

CALIFORNIA INSTITUTE OF TECHNOLOGY

EARTHQUAKE ENGINEERING RESEARCH LABORATORY

EARTHQUAKE GROUND MOTION SIMULATION  
USING NOVEL MACHINE LEARNING TOOLS

BY

ARZHANG ALIMORADI

REPORT No. EERL 2011-01

PASADENA, CALIFORNIA

MAY 2011



A RESEARCH REPORT FOR THE EARTHQUAKE HAZARDS REDUCTION PROFESSIONAL FELLOWSHIP PROGRAM SUPPORTED BY THE EARTHQUAKE ENGINEERING RESEARCH INSTITUTE, THE FEDERAL EMERGENCY MANAGEMENT AGENCY, AND THE CALIFORNIA INSTITUTE OF TECHNOLOGY  
UNDER SUPERVISION OF PROFESSOR JAMES L. BECK

© 2011  
ARZHANG ALIMORADI, PH.D., P.E.  
ALL RIGHTS RESERVED

# Acknowledgements

My fellowship at Caltech has been, undoubtedly, one of the greatest professional experiences I have ever had. I am indebted to Professor James Beck for his constant support and wisdom during this project. Our long Friday afternoon discussions will remain amongst my best memories. Professor Wilfred Iwan, a good friend and a mentor, was instrumental in making this fellowship happen, for this and for everything else I have learned from him I am very grateful. EERI-FEMA professional fellowship program provided financial support for the completion of this project and their support is much appreciated. Interaction with faculty, students, and staff at Caltech's Thomas Lab has been a pleasant part of my fellowship and something that I will dearly miss upon my departure. Professor Beck's research group at Caltech, particularly Dr. Konstantin Zuev and Mr. Stephen Wu, provided much insight for which I am very thankful. Caltech's Center for Advanced Computing Research (CACR) was instrumental in providing much needed computational resources for some of the heavier computational tasks involved in this research. Their enthusiastic assistance is very much appreciated. In particular, I should thank

CACR's Operations Manager, Ms. Sharon Brunett, Mr. Jan Lindheim and CACR Operations for providing administrative and technical support (account creation, system stability). The work reported in this volume had to rely on recorded data, which were provided by the PEER-NGA program, Dr. Yousef Bozorgnia of UC-Berkeley, and by Professor Masumi Yamada of Kyoto University. I am grateful to both Masumi and Yousef for sharing their valuable data with me. Professor Sami Masri of University of Southern California also provided great advice during my study at Caltech for which I am thankful.

This work is dedicated to my family who patiently waited long periods of time between our short visits in the past few years so that their son and brother could focus on his academic interests. It is also dedicated to the love of my life, Olivia Grace. Her confidence in me and her encouragements were the reasons I did not give up work during some of the most turbulent times of my life.

## TABLE OF CONTENTS

<b>Abstract</b>	<b>1</b>
<b>Chapter 1</b>	
<b>Introduction</b>	<b>3</b>
Probabilistic Seismic Hazard Assessment .....	3
Ground Motion Processes as a Complex System.....	12
Research Motivation .....	17
Methodology for This Study .....	19
Literature Review for Ground Motion Simulation .....	22
<b>Chapter 2</b>	
<b>Formulation</b>	<b>31</b>
Principal Component Analysis .....	31
PCA of a Ground Motion Database and the Concept of Eigenquakes....	37
Gaussian Processes for Regression.....	40
Optimal Estimation of the Coefficients of the Eigenquakes.....	45
<b>Chapter 3</b>	
<b>Implementation</b>	<b>49</b>
Ground Motion Database .....	51
Acquired Eigenquakes and Their Properties .....	55
Development of Target Hazard Spectra .....	75
Computational Implementation Issues.....	79
Genetic Algorithms and Parallel Computing Schemes.....	80

<b>Chapter 4</b>	
<b>Development of Ground Motion Ensembles</b>	<b>91</b>
Example 1: Downtown Los Angeles .....	93
Example 2: Downtown San Francisco .....	106
<b>Chapter 5</b>	
<b>Conclusions</b>	<b>118</b>
Concluding Remarks .....	118
Future Work .....	120
<b>Bibliography</b>	<b>122</b>
<b>Appendix I</b>	
<b>PEER-NGA Data Sample</b>	<b>134</b>
<b>Appendix II</b>	
<b>Acquired Eigenquakes</b>	<b>157</b>

# Abstract

A novel method of model-independent probabilistic seismic hazard analysis (PSHA) and ground motion simulation is presented and verified using previously recorded data and machine learning. The concept of “eigenquakes” is introduced as an orthonormal set of basis vectors that represent characteristic earthquake records in a large database. Our proposed procedure consists of three phases, (1) estimation of the anticipated level of shaking for a scenario earthquake at a site using Gaussian Process regression, (2) extraction of the eigenquakes from Principal Component Analysis (PCA) of data, and (3) optimal combination of the eigenquakes to generate time-series of ground acceleration with spectral ordinates obtained in phase (1).

The benefits of using a model-independent method of PSHA and ground motion simulation, particularly in large urban areas where dense instrumentation is available or expected, are argued. The effectiveness of the proposed methodology is exhibited using eight scenario examples for downtown areas of Los Angeles and San Francisco where it is shown that no dependency on specific ground motion predic-

tion equations or processes of selection and scaling would be needed in our procedure. Furthermore, PCA allows systematic analysis of large databases of ground motion records that are otherwise very difficult to handle by conventional methods of data analysis.

Advantages, disadvantages, and future research needs are highlighted at the end.

# Chapter 1     Introduction

## Probabilistic Seismic Hazard Assessment

The problem of probabilistic seismic hazard assessment has been subject of many studies (Cornell 1968; Kramer 1996; Bazzurro and Allin Cornell 1999; Bozorgnia and Bertero 2004; McGuire 2004; McGuire 2008; Delavaud, Scherbaum et al. 2009; Musson 2009; Scherbaum, Delavaud et al. 2009; Wang and Takada 2009) and many debates (Bommer and Abrahamson 2006; Bommer and Abrahamson 2007; Bommer and Abrahamson 2007; Klugel 2007; Wang and Zhou 2007; Klugel 2009; Strasser and Bommer 2009; Geller 2011) in the past forty years. Scarcity of recorded data from large earthquakes, the harsh consequences associated with rare seismic events, and uncertainties associated with the source, path, and site, make the problem of seismic hazard assessment important from an academic point of view and from considerations of the societal needs. This problem is further exacerbated by the vast urban developments spreading into seismogenic regions and the aging infra-

structure (ASCE 2011)<sup>1</sup>, the result of which is evident from the recent global earthquake losses in the past two years<sup>2</sup>. Probabilistic methods of seismic hazard analysis (PSHA) have been developed to provide quantitative measures of the likelihood of the occurrence of seismic events (or exceedance of a certain level of intensity of shaking such as spectral acceleration) at a site over a period of time. Generally, the outcome of these studies is expressed in the form of a set of uniform hazard spectra that show the level of intensity at different periods corresponding to different probabilities of exceedance. PSHA procedures rely on regression models, often called Ground Motion Prediction Equations (GMPE) or attenuation relationships, that estimate the measure of intensity as a function of magnitude, distance, and other model-specific parameters depending on the model and its intended use (*e.g.*, see Figure 1.1).

---

<sup>1</sup> The American Society of Civil Engineers in their latest evaluation of the state of infrastructure in America reported the grand point average of America's infrastructure as "D" on a traditional letter grade scale from A to F.

<sup>2</sup> The January 12, 2010, M7.0 Haiti earthquake killed more than 220,000 people and injured 300,000 more. The February 27, 2010, M8.8 Maule, Chile earthquake collapsed 4 buildings and severely damaged 50 other buildings. The September 3, 2010, M7.1 and its February 22, 2011, M6.3 aftershock in Canterbury, New Zealand, caused about \$12 billion in loss. The March 11, 2011, M9.0 megathrust earthquake in the Sendai region, Japan, is estimated to have caused about 20-30 billion dollars in loss. All of these events occurred at or near the center of urban developments.

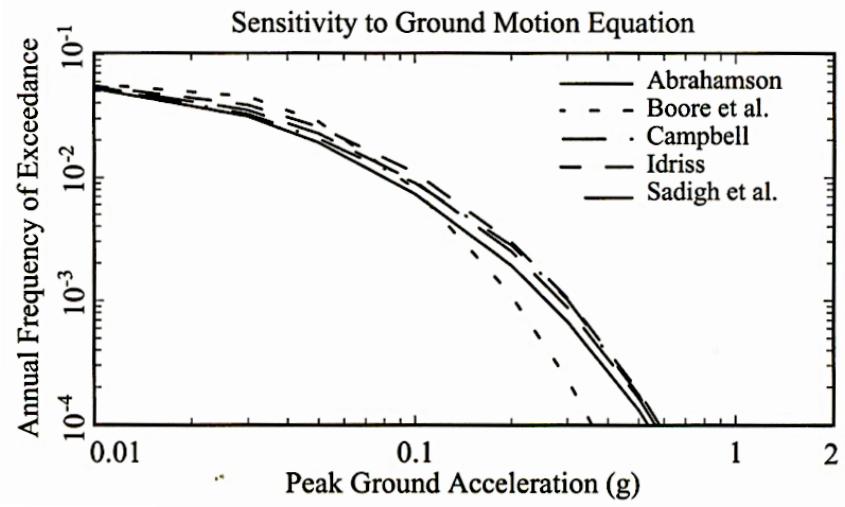


Figure 1.1. Seismic hazard curves in Berkeley and their sensitivity to different ground motion prediction equations (after (McGuire 2004))

The analytical procedures of PSHA can be summarized into four steps (Figure 1.2):

1. Identification and geometrical characterization of all sources of seismicity affecting the site of interest and probabilistic description of source to site distances (Step 1).
2. Development of a recurrence relationship for temporal distribution of event recurrence at each source. The Gutenberg-Richter relationship is often developed with consideration of characteristic earthquakes, from recorded data and catalogues of historical events, for probabilistic descrip-

- tion of the rate of occurrence of different magnitude of events at the site (Step 2).
3. Selection and implementation of a proper GMPE to estimate the measure of shaking for pairs of magnitude and distance from Steps 1 and 2 (Step 3).
  4. Use of Total Probability Theorem to consider uncertainty in earthquake size, location, and ground motion parameter by integrating over all magnitudes and distances to arrive at the probability of shaking exceeding a predefined value over a period of time (Step 4).

With information obtained in Step 4, one can construct a set of uniform hazard spectra with different probabilities of exceedance that would serve as a basis for design of a facility (*i.e.*, the minimum required level of shaking to be considered in the design for events with certain probabilities of occurrence<sup>3</sup>) (Newmark and Hall 1982).

---

<sup>3</sup> A *Design Basis Earthquake (DBE)*, is defined as a level of ground shaking that has less than 10% probability of being exceeded in 50 years (normal life of a building structure) and a *Maximum Credible Earthquake* or a *Maximum Considered Earthquake (MCE)*, is defined as a level of ground shaking that has less than 2% probability of being exceeded in 50 years. DBE and MCE events are commonly used in Performance-based Earthquake Engineering studies.

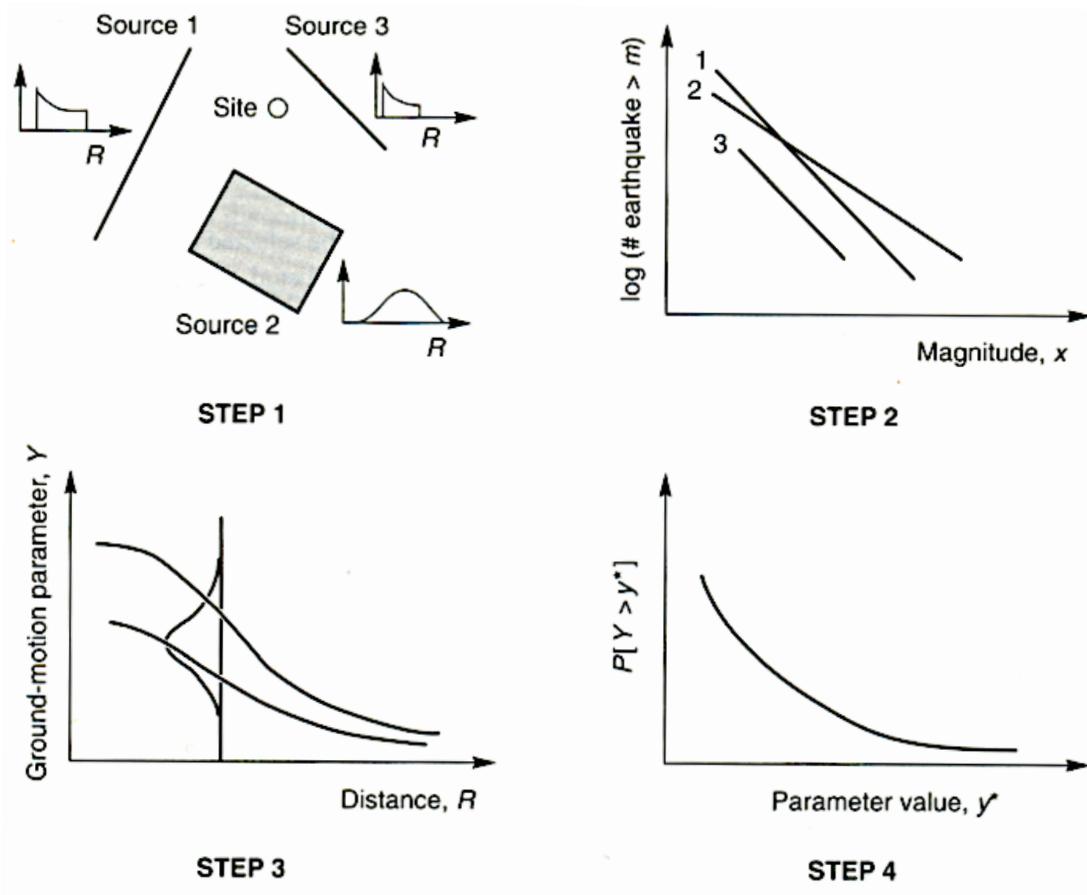


Figure 1.2. PSHA Steps (after (Kramer 1996))

Although development of design spectra is the ultimate goal of PSHA, these spectra by themselves do not provide enough information for modern methods of seismic design and performance assessment. This is because such performance assessment procedures require nonlinear dynamic response history analysis of a model subject to sets of ground motion time-series, natural or synthetic, that represent levels of shaking intensity predicted by the PSHA with considerations of site,

source, and path. Ground motion selection and modification procedures<sup>4</sup> have been developed for proper assembly of the ensembles of ground motion records and their adjustment to match intensities of shaking obtained in PSHA (Shome, Cornell et al. 1998; Naeim, Alimoradi et al. 2004; Watson-Lamprey and Abrahamson 2006; Beyer and Bommer 2007; Luco and Bazzurro 2007; Kottke and Rathje 2008; Shahbazian and Pezeshk 2010; Grigoriu 2011; PEER 2011). Building codes do not require use of a specific ground motion selection and modification method but they do provide brief guidelines on the adequacy of such procedures mainly from the consensus of the earthquake engineering community<sup>5</sup> (ASCE 2006). Consequently, there are different methods of ground motion selection and scaling in use with large discrepancies between their estimates of structural response quantities resulting from their application (PEER 2011). Furthermore, it has been known for a long time that excessive scaling of ground motion records and spectral alteration can be problematic (see as an example Figure 1.3). There is more evidence in recent years against spectral match-

---

<sup>4</sup> in time and / or frequency domain including scaling methods

<sup>5</sup> ASCE 7-05 under Chapter 21, Site-Specific Ground Motion Procedures for Seismic Design, reads, "At least five recorded or simulated horizontal ground motion acceleration time histories shall be selected from events having magnitudes and fault distances that are consistent with those that control the MCE. Each selected time history shall be scaled so that its response spectrum is, on average, approximately at the level of the MCE rock response spectrum over the period range of significance to structural response."

ing and excessive scaling and there are procedures to minimize the adverse effects of altering recorded ground motions (Naeim, Alimoradi et al. 2004; Grigoriu 2011).

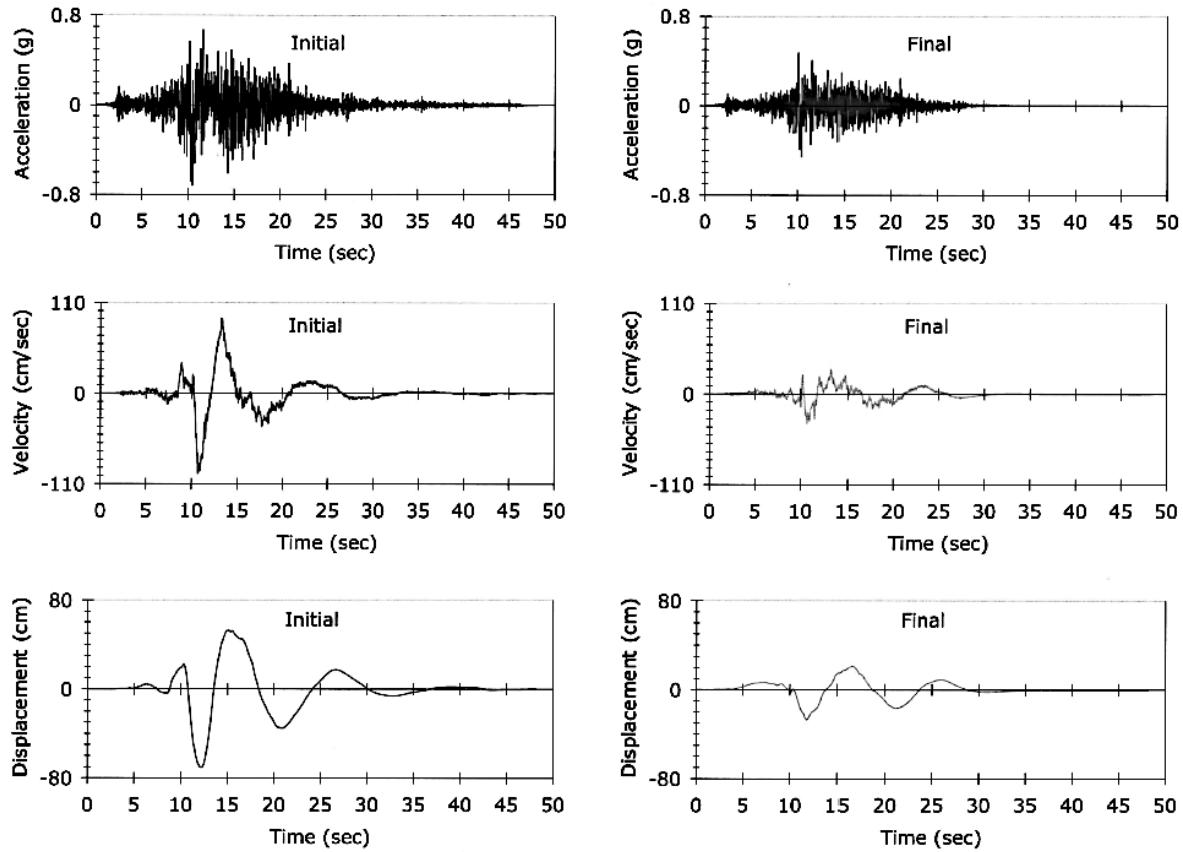


Figure 1.3. An example of ground motion modification to match a design spectrum - note that the near-source pulse is completely removed in the process of modification. Structural response evaluation subject to this record will likely produce questionable results.

An alternative to methods of recorded ground motion selection and modification is stochastic ground motion simulation, which is particularly suitable for stable

continental regions with infrequent seismicity but capable of generating large events (Housner and Jennings 1964; Amin and Ang 1968; Saragoni and Hart 1973; Ahmadi 1979; Ahmadi 1980; Der Kiureghian and Crempien 1989; Conte, Pister et al. 1992; Papadimitriou and Beck 1992; Atkinson and Beresnev 2002; Mobarakeh, Rofooei et al. 2002; Gu and Wen 2007; Rezaeian and Der Kiureghian 2008; Giaralis and Spanos 2009; Yamamoto and Baker 2011).

Stochastic ground motion simulation methods can generate random processes with temporal and spectral nonstationarities for a given set of seismic hazard parameters (magnitude, distance, etc) empirically, or by modifying the source spectrum of an earthquake over the path of seismic waves to account for the attenuation of waves analytically. Some work by passing a random process through a set of adjusted filters (mathematical models - not to be confused with models of physical processes) that change the random process in time and frequency domains in a way that the resulting motion has a form and intensity appropriate for the site. Historically, stochastic simulation methods have received many applications in the design of critical infrastructure such as nuclear power plants.

Both PSHA and stochastic simulation methods rely on regression-type models. With available data, predictive models are easy to set, understand, and use; but similar to implementation of any kind of model, care must be practiced when making predictions in the real world. Models are developed using observations on a

system's output considering a set of intuitively assumed parameters in an assumed model structure. Their utility is limited to the type of information contained in the regression data provided that the model class chosen has the right form. If future observations on data suggest existence of other important system parameters that were originally ignored in the model, or exhibit large prediction errors, the utility of the model would be questionable. In the Bayesian framework of system analysis, one's knowledge of a system is improved by data available from the system. Any effort to represent an actual process with an equivalent mathematical model should be conditioned based on the information that is available from the system's data and prior theoretical knowledge about the process. Furthermore, there are usually more than one possible model class to represent a system. Which model class should be used in a given situation? Note there are many ground motion prediction equations in the literature and in use; for a summary see (Abrahamson, Atkinson et al. 2008). Logic trees are used to reflect our confidence in different models by allowing a weighting scheme applied to different branches of the tree to find the most critical case or to arrive at an average of different model outputs. The problem with this way of treating our confidence in models is that weighting schemes should prefera-

bly be justified using information available from a system's data otherwise arbitrary or intuitive schemes may bias the final output of the tree in one way or another<sup>6</sup>.

There are also other challenges in using simplified models of reality in engineering analysis. Predictive models do not typically get automatically updated with arrival of new data. It would not be clear, when new observations are made, how these new observations could affect the prediction of a model and by how much.

These modeling limitation added to the problem of ground motion selection and modification, provide a justification for development of model-independent, data-driven, methods of seismic hazard analysis and ground motion simulation, as presented in this report.

## Ground Motion Processes as a Complex System

The problem of seismic hazard analysis and ground motion simulation have been traditionally treated in a forward fashion with reductionism applied to decompose the entire process into a number of smaller, better-understood, probabilistically well-defined components (such as seismicity rate, attenuation of waves, and source-to-site distance in PSHA; and stress drop, quality factor, site attenuation, shear wave

---

<sup>6</sup> "There is no true model since reality is always more complex than any model we make of it. Statistician D.R. Cox stated: "All models are wrong but some models are better than others!"" - Personal communication with Professor James Beck.

velocity, etc, in stochastic simulation methods.) Equations (1.1) and (1.2) show this simplifying process symbolically:

$$\log(IM|M,R,SF,SE,NS) = f_1(M) + f_2(R) + f_3(SF) + f_4(SE) + f_5(NS) + \varepsilon \quad (1.1)$$

$$FFT(y|M,R) = g_1(\text{Source}) \times g_2(\text{Path and Site}) \quad (1.2)$$

where  $IM$  = measure of intensity such as spectral acceleration,  $M$  = magnitude,  $R$  = distance,  $SF$  = style of faulting,  $SE$  = local site soil effects,  $NS$  = near source effect, and  $\varepsilon$  = random error. Equation (1.1) gives an estimate of the level of shaking (Step 3 in a PSHA) and Equation (1.2) is a synthetic signal in the frequency domain. The main advantage of this way of analysis is its simplicity that allows systematic investigation of a complicated process. There is a growing evidence, however, that the processes that generate ground motions and affect the propagation of waves are very complex (Mossotti, Barragan et al. 2002; Turcotte and Malamud 2002; Corral 2004; Stein, Liu et al. 2009). A complex system, in this context, is one with many interacting, dynamic, sometimes nonlinear components, in which the behavior of individual system components does not reveal the system's behavior that merges from the interconnection of the constituent components, *i.e.*, reductionism is inapplicable (Gallagher and Appenzeller 1999)<sup>7</sup>. One way of studying a complex system is

---

<sup>7</sup> Chaotic response could also be characteristic of a complex system.

through careful analysis of its behavior observed solely from its output rather than direct investigation of the system, its models, or its components since they could be in general unidentifiable. This way of analysis is data-driven and independent of any assumed model for the complex system. We take this approach in this study and treat ground motions as the output of a complex system that generates and propagates seismic waves from the source to the site. By performing Principal Component Analysis of a ground motion database we develop the concept of Eigen-quakes<sup>8</sup>. A large database of ground motion records can be taken as samples from a complex system of stochastic processes. An ensemble of ground motion records from some of the earliest observations showing large variability in the data is shown in Figure 1.4.

We make no assumptions with regards to details, number, and nature of dynamic processes that work together to generate an array of ground motion observations during a seismic event; however, we hypothesis that whatever the number and nature of these processes might be, their effects must be reflected in the waveforms generated by them (see Figure 1.5). Therefore, careful study of the generated data without imposing any predefined mathematical model may reveal information that is not available in common methods of ground motion and seismic hazard assessment (Alimoradi, Pezeshk et al. 2005). A notable example is the influence of local

---

<sup>8</sup> As described in Chapters 2 and 3.

topography on ground motion observations, a factor that is almost universally ignored in major ground motion prediction equations [see, for example, (Hough, Altidor et al. 2010)] but could significantly alter the characteristics of motion observed at a site. The same is true about other phenomena and processes (basin edge effects, trampoline effects, etc.) that are relatively newly discovered but have important effects on the intensity of motion observed at a site (Choi, Stewart et al. 2005; Yamada, Mori et al. 2009). We use Principal Component Analysis (PCA) (Jolliffe 2002) to extract useful information from a large database of ground motion records for our ground motion simulations. We then use recorded data to develop estimates of ground motion intensity at a site, using acceleration spectra over a wide range of periods, for different probabilities of exceedance in a Gaussian Process regression formulation (Bishop 2006; Rasmussen and Williams 2006). Two examples presented in Chapter 4 show the effectiveness of the procedure.

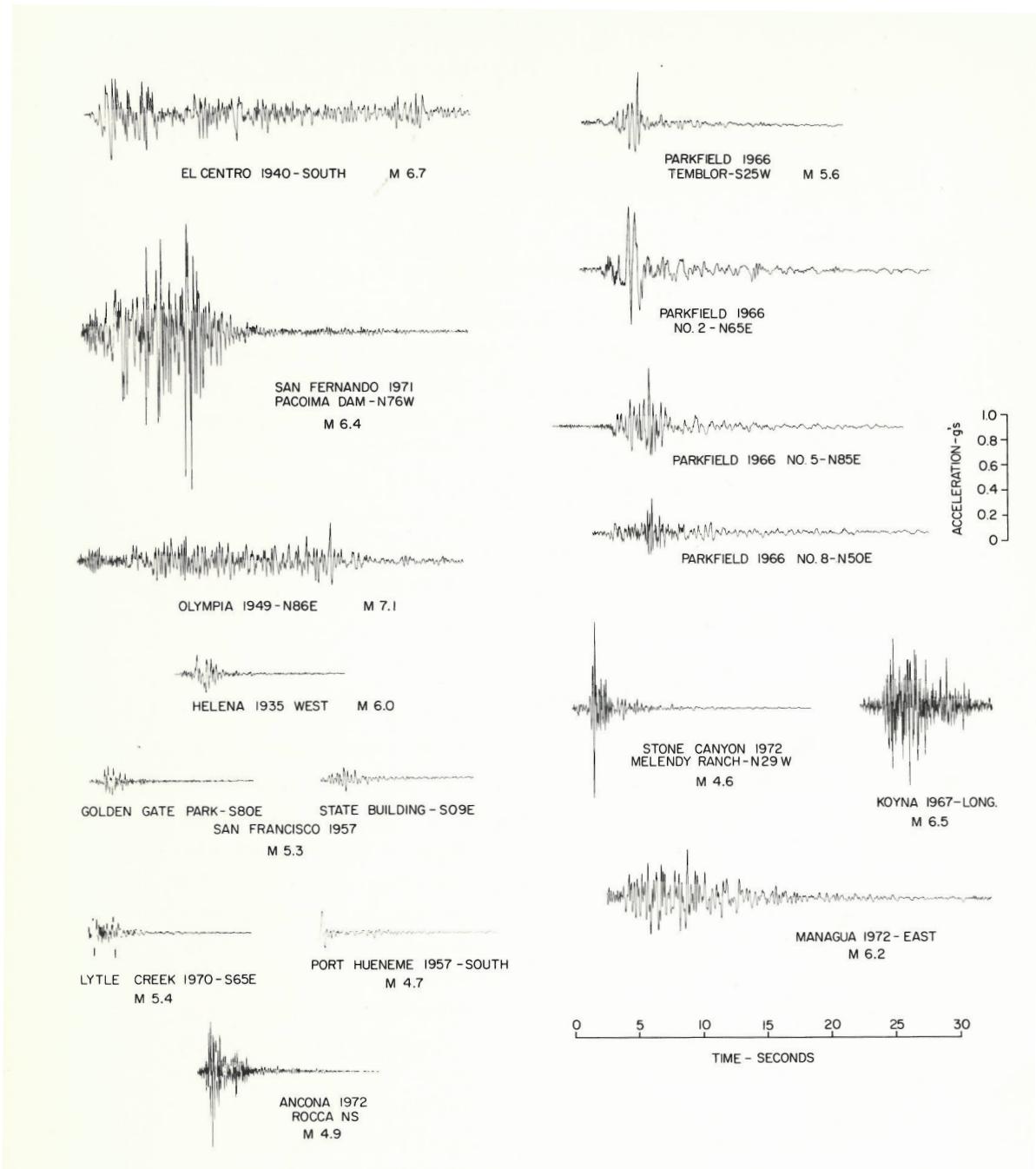


Figure 1.4. The classic example of large variation in ground motion records (Hudson 1979)

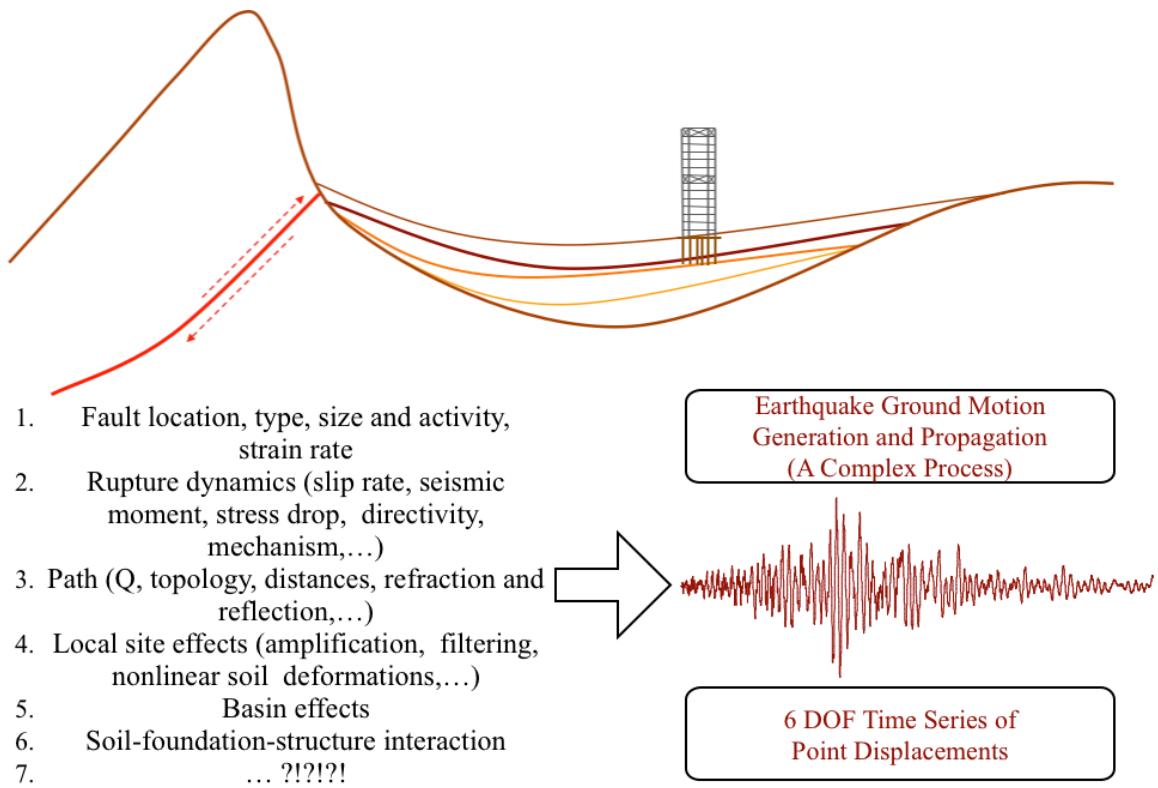


Figure 1.5. Reductionism could be avoided in simulation of ground motion records through application of machine learning techniques to a large database of ground records, *i.e.*, the output of a complex system.

## Research Motivation

As of February 2010, the global ground motion databases exceed  $10^5$  records (Anderson 2010). With the growing number of high-quality ground motion records available worldwide, instigated by the availability of low-cost sensors and major instrumentation programs such as ANSS<sup>9</sup>, application of machine learning methods for analysis, classification, and simulation of data becomes inevitable. The vast

---

<sup>9</sup> Advanced National Seismic System <<http://earthquake.usgs.gov/monitoring/anss/>>

amount of data generated and archived around the globe on a daily basis exceeds the processing capability of human beings but is suitable for the application of machine learning and data mining<sup>10</sup>. Most applications in seismic hazard analysis and ground motion simulation only make use of a small number of previously recorded data. There is a crucial need for creation of systematic procedures for analysis of large databases of ground motion. Besides, it is desired to develop a model-independent method of stochastic ground motion simulation, using information learned from a large number of previously recorded data, that does not depend on ground motion scaling and modification procedures.

A general concern in data-driven methods is dealing with situations where data is scarce, such as large events recorded close to the source. We anticipate that this will not be a problem in future as the archives of ground motion records con-

---

<sup>10</sup> It is important to note that the emphasis here is on the analysis of high-dimensional, high-quality recorded data rather than simple characteristics of motion for which automated rapid machine assessment tools are available (such as in Shakemaps). We are concerned about temporal and spectral forms of data in a large pool, what they represent, and the way they are correlated with the earthquake hazard parameters, *i.e.*, creating computational tools to answer questions such as this: “How is the occurrence of a recent earthquake in region X has contributed to our understanding of the global ground motion variation?” or “Are there any records of ground motion that had not been observed before and if so what kind of conditions do they represent?”

tinue to grow. Many urban areas in the world where earthquakes cause significant losses have already been heavily instrumented or are expected to be instrumented. As a result, the data-driven methods presented here are particularly suitable for seismic hazard assessment and ground motion simulation in large urban areas.

## Methodology for This Study

We look into utility of modern earthquake ground motion databases for probabilistic seismic hazard analysis and ground motion simulation by analyzing a large amount of data using machine learning techniques. Our objective is not to explicitly model the underlying physical processes that contribute to ground shaking and seismic hazard (Song and Somerville 2010), nor it is to rely on regression-type ground motion prediction equations (Abrahamson and Silva 2008; Boore and Atkinson 2008; Campbell and Bozorgnia 2008; Chiou and Youngs 2008; Idriss 2008). Physical modeling, although valuable in development of our understanding of earthquake processes, requires detailed information about the seismogenic, tectonic, and geotechnical settings at a site that is not, in general, available for most engineering projects. Regression-type ground motion prediction equations, on the other hand, pose other difficulties as was discussed earlier in this chapter.

Our proposed methodology is data-driven. From recorded ground motion data, we extract a set of orthonormal basis vectors representing the predominant

variations in the time-series with temporal and spectral non-stationarities considered. From the same data, we develop estimates of intensity of shaking at the site for a given scenario event using Gaussian Processes regression (see Figure 1.6). A scenario is considered an event with a certain moment magnitude and source-to-site distance recorded on a site with a particular shear wave velocity at its top sedimentary layers.

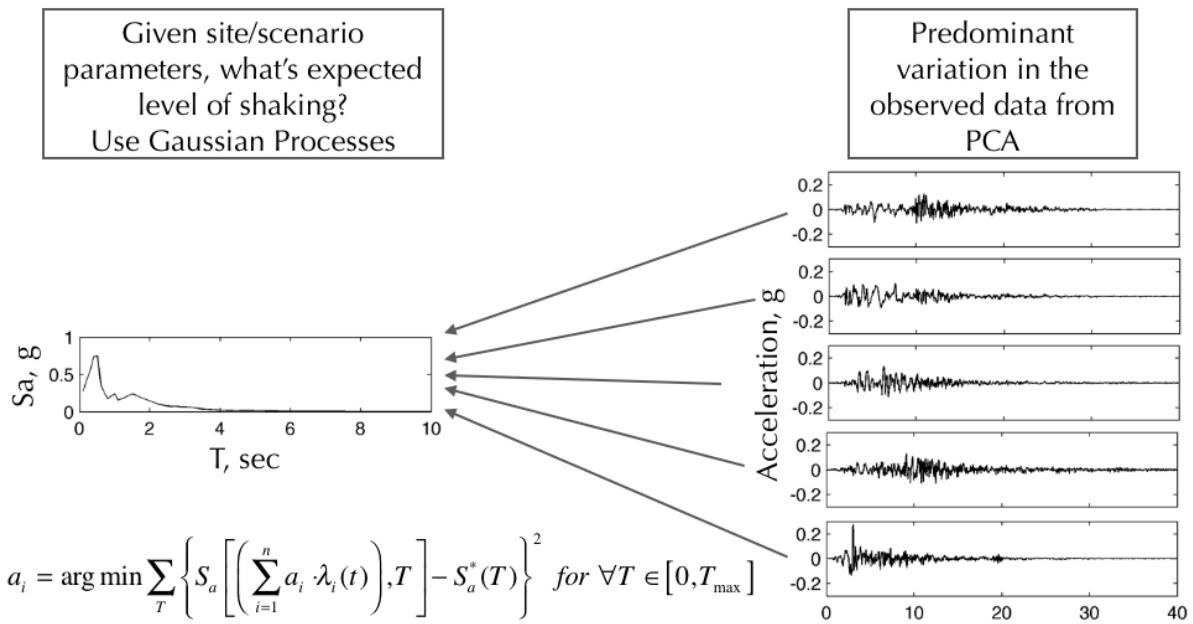


Figure 1.6. Schematic description of the proposed procedure.

The following example describes our proposed solution to the problem of seismic hazard analysis and ground motion simulation: suppose synthetic seismic waveforms (acceleration time-histories) are needed at a site for a specified probabil-

ity of exceedance of shaking over a number of years. Data associated with the location, acceleration response spectra of all magnitudes and distances previously recorded, can be used in a Gaussian Process regression to generate estimates of acceleration spectra of the motion anticipated at the site. Model-independent disaggregation of seismic hazard can be used to arrive at the pair of magnitude and distance given the probability of shaking at the site for use in Gaussian Process regression. Using an orthogonal basis of characteristic earthquake records, hereinafter referred to as *eigenquakes*<sup>11</sup>, the anticipated waveforms can be synthesized by finding an optimal set of coefficients such that the linear combination of the eigenquakes would create a signal that has spectral acceleration ordinates as close as possible to that obtained from the Gaussian Process regression. The generated ground motion records so obtained are verified with recorded data when possible.

---

<sup>11</sup> The prefix “eigen” is “own” in German and has been used in other contexts in earthquake engineering. The proper phrase for eigenquakes, one may argue, should be eigenrecords since the characteristic basis vectors are obtained from a large number of earthquake records with different earthquakes contributing different number of records. We opt for eigenquakes instead of eigenrecords. The interchange between records and earthquakes in earthquake engineering literature is nothing new - see, for example Housner, G. W. and P. C. Jennings (1964). "Generation of artificial earthquakes." Journal of Engineering Mechanics Division, ASCE **90**(1): 113-150.

The concept of eigenquakes as a basis of earthquake ground motion records in a database is studied, along with their utility for analysis and synthesis of data. Different motion properties of the eigenquakes in time and frequency domains (such as velocity and displacement time histories, acceleration, velocity, and displacement spectra, power spectral density spectrograms, and root mean square of cumulative acceleration) exhibit their unique characteristics and their resemblance to natural records of ground motion.

## Literature Review for Ground Motion Simulation

The state-of-the-art in ground motion simulation, selection, scaling, and modification for dynamic response analysis is briefly introduced here. The emphasis is on some of the more significant contributions and some historical developments, therefore, the list of reviewed articles below is by no means conclusive.

### *Historical and Noteworthy Developments*

Jennings and Housner presented one of the earliest proposals for ground motion simulation at the Fourth World Conference on Earthquake Engineering at the time that records of actual earthquake ground shaking were scarce (Jennings, Housner et al. 1969). Their procedure consisted of creating sections of a random process with a prescribed power spectral density and modulated by an envelop function in time domain.

Using a Bayesian probabilistic framework, Papadimitriou and Beck proposed a parsimonious probabilistic ground motion model that captures the essential features of the motion with a few parameters. Their model is formulated to be suitable for structural response computations. They also studied the random response of linear and nonlinear oscillators subject to the proposed ground motion model to investigate the effects of ground motion nonstationarities. (Papadimitriou and Beck 1992)

The role of ground motion  $\varepsilon$ -value<sup>12</sup> in structural response was studied by Baker and Cornell (Baker and Cornell 2006). They developed a method for finding the conditional response spectrum of a ground motion given spectral acceleration at the first-mode period of the structure,  $S_a(T_1)$ , as a measure of intensity, and its associated mean causal magnitude, distance and  $\varepsilon$ -value (CMS- $\varepsilon$ ). They observed that inclusion of  $\varepsilon$  to magnitude and distance reduces estimates of structural response relative to the estimates obtained with  $S_a(T_1)$  alone.

