Effects of Earthquake Source Recurrence on the Conditional Seismic Hazard Analysis

by

Ismail Bahadir Kuzucu

Department of Civil and Environmental Engineering
Duke University

Date: ____________________

Approved:

________________________
Henri P. Gavin, Supervisor

________________________
Tomasz A. Hueckel

________________________
Omar M. Knio

Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Department of Civil and Environmental Engineering in the Graduate School of Duke University

2012
Abstract

(Civil and Environmental Engineering)

Effects of Earthquake Source Recurrence on the Conditional Seismic Hazard Analysis

by

Ismail Bahadir Kuzucu

Department of Civil and Environmental Engineering
Duke University

Date: ______________________

Approved:

__________________________
Henri P. Gavin, Supervisor

__________________________
Tomasz A. Hueckel

__________________________
Omar M. Knio

An abstract of a thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Department of Civil and Environmental Engineering in the Graduate School of Duke University

2012
Abstract

In earthquake engineering, the selection of input ground motions for nonlinear structural dynamic analysis is an important element in performance based structural design. Probabilistic seismic hazard analysis (PSHA) provides a basis for determination of ground motion characteristics by incorporating regressions on ground motion metrics from past recorded earthquakes for known seismic sources, propagation paths and local site conditions. Due to aleatoric variability in the processes of fault rupture, seismic wave propagation, and local site response; and, the epistemic uncertainty in models of these phenomena due in part to limited data, there exist uncertainties in seismic hazard assessments. This uncertainty should be interpreted carefully before the resulting ground motion predictions are adopted for dynamic analysis and design of structures. This thesis examines the sensitivity of algorithmic parameters and the choice of earthquake magnitude recurrence relations on the conditional ground motion characteristics. For this purpose, new correlation coefficients between pseudo-spectral acceleration (PSA), peak ground velocity (PGV), and cumulative absolute velocity (CAV) are derived using data from the PEER-NGA earthquake database. Finally, the sensitivity of the conditional mean spectrum to earthquake recurrence models is investigated. It is concluded that the choice of the Gutenberg-Richter or Characteristic Magnitude model can significantly affect the conditional mean spectra.
Contents

Abstract iv

List of Tables vii

List of Figures viii

1 Introduction 1

2 Basics of Probabilistic Seismic Hazard Analysis (PSHA) 4

2.1 Review of Probability Theory 4

2.1.1 Random Events 5

2.1.2 Conditional Probability and Independence 5

2.1.3 Random Variables 6

2.1.4 Multiple Random Variables 8

2.2 Magnitude Recurrence Relationships 9

2.2.1 Bounded Gutenberg-Richter Model: 11

2.2.2 Characteristic Magnitude Model: 13

2.3 Ground Motion Prediction Equations 14

2.3.1 Intensity Measures 15

2.4 Hazard Curve for Site Response 21

3 Development of New Correlation Coefficients Between PSA and PGV, CAV 26

3.1 Analysis Procedure 26

3.1.1 Methodology for Correlation Coefficient Computation 27
List of Tables

2.1 Input parameters for Ground motion prediction equations and response variables of the GMPEs .................................................. 19

3.1 Datasets utilized for correlation computation .................................................. 31
List of Figures

1.1 Steps of probabilistic seismic hazard analysis ........................................ 2
2.1 PDF and CDF of the G-R and the C-H models ..................................... 14
3.1 $Z_{TOR}$ and $Z_{2.5}$ effects in dataset-1 .................................................. 31
3.2 Correlation results using the BA08 and CB10 GMPE models ................. 32
3.3 Effects of using different GMPEs for dataset-1 and dataset-2 ............... 33
3.4 Effects of using different GMPEs for dataset-3 and dataset-4 ............... 34
3.5 Correlation coefficients between $Sa$ and $PGV$ ...................................... 36
3.6 An example of the conditional mean spectrum ....................................... 37
3.7 Correlation coefficients between $Sa$ and $CAV$ ..................................... 38
3.8 An example of the conditional mean spectrum ....................................... 39
4.1 Illustration of seismic hazard deaggregation ......................................... 41
4.2 Illustration of the example site, with one earthquake source ............... 44
4.3 An example of hazard curve and deaggregation plots for the AS08, BA08, CB08, and CY08 GMPE models .................................................... 45
4.4 Hazard curves for 1, 2, 5, and 10 km distance binning sizes .................. 46
4.5 Hazard curves for 1, 2, 5, and 10 km distance binning sizes .................. 48
4.6 Conditional mean spectra ................................................................. 49
4.7 Hazard deaggregation with different GMPE models .............................. 51
4.8 Hazard deaggregation with different GMPE models .............................. 52
4.9 Conditional response spectrum conditioned on $(PGV_{100}^{*})$ along with the Bounded Gutenberg-Richter recurrence model .... 53

4.10 Probability density function of earthquake magnitudes with the G-R and C-H models .................................................. 55

4.11 Probability density function of earthquake magnitudes along with the G-R and C-H models .................................................. 56

4.12 Hazard curves for cumulative absolute velocity based on the G-R and C-H models .................................................. 56

4.13 Conditional Mean Spectra given $CAV_{1000}$ for the example site .... 58
1

Introduction

Probabilistic seismic hazard analysis (PSHA) [Cornell (1968), McGuire (2008)] is widely used for evaluating the seismic risk at a certain site of an engineering project, and for estimation of seismic design loads. The goal of PSHA is to predict ground motions by integrating the probabilities of all earthquake scenarios across locations and sizes of future potential earthquakes. In other words, given all possible seismic sources at a known site, PSHA quantifies the probability of exceeding various ground shaking levels (ground motion intensities) and allows one to compute the earthquake hazard for that given site [McGuire (2004)]. Ground shaking levels (or, in other words, intensity measures) are typically represented as pseudo-spectral acceleration (denoted as PSA or simply Sa), peak ground acceleration (PGA) or peak ground velocity (PGV). Pseudo-spectral acceleration is defined as the intensity experienced by a simple oscillator of a given period and damping ratio (usually 5 percent). Peak ground velocity is the maximum value of velocity-time histories of earthquakes. And cumulative absolute velocity is defined as the integral of the absolute value of the acceleration-time histories of earthquakes. The accurate representation of ground shaking is a critically important factor in determining response of structures, so the
effects of using different intensity measures will be discussed in this study. Since there exist uncertainties in the size, location, and frequency of earthquakes, and also in the resulting estimation of ground motions, the earthquake hazard in PSHA incorporates all of these uncertainties.

PSHA involves the identification of possible seismic sources, recurrence relations for earthquake magnitude, models for seismic wave propagation, and the integration of these three steps in order to compute the hazard [see figure 1.1]. Firstly, identification and delineation of all seismic sources which are likely to affect a given site requires a spatial representation of seismic sources as a distribution due to the variability of the location of the rupture. Seismologic sources are commonly classified as fault (or line), distributed, and point sources [Cornell (1968)]. For example, if the source is a fault, the whole fault or a segment of fault may rupture. Therefore,
this creates geographical uncertainty for the location of rupture, and the so-called source-to-site distance of earthquakes. Secondly, frequency-magnitude distributions for each source are modeled as earthquake magnitude recurrence relations. Next, for the propagation of seismic energy, empirical ground motion prediction models (that are derived by regression analysis of strong-motion data as a function of earthquake size, source-to-site distance, and often specified other parameters are used. these relations provide distributions of potential ground motions for a given site, for one or more intensity measures. Finally, the seismic hazard is computed by integrating these previous steps: probabilities of all possible seismic sources are defined with distributions of possible magnitudes, locations, and possibly some other parameters; with resulting estimation of ground motions represented as the spectral peak response of a single degree of freedom oscillator (called the pseudo acceleration response spectra) [Newmark and Hall (1982)], are combined in the seismic hazard which is quantified as the rate of exceedance of various levels of intensity measures for a given site. These steps, with the assumptions and hypothesis adopted for seismic hazard analysis, are scrutinized in the following sections of this study. For example, effects of using different earthquake magnitude relations (the Gutenberg-Richter and the Characteristic Earthquake relations which are two competing magnitude recurrence models) on the conditional mean spectrum are examined. Before a detailed discussion of these two magnitude recurrence models, the concept of PSHA, relations between intensity measures, hazard curves, deaggregation of hazard curves, and the computation of the conditional mean spectra are addressed in the course of this study.
Basics of Probabilistic Seismic Hazard Analysis (PSHA)

PSHA is founded upon probability theory by which random variables are well-characterized and distributions of uncertainties are propagated through a hazard analysis. Many engineering problems, including earthquake related problems, involve large variability due to the random nature of environmental conditions. Therefore, a comprehensive understanding of probability concept is crucial in PSHA. Before a detailed discussion of the elements of PSHA, the concept of probability will be introduced in the following sections.

2.1 Review of Probability Theory

This review is intended to give a brief outline of the concept of probability that will allow deeper understanding of probabilistic seismic hazard analysis. Often uncertainties are classified into two categories as 'aleatoric' and 'epistemic' uncertainties [Bommer et al. (2005)]. The inherent random nature of earthquake occurrences, which is called aleatoric uncertainty, cannot be reduced by adding more information.
Earthquake scenarios (in terms of distance, magnitude, and possibly some other parameters) for a given site is an example for the aleatoric uncertainty because there may be a number of different earthquake scenarios. Another type of uncertainty, namely epistemic uncertainty, stems from imperfect or incomplete knowledge about real problems. Variability in the ground motion prediction equations is an example of epistemic uncertainties. Due to intrinsic variability and additionally due to incomplete knowledge and/or data, there is in sum a large variability in analysis of earthquake ground motions. These uncertainties consequently lead to the decisions with an imperfect confidence. So, handling these uncertainties and their interpretation is important for decision making and design in structural and earthquake engineering problems. In this section some concepts in probability theory such as random variables, the total probability theorem, joint probability, conditional probability, correlation and correlation coefficients are briefly introduced.

2.1.1 Random Events

Probability theory is used to portray an event or variable for which the exact value cannot be known. The random event (e.g., an experiment) therefore has more than one possible outcome, and all the possible outcomes are referred to as sample space of the event (denoted \( E \)). If a particular outcome of an event is unknown, this unknown quantity is called the random variable (usually denoted with a capital letter such as \( X \)).

2.1.2 Conditional Probability and Independence

If the occurrence of an event is independent of another event, these two events are said to be independent events; however, if a depedence exist between these two events, they are said to be dependent and associated probability is called conditional probability [Ang and Tang (2007)]. The conditional probability of event \( A \), given
event $B$ is

$$P(A \mid B) = \frac{P(AB)}{P(B)} \quad A, B \subset E \text{ and } P(B) > 0 \quad (2.1)$$

If the events $A$ and $B$ are independent of each other, then equation (2.1) becomes

$$P(A \mid B) = P(A) \quad A, B \subset E \text{ and } P(B) > 0 \quad (2.2)$$

If events $B_i$ are finite or countably infinite partition of sample space $E$, the probability of event $A$ is given as

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i)P(B_i) \quad A, B_i \subset E \text{ and } P(B_i) > 0 \quad (2.3)$$

Equation (2.3) is known as the law of total probability in probability theory. For all possible events, Bayes’ theorem gives the probability of a particular $i^{th}$ event $B_i$, given event $A$ already occurred. This is similar to total probability theorem except Bayes’ theorem is considered as the inverse of the total probability theorem, and is given by

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \quad A, B \subset E \text{ and } P(B) > 0 \quad (2.4)$$

2.1.3 Random Variables

A random variable $X$ is a mapping defined as $X : E \to \mathbb{R}$. Possible values that the random variable $X$ may have are denoted lower-case ($x$ in this case). If a random variable $X$ takes finite or countably finite number of possible values, it is said to be a discrete random variable; on the other hand, if $X$ is defined in an uncountably infinite horizon, then it is called a continuous random variable. There are different ways to represent the outcomes of events (or random variables). These, so-called
probability distribution functions, are probability mass function (PMF), probability density function (PDF), and cumulative distribution function (CDF). Probability distribution functions are briefly described in following subsections.

