Beyond Button Pushing
Seismic Risk Assessment for California

Report Prepared by
the Global Earthquake Model Foundation (GEM)

Authors:
Anirudh Rao, Graeme Weatherill,
Vitor Silva, John Schneider

CALIFORNIA SEISMIC SAFETY COMMISSION
1755 Creekside Oaks Drive, Suite 100
Sacramento, CA 95833
www.seismic.ca.gov

CSSC Publication No. 17-05

Table of Contents

Glossary

List of acronyms

AAL: Average annual loss

AALR: Average annual loss ratio

DM: Deformation model

DSR: Slip along rupture

FM: Fault model

GEM: Global Earthquake Model (https://globalquakemodel.org)

GMM: Ground motion model

GSIM: Ground motion simulation model

Hazus: A multi-hazard loss estimation software application provided by the Federal Emergency Management Agency (https://fema.gov/hazus)

M5: Total rate of M≥5 events in the region

MMAX: Maximum magnitude for events occurring off the modeled faults

NGAWEST2: Next Generation Attenuation Relationships for the Western US, Version 2

NSHM: National Seismic Hazard Model

PGA: Peak Ground Acceleration

SM: Site-conditions model

SPATIALPDF: Spatial distribution of off-fault gridded seismicity set by choosing one of the spatial probability density maps

SR: Magnitude scaling relationship

UCERF3: Uniform California Earthquake Rupture Forecast, Version 3

USGS: United States Geological Survey (https://usgs.gov)

Vs_{30}: time-averaged shear-wave velocity in the top 30 m soil layer

WGCEP: Working Groups on California Earthquake Probabilities

List of terms

Earthquake rupture forecast: An earthquake rupture forecast model provides estimates of the magnitude, location, and likelihood of earthquake fault ruptures in a region.

Epistemic uncertainty: Epistemic uncertainty refers to the uncertainty associated with lack of knowledge or insufficient understanding of the underlying processes. Epistemic uncertainty is typically accounted for in probabilistic seismic hazard and risk analysis by using alternative models to represent the underlying process.

Exposure: Exposure describes the situation of people, infrastructure, housing, production capacities and other tangible assets located in hazard-prone areas (Ref: UNISDR 2017 terminology on disaster risk reduction http://preventionweb.net/go/7822).

Stochastic event-set: A simulated catalog of earthquake events in a region within a certain time-span, each associated with a magnitude, location and probability of occurrence.

Probabilistic risk analysis: A probabilistic earthquake risk analysis considers all possible earthquake events that could occur in a region within a certain time span.

Scenario risk analysis: A scenario analysis addresses the risk associated with a single earthquake, such as a repetition of the 1906 San Francisco Earthquake.

Vulnerability: Vulnerability describes the conditions determined by physical, social, economic and environmental factors or processes which increase the susceptibility of an individual, a community, assets or systems to the impacts of hazards (Ref: UNISDR 2017 terminology on disaster risk reduction http://preventionweb.net/go/508). In particular, a seismic vulnerability model describes the susceptibility of buildings and other infrastructure elements to physical damage leading to potential economic and human losses due to earthquake ground shaking.

Executive Summary

Introduction

According to the Working Group on California Earthquake Probabilities, the likelihood of magnitude 6.7 and larger earthquakes occurring somewhere in California in the next 30 years is near certainty. Given the high level of earthquake risk in California, it is crucial to develop policies for risk mitigation, emergency preparedness, risk transfer and insurance, and buildings and infrastructure design that are informed by the best available science on earthquakes. Consequently, there is a need for improved understanding of the seismic risk in the State, including a better characterization of elements that comprise the risk: hazard, exposure, and vulnerability.

There are too few past earthquakes to use history as the only judge of underlying earthquake risk. Earthquake risk models allow us to bridge that gap by combining with history the best available science and engineering knowledge about earthquakes and their effects on society. In so doing, risk models can provide realistic estimates of the likelihood of future earthquakes and the potential societal impacts, such as damage to structures, economic losses, casualties, and business disruption. However, proprietary earthquake risk models are protected intellectual property, containing countless assumptions and methods that are disclosed only as necessary, as such, these models are 'black boxes.' Model users can 'push buttons' (vary the input) to produce results, but the effect of each 'button' remains mostly hidden. The lack of transparency of these models means the uncertainties in the assumptions and model results are not known to users and stakeholders, which diminishes their credibility and hinders their adoption in earthquake risk management.

The GEM (Global Earthquake Model) Foundation collaboratively develops and shares reliable information on earthquake hazards and risks to vulnerable communities worldwide, and promotes risk reduction and mitigation measures. Through global projects, open-source development, and partnerships with institutions worldwide, the GEM Foundation and its collaborators are developing common datasets and models for seismic hazard and risk assessment, cutting-edge software tools, best practice guidelines, and capacity and knowledge sharing mechanisms. The GEM Foundation is supported by public institutions representing national governments, as well as private companies; altogether more than 50 organizations worldwide partner with GEM with a common interest in GEM's principles for collaboration, credibility, openness and public good.

The Alfred E. Alquist Seismic Safety Commission (SSC) engaged the GEM Foundation to quantify and discuss the impact of various assumptions on earthquake model results for California, and also to investigate the treatment of uncertainty within the model. The results of this study will help identify the key factors that influence the risk results in

California and improve understanding of the seismic risk in the state. Such improved knowledge can be leveraged by a wide range of stakeholders within the community to: enhance earthquake risk mitigation strategies; develop appropriate emergency preparedness and response plans; inform risk transfer and insurance mechanisms; and better manage risk-sensitive investment portfolios.

The SSC leveraged over 20 million dollars in funding from GEM supporters that has been used to develop the OpenQuake software package and related datasets which the Beyond Button Pushing project uses.

Project aims and goals

The goal of the project is to show how important the quantification of uncertainty is in estimating and understanding California's earthquake risk using OpenQuake — GEM Foundation's state-of-the-art open source earthquake hazard and risk assessment software. With OpenQuake's plug-and-play capabilities, expert users can individually select or substitute every model component, data, and assumption. This feature will help model users and decision makers to: 1) 'ask the right questions' when evaluating model results; 2) better interpret risk assessment results and gain trust in model results; and 3) make better risk management decisions.

Specifically, this project aims to:

- Establish representative sets of exposure:
    - for the San Francisco Bay Area;
    - for the Southern California region affected by the Shakeout Scenario (See Note 1 below);
    - worldwide (see Note 2 below)
- Choose specific results (risk metrics) to use as a basis of comparison.
- Produce 'baseline' results from OpenQuake, using a 'control' set of assumptions.
- Re-run OpenQuake multiple times, each time varying one assumption or parameter, such as:
    - earthquake probabilities (controlled by assumptions about fault geometries, slip rates, maximum magnitudes);
    - shaking intensity (ground motion model selection);
    - damageability of individual buildings (vulnerability curves);
    - site conditions; and
    - statistical treatment of uncertainty and correlation.

Note 1: The region covered by the ShakeOut Scenario was found to be too large to yield meaningful results if used directly in a sensitivity study, because any geographical influences that are apparent in smaller portfolios are diminished if the entire Shakeout

area is considered. Thus, the regions chosen for sensitivity analysis and comparison in this study included the five metropolitan statistical areas of California with the highest population concentrations: the San Francisco Bay Area, the Greater Sacramento Area, the Los Angeles Metropolitan Area, the San Diego Metropolitan Area, and the Inland Empire[1]. The smaller region of Napa County was also included in the analysis following the 2014 Napa earthquake. Although the sensitivity analysis part of this study focused on the above-mentioned metropolitan statistical areas, the calculation of average annual losses was undertaken for the entire state of California.

Note 2: The worldwide database on building exposure compiled by the GEM Foundation is optimized for use in earthquake risk assessment studies and is available for download through the OpenQuake platform at https://platform.openquake.org/exposure.

Beyond the aims stated at the outset of the project as listed above, several additional objectives were achieved during the course of the project, including the following:

- Implement within OpenQuake the latest seismic hazard model for California based on the recently published Uniform California Earthquake Rupture Forecast version 3 (UCERF3), produced by the U.S. Geological Survey and the Working Group on California Earthquake Probabilities.
- Calculate the average annual loss[2] estimates for all 8,057 census tracts in California, using the seismic hazard model based on UCERF3.
- Establish the range (distribution) of scientifically viable results for the chosen risk metrics by accounting for the various uncertainties in the hazard model.
- Identify the components of the hazard model contributing most to the overall uncertainty in the risk metrics for the different exposure portfolios.
- Implement a model simplification ('logic-tree trimming') software tool to reduce the number of computer runs and greatly speed up the time required for running the risk model for California.

Uniqueness of the project

The OpenQuake implementation of the 2014 seismic hazard model for California is based on the highly complex and scientifically advanced UCERF3 earthquake occurrence (or 'rupture forecast') model and is the first implementation of UCERF3 into an open-source

[1] The **San Francisco Bay Area** includes Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Solano, and Sonoma counties
The **Greater Sacramento Area** includes El Dorado, Nevada, Placer, Sacramento, Sutter, Yolo, and Yuba counties
The **Los Angeles Metropolitan Area** includes Los Angeles and Orange counties
The **San Diego Metropolitan Area** includes San Diego county
The **Inland Empire** includes San Bernardino and Riverside counties
[2] The average annual loss (AAL) estimate is the expected loss per year, averaged over a large number of years.

risk modeling software. The transparent nature of this implementation makes it possible for stakeholders to inspect, analyze and modify all of the individual components and assumptions contained in the model. Whereas Hazus[3] and other commercial risk modeling software typically provide only the average risk metrics (e.g., average estimated losses), the OpenQuake implementation of the California hazard model can provide stakeholders with information about the full range of scientifically viable risk results.

The sensitivity analysis performed in this study is also the first open analysis of the risk to different building portfolios performed using UCERF3. The model simplification ('logic-tree trimming') tool developed by Porter et al (2012)[4] for the hazard model based on UCERF2 has been extended in this study to handle the significantly more complex model based on UCERF3. Whereas the tool introduced by Porter et al. considered only the average annual loss for model simplification, the tool developed as part of this study is more flexible as it allows stakeholders the choice of other risk metrics to be used for the simplification. Specifically, the loss estimates for any return-period, such as the '100-year' or '250-year' loss estimates[5], can be used as the basis for simplifying the model with this tool.

Intended audience and how to use this report

In order to fully benefit from the results of this project, this report should be used in conjunction with the SSC's related projects. This will ensure that the SSC's recommendations to the Governor and Legislature will be holistic, actionable, and measurable policies that will help the regional economy prepare, recover, and become resilient to earthquakes.

This project also benefits diverse stakeholders beyond the SSC. For example:

- Insurers stand to benefit because they will better understand the variability in losses to better distinguish 'good risks' from 'bad risks,' and spur innovation.
- Consumers stand to benefit with more clarity on how to become financially prepared.
- Regulators stand to benefit by having an objective, third-party 'common ground' to increase trust and collaboration with insurers.
- State legislators stand to benefit by having access to a wide range of results from a state-of-the art risk assessment for California.

[3] Hazus is a natural hazards risk assessment software tool developed by the U.S. Federal Emergency Management Agency (FEMA).
[4] Porter, K. A., Field, E. H., & Milner, K. R. (2012). Trimming the UCERF2 Hazard Logic Tree. *Seismological Research Letters*, 83(5), 815–828
[5] The 100-year return period loss is the loss value that is exceeded once every 100 years, averaged over a large number of years. Similarly, the 250-year return period loss is the loss value that is exceeded once every 250 years, averaged over a large number of years.

The transparency of the earthquake risk models allows model users to inspect all assumptions and input variables, and test alternative implementations. The openness of the datasets and models developed makes external peer review possible. Increased transparency and control over the model assumptions instills a greater confidence in the model results among all stakeholders. With deeper insight into the drivers of risk, the public, as consumers of risk information, would have access to more robust information, and therefore be able to make more informed risk management decisions, ultimately leading to greater safety and economic resilience for California.

Appendix C of the main report describes how to run the seismic hazard and risk models for California using OpenQuake, and describes the datasets required to run the various other calculations described in this report.

The current version of OpenQuake is available for users of Windows, MacOS, and Linux operating systems. The installers can be downloaded from: https://github.com/gem/oq-engine/blob/master/doc/installing/overview.md

The datasets and models developed as part of this study are available at: https://gitlab.openquake.org/risk/usa/tree/master/models

Datasets used and methods developed

The seismic risk calculations performed within this project are based upon the time-independent[6] version of the recently published Uniform California Earthquake Rupture Forecast version 3 (UCERF3[7]), produced by the U.S. Geological Survey and the Working Group on California Earthquake Probabilities (WGCEP). This earthquake forecast model has been implemented within OpenQuake. In addition, the suite of five ground motion models identified for use in the western United States within the 2014 National Seismic Hazard Mapping Program have also been incorporated into OpenQuake and were used in this project.

[6] 'Time-independent' means that the probability of earthquake occurrence does not change with time, independent of whether earthquakes have or have-not occurred. This is a standard assumption in nearly all earthquake hazard models. The USGS and WGCEP more recently issued a 'time-dependent' model, which captures additional knowledge on the behavior of earthquake faults which results in changes in in the average time to failure of faults over time.

[7] The UCERF3 (http://www.WGCEP.org/UCERF3) rupture forecast model provides authoritative estimates of the magnitude, location, and likelihood of earthquake fault rupture throughout California.

For purposes of damage and loss calculations described in this project, a residential exposure[8] model was constructed for California at the census tract level, starting from the number of housing units within each tract as reported in the 2010 Decennial Census. The number of housing units was then transformed into estimates of the number of structures for each of the 36 building classes identified by Hazus. In addition to the residential exposure, the inventory of nonresidential buildings from the Hazus database was used to create exposure models of commercial and industrial structures in the San Francisco Bay Area at the census tract level. The seismic vulnerability[9] models derived for the different building classes in California were also based on the information provided by Hazus.

The methods developed in this project are intended to help stakeholders develop a sense for how the variability or uncertainties in different components of the OpenQuake earthquake risk model contribute to variability or uncertainty in the risk. It is the intent of this study to encourage the creation of transparent risk-models, where the large uncertainties inherent in seismic risk modeling are not concealed, but rather highlighted and propagated through to the final risk-metrics. The project also demonstrates the value, use, and limitations of a model simplification software tool that is intended to reduce the computational complexity as well as greatly reduce the time required for running the model.

Sensitivity analysis and model simplification tool

GEM's implementation of the latest seismic hazard model for California in OpenQuake is used in this project to conduct a thorough sensitivity analysis to obtain the following information:

- Identify the full distribution of various risk metrics that can be considered feasible from a scientific point of view.
- Quantify the relative contribution of various parts of the model to the overall uncertainty in the seismic risk results.
- Propose criteria to reduce the complexity of the model by eliminating components of the model that do not contribute significantly to the spread in the risk results.

[8] *Exposure* describes the situation of people, infrastructure, housing, production capacities and other tangible assets located in hazard-prone areas (Ref: UNISDR 2017 terminology on disaster risk reduction http://preventionweb.net/go/7822).

[9] *Vulnerability* describes the conditions determined by physical, social, economic and environmental factors or processes which increase the susceptibility of an individual, a community, assets or systems to the impacts of hazards (Ref: UNISDR 2017 terminology on disaster risk reduction http://preventionweb.net/go/508). In particular, a *seismic vulnerability* model describes the susceptibility of buildings and other infrastructure elements to physical damage leading to potential economic and human losses due to earthquake ground shaking.

These components can potentially be excluded from the model without significantly affecting the distributions of the risk metrics.

A software tool has been developed as part of this project that allows stakeholders to reduce the complexity of hazard and risk models based on the criteria identified during the sensitivity analysis. Details about the implementation of this tool are available in Appendix C of the report, along with instructions on how to use it for simplification of models other than those developed in this project.

Key findings

As part of this study, two types of analyses were undertaken: scenario and probabilistic. A scenario analysis addresses the risk associated with a single earthquake, such as a repetition of the 1906 San Francisco Earthquake, and a probabilistic analysis considers all possible events that could occur, each associated with a magnitude, location and probability of occurrence. The latest seismic hazard model for California, based on UCERF3, is probabilistically based. Some of the key findings from these analyses include:

1. 1906 San Francisco scenario

Finding 1.1: The uncertainty in estimated damage or loss for the San Francisco Bay Area is generally dominated by the uncertainty in the propagation of seismic waves from the earthquake fault to the soils beneath the site. Generally speaking, less uncertainty in damage is contributed by the uncertainty in amplification of ground motion at the building site itself. In San Francisco County, however, the uncertainty in soil amplification contributes equally to the uncertainty in the estimate of damage or loss.

2. Annual collapse risk

Finding 2.1: For different classes of buildings, the relative likelihood of collapse is highest for low-rise structures (three stories or fewer), lower for mid-rise structures (between four and eight stories), and lowest for high-rise structures (above eight stories), with the exception of unreinforced masonry where mid-rise structures are more likely to collapse than low-rise ones.

Finding 2.2: Across all building materials, wood-frame structures, which comprise more than 90% of all buildings in the San Francisco Bay Area, are least likely to collapse in comparison to other building materials. In absolute numbers, however, collapses of wood-frame residential structures comprise the majority of the total building collapses.

Finding 2.3: The building classes with the greatest likelihood for collapse include low-code[10] and pre-code[11] versions of low-rise precast concrete frames with concrete shear walls, low-rise concrete frames with unreinforced masonry infill walls, low-rise concrete shear walls, low-rise concrete moment frames and low- and mid-rise unreinforced masonry bearing walls. Pre-code versions of structures in these building classes are highly likely to collapse with significant ground shaking, whereas the high-code versions of the same structures are very unlikely to collapse.

Finding 2.4: For the same building class compared across different cities, the likelihood of collapse is lower in the cities with lower seismic hazard, as expected.

3. Probabilistic seismic risk

Finding 3.1: The average annual losses are highest in the Los Angeles Metropolitan Region and the San Francisco Bay Area, both in absolute dollar values as well as in proportion to the total value of the building stock in each area.

Finding 3.2: For the smaller region of Napa County, the average annual losses in absolute terms are negligible in comparison to bigger regions such as Greater Sacramento and the San Diego Metropolitan Region. However, as a proportion of the total value of buildings in each area, the seismic risk is higher for Napa County in comparison to Greater Sacramento and the San Diego Metropolitan Region.

Finding 3.3: There is a high degree of uncertainty in the estimates of risk metrics commonly used by the (re)insurance industry, such as the average annual loss and the '100-year' or '250-year' loss estimates, as demonstrated for the residential exposure in the San Francisco Bay Area.

4. Sensitivity

Finding 4.1: The recalculation of risk under multiple alternative assumptions in the hazard model resulted in the following conclusions regarding the biggest contributors to uncertainty in the estimate of the average annual loss for the San Francisco Bay Area residential portfolio, in order of importance: 1) the association of earthquake magnitudes with the area and length of the ruptures; 2) the frequency of earthquakes larger than magnitude 5 in the region; and 3) the propagation of seismic waves from the earthquake fault. **Figure 1** shows a tornado diagram that illustrates the relative influence of the

[10] Buildings in older areas of construction, not conforming to modern design standards are classified as 'low-code'.

[11] Buildings constructed before 1940 (i.e., before seismic design codes existed) are classified as 'pre-code'. Buildings that are not designed for earthquake loads are also classified as 'pre-code'.

different components of the model to the average annual loss estimate for the San Francisco Bay Area.

Finding 4.2: The factors affecting the overall variability and uncertainty in the risk estimates depends on the risk metric being estimated. For the San Francisco Bay Area residential portfolio, the top three contributors to the uncertainty in the estimate for the 100-year loss are the assumptions concerning the propagation of seismic waves from the earthquake fault, the association of earthquake magnitudes with the area and length of the ruptures, and the fault slip rates. For the 250-year loss estimate for the same portfolio, the top three contributors to the uncertainty are the assumptions concerning the propagation of seismic waves from the earthquake fault, the fault slip rates, and the frequency of earthquakes larger than magnitude 5 in the region.

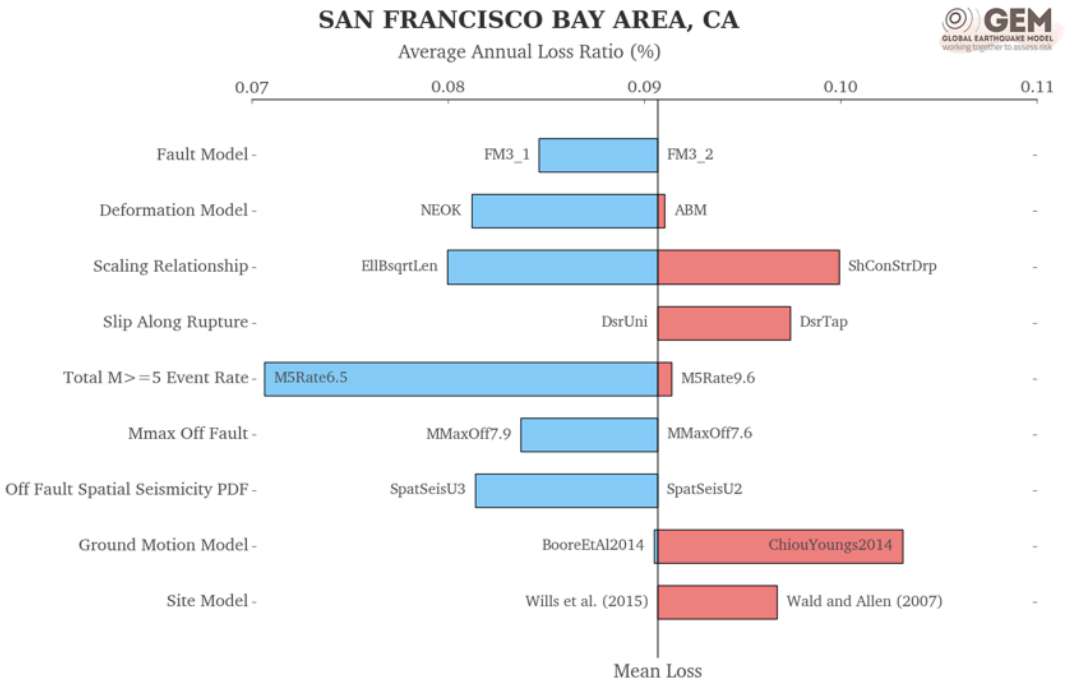


Figure 1. Tornado diagram displaying the sensitivity of the estimated average annual loss ratio (AALR) to the different components comprising the full model for the San Francisco Bay Area. The relative lengths of the bars indicate the relative influence of the different components on the average annual loss estimate.

Finding 4.3: The list of the most influential assumptions also differs for different exposure portfolios. For instance, for the San Diego residential portfolio, the top three contributors to the uncertainty in the estimate for the average annual loss are the assumptions concerning the frequency of earthquakes larger than magnitude 5, association of earthquake magnitudes with the area and length of the ruptures, and the fault slip rates.

5. Model simplification

Finding 5.1: One can reduce the runtime for the model by a factor of more than 100, and still retain its accuracy, by reducing the number of variables in the earthquake hazard calculations. For example, in the calculation of average annual loss for the San Francisco Bay Area, the runtime is reduced by a factor of 200 by considering only uncertainty in fault slip rate, ground shaking intensity, and frequency of earthquakes larger than magnitude 5. The exact combination of variables required in this process is sensitive to the risk metric, exposure and region of interest.

Known limitations

This study specifically examines sources of uncertainty in the earthquake hazard (e.g., earthquake magnitudes and ground shaking) and their relative contribution to the spread in risk estimates. Not addressed are the contributions to the uncertainty in risk from uncertainties in the assumptions made in the exposure and vulnerability models. These two sources of uncertainty were out of scope of the present study and should be considered in future work.

The construction of the exposure model involves assumptions regarding the mapping from building occupancy types to specific construction classes, the assignment of design-eras and height classifications to buildings, and the geographic locations of the buildings within the region of study. Exposure models are arguably the component with the highest uncertainty in most risk models due to the large number of assumptions that are required to develop them. The impact of each of these assumptions can be studied by comparing results obtained using coarse exposure models that involve aggregation assumptions with those obtained by using high-resolution, building-by-building exposure models.

Vulnerability models for building classes typically include large uncertainties due to the assumptions involving the geometric and material properties of buildings, the variability of the building's response to different earthquakes, the variability in the onset of damage due to deformation, and uncertainties in estimating the cost of repair and reconstruction following damage. In addition, there is a possibility that all buildings of a particular class may perform better or worse than the average during an earthquake, and such correlations should be accounted for in a risk model. OpenQuake can already handle these uncertainties and correlations in the vulnerability model; but, their impact on risk results merits further exploration.

The scenario-type analyses in this study focus primarily on the physical damage to the structural components of buildings and the consequent direct economic losses due to such damage. The sensitivity analysis focuses primarily on understanding the sensitivity of loss estimates to alternative input assumptions regarding the earthquake hazard model. Not considered in the analyses are the following:

- damage to building contents and indirect losses due to business interruption;
- losses due to secondary hazards such as landslides, liquefaction, tsunamis, and fire-following earthquake;
- impacts to the infrastructure, lifelines, and utilities; and
- aftershocks, which can cause damage to structures that were previously weakened during a mainshock earthquake.

Such additional sources of losses should be considered in a future study in order to identify all of the potential consequences of a large earthquake.

Recommendations

Recommendation 1: Uncertainties in the variables that make up the hazard model should be reflected in the estimates of risk. Decisions undertaken based on mean estimates alone may thus be based upon severely under- or over-estimated values, relative to a realistic range of uncertainty in the mean loss estimate. The body and full range of the estimates should be available to decision makers.

Recommendation 2: Consideration of all uncertainties through the risk modeling process is often prohibitively expensive and presents the primary hurdle that precludes exploring the full range of risk results. Sensitivity analysis such as the one conducted in this study can aid in simplifying the computational effort, allowing for the approximate calculation of the distribution of risk results in significantly less time. The model simplification tool developed as part of this project can be used to achieve this objective.

Recommendation 3: The findings and methods used in the study may be appropriate to help develop seismic risk models that are transparent and address the full range of uncertainties in the input variables of the risk model. This study has explored the uncertainties in the hazard components in a comprehensive manner. Future work should extend the same level of detail to analyzing the uncertainties in the vulnerability and exposure components.

Recommendation 4: The OpenQuake development of the highly complex and scientifically advanced UCERF3 earthquake hazard model is the first open-source implementation of UCERF3 into a risk modelling software. Given the open and transparent nature of this product, its use to inform the development of policies for risk mitigation, emergency preparedness, and the design of buildings and infrastructure in California is highly recommended.

Recommendation 5: The tools and methods used in this study should be used to understand the seismic risk for exposure portfolios in regions of California outside of

those covered by this study, as well as for the rest of the United States. This could require the use of additional information for estimating the likelihood of earthquakes or the resulting ground shaking intensity, particularly for studies outside of California.

Recommendation 6: Future work should explore the creation of a robust decision-making framework that incorporates uncertainties into the development of programs for reducing risk cost-effectively, thereby improving public safety and community resilience. This framework should include communication of the science of earthquake risk and its uncertainty to non-experts. In particular, improving the interaction between the public, policymakers and the scientific community is necessary for facilitating more effective decision making in earthquake risk management.

1 Introduction

California is the most populous state in the United States, and with a GDP of $2.5 trillion, it is the sixth largest economy in the world. California has a history of damaging earthquakes, and disruptions to the economy due to moderate-to-large earthquakes are likely to have major impacts across the world. According to the Working Group on California Earthquake Probabilities, the likelihood for magnitude 6.7 and larger earthquakes occurring somewhere in California in the next 30 years is near certainty. The likelihood of an earthquake of magnitude 7.5 or higher occurring somewhere in the state over the next 30 years is 48%. Given the high level of exposure and high seismic hazard in California, it is crucial that policies adopted for risk mitigation, emergency preparedness, and the design of buildings and infrastructure should be informed by the best available science on earthquakes. Consequently, there is a need for improved understanding of the seismic risk in the State, including a better characterization of elements that comprise the risk: hazard, exposure, and vulnerability.

There are too few past earthquakes to use history as the only judge of underlying earthquake risk. Earthquake risk models allow us to bridge that gap by combining with history the best available science and engineering knowledge about earthquakes and their effects on society. In so doing, risk models can provide realistic estimates of the likelihood of future earthquakes and the potential societal impacts, such as damage to structures, economic losses, casualties, and business disruption. However, proprietary earthquake risk models are protected intellectual property, containing countless assumptions and methods that are disclosed only as necessary, as such, these models are 'black boxes.' Model users can 'push buttons' (vary the input) to produce results, but the effect of each 'button' remains mostly hidden. The lack of transparency of these models means the uncertainties in the assumptions and model results are not known to users and stakeholders, which diminishes their credibility and hinders their adoption in earthquake risk management.

The Alfred E. Alquist Seismic Safety Commission (SSC) engaged the GEM Foundation to quantify and discuss the impact of various assumptions on earthquake model results for California, and to investigate the treatment of uncertainty within the model. The results of this study will help identify the key factors that influence the risk results in California and improve understanding of the seismic risk in the state. Such improved knowledge can be leveraged by a wide range of stakeholders within the community to: enhance earthquake risk mitigation strategies; develop appropriate emergency preparedness and response plans; inform risk transfer and insurance mechanisms; and better manage risk-sensitive investment portfolios.

The United Nations Office for Disaster Risk Reduction (UNISDR) defines disaster risk[1] as 'the potential loss of life, injury, or destroyed or damaged assets which could occur to a system, society or a community in a specific period of time, determined probabilistically as a function of hazard, exposure, vulnerability and capacity." Seismic hazard, exposure, and physical vulnerability comprise the components of a typical physical seismic risk model, as illustrated in Figure 2.

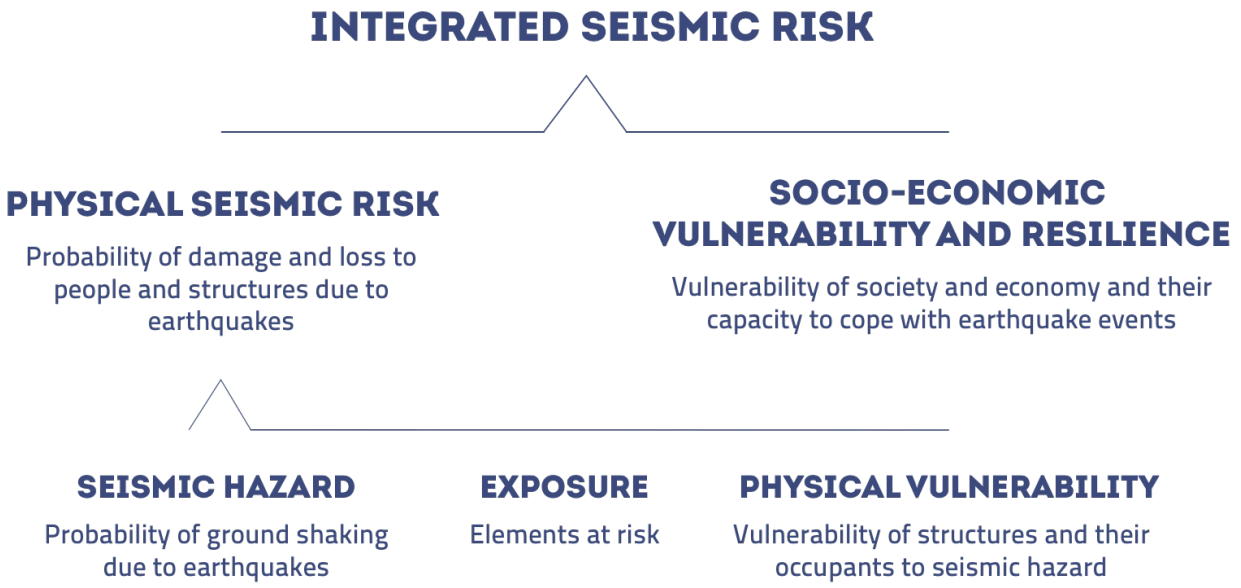


Figure 2. Schematic diagram illustrating the different components of a typical integrated seismic risk model.

