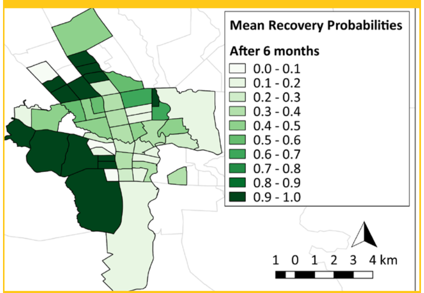
BACK TO NORMAL

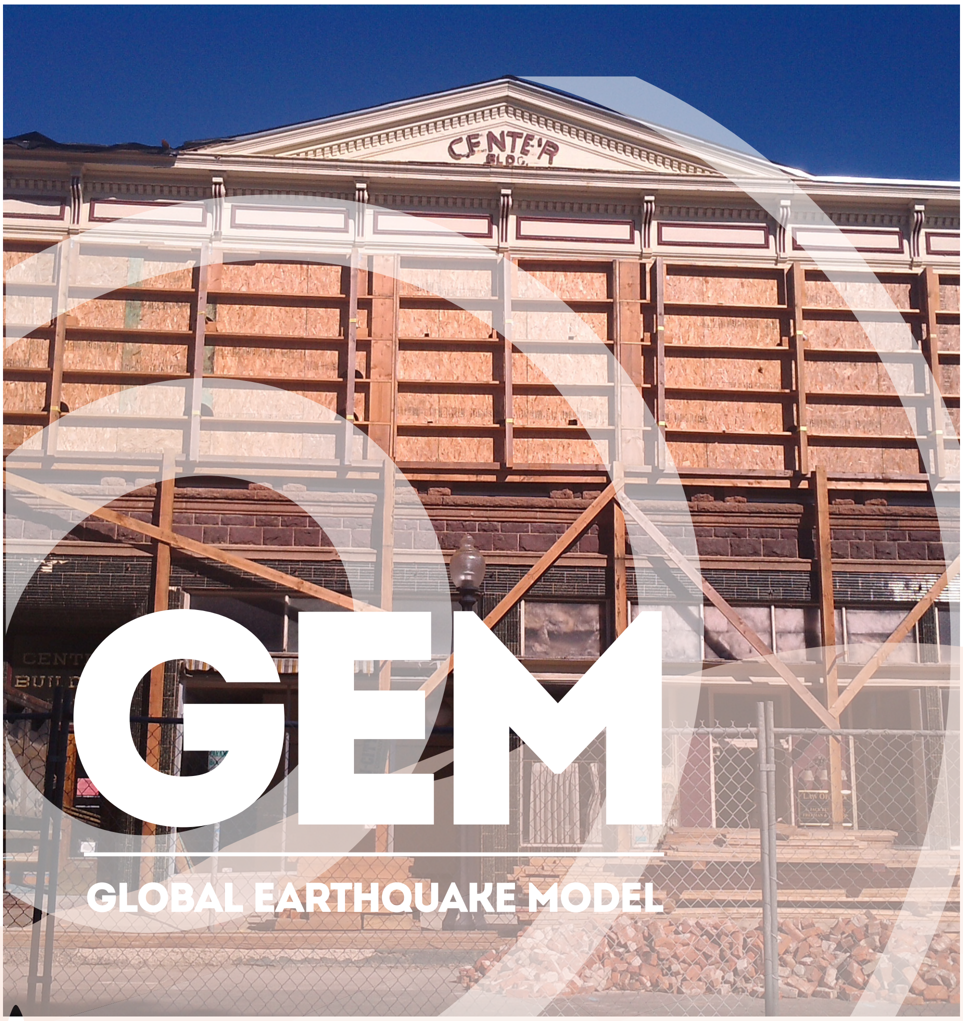
Earthquake Recovery Modelling

GEM TECHNICAL REPORT
2017 V1.0.0

Global Earthquake Model Foundation (GEM)
California Seismic Safety Commission (CSSC)
University of California at Los Angeles (UCLA)



NAPA, CALIFORNIA
MEAN RECOVERY
PROBABILITIES



THE GLOBAL EARTHQUAKE MODEL

The mission of the Global Earthquake Model (GEM) collaborative effort is to increase earthquake resilience worldwide.

To deliver on its mission and increase public understanding and awareness of seismic risk, the GEM Foundation, a non-profit public-private partnership, drives the GEM effort by involving and engaging with a very diverse community to:

- Share data, models, and knowledge through the OpenQuake platform
- Apply GEM tools and software to inform decision-making for risk mitigation and management
- Expand the science and understanding of earthquakes

GEM Foundation

Via Ferrata 1
27100 Pavia, Italy
Phone: +39 0382 5169865
Fax: +39 0382 529131
info@globalquakemodel.org
www.globalquakemodel.org

JULY 2017

The GEM Foundation wishes to acknowledge the following institutions/organizations, for their contributions to the development of this report:

- California Seismic Safety Commission (CSSC)
- University of California at Los Angeles (UCLA)
- Diana Maria Contreras Mojica, GEM Social Vulnerability and Resilience Coordinator

GLOBAL EARTHQUAKE MODEL
working together to assess risk

“Back to Normal”: Earthquake Recovery Modelling

Report Prepared by
The Global Earthquake Model Foundation (GEM)

Authors:
Venetia Despotaki, Henry V. Burton, John Schneider, Christopher G. Burton

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

CSSC Publication No. 2017-06

ACRONYMS

ACS	American Community Survey
FEMA	Federal Emergency Management Agency
GEM	Global Earthquake Model
IM	Intensity Measure
IRMT	Integrated Risk Modelling Toolkit
LS	Limit State
MMI	Modified Mercalli Intensity
Mw	Moment Magnitude
NOcc	Building unsafe to occupy
OccFull	Building fully functional
OccLoss	Building safe to occupy, but unable to facilitate normal operations
PBEE	Performance-based Earthquake Engineering
PEER	Pacific Earthquake Engineering Research Center
PHEALTHINS	Percentage of civilian non-institutionalized population with any type of health insurance
PH-ENGLISH	Percentage of households where they speak English
PHU-OWNED	Percentage of occupied housing units that are owner occupied
QGIS	Quantum Geographic Information System
SPUR	San Francisco Planning and Urban Research Association
SSC	Seismic Safety Commission
$\mathbf{T}_{NOcc}$	Time spent in the NOcc functioning state

$\mathbf{T}_{OccFull}$	Time spent in the OccFull functioning state
$\mathbf{T}_{OccFull}$	Time spent in the OccLoss functioning state
$\mathbf{T}_{LC}$	Control Time: a pre-defined period that the recovery path is assessed
UCLA	University of California at Los Angeles
USGS	United States Geological Survey
Std Dev	Standard Deviation

Executive Summary

Introduction

Given the high level of earthquake risk in California, all communities need to be prepared to respond to and recover from the impacts of a potentially devastating earthquake. Although there has been significant research on the estimation of direct economic losses immediately after an earthquake, there has not been enough research about long-term recovery, and even less in the development and application of computer simulation models. This is to some extent because data on building repair and recovery times from past earthquakes have not been systematically documented (Comerio M. , 2006). Moreover, the models developed so far have not successfully captured the complexity of the recovery process. Recovery depends on many factors (such as the socio-economic conditions of the affected area) that are usually difficult to measure, understand and apply to predicting or modelling the recovery process.

We acknowledge that the parameters taken into account in the present analysis may not be appropriate for application to potential earthquakes in other regions due to many factors that can affect the post-disaster recovery process. The methodology was developed using the city of Napa and the 2014 M6 South Napa Earthquake as a real-world case study. Therefore, the results are expected to apply to earthquakes with similar impact and to communities with similar socio-economic and building characteristics. To apply the resulting model in regions with different characteristics, additional data collection and validation would be necessary.

To address some of the key factors that influence recovery, the Alfred E. Alquist Seismic Safety Commission (SSC) engaged the GEM (Global Earthquake Model) Foundation and the University of California at Los Angeles (UCLA), Department of Civil and Environmental Engineering, a) to develop a methodology and an open-source and transparent software tool to estimate recovery states and recovery times following an earthquake; and b) to investigate the effect of external socio-economic factors on these recovery times. The SSC leveraged over 20 million dollars in funding from GEM supporters that has been used to develop the OpenQuake[1] software package and related data sets, which the 'Back to Normal": Earthquake Recovery Modelling project uses.

[1] Seismic hazard and risk calculation software, developed by the Global Earthquake Model (GEM) Foundation.

Project Aims and Goals

This project aims to quantify the effectiveness of specific actions to speed recovery of the building stock. In order to accomplish this, the project produced the necessary technical science for the development of computer simulation models to estimate building recovery states and recovery times following an earthquake.

In particular, this project has achieved the goals listed below:

1. Developed a user-friendly, non-commercial and transparent software tool, herein referred as the Integrated Risk Modelling Toolkit (IRMT), to make map-based comparisons showing the effect of different resilient actions on the recovery times.
2. Developed an analytic methodology, herein referred as the Reconstruction Recovery Model, to estimate post-earthquake recovery. Specifically:
   - Recovery curves[2] of individual buildings were aggregated to estimate community-wide recovery, while properly accounting for statistical correlation and uncertainty.
   - Thresholds of building damage were defined and quantified that relate to resilience: functionality, livability and reparability.
   - An event-tree[3] method was developed to analyze a building's state of recovery over time, based on its initial damage state.
   - Equations were developed that predict how a building recovers functionality over time.
   - External factors were identified that affect the process of a building's recovery. Although not directly requested from the SSC, following the 2014 M6 South Napa Earthquake, three field surveys were conducted and the recovery of 356 individual damaged buildings in the city of Napa was documented, over a period of 18 months, at 6-month intervals (February and August 2015 and March 2016). Consequently, an additional methodology, the Socio-Economic Recovery Model was also developed. This model provides predictions of community recovery based mainly upon socio-economic factors that influence the recovery trajectory. Because this model was developed outside the scope of the original study, it has not yet been implemented into the IRMT.

[2] Curves that relate building recovery times to the level of functionality.
[3] A graph that indicates the possible recovery stages of a building until it is fully recovered based on its post-earthquake level of damage.

3. Integrated the Reconstruction Recovery Model for practical use into GEM's OpenQuake modelling platform.
4. Demonstrated reasonable results by using the methodologies to model past earthquakes. In this context, the 2014 M6 South Napa Earthquake was used as a case study.
5. Conducted a case study for the Southern California ShakeOut scenario[4] using the Reconstruction Recovery Model, considering each of the counties in the Scenario as a 'community snapshot" (namely: Riverside, Orange, Kern, San Diego, San Bernardino, Ventura, Los Angeles, Imperial). Thus, recovery times from a large simulated earthquake were estimated.
6. Quantified how well specific actions can speed recovery of the building stock, by running the models 'before" and 'after" taking action.

Uniqueness of Project - Intended Audience

Key stakeholders and decision makers have a major responsibility following an earthquake to take actions that best serve their communities and increase resilience. They need to be well informed about the factors that influence the recovery process and the best practices to achieve successful results. Therefore, there is a need for long-term recovery studies to develop user-friendly, open-source tools that enable stakeholders to test how different resilience actions (such as building retrofitting) affect recovery times. Previously, various researchers have developed indicators to monitor the recovery process following a disaster (e.g., Chang, 2009, Brown et al., 2010 or Contreras et al., 2014). Other studies have focused on how socio-economic conditions affect recovery, but have provided only qualitative conclusions (e.g., Burton C., 2015). None of the studies has provided a user-friendly tool to predict recovery times or to relate building recovery stages to socio-economic factors in a quantitative manner.

The results of this project will be useful to various stakeholders, including government, industry and academics that are concerned with enhancing post-disaster recovery of California communities. Particularly:

[4] The ShakeOut is an exercise based on a potential magnitude 7.8 earthquake on the southern San Andreas Fault with the goal to identify the physical, social and economic consequences of a major earthquake in southern California (Jones, et al., 2008).

- The IRMT, which incorporates the Reconstruction Recovery Model, would be useful to assist policy-makers, municipal governments and disaster managers to evaluate if and how factors, such as changes in building design and construction, affect the immediate impact and recovery trajectory.
- The Socio-Economic Recovery Model would be valuable to key decision makers to make initial predictions of the recovery time and to identify those areas where high social vulnerability may prolong recovery times. By identifying the most vulnerable areas either before or after an earthquake, the model could be used as a guide to allocate (post) earthquake financial resources to people in most need of assistance.
- Knowledge of the socio-economic parameters that most significantly influence the recovery trajectory contributes to a better understanding of the complex nature of recovery processes and allows decision-makers to identify the strengths and weaknesses within their communities. Thus, decision makers could develop informed pre-disaster recovery plans to address specific issues to facilitate recovery in accordance with the actual needs of the population.
- The database developed during the field surveys in the city of Napa includes detailed information on: a) the location; b) the structural characteristics; c) the age of construction; d) the number of stories; e) colour-tagging information (red or yellow); f) the damage description; and, g) the quantified recovery progress over time for 356 damaged buildings. This database will be very useful to various stakeholders, including insurance companies, engineering companies or researchers. For example, the database could be used to analyse how the recovery times vary between different levels of damage and building typologies.

Methods developed and data sets used

This study presents two methodologies and a user-friendly, open-source software tool to model recovery trajectories at the building and community level following an earthquake. In both cases, the recovery time is predicted to be a function of the degree of damage from the earthquake. In the first approach, the recovery time is predicted by analyzing the steps in the reconstruction and permitting process for construction and repair of buildings of different types. This approach results in the Reconstruction Recovery Model, which provides a framework to estimate both **single building and community recovery** times and can be used to establish minimum bounds of reconstruction based on standard processes, such as the time required for inspection and/or damage evaluation, design, mobilization, permitting and repair. This is a process-based approach, which is based on normal building practice and approval processes. As such, it does not necessarily capture in a detailed way other factors that may prolong the duration of post-earthquake recovery.

In order to explore this issue, data on the rebuilding process from an actual earthquake was required. By chance, the 2014 M6 South Napa Earthquake occurred part way through this study. This unfortunate event was used as an opportunity to collect data on housing recovery from the city of Napa. On-site field surveys were conducted over a period of 18 months at 6-month intervals after the event, and the actual recovery data were then analyzed in relation to socio-economic data for the city of Napa. Based on these findings, an empirical Socio-Economic Recovery Model was developed to predict earthquake recovery at the community level. Both the Reconstruction Recovery Model and the Socio-Economic Recovery Model start from the initial condition of the distribution of building damage after the earthquake. The latter model was developed and further validated using actual data from the recovery of the city of Napa. This model provides **community recovery** predictions based on socio-economic variables and the level of post-earthquake physical damage. Thus, this model is mainly focused on how socio-economic conditions influence recovery times.

The Reconstruction Recovery Model was incorporated into the GEM IRMT, which allows the user to estimate the recovery time from an earthquake. The IRMT provides default values, but it provides the option to adjust these factors based on other inputs, such as from the results of the Socio-Economic Recovery Model. The IRMT is very flexible and allows users to experiment with different assumptions and factors that could alter the recovery process. The IRMT is a plugin of the open-source *QGIS*[5] software and in order to run it the user needs to download *QGIS* on his/her computer. Then, the IRMT can be installed using the *QGIS* Plugins Manager, which is accessible through the *QGIS* menu as **Plugins » Manage and install plugins**. Additional information about the IRMT can be found here: https://plugins.qgis.org/plugins/svir/. Appendix 2 describes the basic workflow to develop an end-to-end recovery prediction.

[5] A free and open-source Geographic Information System to create, edit, visualize, analyze and publish geospatial information (http://www.qgis.org/en/site/forusers/download.html).

Key Findings

Finding 1: The Reconstruction Recovery and Socio-Economic Recovery methodologies were applied to estimate community recovery times using data from the city of Napa and recovery from the 2014 M6 South Napa Earthquake. By comparing the outputs, it was found that the recovery is significantly influenced by pre-existing socio-economic factors, which should be included in recovery prediction models to obtain more realistic recovery predictions.

Finding 2: The Reconstruction Recovery Model was applied using the Southern California ShakeOut Scenario and considering each of the counties in the Scenario as a 'community snapshot" (namely: Riverside, Orange, Kern, San Diego, San Bernardino, Ventura, Los Angeles, Imperial). However, when the model is applied at the county scale, the single aggregated recovery curve may not provide realistic estimates of recovery durations. Partially this is because: a) the lack of detailed building exposure data required the data to be aggregated at a coarse level; and b) the model was not calibrated for large-scale earthquakes, where various factors, such as shortage of labor or materials and loss of critical infrastructure, might delay the recovery process.

Finding 3: Using the Socio-Economic Recovery Model for the city of Napa, it was determined that the level (or amount) of earthquake building damage together with seven socio-economic variables contribute most to the prediction of building recovery. These variables are listed below **in order of importance**, with the level of damage being the most significant factor that influences recovery:

a. Level (or amount) of building damage
b. Homeownership
c. Percentage of households that have a male householder
d. Presence of health insurance coverage
e. Employment status
f. Percentage of households that have any type of available income
g. Percentage of buildings constructed after 1950 (which are considered to be seismically designed and consequently behave better during an earthquake)
h. Percentage of English speaking households

It should be noted that the earthquake insurance penetration rate was not included in the model because data was not available at a sufficient spatial resolution (i.e., block group level rather than zip code), nor were dates available associated with the insurance approval process.

Finding 4: Based on qualitative analyses, such as personal interviews, it was found that in the city of Napa pre-existing earnings and wealth of the population were the main sources and drivers of recovery. Since earthquake insurance penetration in the city was very low and there was a considerable delay in the authorization of the federal Individual Assistance program, most of the homeowners initiated the payment of repairs from their personal resources and savings. This qualitative finding is consistent with socio-economic factors b, d, e, and f above.

Finding 5: Populations that do not speak English, or that speak English as a second language, such as residents of the Latino communities in Napa, constitute a rather vulnerable group following a disaster. The municipal government of Napa faced many challenges in identifying and documenting housing damage and estimating monetary losses. One of the main factors contributing to this issue was that some people with English as a second language were particularly reluctant to report building damage due to lack of trust in authorities, possibly for fear of losing their housing rights. This lack of trust may have significantly affected the progress of recovery, and is therefore consistent with socio-economic factor h above.

Finding 6: Based on the field surveys and the documentation of the building recovery in the city of Napa, it was observed that about 60% of the yellow-tagged buildings were fully recovered within the first six months, and 83% were fully recovered after 18 months. On the other hand, nearly 50% of the red-tagged buildings were still not recovered 18 months after the earthquake. In addition, nearly 80% of the surveyed structures that were not seismically designed sustained significant damage and were assigned a red tag.

Known Limitations

Reconstruction Recovery Model

The use of the default parameters (such as the assessment times, inspection times and mobilization times) provided currently by the IRMT would not be appropriate for application to potential earthquakes in other regions without additional post-disaster studies. More research is needed to better define the reconstruction times of various building types and at different levels of damage, in order to improve the applicability of model.

Socio-Economic Recovery Model

The methodology was developed using the city of Napa and the 2014 M6 South Napa Earthquake as a real-world case study. Therefore, it is expected to be applicable to communities with similar socio-economic and building characteristics (such as other California cities). In order to apply the model in regions with different characteristics, additional data collection and validation would be necessary.

The selection of the socio-economic variables used in the analysis and their potential to predict recovery outcomes was based on previous research (e.g., Cutter et al., 2010) demonstrating their influence on the recovery of different communities. However, the model is flexible and can accommodate different parameters as predictors of recovery, such as the earthquake insurance penetration rate, which could be a very significant factor in the recovery process. In the case of the city of Napa, earthquake insurance was not considered in the analysis, because data were not available at the desired resolution, nor was there knowledge of the dates that insurance claims were approved.

Additional Limitations

There are other factors that influence the recovery process, such as government intervention in areas of high damage (a.k.a., the red zone), building permit processes, legal and financial aspects of demolition, debris management, and insurance availability.

The government policy of the red zone will depend on the degree of damage after the earthquake and decisions taken by emergency managers in this respect. The amount of debris will be also a function of the degree of damage and their management after the earthquake must be described in the Emergency Response Plan of the city.

Debris management will influence the speed of the recovery process, as well as administrative processes such as permits and legal and financial aspects of demolitions, which could be also addressed either by the Emergency Response Plan of the city or by a pre-impact recovery plan. Another aspect that contributes to recovery is insurance availability. Although we received some data on insurance penetration in the city of Napa, the data provided were not adequate to determine whether it significantly influenced the building reconstruction process.

Recommendations

Recommendation 1: More long-term recovery studies from other earthquakes are required to refine both the Reconstruction Recovery Model and the Socio-Economic Recovery Model. It is recommended to develop a robust tool for earthquake recovery modelling and planning by merging the 2 models and implementing them into the IRMT. In addition, more long-term recovery studies are essential to identify and investigate indirect losses following earthquakes.

Recommendation 2: Extend the methodologies beyond residential buildings to model recovery of critical facilities, such as hospitals, fire and police stations, power plants, water treatment plants and telecommunication networks.

Recommendation 3: Based on the identified significant variables from the Socio-Economic Recovery Model that positively contribute to the recovery process, it is recommended to:

1. Facilitate access to assistance for vulnerable groups of the population, such as residents that do not speak English.
2. Conduct further investigations into the relationships between the variables that correlate most positively with recovery (e.g., homeownership and health insurance) to determine the underlying causes.
3. Conduct more extensive research on cost-benefit analysis of retrofitting buildings because the buildings not seismically designed in the city of Napa sustained significantly more damage compared to stronger structures.
4. Improve access to financial mechanisms, such as earthquake insurance, to residents exposed to high earthquake risk, as well as investigate and promote alternative

post-earthquake resources, such as grants, which will support residents in the rebuilding process.

Recommendation 4: Facilitate the involvement of insurance industry partners in future projects as advisors to improve the model and to gain access to more detailed earthquake insurance data. This would facilitate, for instance, a better understanding of the degree to which access to earthquake insurance influences the recovery process.

TABLE OF CONTENTS

LIST OF FIGURES

1 Introduction

Given the high level of earthquake risk in California, all communities need to be prepared to respond to and efficiently recover from the impacts of a potentially devastating earthquake. Key stakeholders and decision-makers have a major responsibility following an earthquake to take actions that best serve their communities and increase resilience. They need to be well informed about the factors that influence the recovery process and the best practices to achieve successful results.

Although there is significant research into developing computer models to predict direct economic losses immediately after an earthquake, long-term recovery studies and the development of computer simulation recovery models have received limited attention across governments and the scientific community. This is to some extent because: a) data on building repair and recovery times from past earthquakes have not been systematically documented, (Comerio, 2006); and b) recovery is a complex and multidimensional process that depends on many factors (such as the socio-economic conditions of the affected region). These factors are usually difficult to measure, understand and apply to predicting or modelling the recovery process.

