VECTOR-VALUED GROUND MOTION INTENSITY MEASURES FOR PROBABILISTIC SEISMIC DEMAND ANALYSIS

A DISSERTATION
SUBMITTED TO THE DEPARMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Jack W. Baker
September 2005

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

C. Allin Cornell
(Principal Adviser)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Helmut Krawinkler
(Co-adviser)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Gregory Deierlein

Approved for the University Committee on Graduate Studies.

Acknowledgements

This work was supported primarily by the National Science Foundation, under Award Number EEC-9701568 through the Pacific Earthquake Engineering Research Center (PEER). Additional support was provided by the Shah family fellowship and the East Asia Pacific Summer Institutes (funded by the Japan Society for Promotion of Science and the National Science Foundation). This support is greatly appreciated.

The author is grateful to Dr. Luis Ibarra, Dr. Fatemeh Jalayer and Dr. Ricardo Medina for generously sharing the structural analysis models and results that were used extensively in this dissertation. Thanks to Dr. Paul Somerville and Dr. Hong Kie Thio for providing vector-valued ground motion hazard analysis results. The author would like to thank Professor Greg Deierlein, Professor Helmut Krawinkler and Professor Eduardo Miranda for taking the time to review this work and provide helpful feedback.

Thanks to all of the students, staff and faculty associated with the John A. Blume Earthquake Engineering Center. The atmosphere of academic excellence and collegiality fostered there played a big role in my productivity and my enjoyment of the PhD experience. In particular, Hesaam Aslani, Paolo Bazzurro, Paul Cordova, Curt Haselton, Nico Luco, Polsak Tothong, Gee Liek Yeo, and Farzin Zareian have provided valuable support and feedback on this research.

Most importantly the author thanks his advisor and mentor, Professor C. Allin Cornell, whose insight and enthusiasm have made this research experience incredibly rewarding. It was an honor to be his student.

Contents

Chapter 8 – Correlation of Response Spectral Values for Multi-Component Ground Motions **203**

Chapter 9 – Conclusions **225**

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Quantitative assessment of risk to a structure from earthquakes poses significant challenges to analysts. It is a multi-disciplinary problem, incorporating seismology and geotechnical engineering to quantify the shaking which a structure might experience at its base, structural engineering to quantify the structure's response and the resulting damage, as well as finance, public policy, construction cost estimating, etc., to estimate social and economic consequences of this damage. The uncertainties present in many aspects of this problem also require that the assessment be made in terms of probabilities, adding a further layer of complexity.

A formal process for solving this problem has been developed by the Pacific Earthquake Engineering Research (PEER) Center. There are several stages to this process, consisting of quantifying the seismic ground motion hazard, structural response, damage to the building and contents, and resulting consequences (financial losses, fatalities, and business interruption). Each stage of the process is performed in formal probabilistic terms. The process is also modular, allowing the stages to be studied and executed independently, and then linked back together, as illustrated in Figure 1.1. For this method to be tractable and transparent, it is helpful to formulate the problem so that each part of the assessment is effectively independent. The independent assessment modules are then linked together using intermediate output variables, or "pinch point" variables (Kaplan and Garrick 1981). In the PEER methodology the intermediate variables are termed Intensity Measure (IM), Engineering Demand Parameter (EDP) and Damage Measure (DM). The final consequences, termed Decision Variable (DV), could also be considered a pinch point. An important assumption in this methodology is that what follows in the analysis is only dependent on the values of the pinch point variables, and not on the scenario by which it was reached (e.g., the response of the structure depends only upon the intensity measure of the ground motion, with no further dependence on variables such as the magnitude or distance of the causal earthquake). Further, the relationship between each of the stages is Markovian: given knowledge of EDP, the damage to building elements is independent of IM. This model relies heavily on assumptions of conditional independence between analysis stages, and these assumptions should be verified before the model is applied to a given problem. If the assumptions are not valid, then modifications to the model are required before proceeding. The use of vector-valued intensity measures is one such modification.

This multi-stage methodology has had success in other complex Probabilistic Risk Assessment problems (Kaplan and Garrick 1981, NUREG 1983, Garrick 1984), and a significant effort has been made to prepare this methodology for practical earthquake risk assessment applications (Cornell and Krawinkler 2000, Moehle and Deierlein 2004).

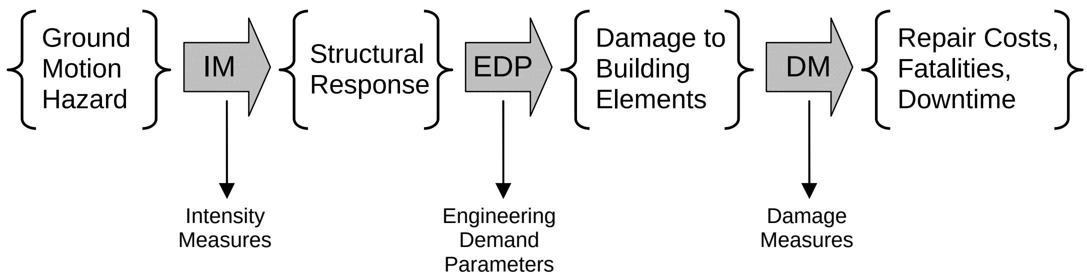


Figure 1.1: Schematic illustration of the Performance-Based Earthquake Engineering model and the pinch points IM, EDP and DM.

1.2 The role of intensity measures

This dissertation focuses on the “intensity measure” pinch-point of Figure 1.1, which links the ground motion hazard with the structural response. Ground motion hazard is computed using Probabilistic Seismic Hazard Analysis, and the typical output is the mean annual frequency of exceeding various levels of peak ground acceleration or a spectral response value at a given period. Structural analysis is then performed in order to obtain a probabilistic model for structural response as a function of this intensity measure parameter (peak ground acceleration or the spectral response value). Spectral acceleration at the first-mode period of the structure ($Sa(T_1)$) is the intensity measure used most frequently today for analysis of structures.

1.2.1 Concerns regarding intensity measures

The intensity measure approach is appealing because it allows the analysis stages shown in Figure 1.1 to be performed independently. However, the assumption that structural response depends only upon the IM parameter or parameters, and not on any other properties of the ground motion, must be carefully verified. Luco and Cornell (2005) have termed this required condition “sufficiency.” If the sufficiency condition is not met, then the estimated EDP distributions will not only depend upon IM, but also upon the properties of the records selected for analysis. If the properties of the selected records do not in some sense match the properties of the records that the real structure will be subjected to, then a biased estimate of structural response will result. This will be demonstrated more precisely in the following section.

Several researchers have proposed alternative methods of linking ground motion hazard and structural response that do not require the use of intensity measures (Han and Wen 1997, Bazzurro et al. 1998, Jalayer et al. 2004). These procedures should not face the same sufficiency concerns as the intensity-measure-based procedure. However, the fully probabilistic methods that avoid use of intensity measures tend to require a much greater number of ground motions and structural analyses. Further, the need for a large number of ground motions often necessitates the simulation of artificial ground motions rather than use of recorded ground motions. Simulation of ground motions is commonly accepted in areas such as the Eastern United States where the number of recorded motions is very small. But in seismically active areas such as the Western United States where recorded ground motions are more plentiful, the use of simulated ground motions is more questionable, as most simulation procedures have not been fully validated to confirm that simulated ground motions are consistent with observed ground motions. Further, some codes (e.g., ASCE 2005) require that real ground motions be used for dynamic analysis. The concerns with simulated ground motions, the added computational expense of performing many analyses and the complications involved when the ground motion hazard and structural response cannot be treated independently mean that the intensity measure approach still has many advantages as a method for assessing seismic risk to structures.

1.2.2 Benefits of vector-valued intensity measures

One way to address concerns about intensity measures is to increase the number of parameters in the intensity measure, so that it more completely describes the properties of the ground motions. The benefit of these so-called "vector-valued intensity measures" can be explained in more detail using the following argument, which has also been made by others (Shome and Cornell 1999, Bazzurro and Cornell 2002). Consider a structural response parameter EDP, which is potentially dependent upon a vector of two ground motion parameters: IM_1 and IM_2. The rate of exceeding a specified value of EDP, z, can be computed using knowledge of the conditional distribution of EDP given IM_1 and IM_2, along with the joint rates of occurrence of the various levels of IM_1 and IM_2. Mathematically, this can be written as

$$\lambda_{EDP}(z) = \int_{IM_1}\int_{IM_2} G_{EDP|IM_1,IM_2}(z \mid im_1, im_2) f_{IM_2|IM_1}(im_2 \mid im_1) \left| d\lambda_{IM_1}(im_1) \right| \tag{1.1}$$

where $\lambda_{EDP}(z)$ is the annual rate of exceeding the EDP level z. The term $G_{EDP|IM_1,IM_2}(z \mid im_1, im_2)$ denotes the probability that EDP is greater than *z*, given an earthquake ground motion with intensity such that $IM_1=im_1$ and $IM_2=im_2$. The term $f_{IM_2|IM_1}(im_2 \mid im_1)$ denotes the conditional probability density function of IM_2 given IM_1, *at the site being considered*, and $\lambda_{IM_1}(im_1)$ is the annual rate of IM_1 exceeding im_1 at the site being considered. This vector-IM-based calculation will be used extensively in the dissertation that follows. A fundamental question that motivates the work of this dissertation is, "why is this vector-IM-based formulation any better than one which uses only a scalar IM?" For example, what if the ground motion parameter IM_2 were ignored, and only IM_1 was used in the assessment? Then the distribution of EDP would not depend (explicitly) on IM_2, and Equation 1.1 would instead look like

$$\lambda_{EDP}(z) = \int_{IM_1} G_{EDP|IM_1}(z \mid im_1) \left| d\lambda_{IM_1}(im_1) \right| \tag{1.2}$$

This scalar-IM-based equation is simpler than Equation 1.1, which makes it appealing if it is accurate. Consider a case where IM_2 actually did not affect structural response (given IM_1). That is, the conditional distribution of EDP did not depend upon IM_2 (given IM_1). Then $G_{EDP|IM_1,IM_2}(z \mid im_1, im_2)$ would be equal to $G_{EDP|IM_1}(z \mid im_1)$, and Equations 1.1 and 1.2 would be exactly equal. This should be intuitive: if the ground motion parameter IM_2 has no effect on structural response, then a calculation which considers IM_2 (Equation 1.1) should not produce a different answer than a calculation which does not (Equation 1.2).

However, if the parameter IM_2 *does* affect response, then more care is needed. If one uses only a scalar IM to estimate structural response in this case, then the estimate of EDP given IM_1 depends implicitly upon the distribution of IM_2 values present in the record set used for analysis. This can be seen by expanding Equation 1.2 using the total probability theorem (Benjamin and Cornell 1970), to explicitly note that EDP is a function of both IM_1 and IM_2

$$\lambda_{EDP}(z) = \int_{IM_1} \int_{IM_2} G_{EDP|IM_1,IM_2}(z \mid im_1, im_2) \tilde{f}_{IM_2|IM_1}(im_2 \mid im_1) \left| d\lambda_{IM_1}(im_1) \right| \tag{1.3}$$

This is similar to Equation 1.1, but with one important difference. The response as a function of IM_2 is dependent upon the IM_2 values of the records used, rather than on the actual IM_2 values occurring at the site of interest[1]. That is, the conditional distribution of IM_2 given IM_1 occurring

[1] If a recording instrument had been located at the site of interest for thousands of years, then it would be possible to use the recorded ground motions in conjunction with Equation 1.2 and get the correct answer. That is, the total probability theorem on which Equation 1.2 is based does not cause any errors. It is the substitution of records which may not be representative of ground motions at the specified site that can potentially cause an error.

at the site, $f_{IM_2|IM_1}(im_2 \mid im_1)$, has been replaced by the conditional distribution of IM_2 given IM_1 *within the record set used for analysis*, which is denoted $\tilde{f}_{IM_2|IM_1}(im_2 \mid im_1)$.

By comparing Equation 1.1 to Equation 1.2 (and its equivalent expanded form in Equation 1.3), it is clear that there are two ways to obtain the "correct" answer of Equation 1.1 without using a vector IM. First, if the structural response, EDP, is not dependent upon the ground motion parameter IM_2, then the two approaches are equivalent. Second, if EDP *is* dependent upon IM_2, then the scalar-IM-based answer will equal the vector-IM-based answer only if the conditional distribution of IM_2 given IM_1 within the record set equals the conditional distribution of IM_2 given IM_1 at the site of interest. Therefore, if one would like to use a scalar IM for analysis, either conditional independence of EDP and IM_2, given IM_1, must be verified (i.e., IM_1 must be sufficient with respect to IM_2), or the records used for analysis must be carefully selected so that they match the required conditional distribution of IM_2 as described above[2]. If one is willing to perform a vector-IM-based analysis, then concerns about proper record selection (with respect to the additional ground motion parameter) are no longer necessary. This discussion all assumes that there is no remaining dependence upon any other parameters—an assumption that will be investigated.

Viewed in this way, the use of vector-valued intensity measures is closely related to the issue of IM sufficiency. The search for additional useful IM parameters is closely related to the verification of sufficiency with respect to the same parameters. Thus, vector-valued intensity measures are a direct way of reducing insufficiency problems, and thus reducing the potential for bias in structural response estimates.

A further goal that can be achieved using a vector-valued IM is increased estimation accuracy. If by adding IM_2 to the analysis, one can explain a large portion of a ground motion's effect on a structure, then the remaining "variability" in EDP given IM_1 and IM_2 will be reduced. This means fewer nonlinear dynamic analyses will be needed to characterize the relationship between structural response and the intensity measure. An IM that results in small variability of EDP given IM is termed "efficient" by Luco and Cornell (2005). A vector-valued IM can achieve significant gains in efficiency, as will be seen in this dissertation. Thus, an effectively chosen

[2] This suggests that careful record selection is a valid alternative to adoption of vector IMs. This result has guided previous attempts to select records with proper distributions of Magnitude, Distance, or other parameters (e.g., Shome 1999, Somerville 2001, Jalayer 2003). In Chapter 6, empirical results will also be used to show that careful record selection can be used in place of vector IMs. Note, however, that the target distribution can and does change as a function of the IM level, meaning that many record sets may need to be selected over a range of IM levels.

vector IM can reduce the number of dynamic analyses that must be performed to assess a structure's performance.

For the above reasons, the vector-valued IM approach can bring significant benefits to the seismic performance assessment approach described in Figure 1.1. By "widening" the IM pinch point to incorporate multiple parameters, more information can be transferred between the ground motion hazard and structural response stages of the analysis. This will reduce the possibility of biasing the structural response estimates, as well as reduce the number of dynamic analyses needed to estimate the relationship between IM and the structural response. Along with these gains, the benefits of keeping the assessment procedure in a modular form are maintained.

Several researchers have recognized the benefits of improved sufficiency and efficiency in an intensity measure, and have investigated the use of vector IMs for this purpose. Investigations for prediction of building response have been performed by several authors (Shome and Cornell 1999, Bazzurro and Cornell 2002, Vamvatsikos 2002, Conte et al. 2003, Luco et al. 2005). A similar approach has been applied to prediction of rock overturning during earthquake shaking (Purvance 2005). Other researchers have taken a slightly different approach and attempted to develop more advanced scalar parameters, rather than switching to a vector parameter (Cordova et al. 2001, Taghavi and Miranda 2003, Luco and Cornell 2005), but the goal is essentially the same. This dissertation aims to extend the findings from these earlier investigations, as will be discussed in the following section.

1.3 Contributions to the vector-valued intensity measure approach

There are three major challenges associated with the selection and use of a vector-valued intensity measure, and in this dissertation contributions are made to all three.

1.3.1 Choice of intensity measure parameters

The parameters in the vector-valued intensity measure should be chosen in order to convey the most possible information between the ground motion hazard and the structural response stages of analysis. This requires identifying parameters that most affect the structure under consideration. It is also desirable that the parameters have low correlation, so that they are not describing the same properties of the ground motion, and so that their individual effects are easily separable when predicting structural response. In this dissertation, a new parameter termed ε is proposed for use in an intensity measure, and its effectiveness is demonstrated. Following the work of others, a vector of parameters consisting of spectral acceleration values at multiple periods is also considered, but here a new method for selecting an optimal set of periods is proposed. Use of these vector IMs to characterize near-fault ground motions is also considered for the first time. These investigations were performed on a range of 2D frame structures, using models developed by others (Jalayer 2003, Ibarra 2003, Medina and Krawinkler 2003), in order confirm the findings for a range of structures.

1.3.2 Ground motion hazard analysis

In order to use a vector-valued intensity measure in this methodology, one must perform probabilistic ground motion hazard analysis for this vector. Bazzurro and Cornell (2002) describe the mathematics of this procedure, especially for a vector consisting of multiple spectral acceleration values. Knowledge of the correlations among these spectral values is needed in order to perform the analysis, and in this dissertation a new predictive model for these correlations is presented. The model improves upon previous models for spectral values of a single component of ground motion, but also provides the first predictions of correlations among perpendicular horizontal components, or among horizontal and vertical components of a ground motion. This will allow vector IMs to be used for analysis of three-dimensional structures.

Difficulties associated with computing the ground motion hazard may also influence the choice of IM parameters. For example, Shome and Cornell (1999) considered a vector-valued intensity measure consisting of spectral acceleration plus magnitude or distance, and they were able to easily compute the hazard for this vector by using standard spectral acceleration hazard analysis along with disaggregation on magnitude or distance, rather performing a more difficult vector-valued hazard analysis. This disaggregation-based hazard analysis will be used for the newly proposed ε parameter, allowing for the ground motion hazard analysis to be simplified in this case.

1.3.3 Estimation of structural response based on a vector of intensity measure parameters

In order to take advantage of a vector of parameters, structural response must be predicted as a *joint* function of *each* of the IM parameters (i.e., one must estimate $G_{EDP|IM_1,IM_2}(z\,|\,im_1,im_2)$ in Equation 1.1). Structural response is estimated statistically based on the IM values associated with each of a set of ground motions and the resulting structural responses obtained by performing nonlinear dynamic analysis of a structure with the ground motions as inputs. Several methods of characterizing the relationship between the IM parameters and the resulting structural response are proposed in this dissertation, and their relative advantages and disadvantages are discussed. Several of these estimation procedures involve scaling of ground motions, and the validity of this scaling is investigated. Another challenge related to prediction of structural response is combining these predictions with the ground motion hazard analysis to compute the annual rates of exceeding various structural response levels. This problem is discussed as well.

In this dissertation, progress is made in all three of the above areas. The primary focus, however, is on selection of parameters for the vector and on use of a vector of parameters to predict structural response.

1.4 Organization

This dissertation addresses several issues related to vector-valued intensity measures for assessment of structural performance due to seismic hazard. Chapter 2 discusses several methods for estimation of structural response based on a vector-valued intensity measure. Chapters 3 through 6 deal primarily with selection of improved vector-valued intensity measures, and their effectiveness in various situations. Chapters 6 through 8 address questions associated with ground motion hazard analysis.

Chapter 2 focuses on the mathematical implementation of vector-valued intensity measures in estimation of structural response. Depending on the desired accuracy and allowable computational effort, several options exist for modeling structural response at a given level of earthquake intensity. The choice of a model is somewhat independent of the actual IM parameters used. Various options have been developed for traditional scalar intensity measures, and these are reviewed briefly. The options for vector-valued intensity measures are then introduced, and their various advantages and disadvantages are described. In later chapters, typically only a single method is used, so this chapter provides an opportunity to compare different methods, and justify the methods used later in the dissertation.

Chapter 3 presents a novel vector IM consisting of spectral acceleration at the first-mode period of the structure, denoted $Sa(T_1)$, plus a parameter termed ε. It is seen that ε has a significant effect on the response of structures, because it is an implicit indicator of spectral shape (more specifically, it tends to indicate whether $Sa(T_1)$ is in a peak or a valley of the spectrum). This vector is particularly appealing professionally because it is easy to find the joint distribution of $Sa(T_1)$ and ε from standard PSHA and its associated disaggregation, with no need for specialized hazard analysis software.

Chapter 4 also investigates a vector IM that attempts to capture spectral shape, but this time it is done directly by combining $Sa(T_1)$ with information about spectral acceleration at a second period. This IM is seen to provide significant predictive power, especially when the second period is chosen properly with regard to the structure and nonlinearity level of primary interest.

Chapter 5 considers a special class of ground motions termed near-fault pulse-like ground motions. These motions, which sometimes occur at certain site/fault geometries, are potentially very damaging to structures. Further, an IM consisting of spectral acceleration alone is not effective in predicting the unusual effects of these motions on structures. In this chapter, the previously proposed vector-valued IMs are evaluated with regard to their effectiveness in predicting the response of near-fault ground motions. It is seen that while the vector consisting of $Sa(T_1)$ and ε is not useful in this situation, the vector considered in Chapter 4 can be quite effective. The implications for incorporating the effect of near-fault ground motions using this IM are investigated.

The vector-valued IMs considered in the early chapters of this dissertation provide insight into which properties of earthquake ground motions have the most effect on structural response. In **Chapter 6**, this knowledge is applied to the problem of appropriately selecting earthquake ground motions for use in structural analysis. By carefully selecting records, one can obtain the benefits of a vector-valued IM without introducing any complexity in the analysis (i.e., one can use scalar-IM-based methods rather than the somewhat more complicated vector-based methods outlined in Chapter 2). But in order to select records to match some target, one must first identify what the target is. This problem is discussed, and new guidelines for record selection are proposed, using intuition gained from the earlier vector-valued IM work.

In **Chapter 7**, an issue is addressed that is relevant to all PBEE assessments, regardless of whether a vector or scalar IM is used. As discussed throughout this dissertation, the PBEE procedure relies heavily on the IM pinch point, which typically includes a spectral acceleration in both the scalar and vector IM formulations (at least for assessment of buildings and bridges). However, in past practice, an error has been made frequently because seismologists and structural engineers use slightly different definitions for “spectral acceleration.” This error typically results in unconservative conclusions about the risk to the structure of interest. The nature of this error is explored in Chapter 7, and several fairly simple methods of correcting it are presented. The correction is alluded to in many of the other chapters, and so this chapter explains the correction in more detail.

Chapter 8 provides supporting material that will be helpful for assessments utilizing a vector IM consisting of multiple spectral acceleration values (such as that used in Chapter 4). In order to compute the *joint* ground motion hazard for these multiple spectral acceleration values, it is necessary to know the correlation of these values within a given record. This chapter presents new predictions of correlation coefficients for single-component ground motions measured at two differing periods, and also across orthogonal components of three-dimensional ground motions.

For correlations within a vertical ground motion or across orthogonal components of a ground motion, these predictions are believed to be the first of their kind. Increased knowledge of response spectrum correlations has facilitated much of the work in earlier chapters, and will also allow vector-valued IM hazard analysis to be extended to new areas such IMs for 3-dimensional ground motions.

Finally, **Chapter 9** will summarize the important contributions and findings of this dissertation and discuss future extensions of this research. Recommendations will be made regarding record-selection criteria and use of vector-valued IMs in research and practice, at least within the scope of the work in this dissertation.

The chapters of this dissertation are designed to be largely self-contained because they have been or will be published as individual journal articles. Because of this, there is some repetition of background material. In addition, notational conventions were chosen to be simple and clear for the topic of each chapter rather than for the dissertation as a whole; because of this, the notational conventions may not be identical for each chapter. Apologies are made for any distraction this causes when reading the dissertation as a continuous document.

Chapter 2

Estimation of EDP given IM

2.1 Abstract

There are many possible methods to estimate probabilistic structural response as a function of ground motion intensity using the results of nonlinear dynamic analyses. Options include using regression analysis on structural analysis results from a set of unscaled (or uniformly scaled) ground motions, or fitting a probability distribution to the analysis results from a set of records, each of which has been scaled to a target ground motion intensity level. Most studies use only a single method among the several options, so it is difficult to compare the relative advantages and disadvantages of alternative methods. In this chapter, several methods for estimating probabilistic structural response are described and discussed. Several novel methods for utilizing vector-valued intensity measures are presented, and the differences between methods for scalar- and vector-valued intensity measures are highlighted. The recommended method for using a vector-valued intensity measure consists of scaling records to the primary intensity measure parameter (e.g., spectral acceleration at the first-mode period of the structure), and then using regression analysis to measure the effect of the additional intensity measure parameters. This method does not severely restrict the functional form of the median response versus intensity measure relationship, while also not requiring excessive numbers of structural analyses to be performed. "Hybrid" estimation methods that obtain the benefit of vector-valued intensity measures through simplified methods such as informed record selection are also discussed. In general, there is a tradeoff between the restrictiveness of the assumptions made regarding the response behavior and the number of structural analyses needed for estimation.

2.2 Introduction

Estimation of probabilistic structural response as a function of ground motion intensity is a major step in performance-based earthquake engineering assessments (Cornell and Krawinkler 2000, Deierlein 2004). The objective of this step is to estimate the probability distribution of a structural response parameter as a function of the earthquake ground motion intensity (quantified by one or several parameters termed an "intensity measure"). Several methods are available for obtaining this estimate, and they are discussed in this chapter. The methods have been developed elsewhere for scalar intensity measures consisting of a single parameter (Jalayer 2003, Chapters 4 and 5). These methods are reviewed and then extended to *vector-valued* intensity measures consisting of two or more parameters.

Following the terminology conventions of the Pacific Earthquake Engineering Research (PEER) center, the structural response parameter is termed an Engineering Demand Parameter, or *EDP*, in the discussion that follows. This *EDP* can be any response parameter of interest, although in the examples here the primary *EDP* used will be the maximum interstory drift ratio observed on any story of a building during the ground motion shaking. The intensity measure will be abbreviated as *IM* in much of the discussion that follows.

It is seen that the optimal estimation method for a particular application depends on the sample size used for estimation (i.e., the number of nonlinear dynamic analyses performed), the level of nonlinearity in the structure, whether the intensity measure is scalar-valued or vector-valued, and the level to which the response can accurately be parameterized.

The estimate of *EDP*|*IM* can be combined with a ground motion hazard curve to compute the mean annual rate of exceeding a given *EDP* level ($\lambda_{EDP}(y)$) using the equation

$$\lambda_{EDP}(y) = \int_{IM} G_{EDP|IM}(y \mid im) \left| \frac{\mathrm{d}\lambda_{IM}(im)}{\mathrm{d}im} \right| \mathrm{d}im \tag{2.1}$$

where $G_{EDP|IM}(y \mid im)$ is the Complementary Cumulative Distribution Function (CCDF) of *EDP*|*IM* (the probability that $EDP > y$, given that $IM = im$) obtained using the estimation methods described in this chapter. The term $\lambda_{IM}(im)$ is the mean annual rate of exceeding the *IM* level *im*, obtained using Probabilistic Seismic Hazard Analysis, a procedure which is well documented elsewhere (e.g., Kramer 1996, McGuire 2004). Equation 2.1 is used often in research, but the estimate of *EDP*|*IM* is normally obtained using a single method. In this chapter, *EDP*|*IM* is obtained using several methods, and Equation 2.1 is evaluated using each estimate to determine whether the resulting estimates of $\lambda_{EDP}(y)$ are equivalent.

2.3 Estimation from a statistical inference perspective

Estimation of the properties of a random variable (in this case, *EDP* given an *IM* level) from a finite sample of data is referred to as statistical inference. In this chapter, statistical inference tools are applied while highlighting particular concerns of this specific application in performance-based engineering.

There are two classes of approaches for estimating probability distributions. Using a *parametric* approach, one assumes that the random variable *EDP* has some probability distribution (e.g. normal or lognormal) which is defined by only a few parameters. One then estimates these parameters to define the distribution (Rice 1995). Generally the choice of a distribution is made based on past experience, and statistical tests exist to identify when a data set is not well represented by the specified distribution (although with a small data set it is difficult for such methods to confidently identify a lack of fit).

Non-parametric approaches are also available, which do not require assumptions about the distribution of the data (Lehmann and D'Abrera 1998). These approaches have the advantage being robust when the data does not fit a specified parametric distribution, at a cost of reduced efficiency when the data *does* fit a specified parametric distribution.

Complicating the *EDP* estimation problem relative to simple statistical inference situations is the fact that the distribution of structural response (*EDP*) is a function of the ground motion intensity measure (*IM*) (i.e., structural response tends to be larger when the intensity of ground motion is greater). For some applications the distribution of *EDP* is needed over a continuous range of *IM*'s, or at least a somewhat finely discretized set. In other applications, the distribution of *EDP* may be needed for only a single *IM* level. The choice of an estimation method used will depend on whether distribution estimates are needed for a (wide or narrow) range of *IM* levels or for only a single level.

2.4 Analysis using a scalar intensity measure

Jalayer (2003, Chapters 4 and 5) has explored in detail the topic of *EDP* estimation using scalar intensity measures. A brief summary of methods is presented here, and they will be generalized in Section 2.5 to incorporate vector-valued intensity measures.

To facilitate comparison of the methods, illustrative results are provided for an example structure. The structure is a 1960s-era reinforced-concrete moment-frame building which will be used extensively throughout this dissertation. A nonlinear, stiffness and strength degrading 2D model of the transverse frame created by Jalayer (2003) is used here. Spectral acceleration at 0.8s (the first-mode period of the building) is used as the *IM*, and maximum interstory drift ratio is used as the *EDP*. These results are presented only as illustration of the methods of analysis; the effectiveness of intensity measures will not be considered until later chapters.

2.4.1 Regress on a "cloud" of ground motions

With this method, the structure is subjected to a set of ground motions with associated *IM* values. The records are either left unmodified, or all records are scaled by a constant factor if the unmodified records are not strong enough to induce the structural response level of interest. The set of *IM* values and their associated *EDP* values resulting from nonlinear dynamic analysis are sometimes referred to as a "cloud," because they form a rough ellipse when plotted (see Figure 2.1). Regression can be used with this cloud of data to compute the conditional mean and standard deviation of *EDP* given *IM*. A linear relationship (determined using regression) between the logarithms of the two variables often provides a reasonable estimate of the mean value of ln*EDP* over a small range, yielding the model

$$\ln EDP = a + b \ln IM + e \tag{2.2}$$

where a and b are constant coefficients to be estimated from linear regression (Neter et al. 1996) and e is a zero-mean random variable representing the remaining variability in ln*EDP* given *IM*. Then the mean value of ln*EDP* given a specified *IM* value *im* is

$$E[\ln EDP \mid IM = im] = \hat{a} + \hat{b} \ln im \tag{2.3}$$

where $\hat{a}$ and $\hat{b}$ are the regression estimates of the coefficients a and b. If e is assumed to have a constant variance for all *IM* (a reasonable assumption over a small range of *IM*, but usually less appropriate for a wide range of *IM*), then its variance can be estimated as

$$V\hat{a}r[e] = \sqrt{\frac{\sum_i^n \left(\ln edp_i - \left(\hat{a} + \hat{b} \ln im_i\right)\right)}{n-2}} \tag{2.4}$$

where edp_i and im_i are the *EDP* and *IM* values of record i, and n is the number of records. If ln*EDP*|*IM* is further assumed to have a Gaussian distribution, then the estimated conditional probability of exceeding an *EDP* level y given *IM*=*im* is

$$G_{EDP|IM}(y \mid im) = 1 - \Phi\left(\frac{\ln y - (\hat{a} + \hat{b} \ln im)}{\sqrt{V\hat{a}r[e]}}\right) \tag{2.5}$$

where $G_{EDP|IM}(y \mid im)$ is the complementary cumulative distribution function (CCDF) of *EDP* given *IM* and $\Phi(\bullet)$ is the cumulative distribution function of the standard Gaussian distribution.

To illustrate the data used in this procedure and the resulting mean estimate, a regression is performed on the results of 60 records that have been scaled by a factor of two and applied to the example structure. The resulting cloud of data is shown in Figure 2.1[3].

[3] Throughout this dissertation, *IM* is displayed on the on the x-axis (abscissa) of figures and *EDP* on the y-axis (ordinate), in contrast with many publications elsewhere. This is done to emphasize that *EDP* is a function of *IM*. Mathematical convention suggests that the independent variable (*IM* in this case) should be

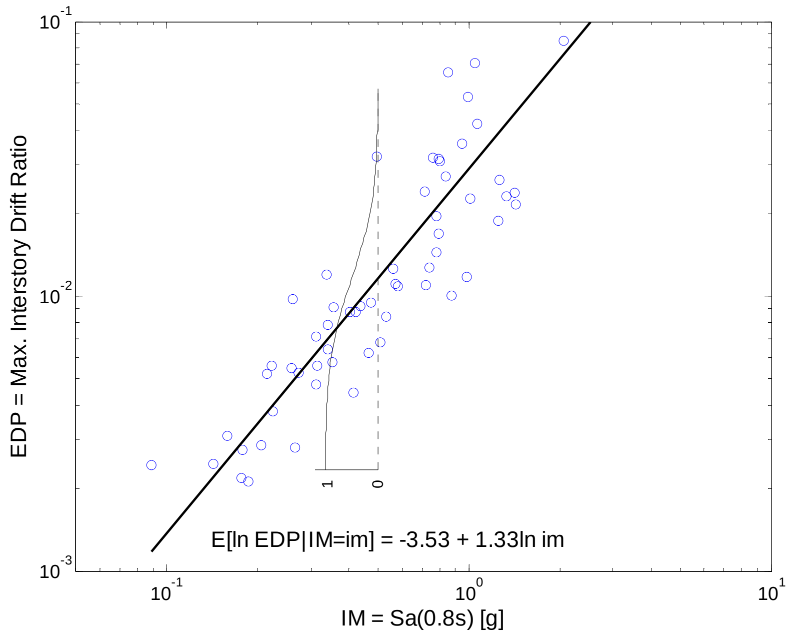


Figure 2.1. A cloud of *EDP*|*IM* data, the conditional mean value from linear regression, and the CCDF of *EDP* given *Sa*(0.8s) = 0.5g.

This method has more restrictions on the form of the conditional distributions than any other method to follow. In particular, it requires the relationship between *EDP* and *IM* to be linear with constant variance (more precisely, linear with constant variance after transformations—most commonly performed by taking logarithms). These restrictions may be appropriate only over a limited range of *IM* levels, but they have the benefit of reducing the computational expense of the estimation (fewer data are needed because there are fewer parameters to estimate). In this basic format, the method also has the great advantage of being associated with a closed-form solutions to Equation 2.1 (Cornell et al. 2002). For these reasons this is the preferred analysis procedure for some situations. This method will be used in the examples of Chapter 7.

If the assumptions discussed above are not believed to be valid, and a closed-form solution to Equation 2.1 is not needed, then these assumptions can be relaxed. For example, if the conditional variance is not constant, then linear regression can still be used (perhaps with weights on the data corresponding to the estimated variance), and the variance can be estimated separately for several *IM* ranges or estimated continuously using a windowed analysis. Also, a linear relationship may not be reasonable for the entire *IM* range of interest, and so the regression should be either limited

---

plotted on the abscissa with the dependent variable (*EDP*) on the ordinate. Once the *IM* becomes a vector of two parameters, this convention is an even more natural choice. It is hoped that the intuition provided by following mathematical convention overcomes the distraction of breaking with the typical display of *IM*-

to the *IM* regime where its form is adequate or conservative (e.g., SAC 2000a, b, Jalayer 2003, Luco and Cornell 2005, Aslani 2005) or a piecewise linear relationship may be fit (e.g., Mackie and Stojadinovic 2003, p50). Finally, the simple framework outlined above does not include the possibility that a record may cause "collapse" of the structure (as indicated by large calculated displacements that would likely cause collapse of the real structure). Collapse can be incorporated by adding a supplemental probability-of-collapse estimation as a function of *IM*. The non-collapse responses are then accounted for by performing regression with only the records that do not cause collapse, and then multiplying the resulting estimated distribution by the probability of non-collapse estimated from the first step (e.g., Shome and Cornell 2000). With these modifications, many of the assumptions of the most basic cloud method can be relaxed. The resulting assessment then lies somewhere between the basic cloud method and the less restrictive methods described below, both in terms of the required number of analyses and the number of parametric assumptions made.

2.4.2 Scale records to the target IM level and fit a parametric distribution to response

If one is interested in estimating the distribution of *EDP* at a specified *IM* level *im*, another possible method consists of scaling a suite of records so that the *IM* value of each record is equal to *im*. (Record scaling is assumed to be valid for now, but investigated further in Section 2.8 and Chapter 6.) Nonlinear dynamic analysis can then be performed with each of these scaled records to obtain *EDP* values. The resulting plot of *EDP* vs *IM* looks like a "stripe" (see Figure 2.2). To estimate a distribution from this data, one can estimate the mean and standard deviation of the responses and use these values to fit a distribution[4]. For example, one can fit a normal distribution to ln*EDP* values, giving a Complementary Cumulative Distribution Function defined as

$$G_{EDP|IM}(y \mid im) = \Phi\left(\frac{\ln y - \hat{\mu}}{\hat{\beta}}\right) \quad (2.6)$$

where $\hat{\mu}$ and $\hat{\beta}$ are the sample mean and standard deviation, respectively, at the stripe of data with $IM = im$. An example of a stripe of 40 records with a fitted CCDF is shown in Figure 2.2.

EDP relationships used by others. That display is intended to maintain the structural convention of "force" versus deformation.

[4] It is also possible to fit the distribution using maximum-likelihood or by making the CCDF pass through two counted fractiles. The choice of fitting should not significantly affect the results when a large number of dynamic analyses are available. The method of moments is used here because it is the natural choice with the cloud regression method, and is also straightforward for the stripe-based method.

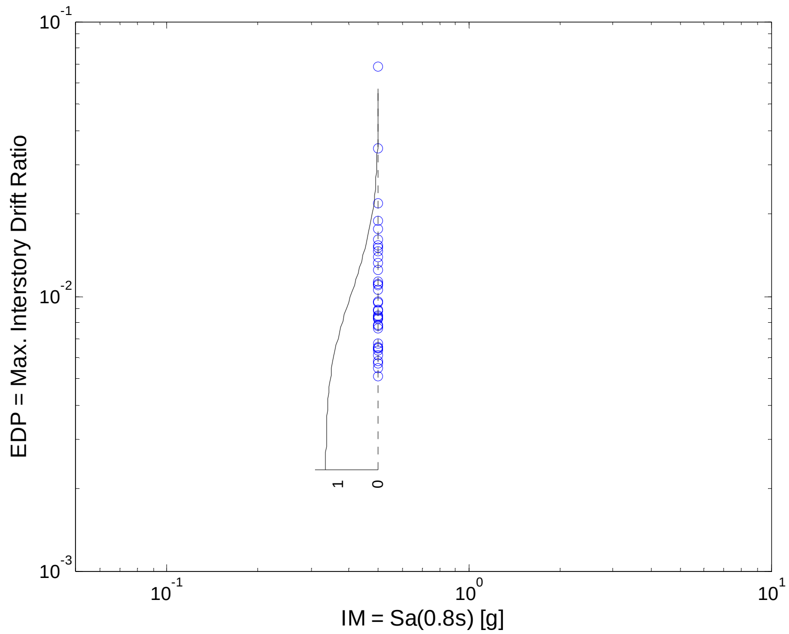


Figure 2.2. A stripe of data and its estimated complementary cumulative distribution function from a fitted parametric distribution.

In order to estimate *EDP* distributions at a range of *IM* values, one simply repeats this procedure at a range of *IM* values (either at every *IM* value of interest, or by analyzing a few *IM* values and interpolating). Multiple stripes of data are shown in Figure 2.3 (using a suite of 40 ground motions scaled to 25 spectral acceleration levels between 0.1g and 5g). It is apparent visually that the standard deviation of ln*EDP* is not constant over the range of *IM* considered here. It also appears that the mean value of ln*EDP* is not a linear function of ln*IM*. Thus, the most basic cloud method would need to be generalized in order to accurately model responses over the entire *IM* range shown here.

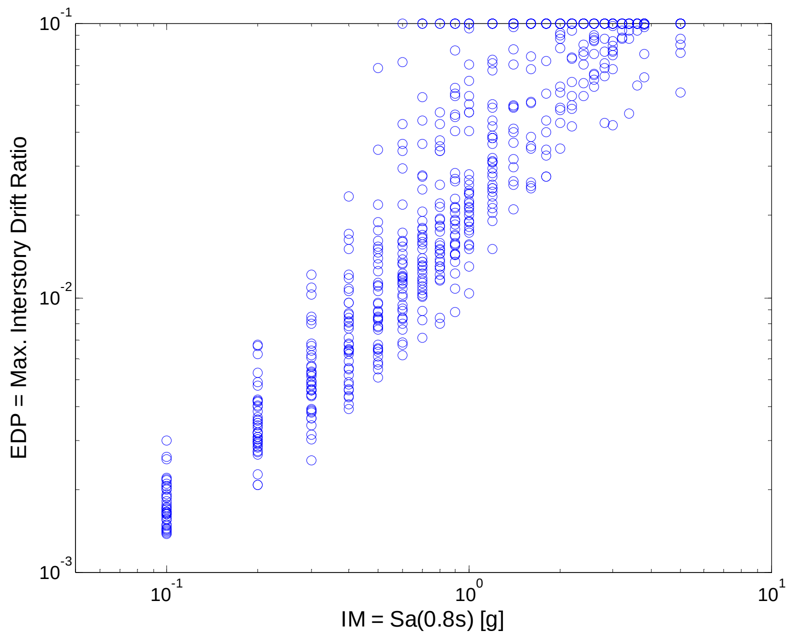


Figure 2.3. Multiple stripes of data used to re-estimate distributions at varying *IM* levels. Note that records that cause maximum interstory drift ratios of larger than 0.1 are displayed with interstory drift ratios of 0.1.

This method can easily account for records that cause collapse of the structure. This is done by first estimating the probability of collapse as the fraction of records in a stripe that cause collapse, and then fitting a parametric (often lognormal) distribution to the remaining records that do not collapse. With this formulation, the lognormal complementary cumulative distribution function is given by

$$G_{EDP|IM}(y \mid im) = 1 - \mathrm{P}(NC \mid IM = im)\Phi\left(\frac{\ln y - \hat{\mu}}{\hat{\beta}}\right) \tag{2.7}$$

where $\hat{\mu}$ and $\hat{\beta}$ are the sample mean and standard deviation, respectively, of the non-collapsing responses, and $\mathrm{P}(NC \mid IM = im)$ is the counted fraction of records at the $IM = im$ stripe that do not cause collapse. When estimating the probability of collapse at multiple *IM* levels, the probability can be counted at each $IM = im$ stripe, or a parametric function can be fit over a range of *IM* levels (Shome and Cornell 2000).

If response at only a single *IM* level is of interest (e.g., if answering “what is probability of exceeding *EDP* level *y* at the 10% in 50 year ground motion hazard level?”), then this is a direct method. The disadvantage of this method is that is potentially requires more structural analyses than the cloud method if response estimates are needed at many *IM* levels. One thousand dynamic analyses were used to produce Figure 2.3 for research purposes, although in practice this number could be reduced significantly.

2.4.3 Scale records to the target IM level and compute an empirical distribution for response

With this non-parametric method, a stripe of data is obtained in the same manner as in the previous section. However, instead of using a parametric CCDF to represent the distribution, an empirical CCDF is fitted (Lehmann and D'Abrera 1998). With this method the probability of exceeding an *EDP* level *y* is estimated by simply counting the fraction of records that cause a response larger than *y*:

$$G_{\theta|IM_i}(y \mid im) = \frac{\text{number of responses } > y}{\text{total number of records}} \tag{2.8}$$

An empirical CCDF is superimposed on a stripe in Figure 2.4. This method has the desirable property that that no assumptions are made regarding distributions or functional relationships between *EDP* and *IM*—the data provides all of the information. The eliminated assumptions have a cost, however: more data are typically needed to characterize well the conditional distributions. In addition, empirical distributions have can have difficulties estimating extreme values[5]. It is important to remember this in situations where extreme values are of concern. In Section 2.4.5, however, it will be seen that this method provides good accuracy with the large dataset used here.

[5] For example, an empirical distribution estimates zero probability of exceeding the largest response value observed, while we might expect larger values to in fact be possible. It should be noted however, that the expected value of $G_{EDP|IM}(y|im)$ estimated from an empirical distributions is unbiased even for these large response values. Methods exist to combine empirical distributions with models for extreme values (Deutsch and Journel 1997, p134), but the amount of effort required is likely not worth any gains that could be achieved for this problem.

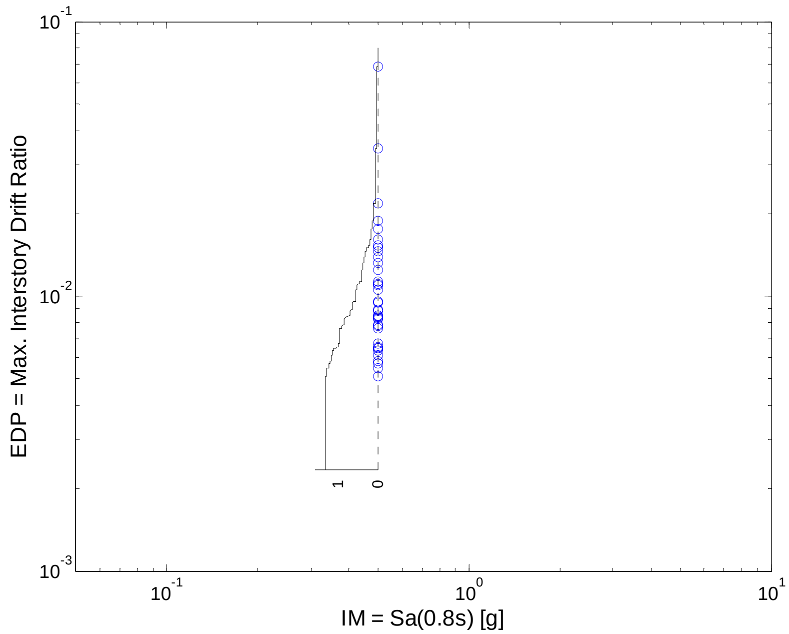


Figure 2.4. A stripe of data and its empirical CCDF.

2.4.4 Fit a distribution for IM capacity to IDA results

Unlike previous methods, with this method the distribution of *EDP*|*IM* is not estimated directly. Rather, the results from Incremental Dynamic Analysis (IDA) are used to determine the distribution of *IM* values that cause a given *EDP* level to be reached. In previous methods, our goal was to obtain

$$G_{EDP|IM}(y \mid im) = \mathrm{P}(EDP > y \mid IM = im) \tag{2.9}$$

With this method, we instead obtain

$$F_{IM_{cap}|EDP}(im \mid y) = \mathrm{P}(IM_{cap} < im \mid EDP = y) \tag{2.10}$$

where IM_{cap} is a random variable representing the distribution of *IM* values that result in an *EDP* level *y* occurring in the structure (i.e., the "Capacity" of the structure to resist a given *EDP* level, defined in terms of the *IM* of the ground motion). The term $F_{IM_{cap}|EDP}(im \mid y)$ is the probability that the *IM* level of a ground motion is less than *im*, given that the ground motion caused a level of response *EDP*=*y*. This method has been used by several others (Kennedy et al. 1984, Bazzurro and Cornell 1994a, b).

Using Equation 2.10, the structural response hazard can be computed as

$$\lambda_{EDP}(y) = \int_{IM} F_{IM_{cap}|EDP}(im \mid y) \left| \frac{\mathrm{d}\lambda_{IM}(im)}{\mathrm{d}im} \right| \mathrm{d}im \tag{2.11}$$

This is an alternative to Equation 2.1 and produces comparable results, as will be seen in the next section for the example structure considered in this chapter.

An example of IDA results and a fitted distribution for *EDP* = 0.01 is shown in Figure 2.5. For ease of comparison, the IDAs in this figure were obtained by interpolating the stripes from Figure 2.3. But with wise selection of analysis points (Vamvatsikos and Cornell 2004) it is not necessary to perform as many dynamic analyses as were needed to generate Figure 2.3. In addition, if one is only interested in a single structural response level y, then Equation 2.10 need only be evaluated at a single *EDP* value. This will further reduce the number of dynamic analyses that need to be performed in order to estimate the mean annual rate of exceeding a given response level[6]. This method also avoids the need to treat separately the records that cause collapse.

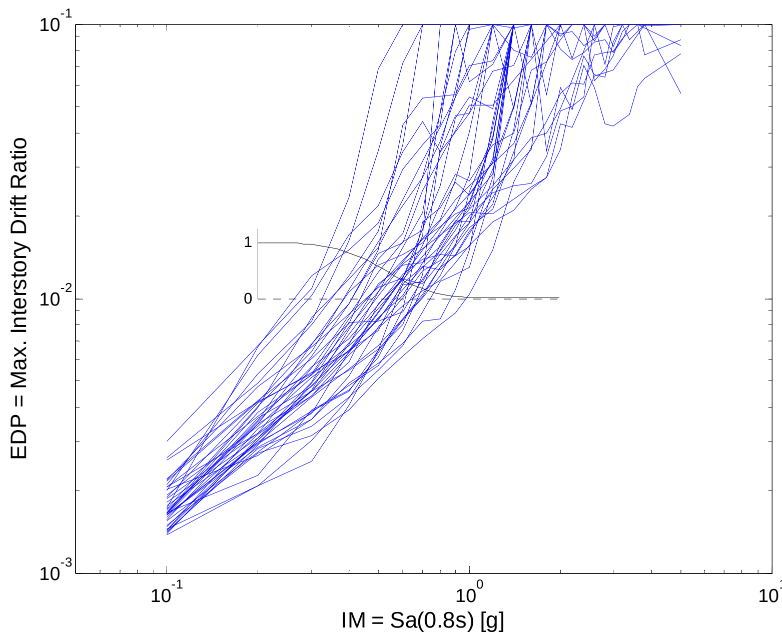


Figure 2.5. Incremental dynamic analysis curves, and an estimated (lognormal) CDF of IM_{cap} given maximum interstory drift ratio = 0.01.

A drawback of this method is that it will likely still require more analyses than a cloud analysis (although it will likely provide more accuracy than the cloud method over a large range of *IM*s). The need to have continuous IDAs also means that one must use the same records for analysis at all *IM* levels, rather than re-selecting different records at increasing *IM* levels in order to reflect the differing causal earthquake events (this will be discussed later in Section 2.6.1).

[6] For example, Ibarra (2003) used this formulation to estimate the mean annual rate of collapse, meaning he only had to perform dynamic analyses until he identified the *IM* level at which each considered record caused collapse, rather than finding the entire IDA curve for each record.

Another issue that arises with this method is that the *IM* value of a record which causes the response *EDP* = *y* is not necessarily unique. This can be observed by drawing a horizontal line through Figure 2.5 and noting that sometimes an IDA trace will cross the line more than one time (i.e., some IDA traces are not monotonically increasing in some ranges of *IM*). The obvious remedy is to define IM_{cap} as the smallest (or largest) value of *IM* such that *EDP* = *y*.

2.4.5 Comparison of results from alternative methods

Using the results shown graphically in the previous sections, resulting estimates can be compared in several ways. The estimated mean and standard deviation of ln*EDP*|*IM* from stripe and cloud methods are shown in Figure 2.6 and Figure 2.7, respectively. With the stripes method, the values at each stripe are interpolated linearly to obtain a continuous estimate. Only the most basic stripe method is used, without piecewise linear fits or probability of collapse estimates (if collapses were of interest in practice the cloud record set would need to be scaled by a greater factor, as none of the records in this example caused a collapse) We see in Figure 2.6 that the simple cloud method can only provide a linear estimate of the mean (in log space), while the stripes method can provide an estimate that varies in any possible way, because the mean is re-estimated at each stripe. Because a large number of dynamic analyses were performed for the stripes in this example, the stripes estimates can be reasonably treated as the “truth.” For spectral acceleration values of approximately 1g, the stripes method reveals an increase in mean *EDP* values as the structure reaches large levels of nonlinearity. These results are for only the non-collapsing cases, where here collapse is assumed to occur after exceedance of a 10% maximum interstory drift ratio. Thus the stripes produce a mean estimate that peaks at about 10% for the extremely high ground motion levels (greater than 2g). The cloud estimate predicts mean values greater than 10%, however, when spectral acceleration is greater than about 2.5g, even though no records produce responses that large.

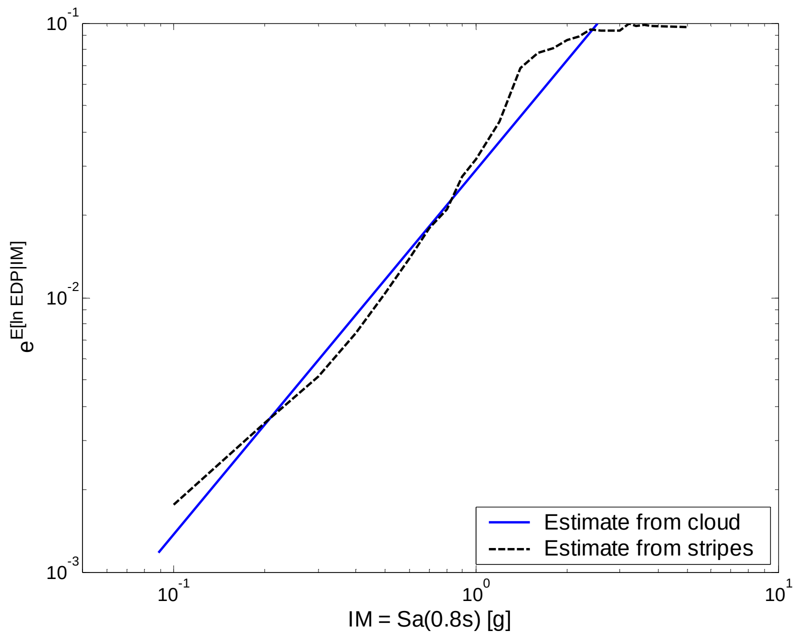


Figure 2.6. Mean values of *EDP*|*IM* for non-collapsing records, estimated using the cloud and stripe methods. For the stripe method, the values between stripes are determined using linear interpolation.

A comparison of the estimated log standard deviations of *EDP* using the cloud and stripe methods is shown in Figure 2.7. The simple cloud method provides a constant estimated standard deviation of ln*EDP* for all spectral acceleration levels, but the stripe estimates indicate that this is not correct. The stripes suggest that log standard deviation is increasing with increasing IM level. It should be noted that the shortcomings of the stripe method identified here can be addressed using the generalizations of the method discussed at the end of Section 2.4.1. However, the number of records needed will begin to increase as the number of model parameters increases. Alternatively, the stripe method can be fit over a more narrow range of *IM* than was used here, where the assumptions are more likely to be reasonable.

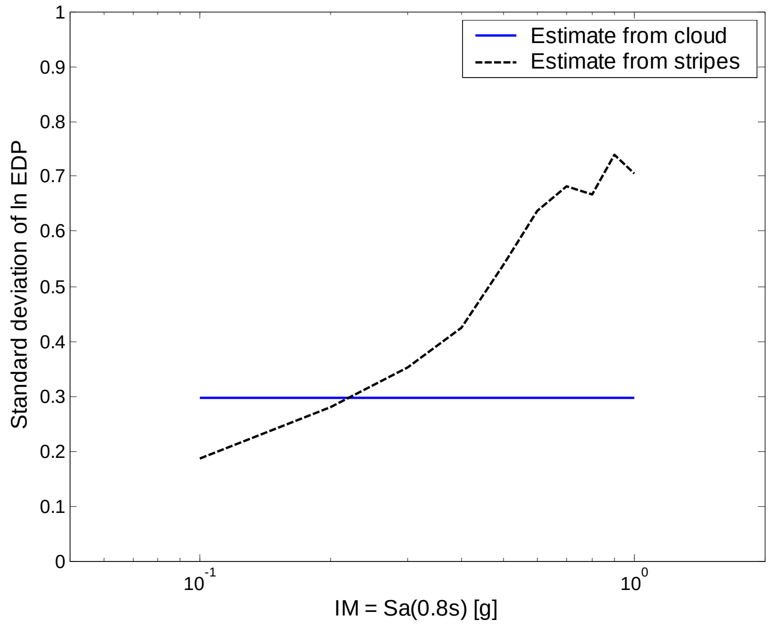


Figure 2.7. Standard deviation of ln*EDP*|*IM* for non-collapse responses, estimated using the cloud and stripe methods. Only IM levels where less than 50% of records cause collapse are displayed because at levels with higher probability of collapse, the statistics estimated using only non-collapse responses are less meaningful.

The means and standard deviations from the stripe data can either be used as the parameters for a normal distribution, or all of the data points from each stripe can be used to create a non-parametric distribution as was done in Figure 2.4.

In order to see the effect of the complete distributional estimates for all spectral acceleration levels, we can combine the probabilistic estimates of *EDP*|*IM* with an *IM* hazard curve to compute an *EDP* hazard curve, per Equation 2.1 (and Equation 2.11 for the capacity formulation). The ground motion hazard is computed for the location in Van Nuys, California, where the example structure is located. The mean spectral acceleration hazard curve is shown in Figure 2.8. Using each of the above methods to estimate *EDP*|*IM*, the drift hazard curve is computed and displayed in Figure 2.9.

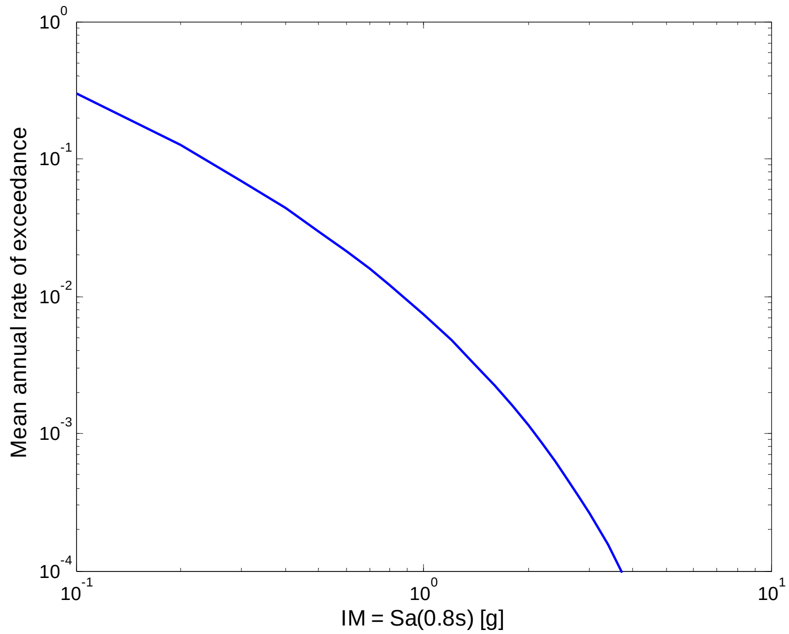


Figure 2.8. Ground motion hazard at the Van Nuys, California site for spectral acceleration at 0.8 seconds.

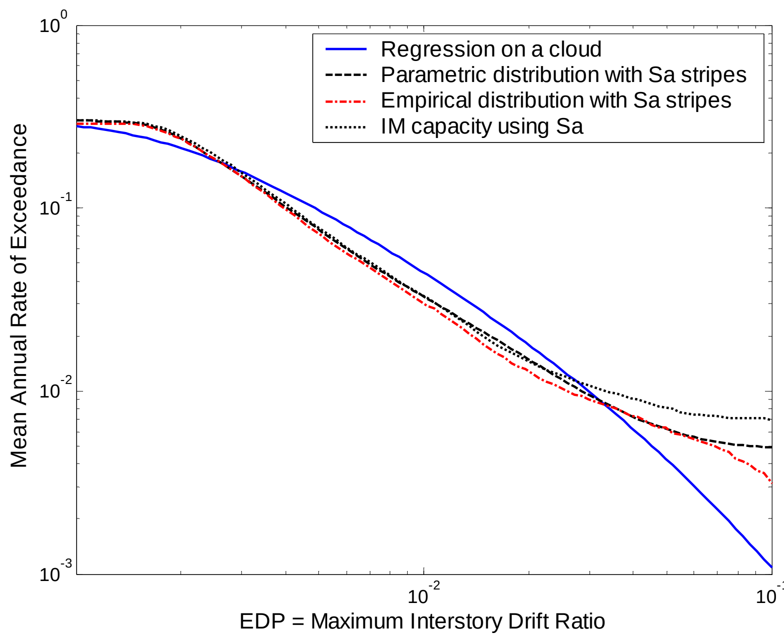


Figure 2.9. Comparison of drift hazard results for the example structure using the four estimation methods described above.

Because of the large number of analyses used, the two stripe methods and the capacity method all capture the distribution of $EDP|IM$ very well, and the resulting drift hazard curves show reasonable agreement. Thus it appears that the stripe methods and the capacity method will produce approximately equivalent results, given sufficient data for estimation, lognormality of the conditional distributions, and monotonicity of the IDA traces. In statistical inference, the property of convergence for large sample sizes is termed *consistency*—this is the property that we are considering in an informal way here. Connections and comparisons between these methods have also been made by others (Shome and Cornell 1999, Jalayer 2003, Vamvatsikos and Cornell 2004).

For this example the cloud method agrees somewhat with the other methods at *EDP* values less than 0.04, but the functional form constraints of this method limit its accuracy at large *EDP* levels. The stripe method was not used with records causing large responses, and probability-of-collapse generalizations were not considered. With a greater number of analyses and generalization to remove the basic assumptions, the stripe method could be made to agree more closely with the other methods. At any rate, the stripe method agrees reasonably well with the other methods over the range where the cloud estimate was fit.

In practice it is possible to reduce the number of stripe levels and the number of records and still retain most of the accuracy achieved using this large number of analyses. But it is unlikely that the number of dynamic analyses needed for stripe analysis could be reduced to less than the number used for cloud analysis (the number of records necessary for the cloud method can also be reduced from the 60 used here). Thus the cloud analysis method remains an important option because of its reduced computational expense. For checking response at a single *IM* level, however, a single stripe may require fewer analyses than a cloud. Thus, the cloud, stripe and capacity formulations each have unique characteristics that may be advantageous in certain cases.

2.5 Analysis using a vector-valued intensity measure

Throughout this dissertation, intensity measures consisting of multiple parameters (i.e., vector-valued intensity measures) are considered and shown to have desirable attributes. As will be shown, the additional parameters provide more information about the earthquake ground motion and thus allow for more accurate prediction of structural response. This enables estimation of conditional *EDP* distributions with smaller standard deviations (increased efficiency), and reduces the possibility that the estimates are biased due to ground motion properties not included in the *IM* (increased sufficiency).

To incorporate these improved intensity measures, the methods of the previous section can all be generalized to incorporate prediction based on multiple parameters. Because the intensity measure is a vector in this section, a new notational convention is used. The vector of intensity measure parameters will be denoted **IM**, and the individual parameters will be denoted IM_1, IM_2, etc. Several of the vector-based methods treat one of the **IM** parameters, IM_1, as the dominant predictor of structural response and treat IM_2, IM_3, etc. as subordinate predictors. This allows us to treat IM_1 differently than we treat the other parameters (e.g., by scaling to match IM_1 and using regression to account for the others). Using this vector **IM**, the integral of Equation 2.1 becomes

$$\lambda_{EDP}(y) = \int_{\mathbf{IM}} G_{EDP|\mathbf{IM}}(y \mid \mathbf{im}) \left| \frac{\mathrm{d}\lambda_{IM}(\mathbf{im})}{\mathrm{d}\mathbf{im}} \right| \mathrm{d}\mathbf{im} \tag{2.12}$$

In many instances, the vector **IM** will consist of only two parameters. Further, the joint hazard $\lambda_{IM}(\mathbf{im})$ will be expressed as a typical scalar hazard in terms of IM_1, along with a conditional distribution of IM_2 given IM_1. In this case, Equation 2.12 can be expressed as

$$\lambda_{EDP}(y) = \int_{IM} G_{EDP|\mathbf{IM}}(y \mid im_1, im_2) f(im_2 \mid im_1) \left| \frac{\mathrm{d}\lambda_{IM}(im_1)}{\mathrm{d}im_1} \right| \mathrm{d}im_1 \tag{2.13}$$

where $\lambda_{IM}(im_1)$ is the ground motion hazard for parameter IM_1, $f(im_2 \mid im_1)$ is the conditional distribution of IM_2 given $IM_1 = im_1$ (obtained from vector-valued hazard analysis) and $G_{EDP|\mathbf{IM}}(y \mid im_1, im_2)$ is the CCDF of *EDP* given the vector of **IM** parameters. The goal of this section is to evaluate methods for obtaining $G_{EDP|\mathbf{IM}}(y \mid im_1, im_2)$.

The same example structure will continue to be used in this section. The same dynamic analysis results will be used as well; the only change is the use of additional ground motion parameters to explain the variability in structural responses. For illustration, a vector **IM** consisting of Sa(0.8s) and a parameter denoted ε will be used for illustration. The parameter ε has been found to be an effective predictor of structural response (Chapter 3). This is because it is an implicit measure of spectral shape; more specifically, it tends to indicate whether $Sa(T_1)$ is in a peak or a valley of the response spectrum. Records with positive ε values tend to have smaller responses than records with negative ε values, given that they have the same $Sa(T_1)$ values. This ground motion parameter will be considered in more detail in Chapter 3 and Chapter 6, but for now it will serve as a useful parameter to demonstrate structural response prediction as a function of a vector **IM**.

2.5.1 Use multiple linear regression on a cloud of ground motions

The cloud method of Section 2.4.1 can be easily adapted to use vector-valued **IM**s. Regression analysis using multiple predictor variables (*IM* parameters) is called *multiple linear regression* and there exists a well-developed theory regarding model selection, confidence intervals for regression coefficients, etc. (Neter et al. 1996). With this method, a linear functional form is assumed (after variable transformations) that specifies the relationship between the mean value of *EDP* and *IM*. For example

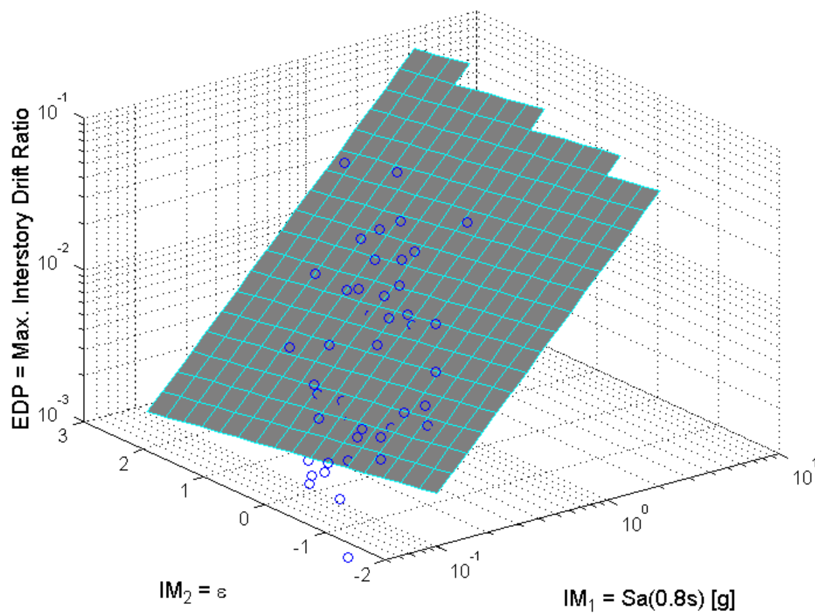
$$\ln EDP = a + b\ln IM_1 + c\ln IM_2 + e \tag{2.14}$$

where a, b and c are coefficients to be estimated from the multiple linear regression, and e is a zero-mean random variable representing the remaining variability in ln*EDP* given **IM**. An example of an estimated mean prediction is shown in Figure 2.10, using the example **IM** parameters ln*Sa*(0.8s) and ε. Assuming that ln*EDP* is normally distributed, then the probability of exceeding an *EDP* level y given **IM** = (im_1, im_2) can be estimated as

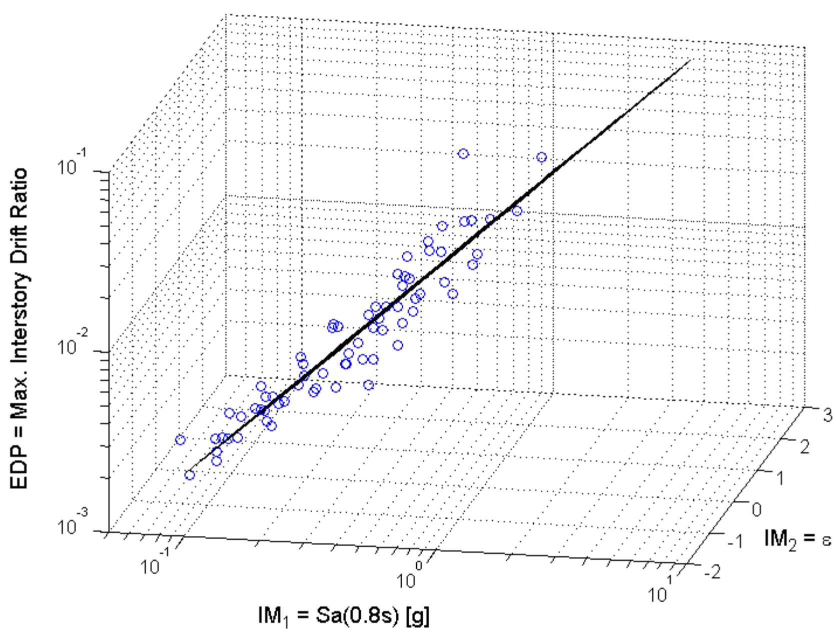
$$G_{EDP|\mathbf{IM}}(y\,|\,\mathbf{im}) = 1 - \Phi\left(\frac{\ln y - (\hat{a} + \hat{b}\ln im_1 + \hat{c}\ln im_2)}{\sqrt{V\hat{a}r[e]}}\right) \tag{2.15}$$

where the mean value of *EDP* given (im_1, im_2) is estimated as $\hat{a} + \hat{b}\ln im_1 + \hat{c}\ln im_2$, and $V\hat{a}r[e]$ is estimated as the sample variance of the residuals.

This method has the advantage of easily accommodating many **IM** parameters by simply adding additional terms to Equation 2.14 (e.g., $d\ln IM_3$). It also has the great advantage of requiring many fewer parameters to be estimated than with any of the other methods, typically allowing estimates to be obtained from fewer analyses—a significant benefit when analyses are computationally expensive. The model can be generalized, as in Section 2.4.1, to incorporate features such as non-constant variance or records that cause collapse.



(a)



(b)

Figure 2.10. (a) The mean fit from a cloud of data, predicted using multiple linear regression with *Sa*(0.8s) and ε as predictors. (b) The same mean from a different viewpoint, to emphasize that the prediction is a plane.

Because of these appealing features, this was the analysis method of choice for many previous investigations of vector-valued intensity measures (e.g., Bazzurro 1998, Shome and Cornell 1999, Luco et al. 2005).

One difficulty with this method is that when two parameters IM_1 and IM_2 are highly correlated, it is difficult to separate their effects and estimate with confidence the coefficients b and c in Equation 2.14. This condition is termed *collinearity*, and is a particular problem when using vector intensity measures consisting of two spectral acceleration values, or when using the vector consisting of spectral acceleration and ε described briefly above and discussed in more detail in Chapter 3.

An under-appreciated difficulty with multiple regression is that extrapolation is much more common in higher dimensions. Although marginally one might have predictor variables that span the **IM** range of interest, their joint distribution may not cover the range of data expected. This can be seen easily in Figure 2.10a: although there are data points with $2<\varepsilon<3$, and there are data points with $0.1\text{g}<Sa<0.3\text{g}$, there are no data points with $2<\varepsilon<3$ *and* $0.1\text{g}<Sa<0.3\text{g}$ (i.e., the left side of the fitted plane). So this region is actually an extrapolation, although it might not be noticeable at first inspection. As the number of predictors grows, this problem quickly becomes very significant. Bellman (1961) named this phenomenon the "curse of dimensionality."

A final problem is that it is difficult to model interactions between intensity measure parameters. For example, if IM_1 is spectral acceleration at the first-mode period of the structure and IM_2 is spectral acceleration at a lower frequency (which captures nonlinear response), then IM_2 may have a small effect on response at low IM_1 levels when the structure is linear and a large effect on response at large IM_1 levels when the structure is very nonlinear. This would imply a small value for the c coefficient in Equation 2.14 for low IM_1 and a large value of c for high IM_1 (this changing functional form will be seen in Figure 2.15 below). In the formulation of Equation 2.14, however, the value of c is constrained to be constant for all levels of IM_1. One could attempt to fix this by adding an *interaction* term to the equation (e.g., $d \ln IM_1 IM_2$), but in practice it can be difficult to determine and incorporate the correct functional form without a large amount of data.

This method has several appealing features, but the assumptions made regarding the functional form should be verified to make sure they are appropriate for a given analysis.

2.5.2 Scale records to a specified level of the primary IM parameter, and regress on additional parameters

With this method, records are scaled to the primary IM parameter (e.g., $Sa(T_1)$) as in Section 2.4.2, and then regression analysis is used on the stripe of data to determine the effect of the remaining IM parameters. Logistic regression is used to compute the probability of collapse, and linear regression is used to model the non-collapsing responses. The procedure is explained here for a two-parameter *IM*, but it is straightforward to add additional parameters by using multiple regression. This method is used extensively in Chapter 3; a more detailed description of the procedure is given there.

Rather than estimating the probability of collapse as simply the fraction of records at an IM_1 stripe that cause collapse, IM_2 is used to predict the probability of collapse using logistic regression (Neter et al. 1996). With this procedure, each record has a value of IM_2 which is used as the predictor variable. Using the indicator variable *C* to designate occurrence of collapse (*C* equals 1 if the record causes collapse and 0 otherwise), the following functional form is fitted

$$\hat{P}\left(C \mid IM_1 = im_1, IM_2 = im_2\right) = \frac{e^{a+bim_2}}{1+e^{a+bim_2}} \tag{2.16}$$

where *a* and *b* are coefficients to be estimated from regression on a record set that has been scaled to $IM_1=im_1$ (because of this, *a* and *b* are implicitly dependent on IM_1). An example of this data and a fitted logistic regression curve is shown in Figure 2.11. By repeating this regression for multiple IM_1 levels, one can obtain the probability of collapse as a function of both IM_1 and IM_2. This is illustrated in Figure 2.12.

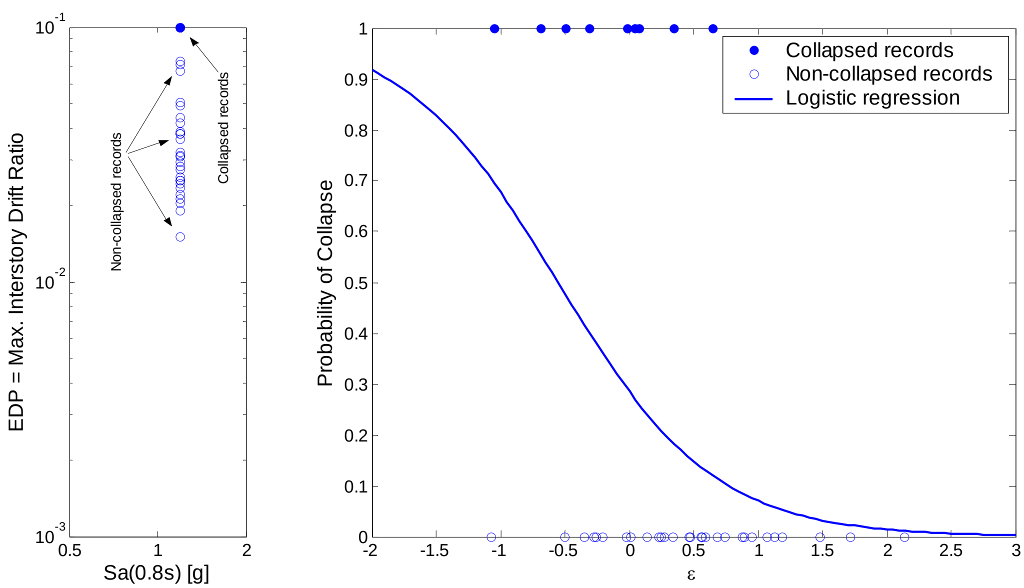


Figure 2.11. Prediction of the probability of collapse using logistic regression applied to binary collapse/non-collapse results at Sa(0.8s) = 1.2g.

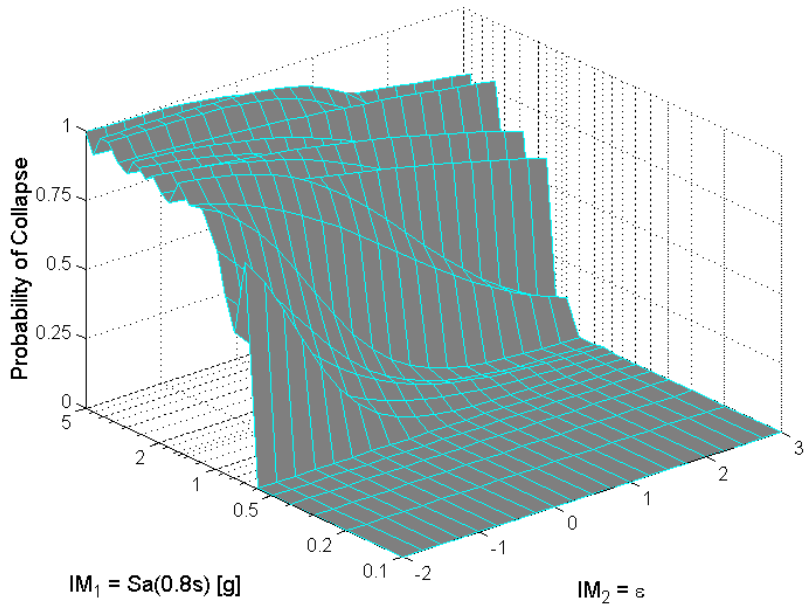


Figure 2.12. Probability of collapse fit from stripes of data, predicted by scaling to Sa(0.8s) and using repeated logistic regressions with ε as the predictor.

To incorporate IM_2 in the prediction of non-collapse response, we use linear regression. At an IM_1 stripe, the relationship between IM_2 and EDP is modeled by a linear function (after transformations). For example

$$\ln EDP = c + d \ln IM_2 + e \tag{2.17}$$

where c and d are constant coefficients, and e is a zero-mean random variable modeling the residuals as before. In contrast to the cloud method, these coefficients and the residual variance are re-estimated at every IM_1 stripe. So assuming a Gaussian distribution for the residuals, the probability that EDP exceeds y, given $IM_1 = im_1$, $IM_2 = im_2$, and no collapse (NC) can be expressed as

$$G_{EDP|\mathbf{IM}}\left(y \mid im_1, im_2, NC\right) = 1 - \Phi\left(\frac{\ln y - \left(\hat{c} + \hat{d}\ln im_2\right)}{\sqrt{V\hat{a}r[e]}}\right) \tag{2.18}$$

Where $\hat{c}$, $\hat{d}$ and $\sqrt{V\hat{a}r[e]}$ are estimated using response data associated with records scaled to IM_1=im_1. A plot of the data, the regression fit and the estimated distribution is shown in Figure 2.13.

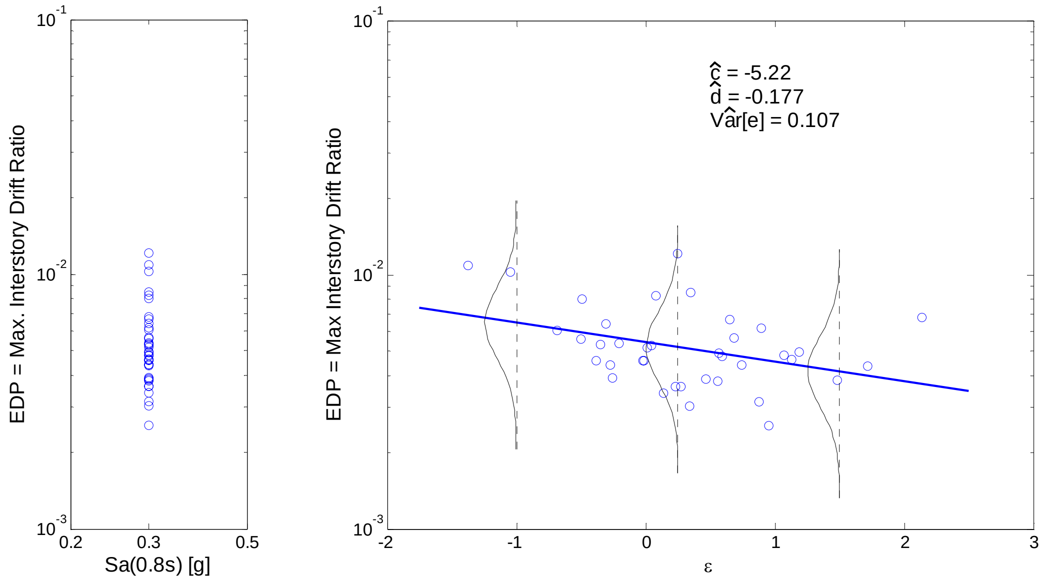


Figure 2.13. Prediction of response given no collapse, at *Sa*(0.8s) = 0.3g, with the distribution of the residuals superimposed over the data.

By repeating this scaling and regression for multiple IM_1 levels, one can obtain distribution of non-collapse responses as a function of both IM_1 and IM_2. The mean value of non-collapse responses as a function of *Sa*(0.8s) and ε shown in Figure 2.14. In this case, the fit is close to planar, and does not look significantly different than Figure 2.10 except at high *Sa*(0.8s) levels. In Figure 2.15 a vector **IM** is shown where there is an interaction effect between IM_1 and IM_2. A cloud regression would have difficulty capturing this behavior.

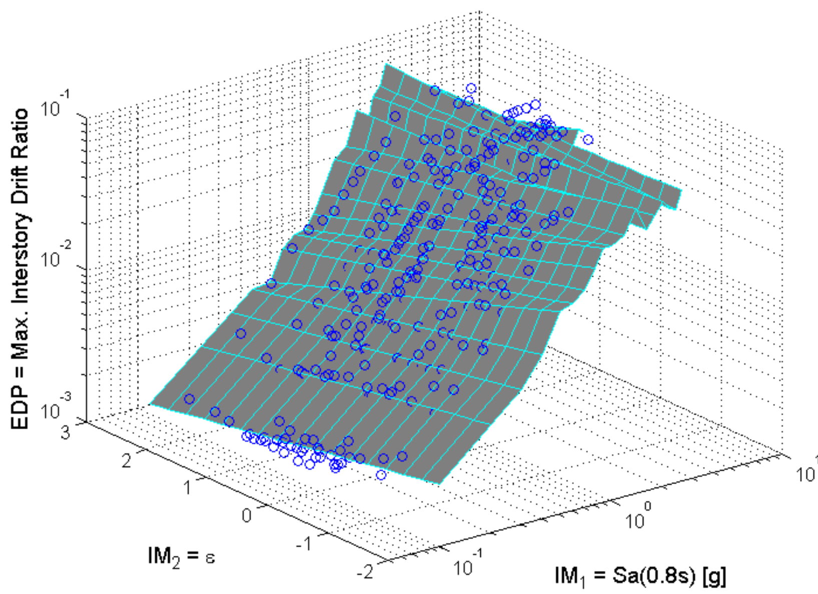


Figure 2.14. Conditional mean fit of non-collapse responses from stripes of data, predicted by scaling records to *Sa*(0.8s) and using repeated regressions with ε as the predictor.

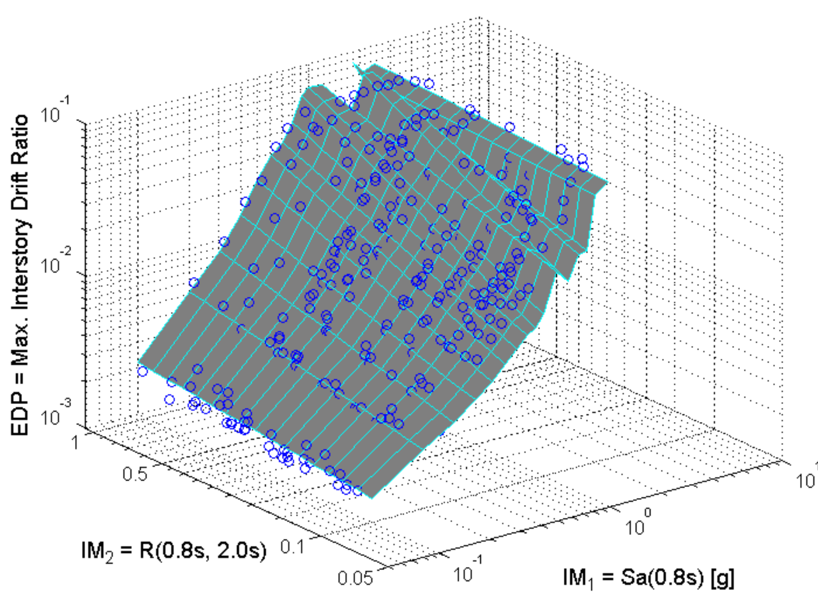


Figure 2.15. Conditional mean fit of non-collapse responses from stripes of data, predicted by scaling records to Sa(0.8s) and using repeated regressions with R(0.8s, 2.0s) as the predictor. In this case, the predicted mean is not planar, so it would be difficult to capture the true behavior using a cloud regression.

The collapse and non-collapse cases are combined using the Total Probability Theorem and Equations 2.14 and 2.16, to compute the conditional probability that *EDP* exceeds *y*

$$G_{EDP|\mathbf{IM}}\left(y\,|\,im_1,im_2\right)=\hat{P}(C)+\left(1-\hat{P}(C)\right)\left(1-\Phi\left(\frac{\ln y-\left(\hat{c}+\hat{d}im_2\right)}{\sqrt{V\hat{a}r[e]}}\right)\right)$$
$$\text{where } \hat{P}\left(C\right)=\frac{\mathrm{e}^{\hat{a}+\hat{b}im_2}}{1+\mathrm{e}^{\hat{a}+\hat{b}im_2}} \tag{2.19}$$

where $\hat{a}$, $\hat{b}$, $\hat{c}$, $\hat{d}$ and $Var[e]$ are all estimated using records scaled to $IM_1 = im_1$.

This method has several desirable attributes. The method of scaling to stripes is already used frequently for scalar *IM*s (Section 2.4.2), so the record processing is not new. We merely add a secondary regression analysis to the procedure. In fact, the scalar parametric stripe case is merely a special case of this method, with the *b* and *d* coefficients omitted from Equation 2.17. In regression terms, only two additional degrees of freedom per stripe are used for estimation by moving to a vector, and this “loss” of degrees of freedom should be offset by the increased efficiency (i.e., decreased $\sqrt{V\hat{a}r[e]}$) obtained from the vector. In addition, because the regression coefficients are re-estimated at each stripe, the interaction between IM_1 and IM_2 is captured automatically—a feature that cannot be achieved using the cloud method of Section 2.5.1. Because of the advantages discussed here, this is the primary procedure adopted throughout this dissertation.

2.5.3 Fit a conditional distribution for IM$_1$ capacity to IDA results

Here, the *IM* capacity method of Section 2.4.4 is generalized to incorporate a vector **IM**. Although only a single **IM** parameter (IM_1) is scaled during the IDA, it is still possible to determine the effects of other **IM** parameters as follows. First, perform Incremental Dynamic Analysis by varying IM_1 until the *EDP* value of interest is obtained. This will provide the distribution of IM_{1cap} values as in Section 2.4.4. Then display these IDAs along with IM_2, as shown in Figure 2.16.

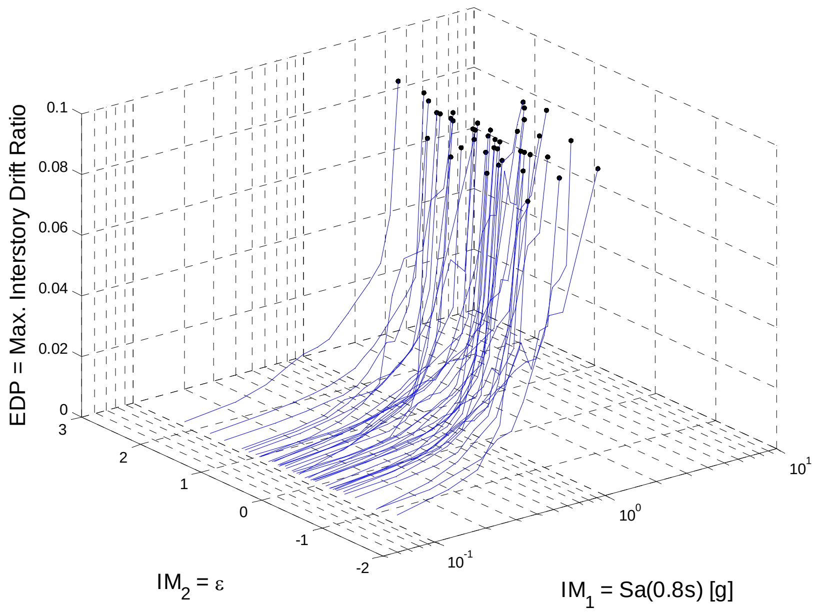


Figure 2.16. Incremental Dynamic Analysis with two intensity parameters, used to determine capacity in terms of the intensity measure.

The point at which each IDA trace first reaches the *EDP* level of interest (max interstory drift ratio = 0.1 for this example) defines a set of **IM** capacity values. These points are plotted in Figure 2.17 for exceedance of *EDP* = 0.1 maximum interstory drift ratio (these are the same points as the tips of the IDA traces in Figure 2.16).

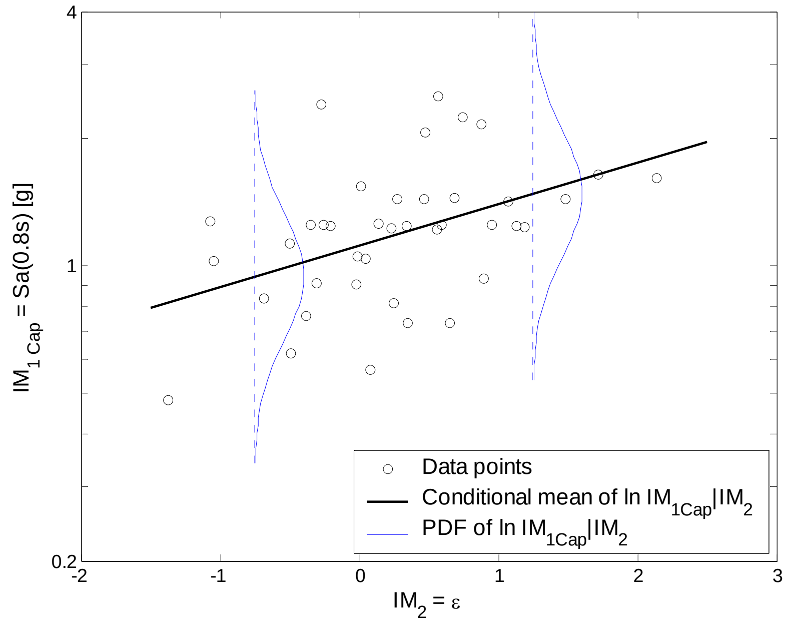


Figure 2.17. *Sa*(0.8s), ε pairs corresponding to occurrence of 0.1 maximum interstory drift ratio.

It is apparent in Figure 2.17 that IM_2 can explain part of the variation of IM_1 capacity (that is, the IM_{1Cap} values tend to be larger for positive values of ε). So the probability of exceeding the **IM** capacity associated with the target *EDP* level can be expressed in terms of a conditional distribution of IM_{1Cap} given IM_2. Here, the conditional distribution of $\ln IM_{1Cap}$ appears to be linearly dependent upon IM_2. Therefore, linear regression can be used to find the conditional mean of $\ln IM_{1Cap}$ given IM_2

$$\mu_{\ln IM_{1Cap}|IM_2=im_2} = \beta_0 + \beta_1 IM_2 \tag{2.20}$$

where β_0 and β_1 are coefficients to be estimated from linear regression using, e.g., the data points from Figure 2.17. Further, the conditional standard deviation of $\ln IM_{1Cap}$ given IM_2 can be estimated by computing the standard deviation of the regression residuals, as was done using the other regression-based methods discussed earlier. This conditional standard deviation is denoted $\sigma_{\ln IM_{1Cap}|IM_2=im_2}$. If the conditional distribution of $\ln IM_{1Cap}$ given IM_2 is assumed to be Gaussian, then the conditional mean and standard deviation computed above completely define the conditional distribution of $\ln IM_{1Cap}$ associated with reaching a given *EDP* level y. The probability density function (PDF) of this distribution is shown in Figure 2.17 for two different values of IM_2. The CDF of this conditional distribution can be computed as

$$P\left(IM_{1Cap} < im_1 \mid EDP = y, IM_2 = im_2\right) = \Phi\left(\frac{\ln im_1 - \mu_{\ln IM_{1Cap}|IM_2 = im_2}}{\sigma_{\ln IM_{1Cap}|IM_2 = im_2}}\right) \tag{2.21}$$

where $\Phi(\cdot)$ is the CDF of the standard Gaussian distribution. Using Equation 2.21, the drift hazard can then be computed as

$$\lambda_{EDP}(y) = \int_{im_1}\int_{im_2} P\left(IM_{1Cap} < im_1 \mid EDP = y, IM_2 = im_2\right) f_{IM_2|IM_1}\left(im_2 \mid im_1\right)\left|\frac{\mathrm{d}\lambda_{IM}(im_1)}{\mathrm{d}im_1}\right| \mathrm{d}im_1 \tag{2.22}$$

where the IM_{1cap} points were defined as the IM_1 level associated with occurrence of the *EDP* level *y*. This procedure can be extended to larger vectors by using additional *IM* parameters to the conditional distribution of IM_{1cap}.

As an example, $P\left(IM_{1Cap} < im_1 \mid IM_2 = im_2\right)$ is displayed in Figure 2.18 for a range of im_1 and im_2, where IM_{1Cap} is defined as the *Sa*(0.8s) level that causes first exceedance of 0.1 maximum interstory drift ratio (which is considered to indicate collapse for this example), given ε, which is defined to be IM_2. This plot can be compared to Figure 2.12, which was obtained using the stripes method (a comparison between the two will be made in Section 2.5.6).

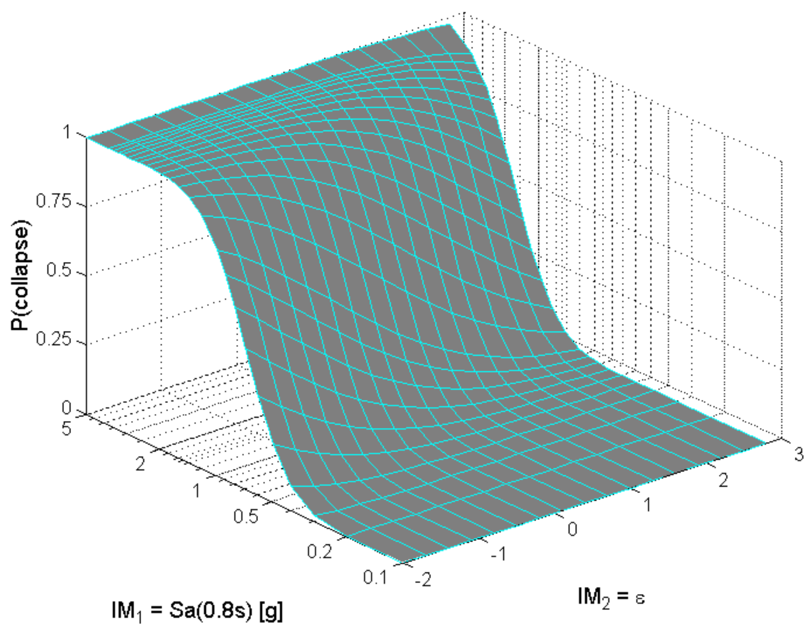


Figure 2.18. Probability of collapse predicted by fitting a joint normal distribution to collapse capacity as a function of ln*Sa*(0.8s) and ε.

The advantage of this method relative to the stripes is that the IDAs potentially require fewer analyses if the data points are chosen wisely. Unlike the cloud method, interactions among the **IM** parameters are captured because, the conditional capacity distribution (as a function im_2) is re-estimated at each *EDP* level. The disadvantage is that wise selection of IDA analysis points and interpolation to determine capacity points is slightly more difficult than simply scaling to target intensity measure values to create stripes.

In this example, assumptions of linear dependence between transformed **IM** parameters, constant standard deviations, and Gaussian distributions of IM_{1Cap} were made. The reasonableness of these assumptions should be verified before proceeding with this method. As with the previous methods, it is possible to relax these assumptions if appropriate. Relaxation of these assumptions, however, may require additional analysis effort as well as an increased number of dynamic analyses for estimation.

2.5.4 Scale records to specified IM levels and calculate an empirical distribution

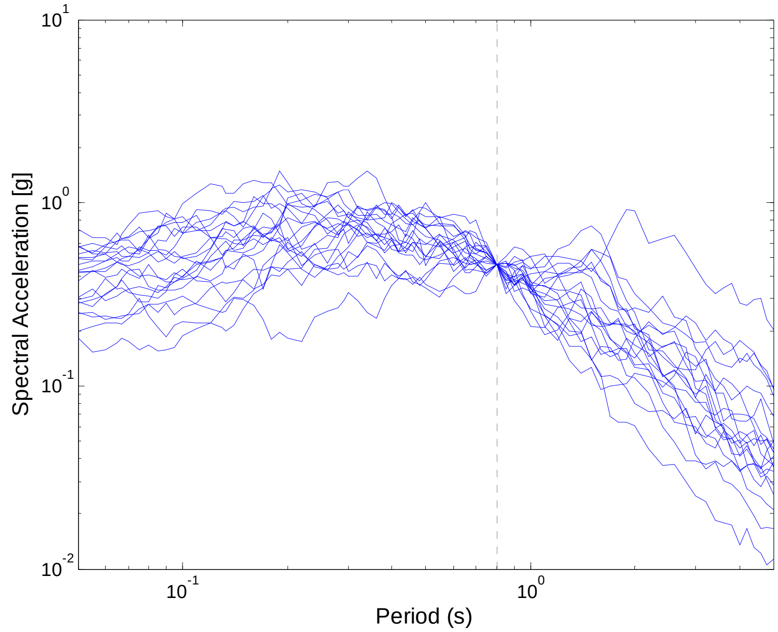
With this method, records are scaled to target IM_1 levels and then bins of IM_2 values are selected. Within each bin, an empirical CCDF is computed as in Section 2.4.3. This is then used as the $G_{EDP|\mathbf{IM}}(y \mid im_1, im_2)$ in drift hazard computations. It can be shown that this method is equivalent to fitting a scalar-*IM* empirical CCDF after using the re-weighting procedure that will be discussed in Section 2.6.2. It is conceptually simpler to present this method in the context of re-weighting, so discussion will be postponed until that later section.

2.5.5 Process records to match target values of all IM parameters

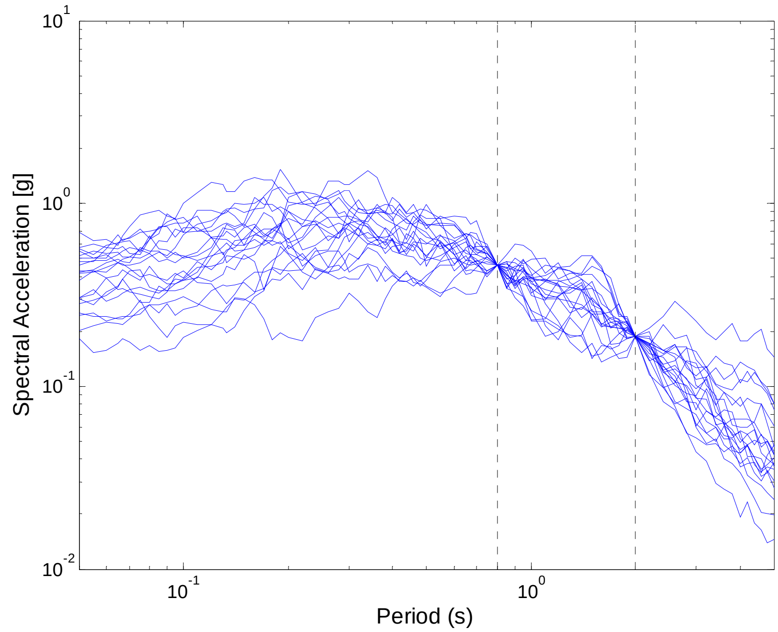
Another potential method, which has not been attempted for vector *IM*s to the author's knowledge, is the processing of a suite of records to match all parameters specified in the **IM** vector, *leaving the other record properties random*. The conditional distribution of *EDP* could then be estimated directly from the modified records. This is done for scalar *IM*s by simply scaling records. With vector **IM**s it will require a more complicated scheme. For example, when the **IM** consists of spectral acceleration plus a measure of spectral shape (e.g., Chapter 4), one would scale the record to match spectral acceleration, and then use a spectrum-compatibilization scheme (Abrahamson 1993, Silva 1987) to match the spectral shape. The spectrum should not be completely smoothed, as is sometimes done to match a Uniform Hazard Spectrum. Rather, the spectrum roughness should be retained and only the general shape of the spectrum should be modified, in accordance with the target spectral shape parameter.

This type of spectrum modification scheme is sometimes referred to as a "one-pass" compatibilization (e.g., McGuire et al. 2001), because it requires only a single iteration of spectrum modification as opposed to the multiple iterations required to create a smooth spectrum. This scheme was used by Carballo and Cornell (1998) to create records representative of Eastern U.S. earthquakes from Western U.S. records. To date, however, it has not been used to estimate *EDP* using a vector **IM** as proposed here. An illustration of this procedure is shown below. In

Figure 2.19a, spectra are shown for records scaled to *Sa*(0.8s). In Figure 2.19b, records are shown scaled to *Sa*(0.8s) and compatibilized to match a target spectral acceleration at 2.0s as well. These figures were created using simulated spectra for illustration—the actual processing and analysis of records using this scheme is a topic of current research.



(a)



(b)

Figure 2.19. (a) Response spectra of records scaled to match spectral acceleration at a 0.8s. (b) Response spectra of records scaled to match spectral acceleration at 0.8s and processed to match spectral acceleration at 2.0s.

If the same suite of records is used for a range of **IM** values by repeatedly re-scaling and re-compatibilizing the records, then this could be thought of as a generalization of Incremental Dynamic Analysis, where there are now two or more scaling parameters rather than the single parameter used in standard IDA. This processing scheme should be possible when the intensity measure consists of spectral acceleration plus a measure of spectral shape (as is the case for the intensity measures in this dissertation). It is not so obvious, however, how one would process a record to match other parameters such as duration. This method would likely also require a great number of dynamic analyses, which may make it impractical in cases where computational expense is a concern.

Rather than processing records to have specified values of each **IM** parameter, one could approximately reproduce this approach by selecting real records whose **IM** values closely match the target values. Or one could still use simple record scaling to match the first spectral acceleration parameter, and then select only those records whose other **IM** values closely match the target values. This approach will be used in Chapter 6, and is closely related to the method proposed in Section 2.6.1.

2.5.6 Comparison of results from alternative methods

Using an intensity measure consisting of *Sa*(0.8s) and ε, several comparisons can be made among the results from the above methods. Mean values of *EDP* given *Sa*(0.8s) and ε were shown for the cloud and stripe methods in Figure 2.10 and Figure 2.14, respectively, and seen to match reasonably well at least for values of *Sa*(0.8s) less than 1 or 2g. Probability of collapse estimates were shown for the stripe and capacity methods in Figure 2.12 and Figure 2.18, respectively. They were also seen to match reasonably well, although the estimate from stripes is not "smooth[7]." The *EDP* predictions can also be combined with ground motion hazard using Equation 2.13 to predict the mean annual rate of exceeding various *EDP* levels. A comparison of this result for each of the methods is shown in Figure 2.20. With the exception of the cloud method, the results are in good agreement. Note also that the curves have shifted down with respect to the curve based on the scalar *IM* *Sa*(0.8s) alone. This tells us that the vector intensity measure has provided more information about structural response. This conclusion will be discussed further in Chapter 3.

[7] One would expect these probabilities of collapse to be monotonic functions of the *IM* parameters, as is seen in Figure 2.18, but not in Figure 2.12. Although this might suggest that the capacity method is preferable, its prediction of probability of exceeding *EDP* levels is not necessarily continuous as a function of the EDP level. In contrast the stripe method prediction is monotonic as a function of *EDP* level, but not as a function of IM_2. So the two methods in fact provide comparable prediction "accuracy" over the **IM**-*EDP* space, even though this is not apparent from Figure 2.12 and Figure 2.18.

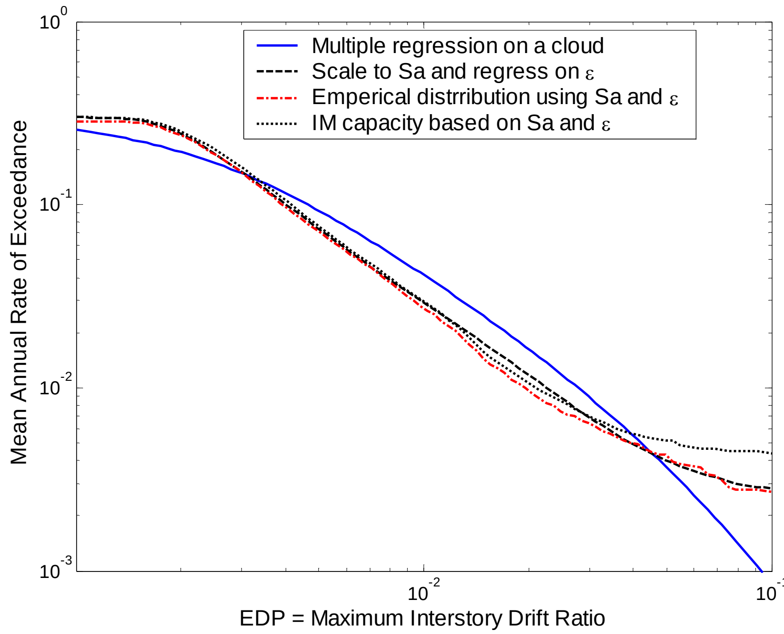


Figure 2.20. Comparison of drift hazard results using a vector-valued intensity measures consisting of $Sa(T_1)$ and ε.

2.6 "Hybrid" methods for capturing the effect of a vector of intensity measures

Although vector-valued intensity measures facilitate improved estimation of *EDP* distributions, they can also add complexity to the analysis. In this section, several methods are described that can achieve the gains of a vector-valued *IM* while simplifying the analysis procedure. The common theme among these methods is that they try to mimic the conditional distribution of $IM_2|IM_1$ that is expected at the site in the ground motion dataset used for analysis. Then only the scalar parameter IM_1 need be used for estimating *EDP*, and the underlying distribution of $IM_2|IM_1$ is consistent between the dataset and the true ground motion hazard. That is, if the distribution of $IM_2|IM_1$ is consistent between the dataset and the true ground motion hazard, then the following is true

$$\int_{im_1}\int_{im_2} G_{EDP|IM_1,IM_2}(y\,|\,im_1,im_2)\,f_{IM_2|IM_1}(im_2\,|\,im_1)\left|\frac{d\lambda_{IM}(im_1)}{dim_1}\right|dim_1 = \int_{im_1} G_{EDP|IM_1}(y\,|\,im_1)\left|\frac{d\lambda_{IM}(im_1)}{dim_1}\right|dim_1 \tag{2.23}$$

2.6.1 Scale records to a specified level of IM$_1$, selecting records to match the desired distribution of secondary parameters

With this method, we carefully select records to match the target $IM_2|IM_1$ distribution *at each* IM_1 *level*. This is often done today, with IM_1 as spectral acceleration and IM_2 as magnitude and/or distance (Stewart et al. 2001). The target conditional distribution is obtained from disaggregation of probabilistic seismic hazard analysis results, and records with the target magnitude and distance values are identified and used for analysis. This same method has also recently been used for the more effective intensity measures (e.g., ε) discussed in this dissertation (Chapter 6 and Haselton et al. 2005).

The difficulty with this method is that the conditional distribution of $IM_2|IM_1$ typically changes as the value of IM_1 changes (e.g., as the spectral acceleration level increases large positive ε values become more common). Thus, records need to be re-selected at differing IM_1 levels. This adds an extra step in the record selection process, and makes IDA methods more complicated because the records change as the *IM* level changes. The limited number of available ground motion records may also provide a practical impediment to the adoption of this procedure. Nonetheless, we will see in Chapter 6 that the advantages of this method may justify the extra work.

2.6.2 Fit a distribution to a stripe of data, after re-weighting to match a target distribution of IM$_2$|IM$_1$

In an attempt to avoid reselecting records at each stripe, one could select a single set of records with a range of IM_2 values and then re-weight the data after scaling to IM_1 so that the re-weighted dataset has the proper distribution of $IM_2|IM_1$ at each IM_1 level. This method has been proposed by Shome and Cornell (1999, p208) and Jalayer (2003, Chapter 6) After discretizing IM_2 into a set of "bins," the weight for a record with an IM_2 value in bin j would be

$$weight_j = \frac{f_{IM_2|IM_1}(im_{2,j} \mid im_1)}{n_j} \tag{2.24}$$

where $f_{IM_2|IM_1}(im_{2,j} \mid im_1)$ is the target probability that $IM_2 = im_2$ (in bin j), given $IM_1 = im_i$ (obtained from vector-valued PSHA, or disaggregation if $IM_2 = \varepsilon$), and n_j is the number of records in the record-set with IM_2 in bin j. The weighted record-set will have a probability distribution equal to the target $f_{IM_2|IM_1}(im_2 \mid im_1)$. The scalar stripe methods of Sections 2.4.2 and 2.4.3 can then be applied to this weighted dataset, and the procedure works as before.

