve a GR MFD from observed seismicity; one of the most common is the maximum likelihood method of Weichert (1980). A major advantage of this method is that it is possible to include earthquake magnitudes with unequal observation periods, allowing the modeller to include small earthquakes that were detected more recently, while utilizing a longer duration—and thus computing more robust rates—for larger magnitudes. The Weichert (1980) method, as well as other maximum likelihood methods, is described in Chapter 2.4 of the Hazard Modeller's Toolkit User's Guide.

In PSHA, common practice is to constrain MFDs to model only independently occurring earthquakes (e.g., mainshocks; review Section 4.1.4 about declustering earthquake catalogues). Thus, when using catalogue observations to constrain a GR MFD, the declustered catalogue should be used. We emphasize that the decision to include only mainshocks is a topic of debate in the hazard community, however, we present the standard approach as of this writing.

Characteristic MFDs

The characteristic MFD, first proposed by (Schwartz and Coppersmith, 1984), hypothesizes that seismic moment for individual earthquake sources (i.e. faults) is predominantly released through large earthquakes that occur cyclically, contrasting the GR MFD by concentrating earthquake occurrences around upper magnitudes. The characteristic MFD is predominantly reserved for seismic sources that are geometrically well constrained and thought to produce large earthquakes, e.g., fault sources or subduction interfaces.

The simplest characteristic MFD assigns all occurrences of a source to a single magnitude—the magnitude that corresponds to full-fault rupture—but MFDs of this type more commonly use boxcar or truncated Gaussian functions centered on the characteristic magnitude, e.g. Figure 4.4. Other characteristic MFDs allocate the majority of seismic moment release to the

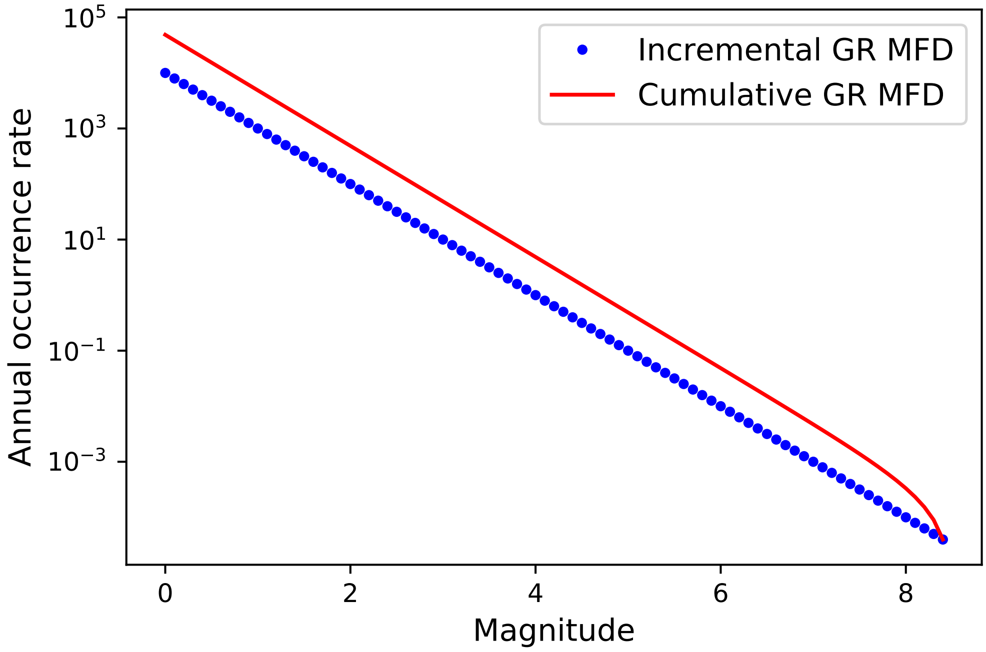


Figure 4.3 – Example of a GR MFD.

characteristic earthquake magnitude/magnitude range, but also include a GR component that allows some smaller magnitude earthquakes, e.g., the Youngs and Coppersmith (1985) MFD (see Figure 4.5). The mathematics of these MFD types are explained in the Hazard Modeler's Toolkit tutorial in Section 4.1.

The characteristic earthquake magnitude used as the basis for constraining characteristic MFDs is typically computed using a magnitude scaling relationship (MSR) and the dimensions (fault surface area, length, or down-dip width), finding the maximum magnitude rupture that the source can theoretically support; review Sections 3.4.4 and 3.2.1 on fault databases and fault investigations. As a “sanity check", the maximum magnitudes computed in this manner must be compared to historical and recorded earthquakes attributed to the source.

Characteristic MFDs are not constrained using seismicity observations, but instead by using a simple moment balance based on the total moment release rate, $\dot{M}_o$, determined by the slip rate and fault surface area following Equation 4.2:

$$\dot{M}_o = c\mu A\dot{s} \tag{4.2}$$

A is the fault surface area, μ is the shear modulus, c the seismogenic coupling coefficient, and $\dot{s}$ is the fault slip rate. The moment balance distributes $\dot{M}_o$ over the range of permitted magnitudes following Equation 4.3.

$$\dot{M}_0 = \int_{M_{min}}^{M_{max}} \lambda(m) M_0(m) dm \tag{4.3}$$

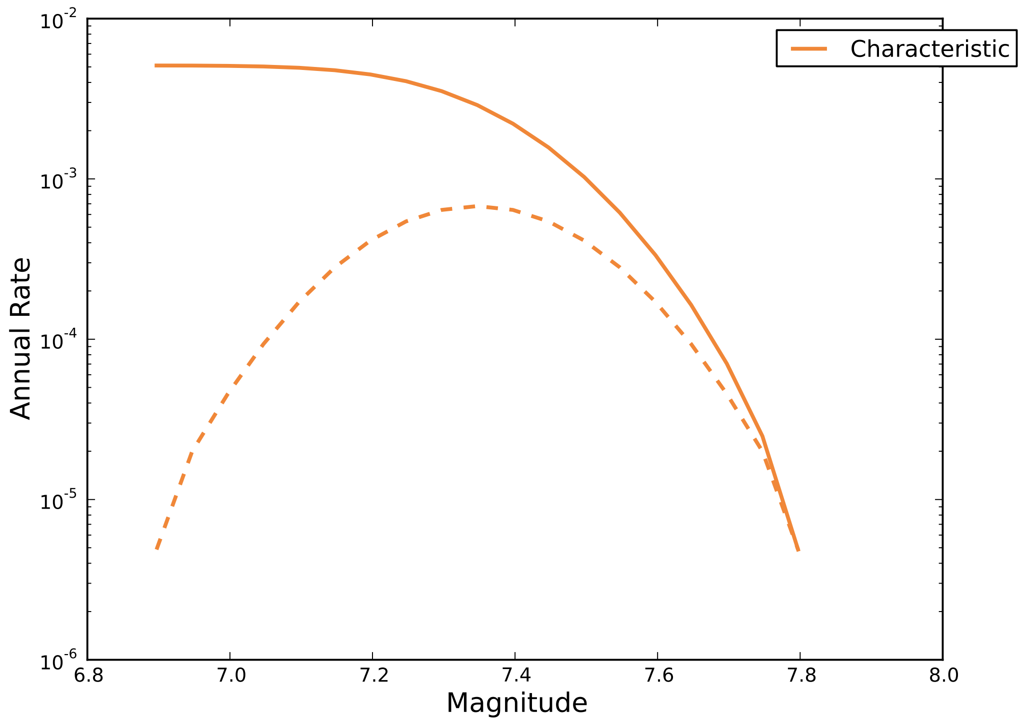


Figure 4.4 – Example of a characteristic MFD using a truncated Gaussian.

4.4.2 Smoothed seismicity

For some source typologies, a single MFD per source sufficiently models the occurrences; this is the case for fault sources and distributed seismicity that is modelled by area sources. Alternatively, distributed seismicity can be modeled using **smoothed gridded seismicity**, which heterogenously distributes the MFD occurrence rates across a source zone using a smoothing kernel and the distribution of observed seismicity. If a GR MFD is used, the b-value remains constant, while the a-value is distributed among of grid of point sources that discretize the source zone. In PSHA, smoothed seismicity is often used to account for the spatial variability in intraslab seismicity or background seismicity in shallow crustal tectonic regions.

One of the most common smoothing algorithms used in modern PSHA is the method of Frankel (1995). This method is described in detail in the HMTK Tutorial Section 2.6.

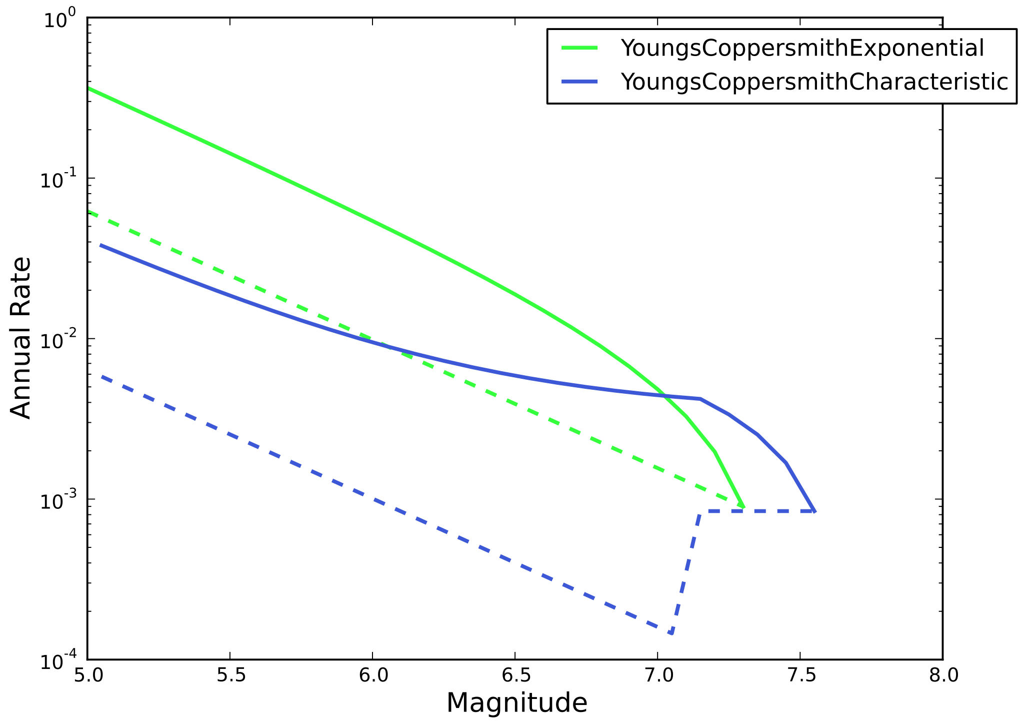


Figure 4.5 – Example of a Youngs and Coppersmith (1985) MFD.

4.4.3 Maximum magnitude

One parameter that must be considered while assigning occurrence rates to a seismic source is the maximum magnitude (M_{max}) earthquake that might occur. Choosing M_{max} is extremely difficult and uncertain, since hypotheses can be rejected by observations, but never confirmed.

There are a few approaches used to determine M_{max} for a seismic source, which consider catalogue observations, history of moment accumulation and release, or the physical dimensions of the source. Here, we introduce four approaches.

The Electric Power Reseach Institute (EPRI) developed a Bayesian approach for determining M_{max} using a prior distribution of M_{max} obtained from a global database of earthquakes that occurred in stable areas, and then the likelihood function using the maximum observed magnitude included in the earthquake catalogue for the investigated area (Johnston et al., 1994).

A method proposed by Kijko (2004) statistically analyses the catalogue for a seismogenic zone or source zone to find a magnitude increment δ to be added to the magnitude of the largest observed earthquake, thus admitting a magnitude that is slightly larger than the largest observed events. In some cases, modellers instead assign an arbitrary δ value (e.g., 0.3 or 0.5).