**Cumulative distribution function (CDF):** CDF is a mapping such that $F_X : \mathbb{R} \rightarrow [0, 1]$, and given as

$$F_X(x) = P(X \leq x)$$  \hspace{1cm} (2.5)

**Probability mass function (PMF) and Probability density function (PDF):**

Probability mass function of a discrete random variable $X$ is a function defined as $p_X : E \rightarrow \mathbb{R}$ such that

$$p_X(x) = P(X = x)$$  \hspace{1cm} (2.6)

Let $X$ be a continuous random variable such that $F_X(x)$ is differentiable with respect to $x$. There exists a nonnegative mapping $f_X : \mathbb{R} \rightarrow [0, \infty)$ such that

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f(y) \, dy$$  \hspace{1cm} (2.7)

where $F$ is probability density function of $X$. It can also be shown that

$$f_X(x) = \frac{dF_X(x)}{dx}$$  \hspace{1cm} (2.8)

**Expectation:** Let $X$ be a random variable and an arbitrary real-valued function $g$ be $g : \mathbb{R} \rightarrow \mathbb{R}$. The expected value (also known as average or mean) of $g(X)$, in
case $X$ is discrete, is given by

$$E[g(X)] = \sum_{x \in X} g(x) p(x)$$ (2.9)

If random variable is continuous, expectation of $g(X)$ is given by

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)dx$$ (2.10)

**Variance of a random variable:** Variance is a measure of dispersion of a random variable about its mean. Variance and standard deviation are respectively defined by:

$$Var[X] = E[(X - \mu_X)^2]$$ (2.11)

$$\sigma_X = \sqrt{Var[X]}$$ (2.12)

where $\mu_X$ is the mean or average of $X$, and $\sigma_X$ is standard deviation of $X$.

### 2.1.4 Multiple Random Variables

Let $X_1, ..., X_N \subset E$ be discrete random variables. We can write the joint PMF of $X_1, ..., X_N$ as

$$p_{X_1, ..., X_N}(x_1, ..., x_N) = P(X_1 = x_1, ..., X_N = x_N)$$ (2.13)

In the continuous case, we can write the joint probability of an event $A \subset \mathbb{R}^N$ as

$$P[(X_1, ..., X_N) \in A] = \int ... \int_{(x_1, ..., x_N \in A)} f(x_1, ..., x_N)dx_1...dx_N$$ (2.14)

where $f(x_1, ..., x_N)$ is joint PDF of random variables $X_1, ..., X_N$.  

8
Let’s say we have two random variables \( X \) and \( Y \). These are jointly distributed with

\[
P(X \leq a, Y \leq b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f(x, y) \, dx \, dy \quad (2.15)
\]

If there exists a joint probability density of random variables \( X \) and \( Y \), then the covariance of \( X \) and \( Y \) is given by

\[
\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - E[X]E[Y] \quad (2.16)
\]

Covariance is a measure of the relative change of a random variable with respect to another. If two random variables are independent of each other, it means they are uncorrelated, therefore covariance is equal to zero. This can be seen in equation (2.16) such that if \( X \) and \( Y \) are independent random variables, then \( E[XY] = E[X]E[Y] \), and therefore \( \text{Cov}(X, Y) = 0 \). If \( X \) and \( Y \) are dependent variables, then the Pearson correlation coefficient of \( X \) and \( Y \) given as

\[
\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \quad (2.17)
\]

For more detailed information on probability, readers are referred to [Ang and Tang (2007)] or any elementary probability theory text.

2.2 Magnitude Recurrence Relationships

In PSHA, there is a need to represent the frequency of exceedance of earthquakes in terms of magnitudes. This is the magnitude recurrence relation. Representing the size of earthquakes in terms of the standard Richter magnitude, earthquake magnitude recurrence relations explain how frequently earthquakes occur above a certain range of magnitude (usually expressed annually). The magnitude recurrence rela-
tion is typically represented as a cumulative recurrence rate of occurrences of earth- 
quakes with respect to earthquake magnitude. There exists aleatoric uncertainty in 
the earthquake magnitude recurrence. That is, an earthquake source may produce 
earthquakes with a range of sizes. Therefore, the magnitude of an earthquake is 
described as a probability distribution, not a single magnitude value. Well-known 
magnitude recurrence models are the Bounded Gutenberg-Richter (G-R) and the 
Characteristic-Magnitude (C-H) models, and both models are widely employed in 
seismic analysis [Wesnousky (1994); Parsons and Geist (2009); Musson (2012)]. How 
seismic hazard analysis is affected by the choice of the magnitude recurrence model 
and by the parameters used in these models is to be addressed in this thesis.

Gutenberg and Richter (1956) developed a model to describe the earthquake 
magnitude distribution, as represented by power-law distribution in very large seis-
mic regions containing multi-faulting systems. On the other hand, some researchers 
[Schwartz and Coppersmith (1984); Youngs and Coppersmith (1985); Wesnousky 
(1994)] propose that magnitude distributions of earthquakes do not always follow the 
Gutenberg-Richter model. Especially, if the seismic region consists of individual 
faults or fault segments with regular geometries, such as the San Andreas seismic 
zone in CA, the Gutenberg-Richter model may not reflect the distribution of the 
seismicity over the whole range of magnitudes up to largest earthquake magnitudes. 
This discrepancy between the G-R model and real earthquakes is explained by the 
idea that larger earthquakes tend to have similar magnitudes. While the G-R re-
currence model may well represent the distribution of small earthquake magnitudes, 
the G-R model underestimates the frequency of large earthquakes [Kramer (1996); 
Youngs and Coppersmith (1985); Parsons and Geist (2009)]. To prevent this under-
estimation of larger earthquakes, the characteristic earthquake model represents the 
distribution of small earthquakes (say, moment magnitude 4-6) as an exponential 
distribution (G-R model) and larger earthquakes, which rupture nearly the entire
fault with complete energy release, as a uniform distribution. From this definition, a characteristic earthquake can simply be considered as the maximum earthquake which may be generated by a seismic source, as estimated by paleoseismic (geologic) observations. Because the characteristic earthquake model (C-H) constitutes a distribution dominated by smaller and larger size earthquakes, earthquakes of intermediate sizes have a relatively smaller frequency of occurrence.

In typical sites, where there are multiple active faults in a large region, the Gutenberg-Richter magnitude recurrence model better represents earthquake magnitude recurrence. On the other hand, for individual seismic sources or fault segments, the characteristic magnitude recurrence model may be favored for PSHA analysis [Schwartz and Coppersmith (1984)]. These two models serve the same purpose, so one can ask the question: which one is better? Some researchers propose that the G-R model still be applied to such seismic regions as the San Andreas and some believe that it may not be appropriate; so, there has been a scientific debate on favoring GR or CH model [Parsons and Geist (2009); Field and Page (2011)]. The effect of choosing either the G-R and the C-H recurrence models on the ground motion response will be provided in Chapter 4. Empirical cumulative distribution functions (CDF), probability distribution functions (PDF), and the mean annual rate of exceedance $\lambda_m$ equations for both models are provided in the following subsections.

2.2.1 Bounded Gutenberg-Richter Model:

Magnitude recurrence relation model proposed by Gutenberg and Richter (1956) is

$$\log(\lambda_m) = a - bm$$

where $\lambda_m$ is the rate of earthquakes with magnitudes greater than $m$. The constant $a$ and $b$ values are estimated from historical observations. The constant $a$ is associated with the seismicity of a given region. The $b$ value physically denotes the relative ratio
of the number of small earthquakes to large earthquakes; that is, higher values of 
b indicate that the distribution incorporates a greater number of small earthquakes 
than larger earthquakes in size.

The CDF and PDF of the doubly-bounded (having minimum and maximum 
magnitude threshold) G-R model is given by Kramer (1996) as follows,

\[
P[M \leq m \mid m_{\text{min}} \leq m \leq m_{\text{max}}] = F_M(m) = \frac{1 - 10^{-b(m-m_{\text{min}})}}{1 - 10^{-b(m_{\text{max}}-m_{\text{min}})}}
\]

(2.19)

\[
f_M(m) = \frac{d}{dm} F_M(m) = \frac{b \ln 10 \ 10^{-b(m-m_{\text{min}})}}{1 - 10^{-b(m_{\text{max}}-m_{\text{min}})}}
\]

(2.20)

where \(m_{\text{min}}\) and \(m_{\text{max}}\) are minimum and maximum credible earthquake magnitudes. 
Truncation bounds, \(m_{\text{min}}\) and \(m_{\text{max}}\), are possible credible earthquake magnitude values 
for a given seismic zone, and \(b\) is the slope of equation (2.20). The \(m_{\text{min}}, m_{\text{max}},\) 
and \(b\) parameters must be judiciously selected because their effect on ground motion 
predictions and response of structures to such ground motion predictions may be 
significant. There has been scientific debate on the assessment of these parameters 
[Weichert (1980); Bender and Campbell (1989); McCann and Reed (1990); Beauval and Scott (1989); Makropoulos and Burton (1983); Kijko (2004)]. The typical 
values of \(b = 1\) (\(b\) usually takes values between 0.8 and 1) is discussed in detail by 
Kagan (2002). For typical buildings, a minimum earthquake magnitude below which 
significant damage is unlikely to occur is often selected as 5 [Petersen et al. (2008)], 
and for more important and seismically vulnerable structures such as nuclear power 
plants and hospitals, this minimum magnitude threshold can be selected smaller due 
to higher safety concerns [McCann and Reed (1990)]. Ultimately, values of the \(m_{\text{min}},\) 
\(m_{\text{max}},\) and \(b\) are to be carefully estimated for a given fault or a fault system (note 
that the type of structure should be considered for \(m_{\text{min}}\) and \(m_{\text{max}}\)).
2.2.2 Characteristic Magnitude Model:

The CDF and PDF of the characteristic earthquake model is given by Youngs and Coppersmith (1985) as

\[
F_C(m) = \begin{cases} 
(1 - p) \frac{1 - 10^{-b(m - m_{\min})}}{1 - 10^{-b(m_t - m_{\min})}} & \text{for } m_{\min} \leq m \leq m_t \\
(1 - p) \frac{m_{\max} - m}{m_{\max} - m_t} & \text{for } m_t \leq m \leq m_{\max}
\end{cases} \tag{2.21}
\]

\[
f_C(m) = \begin{cases} 
(1 - p) b \ln 10 \frac{10^{-b(m - m_{\min})}}{1 - 10^{-b(m_t - m_{\min})}} & \text{for } m_{\min} \leq m \leq m_t \\
p \frac{1}{m_{\max} - m_t} & \text{for } m_t \leq m \leq m_{\max}
\end{cases} \tag{2.22}
\]

where \(m_t = m_{\max} - 0.5\) and the parameter \(p\) is equal to the area between \(m_t\) and \(m_{\max}\). The area equal to \(p\) in the probability density of the characteristic magnitude model represents the characteristic part and the remainder represents the exponential part of the distribution. The PDF and CDF of the G-R and the C-H models are illustrated in figure 2.1. See Chapter 3 for discussions about these magnitude recurrence models.