The UNISDR definitions of the terms hazard, exposure, and vulnerability are listed below, along with brief descriptions of these terms as they are used in the report.

- The UNISDR defines *hazard*[2] as 'a process, phenomenon or human activity that may cause loss of life, injury or other health impacts, property damage, social and economic disruption or environmental degradation", whereas 'the manifestation of a hazard in a particular place during a particular period of time" is referred to as a hazardous event[3]. The hazard component of an earthquake risk model describes the range of possible events of different magnitudes and their corresponding probabilities of occurrence in a given time-span in the region of interest, and the expected ground shaking associated with these events.
- The term *exposure*[4] is defined by the UNISDR as 'the situation of people, infrastructure, housing, production capacities and other tangible human assets

[1] 2017 UNISDR terminology on disaster risk reduction. Disaster risk: http://preventionweb.net/go/7818
[2] Hazard: http://preventionweb.net/go/488
[3] Hazardous event: http://preventionweb.net/go/51759
[4] 2017 UNISDR terminology on disaster risk reduction. Exposure: http://preventionweb.net/go/7822

located in hazard-prone areas". The exposure component of an earthquake risk model describes the elements that are at risk due to earthquakes; it includes information about the geographic location of population and infrastructure elements in the region of interest, as well as their value. The occupancies of the structures and various attributes describing their construction are also often included in an exposure model.

- *Vulnerability*[5] is defined as 'the conditions determined by physical, social, economic and environmental factors or processes which increase the susceptibility of an individual, a community, assets or systems to the impacts of hazards". The physical vulnerability component of an earthquake risk model relates the level of ground shaking to the expected damage or loss to the structure. It describes the susceptibility of buildings and other infrastructure elements to physical damage leading to potential economic and human losses due to earthquake ground shaking.

2 Project Aims

The goal of the project is to show how important the quantification of uncertainty is in estimating and understanding California's earthquake risk using OpenQuake — GEM Foundation's state-of-the-art open source earthquake hazard and risk assessment software. With OpenQuake's plug-and-play capabilities, expert users can individually select or substitute every model component, data, and assumption. This feature will help model users and decision makers to: 1) 'ask the right questions' when evaluating model results; 2) better interpret risk assessment results and gain trust in model results; and 3) make better risk management decisions.

Specifically, this project aims to:

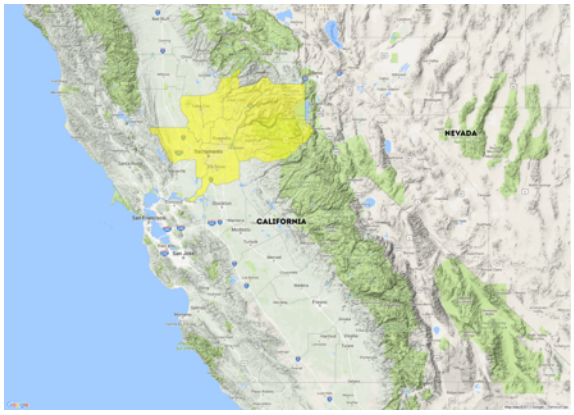
- Establish representative sets of exposure:
  - for the San Francisco Bay Area;
  - for the Southern California region affected by the Shakeout Scenario (See Note 1 below);
  - worldwide (see Note 2 below)
- Choose specific results (risk metrics) to use as a basis of comparison.
- Produce 'baseline' results from OpenQuake, using a 'control' set of assumptions.
- Re-run OpenQuake multiple times, each time varying one assumption or parameter, such as:
  - earthquake probabilities (controlled by assumptions about fault geometries, slip rates, maximum magnitudes);
  - shaking intensity (ground motion model selection);

[5] Vulnerability: http://preventionweb.net/go/508

- damageability of individual buildings (vulnerability curves);
- site conditions; and
- statistical treatment of uncertainty and correlation.

Note 1: The region covered by the ShakeOut Scenario was found to be too large to yield meaningful results if used directly in a sensitivity study, because any geographical influences that become apparent in smaller portfolios are substantially diminished if the entire Shakeout area is considered. Thus, the regions chosen for sensitivity analysis and comparison in this study included the five metropolitan statistical areas of California with the highest population concentrations: the San Francisco Bay Area, the Greater Sacramento Area, the Los Angeles Metropolitan Area, the San Diego Metropolitan Area, and the Inland Empire[6]. The smaller region of Napa County was also included in the analysis following the 2014 Napa earthquake. Maps of these six selected regions are provided below in Figure 3. Although the sensitivity analysis part of this study focused on the above-mentioned metropolitan statistical areas, the calculation of average annual losses was undertaken for the entire state of California.

Note 2: The worldwide database on building exposure compiled by the GEM Foundation is optimized for use in earthquake risk assessment studies and is available for download through the OpenQuake platform at https://platform.openquake.org/exposure.



[6] The **San Francisco Bay Area** includes Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Solano, and Sonoma counties
The **Greater Sacramento Area** includes El Dorado, Nevada, Placer, Sacramento, Sutter, Yolo, and Yuba counties
The **Los Angeles Metropolitan Area** includes Los Angeles and Orange counties
The **San Diego Metropolitan Area** includes San Diego county
The **Inland Empire** includes San Bernardino and Riverside counties

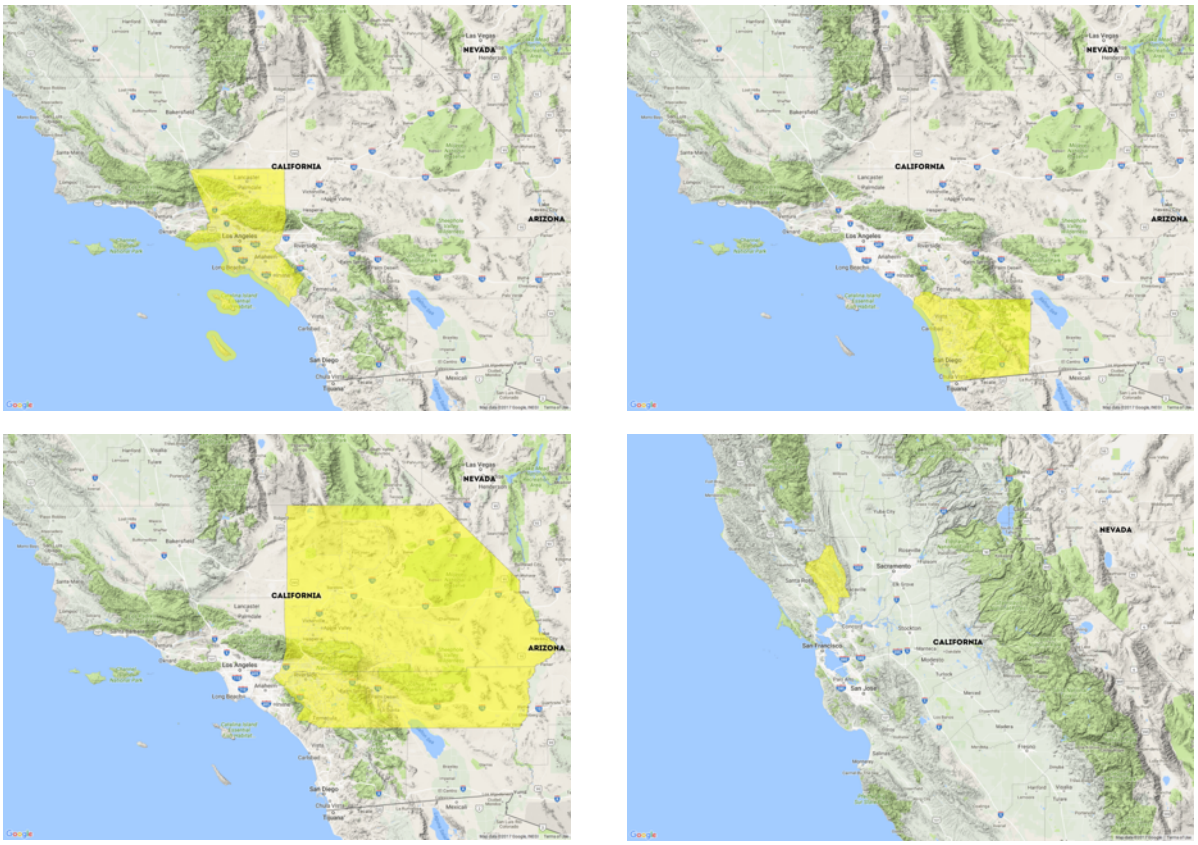


Figure 3. Study regions selected for the sensitivity analysis: the San Francisco Bay Area (top left), the Greater Sacramento Area (top right), the Los Angeles Metropolitan Area (mid left), the San Diego Metropolitan Area (mid right), the Inland Empire (bottom left), and Napa County (bottom right)

Beyond the aims stated at the outset of the project as listed above, several additional objectives were achieved during the course of the project, including the following:

- Implement within OpenQuake the latest seismic hazard model for California based on the recently published Uniform California Earthquake Rupture Forecast version 3 (UCERF3), produced by the U.S. Geological Survey and the Working Group on California Earthquake Probabilities.
- Calculate the average annual loss[7] estimates for all 8,057 census tracts in California, using the seismic hazard model based on UCERF3.
- Establish the range (distribution) of scientifically viable results for the chosen risk metrics by accounting for the various uncertainties in the hazard model.
- Identify the components of the hazard model contributing most to the overall uncertainty in the risk metrics for the different exposure portfolios.

[7] The average annual loss (AAL) estimate is the expected loss per year, averaged over a large number of years.

- Implement a model simplification ('logic-tree trimming') software tool to reduce the number of computer runs and greatly speed up the time required for running the risk model for California.

3 How to Use the Report

In order to fully benefit from the results of this project, this report should be used in conjunction with the SSC's related projects. This will ensure that the SSC's recommendations to the Governor and Legislature will be holistic, actionable, and measurable policies that will help the regional economy prepare, recover, and become resilient to earthquakes.

The transparency of the earthquake risk model developed in this project allows users to inspect all assumptions and input variables, and test alternative implementations.

0 describes how to run the seismic hazard and risk models for California using OpenQuake, and describes the datasets required to run the various other calculations described in this report.

The datasets and models developed as part of this study are available at: https://gitlab.openquake.org/risk/usa/tree/master/models

This report is intended to serve as a guidance document summarizing factors that model users should consider when examining the results of their analyses. The following sections of the report thus provide:

- a description of the methodology and data sets compiled as a part of this project;
- details of each parameter and assumption that was varied in the sensitivity analysis, and its measured effect on the results given a risk metric; and
- implications for interpreting model results from complex risk models such as the one based on the most recent seismic hazard model for California contained within the National Seismic Hazard Model (NSHM) for the United States.

4 Datasets and Software Implementation

4.1 Seismic hazard

The seismic risk calculations performed within this project are based upon the time-independent[8] version of the recently published Uniform California Earthquake Rupture

[8] 'Time-independent' means that the probability of earthquake occurrence does not change with time, independent of whether earthquakes have or have-not occurred. This is a standard assumption in nearly all earthquake hazard models. The USGS and WGCEP more recently issued a 'time-dependent' model, which captures additional knowledge on the behavior of earthquake faults which results in changes in in the average time to failure of faults over time.

Forecast version 3 (UCERF3[9]), produced by the U.S. Geological Survey and the Working Group on California Earthquake Probabilities (WGCEP). The state-of-the-art UCERF3 rupture forecast provides the most recent evaluation of the probabilities of occurrence of potentially damaging earthquakes in California over a time-span. To fully represent the uncertainties in the forecast arising from incomplete knowledge of the earthquake process and alternative modelling choices (e.g. different models describing the current tectonic deformation in California), UCERF3 comprises the integration of all possible combinations of the values of different variables in the occurrence model (or rupture forecasts). The time-independent version of UCERF3 comprises: (1) two alternative *fault models*, which describe the geometry of the active faults in California; (2) four alternative *deformation models*, which describe slip-rates on each fault-segment; (3) and various combinations of maximum magnitude, spatial distribution of earthquakes and rate of large-magnitude earthquakes, which results in 180 alternative *earthquake rate models*. **The combination of the two *fault models*, four *deformation models*, and 180 *earthquake rate models* results in (2×4×180 =) 1,440 alternative realizations of the earthquake rupture forecast model.** Further details about the UCERF3 rupture forecast and its implementation within OpenQuake are provided in Appendix A.1.

Five different ground motion models (GMMs) are employed in the 2014 update to the National Seismic Hazard Model (NSHM) for California. This suite of GMMs consists of the five models developed within the Next Generation Attenuation (NGA) West 2 project (Bozorgnia et al., 2014). The five models selected for application in the western US are listed in Table 5, along with their respective weights in the logic tree. Appendix A.2 provides further details about the derivation of these GMMs and their implementation within OpenQuake. **Evaluating all 1,440 possible combinations of the rupture model together with the five alternative ground motion models results in a total of (1,440×5 =) 7,200 model combinations.**

The use of alternative Vs_{30} models (time-averaged shear-wave velocity in the top 30 m soil layer) to approximate site-conditions allows for modeling the epistemic uncertainty[10] in the soil conditions. Two different site condition models were used for the calculations in this project: (1) the Vs_{30} model from Wald and Allen (2007) based on topographic slope; and (2) the Wills et al. (2015) Vs_{30} model based on surficial geology and topology. Seismic site condition maps based on these two models are shown below in Figure 28, and more details about the two models are provided in Appendix A.3.

[9] The UCERF3 (http://www.WGCEP.org/UCERF3) rupture forecast model provides authoritative estimates of the magnitude, location, and likelihood of earthquake fault rupture throughout California.

[10] *Epistemic uncertainty* refers to the uncertainty associated with lack of knowledge or insufficient understanding of the underlying processes. Epistemic uncertainty is typically accounted for in probabilistic seismic hazard and risk analysis by using alternative models to represent the underlying process.

4.2 Exposure

For purposes of damage and loss calculations described in this project, a residential exposure[11] model was constructed for the entire state of California at the census tract level, starting from the number of housing units within each tract as reported in the 2010 Decennial Census. The residential exposure model described classifies buildings following the Hazus methodology into seven general typologies based on the primary material of construction — wood, steel, reinforced concrete, precast concrete, reinforced masonry, unreinforced masonry, and manufactured homes. Each of these general typologies is further classified based on the building height and primary lateral load resisting system as described in Table 6 to obtain the number of housing units in 36 specific building classes. The number of *housing units* was then transformed into estimates of the number of *structures* for each of the 36 building classes identified by Hazus. Finally, based on the approximate age of the building and the seismic design level in the region, the buildings are further categorized as pre-code, low-code, moderate-code, or high-code, yielding a total of 128 building classes that take into account changes in the building code over time.

In addition to the residential exposure, the inventory of non-residential buildings from the Hazus database was used to create exposure models of commercial and industrial structures in the San Francisco Bay Area at the census tract level. Figure 4 shows the estimated replacement costs for the structural and non-structural components of the residential buildings per census tract in the San Francisco Bay Area. Similar maps for the commercial and industrial exposure models are provided in Appendix A.4. Further details about the process of creating the residential exposure model are also provided in Appendix A.4, as are descriptions of the 36 building classes used in the model.

[11] *Exposure* describes the situation of people, infrastructure, housing, production capacities and other tangible assets located in hazard-prone areas (Ref: UNISDR 2017 terminology on disaster risk reduction http://preventionweb.net/go/7822).

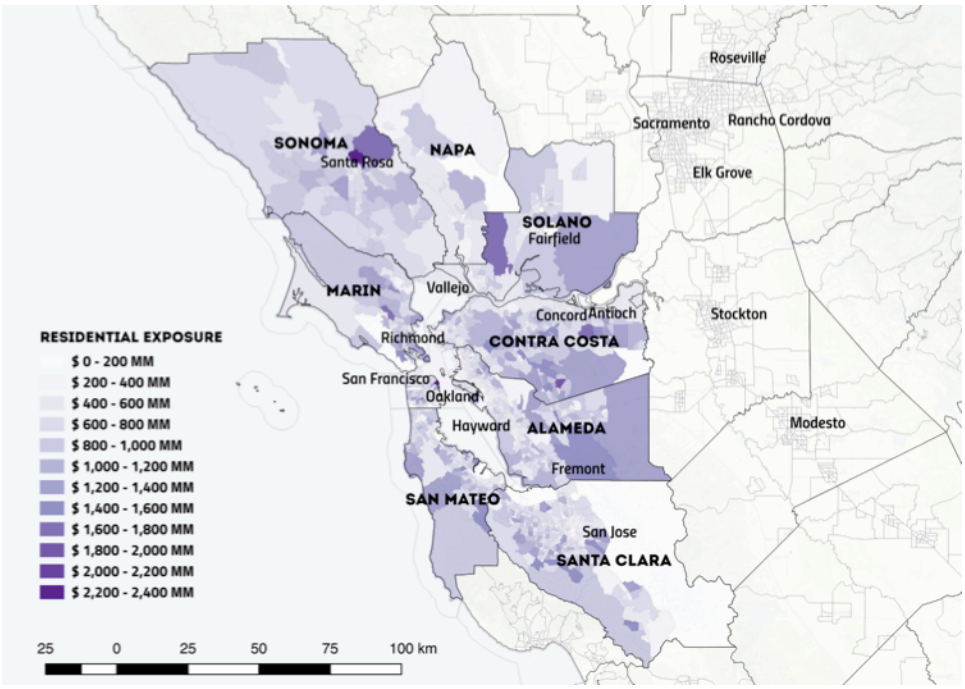


Figure 4. Residential exposure for the San Francisco Bay Area. The map displays the estimated total structural and non-structural replacement cost for residential buildings in each census tract.

Note 1: The study covers structural and non-structural damage to buildings. However, it does not cover some kinds of infrastructure and site examples, such as power plants, gas pipelines, water and sewer pipelines, substations and switchyards, ports and dams, and roadways and bridges. The study focuses on the seismic risk to portfolios of residential, commercial, and industrial buildings, but the tools and methodologies developed are easily extensible to other kinds of infrastructure.

Note 2: The residential, commercial, and industrial exposure datasets compiled using information from the Decennial Census and Hazus inventories includes estimates of the building contents. However, damage to building contents is not estimated in the scenario or probabilistic analyses in this study.

4.3 Seismic vulnerability

The seismic vulnerability[12] models derived for the different building classes in California were also based on the information provided by Hazus. Structural parameters affecting

[12] *Vulnerability* describes the conditions determined by physical, social, economic and environmental factors or processes which increase the susceptibility of an individual, a community, assets or systems to the impacts of hazards (Ref: UNISDR 2017 terminology on disaster risk reduction http://preventionweb.net/go/508). In particular, a *seismic vulnerability* model describes the susceptibility of buildings and other infrastructure elements to physical damage leading to potential economic and human losses due to earthquake ground shaking.

structural capacity and response during earthquakes are provided by Hazus for each of the 128 building classes identified in the exposure models. These structural parameters are used to build numerical models representing the structures from each building class. The numerical models are subjected to simulated earthquake ground shaking using a catalog of ground motion records obtained from large magnitude earthquakes in California, Japan, Turkey, Italy, and Taiwan. All the records were from earthquakes on strike-slip and thrust faults, consistent with the seismicity in California. The simulated structural displacement and deformation of buildings is then used to derive the fragility functions for each building class, using assumptions about damage thresholds from Hazus. The fragility functions are then combined with damage-to-loss ratios suggested by Hazus to derive vulnerability functions for all of the building classes. Additional details about the methods employed for deriving the fragility and vulnerability functions are provided in Appendix A.5.

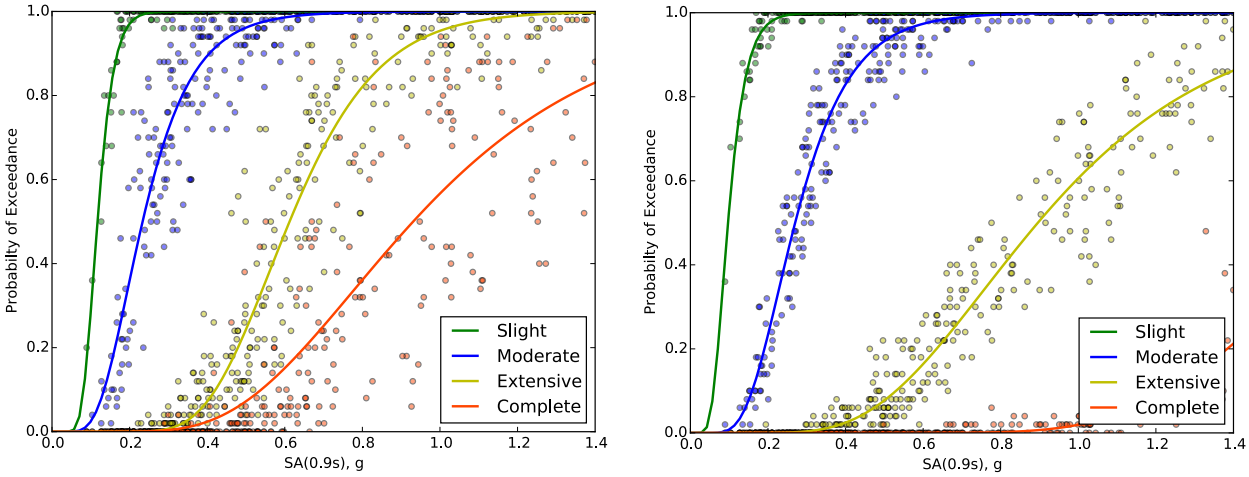


Figure 5. Fragility models derived for light-frame residential wood structures. Left: Low-code wood-frame structures; Right: High-code wood-frame structures.

4.4 UCERF3 calculators for the OpenQuake-engine

The large number of alternative model choices and the associated computational expense makes risk analysis based on the full UCERF3 model suite particularly challenging. The software development team at the GEM Foundation, in close collaboration with the hazard scientists and risk engineers, has developed new hazard and risk calculators for the OpenQuake-engine, tailored specifically to run the full suite of models based on UCERF3. Appendix A.6 provides an overview of the architecture of the hazard and risk modules of the OpenQuake-engine, and Appendix A.7 describes the details of the implementation of the new calculators tailored for UCERF3-specific applications in the OpenQuake-engine. With a view toward examining the contribution of the different components of UCERF3 to the overall uncertainty in the risk results, these newly

implemented UCERF3 calculators of the OpenQuake–engine are employed to fully evaluate the probabilistic hazard and risk for all 7,200 model combinations for seven different statistical regions in California.

4.5 Logic-tree trimming tool

The sensitivity analysis performed in this study is also the first open analysis of the risk to different building portfolios performed using UCERF3. The model simplification ('logic-tree trimming') tool developed by Porter et al (2012)[13] for the hazard model based on UCERF2 has been extended in this study to handle the significantly more complex model based on UCERF3. Whereas the tool introduced by Porter et al. considered only the average annual loss for model simplification, the tool developed as part of this study is more flexible as it allows stakeholders the choice of other risk metrics to be used for the simplification. Specifically, the loss estimates for any return-period, such as the '100-year' or '250-year' loss estimates[14], can be used as the basis for simplifying the model with this tool. This logic-tree trimming tool allows stakeholders to reduce the complexity of hazard and risk models based on the criteria identified during the sensitivity analysis.

5 Key Findings

As part of this study, two types of analyses were undertaken: scenario and probabilistic. A scenario analysis addresses the risk associated with a single earthquake, such as a repetition of the 1906 San Francisco Earthquake, and a probabilistic analysis considers all possible events that could occur, each associated with a magnitude, location and probability of occurrence. The latest seismic hazard model for California, based on UCERF3, is probabilistically based. This section summarizes several of the key findings from these analyses.

1. Section 5.1 presents the influence of various modeling assumptions on damage, loss, and casualty estimates involving a single earthquake event (a 'scenario'), a repeat of the 1906 M7.9 San Francisco Earthquake.
2. Section 5.2 provides results of a comparative collapse risk study for 128 building types across six different cities in California: Oakland, San Jose, Los Angeles, San Francisco, San Diego, and Sacramento.
3. Section 5.3 discusses average annual losses and loss exceedance curves derived from the probabilistic risk model for all of California.
4. Section 5.4 presents issues concerning sensitivity analysis of all of the model variables for a subset of the probabilistic model, specifically the San Francisco Bay Area.

[13] Porter, K. A., Field, E. H., & Milner, K. R. (2012). Trimming the UCERF2 Hazard Logic Tree. *Seismological Research Letters*, 83(5), 815–828

[14] The 100-year return period loss is the loss value that is exceeded once every 100 years, averaged over a large number of years. Similarly, the 250-year return period loss is the loss value that is exceeded once every 250 years, averaged over a large number of years.

5. Section 5.5 discusses techniques for trimming complex logic-trees, which can significantly reduce the computational resources required for this complex model.

When combined, the results provide a comprehensive description of the California earthquake risk model developed for the study and provide an approach for implementing the model in a cost-effective and efficient manner.

5.1 Scenario risk results

Scenario earthquake simulations involve estimating the level of ground shaking and subsequent damage or loss due to a single earthquake. An earthquake scenario is defined as an earthquake where the magnitude, hypocenter and fault geometry are specified. These parameters or factors define the earthquake rupture. However, the estimation of the ground motion as well as the damage and loss may involve uncertainties. Hence, even though the scenario calculators estimate risk due to a single seismic event (or earthquake rupture), the results provided by the OpenQuake-engine are probabilistic in the sense that the results still capture uncertainties in the estimation of the impact of that event.

The use of alternative models for simulating the ground shaking or to account for local site effects, for instance, can lead to differing estimates of the damage and losses due to a scenario earthquake. In the following sub-sections, we examine the impact of such assumptions on the estimates of the number of: 1) building collapses, 2) the number of casualties, and 3) direct economic losses.

The earthquake scenario is a simulation of the 1906 magnitude 7.9M_w rupture on the San Andreas Fault in northern California (aka the Great San Francisco Earthquake). This scenario involves a rupture of the entire northern California segment of the San Andreas fault. Per the Working Group on California Earthquake Probabilities, the probability of such a scenario earthquake occurring in the next 30 years is estimated to be 4.7 percent.

The median ground shaking intensity for this event as estimated by OpenQuake is shown in the map in Figure 6 below. The PGA intensities predicted by OpenQuake can be compared with the MMI intensities predicted by the WGCEP for the same scenario in 2002, which are shown in the map in Figure 7.

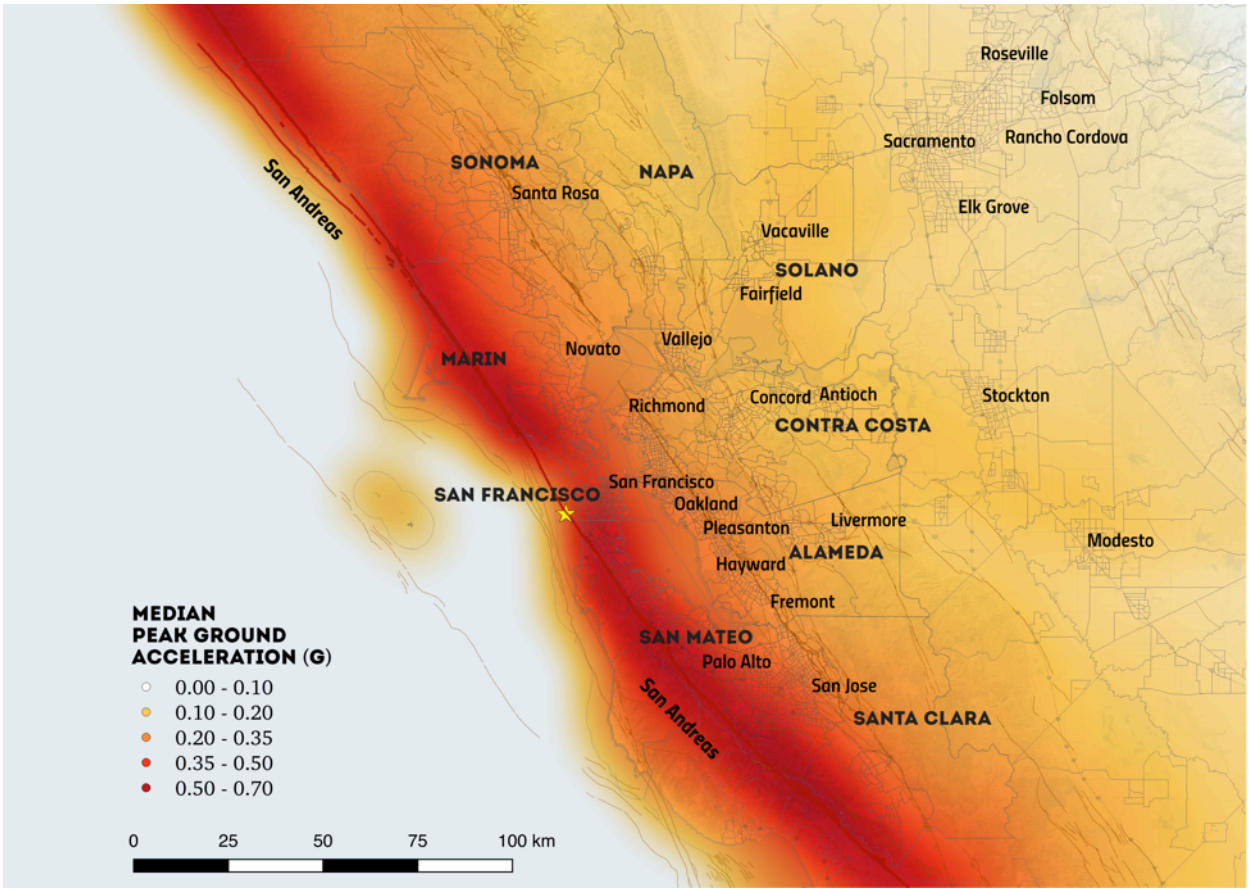
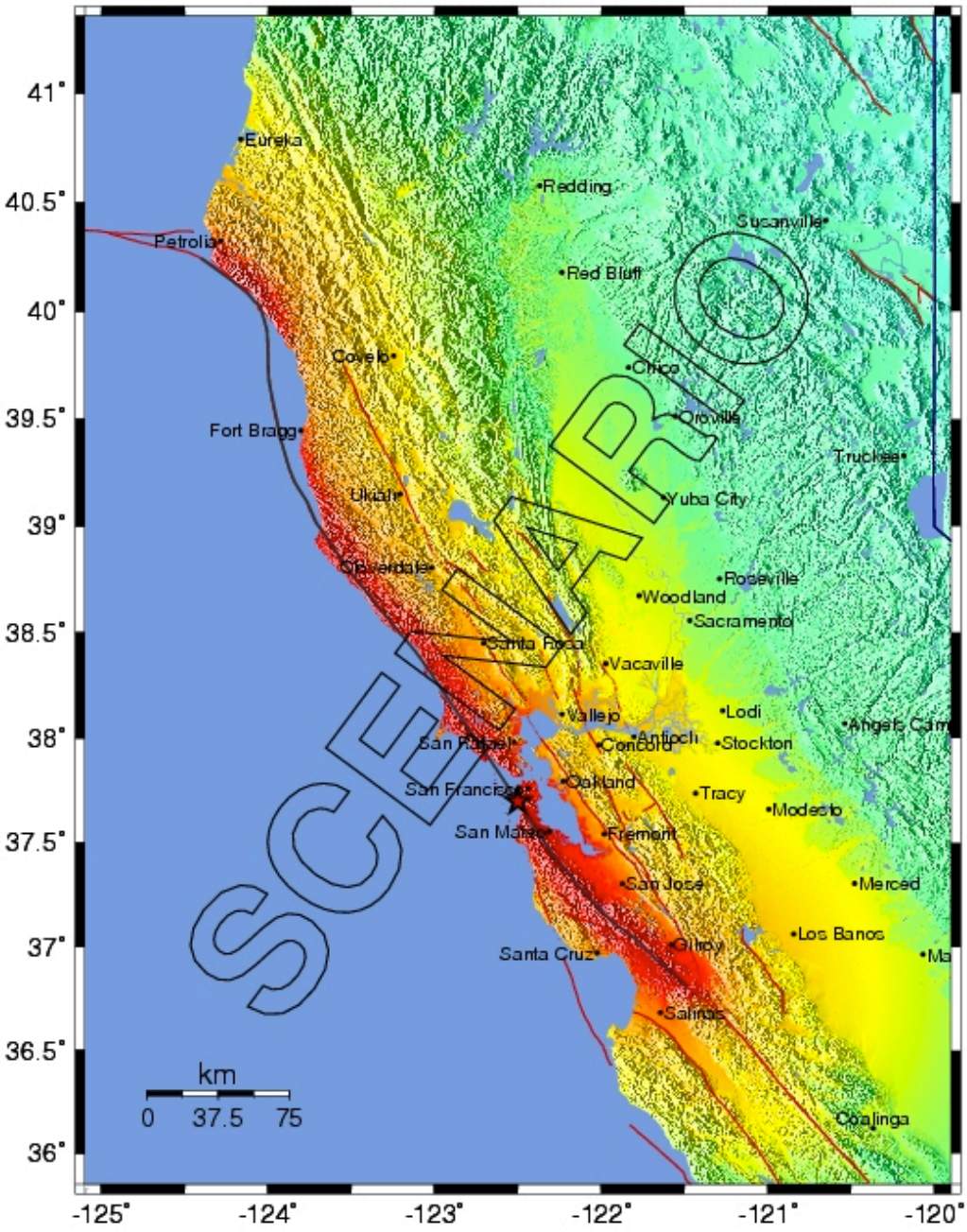


Figure 6. *Median* ground shaking intensity for a simulation of the 1906 magnitude 7.9 San Francisco earthquake, predicted using OpenQuake



PERCEIVED SHAKING	Not felt	Weak	Light	Moderate	Strong	Very strong	Severe	Violent	Extreme
POTENTIAL DAMAGE	none	none	none	Very light	Light	Moderate	Moderate/Heavy	Heavy	Very Heavy
PEAK ACC.(%g)	*<.17*	*.17-1.4*	*1.4-3.9*	*3.9-9.2*	*9.2-18*	18-34	34-65	65-124	>124
PEAK VEL.(cm/s)	<0.1	0.1-1.1	1.1-3.4	3.4-8.1	*8.1-16*	*16-31*	*31-60*	*60-116*	*>116*
INSTRUMENTAL INTENSITY	I	II-III	IV	V	VI	VII	VIII	IX	X+

Figure 7. Ground shaking intensity for the 1906 magnitude 7.9 San Francisco earthquake, predicted by the WGCEP (2002). Source: https://earthquake.usgs.gov/earthquakes/events/1906calif/shakemap/img/1906_Scenario_intensity.jpg

Note: The 1906 earthquake affected regions of California ranging from Eureka on the North Coast to Gilroy in Santa Clara County and Salinas in Monterey County. The San Francisco Bay Area exposure described in Section 4.2 is used for the damage and risk assessment for this scenario. Although the 1906 earthquake remains one of the most damaging earthquakes in US history, other scenarios would be more appropriate to explore the damage and loss estimates for other metropolitan regions in California. For instance, the magnitude 7.8M_w Great Southern California ShakeOut scenario[15] developed by Jones and others in 2008 would be more appropriate for examining the effects of a major earthquake in the Los Angeles and San Diego Metropolitan Areas of Southern California. Similar events are included in the catalog of earthquakes (essentially a set of rupture forecasts) used in the probabilistic analyses of all metropolitan regions in California.