An algorithm by Makropoulos and Burton (1983) considers the amount of seismic moment that has accumulated in a source zone or on a fault compared to the moment that has been released in earthquakes, thus resolving the seismic moment budget.

Lastly, when fault sources are used to model seismicity, the dimensions of the fault surface can be used to constrain M_{max} using magnitude scaling relationships, the same equations that are used to generate ruptures from sources. The magnitude scaling relationships included in the OpenQuake Engine are summarised here.

The algorithms of Kijko (2004) and Makropoulos and Burton (1983) are explained in Section 2.5 of the HMTK Tutorial.

4.5 Examples

- **1_Catalogue_Basics.ipynb**: introduces common catalogue formats, and the format used by the catalogue module (*cat*) of the OpenQuake MBTK
- **2_Declustering.ipynb**: uses the HMTK to demonstrate catalogue declustering
- **3_Completeness.ipynb**: uses the HMTK to demonstrate completeness analysis
- **4_Magnitude_homogenization.ipynb**: uses the *cat* module of the MBTK to demonstrate magnitude homogenization
- **5_Maximum_magnitude.ipynb**: shows methods from the HMTK that are used to determine the maximum magnitude of a source
- **6_MFD_from_seismicity.ipynb**: demonstrates how to use catalogue seismicity to derive a source magnitude frequency distribution using the HMTK and MBTK
- **7_hypocentral_depth_distribution.ipynb**: uses the HMTK and MBTK to compute the probability mass function of depths for seismicity produced by a source zone
- **8_nodal_plane_distribution.ipynb**: uses the HMTK and MBTK to show the relative proportion of recorded earthquakes with each simplified cateogry of focal mechanisms
- **9_MFD_from_slip_rate.ipynb**: uses the HMTK and MBTK to use fault slip rate and surface area to derive various types of MFDs

Part III

Ground motion characterisation

Ground motion—movement of the Earth's surface—occurs during earthquakes when sudden slip on a fault produces seismic waves. Strong motion is ground motion that may be severe enough to cause structural damage.

The second component of a PSHA input model is the ground motion characterization, which accounts for the ground motion model–the set of ground motion prediction equations—used to compute the level of ground motion that will occur due to the ruptures permitted by the seismic source characterizations. Ground-motion prediction equations (or ground-motion models) are empirically calibrated models that use parameters such as earthquake magnitude and source-to-site distance to compute the ground motion at a given site with the associated uncertainty.

In this chapter, we introduce:

1. types of data and parameters that are commonly used in ground motion characterization, including the open-source databases that compile these data
2. ground motion prediction equations
3. the logic and methodologies used to select ground motion prediction equations to be used in a PSHA model for a region of interest

5 Basic data and parameters

This chapter introduces the main data types that are used in ground motion characterization. These can be categorized in two groups: recordings of the shaking that occurred during past earthquakes, and information on site conditions.

5.1 Ground motion records

5.1.1 Strong motion data

The seismic recordings that are most useful to seismic hazard modelling are seismic waves due to nearby earthquakes that are strong enough to cause damage to structures: strong ground motions. These are recorded on strong-motion sensors (accelerometers) designed to capture strong shaking within frequency and amplitude ranges that are of interest for data applications (e.g. 0 to 100 Hz and 0.001 to >2 g, respectively), such as PSHA.

The most common parameters used in PSHA to describe strong ground motion are peak ground acceleration PGA, peak ground velocity PGV, and spectral acceleration SA. The first two parameters—the "peak" values—indicate the highest acceleration (and velocity) experienced at the site of the recording station during an earthquake. Several different methods are used to compute these peak values (e.g. orientation-independent measures proposed by Boore et al., 2006).SA indicates the peak amplitude of seismic waves with a certain frequency (or period), the acceleration that will be imposed on a particle atop an oscillator with the corresponding period. The spectral acceleration as a function of period is called the response spectra. PGA is a valuable parameter for evaluating or designing shorter buildings, while for taller buildings, SA of a period corresponding to that of the building is more useful.

Strong motion stations usually record data on three components: one oriented to record each of north-south, east-west, and vertical motion. In most cases, PSHA modellers are interested in the horizontal components of ground motion, since these components include the seismic waves that are most hazardous to the built environment (e.g. Figure 5.1).

Strong motion data can be retreived in a few ways. In the optimal case, databases provide PGA, SA, and other parameters from accelerograms that have already been processed. Datasets of this type are usually only available for the largest or most impacting earthquakes. More commonly, the accelerograms corresponding to the same event have been grouped and clipped to the duration of that event, simplifying the data compilation for users. For earthquakes that are less widely studied, the time-series of raw accelerograms must be collected. Strong motion data sources are discussed more in Chapter 5.3.1.

Utility in PSHA

Strong ground motions are the shaking levels that modellers aim to replicate in the ground motion characterization. One way that strong motion data are used in PSHA is to derive a Ground Motion Prediction Equation (GMPE) for an area of interest; this is a significant and time-consuming procedure in itself, and is beyond the scope of this training material. We refer interested readers to a review by Douglas and Edwards (2016). More commonly, strong motion data are used in PSHA as a way to choose a GMPE from the existing database to use for a model. In summary, this is done by comparing the observed ground motions to the intensity levels predicted by each GMPE, and choosing that which performs best. One such workflow is presented in more detail in Chapter 6.2.

5.2 Site information

The ground motions experienced by a site due to an earthquake are impacted by the earthquake source parameters (e.g., the magnitude), the path between the source and the site (e.g., the source-to-site distance), and the seismic site conditions (e.g. geology). Site conditions can change the frequency content and the amplitude of the motion. This chapter introduces some of the parameters used in PSHA to characterize site conditions in order to improve seismic hazard estimates.

5.2.1 vs30

Shear wave velocity is a mechanical parameter linked to the rigidity of the medium. Within the Earth, on average, the shear wave velocity decreases as depth decreases (i.e. the shear

Figure 5.1 – An example of an accelerogram, indicating the peak ground acceleration.

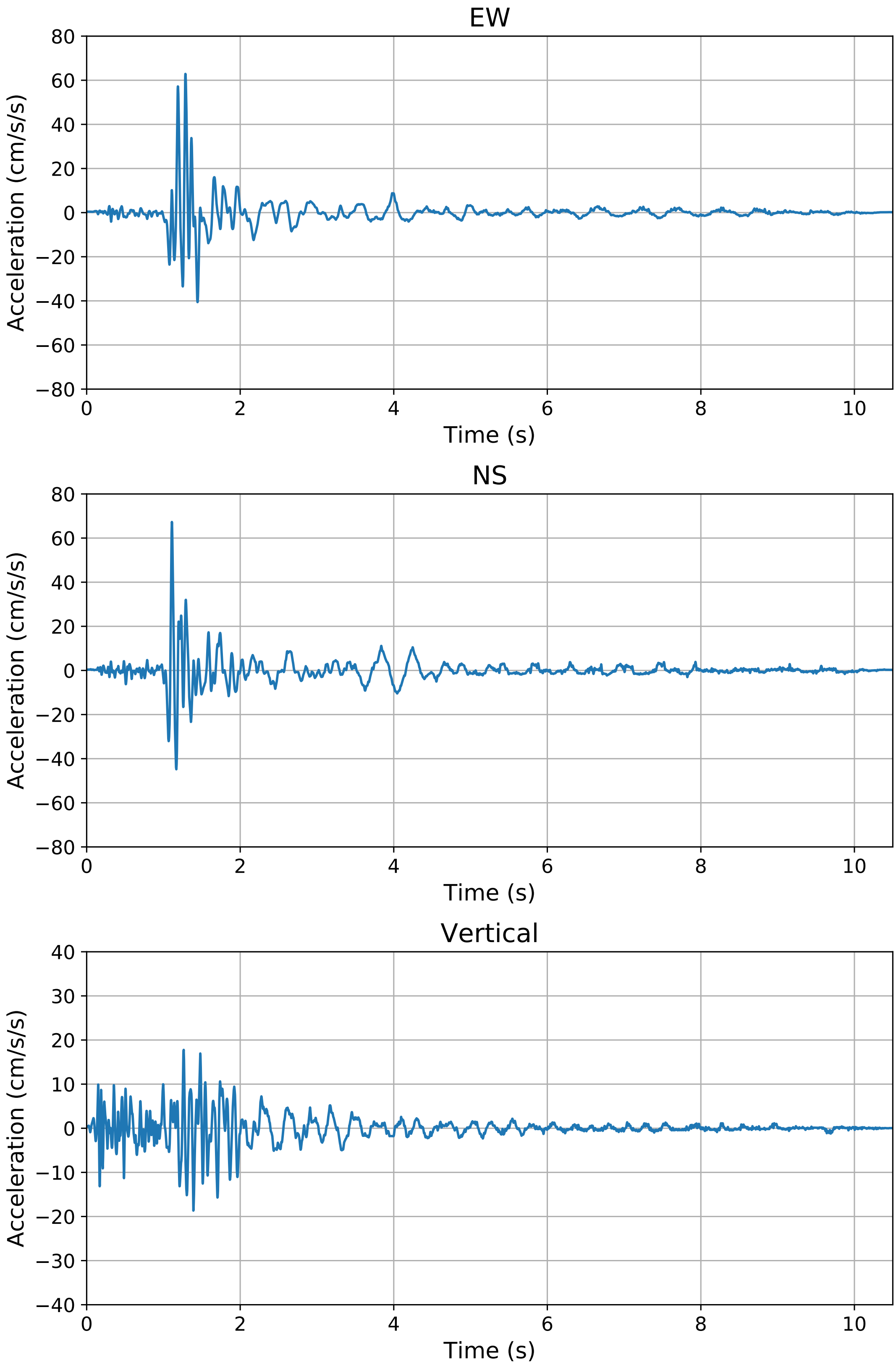
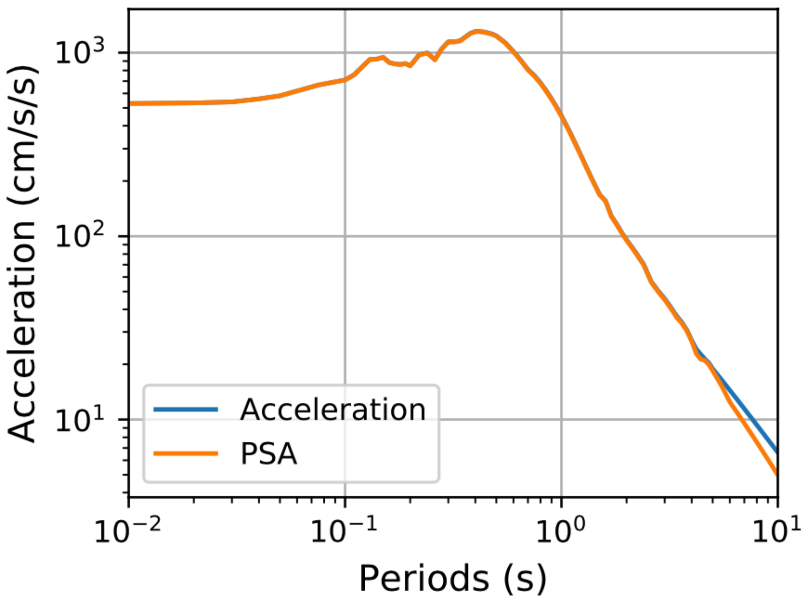


Figure 5.2 – An example of a response spectrum derived from an accelerogram.



wave velocity is lower close to the surface than at great depth). A common parameter used for site characterization in PSHA is $V_{S,30}$ (Borcherdt, 1994), the average shear wave velocity of the materials in the uppermost 30 m of the soil column, inferred by the travel time of a shear wave from the surface to 30 m depth. In general, $V_{S,30}$ increases for more consolidated and harder rock. $V_{S,30}$ first emerged as a means of divising site classes that correlate with shallow rock or soil type, ranging from hard rock to soft soils, as part of the US National Earthquake Hazards Reduction Program (NEHRP) building code (Borcherdt, 1994; Borcherdt and Glassmoyer, 1992), and other building codes have since incorporated similar $V_{S,30}$-based site classifications.