The mean annual rate of exceedance for the exponential model (G-R model) is given by McGuire and Arabasz (1990) as

\[
\lambda_m (M > m) = \lambda_{\min} P[M > m \mid m_{\min} \leq m \leq m_{\max}] = \lambda_{\min} (1 - F_M(m)) \tag{2.23}
\]

where the mean annual rate of minimum earthquake magnitude is given as \(\lambda_{\min} = 10^{a - b m_{\min}}\), and \(a\) and \(b\) are constant parameters as discussed earlier.

Similarly, the mean annual rate of exceedance for the characteristic earthquake model can be computed as

\[
\lambda_m (M > m) = \lambda_{\min} P[M > m \mid m_{\min} \leq m \leq m_{\max}] = \lambda_{\min} (1 - F_C(m)) \tag{2.24}
\]
In equations (2.23) and (2.24), $F_M(m)$ and $F_C(m)$ denote CDF of magnitudes with the exponential magnitude recurrence model and the characteristic magnitude recurrence model, respectively.

### 2.3 Ground Motion Prediction Equations

Ground motion prediction equations estimate the distribution of possible ground motions at a site produced by a nearby fault rupture of known characteristics. The intensity of ground motions are typically characterized by peak ground acceleration (PGA), pseudo-spectral acceleration (PSA or sometimes simply denoted as Sa), and peak ground velocity (PGV). Other widely used intensity measures are peak ground displacement (PGD), cumulative absolute velocity (CAV), arias intensity ($I_a$), and significant earthquake duration (Ds). Intensity measures are important factors in
the estimation of ground motions, and therefore are important in the estimation of structural response because each ground motion intensity possesses distinct information about the shaking. For instance, earthquake duration related intensity measures may be useful if degrading structural response could lead to collapse, in other words, if the cumulative effects of ground motions is of interest.

2.3.1 Intensity Measures

In this thesis, the correlation of PSA (or sometimes denoted as Sa) with PGV and CAV are of interest. PGV has been a widely used peak intensity measure for damage prediction of buildings, macroseismic intensity estimates, evaluation of liquefaction potential of soil due to its strong correlation to ground strain, and estimation of earthquake damage on buried pipelines [Bommer and Alarcon (2006)]. There have been a few different definitions of CAV and PGV in the literature for specific engineering applications such as the geometric mean of horizontal components of ground motions (which is widely used) and the larger of the two horizontal components of ground motions. Specifically, popular CAV definitions are the geometric mean cumulative absolute velocity ($CAV_{GM}$, [Campbell and Bozorgnia (2010)]) and standardized absolute velocity ($CAV_{std}$, O’Hara and Jacobson (1991)). $CAV_{std}$ was first proposed in EPRI-1991 in order to filter out the non-damaging contributions of long and low-amplitude earthquakes. The focus in this study is on the integral of absolute acceleration time history of earthquake records, which is obtained by computing the geometric mean of CAVs of the as-recorded horizontal components of each record. According to the definition of cumulative absolute velocity, the equation is given by Reed and Kennedy (1988) as

$$CAV^{2}_{GM} = \int_{0}^{t_{max}} |a_{x}(t)| \, dt \int_{0}^{t_{max}} |a_{y}(t)| \, dt$$

(2.25)
where $t_{\text{max}}$ is the duration of each record, and $|a_x(t)|$ and $|a_y(t)|$ are the absolute values of two horizontal components ($x$ and $y$) of acceleration time histories. So, CAV represents the total energy measure of ground shaking, and clearly cumulative absolute velocity is non-decreasing with time and accounts for the cumulative effects of ground motions in time.

It is worth noting that, although CAV is believed to be a superior integral damage estimator particularly for nuclear power plants and for the assessment of liquefaction potential of soil, CAV has shortcomings because it does not account for the high phase energy releases of some earthquakes. However, the use of CAV along with another intensity measure (e.g., one that accounts for instances of high amplitude energy releases) may give better estimates of severity of earthquake damages. The vector or conditional intensity measures (Baker and Cornell (2005); Bradley (2010)) may be applied to this problem.

Ground motion intensity measures have uncertainties and are assumed log-normal with mean and variance dependent upon magnitude, distance, and other characteristics of sources and the site. These relations are called ground motion prediction equations and are empirical fits to data from a large number of recorded historical ground motions. The general form of a GMPE is

$$\ln IM = \mu_{\ln IM}(M, R, \Theta) + \sigma_{\ln IM}(M, R, \Theta) \epsilon$$

(2.26)

where $\epsilon$ is the number of logarithmic standard deviations from the logarithmic mean estimation. $\ln IM$ is the natural logarithm of a ground motion intensity measure, and $M$, $R$, and $\Theta$ respectively denote moment magnitude, source-to-site distance, and other site, path, and source parameters. The functional relations in GMPEs, $\mu_{\ln IM}(M, R, \Theta)$ and $\sigma_{\ln IM}(M, R, \Theta)$, are respectively the mean of the natural logarithm and the standard deviation of the mean logarithmic intensity mea-
sure. So, due to the lognormal distribution assumption, the exponential of the mean value of the log intensity measure gives the median of the intensity measure \( (\exp(\mu_{lnIM}(M,R,\Theta)) = IM_{med}) \).

Ground motion prediction equations (GMPEs) present some advantages and shortcomings. GMPEs are advantageous because they provide an easy way of predicting ground motions for a given site as a function of seismic source, path and site parameters; however, there is large uncertainty in the prediction models due to the intrinsic variability and lack of enough earthquake data to create better estimates of ground motions. For instance, ground motion prediction equations are obtained by regressing the strong-motion database involving records mostly from highly active and/or intermediate seismic regions, so they are successful in the application of intermediate and highly active seismic regions. On the other hand, because there are few earthquake records in low-seismic regions, it is difficult to obtain good estimates of ground motions by adopting a regression analysis by using the records only from the associated low-seismic site. Also, ground motion prediction equations do not provide sufficient estimation of very large and rare events (e.g. paleoearthquakes) due to scarcity of data for these events. For these reasons, for the ground motion prediction of low-seismic regions, records are combined from other regions which have similar seismotectonic regimes. A discussion about this issue is provided by Egozcue et al. (1991). Even if the earthquake database is assembled from seismically active regions, the uncertainty coming from natural variability of earthquakes and local site behavior (because each record measures the response of its own point location), GMPEs ultimately estimate the average ground motion with a standard deviation provided in each model. This allows engineers to consider uncertainty in the estimation of ground motions when using current prediction methods.

In 2008, by using the Pacific Earthquake Engineering Research Center (PEER) strong-motion database [Chiou et al. (2008)], five different groups of researchers in
the Next Generation Ground-Motion Attenuation Models (NGA) project developed empirical ground motion prediction equations which are applicable to active shallow crustal tectonic regions in the Western United States [Power et al. (2008)]. Four of the NGA ground motion prediction models are adopted in the course of this study: Abrahamson and Silva (2008), Boore and Atkinson (2008), Campbell and Bozorgnia (2008), and Chiou and Youngs (2008) (also see the errata for AS08, Abrahamson and Silva (2009)). Note that for the sake of brevity they are hereafter referred by the abbreviations AS08, BA08, CB08, and CY08 respectively. These prediction models provide an estimation of ground motions in terms of Sa, PGA, and PGV intensity measures. In addition to the aforementioned four NGA models, the Campbell and Bozorgnia (2010) model is adopted for the estimation of ground motions in terms of CAV in this study, and CB10 is used to refer this model. Another recent model for CAV prediction Danciu and Tselentis (2007) is not adopted in this study due to the limited applicability of the model in terms of magnitude range (magnitude < 6) [Campbell and Bozorgnia (2010)]. For response intensity measures available for the GMPEs, see Table 2.1.

Each ground motion prediction equation is derived from specific subsets of the NGA database, so the applicability of each GMPE model is limited. AS08 suggests that the model is applicable in magnitude range, 5 ≤ M ≤ 8.5 and distance range (km), Rrup ≤ 200; BA08 suggests magnitude range, 5 ≤ M ≤ 8, distance range (km), Rjb ≤ 200, and 180 ≤ Vs30 ≤ 1300 (m/s); CB08 is applicable for magnitude ranges 4 ≤ M ≤ Mmax, (Mmax equals to 7.5 for normal, 8 for reverse and 8.5 for strike-slip faulting mechanisms), distance range (km) Rrup ≤ 200, site shear wave velocity 150 ≤ Vs30 ≤ 1500 (m/s), the rupture depth Ztor ≤ 15 km, the soil depth Z2.5 ≤ 10 km, and dip angle 15 ≤ δ ≤ 90 (degrees); CY08 is suggested to be applicable in magnitude ranges 4 ≤ M ≤ Mmax, (Mmax equals to 8 for normal and reverse, and 8.5 for strike-slip faulting mechanisms), distance range (km) Rrup ≤
Table 2.1: Input parameters for Ground motion prediction equations and response variables of the GMPEs

<table>
<thead>
<tr>
<th>Parameters</th>
<th>GMPEs</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS08</td>
<td>BA08</td>
</tr>
<tr>
<td>(M)</td>
<td>Moment magnitude</td>
</tr>
<tr>
<td>(R_{rup})</td>
<td>Shortest distance to coseismic rupture (km)</td>
</tr>
<tr>
<td>(R_{jb})</td>
<td>Joyner-Boore distance (km)</td>
</tr>
<tr>
<td>(R_x)</td>
<td>Horizontal distance from top edge of rupture (km)</td>
</tr>
<tr>
<td>(\delta)</td>
<td>Dip angle of fault (degrees)</td>
</tr>
<tr>
<td>(Z_{tor})</td>
<td>Depth to top of rupture (km)</td>
</tr>
<tr>
<td>(Z_{1.0})</td>
<td>Depth to (V_s = 1.0\text{ km/s}) at site (m)</td>
</tr>
<tr>
<td>(Z_{2.5})</td>
<td>Depth to (V_s = 2.5\text{ km/s}) at site (km)</td>
</tr>
<tr>
<td>(V_{s30})</td>
<td>Average shear wave velocity for top 30m (m/s)</td>
</tr>
<tr>
<td>(A_{1100})</td>
<td>PGA on a rock outcrop with (Vs30=1100\text{ m/s}) (g)</td>
</tr>
<tr>
<td>(W)</td>
<td>Down-dip rupture width (km)</td>
</tr>
<tr>
<td>(F_f)</td>
<td>Faulting mechanism flag</td>
</tr>
<tr>
<td>(F_{hw})</td>
<td>Hanging-wall flag</td>
</tr>
<tr>
<td>(F_{as})</td>
<td>Aftershock flag</td>
</tr>
<tr>
<td>(PSA)</td>
<td>Pseudo-spectral acceleration</td>
</tr>
<tr>
<td>(PGA)</td>
<td>Peak ground acceleration</td>
</tr>
<tr>
<td>(PGV)</td>
<td>Peak ground velocity</td>
</tr>
<tr>
<td>(PGD)</td>
<td>Peak ground displacement</td>
</tr>
<tr>
<td>(CAV)</td>
<td>Cumulative absolute velocity</td>
</tr>
</tbody>
</table>

200, and \(150 \leq V_{s30} \leq 1500\text{ m/s}\); and CB10 is applicable in magnitude ranges \(5 \leq M \leq M_{max}\) (\(M_{max}\) equals to 7.5 for normal, 8 for reverse, and 8.5 for strike-slip faulting mechanisms), distance ranges (km) \(R_{rup} \leq 100\) for \(M < 7\) and \(R_{rup} \leq 200\) for larger magnitudes, \(150 \leq V_{s30} \leq 1500\text{ (m/s)}\), \(Z_{tor} \leq 15\text{ (km)}\), \(Z_{2.5} \leq 10\text{ (km)}\), and \(15 \leq \delta \leq 90\text{ (degrees)}\). These constraints should be considered in applications of the GMPEs.