Previously, various researchers have developed indicators to monitor the recovery process following a disaster, e.g., (Chang, 2009), (Brown, Platt, & Bevington, 2010) or (Contreras, Blaschke, Kienberger, & Zeil, 2014). Other studies have focused on how socio-economic conditions affect recovery, but have provided only qualitative conclusions, e.g., (Burton C. , 2015). None of the studies has provided a user-friendly tool to predict recovery times or to relate building recovery stages to socio-economic factors in a quantitative manner.

To address this gap, the Alfred E. Alquist Seismic Safety Commission (SSC) engaged the GEM (Global Earthquake Model) Foundation and the University of California at Los Angeles (UCLA), Department of Civil and Environmental Engineering, to develop: a) a methodology and an open-source software tool to estimate recovery states and recovery times following an earthquake, and b) to investigate the effect of external socio-economic factors on these recovery times. The SSC leveraged over 20 million dollars in funding from GEM supporters that has been used to develop the OpenQuake software package and related data sets, which the 'Back to Normal": Earthquake Recovery Modelling project uses. The outputs of this project will be useful to various stakeholders, including government, industry and academics, which are concerned with enhancing post-disaster recovery of California communities.

We acknowledge that the parameters taken into account in the present analysis may not be appropriate for application to potential earthquakes in other regions due to many factors that can affect the post-disaster recover process. The methodology was developed using the city of Napa and the 2014 M6 South Napa Earthquake as a real-world case study. Therefore, the results are expected to apply to earthquakes with similar impact and to communities with similar socio-economic and building characteristics. To apply the resulting model in regions with different characteristics, additional data collection and validation would be necessary.

This report is organized as follows:

Section 2 briefly introduces the main outputs of the project and identifies the possible users and applications of the results.

Section 3 provides a description of the two models developed within this project, the Reconstruction Recovery Model and the Socio-Economic Recovery Model. In addition, it describes the integration of the Reconstruction Recovery Model framework into the GEM's Integrated Risk Modelling Toolkit (IRMT), a software tool designed to develop composite indicators, which is currently further extended to estimate the time required for a community or an individual building to recover following an earthquake.

Section 4 discusses conclusions derived from applying the Reconstruction Recovery Model to the Southern California ShakeOut Scenario. In addition, it presents a comparison of the results of the Reconstruction Recovery Model and the Socio-Economic Recovery Model, which were both applied to estimate community recovery times of the residential building stock of the city of Napa, following the 2014 M6 South Napa Earthquake.

Section 5 describes known limitations of both frameworks and **Section 6** and **Section 7** summarize key findings and recommendations.

2 Intended Audience

This project produced the necessary technical science for the development of a framework (herein referred as the Reconstruction Recovery Model) and an open-source and transparent computer simulation model (herein referred as the Integrated Risk Modelling

Toolkit) to estimate building and community recovery states and recovery times following an earthquake. In addition, the effect of the socio-economic conditions of an area on the building recovery times was investigated using the city of Napa and the building recovery from the 2014 M6 South Napa Earthquake as a real-world case study. In this context, field surveys were conducted over a period of 18 months at 6-month intervals (February and August 2015 and March 2016) and building recovery data were collected and analyzed in relation to socio-economic indicators for the city. This work resulted in the Socio-Economic Recovery Model, which provides community level recovery predictions mainly based on socio-economic variables.

The results of this project will be useful to various stakeholders and particularly:

- The Integrated Risk Modelling Toolkit (IRMT), which incorporates the Reconstruction Recovery Model, would be useful to assist policy-makers, municipal governments and disaster managers to evaluate if and how factors, such as changes in building design and construction, affect the immediate impact and recovery trajectory.
- The Socio-Economic Recovery Model would be valuable to key decision makers to make initial predictions of the recovery time and to identify those areas where high social vulnerability may prolong recovery times. By identifying the most vulnerable areas either before or after an earthquake, the model could be used as a guide to allocate (post) earthquake financial resources to people in most need of assistance.
- Knowledge of the socio-economic parameters that most significantly influence the recovery trajectory contributes to a better understanding of the complex nature of recovery processes and allows decision-makers to identify the strengths and weaknesses within their communities. Thus, decision makers could develop informed pre-disaster recovery plans to address specific issues to facilitate recovery in accordance with the actual needs of the population.
- The database developed during the field surveys in the city of Napa includes detailed information on: a) the location; b) the structural characteristics; c) the age of construction; d) the number of stories; e) color-tagging information (red or yellow); f) the damage description; and, g) the quantified recovery progress over time for 356 damaged buildings. This database will be very useful to various stakeholders, including insurance companies, engineering companies or researchers. For example, it could be used to analyze how the recovery times vary between various levels of damage and building typologies.

3 Methods Developed

This section presents the two methodologies and the open-source software tool that were developed to model recovery trajectories at the building and community level following an earthquake. The description of the Reconstruction Recovery Model framework and its implementation into the IRMT is followed by the description of the development of the Socio-Economic Recovery Model.

Reconstruction Recovery Model

The performance-based earthquake engineering (PBEE) framework, developed by the Pacific Earthquake Engineering Research Center (PEER) provides a rigorous methodology for assessing the seismic performance of individual buildings and it works in four stages: hazard analysis, structural analysis, damage analysis (whereby measures of damage to building components are determined), and finally, estimation of measures of performance, such as associated repair/replacement costs (referred as decision variables). In the Reconstruction Recovery Model, for each individual building in a community the PBEE framework is applied, but incorporating new measures of damage and a new decision variable, the outcome of which is a building-level recovery function. Then, the individual building recovery functions are aggregated to develop a community-level recovery curve, which relates recovery time to the level of functionality of a building. The measure of a building's functionality, which is the new decision variable, can be expressed in several ways, such as the recovery of the housing capacity in the case of a residential building. The conceptual overview of the framework for a residential community is illustrated in Figure 0-1.

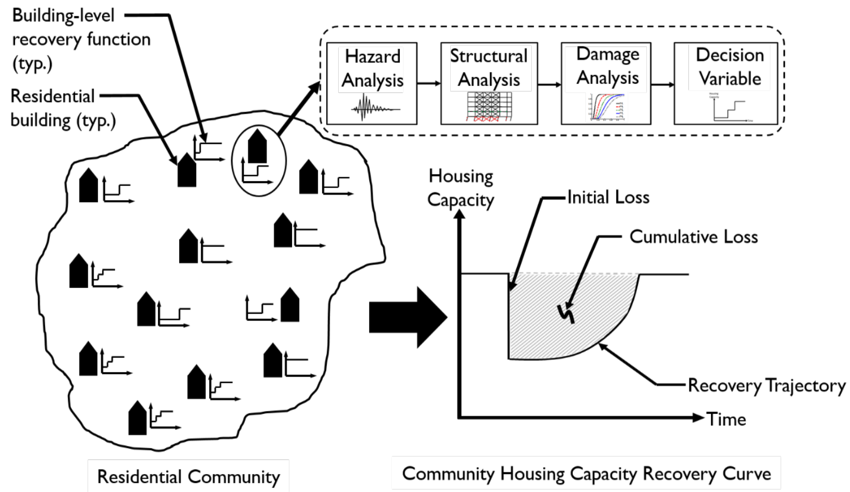


Figure 0-1. Conceptual illustration of the Reconstruction Recovery Model framework. (Burton, Deierlein, Lallemant, & Lin, 2015).

To exemplify, new building measures of the level of damage (limit states) that are intrinsically linked to recovery activities are incorporated in the framework and include: (i) damage triggering inspection (LS_1); (ii) occupiable damage with loss of functionality (LS_2); (iii) unoccupiable damage (LS_3); (iv) irreparable damage (LS_4); and (v) collapse (LS_5). These limit states have been adapted from the building performance categories defined by the San Francisco Planning and Urban Research Association (SPUR) (Poland, 2009).

Each of the above recovery-based (limit) states is associated with a specific recovery path that the building follows until it recovers. The path is described by using discrete functioning states and the time that the building spends in each state. The functioning states include: (i) the building is unsafe to occupy (*NOcc*); (ii) the building is safe to occupy but unable to facilitate normal operations (*OccLoss*); and (iii) the building is fully functional (*OccFull*). For example, if a building after an earthquake falls into the LS_2 (occupiable damage with loss of functionality), it first enters the *NOcc* state until inspections are complete. Then, it enters the *OccLoss* state until functionality is restored and the completion of these repairs brings the building back to the *OccFull* state. The structural repair time of a building can be calculated based on a methodology developed by Mitrani-Reiser, (2007) and other time parameters (such as inspection times or assessment times) can be empirically derived by monitoring the post-earthquake recovery of individual buildings of various typologies and levels of damage.

Each of the above functioning states is linked to a quantifiable level of functionality of the building. As previously mentioned, the level of functionality can be expressed in several ways, such as the recovery of the housing capacity in the case of a residential building. Knowing the level of functionality associated with each functioning state, the recovery path for each limit state can be related to a recovery function, as illustrated in Figure 0-2 and calculated as follows:

$$\left[q(t) \mid LS_i\right] = \begin{cases} \left[q(t) \mid NOcc\right] & t < \left[T_{NOcc} \mid LS_i\right] \\ \left[q(t) \mid OccLoss\right] & \left[T_{NOcc} \mid LS_i\right] \leq t < \left[T_{NOcc} + T_{OccLoss} \mid LS_i\right] \\ \left[q(t) \mid OccFull\right] & \left[T_{NOcc} + T_{OccLoss} \mid LS_i\right] \leq t < T_{LC} \end{cases} \quad 3\text{-}1$$

where $\left[q(t) \mid LS_i\right]$ is the time dependent building functionality given its immediate post-earthquake limit state *LSi*; $\left[q(t) \mid NOcc\right]$, $\left[q(t) \mid OccLoss\right]$ and $\left[q(t) \mid OccFull\right]$ represent the level of functionality associated with the *NOcc*, *OccLoss* and *OccFull* states respectively; $\left[T_{NOcc} \mid LS_i\right]$ is the time between the earthquake and the end of the *NOcc* phase associated with limit state *LSi*; $\left[T_{NOcc} + T_{OccLoss} \mid LS_i\right]$ is the time between the earthquake to the end of the OccLoss phase for limit state LS_i; and T_{LC} is the predefined period that the recovery path (and recovery function) is assessed.

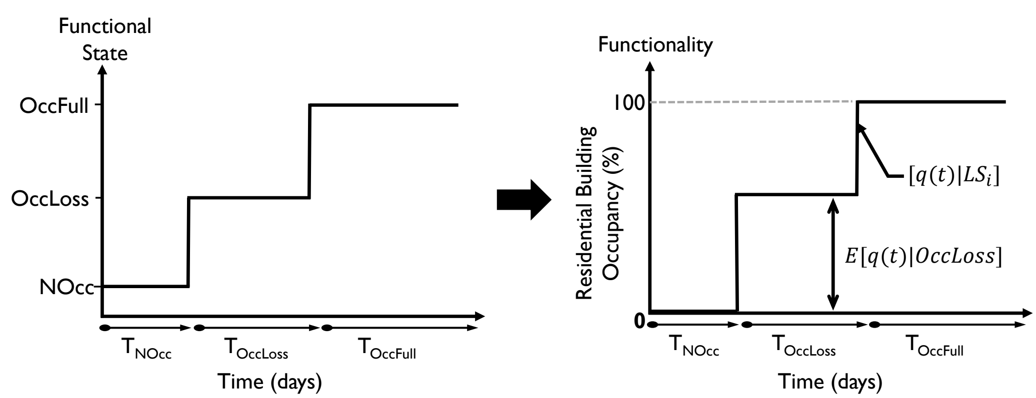


Figure 0-2. Conversion from recovery path to recovery function for a residential building.

(Burton, Deierlein, Lallemant, & Lin, 2015).

In other words, Figure 0-2 demonstrates that if the building is in the *NOcc* functioning state (Figure 0-2 - left), its functionality is zero (Figure 0-2 - right); when it enters the

OccLoss state its functionality increases, until the building is fully functional (*OccFull*), where its functionality is 100%.

The final individual building recovery function is computed accounting for the likelihood of the building being in each of the five limit states for a given ground shaking intensity. This is illustrated in the event tree, shown in Figure 0-3, where each limit state is associated with a unique recovery function (computed from Equation 3-1). Figure 0-3 also incorporates a sixth event that corresponds to damage below the threshold level that triggers inspection. The uncertainty associated with the building's limit state following the earthquake and the expected building recovery is determined by the following equation:

$$E\left[q(t) \mid IM\right] = \sum_{1}^{n_{ls}} \left[q(t) \mid LS_i\right] P\left[LS_i \mid IM\right] \qquad 3\text{-}2$$

where n_{ls} is the number of the limit states; $E\left[q(t) \mid IM\right]$ is the expected individual building recovery function given an intensity measure level (*IM*); and $P\left(LS_i \mid IM\right)$ is the probability that the building is in the i^{th} limit state for a given *IM* level.

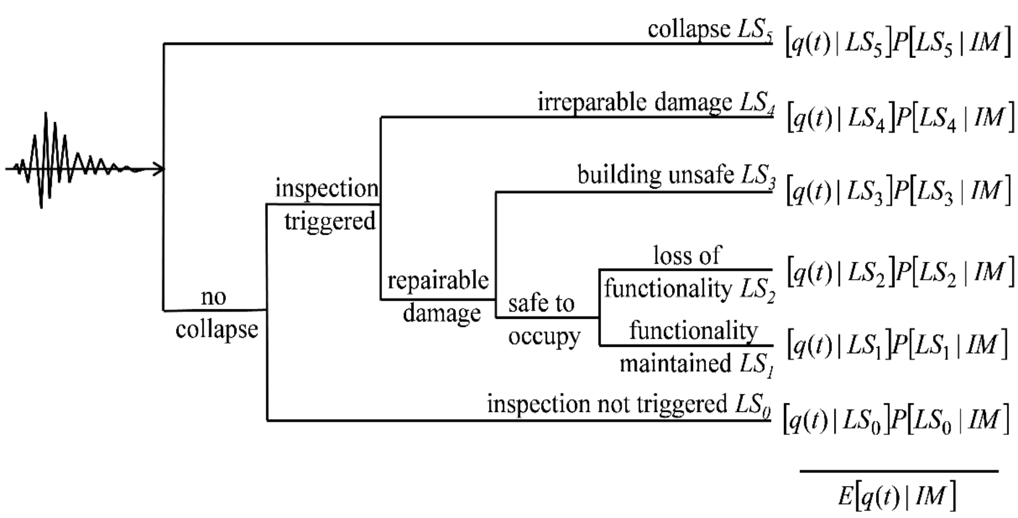


Figure 0-3. Event tree used to assess building-level recovery.

(Burton, Deierlein, Lallemant, & Lin, 2015).

In the final stage of the methodology, individual building recovery functions can be aggregated to develop community-level recovery curves, as shown in Figure 0-1. For more technical and thorough details on the development of this framework readers are referred to Appendix 1.

Integrated Risk Modelling Toolkit (IRMT)

The Reconstruction Recovery Model was implemented into the existing OpenQuake Integrated Risk Modelling Toolkit (IRMT) developed by GEM, which is a software tool that allows the user to develop composite indicators and to integrate them with physical risk estimations. More specifically, the IRMT is a plugin of the open-source *QGIS*[6] software and in order to run it the user needs to download *QGIS* on his/her computer. Then, the IRMT can be installed using the *QGIS* Plugins Manager, which is accessible through the *QGIS* menu as **Plugins » Manage and install plugins**. Additional information about the IRMT can be found here: https://plugins.qgis.org/plugins/svir/.

The IRMT's capabilities were extended (following the implementation of the Reconstruction Recovery Model) to allow users to generate recovery curves at the building and community level. The tool is open-source and transparent and users can access, change and adjust the source code to their needs. The input variables necessary to run the Reconstruction Recovery Model are listed in Table 0-1. These are pre-defined variables that need to be manually introduced and should be adjusted to the study region. Appendix 2 describes in more detail the steps on how to use the tool, along with associated screenshots for clarification purposes.

Table 0-1. Input variables to use the IRMT and run the Reconstruction Recovery Model.

Input Variables	Short Explanation
Damage by asset	File that contains the mean probabilities of exceedance of each limit state for each individual building; output of the OpenQuake-engine
Inspection Time	Time to complete inspections
Assessment Time	Time to conduct engineering assessments
Mobilization Time	Time to mobilize for construction
Lead Time Dispersion	Level of uncertainty associated with the lead time[7]
Repair Time	Time to repair a building
Recovery Time	Period between the occurrence of an earthquake and the restoration of full functionality of the building
Transfer Probabilities	Conditional probability of being in a particular recovery-based (limit) state, given the occurrence of a loss-based damage state (e.g., slight, moderate, extensive, complete)

[6] A free and open-source Geographic Information System to create, edit, visualize, analyze and publish geospatial information (http://www.qgis.org/en/site/forusers/download.html).

[7] Lead time is the time required for building inspection and/or evaluation, finance planning, architectural/engineering consultations, a competitive bidding process, and mobilizing for construction (Mitrani-Reiser, 2007).

Socio-economic Recovery Model

The Reconstruction Recovery Model only **implicitly** accounts for external factors (such as the socio-economic conditions) that may prolong the duration of earthquake recovery (for example, by adjusting the lead time[8] in the model). However, additional research and long-term recovery studies are required to determine the socio-economic parameters that most significantly influence the recovery times and subsequently define ways to incorporate them into the Reconstruction Recovery Model in a more robust and detailed way.

In order to explore this issue, data on the rebuilding process from an actual earthquake was required. Therefore, an earthquake recovery case study was developed using the city of Napa, California and the building recovery from the 2014 M6 South Napa Earthquake. In this case study, the socio-economic factors were identified that may affect recovery times and trajectories. This work resulted in the development of a second model, the Socio-Economic Recovery Model, which provides predictions of post-earthquake community recovery over time, considering both the effect of the post-earthquake damage to the building stock and the effect of the socio-economic conditions of the affected area on the recovery trajectory. Because this model was developed outside the scope of the original study, it has not yet been implemented into the IRMT, described in Section 3.2.

More specifically, a set of socio-economic indicators for the city of Napa (Table 4-1 in Appendix 4) and a post-earthquake damage indicator (the Modified Mercalli Intensity[9] (MMI) observed at the time of the 2014 M6 South Napa Earthquake) were used as predictors of community level recovery at different time instances in a probabilistic recovery modelling framework (Despotaki, Sousa, & Burton, 2017). To accomplish this, it was necessary to: a) monitor and quantify the building recovery of the city of Napa over time; b) assess and quantify the socio-economic conditions of the area of interest; and finally, c) relate the two concepts -recovery and socio-economic conditions- using a statistical regression tool, which provides community level recovery predictions.

[8] Lead time is the time required for building inspection and/or evaluation, finance planning, architectural/engineering consultations, a competitive bidding process, and mobilizing for construction (Mitrani-Reiser, 2007).
[9] MMI: An arbitrary ranking of the intensity of an earthquake based on observed effects (Wood & Newmann, 1931).

Quantification of the Post-Earthquake Building Recovery

In the aftermath of the 2014 M6 South Napa Earthquake, city officials implemented a rapid structural assessment program throughout the city and the building damage observations were geocoded and made available via a web Geographic Information System (*GIS*) by the City of Napa. These included: a) the location of the buildings; b) the structural characteristics; c) the age of construction; d) the number of stories; e) colour-tagging information (red or yellow); and f) the associated building damage description.

In order to monitor and quantify the progress of the building recovery occurring in the city and further use it in a predictive model, a building-by-building detailed inspection via three separate field surveys was conducted in February 2015, August 2015 and March 2016 (6, 12 and 18 months following the earthquake). As part of the evaluation process, the recovery progress of 356 damaged structures (which included all the red and a random sample of the yellow-tagged buildings initially identified by the city officials) was documented and photographed. The spatial distribution of the surveyed buildings in the city of Napa is presented in Figure 0-4.

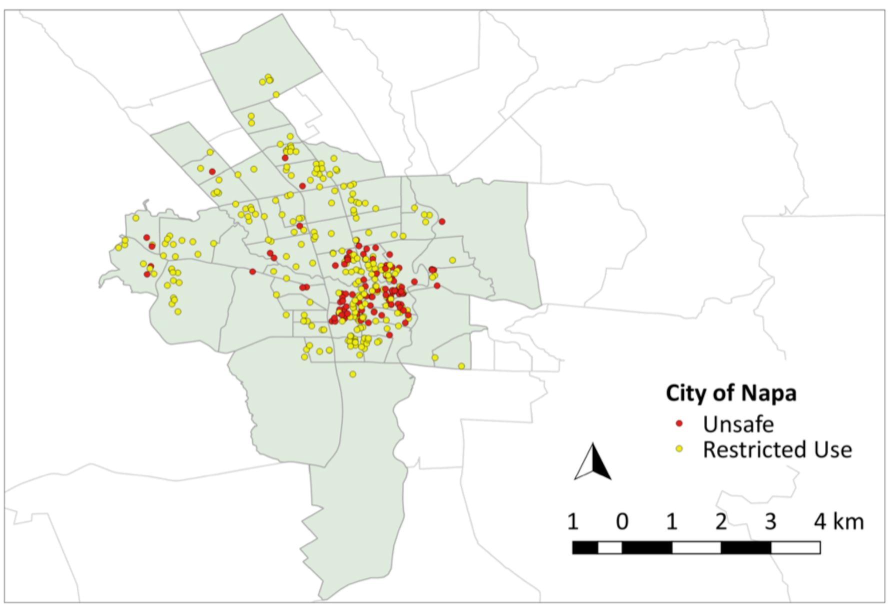


Figure 0-4. The 356 evaluated red and yellow -tagged (ATC, 2005) buildings in the city of Napa.

During each field survey, each individual building (of the 356) was assigned a code (0 or 1) indicating its stage of recovery relative to its initial damage following the event. In this framework, 0 represents a 'No Recovery" stage (no retrofit and/or rebuild operations with respect to the immediate post-earthquake condition); and 1 is associated with the building's 'Full Recovery" (building fully repaired/reconstructed and re-occupied). In this way, for each time instance a database was created indicating the recovery stage of each building and its recovery evolution over time. An example of the recovery progress of a building in the city of Napa over a period of 18 months is presented in Figure 0-5: Figure 0-5a) corresponds to the first inspection, Figure 0-5b) shows the recovery stage at the time of the second survey, and Figure 0-5c) refers to the recovery stage at the time of the third field survey.