There are a few disadvantages to consider when using this method. It is necessary to have records in every IM_2 bin considered, or else the denominator of Equation 2.22 will be equal to zero for some IM_2 bins. Ensuring that there are records in each IM_2 bin will either require careful record selection or course discretization, which may mask the effect of the underlying intensity measure parameter. In addition, some records may be assigned weights close to zero, meaning that some data is essentially discarded for the purposes of estimation. Given the large expense of obtaining response data, this is undesirable. This problem is of some concern when incorporating a second parameter, but becomes nearly insurmountable if the vector consists of more than two parameters. This is because the number of bins increases exponentially with the number of *IM* parameters, and thus the average number of records in each bin decreases very quickly to zero even with a large record set. This so-called *curse of dimensionality* is a concern for all of the methods considered, but is especially problematic for this method because of its nonparametric nature (Bellman 1961, Hastie et al. 2001, p22). Whereas parametric methods require a few model parameters to be estimated for each additional *IM* parameter, the number of parameters (i.e., weights) estimated with this method is equal to the (exponentially increasing) number of bins. Shome and Cornell (1999) and Jalayer (2003) use a slightly different weighting scheme than the one in Equation 2.24, but it too will suffer from the curse of dimensionality. The large number of estimated parameters reduces the free degrees of freedom available for estimating the means and variances of response, and thus reduces or even eliminates any efficiency gained by using a vector-valued **IM**. Therefore, this method is essentially limited to a two-parameter vector where the second parameter is coarsely discretized (which is perhaps not very restrictive in practice, although it is a shortcoming for exploratory research).

Despite these disadvantages, the ability to use the same set of records at all IM_1 levels is very desirable in some situations. This is of particular concern if one is trying to account for a second *IM* parameter using Incremental Dynamic Analysis, where the records must stay the same as the IM_1 level changes. For the IDA situation especially, this may be a preferred method of incorporating vector-valued *IM*s.

2.6.3 Comparison of results from alternative methods

To demonstrate the effect of the hybrid methods, several response estimates are compared to the original estimates based on *Sa*(0.8s) alone, and displayed in Figure 2.21. These results come from the same structural model and record set as used before. The exception to this is for the special record-selection scheme of Section 2.6.1 where records are re-selected at each *Sa* level to match the ε distribution obtained from disaggregation (record-selection details will be given in

Chapter 6). We see that all three of the methods that account for the effect of ε result in a lowering of the drift hazard curve. Although the agreement is not perfect, clearly all three methods are capturing the effect of ε to some extent. Thus the hybrid methods based on record selection or reweighting can achieve similar results to the vector IM method that uses regression on ε. This result also shows that the scalar *IM Sa*(0.8s) is biased ("insufficient") at large levels of EDP (i.e., incorporating ε in the prediction changes the answer, suggesting that the scalar IM consisting of *Sa* alone does not provide complete information about the effect of the ground motions). Further evidence for this will be presented in Chapters 3 and 6.

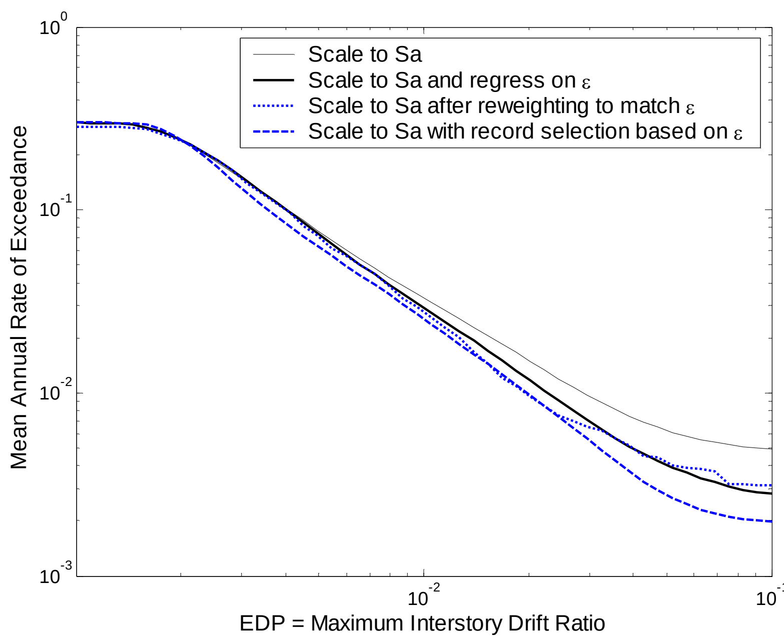


Figure 2.21. Comparison of drift hazard results using the vector-IM based regression procedure of Section 2.5.2, the record selection procedure of Section 2.6.1 and the reweighting procedure of Section 2.6.2. The scalar intensity measure used is *Sa*(0.8s), and the vector is *Sa*(0.8s) and ε.

2.7 Vector-valued EDPs

This chapter has focused on incorporating vector-valued *IM*s into structural analyses, but vector-valued *EDP*s are also a useful quantity in many assessments. When performing loss estimation at a detailed level, it may be important to know the distribution of peak interstory drift ratios and peak floor accelerations at each level of a multi-story building. Fortunately this extension is not a difficult one.

The conditional distribution of individual *EDP*s (given the *IM* value) are often assumed to be lognormally distributed, so that ln*EDP* is normally distributed (Shome and Cornell 1999, Aslani and Miranda 2003). Making an additional assumption that multiple ln*EDP*s have a joint normal distribution allows for a fairly simple estimation method. We compute the sample mean and variance for each of the individual *EDP* parameters as before. For the joint normal distribution, the only additional information that is needed is the correlation between the residuals for each ln*EDP* parameter, which can be easily estimated from the same data used to estimate means and variances. This has been done using the parametric stripe method by Baker and Cornell (2003) and using the cloud method by Luco et al (2005).

If one assumes constant correlations for all *IM* levels (as is typically done with cloud methods), then all of the prediction residuals can be used simultaneously to estimate correlations. If one would like the correlations to vary as a function of a scalar *IM*, then residuals from individual stripes would be used separately for estimates. If one would like the correlations to vary with a vector of *IM* parameters, estimations of correlations will be more difficult. But such a complicated model for correlations may be unnecessary. Merely assuming correlations to be constant or a function of a scalar *IM* parameter is a positive step and is likely to be sufficient for many applications.

2.8 Record scaling for response estimation

Several of the above procedures (which use “stripes” or “IDAs”) require records to be scaled. Even with the cloud methods, which do not require scaling, the records are often scaled by a factor of two or more in order to induce greater nonlinearity in the structure during assessment (as they were in the example above). The use of scaling leads to questions about whether the scaled ground motions are truly representative of ground motions with the given *IM* level. From a structural engineer’s perspective, the question can be phrased, “will ground motions scaled to a specified *IM* level produce the same structural responses as unscaled ground motions naturally at that *IM* level?”

Previous work indicates that a suite of records may be safely scaled to the median *Sa* value of that suite, without biasing the median structural response value (Shome et al. 1998, Iervolino and Cornell 2005a). But Luco and Bazzurro (2005) find that in some other situations record scaling may induce some bias in structural response, a conclusion also reached in Chapter 6 of this dissertation. In Chapter 6 we will conclude that this bias results from the scaled records having inappropriate values of other intensity parameters (such as ε or the spectral shape parameter discussed in Chapter 4). Therefore, despite intuitive concerns about the ground-motion scaling used with some of the above methods, Chapter 6 will show that using an appropriate record-selection scheme (such as that described in Section 2.6.1) or a proper vector-valued *IM* will prevent scaling from causing any significant biases in estimated structural response.

2.9 Summary

There are a variety of ways in which the distribution of *EDP* given a scalar *IM* can be estimated. Several possibilities have been discussed above, and are summarized in Table 2.1. In general, one must make trade-offs between the accuracy of the estimates and the amount of data required for estimation.

In addition, *EDP* estimation methods that utilize vector-valued intensity measures are described. The same tradeoffs between speed and accuracy tend to hold. Cloud methods require relatively few dynamic analyses and most effectively avoid the “curse of dimensionality,” but the simple linear regression model may sometimes need modification in order to capturing the potentially complex relationships between the *IM* parameters and the response parameter of interest. At the other extreme, nonparametric methods such as empirical cumulative distribution functions are potentially very accurate, but may require a prohibitive number of dynamic analyses, especially for vectors containing many parameters. A middle ground is found by scaling records to the dominant IM_1 parameter and using regression analysis to predict the remaining effect of IM_2, IM_3, etc. This allows more freedom in functional forms than cloud methods, but constrains the number of model parameters to a more reasonable number than with fully nonparametric methods. For this reason, it is adopted for most of the investigations in this dissertation.

Table 2.1. Summary of *EDP|IM* estimation methods considered.

Method	**Pros**	**Cons**	**“Vectorization” Potential**
Linear regression on a cloud	Can avoid record scaling. Generally requires fewer records than other methods. Compatible with closed-form solutions for drift hazard.	Approximations (linear conditional mean and constant variance) can cause difficulties over large ranges of *IM*. Relaxation of the approximations is possible, but removes some of the advantages of this choice.	Multiple regression on vector *IM*s is a simple extension. Collinearity among predictors may be a problem.
Parametric distributions on *IM* stripes	Fewer parametric assumptions than with clouds. *IM* dependence can vary by *IM* level.	Typically requires more dynamic analyses than cloud methods.	Regression on stripes is less restrictive than regression on all *IM* parameters simultaneously, while also limiting the required number of dynamic analyses.
Empirical CCDF on *IM* stripes	No parametric assumptions about response distribution	Requires a significant number of records for estimation.	Curse of dimensionality is a significant problem. Application to vectors requires a significant number of records, carefully selected
IM capacity from IDAs	Requires fewer runs than *IM* stripes if IDAs are formed carefully.	Requires a few extra steps to create IDAs and interpolate to compute capacity distributions	Generalization to vectors is straightforward.
Hybrid method: Scale ground motions to IM_1, while carefully selecting the records to match other *IM* parameters	Achieves the gains of vector *IM* methods while requiring only the scalar-*IM* processing procedure.	Requires careful record selection. Records likely need to be re-selected at differing IM_1 levels.	Simpler post-processing than explicit vector procedures.
Hybrid method: Scale ground motions to IM_1, while re-weighting the data to match the target distribution of other *IM* parameters.	After re-weighting, simple scalar-IM methods can be applied but a vector-IM result is obtained. Records do not need to be re-selected at each IM_1 level.	Weights for results from some records will be zero or nearly zero, essentially throwing away data. The curse of dimensionality is a problem for vectors with many elements.	Simpler post-processing than explicit vector procedures.

Chapter 3

A Vector-Valued Ground Motion Intensity Measure Consisting of Spectral Acceleration and Epsilon

Baker, J.W. and Cornell, C.A., (2005) “A Vector-Valued Ground Motion Intensity Measure Consisting of Spectral Acceleration and Epsilon.” *Earthquake Engineering & Structural Dynamics* **34**(10). 1193-1217.

3.1 Abstract

The “strength” of an earthquake ground motion is often quantified by an Intensity Measure (*IM*), such as peak ground acceleration or spectral acceleration at a given period. This *IM* is used to predict the response of a structure. In this paper an intensity measure consisting of two parameters, spectral acceleration and epsilon, is considered. The *IM* is termed a vector-valued *IM*, as opposed to the single parameter, or scalar, *IM*s that are traditionally used. Epsilon (defined as a measure of the difference between the spectral acceleration of a record and the mean of a ground motion prediction equation at the given period) is found to have significant ability to predict structural response. It is shown that epsilon is an indicator of spectral shape, explaining why it is related to structural response. By incorporating this vector-valued *IM* with a vector-valued ground motion hazard, we can predict the mean annual frequency of exceeding a given value of maximum interstory drift ratio, or other such response measure. It is shown that neglecting the effect of epsilon when computing this drift hazard curve leads to conservative estimates of the response of the structure. These observations should perhaps affect record selection in the future.

3.2 Introduction

As nonlinear dynamic analysis becomes a more frequently used procedure for evaluating the demand on a structure due to earthquakes, it is increasingly important to understand which properties of a recorded ground motion are most strongly related to the response caused in the structure. A value that quantifies the effect of a record on a structure is often called an Intensity Measure (*IM*). The Peak Ground Acceleration of a record was a commonly used *IM* in the past. More recently, spectral response values (e.g., spectral acceleration at the first-mode period of vibration – $S_a(T_1)$) have been used as *IM*s. Spectral acceleration at T_1 has been found to be an effective *IM* (Shome et al. 1998), but among records with the same value of $S_a(T_1)$, there is still significant variability in the level of structural response in a multi-degree-of-freedom, nonlinear structural model. If some of this remaining record-to-record variability could be accounted for by an improved intensity measure, then the accuracy and efficiency of structural response calculations could be improved.

In this paper, vector-valued intensity measures are considered as potential improvements on current intensity measures. The intensity measures considered consist of $S_a(T_1)$ as before, but also include a second parameter: the magnitude, distance or ε (“epsilon”) associated with the ground motion. The *IM* is termed vector-valued because it now has two parameters, as opposed to traditional scalar, or single-parameter, *IM*s. It is found that the vector-valued *IM* consisting of $S_a(T_1)$ and ε is significantly superior to the *IM* consisting of $S_a(T_1)$ alone. The predictive power of ε is demonstrated, and an intuitive understanding is developed about the source of this predictive power.

3.3 What is epsilon?

Magnitude and distance are familiar quantities to any earthquake engineer, but understanding of the ε parameter may be less common. Epsilon is defined by engineering seismologists studying ground motion as the number of standard deviations by which an observed logarithmic spectral acceleration differs from the mean logarithmic spectral acceleration of a ground-motion prediction (attenuation) equation. Epsilon is computed by subtracting the mean predicted $\ln S_a(T_1)$ from the record’s $\ln S_a(T_1)$, and dividing by the logarithmic standard deviation (as estimated by the prediction equation). Epsilon is defined with respect to the unscaled record and will not change in value when the record is scaled. We will see later that ε is an implicit indicator of the “shape” of the response spectrum, and the shape of the spectrum does not change with scaling, providing intuition as to why ε would not vary with scaling.

Because of the normalization by the mean and standard deviation of the ground motion prediction equation, ε is a random variable with an expected value of zero, and a unit standard deviation. In fact, the distribution of ε is well represented by the standard normal distribution, at least within values of ±3 (Abrahamson 1988, 2000b). Thus, a sample of randomly chosen records will have an average ε value near zero, as can be seen visually in Figure 3.1b, although an average value of zero is not required for our vector-valued *IM* work below. It should be noted that for a given ground motion record, ε is a function of T_1 (i.e., ε will have different values at different periods) and the ground motion prediction model used (because the mean and standard deviation of $\ln S_a(T_1)$ vary somewhat among models). Epsilon is also a function of damping, because both the observed and predicted spectral acceleration values are functions of the damping used; throughout this dissertation, 5% damping will be used for consistency with most ground motion prediction models. The definition of ε is valid for any ground motion prediction model, but the model of Abrahamson and Silva (1997) is the only one used in calculations here. If one would like to use ε in a vector *IM* to compute drift hazard, the model used to compute ε should be the same as the model used to perform the ground motion hazard assessment and disaggregation.

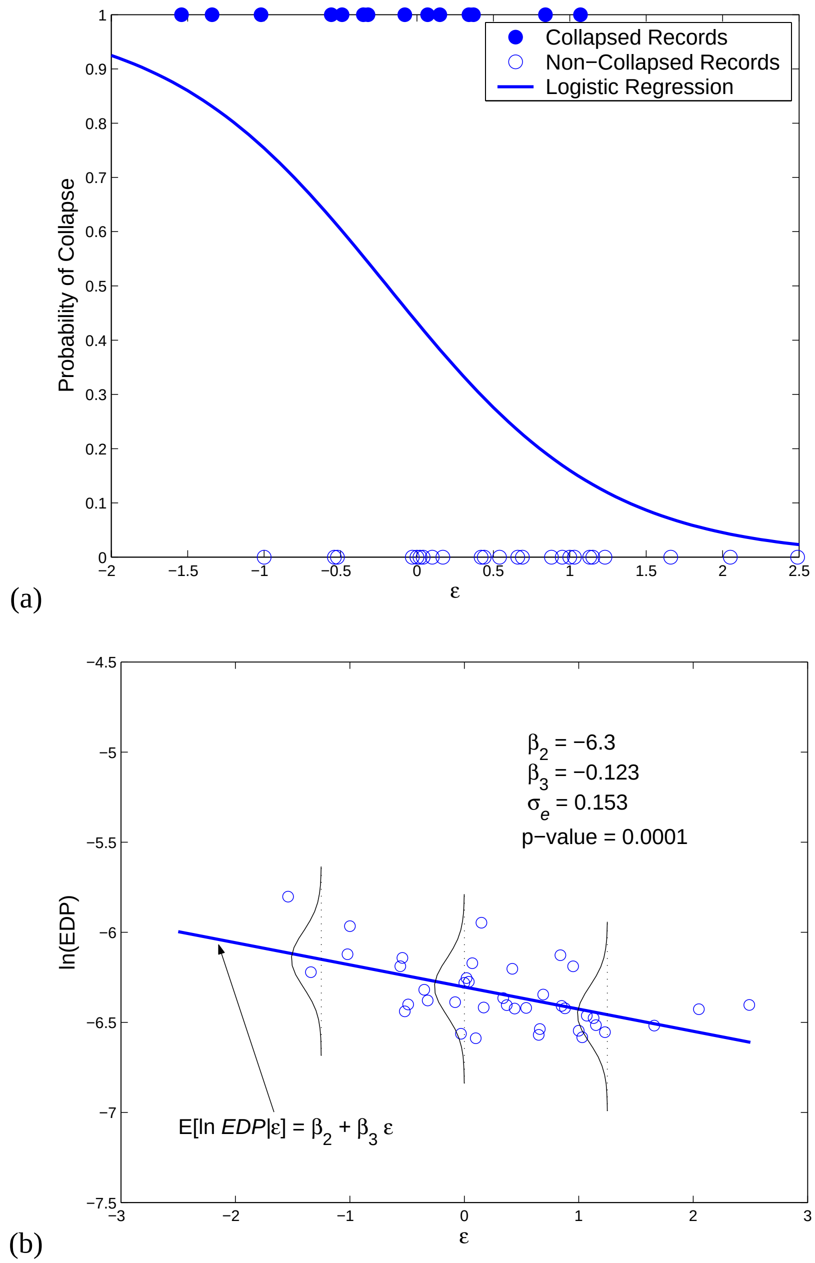


Figure 3.1. Analysis of data at a fixed value of S_a, (a) Prediction of the probability of collapse using logistic regression applied to binary collapse/non-collapse results; and (b) prediction of response given no collapse, with the distribution of the residuals superimposed over the data.

It should also be noted that there is more than one way to define spectral acceleration, and that the choice of definition will affect the computation of ε. Many ground motion prediction equations provide mean and standard deviation values for the average $\ln S_a(T_1)$ of the two horizontal components of a ground motion (Abrahamson and Silva 1997). However, in this work we analyze only 2D frames, and thus use only one (arbitrarily chosen) component of a given ground motion. Thus, we use $\ln S_a(T_1)$ of an arbitrary horizontal component of the ground motion as our *IM*. It is therefore important that we compute our ground motion hazard for $\ln S_a(T_1)$ of an arbitrary component (as opposed to the average $\ln S_a(T_1)$ provided by the ground motion prediction equation). In this study, the ground motion hazard was computed using Abrahamson and Silva's equation, but the standard deviation of $\ln S_a(T_1)$ was inflated to reflect the increased variability of an arbitrary component of $\ln S_a(T_1)$ rather then the average of two components. The inflation factor was determined from another ground motion prediction model (Boore et al. 1997), which presents standard deviation values for both definitions of spectral acceleration[8]. This inflated standard deviation must also be used when computing the ε values for ground motion recordings as well. This issue is explained more thoroughly in Chapter 7.

Now that ε has been defined, we discuss how it and other *IM*s are used to predict drift, and then incorporated with ground motion hazard results to compute a drift hazard curve.

3.4 Calculation of the drift hazard curve using a scalar IM

Once an intensity measure is defined, the predicted structural response given an intensity measure level can be combined with Probabilistic Seismic Hazard Analysis (PSHA) to calculate the mean annual rate of exceeding a given structural response level. An example of the need for this calculation is seen in the work of the Pacific Earthquake Engineering Research (PEER) Center (Cornell and Krawinkler 2000). Here, following PEER practice, the response of a structure is termed an Engineering Demand Parameter, or *EDP* (in this paper, the only *EDP* considered is maximum interstory drift ratio, although the methodology is directly applicable to any *EDP* of interest). The annual frequency of exceeding a given level of the *EDP* is calculated as follows:

$$\begin{aligned}\lambda_{EDP}(z) &= \int_x P(EDP > z \mid IM = x) \cdot \left| d\lambda_{IM}(x) \right| \\ &\cong \sum_{\text{all } x_i} P(EDP > z \mid IM = x_i) \cdot \Delta\lambda_{IM}(x_i)\end{aligned} \tag{3.1}$$

where $\lambda_{EDP}(z)$ is the mean annual frequency of exceeding a given *EDP* value z, $\lambda_{IM}(x_i)$ is the mean annual frequency of *IM* exceeding a given value x_i (this is commonly referred to as the seismic hazard curve), and $\Delta\lambda_{IM}(x_i) = \lambda_{IM}(x_i) - \lambda_{IM}(x_{i+1})$ is approximately the annual frequency of $IM=x_i$. The term $P(EDP>z|IM=x_i)$ represents the probability of exceeding a specified *EDP* level, z, given $IM=x_i$. In this paper we will use numerical integration to compute results, making use of the discrete summation approximation. We see that the rate of exceeding a given *EDP* level is found

[8] The inflation factor can now be determined from the models of Chapter 8 as well, but that work had not been completed at the time this research was performed. Changing the inflation factor would not have affected any of the conclusions drawn in this chapter.

by assessing the ground motion hazard and the response of the structure, and coupling these two parts together with the use of an *IM*. Note that methods of computing drift hazard with techniques other than the *IM*-based method have been proposed (Bazzurro et al. 1998, Jalayer et al. 2004), but the *IM*-based method is the focus of this paper. If the *IM* is a vector, Equation 3.1 must be generalized, as will be discussed below.

3.5 Prediction of structural response using a scalar IM

The *IM*-based procedure described above requires estimation of the probability distribution of structural response at a given *IM* level (i.e., $P(EDP>z|S_a(T_1)=x_i)$ in Equation 3.1). An estimation procedure is now described for the scalar case, and later generalized to the case of a vector-valued *IM*. The scalar *IM* $S_a(T_1)$ is used in this section, both because of its wide use elsewhere, and because it will be easily generalized to our vector case: $S_a(T_1)$ and ε. Standard terminology from regression analysis is used in this section, to allow for quick descriptions of some concepts from statistics. The reader desiring a more detailed explanation is referred to a previous related publication (Baker and Cornell 2004).

The method used in this paper requires a suite of earthquake accelerograms, all at the same *IM* value, $S_a(T_1) = x$ (e.g., in this study, 40 records are used at each *IM* level). We scale a suite of recorded earthquake accelerograms to the given $S_a(T_1)$ value (e.g., Shome et al. 1998). In this study, we use the same suite of records for different $S_a(T_1)$ levels, although one could use different record suites at different levels if PSHA disaggregation suggested that, for example, the representative magnitude level was changing (Stewart et al. 2001)—we shall return to the record selection subject below. This suite of records is used to perform nonlinear dynamic analysis on a model of the structure. Now we have n records, all with $S_a(T_1) = x$, and n corresponding values of *EDP*. So *EDP* given $S_a(T_1) = x$ is a random variable that we need to characterize.

3.5.1 Characterizing the Collapses

When predicting nonlinear response of structures, it is necessary to account for the possibility that some records may cause collapse of the structure at higher levels of IM. For the purpose of illustration here, collapse is defined to have occurred if the dynamic analysis algorithm fails to converge due to numerical instability or if the drift ratio at any story exceeds 10%. Such a finite cutoff is used because the validity of current nonlinear models is not well confirmed beyond large deformation levels, which a real structure may not be able to reach before collapsing. For example, for the particular older reinforced concrete frame considered below, one might in fact expect axial failure of columns in the 3-5% interstory drift ratio range. This failure mechanism is difficult to model and was not incorporated in the computer model, resulting in larger displacements being obtained before collapse was signaled by numerical instability of the program. Other collapse criteria that have been used in such studies include dynamic instability, defined as the ratio of the increment in displacement to the increment in spectral acceleration level exceeding a specified threshold. The simpler 10% drift criterion is used here for illustration, but the proposed procedure applies universally, regardless of the structural model or the specified collapse criteria. Future research to more precisely model and identify response levels associated with collapse will be helpful to the drift hazard procedure presented here.

To account for these collapses, we separate our realizations of *EDP* into collapsed and non-collapsed data. We then estimate *P*, the probability of collapse, *C*, at the given $S_a(T_1)$ level. This estimate is denoted $\hat{P}$ and calculated as:

$$\hat{P}(C \mid S_a(T_1) = x) = \frac{\text{number of records causing collapse}}{\text{total number of records}} \tag{3.2}$$

We then return to the non-collapse responses for the remainder of the response prediction.

3.5.2 Characterizing the non-collapse responses

The distribution of the non-collapsed responses for our *EDP*, maximum interstory drift ratio, has been found to be well represented by a lognormal distribution (the Kolmogorov-Smirnov test, Neter et al. 1996, was used to verify this supposition, and the same conclusion has been reached elsewhere, Shome 1999, Aslani and Miranda 2003). For this reason we work with the natural logarithm of *EDP*, which then follows the normal distribution. We can estimate the parameters for this normal distribution using the method of moments (Benjamin and Cornell 1970). For each *IM* level, we denote the estimated mean of ln*EDP* as $\hat{\mu}_{\ln EDP|S_a(T_1)=x}$ and the estimated standard deviation as $\hat{\beta}_{\ln EDP|S_a(T_1)=x}$ (this logarithmic standard deviation is sometimes referred to as "dispersion"). The probability that *EDP* exceeds *z* given *IM* = *x* and no collapse can now be calculated using the normal complimentary cumulative distribution function:

$$P(EDP > z \mid S_a(T_1) = x, \text{ no collapse}) = 1 - \Phi\left(\frac{\ln z - \hat{\mu}_{\ln EDP|S_a(T_1)=x}}{\hat{\beta}_{\ln EDP|S_a(T_1)=x}}\right) \tag{3.3}$$

where $\Phi(\bullet)$ denotes the standard normal cumulative distribution function.

3.5.3 Combining collapse and non-collapse results

We now combine the characterizations of collapses and non-collapse responses using the total probability theorem. Our estimate of the probability that *EDP* exceeds *z* given *IM* = *x* is:

$$P(EDP > z \mid S_a(T_1) = x) = \hat{P}(C \mid S_a(T_1) = x) + \left(1 - \hat{P}(C \mid S_a(T_1) = x)\right)\left(1 - \Phi\left(\frac{\ln z - \hat{\mu}_{\ln EDP|S_a(T_1)=x}}{\hat{\beta}_{\ln EDP|S_a(T_1)=x}}\right)\right) \tag{3.4}$$

We can now proceed to work with this estimate.

3.6 Calculation of the drift hazard curve using a vector-valued IM

A vector-valued *IM* can also be used to compute a drift hazard curve using a generalization of Equation 3.1, as given in Equation 3.5 (Bazzurro and Cornell 2002).

$$\begin{aligned} \lambda_{EDP}(z) &= \int_{x_1}\int_{x_2} P(EDP > z \mid S_a(T_1) = x_1, \varepsilon = x_2) \cdot \left| \frac{\partial^2 \lambda_{IM}(x_1, x_2)}{\partial x_1 \partial x_2} \right| dx_1 dx_2 \\ &\cong \sum_{\text{all } x_{1,i}} \sum_{\text{all } x_{2,j}} P(EDP > z \mid S_a(T_1) = x_{1,i}, \varepsilon = x_{2,j}) \cdot \Delta\lambda_{IM}(x_{1,i}, x_{2,j}) \end{aligned} \tag{3.5}$$

We see first that the scalar-*IM* drift prediction $P(EDP>z|IM=x)$ has been replaced with the vector-*IM* drift prediction $P(EDP>z|S_a(T_1)=x_1,\varepsilon=x_2)$, which will be expanded in a means analogous to Equation 3.4 in Equation 3.11 below. In addition, the scalar-*IM* ground motion hazard has been replaced by the joint hazard of the vector-valued *IM*. Defining $\Delta\lambda_{IM}(x_{1,i}, x_{2,j})$ as $\lambda_{Sa\in[x_{1,i},x_{1,i+1}],\varepsilon\in[x_{2,j},x_{2,j+1}]}$, we take advantage of the fact that we could also express this as the marginal rate density of $S_a(T_1)$, and the conditional probability distribution of ε given $S_a(T_1)$:

$$\Delta\lambda_{IM}(x_{1,i}, x_{2,j}) = P(x_{2,j} < \varepsilon < x_{2,j+1} \mid S_a(T_1) = x_{1,i}) \cdot \Delta\lambda_{S_a(T_1)}(x_{1,i}) \tag{3.6}$$

Then Equation 3.5 can be restated as:

$$\lambda_{EDP}(z) = \sum_{\text{all } x_{1,i}} \sum_{\text{all } x_{2,j}} P(EDP > z \mid S_a(T_1) = x_{1,i}, \varepsilon = x_{2,j}) \cdot P(x_{2,j} < \varepsilon < x_{2,j+1} \mid S_a(T_1) = x_{1,i}) \cdot \Delta\lambda_{S_a(T_1)}(x_{1,i}) \tag{3.7}$$

We state the equation in this way because we obtain the distribution of $S_a(T_1)$ and ε from PSHA in this form: $\Delta\lambda_{S_a(T_1)}(x_{1,i})$ comes from the standard PSHA hazard curve, and $P(x_{2,j} < \varepsilon < x_{2,j+1} \mid S_a(T_1) = x_{1,i})$ is a standard disaggregation result. Note that the disaggregation can be presented in more than one way. Some PSHA codes provide $P(x_{2,j} < \varepsilon < x_{2,j+1} \mid S_a(T_1) = x_{1,i})$ (see McGuire 1995), while others provide $P(x_{2,j} < \varepsilon < x_{2,j+1} \mid S_a(T_1) \geq x_{1,i})$ (see Bazzurro and Cornell 1999). For instance, Abrahamson's code (2004) provides $P(x_{2,j} < \varepsilon < x_{2,j+1} \mid S_a(T_1) = x_{1,i})$, while the U.S. Geological Survey provides $P(x_{2,j} < \varepsilon < x_{2,j+1} \mid S_a(T_1) \geq x_{1,i})$. It is fairly simple matter to convert the results between the two forms (Bazzurro 1998, p195), but one should be aware of which version is provided by the software in use, and convert the results if necessary. The hazard assessments in this study were performed using the Abrahamson code, which provides results (e.g., Figure 3.6) directly in the form needed for drift hazard calculations.

3.7 Prediction of building response using a vector-valued IM

We now generalize the prediction procedure of the preceding section for use with a vector-valued *IM*. Consider the vector consisting of $S_a(T_1)$ and ε. We are now trying to estimate $P(EDP > z \mid S_a(T_1) = x_1, \varepsilon = x_2)$. The simplest solution, if possible, would be to scale our records to both parameters of our *IM*: $S_a(T_1) = x_1$ and $\varepsilon = x_2$. However, ε is defined with respect to the unscaled record, and does not change with scaling (similarly, magnitude and distance do not change with scaling). Because of this, we need a supplement to scaling for the vector-valued *IM* procedure, in order to predict response as a function of the second *IM* parameter.

The solution we adopt is to scale to $S_a(T_1)$ as before, and then apply regression analysis to estimate *EDP* as a function of ε (Neter et al. 1996). Thus our treatment of $S_a(T_1)$ remains the same as in the scalar case, but we now incorporate information from the regression on ε. Our approach with the vector case is the same as the scalar case in that we separate out the collapse responses first, and then deal with the remaining non-collapse responses.

3.7.1 Accounting for collapses with the vector-valued IM

When using a vector-valued *IM*, instead of taking the probability of collapse to be simply the fraction of records that cause collapse, we can take advantage of the second *IM* parameter to predict the probability of collapse more accurately. We do this using logistic regression, which is a commonly used tool for analyzing binary data (Neter et al. 1996). It should be noted that this is not the only method for quantifying the probability of collapse. For example, a bivariate normal model for collapse *capacity* of the structure could be defined and used to estimate probability of collapse. The results from the two models will be similar (e.g., Figure 2.20), but it may be more convenient to adopt one or the other in certain cases, as discussed in Chapter 2.

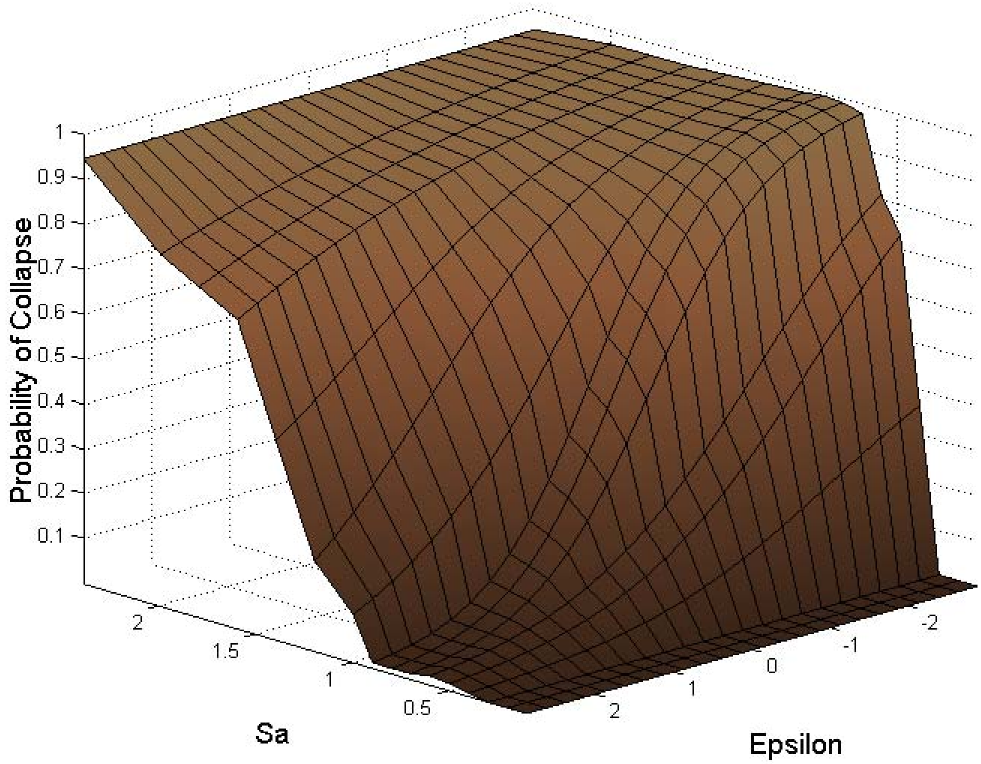


Figure 3.2. Prediction of the probability of collapse as a function of both $S_a(T_1)$ and ε.

With the logistic regression procedure used here, each record has a value of ε, which we use as our predictor variable. We designate C as an indicator variable for collapse (C is equal to 1 if the record causes collapse and 0 otherwise). We then use the logistic regression to predict collapse:

$$\hat{P}\left(C \mid S_a(T_1) = x_1, \varepsilon = x_2\right) = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_2)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_2)} \tag{3.8}$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are coefficients to be estimated from regression on a dataset that has been scaled to $S_a(T_1) = x_1$ (i.e., $\hat{\beta}_0$ and $\hat{\beta}_1$ will be different for different values of $S_a(T_1)$). An example of this data and a fitted logistic regression curve is presented in Figure 3.1a. The tendency for the probability of collapse to decay with increasing ε is common; such observations will be discussed below. By performing this regression for all $S_a(T_1)$ levels, one can obtain the probability of collapse as a function of both $S_a(T_1)$ and ε, as seen in Figure 3.2.

It should be noted that the level of confidence in the result from Equation 3.8 depends on the nature of the data used in the regression analysis. If there are significant numbers of both collapses and non-collapses in the dataset, then the regression should be very stable. However, if for a given S_a level, the dataset consists of, for instance, 39 records that do not cause collapse, and one record that causes collapse, the logistic regression will be strongly influenced by the single collapse data point, and may indicate a different trend that in fact exists (if the exercise were to be repeated with more records or different records). For this reason, it is suggested that if there are two or fewer collapse data points, then the probability of collapse should be taken as a simple constant (i.e., $1/n$ or $2/n$, where n is the number of records) for all levels of x_2. This is equivalent to using the scalar-*IM* procedure for the collapse portion of the prediction (Equation 3.2). The same should be done in the case where all but one or two records cause collapse. It should be noted that in these cases, the probability of collapse will already be very high or very low, so the second *IM* parameter would not have much effect anyway. This modification to the procedure will merely prevent the logistic regression prediction from producing unstable results which are strongly influenced by a single data point.

3.7.2 Characterizing non-collapses with the vector-valued IM

We now incorporate the second parameter of our *IM* in prediction of response for the non-collapse results. Note that each of these records has been scaled to $S_a(T_1) = x_1$. Each of the records has a value of ε and a value of *EDP*. We have found that there tends to be a relationship between ε and *EDP* of the form $\ln EDP = \beta_2 + \beta_3\varepsilon + e$, where β_2 and β_3 are constant coefficients, and e is the prediction error ("residual"). We can use linear least-squares regression (Neter et al. 1996) to obtain estimates of the two regression coefficients, $\hat{\beta}_2$ and $\hat{\beta}_3$ (again, these values will vary for different $S_a(T_1)$ levels). A graphical example of this data and the regression fit is shown in Figure 3.1b.

When using linear least-squares regression on a dataset, several assumptions are normally implicitly made, and the accuracy of the results depends on the validity of these assumptions. The prediction error of record i (the difference between the predicted value of $\ln EDP_i$ and the actual value) is termed the "residual" of record i, and is assumed to be mutually independent from the prediction error of record j for all $i \neq j$. In addition, we will later assume the residuals to be normally distributed with constant variance (i.e., homoscedastic). The assumptions of independent normal residuals with constant variance have been examined for the data in this study, and found to be reasonable. An estimate of the variance of the residuals is also available from the analysis software, and we denote it $V\hat{a}r[e] \equiv \hat{\sigma}_e^2$. This variance in the residuals is displayed graphically in Figure 3.1b, by superimposing the estimated normal distribution of the residuals over the data. From regression, we now know that given $S_a(T_1) = x_1$, $\varepsilon = x_2$ and no collapse, the mean value of ln*EDP* is:

$$E\left[\ln EDP\right] = \hat{\beta}_2 + \hat{\beta}_3 x_2 \tag{3.9}$$

where $\hat{\beta}_2$ and $\hat{\beta}_3$ have been obtained by regressing on records scaled to $S_a(T_1) = x_1$. We also know that, conditional on $S_a(T_1)$ and ε, ln*EDP* is normally distributed with variance equal to the $\hat{\sigma}_e^2$. So the probability that *EDP* exceeds z, given $S_a(T_1) = x_1$, $\varepsilon = x_2$, and no collapse can be expressed:

$$P\left(EDP > z \mid S_a(T_1) = x_1, \varepsilon = x_2, \text{no collapse}\right) = 1 - \Phi\left(\frac{\ln z - \left(\hat{\beta}_2 + \hat{\beta}_3 x_2\right)}{\hat{\sigma}_e}\right) \tag{3.10}$$

Recall that $\hat{\beta}_2$, $\hat{\beta}_3$ and $\hat{\sigma}_e^2$ are all functions of $S_a(T_1)$ or x_1. This equation is very similar to Equation 3.3 used in the scalar case. We previously estimated the mean of the normal distribution by the average logarithmic response of all records, but now we use a result from regression on ε. We have also replaced the standard deviation of the records by the standard deviation of the regression residual. Otherwise, the equation is the same.

As with the estimation of collapse, there is more than one way to incorporate the vector *IM*. For example, rather than using regression, one could use a weighted scheme where the weights of each record are determined according to the results of the PSHA analysis (Shome 1999). The results from these two schemes will agree closely, but one may be more appropriate than the other based on the amount of data available and the extent to which the effect of the *IM* can be parameterized. This will be described in a future publication by the authors.

We now combine the possibilities of collapse or no collapse, using the Total Probability Theorem and Equations 3.8 and 3.10, to compute the conditional probability that *EDP* exceeds z:

$$P(EDP > z \mid S_a(T_1) = x_1, \varepsilon = x_2) = \hat{P}(C) + \left(1 - \hat{P}(C)\right)\left(1 - \Phi\left(\frac{\ln z - \left(\hat{\beta}_2 + \hat{\beta}_3 x_2\right)}{\hat{\sigma}_e}\right)\right) \tag{3.11}$$

$$\text{where } \hat{P}\left(C\right) = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_2)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_2)}$$

Although x_1 does not appear in Equation 3.11, our estimate is implicitly a function of x_1, because the data used to estimate $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ and $\hat{\sigma}_e$ all come from records scaled to $S_a(T_1) = x_1$. This gives us a response prediction that is similar to the original prediction of Equation 3.4, but that now incorporates a two-element vector.

3.8 Investigation of magnitude, distance and epsilon as IM parameters

The procedure for evaluating drift hazard using a vector-valued *IM* can now be used to assess the response-predicting effectiveness of ε as an element in a vector with $S_a(T_1)$. Additionally, we will examine magnitude (M) and distance (R) as elements in a vector with $S_a(T_1)$ using the same procedure, to evaluate whether they have any significant effect on structural response after conditioning on $S_a(T_1)$. This study considers only M, R and ε because the conditional distribution $P(x_{2,j}<IM_2<x_{2,j+1}|S_a(T_1)=x_{1,i})$ is then easily available from standard PSHA software (where IM_2 is used to represent either M, R or ε). If one is interested in the effect of other parameters such as spectral shape or duration, special modifications to the PSHA analysis (Bazzurro and Cornell 2002) are needed in order to obtain the conditional distribution $P(x_{2,j}<IM_2<x_{2,j+1}|S_a(T_1)=x_{1,i})$.

To test the effectiveness of M, R or ε as predictors, we examine the results from our response regressions on these variables. The seven story concrete frame structure described below was used to generate response data. At each level of $S_a(T_1)$, we predict collapse using logistic regression on M, R or ε, and we use linear regression on these variables to model the non-collapse responses. An effective predictor should show a trend in one or both of these regressions and the trend (i.e., slope of the regression) should be statistically significant. A standard way of measuring statistical significance is with the "p-value" for the regression coefficient. Typically, a p-value smaller than 0.05 is interpreted to indicate that the predictive variable is significant (Neter et al. 1996), although when a slightly larger p-value shows up repeatedly in separate tests of the same predictor variable, this can also be interpreted as an indicator of significance. A p-value can also be computed for both the logistic and linear regression results.

When distance is considered as a candidate *IM* parameter, we find no statistical significance (median p-values of 0.34 and 0.29 for linear and logistic regression respectively, considering 13 levels of $S_a(T_1)$). Both ε and magnitude show some significance for the linear response regression (median p-values of 0.05 and 0.06 respectively), although ε shows more significance than M for the logistic collapse probability regression (median p-values of 0.14 and 0.37 respectively). The lack of significance of R and slight significance of M are consistent with results from previous work on this topic (e.g., Shome et al. 1998). The significance of ε and the evaluation of significance with respect to collapse prediction are believed to be new results.

The p-values from both regressions using M and ε as predictors at 13 levels of spectral acceleration are given in Table 3.1 (distance has been omitted because of the consistent lack of significance it demonstrated). Values for logistic regression are omitted when fewer than three records cause collapse, preventing the regression from being performed. At high $S_a(T_1)$ levels, when nearly all of the records cause collapse and there is little data for either the linear or logistic regression, the regressions are less useful, and the results are omitted.

Table 3.1. P-values from linear and logistic regression on magnitude and epsilon.

		Magnitude				Epsilon			
Sa	Percent Collapsed	Linear Regression p-value	Logistic Regression p-value	Linear Regression Coefficient	Potential Prediction Error (M=6.5 vs M=7)	Linear Regression p-value	Logistic Regression p-value	Linear Regression Coefficient	Potential Prediction Error (ε=0 vs ε=1.5)
0.1	0%	0.04	-	0.13	7%	0.00	-	-0.15	24%
0.2	0%	0.09	-	0.16	8%	0.01	-	-0.15	25%
0.3	0%	0.05	-	0.25	13%	0.07	-	-0.14	24%
0.4	0%	0.04	-	0.29	16%	0.05	-	-0.19	32%
0.5	0%	0.06	-	0.35	19%	0.01	-	-0.30	56%
0.6	3%	0.06	-	0.37	20%	0.13	-	-0.20	34%
0.7	8%	0.05	0.36	0.33	18%	0.03	0.14	-0.24	42%
0.8	9%	0.20	0.45	0.23	12%	0.01	0.26	-0.30	57%
0.9	14%	0.10	0.29	0.36	20%	0.02	0.20	-0.32	61%
1.0	16%	0.04	0.37	0.41	22%	0.08	0.14	-0.24	43%
1.2	24%	0.10	0.10	0.28	15%	0.13	0.04	-0.19	32%
1.4	56%	0.10	0.98	0.49	28%	0.17	0.09	-0.23	42%
1.6	68%	0.51	0.66	0.27	14%	0.69	0.05	-0.09	14%
	Median	0.06	0.37	0.29	16%	0.05	0.14	-0.20	34%

In addition to standard p-values, a value termed the "potential prediction error" is also reported in Table 3.1. This value is defined as the percentage difference in predicted response for a reasonable range of values in the second *IM* (i.e., what is the difference in response between magnitude 6.5 and 7 records, given the same $S_a(T_1)$ level). In Table 3.1 we see that for the given range of variation, a change in magnitude produces a potential prediction error of 16%, while a change in ε produces a potential prediction error of 34%. This major difference is because, while the slopes of the two trends are similar, there is more room for variation with ε (records are typically selected to be within 0.5 magnitude units of the target value obtained from PSHA, but using zero-epsilon records in place of 1.5-epsilon records is not uncommon). Both the statistical significance and this potential prediction error are relevant, and jointly they suggest that ε, and to a lesser extent magnitude, should be considered when predicting the response of a structure.

Note that the trend with ε would be more difficult to discern if we had not already scaled the records to $S_a(T_1)$. This is because ε and $S_a(T_1)$ tend to be correlated, making it more difficult to separate effects due to $S_a(T_1)$ and effects due to ε. By first scaling to $S_a(T_1)$ we have eliminated this problem (termed "collinearity", Neter et al. 1996) and allowed the effect of ε to be seen more clearly.

Although Table 3.1 appears to show empirically that ε has an effect on structural response, this conclusion would be much more convincing if supported by an intuitive understanding as to why ε might matter. This understanding is developed in the following section.

3.9 Why does epsilon affect structural response?

When considering why ε could affect structural response, we should consider current understanding about nonlinear response of multiple-degree-of-freedom structures. The first parameter of our *IM*, $S_a(T_1)$, provides the response of a linear single-degree-of-freedom structure with a period of vibration approximately equal to the first-mode period of the MDOF structure under consideration. Given $S_a(T_1)$, the shape of the response spectra is known to be a significant factor in the response of nonlinear MDOF structures (e.g., Kennedy et al. 1984, Baker and Cornell 2004). This is because the response of an MDOF structure is also affected by excitation of higher modes of the structure at periods shorter than T_1 (as implied, e.g., by response spectrum analysis, Chopra 2001). In addition, S_a at periods longer than T_1 affects nonlinear structures, because as the structure starts behaving nonlinearly, the effective period of its first mode increases to a period larger than T_1 (Iwan 1980, Kennedy et al. 1985). Thus, given two records with the same $S_a(T_1)$ value, the record with higher S_a values at periods other than T_1 will tend to cause larger responses in a nonlinear MDOF system. So given $S_a(T_1)$, we would like to know about S_a at other periods. We could do this by measuring that S_a at other periods directly (Baker and Cornell 2004, Cordova et al. 2001, Vamvatsikos 2002). Alternatively, we will see that ε is a convenient implicit measure of spectral shape.

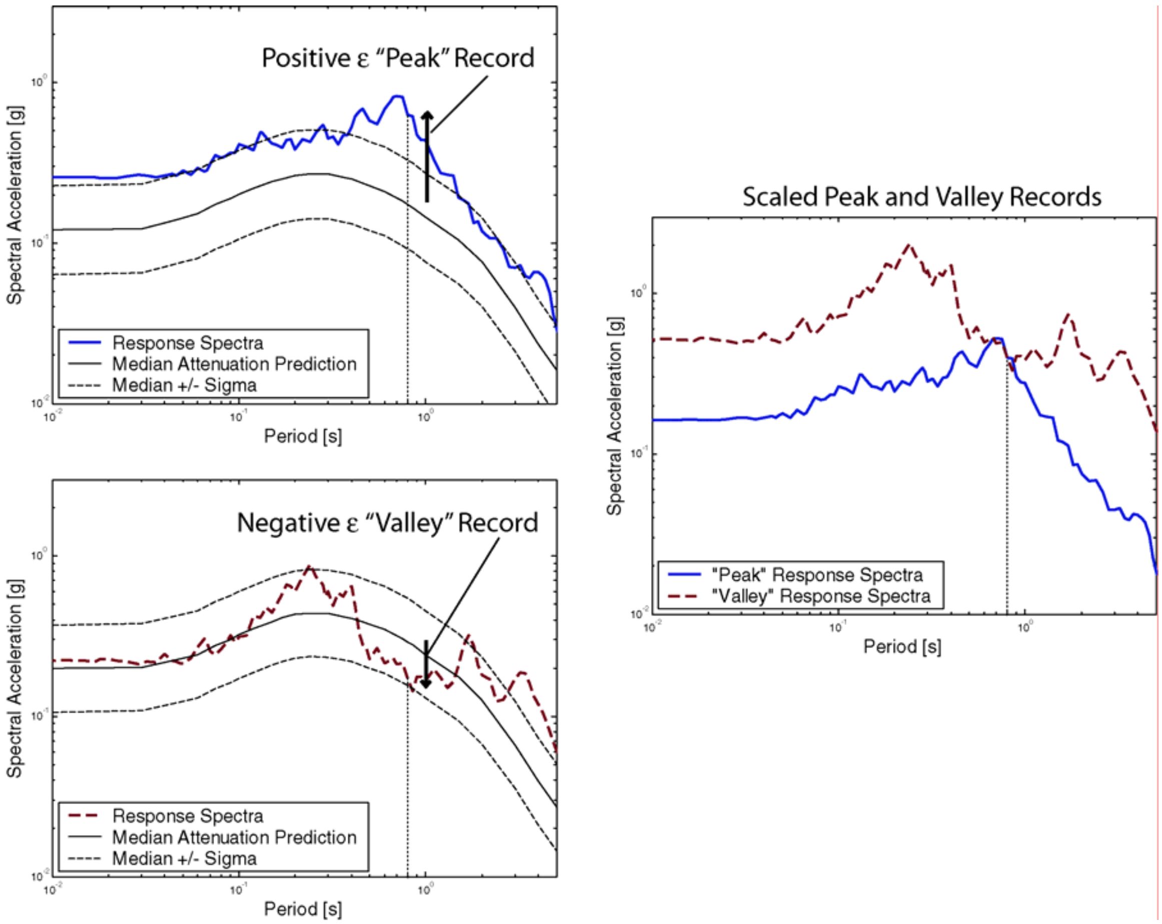


Figure 3.3. Scaling a negative ε record and a positive ε record to the same $S_a(T_1)$: an illustration of the peak and valley effect. In this case, T_1 = 0.8 seconds.

A record with a positive ε value is one that has a larger-than-expected spectral acceleration at the specified period. But what does it tell us about the spectral acceleration at other periods? It may be that the record is stronger than expected at all periods, or it may be that the record is stronger than expected in only a nearby range of periods, and that the spectral acceleration values at other periods are not as strongly related. We are interested in the possibility that only a narrow range of periods have comparatively large S_a values. We term this a record with a spectral "Peak" at $S_a(T_1)$. Conversely, a record that is lower than expected in only a narrow range of periods has a spectral "Valley" at $S_a(T_1)$. Now consider scaling a record with a peak and a record with a valley to the same $S_a(T_1)$ level. At T_1, the two records will have the same spectral acceleration by construction, but at other periods the valley record will tend to have larger spectral accelerations than the peak record. This is seen by examining the two sample response spectra shown in Figure 3.3. If a record has a peak or a valley at the period considered, then ε (which measures deviation from expected spectral values) may be an indicator of this condition, and if so it would be useful for predicting structural response.

3.9.1 The effect of epsilon, as seen using a second-moment model for logarithmic spectral acceleration

Anecdotal evidence of ε indicating a spectral peak or valley is seen in Figure 3.3, but there is more concrete evidence to show the connection between ε and spectral shape. We note that for a given magnitude, distance, site classification and faulting mechanism, logarithmic spectral acceleration at a given period is a random variable with a mean and standard deviation specified by a ground motion prediction equation (Abrahamson and Silva 1997). Lines indicating the mean value ± one standard deviation at all periods for a scenario event are shown in Figure 3.4a. Now consider logarithmic spectral accelerations at two periods simultaneously. We can obtain the means and standard deviations from the ground motion prediction equation. In order to completely specify the first and second moments of this pair, we also need to know the correlation between $\ln S_a$ values at the two periods. An empirically determined relationship for this correlation is given by Inoue and Cornell (1990):

$$\rho_{\ln Sa(T_1),\ln Sa(T_2)} = 1 - 0.33\left|\ln(T_1 / T_2)\right| \qquad 0.1s \le T_1, T_2 \le 4s \tag{3.12}$$

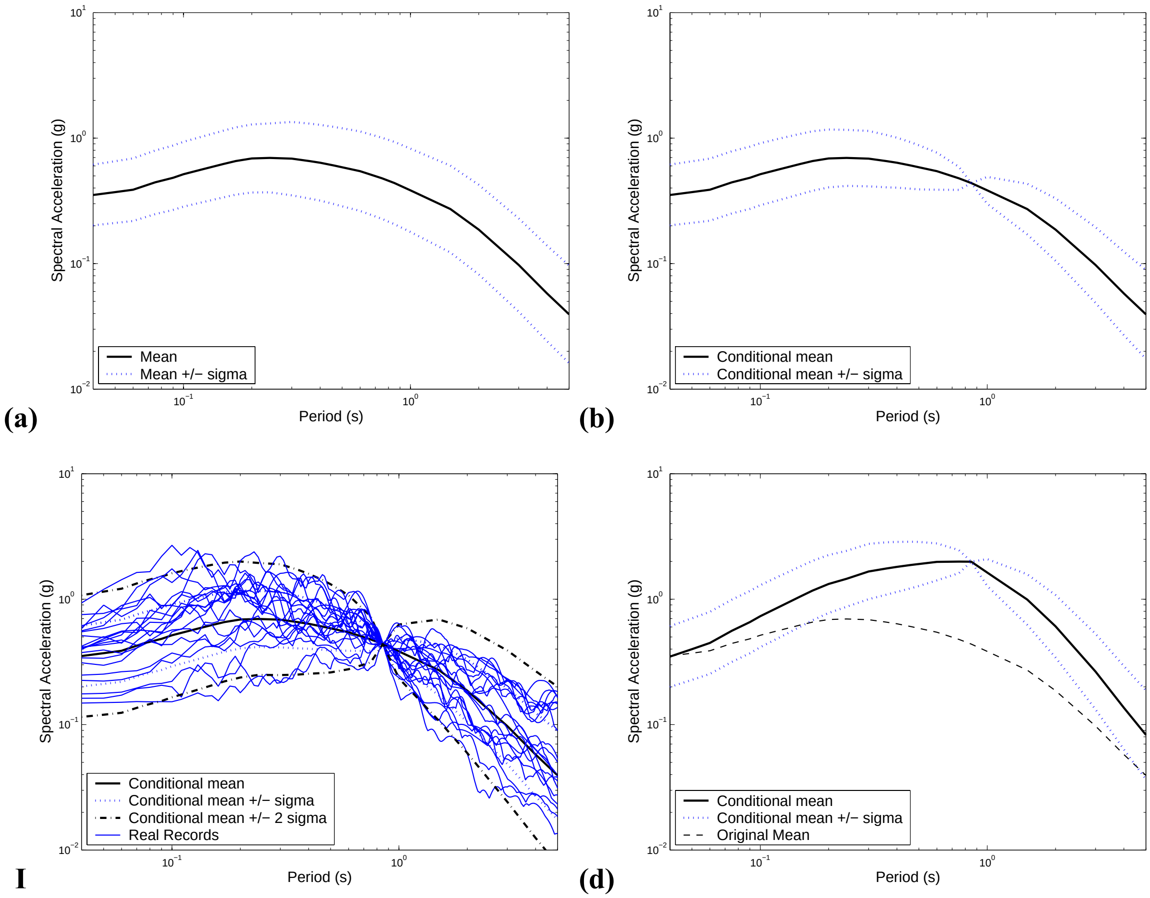


Figure 3.4. The mean value and the mean ± sigma values of $\ln S_a$ for a Magnitude = 6.5, Distance = 8 km event: (a) unconditioned values; (b) conditioned on $\ln S_a(0.8\text{s})$ equal to the mean value of the ground motion prediction equation; (c) conditioned on $\ln S_a(0.8\text{s})$ equal to the mean value of the ground motion prediction with actual response spectra superimposed; (d) conditioned on $\ln S_a(0.8\text{s})$ equal to the mean value of the ground motion prediction plus two standard deviations.

This correlation coefficient approaches one when the two periods are nearly equal, and decreases as the periods get further apart from each other. We have now fully defined the mean and covariance of this pair of response spectra values. Note that for expository purposes, we will later assume that this equation is valid over the range $0.05\text{s} \le T_1, T_2 \le 5\text{s}$.

Previous research has established that $\ln S_a(T_1)$ and $\ln S_a(T_2)$ are each marginally normally distributed (Abrahamson 1988, 2000b). Under the mild assumption that they are jointly normally distributed, we obtain the conditional mean of $\ln S_a(T_2)$, given $\ln S_a(T_1)$, as given in Equation 3.13:

$$\begin{aligned}\mu_{\ln Sa(T_2)|\ln Sa(T_1)=x} &= \mu_{\ln Sa(T_2)} + \rho_{\ln Sa(T_1),\ln Sa(T_2)} \cdot \sigma_{\ln Sa(T_2)} \left(\frac{x - \mu_{\ln Sa(T_1)}}{\sigma_{\ln Sa(T_1)}} \right) \\ &= \mu_{\ln Sa(T_2)} + \rho_{\ln Sa(T_1),\ln Sa(T_2)} \cdot \sigma_{\ln Sa(T_2)} \cdot \varepsilon_{T_1}\end{aligned} \tag{3.13}$$

where ε_{T_1} is the ε value of the record at the period T_1 (this equation is derived from the result $E[\varepsilon_{T_2} \mid \varepsilon_{T_1}] = \rho_{\ln Sa(T_1),\ln Sa(T_2)} \cdot \varepsilon_{T_1}$). We see that the conditional mean of $\ln S_a(T_2)$ is shifted up if $\varepsilon_{T_1} > 0$ or down if $\varepsilon_{T_1} < 0$. We can also obtain the conditional standard deviation of $\ln S_a(T_2)$:

$$\sigma_{\ln Sa(T_2)|\ln Sa(T_1)=x} = \sigma_{\ln Sa(T_2)} \sqrt{1-\rho^2_{\ln Sa(T_1),\ln Sa(T_2)}} \tag{3.14}$$

We see that, as might be expected, the conditional standard deviation of $\ln S_a(T_2)$ is reduced as the correlation increases between $\ln S_a(T_1)$ and $\ln S_a(T_2)$. In Figure 3.4b, for $T_1 = 0.8$ sec, we condition on $\ln Sa(0.8s) = \mu_{\ln Sa(T_1)}$ and plot the mean and mean ± sigma for a scenario event (for each value of T, we use the conditional mean and standard deviation from Equations 3.13 and 3.14). This plot represents the theoretical distribution of the response spectra for all records with $\ln Sa(0.8s) = \mu_{\ln Sa(T_1)}$.

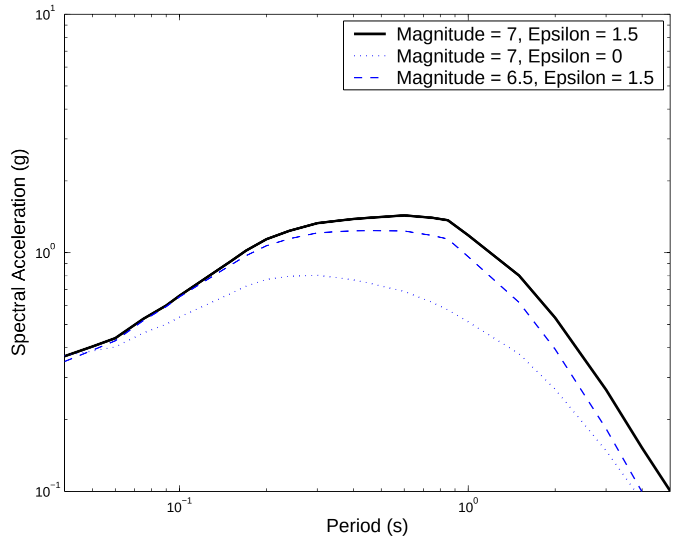
We see in Figure 3.4b that there is less dispersion in S_a for periods close to T_1, but for periods at some distance from T_1 there is little reduction in dispersion gained by conditioning on T_1. While it is not strictly a verification, we can confirm that this multivariate distribution model is representative of reality by scaling real records to $\ln S_a$(0.8s) and superimposing them over the model distribution. We see in Figure 3.4c that the real records match the model reasonably well (i.e. the mean value of the records is close to the predicted mean value, and the model prediction that 95% of the records should fall between the mean $\pm$ two sigma is reasonable).

In Figure 3.4d we plot the expected S_a value and +/- sigma, but now conditioned on $\ln Sa(0.8s) = \mu_{\ln Sa(T_1)} + 2\sigma_{\ln Sa(T_1)}$ (i.e. an "ε = 2" record). The original mean is also plotted as a reference. In this figure, we see that for periods near to T_1 we have larger spectral values than originally expected, but as the period gets much larger or smaller than T_1, the expected value of $\ln S_a(T)$ goes back towards the original mean value, reflecting Equations 3.12 and 3.13.

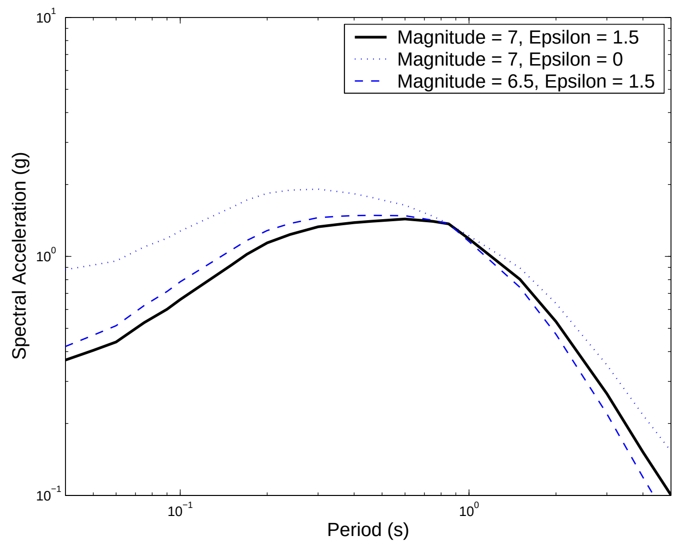
The results of this analysis can be restated in words as follows. A record with a positive ε has a higher than expected S_a value at the specified period. But S_a values are not perfectly correlated, so a higher-than-average value at one period does not imply correspondingly higher-than-average values at all periods—in fact, the conditional expected values of S_a at other periods tend back towards the marginal expected value. Thus, records with positive ε values tend to have peaks in the response spectrum at the specified period, and records with negative ε values tend to have valleys. Therefore, ε is an indicator of spectral shape, and this is why it is effective in predicting the response of nonlinear MDOF models.

3.9.2 Consideration of other candidate IM parameters

It has been shown that ε is an indicator of spectral shape, supporting its utility in predicting nonlinear response. But the question naturally arises, is it the *best* predictor of nonlinear response? Other candidates have been considered. Magnitude has often been considered as an indicator of spectral shape as discussed above, but it effectively has a weaker relationship with spectral shape than ε. We demonstrate next via Figure 3.5 this difference. Consider a target event with $M = 7$ and $\varepsilon = 1.5$. Records with these parameters are rare; there are few recordings available from large magnitudes, and of those recordings, only 7% are expected to have an $\ln S_a(T_1)$ at least 1.5 standard deviations larger than the mean. Suppose instead we have two records available for analysis: an $M = 7$, $\varepsilon = 0$ record and an $M = 6.5$, $\varepsilon = 1.5$ record (i.e., we can match either magnitude or ε, but not both). The expected spectra for these two records are shown in Figure 3.5a, along with the expected spectrum for the target event. In Figure 3.5b, the three "records" have been scaled so that their $S_a(T_1)$ values match. We see that the record with the correct ε and incorrect magnitude comes closer to matching the target spectral shape than does the record with the correct magnitude but incorrect ε. So a difference of 0.5 units in magnitude makes much less difference in spectral shape than the difference of 1.5 units of ε. Therefore it is anticipated that ε will prove more effective than M. This conclusion will be confirmed below.



(a)



(b)

Figure 3.5. (a) Expected response spectra for three scenario events; (b) expected response spectra for three scenario events, scaled to have the same $S_a(0.8s)$ value.

As mentioned earlier, it is also possible to measure spectral shape in a direct manner by considering spectral acceleration at an additional period as a second parameter in the vector *IM*. In previous work, the drift estimation method presented in this paper has been used with such a direct measure of spectral shape. With an informed choice of the second period, the direct measure of spectral shape has been shown to be an effective *IM* as well (Baker and Cornell 2004). The comparative advantage of ε is that it is easy to find the joint distribution of S_a and ε from PSHA disaggregation, with no need for specialized hazard analysis software. Further consideration of spectral shape is outside of the scope of this paper, but it should also be considered a promising candidate for a vector-valued *IM*.

Spectral acceleration averaged over a range of periods may be an effective *IM* in the case where a structure is sensitive to several periods, and may reduce the peak/valley effects seen when using spectral acceleration at only a single period (Shome 1999). The disadvantage is that when structural response is predominantly governed by a single period, averaging spectral values over a range of periods will tend to reduce the efficiency of response predictions.

One might also wonder if there is a better epsilon-based measure of spectral shape. For example, it is possible that a record could have a positive ε at the period considered, but also have large ε values at all other periods. In this case, the positive ε value would incorrectly suggest the presence of a peak and thus it would incorrectly predict the relatively smaller level of response we have seen to be typically associated with positive ε values. It would seem plausible that more sophisticated "epsilon-type" parameter might do a better job than one considering simply ε itself. The authors have tried several approaches to develop an improved epsilon-based measure. A measure separating the inter-event epsilon from the intra-event epsilon (see Abrahamson and Silva 1997) was attempted, anticipating that the intra-event epsilon would tend to represent peaks and valleys, while the inter-event epsilon would reflect the overall increase or decrease in the response spectra already accounted for by $S_a(T_1)$. Measures attempting to separate "global" shape effects from "local" shape effects (see Carballo 2000) were also examined. However, neither of these measures showed significant improvement over ε, and thus were rejected in favor of the simpler ε parameter.

For these reasons, the ε predictor is believed to have advantages over alternative response predictors, making it an interesting candidate as an effective intensity measure.

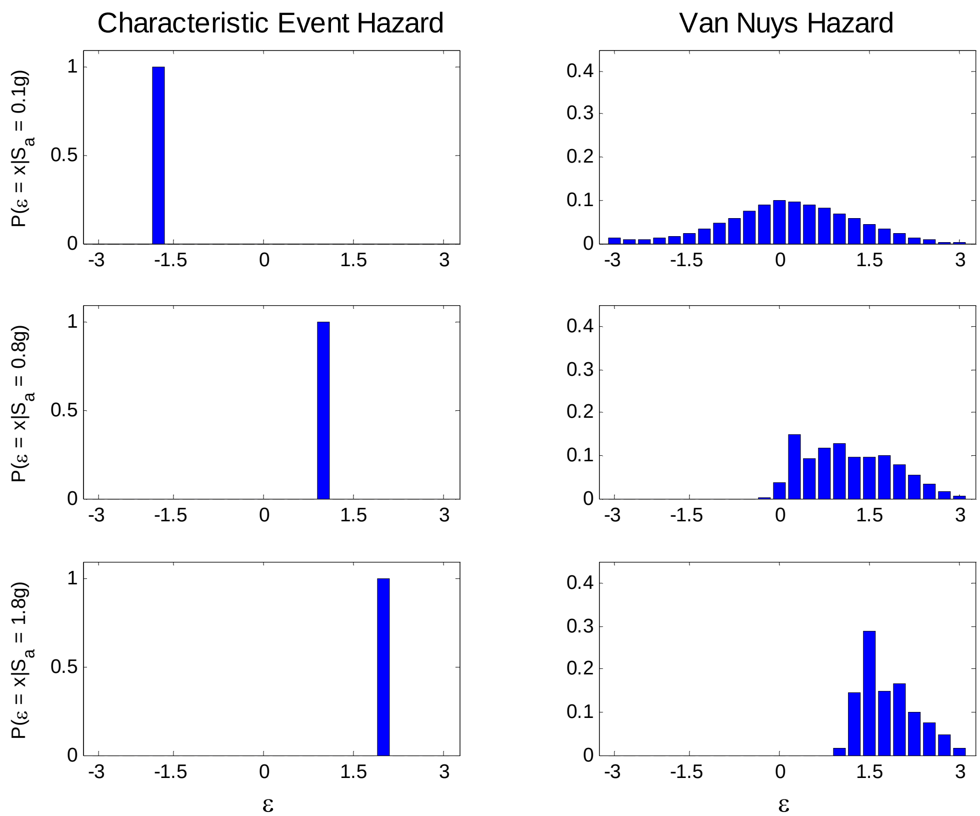


Figure 3.6. Disaggregation of PSHA results. The conditional distribution of ε given $S_a(0.8\text{s})=x$ is shown for both fault models at three different hazard spectral acceleration levels associated with three different mean annual frequencies of exceedance.

3.9.3 Epsilon and ground motion hazard

Having established ε as an effective predictor of structural response supplemental to $S_a(T_1)$, it is useful to consider what values of ε are typically to be anticipated in ground motions that are of interest to structural engineers. For a given site and a given fault, the three parameters that can vary are magnitude, distance, and ε. In a Probabilistic Seismic Hazard Analysis, the possible values of these three parameters and their likelihoods are integrated over for each fault, and the hazard contributions of all faults are summed (Kramer 1996). The result is a curve specifying the annual rate of exceeding varying levels of ground motion intensity. For purposes of illustration, consider a hypothetical site that has a single fault, capable of producing only magnitude 6.5 events at a distance of 8 km. At this site, magnitude and distance are fixed for all events, so ε is the only free variable in the PSHA analysis; thus the effect of ε can be more clearly seen. A single event model is also quite representative of the ground motion hazard situation at many sites located near a single large fault (i.e., some urban locations near the San Andreas or Hayward faults in northern California). This model will be referred to as the "Characteristic Event model."

With this model, the median value of spectral acceleration (i.e., the exponential of the mean of $\ln S_a$) is 0.46g, so any ground motions larger than 0.46g have a positive ε value. Thus, at low annual frequencies (where S_a values greater than 0.46g are seen) the ground motion hazard is governed exclusively by records with positive epsilons. This is seen in Figure 3.6: as the ground motion level is increased, the ε value seen in the disaggregation shows a corresponding increase.

Now consider the ground motion hazard at a real site surrounded by several faults, each of which is capable of producing events with a variety of magnitudes and distances. The hazard assessment used here is that for Van Nuys, California, at the site of the example structure which will be analyzed below. (The activity rate of the Characteristic Event model is specified such that both of these models have the same ground motion level at the 10% in 50 year hazard: $S_a(0.8s)$ = 0.6g. In addition, the magnitude and distance values for the Characteristic Event equal the mean values of the Van Nuys disaggregation at the 10% in 50 year level.) The hazard there is simply a summation of many hazard contributions each similar to the Characteristic Event model. There is a limit on the maximum magnitude and minimum distance of an event, so it is still true that large ground motion events will have positive ε values. In fact, as noted earlier, we can determine the ε values at a given hazard level by examining the disaggregation of the PSHA results. The disaggregation-based distribution of ε at the Van Nuys site is shown in Figure 3.6 for several levels of S_a. As the annual rate of exceedance decreases (i.e., as the ground motion level increases) the epsilons contributing to the hazard are seen to shift to larger values. Thus in general, for any hazard environment, the ground motion hazard at long low annual frequencies will be dominated by positive ε events.

3.10 Epsilon and drift hazard

In the previous section, it was established that at ground motion intensity levels with low annual rates of exceedance, records tend to have positive ε values. Further, we have seen that for a given S_a value, records with larger ε values tend to cause smaller responses in structures, due to the fact that ε is an indicator of peaks and valleys in the response spectra at the period of interest. But a typical random sample of records would have an average ε value of zero. Now consider the usual practice of using simply a scalar S_a with a suite of (on average) zero-epsilon records. When these records are used to estimate the response of a structure at (low annual frequency) high ground motion levels (typically characterized by positive epsilons), estimates of the frequency of exceeding large structural drifts are likely to be conservatively biased. This expectation can be confirmed by using the drift hazard procedure outlined earlier. The traditional approach using Equation 3.1 and $S_a(T_1)$ alone as the intensity measure is referred to here as the "scalar-based approach." The alternative approach using a vector-valued *IM* consisting of $S_a(T_1)$ and M or ε, and using Equation 3.7, is referred to as the "vector-based approach." A complete drift hazard curve is computed for a variety of structures using both of these approaches, and it is seen that neglecting to account for the ε value of a record nearly always results in overestimation of the mean annual rates of exceeding large levels of drift.

3.10.1 Description of the structures analyzed

The primary structure analyzed is a reinforced-concrete moment-frame building. The building has 1960's era construction and is serving as a test-bed for PEER research activities 2004. The actual structure is located in Van Nuys, CA, at the same site for which the ground motion hazard assessment above was conducted. A 2D model of the transverse frame created by Jalayer (2003) is used here. This model has a first-mode period of 0.8 seconds, and contains nonlinear elements that degrade in strength and stiffness, both cyclically and in-cycle (Pincheira et al. 1999). Forty historical earthquake ground motions from California are used to analyze this structure. The events range in magnitude from 5.7 to 7.3, and the recordings are at distances between 6 and 36 km. Directivity effects are not expected at this site (Van Nuys Testbed Report, 2004), so these effects were avoided by choosing records with small distances only when the rupture/site geometry suggested that near-fault effects would be unlikely; the ground motion velocity histories were not observed to contain pulse-like intervals. Epsilon was not calculated before the records were selected—therefore they have been effectively randomly selected with respect to ε. The set of 40 records was then scaled to 16 levels of $S_a(T_1)$ between 0.1g and 2.4g.

To supplement the data from this structure, a series of generic frame structures was evaluated as well. Fifteen generic frame models were analyzed, with a variety of configurations, periods, and degradation properties. The specific model parameters are summarized in Table 3.2. All of the structures are single-bay frames, with stiffnesses and strengths chosen to be representative of typical structures. Five structural configurations were considered, with varying numbers of stories and first-mode periods. A set of non-degrading models designed and analyzed by Medina and Krawinkler (2003) was considered. These models do not have degrading elements, but the more flexible structures still have the potential to collapse due to P-Δ effects. A set of degrading models designed and analyzed by Ibarra (2003) was also considered. These structures are identical to the models of Medina and Krawinkler, except for incorporation of elements that degrade in stiffness and strength. For each of the five building configurations, a non-degrading model and two degrading models were considered. A second set of forty records were used by those authors to analyze these structures, ranging in magnitude from 6.5 to 6.9, and ranging in distance from 13 to 40 km. It is difficult to generalize conclusions to all possible structures, but it is believed that by considering this wide range of models, the consistent effect of ε is apparent.

Table 3.2. Percent change in mean annual collapse rate and in the 10% in 50 year drift demand on a series of structural models when using the improved vector-based procedure versus the scalar-based procedure.

Number of Stories	Period (s)	Degradation Model	Reduction in mean annual rate of collapse	Reduction in drift at 10% in 50 year hazard
3	0.3	a	55%	93%
3	0.3	b	58%	91%
3	0.3	none	0%	17%
3	0.6	a	40%	*n.a.*
3	0.6	b	51%	73%
3	0.6	none	99%	25%
7	0.8	c	43%	*n.a.*
9	0.9	a	72%	42%
9	0.9	b	80%	41%
9	0.9	none	99%	32%
9	1.8	a	5%	18%
9	1.8	b	49%	20%
9	1.8	none	71%	25%
15	3.0	a	41%	5%
15	3.0	b	40%	*n.a.*
15	3.0	none	46%	13%

Degradation models:

a) Ibarra degradation parameters: peak oriented model,
$\delta_c / \delta_y = 4$, $\alpha_c = -0.10$, $\alpha_s = 0.03$, $\gamma_{s,c,k,a} = \infty$

b) Ibarra degradation parameters: peak oriented model,
$\delta_c / \delta_y = 4$, $\alpha_c = -0.05$, $\alpha_s = 0.03$, $\gamma_{s,c,k,a} = 50$

c) Pincheira model (1999), with parameters calibrated to reflect concrete connections with representative detailing

3.10.2 Drift hazard results

The first case considered is the reinforced concrete frame structure. This structure was evaluated for both the Characteristic Event and the Van Nuys ground motion hazard environments. The results are shown in Figure 3.7. We see that in both cases, versus the scalar approach, inclusion of

ε in the intensity measure results in lower mean annual frequencies for high levels of drift. While inclusion of M with S_a has the same effect, it is much less pronounced, as anticipated.

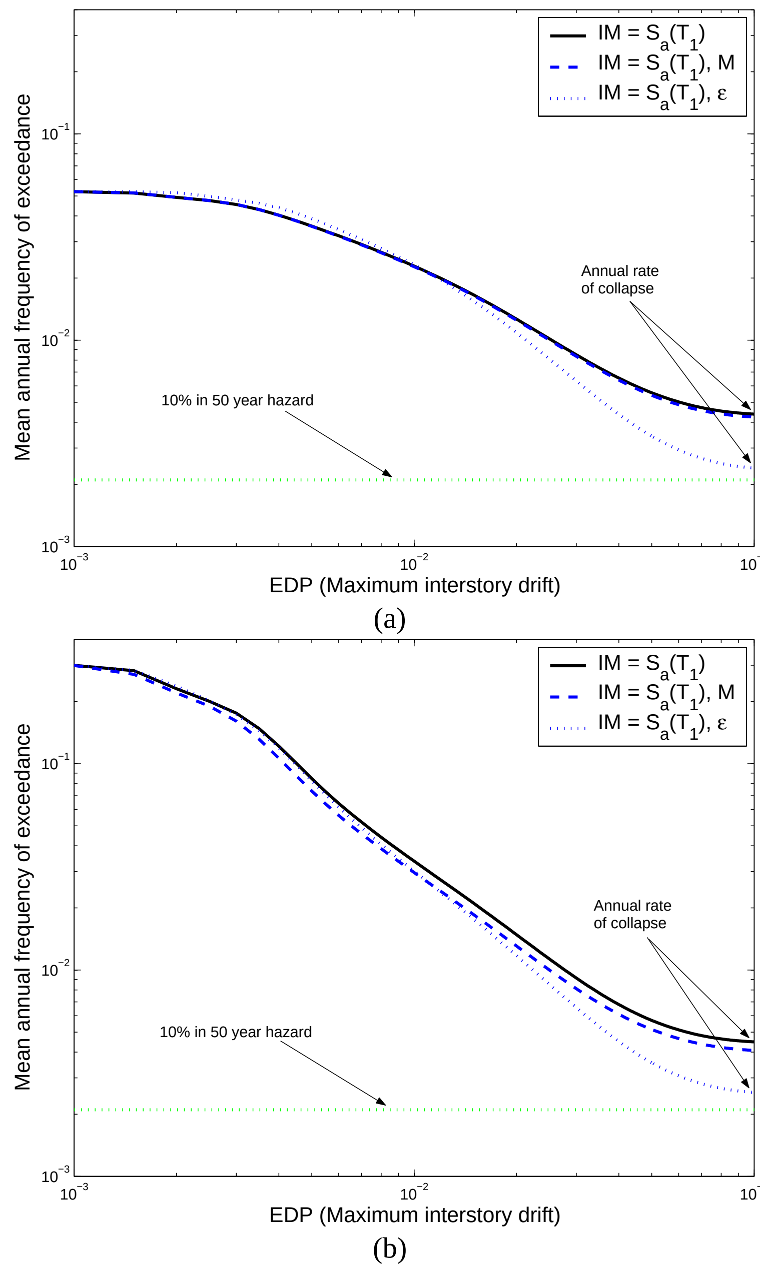


Figure 3.7. Seven story reinforced concrete frame. Mean annual frequency of exceedance versus maximum interstory drift. (a) Scalar-based and vector-based drift hazard curves for the Characteristic Event hazard; and (b) Scalar-based and vector-based drift hazard curves for the Van Nuys hazard.

This same procedure was repeated for the 15 generic frames considered. The drift hazard curves look similar to those of Figure 3.7. In order to present the results in a concise way, two important values for each structure are considered: the annual rate of exceeding 10% maximum interstory drift ratio (this amount of drift could be interpreted to indicate collapse of the structure), and the drift that has a 10% chance of being exceeded in 50 years (i.e., the drift level z such that $\lambda(EDP \geq z) = 0.0021$). The percentage change in these values between the scalar-based (S_a) result and the vector-based (S_a, ε) result is listed in Table 3.2. For some cases, the structure collapses with a greater than 10% probability in a 50 year period (e.g., this is seen in Figure 3.7). In these cases, there is no drift level associated with the 10% in 50 year hazard level, so the corresponding cell in Table 3.2 is marked "n.a." (not applicable). In addition, the non-degrading structure with a period of 0.3 seconds is not predicted to collapse at the ground motion levels present at the site, so there is no change in the frequency of collapse. We see that in every case (besides these special cases) the improved vector-based procedure produces lower demands on the structure. One interesting insight: it can be noted that the collapse rate reduction is typically less for the longer period version of otherwise similar models; the absolute rates are, however larger in such cases implying smaller values of ε, and hence less effect.

3.11 Discussion

It has been shown that ε is an implicity indicator of the shape of a record's response spectra, and thereby an important effect on the response of nonlinear MDOF models. This being the case, it is desirable to account for the effect of ε when predicting the annual rate of exceeding a given drift level. There is more than one way to accomplish this. One approach is to consider ε values carefully when selecting ground motions to use in analysis. In current practice, ground motions are selected so that they match the magnitude and distance values of the events that dominate the disaggregation, and the soil conditions present at the site under consideration (Stewart et al. 2001). This paper suggests that ε should also be considered when selecting ground motions. Unfortunately, the existing library of recorded ground motions is not large enough that all desired parameters can be matched simultaneously, especially for a sample of nominal size. In light of the relatively greater effect that ε has on structural response, the desire to closely match distance and magnitude should probably be relaxed in favor of matching ε levels. For example, one could match ε while also trying to match magnitude, but allow records from a wide range of distances to be used. When selecting records to match ε, one needs to remain aware that target ε values will increase as the mean annual frequency considered decreases (the target magnitude and distance may also change as the annual frequency decreases, and this is sometimes accounted for in the selection process, e.g., Stewart et al. 2001).

The method to address the effect of ε proposed here is to adopt an *IM* that accounts for the effect of ε (i.e., an *IM* that is sufficient with respect to ε, Luco and Cornell 2005). The most obvious *IM* that accomplishes this is the vector-valued *IM* presented in this paper, with $S_a(T_1)$ and ε as parameters. One first assesses the dependence of drift on ε at one or more levels of S_a, as described herein. One level at or near the mean annual frequency of interest (e.g., 2% in 50 years) may be sufficient in some cases. The sample size can probably be limited to the order of 10 records, especially if record selection is designed to capture the dependence of ε (note that with this approach, extreme positive *and* negative ε values in the suite of records are desirable in order to improve the slope fit, in contrast to the previous approach where only records with specific ε values are desired). Then the drift hazard curve is computed using the vector-based procedure of Equation 3.7, as was done to produce Figure 3.7.

Finally, there is a possibility that alternative scalar intensity measures may account for the effects of ε more sufficiently than spectral acceleration (for instance, some of the improved scalar *Im*s that have been proposed recently Cordova et al. 2001, Luco and Cornell 2005, Mori et al. 2004). If this were the case, then ε would not need to be included as a second parameter of the *IM*. However, this hypothesis remains to be tested. In the meantime, the most direct way to address the observed effect of ε is to use the vector-valued *IM* presented in this paper.

3.12 Conclusions

A method for calculating the probabilistic response of structures with a vector-valued *IM* is used to evaluate the significance of magnitude, distance and ε on the response of structures, conditioned on spectral acceleration. It is seen that ε has a significant effect on the response of structures, because it is an indicator of spectral shape (more specifically, it tends to indicate whether S_a at a specified period is in a peak or a valley of the spectrum). For a fixed $S_a(T_1)$, records with positive ε values tend to cause smaller demands in structures than records with negative ε values. The effect of ε on structural response given $S_a(T_1)$ is seen to be greater than the effect of magnitude or distance.

In addition, by examining disaggregation of the ground motion hazard, it is seen that at low mean annual frequency of exceedance the ground motions are all positive-epsilon motions. Therefore, the practice of scaling up zero-epsilon (on average) records to represent records with positive epsilons is likely to result in overestimation of the demand on the structure. This will lead to overestimation of drift at a given hazard level, or overestimation of the mean annual frequency of collapse.

A vector-valued *IM* consisting of $S_a(T_1)$ and ε has been proposed in this paper. The proposed *IM* will account for the effect of ε on structural response. Alternatively, one could correct for the effect of ε on structural response by intelligently choosing records that have the proper ε value, and then using $S_a(T_1)$ as a scalar *IM*.

The results presented here are based on a suite of structural models that are similar in behavior to many frame structures. However, these models are by no means representative of all classes of structures in existence. It is believed that the effect of ε will be seen in a broad class of structures, but further work is needed to further confirm and quantify this effect.

Chapter 4

A Vector-Valued Ground Motion Intensity Measure Consisting of Spectral Acceleration and a Measure of Spectral Shape

4.1 Abstract

Among the challenges of performance-based earthquake engineering is estimation of the demand on a structure from a given earthquake. This is often done by calculating the distribution of structural response as a function of the ground motion intensity (as measured by an "Intensity Measure," or *IM*), and then combining this with the rates of exceeding various *IM* levels, as quantified by the probabilistic ground motion hazard. Traditional *Im*s include peak ground acceleration and spectral acceleration at the first-mode period of vibration. These *Im*s consist of a single parameter. In this chapter, we investigate *Im*s consisting of two parameters: spectral acceleration at the first-mode period of vibration along with a measure of spectral shape (the ratio of spectral acceleration at a second period to the original spectral acceleration value). A method for predicting the probability distribution of demand using a vector *IM* is presented that accounts for the effect of collapses on the distribution of demand. Two complimentary methods for determining the optimum second period at a given intensity level are described, and an improvement in the efficiency of demand predictions is shown. Finally, the parameter ε is included, to create a three-parameter vector. It is seen that although the spectral shape parameter has increased the efficiency of response predictions, it does not fully account for the effect of ε. Thus, ε should still be accounted for in analysis, either through informed record selection or by using this larger vector of *IM* parameters.

4.2 Introduction

In this paper we present a method for determining an optimal vector "Intensity Measure" for predicting the demand in a structure, (defined as a measure of the structural response such as the maximum inter-story drift angle seen in the structure). To predict this response effectively, one needs an Intensity Measure (*IM*) for the given earthquake. In the past, the Peak Ground Acceleration (*PGA*) of the earthquake was commonly used as an *IM*. More recently, spectral response values (i.e., spectral acceleration at the first-mode period of vibration – $S_a(T_1)$) have been used as *Im*s. These *Im*s are generally a single parameter. In this paper, we propose intensity measures containing two or three parameters. These intensity measures are called "Vector *Im*s," as opposed to the "Scalar *Im*s" that contain only a single parameter. One would expect intuitively that a Vector *IM* would contain more information about the ground motion than a Scalar *IM*, and would thus be more effective at predicting the response of a structure. This will be shown, and criteria for finding an optimal set of parameters will be described.

An example of the need for prediction of structural response is seen in the work of the Pacific Earthquake Engineering Research (PEER) Center (Cornell and Krawinkler 2000). Here, the response of a structure is termed an Engineering Demand Parameter, or *EDP*. The annual frequency of exceeding a given *EDP* is calculated as follows

$$\lambda_{EDP}(z) = \sum_{\text{all } x_i} P(EDP > z \mid IM = x_i) \cdot \Delta\lambda_{IM}(x_i) \tag{4.1}$$

Where $\lambda_{EDP}(z)$ is the annual frequency of exceeding a given *EDP* value z, $\lambda_{IM}(x_i)$ is the annual frequency of exceeding a given *IM* value x_i (this is commonly referred to as a ground motion hazard curve), and $\Delta\lambda_{IM}(x_i) = \lambda_{IM}(x_i) - \lambda_{IM}(x_{i+1})$ is approximately the annual frequency of $IM = x_i$. The final element of this equation is $P(EDP{>}z|IM{=}x_i)$, the probability of exceeding a specified *EDP* level, given a level of *IM*. This is the value that we are interested in estimating efficiently using a vector *IM*. If the *IM* becomes a vector, Equation 4.1 must be generalized, as will be discussed below.

In this paper, we will consider only a specific class of candidates as potential vector *Im*s. Given that $S_a(T_1)$ has been verified as an effective predictor of structural response for a wide class of structures, we will continue to use this as the first element of our vector. For the second element of our vector, we will consider the predictor $R_{T1,T2} = S_a(T_2)/ S_a(T_1)$ (see Figure 4.1 for an illustration). The predictor $R_{T1,T2}$ is a measure of spectral shape. Together the vector $S_a(T_1)$ and $R_{T1,T2}$ define two points on the spectrum of an accelerogram. The first-mode elastic period of vibration of the building will always be used as T_1, but T_2 will be allowed to vary and chosen to optimally predict the response of the structure. (A range of T_1 values was examined at an early stage in the research, but at low to moderate levels of nonlinearity, other periods did not show a significant improvement over the first-mode period of the structure.) The spectral shape predictor $R_{T1,T2}$ has been found by others to be a useful predictor of structural response (e.g., Cordova et al. 2001, Vamvatsikos 2002). A two-element vector is the primary focus of this study, although a three element vector will be considered briefly as well. This study also considers only maximum interstory drift (referred to as θ_{max}) as an *EDP*, although an identical study could be performed on any other *EDP* value.

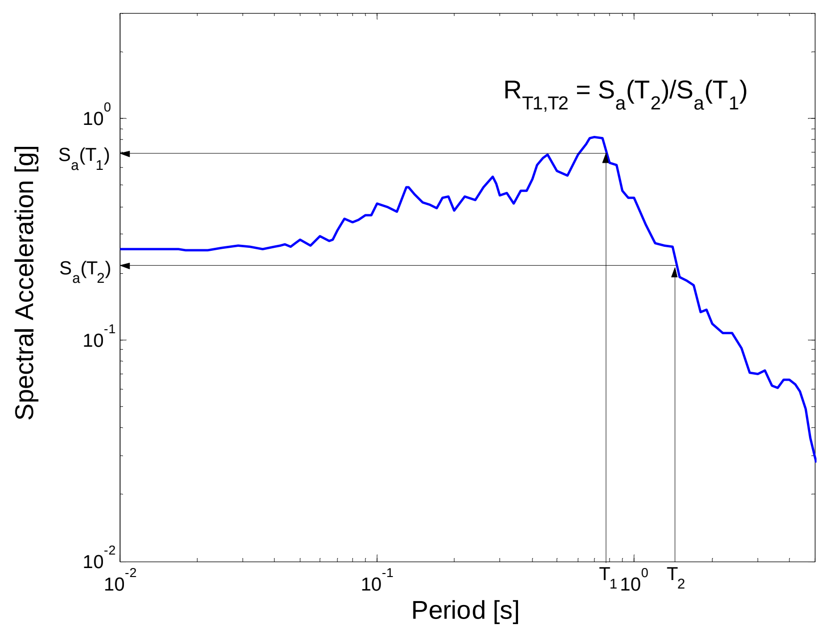


Figure 4.1: Illustration of the calculation of $R_{T1,T2}$ for a given response spectrum.

4.3 Prediction of building response using a scalar IM

The prediction of building response requires estimation of the term $P(EDP>z|IM=x_i)$ in Equation 4.1. This is true for both the scalar and vector *IM* cases. In this section, an estimation method is presented for the scalar *IM* case, and in the next section it is modified for use with vector *Im*s. The scalar *IM* $S_a(T_1)$ is used in this section, both because of its wide use elsewhere, and because it will be easily generalized to our vector case, $IM = (S_a(T_1), R_{T1,T2})$.

The method used in this paper requires a suite of earthquake of records, all at the same *IM* value, $S_a(T_1)=x$ (e.g., in this study, 40 records are used at each *IM* level). We scale a suite of historical earthquake records to the given $S_a(T_1)$ value (e.g., Shome et al. 1998). (We use the same suite of records for different $S_a(T_1)$ levels, although one could use different record suites at different levels if PSHA disaggregation suggested that, for example, the representative magnitude level was changing. This method will be explored in Chapter 6.) This suite of records is used to perform nonlinear dynamic analysis on a model of the structure. Now we have *n* records, all with *IM* = *x*, and *n* corresponding values of *EDP* (see an illustration in Figure 4.2). A perfect *IM* would be one such that all records with *IM* = *x* would have an identical value of *EDP*. However, with our less-than-perfect *Im*s, there will be a distribution of *EDP*'s. So in fact, *EDP* given *IM* = *x* is a random variable with an unknown distribution, and our *n* values of *EDP* are a sample from this distribution. We need to estimate this distribution in order to calculate the probability that *EDP* is greater than a given value. Our *EDP*, maximum interstory drift ratio, has been found to be well represented by a lognormal distribution (e.g. Shome 1999, Aslani and Miranda 2003). Because of this we work with the natural logarithm of *EDP*, which then has the normal distribution. We can estimate the parameters for this normal distribution using the method of moments (Benjamin and Cornell 1970). For each *IM* level, estimate the mean of ln*EDP* as the sample average of the ln*EDP*s, and denote this $\hat{\mu}_{\ln EDP|IM=x}$. Estimate the standard deviation of ln*EDP*, which we call the "dispersion" and denote $\hat{\beta}_{\ln EDP|IM=x}$, as equal to the sample standard deviation. These two parameters fully define the normal distribution. The probability that *EDP* exceeds *z* given *IM* = *x* can now be calculated using the Gaussian Complimentary Cumulative Distribution Function

$$P(EDP > z \mid IM = x) = 1 - \Phi\left(\frac{\ln z - \hat{\mu}_{\ln EDP|IM=x}}{\hat{\beta}_{\ln EDP|IM=x}}\right) \tag{4.2}$$

where $\Phi(\cdot)$ denotes the standard normal distribution.

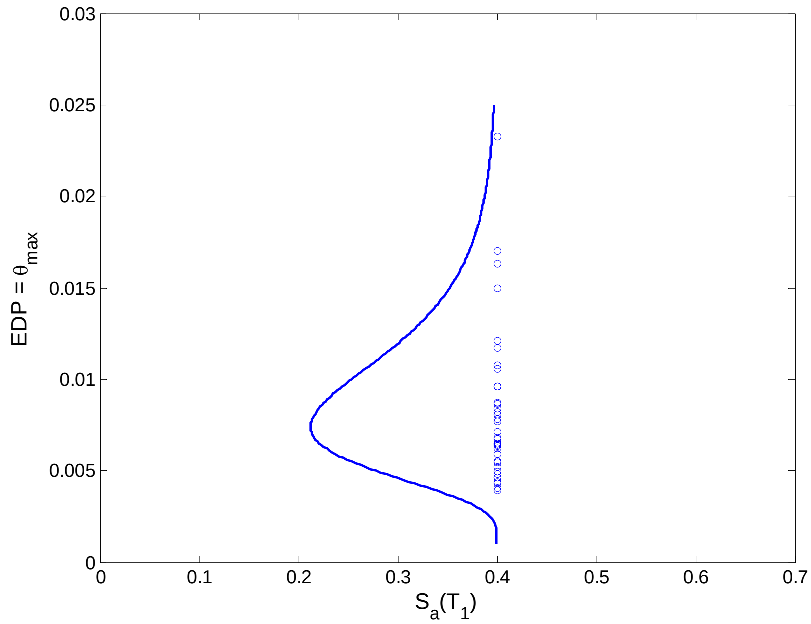


Figure 4.2: Structural response results from dynamic analysis using forty records scaled to $S_a(T_1) = 0.4g$, and a superimposed lognormal probability density function, generated using the method of moments.

4.3.1 Accounting for collapses

In the previous section, it was stated that *EDP* given $IM = x$ has a lognormal distribution. While this may be a valid assumption at lower ground motion levels, it does not explicitly account for the possibility that some records may result in a collapse of the structure at higher levels of *IM*. For these records, we would say that EDP_i is greater than z for any value of z (or equivalently, $EDP_i = \infty$). This causes two problems: the probability that $EDP_i = \infty$ is zero in the lognormal distribution, and collapses cause our estimates of the lognormal mean and standard deviation to be infinite. To address this issue, we need to make a modification to our procedure, as proposed by Shome and Cornell (2000). First, we separate our realizations of *EDP* into collapsed and non-collapsed data. We then estimate the probability of collapse at the given *IM* level as

$$P(C \mid IM = x) = \frac{\text{number of records causing collapse}}{\text{total number of records}} \tag{4.3}$$

We then use the method of moments to estimate $\mu_{\ln EDP|IM=x}$ and $\beta_{\ln EDP|IM=x}$ using *only* the non-collapsed records. Combining the two possibilities, our estimate of the probability that *EDP* exceeds z given $IM = x$ is now

$$P(EDP > z \mid IM = x) = P(C \mid IM = x) + \left(1 - P(C \mid IM = x)\right)\left(1 - \Phi\left(\frac{\ln z - \hat{\mu}_{\ln EDP|IM=x}}{\hat{\beta}_{\ln EDP|IM=x}}\right)\right) \tag{4.4}$$

We can now proceed with this estimate.

4.4 Prediction of building response using a vector IM

We now adapt the procedure of the preceding section for use with a vector *IM*. If we label the two elements of our vector *IM* as $IM_1 = S_a(T_1)$ and $IM_2 = R_{T1,T2}$, then we are trying to estimate $P(EDP > z \mid IM_1 = x_1, IM_2 = x_2)$. Ideally, we would like to scale our records to both $IM_1 = x_1$ and $IM_2 = x_2$. However, when we scale our record by a factor y, each spectral acceleration value is changed by the same factor y. Therefore, $R_{T1,T2} = S_a(T_2)/\ S_a(T_1)$ is unchanged by scaling. (In general, because there is only one factor that we are scaling by—a uniform scale factor on the entire record—we cannot match two *Im*s simultaneously, even if we consider other classes of IM_2 than $R_{T1,T2}$.) So we need a supplement to scaling.

The solution adopted here is to scale on IM_1 ($S_a(T_1)$) as before, and then apply regression analysis to estimate *EDP* versus IM_2 ($R_{T1,T2}$) for each IM_1 level. In Section 4.7, we will explain the reasoning behind scaling to IM_1, rather than using regression analysis on both variables (the advantages and disadvantages of a variety of methods are also discussed in depth in Chapter 2). Our approach will parallel the scalar *IM* case, in that we separate out the collapsing records first, and then deal with the remaining non-collapsed records.

4.4.1 Accounting for collapses with the vector IM

As with the scalar *IM*, there is a possibility that some records may cause collapse of the structure. However, instead of taking the probability of collapse to be simply the fraction of records that collapse, we would like to take advantage of our vector *IM* to predict the probability of collapse more accurately. We do this using logistic regression, which is commonly used to generate predictions for binary data (Neter et al. 1996). Each record has a value of $R_{T1,T2}$, which we denote more simply as R_i and use as our predictor variable. We then build an indicator variable for collapse (call this Y_i and set it to 1 if the record causes collapse and 0 otherwise). We then use the logit transform to predict collapse

$$Y_i = \frac{\exp(\beta_0 + \beta_1 R_i)}{1 + \exp(\beta_0 + \beta_1 R_i)} \tag{4.5}$$

where β_0 and β_1 are parameters to be estimated using our set of data points (Y_i, R_i). Logistic regression is a built-in function in many mathematical and statistical software packages. We then use this function (and our estimated parameters) to predict the probability of collapse, given $S_a(T_1)$ and $R_{T1,T2}$

$$P\left(C \mid S_a(T_1) = x_1, R_{T1,T2} = x_2\right) = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_2)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_2)} \tag{4.6}$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ denote the estimates of β_0 and β_1 obtained from regression on a dataset that has been scaled to $S_a(T_1) = x_1$ (i.e. $\hat{\beta}_0$ and $\hat{\beta}_1$ will be different for different values of $S_a(T_1)$). We now have a probability of collapse prediction that varies with $R_{T1,T2}$, rather than being constant as in the scalar case. An example of this data and a fitted logistic regression curve is presented in Figure 4.3.

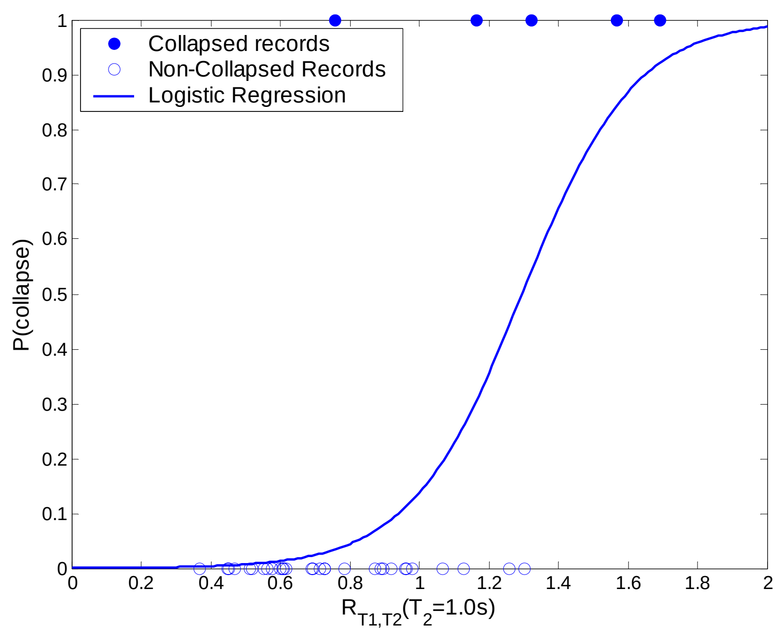


Figure 4.3: An example of prediction of the probability of collapse using logistic regression applied to binary collapse/non-collapse results ($S_a(T_1) = 0.9g$).

4.4.2 Accounting for non-collapses with the vector IM

Once the probability of collapse has been estimated, it is necessary to quantify the behavior of the non-collapse records. We now separate out the non-collapsed records. Note that each of these records has been scaled to $S_a(T_1) = x_1$. Each of the records has a value of $R_{T1,T2}$, which we again refer to as R_i, and it has a value of EDP, which we refer to as EDP_i. We have found that there tends to be a relationship between $R_{T1,T2}$ and EDP of the form $EDP \mid S_a(T_1), NC \approx a(R_{T1,T2})^b \tilde{e}$, where a and b are constant coefficients, and $\tilde{e}$ is a random variable representing the randomness in the relationship. This becomes a linear relationship after we take logarithms of both sides: $\ln EDP \mid S_a(T_1), NC \approx \ln a + b \ln R_{T1,T2} + e$ (in which a new random variable, $e = \ln \tilde{e}$, has been defined). Linear least-squares regression (Neter et al. 1996) is used to obtain estimates of the two regression coefficients, $\beta_2 = \ln a$ and $\beta_3 = b$

$$\ln EDP_i = \beta_2 + \beta_3 \ln R_i + e_i \tag{4.7}$$

Linear regression is available in many mathematical and statistical software packages, and can be used to obtain estimates of the coefficients, $\hat{\beta}_2$ and $\hat{\beta}_3$ (again, these values will vary for different $S_a(T_1)$ levels). A graphical example of this data and the regression fit is shown in Figure 4.4.

When using linear least-squares regression on a dataset, several assumptions are implicitly made, and the accuracy of the results depends on the validity of these assumptions. If we denote our predicted value of ln*EDP* for record *i* as $\ln E\hat{D}P_i$, then the prediction error for this record is called the "residual" from record *i*

$$e_i = \ln EDP_i - \ln E\hat{D}P_i \tag{4.8}$$

These residuals are assumed to be mutually independent. In addition, when estimating the distribution of ln*EDP* below, we will assume the residuals to be normally distributed with constant variance (this condition is termed homoscedasticity). The assumptions of independent normal residuals with constant variance have been examined for the data in this study, and found to be reasonable. An estimate of the variance of the residuals is also available from the analysis software, and we will denote it $V\hat{a}r[e] \equiv \hat{\sigma}_e^2$. This variance in the residuals is represented graphically in Figure 4.4 by superimposing the estimated normal distribution of the residuals over the data.

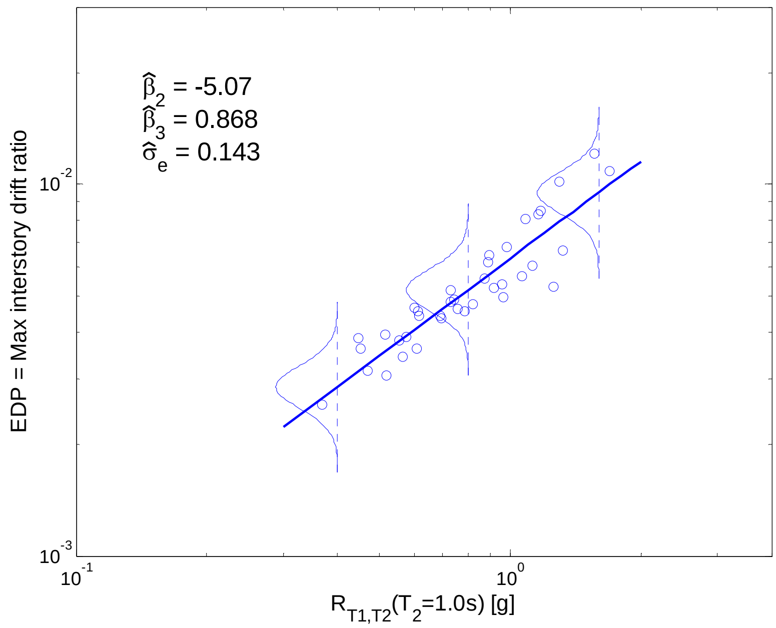


Figure 4.4: An example of non-collapse data, and a fit to the data using linear regression. The data comes from records scaled to $S_a(T_1) = 0.3g$. The estimated distributions of the residuals has been superimposed over the data.

From regression, we now know that given $S_a(T_1) = x_1$ and $R_{T1,T2} = x_2$, and given no collapse, the mean value of ln*EDP* is

$$E[\ln EDP \mid R_{T1,T2} = x_2] = \hat{\beta}_2 + \hat{\beta}_3 \ln x_2 \tag{4.9}$$

where $\hat{\beta}_2$ and $\hat{\beta}_3$ have been obtained by regressing on records scaled to $S_a(T_1) = x_1$. We also know that ln*EDP* is normally distributed, and that it has a variance equal to the $\hat{\sigma}_e^2$. So the probability that ln*EDP* is greater than z, given $S_a(T_1) = x_1$, $R_{T1,T2} = x_2$, and no collapse can be expressed

$$P\left(EDP > z \mid S_a(T_1) = x_1, R_{T1,T2} = x_2, \text{no col.}\right) = 1 - \Phi\left(\frac{\ln z - \left(\hat{\beta}_2 + \hat{\beta}_3 \ln x_2\right)}{\hat{\sigma}_e}\right) \tag{4.10}$$

This equation is very similar to Equation 4.2 used in the scalar case. We previously estimated the mean of the normal distribution by the average response of all records, but now we use a result from regression on $R_{T1,T2}$. We have also replaced the standard deviation of the records by the standard deviation of the regression residual. But otherwise, the equation is the same.