Most of the models use the moment magnitude \((M)\), source-to-site distance \((R)\), faulting mechanism (normal, reverse etc.), and local site characteristics as input parameters (or variables) for the prediction of ground motions. As mentioned earlier, GMPEs are function of some source, path and site variables such as \(M\) and \(R\) which are uncertain and practically modeled with probability distributions. For the input parameters for each GMPE used in this study, see Table 2.1.

The dataset adopted in fitting model parameters is an important factor that affects the estimation of ground motions from different GMPEs. As a result of each author’s preference and interpretation of the earthquake database, GMPEs are
modeled with different parameters and this represents another source of variability of ground motion predictions from different GMPEs. A comprehensive understanding of the input parameters and understanding of pros and cons of each model is essential for a decent ground motion prediction and a consistent comparison of the GMPE models. Discussions about the comparison of GMPEs are presented by Abrahamson et al. (2008) and Scasserra et al. (2009).

Parameters illustrated in Table 2.1 may be classified into three categories as source parameters: $M$, $\delta$, $Z_{tor}$, $W$, $F_f$, and $F_{as}$; path parameters: $R_{rup}$, $R_{jb}$, $R_x$, and $F_{hw}$; and site parameters: $Z_{1.0}$, $Z_{2.5}$, $V_s30$, and $A_{1100}$. As observed in GMPE models, three different distance measures are adopted: the closest distance from the site to the coseismic rupture plane, $R_{rup}$, the closest horizontal distance from site to the vertical projection of rupture, $R_{jb}$, and horizontal distance to the surface projection of the top edge of the rupture plane measured perpendicular to the strike, $R_x$. Relations among these distance definitions and between rupture length and width can be obtained empirically and from geometric relations. So, that allows one to make a consistent comparison of different GMPEs’ results. For a more thorough explanation on these conversion methods, see Wells and Coppersmith (1994), Scherbaum et al. (2004), and Chiou and Youngs (2008). So, if one of the distance parameters is missing, conversion relations may be utilized. Similarly, if one of the site parameters $Z_{1.0}$ and/or $Z_{2.5}$ is missing, conversion relations provided by Abrahamson and Silva (2008), Chiou and Youngs (2008), and Campbell and Bozorgnia (2007) may be used. Abrahamson and Silva (2008) and Chiou and Youngs (2008) provided equations (2.27) and (2.28) derived by the relation between $Z_{1.0}$ and $V_s30$, and Campbell and Bozorgnia (2007) provided equations (2.29) and (2.30) derived by the relation of $Z_{2.5}$ to $Z_{1.0}$ and $Z_{1.5}$. $Z_{1.5}$ is defined as the depth to 1.0 km/sec shear-wave velocity horizon (m). Note that in this study a vertical strike-slip fault is assumed because of the simplicity of conversions of parameters in this case.
\[ Z_{1.0} = \begin{cases} 
\exp(6.745) & \text{for } V_{s30} < 180 \text{ m/s} \\
\exp\left[6.745 - 1.35 \ln\left(\frac{V_{s30}}{180}\right)\right] & \text{for } 180 \leq V_{s30} \leq 500 \text{ m/s} \\
\exp\left[5.394 - 4.48 \ln\left(\frac{V_{s30}}{500}\right)\right] & \text{for } V_{s30} > 500 \text{ m/s} 
\end{cases} \quad (2.27) \]

\[ Z_{1.0} = \exp\left[28.5 - \frac{3.82}{8} \ln(V_{s30}^8 + 378.78)\right] \quad (2.28) \]

\[ Z_{2.5} = 636 + 1.549 Z_{1.5} \quad (2.29) \]

\[ Z_{2.5} = 519 + 3.595 Z_{1.0} \quad (2.30) \]

Note that an updated velocity model is used in CY08, so using (2.28) would better estimate \( Z_{1.0} \). In case \( Z_{1.5} \) is known equation (2.29) can be used to estimate \( Z_{2.5} \) or if \( Z_{1.0} \) is known, equation (2.30) can be used. However, if both \( Z_{1.0} \) and \( Z_{1.5} \) are unknown, either AS08 and CY08 (preferred because of the updated velocity model) may be exploited. Also note that these conversion equations for source and path parameters are used in the development of correlation coefficients of PGV and Sa, and CAV and Sa in the following sections of this study.

### 2.4 Hazard Curve for Site Response

The purpose of PSHA, as explained in previous sections, is to compute the probability of exceedance of various ground motion intensity levels given all possible sources and recurrence relations of ground shakings, with the resulting ground motion. So PSHA yields a rate of exceedance curve as a function of a particular ground motion intensity. An intensity measure is a quantity that measures the effects of ground motions on structures, so the choice of intensity measure (IM) is an important factor
to accurately reflect the response of the structure (or ground response). Peak ground acceleration PGA, peak ground velocity (PGV), and spectral acceleration at fundamental periods of structures -usually at the fundamental mode period- \( (Sa(T_1)) \) are often used as intensity measure of ground motions. More sophisticated methods to represent measured of response of structures (or response of ground) have also been developed such as vector-valued intensity measures [Baker and Cornell (2005)] and generalized conditional intensity measures [Bradley (2010)]. In studies, it is offered that the use of conditional ground motion intensity (therefore conditional ground motion prediction) may better reflect the seismic hazard when it is conditioned upon a specific intensity measure (which is a conditional hazard in this case). Baker (2007) also showed that the cost of nonlinear structural dynamic analysis is reduced with the use of vector-valued intensity measures. Typically the case in conditional hazard analysis is that one primary intensity measure is conditioned upon another intensity measure. However, in conventional PSHA analysis Sa or PGA values are adopted, it may be the case that cumulative effects of ground motions is of interest and therefore \( CAV, I_a \) etc. may give important information about the structural response. For example, if nonlinear effects are of importance in the response of a structure for which the duration of ground motion is significant, the use of integral ground motion intensity measures such as CAV may be appropriate. Since cumulative effects are of primary interest in this scenario, another intensity measure that accounts for the instant effects of ground motion such as PGA may be conditioned upon the primary intensity measure, CAV. This kind of scenario may be seen at sites in which long duration but small and/or intermediate earthquakes occur, and the use of vector-valued PSHA analysis have been proposed to provide better estimates for the hazard (Baker and Cornell (2005); Bradley (2010)). As long as an IM sufficiently estimates the response of a structure (or ground motion response at a site), it is an appropriate IM. For more detailed discussion about various intensity measures and their
efficiency and sufficieny, see Riddell (2007) and Bradley et al. (2009).

In 1970s the basic form of PSHA, which evaluates the total hazard by using the total probability theorem, was established as the rate of exceedance of a specified ground motion intensity as

\[
\lambda(IM > im) = \sum_{i=1}^{N} v_i \int \int P(IM > im | m, r) f_{M,R}(m, r) \, dm \, dr
\]

where \( IM \) is intensity measure of earthquake ground motion and \( im \) is intensity measure that is exceeded. \( M \) and \( m \) are earthquake magnitudes, \( R \) and \( r \) distances (capital letters denote random variables of associated parameters). \( v_i \) is mean annual rate of occurrence of the smallest magnitude earthquakes considered, from the the \( i^{th} \) source. \( N \) is the number of sources contributing to the hazard. \( f_{M,R}(m, r) \) denotes the joint probability distribution of earthquake magnitude (\( M \)) and source-to-site distance (\( R \)), and \( P(IM > im | M, R) \), which obtained directly from GMPEs, is conditional probability of exceedance of a specified ground motion intensity for given \( m \) and \( r \). In the past, integration of equation (2.31) required closed-form solutions [Cornell (1968)]; however, the advent of computers allowed numerical integration techniques for seismic hazard integrals in equation (2.31), and also allowed one to consider more complex ground motion prediction equations [McGuire (2008)]. Modern PSHA method, however, is referred to as the Cornell-McGuire method [Cornell (1968), McGuire (2007)], and is shown to be in the following form:

\[
\lambda(IM > im) = \sum_{i=1}^{N} v_i \left\{ \int \int \int P(IM > im | m, r, \epsilon) f_{M,R,\epsilon}(m, r, \epsilon) \, dm \, dr \, d\epsilon \right\}_i
\]

where \( N \) denotes the number of seismic sources contributing to hazard, and \( \epsilon \) is
the number of logarithmic standard deviations that the logarithmic ground motion estimation deviates from the median. One of the primary advantages of PSHA is that it provides an understanding of major contributors to the total hazard with discretizing or factorizing equation (2.32) in terms of magnitude \( (M) \), source-to-site distance \( (R) \), and measure of the deviation of ground motion from its median \( (\epsilon) \). The resulting rate of exceedance values evaluated by integrating all contributions from various sources for each associated \( M, R, \) and \( \epsilon \) are called the deaggregation of PSHA (also called disaggregation of PSHA). The deaggregation provides the contribution of earthquake events in terms of their location and distance. The deaggregation is the joint PDF of \( m, r, \) and \( \epsilon \) conditioned upon exceedance of a specified significant intensity measure. This conditional PDF can be written as

\[
f_{M,R,\epsilon}(m,r,\epsilon \mid IM > im) = \frac{\sum_{i=1}^{N} v_i P(IM > im \mid m,r,\epsilon) f_{M,R,\epsilon}(m,r,\epsilon)}{\lambda(IM > im)} \tag{2.33}
\]

A more thorough discussion about different deaggregation techniques and their effects is examined in Bazzurro and Cornell (1999). In PSHA deaggregation, it is shown that the binning size of deaggregation, the variables used to discretize the total hazard, and the representation of distribution parameters (PMF versus PDF) in evaluation of equation (2.32) may significantly affect the hazard contributions. The binning size issue is briefly discussed in the following sections of this thesis.

Considering different models for predicting earthquake ground motion, and source, path and site parameters with their weightings in PSHA gives a suite of earthquake hazard curves. This allows one to incorporate epistemic uncertainties in PSHA. As a standard practice, logic tree methods are constructed to include epistemic uncertainties in PSHA applications [Kulkarni et al. (1984); Scherbaum et al. (2005)]. In logic trees, each branch represents a set of values of variables chosen for a seismic hazard model. Then, for each seismic hazard model, hazard calculations are performed and
a single hazard curve is obtained. The relative weighting of each hazard curve is determined by multiplying the weights in each of the branches. From this set of hazard curves, a mean (or median) curve can be obtained.