(a) Recovery Category '0"   (b) Recovery Category '0"   (c) Recovery Category '1"

Photos by Despotaki Venetia and Christopher Burton.

Figure 0-5. Recovery progress of one building in the city of Napa at (a) 6; (b) 12; and (c) 18 months following the earthquake.

In sum, nearly 83% of the yellow-tagged surveyed buildings were fully recovered after 18 months (~60% within the first six months). On the other hand, approximately 50% of the red-tagged buildings were still not recovered at the end of the three surveys.

Quantification of the Socio-Economic Conditions

In order to quantify the pre-existing socio-economic conditions within the city of Napa, a set of 38 proxy parameters were collected, based on two criteria for their selection: 1) they must be retrievable from publically available data sources; and 2) they must have been previously recognized in the literature as being associated with the capacity of a community to recover from an earthquake event. The variables were collected at the census block group level of geography as defined by the U.S. Census Bureau, which provides a relevant proxy for neighbourhoods within an area (Burton C. , 2015). The 5-year estimates provided by the 2013 American Community Survey (ACS) were the primary source of statistical data pertaining to the 38 selected variables (the variables are presented in Table 4-1 in Appendix 4).

Relate Building Recovery and Socio-Economic Conditions

Following the assessment/quantification of the building recovery over time and the assessment/quantification of the pre-existing socio-economic conditions of the city of Napa, the objective is to establish a quantitative relationship between the two (building recovery and socio-economic conditions). Methods such as the multivariate linear regression can be used to describe the relationship between an outcome (dependent variable) that one is interested in predicting (in this case, the building recovery stage), and a set of explanatory (independent) variables known to influence that outcome (in this case, the socio-economic parameters presented in Table 4-1 in the Appendix 4). However, when the dependent variable is dichotomous (such as 0 and 1 in the present case) a multivariate logistic regression model is commonly used (Hosmer & Lemeshow, 2000). According to the logistic regression model, if x_1, ..., x_n are the independent variables and Y is the dependent, the probability of occurrence of the outcome of interest is computed as:

$$P(Y=1) = \frac{e^{(a+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n)}}{1+e^{(a+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n)}} \tag{3-3}$$

where is the intercept of the regression and $_1$ to $_n$ are the regression coefficients corresponding to each of the x_1 to x_n independent variables.

The 38 indicators presented in Table 4-1 in Appendix 4 constitute the independent variables of the logistic regression (x_1, ..., x_{38}) and $P(Y=1)$ represents the probability of a given building (in a certain census block group) being fully recovered. As previously discussed, these variables were used as proxies to represent and quantify the socio-economic conditions of the study area. However, if two particular block groups with similar social conditions are considered, their recovery trajectory may mostly depend on the corresponding extent of the initial observed post-earthquake damage. As a result, the 38 socio-economic parameters were complemented with a 39th independent variable, the Modified Mercalli Intensity (MMI) (Wood & Newmann, 1931) observed in each block group, which was considered as a measure of the seismic induced damage. The distribution of the MMI was acquired from the United States Geological Survey (USGS) ShakeMap platform (U.S. Geological Survey (USGS), 2015). In addition, because the temporal evolution of recovery was also of interest, an independent variable representing the time (T) was included in the dataset, assuming values of 6, 12 and 18 months (which correspond to the time of the three field surveys) (Despotaki, Sousa, & Burton, 2017).

In the next step, the logistic regression was performed and sets of regression coefficients ($_{1}$ to $_{40}$) were produced, which could be used to predict community recovery probabilities of other regions (apart from the city of Napa) by simply updating the values of the independent variables (x_1 to x_n) in Equation 3-3, accordingly. It should be also highlighted that both the mean predicted values of $P(Y=1)$ and associated uncertainty were calculated for each block group, allowing policy-makers to plan for various scenarios. More technical details on the way the regression was performed and the test of the goodness-of-fit and accuracy of the model are presented in Appendix 4.

Discussion of Results

The recovery probabilities provided by the Socio-Economic Recovery Model can be understood as the likelihood of a building located in a specific census block group, being fully recovered at some time *T* after the earthquake and can be obtained using Equation 3-3. The sets of the produced regression coefficients ($_{1}$ to $_{40}$) could be used to generate recovery probabilities for other areas by updating the values of the independent variables of the regression, accordingly.

The Socio-Economic Recovery Model could further be used to predict the evolution of community recovery at any time in the future. To exemplify, Figure 0-6 illustrates the recovery probabilities in the city of Napa at T=24 and 30 months following the event (for which actual recovery data were not obtained) by replacing the values of the variable representing the time (*T*), accordingly.

a) b)

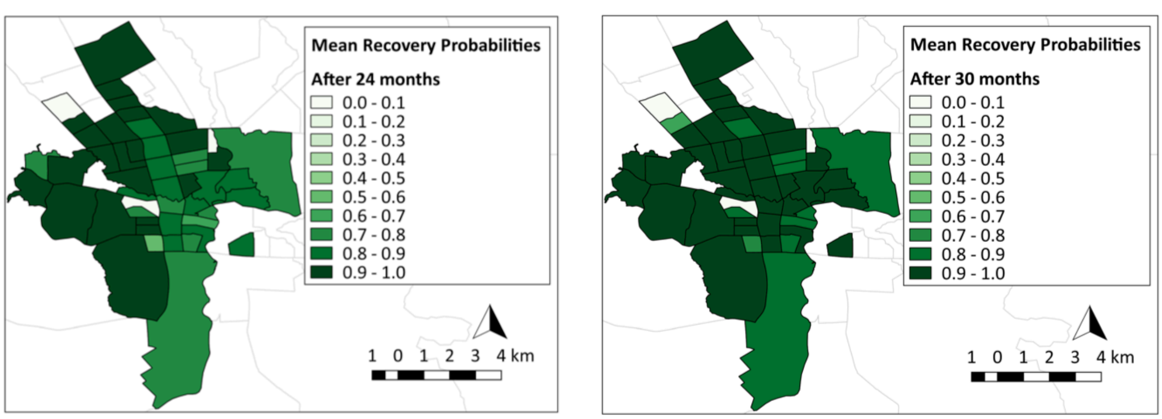


Figure 0-6. Mean predicted recovery probabilities at a) 24; and b) 30 months after the earthquake, for the city of Napa.

(Despotaki, Sousa, & Burton, 2017).

In addition, it is possible to use the Socio-Economic Recovery Model to identify which of the variables used in the analysis have the most significant effect on the recovery trajectory (more technical details on the significance test can be found in Appendix 4).

In the case of the city of Napa, the indicators found to influence the trajectory of the recovery to the greatest extent are listed below in order of importance, with the level of building damage being the most significant:

a. Level (or amount) of building damage
b. Homeownership
c. Percentage of households that have a male householder
d. Presence of health insurance coverage
e. Employment status
f. Percentage of households that have any type of available income
g. Percentage of buildings constructed after 1950 (which are considered to be seismically designed and consequently behave better during an earthquake)
h. Percentage of English speaking households

The influence of homeownership is justified by the fact that homeowners are usually in a better financial condition to respond to the impacts of a devastating earthquake (Burton C. , 2015) and they are more likely to have better access to government assistance programs (Comerio M. C., 1997). In addition, as it was verified during the field surveys, homeowners are more sentimentally attached to their homes, which provide a further incentive to promptly initiate recovery activities (Despotaki, Sousa, & Burton, 2017).

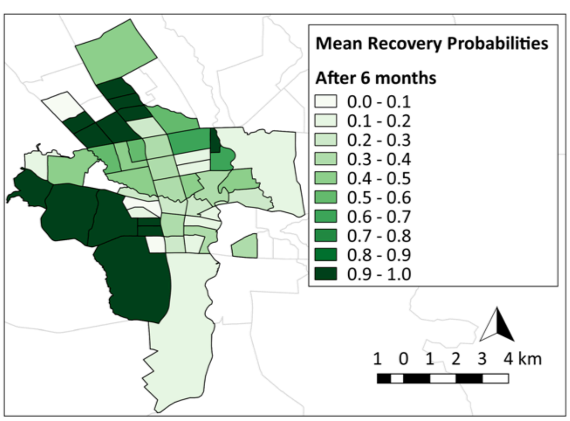
Particularly in the city of Napa, where the employment rate is high and most of the residents have some type of available income, many homeowners paid for repairs from their personal resources and savings. This contributed to a more expedite recovery, since they did not have to wait until other recovery funding resources were made available. Similarly, the presence of health insurance facilitates the recovery processes, since access to medical care is an important post-event source of relief.

Populations that do not speak English, or that speak English as a second language, like many residents of the Latino communities in the city of Napa, constitute a rather vulnerable group following a disaster. In the particular case of Napa, the municipal government faced many challenges in identifying and documenting housing damage and estimating monetary losses. One of the main factors contributing to this issue was that some people with English as a second language were particularly reluctant to report building damage due to lack of trust in authorities, possibly for fear of losing their housing rights.

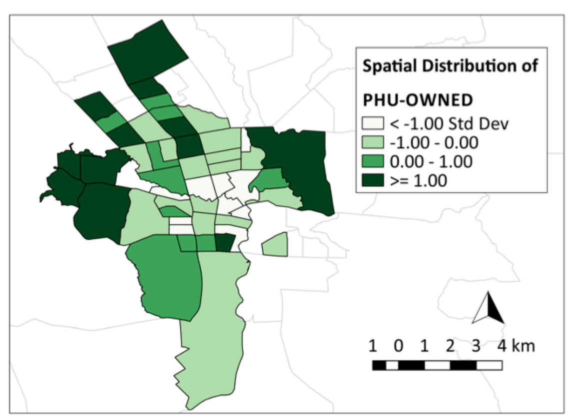
Despite this limitation, the city of Napa is a resourceful and resilient community (Rabinovici, 2017), which significantly contributed to the overall timely and satisfactorily pace of recovery, as shown in Figure 0-7, where the mean recovery probabilities provided by the Socio-Economic Recovery Model at 6, 12 and 18 months after the earthquake are illustrated (Figure 0-7a, b and c, respectively).

In this figure, it is possible to verify that the recovery is constantly progressing over time, albeit with a slower pace in the central and south-eastern block groups. This pattern is in accordance with the spatial distribution of the aforementioned significant variables (identified by the model), according to which lower values of the variables (which are known to hinder recovery processes) are also found in the central, south-eastern block groups. This trend is evident in Figure 0-7, where for illustration purposes, the spatial distribution of three significant variables is mapped; specifically, the percentage of homeownership (Figure 0-7d - PHU-OWNED), the percentage of the population with any type of health insurance (Figure 0-7e - PHEALTHINS), and the percentage of English speaking households (Figure 0-7f - PH-ENGLISH). Indeed, it can be observed that the spatial distribution of these variables is in good agreement with the plotted probabilities of recovery, with lower values in areas where a slower recovery is taking place (Despotaki, Sousa, & Burton, 2017).

a)



d)



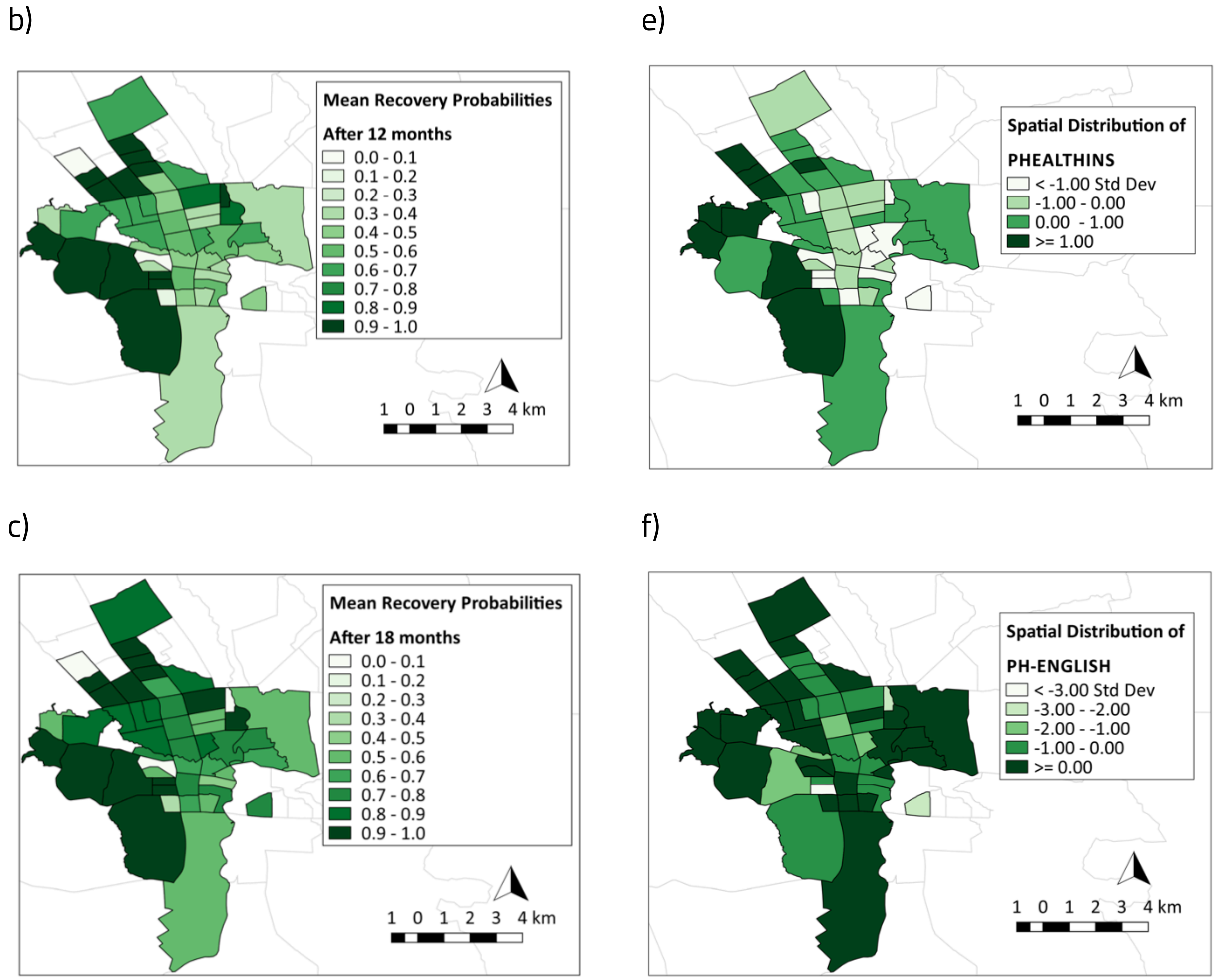


Figure 0-7. a, b, c) Mean predicted recovery probabilities at 6, 12 and 18 months after the event; d, e, f) the spatial distribution of the percentage of homeownership, the percentage of population with any type of health insurance and the percentage of English speaking households across the city of Napa, classified in standard deviation.

(Despotaki, Sousa, & Burton, 2017).

Socio-Economic Recovery Model Results "before" and "after" taking action

It is of great importance for decision-makers to be able to test how different resilience actions influence the recovery trajectory and times following an earthquake. The earthquakes cannot be predicted and their intensity cannot be controlled, thus stakeholders and community residents need to be prepared to effectively response to potential catastrophic events and focus on improving conditions that will facilitate a speedy return to normality.

As previously discussed, the Socio-Economic Recovery Model provides community recovery predictions at different time instances in the future, based on a set of parameters that describe the socio-economic conditions of the affected area and a variable that represents the level of the physical post-earthquake damage. The model further identifies which are the factors that most significantly affect the recovery, providing valuable information to decision makers on the social conditions and potential vulnerable groups of their communities that require special attention following an event.

By experimenting with different values of the parameters that are used as predictors of recovery (Table 4-1 in Appendix 4) and running the Socio-Economic Recovery Model again, stakeholders could identify to what extent and how measures to increase community resilience affect recovery times. To exemplify, it can be identified what is the impact of increasing the health insurance penetration and/or what is the impact of increasing/improving a combination of parameters on the recovery trajectory. In addition, it is possible to change the values of the variables only for specific communities and subsequently identify the overall impact of these changes to the area of interest. Thus, depending on the scope of the analysis, the model provides the opportunity to test various scenarios and define appropriate resilience interventions.

In order to better understand how the model and its predictions could be used in such a way, an exercise using the building recovery in the city of Napa was conducted. For the purposes of this exercise, two of the total identified (by the Socio-Economic Recovery Model) significant variables, were equally increased across all the communities. More specifically, the percentage of the population with any type of health insurance and the employment rate were increased by 10% for each census block group, while all the remaining 38 variables (40 in total) of the model were left as before. Both variables have been previously verified in the literature to facilitate recovery processes, e.g., (H. John Heinz III Center for Science, 2002), (Cutter, Burton, & Emrich, 2010). Following, updated recovery probabilities were generated at 6, 12 and 18 months following the earthquake, as illustrated in Figure 0-8.

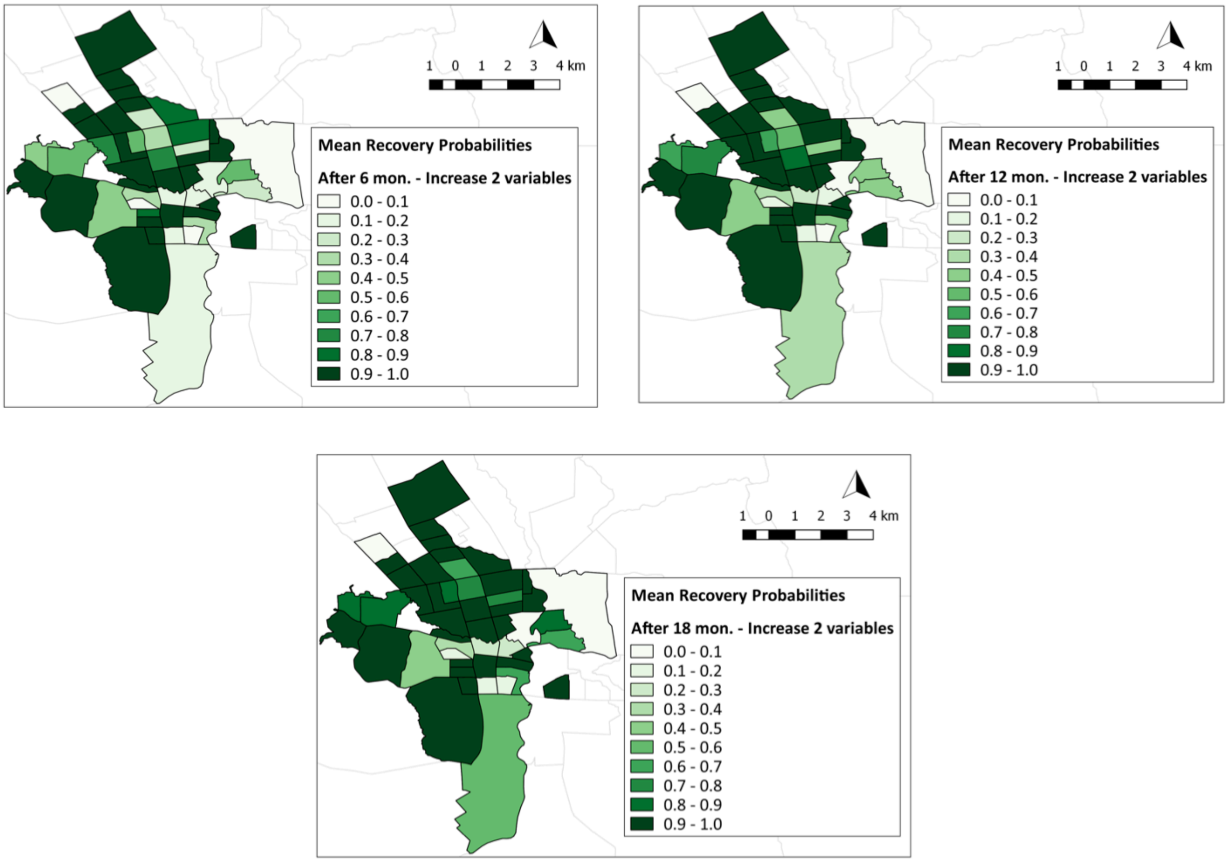


Figure 0-8. Mean recovery probabilities of the city of Napa, after 6, 12, and 18 months following the earthquake, by increasing the percentage of the population with health insurance and the employment rate by 10%, for each block group within the city.

By comparing Figure 0-8 with the respective recovery probabilities illustrated in Figure 0-7, it can be observed that the 10% increase of the two parameters seems to have a crucial positive impact on the recovery progress of the city. Specifically, after 6 months of the event, an overall average increase of nearly 30% of the recovery probabilities is observed (i.e., from x percent to y percent), an average of nearly 12% increase after 12 months (i.e., from x percent to y percent) and an increase of approximately 1% after 18 months, compared to the initial model. This kind of analysis could provide stakeholders valuable insights into the effect of various actions and decisions and assist them in properly identifying the measures/interventions that bring the most successful results.

At this point, it is important to highlight that the model and its predictions need to be further validated and tested in different regions and earthquake events, before these results are used for policy and decision making.

4 Applications of the Reconstruction and Socio-economic Recovery Models

This section describes the application of the Reconstruction Recovery Model to different earthquake scenarios. A case study for the Southern California ShakeOut Scenario is described and conclusions on applying the model to a large-scale earthquake are discussed. Following, a comparison of the results of the Reconstruction Recovery Model and the Socio-Economic Recovery Model is presented. Both models were applied to estimate community recovery times of the residential building stock of the city of Napa following the 2014 M6 South Napa Earthquake.

Southern California Shakeout Scenario

The Reconstruction Recovery Model was applied using the Southern California ShakeOut Scenario and considering each of the counties in the Scenario as a 'community snapshot" (namely: Riverside, Orange, Kern, San Diego, San Bernardino, Ventura, Los Angeles and Imperial). The ShakeOut is an exercise based on a potential magnitude 7.8 earthquake on the southern San Andreas Fault with the goal to identify the physical, social and economic consequences of a major earthquake in southern California (Jones, et al., 2008).

The IRMT was used to develop an aggregated recovery curve for each of the aforementioned counties. Initially, the ShakeOut Earthquake was simulated and damage distribution statistics were computed for all the residential buildings across the eight counties, by running a Scenario Damage Assessment. The latter is a type of analysis supported by the risk component of the OpenQuake-engine (Silva, Crowley, Pagani, Monelli, & Pinho, 2014). The building exposure data that were used in the Scenario Damage Assessment were obtained from the 'Beyond button Pushing – Seismic Risk Assessment for California" project of GEM supported by the California Seismic Safety Commission. Additional information on this exercise can be found in Appendix 3.