Iervolino and Cornell investigated the issues of ground motion selection and amplitude scaling for nonlinear seismic response assessment of structures and con-

---

<sup>12</sup>  $\varepsilon$  is the difference between the spectral acceleration of a record and the mean of a ground motion prediction equation at a given period. It can be taken as an indicator of spectral shape.

cluded that neither careful record selection, based on magnitude and distance, nor record scaling plays a significant role in affecting nonlinear displacement response of a suite of model structures that they considered in their study (Iervolino and Cornell 2005). Their conclusions were limited to firm soil sites with no directivity-affected ground motion records. Subsequently, Iervolino studied the effects of spectral shape and spectral matching of accelerograms on inelastic seismic response of single-degree-of-freedom oscillators (Iervolino, De Luca et al. 2010). They concluded that “the linearly scaled records do not show any systematic trend with respect to the unscaled record results... suggesting that scaling is a legitimate technique”.

The early history of accelerograph design and record processing is reviewed in a paper by Brady (Brady 2009). It covers developments of different ground motion recording instruments in the U.S. (in the 1930’s), Japan (in the 1950’s), and New Zealand (in the 1960’s), as well as the Earthquake Engineering Research Laboratory’s project on using an analog computer to calculate response spectra<sup>13</sup>. On the issue of accelerogram processing to retain important displacement characteristics, such as

---

<sup>13</sup> Brady talks about the tedious task of digitizing film records, “In addition to Caltech undergraduates, a team of women from the Jet Propulsion Laboratory in Pasadena was asked to participate in the digitizing, which proceeded at up to 18 h a day for 5 days a week for a few months.” and “the efforts of the small group of people dedicated to ensuring the quality of these early accelerographs.”

permanent offset in the near-source records, Iwan and colleagues performed some of the earliest work at the California Institute of Technology (Iwan, Moser et al. 1985).

Stein and Liu discuss the paradox of large earthquakes in the New Madrid Seismic Zone with indistinguishable strain build up for future large events and attribute it to complex system behavior of interacting faults (Stein, Liu et al. 2009). They observe that, “mid-continental faults ‘turn on’ and ‘turn off’ on timescales of hundreds or thousands of years, causing large earthquakes that are episodic, clustered, and migrating” and hypothesize that the spatio-temporal variability from interactions among the faults in a region is the result of a complex system behavior in that the fault system’s evolution cannot be understood by studying an individual fault alone.

Using orthogonal Karhunen-Loeve expansion, Masri and his colleagues present a direct procedure for expressing the covariance kernels of random processes in a compact form that is particularly suitable for analytical random vibration studies. They use an ensemble of records from the 1971 San Fernando earthquake to show the utility of their procedure.(Masri, Miller et al. 1990).

Scherbaum and his colleagues present a data-driven information-theoretic methodology for the problem of ground motion model selection based on Kullback–Leibler distance (Scherbaum, Delavaud et al. 2009). Implementation of Kullback–Leibler distance allows for ground motion model selection to be independent

of any *ad hoc* assumptions and the appropriateness of models can be expressed in terms of physically meaningful units (bits).

### *Stochastic simulation methods*

Following earlier work on strong plasma turbulence, Ahmadi offered a complete representation of earthquake ground acceleration using Wiener-Hermite expansion, *i.e.*, expansion of a stochastic function using a set of orthogonal random bases similar to polynomial chaos (Ahmadi 1980). The representation is complete in the sense that other stochastic ground motion representations, such as stationary filtered white noise, stationary multifiltered white noise, and nonstationary stochastic models can be taken as approximations of the Wiener-Hermite expansion. Ahmadi's work is interestingly similar to the procedure presented in this report, where, instead of orthogonal random bases of Gaussian white noise types, the eigenvectors of the covariance of a large ground motion database are used.

Recently Rezaeian and Der Kiureghian developed a fully non-stationary model with separate spectral and temporal non-stationarities “based on the modulation of the response of a linear filter with time-varying characteristics to a discretized white-noise excitation” (Rezaeian and Der Kiureghian 2008; Rezaeian and Der Kiureghian 2010). Their model consists of a modulating function characterized by three parameters and a linear filter represented by another set of three parameters. The parameters are identified from accelerograms and predictive equations are devel-

oped that relate model parameters to hazard and site. They present several examples that show "faithfulness of the model in reproducing realizations with statistical characteristics similar to those of the target motion." An example of a set of realization from their method is shown in Figure 1.7.

Another pioneering work on stochastic ground motion simulation was by Saragoni and Hart where sample functions of a stochastic ground motion process were obtained by modulating contiguous regions of filtered Gaussian white noise with a deterministic time function (Saragoni and Hart 1973). The nonstationary nature of the amplitude and frequency content of ground motion was preserved in their procedure.

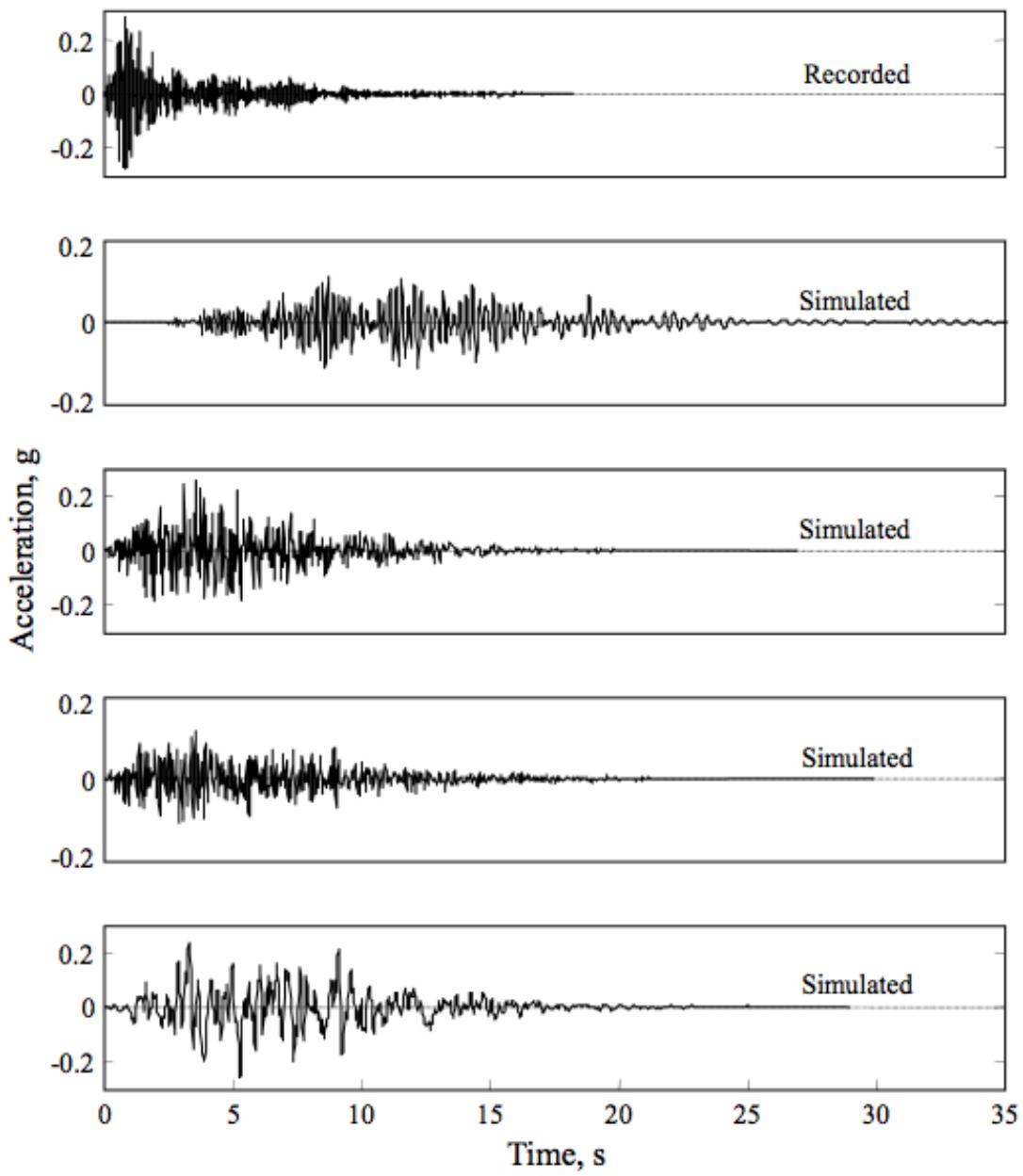


Figure 1.7. An example of four ground motion realizations for reverse faulting,  $M = 6.61$ ,  $R = 19.3$  km, and  $V_{30} = 602$  m/s. Recorded motion is component 291 of the 1971 San Fernando earthquake at the Lake Hughes #12 station (Rezaeian and Der Kiureghian 2010).

## *Physics-based simulation methods*

Our understanding of the faulting processes and rupture dynamics has greatly improved in the past thirty years, mainly through the study of recorded data after an event in an inverse problem setting that finds optimal set of parameters pertaining to the fault rupture process (Kikuchi and Kanamori 1982; Kikuchi and Kanamori 1986; Kikuchi and Kanamori 1991; Turcotte and Malamud 2002; Ammon, Ji et al. 2005; Lay, Ammon et al. 2010). It is anticipated that these advancements will mostly remain within the realm of advanced seismological investigations and will not likely be adopted as a standard of practice for typical structural analysis in the near future. One reason for this anticipation is that prescription of input excitation for evaluation and design of infrastructure should be stochastic in a way that considers all sources of uncertainty with regards to the source, path and site. This is because of the large uncertainty in our understanding of the existence, geometry, mechanics, activity, and past history of all sources of seismicity in the vicinity of a site and in our understanding of the path and site conditions and all other factors that influence the level of shaking (near-source, basin, and slap-down effects, for example).

Song and Somerville recently presented a method of source modeling to imitate physics of earthquake rupture by developing pseudo dynamic (kinematic) source tools that are compatible with rupture dynamics yet are computationally effi-

cient (Song and Somerville 2010). They analyzed synthetic and real dynamic rupture models and demonstrated that important features of earthquake rupture, that are otherwise difficult to study in zero-offset coherence, can be captured in their method.

# Chapter 2     Formulation

## Principal Component Analysis

*Principal Component Analysis* (PCA), also known as discrete Karhunen-Loéve transform or Proper Orthogonal Decomposition, is a common method of multivariate analysis first introduced by the British statistician Karl Pearson in 1901 (Jolliffe 2002). PCA has become a popular machine learning technique in recent years due to advancements in computational sciences and availability of fast processors on the one hand, and the increasing sizes of the archives of various recorded data types that a modern society collects on the other hand. PCA has been used widely in applications such as pattern recognition (face, voice, fingerprint, handwriting, etc.), regression analysis (for feature selection), cluster analysis and classification, and dimensionality reduction, amongst other applications.

PCA allows systematic analysis of large amounts of data by methodically finding directions in the space of the original data (usually high-dimensional) where predominant independent variations may be observed. By finding such sets of or-

thogonal directions of variability, the complex high-dimensional data can be reduced to a smaller set of important and independent variations (see Figure 2.1 below). "The central idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set" (Jolliffe 2002). Therefore, the goal of PCA is to compactly express data in the most meaningful basis.

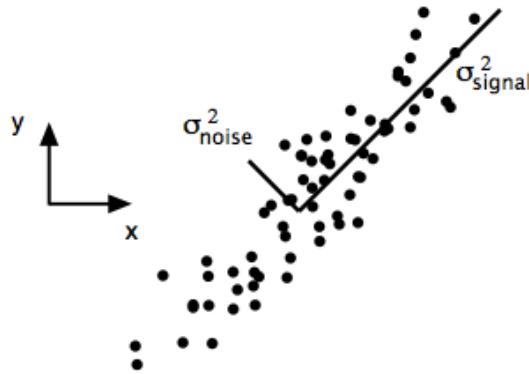


Figure 2.1. PCA is concerned with finding principal directions of variation in the data, after (Shlens 2009).

In its original framework, there are two interpretations of the PCA solution: the covariance interpretation, and the Singular Value Decomposition (SVD) or the change of basis interpretation. The underlying theories of both interpretations are solely built on the rigorous rules of linear algebra.

## Covariance Interpretation of PCA

Assume  $n$  records of  $m$ -dimensional data ( $n$  records with a maximum sample size of  $m$ ), presented in an  $n \times m$  matrix  $A$ :

$$A = \begin{bmatrix} \underline{a}_1 \\ \vdots \\ \underline{a}_n \end{bmatrix} \quad (2.1)$$

where  $\underline{a}_i$  is the  $i^{\text{th}}$  record in the data set, uniformly sampled at time interval  $\Delta t$  to give  $m$  data points and corrected, if necessary, to give zero mean. The sample covariance of data is<sup>14</sup>:

$$C_A = \frac{1}{m} A A^T \quad (2.2)$$

The  $n \times n$  matrix  $C_A$  represents the degree of linear relationship between all pairs of records, and therefore reflects the level of noise and redundancy in the data (see Figure 2.1) (Shlens 2009). It is desired to find a rotation operator,  $R$ , that maximizes the signal and minimizes the redundancy in the rotated database by diagonalizing the covariance of the rotated data. Therefore, if:

$$Y = R A \quad (2.3)$$

is the rotated data, the covariance matrix of the new dataset would be:

---

<sup>14</sup> the normalization constant of  $m$  or  $(m-1)$  does not affect our arguments here.

$$C_Y = \frac{1}{m} YY^T = \frac{1}{m} (RA)(RA)^T = R \left( \frac{1}{m} AA^T \right) R^T \quad (2.4)$$

$$C_Y = RC_A R^T \quad (2.5)$$

The covariance matrix of  $A$  is symmetric and so it can always be diagonalized using a similarity transformation matrix. Choosing this orthogonal matrix the orthonormal eigenvectors of  $C_A$  arranged column-wise to be  $R$ , Equation (2.5) can be rewritten as:

$$C_Y = R(R^T \Lambda R)R^T \quad (2.6)$$

$$C_Y = (RR^{-1})\Lambda(RR^{-1}) = \Lambda \quad (2.7)$$

where  $\Lambda$  is the diagonal matrix of the eigenvalues of the covariance of  $A$  arranged diagonally and using the fact that the inverse of an orthogonal matrix is its transpose,  $R^T = R^{-1}$ . Therefore,  $C_Y$  is a diagonal matrix whose non-zero elements are the principal variances of the data. The rows of the rotation operator  $R$  are called the *Principal Components* of  $A$  (Shlens 2009). Equation (2.3) shows  $A = R^T Y$  and so the data matrix  $A$  may be expressed in terms of the  $n$  principal components which form a basis of the space of  $n \times 1$  column vectors.

## SVD Interpretation of PCA

As before, assume  $n$  records of  $m$ -dimensional data are presented in  $n \times m$  matrix  $A$  of rank  $r$ . Eigenvalues of  $A^T A$  are  $\lambda_i \geq 0, i \in \{1, \dots, m\}$ <sup>15</sup>, and together with a set of  $m$  orthonormal  $m$ -dimensional eigenvectors,  $\underline{\phi}_i$ , they are defined by:

$$(A^T A) \underline{\phi}_i = \lambda_i \underline{\phi}_i, \quad i = 1, \dots, m \quad (2.8)$$

where  $\lambda_i = 0$  for  $i = r + 1, \dots, m$ , since the nullspace of  $A$ , which is also the nullspace of  $A^T A$ , has dimension  $(m - r)$ . The *singular values* of  $A$  are defined as

$\sigma_i \triangleq \sqrt{\lambda_i} > 0, i = 1, \dots, r$ . A new set of  $n$  orthonormal vectors in the  $n$ -dimensional

space can be defined as:

$$\underline{u}_i \triangleq \frac{1}{\sigma_i} A \underline{\phi}_i, i \in \{1, \dots, r\}, \quad \text{i.e. } A^T \underline{u}_i = \sigma_i \underline{\phi}_i \quad (2.9)$$

and

$$A^T \underline{u}_i = \underline{0}, \quad i \in \{r + 1, \dots, n\} \quad (2.10)$$

It is readily shown that  $\underline{u}_i$  are  $n$  orthonormal eigenvectors of  $AA^T$  and so are the principal components of  $A$ . In addition,

$$A\Phi = U\Sigma \quad (2.11)$$

---

<sup>15</sup>  $r$  is the number of  $\lambda_i > 0$

where  $\Phi$  and  $U$  are square matrices of order  $m$  and  $n$ , respectively, and  $\Sigma$  has all zeros except for  $\sigma_i$  as the first  $r$  diagonal values. Note that  $\underline{\phi}_i$ 's and  $\underline{u}_i$ 's are bases for different spaces. ( $m$ - and  $n$ -dimensional real column spaces, respectively).

### *Assumptions and limitations of PCA*

A potential limitation in classical PCA is its linear transformation (Equation 2.3). There are other methods, such as kernel PCA, presented in the literature for analysis of data that can best be presented in a curvilinear basis (for example, see (Bishop 2006)). The challenge for the data analyst in such cases is a quantitative way of comparing various bases for the most meaningful and efficient analysis of principal components.

Another limitation of the classical PCA is its inability to update the principal components and their associated variance with the availability of new data. Variational PCA methods have been proposed as a remedy (Bishop 1999).

One should note that the implied assumption in using PCA is that principal components associated with larger variances in the data present the most important contributions and those associated with smaller variances are influenced largely by noise and so can be dropped. It will be shown later in this report that the temporal characteristics of the higher (lower variance) principal components of a large ground motion database have the characteristics of noise in the data - see Appendix II.

# PCA of a Ground Motion Database and the Concept of Eigenquakes

In the context of earthquake engineering research and application, there is a contemporary belief that for all practical purposes, ground motion processes and their severity can be presented by a few simple parameters (historically, the peak ground acceleration, PGA, had been used solely for characterization of ground motion with all its shortcomings, but more recently spectral acceleration at the first mode period,  $S_a(T_1)$ , has been used widely. Vector-based intensity measures have also been proposed recently (Baker and Cornell 2005; Baker and Cornell 2008). The parameters used to characterize ground motion records and their severity are commonly referred to *Intensity Measures*, or IMs, in the performance-based earthquake engineering literature (Krawinkler 2001). The viewpoint of presenting data as complex as ground motions by only a few simple parameters is in a direct contrast with findings from some of the earliest studies on ground motions that a single measure of intensity cannot reliably predict the damage potential of a process as complex as strong ground motion shaking (Housner 1975; Housner and Jennings 1982). In the present investigation, we propose a new method of characterization of ground motion records using a large number of independent basis vectors which we call eigenquakes.

*Definition of Eigenquakes:*

Eigenquakes are the PCA time histories that characterize earthquake records in a database. Because they are principal components, they form an orthonormal basis in the space of ground motion records where every record in the database can be expanded using the eigenquakes, or alternatively, new records can be created using a linear combination of them:

$$a(t) = \sum_{i=1}^n a_i \lambda_i(t) \quad (2.12)$$

where  $a_i$  and  $\lambda_i(t)$  are the basis coefficient and the time-history of the  $i^{\text{th}}$  eigenquake. Eigenquakes are ordered and their ranking is based on the predominance of their form in the data, *i.e.*, too many records with similar features will be ranked lower than other records. Features of data in our analyses are the values of the ground acceleration at every sampled time instance, that is, the entire history of ground motion.

Eigenquakes resemble the temporal and spectral nonstationarities of the actual earthquake records. (as will be exhibited in Chapter 3). This property makes them particularly suitable as a set of basis vectors for ground motion simulation. They also provide the utility of dimensionality reduction in a large database by extracting only the useful information from previously recorded data and by discarding redundancies and noise.

A sample of the eigenquakes that are used in our simulations is shown in Figure 2.2. Specific information on the data used for their extraction and their properties are explained in Chapter 3.

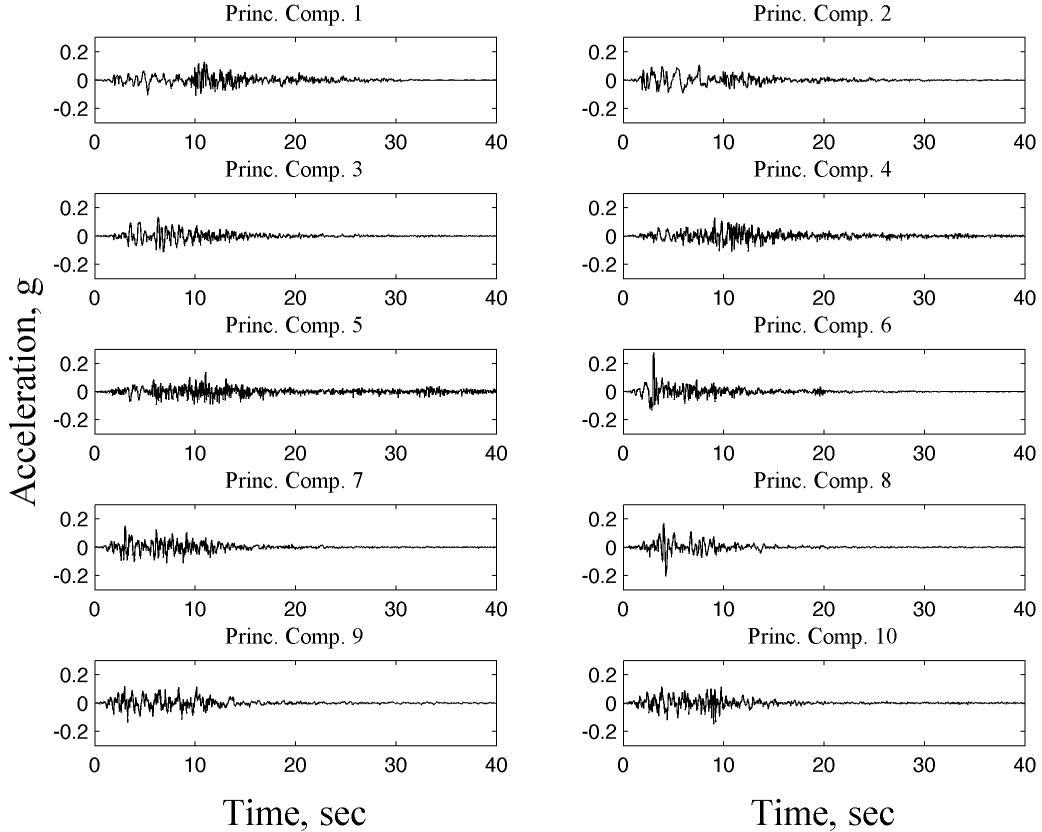


Figure 2.2. A sample of eigenvectors in a ground motion database.

To better treat the complexity and lack of detailed understanding of the processes that generate and propagate ground motion records, the principal components of data can be taken as a set of IMs in place of simpler scalar quantities.

# Gaussian Processes for Regression

Empirical (*i.e.* purely data-based) model development has been an integral part of all branches of applied sciences where mathematical models are developed to describe a complex system's response given some observations of the system variables. Classical parametric regression analysis has been used widely in the past; however, with a large amount of data, which is usually the case in modern applications, the task of model development by the application of classical regression methods becomes challenging. The following issues require careful attention:

1. A set of observed data can be described with different models. Which model should be chosen as most appropriate?
2. It is possible to reduce the scatter around a prediction model by increasing its order but with the risk of over-fitting the data so that it does not do a good job when predicting new data. How should the problem of over-fitting be avoided automatically?

Non-parametric methods, such as artificial neural networks, do not require a predefined mathematical form, although a good choice of the network architecture is often necessary. They make generalizations based on some ‘training’ data from the system. On the other hand, non-parametric methods are also subject to the challenging issues of model selection and over-fitting, as mentioned above.

Gaussian processes (GPs) provide a rational framework for dealing with the problems of model selection and over-fitting, as will be demonstrated later on in this chapter. Their history of development goes back to Kriging methods in geostatistics. Rasmussen defines GPs as “a generalization of multivariate Gaussian distribution to infinitely many variables” and “a distribution over functions [with] inference taking place directly in the space of functions” (Rasmussen 2006; Rasmussen and Williams 2006).

Similar to a Gaussian distribution, a GP is defined by a mean and covariance functions  $m(x)$  and  $k(x,x')$ , respectively:

$$p[f(x)] \sim \mathcal{GP}[m(x), k(x, x')] \quad (2.13)$$

which are functions of the input variables  $x$ . Note that Equation (2.13) defines a distribution over functions of  $x$ . There are different choices available for the covariance function of a GP. Depending on the application, some of these choices are (Snelson 2006):

$$\text{Squared exponential (SE) covariance: } k(x, x') = \sigma_f^2 \exp\left[-\frac{1}{2}\left(\frac{x - x'}{l}\right)^2\right] \quad (2.14)$$

$$\text{Matérn's class: } k(x, x') = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}|x - x'|}{l}\right)^\nu K_\nu\left(\frac{\sqrt{2\nu}|x - x'|}{l}\right) \quad (2.15)$$

$$\text{Brownian motion (Wiener process): } k(x, x') = \min(x, x') \quad (2.16)$$

$$\text{Linear covariance: } k(x, x') = \sigma_f^2 + xx' \quad (2.17)$$

amongst others. The SE covariance is smooth and stationary (invariant to translation) and will be used in this study.

To use GPs for regression in a supervised learning scheme based on training data  $(\mathbb{X}, \mathbb{Y})$ , where  $(\mathbb{X}, \mathbb{Y})$  consists of pairs of system input and output  $(x_i, y_i), i = 1, 2, \dots$ , and model  $y = f_\theta(x) + \varepsilon$ , the Gaussian likelihood would be:

$$p(\mathbb{Y} | \mathbb{X}, \theta, \mathcal{M}) \propto \prod_i \exp \left[ -\frac{1}{2} (y_i - f_\theta(x_i))^2 / \sigma_{\text{noise}}^2 \right] \quad (2.18)$$

with a prior of  $p(\theta | \mathcal{M})$  over the vector of the hyperparameters. An estimate of the vector of the hyperparameters can be obtained from the Bayes' rule:

$$p(\theta | \mathbb{X}, \mathbb{Y}, \mathcal{M}) = \frac{p(\mathbb{Y} | \mathbb{X}, \mathcal{M}, \theta) p(\theta | \mathcal{M})}{P(\mathbb{Y} | \mathbb{X}, \mathcal{M})} \quad (2.19)$$

and predictions can be made considering:

$$p(\mathbb{Y}^* | \mathbb{X}^*, \mathbb{X}, \mathbb{Y}, \mathcal{M}) = \int_{\theta} p(\mathbb{Y}^* | \mathbb{X}^*, \mathbb{X}, \mathbb{Y}, \mathcal{M}, \theta) p(\theta | \mathcal{M}, \mathbb{X}, \mathbb{Y}) d\theta \quad (2.20)$$

Throughout this report, an asterisk superscript is used to identify an estimated quantity.

For a nonparametric formulation, the GP predictive density for scalar  $y^*$  is given as (Rasmussen 2006):

$$p(y^*|x^*, \mathbb{X}, \mathbb{Y}, \mathcal{M}) \sim \mathcal{N} \left( \begin{array}{l} K(x^*, \mathbb{X})^T [K(\mathbb{X}, \mathbb{X}) + \sigma_{\text{noise}}^2 I]^{-1} \mathbb{Y}, \\ K(x^*, x^*) - K(x^*, \mathbb{X})^T [K(\mathbb{X}, \mathbb{X}) + \sigma_{\text{noise}}^2 I]^{-1} K(\mathbb{X}, x^*) \end{array} \right) \quad (2.21)$$

where  $K(\mathbb{X}, x^*)$  is an  $n \times n^*$  matrix of the covariances evaluated at all pairs of training and estimated points, for  $n$  training points and  $n^*$  estimated points. The other  $K$  terms are defined similarly. To estimate spectral acceleration of the ground motion as a function of a scenario event that has a moment magnitude  $m$ , recorded at a hypocentral distance from the causative fault  $r$ , on a site with a shear wave velocity of its top 30 meter layers  $v30$ , the logarithm of the spectral acceleration at all periods of interest is:

$$y = \log[Sa^*(T)] = f(m, r, v30) + \varepsilon, \quad \varepsilon = \mathcal{N}(0, \sigma_{\text{noise}}^2), \quad \text{for } \forall T \in [0, T_{\max}] \quad (2.22)$$

Let  $x = (x_1, x_2, x_3) = (m, r, v30)$ , then a squared exponential kernel for the covariance of the GP would be:

$$k(x, x' | \theta) = \prod_{i=1}^3 \sigma_{f,i}^2 \exp \left[ \sum_{i=1}^3 \frac{-(x_i - x'_i)^2}{2l_i^2} \right] + \sigma_{\text{noise}}^2 \prod_{i=1}^3 \delta(x_i, x'_i) \quad (2.23)$$

to which a noise term has been added (contribution with  $\sigma_{\text{noise}}$ ).  $l$  and  $\sigma_f$  are the hyperparameters for the length-scale and the amplitude; therefore, the vector of the hyperparameters:

$$\theta = \{\sigma_{f,i}, l_i, \sigma_{\text{noise}}\}, i = 1, \dots, 3 \quad (2.24)$$

can be identified using Bayes' theorem which gives the maximum a posteriori estimate of  $\theta^*$ , the regression parameters, through maximization of:

$$\log p(\log[\mathbb{S}a(T)] | \mathbb{X}, \theta) = -\frac{1}{2} \log [\mathbb{S}a(T)]^T K^{-1} \log [\mathbb{S}a(T)] - \frac{1}{2} \log \|K\| - \frac{n}{2} \log 2\pi \quad (2.25)$$

where a broad uniform prior is taken for  $\theta$ .

$\mathbb{X}$  consists of:

$$\{m_i, r_i, v30_i : i = 1, \dots, N = \text{no. of data}\} \quad (2.26)$$

Finally, to estimate the probability of observing spectral acceleration  $S_a^*$  for the target spectrum using recorded data  $\mathbb{S}_a$  and the vector of the identified hyperparameters,  $\theta^*$ , Equation (2.21) gives:

$$P(\log[Sa^*(T)] | \log[\mathbb{S}a(T)], \theta^*) \sim \mathcal{N}(K^* K^{-1} \log[\mathbb{S}a(T)], K^{**} - K^* K^{-1} K^{*T}) \quad (2.27)$$

for  $\forall T \in [0, T_{\max}]$

# Optimal Estimation of the Coefficients of the Eigenquakes

Having specified the anticipated intensity of shaking at a site from the GP regression, one is faced with finding an optimal set of coefficients for the linear combination of the eigenquakes such that the resulting motion's acceleration spectrum is minimally deviated from that obtained from the GP regression,  $S_a^*(T)$ :

$$a_i = \arg \min_T \sum \left\{ S_a \left[ \left( \sum_{i=1}^n a_i \cdot \lambda_i(t) \right), T \right] - S_a^*(T) \right\}^2 \text{ for } \forall T \in [0, T_{\max}], a_i \in \mathbb{R} \quad (2.28)$$

where  $a_i$  is the  $i^{th}$  coefficient associated with the eigenquake  $\lambda_i$  as in Equation (2.12) (see Figure 1.6 in Chapter 1). The operations required in Equation (2.28) cannot be performed directly in the spectral frequency domain, *i.e.*, the fact that every eigenquake has a unique acceleration spectrum is not enough to assume that Equation (2.28) can be simplified in matrix form to:

$$\{S_a^*\}_{np \times 1} = [S_a(\lambda)]_{np \times nEigq} \{a\}_{nEigq \times 1} \quad (2.29)$$

This is because the transformation of a linear combination of the eigenquakes in the time domain  $S_a[a_1\lambda_1 + a_2\lambda_2]$  is not equal to  $a_1S_a(\lambda_1) + a_2S_a(\lambda_2)$  in the spectral frequency domain, that is, the transformation is nonlinear. In this context,  $S_a(x)$  is taken as a function that generates an acceleration response spectrum from the time-

series  $x$ . The nonlinear mapping can be easily verified by considering  $\lambda_1$  and  $\lambda_2$  as

simple pulses of accelerations with different polarities and by taking  $a_1$  and  $a_2$  equal

to one. As shown in Figure 2.3,  $S_a(\lambda_1 + \lambda_2)$  gives a zero spectrum but

$S_a(\lambda_1) + S_a(\lambda_2) = 2S_a(\lambda_1)$  does not. Furthermore, the mapping  $S_a(x)$  is not one-to-

one because there can be many time-series  $x$  that map into a specified spectrum.

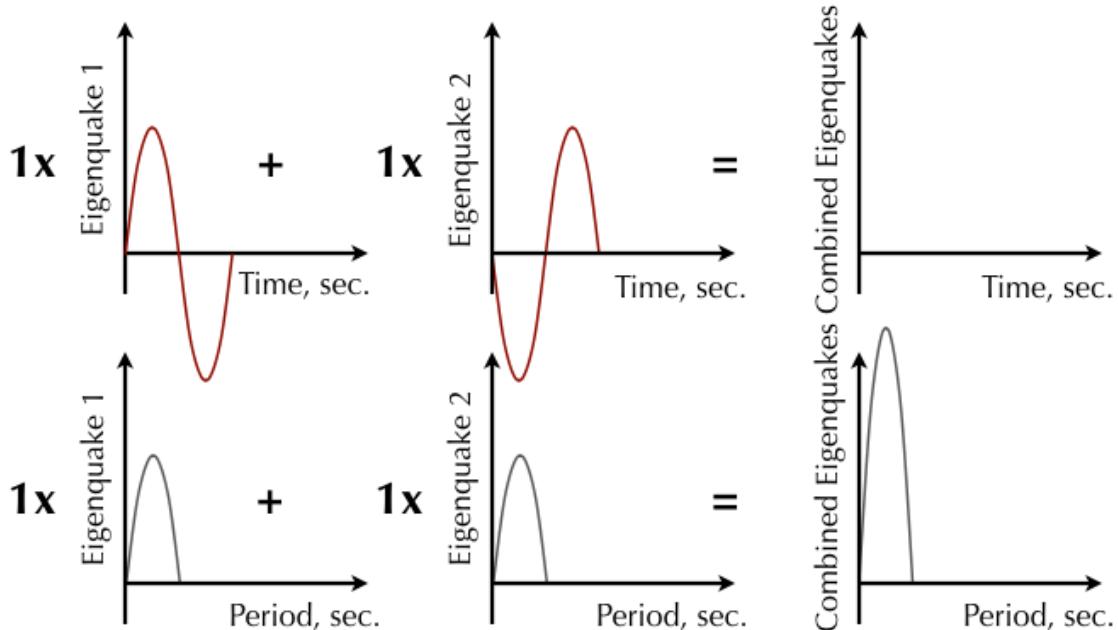


Figure 2.3. The mapping of a linear combination of the eigenquakes to the spectral frequency domain is not linear.

From the total probability theorem, the probability of having spectral acceleration  $S_a^*$  (generated from the GP regression) for  $\forall T \in [0, T_{\max}]$  from the data and a given set of eigenquakes is:

$$\begin{aligned} p(S_a^* | \Lambda, \mathcal{D}, \mathfrak{M}) &= \int_a p(S_a^* | \mathbb{S}_a, \Lambda, a, \mathfrak{M}) p(a | \mathcal{D}, \mathfrak{M}) da \\ &\approx p(S_a^* | \mathbb{S}_a, \Lambda, a^*, \mathfrak{M}) \end{aligned} \quad (2.30)$$

where  $\Lambda$  is the set of eigenquakes.  $\mathbb{S}_a$  is the spectral acceleration values observed from the data. The probability of having coefficients  $a$  from the Bayes' formulae would be:

$$p(a | \mathcal{D}) = \frac{p(S_a | \lambda, a) p(a)}{p(S_a | \lambda)} \quad (2.31)$$

where the probability model for  $S_a$  is defined as:

$$S_a = S_a \left( \sum_i a_i \lambda_i \right) + \varepsilon; \quad \varepsilon \sim \mathcal{N}(0, \sigma_p^2) \quad (2.32)$$

To estimate the coefficients, maximum a posteriori estimation or MAP gives (Beck and Katafygiotis 1998; Katafygiotis and Beck 1998):

$$a^* = \arg \max_{a \in \mathbb{R}^n} p(a | \mathcal{D}) = \arg \max_{a \in \mathbb{R}^n} p(S_a | \Lambda, a) p(a) \quad (2.33)$$

in which  $p(a)$  is the prior probability distribution based on a *priori* information about the distribution of the coefficients. In lieu of using prior information about the distribution of the coefficients, maximum likelihood estimation of the coefficients can be formulated as expressed in Equation (2.28). Stochastic search methods, such as genetic algorithms, have been shown in general as good candidates for robust optimization in large dimensional spaces and are used in this study to locate the optimal set of the coefficients (see Computational Implementation Issues in Chapter 3 for a description of the genetic algorithm implementation).

# Chapter 3 Implementation

Development of model-independent site-specific ground motion acceleration time-series can be summarized in three computational modules, two of which are independent of each other and can be performed in parallel, as is shown in Figure 3.1:

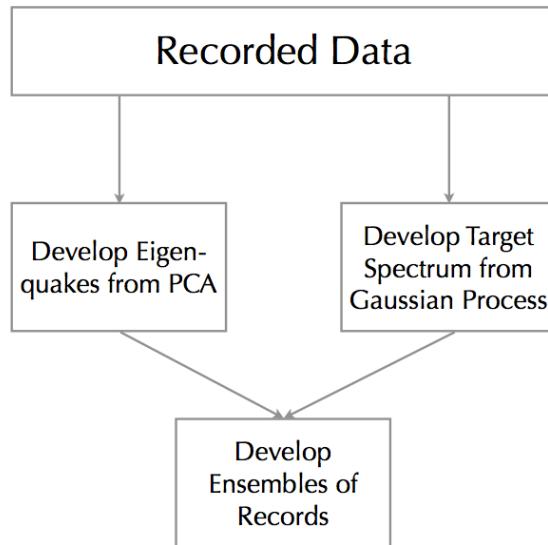


Figure 3.1. Schematic description of the site-specific procedure.

The description of the modules is as follows:

1. *Development of Target Spectrum from a Gaussian Process (GP) Regression:* To estimate intensity of shaking for a given probability of exceedance using previously recorded data, a GP is used to probabilistically express spectral acceleration values anticipated at the site for a given scenario event. A scenario is taken as an event with a given moment magnitude occurring at a given hypocentral distance<sup>16</sup>. Local data, used in the GP regression, is a compilation of spectral acceleration ordinates from all available magnitude-distance data pairs relevant to the site. The target acceleration spectrum so obtained is site-specific because it is derived from local records that contain local site effects due to the top sedimentary layers.

2. *Development of Eigenquakes from PCA:* Constructs an orthonormal set of basis vectors representing time-varying characteristics of actual ground motion records, for ground motion simulation.

3. *Ensemble Development:* Time histories of ground acceleration, given the response spectrum obtained in Step 1, are developed using mean-least square error estimates of a linear combination of the eigenquakes. The optimal coefficients of the eigenquakes are those that result in a linear combination of the eigenquakes

---

<sup>16</sup> Different scenarios can be determined from de-aggregation of seismic hazard at the site.

with minimal mean-square deviation from the target acceleration spectrum obtained in Step 1.

The entire process is data-driven, *i.e.*, it does not require arbitrary or empirical models and factors for its performance. The top two computational modules are also independent of each other and can be completed in parallel<sup>17</sup>.

## Ground Motion Database

PEER-NGA database of strong ground motion records (Chiou, Darragh et al. 2008; PEER 2011) is used in this study although methodologies developed are equally applicable to other databases of earthquake ground motion records<sup>18</sup>. The

---

<sup>17</sup> In fact, it may not be needed to develop the eigenquakes specifically for each simulation. A global set of eigenquakes, representing characteristics of strong ground motion variation globally, may be envisioned as an attribute of modern ground motion databases. Such archives could create and periodically update the list of their eigenquakes from all collected data in the past. Therefore, Step 2 for Development of the Eigenquakes, may be bypassed by resorting to the use of the “global” eigenquakes. Alternatively, one may opt for site-specific eigenquakes when enough local data is available.

<sup>18</sup> COSMOS Virtual Data Center from the Consortium of Organizations for Strong-Motion Observation Systems is a noteworthy example. It can be accessed at:

<http://db.cosmos-eq.org/scripts/default.plx>

important questions that should be addressed when choosing a database of ground motion records for data simulation and seismic hazard analysis are:

- *Sufficiency of Information:* Are all essential data attributes available from the database? The essential data attributes consist of tectonic, seismological, and geotechnical information; as well as station and instrumentation specifications, corrected time-series of the recorded components, and their orientation.
- *Data Quality:* Are data collection and processing protocols up to modern standards and uniform?
- *Richness of the Database:* Is the database rich with many different events of different sizes recorded at different places with different characteristics?
- *Applicability of the Database:* Is the database local or global? Should a global database be used with many records representing diverse characteristics that are likely to happen at a local site that has limited recorded data (although some data characteristics may have not been observed in the past), in place of a regional database that well represents the site (due to site-specific recordings) but has rather a much smaller number of records?
- *Database Growth:* How quickly is the database updated with availability of new data? Inclusion of new recorded ground motions is important for improving quality of the simulations in a data-driven procedure; it is also important for

verification purposes, that is, "how would the occurrence of new events in future change the properties of the simulated records and the estimates of hazard?"

The PEER-NGA database contains 3551 uniformly processed<sup>19</sup> ground motion data from 173 worldwide events recorded between 1935 to 2003 (Chiou, Darragh et al. 2008). For each entry in the database there are over 35 attributes available in time and frequency domains, as well as seismological, tectonic, and geotechnical properties (Strasser and Bommer 2009). The events in the PEER-NGA range from Mw 4.27 to Mw 7.90 recorded within the hypocentral distances of 5.02 to 557.85 km on sites with the top layers' shear wave velocity ranging from 116.4 to 2016.1 m/s. The distribution of magnitude and distance from the PEER-NGA data is shown in Figure 3.2 (Chiou, Darragh et al. 2008).