We can combine Equations 4.6 and 4.10 to compute the probability that *EDP* exceeds z:

$$P(EDP > z \mid S_a(T_1) = x_1, R_{T1,T2} = x_2) = P(C) + \left(1 - P(C)\right)\left(1 - \Phi\left(\frac{\ln z - \left(\hat{\beta}_2 + \hat{\beta}_3 \ln x_2\right)}{\hat{\sigma}_e}\right)\right) \quad (4.11)$$

$$\text{where } P\left(C\right) = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_2)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_2)}$$

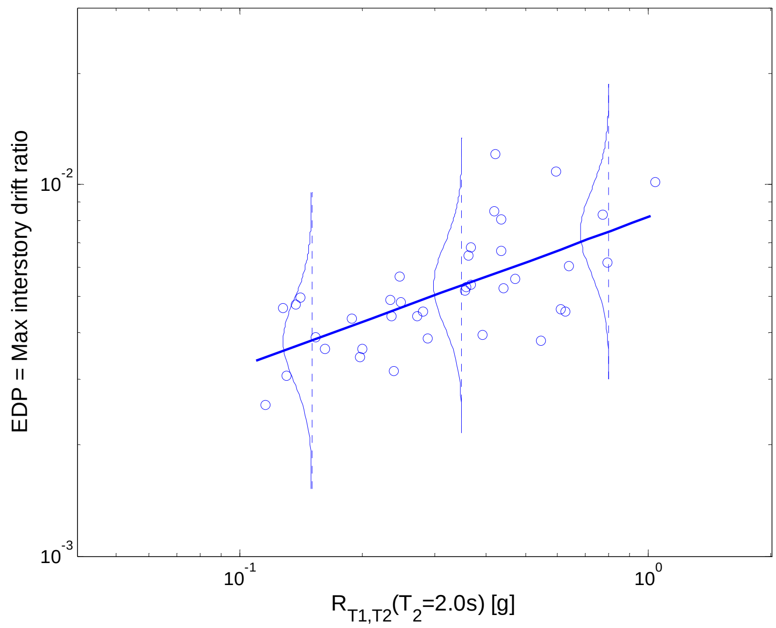
Although x_1 does not appear in Equation 4.11, our estimate is implicitly a function of x_1, because the data used to estimate $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ and $\hat{\sigma}_\varepsilon$ all comes from records scaled to $S_a(T_1) = x_1$. This gives us a response prediction that is similar to the original prediction of Equation 4.4, but that now incorporates a two-element vector. It is a simple matter to generalize this to a larger vector, by using regression on multiple variables in Equations 4.5 and 4.7, as will be done in Section 4.8.

4.5 Choice of a vector

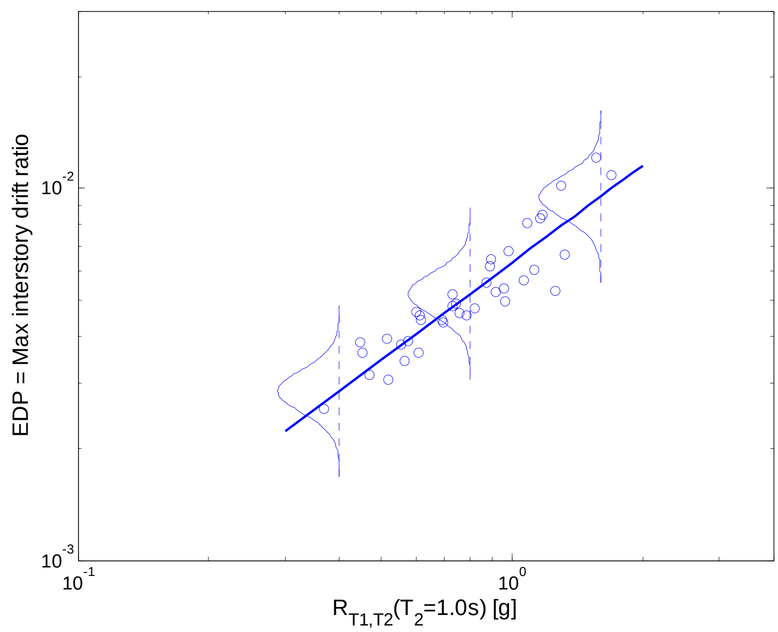
Once the above method has been established for predicting drift, we can proceed to choose an "optimal" vector for use in prediction. The general goals to be considered when choosing an *IM* are efficiency (minimum variance in *EDP* for records with the same *IM* value), sufficiency (Luco and Cornell 2005), and ease of calculation (e.g., it should not be too difficult to obtain the ground motion hazard for the *IM*, or to determine the *IM* value of a given record). As noted earlier, in this study the only class of *IM* considered is $S_a(T_1)$ and $R_{T1,T2}$, with T_2 free to vary over a range of possible values. Here T_2 will be selected to maximize efficiency. The scalar $IM = S_a(T_1)$ has been found to be a sufficient predictor with respect to magnitude and distance (Shome et al. 1998) but in Chapter 3 it was seen to be insufficient with respect to ε. The vector *IM* with the added parameter $R_{T1,T2}$ retains sufficiency with respect to magnitude and distance, and its effect on sufficiency with respect to ε will be considered in Section 4.8. The ground motion hazard for this *IM* can be computed (Bazzurro and Cornell 2002), and spectral acceleration values are easy to calculate and familiar to many engineers. Thus, the ease of calculation criterion is met.

The question of efficiency arises when we examine, for example, the estimated mean of ln*EDP*. We are trying to estimate the mean of a population based on n samples from that population. A basic result in statistics tells us that the standard deviation of that estimate is proportional to $\beta_{\ln EDP|IM=x}\big/\sqrt{n}$ (Benjamin and Cornell 1970). If $\beta_{\ln EDP|IM=x}$ can be decreased by using a more efficient *IM*, then n can be decreased without increasing the standard deviation of the estimated mean ln*EDP*. Reducing n means reducing the number of required nonlinear dynamic analyses, which reduces the effort and computational expense associated with the assessment procedure. This same principle holds in the vector case—to increase the efficiency of our regression estimate, we need to decrease the standard deviation of the regression residuals.

For example, in Figure 4.5a an IM_2 with a large residual standard deviation is shown, whereas in Figure 4.5b an IM_2 with a small residual standard deviation is displayed. If we are trying to estimate a trend in the data, clearly fewer data points will be needed to obtain an estimate of ln*EDP* when using the IM_2 from Figure 4.5b.



(a)



(b)

Figure 4.5: Comparison of the effectiveness of IM_2 with two potential T_2 values. (a) An IM_2 choice with low efficiency; and (b) An IM_2 with high efficiency

To choose the optimal T_2 value at a given $S_a(T_1)$ level, we regress on $R_{T1,T2}$ for a range of T_2 values. We will then compute the standard deviation of the residuals for each of these regressions. To normalize these results, we will compute the fractional reduction in standard deviation relative to the standard deviation from the scalar *IM* case with $S_a(T_1)$ alone. If the fractional reduction is zero, then we have gained no efficiency by including the given IM_2. If the fractional reduction is one, then we have a perfect relation between IM_2 and *EDP*, and there is no remaining randomness. We will choose the T_2 that has the largest fractional reduction among all possible T_2 values.

4.6 Building models and earthquake ground motions

To perform the analysis as described above, one needs both a structural model and suite of earthquake records. The primary structure used in this study is a reinforced concrete moment frame building. The building has 1960's era construction and is serving as a test-bed for PEER research activities (PEER 2004). A 2D model of the transverse frame is used, which was created by Jalayer (2003) and contains nonlinear elements that degrade in strength and stiffness, in both shear and bending (Pincheira et al. 1999). The frame has seven stories and three bays. The first mode of the model has a period of 0.8 seconds, and the second mode has a period of 0.28 seconds. Maximum interstory drift ratios in this structure are predominantly controlled by first-mode response. Forty historical earthquake ground motions are used for the analysis of the primary structure (see Appendix A, Table A.1 for the list of records). All of the records come from California, and the events range in Magnitude from 5.7 to 7.3. The distances vary from 6.5 km to 56 km. Attempts were made to avoid directivity effects by choosing records with small distances only when the rupture and site geometry suggested that near-fault effects would be unlikely, and velocity histories were not observed to contain pulse-like intervals. This set of 40 records was then scaled to 12 levels of $S_a(T_1)$ between 0.02*g* and 1.0*g*.

To supplement this primary structure, a suite of generic frame structures was also evaluated. Twenty generic frame models were analyzed, with a variety of configurations, periods, and degradation properties. The specific model parameters are summarized in Table 3.1. All of the structures are single-bay frames, with stiffnesses and strengths chosen to be representative of typical structures. Six structural configurations were considered, with varying numbers of stories and varying first-mode periods. A set of non-degrading models designed and analyzed by Medina and Krawinkler (Medina and Krawinkler 2005) was considered. These models do not have degrading elements, but can still collapse due to P-Δ effects. A set of degrading models designed and analyzed by Ibarra (2003) was also considered. These structures are identical to the models of

Medina and Krawinkler, except for incorporation of elements that degrade in stiffness and strength. For each of the six building configurations, a non-degrading model and two degrading models were considered. For the supplemental generic structures, a different set of 40 records was used for analysis (see Appendix A, Table A.3 for the list of records). They range in magnitude from 6.5 to 6.9, and range in distance from 13 to 40 km.

Table 4.1. Model parameters for the 20 generic frame structures considered. All parameters specify element properties. The parameter δ_c/δ_y refers to the ductility capacity (the peak displacement at peak strength divided y the yield displacement). The parameter α_c refers to the post-capping stiffness. The cyclic degredation parameters $\gamma_{s,c,k,a}$ quantify the rate of (hysteretic-energy-based) deterioration. All models have a strain hardening stiffness of 0.03 times the elastic stiffness, and all models have a peak-oriented hysteretic model. The details of this hysteretic model are given by Ibarra (2003).

Number of stories	T_1	δ_c/δ_y	α_c	Cyclic deterioration parameters
3	0.3	4	-0.1	$\gamma_{s,c,k,a}=\infty$
3	0.3	4	-0.5	$\gamma_{s,c,k,a}=50$
3	0.3	∞	-	$\gamma_{s,c,k,a}=\infty$
3	0.6	4	-0.1	$\gamma_{s,c,k,a}=\infty$
3	0.6	4	-0.5	$\gamma_{s,c,k,a}=50$
3	0.6	∞	-	$\gamma_{s,c,k,a}=\infty$
6	0.6	4	-0.1	$\gamma_{s,c,k,a}=\infty$
6	0.6	4	-0.5	$\gamma_{s,c,k,a}=50$
6	0.6	∞	-	$\gamma_{s,c,k,a}=\infty$
9	0.9	2	-0.1	$\gamma_{s,c,k,a}=\infty$
9	0.9	4	-0.1	$\gamma_{s,c,k,a}=\infty$
9	0.9	6	-0.1	$\gamma_{s,c,k,a}=\infty$
9	0.9	4	-0.5	$\gamma_{s,c,k,a}=50$
9	0.9	∞	-	$\gamma_{s,c,k,a}=\infty$
9	1.8	4	-0.1	$\gamma_{s,c,k,a}=\infty$
9	1.8	4	-0.5	$\gamma_{s,c,k,a}=50$
9	1.8	∞	-	$\gamma_{s,c,k,a}=\infty$
15	3	4	-0.1	$\gamma_{s,c,k,a}=\infty$
15	3	4	-0.5	$\gamma_{s,c,k,a}=50$
15	3	∞	-	$\gamma_{s,c,k,a}=\infty$

4.7 Results

Using the primary test structure and associated ground motions, we can now use the vector *IM* methodology to test candidate T_2 values for $R_{T1,T2}$. A plot of the fractional reduction in residual standard deviation is shown in Figure 4.6, where all records have been scaled to $S_a(T_1) = 0.3g$. We see that the optimal T_2 is one second (note, this optimal T_2 at this spectral acceleration level was shown in Figure 4.5b above, and a non-optimal T_2 was shown in Figure 4.5a). We see that the reduction in dispersion was approximately 60%. We can make a comparison of the standard errors of estimation

$$\frac{\sigma_{scalar}}{\sqrt{n_{scalar}}} = \frac{\sigma_{vector}}{\sqrt{n_{vector}}} \tag{4.12}$$

If we can reduce our dispersion by 60%, ($\sigma_{vector} = 0.4\sigma_{scalar}$), then $n_{vector} = 0.16 n_{scalar}$. That is, by adopting the most efficient vector, we could potentially reduce the number of records used by a factor of approximately six and still maintain the same confidence in our estimate of the mean response of the structure. This could lead to a great reduction in computational expense. Or, for a fixed n, the estimated standard deviation (which is proportional to the confidence interval) could be reduced by 60%.

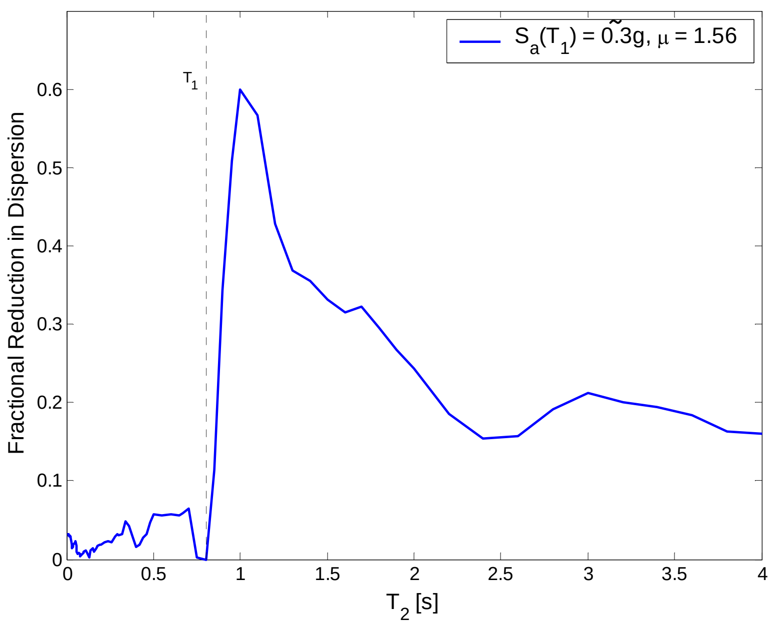


Figure 4.6: Fractional reduction in dispersion vs. T_2 for T_2 between 0 and 4 seconds for $S_a(T_1)$=0.3g.

This optimal T_2 value is only relevant for a single level of $S_a(T_1)$. We repeat this same calculation for two additional levels of $S_a(T_1)$ and show the results in Figure 4.7. It is apparent that the optimal T_2 value varies depending on the level of $S_a(T_1)$. An additional value, $\tilde{\mu}$ is given in the legend. This is the ratio of the average *EDP* among the 40 records to the approximate *EDP* at yielding (as determined from a pushover analysis of the structure). This value is analogous to a ductility level for the structure.

For $S_a(T_1)$ = 0.1g, we find that the optimal T_2 is 0.36 seconds. This is near 0.28 seconds: the second-mode period of the structure. In fact, 0.28 seconds shows a reduction in dispersion that is nearly as large as the reduction at 0.36 seconds. If 0.28 seconds were indeed the best T_2, then this vector *IM* would be somewhat analogous to the modal analysis method of estimating linear response. The modal analysis method (with two modes) would take the spectral acceleration at the first two modes of the building and use them to estimate the response of the structure. Note that at this level of spectral acceleration, our estimate of ductility using the above definition is 0.533, implying that for most of the records, the structure stays linear.

For $S_a(T_1)$ = 0.3g, we find that the optimal T_2 is 1.0 seconds. This is the level that was discussed previously. It is noted now that $\tilde{\mu} = 1.56$, suggesting that most of the records cause some level of nonlinear behavior in the structure. For $S_a(T_1)$ = 0.7g, we find that the optimal T_2 is 1.5 seconds. At this level of spectral acceleration, $\tilde{\mu} = 4.66$. This suggests that most of the records experience large levels of nonlinearity.

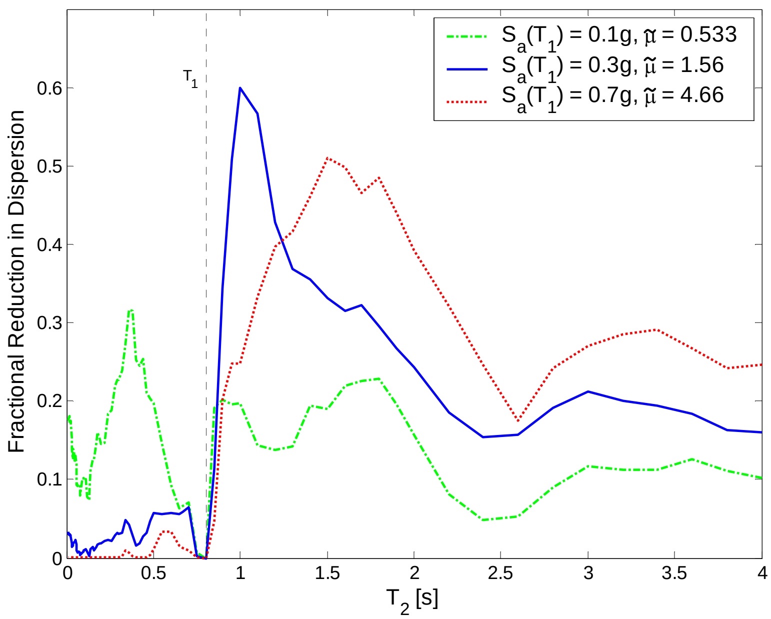


Figure 4.7: Fractional reduction in dispersion vs. T_2 for three levels of $S_a(T_1)$.

If one were to combine engineering intuition with the results of Figure 4.7, the following conclusion might be drawn, keeping in mind that response of this seven-story structure is first-mode dominated: if $S_a(T_1)$ is low enough that few or no records cause nonlinearity, then the optimal second period to incorporate would be near the second-mode period of the structure, and if $S_a(T_1)$ is large enough that most records cause nonlinearities, then the optimal T_2 will be larger than T_1. Figure 4.8 below shows this pattern as well. In this figure, the level of $S_a(T_1)$ is plotted on the x-axis. On the y-axis is T_2, plotted as a ratio of T_2 to T_1. A circle is placed in the plot at the location of the optimal T_2 for a given $S_a(T_1)$. In addition, a line is plotted over the range of T_2/T_1

where the reduction in dispersion is at least 75% of the reduction seen at the optimal T_2. This line is shown to indicate the breadth of the optimal solution (i.e., are there only a few effective T_2's, or is there a large range of T_2 that reduces dispersion comparably?). Other authors have examined choices of T_2, and have also recognized the dependence on the level of nonlinearity (Cordova et al. 2001, Vamvatsikos 2002).

The average $\tilde{\mu}$ is less than one for $S_a(T_1)$ between 0 and 0.1g, and this is where we see an optimal T_2 that is near to the second-mode period of the building. For $S_a(T_1) \geq 0.2g$ (and $\tilde{\mu} > 1$), the optimal T_2 is larger than T_1, and shows an increasing trend as $S_a(T_1)$ increases (and as levels of nonlinearity increase). This increase in T_2 with increasing levels of nonlinearity may be related to the idea of an equivalent linear system (e.g., Iwan 1980, Kennedy et al. 1985). The theory is that a nonlinear Single-Degree-Of-Freedom (SDOF) system may be represented by an "equivalent" linear system with a longer period. As the level of nonlinearity increases, the period of the equivalent linear system increases. A schematic force-deformation diagram of a nonlinear system and its equivalent linear system is shown in Figure 4.9. The original SDOF has elastic stiffness k_e, and the equivalent linear system has a reduced stiffness k^* that is dependant on the ductility demand (μ_{max}) of the nonlinear system. Methods for determining an equivalent nonlinear system based on level of ductility have been proposed by, for example, Iwan (1980) and Kennedy (1985). The suggested equivalent periods from these two papers are plotted on Figure 4.8, and they show trends similar to the ideal T_2's seen in this study. It should be noted, however, that there is a difference between the period of an equivalent linear system and the IM_2 being selected for a vector. An equivalent linear system is used to *replace* $S_a(T_1)$, while the IM_2 is used to *supplement* $S_a(T_1)$. The discrepancy between these two goals is most apparent in Figure 4.8 when $S_a(T_1)$ is 0.3 or 0.4g. Here, the equivalent nonlinear system has a period almost identical to the elastic period of structure. However, the optimal T_2 from this study is at a longer period. This is because if T_2 is near to T_1, then $S_a(T_1)$ is very highly correlated with $S_a(T_2)$ (see Chapter 8). Thus, $R_{T1,T2} = S_a(T_2)/S_a(T_1)$ is essentially constant for all records and so there is little or no predictive ability provided by $R_{T1,T2}$. The optimal T_2 must be significantly different than T_1 in order to decrease the correlation between $S_a(T_1)$ and $S_a(T_2)$. This can be seen in Figure 4.7, where the fractional reduction in dispersion is zero or nearly zero for T_2 values that are very close to T_1. So, although the equivalent linear systems somewhat match the optimal T_2 values at moderate to large levels of ductility, there is a difference between the two at low levels of nonlinearity.

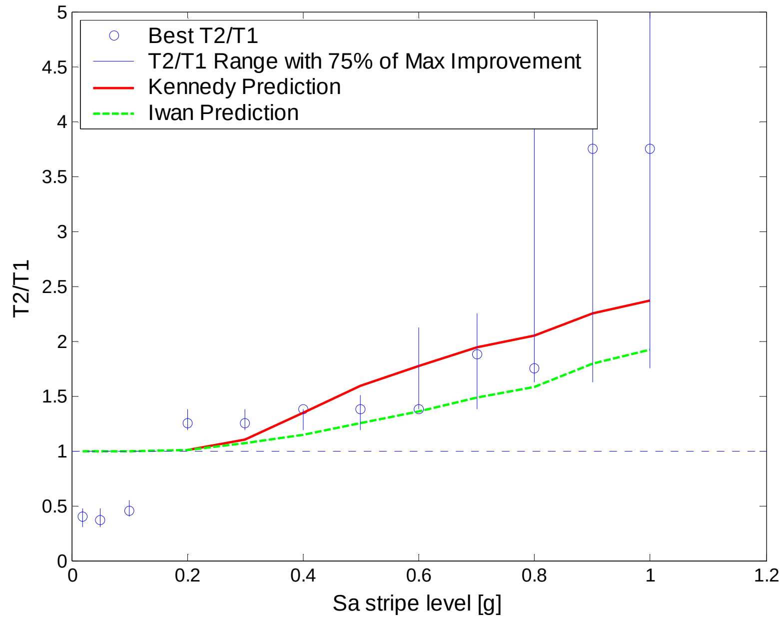


Figure 4.8: The optimum second period T_2, versus the level of $S_a(T_1)$.

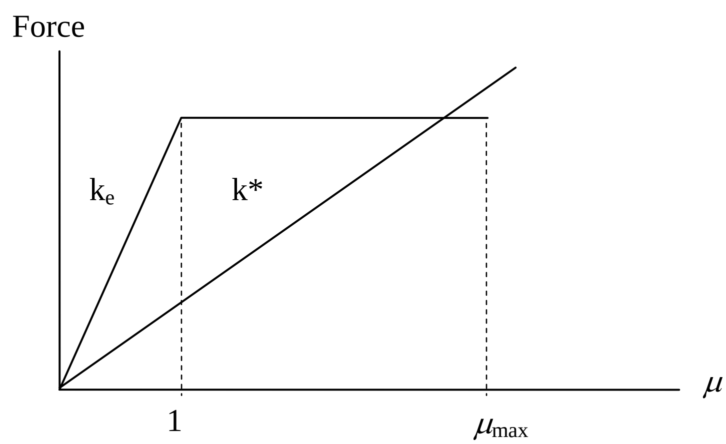


Figure 4.9: Schematic illustration of a non-linear SDOF and a potential equivalent linear system based upon maximum observed ductility.

The conclusions from the Van Nuys testbed structure can be generalized by examining the generic frame structures described earlier and listed in Table 3.1. The same search for optimal periods was performed, but for display purposes, $S_a(T_1)$ is converted to the normalized value $(S_a(T_1)/g)$/(yield base shear/weight), so that the nonlinear behavior starts occurring at a normalized value of approximately one (Medina and Krawinkler 2003). Further, the optimal T_2 is normalized by the first-mode period of the structure as was done in Figure 4.8. These normalizations allow multiple structures to be plotted on the same figure for comparison, independent of their first-mode period or yield strength. The structures are separated into three groups for consideration.

The first group, shown in Figure 4.10, consists of three-story structures with peak interstory drift ratios controlled primarily by first-mode response (a description which also fits the primary building considered previously). For most of these structures, the optimal T_2 is approximately equal to the second-mode period of the structure when the structure is linear (normalized spectral acceleration<1) or slightly nonlinear. In the moderately nonlinear range, however, the optimal period is typically larger than the first-mode period, and increases as the ground motion intensity increases. The general trend in Figure 4.10 is similar to the trend in Figure 4.8.

The second group of structures, shown in Figure 4.11, have peak interstory drift ratios that are moderately sensitive to second-mode response. In the elastic range, the optimal T_2 is again near the second-mode period. However, here the optimal T_2 is less than the first-mode period for normalized *Sa* values up to four in several cases. The optimal T_2 in this nonlinear response range is larger than the elastic second-mode period, however. This suggests that the effective second-mode period may be lengthening due to nonlinearity (Fu 2005), although no further analysis was performed to confirm this.

The third group of structures, shown in Figure 4.12, have peak interstory drift ratios that are significantly affected by second-mode response—more-so than most typical buildings (Krawinkler 2005). Here, the optimal T_2 is nearly always less than the first-mode period. Again, the optimal T_2 is greater than the elastic second-mode period for some structures, perhaps indicating nonlinearity in the second mode.

The fourth group of structures, shown in Figure 4.13, all have nine stories and a first-mode period of 0.9 seconds, but have varying ductility capacities (δ_c/δ_y). Each of these structures undergoes a transition from having an optimal T_2 less than T_1 to having an optimal T_2 greater than T_2, once the normalized *Sa* level becomes large enough. The structures with lower ductility capacity undergo this transition at a lower *Sa* level. This is likely because structures with low ductility are more sensitive to long-period excitations that may drive them to extreme responses. Ductile structures may be more able to withstand these long-period excitations, and thus information about higher mode response may be more important in this transition region. At large enough *Sa* levels, all four of these structures are sensitive to long-period T_2's; the low-ductility structures merely make the transition sooner.

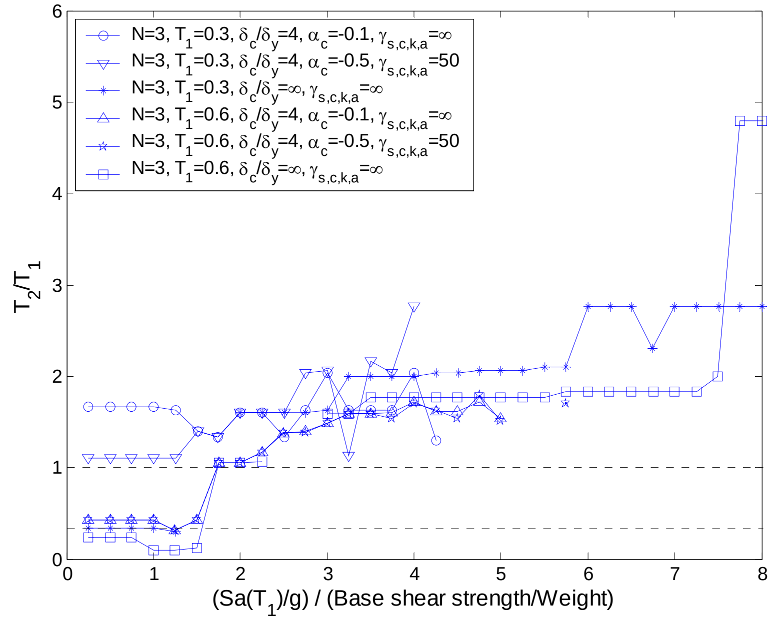


Figure 4.10: The optimal T_2, normalized by first-mode period (T_1) for a set of three-story structures dominated by first-mode response.

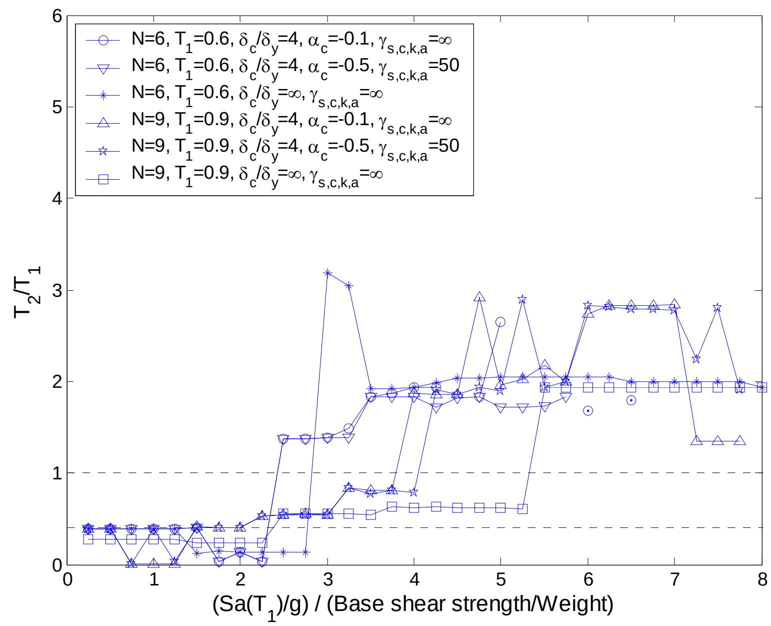


Figure 4.11: The optimal T_2, normalized by first-mode period (T_1) for a set of six- and nine-story structures with moderate contribution from second-mode response.

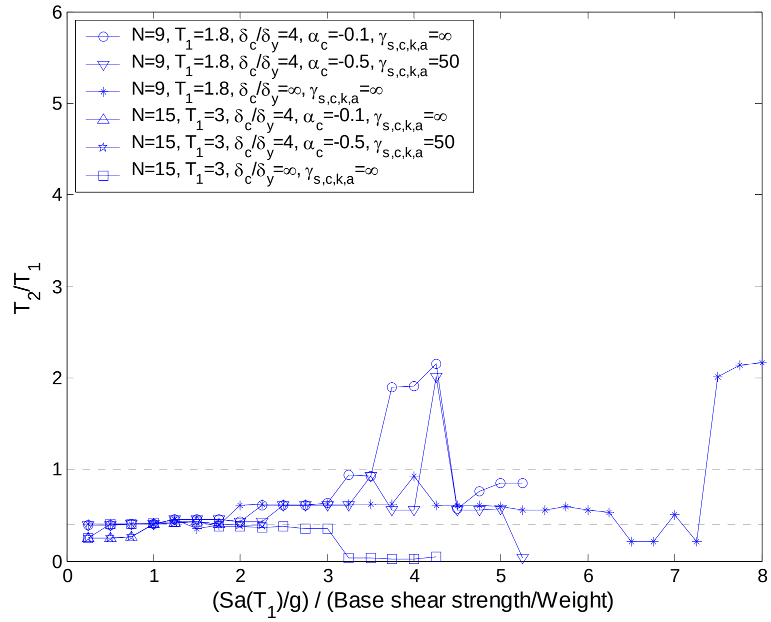


Figure 4.12: The optimal T_2, normalized by first-mode period (T_1) for a set of nine- and fifteen-story structures with significant contribution from second-mode response.

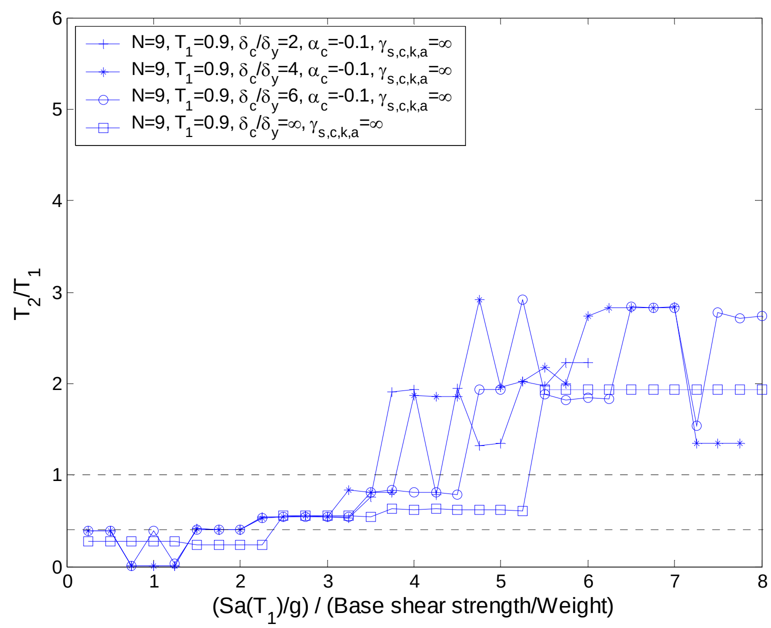


Figure 4.13: The optimal T_2, normalized by first-mode period (T_1) for a set of nine-structures with varying levels of element ductility.

Some general conclusions might be inferred from these empirical results. When the structure is significantly affected by second-mode response (either the structure is behaving linearly or second-mode response is a strong contributor to response even in the nonlinear range), then the optimal T_2 is typically near the elastic second-mode period. When the structure is behaving nonlinearly and second-mode response is not critical, the optimal T_2 is typically larger than the elastic first-mode period. In this case, an optimal T_2 might be estimated using an equivalent linear system based on the predicted ductility level. In addition, low-ductility structures may have optimal T_2 values longer than the first-mode period at lower *Sa* levels than high ductility structures. These trends are all consistent with engineering intuition about dynamic structural response, but the results are not general enough to develop concrete rules for optimal T_2 values at this time.

With these results in mind, it is valuable to reconsider our original scheme for predicting *EDP* as a function of a two *IM* parameters. As discussed in Chapter 2, there are a variety of choices for making this prediction, but here we chose to first scale records to $S_a(T_1)$ and then regress on $R_{T1,T2}$. Therefore, all of our regression coefficients were allowed to vary at each $S_a(T_1)$ level. This allows for full interaction between $S_a(T_1)$ and $R_{T1,T2}$. If we had not scaled the records, but instead used un-scaled records and regressed on both $S_a(T_1)$ and $R_{T1,T2}$ simultaneously, then it would have been more difficult to identify the interaction between the two parameters. In addition, had we regressed on both $S_a(T_1)$ and $R_{T1,T2}$ simultaneously without scaling, we would only find a single optimal T_2 for all $S_a(T_1)$ levels, rather than an optimal T_2 that varies with the level. For these two reasons, scaling to $S_a(T_1)$ and then regressing on $R_{T1,T2}$ was chosen as the methodology. It should be noted however, that regressing on two *Im*s simultaneously has been done by Shome (1999). The advantage of regressing on two *Im*s simultaneously is that there are fewer parameters to estimate (one set of coefficients for the entire range of $S_a(T_1)$ and $R_{T1,T2}$, rather than a new set of coefficients for $R_{T1,T2}$ at each $S_a(T_1)$ stripe), so potentially fewer analyses need to be performed in order to estimate the needed parameters.

4.7.1 Optimization of the choice of IM$_2$ using the bootstrap and the drift hazard curve

Another issue with the proposed methodology that should be mentioned is the criterion for choosing an optimal $R_{T1,T2}$. Recall that the quantity being optimized in Figure 4.6 is the reduction in residual dispersion from regression on the *non-collapse* records only. This quantity does not measure the ability of $R_{T1,T2}$ to predict the probability of collapse. Perhaps if the collapse capacity of the structure is of primary interest, the quantity to optimize would be the reduction in residual dispersion from Equation 4.6. Ideally, the quantity to optimize should incorporate the improvements made in prediction of both collapse *and* non-collapse records. Because of this, the plot of Figure 4.8 is limited to $S_a(T_1) \leq 1g$. For $S_a(T_1) > 1g$, more than one-quarter of the records cause the structure to collapse, so it would not be appropriate to neglect the records that cause collapse as has been done above.

It would be possible to address this issue by computing a weighted sum of the reduction in dispersion from the collapse prediction (Equation 4.6) and the reduction in dispersion from regression on the non-collapse records (Equation 4.7). But the weights used should be a function of the relative importance of collapse and non-collapse behavior, and it is not clear what that function should be. One natural way to resolve this issue is to carry the prediction of *EDP* all the way through to the computation of the mean rate of exceeding an *EDP* value z, as discussed in the introduction. The vector version of Equation 4.1 is given below:

$$\lambda_{EDP}(z) = \sum_{\text{all } x_{1,i}} \sum_{\text{all } x_{2,i}} P(EDP > z \mid S_a(T_1) = x_{1,i}, R_{T1,T2} = x_{2,i}) \cdot \Delta\lambda_{IM}(x_{1,i}, x_{2,i}) \tag{4.13}$$

This rate is the final result we desire, and this calculation incorporates both the collapse prediction and the non-collapse response prediction. The estimate will vary depending upon the records chosen for analysis, although with an efficient *IM*, hopefully the estimate will not vary by much. In fact, our ultimate goal is to choose an *IM* such that this result does not depend upon the records chosen, and so a reduction in sensitivity to the records chosen here would be a good criterion to use in selecting $R_{T1,T2}$. To measure the statistical variability of the estimate of $\lambda_{EDP}(z)$, one can employ the bootstrap (Efron and Tibshirani 1993). The procedure is as follows: select n records *with replacement* from the original set of n records (some records will be duplicated and others will not be present at all). With this new record set and a candidate $R_{T1,T2}$, use the vector *IM* to predict the building response as outlined above. Using this new estimate, re-compute $\lambda_{EDP}(z)$ using Equation 4.13. Repeat this process B times (typically B = 25 to 200). The standard deviation of these B values is an estimate of the standard error of estimation of $\lambda_{EDP}(z)$. No new structural analyses are needed for this process, so the computational expense is not large. A good T_2 value for $R_{T1,T2}$ will result in a $\lambda_{EDP}(z)$ with significantly reduced variability relative to the $\lambda_{EDP}(z)$ using a scalar *IM* (e.g., Equation 4.1). There are two advantages introduced by this method. The first is that it incorporates gains in efficiency from both the collapse and non-collapse predictions, but rather than using a potentially arbitrary weighting scheme, it incorporates them naturally into the final computation where the results will be used. The second advantage is that the probability of exceeding a given *EDP* value does not come from a single $S_a(T_1)$ level, but from a range of levels. Thus, a good predictor according to this criterion will show efficiency gains over the range of $S_a(T_1)$ values that contribute significantly at the given *EDP* level. Note that this calculation requires knowledge of vector-valued ground motion hazard, $\lambda_{IM}(x_{1,i}, x_{2,i})$ (the rate of jointly exceeding both $S_a(T_1)$=$x_{1,I}$ and $R_{T1,T2}$=$x_{2,i}$). This result is available

(Bazzurro and Cornell 2002, Somerville and Thio 2003), but not yet in widespread use in engineering practice.

As an illustration, the complete drift hazard curve $\lambda_{EDP}(z)$ is computed using the scalar and vector procedures (Equations 4.1 and 4.13 respectively). The ground motion hazard is calculated for the actual location of the structure being studied (a soil site in the Los Angeles area). The vector *IM* used is ($S_a(T_1)$, $R_{T1,T2}(T_2$=1.0s)). The scalar and vector-based curves are shown in Figure 4.14. Note that the flattening of the curve towards the right occurs because the exceedance of these *EDP* values is dominated by collapses (P(collapse) $\cong$ 3*10^{-3}). There is not a large difference between the two curves, but what cannot be seen in this figure is that there is much less uncertainty in the curve estimated using the vector *IM*. To measure this uncertainty, we now use the bootstrap. Four example bootstrap replicates of the record set have been generated, and their corresponding vector-*IM*-based drift hazard curves are shown in Figure 4.15. To display the variability in our estimates, we compute histograms of the $\lambda_{EDP}(z)$ values for a given z, using both the vector *IM* and scalar *IM* methods. These curves use maximum interstory drift ratio as the *EDP* of interest.

We now examine histograms at a value of z of 0.01, to compare our scalar and vector *Im*s (see Figure 4.16). We see that the results from the vector-based drift hazard curve are much more tightly bunched around their central value than the results from the scalar-based drift hazard curve. This means that we can be more confident about the value obtained from the vector-based drift hazard calculation. Or equivalently, if we adopt the vector-based drift hazard calculation, we do not need to use as many records in the analysis to obtain a level of uncertainty comparable to the scalar method. (In this case, the standard deviation of the bootstrapped replicates is cut by a factor of about two, so in principle the number of records used could be reduced by a factor of about four. However, the method does demand a minimum of approximately five to ten records per stripe in order to estimate the needed regression coefficients.)

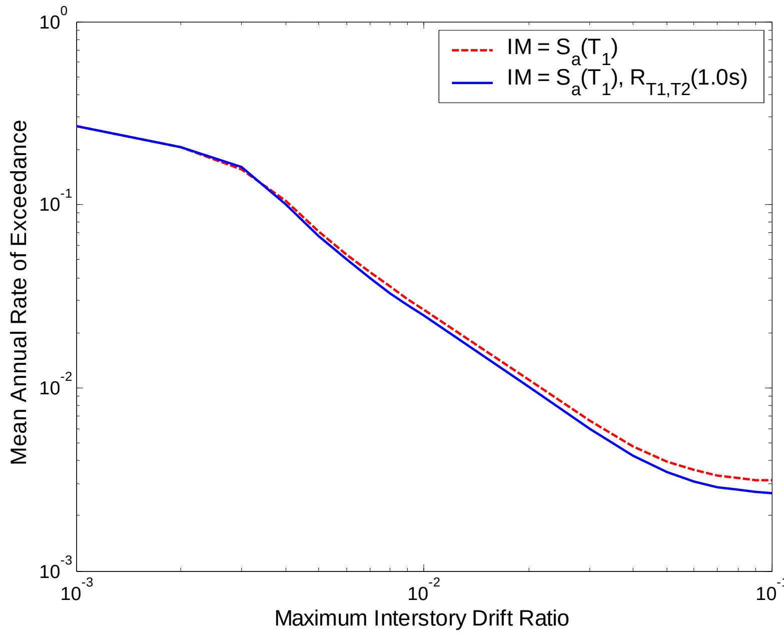


Figure 4.14: Maximum interstory drift hazard curves computed using a scalar *IM* and a vector *IM*

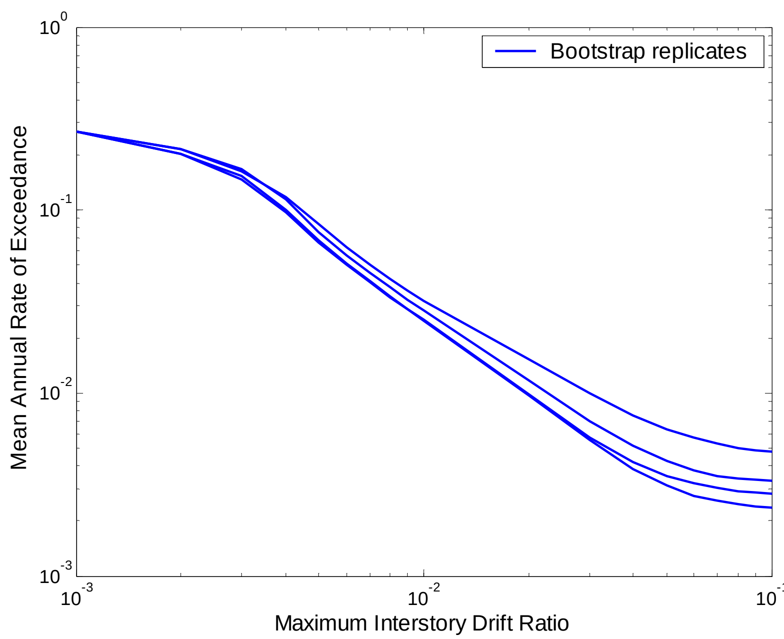


Figure 4.15: Bootstrap replicates of the vector *IM* drift hazard curve

The histogram of Figure 4.16 only shows results for one vector *IM* at a single *EDP* value, but we can extend this calculation to and additional values of *EDP*. We can also consider additional candidate *Im*s. Consideration of candidate vector *Im*s is slightly more complicated now, because the vector ground motion hazard $\lambda_{IM}(x_1,x_2)$ needs to be recomputed for each candidate *IM*. For this reason, the analysis here is limited to three vectors that appeared promising in Figure 4.8: T_2 = 0.28s, 1.0s and 2.0s. The coefficient of variation of the replicate results is used as a measure of dispersion of the bootstrap replicates (because the mean values of $\lambda_{EDP}(z)$ vary by two orders of magnitude depending on z, standard deviations of $\lambda_{EDP}(z)$ are much more variable than the coefficient of variation). A plot of the coefficient of variation versus *EDP* level is shown for the three candidate vector *IM*s in Figure 4.17, along with the coefficient of variation using the scalar *IM*. We see that the vector ($S_a(T_1)$, $R_{T1,T2}(T_2$=1.0s)) produces a significant reduction in coefficient of variation for nearly all levels of *IM*, but the 50% reduction is limited to the range 0.003<*EDP*<0.01. The other two vectors do not show a significant improvement. This result fits with results seen earlier. We saw in Figure 4.8 that the vector with T_2 = 0.28s was only helpful for very small levels of $S_a(T_1)$ (and in fact we see that for very small levels of *EDP*, this *IM* produces a small improvement). The vector with T_2 = 2.0s was helpful as $S_a(T_1)$ levels got very large, but these large-intensity events are rare enough that they do not significantly affect the *EDP* hazard curve except at large levels of *EDP* (and here we see a slight improvement). The vector with T_2 = 1.0s showed a significant improvement over a large range of important $S_a(T_1)$ levels, and thus it is the most useful. These results are consistent with earlier results, but perhaps this method reveals more information about the overall usefulness of a candidate vector than the previous method did. Note that to select a vector, we still need to specify a value of z to consider. But this choice of z may not be too hard if we are concerned with specific limit states (e.g. collapse, or *EDP* = 1% drift).

This bootstrap procedure requires more computation than the regression procedure. However, it has the advantage of directly measuring uncertainty in the value of interest ($\lambda_{EDP}(z)$), and incorporating estimates from both collapse and non-collapse prediction at multiple *IM* levels simultaneously. We saw from the example calculation above that the results from the two procedures appears consistent. Research to date seems to show that the two alternative techniques agree at levels of *EDP* where collapses are not frequent.

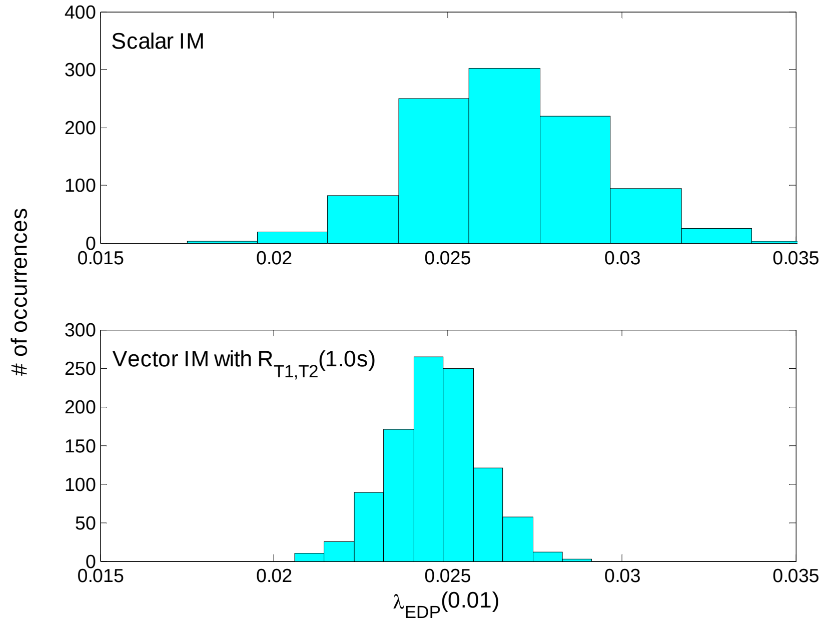


Figure 4.16: Histograms of the scalar and vector drift hazard curves, for z = 0.01.

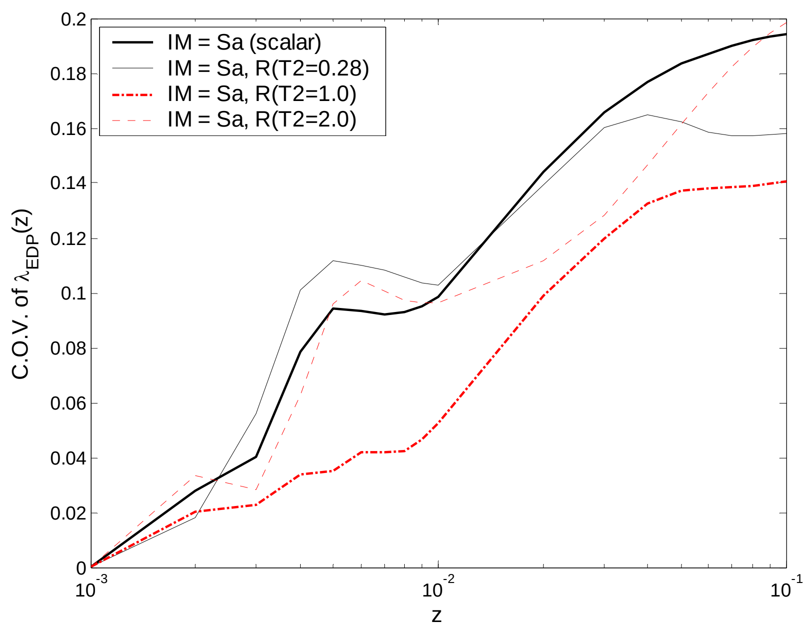


Figure 4.17: Coefficient of variation vs. *EDP* level for three candidate vector *Im*s and the scalar *IM*.

4.8 A three-parameter vector consisting of Sa(T_1), $R_{T1,T2}$ and ε

In Chapter 3, the ground motion parameter ε was investigated, and found to have a significant effect on structural response given $Sa(T_1)$. Both ε and $R_{T1,T2}$ are related to spectral shape, so the possibility exists that they are accounting for the same effect. If they do account for the same effect, then use of a vector with $R_{T1,T2}$ would eliminate the need to carefully account for the effect of ε. In order to investigate this possibility, we consider a vector consisting of three parameters: $Sa(T_1)$, $R_{T1,T2}$ and ε (where ε is calculated at T_1, as in Chapter 3). The question to be answered is whether this three-parameter *IM* is better than the two parameter *IM* consisting of $S_a(T_1)$ and $R_{T1,T2}$. To test this, we first perform scaling on $Sa(T_1)$ stripes, as in Section 4.4. Then, at each stripe, instead of performing regression on $R_{T1,T2}$ alone, we perform multiple regression, using both $R_{T1,T2}$ and ε as predictors. The prediction is analogous to Equation 4.7, but now with an added term for ε. The equation becomes

$$\ln EDP_i = \beta_2 + \beta_3 \ln R_i + \beta_4 \varepsilon_i + e_i \tag{4.14}$$

where EDP_i is the *EDP* value of record *I* at a given $Sa(T_1)$ level, as before, and R_i and ε_I are the records' associated $R_{T1,T2}$ and ε values, respectively. In the same way, we can use the extra ε parameter to aid in predicting collapse, using a generalization of Equation 4.5

$$Y_i = \frac{\exp(\beta_0 + \beta_1 R_i + \beta_5 \varepsilon_i)}{1 + \exp(\beta_0 + \beta_1 R_{ii} + \beta_5 \varepsilon_i)} \tag{4.15}$$

where again Y_i is equal to 1 if the record causes collapse at the given $Sa(T_1)$ level and 0 otherwise.

The question to be answered is whether the predictions incorporating ε (Equations 4.14 and 4.15) are significantly superior to predictions that do not include ε (Equations 4.5 and 4.7). A statistical test known as the F-test (Neter et al. 1996) helps to answer this question. For a given structure and $S_a(T_1)$ level, both the full model (Equations 4.14 and 4.15) and the reduced model (Equations 4.5 and 4.7) are fitted. A normalized measure of the improvement in fit from the reduced model to the full model, called an F statistic, is computed. The probability of exceeding that F statistic under the assumption that ε is not significant is called a "p-value." Low p-values suggest that ε is in fact significant, because they indicate that the observed improvement in fit is unlikely to occur under the assumption that ε is not significant. A p-value of less than 0.05 is typically interpreted as indicating that ε has a statistically significant effect. P-values from a subset of the test results are displayed in Table 4.2. In all, approximately 30% of the tests show p-values below 0.05, versus the predicted 5% under the assumption that ε is not significant. This suggests that ε is still a mildly significant predictor of structural response as a third parameter in addition to $S_a(T_1)$ and $R_{T1,T2}$. One pattern in the results of these tests can be seen in the linear regression tests for approximately linear ($S_a(T_1)/g$)/(Base shear strength/Weight) ≤ 1) structures is that when T_2 is chosen to be longer than the first-mode period of the building, ε is very significant for the 0.6s and 0.9s structures, while when T_2 is chosen at the second-mode period of the building, ε is no longer significant for these same structures. This suggests that ε is predicting response of the second mode: ε is only significant if the $R_{T1,T2}$ is not chosen to account for the second-mode response. The same effect is not present with the 0.3 second structure; this is probably because the second mode of this structure does not contribute significantly to the response parameter of interest.

Table 4.2. P-values for tests of significance of ε, given $S_a(T_1)$ and $R_{T1,T2}$. Tests were performed by predicting the response of the generic frame structures using the candidate intensity measures. All structures considered above have the following parameter values: δ_c/δ_y=4, α_c=-0.1 and $\gamma_{s,c,k,a}$=∞ (see Table 3.1). Linear regression tests were only performed when at least 10 records did not cause collapse. Logistic regression tests were only performed when at least 5 records caused collapse *and* at least 5 records did not cause collapse. The field is marked with "-" if the test was not performed. Tests indicating statistical significance (p<0.05) are marked in bold.

	Linear Regression						**Logistic Regression**					
	$T_2 = T_1/3$			**$T_2 = T_1*2$**			**$T_2 = T_1/3$**			**$T_2 = T_1*2$**		
# of stories	3	6	9	3	6	9	3	6	9	3	6	9
First-mode period, T_1	0.3	0.6	0.9	0.3	0.6	0.9	0.3	0.6	0.9	0.3	0.6	0.9
T_2 used in $R_{T1,T2}$	0.1	0.2	0.3	0.6	1.2	1.8	0.1	0.2	0.3	0.6	1.2	1.8
$Sa(T_1)$/(base shear coefficient*g)												
0.25	0.61	0.97	0.97	0.87	**0.00**	**0.00**	-	-	-	-	-	-
0.5	0.61	0.97	0.97	0.87	**0.00**	**0.00**	-	-	-	-	-	-
0.75	0.61	0.97	0.97	0.87	**0.00**	**0.00**	-	-	-	-	-	-
1	0.61	0.97	0.97	0.87	**0.00**	**0.00**	-	-	-	-	-	-
1.25	0.91	0.42	0.50	0.99	**0.02**	**0.00**	-	-	-	-	-	-
1.5	**0.02**	0.15	0.57	**0.03**	**0.01**	**0.00**	-	-	-	-	-	-
1.75	**0.02**	**0.02**	0.61	**0.02**	**0.01**	**0.00**	-	-	-	-	-	-
2	**0.05**	**0.02**	0.93	**0.03**	**0.02**	**0.00**	-	-	-	-	-	-
2.25	0.07	**0.02**	0.76	0.06	**0.03**	**0.01**	-	-	-	-	-	-
2.5	0.41	**0.03**	0.64	0.31	0.11	**0.03**	-	-	-	-	-	-
2.75	0.30	**0.04**	0.86	0.19	0.21	**0.03**	-	-	-	-	-	-
3	0.24	0.05	0.88	0.18	0.25	0.06	0.25	-	-	0.52	-	-
3.25	0.14	0.07	0.50	**0.04**	0.26	0.10	0.26	-	-	0.71	-	-
3.5	0.77	0.14	0.13	0.59	0.59	**0.03**	**0.05**	-	-	0.06	-	-
3.75	0.57	0.20	0.08	0.38	0.63	**0.04**	0.09	-	-	0.09	-	-
4	0.62	0.19	**0.04**	0.45	0.07	0.08	0.09	-	-	0.12	-	-
4.25	0.70	0.25	0.05	0.11	0.07	0.09	**0.04**	-	-	0.06	-	-
4.5	-	0.07	**0.04**	-	**0.01**	0.15	**0.03**	-	-	**0.03**	-	-
4.75	-	0.22	0.28	-	0.08	0.44	**0.03**	0.17	**0.04**	**0.03**	0.16	0.29
5	-	0.74	0.12	-	0.27	0.46	**0.01**	0.07	0.11	**0.02**	**0.04**	0.67
5.25	-	0.46	0.10	-	0.23	0.39	**0.01**	0.14	0.11	**0.02**	**0.04**	0.67
5.5	-	0.91	0.16	-	0.67	0.32	**0.04**	**0.04**	0.07	**0.03**	**0.02**	0.61
5.75	-	0.89	**0.03**	-	0.94	0.10	0.07	0.11	0.16	0.07	**0.03**	0.97
6	-	0.45	0.06	-	0.41	0.16	0.07	0.14	0.16	0.07	0.08	0.97
7	-	-	0.60	-	-	0.59	-	0.11	0.06	-	0.11	0.38
8	-	-	-	-	-	-	-	0.16	0.06	-	0.87	0.30
9	-	-	-	-	-	-	-	0.17	0.23	-	0.97	0.37
10	-	-	-	-	-	-	-	**0.05**	0.18	-	0.15	0.40
11	-	-	-	-	-	-	-	**0.05**	0.59	-	0.15	0.91
12	-	-	-	-	-	-	-	-	0.86	-	-	0.26

The response predictions of Equations 4.14 and 4.15 can also be combined with ground motion hazard results to re-compute a drift hazard curve, similar to Figure 4.14. The ground motion hazard is calculated by performing vector-valued hazard analysis as before, but also disaggregating on ε to obtain the conditional distributions of ε given $Sa(T_1)$ and $R_{T1,T2}$. The resulting hazard curves using several candidate *Im*s are shown in Figure 4.18. The *Im*s including ε as a parameter (the heavy lines) result in lower estimates of the mean annual rate of exceedance at large maximum interstory drift ratios. The *Im*s consisting of $R_{T1,T2}$, but not ε, do not show the same lower mean annual rate of exceedance. If the addition of ε to the *IM* affects the drift hazard results, then it can be inferred that $R_{T1,T2}$ is not fully accounting for the effect of ε.

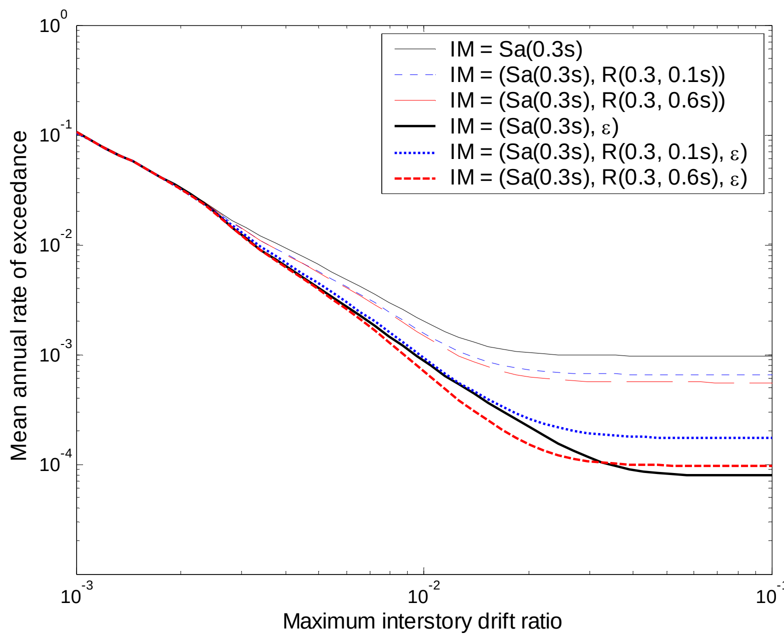


Figure 4.18: Maximum interstory drift hazard curves computed using a scalar *IM*, two-parameter *Im*s, and three-parameter *Im*s. Results shown are for the 3-story structure with T_1=0.3s, δ_c/δ_y=4, α_c=-0.1 and $\gamma_{s,c,k,a}$=∞ (see Table 3.1).

A possible explanation as to why $R_{T1,T2}$ is not able to account for the effect of ε is shown in Figure 4.19. Here, the ability to predict spectral shape at a range of periods is measured using $R_{T1,T2}$ and/or ε. The procedure used to create this plot is as follows. The set of 40 ground motions used for analysis with the generic frames was selected and scaled so that they all had the same Sa(0.8s) value. The parameters ε and $R_{T1,T2}$ were then used to predict spectral acceleration values at a range of periods. The standard deviation of the residuals after prediction was compared to the standard deviation of the spectral values before prediction. The level of reduction in standard deviation indicates the second parameter's predictive power. In Figure 4.19, we see the reduction in standard deviation resulting from prediction using the two possible parameters. $R_{T1,T2}$ explains 100% of the variation at the period T_2 (1.0s in this figure). But at other periods it is less effective. In particular, $R_{T1,T2}$ with this T_2>T_1 has essentially zero predictive power at periods less than T_1, as might be expected. On the other hand, ε has at least some predictive power over almost the entire range of periods considered. Judging from this picture, it appears that ε and $R_{T1,T2}$ are not explaining the same features of spectral shape. In particular, ε predicts spectral values on either side of T_1, while $R_{T1,T2}$ predicts only on one side. This would explain why the predictions based on the three-parameter vector of $Sa(T_1)$, $R_{T1,T2}$ and ε are different than the predictions based on $Sa(T_1)$ and $R_{T1,T2}$.

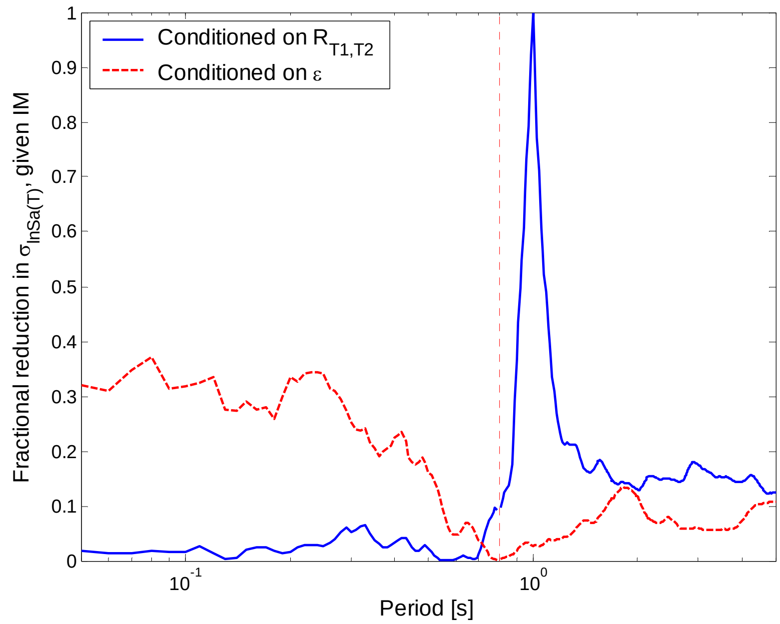


Figure 4.19: Fractional reduction in standard deviation of $Sa(T)$ given $Sa(T_1)$ after conditioning on $R_{T1,T2}$ or ε. T_1=0.8s, T_2=1.0s.

Additionally, unpublished research by the author indicates that when the median spectral shape of the records used for structural analysis is inconsistent with the spectral shape predicted by the ground motion prediction (attenuation) model used for hazard analysis, an inconsistency in the resulting drift hazard is introduced. This may be the cause of the two-parameter and three-parameter drift hazard curves in Figure 4.18 not agreeing. Further investigation is needed into this phenomenon, however, before firm conclusions can be drawn.

Another test that can be performed is to again use bootstrap to compute the coefficient of variation of estimation for each drift hazard curve in Figure 4.18. The surprising result from this figure is that the coefficients of variation of $\lambda_{EDP}(z)$ for the intensity measures containing ε are much larger than even the scalar IM $S_a(T_1)$. This increased coefficient of variation appears to result from the fact that ε-based predictions (i.e., the regression of EDP and $P(Collapse)$ based on ε or $R_{T1,T2}$ and ε at a $S_a(T_1)$ stripe) normally require some extrapolation, because mean ε values for randomly selected ground motions are near zero, while ground motion hazards and *EDP* hazards of interest for safety assessment (e.g., $<10^{-3}$) can often be dominated by ε values of one to two[9]. Further, the prediction based on ε explains less of the record-to-record variability than does $R_{T1,T2}$ (compare, e.g., Figure 4.4 to Figure 3.1b). With the use of three predictors the curse of dimensionality (Bellman 1961, Hastie et al. 2001) becomes more significant, and so although the efficient parameter $R_{T1,T2}$ reduces the coefficient of variation relative to the vector with ε, it cannot reduce the overall coefficient of variation back to the level of the scalar *IM*.

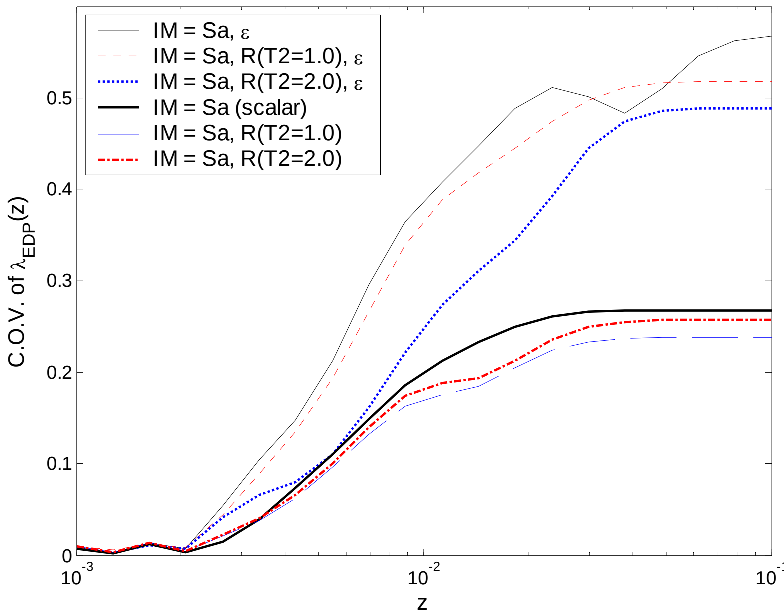


Figure 4.20: Coefficient of variation vs. *EDP* level for six candidate *Im*s consisting of one, two or three parameters.

Although ε increases the coefficient of variation of estimation for of the drift hazard curve when used in an *IM*, it was seen in Chapter 3 to eliminate a source of bias associated with use of $S_a(T_1)$ alone. And Figure 4.18 suggests that $R_{T1,T2}$ by itself is not able to eliminate this bias. Thus, it still suggested that ε be included in probabilistic performance assessments. The increased coefficient of variation of estimation of $\lambda_{EDP}(z)$ due to ε could be avoided by using ε-based record selection (see Chapter 6), rather than random selection with vector-valued *Im*s.

The results from this section suggest that use of a vector-valued *IM* consisting of $S_a(T_1)$ and $R_{T1,T2}$ does not desensitize the analysis to the effects of ε. Although ε and $R_{T1,T2}$ are both effective

[9] For loss assessment, ground motions with mean annual rates of exceedance of 10^{-2} may also be important, and at this hazard level, ground motions are more likely to have mean epsilon values close to zero. Thus, the effect of ε may be less important for loss estimation.

IM parameters because they are related to spectral shape, they are each describing somewhat different properties of the shape of the spectrum. Further, these two parameters have different effects on estimated drift hazard curves: a vector *IM* $R_{T1,T2}$ tends to reduce slightly the uncertainty in the curve relative to use of a scalar *IM*; a vector *IM* with ε tends to eliminate a bias in the drift hazard, while actually increasing the uncertainty in the drift hazard curve[10]. The parameter ε should thus still be accounted for, as its effect is significant, especially at large levels of response. This is seen in Chapter 3 and in the results of Figure 4.18.

4.9 Conclusions

A method for selecting an efficient vector intensity measure using regression analysis has been presented. This method is based on scaling to the first *IM* element ($S_a(T_1)$) and then using regression on the second element ($R_{T1,T2}$) to predict the response of the structure. The optimal second period (T_2) for use in the vector was chosen by minimizing the standard deviation of the prediction errors. It was shown that the proper choice of $R_{T1,T2}$ can significantly reduce prediction error when compared to prediction using $S_a(T_1)$ alone. A range of nonlinear MDOF structures were tested to find the optimal second period for use in this *IM*. When the building response was linear or when nonlinear building response has a significant contribution from the second mode of vibration, the optimal second period for use in $R_{T1,T2}$ was shown to be approximately equal to the second-mode period of the building. When $S_a(T_1)$ was large enough to cause nonlinear behavior and second-mode response was not critical, the optimal second period was larger than the first-mode period of the building, and increased as the level of nonlinearity increased. The lengthened period appears to be related to the effective period of an equivalent linear system. Further, increasing the ductility of a structure tends to decrease its sensitivity to long period T_2 values, relative to short period T_2 values.

A method for evaluating vector *Im*s that utilizes bootstrap replications of the drift hazard curve was also presented. This method has the advantage of directly computing the statistical variability in estimates of the drift hazard curve, and it accounts for the increased prediction efficiency of both collapse and non-collapse responses at many *IM* levels simultaneously. The disadvantage of this method is that it requires a vector-valued ground motion hazard for each candidate *IM*, and it also requires slightly increased computational time. Because of this, it is suggested that the regression analysis method be used to narrow down a broad range of potential

[10] This conclusion is based on the currently common practice of ignoring ε in selecting records. As will be seen in Chapter 6, ε-sensitive record selection can also address this source of bias and may also avoid inflating the uncertainty in $\lambda_{EDP}(z)$ by avoiding regression extrapolation.

vector *Im*s to a few promising candidates. The bootstrap method can then be used to examine these few in detail.

A three-parameter *IM* consisting of $S_a(T_1)$, $R_{T1,T2}$ and ε was also considered, in order to determine whether $R_{T1,T2}$ accounted for the effect of ε discussed in Chapter 3. It was found that $R_{T1,T2}$ does not fully account for the effect of ε, and that neglecting ε in analysis results in conservative estimates of the annual rate of exceeding a given maximum interstory drift ratio, especially at large levels of structural response. This unconservative bias is consistent with the effect seen in Chapter 3. So although the parameter $R_{T1,T2}$ produces significant increases in estimation efficiency, it does not account for the effect of ε.

A vector-valued intensity measure consisting of $S_a(T_1)$ and $R_{T1,T2}$ has the potential to produce a drift hazard curve with significantly narrower confidence bands than the equivalent curve computed using the scalar intensity measure $S_a(T_1)$ and the same number of nonlinear analyses. Analysis of the structure evaluated in this paper shows the potential for a reduction in the standard deviation of $\lambda_{EDP}(z)$ of as much as a factor of two. This implies that in principle the required number of analyses could be reduced by a factor of as much as four without increasing the standard deviation of the $\lambda_{EDP}(z)$ result. This decrease in computational expense is very appealing, and may justify the use of the vector intensity measure.

The subjects of vector *Im*s, estimates of uncertainty in $\lambda_{EDP}(z)$, and record selection with respect to ε are all new. The joint effects of these topics, such as $\lambda_{EDP}(z)$ estimation uncertainty when ε-based record selection is used, require further study.

Chapter 5

Vector-valued intensity measures for pulse-like near-fault ground motions

5.1 Introduction

Pulse-like near-fault ground motions are a special class of ground motions that are particularly challenging to characterize for earthquake hazard assessment. These motions are characterized by a “pulse” in the velocity time history of the motion, in the direction perpendicular to the fault (see Figure 5.1). It is particularly important to characterize these ground motions, because they are typically very intense and have been observed to cause severe damage to structures in past earthquakes.

The source of the severe damage has been divided into two parts. First, it has been noted that these motions have, on average, larger elastic spectral acceleration values at moderate to long periods. This has been accounted for by modifying ground motion prediction (“attenuation”) models (e.g., Somerville et al. 1997), and improved modifications are currently in development. Second, it has been noted that these motions cause severe response in nonlinear multi degree of freedom structures (e.g., Mahin et al. 1976, Bertero et al. 1978. A more detailed history of the research surrounding this topic is provided by Alavi and Krawinkler 2001.). Further, this severe response is not entirely accounted for by measuring the intensity of the ground motion using spectral acceleration of the elastic first-mode period of a structure ($Sa(T_1)$), as has been noted by several researchers (e.g. Mehanny and Deierlein 2000, Baez and Miranda 2000, Alavi and Krawinkler 2001, Ruiz-Garcia 2004, Luco and Cornell 2005).

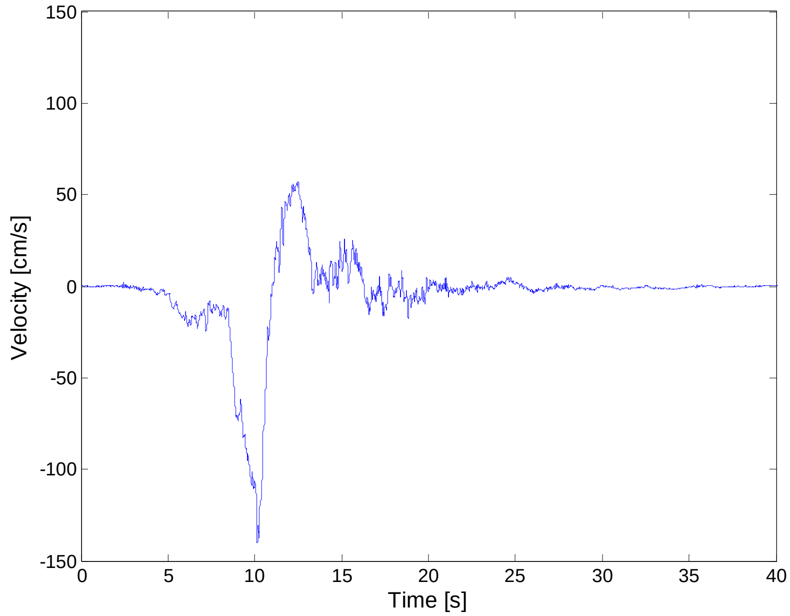


Figure 5.1. The velocity time history of the fault-normal horizontal ground motion recorded at Lucern during the 1992 Landers earthquake. A “pulse” is clearly present.

Given these challenges, Luco and Cornell (2005) suggest that when using the probabilistic performance assessment procedure used throughout this dissertation, the intensity measure $Sa(T_1)$ has shortcomings in capturing the effect of pulse-like ground motions. Luco and Cornell (2005) have proposed a replacement scalar IM for use with pulse-like ground motions, which is a combination of inelastic spectral displacement at the first-mode period and elastic spectral displacement at the second-mode period. As an alternative, the vector-valued Ims considered in Chapters 3 and 4 may be effective at assessing the effect of these motions. In Chapters 3 and 4, only “ordinary” ground motions were considered, while here pulse-like ground motions will be considered as well. The ability of these intensity measures to characterize pulse-like ground motions will be evaluated. Analysis of vector-valued probabilistic ground motion hazard will require some modifications for use with pulse-like ground motions, and these modifications are discussed as well.

5.2 Pulse-like ground motions

The earthquake engineering community has recognized in recent years that sites located near an earthquake fault rupture may experience ground shaking that includes a velocity “pulse.” This pulse is most likely to occur in specific site-source geometrical configurations (see, e.g., Somerville et al. 1997). Roughly speaking, a velocity pulse is likely to occur in the fault-normal direction at sites where the earthquake rupture is propagating towards the site rather than away from it. Ground motion shaking in the fault-parallel direction is typically less intense. In the discussion that follows, the term "pulse-like ground motion" is used to refer to fault-normal ground motions with an observed velocity pulse, typically occurring within 20 or 30 km of the fault.

A set of 70 fault-normal near-fault ground motions collected by Tothong and Cornell (2005a) is utilized in this study. This set is an aggregation of records identified by Mavroeidis and Papageorgiou (2003), Fu and Menun (2004), and Luco (2002). All of the records from these studies recorded on firm soil or rock and including clearly identifiable velocity pulse are included. The processed ground motion time histories come from the Next Generation Attenuation project database (2005). A particularly important property of pulse-like ground motions is the period of the velocity pulse (denoted T_p); following Alavi and Krawinkler (2001), Tothong and Cornell (2005a) have defined T_p as the period associated with the maximum of the velocity response spectrum. A table of the records and their associated properties is presented in Appendix A, Table A.4.

A set of 40 "ordinary" ground motions (i.e., ground motions with no velocity pulses) is also used for comparison with the pulse-like record set. The properties of these records are given in Appendix A, Table A.3.

5.3 Spectral acceleration as an intensity measure

At large periods, pulse-like ground motions tend to cause larger elastic spectral acceleration (*Sa*) levels than standard ground motion (attenuation) models predict. A model for this effect was proposed by Somerville et al. (1997), in the form of a modification to a popular ground motion prediction model (Abrahamson and Silva 1997), with the intention that the correction could also be applied to other prediction models as well. This model adjusts a broad range of spectral acceleration values at periods greater than 0.6 seconds. Future "narrow-band" models may only adjust a smaller range of spectral acceleration values depending upon the period of the velocity pulse, which is related to the magnitude of the earthquake (Somerville 2003).

Regardless of the exact form of the ground motion prediction model, it is clear that *Sa* values tend to be larger, at least at longer periods, for ground motions with velocity pulses than for ordinary ground motions with similar magnitudes and distances. The question to be answered then is whether the larger *Sa* values completely account for the larger structural responses observed from these records, or whether there is an additional effect from pulse-like ground motions that is not described by the *IM* $Sa(T_1)$. It will be seen below that $Sa(T_1)$ does not completely account for the effect of these ground motions, and so vector-valued *Im*s will be used to better quantify the effects of these ground motions.

5.4 Structural response from pulse-like ground motions

To quantify the effect of pulse-like ground motions, a set of multi-degree-of-freedom nonlinear structures are used for evaluation. A set of generic frame structures designed by Ibarra (2003) is used for analysis. All of the structures are single-bay frames, with stiffnesses and strengths chosen to be representative of typical structures. Four structures were considered, with varying numbers of stories and first-mode periods. Their properties are summarized in Table 5.1. Peak interstory drift ratios in these structures are known to be significantly affected by second-mode response. This will help demonstrate the ability of a vector IM to account for higher-mode response, but the effect of higher-mode response is likely less significant for most typical structures with comparable numbers of stories and/or comparable first-mode periods.

Table 5.1. Model parameters for the four generic-frame structures considered in this chapter.

Elastic first-mode period (T_1)	Number of stories (N)	Ductility capacity (δ_c/δ_y)	Post-capping stiffness coefficient (α_c)	Cyclic deterioration parameters ($\gamma_{s,c,k,a}$)
0.3	3	4	-0.5	50
0.9	9	4	-0.5	50
1.2	6	4	-0.5	50
1.8	9	4	-0.5	50

In order to summarize results from these various structures, structural response data is computed for ground motions scaled such that the records' $Sa(T_1)$ level in units of g is a specified multiple of the structure's base shear coefficient γ (where γ=yield base shear/weight). This ratio of first-mode spectral acceleration to base shear coefficient is analogous to an R-factor in present building codes, if there were no overstrength in the structure. Here, this normalized spectral acceleration value will be referred to as R_μ. Using this normalized ground motion intensity measure, all four structures will yield at R_μ factors of approximately one (recognizing that higher-mode response may or may not induce yielding at this R_μ level). Increasing R_μ levels will represent increasing levels of nonlinearity in the structures.

Several levels of $Sa(T_1)$ (or, equivalently, R_μ) were considered for each structure. At each $Sa(T_1)$ level, all of the ground motions were scaled so that they had the same $Sa(T_1)$ value. They were then input into the structure in order to compute structural response. For this study, the structural response parameter of interest is the maximum interstory drift ratio observed in any story, as it is a good indicator of the ability of a structure to resist P-Δ instability and collapse, as well as maximum rotation demands on beams, columns and connections (FEMA 2000a).

An important principle for all of the response studies that follow is that not all pulse-like ground motions are equally severe with respect to a given structure. One important factor that affects structural response is the period of the velocity pulse with respect to the modal periods of the structure (Alavi and Krawinkler 2001, Fu 2005, Tothong and Cornell 2005a), as is illustrated in Figure 5.2. On the y axis, the maximum interstory drift ratio of the 9 story structure with a period of 0.9 seconds is plotted for each of the 70 pulse-like ground motions. On the x axis, the period of each record's velocity pulse (T_p) divided by the first-mode period of the structure (T_1) is plotted[11]. In addition to the individual data points, a general trend is plotted that was computed using the Nadaraya-Watson kernel-weighted average, with a tri-cube kernel weight function and an adaptive window that includes the 10 nearest neighbors (Hastie et al. 2001). All of this data is shown for ground motions scaled such that the records' $Sa(T_1)$ level in units of g is 4 times the structure's base shear coefficient γ, where γ=yield base shear/weight (i.e., the R_μ factor is equal to 4). Note also that three records are highlighted in Figure 5.2, and their associated response acceleration and velocity spectra are shown in Figure 5.3. The velocity spectrum ($Sv(T) = Sa(T) \cdot (T/2\pi)$) is plotted because the periods of the near-fault pulses are most clear in this figure. Use of spectral velocity values as the IM parameters would give the same results as the spectral acceleration values, but spectral acceleration values are used for consistency with standard ground motion prediction models and hazard maps.

[11] Some of the near-fault records have T_p values as large as 9 seconds, but here is assumed that when T_p/T_1 is greater than 5, the "pulse-like" properties of the records are not as important with regard to estimating structural response. Further, no clear trends were observed between T_p/T_1 and structural response for these records. For this reason, Figure 5.2 and other similar figures are truncated at T_p/T_1=5.

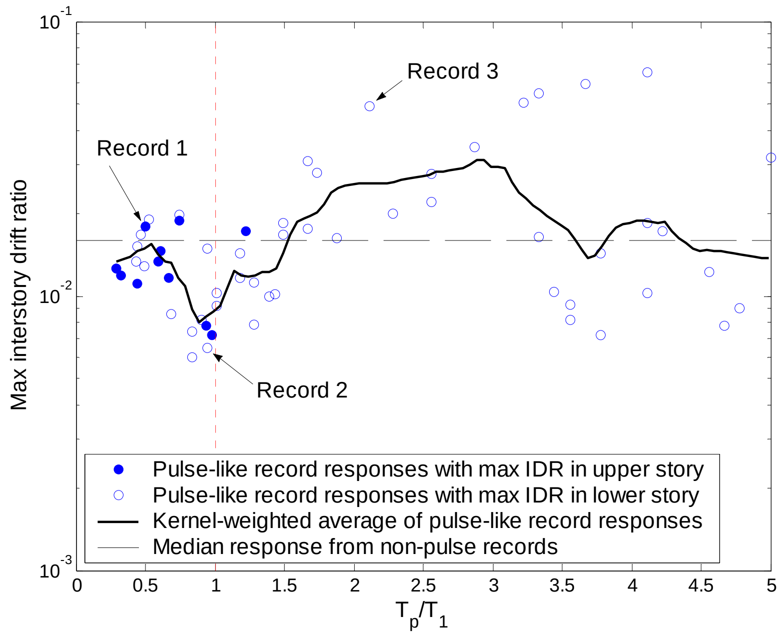
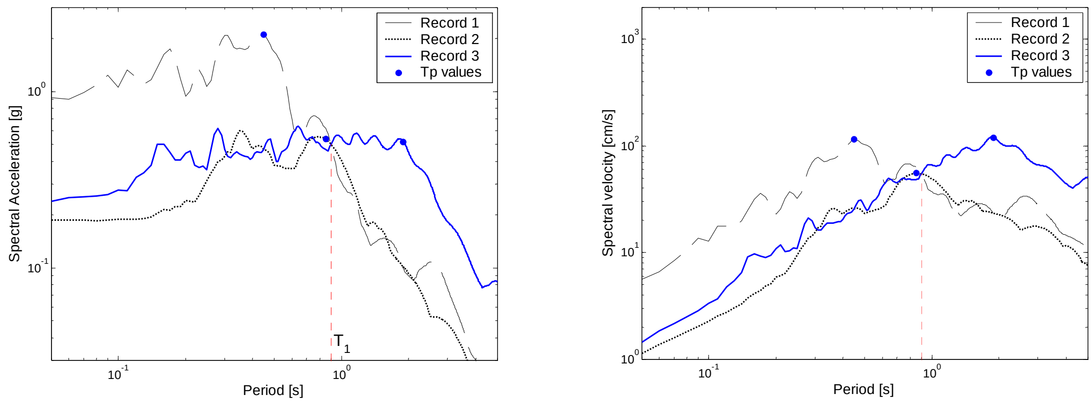


Figure 5.2. Maximum interstory drift ratio versus T_p/T_1 for the generic frame with 9 stories and a first-mode period of 0.9 seconds, at an R_μ-factor level of 4. Record 1: Morgan Hill, Anderson Dam (Magnitude = 6.2, Distance = 3km). Record 2: Kobe, KJMA (Magnitude = 6.9, Distance = 1km). Record 3: Superstition Hills, Parachute Test Site (Magnitude = 6.5, Distance = 1 km).

Some trends apparent in this figure for this particular structure and R_μ value are observed systematically over a range of structures and R_μ factors, and these trends are consistent with the observations of earlier researchers. For T_p/T_1 values larger than 1.5 or 2, the response is relatively large compared to the response from records with shorter-period pulses. This is because $Sa(T_1)$ is only measuring the intensity of the ground motion at T_1. As the structure behaves nonlinearly and its effective period lengthens, it is greatly affected by the velocity pulse at the longer period (see the response spectrum of Record 3 in Figure 5.3). Conversely, the minimum responses are observed for T_p/T_1 values of approximately 1. In this case a record's $Sa(T_1)$ value will be large because of the energy from the pulse with a period of approximately T_1, implying that the record is very intense as measured by $Sa(T_1)$. But as the structure begins to behave nonlinearly, its period lengthens into a range where there is comparatively lesser energy (see the response spectrum of Record 2 in Figure 5.3). Finally, for T_p/T_1 values of approximately 0.3, the pulse falls at a period that excites the higher mode of the structure[12], although the IM $Sa(T_1)$ cannot detect it. Thus, records with T_p/T_1 values in this range can also cause large responses (see the response spectrum

[12] The second mode of this structure has a period of $0.39T_1$. Maximum interstory drift ratios in this structure are particularly sensitive to higher-mode response (Krawinkler, 2005), making the effect of short-period pulses particularly pronounced for this example.

of Record 1 in Figure 5.3). The effect of short-period pulses can also be confirmed by noting that for records with T_p/T_1<1 the maximum responses are often observed in the upper stories (as shown in Figure 5.2), indicating that the higher modes of vibration are contributing significantly to response. Conversely, for T_p/T_1 values greater than one, maximum responses nearly always occur in the lower stories, indicating that first-mode response is controlling peak displacements.



(a) (b)

Figure 5.3. (a) Acceleration, and (b) velocity spectra of the three highlighted records from Figure 5.2, after scaling each record so that Sa(0.9s) = 0.5g.

The velocity time histories of these three records are shown in Figure 5.4, to illustrate the velocity pulses present in these ground motions.

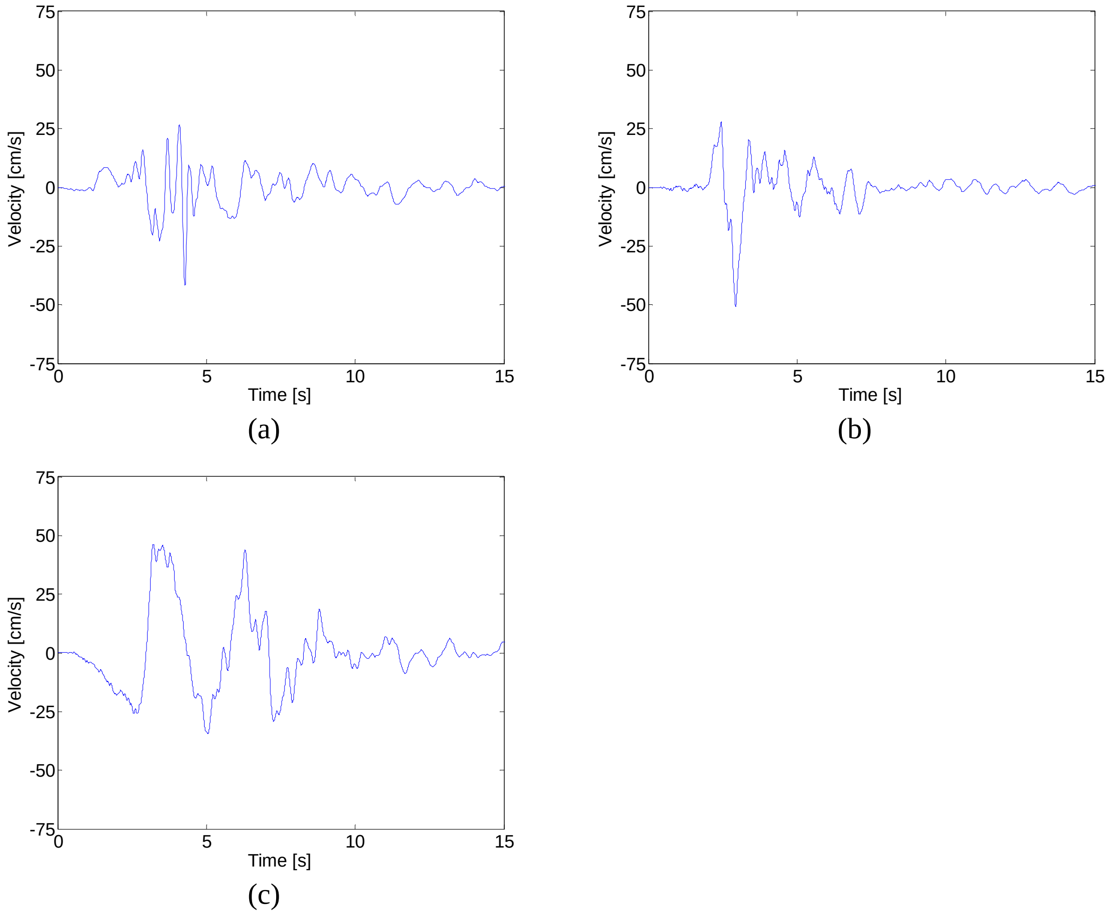


Figure 5.4. Velocity time histories of the three example pulse-like ground motions, after scaling each so that Sa(0.9s)=0.5g. (a) Record 1: Morgan Hill, Anderson Dam (T_p = 0.4s, Magnitude = 6.2, Distance = 3km). (b) Record 2: Kobe, KJMA (T_p = 0.9s, Magnitude = 6.9, Distance = 1km). (c) Record 3: Superstition Hills, Parachute Test Site (T_p = 2.9s, Magnitude = 6.5, Distance = 1 km).

The effect of pulses on higher modes is also seen in Figure 5.5, which is identical to Figure 5.2 except that the records have been scaled to an R_μ factor of 2. At this R_μ factor, structural nonlinearity is minor and so T_p/T_1 values larger than one do not affect the structure significantly. However, T_p/T_1 values near the second-mode period of the structure do cause much larger responses, and peak responses always occur in the upper stories of the structure, indicating the effect of higher modes.

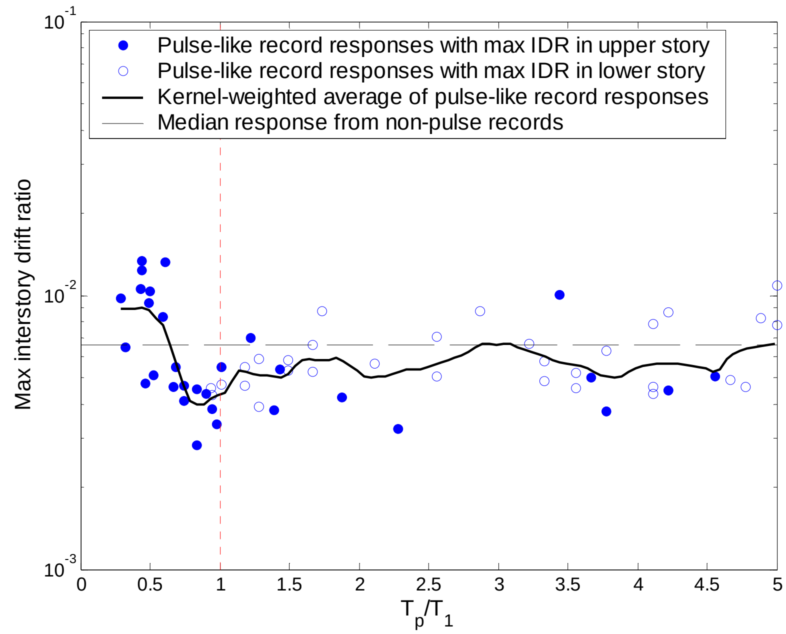


Figure 5.5. Maximum interstory drift ratio versus T_p/T_1 for the generic frame with 9 stories and a first-mode period of 0.9 seconds, at an R_μ level of 2.

The results of Figure 5.2 and Figure 5.5 come from records scaled to given $Sa(T_1)$ levels. If one uses $Sa(T_1)$ as an intensity measure, the assessment of structural response is much more robust (e.g., insensitive to the ground motions selected for analysis) if other properties of a ground motion do not affect structural response, given $Sa(T_1)$. This condition is termed "sufficiency" by Luco and Cornell (2005). As seen in Figure 5.2 and Figure 5.5, $Sa(T_1)$ is not sufficient with respect to T_p (i.e., given $Sa(T_1)$, a record's ground motion property T_p is still a useful predictor of structural response). Without sufficiency, the results of the $Sa(T_1)$-based probabilistic response assessment are affected by the particular ground motions used for analysis, and so currently the procedure has problems when applied to structures located at sites where pulse-like ground motions may occur. If a vector-valued intensity measure can be shown to be sufficient with respect to T_p, then it would be a significant improvement over $Sa(T_1)$ and would allow the probabilistic response assessment to proceed as before. It is this possibility that will be investigated below.

Given knowledge of the effect of T_p/T_1, two subsets of the pulse-like records are also compared with the ordinary ground motions. The records with T_p/T_1>2 are separated and labeled the "worst case" pulse-like ground motions, and the records with 0.5<T_p/T_1<1.5 are separated and labeled "best case" pulse-like ground motions[13]. These records are compared to the complete set of pulse-like ground motions, and to the ordinary ground motions with no velocity pulse. In Figure 5.6 the counted median of the maximum interstory drift ratio is plotted versus R_μ for all four groups of ground motions. In Figure 5.7, the probability of collapse is plotted for the same four groups of ground motions, where collapse of these deteriorating structures is indicated by large interstory drift ratios that cause non-convergence of the analysis program (Ibarra 2003). In general, some trends are apparent. The worst case pulse-like records have the largest median responses for R_μ factors greater than 2, and greater probabilities of collapse for R_μ factors greater than 6. The best case pulse-like ground motions have the smallest median responses and the smallest probabilities of collapse at all R_μ factor levels. However, the set of all pulse-like records does not have median responses that differ substantially from the ordinary ground motions at most R_μ levels. The median probabilities of collapse are nearly equal for the ordinary ground motions, and the set of all pulse-like ground motions. But the pulse like ground motions collapse distribution has heavier tails, due to the presence of the best-case and worst-case ground motions, which cause greater record-to-record variability than the ordinary ground motions.

Clearly there are some pulse-like ground motions that cause relatively severe responses in a structure and some that do not. If a vector-valued IM was able to supplement $Sa(T_1)$ with information that would distinguish between the severe and non-severe records, then the sufficiency problems of $Sa(T_1)$ would be addressed and standard probabilistic performance assessment would be feasible even when pulse-like ground motions are considered.

[13] A "best case" ground motion is likely have a naturally large $Sa(T_1)$ values due the presence of a pulse with a period close to T_1. But a response prediction based on this (large) $Sa(T_1)$ value will tend to overestimate the response from this record, because spectral values at other periods are likely not as large. Thus, *given* $Sa(T_1)$, a best case record will tend to have smaller responses than other records, and that is why it is called a "best case" ground motion.

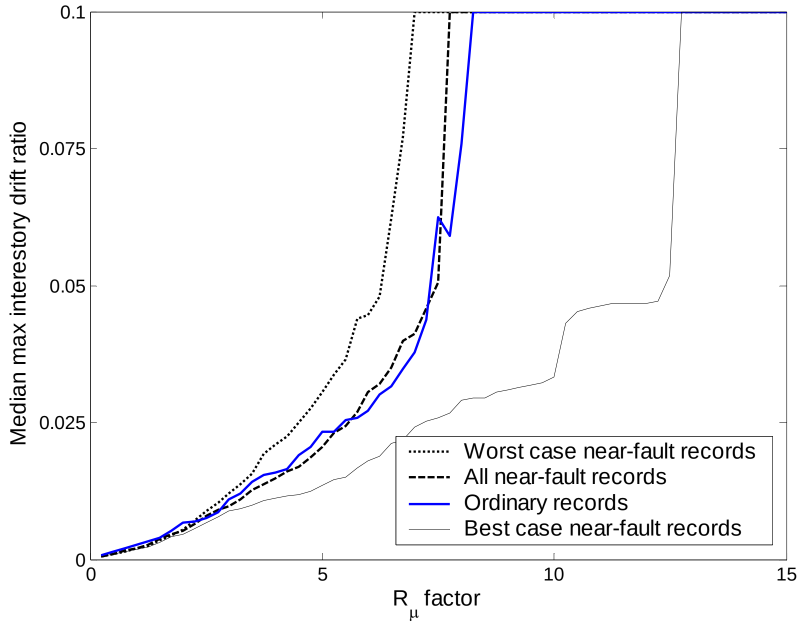


Figure 5.6. Median maximum interstory drift ratio versus normalized spectral acceleration (R_μ) for the generic frame with 9 stories and a first-mode period of 0.9 seconds.

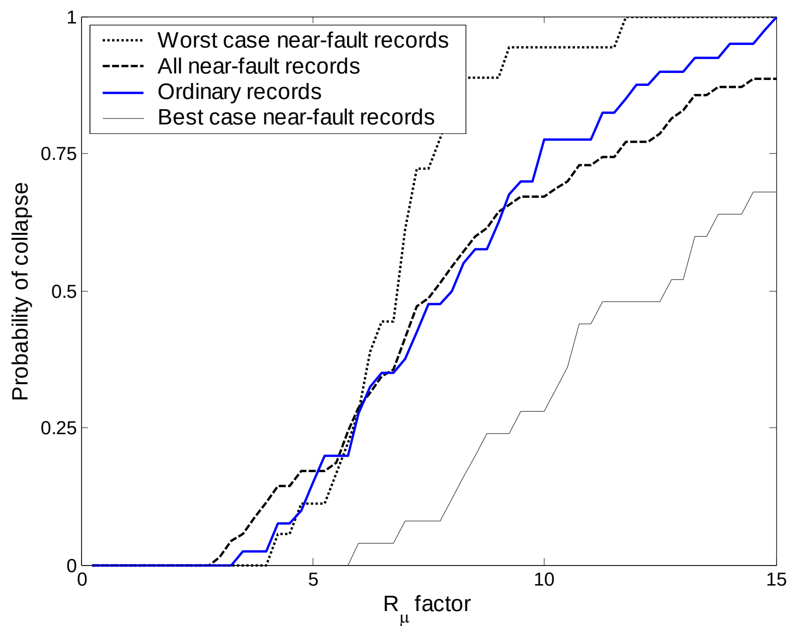


Figure 5.7. Counted probabilities of collapse versus normalized spectral acceleration for the frame with 9 stories and a first-mode period of 0.9 seconds.

5.5 Vector-valued Ims for prediction of response of pulse-like ground motions

Two ground motion parameters, ε and $R_{T1,T2}$, have been considered in Chapters 3 and 4 for inclusion in a vector IM, and both were found to be useful for prediction of response from ordinary ground motions. Here they will be evaluated for prediction of response from pulse-like ground motions. Specifically, they will be evaluated to see if they have fixed the sufficiency problems with respect to T_p.

5.5.1 A vector-valued IM with Sa(T$_1$) and ε

The ground motion parameter epsilon (ε) is a measure of the difference between the spectral acceleration of a record and the mean value of a ground motion prediction model at the given period. Epsilon is an indirect measure of spectral shape (specifically, it tends to indicate whether $Sa(T_1)$ is in a peak or a valley of the spectrum at a given period), and so it is an effective predictor of structural response. In Chapter 3, it was seen that using ε in a vector IM with $Sa(T_1)$ can reduce potential biases in prediction of structural response from ordinary ground motion records. Here it will be considered for prediction of response from pulse-like ground motions.

When computing ε values of pulse-like ground motions, the mean predicted spectral acceleration value should account for near-fault effects that are on average present in the ground motions. Somerville et al. (1997) proposes a modification factor for standard prediction models (which typically do not consider pulse-like motions) in order to account for these pulse-like effects. Thus, for the investigation of this section, ε values are computed based on the ground motion prediction model of Abrahamson and Silva (1997), with modification as specified by Somerville et al. (1997). The effect of this modification is discussed in Appendix C.

Before proceeding to examine the relationship between ε and T_p, it should be verified that ε predicts structural response in the same way for both ordinary and pulse-like records. To perform this test, we compare two prediction alternatives. The simpler alternative is that ordinary and pulse-like ground motions cause the same mean logarithmic structural response as a function of ε. Here the structural response parameter (which is termed an Engineering Demand Parameter, or *EDP*, by the Pacific Earthquake Engineering Research Center) is maximum interstory drift ratio. The predictive equation for this proposed model is

$$E[\ln EDP \mid \ln Sa(T_1), \varepsilon] = a + b \ln \varepsilon \tag{5.1}$$

where a and b are coefficients to be estimated from regression analysis used to predict ln*EDP* as a function of ε using records scaled to a specified level of $Sa(T_1)$. A fitted prediction using this model is shown in Figure 5.8a for an $Sa(T_1)$ level such that the structure has an R_μ factor value of 4.

The more complicated alternative is that records with and without a pulse-like velocity pulse have differing functional relationships between ε and mean structural response. The predictive equations for this alternative could be expressed as

$$\begin{aligned} E[\ln EDP \mid \ln Sa(T_1), \varepsilon, \text{no pulse}] &= a + b \ln \varepsilon \\ E[\ln EDP \mid \ln Sa(T_1), \varepsilon, \text{pulse}] &= c + d \ln \varepsilon \end{aligned} \tag{5.2}$$

where a and b are coefficients to be estimated from regression on *ordinary* records scaled to a specified level of $\ln Sa(T_1)$, and c and d are coefficients to be similarly estimated using *pulse-like* records. Equation 5.2 can also be stated in the following form, which is convenient for regression analysis

$$E[\ln EDP \mid \ln Sa(T_1), \varepsilon, \text{pulse}] = a + b \ln \varepsilon + (c - a) I_{NF} + (d - b) I_{NF} \ln \varepsilon \tag{5.3}$$

where I_{NF} is an indicator variable equal to 1 if the given record has a near-fault velocity pulse and equal to 0 otherwise. A fitted prediction using this model is shown in Figure 5.8b.

A classical statistical test referred to as an F test (Neter et al. 1996) can be used to choose between the models of Equations 5.1 and 5.3. The simpler model of Equation 5.1 is assumed to be the truth until evidence is found to the contrary (this model is called the null hypothesis). The model of Equation 5.3 is tested, and if the prediction errors using this model (called the alternative hypothesis) are significantly smaller than the errors from Equation 5.1, this is taken as evidence that the more complex model is appropriate. An F statistic is computed by comparing the aggregate prediction errors from the two models, and a “p-value” associated with the statistic provides the probability that the apparent improvement from the null hypothesis to the alternative hypothesis would be observed, even though (i.e., “given that”) the null hypothesis is actually true. Low p-values indicate significant improvement using the more complex model, and suggest that this model should be adopted. P-values less than 0.05 are usually taken as support for the more complex model. In the example of Figure 5.8, the associated p-value is 0.49, suggesting that little predictive accuracy was gained by using the more complex model and thus the simpler prediction associated with Figure 5.8a (and Equation 5.1) is the appropriate model. P-values for a range of structures and *Sa* levels are tabulated in Table 5.2. Even if the simple model holds, 5% of p-values will be below 0.05. In Table 5.2, approximately 4% of the tests have p-values below 5%, providing strong support for the simpler model of Equation 5.1: no distinction need be made between pulse-like and ordinary records when predicting response based on ε. This is important because one of the desirable attributes of an improved IM is that an analyst should not have to distinguish between pulse-like and ordinary records when predicting response as a function of

IM. In different words, predictions based on such an IM will be robust to the records selected for analysis.

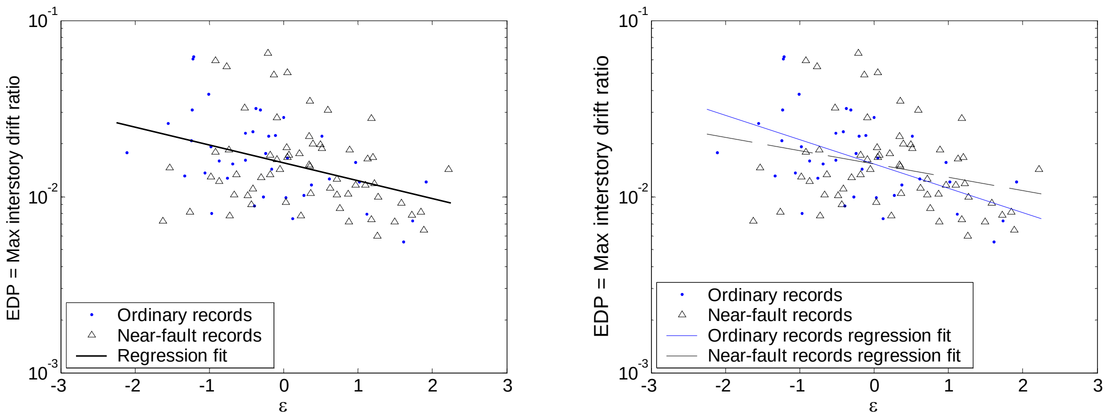


Figure 5.8. Prediction of response as a function of ε using linear regression on records scaled to have an R_μ factor level of 4. (a) Estimate of ordinary and pulse-like ground motion responses using the same prediction equation. (b) Estimate of ordinary and pulse-like ground motion responses using separate prediction equations.

Table 5.2. P-values from F tests to test the hypothesis that ε has a different effect on responses to pulse records and non-pulse records, for each of the four structures listed in Table 5.1. The lack of statistically significant results suggests that it does not. Results are omitted for R_μ levels where more than 50% of the records caused collapse.

	Structure			
R_μ Factor	**N=3, T=0.3**	**N=9, T=0.9**	**N=6, T=1.2**	**N=9, T=1.8**
0.5	0.48	0.43	0.59	0.24
1.0	0.48	0.52	0.54	0.20
1.5	0.56	0.28	0.18	0.25
2.0	0.24	0.66	0.65	0.65
2.5	0.82	0.95	0.97	0.30
3.0	0.40	0.63	0.83	**0.05**
3.5	**0.05**	0.65	0.70	0.27
4.0		0.49	0.50	0.63
4.5		0.59	0.63	0.95
5.0		0.90	0.91	0.56
5.5		0.73	0.89	
6.0		0.70	0.73	
6.5		0.34		
7.0		0.51		
7.5		0.55		

Epsilon and sufficiency with respect to Tp

The next test that must be performed when selecting a vector-valued IM for pulse-like ground motions is to test for T_p sufficiency. It was seen in Figure 5.2 and Figure 5.5 that $Sa(T_1)$ is unable to fully account for the effect of the pulse period (T_p) of the pulse-like ground motions, but perhaps the effect of T_p could be accounted for by adding the parameter ε to the vector. In order to test this for the IM consisting of $Sa(T_1)$ and ε, the prediction residuals are studied. First, consider the scalar IM case of $Sa(T_1)$ again. Given an $Sa(T_1)$ level, the median structural response obtained from the record set forms the median structural response prediction (under the common assumption that structural response given IM is lognormally distributed, the median can be estimated using the geometric mean of the data, or the mean of the logarithmic structural responses). Thus, the difference between an individual record's associated response value and the median response value of all records is a prediction error (termed a *residual* in statistical analysis). These residuals are shown in Figure 5.9 for the 9-story structure with a first-mode period of 0.9s, and $Sa(T_1)$ values such that the structure has an R_μ factor level of 4. The data in this plot are identical to the data from Figure 5.2, only the y axis values have been rescaled based on the mean log response value (0.153) of all of this data. Again it is clear that there is a systematic trend with T_p/T_1.

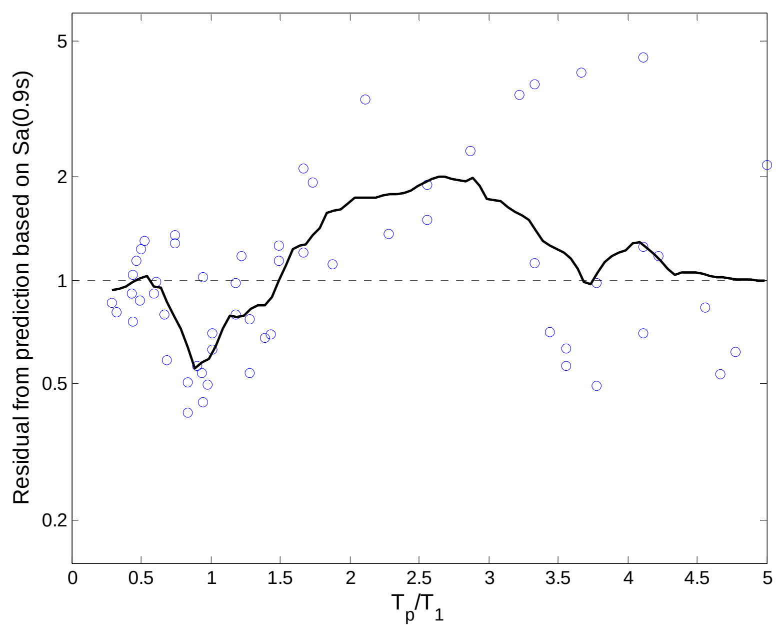


Figure 5.9. Residuals from response prediction based on $Sa(T_1)$ plotted versus T_p/T_1 for the generic frame with 9 stories and a first-mode period of 0.9 seconds, at an R_μ factor level of 4.