However, when the Reconstruction Recovery Model is applied at the county scale, the single aggregated recovery curve may not provide realistic estimates of recovery durations. Partially this is because: a) the lack of detailed building exposure information required the data to be aggregated at a coarse level; and b) the model was not calibrated for large-scale earthquakes, where various factors, such as shortage of labor or materials and loss of critical infrastructure, might delay the recovery process.

2014 M6 South Napa Earthquake

The Reconstruction Recovery Model was applied in order to estimate the recovery time of the residential building stock of the city of Napa following the earthquake. Then, aggregated recovery curves for each census block group were developed, using the IRMT previously described. The city of Napa, the census block group subdivisions and the spatial distribution of the residential buildings is presented in Figure 0-1a, while Figure 0-1b illustrates an example of the recovery curves for a specific block group (zone 5042) produced by the two models (Reconstruction and Socio-Economic Recovery Model). It can be observed that the Reconstruction Recovery Model methodology underestimates the recovery time relative to the empirically-based Socio-Economic Recovery Model. This difference highlights that external socio-economic parameters have a considerable effect on the recovery time and should be taken into account in recovery modelling frameworks.

The Reconstruction Recovery Model considers the effect of the amount of the physical damage in a detailed and robust way (as described in Section 3 and further in Appendix 1), but only implicitly accounts for the effect of the socio-economic conditions on the recovery times by adjusting the lead time for recovery. For example, the buildings' assessment time could be increased in order to capture a possible lack of engineering staff in the study area. On the other hand, the Socio-Economic Recovery Model was developed and further validated based on actual earthquake recovery data and provides community recovery predictions using socio-economic parameters; this model implicitly accounts for the level of physical damage using MMI as a proxy. In the case of the city of Napa, the spatial distribution of MMI proved to be a good indicator of the post-earthquake induced damage, but this may not be the case in other regions. Thus, further research and long-term recovery studies are required in order to merge these 2 models into a hybrid one, in which both physical damage and socio-economic parameters are taken into account directly and in a robust and detailed way.

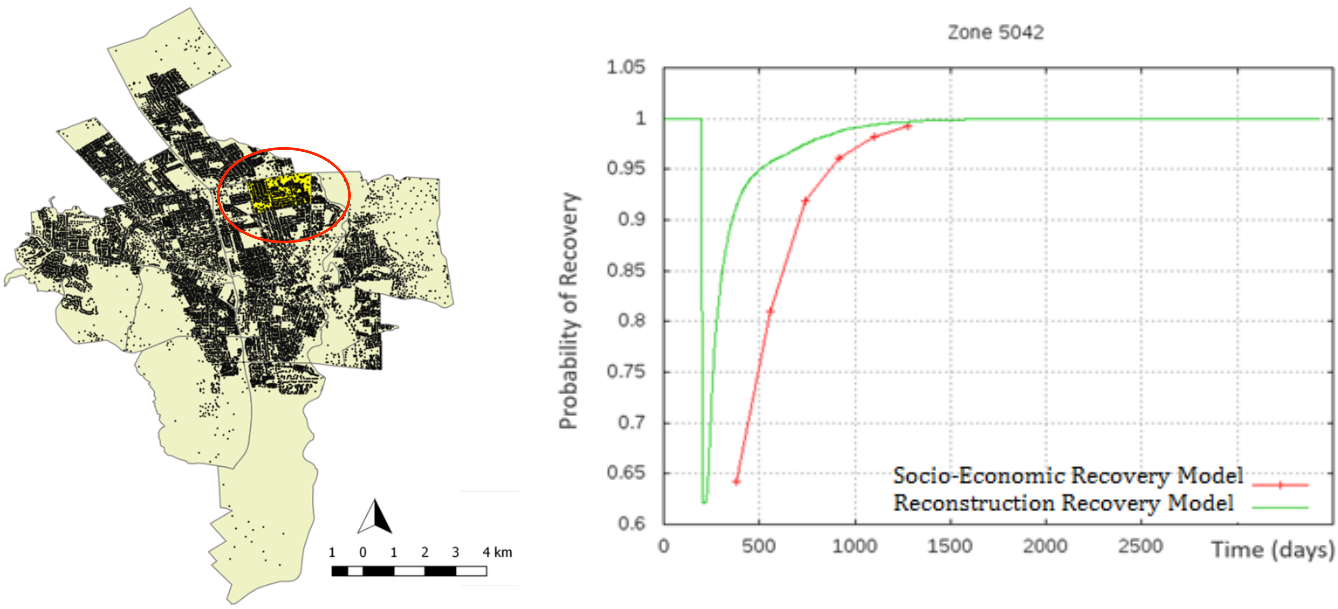


Figure 0-1. (a) The city of Napa and the spatial distribution of the residential buildings; (b) Recovery curves developed by the 2 models for the zone highlighted in yellow.

5 Known Limitations

Reconstruction Recovery Model

The use of the default parameters (such as the assessment times, inspection times and mobilization times) provided currently by the IRMT would not be appropriate for application to potential earthquakes in other regions without additional post-disaster studies. More research is needed to better define the reconstruction times of various building types and at different levels of damage, in order to improve the reliability of model.

Socio-economic Recovery Model

The methodology was developed using the city of Napa and the 2014 M6 South Napa Earthquake as a real-world case study. Therefore, it is expected to be applicable to communities with similar socio-economic and building characteristics (such as other California cities). In order to apply the proposed model in different regions, additional data collection and validation would be necessary.

The selection of the socio-economic variables used in the analysis and their potential to predict recovery outcomes was based on previous research (e.g., Cutter et al., 2010) demonstrating their influence on the recovery of different communities. However, the methodology is flexible and can accommodate different parameters as predictors of recovery, such as the earthquake insurance penetration rate, which could be a very significant factor in the recovery process. In the case of the city of Napa, earthquake insurance was not considered in the analysis, because data were not available at the desired resolution nor was there knowledge of the dates that the insurance claims were approved.

Additional Limitations

There are other factors that influence the recovery process, in areas of high damage (a.k.a., the red zone), building permit processes, legal and financial aspects of demolition, debris management, and insurance availability.

The government policy of the red zone will depend on the degree of damage after the earthquake and decisions taken by emergency managers in this respect. In some cases, one or more zones may be cordoned off while debris is cleaned, heavily damaged buildings are demolished or buildings are repaired or reconstructed. However, the time should be minimized in order, to avoid a traumatic relocation process that goes from temporal to permanent, without considering the breaking of pre-existing collaborative networks (Contreras, Blaschke, & Hodgson, 2017). Another undesirable effect is the breaking down of the urban fabric and creation of marginal zones into the city (Contreras, Blaschke, Kienberger, & Zeil, 2013).

The amount of debris will be also a function of the degree of damage. Taking into account the urban morphology of Napa city, which is mainly made up of buildings with no more than 3 stories, debris was not a critical problem in this earthquake. The implementation of either efficient mitigation plans before the earthquake, or a debris management plan after can prevent the need for a red zone or at least reduce the period when it needs to be established.

There are new tendencies in debris management such as the pilot program released by Federal Emergency Management Agency (FEMA) in 2007, which provides incentives for communities to recycle by allowing them to retain revenue from the sale of disaster debris (Fetter & Rakes, 2012). Xiao et al. (2012) proposed a mathematical model to forecast the amount of building waste, considering the characteristics of the buildings to suggest potential reuse and recycling.

Debris management will influence the speed of the recovery process, as well as administrative processes such as permits and legal and financial aspects of demolitions, which could also be addressed either by the Emergency Response Plan of the city or a pre-impact recovery plan. Complex legal requirements for management of waste hampers efforts to quickly remove, recover and dispose of the waste, delaying the progress of the recovery process (Brown, Milke, Seville, & Giovinazzi, 2010).

Another aspect that contributes to the recovery, is insurance availability. In Napa only 6% of houses were covered by earthquake insurance (Farr & Respaut, 2014). Although, proper data about insurance penetration in the city of Napa was requested it was not possible to get sufficiently detailed information due to privacy restrictions. The results of the 'Back to Normal" report show where earthquake risks are greatest, and therefore where risk mitigation and risk transfer such as through insurance would be most effective. This information could be used to promote the acquisition of insurance not only for reconstruction of houses but also to cover losses for business interruption due to earthquakes (Business interruption insurance), e.g. in the wineries closest to Napa and wine shops in the city (Campbell, 2014); (Farr & Respaut, 2014). Thus, information on insurance coverage would be useful for future planning.

6 Key Findings

Finding 1: The Reconstruction Recovery and Socio-Economic Recovery methodologies were applied to estimate community recovery times using data from the city of Napa and recovery from the 2014 M6 South Napa Earthquake. By comparing the outputs, it was found that the recovery is significantly influenced by pre-existing socio-economic factors, which should be included in recovery prediction models to obtain more realistic recovery predictions.

Finding 2: The Reconstruction Recovery Model was applied using the Southern California ShakeOut Scenario and considering each of the counties in the Scenario as a 'community snapshot" (namely: Riverside, Orange, Kern, San Diego, San Bernardino, Ventura, Los Angeles, Imperial). However, when the model is applied at the county scale, the single aggregated recovery curve may not provide realistic estimates of recovery durations. Partially this is because: a) the lack of detailed building exposure data required the data to be aggregated at a coarse level; and b) the model was not calibrated for large-scale earthquakes, where various factors, such as shortage of labor or materials and loss of critical infrastructure, might delay the recovery process.

Finding 3: Using the Socio-Economic Recovery Model for the city of Napa, it was determined that the level (or amount) of earthquake building damage together with seven socio-economic variables contribute most to the prediction of building recovery. These variables are listed below **in order of importance**, with the level of damage being the most significant factor that influences recovery:

a. Level (or amount) of building damage
b. Homeownership
c. Percentage of households that have a male householder
d. Presence of health insurance coverage
e. Employment status
f. Percentage of households that have any type of available income
g. Percentage of buildings constructed after 1950 (which are considered to be seismically designed and consequently behave better during an earthquake)
h. Percentage of English speaking households

It should be noted that the earthquake insurance penetration rate was not included in the model because data was not available at a sufficient spatial resolution (i.e., block group level rather than zip code), nor were dates available associated with the insurance approval process.

Finding 4: Based on qualitative analyses, such as personal interviews, it was found that in the city of Napa pre-existing earnings and wealth of the population were the main sources and drivers of recovery. Since earthquake insurance penetration in the city was very low and there was a considerable delay in the authorization of the federal Individual Assistance program, most of the homeowners initiated the payment of repairs from their personal resources and savings. This qualitative finding is consistent with socio-economic factors b, d, e, and f above.

Finding 5: Populations that do not speak English, or that speak English as a second language, such as residents of the Latino communities in Napa, constitute a rather vulnerable group following a disaster. The municipal government of Napa faced many challenges in identifying and documenting housing damage and estimating monetary losses. One of the main factors contributing to this issue was that some people with

English as a second language were particularly reluctant to report building damage due to lack of trust in authorities, possibly for fear of losing their housing rights. This lack of trust may have significantly affected the progress of recovery, and is therefore consistent with socio-economic factor h above.

Finding 6: Based on the field surveys and the documentation of the building recovery in the city of Napa, it was observed that about 60% of the yellow-tagged buildings were fully recovered within the first six months, and almost 83% were fully recovered after 18 months. On the other hand, nearly 50% of the red-tagged buildings were still not recovered 18 months after the earthquake. In addition, nearly 80% of the surveyed structures that were not seismically designed sustained significant damage and were assigned a red tag.

7 Recommendations

Recommendation 1: More long-term recovery studies from other earthquakes are required to refine both the Reconstruction Recovery Model and the Socio-Economic Recovery Model. It is recommended to develop a robust tool for earthquake recovery modelling and planning by merging the 2 models and implementing them into the IRMT. In addition, more long-term recovery studies are essential to identify and investigate indirect losses following earthquakes.

Recommendation 2: Extend the methodologies beyond residential buildings to model recovery of critical facilities, such as hospitals, fire and police stations, power plants, water treatment plants and telecommunication networks.

Recommendation 3: Based on the identified significant variables from the Socio-Economic Recovery Model that positively contribute to the recovery process, it is recommended to:

1. Facilitate access to assistance for vulnerable groups of the population, such as residents that do not speak English.
2. Conduct further investigations into the relationships between the variables that correlate most positively with recovery (e.g., homeownership and health insurance) to determine the underlying causes.
3. Conduct more extensive research on cost-benefit analysis of retrofitting buildings because the buildings not seismically designed in the city of Napa sustained significantly more damage compared to stronger structures.
4. Improve access to financial mechanisms, such as earthquake insurance, to residents

exposed to high earthquake risk, as well as investigate and promote alternative post-earthquake resources, such as grants, which will support residents in the rebuilding process.

Recommendation 4: Facilitate the involvement of insurance industry partners in future projects as advisors to improve the model and to gain access to more detailed earthquake insurance data. This would facilitate, for instance, a better understanding of the degree to which access to earthquake insurance influences the recovery process.

8 References

ATC. (2005). *Field Manual: Post Earthquake Safety Evaluation of Buildings ATC-20-1, Second Edition.* Redwood City, California: Applied Technology Council.

Brown, C., Milke, M., Seville, E., & Giovinazzi, S. (2010). Disaster Waste Management on the Road to Recovery: L'Aquila earthquake case study. *14ECEE2010.* Republic of Macedonia.

Brown, D., Platt, S., & Bevington, J. (2010). *Disaster Recovery Indicators: guidelines for monitoring and evaluation.* CURBE.

Burton, C. (2015). A Validation of Metrics for Community Resilience to Natural Hazards and Disasters Using the Recovery from Hurricane Katrina as a Case Study. *Annals of the Association of American Geographers, 105*(1), 67-86.

Burton, H., Deierlein, G., Lallemant, D., & Lin, T. (2015). Framework for Incorporating Probabilistic Building Performance in the Assessment of Community Seismic Resilience. *J.Struct.Eng.* doi:10.1061/(ASCE)ST.1943-541X.0001321

Campbell, C. (2014, August 26). *The California Quake May Cost Wine Country Billiion.* Retrieved from TIME: http://time.com/3180988/california-quake-economic-napa-valley-wine-country-billions/

Chang, S. (2009). Urban disaster recovery: a measurement framework with application to the 1995 Kobe earthquake. *Disasters, 34*(2), 303-327.

Comerio, M. (2006, May). Estimating downtime in loss modeling. *Earthquake Spectra, 22*(2), 349-365.

Comerio, M. C. (1997, September). Housing Issues After Disasters. *Journal of Contingencies and Crisis Management, 5*(3), 166-178. doi:10.1111/1468-5973.00052

Contreras, D., Blaschke, T., & Hodgson, M. (2017). Lack of spatial resilience in a recovery process: Case of L'Aquila, Italy. *Technological Forecasting and Social Change*. doi:http://dx.doi.org/10.1016/j.techfore.2016.12.010

Contreras, D., Blaschke, T., Kienberger, S., & Zeil, P. (2013). Spatial connectivity as a recovery process indicator: The L'Aquila earthquake. *Technological Forecasting and Social Change, 80*(9).

Contreras, D., Blaschke, T., Kienberger, S., & Zeil, P. (2014). Myths and realities about the recovery of L'Aquila after the earthquake. *International Journal of Disaster Risk Reduction, 8*(0), 125-142. doi:doi:http://dx.doi.org/10.1016/j.ijdrr.2014.02.001

Cutter, S., Burton, C., & Emrich, C. (2010). Disaster Resilience Indicators for Benchmarking Baseline Conditions. *Journal of Homeland Security and Emergency Management, 7*(1), 1-22.

Despotaki, V., Sousa, L., & Burton, C. (2017). Using resilience indicators in the Prediction of Earthquake Recovery. *Under Review in Earthquake Spectra*.

Farr, C., & Respaut, R. (2014, August 25). *California wine country quake losses seen in the billions*. Retrieved from Reuters: http://reut.rs/2syTJ1O

Fetter, G., & Rakes, T. (2012). Incorporating recycling into post-disaster debris disposal. *Socio-Economic Planning Sciences, 46*(1), 14-22.

H. John Heinz III Center for Science, E. a. (2002). *Human Links to Coastal Disasters.* Washington, DC: H. John Heinz Center.

Hosmer, D., & Lemeshow, S. (2000). *Applied Logistic Regression, Second Edition.* United States of America: John Wiley & Sons, Inc.

Jones, L., Bernknopf, R., Cox, D., Goltz, J., Hudnut, K., Mileti, D., . . . Wein, A. (2008). *The ShakeOut Scenario.* U.S. Geological Survey OpenFile Report 2008-1150 and California Geological Survey Preliminary Report 25.

Mitrani-Reiser, J. (2007). *An ounce of prevention: Probabilistic loss estimation for performance-based earthquake engineering.* Ph.D. dissertation, Department of Applied Mechanics, California Institute of Technology, Pasadena, CA.

Poland, C. (2009, February 1). *The resilient city: Defining what San Francisco needs from its seismic mitigation policies*. Retrieved from http://www.spur.org/featured-project/resilient-city

Rabinovici, S. (2017). *South Napa Home Impact Study.* Sacramento, California: California Earthquake Authority.

Silva, V., Crowley, H., Pagani, M., Monelli, D., & Pinho, R. (2014). Development of the OpenQuake engine, the Global Earthquake Model's open-source software for seismic risk assessment. *Natural Hazards, 72*(3), 1409-1427.

U.S. Geological Survey (USGS). (2015). MMI ShakeMaps for the 2014 South Napa Earthquake. Retrieved from: https://on.doi.gov/2ruK6md

Wood, H., & Newmann, F. (1931). Modified Mercalli Intensity scale of 1931. *Bull. Seism. Soc. Am., 21*, 277-283.

Xiao, J., Xie, H., & Zhang, C. (2012). Investigation on building waste and reclaim in Wenchuan earthquake disaster area. *Resources, Conservation and Recycling, 61*(0), 109-117. doi:http://dx.doi.org/10.1016/j.resconrec.2012.01.012

APPENDIX 1 - Reconstruction Recovery Model

This appendix presents the scientific recovery modelling framework that was developed to quantify the effectiveness of specific resilience-building actions (preparedness, mitigation, and response) that would increase the speed of recovery following an earthquake. The overall approach is based on the methodology developed by Burton et al. (2015) and consists of four main components: (1) recovery-based limit state fragility function development, (2) developing building-level time dependent recovery functions, (3) accounting for the effect of externalities and socio-economic vulnerability and (4) developing community/regional level recovery functions. These are discussed in detail in the following subsections.

1.1. Recovery-Based Fragility Function Parameters

A rigorous evaluation of seismic resilience requires probabilistic methods for assessing limit states that influence post-earthquake functionality, which can be incorporated in modelling the recovery of the building stock. The methodology incorporates a set of building performance limit states that specifically inform community seismic resilience (Figure 1-1). These limit states have been adapted from the building performance categories defined by the San Francisco Planning and Urban Research Association (SPUR) (Poland et al. 2009). They include (i) damage triggering inspection, (ii) occupiable damage with loss of functionality, (iii) unoccupiable damage, (iv) irreparable damage and (v) collapse. These limit states are different from those that are currently used in OpenQuake and other risk modelling platforms. This sub-section intends to document the methodology used to map the fragility function parameters from the loss-based limit states used in OpenQuake to those of the recovery-based limit states used to model recovery.

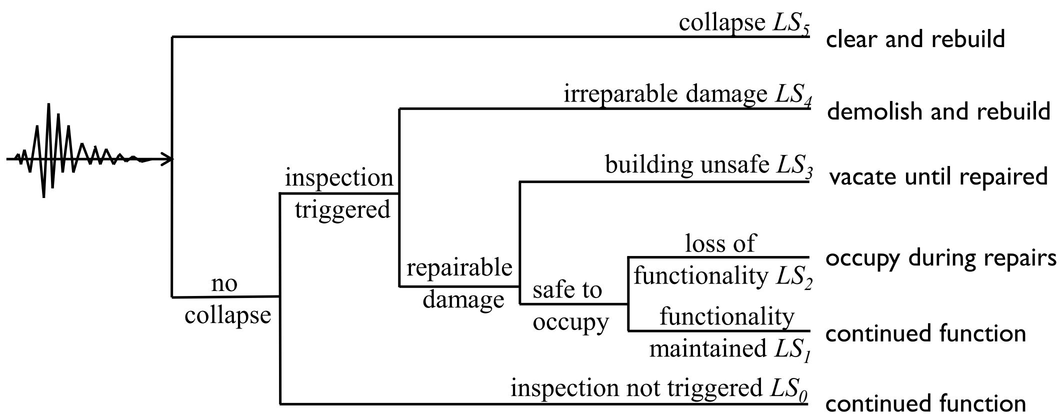


Figure 1-1. Event tree showing building performance limit states and recovery actions.

1.2. Recovery-Based Building Performance Limit States

Five discrete limit states (*LS_0 through LS_5*) are used, which are explicitly linked to post-earthquake recovery-related activities. Each limit state is associated with a unique combination of the following consequent actions to restore building function:

- Assessment and planning activities i.e. post-earthquake inspection and/or evaluation, preparation of plans and designs, financing and bidding preparation for construction work;
- Repairs needed to make building occupiable and repairs needed to restore functionality for repairable buildings; and
- Demolition and building replacement for non-repairable buildings.

LS_0 - Damage below the threshold that would trigger inspection.

LS_1 - Damage Triggering Inspection with Functionality Maintained: This represents the minimum damage threshold that would require post-earthquake inspection and/or evaluation. It is also used to imply a level of damage where, despite the need for post-earthquake inspection, the structural safety and critical subsystems essential to the functionality of the building are not compromised. However, operations may be impacted if the owner/operator decides to close the facility until inspections are completed. This decision is prompted by visible damage to structural (cracking of concrete members) or non-structural elements (partitions, facades etc.). This type of damage occurs at low drift levels and affects structural and non-structural components with low deformation capacities.

LS_2 - Occupiable Damage with Loss of Functionality: This implies that the building is structurally safe, occupiable and accessible but unable to carry out its primary function. This loss of functionality can occur despite the preservation of structural integrity as a result of damage to building systems, non-structural components or contents, which are critical to the operations of the facility. There may also be damage to structural components whose repair actions hinder normal building operations.

LS_3 - Unoccupiable Damage: This infers that the building is either inaccessible or not safe to occupy following an earthquake. The loss of structural safety will likely be due to a substantial loss in the load carrying capacity of the gravity or lateral system that poses a life safety threat in the event of an aftershock. It is also possible but less likely for non-structural damage to compromise the safety or prevent access to the building. This is usually in the form of some type of falling hazard (e.g. brick façade or infill panels); however, these types of dangers can be mitigated in a short period of time. LS_3 is of particular importance to residential buildings as it is directly related to the shelter-in-place performance goal emphasized in SPUR's resilient city initiative.