Most GMPEs include $V_{S,30}$ as a variable that accounts for the site class, and as a result, computed hazard corresponds to only that site class. Regional or national hazard maps are usually computed for a single site class either to provide reference conditions for a wide area, or because detailed site characteristics are not available.

Measured $V_{S,30}$ values can be obtained via both invasive and non-invasive shallow geophysical methods, including borehole analysis and surface wave inversions techniques. Where these are not available, proxies based on geology or topography must be used; the topography proxy developed by Allen and Wald (2007) is described in Chapter 5.3.2.

5.2.2 Shear wave isosurfaces

$V_{S,30}$, while a useful proxy for site characterisation and its impacts on ground motion amplification, is most effective for shorter spectral periods. At longer spectral periods ($> \sim 1.0$ s, e.g. Zhu et al. (2019)), the subsurface beyond 30 m may have greater significance. To account for this, some recent GMPEs have added terms that serve as proxies for basin structure, utilizing shear wave isosurfaces that indicate the local depth to a fixed shear wave velocity: the depth to shear wave velocity $V_s = 1.0$ and 2.5 km/s. These values can be obtained using similar methods to those used for $V_{S,30}$: borehole and seismic inversion techniques, and—in their absence—$V_{S,30}$-based conversion equations. $V_s = 1.0$ km/s is usually on the order of tens to hundreds of meters, while $V_s = 2.5$ km/s ranges from hundreds of meters to a few kilometers.

Computing ground motion within the OpenQuake Engine

The OpenQuake Engine includes more than 250 GMPEs. The site parameters required by each GMPE are listed under **REQUIRES_SITES_PARAMETERS** in the OpenQuake GSIM Package Documentation.

Table 5.1 – Agencies collecting and managing strong motion data.

Network	Agencies	Region of oversight	Distribution
Advanced National Seismic Network (ANSS)	USGS	United States of America	IRIS
California Integrated Seismic Netowrk (CISN)	CGS, CSL, BSL, USGS, Cal OES	California,	CESMD
Global Seismographic Network (GSN)	USGS	Global	IRIS
National Strong Motion Project (NSMP)	USGS	United States of America	CESMD
KIK-Net	NIED	Japan	
Engineering Strong-Motion Database	INGV	Europe, Global	

5.3 Open datasets and models

5.3.1 Strong motion databases

Strong motion data collected by seismic networks is typically managed and distributed by the agency operating that network. Usually these agencies oversee national or regional domains, but sparser global networks also exist. Table 5.1 summarizes some of the seismic networks that are currently operating to monitor strong ground motion and maintain openly available data policies. However, we note that this is not a comprehensive list of the seismic networks operating at the time of writing, and that at any given time, a number of temporary networks are also in place.

In the name of coordination among seismic networks, data collection, data distribution, and other goals pertaining to seismic data, a few collaboratives, consortiums, and organizations have emerged, including:

- the Incorporated Research Institutions for Seismology (IRIS)
- the International Federation of Digital Seismograph Networks (FDSN)
- the Center for Engineering and Strong Motion Data (CESMD)
- the Consortium of Organizations for Strong Motion Observation Systems (COSMOS)
- **other to add?**

These groups have worked toward certain data distribution goals, such as to establish systematic station coding and to develop common waveform data formats like the Standard for Exchange of Earthquake Data (SEED) format and the COSMOS format.

In addition to time series data, some of the listed organizations provide processed and

Figure 5.3 – An example strong motion data summary from CESMD (truncated) for the 30 July 2020 Pacoima earthquake in southern California.

Peak Ground Motion Data (Distance Order) for
Pacoima Earthquake of 30 Jul 2020, 06:48:19 PDT, 3.8MW
34.2937N 118.4355W, 7.2 km depth Event Id: 39322767CI
CESMD Engineering Strong Motion Data Center
Table Last Updated: 2020-07-30 07:47:53

Network Id	Name	Statn Nmbr	Station Name	N.Lat	W.Long	Distance Epic	Fault	PGAv1 (g)	PGAv2 (g)	PGV (cm/s)
CE	CGS	24514	Sylmar - 6-story County Hospital	34.326	118.446	3.7	(--)	.060	.060	1.68
CE	CGS	24800	Mission Hills - Sepulveda & San Fernando	34.273	118.467	3.7	(--)	.039	.038	1.24
CI	SCSN	RINB	San Fernando, junction of I-405 and I-5	34.282	118.479	3.9	(--)	.052	.043	.95
CI	SCSN	RIN	Rinaldi	34.282	118.479	3.9	(--)	.192	.191	2.96
CI	SCSN	LFP	Sylmar, I-5 at Los Angeles Resevoir	34.305	118.488	4.7	(--)	.024	.023	.82
CE	CGS	24088	Pacoima - Kagel Canyon Fire Sta	34.296	118.376	5.5	(--)	.009	.010	.33
CE	CGS	24407	Sylmar - Pacoima Dam Upper Left Abutment	34.334	118.397	5.7	(--)	.041	.041	1.28
CE	CGS	24765	Panorama City - Roscoe & Woodman	34.220	118.430	8.2	(--)	.034	.034	1.11
CE	CGS	24386	Van Nuys - 7-story Hotel	34.220	118.471	8.8	(--)	.018	.018	.82
CI	SCSN	DEC	Green Verdugo Microwave Site	34.254	118.334	10.4	(--)	.004	.004	.23
CE	CGS	24051	Northridge - Lassen & Reseda	34.250	118.537	10.5	(--)	.031	.031	1.16
CE	CGS	24815	North Hollywood - Laurel Cyn & Sherman	34.199	118.397	11.0	(--)	.009	.009	.47
CE	CGS	24805	Northridge - Parthenia & Lindley	34.229	118.529	11.2	(--)	.016	.016	.66
CI	SCSN	GCC	Gold Creek Center	34.328	118.319	11.4	(--)	.015	.014	.30
CI	SCSN	Q0057	Densmore Ave, Lake Balboa	34.188	118.477	12.2	(--)	.043	.044	1.24
NP	NSMP	5492	CA:Burbank;Burbank Arpt FS; Ground	34.200	118.360	12.6	(--)	.015	.016	.47
CI	SCSN	QUG	Valencia, Sierra Hwy	34.396	118.498	12.7	(--)	.003	.003	.14
CE	CGS	24198	Chatsworth - 2-story Commercial Bldg	34.240	118.565	13.0	(--)	.030	.030	.87
CI	SCSN	NOT	Northridge, Parthenia near Reseda	34.229	118.558	13.1	(--)	.017	.017	.63

summarised strong motion data for select earthquakes, e.g. CESMD provides tables of peak ground motion at each station (example in Figure 5.3).

We encourage readers to explore the data and data products provided by each of the strong motion data agencies and organizations.

5.3.2 Site characterisation data

While some openly available site characterisation data exists, only small geographic regions are covered, and in general there is not a global resource for these data. Instead, several proxies have been developed for estimating these parameters based on more readily available parameters, e.g., parameters availble from remote sensing.

One of the most common proxies for estimating $V_{S,30}$ is using the Allen and Wald (2007) method that is based on topography, or - more specifically - slope. This method works on the assumption that flatter topography more likely has a surface cover composed of unconsolidated sediment than of rock. Thus, steeper slopes correspond to harder rock and lesser amplification of strong ground motions due to resonance effects.

The most common algorithms for estimating the depths of isosurfaces at which shear waves travel 1.0 and 2.5 km/s are based on $V_{S,30}$ (measured or computed using Allen and Wald (2007)), while some proxies for the 2.5 km/s isosurface are based on the 1.0 km/s isosurface. Those implemented in the GMPE-SMTK are:

- Depth to 1.0 km/s from Vs30: (Abrahamson and Silva, 2008; Campbell and Bozorgnia, 2007; Chiou and Youngs, 2008; Chiou and Youngs, 2014)
- Depth to 2.5 km/s from Vs30 or z1pt0: (Campbell and Bozorgnia, 2007; Campbell and Bozorgnia, 2014)

6 Ground motion characterisation

6.1 What are ground motion models?

Ground Motion Prediction Equation (GMPE), also called GMMs, are attenuation relationships used to predict the expected ground motion at a site for a given rupture. Given information about the earthquake source, the site, and propagation path, a GMPE calculates the intensity measure level (the amplitude) of shaking for an intensity meausre type—a spectral period. At minimum, a GMPE uses the following parameters about the rupture and site.

1. Magnitude of the rupture, usually as moment magnitude. Ground shaking increases with increasing magnitude.
2. Distance between the rupture and the site. Ground shaking decreases with increasing distance.
3. $V_{S,30}$ of the site. Ground shaking decreases with increasing $V_{S,30}$.

GMPEs may also consider other information such as the rupture depth and the shallow geology at the site.

In the context of PSHA model inputs, the Ground Motion Characterization (GMC) is the set of GMPEs used to compute ground-shaking to be expected upon the incidence of sources in the seismic source model. In the simplest case, the GMC may consist of a single GMPE, while more complex PSHA models that encompass multiple tectonic region types will include one or more GMPEs for each.

6.1.1 Characteristics of GMPEs

Hundreds of GMPEs have been developed for PSHA and other utilities. The GMPEs that are typically used in PSHA are physical, empirical models derived by calibrating some functional form using strong-motion data from past earthquakes.

The applicability of each empirical GMPE varies and is defined by the GMPE authors and the qualities of the data used to define the GMPEs. The ground motions produced by ruptures of the same magnitude vary due to the mechanical, thermal, compositional, and other properties of the two fault blocks, as well as the source-to-site path and the local site conditions. While these properties are difficult to include quantitatively in ground motion modelling, independent GMPEs are developed for the different tectonic region types

Table 6.1 – Rupture-site distances supported by the OpenQuake-engine (oq-engine).

Distance definition	Symbol	Description
Epicentral	R_{epi}	Distance between the epicenter and the site.
Hypocentral	R_{hypo}	Distance between the hypocenter and the site.
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
Horizontal off-end distance	R_{y0}	Horizontal distance off the end of the rupture measured parallel to strike
Top-of-Rupture depth	Z_{tor}	Depth to the top edge of the rupture

(e.g. subduction interface, active shallow crustal, volcanic, etc.) as a means of categorizing sources with common properties. Some GMPEs go beyond this, adding geographical constraints to their applicability. These categorical factors are imposed by the data used to develop the GMPE; e.g. only seismic recordings for subduction interface earthquakes are used to constrain the respective GMPEs.

A comprehensive summary by John Douglas of the epirical GMPEs developed since 1964 is available http://www.gmpe.org.uk/gmpereport2014.html.