McGuire (1995) offered a new method for applying PSHA to obtain a design earthquake ground motion that accurately represents the uniform hazard spectrum from PSHA. The contribution of each source is evaluated separately by different ground motion prediction equations, then the seismic hazard is computed as a function of magnitude, source-to-site distance, and epsilon. As mentioned earlier, this allowed incorporating epistemic uncertainties into PSHA. Then, the hazard deaggregation is examined to see if one source or more than one source dominates the hazard. If there is only one dominant earthquake source for each natural frequency of ground motion, the most likely magnitude, source-to-site distance, and epsilon accurately represents the majority of the hazard. However, if there exist more than one dominant source (say, 2 dominant sources), hazard curves from each source are computed separately resulting in independent values of \( m \), \( r \), and \( \epsilon \) obtained for each dominant source. Then, from these weighted modal values, two separate ground motions are estimated to assemble the hazard accurately.
3 Development of New Correlation Coefficients Between PSA and PGV, CAV

3.1 Analysis Procedure

In order to find statistical relations between earthquake ground motion intensity measures, a suite of earthquake records are required. Firstly the ground motion dataset is determined by considering the availability of the ground motion prediction equations’ input parameters in the PEER-NGA ground motion database (Chiou et al. (2008)). For this purpose, four sets of earthquake records are scrutinized so that the dataset will be representative of temporal and spatial behavior of earthquakes over a broad range of input parameters (e.g. magnitude, distance etc.). Then, the method for computation of intensity measure relations is established based on the assumption that intensity measures are well represented by lognormal distributions as ground motion equations are modeled under the same assumption. Finally, empirical correlation coefficients are computed, and their parametric equations, derived by simple polynomial fits, are provided.
3.1.1 Methodology for Correlation Coefficient Computation

In this study, the correlation coefficients of PSA with PGV and CAV are of interest. Rewriting the general form of ground motion prediction equations ((2.26))

\[ \ln IM = \mu_{ln IM}(M, R, \Theta) + \sigma_{ln IM}(M, R, \Theta) \epsilon \]  

(3.1)

where \( M, R, \Theta, \) and \( \epsilon \) are random variables. Thence, equation (3.1) is a random function of the joint distribution of \( M, R, \Theta, \) and \( \epsilon. \) \( \mu_{ln IM_j} \) and \( \sigma_{IM_j} \) are the mean prediction of observed ground motion and standard deviation of it with its associated parameters, \( M_j, R_j, \) and \( \Theta_j. \) Here \( \epsilon \) (which is assumed to be standard gaussian) denotes the variability in the intensity measure. In simple words, \( \epsilon \) is a random variable that accounts for the number of logarithmic standard deviations that ground motion prediction, \( \ln IM, \) deviates from its mean.

Given measurements of an IM and their associated source, path and site parameters, and employing GMPEs, the gaussian residual \( \epsilon \) in equation (3.1) may be computed as

\[ \epsilon_j = \frac{\ln IM_j - \mu_{ln IM_j}(M_j, R_j, \Theta_j)}{\sigma_{IM_j}(M_j, R_j, \Theta_j)} \]  

(3.2)

where \( \ln IM_j \) is the \( j^{th} \) observation of ground motion. For each ground motion observation, this equation can then be computed easily. Cornell (2002) demonstrated that the correlation coefficients among intensity measures can be computed by utilizing the residuals \( (\epsilon_j) \) of each ground motion intensity measure. In other words, the correlation coefficients between intensity measures may be obtained by computing correlation coefficients of residuals of associated intensity measures. For this purpose, the Pearson product-moment correlation coefficient [Ang and Tang (2007)] is utilized in this study, and results are shown in section 3.1.2.
Correlation coefficients between intensity measures may be utilized to compute
conditional intensity measures and therefore conditional response spectra. Condi-
tional mean intensity measures and standard deviations, as discussed by Baker and
Cornell (2005) and Bradley (2010), may be written in this form

\[
\mu_{\ln IM_i | \ln IM_j} = \mu_{\ln IM_i} + \rho_{\ln IM_j, \ln IM_i} \sigma_{\ln IM_i} \epsilon_{\ln IM_j}
\]  

(3.3)

\[
\sigma_{\ln IM_i | \ln IM_j} = \sigma_{\ln IM_i} \sqrt{1 - \rho^2_{\ln IM_j, \ln IM_i}}
\]  

(3.4)

where \(\mu_{\ln IM_i | \ln IM_j}\) is the mean conditional intensity measure of \(\ln IM_i\) conditioned
upon \(\ln IM_j\), and \(\mu_{\ln IM_i}\) is the mean of logarithmic intensity measure computed by
using ground motion prediction equations. \(\rho_{\ln IM_j, \ln IM_i}\) is the correlation coefficient
between two intensity measures, which are obtained for Sa, PGV, and CAV in Chap-
ter 3. \(\sigma_{\ln IM_i}\) is the standard deviation of the unconditioned intensity measure, and
\(\sigma_{\ln IM_i | \ln IM_j}\) is the standard deviation of the mean conditional intensity measure. The
use of these formulations and applications of them will be illustrated in the following
sections in this study.

3.1.2 Discussion About the Ground Motion Database Considered and Regression
Results

In order to obtain an unbiased correlation coefficient, recorded real earthquake ob-
servations need to be selected carefully by considering the dataset used for the devel-
opment of ground motion prediction equations. The PEER-NGA database contains
3551 multi-component earthquake records from 173 active shallow crustal earth-
quakes for which the greater part come from the Western US and Taiwan. The key
difference in the NGA subsets from which each GMPE derived is the treatment of
aftershocks. The AS08 and CY08 models are fit using data from aftershocks, but
BA08 and CB08 models exclude aftershock data such as aftershocks of Chi-Chi-1999,
Northridge-1994, and Duzce-1999 earthquakes. The number of records used in fitting the AS08 and CY08 models is therefore greater than the BA08 and CB08 models. Almost one half of the records in the database represent aftershock of the Chi-Chi-1999 earthquake (1813 records), and it is believed that the inclusion that much of the earthquake recursion from one specific earthquake may lead to over-representation of this earthquake, and therefore it may result in decreased predictive capabilities for mainshocks. For these reasons, aftershocks of Chi-Chi-1999 and Northridge-1994 are not included in the dataset used for correlation coefficient computations. The effects of using different datasets are addressed in the following sections. For the identification of aftershocks and mainshocks that are required for the AS08 and CY08 models, the information is not provided in the NGA database; however, the AS08 provided a table that indicates which earthquakes are classified as mainshocks, aftershocks, foreshocks, and swarms [see Abrahamson and Silva (2008)]. Note that mainshocks, foreshocks, and swarms are categorized into one category in AS08, but this affects only a few records.

In the PEER-NGA dataset, not all input parameters required to evaluate the GMPEs are provided. So, according to the availability of GMPE input parameters in the database or estimability of them from empirical equations as described in Chapter 2, four different sets of earthquake records (which are subsets of the NGA database) are described and thoroughly scrutinized as follows,

Dataset-1: Excluding aftershocks of the Chi-Chi-1999 and Northridge-1994 earthquakes, 483 records were found to have all necessary input parameters (including $Z_{2.5}$, $R_{rup}$, $R_{jb}$, and $Z_{tor}$) for the CB-2010 $CAV_{GM}$ attenuation model (see Table 2.1). This sample of 483 records is used for correlation analysis. Note that if $Z_{tor}$ (the depth to the top of coseismic rupture) is not available for a record in the database, it may be taken as 1km. Indeed, for records in the NGA database, the average value of the depth to the top of rupture, $\overline{Z_{tor}}$, is approximately 4.5km. This value could be
used, however, changing $Z_{tor}$ from 1km to 4.5km does not affect the $CAV_{GM}$ prediction, as can be seen in Figures 3.1(a) and 3.1(b). Excluding records for which $Z_{2.5}$ values are not provided, would eliminate many important earthquake records such as Chi-Chi-1999 (1813 records), Coalinga-1983 (100 records) and Kocaeli-1999 (32 records).

Dataset-2: $R_{rup}$ and $R_{jb}$ parameters are important factors for prediction models, and if these parameters are missing for a record in the database, they could be estimated from existing empirical formulations. Estimation of $R_{rup}$ and $R_{jb}$ parameters from empirical relations is not preferred since doing so brings additional uncertainty into estimations. So, 2813 records for which $R_{rup}$ and $R_{jb}$ values are given in the NGA database are used in the dataset-2. In this subset, if $Z_{2.5}$ is not provided, its mean value $\overline{Z_{2.5}} = 2.2km$ is used instead. However, using the mean, $\overline{Z_{2.5}}$, may affect the correlation analysis. As can be seen in Figure 3.1(b), CAV increases significantly with $Z_{2.5} > 3km$. Ultimately, dataset-2 includes 2813 records in which $Z_{2.5}$ values are provided for 483 records, and for the remaining 2330 records the mean value, $\overline{Z_{2.5}} = 2.2km$, is used in the correlation computation.

Dataset-3: The number of records with both $R_{rup}$ and $R_{jb}$ parameters is 2813 in the database, as stated previously. If $Z_{2.5}$ is not provided in the database for any record, it is estimated from the CB08 model (equations (2.29) and (2.30)) which requires $Z_{1.0}$ or $Z_{1.5}$. If $Z_{1.0}$ or $Z_{1.5}$ is not provided in the database, the CY08 model (equation (2.28)) is utilized for estimation of $Z_{1.0}$.

Dataset-4: Using the CY08 model for $Z_{1.0}$ and the CB08 for $Z_{2.5}$ and excluding aftershocks of Chi-Chi-1999 and Northridge-1994 earthquakes, dataset-4 involves 1420 records. Note that earthquake records which are non-applicable since they do not meet the applicability criterias of GMPEs were excluded (see section 2.3) from datasets. See Table 3.1 for summary of these datasets.

In order to compare the use of different datasets for correlation coefficients, pre-
Table 3.1: Datasets utilized for correlation computation

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Aftershocks</th>
<th>$R_{rup}$, $R_{jb}$</th>
<th>$V_{S30}$</th>
<th>$Z_{2.5}$</th>
<th>N. of Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset-1</td>
<td>No</td>
<td>data only</td>
<td>data only</td>
<td>data only</td>
<td>483</td>
</tr>
<tr>
<td>Dataset-2</td>
<td>Yes</td>
<td>data only</td>
<td>data only</td>
<td>data and $Z_{2.5} = 2.2$ (km)</td>
<td>2813</td>
</tr>
<tr>
<td>Dataset-3</td>
<td>Yes</td>
<td>data only</td>
<td>data only</td>
<td>data and empirical equation</td>
<td>2813</td>
</tr>
<tr>
<td>Dataset-4</td>
<td>Chi-Chi-1999 and Northridge-1994</td>
<td>data only</td>
<td>data only</td>
<td>data and empirical equation</td>
<td>1420</td>
</tr>
</tbody>
</table>

Figure 3.1: $Z_{TOR}$ and $Z_{2.5}$ effects in dataset-1: (a) $Z_{TOR}$ and (b) $Z_{2.5}$ effects on $CAV_{GM}$.

Preliminary results for the correlation between $Sa$ and $CAV_{GM}$ are shown in figure 3.2 by only using the BA08 and CB10 GMPEs. Figures 3.2(a) and 3.2(b) show the effect of using tabulated values of $Z_{2.5}$ from the data and using the mean value of it $Z_{2.5} = 2.2$ km, which is actually used in dataset-2.

