Reduction of Uncertainty in Post-Event Seismic Loss Estimates Using Observation Data and Bayesian Updating

Maura Torres

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

Columbia University
2017

Abstract

Reduction of Uncertainty in Post-Event Loss Estimates Through Observation Data

Maura Torres

The insurance industry relies on both commercial and in-house software packages to quantify financial risk to natural hazards. For earthquakes, the initial loss estimates from the industry's catastrophe risk (CAT) models are based on the probabilistic damage a building would sustain due to a catalog of simulated earthquake events. Based on the occurrence rates of the simulated earthquake events, an exceedance probability (EP) curve is calculated, which provides the probability of exceeding a specific loss threshold. Initially these loss exceedence probabilities help a company decide what insurance policies are most cost efficient. In addition they can also provide insights into loss predictions in the event that an actual natural disaster takes place, thus they are prepared to pay out their insured parties the necessary amount. However, there is always an associated uncertainty with the loss calculations produced by these models. The goal of this research is to reduce this uncertainty by using Bayesian inference with real time earthquake data to calculate an updated loss. Bayes theory is an iterative process that modifies the loss distribution with every piece of incoming information. The posterior updates are calculated by multiplying a baseline prior distribution with a likelihood function and normalization factor. The first prior is the initial loss distribution from the simulated events database before any information about a real earthquake is available. The crucial step in the update procedure is defining a likelihood function that establishes a relative weight for each simulated earthquake, relating how alike or dislike the attributes of a simulated earthquake are to those of a real earthquake event. To define this likelihood function, the general proposed approach is to quantify real time earthquake attributes such as magnitude, location and damage, and compare them to an equivalent value for each simulated earthquake from the CAT model database. In order to obtain the simulated model parameters, the catastrophe risk model is analyzed for different

building construction types, such as steel and reinforced concrete. For every model case, the loss, peak ground acceleration per building and simulated event magnitude and locations are recorded. Next, in order to calculate the real earthquake attributes, data was collected for two case studies, the 7.1 magnitude 1997 Punitaqui and the 8.8 magnitude 2010 Chile earthquake. For each of these real earthquake events, the magnitude, location, peak ground acceleration at every available accelerometer location, and qualitative damage descriptions were recorded. Once the data was collected for both the real and simulated events, they were quantified so they could be compared on equal scales. Using the quantified parameter values, a likelihood function was defined for each update step. In general, as the number of updates increased, the loss estimates tended to converge to a steady value for both the medium and large event. In addition, the loss for the 7.1 event converged to a smaller value than that of the 8.8 event. The proposed methodology was only applied to earthquakes, but is broad enough to be applied to any type of peril.

Contents

List of Tables

List of Figures

List of Equations

Acknowledgements

I want to thank my family and boyfriend for their support and encouragement during my entire educational journey. I would also like to acknowledge my advisor for being patient and a great mentor throughout my entire stay at Columbia. He has been and will be the best boss I'll ever have. Not only have I had great guidance through my Ph.D. journey, but also an excellent project as well. Guy Carpenter gave me the opportunity to solve a challenging and stimulating research problem. Everyone in the company was helpful and served as additional mentor on my project. Madeleine, Guillermo, and Sergio were especially helpful because they provided feedback and guidance during my internship with Guy Carpenter and any other time I needed them. As a first generation college student it is a blessing to have received an ivy league education for free with the help of so many intelligent and genuinely kind people.

Chapter 1

Introduction

1.1 Past Applications of Bayesian Theory

Bayesian Inference is a powerful tool that is used in multiple fields to help build predictive models for different phenomena. The basic premise behind Bayes theory is that there exists some prior distribution or probability that something will occur. Then some new information becomes available that is relevant to the previous distribution. Using the new data, the prior distribution can be modified via a likelihood function to calculate an improved posterior distribution. This general framework can be applied to refine initial estimates of any given parameter by using pertinent data as it becomes available.

A plethora of Bayesian inference applications exist in technical fields like science and engineering, but there have also been applications in social science and even politics. One political application of Bayes theory was to forecast a winner in a presidential election. Linzer defined a methodology that used a combination of existing early forecasting models (prior) and real time election polls (new information) to predict the 2008 presidential election results [9]. He estimates the Bayesian model with a Markov Chain Monte Carlo sampling procedure to implement the forecast model. Election polls were used 6 months prior to the election and updates were calculated every two weeks, incorporating the most recent polls

results for each state. Using his model Linzer was able to successfully predict the election of Barack Obama in the 2008 presidential election.

Bayesian analysis can help organizations make informed decisions that mitigate exposure to natural disasters. One application in risk analysis used real time flood forecasting for the river Rhine [10]. In this case the modeled prior phenomena was the probability of exceeding a certain critical water level at a given location and time. The prior distributions were constructed using a linear regression models that used water levels of upstream stations. By using information on real observed water heights, an improved estimate of posterior water levels were predicted that could be used for risk assessments.

There have also been many applications to predict losses of buildings subjected to earthquake hazards. Usually this past research involving Bayesian loss updates tend to focus on one building at a time and defines an earthquake model, a structural model, subjects it to a simulated event and then measures the response of different components to asses damage and a monetary loss.

One Bayesian model predicted damage and loss of a seven-story reinforced concrete moment-frame building subjected to the 1971 San Fernando Earthquake [11]. Initially, the analysis used a structural analysis model to calculate predicted displacements due to the earthquake. Then it incorporated real time displacement and drift of an instrumented building via Bayes' theorem to calculate an improved posterior predicted response. Another application used reported insured monetary building loss to improve prior estimates of average losses for buildings in a given zip code when subjected to the 1994 Northridge earthquake [12].

These predictive models show the versatility of using a Bayesian framework to improve estimates in multiple fields using any and all available information that can inform the initial prior estimates. The proposed research methodology will generalize a Bayesian analysis application that can take into account multiple types of input data to improve initial loss estimate for a given portfolio of buildings due to a real time earthquake.

1.2 Research Objectives

The goal of this research is to estimate the monetary loss a portfolio of insured buildings will sustain after a major earthquake event by using all available earthquake attributes. This information is crucial for insurance companies after an earthquake strikes because they need to ensure they have enough money to pay their insured parties for their damages.

Traditionally, insurance companies provide financial protection via contracts or policies that promise partial or full reimbursement for a loss due to different kinds of incidents like work injuries, car accidents, or natural disasters. [13] The goal of an insurance policy is to share financial risk between an insured party and an insurance company. Usually the risk holder is either a person or a business but it can also be another insurance company, in which case the insurance company is insured by a (re)insurance company.

Insurance companies can choose to purchase (re)insurance to mitigate the risk that comes with insuring a large portfolio of buildings for natural disasters. In the case of a large destructive event the (re)insurance company would pay the insurance company, according to their policy, so they can pay all of their insured parties. In order to make an informed decision about what policies are most cost effective (re)insurance companies use catastrophe risk (CAT) models. CAT models estimate the amount of damage buildings within a portfolio will sustain when exposed to specific hazards and the resulting monetary loss given different policies. The model outputs average expected losses as well as the probability that a certain monetary loss will be exceeded. These models results helps a company choose a policy that provides the desired coverage for a specific level of projected risk. The insurance company decides how much they're willing to pay for an insurance so that the probability that they will exceed a target loss is below a desirable threshold.

CAT models help an insurance company chose an initial policy coverage but can also help when a real natural disaster occurs. For example, in the event of a real earthquake the insured company needs an accurate loss estimate because they need to have enough money to

pay all their insured parties. The model losses due to the set of CAT simulated earthquakes along with attributes about the real earthquake can be use used to calculate a new loss for the real event. It is expected that as additional information becomes available about the real earthquake the uncertainty in loss estimates will be reduced.

As discussed in the previous section, Bayesian analysis is a powerful tool that can help solve this very problem of updating an initial estimate based on available data. Although previous applications have been used to calculate risk for flood analysis and loss for individual buildings, applications to calculating loss for a large set of buildings has not been explored.

The proposed approach is to calculate a new loss based on the initial CAT model loss estimates for the simulated events and real earthquake parameters to calculate a new posterior loss estimate. The final posterior loss will be a combination of all the catalog losses, where each event contributes an amount proportional to its similarities with that of the real event. To quantify this contribution a comparison is established, where attributes of the real event are quantified and compared to equivalent attributes of the simulated events. Once this comparison is defined the relative loss contribution of each simulated event to the expected loss due the real event can be calculated. Assuming that all simulated catalog events are mutually exclusive and collectively exhaustive, the sum of the loss contributions will add up to unity. The calculated loss contribution of each simulated event can be interpreted as the relatively probability that an event like the simulated will occur, given the real event.

Any available attribute of the real event can be used as a basis of comparison but as a start only magnitude, location,peak ground acceleration, building tagging and damage information were used. Once the comparison is complete an updated probability of occurrence can be calculated for each simulated event. The loss due to each simulated event is given by the initial CAT model analysis, and its contribution to the total loss will change with each update step. While the proposed framework is general enough to be applied to any peril, earthquake risk is used as a prototype for the initial development process.

As additional attributes describing the real earthquake become available, the updated

estimate based on the CAT model loss output can be refined, thereby reducing the level of uncertainty. Since incoming information will not be available immediately after the earthquake strikes, there will be a subsequent loss update when each additional attribute of the real earthquake becomes available. Before a loss update can be calculated, the CAT model analysis is required, thus a general overview of the model methodology will be described in the following section.

1.3 Catastrophe Risk Models

CAT models are important tools for (re)insurance companies that help establish what the most probable financial risk is for different natural disasters. There are both private and open source CAT models that model different perils for different geographical regions. The proposed Bayesian update methodology can be applied any model, even open source ones like Global Earthquake Model. Given certain inputs these models can calculate a loss a set of assets will sustain due to a peril in a specific region. In general the seismic CAT models are composed of four main modules: a stochastic event module, a hazard module, a vulnerability module, and a financial analysis module.

Before the analysis is conducted there is set of user inputs that need to be defined. The building portfolio must be defined, which includes the physical location or coordinates of the structure as well as building configurations such as construction materials, year built and occupancy. The second user input includes information about the specific policy that the buildings are insured with, such as premium and deductibles. Once these initial inputs are defined, the CAT model can begin analyzing the buildings risk due to seismic events.

The first step in calculating the loss due to an earthquake is to have a representative set of simulated events (event catalog) for the location of your insured assets. Different regions have their own distinct set of stochastic events, for example California's event catalog would differ from the one for Japan. In the case of earthquake risk, the event module defines

different seismic sources that theoretically produce all possible seismic events that could effect the region of interest. These event catalogs can account for multiple seismic sources such as subduction zones, line sources, and background seismicity [14]. Given these sources, a stochastic catalog is formed with seismic events of magnitude, location and occurrence rates.

Once the stochastic earthquake events are defined, the hazard models are used to calculate the earthquake intensity at every property location due to every simulated earthquakes, via an attenuation model. The hazard intensity can be measured as peak ground acceleration (pga), spectral acceleration(Sa) or Modified Mercalli Intensity (MMI) Scale. An attenuation model calculates a hazard intensity at a building location due to a catalog event by using principle attributes of the simulated event. These principle event attributes vary between models, but usually include fault type, magnitude of the event and epicenter distance. Despite different attenuation models, in the general the greater the event magnitude and the closer you are the event epicenter the larger the hazard intensity and the further away the smaller the hazard intensity. After the attenuation model is used, additional factors like local site conditions are used to calculate what the amplification of the intensity will be at each given site [14].

Once the hazard demand is defined, the vulnerability module uses the specific attributes of each building to convert the hazard intensity into a mean damage ratio. This is where the specific characteristics of the insured building come into play. The building construction material, age, number of stories, occupancy class and secondary characteristics help define specific damage curves for each asset. These damage curves define what type of damage a building with specific attributes would sustain when subjected a hazard intensity. The specific damage curves vary between models, but they are validated with physical tests of instrumented buildings and building components that are subjected to different seismic events. Given the demand(seismic intensity) and capacity of the each property the damage curve calculates a mean damage ratio for every simulated catalog event [14].

The last step is the financial analysis module, which uses mean damage ratios along with information about the assets value and policy types to calculate a monetary loss due to each stochastic seismic event [14]. The final CAT model(the financial model.) output is a monetary loss value for every building due to every simulated earthquake. From these computed losses and the rates of each simulated earthquake, different loss statistics can be calculated such as average annual loss and exceedance probability curves. Ultimately the model calculates a total loss due to each simulated event, which is associated with different return periods. By looking at the overall distribution of losses, an insurance company can make an informed decision about which policy terms result in an acceptable loss margin.

Chapter 2

Methodology at the Conceptual Level

2.1 Utility Theory

The goal of this research is to calculate a monetary loss for a portfolio of buildings due to a real earthquake event. One of the primary tools used to calculate this loss is the existing loss output of a catastrophe (CAT) risk model. Theoretically the catastrophe risk model has already been used to calculate a loss for the existing portfolio due its own set of simulated catalog of events. Since it is unlikely that the real event will be identical to any of the simulated events, the goal is use the losses due to all of the simulated events $1 : N_{EQ}$ to calculate the new loss due the real event. If the set of catalog events $1 : N_{eq}$ is collectively exhaustive and mutually exclusive then the target loss prediction due to the real event can be calculated as a linear combination of the simulated losses via the use of utility theory.

Utility theory is used in economics to make decisions by calculating an expected cost of an event given some uncertainty. The expected utility of a primary event B is the sum of all the probabilities that a sub-event n=1:N will occur ,P_n, times its associated cost, C_n [15]. The set of sub events n=1:N are mutually exclusive and collectively exhaustive, thus together

they span the entire subspace of primary event B.

$$E[B] = \sum_{n=1}^{N} P_n * C_n \tag{2.1}$$

Let us assume that someone is trying to decide whether or not to accept a bet. They are told if they pay \$4 they can roll a die and will be give the dollar equivalent of the number on the die. In order to make this decision they decide to calculate the expected profit if the proposal to gamble is either accepted or rejected. In order to apply the theory the total number of possible outcomes must be identified and they must be both mutually exclusive and collectively exhaustive. There are a total of six outcomes; rolling a one, two, three, four, five or six since there are only six faces to a die. It is impossible to roll one number and another number at the same time thus the 6 outcomes are mutually exclusive. In addition six and only six outcomes are possible because you are constrained to roll one of the six given faces of a die, thus the six outcomes are collectively exhaustive. Once these two initial criterion are met the utility can be calculated for both outcomes.

In order to apply the expected utility, first the probability of rolling a specific die face must be calculated. Since the probability of rolling any given number is equal and there are a total of six faces to a die, the probability of rolling any given number is 1/6. Next the utility for each of the 6 outcomes must also be calculated, which is given as \$1, \$2,\$3, \$4,\$5 and \$6. If equation 2.2 is used with the given information the expected utility for accepting the bet is given by the following equation.

$$E[Bet] = 1/6 * \$1 + 1/6 * \$2 + 1/6 * \$3 + 1/6 * \$4 + 1/6 * \$5 + 1/6 * \$6 = \$3.50$$

Since the expected profit of accepting the bet exceeds the cost to enter the bet the rational decision would be to reject the initial offer. This general framework of calculating total cost or loss via contribution of a set of mutually exclusive and collectively exhaustive events can easily be applied to calculate an expected loss of a portfolio due to a real event. Just as

calculating an expected value for the the gambling problem helped the user make a decision, calculating an expected loss due to an earthquake can help an (re)insurance make a decision of whether or not to borrow additional money to pay their insured parties. As before in order to apply the theory, it is required that the catalog earthquakes form a set of mutually exclusive and collectively exhaustive events. Equation 2.2 can be used to calculate an average expected cost or loss due to a real earthquake B by using a set of catalog earthquake events $A_1 : A_{N_{EQ}}$ with a given loss $Loss(A_n)$ and probability of occurrence, $P(A_n)$.

$$E[Loss(B)] = \sum_{n=1}^{N_{EQ}} Loss(A_n) * P(A_n) \tag{2.2}$$

It is assumed that the CAT model has generated a set of simulated earthquakes denoted by A_n where $n = 1 : N_{EQ}$ that can occur in a given geographical area of interest (ie:location of real event B). The given set of A_n are mutually exclusive and collectively exhaustive. $Loss(An)$ represents the loss of the portfolio buildings due to simulated event An, which is given by the CAT model for all events. $P(A_n)$ represents the probability that simulated catalog event A_n will occur. Calculating this probability is determined by the relation between the simulated events and the real earthquake, event B. These probabilities can be considered the weight or contribution each simulated event has to the total real loss.

Before the event occurs there is no information to quantify the simulated event's contribution to the total real loss, thus it is assumed that they contribute equally. The initial probability that a given event will occur is equal to $1/N_{EQ}$. Once information becomes available for event B, a comparison between attributes of the simulated events and the real earthquake can be used to define the probability of occurrence for each simulated event. The objective of the Bayesian framework is to determine via updating better estimates of $P(A_n)$ using observed attributes of the real earthquake.

2.2 Bayesian Framework

Bayesian analysis is a useful analytical tool to calculate the loss estimate of a portfolio of buildings to a real event. The foundation for Bayesian updating is Bayes' Theorem, which states that a posterior probability for update step m ($P_m(A_n|B_m)$) is equal to the product of a prior probability ($P_m(A_n)$) and a likelihood function ($P_m(B_m|A_n)$) divided by a normalization factor ($P_m(B_m)$) [16]:

$$P_m(A_n|B_m) = \frac{P_m(A_n)P_m(B_m|A_n)}{P_m(B_m)} = \frac{P_m(A_n)P_m(B_m|A_n)}{\sum_{n=1}^{N_{EQ}} P_m(A_n)P_m(B_m|A_n)};\ m = 1, 2, 3..., M \quad (2.3)$$

The components of Bayes' theorem are defined in terms of conditional probabilities and the general idea is to use knowledge of observed attributes to improve an initial estimate of some unknown probability distribution. These attributes can include an earthquakes magnitudes, location, peak ground acceleration and corresponding damage to buildings. The update procedure can be an iterative process because every time there is new information about an attribute, there will be a new update. Given M attributes of a real earthquake, there will be M corresponding updates. In order to calculate the posterior estimate for a given update step m, the prior, likelihood and normalization factor must be properly defined.

The posterior probability ($P_m(A_n|B_m)$) is the conditional probability of occurrence for simulated earthquake event A_n, given all the available observed attributes of the actual event B_m at update step m. An important goal of this research is to calculate the posterior probability distribution using all available observed information about the real earthquake.

The initial prior ($P_0(A_n)$) is the probability that a given simulated earthquake event A_n will occur, before having any information about real earthquake event B. Before the updating scheme begins, $P_0(A_n) = 1/N_{EQ}; n = 1 : N_{EQ}$ assuming that all simulated events in the catalog are equally likely to occur. With each subsequent update step m, the prior probability $P_m(A_n)$ is set equal to the posterior probability $P_{m-1}(A_n|B_{m-1})$ from the previous

step $(m-1)$. After the last update, the final posterior probability incorporates all the available observed information about the real earthquake event B.

The likelihood function ($P_m(B_m|A_n)$) is the conditional probability of the real earthquake, event B, given the simulated event A_n. The likelihood function (LF) is a measure of how alike the attributes (magnitude, location, damage, etc.) of the simulated earthquake event A_n are to the corresponding ones of the real earthquake event B. The normalization factor $P_m(B)$ is the standard total probability of having a real earthquake event B and if N_{EQ} is defined to be the total number of simulated events, then it can be easily calculated by equation 2.4.

$$P_m(B) = \sum_{n=1}^{N_{EQ}} P_m(A_n) P_m(B|A_n) \tag{2.4}$$

The normalization factor ensures that the integral of the posterior probability density function will equal one, representing a true PDF probability.

Because the initial prior probability is set equal to a common constant value $1/N_{EQ}$ for all N_{EQ} in the catalog and the posterior probabilities are computed using Bayesian updating, the main challenge of this research is therefore to define the LF. Defining the LF should reflect how the actual earthquake attributes relate to the corresponding attributes of the catalog events. The earthquake attributes can include any quantity describing an earthquake and its consequences, but for simplicity, in this study, they were selected as the magnitude, location, peak ground acceleration and various forms of damage. If there were a simulated event in the catalog with the exact same attributes as the real time earthquake, the LF would take on its maximum value. But since the earthquake attributes of the simulated events usually differ from those of the real event, the LF is selected to decay and asymptotically approach a zero value away from the true values. Many continuous functions could satisfy these properties, but a Gaussian (normal) probability distribution was chosen to represent the LF in the proposed methodology.