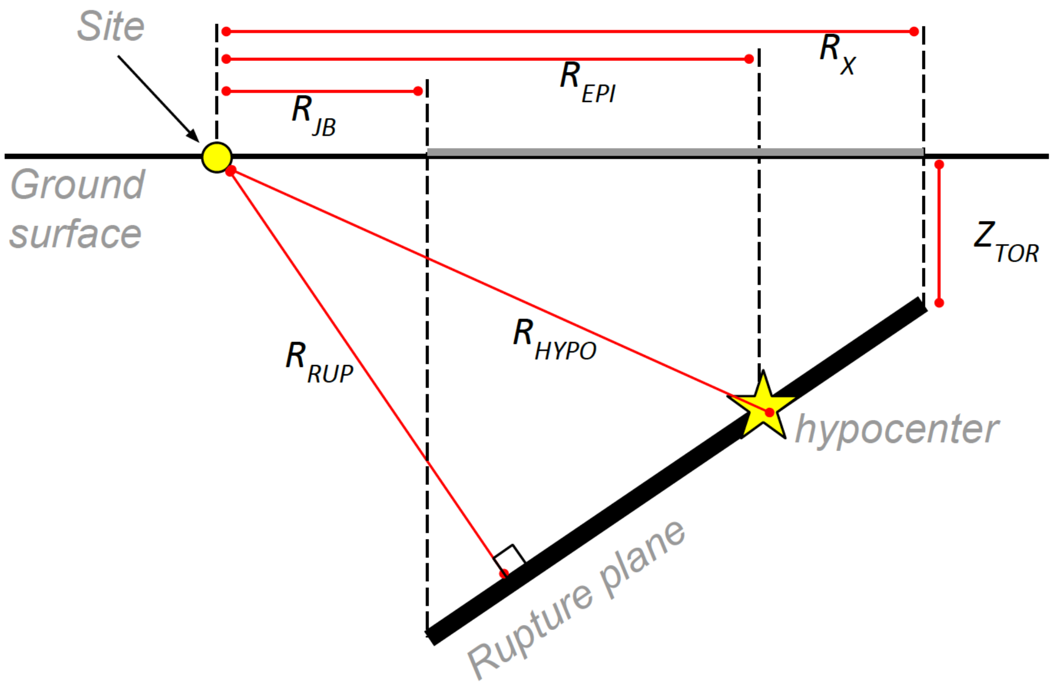
6.1.2 Distance metrics

One of the standard metrics required by any GMPE to compute ground motion is the source-to-site distance. While always required, the source-to-site distance is not always measured in the same way; this is important to understand when defining a ground motion model, since the metric used by a selected GMPE can have a large impact on the computed intensity measure levels for a source-site pair. The most common distance metrics, and those that are included in the OpenQuake Engine, are depicted in Figure 6.1 and summarised in Table 6.1. A GMPE may use more than one distance metric.

6.1.3 Mean and standard deviation

A GMPE yields two values for each intensity measure type and each considered rupture: the mean of the logarithm of ground-motion and its standard deviation. Thus, GMPEs for a given set of descriptive parameters admit a set of intensity measure levels described by a Gaussian-shaped probability distribution centered on the mean predicted intensity measure level logarithm; in some cases, this distribution can be truncated to some number of standard deviations.

Figure 6.1 – Distance metrics that are commonly used in GMPEs; see Table 6.1.



The mean value predicted by a GMPE results from the regression of many strong motion data to a functional form. The data maintain a scattered distribution about the mean result and are the basis of the statistical standard deviation. The standard deviation permits variability in the ground motions for the same rupture parametrization due to factors that cannot reasonably be accounted for by GMPE variables, e.g. rupture velocity, slip distribution, source-to-site path, etc.

The standard deviation of a GMPE can be attributed to two types of variability, intra- and inter-event, which together capture the total standard deviation (see Atik et al. (2010)). Intra-event standard deviation is the variation in intensity measure level that occurs in a single earthquake for equal source-to-site distances. For example, during a single earthquake, the PGA recorded for events 50 km from the earthquake hypocenter may have scatter that depends on the geology between the rupture and the station, the strike of the fault compared to the station azimuth, and the direction of rupture propagation. On the other hand, inter-event standard deviation is the change in intensity measure levels recorded at the same station during different earthquakes. For example, the state of stress in the crust may have changed between two consecutive co-located earthquakes of comparable magnitude, impacting the energy in the seismic waves that propagate away from the rupture.

6.2 Choosing ground motion models for PSHA

At least one GMPE must be selected for each large-scale tectonic region included in the seismic source characterisation of a PSHA input model. Given the large number of available GMPEs, selecting them is not a simple task. To date, there is not a rigorous and uniform methodology used by experts to choose GMPEs, which is in part due to high variability of strong motion data availability globally. However, some "best practice" guidelines and quantitative approaches are widely used. In this chapter, we introduce the most common ways that hazard modellers select GMPEs. We discuss the following approaches as if the aim were to select a single GMPE for each tectonic region, however, PSHA ground motion characterizations usually incorporate epistemic uncertainty by weighting several GMPEs per tectonic region using a logic tree; this is discussed more in Part IV.

6.2.1 Selection and exclusion criteria

One method of selecting GMPEs is to eliminate GMPEs from the possible choices until only one or a few remain. This is done by evaluating the applicability of the each GMPE to the seismic source model in question as well as the robustness of the GMPE derivations, and has led to some authors defining a set of "exclusion criteria". In this section, we introduce the most straightforward exclusion criteria and considerations, and refer the reader to (Bommer et al., 2010; Cotton et al., 2006) for a more comprehensive description.

1. Tectonic environment. First and foremost, the selected GMPE must be derived for a tectonic region type that corresponds to the seismic source to which it is applied. For example, subduction interface sources must use GMPEs derived for subduction interfaces

2. Magnitude and distance. The strong motion data used to derive the selected GMPE should encompass the largest and smallest earthquake magnitudes and source-to-site distances that are included in the seismic source model. Preferably, the selected GMPE was derived using a large data set that includes many earthquakes of each magnitude, each recorded at multiple stations over widespread distances. However, this is difficult and often impossible to achieve given the low number of seismic recordings for large earthquakes and the variable density of seismic stations around the globe. In these cases, one must consider how the functional form of the GMPE extrapolates beyond the magnitudes and distances that are well constrained.

3. Intensity measure types. The GMPE must output the intensity measure types that are required for the PSHA applications. For example, the modeller might want, at minimum, a GMPE that covers spectral periods ranging from 0.0 to 2.0 s.

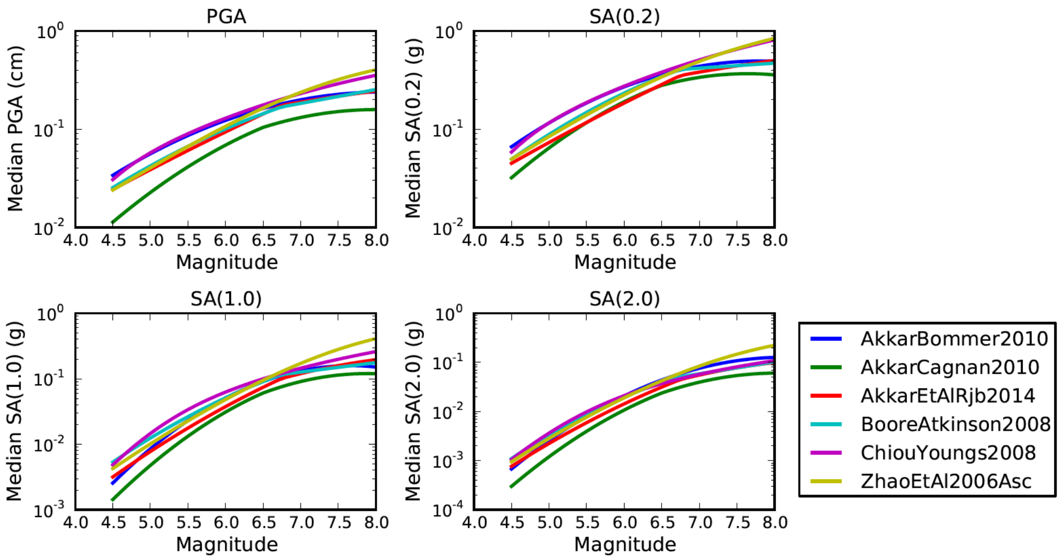
4. Geographic coverage. A few GMPEs have been derived using global data sets, while others only considered data for a confined geographic area (a country or region). The advantage of the former is that much more data is available, while the latter is able to better capture local or regional characteristics that are not included as GMPE variables. GMPEs derived for specific areas should not be used outside of their intended region without careful consideration and a defensible explanation.

The GMPE qualities required to apply these criteria are summarised in a GMPE compendium by John Douglas, which can be found at http://www.gmpe.org.uk.

6.2.2 Comparative scaling

Comparative scaling is a qualitative analysis that demonstrates the relative behavior of GMPEs against each other with respect to the source parameters (magnitude and distance) and spectral periods. A common way to make comparative analyses is via trellis plots, which show how each GMPE in a set scales for one of magnitude, distance, and period while the other two parameters are fixed. For example, a trellis plot may compare how PGA estimates scale with distance for subduction interface GMPEs computing intensity measure levels for a M_w 8.0 earthquake. A single comparative scaling analysis usually results in dozens of trellis plots that represent each of the likely combinations of magnitude, distance, and spectral period. Figure 6.2 shows an example set of trellis plots.

Figure 6.2 – An example of trellis plots showing how ground motion levels predicted by five GMPEs scale with magnitude for four spectral periods for an earthquake ~20 km away from a site on rock.



Trellis plots can be used to compare both the mean ground motions predicted by GMPEs and the associated sigma (standard deviation). Again, there is no firm "recipe" for selecting a GMPE from a set of candidates by using trellis plots, however, there are a number of characteristics of the comparative plots to keep in mind:

1. Are any GMPEs distinct outliers for the majority of trellis plots?
2. Which GMPEs best represent the median and upper and lower bounds for the distances

and magnitudes in the source model?

3. How do the GMPEs behave at the extremes, e.g. the largest distances and magnitudes included in the seismic source model? The sigmas?
4. Does a GMPE have a striking abnormality that may defy physical practicality for some realistic parameter?
5. What is the behaviour of each model within and outside its range of applicability?

Trellis plots were widely used in the GEM-PEER Global GMPEs Project (Stewart et al., 2012), which aimed to narrow down the list of available GMPEs to a more managable selection to be considered for global hazard assessments, and are also used frequently during developement of regional or national hazard models. We recommend additionally reading additional text, e.g. Stewart et al. (2012), for a thorough demonstration of trellis plotting.

6.2.3 GMPE-data comparisons

In the optimal case, GMPEs are compared with strong motion data covering the region of interest, allowing the modeller to select GMPEs that best predict the ground motions that occurred during past earthquakes. These GMPE-data comparisons can validate GMPE selections more effectively than other methods, depending on the completeness of the available data.

The strong motion data used in these comparisons must be carefully selected. At minimum, the recordings must correspond to earthquakes with the correct tectonic environment (e.g. subduction interface strong motion data are used to pick subduction interface GMPEs). In addition, the strong motion data available for the comparison will ideally:

- geographically represent the sites and sources included in the PSHA model
- be made up of mostly strong motion data that were not used to derive the GMPE
- cover a wide parameter space, and preferably one that is wider (e.g. more magnitudes and distances) than the data used to derive the GMPE
- include multiple earthquakes, each with several recordings, such that event biases can be observed.
- incorporate a range of spectral periods and $V_{S,30}$ values.

The most simple GMPE-data comparison is qualitative, imposing strong motion data on trellis plots used for comparative analysis and visually determining which GMPE best predicts the data.

More robust analyses use residual analysis that quantitatively identifies the models that perform the best. These methods compute the difference between the predicted ground motions and the observations and assesses them together, usually using a maximum likelihood approach that determines for which GMPE the data is most probable. Some residual analysis methodologies consider only the total likelihood, while others incorporate inter- and intra-event terms. The GMPE-SMTK includes residual analysis methods by Mak et al. (2017), Scherbaum et al. (2004), and Scherbaum et al. (2009), and a Euclidean distance-based methodology by Kale and Akkar (2013).

6.2.4 Cases where no data is available

Often, little or no new strong motion data is available for use in GMPE-data comparisons, and in some geographic locations the strong motion database is in any case sparse. This condition is challenging, and—again—there is not a straightforward solution. In these cases, modelling of epistemic uncertainty is extremely important, as best practice indicates that multiple GMPEs will be used to estimate the range of possible ground motions. The modeller may take the following considerations:

- is there a more data-rich geographic region that may be able to serve as an analog?
- which GMPEs, when considered together, are *mutually exclusive but collectively exhaustive* (e.g. Bommer and Scherbaum, 2008) in that they cover a range of predicted values while no two GMPEs are redundancies of each other

This latter concept is discussed more in Part IV.

