Title

Non-ergodic ground-motion models for California, Ground-motion embedment factors for the Seattle Region, and Global fault displacement model

Author

Lavrentiadis, Grigorios

Publication Date

2021

Non-ergodic ground-motion models for California, Ground-motion embedment factors for the Seattle Region, and Global fault displacement model

by

Grigorios Lavrentiadis

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Civil and Environmental Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Norman A. Abrahamson, Co-chair
Professor Adda Athanasopoulos-Zekkos, Co-chair
Professor Nicholas Sitar
Professor Douglas S. Dreger

Spring 2021

Non-ergodic ground-motion models for California,
Ground-motion embedment factors for the Seattle Region, and
Global fault displacement model

Abstract

Non-ergodic ground-motion models for California,
Ground-motion embedment factors for the Seattle Region, and
Global fault displacement model

by

Grigorios Lavrentiadis

Doctor of Philosophy in Engineering - Civil and Environmental Engineering

University of California, Berkeley

Professor Norman A. Abrahamson, Co-chair

Professor Adda Athanasopoulos-Zekkos, Co-chair

Three topics related to seismic hazard estimation are presented in this study: a non-ergodic ground-motion model (GMM) for California, an approach to adjust a GMM for the site conditions in Pacific North West, and a fault displacement model.

A new approach is used in the development of a fully non-ergodic GMM for pseudo-spectral accelerations (PSa). First, a non-ergodic effective Fourier amplitude spectrum (EAS) GMM is derived, and then, through random vibration theory (RVT), it is converted to a PSa non-ergodic GMM. The advantage of this two-step approach is that it can better capture the non-ergodic source, path, and site effects through the small magnitude earthquakes. Fourier transform is a linear operator, and therefore, the non-ergodic effects from the small magnitude events can be applied directly to the large magnitude earthquakes. The Bayless and Abrahamson (2019b) ergodic EAS GMM is used as a backbone for the non-ergodic EAS GMM; the non-ergodic effects related to the source and site are modeled as spatially varying coefficients, and the non-ergodic effects related to the path are captured through a cell-specific anelastic attenuation. Two non-ergodic PSa GMMs are developed using the ASK14 (Abrahamson et al., 2014) and CY14 (Chiou and Youngs, 2014) ergodic GMMs as backbone models, respectively. The PSa non-ergodic effects are expressed as ergodic to non-ergodic PSa ratios, which is the adjustment that needs to be applied to an ergodic PSa GMM to incorporate the non-ergodic effects. To calculate these ratios, first both the ergodic and non-ergodic EAS are calculated for a scenario of interest (M, R_{rup}, V_{S30}, x_{eq}, x_{site}, etc.) and then, with RVT, the equivalent PSa values are computed. It is the second step that introduces the magnitude dependence in the non-ergodic PSa terms. This approach leads to a 30 to 35% reduction in the total aleatory standard deviation compared to the corresponding

ergodic GMMs. The epistemic uncertainty associated with the PSa ratios is small in areas close to stations and past events; in areas with sparse data, the mean of the non-ergodic ratios goes to zero implying ergodic scaling and the epistemic uncertainty increases.

The site amplification in most GMMs is quantified by the time-average shear-wave velocity of the top $30m$ (V_{S30}). However, V_{S30} is not a fundamental physical property that controls site amplification. It works as predictor variable for site amplification due to its correlation with the site velocity profile ($V_S(z)$), which depends on the empirical data set used to develop each GMM. The V_{S30} scaling of the NGAWest2 and subduction GMMs may not be applicable to Seattle, as the geological environment in Seattle is different to the geological environments in California and Japan, where most of the data in NGAWest2 and subduction GMMs were recorded, respectively. GMM-to-site scale factors are developed to adjust the NGAWest2 and subduction GMM to the site conditions in Pacific North West. The amplification ratios between the $V_S(z)$ profiles implied by the GMMs and $V_S(z)$ profiles representative for Seattle are estimated with $1D$ site response analyses.

A model in the wavenumber domain is developed to describe the surface fault displacements for use in probabilistic fault rupture hazard analysis (PFDHA). The advantages of this method are that it avoids the surface-rupture length normalization and that it considers the along strike correlation of displacements. A regularized Fourier Transform (RFT) is used to compute the Fourier spectra from unevenly sampled surface-slip data which could be potentially biased towards the peaks. The amplitude spectrum is based on the Somerville et al. (1999) model used for the generation of slip distributions in kinematic simulations, and the phase-derivative model is defined as a logistic distribution. Compared to previous models, the wavenumber-spectrum method leads to narrower tails of the slip distribution which is important for PFDHA at long return periods. Near the center of the rupture, the wavenumber-spectrum method gives slip distributions that are consistent with the distributions from the empirical data, but at the ends of the rupture, the wavenumber-spectrum method underestimates the range of the slip. This discrepancy may reflect limitations of the current data sets in terms of mapping the slip near the ends of the ruptures. Improved surface rupture data sets that are currently being compiled will provide improved constrains at the ends of the ruptures.

Dedication

To my parents Constantinos and Maria,
my elementary teacher Constantinos Michos,
and undergraduate professor Demos C. Angelides

Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my sincere gratitude to my advisor, Professor Norman Abrahamson, for his mentoring and support throughout the course of this work. His kindness, modesty, selfless, and dedication to someone else learning are inspiring. I am also thankful to my committee members: Professors Adda Athanasopoulos-Zekkos, Douglas Dreger, and Nicholas Sitar for their review and constructive comments.

A big thank you to Albert Kottke, Nicolas Kuehn, and Chris Madugo for their knowledge and experience. I would like to thank the GeoSystem students: Christie Hale, Zorana Mijic, Franklin Olaya Trinidad, Tessa Williams, and Zhen Yang for their friendship and support.

Partial funding for this research was provided by the PG&E Geosciences Department Long-Term Seismic Program and by Caltrans funding to PEER under PEER TSRP Solicitation 18-01.

Chapter 1

Introduction

1.1 Overview

In seismically active regions, the estimation of earthquake hazard is of paramount importance due to the treat to human life and significant financial losses. Any given earthquake can lead to a number of different hazards including: ground shaking and surface rupture.

Traditionally, the hazard due to ground shaking is estimated through a probabilistic seismic hazard analysis (PSHA), which computes the annual rate of exceeding a ground-motion parameter at a site of interest. PSHA breaks the problem in two parts: the seismic source characterization (SSC), and the ground-motion estimation. SSC involves the modeling of the earthquake scenarios which includes the rate, magnitude distribution and location of earthquakes that affect the seismic hazard at the site of interest. For a given earthquake scenario, the level of ground motion at the site of interest is estimated through Ground-motion models (GMMs). Due to the small number of regional data, GMMs are typically derived with global ground-motion data sets based on the ergodic assumption. These models have robust median estimates due to the larger data sets of global models, but they also have large aleatory variability, which is the range of ground motion for a given earthquake scenario, due to combining data from different regions with different medians. The large aleatory variability of the ergodic GMMs leads to flat hazard curves which, for important structures that are designed for large return periods, results to large values of design ground motions. In the non-ergodic approach, it is recognized that different parts of the world have systematic differences in the source, path, and site effects, which leads to a reduction aleatory variability, as some of the assumed ergodic aleatory variability is actually repeatable and hence predictable. The repeatable effects lead to a shift in the median values (either increase or decrease). The range of systematic effects is quantified by the epistemic uncertainty, which in areas with little data, where the systematic effects cannot be estimated, accommodates the reduction of aleatory variability. Non-ergodic GMMs are a promising development in PSHA because their smaller aleatory variability can have a large impact on the seismic hazard at large return periods. Non-ergodic GMMs for the state of California are developed in this study.

In the Seattle region, it is common to use the NGAWest2 GMMs to estimate the ground motion from shallow crustal events, and subduction GMMs to estimate the ground motion from the subduction zones. Both GMM types use the time-averaged shear-wave velocity measured from the surface to a depth of 30m, V_{S30}, as the main site parameter. However, it may not be appropriate to use the site V_{S30} in the NGA-W2 GMMs in Seattle, which often done by the state of practice, as V_{S30} is not a fundamental physical property that controls site amplification. V_{S30} works as a predictor of site amplification due to its correlation with the site $V_S(z)$ profile, but the $V_{S30} - V_S(z)$ correlation depends on the empirical ground-motion data set used to develop each GMM. The geological environment in Seattle, in the scale of kilometers, is different to the geological environment in California, where most of the data in NGAWest2 data set were recorded, and Japan, where most of the data used in subduction GMMs were recorded. An approach to adjust shallow crustal and subduction GMM for the site conditions in the Pacific North West is presented in this study.

The surface rupture hazard is estimated in a framework that is similar to ground-motion hazard. The main difference between PSHA and probabilistic fault displacement hazard analysis (PFDHA) is that instead of a GMM a fault displacement model (FDM) is used. Current FDMs are derived with small data-sets and simplified statistical assumptions, which assume that the fault displacements along the strike of the fault are uncorrelated. Various tapering functions (i.e. elliptical, quadratic, triangular, etc.), which have been adopted in existing FDMs to model the reduction of slip towards the ends of ruptures, are based on empirical observations but without a underlying physical basis to constrain their extrapolation to larger magnitude earthquakes. Furthermore, existing FDMs normalize the along-strike distance (x) by the surface rupture length (SRL) implying that the tapering function scales linearly with SRL, which may not be appropriate. The wavenumber model developed in this study considers the along-strike slip correlation, avoids the x/SRL normalization, and is consistent with the seismological modeling of slip on rupture planes.

1.2 Organization

This thesis is organized in the following chapters:

- Chapter 2 describes the formulation of a non-ergodic ground-motion model (GMM) for effective amplitude spectral (EAS) values for California. EAS, which is defined in Goulet et al. (2018), is a smoothed rotation-independent Fourier amplitude spectrum of the two horizontal components of an acceleration time history. The main motivation for developing a non-ergodic EAS GMM, rather than a spectral acceleration GMM, is that the scaling of EAS does not depend on spectral shape, and therefore, the more frequent small magnitude events can be used in the estimation of the non-ergodic terms. The model is developed using the California subset of the NGAWest2 dataset (Ancheta et al., 2013). The Bayless and Abrahamson (2019b) (BA18) ergodic EAS GMM was used as backbone to constrain the average source, path, and site scaling. The non-ergodic GMM is formulated as a Bayesian hierarchical model: the non-ergodic source and site terms are modeled as spatially varying coefficients following the approach of Landwehr et al. (2016), and the non-ergodic path effects are captured by the cell-specific anelastic attenuation attenuation following the approach of Dawood and Rodriguez-Marek (2013). Close to stations and past events, the mean values of the non-ergodic terms deviate from zero to capture the systematic effects and their epistemic uncertainty is small. In areas with sparse data, the epistemic uncertainty of the non-ergodic terms is large, as the systematic effects cannot be determined. The non-ergodic total aleatory standard deviation is approximately 30 to 40% smaller than the total aleatory standard deviation of BA18. This reduction in the aleatory variability has a significant impact on hazard calculations at large return periods. The epistemic uncertainty of the ground motion predictions is small in areas close to stations and past events.

- Chapter 3 describes a new approach for creating a non-ergodic PSa ground-motion model which account for the magnitude dependence of the non-ergodic effects. In this approach, the average PSa scaling is controlled by an ergodic PSa GMM, and the non-ergodic effects are captured with non-ergodic PSa ratios, which are the adjustment that needs to be applied to an ergodic PSa GMM to incorporate the non-ergodic effects. The non-ergodic PSa ratios are based on EAS non-ergodic effects and are converted to PSa through Random Vibration Theory (RVT). The advantage of this approach is that it better captures the non-ergodic source, path, and site effects through the small magnitude earthquakes. Due to the linear properties of Fourier Transform, the EAS non-ergodic effects of the small events can be applied directly to the large magnitude events. This is not the case for PSa, as response spectrum is a non-linear operator, making PSa non-ergodic effects depended on the spectral shape. Two PSa non-ergodic GMMs are derived using the ASK14 (Abrahamson et al., 2014) and CY14 (Chiou and Youngs, 2014) GMMs as backbone models, respectively. The non-ergodic EAS effects are estimated with the $\mathrm{LAK21_{EAS}}$ GMM, presented in Chapter 3. The RVT calculations are performed with the V75 (Vanmarcke, 1975) peak factor model, the D_{5-75a} estimate of AS96 (Abrahamson and Silva, 1996) for the ground-motion duration, and BT15 (Boore and Thompson, 2015) oscillator duration model. The California subset of the NGAWest2 database (Ancheta et al., 2014) is used for both models. The total aleatory standard deviation of the two non-ergodic PSa GMMs is approximately 30 to 35% smaller than the total aleatory standard deviation of the corresponding ergodic PSa GMMs. This reduction has a significant impact on hazard calculations at large return periods. In remote areas, far from stations and past events, the reduction of aleatory variability is accommodated by an increase of epistemic uncertainty.

- Chapter 4 describes an approach to adjust a ground-motion model for the site conditions in Pacific North West. In most GMM, the site conditions are quantified by the time-averaged shear-wave velocity measured from the surface to a depth of 30m (V_{S30}).

However, the V_{S30} works as a predictor of site amplification due to the correlation of V_{S30} and the site $V_S(z)$ profile. This correlation between V_{S30} and $V_S(z)$ implicit in GMMs depends on the empirical ground-motion data set used to develop the V_{S30} dependence of the site amplification in the GMM. The NGA-W2 GMMs (Gregor et al., 2014) for shallow crustal events are dominated by data from California, so the correlation between the site amplification and the V_{S30} in the NGA-W2 GMMs represents the depositional environment in California. Similarly, subduction GMMs are derived with ground-motion data that is mainly from Japan; as a result, the correlation between the site amplification and the V_{S30} in the subduction GMMs represents the depositional environment in Japan. In the scale of kilometers which affects ground-motion amplification of long periods, the depositional environment in Seattle is different to the depositional environments in California and Japan. In this chapter, GMM-to-site scale factors for shallow crustal and subduction GMMs are developed for structure in Seattle with a fundamental period greater than 1 *sec*. For this, the amplification ratios between the host and target profiles are estimated with $1D$ site response analyses. The host profiles are $V_S(z)$ that consistent with the CA and JP V_{S30} scaling, while the target profiles are representative for the site conditions in Seattle.

- Chapter 5 describes a new wavenumber-domain methodology to model surface slip profiles and generate potential displacement profiles for use in PFDHA. The benefits of this approach are that it captures the correlation of the surface slip variability along strike and it avoids the surface rupture length normalization. A regularized Fourier Transform (RFT) approach is used to compute the Fourier spectra from uneven sampling and biased sampling, typical of surface-slip data sets. The wavenumber amplitude spectrum for surface displacements is modeled by a functional form based on the shape of the Butterworth filter. The proposed RFT approach is validated using synthetic data-set with known model parameters which are down-sampled to be consistent with the sampling in empirical surface-rupture data sets. Preliminary models for the scaling of the amplitude and phase derivative as a function of the surface rupture length are developed using a subset of the earthquakes compiled by Wesnousky (2008). The analyzed events range from magnitude 6.1 to 7.9 and include both single and multi-segment ruptures.

Chapter 2

A Non-ergodic Effective Amplitude Ground-Motion Model for California

The contents of this chapter are primarily from a journal article submitted to the Bulletin of Earthquake Engineering by Lavrentiadis, G., Abrahamson, N.A., and Kuehn, M.N., entitled: "A Non-ergodic Effective Amplitude Ground-Motion Model for California", which is currently under review.

2.1 Introduction

Probabilistic seismic hazard analyses (PSHA) estimates the annual rate of exceeding a ground-motion parameter at a site of interest. It typically breaks the problem in two parts: the seismic source characterization and the ground-motion model (GMM). The first part defines the rate, the magnitude distribution, and the location distribution of earthquakes in a region, and the second part, which is the focus of this study, provides the probability of exceeding the ground-motion for a specific earthquake scenario. Most GMM are derived empirically using strong-motion datasets and define the distribution of the ground-motion parameter as a function of source, path, and site parameters such as magnitude (M), closest rupture distance (R_{rup}) and time-averaged shear-wave velocity in the top $30m$ (V_{S30}). Traditionally, due to the scarcity of regional data, GMM were developed under the ergodic approach which assumes the statistical properties of the ground-motion parameter do not change in space (Anderson and Brune, 1999). These ergodic models tend to have large aleatory variability as they treat some of the systematic effects for a specific site/source location as random variability that can occur anywhere. Examples of models that have been developed under this approach are the NGA-West GMMs for California (Abrahamson et al., 2008), and the Douglas et al. (2014) GMM for Europe; however, as more data are collected, the ergodic assumption can be relaxed, and repeatable effects related to the source, path and site can be properly modeled, which leads to a decrease in the aleatory variability. This reduction has a large impact on the hazard at large return periods because, the ground-motion aleatory variability controls the slope of the hazard curves which has a large influence on the hazard at large return periods. Al Atik et al. (2010) describes how the aleatory variability of an ergodic model can be separated into epistemic uncertainty related to the systematic source, path, and site terms, and the reduced aleatory variability of a non-ergodic GMM. The epistemic uncertainty refers to the range by which the non-ergodic terms vary in areas with no available data to constrain them. In areas with data from past earthquakes, the non-ergodic terms can be estimated, and their epistemic uncertainty can be reduced.

The first step in this new paradigm was to create a regional GMM or a global GMM with some regionalized terms. Regional GMMs are developed with smaller regional data-sets, for instance Akkar and Çağnan (2010) for Turkey, Bindi et al. (2011) for Italy, Bragato and Slejko (2005) for Eastern Alps, and Danciu and Tselentis (2007) for Greece. These models have smaller aleatory variability than global ergodic GMMs, but they suffer from weaker constrains on the scaling due to the smaller size of the regressed data sets. Global GMMs with regionalized terms are developed with large global datasets, the same way ergodic GMM are developed but with the difference that some of the scaling terms are estimated separately for each region. The NGAWest2 GMM (Bozorgnia et al., 2014) followed this approach. For example, in Abrahamson et al. (2014), both the V_{S30} and anelastic attenuation scaling terms were regionalized: they have a different set of coefficients for Califonia, China, Japan, and Taiwan.

Partially non-ergodic GMMs that only capture the systematic site effects, known as single station GMMs, lead to an approximately 30% reduction in the aleatory variance compared to an ergodic GMM (Coppersmith et al., 2014; Bommer et al., 2015; Tromans et al., 2019). Similarly, other researchers have developed partially non-ergodic GMMs that capture repeatable effects related to the source (Yagoda-Biran et al., 2015), path (Dawood and Rodriguez-Marek, 2013; Kuehn et al., 2019), and single source/single site (Hiemer et al., 2011). Fully non-ergodic GMMs include non-ergodic terms to capture simultaneously all the aforementioned systematic effects (source, path, and site); these type of models have 60 to 70% smaller aleatory variance than ergodic GMMs (Lin et al., 2011; Landwehr et al., 2016; Abrahamson et al., 2019).

The model presented in this study is a fully non-ergodic GMM that captures the systematic effects of the source, site, and anelastic attenuation from the path. It is developed as spatially varying coefficient model (VCM), following the methodology used in Bussas et al. (2017) and Landwehr et al. (2016). The non-ergodic anelastic attenuation is modeled with the cell-specific anelastic attenuation similar to Dawood and Rodriguez-Marek (2013) and Abrahamson et al. (2019). VCMs impose a spatial correlation on the coefficients, so that they vary continuously from location to location; some of the methods to impose the spatial correlation are: splines, Gaussian Processes (GPs), or neural networks. This is done in order to reduce the number free parameters and to get a more stable estimate of the non-ergodic effects. In the cell-specific anelastic attenuation approach, the domain of interest is broken into cells which all have their own attenuation coefficient. The attenuation along a path from a source to site is equal to the sum of attenuation of the cells that it traverses.

One distinction of the current model from other non-ergodic GMMs is that it is developed for effective amplitude spectral (EAS) values instead of response spectral accelerations. This is done so that the more frequent smaller magnitude earthquakes (less than magnitude 3) can be used in the estimation of the non-ergodic terms for linear effects and still be applicable to larger magnitudes. The response spectrum is a non-linear operator and the scale factors for response spectral values depend on the spectral shape. This dependence in on the spectral shape makes it difficult to use small magnitude earthquakes in the development of a GMM that is applicable to larger magnitudes due to the differences in the spectral shapes of small and large magnitude earthquakes. This happens because the response spectral value is the peak in the time domain and the frequencies that affect the peak value depend on the spectral shape. The shape of a response spectrum and period of the peak change with magnitude: as the magnitude of an earthquake increases, the period of the response spectrum peak also increases, making the short period scaling of a PSa GMM magnitude dependent. Therefore, the short-period non-ergodic terms from smaller magnitudes may not be applicable to larger magnitudes without modification.

Figure 2.1 shows the average shape of a normalized response spectrum for a $M3$ and $M7.5$ earthquake: the $M3$ response spectrum has its peak at $T = 0.1sec$ whereas, the $M7$ response spectrum from has its peak at $T = 0.3sec$. Due to this spectral shape difference, the PGA scaling will be consistent with the scaling of the $PSa(T = 0.1sec)$ at small magnitudes and with the scaling of the $PSa(T = 0.3sec)$ at large magnitudes. For example, to properly capture the magnitude dependence in a PSa GMM, the V_{S30} coefficient for PGA would be close to the V_{S30} coefficient for $T = 0.1sec$ at small magnitudes, and it should gradually shift towards the V_{S30} coefficient for $T = 0.3sec$ as magnitude increases.

The EAS is defined in Goulet et al. (2018) as the smoothed power-averaged Fourier amplitude spectrum (FAS) of the two horizontal components. The EAS does not suffer from the same problem as PSa because the Fourier transform is a linear operation, and the scaling coefficients and non-ergodic terms estimated from small magnitude earthquakes can be applied directly to large events for linear source, path, and site effects. To ensure that the proposed model extrapolates reasonably to scenarios outside the rage of events in the regression data set, we use the Bayless and Abrahamson (2019b) ergodic EAS GMM (BA18) as a backbone model; we selected BA18 because it was developed on a large global data-set, and it includes seismological constraints to avoid oversaturation at short distances and large magnitudes. The non-ergodic terms and cell-specific anelastic attenuation coefficients were estimated with the total residuals of BA18 from the NGAWest2 California subset.

Lastly, one common comment regarding the usage of an EAS GMM is that, in most seismic design methods, the intensity of the ground-motion shaking is specified in terms of PSa and not EAS. We can use an EAS GMM and Random Vibration Theory (RVT) to compute the equivalent PSa. RVT uses the FAS and duration of an SDOF oscillator response to a ground motion to compute the root-mean-square amplitude of the response (x_{rms}) using Parseval's theorem, and it uses the product of x_{rms} with a peak factor to estimate the peak response of the SDOF, which is the definition of spectral acceleration. Boore (1983) used RVT with FAS from seismological theory to calculate the equivalent PSa, and Bora et al. (2015, 2019) derived a duration model which, when used with a Brune (1970, 1971) omega-squared

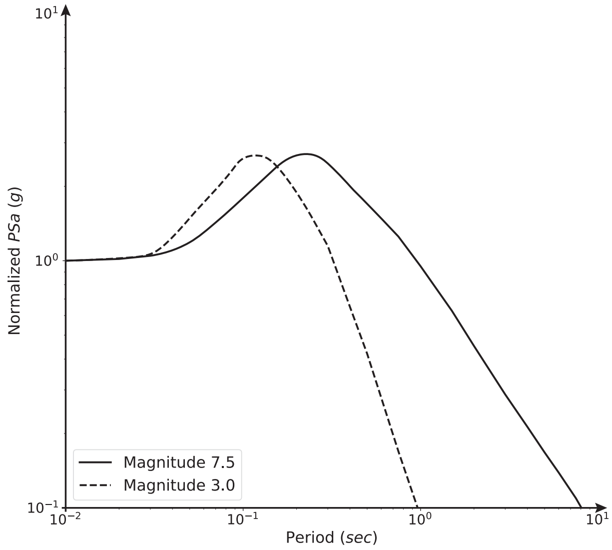


Figure 2.1: Schematic of normalized response spectra for M 3.0 and 7.5 earthquakes

FAS model, gives predictions that are consistent with the NGAWest2 data set. Converting the non-ergodic EAS GMM into an equivalent PSa GMM is not in the scope of this paper; it is covered in the second part of this study.

2.2 Ground-Motion Data

The NGAWest2 data-base (Ancheta et al., 2014) includes more than 21000 recordings covering a magnitude range from 3 to 7.9 and a closest distance range (R_{rup}) from 0.05 to 1500 km. For this study, a subset of this data-base was used which included the earthquakes and stations located in California, western Nevada, and northern Mexico. The recordings that were identified as questionable in Abrahamson et al. (2014) were not used in the regressions. The final data-set contains 8916 recordings from 188 earthquakes recorded at 1497 stations. Figure 2.2 shows the magnitude-distance distribution of the data and the number of recordings per frequency. The earthquake magnitudes in the selected data range from 3.1 to 7.3 and the distances range from 0.1 to 300 km. The usable frequency range of the majority of the recordings spans from 1 and $10hz$.

Figure 2.3 shows the spatial distribution of the data: most of the stations are located in Los Angeles, Bay Area, and San Diego metropolitan areas, whereas in less populated areas, such as northern-eastern California the spatial density of the stations is lower. This difference in the density of stations has a large impact on the distribution of epistemic uncertainty of the non-ergodic GMM: the epistemic uncertainty is higher is areas with lower station density which makes a case for expanding the strong-motion networks in these regions. In this study,

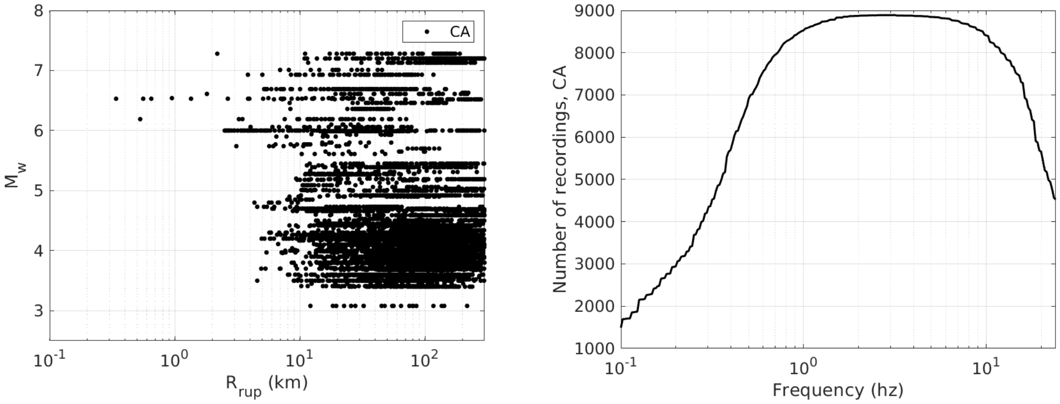


Figure 2.2: Selected data from the NGAWest2 database. The left figure shows Magnitude - Distance distribution of the subset used in the regression analysis. The right figure the number of recordings per frequency used in the regression analysis.

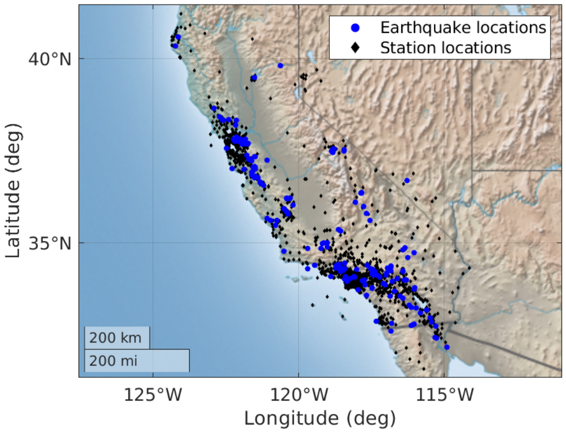


Figure 2.3: Spatial distribution for earthquakes and station used in this study.

the location of the earthquakes and stations is defined in kilometers in UTM coordinates; the longitude/latitude coordinates were transformed to UTM coordinates using the WGS84 reference ellipsoid and 11S UTM zone.