The ground motion records in the PEER-NGA database are processed and corrected, when needed, to account for the response of the ground motion instrument and to remove the effects of random noise. Both low- and high-pass Butterworth filters were applied in the frequency domain with baseline correction<sup>20</sup> when filtering did not remove non-physical trends in the displacement time series (Chiou,

---

<sup>19</sup> When needed, some records were entered into the database unprocessed. See Chiou, B., R. Darragh, et al. (2008). "NGA Project Strong-Motion Database." *Earthquake Spectra* 24(1): 23-44.

<sup>20</sup> By fitting a polynomial to the displacement time series and subtraction of the corresponding acceleration from the filtered acceleration time series.

Darragh et al. 2008). The PEER-NGA records vary in the Peak Ground Acceleration (PGA) values from 0.0009 g to 1.6615 g. The range of variation for the Peak Ground Velocity (PGV) is from 0.10 cm/sec to 169.96 cm/sec, and for the Peak Ground Displacement (PGD) is from 0.01 cm to 232.39 cm.

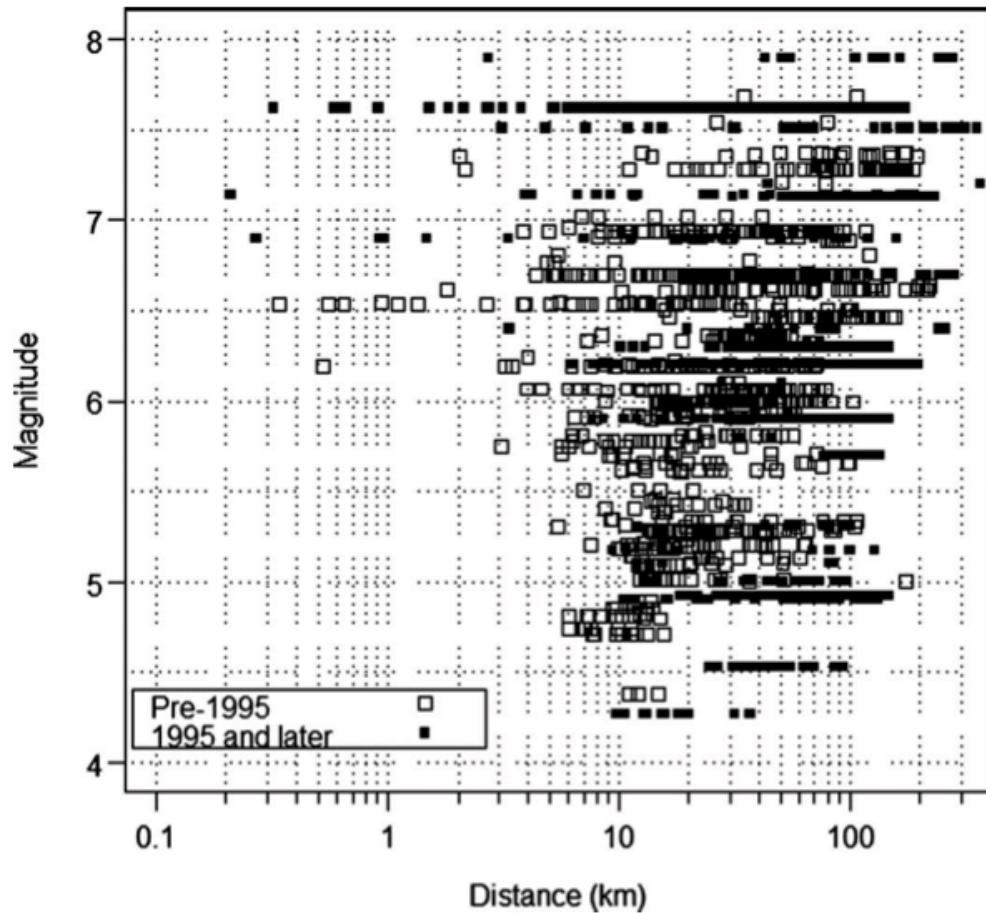


Figure 3.2. Distribution of magnitude and distance in the PEER-NGA database, after (Chiou, Darragh et al. 2008).

We randomly select 530 records from 87 worldwide events from the PEER-NGA database for derivation of the eigenquakes. These records are shown in Appendix I along with their filename. The large diversity in ground motion histories is clear from a simple visual inspection of this sample. It includes, for example, low-amplitude long-duration records (such as 0523c090.AT2), high-amplitude short-duration records (DUR--T.AT2), and variations in between. Although most records appear roughly symmetric (peak positive and negative acceleration in the same order) there are some asymmetric records in the sample (40O07NS.AT2). Some records show very large amplitudes (ABBAR--T.AT2) while others exhibit somewhat stationary fluctuations (HPB270.AT2). There are also uncommon profiles of acceleration evolution in the sample, such as successive asymmetric spikes (GIC180.AT2) and bimodal (phases of buildup, strong motion, decay) records (TARZA360.AT2).

## Acquired Eigenquakes and Their Properties

Following procedures of Chapter 2, PCA of the sample ground motion histories was performed to obtain the eigenquakes. Eigenquakes are characteristic records of a database and their form and rank change as new data arrives in the database. To show this data-dependency of the eigenquakes, PCA of the previously selected sample of 530 records was compared to PCA of two of its smaller subsets; a sample of 100 records and a set of 300 records. Each database was analyzed to cap-

ture enough principal components so that 95% variability in the data is captured, *i.e.*, cumulative sum of the principal components reaching 0.95. The results are shown in Figures 3.3 and 3.4 for the 100-record sample, in Figures 3.5 and 3.6 for the 300-record sample, and in Figures 3.7 and 3.8 for the 530-record sample. Figures 3.3, 3.5, and 3.7 show the top ten eigenquakes for the 100-, 300-, and 530-record samples, respectively. Principal components in these figures are eigenquakes of data. Note that some eigenquakes that are nonexistent in the 100-record sample start showing up when more data is added in larger databases. For example, no eigenquake of the 100-record sample resembles the type of ground motion variation that eigenquake number 2 in the 300-sample represents. As more data is added to the database, this eigenquake subsequently moves to eigenquake number 6 in the 530-sample case due to predominance of other types of variation brought in by the arrival of new data.

Figures 3.4, 3.6, and 3.8 present the number of eigenquakes required in each analysis case to capture 95% variability in the data and the contribution of individual eigenquakes. The bar chart at the bottom of these figures shows the contribution of each eigenquake to the cumulative sum. Dimensionality reduction in PCA is evident as the number of required records has reduced to 39, 103, and 140 in the 100-, 300-, and 530 record samples.

A summary of the analyses is presented in Table 3.1.

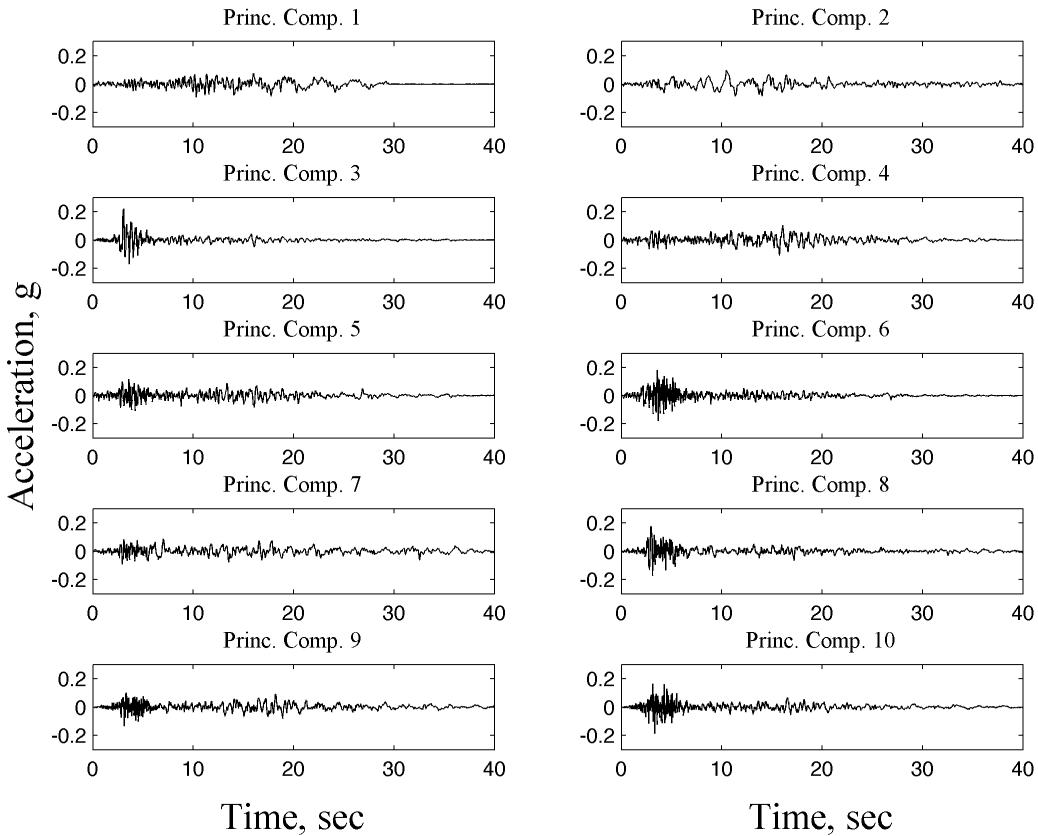


Figure 3.3. Top ten extracted eigenquakes from a sample of 100 NGA records.

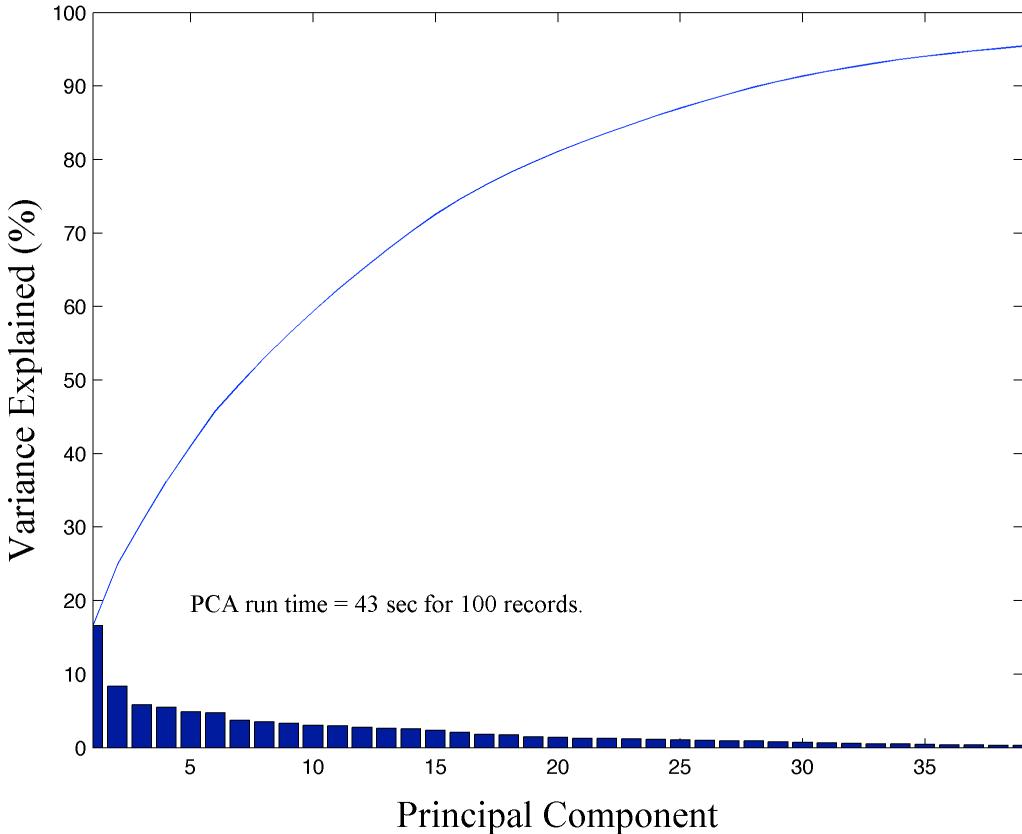


Figure 3.4. Number of eigenquakes needed to account for 95% variation in data in a sample of 100 NGA records.

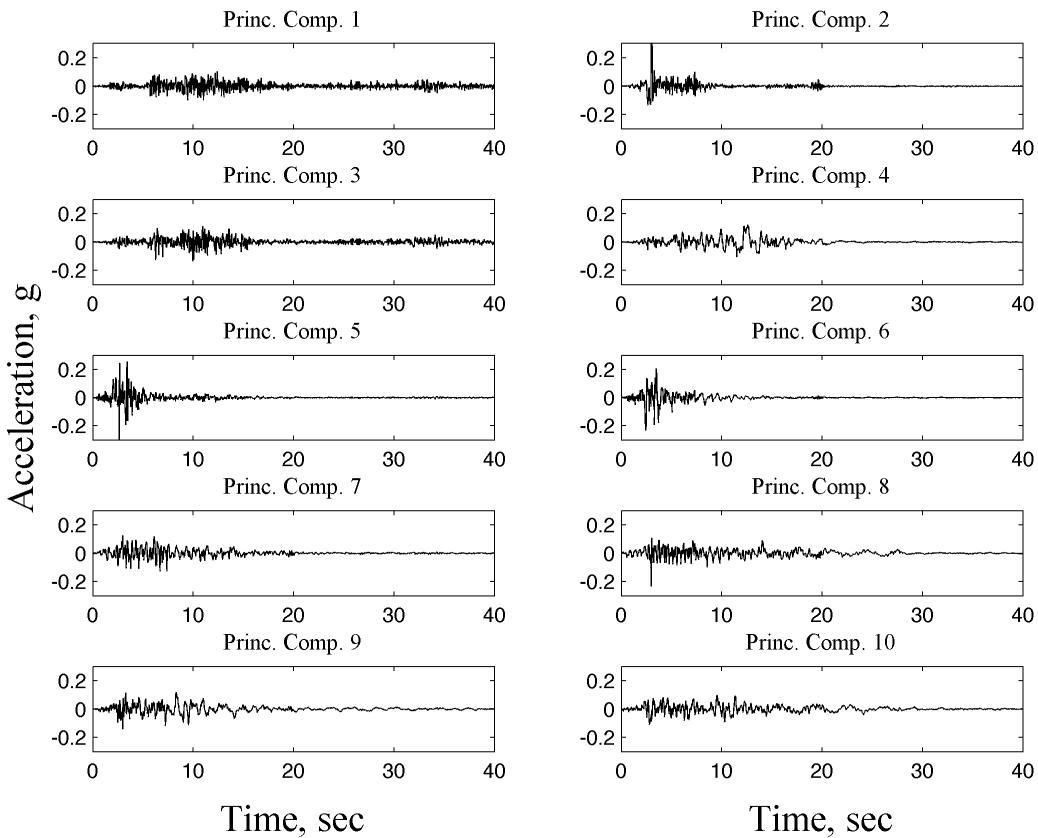


Figure 3.5. Top ten extracted eigenquakes from a sample of 300 NGA records.

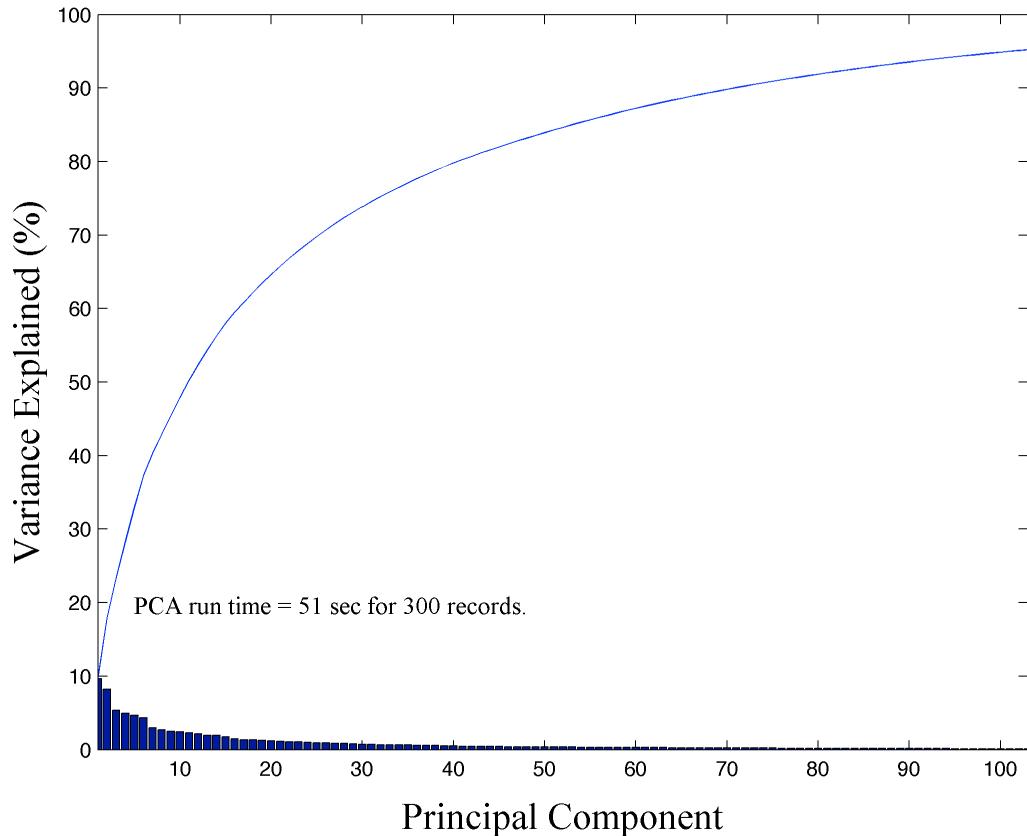


Figure 3.6. Number of eigenquakes needed to account for 95% variation in data in a sample of 300 NGA records.

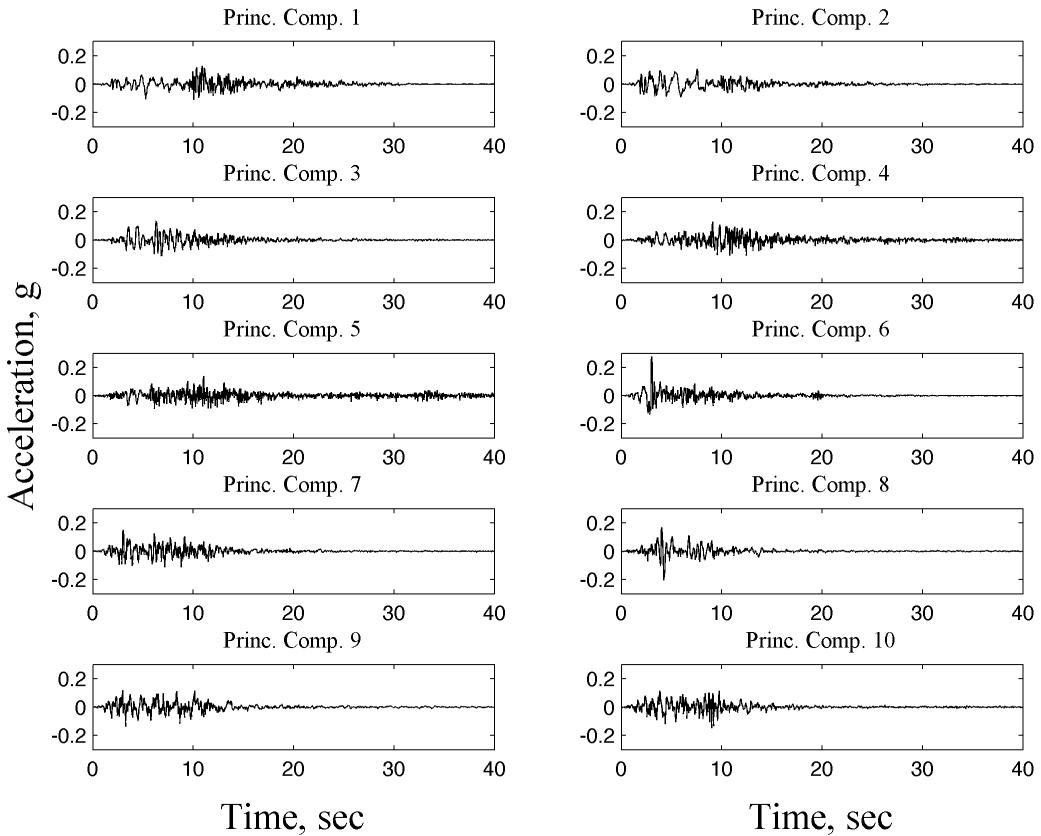


Figure 3.7. Top ten extracted eigenquakes from a sample of 530 NGA records.

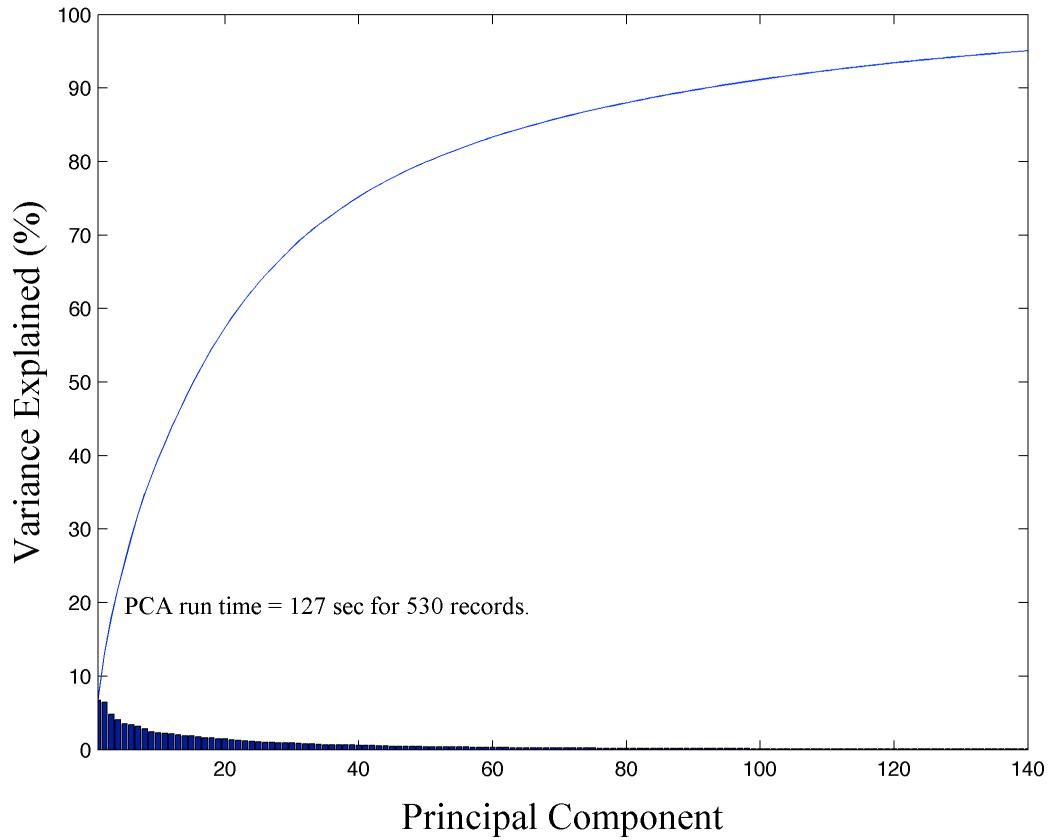


Figure 3.8. Number of eigenquakes needed to account for 95% variation in data in a sample of 530 NGA records.

Table 3.1. Summary of PCA Runs.

Analysis Case Number	Database Size	Number of Eigenquakes to Capture 95% Variability in Data	PCA Run Time, sec.
1	100	39	43
2	300	103	51
3	530	140	127

The complete set of eigenquakes obtained from the 530-record database is presented in Appendix II. As the rank of an eigenquakes increases its contribution to describing variability in the data decreases. As a result, the last few eigenquakes appear largely non-discriminative and very much representative of small variations in the data.

### *Time domain properties of the eigenquakes*

The acceleration time histories of the top 10 eigenquakes, shown in Figure 3.7, are integrated using the trapezoidal rule to arrive at time histories of velocity shown in Figure 3.9. Similarly, integration of the velocity time histories is performed to obtain time histories of the eigenquakes' displacement, shown in Figure 3.10. Considering that eigenquakes are normalized, the absolute peak values of these histories are unimportant, but the evolution of the signals, the shape of the envelops, their duration, and the number of their peaks are consequential. The most essential prop-

erty to note when studying eigenquakes is that, (1) they are independent of each other, and (2) they represent the essence of recorded data. Therefore eigenquakes are ideal vectors to be used as seeds in ground motion simulation.

The strong motion phase of the eigenquakes could start in just a couple of seconds into the signal (as in eigenquakes 6) or could take longer (10 seconds to complete the buildup as in eigenquake 1). It could be short, such as in eigenquake 8, or could be long as in eigenquakes 4 and 5. No offset is observed at the end of the velocity time histories due to PEER-NGA prior data correction schemes. The duration of significant shaking appears to be from less than 10 seconds to around 35 seconds. A few of the eigenquakes exhibit characteristics of motion in the near-source region<sup>21</sup>.

---

<sup>21</sup> Given proper amplification, the near-source pulses should be clear from the time histories of velocity and displacement in eigenquakes 6 and 8.

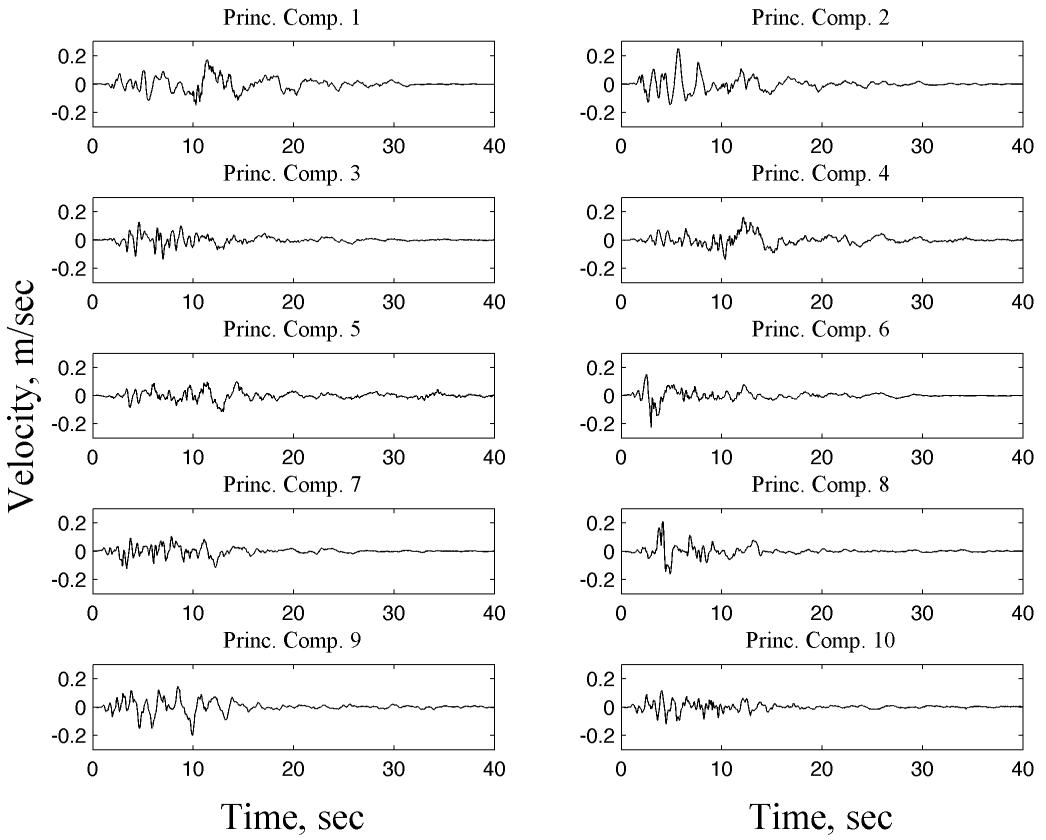


Figure 3.9. Velocity time histories of the top ten eigenquakes from numerical integration.

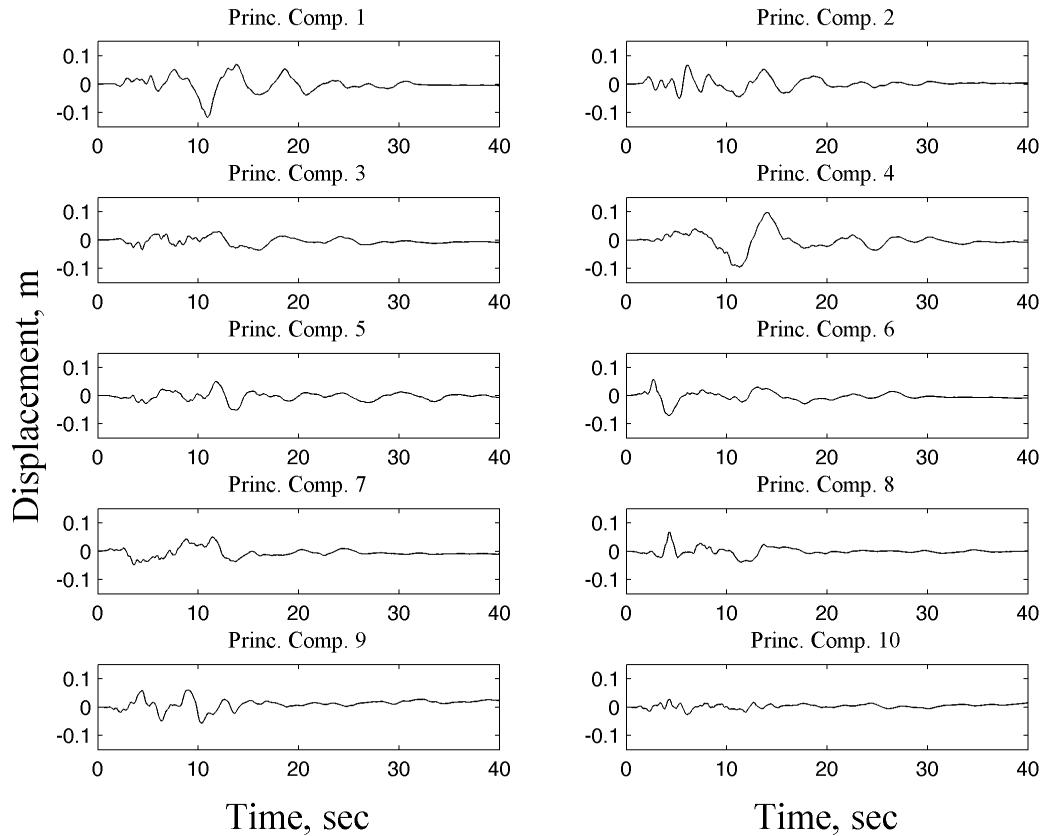


Figure 3.10. Displacement time histories of the top ten eigenquakes from numerical integration.

We also study the evolution of the signals by checking the envelop of acceleration histories using the square root of the cumulative square of accelerations:

$$CA(T) = \left[ \int_0^T a^2(t) dt \right]^{\frac{1}{2}} \quad (3.1)$$

The square root of the cumulative squared acceleration curves of the top ten eigenquakes are shown in Figure 3.11. It should be clear from this figure that the eigenquakes are normal vectors, since:

$$\lim_{T \rightarrow T_{\max}} [CA(T)] = \|a\|_2 \quad (3.2)$$

where  $\|a\|_2$  is the  $l^2$ -norm of the acceleration vector.

When computed for a much larger number of eigenquakes, the square root of the cumulative squared acceleration curves show how the eigenquakes differ from each other with their evolution over time (see Figure 3.12).

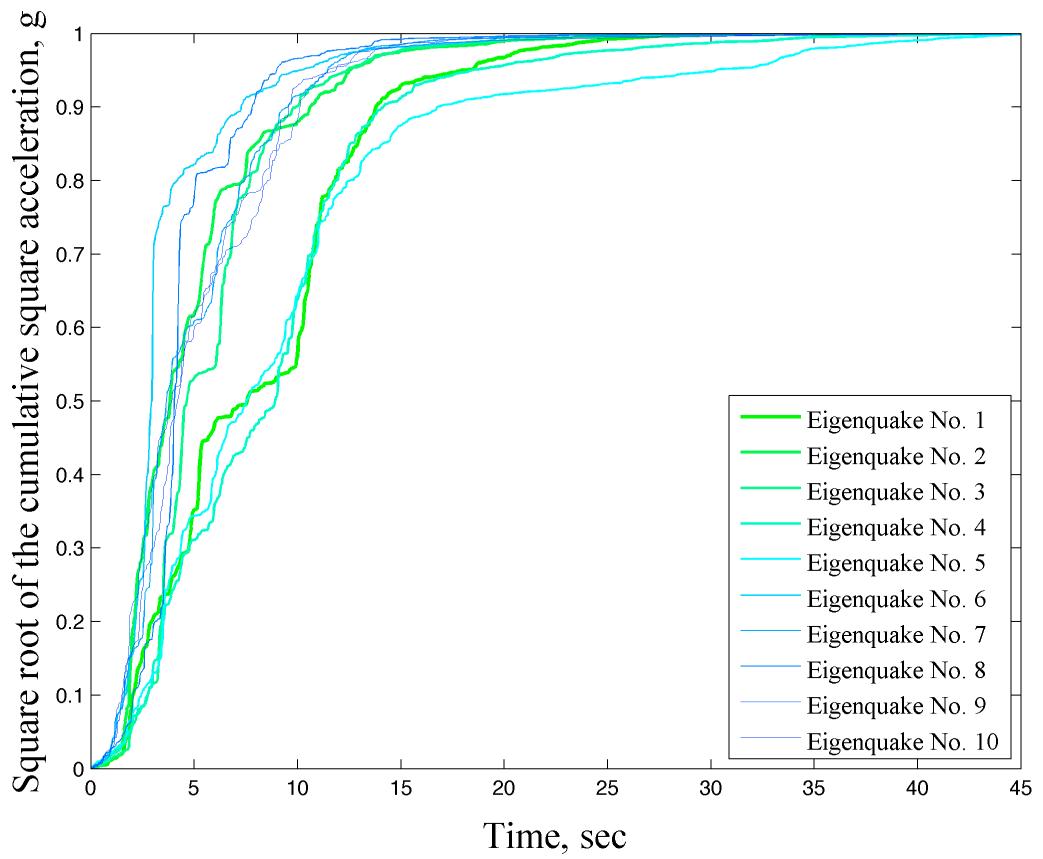


Figure 3.11. The square root of the cumulative squared acceleration for the top ten eigenquakes.

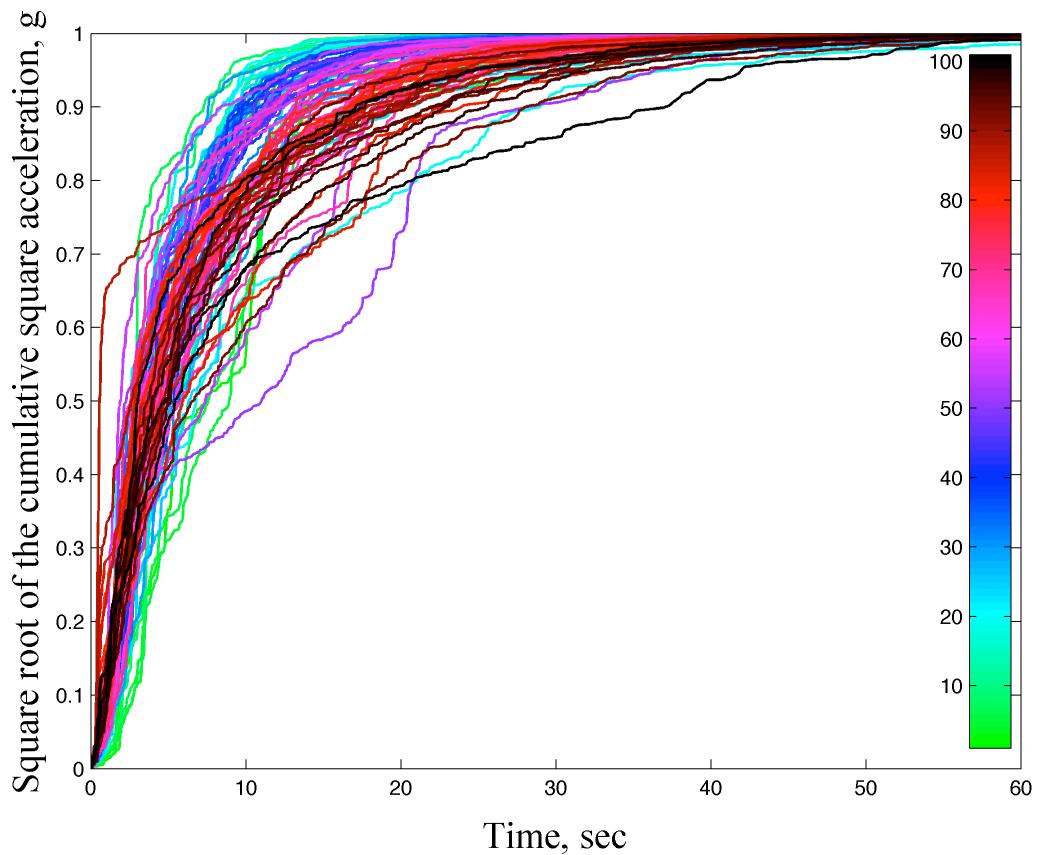


Figure 3.12. The square root of the cumulative squared acceleration for the top one hundred eigenquakes. Color bar on the side shows eigenquake numbers.

### *Frequency domain properties*

Similar observations about the independence of the eigenquakes and the rich dynamics that they jointly present are also valid in the frequency domain. Acceler-

tion spectra of the top ten eigenquakes are shown in Figure 3.13. For the data sample used in the derivation of these eigenquakes, the predominant period where the maximum spectral acceleration is observed in a SDOF system, occurs at periods shorter than 2 second (or frequencies higher than 0.5 Hz)

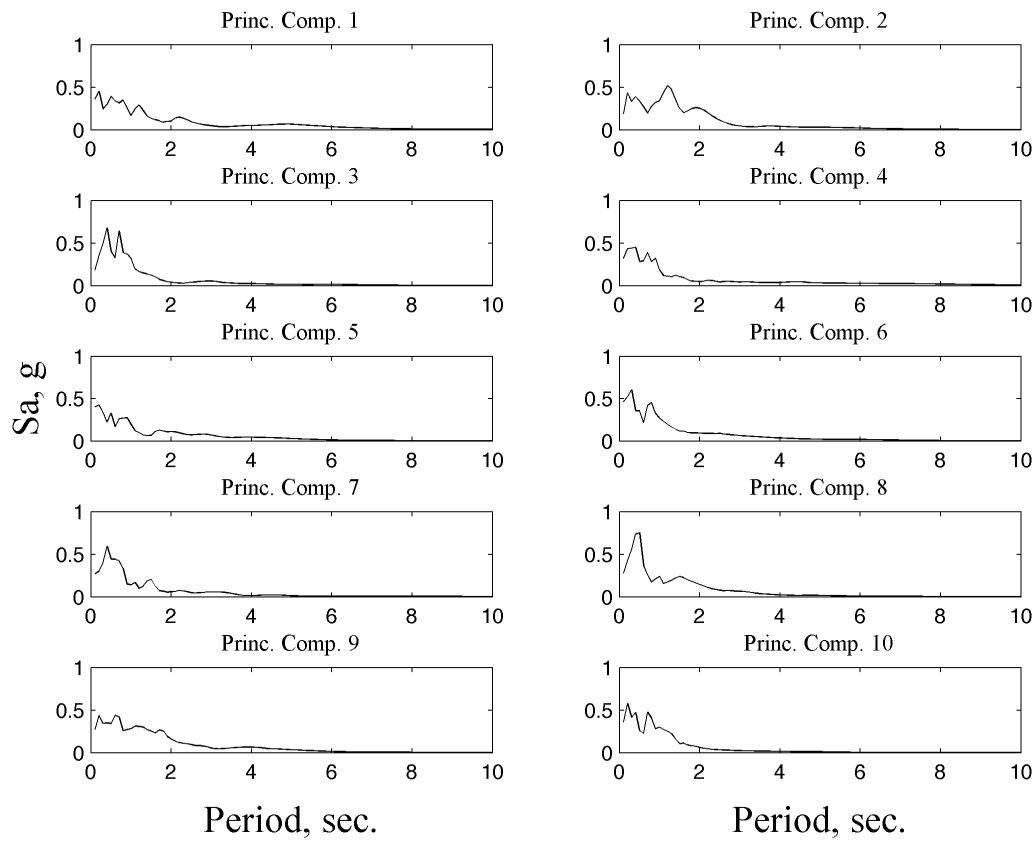


Figure 3.13. Acceleration spectra of the top ten eigenquakes.

Velocity spectra of the top ten eigenquakes are shown in Figure 3.14. While eigenquake 1 imposes a large velocity demand on oscillators with natural period of about 5.0 sec, eigenquakes 3 and 10 are most demanding to oscillators with periods shorter than 2.0 sec.