LS_4 - Irreparable Damage: LS_4 pertains to cases where the building is damaged to such an extent that repair becomes technically or cost prohibitive, necessitating demolition and replacement. The three main earthquake-related situations that can lead to demolition include (1) large permanent deformations and story drifts that make repairs unfeasible, (2) direct economic losses that exceed the limit set by insurance providers triggering full-value pay-out leading to complete replacement and (3) damage to key structural components that could significantly impede the repair process.

LS_5 - Collapse: LS_5 is associated with complete or partial collapse, which is generally associated with either excessive lateral deformations (sideways collapse) or the local or global loss of vertical load carrying capacity.

1.3. Loss-Based Building Performance Limit States

Risk modelling platforms such as OpenQuake and HAZUS use limit state fragility functions that relate earthquake ground shaking intensity to building damage. These limit states are used to link ground motion intensity to direct economic losses (vulnerability curves) that result from having to repair or replace damaged buildings. The limit states are classified based on construction type and are described in terms of the type and extent of physical damage to the building. The following is a description of the limit states (taken from HAZUS), which are relevant to wood frame single- and multi-family residential buildings found in Napa and California in general:

- *Slight Damage:* Small cracks in non-structural elements (window, wall intersections, masonry chimneys, masonry veneer and stucco) and slippage in bolted connections.
- *Moderate Damage:* Small cracks across shear wall, large cracks at doors, windows and masonry veneer, topping of tall masonry chimneys, minor slack in diagonal rod bracing and small cracks and split in bolted connections.
- *Extensive Damage:* Large cracks across shear wall plywood joints, large slack at diagonal and broken braces, permanent lateral movement at floors and roof, topping of most brick chimneys, small cracks in foundations, split and/or slippage of sill plates and partial collapse at garage with soft-story configurations.
- *Complete Damage:* Large permanent lateral displacement, may collapse, imminent collapse, some structures slip off foundations, large foundation cracks, 3% total area collapsed, broken brace rod or failed framing connections.

1.4. Methodology for Mapping Fragility Function Parameters from Loss-Based to Recovery-Based Building Performance Limit States

There is an obvious correlation between the loss-based and recovery-based building performance limit states. In both cases, the limit states are discrete, sequential and mutually exclusive with the higher limit states being associated with more extensive damage. This obvious link was used as the basis for mapping the fragility function parameters between the two types of limit states.

The fragility function for each of the loss-based limit states is assumed to take on a lognormal distribution and is defined by the following relationship (Singhal & Kiremidjian, 1996); (FEMA, 2016):

$$P(DS > ds_i \mid S_d) = \Phi\left[\frac{1}{\beta_{ds_i}}\ln\left(\frac{S_d}{\bar{S}_{d,ds_i}}\right)\right] \qquad 1\text{-}1$$

where $\bar{S}_{d,ds_i}$ is the median value of the spectral displacement at which the building reaches the threshold of damage state ds_i; β_{ds_i} is the standard deviation of the natural logarithm at which the building reaches the threshold of the damage state ds_i and Φ is the standard normal cumulative distribution function. For a given building construction type, HAZUS provides the median spectral displacement and log standard deviation for each of the four loss-based limit states. The inter-story drift at the threshold of each limit state is also provided. In addition to the building construction type, the parameters also vary based on the seismic code design level, which is directly related to the age of the buildings. Four code design levels are included, high-code, moderate-code, low-code and pre-code. The fragility function parameters for the two construction types for wood frame buildings (W1, wood light frame and W2, wood commercial and industrial) can be found in Table 5.9 of HAZUS 3.2 (2016).

The primary objective of this part of the overall study was to map the loss-based fragility parameters shown in Table 5.9 of HAZUS to recovery-based limit state parameters. The first step was to estimate the conditional probability of being in a particular recovery-based limit state given the occurrence of a loss-based damage state. This conditional probability relationship is defined as follows:

$$P(RBDS = rbds_i \mid LBDS = lbds_j) \qquad \text{1-2}$$

where $P(RBDS = rbds_i \mid LBDS = lbds_j)$ is the probability that the recovery-based damage state $(RBDS)_i$ occurs given that the loss-based damage state $(LBDS)_j$ has been observed. Estimates of these conditional probabilities are provided in Table 1-1. The current values are based on engineering judgment. They were obtained by examining the physical description of damage provided for the loss-based limit states and inferring the likelihood that this type of damage would trigger each of the six (LS_0 through LS_5) recovery-based limit states. Later on in the development of the framework, these estimates were refined based on the results from nonlinear response history analyses of typical wood-frame buildings using the *OpenSees* modelling platform. The results from the structural response simulation were used to establish analytical fragility functions for both types of limit states. This process enabled the development of a more explicit relationship between the fragility parameters for the two types of limit states. The conditional probabilities estimates can be further refined using heuristic data obtained from expert opinion. However, this approach is outside the scope of the current project.

In Table1-1, each row provides the probability of being in each of the recovery-based limit states given the occurrence of the loss-based limit state in the first column of that row. For example, it can be observed that for a building that is in the loss-based limit state corresponding to moderate damage, the probability of being in recovery-based limit states LS_0, LS_1, LS_2 and LS_3, is 0.2, 0.4, 0.3 and 0.1 respectively with a zero probability of being in the remaining limit states (LS_4 and LS_5). Given that the recovery-based limit states are mutually exclusive and collectively exhaustive, each row must sum to one.

Table 1-1. Conditional probabilities used to map fragility parameters for loss-based to recovery-based limit states.

Loss-Based Damage States	$P(RBDS = rbds_i \mid LBDS = lbds_j)$					
	LS0 Inspection not Triggered	*LS1* Inspection	*LS2* Loss of Functionality	*LS3* Unsafe to Occupy	*LS4* Damaged Beyond Repair	*LS5* Collapse
None	1.0	0.0	0.0	0.0	0.0	0.0
Slight	0.6	0.4	0.0	0.0	0.0	0.0
Moderate	0.2	0.4	0.3	0.1	0.0	0.0
Extensive	0.0	0.0	0.2	0.4	0.3	0.1
Complete	0.0	0.0	0.0	0.0	0.2	0.8

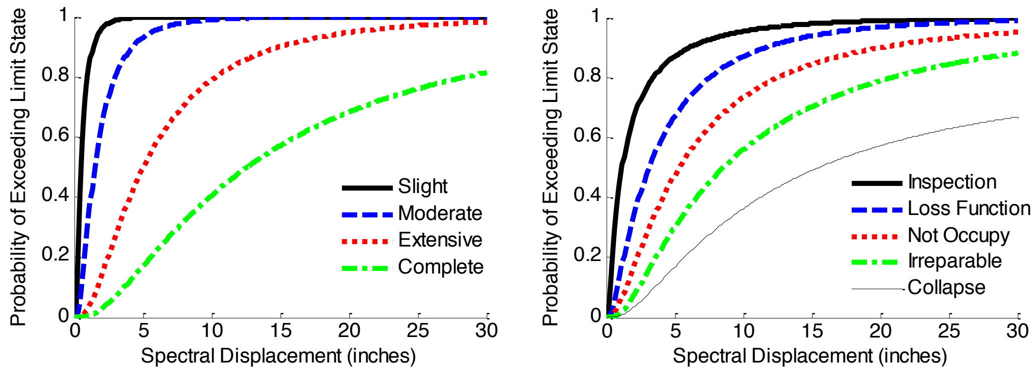
Given the loss-based fragility function parameters in Table 5.9 of HAZUS and the conditional probability estimates in Table 1-1, the probability of occurrence of a particular recovery-based limit state can be obtained using the total probability theorem:

$$P(RBDS = rbds_i \mid S_d) = \sum_{j=1}^{n_{lbds}} P\left(PRBDS = rbds_i \mid LBDS = lbds_j\right) \cdot P\left(LBDS = lbds_j \mid S_d\right) \qquad \text{1-3}$$

where $P(RBDS = rbds_i \mid LBDS = lbds_j)$ is taken from Table 1-1 and $P\left(LBDS = lbds_j \mid S_d\right)$ is obtained from the fragility functions of the loss-based limit states. Given the probability of being in recovery-based limit state i, the probability of exceeding that limit state is taken as the sum of the probabilities of occurrence of all limit states equal to and greater than i.

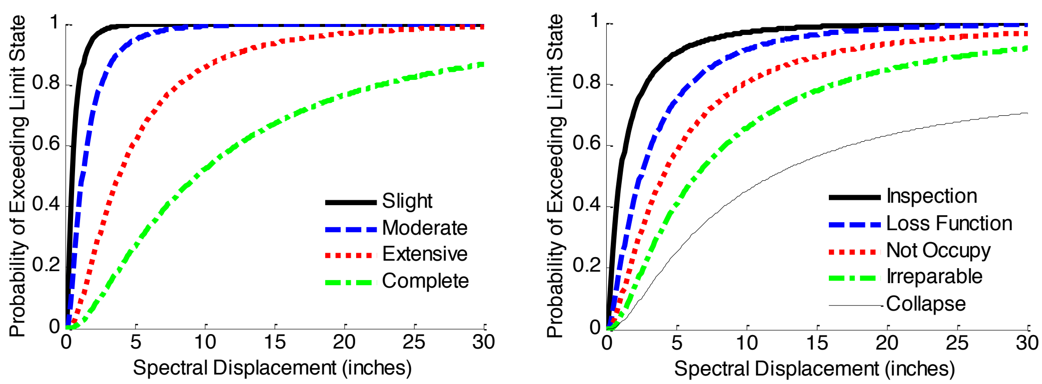
$$P(RBDS > rbds_i \mid S_d) = \sum_{i}^{n_{rbds}} P\left(RBDS = rbds_i \mid S_d\right) \qquad \text{1-4}$$

We can then use equation 1-4 to compute the median spectral displacement, $\bar{S}_{d,rbds_i}$ and dispersion β_{rbds_i} for the recovery-based limit state fragilities. Figure 1-2, Figure 1-3, Figure 1-4 and Figure 1-5 provide a comparison of the recovery- and loss-based fragility functions. They were generated as part of the development of the methodology and reflect the differences between the loss-based and the recovery-based fragility parameters, both of which are built into the new OpenQuake tool. The parameters that define the recovery-based fragility functions for the wood light frame construction type, W1 (all code levels included) is summarized in Table 1-2.



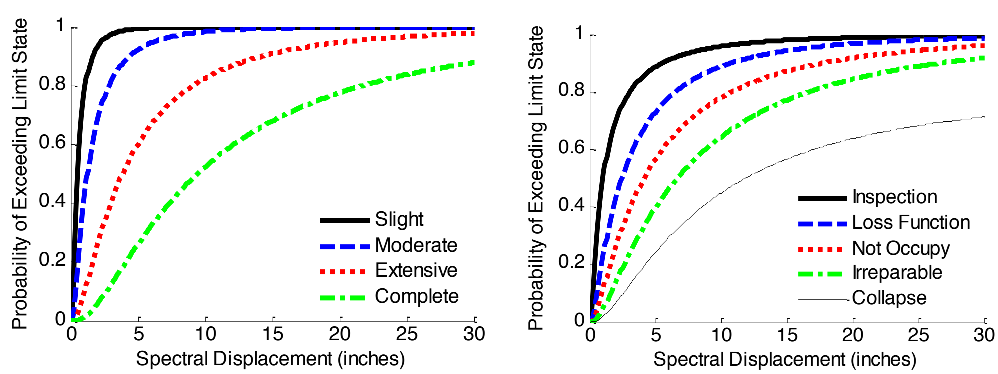
(a) **(b)**

Figure 1-2. Fragility curves for (a) loss-based and (b) recovery based limit states for light wood frame buildings (W1) with high-code seismic design.



(a) **(b)**

Figure 1-3. Fragility curves for (a) loss-based and (b) recovery based limit states for light wood frame buildings (W1) with moderate-code seismic design.



(a) **(b)**

Figure 1-4. Fragility curves for (a) loss-based and (b) recovery based limit states for light wood frame buildings (W1) with low-code seismic design.

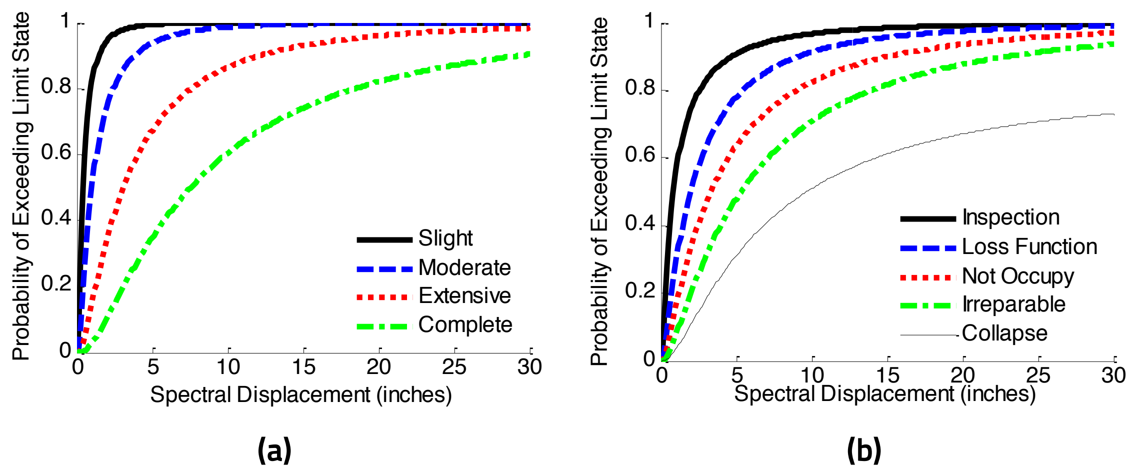


Figure 1-5. Fragility curves for (a) loss-based and (b) recovery based limit states for light wood frame buildings (W1) with pre-code seismic design.

Table 1-2. Fragility function parameters for recovery-based limit states for light woodframe buildings, W1: (a) Median spectral displacement at limit state exceedance and (b) log-standard deviation.

(a)

Code Level	Median Spectral Displacement for Exceeding Limit State				
	LS1 Inspection Triggered	*LS2* Loss of Functionality	*LS3* Unsafe to Occupy	*LS4* Damaged Beyond Repair	*LS5* Collapse
High-Code	1.2	3.1	5.4	8.5	15.6
Moderate-Code	1.1	2.5	4.1	6.4	11.9
Low-Code	1.0	2.5	4.0	6.1	11.9
Pre-Code	0.8	2.0	3.4	5.3	9.8

(b)

Code Level	Log-Standard Deviation of Spectral Displacement at Limit State Exceedance				
	LS1 Inspection Triggered	*LS2* Loss of Functionality	*LS3* Unsafe to Occupy	*LS4* Damaged Beyond Repair	*LS5* Collapse
High-Code	0.80	0.81	0.85	0.97	0.99
Moderate-Code	0.84	0.86	0.89	1.04	1.07
Low-Code	0.93	0.98	1.02	0.99	0.99
Pre-Code	1.01	1.05	1.07	1.06	1.08

1.5. Building-Level Recovery Model

Overview

In the previous section, it was noted that the recovery modelling methodology incorporates a set of building performance limit states that specifically inform community seismic resilience including (i) damage triggering inspection, (ii) occupiable damage with loss of functionality, (iii) unoccupiable damage, (iv) irreparable damage and (v) collapse. The link to post-earthquake recovery is established by defining a characteristic recovery path that is associated with each state and the level of functionality associated with each state. A building recovery function was computed accounting for the uncertainty in the occurrence of each recovery path and its associated limit state. The outcome is a probabilistic assessment of recovery of functionality at the building level for a given ground motion intensity.

1.5.1. Building Recovery Paths

Five distinct recovery paths were defined based on the limit states discussed previously. The recovery paths are described using discrete functioning states and the time spent in each state. The functioning states represent the changing condition of the building with respect to its ability to facilitate its intended operation. The functioning states for modelling the recovery of shelter-in-place housing capacity include (1) the building is unsafe to occupy (*NOcc*), (2) the building is safe to occupy but unable to facilitate normal operations (*OccLoss*) and (3) the building is fully functional (*OccFull*). Note that these three states are specific to the shelter-in-place metric and would need to be re-defined for other measures of functionality. The key to define the functioning states are that (1) they must be explicitly linked to the building level limit states described earlier and (2) each functioning state must be associated with a quantifiable measure of functionality.

The building level recovery path is conceptually shown in Figure 1-6. It is a step function that describes the time spent in each of the discrete functioning states. The recovery path (and recovery function discussed later) is assessed over a pre-defined period of time referred to as the control time, T_{LC}; and T_{NOcc}, $T_{OccLoss}$ and $T_{OccFull}$ denote the time spent in the *NOcc*, *OccLoss* and *OccFull* functioning states, respectively. It is important to note that the functioning states that comprise the recovery path for a given building depend on the limit state of that building immediately following the earthquake. For example, a building that is in limit state LS_1 will only experience the *NOcc* and *OccFull* functioning states. On the other hand, a building that is in limit state LS_2 or LS_3 will experience all three functioning states. This is illustrated later in the discussion of building recovery paths. The time spent in each functioning state will also vary depending on the level of damage. For example, a building that is in limit state LS_4, which must be demolished and rebuilt, will spend a significantly greater amount of time in the *NOcc* state than a building in limit state LS_3, which only requires repairs.

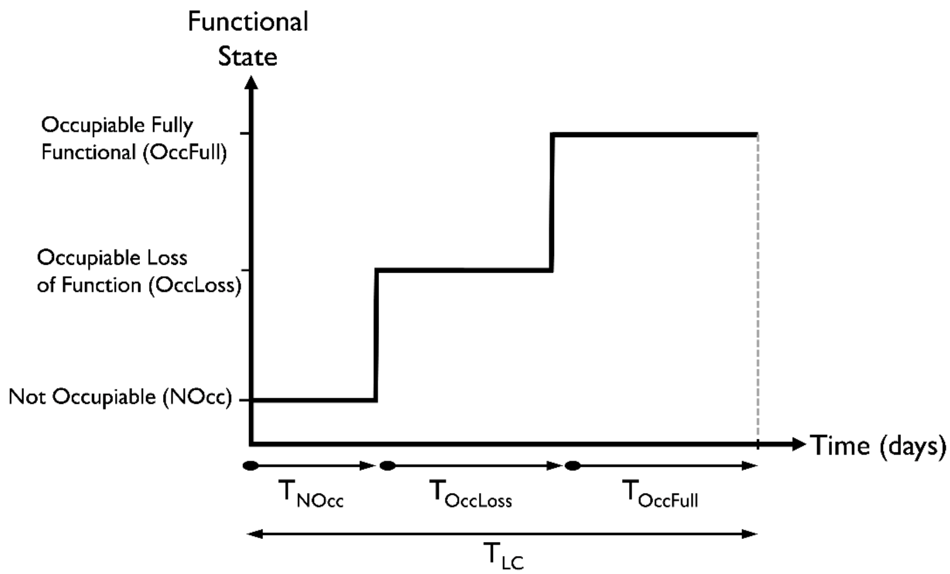


Figure 1-6. Conceptual illustration of recovery path for an individual building. (Burton, Deierlein, Lallemant, & Lin, 2015).

The recovery time for an individual building is defined as the period of time between the occurrence of the earthquake and the restoration of full functionality. The recovery time includes (1) the lead time which is the time required for building inspection and/or evaluation, finance planning, architectural/engineering consultations, a competitive bidding process and to mobilize for construction (Mitrani-Reiser, 2007), (2) the repair time needed to restore occupiability and (3) the repair time needed to restore functionality. The time needed to restore occupiability is taken as the time to complete repairs related to structural safety and internal access, whereas the time needed to restore functionality includes the additional time needed to repair/replace building systems, non-structural components and contents that are essential to the building functionality. Both the lead and repair times for structural and non-structural components depend on the limit state of the building immediately following the event. For example, a building that is in limit state LS_1 following an event (damage triggers inspection but the building is found to be safe to occupy and functional) will likely be green tagged and only be out of service for the time it takes to complete the inspection. On the other hand, a building that is in limit state LS_2 (building is safe to occupy but not functional) may receive a yellow tag, which would require detailed evaluations by a professional engineer prior to re-occupancy. A building that is red tagged (*LS_3*, *LS_4* and *LS_5*) may require demolition or extensive repairs, triggering additional lead time for planning, architectural/engineering consultations, possible competitive bidding and mobilization for construction. Mitrani-Reiser (2007) developed a performance-based approach to estimating repair times for both structural and non-structural damage, which incorporates the lead times for different tagging scenarios as well as the sequencing of repairs. In this study, Mitrani-Reiser's method is used to compute both the repair time needed to restore safety/accessibility and the repair time needed to restore functionality.

The recovery paths for each limit state were derived from the information provided in Table 1-3, which shows the relevant activities and time spent in each functioning state. The recovery paths are described as follows:

- *Recovery Path for LS_0*: This implies that the functionality of the building is not disrupted and the *OccFull* state is maintained throughout the period following the earthquake.
- *Recovery Path for LS_1*: This path is associated with the occurrence of *LS_1* where the extent of damage triggers inspection but does not compromise the functionality of the building. It is comprised of the *NOcc* and *OccFull* states. The time spent in the *NOcc* state is the time to complete inspections. Following inspections, the building is deemed occupiable and fully functional, immediately entering fully functional *OccFull* state.
- *Recovery Path for LS_2*: For *LS_2*, the recovery path includes all three functioning states. Like recovery path 1, the building initially enters the *NOcc* state until inspections are complete. Following inspections, the building enters the *OccLoss* state because, despite being safe to occupy, repairs will be needed to restore functionality. The time spent in the *OccLoss* state is determined by the repair time for those building systems, non-structural components and content that is essential to the building function. Completion of these repairs returns the building to the fully functional *OccFull* state.
- *Recovery Path for LS_3*: The recovery path for *LS_3* also includes all three functioning states. Initially, the *NOcc* state includes the inspection and other lead times, along with the time to complete structural repairs needed to restore occupiability. Since *LS_3* is associated with significant structural and non-structural damage, the lead time will include planning, design consultations, bidding and the time to mobilize

for construction. Following the completion of structural repairs, the recovery will enter the *OccLoss* state during which the repairs needed to restore functionality are completed. The completion of these repairs would return the building to the *OccFull* state.