2.3 Non-ergodic Model Development

Rather than developing the non-ergodic GMM from scratch, the Bayless and Abrahamson (2019b) (BA18) ergodic EAS GMM was used as a back-bone model to describe the average ground-motion scaling. The main reasons for that decision were that: i) the local data may non be adequate to estimate the scaling of complex terms, and ii) the adoption of the constraints built into BA18 ensures that it extrapolates properly outside the range of data.

2.3.1 Functional Form

A GMM (Equation (2.1)) is composed of the median model (y_{med}) and the aleatory variability. The median model describes the average scaling of a ground-motion parameter with magnitude, distance, site conditions, etc.; it includes any systematic effects related the source, site, and path. The aleatory variability describes the misfit between a ground-motion observation and y_{med} which is related to the true or apparent (due to simplified modeling) stochastic behavior of the source, site and path. It is typically expressed as the sum of the between-event (δB_e) and within-event ($\delta W_{e,s}$) terms. $\delta B_{e,s}$ describes average shift of the ground motion for an earthquake, e, from y_{med}, and $\delta W_{e,s}$ describes the variability of the ground motion at site, s from the median ground motion of that earthquake.

$$ln(EAS) = y_{med} + \delta B_e + \delta W_{e,s} \tag{2.1}$$

The median EAS model of Bayless and Abrahamson (2019b) is formulated as:

$$y_{med} = f_M + f_R + f_S + f_{ztor} + f_{NM} + f_{Z1} \tag{2.2}$$

where f_M is the magnitude scaling, f_R is the path scaling, f_S is the site scaling, f_{ztor} is the top of rupture scaling, and f_{NM} is the normal-fault scaling. The f_M, f_R and f_S terms were modified to include the additional non-ergodic terms, whereas f_{ztor}, f_{NM} and f_{Z1} were kept fixed. The different terms and coefficients are described in detail in Bayless and Abrahamson (2019b) and Chiou and Youngs (2014).

Magnitude Scaling

The f_M in the non-ergodic model is:

$$f_M = c_1 + c_2(M-6) + \frac{c_2 - c_3}{c_n} \ln(1 + e^{c_n(c_M - M)}) + \delta c_0 + \delta c_{0N/S} + \delta c_{1a}(x_{eq}) \tag{2.3}$$

As an example, Figure 2.4 shows the magnitude scaling at $5hz$. In summary, c_1 is the intercept of the model, c_2 controls the magnitude scaling for large magnitudes where the corner frequency is smaller than the frequency of interest, $(c_2 - c_3)/c_n$ describes the magnitude scaling at small magnitudes, where the corner-frequency is larger than the frequency of interest, c_n controls the width of the transition zone between small and large magnitude earthquakes, and c_M is the magnitude at the center of the transition zone.

The additional coefficients in the non-ergodic model are δc_0, $\delta c_{0N/S}$, and δc_{1a}. δc_0 is added to allow a small constant shift in the non-ergodic model due the difference in the weighting of residuals between the ergodic and non-ergodic GMMs. δc_{1a} is a function of the earthquake

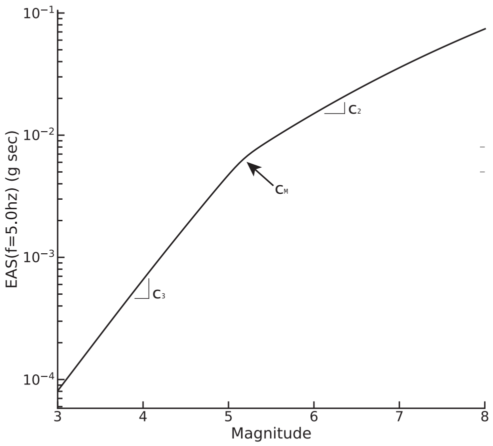


Figure 2.4: Sketch of Bayless and Abrahamson (2019b) magnitude scaling at $5hz$ for $V_{S30} = 500m/sec$ and $R_{rup} = 20km$

coordinates, x_{eq}, and is intended to capture repeatable non-ergodic effects related to the source location. For instance, regions with a higher than average median stress drop will have higher than average median ground-motions resulting in a positive δc_{1a}. $\delta c_{0N/S}$ is a regional term that corrects a potential bias in the magnitude estimation of small earthquakes between northern and southern California. $\delta c_{0N/S}$ is applied to earthquakes less than magnitude 5 and frequencies less than $5hz$. The vertices which define the polygons for the northern and southern CA subregions are summarized in Table 2.1. The border between northern and southern CA corresponds approximately to the boundary between the Northern California Seismic Network (NCSN) and Southern California Seismic Network (SCSN).

The $\delta c_{0N/S}$ term is believed to address potential magnitude inconsistencies between northern and southern CA because the NCSN/SCSN boundary was also used in the NGAWest2 dataset Ancheta et al. (2013) for the selection of source parameters (magnitude, hypocenter location, etc.) for small-to-moderate (less than M 5) earthquakes. If a small earthquake was located north of the boundary, the NCSN catalog was used for the source parameters, whereas if a small earthquake was located south of the boundary, the SCSN/CIT catalog was used for the source parameters. Preliminary regressions which did not include the $\delta c_{0N/S}$ term showed significant differences in δc_{1a} between northern and southern CA at small frequencies (Section 2.4.5). It was found that these differences were caused by a noticeable bias in the total residuals of BA18 between northern and southern CA for small magnitude events. Chiou et al. (2010) made a similar observation for the total residuals of CY08: they found a regional difference in median ground-motion amplitude between north and south CA earthquakes which was more noticeable at small magnitude events. The results in section 2.4 show that the difference in the median ground-motion of small events between northern and southern CA is approximately 0.4 in natural-log units at frequencies between 0.2 and 5.0 hz. At frequencies well below the corner frequency, a unit change in magnitude leads to a factor

Table 2.1: Vertices of Northern and Southern CA regions for $\delta c_{0N/S}$

Northern CA		Southern CA	
lat. (deg)	lon. (deg)	lat. (deg)	lon. (deg)
34.5175	-121.5250	37.9775	-116.6225
39.8384	-125.2341	35.2944	-113.4142
41.3595	-124.1684	31.4772	-115.0250
41.3995	-120.7227	31.0082	-117.6898
37.9775	-116.6225	34.5175	-121.5250

of 32 change in the amplitude of the ground motion; thus, a 0.4 natural-log difference in ground motion can be caused by a 0.11 bias in the magnitude estimation between the NCSN and SCSN networks, which could be due to different assumptions in the velocity models or other input parameters used to determine the magnitude of an event. Currently, the $\delta c_{0N/S}$ term is not applied to larger than M 5 events as the magnitude of those events is determined with fault inversion using data from global networks; however, this issue should be further investigated in future studies to find the exact cause of this regional difference.

Path Scaling

The functional form for f_P in the model is:

$$\begin{aligned} f_P =& c_4 \ln \left(R_{rup} + c_5 \cosh \left(c_6 \max(M - c_{hm}, 0) \right) \right) + (-0.5 - c_4) \ln(\hat{R}) \\ &+ \sum_{i}^{n_{cell}} c_{cA\ i}(x_{cell\ i}) \Delta R_i \end{aligned} \tag{2.4}$$

where $\hat{R} = \sqrt{R_{rup}^2 + 50^2}$. The coefficient c_4, which corresponds to the geometrical attenuation, manages how the ground motion attenuates at short distances. The coefficient c_5 describes the short-distance saturation, this term increases the effective distance for large magnitudes to capture the finite-fault effects (i.e. as the earthquake magnitude increases, the size of the rupture increases resulting in the seismic energy being spread out over more distant segments of the rupture leading to a larger effective rupture distance). Coefficients c_4 and c_6 control the magnitude saturation at short distances. Full saturation at zero distance is achieved when $c2 = -c_4 c_6$ is satisfied, non full saturation (i.e. positive magnitude scaling) is achieved when $c_2 > -c_4 c_6$. At distances greater than $50km$, the term $(-0.5 - c_4)$ cancels the empirically estimated geometrical attenuation and fixes it to 0.5 which is the theoretical geometrical attenuation of surface waves. To maintain proper distance scaling and magnitude saturation in the non-ergodic model, all the aforementioned coefficients were fixed to their $BA18$ values. The non-ergodic distance scaling is captured with the cell-specific anelastic attenuation.

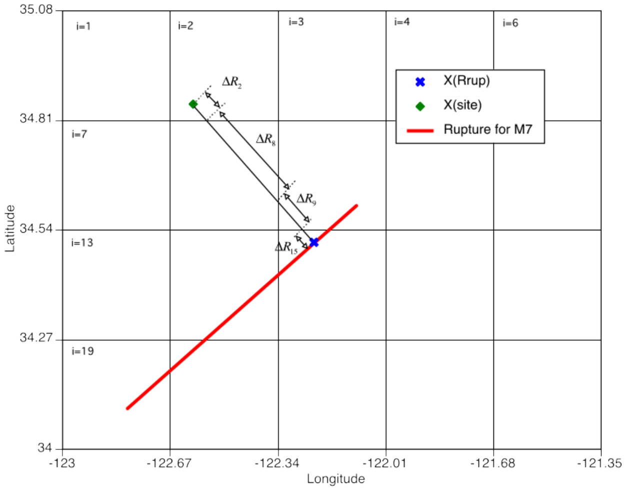


Figure 2.5: Schematic showing the calculation of the cell-path segments for the cell-specific anelastic attenuation. x_{site} is the site location, x_{cls} is the closest point on the rupture to the site, the dashed line indicates the source-to-site path, and the ΔR_i of the i^{th} cell (reproduced from Abrahamson et al. (2019)).

The anelastic attenuation is modeled with the cell-specific anelastic approach, first proposed by Dawood and Rodriguez-Marek (2013) and then extended by Kuehn et al. (2019) and Abrahamson et al. (2019). In this method, the states of California and Nevada are broken into $25 \times 25km$ cells and, for each record, the ray path which connects the site (x_{site}) to the closest point on the rupture (x_{cls}) is broken into cell-path segments (ΔR_i) which are lengths of the ray within each cell (Figure 2.5). For each record, the sum of cell-path segment lengths $\sum_{i=0}^{N_c} \Delta R_i$, is equal to R_{rup}.

The cell-specific anelastic attenuation is modeled by $\sum_{i=0}^{N_c} c_{cA\ i} \Delta R_i$ where $c_{cA\ i}$ is the cell attenuation of the i^{th} cell. In this GMM, c_{cA} is modeled so that, in areas away from past paths $c_{cA\ i}$ reverts to c_7 which is the anelastic attenuation coefficient in BA18, making the anelastic attenuation of the non-ergodic model equal to the anelastic attenuation of BA18 $(\sum_{i=0}^{N_c} c_{cA\ i} \Delta R_i = c_7 R_{rup})$; while in areas that are covered by the paths in NGAWest2 dataset, c_{cA} deviates from c_7 to capture the regional attenuation. Figure 2.6 shows the cells and the path coverage in the selected subset of the NGA-West2 dataset, as well as, the number of paths per cell.

Site Scaling

The functional form of the f_S the model is:

$$f_S = c_8 \ln \left(\frac{\min(V_{s30}, 1000)}{1000} \right) + f_{NL} + \delta c_{1b}(x_{site}) + \delta S2S \tag{2.5}$$

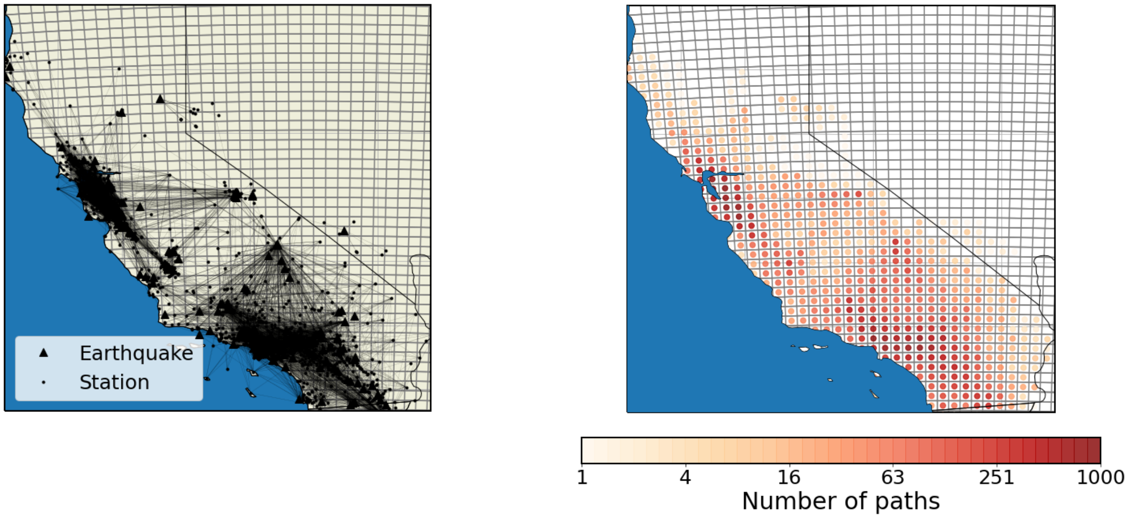


Figure 2.6: The left figure shows the path coverage for the cell-specific anelastic attenuation in the CA subset of NGAWest2. The right figure shows the number of paths per cell.

The ergodic components of the site term are: c_8 which controls the V_{S30} scaling of the ground motion, and f_{NL} which is the non-linear site amplification term. The non-ergodic effects related to the site are expressed by the $\delta S2S$ and δc_{1b} coefficients. The station constant, δc_{1b}, which has a finite correlation length, describes the broader adjustments to the backbone model to express the regional site effects. $\delta S2S$ has a zero correlation length and acts on top of δc_{1b} to describe the site specific adjustments. Coefficients with a finite correlation length vary continuously across the domain of interest, whereas coefficients with zero correlation length vary independently from location to location.

The remaining terms f_{ztor}, f_{NM} and f_{Z1} where kept as they are in the ergodic GMM.

2.3.2 Formulation of spatially varying coefficient model

The non-ergodic terms, cell-specific coefficients, and aleatory terms, hereafter collectively called model parameters ($\vec{\theta}$), were estimated by describing the GMM as a hierarchical Bayesian model using the computer software STAN (Stan Development Team, 2019). In Bayesian statistics, the posterior distribution of the parameters is proportional to the likelihood times the prior distribution of the parameters:

$$p(\vec{\theta}|y,x) \propto \mathcal{L}(\vec{\theta})p(\vec{\theta}) \tag{2.6}$$

The prior distributions are the distributions that the model parameters are assumed to follow in the absence of data; the likelihood function, in general terms, is the probability of observing the data given the model parameters; and the posterior distributions are the model-parameter distributions informed by the data.

The likelihood can be estimated from the density function of the ground motion:

$$\mathcal{L}(\vec{\theta}) = pdf(\vec{y}|f(x,\vec{\theta}), \phi_0^2 + \tau_0^2) \tag{2.7}$$

where $f(x,\vec{\theta})$ is the functional form for the median non-ergodic ground-motion:

$$\begin{aligned} f(x,\vec{\theta}) = &\left(f_{erg}(M, R_{rup}, V_{S30}, ...) - c_7 R_{rup}\right) \\ &+ \delta c_0 + \delta c_{0MS} + \delta c_{1a} + \delta c_{1b} + \delta S2S + \sum_{i}^{n_{cell}} c_{cA\ i}(x_{cell\ i}) \Delta R_i \end{aligned} \tag{2.8}$$

It is equal to the ergodic backbone model without the effect of anelastic attenuation $f_{erg}(M, R_{rup}, V_{S30}, ...) - c_7 R_{rup})$, plus the non-ergodic spatially varying constants that have been described the previous section (δc_i), and the cell-specific anelastic attenuation.

The model is called hierarchical because the prior distributions are defined in multiple levels. At the lower level, $\vec{\theta}$ follow some prior distributions, which are defined in terms of a different set of parameters, hereafter called hyper-parameters $\vec{\theta}_{hyp}$, which, in turn, either follow some other prior distributions, or they are fixed. In this study, the non-ergodic ergodic regression was performed in two phases: in the first phase, which included a smaller number of frequencies, $\vec{\theta}_{hyp}$ were defined by their own prior distributions that are described later in this section, whereas, in the second phase, most of $\vec{\theta}_{hyp}$ were fixed to their smoothed values, estimated from first phase, and the remaining hyperparameters were left free to follow the same prior distributions as in the first phase. The main reasons for fixing $\vec{\theta}_{hyp}$ in the second phase were to ensure that there are no abrupt changes in the non-ergodic terms between frequencies, to constrain the model to a more physical behavior, and to reduce the computational cost of the second phase which included more frequencies. Table 2.2 summarizes the parameters that were classified as $\vec{\theta}$ and $\vec{\theta}_{hyp}$; the parameters composing $\vec{\theta}$ have been defined in Section 2.3.1, the hyper-parameters composing $\vec{\theta}_{hyp}$ have defined later in this section. Table 2.3 summarizes the $\vec{\theta}_{hyp}$ that were free at each phase. If $\vec{\theta}_{hyp}$ is free, the prior distribution of a model parameter, θ_i, can be explicitly defined in terms of $\vec{\theta}_{hyp}$ as follows:

$$p(\theta_i) = f(\theta_i) d\theta_i = \left(\int f(\theta_i|\vec{\theta}_{hyp}) f(\vec{\theta}_{hyp})\ d\vec{\theta}_{hyp} \right)\ d\theta_i \tag{2.9}$$

More specifically, the δc_0 constant has a normal prior distribution with a zero mean and a 0.1 standard deviation:

$$\delta c_0 \sim \mathcal{N}(0, 0.1) \tag{2.10}$$

The mean is set to zero because in absence of data, there should be no shift between the ergodic and non-ergodic GMM. The standard deviation is small because any constant shift informed by the regional data or the re-weighting of the residuals is expected to be small.

Table 2.2: Summary of model parameters and hyper-parameters

Group Name	Group Notation	Components
Model parameters	$\vec{\theta}$	δc_0, $\delta c_{N/S}$, δc_{1a}, δc_{1b}, $\delta S2S$, c_{cA}, $\delta W^0_{e,s}$, δB^0_e
Model hyperparameters	$\vec{\theta}_{hyp}$	ρ_{1a}, ψ_{1a}, ψ_{1b}, ρ_{1b}, ϕ_{S2S}, ρ_{cA}, ψ_{cA}, σ_{cA}, ϕ_0, τ_0

Table 2.3: Free hyper-parameters at each regression phase

Phase	Free hyper-parameters
1	ρ_{1a}, ψ_{1a}, ψ_{1b}, ρ_{1b}, ϕ_{S2S}, ρ_{cA}, ψ_{cA}, σ_{cA}, ϕ_0, τ_0
2	ϕ_0, τ_0

For earthquakes with magnitudes less than M 5 and frequencies less than 5 hz, $\delta c_{0N/S}$ follow a normal prior distribution with a zero mean and a 0.2 standard deviation. Preliminary analyses, which did not include $\delta c_{0N/S}$, had a 0.2 to 0.4 regional difference in c_{1a} between northern and southern California, which have a 16 and 3% probability of being exceeded with the selected standard deviation. Therefore, the posterior distribution of $\delta c_{0N/S}$ will deviate from zero to reach a similar range only if there is significant support by the data; otherwise, $\delta c_{0N/S}$ will stay be close to zero implying no systematic difference between northern and southern CA at small magnitude events.

$$\delta c_{0NS} \sim \begin{cases} 0 & for \quad M > 5 \quad or \quad\;\; f > 5hz \\ \mathcal{N}(0, 0.2) & for \quad M < 5 \quad and \quad f < 5hz \end{cases} \tag{2.11}$$

The non-ergodic constants $\delta c_{1a}(x_{eq})$ and $\delta c_{1b}(x_{site})$ follow multivariate normal prior distributions with zero mean and Matern (negative exponential) covariance functions (κ). κ imposes the spatial correlation on δc_{1a} and δc_{1b}: it ensures that the values of δc_{1a} and δc_{1b} will be similar for earthquakes or sites in close proximity and that δc_{1a} and δc_{1b} would vary continuously from location to location.

$$\begin{aligned} \delta c_{1a} &\sim \mathcal{N}(0, \kappa_{1a}) \\ \delta c_{1b} &\sim \mathcal{N}(0, \kappa_{1b}) \end{aligned} \tag{2.12}$$

The covariance function between a pair of earthquakes for $\kappa(x_{eq}, x'_{eq})$ or between a pair of stations for $\kappa(x_{sta}, x'_{sta})$ is defined in equation (2.13); x and x' are the coordinates of the two earthquakes or sites depending on the coefficient, ψ_i is the standard deviation, and ρ_i is the correlation length. The ψ_i controls the variability of the non-ergodic coefficients, that is, how much the values of the coefficients could vary between locations that are far from each other. ρ governs the length scale of the spatial variation of δc_i; increasing ρ_i makes δc_i to vary more gradually with distance.

$$\begin{aligned} \kappa_{1a}(x, x') &= \psi_{1a}^2 e^{-\frac{\sqrt{x^2 - x'^2}}{\rho_{1a}}} \\ \kappa_{1b}(x, x') &= \psi_{1b}^2 e^{-\frac{\sqrt{x^2 - x'^2}}{\rho_{1b}}} \end{aligned} \tag{2.13}$$

Both ρ_{1a} and ρ_{1b} have inverse gamma prior distributions with distribution parameters α and β equal to 2 and 50 which corresponds to a mode and mean of 16.7 and 50 km, respectively.

$$\begin{aligned} \rho_{1a} &\sim InvGamma(2.0, 50.0) \\ \rho_{1b} &\sim InvGamma(2.0, 50.0) \end{aligned} \tag{2.14}$$

Inverse gamma distributions are defined only for positive real numbers which is a desirable property for the prior distributions of the correlation lengths as negative correlation lengths do not have any physical meaning. The ρ_{1a} and ρ_{1b} correlation lengths are expected to be around 10 to 50 km, where the inverse gamma distribution has most of the mass, but larger values are also possible, if they are supported by the data, due to its exponential tail.

The prior distribution of ψ_{1a} and ψ_{1b} is an exponential distribution with a rate of 20

$$\begin{aligned} \psi_{1a} &\sim exp(20) \\ \psi_{1a} &\sim exp(20) \end{aligned} \tag{2.15}$$

These prior distributions were chosen for ψ_{1a} and ψ_{1b} to penalize unnecessary model complexity (Simpson et al., 2017). A ψ_i equal to zero implies no variability for δc_i, meaning that there are no systematic effects related to that parameter. In an exponential distribution, most of the mass is near zero, which allows δc_i to deviate from zero to capture systematic effects related to that parameter only if there is significant support by the ground-motion observations. For the same reason, exponential prior distributions were used in Kuehn et al. (2020) to model the standard deviations of the regional terms in the KBCG20 partially non-ergodic subduction-zone GMM.

The site-specific adjustment, $\delta S2S$ follows a normal distribution with a zero mean and a ϕ_{S2S} standard deviation:

$$\delta S2S \sim \mathcal{N}(0, \phi_{S2S}) \tag{2.16}$$

$\delta S2S$ is a function of the site location, the same adjustment is applied to all ground motions recorded at the same station.

The prior distribution for ϕ_{S2S} is a log-normal distribution with a logmean of -0.8 and a standard deviation of 0.3 natural log units:

$$\phi_{S2S} \sim \mathcal{LN}(-0.8, 0.3) \tag{2.17}$$

This prior distribution has a median value of 0.45, and a 16^{th} and 84^{th} percentile of 0.33 and 0.6, respectively. Bayless and Abrahamson (2019b) found ϕ_{SS} to range from 0.4 to 0.6 which is consistent with the prior distribution for ϕ_{SS}.

The prior distribution of c_{cA} is a multivariate normal distribution with an upper truncation limit at zero:

$$c_{cA} \sim \mathcal{N}(\mu_{cA}, \kappa_{cA})\mathcal{T}(,0) \tag{2.18}$$

where μ_{cA} is the mean of the distribution, and κ_{cA} is the covariance function. To ensure the physical extrapolation of the GMM, c_{cA} is limited to be less or equal to zero, which is satisfied by setting the upper limit of the normal distribution at zero ($T(,0)$). Two key differences from the Abrahamson et al. (2019) approach when modeling the cells are the different mean and the different covariance function of the prior distribution. In this model, the mean of the prior is equal to the value of the anelastic-attenuation coefficient in Bayless and Abrahamson (2019b) ($\mu_{cA} = c_{7\ BA18}$); thus, in areas with sparse data, the non-ergodic attenuation goes back to the ergodic attenuation to ensure reasonable extrapolation at large distances. This decision was made because the local data may not be sufficient to estimate both the median shift and the spatial variability of the anelastic attenuation. The covariance function (equation (2.19)) is the sum of a Matern kernel scaled by ψ_{cA}^2 and a diagonal kernel scaled by σ_{cA}^2. The Matern kernel controls the underlining continuous variation of anelastic attenuation over large areas, whereas, the diagonal kernel allows for some independence in the attenuation from cell to cell. ψ_{cA} controls the size of the underling variability of c_{cA} over large distances, the correlation length ρ_{cA} controls how fast the underling component of c_{cA} varies with distance, and σ_{cA} controls the size of the independent variability.

$$\kappa_{cA}(x_i, x_j) = \psi_{cA}^2 e^{-\frac{\sqrt{x_i^2 - x_j'^2}}{\rho_{cA}}} + \sigma_{cA}^2 \delta(i-j) \tag{2.19}$$

ρ_{cA} has an inverse gamma distribution with the same parameters as the prior distributions for ρ_{1a} and ρ_{1b}. ψ_{cA} and σ_{cA} have an exponential prior distribution with the same parameters as the prior distributions for θ_{1a} and ψ_{1b}.

$$\begin{aligned} \rho_{cA} &\sim InvGamma(2.0, 50.0) \\ \psi_{cA} &\sim exp(20) \\ \sigma_{cA} &\sim exp(20) \end{aligned} \tag{2.20}$$

The non-ergodic within-event residuals, $\delta W_{e,s}^0$, follow a normal distribution with a zero mean and ϕ_0 standard deviation.

$$\delta W_{e,s}^0 \sim \mathcal{N}(0, \phi_0^2) \tag{2.21}$$

The prior distribution of ϕ_0 is a log-normal distribution with a logmean of -1.3 and a standard deviation of 0.3 natural log units (equation (2.22)). This set of parameters leads to a prior mean of 0.27 and a 16^{th} to 84^{th} percentile range of 0.2 to 0.37. BA18 is an ergodic model, and so an estimate of ϕ_0 is not available to inform the prior distribution of ϕ_0 of this model. However, ϕ_{SS}, which is available in BA18, is about 0.4 for most frequencies, and because ϕ_0 is smaller than ϕ_{SS} by definition (Al Atik et al., 2010), the range of the prior distribution is reasonable.

$$\phi_0 \sim \mathcal{LN}(-1.3, 0.3) \tag{2.22}$$

The non-ergodic between-event residuals, δB_e^0, follow a normal distribution:

$$\delta B_e^0 \sim \mathcal{N}(0, \tau_0) \tag{2.23}$$

with a zero mean and τ_0 standard deviation. δB_e^0 is a function of the earthquake id, e; that is, the same δB_e^0 is applied to all recordings of the same earthquake.