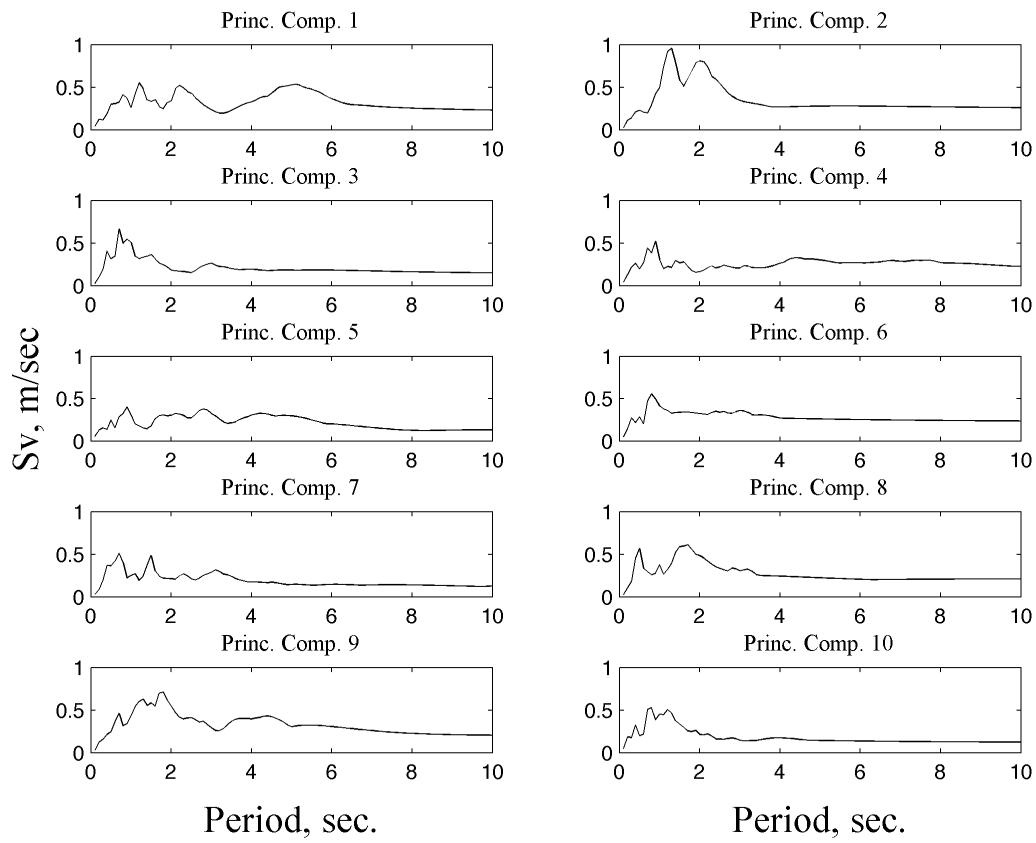


Figure 3.14. Velocity spectra of the top ten eigenquakes.

The displacement spectra of the top ten eigenquakes also show a large diversity between spectral characteristics of the eigenquakes (Figure 3.15), and so is the case with the Fourier amplitude spectra of the eigenquakes (Figure 3.16).

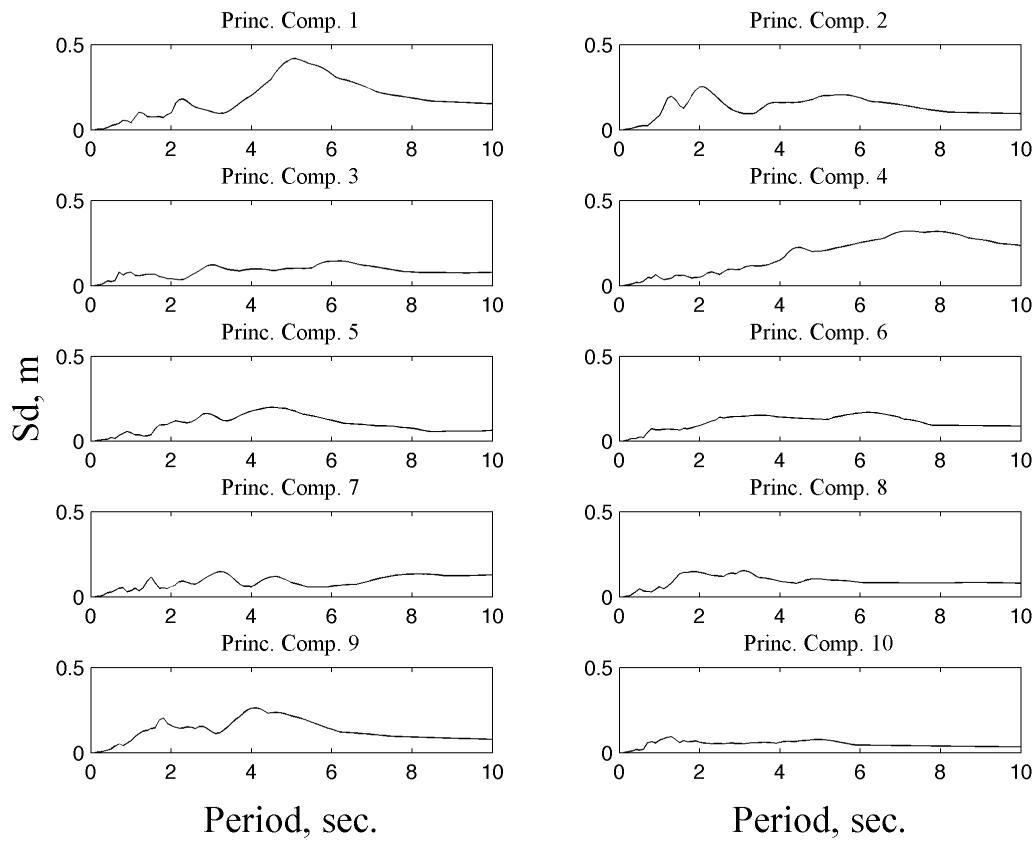


Figure 3.15. Displacement spectra of the top ten eigenquakes.

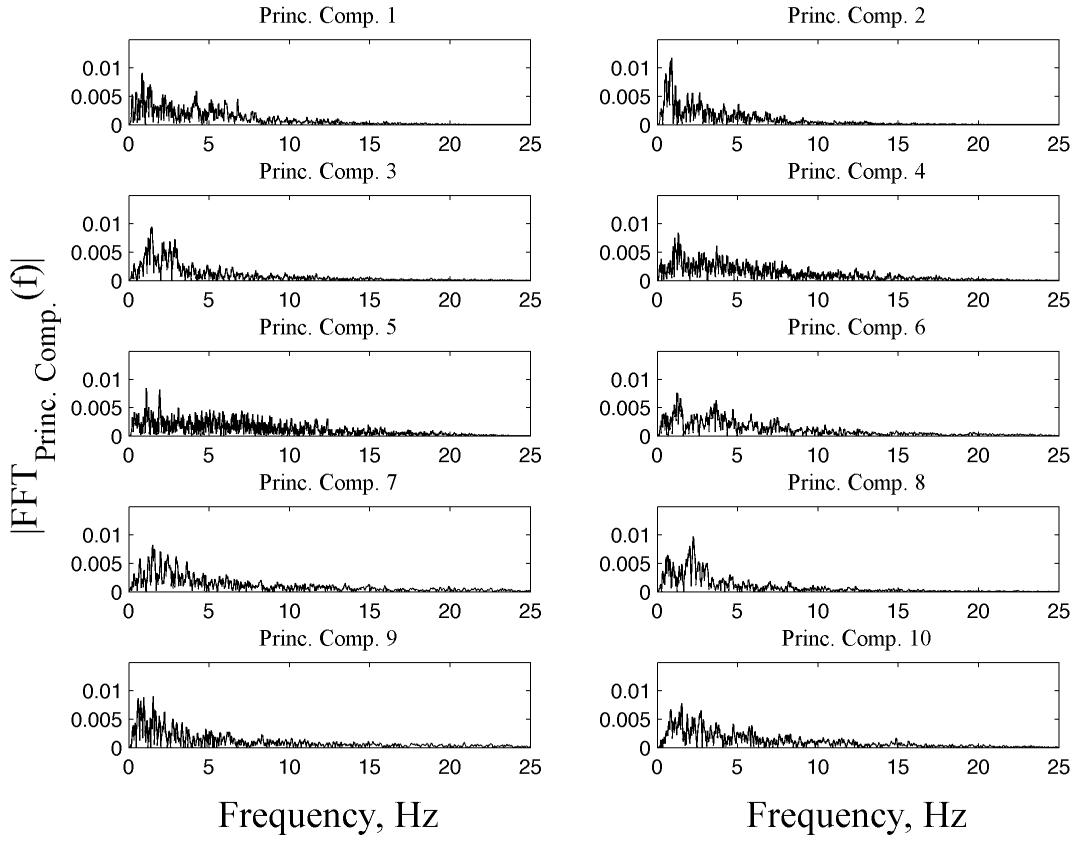


Figure 3.16. Fourier amplitude spectra of the top ten eigenquakes.

### *Time-Frequency properties*

The power spectral density spectrograms obtained from short-time Fourier transforms of the top 10 eigenquakes are shown in Figure 3.17. These spectrograms are calculated for the first 40 seconds of the eigenquakes with a Kaiser window of length  $2^8$  data points, making approximately eight moving windows along the

length of the signals. A sidelobe attenuation parameter of 0.5 is used with 78% overlap and FFT length of  $2^9$  data points (MathWorks). The sampling frequency is 50 Hz.

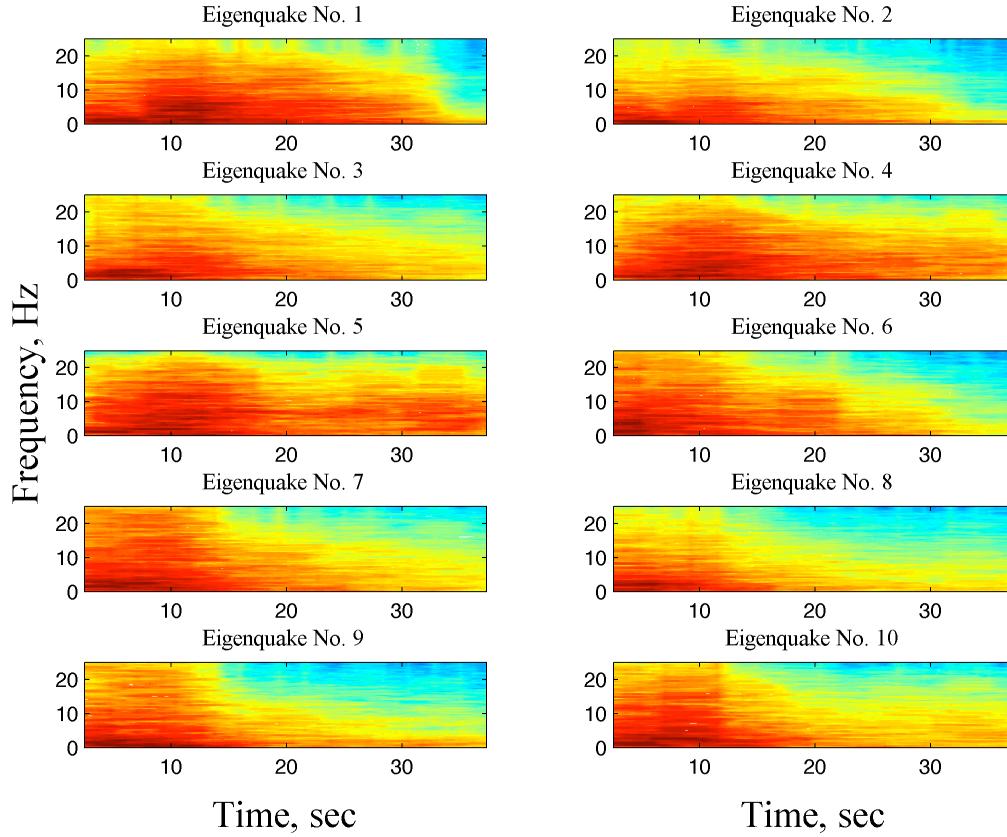


Figure 3.17. Power spectral density spectrograms for the top ten eigenquakes.

Similar to other properties in time and frequency domains already discussed, the time-frequency analysis also demonstrates the space of very diverse dynamics that could be spanned by the vectors of eigenquakes collectively due to their independent rich properties.

## Development of Target Hazard Spectra

It was mentioned earlier in Chapter 2 that Gaussian Processes (GPs) provide a rational framework for dealing with problems of model selection and over-fitting. With availability of data, a GP regression is easy to set up and use provided that posterior estimation is done for interpolation purposes as extrapolation into unknown territory would have unknown outcomes. The set of training data that the GP uses to adjust parameters of its kernel covariance matrix<sup>22</sup> consists of an ensemble of spectral acceleration values, for different magnitudes, distances, and site's shear wave velocity values, at a given site from previously recorded data. Examples of such posterior estimation are presented in Figures 3.18 to 3.20. These figures exemplify the capabilities of carefully designed GPs in reliably estimating the spectrum of motion of a process as complex as earthquakes using solely magnitude, distance, and the shear wave velocity as input parameters and assuming no parametric form as in Ground Motion Prediction Equations (GMPEs). The quality of estimates are good even at the lower tail of the distribution of spectral acceleration, as is shown for a small magnitude event at a far distance in Figure 3.19.

---

<sup>22</sup> See Equation (2.23)

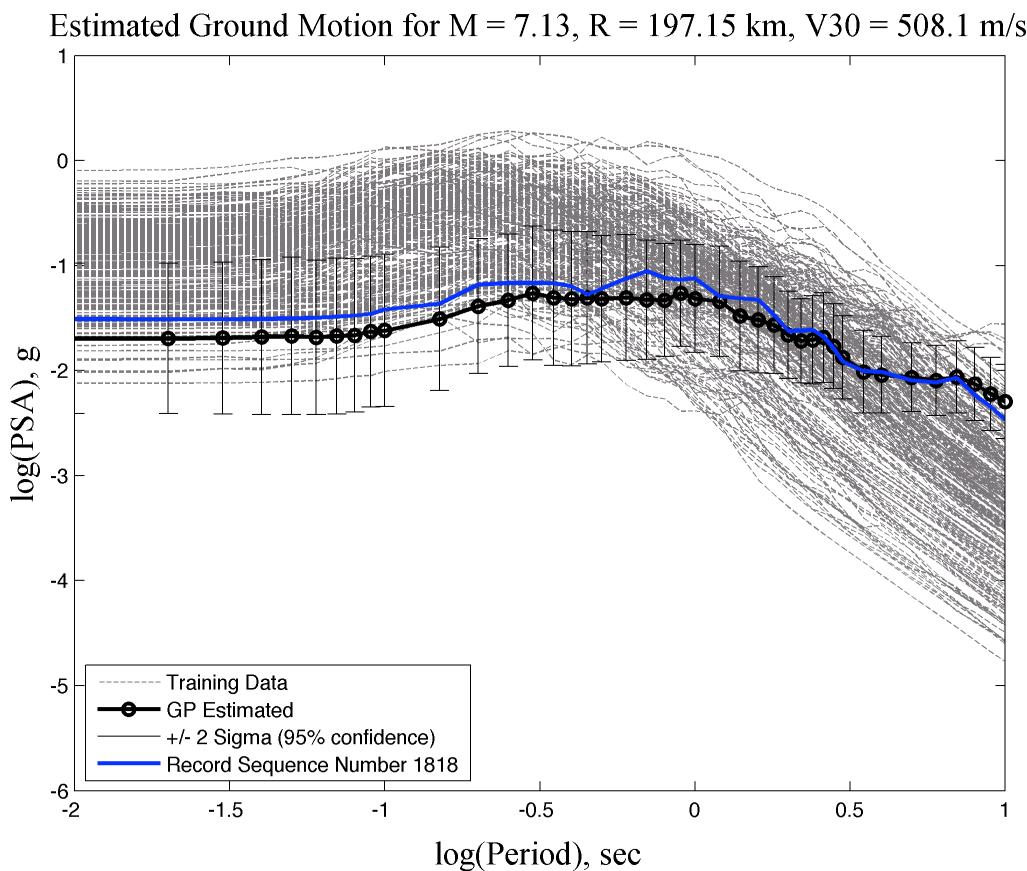


Figure 3.18. Estimated acceleration spectrum and confidence intervals for a scenario motion with its comparison with an actual event (Record Sequence Number 1818 is not used in training)<sup>23</sup>. The background dashed lines are previously recorded data used for the derivation of the kernel covariance matrix.

---

<sup>23</sup> M = Moment magnitude, R = Hypocentral distance, and V30 = Average shear wave velocity of the top 30 meters of the sedimentary layers at the site.

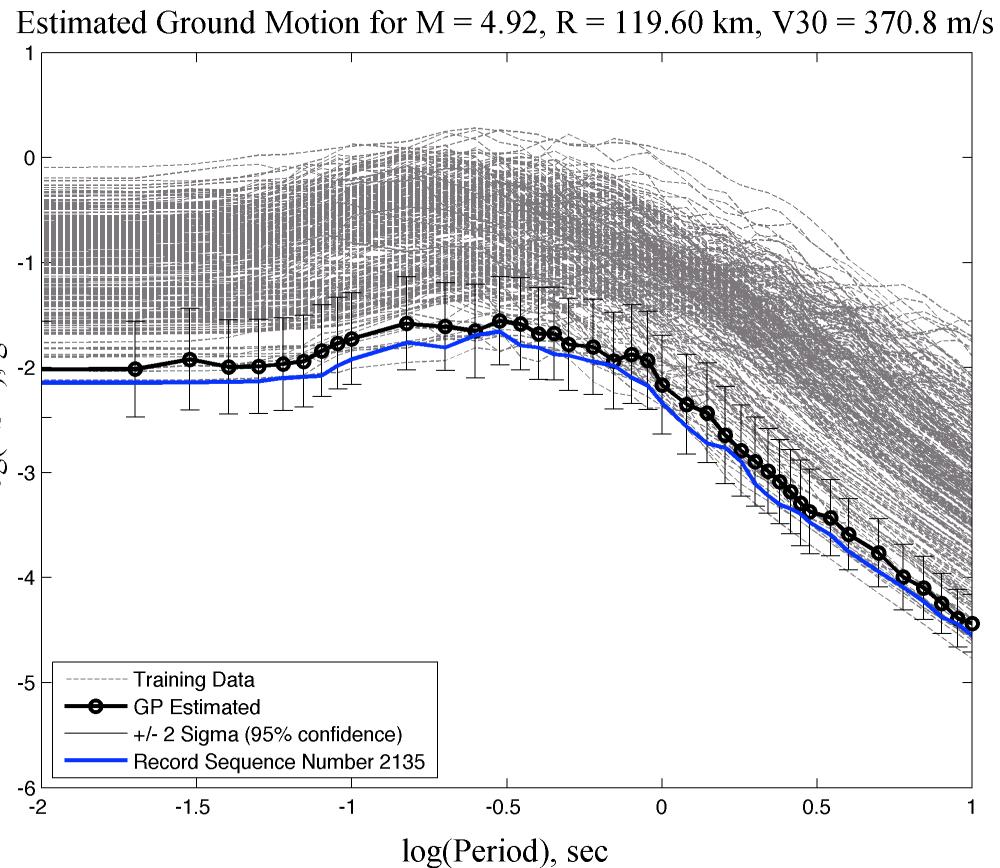


Figure 3.19. Estimated acceleration spectrum and confidence intervals for a scenario motion with its comparison with an actual event (Record Sequence Number 2135 is not used in training). The background dashed lines are previously recorded data used for the derivation of the kernel covariance matrix.

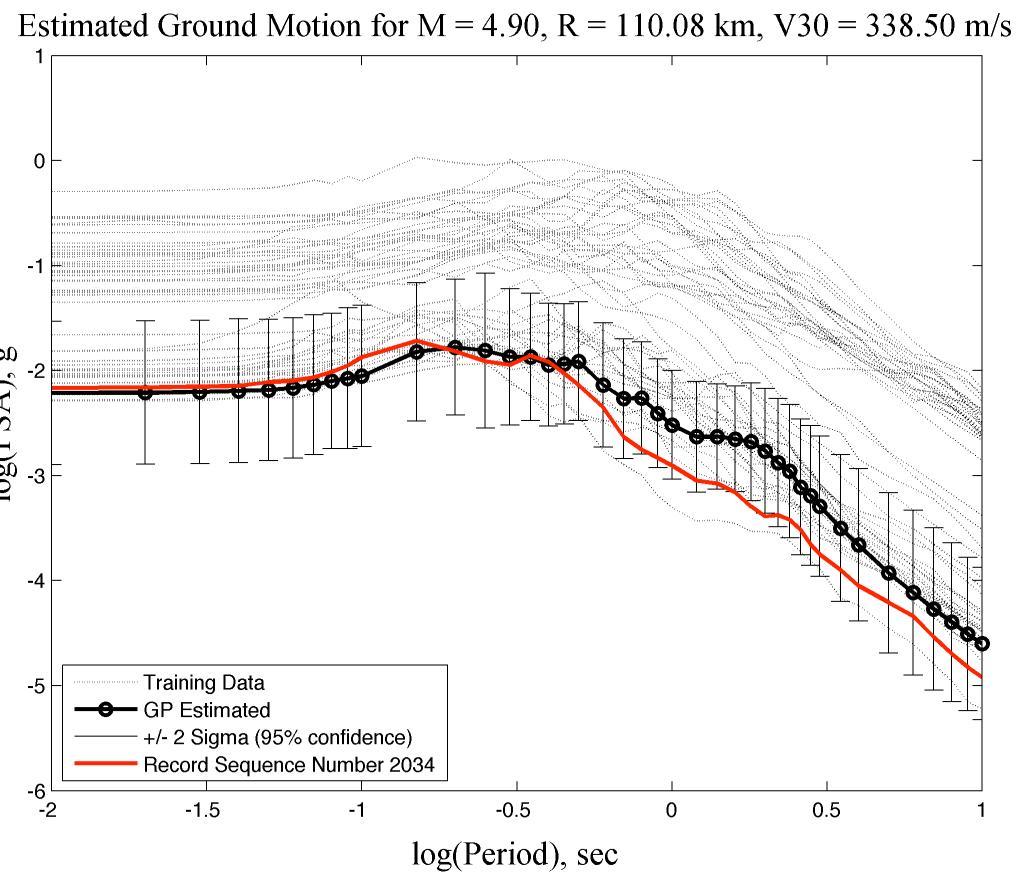


Figure 3.20. Estimated acceleration spectrum and confidence intervals for a scenario motion with its comparison with an actual event (Record Sequence Number 2034 is not used in training). The background dashed lines are previously recorded data used for the derivation of the kernel covariance matrix.

# Computational Implementation Issues

The number of required eigenquakes to represent significant variations in a large database could be generally sizable. Consequently, estimation of the coefficients of the eigenquakes would require large-scale robust global optimization which is computationally expensive. The estimation of the coefficients would have been much more efficient in the frequency domain, as depicted in Figure 3.21, but the non-uniqueness of the mapping between spectral accelerations and the acceleration time histories would require the use of optimization in time domain.

We use Genetic Algorithms (GA), known as robust stochastic optimization methods, to solve for the coefficients of the eigenquakes in high-dimensional space:

$$a_i = \arg \min_T \sum \left\{ S_a \left[ \left( \sum_{i=1}^n a_i \cdot \lambda_i(t) \right), T \right] - S_a^*(T) \right\}^2 \text{ for } \forall T \in [0, T_{\max}], a \in \mathbb{R}, i = 1, \dots, 140 \quad (3.3)$$

GAs have been used extensively in many structural mechanics and earthquake engineering applications over the past 25 years (Chan 1997; Raich and Ghaboussi 2000; Chou and Ghaboussi 2001; Kim and Ghaboussi 2001; Alimoradi, Miranda et al. 2006; Alimoradi and Naeim 2006; Alimoradi, Pezeshk et al. 2007; Foley, Pezeshk et al. 2007). GAs theory and operation are presented elsewhere (Goldberg 1989), nevertheless, a brief description is given here for comprehensiveness.

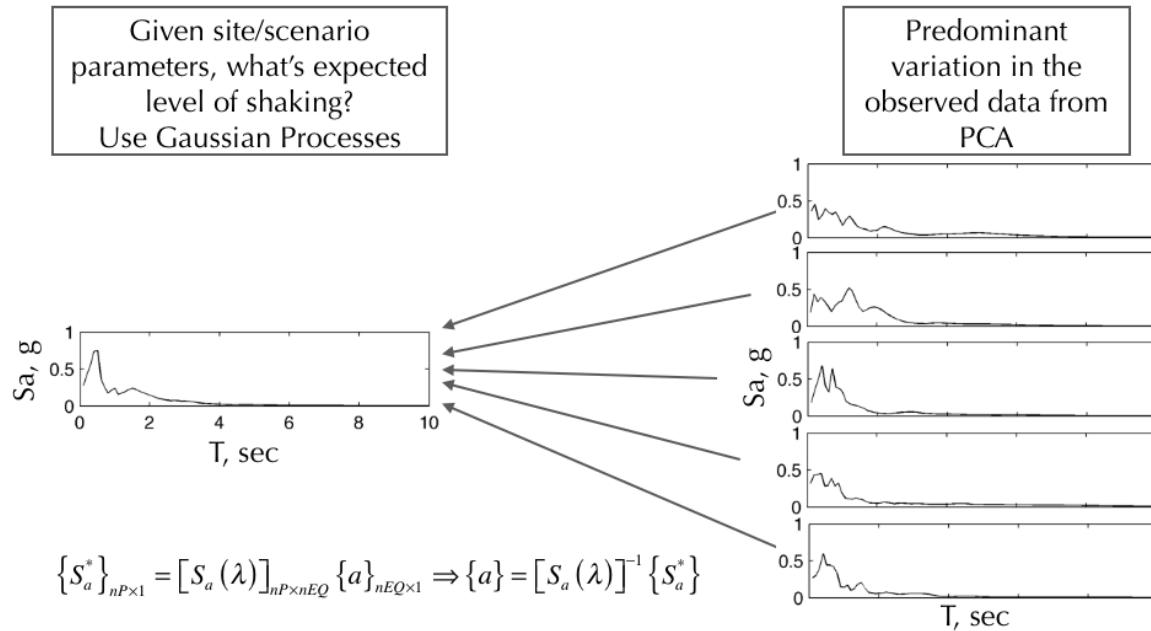


Figure 3.21. Optimization could have been avoided for the estimation of the coefficients of the eigenquakes in frequency domain by using simple matrix operations if the mapping between spectral accelerations and the acceleration time histories was linear and one-to one - see Chapter 2.

## Genetic Algorithms and Parallel Computing Schemes

Genetic algorithms belong to a group of numerical optimization methods<sup>24</sup> collectively called *Nature-Inspired Algorithms* which includes other methods such as Artificial Immune Systems, Ant Colony Optimization, Harmony Search, Big Bang-

---

<sup>24</sup> As one of the earliest variations of such methods.

Big Crunch Optimization, Swarm Optimization, etc. These algorithms use adaptation and simulation of different natural processes effective in search and optimization on a computer. Genetic algorithms are rough computational models of natural evolution. Survival of the fittest and evolution of the population of species through generations is an effective search strategy. Genetic algorithms start by initiating a randomly generated *population of chromosomes*, which when decoded, create the optimization variables in a search space. The chromosomes must compete for survival in an environment with the measure of their fitness being their optimality. Each chromosome is, therefore, tested in a *fitness function* to determine its likelihood for survival to the next generation. Those that survive, stochastically selected from the previous generation, reproduce the population of offspring by the actions of *crossover* and random occasional *mutation*. The process of creating new generations is continued until a measure of convergence to an optimal solution is met.

Genetic algorithms are considered a powerful and efficient class of methods in stochastic search. They are efficient because of the action of the populations of chromosomes (different candidate solutions) can be done in parallel, which is particularly suitable for high-performance computing. They are also powerful because they implement both exploitation and exploration in the search space to locate the optima. As a result, eventual convergence to a solution is almost always guaranteed.

A pseudocode for the GA operations would look like this<sup>25</sup>:

```
Initialize the population
Evaluate initial population
Repeat
    Perform competitive selection
    Apply genetic operators to generate new solutions
    Evaluate solutions in the population
Until some convergence criteria is satisfied
```

For the problem of optimal estimation of the coefficients of the eigenquakes using a genetic algorithm, the flowchart in Figure 3.22 is used to solve Equation (3.3). A population size of twice the number of the eigenquakes was used in the simulations. The crossover ratio was set to 0.8 (the fraction of the population, excluding elite chromosomes, created by the crossover operation) and a Gaussian mutation was used (to alter the chromosomes by adding a random number from a Gaussian distribution as a means of exploration in the space). The value of these parameters was chosen empirically and by trial and error. Performance of a genetic algorithm depends on these parameters but development of strategies for fine tuning the performance of genetic algorithms is outside the scope of this study. Generally, performing robust global high-dimensional optimization of a nonlinear objective function is a very challenging task. The large number of optimization variables (coefficients of 140 eigenquakes) in a highly nonlinear objective function whose

---

<sup>25</sup> From Evolutionary Algorithms: Genetic Algorithms, Evolutionary Programming and Genetic Programming, <<http://www.cs.sandia.gov/opt/survey/ea.html>> accessed on May 2011.

evaluation requires many dynamic analysis runs make the problem even more daunting. This reliance on high-dimensional optimization is a drawback of the methodology. There may be cases of convergence to local optima, or in other cases, long computational times may be needed for converge to alternative solutions. Some examples of good convergence and poor convergence are shown in Figures 3.23 to 3.26.

### *Parallel Computing Scheme*

It was mentioned earlier that individual chromosomes in a population of a genetic algorithm can be evaluated in parallel. This property makes genetic algorithms particularly suitable for parallel computing. We used Caltech’s “Mind-Meld”, a local experiments server, for our simulations (CACR 2011). Mind-Meld specifications are given in Table 3.2 (CACR 2011). Input files and code can be created locally on a machine and communicated with the run server via secure shell protocol using rather easy syntax:

```
ssh -X -o ServerAliveInterval=60 mind-meld.cacr.caltech.edu
      (requires a pass-phrase!)
use matlab
matlab (to run matlab)
do what you normally do in matlab
quit (matlab)
logout (from mind-meld)
exit (from ssh)
```

The genetic algorithm should also be configured so that it uses a pool of processors for population fitness evaluations. This can be done in Matlab by setting up the solver options to (MathWorks):

```
options = gaoptimset('UseParallel', 'always', 'Vectorized', 'off');
```

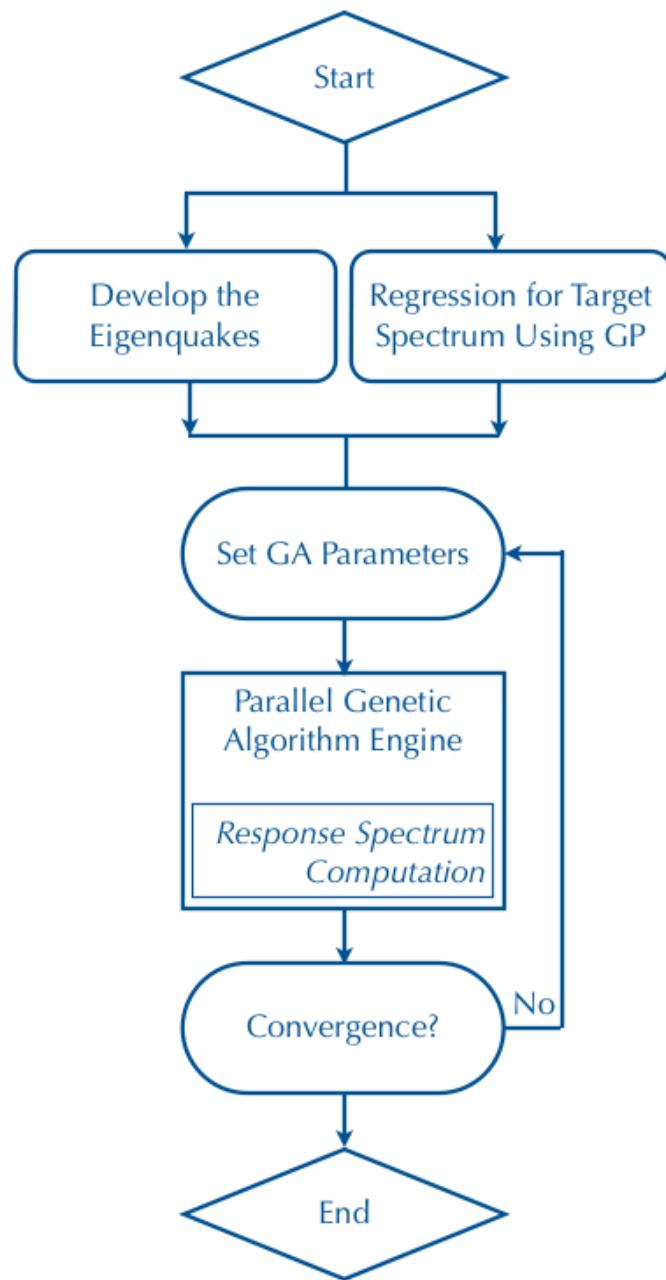


Figure 3.22. Flowchart of the procedure.

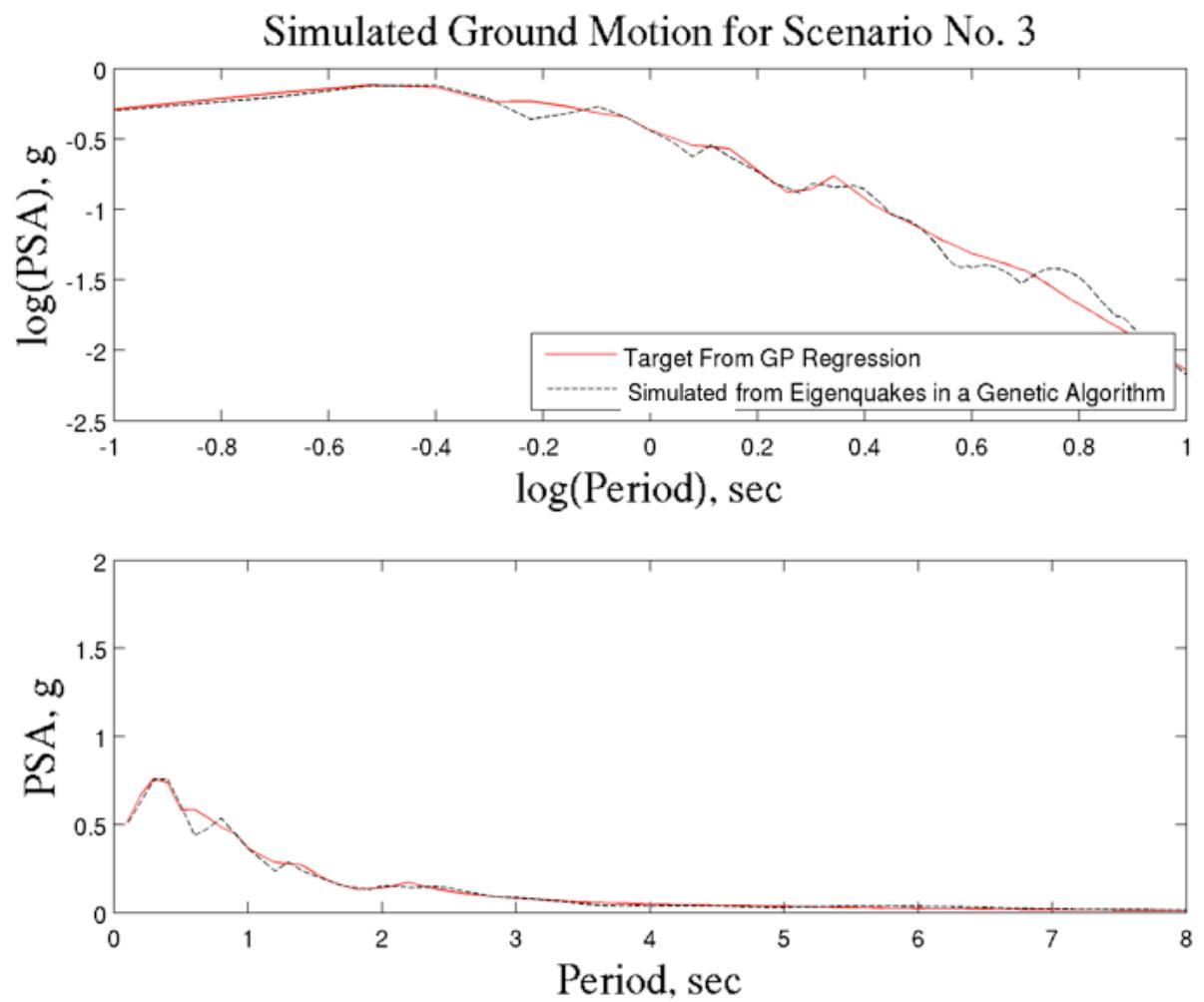


Figure 3.23. An example of convergence to a good solution.

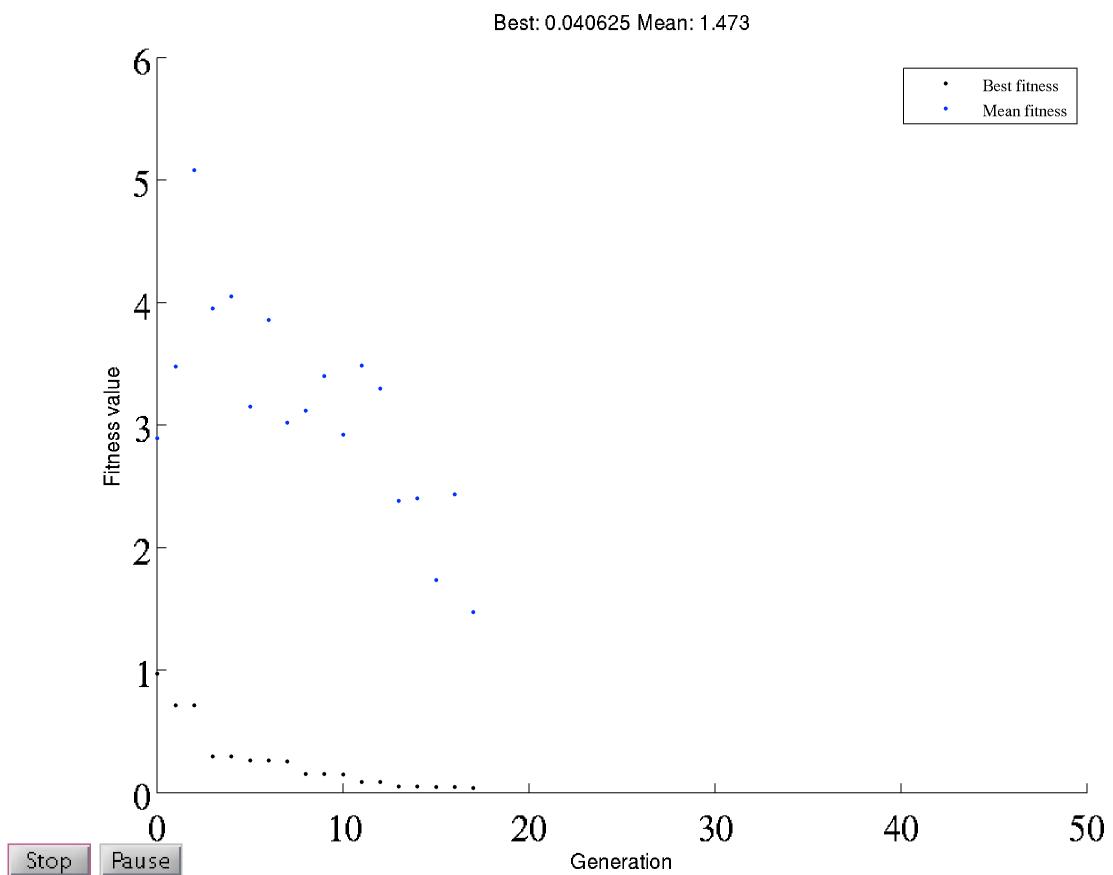


Figure 3.24. Trajectories of solutions (best and average of each population) showing rapid convergence to a good solution for the simulation in Figure 3.23 (MathWorks).