2.3 Likelihood Function Definitions

A major hurdle in Bayesian analysis is choosing an appropriate form for the likelihood function. There are many possible choices depending on the data being represented. A common choice for the likelihood function is a Gaussian PDF, which was ultimately chosen for this application. An example where this is used is for the previously mentioned case where an instrumented building and analytical model were used to calculate an updated loss estimate due to the 1971 San Fernando Earthquake [11]. In this application multiple stochastic model trials were ran in order to find the demand on the building structural components. Initially each trial was given an equal weight, constant prior equal to 1 divided by number of trials. In order to calculate an appropriate weight for each Monte Carlo simulation, a Gaussian likelihood was constructed. The likelihood function measured how close the simulation demand was to those recorded using the building instrumentation. This Bayesian methodology is very similar to the earthquake loss methodology presented. Just as the initial model weights are set equal to a constant, the prior probability for each simulated CAT model is set equal to a constant. Also, the goal of the updates is to reconcile the differences between the Monte Carlo model and the buildings real behavior, and for the EQ loss estimation the goal is to reconcile the difference between the attributes of the real event and the simulated CAT model events. [11]

Another potential form of a likelihood function is an exponential PDF. An example of this application was a study that used Bayesian analysis to calculate life cycle loss for timber buildings by performing Monte Carlo simulations using component level fragility curves and loss distributions [17] . The goal was to model different parts of the building structure, such as shear walls or windows, in order to define the probability of exceeding a specific damage state given a demand displacement or acceleration. Initially the probability of a component exceeding any given damage state was equally likely and set equal to a constant (constant Prior). Next, in order to modify the models damage prediction case studies of real lab test

of timber buildings were used. By comparing the simulated Monte Carlo response with the test data, exponential likelihood functions were constructed to modify the initial constant prior distributions.

Both cases had some similarities with the proposed methodology. In both cases constant priors were used because before any available data it was assumed that every possible scenario was equally likely to occur. Even though the timber loss estimation used a exponential likelihood function, the basic goal is the same: establish a weight for an experimental outcome given available data on a real event and real response.

2.4 Bayesian Update Example

To illustrate the Bayesian process, a sample portfolio of 5 simulated earthquake events (the catalog events) were used in a single update step. Five randomly selected earthquake magnitudes, equal to $M_{A_n} = [0.35, 3.00, 4.39, 7.00, 9.50]$, were used. Before any information about a real event is available, every event is equally likely to occur. The sum of the discrete prior density function must equal unity and be equal to a constant, thus the prior $P_1(A_n)$ for update step 1 is equal to $1/N_{EQ} = 1/5 = 0.2$ for all of these 5 events. This non-informative prior probability density distribution is shown in Figure 2.1. In addition to information about the event magnitude, an analysis has also been conducted that indicates the loss a single building C would sustain if subjected to the catalog events $L^C_{A1}, L^C_{A2}, L^C_{A3}, L^C_{A4}, L^C_{A5}$. Table 2.1 summarizes the preliminary model losses for all 5 simulated events for building C. If there is no additional information then an expected loss can be calculated with the baseline prior distribution and CAT model losses. $E[L] = 0.2 * (0 + 200 + 500 + 1000 + 4000) = \$1140$. In this case all events contribute equally to the predicted loss.

Table 2.1: Model Losses

Event	A1	A2	A3	A4	A5
$\mathbf{L^C_{An}}$	0	200	500	1000	4000

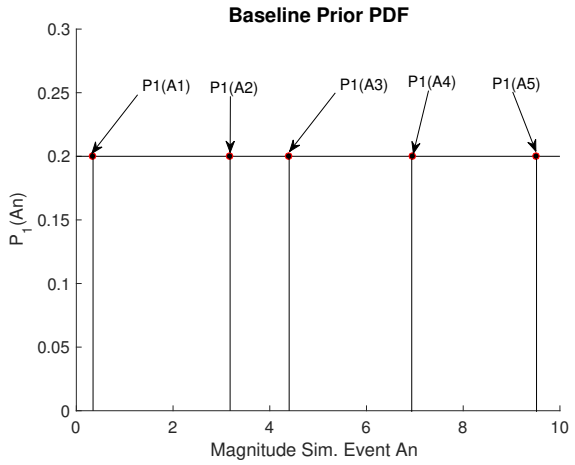

Figure 2.1: Prior Probability Density Function

After some time, a real earthquake, event B, with magnitude $M_B = 3.82$ occurs. The goal becomes to use the initial loss estimates, from the previous analysis, along with information about event B to calculate a new loss estimate for building C due to event B, L_B^C. Any information about the earthquake can potentially be used as an update parameter. If initially only information about the event magnitudes is available, it can be used to calculate a single updated loss. The basic idea is to calculate the new loss as a linear combination of the losses due the event catalog using different weights. These weights are defined via a probability density function that is calculated using Bayesian analysis.

As mentioned in the previous section, in order to calculate an updated posterior PDF a likelihood function(LF) must be defined. A normal distribution was chosen as the likelihood $P_1(B_1|A_n)$ centered at the magnitude value of the real event $\mu_1 = M_B$ with standard deviation equal to $\sigma_1 = 0.5 * [max(M_n) - min(M_n)] = 0.5 * (9.5 - 0.35) = 4.58$. Once the LF is defined it must be evaluated at the magnitude values of the simulated events M_{A_n}. The catalog event magnitudes closer to the real event are weighted more than those further away. Figure 2.2 shows the LF and the evaluated points for the five simulated catalog events.

Once the LF is defined and evaluated, the next step is is to calculate the unnormalized posterior probability density function by multiplying the prior with the likelihood values for all 5 simulated events. Initially, the prior is a constant so the posterior pdf is the same shape

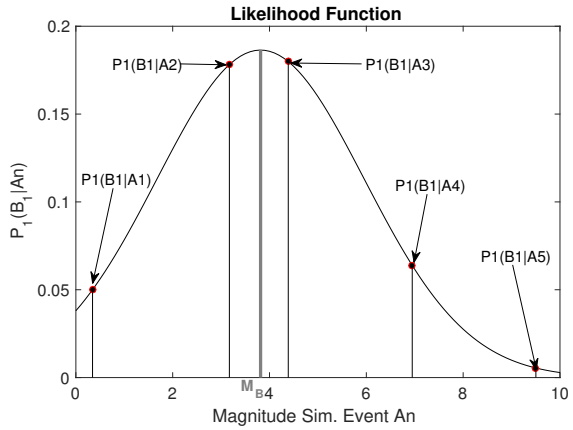

Figure 2.2: Likelihood Probability Density Function

as the LF, but as additional updates occur this will change. Figure 2.3 shows the discrete posterior pdf values for all 5 simulated events.

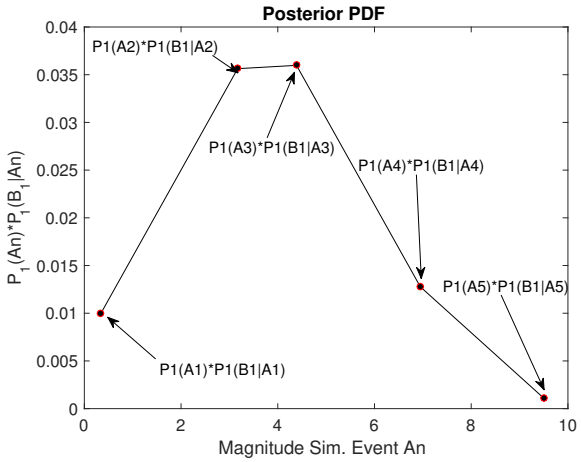

Figure 2.3: Posterior Probability Density Function

The last step of Bayesian approach is to normalize the distribution by the sum of all the unnormalized posteriors to ensure the sum of the final discrete posterior PDF adds up to unity. The final normalized posterior pdf is shown in Figure 2.4 This final step is completed to ensure the final distribution is a true probability density function. In addition it ensures set of simulated events are mutually exclusive collectively exhaustive, since the events make up the entire range of possible outcomes without any overlap. Even though the magnitude was chosen as an event attribute, any available information of the real earthquake can be

used to perform an update.

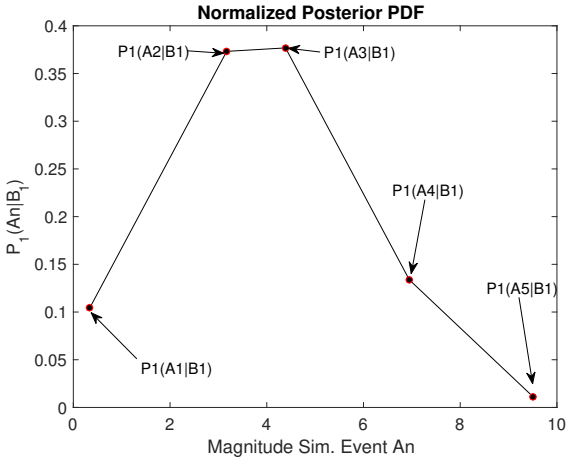

Figure 2.4: Normalized Posterior Probability Density Function

After the Bayesian analysis is complete data post processing is performed to calculate a loss estimate for building C due to event B, L_B^C. For simplicity a portfolio with a single building was considered with a corresponding loss for all five simulated events (Table 2.1). By applying the basic utility theory, the expected loss for the real earthquake event B can be calculated using the posterior PDF calculated using the Bayesian approach and the model loss values. The contribution of each earthquake to the total loss of building C is the simulated loss from the initial analysis, L_{An}^C, times the posterior PDF, $P_1(A_n|B_1)$, for all five events.

$$L_B^C = E[Loss(B_1)] = \sum_{n=1}^{N_{EQ}} L_{An}^C * P_1(A_n|B_1) \tag{2.5}$$

Table 2.2 shows the contribution of each simulated event and the total estimate for the first update step. In general a Bayesian loss estimate will be larger or equal to the smallest analysis loss and smaller or equal to the largest analysis loss. The loss contribution of event A_3 is the largest because its posterior PDF value is the largest. Yet events that have a small posterior value, but still have a large initial loss estimate, can still contribute a large loss like the case of event A_4. The final loss estimate for building C due to event B is between the initial loss estimate of event three and four, which makes sense because they are closest in magnitude with the real event. This value is much lower than the initial estimate of over

Table 2.2: Bayesian Loss Estimation Example

Simulated Event	$\mathbf{L^C_{A_n}}$	$\mathbf{P_1(A_n|B_1)}$	**Loss Contribution**
A1	0	0.10	0
A2	200	0.37	74
A3	500	0.38	190
A4	1000	0.13	130
A5	4000	0.011	44
		$\mathbf{E[Loss(B_1)]}$	364

$1000.

Although this example used only one building and five events in the catalog, it could easily be expanded to include multiple buildings and events. The basic Bayesian procedure remains the same for ever update step, any given earthquake attribute, number of simulated events and number of buildings. As long as there is an equivalent attribute for the CAT model events, any available information from a real event can be used to perform a loss update.

Chapter 3

Methodology at the Application Level

3.1 Potential Forms of Reported Earthquake Data

The goal of the Bayesian update procedure is to use all available real information to update an initial loss estimate. The first available earthquake information will usually be the latitude, longitude and magnitude of the event, which is already in quantitative form, thus it can be used directly for the first update. Additional information about the earthquake will continue to arise, enabling subsequent updates of the loss estimate and further reduction in its uncertainty. The type of information and the time at which it becomes available will vary, but it will be helpful to represent all information in quantitative forms so it can be incorporated into the Bayesian update process.

In general there may be different preliminary qualitative descriptions of damage for specific buildings. In order to incorporate them into the Bayesian update framework they have to be mapped to a damage level between 0 and 1 where 0 would correspond to no damage and 1 to collapse. There will be different qualitative descriptions, so it may be difficult to standardize the way the descriptor is mapped to the damage index. In order to standardize this mapping, a guideline can be used that defines different damage ranges and defines what damage is characteristic of each range.

Another piece of information that becomes available post-event is the time history of the earthquake at locations where accelerometers have been installed. At these locations, the peak ground acceleration (PGA) can be extracted. The CAT models also contain similar hazard intensity estimates for each simulated earthquake event to which the real PGA values can be compared to. If the frequency time history is available for both the real and simulated earthquakes it can be used as a parameter to compute new likelihood functions to update the current loss estimate.

Satellite images and drone footage of the terrain may also be available. They can come from different organizations like USGS remote sensing, NASA and private organizations. Processing this information to a corresponding damage level is possible but may be computationally challenging because of large size of the images. Currently, programs and algorithms exist that help quantify a loss incurred by a structure by comparing before and after images. If the damage is calculated and converted into a damage scale of zero to one it can be used for one of the loss updates.

Loss estimates from insurance companies may also become available. In order to normalize these values, they can be separated by building types and locations. If additional information about the building type isn't available, the loss can be normalized by the number of building in the portfolio and used as rough reference point for the loss.

In addition, nonstructural loss is associated with loss of use when a business cannot operate due to the seismic event (business interruption). Newspapers and news chains may share information on which businesses are closed and for how long they remain closed. The monetary loss for each day it remains closed would be equal to the projected profit per day of the business, which can be added to the structural monetary loss. Another source of data may be social media sites such as Facebook, Twitter and Instagram. After the earthquake occurs, it would be likely that pictures and descriptions of the subsequent damage would be posted in the affected areas. These qualitative descriptions and images would have to be collected and quantified into a damage index scale of zero to one similar to any other

Table 3.1: Types of Earthquake Data

Order	Information	Source	Quantification
1	Magnitude, Location, Depth	USGS	Magnitude, latitude, longitude, depth
2	Initial Damage Reports and First Loss Estimations	News reports, engineering surveys, insurance companies	Map qualitative description [no damage, collapse] to scale [0, 1]
3	EQ Time Histories	USGS	PGA (or other intensity metric) at accelerometer locations
4	Satellite Images	USGS remote sensing, NASA, Private Orgs.	Before and after SAR images, complex coherence and intensity correlation, convert to [0,1]
5	Drone Footage	Government agencies, insurance companies	Construct 3D image of building, also quantify damage by visual inspection of video and convert [0,1]
6	Business Interruption	Newspaper	Loss of use/day= projected profits/day; number of days the business is closed
7	Reports from Social Media	Facebook, Twitter, Instagram	TBD (convert descriptions and pictures to quantification)
8	Final Real Loss Experience	Inspections and Claims	Last official monetary loss based on claims data

observed damage for the building portfolio.

The last level of information to become available would be the true losses suffered by an insurance company, assessed after processing all actual claims. By the time this information is known, an adequate estimate should already be established by the Bayesian update process. Nevertheless, the real loss estimate can be used to validate or calibrate the Bayesian model to improve further estimates, particularly in the development process. Ultimately, incorporating information about the real earthquake via the use of Bayesian updates will help improve the loss estimate for any portfolio of insured buildings in a timely manner. All of the proposed update information is summarized in Table 3.1.

3.2 Required Analysis Input

The Bayesian update procedure requires two set of parameters, one from the simulated events ($\mathbf{X^{(n)}} = [X_1^{(n)}, X_2^{(n)} \dots X_M^{(n)}]$) and an equivalent vector from an observed real time event ($\mathbf{X^{(*)}} = [X_1^{(*)}, X_2^{(*)} \dots X_M^{(*)}]$). These parameters will vary depending on the update step and available real event information. The first update parameters will be the earth-

quake magnitude and location. The observed parameter is the magnitude and location reported by USGS ($X_1^{(*)} = [M_1^{(*)}, D_1^{(*)}]$) and the equivalent model values are each simulated event's magnitude and location ($X_1^{(n)} = [M_1^{(n)}, D_1^{(n)}]$). An additional parameter used was peak ground acceleration both from a real recorded event ($X_m^{(*)} = PGA_m^{(*)}$) and the model simulated earthquakes($X_m^{(n)} = PGA_m^{(n)}$). The next set of parameters were loss by location by event and building type. In this case, the observed real time parameter would be a reported level of damage for a specific region for a specific building type ($X_m^{(*)} = DI_m^{(*)}$). The equivalent model output would be damage for each simulated earthquake events for that same building exposure and the same location where the actual damage was reported ($X_m^{(n)} = DI_m^{(n)}$). One last real time parameter used was building tagging inspections of red, yellow, and green corresponding to unsafe to enter, limited entry and no apparent hazard [8]. These color designations were converted to a damage index between zero and one ($X_m^{(*)} = BT_m^{(*)}$). This building tagging indice was compared to damage for equivalent model buildings in the same region as the tagged ones ($X_m^{(n)} = BT_m^{(n)}$). In order to implement this proposed update procedure, these two set of parameters need to be collected and quantified. The real earthquake parameters were obtained using a medium and large magnitude earthquake that occurred in Chile and one medium event in Southern California. To generate the equivalent simulated parameters two CAT model portfolios were analyzed, one for California and another for Chile.

3.3 Analysis Procedure

The final output of the Bayesian update procedure is the contribution each simulated events has to the loss due the real event being considered. Every update incorporates a new observed characteristic of the real event $X_m^{(*)}$ and compares it to an equivalent simulated model characteristic $X_m^{(n)}$. The comparison of these two sets of parameters via Bayesian analysis establishes a independent weight for each catalog event: a discrete probability density function

Table 3.2: Real and Simulated Input Parameters

Real Earthquake Parameter ($X_m^{(*)}$)	Simulated Earthquake Parameter ($X_m^{(n)}$)
1. Magnitude and Location	**1. Magnitude and Location**
$[\widetilde{M}_1^{(*)}, \widetilde{D}_1^{(*)}] \xrightarrow{\text{scale [0,1]}} [M_1^*, D_1^{(*)}]$	$\begin{bmatrix} \widetilde{M}_1^{(1)} & \widetilde{D}_1^{(1)} \\ \widetilde{M}_1^{(2)} & \widetilde{D}_2^{(2)} \\ \vdots & \vdots \\ \widetilde{M}_1^{(N)} & \widetilde{D}_1^{(N)} \end{bmatrix} \xrightarrow{\text{scale [0,1]}} \begin{bmatrix} M_1^{(1)} & D_1^{(1)} \\ M_1^{(2)} & D_1^{(2)} \\ \vdots & \vdots \\ M_1^{(N)} & D_1^{(N)} \end{bmatrix}$
2. Peak Ground Acceleration	**2. Model PGA for building closest to station**
$[PGA_m^*]$	$\begin{bmatrix} PGA_m^{(1)} \\ PGA_m^{(2)} \\ \vdots \\ PGA_m^{(N)} \end{bmatrix}$
3. Reported Damage: Bld type i and Region j	**3. Ground up loss**
$\begin{bmatrix} damage \\ description \end{bmatrix} \xrightarrow{\text{scale [0,1]}} \begin{bmatrix} DI_m^{(*)} \end{bmatrix}$	$\begin{bmatrix} \widetilde{L}_m^{(1)} \\ \widetilde{L}_m^{(2)} \\ \vdots \\ \widetilde{L}_m^{(N)} \end{bmatrix} \xrightarrow{\text{scale [0,1]}} \begin{bmatrix} DI_m^{(1)} \\ DI_m^{(2)} \\ \vdots \\ DI_m^{(N)} \end{bmatrix}$
4. Reported Bld Tagging: Bld type i and Region j	**4. Ground up loss**
$\begin{bmatrix} R, Y, G \\ Tag \end{bmatrix} \xrightarrow{\text{scale [0,1]}} \begin{bmatrix} BT_m^{(*)} \end{bmatrix}$	$\begin{bmatrix} \widetilde{L}_m^{(1)} \\ \widetilde{L}_m^{(2)} \\ \vdots \\ \widetilde{L}_m^{(N)} \end{bmatrix} \xrightarrow{\text{scale [0,1]}} \begin{bmatrix} BT_m^{(1)} \\ BT_m^{(2)} \\ \vdots \\ BT3_m^{(N)} \end{bmatrix}$

(PDF) of the occurrence for each simulated event. The final step is to link each simulated event to a real portfolio of loss, which results in a PDF of the probability of reaching a specific portfolio loss. Figure 3.1 shows the steps necessary to perform the Bayesian update procedure.