The prior distribution of τ_0 is a log-normal distribution with a -1.0 logmean and 0.3 log-standard deviation.

$$\tau_0 \sim \mathcal{LN}(-1.0, 0.3) \tag{2.24}$$

The τ_0 distribution parameters are judged to be reasonable because the mean and 16^{th} and 84^{th} percentiles (0.38, 0.27 and 0.50) are in agreement with other non-ergodic studies where the total non-ergodic standard deviation ($\sqrt{\phi_0^2 + \tau_0^2}$) ranges from 0.36 to 0.55.

2.3.3 Predictive distributions of coefficients at new locations

The non-ergodic coefficients can be estimated at new locations ($\vec{x}^*$) by conditioning them on the non-ergodic coefficients at the existing locations ($\vec{x}$); that is, the location of stations or past events depending on the coefficient. Since all non-ergodic coefficients follow multivariate normal distributions, for known values of the non-ergodic coefficients ($\vec{\delta c_i}$) at the $\vec{x}$, the non-ergodic coefficients at $\vec{x}^*$ also follow multivariate normal distributions with the mean and covariance matrix (Rasmussen and Williams, 2006; Landwehr et al., 2016):

$$\vec{\mu}_{\delta c_i^* | \delta c_i} = \mathbf{k}_i^\intercal \mathbf{K}_i^{-1} \vec{\delta c_i} \tag{2.25}$$

$$\mathbf{\Psi}_{\delta c_i^* | \delta c_i} = \mathbf{K^*}_i - \mathbf{k}_i^\intercal \mathbf{K}_i^{-1} \mathbf{k}_i \tag{2.26}$$

where $\vec{\mu}_{\delta c_i^* | \delta c_i}$ mean of non-ergodic ergodic coefficients at $\vec{x}'$ conditioned on $\vec{\delta c_i}$, $\mathbf{\Psi}_{\delta c_i^* | \delta c_i}$ is the covariance of non-ergodic coefficients at $\vec{x}'$ conditioned on $\vec{\delta c_i}$, $\mathbf{K}$ is the covariance between the non-ergodic coefficients at the existing locations ($\mathbf{K}_i = \kappa_i(\vec{x}, \vec{x})$), $\mathbf{k}$ is the covariance between the non-ergodic coefficients at the existing and new locations ($\mathbf{k}_i = \kappa_i(\vec{x}, \vec{x}^*)$), and $\mathbf{K}^*$ is the covariance between the non ergodic coefficients at the new locations ($\mathbf{K^*}_i = \kappa_i(\vec{x}^*, \vec{x}^*)$).

However, the non-ergodic coefficients at $\vec{x}$ are not known. There is some epistemic uncertainty associated with $\vec{\delta c_i}$ which is quantified by their posterior distribution. To simplify the calculations and obtain a closed-form solution, $\vec{\delta c_i}$ is assumed to follow a multivariate normal posterior distribution ($\vec{\delta c_i} \sim \mathcal{N}(\vec{\mu}_{\delta c_i}, \mathbf{\Psi}_{\delta c_i})$); $\vec{\mu}_{\delta c_i}$ is the posterior mean, and $\mathbf{\Psi}_{\delta c_i}$ is a diagonal matrix with the posterior variances across the diagonal. This assumption is deemed reasonable as all $\vec{\delta c_i}$ have multivariate normal prior distributions. If all hyper-parameters were fixed, the posterior distributions of $\vec{\delta c_i}$ would indeed be multivariate normal distributions, but because some of the hyper-parameters are free, the posterior distributions of $\vec{\delta c_i}$ may slightly deviate from the assumption. The epistemic uncertainty of δc_i can be accounted in predicting δc_i^* by using the marginal distribution of δc_i^*:

$$p(\delta c_i^*) = \int p(\delta c_i^* | \delta c_i) p(\delta c_i) \; d\delta c_i \tag{2.27}$$

Due to the previous assumption, $p(\delta c_i^*)$ is also a multivariate normal distribution with mean and covariance matrix given in Equations (2.28) and (2.29), respectively (Bishop, 2006).

$$\vec{\mu}_{\delta c_i^*} = \mathbf{k}_i^\intercal \mathbf{K}_i^{-1} \vec{\mu}_{\delta c_i} \tag{2.28}$$

$$\mathbf{\Psi}_{\delta c_i^*} = \mathbf{K^*}_i - \mathbf{k}_i^\intercal \mathbf{K}_i^{-1} \mathbf{k}_i + \mathbf{k}_i^\intercal \mathbf{K}_i^{-1} \mathbf{\Psi}_{\delta c_i} (\mathbf{k}_i^\intercal \mathbf{K}_i^{-1})^\intercal \tag{2.29}$$

2.3.4 Inter-frequency Correlation

The main motivation behind the development of this non-ergodic EAS GMM is to use it with RVT to create an equivalent non-ergodic PSa GMM; in doing this, it is important to capture the inter-frequency correlation of the non-ergodic terms, otherwise, as it was demonstrated by Bayless and Abrahamson (2018), the variability of the PSa values is underestimated. The correlation coefficient (ρ) is a measure of the linear relationship of two random variables X_1 and X_2. A ρ that is equal to one implies that X_2 can be perfectly defined as a linear function of X_1, and vice versa; a zero ρ implies that the two random variables are linearly independent. In ground-motion studies, the inter-frequency correlation coefficient is a measure of the width of the peaks and troughs of a PSa or EAS spectrum: the stronger the correlation of the amplitudes between frequencies, the wider the peaks and troughs of the spectra will be.

The correlation coefficient for a non-ergodic term δc_i, at frequencies f_1 and f_2, is defined as:

$$\rho_{\delta c_i}(f_1, f_2) = \frac{cov\left(\delta c_i(f_1), \delta c_i(f_2)\right)}{\sigma_{\delta c_i(f_1)} \sigma_{\delta c_i(f_2)}} \tag{2.30}$$

where cov is the covariance of δc_i at the two frequencies, and σ_i is the standard deviation of δc_i. ρ can be determined from the data using the maximum likelihood estimator (Kutner et al., 2005):

$$\rho_{\delta c_i}(f_1, f_2) = \frac{\sum_{j=1}^{n} \left(\delta c_{i,j}(f_1) - \overline{\delta c_i}(f_1)\right) \left(\delta c_{i,j}(f_2) - \overline{\delta c_i}(f_2)\right)}{\sqrt{\sum_{j=1}^{n} \left(\delta c_{i,j}(f_1) - \overline{\delta c_i}(f_1)\right)^2} \sqrt{\sum_{j=1}^{n} \left(\delta c_{i,j}(f_2) - \overline{\delta c_i}(f_2)\right)^2}} \tag{2.31}$$

where n is the number of observations, $\delta c_{i,j}$ is the j^{th} sample of δc_i, and $\overline{\delta c_i}$ is the mean value of δc_i. For a large number of samples ($n > 25$), ρ can be transformed into a random variable z that follows a normal distribution with equation (2.32) (Kutner et al., 2005); the standard deviation of z is given in equation (2.33).

$$z = \tanh^{-1}(\rho) = \frac{1}{2}\ \ln\left(\frac{1+\rho}{1-\rho}\right) \tag{2.32}$$

$$\sigma(z) = \sqrt{\frac{1}{n-3}} \tag{2.33}$$

The same functional form that was used to model the correlation of the total EAS residuals in Bayless and Abrahamson (2019a) was used here to fit the empirical correlations of the non ergodic terms:

$$\rho(f_r) = \left\{ \begin{array}{cc} 1 & for \quad f_r = 0 \\ \tanh\left(Ae^{Bf_r} + Ce^{Df_r}\right) & for \quad f_r \neq 0 \end{array} \right. \tag{2.34}$$

$$f_r = \left| \ln\left(\frac{f_1}{f_2}\right) \right| \tag{2.35}$$

A, B, C, D are the model parameters, and f_r is the absolute value of the natural log of the ratio of the two frequencies. This functional form allows for a two-term exponential decay as a function of the logarithm of frequency; this behavior is required because both the correlation of the total residuals in Bayless and Abrahamson (2019a) and the correlation of the epistemic uncertainty terms presented in section 2.4 exhibit a steep decay at frequencies near the conditioning frequency which then flattens at frequencies that are further from the conditioning frequency. The model parameters were estimated with a non-linear least-squares regression on z using the MINPACK.LM package (Elzhov et al., 2016) in the statistical software R (R Core Team, 2020); $\sigma(z)$ was used as weights in the least-squares regression emphasising the fit to the higher correlation values which have more samples.

In this study, one difference from the Bayless and Abrahamson (2019a) is it that the inter-frequency correlation of all epistemic uncertainty terms was modeled as frequency independent (i.e. A, B, C, D are constants). This was done, because at it can be seen in section 2.4, δc_{1b} and $\delta S2S$, which are the biggest contributors to the total non-ergodic effects, have an almost frequency independent inter-frequency correlation. δc_{cA} show the most noticeable frequency dependence in inter-frequency correlation, but it becomes more stable at intermediate and large frequencies which is the frequency range where it has the biggest impact. The assumption of frequency independence should be re-examined in future studies with a larger dataset.

2.4 Results

2.4.1 Hyperparameters

Figure 2.7 presents the mean, 5^{th} and 95^{th} percentiles of the posterior distribution of the hyper-parameters of the non-ergodic terms; the proposed smoothed values are also presented in the same figure. As mentioned in section 2.3.2, the regression for this model was performed in two phases; in the first phase, all model hyperparameter were free and estimated based on the data and prior distributions, whereas in the second phase, the hyper-parameters of the non-ergodic terms were fixed to their smoothed values, and τ_0 and ϕ_0 were reestimated for the new set of smoothed hyper-parameters.

Figure 2.7a shows the variation of the correlation length for δc_{1a} with frequency. The posterior distribution of ρ_{1a} is wide due to the small number of unique earthquakes. Overall, the mean estimate of ρ_{1a} is around $40km$ except at low frequencies where at $0.3hz$ it jumps up to $85km$. Because there is no physical reason for ρ_{1a} to increase at low frequencies, it was fixed to the average ρ_{1a} over all frequencies. Furthermore, there are less data at low and high frequencies making the estimates of the hyperparameters at these frequency ranges less stable.

Figure 2.7b shows how ψ_{1a} changes with frequency. In this case, the posterior mean of ψ_{1a} stays constant at low and intermediate frequencies, and it exhibits an increase at high frequencies. Similarly to ρ_{1a}, there is no physical reason for ψ_{1a} to increase at high frequencies so it was fixed to the average value over all frequencies. One possible cause of the apparent increase of ψ_{1a} at high frequencies is that some of the non-ergodic site effects could have been mapped into non-ergodic source effects in the regression. It is expected that the non-ergodic site effects will increase at high frequencies because the regional differences in site amplification tend to have a larger impact on the high frequencies. This assumption is consistent with the behavior of ψ_{1b} and ϕ_{SS}, which both show an increase with frequency. For this reason, the difference between the estimated and smoothed ψ_{1a} was moved to ψ_{1b}.

In smoothing ψ_{1b}, up to the frequency of $15hz$, a piece-wise linear model was fit to the estimated mean values; whereas beyond $15hz$, it was fit to the square root of the sum of squares of the estimated mean ψ_{1b} and the difference between the estimated mean and smoothed ψ_{1a}. Minimal smoothing was applied to ρ_{1b} and ϕ_{SS} as they show a relatively small variation between neighbouring frequencies.

The smoothed ρ_{cA} was fixed to the average of the mean estimates that are less than $75km$; this upper limit was imposed because the ρ_{cA} with large mean estimates also had wide posterior distributions, meaning that ρ_{cA} could not be reliably estimated at those frequencies.

ψ_{cA} and σ_{cA} exhibit similar characteristics in their variation with frequency: they are very small at low frequencies, they show an approximately linear increase with the log of frequency at intermediate frequencies, and they reach a plateau at high frequencies. This happens because the effects of anelastic attenuation are more noticeable at high frequencies, and likewise, spatial changes in the anelastic attenuation will have a larger effect on higher frequencies. This behavior was also observed by Kuehn et al. (2019) who found that the standard deviation of the cell-specific attenuation coefficients of their non-ergodic PSa GMM was smaller at long periods .

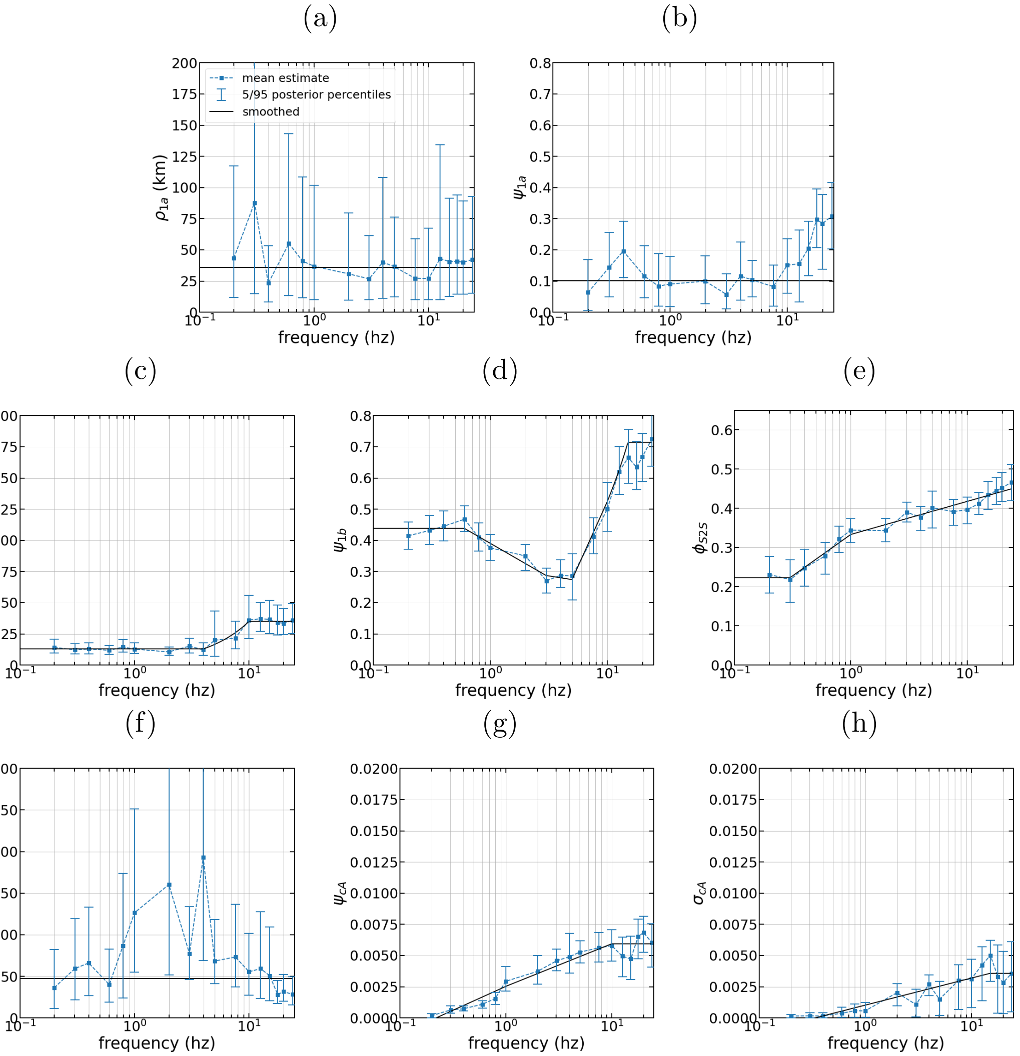
Figure 2.8 presents the hyperparamters that were reestimated in the second step. For most frequencies, ϕ_0 is about 0.47, and τ_0 ranges from 0.35 to 0.45. Based on this range, the total aleatory standard deviation of the non-ergodic GMM ranges from 0.57 to 0.65 which is about a 40 to 30% reduction from the standard deviation of the ergodic GMM.

Figure 2.9 shows the variation of δc_0 and $\delta c_{0N/S}$ with frequency. The coefficient δc_0 corresponds to the shift of the non-ergodic GMM due to the reweighting of the residuals. For most frequencies, the change in the constant from the ergodic model is less than 10%. The regional constant $\delta c_{0N/S}$, which corrects for the bias due to the differences in the small magnitude conversion, is about 0.4 for the northern California and zero for the southern California. In the NGAWest2 dataset, most of the data are located in southern California, so it is expected that the base model would be consistent with the southern part of the state with the main correction applied to the northern part of California.

2.4.2 Spatially varying coefficients and cell-specific anelastic attenuation

Figure 2.10 shows the spatial distribution of the mean estimate and epistemic uncertainty of δc_{1a}, δc_{1b}, and $\delta S2S$ for $f = 5hz$. As mentioned in Section 2.3, the δc_{1a} varies as a function of the source coordinates, whereas δc_{1b} and $\delta S2S$ as a function the site coordinates. In areas with past earthquakes, the mean estimate of δc_{1a} deviates from zero and its epistemic uncertainty is small. δc_{1a} is positive if the earthquakes in a region have systematically above average source effects and negative if the source effects are below the average. In areas with sparse or no data, the systematic effects related to the source cannot be reliably estimated, thus, δc_{1a} approaches zero and its epistemic uncertainty is large. The same behavior is observed in the spatial distribution of δc_{1b}: in large metropolitan areas, were most of the station are located, the δc_{1b} mean estimate deviates from zero, and its epistemic uncertainty is small; in remote areas, the δc_{1b} mean estimate approaches zero and, its epistemic uncertainty is large. $\delta S2S$ is only plotted at the station locations as it has a zero correlation length, meaning that as we move away from a station it will directly go to zero. The mean estimates of $\delta S2S$ do not exhibit any spatial correlation (i.e. there are no regions where $\delta S2S$ are systematically positive or negative) meaning that spatially correlated component of the site effects was properly captured by δc_{1b}

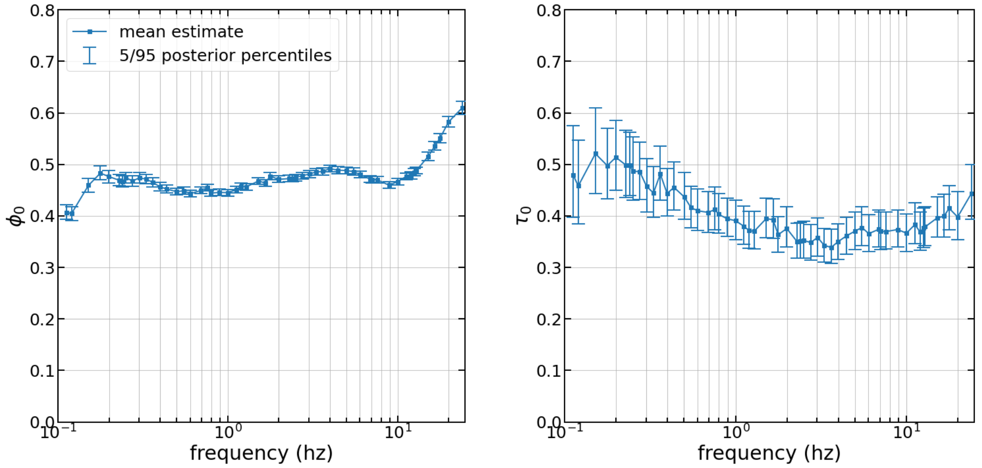
Figure 2.11 illustrates the spatial distribution of the cell-specific anelastic attenuation. The mean of c_{cA} deviates from $c_{7\ BA18}$ in cells that are crossed by many paths, whereas it stays close to $c_{7\ BA18}$ in cells crossed by few or zero paths. In addition, cells that are crossed by few paths have large epistemic uncertainty in c_{cA}. Overall, the epistemic uncertainty is low in Bay Area and Los Angeles and high in the northern part of California and the state of Nevada. The main features that stand out in Figure 2.11a are the higher than average anelastic attenuation north of the San Francisco Bay Area and east of San Diego, and the


(a) (b)

(c) (d) (e)

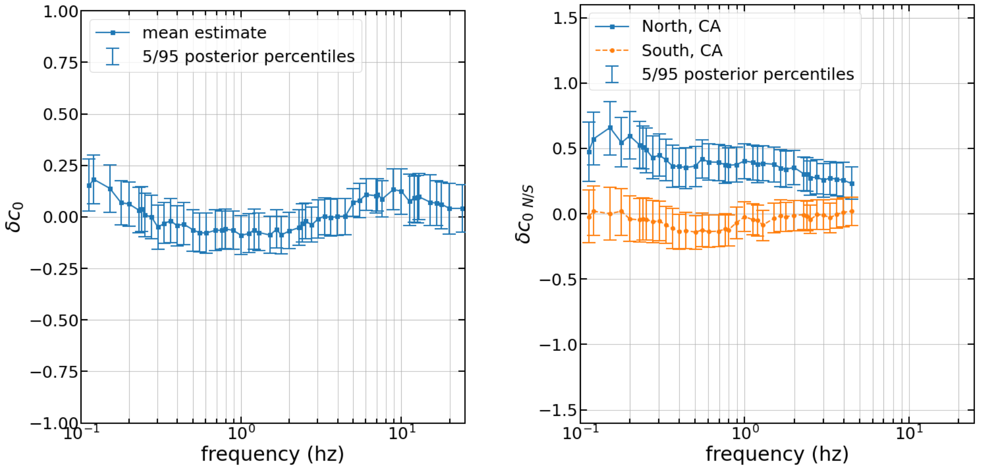
(f) (g) (h)

Figure 2.7: Estimated and smoothed hyperparameters versus frequency; the circular line depicts the mean estimate of the hyperparameters from the original regression, the vertical bars correspond to the 5/95 percentiles of the posterior distribution, and the solid line represent to the smoothed hyperparameters. (a) correlation length of the source constant, ρ_{1a}, (b) standard deviation of source constant, ψ_{1a}, (c) correlation length of the site constant with finite correlation length, ρ_{1b}, (d) standard deviation of the site constant with finite correlation length, ψ_{1b}, (e) standard deviation of the site term with zero correlation length, ϕ_{S2S}, (f) correlation length of the cell-specific anelastic attenuation, ρ_{cA}, (g) standard deviation of the correlated component of the cell-specific anelastic attenuation, ψ_{cA}, and (h) standard deviation of the correlated component of the cell-specific anelastic attenuation, σ_{cA}.



(a) (b)

Figure 2.8: Estimated hyperparameters of final regression versus frequency; the circular marker corresponds to the mean estimate, the vertical bars represents the 5/95 percentiles of the posterior distribution (a) within-event aleatory standard deviation, ϕ_0, and (b) between-event aleatory standard deviation τ_0.



(a) (b)

Figure 2.9: Estimated δc_0 and $\delta c_{0\ N/S}$ versus frequency; the circular marker corresponds to the mean estimate, the vertical bars represents the 5/95 percentiles of the posterior distribution (a) constant shift, δc_0, (b) regional shift, $\delta c_{0\ N/S}$, the solid line with the square markers corresponds to the Northern CA, the dashed line with the circular markers corresponds to the Southern CA

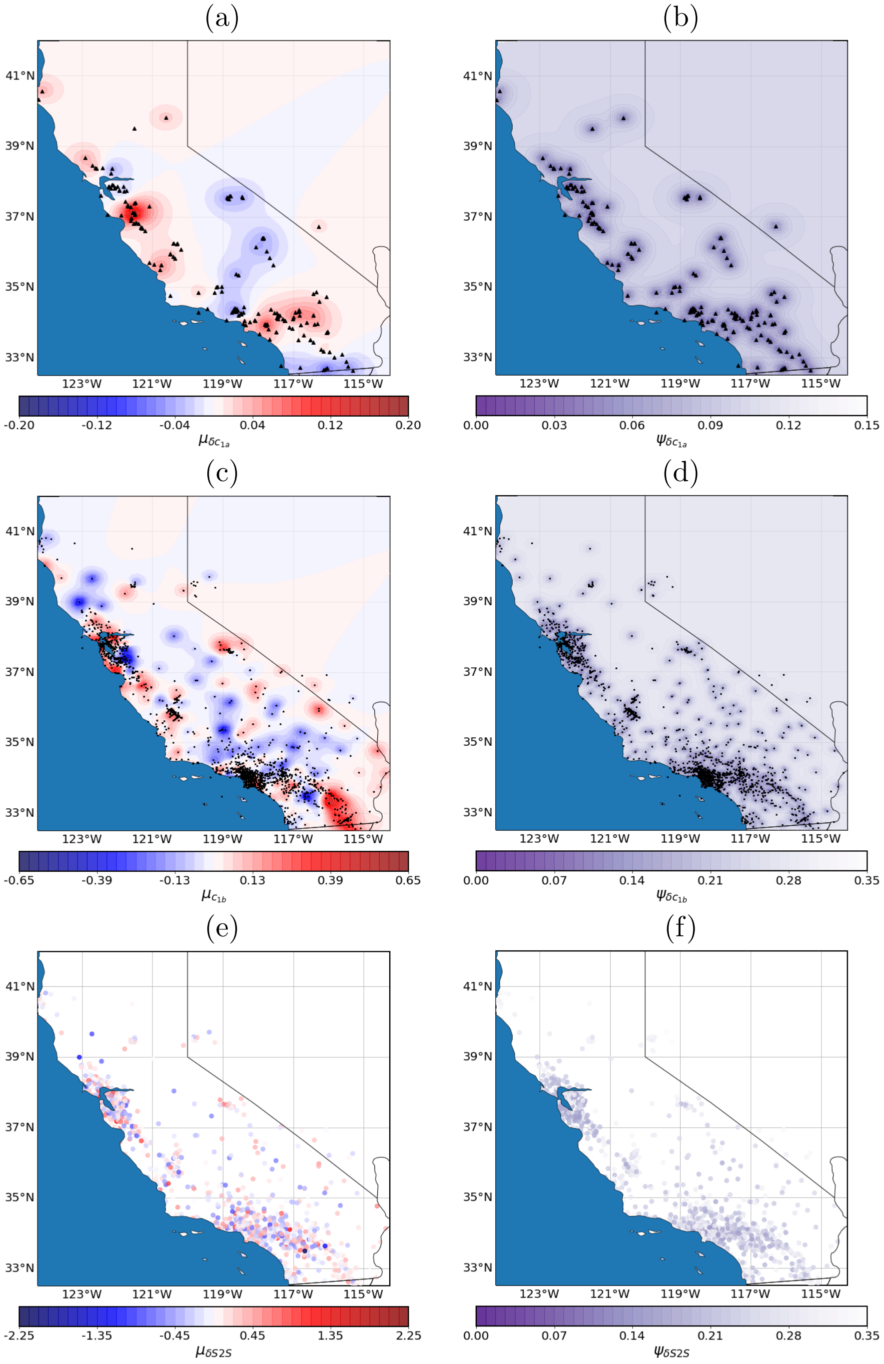


Figure 2.10: Spatial distribution of source and site constants at $f = 5hz$. Triangle markers show the location of earthquakes, dots show the location of stations. (a) mean estimate of δc_{1a}, (b) epistemic uncertainty of δc_{1a}, (c) mean estimate of δc_{1b}, (d) epistemic uncertainty of δc_{1b}, (e) mean estimate of $\delta S2S$, and (f) epistemic uncertainty of $\delta S2S$.