### Simulated Ground Motion for Scenario No. 2

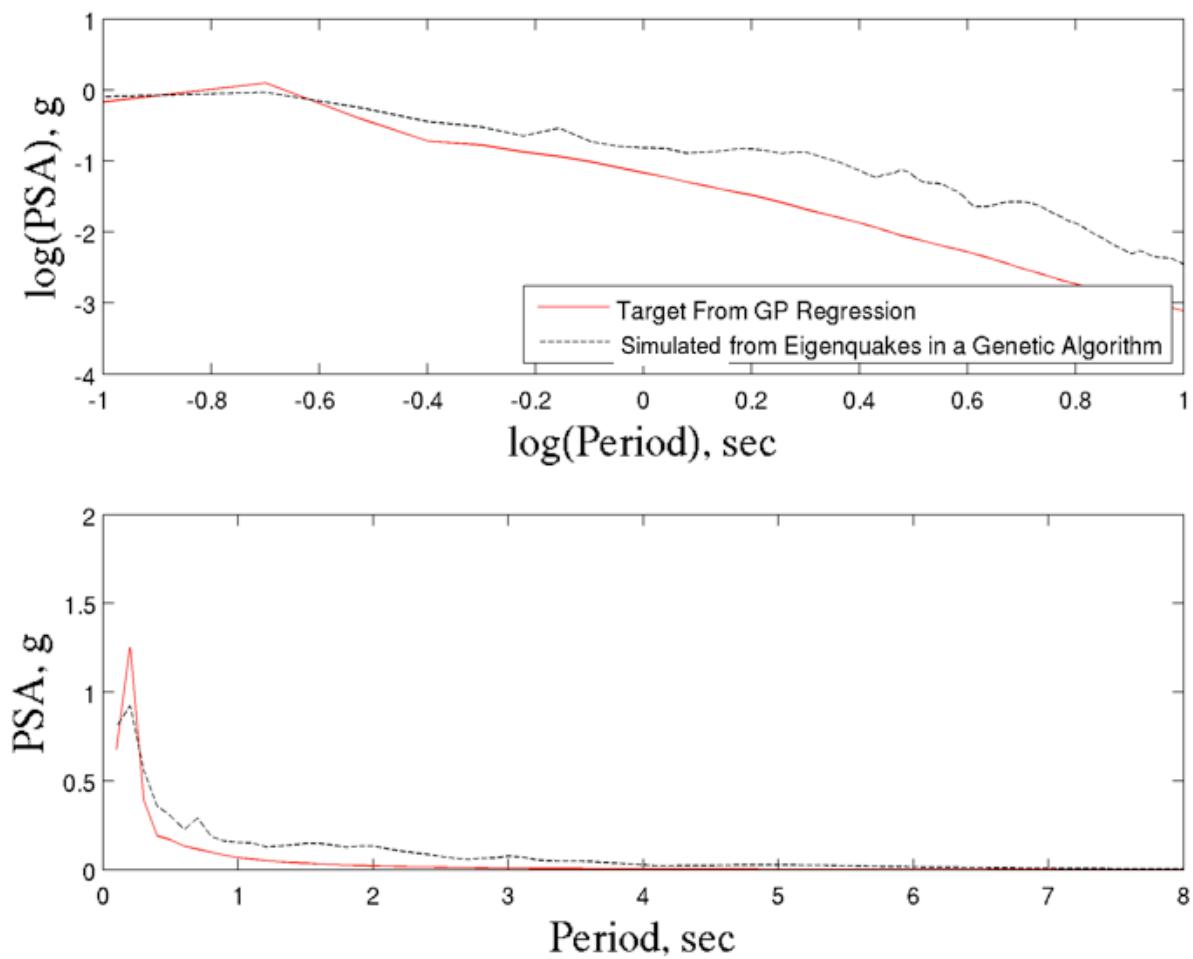


Figure 3.25. An example of convergence to a poor solution.

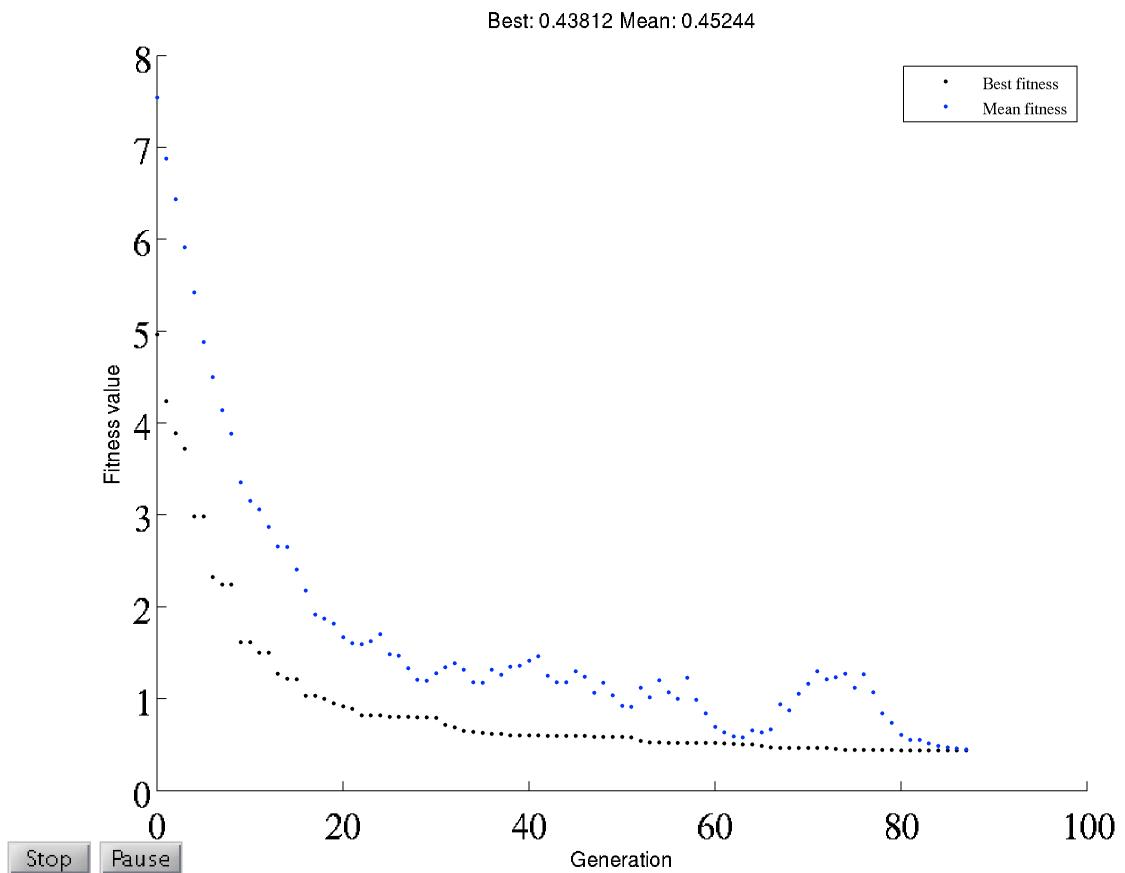


Figure 3.26. Trajectories of solutions (best and average of each population) showing good convergence to a poor solution for the simulation in Figure 3.25 (MathWorks).

Table 3.2. Caltech CACR's Mind-Meld Sun Fire X4600 specifications (CACR 2011).

Processor	8 AMD Opteron 8220 dual core processors, 2.8 GHz
CPU Interconnect	HyperTransport speed: 1GHz, 8 GB/second
Cache	1 MB Level 2 cache per core
Operating System	Red Hat Enterprise Linux

# Chapter 4 Development of Ground Motion Ensembles

To demonstrate the capabilities of the proposed methodology, simulation of ground motion acceleration histories for eight scenario events<sup>26</sup> are performed in this Chapter. The scenarios are shown in Table 4.1. A pair of a moment magnitude and a hypocentral distance constitutes a scenario. The scenarios are planned for two sites; one in downtown Los Angeles and the other in downtown San Francisco. At each site two sets of records are simulated, one for a frequent event with less than 50% probability of exceedance (POE) in 50 years and the other for a rare event that has less than 2% probability of being exceeded in 50 years. Also, for each probability of exceedance, simulations are presented corresponding to two different periods; 0.2 seconds representing a typical short structure<sup>27</sup>, and 2.0 seconds representing a mid-rise building<sup>28</sup>. Ground motion scenarios can be obtained from model-independent

---

<sup>26</sup> Scenarios are controlling events for design.

<sup>27</sup> Approximately a two story building.

<sup>28</sup> Approximately a twenty story building.

deaggregation of seismic hazard. For example, USGS interactive deaggregations calculator<sup>29</sup>, shown in Figure 4.1, uses coordinates of a site in addition to probability of exceedance to generate a set of scenarios. By choosing “No” for the “GMPE Deaggs” option, mean hazard is deaggregated independently from any model of ground motion prediction equations. A specification of spectral period and the site’s soil condition is also required. The process of generating the simulations and the results are discussed next.

Table 4.1. The list of scenarios considered for ground motion simulation; T is period, M is the moment magnitude, R is the hypocentral distance, and V<sub>30</sub> is the average shear wave velocity at the top 30 meters of the site’s sedimentary layers.

Scenario	Location	POE	T, sec	M	R, km	V <sub>30</sub> , m/s	Elapsed Time, min, (Min. Function Value, No. of Generations, No. of Eigenquakes)
S1	Downtown Los Angeles	2% /50Y	0.2	6.7	6.6	388	327 (0.14, 48, 40)
S2a			2.0	6.9	11	388	536 (0.44, 87, 100)
S2b			0.2	6.6	24		504 (0.31, 81, 20)
S3		50% /50Y	0.2	6.6	24	388	123 (0.04, 17, 40)
S4a			2.0	7.1	44	388	256 (0.98, 37, 40)
S4b			0.2	7.3	13		416 (0.18, 67, 20)
S5	Downtown San Francisco	2% /50Y	0.2	7.8	15	428	355 (0.14, 58, 20)
S6			2.0	6.9	20	428	463 (0.27, 76, 20)
S7		50% /50Y	0.2	7.2	33	428	275 (0.06, 44, 20)
S8			2.0	7.2	428	353 (0.09, 56, 20)	

<sup>29</sup> <https://geohazards.usgs.gov/deaggint/2008/> last accessed on May 2011.

FAQ   Documentation   1996 Update   2002 Update   Feedback

Site Name

[Enter address instead](#)

Latitude  Longitude

Exceedance Probability  in

Spectral Period

$V_{s30}$  (m/s)  [What values can I use at various locations?](#)

Run GMPE Deags?  Yes  No [What's this?](#)

Additional Output  Geographic Deagg [What's this?](#)  Conditional Mean Spectra  None

[\(Show Map\)](#)

Figure 4.1. USGS interactive deaggregations calculator<sup>30</sup>.

## Example 1: Downtown Los Angeles

The site is located at the intersection of 7<sup>th</sup> Street and Metro with a latitude of 34.04869 and a longitude of -118.258775 (Figure 4.2). Scenarios S1 to S4 correspond to this site with results of the deaggregation of seismic hazard shown in Figures 4.3 to 4.6. From these figures, the mean values of magnitude and distance are used for the simulations and are shown in Table 4.1. There are two sets of simulations for scenarios S2 and S4 because these scenarios were more challenging during high-dimensional optimization of the coefficients of the eigenquakes. Therefore, the

---

<sup>30</sup> <https://geohazards.usgs.gov/deaggint/2008/> last accessed on May 2011.

number of required eigenquakes was lowered from 100 in scenario S2a to 20 in scenario S2b and from 40 in scenario S4a to 20 in scenario S4b, for better convergence.

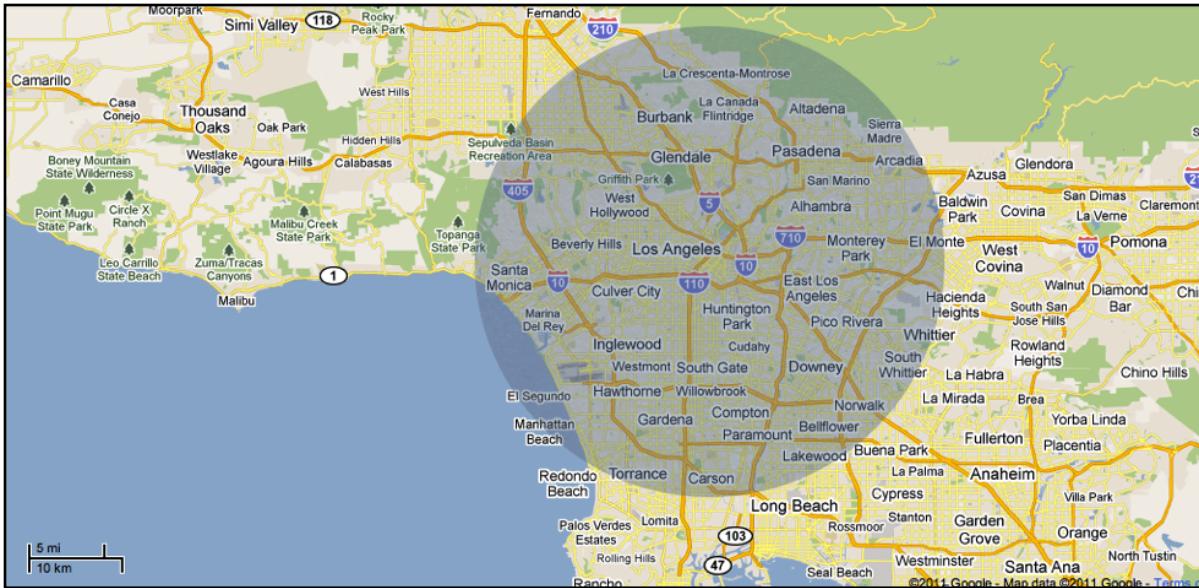


Figure 4.2. Location of the stations used in generating simulations S1 to S4<sup>31</sup>.

<sup>31</sup> Modified after GoogleMaps <<http://maps.google.com>> retrieved on May 2011.

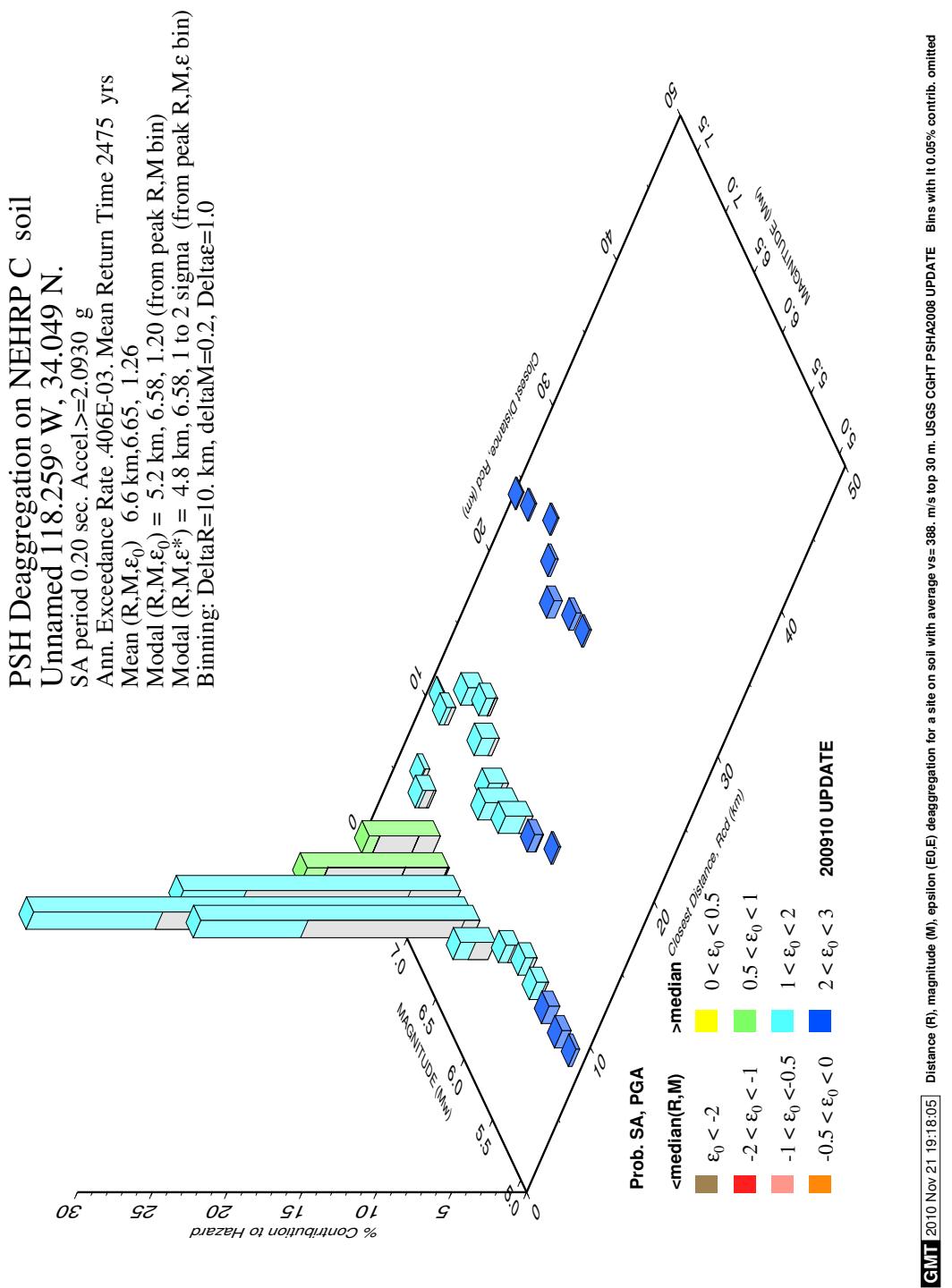


Figure 4.3. Deaggregation of seismic hazard for downtown Los Angeles for probability of exceedance of less than 2% in 50 years and period of 0.2 sec.

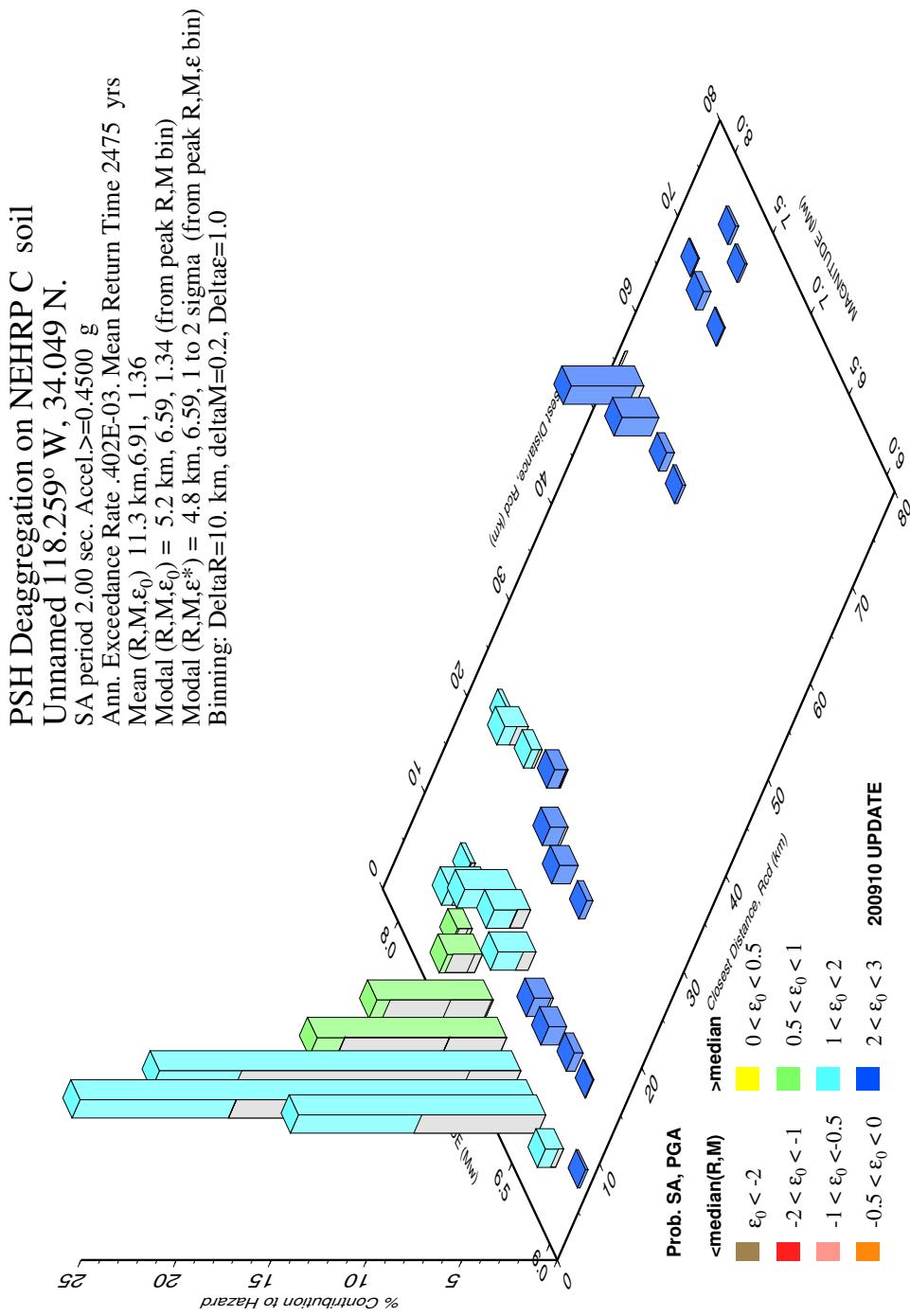


Figure 4.4. Deaggregation of seismic hazard for downtown Los Angeles for probability of exceedance of less than 2% in 50 years and period of 2.0 sec.

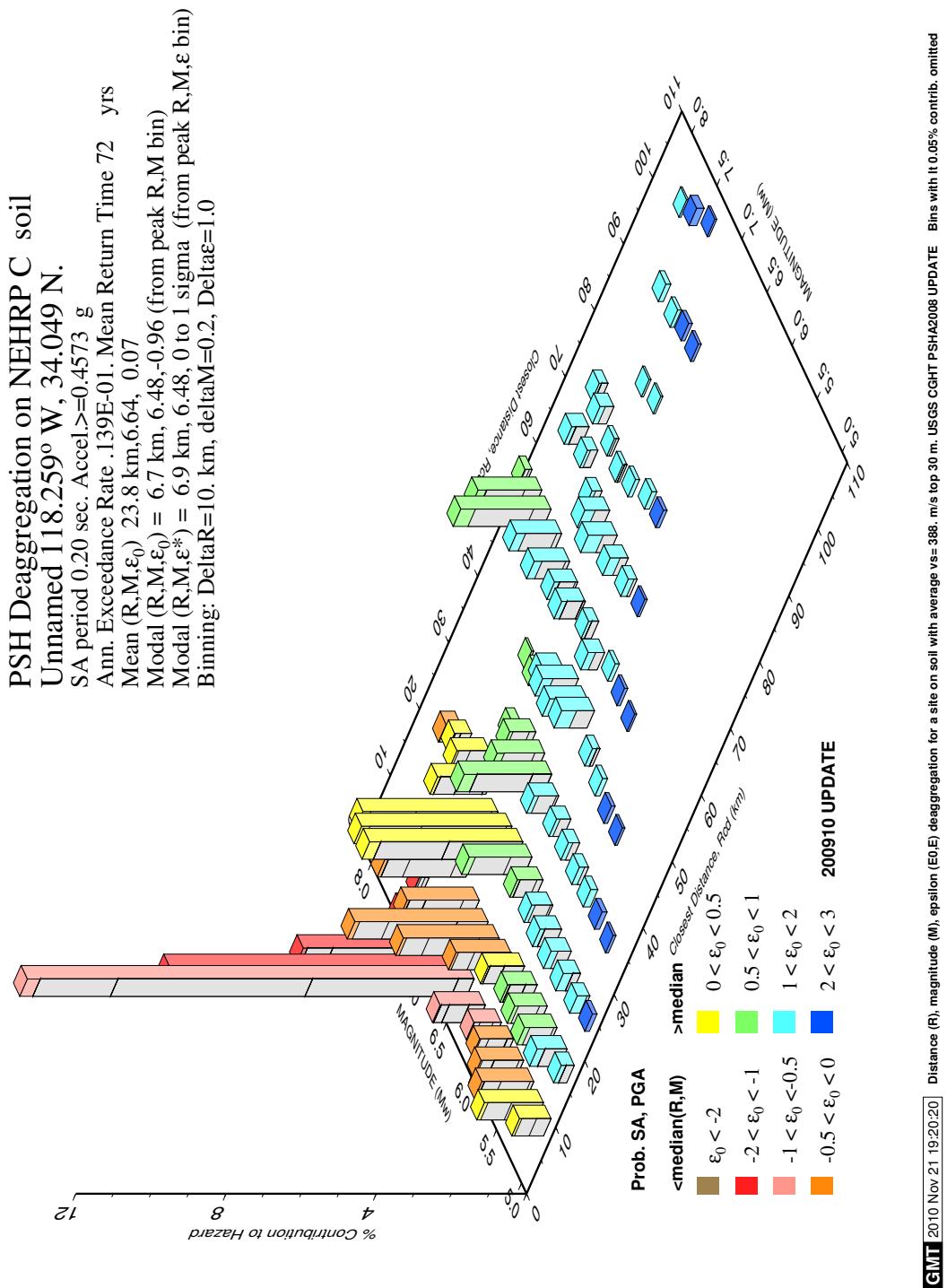


Figure 4.5. Deaggregation of seismic hazard for downtown Los Angeles for probability of exceedance of less than 50% in 50 years and period of 0.2 sec.

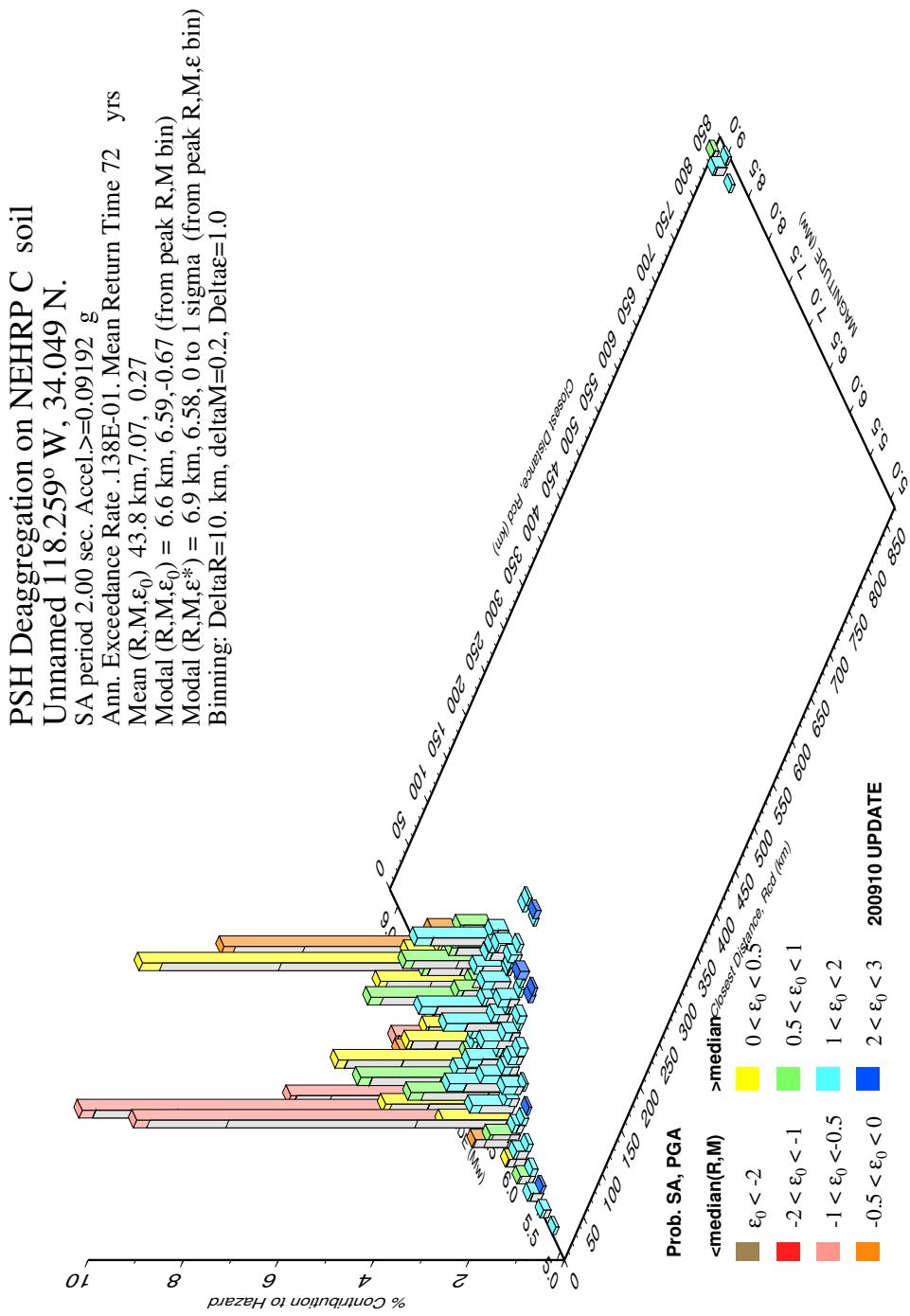


Figure 4.6. Deaggregation of seismic hazard for downtown Los Angeles for probability of exceedance of less than 50% in 50 years and period of 2.0 sec.

Earthquake events within a radius of 0.5 degree (approximately 50 km) from the PEER-NGA database were collected<sup>32</sup> and used in the Gaussian Process regression to arrive at mean spectral acceleration ordinates for different scenarios. The data that was implemented included events with moment magnitudes from 4.27 to 7.36 recorded from sources between 8.88 to 226.98 km<sup>33</sup>. The results are shown in Figure 4.7. Note different spectral shapes that are present at this site for different scenarios.

The optimal set of the coefficients of the eigenquakes for each scenario are shown graphically in Figure 4.8. Note that scenarios S2b and S4b were calculated with a set of 20 eigenquakes. The mean target spectra in Figure 4.7 show different spectral shapes and this difference is also reflected in the magnitudes of the optimal coefficients of the eigenquakes presented in Figure 4.8.

---

<sup>32</sup> 266 records.

<sup>33</sup> Not to be confused with the 50 km radius for selection of the stations around the site. The source to site distance can be larger than 50 km.

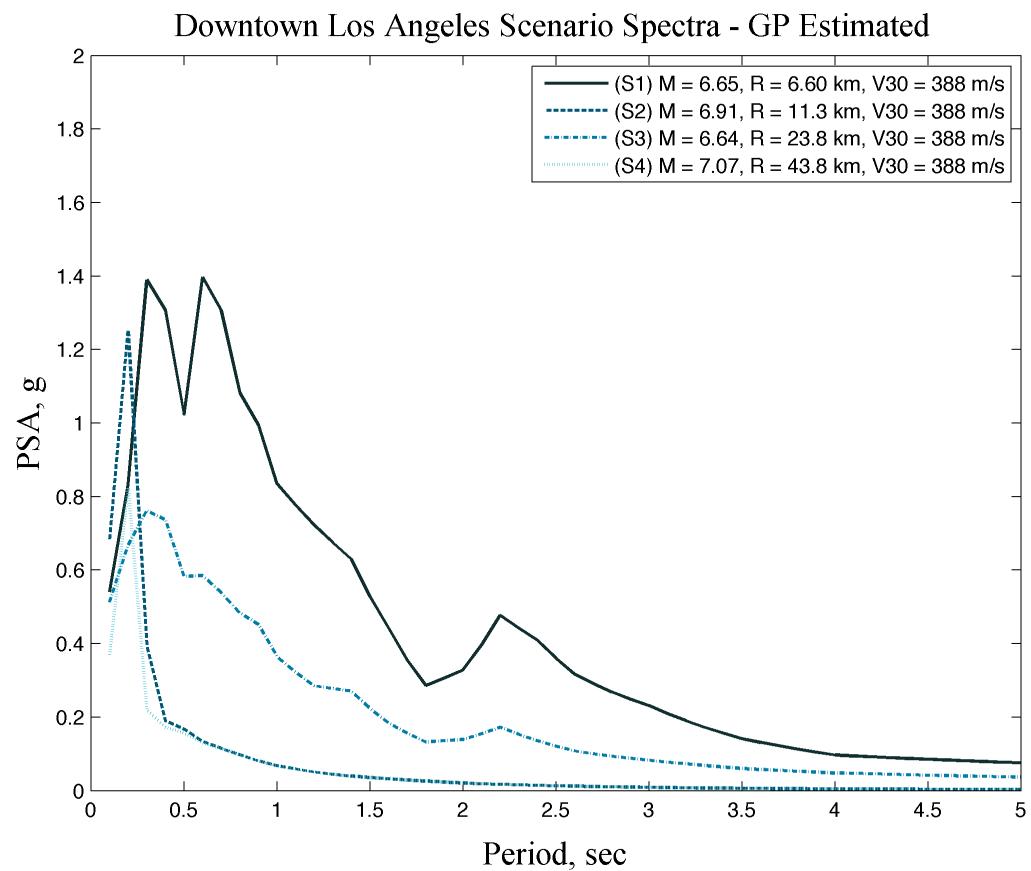


Figure 4.7. Mean target (controlling) response spectra for downtown Los Angeles for different scenarios.

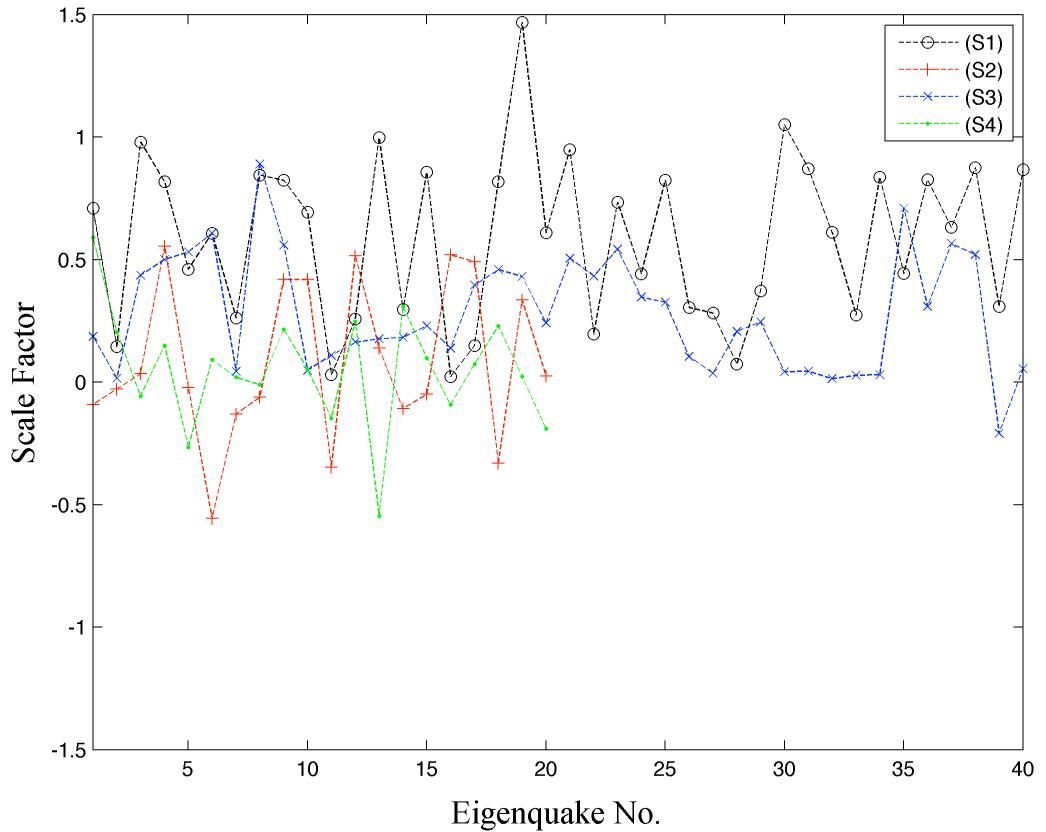


Figure 4.8. The optimal set of coefficients of the eigenquakes for different scenarios for downtown Los Angeles.

With an optimal set of coefficients, simulation of ground motion histories can be done simply by a linear combination of the eigenquakes. The results are shown in Figures 4.9 to 4.12 for spectral acceleration, acceleration, velocity, and displacement time histories, respectively. The difficulty in arriving at a global solution for optimization of the coefficients of the eigenquakes for scenarios S2 and S4 is evident in Figure 4.9.

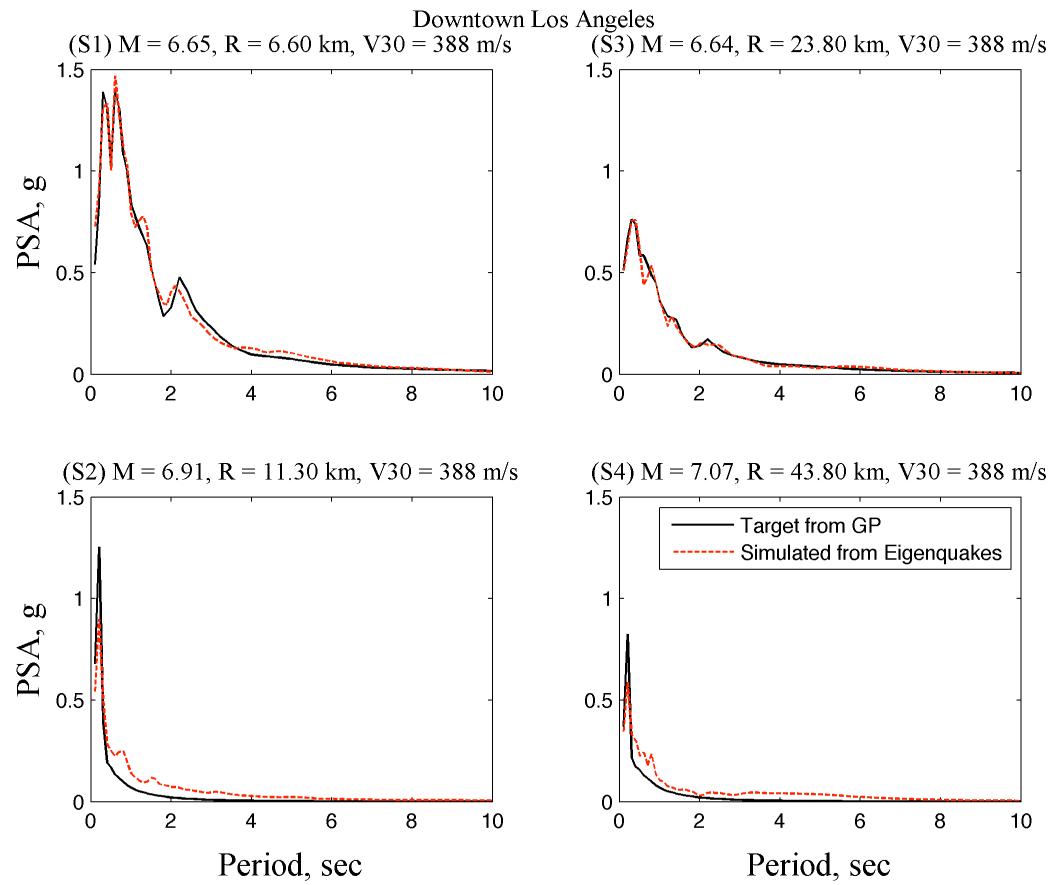


Figure 4.9. Target response spectra for downtown Los Angeles (black lines) and the response spectra from the simulated motions using an optimal combination of the eigenquakes (red dashed lines).

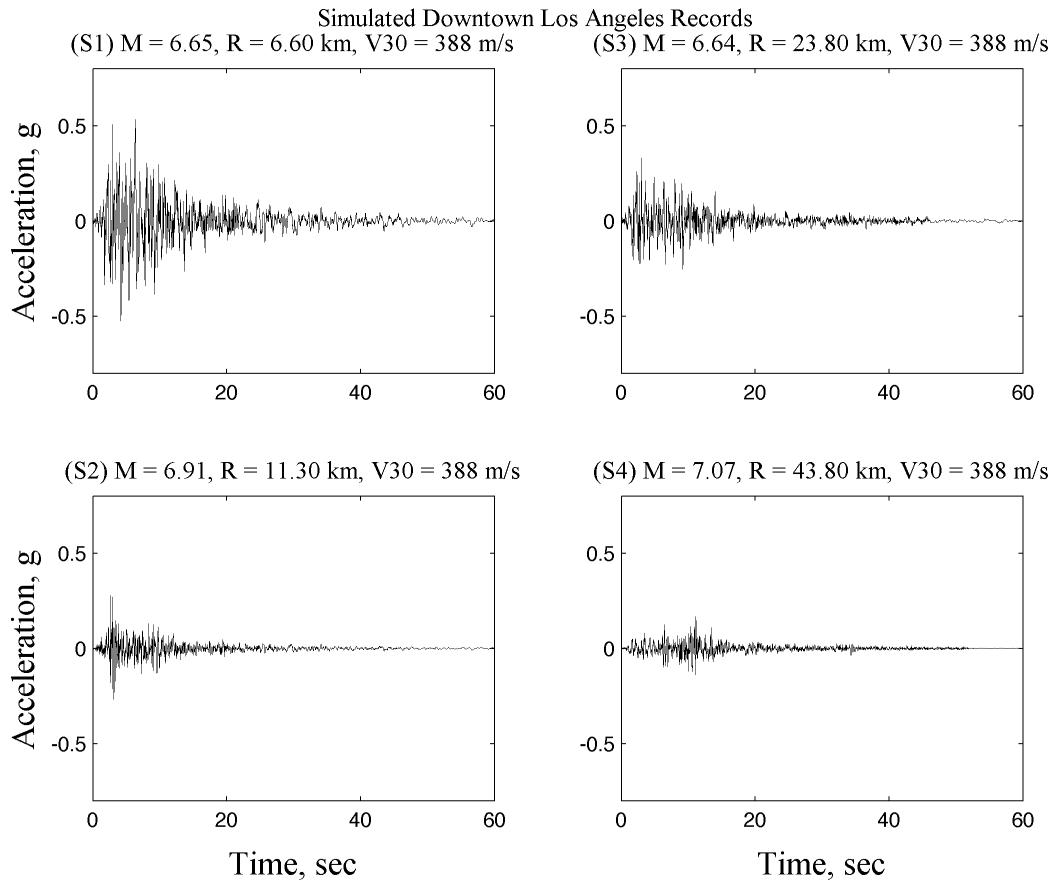


Figure 4.10. Simulated ground motions (acceleration time histories) for downtown Los Angeles.