The equivalent test that can be performed based on a vector IM with $Sa(T_1)$ and ε is to compare the prediction residuals based on that IM to T_p/T_1. In this case, the prediction residuals[14] are obtained from the prediction of Equation 5.1: that is, the distance from each response data point in Figure 5.8a to the regression fit in that figure. These new residuals are plotted vs T_p/T_1 in Figure 5.10. Unfortunately there is little improvement in the dependence of the residuals vs T_p/T_1: ε has *not* accounted for the effect of T_p on structural response.

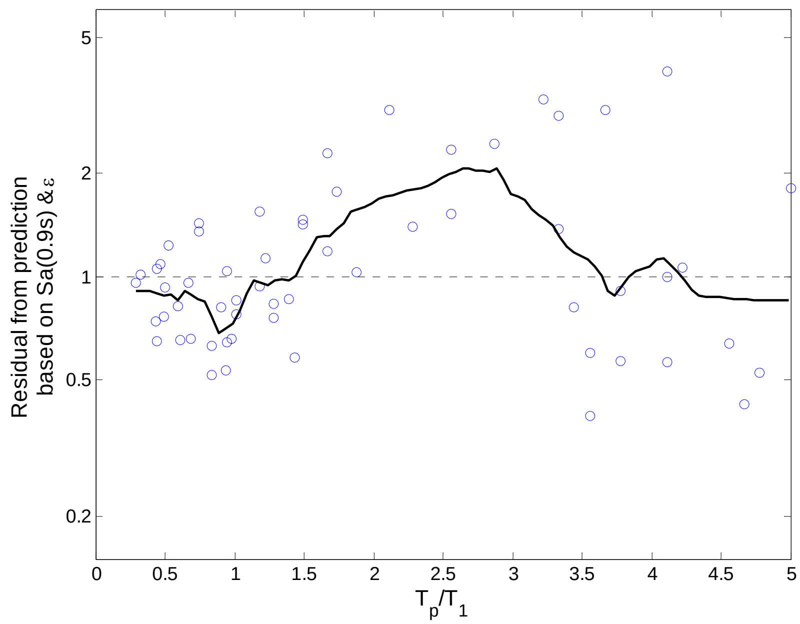


Figure 5.10. Residuals from response prediction based on $Sa(T_1)$ and ε plotted versus T_p/T_1 for the generic frame with 9 stories and a first-mode period of 0.9 seconds, at an R_μ factor level of 4. No significant reduction in bias is seen, relative to Figure 5.9.

Further insight into why ε could not account for the effect of T_p is provided in Figure 5.11. When T_p/T_1=1 (or equivalently T_p=T_1), ε tends to be positive. This is because when T_p=T_1, the elastic spectral acceleration value is likely to be unusually large, and thus ε would indicate that the record has a "peak" in the spectrum at T_1. Because of this, the vector with ε does improve the prediction error at $T_p/T_1\approx1$ (if the region near T_p/T_1=1 is compared in Figure 5.9 and Figure 5.10, it is seen that the residuals tend to be closer to one, and thus less biased, in Figure 5.10). However, ε does not vary with other T_p/T_1 values, and thus is not able to predict the effect of records with varying T_p values. Critically, ε is not able to identify pulses in the important range

[14] Because the predictions were made for log response values, the residual can be computed as the ratio of non-log actual response divided by predicted response, and then plotted on logarithmic paper. This is equivalent to plotting the differences of the log predictions on non-log paper, and the units on the y axis may be more intuitive when non-log response values are shown.

near T_p/T_1=2. To do so, ε would need to be negative when $T_p/T_1 \approx 2$ (i.e., there would need to be a reason why Sa values at $T_p/T_1 \approx 2$ were less than predicted by Somerville 1997); this is clearly not the case.

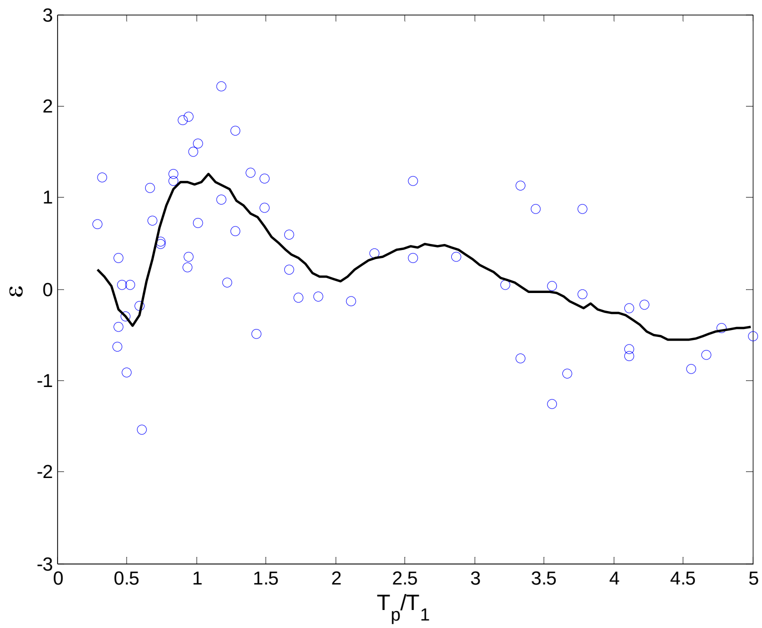
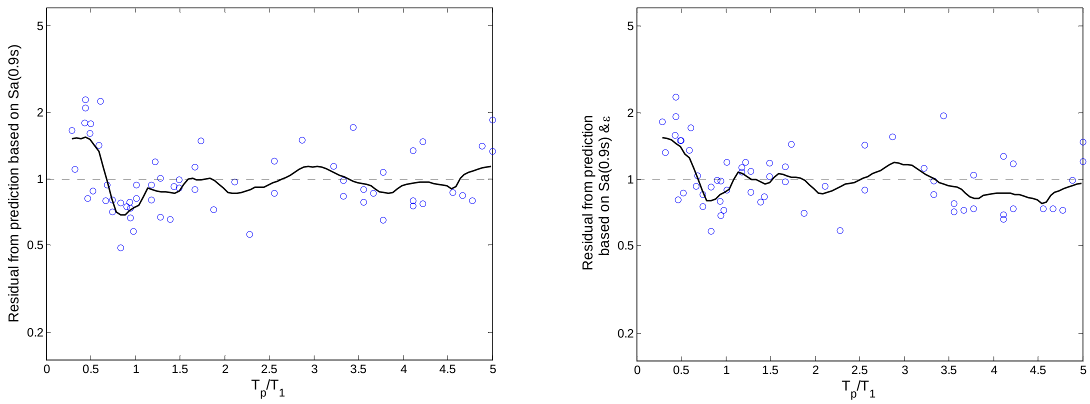


Figure 5.11. Epsilon values at 0.9s plotted versus T_p/T_1.

This same test can be performed at other spectral acceleration levels as well. In Figure 5.12, another comparison is made between the residual predictions using either $Sa(T_1)$ alone or $Sa(T_1)$ and ε as predictors. In this figure, the test is performed on the same structure as above, but for an R_μ factor of two. Again it is seen that adding the parameter ε to the vector did not significantly improve the Ims ability to account for the effect of pulse period. Further examination of all of the structures listed above at many levels of ground motion intensity suggests that ε has no significant ability to account for the effect of T_p. Therefore, a vector IM consisting of $Sa(T_1)$ and ε should *not* be considered an effective IM for use with pulse-like ground motions.



(a) (b)

Figure 5.12. Residuals from response prediction based on (a) $Sa(T_1)$ and (b) $Sa(T_1)$ and ε plotted versus T_p/T_1 for the generic frame with 9 stories and a first-mode period of 0.9 seconds, at an R_μ level of 2. No significant reduction in bias is seen when ε is incorporated in the intensity measure.

5.5.2 A vector valued IM with Sa(T_1) and $R_{T1,T2}$

A vector-valued IM based on spectral acceleration values at two periods was proposed in Chapter 4 and found to be an effective predictor for ordinary ground motions. The IM consists of the parameters $Sa(T_1)$ and $R_{T1,T2}=Sa(T_2)/Sa(T_1)$, where T_1 is constrained to equal the first-mode period of the structure, but T_2 can be chosen freely. This intensity measure can provide information about excitation of higher modes or information about the severity of response once the structure becomes nonlinear and its effective period lengthens. Because the effect of the velocity pulse period is related to these same effects, this vector was expected to be more effective than the vector with ε at accounting for near-fault effects.

As with the previous vector IM, statistical tests are first performed to determine whether the relationship between $R_{T1,T2}$ and structural response is the same for both ordinary and pulse-like ground motions. In this case, however, T_2 must be specified before the test can be performed. In much of the following investigation, T_2 will be specified as twice the elastic first-mode period. This was seen to be an effective predictor for nonlinear MDOF structures in Chapter 4, and twice the elastic first-mode period is a period range of particular concern for pulse-like ground motions as explained in Section 5.4. Thus this choice of T_2 seems natural. In Figure 5.13, predictions based on the two potential models are displayed; the predictive equations for these models are the same as Equations 5.1 and 5.3, but with $\ln R_{T1,T2}$ substituted for ε as the predictor. The first model again treats pulse-like and ordinary motions the same, while the second distinguishes between these two types of ground motions. In Figure 5.13, the predictions based on these two models are similar visually, which suggests that the simpler model is appropriate. F tests are again performed here to determine which model is appropriate. The tests are performed for a suite of structures, using the following two choices for T_2: $T_2=2T_1$ (results shown in Table 5.3a) and $T_2=T_1/3$ (results shown in Table 5.3b). There are no clear characteristics of statistical significance apparent in Table 5.3 (8.5% of the tests indicate significance, which is relatively close to the expected 5% if the simpler model is appropriate). Thus we will proceed with the simpler model which treats pulse-like and ordinary ground motions together.

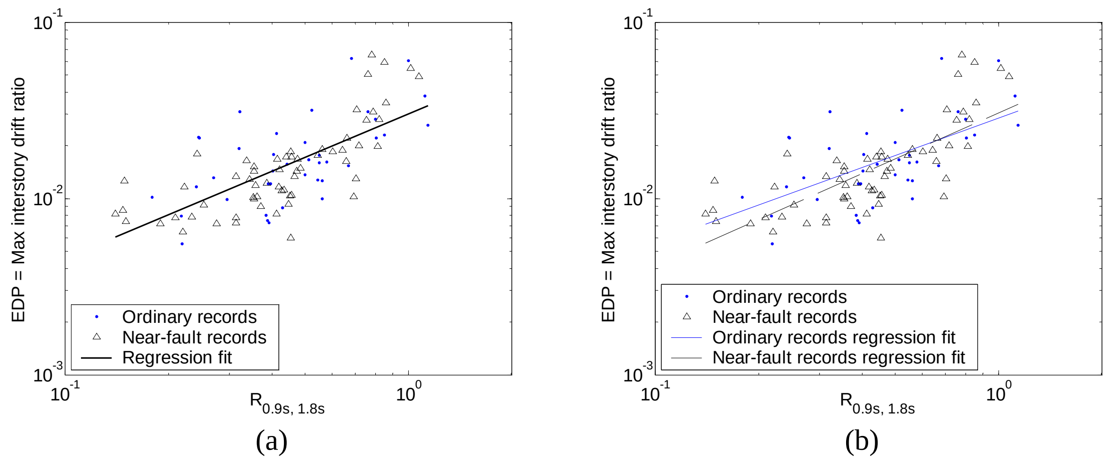


Figure 5.13. Prediction of response as a function of $R_{0.9s,1.8s}$ using linear regression on records scaled to an $Sa(T_1)$ level such that the structure's R_μ factor is 4. (a) Estimate of ordinary and pulse-like ground motion responses using the same prediction equation. (b) Estimate of ordinary and pulse-like ground motion responses using separate prediction equations.

Table 5.3. P-values from F tests to test the hypothesis that $R_{T1,T2}$ has a different effect on pulse records and non-pulse records. The tests are performed for two choices of T_2: (a) T_2=2T_1 and (b) T_2=T_1/3. If more than 50% of the records caused collapse, then a p-value is not reported.

R_μ Factor	Structure: N=3, T=0.3	N=9, T=0.9	N=6, T=1.2	N=9, T=1.8
0.5	0.14	0.19	0.08	0.95
1.0	0.14	0.29	0.07	0.83
1.5	0.47	0.69	0.06	0.70
2.0	**0.04**	0.65	0.27	0.91
2.5	**0.05**	0.42	0.21	0.65
3.0	0.13	0.22	0.66	0.21
3.5	**0.02**	0.30	0.33	0.66
4.0	0.32	0.52	0.30	0.71
4.5		0.31	0.50	0.98
5.0		0.36	0.90	0.53
5.5		0.27	0.88	
6.0		0.80	0.93	
6.5		0.28	0.54	
7.0		0.48		
7.5		0.57		

(a)

R_μ Factor	Structure: N=3, T=0.3	N=9, T=0.9	N=6, T=1.2	N=9, T=1.8
0.5	0.20	0.40	0.31	**0.01**
1.0	0.20	0.35	0.14	**0.04**
1.5	0.73	0.85	**0.01**	0.16
2.0	0.47	0.28	**0.02**	0.25
2.5	0.46	0.27	0.13	0.15
3.0	0.23	0.06	0.19	0.17
3.5	**0.03**	0.13	0.24	0.39
4.0	0.50	0.12	0.24	0.90
4.5		0.24	0.08	0.90
5.0		0.41	0.62	0.67
5.5		0.51	0.96	
6.0		0.44	0.90	
6.5		0.70	0.29	
7.0		0.98		
7.5		0.85		

(b)

$R_{T1,T2}$ and sufficiency with respect to Tp

The parameter $R_{T1,T2}$ is next considered to determine whether it can account for the effect of T_p. The procedure used to test ε in the previous section is repeated here. The residuals from the prediction based on $Sa(T_1)$ are used as the baseline. The 9 story structure with a first-mode period of 0.9 seconds is again used here, with an R_μ factor of 4. Thus, the residuals are the same as in Figure 5.9 earlier. These residuals are re-plotted in Figure 5.14a for convenience. The dependence on T_p is very clear. Next, prediction based on $R_{T1,T2}$ is considered. For this significantly nonlinear case, one reasonable choice for T_2 is $2T_1$, or 1.8 seconds. Prediction based on this IM was shown in Figure 5.13a. The residuals from this prediction are plotted in Figure 5.14b, and there is a dramatic reduction in the relationship between T_p and structural response. Further, the standard deviation of the residuals is also reduced significantly (e.g., the standard deviation of the residuals in Figure 5.14b is 34% less than the standard deviation in Figure 5.14a), and so gains have been made in prediction efficiency (the goal of Chapter 4) as well as in bias reduction with respect to pulse-like records.

An explanation for the effectiveness of $R_{T1,T2}$ in this case can be seen by examining the relationship between T_p and $R_{T1,T2}$. These two values are plotted in Figure 5.15. The $R_{T1,T2}$ has nearly the same relationship with T_p as structural response does (i.e., the shapes of the kernel-weighted average trends from Figure 5.14a and Figure 5.15 are very similar). When $T_p/T_1 \approx 1$, $R_{T1,T2}$ tends to be small, because $Sa(T_2)$ is lower than the peak caused by the pulse. But when $T_p/T_1 \approx 2$, $R_{T1,T2}$ tends to be large because $Sa(T_2)$ is on a peak caused by the pulse. This is encouraging for dealing with T_p sufficiency in a vector-valued IM.

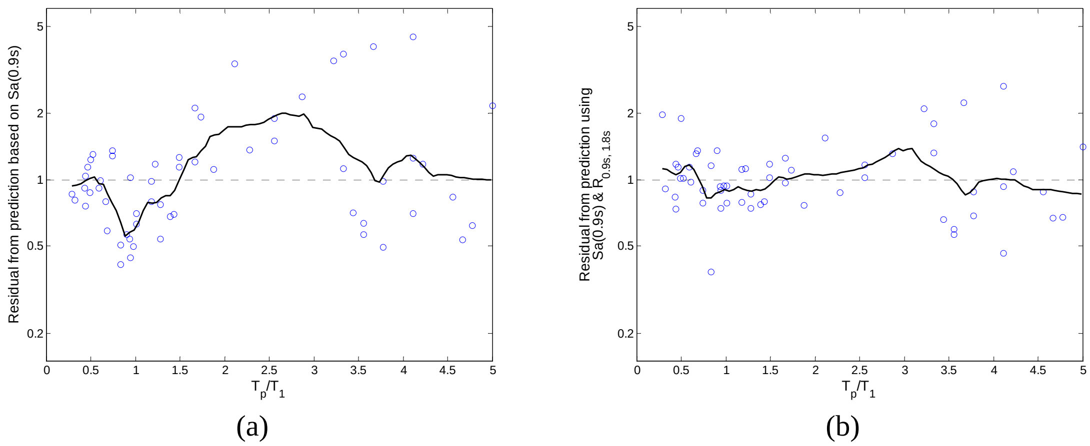


Figure 5.14. Residuals from response prediction based on (a) $Sa(T_1)$ only, and (b) both $Sa(T_1)$ and $R_{T1,T2}$, plotted versus T_p/T_1 for the generic frame with 9 stories and a first-mode period of 0.9 seconds, at an R_μ factor level of 4. T_2=1.8s for this plot.

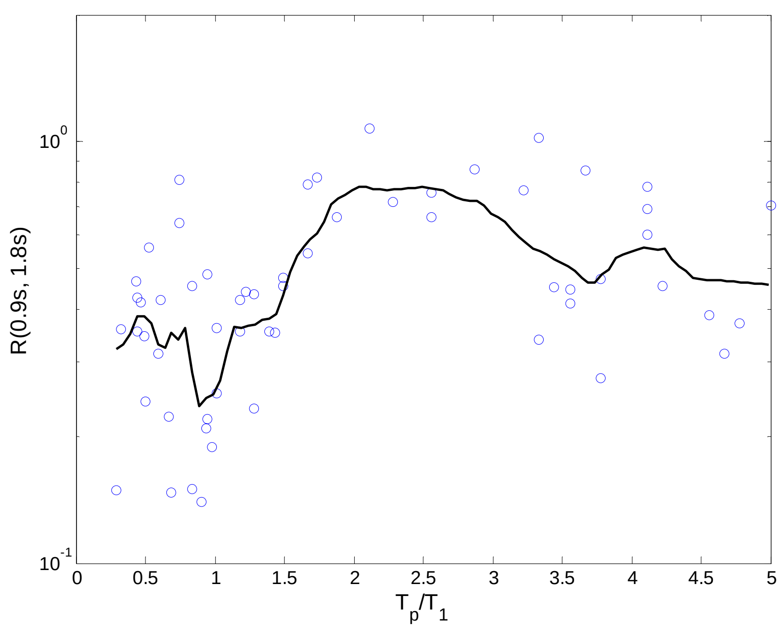
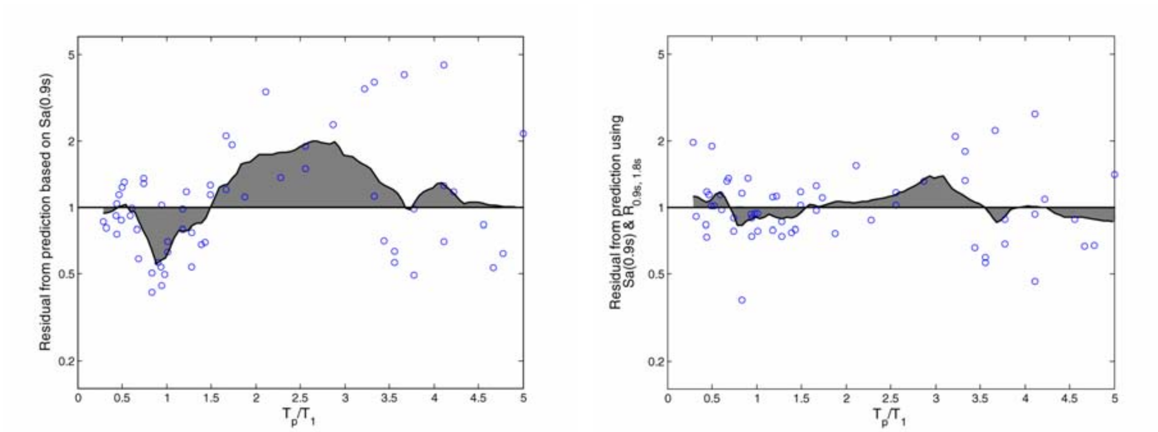


Figure 5.15. T_p versus $R_{T1,T2}$ for the pulse-like ground motions, where T_1=0.9s and T_2 =1.8s.

The choice of T_2=2T_1 was made based on intuition gained from Chapter 4, but other periods may also account for the effect of velocity pulses. To examine this, consider the following statistic, which is designed to measure the potential bias due to an IM's inability to account for T_p. The area between the kernel-weighted average line and the zero residual line is taken as a proxy for the total bias unaccounted for by the IM, as illustrated in Figure 5.16 (the total area is summed to create the statistic, so that positively and negatively biased regions do not cancel each other out). In this example, it is clear that the IM consisting of $Sa(T_1)$ and $R_{T1,T2}$ has less total bias than the IM consisting of $Sa(T_1)$ alone, and that this is reflected in the reduced shaded area. To measure the improvement gained by adding $R_{T1,T2}$, we compute the fractional reduction in area between these two figures – 0.65 in this case. This fractional reduction is a function of the T_2 value chosen. In Figure 5.17, this fractional reduction is plotted for a range of possible T_2 values. The optimal T_2 in this case is exactly $2T_1$, but other nearby periods also cause significant reduction in bias: T_2 values between $1.7T_1$ and $2.4T_1$ all provide at least 75% of the improvement attained using the optimal T_2. Further, this same test can be performed for the four structures considered, each with an R_μ factor of 4. As seen in Figure 5.18, in each case the value T_2=$2T_1$ is close to the optimal T_2.



(a) (b)

Figure 5.16. Area under the kernel-weighted average line, to be used as a proxy for the total bias as a function of T_p. Results are for the generic frame with 9 stories and a first-mode period of 0.9 seconds, T_2=1.8s, and R_μ=4. (a) Bias when using $Sa(T_1)$ as the IM. (b) Bias when using $Sa(T_1)$ and $R_{T1,T2}$, as the IM. The total area is reduced by 65% when the vector IM is adopted.

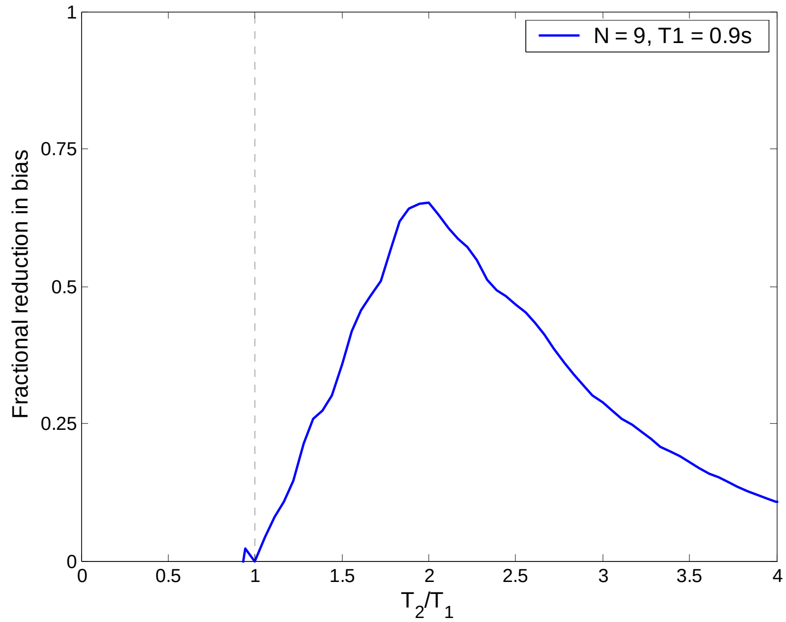


Figure 5.17. Percent reduction in the proposed bias statistic versus the T_2 value used in $R_{T1,T2}$. Results are shown for the N=9, T_1 = 0.9s structure with an R_μ factor of 4.

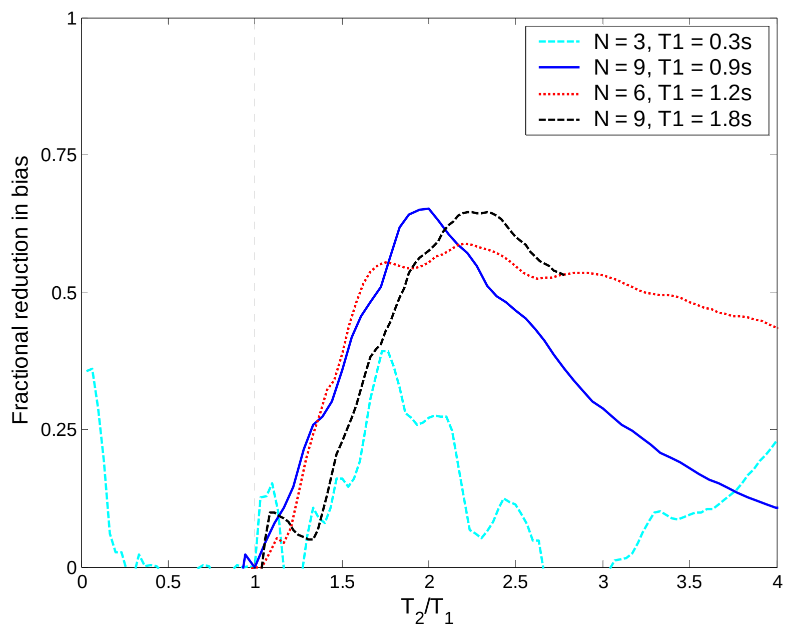


Figure 5.18. Percent reduction in the proposed bias statistic versus the T_2 value used in $R_{T1,T2}$. Results are shown for all four structures, each with an R_μ factor of 4.

To further test the effect of $R_{T1,T2}$, the above test was repeated when the ground motions were scaled to an $Sa(T_1)$ level such that the R_μ factor of the structure was 2. Remember that for this structure and at this low level of nonlinearity, higher mode effects were more important than nonlinear effects (this might be predicted based on the results of Chapter 4, and direct evidence from Figure 5.5 is also available in this case). For this reason, T_2=T_1/3 was selected as the second period, in order to capture second-mode response. The residuals from prediction based on $Sa(T_1)$ alone and based on $Sa(T_1)$ and $R_{T1,T2}$ are shown in Figure 5.19. Again, a substantial gain is made in reducing the effect of T_p on structural response. A plot of the fractional reduction in bias versus T_2 is shown in Figure 5.20, and it is seen that only T_2 values less than T_1 produce a reduction in bias. For this low-nonlinearity case, the period T_2=2T_1 cannot effectively account for the effect of T_p, because it is the pulses at short periods that affect the structure significantly, and so a lengthened T_2 value is not effective. However, for the moderate to high nonlinearity cases most likely to be of engineering interest when pulse-like ground motions are present, a T_2 value greater than the first-mode period of structure is likely to be effective. In general, however, the choice of whether nonlinear first-mode response or higher-mode response is more important will depend upon the structure of interest and the level of nonlinearity.

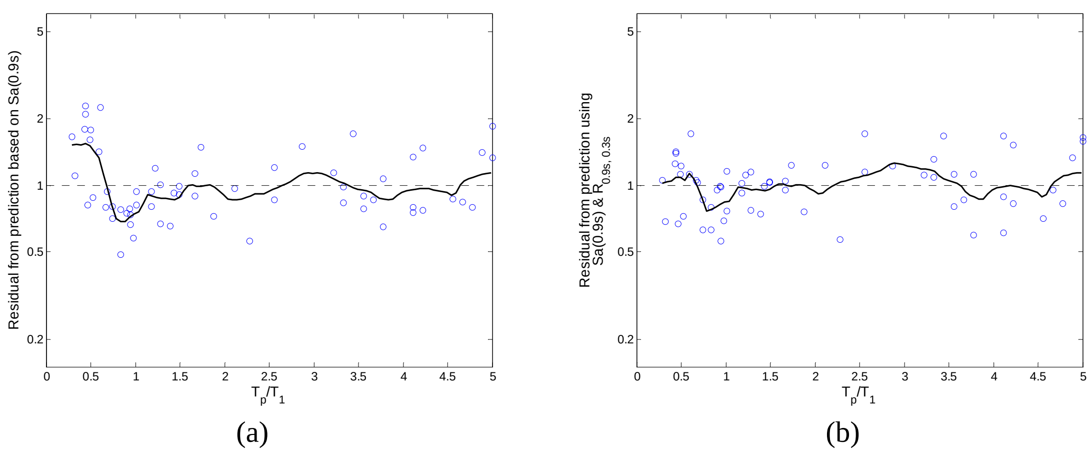


Figure 5.19. Residuals from response prediction based (a) $Sa(T_1)$ only, and (b) both $Sa(T_1)$ and $R_{T1,T2}$, plotted versus T_p/T_1 for the generic frame with 9 stories and a first-mode period of 0.9 seconds, at an R_μ factor level of 2. T_2 =0.3s for this plot.

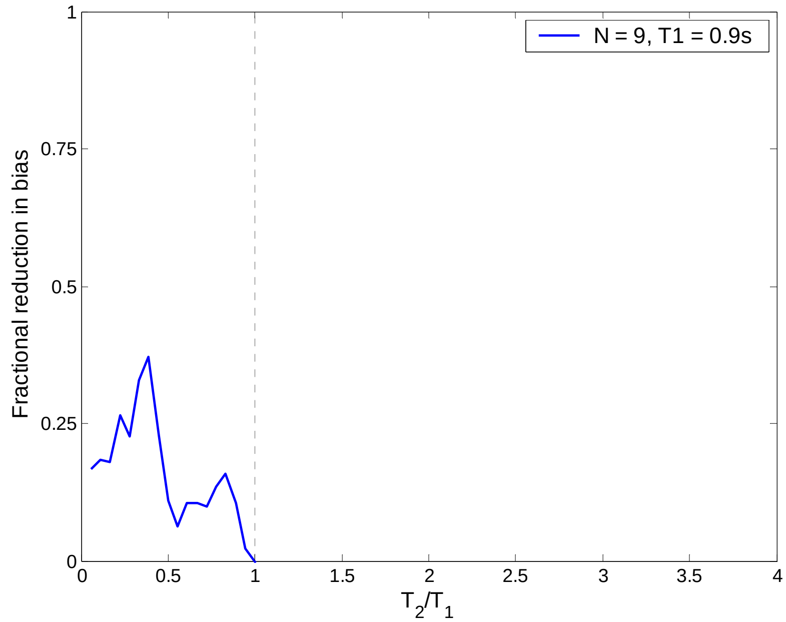


Figure 5.20. Percent reduction in the proposed bias statistic versus the T_2 value used in $R_{T1,T2}$. Results are shown for the N=9, T_1 = 0.9s structure with an R_μ factor of 2.

The same tests were performed for each of the structures at a range of $Sa(T_1)$ levels, and $R_{T1,T2}$ was seen to consistently reduce bias due to the effect of T_p, given reasonable selection of the period T_2. These results are encouraging for engineering assessment of pulse-like ground motions: they suggest that a structural analyst's primary concern be the ground motion parameters $Sa(T_1)$ and $R_{T1,T2}$, with reduced need to worry about the presence and period of a velocity pulse in the record.

To illustrate this, probabilistic structural response predictions are combined with ground motion hazard analysis to compute the "structural response hazard" for an example structure, using the scalar- and vector-IM based procedures explained in detail in Chapter 4. The procedure is repeated with the four sets of ordinary and near-fault ground motions used earlier (in Figure 5.6 and Figure 5.7), to determine whether the calculated result depends upon the ground motions used. The same nine-story structure as was used earlier is considered for illustration. Ground motion hazard is computed for two intensity measures: the scalar IM Sa(0.9s), and the vector IM consisting of Sa(0.9s) and $R_{0.9s,1.8s}$. As was seen in Figure 5.6 and Figure 5.7, the four sets of ground motions cause different levels of response in the structure for a given Sa(0.9s) level, so it seems reasonable that the structural response hazard curves computed using Sa(0.9s) will vary depending upon which of these ground motion sets are used. This is in fact observed in Figure 5.21a: the "best case pulse-like records" which were observed to cause smaller responses at a given *Sa*(0.9s) level, also result in lower estimates of mean rates of exceeding large maximum interstory drift ratios. Similarly, the "worst case pulse-like records" result in higher estimates of the mean rates of exceedance. However, when the vector IM consisting of *Sa*(0.9s) and $R_{0.9s,1.8s}$ is used, the variations in estimated structural response hazard are reduced, as seen in Figure 5.21b. Similar results are observed for the other example structures as well, suggesting that the use of the vector IM consisting of $Sa(T_1)$ and $R_{T1,T2}$ desensitizes interstory drift ratio estimates to the presence of velocity pulses in the near-fault ground motions.

The ground motion hazard is for the Van Nuys site used throughout this dissertation. The site is not expected to experience pulse-like ground motions, but it still provides a useful test as to whether the vector IM proposed here can desensitize the resulting estimates to the presence of velocity pulses. To accurately perform this assessment at a site where pulse-like ground motions can potentially occur, revisions are needed to standard ground motion hazard analysis, as will be described below.

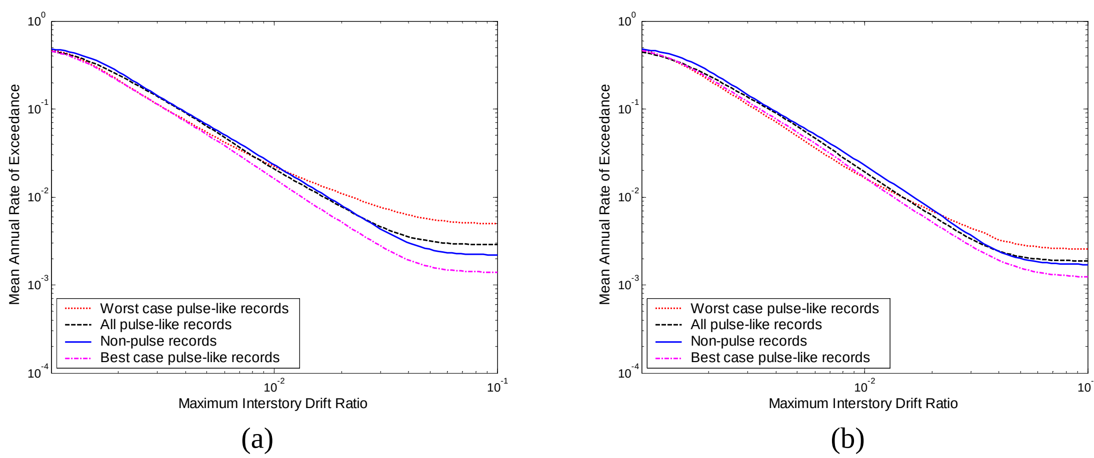


Figure 5.21. Drift hazard curves using the four considered record sets. (a) Curves computed using the scalar IM Sa(0.9s). (b) Curves computed using the vector IM consisting of *Sa*(0.9s)

The vector IM consisting of $Sa(T_1)$ and $R_{T1,T2}$ is useful for the reasons discussed above, but some significant limitations remain. This vector IM has so far only been shown to robustly predict *maximum* interstory drift ratios. Alavi and Krawinkler (2001) noted that pulse-like ground motions can cause different distributions of interstory drift ratios over the height of the structure, and this IM may not account for that effect. Further work is needed to test the robustness of vector Ims in predicting a range of other structural response parameters, and predicting response in other structural systems. Further, in order to use response predictions based on $Sa(T_1)$ and $R_{T1,T2}$ in the presence of pulse-like ground motions, the challenge still remains to compute a vector-valued ground motion hazard that incorporates the effect of pulse-like ground motions on the probability of given $Sa(T_1)$ and $R_{T1,T2}$ values occurring at the site of interest. This challenge will be discussed in Section 5.6 below.

5.5.3 Additional IMs

Before concluding the consideration of Ims, two more potential Ims should be mentioned. Luco and Cornell (2005) have proposed an improved scalar intensity measure that was shown to be effective for predicting the effect of pulse-like ground motions. This intensity measure is a combination of inelastic spectral displacement at the first-mode period and elastic spectral displacement at the second-mode period. It was seen to be an effective at accounting for the effect of pulse-like records, but as with the vector IM proposed here, accounting for pulse-like records in the ground motion hazard for this IM will be a challenge. Ongoing research should address this problem soon (Tothong and Cornell 2005b).

Another IM that might be considered is a vector consisting of $Sa(T_1)$ along with the pulse period of the ground motion. This would certainly account for T_p dependence in a direct way. However, there are several difficulties which keep this from being an appealing option. First, relationship between T_p and structural response could not be quantified through a simple linear equation (e.g., Figure 5.14 or Figure 5.19). This will cause some difficulties when creating a predictive model, although one could be determined if necessary. Also, only pulse-like ground motions have an associated pulse period. This means that pulse and non-pulse records will need to be treated separately, which is an inconvenience. On the other hand, the relationship between $R_{T1,T2}$ and maximum interstory drift ratio is essentially the same for both ordinary and pulse-like motions if T_2 is chosen effectively (e.g., Figure 5.13a), and so the two classes of ground motions may not need to be distinguished when $R_{T1,T2}$ is used. In addition, in Chapter 4 $R_{T1,T2}$ was shown to provide increased efficiency for predicting response from ordinary ground motions, and so the effectiveness of $R_{T1,T2}$ as an intensity measure is much broader than the near-fault-specific T_p value. For these reasons, $R_{T1,T2}$ appears to be a more effective IM parameter than T_p for use with $Sa(T_1)$ to predict response from pulse-like ground motions.

5.6 Vector-valued PSHA for pulse-like ground motions

In the preceding section it was demonstrated that a vector-valued IM consisting of $Sa(T_1)$ and $R_{T1,T2}$ can efficiently predict the effect of pulse-like ground motions on structures, and that this vector is sufficient with respect to T_p. This is very appealing for analysis of structures, because it means that pulse-like and ordinary ground motions do not need to be identified separately, at least for estimation of maximum interstory drift ratio. Given the values of $Sa(T_1)$ and $R_{T1,T2}$, the presence of a velocity pulse in a ground motion does not further affect the maximum interstory drift ratio caused in a structure. However, to perform a full probabilistic assessment of a structure, one hurdle remains: estimation of the joint probabilities of $Sa(T_1)$ and $R_{T1,T2}$ values at the site, accounting for the fact that pulse-like ground motions might occur. Loosely speaking, we now know how to predict structural response from both ordinary and pulse-like ground motions using $Sa(T_1)$ and $R_{T1,T2}$, but for a given site we need to know how often ground motions with given values of those two parameters will occur. Bazzurro and Cornell (2002) have described this procedure, termed Vector-valued Probabilistic Seismic Hazard Analysis (VPSHA) for sites subjected to ordinary ground motions, but it requires some modification for use at sites with pulse-like ground motions.

Ground motion prediction models for pulse-like ground motions are in active development at present and so improved models should be available in the near future. The use of these improved models for performing a probabilistic hazard analysis is described by Tothong et al. (2005). In the present period of transition to these new models, it may be helpful to think about a reasonable approach that could be used immediately, as well as an improved approach that would be preferable given results from anticipated future research. These possibilities are considered below.

5.6.1 Minor modification to standard VPSHA – modify the means and covariances of the ground motion prediction model

Bazzurro and Cornell (2002) outline VPSHA using joint normal random variables for $\ln R_{T1,T2}$ and $\ln Sa(T_1)$, where the conditional dependence between the two is fully defined by a correlation coefficient. The simplest modification to VPSHA would be to use this same formulation with the means, variances and correlation coefficients modified to account for the fact that pulse-like motions might be present in some situations. This could be implemented immediately, using existing attenuation models for pulse-like ground motions. Today, the only existing prediction model for pulse-like ground motions is the one proposed by Somerville et al. (1997), and modified by Abrahamson (2000a) and Bozorgnia and Bertero (2004, Chapter 5) for application in PSHA. Further, the correlation coefficient between $R_{T1,T2}$ and $Sa(T_1)$ may need to be modified. Empirical correlation coefficients obtained from the pulse-like ground motions used here tend to be lower than the correlation coefficients observed in ordinary ground motions (see Appendix C).

For example, consider Figure 5.22. In this figure, the joint distribution of $Sa(T_1)$ and $R_{T1,T2}$ values is shown, given occurrence of a Magnitude 7 event at a distance of 5km, with the near-fault parameters X=0.5 and θ=5% (parameters from Somerville et al. 1997). The prediction is based on the ground motion prediction model of Abrahamson and Silva (1997) with near-fault modification as described in Bozorgnia and Bertero (2004, Chapter 5) and an empirical correlation coefficient from the data set used here (Appendix C). Considering near-fault effects increases the probability of observing large $Sa(T_1)$ and/or $R_{T1,T2}$ values.

To perform VPSHA, one would simply repeat this calculation for a range of magnitude, distance, X and θ values, where the distribution of magnitudes and distances specified in the same way as for standard PSHA (McGuire 2004) and the distributions of X and θ values come from assuming random earthquake hypocenter locations within the fault rupture. This procedure has been implemented for scalar intensity measures (Abrahamson 2000a), so the only modification would be to compute joint distributions for $Sa(T_1)$ and $R_{T1,T2}$ (as in Figure 5.22a) rather than marginal distributions for $Sa(T_1)$. This approach to VPSHA requires the least modification from current ground motion hazard analysis practice, and can be implemented in the near future (Somerville and Thio 2005).

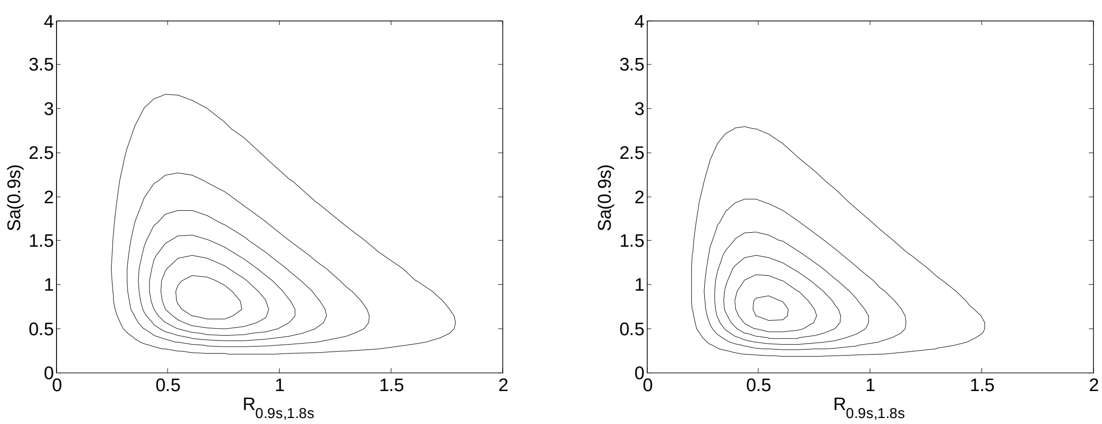


Figure 5.22. Joint conditional probability density function of $Sa(T_1)$ and $R_{T1,T2}$, given a Magnitude 7 event at a distance of 5km, with the near-fault parameters X=0.5 and θ=5%. T_1=0.9s and T_2=1.8s, in the fault-normal direction. (a) With near-fault effects considered. (b) Without near-fault effects considered.

5.6.2 Major modification to standard VPSHA – explicitly incorporate the probability of a velocity pulse and the distribution of pulse periods

The procedure of the preceding section is desirable from a practical viewpoint, but it has theoretical shortcomings. Earlier in this chapter, it was seen that pulse-like ground motions may or may not cause severe responses, based on the period of the pulse. Further, the occurrence of a velocity pulse is not certain even at source-to-site geometries where a pulse might be comparatively likely. So given a site where a pulse might occur, if the occurrence of a pulse is a random event and the period of the pulse is a random variable, then the resulting $Sa(T_1)$ value may not have a lognormal distribution, as was assumed in the previous section. In fact, the distribution may be bimodal, with one peak associated with records that have $T_p \approx T_1$, and another peak associated with the other records. Further, if the pulse only increases the elastic response spectrum in a narrow range, then the distribution of $Sa(T_1)$ and $R_{T1,T2}$ might not be jointly lognormal, as was assumed in the previous section. In this section, a VPSHA procedure is described that incorporates the occurrence of pulses and their resulting T_p values explicitly. The procedure consists of the following steps:

1. For each magnitude, distance and hypocenter location considered, does the geometry suggest pulse-like effects?
   a. If no, compute the joint distribution of $Sa(T_1)$ and $R_{T1,T2}$ given the magnitude, distance, site conditions, etc.
   b. If yes, compute the conditional joint distribution of $Sa(T_1)$ and $R_{T1,T2}$ for both the pulse and non-pulse possibilities
      i. Compute the joint distribution of $Sa(T_1)$ and $R_{T1,T2}$ given the magnitude, distance, site conditions, etc., given no velocity pulse (this might be the same as, or similar to, the non-pulse-like case)
      ii. Given that a pulse does occur, integrate over the distribution of possible T_p values. For each T_p value, compute the joint distribution of $Sa(T_1)$ and $R_{T1,T2}$ given the magnitude, distance, site conditions, etc., and given a velocity pulse with period T_p.
      iii. Combine the pulse and non-pulse predictions by weighting the conditional distributions, given a pulse and given no pulse, by their associated probabilities of occurrence.
2. Integrate these over the random hypocenter locations for each earthquake event (i.e., magnitude and distance) considered
3. Integrate over the joint rates of occurrence of magnitude and distance values (as with standard PSHA)

This hazard analysis model should provide the most accurate distribution of $Sa(T_1)$ and $R_{T1,T2}$ values for hazard analysis. Further, by using disaggregation (McGuire 1995, Bazzurro and Cornell 1999), it would be straightforward to compute the probability that a pulse-like record caused the occurrence of a given $Sa(T_1)$ and $R_{T1,T2}$ level, as well as the distribution of T_p values associated with the causal pulse-like record. This information would be helpful in identifying a representative “scenario event” that has a high probability of causing the target IM level to occur. It should not be necessary to select specific ground motions based on this information, because $Sa(T_1)$ and $R_{T1,T2}$ are already sufficient with respect to pulse-like record properties. However, if one wanted to, one could select ground motions with the scenario event in mind.

This procedure is not presently ready for implementation, however, because currently not all of the required steps can be performed. The probability of occurrence of a velocity pulse is needed in step 1.b of the procedure, but no prediction of this probability has been published (Iervolino and Cornell 2005b have a model in preparation). Several models have been proposed to predict the distribution of pulse periods given existence of a pulse and given the earthquake magnitude needed in step 1.b.ii (Somerville et al. 1997, Alavi and Krawinkler 2001, Mavroeidis and Papageorgiou 2003, Somerville 2003, Fu and Menun 2004). No models currently predict the distribution of $Sa(T_1)$ given magnitude, distance and T_p, although again work is known to be underway within the PEER NGA project. Even more challenging than prediction of a single $Sa(T_1)$ value (given magnitude, distance and T_p) is prediction of the joint distribution of $Sa(T_1)$ and $R_{T1,T2}$ (i.e., the distribution needed in step 1.b.ii of this procedure). Although this model is not yet ready for practice, discussion of the proposed procedure may help to guide current research efforts, as well as motivate new research on topics required for execution of this procedure. The ideas presented in this section were developed with significant assistance from Polsak Tothong and Iunio Iervolino, and are described more rigorously in an upcoming paper (Tothong et al. 2005).

VPSHA for pulse-like ground motions is a challenging problem with more research required, but the vector-valued-IM approach appears to be a promising solution for characterizing the effect of pulse-like ground motions. This is because once the VPSHA challenges are overcome, the advantages of the IM-based procedure can be applied to near-fault environments. Because the vector IM is sufficient with respect to pulse-like ground motions, it is less important to identify representative pulse-like records on a building-by-building basis, as is sometimes done today (Stewart et al. 2001, Somerville 2001a). Record selection problems will become much less critical, as is the case today for ordinary ground motions. By de-coupling the ground motion occurrence problem from the structural response problem, the probabilistic assessment procedures described elsewhere in this dissertation can be used for pulse-like ground motions as well.

5.7 Conclusions

Vector-valued intensity measures have been considered for predicting the effects of pulse-like ground motions, where here the term pulse-like ground motion is used to refer to near-field ground motions in the fault-normal direction which exhibit a velocity pulse. These ground motions are of particular concern to structural engineers because of their potential to cause large levels of structural response. Further, their effects are not well-predicted by traditional measures of ground motion intensity such as spectral acceleration.

To address these problems, two vector-valued intensity measures were considered for prediction of these ground motions. The first was a vector consisting of spectral acceleration at the first-mode period of the building (denoted $Sa(T_1)$) along with a parameter termed ε. While ε was found in Chapter 3 to be an effective predictor of structural response because it indicates whether the spectrum is in a peak or valley at the specified period, here it has been found to be ineffective at identifying the effect of velocity pulses in near-fault ground motions. The second IM considered was one consisting of $Sa(T_1)$ plus a parameter $R_{T1,T2} = Sa(T_2)/ Sa(T_1)$ that describes the shape of the response spectrum, and where the period T_2 is specified by the user. In Chapter 4 this IM was found to efficiently predict maximum interstory drift ratios from ordinary ground motions, especially given wise choice of T_2. Here this vector has been found to be effective at predicting maximum interstory drift ratios resulting from pulse-like ground motions as well. This IM is sufficient with respect to near-fault effects if T_2 is chosen reasonably well. That is, given $Sa(T_1)$ and $R_{T1,T2}$, there is generally no statistically significant difference in maximum interstory drift ratios between ordinary and pulse-like records. Further, given $Sa(T_1)$ and $R_{T1,T2}$, the resulting maximum interstory drift ratio is significantly desensitized to the period of a record's velocity pulse. These desirable properties do not hold when $Sa(T_1)$ alone is used as an intensity measure. This finding suggests that the vector consisting of $Sa(T_1)$ and $R_{T1,T2}$ is an effective measure of ground motion intensity for predicting structural performance (in terms of max. interstory drift ratio) when pulse-like ground motions may occur at a site. For the example structures considered here, a second period (T_2) equal to twice the first-mode period (T_1) of the structure was seen to effectively predict moderate-to-severe nonlinear structural response, although this should be confirmed on a wider range of structural types.

In addition to predicting response given $Sa(T_1)$ and $R_{T1,T2}$, the ground motion hazard must be computed for this intensity measure in order to assess the performance of a structure at a given site. Two possible methods were proposed for capturing near-fault effects in ground motion hazard analysis. The first is a slightly modified version of the method of Bazzurro and Cornell (2002), where random hypocenter locations must be considered and where the means, variances and correlations of the response spectral values are computed based on a ground motion prediction model that accounts for near-fault effects. Results obtained using this method will be available soon. In the second method of computing ground motion hazard, the probability of pulses occurring and their associated pulse periods are considered explicitly as conditional random variables, in the same way as magnitude and distance are in standard seismic hazard analysis. The ground motion prediction model could then be dependent upon the presence of a pulse and its associated period. This method should be more accurate, but requires further development of models for the probability of a pulse occurring, as well as so-called narrow-band ground motion prediction models which account for pulse period explicitly.

By adopting the vector-valued IM consisting of $Sa(T_1)$ and $R_{T1,T2}$, the problem of rigorously estimating probabilistic structural response in the presence of near-fault effects is less critical. This IM is sufficient with respect to the presence of velocity pulses and their associated pulse-periods, and selection of appropriate ground motions for analysis becomes much less important. Challenges remain with respect to computing the probabilistic ground motion hazard in a way that accounts for near-fault effects, but research in progress promises to resolve those concerns in the near future.

Chapter 6

Spectral shape, epsilon and record selection

6.1 Abstract

Selection of earthquake ground motions is considered with the goal of accurately estimating the response of a structure at a specified ground motion intensity, as measured by spectral acceleration at the first-mode period of the structure ($Sa(T_1)$). Consideration is given to magnitude, distance and epsilon (ε) values of the ground motions. First, it is seen that selecting records based on their ε values is more effective than selecting records based on magnitude and distance. Second, a method is discussed for finding the conditional response spectrum of a ground motion, given a level of $Sa(T_1)$ and its associated mean (disaggregation-based) causal magnitude, distance and ε value. Records can then be selected to match the mean of this target spectrum, and the same benefits are achieved as when records are selected based on ε. This mean target spectrum differs from a Uniform Hazard Spectrum, and it is argued that this spectrum is a more appropriate target for record selection than a Uniform Hazard Spectrum. When properly selecting records based on either spectral shape or ε, the improved accuracy in estimated structural response is comparable to that gained by using a vector-valued measure of earthquake intensity. Here accuracy includes both reduced bias and variance. These improved record-selection methods also allow records to be scaled without introducing bias, unlike the less effective methods of record selection.

6.2 Introduction

Selection of recorded earthquake ground motions as inputs for dynamic analysis is an important consideration when assessment of structures is based on nonlinear dynamic analysis. As described in Chapter 2, careful selection of earthquake records can achieve same the reduction in bias and variance of structural response gained by use of improved vector-valued intensity measures, while allowing the user to process the records using a simple scalar IM. In Chapter 3, the ground motion parameter epsilon (ε) was seen to be an important predictor of structural response as part of a vector-valued IM, and it was suggested that ε should be considered when selecting records. The effect of selecting records based on ε was not studied, however.

Here we will consider record-selection criteria more thoroughly, and identify the record properties that should be considered when selecting records for analysis. These record-selection criteria are considered in the context of probabilistic assessment of structures, where both the ground motion hazard and the response given the ground motion are quantified in a formal probabilistic manner. It will be seen that consideration of ε values when selecting records is important, as ε is an effective predictor of structural response (and a more effective predictor than the magnitude or distance of the ground motion).

Alternatively, a method is proposed for developing a target spectrum that accounts for the magnitude (M), distance I and ε values likely to cause a given target ground motion intensity at a given site. The spectrum obtained in this way is termed a *Conditional Mean Spectrum, considering* ε (CMS-ε). This target spectrum possibly widens the range of acceptable records for analysis, because the selected records do not have to match the causal magnitude, distance and ε values, but rather the records need only have a spectral shape which looks like the mean spectrum from the causal event. The proposed target spectrum is also compared to a Uniform Hazard Spectrum, and seen to be superior for obtaining unbiased estimates of structural response. The CMS-ε spectrum is similar to target response spectra used in the nuclear safety industry (DOE 1996, NRC 1997, ASCE 2005), except that the effect of ε has been incorporated as well, given the new findings about the effect of ε on structural response (Chapter 3).

The effect of record scaling is also considered, and it is seen that if records are selected based on ε or the CMS-ε, then the records can be scaled without inducing a bias in structural response. However, for records selected without respect to these values, scaling the records by large scale factors may induce a bias.

For the above reasons, it is suggested that earthquake motions for dynamic structural analysis be selected based on ε or the CMS-ε. The magnitude and distance values associated with the records are relatively less important, although they are accounted for in the CMS-ε. Using these selection criteria, more efficient estimates of structural response can be obtained, and biases due to scaling of records can be avoided. These results are relevant for estimating the mean response at a given ground motion level (as used in, e.g., ICC 2003) as well as probabilistic assessments of structural response (as used in, e.g., FEMA 2000a, Deierlein 2004).

6.3 The effect of epsilon on spectral shape

In Chapter 3, ε was identified as an indicator of spectral shape (spectral shape was formally measured using $Sa(T_2)/Sa(T_1)$ for some T_1 and T_2; more generally it can be thought of as the relative values of spectral accelerations at other periods, given Sa at T_1). The parameter ε is a measure of the difference between the spectral acceleration of a record and the mean of a ground motion prediction equation at the given period. That finding is confirmed here by observing patterns seen in a large set of recorded ground motions taken from the PEER Strong Ground Motion Database (2000). This relationship between ε and spectral shape justifies the conclusions regarding record selection that will be made below. All of the records from this database that met the following criteria were selected:

1. The site was classified as stiff soil: USGS class B-C or Geomatrix class C-D.
2. The recording was made in the free field or the first story of a structure.
3. The earthquake magnitude was greater than 5.5.
4. The source-to-site distance was less than 100 km.
5. The vertical and both horizontal components of the recording were available and each had high-pass filter corner frequencies less than 0.2 hertz and low-pass filter corner frequencies greater than 18 hertz.

These criteria were used to identify records believed to be most relevant for engineering purposes. A total of 191 recordings met the above criteria, resulting in 382 horizontal ground motion components available for use below (record details are given in Appendix A, Table A.5). The records were examined to search for a relationship between ε and spectral shape. First, ε values were computed for each record at a range of periods using the ground motion prediction model of Abrahamson and Silva (1997). The structure considered below has a first- mode period of 0.8 seconds, so here the ε values at 0.8s are investigated as an example. The twenty records with the largest ε values at a period of 0.8s were identified (these records have ε values greater than 2.25), as well as the twenty closest to zero (these records have ε values between -0.06 and 0.06) and the twenty with the smallest ε values (these records have ε values less than -2.25). The spectra of the records with large ε values are scaled to have the same Sa(0.8s) value and plotted in Figure 6.1a, along with the geometric mean of the twenty spectra (the records are scaled to Sa(0.8s)=0.5g, but the relative shapes of the spectra are not affected by the Sa(0.8s) value and so the actual value is not important here). In Figure 6.1b, the spectra with small ε values are shown, along with their mean. In Figure 6.2, the geometric means of these two sets, along with the geometric mean of the set with ε values close to zero, are displayed. It is clear that the average shapes of these record sets differ, even though each set has a wide range of magnitudes and distances, and the distribution of magnitude and distance values of the records do not differ appreciably among the sets. The same exercise was repeated at a period of 0.3s (the selected records depend upon the period, because records with a large ε value at 0.8s may not have an equally large ε value at 0.3s), and the resulting median spectral shapes are shown in Figure 6.3, and the same effect is seen. Thus, we see empirically that ε is accounting for differences in spectral shape, providing further evidence for the effect proposed in Chapter 3. Note that although both positive and negative epsilons are considered here, the most important distinction is that between positive-ε records and zero-ε records because the former are associated with long-return period ground motions while the latter are associated with typical record sets chosen without regard to ε values.

It is widely known that for records with the same $Sa(T_1)$ value, spectral shape will affect the response of multi-degree-of-freedom and nonlinear structures (because spectral values at other periods affect response of higher modes of the structure as well as nonlinear response once the structure's effective period has lengthened). It is also widely recognized that magnitude and distance affect the spectral shape of records. In Chapter 3, however, it was seen that ε also affects spectral shape (and that its effect is greater than the effect of magnitude or distance, as was seen in Chapter 3). In the next section, we will use knowledge of the effect of M, R and ε on structural response as a guide when selecting appropriate records for nonlinear analysis.

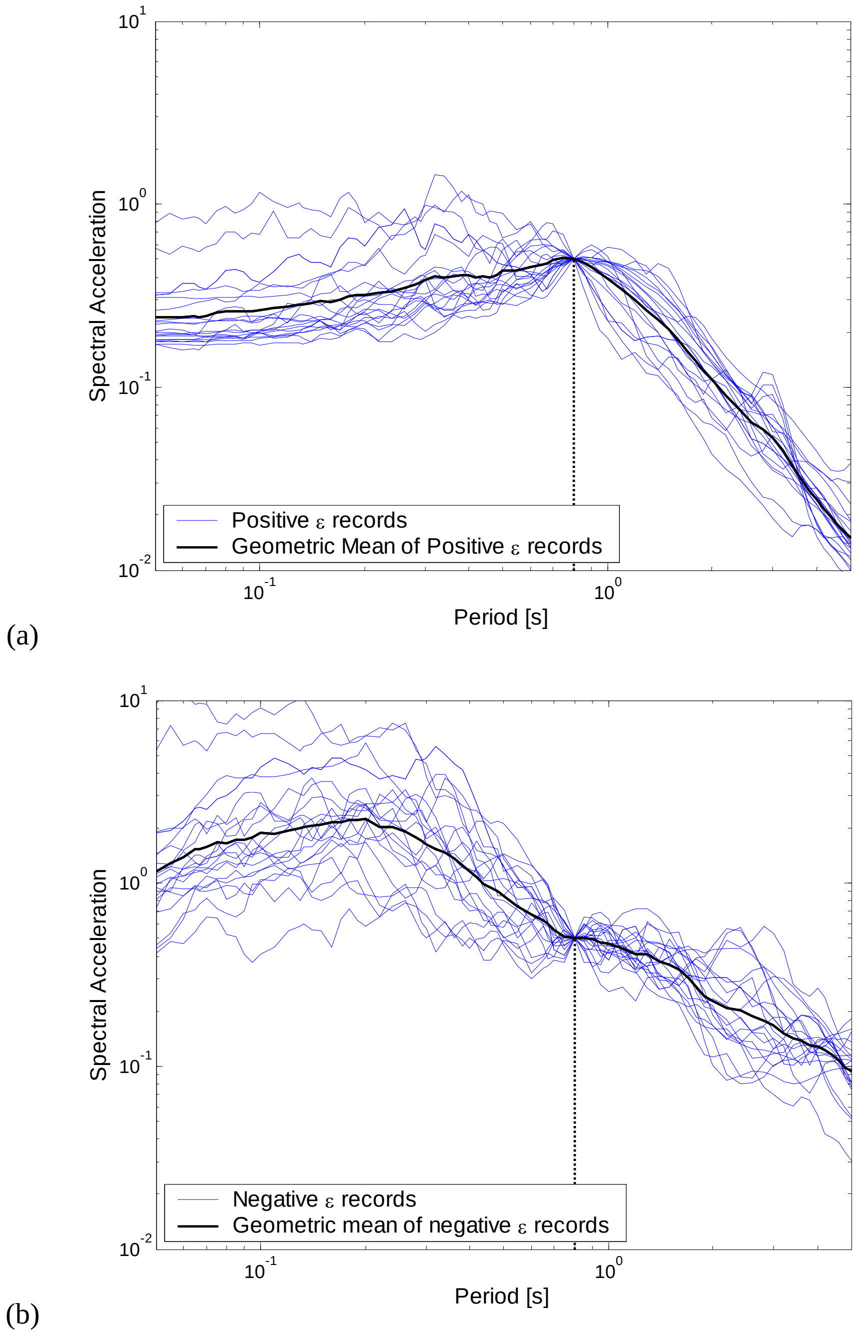


Figure 6.1: (a) Response spectra of records with the 20 largest ε values at 0.8s, and the geometric mean of the set, after scaling all records to Sa(0.8s)=0.5g. (b) Response spectra of records with the 20 smallest ε values at 0.8s, and the geometric mean of the set, after scaling all records to Sa(0.8s)=0.5g.

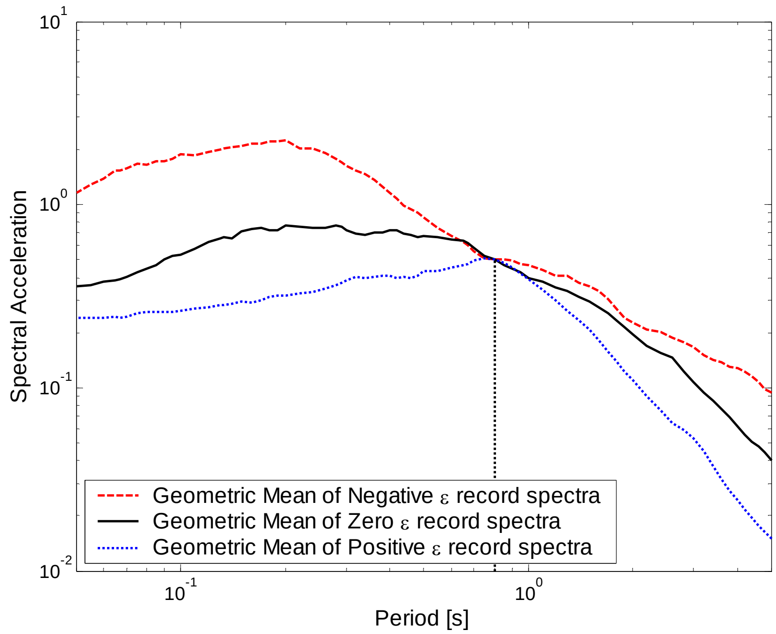


Figure 6.2: The geometric mean of response spectra for negative-ε, zero-ε and positive-ε record sets, after each record's spectrum has been scaled to Sa(0.8s)=0.5g.

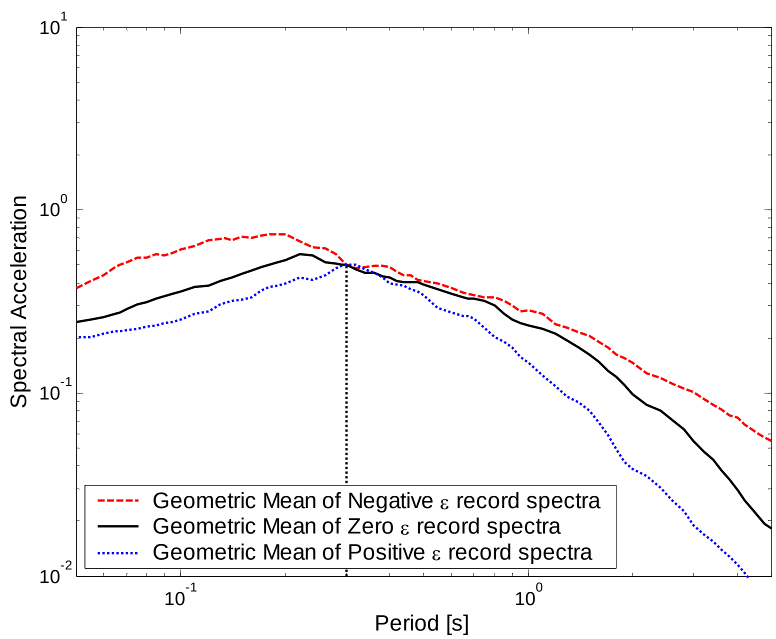


Figure 6.3: The geometric mean of response spectra for negative-ε, zero-ε and positive-ε record sets, after each record's spectrum has been scaled to Sa(0.3s)=0.5g.

6.4 A predictive model for spectral shape

Given our growing knowledge of the role of ε in spectral shape, we propose here a new method of developing a target spectrum. To develop this target spectrum, we first find the $Sa(T_1)$ value corresponding to the target probability of exceedance at the site using PSHA, denoted $Sa(T_1)^*$. We then use disaggregation to find the mean of the M, R and ε values (denoted $\bar{M}$, $\bar{R}$ and $\bar{\varepsilon}$) that cause the occurrence of $Sa(T_1)^*$ level (e.g., McGuire 1995)[15]. $\bar{M}$ and $\bar{R}$, in turn, via ground motion prediction models, determine the means and standard deviations of the response spectral values for all periods, and $\bar{\varepsilon}$ specifies the number of standard deviations away from the mean the ground motion is *at the first-mode period,* T_1. Given knowledge of the mean ε at T_1, denoted $\bar{\varepsilon}(T_1)$, we can calculate the conditional distribution of Sa values at other periods using only the disaggregation data and knowledge of correlations of ε values at a range of periods, as will be shown below.

This scheme for developing a target spectrum follows closely from procedures to develop target spectra for analysis of nuclear facilities (DOE 1996, NRC 1997, ASCE 2005), except that those methods incorporate only the causal M and R values from disaggregation. The target spectra must then be scaled up to match the specified Sa value. Here, the effect of ε is incorporated as well, given the finding in Chapter 3 that ε is an important predictor of structural response. A similar idea was used by Toro and Silva (2001) to develop target response spectra that accounted somewhat for the effect of ε.

The target response spectrum based on $\bar{M}$, $\bar{R}$ and $\bar{\varepsilon}$ can be computed in the following manner: as was outlined in Chapter 3, given certain assumptions the conditional mean and standard deviation of the response spectrum can be computed using the following equations[16].

$$\mu_{\ln Sa(T_2)|\ln Sa(T_1)=\ln Sa(T_1)^*} = \mu_{\ln Sa}(\bar{M},\bar{R},T_2) + \sigma_{\ln Sa}(\bar{M},T_2)\rho_{\ln Sa(T_1),\ln Sa(T_2)} \cdot \bar{\varepsilon}(T_1) \qquad (6.1)$$

[15] For record-selection purposes, one should use the McGuire (1995) definition of disaggregation, which provides the distribution of M, R and ε given $Sa(T_1)$ *equals* $Sa(T_1)^*$, rather than the Bazzurro and Cornell (1999) definition, which provides the distribution of M, R and ε given $Sa(T_1)$ *exceeds* $Sa(T_1)^*$. Bazzurro and Cornell discussed the differences between these two definitions and expressed a preference for the latter, but they were not considering this target spectrum question.

[16] These equations are in fact an approximation obtained by substituting $\bar{M}$, $\bar{R}$ and $\bar{\varepsilon}$ for the random values of M, R and ε obtained from disaggregation given the target $Sa(T_1)$ value. The approximation slightly modifies the mean and standard deviation, but the difference is believed to be small in most cases. Further, when substituting $\bar{M}$, $\bar{R}$ and $\bar{\varepsilon}$ into Equation 6.1, one does not necessarily obtain the target $Sa(T_1)$ value back again. This can be addressed by re-assigning $\bar{\varepsilon}$ to the ε value that results in a prediction of the $Sa(T_1)$ target value; the modification will be small and this is consistent with the treatment of ε by McGuire (1995). More details regarding this approximation are given in Appendix E.

$$\sigma_{\ln Sa(T_2)|\ln Sa(T_1)=x} = \sigma_{\ln Sa}(\bar{M}, T_2)\sqrt{1-\rho^2_{\ln Sa(T_1),\ln Sa(T_2)}} \tag{6.2}$$

Where $\bar{M}$, $\bar{R}$ and $\bar{\varepsilon}(T_1)$ come from disaggregation given $Sa(T_1) = Sa(T_1)^*$. The terms $\mu_{\ln Sa}(\bar{M}, \bar{R}, T_2)$ and $\sigma_{\ln Sa}(\bar{M}, T_2)$ are the marginal mean and standard deviation of ln*Sa* at T_2, as predicted by a ground motion prediction (attenuation) relationship (e.g., Abrahamson and Silva 1997), and a model for $\rho_{\ln Sa(T_1),\ln Sa(T_2)}$ is given in Equation 8.9. Equations 6.1 and 6.2 describe the distribution of response spectra at the specific site of interest, given that $Sa(T_1)$ is to the target value $Sa(T_1)^*$. Actually, to completely specify the conditional distribution, one also needs the conditional correlations; it is also possible to calculate those, but they are not needed for what follows.

In Figure 6.4, a distribution for the response spectrum is given, conditioned on occurrence of an *Sa*(0.8s) level of 1.6g (the 2% in 50 years ground motion at the example site), and $\bar{M}$ = 6.4, $\bar{R}$ = 11.5 km and $\bar{\varepsilon}$ = 2.1 from the PSHA disaggregation. Note that the distribution of *Sa* at 0.8s has zero standard deviation (because we have conditioned knowing this value), while the *Sa* values at other periods have some uncertainty because they are only partially correlated with *Sa*(0.8s). Note also that the spectrum has a slight "peak" at 0.8s. This phenomenon was identified in Chapter 3 and observed empirically in Figure 6.1a.

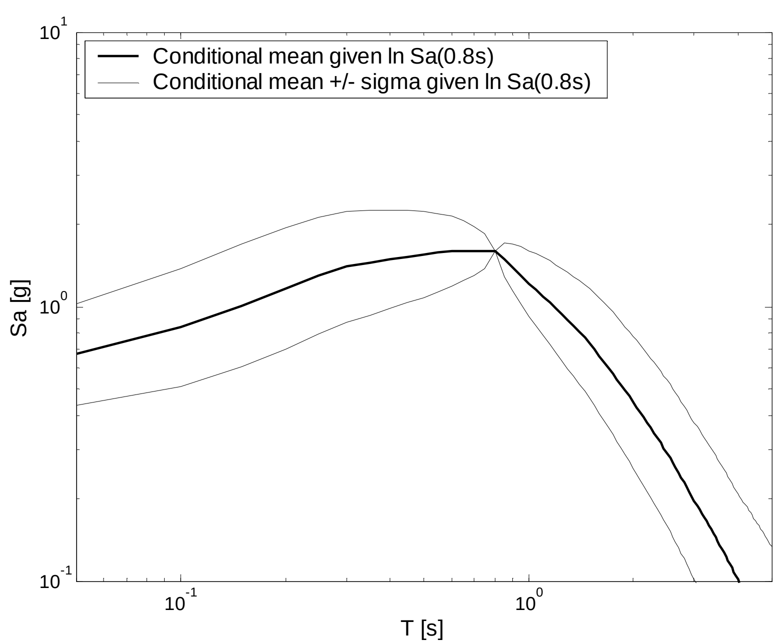


Figure 6.4: The distribution of spectral acceleration values, given *Sa*(0.8s) = 1.6g, and given $\bar{M}$, $\bar{R}$, $\bar{\varepsilon}$ from PSHA disaggregation.

It is reasonable to assume that *M*, *R* and ε are important to structural response only indirectly as proxies for spectral shape, because, beyond $Sa(T_1)$, spectral shape is the dominant factor affecting structural response (this assumption will be further justified below). This is because spectral shape dictates higher mode response and describes the relative ground motion strength at longer periods of concern in nonlinear behavior. Thus, rather than trying to match target *M*, *R* and ε values when selecting records, one might use the target *M*, *R* and ε values to determine a target spectral shape, and select records based on this target spectral shape alone. This should increase the number of available records, because some records with incorrect *M*, *R* or ε values may have the "correct" spectral shape. The natural choice for this spectrum is the conditional mean spectrum computed in Equation 6.1.

We will take only the mean value of this conditional spectrum as the target, rather than attempting to match the entire conditional distribution. This is done under the justification that $Sa(T_1)$ is the primary predictor of structural response, and spectral values of other periods are of secondary importance. Thus we account for $Sa(T_1)$ fully probabilistically using probabilistic hazard analysis, and we account for the other spectral values by taking their conditional means. This approach follows probabilistically based load combination rules used in practice elsewhere (e.g., Norwegian Technology Standards Institution 1999, p17-18). This condition spectrum is termed the *Conditional Mean Spectrum, considering* ε (or *CMS-*ε). The "considering ε" term is included to emphasize the modification from nuclear industry conditional mean spectrum, which are computed without considering the effect of ε. A record selection procedure based on the *CMS-*ε spectrum will be considered below.

6.5 Potential record-selection strategies

Given the effect that ε has on structural response, ε values should be considered carefully when choosing ground motions for estimation of structural response. Ideally, the target distribution of magnitude, distance *and* ε values in the record set should equal the condition distribution of magnitude, distance and ε values seen at the site of interest, given $Sa(T_1)$, as determined from standard PSHA disaggregation. Note that this conditional distribution will change as a function of $Sa(T_1)$, and so different records will need to be selected for different $Sa(T_1)$ levels.

To test the effect of the *M*, *R* and ε values in a selected record set, several record-selection methods are now considered, and the resulting structural response outputs compared. Ideally, we would match the target distribution of all of these parameters simultaneously, but this can be difficult in practice due to the finite number of available recorded ground motions. Because of this limitation, priority should be given to matching the most important parameters, and so the most important parameters need to be identified. Alternatively, one could select records that have a spectral shape representative of a spectrum given the target M, R and ε values, as discussed above. Four record-selection methods are used, in order to investigate the effect of these different record-selection strategies on resulting estimated structural responses:

1. Select records at random from a record library, without attempting to match any specific record properties (this will be abbreviated as the **AR Method**, as it uses Arbitrary Records).
2. Select records with magnitude (M) and distance I values representative of the site hazard, without attempting to match the ε values (this will be abbreviated as the **MR-BR Method**, as it uses M, R-Based Records).
3. Select records with ε values representative of the site hazard, without attempting to match the magnitude and distance values (this will be abbreviated as the **ε-BR Method**, as it uses ε-Based Records).
4. Select records with spectral shapes that match the conditional mean spectral shape given $\bar{M}$, $\bar{R}$ and $\bar{\varepsilon}$ as discussed earlier, but make no attempt to directly match the M, R or ε values (this will be abbreviated as the **CMS-ε Method**, as it uses the Conditional Mean Spectrum, considering ε).

With all four methods, we make the additional restriction that the record site be classified as soil, to match the conditions present at the site of interest. Method 1 can be considered as a base case, where information about the causal events is ignored when selecting records (i.e., 40 records at random were selected from the library of 382). Method 2 reflects state-of-the-art practice today (e.g., Stewart et al. 2001, Bommer and Acevedo 2004). For Method 2, records were selected to have a minimum difference in magnitude and distance relative to the $\bar{M}$, $\bar{R}$ obtained from disaggregation (where a difference of one unit in magnitude was treated as equal to a difference of 40km in distance). Method 3 is used to test the effect of ε. With this method, the 40 records with ε values closest to $\bar{\varepsilon}$ were selected. One other study has been performed to date where records were selected based on M, R and ε simultaneously (Haselton et al. 2005), but here we choose to match only subsets of the parameters (i.e., only ε in method 3, and only M and R in method 2), in order to clearly distinguish the effects of these parameters. Later, a vector-valued intensity measure will be used to account for magnitude, distance and ε simultaneously.

Finally, Method 4 is used to investigate the possibility that M, R and ε are proxies for spectral shape, and that spectral shape is the factor directly influencing structural response. If this is true, then records with a spectral shape matching the CMS-ε spectrum for a given M, R and ε will be accurate predictors of structural response, regardless of their actual M, R and ε values. With the CMS-ε method, records were selected that had a minimum sum of squared differences between their spectrum and the CMS-ε spectrum at seven periods in the range 0.16 to 2.4 seconds (0.16, 0.3, 0.5, 1.2, 1.5, 1.9, and 2.4 seconds), after scaling the records to match the target *Sa*(0.8s). The periods were selected to include shorter periods corresponding to higher modes of oscillation of the structure, and longer periods corresponding to structure softening as it becomes nonlinear. The suggestion of ASCE 7-02 (2002, section 9.5.7.2.1) to use $0.2T_1$ and $1.5T_1$ as the period range of interest was adopted, although periods as large as $3T_1$ were also included because the structure will be driven to high levels of nonlinearity where effective-period lengthening may cause the structure to be sensitive to motion at much longer periods. The spectra of selected records are displayed in Figure 6.5, along with the conditional distribution of the target spectrum. For most periods the standard deviation of the records' spectra is smaller than the conditional spectrum distribution, but the mean spectrum of the records is approximately equal to the target mean[17].

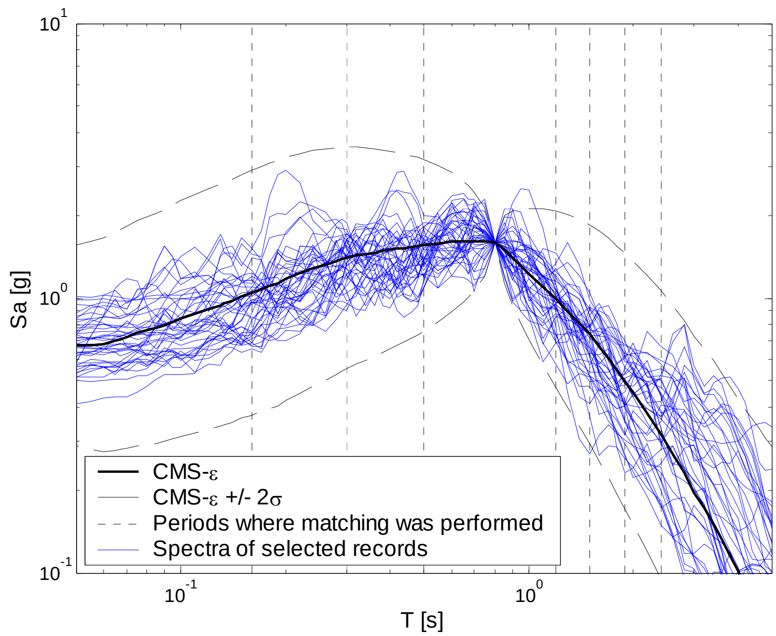


Figure 6.5: The Conditional Mean Spectrum considering ε, the CMS-ε +/- two σ, and response spectra of records selected based on their match with this spectral shape. All spectra are conditioned upon *Sa*(0.8s)=1.6g.

The records selected using Methods 2, 3 and 4 will depend (to different degrees) on the site of interest because the causal *M*, *R* and ε values depend on the surrounding faults and their rates

---

[17] Underestimating the standard deviation of the records spectra suggests that the standard deviation of structural response will be underestimated as well, as will be seen below. This is similar to the effect seen when spectrum-compatibilizing records, only it is seen to a lesser extent here because the records are not fully smooth. This effect may be helpful for code-based procedures when the mean-response is desired, because the lower standard deviation of structural response implies that the mean can be estimated with greater confidence. The reduced standard deviation of response may cause problems for probabilistic

of activity. The records selected will also depend upon the hazard level of interest, because small frequent levels of ground motion typically have different associated magnitudes and distances than large rare levels of ground motions, but most importantly, ε changes radically as the ground motion level increases. This can be seen in the changing conditional mean spectral shapes (and associated M, R and ε values) at three hazard levels in Figure 6.6 for a site in Van Nuys, California. Thus, if one is analyzing the structure at multiple ground motion intensities, as will be done here, then the records should be re-selected at each level to reflect the changing sources. For this exercise, we consider a large range of ground motion intensities (as measured using *Sa*(0.8s)—the first-mode period of the example structure considered below). Note that if we were able to simultaneously match target M, R and ε values in our records then they would also have the approximately the correct *Sa*(0.8s) value. (To match the target *Sa*(0.8s) value exactly, a minor modification is needed, as explained in Appendix E), but because we are not able to match all properties we will have to employ some degree of record scaling to match the target *Sa*(0.8s) values.

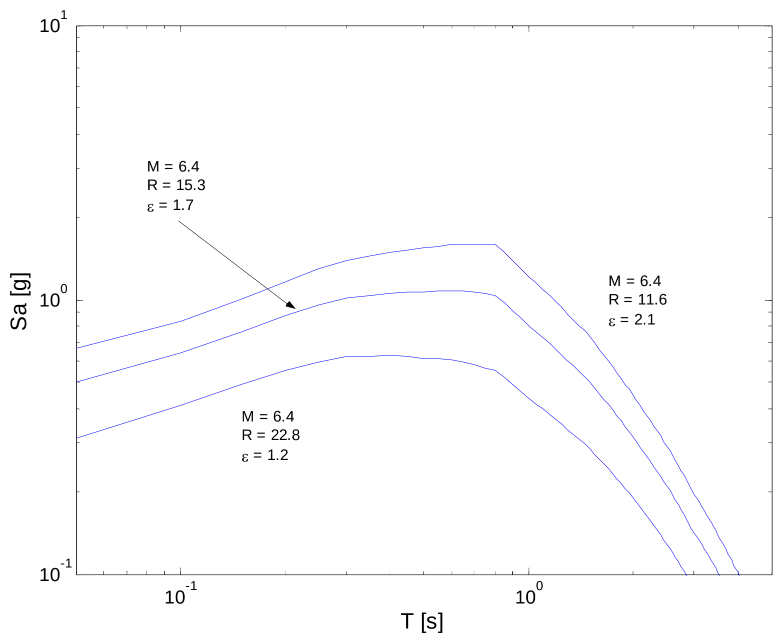


Figure 6.6: Mean values of the conditional response spectrum for a site in Van Nuys, California, given occurrence of *Sa*(0.8s) values exceeded with 2%, 10% and 50% probabilities in 50 years.

Using Method 1, we use the same set of records at all *Sa*(0.8s) levels. Using methods 2, 3 and 4, we re-select records at several levels of ground motion intensity according the disaggregation

assessment procedures, although the effect is not obvious in the example results below.

results at each level. The selected records used are listed in Appendix A. For all four methods, 40 records were selected at 12 $Sa(T_1)$ levels between 0.1 and 4.0g.

In Table 6.1 through Table 6.3, the magnitude, distance and ε values of the selected records are given, along with the target values obtained from disaggregation. A few observations can be made. When the record property is being matched explicitly the selected records tend to match the target value, and otherwise the records tend to match the mean value of the record library. This is most obvious in Table 6.2, where the distance-specified records closely match the target distances, but the other record sets have mean distance values of approximately 33km: the mean of the record library. An interesting phenomenon also arises with the records selected based on spectral shape. The magnitudes and distances of the selected records do not change appreciably as the *Sa* level varies, but the ε values do change. This implies that in order to match the spectral shape associated with a given *M*, *R* and ε, it is possible to substitute records with differing magnitudes and distances, but the spectral shapes associated with large ground motion levels tend to have large ε values. Note, however, that the mean ε value of the spectral-shape selected records is lower than the mean ε value from disaggregation, meaning that it is not entirely necessary to rely on only the largest (and rarest) ε records for selection of these more peaked conditional mean spectra.

Table 6.1. Mean magnitude values from disaggregation of the Van Nuys site, and the mean magnitude values of the records selected using each of the four proposed methods. The mean magnitude value of the record library was 6.7.

	Magnitude				
Sa(0.8s) [g]	Target from disagg	1. AR Method	2. MR-BR Method	3. ε-BR Method	4. CMS-ε Method
0.1	6.3	6.7	6.5	6.8	6.6
0.2	6.4	6.7	6.4	6.7	6.7
0.4	6.4	6.7	6.4	6.6	6.9
0.6	6.4	6.7	6.4	6.8	6.9
0.8	6.4	6.7	6.4	6.7	6.9
1.0	6.4	6.7	6.5	6.7	6.9
1.4	6.4	6.7	6.5	6.6	6.8
1.6	6.4	6.7	6.5	6.6	6.9
1.8	6.4	6.7	6.5	6.6	6.8
2.4	6.5	6.7	6.5	6.6	6.8
3.0	6.6	6.7	6.6	6.6	6.7
4.0	6.8	6.7	6.7	6.6	6.7

Table 6.2. Mean distance values from disaggregation of the Van Nuys site, and the mean distance values of the records selected using each of the four proposed methods. The mean distance value of the record library was 33 km.

	Distance				
Sa(0.8s) [g]	Target from disagg	1. AR Method	2. MR-BR Method	3. ε-BR Method	4. CMS-ε Method
0.1	57.6	35.5	52.4	36.8	33.1
0.2	40.6	35.5	36.4	28.5	31.9
0.4	27.5	35.5	29.5	30.9	34.7
0.6	21.7	35.5	23.3	32.9	35.2
0.8	18.1	35.5	20.8	35.2	33.5
1.0	15.6	35.5	16.9	39.8	33.9
1.4	12.5	35.5	10.8	48.0	33.5
1.6	11.5	35.5	10.3	48.9	36.6
1.8	10.8	35.5	7.4	49.7	35.8
2.4	9.7	35.5	7.7	49.7	37.0
3.0	9.0	35.5	9.2	49.7	37.2
4.0	9.1	35.5	11.4	49.7	37.1

Table 6.3. Mean ε values (at 0.8s) from disaggregation of the Van Nuys site, and the mean εvalues of the records selected using each of the four proposed methods. The mean ε value of the record library was 0.2.

	Epsilon (ε)				
Sa(0.8s) [g]	Target from disagg	1. AR Method	2. MR-BR Method	3. ε-BR Method	4. CMS-ε Method
0.1	0.0	0.0	0.6	0.0	0.1
0.2	0.5	0.0	0.2	0.5	0.2
0.4	1.0	0.0	0.2	0.9	0.4
0.6	1.3	0.0	0.0	1.3	0.5
0.8	1.5	0.0	0.0	1.4	0.5
1.0	1.7	0.0	-0.2	1.6	0.7
1.4	2.0	0.0	-0.4	1.9	0.9
1.6	2.1	0.0	-0.5	1.9	1.0
1.8	2.2	0.0	-0.4	2.0	1.0
2.4	2.4	0.0	-0.5	2.0	1.2
3.0	2.5	0.0	-0.1	2.0	1.2
4.0	2.8	0.0	0.1	2.0	1.3

The selected records can also be compared by examining their mean response spectra at a specified *Sa* level relative to the target mean from Equation 6.1. This is shown in Figure 6.7. At *Sa*(0.8s)=1.6g, the ε-BR and CMS-ε records have mean spectra close to the target, while the MR-BR records have larger *Sa* values at some periods, and the AR records have particularly high *Sa* values at nearly all periods other than 0.8s.

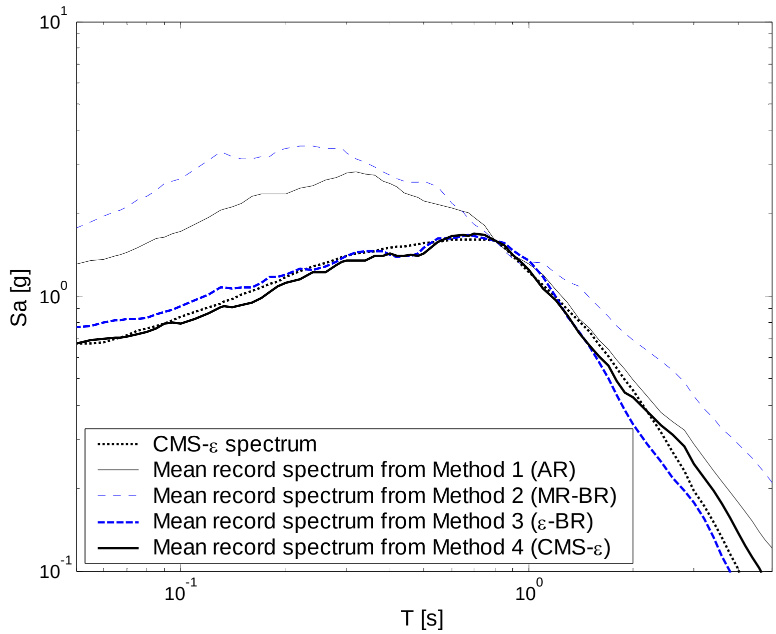


Figure 6.7: Conditional Mean Spectrum at *Sa*(0.8s)=1.6g (given $\bar{M}$ =6.4, $\bar{R}$ =11.5 km and $\bar{\varepsilon}$ =2.1) and the mean response spectra of record sets selected using each of the four proposed record selection methods.

6.6 Structural analysis

The records selected in the previous section were then used as inputs to the Van Nuys testbed model used earlier in this dissertation, in order to investigate any differences in resulting structural responses. For clarity of presentation, only one structure is considered in this chapter; in Appendix F, two additional structures are analyzed using the same procedure and similar results are observed. The maximum observed interstory drift ratio was used as the structural response parameter of interest. This structure has a first-mode period of 0.8 seconds, which is the reason why *Sa*(0.8s) has been used as the intensity measure in this study. A plot of the calculated maximum interstory drift ratios versus *Sa*(0.8s) using the records of Method 1 is shown in Figure 6.8.

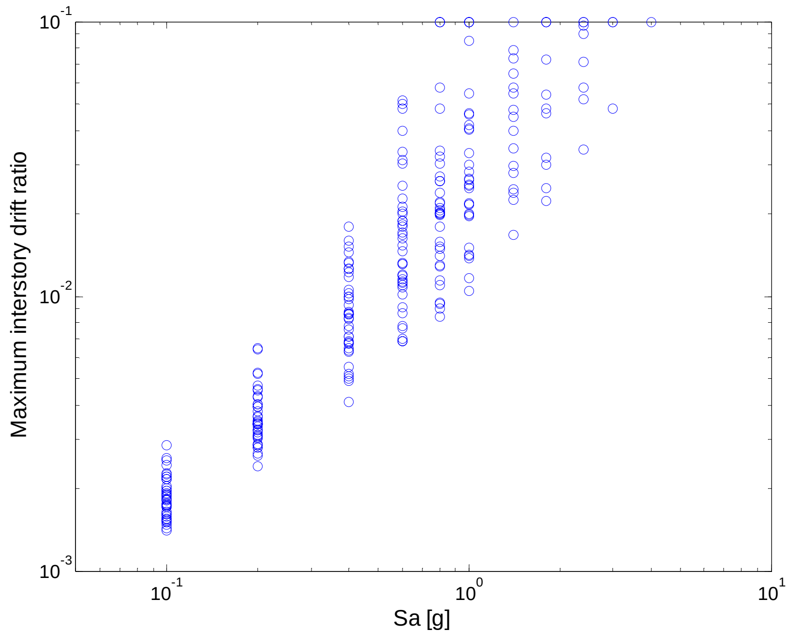


Figure 6.8: Maximum interstory drift ratio vs. *Sa*(0.8s). The results are shown for records from the Arbitrary Records method (i.e., Method 1). Similar results were obtained for the other three methods. Collapses are plotted as 10^{-1} maximum interstory drift ratio, and the fraction of records causing collapse are indicated in Figure 6.11.

The geometric mean of *EDP* as a function of *IM* is shown in Figure 6.9, using the results from the four record-selection methods. We see that the matched-ε records produce the lowest mean responses, with the spectral-shape-matched records producing approximately the same response. The other two methods produce slightly larger mean responses. This comparison of medians is of particular interest to current practice (e.g., ICC 2003), where the objective is to estimate the mean response at a specified ground motion intensity level. Note that the x axis of this plot is limited to *Sa* values less than 1*g*, because at larger *Sa* levels, a significant portion of records cause collapse and thus comparisons of non-collapse responses are less meaningful.

The standard deviations of log maximum interstory drift ratio from the three record sets are shown in Figure 6.10, where we see that the ε-matched records produce slightly smaller dispersions than the non-specified and *M*,*R*-matched records, and the CMS-ε records produce significantly lower dispersions, in part because the spectral-shape-matching procedure used to select these records causes them to be "smoother" than the model specifies them to be. This was observed in Figure 6.5, where the selected records had smaller standard deviations over a range of periods than the predicted conditional standard deviations.

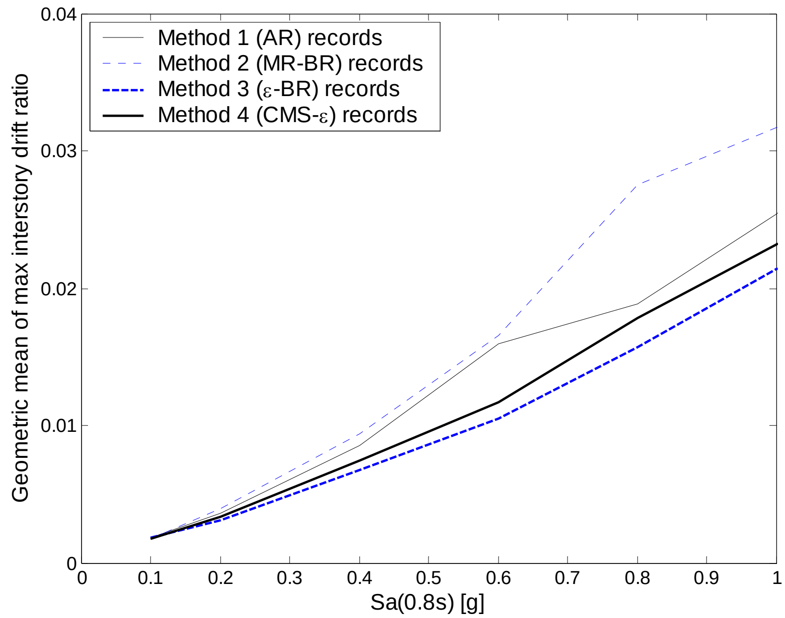


Figure 6.9: The geometric mean of maximum interstory drift ratio for records that do not cause collapse, plotted versus $Sa(T_1)$ for the four record-selection methods considered. The x axis of the figure is truncated at 1g, because at higher levels of spectral acceleration, a significant fraction of records cause collapse.

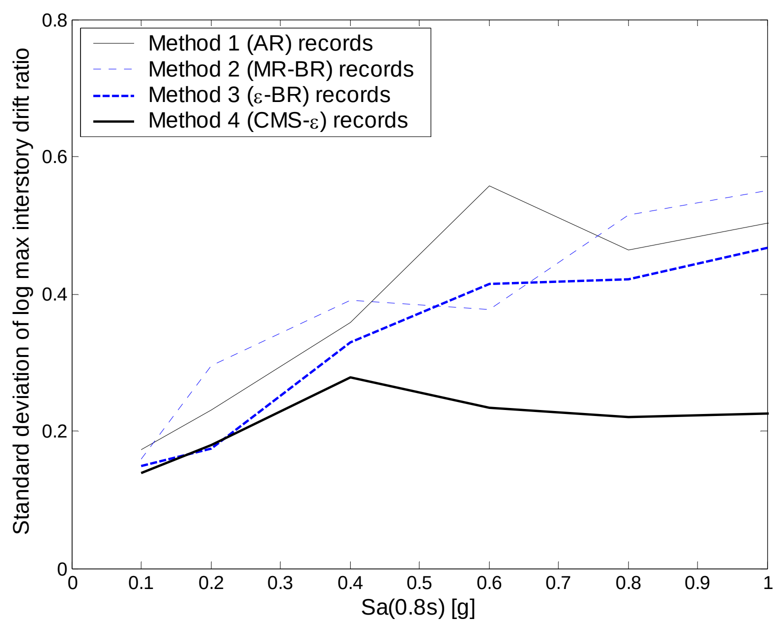


Figure 6.10: The standard deviation of log maximum interstory drift ratio for records that do not cause collapse, plotted versus $Sa(T_1)$ for the four record-selection methods considered.

In addition, we can compute the probability of collapse versus *Sa* (i.e., the collapse "fragility curve") for the four record selection methods, by counting the fraction of records that cause collapse at each *Sa* level and fitting a lognormal distribution to the results (see Appendix D for details). The result is seen in Figure 6.11. The non-specified and *M*,*R*-specified records have much larger probabilities of collapse, particularly in the left tail of the distribution, which tends to be the most critical (because smaller *Sa* values occur much more frequently than the larger *Sa* values at the right end of the distribution). The estimated probabilities of collapse versus *Sa* from the spectral-shape-selected records and the ε-selected records are nearly identical.

Judging from the non-collapse geometric means and log standard deviations, and the probability of collapse estimates, it appears that the method used to select records can have an effect on the resulting estimates of structural response. Further, the spectral-shape-matched records (which should be accounting for the effects of *M*, *R* and ε) produce responses comparable to the ε-matched records.

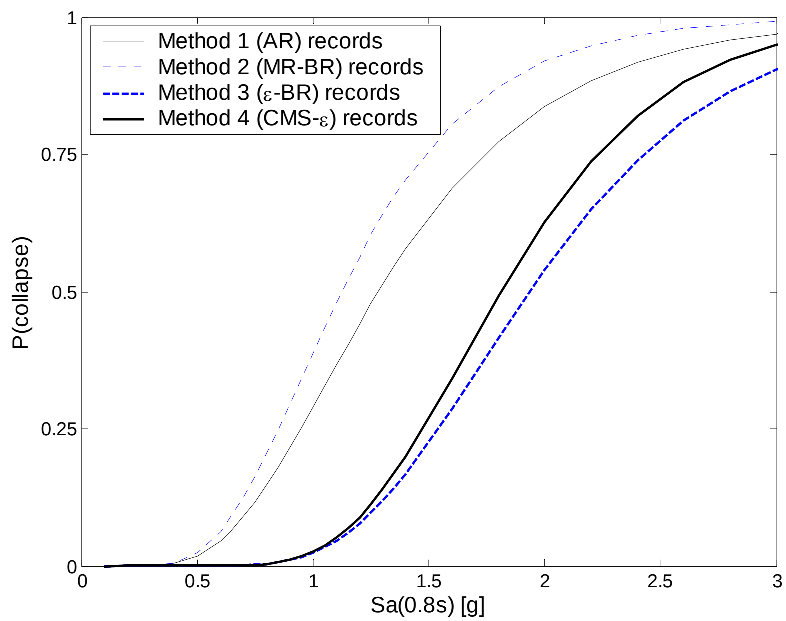


Figure 6.11: Estimated probability of collapse vs. $Sa(T_1)$ (i.e., the collapse “fragility curve”) using the four record-selection methods considered.

Incorporating ground motion hazard to compute drift hazard

An additional way to evaluate the results from each of the record-selection methods is to combine the estimated distributions of response as a function of $Sa(T_1)$ with the probability of exceedance of each $Sa(T_1)$ level (shown in Figure 6.12), to compute the mean annual frequency of exceeding a given structural response level (sometimes referred to as a “drift hazard curve”). The mathematical details of this simple integration are explained in detail in, e.g., Chapter 3 and Chapter 4.

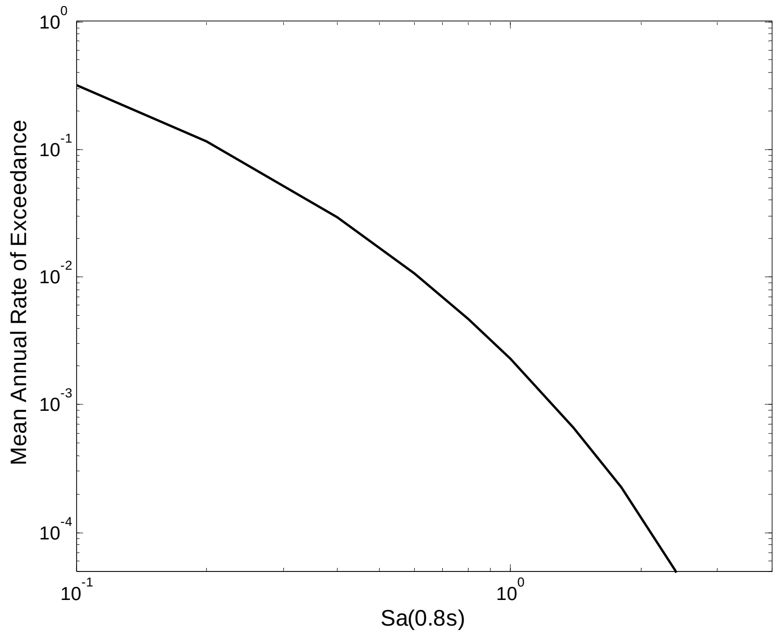


Figure 6.12: Mean annual frequency of exceeding various levels of *Sa*(0.8s) (i.e., the ground motion hazard curve), for the Van Nuys, California, site of interest.

The resulting drift hazard curves are shown in Figure 6.13. We see that the AR Method and MR-BR Method records produce higher estimated probabilities of exceedance than the other two methods, particularly at large response levels. These results are in good agreement with the results of Chapter 3, but here we are investigating the results of various record properties through careful record selection rather than with a vector-valued intensity measure as was done in Chapter 3. We see that results from records selected using the AR Method and MR-BR Method do not differ significantly, while the records selected using the ε-BR Method produce significantly lower mean annual frequencies of exceeding large levels of structural response. Further, we see here that the CMS-ε Method records produce nearly identical results to the ε-BR Method records.

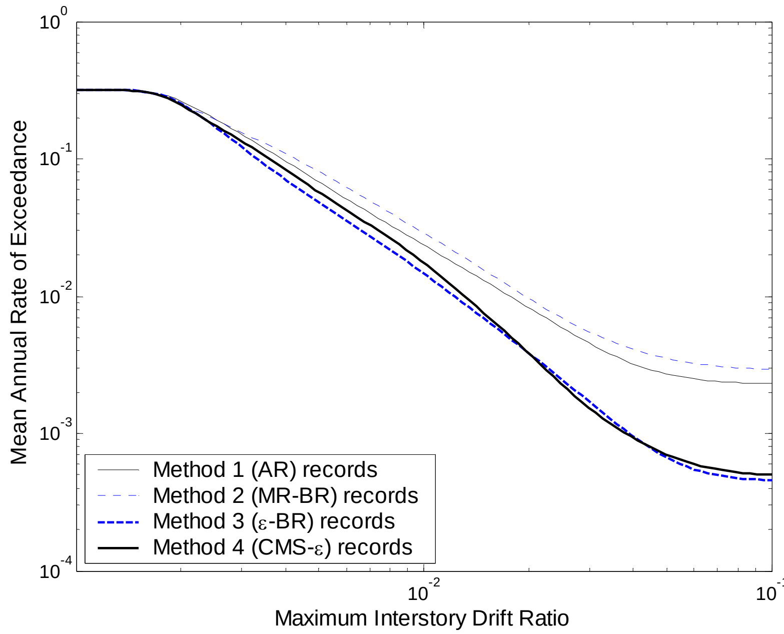


Figure 6.13: Mean annual frequency of exceeding various levels of maximum interstory drift ratio, as computed using the scalar intensity measure $Sa(T_1)$.

We can further verify that the variation among the results in Figure 6.13 is due to variation in spectral shape (which is explained by either ε or the CMS-ε) by incorporating a vector valued intensity measure consisting of $Sa(T_1)$ and ε, using the method of Chapter 3. As discussed in Chapter 2, records selected based on ε values (or equivalently, the CMS-ε spectrum) should produce comparable drift hazard results as records selected randomly but analyzed using a vector valued intensity measure which includes ε. In Figure 6.14, we see that this is indeed the case—the curves lay on top of each other. Unfortunately the curve from the MR-BR Method records does not agree perfectly with the others, but the difference is now less than it was in Figure 6.13 before the effect of ε was accounted for. The results of Figure 6.9 through Figure 6.11 can also be corrected with the vector IM including ε, and here the ε parameter also causes these results to agree more closely.

Vector valued intensity measures can also be used to incorporate the effect of magnitude or distance with the ε-specified records. However, incorporating magnitude or distance in the intensity measure does not result in appreciable changes in the drift hazard curve (consistent with the findings in Chapter 3). This suggests that ε (or its implied effect on mean spectral shape) should be given primary consideration when selecting records, with lesser consideration given to magnitude or distance.

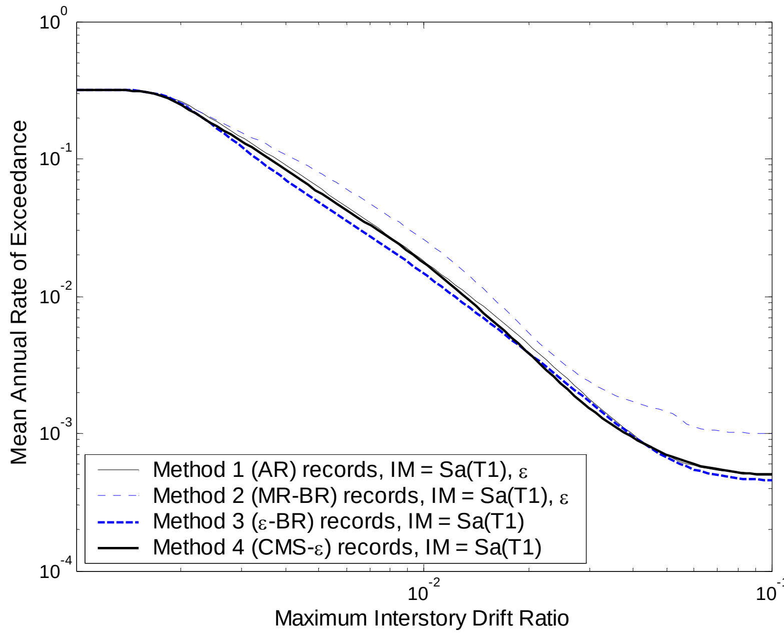


Figure 6.14: Mean annual frequency of exceeding various levels of maximum interstory drift ratio, as computing using either the scalar intensity measure $Sa(T_1)$ or the vector intensity measure $Sa(T_1)$ and ε.

6.7 Conditional mean spectra vs. uniform hazard spectra

The target spectrum used most frequently today for analysis of buildings is the Uniform Hazard Spectrum (UHS) (Kramer 1996, ASCE 2002, McGuire 2004). A uniform hazard spectrum is defined as the locus of points such that the spectral acceleration value at each period has an exceedance probability equal to the specified target probability. When developing this spectrum, the PSHA analysis at each period is performed independently of all other periods, so it is important to remember that nothing can be said about the joint occurrence or exceedance of all of these spectral values simultaneously. Because of this, treating the UHS as the spectrum of a single earthquake event is questionable, as has been noted by others (e.g., Reiter 1990, Naeim and Lew 1995, Bommer et al. 2000). Because of these concerns regarding the reasonableness of the UHS as a design spectrum, nuclear industry design procedures specify a design spectrum similar to the CMS-ε proposed here, but without considering ε (DOE 1996, NRC 1997, ASCE 2005). The difference in requirements between the nuclear and building design procedures indicates that no consensus has yet emerged regarding appropriate design spectra.