6.3 Examples

- **10_GMPEs_Introduction.ipynb**: uses the OpenQuake Engine and GMPE-SMTK to overview the basic properties of GMPEs
- **11_Trellis_Plots.ipynb**: uses the GMPE-SMTK to demonstrate aspects of comparative scaling

Part IV

Modelling epistemic uncertainty

7 Epistemic uncertainty and logic trees

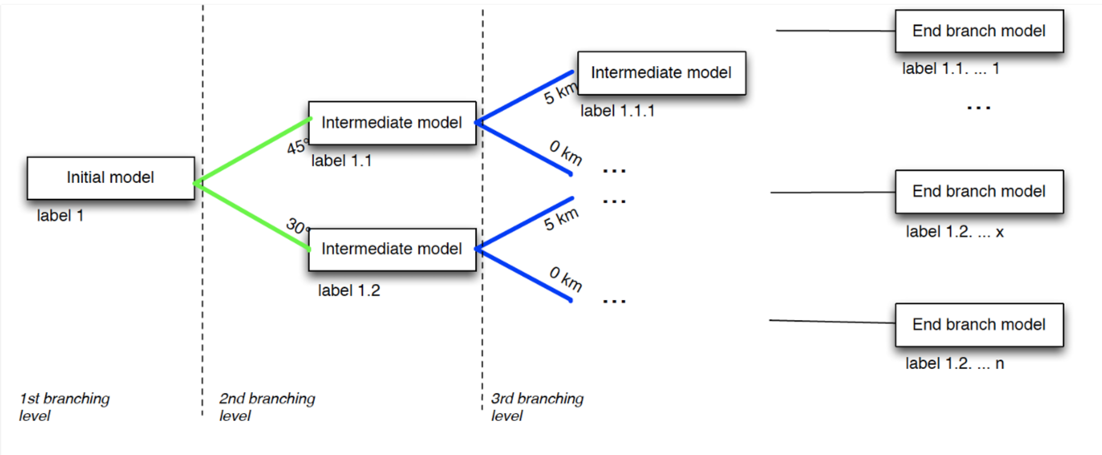
Parts II and III introduced procedures for constructing seismic source models and choosing ground motion models for use in PSHA, but without considering the epistemic uncertainty in modelling decisions. Recall from Chapter 2 that epistemic uncertainty refers to our incomplete scientific knowledge about a phenomenon, rendering many modelling components impossible to confidently and indisputably parametrize.

Epistemic uncertainty pertaining to seismic hazard analysis exists for several reasons. While the last decades have yielded major advances in the scientific understanding of earthquake occurence and strong ground motions, there is still a general scarcity of data relative to the duration of the seismic cycle and global variability in tectonic processes; this challenge was also discussed in Parts II and III. For example, strong-motion stations have only existed for a few decades, and so in many cases they have not have recorded ground shaking during the largest and most impactful earthquakes at a site (i.e. the most significant earthquake occurred during the pre-instrumental period), and thus the expected ground motions are unknown, and must be predicted by ground motion models that extrapolate or use large magnitude earthquakes from other areas. Even in cases where the largest observed earthquake was recorded, it is impossible to assert whether it represents the largest magnitude earthquake that the causative source can produce. This latter point corresponds to another source of epistemic uncertainty: theoretically, an infinite data colletion period is required to reject alternative hypotheses.

While the examples listed so far represent data challenges that lead to epistemic uncertainty, another cause is disagreement among experts, either due to differing assumptions, interpretations, methodologies, or theoretical bases regarding seismic processes. These factors can result in input models that yield a range of projected hazard levels, even when using overlapping or identical datasets, as we demonstrated using the examples in Parts II and III of this tutorial.

In classical approaches to seismic hazard analysis (e.g. PSHA), epistemic uncertainty is handled using logic trees: data structures that are used to systematically define reasonable alternatives to modelling decisions or parameters. Logic trees consist of alternative models or parameters that are assigned weights according to their validity as presumed by the model author/s. First introduced by Kulkarni et al. (1984), the logic tree approach has become the state-of-practice in PSHA (e.g. Bommer and Scherbaum, 2008). Figure 7.1 depicts a generic logic tree.

Figure 7.1 – An example of a logic tree from the OpenQuake User's Manual, labelled using the OpenQuake terminology. The logic tree is divided into hierarchical branching levels, each of which represents a specific epistemic. The model begins at Branching level 1. Then, Branching Level 2 introduces an epistemic uncertainty with two branches, the green ones, each of which have two alternatives in branching Level 3, the ones colored blue, for a total of four branches. The schematic shows how logic trees can easily grow to have a large number of end branches



The complexity and comprehensiveness of logic trees used in PSHA varies based on the utility of the model, its maturity (e.g., its version), and the agencies and experts involved in costructing the input model. For example, site-specific hazard studies, and especially those produced for critical facilities, are typically more comprehensive in their logic tree structures than national or regional PSHA. The most comprehensive logic trees aim to be mutually exclusive and collectively exhaustive (e.g Bommer and Scherbaum, 2008; Jaynes, 2003) for each included epistemic uncertainty (in OpenQuake Engine terminology, for the branches in a branchset), meaning that the alternative parametrizations inherently reject each other, and together cover the range of possible solutions.

In this tutorial, we introduce the basics of the two logic trees used in the OpenQuake Engine: the seismic source logic tree and the ground-motion logic tree. We additionally recommend the following resources:

- the Recommendations prepared by the Senior Seismic Hazard Analysis Committee (Budnitz et al., 1997)
- Chapter 5 of the OpenQuake Engine Underlying Hazard Science and references therein

8 Seismic source model logic tree

The epistemic uncertainty considered in a seismic source characterisation can be generalized in two groups:

1. alternative “initial" seismic source models
2. alternative source parameterizations

8.0.1 Alternative seismic source models

A seismic source model is meant to account for all the seismic sources that could produce ground motions stronger than a threshold of concern for sites within a region of interest. The logic tree structure allows for two (or more) seismic source models the have been characterized entirely independently. Thus, the source models must be weighted against each other at the highest branching level of the logic tree (e.g. at the "root" of the tree).

There are a few fundamental aspects of seismic source models that could merit including more than one in a PSHA input model, including the following:

- **source typology**. For example, one model could use faults plus smoothed seismicity, while the alternative model uses area sources with internally homogenous activity rates. In this case, the alternative models make different interpretations of how strongly the known faults and recorded seismicity will correspond the positions of seismicity within the investigation time.
- **resolution**. For example, one source models may comprise a more sources/source zones with smaller dimensions. Smaller or more discrete sources and source zones define the positions of seismicity more precisely, but suffer from lesser data available to characterise their rates (i.e. they might have larger epistemic uncertainty affecting the parameters describing their earthquake occurrence).
- **basis for source delineation**. For example, one source model may be based on geodesy, while another is based on geology or seismicity.

8.0.2 Alternative source parameterizations

In some cases, epistemic uncertainty is more appropriately incorporated into the parametrization of individual seismic sources, rather than by changing the seismic source model altogether. This strategy is advantageous when discrete source parameters are difficult to constrain with high confidence, for example:

- when uncertainties affect one specific source

- when modelling assumptions largely impact source parameter values, and ultimately the hazard results, or
- when imprecise or sparse data are used to resolve source parameters.

While there are certain state-of-practice methodologies used to construct a seismic source model, many of these methodologies still depend on some modeller input, for example, the selection of magnitude scaling relationships, the type of MFD to assign to a source, and the maximum magnitude. The examples in Part II demonstrated that choices made during catalogue processing—such as the completeness table and the declustering algorithm—can produce significantly different subcatalogues, and subsequently very different MFDs. Thus, a modeller may choose to include two or more alternative MFDs for one seismic source. (NB: this would result in numerous alternative initial SSMs.)

Uncertainty is inherent in many data types used to build seismic hazard models, often due to short observation periods relative to the processes observed, sub-optimal measurement positioning, or very low numbers of data points. For example, fault slip rates derived from geodesy (e.g. GPS), paleoseismic trenching, or other geochronology often consist of a minimum, maximum, and preferred value that together span a few mm/yr. For slow moving faults, the range of slip rates may be larger than the preferred slip rate itself. As in the former example, a modeller may choose to include multiple alternative MFDs for a fault source.

The OpenQuake Engine seismic source logic tree supports several epistemic uncertainties that are commonly incorporated into seismic source characterisations, including seismic source models, Gutenberg-Richter a- and b-values, and maximum magnitude. Please read the OpenQuake Engine User Instruction Manual and the OpenQuake Engine Underlying Hazard Science for a complete list of supported uncertainties and further information on constructing a seismic source logic tree for the OpenQuake Engine.

8.0.3 Sensitivity analysis

Treatment of epistemic uncertainty is fundamental to robustly calculating seismic hazard. However, each uncertainty added to a seismic hazard model increases the computational demands of the seismic hazard model. In order to keep the uncertainties managable and meaningful, the modeller must conduct sensitivity tests, which are used to to identify parameters that, when varied, change the hazard results. The results of sensitivity analysis can help the modeller to prioritize the impactful parameters in the logic tree.

9 Ground motion model logic tree

The ground motion logic tree is simpler in that only one epistemic uncertainty is supported: that of the ground motion model for any of the teconic region in the model. The state-of-practice in PSHA is to include at least two ground motion models per tectonic region. While individual ground motion models are accompanied by a term σ that represents aleatory uncertainty, the true variability or randomness in observed ground motion for the same parametrization, σ is usually defined for a given spectral period or magnitude, and effectively shifts an attenuation curve vertically in log space. On the other hand, alternative ground-motion models may have different functional forms, and so using a set of GMPEs may yield varying rates of attenuation and relative ground-motions among magnitudes or periods; this means the logic tree is closer to “mutually exclusive and collectively exhaustive" (e.g Bommer and Scherbaum, 2008; Jaynes, 2003).

There is no universally used method for weighting ground motion models, in particular when strong-motion data is unavailable. However, some quantitative approaches exist for data-driven ground motion model selection. The Scherbaum et al. (2009) method, which uses likelihood priniciples, and the Kale and Akkar (2013), which uses Euclidean distance-based ranking, are implemented in the GMPE-SMTK.

Part V

PSHA results

9.0.1 Hazard results

Together, the components described in Parts II–IV comprise the input required for PSHA. Here, we introduce the major types of hazard results produced by Classical PSHA:

1. hazard curves
2. hazard maps
3. uniform hazard spectra

This manual focuses on understanding the different result types. Please read the OpenQuake Engine User Instruction Manual to learn how to set up and execute OpenQuake Classical PSHA jobs and the OpenQuake Underlying Hazard Science Chapter 2 to understand the mathematics used by the OpenQuake calculators.

9.0.2 Probability versus rate of exeedance

The state-of-practice PSHA softwares compute hazard results either in terms of annual rates of exceedance, also called annual frequency of exceedence (AFE), or probabilities of exceedance in a given time span. They differ in that the former is independent of the investigation time, while probabilities correspond to a specific time period. AFE is the inverse of a return period. For an investigation time of 1 year, probabilities and rates of exceedance are nearly equal, and thus the analyst must be attentive to the spefications of their software. The OpenQuake Engine computes results in terms of probabilities of exceedance.

Equation 9.1 shows the conversion from rate to probability, where $P(t)$ is the probability given investigation time t and λ is the rate of occurrence. Table 9.1 lists some of the numbers that are of interest to engineering applications in terms of annual probability of exceedance, rate of exceedance, and return period.

$$P(t) = 1 - e^{-\lambda t} \tag{9.1}$$

Table 9.1 – Common recurrences in terms of annual probability of exceedance (POE), rate of exceedance, and return period.