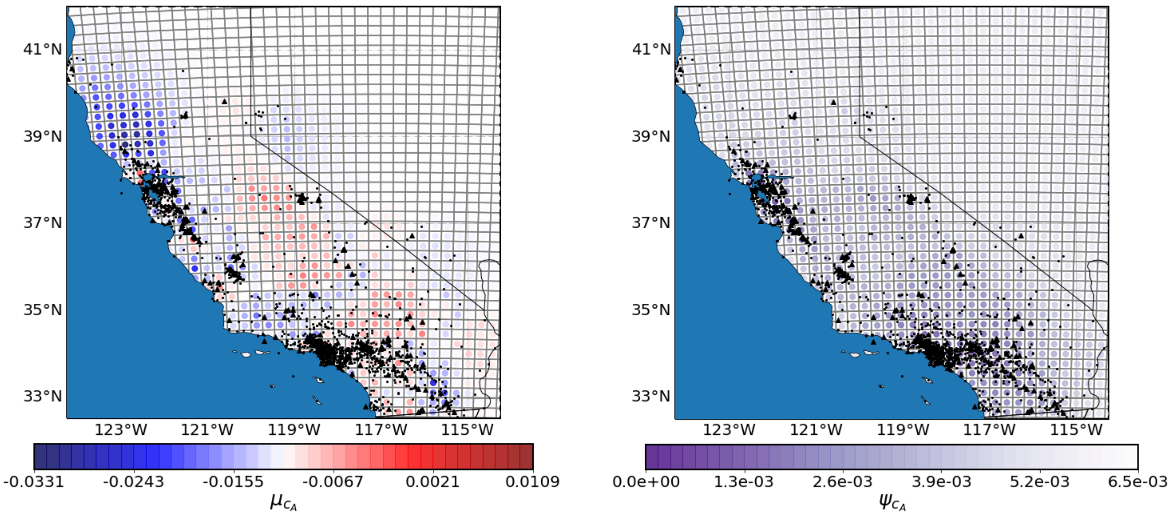

(a) (b)

Figure 2.11: Spatial distribution of cell specific anelastic attenuation at $f = 5hz$. (a) mean estimate of c_{cA}, and (b) epistemic uncertainty of c_{cA}.

less than average anelastic attenuation in the Central Valley and east of Los Angeles.

These findings are consistent with published attenuation models shown in Figure 2.12; Figure 2.12a corresponds to the Eberhart-Phillips (2016) Q model for frequencies 6 to 12 hz, and 2.12a corresponds to the Phillips et al. (2014) Q model for the S-waves at $4km$ depth. The quality factor, Q, is inversely proportional to the anelastic attenuation: high Q means low anelastic attenuation, and vice versa. Both models show small Q values north of Bay Area, and large Q values in Central Valley; additionally, Eberhart-Phillips (2016), which covers the entire state of California and Nevada, shows small values of Q east of San Diego and large values of Q east of Los Angeles. The mean value of c_{cA} and the Q model of Eberhart-Phillips (2016) differ in Nevada because there are no paths that cover that region. The large epistemic uncertainty of c_{cA} in Nevada means that the cell-specific anelastic attenuation cannot be estimated in that region with the current data set. This comparison shows that the cell specific anelastic attenuation has a physical basis.

Figure 2.13 shows the epistemic uncertainty of the non-ergodic terms as a function of the number of records for δc_{1a}, δc_{1b} and $\delta S2S$, and as a function of the number of paths for c_{cA}. The epistemic uncertainty of δc_{1a} and δc_{1b} is not sensitive to the number of records, whereas the epistemic uncertainty of $\delta S2S$ and c_{cA} decreases as the number of records and number of paths increases. This happens because $\delta S2S$ is spatially uncorrelated, and c_{cA} has a spatially uncorrelated component; $\delta S2S$ can be estimated more accurately as the number of ground motions recorded at a station increases, and c_{cA} can be estimated more accurately as the number of paths crossing a cell increases. δc_{1a}, δc_{1b} are spatially correlated and so the location of an event or a station is also important. That is, δc_{1b} can have less epistemic

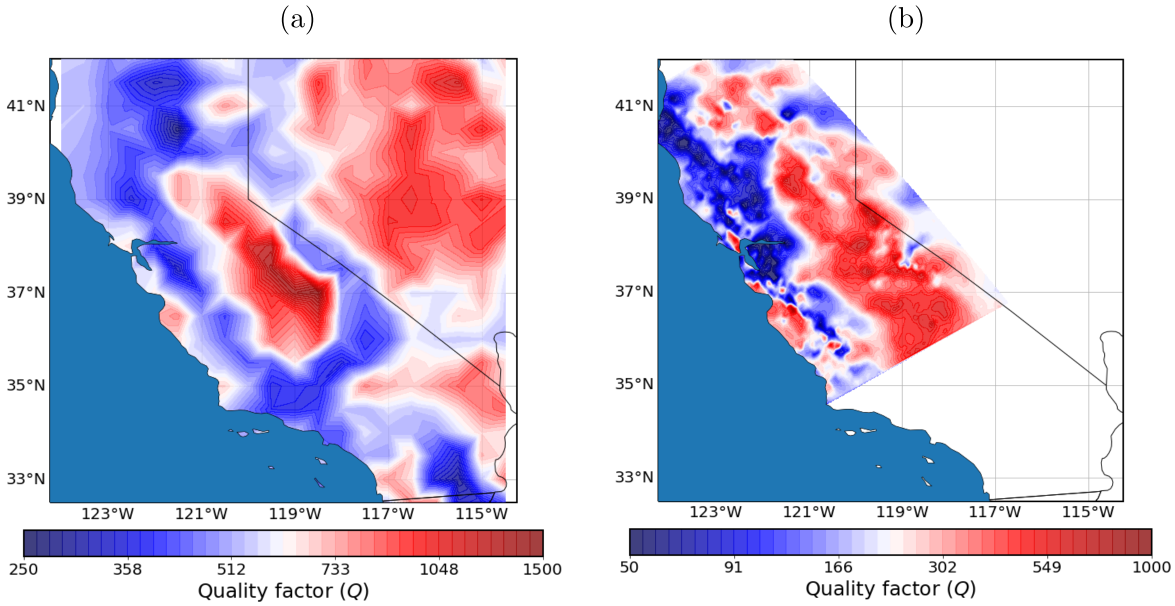


Figure 2.12: Seismic attenuation models for California from seismic inversions. (a) Phillips et al. (2014) attenuation model for frequencies between 6 and 12 hz, and (b) Eberhart-Phillips (2016) S-wave attenuation model for northern California at depth of $4km$.

uncertainty near a group of stations, with few records at each station, than near a remote station with a large number of records, if collectively, the group of stations has more data to constrain δc_{1b}. The same holds true for δc_{1a} regarding the spatial distribution of events.

2.4.3 Non-ergodic residuals

The residuals of the non-ergodic model at $f = 5hz$ are presented in Figure 2.14: the dots represent the residuals, the solid line corresponds to the moving average, and the error bars correspond to the standard deviation. δB_e shows no trend and an approximately constant standard deviation with magnitude; δW_{es} also shows no trend, but the standard deviation standard deviation reduces with magnitude. Additionally, δW_{es} shows no trend and a constant standard deviation with R_{rup} and V_{S30}

2.4.4 Standard deviation

In the model development, for simplicity, the aleatory standard deviations, τ_0 and ϕ_0, were modeled as magnitude independent. Any magnitude dependence of τ_0 and ϕ_0 was determined in post processing based on the non-ergodic residuals. τ_0 was modeled as constant, as, for the most part, the δB_e^0 residuals did not exhibit any reduction in standard deviation with magnitude (Figure 2.14). Because the number of events greater than M 6.5 is small, the model

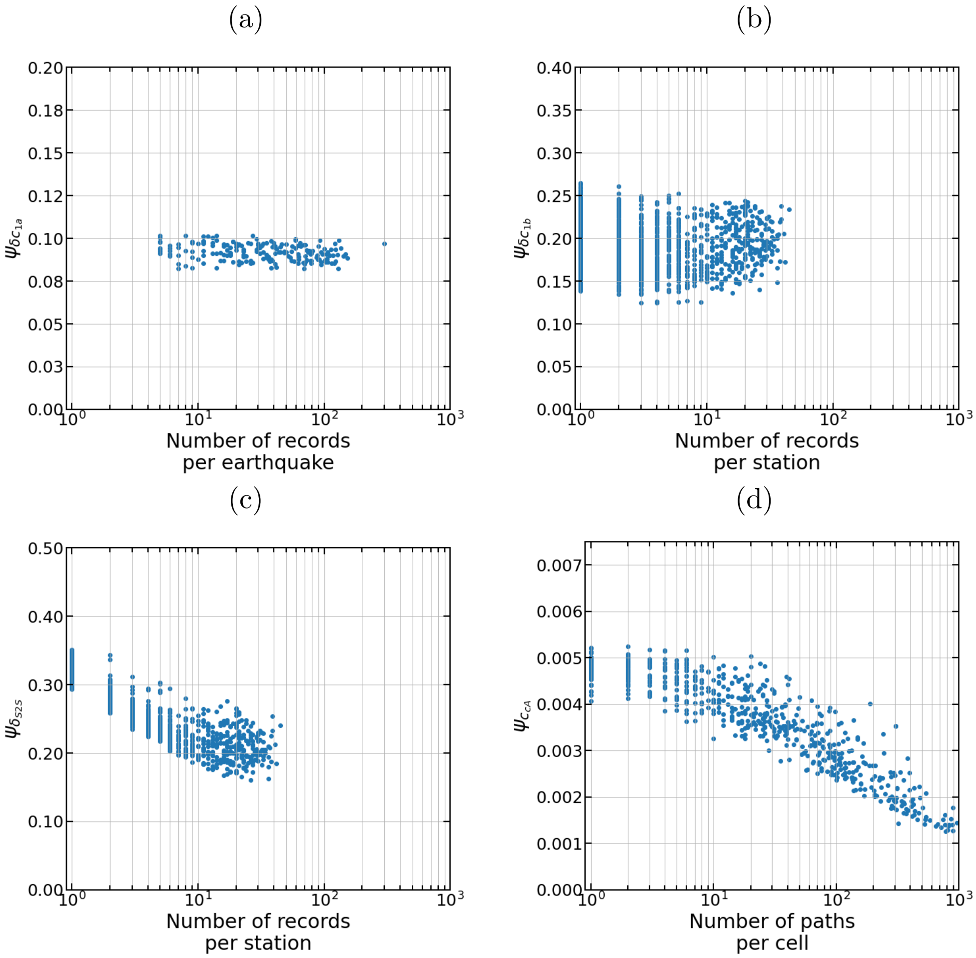


Figure 2.13: Standard deviations of posterior distributions of non-ergodic terms; (a) δc_{1a}, (b) δc_{1b} (c) $\delta S2S$, and (d) c_{cA}.

for τ_0 did not follow the reduction of the empirical standard deviation at large magnitudes, but instead it followed the standard deviation of the small events. ϕ_0 was modeled as a piece-wise linear function (equation (2.36)), as δW^0_{es} residuals exhibit some reduction in the standard deviation with increasing magnitude: ϕ_{0M_1} is the within-event standard deviation for magnitudes less than 5, and ϕ_{0M_2} is the within-event standard deviation for magnitudes greater than 6.5.

The aleatory parameters (τ_0, ϕ_{0M_1}, and ϕ_{0M_2}) were smoothed in order to ensure that the resulting EAS will have a reasonable shape (Figures 2.15 and 2.16). The smoothing was performed by fitting the aleatory parameters with a fourth order polynomial. The value of τ_0 decreases from small to intermediate frequencies and increases again after $f = 3hz$, which is consistent with the behaviour of BA18 and other PSa GMM, such as Abrahamson et al. (2014). The magnitude dependence of ϕ_0 is more pronounced at high frequencies.

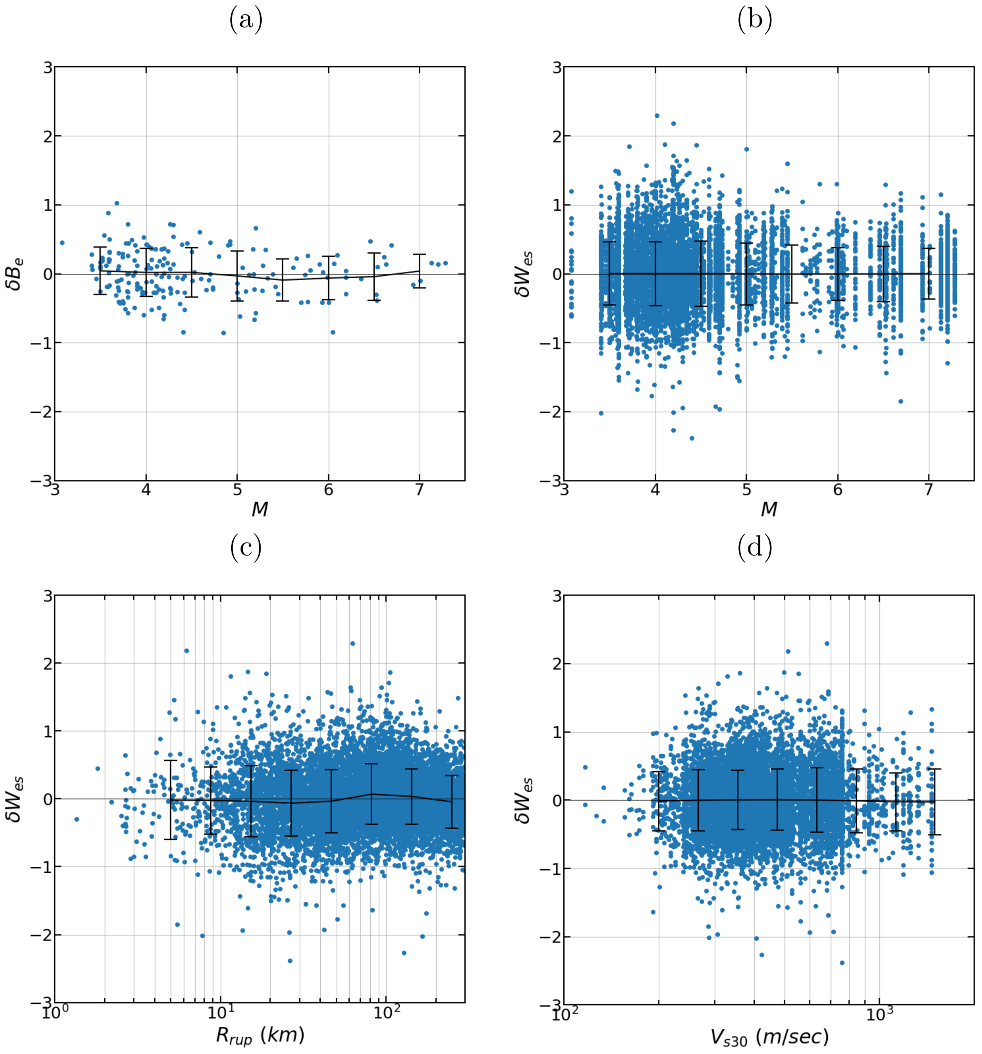


Figure 2.14: Non-ergodic within-event and between-event residuals at $f = 5hz$. (a) δB_e versus magnitude, (b) δW_{es} versus magnitude, (c) δW_{es} versus rupture distance, and (d) δW_{es} versus V_{S30}.

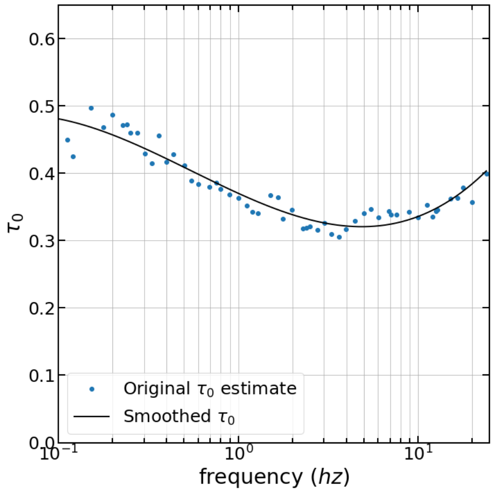


Figure 2.15: Between-event standard deviation, τ_0, versus frequency; circular markers correspond to the estimated τ_0 at every frequency, solid line corresponds to smoothed τ_0.

The higher ϕ_0 of small events at high frequencies is believed to be due to an increased effect of the radiation patterns. At large events, the effect of radiation patterns is smaller as seismic rays originate from more locations along the fault, which increases the range of azimuthal angles, and leads to destructive interference of the radiation patterns resulting in less ground-motion variability. Figure 2.17 compares the magnitude relationships for τ_0 and ϕ_0 with the empirical standard deviations at $f = 5hz$. Overall, there is a good fit between the τ_0 and ϕ_0 relationships and the standard deviations of the binned residuals. The differences at large magnitudes should be reevaluated with a dataset which includes a grater number of large magnitude events.

$$\phi_0 = \begin{cases} \phi_{0M_1} & for \quad M < 5 \\ \phi_{0M_1} + (\phi_{0M_2} - \phi_{0M_2})(M-5)/(6.5-5) & for \quad 5 < M < 6.5 \\ \phi_{0M_2} & for \quad M > 6.5 \end{cases} \tag{2.36}$$

2.4.5 Regional source term $\delta c_{0N/S}$

Figure 2.18 shows the earthquake constant, δc_{1a}, from a preliminary analysis which did not include $\delta c_{0N/S}$, for different frequencies. At low frequencies, $f = 0.2$ and $1.0hz$, the preliminary δc_{1a} shows a significant regional difference between northern and southern CA, whereas at higher frequencies, $f = 5.0$ and $10.0hz$, such regional difference is not observed. These trends were also observed at other frequencies: at frequencies less than $5hz$, the spatial distribution of the preliminary δc_{1a} exhibited a regional difference between northern and southern CA, while at frequencies greater than $f = 5hz$, any spatial variability of δc_{1a} was more localized.

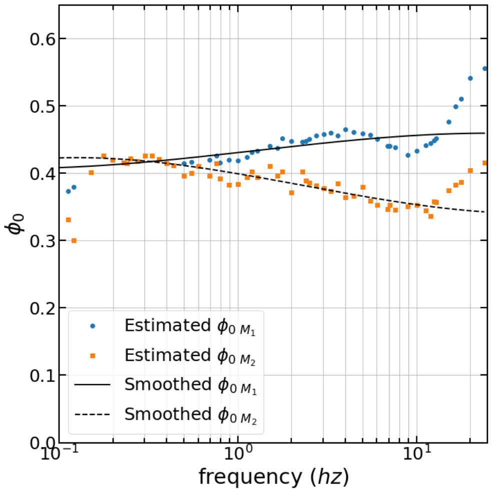
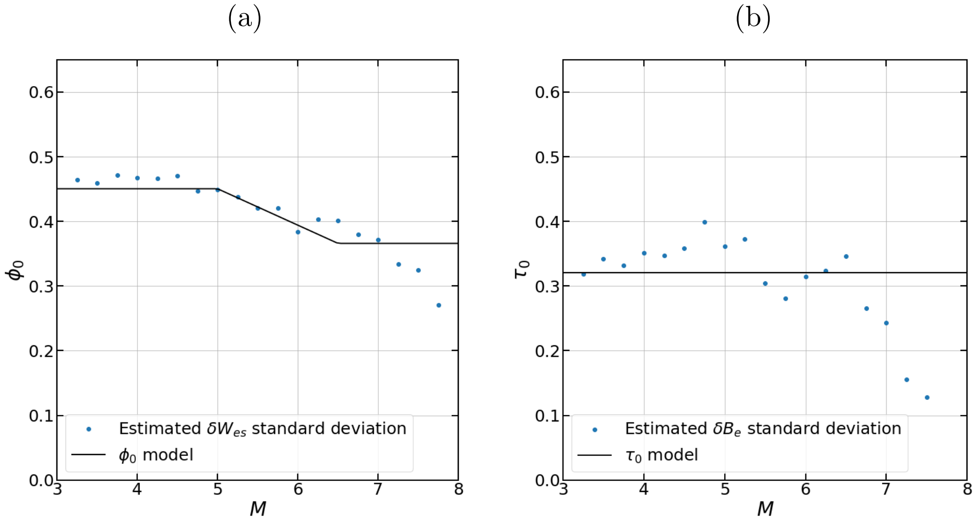


Figure 2.16: Within-event standard deviation, ϕ_0, versus frequency; circular markers correspond to the estimated ϕ_0 at small magnitudes, square markers correspond to the estimated ϕ_0 at large magnitudes, solid line corresponds to smoothed ϕ_0 for small magnitudes, dashed line corresponds to the smoothed ϕ_0 for large magnitudes.


(a)

(b)

Figure 2.17: Magnitude scaling of τ_0 and ϕ_0 for $f = 5hz$; circular markers denote the standard deviations of the binned residuals, and solid lines correspond to the standard deviation models.

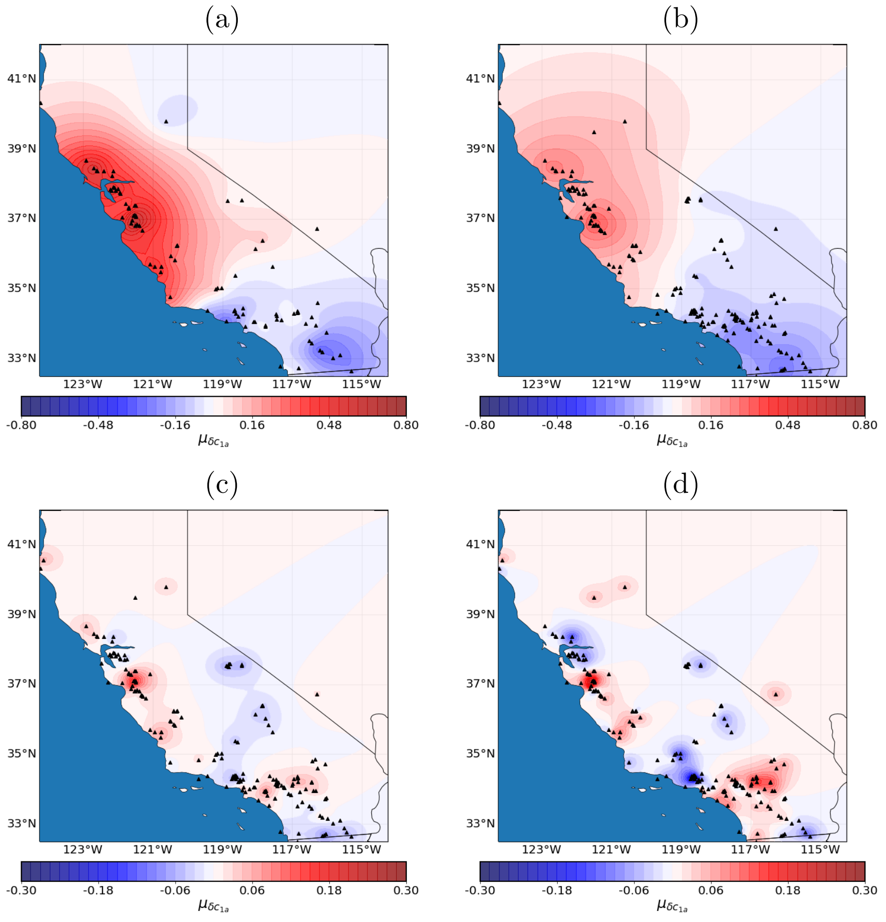


Figure 2.18: Spatial distribution of source constant, δc_{1a}, from preliminary analyses, which did not include $\delta c_{0N/S}$, for different frequencies. (a) $f = 0.2hz$, (b) $f = 1.0hz$, (c) $f = 5.0hz$, (d) $f = 10.0hz$.

The threshold between the northern and southern range of δc_{1a} at small frequencies seems to coincide with the NCSN/SCSN border used in the NGAWest2 data set (Ancheta et al., 2014) for the selection of the source parameters for the small-to-moderate magnitude events (Figure 2.19). This is why the regional difference of δc_{1a} is believed to be caused by potential inconsistencies in the magnitude estimation between the Northern and Southern California seismic networks. The low frequencies are more sensitive to a bias in the magnitude estimation, because for frequencies much smaller than the corner frequency, a 0.1 change in magnitude leads to a 40% change in amplitude, whereas at frequencies much larger than the corner frequency, a 0.1 in magnitude leads to a 10% change in amplitude. In the final analysis, the regional source differences are captured by the $\delta c_{0N/S}$ term. The northern and southern regions for $\delta c_{0N/S}$ are also shown in Figure 2.19.

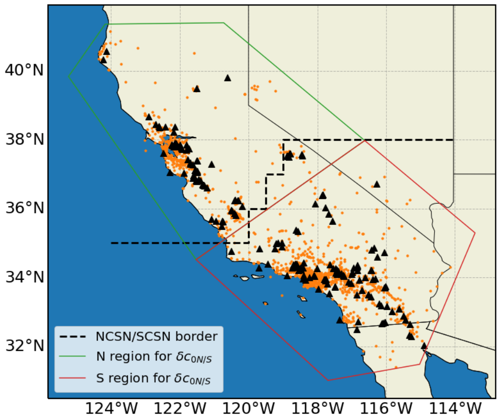


Figure 2.19: Border between Norther California Seismic Network (NCSN) and Seismic California Seismic Network (SCSN), and Norther and Southern CA regions for $\delta c_{0N/S}$.