It is well-known that a Uniform Hazard Spectrum has difficulties in representing a single earthquake at a given site. This problem is most often explained as follows: the high-frequency portion of the UHS is frequently dominated by small nearby earthquakes, while the low-frequency portion is dominated by larger, more distant earthquakes. Because the high-frequency and low-frequency portions come from different events, no single earthquake will produce a response spectrum as high as the UHS throughout the frequency range considered (Reiter 1990, NRC 1997, DOE 1996). While it is true that no single earthquake is likely to produce a spectrum as high as the UHS, an underappreciated reason for this is the variability in spectral values (for a given magnitude and distance). At each period, the spectral acceleration value given a magnitude and distance is random—it could be higher or lower than the mean prediction. For low annual probability (long return period) design criteria, *the uniform hazard spectrum is almost always higher than the mean spectrum (or spectra) for the dominant event (or events) at a site* (i.e., ε is greater than 0, as explained in Chapter 3). But even if a particular record has a spectral acceleration value as large as the UHS at a given period, it is unlikely to be as high as the uniform hazard spectra at all periods. Thus, a spectrum that is equal to the UHS at a range of periods has a much lower probability of exceedance than the probability level assigned to the UHS. Therefore, using a UHS as a target spectrum for probabilistic analysis would be conservative, due to variations in causal magnitudes, distances and epsilons from period to period, as discussed above. An alternative target spectrum is needed if this conservatism is to be eliminated.

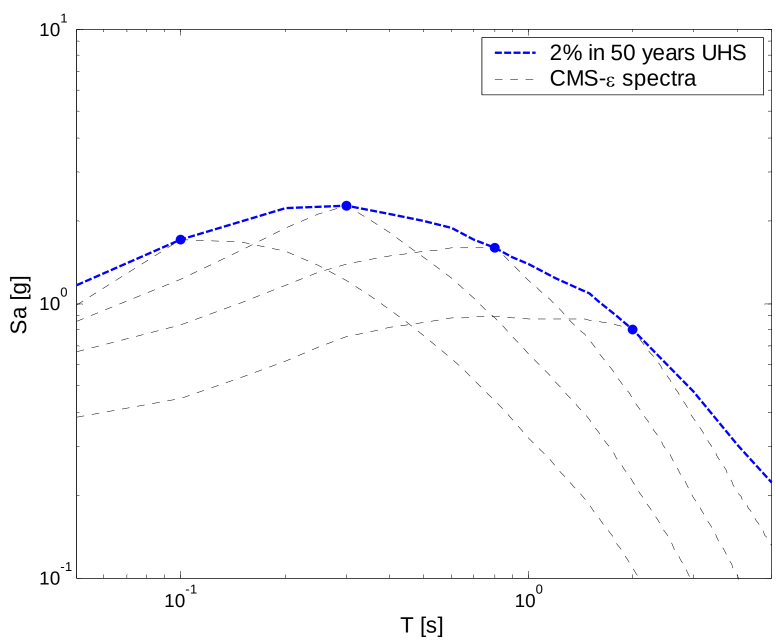


Figure 6.15: 2% in 50 years uniform hazard spectrum at the Van Nuys site, along with several Conditional Mean Spectra, considering ε (CMS-ε), conditioned on $Sa(T)$ at four different values of T (0.1, 0.3, 0.8 and 2 seconds).

The fact that a uniform hazard spectrum is always higher than individual CMS-ε spectra, as illustrated in Figure 6.15, suggests that a UHS may be a reasonable target spectrum in some cases. If one is performing expensive analyses or experiments on a system with an unknown period or many sensitive periods *and* cannot run many tests, the UHS could safely be used as a target spectrum, recognizing that the results will not be unconservative, although they may be quite conservative. But if one can afford to perform many analyses, a less biased estimate of responses can be obtained by using CMS-ε spectra conditioned on target *Sa* values at several periods, and taking the envelope (or some other combination) of responses estimated with records based on these spectra. That is, rather than estimating the response from an envelope of spectral values (the UHS), one could estimate the envelope of responses from several CMS-ε spectra. Using several CMS-ε spectra should be less conservative than using the UHS, and more suited for use in probabilistic assessments. Note that this procedure follows the nuclear industry design guidelines (e.g., DOE 1996), except that the important effect of ε has been included in this derivation of target conditional mean spectral shapes.

6.8 Bias from record scaling

The analysis above required use of (generally upward) record scaling in order to obtain records with the desired target $Sa(T_1)$ values. The use of scaling leads to questions about whether the scaled ground motions are truly representative of ground motions with the given *IM* level. Seismologists and engineers sometimes worry that simple record scaling violates the physics of the earthquake ground motion process. Here we will take a more pragmatic view, and conclude that record scaling is "valid" and useful as long as records scaled to a given *IM* level produce the same levels of structural response as ground motions naturally at the target *IM* level (i.e., scaled records do not produce biased estimates of structural response).

One very straightforward way to detect a bias from scaling is to take a suite of records that have been scaled to a given *Sa* level, and plot the structural response caused by each record versus the factor by which the record was scaled. This is done for an *Sa*(0.8s) value of 0.6g with each of the four proposed record selection methods, and displayed in Figure 6.16. By examining the data visually and considering the regression predictions plotted on the figures, we can examine whether the records with large scale factors induce different responses in the structure than records with a scale factor near one. If the regression line has a slope of zero, then records with large scale factors are unbiased (i.e., the mean estimated response does not depend on whether the record has been scaled). We see in Figure 6.16 that the record sets selected with the AR Method and the MR-BR Method show some bias, while the record sets selected with the ε-BR Method and the CMS-ε Method show no bias.

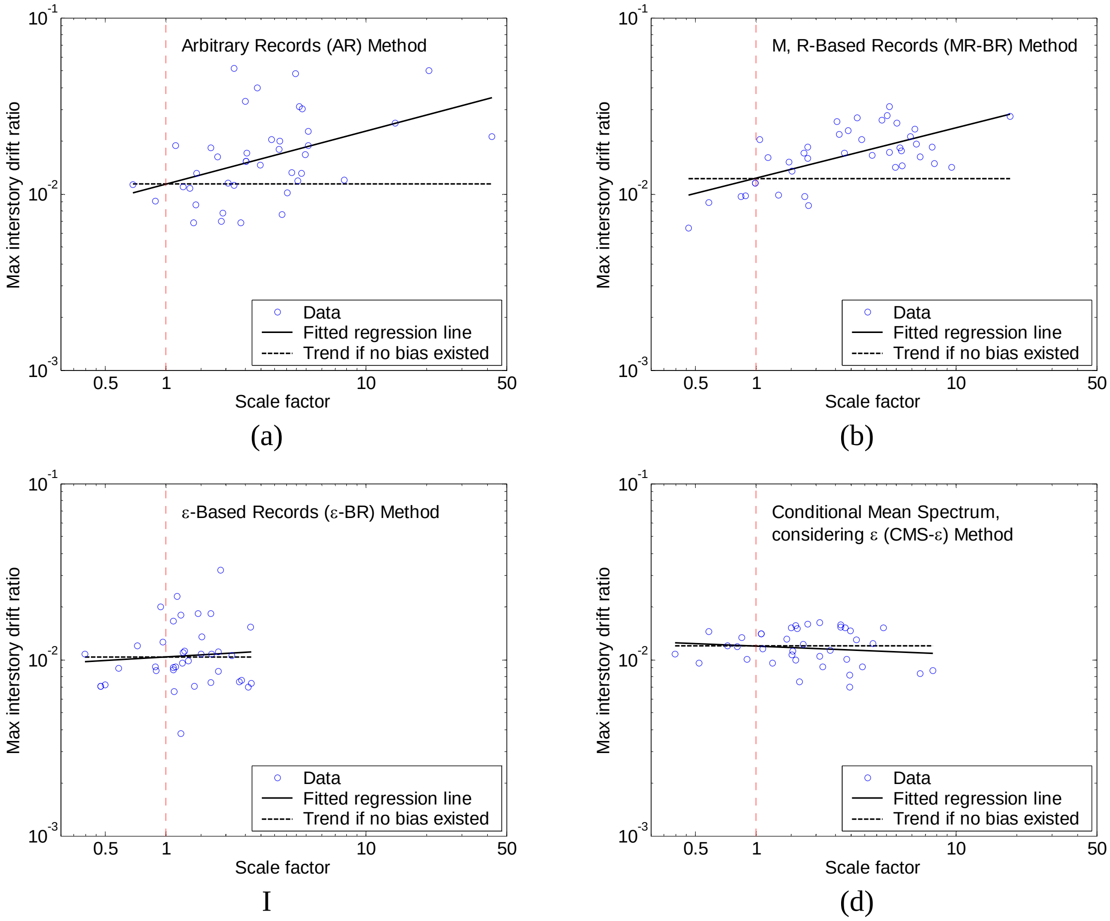


Figure 6.16: Maximum interstory drift ratio versus record scale factor for each of the four selection methods considered, at an *Sa*(0.8s) level of 0.6g. Regression fits based on scale factor are shown with solid lines. Dashed horizontal lines corresponding to the mean prediction at a scale factor of one are shown for comparison. (a) Records using the AR Method. (b) Records using the MR-BR Method. (c) Records using the ε-BR Method. (d) Records using the CMS-ε Method.

The significance of the slopes from the regression analyses of Figure 6.16 can be measured using a common statistical diagnostic tool known as the F-test (Neter et al. 1996). This test produces a probability (referred to as a p-value) that the estimated slope would as large as, or larger than, the observed slope given that there was no underlying trend in the data (i.e., the probability of erroneously estimating a given slope due to an imprecise estimate from a finite data sample). P-values for the four record-selection methods are reported for six *Sa* levels in Table 6.4 (higher *Sa* levels are not included because the number of records causing collapse is significant and so regression on non-collapsing records becomes less meaningful). The row of this table associated with *Sa*(0.8s) = 0.6g provides the p-values for the regressions of Figure 6.16. For this *Sa* level, the method using a range of *M*, *R* and ε values and the method based on *M*, *R* record selection have low p-values, indicating that the observed trend is statistically significant. The large p-values for the other two methods indicate that there is likely no underlying trend. These conclusions fit with intuitive observations made by visually examining Figure 6.16. When examining all levels of *Sa* in this table, some trends are apparent: methods 1 and 2 show significant trends with scale factor, while methods 3 and 4 do not. Thus, we can conclude that if records are selected based on ε or spectral shape, then they can be safely scaled without inducing a bias in the estimated responses (this is supported by additional results from the two structures considered in Appendix F).

Table 6.4. P-values from regression prediction of max interstory drift ratio as a function of scale factor for four methods of record selection, at six levels of *Sa*(0.8s). P-values of less than 0.05 are marked in boldface.

Sa(0.8s)	Method 1: Arbitrary Records	Method 2: M, R-Based Records	Method 3: ε-Based Records	Method 4: CMS-ε Method
0.1	**0.00**	**0.01**	0.21	0.68
0.2	**0.01**	0.89	0.33	0.07
0.4	**0.01**	0.46	0.73	0.51
0.6	**0.00**	**0.00**	0.67	0.44
0.8	**0.00**	**0.00**	0.73	0.30
1	**0.05**	**0.04**	0.23	0.37
median p-value	**0.01**	**0.02**	0.50	0.40

Several previous studies have investigated the effect of record scaling, and so their conclusions can be compared the results in this paper. Shome et al. (1998) found that an *M*, *R* “bin” of records (a set of records with similar magnitude and distance values) scaled to the bin-median spectral acceleration at the fundamental period of the structure will not produce biased median structural response values relative to unscaled records; a conclusion further confirmed by Iervolino and Cornell (2005a). The magnitude-distance bin approach for selecting records used in that paper is very similar to the “Method 2 (*M*, *R* selection)” approach used here, but in this chapter the records were not necessarily scaled to the bin-median spectral acceleration value. Luco and Bazzurro (2005) found that scaling an *M*, *R* “bin” of records to *Sa* values other than the median of the suite may induce some bias in structural response (this scaling procedure was termed “intra-bin” scaling by the authors). This finding is consistent with the results seen using the *M*, *R* -based record selection scheme in Figure 6.16b. The conclusions from these studies do not conflict with the findings here.

The conclusion here that inappropriate record scaling can bias the estimated structural response supports the concern expressed by others that record scaling might fail to modify all ground motion properties in an appropriate way. For example, Han and Wen (1994) state that "scaling an earthquake to attain a target damage level of different intensity is questionable since scaling a ground motion does not account for variations in ground motion characteristics (e.g., frequency content) which change with intensity." Through the exploration of conditional mean response spectra above, we have seen that the frequency content of ground motions does in fact change as the intensity (i.e., $Sa(T_1)$) changes. What is probably unexpected for most readers, however, is that the frequency content is more affected by the variation of ε than by the variation of magnitude or distance. Further, if we select records with the desired spectral shape through a careful record selection scheme (i.e., ε-BR or CMS-ε selection), then we *can* scale records without inducing bias.

6.9 Do we really want records with a peak in their spectrum? Consideration of spectral acceleration averaged over a period range

The previous discussion depends upon measuring earthquake intensity with $Sa(T)$—that is, spectral acceleration at a single period. This intensity measure is a perfect predictor of structural response for elastic single-degree-of-freedom systems with natural period T. For multiple-degree-of-freedom structures, $Sa(T)$ has also been seen to effectively predict even nonlinear structural response when the period T is chosen to equal the first-mode period of the structure, especially if the structure is "first-mode dominated" (Shome et al. 1998).

Given that earthquake intensity will be measured with $Sa(T_1)$, we have seen above that extreme (rare) values of $Sa(T_1)$ are not associated with equally extreme Sa values at other periods, due to a lack of perfect correlation among response Sa values at differing periods. This phenomenon results in the spectrum of rare ground motions (as defined by their $Sa(T_1)$ level) having a peak at $Sa(T_1)$, as was seen earlier, and in Chapter 3. But if the structural response parameter of interest is sensitive to Sa values at multiple periods, then perhaps this peaked spectrum should not be of primary concern (loosely speaking, rather than worrying about a spectrum that is 'very' strong at a single period, one might worry about a spectrum that is 'somewhat' strong at several periods). Perhaps an intensity measure which averages spectral acceleration values over a range of periods is a better indicator of structural response (this class of intensity measure was first used by Kennedy et al. 1988). In this section we will discuss what happens to target spectra if our IM is Sa averaged over a range.

Here we will consider the case of the geometric mean of spectral acceleration values at a set of periods:

$$Sa_{avg}(T^1,...,T^n) = \sqrt[n]{\prod_{i=1}^{n} Sa(T^i)} \tag{6.3}$$

where $T^1,...,T^n$ are the n periods of interest. This can also be expressed as an (arithmetic) mean of logarithmic spectral acceleration values

$$\ln Sa_{avg}(T^1,...,T^n) = \frac{1}{n}\sum_{i=1}^{n} \ln Sa(T^i) \tag{6.4}$$

This formulation is convenient, because attenuation laws can be easily developed for an $\ln Sa_{avg}$ with an arbitrary set of periods, $T^1,...,T^n$, using existing attenuation models and the correlation model of Chapter 8. The mean and variance of $\ln Sa_{avg}(T^1,...,T^n)$ are given by

$$E\left[\ln Sa_{avg}(T^1,...,T^n)\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[\ln Sa(T^i)\right] \tag{6.5}$$

$$Var\left[\ln Sa_{avg}(T^1,...,T^n)\right] = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \rho_{\ln Sa(T^i),\ln Sa(T^j)}\sigma_{\ln Sa(T^i)}\sigma_{\ln Sa(T^j)} \tag{6.6}$$

where $E\left[\ln Sa(T^i)\right]$ and $\sigma_{\ln Sa(T^i)}$ are the conditional mean and standard deviation of $\ln Sa(T)$, available from popular ground motion attenuation models (e.g., Abrahamson and Silva 1997), and $\rho_{\ln Sa(T^i),\ln Sa(T^j)}$ is available from Equation 8.9 of this dissertation. (Note that Equations 6.5 and 6.6 are the conditional mean and variance given magnitude, distance, etc., as with standard ground motion prediction models for $\ln Sa(T)$.) Further, if individual $\ln Sa(T^i)$ values assumed to be jointly Gaussian, as is often done (Bazzurro and Cornell 2002, Stewart et al. 2002), their sum is also Gaussian, and thus the mean and variance of Equations 6.5 and 6.6 completely define the distribution of $\ln Sa_{avg}(T^1,...,T^n)$. We can then proceed to perform probabilistic seismic hazard analysis for this intensity measure, exactly as we would for any single spectral acceleration value.

Further, we can again develop the Conditional Mean Spectrum, considering ε (CMS-ε) for this intensity measure, as was done above for $\ln Sa(T)$ when conditioned on the $\ln Sa(T_1)$ level. Analogously, we are now interested in the conditional mean of $\ln Sa(T)$ versus T, conditioned on the $\ln Sa_{avg}(T^1,...,T^n)$ level. The critical value to be determined for this calculation is the correlation coefficient between any $\ln Sa(T)$ and $\ln Sa_{avg}(T^1,...,T^n)$ (because $\ln Sa(T)$ values at multiple periods are assumed to be jointly Gaussian, then a linear correlation coefficient completely defines the dependence between $\ln Sa(T)$ values or linear functions of them). This correlation coefficient can be computed as

$$\rho_{\ln Sa(T),\ln Sa_{avg}(T^1,...,T^n)} = \frac{\sum_{i=1}^{n} \rho_{\ln Sa(T),\ln Sa(T^i)} \sigma_{\ln Sa(T^i)}}{\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} \rho_{\ln Sa(T^i),\ln Sa(T^j)} \sigma_{\ln Sa(T^i)} \sigma_{\ln Sa(T^j)}}} \qquad (6.7)$$

This correlation coefficient, along with the results from hazard analysis for $\ln Sa_{avg}(T_1,...,T_n)$, will allow us to develop a target spectrum analogous to the CMS-ε presented above[18].

For illustration, we consider a simple case of Equation 6.4 consisting of periods at only T_1 (the first-mode period of the considered structure) and $2T_1$. This is in fact equivalent to the intensity measure considered by Cordova et al. (2001). We will again consider a T_1 value of 0.8s, so that $2T_1$=1.6s. Using the Abrahamson and Silva (1997) ground motion prediction model for $\ln Sa(0.8s)$ and $\ln Sa(1.6s)$, and the prediction for $\ln Sa_{avg}(0.8s,1.6s)$ based on the above equations, three hazard analyses were performed for the Van Nuys site described above[19]. Their hazard curves are shown in Figure 6.17. Further, at the 2% in 50 year hazard level, the CMS-ε is computed for each of the three IMs using their respective IM values and associated disaggregations on M, R and ε. The spectra are displayed in Figure 6.18. Note that with the $\ln Sa_{avg}(0.8s,1.6s)$ CMS-ε spectrum, the peaks at individual periods no longer exist, but this smoother spectrum is different than the uniform hazard spectrum (it will always be lower than the uniform hazard spectrum).

[18] Although Equations 6.5, 6.6 and 6.7 appear complicated, they are nearly always computed within a computer program by using "for" loops and calls to functions previously defined for other purposes. Thus, in practice the computation is rather simple. Further, for most ground motion prediction models, these equations simplify to the original prediction equation, only with new coefficients. Therefore, probabilistic seismic hazard analysis software (the application where these equations would be used) need only call the same ground motion prediction subroutine but with a modified set of coefficients than those for a fixed single T value.

[19] This hazard analysis was also performed in the paper by Cordova et al. (2001).

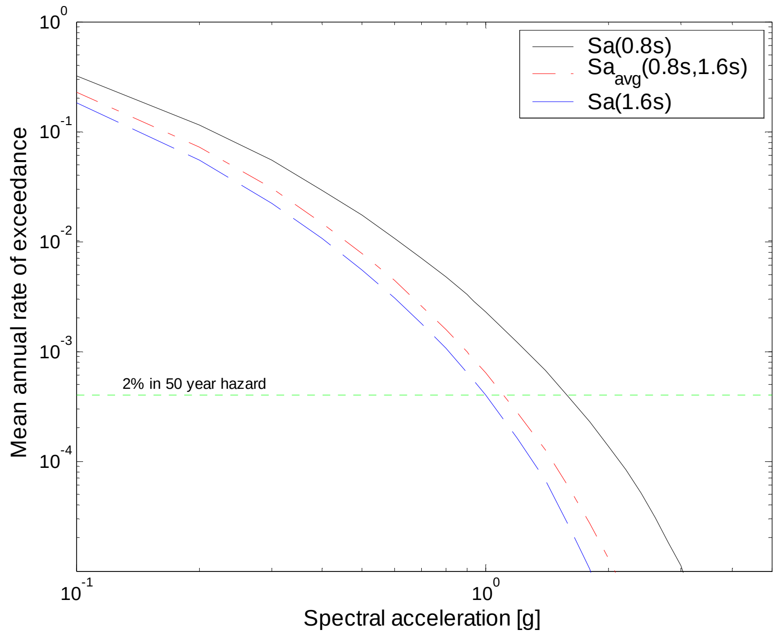


Figure 6.17: Hazard curves for $\ln Sa(0.8s)$, $\ln Sa(1.6s)$, and $\ln Sa_{avg}(0.8s,1.6s)$ at the Van Nuys site.

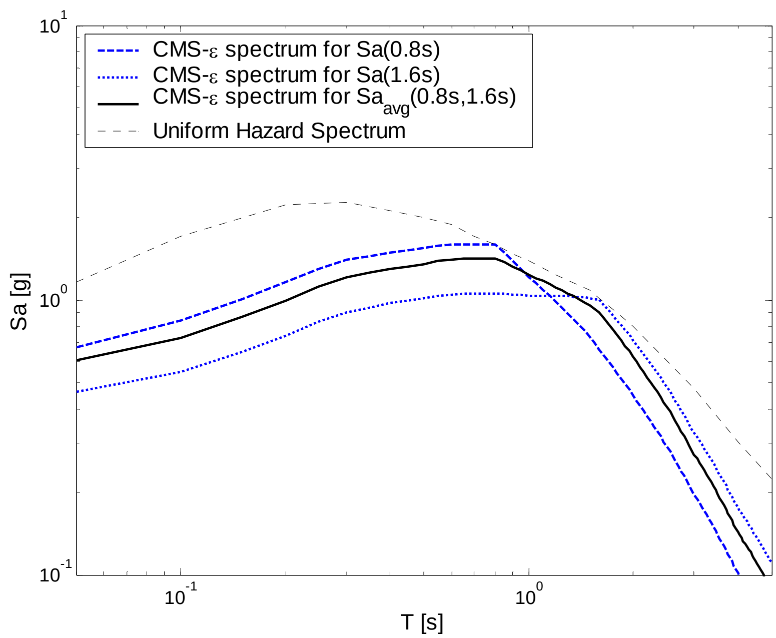


Figure 6.18: CMS-ε spectra for $\ln Sa(0.8s)$, $\ln Sa(1.6s)$, and $\ln Sa_{avg}(0.8s,1.6s)$ at the 2% in 50 year hazard level, and the 2% in 50 year uniform hazard spectrum.

In addition to the conditional mean, we can examine the conditional variance of spectral values, given each of these IMs. The conditional mean +/- one standard deviation is shown for

$\ln Sa(0.8s)$ and $\ln Sa_{avg}(0.8s,1.6s)$ in Figure 6.19. Note that the conditional standard deviation of $\ln Sa(0.8s)$ decreases to zero at 0.8s. The conditional standard deviation of $\ln Sa_{avg}(0.8s,1.6s)$ is nonzero everywhere, but it is reduced through the entire range from 0.8s to 1.6s. So by giving up perfect knowledge of *Sa* at 0.8s, the $\ln Sa_{avg}(0.8s,1.6s)$ IM is able to provide some improved information about *Sa* values at a range of periods. This can also be seen in Figure 6.20, where actual records have been scaled to these target IM values. While the spectra in Figure 6.20a display the characteristic "pinch" at 0.8s, the spectra in Figure 6.20b are never equal at any period. It should be emphasized that although the spectra in Figure 6.20b appear to be casually scaled using a conventional method to minimize the RMS error between the record spectra and a target spectrum over a range of periods (e.g., ASCE 2002), they have actually been scaled precisely so that each record has a common specified value of $Sa_{avg}(0.8s,1.6s)$. Therefore they can be used unambiguously to obtain a drift hazard curve as was done earlier when $Sa(0.8s)$ was the IM.

Although some insight can be gained from examining response spectra as is done below, no structural assessments are performed as part of this study. Whether this intensity measure is a more efficient predictor of response than the traditional $Sa(T_1)$ will depend upon the level of nonlinearity in the structure, and whether the structure is sensitive to multiple periods. An investigation of this type has been performed by Cordova et al. (2001) for a single structure, but more work is needed to draw general conclusions.

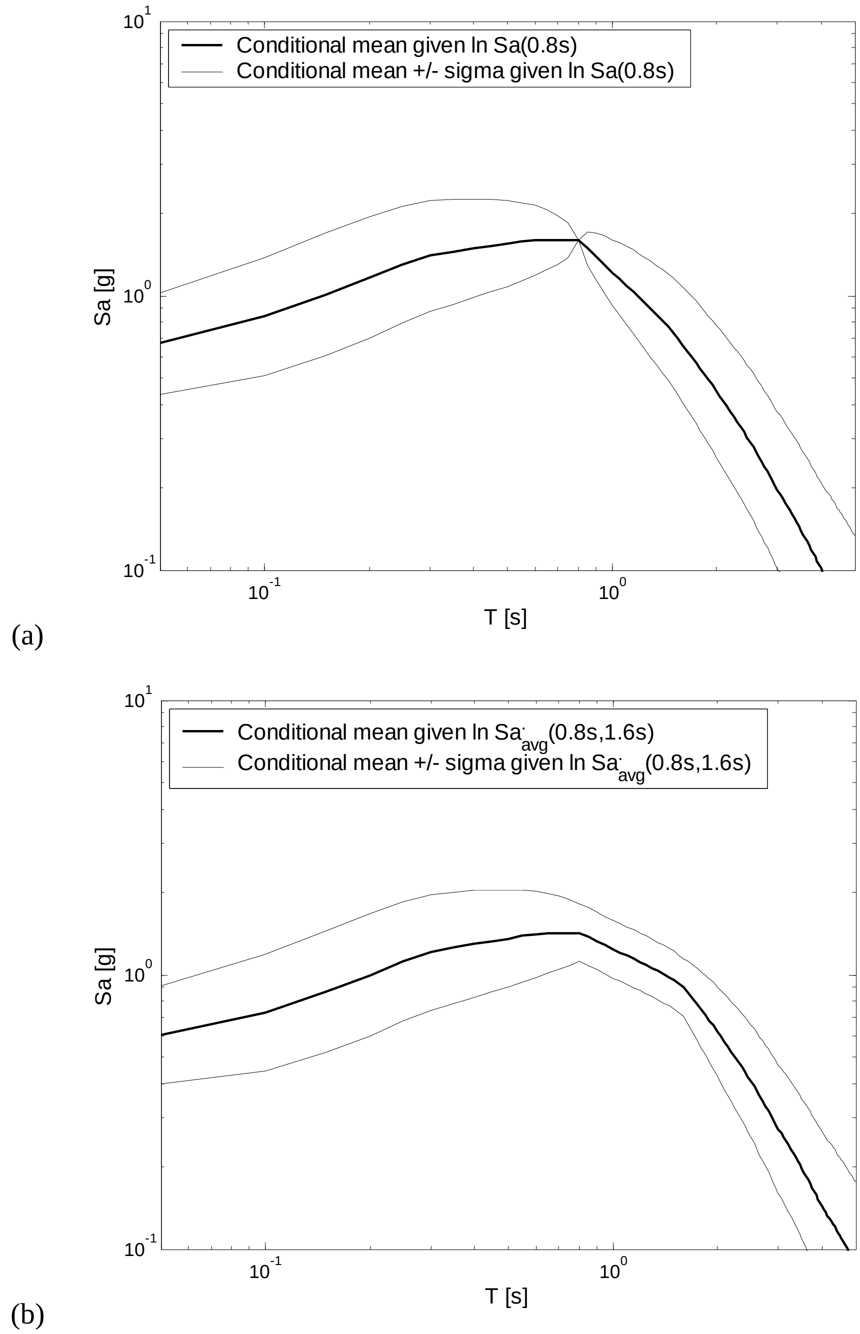
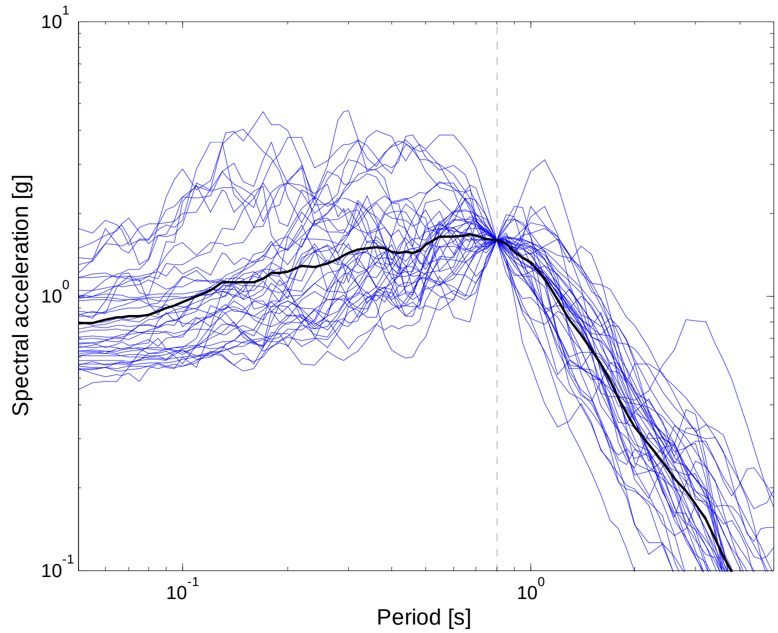
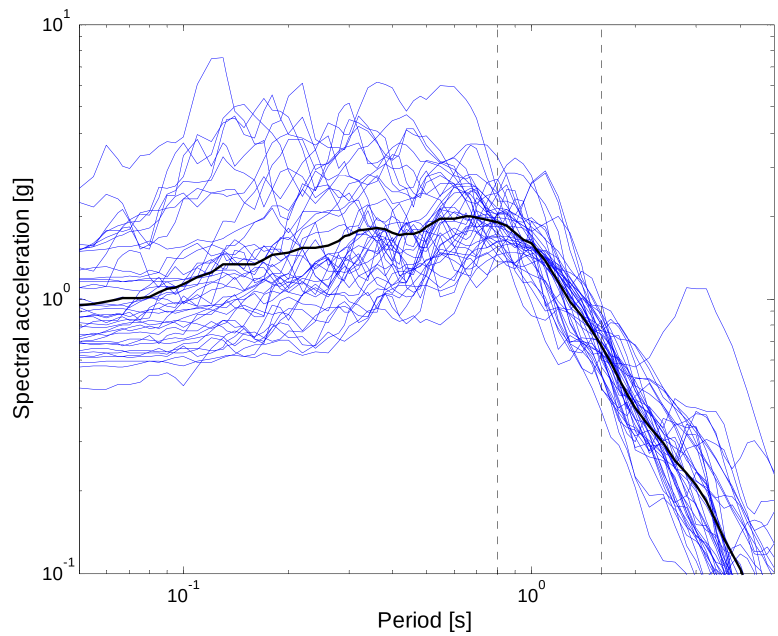


Figure 6.19: Conditional Mean Spectrum, considering ε, and +/ one standard deviation at the 2% in 50 year hazard level given (a) $\ln Sa(0.8s)$, and (b) $\ln Sa_{avg}(0.8s, 1.6s)$.



(a)



(b)

Figure 6.20: Records scaled to target 2% in 50 year IM levels. (a) Records scaled to $\ln Sa(0.8s)$. (b) Records scaled to $\ln Sa_{avg}(0.8s, 1.6s)$.

If a structure is sensitive to a range of periods, the possibility exists that the IM $\ln Sa_{avg}(T^1, ..., T^n)$ could be designed specifically to incorporate these periods. The Sa values at different periods could even have differing weights to reflect to some degree the relative importance of each to structural response (Shome and Cornell 1999, Cordova et al. 2001, Mori et al. 2004). Using a measure such as this, it may be possible to desensitize the structural response to remaining spectral shape variation, so that differences between the record-selection schemes considered above are not as critical. However, care would need to be taken to choose the intensity measure appropriately for that structure. Further, the ground motion hazard analysis would need to be performed specifically for the new intensity measure. This possibility is to be considered in future research.

The CMS-ε spectrum can be easily computed when measuring ground motion intensity as the average of spectral acceleration values at a set of periods. The resulting CMS-ε are not as 'peaked' as CMS-ε based on spectral acceleration at a single period. These smoother spectra might be desirable if a structure's response is sensitive to a range of periods. However, the CMS-ε obtained in this way still differs from a uniform hazard spectrum.

6.10 Conclusions

Selection of earthquake ground motions as input to dynamic analysis of structures has been investigated, with the aim of accurately measuring the distribution of structural response associated with earthquake ground motions of a specified intensity, as measured by $Sa(T_1)$. The structural response parameter of interest here was the maximum interstory drift ratio observed in the building. The goal was to identify which properties of ground motions affect the response of a nonlinear multi-degree-of-freedom structure, given that the motion has a specified $Sa(T_1)$ level. To investigate this question, records were selected using several methods: use arbitrary records, select records to match the causal magnitude and distances at the site of interest, or select records to match the causal ε values at the site of interest. Causal values of magnitude, distance *and* ε depend upon the site of interest and the earthquake intensity level of interest, and are determined from probabilistic seismic hazard disaggregation. A method for calculating the expected spectral shape given a specified $Sa(T_1)$ level and its associated causal magnitude, distance and ε values was presented. The resulting spectrum was termed the Conditional Mean Spectrum, considering ε, or CMS-ε. As a fourth record-selection alternative, records were selected to match this CMS-ε, without trying to match those values directly. This CMS-ε is very similar to spectra used for design of nuclear facilities, except that the effect of ε is not considered in those spectra.

The response spectra of records selected using these four methods were examined, and it was seen that the CMS-ε spectrum at a rare, extreme ground motion intensity (as measured by $Sa(T_1)$) has a 'peak' at the T_1. Records selected based on their ε values, or based on their match with the CMS-ε, displayed this peak, while records selected at random or selected based on magnitude and distance did not. For each of these of these four record-selection methods, the median structural response, standard deviation of structural response, probability of collapse, and drift hazard were computed at a range of ground motion intensity levels. It was found that the records selected based on ε or based on their match with the CMS-ε produced unbiased structural response results using only $Sa(T_1)$ as an intensity measure. However, for the records selected at random and the records selected based on magnitude and distance, it was necessary to use a vector IM including $Sa(T_1)$ and ε in order to obtain unbiased structural response results.

Because records were scaled for analysis, the records were studied to determine whether scaling had induced any bias in the resulting structural responses. It was seen that the records selected based on ε or spectral shape could be safely scaled without introducing any bias, whereas the records selected at random or selected based on magnitude and distance had biased structural responses when scaled.

These conclusions can be explained by recognizing that spectral shape (i.e., spectral acceleration values at other periods, given $Sa(T_1)$) is the record property that directly affects structural response, whereas magnitude, distance and ε are merely proxies for spectral shape. However, variation among ε values has a greater affect on spectral shape than variation in either magnitude or distance, explaining why matching ε was seen to be more critical in obtaining correct structural response levels. All selected records were recorded on stiff soil sites, in order to match the soil conditions at the example site of interest. It is likely also important to match soil type, especially at soft soil sites, due to its effect on spectral shape (Stewart et al. 2001).

Response spectra of ground motions whose intensity is defined based on spectral acceleration averaged over a range of periods is also considered. It is seen that the mean spectra of motions defined in this way do not have the ‘peak’ seen in the previous mean conditional spectra. This definition of intensity provides improved but limited information about spectral values at a defined set of periods, at the cost of giving up perfect information about spectral acceleration at T_1. Probabilistic seismic hazard analysis, record selection and target spectra could all be obtained for this intensity measure in the same way as for the standard $Sa(T_1)$ intensity measure. The question of which of these two intensity measures is more efficient for analysis of a given structure was not considered in this paper.

Based on the findings in this chapter, the following suggestions for record selection can be made: the record property ε at the period T_1 is the most important property to try to match when selecting ground motions for analysis, where the ground motion intensity is measured using $Sa(T_1)$. This applies for estimation of mean response at a given $Sa(T_1)$, as well as for fully probabilistic “drift hazard” assessment. The magnitude and distance values of the ground motions may also be important, but when selecting records, ε values should not be ignored in an effort to match magnitude and distance. As an alternative to selecting earthquake records based on magnitude, distance and/or ε, one can instead use the procedure outlined above (Method 4) to determine the Conditional Mean Spectrum, considering ε, given the $Sa(T_1)$ level of interest. Records can then be selected to match this spectrum without worrying further about other record properties (magnitude, distance and ε, for example). This spectral-shape-matching approach is appealing not only because it avoids response prediction bias by accounting for ε, but also 1because there are by definition few records with ε values equal to the ε values of rare intense ground motions of engineering interest. This lack of records limits the application of Method 3. There are, however, relatively more records with a spectral shape similar to the spectral shape of rare ground motions, which means that there are potentially a larger number of appropriate ground motions available with this approach.

Chapter 7

Which Spectral Acceleration Are You Using?

Baker J.W., and Cornell C.A., 2006. "Which Spectral Acceleration Are You Using?" *Earthquake Spectra*, **22** (2), (in press).

7.1 Abstract

Analysis of the seismic risk to a structure requires assessment of both the rate of occurrence of future earthquake ground motions (hazard) and the effect of these ground motions on the structure (response). These two pieces are often linked using an intensity measure such as spectral acceleration. However, earth scientists typically use the geometric mean of the spectral accelerations of the two horizontal components of ground motion as the intensity measure for *hazard* analysis, while structural engineers often use spectral acceleration of a single horizontal component as the intensity measure for *response* analysis. This inconsistency in definitions is typically not recognized when the two assessments are combined, resulting in un-conservative conclusions about the seismic risk to the structure. The source and impact of the problem is examined in this paper, and several potential resolutions are proposed. This discussion is directly applicable to probabilistic analyses, but also has implications for deterministic seismic evaluations.

7.2 Introduction

Calculation of the risk to a structure from future earthquakes requires assessment of both the probability of occurrence of future earthquakes (hazard) and the resulting response of the structure due to earthquakes (response). The analysis of hazard is typically performed by earth scientists (e.g., seismologists or geotechnical engineering scientists), while the analysis of response is typically performed by structural engineers. The results from these two specialists must then be combined, and this is often done by utilizing an intensity measure (IM) (Banon et al.

2001, Cornell et al. 2002, Moehle and Deierlein 2004). Earth scientists provide the probability of occurrence of varying levels of the IM (through hazard maps or site-specific analysis), and structural engineers estimate the effect of an earthquake with given levels of the IM (using dynamic analysis or by associating the IM with the forces or displacements applied in a static analysis).

Spectral acceleration, *Sa*, is the most commonly used intensity measure in practice today for analysis of buildings. This value represents the maximum acceleration that a ground motion will cause in a linear oscillator with a specified natural period and damping level. (In fact the true measure is pseudo-spectral acceleration, which is equal to spectral displacement times the square of the natural frequency, but the difference is often negligible and the name is often shortened to simply "spectral acceleration,"). But *Sa* is often defined differently by earth scientists and structural engineers. The difference originates from the fact that earthquake ground motions at a point occur in more than one direction. While structural engineers often use the *Sa* caused by a ground motion along a single axis in the horizontal plane, earth scientists often compute *Sa* for two perpendicular horizontal components of a ground motion, and then work with the geometric mean of the *Sa*'s of the two components. Both definitions of *Sa* are valid. However, the difference in definitions is often not recognized when the two pieces are linked, because both are called "spectral acceleration." Failure to use a common definition may introduce an error in the results.

In this paper, the differences in these two definitions are examined, along with the reasons why earth scientists and structural engineers choose their respective definitions. Examples of the use of these definitions are presented, along with the potential impact of failing to recognize the discrepancy. Several procedures for addressing the problem are examined, and the relative advantages and disadvantages of each are considered. Analysis is sometimes performed for each axis of a structure independently, and other times an entire 3D structural model is analyzed at once. Both of these cases are considered, and consistent procedures for each are described. These procedures should be helpful for analysts performing seismic risk assessments of structures.

7.3 Spectral acceleration: two definitions

7.3.1 Treatment of spectral acceleration by earth scientists

The earth scientist's concern with spectral acceleration is in predicting the distribution of spectral acceleration at a site, given an earthquake with a specified magnitude, distance, faulting style, local soil classification, etc. This prediction is made in the form of an attenuation model. Many attenuation models are empirically developed using analysis of recorded ground motions (see Abrahamson and Silva 1997, Boore et al. 1997, Campbell 1997, Sadigh et al. 1997, and Spudich et al. 1999 among many others). There is scatter in this recorded data (due to path effects, variation in stress drop, and other factors which are not captured by the attenuation model), which must be dealt with during development of the attenuation model.

The observed variability in spectral acceleration is well represented by a lognormal distribution (Abrahamson 1988, 2000b). Thus, attenuation models work with the mean and standard deviation of the logarithm of *Sa*, which can be represented by a Gaussian distribution. The broad variability of the distribution hinders estimation of the mean value of ln*Sa* needed for the attenuation law. The log *Sa*'s of two perpendicular components of the ground motion are thus averaged, reducing the variance and allowing the mean value of ln*Sa* to be estimated with greater confidence. For example, in Figure 7.1 it is seen that arbitrary-component spectra vary more about the estimated mean than their geometric mean does.

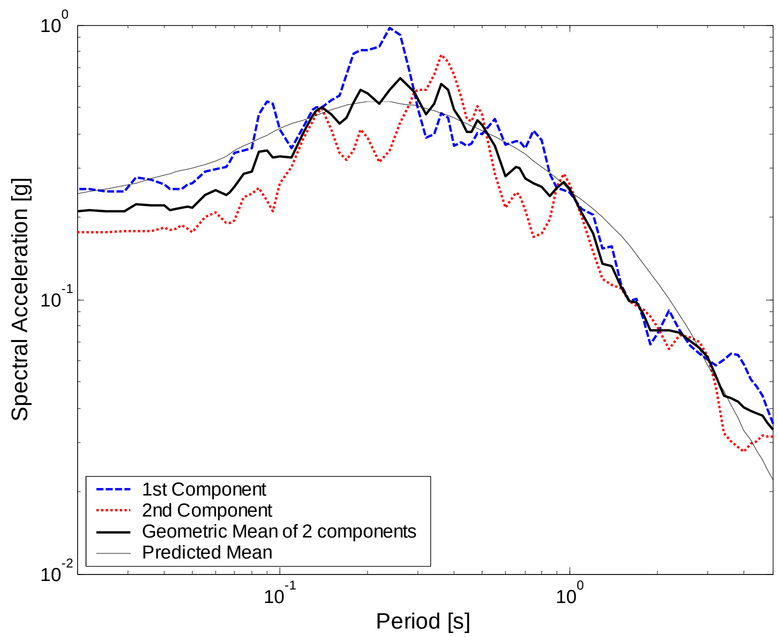


Figure 7.1. Response spectra from the magnitude 6.2 Chalfant Valley earthquake recorded at Bishop LADWP, 9.2 km from the fault rupture. Response spectra for the two horizontal components of the ground motion, the geometric mean of the response spectra, and the predicted mean for the given magnitude and distance using the prediction of Abrahamson and Silva (1997).

The exponential of the mean of the logarithms of two numbers is termed the geometric mean because it is the square root of their product (this is also the same as the SRSS spectral values referred to in section 9.5.7.2.2 of ASCE 2002). For conciseness, we will refer to the geometric mean of spectral acceleration of two components as $Sa_{g.m.,}$ and the spectral acceleration of an arbitrary component will be referred to as Sa_{arb}. The logarithms of these values will be referred to as $\ln Sa_{g.m.}$ and $\ln Sa_{arb}$, respectively. The terms Sa and $\ln Sa$ will be used to refer to spectral acceleration and its logarithm, without specification as to which definition is used. And the standard deviation of $\ln Sa$ will be referred to as the "dispersion" of Sa, following common practice elsewhere. It is noted again that these values are functions of the period and damping level specified, but this is not stated explicitly in the notation because consideration of a particular period and damping are not needed for this discussion.

Attenuation models typically provide a predicted mean and standard deviation for the conditional random variable $\ln Sa_{g.m.}$, given an earthquake magnitude, distance, etc. These estimates for $\ln Sa_{g.m.}$ can be made directly from the data, because the averaging of the two components transformed the observed data into values of $Sa_{g.m.}$. For a given earthquake, the mean of the conditional random variable $\ln Sa_{arb}$ is equal to the mean of $\ln Sa_{g.m.}$. But the standard deviation of $\ln Sa_{arb}$ is greater than that of $\ln Sa_{g.m.}$ by a factor that could be as large as $\sqrt{2}$ if the two components were uncorrelated (because the standard deviation of the mean of 2 uncorrelated random variables with common standard deviation σ is equal to $\sigma/\sqrt{2}$). Calculating the standard deviation of $\ln Sa_{arb}$ thus takes an additional step of going back to the non-averaged data and examining the standard deviation there. This step is taken by some researchers (e.g., Boore et al. 1997, Spudich et al. 1999), but many others have not because it was not recognized as important. However, the difference in standard deviations is in fact relevant for ground motion hazard analysis, as will be seen in the next section.

Deterministic ground motion hazard analysis

The effect of the standard deviation of $\ln Sa$ is easily seen in deterministic seismic hazard analysis. Often in a deterministic hazard analysis, a target spectral acceleration for a "Maximum Considered Event" is computed by specifying a scenario event (magnitude and distance), and then computing the value of $\ln Sa$, at a given period and damping level, that is one standard deviation greater than the mean prediction for that event (Reiter 1990, Anderson 1997). But the value of the standard deviation depends upon whether $Sa_{g.m.}$ or Sa_{arb} is being used as the IM. Because of its greater dispersion (logarithmic standard deviation), the target value of Sa_{arb} will thus be larger than that for $Sa_{g.m.}$. So the target spectral acceleration depends on the definition of Sa being used, even though both definitions have the same mean value of $\ln Sa$. For a "mean plus one sigma" ground motion, Sa_{arb} will thus be larger than that for $Sa_{g.m.}$ by a factor of $\exp(\sigma_{\ln Sa_{arb}} - \sigma_{\ln Sa_{g.m.}})$. For example, using the model of Boore et al. (1997, Boore 2005), this difference is $\exp(0.047)$ at a period of 0.8 seconds with 5% damping, implying that if Sa_{arb} is to be used as the IM, the target spectral acceleration would be about 5% larger than if $Sa_{g.m.}$ is used.

Another method used in deterministic hazard maps is to take as the hazard value 150% of the median spectral acceleration value for a characteristic event (ASCE 2002). One of the justifications for the 150% rule is that this will capture a reasonable fraction of the *Sa* values that could result from occurrence of this characteristic event. However, the fraction captured will vary based on which of the two definitions is used. Consider *Sa* at a period of 0.8 seconds. Per the Boore et al (1997, Boore 2005) attenuation relationship used above, Sa_{arb} has a 23% chance of exceeding 150% of the median *Sa* value given the event, while $Sa_{g.m.}$ has a 21% chance of exceeding 150% of the median *Sa* given the event. Thus the level of conservatism resulting from this rule varies slightly depending on the *Sa* definition used. Spectral acceleration is merely a tool used to simplify the analysis problem; therefore the factor of safety should not vary based on the definition used. In principle the 150% rule for $Sa_{g.m.}$ should be a "156% rule" for Sa_{arb}, in order to provide the same level of conservatism.

Probabilistic ground motion hazard analysis

The variation in standard deviation is also seen in probabilistic seismic hazard analysis. In Figure 7.2, hazard curves for the two definitions of *Sa* are shown for a hypothetical site 8 kilometers away from a recurring magnitude 6.5 earthquake (this simple hazard environment is representative of sites near a single large fault). Again the attenuation model of Boore et al. (1997, Boore 2005) is used, because it provides dispersions for both Sa_{arb} and $Sa_{g.m.}$. We see that the hazard curve for Sa_{arb} is greater than $Sa_{g.m.}$ due to the larger dispersion in Sa_{arb}. This is because ground motion hazard, especially at long return periods, is driven by ground motions that are larger-than-average. So even though Sa_{arb} and $Sa_{g.m.}$ have the same median value for each magnitude/distance considered, the larger-than-average values for Sa_{arb} will be greater than those for $Sa_{g.m.}$. This is the same effect as is seen in the deterministic hazard analysis case. (It is also true that the smaller-than-average values will be smaller for Sa_{arb} than for $Sa_{g.m.}$, but these events make a relatively smaller contribution to hazard, so the larger-than-average and smaller-than-average events are not offsetting.) For this site the Sa_{arb} with a 2% probability of exceedance in 50 years is 12% larger than the corresponding $Sa_{g.m.}$. The hazard for sites near multiple faults is simply a weighted sum of hazard curves similar to that shown in Figure 7.2, so we expect hazard curves at all sites to show this pattern.

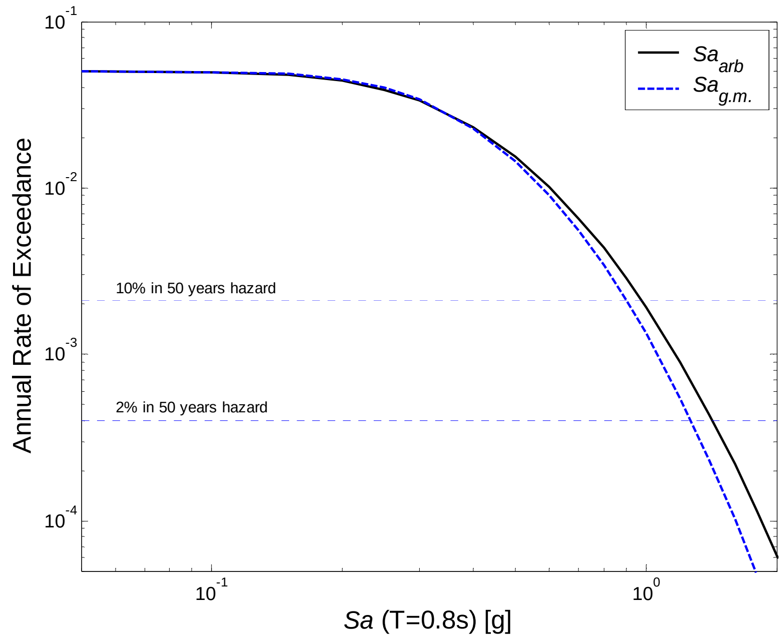


Figure 7.2. Ground motion hazard from a recurring magnitude 6.5 earthquake at a distance of 8 kilometers, for Sa_{arb} and $Sa_{g.m.}$ at a period of 0.8 seconds with 5% damping

The conclusion to be drawn from these examples is that the ground motion hazard is dependent on the definition of spectral acceleration. Even though $Sa_{g.m.}$ and Sa_{arb} have the same median value (for a given magnitude, distance, etc.), the difference in dispersion will cause a difference in ground motion hazard.

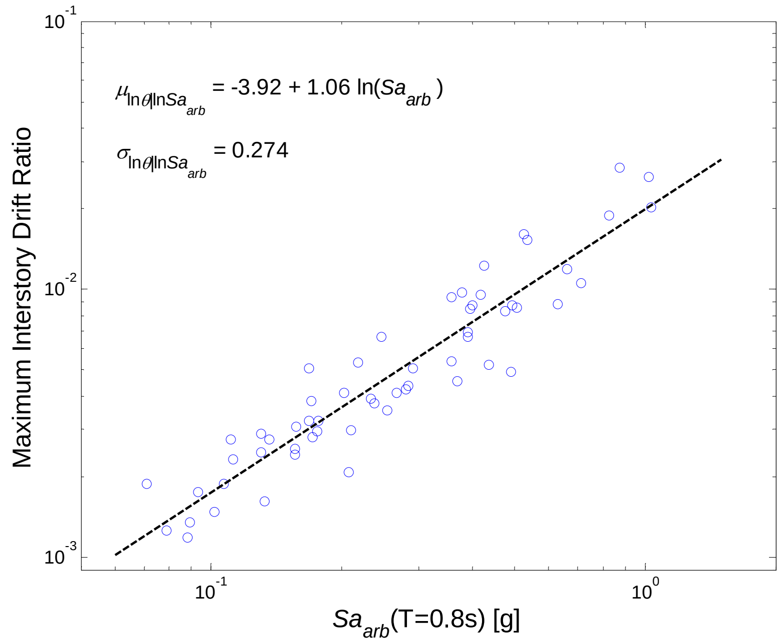
This result is important, because it means that ground motion hazard cannot be used interchangeably for both Sa_{arb} and $Sa_{g.m.}$. This brings to light a problem with the U.S. Geological Survey maps of spectral acceleration hazard (Frankel et al. 2002). The maps are produced using results from several attenuation models, some of which emphasize the dispersion in Sa_{arb}, some of which provide only dispersion for $Sa_{g.m.}$. The U.S. Geological Survey has used the dispersions emphasized by the models' authors, resulting in a mix of both definitions being used. Thus the current maps are not strictly interpretable as the ground motion hazard for either Sa_{arb}, or $Sa_{g.m.}$. This will be addressed in future revisions to the maps, in light of the new recognition of the importance of this issue (Frankel 2004).

7.3.2 Treatment of spectral acceleration by structural engineers

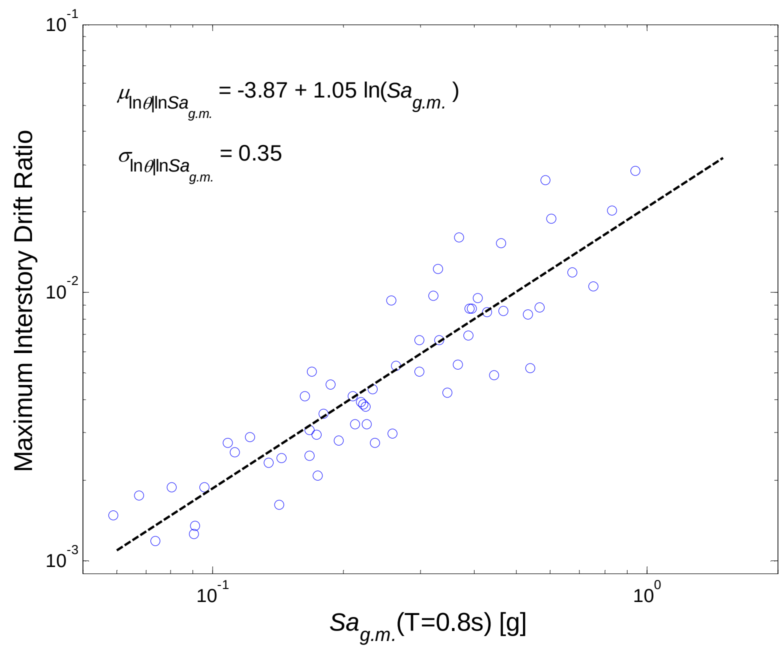
Structural engineers also utilize spectral acceleration as a basis for analysis of structural response. Let us first consider analysis of a single two-dimensional frame of a structure—a common situation in practice. In this case, only a single horizontal component of earthquake ground motion is needed for analysis. Therefore, spectral acceleration is computed only for the selected component at a period equal to the elastic first-mode period of the structure, and that is used as the intensity measure. In most cases, no distinction is made between the two components of a ground motion, so using a single component in this case is equivalent to using Sa_{arb} as the intensity measure. To compute $Sa_{g.m.}$ using both horizontal components of the ground motion, but then use only one of the components, the stronger or the weaker, for analysis would only introduce unnecessary scatter into the relationship between the IM and structural response.

This increased scatter resulting from prediction with $Sa_{g.m.}$ is illustrated in Figure 7.3. Prediction of response of a structure is made using both Sa_{arb} and $Sa_{g.m.}$. A model of an older seven-story reinforced concrete frame, described by Jalayer (2003), is used for analysis. Sixty unscaled recorded ground motions were used to perform nonlinear dynamic analysis. Figure 7.3 shows that the prediction of the mean log maximum interstory drift ratio, denoted $\ln\theta$, is very similar for both intensity measures, but use of $Sa_{g.m.}$ as the IM results in increased dispersion of θ relative to the use of Sa_{arb}, as was anticipated above. The larger dispersion implies that there is greater uncertainty in the estimate of median response (i.e., if $Sa_{g.m.}$ is used as the IM, a greater number of analyses would need to be performed to achieve the same confidence in the mean $\ln\theta$). Thus, the use of Sa_{arb} as the IM is preferable for the structural engineer, in order to minimize the number of nonlinear dynamic analyses performed.

Many examples of the use of *Sa* as an intensity measure exist in the literature. For example, Modal Analysis (Chopra 2001), the SAC/FEMA methodology (FEMA 2000a, b, c), and Incremental Dynamic Analysis (Vamvatsikos and Cornell 2002) all use *Sa* as a predictor of structural response in some cases. In virtually every application of these procedures, Sa_{arb} (as opposed to $Sa_{g.m.}$) is used as the intensity measure for analysis of a single frame of a structure.



(a)



(b)

Figure 7.3. Prediction of the response of a single frame of a structure using (a) the spectral acceleration of the ground motion component used (Sa_{arb}), and (b) the spectral acceleration of the average of both components ($Sa_{g.m.}$).

7.4 Incorrect integration of hazard and response

Spectral acceleration hazard is coupled with response analysis during performance-based analysis procedures (e.g., Cornell and Krawinkler 2000, Cornell et al. 2002). In past application of these procedures, frequently the ground motion hazard analysis has been unwittingly performed with $Sa_{g.m.}$ (to utilize existing attenuation models), and the response analysis has performed with Sa_{arb} (to minimize dispersion in the response prediction), resulting in the inconsistency discussed in this paper. Examples where the authors know only too intimately that a hazard analysis based on $Sa_{g.m.}$ was inadvertently coupled with a response analysis based on Sa_{arb} include Baker and Cornell (2004), Jalayer and Cornell (2003), Yun et al. (2002), and Shome and Cornell (1999). In the work of others it is seldom clear because the question was not discussed, but it can be suspected that if the hazard analysis was based on the USGS hazard maps or popular attenuation laws such as Abrahamson and Silva (1997) and Sadigh et al (1997) where the only reported dispersion is that for $Sa_{g.m.}$, then an inconsistency is likely to exist.

In addition, other design and analysis procedures (e.g., ASCE 2000d, 2002), utilize the U.S. Geological Survey maps of spectral acceleration hazard (Frankel et al. 2002) to attain target Sa values at which the performance of the structure should be checked. Although in this case there is no explicit statement of the reliability of a structure analyzed in this manner, Sa is still used as a link between hazard and response. Thus, it is preferable to define Sa consistently in both the hazard and response. Possibilities for a consistent treatment of the problem are discussed in the following section.

7.5 Valid methods of combining hazard and response

For performance-based analysis procedures, it is necessary that the median and dispersion of response at a given IM level be consistent with the IM definition used for hazard analysis. This can be achieved in several ways, the choice of which may depend in part on the situation and available information. Three proposed solutions for use in analyzing a structure along a single axis are outlined below. The common characteristic of each method is that the IM used for hazard analysis and the IM used for response analysis are consistently defined.

7.5.1 Calculate the ground motion hazard for Sa_{arb}

With this method, the structural response analysis described above is unchanged, but the ground motion hazard analysis is performed for the consistent intensity measure, Sa_{arb}. This allows for the estimation of structural response with less dispersion than when $Sa_{g.m.}$ is used (e.g., see Figure 7.3). And once more attenuation models are developed with dispersion for Sa_{arb}, the hazard analysis is no more difficult than hazard analysis for $Sa_{g.m.}$. The disadvantage is that few current attenuation models provide the dispersion for Sa_{arb}, meaning that many models, and the resulting hazard analysis, cannot be used without modification.

7.5.2 Predict structural response using $Sa_{g.m.}$

With this method, the ground motion hazard is unchanged from current $Sa_{g.m.}$-based practice. Instead, the response analysis is modified, using $Sa_{g.m.}$ as the IM rather than Sa_{arb}. To do this, one would compute the IM of a record as the geometric mean of Sa of the two components of the ground motion, even though only one component will be used for analysis. This method has the advantage of not requiring new attenuation laws or hazard analysis. Unfortunately, it will introduce additional dispersion into the response prediction, as was seen in Figure 7.3, and hence be less efficient as an IM.

7.5.3 Perform hazard analysis with $Sa_{g.m.}$, response analysis with Sa_{arb}, and inflate the response dispersion

This method takes advantage of the fact that the median structural response for a given Sa level is the same whether $Sa_{g.m.}$ or Sa_{arb} is used. Only the dispersion is increased if $Sa_{g.m.}$ is used, as was seen in Figure 7.3. Structural response is thus performed using Sa_{arb} (as per standard practice) to obtain the median response. Then the dispersion in response is inflated to reflect that which would have been seen if $Sa_{g.m.}$ had been used as the intensity measure instead. An estimate of the amount by which the dispersion should be increased can be obtained using a first-order approximation. For this procedure, we assume the following model for the relationship between spectral acceleration and response:

$$\ln\theta = a + b\ln Sa_{arb} + \varepsilon_{arb} \tag{7.1}$$

$$\ln\theta = a + b\ln Sa_{g.m.} + \varepsilon_{g.m.} \tag{7.2}$$

Where θ is the structural response value of interest, a and b are coefficients to be estimated from the data using least-squares regression, and ε's are zero-mean random variables (note that a and b have the same expected value in Equations 7.1 and 7.2, as will be demonstrated in Equations 7.3 and 7.11, and as is supported by the empirical estimates in Figure 7.3). The model is seen to fit well in Figure 7.3, as in many other cases (at least locally). We are interested in estimating the standard deviation of lnθ given $Sa_{g.m.}$, in the case where we know only the standard deviation of lnθ given lnSa_{arb}. From Equation 7.13, we can find the ratio of the two conditional standard deviations. For the example dataset here, $\rho_{\ln Sa_x, \ln Sa_y} = 0.797$ and $\rho_{\ln \theta_x, \ln Sa_x} = 0.942$, implying a ratio between standard deviations of 1.34. Thus the predicted conditional standard deviation for the example problem would be 1.34*0.27=0.36, approximately matching the standard deviation in 2b (0.35).

The SAC procedure (FEMA 2000a, b, c) uses the model of structural response adopted in Equation 7.1, but with b = 1. So an analysis using the SAC procedure would be a natural candidate for this method. One would simply perform analysis using Sa_{arb} as before, but inflate the dispersion in structural response using Equation 7.13 before continuing with the SAC methodology.

In the short term, this third method is attractive because it leaves existing hazard and response procedures unmodified, and instead makes a correction before the two analyses are combined. However, in the long term one of the two more direct methods using a consistent IM for both hazard and response would be more expeditious.

7.5.4 Results from the proposed methods

All three of the methods described should result in the same answer for the probability of exceedance of a given limit state in the structure, aside from the inherent variability in the answer resulting from the statistically uncertain estimates of hazard and response (Baker and Cornell 2003). Additional methods can also be conceived using alternative intensity measures, but the above methods are expected to be the simplest and most similar to current practices.

To illustrate the results of the proposed methods, a drift hazard analysis is performed using the three proposed methods and the previous inconsistent method. This analysis combines the ground motion hazard from Figure 7.2 with the structural response analysis from Figure 7.3 to determine the rate of exceeding a given response level in the structure. The procedure for computing this drift hazard curve is described by Bazzurro et al. (1998). The results are displayed in Figure 7.4. It is seen that the three proposed methods produce comparable results, while the inconsistent method results are unconservative (e.g., the drift level exceeded with a 2% probability in 50 years is underestimated by approximately 10%). While the magnitude of the error is not overwhelmingly large, it nonetheless represents an easily correctible systematic flaw in the procedure.

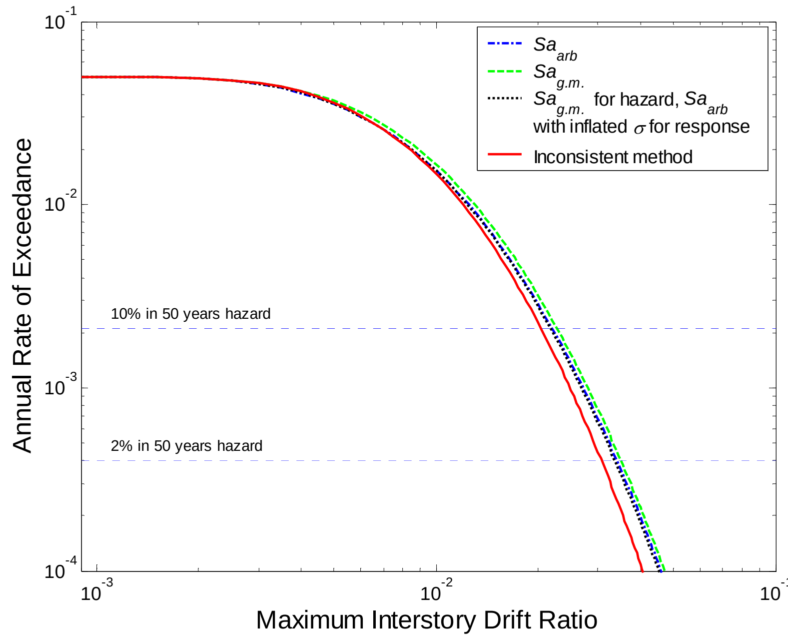


Figure 7.4. Drift hazard as computed using the three methods proposed above and the inconsistent method.

In many applications today, the drift hazard curve is not computed. Rather, the ground motion hazard is used to specify a target spectral acceleration to use in analyzing structural response (i.e., the spectral acceleration associated with a 2% probability of exceedance in 50 years). An inconsistent approach will cause a comparable bias in these calculations as well.

7.6 Analysis of 3D structural models: Combining hazard and response

When analyzing a 3D structural model, both horizontal components of the ground motion are used, so the above procedures using a single component are not necessarily applicable. In this case, as before, the concern is that the intensity measures used in the hazard analysis and response analysis should be consistent. Fortunately, the preferred method for analysis in this case is also a method which is apparently often used in practice. Several potential procedures are discussed here:

7.6.1 Use $Sa_{g.m.}$ as the intensity measure

In this case, ground motion hazard analysis is performed for $Sa_{g.m.}$, as is standard practice today. The intensity measure used for structural response is also $Sa_{g.m.}$, computed for the two components of ground motion used in the analysis. This method is in use today (e.g., Stewart et al. 2001) and appears to be the most straightforward for 3D structures. For this reason, it is currently recommended by the authors when a scalar intensity measure is used (see below). The preferred choice of an IM in the case where the two axes of the structure have different fundamental periods has not been extensively examined. In the absence of further research, one obvious possibility is to use $Sa_{g.m.}$ at an intermediate period (e.g., the geometric mean of the two periods).

7.6.2 Use Sa_{arb} as the intensity measure

When using Sa_{arb} as an intensity measure in this case, it is necessary to specify the component of the ground motion being measured, as there are now two horizontal components used in the analysis, each with a differing value of Sa. If the objective is only a scalar drift hazard curve (e.g., there is only a single response parameter of interest), then the practitioner may obtain the most efficient estimate by performing the regression analysis shown in Figure 7.3 one time for each candidate IM (e.g., for Sa_{arb} oriented along the "X-X" axis of the structure, for Sa_{arb} oriented along the "Y-Y" axis of the structure, and for $Sa_{g.m.}$). Because all three IM choices should lead to the same answer (in the limit with a very large sample of dynamic analyses), the engineer is free to choose the IM which results in the most efficient estimation; that is, the IM which results in the smallest standard deviation of response prediction. In some cases, the optimal IM will be apparent a priori: if the response parameter of interest is a drift in the X-X axis, then it is likely that the optimal IM is Sa_{arb} oriented along the X-X axis. In other cases, for example when assessing the axial force in a corner-column of a structure or when there is significant torsion in the structure, the optimal IM may be less obvious.

In some cases it is necessary to select a common IM for estimation of more than one response parameter simultaneously. For example, in loss estimation procedures associated with performance-based engineering it is often desirable to know the probability distribution of a set of story drifts and floor accelerations *simultaneously*. In this case, it may again be useful to consider several candidate IMs and examine the tradeoffs in efficiency. For instance, Sa_{arb} oriented along the X-X axis is likely to estimate the responses along the X-X axis efficiently but the responses along the Y-Y axis less efficiently and vice versa for Sa_{arb} along the Y-Y axis. Choosing an IM in this case will depend upon the relative importance of the various response parameters of interest, and an understanding of which IM is most efficient for predicting the important response parameters. Note that the results for the less-important response parameters will not be incorrect, but only estimated with less statistical precision.

The above procedure appears to be valid in the case where no record-scaling is used as part of the response predictions. If the records are scaled before performing the structural analysis, the scaling procedure must be carefully considered. Previous studies of the implications of scaling a single component of ground motion may not be applicable to scaling of two components. The authors are particularly concerned about the choice of a scale factor for the orthogonal component of a ground motion when a selected component has been scaled by a specified factor. This problem is currently under investigation.

7.6.3 Use a vector intensity measure representing the two components individually

This approach uses a two-parameter intensity measure, consisting of the spectral accelerations in both the X-X and Y-Y directions. Vector-valued ground motion hazard analysis (Bazzurro and Cornell 2002) is used to compute the joint hazard for the spectral acceleration values of the two components of ground motion. Response prediction can then be an explicit function of the two components independently. This approach should reduce the dispersion in structural response, and may be useful in some situations (e.g., the two axes of the structure have differing periods, or Sa_{arb} along a given axis is not effective at estimating responses along the opposite axis). However, this method is not ready for widespread adoption until use of vector ground motion hazard analysis becomes more common.

7.7 Application to current practice

The approaches described above rationally combine the uncertainty in both ground motion hazard and structural response, but they differ from current US building-code-based design practice (i.e., ASCE 2002). When dynamic time history analyses are utilized in practice today, a suite of ground motions (typically three or seven) is scaled to a target response spectrum (obtained using the deterministic or probabilistic hazard analysis methods described above). These motions are then used to analyze a structure, and evaluation is based on either the maximum structural response (if fewer than seven motions are used) or the average response (if at least seven motions are used). But again, the definition of spectral acceleration is not stated explicitly. An inconsistent basis (e.g., $Sa_{g.m.}$ spectra and Sa_{arb} scaling) is to be discouraged. As shown above, the use of $Sa_{g.m.}$ for two-dimensional analysis will result in lower target spectra than when Sa_{arb} is used, but higher variation among the structural response. If less than seven records are used, then the higher variation of structural response values will be implicitly (but not accurately) captured by the current rule because the maximum response value is likely to be larger. If, however, seven records are used and the average structural response is taken, then there is no penalty paid for the higher variation of structural response that results from using $Sa_{g.m.}$ rather than Sa_{arb}. Therefore, consistent use of $Sa_{g.m.}$ would be somewhat unconservatively biased with seven or more records under the current rule. The ideal solution to this inconsistency would be to incorporate the structural response uncertainty explicitly (as is done explicitly in, e.g., FEMA 2000a, b, c and Banon et al. 2001). Short of this, the code should require that target response spectrum be based on Sa_{arb} when at least seven records are used for two-dimensional analysis, so that the analysis will include the extra variability in the ground motion intensity. In this case structural engineers should explicitly request that the hazard analysis and the scaled ground motion records be based on the Sa_{arb} definition of Sa. For three-dimensional analysis, $Sa_{g.m.}$ is probably the natural choice for the reasons outlined in the previous section.

7.8 Conclusions

Although intensity-measure-based analysis procedures have proven to be useful methods for linking the analyses of earth scientists and structural engineers, care is needed to make sure that the link does not introduce errors into the analysis. Two definitions of "spectral acceleration" are commonly used by analysts, and the distinction between the definitions is not always made clear. Because of this, a systematic error has been introduced into the results from many risk analyses, typically resulting in unconservative conclusions. For an example site and structure located in Los Angeles, the error resulted in a 12% underestimation of the spectral acceleration value exceeded with a 2% probability in 50 years, and a 10% underestimation of the structure's maximum interstory drift ratio exceeded with a 2% probability in 50 years.

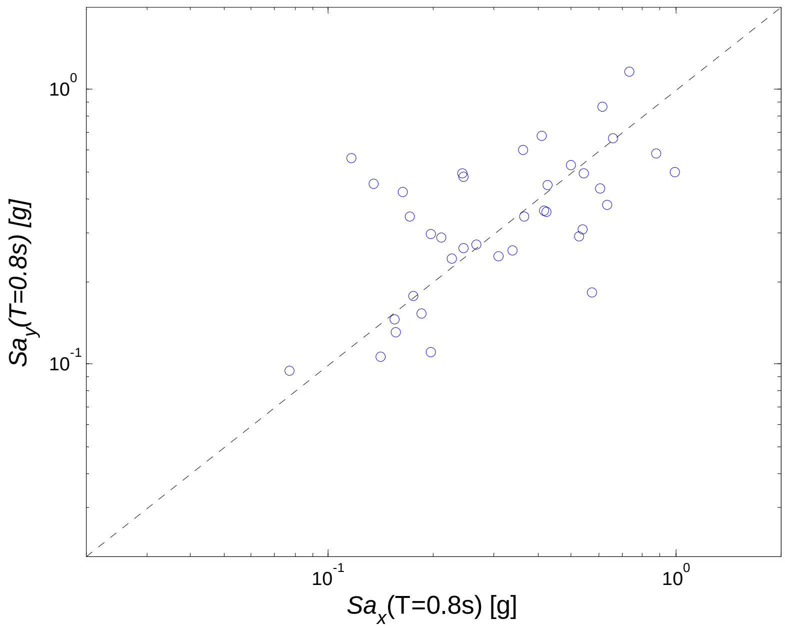
This problem is, however, merely one of communication, and not a fundamental flaw with the intensity measure approach. It is not difficult to use intensity measures in ways that produce correct results. For analysis of a single frame of a structure, the authors see three paths to the correct answer: (1) use Sa_{arb} for both parts of the analysis, (2) use $Sa_{g.m.}$ for both parts of the analysis, (3) perform hazard analysis with $Sa_{g.m.}$, structural response analysis with Sa_{arb}, but inflate the dispersion in the structural response prediction to represent the dispersion that would have been seen if $Sa_{g.m.}$ had been used. If a three-dimensional model of a structure is to be analyzed, the most straightforward method is to use $Sa_{g.m.}$ as the intensity measure for both the ground motion hazard and the structural response. In the absence of a single standard procedure, both earth scientists and structural analysts are encouraged to explicitly state which Sa definition they are using for evaluation, in the interest of transparency.

The methods described above will all produce valid estimates of the annual frequency of exceeding a given structural response level. In the future it would be desirable to have attenuation models which estimate the dispersion of both $Sa_{g.m.}$ and Sa_{arb}, in order to allow flexibility in the definition of the spectral acceleration used for analysis. Finally, vector-based methods of hazard and response analysis should improve upon the current situation in the future.

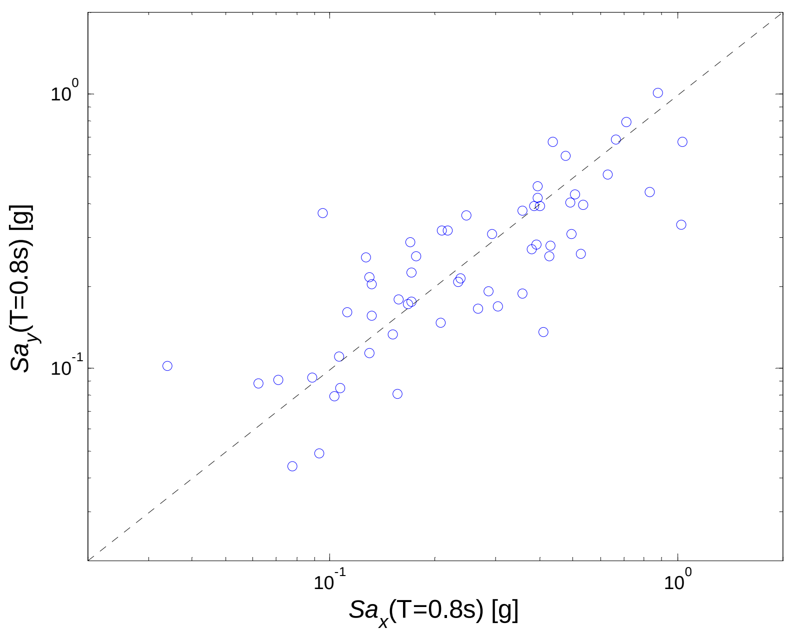
7.9 Appendix: slopes and standard deviations of regression predictions

This appendix explores in more detail the prediction of structural response as a function of either spectral acceleration of an arbitrary component or an average component. Consider a set of earthquake ground motions consisting of two components. We will refer to these two components as the "X" and "Y" components for clarity (the common assumption of no preferential orientation of motion is made here, which is typically valid when near-fault directivity effects are not present). Now consider the probabilistic distribution of the spectral acceleration values of these ground motion components. Logarithms of spectral acceleration and structural response are used, to take advantage of the linear relationships in the logarithmic domain often observed between these variables (e.g., Figure 7.3 and Figure 7.5).

The ln*Sa* values of the X and Y components have means, denoted $\mu_{\ln Sa_x}$ and $\mu_{\ln Sa_y}$, and standard deviations, denoted $\sigma_{\ln Sa_x}$ and $\sigma_{\ln Sa_y}$. Because there is no preferential direction to these motions, $\mu_{\ln Sa_x} = \mu_{\ln Sa_y}$ and $\sigma_{\ln Sa_x} = \sigma_{\ln Sa_y}$ (although our *estimates* of these values for a particular dataset might not be exactly equal, the underlying *true values* are assumed to be). We can compute a correlation coefficient between the two components, denoted $\rho_{\ln Sa_x, \ln Sa_y}$. The dependence between $\ln Sa_x$ and $\ln Sa_y$ is purely linear due to the lack of preferential orientation, as can be seen for example in Figure 7.5. We make the further mild assumption that the conditional variances of $\ln Sa_y$ given $\ln Sa_x$ and $\ln Sa_x$ given $\ln Sa_y$ are constant.



(a)



(b)

Figure 7.5. Samples of ($\ln Sa_x$, $\ln Sa_y$) pairs from (a) a set of ground motions with magnitude ≈ 6.5 and distance ≈ 8km, and (b) a set of ground motions with a wider range of magnitudes and distances, used to perform the structural analyses displayed in Figure 7.3.

Consider now the analysis of a structural frame, oriented along the X axis of the ground motions. We perform nonlinear dynamic analysis with the X component of each ground motion to calculate a set of structural response values. The logarithmic response values have a mean, $\mu_{\ln\theta_x}$, and a standard deviation, $\sigma_{\ln\theta_x}$. There is a relationship between $\ln Sa_x$ (log spectral acceleration in the X direction) and $\ln\theta_x$ (log response of the structure, oriented in the X direction) which can be represented by a linear correlation, and measured with a correlation coefficient denoted $\rho_{\ln\theta_x,\ln Sa_x}$. This linear relationship is represented by, for instance, the regression line shown in Figure 7.3a. But sometimes the relationship between ground motion intensity and structural response is represented as a function of the geometric mean of the *Sa*'s of the two components, as in Figure 7.3b. We can represent this relationship with the correlation coefficient denoted $\rho_{\ln\theta_x,\ln Sa_{g.m.}}$. We make the assumption that the dependence between $\ln\theta_x$ and $\ln Sa$ is purely linear and the conditional variance of $\ln\theta_x$ given $\ln Sa$ is constant (for both $\ln Sa_x$ and $\ln Sa_{g.m.}$). We are interested in the relationship between the slopes and standard deviations of prediction errors resulting from prediction using these two predictors $\ln Sa_x$ and $\ln Sa_{g.m.}$.

First, recalling that $\ln Sa_{g.m.}$ is the average of $\ln Sa_x$ and $\ln Sa_y$, we note that the spectral acceleration of the Y component has no direct physical effect on the response of the frame oriented in the X direction (indeed, the Y component of the ground motion is not even used in analysis). By itself, the parameter $\ln Sa_y$ does have, however, an indirect ability to predict θ_x simply due to its correlation with the predictor $\ln Sa_x$. Therefore, while $\ln\theta_x$ and $\ln Sa_y$ are statistically correlated, it is true that *given* $\ln Sa_x$, $\ln Sa_y$ provides no *additional* information about the distribution of $\ln\theta_x$. That is, $\ln\theta_x$ and $\ln Sa_y$ are conditionally independent given $\ln Sa_x$.

Using the above information, we can calculate conditional means and variances, using best linear predictors. For example, the mean value of $\ln\theta_x$ given $\ln Sa_x$ is:

$$\begin{aligned}\mu_{\ln\theta_x|\ln Sa_x=x} &= \left(\mu_{\ln\theta_x} - \frac{\rho_{\ln\theta_x,\ln Sa_x}\sigma_{\ln\theta_x}\mu_{\ln Sa_x}}{\sigma_{\ln Sa_x}}\right) + \left(\frac{\rho_{\ln\theta_x,\ln Sa_x}\sigma_{\ln\theta_x}}{\sigma_{\ln Sa_x}}\right)x \\ &\equiv a + bx\end{aligned} \tag{7.3}$$

It can be shown that *a* and *b* are the expected values of the coefficients estimated from linear least squares regression of $\ln\theta_x$ on $\ln Sa_x$. The conditional variance of $\ln\theta_x$ given $\ln Sa_x$ is:

$$\sigma^2_{\ln\theta_x|\ln Sa_x} = \sigma^2_{\ln\theta_x}\left(1-\rho^2_{\ln\theta_x,\ln Sa_x}\right) \tag{7.4}$$

We are interested in computing the conditional mean and variance of $\ln\theta_x$ given $\ln Sa_{g.m.}$ for comparison, but several intermediate results are needed first. The marginal mean and variance of $\ln Sa_{g.m.}$ are:

$$\begin{aligned}\mu_{\ln Sa_{g.m.}} &= E\left[1/2\left(\ln Sa_x + \ln Sa_y\right)\right] \\ &= \mu_{\ln Sa_x}\end{aligned} \tag{7.5}$$

$$\begin{aligned}\sigma^2_{\ln Sa_{g.m.}} &= Var\left[1/2\left(\ln Sa_x + \ln Sa_y\right)\right] \\ &= 1/4\left(\sigma^2_{\ln Sa_x} + \sigma^2_{\ln Sa_y} + \rho_{\ln Sa_x,\ln Sa_y}\sigma_{\ln Sa_x}\sigma_{\ln Sa_y}\right) \\ &= \sigma^2_{\ln Sa_x}\frac{1+\rho_{\ln Sa_x,\ln Sa_y}}{2}\end{aligned} \tag{7.6}$$

The conditional variance of $\ln\theta_x$ given $\ln Sa_y$ is:

$$\begin{aligned}\sigma^2_{\ln\theta_x|\ln Sa_y} &= \mathrm{E}\left[\mathrm{Var}\left[\ln\theta_x \mid \ln Sa_x, \ln Sa_y\right] \mid \ln Sa_y\right] + \mathrm{Var}\left[\mathrm{E}\left[\ln\theta_x \mid \ln Sa_x, \ln Sa_y\right] \mid \ln Sa_y\right] \\ &= \mathrm{E}\left[\sigma^2_{\ln\theta_x}\left(1-\rho^2_{\ln\theta_x,\ln Sa_x}\right) \mid \ln Sa_y\right] \\ &\quad + \mathrm{Var}\left[\mu_{\ln\theta_x} + \rho_{\ln\theta_x,\ln Sa_x}\sigma_{\ln\theta_x}\left(\frac{\ln Sa_x - \mu_{\ln Sa_x}}{\sigma_{\ln Sa_x}}\right)\middle| \ln Sa_y\right] \\ &= \sigma^2_{\ln\theta_x}\left(1-\rho^2_{\ln\theta_x,\ln Sa_x}\right) + \rho^2_{\ln\theta_x,\ln Sa_x}\frac{\sigma^2_{\ln\theta_x}}{\sigma^2_{\ln Sa_x}}\mathrm{Var}\left[\ln Sa_x \mid \ln Sa_y\right] \\ &= \sigma^2_{\ln\theta_x}\left(1-\rho^2_{\ln\theta_x,\ln Sa_x}\rho^2_{\ln Sa_x,\ln Sa_y}\right)\end{aligned} \tag{7.7}$$

This implies that the correlation coefficient between $\ln\theta_x$ and $\ln Sa_y$ is $\rho_{\ln\theta_x,\ln Sa_x}\rho_{\ln Sa_x,\ln Sa_y}$, showing that it is weaker than the correlation between $\ln\theta_x$ and $\ln Sa_x$, and only as strong as the correlation $\ln Sa_y$ has with $\ln Sa_x$ permits. This simple product form is a result. Next we compute the covariance between $\ln\theta_x$ and $\ln Sa_{g.m.}$:

$$\begin{aligned}\mathrm{Cov}\left[\ln\theta_x, \ln Sa_{g.m.}\right] &= 1/2\,\mathrm{Cov}\left[\ln\theta_x, \ln Sa_x\right] + 1/2\,\mathrm{Cov}\left[\ln\theta_x, \ln Sa_y\right] \\ &= 1/2\left(\rho_{\ln\theta_x,\ln Sa_x}\sigma_{\ln\theta_x}\sigma_{\ln Sa_x} + \rho_{\ln\theta_x,\ln Sa_x}\rho_{\ln Sa_x,\ln Sa_y}\sigma_{\ln\theta_x}\sigma_{\ln Sa_y}\right) \\ &= 1/2\,\rho_{\ln\theta_x,\ln Sa_x}\sigma_{\ln\theta_x}\sigma_{\ln Sa_x}\left(1+\rho_{\ln Sa_x,\ln Sa_y}\right)\end{aligned} \tag{7.8}$$

Finally, using the results of Equations 7.6 and 7.8, we compute the correlation coefficient between $\ln\theta_x$ and $\ln Sa_{g.m.}$:

$$\begin{aligned}\rho_{\ln\theta_x,\ln Sa_{g.m.}} &= \frac{\mathrm{Cov}\left[\ln\theta_x \mid \ln Sa_{g.m.}\right]}{\sigma_{\ln\theta_x}\sigma_{\ln Sa_{g.m.}}} \\ &= \rho_{\ln\theta_x,\ln Sa_x}\sqrt{\frac{1+\rho_{\ln Sa_x,\ln Sa_y}}{2}}\end{aligned} \tag{7.9}$$

We are now ready to find the conditional mean of $\ln\theta_x$ given $\ln Sa_{g.m.}$:

$$\mu_{\ln\theta_x|\ln Sa_{g.m.}=z} = \mu_{\ln\theta_x} + \rho_{\ln\theta_x,\ln Sa_{g.m.}}\sigma_{\ln\theta x}\left(\frac{z-\mu_{\ln Sa_{g.m.}}}{\sigma_{\ln Sa_{g.m.}}}\right) \tag{7.10}$$

Substituting from Equations 7.5, 7.6 and 7.9 gives:

$$\mu_{\ln\theta_x|\ln Sa_{g.m.}=z} = \left(\mu_{\ln\theta_x} - \frac{\rho_{\ln\theta_x,\ln Sa_x}\sigma_{\ln\theta_x}\mu_{\ln Sa_x}}{\sigma_{\ln Sa_x}}\right) + \left(\frac{\rho_{\ln\theta_x,\ln Sa_x}\sigma_{\ln\theta_x}}{\sigma_{\ln Sa_x}}\right)z \quad (7.11)$$
$$= a + bz$$

where a and b are the same as in Equation 7.3. Therefore*, the expected slope and intercept of a regression analysis will be the same regardless of whether* ln$Sa_{g.m.}$ *or* ln$Sa_{x.}$ *is used to predict* lnθ_x. Now consider the conditional variance of lnθ_x given ln$Sa_{g.m.}$:

$$\begin{aligned}\sigma^2_{\ln\theta_x|\ln Sa_{g.m.}} &= \sigma^2_{\ln\theta_x}\left(1-\rho^2_{\ln\theta_x,\ln Sa_{g.m.}}\right)\\ &= \sigma^2_{\ln\theta_x}\left(1-\rho^2_{\ln\theta_x,\ln Sa_x}\frac{1+\rho_{\ln Sa_x,\ln Sa_y}}{2}\right)\end{aligned} \quad (7.12)$$

Therefore, the conditional standard deviation of of lnθ_x given ln$Sa_{g.m.}$ is greater than the conditional standard deviation given *ln*Sa_x by a factor equal to:

$$\frac{\sigma_{\ln\theta_x|\ln Sa_{g.m.}}}{\sigma_{\ln\theta_x|\ln Sa_x}} = \sqrt{1+\frac{1-\rho_{\ln Sa_x,\ln Sa_y}}{2}\left(\frac{\rho^2_{\ln\theta_x,\ln Sa_x}}{1-\rho^2_{\ln\theta_x,\ln Sa_x}}\right)} \quad (7.13)$$

Noting that correlation coefficients always lie on the interval [-1, 1], we see that the ratio in Equation 7.13 is always greater than or equal to 1, and increases with decreasing $\rho_{\ln Sa_x,\ln Sa_y}$ or increasing $\rho_{\ln\theta_x,\ln Sa_x}$. The term $\rho_{\ln Sa_x,\ln Sa_y}$ is dependent on the record set in use, but typically falls between 0.8 and 0.9, depending on the range of magnitudes and distances of the records (letting the magnitude and distance vary in the recordset increases the correlation between components relative to a recordset selected from a narrow range of magnitude and distance values, as is seen in Figure 7.5).

Note that this result only holds under the conditions described above. The most critical assumption is that response in the X direction is unaffected by ground motion input in the Y direction. This may not be the case in, for example, torsionally coupled structures. In addition, the assumption of linear dependence between lnθ_x and lnSa may not always hold. In these cases, the simple inflation factor is not applicable.

Chapter 8

Correlation of Response Spectral Values for Multi-Component Ground Motions

Baker, J.W. and Cornell, C.A., (2005) "Correlation of Response Spectral Values for Multi-Component Ground Motions." *Bulletin of the Seismological Society of America* (In press).

8.1 Abstract

Ground motion prediction (attenuation) models predict the probability distributions of spectral acceleration values for a specified earthquake event. These models provide only marginal distributions, however; they do not specify correlations among spectral accelerations with differing periods or orientations. In this paper a large number of strong ground motions are used to empirically estimate these correlations, and nonlinear regression is used to develop approximate analytical equations for their evaluation. Because the correlations apply to residuals from a ground motion prediction, they are in principle dependent on the ground motion prediction model used. The observed correlations do not vary significantly when the underlying model is changed, however, suggesting that the predictions are applicable regardless of the model chosen by the analyst. The analytical correlation predictions improve upon previous predictions of correlations at differing periods in a randomly oriented horizontal ground motion component. For correlations within a vertical ground motion or across orthogonal components of a ground motion, these results are believed to be the first of their kind.

The resulting correlation coefficient predictions are useful for a range of problems related to seismic hazard and the response of structures. Past uses of previous correlation predictions are described, and future applications of the new predictions are proposed. These applications will

allow analysts to better understand the properties of single- and multi-component earthquake ground motions.

8.2 Introduction

Spectral acceleration (*Sa*) values of earthquake ground motions are widely used in seismic hazard analysis and evaluation of structural response. Relatively little work has been done, however, to measure the joint distributions of multiple simultaneous spectral acceleration values. In this paper, the authors measure correlation coefficients of spectral acceleration values to gain insight into these joint distributions and to facilitate other research. Correlations are presented for spectral acceleration values of a single ground motion component at two differing periods, and also for spectral accelerations of orthogonal components (horizontal/horizontal or horizontal/vertical) at two periods.

This analysis was performed in recognition of the many potential applications of the results. Several predictions have been previously developed for correlations of spectral acceleration values of a single ground motion component, and past uses of those models are mentioned below. The models for spectral accelerations of orthogonal components are new, and so several potential applications are described.

8.3 Motivation

Knowledge of correlation of spectral values has been used in several past studies to gain insight into seismic hazard and structural performance. Several past uses of the type of models presented here are discussed in this section, and potential future applications will be discussed later.

Conventional probabilistic seismic hazard analysis (PSHA) (Kramer 1996) provides the mean annual rate of exceeding a specified value of a single ground motion parameter, such as spectral acceleration at a given period. These hazard analyses can be repeated for spectral acceleration at several periods and presented simultaneously as uniform hazard spectra. But these uniform hazard spectra, being the locus of results from a suite of marginal hazard analyses for individual spectral values, should not be interpreted as providing any knowledge about the joint occurrence of spectral values at differing periods. In order to obtain knowledge about the joint or simultaneous occurrence of spectral acceleration at multiple periods, it is necessary to perform a vector-valued probabilistic seismic hazard analysis (VPSHA) (Bazzurro and Cornell 2002). This analysis is a direct extension of traditional PSHA, using the same information about the magnitudes, locations, and recurrence rates of earthquakes and the same ground motion prediction (attenuation) models. The only additional information requirement is knowledge of the joint distribution of the spectral values for a given magnitude and distance. Logarithmic spectral acceleration values have been observed to be well-represented by the normal distribution marginally, so the mild assumption that pairs of values are well-represented by the joint normal distribution (and/or that the conditional distributions of one given the other are normal) is probably a reasonable one, but has not been investigated as yet to the authors' knowledge. Under this assumption, only correlation coefficients between spectral values at two periods are needed to define the joint distribution and proceed with VPSHA. The models presented in this paper will provide improved predictions of these correlation coefficients, furthering the development of the vector-valued probabilistic seismic hazard analysis. Once a vector-valued PSHA has been performed, structural engineers can use this new information to improve the efficiency of probabilistic performance assessments of structures (e.g., Baker and Cornell 2004, 2005a). Because engineers are provided with more information about the spectral content of the ground motions occurring at a site, they are able to increase the precision of their structural response analyses.

Correlation of spectral acceleration values also arises implicitly in the development of ground motion prediction models. The seismologists who develop these models often average the (log) spectral acceleration values of two perpendicular horizontal components of a ground motion and use this averaged data for fitting regression lines. This averaging decreases the noise in the data, allowing for more accurate estimates of the model parameters. Thus, PSHA calculations using these ground motion prediction models provide mean exceedance rates for averaged spectral acceleration values, or more precisely for the geometric mean of the two horizontal components. But structural engineers often do not perform this averaging across components, resulting in an inconsistency between PSHA and structural analysis. The work of the seismologist and engineer can be properly re-connected, however, once knowledge of correlations of spectral acceleration values in perpendicular ground motion components is known (Baker and Cornell 2005b).

In addition to averaging across perpendicular components, averaging spectral accelerations across a range of periods is sometimes also performed, with a similar goal of reducing the variability of the resulting ground motion intensity parameter for a given magnitude and distance (e.g., Pacific Gas & Electric 1988, Shome and Cornell 1999, Abrahamson et al. 2003). With this work, knowledge of correlations of spectral accelerations is again needed, in order to quantify the effect of the averaging procedure.

A measure of ground motion intensity proposed by Cordova et al. (2001) is a function of spectral acceleration at two periods. This measure was found to be useful for predicting response of the structure under consideration. A custom ground motion prediction model was needed to complete the assessment of the structure, and was derived from an existing model by making use of an estimate of correlation between spectral acceleration values at the two periods of interest.

In addition to these past applications, there are several easily envisioned future applications that make use of the new information in this paper regarding correlations across multiple components of ground motion. Examples are presented below, after the development of the predictive equations.

It should be noted that there are other studies of earthquake ground motions that at first glance might appear similar to this work, but are in fact not related. For example, Penzien and Watabe (1975) observed that temporal cross-correlations of ground motion accelerations at an instant in time are approximately zero along specified principle axes. That work does not imply anything about the phenomenon examined in this paper: the correlation of peak spectral values (i.e., frequency content measures) at differing frequencies and orientations.

8.4 Analysis procedure

8.4.1 Record selection

The results presented in this study were derived empirically from a strong motion data set based on worldwide recordings of shallow crustal earthquakes. The records for this study were taken from the PEER Strong Motion Database (2000). Records were selected based on the following criteria:

1. The site was classified as stiff soil: USGS class B-C or Geomatrix class B-D.
2. The recording was made in the free field or the first story of a structure.
3. All three components (two horizontal and one vertical) were available and had high-pass filter corner frequencies less than 0.2 hertz and low-pass filter corner frequencies greater than 18 hertz.
4. The earthquake magnitude was greater than 5.5.
5. The source-to-site distance was less than 100 km.

The recordings were left oriented as recorded rather than rotated into fault normal and fault parallel components, so they have effectively random orientations with respect to fault direction. This is analogous to the record orientations used to develop typical ground motion prediction models. The Next Generation Attenuation project will produce a record library and predictive models that treat fault normal and fault parallel ground motions separately. When that project is completed it will be useful to compute correlations for fault normal and fault parallel ground motions separately, but at present only randomly oriented ground motions are considered.

A total of 469 records from 31 earthquakes met the selection criteria, each consisting of three components of ground motion recordings. Of the 469 records, 202 were from the 1999 Chi-Chi, Taiwan earthquake. The Chi-Chi records were removed from the initial analysis to ensure that the results would not be excessively influenced by any peculiarities of the records from this single earthquake. The Chi-Chi records were later used to cross-validate the predictive equations.

8.4.2 Computation of correlations

Using the 267 remaining three-component records, correlations were computed for the two horizontal and vertical components. At this point, the variable for which the correlation is estimated should be defined more clearly. The logarithmic spectral accelerations of the three ground motion components can be represented by the following model:

$$\ln Sa_x(T) = f_H(M,R,T,\theta) + \sigma_H(M,T)\varepsilon_x(T) \tag{8.1}$$

$$\ln Sa_y(T) = f_H(M,R,T,\theta) + \sigma_H(M,T)\varepsilon_y(T) \tag{8.2}$$

$$\ln Sa_z(T) = f_V(M,R,T,\theta) + \sigma_V(M,T)\varepsilon_z(T) \tag{8.3}$$

where x and y are used to denote the two horizontal directions of the recording, and z is used to denote the vertical direction. The functions $f_H(M,R,T,\theta)$ and $f_V(M,R,T,\theta)$ are mean ground motion predictions for the horizontal and vertical logarithmic response spectral values, respectively. These predictions are a function of the earthquake magnitude (M), distance (R), period (T) and other parameters (θ) such as the local soil conditions and faulting mechanism. These mean ground motion predictions are deterministic, given the input parameters. The terms $\sigma_H(M,T)$ and $\sigma_V(M,T)$ account for the observed standard deviation of the logarithmic horizontal and vertical spectral accelerations, respectively. The standard deviations are observed to be dependent on the magnitude of the earthquake and the period of interest. Finally, the random variables $\varepsilon_x(T)$, $\varepsilon_y(T)$, and $\varepsilon_z(T)$ account for the randomness of the observations. Because the other terms in Equations 8.1 through 8.3 have already accounted for the means and standard deviations of logarithmic spectral acceleration, the ε terms have means of zero and unit standard deviations.

The $f(M,R,T,\theta)$ and $\sigma(M,T)$ functions are completely defined by previously published ground motion prediction models. What is not defined by standard ground motion prediction models is the correlation between ε terms at different frequencies or for different components. For example, the correlation between the ε values of the two horizontal components of a given record is of interest: $\rho_{\varepsilon_x(T),\varepsilon_y(T)}$. The ε values of a record also vary as the period varies, and so the correlation of the ε values of a single component of a record at two periods is also of interest:

$\rho_{\varepsilon_x(T_1),\varepsilon_x(T_2)}$. Finally, one might be interested in the correlations between two components at two differing periods: $\rho_{\varepsilon_x(T_1),\varepsilon_y(T_2)}$. These correlation values are estimated in this paper.

It can be seen from Equations 8.1 through 8.3 that the computed ε values will vary somewhat for a given record depending on the ground motion prediction model chosen. That is, ln*Sa*(*T*) of a record is given and $f(M,R,T,\theta)$ and $\sigma(M,T)$ vary slightly among models, and so the $\varepsilon(T)$ value of a record must also vary among models to maintain equality. The correlations of ε values, however, were observed to be insensitive to the ground motion prediction model considered. Correlations were computed here using the model of Abrahamson and Silva (1997), but the results were found to be nearly identical when other models (specifically, Boore et al. 1997 and Campbell 1997) were compared.

Once correlations of the ε values have been determined, we note that $\ln Sa(T)$ is simply a linear function of $\varepsilon(T)$, with no other sources of uncertainty. Therefore, the correlation between, for example, $\ln Sa_x(T)$ and $\ln Sa_y(T)$ (for a given record) is equal to the correlation between $\varepsilon_x(T)$ and $\varepsilon_y(T)$. Thus the procedure used here is to compute the ε values for all records, in order to remove the effect of magnitude, distance, etc., from the variation in observed spectral values. Correlations can be computed for these ε values, which are then appropriate to represent the correlations between ln*Sa* values for a given magnitude, distance, etc. For the same reason, the predictions can be used to represent logarithmic spectral velocity or spectral displacement. The correlation of $\ln Sa_x(T)$ and $\ln Sa_y(T)$ is also generally a reasonable approximation for the correlation between $Sa_x(T)$ and $Sa_y(T)$ (Liu and Der Kiureghian 1986). In applications such as vector-valued PSHA, however, it is often the logarithms of response spectral values that are used in the joint distributions, so the more precise correlation between $\ln Sa_x(T)$ and $\ln Sa_y(T)$ is sufficient in many cases.

To estimate correlation coefficients, we use the maximum likelihood estimator, sometimes referred to as the Pearson product-moment correlation coefficient (Neter et al. 1996):

$$\hat{\rho}_{A,B} = \frac{\sum_{i=1}^{n}\left(A_i - \overline{A}\right)\left(B_i - \overline{B}\right)}{\sqrt{\sum_{i=1}^{n}\left(A_i - \overline{A}\right)^2 \sum_{i=1}^{n}\left(B_i - \overline{B}\right)^2}} \tag{8.4}$$

where *A* and *B* are the random variables of interest (e.g., $\varepsilon_x(T)$ and $\varepsilon_y(T)$), $\overline{A}$ and $\overline{B}$ are their sample means, A_i is the *i*th observation of variable *A*, and *n* is the total number of observations (records). We perform this correlation computation for each pair of orientations (horizontal/horizontal in the same direction, horizontal/horizontal in perpendicular directions, vertical/vertical, and vertical/horizontal) and for each pair of periods of interest (75 periods between 0.05 and 5 seconds). The matrix representing correlations for all combinations of the 75 periods is most compactly displayed using a contour plot as a function of the periods T_1 and T_2 (e.g., Figure 8.1).

Before performing further analysis, the correlation coefficients estimated using Equation 8.4 were smoothed using a simple averaging with correlation coefficients in a nearby neighborhood of periods, in order to remove some of the noise in the estimates and make the underlying patterns in the correlation matrix clearer. A comparison of contours of the correlation matrix before and after smoothing is displayed in Figure 8.1.

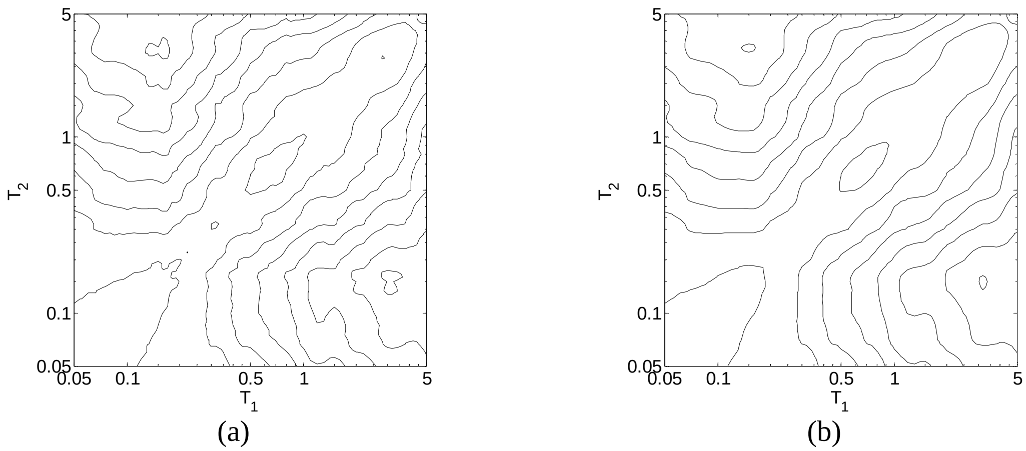


Figure 8.1. Effect of smoothing on the empirical correlation matrix for horizontal epsilons in perpendicular directions at two periods (T1 and T2). (a) Before smoothing. (b) After smoothing.

8.4.3 Nonlinear regression

Nonlinear least squares regression was utilized to condense the data from a large empirical correlation matrix into a relatively simple predictive equation. Functional forms were chosen based on inspection of the correlation matrices, and coefficients for the functions were determined using nonlinear regression. Correlation coefficients estimated from empirical data have non-constant standard errors that are dependent on the true underlying correlation coefficient. For this reason, minimizing the squared error between the empirical correlation matrix and the predictive function would not be the optimal criteria for fitting the predictive function (i.e., fitting a correlation coefficient of 0.9 with an estimate of 0.8 is a worse error than fitting a correlation coefficient of 0.1 with an estimate of 0). For this reason the Fisher z transformation (Neter et al. 1996) was applied to the correlation coefficients:

$$z = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right) \tag{8.5}$$

where ρ is an estimated correlation coefficient and z the transformed data with a constant standard error. Simple least squares regression could then be applied to these z values. The coefficients for the prediction equations were selected such that the squared prediction errors were minimized over the range of periods of interest:

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{1}{2}\ln\left(\frac{1+\rho_{i,j}}{1-\rho_{i,j}}\right) - \frac{1}{2}\ln\left(\frac{1+\hat{\rho}_{i,j}(\boldsymbol{\theta})}{1-\hat{\rho}_{i,j}(\boldsymbol{\theta})}\right)\right)^2 \tag{8.6}$$

where $\rho_{i,j}$ is the empirical correlation coefficient at the period pair (T_i, T_j) and $\hat{\rho}_{i,j}(\boldsymbol{\theta})$ is its predicted value using the functional forms shown below with a vector of coefficients $\boldsymbol{\theta}$. The resulting models are strictly empirical and thus should not be extrapolated beyond the range over which they were fit (periods between 0.05 and 5 seconds, earthquake magnitudes between 5.5 and 7.6, and distances between 0 and 100 km).

8.5 Results

The results of the above analyses are presented in the form of simple equations. They are broken into three separate cases, presented individually below.

8.5.1 Cases at a single period

Correlations between response spectral values with the same period but differing orientations are presented first. The correlation between horizontal orthogonal ε values at the period T is estimated by the equation:

$$\rho_{\varepsilon_x,\varepsilon_y} = 0.79 - 0.023 \cdot \ln(T) \tag{8.7}$$

The predictions from this function and the empirical correlations from the data set are displayed in Figure 8.2. By using the bootstrap to resample ground motion records (Efron and Tibshirani 1993), the slope of the regression line was found to be statistically significant, with a p-value of 0.001.

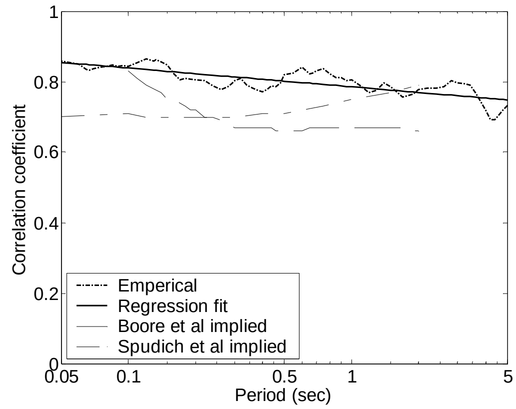


Figure 8.2. Correlation coefficients for perpendicular horizontal epsilons at the same period. Empirical results, the prediction from Equation 8.7, and the correlations implied from the ratios of standard deviations in Boore et al. (1997) and Spudich et al. (1999).

The correlation between a horizontal ε value and a vertical ε value at the period T is estimated by the constant:

$$\rho_{\varepsilon_x, \varepsilon_z} = 0.63 \tag{8.8}$$

A regression fit was performed with a prediction as a function of T, but the slope was not statistically significant and so the prediction is a constant value for all periods. The empirical correlations from the data set and the predicted value are displayed in Figure 8.3.

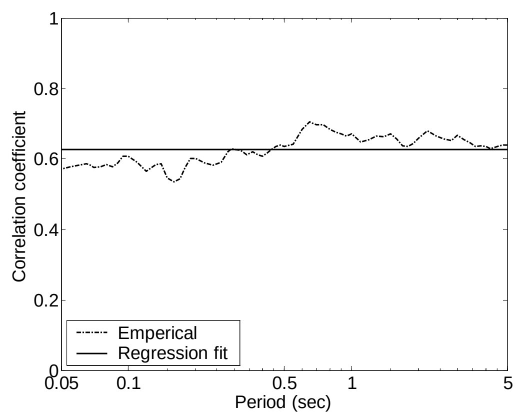


Figure 8.3. Correlation coefficients between horizontal epsilons and vertical epsilons at the same period. Empirical results and the prediction from Equation 8.8.

8.5.2 Cases with differing periods but the same orientation

When the two periods of interest differ, more complex functional forms are needed. The correlation between the ε values of a single horizontal ground motion component at two differing periods is estimated by the function:

$$\rho_{\varepsilon_x,\varepsilon_x} = 1 - \cos\left(\frac{\pi}{2} - \left(0.359 - 0.163 I_{(T_{\min<0.189})} \ln\frac{T_{\min}}{0.189}\right)\ln\frac{T_{\max}}{T_{\min}}\right) \tag{8.9}$$

where $I_{(T_{\min<0.189})}$ is an indicator function equal to 1 if $T_{\min} < 0.189$ second and equal to 0 otherwise, implying that the form of the equation is simply $1-\cos\left(a - b\ln\left(T_{\max}/T_{\min}\right)\right)$ for periods larger than 0.189 second. The variables $T_{\min}$ and $T_{\max}$ are used to denote to the smaller and larger of the two periods of interest, respectively. The empirical correlations from the data set and the predictions from Equation 8.9 are displayed in Figure 8.4a and Figure 8.4b, respectively. Note that both the empirical data and this prediction imply that correlation does not always decrease with increasing separation of periods (e.g., the correlation between *Sa* values at {$T_{\min}$=0.05s, $T_{\max}$=1s} is greater then the correlation at {$T_{\min}$=0.2s, $T_{\max}$=1s}). The authors see no obvious reason for this phenomenon, but it is seen clearly in the data.