Results in Figure 3.2 show that the relation between $Sa$ and $CAV_{GM}$ is affected by the selection of data. In the results of dataset-1 (Figures 3.2(a) and 3.2(b)), $\rho_{lnSa(T),lnCAV_{GM}}$ varies between 0.5 to 0.2 with respect to the period range of (0.01-10) seconds. Specifically, in the 0.5 to 3.0 seconds period range, correlations from dataset-1 show a decreasing trend with increasing period from 0.46 to 0.25; conversely correlation coefficients from dataset-2 and dataset-3 are very similar to one another and increase from 0.5 to 0.6. Finally, correlation coefficients from dataset-4 decrease significantly from 0.6 to 0.4. Comparing results from datasets 2 and 3 shows that
Figure 3.2: Correlation results using the BA08 and CB10 GMPE models: (a) dataset-1: using data for $Z_{TOR}$, and if not provided, $Z_{TOR} = 1$ km; (b) dataset-2; (c) dataset-3; (d) dataset-4

Using an empirical formula for $Z_{2.5}$ or using its mean value does not have a major influence on the correlation coefficients. Moreover, inclusion (in datasets 2 and 3) or exclusion (in the datasets 1 and 4) of aftershocks have significant effects on the relation between $Sa$ and $CAV_{GM}$ (or similarly on the relation between $Sa$ and $PGV$). Including aftershock data in the analysis increases correlations $\rho_{lnSa(T),lnCAV_{GM}}$ at periods greater than 0.5 seconds.

Using a set of ground motion prediction equations incorporates epistemic uncertainties into correlation results. Also, examining the effects of using different ground motion prediction equations on correlation coefficients may help in determining which
dataset is more appropriate for the correlation analysis. For that purpose, correlation coefficients are firstly computed by using the CB08 and CB10 GMPE models, and results are compared with previously obtained correlation coefficients (by using the BA08 and CB10 models). This comparison is shown in figures 3.3 and 3.4. As can be seen from these figures, correlations computed from the CB08 are somewhat higher than those for the BA08 at shorter periods ($T < 0.5$ sec). Otherwise, they are similar.

To briefly summarize the procedure adopted so far: four different subsets of the
The purpose here is to attempt to define an appropriate database for computing correlation coefficients between intensity measures $S_a$, $CAV_{GM}$, and $PGV$. We seek to identify the dataset for which $CAV-S_a(T)$ correlations are least affected by the GMPE selected. As can be seen in figures 3.3 and 3.4, correlations $\rho_{\ln S_a(T), \ln CAV_{GM}}$ using the BA08 and CB10 GMPEs are closer to those using the CB08 and CB10 GMPEs with data from dataset-4. In other words, correlations computed from dataset-4 affected...
less by the choice of the GMPEs considered than the other datasets considered.

3.2 Correlation Coefficients of Spectral Acceleration to Peak Ground Acceleration and Cumulative Absolute Velocity, and Examples of the Conditional Mean Spectra

In order to compute the conditional spectral acceleration conditioned upon a peak ground velocity value, the statistical relation between $S_a$ and $PGV$ is necessary. Correlation coefficients, $\rho_{\ln S_a, \ln PGV}$ and $\rho_{\ln S_a(T), \ln CAV_{GM}}$ are obtained by using a subset of PEER-NGA database: the 1420 records from dataset-4 as described in the previous subsection. Empirical ground motion prediction equations BA08, CB08, AS08, and CY08 are employed in a correlation analysis between $S_a$ and $PGV$. Correlation coefficients $\rho_{\ln S_a(T), \ln PGV}$ with respect to spectral acceleration periods are computed by taking the mean of the four correlation coefficient curves obtained by using four attenuation relationship equations.

As can be seen in figure 3.5 the correlation between $S_a$ and $PGV$ varies between 0.5-0.8 with respect to the spectral acceleration period. Correlation coefficients with four GMPEs have generally a similar trend, and in the 0.5-3.0 seconds period range $S_a$ and $PGV$ have a high correlation ($\approx 0.7-0.8$). In the short period range, however, correlation coefficients are more sensitive to the ground motion prediction equation selected. The CY08 model gives lower correlations than other GMPEs. The variability of correlation coefficients and their relative deviations can also be seen in figure 3.5(c). Mean and 90% confidence intervals are provided in figure 3.5(b). The mean correlation curve lies within 90% percentiles of all GMPEs, except the CB08 in the short period range. Finally, in figure 3.5(d), a simple polynomial fit of the mean correlation curve is given as a piecewise continuous function. This polynomial equation for computation of empirical correlation coefficients between $S_a$ and $PGV$ is given in equation (3.5).
Figure 3.5: Correlation coefficients between $Sa$ and $PGV$: (a) results obtained by using BA08, CB08, AS08, and CY08 GMPE models, (b) Results with mean and 90% confidence intervals, (c) box plots representing the distribution of uncertainties of each GMPE, and (d) Simple polynomial fit of the mean correlation curve

$$\rho_{\ln Sa, \ln PGV} = \begin{cases} 
-0.047 (\ln T)^2 - 0.4041 \ln T - 0.192 & \text{for } T \leq 0.12 \text{ (s)} \\
-0.0646 (\ln T)^3 - 0.204 (\ln T)^2 + 0.0078 \ln T + 0.782 & \text{for } 0.12 < T \leq 1.0 \text{ (s)} \\
0.0463 (\ln T)^3 - 0.18 (\ln T)^2 + 0.0905 \ln T + 0.777 & \text{for } 1.0 < T \leq 10.0 \text{ (s)} 
\end{cases}$$ (3.5)

To illustrate the conditional mean spectrum, an example with parameters $M = 6.5$, $R_{rup} = 25$ km, and exceedance level of peak ground velocity $PGV^\star = 22.5$ cm/s is presented in figure 3.6. 3.6(a) shows the conditional mean spectrum computed by using the correlation coefficients developed. As can be seen, four GMPEs have a similar trend in the moderate to long period range, but have differences in the short
Figure 3.6: An example of the conditional mean spectrum conditioned upon \(PGV^* = 22.5\) cm/sec and \(M = 6.5\), \(R_{rup} = 25\) km. (a) The conditional acceleration spectra obtained by using BA08, CB08, AS08, and CY08 GMPE models, (b) The same conditional acceleration spectra with samples of 13 response spectra from real earthquake records selected from the database.

period range. Figure 3.6(b) also shows the same conditional spectral acceleration conditioned upon \(PGV^* = 22.5\) cm/s with 13 real earthquake records which posses similar parameters as assumed (\(M \approx 6.5\), \(R_{rup} \approx 25\) (km), and \(PGV^* \approx 22.5\) cm/s). This figure indicates a positive bias in the predicted conditional spectral acceleration in the long-period range (> 4 sec), with respect to the 13 records chosen.

Similarly, for conditional spectral acceleration conditioned upon cumulative absolute velocity, the correlation between \(Sa\) and \(CAV_{GM}\) is necessary. The empirical ground motion prediction equations BA08, CB08, AS08, and CY08 are utilized to compute \(Sa\) and the CB10 model is used to compute \(CAV_{GM}\) by using the dataset-4. By employing these five GMPEs, correlation coefficients, \(\rho_{lnSa,lnCAV_{GM}}\), with respect to spectral acceleration period are computed by taking the mean of the four correlation coefficient curves obtained.

As can be seen in figure 3.7, there is only moderate correlation between \(Sa\) and \(CAV_{GM}\) ranging \(\approx 0.4 - 0.5\). Correlation coefficients have generally a similar trend in each pair of GMPEs (e.g, correlations obtained by AS08-CB10 are similar to the results obtained by BA08-CB10 models) used in 0.5-10.0 seconds period range. In the
short period the CB08 GMPE model gives higher correlation than the other models. The variability of correlation coefficients and relative deviations can also be seen in figure 3.7(c). Mean and the 90% confidence intervals are provided in figure 3.7(b).

The mean correlation curve lies within the 90% percentiles for all GMPEs. Finally, in figure 3.7(d) simple polynomial fit of the mean correlation curve is derived as a piecewise continuous function, and is given in equation (3.6).

\[
\rho_{ln \text{Sa}, ln \text{CAV}} = \begin{cases} 
0.0169 (\ln T)^3 + 0.143 (\ln T)^2 + 0.3287 \ln T + 0.617 & \text{for } T \leq 0.2 \ (s) \\
0.0057 (\ln T)^3 - 0.044 (\ln T)^2 - 0.0266 \ln T + 0.489 & \text{for } 0.2 < T \leq 10.0 \ (s)
\end{cases}
\]
Figure 3.8: An example of the conditional mean spectrum conditioned upon $CAV_{GM}^* = 0.5 \text{ g-s}$ and $M = 6.5$, $R_{rup} = 25 \text{ km}$. (a) Conditional acceleration spectra obtained by using BA08, CB08, AS08, and CY08 GMPE models, (b) Conditional acceleration spectra with samples of 21 response spectra from real earthquake records selected from the database.

An example with parameters $M = 6.5$, $R_{rup} = 25 \text{ km}$, and the exceedance level of cumulative absolute velocity $CAV_{GM}^* = 0.5 \text{ g-s}$ is illustrated in figure 3.8. Figure 3.8(a) shows the conditional mean spectrum obtained using the correlation coefficients developed. Figure 3.8(b) shows the same conditional spectral acceleration conditioned upon $CAV_{GM}^* = 0.5 \text{ g-s}$ with 21 real earthquake records which possess similar parameters as assumed ($M \approx 6.5$, $R_{rup} \approx 25 \text{ km}$, and $CAV_{GM}^* \approx 0.5 \text{ g-s}$). This figure illustrates that the conditional mean spectrum predicts the spectra from actual records well.

Correlations between $Sa$ and PGV are higher ($\approx 0.65$) than correlations between $Sa$ and $CAV_{GM}$ ($\approx 0.45$) (figures 3.5 and 3.7). This is supported by the dispersion of $Sa$ of records selected for PGV (figure 3.6(b)) as compared to the dispersion of $Sa$ of records selected for CAV (figure 3.8(b)).

The correlation coefficients between $Sa$ and PGV and between $Sa$ and CAV presented here are very similar to those recently presented by Bradley (Bradley (2012a) and Bradley (2012b)). This thesis addresses the importance of the selection of the ground motion database used in computing these correlations.
Deaggregation (or disaggregation) of seismic hazard, as described in section 2.4, is useful for identification of earthquake scenarios. Deaggregation accounts for the distribution of the seismic hazard with respect to uncertainty in site, path, and source parameters. In deaggregation analysis, the hazard is spatially factorized with respect to significant parameters of a PSHA such as magnitude, distance, and $\epsilon$ with a certain binning size of factorization. So, the deaggregation reveals the relative contributions of various earthquakes to the prescribed exceedance level of a significant ground motion intensity measure, which is in turn related to the hazard level. Results of deaggregation and its interpretation is important. The mean (expected) and mode of deaggregation parameters are often employed in earthquake engineering applications. In a case where there exist multiple earthquake sources with distinct characteristics, the seismic hazard may originate from different potential earthquake ruptures. As a result, the distribution of potential causal earthquakes may be observed.
as a bimodal or multimodal distribution in deaggregation simulation (see figure 4.1). In such a case, the use of mean values of variates or a single mode of deaggregation may not be sufficient to capture all aspects of the hazard, and therefore may falsify the design earthquake selection. To prevent this problem in hazard analysis, it has been proposed that in areas where one single source dominates the seismic hazard for exceedance of a specified intensity measure level with creating a unimodal distribution of hazard, identifying one single design earthquake with respect to mean and/or mode of the deaggregation may be a sufficient estimation; however, in case the hazard is dominated by multiple potential earthquake scenarios (with multiple mountains in the distribution of hazard - in a deaggregation plot), the hazard may better be reflected by identifying a few design earthquakes which then sum to multiple causal earthquake scenarios pointed by mean or modal values of each dominant source [McGuire (1995), Bazzurro and Cornell (1999)]. For these reasons, it may be preferable to deaggregate the seismic hazard for each seismic source, and handle each source individually.