2.4.6 Inter-frequency correlation

The inter-frequency correlation of the non-ergodic terms is presented in Figure 2.20 and the model parameters are summarized in Table 2.4. Currently, the correlation of all non-ergodic terms is modeled as frequency independent: that is, the width of the EAS peaks and troughs does not depend their central frequency. A frequency independent correlation would mean that the width of the correlation ridges in Figure 2.20 does not change along the diagonal, whereas a non-constant width would mean that the correlation changes with frequency. Out of all the non-ergodic terms, δc_{1a} has the widest confidence intervals because the number of unique earthquakes is smaller than both the number of unique stations or the number of anelastic attenuation cells. δc_{1a} has a relatively wide correlation, meaning that if δc_{1a} is positive at one frequency, it is highly likely that it will also be positive over a wide range of neighbouring frequencies. The correlation of this term also exhibits some frequency dependence similar to the δB_e frequency dependence found in Bayless and Abrahamson (2019a): the correlation is the widest at $f = 0.5hz$, it narrows at intermediate frequencies, and it widens again at frequencies larger than $8hz$. Both δc_{1b} and $\delta S2S$ have narrow, mostly frequency independent, inter-frequency correlations, which are similar to the correlation structure of $\delta S2S$ in Bayless and Abrahamson (2019a). The correlation of c_{cA} shows the strongest frequency dependence; there is very little correlation at frequencies less than $1hz$, it gradually increases and reaches the widest point at $5hz$, and then, it narrows again. The narrow frequency correlation at low frequencies is expected as the anelastic attenuation is very weak for that frequency range; however, it is unclear why the inter-frequency correlation narrows at high frequencies. It could be an artifact of poor sampling. For now, it is modeled as frequency independent, but in future studies, this assumption would need to be reevaluated. As a point of comparison, in seismic numerical simulations, a deterministic velocity model would imply a perfect inter-frequency correlation on c_{cA}, which is more similar to the width

Table 2.4: Interfequency model coefficients for non-ergodic terms

	A	B	C	D
δc_{1a}	1.94	0.77	0.96	19.49
δc_{1b}	1.30	0.92	1.36	30.85
$\delta S2S$	1.83	1.86	2.77	63.96
c_{cA}	1.85	0.41	0.27	10.00

of the correlation of data at $f = 5hz$

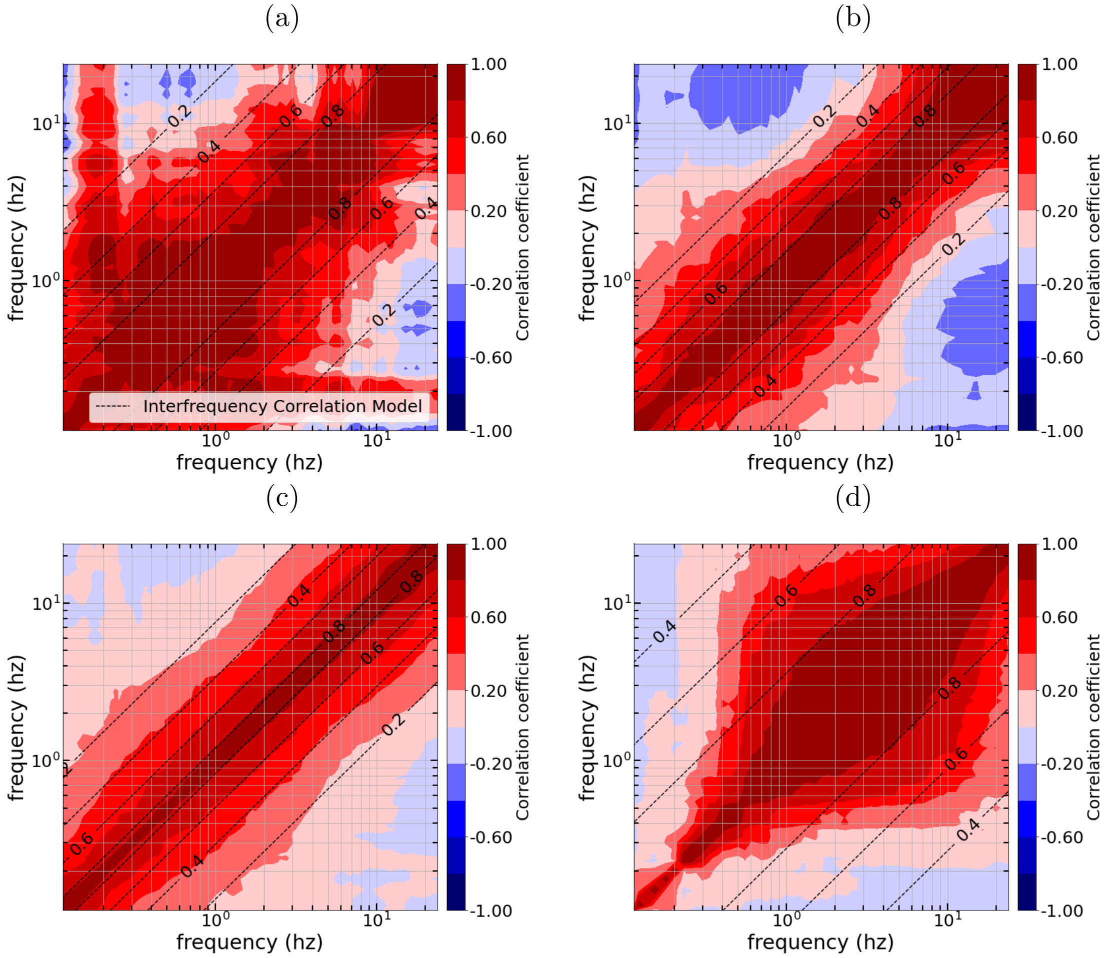
2.4.7 Examples

Figure 2.21 shows the effect of inter-frequency correlation in sampling the non-ergodic terms for an M 7 earthquake in Hayward fault 10 km from a site in Berkeley, CA. Figure 2.21a shows the median non-ergodic EAS, the 16^{th} to 84^{th} percentile range of epistemic uncertainty, and a representative EAS sample with epistemic uncertainty for zero inter-frequency correlation. Figure 2.21b shows the same information, but in this case, the ground motions were generated with the inter-frequency correlation model described previously. The median EAS and range of epistemic uncertainty is the same in both cases; what is different are the representative EAS samples. The EAS sample with zero inter-frequency correlation has zero width in the peaks and the troughs, whereas EAS sample with inter-frequency correlation has peaks and troughs that span approximately a quarter of a decade. It should be noted that these samples do not include aleatory variability, the inter-frequency correlation of the aleatory variability will influence the final width of the peaks and troughs.

The distance scaling of the model for $f = 5hz$ for a site in San Jose, CA (SJ) and a site in Northeastern California (NE) is presented in Figure 2.22; the site in SJ has a station which has recorded ground motions from past earthquakes to constrain $\delta S2S$, whereas the site in NE does not have one, so $\delta S2S$ is unconstrained. In both cases, the earthquakes are located north of sites. North of SJ, c_{cA} is less than average (Figure 2.11a) which causes the non-ergodic GMM to have higher attenuation than BA18. Due to the small number of paths in NE, c_{cA} is very close to the mean value which is why the non-ergodic GMM and BA18 have similar distance scaling. The epistemic uncertainty is less in SJ as there are more earthquakes and stations to constrain the non-ergodic terms.

2.5 Model Validation

The performance of the non-ergodic GMM was evaluated with a 10-fold cross validation test. In each of the 10 iterations of the cross-validation test, the data of NGAWest2 CA dataset were randomly split into a training and test datasets, 80% of the earthquakes composed



(a)

(b)

(c)

(d)

Figure 2.20: Inter-frequency correlation of non-ergodic terms; contour plot corresponds to the inter-frequency correlation of the data, dashed lines corresponds to the inter-frequency correlation model. (a) inter-frequency correlation of δc_{1a}, (b) inter-frequency correlation of δc_{1b}, (c) inter-frequency correlation of $\delta S2S$, and (d) inter-frequency correlation of c_{cA}.

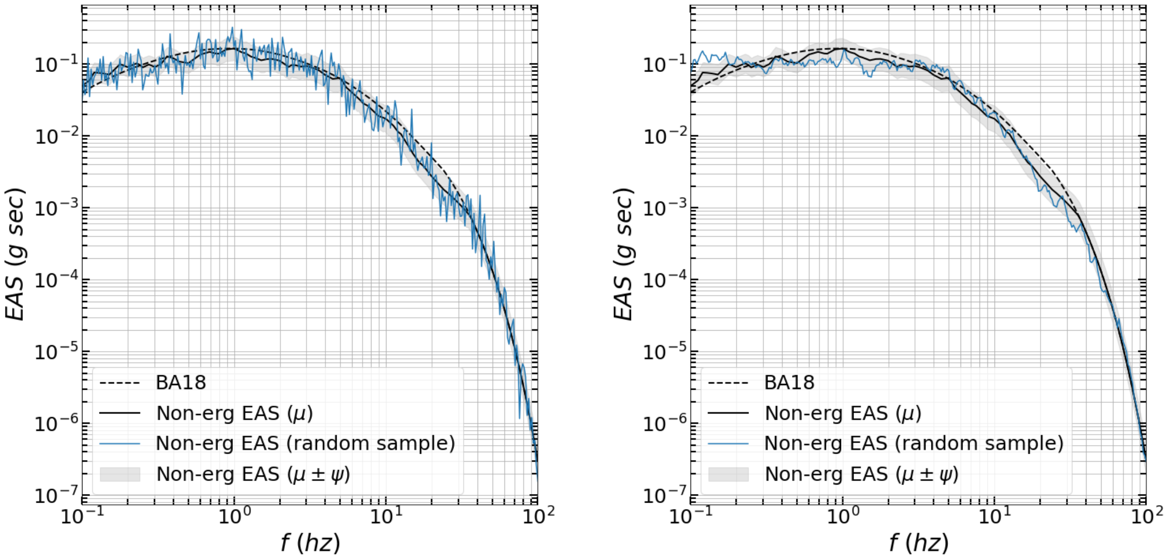

(a) (b)

Figure 2.21: Effective amplitude spectra for a M 7 earthquake, 10 km away for a site located in Berkeley CA. (a) without inter-frequency correlation, and (b) with inter-frequency correlation.

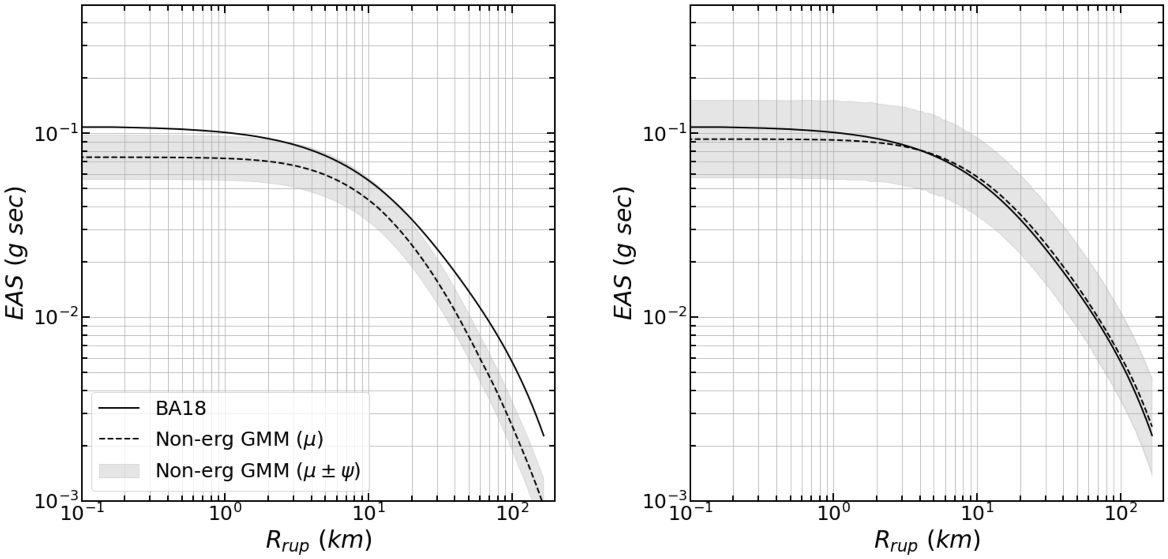

(a) (b)

Figure 2.22: Distance scaling of $EAS(f = 5hz)$ for (a) a site in San Jose and (b) a site in northeastern California.

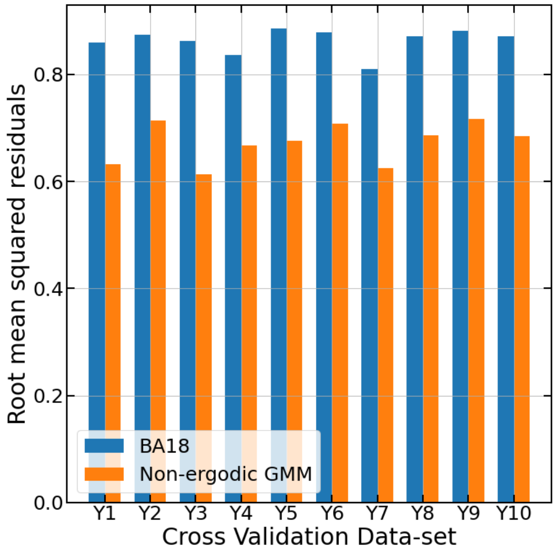


Figure 2.23: Root-mean-square error of 10-fold cross-validation test

training dataset and the remaining 20% of earthquakes composed the test dataset. The training set was used to estimate the coefficients of the non-ergodic model, and the test dataset was used to evaluate the accuracy of the predictions with the estimated coefficients. The NGAWest2 CA dataset was split based on the earthquakes so that the non-ergodic GMM is evaluated on events that were not used in the parameter estimation. Figure 2.23 shows the root-mean-square error ($rmse$) of the non-ergodic GMM and BA18 for all iterations. The average $rmse$ of the non-ergodic GMM and BA18 is 0.67 and 0.86, respectively, which indicates that incorporating the non-ergodic terms improves the ground-motion prediction for events that were not part of the regression dataset.

2.6 Conclusions and Discussion

A fully non-ergodic EAS GMM is presented in this study. The non-ergodic source and station effects are captured by spatially varying coefficients; the non-ergodic path effects are captured with the cell-specific anelastic attenuation. A regional term that accounts for the differences in the ground motion of small earthquakes between northern and southern California is also added in the non-ergodic GMM; this term is applied to events less than M 5, and frequencies less than 5 hz. The exact cause of these differences could not be identified, but it could be related to a potential bias in the magnitude estimation between the NCSN and SCSN networks. Future studies should further investigate the cause of these differences.

The proposed non-ergodic GMM has a 30 to 40% smaller standard total aleatory standard deviation than BA18. Furthermore, the cross-validation test shows that the non-ergodic

GMM performs better than BA18 in predicting the ground motion for events that were not part of the regression dataset.

The next step is to use this non-ergodic EAS GMM with RVT to develop an equivalent non-ergodic PSa GMM. The advantage of this approach is that it is easier to transfer the estimated non-ergodic terms, which are primarily based on small magnitude events, to the non-ergodic terms for the scenarios of interest, which typically are large magnitude events, using RVT than it is to estimate the magnitude dependence during the development of the non-ergodic EAS GMM. For the scenarios of interest, the PSa non-ergodic terms can be estimated by combining the EAS predictions, for the same scenarios, with RVT. In this approach, the magnitude dependence of the non-ergodic PSa terms is captured.

As larger data sets become available, future studies should consider the addition of a spatially varying term for geometrical spreading and test the frequency dependence of the inter-frequency correlation of the non-ergodic terms. A spatially-varying geometrical-spreading coefficient may be able to better capture the non-ergodic path effects at short distances; however, if such a coefficient is added, it should be constrained so that the GMM does not over-saturate at short distances. Currently, the inter-frequency correlation of the non-ergodic terms was assumed to be frequency independent, future studies should reevaluate if this assumption is valid.

Currently, the path for the cell-specific anelastic attenuation connects the site with closest point on the rupture. This path was chosen because it is the same path that is used in the R_{rup} calculation; however, it has not been tested whether a path connecting the site and a different point on the rupture would be more appropriate for the cell-specific anelastic attenuation. A large number of broadband earthquake simulations that include $3D$ velocity structure effects up to high frequencies (e.g. 5 Hz) would be ideal for solving this problem.

Chapter 3

A Non-ergodic Spectral Acceleration Ground Motion Model for California Developed with RVT

The contents of this chapter are primarily from a journal article submitted to the Bulletin of Earthquake Engineering by Lavrentiadis, G., and Abrahamson, N.A., entitled: "A Non-ergodic Spectral Acceleration Ground Motion Model for California Developed with Random Vibration Theory", which is currently under review.

3.1 Introduction

Ground-motion models (GMMs) are used to estimate the distribution of a ground-motion intensity measure (IM) for a given earthquake scenario. The most common IM is pseudo-spectral acceleration (PSa) as it is a good estimator of seismic loading for a wide range of structures. PSa is defined as the absolute maximum response of a single degree of freedom oscillator (SDOF) to an input ground motion. SDOFs are defined by their natural period (T_0) or natural frequency ($f_0 = 1/T_0$) and damping (ζ); in GMMs, typically, T_0 ranges from 0.01 to 10 *sec* and, ζ is equal to 5%. The response of the oscillator depends on the frequency content and timing (compactness of energy) of the ground motion. From the entire frequency content of the ground motion, the response of the oscillator mainly depends on the amplitudes of the frequencies near and below f_0. Therefore, at small T_0 (high f_0), the response of the oscillator depends on the entire frequency content of the ground motion (i.e. spectral shape) and not just a narrow frequency bin. This makes the coefficients of a PSa GMM at small T_0 magnitude dependent, as the shape of spectral acceleration response spectrum changes with magnitudes. The peak of a spectral acceleration response spectrum will be at 0.1 *sec* for a magnitude (M) 3 event and at 0.3 *sec* for a M 7.5 event (Figure 3.1); this means that at small magnitudes, the PGA scaling (e.g. V_{S30} coefficient) will be consistent with the scaling of $T_0 = 0.1$ *sec*, while at large magnitudes, the PGA scaling will be consistent with

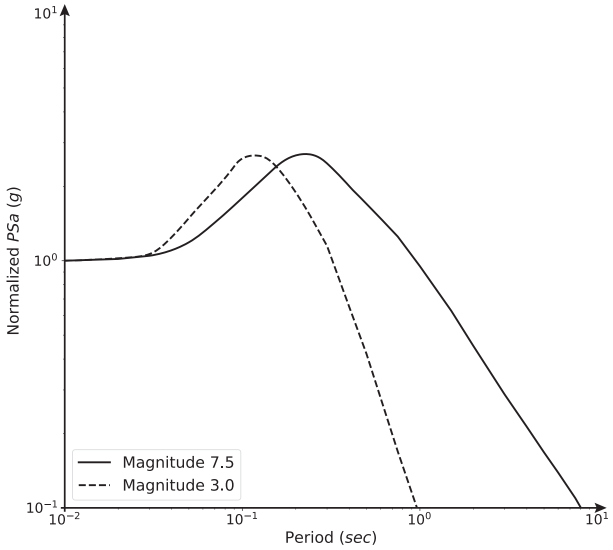


Figure 3.1: Schematic of normalized response spectra for M 3.0 and 7.5 earthquakes

$T = 0.3 sec$.

Most PSa GMMs do not explicitly account for the magnitude dependence of the coefficients, such as the VS30 scaling or distance scaling; instead, they often use a limited range of magnitudes where the magnitude dependence of the coefficients is not pronounced. For instance, the data-set that was used in the development of the NGAWest-1 GMMs had a limited set of magnitudes that ranged from M 4.5 to M 8 Power et al. (2008). The solution of using a smaller range of magnitudes works when developing an ergodic GMM, as there is enough number of moderate-to-large magnitude events globally to estimate the coefficients, but it can be problematic when developing a non-ergodic GMM.

For the NGA-W2 GMMs, the data set was extended to down to M3 with the objective of setting the reference ergodic model that could be used to evaluate regional differences in the site, path, and source terms based on small magnitude data. The NGA-W2 GMMs modified the magnitude scaling to capture the average effect of the magnitude dependence of the coefficients, but this does not accurately model the magnitude dependence of the site and path effects.

GMMs fall into two main categories: ergodic GMM and non-ergodic GMM. Ergodic GMMs assume that the statistical properties of a ground motion IM do not change in space (Anderson and Brune, 1999), and therefore, earthquakes and recordings from all around the world can be merged into a single dataset to estimate the GMM coefficients. Models developed under this assumption tend to have stable median estimates but large aleatory variability. Some models developed with the ergodic approach are: the NGA West GMMs for California Abrahamson et al. (2008), and the Douglas et al. (2014) GMM for Europe. Non-ergodic GMMs recognize that source, path, and site effects are systematically different at different parts of the world and account for these differences in the model development. Non-ergodic GMMs have smaller aleatory variability than ergodic GMMs, but in areas with sparse data, where the systematic effects are unknown, the reduced aleatory variability is accompanied by an increase in the epistemic uncertainty. The use of non-ergodic GMMs in Probabilistic Seismic Hazard Analysis (PSHA) is very promising, as the reduction in aleatory variability can have a large impact on the seismic hazard at large return periods. A more in-depth discussion of ergodic and non-ergodic GMM is provided in Chapter 3.

The estimation of the non-ergodic terms requires a large set of regional data. To achieve that, the datasets used in the development of non-ergodic GMM need to have a wider range of magnitudes to include the more frequent small-to-moderate earthquakes. It is this expansion of the magnitude range that makes the magnitude dependence of the GMM coefficients a more significant issue in non-ergodic GMMs. One solution to this problem is, first, develop a non-ergodic GMM for an IM whose scaling does not suffer from the magnitude dependence, as PSa does, and then for a scenario of interest, calculate the non-ergodic PSa based on the non-ergodic IM estimate.

The effective amplitude spectrum (EAS), defined in Goulet et al. (2018), is one such IM: the EAS is a smoothed rotation-independent average power of the Fourier amplitude spectrum (FAS) of the two horizontal components of an acceleration time history. In EAS, the amplitude at each frequency is independent of the amplitudes of the adjacent frequencies making the coefficients of an EAS GMM magnitude independent. Random vibration theory (RVT) provides a framework to calculate PSa from EAS. It relies on extreme value statistics to estimate the peak response of the oscillator directly in the Fourier domain; it does not require a phase-angle spectrum to first convert the ground motion in the time domain to compute the peak oscillator response. RVT has been used in the past to compute PSa based on FAS from seismological theory (Hanks and McGuire, 1981; Boore, 1983, 2003) Other studies, such as Boore and Joyner (1984), Liu and Pezeshk (1999) Bora et al. (2015) and, Boore and Thompson (2012), focused on semi-empirical adjustments to the RVT framework to correct for the assumptions not satisfied by ground motions, mainly the fact that acceleration time histories are not stationary signals. More recently, Kottke et al. (2018) used RVT to develop an ergodic PSa GMM for the eastern US based on an ergodic EAS GMM for the same region.

In this study we developed two non-ergodic PSa GMM. The average PSa scaling is determined by backbone ergodic PSa GMMs. The non-ergodic effects are defined in terms of non-ergodic PSa factors which are estimated by combining the non-ergodic EAS GMM in Chapter 2 with RVT.

3.2 Ground-Motion Data

A subset of the NGAWest2 data-set (Ancheta et al., 2014) was used in this study. The selected subset contains the earthquake and stations that are located in California, western Nevada, and northern Mexico. Recordings that were flagged as questionable in Abrahamson et al.

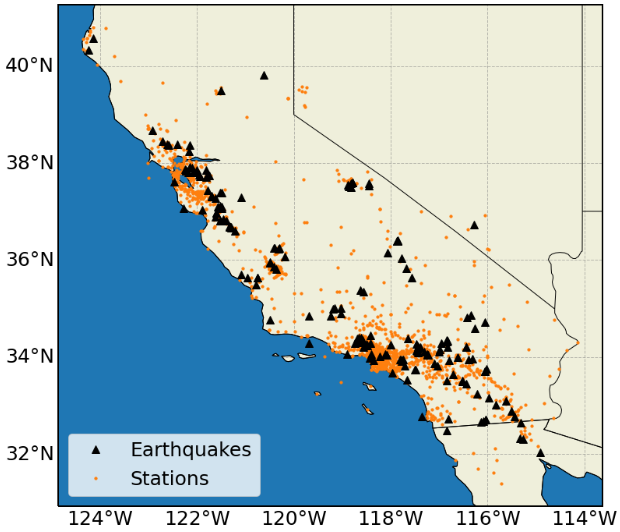


Figure 3.2: Spatial distribution for earthquakes and station used in this study.

(2014) were removed from the regression subset Figure 3.2 shows the spatial distribution of earthquakes and stations. Most of the stations are located in Los Angeles, Bay Area, and San Diego metropolitan areas, whereas spatial density of the stations is lower in less populated areas, such as northern-eastern California. The regression data-set contains 7520 records from 185 earthquakes recorded at 1410 stations. Figure 3.3 shows the magnitude-distance distribution of the data and the number of records per frequency. The magnitude of the earthquakes ranges from 3.1 to 7.3, and the distance of most records ranges from 10 to 200 km. The usable frequency range of the majority of EAS records spans from 0.4 and $20hz$. The minimum usable frequency of most PSa records is 0.5 hz.

3.3 Model development

3.3.1 Random-Vibration Theory

RVT uses Parseval's theorem and extreme value statistics (EVS) to estimate the PSa based on the frequency content (i.e. FAS) and duration of a ground motion. Parseval's theorem is used to calculate the root-mean-square of the oscillator's response (x_{rms}) to the input ground motion, and a peak factor (PF), which is based on EVS, is used to estimate the absolute peak response of the oscillator, which is the definition of PSa, based on x_{rms}. PFs assume that the ground motion is a stationary stochastic process, and that it can be described as a band-limited white Gaussian noise with zero mean. The first assumption means that the amplitudes of the ground motion are identically distributed, and the second assumption means that the phase angles of the ground motion are randomly distributed. Although,

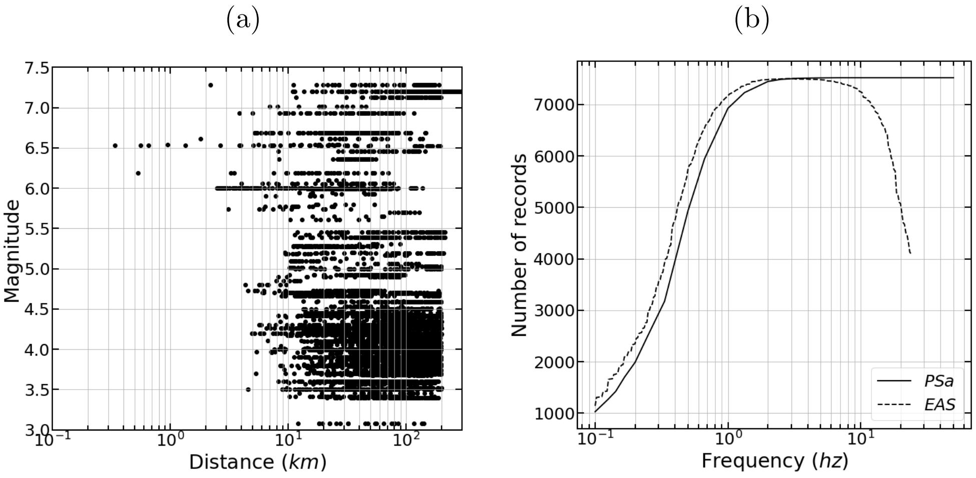

(a) (b)

Figure 3.3: Selected data from the NGAWest2 database. (a) Magnitude - Distance distribution, (b) number of PSa and EAS recordings per frequency used in the regression analysis

earthquake ground motions violate both assumptions, numerous studies have shown that RVT provides PSa estimates that are in agreement with observed ground motions (Hanks and McGuire, 1981; Boore, 1983, 2003)

Oscillator Response

The response of an oscillator to a ground motion can be computed by convolving the ground motion with the impulse response (IR) of the oscillator. IR is the response an oscillator to a very brief acceleration pulse; that is a Dirac delta function. For an SDOF oscillator, the Fourier transform of the impulse response of the spectral acceleration is:

$$IR(f, f_0, \zeta) = \frac{-f_0^2}{f^2 - f_0^2 - 2j * \zeta * f_0 * f} \tag{3.1}$$

where, f_0 is the natural frequency of the oscillator, and ζ is the oscillator's damping. As an example, Figure 3.4 shows the PSa impulse response, in time and Fourier domain, for an SDOF oscillator with $f_0 = 2hz$ and $\zeta = 5\%$ In Fourier domain, the convolution is performed by multiplying the ground motion's FAS with IR; therefore, the response of an SDOF oscillator to a ground motion is:

$$X(f) = FAS(f)\ IR_{SD}(f, f_0, \zeta) \tag{3.2}$$

The x_{rms} of the oscillator's response is defined as:

$$x_{rms} = \sqrt{\frac{1}{D_{rms}} \int_{-\infty}^{+\infty} x(t)^2 dt} \tag{3.3}$$

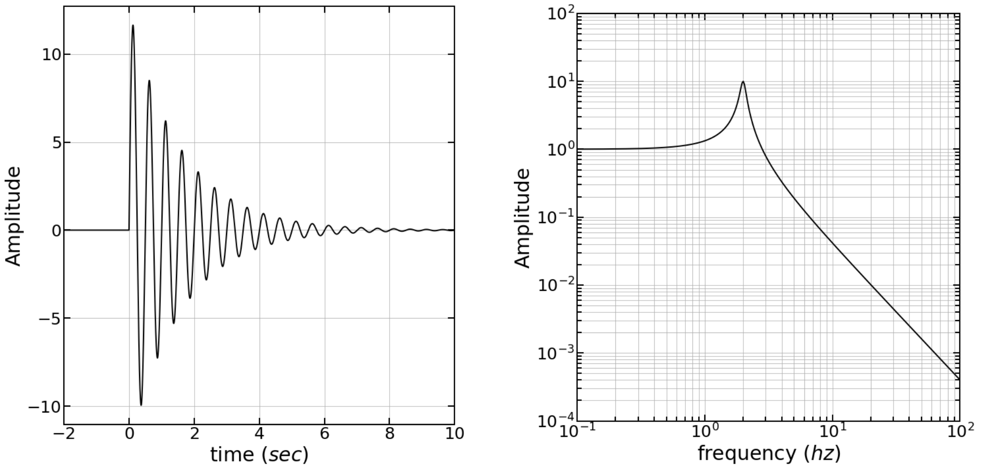

(a) (b)

Figure 3.4: Impulse response of a single degree of oscillator; (a) Time domain, (b) Fourier domain

where D_{rms} is a measure of the duration which is defined in Section 3.3.1. Parseval's theorem states that the amount of energy in the time domain is equal to the amount of energy in the Fourier domain ($\int_{-\infty}^{+\infty} x(t)^2 dt = 2\int_0^{+\infty} X(f)^2 df$) which allows to compute x_{rms} directly in Fourier domain:

$$x_{rms} = \sqrt{\frac{1}{D_{rms}} 2 \int_0^{+\infty} X(f)^2 df} = \sqrt{\frac{m_0}{D_{rms}}} \tag{3.4}$$

with m_0 being the zeroth moment of FAS. The k^{th} moment of FAS is defined as:

$$m_k = 2 \int_0^{+\infty} (2\pi f)^k X(f)^2 df \tag{3.5}$$

Peak Factor

The peak factor relates the x_{rms} with the maximum response of the oscillator (x_{max}), which is the definition of the PSa.

$$PSa = PF\ x_{rms} \tag{3.6}$$

In general, PFs fall into two main categories: those based on the Cartwright (1956) peak factor, abbreviated as CLH56, and those that are based on the Vanmarcke (1975) peak factor, abbreviated as V75.