- *Recovery Path for LS_4*: In LS_4, where the building is irreparably damaged, the recovery path includes the *NOcc* and *OccFull* states, where the *NOcc* state includes the time to demolish and replace the damaged building. As the recovery of this building involves new construction, occupancy is not likely to be restored prior to full completion, which is why this path does not include the *OccLoss* phase.
- *Recovery Path for LS_5*: The recovery path associated with partial or complete collapse is very similar to that of the demolition case, the only difference being that LS_5 would not require any time to assess whether or not the building could or would be repaired. However, this additional time is likely to be insignificant compared to the time needed to replace the building, hence the recovery paths associated with LS_4 and LS_5 are essentially the same.

Table 1-3. Recovery path activities and times for each functioning state.
(Burton, Deierlein, Lallemant, & Lin, 2015).

Recovery Path No.	Time/Acitivies in Functional State		
	NOcc	OccLoss	OccFull
0	0	0	T_{LC}
1	T_{INSP}	0	$T_{LC} - T_{INSP}$
2	T_{INSP}	T_{FUNC}	$T_{LC} - T_{INSP} - T_{FUNC}$
3	$T_{INSP} + T_{ASMT} + T_{MOB} + T_{OCC}$	T_{FUNC}	$T_{LC} - T_{INSP} - T_{ASMT} - T_{MOB} - T_{OCC} - T_{FUNC}$
4	$T_{ASMT} + T_{MOB} + T_{REP}$	0	$T_{LC} - T_{ASMT} - T_{MOB} - T_{REP}$
5	$T_{MOB} + T_{REP}$	0	$T_{LC} - T_{MOB} - T_{REP}$

T_{INSP} - Time to complete inspections
T_{FUNC} - Time to restore functionality
T_{ASMT} - Time to conduct engineering assessment
T_{MOB} - Time to moblize for construction
T_{OCC} - Time to complete repairs needed to restore occupiability/structural safety
T_{REP} - Time to replace building

1.5.2. Probabilistic Assessment of Recovery of Functionality at the Building-Level

Each functioning state can be linked to a quantifiable level of functionality. The functionality will typically be specified based on building owner/stakeholder and community resilience needs. For example, in the case of residential buildings in a community, where, from the perspective of the policy-makers, functionality is measured by housing capacity or number of persons housed. Where loss of certain building functions would not preclude short-term shelter-in-place requirements, the shelter-in-place functionality could be determined by assuming the full pre-earthquake housing capacity is achieved for both the *OccFull* and *OccLoss* states. Alternatively, the expected housing capacity at the *OccLoss* state may need to account for the likelihood that the building is evacuated by its residents given the loss of a particular service, i.e.,

$$E\left[q(t) \mid OccLoss\right] = \left[1 - P\left(Evac \mid OccLoss\right)\right]\left[q(t) \mid OccFull\right] \tag{1-5}$$

where $E\left[q(t) \mid OccLoss\right]$ is the expected housing capacity for a residential building in the *OccLoss* functioning state; $P\left(Evac \mid OccLoss\right)$ is the probability that the building is evacuated, given that it is safe to occupy but without some of its services; and $\left[q(t) \mid OccFull\right]$is the housing capacity associated with the *OccFull* state or the pre-earthquake housing capacity. The $P\left(Evac \mid OccLoss\right)$ can be determined based on judgment informed by observations from past earthquakes. Knowing the level of functionality associated with each functioning state, the recovery paths for each limit state can be related to recovery functions, as illustrated in Figure 1-7 and calculated as follows:

$$\left[q(t) \mid LS_i\right] = \begin{cases} \left[q(t) \mid NOcc\right] & t < \left[T_{NOcc} \mid LS_i\right] \\ \left[q(t) \mid OccLoss\right] & \left[T_{NOcc} \mid LS_i\right] \leq t < \left[T_{NOcc} + T_{OccLoss} \mid LS_i\right] \\ \left[q(t) \mid OccFull\right] & \left[T_{NOcc} + T_{OccLoss} \mid LS_i\right] \leq t < T_{LC} \end{cases} \tag{1-6}$$

where $\left[q(t) \mid LS_i\right]$ is the time dependent building functionality given its immediate post-earthquake limit state *LSi*; $\left[q(t) \mid NOcc\right]$, $\left[q(t) \mid OccLoss\right]$ and $\left[q(t) \mid OccFull\right]$represent the level of functionality associated with the *NOcc*, *OccLoss* and *OccFull* states respectively; $\left[T_{NOcc} \mid LS_i\right]$is the time from the earthquake to the end of the *NOcc* phase associated with limit state *LSi*; $\left[T_{NOcc} + T_{OccLoss} \mid LS_i\right]$ is the time from the earthquake to the end of the *OccLoss* phase for limit state *LSi*.

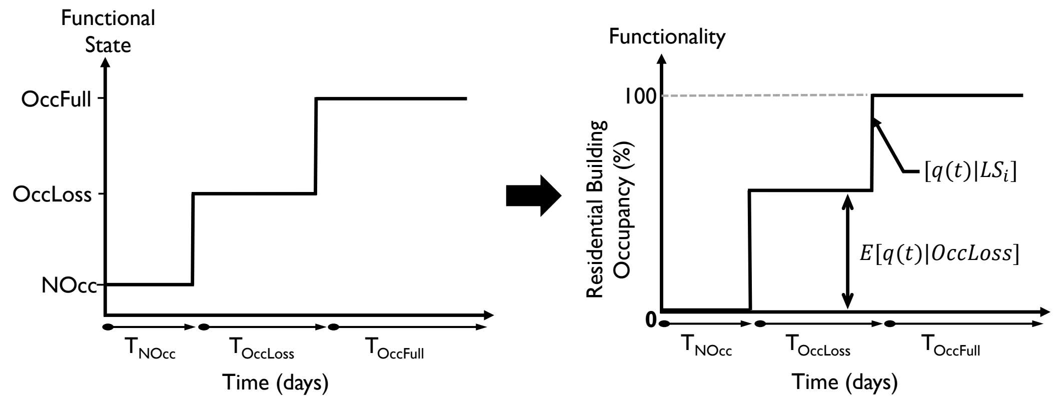


Figure 1-7. Conversion from recovery path to recovery function for residential building.
(Burton, Deierlein, Lallemant, & Lin, 2015).

The building recovery function is computed accounting for the likelihood of the building being in each of the five limit states for a given ground shaking intensity. This is illustrated in the event tree, shown in Figure 1-8, where each limit state is associated with a unique recovery function (computed from Equation 1-6). Figure 1-8 also incorporates a sixth event that corresponds to damage below the threshold level that triggers inspection. The uncertainty in the building limit state and expected recovery is determined by the following:

$$E[q(t) \mid IM] = \sum_{1}^{n_{ls}} [q(t) \mid LS_i] P[LS_i \mid IM] \qquad \text{1-7}$$

where $E[q(t) \mid IM]$ is the expected recovery function given *IM* and $P(LS_i \mid IM)$ is the probability that the building is in the *ith* limit state for a given *IM* level.

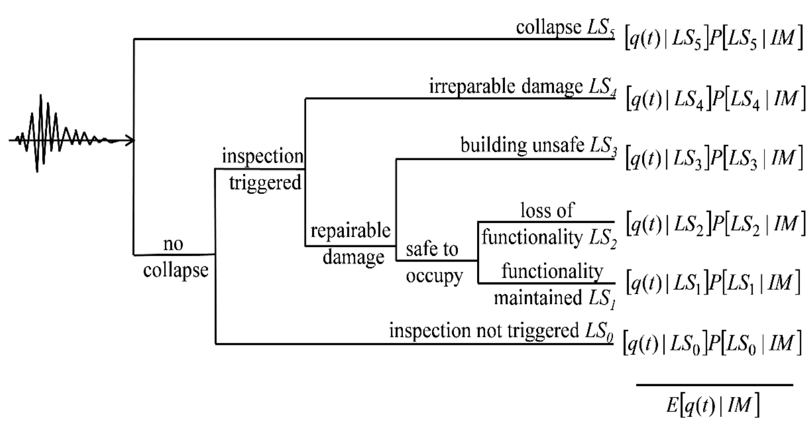


Figure 1-8. Limit state event tree used to assess building-level recovery.
(Burton, Deierlein, Lallemant, & Lin, 2015).

1.6. Statistical Models for Linking the Rate of Recovery to Socio-Economic Variables

Overview

Externalities and socio-economic vulnerability will have a significant effect on the rate of recovery of communities. This section presents several statistical models in addition to logistic regression (Section 4, main text) that could inform the relationship between the pace of recovery and some combination of predictor variables that are related to socio-economic factors. Three linear regression methods were utilized using the Napa recovery and resilience variables including (i) Ordinary Least Squares, (ii) Ridge, and (iii) The Least Absolute Shrinkage and Selector Operator. Among those methods, statistical significance is attained when a *p-value* is less than the significance level, which indicates that the statistic is reliable, i.e. that predictor has a strong influence on recovery. Two machine learning methods were also incorporated: (i) Random Forest, and (ii) Support Vector Machine.

Ordinary Least Squares Regression (OLS)

OLS is one of the most basic and commonly used prediction techniques with applications in fields as diverse as statistics, finance, economics and engineering. It uses a linear combination of predictors to estimate the dependent variable, which can be taken using the formula:

$$Y = X\beta + \varepsilon \qquad \text{1-8}$$

Y is a $n \times 1$ vector dependent variable, where n is the number of data points; X is a $n \times p$ matrix of the explanatory variables, where p is the number of predictor variables; β is a $p \times 1$ vector of the regression coefficient; and ε is a $n \times 1$ vector describing the random component of the linear relationship between X and Y.

The objective of *OLS* is to minimize the difference (residual) between the observed value of the dependent variable and predicted value by the linear approximation of the data.

$$\hat{\beta} = \ argmin \sum_i (y_i - x_i^T \beta)^2 = (X^T X)^{-1} X^T Y \qquad \text{1-9}$$

To evaluate the statistical significance of the regression coefficient, the t-statistic is used, where implicit statistical hypotheses are $H_0 : \beta = 0$ and $H_1 : \beta \neq 0$. The t-statistic, which follows the t distribution with $(n-p)$ degrees of freedom, is computed as the ratio between the estimated regression coefficient and its standard error:

$$T = \frac{\widehat{\beta}_J}{se(\widehat{\beta}_J)} \sim t(n-p) \tag{1-10}$$

where $se(\widehat{\beta}_J) = \sqrt{s^2(X^TX)^{-1}_{jj}}, s^2 = \frac{\hat{\varepsilon}^T\hat{\varepsilon}}{n-p}, \hat{\varepsilon} = \mathrm{Y} - \mathrm{X}\hat{\beta}$

The *p-value* is the probability of obtaining at least as extreme results given that null hypothesis is true, i.e. $p = Probability(t > |T|)$. If *p-value* is smaller than significance level α, the predictor attains statistical significance and therefore rejecting the null hypothesis. α is mostly often set at 0.05.

Ridge Regression

Ridge Regression is employed when analyzing multiple regression data that suffers from multicollinearity (Groetsch, 1984). Multicollinearity, which is the existence of strong correlations between predictor variables, can lead to inaccurate estimates of the regression coefficients, inflate the standard errors of the regression coefficients and deflate the partial t-tests for the regression coefficients.

Ridge regression penalizes the size of the regression coefficients by imposing a constraint on the sum of the squared values (the L_2 norm) of the predictor coefficients:

$$\hat{\beta} = \ argmin\left[\sum_i (y_i - x_i^T\beta)^2 + \lambda \sum_{j=1}^{p} \beta_j^2\right] = (X^TX + \lambda I)^{-1}X^TY \tag{1-11}$$

To select λ, cross validation is conducted whereby, the data is split into K folds. The regression model is developed using $\mathrm{K}-1$ folds and the test error is evaluated using the fold that was excluded from the model fitting. The best choice of λ would be the one that provides the least test error. T-statistic in ridge is computed in a similar manner to OLS (Equation 1-10):

$$where\ se(\widehat{\beta}_J) = \sqrt{s^2[(X^TX + \lambda I)^{-1}X^TX(X^TX + \lambda I)^{-1}]_{jj}}, s^2 = \frac{\hat{\varepsilon}^T\hat{\varepsilon}}{n-p}, \hat{\varepsilon} = \mathrm{Y} - \mathrm{X}\hat{\beta}$$

The Least Absolute Shrinkage and Selector Operator Regression (LASSO)

The *LASSO* regression is another method that addresses multicollinearity by doing both model selection and coefficient shrinkage (Tibshirani, 1996). Using L_1-penalization,

$$\hat{\beta} = argmin\left[\sum_i (y_i - x_i^T \beta)^2 + \lambda \sum_{j=1}^{p} \left|\beta_j\right|\right] = (X^T X + \lambda I)^{-1} X^T Y \qquad 1\text{-}12$$

LASSO is preferred over ridge regression when there is an assumption that the solution is sparse, i.e. many β_i= 0, because L_1 regularization shrinks some of the predictor coefficients to zero. If there is no assumption of the sparse feature, ridge regression is usually preferred since ridge only shrinks the regression coefficients.

Random Forest

The random forest is a machine learning algorithm, which combines the strategies of bagging, and randomly selecting features to classify and regress data (Breiman, 2001). Specifically, random forest constructs a multitude of decision trees using training data and outputs the mode of the classes (classification) or mean prediction (regression) of the individual trees. Bagging involves applying a majority vote (selecting the path with the greatest number of outcomes) for classification or prediction after many large trees are independently constructed using bootstrap resampled versions of the training data. Random forests change how the classification and regression trees are constructed. In normal trees, each node is split using the best split among all variables. In a random forest, each node is split using the best among a subset of predictors randomly chosen at that node.

The algorithm of random forest is as follows:

1. Draw bootstrap samples from training dataset.
2. For each of sample tree, the Classification and Regression Tree (CART) is applied first. The idea of CART is to recursively divide the space into rectangular subspaces until satisfying some criteria of classification or regression.
3. In each tree, randomly selected features need to be incorporated to modify CART. The best split (based on the Gini index (Menze, et al., 2009) among a subset of predictors is chosen randomly rather than using all variables.
4. Classify or regress data by major vote on all individual tress.

Support Vector Machine (SVM)

SVM is another widely used supervised learning method that is used to analyze data and perform classification and regression. In *SVM* regression, the input *n*-dimension features *x* are first mapped onto a high *m*-dimensional space using some fixed mapping (Smola & Vapnik, 1997). In this study, the Gaussian Radial Basis mapping is used, then linear regression is used to construct the model in this space. The linear model in the *m*-dimension feature space could be defined as:

$$f(x,\omega) = \sum_{i=j}^{m} \omega_j . g_j(x) + b \qquad \text{1-13}$$

where, $g_i(x), i = 1, \dots, m$ denotes the transformation using Gaussian Radial Basis.

SVM regression uses the ε-insensitive loss function proposed by (Smola & Vapnik, 1997):

$$L\big(y, f(x,w)\big) = \max\,(0, |y - f(x,w)| - \varepsilon) \qquad \text{1-14}$$

where ε is the tolerable bandwidth. At the same time, it also tries to reduce model complexity by minimizing $||w||^2$. Combining these two techniques, the following primal optimization problem is formed:

$$min\left\{\frac{1}{2}||w||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)\right\} \qquad \text{1-15}$$

$$\text{s.t.}\begin{cases} y_j - f\big(x_j, w\big) - \varepsilon \le \xi_i \\ f\big(x_j, w\big) - y_j - \varepsilon \le \xi_i^* \\ \xi_i, \xi_i^* \ge 0,\ \ i = 1, \dots n \end{cases}$$

The solution to the above dual problem is given by:

$$f(x) = \sum_{i=1}^{k}(\alpha_i - \alpha_i^*)K(x_i, x) \qquad \text{1-16}$$

$$\text{s.t.}\begin{cases} 0 \le \alpha_i \le C \\ 0 \le \alpha_i^* \le C \end{cases}$$

where, $K(x_i, x) = \sum_{j=1}^{m} g_j(x_i) g_j(x)$, and *k* is the number of support vectors.

1.7. Community-Scale Recovery Functions

An overview of the framework used to generate the community-level recovery functions is shown in Figure 1-9. The performance-based earthquake engineering framework is applied to each building within the target community, incorporating the limit states described earlier. The outcome is a recovery function that is generated for individual buildings. The function describing community-level recovery is obtained by aggregating the recovery curves for the individual buildings after accounting for the variation and spatial correlation of shaking intensity at each site, the effect of externalities and other socio-economic factors. Note that the contribution of individual buildings to the functionality of the region depends on the type of building and measure of functionality. This aggregation of building-level functionality would require quantifying the contribution of each building to the defined measure of community function. The housing recovery function is described by the following equation:

$$\left[Q(t) \mid EQ_j\right] = \sum_{i=1}^{n_{bldg}} E\left[q_i(t) \mid IM_i, EQ_j\right] \tag{1-16}$$

where $\left[Q(t) \mid EQ_j\right]$ describes community recovery for scenario earthquake *j*, $E\left[q_i(t) \mid IM_i, EQ_j\right]$ describes the expected recovery curve for building *i* at a given ground motion *IM* level resulting from scenario earthquake *j*, and n_{bldg} is the number of buildings in the community.

The long-term effects of an earthquake on a community can also be described by the cumulative loss of functionality over the course of the recovery period. For example, the loss of housing capacity over the recovery period measured in 'person-days" can be computed from a community-level recovery curve that has number of residents housed by the community as the measure of functionality. This cumulative loss in functionality for a particular earthquake event is illustrated in Figure 1-9 and can be described by the following equation:

$$\left[LQ_{RE} \mid EQ_j\right] = \int_{T_E}^{T_{RE}} (Q_0 - Q(t))dt \tag{1-17}$$

where $\left\lfloor LQ_{RE} \mid EQ_j \right\rfloor$ is the loss of functionality over the recovery period for scenario earthquake *j*, Q_0 is the pre-earthquake level of functionality, T_E is the time of the earthquake and T_{RE} is the time at full recovery.

Equation 1-17 describes the cumulative loss of functionality for a single scenario earthquake. Multiple scenario earthquakes can be considered and used to describe the annual exceedance rate for specified amounts of cumulative loss. This is obtained by computing the cumulative loss for multiple earthquake scenarios each with a different magnitude, location and annual rate of occurrence. The rate of exceedance, λ, for specified loss levels is estimated by summing the occurrence rate for all scenarios in which the loss threshold on interest is exceeded.

$$\lambda_{LQ_{RE} \geq lq} = \sum_{j=1}^{J} w_j I(LQ_{RE} \geq lq) \tag{1-18}$$

where w_j is the occurrence rate for scenario j and lq is the cumulative loss threshold. The indicator function $I(LQ_{RE} \geq lq)$ is set equal to 1 if the argument $LQ_{RE} \geq lq$ is true and 0 otherwise.

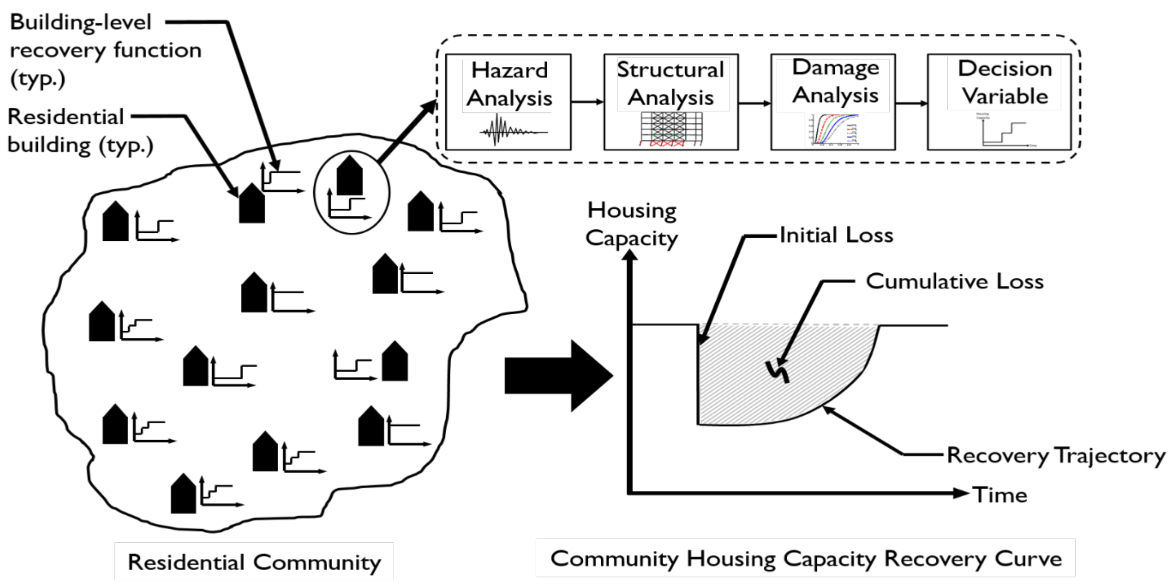


Figure 1-9. Conceptual illustration of recovery modelling framework.
(Burton, Deierlein, Lallemant, & Lin, 2015).

APPENDIX 2 – Integrated Risk Modelling Toolkit (IRMT)

In the following steps, a brief description of the basic workflow to develop building and community level recovery curves is presented, where the main features and capabilities of the tool are highlighted. For the sake of demonstration, the recovery of a random sample of the residential buildings of the city of Napa, California is utilized as a case study.

STEP 1: Preparation of the input files to launch an OpenQuake-engine analysis

In order to run the Reconstruction Recovery Model users are required to provide a CSV file containing the probability of exceedance of each limit state for each individual building in the exposure model. The latter can be computed by running a Scenario Damage Assessment, which is a type of analysis supported by the risk component of the OpenQuake-engine. The input files necessary for running a scenario damage calculation and the resulting output files are depicted in Figure 2-1. For technical details, definitions and examples of each component, readers are referred to Silva, Crowley, Pagani, Monelli, & Pinho (2014).