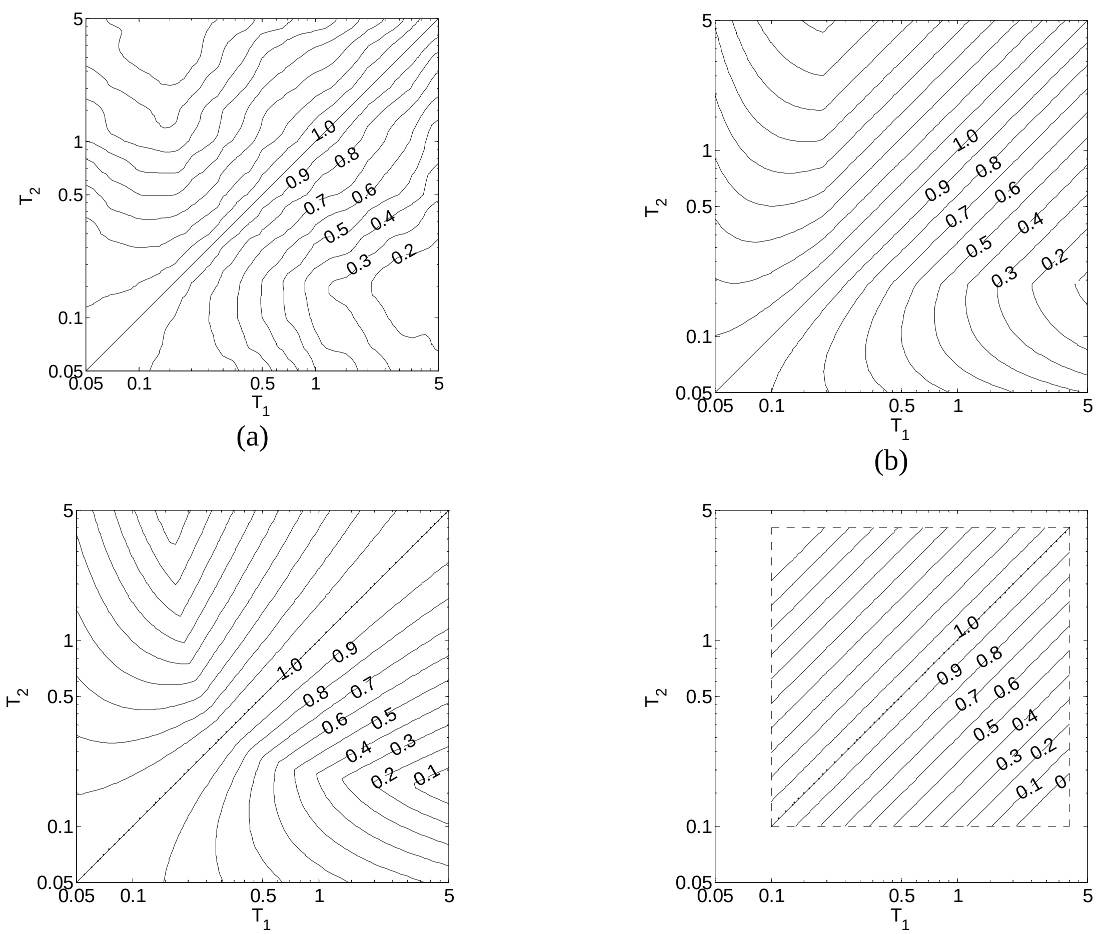


Figure 8.4. Correlation contours for horizontal epsilons in the same direction at two periods (T1 and T2). (a) Smoothed empirical results. (b) The prediction from Equation 8.9. (c) The prediction from Abrahamson et al. (2003). (d) The prediction from Inoue and Cornell (1990).

Although Equation 8.9 was fit for ε values of individual components, it is equally valid for ε values of geometric mean spectral acceleration values. This is shown both theoretically and empirically in Appendix B. The equation could also be used to approximately represent the correlation of inter-event ε values (which are of interest for modeling losses to portfolios of spatially distributed buildings) but the agreement is not as good in this situation. The definition of inter-event ε values can be found in Abrahamson and Silva (1997).

The correlation between the ε values of a vertical ground motion component at two different periods is estimated by the function:

$$\rho_{\varepsilon_z,\varepsilon_z} = 1 - 0.77\ln\frac{T_{\max}}{T_{\min}} + 0.315\left(\ln\frac{T_{\max}}{T_{\min}}\right)^{1.4} \tag{8.10}$$

A comparison of this prediction with the empirical correlation coefficients is shown in Figure 8.5.

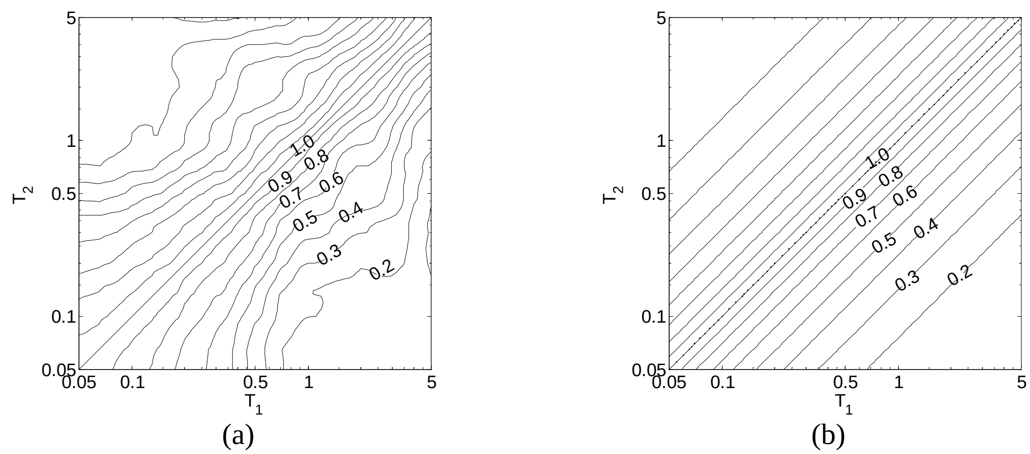


Figure 8.5. Correlation contours for vertical epsilons in the same direction at two periods (T1 and T_2). (a) Smoothed empirical results. (b) The prediction from Equation 8.10.

8.5.3 Cases with differing periods and differing orientations

For correlations of ε values between two horizontal components in perpendicular directions, it was hypothesized that perhaps the correlation coefficient could be represented as a product of the correlation due to perpendicular orientation and the correlation due to differing periods. The model of Equation 8.7 was used to represent the perpendicular orientations, evaluated at $T = \sqrt{T_{\min}T_{\max}}$, the geometric mean of the two periods of interest, and the model of Equation 8.9 was used to represent the correlations at differing periods. The resulting product-form equation is:

$$\begin{aligned}\rho_{\varepsilon_x,\varepsilon_y} = &\left(0.79 - 0.023\cdot\ln\sqrt{T_{\min}T_{\max}}\right)\\ &\cdot\left(1-\cos\left(\frac{\pi}{2}-\left(0.359-0.163 I_{(T_{\min}<0.189)}\ln\frac{T_{\min}}{0.189}\right)\ln\frac{T_{\max}}{T_{\min}}\right)\right)\end{aligned} \tag{8.11}$$

A comparison of this prediction with the empirical correlation contours is shown in Figure 8.6. This form fits the empirical results well, and agrees with Equation 8.7 in the special case $T_{\min} = T_{\max}$. This "product of correlation coefficients" form implies that a Markov-process-like relationship exists, where $\varepsilon_x(T_1)$ and $\varepsilon_y(T_2)$ are conditionally (linearly) independent given either $\varepsilon_x(T_2)$ or $\varepsilon_y(T_1)$ (Ditlevsen 1981, p339). Approximate conditional independence is observed in both the empirical data and in Equation 8.11 (making the approximation $\rho_{\varepsilon_x(\sqrt{T_1T_2}),\varepsilon_y(\sqrt{T_1T_2})} \cong \rho_{\varepsilon_x(T_1),\varepsilon_y(T_1)} \cong \rho_{\varepsilon_x(T_2),\varepsilon_y(T_2)}$).

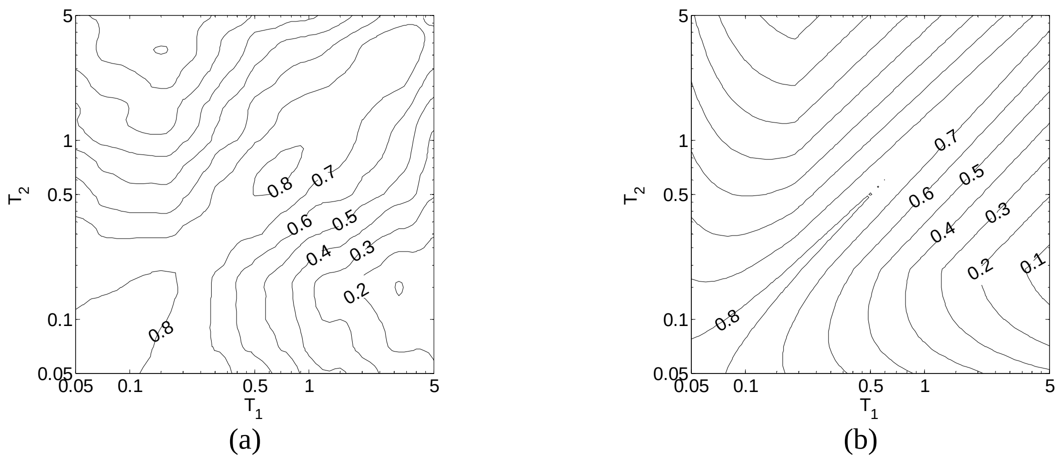


Figure 8.6. Correlation contours for horizontal epsilons in perpendicular directions at two periods (T_1 and T_2). (a) Smoothed empirical results. (b) The prediction from Equation 8.11.

The same procedure was used to predict the correlations of epsilons between a horizontal and a vertical component of a ground motion at differing periods. The estimate of Equation 8.8 at the geometric mean of the two periods was multiplied by a term with the functional form of Equation 8.9. In this case, the coefficients of Equation 8.9 did not provide a good fit to empirical results (which was expected because the coefficients are not associated with vertical motions), so two coefficients were re-estimated using nonlinear regression (the additional improvement from re-fitting all coefficients was negligible). The final estimate is given by:

$$\rho_{\varepsilon_x,\varepsilon_z} = \left(0.64 + 0.021 \cdot \ln\sqrt{T_{\min}T_{\max}}\right) \cdot \left(1 - \cos\left(\frac{\pi}{2} - \left(\ln\frac{T_{\max}}{T_{\min}}\right)\left(0.29 - 0.094 I_{(T_{\min}<0.189)} \ln\frac{T_{\min}}{0.189}\right)\right)\right) \tag{8.12}$$

This prediction is compared to empirical results in Figure 8.7. One notable feature of Figure 8.7a is that the contours are not symmetric about the line T_1=T_2. That is, the correlation between *Sa* of the horizontal component at T_1 with the *Sa* of the vertical component at T_2 is *not necessarily equal to* the correlation of the *Sa* of the horizontal component at T_2 with the *Sa* of the vertical component at T_1. When the empirical correlations from the primary dataset were compared with the empirical correlations from the Chi-Chi data set, however, variations between the two indicated that any asymmetries in the empirical correlations were likely due to sampling variability, rather than from some underlying trend. For this reason, the form of Equation 8.12 is left such that predictions are symmetric about the line T_1=T_2.

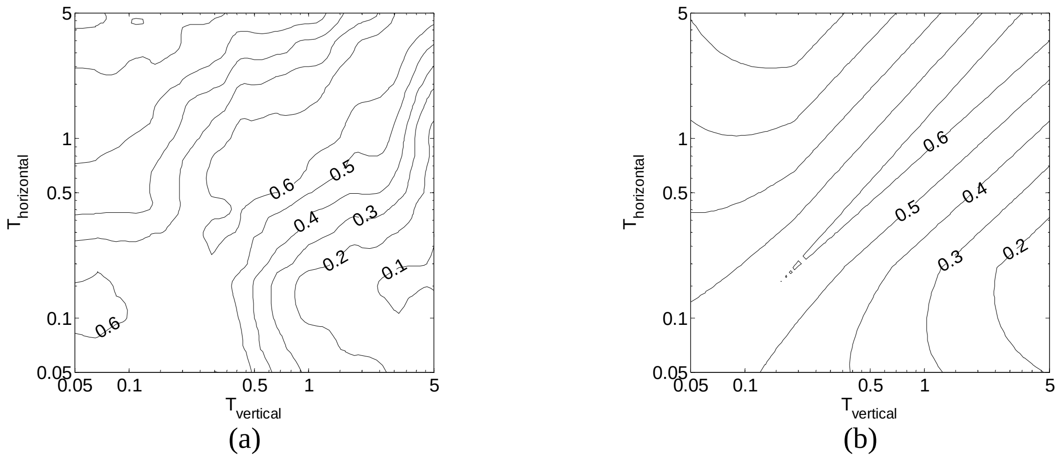


Figure 8.7. Correlation contours of vertical epsilons with horizontal epsilons at two periods (T_1 and T_2). (a) Smoothed empirical results. (b) The prediction from Equation 8.12.

Among the horizontal/vertical correlations, certain period pairs will be of more engineering interest than others. For instance, periods of vibration of buildings are typically much shorter in the vertical direction than the horizontal direction. Using a simple estimate of 0.1s for the vertical period of a typical building, it would be interesting to know the correlation of vertical *Sa*'s at 0.1s with horizontal *Sa*'s at a range of periods (corresponding to varying horizontal periods of vibration). This is shown in Figure 8.8, for both the empirical and predicted correlations.

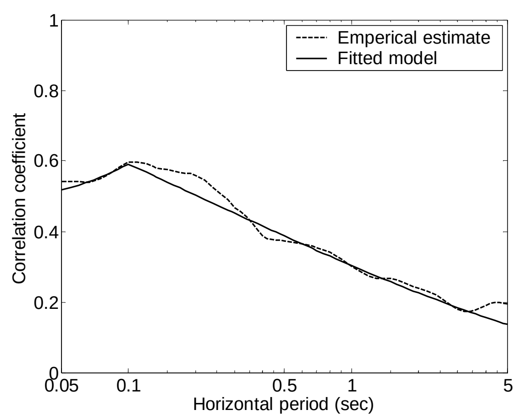


Figure 8.8. Correlation coefficient between vertical epsilons and horizontal epsilons when the period in the vertical direction is 0.1 seconds.

For all of the predictions above, several checks were performed. The positive definiteness of the predicted correlation matrices (when computed for a large array of periods simultaneously) was verified. A joint correlation matrix consisting of predictions in all three directions simultaneously was also found to be positive definite. This is a required property of a correlation matrix, and is necessary if one needs the joint distribution of *Sa* at many periods simultaneously (e.g., the "simulation of response spectra" application below). In addition, it was verified that the empirical correlations do not depend on magnitude or distance. This was done by taking windows of magnitude or distance values, and comparing the computed correlation coefficients as the window moved to different magnitude or distance values. No trends were seen, and so the above models were left functionally independent of magnitude and distance. Supporting calculations for these conclusions can be found in Appendix B.

Some general observations can be made from the above data and analytical predictions. When both periods are the same, the correlation between the *Sa*'s of two perpendicular horizontal components is roughly 0.8. For frame-type buildings, the first two periods of vibration in the same axis typically have a ratio of approximately 3 to 1. Spectral acceleration values at these two periods are often used by engineers (e.g., in response spectrum analysis, Chopra 2001), and we see that if the periods are greater than 0.189s, the correlation coefficient between these two *Sa*'s is approximately 0.6. When considering two periods with a ratio of 3 to 1 in orthogonal horizontal directions, we can use the Markov approximation and estimate the correlation coefficient as 0.8*0.6=0.48. When considering vertical ground motions, if we assume that the vertical period of interest is 0.1s and the horizontal period of interest is 0.5 to 1s (for mid-rise buildings) then we see that the correlation coefficient is approximately 0.3 to 0.4. If we define the correlation distance as the ratio of periods T_{max}/T_{min} such that the correlation coefficient between the two is e^{-1}=0.37, then the correlation distance is approximately 5 for vertical records and 6.5 for horizontal records (if T_{min}>0.189). These numbers may serve as useful rules-of-thumb for quick estimates.

8.6 Comparisons with previous work

The predictions provided by Equations 8.7 through 8.12 are believed to be the first of their kind in most cases, except for two previous models analogous to Equation 8.9. Inoue and Cornell (1990) proposed the following analogous model:

$$\rho_{\varepsilon_x,\varepsilon_x} = 1 - 0.33 \cdot \ln(T_{\max} / T_{\min}) \tag{8.13}$$

The function was fit over a period range between 0.1 seconds and 4 seconds, using 64 record components. Its contours are plotted in Figure 8.4c. This prediction agrees reasonably well with empirical results within the range over which it was fitted, but it would perform poorly if extrapolated to the larger period range used by the newer models.

Another model for the correlations of epsilons at two periods along the same component is provided by Abrahamson et al. (2003):

$$\rho_{\varepsilon_x,\varepsilon_x} = \begin{cases} \tanh\left(1-1.32\cdot\log_{10}\left(X\right)+0.072\left(\log_{10}\left(X\right)\right)^2-0.05-0.5\cdot\log_{10}(f_c)\right) & \text{for } f_c \le 3.16\,Hz \\ \tanh\left(1-1.32\cdot\log_{10}\left(X\right)+0.072\left(\log_{10}\left(X\right)\right)^2-0.95+1.3\cdot\log_{10}(f_c)\right) & \text{for } f_c > 3.16\,Hz \end{cases} \tag{8.14}$$

where f_1 and f_2 are the larger and smaller frequencies respectively, f_c=0.5(f_1+f_2), and X=ln(f_1/f_2). The function was fit over a range of periods between 0.03 seconds and 5 seconds, using a record set similar to the one used here. The contours of this prediction are shown in Figure 8.4d. It is not dramatically different from that of Equation 8.9, displayed in Figure 8.4b. Selected contours from Figure 8.4a, b and c are overlaid in Figure 8.9 to aid comparisons of the predictions. It should be noted that one desirable attribute missing from the Abrahamson et al. prediction is positive definiteness when the correlation matrix is computed for multiple periods simultaneously. This violates a required property of correlation matrices as discussed above.

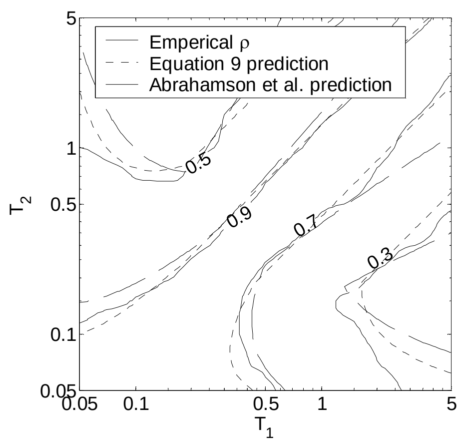


Figure 8.9. Overlaid contours at four correlation levels for the empirical correlations, the prediction from Equation 8.9 and the prediction from Abrahamson et al. (2003).

In general, these previous models agree reasonably well with Equation 8.9 proposed here. Equation 8.9 may be considered an improvement due to its increased range of periods as compared with the model of Inoue and Cornell and due to its positive definiteness property, which the Abrahamson et al. model does not possess. Equation 8.9 also produces smaller residuals than these two models when predicting correlations of either the primary record set above or the Chi-Chi record set (which was not used to fit any of the three models).

One additional study of spectral acceleration correlations used 30 records from the 1999 Chi-Chi, Taiwan, earthquake (Wang et al. 2001). The results are not directly comparable to the work here, however, and so this work is not considered further.

Previous work also exists that can be indirectly compared to Equation 8.7. Some ground motion prediction models provide standard deviations of residuals for both the logarithmic spectral acceleration of a single horizontal component of a ground motion, or for the geometric mean of two orthogonal components (e.g., Boore et al. 1997, Spudich et al. 1999). By examining the ratios of the two standard deviations, one can back-calculate the implied correlation coefficient between the two components using the following equation:

$$\rho_{\varepsilon_x,\varepsilon_y} = 2\sigma^2_{g.m.} \big/ \sigma^2_{arb} - 1 \tag{8.15}$$

where $\sigma_{g.m.}$ is the logarithmic standard deviation of the geometric mean of the two horizontal components, and σ_{arb} is the logarithmic standard deviation of an arbitrary component. The correlation coefficients implied by the models of Boore et al. (1997, 2005) and Spudich et al. (1999) are displayed in Figure 8.2. These models underestimate the correlation seen empirically in this study, but estimation of correlations was not a goal of these studies. The value of interest to these authors is $\sigma^2_{arb} \big/ \sigma^2_{g.m.}$, but a slight change in this ratio can produce a large change in the correlation coefficient. Given that these authors were not interested in correlations and that the correlations calculated using Equation 8.15 are very sensitive to the ratio $\sigma^2_{arb} \big/ \sigma^2_{g.m.}$, it is perhaps not surprising that there is a slight discrepancy between the direct calculations of this paper and the indirect back-calculations from previous work.

8.7 Applications

To demonstrate the usefulness of these new predictions and perhaps inspire new uses, several applications of the new correlation predictions are briefly described here.

8.7.1 Vector-valued hazard analysis for horizontal and vertical components of ground motion

The vector-valued hazard analysis methodology of Bazzurro and Cornell (2002) can now be easily applied to analysis of horizontal and vertical ground motions simultaneously. Consider a hypothetical two-dimensional building frame with a first-mode period of 1 second in the horizontal direction and a first-mode period of 0.1 seconds in the vertical direction. Using Equation 8.12, we estimate a correlation coefficient between the *Sa*'s at these two periods of 0.30. We assume that the building is located 8 km from a single fault which produces only (characteristic) magnitude 6.5 earthquakes with a mean return period of 500 years. Using the Abrahamson and Silva (1997) ground motion prediction model and the correlation coefficient predicted here, we can compute the joint distribution of horizontal and vertical spectral acceleration values at the specified first-mode periods. Contours of that hazard are displayed in Figure 8.10. It is interesting to note that because of the low correlation, extreme values of *Sa* are unlikely to occur in the horizontal and vertical directions simultaneously. For example, we note that a horizontal *Sa* of 0.65g has a 2% probability of exceedance in 50 years, and a vertical *Sa* of 0.98g has a 2% probability of exceedance in 50 years. But the probability of exceeding *both* a horizontal *Sa* of 0.65g and a vertical *Sa* of 0.98g *simultaneously* is only 0.65% in 50 years. This suggests that designing for extreme ground motions in all directions simultaneously (e.g., by applying a horizontal Uniform Hazard Spectrum and a vertical Uniform Hazard Spectrum simultaneously), may be more conservative than intended. If one is primarily concerned with horizontal motions and thus uses the 2% in 50 years horizontal *Sa* value, the preferred vertical *Sa* design value would be that associated with the mean ln*Sa* in the vertical direction, given that *Sa* in the horizontal direction has exceeded 0.65g. This choice is consistent with load combination rules used in practice elsewhere (e.g., Norwegian Technology Standards Institution 1999, p17-18). For this example the design value of the vertical *Sa* was determined to be 0.66g, which is approximately 30% less than the *Sa* with a 2% probability of exceedance in 50 years.

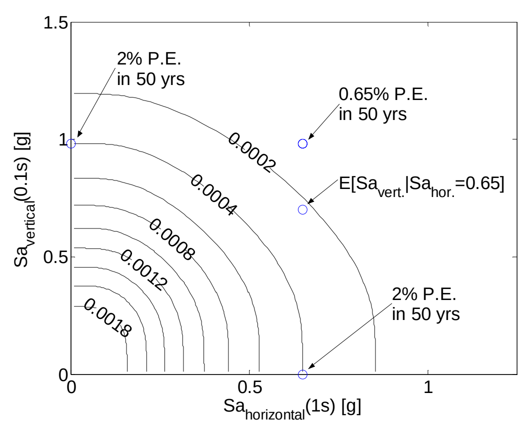


Figure 8.10. Contours of vector-valued probabilistic seismic hazard analysis. The contours denote the mean annual rate of exceeding *both* the $Sa_{vertical}$ *and* the $Sa_{horizontal}$ values.

Only a single magnitude/distance pair was used in this example for computational simplicity. Generalization of the analysis to incorporate multiple faults with multiple magnitudes and distances is merely a matter of coding the correlation prediction into a vector-valued hazard analysis program (Somerville and Thio 2003). (As an approximation, one can also use the dominant value of ε obtained by disaggregation to obtain the associated ε's for other components.) No further mathematical developments are needed.

8.7.2 Ground motion prediction model for the geometric mean of orthogonal spectral accelerations at two periods

The geometric mean of spectral acceleration values in two orthogonal horizontal directions is often computed in ground motion prediction models. This quantity is useful for analyzing a three-dimensional structure subjected to ground motions in two horizontal directions, because it describes the intensity of ground motion in two directions using only a single parameter (Stewart et al. 2001, Baker and Cornell 2005b). The geometric mean of spectral acceleration provided by ground motion prediction models uses the same period of vibration in both directions, however, while a structure commonly has different periods of vibration in its two principal directions. Using the correlation models presented above, one can easily develop a "custom" correlation model incorporating the two periods of interest in a particular application.

Ground motion prediction models provide the $f_H(M,R,T,\theta)$ and $\sigma_H(M,T)$ terms for Equation 8.1. We are interested in determining an analogous equation of the form:

$$\ln Sa_{g.m.}(T_1,T_2) = f_{g.m.}(M,R,T_1,T_2,\theta) + \sigma_{g.m.}(M,T_1,T_2)\varepsilon_{g.m.}(T_1,T_2) \tag{8.16}$$

where $Sa_{g.m.}(T_1,T_2)$ is the geometric mean of two orthogonal horizontal spectral accelerations at two periods T_1 and T_2. The function $f_{g.m.}(M,R,T_1,T_2,\theta)$ is the mean value of $\ln Sa_{g.m.}(T_1,T_2)$ and $\sigma_{g.m.}(M,T_1,T_2)$ is the standard deviation. Recognizing that $\ln Sa_{g.m.}(T_1,T_2) = 1/2\left(\ln Sa_x(T_1) + \ln Sa_y(T_2)\right)$, the mean and standard deviation terms can be derived from existing models and the correlation predictions presented above:

$$f_{g.m.}(M,R,T_1,T_2,\theta) = 1/2\left(f_H(M,R,T_1,\theta) + f_H(M,R,T_2,\theta)\right) \tag{8.17}$$

$$\sigma_{g.m.}(M,T_1,T_2) = \sqrt{\frac{\sigma_H^2(M,T_1)}{4} + \frac{\sigma_H^2(M,T_2)}{4} + \frac{\rho_{\varepsilon_x,\varepsilon_y}(T_1,T_2)\sigma_H(M,T_1)\sigma_H(M,T_2)}{2}} \tag{8.18}$$

where $\rho_{\varepsilon_x,\varepsilon_y}(T_1,T_2)$ comes from Equation 8.11. Remembering that many ground motion prediction models provide the standard deviation of the geometric mean of two horizontal *Sa*'s, rather than the standard deviation of a single component *Sa*, it may be first necessary to make the following conversion before calculating Equation 8.18:

$$\sigma_H^2(M,T) = \frac{2\sigma_{g.m.}^2(M,T)}{1+\rho_{\varepsilon_x,\varepsilon_y}} \tag{8.19}$$

where $\rho_{\varepsilon_x,\varepsilon_y}$ comes from Equation 8.7, $\sigma_{g.m.}(M,T)$ is the standard deviation of the geometric mean of two horizontal *Sa*'s (the quantity presented in most ground motion prediction models), and $\sigma_H(M,T)$ is the standard deviation of a single component *Sa* (the quantity used in Equations 8.1, 8.2 and 8.18). The $f_H(M,R,T,\theta)$ term is unchanged regardless of whether the geometric mean or arbitrary component *Sa* definition is adopted, so it can be taken from the ground motion prediction model without modification. Equations 8.17 through 8.19 can be easily implemented in a computer code alongside an existing ground motion prediction, and the output used in the same way as the output from any other ground motion prediction (e.g. for PSHA analysis). It is expected that this "custom" model and corresponding hazard analysis will allow an engineer to increase the precision of response analyses in cases where the structure of interest has different periods of vibration in its two principle directions.

8.7.3 Simulation of response spectra

The correlation predictions derived above can be used to simulate response spectra given an earthquake scenario. For a given earthquake magnitude (*M*), distance (*R*), and other parameters (θ), the distribution of horizontal ln*Sa* values at a range of periods (T_1, T_2, …, T_n) can be obtained using the model of Equation 8.1. The mean of ln$Sa(T_i)$ is equal to $f_H(M,R,T_i,\theta)$ and its standard deviation is equal to $\sigma_H(M,T_i)$. The covariance of ln$Sa(T_i)$ and ln$Sa(T_j)$ is equal to $\sigma_H(M,T_i)\sigma_H(M,T_j)\rho_{\varepsilon_x,\varepsilon_x}(T_i,T_j)$, where $\rho_{\varepsilon_x,\varepsilon_x}(T_i,T_j)$ is computed using Equation 8.9. If we again assume a multivariate normal distribution for ln*Sa* values, then these means and covariances fully define the distribution which can be used for simulation. A comparison of empirical spectra and simulated spectra is shown in Figure 8.11. In Figure 8.11a and b, spectra are simulated using zero correlation and perfect correlation, respectively. In Figure 8.11c, spectra are simulated using the correlation prediction from Equation 8.9. In Figure 8.11d, real spectra from recorded ground motions are shown for comparison. It is clear that the results in Figure 8.11a and b are not accurate representations of real record spectra, and so a model such as that presented here is needed for accurate simulation. Note that this simulation procedure can also be applied to simulation of multiple-component response spectra.

These simulated spectra may also be compared with synthetic ground motions to verify that the spectra of the synthetic motions show sufficient variability (or "roughness"). In cases where synthetic spectra are not "rough" enough corrective actions could be taken to increase the variability (e.g., "rough" spectra were generated in Sewell et al. 1996).

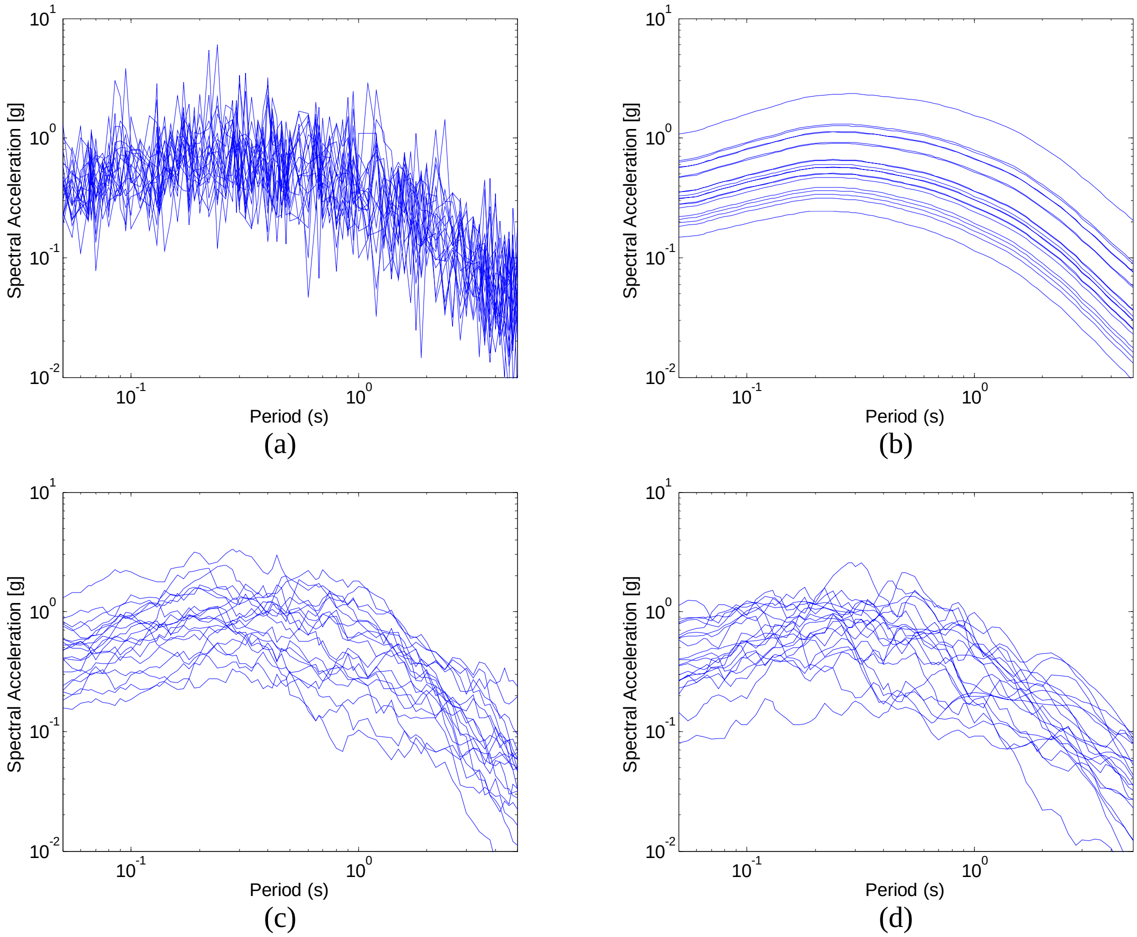


Figure 8.11. Samples of 20 response spectra from magnitude 6.5 earthquakes with a source-to-site distance of 8km. The simulated spectra use means and variances from Abrahamson and Silva (1997). (a) Simulated spectra using correlation coefficients equal to zero between all periods. (c) Simulated spectra using correlation coefficients equal to one between all periods. (c) Simulated spectra using correlation coefficients from Equation 8.9. (d) Real spectra from recorded ground motions with magnitude $\cong$ 6.5 and distance $\cong$ 8km.

8.8 Conclusions

Models have been presented for correlations of spectral response values of earthquake ground motions. The models include predictions of correlations for single-component ground motions measured at two differing periods, and also across orthogonal components of three-dimensional ground motions. The predictive models improve upon previous predictions of correlations at differing periods in a single horizontal ground motion. For correlations within a vertical ground motion or across orthogonal components of a ground motion, these predictions are believed to be the first of their kind. It is seen that correlations of *Sa* at differing periods across orthogonal horizontal components of a ground motion can be approximated as a product of the correlation of

Sa's across differing components (at a single period) and the correlation at differing periods (in the same component).

The predictions are presented in the form of correlations of standardized residuals ("epsilons") from established empirical ground motion prediction models. These predictions provide information about the correlations of two logarithmic *Sa* values for a given magnitude and distance. Although the observed residuals are in principle dependent on the ground motion prediction model chosen, it is observed that the correlations do not vary significantly when the underlying model is changed. Thus the correlation predictions are applicable regardless of the ground motion prediction model used by the analyst. This suggests that although one might repeat this exercise when the models from the current Next Generation Attenuation project are released, the functional forms and even parameter values should not change appreciably, at least for vertical or randomly oriented horizontal components. (Correlations among or within fault normal and fault parallel components were not examined as part of this study, but it will be possible to examine them upon completion of the Next Generation Attenuation project.)

Several approximate "rule-of-thumb" correlation values can be determined from the above models. It is seen that correlation of orthogonal horizontal *Sa* values with the same period are comparatively highly correlated ($\rho \cong 0.8$), while horizontal and vertical *Sa* values at typical first-mode periods of mid-rise buildings are less correlated ($\rho \cong 0.3$ to 0.4). Several past and potential future applications are presented, illustrating that the correlations shown here are useful for a variety of earthquake hazard and engineering problems. Increased knowledge of response spectrum correlations will facilitate the further development of vector-valued probabilistic seismic hazard analysis, as well as allow simple modification of existing ground motion prediction models to develop custom predictions for any combination of periods and orientations. These applications will allow analysts to better understand the properties of multi-component earthquake ground motions.

Chapter 9

Conclusions

In this dissertation, vector-valued Intensity Measures (IMs) have been considered for use in probabilistic assessments of the seismic performance of structures. Use of vector-valued IMs requires methods for predicting structural response as a function of the IM parameters, calculating a vector-valued ground motion hazard, and choosing effective parameters to include in the IM. Contributions have been made in all three of these areas. The following subsections summarize briefly the important findings of this work, the limitations of this work and suggested future work related to this dissertation.

9.1 Practical implications

9.1.1 Structural response prediction given a vector IM

Several methods for prediction the probability distribution of structural response given a scalar IM have been considered previously (see, e.g., Jalayer 2003 for a summary). In Chapter 2 of this dissertation, extensions of those methods to vector IMs have been proposed and discussed. In particular, while most previous work with vector IMs has been based on joint "cloud" regression on multiple IM parameters simultaneously (Shome and Cornell 1999, Bazzurro and Cornell 2002, Luco et al. 2005), a new method is proposed and used extensively in this dissertation. This method consists of scaling records to the first IM parameter (typically $Sa(T_1)$) and then using regression analysis to predict response as a function of one or more additional parameters. This method is advantageous for two reasons. First, the interaction between the first and subsequent IM parameters is not parameterized, making this interaction easier to detect than with multiple regression alone, where the interaction must be explicitly parameterized. Second, it allows one to see more clearly the separate effects of correlated IM parameters (such as $Sa(T_1)$ and ε, as was seen in Chapter 3). The downfall of this new method is that it generally requires more dynamic analyses than the "cloud" method, which may reduce its usefulness in practice in the near future. However, because of the advantages listed above, it is believed to be superior to any alternative methods for exploratory research purposes. Other methods of incorporating vector IMs, such as carefully selecting records to have the appropriate distribution of IM values, have also been explored in depth.

9.1.2 The parameter ε as a predictor of structural response

The ground motion parameter ε, defined as a measure of the difference between an accelerogram's recorded spectral acceleration value at a specified period and the predicted value from a ground motion prediction model, has been found to be an effective predictor of structural response. The parameter ε is important because it provides information about the shape of a record's response spectrum (i.e., whether *Sa* at the specified period is in a peak or a valley of the spectrum). The effect of ε is seen to be greater than the effect of either magnitude or distance. Further, rare ground motions (which are likely to be of most engineering interest) are associated with positive ε values. Positive-ε ground motions tend to cause smaller demands in nonlinear or multi-degree-of-freedom structures, given the same $Sa(T_1)$ value. This implies that neglecting the effect of ε will result in conservative conclusions being drawn about the response of a structure: over-estimation of drift associated with a specified ground motion level, and over-estimation of the mean annual rate of exceeding a given response limit state such as collapse. When ε was neglected for the example cases considered in Chapter 3, the estimated mean annual rate of collapse was often biased high by a factor of two, and sometimes by a factor of more than ten.

Once the effect of ε was identified, two methods were proposed to incorporate its effect in a probabilistic structural response assessment. The first method, considered in Chapter 3, consists of including ε as a parameter in the intensity measure. This ensures that structural response is predicted as a function of ε, and the ε values associated with a given ground motion level are included through the ground motion hazard. The (vector) ground motion hazard for this IM is simple to obtain, as it consists of the standard scalar ground motion hazard for $Sa(T_1)$ along with the conditional distributions of ε given $Sa(T_1)$, which can be obtained from PSHA disaggregation. The second method of incorporating ε, considered in Chapter 6 and discussed further in Section 9.1.5, consists of using the PSHA disaggregation to determine the mean value of ε for each $Sa(T_1)$ level of interest. Records for analysis are then carefully selected at each $Sa(T_1)$ level to have ε values as close to this mean value as possible. With this method, the responses from these records are naturally at the level associated with the proper ε value, and so no vector IM is necessary.

Additionally in Chapter 6, a method of selecting records based on their consistency with the spectral shape implied by ε, rather than by their actual ε values, was considered. It was seen that either of these record-selection procedures can produce results comparable to those obtained using a vector IM with ε.

The target ε value associated with a given hazard level is site dependent. For highly seismic areas such as coastal California, typical target ε values are in the range of 1 to 2 at the 10% in 50 year hazard level, and 2 or greater at the 2% in 50 year hazard level. The limited library of ground motion recordings makes it difficult to find many records with ε values larger than 2, so when selecting records to analyze a structure for large ground motion intensities, one might practically select those records with the largest available ε values. In regions with low seismicity such as some parts of the Eastern United States, the ground motion intensities exceeded with 2% in 50 year probability might have associated ε values near one[20].

Consideration of ε when performing structural analysis is important in order to avoid conservative conclusions about structural performance, especially at rare levels. To date, ε has been considered in one other research project by selecting records based on ε values, in a similar manner to that proposed in Chapter 6 (Haselton et al. 2005). Given the findings in this dissertation, it is suggested that this practice be adopted widely.

9.1.3 Optimal vector-valued IMs consisting of spectral acceleration values at multiple periods

In Chapters 4 and 5, a vector was considered that consists of two parameters: $Sa(T_1)$ along with a measure of spectral shape (the ratio of spectral acceleration at a second period to the original spectral acceleration value). The extra IM parameter is denoted $R_{T1,T2}$, where T_2 is the second period at which spectral acceleration is measured. In order to be most effective, T_2 should be chosen based on the level of nonlinearity in the structure and the structure's sensitivity to higher-mode response. This IM has been considered by others, but a more thorough investigation has been made here. Two methods of choosing an optimal second period were proposed (fixing T_1 at the first-mode period of the structure). The first method consists of choosing T_2 in order to minimize the standard deviation of structural responses after regressing on $R_{T1,T2}$ at a given $Sa(T_1)$ level. The second method consists of minimizing the standard error of estimation of the drift hazard curve calculated using the vector IM $\{Sa(T_1), R_{T1,T2}\}$. With the second method, the

[20] In these less seismically active regions, ε values might even be negative at some low ground motion levels, making it possible that in some situations neglecting ε could be unconservative. These situations may not be of engineering interest, but the possibility that they could arise should be kept in mind.

standard error is estimated by using the bootstrap to repeatedly re-sample results from the available dynamic analyses. The first method is simpler, but the second method has the advantage of directly computing the statistical variability in estimates of the drift hazard curve.

Prediction of structural response with this vector shows significantly reduced standard deviations relative to prediction using $Sa(T_1)$ alone (reductions of as great as 50% were observed). This suggests that the number of dynamic analyses could be reduced if this vector is adopted. Further, this vector IM is effective at characterizing the effect of near-fault ground motions, as will be discussed in Section 9.1.4. Because $R_{T1,T2}$ describes the shape of the response spectrum, it was hoped that this parameter might account for the effect of ε (which also describes spectral shape). However, it was found that $R_{T1,T2}$ does not fully account for the effect of ε. Specifically, when ε is included in an IM it removes a bias in structural response predictions, and when $R_{T1,T2}$ is included in an IM it does not remove this bias. This shortcoming of $R_{T1,T2}$ occurs because ε describes somewhat different properties of spectral shape than $R_{T1,T2}$. In particular, $R_{T1,T2}$ measures spectral shape on only one side of T_1, while ε measures spectral shape at both higher and lower periods simultaneously. Thus, the effect of ε still needs to be considered explicitly when using this IM, either through careful record selection or through use of a three-parameter vector consisting of $Sa(T_1)$, $R_{T1,T2}$ and ε.

9.1.4 Vector-valued IMs for predicting response from pulse-like ground motions

In Chapter 5, the effectiveness of vector-valued IMs for predicting response from pulse-like ground motions was considered (where the term "pulse-like ground motions" is used to refer to those ground motions occurring nearby a fault where a velocity pulse is detected in the fault-normal direction). These ground motions can cause severe levels of structural response, which is a significant concern for engineers. The effect of these ground motions was not well-captured by $Sa(T_1)$ alone, but the vector consisting of $Sa(T_1)$ and $R_{T1,T2}$ was better able to account for the effect of these motions. Given $Sa(T_1)$ and $R_{T1,T2}$, there was found to be no significant difference in the maximum interstory drift ratios of frame structures caused by ordinary and pulse-like ground motions. This IM also substantially accounted for the effect of the velocity pulse period on the resulting maximum interstory drift ratio. This suggests that the intensity-measure-based procedure might be used effectively even when pulse-like ground motions may occur at a site, as long as a vector of $Sa(T_1)$ and $R_{T1,T2}$ is used as the intensity measure.

In order to perform the IM-based procedure with this IM, however, it is still necessary to compute the ground motion hazard for $Sa(T_1)$ and $R_{T1,T2}$ while considering the effect of pulse-like ground motions on the IM levels that are likely to occur. Two methods for computing this hazard were proposed. The first is based on the method proposed by Bazzurro and Cornell (2002), except that the ground motion prediction model should be modified to account for near-fault effects. The second method incorporates explicit consideration of the probabilities of occurrence of velocity pulses, along with the distribution of their associated pulse periods, conditioned on each possible magnitude and distance. Results obtained using the first method will be available soon (Somerville and Thio 2005). The second method should produce more accurate hazard estimates, but it requires new ground motion models that do not yet exist. Once these ground motion hazard analysis results are available, they can be combined with structural response analyses of the type performed in Chapter 5 in order to assess the performance of a structure located at a site where pulse-like ground motions might occur.

The IM consisting of $Sa(T_1)$ and ε was also considered for prediction of response from pulse-like ground motions, but it was found to be ineffective at distinguishing between pulse-like and ordinary ground motions. This result is consistent with the current understanding of the effect of ε. It suggests that while the IM consisting of $Sa(T_1)$ and ε is useful in general, it is *not* able to quantify the effect of pulse-like ground motions.

9.1.5 Record selection for dynamic analysis

Based on the understanding gained in the previous chapters, new techniques for selecting "appropriate" ground motion records were considered in Chapter 6, with the goal of obtaining the benefits of vector IMs without their added analysis complexity. Given that ε is a ground motion parameter affecting structural response, record selection incorporating this knowledge was considered, and successfully accounted for the effect of ε without using a vector IM. Further, knowledge of the mean magnitude, distance and ε values associated with a given $Sa(T_1)$ level (as determined from PSHA disaggregation) was used to determine the conditional mean value of a response spectrum at the given Sa level. The spectrum was named a *Conditional Mean Spectrum, considering* ε (abbreviated as CMS-ε). Records can then be selected based that match this CMS-ε spectrum, without considering further the magnitude, distance or ε values of the records. Records selected in this way were shown to account for the effect of ε, but the number of records available using this method is potentially larger than the number of records with the exact target ε value. Further, records selected based on either their ε values or their match with the CMS-ε spectrum were able to be scaled without introducing a bias in structural response, unlike records selected using other methods. The CMS-ε spectrum is somewhat analogous to target response spectra used in the nuclear safety industry (DOE 1996, NRC 1997, ASCE 2005), except that the effect of ε has been incorporated as well, given the new findings in this dissertation. In Section 6.7, the CMS-ε

spectrum was compared to a Uniform Hazard Spectrum (UHS), in order to demonstrate the conservativeness of the UHS. The CMS-ε spectrum is a useful tool for use in selecting records for dynamic analyses associated with probabilistic performance assessments, but it could also be a useful target spectrum for other code-based procedures as well.

9.1.6 Consistency in ground motion hazard and structural response prediction for probabilistic structural assessment

The probabilistic assessment procedure used throughout this dissertation relies on combining the ground motion hazard (in terms of the specified IM) with conditional structural response predictions (structural response given IM). In practice today, the IM most often used is $Sa(T_1)$. However, the ground motion hazard is typically computed for the geometric mean of the spectral accelerations of the two horizontal components of ground motion, while prediction of the response of (two-dimensional) structural models is made using a single horizontal component. This inconsistency is identified in Chapter 7, and its implications are discussed. The error results in unconservative estimates of the rates of exceeding given structural response levels. For an example site and structure located in Los Angeles, inconsistent treatment of $Sa(T_1)$ resulted in a 12% underestimation of the spectral acceleration value exceeded with a 2% probability in 50 years, and a 10% underestimation of the structure's maximum interstory drift ratio exceeded with a 2% probability in 50 years. Several methods of correcting the error are discussed, and suggestions are made regarding permanent correction of this error, such as modifying the U.S. Geological Survey hazard maps (2002) to be defined in terms of the $Sa(T_1)$ parameter used most often for structural response prediction.

9.1.7 Correlation of response spectral values for use in hazard analysis

An important part of the vector-valued IM analysis process is computation of ground motion hazard for the vector IM. When the vector IM consists of multiple spectral acceleration values, it is necessary to know the correlation coefficients between these spectral acceleration values in a given ground motion (Bazzurro and Cornell 2002). In Chapter 8, new predictive equations were proposed for these correlation coefficients. Several approximate "rule-of-thumb" correlation values can be determined from the models: correlation of orthogonal horizontal Sa values with the same period are comparatively highly correlated ($\rho \cong 0.8$), while horizontal and vertical Sa values at typical first-mode periods of mid-rise buildings are less correlated ($\rho \cong 0.3$ to 0.4).

The predictive equation for correlations within a single-component ground motion improves upon two previous models of the same type. The model proposed here is valid for a larger range of periods than the model of Inoue and Cornell (1990), and the matrix of correlations at multiple periods predicted using this model has the required property of positive definiteness, unlike the prediction of Abrahamson et al. (2003). Also, the first predictive equations are developed for correlations among different components of a three-dimensional ground motion (i.e., between perpendicular horizontal components or between a horizontal and a vertical component). These new models will facilitate analysis of three dimensional structures by allowing the incorporation of spectral acceleration values from orthogonal ground motion components into a vector IM. Software for vector-valued probabilistic seismic hazard analysis that incorporates these new models should be available in the near future (Somerville and Thio 2005).

9.2 Limitations and future work

9.2.1 Structural response parameters of interest

The only structural response parameter considered in this dissertation is maximum interstory drift ratio in frame structures. This parameter has been used in, e.g., the FEMA-350 guidelines (2000a), because it is a good indicator of the ability of a frame structure to resist P-Δ instability and collapse, as well as and indicator of peak rotation demands on beams, columns and connections. However, modern probabilistic loss estimation procedures (e.g., Aslani and Miranda 2003, Haselton et al. 2005) often require knowledge of floor accelerations as well. Further, these loss estimation procedures typically require a vector of response parameters at each story of a structure, rather than a peak value over all stories. Thus, maximum interstory drift ratio may not be satisfactory as a structural response parameter. It is not obvious *a priori* that the vector IMs proposed here will be equally effective for predicting these other response parameters of interest. In fact, Taghavi and Miranda (2003) have found that effective IMs for prediction of peak floor accelerations may differ from effective IMs for prediction of interstory drift ratios. Further investigation of this observation could be performed using the same procedures proposed in this dissertation. If one is computing expected annual losses only and total losses are a sum of element losses, then separate IMs could be used to predict displacements and accelerations, and the mean annual losses from drift-sensitive and acceleration-sensitive elements could be computed separately and summed. However, if complete distributions of losses are desired, then a joint prediction of all response parameters is needed, and a single intensity measure must be used. However, the question of which IM is most effective for joint prediction of a vector of response parameters will need to be posed carefully, because tradeoffs will need to be made between effective predictions of the various structural response parameters.

9.2.2 Intensity measure parameters

The intensity measure parameters considered in this dissertation consisted of spectral acceleration values at multiple periods, as well as the ground motion parameter ε. These parameters have been shown to be effective in this dissertation, and other researchers have also confirmed the effectiveness of multiple spectral parameters (Shome and Cornell 1999, Bazzurro and Cornell 2002, Vamvatsikos 2002). However, other parameters may also be useful. In Chapter 6, an IM parameter was proposed consisting of spectral accelerations averaged over a range of periods. This intensity measure has been suggested elsewhere (Kennedy et al. 1984, Shome and Cornell 1999, Cordova et al. 2001) and it might be more robust with respect to the effect of ε identified in Chapter 3. Further, given correlation information such as that in Chapter 8, the ground motion hazard for this IM can be easily developed based on existing ground motion prediction models for single spectral acceleration values, as was discussed in Chapter 6. Many of the studies presented here could be repeated on this IM, in order to determine its robustness with respect to ε, its efficiency in predicting structural response and its role in record selection.

Intensity measures consisting of inelastic response spectral values may also prove effective as structural response predictors. Intensity measures of this type have been seen to have some of the same benefits as the IMs considered here (Mori et al. 2004, Luco and Cornell 2005, Tothong and Cornell 2005b, Luco et al. 2005). More research is needed to determine, for example, whether these intensity measures account for the effect of ε.

Only two-element vectors were considered in depth in this dissertation, although a three-element vector was considered briefly in Chapter 4. As seen in Chapter 4, larger vectors may provide additional insight into the relationship between record properties and structural response. For example, questions remain about the relationship between the $R_{T1,T2}$ parameter and the ε parameter, given that both are descriptors of spectral shape. It is anticipated that extremely large vectors of parameters will be impractical (due to the curse of dimensionality), but more work is needed to determine a limiting size. Three-element vectors may be useful in some situations, but that was not determined with certainty in this work.

9.2.3 Structural models considered

The models used for evaluation here represent only a small subset of the range of possible models (and their associated physical structures). Only two-dimensional frame models were considered. Three dimensional models will provide a rich set of new research problems related to this work. For example, a vector IM consisting of *Sa* values in two perpendicular directions may be very effective for predicting response of a three-dimensional structure, but testing of this IM is currently quite limited (e.g., Vamvatsikos and Sigalas 2005), and response prediction using this IM has never been coupled with a ground motion hazard analysis. The new models of Chapter 8 will allow ground motion hazard analysis to be performed for this IM in the future.

Additionally, while the structural models considered include strength degradation, they do not include other effects known to influence structural response such as soil-structure interaction, irregularities in elevation, and axial failure of columns. Further, other structural systems such as walls and dual systems have not been considered. Intuition developed in this dissertation about the relationship between the proposed IMs and spectral shape, plus our confidence that spectral shape will affect response of all structures, suggests that these improved vector IMs will be effective for other types of structures. It will be important to verify that hypothesis empirically, however.

9.2.4 Hazard analysis for pulse-like near-fault ground motions.

The intensity measure consisting of $Sa(T_1)$ and $R_{T1,T2}$ was found to be effective at characterizing the effect of pulse-like near-fault ground motions, given a (ductility-dependent) effective choice of T_2. However, there are still a variety of related unresolved questions. Challenges remain in the computation of ground motion hazard for a vector IM at a site where near-fault motions may occur. For this to be done using the more accurate of the two methods proposed in Chapter 6, it is necessary to know the probability of a velocity pulse occurring in a record for a given magnitude, distance and source-to-site geometry. It is also necessary to have a prediction model for the distribution of $Sa(T_1)$ and $R_{T1,T2}$, given a ground motions with a velocity pulse and a given pulse period. Further, it would be helpful to have a quantitative method for identifying whether a given record is actually a near-fault record or an ordinary record. This classification is made in a variety of ways today, but all of them require at least some level of subjective judgment on the part of the analyst. An objective classification method will be helpful for a range of research problems, such as investigation of IMs for predicting structural response and development of ground motion prediction models.

In addition, the effect of ε was not considered in this study of near-fault ground motions, even though it was seen to be important in other sections of the dissertation. Once models for hazard analysis are developed that consider pulse-like ground motions, it will be possible to measure the changes to spectral shape from consideration of both ε and pulse-like ground motions. Consideration of ε tends to reduce the estimated demand on structures, while pulse-like ground motions tend to increase the demand. The net effect when both of these phenomena are considered has not yet been determined.

9.2.5 Adoption of vector IMs in code procedures

While the proposed vector IMs can be adopted in the assessment methodology of the Pacific Earthquake Engineering Research center, they do require much more effort than is allowable for a more common code-based assessment procedure. In order to achieve widespread adoption of the ideas presented in this dissertation, simplified procedures for using these intensity measures are needed. On particularly promising method is environmental contours (e.g., Winterstein et al. 1993, Bazzurro 1998), a procedure which has been adopted for similar problems in assessment of offshore structures. Future work by the author will further explore this opportunity.

9.3 Concluding remarks

The summaries and observations made in this chapter are based on empirical analysis of two dimensional nonlinear frame structures subjected to a variety of recorded earthquake ground motions of varying intensities. Maximum interstory drift ratio was the only response parameter considered, and only a subset of potential intensity measures was considered when searching for an “optimal” intensity measure. Interpretation of the conclusions drawn within this dissertation should be made while keeping these limitations in mind. Given the success of vector valued intensity measures within these limitations, further related research appears warranted.

Appendix A

Earthquake Ground Motion Records

Several sets of earthquake ground motions were used for various purposes throughout this dissertation. The record properties are given in the tables below.

All records except those from Table A.4 come from the PEER Strong Motion Database (2000). Column headings match fields from that database. The records from Table A.4 come from the Next Generation Attenuation (NGA) project strong motion database (2005). Column headings match fields from that database. Empty fields are either not applicable for the given record, or not provided by the source database.

Table A.1. Primary recordset used in stripe analysis of the Van Nuys Testbed building. These records were used in Chapters 2, 3, 4, 6 and 7.

#	Event	Mechanism	Year	M_w	Station Name	Distance	USGS Soil Type	FileName[1]	HP	LP
1	Loma Prieta	RV/OB	1989	6.9	Agnews State Hospital	28.2	D	LOMAP/AGW090	0.2	30
2	Loma Prieta	RV/OB	1989	6.9	Anderson Dam (Downstream)	21.4	D	LOMAP/AND270	0.2	41
3	Northridge	TH	1994	6.7	Arleta - Nordhoff Fire Sta	9.2	D	NORTHR/ARL090	0.12	23
4	Landers	SS	1992	7.3	Coolwater	21.2	D	LANDERS/CLW-LN	0.1	30
5	Landers	SS	1992	7.3	Desert Hot Springs	23.2	D	LANDERS/DSP000	0.07	23
6	Cape Mendocino	-	1992	7.1	Fortuna - Fortuna Blvd	23.6	D	CAPEMEND/FOR000	0.07	23
7	Imperial Valley	SS	1979	6.5	Aeropuerto Mexicali	8.5	D	IMPVALL/H-AEP315	0.05	-
8	Imperial Valley	SS	1979	6.5	Agrarias	12.9	D	IMPVALL/H-AGR273	0.05	-
9	Imperial Valley	SS	1979	6.5	Chihuahua	28.7	D	IMPVALL/H-CHI282	0.05	-
10	Imperial Valley	SS	1979	6.5	El Centro Array #13	21.9	D	IMPVALL/H-E13140	0.2	40
11	Imperial Valley	SS	1979	6.5	Holtville Post Office	7.5	D	IMPVALL/H-HVP225	0.1	40
12	Coalinga	RV/OB	1983	6.4	Pleasant Valley P.P. - yard	8.5	D	COALINGA/H-PVY045	0.2	40
13	Loma Prieta	RV/OB	1989	6.9	Hollister - South & Pine	28.8	D	LOMAP/HSP000	0.1	29
14	Imperial Valley	SS	1940	7.0	El Centro Array #9	8.3	D	IMPVALL/I-ELC180	0.2	15
15	Landers	SS	1992	7.3	Joshua Tree	11.6	C	LANDERS/JOS000	0.07	23
16	Landers	SS	1992	7.3	North Palm Springs	24.2	D	LANDERS/NPS090	-	-
17	N Palm Springs	RV/OB	1986	6.0	Palm Springs Airport	16.6	D	PALMSPR/PSA000	0.2	50
18	Northridge	TH	1994	6.7	Sun Valley - Roscoe Blvd	12.3	D	NORTHR/RO3090	0.1	30
19	Loma Prieta	RV/OB	1989	6.9	Sunnyvale - Colton Ave.	28.8	D	LOMAP/SVL360	0.1	32
20	Landers	SS	1992	7.3	Yermo Fire Station	24.9	D	LANDERS/YER270	0.07	23
21	Borrego	-	1942	6.5	El Centro Array #9	49.0	D	BORREGO/B-ELC000	0.1	15
22	Coalinga	RV/OB	1983	6.4	Parkfield - Cholame 8W	50.7	D	COALINGA/H-C08000	0.2	23
23	Coalinga	RV/OB	1983	6.4	Parkfield - Gold Hill 1W	46.5	D	COALINGA/H-PG1090	0.2	22
24	Coalinga	RV/OB	1983	6.4	Parkfield - Fault Zone 3	36.4	D	COALINGA/H-COH000	0.1	27
25	Imperial Valley	SS	1979	6.5	Compuertas	32.6	D	IMPVALL/H-CMP015	0.2	-
26	Imperial Valley	SS	1979	6.5	Victoria	54.1	D	IMPVALL/H-VCT075	0.05	-
27	Northridge	TH	1994	6.7	Covina - W. Badillo	56.1	D	NORTHR/BAD000	0.2	30
28	Northridge	TH	1994	6.7	LA - Pico & Sentous	32.7	D	NORTHR/PIC180	0.2	46
29	Northridge	TH	1994	6.7	LA - E Vernon Ave	39.3	D	NORTHR/VER180	0.1	30
30	Coalinga	RV/OB	1983	5.8	Pleasant Valley P.P. - yard	17.4	D	COALINGA/D-PVY045	0.08	30
31	Coyote Lake	SS	1979	5.7	Gilroy Array #2	7.5	D	COYOTELK/G02050	0.2	40
32	Coyote Lake	SS	1979	5.7	Gilroy Array #3	6.0	D	COYOTELK/G03050	0.2	40
33	Coyote Lake	SS	1979	5.7	San Juan Bautista	15.6	D	COYOTELK/SJB213	0.2	20
34	Livermore	SS	1980	5.8	Antioch - 510 G St	20.3	D	LIVERMOR/A-ANT270	0.2	13
35	Whittier Narrows	RV	1987	6.0	Arcadia - Campus Dr	12.2	D	WHITTIER/A-CAM009	0.15	25
36	Whittier Narrows	RV	1987	6.0	LA - 116th St School	22.5	D	WHITTIER/A-116360	0.2	30
37	Whittier Narrows	RV	1987	6.0	Carson - Water St	24.5	D	WHITTIER/A-WAT180	0.2	25
38	Northridge	TH	1994	6.7	Northridge, Roscoe #1	13.7	-	*NR-nrr1/transverse	-	-
39	Northridge	TH	1994	6.7	Van Nuys, Sherman Circle #1	12.8	-	*NR-vnsc/longitudinal	-	-
40	Whittier Narrows	RV	1987	6.0	Cal Tech, Brown Athletic Building	16.6	-	*WH-athl/transverse	-	-

[1] Filenames preceded by an asterisk come from the Somerville ground motion set for the Van Nuys tesbed project (2001b).

Table A.2. Supplemental records that were added to the primary recordset of Table A.1 for cloud analysis of the Van Nuys Testbed building. These records were used in Chapter 7.

#	Event	Mechanism	Year	M_w	Station Name	Distance	USGS Soil Type	FileName[1]	HP	LP
1	Superstitn Hills	SS	1987	6.7	Wildlife Liquef. Array	24.4	D	SUPERST/B-IVW360	0.1	40
2	Loma Prieta	RV/OB	1989	6.9	Coyote Lake Dam (Downst)	22.3	D	LOMAP/CLD285	0.1	29
3	Northridge	TH	1994	6.7	Canoga Park - Topanga Can	15.8	D	NORTHR/CNP106	0.05	30
4	Loma Prieta	RV/OB	1989	6.9	Gilroy Array #4	16.1	D	LOMAP/G04000	0.2	28
5	Imperial Valley	SS	1979	6.5	Bonds Corner	2.5	D	IMPVALL/H-BCR230	0.1	40
6	Coalinga	RV/OB	1983	6.4	Cantua Creek School	25.5	D	COALINGA/H-CAK270	0.2	23
7	Imperial Valley	SS	1979	6.5	Calexico Fire Station	10.6	D	IMPVALL/H-CXO315	0.2	40
8	Imperial Valley	SS	1979	6.5	El Centro Array #12	18.2	D	IMPVALL/H-E12230	0.1	40
9	Northridge	TH	1994	6.7	LA - Hollywood Stor FF	25.5	D	NORTHR/HOL360	0.2	23
10	Imperial Valley	SS	1979	6.5	Cucapah	23.6	D	IMPVALL/H-QKP085	0.05	null
11	Imperial Valley	SS	1979	6.5	SAHOP Casa Flores	11.1	C	IMPVALL/H-SHP270	0.2	null
12	San Fernando	RV	1971	6.6	Lake Hughes #1	25.8	C	SFERN/L01111	0.5	35
13	San Fernando	RV	1971	6.6	Palmdale Fire Station	25.4	D	SFERN/PDL210	0.5	35
14	Cape Mendocino	-	1992	7.1	Rio Dell Overpass - FF	18.5	C	CAPEMEND/RIO270	0.07	23
15	Northridge	TH	1994	6.7	Sepulveda VA	8.9	D	NORTHR/SPV270	0.1	null
16	Loma Prieta	RV/OB	1989	6.9	WAHO	16.9	D	LOMAP/WAH090	0.1	null
17	Coalinga	RV/OB	1983	6.4	Parkfield - Cholame 12W	55.2	D	COALINGA/H-C12270	0.2	23
18	Coalinga	RV/OB	1983	6.4	Parkfield - Fault Zone 1	40.4	D	COALINGA/H-COW000	0.2	20
19	Imperial Valley	SS	1979	6.5	Coachella Canal #4	49.3	D	IMPVALL/H-CC4045	0.2	40
20	Imperial Valley	SS	1979	6.5	Delta	43.6	D	IMPVALL/H-DLT262	0.05	
21	Imperial Valley	SS	1979	6.5	Niland Fire Station	35.9	D	IMPVALL/H-NIL090	0.1	30
22	Northridge	TH	1994	6.7	Bell Gardens - Jaboneria	46.6	D	NORTHR/JAB310	0.13	30
23	Northridge	TH	1994	6.7	Lawndale - Osage Ave	42.4	D	NORTHR/LOA092	0.13	30
24	Northridge	TH	1994	6.7	LB - Rancho Los Cerritos	54.3	D	NORTHR/LBR000	0.16	23
25	Coyote Lake	SS	1979	5.7	Gilroy Array #4	4.5	D	COYOTELK/G04360	0.12	25
26	Point Mugu	RV	1973	5.8	Port Hueneme	25.0	D	PTMUGU/PHN180	0.2	25
27	Whittier Narrows	RV	1987	6.0	Compton - Castlegate St	16.9	D	WHITTIER/A-CAS000	0.09	25
28	Whittier Narrows	RV	1987	6.0	Carson - Catskill Ave	28.1	D	WHITTIER/A-CAT090	0.18	25
29	Northridge	TH	1994	6.7	Encino, Ventura Blvd. #1	17.7	-	*NR-env1/transverse	-	-
30	Northridge	TH	1994	6.7	Encino, Ventura Blvd. #9	17.9	-	*NR-env9/transverse	-	-
31	Northridge	TH	1994	6.7	N. Hollywood, Landershim Blvd #1	18.4	-	*NR-nhl2/longitudinal	-	-
32	Northridge	TH	1994	6.7	Van Nuys, Sherman Way #1	12.8	-	*NR-vns1/transverse	-	-
33	Northridge	TH	1994	6.7	Van Nuys, 7-story Hotel	11.3	-	*NR-vnuy/longitudinal	-	-
34	Northridge	TH	1994	6.7	Woodland Hills, Oxnard Street #4	20.0	-	*NR-whox/transverse	-	-
35	San Fernando	RV	1971	6.6	Los Angeles, 14724 Ventura Blvd	16.3	-	*SF-253/longitudinal	-	-
36	San Fernando	RV	1971	6.6	Los Angeles, 15910 Ventura Blvd	16.2	-	*SF-461/longitudinal	-	-
37	San Fernando	RV	1971	6.6	Los Angeles, 15250 Ventura Blvd	16.4	-	*SF-466/transverse	-	-
38	San Fernando	RV	1971	6.6	Glendale, Muni. Bldg. 633 E Brdwy	18.8	-	*SF-glen/transverse	-	-
39	San Fernando	RV	1971	6.6	Van Nuys, 7-story Hotel	9.5	-	*SF-vnuy/longitudinal	-	-

[1] Filenames preceded by an asterisk come from the Somerville ground motion set for the Van Nuys tesbed project (2001b).

For initial studies of ground motions intensity measures, it was desirable to have earthquake ground motions with a range of intensity measure values. Thus, as seen in Figure A.1, the records of Table A.1 and Table A.2 were selected to have a range of magnitude and distance values.

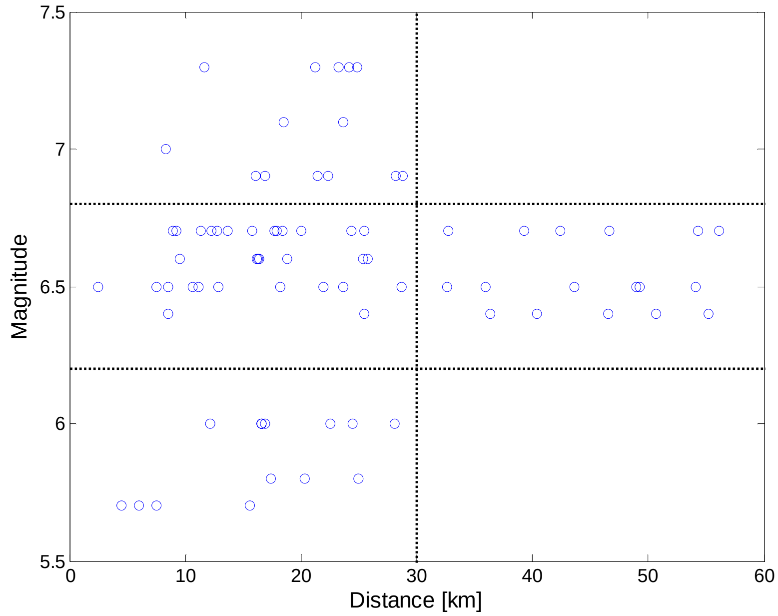


Figure A.1. Magnitudes and distances of the primary and secondary datasets (Table A.1 and Table A.2) used for analysis of the Van Nuys Testbed building.

Table A.3. The "LMSR-N" Record set from the work of Medina and Ibarra. These records were used in Chapters 3, 4 and 5.

#	Event	Mechanism	Year	M_w	Station Name	Dist.	USGS Soil Type	FileName	HP	LP
1	Imperial Valley	SS	1979	6.5	Calipatria Fire Station	23.8	C	IMPVALL\H-CAL315	0.1	40
2	Imperial Valley	SS	1979	6.5	Chihuahua	28.7	C	IMPVALL\H-CHI012	0.05	
3	Imperial Valley	SS	1979	6.5	Compuertas	32.6	C	IMPVALL\H-CMP015	0.2	
4	Imperial Valley	SS	1979	6.5	El Centro Array #1	15.5	C	IMPVALL\H-E01140	0.1	40
5	Imperial Valley	SS	1979	6.5	El Centro Array #12	18.2	C	IMPVALL\H-E12230	0.1	40
6	Imperial Valley	SS	1979	6.5	El Centro Array #13	21.9	C	IMPVALL\H-E13230	0.2	40
7	Imperial Valley	SS	1979	6.5	Niland Fire Station	35.9	C	IMPVALL\H-NIL090	0.1	30
8	Imperial Valley	SS	1979	6.5	Plaster City	31.7	C	IMPVALL\H-PLS135	0.1	40
9	Imperial Valley	SS	1979	6.5	Cucapah	23.6	C	IMPVALL\H-QKP085	0.05	
10	Imperial Valley	SS	1979	6.5	Westmorland Fire Sta	15.1	C	IMPVALL\H-WSM180	0.1	40
11	Loma Prieta	RV/OB	1989	6.9	Agnews State Hospital	28.2	C	LOMAP\AGW000	0.2	30
12	Loma Prieta	RV/OB	1989	6.9	Capitola	14.5	C	LOMAP\CAP090	0.2	40
13	Loma Prieta	RV/OB	1989	6.9	Gilroy Array #3	14.4	C	LOMAP\G03090	0.1	40
14	Loma Prieta	RV/OB	1989	6.9	Gilroy Array #4	16.1	C	LOMAP\G04090	0.2	30
15	Loma Prieta	RV/OB	1989	6.9	Gilroy Array #7	24.2	C	LOMAP\GMR000	0.2	40
16	Loma Prieta	RV/OB	1989	6.9	Hollister City Hall	28.2	C	LOMAP\HCH090	0.1	29
17	Loma Prieta	RV/OB	1989	6.9	Hollister Diff. Array	25.8	C	LOMAP\HDA255	0.1	33
18	Loma Prieta	RV/OB	1989	6.9	Halls Valley	31.6	C	LOMAP\HVR000	0.2	22
19	Loma Prieta	RV/OB	1989	6.9	Salinas - John & Work	32.6	C	LOMAP\SJW250	0.1	28
20	Loma Prieta	RV/OB	1989	6.9	Palo Alto - SLAC Lab	36.3	B	LOMAP\SLC270	0.2	33
21	Loma Prieta	RV/OB	1989	6.9	Sunnyvale - Colton Ave.	28.8	C	LOMAP\SVL270	0.1	40
22	Northridge	RV	1994	6.7	LA - Centinela St	30.9	C	NORTHR\CEN245	0.2	30
23	Northridge	RV	1994	6.7	Canoga Park - Topanga Can	15.8	C	NORTHR\CNP196	0.05	30
24	Northridge	RV	1994	6.7	LA - N Faring Rd	23.9	C	NORTHR\FAR000	0.13	30
25	Northridge	RV	1994	6.7	LA - Fletcher Dr	29.5	C	NORTHR\FLE234	0.15	30
26	Northridge	RV	1994	6.7	Glendale - Las Palmas	25.4	C	NORTHR\GLP267	0.1	30
27	Northridge	RV	1994	6.7	LA - Hollywood Stor FF #	25.5	C	NORTHR\PEL090	0.2	23
28	Northridge	RV	1994	6.7	Lake Hughes #1 #	36.3	B	NORTHR\L01000	0.12	23
29	Northridge	RV	1994	6.7	Leona Valley #2 #	37.7	C	NORTHR\LV2090	0.2	23
30	Northridge	RV	1994	6.7	Leona Valley #6 #	38.5	C	NORTHR\LV6090	0.2	23
31	Northridge	RV	1994	6.7	La Crescenta - New York	22.3	C	NORTHR\NYA180	0.1	30
32	Northridge	RV	1994	6.7	LA - Pico & Sentous #	32.7	C	NORTHR\PIC180	0.2	46
33	Northridge	RV	1994	6.7	Northridge - 17645 Saticoy St	13.3	C	NORTHR\STC090	0.1	30
34	Northridge	RV	1994	6.7	LA - Saturn St	30	C	NORTHR\STN020	0.1	30
35	Northridge	RV	1994	6.7	LA - E Vernon Ave	39.3	C	NORTHR\VER180	0.1	30
36	San Fernando	RV	1971	6.6	LA - Hollywood Stor FF	21.2	C	SFERN\PEL180	0.2	35
37	Superstitn Hills	SS	1987	6.7	Brawley Airport	18.2	C	SUPERST\B-BRA225	0.1	23
38	Superstitn Hills	SS	1987	6.7	El Centro Imp. Co. Cent	13.9	C	SUPERST\B-ICC000	0.1	40
39	Superstitn Hills	SS	1987	6.7	Plaster City	21	C	SUPERST\B-PLS135	0.2	18
40	Superstitn Hills	SS	1987	6.7	Westmorland Fire Sta	13.3	C	SUPERST\B-WSM090	0.1	35

Table A.4. Near-fault records used for Chapter 5. Note that these records came from the Next Generation Attenuation (NGA) ground motion database (2005) rather than the PEER ground motion database.

#	Location	Mechanism	Year	Mw	Station Name	Distance	GMX Soil Type
1	Parkfield	SS	1966	6.2	Cholame - Shandon Array #2	6.28	D
2	Parkfield	SS	1966	6.2	Temblor pre-1969	15.96	A
3	San Fernando	RV	1971	6.6	Pacoima Dam (upper left abut)	1.81	A
4	Gazli, USSR	RV	1976	6.8	Karakyr	5.46	A
5	Tabas, Iran	RV	1978	7.4	Tabas	2.05	A
6	Coyote Lake	SS	1979	5.7	Gilroy Array #6	3.11	A
7	Imperial Valley-06	SS	1979	6.5	Brawley Airport	10.42	D
8	Imperial Valley-06	SS	1979	6.5	EC County Center FF	7.31	D
9	Imperial Valley-06	SS	1979	6.5	EC Meloland Overpass FF	0.07	D
10	Imperial Valley-06	SS	1979	6.5	El Centro Array #10	6.17	D
11	Imperial Valley-06	SS	1979	6.5	El Centro Array #4	7.05	D
12	Imperial Valley-06	SS	1979	6.5	El Centro Array #5	3.95	D
13	Imperial Valley-06	SS	1979	6.5	El Centro Array #6	1.35	D
14	Imperial Valley-06	SS	1979	6.5	El Centro Array #7	0.56	D
15	Imperial Valley-06	SS	1979	6.5	El Centro Array #8	3.86	D
16	Imperial Valley-06	SS	1979	6.5	El Centro Differential Array	5.09	D
17	Imperial Valley-06	SS	1979	6.5	Westmorland Fire Sta	15.25	D
18	Coalinga-01	RV	1983	6.4	Pleasant Valley P.P. - bldg	8.41	D
19	Morgan Hill	SS	1984	6.2	Anderson Dam (Downstream)	3.26	D
20	Morgan Hill	SS	1984	6.2	Coyote Lake Dam (SW Abut)	0.53	A
21	Morgan Hill	SS	1984	6.2	Gilroy Array #6	9.86	A
22	Morgan Hill	SS	1984	6.2	Halls Valley	3.48	C
23	Nahanni, Canada	RV	1985	6.8	Site 1	9.60	A
24	Nahanni, Canada	RV	1985	6.8	Site 2	4.93	A
25	N. Palm Springs	RV/OB	1986	6.1	Desert Hot Springs	6.82	D
26	N. Palm Springs	RV/OB	1986	6.1	North Palm Springs	4.04	D
27	N. Palm Springs	RV/OB	1986	6.1	Whitewater Trout Farm	6.04	C
28	Whittier Narrows-01	RV/OB	1987	6.0	Bell Gardens - Jaboneria	17.79	D
29	Whittier Narrows-01	RV/OB	1987	6.0	Downey - Co Maint Bldg	20.82	D
30	Whittier Narrows-01	RV/OB	1987	6.0	Norwalk - Imp Hwy, S Grnd	20.42	D
31	Whittier Narrows-01	RV/OB	1987	6.0	Santa Fe Springs - E.Joslin	18.49	D
32	Superstition Hills-02	SS	1987	6.5	El Centro Imp. Co. Cent	18.20	D
33	Superstition Hills-02	SS	1987	6.5	Parachute Test Site	0.95	D
34	Loma Prieta	RV/OB	1989	6.9	Gilroy - Gavilan Coll.	9.96	B
35	Loma Prieta	RV/OB	1989	6.9	Gilroy - Historic Bldg.	10.97	D
36	Loma Prieta	RV/OB	1989	6.9	Gilroy Array #1	9.64	A
37	Loma Prieta	RV/OB	1989	6.9	Gilroy Array #2	11.07	D
38	Loma Prieta	RV/OB	1989	6.9	Gilroy Array #3	12.82	D
39	Loma Prieta	RV/OB	1989	6.9	Gilroy Array #4	14.34	D
40	Loma Prieta	RV/OB	1989	6.9	LGPC	3.88	A

Table A.4 (continued). Near-fault records used for Chapter 5.

#	FileName1	HP	LP	T_p[2]	X[3]	θ[3]	$X\cos(\theta)$[3]	Y[3]	ϕ[3]	$Y\cos(\phi)$[3]
1	PARKF/C02065.at2	0.20	10	0.67	1	4.74	1.00	0.83	4.05	
2	PARKF/TMB_051_FN.at2	0.20	15	0.4	1	10.43	0.98	0.83	34.39	
3	SFERN/PUL_195_FN.at2	0.10	35	1.34	0.25	63.42		0.8	7.53	0.79
4	GAZLI/GAZ_177_FN.at2	0.05	38	1.06	0	89.96		1	3.15	
5	TABAS/TAB-TR.at2	0.05		4.7	0.6	12.56		0.32	0.58	0.32
6	COYOTELK/G06_246_FN.at2	0.08	25	0.91	0.62	16.99	0.59	0.71	0.93	
7	IMPVALL/H-BRA_233_FN.at2	0.10	40	4.8	0.76	10.54	0.75	0.65	48.4	
8	IMPVALL/H-ECC_233_FN.at2	0.10	40	3.7	0.55	18.15	0.52	0.78	32.28	
9	IMPVALL/H-EMO_233_FN.at2	0.10	40	3	0.39	5.37	0.39	0.78	0.34	
10	IMPVALL/H-E10_233_FN.at2	0.10	40	6.1	0.5	17.53	0.48	0.78	28.5	
11	IMPVALL/H-E04_233_FN.at2	0.10	40	3.7	0.53	11.49	0.52	0.68	38.49	
12	IMPVALL/H-E05_233_FN.at2	0.10	40	3.4	0.55	4.68	0.55	0.72	22.82	
13	IMPVALL/H-E06_233_FN.at2	0.10	40	3.3	0.55	0.78	0.55	0.76	7.85	
14	IMPVALL/H-E07_233_FN.at2	0.10	40	3.2	0.55	4.8	0.55	0.78	3.07	
15	IMPVALL/H-E08_233_FN.at2	0.10	40	4.2	0.55	11.52	0.54	0.78	19.39	
16	IMPVALL/H-EDA_233_FN.at2	0.10	40	3.7	0.53	14.55	0.51	0.78	24.48	
17	IMPVALL/H-WSM_233_FN.at2	0.10	40	4.6	0.76	3.04	0.76	0.72	25.71	
18	COALINGA/H-PVB_047_FN.at2	0.20	20	1.1	0.15	76.24		0.21	4.61	0.21
19	MORGAN/AND_058_FN.at2	0.10	30	0.45	0.61	11.12	0.60	0.7	20.73	
20	MORGAN/CYC_058_FN.at2	0.10	39	0.75	0.91	0.4	0.91	0.7	1.15	
21	MORGAN/G06_058_FN.at2	0.10	35	1.15	0.98	0.95	0.98	0.7	4.07	
22	MORGAN/HVR_058_FN.at2	0.20	26	0.84	0.02	6.74	0.02	0.7	3.11	
23	NAHANNI/S1_070_FN.at2	0.05	63	3.4	0.11	55.73		0.1	79.91	0.02
24	NAHANNI/S2_070_FN.at2	0.10	63	0.55	0.11	56.61		0.47	30.77	0.40
25	PALMSPR/DSP_197_FN.at2	0.50	46	0.42	0.47	7.01		0.66	37.43	
26	PALMSPR/NPS_197_FN.at2	0.15	20	0.91	0.43	36.12		0.73	14.47	0.71
27	PALMSPR/WWT_197_FN.at2	0.10	40	0.53	0.18	32.22		0.71	32.4	0.60
28	WHITTIER/A-JAB_190_FN.at2	0.25	25	0.62	0.5	62.84		0.03	24.36	0.03
29	WHITTIER/A-DWN_190_FN.at2	0.20	30	0.81	0.5	70.33		0.03	14.02	0.03
30	WHITTIER/A-NOR_190_FN.at2	0.15	40	0.6	0.36	76.12		0.03	15.12	0.03
31	WHITTIER/A-EJS_190_FN.at2	0.35	25	0.26	0.15	82.71		0.03	21.45	0.03
32	SUPERST/B-ICC_037_FN.at2	0.10	40	1.5	0.9	8.31	0.89	0.75	29.91	
33	SUPERST/B-PTS_037_FN.at2	0.06	20	1.9	0.8	3.4	0.80	0.75	6.01	
34	LOMAP/GIL_038_FN.at2	0.20	45	0.39	0.5	23.03		0.81	12.96	
35	LOMAP/GOF_038_FN.at2	0.20	38	1.29	0.5	30.89		0.81	19.54	0.76
36	LOMAP/G01_038_FN.at2	0.20	50	0.4	0.5	22.84		0.81	12.45	
37	LOMAP/G02_038_FN.at2	0.20	40	1.56	0.5	25.42		0.81	16.18	
38	LOMAP/G03_038_FN.at2	0.10	33	0.47	0.5	27.13		0.81	19.32	
39	LOMAP/G04_038_FN.at2	0.20	28	1.5	0.5	30.26		0.81	23.02	
40	LOMAP/LGP_038_FN.at2	0.10	80	0.75	0.45	14.37		0.81	5.32	0.81

Table A.4 (continued). Near-fault records used for Chapter 5.

#	Location	Mechanism	Year	Mw	Station Name	Distance	GMX Soil Type
41	Loma Prieta	RV/OB	1989	6.9	Saratoga - Aloha Ave	8.50	D
42	Loma Prieta	RV/OB	1989	6.9	Saratoga - W Valley Coll.	9.31	D
43	Sierra Madre	RV	1991	5.6	Cogswell Dam - Right Abutment	22.00	A
44	Erzican, Turkey	SS	1992	6.7	Erzincan	4.38	D
45	Landers	SS	1992	7.3	Lucerne	2.19	A
46	Northridge-01	RV	1994	6.7	Canoga Park - Topanga Can	14.70	D
47	Northridge-01	RV	1994	6.7	Canyon Country - W Lost Cany	12.44	C
48	Northridge-01	RV	1994	6.7	Jensen Filter Plant	5.43	B
49	Northridge-01	RV	1994	6.7	LA - Sepulveda VA Hospital	8.44	D
50	Northridge-01	RV	1994	6.7	LA Dam	5.92	A
51	Northridge-01	RV	1994	6.7	Newhall - Fire Sta	5.92	D
52	Northridge-01	RV	1994	6.7	Newhall - W Pico Canyon Rd.	5.48	B
53	Northridge-01	RV	1994	6.7	Pacoima Dam (downstr)	7.01	A
54	Northridge-01	RV	1994	6.7	Pacoima Kagel Canyon	7.26	B
55	Northridge-01	RV	1994	6.7	Rinaldi Receiving Sta	6.50	B
56	Northridge-01	RV	1994	6.7	Sylmar - Converter Sta	5.35	D
57	Northridge-01	RV	1994	6.7	Sylmar - Converter Sta East	5.19	B
58	Northridge-01	RV	1994	6.7	Sylmar - Olive View Med FF	5.30	D
59	Kobe, Japan	SS	1995	6.9	KJMA	0.96	B
60	Kocaeli, Turkey	SS	1999	7.5	Arcelik	13.49	B
61	Kocaeli, Turkey	SS	1999	7.5	Duzce	15.37	D
62	Kocaeli, Turkey	SS	1999	7.5	Gebze	10.92	A
63	Kocaeli, Turkey	SS	1999	7.5	Sakarya	3.12	B
64	Kocaeli, Turkey	SS	1999	7.5	Yarimca	4.83	D
65	Chi-Chi, Taiwan	RV/OB	1999	7.6	TCU052	0.66	A
66	Chi-Chi, Taiwan	RV/OB	1999	7.6	TCU065	0.59	D
67	Chi-Chi, Taiwan	RV/OB	1999	7.6	TCU068	0.32	A
68	Chi-Chi, Taiwan	RV/OB	1999	7.6	TCU075	0.91	D
69	Chi-Chi, Taiwan	RV/OB	1999	7.6	TCU076	2.76	D
70	Chi-Chi, Taiwan	RV/OB	1999	7.6	TCU129	1.84	D

Table A.4 (continued). Near-fault records used for Chapter 5.

#	FileName1	HP	LP	T_p[2]	X[3]	θ[3]	Xcos(θ)[3]	Y[3]	ϕ[3]	Ycos(ϕ)[3]
41	LOMAP/STG_038_FN.at2	0.10	38	1.69	0.5	23.2		0.81	11.53	
42	LOMAP/WVC_038_FN.at2	0.10	38	1.15	0.5	27.61		0.81	15.65	0.78
43	SMADRE/chan1_152_FN.at2	0.50	23	0.29	0.13	51.57		0.07	89.22	
44	ERZIKAN/ERZ_032_FN.at2	0.10		2.3	0.31	1.95	0.31	0.75	25.06	
45	LANDERS/LCN_239_FN.at2	0.08	60	4.5	0.66	20.24	0.62	0.47	8.32	
46	NORTHR/CNP_032_FN.at2	0.05	30	2.05	0.25	23.66		0.41	56.34	0.23
47	NORTHR/LOS_032_FN.at2	0.05	30	0.67	0.15	84.31		0.81	6.42	0.80
48	NORTHR/JEN_032_FN.at2	0.08		2.9	0.1	82.22		0.81	13.7	0.79
49	NORTHR/0637_032_FN.at2	0.08	50	0.85	0.19	66.62		0.72	26.01	0.65
50	NORTHR/LDM_032_FN.at2	0.10		1.34	0.02	88.45		0.81	16.04	0.78
51	NORTHR/NWH_032_FN.at2	0.12	23	1.25	0.5	63.44		0.81	3.99	0.81
52	NORTHR/WPI_032_FN.at2	0.05	30	2.3	0.78	41.13		0.81	10.99	
53	NORTHR/PAC_032_FN.at2	0.16	23	0.44	0.22	76.56		0.81	1.46	0.81
54	NORTHR/PKC_032_FN.at2	0.14	23	0.88	0.22	63.46		0.81	5.41	
55	NORTHR/RRS_032_FN.at2	0.06	30	1.06	0.08	82.22		0.81	18.3	0.77
56	NORTHR/SCS_032_FN.at2			3	0.07	84.62		0.81	13.29	0.79
57	NORTHR/SCE_032_FN.at2			3.1	0.03	87.53		0.81	12.18	0.79
58	NORTHR/SYL_032_FN.at2	0.12	23	2.58	0.08	84.98		0.81	6.32	0.81
59	KOBE/KJM_140_FN.at2	0.05		0.85	0.3	7.83	0.30	0.87	13.16	
60	KOCAELI/ARC_184_FN.at2	0.07	50	9.2	0.35	19.92	0.33	0.64	40.92	
61	KOCAELI/DZC_163_FN.at2		20	3.8	0.65	17.4	0.62	0.55	14.72	
62	KOCAELI/GBZ_184_FN.at2	0.03	25	4.3	0.34	23.87	0.31	0.64	37.34	
63	KOCAELI/SKR090.at2			5.7	0.24	3.56	0.24	0.73	25.32	
64	KOCAELI/YPT_180_FN.at2	0.07	50	3.8	0.14	13.92	0.14	0.64	4.57	
65	CHICHI/TCU052_278_FN.at2	0.04	50	6.7	0.43	34.77		0.39	6.59	0.39
66	CHICHI/TCU065_272_FN.at2	0.03	50	4.5	0.22	46.3		0.38	5.99	0.38
67	CHICHI/TCU068_280_FN.at2	0.03	50	9	0.48	33.32		0.39	7.01	0.39
68	CHICHI/TCU075_271_FN.at2	0.03	50	4.4	0.15	54.08		0.36	4.05	0.36
69	CHICHI/TCU076_271_FN.at2	0.10	50	3.2	0.06	63.27		0.32	0.05	0.32
70	CHICHI/TCU129_271_FN.at2	0.03	50	4.1	0.05	73.62		0.29	3.75	0.29

[2] Pulse period as measured by Tothong and Cornell (2005a).

[3] Near-fault ground motion parameters as specified in Somerville et al. (1997). Blanks denote that the parameter is not applicable (e.g., Xcos(θ) is only provided for strike-slip events within the distance range where directivity is suggested).

Table A.5. Expanded record set used for calculation of response spectral correlations in Chapter 8. These records were also used as the library of records in Chapter 6.

#	Event	Mechanism	Year	M_w	Station Name	Distance	USGS Soil Type	GM Soil Type
1	Cape Mendocino	RV	1992	7.1	Eureka - Myrtle & West	44.6	B	D
2	Cape Mendocino	RV	1992	7.1	Fortuna - Fortuna Blvd	23.6	B	D
3	Cape Mendocino	RV	1992	7.1	Petrolia	9.5	C	D
4	Cape Mendocino	RV	1992	7.1	Rio Dell Overpass - FF	18.5	B	C
5	Chalfant Valley	SS	1986	5.9	Bishop - LADWP South St	24	-	D
6	Chalfant Valley	SS	1986	6.2	Bishop - LADWP South St	9.2	-	D
7	Chalfant Valley	SS	1986	5.8	Bishop - LADWP South St	13	-	D
8	Chalfant Valley	SS	1986	6.2	Convict Creek	44.9	-	D
9	Chalfant Valley	SS	1986	6.2	McGee Creek Surface	36.3	-	C
10	Chalfant Valley	SS	1986	5.9	Zack Brothers Ranch	11	-	D
11	Chalfant Valley	SS	1986	6.2	Zack Brothers Ranch	18.7	-	D
12	Coalinga	RV/OB	1983	6.4	Cantua Creek School	25.5	-	D
13	Coalinga	RV	1983	5.8	Palmer Ave	12.2	-	B
14	Coalinga	RV/OB	1983	6.4	Parkfield - Cholame 12W	55.2	-	D
15	Coalinga	RV/OB	1983	6.4	Parkfield - Cholame 2WA	42.8	-	D
16	Coalinga	RV/OB	1983	6.4	Parkfield - Cholame 3W	43.9	-	C
17	Coalinga	RV/OB	1983	6.4	Parkfield - Cholame 4W	44.7	-	C
18	Coalinga	RV/OB	1983	6.4	Parkfield - Fault Zone 1	40.4	-	D
19	Coalinga	RV/OB	1983	6.4	Parkfield - Fault Zone 10	30.4	-	D
20	Coalinga	RV/OB	1983	6.4	Parkfield - Fault Zone 11	28.4	-	B
21	Coalinga	RV/OB	1983	6.4	Parkfield - Fault Zone 12	29.5	-	C
22	Coalinga	RV/OB	1983	6.4	Parkfield - Fault Zone 14	29.9	-	C
23	Coalinga	RV/OB	1983	6.4	Parkfield - Fault Zone 15	29.9	-	B
24	Coalinga	RV/OB	1983	6.4	Parkfield - Fault Zone 16	28.1	-	C
25	Coalinga	RV/OB	1983	6.4	Parkfield - Fault Zone 3	36.4	-	D
26	Coalinga	RV/OB	1983	6.4	Parkfield - Fault Zone 4	34.3	-	B
27	Coalinga	RV/OB	1983	6.4	Parkfield - Fault Zone 6	32.8	-	B
28	Coalinga	RV/OB	1983	6.4	Parkfield - Fault Zone 7	31	-	C
29	Coalinga	RV/OB	1983	6.4	Parkfield - Fault Zone 8	29.6	-	B
30	Coalinga	RV/OB	1983	6.4	Parkfield - Fault Zone 9	31.9	-	B
31	Coalinga	RV/OB	1983	6.4	Parkfield - Gold Hill 1W	46.5	-	D
32	Coalinga	RV/OB	1983	6.4	Parkfield - Gold Hill 2E	32.3	-	D
33	Coalinga	RV/OB	1983	6.4	Parkfield - Gold Hill 2W	36.6	-	B
34	Coalinga	RV/OB	1983	6.4	Parkfield - Gold Hill 3E	29.2	-	D
35	Coalinga	RV/OB	1983	6.4	Parkfield - Gold Hill 3W	38.8	-	B
36	Coalinga	RV/OB	1983	6.4	Parkfield - Gold Hill 4W	41	-	B
37	Coalinga	RV/OB	1983	6.4	Parkfield - Gold Hill 5W	43.7	-	B
38	Coalinga	RV/OB	1983	6.4	Parkfield - Gold Hill 6W	48	-	C
39	Coalinga	RV/OB	1983	6.4	Parkfield - Vineyard Cany 1E	26.7	-	C
40	Coalinga	RV/OB	1983	6.4	Parkfield - Vineyard Cany 2W	30.7	-	C

Table A.5 (continued). Expanded record set used for calculation of response spectral correlations in Chapter 8. These records were also used as the library of records in Chapter 6.

#	FileName (Vertical)	Filename (Horiz. #1)	Filename (Horiz. #2)	HP[4]	LP[4]
1	CAPEMEND\EUR-UP	CAPEMEND\EUR000	CAPEMEND\EUR090	0.16	23
2	CAPEMEND\FOR-UP	CAPEMEND\FOR000	CAPEMEND\FOR090	0.07	23
3	CAPEMEND\PET-UP	CAPEMEND\PET000	CAPEMEND\PET090	0.07	23
4	CAPEMEND\RIO-UP	CAPEMEND\RIO270	CAPEMEND\RIO360	0.07	23
5	CHALFANT\B-LAD-UP	CHALFANT\B-LAD270	CHALFANT\B-LAD180	0.11	20
6	CHALFANT\A-LAD-UP	CHALFANT\A-LAD180	CHALFANT\A-LAD270	0.1	30
7	CHALFANT\D-LAD-UP	CHALFANT\D-LAD160	CHALFANT\D-LAD070	0.2	20
8	CHALFANT\A-CVK-UP	CHALFANT\A-CVK090	CHALFANT\A-CVK000	0.2	30
9	CHALFANT\A-MCG-UP	CHALFANT\A-MCG270	CHALFANT\A-MCG360	0.1	35
10	CHALFANT\B-ZAK-UP	CHALFANT\B-ZAK270	CHALFANT\B-ZAK360	0.11	30
11	CHALFANT\A-ZAK-UP	CHALFANT\A-ZAK270	CHALFANT\A-ZAK360	0.2	33
12	COALINGA\H-CAK-UP	COALINGA\H-CAK270	COALINGA\H-CAK360	0.2	23
13	COALINGA\D-PLM-UP	COALINGA\D-PLM360	COALINGA\D-PLM270	0.2	20
14	COALINGA\H-C12-UP	COALINGA\H-C12270	COALINGA\H-C12360	0.2	21
15	COALINGA\H-C02-UP	COALINGA\H-C02000	COALINGA\H-C02090	0.2	22
16	COALINGA\H-C03-UP	COALINGA\H-C03000	COALINGA\H-C03090	0.2	21
17	COALINGA\H-C04-UP	COALINGA\H-C04000	COALINGA\H-C04090	0.2	21
18	COALINGA\H-COW-UP	COALINGA\H-COW000	COALINGA\H-COW090	0.2	20
19	COALINGA\H-Z10-UP	COALINGA\H-Z10000	COALINGA\H-Z10090	0.2	21
20	COALINGA\H-Z11-UP	COALINGA\H-Z11000	COALINGA\H-Z11090	0.2	21
21	COALINGA\H-PRK-UP	COALINGA\H-PRK090	COALINGA\H-PRK180	0.2	20
22	COALINGA\H-Z14-UP	COALINGA\H-Z14090	COALINGA\H-Z14000	0.2	23
23	COALINGA\H-Z15-UP	COALINGA\H-Z15000	COALINGA\H-Z15090	0.2	20
24	COALINGA\H-Z16-UP	COALINGA\H-Z16000	COALINGA\H-Z16090	0.2	26
25	COALINGA\H-COH-UP	COALINGA\H-COH000	COALINGA\H-COH090	0.1	22
26	COALINGA\H-Z04-UP	COALINGA\H-Z04000	COALINGA\H-Z04090	0.2	22
27	COALINGA\H-Z06-UP	COALINGA\H-Z06000	COALINGA\H-Z06090	0.2	24
28	COALINGA\H-Z07-UP	COALINGA\H-Z07000	COALINGA\H-Z07090	0.2	30
29	COALINGA\H-Z08-UP	COALINGA\H-Z08000	COALINGA\H-Z08090	0.2	21
30	COALINGA\H-Z09-UP	COALINGA\H-Z09000	COALINGA\H-Z09090	0.2	23
31	COALINGA\H-PG1-UP	COALINGA\H-PG1000	COALINGA\H-PG1090	0.2	22
32	COALINGA\H-GH2-UP	COALINGA\H-GH2000	COALINGA\H-GH2090	0.2	30
33	COALINGA\H-PG2-UP	COALINGA\H-PG2000	COALINGA\H-PG2090	0.2	20
34	COALINGA\H-GH3-UP	COALINGA\H-GH3000	COALINGA\H-GH3090	0.2	26
35	COALINGA\H-PG3-UP	COALINGA\H-PG3000	COALINGA\H-PG3090	0.2	30
36	COALINGA\H-PG4-UP	COALINGA\H-PG4000	COALINGA\H-PG4090	0.2	30
37	COALINGA\H-PG5-UP	COALINGA\H-PG5000	COALINGA\H-PG5090	0.2	26
38	COALINGA\H-PG6-UP	COALINGA\H-PG6000	COALINGA\H-PG6090	0.2	30
39	COALINGA\H-PV1-UP	COALINGA\H-PV1000	COALINGA\H-PV1090	0.2	23
40	COALINGA\H-VC2-UP	COALINGA\H-VC2000	COALINGA\H-VC2090	0.2	30

[4] Maximum high-pass frequency among the three components, and minimum low-pass frequency among the three components.

Table A.5 (continued). Expanded record set used for calculation of response spectral correlations in Chapter 8. These records were also used as the library of records in Chapter 6.

#	Event	Mechanism	Year	M_w	Station Name	Distance	USGS Soil Type	GM Soil Type
41	Coalinga	RV/OB	1983	6.4	Parkfield - Vineyard Cany 4W	34.6	-	B
42	Coalinga	RV/OB	1983	6.4	Parkfield - Vineyard Cany 6W	41	-	C
43	Coalinga	RV/OB	1983	6.4	Pleasant Valley P.P. - bldg	8.5	-	D
44	Coalinga	RV	1983	5.8	Pleasant Valley P.P. - FF	17.4	-	D
45	Coalinga	RV/OB	1983	6.4	Pleasant Valley P.P. - yard	8.5	-	D
46	Coyote Lake	SS	1979	5.7	Gilroy Array #2	7.5	C	D
47	Coyote Lake	SS	1979	5.7	Gilroy Array #6	3.1	B	B
48	Coyote Lake	SS	1979	5.7	San Juan Bautista, 24 Polk St	15.6	B	D
49	Duzce, Turkey	SS	1999	7.1	Duzce	8.2	C	D
50	Duzce, Turkey	SS	1999	7.1	Lamont 1058	0.9	B	B
51	Duzce, Turkey	SS	1999	7.1	Lamont 1059	8.5	-	B
52	Duzce, Turkey	SS	1999	7.1	Lamont 1061	15.6	B	B
53	Duzce, Turkey	SS	1999	7.1	Lamont 1062	13.3	-	B
54	Duzce, Turkey	SS	1999	7.1	Lamont 375	8.2	-	B
55	Duzce, Turkey	SS	1999	7.1	Sakarya	49.9	B	B
56	Friuli, Italy	-	1976	6.5	Barcis	49.7	-	B
57	Friuli, Italy	-	1976	6.5	Codroipo	34.6	-	D
58	Friuli, Italy	-	1976	6.5	Tolmezzo	37.7	-	B
59	Imperial Valley	SS	1979	6.5	Bonds Corner	2.5	C	D
60	Imperial Valley	SS	1979	6.5	Brawley Airport	8.5	C	D
61	Imperial Valley	SS	1979	6.5	Calexico Fire Station	10.6	C	D
62	Imperial Valley	SS	1979	6.5	Calipatria Fire Station	23.8	C	D
63	Imperial Valley	SS	1979	6.5	Coachella Canal #4	49.3	C	D
64	Imperial Valley	SS	1979	6.5	EC County Center FF	7.6	C	D
65	Imperial Valley	SS	1979	6.5	EC Meloland Overpass FF	0.5	C	D
66	Imperial Valley	SS	1979	6.5	El Centro Array #1	15.5	C	D
67	Imperial Valley	SS	1979	6.5	El Centro Array #10	8.6	C	D
68	Imperial Valley	SS	1979	6.5	El Centro Array #11	12.6	C	D
69	Imperial Valley	SS	1979	6.5	El Centro Array #12	18.2	C	D
70	Imperial Valley	SS	1979	6.5	El Centro Array #13	21.9	C	D
71	Imperial Valley	SS	1979	6.5	El Centro Array #4	4.2	C	D
72	Imperial Valley	SS	1979	6.5	El Centro Array #5	1	C	D
73	Imperial Valley	SS	1979	6.5	El Centro Array #6	1	C	D
74	Imperial Valley	SS	1979	6.5	El Centro Array #7	0.6	C	D
75	Imperial Valley	SS	1979	6.5	El Centro Array #8	3.8	C	D
76	Imperial Valley	SS	1979	6.5	El Centro Differential Array	5.3	C	D
77	Imperial Valley	SS	1979	6.5	Holtville Post Office	7.5	C	D
78	Imperial Valley	SS	1979	6.5	Niland Fire Station	35.9	C	D
79	Imperial Valley	SS	1979	6.5	Parachute Test Site	14.2	B	D
80	Imperial Valley	SS	1979	6.5	Plaster City	31.7	C	D

Table A.5 (continued). Expanded record set used for calculation of response spectral correlations in Chapter 8. These records were also used as the library of records in Chapter 6.

#	FileName (Vertical)	Filename (Horiz. #1)	Filename (Horiz. #2)	HP[4]	LP[4]
41	COALINGA\H-VC4-UP	COALINGA\H-VC4000	COALINGA\H-VC4090	0.2	27
42	COALINGA\H-VC6-UP	COALINGA\H-VC6000	COALINGA\H-VC6090	0.2	25
43	COALINGA\H-PVB-UP	COALINGA\H-PVB045	COALINGA\H-PVB135	0.2	20
44	COALINGA\D-PVP-UP	COALINGA\D-PVP270	COALINGA\D-PVP360	0.1	30
45	COALINGA\H-PVY-UP	COALINGA\H-PVY045	COALINGA\H-PVY135	0.2	31
46	COYOTELK\G02-UP	COYOTELK\G02050	COYOTELK\G02140	0.2	40
47	COYOTELK\G06-UP	COYOTELK\G06230	COYOTELK\G06320	0.2	25
48	COYOTELK\SJB-UP	COYOTELK\SJB213	COYOTELK\SJB303	0.2	20
49	DUZCE\DZC-UP	DUZCE\DZC180	DUZCE\DZC270	0.08	50
50	DUZCE\1058-V	DUZCE\1058-N	DUZCE\1058-E	0.06	50
51	DUZCE\1059-V	DUZCE\1059-N	DUZCE\1059-E	0.06	50
52	DUZCE\1061-V	DUZCE\1061-N	DUZCE\1061-E	0.07	50
53	DUZCE\1062-V	DUZCE\1062-N	DUZCE\1062-E	0.06	50
54	DUZCE\375-V	DUZCE\375-N	DUZCE\375-E	0.15	50
55	DUZCE\SKR-UP	DUZCE\SKR180	DUZCE\SKR090	0.05	40
56	FRIULI\A-BCS-UP	FRIULI\A-BCS000	FRIULI\A-BCS270	0.2	30
57	FRIULI\A-COD-UP	FRIULI\A-COD000	FRIULI\A-COD270	0.1	25
58	FRIULI\A-TMZ-UP	FRIULI\A-TMZ000	FRIULI\A-TMZ270	0.1	30
59	IMPVALL\H-BCR-UP	IMPVALL\H-BCR140	IMPVALL\H-BCR230	0.1	40
60	IMPVALL\H-BRA-UP	IMPVALL\H-BRA225	IMPVALL\H-BRA315	0.1	40
61	IMPVALL\H-CXO-UP	IMPVALL\H-CXO225	IMPVALL\H-CXO315	0.2	40
62	IMPVALL\H-CAL-UP	IMPVALL\H-CAL225	IMPVALL\H-CAL315	0.1	40
63	IMPVALL\H-CC4-UP	IMPVALL\H-CC4045	IMPVALL\H-CC4135	0.2	40
64	IMPVALL\H-ECC-UP	IMPVALL\H-ECC002	IMPVALL\H-ECC092	0.1	35
65	IMPVALL\H-EMO-UP	IMPVALL\H-EMO000	IMPVALL\H-EMO270	0.1	40
66	IMPVALL\H-E01-UP	IMPVALL\H-E01140	IMPVALL\H-E01230	0.1	40
67	IMPVALL\H-E10-UP	IMPVALL\H-E10050	IMPVALL\H-E10320	0.1	40
68	IMPVALL\H-E11-UP	IMPVALL\H-E11230	IMPVALL\H-E11140	0.2	40
69	IMPVALL\H-E12-UP	IMPVALL\H-E12140	IMPVALL\H-E12230	0.1	40
70	IMPVALL\H-E13-UP	IMPVALL\H-E13140	IMPVALL\H-E13230	0.2	40
71	IMPVALL\H-E04-UP	IMPVALL\H-E04140	IMPVALL\H-E04230	0.1	40
72	IMPVALL\H-E05-UP	IMPVALL\H-E05140	IMPVALL\H-E05230	0.1	40
73	IMPVALL\H-E06-UP	IMPVALL\H-E06140	IMPVALL\H-E06230	0.2	40
74	IMPVALL\H-E07-UP	IMPVALL\H-E07140	IMPVALL\H-E07230	0.1	40
75	IMPVALL\H-E08-UP	IMPVALL\H-E08140	IMPVALL\H-E08230	0.1	40
76	IMPVALL\H-EDA-UP	IMPVALL\H-EDA270	IMPVALL\H-EDA360	0.1	40
77	IMPVALL\H-HVP-UP	IMPVALL\H-HVP225	IMPVALL\H-HVP315	0.1	40
78	IMPVALL\H-NIL-UP	IMPVALL\H-NIL090	IMPVALL\H-NIL360	0.1	30
79	IMPVALL\H-PTS-UP	IMPVALL\H-PTS225	IMPVALL\H-PTS315	0.1	40
80	IMPVALL\H-PLS-UP	IMPVALL\H-PLS045	IMPVALL\H-PLS135	0.1	40

[4] Maximum high-pass frequency among the three components, and minimum low-pass frequency among the three components.