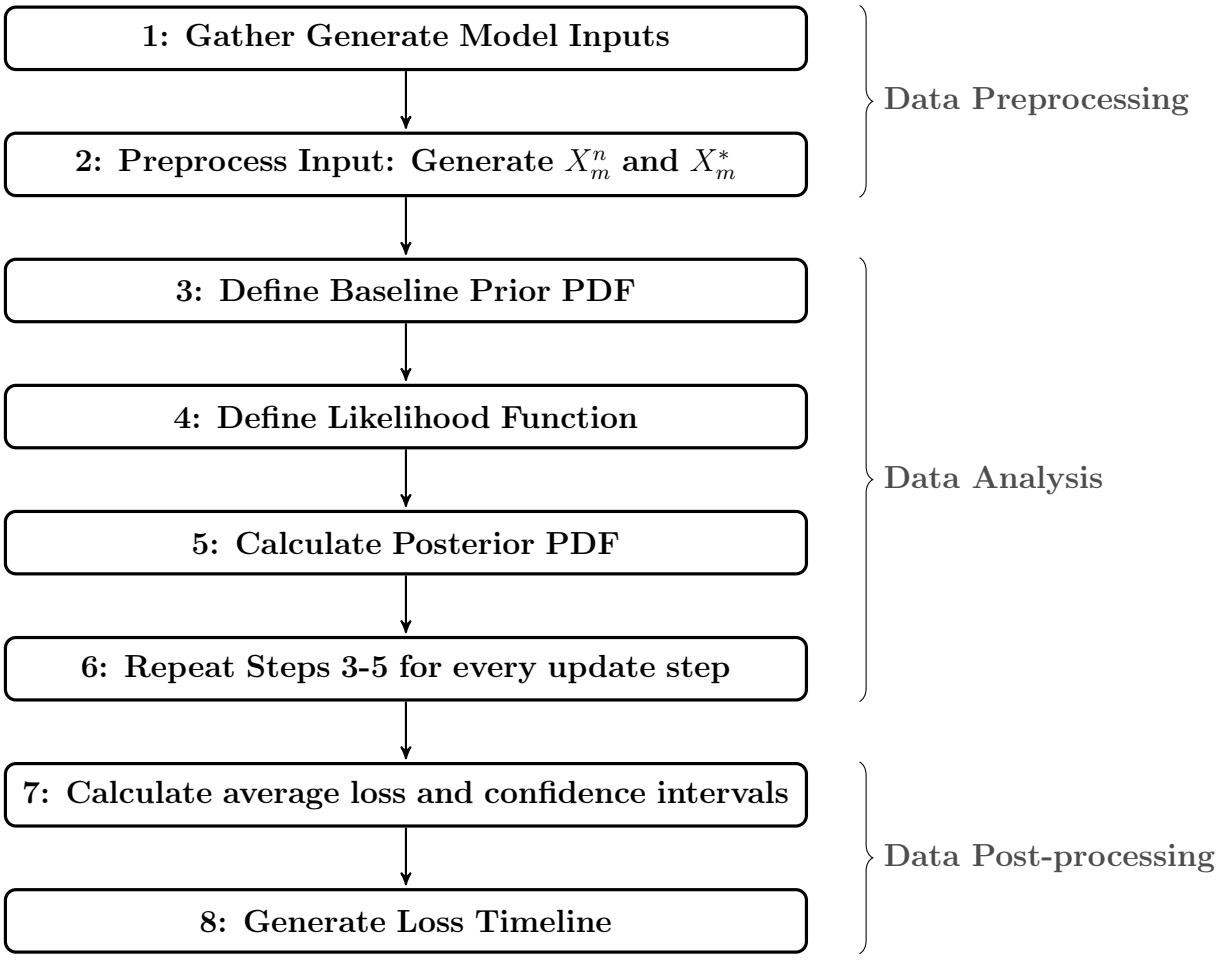

Figure 3.1: Analysis Procedure Flowchart

3.3.1 Gather/Generate Model Input Data

The first step in the Bayesian update procedure is to collect the necessary data to perform the updates. There are always two sets of required input: one from the real event and an equivalent set for each simulated event in the catalog.The real event descriptions were collected from a series of surveys, reports, USGS databases and journal publications. The equivalent parameters for the simulated events were either extracted or generated through CAT models analysis. Table 3.2 shows a comparison of the real and simulated parameters used for the analysis.

3.3.2 Preprocess Input

3.3.2.1 Scaled Magnitude and Location Input

The first step of data preprocessing consists of quantifying and possibly scaling the observed parameter value. The first update step uses the first available earthquake parameters: magnitude and location. The magnitude and location are already in quantitative form but for convenience they are all scaled to fall within a range of approximately 0 to 1. The scaled magnitude is equal to the unscaled magnitude minus the minimum reasonable magnitude divided by the difference of the maximum and minimum reasonable projected magnitude. To calculate the minimum and maximum projected magnitude the event catalog earthquake magnitudes were used.

$$M_m^{(n)} = \frac{\widetilde{M}_m^{(n)} - M_{min}}{M_{max} - M_{min}}; \tag{3.1}$$

The latitude/longitude input was mapped to a single distance in order to reduce the number of variables. The Haversine formula (Equation: 3.2) was used to calculate the distance between the real earthquakes and the simulated earthquakes [18]. The Haversine formula assumes the earth is a perfect sphere with radius R and calculates the arch-length between two points on the surface of the earth. The value R was set to 6371 km and subscripts one and two in equation 3.2 correspond to the simulated earthquake and real time earthquake respectively.

$$D = 2R \arcsin\sqrt{sin^2\left(\frac{lat_2 - lat_1}{2}\right) + cos(lat_1)cos(lat_2)sin^2\left(\frac{long_2 - long_1}{2}\right)} \tag{3.2}$$

Once the distance was calculated, it was also scaled to fall within an approximate range from 0 to 1. The minimum reasonable distance between a simulated earthquake and real earthquake was set to 0 km and the maximum was set equal to the length of Chile 4300km.

Thus the scaled distance is equal to the unscaled distance minus the minimum reasonable distance (0km) divided by the difference of the maximum (4300km) and minimum (0km) reasonable projected distance. For the California catalog, the minimum reasonable distance was also set to 0km but the largest to 700km because the focus was only on damage to southern California.

$$Chile : D_m^{(n)} = \frac{\widetilde{D}_m^{(n)} - 0km}{4300km - 0km}; \quad Northridge : D_m^{(n)} = \frac{\widetilde{D}_m^{(n)} - 0km}{700km - 0km} \tag{3.3}$$

The first set of simulated earthquake parameters are the scaled magnitude and scaled distance: $[X_1^{(n)}] = [M_1^{(n)}, D_1^{(n)}]$. The equivalent first real earthquake parameter $X_1^{(*)}$ is the scaled real earthquake magnitude and distance $[X_1^{(*)}] = [M_1^{(*)}, D_1^{(*)}]$. The scaled real earthquake magnitude is calculated using equation 3.1 by replacing the simulated earthquake magnitude with the real earthquake magnitude. The scaled real distance is equal to 0 because the distance of the real earthquake to itself is 0. Thus the first real time parameter is equal to $X_1^{(*)} = [0.76, 0]$ and $X_1^{(*)} = [0.42, 0]$ for the large and medium Chilean events and $X_1^{(*)} = [0.49, 0]$ for the medium California event.

3.3.2.2 Peak Ground Acceleration

The second type of information used was the peak ground acceleration at different accelerometer locations. Larger destructive events tend to have more recorded data but some accelerometer records were present for all three real earthquakes. The PGA values were in units of g, acceleration due to gravity, and the coordinates of the stations in latitude and longitude.

$$[X_m^{(*)}] = [PGA_m^{(*)}] \tag{3.4}$$

After the real PGA parameters were stored, the corresponding simulated PGA parameters were calculated. First a CAT model analysis case was chosen, which is arbitrary because PGA is independent of the building properties, unlike other factors like spectral accelera-

tion and damage. Once an analysis case was chosen, buildings within a 5km radius of the accelerometer location were identified ($1 : N_{bld}^{5km}$). Ideally a CAT model building should coincide with the real accelerometer location, but a pre-established 5km grid for the analyses made this difficult. After the model buildings were identified, the PGA for each simulated earthquake was averaged for these buildings for each real PGA input.

$$[X_m^{(n)}] = \sum_{i=1:N_{bld}^{5km}}^{N_{bld}^{5km}} \frac{PGA_m^{(n)}(i)}{N_{bld}^{5km}} \tag{3.5}$$

3.3.2.3 Scaled Damage Update Input Data

All subsequent update parameters used real reported damage from survey results and on site investigations. The qualitative observed damage must be mapped to quantitative damage values before they can be used. In general, the damage descriptions included distinctions between structural and non-structural damage. In addition to the reported damage state some also included possible reasons for discrepancies. For example, in general newer, buildings performed better than older ones, but in some cases a building would suffer complete collapse and a possible design or construction problems were cited. In order to quantify the damage state both of these factors must be considered.

First the damage descriptions were mapped to a numerical value of 0 to 1. Damage ratios were mapped to different intervals: [0-0.3] for "non-structural damage and contents damage", [0.3-0.5] for "damaged", [0.5-0.8] for "severe damage" and [0.8-1] for partial collapse to total collapse. Next this value was modified to introduce uncertainty in loss reports or inconsistencies in the building. Uncertainties can include construction or design irregularities that would result in a larger loss than expected for a building with similar exposure attributes.

For the 1997 event, some of the damage reports were already quantified for the number of buildings falling within a damage range of 0 to 5. In order to process this data the damage range was converted to a damage range, d_i, from 0 to 1: [0,1,2,3,4,5]= [0,0.1,0.3,0.5,0.7,0.9].

Next a weighted average was taken where the number buildings falling within each of the six damage ranges (n_i) were multiplied by the damage range (d_i) and divided by the total number of buildings.

$$DI_m^{(*)} = \frac{\sum_{i=0}^{5} d_i * n_i}{\sum_{i=0}^{5} n_i} \tag{3.6}$$

The real damage parameters constituted a vector of numerical values between zero and one with length equal to the number of observed damage entries for update step m.

$$X_m^{(*)} = DI_m^{(*)} \tag{3.7}$$

In order to establish the Bayesian update procedure, the simulated loss data must be consistent over all analysis cases. The CAT model considers every simulated event but it omits results for those that don't produce a loss. Each CAT model analysis, of a specific building portfolio with unique building properties, had a set of catalog earthquakes that produce loss, $EQ_{port,i}$. The union of of these events $EQ_{port,i}$ was calculated for all portfolios $i = 1 : N_{port}$. The final set of unique earthquake catalog events was a $[N_{EQ}, 1]$ matrix, where N_{EQ} was the total number of unique earthquake catalog events that produced a loss, for all portfolios $1 : N_{port}$.

$$EQ_{unique} = \bigcup_{i=1}^{N_{port}} EQ_{port,i}; \quad dim(EQ_{unique}) = [N_{EQ} x 1] \tag{3.8}$$

In order to be consistent with the update procedure, only this unique set of simulated events was considered. For every portfolio, a global loss array ($\mathbf{L_g}$) was constructed with the number of rows equal to the total number of building locations (N_{bld}) and the columns equal to the number of unique simulated events (N_{EQ}) used in the CAT model analysis. For a given analysis the loss for the events that were not included in the analysis were set equal to zero. $\mathbf{L_g}(\mathbf{1}, \mathbf{3})$ would correspond to the loss of building 1 due to simulated event 3.

$$dim(\mathbf{L_G}) = [N_{bld}\ x\ N_{EQ}] \tag{3.9}$$

Once the global loss matrix is generated, calculating the simulated damage indices becomes simple. First the relevant CAT model analysis case must be identified for each observed real damage report. For example, if a real damage index corresponded to a 3 story reinforced concrete building then the CAT model analysis case that corresponds to 3 story reinforced concrete buildings must be identified. Next the simulated buildings in the region where the real damage report was documented must be identified. Thus, the rows in the global loss array,($\mathbf{L_g}$) , with portfolio buildings in the region of damage must be selected. Then the simulated losses for all the buildings in the region of interest in the CAT model of interest for each unique simulated event must be averaged. For example, if there were two portfolio buildings in the reported damage area, then the loss for these two buildings (2 rows in ($\mathbf{L_g}$)) must be added and divided by 2. Lastly, this average loss must be divided by the building values, which were all set equal to \$1,000,000. For each update step n, the simulated damage index parameter is a vector with a DI for each unique simulated event.

$$X_m^{(n)} = \frac{L_m^{(n)}}{1000000} = [X_m^{(1)}; X_m^{(2)}; \ldots; X_m^{(N)}] \tag{3.10}$$

3.3.2.4 Building Tagging Data

After some events, agencies conduct preliminary building inspection where they designate the level of damage via the colors red, yellow and green. Red signifies imminent collapse and access to the building is restricted, yellow signifies some damage and limited access, green signifies no major structural damage [8]. In order to quantify this data into a real parameter, it was converted into a building tagging damage index similar to that for reported damage. The building tagging information was reported for different cities in southern California and for four building types: masonry, steel, concrete and timber [8]. Different values were assigned to each tagging designation Red: 0.8,0.9, yellow: 0.6,0.5 and green: 0.2,0.3 resulting in 6 cases . The analysis with these different factors was conducted until one set yielded an average loss that minimized sudden jumps in values. These results are show in detail

in Section 5.2. In the end 0.8, 0.6 and 0.2 were chosen for the tagging damage factors red yellow and green respectively. In order to calculate the building tagging damage index $BT_m^{(*)}$ for each update step m, a weighted average of the building tagging designations was calculated as shows in equation 3.11. The main variables used were bld_{red}, bld_{yellow}, bld_{green}, which corresponded to the total number buildings that were tagged as red, yellow and green respectively.

$$BT_m^{(*)} = \frac{0.8 * bld_{red} + 0.6 * bld_{yellow} + 0.2 * bld_{green}}{bld_{red} + bld_{yellow} + bld_{green}}; \tag{3.11}$$

The next step is to define an equivalent building tagging damage index for each simulated event. First, the CAT model analysis portfolio, $portj$, corresponding to the same building type is chosen: timber, concrete, steel or masonry. Next, the buildings in the model were identified that were located in the tagging city. Portfolio buildings $bld_1 : bld_{N_{bld}^{zip5km}}$ corresponded to the union of CAT model buildings that were in the zipcode of the tagged city and a 5km radius of the same city center. Once the model analysis and simulated building were identified, their loss, $L_m^{(n)}(bldi, portj)$, was added and divided by the asset value ($1,000,000) to calculate the simulated building tagging damage parameter.

$$BT_m^{(n)} = \frac{\sum_{i=1}^{N_{bld}^{zip5km}} L_m^{(n)}(bldi, portj)}{\$1,000,000 * N_{bld}^{zip5km}}; \tag{3.12}$$

3.3.3 Establish A Baseline Probability Density Function

In order to initiate the Bayesian update procedure, there must be a prior distribution: a baseline probability density function (PDF) for the random variable of interest. The random variable for this procedure is the occurrence of the catalog events. Before any prior knowledge of a real event, the relative occurrence of the events should be equal to a constant (C). The only additional constraint on the prior PDF is that its integral or the sum of its values must equal 1. The initial PDF of the catalog events is therefore equal to one divided by the

number of unique simulated events, N_{EQ}.

$$\int_{n=1}^{N_{EQ}} C dn = 1 \Rightarrow C = \frac{1}{\int_{n=1}^{N} dn} \xrightarrow{\text{discrete}} \frac{1}{N_{EQ}};\ P_1(A_n) = \frac{1}{N_{EQ}} \tag{3.13}$$

3.3.4 Define a Likelihood Function

The next step of the Bayesian update procedure is to define a likelihood function (LF) for each set of observed real earthquake parameters $X_m^{(*)}$. Once the likelihood function is defined, it must be evaluated at the discrete simulated parameter values $X_m^{(n)}$ of each unique simulated event (Figure 3.2). The likelihood function selection depends on the type of observed parameters being used for the update step. A normal distribution was chosen for all update cases, but its degrees of freedom were chosen to correspond to the number of observed parameters being used for the given update step. Although a normal distribution was chosen, any other distribution can be used to define the likelihood function. The first update step uses the scaled earthquake magnitude and distance, thus a bivariate normal distribution was chosen. All subsequent updates consist of either the peak ground acceleration (PGA) or a single damage index for a specific exposure and location, thus a single degree of freedom normal distribution was chosen.

The damage index update steps (2,3,...M) use a one dimensional normal distribution, which is characterized by a single mean(μ_m), and standard deviation (σ_m) . The PDF reaches its maximum value at its mean and decays at a rate set by the standard deviation [19]. The PDF must be centered at the real observed damage index or PGA, thus the mean was set to the DI for the given step: $\mu_m = X_m^{(*)}$. The standard deviation was varied depending on the range of the simulated parameters for each update step m, $maximum(X_m^{(n)}) - minimum(X_m^{(n)})$. A larger standard deviation results in a flatter distribution, similar likelihood values and thus

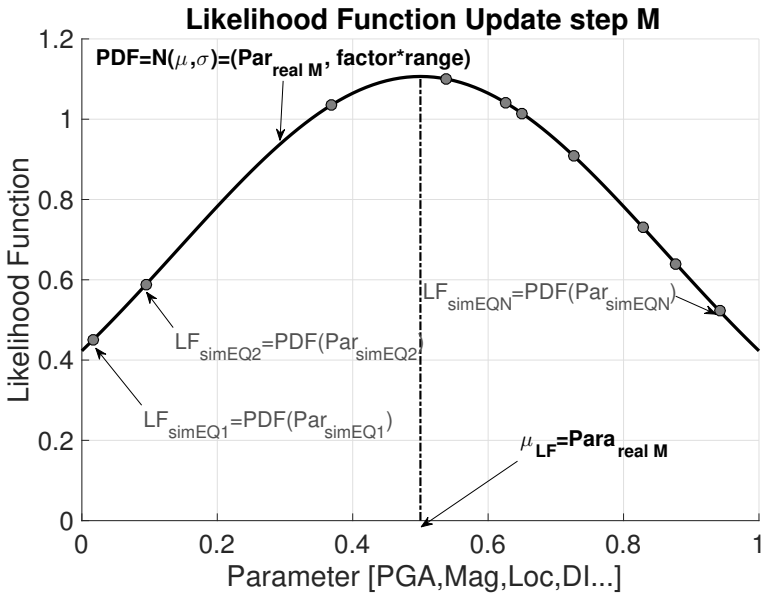

Figure 3.2: General Shape of Normal Likelihood Function

a smaller change in the final updated loss.

$$P_m(B_m|A_n) = f_{X_m^{(n)}}(y|\mu_m, \sigma_m^2) = \frac{1}{\sqrt{2\pi\sigma_m^2}} exp\left[-\frac{y-\mu_m}{2\sigma_m^2}\right] = \frac{1}{\sqrt{2\pi\sigma_m^2}} exp\left[-\frac{y-X_m^{(*)}}{2\sigma_m^2}\right] \tag{3.14}$$

After the LF is defined it must be evaluated at the simulated parameter values for the given update step.

$$f_{X_m^{(n)}}(X_m^{(n)}|\mu_m, \sigma_m^2) = \frac{1}{\sqrt{2\pi\sigma_m^2}} exp\left[-\frac{X_m^{(n)}-X_m^{(*)}}{2\sigma_m^2}\right] \tag{3.15}$$

The first update step uses two observed real parameters: scaled magnitude and scaled distance, thus a two degree of freedom normal distribution was used. Equations 3.16 and 3.17 define the properties of a bivariate Normal PDF. The input for the PDF are two random variables x_1 and x_2, which represent the scaled magnitude and location of the real earthquake respectively. The first model parameter is a vector $(\boldsymbol{\mu_{12}})$ containing the mean of random variable one and two, μ_{1m} and μ_{2m} respectively. The second and last model parameter is a covariance matrix containing the individual standard deviation of x_1 and x_2 separately σ_{1m}, σ_{2m} and the cross correlation between the two random variables, $\sigma_{1m}\sigma_{2m}$ [20].

$$P_m(B_m|A_n) = f_{\mathbf{X_m^{(n)}}}(\mathbf{y}|\mu_\mathbf{m}, \mathbf{V_m}); \quad \mathbf{y} = [x_1, x_2]$$
$$f_{\mathbf{X_m^{(n)}}}(\mathbf{y}|\mu_\mathbf{m}, \mathbf{V_m}) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_m^2}} exp\left[-\frac{z}{2(1-\rho_m^2)}\right] \tag{3.16}$$

$$z = \frac{(x_1-\mu_{1m}^2)^2}{\sigma_{1m}^2} - \frac{2\rho_m(x_1-\sigma_{1m})(x_2-\sigma_{2m})}{\sigma_{1m}\sigma_{2m}} + \frac{(x_2-\mu_{2m})^2}{\sigma_{2m}^2} \tag{3.17}$$

$$\mathbf{V_{12m}} = \begin{bmatrix} \sigma_{1m}^2 & \rho\sigma_{1m}\sigma_{2m} \\ \rho\sigma_{1m}\sigma_{2m} & \sigma_{2m}^2 \end{bmatrix}; \ \mu_\mathbf{12m} = [\mu_{1m}; \mu_{2m}] \tag{3.18}$$

To define the LF $\mu_\mathbf{12m}$ and $\mathbf{V_{12m}}$ must be properly defined according to the real earthquake observed parameters. The mean vector is set equal to the first parameter values of the real time earthquake: $\mu_\mathbf{12m} = [\mu_{1m}; \mu_{2m}] = X_1^{(*)} = [0.76, 0]$ for the large Chilean earthquake, $X_1^{(*)} = [0.42, 0]$ for the small Chilean earthquake and $X_1^{(*)} = [0.49, 0]$ for the California earthquake. The covariance is defined such that the desired level of uncertainty is introduced and the number of variables is reduced to one. The off diagonal terms of the covariance matrix, $\rho\sigma_{1m}\sigma_{2m}$, are set equal to zero and the first and second standard deviations ($\sigma_1 \, and \sigma_1$) are set equal to each other; thus only one standard deviation value is defined for each update.

$$\mathbf{V_{12m}} = \begin{bmatrix} \sigma_m^2 & 0 \\ 0 & \sigma_m^2 \end{bmatrix}; \quad Chile : \mu_{12m} = [0.76; 0] \ or \ [0.42; 0] \quad CA : \mu_{12m} = [0.49; 0] \tag{3.19}$$

$$Chile : z = \frac{(x_1-0.76)^2}{\sigma_m^2} - 0 + \frac{(x_2)^2}{\sigma_m^2}; \quad z = \frac{(x_1-0.42)^2}{\sigma_m^2} - 0 + \frac{(x_2)^2}{\sigma_m^2}$$
$$CA : z = \frac{(x_1-0.49)^2}{\sigma_m^2} - 0 + \frac{(x_2)^2}{\sigma_m^2} \tag{3.20}$$

After it is defined, the continuous LF is evaluated for each simulated earthquake parameter value. For each update step these discrete LF values are used in the Bayesian update

procedure to calculate the posterior probabilities.

$$\mathbf{y} = X_1^{(n)} = [M_1^{(n)}, D_1^{(n)}]$$
$$f_{X_m}(X_1^{(n)}|\mu_{\mathbf{m}}, \mathbf{V_m}) = \frac{1}{2\pi\sigma_m^2} \exp\left[-\frac{1}{2}\left(\frac{(M_1^{(n)} - \mu_{1m})^2}{\sigma_m^2} + \frac{D_1^{(n)}}{\sigma_m^2}\right)\right] \tag{3.21}$$

Defining the standard deviation of the likelihood function is an important user defined input. Different approaches were attempted: a constant standard deviation for every update step and an adaptive standard deviation for each step. One of the goals of the methodology is to reduce the loss updates sudden increase or decrease in value for different update steps. For the case where the standard deviation was constant for all cases there was a large variability in the loss output, thus the second approach was selected. In order to define the adaptive standard deviation of the likelihood function, the range of the simulated earthquake parameters was calculated for each update step. For example if the nonzero maximum and minimum of the DI for update step m was 0.2 and 0.5 then the range would be calculated as 0.5-0.2=0.3. Next 0.25, 0.5, 0.75, 1 and 1.25 multiplied this range in order to come up with four different standard deviation values=[0.075,0.15,0.225,0.3,0.375]. This procedure was an attempt to match the standard deviation of the likelihood function with that of the simulated data.