5.1.1 Number of building collapses

The mean estimate of the structural damage distribution for the possible repetition of the 1906 M7.9 San Francisco earthquake rupture was computed using:

- Two different site condition models described in 0 and illustrated in Figure 28.
- Three different ground motion models: Boore and Atkinson (2008), Chiou and Youngs (2008), and Campbell and Bozorgnia (2008)

The results of this analysis can be used to investigate the effect of different model assumptions on the estimated damage distribution.

Table 1 below shows the estimated number of residential buildings in different damage states for the different modeling assumptions. We observe that the choice of ground motion conditions has a significant impact on damage estimates. Looking at mean estimates of the number of buildings that will be completely damaged, based on the Wald and Allen (2007) site conditions model, we see that the high estimate of 255,575 buildings is over 25% larger than the low estimate of 205,967 buildings. On the other hand, comparing the mean damage estimates coming from the two different site models based on the same ground motion model, we note that the difference in the two estimates is less than 10% for all cases.

[15] Jones, Lucile M., Bernknopf, Richard, Cox, Dale, Goltz, James, Hudnut, Kenneth, Mileti, Dennis, Perry, Suzanne, Ponti, Daniel, Porter, Keith, Reichle, Michael, Seligson, Hope, Shoaf, Kimberley, Treiman, Jerry, and Wein, Anne, 2008, The ShakeOut Scenario: U.S. Geological Survey Open- File Report 2008-1150 and California Geological Survey Preliminary Report 25 [http://pubs.usgs.gov/of/2008/1150/]

Table 1. Damage distribution for the San Francisco Bay Area considering different site and ground motion models

Site Model	GMM	PGA (g)	Damage State				
			No damage	Slight	Moderate	Extensive	Complete
Wald and Allen (2007)	BA2008	1.72	172,318	375,904	673,766	216,691	205,967
	CB2008	2.02	130,207	334,046	680,246	244,572	255,575
	CY2008	1.95	166,429	351,696	653,026	229,781	243,714
Wills et al. (2015)	BA2008	1.58	184,928	395,762	672,055	203,247	188,654
	CB2008	1.89	137,901	352,276	684,597	232,505	237,368
	CY2008	1.77	174,716	369,188	656,387	218,717	225,639

BA2008: Boore and Atkinson (2008); CY2008: Chiou and Youngs (2008); CB2008: Campbell and Bozorgnia (2008)

Although buildings in the complete damage state would require eventual demolition and reconstruction, it should be noted that only a portion of such buildings would be considered as collapsed from a structural perspective. The estimated numbers of collapsed buildings for the same earthquake are listed below in Table 2, and these estimates range from 5,660 to 7,667 collapsed residential structures, with an average estimate across all six models of 6,785 collapsed residential structures.

Table 2. Estimated number of residential building collapses in the San Francisco Bay Area

Site Model	GMM	No. of building collapses
Wald and Allen (2007)	BA2008	6,179
	CB2008	7,667
	CY2008	7,311
Wills et al. (2015)	BA2008	5,660
	CB2008	7,121
	CY2008	6,769

5.1.2 Number of occupant casualties

In addition to structural damage, OpenQuake-engine is used to obtain casualty estimates for the scenario. Such estimates of the approximate number of injuries and fatalities can be very useful to regional emergency responders and medical authorities. As with the estimates of the numbers of building collapses, the estimates for the casualties are also computed for the two different site models and using three different ground motion models. These estimates are listed below in Table 3. Considering the average estimate across all six models, there are expected to be around 1,600 fatalities for this scenario earthquake. There will also be about 1,000 people with very severe injuries in need of

advanced medical care to survive, and approximately 17,500 people are estimated to sustain injuries requiring emergency room care.

Table 3. Estimated number of casualties in the San Francisco Bay Area for a repeat of the 1906 magnitude 7.9 earthquake[16]

Site Model	GMM	PGA (g)	Casualty Severity		
			Emergency Care	Severe Injuries	Fatalities
Wald and Allen (2007)	BA2008	1.72	15,990	908	1,449
	CB2008	2.02	19,530	1,126	1,798
	CY2008	1.95	18,615	1,074	1,714
Wills et al. (2015)	BA2008	1.58	14,739	832	1,328
	CB2008	1.89	18,232	1,046	1,670
	CY2008	1.77	17,328	994	1,587

5.1.3 Scenario loss maps and statistics

The structural losses for the scenario, aggregated within each county, are listed below in Table 4. The mean loss maps computed using the Wald and Allen (2007) site model are shown below in Figure 8 for the Boore and Atkinson (2008), Campbell and Bozorgnia (2008), and Chiou and Youngs (2008) GMMs. Note that only direct structural losses to residential buildings were considered in this analysis.

Similar to what was observed in the case of the damage estimates, the choice of the ground motion model has a larger impact on the loss estimates compared to the choice of the site conditions model.

For example, in Marin County, based on the Wald and Allen (2007) site model, we see that the high estimate of $1.18 billion in residential structural losses is 21% larger than the low estimate of $978 million. On the other hand, comparing the loss estimates coming from the two different site conditions models based on the Boore and Atkinson (2008) GMM, we note that the difference in the two estimates is only 7%. In Contra Costa County, the two site conditions models produce virtually identical loss estimates for all three GMMs.

The effect of the site conditions model on the county-wide loss estimates is most apparent in San Francisco county, where the geology based Wills et al. (2015) site model predicts higher Vs_{30} values in most parts of the county compared to the topographic

[16] Note: The estimates of casualties are obtained with the assumption that the earthquake occurs at night, when almost the entire population is likely to be indoors, in wood-frame residential units. These estimates are likely to be much higher if the same earthquake occurs during working hours, when most of the population is likely to be in commercial structures, several of which have a higher vulnerability to earthquakes in comparison to the light wood-frame residential units.

slope-based Wald and Allen (2007) site model. With the Campbell and Bozorgnia (2008) GMM, the loss estimate from the Wald and Allen (2007) site model for San Francisco county is $4.67 billion and that from the Wills et al. (2015) is $3.93 billion, indicating a difference of 19% between the two estimates. For the same county, the difference between the high and low estimates from the three GMMs using the Wald and Allen (2007) site model is 22%.

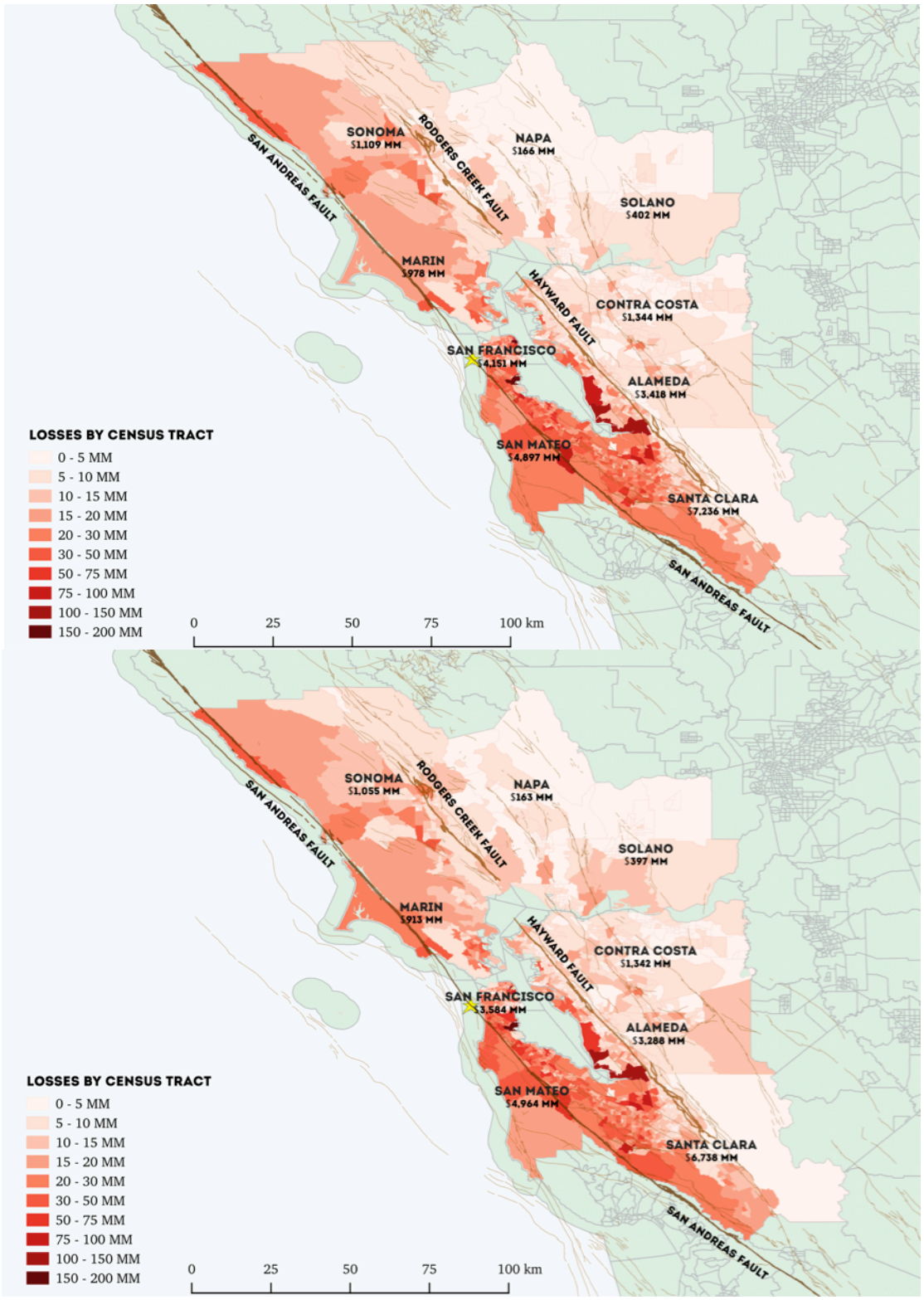


Figure 8. Mean loss maps for the M7.9 San Francisco scenario computed using the Boore and Atkinson (2008) GMM. Top: Using the Wald and Allen (2007) Vs_{30} model; Bottom: Using the Wills et al. (2015) Vs_{30} model.

Table 4. Mean aggregate losses by county for the San Francisco Bay Area considering different site and ground motion models

County	Wald and Allen (2007)			Wills et al. (2015)		
	BA08 ($ MM)	CB08 ($ MM)	CY08 ($ MM)	BA08 ($ MM)	CY08 ($ MM)	CB08 ($ MM)
Alameda	3,418	3,511	3,319	3,288	3,404	3,222
Contra Costa	1,344	1,364	1,244	1,342	1,365	1,243
Marin	978	1,089	1,182	913	1,025	1,099
Napa	166	166	150	163	166	150
San Francisco	4,151	4,670	5,060	3,584	3,929	4,344
San Mateo	4,897	5,477	5,671	4,964	5,605	5,692
Santa Clara	7,236	7,701	7,938	6,738	7,353	7,623
Solano	402	404	352	397	402	346
Sonoma	1,109	1,147	1,130	1,055	1,111	1,095

5.2 Annual collapse risk

The residential exposure model built as part of this study comprises a total of 128 building classes, as described earlier in Section 4.2, and in more detail in Appendix A.4. We conducted a comparative collapse risk study for each of these 128 building classes represented in the residential exposure model for seven different cities in California. The six cities selected for this analysis, in descending order of the seismic hazard, were Oakland, San Jose, Los Angeles, San Francisco, Napa, San Diego, and Sacramento. The annual collapse risk for a particular building class is obtained by the integration of the collapse fragility for the building class with the seismic hazard curve at the location of the building, as described in Eads et al. (2013). In this study, we are interested in comparing the relative collapse risk across different cities, across different building materials, across different building classes using the same construction material, and across different seismic design eras. We do not, however, consider the effect of variation in site conditions in this analysis — we assume similar site conditions for the different locations considered. The annual collapse probabilities calculated for these building classes are shown below in Figure 9–Figure 12.

In general, we observe that within the same general typology, collapse probabilities are highest for the low-rise structures, lower for the mid-rise structures, and lowest for the high-rise structures, with the exception of unreinforced masonry for which the mid-rise structures have a higher collapse probability. Across all building materials, wooden

structures, which comprise more than 90% of all buildings in the San Francisco Bay Area, have the lowest collapse probabilities. The building classes with the highest collapse probabilities include the low-code and pre-code versions of low-rise precast concrete frames with concrete shear walls, low-rise concrete frames with unreinforced masonry infill walls, low-rise concrete shear walls, low-rise concrete moment frames and low- and mid-rise unreinforced masonry bearing walls.

For the same building typology compared across different cities, the annual collapse probability is lower in the cities with lower seismic hazard, as would be expected. The impact of improving seismic design standards is also observable in these charts — for the same structural typology, pre-code versions of the structure exhibit high annual collapse probabilities, whereas the annual collapse probabilities of the high-code versions of the same structures are negligibly low in comparison.

Such studies for appraising the comparative collapse risk of different building types across different cities – and built in different eras according to different seismic design codes - can be quite useful for the development of policies for region-wide risk mitigation strategies for existing structures, or to assess the most adequate seismic design for the region of interest.

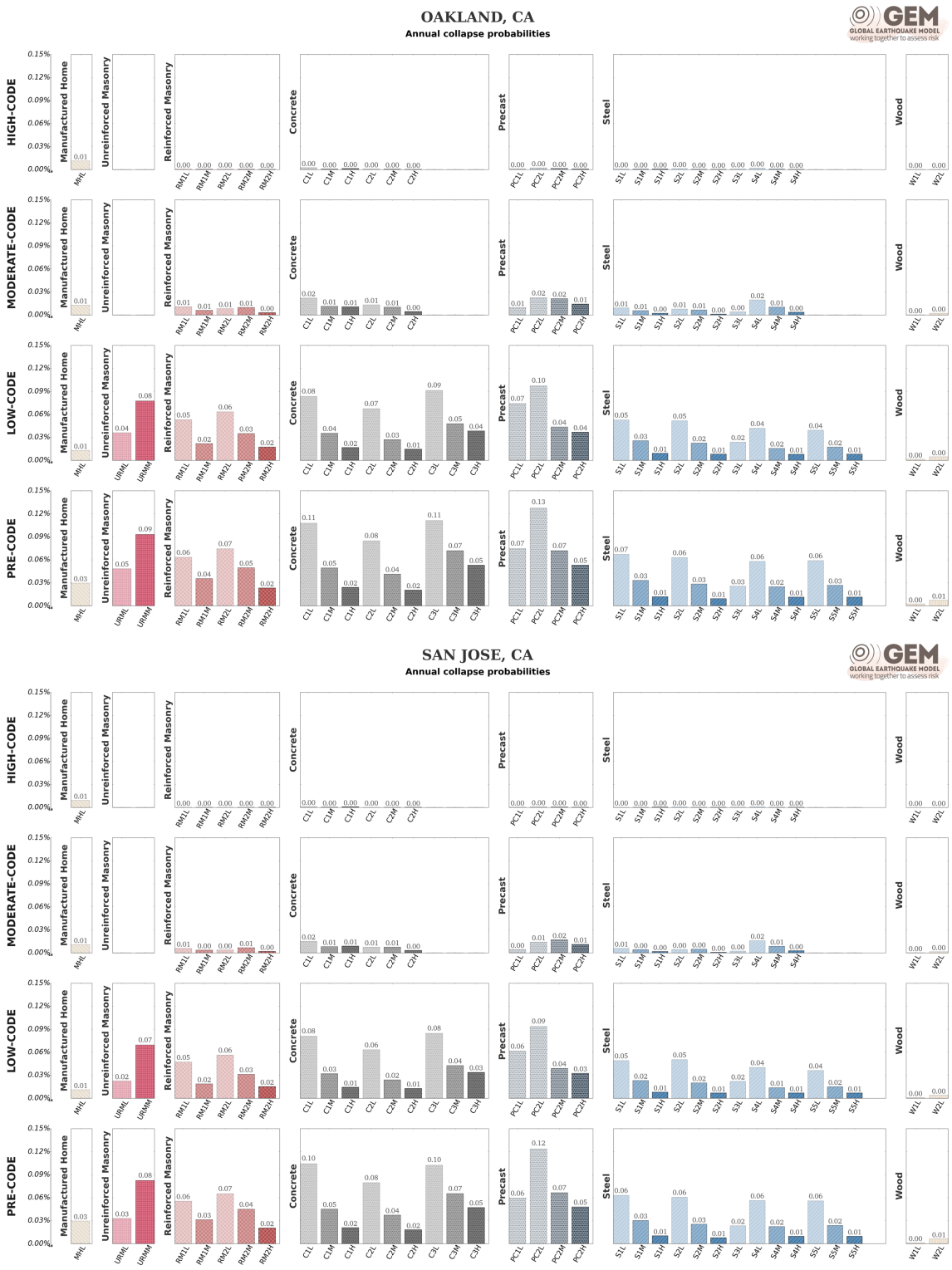


Figure 9. Annual collapse probabilities for 128 building classes in the cities of Oakland and San Jose, CA.

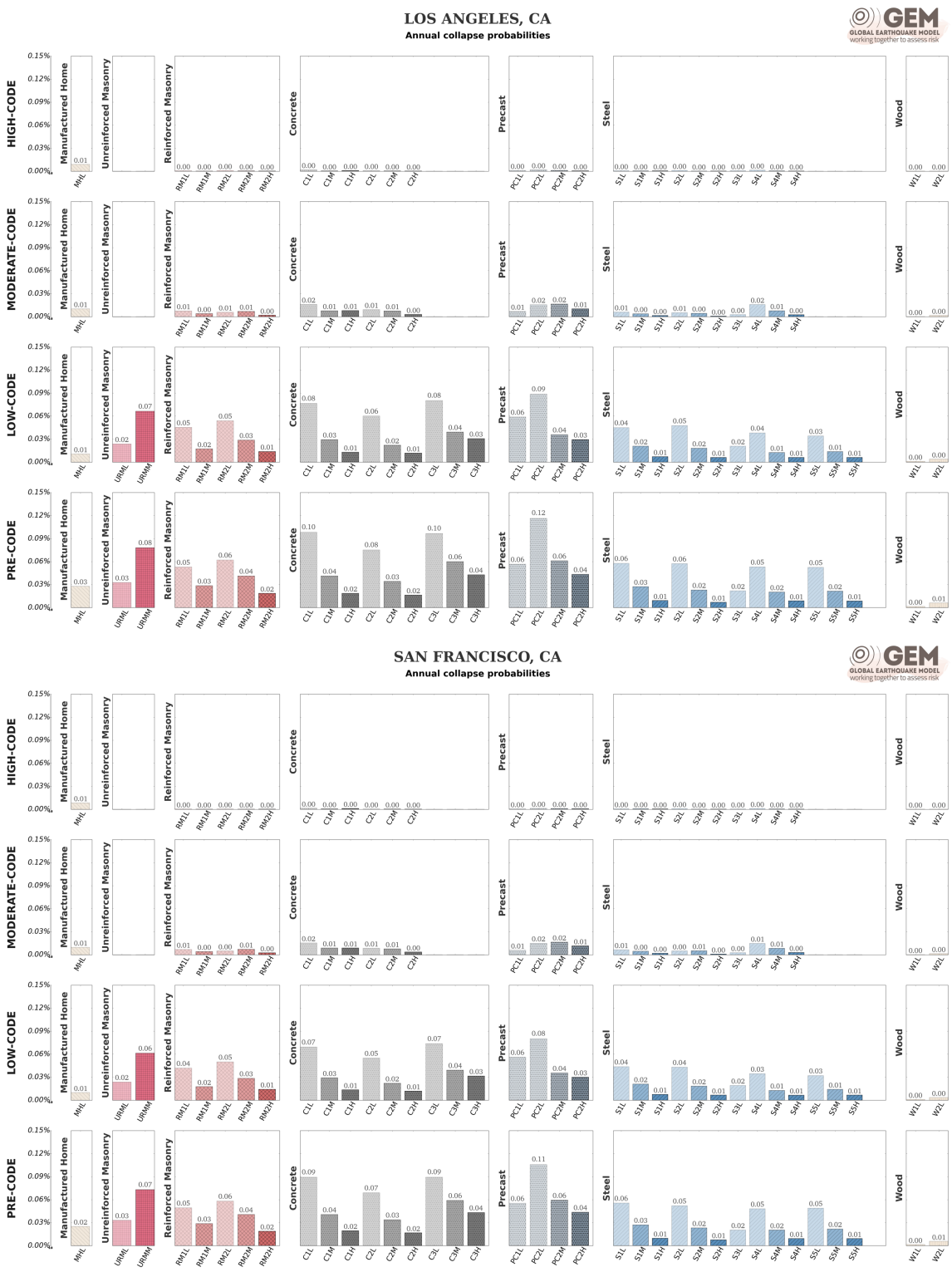


Figure 10. Annual collapse probabilities for 128 building classes in the cities of Los Angeles and San Francisco, CA.

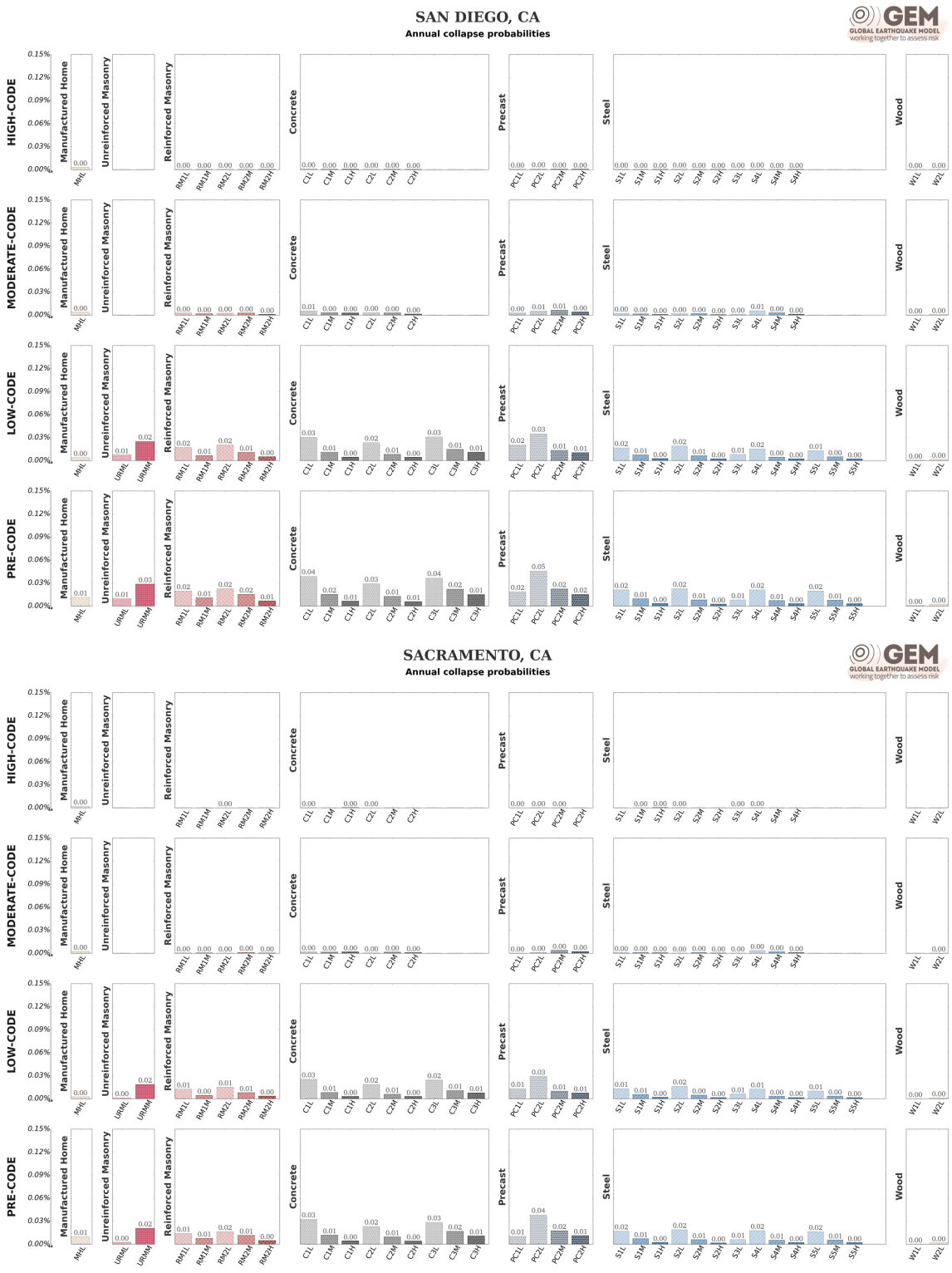


Figure 11. Annual collapse probabilities for 128 building classes in the cities of San Diego and Sacramento, CA.

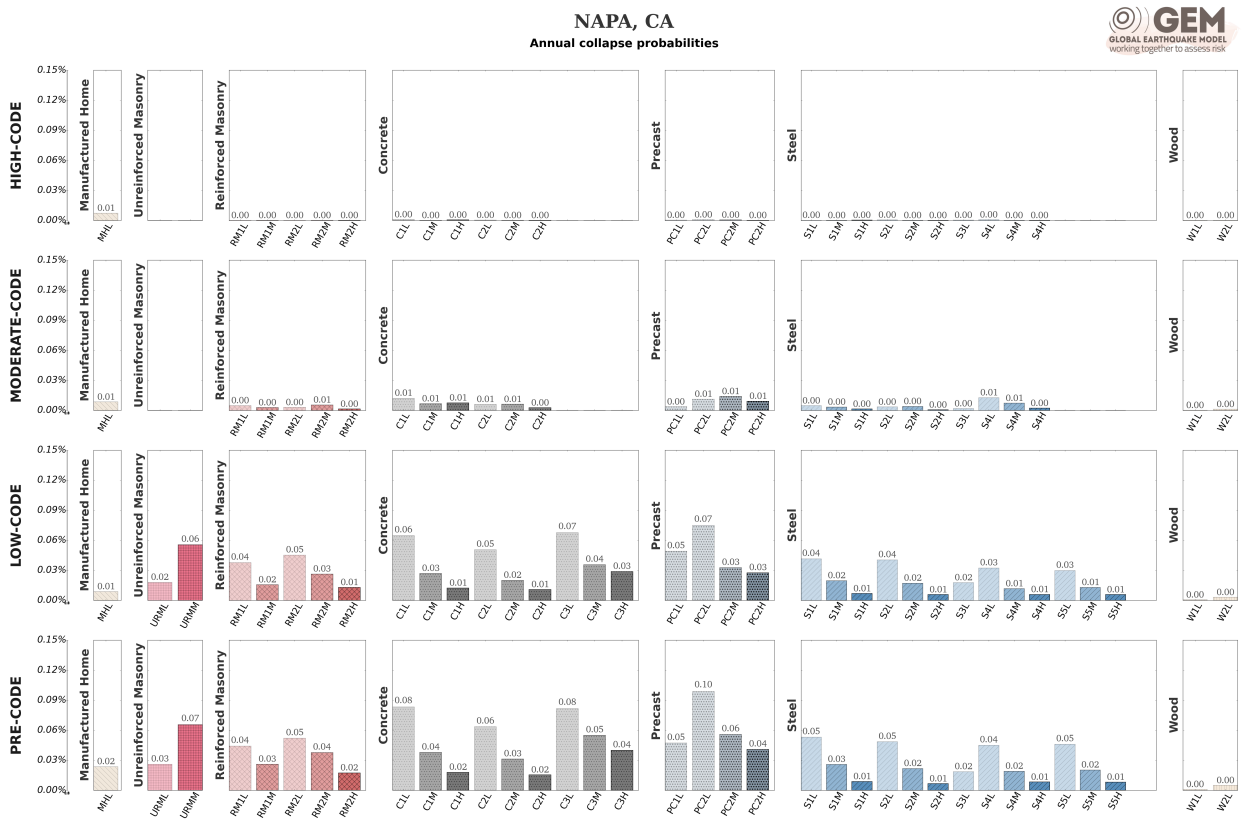


Figure 12. Annual collapse probabilities for 128 building classes in Napa, CA.