Table A.5 (continued). Expanded record set used for calculation of response spectral correlations in Chapter 8. These records were also used as the library of records in Chapter 6.

#	Event	Mechanism	Year	M_w	Station Name	Distance	USGS Soil Type	GM Soil Type
81	Imperial Valley	SS	1979	6.5	Superstition Mtn Camera	26	B	A
82	Imperial Valley	SS	1979	6.5	Westmorland Fire Sta	15.1	C	D
83	Kobe	SS	1995	6.9	Abeno	23.8	C	D
84	Kobe	SS	1995	6.9	Tadoka	30.5	C	D
85	Kocaeli, Turkey	SS	1999	7.4	Arcelik	17	B	B
86	Kocaeli, Turkey	SS	1999	7.4	Atakoy	67.5	C	D
87	Kocaeli, Turkey	SS	1999	7.4	Bursa Tofas	62.7	-	D
88	Kocaeli, Turkey	SS	1999	7.4	Cekmece	76.1	C	B
89	Kocaeli, Turkey	SS	1999	7.4	Fatih	64.5	-	C
90	Kocaeli, Turkey	SS	1999	7.4	Goynuk	35.5	-	B
91	Kocaeli, Turkey	SS	1999	7.4	Iznik	31.8	C	D
92	Kocaeli, Turkey	SS	1999	7.4	Mecidiyekoy	62.3	B	B
93	Kocaeli, Turkey	SS	1999	7.4	Zeytinburnu	63.1	C	D
94	Landers	SS	1992	7.3	Baker Fire Station	88.5	B	D
95	Landers	SS	1992	7.3	Barstow	36.1	B	D
96	Landers	SS	1992	7.3	Boron Fire Station	90.6	B	D
97	Landers	SS	1992	7.3	Coolwater	21.2	B	D
98	Landers	SS	1992	7.3	Desert Hot Springs	23.2	B	D
99	Landers	SS	1992	7.3	Fort Irwin	64.2	B	D
100	Landers	SS	1992	7.3	Hemet Fire Station	69.5	C	D
101	Landers	SS	1992	7.3	Indio - Coachella Canal	55.7	C	D
102	Landers	SS	1992	7.3	Joshua Tree	11.6	B	C
103	Landers	SS	1992	7.3	Palm Springs Airport	37.5	C	D
104	Landers	SS	1992	7.3	Riverside Airport	96.1	B	B
105	Landers	SS	1992	7.3	San Bernardino-E &Hospitality	80.5	C	D
106	Landers	SS	1992	7.3	Yermo Fire Station	24.9	C	D
107	Loma Prieta	RV/OB	1989	6.9	Agnews State Hospital	28.2	C	D
108	Loma Prieta	RV/OB	1989	6.9	Anderson Dam (Downstream)	21.4	B	D
109	Loma Prieta	RV/OB	1989	6.9	Anderson Dam (L Abut)	21.4	B	A
110	Loma Prieta	RV/OB	1989	6.9	APEEL 10 - Skyline	47.8	B	A
111	Loma Prieta	RV/OB	1989	6.9	APEEL 2E Hayward Muir Sch	57.4	C	D
112	Loma Prieta	RV/OB	1989	6.9	APEEL 3E Hayward CSUH	57.1	B	A
113	Loma Prieta	RV/OB	1989	6.9	APEEL 7 - Pulgas	47.7	B	A
114	Loma Prieta	RV/OB	1989	6.9	APEEL 9 - Crystal Springs Res	46.9	B	A
115	Loma Prieta	RV/OB	1989	6.9	Belmont - Envirotech	49.9	B	A
116	Loma Prieta	RV/OB	1989	6.9	Berkeley LBL	83.6	B	A
117	Loma Prieta	RV/OB	1989	6.9	Capitola	14.5	C	C
118	Loma Prieta	RV/OB	1989	6.9	Corralitos	5.1	B	B
119	Loma Prieta	RV/OB	1989	6.9	Fremont - Emerson Court	43.4	C	B
120	Loma Prieta	RV/OB	1989	6.9	Fremont - Mission San Jose	43	B	B

Table A.5 (continued). Expanded record set used for calculation of response spectral correlations in Chapter 8. These records were also used as the library of records in Chapter 6.

#	FileName (Vertical)	Filename (Horiz. #1)	Filename (Horiz. #2)	HP[4]	LP[4]
81	IMPVALL\H-SUP-UP	IMPVALL\H-SUP045	IMPVALL\H-SUP135	0.1	40
82	IMPVALL\H-WSM-UP	IMPVALL\H-WSM090	IMPVALL\H-WSM180	0.1	40
83	KOBE\ABN-UP	KOBE\ABN000	KOBE\ABN090	0.05	40
84	KOBE\TDO-UP	KOBE\TDO000	KOBE\TDO090	0.05	40
85	KOCAELI\ARCDWN	KOCAELI\ARC000	KOCAELI\ARC090	0.08	50
86	KOCAELI\ATK-UP	KOCAELI\ATK000	KOCAELI\ATK090	0.08	40
87	KOCAELI\BUR-UP	KOCAELI\BUR090	KOCAELI\BUR000	0.08	50
88	KOCAELI\CNA-UP	KOCAELI\CNA000	KOCAELI\CNA090	0.02	50
89	KOCAELI\FAT-UP	KOCAELI\FAT000	KOCAELI\FAT090	0.03	50
90	KOCAELI\GYN-UP	KOCAELI\GYN090	KOCAELI\GYN000	0.15	25
91	KOCAELI\IZN-UP	KOCAELI\IZN180	KOCAELI\IZN090	0.1	25
92	KOCAELI\MCD--V	KOCAELI\MCD000	KOCAELI\MCD090	0.1	50
93	KOCAELI\ZYT-UP	KOCAELI\ZYT000	KOCAELI\ZYT090	0.06	50
94	LANDERS\BAK-UP	LANDERS\BAK050	LANDERS\BAK140	0.1	23
95	LANDERS\BRS-UP	LANDERS\BRS000	LANDERS\BRS090	0.07	23
96	LANDERS\BFS-UP	LANDERS\BFS000	LANDERS\BFS090	0.07	23
97	LANDERS\CLW-UP	LANDERS\CLW-LN	LANDERS\CLW-TR	0.1	30
98	LANDERS\DSP-UP	LANDERS\DSP000	LANDERS\DSP090	0.07	23
99	LANDERS\FTI-UP	LANDERS\FTI000	LANDERS\FTI090	0.07	23
100	LANDERS\H05-UP	LANDERS\H05000	LANDERS\H05090	0.16	23
101	LANDERS\IND-UP	LANDERS\IND000	LANDERS\IND090	0.1	23
102	LANDERS\JOS-UP	LANDERS\JOS000	LANDERS\JOS090	0.07	23
103	LANDERS\PSA-UP	LANDERS\PSA000	LANDERS\PSA090	0.07	23
104	LANDERS\RIV-UP	LANDERS\RIV180	LANDERS\RIV270	0.16	23
105	LANDERS\HOS-UP	LANDERS\HOS090	LANDERS\HOS180	0.1	50
106	LANDERS\YER-UP	LANDERS\YER270	LANDERS\YER360	0.07	23
107	LOMAP\AGW-UP	LOMAP\AGW000	LOMAP\AGW090	0.2	30
108	LOMAP\AND-UP	LOMAP\AND270	LOMAP\AND360	0.2	40
109	LOMAP\ADL-UP	LOMAP\ADL250	LOMAP\ADL340	0.1	32
110	LOMAP\A10-UP	LOMAP\A10000	LOMAP\A10090	0.1	20
111	LOMAP\A2E-UP	LOMAP\A2E000	LOMAP\A2E090	0.2	25
112	LOMAP\A3E-UP	LOMAP\A3E000	LOMAP\A3E090	0.2	30
113	LOMAP\A07-UP	LOMAP\A07000	LOMAP\A07090	0.1	22
114	LOMAP\A09-UP	LOMAP\A09137	LOMAP\A09227	0.2	40
115	LOMAP\BES-UP	LOMAP\BES000	LOMAP\BES090	0.2	22
116	LOMAP\BRK-UP	LOMAP\BRK000	LOMAP\BRK090	0.2	18
117	LOMAP\CAP-UP	LOMAP\CAP000	LOMAP\CAP090	0.2	40
118	LOMAP\CLS-UP	LOMAP\CLS000	LOMAP\CLS090	0.2	32
119	LOMAP\FMS-UP	LOMAP\FMS090	LOMAP\FMS180	0.1	31
120	LOMAP\FRE-UP	LOMAP\FRE000	LOMAP\FRE090	0.2	24

[4] Maximum high-pass frequency among the three components, and minimum low-pass frequency among the three components.

Table A.5 (continued). Expanded record set used for calculation of response spectral correlations in Chapter 8. These records were also used as the library of records in Chapter 6.

#	Event	Mechanism	Year	M_w	Station Name	Distance	USGS Soil Type	GM Soil Type
121	Loma Prieta	RV/OB	1989	6.9	Gilroy - Gavilan Coll.	11.6	B	B
122	Loma Prieta	RV/OB	1989	6.9	Gilroy - Historic Bldg.	12.7	-	D
123	Loma Prieta	RV/OB	1989	6.9	Gilroy Array #2	12.7	C	D
124	Loma Prieta	RV/OB	1989	6.9	Gilroy Array #3	14.4	C	D
125	Loma Prieta	RV/OB	1989	6.9	Gilroy Array #4	16.1	C	D
126	Loma Prieta	RV/OB	1989	6.9	Gilroy Array #6	19.9	B	B
127	Loma Prieta	RV/OB	1989	6.9	Gilroy Array #7	24.2	C	B
128	Loma Prieta	RV/OB	1989	6.9	Golden Gate Bridge	85.1	B	A
129	Loma Prieta	RV/OB	1989	6.9	Halls Valley	31.6	C	C
130	Loma Prieta	RV/OB	1989	6.9	Hayward - BART Sta	58.9	B	D
131	Loma Prieta	RV/OB	1989	6.9	Hollister - South & Pine	28.8	-	D
132	Loma Prieta	RV/OB	1989	6.9	Hollister City Hall	28.2	C	D
133	Loma Prieta	RV/OB	1989	6.9	Hollister Diff. Array	25.8	C	D
134	Loma Prieta	RV/OB	1989	6.9	Oakland - Title & Trust	77.4	C	D
135	Loma Prieta	RV/OB	1989	6.9	Palo Alto - 1900 Embarc.	36.1	C	D
136	Loma Prieta	RV/OB	1989	6.9	Palo Alto - SLAC Lab	36.3	B	A
137	Loma Prieta	RV/OB	1989	6.9	Richmond City Hall	93.1	C	D
138	Loma Prieta	RV/OB	1989	6.9	SAGO South - Surface	34.7	B	B
139	Loma Prieta	RV/OB	1989	6.9	Salinas - John & Work	32.6	C	D
140	Loma Prieta	RV/OB	1989	6.9	Saratoga - Aloha Ave	13	B	D
141	Loma Prieta	RV/OB	1989	6.9	SF - Diamond Heights	77	B	A
142	Loma Prieta	RV/OB	1989	6.9	SF - Presidio	83.1	B	A
143	Loma Prieta	RV/OB	1989	6.9	SF Intern. Airport	64.4	C	D
144	Loma Prieta	RV/OB	1989	6.9	Sunnyvale - Colton Ave.	28.8	C	D
145	Loma Prieta	RV/OB	1989	6.9	UCSC Lick Observatory	17.9	B	A
146	Loma Prieta	RV/OB	1989	6.9	Woodside	39.9	B	B
147	Mammoth Lakes	RV/OB	1980	6.3	Convict Creek	9	-	D
148	Mammoth Lakes	SS	1980	6	Convict Creek	17.4	-	D
149	Mammoth Lakes	SS	1980	5.7	Convict Creek	3	-	D
150	Mammoth Lakes	RV/OB	1980	6	Convict Creek	18.6	-	D
151	Morgan Hill	SS	1984	6.2	Anderson Dam (Downstream)	2.6	B	D
152	Morgan Hill	SS	1984	6.2	Capitola	38.1	C	C
153	Morgan Hill	SS	1984	6.2	Corralitos	22.7	B	B
154	Morgan Hill	SS	1984	6.2	Gilroy Array #2	15.1	C	D
155	Morgan Hill	SS	1984	6.2	Gilroy Array #3	14.6	C	D
156	Morgan Hill	SS	1984	6.2	Gilroy Array #4	12.8	C	D
157	Morgan Hill	SS	1984	6.2	Gilroy Array #6	11.8	B	B
158	Morgan Hill	SS	1984	6.2	Gilroy Array #7	14	C	B
159	Morgan Hill	SS	1984	6.2	Halls Valley	3.4	C	C
160	Morgan Hill	SS	1984	6.2	Hollister City Hall	32.5	C	D

Table A.5 (continued). Expanded record set used for calculation of response spectral correlations in Chapter 8. These records were also used as the library of records in Chapter 6.

#	FileName (Vertical)	Filename (Horiz. #1)	Filename (Horiz. #2)	HP[4]	LP[4]
121	LOMAP\GIL-UP	LOMAP\GIL067	LOMAP\GIL337	0.2	35
122	LOMAP\GOF-UP	LOMAP\GOF090	LOMAP\GOF180	0.2	38
123	LOMAP\G02-UP	LOMAP\G02000	LOMAP\G02090	0.2	31
124	LOMAP\G03-UP	LOMAP\G03000	LOMAP\G03090	0.1	33
125	LOMAP\G04-UP	LOMAP\G04000	LOMAP\G04090	0.2	28
126	LOMAP\G06-UP	LOMAP\G06000	LOMAP\G06090	0.2	31
127	LOMAP\GMR-UP	LOMAP\GMR000	LOMAP\GMR090	0.2	35
128	LOMAP\GGB-UP	LOMAP\GGB270	LOMAP\GGB360	0.2	22
129	LOMAP\HVR-UP	LOMAP\HVR000	LOMAP\HVR090	0.2	22
130	LOMAP\HWB-UP	LOMAP\HWB220	LOMAP\HWB310	0.2	31
131	LOMAP\HSP-UP	LOMAP\HSP000	LOMAP\HSP090	0.1	23
132	LOMAP\HCH-UP	LOMAP\HCH090	LOMAP\HCH180	0.1	29
133	LOMAP\HDA-UP	LOMAP\HDA165	LOMAP\HDA255	0.1	33
134	LOMAP\TIB-UP	LOMAP\TIB180	LOMAP\TIB270	0.2	38
135	LOMAP\PAE-UP	LOMAP\PAE000	LOMAP\PAE090	0.2	30
136	LOMAP\SLC-UP	LOMAP\SLC270	LOMAP\SLC360	0.2	28
137	LOMAP\RCH-UP	LOMAP\RCH190	LOMAP\RCH280	0.2	25
138	LOMAP\SG3-UP	LOMAP\SG3261	LOMAP\SG3351	0.1	25
139	LOMAP\SJW-UP	LOMAP\SJW160	LOMAP\SJW250	0.1	28
140	LOMAP\STG-UP	LOMAP\STG000	LOMAP\STG090	0.1	38
141	LOMAP\DMH-UP	LOMAP\DMH000	LOMAP\DMH090	0.2	22
142	LOMAP\PRS-UP	LOMAP\PRS000	LOMAP\PRS090	0.1	31
143	LOMAP\SFO-UP	LOMAP\SFO000	LOMAP\SFO090	0.2	30
144	LOMAP\SVL-UP	LOMAP\SVL270	LOMAP\SVL360	0.1	32
145	LOMAP\LOB-UP	LOMAP\LOB000	LOMAP\LOB090	0.2	40
146	LOMAP\WDS-UP	LOMAP\WDS000	LOMAP\WDS090	0.1	25
147	MAMMOTH\I-CVK-UP	MAMMOTH\I-CVK090	MAMMOTH\I-CVK180	0.2	41
148	MAMMOTH\A-CVK-UP	MAMMOTH\A-CVK090	MAMMOTH\A-CVK180	0.2	30
149	MAMMOTH\B-CVK-UP	MAMMOTH\B-CVK090	MAMMOTH\B-CVK180	0.2	35
150	MAMMOTH\L-CVK-UP	MAMMOTH\L-CVK090	MAMMOTH\L-CVK180	0.1	40
151	MORGAN\AND-UP	MORGAN\AND250	MORGAN\AND340	0.1	30
152	MORGAN\CAP-UP	MORGAN\CAP042	MORGAN\CAP132	0.2	28
153	MORGAN\CLS-UP	MORGAN\CLS220	MORGAN\CLS310	0.2	24
154	MORGAN\G02-UP	MORGAN\G02000	MORGAN\G02090	0.2	31
155	MORGAN\G03-UP	MORGAN\G03000	MORGAN\G03090	0.1	32
156	MORGAN\G04-UP	MORGAN\G04270	MORGAN\G04360	0.1	25
157	MORGAN\G06-UP	MORGAN\G06000	MORGAN\G06090	0.1	27
158	MORGAN\GMR-UP	MORGAN\GMR000	MORGAN\GMR090	0.1	30
159	MORGAN\HVR-UP	MORGAN\HVR150	MORGAN\HVR240	0.2	26
160	MORGAN\HCH-UP	MORGAN\HCH001	MORGAN\HCH271	0.2	19

[4] Maximum high-pass frequency among the three components, and minimum low-pass frequency among the three components.

Table A.5 (continued). Expanded record set used for calculation of response spectral correlations in Chapter 8. These records were also used as the library of records in Chapter 6.

#	Event	Mechanism	Year	M_w	Station Name	Distance	USGS Soil Type	GM Soil Type
161	Morgan Hill	SS	1984	6.2	Hollister Diff Array #1	28.3	C	D
162	Morgan Hill	SS	1984	6.2	Hollister Diff Array #3	28.3	C	D
163	Morgan Hill	SS	1984	6.2	Hollister Diff Array #4	28.3	C	D
164	Morgan Hill	SS	1984	6.2	Hollister Diff Array #5	28.3	C	D
165	Morgan Hill	SS	1984	6.2	Hollister Diff. Array	28.3	C	D
166	Morgan Hill	SS	1984	6.2	San Juan Bautista, 24 Polk St	30.3	B	D
167	N. Palm Springs	RV/OB	1986	6	Cabazon	16.3	-	D
168	N. Palm Springs	RV/OB	1986	6	Hesperia	75.9	B	D
169	N. Palm Springs	RV/OB	1986	6	Indio	39.6	-	D
170	N. Palm Springs	RV/OB	1986	6	Puerta La Cruz	71.9	B	B
171	Northridge	RV	1994	6.7	Alhambra - Fremont School	35.7	B	D
172	Northridge	RV	1994	6.7	Anaverde Valley - City R	38.4	C	D
173	Northridge	RV	1994	6.7	Arleta - Nordhoff Fire Sta	9.2	C	D
174	Northridge	RV	1994	6.7	Bell Gardens - Jaboneria	46.6	C	D
175	Northridge	RV	1994	6.7	Burbank - Howard Rd.	20	B	B
176	Northridge	RV	1994	6.7	Camarillo	36.5	C	-
177	Northridge	RV	1994	6.7	Canoga Park - Topanga Can	15.8	C	D
178	Northridge	RV	1994	6.7	Canyon Country - W Lost Cany	13	C	D
179	Northridge	RV	1994	6.7	Castaic - Old Ridge Route	22.6	B	B
180	Northridge	RV	1994	6.7	Downey - Co Maint Bldg	47.6	C	D
181	Northridge	RV	1994	6.7	Elizabeth Lake	37.2	C	D
182	Northridge	RV	1994	6.7	Hollywood - Willoughby Ave	25.7	B	D
183	Northridge	RV	1994	6.7	Huntington Beach - Lake St	79.6	C	D
184	Northridge	RV	1994	6.7	Inglewood - Union Oil	44.7	B	D
185	Northridge	RV	1994	6.7	LA - Baldwin Hills	31.3	B	B
186	Northridge	RV	1994	6.7	LA - Centinela St	30.9	C	D
187	Northridge	RV	1994	6.7	LA - Century City CC North	25.7	C	D
188	Northridge	RV	1994	6.7	LA - E Vernon Ave	39.3	C	D
189	Northridge	RV	1994	6.7	LA - Hollywood Stor FF	25.5	C	D
190	Northridge	RV	1994	6.7	LA - N Faring Rd	23.9	C	B
191	Northridge	RV	1994	6.7	LA - N Westmoreland	29	B	D
192	Northridge	RV	1994	6.7	LA - Pico & Sentous	32.7	C	D
193	Northridge	RV	1994	6.7	LA - Saturn St	30	C	D
194	Northridge	RV	1994	6.7	LA - Temple & Hope	32.3	B	A
195	Northridge	RV	1994	6.7	LA - UCLA Grounds	14.9	B	-
196	Northridge	RV	1994	6.7	LA - Univ. Hospital	34.6	B	A
197	Northridge	RV	1994	6.7	Lake Hughes #1	36.3	B	C
198	Northridge	RV	1994	6.7	Lake Hughes #12A	22.8	B	C
199	Northridge	RV	1994	6.7	Lake Hughes #4 - Camp Mend	32.3	B	B
200	Northridge	RV	1994	6.7	Lake Hughes #4B - Camp Mend	32.3	B	B

Table A.5 (continued). Expanded record set used for calculation of response spectral correlations in Chapter 8. These records were also used as the library of records in Chapter 6.

#	FileName (Vertical)	Filename (Horiz. #1)	Filename (Horiz. #2)	HP[4]	LP[4]
161	MORGAN\HD1-UP	MORGAN\HD1255	MORGAN\HD1165	0.2	30
162	MORGAN\HD3-UP	MORGAN\HD3255	MORGAN\HD3165	0.2	30
163	MORGAN\HD4-UP	MORGAN\HD4165	MORGAN\HD4255	0.2	30
164	MORGAN\HD5-UP	MORGAN\HD5165	MORGAN\HD5255	0.2	30
165	MORGAN\HDA-UP	MORGAN\HDA165	MORGAN\HDA255	0.2	23
166	MORGAN\SJB-UP	MORGAN\SJB213	MORGAN\SJB303	0.1	21
167	PALMSPR\CAB-UP	PALMSPR\CAB180	PALMSPR\CAB270	0.2	40
168	PALMSPR\HES-UP	PALMSPR\HES002	PALMSPR\HES092	0.2	25
169	PALMSPR\INO-UP	PALMSPR\INO225	PALMSPR\INO315	0.1	35
170	PALMSPR\PLC-UP	PALMSPR\PLC258	PALMSPR\PLC348	0.2	32
171	NORTHR\ALH-UP	NORTHR\ALH090	NORTHR\ALH360	0.12	25
172	NORTHR\ANA-UP	NORTHR\ANA090	NORTHR\ANA180	0.2	46
173	NORTHR\ARL-UP	NORTHR\ARL090	NORTHR\ARL360	0.12	23
174	NORTHR\JAB-UP	NORTHR\JAB220	NORTHR\JAB310	0.13	30
175	NORTHR\HOW-UP	NORTHR\HOW060	NORTHR\HOW330	0.1	30
176	NORTHR\CMR-UP	NORTHR\CMR180	NORTHR\CMR270	0.1	25
177	NORTHR\CNP-UP	NORTHR\CNP106	NORTHR\CNP196	0.1	30
178	NORTHR\LOS-UP	NORTHR\LOS270	NORTHR\LOS000	0.2	30
179	NORTHR\ORR-UP	NORTHR\ORR090	NORTHR\ORR360	0.12	23
180	NORTHR\DWN-UP	NORTHR\DWN090	NORTHR\DWN360	0.2	23
181	NORTHR\ELI-UP	NORTHR\ELI090	NORTHR\ELI180	0.16	46
182	NORTHR\WIL-UP	NORTHR\WIL180	NORTHR\WIL090	0.2	30
183	NORTHR\HNT-UP	NORTHR\HNT000	NORTHR\HNT090	0.2	23
184	NORTHR\ING-UP	NORTHR\ING000	NORTHR\ING090	0.16	23
185	NORTHR\BLD-UP	NORTHR\BLD090	NORTHR\BLD360	0.16	23
186	NORTHR\CEN-UP	NORTHR\CEN155	NORTHR\CEN245	0.2	30
187	NORTHR\CCN-UP	NORTHR\CCN090	NORTHR\CCN360	0.14	23
188	NORTHR\VER-UP	NORTHR\VER090	NORTHR\VER180	0.2	30
189	NORTHR\PEL-UP	NORTHR\PEL090	NORTHR\PEL360	0.2	23
190	NORTHR\FAR-UP	NORTHR\FAR000	NORTHR\FAR090	0.2	30
191	NORTHR\WST-UP	NORTHR\WST000	NORTHR\WST270	0.2	30
192	NORTHR\PIC-UP	NORTHR\PIC090	NORTHR\PIC180	0.2	46
193	NORTHR\STN-UP	NORTHR\STN020	NORTHR\STN110	0.13	30
194	NORTHR\TEM-UP	NORTHR\TEM090	NORTHR\TEM180	0.2	46
195	NORTHR\UCL-UP	NORTHR\UCL090	NORTHR\UCL360	0.08	25
196	NORTHR\UNI-UP	NORTHR\UNI005	NORTHR\UNI095	0.2	46
197	NORTHR\L01-UP	NORTHR\L01000	NORTHR\L01090	0.12	23
198	NORTHR\H12-UP	NORTHR\H12090	NORTHR\H12180	0.13	46
199	NORTHR\L04-UP	NORTHR\L04000	NORTHR\L04090	0.12	23
200	NORTHR\L4B-UP	NORTHR\L4B000	NORTHR\L4B090	0.12	23

[4] Maximum high-pass frequency among the three components, and minimum low-pass frequency among the three components.

Table A.5 (continued). Expanded record set used for calculation of response spectral correlations in Chapter 8. These records were also used as the library of records in Chapter 6.

#	Event	Mechanism	Year	M_w	Station Name	Distance	USGS Soil Type	GM Soil Type
201	Northridge	RV	1994	6.7	Lawndale - Osage Ave	42.4	C	D
202	Northridge	RV	1994	6.7	LB - Rancho Los Cerritos	54.3	-	D
203	Northridge	RV	1994	6.7	Leona Valley #2	37.7	C	-
204	Northridge	RV	1994	6.7	Leona Valley #4	38.1	C	-
205	Northridge	RV	1994	6.7	Leona Valley #5 - Ritter	38.3	C	C
206	Northridge	RV	1994	6.7	Leona Valley #6	38.5	C	D
207	Northridge	RV	1994	6.7	N Hollywood - Coldwater Can	14.6	B	C
208	Northridge	RV	1994	6.7	Neenach - Sacatara Ck	53.2	B	D
209	Northridge	RV	1994	6.7	Newhall - Fire Sta	7.1	C	D
210	Northridge	RV	1994	6.7	Newhall - W Pico Canyon Rd.	7.1	B	C
211	Northridge	RV	1994	6.7	Newport Bch - Irvine Ave. F.S	87.6	B	-
212	Northridge	RV	1994	6.7	Newport Bch - Newp & Coast	84.6	B	B
213	Northridge	RV	1994	6.7	Pacific Palisades - Sunset	26.2	B	B
214	Northridge	RV	1994	6.7	Pacoima Kagel Canyon	8.2	B	B
215	Northridge	RV	1994	6.7	Palmdale - Hwy 14 & Palmdale	43.6	C	C
216	Northridge	RV	1994	6.7	Phelan - Wilson Ranch	86.1	B	D
217	Northridge	RV	1994	6.7	Playa Del Rey - Saran	34.2	B	D
218	Northridge	RV	1994	6.7	Port Hueneme - Naval Lab.	54.3	C	D
219	Northridge	RV	1994	6.7	Rolling Hills Est-Rancho Vista	46.6	B	-
220	Northridge	RV	1994	6.7	Santa Monica City Hall	27.6	C	D
221	Northridge	RV	1994	6.7	Seal Beach - Office Bldg	64.9	B	D
222	Northridge	RV	1994	6.7	Sun Valley - Roscoe Blvd	12.3	C	D
223	Northridge	RV	1994	6.7	Sunland - Mt Gleason Ave	17.7	B	C
224	Northridge	RV	1994	6.7	Sylmar - Olive View Med FF	6.4	C	D
225	Northridge	RV	1994	6.7	Tarzana - Cedar Hill	17.5	C	B
226	Northridge	RV	1994	6.7	Terminal Island - S Seaside	60	C	D
227	Northridge	RV	1994	6.7	West Covina - S Orange Ave	54.1	C	B
228	Parkfield	SS	1966	6.1	Cholame #12	14.7	B	B
229	Parkfield	SS	1966	6.1	Cholame #8	9.2	C	B
230	Point Mugu	RV	1973	5.8	Port Hueneme	25	C	D
231	San Fernando	RV	1971	6.6	2516 Via Tejon PV	65.1	-	C
232	San Fernando	RV	1971	6.6	Cedar Springs Pumphouse	87.6	-	B
233	San Fernando	RV	1971	6.6	Fort Tejon	64.1	B	B
234	San Fernando	RV	1971	6.6	Gormon - Oso Pump Plant	48.1	C	C
235	San Fernando	RV	1971	6.6	LB - Terminal Island	69.2	C	D
236	San Fernando	RV	1971	6.6	Pearblossom Pump	38.9	B	B
237	San Fernando	RV	1971	6.6	Wheeler Ridge - Ground	81.6	C	D
238	San Fernando	RV	1971	6.6	Wrightwood - 6074 Park Dr	60.3	B	B
239	Santa Barbara	RV/OB	1978	6	Santa Barbara Courthouse	14	B	D
240	Superstitn Hills(A	SS	NaN	6.3	Wildlife Liquef. Array	24.7	-	D

Table A.5 (continued). Expanded record set used for calculation of response spectral correlations in Chapter 8. These records were also used as the library of records in Chapter 6.

#	FileName (Vertical)	Filename (Horiz. #1)	Filename (Horiz. #2)	HP[4]	LP[4]
201	NORTHR\LOA-UP	NORTHR\LOA092	NORTHR\LOA182	0.13	30
202	NORTHR\LBR-UP	NORTHR\LBR000	NORTHR\LBR090	0.16	23
203	NORTHR\LV2-UP	NORTHR\LV2000	NORTHR\LV2090	0.2	23
204	NORTHR\LV4-UP	NORTHR\LV4000	NORTHR\LV4090	0.2	23
205	NORTHR\LV5-UP	NORTHR\LV5000	NORTHR\LV5090	0.2	23
206	NORTHR\LV6-UP	NORTHR\LV6090	NORTHR\LV6360	0.2	23
207	NORTHR\CWC-UP	NORTHR\CWC180	NORTHR\CWC270	0.13	30
208	NORTHR\NEE-UP	NORTHR\NEE090	NORTHR\NEE180	0.12	46
209	NORTHR\NWH-UP	NORTHR\NWH090	NORTHR\NWH360	0.12	23
210	NORTHR\WPI-UP	NORTHR\WPI046	NORTHR\WPI316	0.1	30
211	NORTHR\NBI-UP	NORTHR\NBI000	NORTHR\NBI090	0.2	23
212	NORTHR\NEW-UP	NORTHR\NEW090	NORTHR\NEW180	0.17	46
213	NORTHR\SUN-UP	NORTHR\SUN190	NORTHR\SUN280	0.1	30
214	NORTHR\PKC-UP	NORTHR\PKC090	NORTHR\PKC360	0.2	23
215	NORTHR\PHP-UP	NORTHR\PHP000	NORTHR\PHP270	0.2	46
216	NORTHR\PHE-UP	NORTHR\PHE090	NORTHR\PHE180	0.2	46
217	NORTHR\SAR-UP	NORTHR\SAR000	NORTHR\SAR270	0.1	30
218	NORTHR\PTH-UP	NORTHR\PTH090	NORTHR\PTH180	0.14	23
219	NORTHR\RHE-UP	NORTHR\RHE090	NORTHR\RHE360	0.15	25
220	NORTHR\STM-UP	NORTHR\STM090	NORTHR\STM360	0.14	23
221	NORTHR\SEA-UP	NORTHR\SEA000	NORTHR\SEA090	0.16	46
222	NORTHR\RO3-UP	NORTHR\RO3000	NORTHR\RO3090	0.1	30
223	NORTHR\GLE-UP	NORTHR\GLE170	NORTHR\GLE260	0.1	30
224	NORTHR\SYL-UP	NORTHR\SYL090	NORTHR\SYL360	0.12	23
225	NORTHR\TAR-UP	NORTHR\TAR090	NORTHR\TAR360	0.1	23
226	NORTHR\SSE-UP	NORTHR\SSE240	NORTHR\SSE330	0.13	30
227	NORTHR\SOR-UP	NORTHR\SOR225	NORTHR\SOR315	0.2	30
228	PARKF\C12DWN	PARKF\C12050	PARKF\C12320	0.2	20
229	PARKF\C08DWN	PARKF\C08050	PARKF\C08320	0.2	20
230	PTMUGU\PHN-UP	PTMUGU\PHN180	PTMUGU\PHN270	0.2	25
231	SFERN\PVEDWN	SFERN\PVE065	SFERN\PVE155	0.2	20
232	SFERN\CSPDWN	SFERN\CSP126	SFERN\CSP216	0.1	20
233	SFERN\FTJ-UP	SFERN\FTJ000	SFERN\FTJ090	0.1	20
234	SFERN\OPP-UP	SFERN\OPP000	SFERN\OPP270	0.1	23
235	SFERN\TLI-UP	SFERN\TLI249	SFERN\TLI339	0.1	20
236	SFERN\PPPDWN	SFERN\PPP000	SFERN\PPP270	0.2	35
237	SFERN\WRP-UP	SFERN\WRP090	SFERN\WRP180	0.1	23
238	SFERN\WTWDWN	SFERN\WTW025	SFERN\WTW295	0.2	30
239	SBARB\SBA-UP	SBARB\SBA132	SBARB\SBA222	0.1	26
240	SUPERST\A-IVW-UP	SUPERST\A-IVW090	SUPERST\A-IVW360	0.2	50

[4] Maximum high-pass frequency among the three components, and minimum low-pass frequency among the three components.

Table A.5 (continued). Expanded record set used for calculation of response spectral correlations in Chapter 8. These records were also used as the library of records in Chapter 6.

#	Event	Mechanism	Year	M_w	Station Name	Distance	USGS Soil Type	GM Soil Type
241	Superstitn Hills(B	SS	NaN	6.7	El Centro Imp. Co. Cent	13.9	C	D
242	Superstitn Hills(B	SS	NaN	6.7	Westmorland Fire Sta	13.3	C	D
243	Superstitn Hills(B	SS	NaN	6.7	Wildlife Liquef. Array	24.4	-	D
244	Taiwan SMART1(40)	RV/OB	1986	6.4	SMART1 C00	64	-	D
245	Taiwan SMART1(40)	RV/OB	1986	6.4	SMART1 E01	64	-	D
246	Taiwan SMART1(40)	RV/OB	1986	6.4	SMART1 I01	64	-	D
247	Taiwan SMART1(40)	RV/OB	1986	6.4	SMART1 I07	64	-	D
248	Taiwan SMART1(40)	RV/OB	1986	6.4	SMART1 M01	64	-	D
249	Taiwan SMART1(40)	RV/OB	1986	6.4	SMART1 M07	64	-	D
250	Taiwan SMART1(40)	RV/OB	1986	6.4	SMART1 O01	64	-	D
251	Taiwan SMART1(40)	RV/OB	1986	6.4	SMART1 O07	64	-	D
252	Taiwan SMART1(45)	RV	1986	7.3	SMART1 C00	39	-	D
253	Taiwan SMART1(45)	RV	1986	7.3	SMART1 E01	39	-	D
254	Taiwan SMART1(45)	RV	1986	7.3	SMART1 E02	39	-	D
255	Taiwan SMART1(45)	RV	1986	7.3	SMART1 I01	39	-	D
256	Taiwan SMART1(45)	RV	1986	7.3	SMART1 I07	39	-	D
257	Taiwan SMART1(45)	RV	1986	7.3	SMART1 M01	39	-	D
258	Taiwan SMART1(45)	RV	1986	7.3	SMART1 M07	39	-	D
259	Taiwan SMART1(45)	RV	1986	7.3	SMART1 O01	39	-	D
260	Taiwan SMART1(45)	RV	1986	7.3	SMART1 O02	39	-	D
261	Taiwan SMART1(45)	RV	1986	7.3	SMART1 O04	39	-	D
262	Taiwan SMART1(45)	RV	1986	7.3	SMART1 O06	39	-	D
263	Taiwan SMART1(45)	RV	1986	7.3	SMART1 O07	39	-	D
264	Taiwan SMART1(45)	RV	1986	7.3	SMART1 O08	39	-	D
265	Taiwan SMART1(45)	RV	1986	7.3	SMART1 O10	39	-	D
266	Taiwan SMART1(45)	RV	1986	7.3	SMART1 O12	39	-	D
267	Whittier Narrows	RV	1987	6	LA - 116th St School	22.5	B	D

Table A.5 (continued). Expanded record set used for calculation of response spectral correlations in Chapter 8. These records were also used as the library of records in Chapter 6.

#	FileName (Vertical)	Filename (Horiz. #1)	Filename (Horiz. #2)	HP[4]	LP[4]
241	SUPERST\B-ICC-UP	SUPERST\B-ICC000	SUPERST\B-ICC090	0.1	38
242	SUPERST\B-WSM-UP	SUPERST\B-WSM090	SUPERST\B-WSM180	0.1	35
243	SUPERST\B-IVW-UP	SUPERST\B-IVW090	SUPERST\B-IVW360	0.1	40
244	SMART1\40C00DN	SMART1\40C00EW	SMART1\40C00NS	0.2	25
245	SMART1\40E01DN	SMART1\40E01EW	SMART1\40E01NS	0.2	25
246	SMART1\40I01DN	SMART1\40I01EW	SMART1\40I01NS	0.2	25
247	SMART1\40I07DN	SMART1\40I07EW	SMART1\40I07NS	0.2	25
248	SMART1\40M01DN	SMART1\40M01EW	SMART1\40M01NS	0.2	25
249	SMART1\40M07DN	SMART1\40M07EW	SMART1\40M07NS	0.2	25
250	SMART1\40O01DN	SMART1\40O01EW	SMART1\40O01NS	0.2	25
251	SMART1\40O07DN	SMART1\40O07EW	SMART1\40O07NS	0.2	25
252	SMART1\45C00DN	SMART1\45C00EW	SMART1\45C00NS	0.1	25
253	SMART1\45E01DN	SMART1\45E01EW	SMART1\45E01NS	0.1	25
254	SMART1\45E02DN	SMART1\45E02EW	SMART1\45E02NS	0.1	25
255	SMART1\45I01DN	SMART1\45I01EW	SMART1\45I01NS	0.1	25
256	SMART1\45I07DN	SMART1\45I07EW	SMART1\45I07NS	0.1	25
257	SMART1\45M01DN	SMART1\45M01EW	SMART1\45M01NS	0.1	25
258	SMART1\45M07DN	SMART1\45M07NS	SMART1\45M07EW	0.2	25
259	SMART1\45O01DN	SMART1\45O01EW	SMART1\45O01NS	0.1	25
260	SMART1\45O02DN	SMART1\45O02EW	SMART1\45O02NS	0.1	25
261	SMART1\45O04DN	SMART1\45O04EW	SMART1\45O04NS	0.1	25
262	SMART1\45O06DN	SMART1\45O06EW	SMART1\45O06NS	0.1	25
263	SMART1\45O07DN	SMART1\45O07EW	SMART1\45O07NS	0.1	25
264	SMART1\45O08DN	SMART1\45O08EW	SMART1\45O08NS	0.1	25
265	SMART1\45O10DN	SMART1\45O10EW	SMART1\45O10NS	0.1	25
266	SMART1\45O12DN	SMART1\45O12EW	SMART1\45O12NS	0.2	25
267	WHITTIER\A-116-UP	WHITTIER\A-116270	WHITTIER\A-116360	0.2	30

[4] Maximum high-pass frequency among the three components, and minimum low-pass frequency among the three components.

Table A.6. M,R-Based Records for Chapter 5. Selected to match Van Nuys disaggregation with IM=*Sa*(0.8s).

#	0.1g	0.2g	0.4g	0.6g	0.8g	1g	1.4g	1.8g	2.4g	3g	4g
1	2 (1)	2 (1)	2 (1)	2 (1)	2 (1)	2 (1)	2 (1)	2 (1)	2 (1)	2 (1)	2 (1)
2	4 (1)	4 (1)	4 (1)	4 (1)	4 (1)	4 (1)	4 (1)	4 (1)	4 (1)	4 (1)	4 (1)
3	7 (2)	7 (2)	7 (2)	7 (2)	7 (2)	7 (2)	7 (2)	7 (2)	7 (2)	7 (2)	7 (2)
4	9 (1)	9 (1)	9 (1)	9 (1)	9 (1)	9 (1)	9 (1)	9 (1)	9 (1)	9 (1)	9 (1)
5	10 (1)	10 (1)	10 (1)	10 (1)	10 (1)	10 (1)	10 (1)	10 (1)	10 (1)	10 (1)	10 (1)
6	13 (1)	13 (1)	13 (1)	13 (1)	13 (1)	13 (1)	13 (1)	13 (1)	13 (1)	13 (1)	13 (1)
7	18 (1)	18 (1)	18 (1)	18 (1)	18 (1)	18 (1)	18 (1)	18 (1)	18 (1)	18 (1)	18 (1)
8	26 (1)	26 (1)	26 (1)	26 (1)	26 (1)	26 (1)	26 (1)	26 (1)	26 (1)	26 (1)	26 (1)
9	35 (1)	35 (1)	35 (1)	35 (1)	35 (1)	35 (1)	35 (1)	35 (1)	35 (1)	35 (1)	35 (1)
10	41 (1)	41 (1)	41 (1)	41 (1)	41 (1)	41 (1)	41 (1)	41 (1)	41 (1)	41 (1)	41 (1)
11	59 (2)	59 (2)	59 (2)	59 (2)	59 (2)	59 (2)	59 (2)	59 (2)	59 (2)	59 (2)	59 (2)
12	69 (1)	69 (1)	69 (1)	69 (1)	69 (1)	69 (1)	69 (1)	69 (1)	69 (1)	69 (1)	69 (1)
13	77 (2)	77 (2)	77 (2)	77 (2)	77 (2)	77 (2)	77 (2)	77 (2)	77 (2)	77 (2)	77 (2)
14	79 (1)	79 (1)	79 (1)	79 (1)	79 (1)	79 (1)	79 (1)	79 (1)	79 (1)	79 (1)	79 (1)
15	87 (2)	87 (2)	87 (2)	87 (2)	87 (2)	87 (2)	87 (2)	87 (2)	87 (2)	87 (2)	87 (2)
16	88 (2)	88 (2)	88 (2)	88 (2)	88 (2)	88 (2)	88 (2)	88 (2)	88 (2)	88 (2)	88 (2)
17	89 (2)	89 (2)	89 (2)	89 (2)	89 (2)	89 (2)	89 (2)	89 (2)	89 (2)	89 (2)	89 (2)
18	92 (1)	92 (1)	92 (1)	92 (1)	92 (1)	92 (1)	92 (1)	92 (1)	92 (1)	92 (1)	92 (1)
19	97 (1)	97 (1)	97 (1)	97 (1)	97 (1)	97 (1)	97 (1)	97 (1)	97 (1)	97 (1)	97 (1)
20	98 (2)	98 (2)	98 (2)	98 (2)	98 (2)	98 (2)	98 (2)	98 (2)	98 (2)	98 (2)	98 (2)
21	108 (1)	108 (1)	108 (1)	108 (1)	108 (1)	108 (1)	108 (1)	108 (1)	108 (1)	108 (1)	108 (1)
22	114 (1)	114 (1)	114 (1)	114 (1)	114 (1)	114 (1)	114 (1)	114 (1)	114 (1)	114 (1)	114 (1)
23	115 (1)	115 (1)	115 (1)	115 (1)	115 (1)	115 (1)	115 (1)	115 (1)	115 (1)	115 (1)	115 (1)
24	116 (2)	116 (2)	116 (2)	116 (2)	116 (2)	116 (2)	116 (2)	116 (2)	116 (2)	116 (2)	116 (2)
25	133 (2)	133 (2)	133 (2)	133 (2)	133 (2)	133 (2)	133 (2)	133 (2)	133 (2)	133 (2)	133 (2)
26	140 (2)	140 (2)	140 (2)	140 (2)	140 (2)	140 (2)	140 (2)	140 (2)	140 (2)	140 (2)	140 (2)
27	150 (2)	150 (2)	150 (2)	150 (2)	150 (2)	150 (2)	150 (2)	150 (2)	150 (2)	150 (2)	150 (2)
28	152 (1)	152 (1)	152 (1)	152 (1)	152 (1)	152 (1)	152 (1)	152 (1)	152 (1)	152 (1)	152 (1)
29	159 (1)	159 (1)	159 (1)	159 (1)	159 (1)	159 (1)	159 (1)	159 (1)	159 (1)	159 (1)	159 (1)
30	174 (1)	174 (1)	174 (1)	174 (1)	174 (1)	174 (1)	174 (1)	174 (1)	174 (1)	174 (1)	174 (1)
31	181 (2)	181 (2)	181 (2)	181 (2)	181 (2)	181 (2)	181 (2)	181 (2)	181 (2)	181 (2)	181 (2)
32	193 (1)	193 (1)	193 (1)	193 (1)	193 (1)	193 (1)	193 (1)	193 (1)	193 (1)	193 (1)	193 (1)
33	218 (2)	218 (2)	218 (2)	218 (2)	218 (2)	218 (2)	218 (2)	218 (2)	218 (2)	218 (2)	218 (2)
34	230 (1)	230 (1)	230 (1)	230 (1)	230 (1)	230 (1)	230 (1)	230 (1)	230 (1)	230 (1)	230 (1)
35	231 (1)	231 (1)	231 (1)	231 (1)	231 (1)	231 (1)	231 (1)	231 (1)	231 (1)	231 (1)	231 (1)
36	250 (2)	250 (2)	250 (2)	250 (2)	250 (2)	250 (2)	250 (2)	250 (2)	250 (2)	250 (2)	250 (2)
37	251 (2)	251 (2)	251 (2)	251 (2)	251 (2)	251 (2)	251 (2)	251 (2)	251 (2)	251 (2)	251 (2)
38	256 (2)	256 (2)	256 (2)	256 (2)	256 (2)	256 (2)	256 (2)	256 (2)	256 (2)	256 (2)	256 (2)
39	260 (1)	260 (1)	260 (1)	260 (1)	260 (1)	260 (1)	260 (1)	260 (1)	260 (1)	260 (1)	260 (1)
40	266 (1)	266 (1)	266 (1)	266 (1)	266 (1)	266 (1)	266 (1)	266 (1)	266 (1)	266 (1)	266 (1)

The column specifies the *Sa*(0.8s) “stripe” level in units of g. The row refers to the record number within that stripe. For each row and column, the first number specifies the record (as numbered in Table A.5) and the number in parentheses specifies the horizontal component number (as numbered in Table A.5).

Table A.7. ε-Based Records for Chapter 5. Selected to match Van Nuys disaggregation with IM=*Sa*(0.8s).

#	0.1g	0.2g	0.4g	0.6g	0.8g	1g	1.4g	1.8g	2.4g	3g	4g
1	4 (2)	4 (1)	5 (2)	3 (1)	3 (1)	3 (1)	3 (2)	3 (2)	3 (2)	3 (2)	3 (2)
2	7 (1)	10 (1)	12 (1)	3 (2)	3 (2)	3 (2)	11 (1)	11 (1)	11 (1)	11 (1)	11 (1)
3	10 (2)	17 (2)	15 (1)	12 (1)	11 (1)	11 (1)	12 (2)	11 (2)	11 (2)	11 (2)	11 (2)
4	14 (2)	28 (1)	15 (2)	12 (2)	12 (1)	12 (2)	18 (1)	18 (1)	18 (1)	18 (1)	18 (1)
5	38 (1)	28 (2)	16 (2)	18 (1)	12 (2)	18 (1)	22 (1)	22 (1)	22 (1)	22 (1)	22 (1)
6	64 (2)	44 (2)	21 (2)	21 (2)	18 (1)	22 (1)	22 (2)	22 (2)	22 (2)	22 (2)	22 (2)
7	68 (2)	48 (1)	31 (1)	25 (2)	22 (2)	22 (2)	96 (1)	96 (1)	96 (1)	96 (1)	96 (1)
8	76 (1)	48 (2)	39 (1)	45 (1)	25 (2)	25 (2)	130 (1)	130 (1)	130 (1)	130 (1)	130 (1)
9	76 (2)	57 (1)	39 (2)	46 (2)	45 (1)	45 (1)	131 (1)	131 (1)	131 (1)	131 (1)	131 (1)
10	83 (1)	59 (1)	46 (2)	49 (2)	49 (2)	96 (1)	132 (1)	132 (1)	132 (1)	132 (1)	132 (1)
11	86 (2)	74 (2)	49 (2)	63 (2)	96 (1)	129 (1)	132 (2)	132 (2)	132 (2)	132 (2)	132 (2)
12	91 (2)	89 (1)	63 (2)	105 (1)	129 (1)	130 (1)	133 (1)	133 (1)	133 (1)	133 (1)	133 (1)
13	93 (1)	93 (2)	84 (1)	129 (1)	130 (1)	132 (1)	133 (2)	133 (2)	133 (2)	133 (2)	133 (2)
14	95 (2)	102 (2)	97 (2)	130 (1)	132 (1)	132 (2)	134 (1)	134 (1)	134 (1)	134 (1)	134 (1)
15	101 (1)	106 (1)	105 (1)	132 (2)	132 (2)	133 (1)	134 (2)	134 (2)	134 (2)	134 (2)	134 (2)
16	105 (2)	123 (1)	117 (1)	133 (1)	133 (1)	133 (2)	135 (2)	135 (2)	135 (2)	135 (2)	135 (2)
17	107 (2)	124 (1)	117 (2)	133 (2)	133 (2)	134 (1)	137 (1)	137 (1)	137 (1)	137 (1)	137 (1)
18	108 (1)	131 (2)	148 (1)	135 (2)	135 (2)	135 (2)	143 (1)	143 (1)	143 (1)	143 (1)	143 (1)
19	108 (2)	152 (2)	160 (2)	137 (2)	137 (2)	137 (2)	143 (2)	143 (2)	143 (2)	143 (2)	143 (2)
20	111 (2)	162 (1)	161 (2)	160 (2)	143 (1)	143 (1)	182 (1)	182 (1)	182 (1)	182 (1)	182 (1)
21	129 (2)	163 (1)	164 (2)	182 (1)	160 (2)	143 (2)	186 (2)	186 (2)	186 (2)	186 (2)	186 (2)
22	140 (1)	164 (1)	165 (1)	186 (2)	182 (1)	160 (2)	189 (2)	189 (2)	189 (2)	189 (2)	189 (2)
23	144 (2)	173 (1)	177 (2)	189 (2)	186 (2)	182 (1)	193 (1)	193 (2)	193 (2)	193 (2)	193 (2)
24	147 (1)	177 (1)	178 (1)	191 (2)	189 (2)	186 (2)	193 (2)	209 (2)	209 (2)	209 (2)	209 (2)
25	181 (1)	178 (2)	180 (2)	193 (1)	193 (1)	189 (2)	209 (2)	244 (1)	244 (1)	244 (1)	244 (1)
26	183 (1)	181 (2)	187 (2)	193 (2)	193 (2)	193 (1)	244 (1)	244 (2)	244 (2)	244 (2)	244 (2)
27	186 (1)	182 (2)	191 (2)	205 (1)	205 (1)	193 (2)	244 (2)	245 (1)	245 (1)	245 (1)	245 (1)
28	206 (1)	184 (2)	205 (1)	209 (2)	209 (2)	209 (2)	245 (2)	245 (2)	245 (2)	245 (2)	245 (2)
29	216 (1)	189 (1)	217 (1)	217 (1)	220 (1)	220 (1)	246 (1)	246 (1)	246 (1)	246 (1)	246 (1)
30	216 (2)	192 (2)	222 (2)	220 (1)	222 (2)	243 (2)	246 (2)	246 (2)	246 (2)	246 (2)	246 (2)
31	217 (2)	205 (2)	224 (1)	222 (2)	226 (2)	244 (2)	247 (1)	247 (1)	247 (1)	247 (1)	247 (1)
32	221 (2)	208 (2)	224 (2)	226 (1)	243 (2)	245 (2)	247 (2)	247 (2)	247 (2)	247 (2)	247 (2)
33	230 (2)	209 (1)	226 (1)	226 (2)	245 (2)	246 (2)	248 (1)	248 (1)	248 (1)	248 (1)	248 (1)
34	240 (1)	252 (2)	226 (2)	243 (2)	250 (1)	247 (2)	248 (2)	248 (2)	248 (2)	248 (2)	248 (2)
35	240 (2)	255 (2)	253 (2)	253 (2)	251 (2)	250 (1)	249 (2)	249 (1)	249 (1)	249 (1)	249 (1)
36	253 (1)	256 (1)	256 (2)	258 (2)	253 (2)	251 (2)	250 (1)	249 (2)	249 (2)	249 (2)	249 (2)
37	254 (2)	257 (2)	258 (2)	261 (1)	261 (1)	253 (2)	250 (2)	250 (1)	250 (1)	250 (1)	250 (1)
38	261 (2)	264 (1)	261 (1)	262 (2)	262 (2)	262 (2)	251 (1)	250 (2)	250 (2)	250 (2)	250 (2)
39	262 (1)	265 (1)	266 (1)	264 (2)	264 (2)	264 (2)	251 (2)	251 (1)	251 (1)	251 (1)	251 (1)
40	265 (2)	266 (2)	267 (2)	267 (1)	267 (1)	267 (1)	267 (1)	251 (2)	251 (2)	251 (2)	251 (2)

The column specifies the *Sa*(0.8s) "stripe" level in units of g. The row refers to the record number within that stripe. For each row and column, the first number specifies the record (as numbered in Table A.5) and the number in parentheses specifies the horizontal component number (as numbered in Table A.5).

Table A.8. CMS-ε records for Chapter 5. Selected to match spectral shape based on Van Nuys disaggregation with IM=Sa(0.8s).

#	0.1g	0.2g	0.4g	0.6g	0.8g	1g	1.4g	1.8g	2.4g	3g	4g
1	201 (2)	201 (2)	209 (2)	241 (1)	266 (1)	39 (1)	95 (2)	57 (1)	57 (1)	57 (1)	246 (2)
2	240 (1)	242 (1)	215 (1)	266 (1)	39 (1)	95 (2)	39 (1)	95 (2)	249 (2)	246 (2)	247 (2)
3	38 (1)	215 (1)	241 (1)	184 (2)	95 (2)	266 (1)	57 (1)	249 (2)	246 (2)	249 (2)	251 (1)
4	187 (1)	255 (2)	184 (2)	209 (2)	259 (1)	259 (1)	222 (2)	39 (1)	95 (2)	257 (1)	57 (1)
5	245 (2)	209 (2)	174 (2)	259 (1)	241 (1)	252 (1)	252 (1)	222 (2)	39 (1)	244 (2)	258 (2)
6	166 (2)	240 (1)	187 (2)	101 (1)	101 (1)	222 (2)	249 (2)	252 (1)	257 (1)	258 (2)	257 (1)
7	252 (2)	98 (2)	100 (2)	100 (2)	252 (1)	57 (1)	259 (1)	192 (2)	244 (2)	247 (2)	244 (2)
8	231 (2)	166 (2)	98 (2)	39 (1)	184 (2)	131 (1)	131 (1)	166 (1)	222 (2)	95 (2)	249 (2)
9	235 (1)	173 (2)	255 (2)	187 (2)	131 (1)	241 (1)	266 (1)	246 (2)	3 (2)	39 (1)	250 (1)
10	242 (1)	174 (2)	201 (1)	174 (2)	222 (2)	101 (1)	192 (2)	3 (2)	252 (1)	251 (1)	182 (1)
11	4 (1)	163 (1)	83 (1)	215 (1)	209 (2)	263 (1)	166 (1)	131 (1)	258 (2)	3 (2)	133 (1)
12	221 (1)	5 (2)	123 (1)	95 (2)	263 (1)	192 (2)	263 (1)	257 (1)	166 (1)	246 (1)	246 (1)
13	173 (2)	123 (1)	242 (1)	201 (1)	57 (1)	102 (1)	3 (2)	244 (2)	192 (2)	222 (2)	248 (2)
14	208 (1)	187 (2)	266 (1)	252 (1)	102 (1)	184 (2)	257 (1)	22 (1)	22 (1)	22 (1)	245 (1)
15	177 (2)	177 (2)	71 (1)	83 (1)	100 (2)	249 (2)	102 (1)	259 (1)	246 (1)	166 (1)	3 (2)
16	108 (2)	100 (2)	101 (1)	131 (1)	255 (1)	166 (1)	253 (2)	266 (1)	131 (1)	252 (1)	265 (1)
17	215 (1)	83 (1)	93 (1)	255 (1)	192 (2)	255 (1)	101 (1)	263 (1)	247 (2)	133 (1)	132 (1)
18	163 (1)	201 (1)	259 (1)	102 (1)	187 (2)	173 (1)	22 (1)	253 (2)	74 (1)	248 (2)	39 (1)
19	255 (2)	16 (1)	201 (2)	98 (2)	174 (2)	224 (1)	161 (2)	161 (2)	251 (1)	192 (2)	22 (1)
20	42 (2)	184 (2)	255 (1)	123 (1)	173 (1)	100 (2)	74 (1)	74 (1)	159 (2)	265 (1)	244 (1)
21	172 (2)	71 (1)	39 (1)	71 (1)	224 (1)	209 (2)	246 (2)	159 (2)	161 (2)	74 (1)	74 (1)
22	7 (2)	93 (1)	173 (2)	263 (1)	166 (1)	96 (2)	241 (1)	246 (1)	253 (2)	182 (1)	264 (2)
23	240 (2)	38 (1)	16 (1)	222 (2)	201 (1)	253 (2)	65 (1)	258 (2)	133 (1)	49 (2)	95 (2)
24	32 (1)	245 (2)	2 (2)	255 (2)	249 (2)	261 (1)	244 (2)	65 (1)	49 (2)	250 (1)	166 (1)
25	5 (2)	241 (1)	163 (1)	93 (1)	215 (1)	74 (1)	173 (1)	49 (2)	265 (1)	159 (2)	222 (2)
26	75 (2)	252 (2)	5 (2)	197 (1)	83 (1)	161 (2)	96 (2)	102 (1)	248 (2)	131 (1)	49 (2)
27	98 (2)	187 (1)	164 (1)	224 (1)	197 (1)	256 (1)	159 (2)	265 (1)	263 (1)	161 (2)	256 (2)
28	257 (2)	164 (1)	197 (1)	173 (1)	261 (1)	264 (1)	246 (1)	264 (1)	65 (1)	245 (1)	252 (1)
29	15 (1)	221 (1)	6 (1)	137 (2)	96 (2)	257 (1)	256 (1)	256 (1)	182 (1)	253 (2)	192 (2)
30	218 (1)	144 (1)	95 (2)	57 (1)	137 (2)	65 (1)	264 (1)	133 (1)	266 (1)	65 (1)	159 (2)
31	260 (2)	75 (2)	102 (1)	95 (1)	71 (1)	3 (2)	224 (1)	96 (2)	259 (1)	132 (1)	65 (1)
32	139 (1)	108 (2)	252 (1)	192 (2)	256 (1)	187 (2)	261 (1)	21 (2)	264 (1)	264 (2)	130 (1)
33	16 (1)	2 (2)	131 (1)	261 (1)	264 (1)	197 (1)	49 (2)	101 (1)	21 (2)	263 (1)	161 (2)
34	209 (2)	15 (2)	137 (2)	242 (1)	123 (1)	174 (2)	255 (1)	173 (1)	264 (2)	256 (2)	250 (2)
35	15 (2)	172 (2)	100 (1)	6 (1)	95 (1)	22 (1)	21 (2)	251 (1)	16 (2)	264 (1)	131 (1)
36	147 (1)	2 (1)	2 (1)	96 (2)	253 (2)	137 (2)	184 (2)	247 (2)	256 (1)	244 (1)	253 (2)
37	123 (1)	28 (1)	216 (1)	216 (1)	74 (1)	201 (1)	265 (1)	16 (2)	250 (1)	21 (2)	264 (1)
38	189 (1)	76 (1)	95 (1)	166 (1)	230 (1)	83 (1)	258 (2)	261 (1)	132 (1)	130 (1)	74 (2)
39	222 (1)	42 (2)	263 (1)	230 (1)	98 (2)	230 (1)	16 (2)	248 (2)	245 (1)	223 (2)	21 (2)
40	174 (2)	218 (1)	224 (1)	256 (1)	65 (1)	164 (2)	164 (2)	224 (1)	256 (2)	165 (1)	223 (2)

The column specifies the Sa(0.8s) "stripe" level in units of g. The row refers to the record number within that stripe. For each row and column, the first number specifies the record (as numbered in Table A.5) and the number in parentheses specifies the horizontal component number (as numbered in Table A.5).

Appendix B

Supporting details for the correlation model of Chapter 8

In Chapter 8, a several statements were made about properties of spectral acceleration correlations that were not explained in detail in the text. In this Appendix, details to support those statements are presented for interested readers.

B.1 Correlations of geometric mean spectral acceleration values

It was stated in Chapter 8 that the correlation coefficient prediction for ln*Sa* values of arbitrary components with the same orientation could also be used to model correlation coefficients for geometric mean ln*Sa* values. Geometric mean spectral acceleration values are computed using the equation

$$Sa_{g.m.}(T) = \sqrt{Sa_x(T)Sa_y(T)} \tag{B.1}$$

or equivalently

$$\ln Sa_{g.m.}(T) = \frac{\ln Sa_x(T) + \ln Sa_y(T)}{2} \tag{B.2}$$

where $Sa_{g.m.}(T)$ is the geometric mean spectral acceleration at period T and $Sa_x(T)$ and $Sa_y(T)$ are the spectral acceleration values of the two horizontal components (denoted the x and y components) at period T.

In this section, we will show that the correlation coefficient between $\ln Sa_{g.m.}(T_1)$ and $\ln Sa_{g.m.}(T_2)$ is approximately equal to the correlation coefficient between $\ln Sa_x(T_1)$ and $\ln Sa_x(T_2)$, where T_1 and T_2 are two periods of interest.

B.1.1 Mathematical derivation

Let spectral acceleration values of the two horizontal components be defined as follows

$$\begin{aligned} \ln Sa_x(T) &= \mu(T) + \sigma(T)\varepsilon_x(T) \\ \ln Sa_y(T) &= \mu(T) + \sigma(T)\varepsilon_y(T) \end{aligned} \tag{B.3}$$

where $\mu(T)$ and $\sigma(T)$ are deterministic functions from a ground motion prediction model that represent the mean and standard deviation, respectively, of ln*Sa*. The variables $\varepsilon_x(T)$ and $\varepsilon_y(T)$ are random variables with zero mean and unit standard deviation that represent the record-to-record variability of ground motions, given their magnitude, distance, etc. Then the geometric mean of spectral acceleration is defined as

$$\ln Sa_{gm}(T) = \frac{\ln Sa_x(T) + \ln Sa_x(T)}{2} = \mu(T) + \sigma(T)\frac{\varepsilon_x(T) + \varepsilon_y(T)}{2} \tag{B.4}$$

The correlation coefficient of $\ln Sa_{gm}(T)$ at two periods is then

$$\begin{aligned}\rho_{\ln Sa_{gm}(T_1),\ln Sa_{gm}(T_2)} &= \frac{Cov\left[\dfrac{\varepsilon_x(T_1)+\varepsilon_y(T_1)}{2},\dfrac{\varepsilon_x(T_2)+\varepsilon_y(T_2)}{2}\right]}{\sqrt{Var\left[\dfrac{\varepsilon_x(T_1)+\varepsilon_y(T_1)}{2}\right]Var\left[\dfrac{\varepsilon_x(T_2)+\varepsilon_y(T_2)}{2}\right]}} \\ &= \frac{\dfrac{1}{4}\begin{bmatrix} Cov\left[\varepsilon_x(T_1),\varepsilon_x(T_2)\right] + Cov\left[\varepsilon_x(T_1),\varepsilon_y(T_2)\right] \\ +Cov\left[\varepsilon_y(T_1),\varepsilon_x(T_2)\right] + Cov\left[\varepsilon_y(T_1),\varepsilon_y(T_2)\right]\end{bmatrix}}{\sqrt{\dfrac{1+\rho_{\varepsilon_x,\varepsilon_y}(T_1)}{2}\dfrac{1+\rho_{\varepsilon_x,\varepsilon_y}(T_2)}{2}}}\end{aligned} \tag{B.5}$$

Recognizing that $Cov\left[\varepsilon_x(T_1),\varepsilon_x(T_2)\right] = Cov\left[\varepsilon_y(T_1),\varepsilon_y(T_2)\right] = \rho_{\varepsilon_x,\varepsilon_x}(T_1,T_2)$ and that $Cov\left[\varepsilon_x(T_1),\varepsilon_y(T_2)\right] = Cov\left[\varepsilon_y(T_1),\varepsilon_x(T_2)\right] = \rho_{\varepsilon_x,\varepsilon_y}(T_1,T_2)$, Equation B.5 simplifies to

$$\rho_{\ln Sa_{gm}(T_1),\ln Sa_{gm}(T_2)} == \frac{\rho_{\varepsilon_x,\varepsilon_x}(T_1,T_2) + \rho_{\varepsilon_x,\varepsilon_y}(T_1,T_2)}{\sqrt{\left(1+\rho_{\varepsilon_x,\varepsilon_y}(T_1)\right)\left(1+\rho_{\varepsilon_x,\varepsilon_y}(T_2)\right)}} \tag{B.6}$$

Using the approximation that $\rho_{\varepsilon_x,\varepsilon_y}(T_1)$ is a linear function of lnT_1 (per Equation 8.7), the denominator of Equation B.6 can be approximated by

$$\left(1+\rho_{\varepsilon_x,\varepsilon_y}(T_1)\right)\left(1+\rho_{\varepsilon_x,\varepsilon_y}(T_2)\right) \cong \left(1+\rho_{\varepsilon_x,\varepsilon_y}(\sqrt{T_1T_2})\right)^2 \tag{B.7}$$

Substituting B.7 into B.6 gives

$$\rho_{\ln Sa_{gm}(T_2),\ln Sa_{gm}(T_2)} =\cong \frac{\rho_{\varepsilon_x,\varepsilon_x}(T_1,T_2) + \rho_{\varepsilon_x,\varepsilon_y}(T_1,T_2)}{1+\rho_{\varepsilon_x,\varepsilon_y}(\sqrt{T_1T_2})} \tag{B.8}$$

Then we use the approximation

$$\rho_{\varepsilon_x,\varepsilon_y}(T_1,T_2) \cong \rho_{\varepsilon_x,\varepsilon_y}(\sqrt{T_1T_2}) \cdot \rho_{\varepsilon_x,\varepsilon_x}(T_1,T_2) \tag{B.9}$$

as was found to be valid for fitting Equation 8.11. Substituting B.9 into B.8 gives

$$\rho_{\ln Sa_{gm}(T_2),\ln Sa_{gm}(T_2)} \cong \frac{\rho_{\varepsilon_x,\varepsilon_x}(T_1,T_2)\left(1+\rho_{\varepsilon_x,\varepsilon_y}(\sqrt{T_1T_2})\right)}{1+\rho_{\varepsilon_x,\varepsilon_y}(\sqrt{T_1T_2})} \tag{B.10}$$

Canceling terms and recognizing that $\rho_{\varepsilon_x,\varepsilon_x}(T_1,T_2) = \rho_{\ln Sa_x,\ln Sa_x}(T_1,T_2)$, we have

$$\rho_{\ln Sa_{gm}(T_2),\ln Sa_{gm}(T_2)} \cong \rho_{\ln Sa_x,\ln Sa_x}(T_1,T_2) \tag{B.11}$$

Thus, given two assumptions that were found to be reasonable in Chapter 8, we can show that the correlation coefficient between $\ln Sa_{g.m.}(T_1)$ and $\ln Sa_{g.m.}(T_2)$ is approximately equal to the correlation coefficient between $\ln Sa_x(T_1)$ and $\ln Sa_x(T_2)$.

B.1.2 Empirical evidence

To support the mathematical result of Section B.1.1, we also see empirically that correlation coefficients between $\ln Sa_{g.m.}(T_1)$ and $\ln Sa_{g.m.}(T_2)$ are approximately equal to the correlation coefficients between $\ln Sa_x(T_1)$ and $\ln Sa_x(T_2)$. The contours of the correlation coefficients of a single ground motion component as a function of T_1 and T_2 are displayed in Figure B.1 (this is the same as Figure 8.4a, but is repeated here to facilitate comparison with the figure to follow). The contours for correlation coefficients of geometric mean spectral acceleration values are displayed in Figure B.2. The two figures are nearly identical.

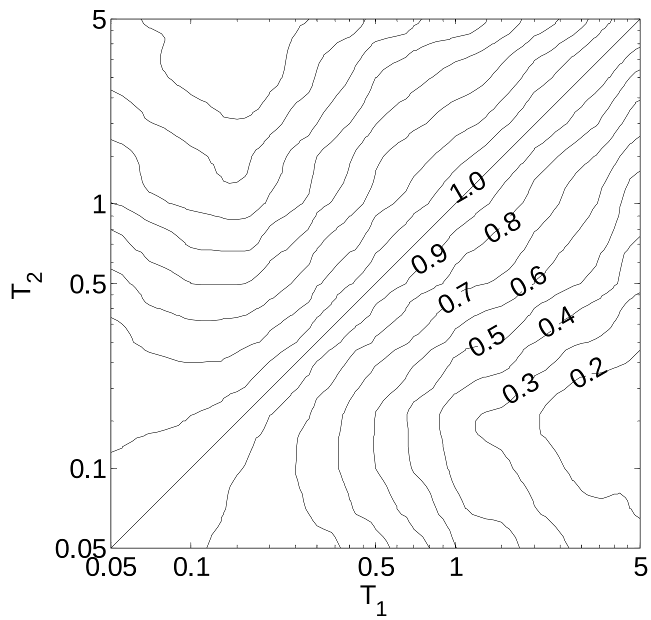


Figure B.1: Empirical spectral acceleration correlation contours for single ground motion components at two periods.

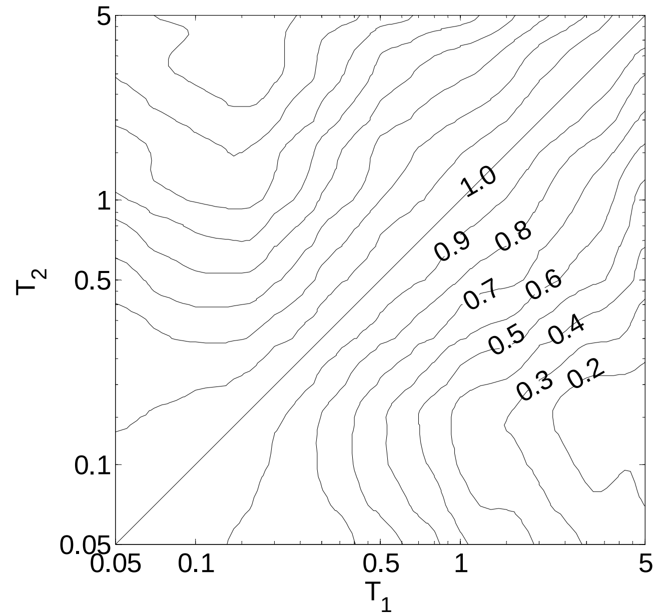


Figure B.2: Empirical spectral acceleration correlation contours for geometric means at two periods.

Another comparison of the correlation coefficient estimates for arbitrary component and geometric mean spectral acceleration values is seen in Figure B.3. For each period pair considered (i.e., each location on Figure B.1 and Figure B.2), the two empirical correlation coefficients are plotted versus each other. We see that the agreement is very good.

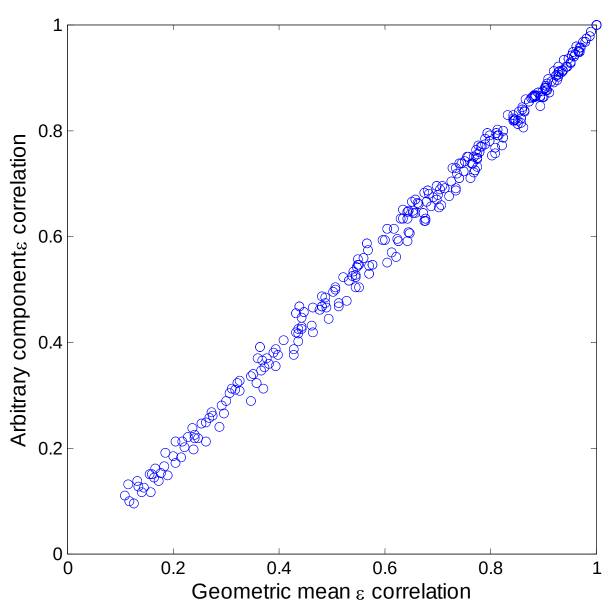


Figure B.3: Comparison of correlation coefficients for geometric mean ε values versus correlation coefficients for arbitrary component ε values.

Given the results of Sections B.1.1 and B.1.2, it seems likely that any minor differences between estimated correlation coefficients are due to the finite samples of data used for the estimates, rather than any underlying differences.

B.2 Correlations of inter-event epsilon values

In Chapter 8 it was also stated that the equation for correlations of spectral acceleration values of a single component can be used to approximate the correlation of inter-event ε values. This statement will now be examined in more detail. For this study, the inter-event terms from the Abrahamson and Silva (1997) attenuation model were used, as provided by Abrahamson (2004). The original set of inter-event terms provided by Abrahamson included events with magnitudes beween 4.4 and 7.4, but only the events with magnitude greater than 5.5 were retained here (no events with magnitude less than 5.5 were used for the other correlation computations, so for consistency they are not used here either).

Contours of the correlation matrix for these terms are shown in Figure B.4. It is clear that there are differences between Figure B.4 and Figure B.1.

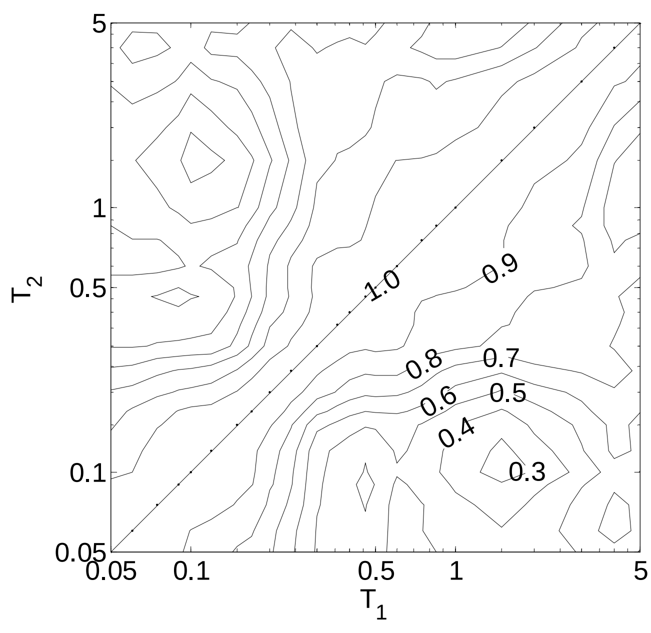


Figure B.4: Empirical correlation contours for inter-event ε values.

Another comparison of the empirical correlation coefficient estimates is seen in Figure B.5. For each period pair considered, the correlation coefficient estimates for the two ε values are plotted versus each other. We see that the agreement between the two is not perfect, and is certainly not as good as the agreement for geometric mean *Sa* values seen earlier in Figure B.3.

There are several possible explanations for this. First, while the same record sets were used earlier for both the geometric mean and arbitrary component correlations, here we do not have that luxury. The inter-event ε values were provided by Abrahamson, and so the events used are not identical. Second, there are a much smaller number of data points available for estimating correlations of inter-event ε values (only 36 events are available, versus 534 total horizontal components used for arbitrary component ε values), which decreases the confidence with which the correlation coefficients can be estimated. Third, and perhaps most importantly, there may be physical reasons why the two sets of correlations are in fact different.

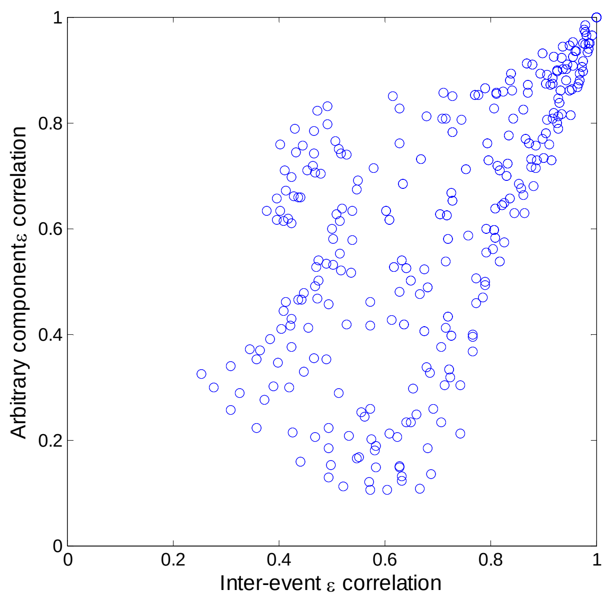


Figure B.5: Comparison of correlation coefficients for inter-event ε values versus correlation coefficients for total ε values.

To determine the true extent of the differences seen in Figure B.5, statistical hypothesis testing was used to determine the significance of apparent differences in the two correlation coefficient estimates (Neter et al. 1996, p644). First the z-transform was taking for each correlation coefficient estimate, in order to stabilize its variance (see Equation 8.5). The test statistic is then

$$z^* = \frac{z_1 - z_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}} \qquad \text{(B.12)}$$

where z_1 and z_2 are the z-transformed correlation coefficients for the inter-event ε values and the arbitrary component ε values, respectively, and n_1 and n_2 are the number of data points used for estimating each (usually 36 and 534, although the number is slightly lower for the inter-event ε

values at a few periods, because not all events were used at all periods due to filter-frequency limitations). If the two correlation coefficients are equal, and as long as n_1 and n_2 are greater than about 25 (which is true here), then z^* is distributed approximately as a standard normal random variable. Hence, if $|z^*|>1.96$, we can conclude that the two correlation coefficients differ with a 5% significance level[21]. This criterion allows us to identify the region of Figure B.5 where apparent differences in the two correlation coefficients are not statistically significant. Setting n_1=36 and n_2=534, we have the criteria

$$\left| \frac{\frac{1}{2}\ln\left(\frac{1+\rho_1}{1-\rho_1}\right)-\frac{1}{2}\ln\left(\frac{1+\rho_2}{1-\rho_2}\right)}{\sqrt{\frac{1}{36-3}+\frac{1}{534-3}}} \right| < 1.96 \tag{B.13}$$

which simplifies to

$$\left| \ln\left(\frac{1+\rho_1}{1-\rho_1}\frac{1-\rho_2}{1+\rho_2}\right) \right| < 0.703 \tag{B.14}$$

The region where this inequality holds is shown in Figure B.6, with the data from Figure B.5 superimposed. If the correlation coefficients for inter-event ε values were in fact equal to correlation coefficients for total ε values, we would expect to see 95% of the points inside the shaded region and 5% outside. In the figure, however, 58% of the points are inside the shaded region, with 42% outside, indicating that significant differences between the two occur more often than expected. However, many of the statistically significant differences occur in the top right corner of the plot, where the difference in absolute correlation values is not large. For example, the difference between correlation coefficients of 0.9 and 0.95 is statistically significant, but may or may not have practical significance in application.

---

[21] Lack of statistical significance does not necessarily imply that the correlation coefficients are equal. It merely indicates that, given the limited data available, it is difficult to detect a significant difference between the two.

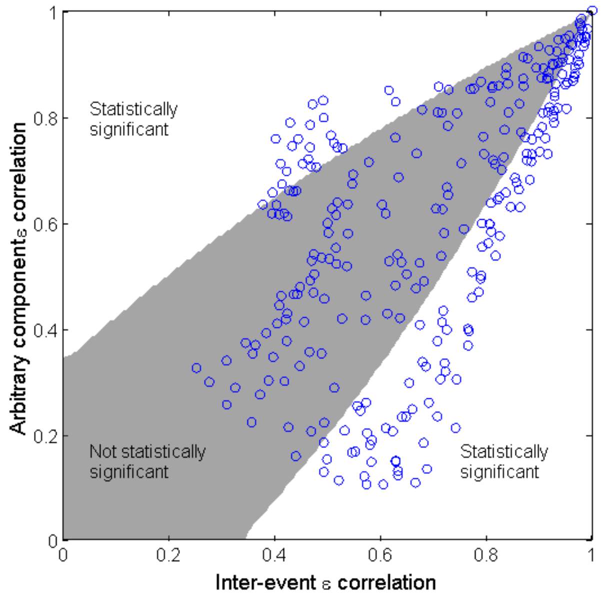


Figure B.6: Comparison of correlation coefficients for inter-event ε values versus correlation coefficients for total ε values, and the region with no statistical significance (at the 5% level) superimposed.

Given the small number of inter-event ε data available, no attempt was made to fit a new correlation model for the inter-event ε data. Using Equation 8.9 as an approximation for the correlation of inter-event ε values may be a reasonable approach in the absence of other models. In the future, a model specific to the inter-event ε data could be fit that would likely provide improved predictions.

B.3 Positive definiteness of correlation matrices

The correlation coefficient model of Chapter 8, when combined with marginal variances provided by the ground motion prediction (attenuation) model, specifies the covariance matrix of a vector of spectral acceleration values with varying periods and orientations. It is necessary that the variance of any linear combination of these spectral acceleration values be non-negative, which leads to the requirement that the covariance matrix be positive definite. It can be shown that the covariance matrix is positive definite if and only if the correlation matrix is positive definite. So to confirm that the specified correlation model is permissible, it suffices to shown that corresponding correlation matrices are positive definite.

A common method for determining the positive-definiteness of a matrix $\mathbf{A}$ involves attempting to factor the matrix into the form $\mathbf{A} = \mathbf{R}^T\mathbf{R}$, where $\mathbf{R}$ is an upper triangular matrix with positive diagonal entries. The factorization is called a Cholesky decomposition, and is possible if and only if **A** is positive definite (Lay 1997, p456).

Correlation matrices were created for each of the correlation models proposed in Chapter 8. For each model (e.g., horizontal/horizontal in the same direction, horizontal/vertical, etc.), a 75x75 correlation matrix was formed by evaluating the appropriate predictive equation at 75 periods between 0.05 and 5 seconds. The Cholesky decomposition was performed for each of these matrices and was found to exist. Further, a 225x225 correlation matrix was constructed consisting of correlations for spectral values at 75 periods in all three directions (two horizontal and one vertical) simultaneously. The Cholesky decomposition also exists for this matrix.

The Cholesky decomposition was attempted and found to exist for the correlation matrix computed using the model of Inoue and Cornell (1990), but it does not exist for the correlation matrix computed using the model of Abrahamson et al. (2003).

B.4 Correlation as a function of magnitude or distance

The proposed correlation coefficient predictions are functions of the two periods considered and the orientations of the ground motion components considered. The possibility exists that they are also functions of the magnitude or distance of the ground motion. In order to consider this possibility, records were selected within magnitude or distance bins, and the calculated correlation coefficient from each bin was compared to the original correlation coefficient calculated using all of the records.

The results are displayed graphically in Figure B.7 below. Here, correlation coefficients are computed at a period of 1s for opposite horizontal components and for horizontal versus vertical components. The correlation coefficients are computed using all records, and this coefficient is compared to correlation coefficients selected using only those records that fall within a magnitude or distance bin (+/-0.2 magnitude units or +/-10km are used as the bin limits). In Figure B.8 and Figure B.9, correlation coefficients are displayed for opposite horizontal components versus magnitude and distance, respectively, for a range of considered periods. No systematic trends are seen in any of these figures, indicating that the predictive equations do not need to include terms for the magnitude or distance of the event.

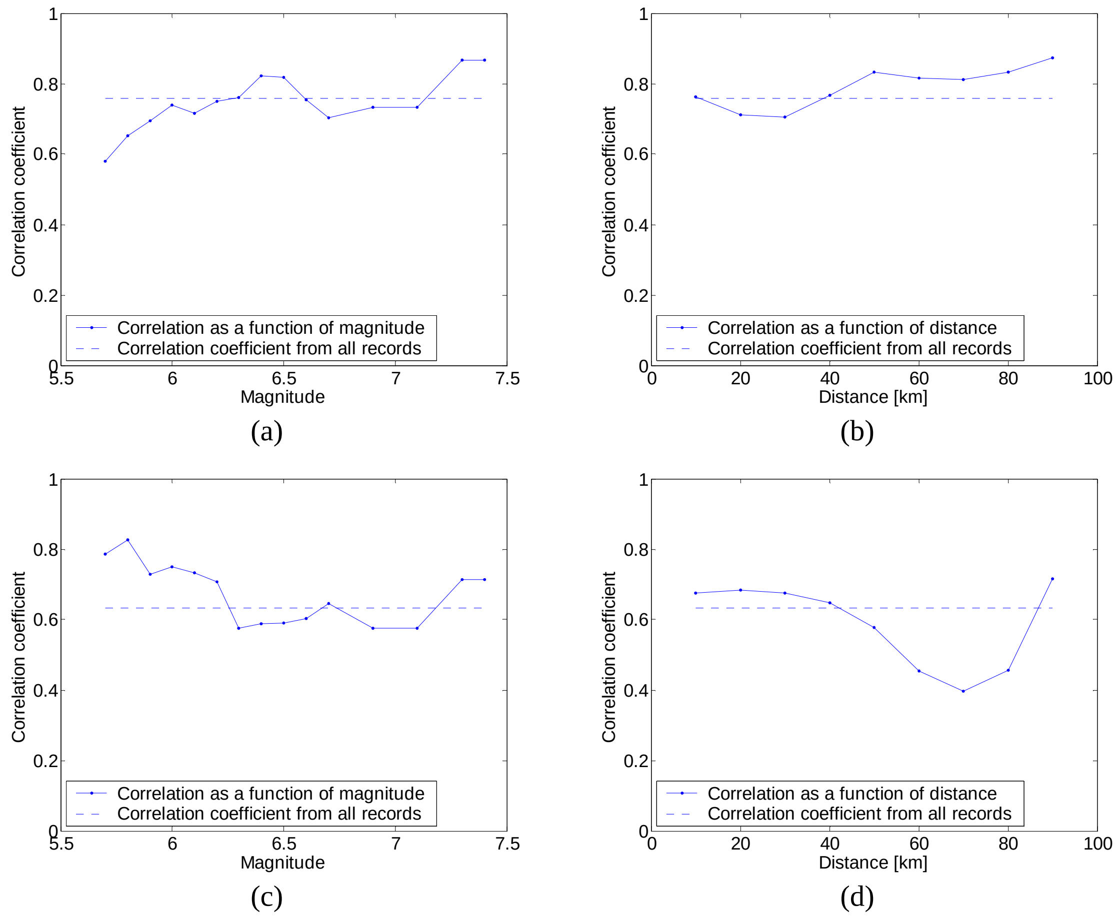


Figure B.7: Correlation coefficients for ε values at a period of 1 second as a function of magnitude and distance. (a) Opposite horizontal components versus magnitude. (b) Opposite horizontal components versus distance. (c) Horizontal/vertical components versus magnitude. (d) Horizontal/vertical components versus distance.

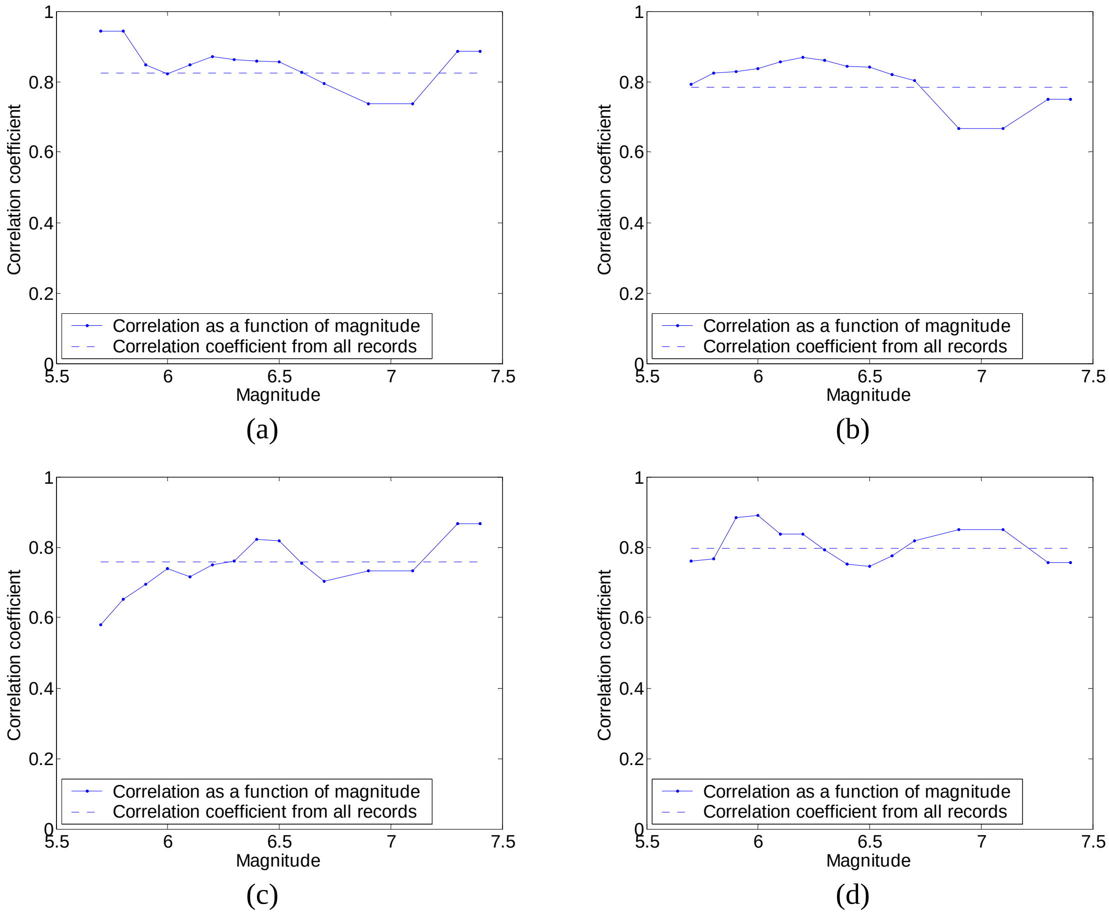


Figure B.8: Correlation coefficients for ε values between opposite horizontal components as a function of magnitude for several periods. (a) Period = 0.05s. (b) Period = 0.2s. (c) Period = 1s. (d) Period = 5s.

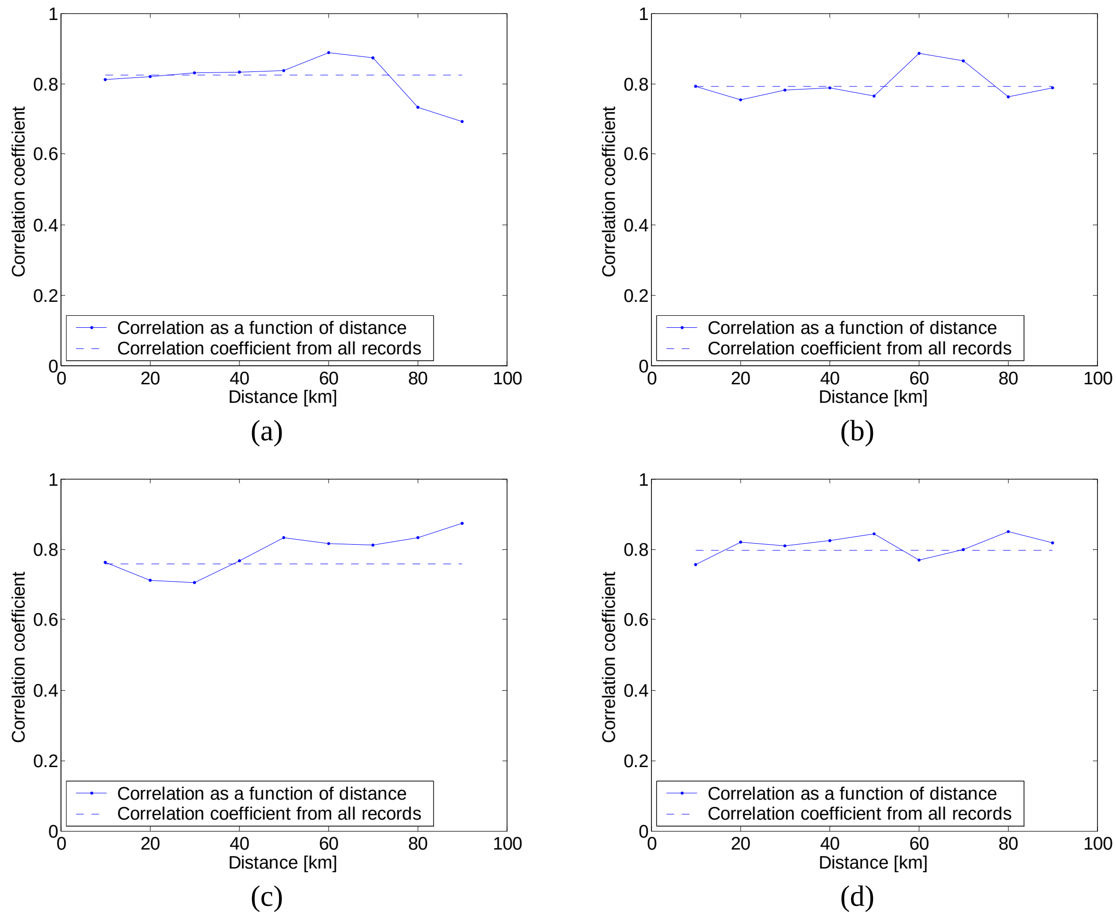


Figure B.9: Correlation coefficients for ε values between opposite horizontal components as a function of distance for several periods. (a) Period = 0.05s. (b) Period = 0.2s. (c) Period = 1s. (d) Period = 5s.

The above computations were performed for the case when the two periods of interest are equal. The same computations could be repeated for correlation coefficients at two differing periods. However, the number of possible combinations is very large. The procedure was repeated for a small number of period combinations, and again no trends were seen. Because of the lack of trends seen in any of these tests, no systematic investigation of all period pairs was performed. At this point it was concluded that there is no detectible trend in correlation coefficients as a function of either magnitude or distance.

B.5 Modification of ground motion prediction (attenuation) models

Throughout this dissertation, it was stated that typical ground motion predictions for geometric mean spectral accelerations can be converted into predictions for arbitrary components of ground motion, given knowledge of correlations of the two horizontal components. The importance of distinguishing between the two was discussed in Chapter 6, and the equation for the conversion was given in (8.19)

$$\sigma_H^2(M,T) = \frac{2\sigma_{g.m.}^2(M,T)}{1+\rho_{\varepsilon_x,\varepsilon_y}} \qquad (8.19)$$

where $\sigma_H(M,T)$ is the standard deviation of a single component spectral acceleration (which is typically a function of magnitude and the period considered), $\sigma_{g.m.}(M,T)$ is the standard deviation of geometric mean spectral acceleration, and $\rho_{\varepsilon_x,\varepsilon_y}$ is the correlation between two orthogonal horizontal components at the period *T*.

Given this equation for conversion, the Abrahamson and Silva (1997) ground motion prediction model can be easily modified for prediction of arbitrary components. The only change necessary is the modification of the b_5 and b_6 components for the standard deviations, presented in Table 4 of Abrahamson and Silva (1997). The replacement values for these coefficients are given in Table 3.1, based on the correlation model of Chapter 8.

Table B.1. Coefficients for the standard deviation of arbitrary component standard deviations, to be used with the Abrahamson and Silva (1997) ground motion prediction model.

Period	b_5	b_6
0.01	0.72	0.139
0.02	0.72	0.139
0.03	0.72	0.139
0.04	0.73	0.139
0.05	0.74	0.139
0.06	0.75	0.139
0.075	0.76	0.139
0.09	0.76	0.139
0.10	0.77	0.140
0.12	0.78	0.140
0.15	0.79	0.140
0.17	0.80	0.140
0.20	0.80	0.141
0.24	0.81	0.141
0.30	0.82	0.141
0.36	0.83	0.142
0.40	0.84	0.139
0.46	0.84	0.137
0.50	0.85	0.134
0.60	0.85	0.134
0.75	0.86	0.130
0.85	0.87	0.128
1.00	0.87	0.125
1.50	0.89	0.117
2.00	0.91	0.112
3.00	0.93	0.104
4.00	0.94	0.099
5.00	0.96	0.094

Appendix C

Accounting for near-fault effects in ground motion prediction

C.1 Modification of mean predictions

As discussed in Section 5.5.1, ε values should be computed with care for pulse-like ground motions. Particularly when one is concerned about the occurency of velocity pulses, one should use a ground motion prediction model that accounts for pulse-like ground motions explicitly. Most prediction models in use today do not account for this effect, unfortunately. However, Somerville et al. (1997)[22] have proposed a modification to the Abrahamson and Silva (1997) prediction model that accounts for near-fault effects. The Somerville et al. model includes modifications for both fault-normal ground motion components and average components. Here the modification for fault-normal components is used, because all of the records used here are known to be oriented in the fault-normal direction.

In Figure C.1, the effect of the near-fault modification is illustrated. The 5% damped response spectrum of the Lucerne recording from the 1992 Landers earthquake is shown. In addition, the predicted median response spectra are plotted using the Abrahamson and Silva (1997) model, both with and without near-fault modifications. It can be seen that for this ground motion, the near-fault modification has the effect of raising the median prediction for periods greater than 0.6s. It should be noted that this near-fault modification may also lower the median prediction if, for example, the fault geometry suggests that the rupture is moving away from the site and

[22] For this work, the model of Bozorgnia (Bozorgnia and Bertero 2004, Chapter 5) is used, which is based on the Somerville 1997 model. The Somerville 1997 model was originally adapted for prediction by Abrahamson (2000), who tapered the modification off as the magnitude decreased or the distance increased. Bozorgnia repeats a description of the Abrahamson model, but corrects mistakes that are present in the Abrahamson manuscript.

negative directivity effects are likely to be present. For the record set considered here, however, the modification typically increases the median prediction because the records have been preferentially selected have positive directivity effects. The modification also modifies the standard deviation of the prediction. Because ε values are computed as deviations of the record response spectrum from the predicted response spectrum, it can be seen that the Lucerne record's ε values will change at periods larger than 0.6 seconds.

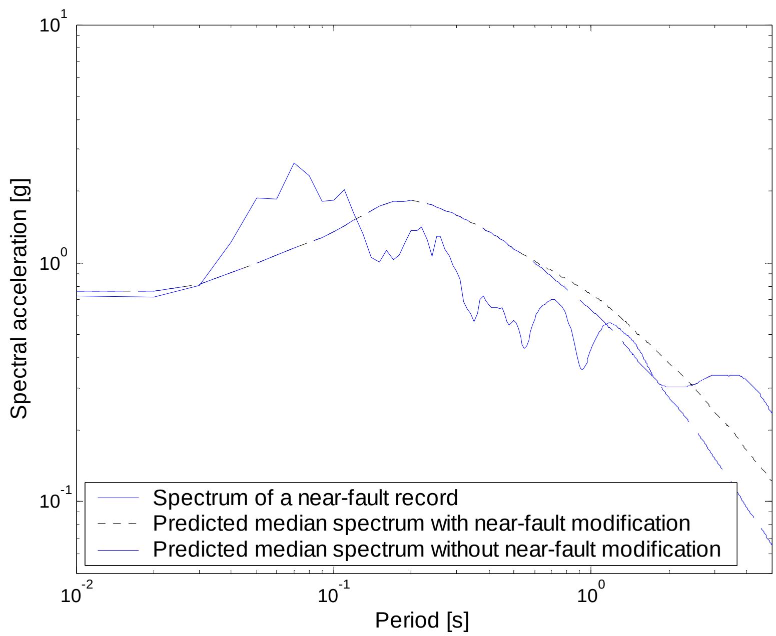
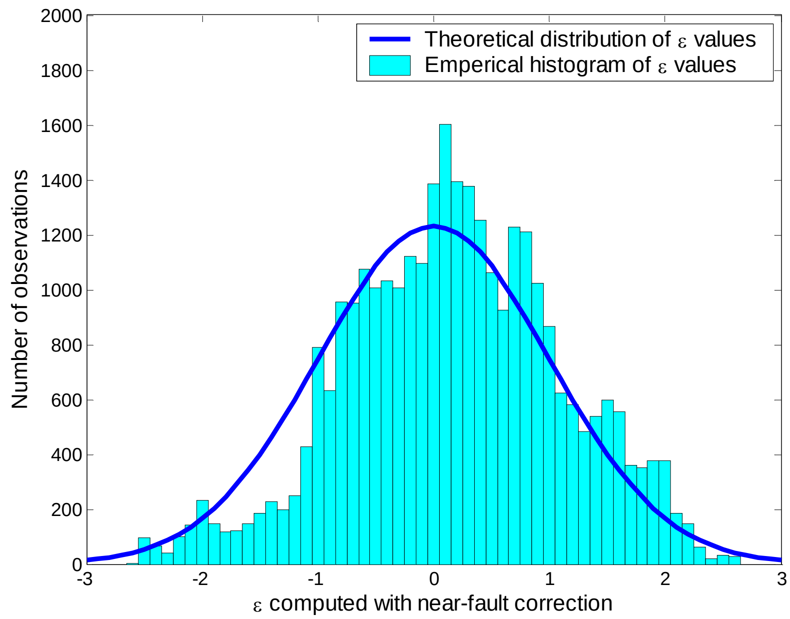
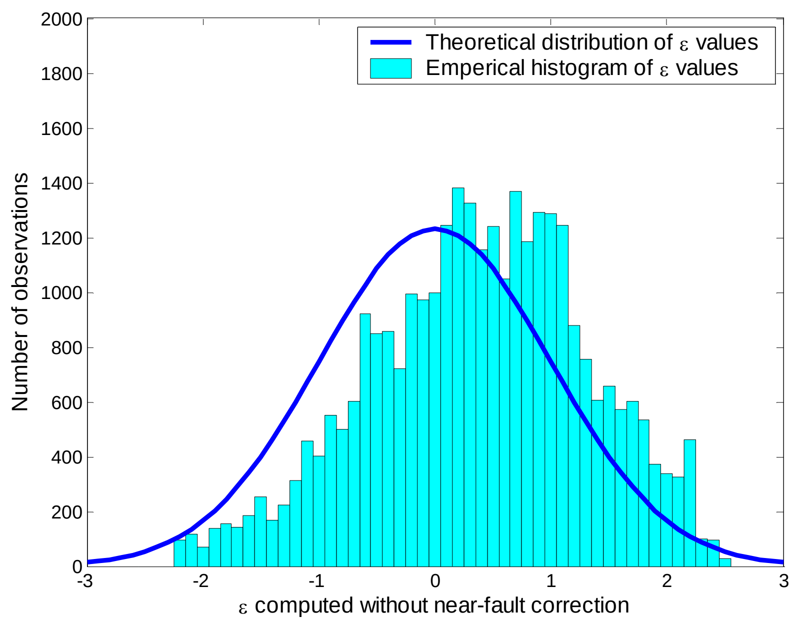


Figure C.1. The response spectrum of the Lucerne recording from the 1992 Landers earthquake, along with predicted median spectra from ground motion prediction models with and without near-fault effects accounted for.

The near-fault modification should improve the accuracy of the ground motion predictions. This can be tested by looking at the residual (ε) values computed from the predictions with and without the near-fault modification. The ε values should theoretically have a standard normal distribution. This can be checked by examining histograms of the ε values from the 70 pulse-like ground motions. These histograms are shown in Figure C.2. Epsilon values are computed for each record using the model of Abrahamson and Silva (1997), both with and without modification for near-fault effects. For each record, ε values were computed from 0.6 to 5.0 seconds, in increments of 0.01 seconds. The ε values for all records and all periods are pooled and used to generate the histograms of Figure C.2. The theoretical distribution is also superimposed for comparison. In Figure C.2b, the histogram tends to fall to the right of the theoretical distribution, suggesting that the unmodified prediction equation is biased for this specific class of records (it tends to under-predict the spectral acceleration values of these pulse-like records). In Figure C.2a, the match between the theoretical distribution and the empirical histogram is improved, indicating that much of the bias has been removed. Although the match is not perfect in Figure C.2a, it is certainly better than in Figure C.2b. Thus, the near-fault modification has improved the fit of the ground motion prediction for near-fault records.



(a)



(b)

Figure C.2. Histograms of ε values computed for the 70 near-fault ground motions at periods from 0.6 seconds to 5.0 seconds, both with (a) and without (b) modification to account for near-fault effects.

C.2 Correlations among spectral values at varying periods

In order to perform vector-valued hazard analysis for the vector IM consisting of $Sa(T_1)$ and $R_{T1,T2}$, it is not enough to modify the mean and standard deviations to account for near-fault effects. One must also know the correlation coefficient between $\ln Sa(T_1)$ and $\ln Sa(T_2)$ values for near-fault ground motions. A correlation coefficient model for ordinary ground motions is presented in Chapter 8, but the unique characteristics of pulse-like ground motions suggest that perhaps their spectral values have different correlation coefficients than ordinary ground motions.

Empirical correlation coefficients were estimated for the 70 near-fault ground motions used in Chapter 5, using the method of Chapter 8 and using the ground motion prediction model discussed in the previous section. Contours of these correlation coefficients as a function of the two periods are shown in Figure C.3, along with the correlation coefficients from ordinary ground motions, for comparison. The near-fault ground motions appear to have lower correlation coefficients for a given T_1 and T_2, and even have negative correlation coefficients for some widely separated period pairs. However, for period ratios likely to be considered (e.g., the T_2=T_1*2 suggestion in Chapter 5), the correlations are in fact not significantly different. This suggests that if the period pairs are not widely separated, the correlation model of Chapter 8 may be a reasonable approximation. The preferred approach for pulse-like ground motions, however, would be to take correlations from the empirical results shown here (this is especially important when the period pairs are widely separated). In the future, a predictive equation could be fit to this data from pulse-like records by using the method of Chapter 8.

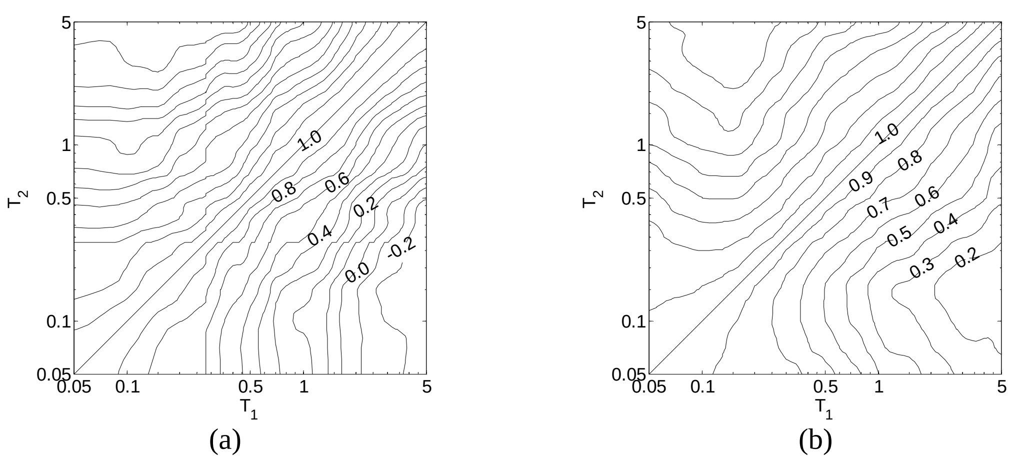


Figure C.3. Empirical correlation contours for horizontal spectral acceleration values in the same direction at two periods (T_1 and T_2) for (a) pulse-like and (b) ordinary ground motions.

Appendix D

Fitting a lognormal distribution for collapse capacity, when different records are used at each IM level

A useful structural response quantity to estimate from dynamic analyses is the probability of collapse as a function of IM (typically *Sa* at the first-mode period of the structure). This result can then be combined with a ground motion hazard to compute the mean annual rate of structural collapse (e.g., Shome and Cornell 1999, FEMA 2000a, Ibarra 2003).