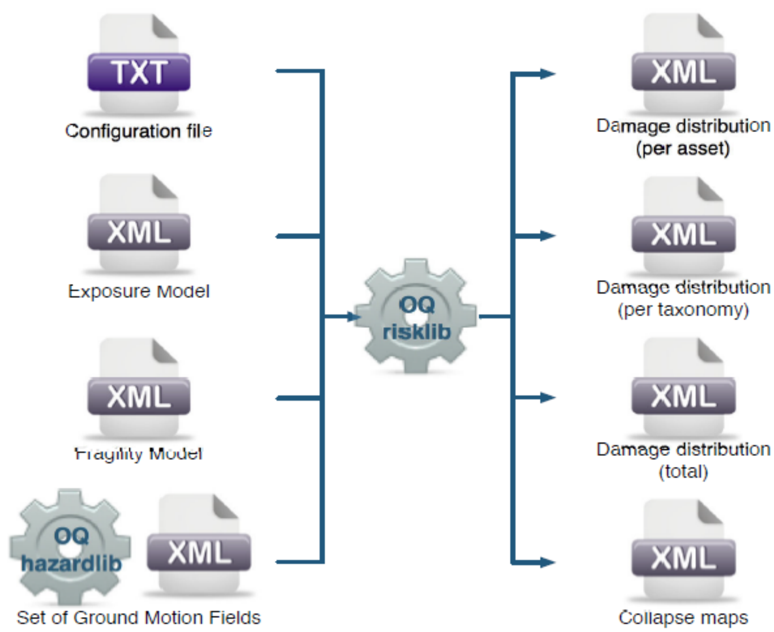


Figure 2-1. Scenario Damage Calculator input/output structure.

The window that requests users to upload the input files and run the scenario damage calculation is shown in Figure 2-2.

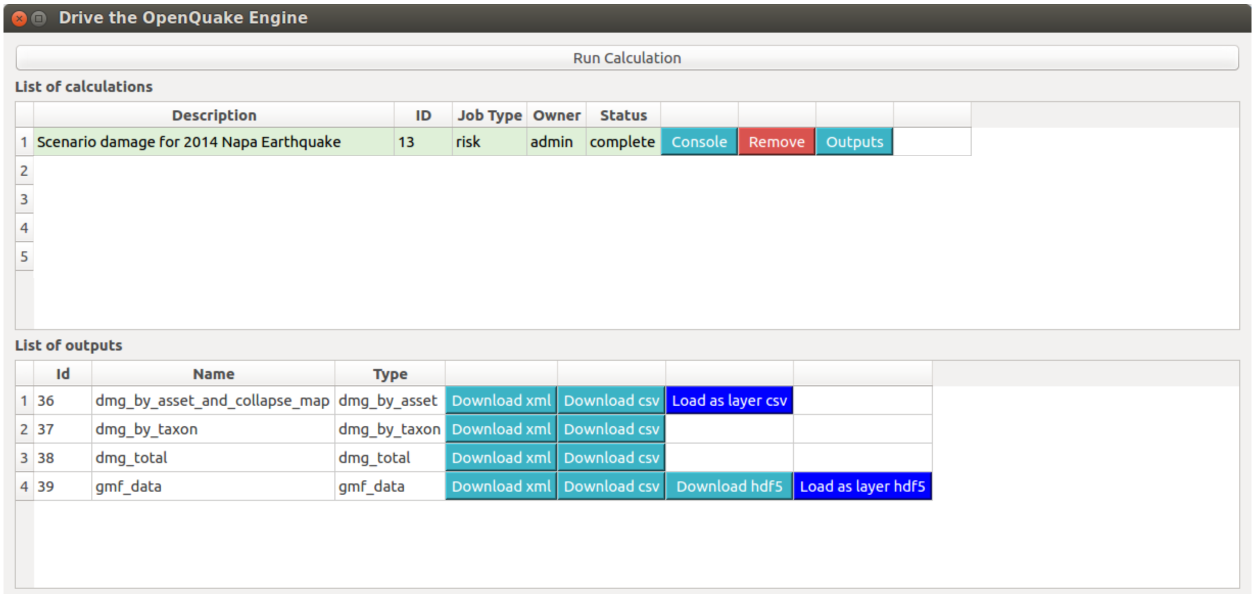


Figure 2-2. Pop-up window to run the OpenQuak*e*-engine server.

It should be noted that to use the OpenQuake-engine from QGIS, the user needs to set up the connection with a working OpenQuake-engine server using the 'IRMT settings" dialog (Figure 2-3); the server can be installed in the same machine where the plugin is used (Figure 2-3, Host). Alternatively, it is possible to use a remote server or cluster.

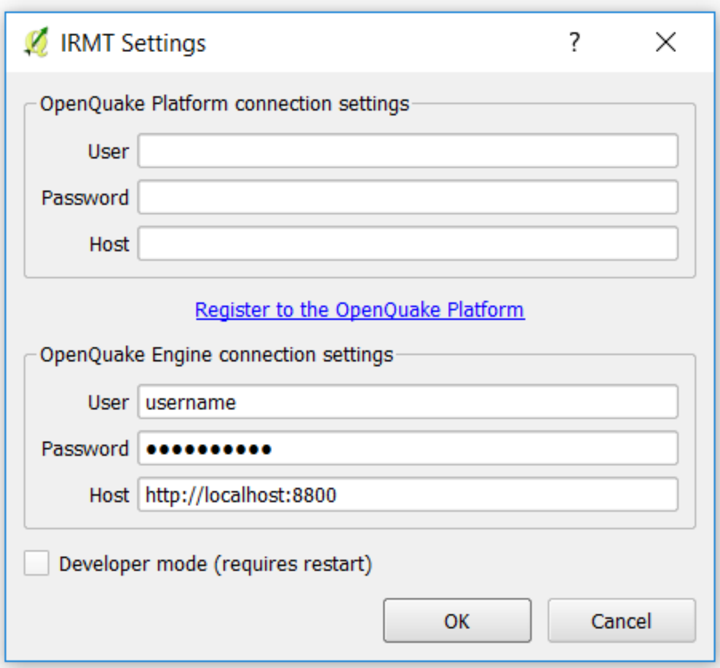


Figure 2-3. Pop-up window to set up the connection with a working OpenQuake-engine server.

STEP 2: Setting up the configuration variables to run the Reconstruction Recovery Model

The configuration variables necessary to perform the recovery modelling analysis are listed in Table 2-1. They should be adjusted to the available data and needs of the user.

Table 2-1. Required configuration variables in order to run the Reconstruction Recovery Model.

Input Variables	Short Explanation
Damage by asset	File that contains the mean probabilities of exceedance of each limit state for each individual building; output of the OpenQuake-engine[10]
Inspection Time	Time to complete inspections
Assessment Time	Time to conduct engineering assessments
Mobilization Time	Time to mobilize for construction
Lead Time Dispersion	Level of uncertainty associated with the lead time[11]
Repair Time	Time to repair a building
Repair Time Dispersion	Level of uncertainty associated with the repair times
Recovery Time	Period between the occurrence of an earthquake and the restoration of full functionality of the building
Transfer Probabilities	Conditional probability of being in a particular recovery-based (limit) state, given the occurrence of a loss-based damage state (e.g., slight, moderate, extensive, complete)

The list of the outputs from the scenario damage calculation can be visualized in Figure 2-2. The tool offers the possibility to load the 'Damage by asset" CSV file (dmg_by_asset as shown in Figure 2-2) as a QGIS vector layer, stored in the user's computer as a shapefile. In addition, it is possible to automatically style the layer with respect to a chosen damage state. Alternatively, the user can upload on QGIS the 'Damage by asset" CSV file, structured in the same format as produced by the OpenQuake-engine. If the user does not need to edit the layer by adding or removing fields to/from it, it is possible to perform the recovery modelling calculation using the CSV-based layer. Otherwise, the layer should be converted and saved as a shapefile. Note that shapefile limitations will reduce the field names to a maximum length of 10 characters each.

At this point, the user may choose between two workflows on how to proceed to the generation of single buildings and/or community level recovery curves.

[10] Seismic hazard and risk calculation software, developed by the Global Earthquake Model (GEM) Foundation.

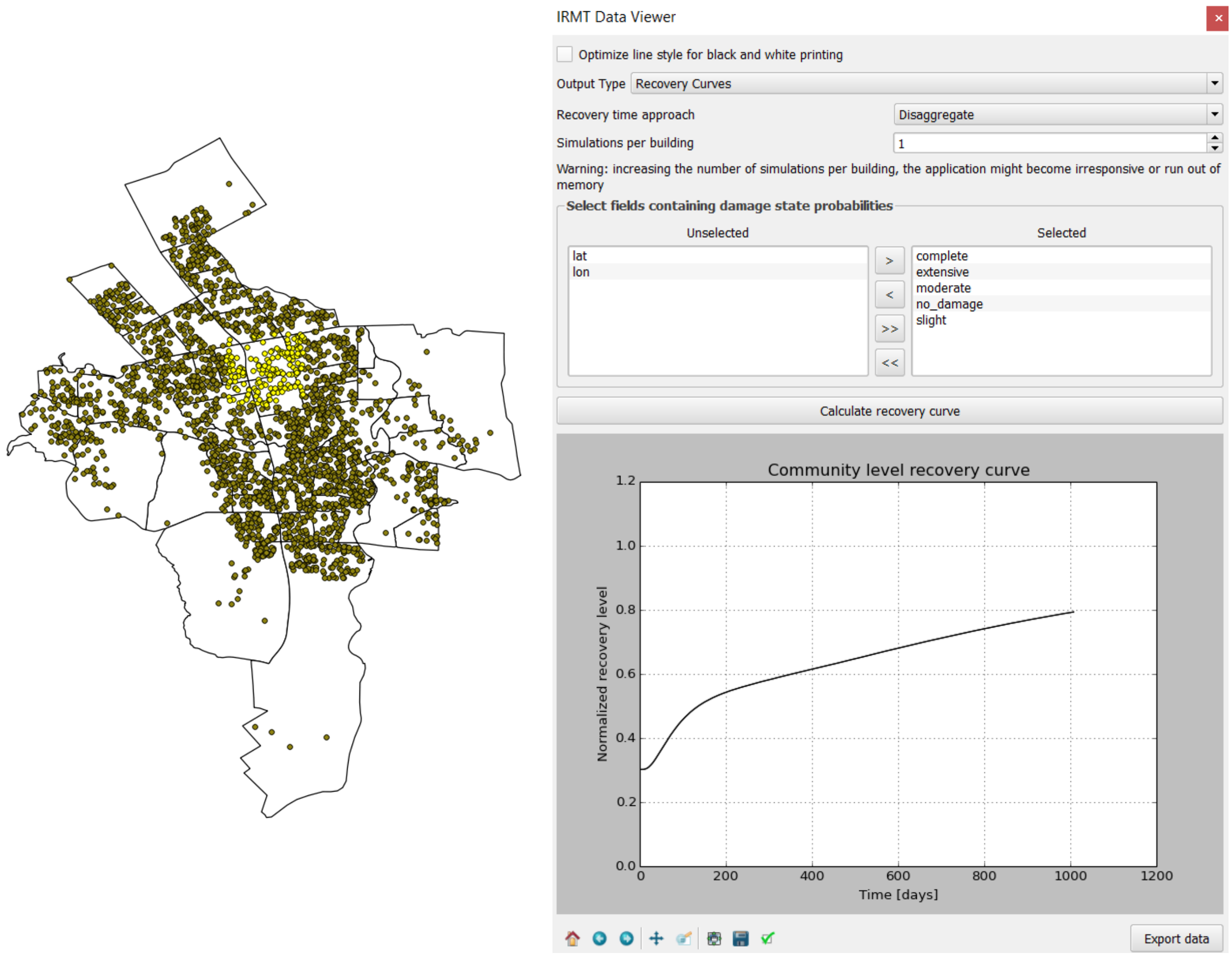
[11] Lead time is the time required for building inspection and/or evaluation, finance planning, architectural/engineering consultations, a competitive bidding process, and mobilizing for construction (Mitrani-Reiser 2007).

Workflow 1

The user can select individual buildings (or a group of buildings) and the respective recovery curve (single or aggregated) is automatically developed. The curve can be edited, digitized and exported as a CSV, as well as saved as an image. The user is required to select one of two available algorithmic approaches regarding the estimation of the recovery time (see Table 2-2) and, more importantly, to request the development of recovery curves by setting the Output Type tab to 'Recovery Curves" (Figure 2-4b). In addition, the user is able to manually select the fields of the file that contain the probabilities of being in each damage state (Figure 2-4b:"Select fields containing damage state probabilities"). If the file with the damage state probabilities is in the same format as produced by OpenQuake, the software pre-selects the appropriate fields for the recovery modelling algorithm. Figure 2-4 illustrates an example of the aggregated recovery curve of a set of selected buildings (highlighted with yellow).

Table 2-2. Aggregated and disaggregated approaches for the estimation of the recovery time.

Approach for estimation of recovery time	**Short Explanation**
Aggregated approach	Building-level recovery model as a single process.
Disaggregated approach	Building-level recovery modelled using four processes: inspection, assessment, mobilization and repair.



a) b)

Figure 2-4. Aggregated recovery curve of a set of selected buildings in the city of Napa.

It should be emphasized that the integration of the recovery modelling algorithm in the *QGIS* software enables the users to adapt the workflow to their needs, leveraging all the features provided by the *QGIS* framework. The *QGIS* Processing Toolbox gives access to a wide variety of geoalgorithms, seamlessly integrating several different open-source resources, such as R, SAGA or GDAL. For instance, a SAGA algorithm, the 'Add Polygon Attributes to Points", can be used to aggregate by zone a set of selected assets, resulting in relating each asset to the identifier of the geographical area (zone) where it belongs. Following, the selection of the set of assets to be considered in the analysis can be performed in several different ways. The user can directly select points by clicking them on the map, or select points by using a formula. If points have been labeled with the identifier of the zone, the selection can be done with respect to the zone identification (or ID).

Workflow 2

Initially, the user must select the layer containing the information regarding the damage state probabilities per asset (see **STEP 1**), after which the specific fields that contain these probabilities shall be chosen (Figure 2-5). Next, the user must select a specific recovery time approach (Aggregate/Disaggregate) and set the number of the simulations per building[12] (Figure 2-5). Here, it is possible to upload the layer of the study area with zonal geometries and generate aggregated recovery curves by zones. To exemplify, the block groups (zones) of the city of Napa and the aggregated recovery curve for the block group with the ID of 8032 are depicted in Figure 2-6.

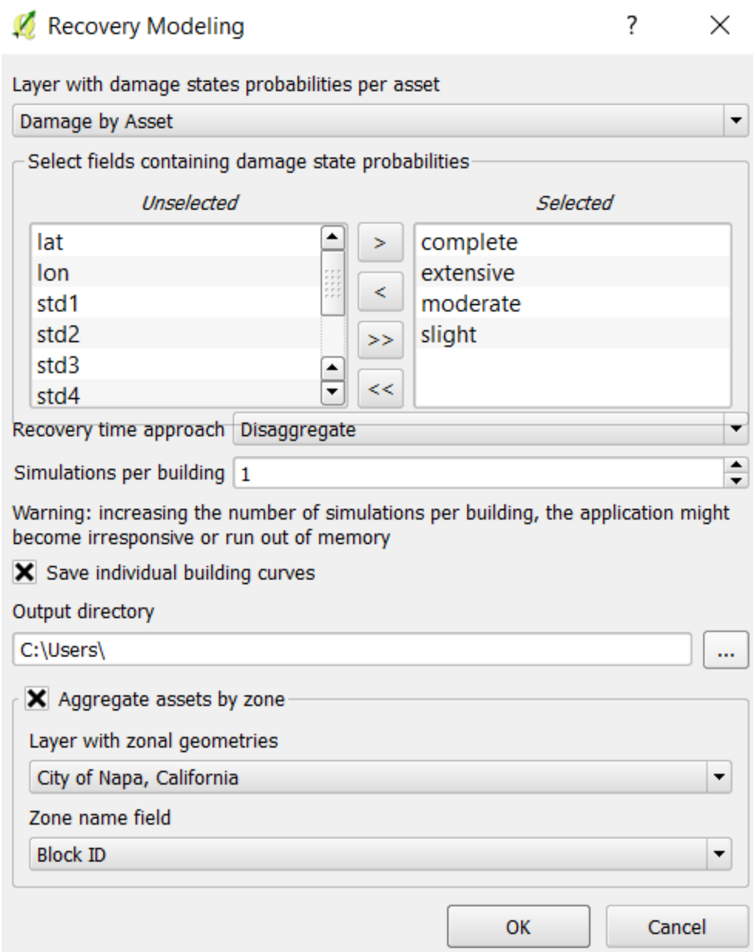


Figure 2-5. Pop-up window to run the Reconstruction Recovery Model.

[12] Number of damage realizations used in Monte Carlo Simulation.

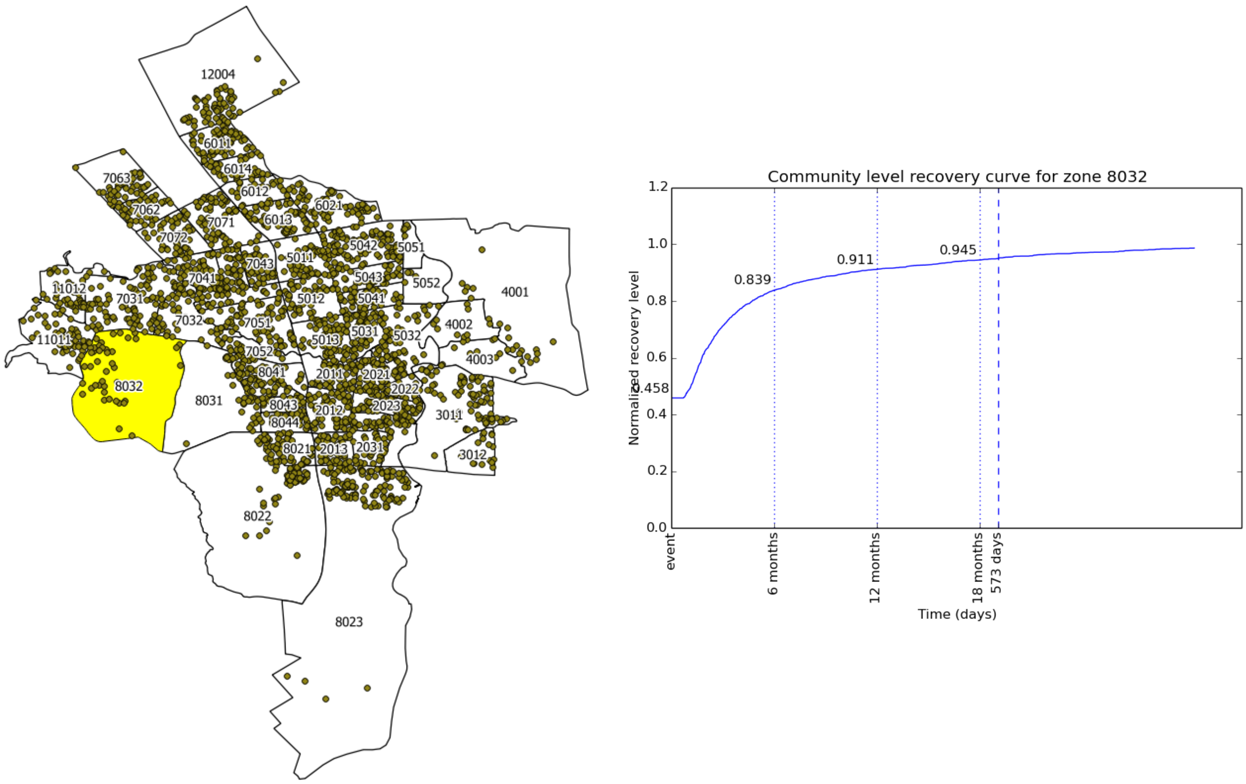


Figure 2-6. Community level recovery curve for the zone (block group) with an ID of 8032, of the city of Napa.

By unchecking the 'Aggregate assets by zone" box (Figure 2-5) the algorithm generates a single community recovery curve by aggregating the recovery curves of all the buildings within the region. The graphs, like the one shown in Figure 2-6, are saved in the output directory designated by the user. In addition, building-by-building recovery curves are digitized and can be saved as text files (.txt) in the same output directory. The user can decide whether or not to generate the building-by-building recovery curves by (un)checking the 'Save individual building curves" box. The data can be further used (e.g., with Microsoft Excel) to generate and visualize individual building recovery curves that may be of interest to the user.

APPENDIX 3 – Recovery from the Southern California ShakeOut Scenario

A case study was conducted to model the recovery of the residential building stock from the 2008 Southern California ShakeOut Scenario. The ShakeOut is an exercise based on a potential magnitude 7.8 earthquake on the southern San Andreas Fault with the goal to identify the physical, social and economic consequences of a major earthquake in southern California and to enable the users of its results to identify what they can change now to avoid catastrophic impact after the inevitable earthquake occurs (Jones, et al., 2008). It involves eight counties, including: Riverside, Orange, Kern, San Diego, San Bernardino, Ventura, Los Angeles and Imperial.

The Integrated Risk Modelling Toolkit (IRMT) was used to develop an aggregated recovery curve for each of the above counties. Initially, the ShakeOut Earthquake was simulated and damage distribution statistics were computed for all the residential buildings across the eight counties, by running a Scenario Damage Assessment. The latter is a type of analysis supported by the risk component of the OpenQuake-engine (Silva et al., 2014). The building exposure data that were used in the Scenario Damage Assessment were obtained from the 'Beyond button Pushing – Seismic Risk Assessment for California" project of GEM supported by the California Seismic Safety Commission.

The recovery curves for each county are depicted in Figure 3-1. The highlighted blue dashed line indicates the estimated number of days needed for 95% of the housing capacity of each county to be restored.