5.3 Probabilistic seismic risk results

Whilst a scenario simulation considers a single earthquake event, stochastic calculations involve simulating seismicity within the region over a long time-period that typically spans thousands of years. Damage and losses are estimated for each of the earthquakes in the simulated catalog of events. Finally, risk metrics such as average annual losses, or losses corresponding to various return periods can be extracted from this set of individual event losses.

The OpenQuake-engine stochastic event-based calculator employs a Monte Carlo simulation technique to estimate the loss distribution for individual assets and aggregated loss distribution for a spatially distributed portfolio of assets within a specified time period. The main results of this calculator are event loss tables, which describe the total loss across the portfolio for each seismic event in the stochastic event-set. Aggregated loss exceedance curves can also be generated for the portfolio and loss maps for the region, which describe the loss values that have a given probability of exceedance over the specified time-period.

Dedicated OpenQuake-engine hazard and risk calculators have been developed to perform such simulations for all branches of the 2014 seismic hazard model of the State of California. The details of the implementation of these new calculators are provided in Appendix A.7. Using these calculators, catalogs of events spanning a period of 10,000 years were generated for each of the 1,440 rupture model branches of the UCERF3 time-independent logic-tree. Two such catalogs are shown below in Figure 13. The five ground motion models, selected for California in the NGAWEST2 project, were used to simulate ground motion fields for each event in the 1,440 catalogs, and two different site models were used to account for site-amplification of ground motion values. Portfolio losses were computed for each of the resultant (1,440 rupture model branches × 5 ground motion model branches × 2 site model branches =) 14,400 branches. The following sub-sections describe the two prominent risk metrics computed for each of these branches: the average annual losses, and loss exceedance curves.

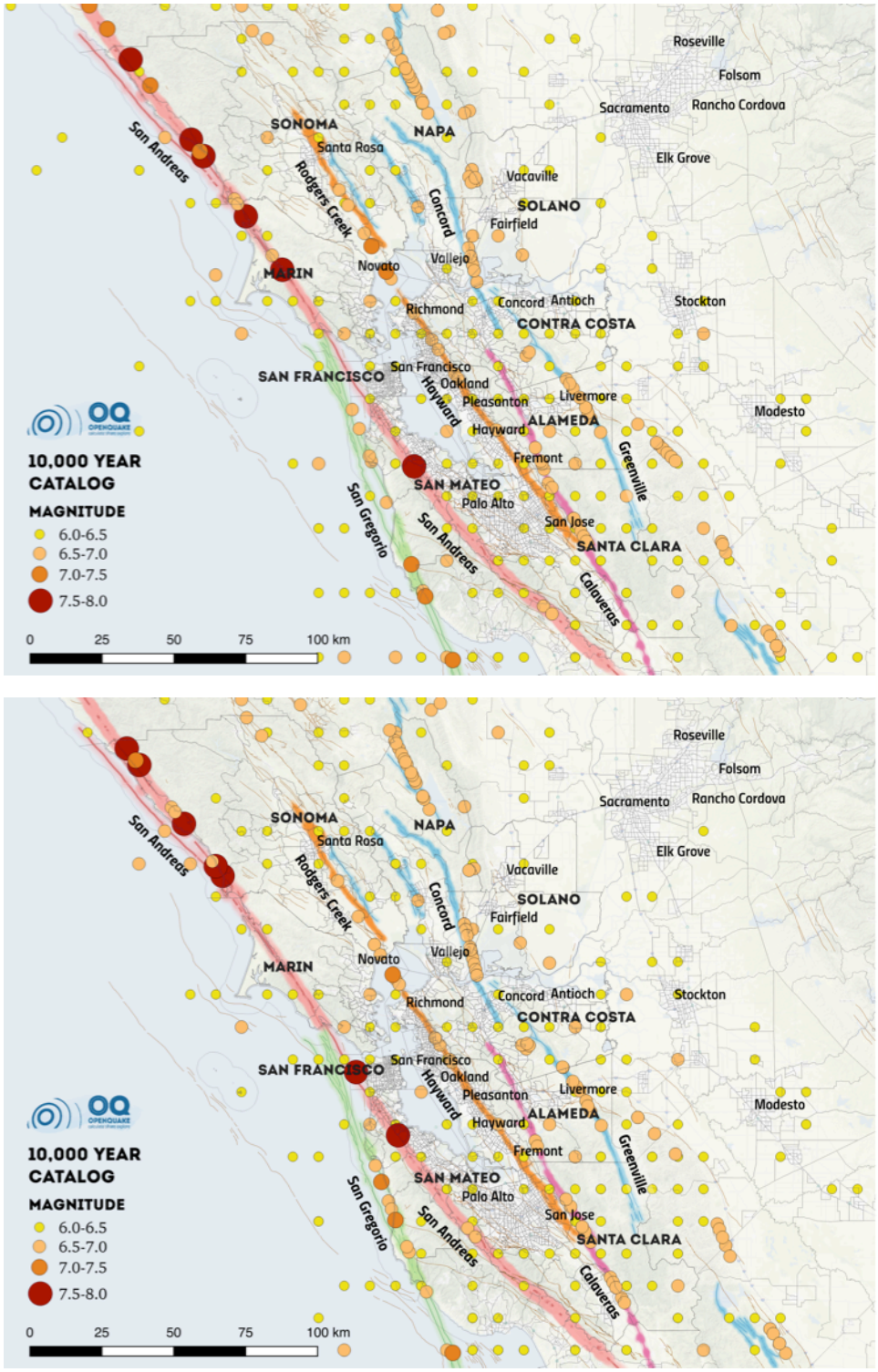


Figure 13. Two stochastic event-sets spanning 10,000 years showing the magnitudes and epicenters of earthquakes simulated in the San Francisco Bay Area. Earthquakes below magnitude M_w6.0 are not displayed on the map for the sake of clarity.

5.3.1 Average annual losses

The average annual loss (AAL) for a portfolio is computed by simply dividing the sum of all losses by the duration (in years) of the stochastic event catalog used for the calculation. Estimates for the AAL for the residential exposure in the San Francisco Bay Area were computed for all branches of the 2014 NSHM for California using the specialized UCERF3 risk calculator in the OpenQuake-engine. The computation was repeated for two different site-condition models, the Vs_{30} models from Wald and Allen (2007) and Wills et al. (2015). Table 8 in 0 displays the AAL values for the individual branches of the logic tree. Due to space constraints, only a selected section of the full AAL table is displayed in Table 8.

The AAL is typically normalized by the total exposed value of the portfolio and the resultant average annual loss ratio (AALR) is reported as a percentage. The histogram for the AAL values for the residential exposure in the San Francisco Bay Area obtained from Table 8 is displayed below in Figure 14. This histogram clearly highlights the importance of tracking the impact of different assumptions throughout the modeling process, starting from basic hazard parameters through to estimation of the final risk metrics. Most earthquake risk modeling tools, including Hazus, do not consider the epistemic uncertainty involved in the modeling process and provide mean or median results only. An understanding of the sensitivity of the results to model assumptions and parameter uncertainties cannot be obtained using such tools.

The weighted mean AAL across all branches in this example is $610 million. However, we observe from the histogram that depending on the model assumptions, the AAL estimate varies widely, ranging from $390 million to $890 million. Decisions undertaken based on mean estimates alone may thus be under- or over-estimated, relative to a realistic range of uncertainty in the mean loss estimate. A direct conclusion of this study is that it is imperative to propagate the uncertainties involved in the modeling process from hazard through to the estimation of the final risk metrics, and the full distribution of estimates should be used for optimal decision-making. The decision maker then has the option select a level of conservatism (or not), depending on the application or degree of risk aversion deemed appropriate.

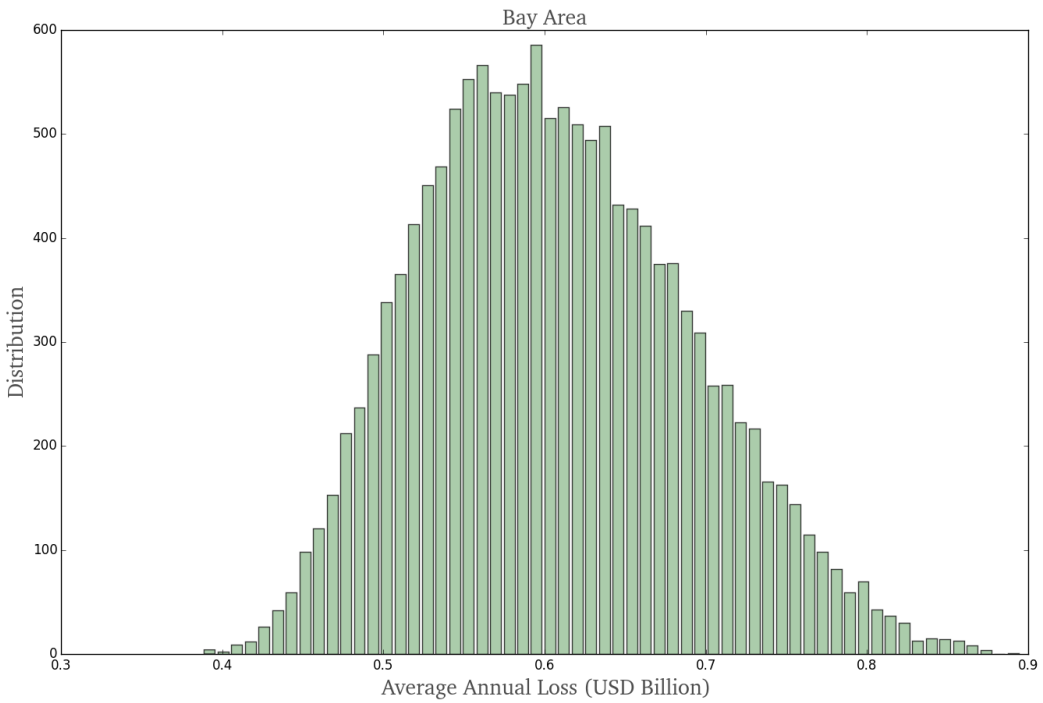


Figure 14. The distribution of the estimated average annual loss ratio for the San Francisco Bay Area for each of the 14,400 branches of the logic-tree.

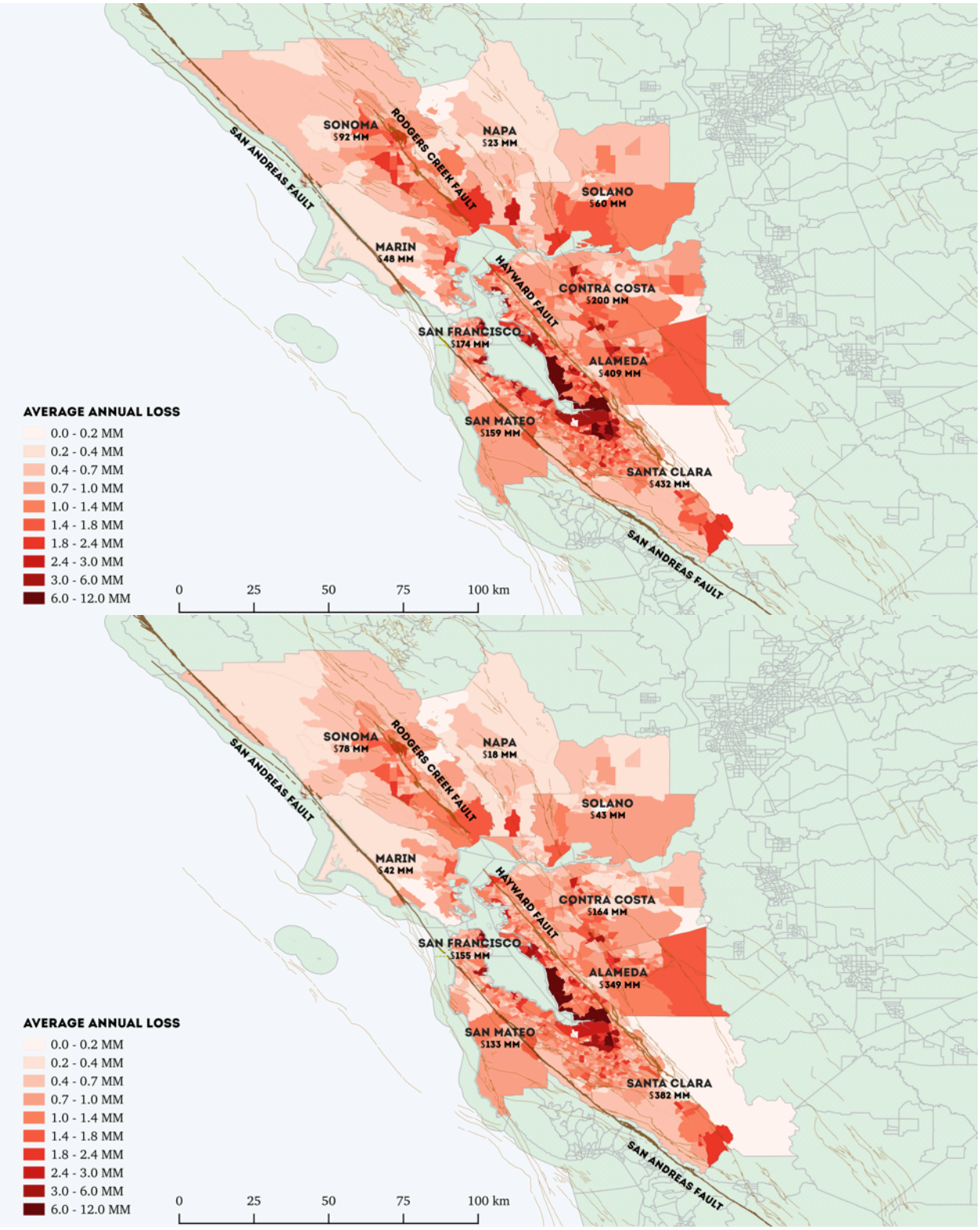


Figure 15. Average annual loss maps for the residential exposure in the San Francisco Bay Area computed using two different GMPEs. Top: Boore and Atkinson (2008); Bottom: Chiou and Youngs (2008).

The average annual loss estimates, when aggregated by census tracts and visualized on maps, can provide risk mitigation planners valuable information concerning the geographic spread of relative economic risk across the region under study. For instance, the distribution of limited economic resources for retrofitting of outdated (i.e., not compliant with the current building code) structures can be prioritized based on such maps. Insurance firms can use such maps to inform optimal policy rate setting. Two such AAL maps for the San Francisco Bay Area, calculated using two different ground motion models are displayed in Figure 15. A comparison of the two maps in Figure 15 reveals that the average losses for each county estimated using the Chiou and Youngs ground motion model are consistently lower than those estimated using the Boore and Atkinson model, although the spatial pattern of the loss distribution across the various census tracts of the Bay Area are similar.

The average annual losses for all branches of the logic-tree were calculated for the residential exposure in six different statistical regions of California. The six regions included the five statistical areas with the highest population agglomerations according to the US Census Bureau: the Los Angeles Metropolitan Area, the San Francisco Bay Area, the Inland Empire Region, the San Diego Metropolitan Area, and the Greater Sacramento Area[17]. The smaller region of Napa County was also included in the analysis. The weighted mean across all 7,200 branches of the logic-tree for these six regions is shown below in Figure 16. The same figure also shows the average annual loss ratios (AALR), which are obtained by normalizing the AAL values for the different regions by the total exposed values for those regions. The AALR plot facilitates a quick visual comparison of the relative seismic risk across different regions.

[17] **The San Francisco Bay Area** includes Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Solano, and Sonoma counties
The Greater Sacramento Area includes El Dorado, Nevada, Placer, Sacramento, Sutter, Yolo, and Yuba counties
The Los Angeles Metropolitan Area includes Los Angeles and Orange counties
The Inland Empire includes San Bernardino and Riverside counties
The San Diego Metropolitan Area includes San Diego county

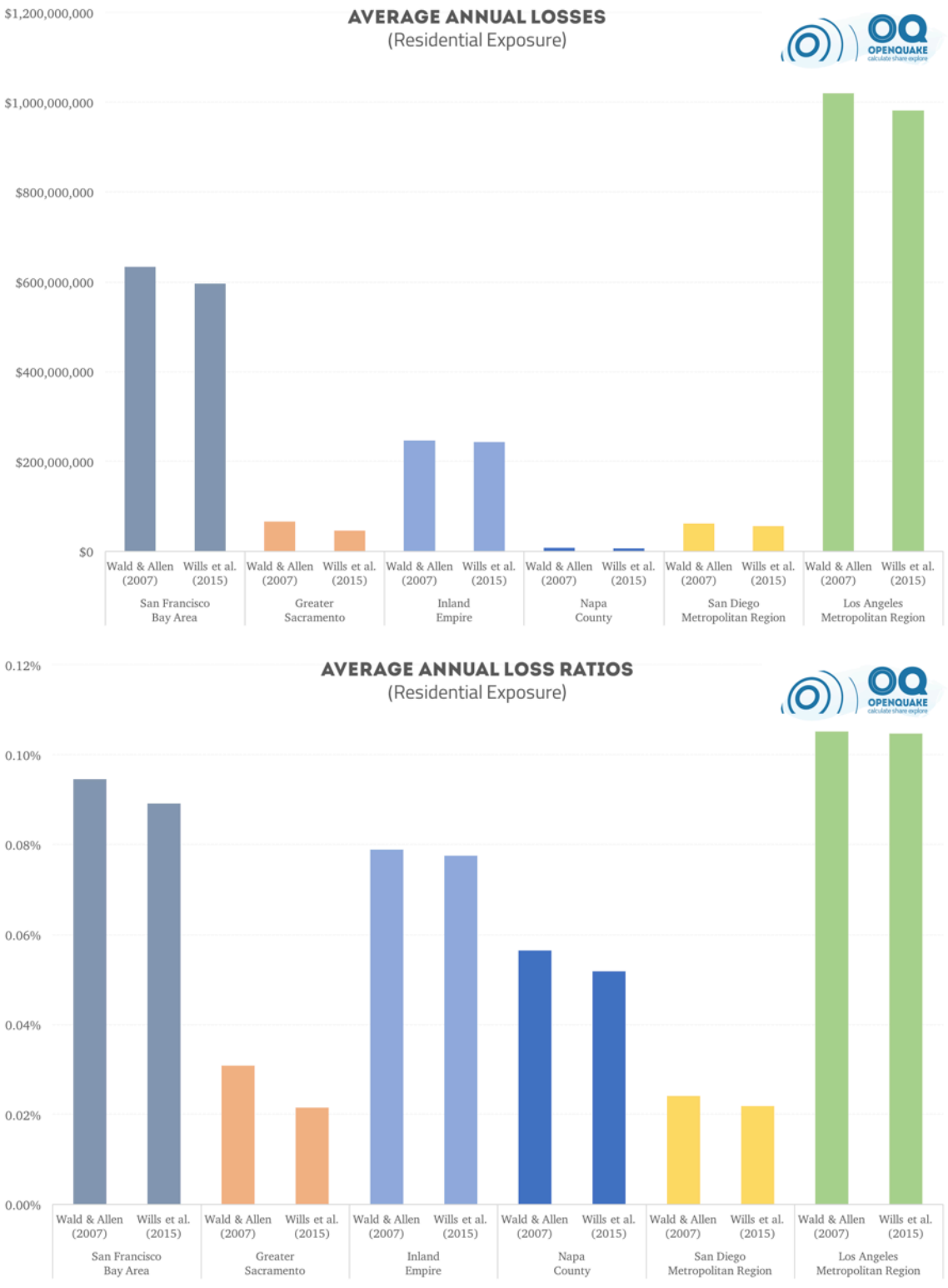


Figure 16. Average annual losses (top) and average annual loss ratios (bottom) for the residential exposure calculated for six statistical regions in California

We observe from Figure 16 that the expected annual losses are highest in the Los Angeles Metropolitan Region and the San Francisco Bay Area, both in absolute dollar values and in terms of the values normalized by the exposed values. For the smaller region of Napa County, the expected annual losses in absolute terms are negligible in comparison to bigger regions such as Greater Sacramento and the San Diego Metropolitan Region, but if we compare the expected annual loss *ratios*, we observe that the seismic risk is higher for Napa County in comparison to Greater Sacramento and the San Diego Metropolitan Region. Another observation from Figure 16 can be made regarding the impact of the choice of the site conditions model on the AAL estimates for the different geographical regions. The difference between the weighted AAL estimate using the two site models is less than 6% for the Los Angeles Metropolitan Region and the San Francisco Bay Area. On the other hand, for Greater Sacramento where the geology based Wills et al. (2015) site model predicts higher Vs_{30} values in most parts of the region compared to the topographic slope-based Wald and Allen (2007) site model, the two estimates differ by more than 40%. Thus, care must be taken in generalizing conclusions regarding parameter sensitivities for different combinations of models or applications to different portfolios and or regions.

The average annual loss map for the entire state of California computed using the UCERF3-based 2014 seismic hazard model is shown below in Figure 17. Figure 18 shows the breakdown of the average annual loss for California according to the general building categories. Losses to wood-frame residential structures dominate the overall losses, as these structures account for over 90% of the residential housing units in California.

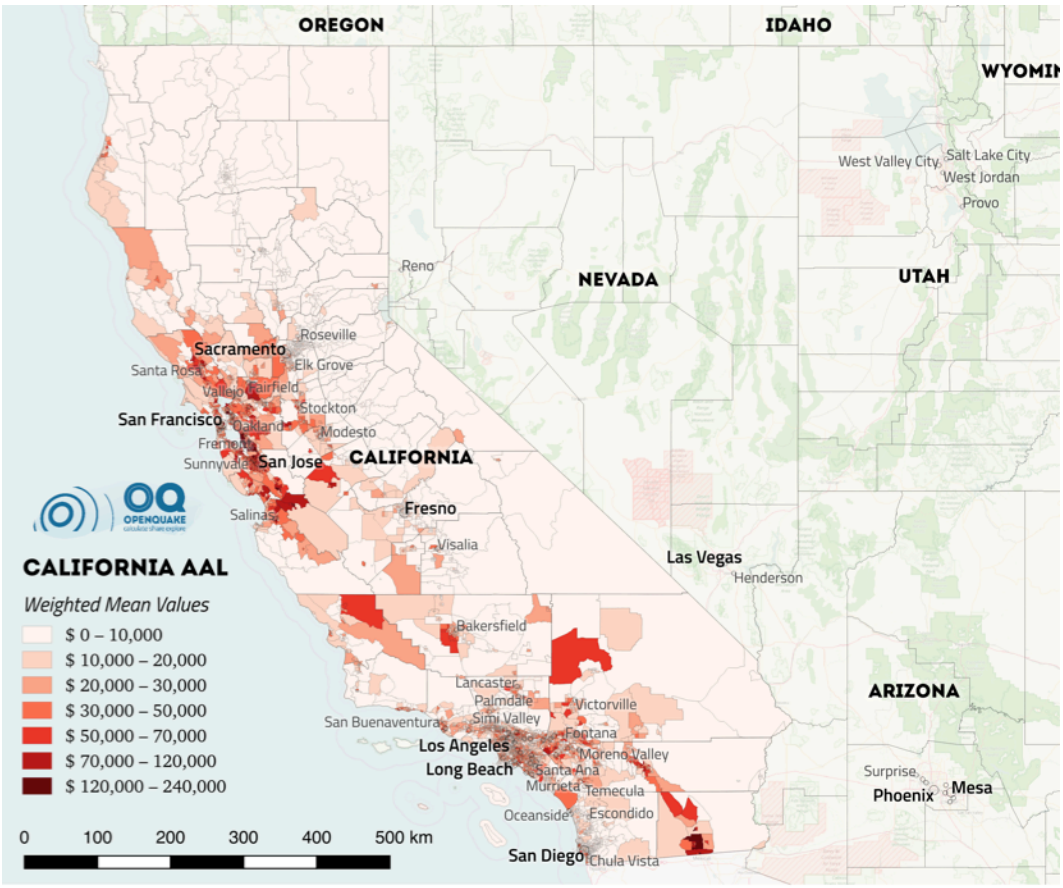


Figure 17. Average annual loss for the residential exposure in California

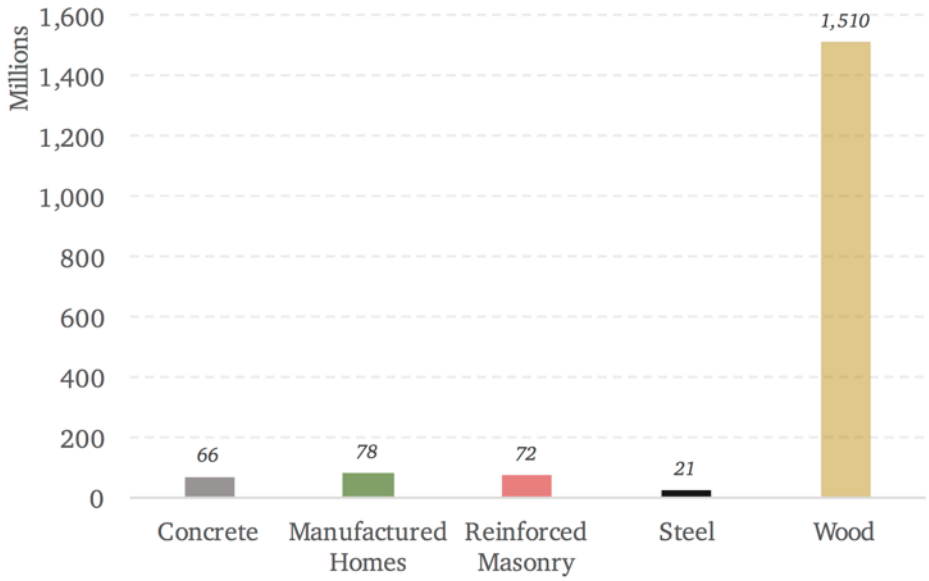


Figure 18. Breakdown of the average annual loss for the residential exposure in California by general building classes

5.3.2 Loss exceedance curves

The aggregated loss exceedance curves describe the probability of exceedance of different loss levels for all assets in the portfolio. These loss exceedance probabilities are typically reported in the form of portfolio loss values (expressed as a percentage of the total exposed value of the portfolio) versus the corresponding 'return periods". The return period is the inverse of the annual probability of exceedance. For instance, for the 250-year return period loss, there would be a 0.4% probability of this loss value being exceeded in any given year.

Loss exceedance probabilities for the residential exposure in the San Francisco Bay Area were computed for all branches using the specialized UCERF3 risk calculator in the OpenQuake-engine. As in the case of the AAL calculations described in the previous section, the computation was repeated for two different site-condition models, the Vs_{30} models from Wald and Allen (2007) and Wills et al. (2015). The resultant loss exceedance curves for the full logic-tree are shown in Figure 19.

Once again, the considerable spread in the estimates of the 100-year and 250-year return-period losses highlights the importance of not relying solely on the mean-branch results for decision-making. A more robust decision-making process would take into consideration not just the mean or median estimate as is typically done, but also ensure that the body and full range of the estimates are considered.

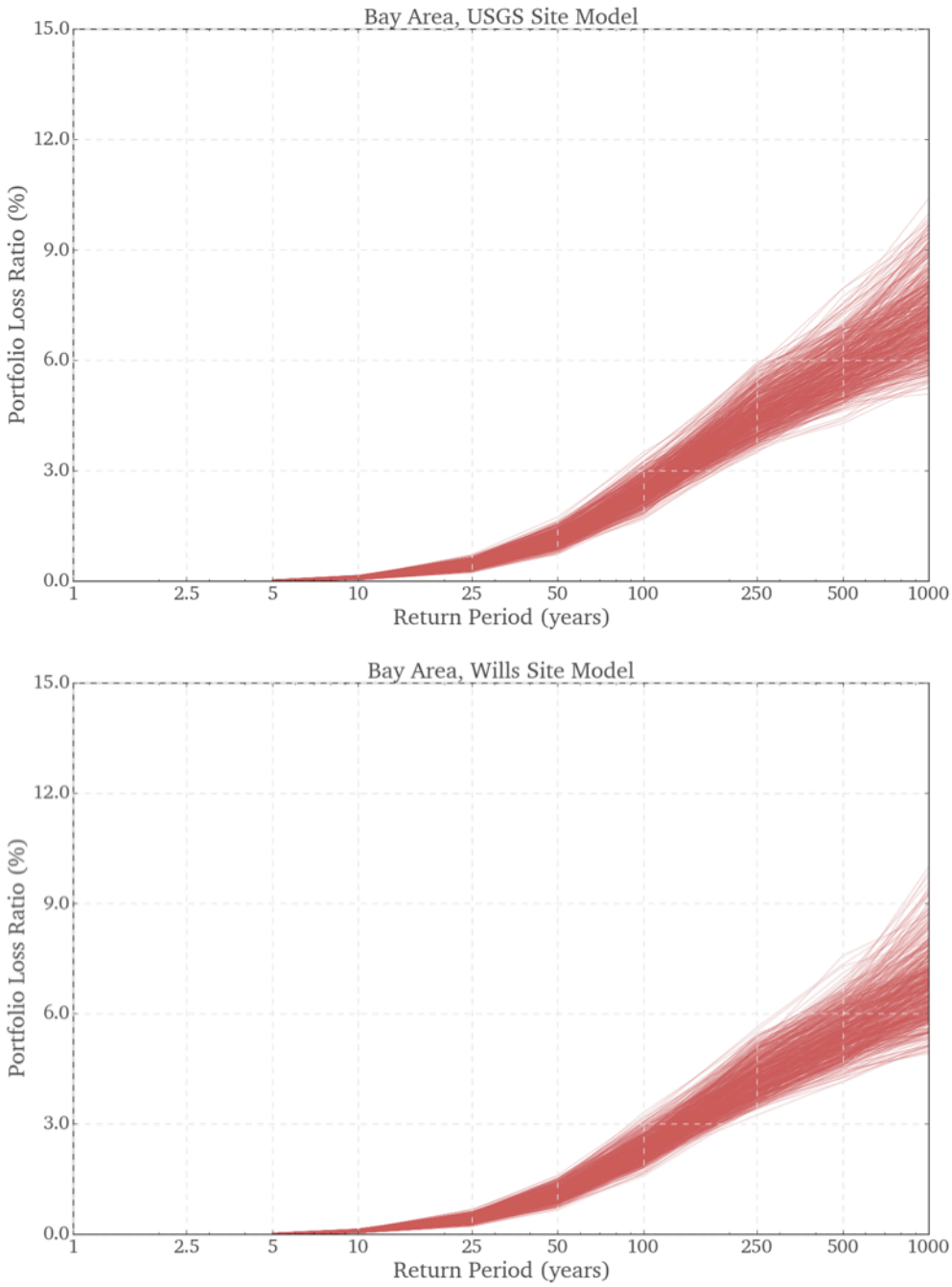


Figure 19. Loss exceedance curves for residential exposure in the San Francisco Bay Area for each of the 7,200 branches of the hazard logic-tree, calculated using two different site-condition models. Top: Using the Wald and Allen (2007) Vs_{30} model; Bottom: Using the Wills et al. (2015) Vs_{30} model.