	Annual POE	Rate of exceedance (1/yr)	Return period (yrs)
10% in 50 yrs	0.002105	0.002107	475
2% in 50 yrs	0.000404	0.000404	2475

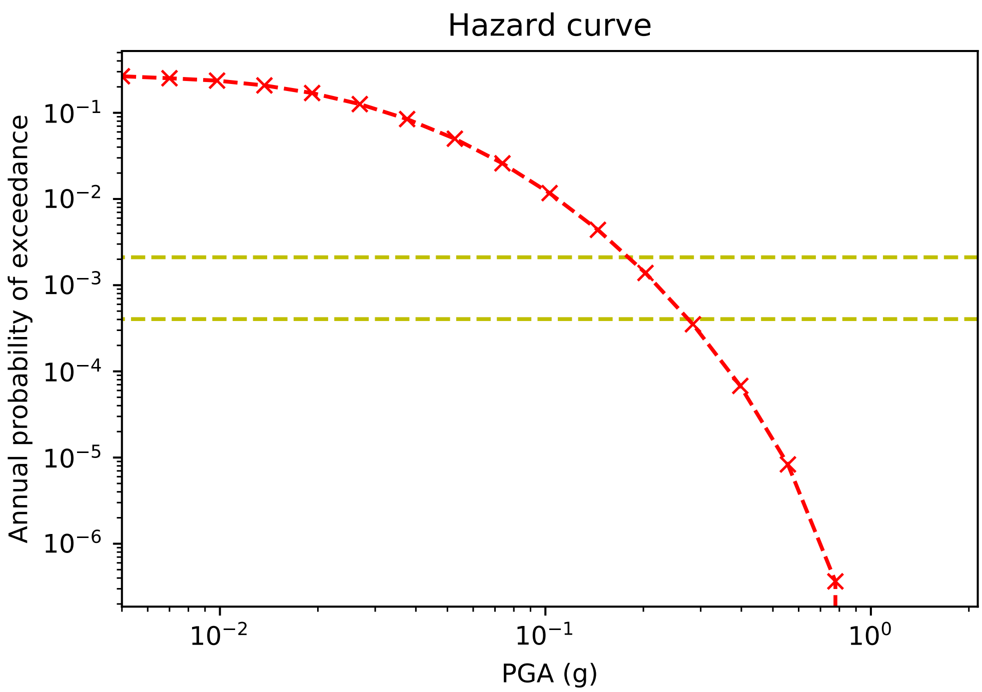
10 Types of hazard results

10.0.1 Hazard curves

Classical PSHA computes hazard curves as the primary calculation result. A hazard curve is the calculated probabilities (or rates) that considering all the sources included in a seismic source model, a set of ground motion parameter levels will be exceeded. Each hazard curve corresponds to a single site, intensity measure type (e.g. PGA, SA at some period), and site condition. Hazard curves reported as probabilities are applicable to the investigation time specified in the calculation, while rates are for annual periods.

Figure 10.1 shows an example of a hazard curve.

Figure 10.1 – An example of hazard curve in terms of probabilities of exceedance (POEs) for PGA levels with a 1 year investigation period (e.g. annual POEs). Common return periods of interest are indicated.



Realizations and mean hazard

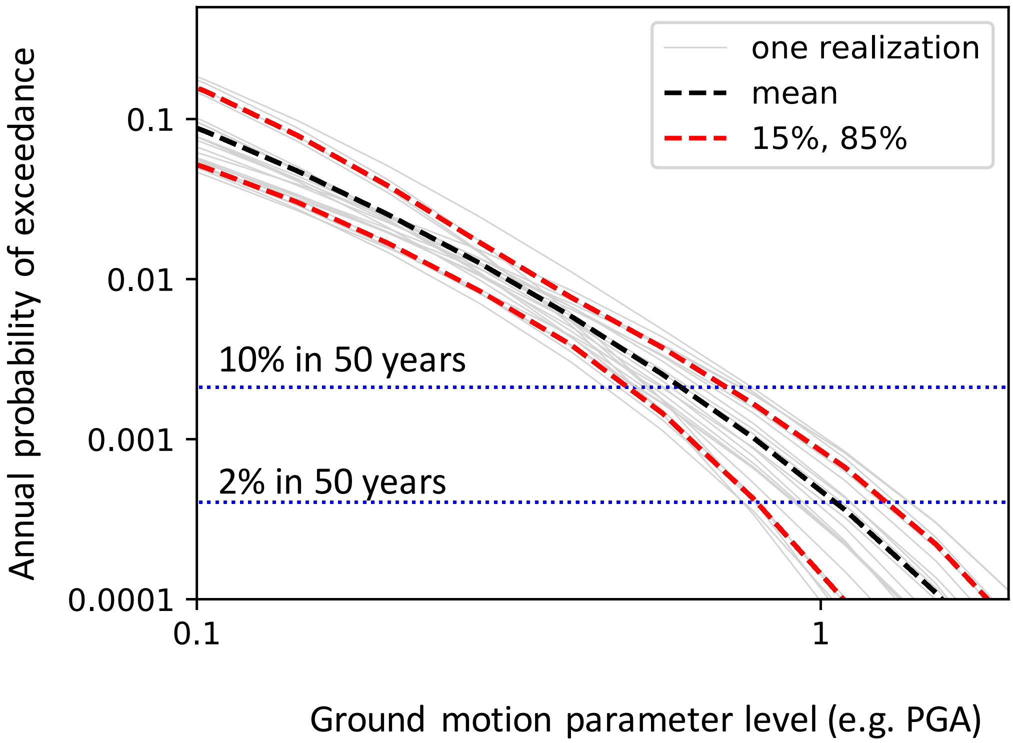
In practice, a classical PSHA calculation computes multiple hazard curves for each ground motion parameter at each site: one for each realization of the model. A realization is a unique pair of end branches, one each from the seismic source and ground motion logic trees, and thus a model has M X N realizations where M is the number of source model logic tree end branches and N is the number of ground motion logic tree end branches.

Together, all realizations represent the range of probabilities for a site's ground motion parameter levels that are permissible by the PSHA input model, thus reflecting parameter uncertainties. However, the applications of PSHA are typically more interested in statistical parameters of the computed results, and in particular the mean hazard and some quantiles.

The mean hazard curve is computed using the individual realization probabilities and the weights of their associated end branches as defined by the logic tree. Quantiles are usually computed by interpolating between the values of the realizations, and sampling accordingly. Please refer to the OpenQuake Underlying Hazard Science Chapter 5.3.3 to see the mathematics of hazard statistics as computed by the OpenQuake Engine.

Figure 10.2 shows an example of the mean hazard curve as well as the 0.15, 0.50, and 0.85 quantiles hazard results for a model with numerous realizations.

Figure 10.2 – An example of mean and quantile hazard curves.

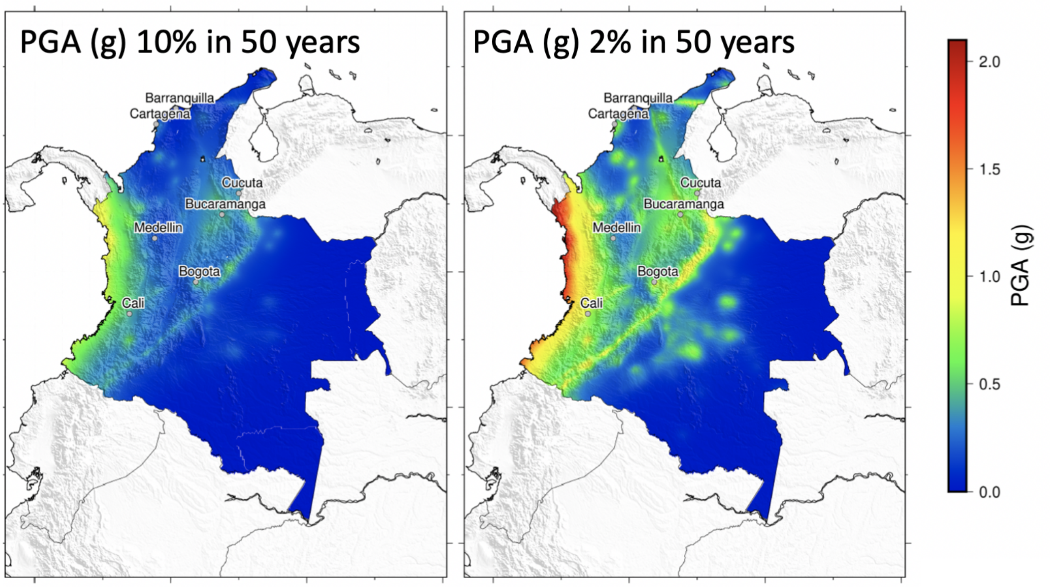


10.0.2 Hazard maps

Hazard curves are useful for understanding how the probability changes for different intensity measure levels. However, they focus on individual sites. To understand how seismic hazard varies geospatially across the region of interest, hazard maps are derived from the hazard curves for all sites. Usually, a seismic hazard map shows how the ground motion parameter (e.g. PGA or SA at some period) varies across the mapped region for a fixed probability of exceedance during an investigation period (e.g. 10% or 2% POE in 50 years). Hazard maps can depict the mean hazard or a quantile of interest, and can be derived for either a reference site condition (e.g. rock or soft soil, see Section 5.2) or for the local site conditions (e.g. using).

Figure 10.3 shows two hazard maps for PGA with 10% and 2% POE in 50 years for a constant computed for a simple seismic hazard model.

Figure 10.3 – Example hazard maps covering Colombia showing mean PGA with (left) 10% POE and (right) 2% POE in 50 years (NB: the maps are for demonstration purposes only).



10.0.3 Uniform Hazard Spectra

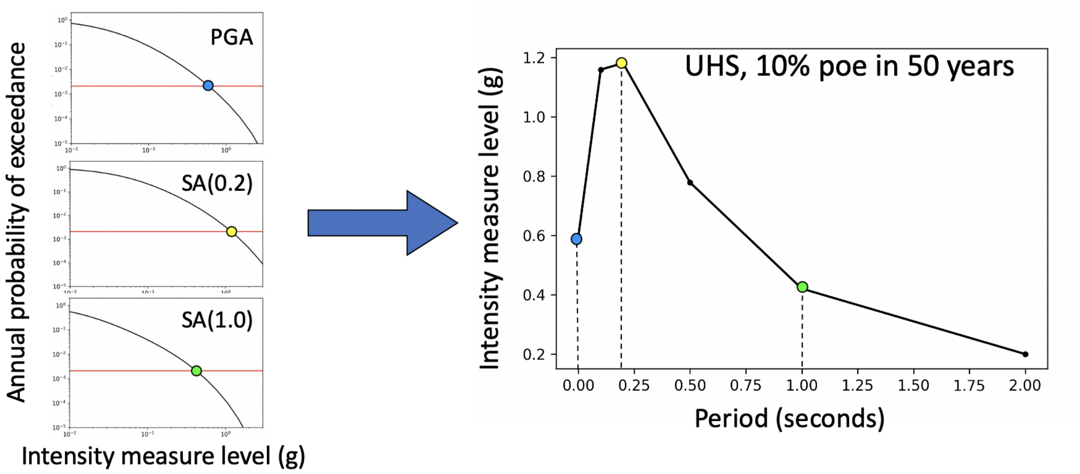
Hazard curves for different spectral periods can be be combined into a uniform hazard spectrum (UHS) that shows, for a certain probability value, the shaking intensity that expected to be imposed on structures with different resonant periods. In this case, hazard curves for sites of interest are computed for a range of spectral periods and fixed site condition, and then the intensity measure levels for a fixed POE are joined. Like hazard maps, uniform hazard spectra correspond to a probability of exceedance for a specified investigation period.

Figure 10.4 shows an example UHS.

10.0.4 Disaggregation

The PSHA outputs described so far have reflected the seismic hazard contributions from all the sources and their possible ruptures and ground motions that may impact a site.

Figure 10.4 – An example UHS with 10% POE in 50 years. Left: Hazard curves for several spectral periods used to construct a UHS, with filled circles indicating where the curve intersects 10% POE in 50 years. Right: the UHS constructed from these (and more) hazard curves.



However, for some applications, i.e. in engineering or risk analysis, it may be useful to focus on independent earthquake ruptures, those that are contriubting most to the computed intensity measure level for a given annual POE and spectral period. For example, one may want to know which rupture distances and magnitudes are contributing most, and which tectonic region types produce those ruptures.