For an illustration of the above discussion about the hazard deaggregation, a hypothetical example is shown in figure 4.1 which displays the joint PDF of magnitude

![Figure 4.1: Illustration of seismic hazard deaggregation with bimodal distribution.](image)
and distance (usually $\epsilon$ as well) conditioned upon the exceedance of a prescribed intensity measure level. In this example, input parameters are purposely chosen to yield significantly different causal earthquake ruptures that contribute to the total hazard. As can be seen, two earthquake scenarios dominate the hazard, resulting in a bivariate, bimodal probability distribution. It can be observed that one rupture scenario is approximately from magnitude 6 and distance 10 km, and the other is approximately from magnitude 7.5 and distance 60 km. If mean values of the variants are used for future potential threats, one would approximately get magnitude 6.75 and distance 40 km. The first nearby scenario is capable of generating smaller earthquakes in size, and another rupture scenario further away generating mainly characteristic earthquakes of larger size. The use of these mean magnitude 6.75 and distance 40 values would not be representative of the true hazard because neither of these mean metrics points to locations of the two major earthquake sources and their potential sizes. For these reasons, a pair of modal magnitude and distance values for each dominant source may be utilized to estimate the potential hazard. Or as mentioned, if the seismic zonation and characteristics of sources surrounding a given site cause multimodal hazard distribution, each of the single source deaggregations may be computed individually, and a pair of mean or mode of magnitude and distance may be adopted for estimation of the hazard. It is worth noting that a hazard deaggregation is drawn conditional upon an exceedance of certain aggregated hazard level (e.g. 500-year return period) which in turn represents the exceedance of an intensity measure (e.g. $CAV_{GM}^* = 0.5$ g-sec). So, exceedance of a specific intensity measure from each individually deaggregated source does not necessarily correspond to the same exceedance rate. Similar observations can be made in cases where the seismic hazard is dominated by multiple potential sources at different spectral periods. The frequency content of ground motions conditioned on exceedance of short period spectral acceleration attenuates quickly with distance, and therefore the haz-
ard is dominated by small-size and nearby events. Similarly, in the exceedance of long period spectral acceleration, larger-size (rarer) events (which could be characteristic events) dominate the hazard. Again, the use of mean values or modes of the entire deaggregation is questionable for seismic scenarios with multiple sources. A similar scenario is thoroughly examined by McGuire (1995) and concluded that the use of a set of design ground motions in areas with two dominant sources near the site of interest, rather than using one single design ground motion, a set of design ground motions for short and long period ranges may be appropriate to capture the entire seismic hazard.

To identify potential future ground motions, standard statistics of the deaggregation distribution such as mean and mode as described above may be utilized, but the use of such fixed statistics for describing a stochastic problem may be criticized regarding the loss of information in the distribution of the hazard. A sampling strategy can be used to incorporate the complete deaggregation information.

Monte Carlo simulation is a widely used approach for incorporating more information in more sophisticated hazard analysis, and Monte Carlo simulations have been exploited in PSHA for handling epistemic uncertainties by sampling random variables which are incorporated into the branches of a logic tree method. The logic tree method [Cramer et al. (1996), Bommer et al. (2005)] is a widely used approach in PSHA and a powerful tool to capture the epistemic uncertainties such as uncertainties associated with the various assumptions of spatial and temporal regime of earthquake occurrences and the ground-motion prediction models. Through the use of the Monte Carlo simulation method, uncertainties can be propagated by sampling the distribution of input parameters. Musson (2000) examined the use of the Monte Carlo method in seismic hazard analysis and made a comparison of conventional PSHA and the Monte-Carlo method of seismic hazard analysis. PSHA may also be performed by combining logic trees with the Monte Carlo method [Cramer
et al. (1996), Secanell et al. (2008)]. Each branch of the logic tree accounts for an alternative assumption or hypothesis that is sampled from its associated probability distribution rather than assuming a single point estimate. For example, four competing GMPEs may be adopted in a logic tree, each along with a branch of the tree so that all four will contribute to the total hazard. Instead of using the mean or median ground motion predictions, Monte Carlo simulation can be exploited to sample from each GMPE. As a result, each branch of the logic tree representing a potential scenario of earthquake occurrence consistent with the past behavior of a given site yields a hazard curve for the predefined exceedance of hazard level (such as spectral acceleration of 475-year, 1000-year, 5000-year etc. return period). Monte Carlo simulation is also an alternate method for estimating site ground motions by sampling the hazard deaggregation, which is a joint probability density function (PDF) of parameters (usually) magnitude, distance, and $\epsilon$.

An example of the annual frequency of exceedance of various hazard levels (namely the hazard curve) and the hazard deaggregations conditioned upon exceedance of 1000-year peak ground velocity ($PGV_{1000}$) utilizing the AS08, BA08, CB08, and CY08 GMPE models and the mean of these four deaggregations are illustrated in
Figure 4.3: (a) An example of hazard curve: Annual frequency of exceedance of various hazard levels; and Hazard deaggregation conditioned on exceedance of 1000-year peak ground velocity (PGV\(_{1000}\)) for the (b) AS08, (c) BA08, (d) CB08, (e) CY08 GMPE models, and (f) Mean Deaggregation from these four GMPE models.

Figure 4.3 for a scenario illustrated in figure 4.2. Figure 4.3(a) illustrates the hazard curve obtained by adopting the bounded G-R magnitude recurrence relation for the temporal occurrence of earthquakes along with the four GMPE models for PGV. Asterisks on the hazard curve denote 100-year, 500-year, 1000-year, 2500-year, and 10000-year hazard levels, respectively. The contribution of the source to a specified exceedance level of hazard may also be observed in the four-dimensional deaggregation plots. As can be seen, the most likely seismic threat for the 1000-year event
Figure 4.4: Hazard curves (mean of deaggregations obtained by four GMPEs) for: (a) 1 km, (b) 2 km, (c) 5 km, and (d) 10 km distance binning sizes for exceedance of 1000-year peak ground velocity ($PGV_{1000}^*$).

comes approximately from magnitude 6.5 events with a source-to-site distance of 15 km. The frequency content of median size earthquakes ($M = 6 - 7$) from this source is larger, so the major contribution to the hazard comes from nearby and median-size earthquakes. The binning size of distance, magnitude, and $\epsilon$, respectively, chosen 5 km, 0.25, and 1 for this example.

The resolution (bin size) of deaggregation affects the deaggregation distribution, so binning sizes and reasons for choosing them should be reported [Bazzurro and Cornell (1999)]. To discover the effect of adopting different binning schemes on conditional spectral acceleration, a suite of binning sizes was chosen and deaggregation plots were generated using the bounded G-R magnitude recurrence model for the same example illustrated in figure 4.2. Along with the Monte Carlo simulation method, each deaggregation is sampled by yielding pseudo-random numbers for use
of conditional ground motion estimation. Hazard deaggregations (the mean of deaggregations obtained by using the four GMPEs) are drawn in figure 4.4 for 1, 2, 5, and 10 km distance binning sizes and annual frequency of exceedance of $PGV_{1000}^*$. Similarly, for the same binning sizes, deaggregations are plotted for annual frequency of exceedance of $PGV_{100}^*$ in figure 4.5. As can be seen from these figures, the binning size of distance slightly affects the distribution of the deaggregation hazard. Comparison of deaggregation plots for varying binning size of distance is difficult here, so the effect of use of different binning size may be better observed in the resulting acceleration spectra. The conditional spectral acceleration conditioned upon 100-year and 1000-year peak ground velocity ($PGV_{100}^*$ and $PGV_{1000}^*$) along with the G-R source magnitude recurrence model are computed in figure 4.6. Potential earthquake scenarios are sampled from hazard deaggregations of 1, 2, 5, and 10 km distance binning sizes. Using the samples of potential earthquake scenarios represented by distance, magnitude and epsilon, the conditional mean spectra are computed by exploiting the newly-developed correlation coefficients in Chapter 3.

To see the effects of distance binning size, the conditional spectra for binning size of 1, 2, 5, and 10 km are shown in figure 4.6 in which the bounded G-R model is adopted. These conditional mean spectra are computed for 100-year and 1000-year hazard levels. As can be seen from these figures, the binning size of the deaggregation remarkably affects the conditional spectral acceleration for both conditioning hazard levels (peak ground velocity of 100-year and 1000-year return period). It can be observed from comparing the exceedance of $PGV_{100}^*$ and $PGV_{1000}^*$ levels (in figures 4.4 and 4.5) that the deaggregation of exceedance of larger peak ground velocity has smaller dispersion because few earthquakes can cause the peak ground velocity to exceed a large threshold value (or exceedance level). Similarly, many earthquakes can cause the peak ground velocity to exceed a smaller value of an exceedance level, therefore the deaggregation of exceedance of smaller peak ground velocity has larger
Figure 4.5: Hazard curves (mean of deaggregations obtained by four GMPEs) for: (a) 1 km, (b) 2 km, (c) 5 km, and (d) 10 km distance binning sizes for exceedance of 100-year peak ground velocity ($PGV_{100}$).

dispersion. For increasing distance binning size, the conditional spectral acceleration has an increasing trend. This is an interesting observation because the effect of binning size is noticeable.

Binning sizes of distance and magnitude are hereafter taken as 5km and 0.25, respectively. Hazard curves can be obtained from one ground motion prediction equation or by combining two or more. The variability of the conditional mean spectra with respect to ground motion prediction models may be investigated by combining various GMPEs (AS08, BA08, CB08, CY08). Hazard curves for PGV derived from four GMPEs are illustrated in figure 4.3(a).

As can be seen, there is not a significant difference among hazard curves with different GMPE models, therefore the mean of these four curves would be a sufficient estimate for exceedance of 100, 500, 1000, 2500, and 10000 year peak ground velocity,
Figure 4.6: Conditional mean spectra: (a) Conditioned upon ($PGV_{100}^*$) and (b) Conditioned upon ($PGV_{1000}^*$).

denoted as asterisks in figure 4.3(a). By deaggregating these hazard curves individually and conditioning them upon the annual frequency exceedance of the 100-year hazard level, the joint probability distribution of potential earthquake scenarios are demonstrated in figures 4.7 and 4.8. Comparing these figures again shows only very slight differences with respect to the GMPE model utilized. Once the deaggregation
of hazard for individual GMPE models is obtained, with the use of correlation coefficients $\rho_{\text{Sa}, \text{PGV}}$, the conditional acceleration spectrum conditioned upon $\text{PGV}_{100}^*$ may be computed by adopting one or more of the AS08, BA08, CB08, and CY08 ground motion prediction models. Figure 4.9 illustrates this conditional response spectrum and the average of these four spectra. As can be seen, in very low and high frequency ranges, there exist some variation on the response with respect to GMPEs used, but this difference is insignificant from any practical point of view. In the moderate frequency range, however, a relatively smaller deviation from the mean ground motion response exists. Of particular note for this figure is that, knowing the fact that both epistemic and aleatoric uncertainties are involved in this conditional mean spectrum, epistemic uncertainties can easily be observed by comparing the variation of ground motion response in very short and long period ranges to the moderate period range with respect to various ground motion prediction models. It is worth noting at this point that these response spectra have smaller standard deviations since they are conditioned upon exceedance of peak ground velocity, so the total uncertainty in the conditional spectrum is reduced (see equation (3.4)).