3.3.5 Calculate Posterior PDF

Once the likelihood function is defined the posterior can be calculated by simple multiplication. The un-normalized posterior for update step m is calculated from the product of the prior PDF times the corresponding likelihood value for the same simulated event.

$$\widetilde{P}_m(A_n|B_m) = P_m(A_n) * P_m(B_m|A_n) \tag{3.22}$$

The final step is to enforce the properties of a PDF and ensure its sum over all possible events equals 1. To ensure this property, the un-normalized posterior is divided by the sum

of its values for all simulated EQ events in the catalog. Equation 3.23 shows the final update components and the recursive nature of the problem. The problem is initiated with a constant prior $P_0(A_n) = \frac{1}{N_{EQ}}$ and after, for any given update step m, the prior becomes the posterior of the previous update step $m-1$.

$$P_m(A_n|B_m) = \frac{\widetilde{P}_m(A_n|B_m)}{\sum_{n=1}^{N_{EQ}} P_m(A_n)P_m(B_m|A_n)} = \frac{P_m(A_n) * P_m(B_m|A_n)}{\sum_{n=1}^{N_{EQ}} P_m(A_n)P_m(B_m|A_n)} = \frac{P_{m-1}(A_n|B_{m-1})P_m(B_m|A_n)}{\sum_{n=1}^{N_{EQ}} P_m(A_n)P_m(B_m|A_n)} \tag{3.23}$$

3.3.6 Calculate Updated Loss

The final step is to use a real portfolio to link the posterior probabilities of having a simulated event occur with a monetary loss. The goal is to have a real portfolio loss for each simulated event $n = 1 : N_{EQ}$ for all portfolio buildings. The model outputs a loss$(L^{(n)}(i))$ for each simulated event at each separate building locations $i = 1 : N_{bld}$. In order to have one loss for each event $(L^{(n)})$, the loss for each simulated earthquake is added up over all locations $i = 1 : N_{bld}$.

$$L^{(n)} = \sum_{i=1}^{N_{bld}} L^{(n)}(i) \tag{3.24}$$

Then after each update step, a new posterior PDF was constructed by plotting the posterior occurrence probability for each simulated event vs. the total loss for the same simulated event. These posterior loss PDF's were used to calculate a loss estimate for each update step.

The first step to calculate the loss confidence intervals was to convert the PDF into a cumulative distribution function (CDF). In the continuous case, the CDF is the integral of a PDF and for the discrete case, it's the sum. The CDF at a fixed loss value L_i is the sum of all the discrete PDF values for a loss equal to or less than L_i.

$$CDF(x) = \int_{-\inf}^{x} PDF(t)dt \xrightarrow{\text{discrete}} CDF_m(L_n) = \sum_{i=1}^{n} PDF_{upd,norm}(L_i) \tag{3.25}$$

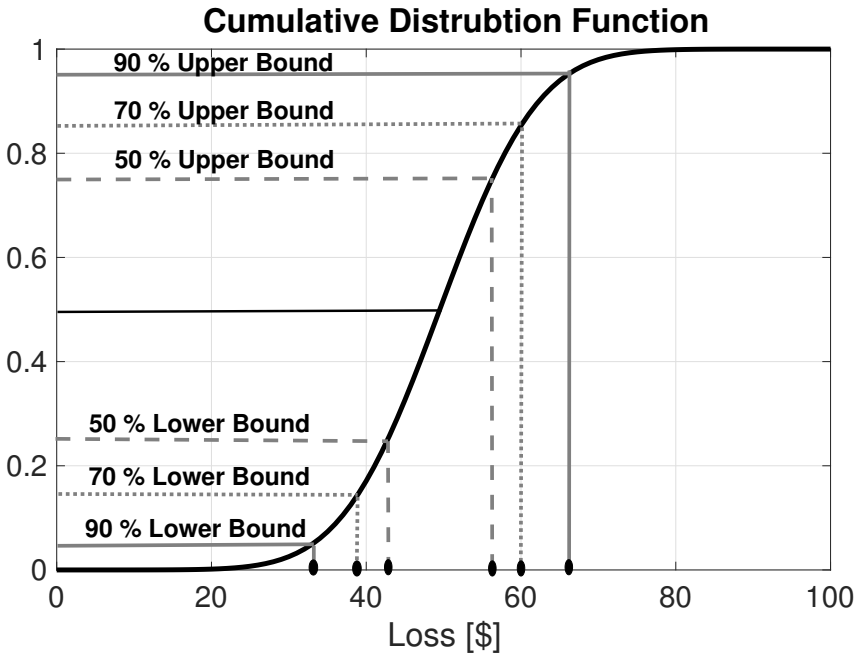

Figure 3.3: Loss Confidence Intervals

Next the goal is to find the 50, 70 and 90 % confidence intervals centered at the mean of the distribution. The probability of being between a loss L_a and L_b is equal to the difference of the CDF evaluated at L_b and L_a. In order to find the bounds for a d% confidence interval centered at the mean, the loss values that results in a CDF value of 50%-d/2 and 50%+d/2 must be calculated.

$$P(L_a < L < L_b) = CDF(L_b) - CDF(L_a) = (50\% + d/2) - (50\% - d/2) = d\% \Rightarrow$$
$$d\% \quad ConfidenceInterval : [L_{50\%-d/2}, L_{50\%+d/2}]$$
$$90\%\ CI : [L_{50\%-90\%}2, L_{50\%+(90\%)/2}] = [L_{5\%}, L_{95\%}]$$
$$70\%\ CI : [L_{50\%-(90\%)/2}, L_{50\%+(90\%)/2}] = [L_{15\%}, L_{85\%}]$$
$$50\%\ CI : [L_{50\%-(90\%)/2}, L_{50\%+(90\%)/2}] = [L_{25\%}, L_{75\%}] \tag{3.26}$$

Another measure of the updated loss is a weighted average loss calculation, following utility theory mentioned in section 2.1. After every update step, a relative occurrence rate is

calculated for each simulated catalog event. In order to consider the loss contribution of every catalog event, the posterior PDF value of each simulated earthquake n at update step m ($P_m(A_n|B_m)$) is multiplied by its corresponding loss over all locations($L^{(n)}$) (eq: 3.27). Then to calculate the total portfolio loss for the real event at each update step m (L_m), each of these weighted loss values is added for every simulated event. This calculation creates natural bounds for the computed updated real earthquake loss at each update step m because it must be larger that the smallest simulated EQ loss ($min(L^{(n)})$) and cannot exceed the largest simulated EQ loss ($max(L^{(n)})$).

$$\begin{aligned} L_m &= \sum_{n=1}^{N_{EQ}} P_m(A_n|B_m) * L^{(n)} \\ min(L^{(n)}) &< L_m < max(L^{(n)}) \end{aligned} \tag{3.27}$$

Chapter 4

Numerical Results

4.1 Case Studies

4.1.1 Maule Chile 2010 Earthquake

February 27, 2010 an 8.8 magnitude earthquake struck off the coast of Chile resulting in extensive damage throughout the country. This large event was followed by numerous aftershocks and a large tsunami that leveled many towns. Chile is prone to some of the largest recorded earthquakes because it is situated next to an active subduction zone. The Nazca and South American tectonic plates move an average 2.8 inches every year, which result in large events every few years [21]. There were extensive damage reports and surveys conducted for the 2010 Chile earthquake making it an ideal case study. In order to implement the Bayesian update procedure, region specific damage information were collected from field data reports and journal publications.

4.1.2 Punitaqui Chile 1997 Earthquake

In order to test the robustness of the Bayesian update procedure it was applied to a smaller event. Chile's proximity to the ring of fire and active fault lines results in an earthquake

nearly every day, which do not always have reported damage. In order to pick an earthquake with sufficient historical data a minimum magnitude threshold of 7 was set and a time frame between 1990 to the 2016. The October 14^{th}, 1997 7.1 magnitude Punitaqui earthquake was chosen because it had sufficient data to compare with the considerably larger 8.8 Chile 2010 event.

4.1.3 Northridge 1994 Earthquake

The January 17^{th}, 1994 6.7 magnitude Northridge earthquake that resulted from a blind thrust fault was the most costly events in U.S. history resulting in over $20 billions in loss [22]. Since it occurred, many studies have been performed and data documented making it an excellent candidate for a case study. One of the major studies by the Geographic Information System Group of the Governor's Office of Emergency Services in Pasadena resulted in two major reports summarizing damage, building inventory [8] and analysis and trend of the initial data [23].

4.2 Real Earthquake Parameters

Three major types of information were used for the update: attributes of the earthquake, peak ground acceleration readings, and building damage reports. The sources varied between the type of information and case study earthquakes. The primary source for a majority of the data was the United States geological service (USGS). The next major source was damage survey results from civil engineering experts that went to Chile to asses and record the damage to different types of buildings in different locations. One unique data set to Northridge included preliminary building tagging in the LA county following the earthquake [8]. The last sources were academic publications of studies on the different earthquakes.

4.2.1 Magnitude and Location

For all three events event the magnitude and location attributes were taken from those reported by USGS [1] [2] [3]. This information is the most readily available and can be accessed hours or a day after the event. Initial reports of the location and magnitude may vary slightly the days following the earthquake, but they can all be incorporated as a separate updates step.

4.2.2 Peak Ground Acceleration

For the 2010 Chile and 1994 Northridge earthquake the reported peak ground accelerations (PGA) were also reported by USGS [4] [3]. Both the location and PGA values were recorded for 58 different stations in Chile. The smaller 1997 event did not have PGA readings available on USGS thus similar information was taken from a journal publication [5]. Since the 1997 event had a smaller magnitude, it only had six available PGA station recordings.

4.2.3 Reported Damage

The 2010 Chile earthquake damage reports were taken from model vendor survey results. After every major hazard CAT model vendors inform insurance companies the details of the event and potential damage to their insured assets. Following the 2010 Chile earthquake the model vendors sent teams to conduct surveys and compile damage information for various regions in the country. The available reports had extensive data on several cities in Chile. Many factors affected the extent of earthquake-induced damage. As expected, cities located closer to the epicenter experienced the most extensive damage. Another factor that influenced building performance was age, with newer buildings sustaining little or no damage. The main reason for this trend is that Chile's newer building codes are more stringent because it is such a seismically active region. With every major event there are revisions and improvements to the Chile Building Code.

Besides proximity to the earthquake and construction year the building class is another major factor in a buildings ultimate performance. In many remote areas of Chile there was a large subset of older adobe residential or commercial buildings. These buildings tended to perform the worst with a majority suffering complete collapse. In addition, soil conditions were a major influence in the final damage. The building attributes reported in the surveys were used to define a set of nominal model portfolios to output average damage per location per specific exposure.

The 1997 Chile earthquake damage reports were taken from technical journal publications because model vendor reports reports were unavailable. The level of detail varied by location and publication but always included a qualitative or quantitative damage description, some construction information and location of reported damage. The broadest descriptions included the region the building found, construction material and some qualitative description [6] [7]. The most detailed type of information was found in a journal in Tectonophysics and included coordinates of location and number of buildings in location falling within a damage scale of one to five [5].

4.2.4 Building Tagging

EQE and the Geographic Information Systems Group of the Governor's Office of Emergency Services compiled a two part report summarizing data and analysis of effect of the Northridge EQ in LA country area [8] [23]. For each city the number of buildings that that were designated red, yellow, and green were reported for masonry, steel, timber and concrete buildings. In addition a building inventory was reported that showed more detailed information about the number of buildings constructed in different years, with different numbers of stories and occupancies [8].

4.2.5 Summary of Collected Real Parameters

4.2.5.1 Chronology of Reported Loss

In order to use the collected data for both the Chilean and Californian earthquakes it was quantified and organized in chronological order. For each earthquake all available information was compiled into separate tables for magnitude, location (Tables 4.1 & 4.2 & 4.3), peak ground acceleration (Tables 4.4 & 4.5 & 4.6 & 4.7 & 4.8), reported damage (Tables 4.9 & 4.10) and building tagging (Table 4.11). Every table includes a column for the date the information was reported, the raw data, and the corresponding model input values. The type of information was similar between earthquakes but varied the most for reported damage. Damage reports for the 2010 Chile earthquake included detailed information about specific building construction material, date, and qualitative descriptions of damage. Damage for the 1997 Chile event included more detailed information for the number of buildings in a region that fell within a quantitative damage scale of 0 to 5. In addition, only the 1994 Northridge Earthquake used building tagging information as an update. The process used to convert the collected data to a model parameter input was discussed in the previous chapter.

Table 4.1: Chile 2010 Magnitude and Location [1]

Update #	**Date**	**Source**	**Damage**	**Index**	**Value**
1	2/27/10	USGS	Earthquake Occurs	n/a	n/a
2	2/28/10	USGS	8.8 Magnitude; 35.909°S, 72.733°W	$X_1^{(*)}$	[0.76,0]

Table 4.2: Chile 1997 Magnitude and Location [2]

Update #	**Date**	**Source**	**Damage**	**Index**	**Value**
0	10/14/97	USGS	Earthquake Occurs	n/a	n/a
1	10/15/97	USGS	7.1 Magnitude; 30.93°S, 71.22°W	$X_1^{(*)}$	[0.42,0]

Table 4.3: Northridge 1994 Magnitude and Location [3]

Update #	**Date**	**Source**	**Damage**	**Index**	**Value**
0	01/17/94	USGS	Earthquake Occurs	n/a	n/a
1	10/18/94	USGS	6.7 Magnitude; 34.213°S, 118.537°W	$X_1^{(*)}$	[0.49,0]

Table 4.4: Chile 2010 PGA [4]

	Accelerometer Data			**Model Input**	
Update #	**Lat**	**Long**	**pga [g]**	**Index**	**Value**
2	32.79	71.19	0.24	$X_2^{(*)}$	0.24
3	35.96	72.32	0.32	$X_3^{(*)}$	0.32
4	37.81	73.40	0.32	$X_4^{(*)}$	0.32
5	36.90	73.04	0.32	$X_5^{(*)}$	0.32
6	36.60	72.11	0.29	$X_6^{(*)}$	0.29
7	33.70	70.77	0.22	$X_7^{(*)}$	0.22
8	37.79	72.71	0.24	$X_8^{(*)}$	0.24
9	36.83	73.05	0.30	$X_9^{(*)}$	0.30
10	35.33	72.42	0.39	$X_{10}^{(*)}$	0.39
11	37.03	73.16	0.32	$X_{11}^{(*)}$	0.32
12	34.98	71.24	0.24	$X_{12}^{(*)}$	0.24
13	34.07	70.72	0.25	$X_{13}^{(*)}$	0.25
14	32.46	71.24	0.24	$X_{14}^{(*)}$	0.24
15	37.28	72.71	0.24	$X_{15}^{(*)}$	0.24
16	33.28	70.90	0.16	$X_{16}^{(*)}$	0.16
17	37.61	73.65	0.37	$X_{17}^{(*)}$	0.37
18	33.02	71.27	0.22	$X_{18}^{(*)}$	0.22
19	35.84	71.59	0.29	$X_{19}^{(*)}$	0.29
20	32.83	70.60	0.20	$X_{20}^{(*)}$	0.20
21	37.46	72.36	0.29	$X_{21}^{(*)}$	0.29
22	34.18	70.67	0.22	$X_{22}^{(*)}$	0.22
23	33.69	71.22	0.30	$X_{23}^{(*)}$	0.30
24	35.11	71.28	0.30	$X_{24}^{(*)}$	0.30
25	37.50	72.68	0.35	$X_{25}^{(*)}$	0.35
26	40.57	73.16	0.10	$X_{26}^{(*)}$	0.10
27	33.82	70.75	0.29	$X_{27}^{(*)}$	0.29
28	36.14	71.83	0.32	$X_{28}^{(*)}$	0.32
29	36.73	73.00	0.35	$X_{29}^{(*)}$	0.35

Table 4.5: Chile 2010 PGA(continued) [4]

	Accelerometer Data			Model Input	
Update #	**Lat**	**Long**	**pga [g]**	**Index**	**Value**
30	33.62	70.92	0.29	$X_{30}^{(*)}$	0.29
31	33.61	70.57	0.22	$X_{31}^{(*)}$	0.22
32	32.88	71.26	0.20	$X_{32}^{(*)}$	0.20
33	33.04	71.50	0.21	$X_{33}^{(*)}$	0.21
34	34.17	70.74	0.25	$X_{34}^{(*)}$	0.25
35	34.17	70.74	0.29	$X_{35}^{(*)}$	0.29
36	34.41	70.87	0.24	$X_{36}^{(*)}$	0.24
37	33.60	71.61	0.22	$X_{37}^{(*)}$	0.22
38	33.60	70.70	0.18	$X_{38}^{(*)}$	0.18
39	35.53	71.49	0.24	$X_{39}^{(*)}$	0.24
40	34.58	70.99	0.24	$X_{40}^{(*)}$	0.24
41	34.58	70.99	0.30	$X_{41}^{(*)}$	0.30
42	35.59	71.74	0.20	$X_{42}^{(*)}$	0.20
43	34.43	71.08	0.35	$X_{43}^{(*)}$	0.35
44	34.43	71.08	0.42	$X_{44}^{(*)}$	0.42
45	34.63	71.37	0.29	$X_{45}^{(*)}$	0.29
46	33.46	70.64	0.22	$X_{46}^{(*)}$	0.22
47	33.66	70.93	0.27	$X_{47}^{(*)}$	0.27
48	33.73	70.75	0.24	$X_{48}^{(*)}$	0.24
49	35.42	71.66	0.30	$X_{49}^{(*)}$	0.30
50	36.74	73.13	0.30	$X_{50}^{(*)}$	0.30
51	38.73	72.58	0.20	$X_{51}^{(*)}$	0.20
52	34.86	71.16	0.20	$X_{52}^{(*)}$	0.20
53	36.61	72.96	0.42	$X_{53}^{(*)}$	0.42
54	39.82	73.23	0.13	$X_{54}^{(*)}$	0.13
55	33.04	71.64	0.21	$X_{55}^{(*)}$	0.21
56	33.04	71.36	0.22	$X_{56}^{(*)}$	0.22
57	36.74	72.30	0.17	$X_{57}^{(*)}$	0.17
58	39.28	72.23	0.21	$X_{58}^{(*)}$	0.21
59	33.01	71.55	0.22	$X_{59}^{(*)}$	0.22

Table 4.6: Chile 1997 PGA [5]

		Station Coordinates		PGA value [g]			Model Input	
Update #	**Date**	**Latitude**	**Longitude**	**Horizontal 1**	**Horizontal 2**	**Vertical**	**Index**	**Value**
2	11/1/00	31.38	71.1	0.27	0.35	0.18	$X_{2}^{(*)}$	0.35
3	11/1/00	32.31	71.27	0.09	0.14	0.04	$X_{3}^{(*)}$	0.14
4	11/1/00	32.34	71.28	0.05	0.06	0.04	$X_{4}^{(*)}$	0.06
5	11/1/00	33.27	70.4	0.02	0.02	0.01	$X_{5}^{(*)}$	0.02
6	11/1/00	33.28	70.39	0.02	0.02	0.01	$X_{6}^{(*)}$	0.02
7	11/1/00	33.26	70.37	0.02	0.02	0	$X_{7}^{(*)}$	0.02

Table 4.7: Northridge 1994 PGA [3]