5.4 Sensitivity analysis

Some questions that one should attempt to answer when faced with complex risk models involving multiple alternate assumptions for the different components include:

- Which of the model components have the most impact on the full distribution of the final risk metrics?
- Can reasonably close approximations of the distributions of the final risk metrics be obtained using a smaller subset of branches of the full logic-tree?

If we succeed in answering these questions, hazard and risk modelers can concentrate their efforts on improving those components that most impact the final risk metrics, whilst expending comparatively less effort on the components that affect the risk metrics less. Furthermore, we can significantly reduce computational run-times if we succeed in identifying a smaller subset of branches of the full logic-tree that yield reasonably close estimates of the distributions of the final risk metrics.

A sensitivity analysis was undertaken to identify the model components that make up the leading sources of uncertainty in the AALR values. We computed the sensitivity starting from the AALR estimates for the residential exposure in the San Francisco Bay Area for the 7,200 branches of the full logic-tree, and applying two different site-condition models. The model components included in the sensitivity analysis comprised the fault models, the deformation models, the magnitude scaling relationships, the models describing slip along ruptures, the different assumptions for the total rate of M≥5 events in the region, the assumptions for the value of Mmax for off-fault events, the off-fault spatial seismicity distribution models, the five ground motion models, and the two site-condition models. These model components are described in greater detail in Appendix A.1–A.3.

First, the end branch with the AALR value closest to the weighted-mean AALR across all branches was employed as the 'control" branch. To test the impact of each of the model components on the AALR value, only the models comprising that component were varied, whilst keeping all other components constant at the control branch values. The analysis was repeated using the end branch with the AALR value closest to the median AALR across all branches as the 'control" branch, to test the robustness of the results under the choice of a different baseline. The results of the sensitivity analysis are displayed graphically in Figure 20 in the form of a tornado diagram. The vertical line in the center of the top chart represents the *weighted mean* AALR across all 14,400 branches; the center line in the bottom chart represents the *median* AALR. The two edges of each bar in the tornado diagrams represent the minimum and maximum values of the range of AALR values for each model component.

From these tornado diagrams, we observe that the magnitude scaling relationship, the assumptions concerning the total rate of M≥5 events in the region, and the choice of the ground motion model are the biggest contributors of uncertainty in the estimate of the AALR for the San Francisco Bay Area residential portfolio. The two tornado diagrams in Figure 20 differ in the model components representing the extreme, but both identify the same top contributors of uncertainty.

The same analysis was also repeated for two other risk metrics: the 100-year and 250-year return period losses. In comparison to the AAL, the distributions for these two metrics are skewed to the left. Thus, we might expect the sensitivity ranking of the model components to be different for these metrics in comparison to the AAL. The tornado diagram for the 250-year return period loss, using the branch with the value closest to the median 250-year return period loss ratio as the control branch is shown below in Figure 21, for the San Francisco Bay Area and for the San Diego Metropolitan Area. Regional differences in the influential components can be observed in Figure 21. For instance, since the San Francisco Bay Area is situated in a high seismicity zone, the choice of the spatial distribution model for off-fault gridded seismicity (aka Off-Fault Spatial Seismicity PDF) has a much higher sensitivity in this region compared with the San Diego Metropolitan Area. These different tornado diagrams illustrate that the components contributing most to the overall uncertainty in the risk metrics can be quite different for different risk-metrics, as well as for different portfolios.

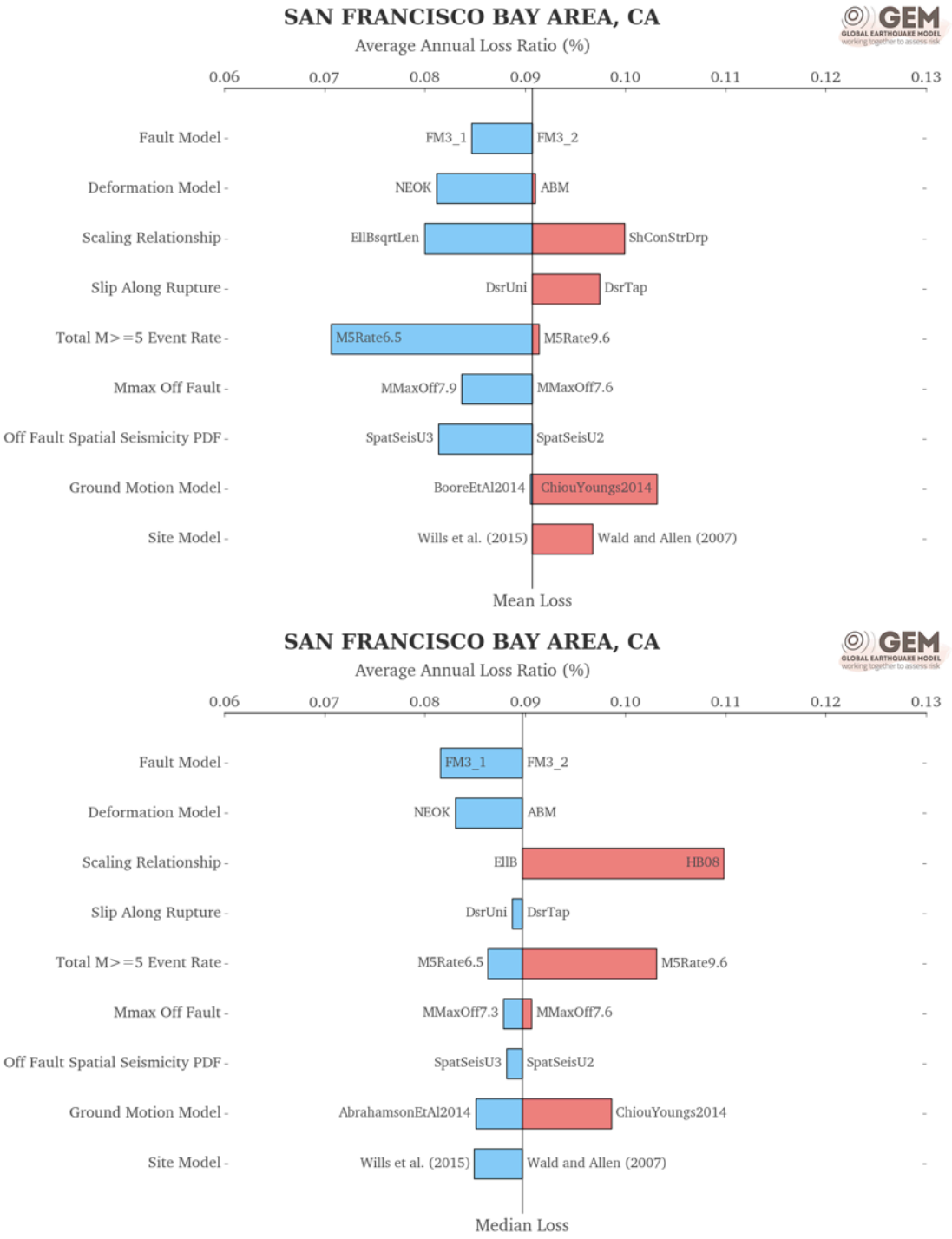


Figure 20. Tornado diagrams displaying the sensitivity of the estimated average annual loss ratio (AALR) to the different components comprising the full logic-tree. Top: using the branch with an AALR value closest to the *weighted mean* AALR computed across all branches as the control branch; Bottom: using the branch with an AALR value closest to the *median* AALR across all branches as the control branch. Please refer to Appendix Appendix A.1–A.3 for descriptions of the various model components.

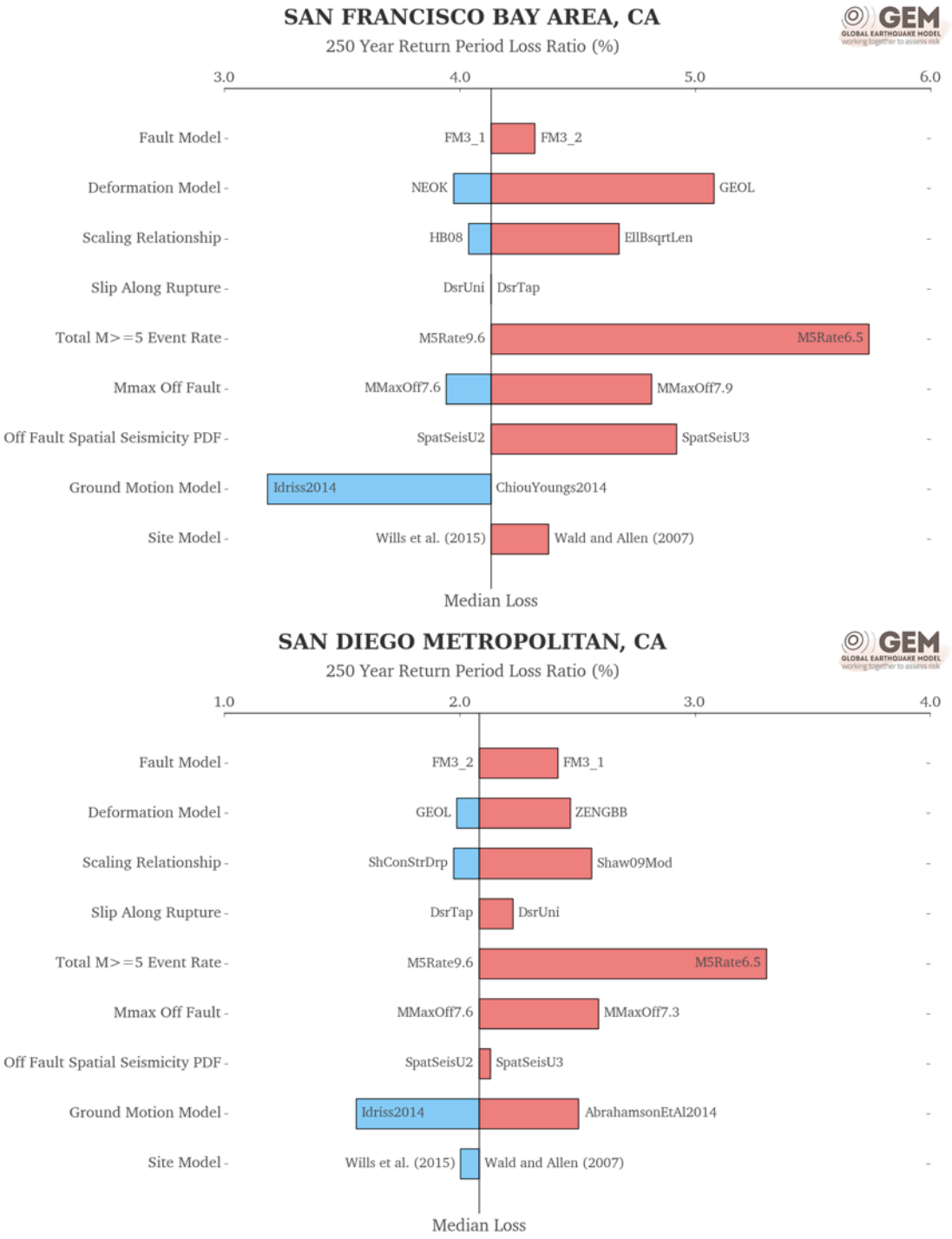


Figure 21. Tornado diagrams displaying the sensitivity of the estimated 250-year return period loss ratio to the different components comprising the full logic-tree, using the branch with the value closest to the median 250-year return period loss ratio as the control branch for two regions of California. Top: The San Francisco Bay Area; Bottom: The San Diego Metropolitan Area. Please refer to Appendix Appendix A.1–A.3 for descriptions of the various model components.

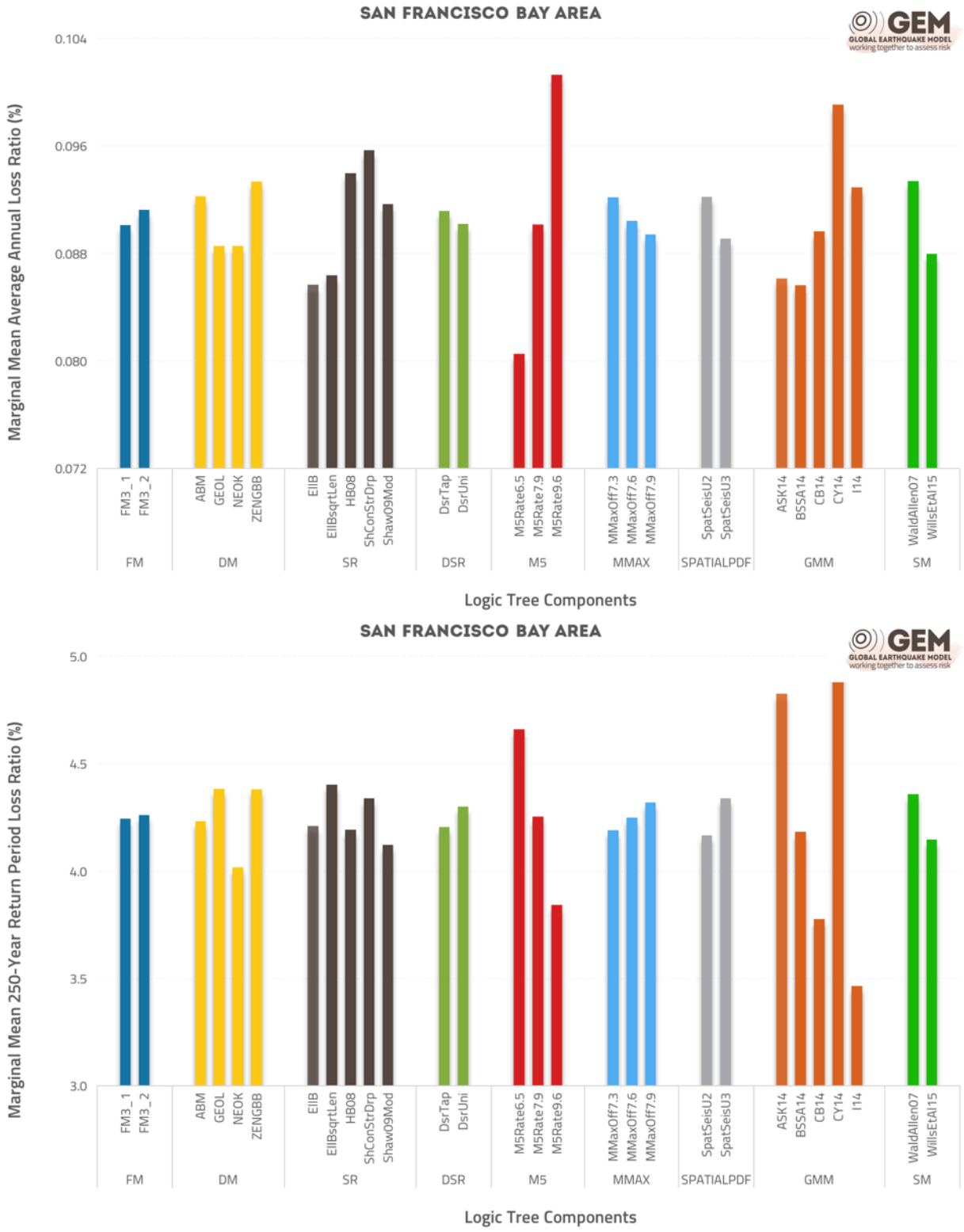


Figure 22. Marginal means plots displaying the sensitivity of the average annual loss ratio (Top) and 250-year return period loss ratio (Bottom) to the different components comprising the full logic-tree for the San Francisco Bay Area.

A second complementary approach to conducting sensitivity analysis involves estimating the marginal mean[18] average annual loss ratios for each component of the full logic-tree. If a component has little influence on the results, the marginal means within that component should be nearly equal, but if a component has a great influence on the results, then its marginal means would be expected to vary considerably. Figure 22 shows the results of such an analysis for the San Francisco Bay Area. The top panel in Figure 22 shows the sensitivity of the average annual loss ratio to the different components comprising the full logic-tree. Similarly, the bottom panel shows the sensitivity of the 250-year loss ratio. From these results, we can conclude that the components describing the fault model, the slip along the rupture, the maximum magnitude for events occurring off the modeled faults, and the off-fault spatial seismicity PDF have relatively low influence on the risk metrics for the San Francisco Bay Area. On the other hand, choice of ground motion model (GMM) and rate of events greater than M5 are shown to have the greatest influence.

5.5 Logic-tree trimming

Plotting the range of the marginal means provides us with a visual measure of the low sensitivity (low range) components and the high sensitivity (high range) components of the logic-tree, across different regions in California. Two such example plots for the structural losses to residential buildings in four different regions of California are shown in Figure 23.

The low sensitivity components can be trimmed off the logic-tree without much effect on the distribution of the risk metric under consideration. For instance, for the San Francisco Bay Area residential exposure, retaining only the components of the deformation model, the total rate of M≥5 events in the region, and the ground motion model provides a distribution of the average annual loss close to that obtained using the full logic tree. The 'pruned" (or 'trimmed") logic-tree in this case comprises only 75 branches in comparison to 14,440 branches for the full logic-tree. Figure 24 below shows the comparison of the histograms of the average annual loss for the residential exposure in the San Francisco Bay Area obtained using the full logic-tree with the approximate distribution arising from the pruned logic-tree. The same figure also shows a similar comparison for the 250-year return period loss.

[18] The marginal mean value of a risk metric for a branch is computed by taking the average of values of that risk metric for all solutions of the logic tree that involve that branch.

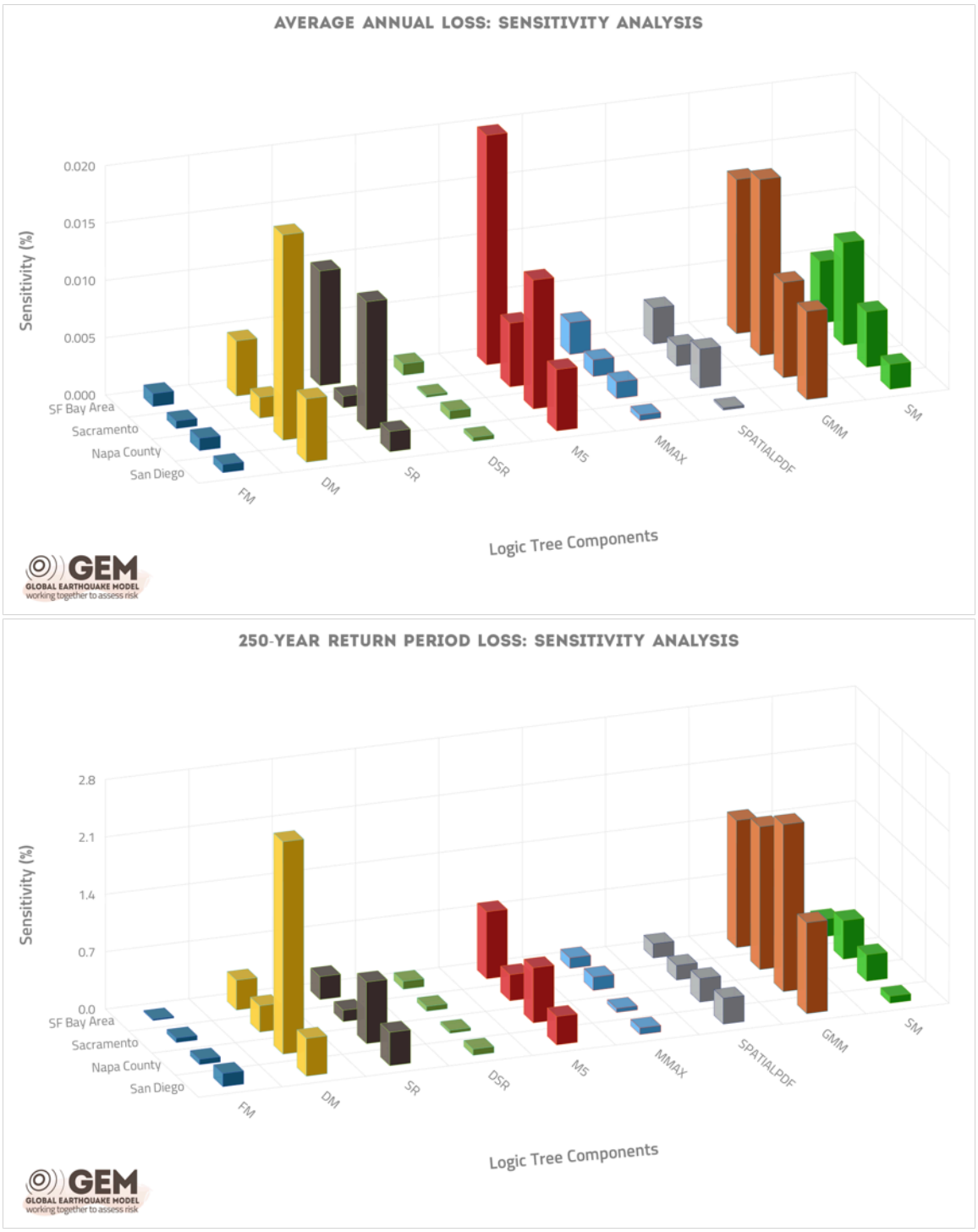


Figure 23. Range of the marginal means for each component of the full logic-tree for four regions in California: The San Francisco Bay Area, Greater Sacramento, Napa County, and the San Diego Metropolitan region. Top: Plot for the average annual loss; Bottom: Plot for the 250-year return period loss.

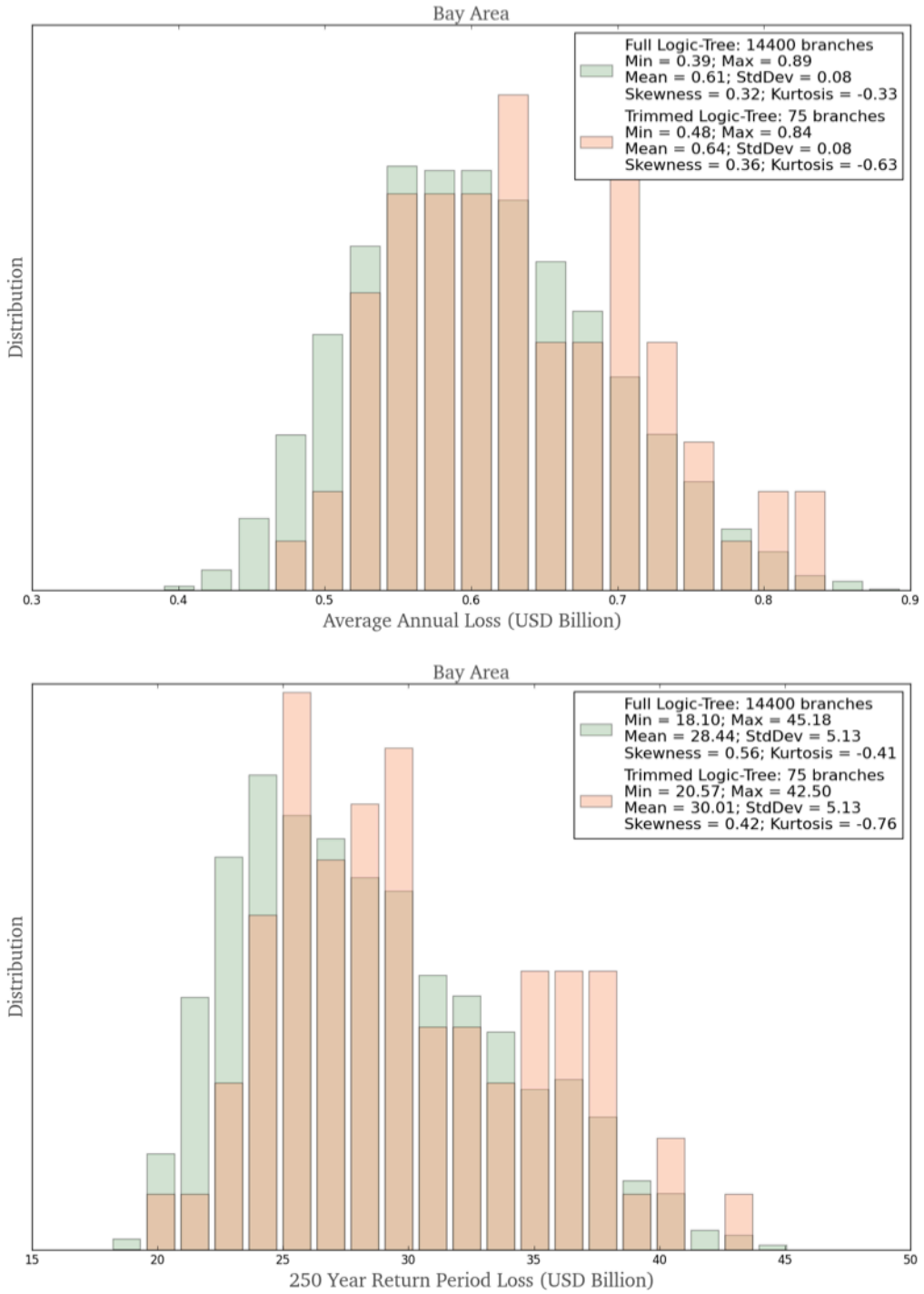


Figure 24. Comparison of the distribution of risk-metrics obtained using the full logic-tree comprising 14,400 branches with the approximate distribution obtained using a selected subset of only 75 branches, for the residential exposure in the San Francisco Bay Area. Top: Plot for average annual loss; Bottom: Plot for the 250-year return period loss.

This result has important consequences in terms of reducing the computational complexity and runtime of the risk model — in this case one can reduce the runtime for the model by a factor of almost 200, and still obtain reasonably close approximations of the distributions of the final risk metrics. It is important to recognize that while such a trimming exercise is extremely powerful for vastly reducing computation time, the trimming process must be tailored carefully to the portfolio and risk metrics of interest.

6 Discussions and recommendations

Earthquake models can be used to perform computer simulations that estimate losses from past or future earthquakes. Results from such model simulations are useful for actuarial (or risk transfer) purposes and for planning disaster risk reduction efforts. To perform calculations, models can be used to simulate tens or hundreds of thousands of earthquakes, over up to a million years of time (or to extremely low annual probabilities). If models incorporated all available scientific complexity and statistical uncertainty, such calculations could take months to complete, even on the fastest computer clusters. Thus, models used in practice make simplifying assumptions, or shortcuts, that allow the model to produce appropriate results on average, at a fraction of the computational expense (computer run-time). There are several trade-offs between mathematical completeness and computer run-time. The effect of shortcuts on results remains hidden, thereby limiting the applicability of the models and the ability to use them to properly inform risk reduction efforts.

The previous sections have looked at the impact of various assumptions made during the modeling process on different risk metrics, using the OpenQuake Engine to perform the calculations. Due to its highly flexible implementation, the OpenQuake Engine makes it possible to vary almost all the modeling assumptions for a hazard or risk calculation and study the effect on the results. Finally, the relative impact of different assumptions is investigated through demonstrative calculations for a residential exposure model constructed for the San Francisco Bay Area, California.

6.1 Recommendations

Recommendation 1: Uncertainties in the variables that make up the hazard model should be reflected in the estimates of risk. Decisions undertaken based on mean estimates alone may thus be based upon severely under- or over-estimated values, relative to a realistic range of uncertainty in the mean loss estimate. The body and full range of the estimates should be available to decision makers.

Recommendation 2: Consideration of all uncertainties through the risk modeling process is often prohibitively expensive and presents the primary hurdle that precludes exploring the full range of risk results. Sensitivity analysis such as the one conducted in this study can aid in simplifying the computational effort, allowing for the approximate calculation of

the distribution of risk results in significantly less time. The model simplification tool developed as part of this project can be used to achieve this objective.

Recommendation 3: The findings and methods used in the study may be appropriate to help develop seismic risk models that are transparent and address the full range of uncertainties in the input variables of the risk model. This study has explored the uncertainties in the hazard components in a comprehensive manner. Future work should extend the same level of detail to analyzing the uncertainties in the vulnerability and exposure components.

Recommendation 4: The OpenQuake development of the highly complex and scientifically advanced UCERF3 earthquake hazard model is the first open-source implementation of UCERF3 into a risk modelling software. Given the open and transparent nature of this product, its use to inform the development of policies for risk mitigation, emergency preparedness, and the design of buildings and infrastructure in California is highly recommended.

Recommendation 5: The tools and methods used in this study should be used to understand the seismic risk for exposure portfolios in regions of California outside of those covered by this study, as well as for the rest of the United States. This could require the use of additional information for estimating the likelihood of earthquakes or the resulting ground shaking intensity, particularly for studies outside of California.

Recommendation 6: Future work should explore the creation of a robust decision-making framework that incorporates uncertainties into the development of programs for reducing risk cost-effectively, thereby improving public safety and community resilience. This framework should include communication of the science of earthquake risk and its uncertainty to non-experts. In particular, improving the interaction between the public, policymakers and the scientific community is necessary for facilitating more effective decision making in earthquake risk management.

6.2 Present shortcomings and future directions

Epistemic uncertainty in earthquake risk assessment refers to our lack of knowledge and insufficient understanding of the different phenomena concerning earthquake processes and the behavior of the built environment under earthquake ground shaking. Such uncertainties are typically expressed through the employment of multiple alternate models or conceptual representations of the underlying phenomena. Although this study examined several sources of epistemic uncertainty and their relative contribution to the spread in risk estimates, several more sources of uncertainty remain unaccounted for and merit more detailed study.

This study specifically examines sources of uncertainty in the earthquake hazard (e.g., earthquake magnitudes and ground shaking) and their relative contribution to the spread in risk estimates. Not addressed are the contributions to the uncertainty in risk from uncertainties in the assumptions made in the exposure and vulnerability models. These two sources of uncertainty were out of scope of the present study and should be considered in future work.

The construction of the exposure model involves assumptions regarding the mapping from building occupancy types to specific construction classes, the assignment of design-eras and height classifications to buildings, and the geographic locations of the buildings within the region of study. Exposure models are arguably the component with the highest uncertainty in most risk models due to the large number of assumptions that are required to develop them. The impact of each of these assumptions can be studied by comparing results obtained using coarse exposure models that involve aggregation assumptions with those obtained by using high-resolution, building-by-building exposure models.

Vulnerability models for building classes typically include large uncertainties due to the assumptions involving the geometric and material properties of buildings, the variability of the building's response to different earthquakes, the variability in the onset of damage due to deformation, and uncertainties in estimating the cost of repair and reconstruction following damage. In addition, there is a possibility that all buildings of a particular class may perform better or worse than the average during an earthquake, and such correlations should be accounted for in a risk model. OpenQuake can already handle these uncertainties and correlations in the vulnerability model; but their impact on risk results was out of scope of scope of the present study.

The scenario-type analyses in this study focus primarily on the physical damage to the structural components of buildings and the consequent direct economic losses due to such damage. The sensitivity analysis focuses primarily on understanding the sensitivity of loss estimates to alternative input assumptions regarding the earthquake hazard model. Not considered in the analyses are the following:

- damage to building contents and indirect losses due to business interruption;
- losses due to secondary hazards such as landslides, liquefaction, tsunamis, and fire-following earthquake;
- impacts to the infrastructure, lifelines, and utilities; and
- aftershocks, which can cause damage to structures that were previously weakened during a *mainshock* earthquake.