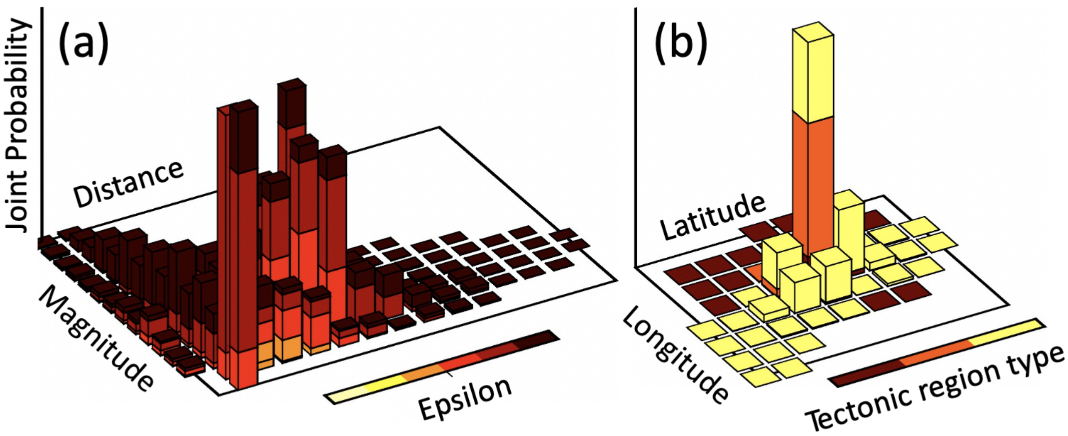
The method used to identify which sources and ruptures contribute most significantly is called seismic hazard disaggregation (sometimes called deaggregation). Disaggregation reveals the relative contributions to a given frequency of exceedance of the different independent variables considered in the hazard integrals, such as magnitude, distance, or epsilon (the coefficient multiplied by the GMPE sigma) or other categorical parameter, e.g. tectonic region type. (Bazzurro and Cornell, 1999) and (Pagani and Marcellini, 2007) describe some of the methodologies used to disaggregate the seismic hazard. The approach used by the OpenQuake Engine is described in the OpenQuake Engine Underlying Hazard Science and the OpenQuake Engine User Instruction Manual.

Figure 10.5 shows two possible configurations of a disaggregation calculation: magnitude, distance, and the GMPE standard deviation "epsilon" (the coefficent multiplied by GMPE sigma); and latitude, longitude, and tectonic region type.

10.1 Examples

- ***12_simple_PSHA_calculations.ipynb***: uses the OpenQuake Engine as a library for performing simple hazard calculations, and reviews common hazard outputs

Figure 10.5 – Two disaggregation configurations. Left: magnitude, distance, and the GMPE standard deviation "epsilon". Right: latitude, longitude, and tectonic region type. In each, the height of the column indicates the contribution from that parameter bin.



11 Bibliography

Books

Campbell, K. W. and Y. Bozorgnia (2007). Campbell-Bozorgnia NGA ground motion relations for the geometric mean horizontal component of peak and spectral ground motion parameters. Pacific Earthquake Engineering Research Center (cited on page 60).

Jaynes, E. T. (2003). Probability theory: The logic of science. Cambridge university press (cited on pages 70, 73).

McGuire, R. K. (2004). Seismic Hazard and Risk Analysis. EERI (cited on page 5).

Articles

Abrahamson, N. A. and W. Silva (2008). "Summary of the Abrahamson & Silva NGA Ground-Motion Relations". In: Earthquake Spectra 24.1, pages 67–97 (cited on page 60).

Atik, L. A., N. Abrahamson, J. J. Bommer, F. Scherbaum, F. Cotton, and N. Kuehn (2010). "The variability of ground-motion prediction models and its components". In: Seismological Research Letters 81.5, pages 794–801 (cited on page 62).

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 page 81).

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 pages 66, 69, 70, 73).

Bommer, J. J., J. Douglas, F. Scherbaum, F. Cotton, H. Bungum, and D. Fah (Sept. 2010). "On the Selection of Ground-Motion Prediction Equations for Seismic Hazard Analysis". In: Seismological Research Letters 81.5, pages 783–793. ISSN: 0895-0695. DOI: 10.1785/gssrl.81.5.783. eprint: https://pubs.geoscienceworld.org/srl/article-pdf/81/5/783/2762378/783.pdf. URL: https://doi.org/10.1785/gssrl.81.5.783 (cited on page 63).

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 96.4A, pages 1502–1511 (cited on page 53).

Borcherdt, R. D. (1994). "Estimates of site-dependent response spectra for design (methodology and justification)". In: Earthquake spectra 10.4, pages 617–653 (cited on page 57).

Borcherdt, R. D. and G. Glassmoyer (1992). "On the characteristics of local geology and their influence on ground motions generated by the Loma Prieta earthquake in the San Francisco Bay region, California". In: Bulletin of the seismological society of America 82.2, pages 603–641 (cited on page 57).

Campbell, K. W. and Y. Bozorgnia (2014). “NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra”. In: Earthquake Spectra 30.3, pages 1087–1115 (cited on page 60).

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 page 60).

Chiou, B. S.-J. and R. R. Youngs (2014). “Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra”. In: Earthquake Spectra 30.3, pages 1117–1153 (cited on page 60).

Cornell, C. A. (1968). “Engineering Seismic Risk Analysis”. In: Bulletin of the Seismological Society of America 58, pages 1583–1606 (cited on page 11).

Cotton, F., F. Scherbaum, J. Bommer, and H. Bungum (Apr. 2006). “Criteria for Selecting and Adjusting Ground-Motion Models for Specific Target Regions: Application to Central Europe and Rock Sites”. In: Journal of Seismology 10, pages 137–156. DOI: 10.1007/s10950-005-9006-7 (cited on page 63).

Douglas, J. and B. Edwards (2016). “Recent and future developments in earthquake ground motion estimation”. In: Earth-Science Reviews 160, pages 203–219 (cited on page 54).

Field, E. H., R. J. Arrowsmith, G. P. Biasi, P. Bird, T. E. Dawson, K. R. Felzer, D. D. Jackson, K. M. Johnson, T. H. Jordan, C. Madden, A. J. Michael, K. R. Milner, M. T. Page, T. Parsons, P. M. Powers, B. E. Shaw, W. R. Thatcher, R. J. Weldon, and Y. Zeng (June 2014). “Uniform California Earthquake Rupture Forecast, Version 3 (UCERF3)—The Time-Independent Model”. en. In: Bulletin of the Seismological Society of America 104.3, pages 1122–1180. ISSN: 0037-1106, 1943-3573. DOI: 10.1785/0120130164. URL: http://bssa.geoscienceworld.org/content/104/3/1122 (visited on 05/05/2016) (cited on page 43).

Frankel, A. (1995). “Mapping Seismic Hazard in the Central and Eastern United States”. In: Seismological Research Letters 66.4, pages 8–21 (cited on pages 47, 89).

Garcia, D., D. J. Wald, and M. Hearne (2012). “A global earthquake discrimination scheme to optimize ground-motion prediction equation selection”. In: Bulletin of the Seismological Society of America 102.1, pages 185–203 (cited on page 32).

Gardner, J. K. and L. Knopoff (1974). “Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian?” In: Bulletin of the Seismological Society of America 64.5, pages 1363 –1367 (cited on page 33).

Gutenberg, B. and C. F. Richter (1944). “Frequency of Earthquakes in California”. In: Bulletin of the Seismological Society of America 34, pages 185 –188 (cited on page 44).

Hale, C, N Abrahamson, and Y Bozorgnia (2018). “Probabilistic seismic hazard analysis code verification”. In: Pacific Earthquake Engineering Research Center, Report 3 (cited on page 11).

Hayes, G. P., D. J. Wald, and R. L. Johnson (2012). “Slab1. 0: A three-dimensional model of global subduction zone geometries”. In: Journal of Geophysical Research: Solid Earth 117.B1 (cited on page 29).

Hayes, G. P., G. L. Moore, D. E. Portner, M. Hearne, H. Flamme, M. Furtney, and G. M. Smoczyk (2018). “Slab2, a comprehensive subduction zone geometry model”. In: Science 362.6410, pages 58–61 (cited on pages 29, 42).

Kale, O. and S. Akkar (2013). “A new procedure for selecting and ranking ground-motion prediction equations (GMPEs): The Euclidean distance-based ranking (EDR) method”. In: Bulletin of the Seismological Society of America 103.2A, pages 1069–1084 (cited on pages 66, 73).

Kijko, A. (2004). “Estimation of the Maximum Earthquake Magnitude, M_{MAX}”. In: Pure and Applied Geophysics 161, pages 1655 –1681 (cited on pages 48, 49).

Mak, S., R. A. Clements, and D. Schorlemmer (2017). “Empirical evaluation of hierarchical ground-motion models: Score uncertainty and model weighting”. In: Bulletin of the Seismological Society of America 107.2, pages 949–965 (cited on page 65).

Makropoulos, K. C. and P. W. Burton (1983). “Seismic Risk of Circum-Pacific Earthquakes I. Strain Energy Release”. In: Pure and Applied Geophysics 121.2, pages 247 –266 (cited on page 49).

Pagani, M. and A. Marcellini (2007). “Seismic-Hazard Disaggregation: A Fully Probabilistic Methodology”. In: Bulletin of the Seismological Society of America 97.5, pages 1688–1701. DOI: 10.1785/0120060126 (cited on page 81).

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 2014). “OpenQuake Engine: An Open Hazard (and Risk) Software for the Global Earthquake Model”. en. In: Seismological Research Letters 85.3, pages 692–702. ISSN: 0895-0695, 1938-2057. DOI: 10.1785/0220130087. URL: http://srl.geoscienceworld.org/content/85/3/692 (visited on 10/20/2014) (cited on pages 11, 17).

Pagani, M, K Johnson, and J. G. Palaez (2020). “Modelling subduction sources for probabilistic seismic hazard analysis”. In: Geological Society, London, Special Publications 501 (cited on page 32).

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 65).

Scherbaum, F., E. Delavaud, and C. Riggelsen (2009). “Model selection in seismic hazard analysis: An information-theoretic perspective”. In: Bulletin of the Seismological Society of America 99.6, pages 3234–3247 (cited on pages 65, 73).

Schwartz, D. P. and K. J. Coppersmith (1984). “Fault Behaviour and Characteristic Earthquakes: Examples from the Wasatch and San Andreas fault zones”. In: Journal of Geophysical Research 89.B7, pages 5681–5698 (cited on pages 44, 45).

Silva, V., H. Crowley, M. Pagani, D. Monelli, and R. Pinho (2014). "Development of the OpenQuake engine, the Global Earthquake Model's open-source software for seismic risk assessment". In: Natural Hazards 72.3, pages 1409–1427 (cited on page 11).

Stewart, J. P., J. Douglas, C. Di Alessandro, Y. Bozorgnia, N. A. Abrahamson, D. M. Boore, K. W. Campbell, E. Delavaud, M. Erdik, and P. J. Stafford (2012). "Selection of a global set of GMPEs for the GEM-PEER Global GMPEs Project". In: (cited on page 65).

Stirling, M., T. Goded, K. Berryman, and N. Litchfield (2013). "Selection of earthquake scaling relationships for seismic-hazard analysis". In: Bulletin of the Seismological Society of America 103.6, pages 2993–3011 (cited on page 42).

Weatherill, G. A., M. Pagani, and J. Garcia (Sept. 2016). "Exploring earthquake databases for the creation of magnitude-homogeneous catalogues: tools for application on a regional and global scale". en. In: Geophysical Journal International 206.3, pages 1652–1676. ISSN: 0956-540X. DOI: 10.1093/gji/ggw232. URL: https://academic.oup.com/gji/article/206/3/1652/2583518 (visited on 03/22/2019) (cited on pages 22, 23, 27, 30).