	Accelerometer Data			Model Input	
Update #	**Lat**	**Long**	**pga [g]**	**Index**	**Value**
2	33.817	-117.95	0.08	$X_2^{(*)}$	0.08
3	34.758	-118.36	0.07	$X_3^{(*)}$	0.07
4	33.847	-118.02	0.16	$X_4^{(*)}$	0.16
5	34.204	-118.3	0.17	$X_5^{(*)}$	0.17
6	33.965	-118.16	0.11	$X_6^{(*)}$	0.11
7	34.009	-118.36	0.24	$X_7^{(*)}$	0.24
8	34.1	-117.97	0.14	$X_8^{(*)}$	0.14
9	33.836	-118.24	0.09	$X_9^{(*)}$	0.09
10	33.914	-117.84	0.19	$X_{10}^{(*)}$	0.19
11	33.899	-118.2	0.13	$X_{11}^{(*)}$	0.13
12	34.212	-118.61	0.41	$X_{12}^{(*)}$	0.41
13	34.15	-117.94	0.08	$X_{13}^{(*)}$	0.08
14	34.26	-118.34	0.22	$X_{14}^{(*)}$	0.22
15	34.662	-118.39	0.15	$X_{15}^{(*)}$	0.15
16	34.093	-118.02	0.17	$X_{16}^{(*)}$	0.17
17	34.15	-118.51	0.20	$X_{17}^{(*)}$	0.20
18	33.79	-118.01	0.11	$X_{18}^{(*)}$	0.11
19	34.2	-118.23	0.40	$X_{19}^{(*)}$	0.40
20	34.05	-118.11	0.18	$X_{20}^{(*)}$	0.18
21	34.088	-118.36	0.26	$X_{21}^{(*)}$	0.26
22	34.09	-118.34	0.39	$X_{22}^{(*)}$	0.39
23	33.697	-118.02	0.12	$X_{23}^{(*)}$	0.12
24	33.905	-118.28	0.10	$X_{24}^{(*)}$	0.10
25	33.656	-117.86	0.06	$X_{25}^{(*)}$	0.06
26	34.309	-118.5	0.84	$X_{26}^{(*)}$	0.84
27	34.608	-118.56	0.23	$X_{27}^{(*)}$	0.23
28	34.65	-118.48	0.08	$X_{28}^{(*)}$	0.08
29	34.3	-118.48	0.36	$X_{29}^{(*)}$	0.36
30	33.84	-118.19	0.07	$X_{30}^{(*)}$	0.07
31	33.996	-118.16	0.27	$X_{31}^{(*)}$	0.27
32	34.238	-118.25	0.23	$X_{32}^{(*)}$	0.23
33	33.946	-118.39	0.18	$X_{33}^{(*)}$	0.18
34	34.053	-118.17	0.32	$X_{34}^{(*)}$	0.32
35	34.115	-118.24	0.26	$X_{35}^{(*)}$	0.26
36	34.111	-118.19	0.17	$X_{36}^{(*)}$	0.17
37	34.042	-118.55	0.49	$X_{37}^{(*)}$	0.49
38	33.896	-118.38	0.19	$X_{38}^{(*)}$	0.19
39	34.739	-118.21	0.08	$X_{39}^{(*)}$	0.08
40	34.04	-118.44	0.52	$X_{40}^{(*)}$	0.52
41	34.046	-118.36	0.49	$X_{41}^{(*)}$	0.49
42	34.045	-118.3	0.18	$X_{42}^{(*)}$	0.18
43	34.059	-118.25	0.18	$X_{43}^{(*)}$	0.18
44	34.062	-118.2	0.49	$X_{44}^{(*)}$	0.49

Table 4.8: Northridge 1994 PGA Cont. [3]

	Accelerometer Data			Model Input	
Update #	**Lat**	**Long**	**pga [g]**	**Index**	**Value**
45	34.594	-118.24	0.09	$X_{45}^{(*)}$	0.09
46	34.595	-118.24	0.09	$X_{46}^{(*)}$	0.09
47	34.596	-118.24	0.11	$X_{47}^{(*)}$	0.11
48	34.598	-118.24	0.08	$X_{48}^{(*)}$	0.08
49	34.6	-118.24	0.15	$X_{49}^{(*)}$	0.15
50	34.604	-118.24	0.18	$X_{50}^{(*)}$	0.18
51	34.005	-118.28	0.29	$X_{51}^{(*)}$	0.29
52	33.846	-118.1	0.14	$X_{52}^{(*)}$	0.14
53	33.886	-118.39	0.16	$X_{53}^{(*)}$	0.16
54	34.001	-118.43	0.47	$X_{54}^{(*)}$	0.47
55	33.99	-118.11	0.19	$X_{55}^{(*)}$	0.19
56	33.634	-117.9	0.06	$X_{56}^{(*)}$	0.06
57	34.848	-118.54	0.07	$X_{57}^{(*)}$	0.07
58	34.194	-118.41	0.32	$X_{58}^{(*)}$	0.32
59	34.564	-118.64	0.57	$X_{59}^{(*)}$	0.57
60	34.148	-118.17	0.10	$X_{60}^{(*)}$	0.10
61	34.334	-118.4	0.43	$X_{61}^{(*)}$	0.43
62	34.145	-119.21	0.10	$X_{62}^{(*)}$	0.10
63	34.46	-118.75	0.27	$X_{63}^{(*)}$	0.27
64	33.89	-117.64	0.14	$X_{64}^{(*)}$	0.14
65	33.746	-118.4	0.07	$X_{65}^{(*)}$	0.07
66	33.787	-118.36	0.12	$X_{66}^{(*)}$	0.12
67	34.104	-117.57	0.07	$X_{67}^{(*)}$	0.07
68	33.951	-117.45	0.06	$X_{68}^{(*)}$	0.06
69	34.442	-119.71	0.04	$X_{69}^{(*)}$	0.04
70	34.065	-117.29	0.10	$X_{70}^{(*)}$	0.10
71	34.097	-118.47	0.46	$X_{71}^{(*)}$	0.46
72	34.106	-118.45	0.38	$X_{72}^{(*)}$	0.38
73	34.31	-118.49	0.80	$X_{73}^{(*)}$	0.80
74	34.106	-118.45	0.38	$X_{74}^{(*)}$	0.38
75	33.944	-118.09	0.14	$X_{75}^{(*)}$	0.14
76	34.236	-118.44	0.34	$X_{76}^{(*)}$	0.34
77	34.091	-118.09	0.27	$X_{77}^{(*)}$	0.27
78	34.011	-118.49	0.88	$X_{78}^{(*)}$	0.88
79	34.743	-118.72	0.10	$X_{79}^{(*)}$	0.10
80	34.084	-118.6	0.33	$X_{80}^{(*)}$	0.33
81	34.004	-118.23	0.15	$X_{81}^{(*)}$	0.15
82	34.063	-118.46	0.19	$X_{82}^{(*)}$	0.19
83	34.249	-118.47	0.94	$X_{83}^{(*)}$	0.94
84	34.05	-118.45	0.39	$X_{84}^{(*)}$	0.39
85	34.052	-118.26	0.13	$X_{85}^{(*)}$	0.13
86	34.314	-117.54	0.04	$X_{86}^{(*)}$	0.04
87	33.916	-117.9	0.11	$X_{87}^{(*)}$	0.11
88	33.99	-117.94	0.08	$X_{88}^{(*)}$	0.08
89	34.064	-117.95	0.07	$X_{89}^{(*)}$	0.07

Table 4.9: Chile 2010 Damage Reports

Update #	**Date**	**Building Type**	**Year Constructed**	**Region**	**Index**	**Value**
60	3/9/10	RC Shear wall	2002	Santiago	$X_{60}^{(*)}$	0.2
61	3/9/10	RC Shear wall	Unknown	Valparaiso	$X_{61}^{(*)}$	0.5
62	3/9/10	RC Shear wall	Unknown	Valparaiso	$X_{62}^{(*)}$	0.7
63	3/9/10	Confined Masonry	very old	Valparaiso	$X_{63}^{(*)}$	0.7
64	3/9/10	Confined Masonry	newer	Valparaiso	$X_{64}^{(*)}$	0.2
65	3/9/10	Confined Masonry	newer	O'Higgins	$X_{65}^{(*)}$	0.1
66	3/9/10	Unreinforced Masonry	Unknown	Valparaiso	$X_{66}^{(*)}$	0.7
67	3/9/10	Unreinforced Masonry	Unknown	Valparaiso	$X_{67}^{(*)}$	0.2
68	3/9/10	Adobe	old	O'Higgins	$X_{68}^{(*)}$	0.8
69	3/9/10	Adobe	Unknown	O'Higgins	$X_{69}^{(*)}$	0.8
70	3/9/10	RC	post 2000	Santiago	$X_{70}^{(*)}$	0.1
71	3/9/10	Steel	165 years old	O'Higgins	$X_{71}^{(*)}$	0.8
72	3/10/10	RC Shear wall	2003	Maule	$X_{72}^{(*)}$	0.6
73	3/10/10	Confined Masonry	Unknown	Maule	$X_{73}^{(*)}$	0.2
74	3/10/10	Unreinforced Masonry	Unknown	Maule	$X_{74}^{(*)}$	0.8
75	3/10/10	Adobe	Unknown	O'Higgins/Maule	$X_{75}^{(*)}$	0.8
76	3/10/10	Adobe	Unknown	Maule	$X_{76}^{(*)}$	0.8
77	3/10/10	Adobe	Unknown	Maule	$X_{77}^{(*)}$	0.8
78	3/10/10	RC	Unknown	Maule	$X_{78}^{(*)}$	0.2
79	3/10/10	Steel	1995	Maule	$X_{79}^{(*)}$	0.8
80	3/10/10	Steel	Unknown	Maule	$X_{80}^{(*)}$	0.7
81	3/12/10	RC	2008	Bio-Bio	$X_{81}^{(*)}$	0.7
82	3/12/10	RC	recently opened	Bio-Bio	$X_{82}^{(*)}$	0.9
83	3/12/10	Steel	Unknown	Bio-Bio	$X_{83}^{(*)}$	0.8
84	3/16/10	Confined Masonry	Unknown	Bio-Bio	$X_{84}^{(*)}$	0.1
85	3/16/10	Unreinforced Masonry	Unknown	Maule	$X_{85}^{(*)}$	0.6
86	3/16/10	Adobe	Unknown	Bio-Bio	$X_{86}^{(*)}$	0.8
87	3/16/10	Adobe	Unknown	Maule	$X_{87}^{(*)}$	0.5
88	3/19/10	Confined Masonry	Unknown	Bio-Bio	$X_{88}^{(*)}$	0.6
89	3/19/10	Confined Masonry	1960	Bio-Bio	$X_{89}^{(*)}$	0.7

Table 4.10: Chile 1997 Damage Reports [6] [7]

			Damage Description						Model Input	
Update #	**Date**	**Bld Type**	**Gr 0**	**Gr 1**	**Gr 2**	**Gr 3**	**Gr 4**	**Gr 5**	**Index**	**Value**
8	11/1/00	Adobe	42	38	10	0	0	0	$X_8^{(*)}$	0.076
9	11/1/00	Adobe	11	3	2	0	0	0	$X_9^{(*)}$	0.056
10	11/1/00	Adobe	32	28	8	1	0	0	$X_{10}^{(*)}$	0.083
11	11/1/00	Adobe	1	18	10	17	7	6	$X_{11}^{(*)}$	0.4
12	11/1/00	Adobe	33	15	5	1	0	0	$X_{12}^{(*)}$	0.065
13	11/1/00	Adobe	25	52	13	2	0	0	$X_{13}^{(*)}$	0.11
14	11/1/00	Adobe	20	16	9	2	0	0	$X_{14}^{(*)}$	0.113
15	11/1/00	Adobe	5	54	284	81	12	0	$X_{15}^{(*)}$	0.32
16	11/1/00	Adobe	15	27	18	12	10	2	$X_{16}^{(*)}$	0.273
17	11/1/00	Adobe	8	4	5	1	0	0	$X_{17}^{(*)}$	0.133
18	11/1/00	Adobe	2	9	10	12	2	1	$X_{18}^{(*)}$	0.339
19	11/1/00	Adobe	3	2	1	0	0	0	$X_{19}^{(*)}$	0.083
20	11/1/00	Adobe	3	8	4	3	0	2	$X_{20}^{(*)}$	0.265
21	11/1/00	Adobe	0	0	0	1	0	8	$X_{21}^{(*)}$	0.856
22	11/1/00	Adobe	7	10	4	3	0	0	$X_{22}^{(*)}$	0.154
23	11/1/00	Adobe	1	12	65	90	37	19	$X_{23}^{(*)}$	0.485
24	11/1/00	Adobe	0	0	18	0	0	2	$X_{24}^{(*)}$	0.36
25	11/1/00	Adobe	1	0	23	9	2	0	$X_{25}^{(*)}$	0.366
26	11/1/00	Adobe	16	20	10	7	0	1	$X_{26}^{(*)}$	0.174
27	11/1/00	Adobe	18	7	3	1	0	0	$X_{27}^{(*)}$	0.072
28	11/1/00	Adobe	10	15	5	0	0	0	$X_{28}^{(*)}$	0.1
29	11/1/00	Adobe	0	0	12	2	1	4	$X_{28}^{(*)}$	0.468
30	11/1/00	Adobe	22	44	32	0	2	0	$X_{29}^{(*)}$	0.154
31	11/1/00	Adobe	6	7	14	0	0	0	$X_{30}^{(*)}$	0.181
32	11/1/00	Adobe	20	4	3	0	0	0	$X_{31}^{(*)}$	0.048
33	11/1/00	Adobe	2	5	7	1	0	0	$X_{32}^{(*)}$	0.207
34	4/2/17	Adobe	16/35 Adobe Schools destroyed						$X_{34}^{(*)}$	0.58
35	4/2/17	Adobe	4/48 schools destroyed and rest diff levels of damage						$X_{35}^{(*)}$	0.54
36	11/9/17	RC	hospital with minor damages						$X_{36}^{(*)}$	0.15
37	11/9/17	Adobe	collapse of adobe structures						$X_{37}^{(*)}$	0.8
38	11/9/17	Adobe	many houses suffered major damage						$X_{38}^{(*)}$	0.6
39	11/9/17	Masonry	façade fell off						$X_{39}^{(*)}$	0.15

Table 4.11: Northridge 1994 Tagging [8]

	Building Tagging Data					Model Input	
Update #	**City**	**Bld Type**	**R**	**Y**	**G**	**Index**	**Value**
90	Agoura Hills	Wood	0	0	86	$X_{90}^{(*)}$	0.20
91	Alhambra	Wood	3	5	46	$X_{91}^{(*)}$	0.27
92	Alhambra	Masonry	0	2	12	$X_{92}^{(*)}$	0.26
93	Beverly Hills	Wood	18	68	475	$X_{93}^{(*)}$	0.27
94	Beverly Hills	Steel	0	0	9	$X_{94}^{(*)}$	0.20
95	Beverly Hills	Concrete	1	1	13	$X_{95}^{(*)}$	0.27
96	Beverly Hills	Masonry	5	1	42	$X_{96}^{(*)}$	0.27
97	Calabasas	Wood	0	1	19	$X_{97}^{(*)}$	0.22
98	Calabasas	Steel	1	212	469	$X_{98}^{(*)}$	0.33
99	Culver City	Wood	12	5	0	$X_{99}^{(*)}$	0.74
100	Glendale	Wood	8	3	773	$X_{100}^{(*)}$	0.21
101	Glendale	Masonry	9	3	75	$X_{101}^{(*)}$	0.28
102	Hiden Hills	Wood	0	43	30	$X_{102}^{(*)}$	0.44
103	Lakewood	Wood	0	0	7	$X_{103}^{(*)}$	0.20
104	Los Angeles	Wood	1411	6746	61273	$X_{104}^{(*)}$	0.25
105	Los Angeles	Steel	7	26	87	$X_{105}^{(*)}$	0.32
106	Los Angeles	Concrete	22	35	100	$X_{106}^{(*)}$	0.37
107	Los Angeles	Masonry	210	483	1773	$X_{107}^{(*)}$	0.33
108	Manhattan Beach	Wood	3	218	25	$X_{108}^{(*)}$	0.56
109	San Fernando	Wood	64	100	872	$X_{109}^{(*)}$	0.28
110	San Fernando	Masonry	22	13	59	$X_{110}^{(*)}$	0.40
111	Santa Clarita	Wood	56	118	2716	$X_{111}^{(*)}$	0.23
112	Santa Clarita	Masonry	6	8	60	$X_{112}^{(*)}$	0.29
113	Santa Monica	Wood	38	128	783	$X_{113}^{(*)}$	0.28
114	Santa Monica	Steel	2	3	6	$X_{114}^{(*)}$	0.42
115	Santa Monica	Concrete	2	2	4	$X_{115}^{(*)}$	0.45
116	Santa Monica	Masonry	17	32	46	$X_{116}^{(*)}$	0.44
117	South Gate	Wood	0	0	22	$X_{117}^{(*)}$	0.20
118	South Gate	Masonry	0	1	8	$X_{118}^{(*)}$	0.24
119	Vernon	Masonry	3	1	4	$X_{119}^{(*)}$	0.48
120	West Hollywood	Wood	0	2	5	$X_{120}^{(*)}$	0.31

4.3 Simulated Earthquake Parameters

4.3.1 CAT Model Analysis

Each update step requires two inputs: one from the model ($\mathbf{X^{(n)}}$) and one from real event observations ($\mathbf{X^{(*)}}$). These two inputs establish a set of parameters that can be compared between the simulated events and the real event. Ultimately this comparison defines the likelihood function used for the Bayesian update procedure. Two types of CAT model outputs were generated: real portfolio of buildings and nominal portfolio for both Chile and California. The nominal portfolio models are used to generate the simulated parameters such as model pga and earthquake loss. The nominal portfolios are used to implement Bayes theorem and establish the contribution each simulated event has toward the total loss. Yet the final goal isn't to rank the contribution of simulated events but rather to calculate a loss, which is why a real portfolio is necessary. Data for a real set of buildings is used to calculate an updated loss estimate based on the contribution of each simulated event.

The CAT model building locations for Chile and California shown in Figures 4.1 and 4.2. The California analysis included 2,694 buildings and Chile's consisted of 30,333 buildings (N_{loc}) both evenly spaced at 5km. The analysis portfolios were defined to include exposure parameters for all the survey damage reports or tagging reports. Exposure for a specific building is defined by building class, occupancy, year built and number of stories. In some cases one or more of these attributes were left unknown because the information was not available from the reports. Tables 4.12 and 4.13 show the nominal portfolio cases for Chile and California respectively.

The CAT model calculates hazard intensities at each location and converts that to a damage ratio using a damage function. The damage function is unique to each building class and occupancy. The year and number of stories also have a damage function for different bands instead of a specific value. The number of stories are separated into groups of 1-3, 4-7,

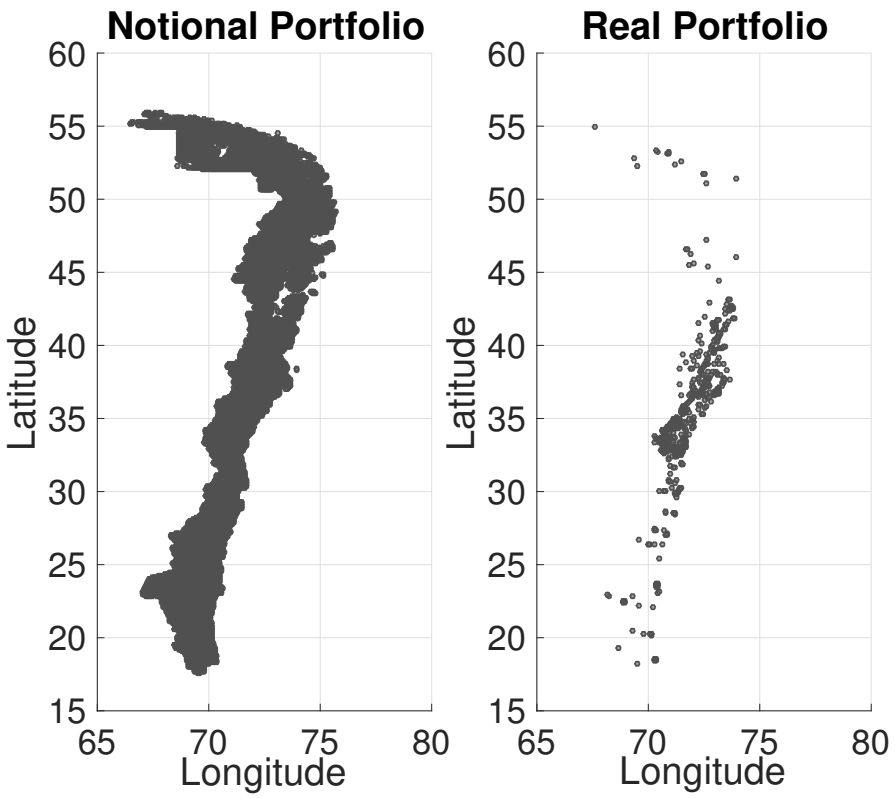

Figure 4.1: Chile CAT model building location

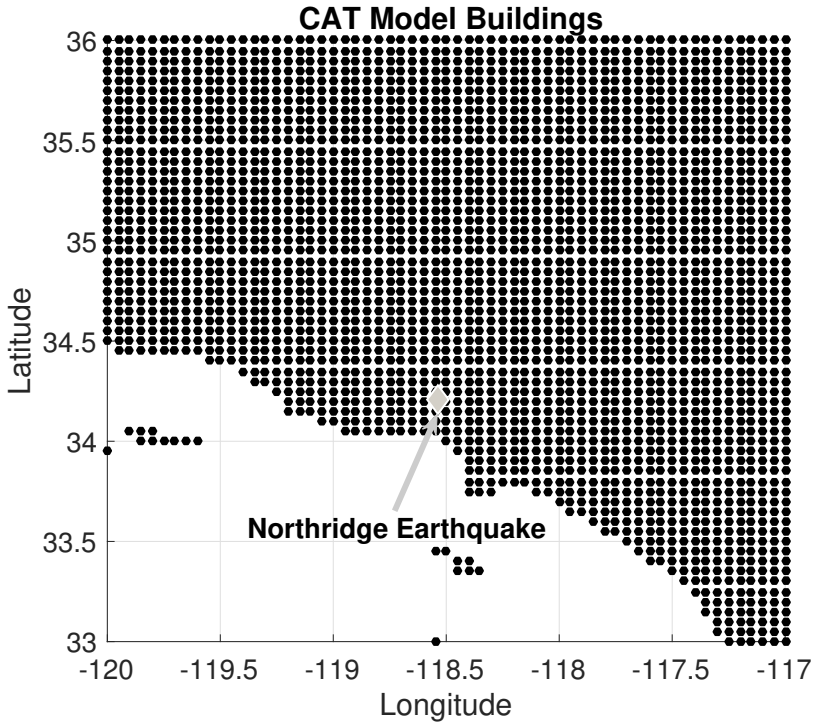

Figure 4.2: California CAT model building location

8-14 and >15 corresponding to low-rise, mid-rise, high-rise and tall respectively. Some real survey damage descriptions included similar descriptions such as low-rise and mid-rise building. These real survey height descriptions were converted to a story height from the corresponding height bands defined in the model. For example if a real damage report described a building as low rise it was defined to between 1 to 3 stories. Year built is treated in a similar fashion where different year bands result in a different damage function. These year bands correspond to major changes in the building code and thus different building resiliency. For cases of similar construction year they were grouped into average value for a given year band.