In the first group, the CLH56 peak factor assumed that the peaks of a time history occur independently according to a Poisson process. BJ83 peak factor (Boore, 1983; Boore and

Joyner, 1984; Boore, 2003) reformulated CLH56 and removed an integrable singularity. Davenport (1964) proposed the D64 peak factor based on an asymptotic form that approximates CLH56 for long time histories.

The main difference between V75 (Vanmarcke, 1975, 1976) and the PFs of the first group is that V75 dropped the Poisson process assumption. Because of this, V75 PF accounts for the time spend outside the threshold, which is important for a narrow-band process, and considers that the peaks could be clustered in time, which is important for a wide-band process. Der Kiureghian (1980) noted that the D64 peak factor overestimates the number of zero crossings, and developed the DK80 PF by modifying D65 PF so that it is asymptotically consistent with V75. V75 and D80 are in general agreement, but they deviate in time histories with a small number of zero crossings.

The V75 PF is selected for the development of the non-ergodic PSa GMM. V75 is preferred over the group of PF that are based on CLH56 due to the simplified assumptions in CLH56, and the complete form of V75 is preferred over the asymptotic forms, as the former is more accurate for the wide range of ground motions considered in this project. This choice is consistent with the PF used in Kottke et al. (2018).

V75 expressed the probability distribution of the peaks as a first-passage problem. For a Gaussian process, the first-passage probability (i.e. the probability of no crossing) a $\pm a$ threshold (type-D barrier) in the time interval $(0, t)$ is equal to:

$$P(|z| < r) = A \ \exp\left(-f_z t \ \exp(-r^2/2) \frac{1 - \exp(-\sqrt{\pi/2} \ \delta_e \ r)}{1 - \exp(-r^2/2)} \right) \tag{3.7}$$

where r is the normalized barrier level ($r = a/x_{rms}$), A is the probability of starting within the thresholds ($A = 1 - \exp(-r^2/2)$), f_z is the average rate of zero crossings, and δ_e is an semi-empirical measure of bandwidth ($\delta_e = \delta^{1+b}$). b a non-negative constant which, in this case, is equal to 0.2, and δ is a measure of bandwidth based on the spectral moments (Vanmarcke, 1972) defined as:

$$\delta = \sqrt{1 - \frac{m_1^2}{m_0 m_2}} \tag{3.8}$$

The cumulative distribution function (CDF) of the peak values is obtained by setting t equal to D_{gm} in equation (3.7); that is, the probability of the peak of the time history being less than $r \times x_{rms}$ is equal to the probability that the time history will remain within the thresholds $\pm r \times x_{rms}$ for the entire ground-motion duration. With that, the CDF of PF is equal to:

$$\begin{aligned} F_{PF}(r) = & \big(1 - \exp(-r^2/2)\big) \\ & \times \exp\left(-f_z D_{gm} \ \exp(-r^2/2) \frac{1 - \exp(-\sqrt{\pi/2} \ \delta_e \ r)}{1 - \exp(-r^2/2)} \right) \end{aligned} \tag{3.9}$$

The expected value of PF can be computed with the probability density function (PDF) of PF (Equation (3.10)), which requires the derivation of the PDF. However, PF is continuous and defined on the positive side of the real line; thus, the expected value of PF can be computed directly from the CDF with equation Equation (3.11).

$$E[PF] = \int_0^{+\infty} r f_{PF}(r)\ dr \tag{3.10}$$

$$E[PF] = \int_0^{+\infty} \left(1 - F_{PF}(r)\right)\ dr \tag{3.11}$$

The mean estimate of the RVT PSa can be computed by substituting the expected value of the $V75$ PF in Equation (3.6).

Ground-Motion Duration

In RVT, a measure of duration is needed in two steps: in the calculation of the peak factor, and in the calculation of x_{rms}. Due to transient nature of a ground-motion, the duration measures used in these two steps are often different. D_{gm} is the ground-motion duration, which is used in the calculation of PF; D_{rms} is the duration measure for the calculation of x_{rms}. which is defined in section 3.3.1.

In seismology, the ground-motion duration is most commonly defined as either the bracketed or the significant duration. Bracketed duration is the time interval between the first and last time the ground motion exceeds a threshold. Significant duration is the difference in time the normalized Arias intensity reaches two specific values. For instance, the $5-75\%$ significant duration is the difference between the time the normalized Arias intensity is 5% and, the time the normalized Arias intensity is 75%. The Arias intensity is defined as integral of the squared acceleration time history:

$$I_a(t') = \frac{2\pi}{g} \int_0^{t'} x^2(t) dt \tag{3.12}$$

The normalized Arias intensity, also known as Husid curve, is the ratio of I_a at time t over I_a at the end of the ground motion:

$$h(t) = \frac{I_a(t)}{I_a(+\infty)} \tag{3.13}$$

In some RVT methods, D_{gm} is set to a measure of significant duration, but in others, D_{gm} is treated as a free parameter with units of time. For instance, Boore (2003) used the $D_{a0.05-0.95}$ significant duration as D_{gm}, while Bora et al. (2015) and Bora et al. (2019) treated D_{gm} as free parameter and developed a duration GMM with the goal to minimize misfit between the observed PSa and the PSa computed with RVT.

In this study, D_{gm} is defined as an interval of significant duration. Different intervals of significant duration were tested as D_{gm} candidates to find the one that minimized the misfit between the PSa of the used dataset (PSa_{NGA}) and the PSa estimated with RVT (PSa_{RVT}); the results of this comparison are shown in Appendix A, Section A.1. The $D_{a0.05-0.85}$ significant duration resulted in the best fit of PSa_{NGA} for the entire frequency range, 0.1 to 100 hz. The Abrahamson and Silva (1996) duration GMM (AS96) was selected for estimating $D_{a0.05-0.85}$ for new scenarios, as to our knowledge, AS96 is the only GMM that provides an estimate for the selected duration interval. Despite the previous results, the D_{a5-75}, D_{a5-95}, D_{v5-75}, and D_{v5-95} estimates of the Kempton and Stewart (2006) duration GMM and D_{a5-75}, D_{a5-95}, and $2D_{a20-80}$ estimates of the Afshari and Stewart (2016) duration GMM were evaluated as candidates for D_{gm}, but the $D_{a0.05-0.85}$ of AS96 resulted to a better fit of PSa_{NGA}. The results of this comparison can be found in Appendix A, Section A.2

The AS96 functional form for the mean estimate or the $D_{0.05-0.75}$ duration is:

$$\ln D_{5-75} \begin{cases} \ln\left(\frac{1}{f_c} + c_1(R_{rup} - R_c) + c_2 S\right) & for \quad R_{rup} \geq R_c \\ \ln\left(\frac{1}{f_c} + c_2 S\right) & for \quad R_{rup} < R_c \end{cases} \tag{3.14}$$

where f_c is the corner frequency of the earthquake:

$$f_c = 4.9\ 10^6 \left(\frac{\Delta\sigma}{10^{1.5M+16.05}}\right) \tag{3.15}$$

β is the shear-wave velocity at the source, and $\Delta\sigma$ is the stress drop. $1/f_c$ is the source duration, $c_1(R_{rup} - R_c)$ captures the distance dependence, and $c_2 S$ captures the site dependence. The scaling of AS96 has a physical basis because the distance and site dependence terms are additive, instead of multiplicative, to the source duration. The rational for an additive distance dependence is that small and large magnitude earthquakes are expected to have a similar increase of duration due to the dispersion of the seismic waves. Similarly, the duration increase due to the site effects is also expected to be independent of the earthquake size. In AS96, other interval of significant duration can be calculated with Equation (3.16).

$$\ln\left(\frac{D_{0.05-I}}{D_{5-75}}\right) = a_1 + a_2 \ln\left(\frac{I - 0.05}{1 - I}\right) + a_3 \ln\left(\frac{I - 0.05}{1 - I}\right)^2 \tag{3.16}$$

Correction for non-stationarity

One of RVT's main assumptions that is violated when applied in ground motions is that the signal stationary. Especially when predicting PSa for large T_0, an SDOF oscillator will not abruptly stop at the end of the ground motion, it will have a transient decaying response, which if not considered, would lead to an overestimation of x_{rms}. To solve this problem, Boore and Joyner (1984) (JB84) proposed to include the oscillator duration (D_o) in D_{rms} as shown in Equation (3.17); D_o is not included in the calculation of the PF because the response of the oscillator follows a steady decay after the end of the excitation. Liu and Pezeshk (1999) (LP99) improved the estimate of D_o by considering the spectral shape of the input time history in the D_o scaling. Boore and Thompson (2012) (BT12), and Boore and

Thompson (2015) (BT15) proposed a relationship for D_{rms}/D_{gm}; they used a more flexible functional form compared to the previous studies and considered the magnitude and distance scaling of D_{rms}/D_{gm}.

$$D_{rms} = D_{gm} + D_o \tag{3.17}$$

The BT15 oscillator duration model was selected for the subsequent analyses, as in preliminary evaluations, the RVT PSa estimates with BT15 provided a better fit to the recorded PSa than the alternative models. Although BT12 performed well in estimating the PSa of medium-to-large earthquakes, it was not selected because its is not applicable to magnitudes less than 4.

Extrapolation of EAS

To ensure that entire frequency content of the ground-motion is captured in the RVT calculations, the EAS spectrum is extrapolated at low and high frequencies. At low frequencies, EAS is extrapolated to $0.01hz$ with an omega-square model (Brune, 1970):

$$\begin{aligned} \Omega(f) &= \frac{f^2}{1 + f^2/f_c^2} \\ EAS(f < f_{min}) &= A_{f_{min}} \Omega(f) \end{aligned} \tag{3.18}$$

where f_c is the corner frequency (Equation (3.15)), and $A_{f_{min}}$ is the amplitude of the omega-squared model at the minimum frequency of the EAS (f_{min}). The stress drop for the calculation of f_c for the omega-squared model is estimated with the Atkinson and Boore (2011) empirical relationship. $A_{f_{min}}$ is estimated based on the EAS amplitudes of $1.00 f_{min}$ to $1.05 f_{min}$ frequency bin:

$$A_{f_{min}} = mean\left(\frac{EAS(f)}{\Omega(f)}\right) \quad for\ f \in [1.0 f_{min}, 1.05 f_{min}] \tag{3.19}$$

At high frequencies, EAS is extrapolated to $100hz$ with a kappa model (Anderson and Hough, 1984):

$$\begin{aligned} D(f) &= exp(-\pi \kappa f) \\ EAS(f > f_{max}) &= A_{f_{max}} D(f) \end{aligned} \tag{3.20}$$

κ defines the rate of decay of the high frequencies, and $A_{f_{max}}$ is the amplitude of the kappa model at the largest EAS frequency, f_{max}. κ can be estimated with the Ktenidou et al. (2014) $\kappa - V_{S30}$ empirical relationship:

$$\ln(\kappa) = -0.4 \ln\left(\frac{V_{S30}}{760}\right) - 3.5 \tag{3.21}$$

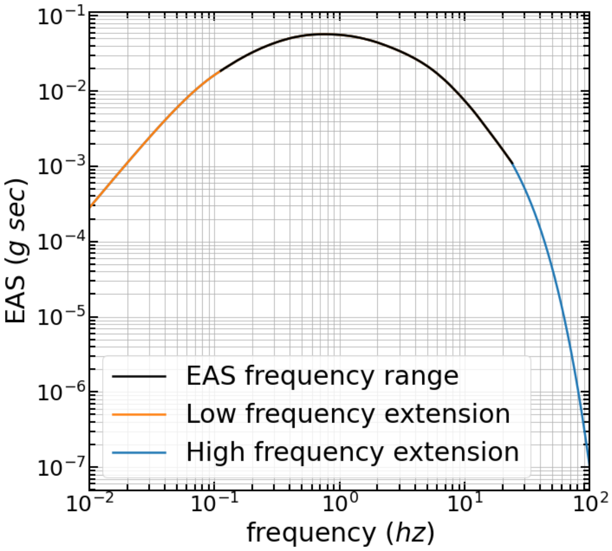


Figure 3.5: Extrapolation of EAS to low and high frequencies. EAS is estimated for $M = 7$, $R_{rup} = 30km$,and $V_{S30} = 400m/sec$.

$A_{f_{max}}$ is estimated based on the EAS amplitudes in the $0.95 f_{max}$ to $1.00 f_{max}$ frequency bin:

$$A_{f_{max}} = mean \left(\frac{EAS(f)}{D(f)} \right) \quad for\ f \in [0.95 f_{max}, 1.00 f_{max}] \tag{3.22}$$

As an example of the extrapolation procedure, the median estimate of EAS for a M 7 event, at a R_{rup} distance of $30km$, and a V_{S30} value of $400m/sec$ is extend to high and low frequencies using the omega-squared and kappa models in Figure 3.5, which shows that the amplitudes of the extended frequencies are in agreement with the EAS over the usable frequency range.

RVT summary and validation

In summary, all subsequent RVT calculations are performed with: the V75 PF, the median estimate of AS96 for $D_{a0.05-0.85}$ as D_{gm}, BT15 for D_{rms}, and the extrapolation procedure described in the previous subsection.

As a validation, Figure 3.6 shows the residuals between the natural-log of PSa_{NGA} and the natural-log of PSa_{RVT} with the recommended RVT procedure. Overall, PSa_{RVT} is in good agreement with PSa_{NGA} for the entire period range ($T_0 = 0.01 - 10sec$) with the fit improving for $M > 5$. Figure 3.7 shows the mean and the standard deviation of the residuals versus T_0. The residuals have a positive bias at $T_0 = 1 - 4sec$; however, this is not propagated in the non-ergodic PSa GMM, as the GMM is developed using non-ergodic ratios, which

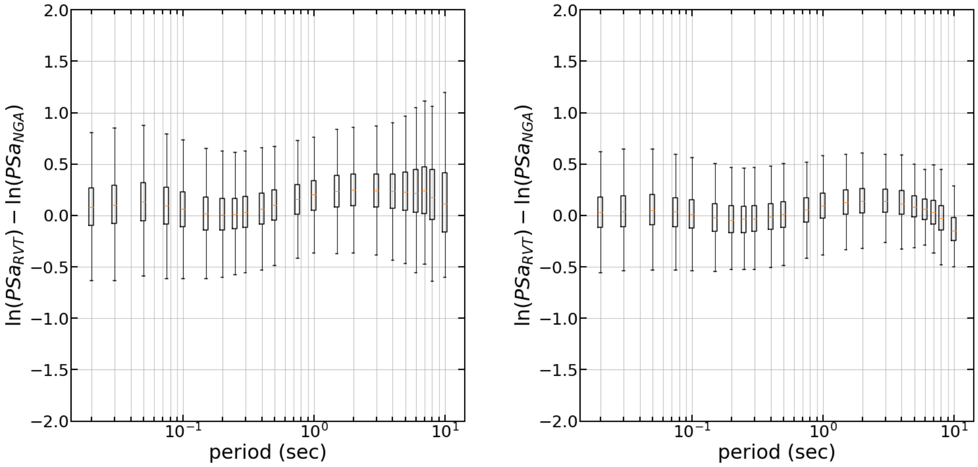

(a) (b)

Figure 3.6: Residuals between the records' PSa and PSa calculated with RVT. (a) residuals of records of all M, (b) residuals of records of $M > 5$

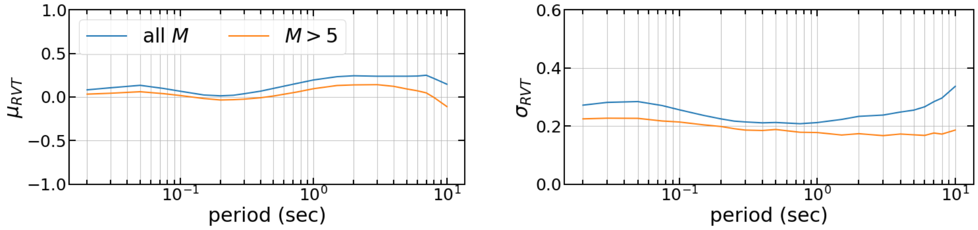

(a) (b)

Figure 3.7: Mean and standard deviation of the residuals between the records' PSa and PSa calculated with RVT.

are defined in the next subsection (Section 3.3.2). The standard deviation or the residuals is approximately 0.2 natural-log units for the entire period range.

3.3.2 Non-ergodic PSa ratios

The non-ergodic effects of the proposed PSa GMM are expressed in terms of non-ergodic PSa ratios (r_{nerg}); that is, the ratio of the log of the non-ergodic PSa estimate for a scenario of interest over the ergodic PSa estimate for the same scenario (Equation (3.23)) The non-ergodic PSa values are calculated with RVT and the non-ergodic EAS GMM (LAK21$_{\text{EAS}}$), and the ergodic PSa values are calculated with RVT and the Bayless and Abrahamson (2019b) ergodic EAS GMM (BA18). The scenarios of interest are defined by the magnitude

(M), closest-rupture distance (R_{rup}), time-average shear-wave velocity at the top $30m$ (V_{S30}), etc., which are input parameters to both the ergodic and non-ergodic EAS GMMs, but also the earthquake and site coordinates, x_{eq} and x_{site}, which define the source, path and site non-ergodic effects in $\text{LAK21}_{\text{EAS}}$. In this formulation, r_{nerg} captures the combined effect of all non-ergodic terms; there are no separate terms for the earthquake, path, and site non-ergodic effects.

$$\begin{aligned} r_{nerg}(T_0, M, R_{rup}, V_{S30}, x_{eq}, x_{site}, ..) = \\ = \ln\left(PSa_{RVT}\left[IR(T_0)\right)\ EAS_{LAK21_{EAS}}(M, R_{rup}, V_{S30}, x_{eq}, x_{site}, ...)\right]\right) \\ - \ln\left(PSa_{RVT}\left[IR(T_0)\ EAS_{BA18}(M, R_{rup}, V_{S30}, ...)\right]\right) \end{aligned} \tag{3.23}$$

The proposed non-ergodic PSa GMM is developed by coupling the aforementioned non-ergodic ratios with an existing ergodic PSa GMM:

$$\begin{aligned} y_{nerg}(M, R, V_{S30}, x_{eq}, x_{site}, ...) = \\ = r_{nerg}(M, R, V_{S30}, x_{eq}, x_{site}, ...) + y_{erg}(M, R, V_{S30}, ...) \end{aligned} \tag{3.24}$$

where y_{nerg} is the natural log of the non-ergodic PSa median estimate, and y_{erg} is the natural log of the ergodic median estimate. The benefit of this approach is that it separates the non-ergodic effects from the average ground-motion scaling. r_{nerg} does not affect the average scaling of the non-ergodic PSa GMM, as $\text{LAK21}_{\text{EAS}}$ is based on BA18, and thus, their average scaling is canceled out. Furthermore, the small bias of RVT is also canceled out in this approach, as the same RVT procedure is used to compute PSa_{erg} and PSa_{nerg} For the average scaling of the non-ergodic PSa GMM, y_{erg}, we chose the Abrahamson et al. (2014) (ASK14) and Chiou and Youngs (2014) (CY14) ergodic PSa GMMs. Hereafter, the non-ergodic GMM that is based on ASK14 is called non-ergodic GMM_1, and the non-ergodic GMM that is based on CY14 is called non-ergodic GMM_2. The main reasons ASK14 and CY14 are selected to develop the non-ergodic GMM are: i) they were developed with the same data-set as BA18, and ii) they include complex scaling terms, such as hanging-wall effects, which can be passed to the non-ergodic GMMs.

The non-ergodic PSa GMM was not developed directly with RVT and $\text{LAK21}_{\text{EAS}}$ as that this approach was leading to an overestimation the median PSa at medium-to-large periods. Figure 3.8 compares the four NGAWest2 GMMs: ASK14, BSSA14, CB14, and CY14 (Abrahamson et al., 2014; Boore et al., 2014; Campbell and Bozorgnia, 2014; Chiou and Youngs, 2014) with the spectral acceleration response spectrum created with RVT and BA18. The NGAWest2 GMMs are in good agreement with the PSa from BA18 for the M 5 event, but the comparison worsens as the size of the earthquake increases. For periods of $T_0 = 2 - 4sec$ from a M 8 earthquake, the PSa from BA18 is a factor of two higher than the NGAWest2 GMMs, indicating that, in this period range, BA18 has a stronger magnitude scaling than the NGAWest2 GMMs. Since $\text{LAK21}_{\text{EAS}}$ is based on BA18, a non-ergodic PSa GMM developed with RVT and $\text{LAK21}_{\text{EAS}}$ will also have a stronger magnitude scaling than the NGAWest2 GMMs. Due to the effort involved in the development of the NGAWest2 GMMs, we judge that their magnitude scaling is more likely to be correct, which is why

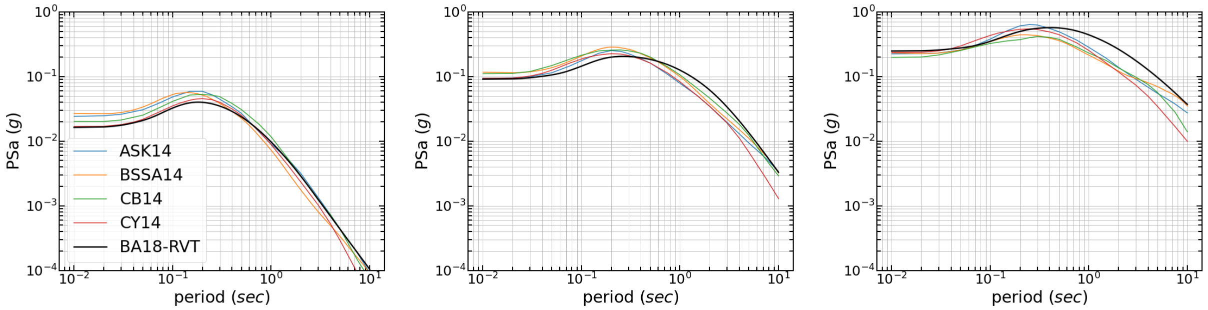

(a) (b) (c)

Figure 3.8: Comparison of PSa spectra developed with the BA18 EAS GMM and RVT, shown with the black line, and PSa spectra estimated using the NGAWest2 GMMs, shown with the colored lines. (a) M 5.0, (b) M 6.5, and (c) M 8.0 earthquake scenario with $R_{rup} = 30\ km$ and $V_{S30} = 400\ m/sec$.

we used the non-ergodic ratios approach to develop the non-ergodic GMM; however, future studies should further investigate the cause of the different magnitude scaling.

3.3.3 Constant Swift and Aleatory Model

The constant shift (δc_0), between-event residuals (δB^0_e), and within-event residuals ($\delta W^0_{e,s}$) are estimated by fitting a mixed-effects linear model to the total residuals of the non-ergodic models:

$$\epsilon_{e,s} = \delta c_0 + \delta B^0_e + \delta W^0_{e,s} \tag{3.25}$$

The magnitude dependence of δB^0_e and $\delta W^0_{e,s}$ of the two non-ergodic PSa GMMs for $T_0 = 0.25 sec$ is evaluated in Figure 3.9. The mean of δB^0_e and $\delta W^0_{e,s}$ shows no trend with M, but their emperical standard deviation decreases with M. Similarly, the R_{rup} and V_{S30} dependence of the $\delta W^0_{e,s}$ for $T_0 = 0.25 sec$ is evaluated in Figures 3.10 and 3.11 where no significant trends are found both in the mean and in the standard deviation.

Figure 3.12 shows the estimated and smoothed δc_0 of the two non-ergodic PSA GMMs. Non-ergodic GMM$_2$, which is based on CY14, is only estimated up to $T_o = 5\ sec$ because at larger periods δc_0 deviated significantly from zero.

Based on the empirical standard deviation of the non-ergodic residuals (Figure 3.9), both ϕ_0 and τ_0 are modeled as magnitude dependent (Equation (3.26) and (3.27)). Figure 3.13 shows the period dependence of ϕ_0 and τ_0 for small and large magnitudes. The magnitude dependence of ϕ_0 and τ_0 is more significant at small periods. The increase of the within-event aleatory variability at the small periods of small magnitudes may be caused by the radiation patterns which make the amplitude of the ground motion sensitive to the azimuthal angle. In large magnitudes, which can be thought as many small events, the radiation patterns

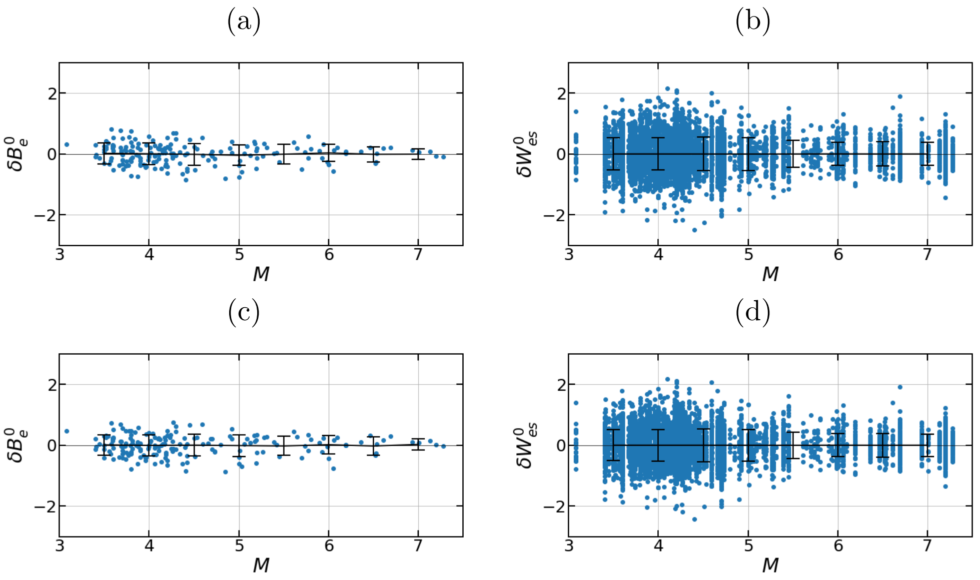

(a) (b) (c) (d)

Figure 3.9: Between-event and within-event residuals for $T_0 = 0.25sec$ versus magnitude. (a) δB_e of non-ergodic GMM$_1$, (b) $\delta W_{e,s}$ of non-ergodic GMM$_1$, (c) δB_e of non-ergodic GMM$_2$, and (d) $\delta W_{e,s}$ of non-ergodic GMM$_2$.