Such additional sources of losses should be considered in a future study in order to identify all of the potential consequences of a large earthquake.

It is the intent of this study to encourage the creation of transparent risk-models, where the large uncertainties inherent in seismic risk modeling are not concealed, but rather

highlighted and propagated through to the final risk-metrics. Emergency management plans, (re)insurance policies, and building codes should be informed by the best available science. In some cases, however, the best available science involves large uncertainties. Such is the case of risk analysis based on UCERF3, where the full hazard model (including the time-dependent model) is really a collection of 86,400 alternative scientifically viable models. Future work should explore the creation of more robust decision-making frameworks that can translate risk results involving such large uncertainties into policies that lead to higher public safety and enhance community resilience.

Data and Models

This appendix contains a description of various datasets and components of the risk model used to perform most of the calculations described in the body of this report, including:

- The earthquake occurrence model
- The ground motion models
- The site models
- The exposure data
- The fragility and vulnerability models

A.1 Earthquake Occurrence Model

The seismic risk calculations performed within this project are based upon the time-independent version of the recently published Uniform California Earthquake Rupture Forecast version 3 (UCERF3), produced by the U. S. Geological Survey (Field et al. 2013; 2014) and the Working Group on California Earthquake Probabilities (WGCEP). The UCERF3 model is revolutionary in many respects, being the first model of its kind to consider the behavior of earthquakes in a region at a system level, rather than as a collection of individual, independent faults. Each of the 1,440 sources models within the logic tree describes an exhaustive set of possible ruptures for California, alongside their associated rates of occurrence. The exhaustive set of ruptures and associated probabilities of occurrence in a specified time period is referred to as the *earthquake rupture forecast* (*ERF*). Each rupture forecast is consistent with the large suite of geological and geodetic observations from California and the surrounding region.

Previous models of earthquake rupture behavior in California, such as version 2 of the Uniform California Earthquake Rupture Forecast (UCERF2) (Field et al., 2009), required complex judgements and assumptions regarding the possible geographic limits of the ruptures (known as *segmentation*) and the potential interactions between fault systems. The result is that whilst a large amount of expert knowledge was input into the model, certain inconsistencies emerged when compared against observed seismicity in California. The most significant of these inconsistencies was a systematic over-estimation of the rate of earthquakes in the range $6.5 \leq M \leq 7.0$ in California (referred to colloquially as 'the bulge problem').

UCERF3 represents a significant change in modelling philosophy, revising the constraints to a set of simpler hypotheses regarding the nature of earthquake ruptures, those hypotheses being well established in the technical literature and/or well-constrained by geological observation (Field et al., 2014). In UCERF2, and previous models., California's detailed and complex fault system was divided into segments corresponding to potential

rupture lengths, as inferred from geological observations and/or the geographic limits of historical earthquakes. As more evidence emerged to challenge the segmentation concept, in this construction, an innovative new modeling strategy emerged. In UCERF3 the entire California fault system is subdivided into equal sized segments of lengths typically on the order of 7 to 8 km, resulting in 2,606 and 2,664 'micro-segments" for the two considered fault models, FM 3.1 and FM 3.2, respectively. Each rupture in the rupture forecast is constructed from two or more micro-segments, subject to a set of plausibility criteria to avoid the creation of ruptures that are in contrast with the established physics of the rupture process. Within the constraints imposed by plausibility, and those provided by both the observations of geological slip at various locations within fault system as well as GPS measurements of tectonic deformation across California, a massive computational inversion process (the 'Grand Inversion") is undertaken to identify earthquake rupture forecasts (i.e. the rupture sizes and long-term rates) that represent the potential long-term behavior of the system (Page et al., 2014).

From the inversion process, each earthquake rupture forecast contains 253,706 and 305,709 unique ruptures for each of the branches associated to fault models 3.1 and 3.2, respectively. Those rupture forecasts found to match well the constraining data from the suite of 1,440 models that describe the *epistemic uncertainty* (i.e. the model-to-model uncertainty) of UCERF3 (see Figure 25). To quantify the scale of the calculation, should one try to evaluate the set of ERFs for the entire logic tree, as each fault model is associated to 720 branches the total number of ruptures that would be evaluated would be on the order of approximately 180 million for FM 3.1 and 220 million for FM 3.2.

The 2014 update the U.S. Geological Survey's National Seismic Hazard Model incorporates (NSHMP) the UCERF3 model as the basis for the seismic hazard assessment within California (Petersen et al., 2014). Due to the complex, and computationally intensive, nature of its logic tree, the version of UCERF3 adopted within the NSHMP is the 'true mean" model. This is a single model in which the activity rate associated with each rupture is taken as the mean of the activity rates from all branches of the logic tree in which the rupture appears, weighted by the probabilities assigned to the respective branch. The result is that when using the 'true mean" model only one single source model needs to be calculated, but the full range of epistemic uncertainty cannot be explored. The 'true mean" model should be equivalent to the mean from the full logic tree. Whilst this approach delivers a single outcome that fulfills the objective of defining a mean hazard map, the 'true mean" model does not allow for the exploration of uncertainties and their implications for hazard. Results from the mean model will be shown for the classical PSHA calculations below, and as a means of verifying the results of the OpenQuake-engine implementation of UCERF3 against those of the USGS. Otherwise, the full logic tree is supported where possible.

The full logic-tree for UCERF3 is illustrated in Figure 25, and the different branching points are described briefly below. There are three main groups of uncertainties: those relating to

the physical geometry of the fault system (the *FAULT MODELS*), those relating to the estimation of seismic slip on the faults (the *DEFORMATION MODELS*) and those relating to the calculated rates of occurrence of earthquakes of different magnitude within the fault system (the *EARTHQUAKE RATE MODELS*). Two alternative fault models and four deformation models are proposed. The deformation models describe four different means by which the long-term seismic slip on a fault can be inferred from geological and geodetic data. Of the four models the 'Geologic" branch presents a model of slip determined exclusively from geological measurements, whilst the remaining three branches estimate slip on the faults with additional constraints from global positioning system (GPS) data and/or plate tectonic motions. As an illustration of a fault and deformation model, Figure 26 shows the distribution of slip across the California fault system for FM3.1 and the 'GEOL" deformation model.

The logic tree branches relating to the earthquake rate models contain epistemic uncertainty models for several different elements of the seismogenic source. Each branch of the logic tree contains two types of source: i) ruptures modeled explicitly on one or more of the identified seismogenic faults, ii) earthquakes occurring in the deforming background, representing as-yet unidentified or slow slipping faults undergoing tectonic deformation. The latter are represented as a 0.1˚ by 0.1˚ grid of seismic activity rates, referred to hereafter as the 'off-fault spatial seismicity". Each complete source model in the logic tree (i.e. each of the 1,440 branches) contains both of these elements, for which

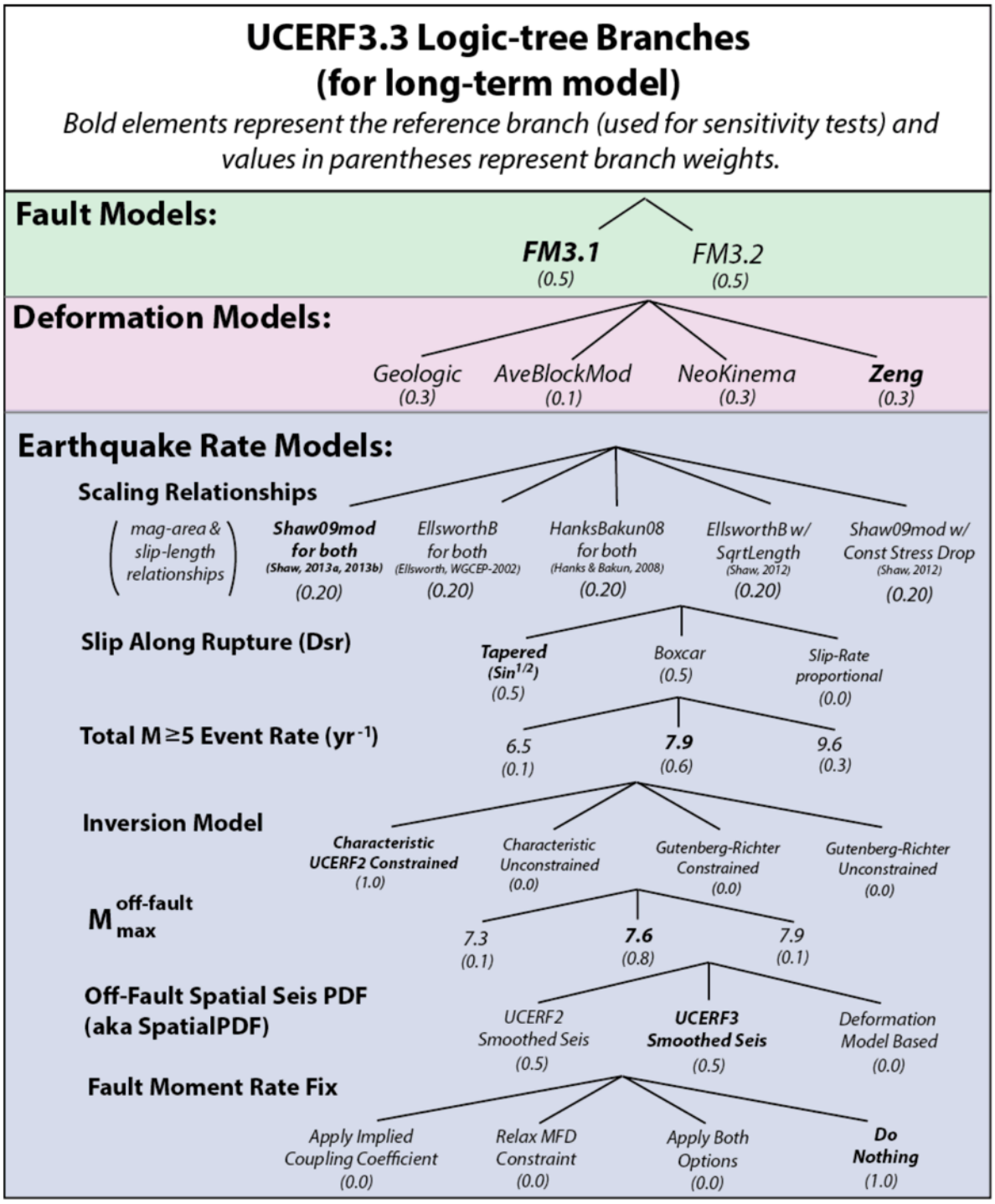


Figure 25. Logic-tree branches of the UCERF3 time-independent model. The full UCERF3 logic-tree comprises a total of 1,440 nonzero branches. Figure from Field et al. (2013).

reasonable agreement between the modeled seismicity and that observed both on a fault-by-fault and regional basis. For each branch the earthquake rupture forecast is

determined by defining the set of possible ruptures, and their long-term activity rates, that would satisfy a series of numerical and theoretical constraints. These include agreement with the long-term observed rate of slip on the fault system as measured on specific sub-sections of the fault, as well as agreement with the long-term rate of paleoseismic events measured at different locations along the fault. In addition to these, numerous other conditions are placed upon the model to prevent unrealistic or unwanted behavior, such as localized spikes of high activity, or connections across fault sections that would be inconsistent with the current knowledge of the fault rupture process. As it can be seen in Figure 16, for some uncertainties, different options have been explored, yet only one branch is assigned a non-zero weighting. The non-trivial uncertainties for the earthquake rate model are therefore the choice of magnitude scaling relation (i.e. the relation between the physical dimensions of an earthquake rupture, such as slip or area, and its magnitude), the characterization of slip along the rupture, the constraining total annual rate of events above M_W 5.0, the upper bound magnitude of the off-fault seismicity, and the actual rate of off-fault seismicity itself.

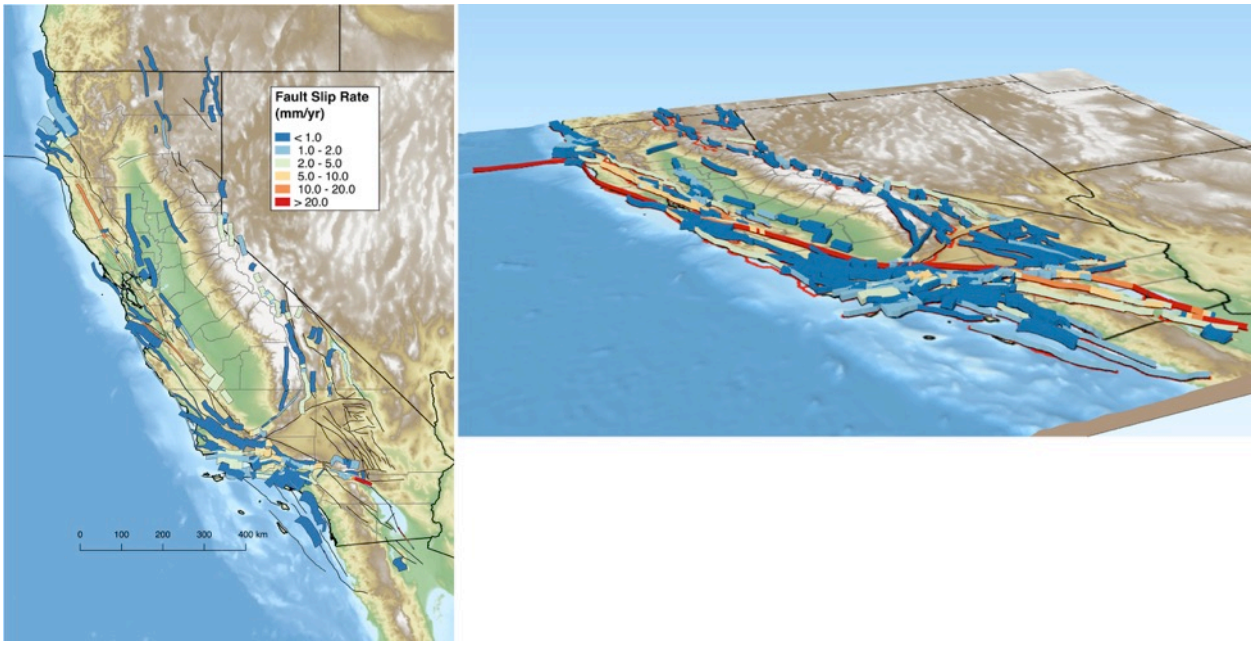


Figure 26: Example of the active fault structure of California from the UCERF3 model (Fault Model 3.1 – Geological Branch)

A.2 Ground motion models

Whilst the UCERF3 model itself does not specifically provide the ground motion model (or models) for application in seismic hazard analysis, within the 2014 NSHMP a suite of ground motion models (GMMs) were identified for use in the western United States. This

suite of GMMs consists of the five models developed within the Next Generation Attenuation (NGA) West 2 project (Bozorgnia et al., 2014). They are derived from an extensive database of ground motion observations from earthquakes originating within the shallow crust of active seismogenic regions (Ancheta et al., 2014). These are supplemented by numerical simulations of strong ground motions to improve the understanding of the behavior of ground motions under conditions that are found only sporadically within the set of observations (such as for large magnitudes in near fault regions). The five models selected for application in the western US are listed in Table 5, along with their respective weights in the logic tree.

The epistemic uncertainty in the choice of ground motion model is further widened within the US NSHMP to take into account the potential sample bias in the number of ground motion records within various magnitude and distances ranges used in the modeling (Petersen et al., 2014). This approach, initially adopted within the 2008 NSHMP (Petersen et al., 2008), used the 'square-root rule", where the ground motion values are increased or decreased by a factor (Δ), where $\Delta = 0.4\times\sqrt{\frac{n}{N}}$, N is the number of earthquakes recorded in each magnitude-distance bin, and n is the number of earthquakes in the bin for magnitude greater than M_W7.0 and distance from the rupture less than 10 km. This adjustment corresponds to an increase/decrease of a factor of about 50 % of the ground motion for magnitudes greater than 7.0 within 10 km of the fault, decreasing to approximately 25 % for well-constrained parts of the ground motion database. The result is that each of the five core GMMs is split into a further three models, the 'mean" model (without the adjustment factor), and the 'upper" and 'lower" models applying the increasing and decreasing adjustment factors respectively. The 'mean" model is assigned a weight of 0.63, and the adjustments 0.185 each. For those models assigned an initial weighting of 0.22 the final weightings are 0.1386 for the mean and 0.0407 for the adjustments, whilst for the Idriss (2014) model (assigned 0.12 in the initial weighting) these correspond to 0.0756 and 0.0222 respectively (Table 1). More detail regarding the derivation of the adjustment factors can be found in pages 145 to 146 of Petersen et al. (2014).

Table 5. Ground motion models for crustal earthquakes and their corresponding weights in the 2014 version of the hazard maps for the Western United States

Crustal WUS GMM	**Abbreviation**	**Weight (*Upper/Mean/Lower*)**
Abrahamson et al. (2014)	ASK14	0.22 (*0.0407/0.1386/0.0407*)
Boore et al. (2014)	BSSA14	0.22 (*0.0407/0.1386/0.0407*)
Campbell and Bozorgnia (2014)	CB14	0.22 (*0.0407/0.1386/0.0407*)
Chiou and Youngs (2014)	CY14	0.22 (*0.0407/0.1386/0.0407*)

Idriss (2014)	I14	0.12 (*0.0222/0.0756/0.0222*)

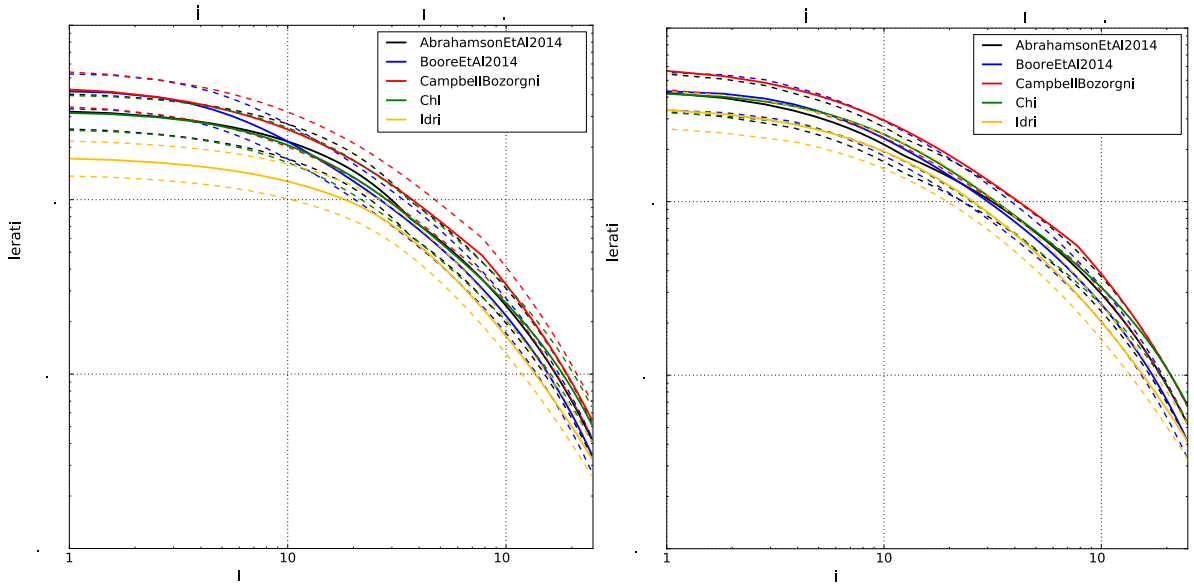


Figure 27. Expected ground motions from the five NGA West 2 GMMs for the Northridge earthquake scenario (left) and Loma Prieta earthquake scenario (right). The mean (unadjusted models) given by continuous lines, their adjustments by dashed lines.

To illustrate the full range of ground motions values expressed by this model of epistemic uncertainty, the expected ground motions (i.e. the decay in ground motion with distance from the source) of the five core GMMs, and the attenuation with distance, are compared for two 'scenario" earthquakes in Figure 27. The scenarios are based on the 1994 Northridge and 1989 Loma Prieta earthquake ruptures respectively, with rupture geometry configurations taken from the database compiled by Ancheta et al. (2014). It is relevant to note that at 10 km from the rupture the total range of ground motions spanned by the full suite varies by a factor of about 3, with greater variability found for the Northridge earthquake scenario at distances closer than 10 km.

A.3 Site model

Local soil conditions are taken into consideration in hazard and risk calculations typically through the use of Vs_{30} values in the ground motion models (GMMs)—where Vs_{30} represents the time-averaged shear-velocity (measured or inferred) within the uppermost 30 m. A few GMMs also require $Z_{1.0}$, the depth (in meters) to the soil layer where the shear-wave velocity first exceeds 1 km/s, and a few others require $Z_{2.5}$, the depth (in km) to the soil layer where the shear-wave velocity first exceeds 2.5 km/s. Two different site condition models were used for the calculations in this project: (1) the Vs_{30} model from Wald and Allen (2007) based on topographic slope; (2) the Wills et al. (2015) Vs_{30} model

based on surficial geology and topology. Seismic site condition maps based on these two models are shown below in Figure 28.

The $Z_{1.0}$ values were estimated using the regression equation recommended by Chiou and Youngs (2014) for California:

$$\ln Z_{1.0} = -\frac{7.15}{4} \ln\left(\frac{V_{S_{30}}^4 + 571^4}{1360^4 + 571^4}\right)$$

The relationship between $Z_{2.5}$ and Vs_{30} values recommended by Campbell & Bozorgnia (2014) for California sites was used to estimate the $Z_{2.5}$ values in this study:

$$\ln Z_{2.5} = 7.089 - 1.144 \ln V_{S_{30}}$$

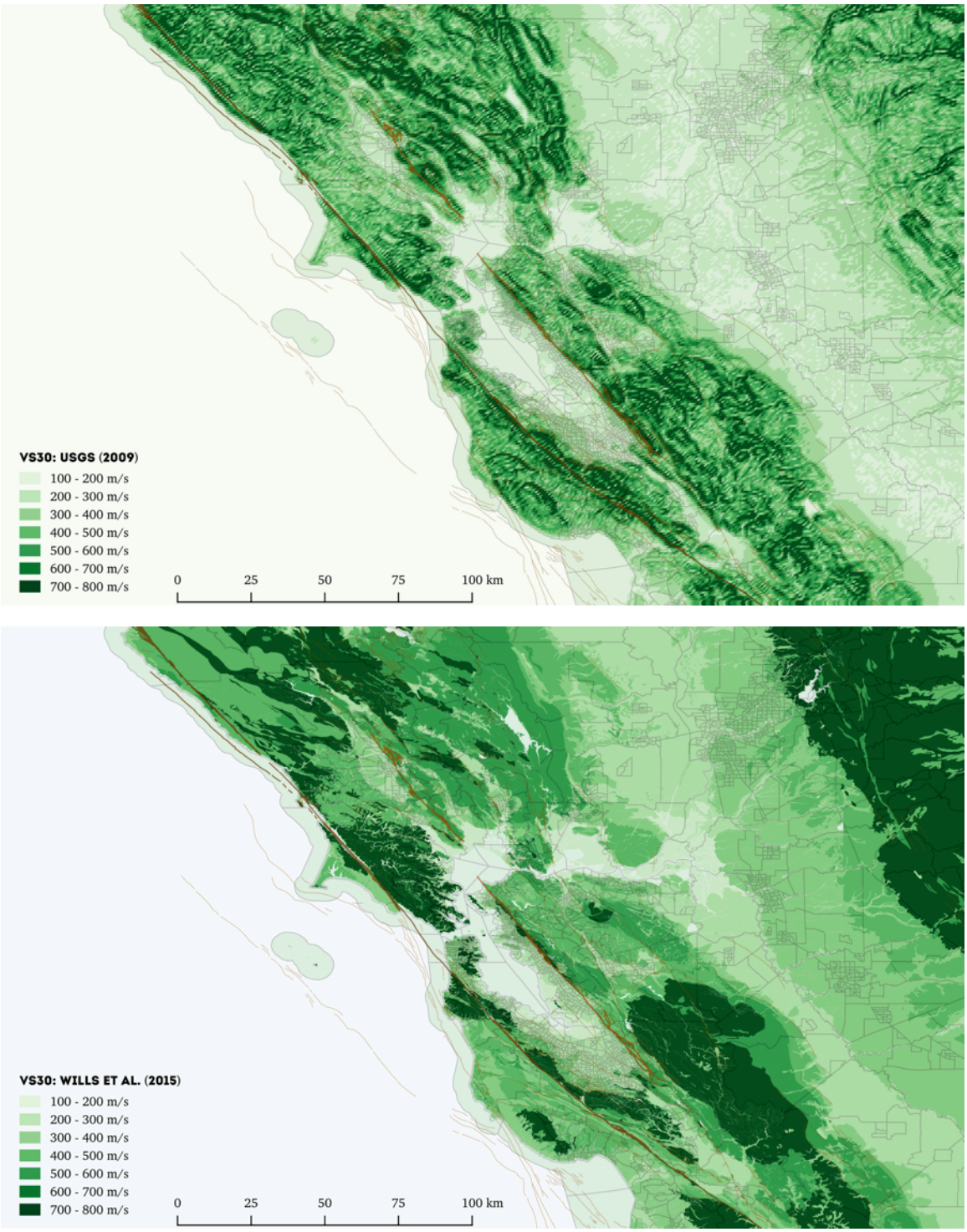


Figure 28. Seismic site condition maps displaying time-averaged shear-wave velocity in the uppermost 30 m (Vs_{30}) for the San Francisco Bay Area. Top: Wald and Allen (2007); Bottom: Wills et al. (2015)

A.4 Exposure data

For the purposes of the damage and loss calculations described in this project, a residential exposure model was constructed for the entire state of California at the census tract level, starting from the number of housing units within each tract as reported in the 2010 Decennial Census. The number of housing units were then transformed into estimates of the number of structures for each of the 36 Hazus building classes, by applying a series of 'mapping schemes" defined for Western U.S. buildings based on information provided in ATC-13. These building classes are listed in

Table 6 along with brief descriptions of each class.

The proportion of buildings in low, mid, and high-rise categories within each census tract was estimated based on the intensity of development identified in the 2011 National Land Cover Database for the conterminous United States. The proportion of buildings according to age in three categories (pre-1950, 1950-1970, and post-1970) was extracted from the housing data profile compiled in the 2010-2014 American Community Survey 5-Year Estimates. Finally, based on the age-profile of the buildings and the seismic design category assigned at the location of the buildings in the 1997 Uniform Building Code, the assets were categorized into pre-code, low-code, moderate-code, and high-code classes. Figure 29 shows the criteria used for the code-level classification.

The buildings within each census tract are assumed to be situated at the centroid of the tract. Overall, the exposure model comprises 10.02 million structures represented as 28,596 assets and categorized into 128 distinct building classes at 1,588 locations.

In addition to the residential exposure, the inventory of non-residential buildings from the Hazus database was used to create exposure models of commercial and industrial structures in the San Francisco Bay Area at the census tract level.

UBC Seismic Zone (NEHRP Map Area)	Post-1975	1941 - 1975	Pre-1941
Zone 4 **(Map Area 7)**	High-Code	Moderate-Code	Pre-Code (W1 = Moderate-Code)
Zone 3 **(Map Area 6)**	Moderate-Code	Moderate-Code	Pre-Code (W1 = Moderate-Code)
Zone 2B **(Map Area 5)**	Moderate-Code	Low-Code	Pre-Code (W1 = Low-Code)
Zone 2A **(Map Area 4)**	Low-Code	Low-Code	Pre-Code (W1 = Low-Code)
Zone 1 **(Map Area 2/3)**	Low-Code	Pre-Code (W1 = Low-Code)	Pre-Code (W1 = Low-Code)
Zone 0 **(Map Area 1)**	Pre-Code (W1 = Low-Code)	Pre-Code (W1 = Low-Code)	Pre-Code (W1 = Low-Code)

Figure 29. Classification of buildings based on approximate year of construction and the 1997 Uniform Building Code seismic zone. Figure source: Hazus-MH 2.1 Technical Manual: Multi-hazard Loss Estimation Methodology — Earthquake Model.

Table 6. Description of the Hazus building typologies used in the exposure model

Taxonomy	General Typology	Description
C1H	Concrete	Concrete Moment Frame High-Rise
C1L	Concrete	Concrete Moment Frame Low-Rise
C1M	Concrete	Concrete Moment Frame Mid-Rise
C2H	Concrete	Concrete Shear Walls High-Rise
C2L	Concrete	Concrete Shear Walls Low-Rise
C2M	Concrete	Concrete Shear Walls Mid-Rise
C3H	Concrete	Concrete Frame with Unreinforced Masonry Infill Walls High-Rise
C3L	Concrete	Concrete Frame with Unreinforced Masonry Infill Walls Low-Rise
C3M	Concrete	Concrete Frame with Unreinforced Masonry Infill Walls Mid-Rise
MH	Manufactured Home	Manufactured Home
PC1	Precast Concrete	Precast Concrete Tilt-Up Walls
PC2H	Precast Concrete	Precast Concrete Frames with Concrete Shear Walls High-Rise
PC2L	Precast Concrete	Precast Concrete Frames with Concrete Shear Walls Low-Rise
PC2M	Precast Concrete	Precast Concrete Frames with Concrete Shear Walls Mid-Rise
RM1L	Reinforced Masonry	Reinforced Masonry Bearing Walls with Wood / Metal Deck Diaphragms Low-Rise
RM1M	Reinforced Masonry	Reinforced Masonry Bearing Walls with Wood / Metal Deck Diaphragms Mid-Rise
RM2H	Reinforced Masonry	Reinforced Masonry Bearing Walls with Precast Concrete Diaphragms High-Rise
RM2L	Reinforced Masonry	Reinforced Masonry Bearing Walls with Precast Concrete Diaphragms Low-Rise
RM2M	Reinforced Masonry	Reinforced Masonry Bearing Walls with Precast Concrete Diaphragms Mid-Rise
S1H	Steel	Steel Moment Frame High-Rise
S1L	Steel	Steel Moment Frame Low-Rise
S1M	Steel	Steel Moment Frame Mid-Rise
S2H	Steel	Steel Braced Frame High-Rise
S2L	Steel	Steel Braced Frame Low-Rise
S2M	Steel	Steel Braced Frame Mid-Rise
S3	Steel	Steel Light Frame
S4H	Steel	Steel Frame with Cast-in-Place Concrete Shear Walls High-Rise
S4L	Steel	Steel Frame with Cast-in-Place Concrete Shear Walls Low-Rise
S4M	Steel	Steel Frame with Cast-in-Place Concrete Shear Walls Mid-Rise
S5H	Steel	Steel Frame with Unreinforced Masonry Infill Walls High-Rise
S5L	Steel	Steel Frame with Unreinforced Masonry Infill Walls Low-Rise
S5M	Steel	Steel Frame with Unreinforced Masonry Infill Walls Mid-Rise
URML	Unreinforced Masonry	Unreinforced Masonry Bearing Walls Low-Rise
URMM	Unreinforced Masonry	Unreinforced Masonry Bearing Walls High-Rise
W1	Wood	Light Frame Wood (≤5000 sq. ft.)
W2	Wood	Commercial and Industrial Wood (>5000 sq. ft.)