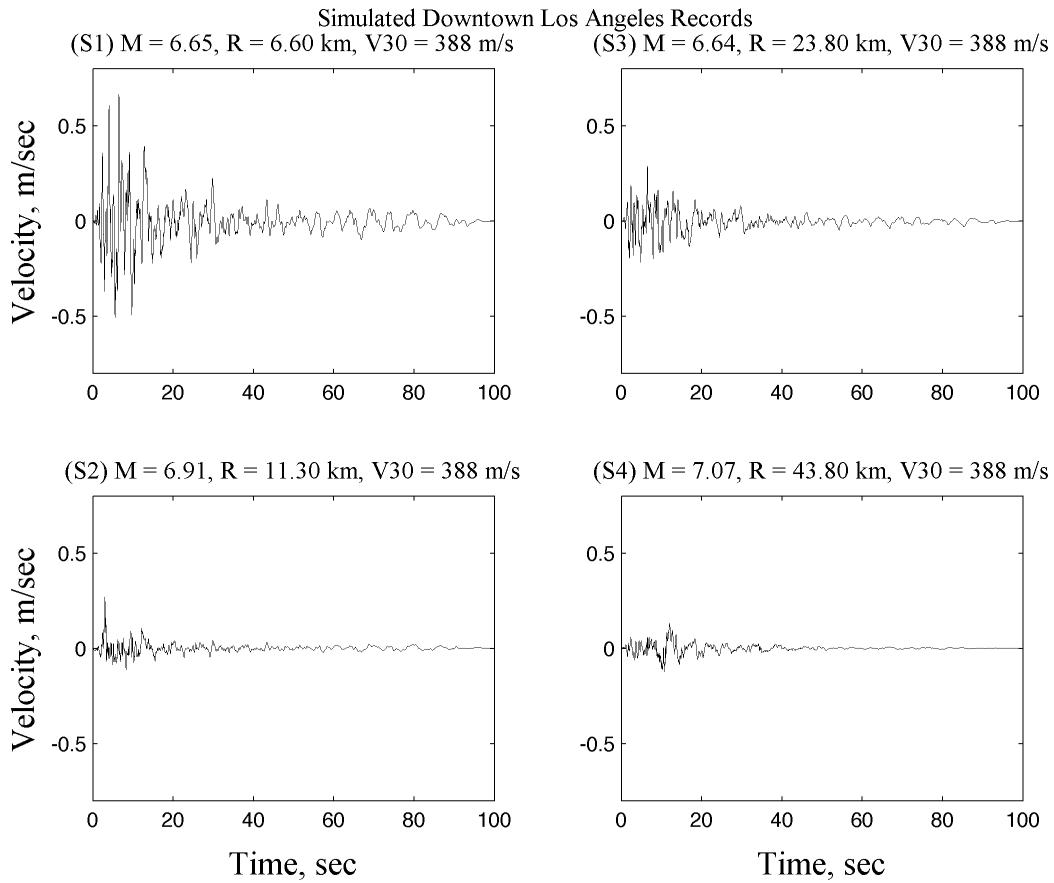


Figure 4.11. Simulated ground motions (velocity time histories) for downtown Los Angeles.

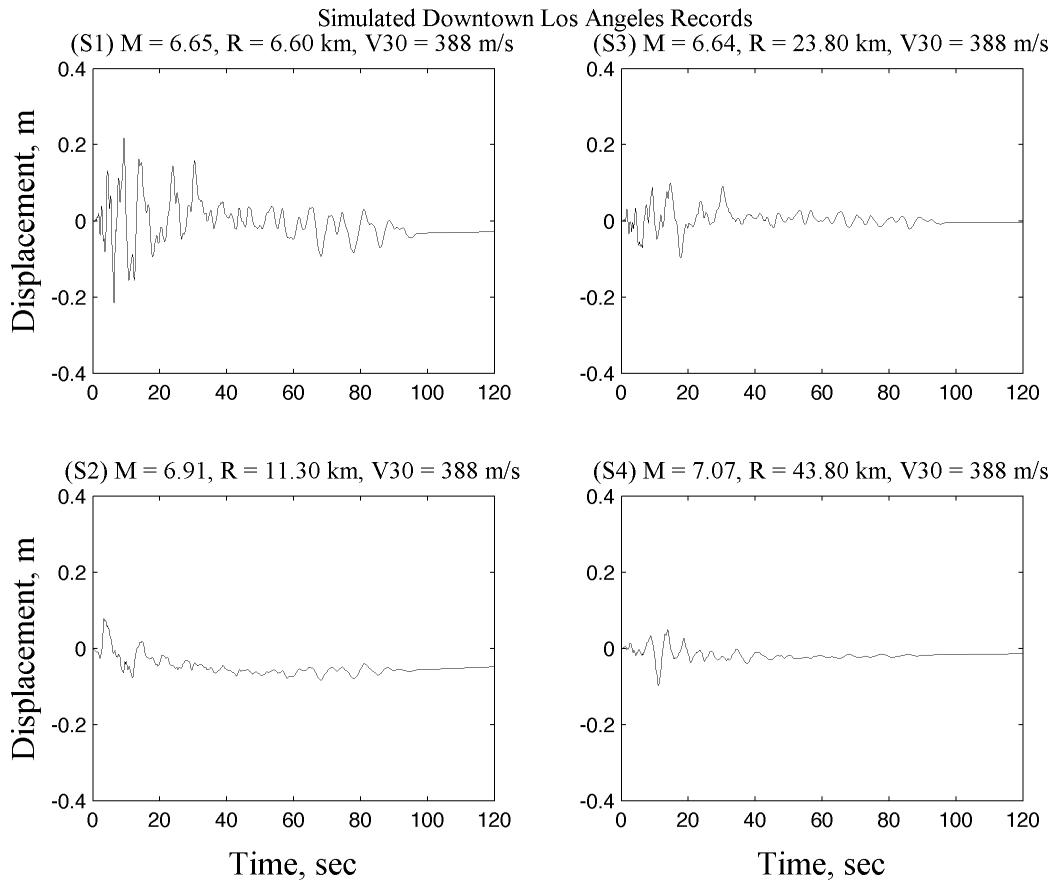


Figure 4.12. Simulated ground motions (displacement time histories) for downtown Los Angeles.

## Example 2: Downtown San Francisco

The site is located at the intersection of 4<sup>th</sup> Street and Market with a latitude of 37.78575 and a longitude of -122.405913 (Figure 4.13). Scenarios S5 to S8 correspond to this site with results of the deaggregation of seismic hazard shown in Figures 4.14 to 4.17. From these figures, the mean values of magnitude and distance are used for the simulations and are shown in Table 4.1.

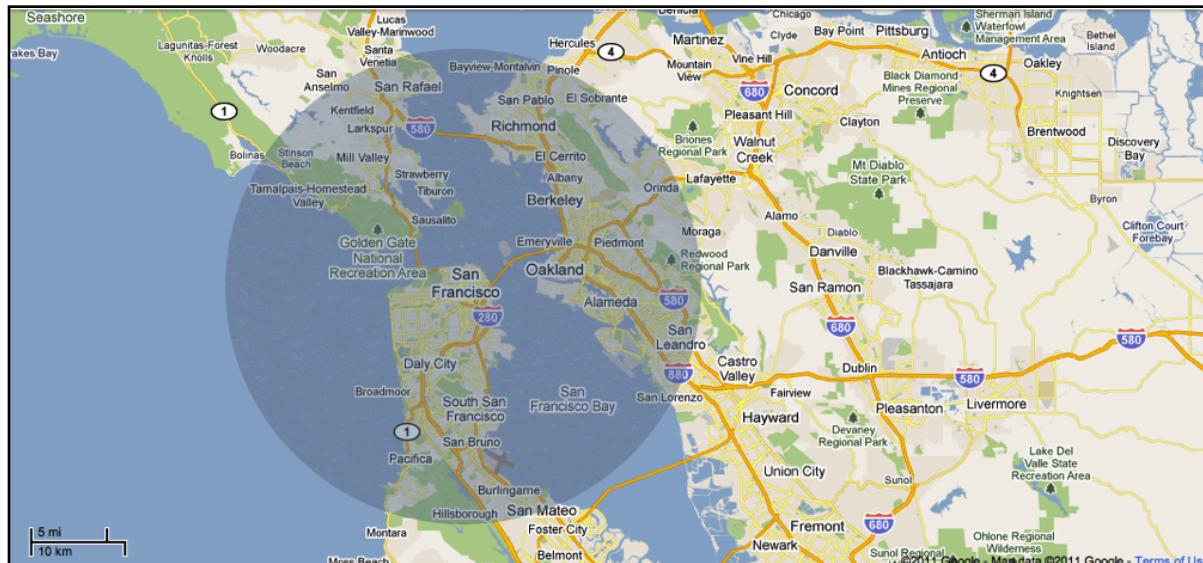
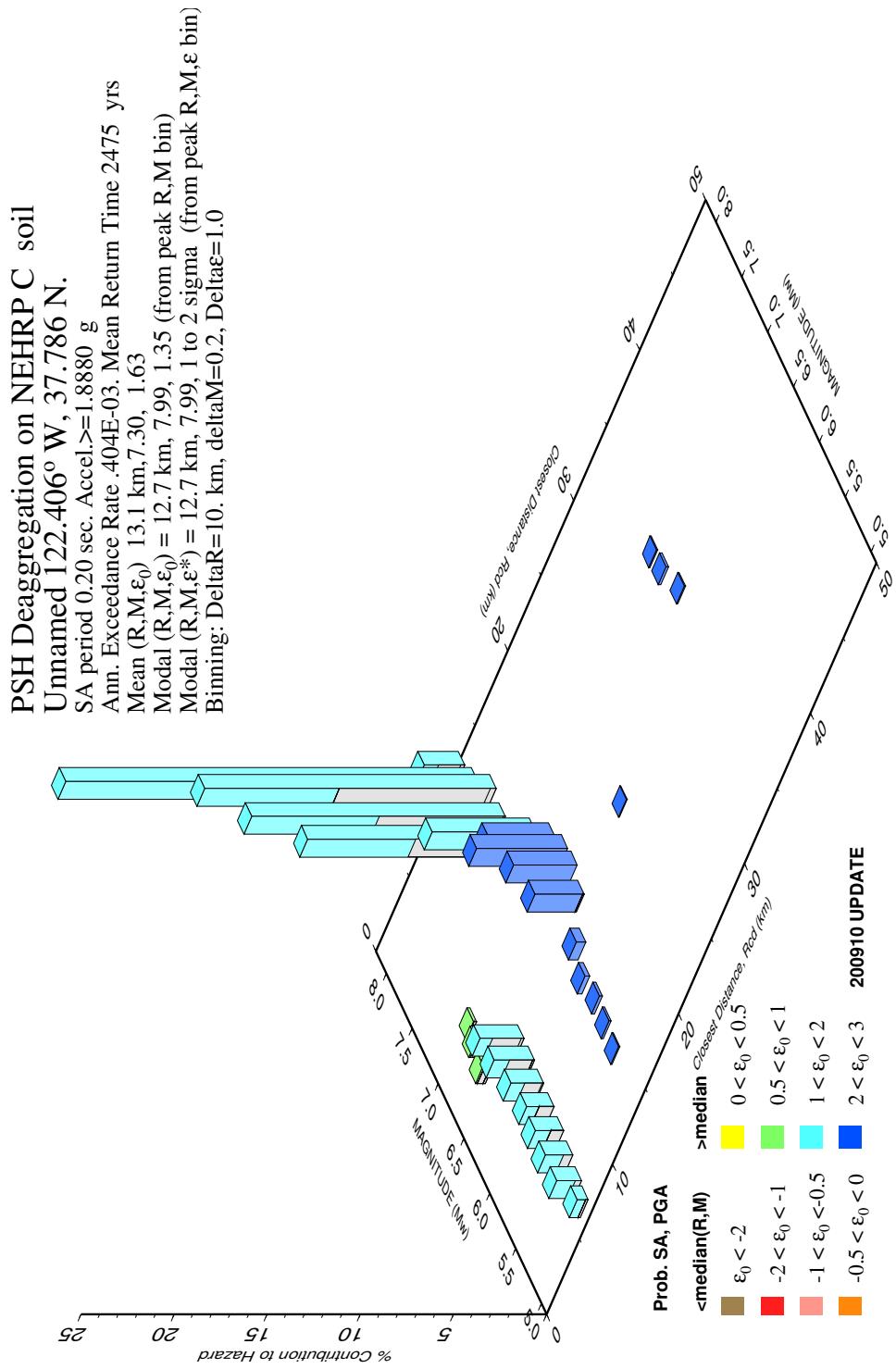


Figure 4.13. Location of the stations used in generating simulations S5 to S8 (from Google Maps).



GMT 2010 Nov 21 18:56:35 | Distance (R), magnitude (M), epsilon (E) deaggregation for a site on soil with average vs=428. m/s top 30 m. USGS CGHT PSHA2008 UPDATE | Bins with 0.05% contrib. omitted

Figure 4.14. Deaggregation of seismic hazard for downtown San Francisco for probability of exceedance of less than 2% in 50 years and period of 0.2 sec.

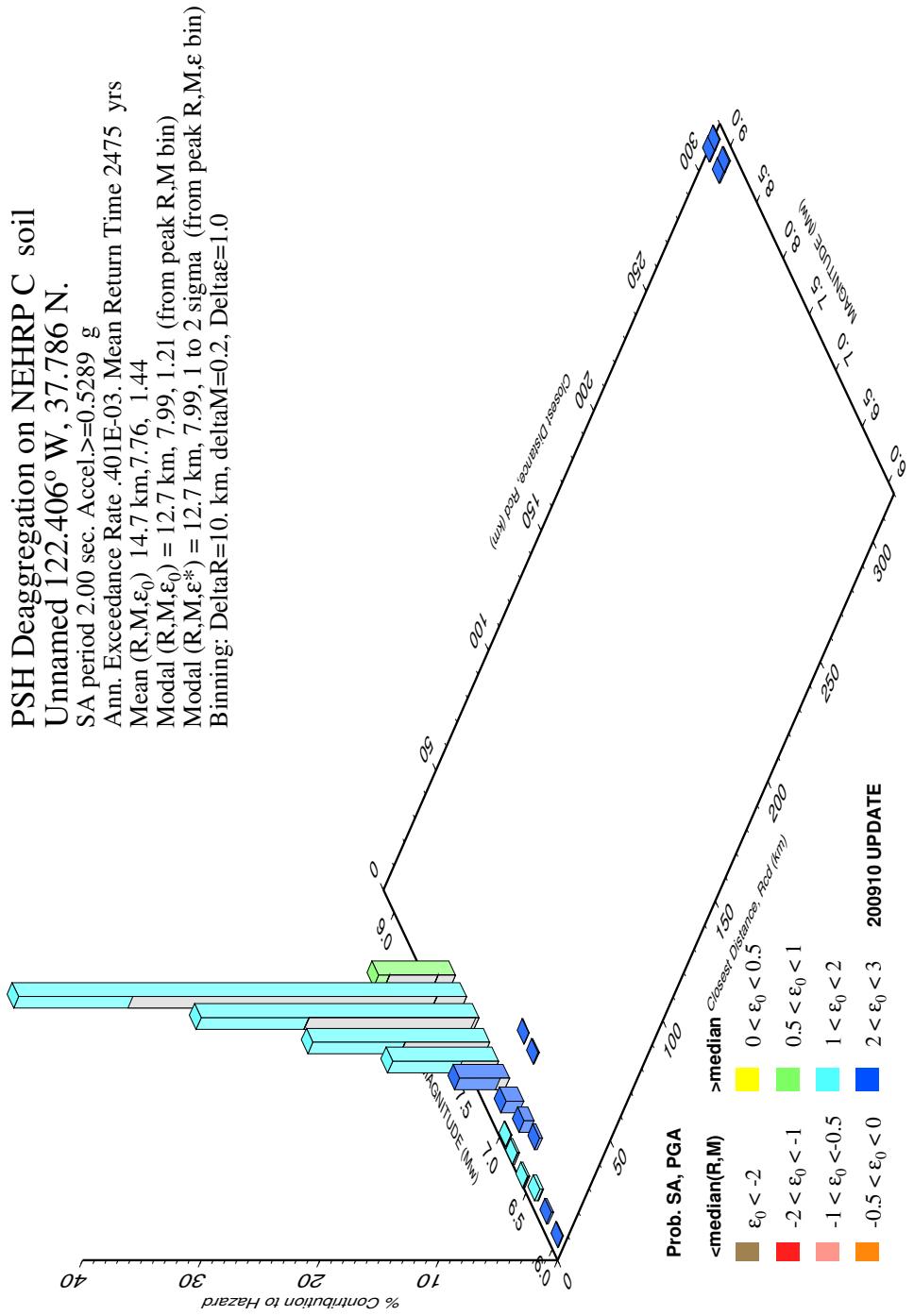


Figure 4.15. Deaggregation of seismic hazard for downtown San Francisco for probability of exceedance of less than 2% in 50 years and period of 2.0 sec.

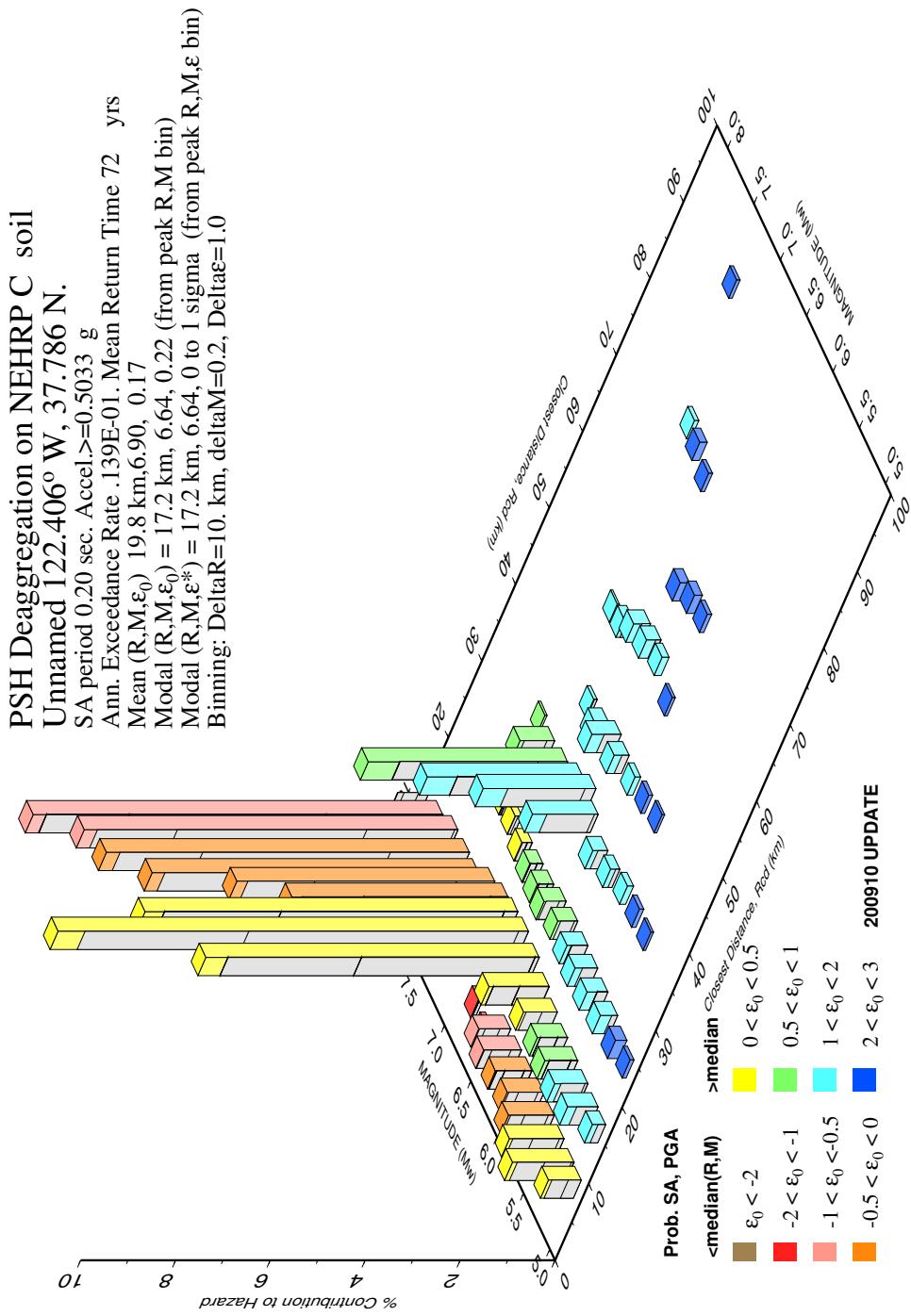


Figure 4.16. Deaggregation of seismic hazard for downtown San Francisco for probability of exceedance of less than 50% in 50 years and period of 0.2 sec.

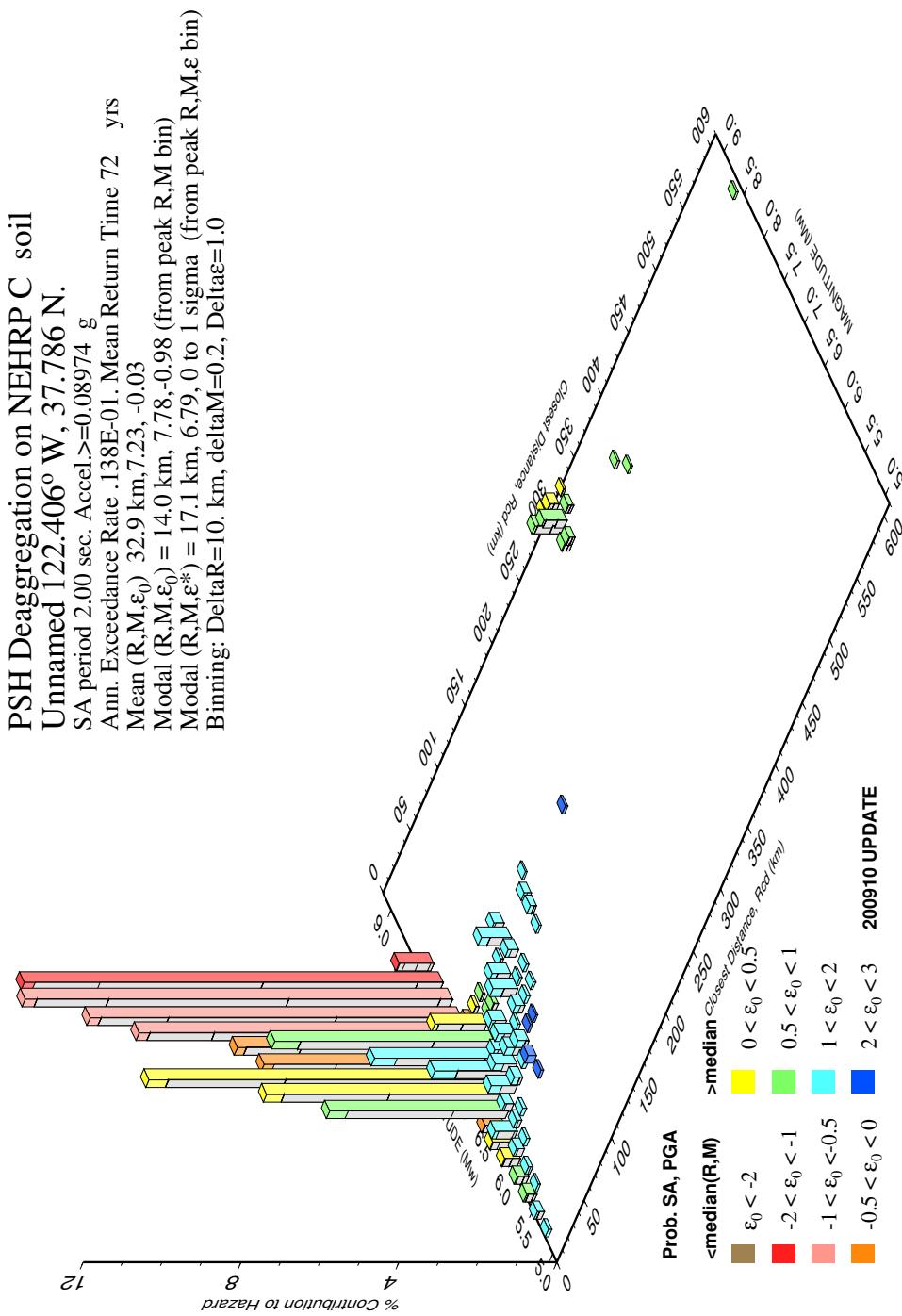


Figure 4.17. Deaggregation of seismic hazard for downtown San Francisco for probability of exceedance of less than 50% in 50 years and period of 2.0 sec.

Earthquake events within a radius of 0.5 degree (approximately 50 km) from the PEER-NGA database were collected<sup>34</sup> and used in the Gaussian Process regression to arrive at mean spectral acceleration ordinates for different scenarios. The data that was implemented included events with moment magnitudes from 4.9 to 6.93 recorded from sources between 13.7 to 131.53 km. The results are shown in Figure 4.18.

The optimal set of the coefficients of the eigenquakes for each scenario are shown graphically in Figure 4.19. The mean target spectra in Figure 4.18 show similarities between spectral shapes and this is also reflected in the magnitudes of the optimal coefficients of the eigenquakes presented in Figure 4.19.

With an optimal set of coefficients, simulation of ground motion histories can be done simply by a linear combination of the eigenquakes. The results are shown in Figures 4.20 to 4.23 for spectral acceleration, acceleration, velocity, and displacement time histories, respectively.

---

<sup>34</sup> 52 records.

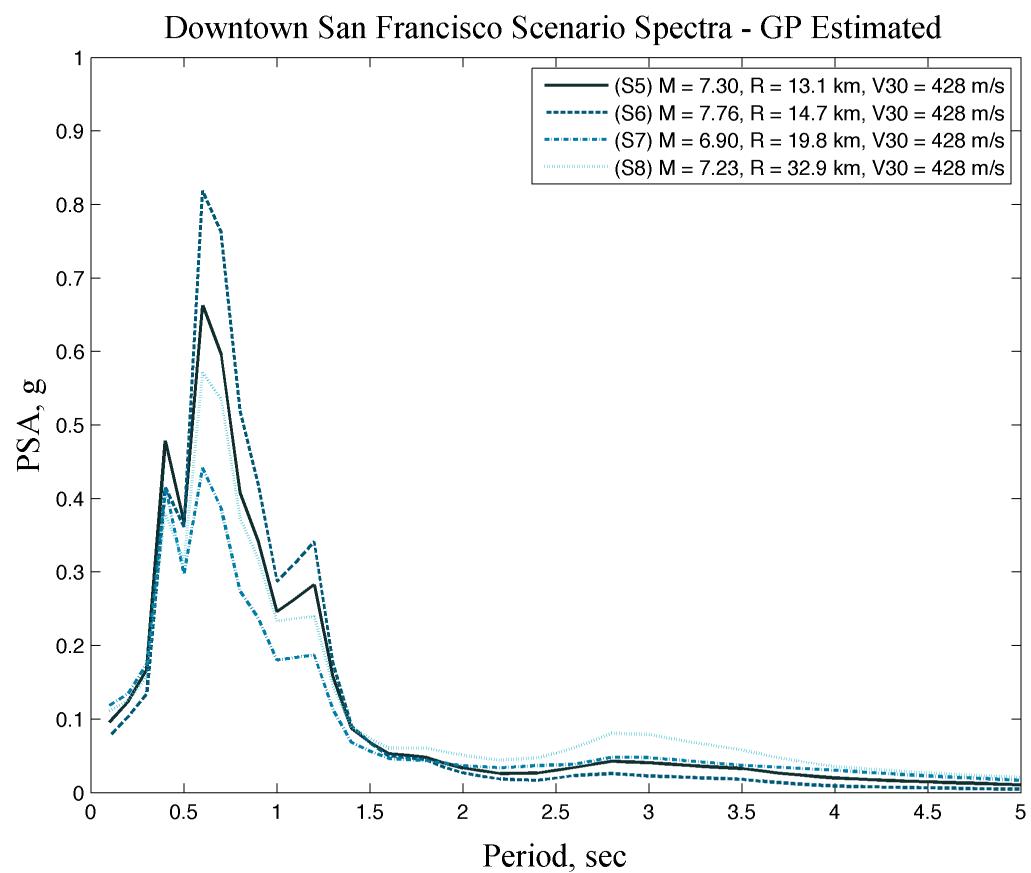


Figure 4.18. Mean target (controlling) response spectra for downtown San Francisco for different scenarios.

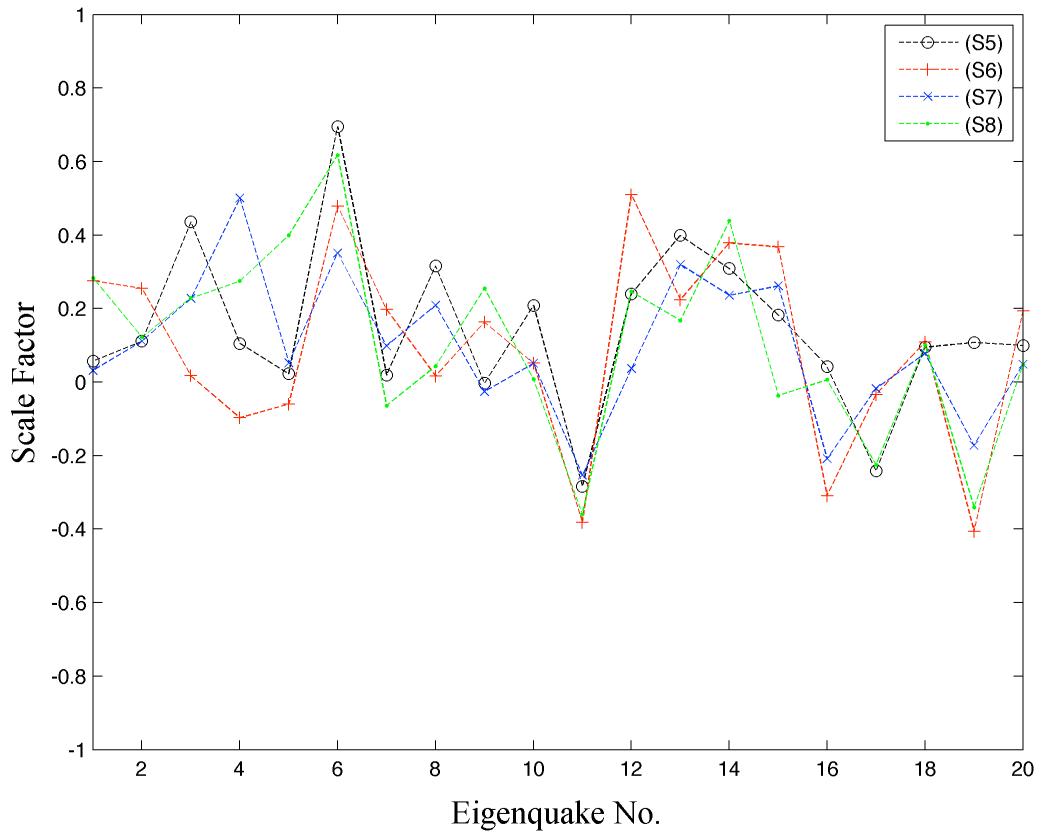


Figure 4.19. The optimal set of coefficients of the eigenquakes for different scenarios for downtown San Francisco.

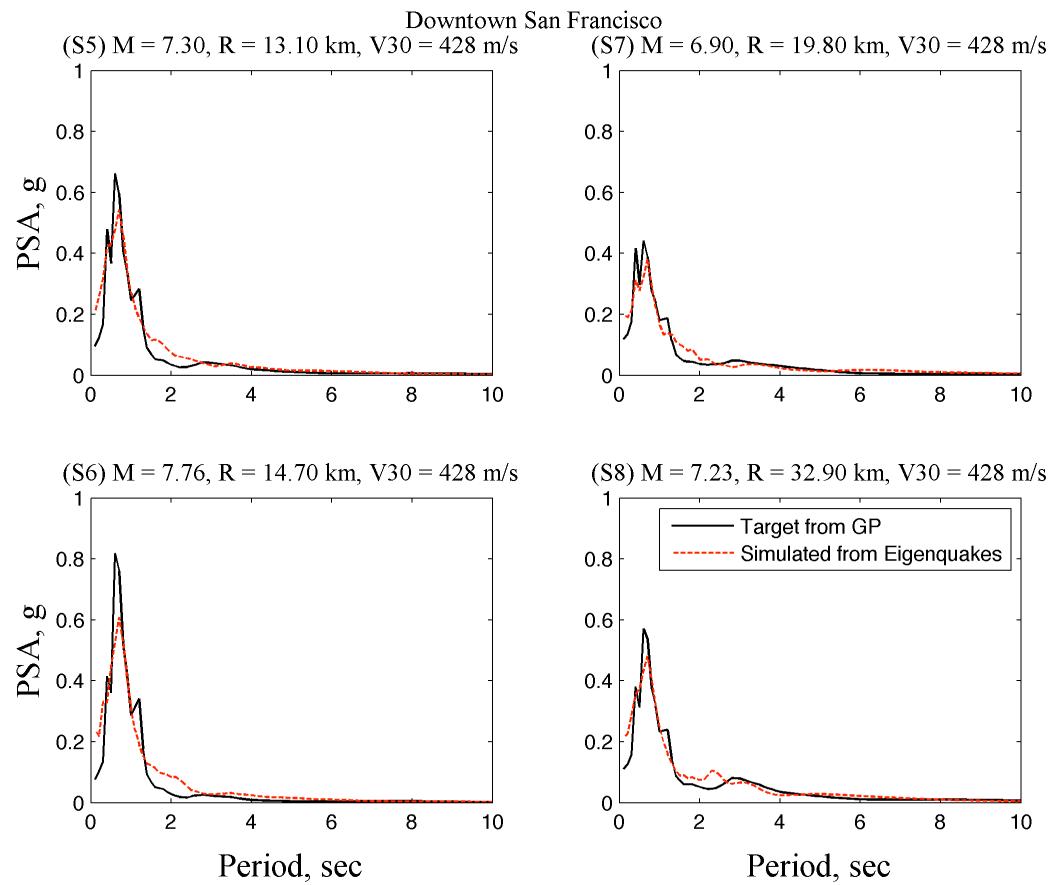


Figure 4.20. Target response spectra for downtown San Francisco (black lines) and the response spectra from the simulated motions using an optimal combination of the eigenquakes (red dashed lines).

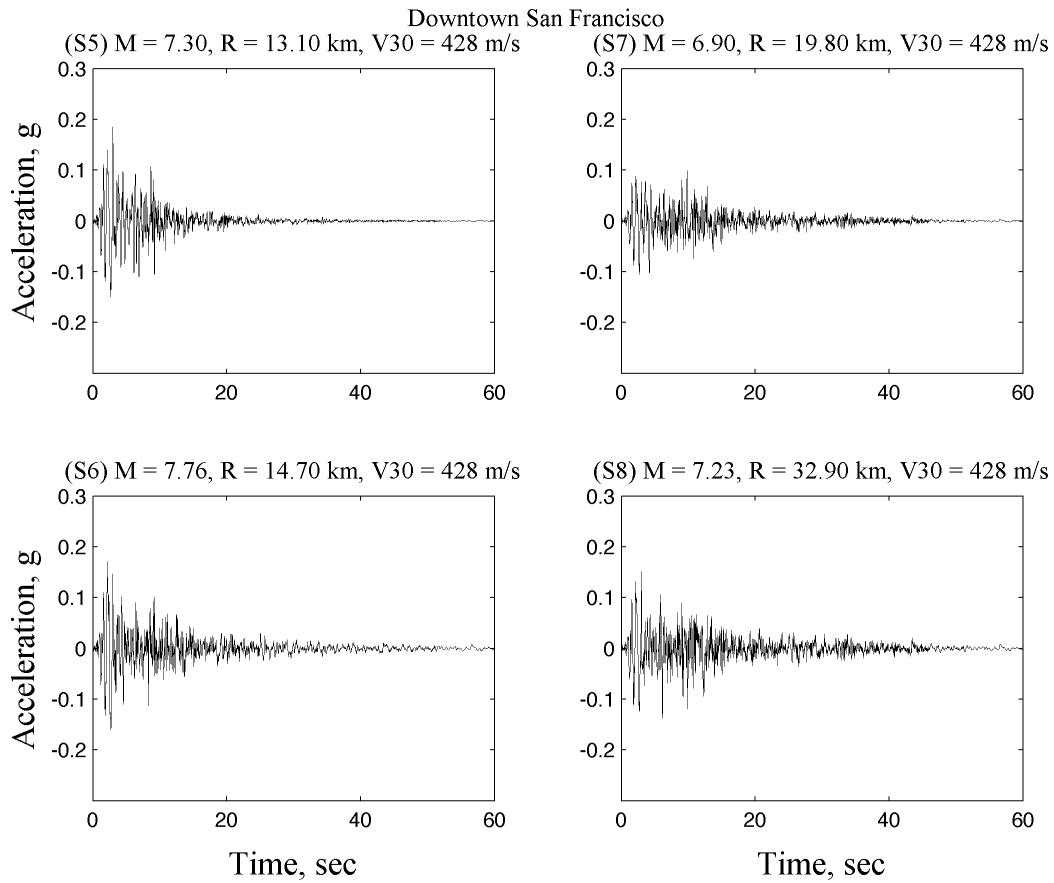


Figure 4.21. Simulated ground motions (acceleration time histories) for downtown San Francisco.

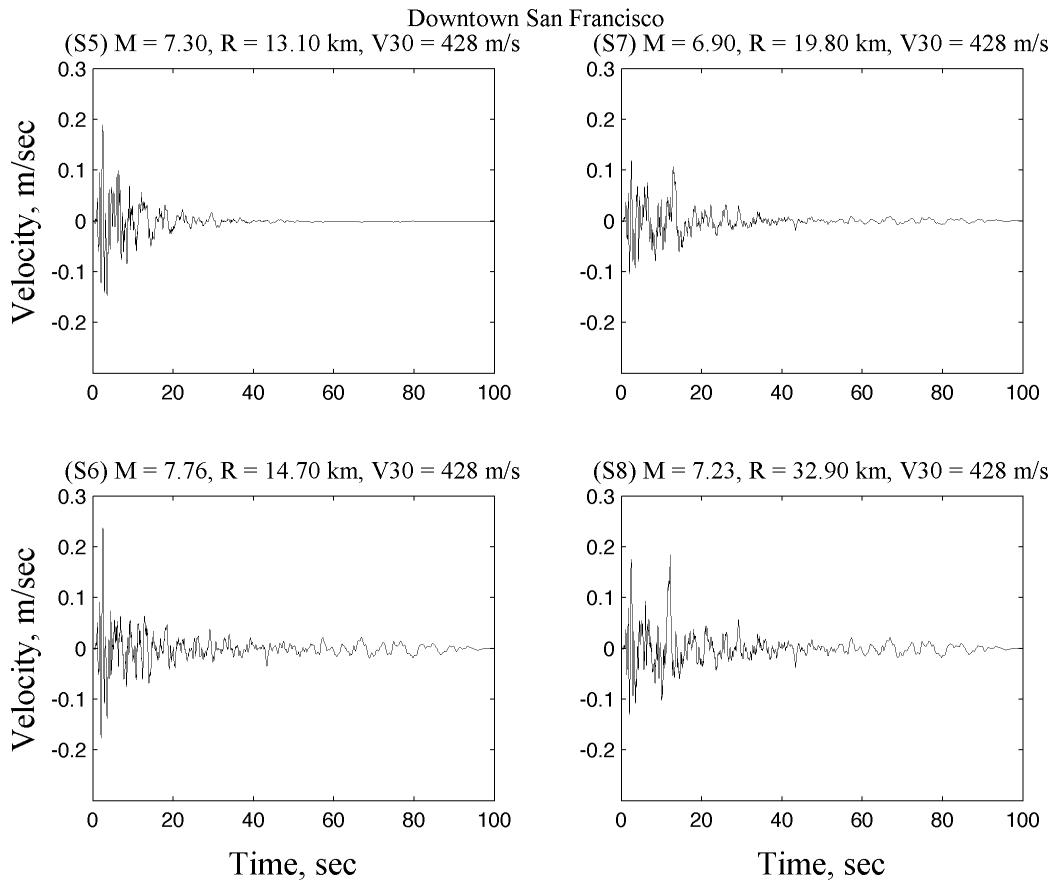


Figure 4.22. Simulated ground motions (velocity time histories) for downtown San Francisco.

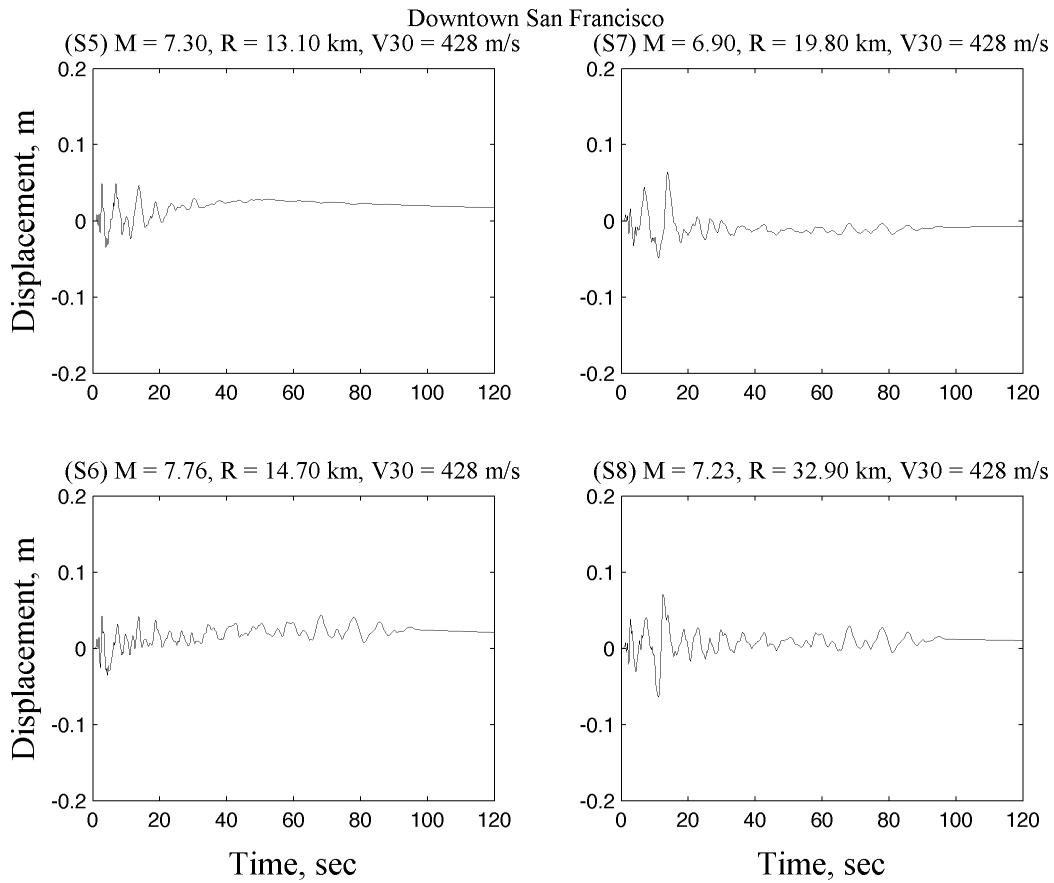


Figure 4.23. Simulated ground motions (displacement time histories) for downtown San Francisco.