Table 4.12: Notional Building Parameters

Portfolio	**Building Class**	**Occupancy**	**Year Built**	**Stories**
1	RC Shear Wall	Business Services	2000	15
2	RC Shear Wall	Unknown	2000	15
3	Confined Masonry	Unknown	1960	2
4	Confined Masonry	Unknown	2000	2
5	Confined Masonry	Business Services	1960	2
6	Confined Masonry	Business Services	2000	2
7	Unreinforced Masonry	Commercial	unknown	3
8	Unreinforced Masonry	Unknown	1960	3
9	Adobe	Unknown	1939	3
10	Reinforced Concrete	Education	2000	Unknown
11	Reinforced Concrete	Commercial	2000	Unknown
12	Reinforced Concrete	Education	Unknown	Unknown
13	Reinforced Concrete	Commercial	Unknown	Unknown
14	Steel	Business Services	Unknown	Unknown
15	Steel	Business Services	Unknown	50
16	Steel	Food and Drug	Unknown	Unknown
17	Steel	Petroleum	Unknown	Unknown
18	Steel	Air	Unknown	Unknown

4.3.2 CAT Model Earthquake Catalog

During each CAT model analysis a building portfolio is subjected to hundreds of thousands of simulated events. Once the loss is calculated the model only outputs the loss due to events that produced a non-zero loss. The simulated events that produce a non-zero loss vary between portfolios because the building properties also vary. For example an adobe building would have a greater number of simulated events that produce a loss than a steel

Table 4.13: Notional Building Parameters

Portfolio	Building Class	Occupancy	Year Built	Stories
1	Masonry	unknown	1950	3
2	Masonry	unknown	1976	3
3	Wood frame	SFPD	1940	2
4	Wood frame	SFPD	1960	2
5	Wood frame	SFPD	1976	2
6	RC	Gen commercial	unknown	3
7	RC	Gen commercial	unknown	5
8	RC	Gen commercial	unknown	8
9	Steel	Gen commercial	unknown	3
10	Steel	Gen commercial	unknown	5
11	Steel	Gen commercial	unknown	8

building would. The simulated events that didn't produce a loss were omitted from the analysis. In order to apply the update methodology, for each distinct simulated event there must be a corresponding magnitude ($M_m^{(n)}$), location ($D_m^{(n)}$) and frequency.

4.3.3 Model Damage Output

The loss output lists a ground up monetary loss value per building location per simulated event. In order to define damage indices (DI) on a scale from zero to one the loss was divided by the building value of $1,000,000. This DI per event per location becomes the simulated event parameter ($DI_m^{(n)}$) that is compared to a user defined DI ($DI_m^{(*)}$) based on reported damage descriptions.

4.3.4 General Loss Trends

The maximum loss over all locations and all events was identified for each CAT model analysis case. For Chile in all analyses the location of maximum loss varied but the event that caused the loss was identical. This catalog event had a magnitude close to the maximum catalog value, thus it is reasonable that it would produce the maximum damage. For California the location and event of maximum Loss varied between two sets of values: one with a large magnitude and one with medium magnitude but small epicenter distance. The magnitude of the maximum damaging earthquakes for Chile and California were both close their respective

maximum catalog magnitude values. Tables 4.14 and 4.15 show the occupancy description and damage in ascending order. For each case the damage was calculated by dividing the maximum loss by the building value of $1,000,000.

As expected, the largest monetary loss was sustained by adobe for Chile and masonry for California. The smallest damage was found in steel buildings, which also tend to be the most ductile. Between cases of identical building class the older the building or more stories the greater the loss it suffered. The newer buildings are built with more recent versions of the building code, which apply more stringent design standards and thus tend to be more resilient. Figures 4.3 and 4.4 show the DI for each building class over all location for the

Table 4.14: Ascending Damage Chile

Building Class	**Occupancy**	**Year Built**	**Stories**
Steel	Food and Drug	Unknown	Unknown
Steel	Petroleum	Unknown	Unknown
Steel	Air	Unknown	Unknown
Confined Masonry	Business Services	2000	2
Confined Masonry	Unknown	2000	2
Confined Masonry	Business Services	1960	2
Confined Masonry	Unknown	1960	2
Steel	Business Services	Unknown	50
Reinforced Concrete	Education	2000	Unknown
Reinforced Concrete	Commercial	2000	Unknown
RC Shear Wall	Business Services	2000	15
RC Shear Wall	Unknown	2000	15
Reinforced Concrete	Education	Unknown	Unknown
Reinforced Concrete	Commercial	Unknown	Unknown
Steel	Business Services	Unknown	Unknown
Unreinforced Masonry	Commercial	unknown	3
Unreinforced Masonry	Unknown	1960	3
Adobe	Unknown	1939	3

event that produced the maximum loss. For all cases there is a large spread of possible DI values, which is expected because distance between the building and earthquake also vary greatly. For every case there are also a large number of events that produce zero loss and thus zero DI. In general the model loss for each case correlates with what is expected for its given exposure properties.

Table 4.15: Ascending Damage California

Building Class	Occupancy	Year Built	Stories
Wood frame	SFPD	1976	2
Wood frame	SFPD	1960	2
RC	Gen commercial	unknown	5
Steel	Gen commercial	unknown	5
Wood frame	SFPD	1940	2
Steel	Gen commercial	unknown	8
RC	Gen commercial	unknown	8
Steel	Gen commercial	unknown	3
RC	Gen commercial	unknown	3
Masonry	unknown	1976	3
Masonry	unknown	1950	3

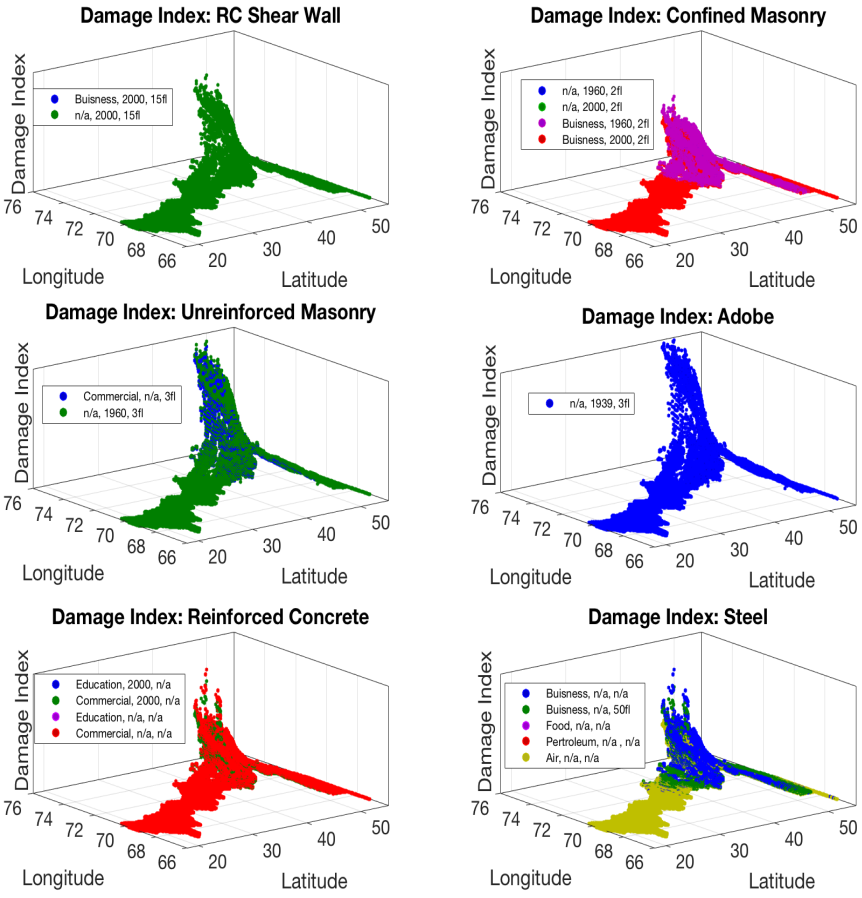

Figure 4.3: Maximum DI Per Building Class Chile

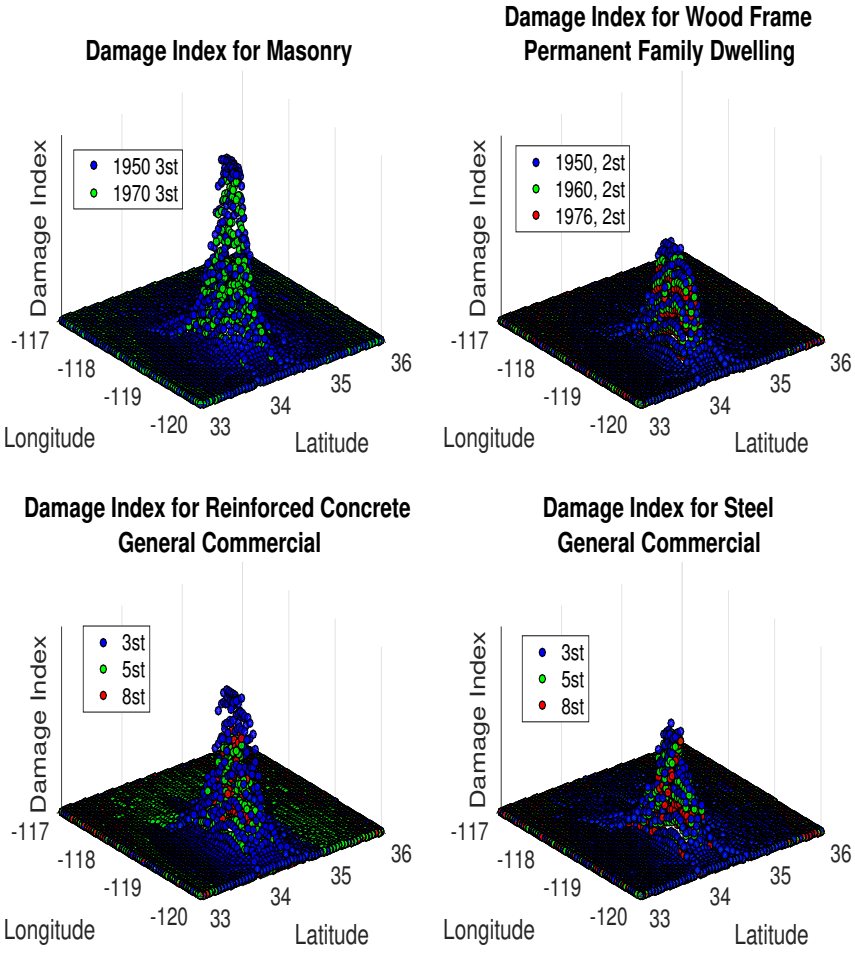

Figure 4.4: Maximum DI Per Building Class Northridge

4.4 Numerical Results for Case Studies

The Bayesian update procedure was applied for 1997 2010 Chile and 1994 California earthquakes using all the available observed real event descriptions resulting in 39, 89 and 120 updates respectively. Tables 4.1-4.11 summarize the real index parameters used for each update step. Figures 4.5 and 4.6 show the earthquakes location and magnitude relative to the catalog of simulated events. Given these properties, the larger 2010 event should result in a much larger loss compared to the 1997 earthquake. In addition it is expected that the 1994 Northridge earthquake would result in medium to large losses.

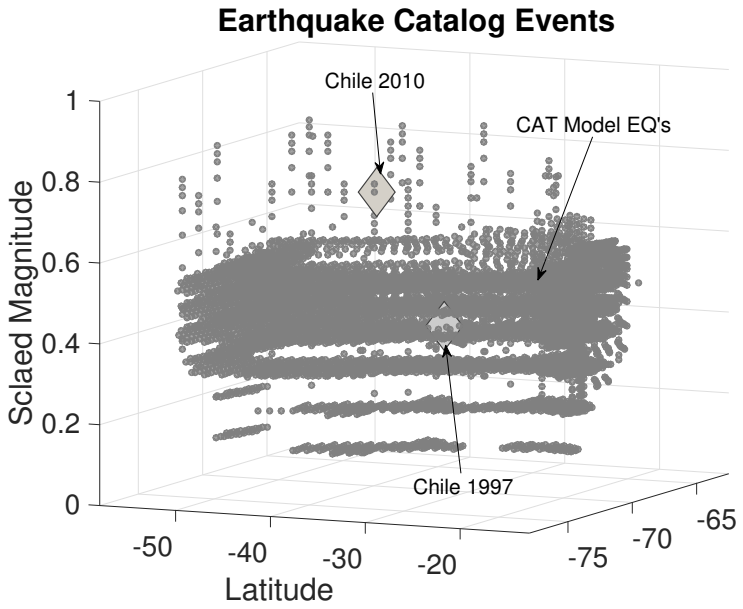

Figure 4.5: Chile CAT model EQ's

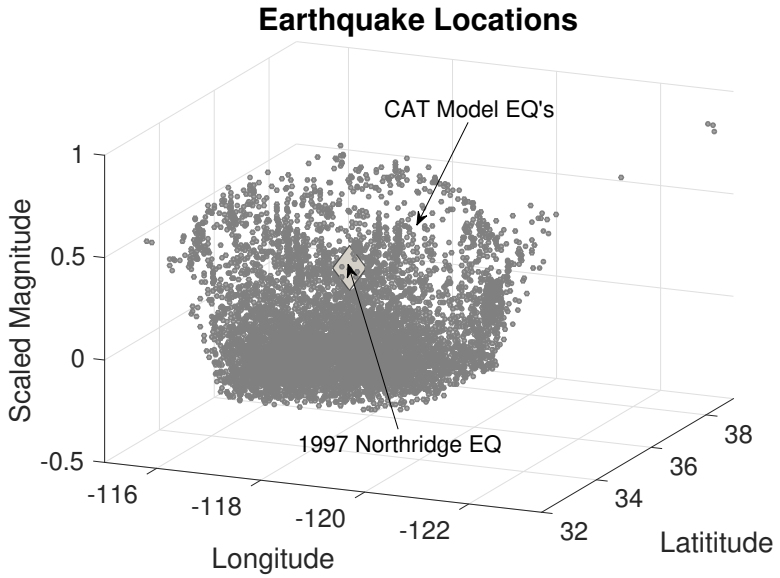

Figure 4.6: California CAT model EQ's

4.4.1 Chile 2010 EQ

Following the procedure outlined in the previous section, the methodology was applied to the 8.8 magnitude, 2010 Chile Earthquake. Figure 4.7 shows the average loss for four different standard deviation values. The monetary loss for the first 59 updates remains relatively

constant and begins to increase until reaching a stead final value. The type of input data affects the loss output differently, with peak ground acceleration (upd2-59) having the smallest effect and damage reports having the greatest influence (upd60-89).

There are a few factors that caused updates involving peak ground acceleration to have little effect on the loss. Firstly the resolution of the model output was 5km, which meant that accelerometer stations were sometimes far away from the buildings location in the model analysis. Secondly peak ground acceleration does not directly correlate to loss because specific building attributes may result in magnifying the effect and resulting in a larger spectral acceleration. Lastly instead of taking the PGA of the closest building all buildings within a 5km radius were averages, which could result in overall smaller value. Despite the jump in loss all four curves converge to similar values despite the different standard deviation values.

A second method of interpreting the results is calculating loss confidence intervals. Figure 4.8 shows a summary of the results for 50% 70% and 90% confidence intervals for four different standard deviations. In general the upper bound of the 90th % confidence interval tend to converge to similar values despite the different standard deviation values. Larger standard deviation values result in a larger range of loss values, whereas the smaller standard deviation result in a relatively narrow band of confidence interval loss values. These results are consistent with the way the likelihood function is defined. A larger standard deviation value results in a flatter likelihood function thus a wider range of values are given similar weights. A smaller standard deviation value results in a likelihood function that magnifies only a few large events and gives little weight to smaller simulated earthquakes.

4.4.2 Punitaqui 1997 EQ

The same method was implemented for the smaller 1997 event. As was expected Figure 4.10 shows a smaller overall loss vs. the larger 2010 events. The results are relatively constant until the end where there is a large increase in values. This can be explained by the

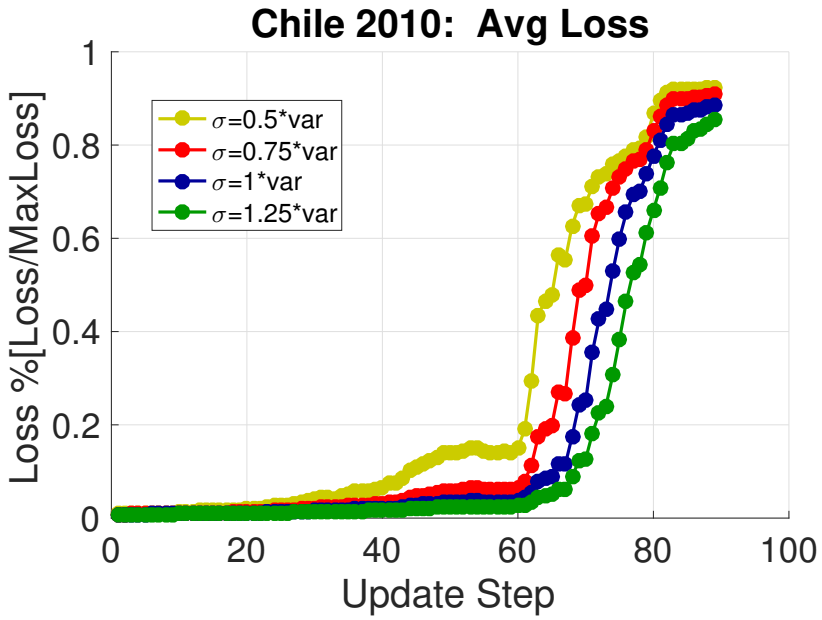

Figure 4.7: Chile 2010 Weighted Average Loss

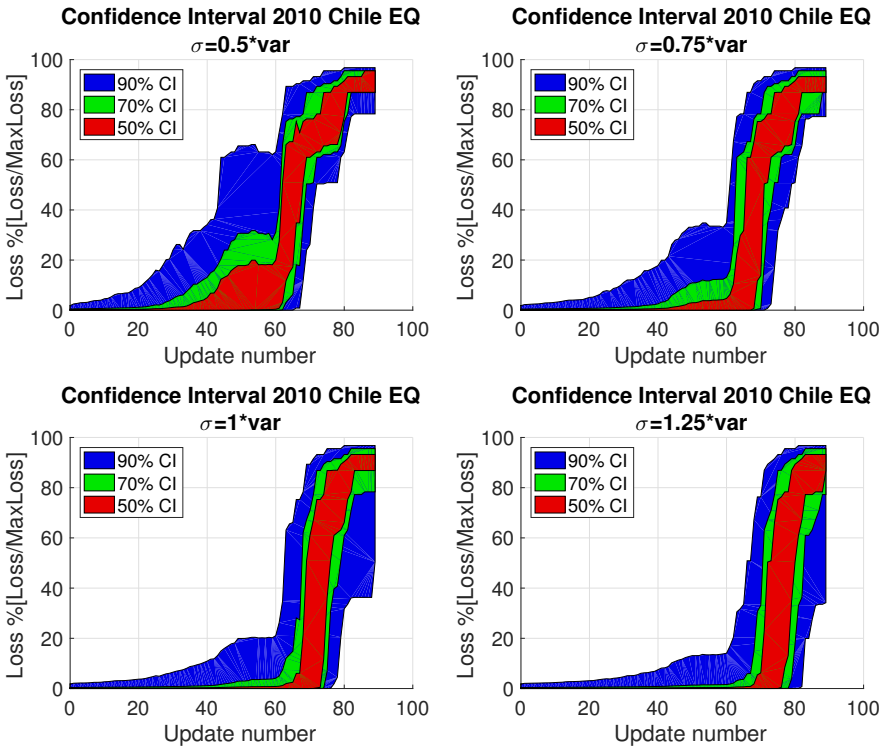

Figure 4.8: Chile 2010 Confidence intervals

real earthquake parameters used for the last few updates, which were damage descriptions resulting in relatively large damage indices (table: 4.10) compared to the previous steps. These large simulated damage indices tended to give a larger weight to events that produced larger loss values.