In most work using this method, the collapse capacity can be found by repeatedly scaling a ground motion (i.e., using incremental dynamic analysis) until the ground motion causes collapse of the structure. Using this method, each ground motion has a single *Sa* value associated with its collapse. By repeating this process for a set of ground motions, one can obtain a set of *Sa* values associated with the onset of collapse, as illustrated in Figure D.1. The probability of collapse at a given *Sa* level, *Sa**, can then be estimated as the fraction of records for which collapse occurs at a level lower than *Sa**. A plot of this estimate is shown in Figure D.2. The distribution of *Sa* levels which cause the structure to collapse is often assumed to be lognormal, and so a lognormal distribution is fitted to the results. This can be done by taking logarithms of each *Sa* value associated with collapse of a record. The mean and standard deviation of the log *Sa* values can then be calculated, using e.g., the method of moments (Rice 1995), and converted in the parameters of the lognormal distribution (see, e.g., Ibarra 2003). Alternatively, counted fractiles could be used to estimate the mean and standard deviation. The resulting fitted distribution is also shown in Figure D.2.

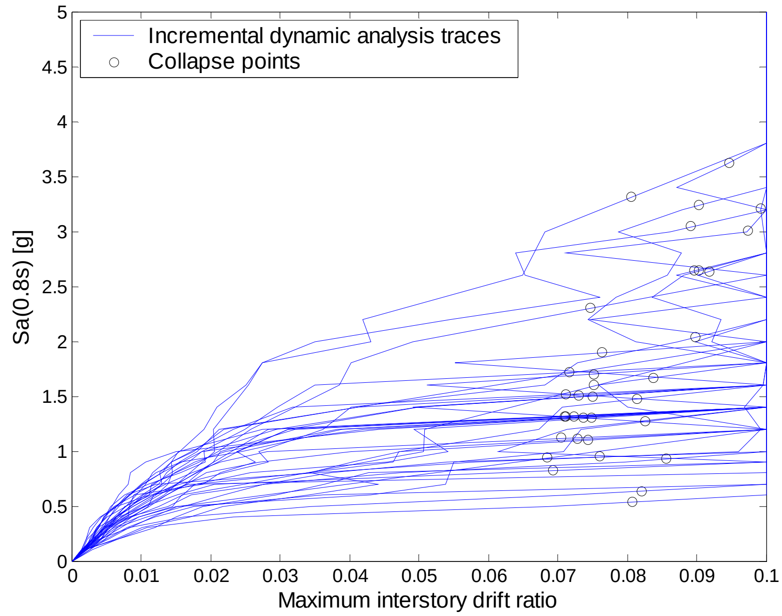


Figure D.1: Incremental dynamic analyses of the Van Nuys building, used to identify *Sa* values associated with collapse for each record.

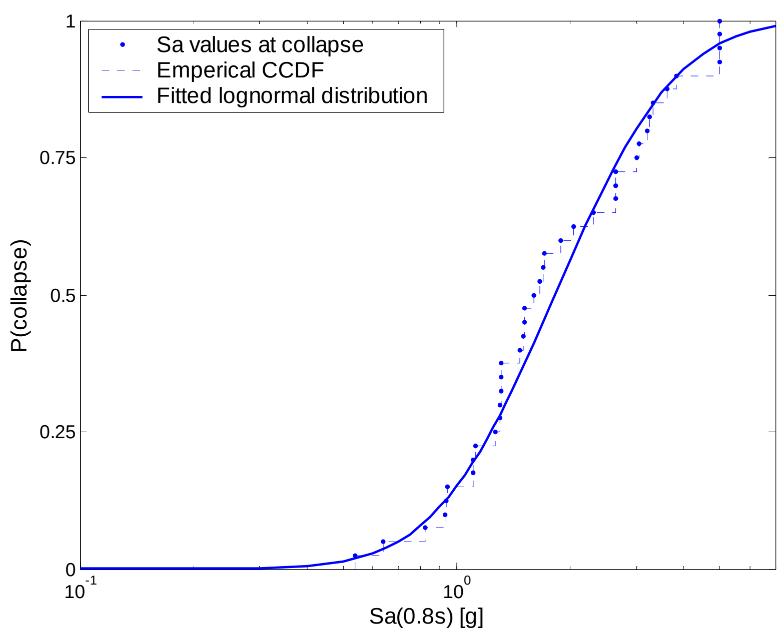


Figure D.2: Probability of collapse using an empirical CCDF and a fitted lognormal CCDF, for a set of records.

The above method is straightforward, but it cannot be used to estimate the probability of collapse for Figure 6.11 of Chapter 6. This is because different records are used at each *Sa* level, and so the incremental dynamic analysis approach illustrated in Figure D.1 cannot be used.

Instead of an *Sa* value associated with the onset of collapse for each record, we have, for each *Sa* level, the fraction of records at that level that caused collapse. This is illustrated in Figure D.3. For this figure, probabilities of collapse were obtained from records specially selected to match target epsilon values at each *Sa* level (i.e., the records selected using Method 3 in Chapter 6).

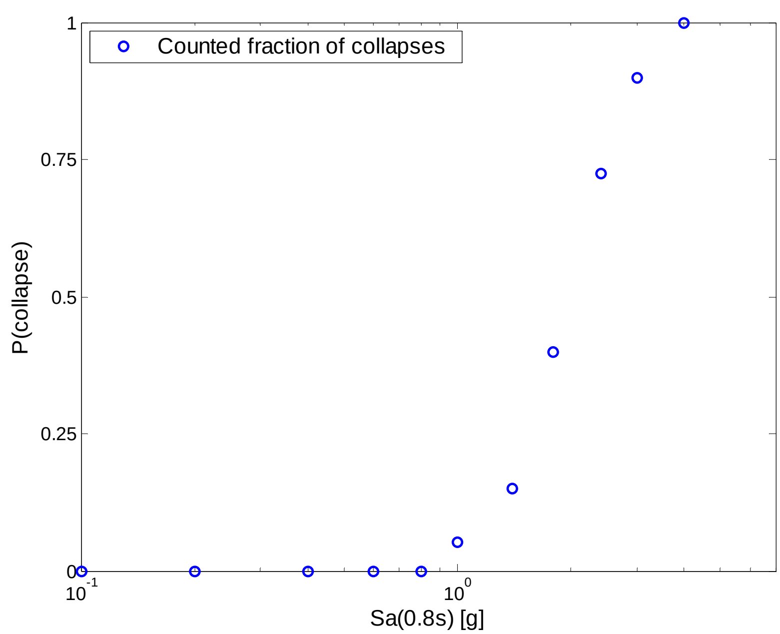


Figure D.3: Empirical probability of collapse for a set of *Sa* values, obtained from records selected to match target epsilon values at each *Sa* level.

It is possible to fit a lognormal distribution to the observations of Figure D.3, but the process is slightly more difficult. The goal is to identify the lognormal distribution parameters so that the fitted distribution predicts probabilities is as close to the observed probabilities of collapse as possible. To illustrate, consider the following. The observed probability of collapse at level Sa_i is

$$P(C\,|\,Sa_i)_{observed} = \frac{\#\text{ of collapses at } Sa \text{ level } Sa_i}{\text{total number of records}} \tag{D.1}$$

and the predicted probabilities of collapse are

$$P(C\,|\,Sa_i)_{pred} = 1 - \Phi\left(\frac{\ln Sa_i - \hat{\mu}_{\ln Sa_{Cap}}}{\hat{\sigma}_{\ln Sa_{Cap}}}\right) \tag{D.2}$$

where $\hat{\mu}_{\ln Sa_i}$ and $\hat{\sigma}_{\ln Sa_i}$ are the parameters of the estimated lognormal distribution. In the basic case above, $\hat{\mu}_{\ln Sa_i}$ and $\hat{\sigma}_{\ln Sa_i}$ could be estimated directly from the IDA results (i.e., the data of Figure D.1). When the records are changing at each *Sa* level, however, another method is needed.

One method for estimating these parameters would be to minimize the total squared errors between the estimated probability of collapse and the observed probability of collapse over all of the *Sa* levels considered. That is,

$$\begin{aligned}\{\hat{\mu}_{\ln Sa_{Cap}}, \hat{\sigma}_{\ln Sa_{Cap}}\} &= \min_{\mu,\sigma} \sum_i \left(P(C \mid Sa_i)_{observed} - P(C \mid Sa_i)_{pred} \right)^2 \\ &= \min_{\mu,\sigma} \sum_i \left(P(C \mid Sa_i)_{observed} - 1 + \Phi\left(\frac{\ln Sa_i - \mu}{\sigma} \right) \right)^2\end{aligned} \tag{D.3}$$

This is a straightforward nonlinear optimization that can be easily performed with a variety of algorithms, taking care to verify (usually visually) that the minimum is actually a global minimum (and not just a local minimum).

However, the least-squares method ignores a fundamental property of the data: the variance of the observations is non-constant. That is, if zero collapses are observed at a given *Sa* level and the fitted probability of collapse is 0.1, then this error is much larger than fitting a probability of collapse of 0.6 at an *Sa* level where 50% collapses are observed. In order to account for the non-constant variance, the method of maximum likelihood can be applied (e.g., Rice 1995). This estimate is then

$$\{\hat{\mu}_{\ln Sa_{Cap}}, \hat{\sigma}_{\ln Sa_{Cap}}\} = \max_{\mu,\sigma} \sum_i \left(1 - \Phi\left(\frac{\ln Sa_i - \mu}{\sigma} \right) \right)^{n_i} \Phi\left(\frac{\ln Sa_i - \mu}{\sigma} \right)^{N - n_i} \tag{D.4}$$

where n_i is the number of collapses observed at *Sa* level Sa_i, and N is the total number of records analyzed at level Sa_i. This method was observed by the author to be more sensitive to local minima than the above least squares method, at least for the examples considered in this dissertation. However, an equivalent way to estimate the lognormal distribution using maximum likelihood is to use generalized linear regression with a Probit link function (Agresti 2002). This is because generalized linear regression uses maximum likelihood as its optimization scheme. This regression function is available in many statistical software packages. The inputs to the function are the number of records at each *Sa* level and number of collapses at each *Sa* level (the “y” data from Figure D.3), along with the associated *Sa* level (the “x” data). The Probit model fits a normal CDF, and so the *Sa* values should actually be input as ln*Sa* values in order to obtain a lognormal CDF in terms of *Sa*. Estimates obtained using this scheme were observed to always be at the global minimum, unlike the manually implemented maximum likelihood method. An illustration of the fitted distribution is displayed in Figure D.4.

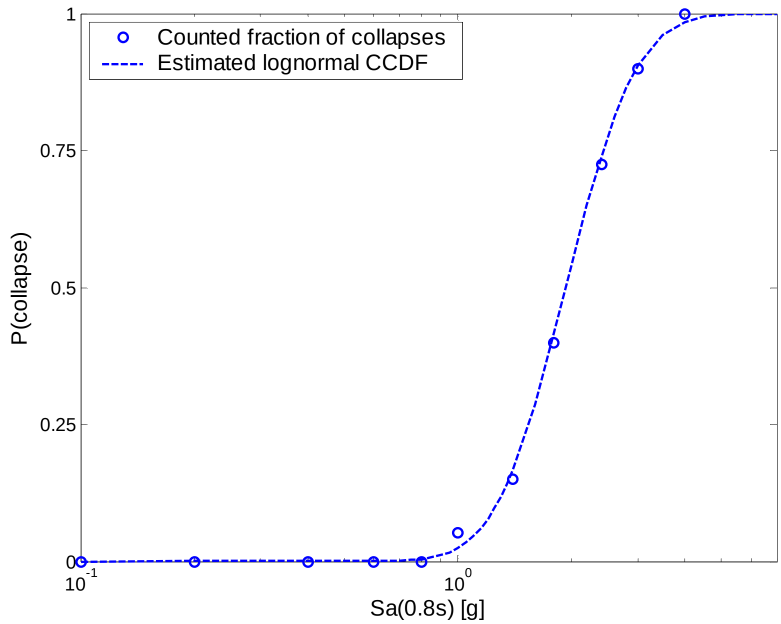


Figure D.4: Empirical probability of collapse and a lognormal distribution fitted using generalized linear regression with a Probit link function.

This method was used to obtain estimated the probabilities of collapse in Figure 6.11 of this dissertation. Note that although this method is slightly more complicated than the standard method used elsewhere, it is also more flexible. For example, if one is primarily interested in the lower tail of the distribution (as is often the case, because lower *Sa* values occur much more frequently), then only this portion of the data need be obtained for fitting. For example, *Sa* "stripes" could be analyzed for increasing *Sa* levels until a certain fraction (e.g. 40% or 50%) of the records cause collapse. Then the analysis can be stopped, and the distribution fitted without knowledge of the effect of larger *Sa* values. This may be advantageous when structural analyses are expensive.

Appendix E

Use of mean disaggregation values to model target spectra

The mean values of *M*, *R* and ε (referred to here as $\bar{M}$, $\bar{R}$ and $\bar{\varepsilon}$) from disaggregation have been used to develop target spectra in Chapter 6. Strictly speaking, however, there are a range of (*M*, *R*, ε) values that could result in occurrence of the target $Sa(T_1)$ value. For example, a disaggregation obtained from the U.S. Geological Survey (2002) is shown in Figure E.1 for the Van Nuys, California site used as an example throughout this dissertation; this disaggregation is for the *Sa*(1s) value exceeded with 2% probability in 50 years. A description of the seismicity at this site is provided by Somerville (2001b). The mean (*M*, *R*, ε) values are (6.8, 9.1km, 1.44), but the plot indicates that a range magnitude, distance and ε values contribute.

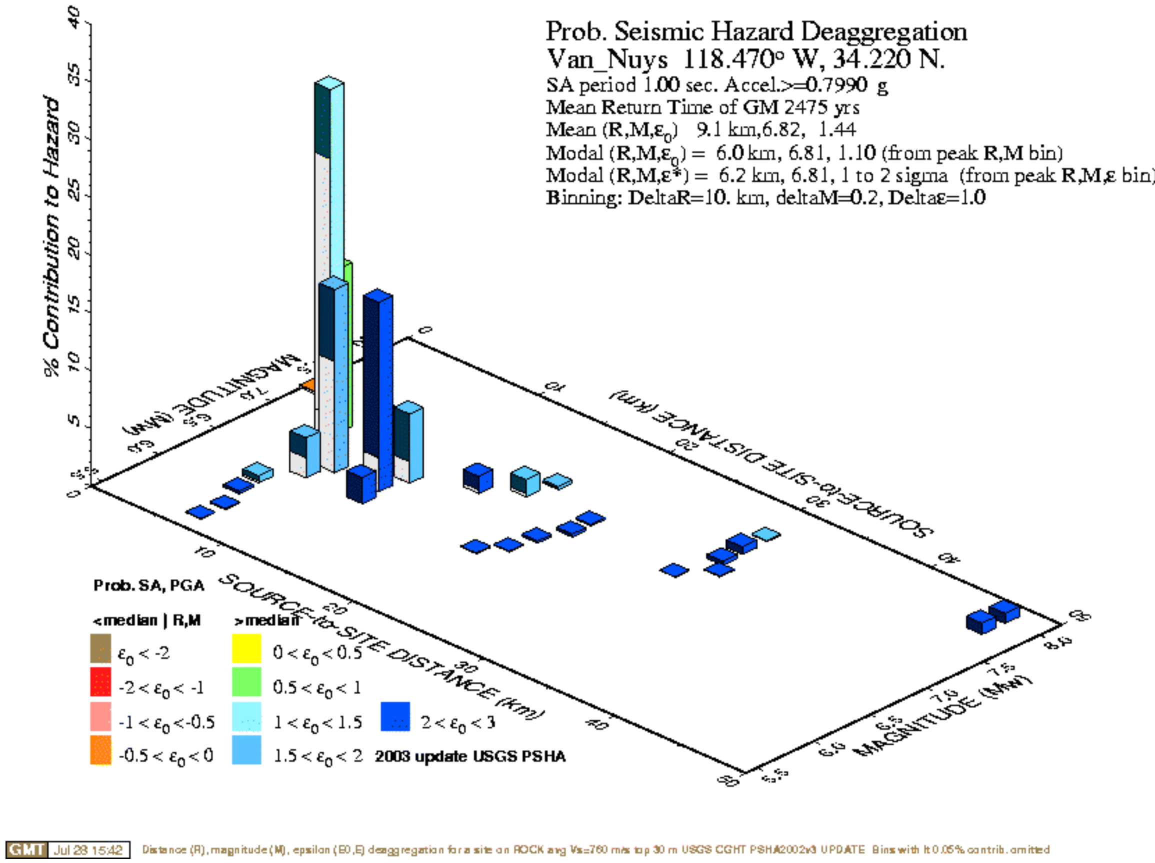


Figure E.1: Example disaggregation for Sa(1.0s) at the Van Nuys site discussed in Chapter six (USGS 2002).

In order to develop the true distribution of conditional spectra, one should consider a weighted combination of conditional spectra from each possible (M, R, ε) triplet, where the weights are assigned according to the probability of that triplet causing the target $Sa(T_1)$ value to be exceeded (these probabilities are readily available from standard PSHA disaggregation). Consider n possible (M, R, ε) values, denoted (M_i, R_i, ε_i), with associated probability p_i. That is

$$P(M = M_i, R = R_i, \varepsilon = \varepsilon_i) = p_i, \quad \text{where} \sum_{i=1}^{n} p_i = 1 \tag{E.1}$$

We can then use conditional means and variances, and the total probability theorem (Benjamin and Cornell 1970), to compute the mean and variance of the resulting spectrum. The conditional mean of $\ln Sa(T_2)$ given $\ln Sa(T_1)=x$ is

$$\mu_{\ln Sa(T_2)|\ln Sa(T_1)=x} = \sum_{i=1}^{n} p_i \cdot \left(\mu_{\ln Sa}(M_i, R_i, T_2) + \sigma_{\ln Sa}(M_i, T_2) \rho_{\ln Sa(T_1), \ln Sa(T_2)} \cdot \varepsilon_i(T_1) \right) \tag{E.2}$$

where $\mu_{\ln Sa}(M_i, R_i, T_2)$ and $\sigma_{\ln Sa}(M_i, T_2)$ are the mean and standard deviation from a ground motion prediction model, following the notation of Chapter 6. The term $\rho_{\ln Sa(T_1), \ln Sa(T_2)}$ is the correlation of spectral values from Chapter 8 (Equation 8.9). Note that $\rho_{\ln Sa(T_1),\ln Sa(T_2)}$ is specified for a given magnitude and distance, but it is found empirically to be insensitive to (i.e., independent of) the actual magnitude or distance value (see Section B.4). Similarly, the conditional variance of $\ln Sa(T_2)$ given $\ln Sa(T_1)=x$ is

$$\begin{aligned} Var\left[\ln Sa(T_2)\,|\,\ln Sa(T_1)=x\right] &= \left(1-\rho^2_{\ln Sa(T_1),\ln Sa(T_2)}\right)\sum_{i=1}^{n} p_i \cdot \sigma^2_{\ln Sa}(M_i,T_2) \\ &\quad + \sum_{i=1}^{n} p_i \left(\left(\begin{array}{l} \mu_{\ln Sa}(M_i,R_i,T_2) \\ \quad + \sigma_{\ln Sa}(M_i,T_2)\rho_{\ln Sa(T_1),\ln Sa(T_2)} \cdot \varepsilon_i(T_1) \end{array}\right) - \mu_{\ln Sa(T_2)|\ln Sa(T_1)=x}\right)^2 \end{aligned} \tag{E.3}$$

In order to achieve practical approximations, we make the first-order approximation that that the marginal mean and standard deviation $\mu_{\ln Sa}(M,R,T)$ and $\sigma_{\ln Sa}(M,T)$ are linear functions of M and R. Then, using Taylor expansions about $\bar{M}$ and $\bar{R}$, we obtain the First-Order, Second Moment (Melchers 1999) approximation of the conditional mean and variance of the target spectrum, given $\ln Sa(T_1)=x$. The conditional mean is

$$\begin{aligned} &\sum_{i=1}^{n} p_i \cdot \left(\mu_{\ln Sa}(M_i,R_i,T_2) + \sigma_{\ln Sa}(M_i,T_2)\rho_{\ln Sa(T_1),\ln Sa(T_2)} \cdot \varepsilon_i(T_1)\right) \\ &\qquad = \mu_{\ln Sa}(\bar{M},\bar{R},T_2) + \sigma_{\ln Sa}(\bar{M},T_2)\rho_{\ln Sa(T_1),\ln Sa(T_2)} \cdot \bar{\varepsilon}(T_1) \end{aligned} \tag{E.4}$$

and Equation E.2 simplifies to Equation 6.1. We make the additional approximation that the variance in conditional $\ln Sa(T_2)$ values is primarily due to ε rather than variations in causal magnitudes and distances. That is,

$$\left(1-\rho^2_{\ln Sa(T_1),\ln Sa(T_2)}\right)\sum_{i=1}^{n} p_i \cdot \sigma^2_{\ln Sa}(M_i,T_2) \gg \sum_{i=1}^{n} p_i \left(\left(\begin{array}{l} \mu_{\ln Sa}(M_i,R_i,T_2) \\ \quad + \sigma_{\ln Sa}(M_i,T_2)\rho_{\ln Sa(T_1),\ln Sa(T_2)} \cdot \varepsilon_i(T_1) \end{array}\right) - \mu_{\ln Sa(T_2)|\ln Sa(T_1)=x}\right)^2 \tag{E.5}$$

Under this approximation, then Equation E.3 simplifies to Equation 6.2. These approximations are convenient because it is much simpler to consider the spectrum at only the mean (M, R, ε) values.

To test the validity of this approximation, consider the following examples. For the site of interest here, the conditional spectral shape developed from $\bar{M}$, $\bar{R}$, $\bar{\varepsilon}$ was very close to the conditional spectral shape developed using the rigorous procedure (the conditional means and variances nearly always differed by less than a few percent between the two methods). Note that when the ground motion hazard is dominated by a single magnitude and distance, Equation E.4 is exact (the site in Figure E.1 looks like this, as will many coastal California sites and certain Central and Eastern US sites near Charleston or New Madrid).

A more difficult test can be performed by considering a hypothetical site where multiple earthquake events contribute to the hazard at a given $Sa(T_1)$ level, and these events have very different mean spectral shapes. Such a site is considered here. The hypothetical site is near two faults, each of which produces only a single characteristic earthquake, as illustrated schematically in Figure E.2.

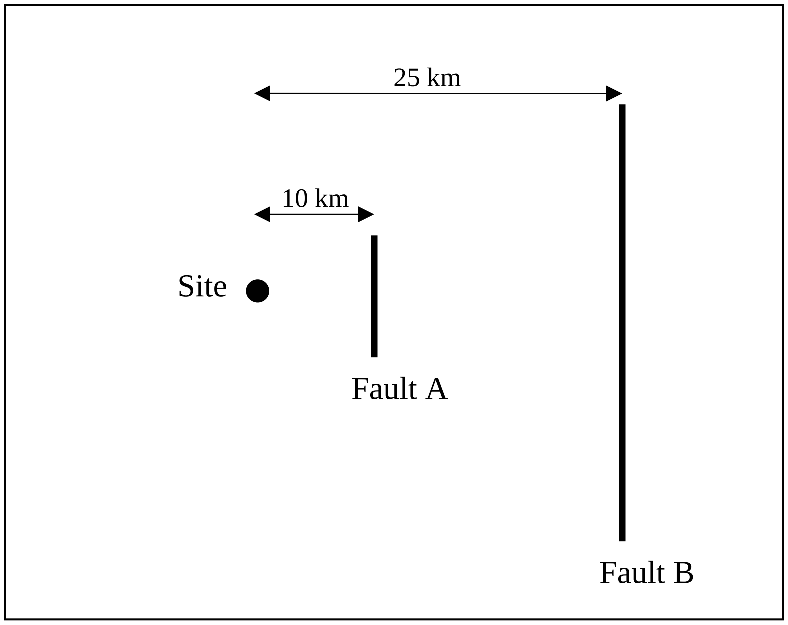


Figure E.2: Schematic illustration of the hypothetical site considered.

The two faults produce earthquake events with the following properties: Fault A produces only Magnitude 6 earthquakes at a distance of 10 km from the site. This is referred to as "Event A." Event A has a mean annual frequency of occurrence of 0.01 (i.e., a mean return period of 100 years). Fault B produces only magnitude 8 events at a distance of 25 km from the site. This event is referred to as Event B, and has a mean annual frequency of occurrence of 0.002 (i.e., a mean return period of 500 years). These two events have very different mean spectra, as predicted by the attenuation model of Abrahamson and Silva (1997) and displayed in Figure E.3.

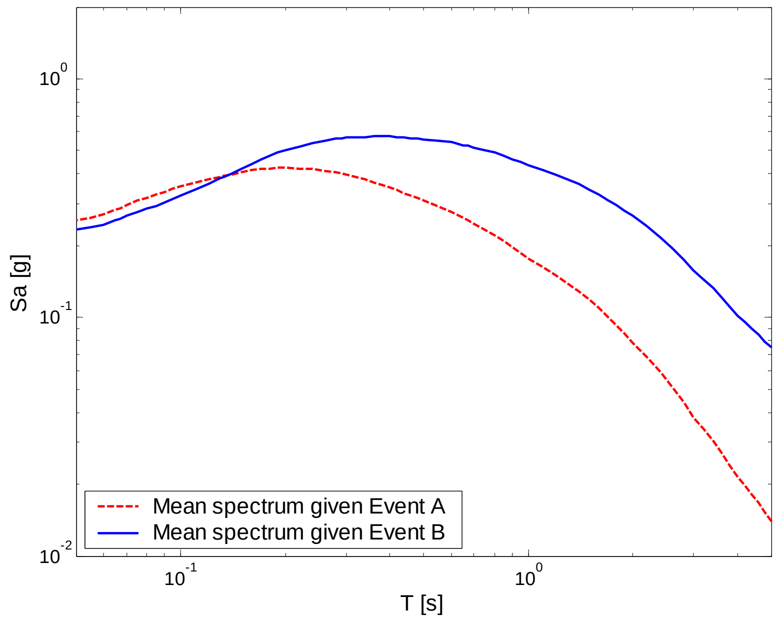


Figure E.3: Predicted mean spectra from the two possible events.

The ground motion intensity resulting from these faults is a random variable, due to variation in the earthquakes' stress drop, slip distribution, etc., even given multiple earthquakes with the same magnitude and distance. The uncertainty in spectral acceleration values is illustrated in Figure E.4, which displays the mean spectra and the mean spectra +/- one standard deviation. Given that the ground motion spectral acceleration is a random variable, the possibility exists that the spectral acceleration resulting from an earthquake is much larger than the mean ground motion associated with the earthquake's magnitude and distance (this would be a "positive epsilon" ground motion). Consider the exceedance of an *Sa*(1s) value of 0.9g. (This level of intensity is chosen because it has a $4*10^{-4}$ annual rate of exceedance at the given site, as will be calculated below. If a Poisson model for earthquake recurrence is assumed, this corresponds to the *Sa*(1s) ground motion level exceeded with 2% probability in 50 years—a common design level in structural engineering.) This *Sa*(0.1s)=0.9g ground motion is also displayed in Figure E.4. It is clear from this figure that the target *Sa* level is slightly more than one standard deviation larger than the mean *Sa* from Event B. It is approximately two standard deviations larger than the mean *Sa* from Event A. However, Event A has a much higher rate of occurrence, and so it will still make a significant contribution to the ground motion hazard at this *Sa* level.

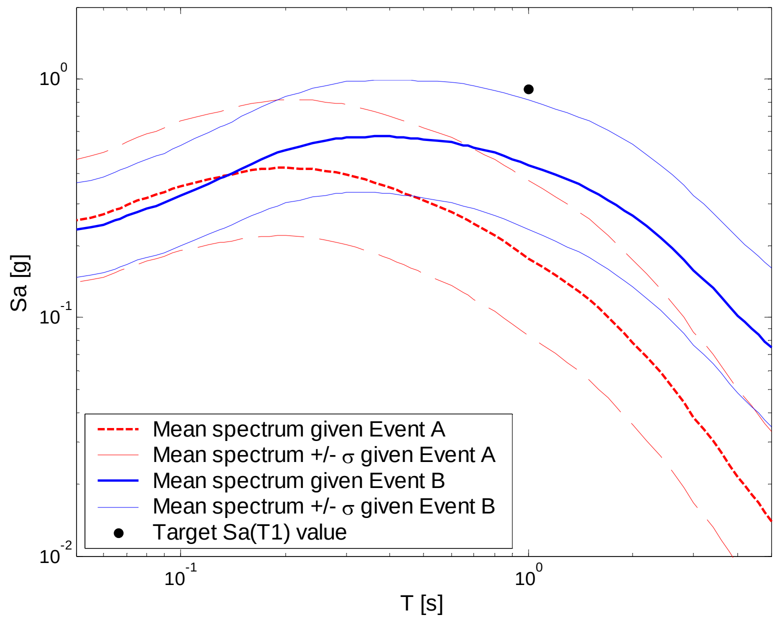


Figure E.4: Predicted mean spectra from the two possible events along with +/- one standard deviation. The target *Sa*(1s) value is also noted.

In fact, the contribution of each event to the hazard at this *Sa* level can be computed using disaggregation (McGuire 1995, Bazzurro and Cornell 1999). Given the rates of occurrence of the two events given above, and the mean and standard deviation of log *Sa*(1s) given each of the events (log *Sa* values, given magnitude and distance, have been seen to be well-represented by Gaussian distributions), disaggregation can be performed. Some of the relevant data is summarized below.

Event A:

Mean of ln*Sa*(1s) given Event A = -1.7356

Standard deviation of ln*Sa* (1s) given Event A = 0.7494

Probability of exceeding *Sa*(1s)=0.9g, given Event A = 0.0148

Mean annual rate of *Sa*(1s)>0.9g due to earthquakes from Fault A = 0.01*0.0148=1.48*10^{-4}

Event B:

Mean of ln*Sa* (1s) given Event B = -0.8329

Standard deviation of ln*Sa* (1s) given Event B = 0.6243

Probability of exceeding *Sa*(1s)=0.9g, given Event B = 0.1219

Mean annual rate of *Sa*(1s)>0.9g due to earthquakes from Fault B = 0.002*0.1219= 2.44*10^{-4}

By summing the mean annual frequencies of *Sa*(1s)>0.9g due to the two events, we get a total mean annual frequency of ~$4*10^{-4}$, as noted above[23]. We can now use disaggregation to determine the contribution of each event to exceedance of 0.9g, along with the associated ε value. Using PSHA disaggregation, we find that given occurrence of a ground motion with *Sa*(1s)>0.9g, the probability is 0.38 that the ground motion came from Event A, and the probability is 0.62 that the ground motion came from Event B (these are the "p_i" probabilities needed in the equations throughout this Appendix). Note that these probabilities are just the relative contributions of the two events to the total mean annual frequency of exceeding 0.9g. Further, the disaggregation values associated with these two events can also be determined. Given Event A, the causal magnitude, distance and ε values are 6, 10 and 2.18, respectively[24]. Given Event B, the causal magnitude, distance and ε values are 8, 25 and 1.17, respectively. These are the M_i, R_i and ε_i values needed in the equations throughout this Appendix. Further, we can weight these two *M*, *R* ,ε triples by their associated probabilities of causing *Sa*(1s)>0.9g to determine the mean *M*, *R* and ε. In this example, they are $\bar{M}$ =7.25, $\bar{R}$ =19.3 km, and $\bar{\varepsilon}$ =1.55. These are the values used for the approximation of Equations 6.1 and 6.2. Note that these values are not associated with either of the two earthquakes that can physically occur at the site; however, they will produce a reasonable approximation of the Conditional Mean Spectrum, considering ε, as will now be demonstrated.

In Figure E.5, the Conditional Mean Spectra, considering ε, are plotted for several situations. The conditional mean spectra are plotted, given *Sa*(1s)=0.9g and occurrence of either Event A or B. These two spectra are then combined using Equation E.2 to compute the CMS-ε, which is also shown in Figure E.5.

---

[23] As can be observed in Figure E.4, the ground motion level associated with the 2% in 50 years hazard is much larger than the mean ground motion from any possible earthquake. Although not always emphasized, this is often the case in real sites as well, particularly in seismically active areas such as the Western United States.

[24] Here the McGuire (1995) disaggregation method is used to obtain ε, rather the alternative method considered by the Bazzurro and Cornell (1999). As mentioned in Chapter 6, the McGuire method is more appropriate for this record-selection problem.

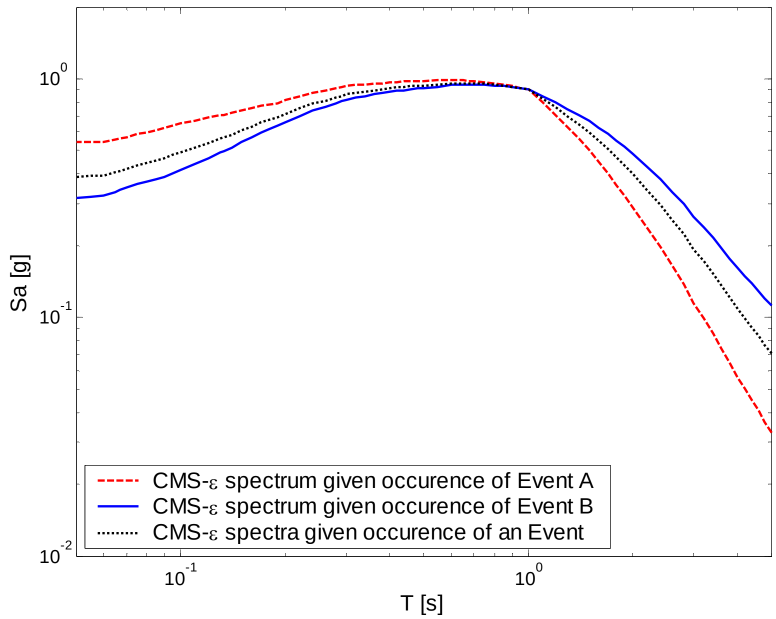


Figure E.5: Conditional mean spectra given occurrence of the individual events, and then combined CMS-ε spectrum that is proposed as a target spectrum for record selection.

The CMS-ε spectrum shown in Figure E.5 is the exact spectrum for this site to use in the record selection procedure proposed in Chapter 6. As an alternative, the $\bar{M}$, $\bar{R}$, and $\bar{\varepsilon}$ values from this site can be used to approximate the CMS-ε spectrum, using Equation 6.1. The two CMS-ε spectra are shown in Figure E.6, and the agreement is seen to be reasonable.

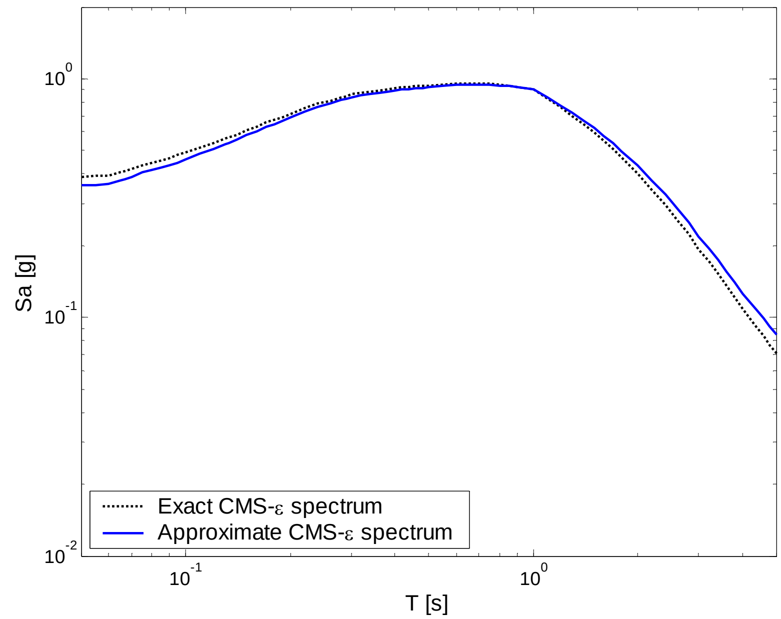


Figure E.6: Exact (using Equation E.2) and approximate (using Equation 6.1) CMS-ε spectra for the given hypothetical site.

Further, one could compare the exact and approximate conditional standard deviations of the response spectrum given Sa(1s)=0.9g. These are displayed graphically in Figure E.7.

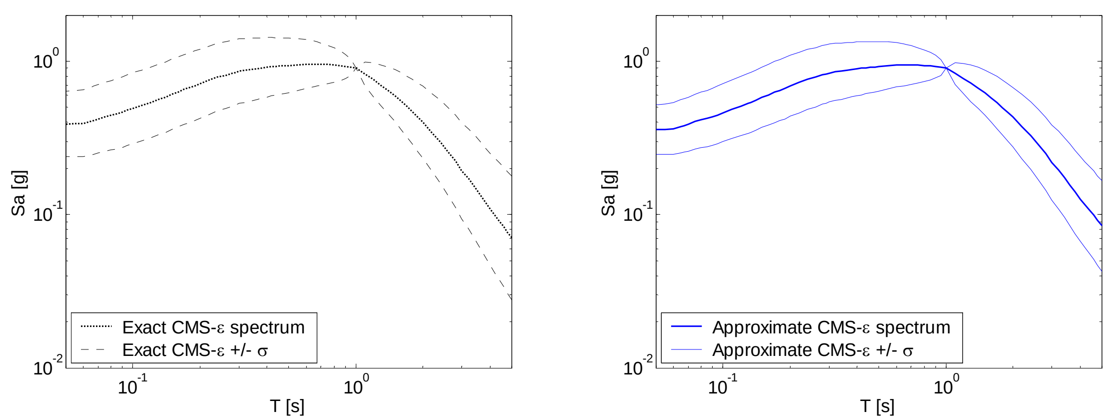


Figure E.7: (a) Exact conditional mean and standard deviation of the response spectrum, given Sa(1s)=0.9g. (using Equations E.2 and E.3). (b) Approximate conditional mean and standard deviation of the response spectrum, given Sa(1s)=0.9g. (using Equations 6.1 and 6.2).

Visually, the results from Figure E.7 look comparable. Within the range T/3 to T*3 (the range of the response spectrum that might be expected to affect a structure with fundamental period T, which is equal to 1s in this example), the approximate mean differs from the exact mean by no more than 12%. The conditional standard deviations differ by no more than 20% over this range.

These errors could be considered an upper bound on the errors that might occur at a real site. This site was specifically constructed to have significant contributions from two events with dramatically different magnitudes. In some situations, this might approximately represent the hazard at a site (e.g., *Sa*(1s) at the 2% in 50 year hazard level for Atlanta, shown in Figure E.8, has significant contributions from two different sources), and by idealizing this as two sources, one could repeat the calculation performed above to compute a more exact CMS-ε spectrum. Otherwise, at sites where the hazard is dominated by events with very similar magnitudes and distances (such as the one displayed in Figure E.1), there will be much less difference between the exact and approximate CMS-ε spectra. For this reason, the approximations used in Chapter 6 are believed to be reasonable in many situations situations.

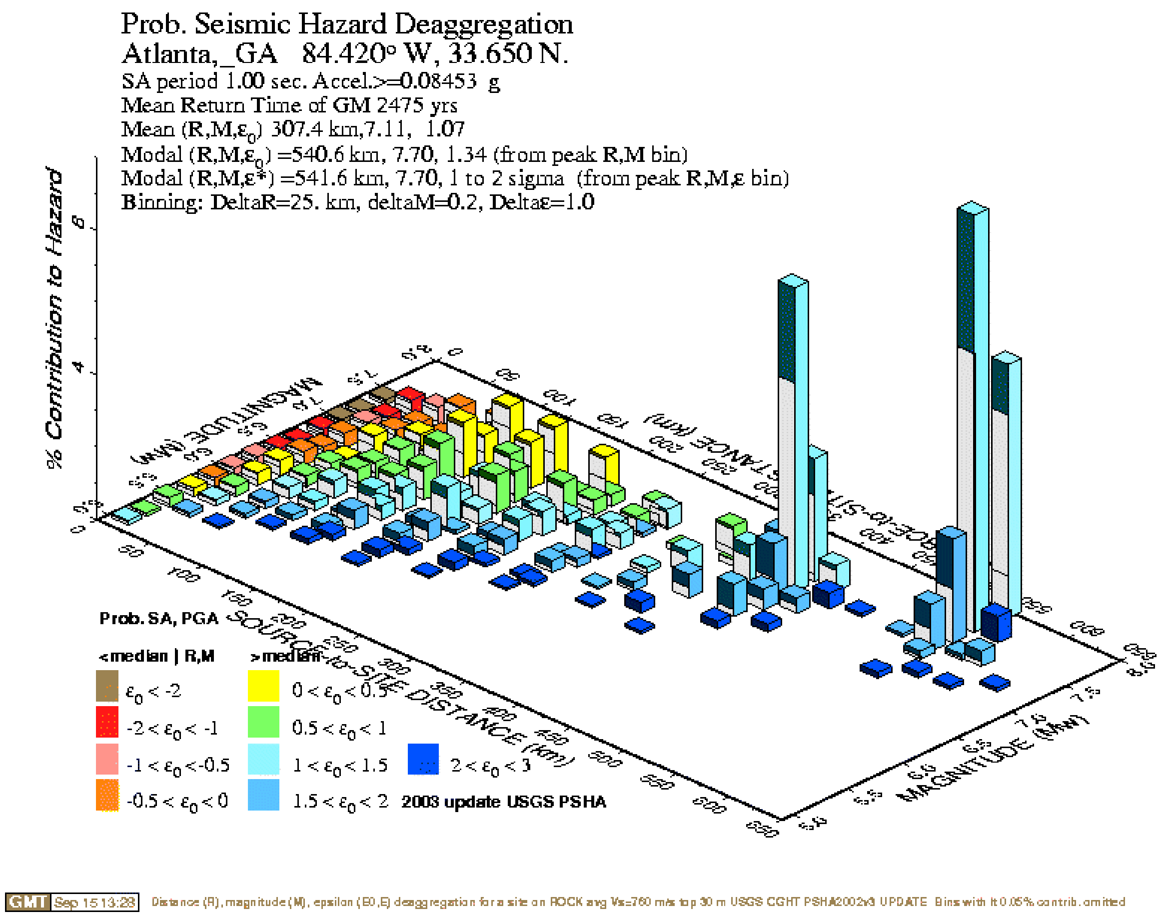


Figure E.8: Disaggregation for *Sa*(1.0s) at the 2% in 50 year hazard level for Atlanta, Georgia (USGS 2002).

The success of this approximation depends on the observation that the conditional variance in ε values at other periods has a much greater effect on variations in spectral shape than differences in mean spectra due to variations in magnitude and distance (because the variance of ε is large, *and* because *Sa* is sensitive to changing values of ε) . Given the close match of the two procedures, the much simpler procedure of using $\bar{M}$, $\bar{R}$, $\bar{\varepsilon}$ is recommended for most practice (the exact answer in the above example was easy to compute because there were only two possible events, but this is not the case at real sites). However, if the analyst has any concern about the accuracy of the approximate result for a given site, there is no obstacle other than programming time and computational effort preventing one from obtaining the exact conditional spectra (within the limits of the discrete disaggregation, which can often be fairly coarse). It should also be mentioned that $\bar{M}$, $\bar{R}$ values from disaggregation sometimes correspond to values that do not occur from any fault in the area surrounding the site (e.g., Bazzurro and Cornell, 1999, note that $\bar{R}$ is sometimes intermediate between the distances to two major faults, as was the case in the example above), causing some to express concern about using $\bar{M}$, $\bar{R}$ as target values for record selection. This concern should be lessened here, because we are only using $\bar{M}$ and $\bar{R}$ to identify a target response spectrum distribution and not to identify target *M* and *R* values. The response spectra of records coming from magnitudes and distances that actually occur at the site and have the specified $Sa(T)^*$ level will agree well with the conditional response spectra based on $\bar{M}$ and $\bar{R}$ given $Sa(T)^*$.

One final concern that should be noted with this simplification is that when $\bar{M}$, $\bar{R}$ and $\bar{\varepsilon}$ obtained from disaggregation of $Sa(T_1)$ are used to predict $Sa(T_1)$, one does not necessarily obtain the target value of $Sa(T_1)$ back again. Although each (M_i, R_i, ε_i) triplet will produce an $Sa(T_1)$ value equal to the target, when one obtains the mean values of *M*, *R* and ε marginally, they do not necessarily produce this target $Sa(T_1)$ value. This results in the mean conditional spectra predicted by Equation 6.1 not actually passing through the $Sa(T_1)$ value that was conditioned on. There are two simple fixes for this. One is to re-assign $\bar{\varepsilon}$ to the ε value that results in a prediction of the $Sa(T_1)$ target value; the modification will be small and this is consistent with the treatment of ε by McGuire (1995). The second option is to slightly re-scale the target spectrum so that it passes through the target $Sa(T_1)$ value. The second option has been adopted in this work, although the modification is so small that the first option will produce essentially identical results.

Appendix F

Response results for two additional structures, using four methods for selecting ground motions

To support the conclusions of Chapter 6, the four record-selection methods proposed in that chapter were used to analyze two additional structures. Both structures were considered to be located at the same Van Nuys, California, site used in Chapter 6. Several summary figures and tables are shown in this appendix, and the equivalent figures and tables from Chapter 6 are noted.

F.1 First structure

The first structure considered is a generic frame structure designed by Ibarra (2003), which has also been used in Chapters 3, 4 and 5. It has the following properties: 3 stories, T_1 = 0.3s, ductility capacity (δ_c/δ_y) = 4, post-capping stiffness coefficient (α_c)= -50%, V/W = 0.75 (i.e., Sa_{yield}(0.3s) = 0.75g).

This first-mode dominated structure has a fundamental period of 0.3 seconds, making it sensitive to shorter periods than the structure tested in Chapter 6. The hazard disaggregation values for Sa(0.3s) are given in Table F.1 through Table F.3, which correspond to Table 6.1 through Table 6.3 in the body of the dissertation. At this shorter period, the hazard tends to be controlled by smaller magnitude events at closer distances than for the 0.8s period considered in Chapter 6.

Table F.1. Mean magnitude values from disaggregation of the Van Nuys site, and the mean magnitude values of the records selected using each of the four proposed methods. The mean magnitude value of the record library was 6.7.

	Magnitude				
Sa(0.3s) [g]	Target from disagg	1. AR Method	2. MR-BR Method	3. ε-BR Method	4. CMS-ε Method
0.1	6.0	6.7	6.3	6.7	6.6
0.4	6.0	6.7	6.1	6.5	6.5
0.8	6.0	6.7	6.1	6.6	6.5
1.2	6.0	6.7	6.1	6.7	6.6
1.6	6.1	6.7	6.2	6.7	6.5
2.2	6.1	6.7	6.2	6.7	6.5
3.0	6.2	6.7	6.3	6.7	6.6
4.0	6.3	6.7	6.4	6.7	6.6
5.0	6.4	6.7	6.5	6.7	6.7

Table F.2. Mean distance values from disaggregation of the Van Nuys site, and the mean distance values of the records selected using each of the four proposed methods. The mean distance value of the record library was 33 km.

	Distance				
Sa(0.3s) [g]	Target from disagg	1. AR Method	2. MR-BR Method	3. ε-BR Method	4. CMS-ε Method
0.1	44.8	35.5	37.0	30.3	27.0
0.4	20.5	35.5	20.0	32.0	26.2
0.8	14.7	35.5	14.2	34.9	27.4
1.2	12.7	35.5	12.7	36.3	27.0
1.6	11.7	35.5	12.5	38.0	27.3
2.2	10.7	35.5	11.7	37.1	28.4
3.0	9.6	35.5	10.6	37.1	30.4
4.0	8.8	35.5	9.0	37.1	31.6
5.0	8.8	35.5	7.3	37.1	33.5

Table F.3. Mean ε values (at 0.3s) from disaggregation of the Van Nuys site, and the mean εvalues of the records selected using each of the four proposed methods. The mean ε value of the record library was 0.2.

	Epsilon (ε)				
Sa(0.3s) [g]	Target from disagg	1. AR Method	2. MR-BR Method	3. ε-BR Method	4. CMS-ε Method
0.1	-0.2	0.2	0.2	-0.2	0.2
0.4	0.6	0.2	0.4	0.6	0.3
0.8	1.2	0.2	0.2	1.1	0.4
1.2	1.5	0.2	0.5	1.4	0.6
1.6	1.8	0.2	0.5	1.7	0.6
2.2	2.1	0.2	0.5	1.7	0.5
3.0	2.4	0.2	0.2	1.8	0.7
4.0	2.6	0.2	0.0	1.8	0.7
5.0	2.8	0.2	-0.2	1.8	0.7

The conditional mean spectra at the *Sa*(0.3s) level corresponding to a 2% in 50 years hazard are shown in Figure F.1 (corresponding to Figure 6.7). The spectra from the Arbitrary Records and M, R-Based Records again tend to be larger at most periods. Note, however, that the mean spectrum from the MR-BR records has a slight “valley” at 0.4 seconds, which might cause these records to produce lower levels of nonlinear response than might otherwise be expected (a result which will be observed below). The valley observed here is believed to be a chance observation due to the finite sample size—there is no reason to expect a spectral shape like this to occur consistently. Note also that the trends in spectra become less clear at periods greater than 1 second; however, this period range will have little or no effect on the short-period structure considered here. In fact, the variation of the spectra at these long periods occurs precisely because we have conditioned on spectral acceleration at a much shorter period.

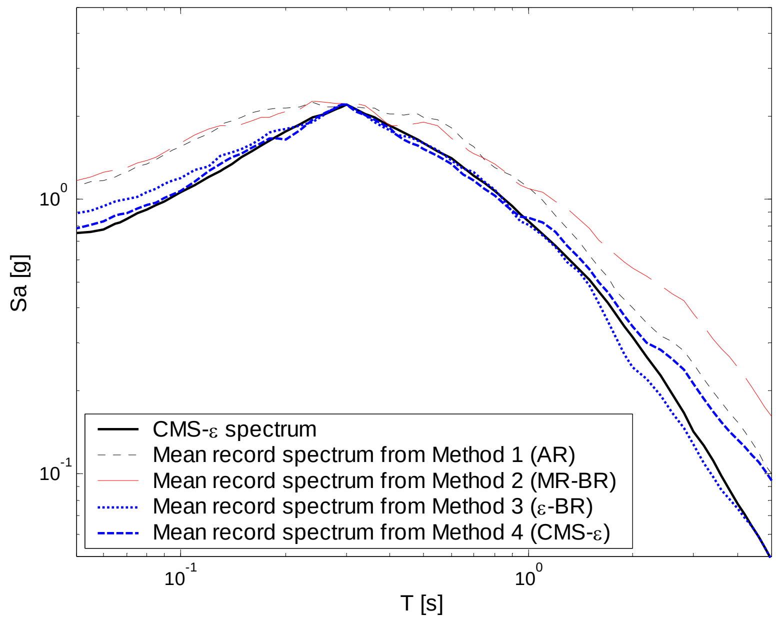


Figure F.1: Conditional Mean Spectra at Sa(0.3s)=2.2g (given $\bar{M}$ =6.1, $\bar{R}$ =10.7 km and $\bar{\varepsilon}$ =2.1) and the mean response spectra of record sets selected using each of the four proposed record selection methods.

The geometric mean of *EDP* as a function of *IM* is shown in Figure F.2 (corresponding to Figure 6.9), using the results from the four record-selection methods. The CMS-ε records produce smaller responses than the other methods. This may be because the record-selection method preferentially selects records with spectra that are somewhat smooth, in order to match the target spectrum; Carballo (2000) observed that records which were artificially processed to have very smooth spectra tend to produce lower levels of structural response, and the same effect may be occuring here to a lesser extent with the partially-smooth records. The MR-BR records produce responses comparable to the ε-BR records, possibly due to the valley in the MR-BR record spectra noticed earlier. Given the results from similar tests on the other structures, it is expected that the MR-BR records will typically produce larger responses (comparable to the AR-records' responses).

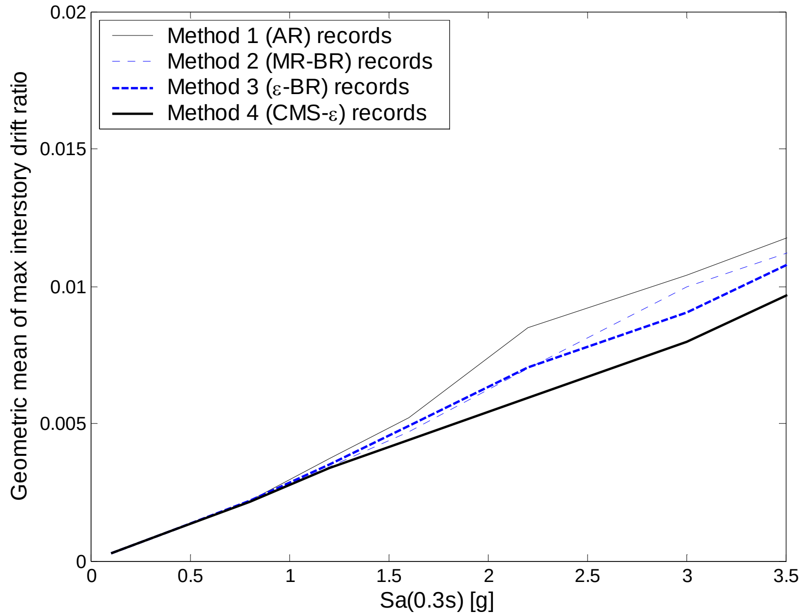


Figure F.2: The geometric mean of maximum interstory drift ratio vs. $Sa(T_1)$ for the four record-selection methods considered.

The standard deviations of log maximum interstory drift ratio from the three record sets are shown in Figure F.3 (comparable to Figure 6.10). These results are quite similar to Figure 6.10. The standard deviations are only reported for *Sa* levels where fewer than 50% of records caused collapse: when a large fraction of records cause collapse, then the standard deviation of non-collapse responses becomes a less meaningful statistic.

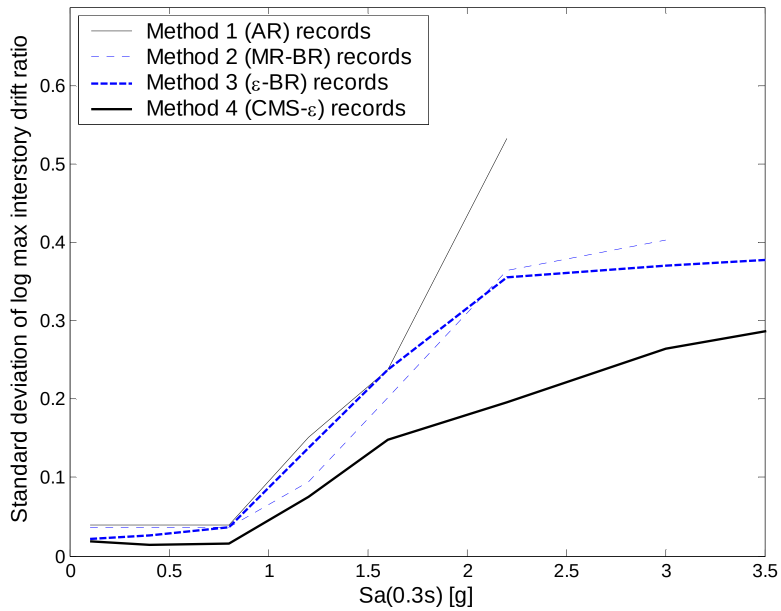


Figure F.3: The standard deviation of log maximum interstory drift ratio vs. $Sa(T_1)$ for the four record-selection methods considered, at $Sa(T_1)$ levels where less than 50% of records cause collapse.

The probability of collapse as a function of *Sa* is shown in Figure F.4 (which is comparable to Figure 6.11). The MR-BR records have relatively lower probabilities of causing collapse than in the other two test cases (Figure 6.11 and Figure F.10), possibly because of the valley in the spectra observed in Figure F.1. If more records were available to generate another sample of MR-BR records, it would be expected that they would have similar probabilities of causing collapse as the AR records.

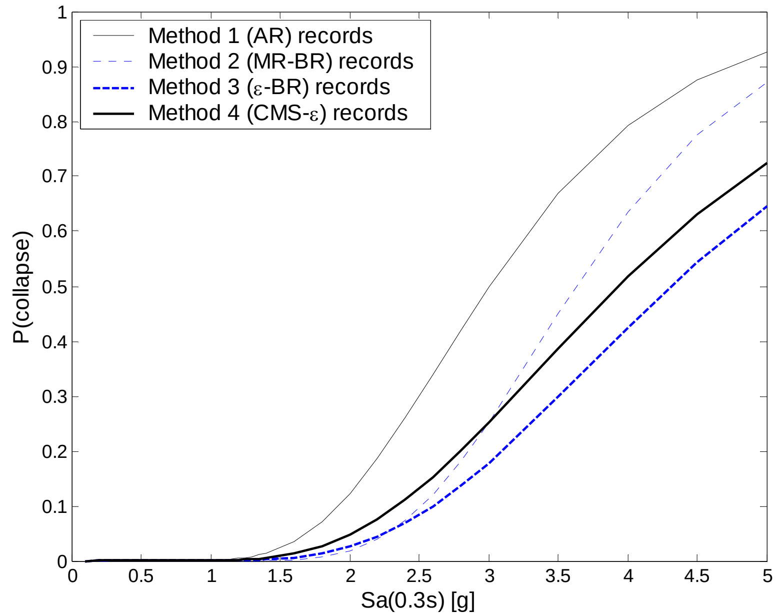


Figure F.4: Estimated probability of collapse vs. $Sa(T_1)$ (i.e., the collapse "fragility curve") using the four record-selection methods considered.

The Sa(0.3s)-based drift hazard curves are shown in Figure F.5. The trends here are comparable to those observed in Figure 6.13, except that the MR-BR records produces lower mean annual rates of exceedance.

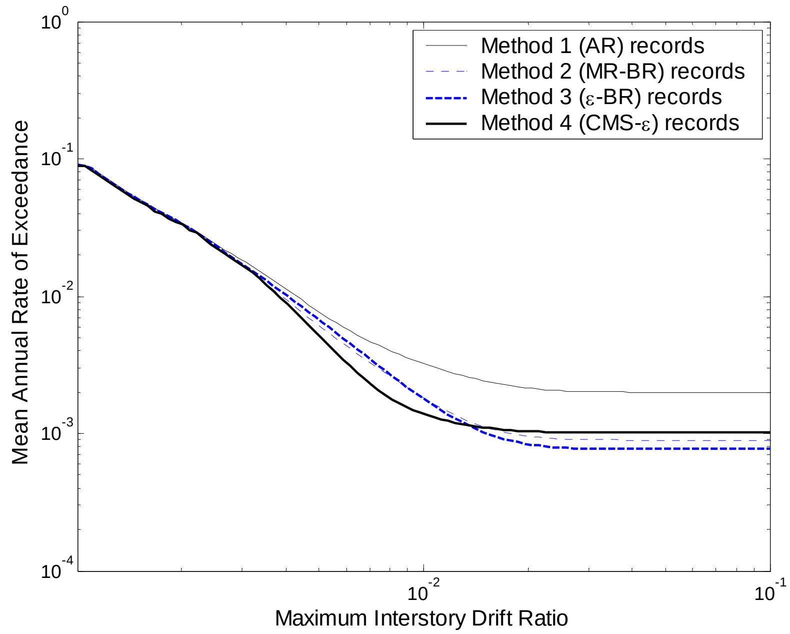


Figure F.5: Mean annual frequency of exceeding various levels of maximum interstory drift ratio, as computed using the *scalar* intensity measure $Sa(T_1)$.

When a vector IM incorporating ε is incorporated with the AR records, the resulting drift hazard curve agrees more closely with the (scalar-based) drift hazard curves estimated using ε-BR records or CMS-ε records. However, when the MR-BR records are used with the vector IM, the result becomes *too* low. Again, it should be noted that with another sample of M,R-Based records, this behavior from the MR-BR records would likely not occur.

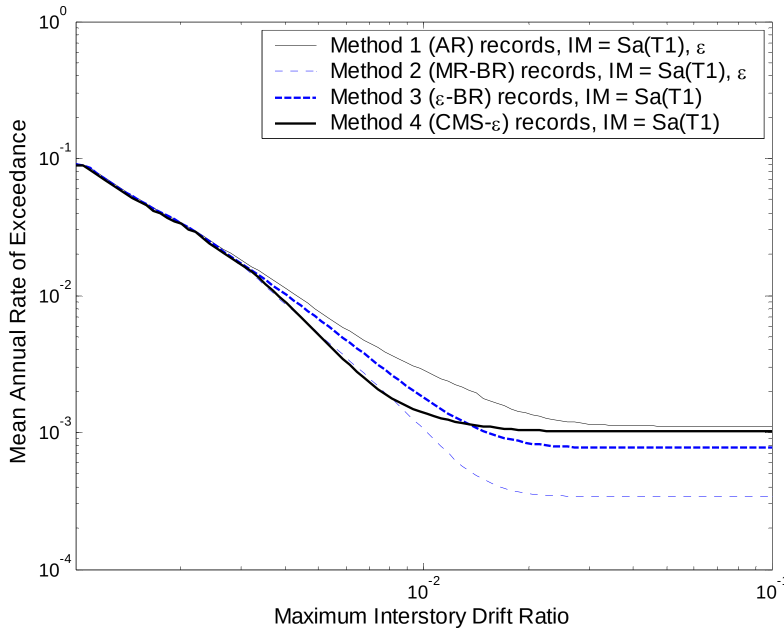


Figure F.6: Mean annual frequency of exceeding various levels of maximum interstory drift ratio, as computing using either the scalar intensity measure $Sa(T_1)$ or the vector intensity measure $Sa(T_1)$ and ε.

None of the four record selection methods showed a statistically significant bias with record scaling. In Table F.5 (analogous to Table 6.4), p-values for maximum interstory drift versus scale factor are given, and no characteristics of statistical significance are evident. This is in contrast to results from Chapter 6 and the structure analyzed below, where Methods 1 and 2 display a statistically significant bias with scale factor. Even if this scaling bias is present with Methods 1 and 2 in only two of the three structures considered, it is preferable to select records using Methods 3 or 4, which displayed no scaling bias in any of the three cases considered.

Table F.4. P-values from regression prediction of max interstory drift ratio as a function of scale factor for four methods of record selection, at seven levels of *Sa*(0.3s). P-values of less than 0.05 are marked in boldface.

Sa(0.3s)	Method 1: Arbitrary Records	Method 2: M, R-Based Records	Method 3: ε-Based Records	Method 4: CMS-ε Method
0.1	0.69	0.23	0.35	0.71
0.4	0.69	0.28	0.36	0.87
0.8	0.68	0.32	0.98	0.80
1.2	0.07	0.47	0.11	0.07
1.6	0.78	0.21	**0.01**	0.58
2.2	0.66	0.83	0.05	0.70
3	0.43	0.24	0.90	1.00
median p-value	0.68	0.28	0.35	0.71

F.2 Second structure

The second structure considered is also a generic frame structure designed by Ibarra (2003). It has the following properties: 6 stories, T_1 = 1.2s, ductility capacity (δ_c/δ_y) = 4, post-capping stiffness coefficient (α_c)= -50%, V/W = 0.2 (i.e., Sa_{yield}(1.2s) = 0.2g).

This structure has a fundamental period of 1.2 seconds. The hazard disaggregation values for *Sa*(1.2s) are given in Table F.5 through Table F.7, which correspond to Table 6.1 through Table 6.3 in the body of the dissertation. The general trends in these tables are very similar to the trends observed in Table 6.1 through Table 6.3.

Table F.5. Mean magnitude values from disaggregation of the Van Nuys site, and the mean magnitude values of the records selected using each of the four proposed methods. The mean magnitude value of the record library was 6.7.

	Magnitude				
Sa(1.2s) [g]	Target from disagg	1. AR Method	2. MR-BR Method	3. ε-BR Method	4. CMS-ε Method
0.1	6.5	6.7	6.5	6.7	6.8
0.2	6.6	6.7	6.6	6.8	6.8
0.4	6.7	6.7	6.6	6.5	6.8
0.6	6.7	6.7	6.6	6.5	6.8
0.8	6.6	6.7	6.7	6.5	6.8
1.2	6.6	6.7	6.6	6.6	6.7
1.6	6.6	6.7	6.5	6.6	6.7
2.0	6.6	6.7	6.6	6.6	6.7
2.4	6.7	6.7	6.6	6.6	6.7

Table F.6. Mean distance values from disaggregation of the Van Nuys site, and the mean distance values of the records selected using each of the four proposed methods. The mean distance value of the record library was 33 km.

	Distance				
Sa(1.2s) [g]	Target from disagg	1. AR Method	2. MR-BR Method	3. ε-BR Method	4. CMS-ε Method
0.1	62.1	35.5	59.0	31.6	28.3
0.2	45.2	35.5	42.7	30.8	30.2
0.4	31.8	35.5	31.9	34.1	34.3
0.6	25.6	35.5	27.2	41.2	38.5
0.8	21.6	35.5	20.9	46.5	42.8
1.2	15.8	35.5	14.9	51.2	42.4
1.6	11.9	35.5	10.4	51.2	42.6
2.0	9.6	35.5	9.7	51.2	42.5
2.4	9.0	35.5	10.0	51.2	42.5

Table F.7. Mean ε values (at 1.2s) from disaggregation of the Van Nuys site, and the mean ε values of the records selected using each of the four proposed methods. The mean ε value of the record library was 0.1.

	Epsilon (ε)				
Sa(1.2s) [g]	Target from disagg	1. AR Method	2. MR-BR Method	3. ε-BR Method	4. CMS-ε Method
0.1	0.2	-0.1	0.8	0.2	0.1
0.2	0.7	-0.1	0.0	0.7	0.4
0.4	1.2	-0.1	0.0	1.2	0.7
0.6	1.6	-0.1	0.2	1.5	0.9
0.8	1.8	-0.1	0.0	1.7	1.0
1.2	2.1	-0.1	-0.2	1.9	1.2
1.6	2.4	-0.1	-0.2	1.9	1.4
2.0	2.5	-0.1	-0.1	1.9	1.4
2.4	2.6	-0.1	0.1	1.9	1.4

The conditional mean spectra at the *Sa*(1.2s) level corresponding to a 2% in 50 years hazard are shown in Figure F.7 (corresponding to Figure 6.7). The spectra from the Arbitrary Records and M,R-Based Records again tend to be larger than the ε-BR and CMS-ε spectra at all periods. Note that the target spectrum does not match the ε-BR and CMS-ε spectra near 2 seconds. Research in progress by the author suggests that the attenuation model used here (Abrahamson and Silva 1997) tends to over-predict *Sa* values in this period range, and thus leads to target spectra which are larger than the observed spectra. This does not significantly affect the results below, and no special effort was made to correct the difference, but it is interesting to note.

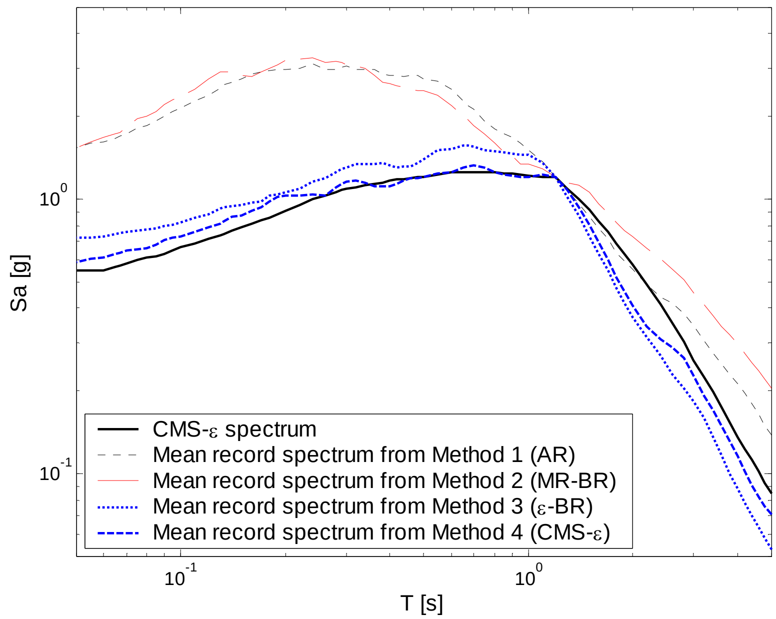


Figure F.7: Conditional Mean Spectrum at $Sa(1.2s)$=1.2g (given $\bar{M}$ =6.6, $\bar{R}$ =15.8 km and $\bar{\varepsilon}$ =2.1) and the mean response spectra of record sets selected using each of the four proposed record selection methods.

The geometric mean of *EDP* as a function of *IM* is shown in Figure F.8 (corresponding to Figure 6.9), for the *Sa* levels at which less than 50% of records caused collapse. The ε-BR and CMS-ε records produce smaller responses, as was observed in Figure 6.9.

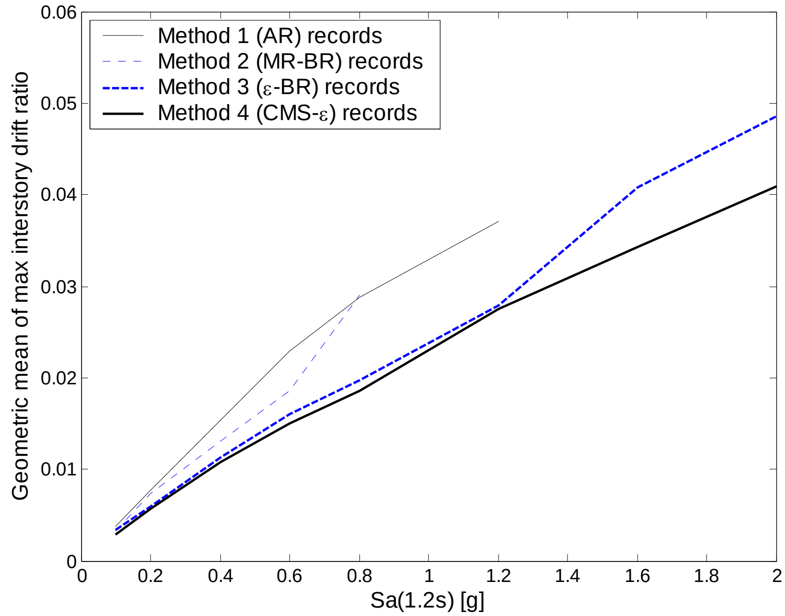


Figure F.8: The geometric mean of maximum interstory drift ratio vs. $Sa(T_1)$ for the four record-selection methods considered.

The standard deviations of log maximum interstory drift ratio from the three record sets are shown in Figure F.9 (comparable to Figure 6.10). These results are quite similar to Figure 6.10.

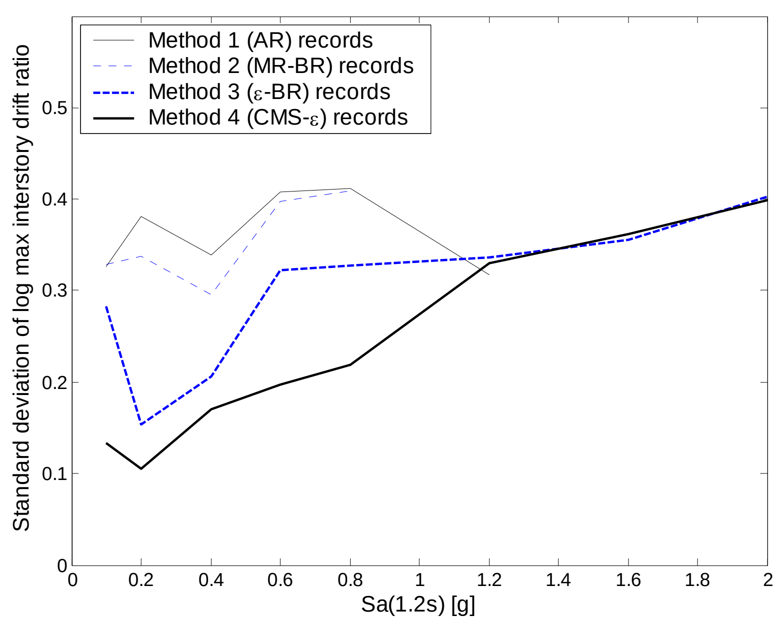


Figure F.9: The standard deviation of log maximum interstory drift ratio vs. $Sa(T_1)$ for the four record-selection methods considered, at $Sa(T_1)$ levels where less than 50% of records cause collapse.

The probability of collapse as a function of *Sa* is shown in Figure F.10 (which corresponds to Figure 6.11). These results are also quite similar to Figure 6.11.

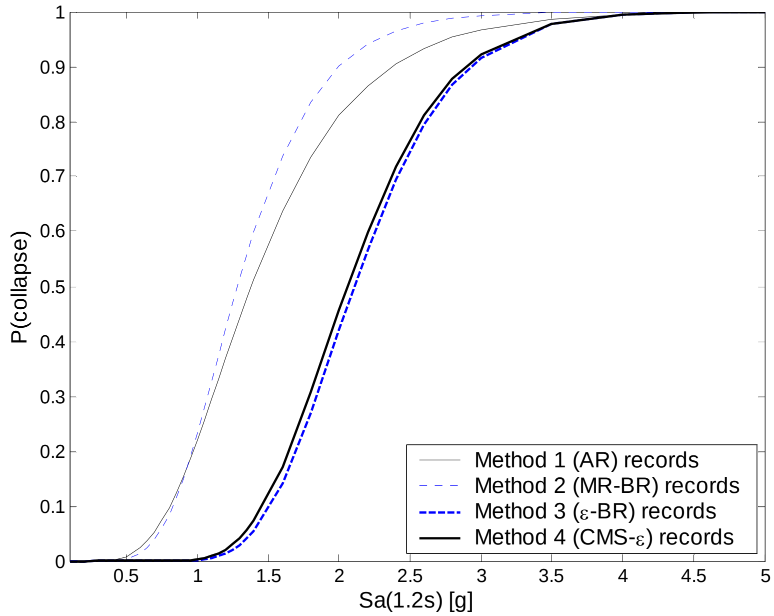


Figure F.10: Estimated probability of collapse vs. $Sa(T_1)$ (i.e., the collapse “fragility curve”) using the four record-selection methods considered.

The *Sa*(1.2s)-based drift hazard curves are shown in Figure F.11. As in Figure 6.13, the MR-BR and AR records produces higher mean annual rates of exceedance because of the higher median responses and probabilities of collapse shown above. In fact, the difference is as great as an order of magnitude in mean rates at for max interstory drift ratios near 0.1. However, when ε is incorporated in a vector IM for the MR-BR and AR records, their resulting estimated drift hazard curves are in much closer agreement with the lower two drift hazard curves, as is seen in Figure F.12.

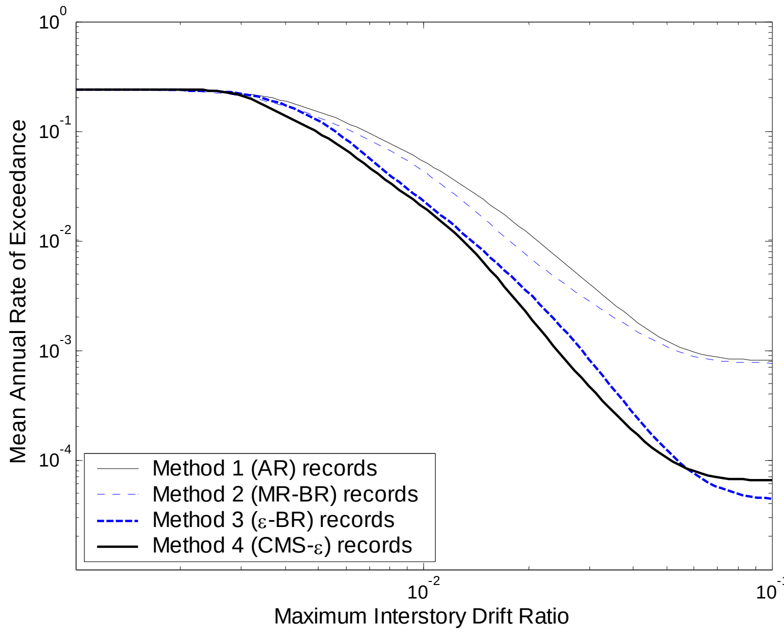


Figure F.11: Mean annual frequency of exceeding various levels of maximum interstory drift ratio, as computed using the *scalar* intensity measure $Sa(T_1)$.

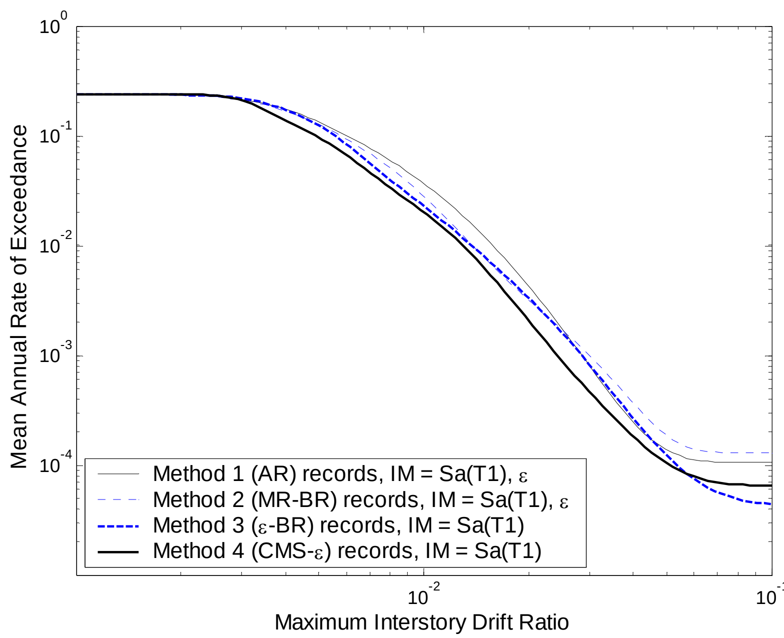


Figure F.12: Mean annual frequency of exceeding various levels of maximum interstory drift ratio, as computing using either the scalar intensity measure $Sa(T_1)$ or the vector intensity measure $Sa(T_1)$ and ε.

In Table F.8 (analogous to Table 6.4), p-values for maximum interstory drift as a function of scale factor are given. As in Table 6.4, Methods 1 and 2 show a statistically significant trend as a function of scale factor. To illustrate this trend graphically, plots of max interstory drift ratio as a function of scale factor ar shown in Figure F.13 (analogous to Figure 6.16) for the *Sa*(1.2s) level of 0.6g.

Table F.8. P-values from regression prediction of max interstory drift ratio as a function of scale factor for four methods of record selection, at seven levels of *Sa*(1.2s). P-values of less than 0.05 are marked in boldface.

Sa(1.2s)	Method 1: Arbitrary Records	Method 2: M, R-Based Records	Method 3: ε-Based Records	Method 4: CMS-ε Method
0.1	**0.00**	**0.02**	0.43	0.72
0.2	**0.00**	**0.00**	0.12	0.27
0.4	**0.01**	**0.00**	0.42	0.80
0.6	**0.02**	**0.01**	0.40	0.66
0.8	**0.10**	**0.01**	0.22	0.78
1.2	0.22	0.32	**0.01**	0.03
1.6	0.47	0.31	**0.01**	**0.21**
median p-value	0.02	0.01	0.22	0.66

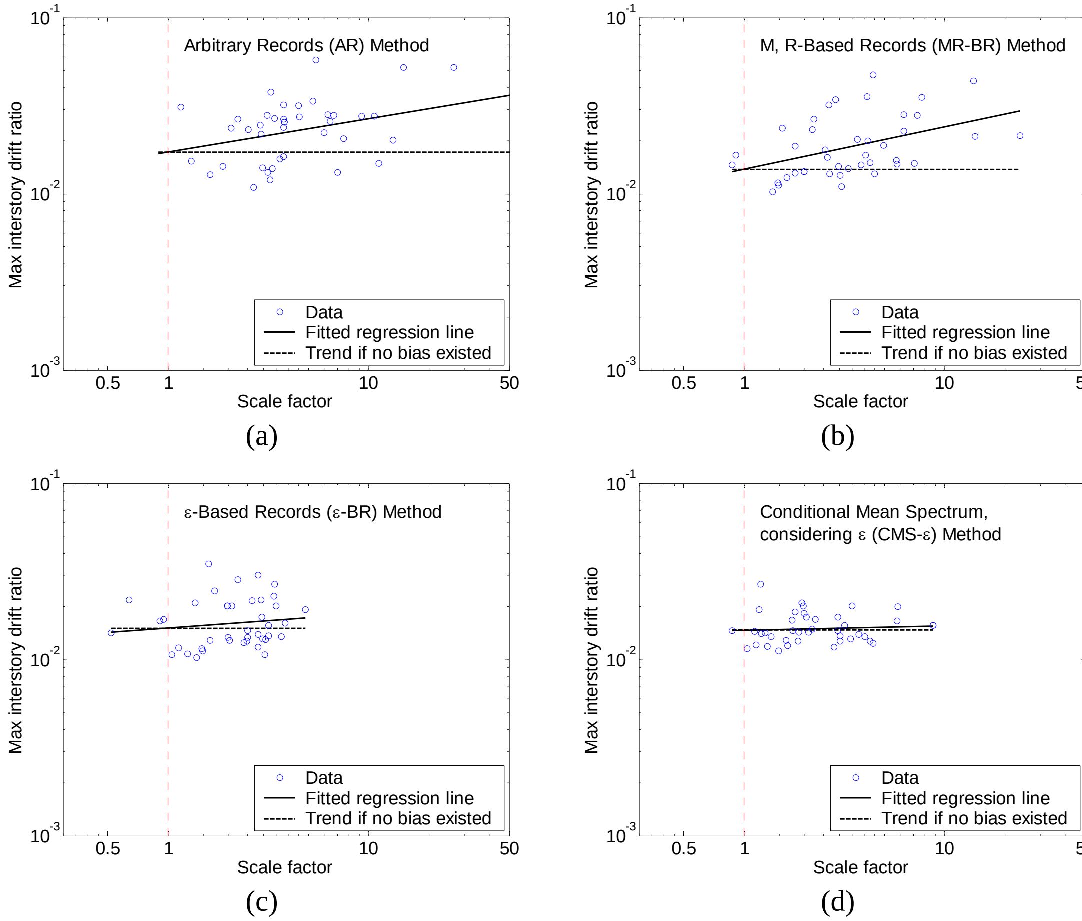


Figure F.13: Maximum interstory drift ratio versus record scale factor for each of the four selection methods considered, at an Sa(1.2s) level of 0.6g. Regression fits based on scale factor are shown with solid lines. Dashed horizontal lines corresponding to the mean prediction at a scale factor of one are shown for comparison. (a) Records using the AR Method. (b) Records using the MR-BR Method. (c) Records using the ε-BR Method. (d) Records using the CMS-ε Method.

The conclusions drawn in Chapter 6 appear to hold for these two structures as well. There will still be some variation in estimated structural response because of the finite number of records used (this might be the source of the MR-BR records causing smaller levels of response in the 0.3s structure than had been expected). On average, however, the ε-BR and CMS-ε records tend to produce smaller mean responses and smaller probabilities of collapse for a given $Sa(T_1)$ level than do the AR and MR-BR records. As was argued in Chapter 6, the smaller responses from the ε-BR and CMS-ε records are more accurate than responses from the other records, because they account for the effect of ε, which has been observed to be a useful predictor of structural response.

References

Abrahamson N.A., 1988. Statistical properties of peak ground accelerations recorded by the SMART 1 array, *Bulletin of the Seismological Society of America*, **78** (1), 26-41.
Abrahamson N.A., 1993. Nonstationary spectral matching program RSPMATCH.
Abrahamson N.A. and Silva W.J., 1997. Empirical response spectral attenuation relations for shallow crustal earthquakes, *Seismological Research Letters*, **68** (1), 94-126.
Abrahamson N.A., 2000a. Effects of rupture directivity on probabilistic seismic hazard analysis, in *Sixth International Conference on Seismic Zonation*, *Oakland, California*.
Abrahamson N.A., 2000b. State of the practice of seismic hazard evaluation, in *GeoEng 2000, Melbourne, Australia*.
Abrahamson N.A., Kammerer A., and Gregor N., 2003. *Summary of scaling relations for spectral damping, peak velocity, and average spectral acceleration*, *Personal Communication*.
Abrahamson N.A., 2004. Personal communication.
Agresti A., 2002. *Categorical data analysis*, Wiley, New York, 710 pp.
Alavi B. and Krawinkler H., 2001. *Effects of near-fault ground motions on frame structures*, Blume Center Report #138, Stanford, California, 301 pp.
American Society of Civil Engineers, 2002. *ASCE standard: Minimum design loads for buildings and other structures*, American Society of Civil Engineers, *SEI/ASCE 7-02*, Reston, VA.
American Society of Civil Engineers, 2005. *Seismic design criteria for structures, systems, and components in nuclear facilities*, Structural Engineering Institute, Working Group for Seismic Design Criteria for Nuclear Facilities, *ASCE/SEI 43-05*, Reston, VA, 81 pp.
Anderson J.G., 1997. Benefits of scenario ground motion maps, *Engineering Geology*, **48**, 43-57.
Aslani H. and Miranda E., 2003. *Probabilistic response assessment for building-specific loss estimation*, Pacific Earthquake Engineering Research Center, University of California at Berkeley, PEER 2003-03, Berkeley, California, 49 pp.
Aslani H., 2005. *Probabilistic earthquake loss estimation and loss deaggregation in buildings*, Stanford University, *Ph.D. Thesis*, Stanford, CA.
Baez J. and Miranda E., 2000. Amplification factors to estimate inelastic displacement demands for the design of structures in the near field, in *Proceedings, 12th World Conference on Earthquake Engineering*, *Auckland, New-Zealand*, 8 pp.
Baker J.W. and Cornell C.A., 2003. *Uncertainty specification and propagation for loss estimation using FOSM methods*, Pacific Earthquake Engineering Research Center, University of California at Berkeley, PEER 2003-07, Berkeley, California, 89 pp.
Baker J.W. and Cornell C.A., 2004. Choice of a vector of ground motion intensity measures for seismic demand hazard analysis, in *Proceedings, 13th World Conference on Earthquake Engineering*, *Vancouver, Canada*, 15 pp.
Baker J.W. and Cornell C.A., 2005a. 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 Cornell C.A., 2005b. Which spectral acceleration are you using?, *Earthquake Spectra* (in press).

Banon H., Cornell C.A., Crouse C.B., Marshall P., Nadim F., and Younan A., 2001. ISO seismic design guidelines for offshore platforms, in *Proceedings, 20th OMAE, OMAE 2001/S&R-2114*, *Rio de Janeiro*.

Bazzurro P. and Cornell C.A., 1994a. Seismic hazard analysis of nonlinear structures I: Methodology, *Journal of Structural Engineering*, **120** (11), 3320.

Bazzurro P. and Cornell C.A., 1994b. Seismic hazard analysis of nonlinear structures II: Applications, *Journal of Structural Engineering*, **120** (11), 3345.

Bazzurro P., 1998. *Probabilistic seismic demand analysis*, Stanford University, Stanford, CA, 329 pp. http://www.stanford.edu/group/rms/

Bazzurro P., Cornell C.A., Shome N., and Carballo J.E., 1998. Three proposals for characterizing MDOF nonlinear seismic response, *Journal of Structural Engineering*, **124** (11), 1281-1289.

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

Bazzurro P. and Cornell C.A., 2002. Vector-valued probabilistic seismic hazard analysis, in *7th U.S. National Conference on Earthquake Engineering*, *Boston, MA*, 10 pp.

Bellman R.E., 1961. *Adaptive control processes: A guided tour*, Princeton University Press, Princeton, N.J.,, 255 pp.

Benjamin J.R. and Cornell C.A., 1970. *Probability, statistics, and decision for civil engineers*, McGraw-Hill, New York, 684 pp.

Bertero V., Mahin S., and Herrera R., 1978. Aseismic design implications of near-fault San Fernando earthquake records, *Earthquake engineering & structural dynamics*, **6** (1), 31-42.

Bommer J.J., Scott S.G., and Sarma S.K., 2000. Hazard-consistent earthquake scenarios, *Soil Dynamics and Earthquake Engineering*, **19**, 219-231.

Bommer J.J. and Acevedo A.B., 2004. The use of real earthquake accelerograms as input to dynamic analysis, *Journal of Earthquake Engineering*, **08**, 43-91.

Boore D.M., Joyner W.B., and Fumal T.E., 1997. Equations for estimating horizontal response spectra and peak acceleration from western North American earthquakes: A summary of recent work, *Seismological Research Letters*, **68** (1), 128-153.

Boore D.M., 2005. Erratum: Equations for estimating horizontal response spectra and peak acceleration from western North American earthquakes: A summary of recent work, *Seismological Research Letters*, **76** (3), 368-369.

Bozorgnia Y. and Bertero V.V., 2004. *Earthquake engineering: From engineering seismology to performance-based engineering*, CRC Press, Boca Raton. Various pagings.

Campbell K.W., 1997. Empirical near-source attenuation relationships for horizontal and vertical components of peak ground acceleration, peak ground velocity, and pseudo-absolute acceleration response spectra, *Seismological Research Letters*, **68** (1), 154-179.

Carballo J.E. and Cornell C.A., 1998. Input to nonlinear structural analysis: Modification of available accelerograms for different source and site characteristics, in *6th U.S. National Conference on Earthquake Engineering*, *Seattle, WA*, 13 pp.

Carballo J.E., 2000. *Probabilistic seismic demand analysis: Spectrum matching and design*, Stanford University, Stanford, CA, 259 pp. http://www.stanford.edu/group/rms/

Chopra A.K., 2001. *Dynamics of structures: Theory and applications to earthquake engineering*, Prentice Hall, Upper Saddle River, NJ, 844 pp.

Conte J.P., Pandit H., Stewart J.P., and Wallace J.W., 2003. Ground motion intensity measures for performance-based earthquake engineering, in *Applications of Statistics and Probability in Civil Engineering: Proceedings of the 9th International Conference*, 1465-1472 pp.

Cordova P.P., Deierlein G.G., Mehanny S.S.F., and Cornell C.A., 2001. Development of a two-parameter seismic intensity measure and probabilistic assessment procedure, in *The Second U.S.-Japan Workshop on Performance-Based Earthquake Engineering Methodology for Reinforced Concrete Building Structures*, *Sapporo, Hokkaido*, 187-206 pp.

Cornell C.A. and Krawinkler H., 2000. Progress and challenges in seismic performance assessment, *PEER Center News*, **3** (2).

Cornell C.A., Jalayer F., Hamburger R.O., and Foutch D.A., 2002. Probabilistic basis for 2000 SAC federal emergency management agency steel moment frame guidelines, *Journal of Structural Engineering*, **128** (4), 526-533.

Deierlein G.G., 2004. Overview of a comprehensive framework for earthquake performance assessment, in *International Workshop on Performance-Based Seismic Design Concepts and Implementation*, *Bled, Slovenia*, 12 pp.

Department of Energy, 1996. *Guidelines for use of probabilistic seismic hazard curves at department of energy sites for department of energy facilities*, *Standard DOE-1024-92*. www.eh.doe.gov/techstds/standard/std1024/std1024.pdf (accessed 8/13/05)

Deutsch C.V. and Journel A.G., 1997. *GSLIB geostatistical software library and user's guide*, Oxford University Press, New York, 369 pp.

Ditlevsen O., 1981. *Uncertainty modeling: With applications to multidimensional civil engineering systems*, McGraw-Hill International Book Co., New York ; London, 412 pp.

Efron B. and Tibshirani R.J., 1993. *An introduction to the bootstrap*, Chapman & Hall, New York, 436 pp.

FEMA 350, 2000a. *Recommended seismic design criteria for new steel moment-frame buildings*, prepared for the Federal Emergency Management Agency, *SAC Joint Venture*, Washington, D.C.

FEMA 351, 2000b. *Recommended seismic evaluation and upgrade criteria for existing welded steel moment-frame buildings*, prepared for the Federal Emergency Management Agency, *SAC Joint Venture*, Washington, D.C.

FEMA 352, 2000c. *Recommended post-earthquake evaluation and repair criteria for welded steel moment-frame buildings*, prepared for the Federal Emergency Management Agency, *SAC Joint Venture*, Washington, D.C.

FEMA 356, 2000d. *Prestandard and commentary for the seismic rehabilitation of buildings*, Federal Emergency Management Agency, *American Society of Civil Engineers*, Washington, D.C.

Frankel A.D., Petersen M.D., Mueller C.S., Haller K.M., Wheeler R.L., Leyendecker E.V., Wesson R.L., Harmsen S.C., Cramer C.H., Perkins D.M., and Rukstales K.S., 2002. *Documentation for the 2002 update of the national seismic hazard maps*, U.S. Geological Survey, Open-File Report 02-420, 33 pp.

Frankel A.D., 2004. Personal communication.

Fu Q. and Menun C., 2004. Seismic-environment-based simulation of near-fault ground motions, in *Proceedings, 13th World Conference on Earthquake Engineering*, *Vancouver, Canada*, 15 pp.

Fu Q., 2005. *Modeling and prediction of fault-normal near-field ground motions and structural response*, Stanford University, *Ph.D. Thesis*, Stanford, CA.

Garrick B., 1984. Recent case studies and advancements in probabilistic risk assessment., *Risk analysis*, **4** (4), 267.

Han S.W. and Wen Y.K., 1994. *Method of reliability-based calibration of seismic structural parameters.*, University of Illinois and Urbana-Champaign, UIUC-ENG-94 2016, Stuctural Research Series No. 595, Champaign, IL, 170 pp.

Han S.W. and Wen Y.K., 1997. Method of reliability-based seismic design I: Equivalent nonlinear systems, *Journal of Structural Engineering*, **123** (3), 256-263.

Haselton C., Mitrani-Reiser J., Goulet C., Deierlein G.G., Beck J.L., Porter K.A., Stewart J.P., and Taciroglu E., 2005. *An assessment to benchmark the seismic performance of a code-conforming reinforced-concrete moment-frame building*, Pacific Earthquake Engineering Research Center, University of California at Berkeley, Berkeley, California (In preparation).

Hastie T., Tibshirani R., and Friedman J.H., 2001. *The elements of statistical learning : Data mining, inference, and prediction*, Springer, New York, 533 pp.

Ibarra L.F., 2003. *Global collapse of frame structures under seismic excitations*, Stanford University, Stanford, CA, 324 pp.

ICC, 2003. *International building code*, International Code Council.

Iervolino I. and Cornell C.A., 2005a. Record selection for nonlinear seismic analysis of structures, *Earthquake Spectra*, **21** (3), 685-713.

Iervolino I. and Cornell C.A., 2005b. Prediction of the occurence of velocity pulses in near-fault ground motions (in preparation).

Inoue T. and Cornell C.A., 1990. *Seismic hazard analysis of multi-degree-of-freedom structures*, Reliability of Marine Structures, RMS-8, Stanford, CA, 70 pp.

Iwan W.D., 1980. Estimating inelastic response spectra from elastic response spectra, *Earthquake Engineering & Structural Dynamics*, **8**, 375-388.

Jalayer F., 2003. *Direct probabilistic seismic analysis: Implementing non-linear dynamic assessments*, Stanford University, *Ph.D. Thesis*, Stanford, CA, 244 pp. http://www.stanford.edu/group/rms/ (accessed 3/14/05)

Jalayer F. and Cornell C.A., 2003. *A technical framework for probability-based demand and capacity factor design (DCFD) seismic formats*, Pacific Earthquake Engineering Research Center, University of California at Berkeley, PEER 2003-08, Berkeley, California, 106 pp.

Jalayer F., Beck J.L., and Porter K.A., 2004. Effects of ground motion uncertainty on predicting the response of an existing RC frame structure, in *13th World Conference on Earthquake Engineering*, *Vancouver, Canada*, 10 pp.

Kaplan S. and Garrick B.J., 1981. On the quantitative definition of risk, *Risk analysis*, **1** (1), 11-27.

Kennedy R.P., Short S.A., Merz K.L., Tokarz F.J., Idriss I.M., Power M.S., and Sadigh K., 1984. *Engineering characterization of ground motion - task I: Effects of characteristics of free-field motion on structural response*, U.S. Nuclear Regulatory Commision, NUREG/CR-3805, Washington, D.C.

Kennedy R.P., Kincaid R.H., and Short S.A., 1985. Prediction of inelastic response from elastic response spectra considering localized nonlinearities and soil-structure interaction, in *8th SMIRT*, 427-434 pp.

Kennedy R.P., Wesley D.A., and Tong W.H., 1988. *Probabilistic evaluation of the Diablo Canyon turbine building seismic capacity using nonlinear time history analysis*, Pacific Gas & Electric Co. Report Number 1643.01.

Kramer S.L., 1996. *Geotechnical earthquake engineering*, Prentice Hall, Upper Saddle River, N.J., 653 pp.

Krawinkler H., 2005. Personal communication.

Lay D., 1997. *Linear algebra and its applications*, Addison-Wesley, Reading, Massechusetts, 486 pp.

Lehmann E.L. and D'Abrera H.J.M., 1998. *Nonparametrics : Statistical methods based on ranks*, Prentice Hall, Upper Saddle River, N.J., 463 pp.

Liu P.L. and Der Kiureghian A., 1986. Multivariate distribution models with prescribed marginals and covariances, *Probabilistic Engineering Mechanics*, **1** (2), 105-112.

Luco N., 2002. *Probabilistic seismic demand analysis, SMRF connection fractures, and near-source effects*, Stanford University, Stanford, CA, 260 pp. http://www.stanford.edu/group/rms/

Luco N. and Cornell C.A., 2005. Structure-specific scalar intensity measures for near-source and ordinary earthquake ground motions, *Earthquake Spectra (Under revision for publication)*.

Luco N. and Bazzurro P., 2005. *Effects of earthquake record scaling on nonlinear structural response*, PEER Lifelines Program, *Report on PEER-LL Program Task 1G00 Addendum (Sub-task 1 of 3)*, 64 pp.

Luco N., Manuel L., Baldava S., and Bazzurro P., 2005. Correlation of damage of steel moment-resisting frames to a vector-valued set of ground motion parameters, in *9th International Conference on Structural Safety and Reliability*, *Rome, Italy*, 8 pp.

Mackie K. and Stojadinovic B., 2003. *Seismic demands for performance-based design of bridges*, Pacific Earthquake Engineering Research Center, University of California at Berkeley, PEER 2003-16, Berkeley, California, 150 pp.

Mahin S., Bertero V., Chopra A.K., and Collins R., 1976. *Response of the Olive View hospital main building during the San Fernando earthquake*, Earthquake Engineering Research Center, University of California, UCB/EERC-76/22, Berkeley, California.

Mavroeidis G.P. and Papageorgiou A.S., 2003. A mathematical representation of near-fault ground motions, *Bulletin of the Seismological Society of America*, **93** (3), 1099-1131.

McGuire R.K., 1995. Probabilistic seismic hazard analysis and design earthquakes: Closing the loop, *Bulletin of the Seismological Society of America*, **85** (5), 1275-1284.

McGuire R.K., Silva W.J., and Costantino C.J., 2001. *Technical basis for revision of regulatory guidance on design ground motions hazard- and risk-consistent ground motion spectra guidelines*, Division of Engineering Technology Office of Nuclear Regulatory Research U.S. Nuclear Regulatory Commission, *NUREG/CR-6728*, Washington, DC.

McGuire R.K., 2004. *Seismic hazard and risk analysis*, Earthquake Engineering Research Institute, Berkeley, 240 pp.

Medina R.A. and Krawinkler H., 2003. *Seismic demands for nondeteriorating frame structures and their dependence on ground motions*, Stanford University, *John A. Blume Earthquake Engineering Center Report No. 144*, Stanford, CA, 347 pp.

Medina R.A. and Krawinkler H., 2005. Evaluation of drift demands for the seismic performance assessment of frames, *Journal of Structural Engineering*, **131** (7), 1003-1013.

Mehanny S.S.F. and Deierlein G.G., 2000. *Modeling and assessment of seismic performance of composite frames with reinforced concrete columns and steel beams*, Stanford University, *John A. Blume Earthquake Engineering Center Report No. 135*, Stanford, CA, 442 pp.

Melchers R.E., 1999. *Structural reliability analysis and prediction*, John Wiley, Chichester ; New York, 437 pp.

Moehle J. and Deierlein G.G., 2004. A framework methodology for performance-based earthquake engineering, in *Proceedings, 13th World Conference on Earthquake Engineering*, *Vancouver, Canada*, 13 pp.

Mori Y., Yamanaka T., Luco N., Nakashima M., and Cornell C.A., 2004. Predictors of seismic demand of SMRF buildings considering post-elastic mode shape, in *13th World Conference on Earthquake Engineering*, *Vancouver, Canada*, 15 pp.

Naeim F. and Lew M., 1995. On the use of design spectrum compatible time histories, *Earthquake Spectra*, **11** (1), 111-127.

Neter J., Kutner M.H., Nachtsheim C.J., and Wasserman W., 1996. *Applied linear statistical models*, McGraw-Hill, Boston, 1408 pp.

Next Generation Attenuation (NGA) project strong motion database. 2005. http://peer.berkeley.edu/nga/ (accessed 8/9/2005).

Norwegian Technology Standards Institution, 1999. *Actions and action effects*, *N-003*, Oslo, Norway, 85 pp. http://www.standard.no/imaker.exe?id=1399 (accessed 3/13/05)

Nuclear Regulatory Commision, 1997. *Identification and characterization of seismic sources and determination of safe shutdown earthquake ground motion*, *Regulatory Guide 1.165*. www.nrc.gov/reading-rm/doc-collections/reg-guides/power-reactors/active/01-165/ (accessed 8/17/2005)

NUREG, 1983. *PRA procedures guide—a guide to the performance of probabilistic risk assessment for nuclear power plants*, Division of Engineering Technology Office of Nuclear Regulatory Research U.S. Nuclear Regulatory Commission, *NUREG/CR-2300*, Washington, DC.

Pacific Gas & Electric, 1988. *Final report of the Diablo Canyon long term seismic program*, US Nuclear Regulatory Commission, *Docket Numbers 50-275 and 50-323*.

PEER strong motion database. 2000. http://peer.berkeley.edu/smcat/ (accessed 6/29/2005).

PEER Testbeds project site, 2004, http://www.peertestbeds.net. (accessed 6/29/2005).

Penzien J. and Watabe M., 1975. Characteristics of 3-D earthquake ground motions, *Earthquake engineering & structural dynamics*, **3**, 365–373.

Pincheira J.A., Dotiwala F.S., and D'Souza J.T., 1999. Seismic analysis of older reinforced concrete columns, *Earthquake Spectra*, **15** (2), 245-272.

Purvance M., 2005. *Overturning of slender blocks: Numerical investigation and application to precariously balanced rocks in southern california*, University of Nevada, Reno, Reno, NV, 233 pp. ftp://dataworks.library.unr.edu/seismo/Data/PrecRock/purvance_dissertation.pdf (accessed 8/1/05)

Reiter L., 1990. *Earthquake hazard analysis: Issues and insights*, Columbia University Press, New York, 254 pp.

Rice J.A., 1995. *Mathematical statistics and data analysis*, Duxbury Press, Belmont, CA, xx, 602, A49 pp.

Ruiz-Garcia J., 2004. *Perfomance-based assessment of existing structures accounting for residual displacements*, 407 pp.

Sadigh K., Chang C.-Y., Egan J.A., Makdisi F., and Youngs R.R., 1997. Attenuation relationships for shallow crustal earthquakes based on california strong motion data, *Seismological Research Letters*, **68** (1), 180-189.

Sewell R.T., Toro G.R., and McGuire R.K., 1996. *Impact of ground motion characterization on conservatism and variability in seismic risk estimates*, U.S. Nuclear Regulatory Commission, NUREG/CR-6467, Washington, D.C.

Shome N., Cornell C.A., Bazzurro P., and Carballo J.E., 1998. Earthquakes, records, and nonlinear responses, *Earthquake Spectra*, **14** (3), 469-500.

Shome N., 1999. *Probabilistic seismic demand analysis of nonlinear structures*, Stanford University, 320 pp. http://www.stanford.edu/group/rms/

Shome N. and Cornell C.A., 1999. *Probabilistic seismic demand analysis of nonlinear structures*, RMS Program, RMS-35, Stanford, CA, 320 pp. http://www.stanford.edu/group/rms/ (accessed 3/14/05)

Shome N. and Cornell C.A., 2000. Structural seismic demand analysis: Consideration of "collapse", in *8th ASCE Specialty Conference on Probabilistic Mechanics and Structural Reliability, University of Notre Dame, South Bend, Indiana*, 7 pp.

Silva W.J., 1987. *WES RASCAL code for synthesizing earthquake ground motions*, *US Army Corps of Engineers State-of-the-Art for Assessing Earthquake Hazards in the United States Report 24*, Paper S-73-1.

Somerville P.G., Smith N.F., Graves R.W., and Abrahamson N.A., 1997. Modification of empirical strong ground motion attenuation relations to include the amplitude and duration effects of rupture directivity, *Seismological Research Letters*, **68** (1), 199-222.

Somerville P.G., 2001a. *PEER testbeds project report: Ground motion time histories for the UC lab building*. www.peertestbeds.net/UCS/Somerville_Dec_2001_UCS_ground_motions.pdf (accessed 8/11/2005)

Somerville P.G., 2001b. *PEER testbeds project report: Ground motion time histories for the Van Nuys building.* http://peer.berkeley.edu/peertestbeds/VNY/PrelimVanNuysHazardReport.pdf (accessed 8/11/05)

Somerville P.G. and Thio H.K., 2003. *Probabilistic vector-valued ground motion intensity measures and engineering demand measures for the PEER Van Nuys Holiday Inn PBEE testbed.*, SCEC 2003 Project Report.

Somerville P.G., 2003. Magnitude scaling of the near fault rupture directivity pulse, *Physics of the earth and planetary interiors*, **137** (1), 12.

Somerville P.G. and Thio H.K., 2005. *Probabilistic vector-valued seismic hazard analysis for near-fault ground motions*, SCEC Project Report (in preparation).

Spudich P., Joyner W.B., Lindh A.G., Boore D.M., Margaris B.M., and Fletcher J.B., 1999. SEA99: A revised ground motion prediction relation for use in extensional tectonic regimes, *Bulletin of the Seismological Society of America*, **89** (5), 1156-1170.

Stewart J.P., Chiou S.-J., Bray J.D., Graves R.W., Somerville P.G., and Abrahamson N.A., 2001. *Ground motion evaluation procedures for performance-based design*, Pacific Earthquake Engineering Research Center, University of California at Berkeley, PEER 2001-09, Berkeley, California, 229 pp.

Stewart J.P., Chiou S.J., Bray J.D., Graves R.W., Somerville P.G., and Abrahamson N.A., 2002. Ground motion evaluation procedures for performance-based design, *Soil dynamics and earthquake engineering*, **22** (9-12), 765-772.

Taghavi S. and Miranda E., 2003. Probabilistic study of peak floor acceleration demands in linear structures, in *Proceedings, Ninth International Conference on Applications of Statistics and Probability in Civil Engineering, San Francisco, California*.

Toro G.R. and Silva W.J., 2001. *Scenario earthquakes for Saint Louis, MO, and Memphis, TN, and seismic hazard maps for the central United States region including the effect of site conditions*, For US Geological Survey, award number 1434-HQ-97-GR-02981. http://www.riskeng.com/PDF/Scen_CEUS_Rept.pdf (Accessed 6/1/2005)

Tothong P. and Cornell C.A., 2005a. Near-fault ground motions for seismic demand analysis (in preparation).

Tothong P. and Cornell C.A., 2005b. An empirical ("attenuation law") ground motion prediction model for inelastic spectral displacement, *Bulletin of the Seismological Society of America (In Preparation)*.

Tothong P., Iervolino I., Baker J.W., and Cornell C.A., 2005. Probabilistic seismic hazard analysis accounting for the possibility of near-fault velocity pulses (in preparation).

U.S. Geological Survey earthquake hazard maps. 2002. http://eqhazmaps.usgs.gov/ (accessed July 2 5, 2005).

Vamvatsikos D., 2002. *Seismic performance, capacity and reliability of structures as seen through incremental dynamic analysis*, Stanford University, Stanford, CA, 152 pp. http://www.stanford.edu/group/rms/

Vamvatsikos D. and Cornell C.A., 2002. Incremental dynamic analysis, *Earthquake Engineering & Structural Dynamics*, **31** (3), 491-514.

Vamvatsikos D. and Cornell C.A., 2004. Applied incremental dynamic analysis, *Earthquake Spectra*, **20** (2), 523-553.

Vamvatsikos D. and Sigalas I., 2005. Seismic performance evaluation of a horizontally curved highway bridge using incremental dynamic analysis in 3D, in *Proceedings, 4th European workshop on the seismic behavior of irregular and complex structures*, 12 pp.

Van Nuys hotel building testbed report: Exercising seismic performance assessment, 2004, Pacific Earthquake Engineering Research Center, University of California at Berkeley, Berkeley, California.

Wang G.-Q., Zhou X.-Y., Ma Z.-J., and Zhang P.-Z., 2001. A preliminary study on the randomness of response spectra of the 1999 Chi-Chi, Taiwan, earthquake, *Bulletin of the Seismological Society of America*, **91** (5), 1358-1369.

Winterstein S.R., Ude T.C., and Cornell C.A., 1993. Environmental parameters for extreme response: Inverse FORM with omission factors, in *Proceedings, 6th International Conference on Structural Safety and Reliability, Innsbruck, Austria*, 551-557 pp.

Yun S.-Y., Hamburger R.O., Cornell C.A., and Foutch D.A., 2002. Seismic performance evaluation for steel moment frames, *Journal of Structural Engineering*, **128** (4), 12.