# Chapter 5 Conclusions

A procedure for ground motion simulation using machine learning techniques was presented. A brief description of the traditional methods of seismic hazard analysis, with implications in structural design, was provided. The benefits of relying on recorded data in the context of a complex systems treatment of ground motion processes was discussed. The following conclusions can be drawn:

## Concluding Remarks

- I. The large amount of recorded ground motion data available worldwide requires systematic methods of analysis. Many convenient methods of data analysis can be borrowed from the field of machine learning for ground motion database analysis and for record simulation.
- II. The problem of ground motion selection and scaling that has been subject of many controversies in the past ten years is avoided in the proposed procedure.

III. The new concept of eigenquakes was introduced as a set of orthonormal basis vectors for simulation of ground motion records.

IV. Eigenquakes can be taken as a means of encapsulating the information that seismic stations collect over time and so represent the state of our knowledge about strong ground motion processes and their variation. Comparison of the eigenquakes, their form and their rank, before and after new events are appended to the global databases would indicate whether the new events have contributed to our state of knowledge about these processes. If “global” eigenquakes or their ranks remain unchanged in the aftermath of an event, it may be argued that, from perspective of data forms<sup>35</sup>, the new event has not generated new information.

V. Gaussian Process regression can reliably estimate spectral shapes of the ground motion expected at a site for a given scenario event, given availability of previously recorded data.

VI. The proposed procedure is particularly suitable for large urban areas where dense instrumentation exists or is expected to grow, thereby producing increasing numbers of records over time.

---

<sup>35</sup> Since the eigenquakes are obtained based on the history of ground acceleration.

## Future Work

- I. *Issues with Dynamic Databases:* a Bayesian framework of analysis is needed for dynamic databases, *i.e.*, updating the eigenquakes with availability of new data. See Point II below.
- II. Variational PCA (Bishop 1999) is particularly suitable for this purpose as it would allow for a probabilistic formulation of the classical PCA. Automatic selection of the appropriate model dimensionality (the number of eigenquakes to retain) would then be possible.
- III. *Challenges of high-dimensional optimization:* A drawback of the procedure is its reliance on robust global optimization in high-dimensional spaces. It was shown in a few cases that the genetic algorithm could be trapped in local minima. The high-dimensional search space may require long computational times to converge because of larger sets of candidate solutions that are needed to effectively search the solution space.
- IV. A variation of the proposed procedure for linear combination of the eigenquakes may be needed to take into account the ranks of the eigenquakes. The current formulation treats different eigenquakes uniformly.



# Bibliography

Abrahamson, N., G. Atkinson, et al. (2008). "Comparisons of the NGA Ground-Motion Relations." Earthquake Spectra **24**(1): 45-66.

Abrahamson, N. and W. Silva (2008). "Summary of the Abrahamson & Silva NGA Ground-Motion Relations." Earthquake Spectra **24**(1): 67-97.

Ahmadi, G. (1979). "Generation of artificial time histories compatible with given response spectra—a review." Solid Mechanics Archives **4**(3): 207-239.

Ahmadi, G. (1980). "A note on the Wiener-Hermite representation of the earthquake ground acceleration." Mechanics Research Communications **7**(1): 7-13.

Alimoradi, A., E. Miranda, et al. (2006). "Evolutionary modal identification utilizing coupled shear-flexural response - implication for multistory buildings. Part I : Theory." The Structural Design of Tall and Special Buildings **15**(1): 51-65.

Alimoradi, A. and F. Naeim (2006). "Evolutionary modal identification utilizing coupled shear-flexural response - implication for multistory buildings. Part II : Application." The Structural Design of Tall and Special Buildings **15**(1): 67-103.

Alimoradi, A., S. Pezeshk, et al. (2007). "Probabilistic Performance-Based Optimal Design of Steel Moment-Resisting Frames. II: Applications." Journal of Structural Engineering **133**(6): 767-776.

- Alimoradi, A., S. Pezeshk, et al. (2005). "Fuzzy Pattern Classification of Strong Ground Motion Records." Journal of Earthquake Engineering **9**(3): 307-332.
- Amin, M. and A. H.-S. Ang (1968). "Nonstationary stochastic model of earthquake motions." Journal of Engineering Mechanics Division, ASCE **94**(2): 559-583.
- Ammon, C. J., C. Ji, et al. (2005). "Rupture Process of the 2004 Sumatra-Andaman Earthquake." Science **308**(5725): 1133-1139.
- Anderson, J. G. (2010). "Source and Site Characteristics of Earthquakes That Have Caused Exceptional Ground Accelerations and Velocities." Bulletin of the Seismological Society of America **100**(1): 1-36.
- ASCE (2006). Minimum Design Loads for Buildings and Other Structures (ASCE Standard No. 7-05), American Society of Civil Engineers.
- ASCE. (2011). "Report Card for America's Infrastructure." Retrieved April 2011, 2011, from <http://www.infrastructurereportcard.org/>.
- Atkinson, G. and I. Beresnev (2002). "Ground motions from large earthquakes in the New Madrid seismic zone." Bulletin of the Seismological Society of America **92**: 1015-1024.
- Baker, J. W. and C. A. Cornell (2005). "A vector-valued ground motion intensity measure consisting of spectral acceleration and epsilon." Earthquake Engineering & Structural Dynamics **34**(10): 1193-1217.
- Baker, J. W. and C. A. Cornell (2006). "Spectral shape, epsilon and record selection." Earthquake Engineering & Structural Dynamics **35**(9): 1077-1095.
- Baker, J. W. and C. A. Cornell (2008). "Vector-valued intensity measures for pulse-like near-fault ground motions." Engineering Structures **30**(4): 1048-1057.

Bazzurro, P. and C. Allin Cornell (1999). "Disaggregation of seismic hazard." Bulletin of the Seismological Society of America **89**(2): 501-520.

Beck, J. L. and L. S. Katafygiotis (1998). "Updating Models and Their Uncertainties. Part I: Bayesian Statistical Framework." Journal of Engineering Mechanics **124**(4): 455-461.

Beyer, K. and J. J. Bommer (2007). "Selection and Scaling of Real Accelerograms for Bi-Directional Loading: A Review of Current Practice and Code Provisions." Journal of Earthquake Engineering **11**(1 supp 1): 13 - 45.

Bishop, C. M. (1999). Variational Principal Components. The 9th International Conference on Neural Networks, ICANN'99: 509-514.

Bishop, C. M. (2006). Pattern Recognition and Machine Learning, Springer.

Bommer, J. J. and N. A. Abrahamson (2006). "Why Do Modern Probabilistic Seismic-Hazard Analyses Often Lead to Increased Hazard Estimates?" Bulletin of the Seismological Society of America **96**(6).

Bommer, J. J. and N. A. Abrahamson (2007). "Reply to 'Comment on 'Why Do Modern Probabilistic Seismic-Hazard Analyses Often Lead to Increased Hazard Estimates?' by Julian J. Bommer and Norman A. Abrahamson' by Jens-Uwe Klugel." Bulletin of the Seismological Society of America **97**(6): 2208-2211.

Bommer, J. J. and N. A. Abrahamson (2007). "Reply to 'Comment on 'Why Do Modern Probabilistic Seismic-Hazard Analyses Often Lead to Increased Hazard Estimates?' by Julian J. Bommer and Norman A. Abrahamson' by Zhenming Wang and Mai Zhou." Bulletin of the Seismological Society of America **97**(6): 2215-2217.

- Boore, D. M. and G. M. Atkinson (2008). "Ground-Motion Prediction Equations for the Average Horizontal Component of PGA, PGV, and 5%-Damped PSA at Spectral Periods between 0.01 s and 10.0 s." *Earthquake Spectra* 24(1): 99-138.
- Bozorgnia, Y. and V. V. Bertero (2004). *Earthquake Engineering: From Engineering Seismology to Performance-Based Engineering*, CRC Press.
- Brady, A. G. (2009). "Strong-motion accelerographs: Early history." *Earthquake Engineering & Structural Dynamics* 38(9): 1121-1134.
- CACR. (2011). "Caltech Center for Advanced Computing Research." Retrieved May 2011, 2011, from <http://www.cacr.caltech.edu/main/>.
- Campbell, K. W. and Y. Bozorgnia (2008). "NGA Ground Motion Model for the Geometric Mean Horizontal Component of PGA, PGV, PGD and 5% Damped Linear Elastic Response Spectra for Periods Ranging from 0.01 to 10 s." *Earthquake Spectra* 24(1): 139-171.
- Chan, E. (1997). Optimal Design of Building Structures Using Genetic Algorithms. *Earthquake Engineering Research Laboratory*. Pasadena, California Institute of Technology.
- Chiou, B., R. Darragh, et al. (2008). "NGA Project Strong-Motion Database." *Earthquake Spectra* 24(1): 23-44.
- Chiou, B. S. J. and R. R. Youngs (2008). "An NGA Model for the Average Horizontal Component of Peak Ground Motion and Response Spectra." *Earthquake Spectra* 24(1): 173-215.
- Choi, Y., J. P. Stewart, et al. (2005). "Empirical Model for Basin Effects Accounts for Basin Depth and Source Location." *Bulletin of the Seismological Society of America* 95(4): 1412-1427.

- Chou, J.-H. and J. Ghaboussi (2001). "Genetic Algorithm in Structural Damage Detection." Computers and Structures **79**: 1335-1353.
- Conte, J. P., K. S. Pister, et al. (1992). "Nonstationary ARMA modeling of seismic motion." Journal of Soil Dynamics and Earthquake Engineering **11**(7): 411-426.
- Cornell, C. A. (1968). "Engineering seismic risk analysis" Bulletin of the Seismological Society of America **58**(5): 1583-1606.
- Corral, A. (2004). "Long-Term Clustering, Scaling, and Universality in the Temporal Occurrence of Earthquakes." Physical Review Letters **92**(10): 108501.
- Delavaud, E., F. Scherbaum, et al. (2009). "Information-Theoretic Selection of Ground-Motion Prediction Equations for Seismic Hazard Analysis: An Applicability Study Using Californian Data." Bulletin of the Seismological Society of America **99**(6): 3248-3263.
- Der Kiureghian, A. and J. Crempien (1989). "An evolutionary model for earthquake ground motion." Structural Safety **6**(2-4): 235-246.
- Foley, C. M., S. Pezeshk, et al. (2007). "Probabilistic Performance-Based Optimal Design of Steel Moment-Resisting Frames. I: Formulation." Journal of Structural Engineering **133**(6): 757-766.
- Gallagher, R. and T. Appenzeller (1999). "Beyond Reductionism." Science **284**(5411): 79-.
- Geller, R. J. (2011). "Shake-up time for Japanese seismology." Nature **472**(7344): 407-409.
- Giaralis, A. and P. D. Spanos (2009). "Wavelet-based response spectrum compatible synthesis of accelerograms--Eurocode application (EC8)." Soil Dynamics and Earthquake Engineering **29**(1): 219-235.

Goldberg, D. E. (1989). Genetic Algorithms in Search, Optimization, and Machine Learning. Boston, MA, Addison-Wesley.

Grigoriu, M. (2011). "To Scale or Not to Scale Seismic Ground-Acceleration Records." Journal of Engineering Mechanics **137**(4): 284-293.

Gu, P. and Y. K. Wen (2007). "A Record-Based Method for the Generation of Tridirectional Uniform Hazard-Response Spectra and Ground Motions Using the Hilbert-Huang Transform." Bulletin of the Seismological Society of America **97**(5): 1539-1556.

Hough, S. E., J. R. Altidor, et al. (2010). "Localized damage caused by topographic amplification during the 2010 M7.0 Haiti earthquake." Nature Geosci **3**(11): 778-782.

Housner, G. W. (1975). Measures of Severity of Earthquake Ground Shaking. First U.S. National Conference on Earthquake Engineering. Ann Arbor, MI, EERI: 25-33.

Housner, G. W. and P. C. Jennings (1964). "Generation of artificial earthquakes." Journal of Engineering Mechanics Division, ASCE **90**(1): 113-150.

Housner, G. W. and P. C. Jennings (1982). Earthquake Design Criteria. Oakland, CA, Earthquake Engineering Research Institute.

Hudson, D. E. (1979). Reading and Interpreting Strong Motion Accelerograms, Earthquake Engineering Research Institute.

Idriss, I. M. (2008). "An NGA Empirical Model for Estimating the Horizontal Spectral Values Generated By Shallow Crustal Earthquakes." Earthquake Spectra **24**(1): 217-242.

Iervolino, I. and C. A. Cornell (2005). "Record Selection for Nonlinear Seismic Analysis of Structures." Earthquake Spectra **21**(3): 685-713.

Iervolino, I., F. De Luca, et al. (2010). "Spectral shape-based assessment of SDOF non-linear response to real, adjusted and artificial accelerograms." Engineering Structures **32**(9): 2776-2792.

Iwan, W. D., M. A. Moser, et al. (1985). "Some observations on strong-motion earthquake measurement using a digital accelerograph." Bulletin of the Seismological Society of America **75**(5): 1225-1246.

Jennings, P. C., G. W. Housner, et al. (1969). Simulated Earthquake Motions for Design Purposes. 4th World Conference on Earthquake Engineering. Santiago de Chile.

Jolliffe, I. T. (2002). Principal Component Analysis, Springer.

Katafygiotis, L. S. and J. L. Beck (1998). "Updating Models and Their Uncertainties. Part II: Model Identifiability." Journal of Engineering Mechanics **124**(4): 463-467.

Kikuchi, M. and H. Kanamori (1982). "Inversion of complex body waves." Bulletin of the Seismological Society of America **72**(2): 491-506.

Kikuchi, M. and H. Kanamori (1986). "Inversion of complex body waves-II." Physics of The Earth and Planetary Interiors **43**(3): 205-222.

Kikuchi, M. and H. Kanamori (1991). "Inversion of complex body waves--III." Bulletin of the Seismological Society of America **81**(6): 2335-2350.

Kim, Y.-J. and J. Ghaboussi (2001). "Direct Use of Design Criteria in Genetic Algorithm-Based Controller Optimization." Earthquake Engineering and Structural Dynamics **30**: 1261-1278.

Klugel, J.-U. (2007). "Comment on "Why Do Modern Probabilistic Seismic-Hazard Analyses Often Lead to Increased Hazard Estimates?" by Julian J. Bommer

and Norman A. Abrahamson." Bulletin of the Seismological Society of America **97**(6): 2198-2207.

Klugel, J.-U. (2009). "Comment on "Sigma: Issues, Insights and Challenges" by F. O. Strasser, N. A. Abrahamson, and J. J. Bommer." Seismological Research Letters **80**(3): 494-498.

Kottke, A. and E. M. Rathje (2008). "A Semi-Automated Procedure for Selecting and Scaling Recorded Earthquake Motions for Dynamic Analysis." Earthquake Spectra **24**(4): 911-932.

Kramer, S. L. (1996). Geotechnical Earthquake Engineering. Upper Saddle River, New Jersey, Prentice Hall.

Krawinkler, H. (2001). Progress and Challenges in Performance-based Earthquake Engineering. Linbeck Distinguished Lecture Series in Earthquake Engineering: Challenges of the New Millennium, University of Notre Dame, Friedman Family Visiting Professional Program, Earthquake Engineering Research Institute.

Lay, T., C. J. Ammon, et al. (2010). "Teleseismic inversion for rupture process of the 27 February 2010 Chile (Mw 8.8) earthquake." Geophysical Research Letters **37**(L13301).

Luco, N. and P. Bazzurro (2007). "Does amplitude scaling of ground motion records result in biased nonlinear structural drift responses?" Earthquake Engineering & Structural Dynamics **36**(13): 1813-1835.

Masri, S. F., R. K. Miller, et al. (1990). "Probabilistic representation and transmission of earthquake ground motion records." Earthquake Engineering & Structural Dynamics **19**(7): 1025-1040.

MathWorks MATLAB, The Language of Technical Computing.

McGuire, R., K. (2004). Seismic Hazard and Risk Analysis. EERI Monograph Series. Oakland, CA.

McGuire, R. K. (2008). "Probabilistic seismic hazard analysis: Early history." Earthquake Engineering & Structural Dynamics 37(3): 329-338.

Mobarakeh, A. A., F. R. Rofooei, et al. (2002). "Simulation of earthquake records using time-varying Arma (2,1) model." Probabilistic Engineering Mechanics 17(1): 15-34.

Mossotti, V. G., J. A. Barragan, et al. (2002). Cx-02 Program Workshop on Modeling Complex Systems. Reno, Nevada, U.S. Geological Survey.

Musson, R. (2009). "Ground motion and probabilistic hazard." Bulletin of Earthquake Engineering 7(3): 575-589.

Naeim, F., A. Alimoradi, et al. (2004). "Selection and scaling of ground motion time histories for structural design using genetic algorithms." Earthquake Spectra Vol. 20(2, pp. 413-426. May).

Newmark, N. M. and W. J. Hall (1982). Earthquake Spectra and Design. EERI Monograph Series. Oakland, CA: 103.

Papadimitriou, K. and J. L. Beck (1992). Stochastic Characterization of Ground Motion and Applications to Structural Response. Tenth World Conference on Earthquake Engineering. Madrid, Spain.

PEER. (2011). "Ground Motion Selection and Modification Program." Retrieved April 2011, 2011, from <http://peer.berkeley.edu/gmsm/index.html>.

PEER. (2011). "NGA Database." Retrieved Jan. 2011, 2011, from <http://peer.berkeley.edu/nga/>.

Raich, A. M. and J. Ghaboussi (2000). "Evolving Structural Design Solutions Using an Implicit Redundant Genetic Algorithm." Structural and Multidisciplinary Optimization 20(3): 222-231.

Rasmussen, C. E. (2006). Advances in Gaussian Processes. NIPS 2006. Vancouver.

Rasmussen, C. E. and C. K. I. Williams (2006). Gaussian Processes for Machine Learning. Cambridge, Massachusetts, the MIT Press.

Rezaeian, S. and A. Der Kiureghian (2008). "A stochastic ground motion model with separable temporal and spectral nonstationarities." Earthquake Engineering & Structural Dynamics 37(13): 1565-1584.

Rezaeian, S. and A. Der Kiureghian (2010). Stochastic Modeling and Simulation of Ground Motions for Performance-Based Earthquake Engineering, Pacific Earthquake Engineering Research Center.

Saragoni, G. R. and G. C. Hart (1973). "Simulation of artificial earthquakes." Earthquake Engineering & Structural Dynamics 2(3): 249-267.

Scherbaum, F., E. Delavaud, et al. (2009). "Model Selection in Seismic Hazard Analysis: An Information-Theoretic Perspective." Bulletin of the Seismological Society of America 99(6): 3234-3247.

Shahbazian, A. and S. Pezeshk (2010). "Improved Velocity and Displacement Time Histories in Frequency Domain Spectral-Matching Procedures." Bulletin of the Seismological Society of America 100(6): 3213-3223.

Shlens, J. (2009, April 22, 2009). "A Tutorial on Principal Component Analysis." from <http://www.snl.salk.edu/~shlens/pca.pdf>.

Shome, N., C. A. Cornell, et al. (1998). "Earthquakes, Records, and Nonlinear Responses." Earthquake Spectra 14(3): 469-500.

Snelson, E. (2006). Tutorial: Gaussian process models for machine learning. Gatsby Computational Neuroscience Unit, UCL.

Song, S. G. and P. Somerville (2010). "Physics-Based Earthquake Source Characterization and Modeling with Geostatistics." Bulletin of the Seismological Society of America **100**(2): 482-496.

Stein, S., M. Liu, et al. (2009). "Mid-Continent Earthquakes as a Complex System." Seismological Research Letters **80**(4): 551-553.

Strasser, F. O. and J. J. Bommer (2009). "Review: Strong Ground Motions--Have We Seen the Worst?" Bulletin of the Seismological Society of America **99**(5): 2613-2637.

Turcotte, D. L. and B. D. Malamud (2002). Earthquakes as a Complex System (Chapter 14). International Handbook of Earthquake & Engineering Seismology. W. H. K. Lee, H. Kanamori, P. Jennings and C. Kisslinger, Academic Press. A: 209- 228.

Wang, M. and T. Takada (2009). "A Bayesian Framework for Prediction of Seismic Ground Motion." Bulletin of the Seismological Society of America **99**(4): 2348-2364.

Wang, Z. and M. Zhou (2007). "Comment on "Why Do Modern Probabilistic Seismic-Hazard Analyses Often Lead to Increased Hazard Estimates?" by Julian J. Bommer and Norman A. Abrahamson." Bulletin of the Seismological Society of America **97**(6): 2212-2214.

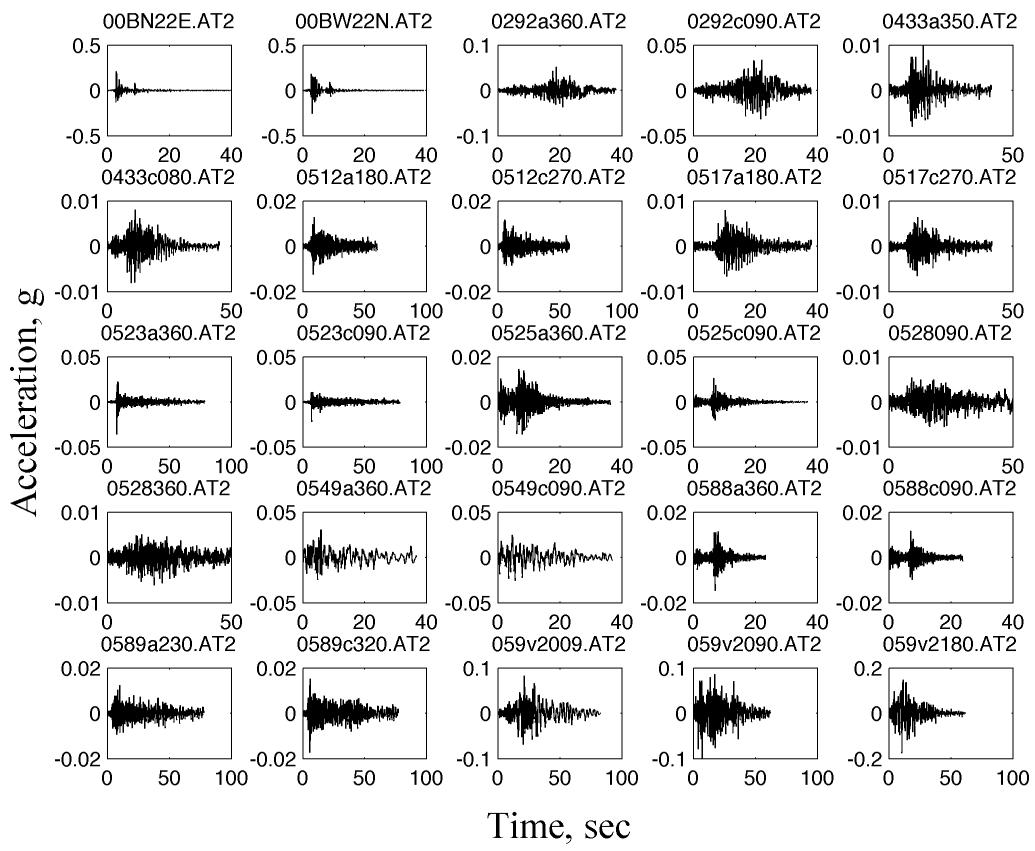
Watson-Lamprey, J. and N. Abrahamson (2006). "Selection of ground motion time series and limits on scaling." Soil Dynamics and Earthquake Engineering **26**(5): 477-482.

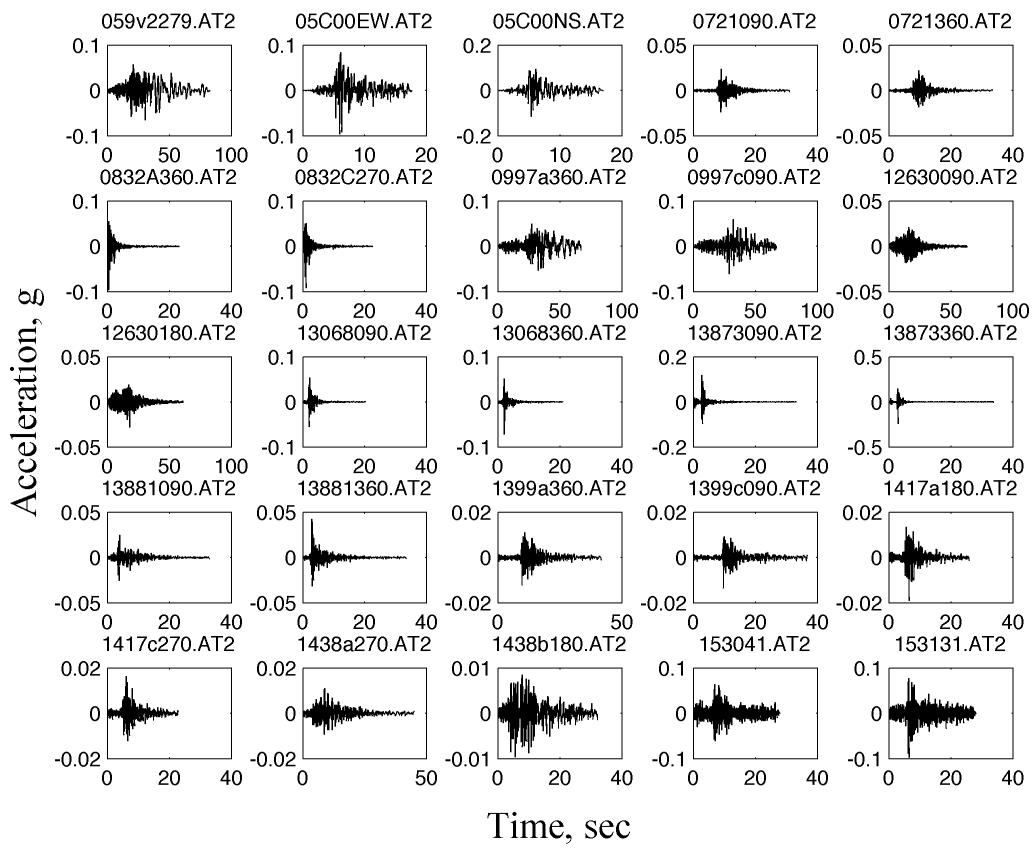
Yamada, M., J. Mori, et al. (2009). "The Slapdown Phase in High-acceleration Records of Large Earthquakes." Seismological Research Letters **80**(4): 559-564.

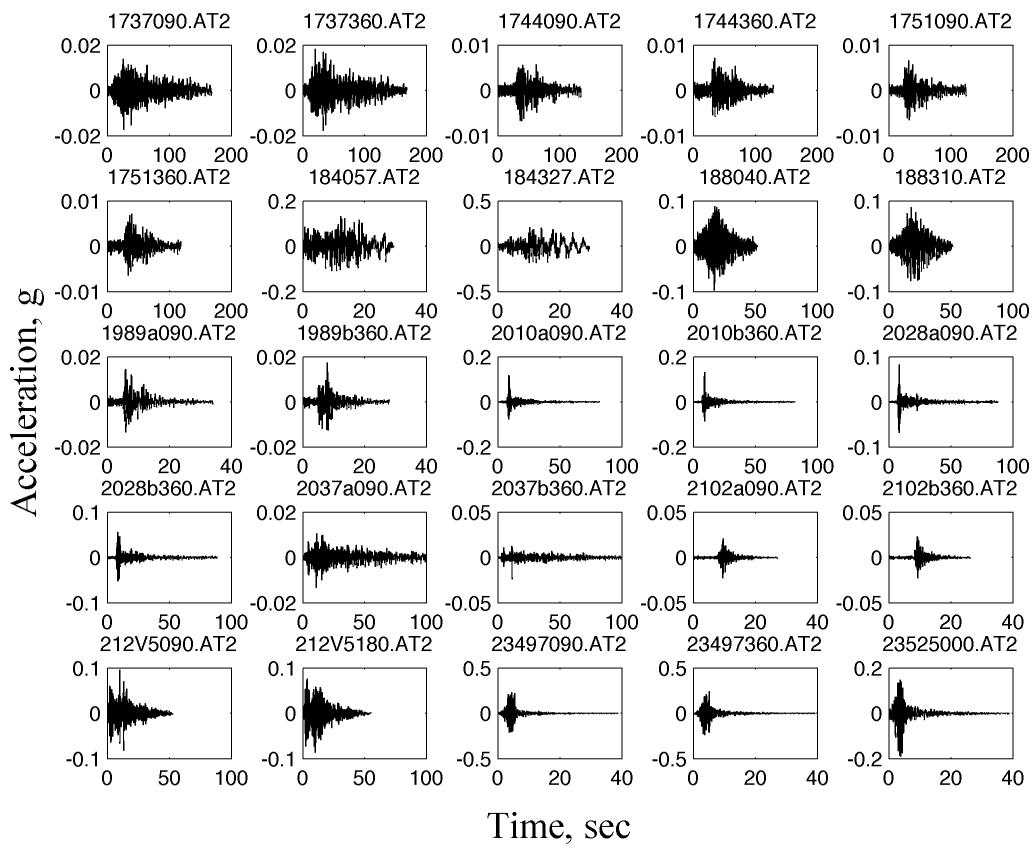
Yamamoto, Y. and J. W. Baker (2011). Stochastic model for earthquake ground motions using wavelet packets. 11th International Conference on Applications of Statistics and Probability in Civil Engineering. Zurich, Switzerland: 8.

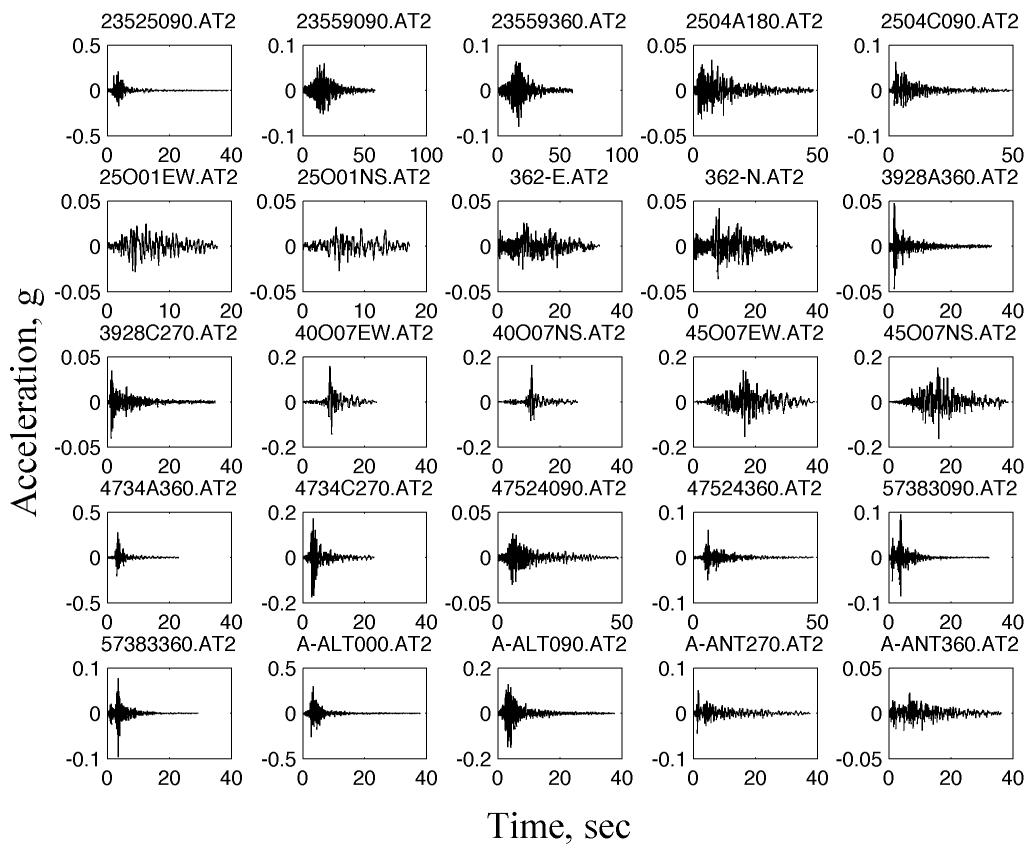
# Appendix I            PEER-NGA Data Sample

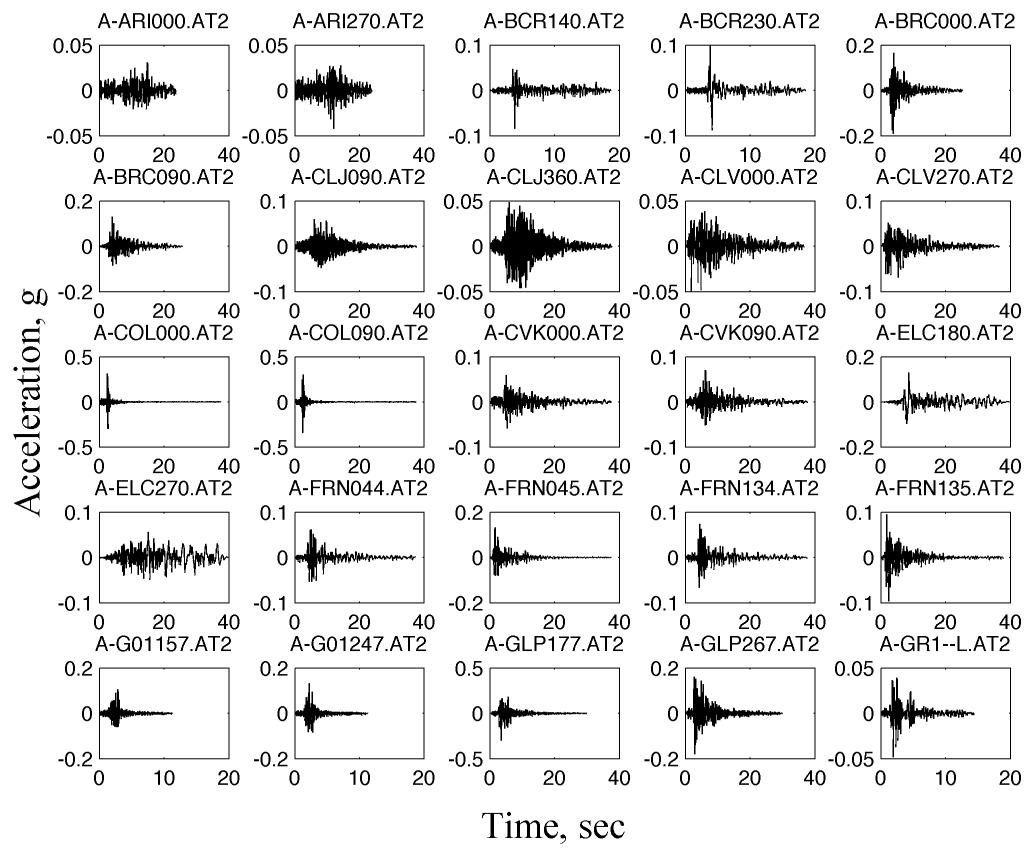
Acceleration time histories of individual components are shown below. File-names are given on the top of each plot.