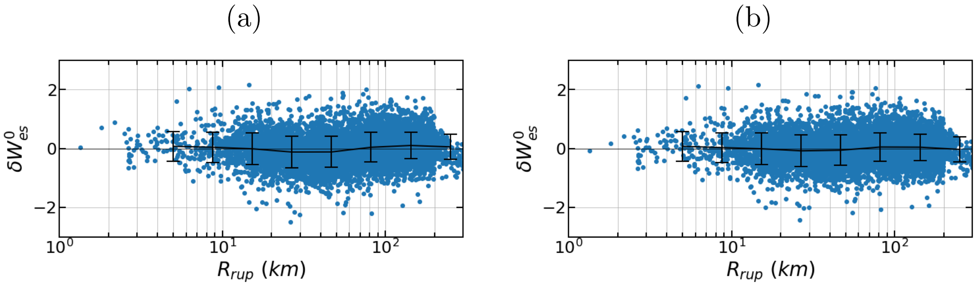

(a) (b)

Figure 3.10: Within-event residuals for $T_0 = 0.25sec$ versus R_{rup}. (a) $\delta W_{e,s}$ of non-ergodic GMM$_1$, (b) $\delta W_{e,s}$ of non-ergodic GMM$_2$.

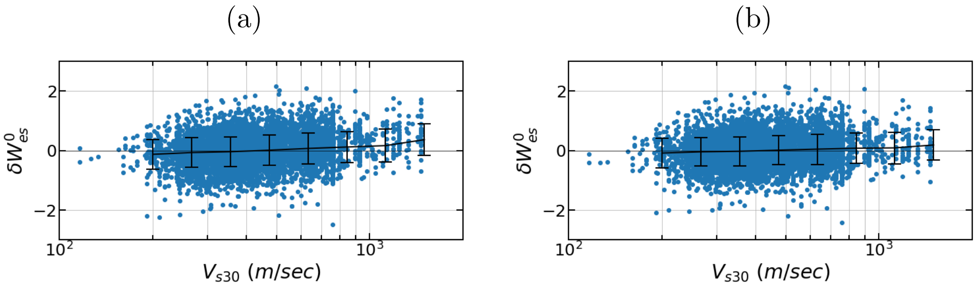

(a) (b)

Figure 3.11: Within-event residuals for $T_0 = 0.25sec$ versus V_{S30}. (a) $\delta W_{e,s}$ of non-ergodic GMM$_1$, (b) $\delta W_{e,s}$ of non-ergodic GMM$_2$.

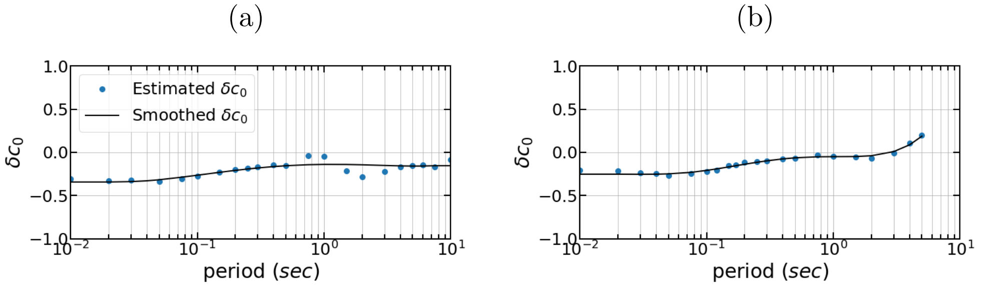

(a) (b)

Figure 3.12: Estimated and smoothed $deltac_0$ versus T_0. (a) non-ergodic GMM$_1$, (b) non-ergodic GMM$_2$

have less impact on the ground-motion variability, because the individual radiation patterns destructively interfere with each other due to the different azimuthal angles. Similarly, the larger between-event aleatory variability at the small periods of small magnitudes is believed to be caused by differences in stress drop which shifts the ground motions at frequencies above the corner frequency of the earthquake. Due to the large‘r rupture dimensions of the large events, any variability in the stress drop along the rupture averages out resulting in reduced between-event variability.

The total standard deviation of the two non-ergodic GMMs are 30 to 35% smaller than the total standard deviation of ergodic GMMs.

$$\phi_0 = \begin{cases} \phi_{0M_1} & for \quad M < 5 \\ \phi_{0M_1} + (\phi_{0M_2} - \phi_{0M_2})(M-5)/(6.5-5) & for \quad 5 < M < 6.5 \\ \phi_{0M_2} & for \quad M > 6.5 \end{cases} \tag{3.26}$$

$$\tau_0 = \begin{cases} \tau_{0M_1} & for \quad M < 5 \\ \tau_{0M_1} + (\tau_{0M_2} - \tau_{0M_2})(M-5)/(6.5-5) & for \quad 5 < M < 6.5 \\ \tau_{0M_2} & for \quad M > 6.5 \end{cases} \tag{3.27}$$

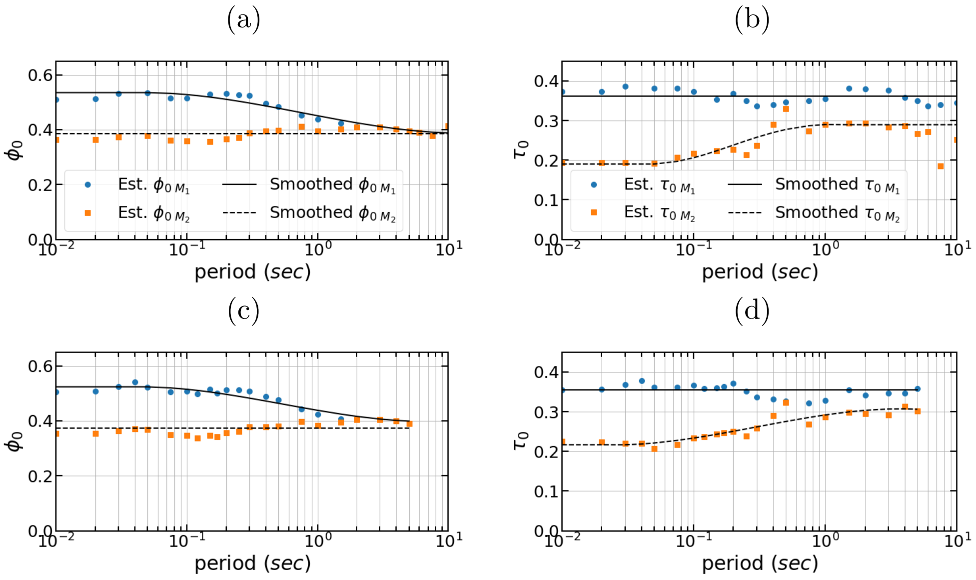


Figure 3.13: Period dependence of aleatory model parameters. (a) period dependence of ϕ_{0M_1}, ϕ_{0M_1} for non-ergodic GMM$_1$ (b) period dependence of τ_{0M_1}, ϕ_{0M_1} for non-ergodic GMM$_1$ (c) period dependence of ϕ_{0M_1}, ϕ_{0M_1} for non-ergodic GMM$_2$ (d) period dependence of τ_{0M_1}, ϕ_{0M_1} for non-ergodic GMM$_2$

Figure 3.14 compares the proposed models for ϕ_0 and τ_0 with the standard deviations of the binned residuals for $T_0 = 0.25sec$. Overall, the aleatory models are in good agreement with the empirical standard deviations. The discrepancy at large magnitudes is considered acceptable, as the number of large magnitude events is small to reliably estimate the empirically standard deviation.

As a comparison with previous non-ergodic models, Figure 3.15 shows the total standard deviation of the two non-ergodic GMMs and the total standard deviation of the SWUS15 partially non-ergodic GMM (Abrahamson et al., 2015). The standard deviations of non-ergodic GMM$_1$ and GMM$_2$ are within the low and high branches of SWUS15 for entire period range for both small-to-moderate and large events. More specifically, for small-to-moderate magnitude events and $T_0 < 1sec$, the total standard deviations of GMM$_1$ and GMM$_2$ are larger than the median branch of SWUS15. One possible reason for this is that σ_{SS} of SWUS15 was estimated with magnitudes greater than 4, whereas σ_0 of GMM$_1$ and GMM$_2$ were estimated with magnitudes greater than 3 which exhibit larger variability at small periods. At large events, the total standard deviations of GMM$_1$ and GMM$_2$ are between the central and lower branch of SWUS15. The GMM$_1$ and GMM$_2$ σ_0 values are expected to be less than SWUS15 σ_{SS} central branch because in addition to the systematic site effects, GMM$_1$ and GMM$_2$ capture the systematic source and path effects. However, the fact that the σ_0 GMM$_1$ and GMM$_2$ are larger than the lower branch of SWUS15 means that the majority

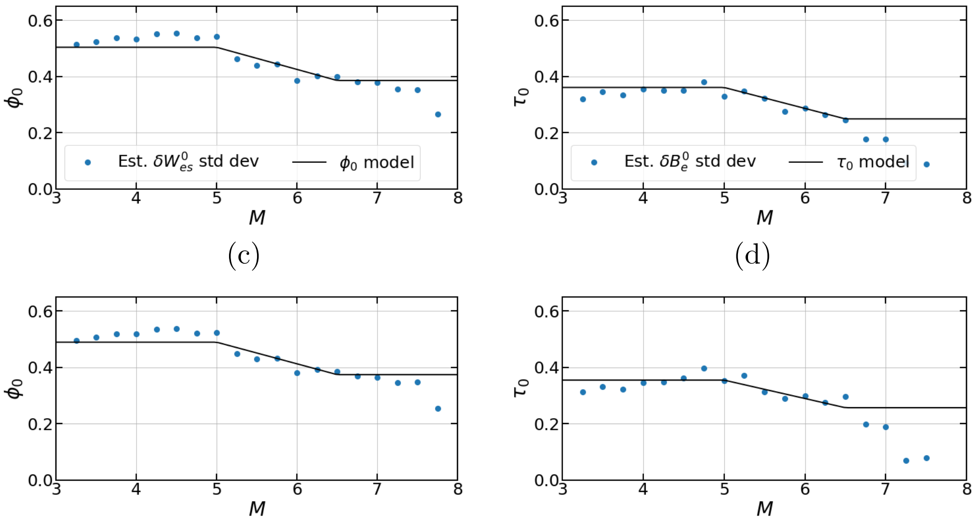

(a) (b) (c) (d)

Figure 3.14: Magnitude dependence of ϕ_0 and τ_0 for $T_0 = 0.25 sec$. Circular markers denote the standard deviations of the binned residuals, and solid lines correspond to the standard deviation models. (a) ϕ_0 for non-ergodic GMM$_1$, (b) τ_0 for non-ergodic GMM$_1$, (a) ϕ_0 for non-ergodic GMM$_2$, and (b) τ_0 for non-ergodic GMM$_2$

of the systematic effects captured by GMM$_1$ and GMM$_2$ are related to the site effects.

3.4 Applications

3.4.1 Effect of EAS inter-frequency correlation in $r_{nerg\ PSa}$

In most GMMs, the ground-motion amplitude (i.e. PSa or EAS) at every frequency is estimated independently; however, an actual ground-motion recording has peaks and troughs. That is the amplitudes of neighbouring frequencies are correlated. For instance, if amplitude of some frequency is above the average, it is likely that amplitudes of the nearby frequencies will also be above the average. This inter-frequency correlation is important in RVT, as the response of an SDOF oscillator does not only depend on the ground-motion amplitude at T_0 but also at the frequency content around T_0. Bayless and Abrahamson (2018) showed that the PSa variability is underestimated if the inter-frequency correlation of FAS is not considered.

To illustrate the effect of the inter-frequency correlation in the calculation of $r_{nerg\ PSa}$, we applied the proposed non-ergodic GMM with and without the inter-frequency correlation in EAS. In both cases, the scenario of interest is a $M7$ earthquake in Hayward Fault $8km$ away from a site in Berkeley, CA. The ergodic and non-ergodic EAS of the two approaches are shown

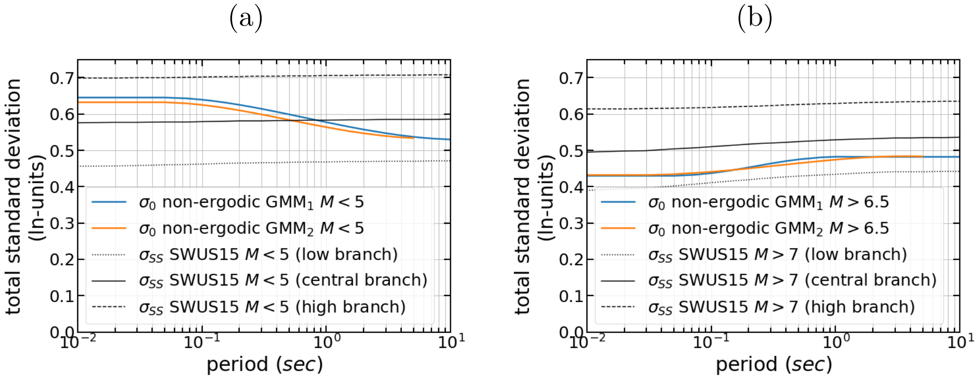


Figure 3.15: Comparison of total standard deviation of non-ergodic GMM$_1$ and GMM$_1$ with total standard deviation of SWUS15 partially non-ergodic GMM. (a) small-to-moderate magnitude comparison, and (b) large magnitude comparison

in Figure 3.16, and the corresponding non-ergodic PSa spectra are shown in Figure 3.17. The non-ergodic EAS in Figure 3.16a are developed without inter-frequency correlation, whereas the non-ergodic EAS in figure 3.16b are developed using the inter-frequency correlation model in Section 2.4.6 of Chapter 2.

In EAS space, both approaches resulted in the same median and epistemic uncertainty range, but in PSa space, only the median is the same. The epistemic uncertainty of PSa is larger when the EAS inter-frequency correlation is considered, because if EAS is at an extreme at T_0 it will generally stay at the extreme over the neighbouring frequencies; thus, all the frequencies which influence the response of the oscillator will constructively interfere leading in a range of PSa amplitudes that is wider. In contrast, if the EAS amplitudes are uncorrelated, they will have negating effect on the response of the oscillator, resulting in a narrower range of PSa. This shows the importance of considering the EAS inter-frequency correlation in the non-ergodic PSa calculations, as otherwise, the epistemic uncertainty of the PSa is underestimated.

3.4.2 Magnitude dependence $r_{nergPSa}$

As an application example, Figures 3.18 and 3.19 present the EAS and PSa non-ergodic ratios for $T_0 = 0.1sec$ ($f_0 = 10hz$) for a $M3$ and $M8$ earthquake in San Andreas fault. The EAS non-ergodic ratios are magnitude independent; the median estimate and epistemic uncertainty of $r_{nerg\ EAS}$ is the same in both events (Figure 3.18). The magnitude independence allows $r_{nergEAS}$ to be estimated from the more frequent small magnitude earthquakes and directly applied to the large magnitude events, which are typically of more interest. This is not the case for the PSa non-ergodic ratios; $r_{nergPSa}$ depend on the spectral shape; which is why $r_{nergPSa}$ are different in the $M3$ and $M8$ earthquakes (Figure 3.19), which illustrates why the non-ergodic PSa GMM is developed with non-ergodic ratios that based on EAS. Most

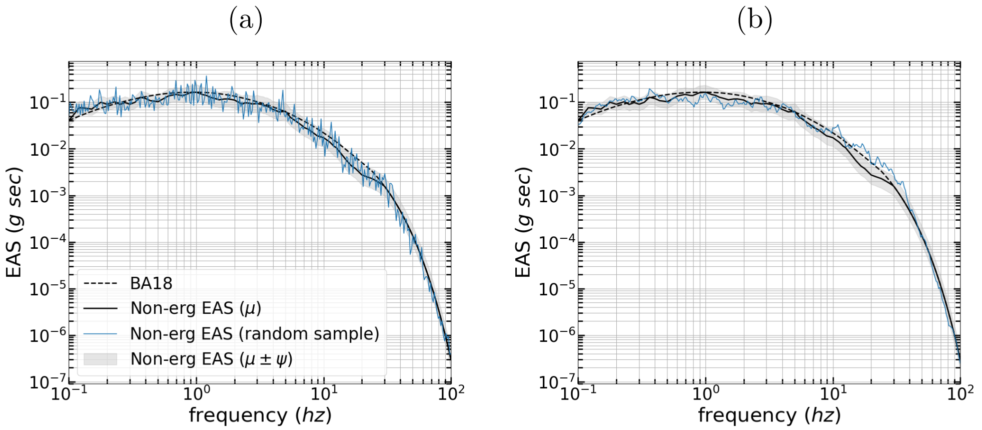


Figure 3.16: Effective amplitude spectra for a M 7 earthquake in Hayward fault, 8 km away from a site located in Berkeley CA. (a) without inter-frequency correlation, and (b) with inter-frequency correlation.

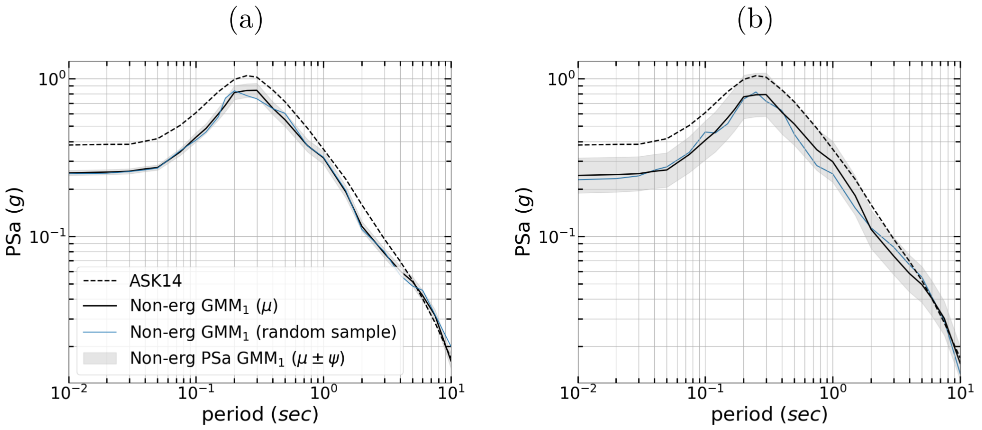


Figure 3.17: PSa spectra for a M 7 earthquake in Hayward fault, 8 km away from a site located in Berkeley CA. (a) without inter-frequency correlation, and (b) with inter-frequency correlation.

of the regional data that are used to estimate the non-ergodic effects are in form of small magnitude events, which couldn't be used if PSa non-ergodic effects were estimated directly.

In addition, Figures 3.18 and 3.19 show the spatial distribution of the epistemic uncertainty. In this example, where the location of the earthquake is fixed, the spatial distribution of the epistemic uncertainty depends on the path and site location. Both the EAS and PSa epistemic uncertainties are small near stations that have recorded past events, whereas in remote areas with no available ground-motion data to constrain the non-ergodic terms, the epistemic uncertainties are larger.

The evaluation of the magnitude dependence of the EAS and PSa non-ergodic ratios is further examined in Figures 3.20 and 3.21. The three scenarios in this comparison are a M 3, 5.5 and 8 event in San Andreas Fault, $105km$ from the site in San Francisco, CA. As mentioned previously, the non-ergodic EAS ratios are the same for all three events (Figure 3.20, while the non-ergodic PSa ratios are different, especially at small periods (Figure 3.21), $T_0 < 0.1sec$. This happens because, for $f_0 > 10hz$ ($T_0 < 0.1sec$), there is little ground-motion content in EAS to resonate the SDOF oscillator, making its response, and subsequently PSa, depended on the peak of each spectrum. Similarly, the non-ergodic PSa ratios for $T_0 < 0.1sec$ depend on the non-ergodic EAS ratios at the peak of each spectrum. In this example, the M 3 event has the largest non-ergodic PSa ratios at $T_0 < 0.1sec$, because the non-ergodic EAS ratios are predominately positive over its peak ($f = 2$ to $6hz$). The M 8 event has the smallest non-ergodic PSa ratios at $T_0 < 0.1sec$ because its peak ($f < 0.1$ to $6hz$) encompasses the dip of the non-ergodic EAS ratios that occur from $f = 0.3$ to $2hz$.

3.5 Conclusions

A new approach to develop non-ergodic PSa GMMs is presented in this study which considers the magnitude dependence of the non-ergodic terms. Due to the linear properties of Fourier Transform, a non-ergodic EAS GMM is used to estimate the non-ergodic effects from the small magnitude events and transfer them to the events of interest. RVT is used to compute the non-ergodic PSa effects based on the non-ergodic EAS effects, while the average scaling of the non-ergodic PSa GMM is controlled by an existing ergodic PSa GMM.

Two non-ergodic PSa GMMs are developed in this study. The first one uses the ASK14 GMM as a backbone model for the average scaling and is applicable to periods $T_0 = 0.01 - 10sec$. The second one uses the CY14 GMM as a backbone model for the average scaling and is applicable to periods $T_0 = 0.01 - 5sec$. The non-ergodic PSa effects are quantified in terms of non-erodic PSa ratios, that is the difference between the log of PSa estimated with RVT and the non-ergodic EAS and the log of PSa estimated with RVT and the ergodic EAS. In both cases, the $\text{LAK21}_{\text{EAS}}$ GMM is used for the non-ergodic EAS and the BA18 GMM is used for the ergodic EAS. The RVT calculations are performed with the V75 PF, the median estimate of D_{a5-85} from AS96 for the ground-motion duration, and the BT15 for the oscillator duration. The RVT components were chosen based on a thorough evaluation of alternative models for the peak factors, ground-motion duration and oscillator

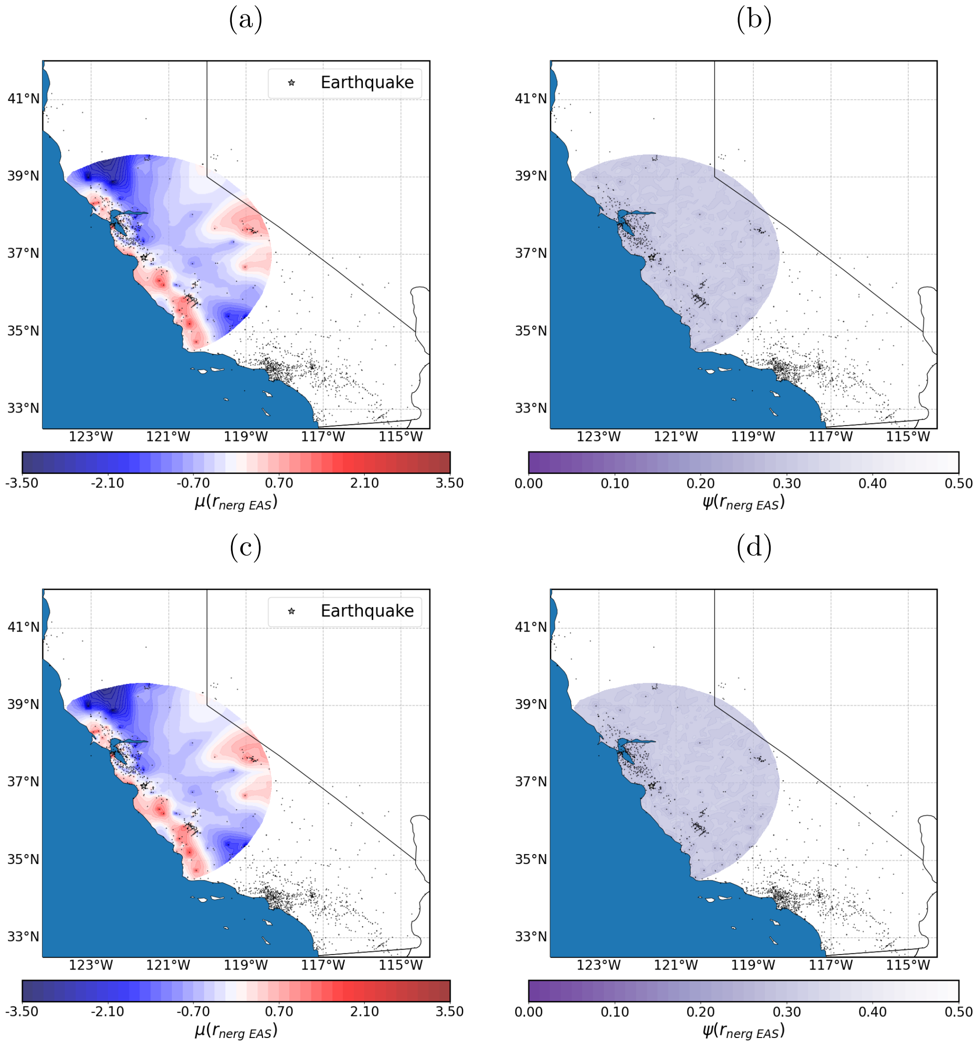


Figure 3.18: *EAS* non-ergodic ratios, $r_{nerg\ EAS}$, for $f_0 = 10hz$ for an earthquake in San Andreas. The star corresponds to the earthquake location, and the dots correspond the location of the stations in the used dataset. (a) mean of $r_{nerg\ EAS}$ for $M = 3.0$, (b) epistemic uncertainty of $r_{nerg\ EAS}$ for $M = 3.0$ (c) mean of $r_{nerg\ EAS}$ for $M = 8.0$, and (d) epistemic uncertainty of $r_{nerg\ EAS}$ for $M = 8.0$

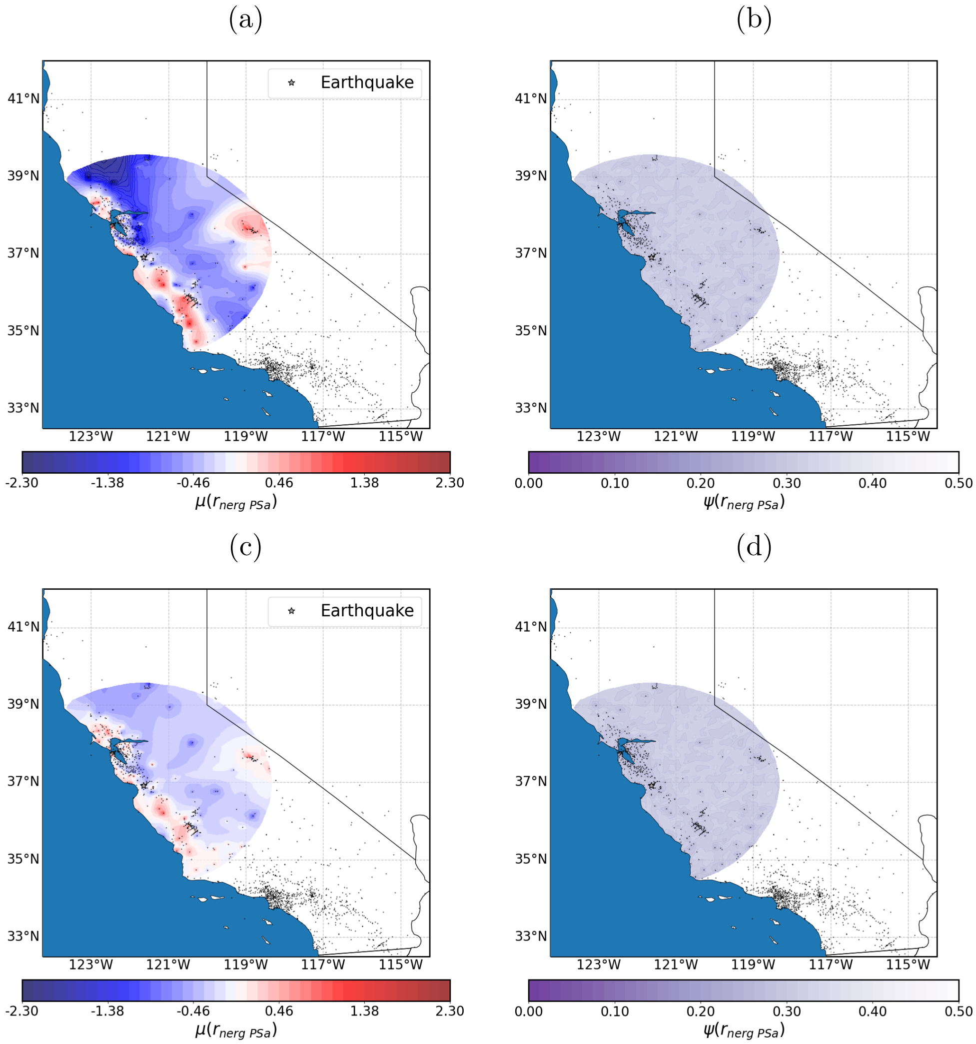

(a) (b) (c) (d)

Figure 3.19: PSa non-ergodic ratios, $r_{nerg\ PSa}$, for $T_0 = 0.1sec$ for an earthquake in San Andreas. The star corresponds to the earthquake location, and the dots correspond the location of the stations in the used dataset. (a) mean of $r_{nerg\ PSa}$ for $M = 3.0$, (b) epistemic uncertainty of $r_{nerg\ PSa}$ for $M = 3.0$ (c) mean of $r_{nerg\ PSa}$ for $M = 8.0$, and (d) epistemic uncertainty of $r_{nerg\ PSa}$ for $M = 8.0$

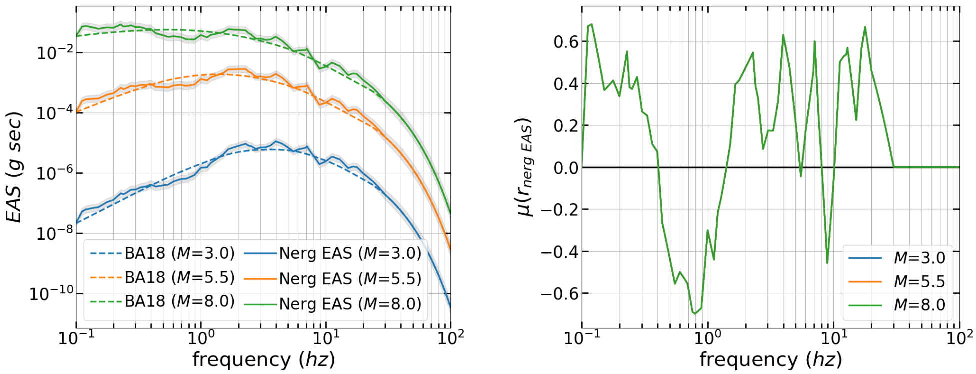

(a) (b)

Figure 3.20: (a) Ergodic and non-ergodic EAS for $M3$, 5.5, and 8 earthquakes in San Andreas fault, $105km$ from a site in San Francisco, CA (b) non-ergodic EAS ratios for the same scenarios.