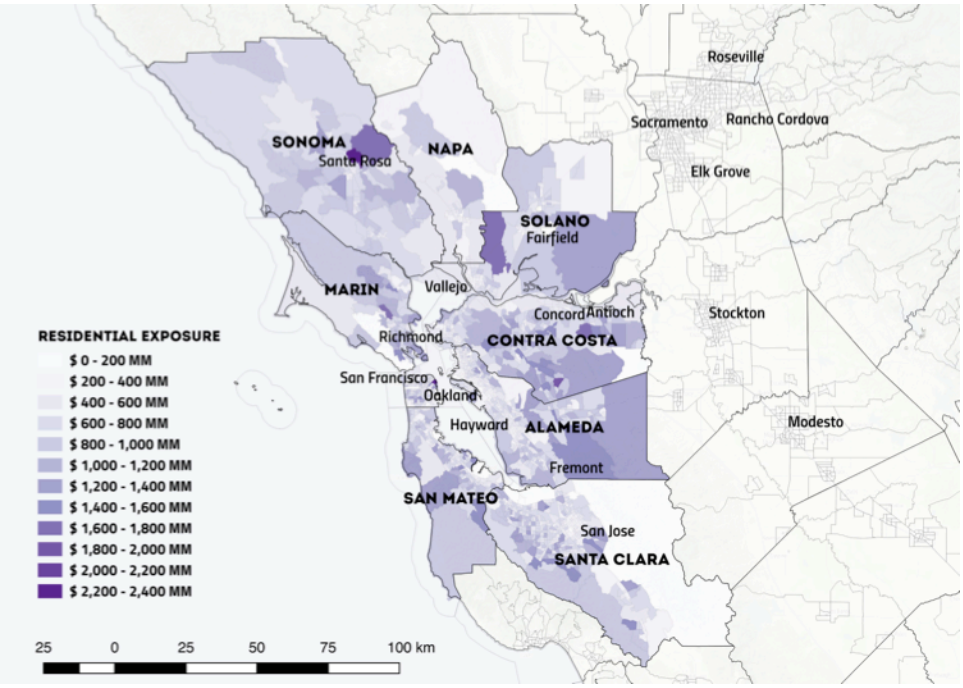


Figure 30. Residential exposure for the San Francisco Bay Area. The map displays the estimated total structural and non-structural replacement cost for residential buildings in each census tract.

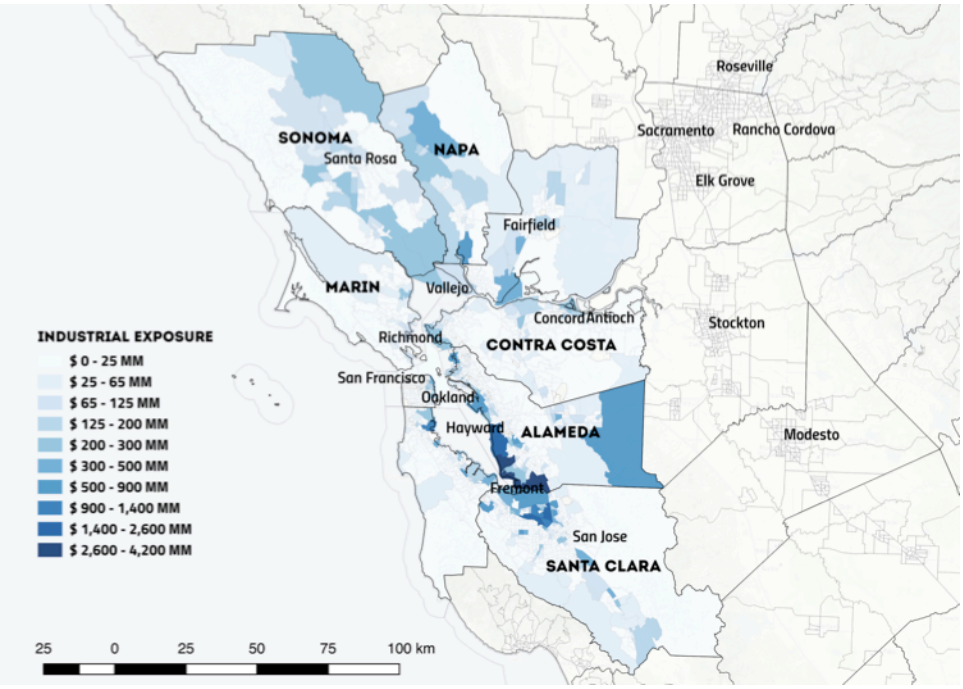


Figure 31. Industrial exposure for the San Francisco Bay Area. The map displays the estimated total structural and non-structural replacement cost for industrial buildings in each census tract.

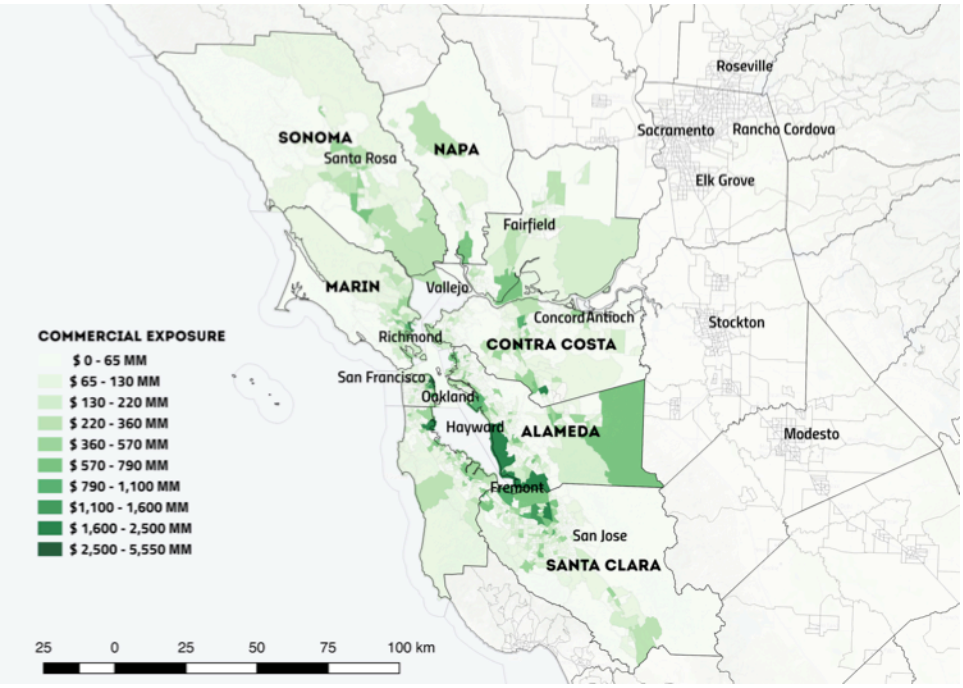


Figure 32. Commercial exposure for the San Francisco Bay Area. The map displays the estimated total structural and non-structural replacement cost for commercial buildings in each census tract.

A.5 Fragility and vulnerability models

In order to perform probabilistic or scenario damage calculations, it is necessary to define a fragility function for each building class present in the exposure model. A fragility function for a building describes the probability of exceeding a set damage states conditional on a set of ground shaking intensity levels.

For this project, the capacity curves provided by Hazus for each of the 128 building classes were converted into spectral acceleration-based continuous lognormal fragility functions suitable for use in risk analysis, using an approach similar to that described in Ryu et al. (2008). The bilinear capacity curves prescribed by Hazus were first adapted for use in nonlinear time-history analysis. Then, several single degree of freedom (SDOF) models were generated for each building class using these adapted capacity curves and a pinching model. These SDOF models were subjected to nonlinear time-history analysis using the FEMA P695 set of far-field records scaled to increasing intensity levels. Using the building response statistics from the time-history analyses and the median and dispersion of damage state thresholds from Hazus, the set of fragility functions was derived. In deriving these functions, the variability in the capacity curve representing each building class, the uncertainty in the damage state threshold, and also the record-to-record variability in the

building response have been considered. Figure 33 shows four of these 128 fragility models, the ones derived for light-frame residential wood structures.

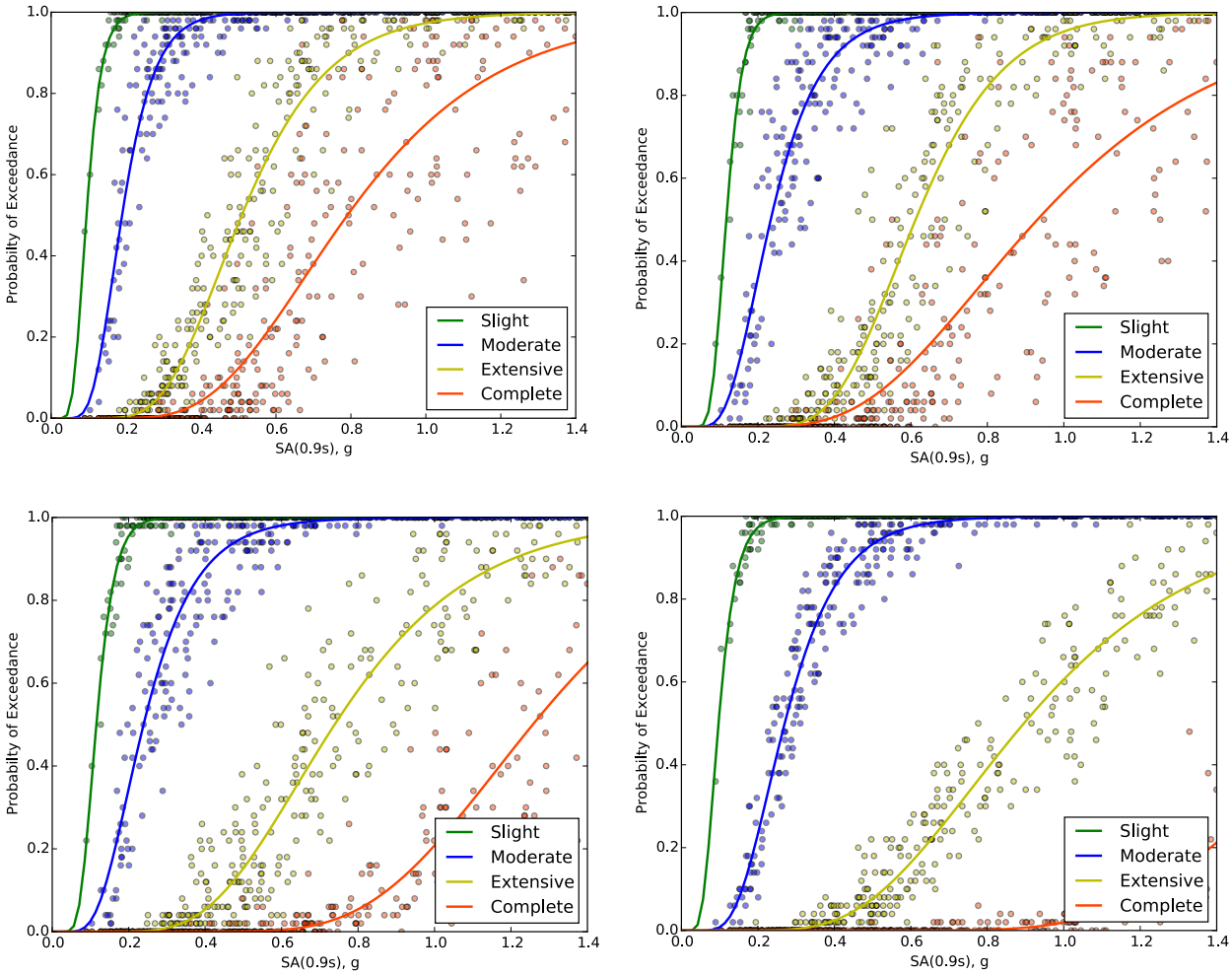


Figure 33. Fragility models derived for light-frame residential wood structures. Top-left: Pre-code structures; Top-right: Low-code structures; Bottom-left: Moderate-code structures; Bottom-right: High-code structures.

A consequence model defines a set of consequence or 'damage-to-loss" functions, describing the distribution of the loss ratio conditional on a set of discrete damage states. In the calculations described in this report, the following consequence ratios recommended by Hazus were assumed for the four damage states: Slight damage: 2%; Moderate damage: 10%; Extensive damage: 50%; Complete damage: 100%.

A vulnerability function prescribes the distribution of loss ratio conditional on the level of ground shaking. The vulnerability model used in this study uses the lognormal distribution to model the uncertainty in the loss ratios at different intensities, and the mean loss ratio

and coefficient of variation are derived based on the fragility and consequence models described above. To derive the vulnerability functions from the fragility, the Monte Carlo simulation process is continued one step further — by applying the Hazus building repair cost ratios for each damage state to the fragility functions.

Software Implementation

A.6 The OpenQuake-engine

The OpenQuake-engine is the state-of-the-practice seismic hazard and risk analysis software developed by the Global Earthquake Model (GEM) Foundation in collaboration with a worldwide community of experts on earthquake engineering and seismology. Development of the software follows the continuous-integration approach and the latest builds are always available at a public web-based repository at the following address: http://github.com/gem/oq-engine. Significant updates to the codebase are packaged and released with the corresponding documentation and user manuals on a monthly basis.

The latest version of the OpenQuake-engine (v2.1) is optimized for efficient performance on both single machines as well as computing clusters comprising several cores. All core modules and functionality of the hazard and risk calculators undergo regular and rigorous quality assurance testing. Contribution to the source code by users of the OpenQuake-engine is highly encouraged through the use of public issue-tracking, pull-requests, and code-reviews.

A.6.1 The OpenQuake-engine risk calculators

The risk component of the OpenQuake-engine can compute both scenario-based and probabilistic seismic damage and risk using various approaches. The following types of analysis are currently supported:

- ***Scenario Damage and Loss Assessment***, for the calculation damage statistics, spatial distribution of damage, or for the calculation of individual asset and portfolio loss statistics from a single earthquake rupture scenario taking into account aleatory and epistemic ground-motion variability.
- ***Classical Probabilistic Seismic Damage and Loss Analysis***, for the calculation of damage state probabilities over a specified time period and probabilistic collapse maps, or for the calculation of loss curves and loss maps, starting from the hazard curves computed following the classical integration procedure. A probabilistic retrofit cost-benefit analysis calculator is also included within the risk component of the OpenQuake-engine.
- ***Event-based Probabilistic Seismic Risk Analysis***, for the calculation of event loss tables starting from stochastic event sets. Other results such as portfolio loss-exceedance curves, probabilistic loss maps, average annual losses, and insured loss statistics can be obtained by post-processing the event loss tables.

A.7 Implementation of the UCERF3 calculators

In addition to these calculators described above, the OpenQuake-engine now includes two new hazard and risk calculators designed specifically to run computations involving the full 7,200 logic-tree branches of the latest hazard model for California. The implementation of these new calculators is described briefly in the subsections below.

A.7.1 New Scientific Features Required for the OpenQuake-engine

Before addressing the new additions and improvements made to the specific UCERF3 calculators, the implementation of the UCERF3 required the addition of several new features in the OpenQuake-engine that have a broader application beyond California and the Western United States. The first of these is the comprehensive implementation of the NGA West 2 ground motion models described in section A.2. These models represent the current state-of-the-art in ground motion modeling, in some cases requiring new parameters describing the properties of the rupture and the source to site distance configuration. Several of the NGA West 2 ground motion models (Abrahamson et al., 2014; Campbell & Bozorgnia, 2014; Chiou & Youngs, 2014) adopt relatively complex equations based on the rupture geometry and source-to-site-distance to characterize the amplification observed at sites very close to the source (so called 'hanging wall effects"). Not only can these models of near-fault amplification have a large influence on the resulting hazard in urban areas close to major active faults, they require a comprehensive description of the finite source geometry that must be applicable to all of the different typologies of finite-rupture that OpenQuake-engine is capable of generating.

As described in section A.1, ruptures occurring on defined active faults (the 'supraseismogenic" sources) are characterized as a set of rectangular planar segments, mostly on the order of approximately 7 to 8 km in length and between 15 to 20 km in down-dip width (shorter segments can be seen in some places). The UCERF3 ruptures introduce an additional complication as these sets of segments are not always continuous (i.e. ruptures can 'jump" over small gaps and discontinuities in the fault) and can, within reason, change the direction of strike. These are referred to as 'discontinuous" and (if the direction of the segments changes significantly) 'discordant" ruptures. Certain metrics of source to site distance are required for the NGA West 2 models that measure the distance not just from the site to the rupture, but also the distances to the up-dip projection of the rupture to the surface measured both perpendicular to the strike (azimuth) of the rupture (R_X) and parallel to the strike of the rupture (R_{y0}). These are illustrated conceptually in Figure 34. When ruptures are discontinuous and discordant these metrics can become complex to determine and may lead to undesirable behavior in the ground motions. To overcome this, Spudich & Chiou (2015) introduced a coordinate system that defines the strike-parallel and strike-normal distances for such rupture types. This is known as the

Generalized Coordinate System 2 (or GC2). An illustration of the GC2 system for a complex rupture (the 2010 El Mayor-Cucapah earthquake) is shown in Figure 35. Incorporation of this system into the OpenQuake-engine for application to the UCERF3 ruptures was a key new development of the work that will potentially allow rupture characterizations similar to those of UCERF3 to be adopted in other regions of the globe.

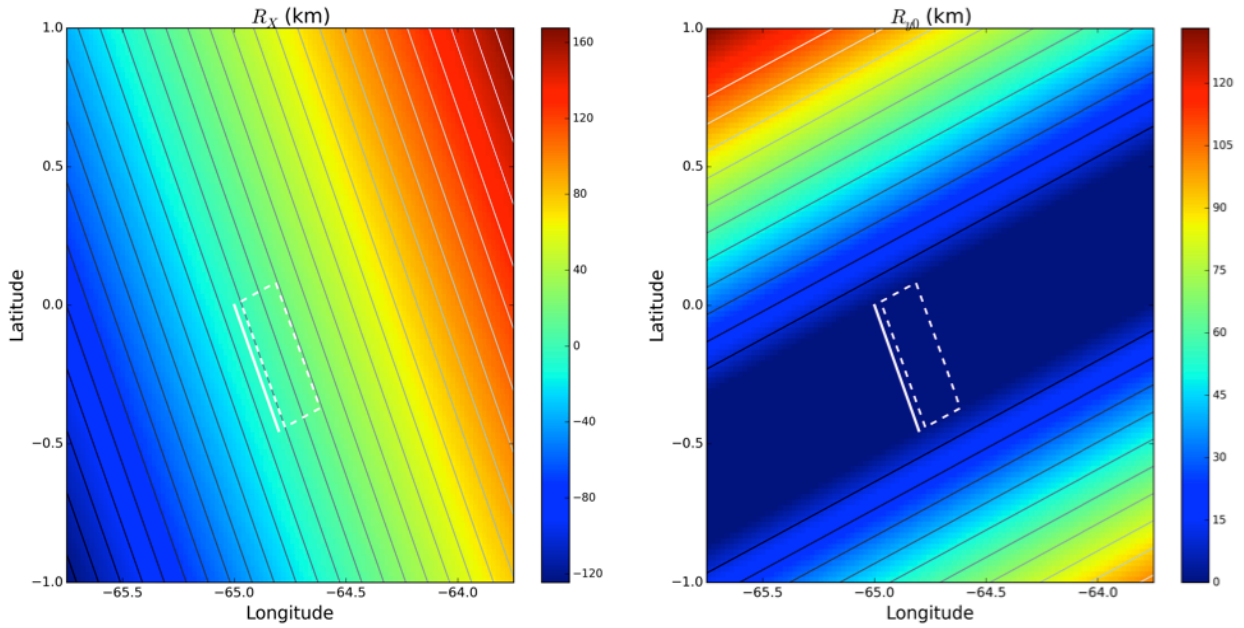


Figure 34. Illustration of near-fault strike-perpendicular distances R_x (left) and strike-parallel distances (R_{y0}) for an idealised single rupture plane indicated by its vertical surface projection (dashed white line) and up-dip projection to the surface (thick, continuous white line)

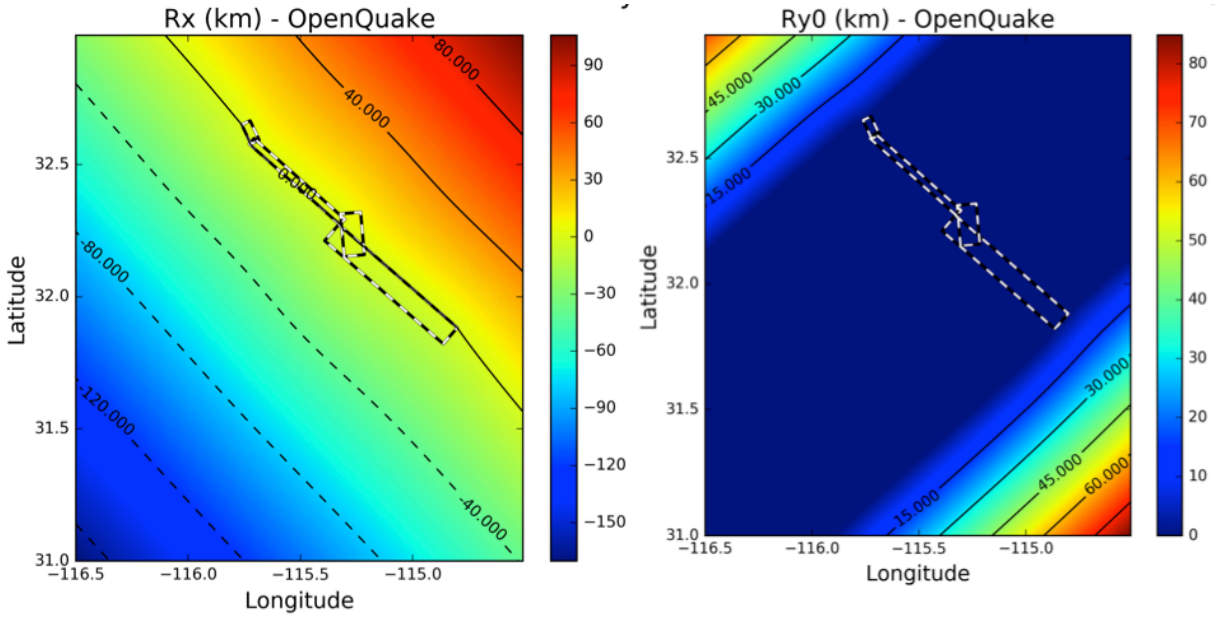


Figure 35. Illustration of near-fault strike-perpendicular distances R_x (left) and strike-parallel distances (R_{y0}) for a the discordant and discontinuous El-Mayor-Cucapah rupture plane (Ancheta et al., 2014)

A.7.2 The UCERF3 ‘classical” PSHA calculator

The ‘classical” PSHA calculator implements the method of Cornell (1965) and McGuire (1976), adapted by Field et al. (2003), to determine the probability of exceeding specific levels of ground motion given the set of earthquake ruptures and their probabilities of occurrence within a specified time-span. This method requires the enumeration of every earthquake within the earthquake rupture forecast, including those occurring in the gridded background. For each rupture the probability of exceeding a given level of ground motion is calculated for each site using the ground motion model. In more conventional seismic hazard models, the earthquake rupture forecast for a given seismogenic source is constructed within the earthquake software, with the user required to specify parameters that allow the software to construct the rupture set. These include the source geometry, the magnitude frequency distribution (i.e. the relation between the magnitudes the source can generate and the rate of occurrence), the magnitude-scaling relationship and several parameters controlling the dimensions of the ruptures generated in the source.

Adaptation of OpenQuake-engine to implement the UCERF3 calculations presents some challenges, primarily from a technological perspective but also from a scientific perspective. The main adaptation required is to enable the software to build the earthquake rupture forecast directly from the input file, rather than allowing the software to build it from a simpler parameterization of the seismogenic source. Furthermore, the large number of logic tree branches prohibits the direct construction of the comprehensive earthquake rupture forecast for each source model, which would require 1,440 individual source model files each likely exceeding 1 gigabyte in size.

Instead, the largest technological change in the UCERF calculator is in the input model definition and the way OpenQuake-engine builds the earthquake rupture forecast from the information provided. For this purpose, the complete UCERF source model is merged into a single high-density binary (HDF5) file. From here the user needs only to input a logic tree branch identification code to enable OpenQuake-engine to retrieve the necessary information from the binaries within the calculation. Effectively this means that unlike the conventional ‘classical” PSHA calculators that will build the earthquake rupture forecast from the source at the beginning of the calculation before analyzing them in turn, the UCERF3 calculator can determine which of the ruptures may be required in the calculation before constructing the rupture geometry and estimating the probability of exceeding the given level of ground motion.

The adaptation described above has several critical advantages that can help to make a calculation of this size efficient to run on a smaller scale distributed computing system.

The primary advantage is that at any time in the calculation the amount of Random Access Memory (RAM) required is minimal. This means that the calculations are well suited to computing environments with many Central Processing Units (CPUs), such as server or cloud environments, and even smaller scale calculations become feasible on an ordinary desktop. Furthermore, OpenQuake-engine may be able to identify a suitable strategy to parallelize the calculations depending on the user's requirements. For example, where the user may wish to run many logic tree branches, it may be more efficient for the engine to parallelize the calculation by logic tree branch. If that is not the case, perhaps when the user wishes to analyze only a single branch, then OpenQuake-engine can parallelize the calculation by groups of ruptures. Both strategies help to ensure that the usage of the available computational resources can be maximized regardless of the type of calculation.

One critical aspect of the UCERF3 hazard calculations often overlooked is the means by which distances to the finite rupture plane are determined for the gridded seismicity sources. It should be recalled that the gridded seismicity model simply defines an activity rate per 0.1˚ by 0.1˚ cell across California. As described previously, however, the ground motion models require the definition of the complete three-dimensional geometry of the rupture, which due to the interactions of the complex near-fault amplifications cannot be inferred from the distance to the centroid of the cell, nor can they be taken as an 'average" of the distances to a randomly oriented rupture plane, as is the practice in the USGS national seismic hazard map (Petersen et al., 2008; 2014). Therefore, it is necessary for the user to define a set of constraining properties that would allow for finite rupture planes to be constructed in a manner that is consistent with the tectonic environment of the region. More detail on this issue, including an illustration of the resulting impact on the hazard assessment, can be found in Pagani et al. (2014) and Monelli et al. (2014).

Table 7. Recommended user configuration parameters for the UCERF3 gridded seismicity sources in OpenQuake-engine

Gridded Seismicity Configuration Parameter	**Adopted Value**
Upper Seismogenic Depth	0.0 (km) – the free surface
Lower Seismogenic Depth	15.0 km
Aspect Ratio	1.5
Magnitude-Area Scaling Relationship	Wells & Coppersmith (1994)
Rupture Orientation Distribution (strike, dip, rake, probability)	(0.0˚, 90.0˚, 0.0˚, 0.15) (45.0˚, 90.0˚, 0.0˚, 0.15) (90.0˚, 90.0˚, 0.0˚, 0.15) (135.0˚, 90.0˚, 0.0˚, 0.15) (0.0˚, 45.0˚, 90.0˚, 0.1) (90.0˚, 45.0˚, 90.0˚, 0.1) (180.0˚, 45.0˚, 90.0˚, 0.1)

	(270.0°, 45.0°, 90.0°, 0.1)
Hypocentral Depth Distribution (depth, probability)	(6.0 km, 1.0)
Minimum Magnitude	5.0

The additional information that is required are the constraints on the upper and lower seismogenic depth (i.e. the depth limits above and below which ruptures cannot propagate), the aspect ratio (i.e. the ratio of length to width) of the rupture, area to magnitude-scaling relationship for the gridded seismicity sources, the distribution of possible rupture orientations (as a set of azimuths, dips and slip directions [*rake*] with their associated probabilities) and the possible depths of the rupture hypocentres. It may be surprising to learn that it is this process, and the assumptions about the physical nature of the system made therein, that can account for differences between PSHA software codes when given the same input model. The user of the software can configure these parameters, and for many of the present calculations the configuration settings are as described in Table 7.

To test the implementation of the UCERF3 classical calculator we run the 'true mean" model with the complete GMM logic tree, and compared against the hazard published by the US Geological Survey for specific locations in California: Los Angeles (Downtown), San Francisco (Downtown), San Diego, and Sacramento (all shown in Figure 36). Generally, the OpenQuake-engine results and those published by the US Geological Survey are in good agreement for the typical range of return periods considered for seismic design (between 100 and 2,500 years). Three of the sites are close to the major faults with higher slip rates (Los Angeles, San Francisco, and San Diego), whilst Sacramento is located further inland at a greater distance from the most active faults. Where the sites are located away from the fault zones it is the gridded seismicity that may control the hazard more than the faults. In these regions, at low annual probabilities of exceedance the assumptions regarding the characterization of the finite ruptures in the grid cells begin to be significant as they strongly influence the source to site distances for large earthquakes generated in these regions. The systematically higher ground motions for lower annual probabilities predicted by OpenQuake-engine have been explored previously by Monelli et al. (2014) and these observations are found to persist across the entire US National Seismic Hazard Model.

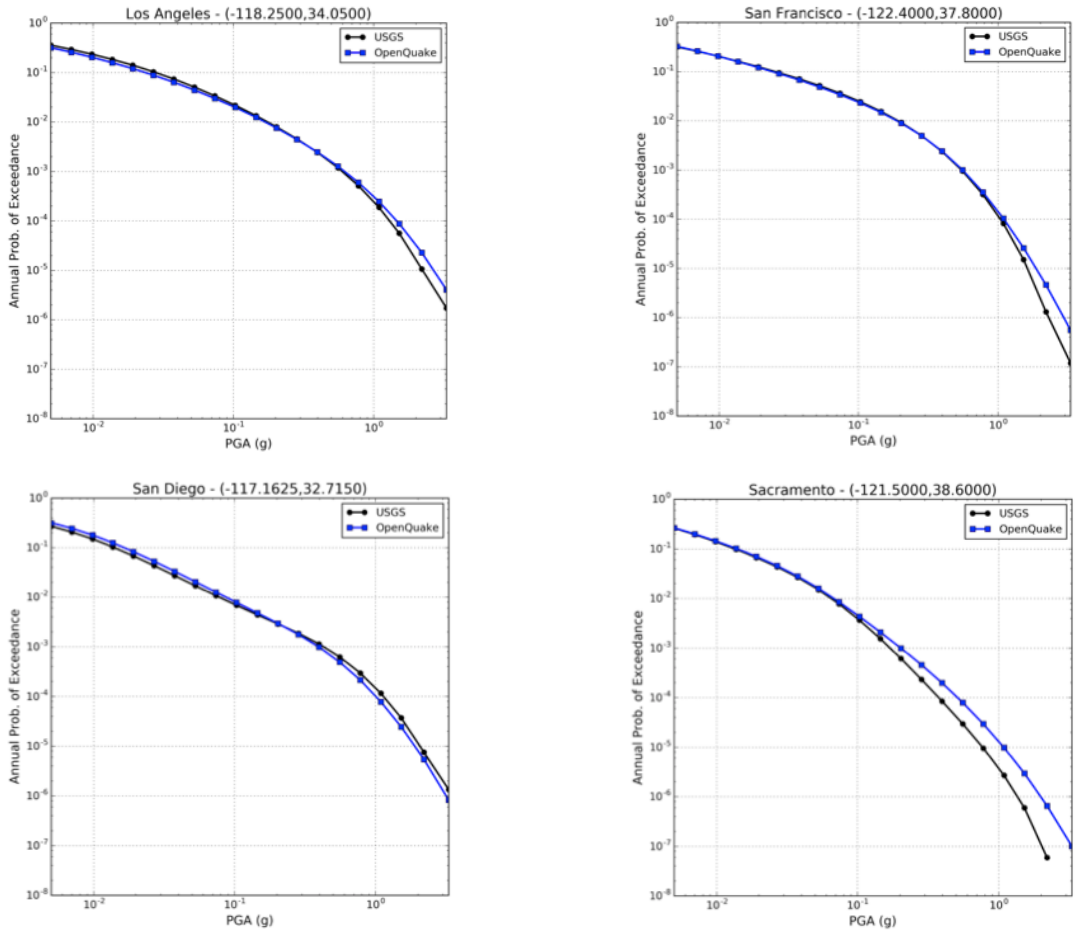


Figure 36. Comparisons of the hazard curves for Peak Ground Acceleration (PGA) between OpenQuake (blue) and the USGS published values (black) for: Los Angeles (top left), San Francisco (top right), San Diego (bottom left), and Sacramento (bottom right)

A.7.3 The UCERF3 stochastic event-based risk calculator

The other of OpenQuake-engine's two primary probabilistic hazard and risk calculators is the event-based module. Unlike the classical calculator, this works not by enumerating each rupture in the earthquake rupture forecast and calculating the probability of exceeding a given level of ground motion within the specified investigation time, T, but instead generates synthetic catalogues of earthquakes of that same duration by randomly sampling the number of occurrences within time T of each earthquake in the earthquake rupture forecast. When the investigation time T is sufficiently large and/or the number of synthetic catalogues is high, both the classical and stochastic event-based approaches should produce equivalent results. Whilst the event-based approaches may be less efficient for estimating the seismic hazard, they offer the possibility to incorporate

features such as spatial correlation in the ground motion model, which has a strong influence when assessing the seismic risk to a spatially distributed portfolio of assets.