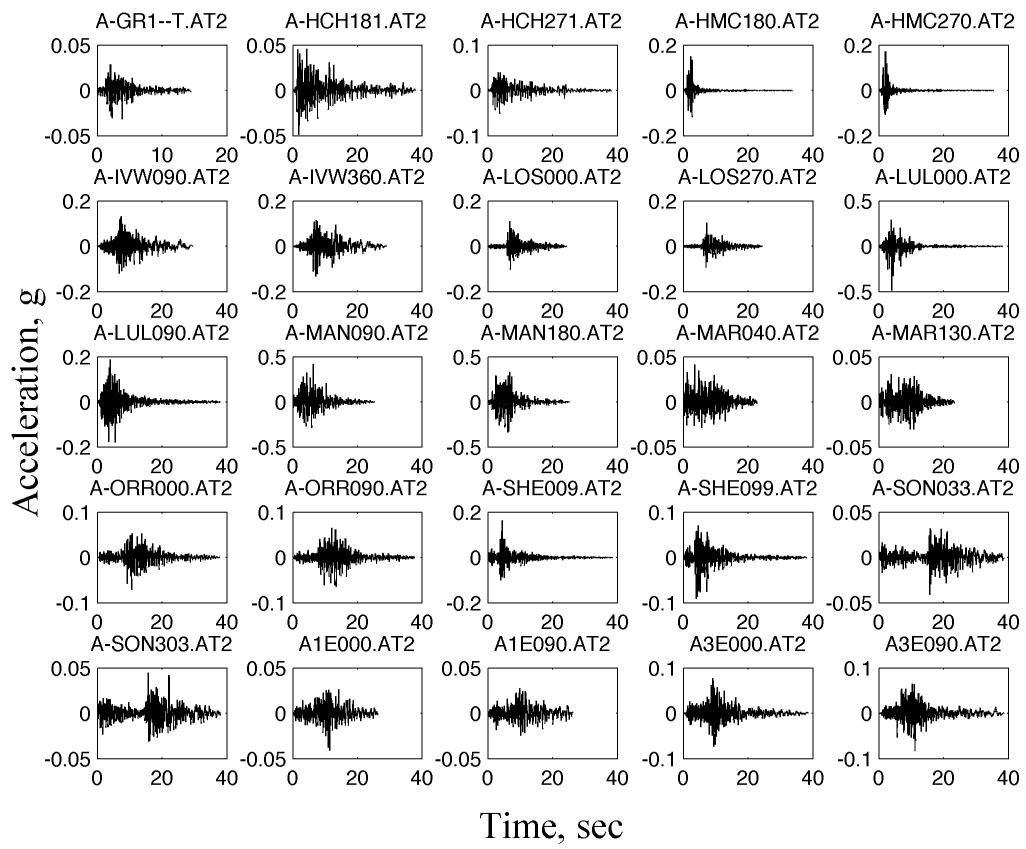


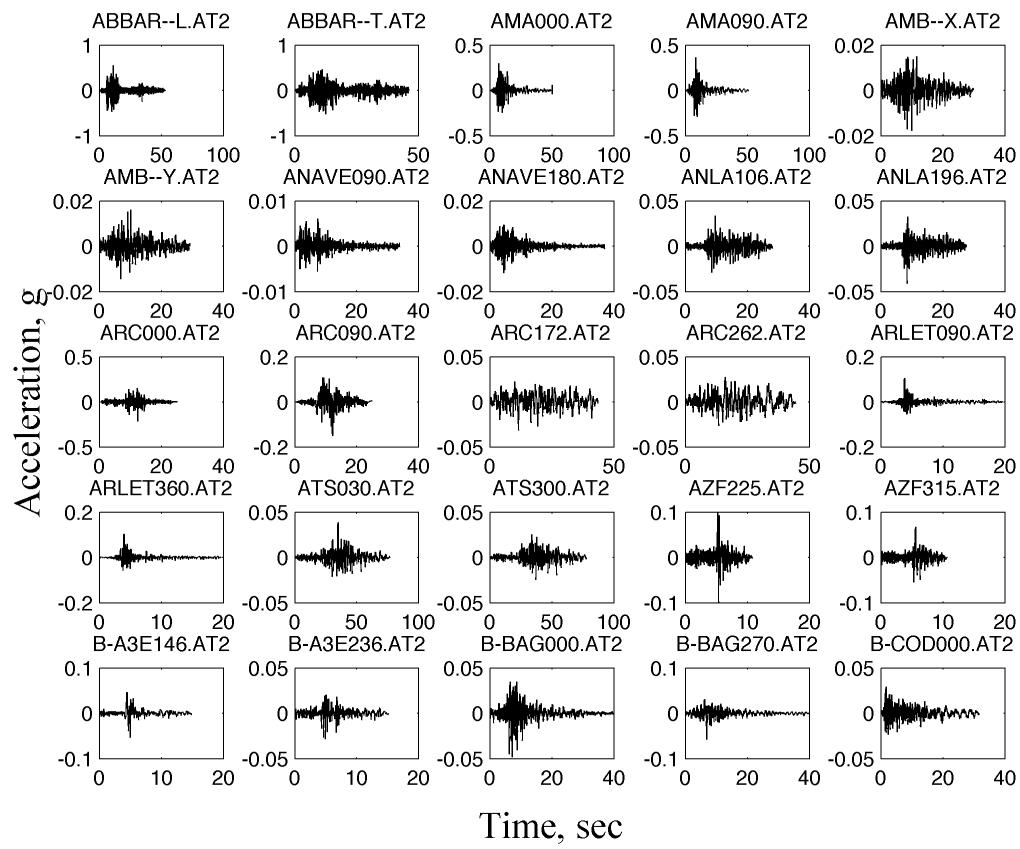


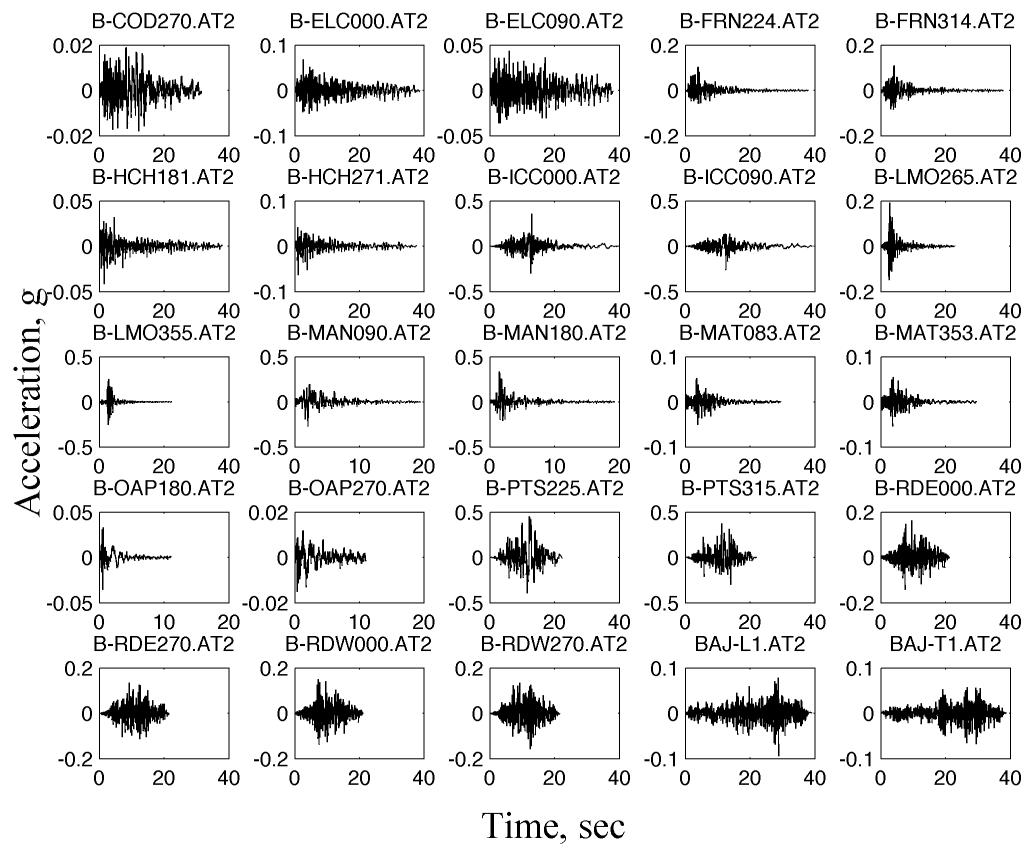


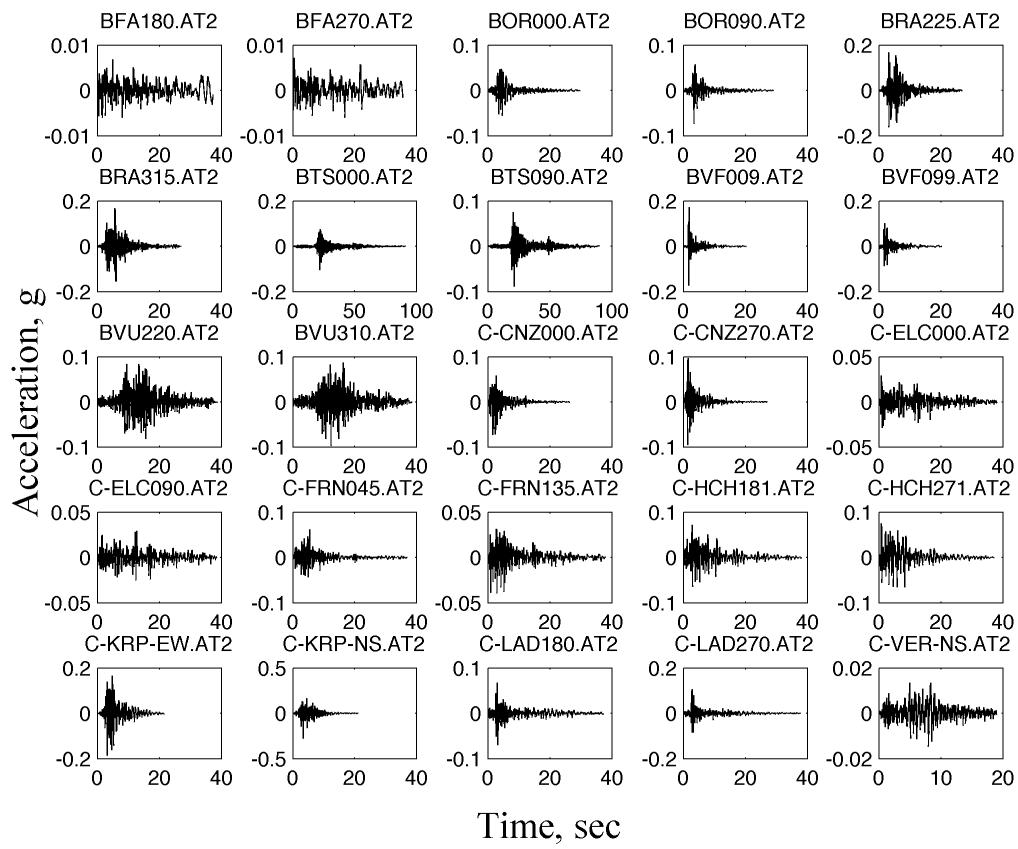


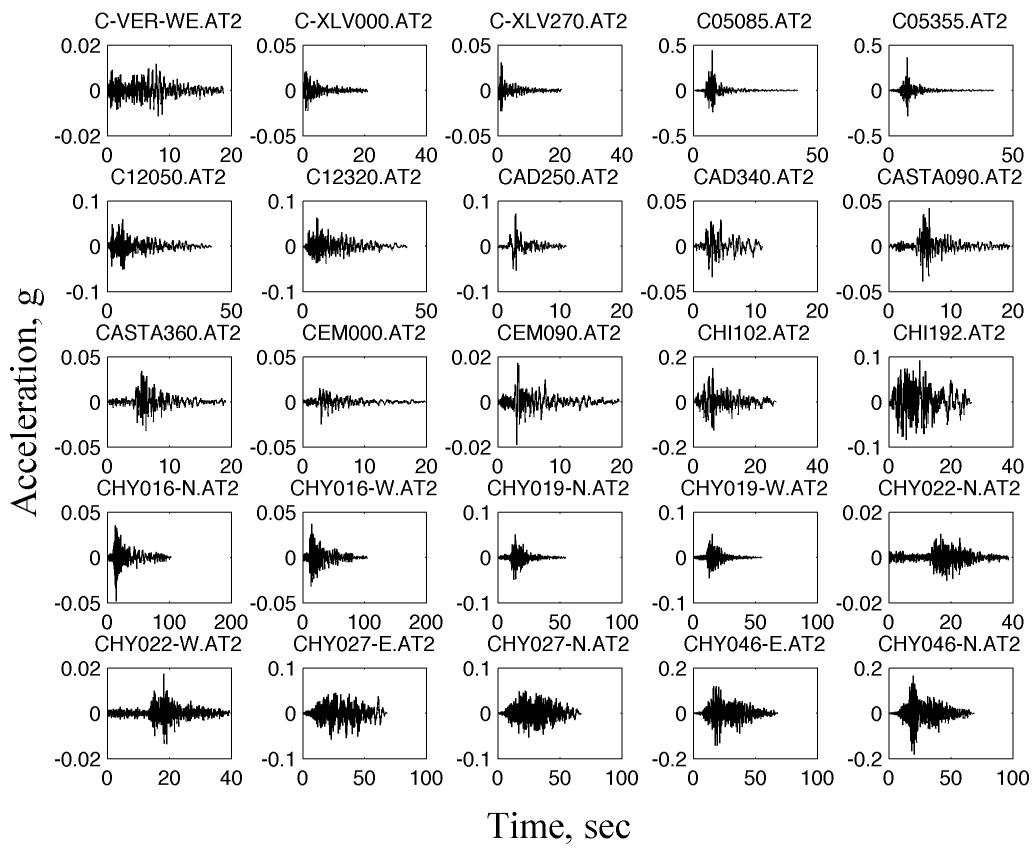


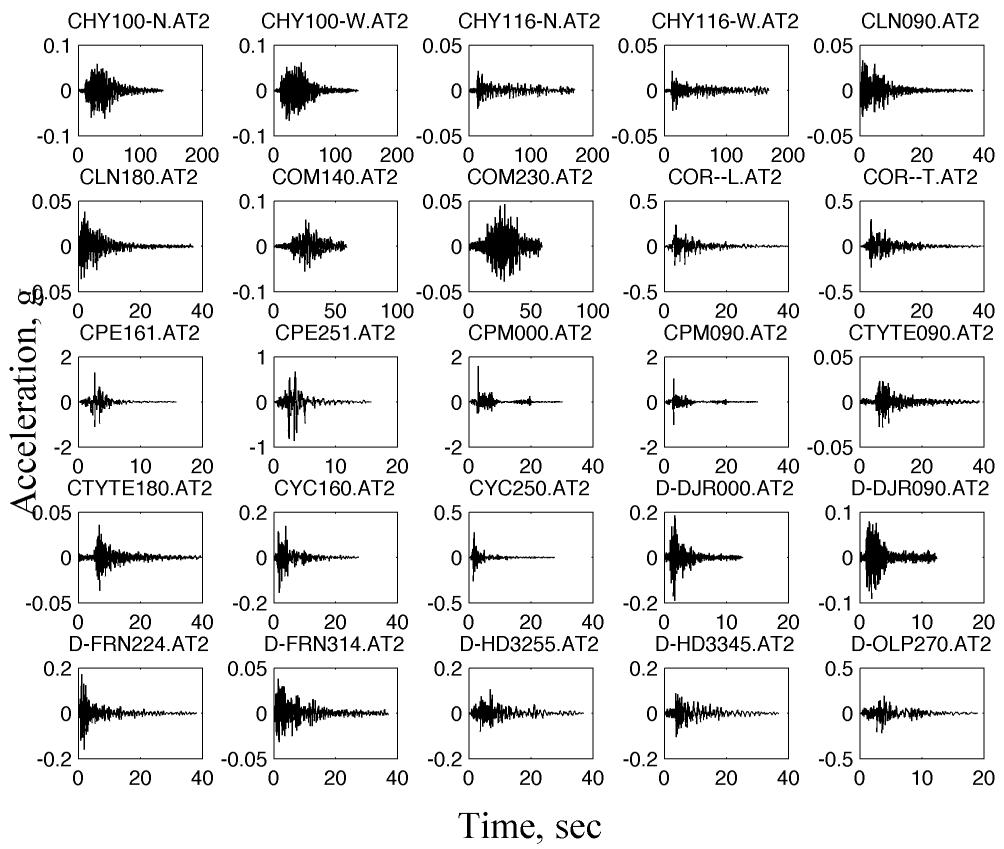


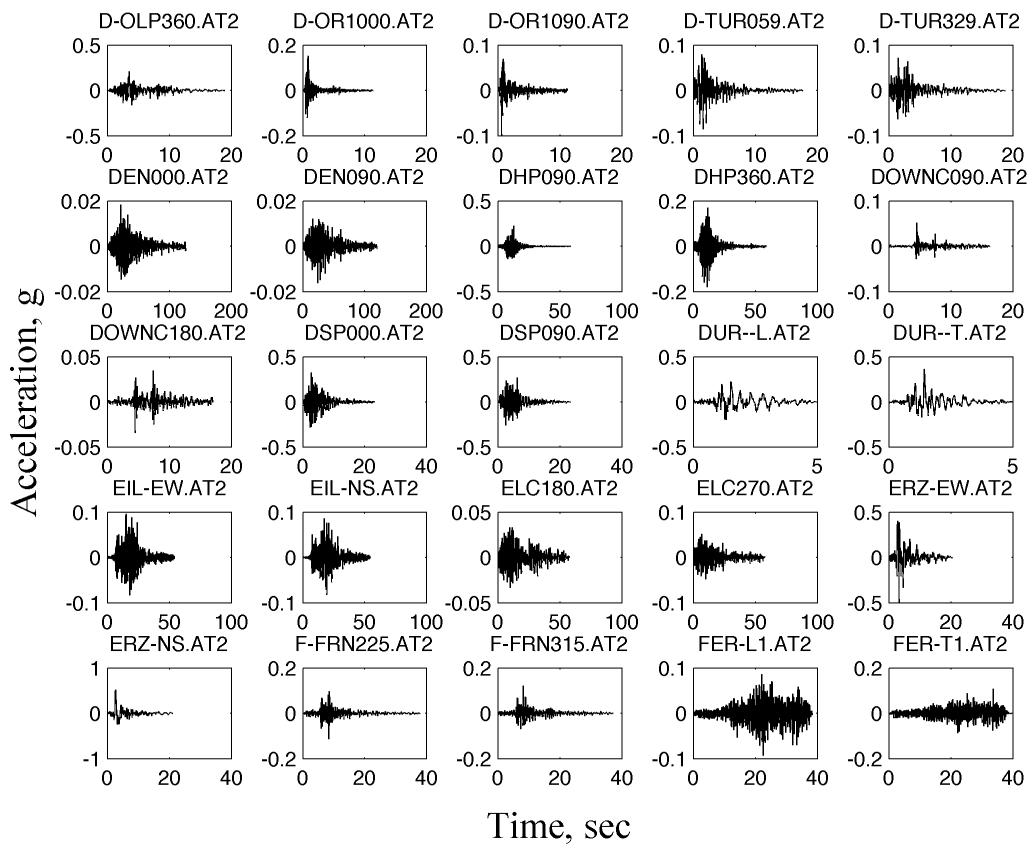


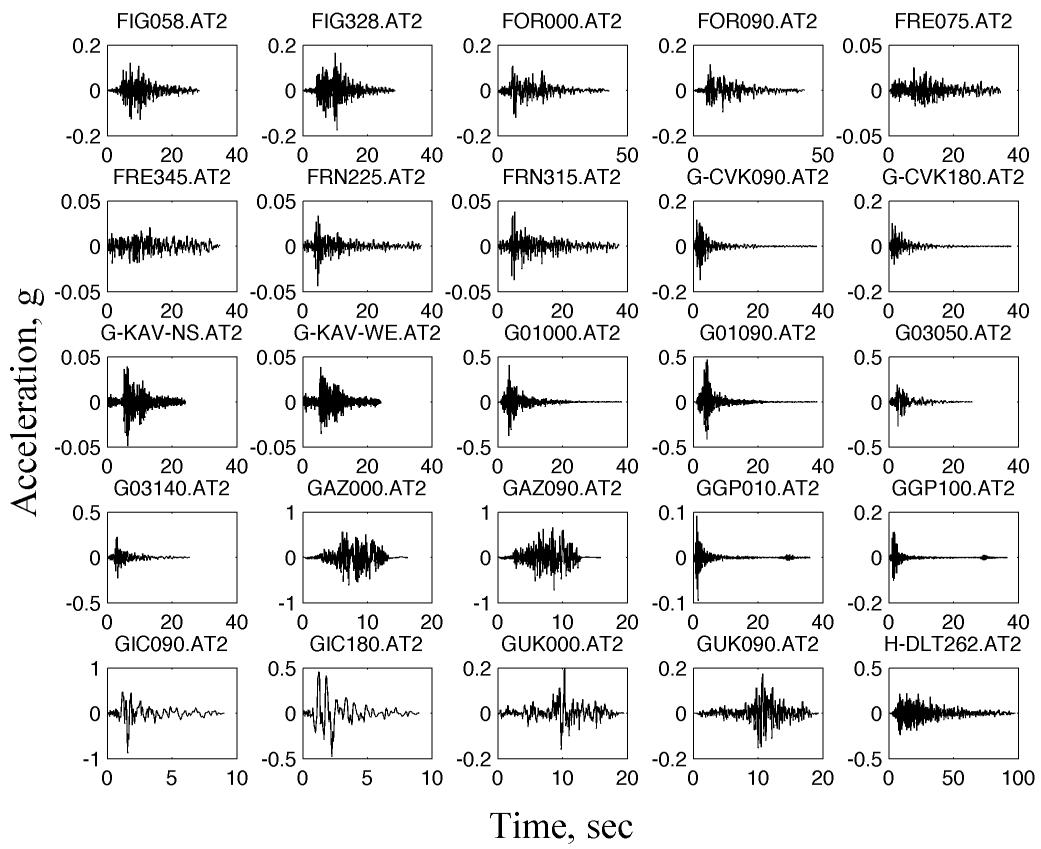


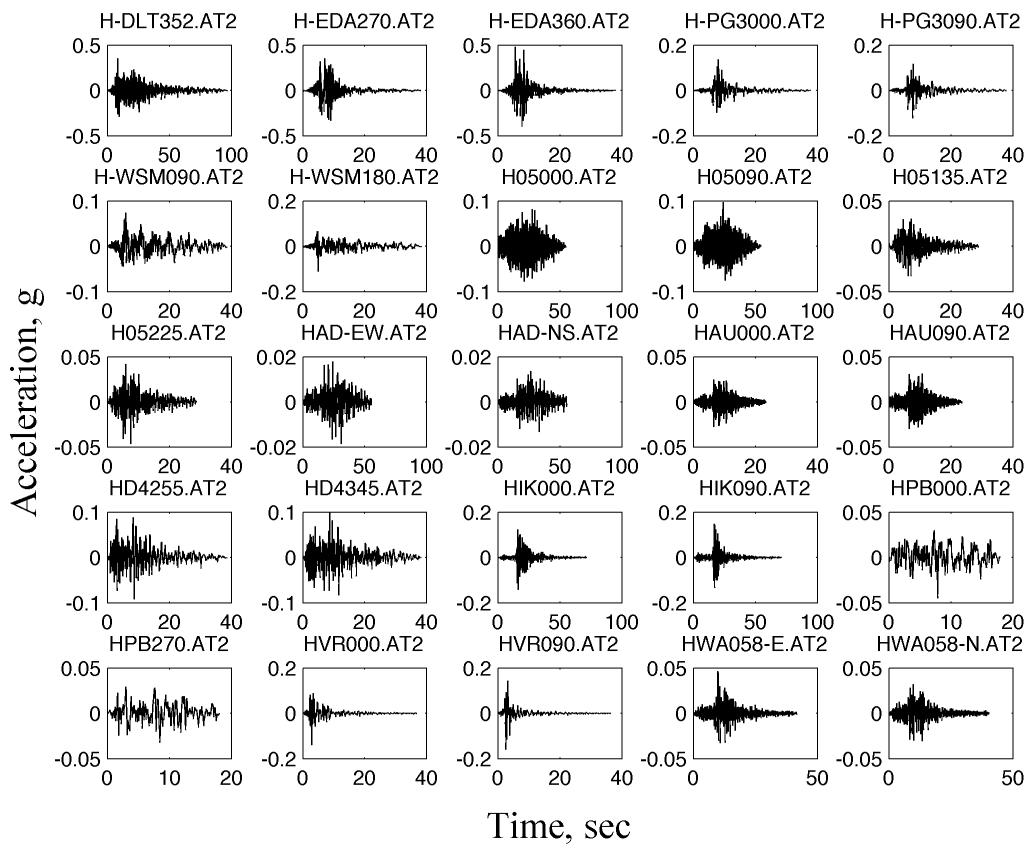


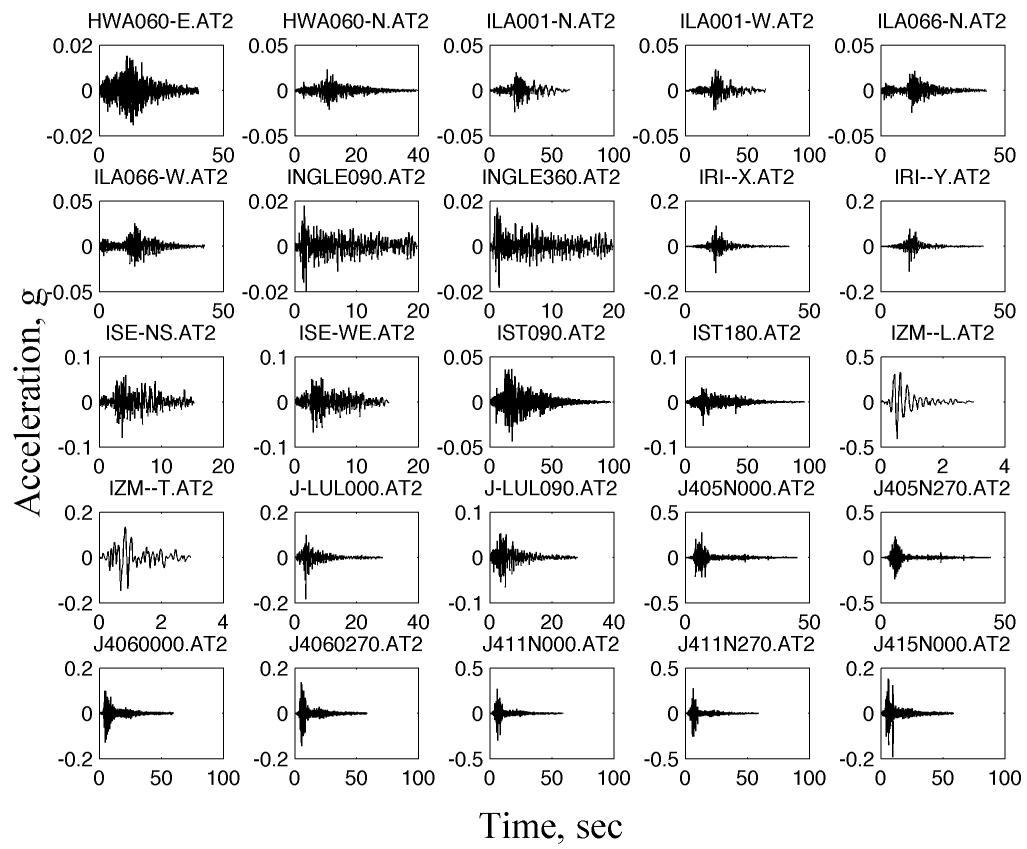


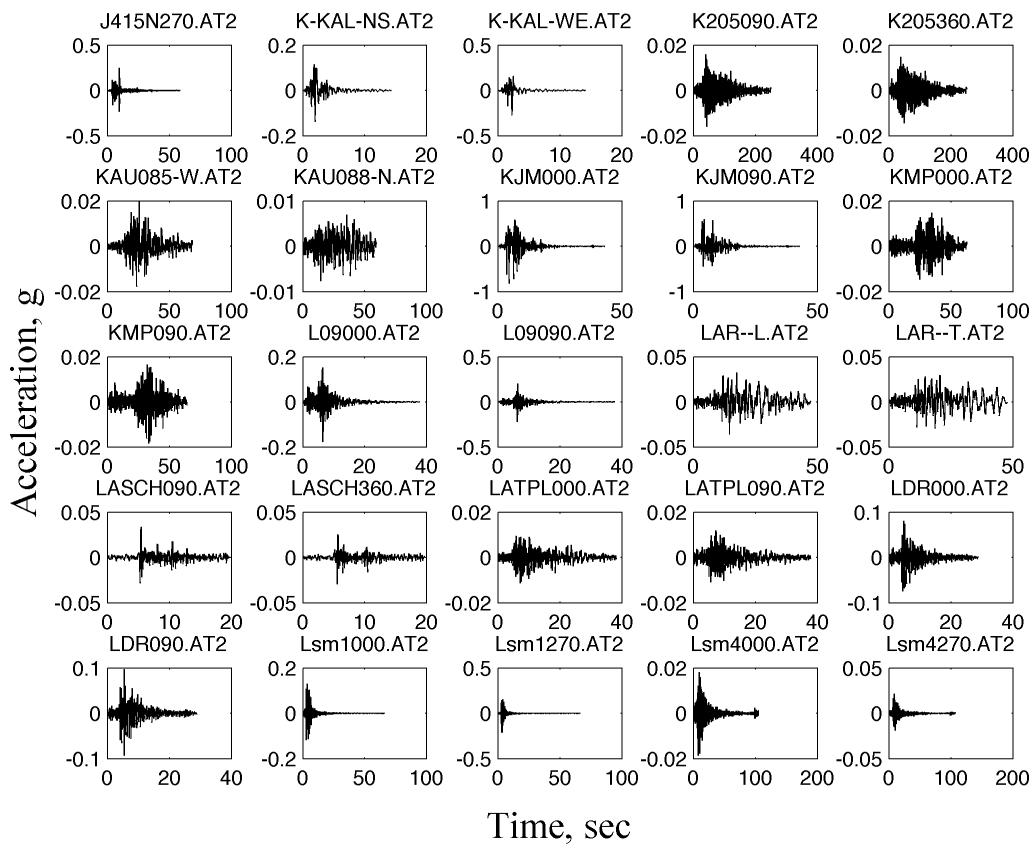


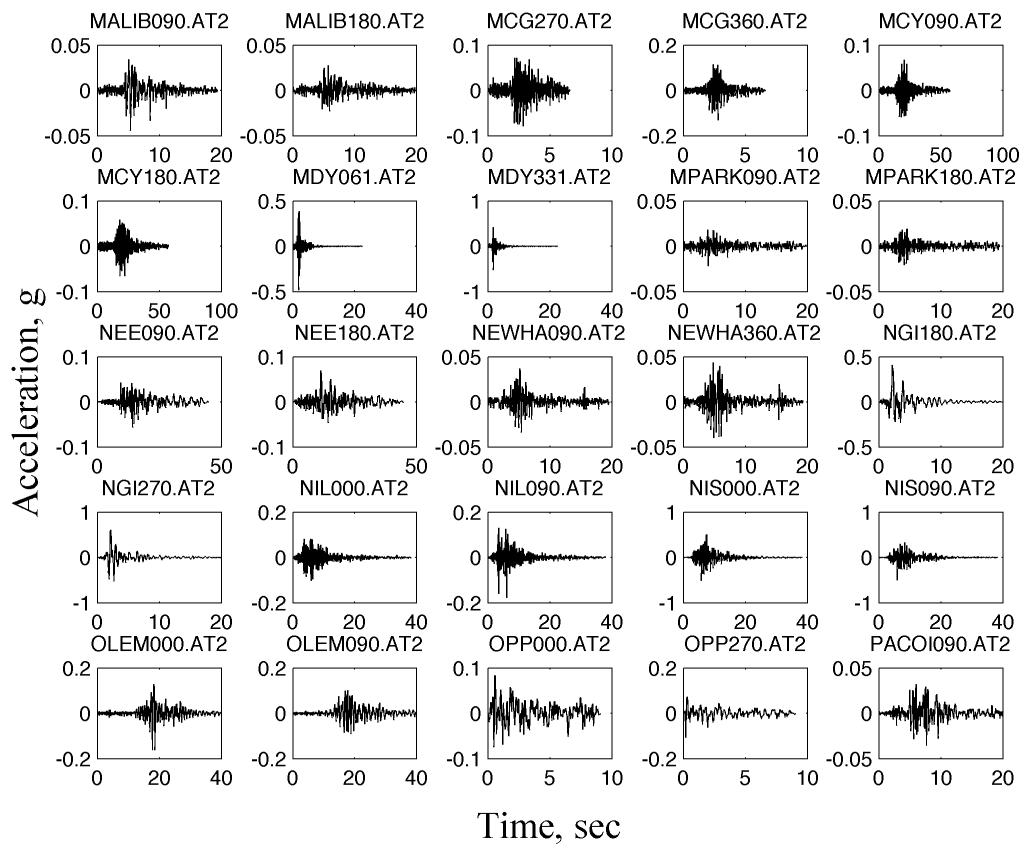


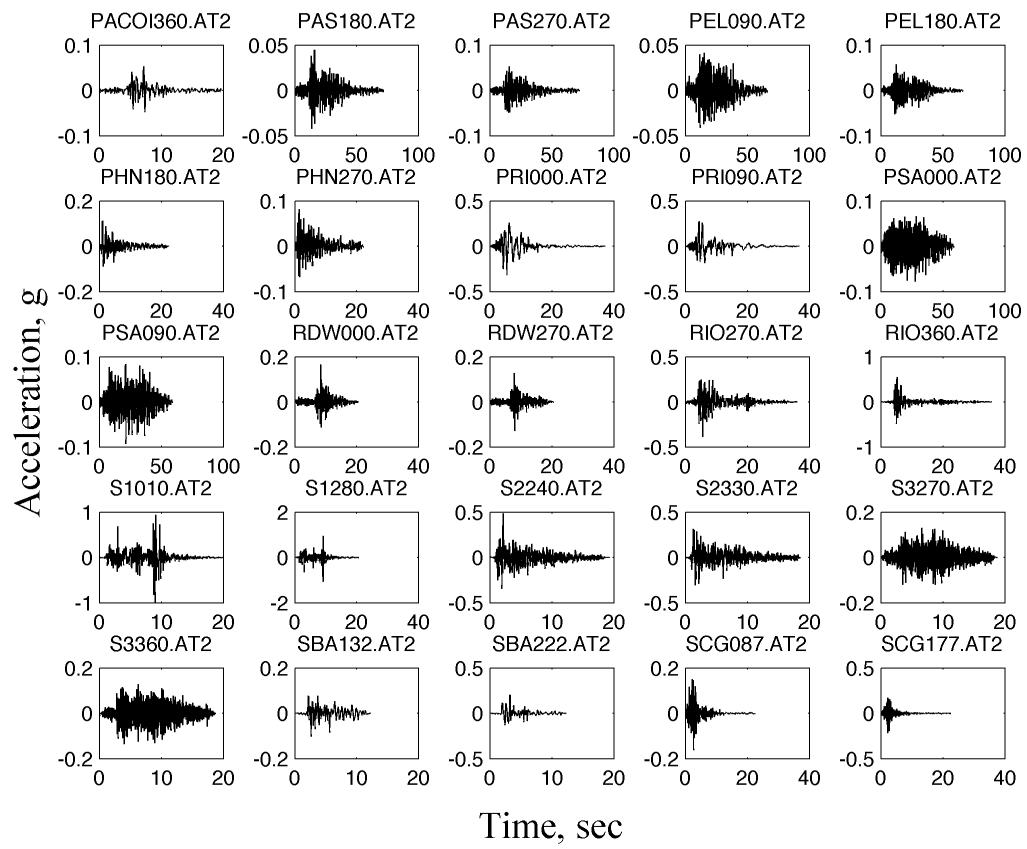


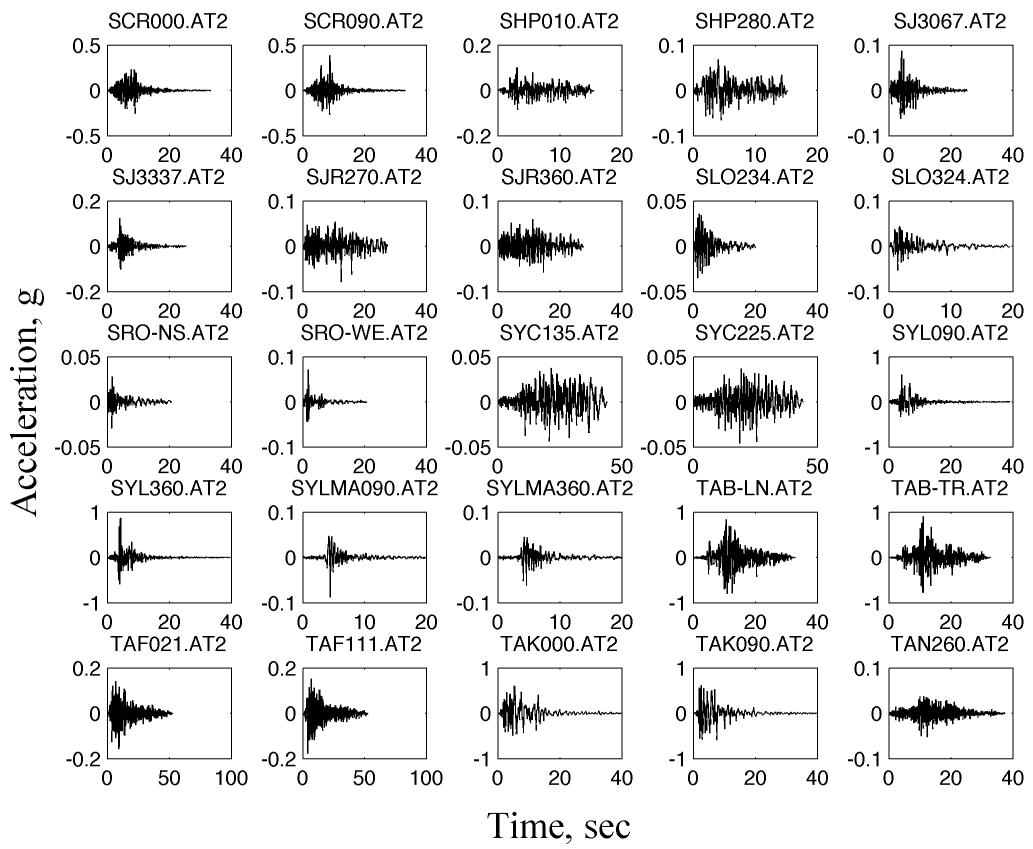


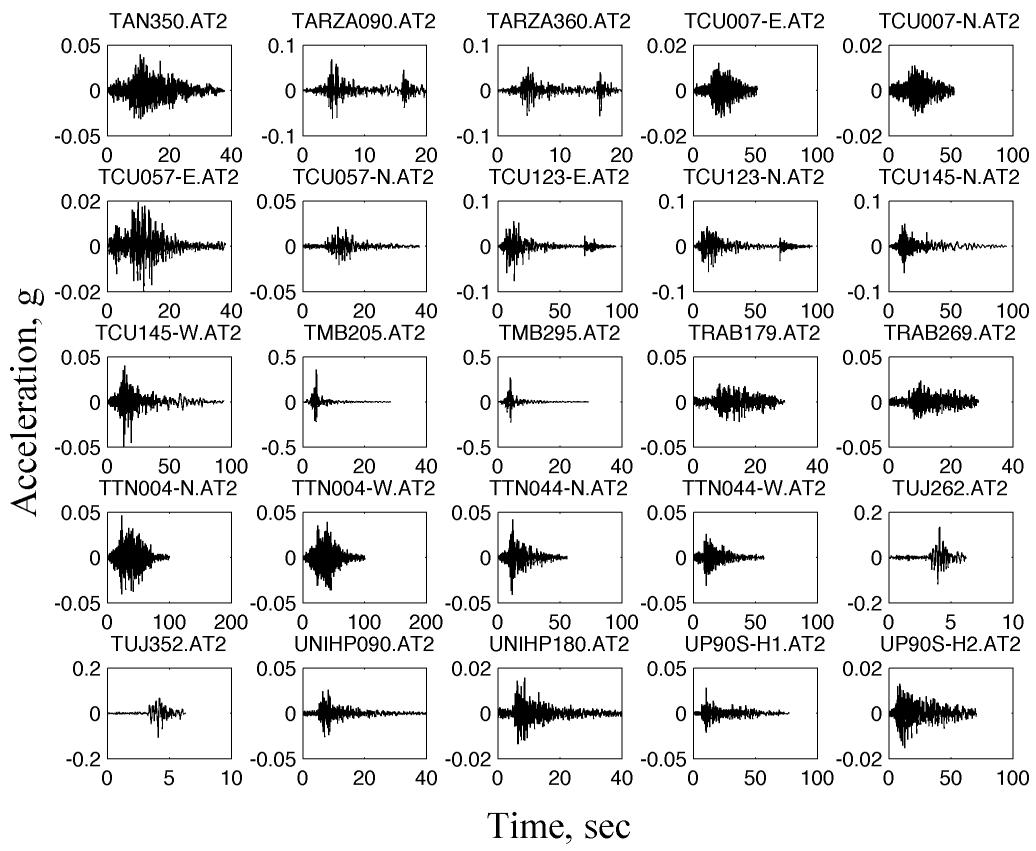


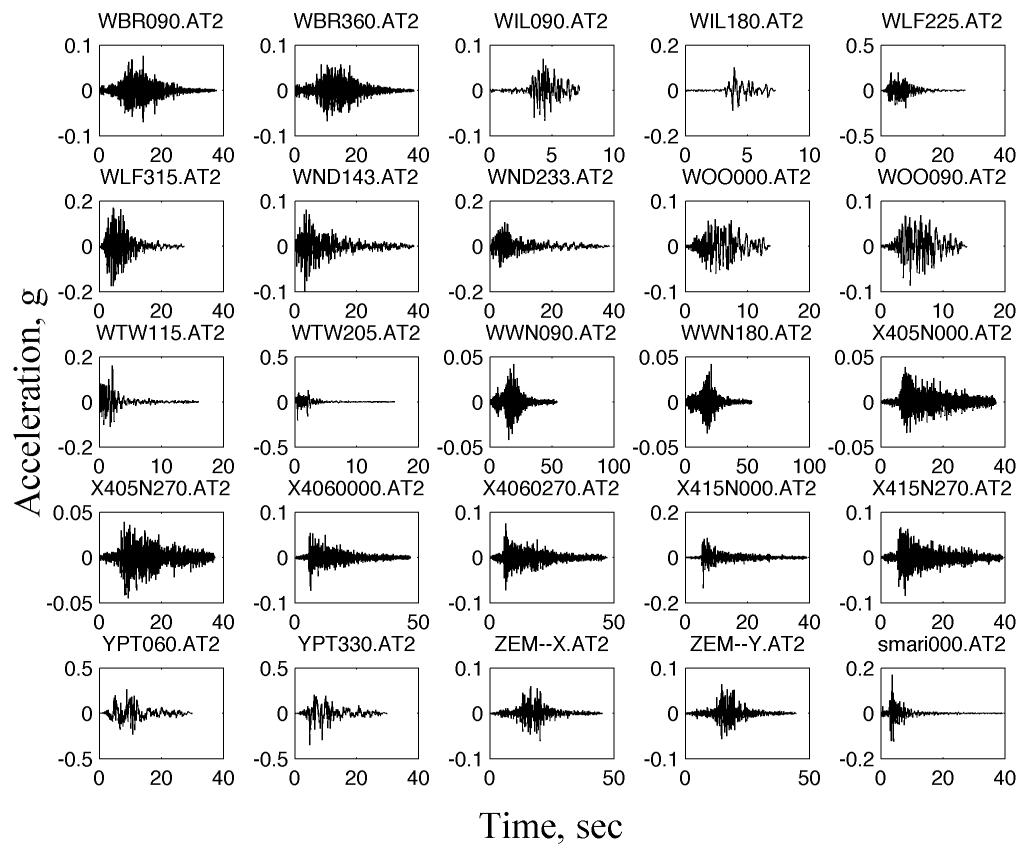


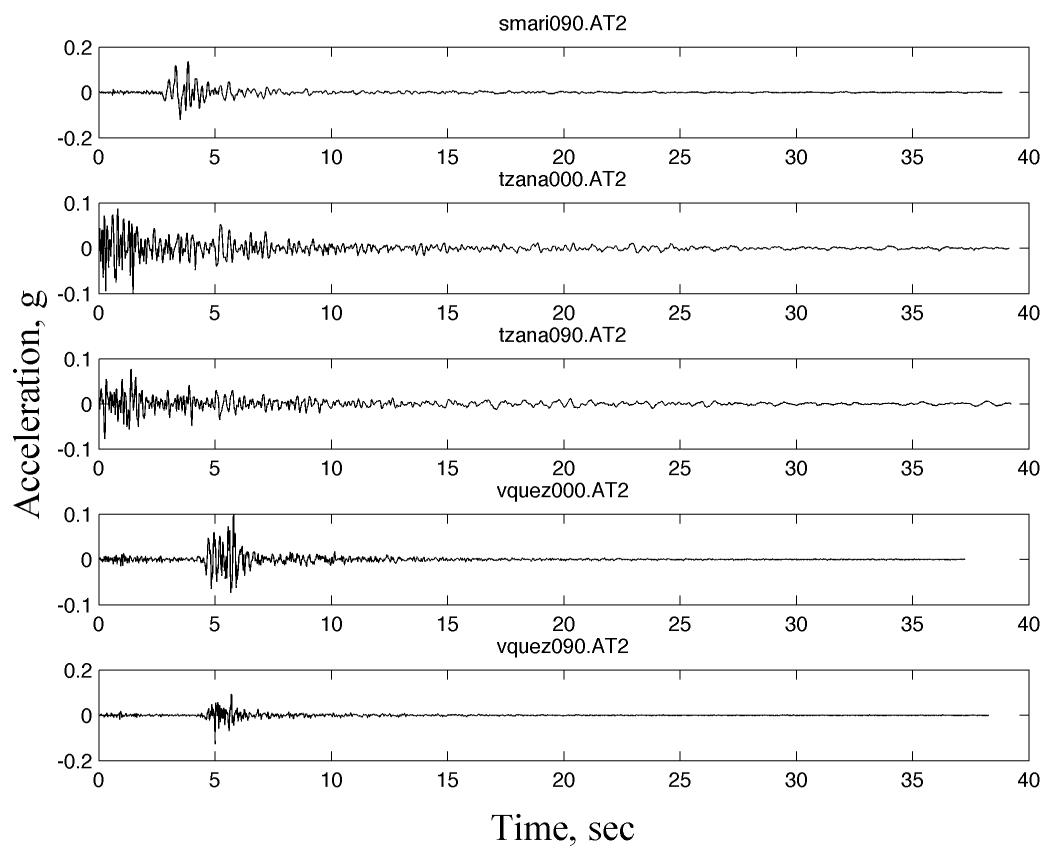












## Appendix II      Acquired Eigenquakes

Acceleration time histories of 140 eigenquakes obtained from the 530-record database in Appendix I are shown below.