Weichert, D. H. (1980). "Estimation of the Earthquake Recurrance Parameters for Unequal Observation Periods for Different Magnitudes". In: Bulletin of the Seismological Society of America 70.4, pages 1337 –1346 (cited on page 45).

Youngs, R. R. and K. J. Coppersmith (Apr. 1985). "Implications of fault slip rates and earthquake recurrence models to probabilistic seismic hazard estimates". In: Bulletin of the Seismological Society of America 75, pages 939–964 (cited on pages 44, 46, 48).

Zhao, J. X., S. Zhou, P. Gao, T. Long, Y. Zhang, H. K. Thio, M. Lu, and D. A. Rhoades (2015). "An earthquake classification scheme adapted for Japan determined by the goodness of fit for ground-motion prediction equations". In: Bulletin of the Seismological Society of America 105.5, pages 2750–2763 (cited on page 32).

Zhu, C., F. Cotton, and M. Pilz (2019). "Testing the depths to 1.0 and 2.5 km/s velocity isosurfaces in a velocity model for Japan and implications for ground-motion modeling". In: Bulletin of the Seismological Society of America 109.6, pages 2710–2721 (cited on page 57).

Other Sources

Allen, T. I. and D. J. Wald (2007). Topographic slope as a proxy for seismic site-conditions (VS30) and amplification around the globe. Technical report. Geological Survey (US) (cited on pages 57, 59).

Budnitz, R., G Apostolakis, and D. Boore (1997). Recommendations for probabilistic seismic hazard analysis: guidance on uncertainty and use of experts. Technical report. Nuclear Regulatory Commission, Washington, DC (United States). Div. of . . . (cited on page 70).

Johnston, A., L. Kanter, K. Coppersmith, and C. Cornell (1994). The earthquakes of stable continental regions. Volume 1, Assessment of large earthquake potential, Final report.

Technical report. Electric Power Research Inst., Palo Alto, CA (United States); Memphis State ... (cited on page 48).

Kulkarni, R., R. Youngs, and K. Coppersmith (1984). "Assessment of confidence intervals for results of seismic hazard analysis". In: Proceedings of the eighth world conference on earthquake engineering. Volume 1, pages 263–270 (cited on page 69).

Mignan, A., J. Woessner, and D. Scherlemmer (2010). Theme IV - Understanding Seismicity Catalogs and their Problems. Technical report doi: 10.5078/corssa-00180805. Community Online Resource for Statistical Seismicity Analysis (cited on page 34).

Stepp, J. C. (1971). "An investigation of earthquake risk in the Puget Sound area by the use of the type I distribution of largest extreme". PhD thesis. Pennsylvania State University (cited on page 34).

Stiphout, T. van, J. Zhuang, and D. Marsan (2010). Theme V - Basic Techiniques/Models for Analysing Seismicity. Technical report. Community Online Resource for Statistical Seismicity Analysis. URL: http://www.corssa.org (cited on page 33).

Glossary

activity rate

See magnitude-frequency distribution.

aleatory uncertainty

Uncertainty in a phenomenon that is due to true variability..

annual frequency of exceedence

The frequency in 1/yr at which a ground motion level is expected to be exceeded..

annual rate of occurrence

The average number of occurences in a one year period, e.g. earthquakes of a given magnitude by a source..

area source

A source type usually adopted to model distributed seismicity. In an area source the seismicity occurrence rate is assumed uniform over the source area; this produces an hazard pattern with a plateau of constant hazard inside the polygon delimiting the area source and values of hazard that tend to decrease as we move away from the border of the source.

centroid

the center of the energy release of an earthquake.

characteristic MFD

An magnitude-frequency distribution in which all or most moment release occurs at one magnitude or over a small range of magnitudes..

classified

grouped based on a category such as tectonic region, i.e. a classified catalogue..

completeness magnitude

The lower magnitude threshold above which all earthquakes occurring within the region covered by a catalogue are thought to be included..

cumulative MFD

An magnitude-frequency distribution in which the given annual occurrence rate corresponds to magnitudes greater than or equal to the respective magnitude..

declustering

The procedure by which independent earthquakes (mainshocks) in a catalogue are identified and separated from other occurrences (foreshocks and aftershocks)..

distance

i.e. the distance between an earthquake and a site. There are several ways to measure distance, e.g. see r_{rup}.

distributed

i.e. distributed seismicity. a collection of possible seismic source with less definitive geometry.

earthquake catalogue

A database of earthquakes, where each entry include the basic information about an earthquake on record, at minimum including the earthquake time, magnitude, and hypocentral location.

earthquake rupture

A 3D surface - representing a portion or the entire fault surface - over which a slip event (i.e. an earthquake) occurs.

epistemic uncertainty

The type of uncertainty that results from limited knowledge or data about a phenomenon.

event

In seismology, a term sometimes used to refer to one earthquake.

fault

Discontinuities within the earth's crust where rock has been crushed or fractured, separating the surrounding earth into separate blocks. See also the OpenSHA website.

fault trace

A curve representing the intersection between the surface containing the fault surface (or its prolongation) and the topographic surface .

focal mechanism

Geometrical or mathematical representations of the faulting that occurs during an earthquake Earthquake, Focal Mechanism.

global catalogue

Earthquake catalogues with global coverage..

grid source

A source typology usually adopted to model distributed seismicity. It is routinely produced by a seismicity smoothing algorithm (one of the most famous algorithm is the one proposed by Frankel (1995)).

ground-motion logic tree

A method used to systematically describe the epistemic uncertainties related to the ground motion models used in the computation of hazard using a specific PSHA input model.

ground-motion model

An object that given a rupture with specific properties computes the expected ground motion at the given site. In simplest case a ground motion model corresponds to a ground-motion prediction equation. In case of complex PSHA input models, the produced ground motion models contains a set of Ground Motion Prediction Equations (GMPEs), one for each tectonic region considered.

ground-motion prediction equation

An equation that - given some fundamental parameters characterizing the source, the propagation path and the site (in the simplest case magnitude, distance and $V_{S,30}$) - computes the value GM of a (scalar) ground motion intensity parameter.

hazard curve

A set of intensity measure levels and their corresponding probabilities of exceedance for a single site, site condition, and investigation time..

hazard map

The spatial distribution of intensity measure levels for a fixed probability of exceedance and site condition..

historical catalogue

A compilation of past earthquakes and their parameters comprising events that occurred before the time of seismometers, but that were felt by humans..

hypocenter

The 3D nucleation point of an earthquake rupture.

incremental MFD

An magnitude-frequency distribution in which the given annual occurrence rate corresponds to magnitudes only of the specified magnitude. In practice, e.g. in the OpenQuake Engine, the specified magnitude is a magnitude range defined by some bin width and center or edge magnitude value..

initial seismic source input model

It is the ensemble of information needed to fully describe the seismic sources composing a seismic source input model. The initial seismic source input model is included in the first branching level of a seismic source logic tree.

input model

An object containing the information necessary to describe the seismic source and the ground motion models - plus the related epistemic uncertainties.

instrumental catalogue

A compilation of past earthquakes and their parameters comprising events for which the event information is obtained by processing seismometer data..

length

i.e. Fault length; the total length of the fault trace..

local catalogue

Earthquake catalogues with higher density station coverage but covering only sections of the globe (<100 km-scale)..

logic tree

Data structure used to systematically describe uncertainties on parameters and models used in a PSHA study.

magnitude

The size of an earthquake.

magnitude homogenization

The process by which the magnitudes for a set of earthquakes are converted to use a single measure of magnitude..

magnitude-frequency distribution

A distribution describing the frequency of earthquakes with a specific magnitude. It can be continuous or discrete. One frequency- magnitude distribution frequently adopted in Classical PSHA (PSHA) is the double truncated Gutenberg-Richter distribution.

magnitude-scaling relationship

An empirical relationship linking the magnitude with a parameter describing the size of the corresponding rupture (e.g. the area of the rupture or the rupture length).

nodal plane

In seismology, the two great circles that separate the extensional and compressional quadrants of a focal mechanism..

paleoseismic trench

An elongate hole excavated across strata offset by a fault and used to reveal a sequence of depositional and offsetting events.

peak ground acceleration

The highest acceleration experienced by a particle on the ground during an earthquake. See also the USGS webpage: Earthquake Hazards 201.

point source

The elemental source typology used in the OpenQuake-engine to model distributed seismicity.

pre-historic catalogue

A compilation of past earthquakes and their parameters comprising events that occurred before humans, which are revealed through geological investigations..

prime

The preferred earthquake origin parameters as determined by the organization reporting the data..

probabilistic seismic hazard analysis

A methodology to compute seismic hazard by taking into account the potential contributions coming from all the sources of engineering importance for a specified site.

probability of exceedance

In PSHA, the probability that a ground-motion value for a given intensity measure type will be exceeded during an investigation time..

rake

The rake is the direction in which a hanging wall block moves during a rupture, measured relative to fault strike on the plane of the fault.

rate of exceedance

The number of times that an intensity measure level (e.g. a ground-motion value generated by an earthquake) is exceeded..

realization

The hazard results computed from a single end branch of the logic tree for a input model..

recurrence interval

The average time between consecutive earthquakes on a fault.

regional catalogue

Earthquake catalogues with higher density station coverage but covering only sections of the globe (>100 km-scale)..

residual analysis

A quantitative method of comparing strong motion data to GMPE predicted values in order to evaluate or select GMPEs for a PSHA model.

response spectra

The magnitude of a parameter describing strong ground motion (e.g. acceleration, velocity, displacement) as a function of the natural vibrational period.

rupture aspect ratio

The ratio between the lenght and the width of an earthquake rupture.

scarp

The surface expression of an earthquake, often manifesting as a step in topography across a fault..

seismic hazard analysis

A procedure used to determine the level of danger that may occur due to earthquakes within an area of interest over a given time frame. See also the OpenSHA website.

seismic hazard disaggregation

A methodology to investigate the contributions to a specific level of hazard in terms of fundamental variables commonly used to characterize seismic sources and ground motion models (e.g. magnitude, source-site distance, epsilon.

seismic source

An object that can generate earthquake ruptures.

seismic source logic tree
Logic tree structure defined to describe in structured and systematic way the epistemic uncertainties characterizing the seismic source model. The first branching level in the logic tree by definition contains one or several alternative initial seismic source input model.

seismic source model
An object containing a list of seismic sources objects.

seismogenic zone
the part of the brittle crust that can fail elastically in an earthquake.

shear wave isosurface
A source typology usually adopted to model shallow structures with an uncomplicated geometry.

slip
The distance that the fault blocks move relative to each other during an earthquake. Slip can also occur aseismically during slow-slip or creep events.

slip rate
the rate at which the two sides of a fault slip relative to each other. See also the USGS Earthquake Glossary: slip rate..

smoothed seismicity
An approach to modelling distributed seismicity, in which occurrence rates over a grid source are variable and based on the locations of observed earthquakes..

source
See seismic source..

source typology
A geometric representations of modelled seismicity, which vary depending on the level of knowledge regarding these sources, and depends on the degree of confidence in the positions of the ruptures that can be generated..

strong motion
Ground motions strong enough that they may cause structural damage..

surface area
The contact area between the two blocks that comprise a fault.

surface rupture
Offset at the ground surface due to an earthquake.

tectonic region
A area on the topographic surface that can be considered homogeneous in terms of tectonic properties such as the prevalent seismogenic properties and/or the seismic wave propagation properties.

time-independent
A modelling assumption that indicates a model is valid for any time frame.

truncated GR MFD

A Gutenberg-Richter MFD that is “cut" at an upper magnitude that is considered the maximum credible earthquake that can be produced by the corresponding seismic source..

uniform hazard spectrum

The response spectra for a single hazard site showing the intensity measure levels for a range of spectral periods with equal probabilities of exceedance..

$V_{S,30}$

Average shear wave velocity of the materials in the uppermost 30m of the soil column.