As with conventional seismic hazard models, OpenQuake-engine ensures that the source model and ground motion model inputs are interchangeable both for the event-based and classical hazard approaches. With the UCERF3 calculators this is also a key requirement. There is one critical adaptation in the event-based calculator for the UCERF3 calculation that improves the efficiency of this calculator in comparison to the classical PSHA approach. Normally, when a seismogenic source is input into the stochastic event-based calculation, OpenQuake-engine still builds the complete earthquake rupture forecast, which is subsequently sampled. In the case of UCERF3, it would be impossible to build the forecast and store it in memory, but as the rates of occurrence of each rupture are stored explicitly in the source model, it is possible to apply the sampling for each event set based only on the annual rates of occurrence. Whilst the probability of occurrence of any individual rupture in the ERF is significantly lower than 1, only a small subset of the possible ruptures will be found in each event set. This means that it is known which ruptures occur and how often in each event set before the actual physical ruptures are built. Thus the process of generating the event sets becomes simpler, and the calculation effort mostly correlates linearly with the investigation time and number of event sets.

Table 8[30]. Partial view of the average annual loss table for the residential exposure in the San Francisco Bay Area for all branches of the 2014 NSHM for California

SM	FM	DM	SR	DSR	M5	MMAX	SPATIALPDF	GSIM	WEIGHT	LOSS
Wald and Allen (2007)	FM3_1	ABM	Shaw09Mod	DsrUni	M5Rate6.5	MMaxOff7.3	SpatSeisU2	AbrahamsonEtAl2014	0.000005	5.21E+12
								BooreEtAl2014	0.000005	5.33E+12
								CampbellBozorgnia2014	0.000005	5.48E+12
								ChiouYoungs2014	0.000005	5.96E+12
								Idriss2014	0.000003	5.52E+12
							SpatSeisU3	AbrahamsonEtAl2014	0.000005	5.29E+12
								BooreEtAl2014	0.000005	5.33E+12
								CampbellBozorgnia2014	0.000005	5.55E+12
								ChiouYoungs2014	0.000005	6.12E+12
								Idriss2014	0.000003	5.71E+12
						MMaxOff7.6	SpatSeisU2	AbrahamsonEtAl2014	0.000044	4.65E+12
								BooreEtAl2014	0.000044	4.80E+12
								CampbellBozorgnia2014	0.000044	4.98E+12
								ChiouYoungs2014	0.000044	5.38E+12
								Idriss2014	0.000024	5.13E+12
							SpatSeisU3	AbrahamsonEtAl2014	0.000044	5.46E+12
								BooreEtAl2014	0.000044	5.53E+12
								CampbellBozorgnia2014	0.000044	5.75E+12
								ChiouYoungs2014	0.000044	6.36E+12
								Idriss2014	0.000024	5.96E+12
						MMaxOff7.9	SpatSeisU2	AbrahamsonEtAl2014	0.000005	5.60E+12
								BooreEtAl2014	0.000005	5.76E+12
								CampbellBozorgnia2014	0.000005	5.90E+12
								ChiouYoungs2014	0.000005	6.53E+12
								Idriss2014	0.000003	6.07E+12
							SpatSeisU3	AbrahamsonEtAl2014	0.000005	5.19E+12
								BooreEtAl2014	0.000005	5.18E+12
								CampbellBozorgnia2014	0.000005	5.40E+12
								ChiouYoungs2014	0.000005	6.03E+12
								Idriss2014	0.000003	5.67E+12

Using the OpenQuake-engine

A.8 Installing the OpenQuake-engine

The OpenQuake-engine is available for Windows, Linux, and MacOS machines. Installation instructions for Windows and MacOS machines are provided below. For the installation instructions for Linux machines, please consult https://github.com/gem/oq-engine/blob/engine-2.3/doc/installing/linux-generic.md.

A.8.1 Windows

The requirements for installing the OpenQuake-engine on a Windows machine are:

- Windows 7 (64bit) or Windows 8 / 8.1 (64bit) or Windows 10 (64bit)
- Minimum 4 GB RAM (at least 8 GB is recommended)
- 1.2 GB of free disk space

[30] Abbreviations used in Table 8. SM: Site-conditions model; FM: Fault model; DM: Deformation model; SR: Magnitude scaling relationship; DSR: Slip along rupture; M5: Total rate of M≥5 events in the region; MMAX: Maximum magnitude for events occurring off the modeled faults; SPATIALPDF: Off-fault spatial seismicity PDF; GSIM: Ground motion simulation model

Download the installer from http://www.globalquakemodel.org/pkgs/windows/oq-engine/OpenQuake_Engine_2.3.0-1.exe using any browser and run the installer, then follow the instructions provided by the wizard on the screen. The screenshots shown in Figure 37 illustrate the steps involved in the installation.

A.8.2 MacOS

The requirements for installing the OpenQuake-engine on a MacOS machine are:

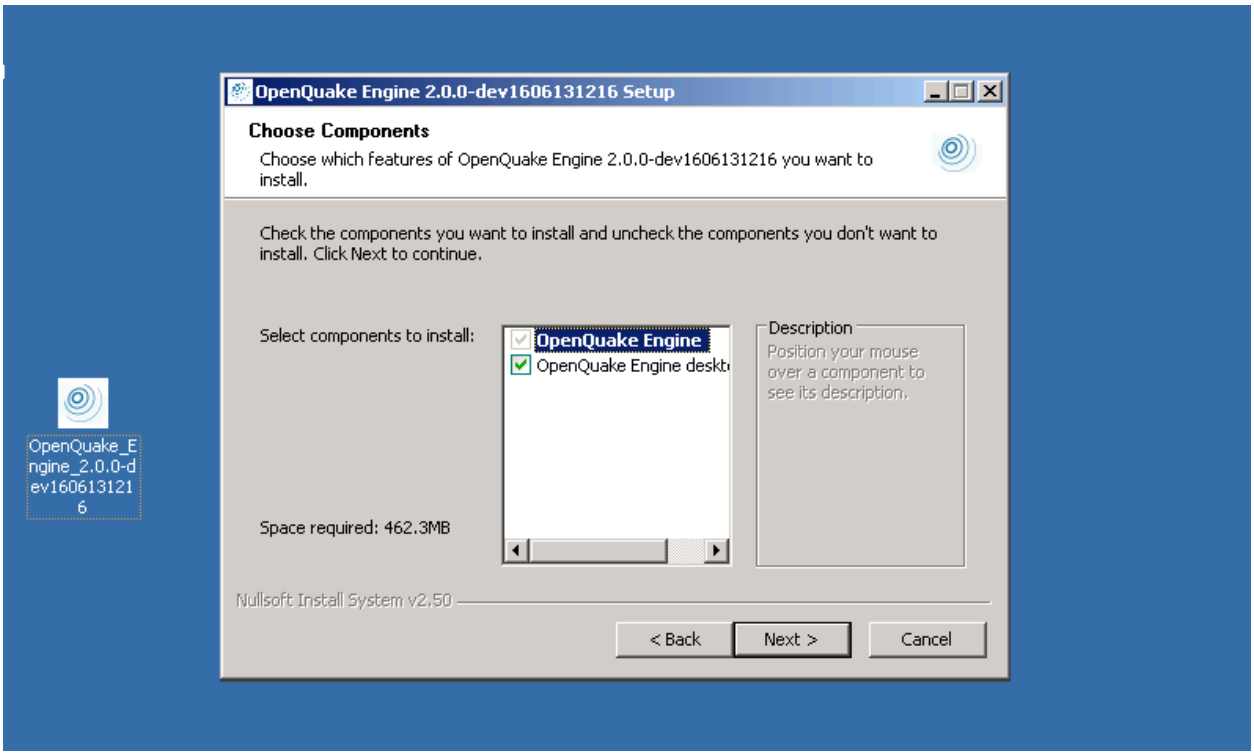
- Mac OS X 10.10 (Yosemite) or Mac OS X 10.11 (El Capitan) or MacOS 10.12 (Sierra)
- Minimum 4 GB RAM (at least 8 GB is recommended)
- 1.2 GB of free disk space
- Terminal or iTerm app

Download the installer from http://www.globalquakemodel.org/pkgs/macos/oq-engine/openquake-setup-macos-2.3.0-1.run using any browser.

From the Terminal app (or using iTerm) run

```
cd Downloads
chmod +x openquake-setup-macos-2.3.0-1.run
./openquake-setup-macos-2.3.0-1.run
```

then follow the wizard on screen. By default the code is installed in ~/openquake.



(Caption on the next page)

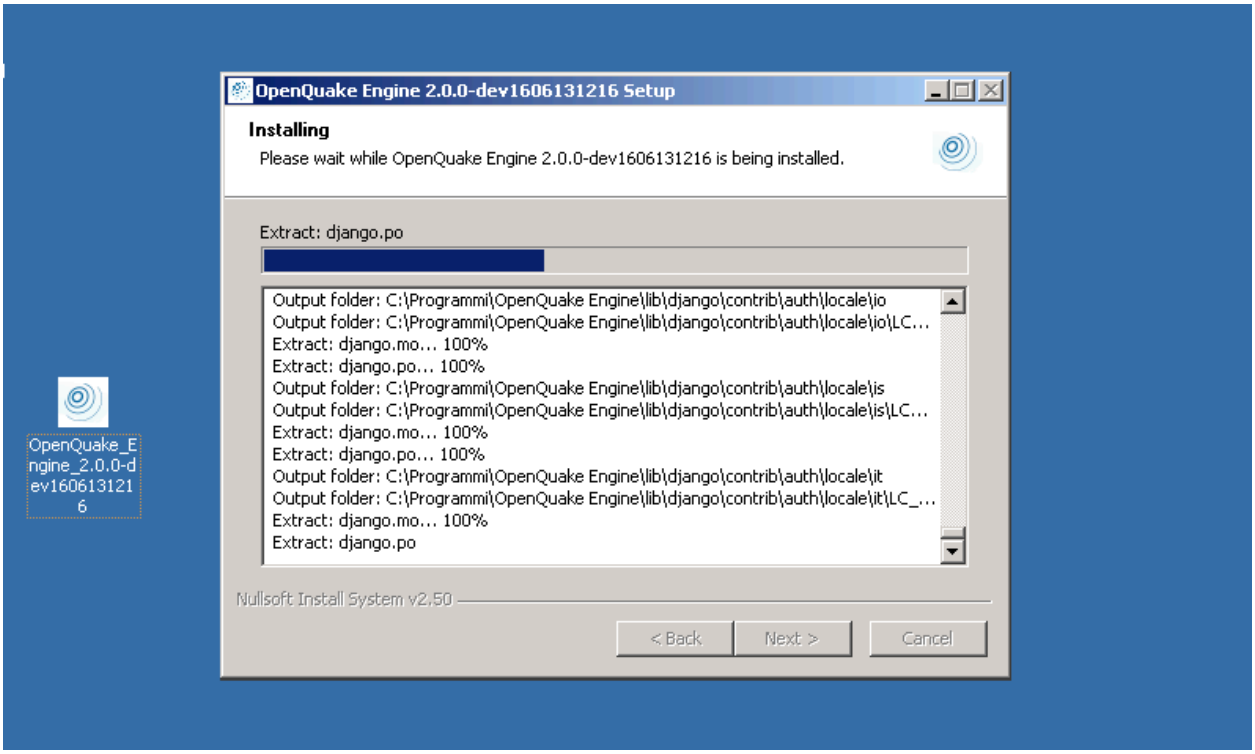


Figure 37. Screenshots of the OpenQuake-engine installation steps on Windows

A.9 Running the OpenQuake-engine

The OpenQuake-engine installation provides a browser-based user interface ('WebUI') for running seismic hazard and risk calculations and for downloading the outputs. The OpenQuake Engine WebUI does not require any special software on the client side except for a browser. The browsers currently supported are:

- Mozilla Firefox ≥ 38
- Google Chrome
- Microsoft Internet Explorer ≥ 10
- Microsoft Edge
- Apple Safari ≥ 6

To start the WebUI, open a browser and load the http://localhost:8800/engine page

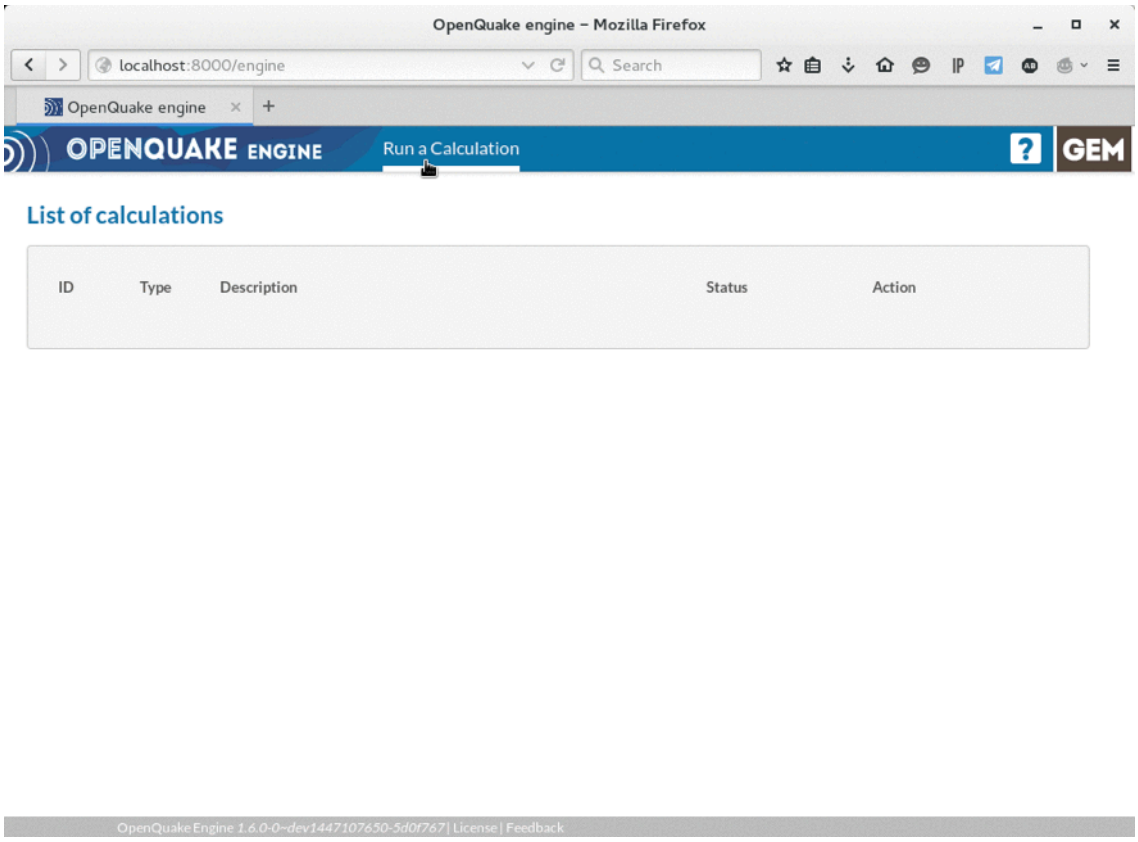


Figure 38. The OpenQuake-engine WebUI: Start page

Before starting a new calculation you need to prepare the input files. All the inputs files must be placed in a compressed zip file. To start a calculation select the **Run a Calculation** button in the header menu. A dialog window will appear: select the zip file to be uploaded and press Open to confirm.

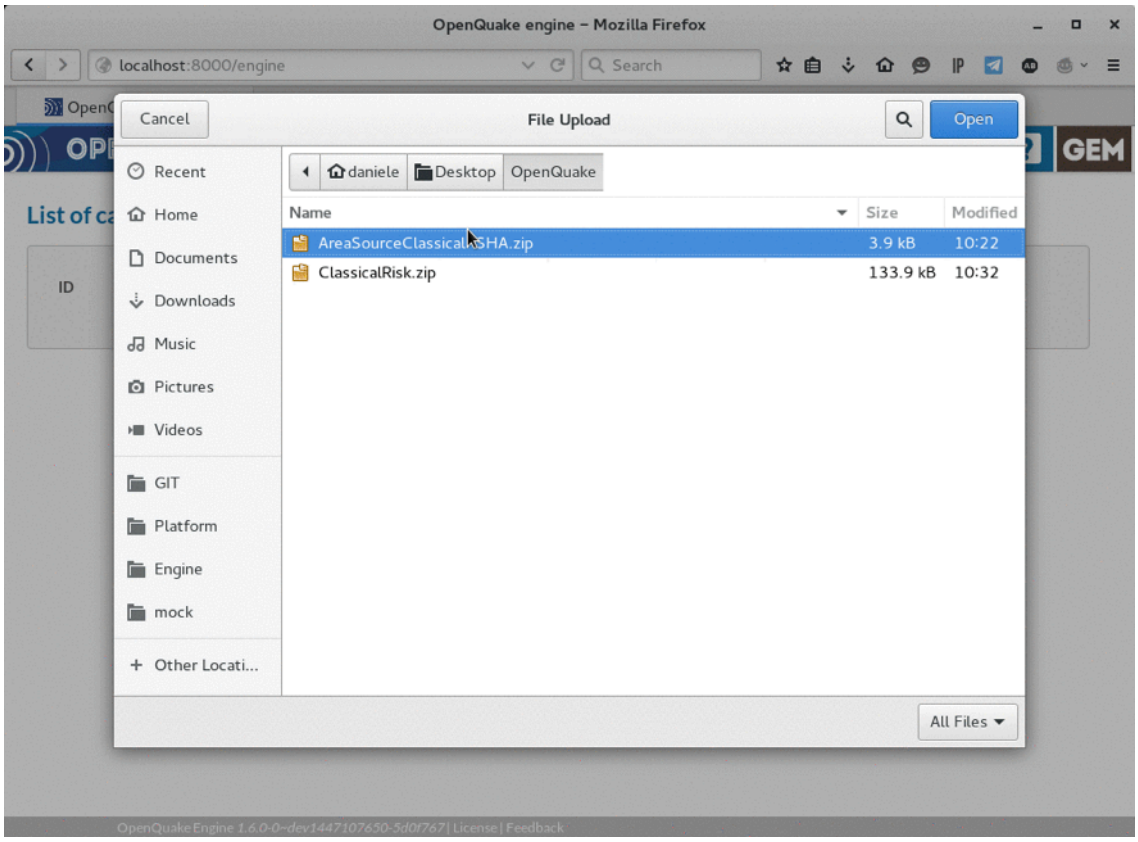


Figure 39. The OpenQuake-engine WebUI: Selecting a calculation to run

If a calculation has run successfully an **Outputs** button is provided. Click the **Outputs** button to see the list of outputs available for the specific calculation type. Output availability depends on the type of calculation run and the types of outputs requested in the configuration file for that calculation. Click on the desired output/format to download the corresponding file.

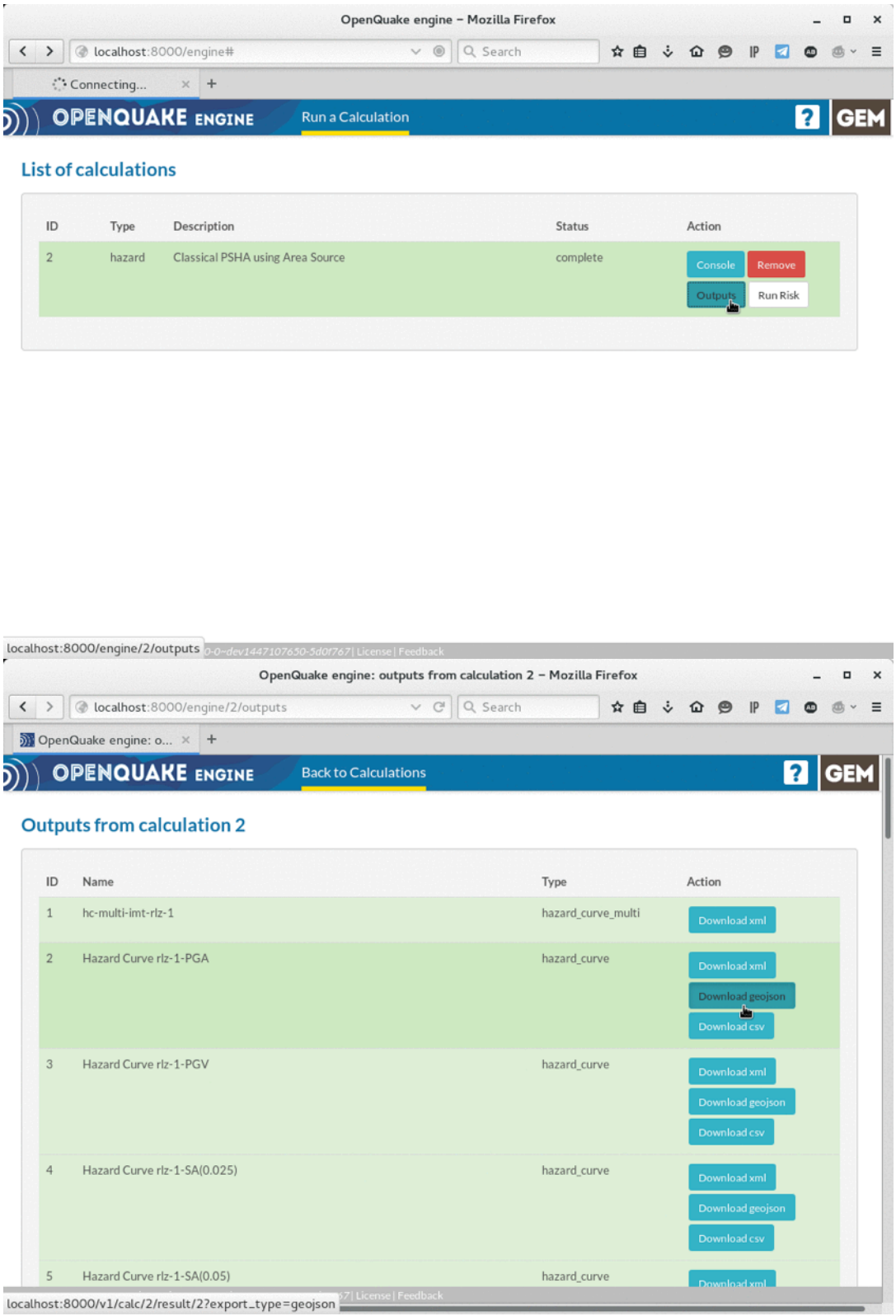


Figure 40. The OpenQuake-engine WebUI: Downloading outputs from a calculation

Please consult the OpenQuake-engine user manual for more details about the different calculators available within the engine, and instructions on how to use them. The user manual for the latest stable version of the OpenQuake-engine is always available at: https://globalquakemodel.org/openquake/support/documentation/engine

References

Abrahamson, N.A., Silva, W.J., Kamai, R., 2014. Summary of the ASK14 ground motion relation for active crustal regions. *Earthquake Spectra,* 30, pp.1025–1055.

Applied Technology Council, 2009. *FEMA P695: Quantification of Building Seismic Performance Factors*, Washington, DC.

Boore, D.M., Stewart, J.P., Seyhan, E., Atkinson, G.M., 2014. NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes. *Earthquake Spectra*, 30, pp.1057–1085.

Campbell, K.W., Bozorgnia, Y., 2014. NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra. *Earthquake Spectra*, 30, pp.1087–1115.

Chiou, B.S.-J., Youngs, R.R., 2014. Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra. *Earthquake Spectra,* 30, pp.1117–1153.

Eads, L., Miranda, E., Krawinkler, H. & Lignos, D. G., 2013. An efficient method for estimating the collapse risk of structures in seismic regions. *Earthquake Engineering & Structural Dynamics* 42**,** 25–41.

Federal Emergency Management Agency, 2012. *Hazus-MH 2.1 Technical Manual: Multi-hazard Loss Estimation Methodology: Earthquake Model*, Washington, DC.

Federal Emergency Management Agency, 2012. *Hazus-MH 2.1 User Manual. Multi-hazard Loss Estimation Methodology: Earthquake Model*, Washington, DC.

Field, E. H. Biasi, G. P., Bird, P., Dawson, T. E., Felzer, K. R., Jackson, D. D., Johnson, K. M., Jordan, T. H., Madden, C., Michael, A. J., Milner, K. R., Page, M. T., Parsons, T., Powers, P. M., Shaw, B. E., Thatcher, W. R., Weldon, R. J., Zeng, Y., 2013. Uniform California Earthquake Rupture Forecast, Version 3 (UCERF3) – The Time-Independent Model, *U. S. Geological Survey Open File Report 2013-1165*, 97p.

Field, E.H., Arrowsmith, R.J., Biasi, G.P., Bird, P., Dawson, T.E., Felzer, K.R., Jackson, D.D., Johnson, K.M., Jordan, T.H., Madden, C., Michael, A.J., Milner, K.R., Page, M.T., Parsons, T., Powers, P.M., Shaw, B.E., Thatcher, W.R., Weldon, R.J., Zeng, Y., 2014. Uniform California Earthquake Rupture Forecast, Version 3 (UCERF3) – The Time-Independent Model. *Bulletin of the Seismological Society of America*, 104, pp.1122–1180.

Field, E.H., Dawson, T.E., Felzer, K.R., Frankel, A.D., Gupta, V., Jordan, T.H., Parsons, T., Petersen, M.D., Stein, R.S., Weldon, R.J., Wills, C.J., 2009. Uniform California Earthquake Rupture Forecast, Version 2 (UCERF 2). *Bulletin of the Seismological Society of America*, 99, pp.2053–2107.

Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., and Megown, K., 2015, Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover

change information. *Photogrammetric Engineering and Remote Sensing*, v. 81, no. 5, pp. 345-354.

Idriss, I.M., 2014. An NGA-West2 empirical model for estimating the horizontal spectral values generated by shallow crustal earthquakes. *Earthquake Spectra*, 30, pp.1155–1177.

Jones, L.M., Bernknopf, R., Cox, D., Goltz, J., Hudnut, K., Mileti, D., Perry, S., Ponti, D., Porter, K., Reichle, M., Seligson, H., Shoaf, K., Treiman, J., and Wein, A., 2008, The ShakeOut Scenario: U.S. Geological Survey Open- File Report 2008-1150 [http://pubs.usgs.gov/of/2008/1150/]

Monelli, D., Pagani, M., Weatherill, G., Danciu, L., Garcia, J., 2014. Modeling Distributed Seismicity for Probabilistic Seismic-Hazard Analysis: Implementation and Insights with the OpenQuake Engine. *Bulletin of the Seismological Society of America*, 104, pp.1636–1649.

Pagani, M., Monelli, D., Weatherill, G., Danciu, L., Crowley, H., Silva, V., Henshaw, P., Butler, L., Nastasi, M., Panzeri, L., Simionato, M., Vigano, D., 2014. OpenQuake Engine: An Open Hazard (and Risk) Software for the Global Earthquake Model. *Seismological Research Letters*, 85, pp.692–702.

Page, M.T., Field, E.H., Milner, K.R., Powers, P.M., 2014. The UCERF3 Grand Inversion: Solving for the Long-Term Rate of Ruptures in a Fault System. *Bulletin of the Seismological Society of America*, 104, 1181–1204.

Petersen, M.D., Frankel, A.D., Harmsen, S.C., Mueller, C.S., Haller, K.M., Wheeler, R.L., Wesson, R.L., Zeng, Y., Boyd, O.S., Perkins, D.M., Luco, N., Field, E.H., Wills, C.J., Rukstales, K.S., 2008. Documentation for the 2008 Update of the United States National Seismic Hazard Maps (USGS Open File Report No. OFR 2008-1128). U.S. Geological Survey, Reston, Virginia.

Petersen, M.D., Moschetti, M.P., Powers, P.M., Mueller, C.S., Haller, K.M., Frankel, A.D., Zeng, Yuehua, Rezaeian, Sanaz, Harmsen, S.C., Boyd, O.S., Field, Ned, Chen, Rui, Rukstales, K.S., Luco, Nico, Wheeler, R.L., Williams, R.A., and Olsen, A.H., 2014, Documentation for the 2014 update of the United States national seismic hazard maps: U.S. Geological Survey Open-File Report 2014–1091, 243 p.

Rezaeian, S., Petersen, M.D. & Moschetti, M.P., 2015. Ground Motion Models Used in the 2014 U.S. National Seismic Hazard Maps. Earthquake Spectra.

Ryu, H., Luco, N., Baker, J. W. & Karaca, E., 2008. Converting Hazus capacity curves to seismic hazard compatible building fragility functions: Effect of hysteretic models. In The 14th World Conference on Earthquake Engineering. Beijing, China, pp. 1–8.

Spudich, P., Chiou, B., 2015. Strike-Parallel and Strike-Normal Coordinate System Around Geometrically Complicated Rupture Traces – Use by NGA-West2 and Further Improvements, U. S. Geological Survey Open-File Report 2015-1028, p20.

U.S. Census Bureau, 2015. Selected housing characteristics, 2011-2015 American Community Survey 5-year estimates. Retrieved from American FactFinder:

http://factfinder2.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_15_5YR_DP04.

Vekshin, Alison. "California Overtakes France to Become Sixth-Largest Economy." http://www.bloomberg.com/politics/articles/2016-06-14/california-overtakes-france-to-become-sixth-largest-economy. Bloomberg, Published: 15 June 2016. Accessed: 01 December 2016.

Wald, D.J. & Allen, T.I., 2007. Topographic Slope as a Proxy for Seismic Site Conditions and Amplification. Bulletin of the Seismological Society of America, 97(5), pp.1379–1395.

Wills, C. J., Gutierrez, C. I., Perez, F. G. & Branum, D. M., 2015. A Next Generation VS30 Map for California Based on Geology and Topography. Bulletin of the Seismological Society of America, 105(6), pp.3083–3091.

The Working Group On California Earthquake Probabilities, 2013. The Uniform California Earthquake Rupture Forecast, Version 3 (UCERF3)—The Time-Independent Model, Reston, VA.