For this smaller event the effect of the standard deviation was more pronounced, with the smallest standard deviation resulting in a large jump in the average loss values. A possible cause for this drastic change is that a likelihood function with a small standard deviation give a larger weight to an earthquake producing a larger loss compared to those with a larger standard deviation value. This shows the importance in choosing an appropriate standard deviation value or range of values in order to achieve steady results. For this case a standard deviation of 0.75, 1 and 1.25 * variance produced similar final loss values.

The loss confidence interval results are summarized in Figure 4.9. In general the lower bound of all three confidence intervals are equal or nearly equal to zero loss. These results are reasonable because a smaller earthquake would tend to produce little monetary loss resulting in a smaller loss confidence interval values. Similar to the trends shown for the average loss values the smaller standard deviation values tended to show the most drastic change in loss values.

4.4.3 1994 Northridge Earthquake

The Bayesian update was also applied to the 1994 Northridge event using magnitude, location, pga values and building tagging data. The weighted average loss for different standard deviation values is shown in Figure 4.11. These result show a larger deviation between different variances of the likelihood function, but the results for 0.25var stabilized for the last 10 update steps. It is expected that if additional information was used for the procedure at the higher standard deviation values they would eventually converge to similar values to that of the smallest standard deviation. The loss confidence interval results are summarized in Figure 4.12. In general the smaller the standard deviation the smaller the bounds of the

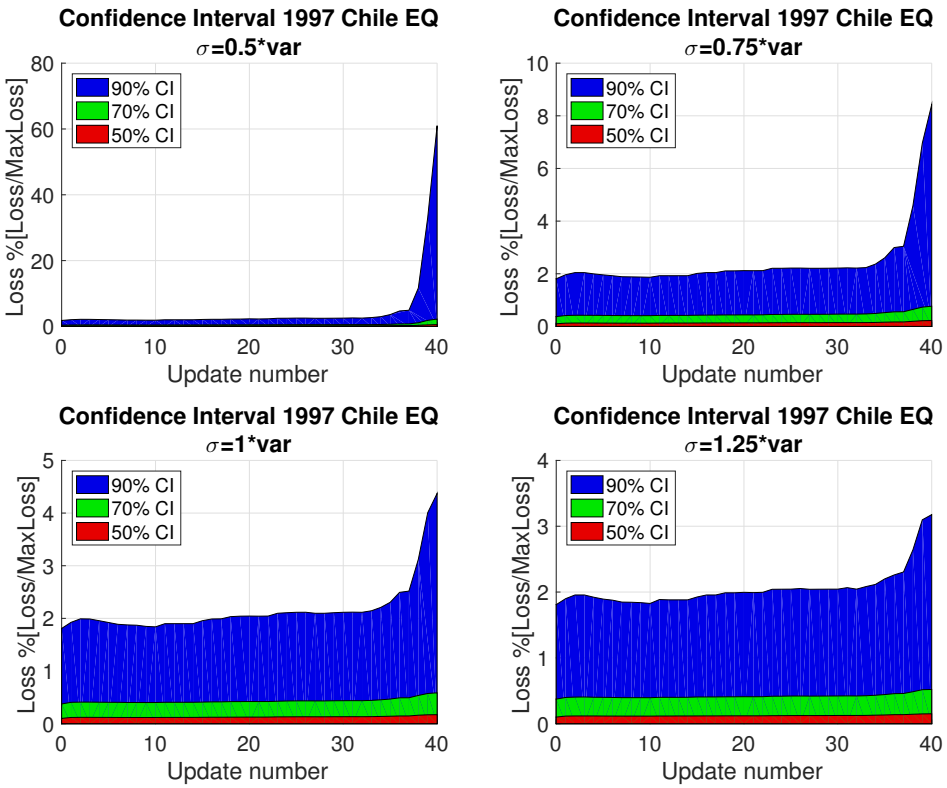

Figure 4.9: Chile 1997 Confidence intervals

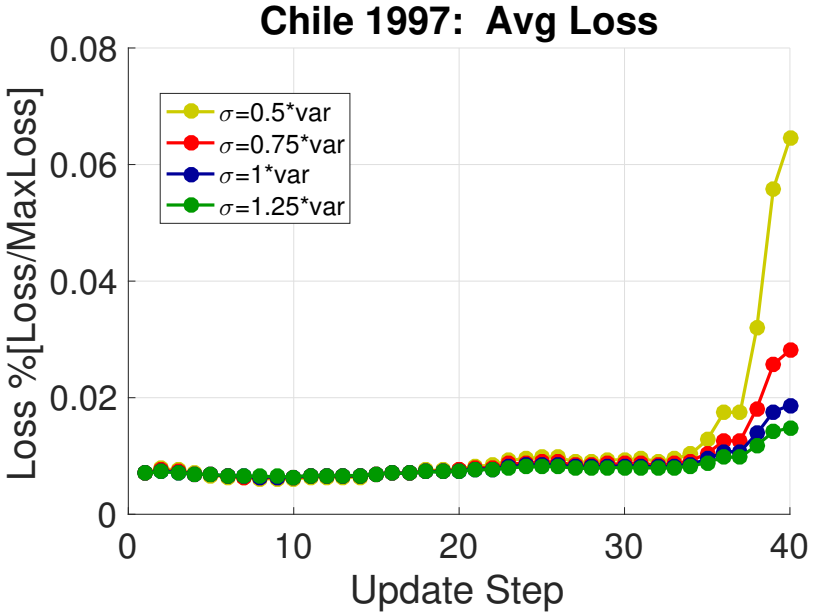

Figure 4.10: Chile 1997 Weighted Average Loss

confidence intervals. These results are reasonable because a smaller standard deviation filters out events faster because the contribution of an event close to the mean of the Gaussian LF is much larger than one further away.

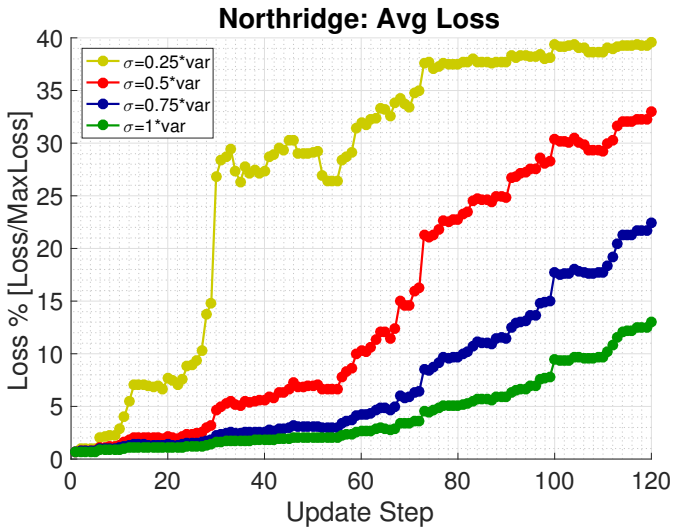

Figure 4.11: Northridge 1994 Average Loss

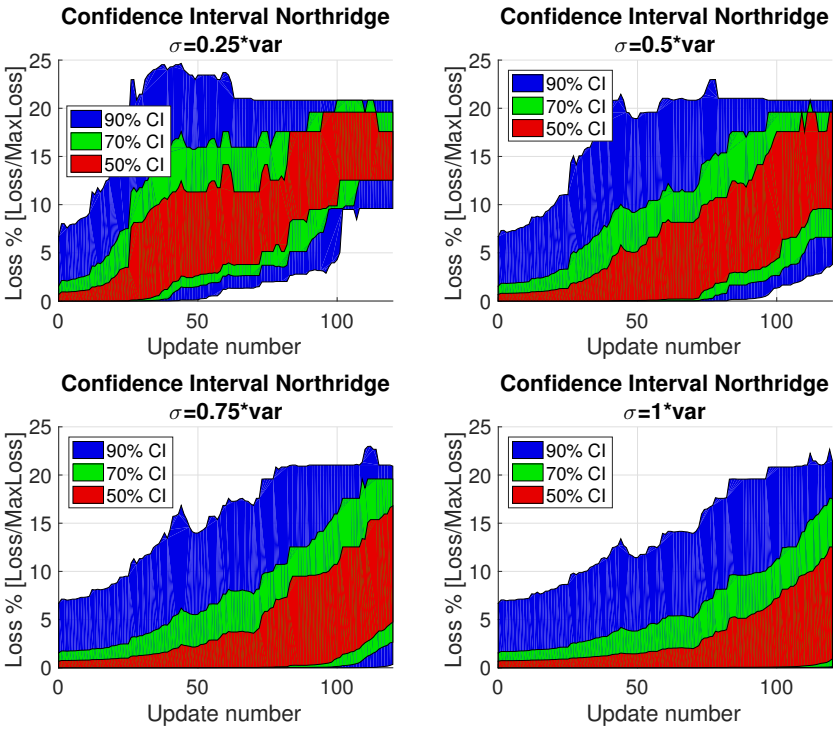

Figure 4.12: Northridge 1994 Confidence intervals

Chapter 5

Discussion

5.1 Challenges

The proposed Bayesian analysis has many benefits because of the versatility of its possible applications, but it also has some associated challenges. One of the challenges is lack of access to necessary data or delay in access. Ideally immediately after the event helpful earthquake attributes would become available so they can be used to quickly calculate updated loss estimates. Unfortunately this is not always the case because it takes time to gather certain types of data, such as building damage assessments. In addition, even if data has been compiled their may be an additional delay to public access to the necessary records. Despite this possible barrier USGS usually provides earthquake magnitude and location immediately after it an event occurs, which can be used to begin the update process.

Once earthquake attributes become available the next challenge is processing the data so it can be used in the methodology. In order to use real earthquake data there must be a comparable attribute for the simulated event catalog. For features like event magnitude and location this is simple because there is already an equivalent parameter for the simulated events, but this is not always the case. The most difficult attributes to process are qualitative measures like damage reports. It is important to be as consistent as possible when

quantifying these parameters, but since the form of the data may vary this process can become difficult. Once this obstacle is overcome any type of data can be incorporated into the update procedure.

After data has been collected and processed the final hurdle is to define an appropriate likelihood function to calculate updated losses. For the case with a normal pdf, defining a standard deviation can be difficult. If the standard deviation is too large the updated loss may not be effected, but if it is too small there could be a large jump in the updated loss. Even though defining this parameter is subjective, in general setting it equal to a factor times the range of data values for the update step produced steady results.

There are also external factors that can effect the methodology and final results. The accuracy of the available real building data has great influence on the CAT models loss output. The more detailed the insured building data the more accurate the models can predict the loss of the building to different simulated events. The proposed methodology determines the weights of the simulated events to the final loss, but if the losses from the models aren't reliable then the loss updates will not be reliable either. Despite all these challenges, for the presented case studies, the methodology was able to converge to reasonable estimates for both Chilean earthquakes.

5.2 Sensitivity Analysis Tagging Factors

A major step in the Bayesian update process starts before the analysis is even implementing: preprocessing data. Depending on the type of data this step can be unbiased or subjective. Quantifying the damage description and building tagging into damage indices is the most subjective process. In order to see the effect of this quantification step different factor value were tested for building tagging. Factors of 0.9 and 0.8 were used for red, 0.6 and 0.5, were used for yellow and 0.2 and 0.3 were used for red tagging designation and shown in Figure 5.1. Modifying the green tagging weight had the greatest effect because the number

of buildings with a tagging designation of green was the largest of all the color. Modifying the red factor had the smallest effect because very few buildings were tagged red. In order to maximize the stability of the results a factor combination of 0.8, 0.6 and 0.2 was chosen for red, yellow and green tagging designation.

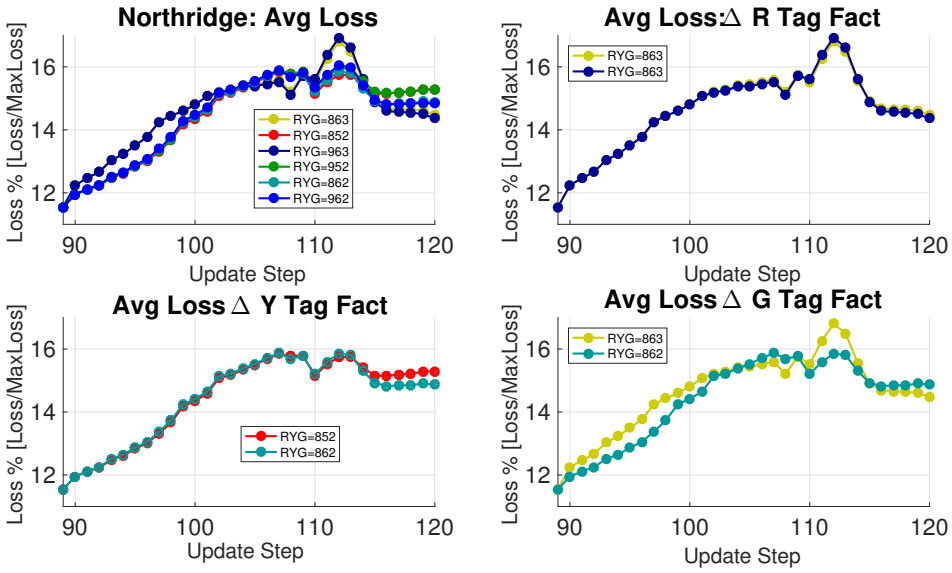

Figure 5.1: Sensitivity Analysis

5.3 Effect of Update Order on Final Posterior Distribution

One of the major benefits of Gaussian analysis is that it is an iterative process the takes into account the cumulative effect of all available input data. The order the available input is used affects the loss path but it should not change the final loss value. In order to test the effect of update order both numerical and analytic calculations were performed.

First an analytical solution was calculated by writing out three complete update step using Bayes theorem. In addition, in order to simplify the expression, for any given step the prior distribution $P_n(A_n)$ was replaced with the posterior of the previous step $P_{n-1}(A_n|B_{n-1})$.

Update Step 1:

$$P_1(A_n|B_1) = \frac{P_1(A_n)P_1(B_1|A_n)}{\sum_{n=1}^{N_{EQ}} P_1(A_n)P_1(B_1|A_n)} \tag{5.1}$$

Update Step 2:

$$
\begin{aligned}
P_2(A_n|B_2) &= \frac{P_2(A_n)P_2(B_2|A_n)}{\sum_{n=1}^{N_{EQ}} P_2(A_n)P_2(B_2|A_n)} = \frac{P_1(A_n|B_2)P_2(B_m|A_n)}{\sum_{n=1}^{N_{EQ}} P_1(A_n|B_2)P_2(B_2|A_n)} \\
&= P_2(B_2|A_n) * \frac{P_1(A_n)P_1(B_2|A_n)}{\sum_{n=1}^{N_{EQ}} P_1(A_n)P_1(B_1|A_n)} * \sum_{n=1}^{N_{EQ}} \left[\frac{\sum_{n=1}^{N_{EQ}} P_1(A_n)P_1(B_1|A_n)}{P_2(B_1|A_n)P_1(B_1|A_n)P_1(A_n)} \right] \\
&= \frac{P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)}{\sum_{n=1}^{N_{EQ}} P_1(A_n)P_1(B_1|A_n)} * \frac{\sum_{n=1}^{N_{EQ}} P_1(A_n)P_1(B_1|A_n))}{\sum_{n=1}^{N_{EQ}} P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)} \\
&\Rightarrow P_2(A_n|B_2) = \frac{P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)}{\sum_{n=1}^{N_{EQ}} P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)}
\end{aligned}
\tag{5.2}
$$

Update Step 3:

$$
\begin{aligned}
P_3(A_n|B_3) &= \frac{P_3(A_n)P_3(B_3|A_n)}{\sum_{n=1}^{N_{EQ}} P_3(A_n)P_3(B_3|A_n)} = \frac{P_2(A_n|B_2)P_3(B_3|A_n)}{\sum_{n=1}^{N_{EQ}} P_2(A_n|B_2)P_3(B_3|A_n)} \\
&= \frac{P_3(B_3|A_n)P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)}{\sum_{n=1}^{N_{EQ}} P_2(B_2|A_n)P_1(B_2|A_1)P_1(A_n)} \sum_{n=1}^{N_{EQ}} \left[\frac{\sum_{n=1}^{N_{EQ}} P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)}{P_3(B_3|A_n)P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)} \right] \\
&= \frac{P_3(B_3|A_n)P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)}{\sum_{n=1}^{N_{EQ}} P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)} * \frac{\sum_{n=1}^{N_{EQ}} P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)}{\sum_{n=1}^{N_{EQ}} P_3(B_3|A_n)P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)} \\
&\Rightarrow P_3(A_n|B_3) = \frac{P_3(B_3|A_n)P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)}{\sum_{n=1}^{N_{EQ}} P_3(B_3|A_n)P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)}
\end{aligned}
\tag{5.3}
$$

After fully simplifying three update steps a generic equation was calculated for a posterior probability calculation. The posterior distribution for any given update step m is equal to the product of the baseline prior times all the likelihood functions for steps 1 through m divided by the sum of this product for every simulated event 1 through N. The likelihood definition is independent of the update step because it is only a function of the update data values. The baseline prior is a user-defined constant that is also independent of the update sequence. Thus, as can be seen by the commutative property of multiplication, equations 32-34 prove that update order does not affect the final answer.

Update Step M:

$$P_m(A_n|B_m) = \frac{P_1(A_n)P_1(B_1|A_n)P_2(B_2|A_n) * ... * P_{m-1}(B_m|A_n)P_m(B_m|A_n)}{\sum_{n=1}^{N_{EQ}} P_1(A_n)P_1(B_1|A_n)P_2(B_2|A_n) * ... * P_{m-1}(B_{m-1}|A_n)P_m(B_m|A_n)} \quad (5.4)$$

After the analytical proof was completed a numerical example was performed to ensure it was consistent with the analytical results. For each case study three cases were calculated: chronological update order and two randomly shuffled update orders. As shown by Figure 5.2 the final loss converged to the same value regardless of the update order. This is an important property because it shows what is important is not the order the earthquake data is received but rather the contents of the data itself.

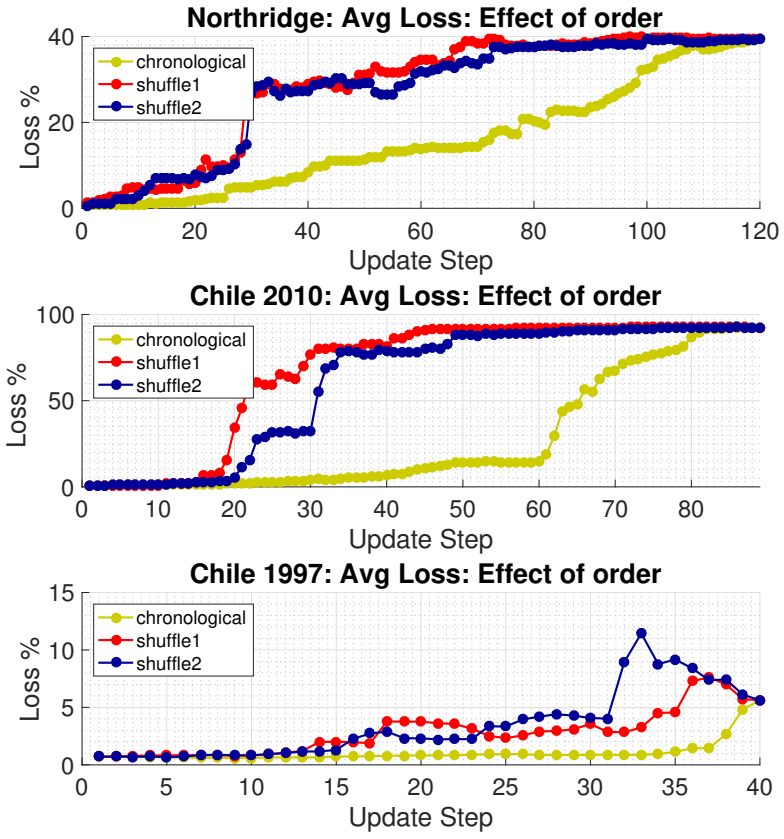

Figure 5.2: Effect of Update Order

Chapter 6

Conclusions

6.1 Conclusion

The update steps are sensitive to the damage quantification steps and the standard deviation of the likelihood function. The conversion between the qualitative damage descriptions to damage indices should be defined carefully because it greatly impacts the updated posteriors. For this analysis the likelihood function was always set to a normal probability distribution function. In general the larger 8.8 magnitude event produced at least an order of magnitude larger loss values compared the 7.1 magnitude event. The rate at which the confidence interval loss values converged to a loss increased with smaller standard deviation values of the Gaussian likelihood functions.