So far, conditional response spectra have been illustrated using four different ground motion prediction models along with bounded Gutenberg-Richter magnitude recurrence model. The conditioning intensity measure is chosen to be the peak ground velocity and for conditional response spectra the correlations between $\text{Sa}$ and $\text{PGV}$ developed in Chapter 3 are utilized. The effect of using different recurrence relations (such as the Bounded Gutenberg-Richter and the Characteristic magnitude models, which were discussed earlier in Chapter 2) may also be of interest. Seismological and geological observations of earthquake occurrences at different seismic locations have shown that large earthquakes from a particular fault might occur with frequencies larger than the exponential G-R model would predict. This idea has led to a scientific debate on which model is more appropriate for a given seismicity [Wes-
Figure 4.7: Hazard deaggregation conditioned on exceedance of 100-year peak ground velocity ($PGV_{100}$) along with the Bounded Gutenberg-Richter recurrence model: (a) with AS08 GMPE model, (b) with BA08 GMPE model
Figure 4.8: Hazard deaggregation conditioned on exceedance of 100-year peak ground velocity \((PGV_{100})\) along with the Bounded Gutenberg-Richter recurrence model: (a) with CB08 GMPE model, and (b) with CY08 GMPE model.
Figure 4.9: Conditional response spectrum conditioned on $(PGV_{100}^*)$ along with the Bounded Gutenberg-Richter recurrence model and GMPE models: AS08, BA08, CB08, and CY08 (color figure can be seen in the online version).

The idea of periodicity of characteristic earthquakes is stimulated by geologi-
cal observations of recurrent behavior of paleoearthquakes and the similar slip rates caused by them. Geologic evidence indicates that slip in large earthquakes is repeatedly released on a plane or across a narrow zone; however, small and moderate size earthquakes are typically generated from sources off of the main fault. For example, Stirling et al. (1996) examined 22 strike-slip faults from California, China, Japan, Mexico, New Zealand, and Turkey, then concluded that 18 of the faults are consistent with the C-H model. Wesnousky (1994) examined paleoseismic occurrences between 1944 and 1992 in Garlock, Elsinore, Newport-Inglewood, San Andreas, and San Jacinto faults and concluded that the San Jacinto fault is consistent with a G-R model and others fit the C-H model. Extrapolation of small and moderate size earthquakes to characteristic earthquakes, which are represented by the successive occurrences of similar-displacement (slip rate) events, yields a discrepancy between small-to-moderate earthquakes and large earthquake regimes. On the contrary, Page et al. (2011) posits that there is an interaction between small-to-moderate earthquake occurrences and large magnitude earthquakes. Ruptures of smaller faults may trigger earthquakes on nearby major faults, and these smaller faults have characteristic behavior at smaller magnitudes; therefore, these small earthquakes and their statistics are relevant to the occurrence of large earthquakes. So, considering the whole picture with the inclusion of main and secondary nearby faults causing a fault to rupture may be a counter indication for the characteristic earthquake model.

Youngs and Coppersmith (1985) proposed, for individual seismic sources or fault segments, that characteristic magnitude recurrence model with a limited size range of large earthquakes may be favored. This earthquake size range varies from fault to fault, so for some faults this range is narrower and for some faults it is broader depending on the paleoseismic capacity of the site of interest. The magnitude distribution of both the G-R and C-H models computed from equations (2.20) and (2.22) are illustrated in figure 4.10. This figure shows the relative frequency of strong earth-
Figure 4.10: Probability density function of earthquake magnitudes with the G-R and C-H models.

Quakes between magnitude 6.5 and 7.0 in a C-H recurrence model. The frequency content of large or strong earthquakes is directly controlled by the parameter $p$ (see equation (2.22)), which is the area between magnitude 6.5 and 7.0 in figure 4.10. For increasing value of $p$, larger earthquake occurrences and therefore larger ground responses are expected.

For application of another correlation between $Sa$ and $CAV_{GM}$ developed in Chapter 3, and for evaluating the effects of favoring the C-H model rather than G-R model, the mean conditional response spectrum conditioned upon $CAV$ along with the G-R and C-H models are computed by following the steps illustrated so far in this study. The mean conditional response spectrum conditioned upon a particular value of peak ground velocity along with the G-R model was illustrated in figure 4.9. Similarly, the conditional response spectrum conditioned upon a particular value of cumulative absolute velocity will be illustrated by utilizing the G-R and C-H magnitude recurrence models. For that purpose, an example site with a
Figure 4.11: Probability density function of earthquake magnitudes along with the G-R and C-H models with various $p$ values (color figure can be seen in the online version).

Figure 4.12: Hazard curves for cumulative absolute velocity based on the G-R and C-H models for the example site (color figure can be seen in the online version).
single fault is assumed. This fault produces earthquakes with minimum magnitude, \( M_{\text{min}} = 4 \), maximum magnitude, \( M_{\text{max}} = 7 \), closest distance, \( R = 15 \) km, and the length of the fault is 100 km (horizontal projection of the location of site to the fault assumed to be in the middle of the fault) as shown in figure 4.4. With the Gutenberg-Richter exponential model (equation (2.18)), earthquakes on this single fault have an annual minimum magnitude occurrence rate \( \lambda_{\text{min}} = 0.1 \). This event is assumed to have vertical strike-slip faulting mechanism, and the site is considered to be NEHRP site class of \( B/C \) with shear wave velocity \( V_{S30} = 760 \) m/s. The probability density function for earthquake magnitudes for the example site, using the G-R and C-H recurrence laws, is illustrated for various \( p \) values in figure 4.11 (y axis in logarithmic scale). As can be easily seen from this figure, the parameter \( p \) accounting for paleoearthquake frequency content in the magnitude distribution is strictly connected with the relative frequency of strong earthquakes. Note that for increasing values of parameter \( p \), the step height (for large earthquakes, magnitude 6.5-7.0) in figure 4.11 increases (dotted lines).

Hazard curves for the example site described with its seismic characteristics are computed by adopting both the G-R and C-H recurrence models for various \( p \) parameters. It is worth to restate here again that these hazard curves present an annual rate of exceedance of cumulative absolute value (or denoted as \( CAV_{GM} \)) obtained by using CB10 (Campbell and Bozorgnia (2010)) ground motion prediction equation. Hazard curves are plotted in figure 4.12. These hazard curves may then be deaggregated individually. By using the Monte Carlo simulation method described earlier, the deaggregation of the seismic hazard for each curve may be sampled in order to compute the conditional mean spectrum. Note that correlation coefficients developed in Chapter 3 and the AS08, BA08, CB08, CY08, CB10 ground motion prediction models are utilized for these conditional mean spectra.

The conditional mean spectrum is plotted in figure 4.13 for 1000-year cumula-
Figure 4.13: Conditional Mean Spectra given $CAV_{1000}$ along with the G-R and C-H models for the example site (color figure can be seen in the online version).

This figure shows how sensitive the ground motion response is to magnitude recurrence relations (the Gutenberg-Richter and Characteristic Magnitude models), and varying probability of characteristic earthquakes in the C-H model. The parameters for both the G-R and C-H models are the same except the parameter $p$, which accounts for the characteristic earthquake probability. As can be seen, there exists a strong relation between response spectra and $p$. The conditional mean spectra are scaled up remarkably for increasing values of $p$. For example around 0.1-0.5 second natural period, the conditional spectral acceleration increases with approximately a factor of 2 (from $\approx 0.2g$ to $0.4g$). The parameter $p$ essentially depends on the nature of the fault and temporal cyclic behavior of potential large earthquakes. If the parameter $p$ is very small, indicating that characteristic earthquake behavior is not really observed, there is not much of an effect on the conditional response spectra obtained by the G-R and C-H models.
Probabilistic seismic hazard analysis is an evolving field in which an analyst selects among many modeling alternatives, such as the modeling of regional seismic activity, the selection of a ground motion attenuation model, the selection of records from ground motion databases, and the level of resolution involved in an analysis. This thesis examines the effects of such decisions on the resulting probabilistic description of infrastructure responses. Specifically, the probabilistic response of infrastructure (as described by spectral accelerations conditioned upon peak ground velocity or cumulative absolute acceleration) to ruptures of a single fault are examined in this thesis.

Determining the seismic hazard to infrastructure requires a model for the return period of earthquakes exceeding specified magnitude levels, that is, the magnitude recurrence relation. There remains scientific debate regarding the suitability of the bounded Gutenberg-Richter and Characteristic Magnitude recurrence models. The Characteristic Magnitude model is generally thought to better reflect the seismicity generated by large faults or fault segments, while the bounded Gutenberg-Richter model is thought to be more suitable for distributed seismicity. Results presented
in this thesis indicate that conditional spectral accelerations are remarkably affected by the choice of the earthquake magnitude recurrence model and are highly sensitive to the probability of characteristic earthquakes in the Characteristic Magnitude recurrence model.

The return period of a specific seismic hazard at a building site is described by the exceedance rate of a particular ground motion intensity measure. For example, the “1000 year earthquake” could be represented by earthquakes with a peak ground velocity (PGV) exceeding a specified threshold value, such as 50 cm/s. Such ground motions may arise from ruptures of various magnitudes and distances, perhaps on distinctly separate faults. The seismic sources contributing to a specified hazard (e.g., PGV greater than 50 cm/s) are described by a distribution function of the relative contribution of sources to the hazard as a function of the magnitude and distance to the source. This is called the hazard deaggregation. Hazard deaggregation is usually computed numerically as a probability mass function, for a specified set of discrete distance and magnitude values. The discretezation of distance and magnitude of the deaggregation (that is, the binning sizes) is found to remarkably affect conditional spectral accelerations. Conditional spectral accelerations increase noticeably with increasing distance bin size.

In order to identify the tectonic sources of potential future ground motions, standard statistics of the deaggregation distribution such as its mean and mode may be utilized. However, if multiple sources contribute roughly equally to the seismic hazard, the use of the mean of the deaggregation distribution will not reflect the true nature of the seismic environment. In this study, a sampling strategy is used to capture the actual seismic environment and its impact on infrastructure.

The computation of spectral accelerations (Sa(T)) conditioned upon peak ground velocity (PGV) or cumulative absolute velocity (CAV) requires period-dependent correlation coefficients between Sa(T) and PGV or Sa(T) and CAV. This study delivers
newly-developed correlation coefficients between $\text{Sa}(T)$ and PGV and between $\text{Sa}(T)$ and CAV. The sensitivity of these correlations to the earthquake dataset from which they are computed is determined by examining results from four overlapping subsets of PEER-NGA database. The inclusion or exclusion of aftershocks significantly affects these correlations. The subset which excluded afterhocks of the 1994 Northridge earthquake and aftershocks of the 1999 ChiChi earthquake was ultimately selected as providing correlations that were least affected by the ground motion attenuation model used. This subset consisted of 1420 ground motion records from active shallow crustal earthquakes. Including aftershock data in the analysis increases correlations $\rho_{\ln \text{Sa}(T) \ln \text{CAV}_{GM}}$ at periods greater than 0.5 seconds. It is found that the correlations between $\text{Sa}(T)$ and PGV are higher ($\approx 0.65$) than correlations between $\text{Sa}(T)$ and CAV ($\approx 0.45$). This is supported by the dispersion of $\text{Sa}(T)$ from records selected for PGV as compared to the dispersion of $\text{Sa}(T)$ from records selected for CAV. By computing correlations from a ground motion dataset for which correlations were relatively insensitive to the ground motion attenuation relationship used, the ensuing conditional spectra computed from different attenuation models were also similar to one another.
Bibliography