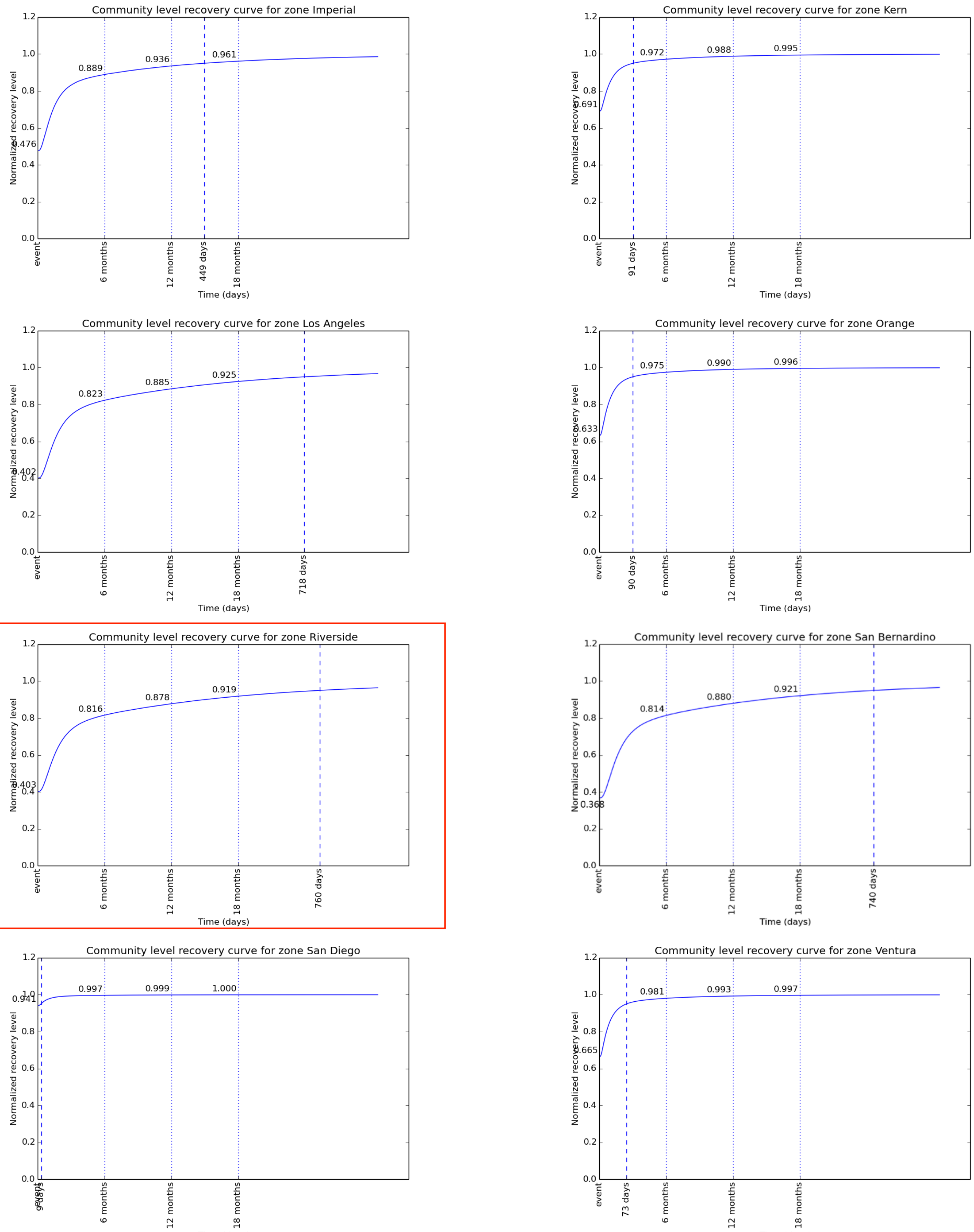


Figure 3-1. Recovery curves for the counties of the Southern California ShakeOut Scenario.

To exemplify, the County of Riverside (in red square in Figure 3-1), will need approximately 760 days to restore 95% of its housing capacity, with a drop to almost 40% following the event. Almost 74% of the total residential buildings of this county consists of wood, light frame with moderate code[13], 11% of wood, light frame with high code[14] and nearly 9% of mobile homes with moderate code ('Beyond button Pushing – Seismic Risk Assessment for California"). As calculated by the OpenQuake-engine, nearly 45% of the buildings that will sustain complete damage consist of wood, light frame buildings constructed according to moderate seismic level of design and nearly 40% of mobile homes. Mobile homes, which represent only 9% from the total residential buildings in the county, constitute the second more impacted building typology.

These kinds of estimates could provide vital information to emergency managers and recovery planners, such as defining the temporary shelter needs and/or prioritizing the allocation of available financial resources and funds, making sure that they reach the areas in higher need.

[13] Buildings of older construction are best represented as Moderate-Code, if built after about 1940 (FEMA, 2016).
[14] Classification system according to the level of seismic design. For example, buildings of newer construction (e.g., post-1973) are best represented as High-Code (FEMA, 2016).

APPENDIX 4 – Socio-Economic Recovery Model

Table 4-1. Variables that were used to represent the socio-economic conditions in the city of Napa.

Variables
Social Sector of the Community
Percentage of households where they speak English
Percentage of housing units with no persons with a disability
Percentage of civilian non-institutionalized population with any type of health insurance
Percentage of occupied housing units with telephone service
Percentage of occupied housing units with vehicle available
Percentage of population 25 years and over that have at least a regular high school diploma
Percentage of total population that is male
Percentage of total population that is above 5 and below 60 years
Percentage of total population that is not a minority (White alone, not Hispanic or Latino)
Number of child care services
Percentage of total households with less than 5 persons
Percentage of single-parent households with a male householder, no wife present
Economic Sector
Percentage of households with earnings
Percentage of population 16 years and over in labour force that is employed
Percentage of population that has income at or above poverty level
Per capita income as a fraction of the highest amongst the block groups
Percentage of the renter-occupied housing units with gross rent less than $1500[15)] (+)
Percentage of civilian employed population 16 years and over that are not employed in food services, accommodation and retail trade[16)]
Percentage of females 20 to 64 years in households that are in labour force
Percentage of occupied housing units that are owner occupied
Percentage of civilian employed population 16 years and over that are employed in healthcare practitioners and technical occupations
Percentage of households with no supplemental security income

[15)] In the U.S., it is commonly accepted that families who pay more than 30% of their income for housing are considered cost burdened. The value of $1500 as a limit of affordability was set according to this rule.
[16)] This variable is used as a proxy for single sector employment dependence.

Variables
Percentage of households with no public assistance income
Infrastructure Sector
Housing density
Percentage of housing units that are built after 1950
Percentage of housing units that are not mobile homes
Number of internet, television, radio and telecommunications broadcasters
Number of schools (primary and secondary)
Number of hotels & motels
Number of banks
Percentage of housing units that are vacant
Number of police, fire, emergency relief services and temporary shelters
Percentage of housing units that are single-family detached homes
Community Capital
Number of civic and social advocacy organizations
Number of churches and religious organizations
Number of arts, entertainment and recreation centres, libraries, museums, parks and historic
Percentage of population that lived in the same Metropolitan Statistical Area 1 year ago
Institutional Sector
Percentage of civilian employed population 16 years and over employed in emergency services (firefighting, law enforcement, protection)

4.1. Community Recovery Predictions over time

In this section, the development of the Socio-Economic Recovery Model is described in more detail, providing information on the statistical model that was used and the methods to validate the accuracy of the results. The overall approach is based on (Despotaki, Sousa, & Burton, 2017), which is a scientific paper currently under review in the Earthquake Spectra Journal, resulted from the work conducted within this project.

As mentioned in Section 3.3.1 in the main text, the recovery progress of only a sample of 356 damaged buildings was monitored over a period of 18 months following the 2014 M6 South Napa Earthquake (out of a total of 1462 observed by the city officials damaged structures at the moment of this study). The spatial distribution of the surveyed buildings is presented in Figure 4-1.

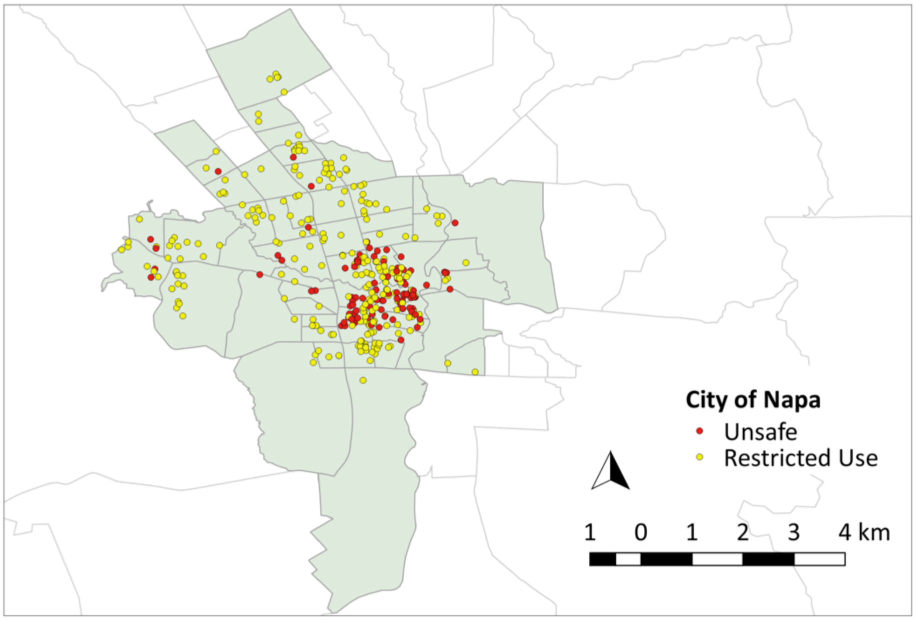


Figure 4-1. The 356 evaluated red and yellow-tagged (ATC, 2005) buildings in the city of Napa.

Thus, the number of assessed structures in several block groups (subdivisions, which can be seen in Figure 4-1) was not considered sufficient to ensure a meaningful prediction of the fraction of recovered buildings inside such block groups. More specifically, because damage data was collected for a limited number of buildings, we were interested in determining the probability of a given structure being in recovery state 1 (and conversely in recovery state 0) as a function of a set of socio-economic indicators. Various methods are commonly used to describe the relationship between outcome parameters and a set of independent[17] parameters. In multivariate linear regression analysis, it is possible to test whether predicted and independent variables are linearly related as established by an equation of the form $Y = \alpha + \beta X$ (Hosmer & Lemeshow, 2000). In this relationship, Y is the variable being predicted; X is a vector of independent variables; α represents the value of Y when $X = 0$ (i.e. the so-called intercept); and β is a vector of estimated regression parameters with length equal to the length of X. Estimates of the intercept α and the regression coefficients β are used in the form of the above regression equation, resulting in a set of predictions - $\hat{Y}$.

[17] The independent parameters represent inputs or causes for variation.

The multivariate linear regression model can easily be extended to accommodate dichotomous independent variables (Fox, 2000); i.e. variables that occur in one of two possible states. However, when the dichotomous variable is the dependent[18] one (such as the present case - No Recovery:0, Full Recovery:1), the interpretation of the regression equation is not as straightforward. In this situation, the predicted $\hat{Y}$ value corresponds to the probability of the outcome falling into one of the two possible categories; that is: $P(Y = 1) = 1 - P(Y = 0)$. However, probabilities of Y being equal to 1 predicted by a linear regression model can, in theory, be infinitely large (therefore>1) or small (thus, <0); which is not in accordance with reality.

A more realistic approach would be that according to which the probability of $Y = 1$ tends to a value of one when X approaches very high values, and to zero for very low values of X, but never exceeds one or is lower than zero. In this case, it is possible to transform the independent variable so that the substantive relationship is nonlinear, but its form remains linear. A particularly relevant example of such transformation is the multivariate logistic model, which resorts to the logit correction (Fox, 2000) to produce independent variables that, varying between negative and positive infinity, ensure that the predicted values of $P(Y = 1)$ lie in the interval $[0, 1]$. If $x_1, x_2, \dots, x_n$ are the independent variables and Y is the dependent one, the logit is the natural logarithm (ln) of the odds of Y, and the odds are defined as the probability (P) of Y occurring, divided by the probability of Y not occurring:

$$Logit(Y) = ln\ \frac{P}{1 - P} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \qquad 4\text{-}1$$

where β_1 to β_n are the regression coefficients correspoing to each of the x_1 to x_n independent variables. After exponentiating both sides, the probability of occurrence of the outcome of interest is derived as follows:

$$P(Y = 1) = \frac{e^{(\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)}}{1 + e^{(\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)}} \qquad 4\text{-}2$$

where α is the intercept of the regression.

For the regression to be possible, each block group was assigned a recovery stage (0 or 1), to make the data analogous to the socio-economic variables that were collected at the

[18] The dependent parameters represent the output or outcome whose variation is being studied.

block group level of geography. To accomplish this, a simulation procedure was devised and for each of a total of 1000 simulations a single building was selected at random from the structures assessed in each block group. Next the selected building's recovery stage was attributed to the respective block group. Accordingly, for each simulation (*S*), the corresponding $P(Y = 1)s$ was determined through Equation 4-2, based on the maximum-likelihood estimation (Hosmer & Lemeshow, 2000) of the corresponding logistic regression parameters a_s and β_s. Thus, both the mean predicted values of *P(Y=1)* and associated uncertainty could be calculated for each block group. An example of these results is presented in Figure 4-2, where the mean, as well as the lower (16% percentile) and upper (84% percentile) bound predicted recovery probabilities were mapped for each block group, 6 months after the earthquake (similar graphs can be produced at any time after the event, even after the time of the last field survey, 18 months after the earthquake). Figure 4-2 highlights the fact that it is possible to communicate the uncertain nature of the recovery process in a simple yet meaningful way, allowing decision and policy-makers to plan for various scenarios.

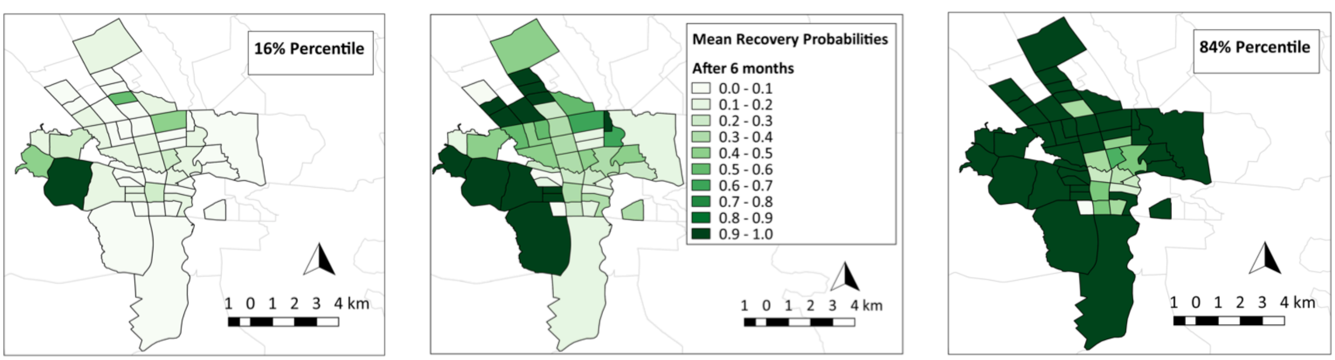


Figure 4-2. Mean, 16% and 84% percentile recovery probabilities in the assessed area, 6 months following the event, as determined by the Socio-Economic Recovery Model (Despotaki, Sousa, & Burton, 2017).

4.2. Goodness-of-Fit and Accuracy of the Model

The methods used to assess the goodness-of-fit of a logistic regression model follow similar principles as those applicable in the evaluation of a linear model. Therefore, to better understand the methodology subsequently described, each step is presented in the context of its equivalent in a general linear regression. In linear regression analysis, the evaluation of the overall model is based on three sums of squares: a) the total sum of squares (SST), which is the sum of squared deviations of all the observed Y_i with respect to the mean predictions ($\bar{Y}$) (i.e. $\sum(Y_i - \bar{Y})^2$); $\sum(Y_i - \hat{Y})^2$; b) the error sum of squares (SSE); and c) the regression sum of squares (SSR) is simply SSR=SST-SSE (Hosmer & Lemeshow, 2000).

Just as the sum of squared errors (SSE) is the quantity used to assess the quality of a linear regression, the log likelihood is the criterion in the case of the logistic model. For convenience, the log likelihood (LL) is usually multiplied by -2 (-2LL), resulting in a value that, for the intercept only model (designated D_0), is equivalent to the total sum of squares in linear regression (SST). Similarly, the value of -2LL for the logistic regression model that includes the independent variables as well as the intercept (full model) that is designated D_M is analogous to the error sum of squares (SSE). In logistic regression, the most direct comparable parameter to the regression sum of squares (SSR) is the difference between D_0 and D_M, which is interpreted as the model chi-square (χ^2) statistic (Hosmer & Lemeshow, 2000).

All the log likelihood terms presented above can be used in the evaluation of different statistical hypotheses in the context of logistic regression. However, D_M is of particular relevance here, as it has been historically used as a measure of the model 'goodness of fit". While (D_0 - D_M) can be used to determine whether a full model provides better predictions than an intercept-only model, D_M compares the full model with a saturated one. The further assumption that D_M follows a χ^2 distribution allows the computation of its statistical significance (or *p-value*), as $p = 1 - \chi^2{}_{CDF}(D_M, df)$. In the latter, df is the number of degrees of freedom of the χ^2 distribution, and CDF stands for its cumulative distribution function. For further details, readers are referred to Hosmer & Lemeshow (2000) regarding matters concerning the definition of a saturated model, the number of degrees of freedom of χ^2, and the limitations inherent to the assumption of D_M following a χ^2 distribution.

In the present case, a significance level (*alpha*) of 0.05 was assumed, proving the limit for the *p-value* below which we reject the null hypothesis that the full model allows us to make predictions with similar quality as the saturated model. It was found that from each of the 1000 regressions the predicted probabilistic distributions of $P(Y = 1)$ were statistically significant for virtually all (95%) simulations, based on the considered *alpha* threshold.

In addition, as a measure of the quality of the results, the mean predicted values of recovery were plotted as a function of the 'surveyed" probabilities in each block group for each of the field surveys. In this context, 'surveyed" probabilities are determined as the fraction of assessed buildings that are fully recovered at the time of interest, in each block group. As illustrated in Figure 4-3, the relationship between mean predicted and surveyed values approaches an almost perfect linear trend for the three survey instances.

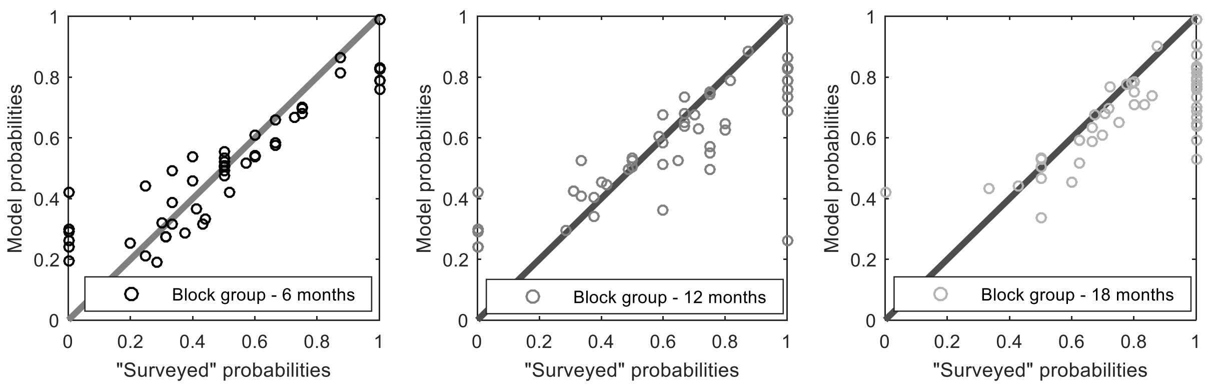


Figure 4-3. Predicted versus 'surveyed" probabilities in each block group (Despotaki, Sousa, & Burton, 2017).

On the other hand, when surveyed probabilities are either zero or one (mainly in block groups where very few buildings were surveyed), the model probabilities differed significantly from the observed ones. This is considered to be a strength of the Socio-Economic Recovery Model, because the observed probabilities in these cases (zero or one) are not an accurate representation of reality, but rather a consequence of the limited number of buildings assessed in these particular block groups. As a result, the correction provided by the model for these probabilities reflects the spatial trend of recovery in the block groups with higher (and more significant) number of assessed structures.

4.3. Contribution of the predictors in the regression analysis

Wald statistics were used to compute the significance (*p-value*) of the regression coefficients obtained for each independent socio-economic variable. The Wald statistic can be calculated as the ratio between the coefficient of a given independent variable and its standard error, in which case it follows a standard normal distribution and its formula parallels the formula for the t ratio for coefficients in linear regression (Hosmer & Lemeshow, 2000). *P-values* lower than 0.10, as suggested by Hosmer and Lemeshow

(2000), are assumed to correspond to the variables that most significantly affect the recovery results. Herein, seven of the 38 (Table 4-1) socio-economic parameters, the variable representing the time, *T*, and the damage indicator (MMI) significantly contribute to the prediction of recovery trajectory, as presented in Section 3.3 in the main text.

9 References of Appendices

ATC. (2005). *Field Manual: Post Earthquake Safety Evaluation of Buildings ATC-20-1, Second Edition.* Redwood City, California: Applied Technology Council.

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Burton, H., Deierlein, G., Lallemant, D., and Lin, T. (2015). 'Framework for Incorporating Probabilistic Building Performance in the Assessment of Community Seismic Resilience," *Journal of Structural Engineering*, 10.1061/(ASCE)ST.1943-541X.0001321 , C4015007.

Despotaki, V., Sousa, L., & Burton, C. (2017). Using resilience indicators in the Prediction of Earthquake Recovery. *Under Review in Earthquake Spectra.*

FEMA. (2016). *"HAZUS 3.2 earthquake loss estimation methodology." Technical Manuals I, II, IIII.* Washington, D.C.: National Institute of Building Sciences for FEMA.

Fox, J. (2000). Multiple and Generalized Nonparametric Regression. Thousand Oaks, CA: SAGE Publications, Inc.

Groetsch, C. W. (1984). *The theory of Tikhonov regularization for Fredholm equations of the first kind* (Vol. 105). Pitman Advanced Publishing Program.

Hosmer, D., & Lemeshow, S. (2000). *Applied Logistic Regression*, Second Edition. United States of America: John Wiley & Sons, Inc.

Jones, L., Bernknopf, R., Cox, D., Goltz, J., Hudnut, K., Mileti, D., . . . Wein, A. (2008). *The ShakeOut Scenario*. U.S. Geological Survey OpenFile Report 2008-1150 and California Geological Survey Preliminary Report 25.

Menze, B. H., Kelm, B. M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., & Hamprecht, F. A. (2009). A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC bioinformatics*, 10(1), 10.1186/1471-2105-10-213.

Mitrani-Reiser, J. (2007). '*An ounce of prevention: Probabilistic loss estimation for performance-based earthquake engineering,"* Ph.D. dissertation, Department of Applied Mechanics, California Institute of Technology.

Poland, C. (2009, February 1). *The resilient city: Defining what San Francisco needs from its seismic mitigation policies*. Retrieved from http://www.spur.org/featured-project/resilient-city.

Silva, V., Crowley, H., Pagani, M., Monelli, D., & Pinho, R. (2014). Development of the OpenQuake engine, the Global Earthquake Model's open-source software for seismic risk assessment. *Natural Hazards*, 72(3), 1409-1427.

Singhal, A., and Kiremidjian, A.S. (1996). 'Method for probabilistic evaluation of seismic structural damage,' *Journal of Structural Engineering*, 122(12), 1459-1467.

Smola, A., & Vapnik, V. (1997). Support vector regression machines. Advances in neural information processing systems, 9, 155-161.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*. Series B (Methodological), 267-288.