6.2 Future Work

A more stringent way of defining the standard deviation of the likelihood function should be defined. In addition multiple shapes for the Likelihood Function need to be explored in order to come up with a less biased update. The general procedure shows promise and could potentially be applied to multiple hazards. The next step could be to implement this methodology to floods or hurricanes and validate results with real documented losses.

Chapter 7

Bibliography

[1] "M 8.8-offshore Bio-Bio, Chile." [Online]. Available: https://earthquake.usgs.gov/earthquakes/eventpage/official20100227063411530_30#executive

[2] "M 7.1-Coquimbo, Chile." [Online]. Available: http://earthquake.usgs.gov/earthquakes/eventpage/usp000892b#executive

[3] USGS Earthquake Hazards Program, "M 6.7 - 1km NNW of Reseda, CA." [Online]. Available: https://earthquake.usgs.gov/earthquakes/eventpage/ci3144585{#}executive

[4] "M 8.8- offshore Bio-Bio, Chile shakemap." [Online]. Available: http://earthquake.usgs.gov/earthquakes/eventpage/official20100227063411530_30#shakemap

[5] M. Pardo, D. Comte, T. Monfret, R. Boroschek, and M. Astroza, "The October 15, 1997 Punitaqui earthquake (Mw=7.1): a destructive event within the subducting Nazca plate in central Chile," *Tectonophysics*, vol. 345, no. 1–4, pp. 199–210, feb 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S004019510100213X

[6] M. Astroza, O. Diaz, M. Pardo, and S. Rebolledo, "LECCIONES DE UN TERREMOTO DE SUBDUCCION DEL TIPO INTRAPLACA. EL TERREMOTO DE PUNITAQUI DEL 14 DE OCTUBRE DE 1997," *VIII*

Jorndada Chilenas de Sismologia e Ingenieria Antisismica. [Online]. Available: https://www.researchgate.net/publication/265293035_LECCIONES_DE_UN_TERREMOTO_DE_SUBDUCCION_DEL_TIPO_INTRAPLACA_EL_TERREMOTO_DE_PUNITAQUI_DEL_14_DE_OCTUBRE_DE_1997

[7] "Sismo Destructivo De Punitaqui, 14 Octubre 1997," p. November 2009. [Online]. Available: http://repositoriodigitalonemi.cl/web/bitstream/handle/2012/1093/SismoDestructivoPunitaqui1997_Coquimbo.pdf?sequence=1

[8] EQE International, GIS, and GOES, "The Northridge Earthquake of January 17, 1994: Report of Data Collection and Analysis. Part A: Damage and Inventory Data," EQE International, Geographic Information Systems Group of Governer's Office of Emergency Services, Tech. Rep., 1995.

[9] D. A. Linzer, "Dynamic Bayesian Forecasting of Presidential Elections in the States," *Journal of the American Statistical Association*, vol. 108, no. 501, pp. 124–134, mar 2013. [Online]. Available: http://dx.doi.org/10.1080/01621459.2012.737735

[10] P. Reggiani and A. H. Weerts, "A Bayesian approach to decision-making under uncertainty: An application to real-time forecasting in the river Rhine," *Journal of Hydrology*, vol. 356, no. 1–2, pp. 56–69, jul 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0022169408001698

[11] K. Porter, J. Mitrani-Reiser, and J. L. Beck, "Near-real-time loss estimation for instrumented buildings," *The Structural Design of Tall and Special Buildings*, vol. 15, no. 1, pp. 3–20, mar 2006. [Online]. Available: http://dx.doi.org/10.1002/tal.340

[12] R. Taylor, C.E. CHang, S.E. Eguchi, "Updating Real-Time Earthquake Loss Estimates: Methods, Problems and Insights," *MCEER: Earthquake Engineering to Extreme Events*, 2001.

[13] M.-W. Dictionary, "Insurance." [Online]. Available: https://www.merriam-webster.com/dictionary/insurance

[14] "Perils Models General Methodology," CoreLogic Inc, Tech. Rep.

[15] A. H.-S. Ang and W. H. Tang, *Probability Concepts in Engineering Planning and Design. Volume II-Decision, Risk and Reliability*, 1st ed., 1990.

[16] M. Kendall, A. Stuart, and J. Ord, "Vol. 1: Distribution theory," *London [etc.]: Arnold [etc.]*, 1994.

[17] S. Pei and J. van de Lindt, "Methodology for earthquake-induced loss estimation: An application to woodframe buildings," *Structural Safety*, vol. 31, no. 1, pp. 31–42, jan 2009. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0167473008000040

[18] W. H. Encycolopedia, "Law of Haversines." [Online]. Available: http://www.worldlibrary.org/articles/law_of_haversines

[19] W. Bryc, *The Normal Distribution*. New York: Springer-Verlag, 1995. [Online]. Available: https://link.springer.com/book/10.1007%2F978-1-4612-2560-7

[20] J. Kenney and E. Keeping, *Mathematics of Statistics*, 2nd ed. Princeton: Van Nostrand, 1951.

[21] "Poster of the Seismicity of the Nazca Plate and South America." [Online]. Available: https://earthquake.usgs.gov/earthquakes/eqarchives/poster/regions/nazca.php

[22] S. C. E. D. Center, "Significant Earthquakes and Faults." [Online]. Available: http://scedc.caltech.edu/significant/northridge1994.html

[23] EQE International, GIS, and GOES, "The Northridge Earthquake of January 17,1994: Report of Data Collection and Analysis. Part B: Analysis and Trends," EQE Inter-

national, Geographic Information Systems Group of Governer's Office of Emergency Services, Tech. Rep., 1997.

Appendices

Appendix A

Notation

Utility Theory	
$E[\ \ ]$	The expected value
$Loss(A_n)$	The loss from the actual Earthquake
$Loss(A_n)$	loss from simulated earthquake A_n given in the catalog
$P(A_n)$	the probability of occurrence of simulated earthquake A_n in the catalog
C_n	the cost for event n
P_n	the probability that event n will occur

Bayesian Notation	
B_m	occurrence of a real earthquake event (either 1997 Punitaqui and 2010 Maule)
A_n	Simulated earthquake event n where $n \in [1, N_{EQ}]$
$P_m(B_m\|A_n)$	conditional posterior probability of having a simulated earthquake event A_n given you have a real earthquake event B
$\widetilde{P}_m(B_m\|A_n)$	unnormalized conditional posterior probability (prior*Likelihood) of having a simulated earthquake event A_n given you have a real earthquake event B
$P_m(A_n)$	Prior probability of update step m
$P_m(B_m)$	Total probability of having an event exactly like real earthquake B

Haversine Formula	
D	distance [km] between two points on a sphere with coordinates $(lat_1, long_1)$ and $(lat_2, long_2)$
R	radius of the earth used in Haversine Formula, taken to be 6371km
$lat_1,\ lat_2$	latitude of location 1 and 2 on the world
$long_1,\ long_2$	longitude of location 1 and 2 on the world

Real Input Parameters	
$X_m^{(*)}$	the real earthquake parameter of update step m
$\widetilde{M_m^{(*)}}$	unnormalized magnitude of a real earthquake event at update step m
$M_m^{(*)}$	normalized magnitude of a real earthquake event a update step m, where $M_m^{(*)} \in [0, 1]$
$\widetilde{D}_m^{(*)}$	unnormalized distance [km] between a real event's epicenter to itself
$D_m^{(*)}$	normalized distance [km] between a real events epicenter to itself $D^{(*)} = 0$
$DI_m^{(*)}$	damage index of real earthquake at update step m, where $DI_m^{(*)} \in [0, 1]$
$PGA_m^{(*)}$	peak ground accelerations at real accelerometer location of update step m
$BT_m^{(*)}$	building tagging damage index at update step m

Simulated Input Parameters	
$X_m^{(n)}$	the simulated input parameter of earthquake n of update step m
$\widetilde{M}_m^{(n)}$	unnormalized magnitude of a simulated earthquake event n and update step m
$M_m^{(n)}$	normalized magnitude of a simulated earthquake event n, where $M_m^{(n)} \in [0, 1]$
$\widetilde{D}_m^{(n)}$	unnormalized distance [km] between simulated event n's epicenter to the epicenter of the real earthquake
$D_m^{(n)}$	normalized distance [km] between simulated event n's epicenter to the epicenter of the real earthquake where $D_m^{(n)} \in [0, 1]$
$DI_m^{(n)}$	damage index (loss/building value) of simulated earthquake n at update step m where $DI_m^{(n)} \in [0, 1]$
$PGA_m^{(n)}$	peak ground acceleration for simulated event n at update step m
$BT_m^{(n)}$	building tagging damage index for simulated event n at update step m

Likelihood Functions	
$f_{X_m^{(n)}}(\quad)$	Normal distribution probability density function of update step n
μ_m	the mean of a single variate normal distribution for update step m
σ_m^2	the standard deviation of a single variate normal distribution for update step m
$f_{\mathbf{X_m^{(n)}}}(\quad)$	bivariate normal probability density function
$\boldsymbol{\mu}_\mathbf{m}$	mean vector of a bivariate normal distribution of update step m
$\mathbf{V_m}$	Covariance matrix of bivariate normal of update step m

Miscellaneous	
$L_m^{(n)}(i)$	total weighted average loss due simulated event n at update step m and building i
$L_m^{(n)}$	the total loss due to simulated earthquake n at update step m
L_m	total weighted average loss due all simulated events at update step m
L_B^C	Loss for building C due to earthquake B
$\mathbf{L_G}$	global loss matrix with $dim = [N_{bld}\ N_{EQ}]$ where $\mathbf{L_G}(i, j)$ is the loss of building i due to simulated event j
N_{bld}^{5km}	number of buildings withing a 5km radius of an accelerometer location
CDF	cumulative distribution function
PDF	probability density function
M_{max}	Maximum probable earthquake magnitude
M_{min}	Minimum probable earthquake magnitude

Appendix B

List of Equations

$$E[B] = \sum_{n=1}^{N} C_n * P_n \tag{B.1}$$

$$E[Loss(b)] = \sum_{n=1}^{N_{EQ}} Loss(A_n) * P(A_n) \tag{B.2}$$

$$P_m(A_n|B_m) = \frac{P_m(A_n)P_m(B|A_n)}{P_m(B_m)} = \frac{P_m(A_n)P_m(B_m|A_n)}{\sum_{n=1}^{N_{EQ}} P_m(A_n)P_m(B_m|A_n)} \tag{B.3}$$

$$P_m(B) = \sum_{n=1}^{N_{EQ}} P_m(A_n)P_m(B|A_n) \tag{B.4}$$

$$M_m^{(n)} = \frac{\widetilde{M}_m^{(n)} - M_{min}}{M_{max} - M_{min}}; \tag{B.5}$$

$$D = 2R \arcsin\sqrt{sin^2\left(\frac{lat_2 - lat_1}{2}\right) + cos(lat_1)cos(lat_2)sin^2\left(\frac{long_2 - long_1}{2}\right)} \tag{B.6}$$

$$Chile : D_m^{(n)} = \frac{\widetilde{D}_m^{(n)} - 0km}{4300km - 0km}; \quad Northridge : D_m^{(n)} = \frac{\widetilde{D}_m^{(n)} - 0km}{700km - 0km} \tag{B.7}$$

$$[X_m^{(*)}] = [PGA^{(*)}] \tag{B.8}$$

$$[X_m^{(n)}] = \sum_{i=1:N_{bld}^{5km}}^{N_{bld}^{5km}} \frac{PGA_m^{(n)}(i)}{N_{bld}^{5km}} \tag{B.9}$$

$$[X_m^*] = [PGA_m^{(*)}] \tag{B.10}$$

$$X_m^{(*)} = DI_m^{(*)} \tag{B.11}$$

$$EQ_{unique} = \bigcup_{i=1}^{18} EQ_{port,i}; \quad dim(EQ_{unique}) = [Nx1] \tag{B.12}$$

$$dim(\mathbf{L_G}) = [N_{bld}\ N_{EQ}] \tag{B.13}$$

$$X_m^{(n)} = \frac{L_m^{(n)}}{1000000} = [X_m^{(1)}; X_m^{(2)}; \ldots; X_m^{N}] \tag{B.14}$$

$$BT_m^{(*)} = \frac{0.8 * bld_{red} + 0.6 * bld_{yellow} + 0.2 * bld_{green}}{bld_{red} + bld_{yellow} + bld_{green}}; \tag{B.15}$$

$$BT_m^{(n)} = \frac{\sum_{i=1}^{N_{bld}^{zip5km}} L_m^{(n)}(bldi, portj)}{\$1,000,000 * N_{bld}^{zip5km}}; \tag{B.16}$$

$$\int_{n=1}^{N_{EQ}} Cdn = 1 \Rightarrow C = \frac{1}{\int_{n=1}^{N} dn} \xRightarrow{\text{discrete}} \frac{1}{N_{EQ}};\ P_1(A_n) = \frac{1}{N_{EQ}} \tag{B.17}$$

$$P_m(B_m|A_n) = f_{X_m^{(n)}}(y|\mu_m, \sigma_m^2) = \frac{1}{\sqrt{2\pi\sigma_m^2}} exp\left[-\frac{y-\mu_m}{2\sigma_m^2}\right] = \frac{1}{\sqrt{2\pi\sigma_m^2}} exp\left[-\frac{y-X_m^{(*)}}{2\sigma_m^2}\right] \tag{B.18}$$

$$f_{X_m^{(n)}}(X_m^{(n)}|\mu_m, \sigma_m^2) = \frac{1}{\sqrt{2\pi\sigma_m^2}} exp\left[-\frac{X_m^{(n)} - X_m^{*}}{2\sigma_m^2}\right] \tag{B.19}$$

$$\begin{aligned} P_m(B_m|A_n) &= f_{X_m}(\mathbf{y}|\mu_\mathbf{m}, \mathbf{V_m}); \quad \mathbf{y} = [x_1, x_2] \\ f_{\mathbf{X_m^{(n)}}}(\mathbf{y}|\mu_\mathbf{m}, \mathbf{V_m}) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_m^2}} exp\left[-\frac{z}{2(1-\rho_m^2)}\right] \end{aligned} \tag{B.20}$$

$$z = \frac{(x_1 - \mu_{1m}^2)^2}{\sigma_{1m}^2} - \frac{2\rho_m(x_1 - \sigma_{1m})(x_x - \sigma_{2m})}{\sigma_{1m}\sigma_{2m}} + \frac{(x_2 - \mu_{2m})^2}{\sigma_{2m}^2} \tag{B.21}$$

$$\mathbf{V_{12m}} = \begin{bmatrix} \sigma_{1m}^2 & \rho\sigma_{1m}\sigma_{2m} \\ \rho\sigma_{1m}\sigma_{2m} & \sigma_{2m}^2 \end{bmatrix};\ \mu_{\mathbf{12m}} = [\mu_{1m}; \mu_{2m}] \tag{B.22}$$

$$\mathbf{V_{12m}} = \begin{bmatrix} \sigma_m^2 & 0 \\ 0 & \sigma_m^2 \end{bmatrix}; \quad \mu_{12m} = [0.76; 0]_{2010EQ}\ or\ [0.42; 0]_{2010EQ} \tag{B.23}$$

$$z_{2010EQ} = \frac{(x_1 - 0.76)^2}{\sigma_m^2} - 0 + \frac{(x_2)^2}{\sigma_m^2}; \quad z_{1997EQ} = \frac{(x_1 - 0.42)^2}{\sigma_m^2} - 0 + \frac{(x_2)^2}{\sigma_m^2} \tag{B.24}$$

$$\mathbf{y} = X_1^{(n)} = [M_1^{(n)}, D_1^{(n)}]$$
$$f_{X_m}(X_1^{(n)}|\mu_{\mathbf{m}}, \mathbf{V_m}) = \frac{1}{2\pi\sigma_m^2} \exp\left[-\frac{1}{2}\left(\frac{(M_1^{(n)} - \mu_{1m})^2}{\sigma_m^2} + \frac{D_1^{(n)}}{\sigma_m^2}\right)\right] \tag{B.25}$$

$$\widetilde{P}_m(A_n|B_m) = P_m(A_n) * P_m(B_m|A_n) \tag{B.26}$$

$$P_m(A_n|B_m) = \frac{\widetilde{P}_m(A_n|B_m)}{\sum_{n=1}^{N_{EQ}} P_m(A_n)P_m(B_m|A_n)} = \frac{P_m(A_n) * P_m(B_m|A_n)}{\sum_{n=1}^{N_{EQ}} P_m(A_n)P_m(B_m|A_n)} = \frac{P_{m-1}(A_n|B_{m-1})P_m(B_m|A_n)}{\sum_{n=1}^{N_{EQ}} P_m(A_n)P_m(B_m|A_n)} \tag{B.27}$$

$$L_m^{(n)} = \sum_{i=1}^{N_{bld}} L_m^{(n)}(i) \tag{B.28}$$

$$CDF(x) = \int_{-\inf}^{x} PDF(t)dt \xrightarrow{\text{discrete}} CDF_{upd}(L_n) = \sum_{i=1}^{n} PDF_{upd,norm}(L_i) \tag{B.29}$$

$$P(L_a < L < L_b) = CDF(L_b) - CDF(L_a) = (50\% + d/2) - (50\% - d/2) = d\% \Rightarrow$$
$$d\%\ Confidence\ Interval : [L_{50\%-d/2}, L_{50\%+d/2}]$$
$$90\%\ CI : [L_{50\%-90\%}2, L_{50\%+(90\%)/2}] = [L_{5\%}, L_{95\%}]$$
$$70\%\ CI : [L_{50\%-(90\%)/2}, L_{50\%+(90\%)/2}] = [L_{15\%}, L_{85\%}]$$
$$50\%\ CI : [L_{50\%-(90\%)/2}, L_{50\%+(90\%)/2}] = [L_{25\%}, L_{75\%}]$$

(B.30)

$$L_m = \sum_{n=1}^{N_{EQ}} P_m(A_n|B_m) * L_m^{(n)} \tag{B.31}$$
$$min(L_m^{(n)}) < L_m < max(L_m^{(n)})$$

$$P_1(A_n|B_1) = \frac{P_1(A_n)P_1(B_1|A_n)}{\sum_{n=1}^{N_{EQ}} P_1(A_n)P_1(B_1|A_n)} \tag{B.32}$$

$$
\begin{aligned}
P_2(A_n|B_2) &= \frac{P_2(A_n)P_2(B_2|A_n)}{\sum_{n=1}^{N_{EQ}} P_2(A_n)P_2(B_2|A_n)} = \frac{P_1(A_n|B_2)P_2(B_m|A_n)}{\sum_{n=1}^{N_{EQ}} P_1(A_n|B_2)P_2(B_2|A_n)} \\
&= P_2(B_2|A_n) * \frac{P_1(A_n)P_1(B_2|A_n)}{\sum_{n=1}^{N_{EQ}} P_1(A_n)P_1(B_1|A_n)} * \sum_{n=1}^{N_{EQ}} \left[\frac{\sum_{n=1}^{N_{EQ}} P_1(A_n)P_1(B_1|A_n)}{P_2(B_1|A_n)P_1(B_1|A_n)P_1(A_n)} \right] \\
&= \frac{P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)}{\sum_{n=1}^{N_{EQ}} P_1(A_n)P_1(B_1|A_n)} * \frac{\sum_{n=1}^{N_{EQ}} P_1(A_n)P_1(B_1|A_n))}{\sum_{n=1}^{N_{EQ}} P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)} \\
&\Rightarrow P_2(A_n|B_2) = \frac{P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)}{\sum_{n=1}^{N_{EQ}} P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)}
\end{aligned}
\tag{B.33}
$$

$$
\begin{aligned}
P_3(A_n|B_3) &= \frac{P_3(A_n)P_3(B_3|A_n)}{\sum_{n=1}^{N_{EQ}} P_3(A_n)P_3(B_3|A_n)} = \frac{P_2(A_n|B_2)P_3(B_3|A_n)}{\sum_{n=1}^{N_{EQ}} P_2(A_n|B_2)P_3(B_3|A_n)} \\
&= \frac{P_3(B_3|A_n)P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)}{\sum_{n=1}^{N_{EQ}} P_2(B_2|A_n)P_1(B_2|A_1)P_1(A_n)} \sum_{n=1}^{N} \left[\frac{\sum_{n=1}^{N_{EQ}} P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)}{P_3(B_3|A_n)P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)} \right] \\
&= \frac{P_3(B_3|A_n)P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)}{\sum_{n=1}^{N_{EQ}} P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)} * \frac{\sum_{n=1}^{N_{EQ}} P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)}{\sum_{n=1}^{N_{EQ}} P_3(B_3|A_n)P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)} \\
&\Rightarrow P_3(A_n|B_3) = \frac{P_3(B_3|A_n)P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)}{\sum_{n=1}^{N_{EQ}} P_3(B_3|A_n)P_2(B_2|A_n)P_1(B_1|A_n)P_1(A_n)}
\end{aligned}
\tag{B.34}
$$

$$
P_m(A_n|B_m) = \frac{P_1(A_n)P_1(B_1|A_n)P_2(B_2|A_n) * ... * P_{m-1}(B_m|A_n)P_m(B_m|A_n)}{\sum_{n=1}^{N_{EQ}} P_1(A_n)P_1(B_1|A_n)P_2(B_2|A_n) * ... * P_{m-1}(B_{m-1}|A_n)P_m(B_m|A_n)} \tag{B.35}
$$