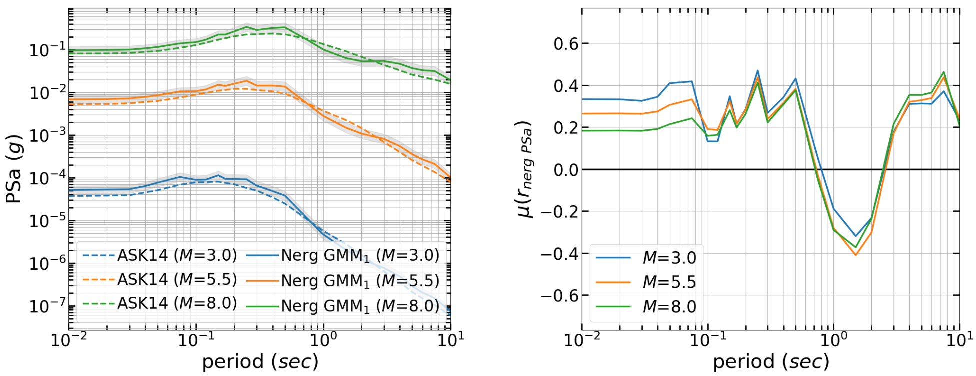

(a) (b)

Figure 3.21: (a) Ergodic and non-ergodic PSa spectra for $M3$, 5.5, and 8 earthquakes in San Andreas fault, $105km$ from a site in San Francisco, CA (b) non-ergodic PSa ratios for the same scenarios.

duration. The objective of the evaluation was to minimize misfit between the observed PSa and the PSa computed with RVT.

The advantages of developing the non-ergodic GMM with an ergodic backbone model and non-ergodic PSa ratios, instead of developing it directly with RVT and the $\mathrm{LAK21_{EAS}}$ are: i) the elimination of the small bias of RVT at $T_0 = 1-4sec$, ii) the separation of the non-ergodic effects from average scaling, and iii) the adoption of complex scaling terms present in ergodic PSa GMMs. Compared to the recorded PSa, the PSa estimated with RVT has a small positive bias at $T_0 = 1-4sec$. This bias is not propagated in the non-ergodic PSa ratios; it is canceled out, as both the ergodic and non-ergodic RVT PSa estimates are calculated with the same approach.

Aleatory aleatory variability of the two non-ergodic PSa GMMs is approximately 30 to 35% smaller than the aleatory variability of an ergodic PSa GMM.

Future studies should reevaluate the RVT and EAS models so that when combined they result in a PSa predictions consistent with PSa GMMs. Furthermore, the proposed non-ergodic GMMs were developed with a subset of the NGAWest2 database which was compiled in 2014. As larger data sets which include more recent and more frequent small magnitude events become available, the proposed models should be assessed and potentially expanded with additional non-ergodic terms. Similarly, $3D$ broadband numerical simulations or inferred intensity measurements from historical earthquakes should be used to evaluate the efficacy of the proposed models.

3.6 Software and Resources

The RVT calculations were performed with the pyRVT library (Kottke, 2020) in the computer language Python (Van Rossum and Drake, 2009). The linear mixed-effects regressions were performed with the lme4 package (Bates et al., 2015) in the statistical environment R (R Core Team, 2020).

Chapter 4

Selection of V_{S30} for Embedded Structures in the Seattle Region

4.1 Introduction

For seismic hazard analyses in the Pacific Northwest, it is common to use ground-motion models (GMMs) that use the time-averaged shear-wave velocity measured from the surface to a depth of 30m, V_{S30}, as the main site parameter. The V_{S30} is not a fundamental physical property that controls site amplification, but in natural soil deposits, the V_{S30} is often correlated with the deeper V_S profile that is the fundamental physical property controlling site amplification. The V_{S30} parameter works as a predictor of site amplification due to the correlation of V_{S30} and the site $V_S(z)$ profile. The correlation between V_{S30} and $V_S(z)$ implicit in GMMs depends on the empirical ground-motion data set used to develop the V_{S30} dependence of the site amplification in the GMM.

For embedded structures, there is controversy about where the V_{S30} should be measured: should it be measured from the surface to a depth of 30 m, from the embedment depth to 30 m below the embedment, or something in between? If the V_{S30} was a physical property that directly caused the site amplification, then using the V_{S30} measured from the embedment depth would be appropriate, but because the site amplification is due to the full V_S profile and not just the time-averaged shear-wave velocity measured over 30 m, this is not the case. Furthermore, because the empirical correlation between the V_{S30} and the site amplification in GMMs is based on the surface V_{S30}, the V_{S30} measured from the embedment depth may not have the same correlation with the site amplification as represented by the V_{S30} in the GMMs. For example, if a structure is embedded 50 m, this does not change the deeper V_s profile that controls the site amplification. Removing the top 50 m from the V_s profile changes the site amplification for periods with wavelengths that are affected by the top 50 m, but it does not change the deeper part of the profile that controls long-period site amplification. Using a V_{S30} measured from the embedment depth to 30 m below the embedment depth as input into the NGA-West2 GMMs will, in general, be inconsistent with the V_{S30} scaling in the GMMs.

In addition to the embedment issue, there is a second, more significant issue to consider for application of GMMs to the Seattle region: the correlation between the V_{S30} and the deeper V_S profile may be different in Seattle than it is in the regions that dominate the empirical ground-motion data sets used to develop the V_{S30} scaling in the GMM. For example, the NGA-W2 GMMs that are commonly used in Seattle to model the ground motion from crustal earthquakes are dominated by data from California, so the correlation between the site amplification and the V_{S30} in the NGA-W2 GMMs represents the depositional environment in California. This correlation may not be applicable to Seattle.

Figure 4.1 compares the V_s profiles for California and Seattle for a V_{S30} of 300 m/s. This figure shows that while the Seattle profile and California profile have the same V_{S30} values, the Seattle profile has a much weaker velocity gradient than the California profile. Some of the NGA-West2 GMMs include an additional site term that captures the depth to rock ($Z_{1.0}$ or $Z_{2.5}$). Including the region specific $Z_{1.0}$ and $Z_{2.5}$ terms will help to account for the difference in site amplification for these two profiles.

For subduction zone earthquakes, the empirical GMMs commonly used in the Seattle region are based on ground-motion data that is mainly from Japan. As a result, the correlation between the site amplification and the V_{S30} in the subduction GMMs represents the depositional environment in Japan. Figure 4.1 also compares the V_s profiles for Japan and Seattle for a V_{S30} of 300 m/s. Again, the Seattle profile has a much weaker velocity gradient than the Japan profile.

4.2 Simplified Approach

The current approach used in seismic hazard studies in the Seattle region uses GMMs with a site-specific V_{S30} value and basin amplification factors (BAF) to account for the basin effects. For crustal earthquakes, the $Z_{2.5}$-based basin factors from the Campbell and Bozorgnia (2014) GMM are used (eq 4.1) where $\bar{Z}_{2.5}(V_{S30})$ is the average $Z_{2.5}$ for a given V_{S30} for sites in California (called the default $Z_{2.5}$ value). For subduction earthquakes, the simulation-based basin factors from the M9 simulations (Wirth et al., 2018) are used (eq 4.2). The period dependence of the basin factors for these two models are shown in Figure 4.2.

$$BAF_{crustal} = \frac{Sa_{CB14}(M, R, V_{S30}, Z_{2.5site})}{Sa_{CB14}(M, R, V_{S30}, \bar{Z}_{2.5}(V_{S30}))} \tag{4.1}$$

$$BAF_{sub}(T) = min(T/2 + 1, 2) \tag{4.2}$$

With these basin factors, the median ground motion is given by

$$ln(SA_{site}(M, R, V_{S30}) = ln\left(SA_{med}\left(M, R, V_{S30}, \bar{Z}_{2.5}(V_{S30})\right)\right) + ln\left(BAF(Z_{2.5}\right) \tag{4.3}$$

The issue being considered in the study is what value of V_{S30} should be used in the GMM for embedded structures? In particular, how can the effect of the embedment be included in a simple manner? The simple approach being considered is to choose the depth at which to

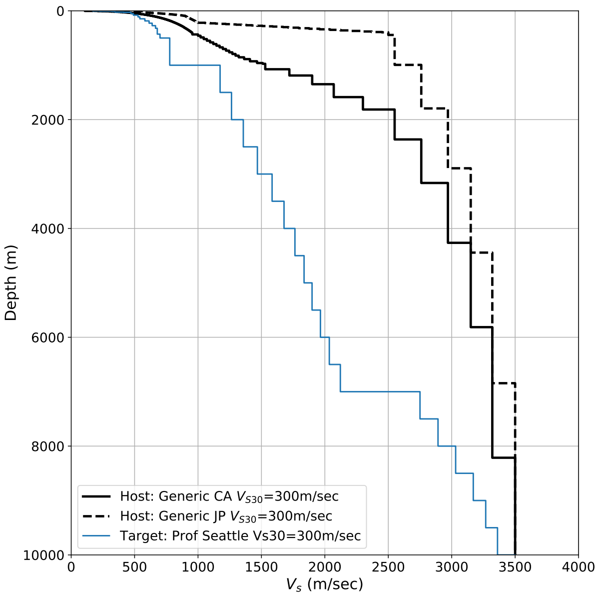


Figure 4.1: Comparison of a site-specific Seattle profile for a deep basin site ($V_{S30} = 300\,m/s$ and $Z_{2.5} = 7\,km$) with a representative V_S profile for California ($V_{S30} = 300\,m/s$ and $Z_{2.5} = 1.8\,km$), and a representative V_S profile for Japan ($V_{S30} = 300\,m/s$ and $Z_{2.5} = 0.4\,km$)

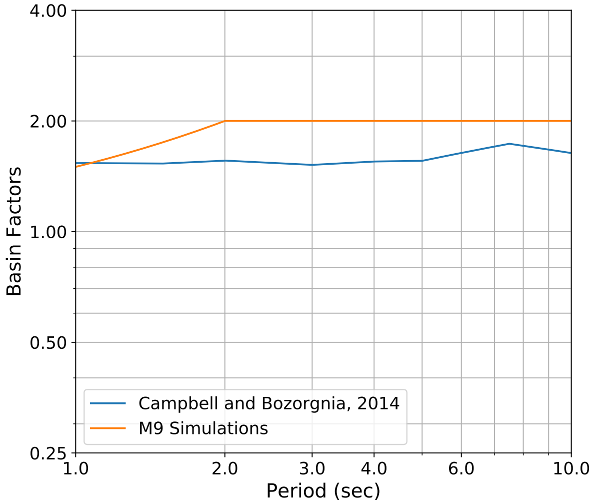


Figure 4.2: Comparison basin factors applied to shallow crustal GMMs and subduction zone GMMs for the Seattle region.

measure the V_{S30} and use that V_{S30} value in the GMM. As discussed in the introduction, the V_{S30} scaling in the GMM is based on the surface V_{S30} and may not apply to V_{S30} measured at depth and, due to different depositional environments, the V_{S30} scaling may also be different for the site region (Seattle) than for the region for which the site factors in the GMM were derived (e.g., California or Japan).

To keep the approach simple, we consider how to choose a V_{S30} value to plug into the scaling in the GMM, which when combined with the Seattle basin terms, will lead to a reasonable total site factor for the Seattle region. The selected V_{S30} value should be related to something that can be measured at the site. Three alternative approaches for choosing the V_{S30} are considered: (1) measure V_{S30} from the surface, (2) measure V_{S30} starting at one-half of the embedment depth, and (3) measure V_{S30} starting at the full embedment depth.

4.3 GMM-to-Site Scale Factors

To address the question of the appropriate V_{S30} value to use in the GMM for embedded structures in Seattle, we need to address both the differences in the $V_S(z)$ profiles due to the different regions and the effect of embedment. To develop the scale factors between the ground motion from the GMM and the ground-motion for Seattle site conditions, the same input motion is propagated through a representative 1-D $V_S(z)$ profile for the GMM and a site-specific 1-D $V_S(z)$ profile for Seattle. To be consistent with the current approach used for seismic hazard in Seattle, the basin factors shown in Figure 4.2 are applied to the ground motion computed using the GMM $V_S(z)$ profile with default basin depths. We then compute the ratio of the outcropping ground motions at the embedment depth from the Seattle profile to the surface motion with basin factors from the GMM profile.

$$Ratio(V_{S30_x}, V_{S30}) = \frac{SA_{Site-VS-embed}(V_{S30})}{SA_{GMM-VS}(V_{S30_x}, \bar{Z}_{2.5}(V_{S30_x}) \times BAF} \tag{4.4}$$

where $SA_{site-VS-embed}(V_{S30})$ is the outcropping motion from the Seattle site-specific profile for a given surface V_{S30} and $SA_{GMM-VS}(V_{S30_x}, \bar{Z}_{2.5}(V_{S30_x}))$ is the surface ground motion from the GMM profile for V_{S30_x} with the default $Z_{2.5}$. The V_{S30_x} is the V_{S30} measured from depth x to depth $x + 30\ m$.

If the ratio in eq 4.4 is near unity, then using V_{S30_x} as input into the GMM will lead to ground motions similar to the outcropping motion at the embedment depth after accounting for (1) the differences in the V_S profile for the site and the V_S profile for the GMM, (2) the differences between the assumed basin factors and the site-specific $V_S(z)$ profile, and (3) the effect of embedment.

The V_{S30_x} value (measured at the surface, one-half of embedment depth, or full embedment depth) that has a ratio closest to unity is selected as the best approach for capturing both the effects of embedment and the differences in the V_S profiles for the site and the GMM.

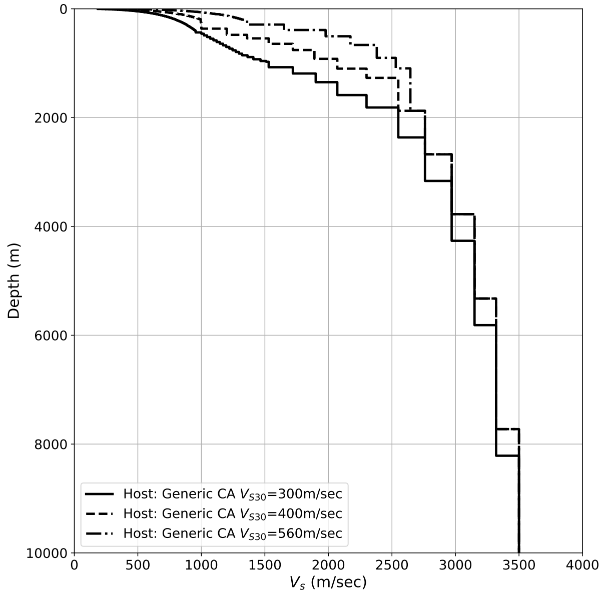


Figure 4.3: Representative profiles for the NGA-W2 GMM (California)

4.4 V_S profiles for GMMs

For the crustal GMMs, the representative $V_S(z)$ for a given V_{S30} and the default $Z_{2.5}$ is developed using the Kamai et al. (2014) California profiles for shallow depths ($< 200\ m$), and the Kamai et al. (2013) profile by geologic unit for the deeper depths. The deep profile was modified so that it was consistent with the default $Z_{2.5}$ values. Representative profiles for the NGA-W2 GMM are developed for three V_{S30} values and are plotted in Figure 4.3. For other V_{S30} values, the spectral values are interpolated from the spectral values for these three V_{S30} cases.

For the subduction GMMs, the representative $V_S(z)$ for a given V_{S30} is developed using the Kamai et al. (2014) Japan profiles for shallow depths ($< 200\ m$), and the deeper part was developed based on NGA-W2 relationships between $Z_{1.0}$, $Z_{2.5}$ and V_{S30} for sites in Japan (Chiou and Youngs, 2014; Campbell and Bozorgnia, 2014). Representative profiles for Japan are developed for the same three V_{S30} values used for the crustal GMMs and are plotted in Figure 4.4. The Japanese profiles have a much smaller $Z_{2.5}$ for soil sites than the California profiles with the same V_{S30} values.

4.5 Site Response Method

The site response is computed in the frequency domain using the Python package pySRA (Kottke, 2019). For the application to tall buildings, the scale factors at large periods

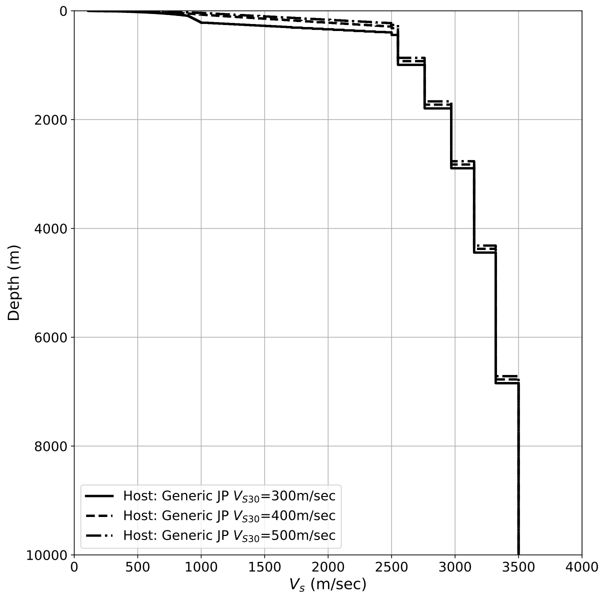


Figure 4.4: Representative profiles for the subduction GMMs (Japan)

($T > 2$ *sec*) are of primary interest. Therefore, a linear site-response analyses is sufficient.

The half-space velocity for all cases was set to 3500 m/sec. The density of each layer was estimated as a function of V_S using equation 4.5. For layers with V_S less than 1000 m/sec, the damping was estimated as function of V_S with equation 4.6. For layers with V_S greater than 1000 m/sec, the damping was set to zero because the effect of damping in the rock is already represented in the input motion by using a kappa value applicable to surface ground motion on rock sites.

$$\rho = 1.742 + 0.2875 V_S \tag{4.5}$$

$$Damp = 0.0396\ e^{(-0.00103 V_S)} \tag{4.6}$$

To smooth the site amplification ratio, both the Seattle profile and the GMM profile were randomized following the Toro (1995) method. The parameters for the randomization are summarized in Table 4.1. Figure 4.5 shows an example of the randomized profiles for $V_{S30} = 400\ m/sec$.

The input ground motion, shown in Figure 4.6, was based on the point-source model for a M 7.2 earthquake at a distance of 5 km and a kappa of 0.035 sec, typical for the western U.S. (Campbell, 2003). The input motion was inserted at the half-space of the profiles. The selection of the earthquake parameters for the input motion is not critical as the site response analysis is linear and we are interested in the long-period response. That is, the amplification

Table 4.1: Parameters for Randomization of Velocity Profiles

Parameter	Ref. Profiles	Target Profiles
σ	0.15	0.1
ρ_0	0.97	0.97
ρ_{200}	1.0	1.0
Δ	3.8	3.8
h_0	0.0	0.0
b	0.293	0.293

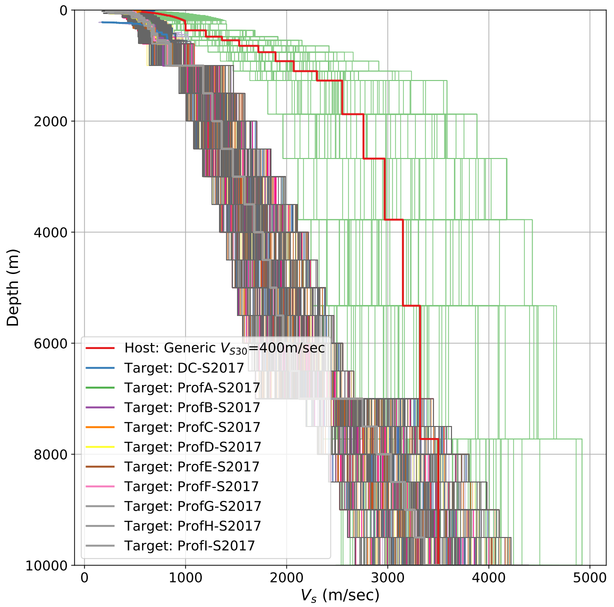


Figure 4.5: Randomized reference and target profile, the reference profiles correspond of $V_{S30} = 400\ m/sec$

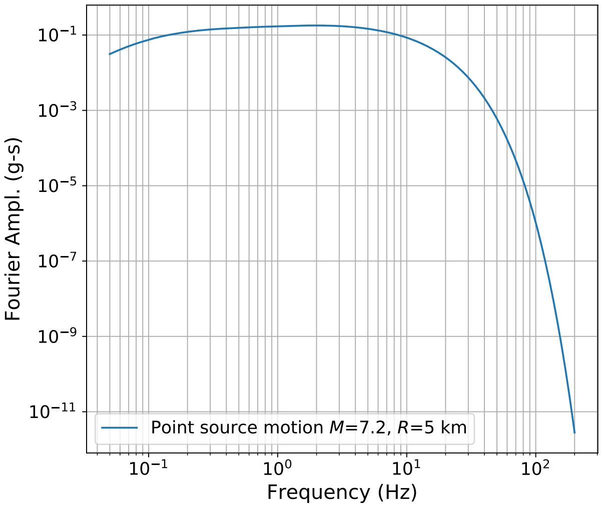


Figure 4.6: Fourier Amplitude spectrum of input motion

between the input motion and the surface motion will be similar if a different input motion was used.

The site response was conducted for GMM profiles corresponding to V_{S30} values of 300 m/s, 400 m/s, and 560 m/s for California and to V_{S30} values of 300 m/s, 400 m/s, and 500 m/s for Japan representative profiles. An example of the surface spectra for the suite of randomized profiles is shown in Figure 4.7.

4.6 Seattle Profiles

The site response was also conducted for the three sets of $V_S(z)$ profiles for Seattle shown in Figure 4.8. The first set is based on shallow $V_S(z)$ profiles from a project in the Seattle region with an average $\bar{V}_{S30} = 420$ m/s that were merged with Stephenson et al. (2017) regional Seattle profile for the deep part; the depth to $V_S = 2500$ m/sec in the deep part of the Seattle basin is approximately 7 km. The embedment depth of the structure was assumed to be 50 m; thus the three depths for measuring the V_{S30} are 0 m, 25 m, and 50 m. The average V_{S30} values at these three depths are listed in Table 4.2. For Set 1, the V_{S30} only increases from 420 to 490 m/s from the surface to the 50 m embedment depth.

The second Seattle set is composed of a single profile which was developed to have a larger difference between the V_{S30} measured from the surface and the V_{S30_x} measured from the embedment depth than the difference for the profiles in Set 1 (Figure 4.8). Starting with the average $V_s(z)$ of the Seattle Set 1 profiles, the velocities in the top 20 m were reduced by a constant shift to give a surface V_{S30} of 300 m/s. For a smooth transition in the $V_S(z)$ profile, the V_S in the next layer below 20 m (layer from 23 to 28 m depth) was reduced by half of the

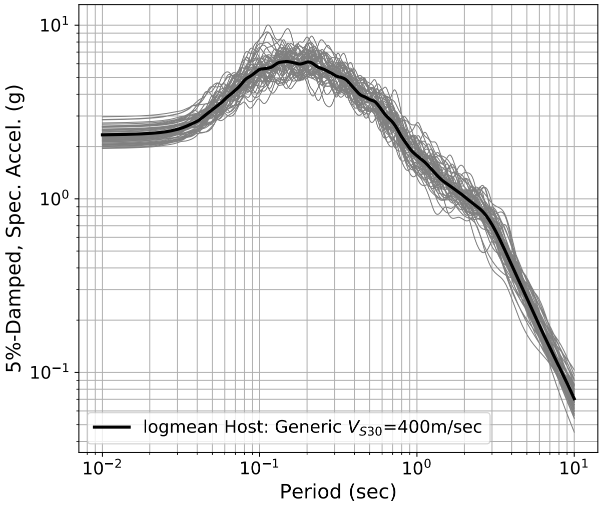


Figure 4.7: Example of the surface spectra from the profile randomization for the California profile with $V_{S30} = 400\ m/sec$

shift in the shallow layers. The V_{S30_x} at the embedment depth of 50 m is unchanged from the average of Set 1. This profile represents a thin soil site which leads to much larger long-period motion if the V_{S30} scaling in the GMM is used, but has little effect on the long-period ground motion from the 1-D site response due to the small thickness of the softer soil relative to the wavelengths of the long-period ground motion. The results for the second set will be more sensitive to the selection of the depth for computing the V_{S30} than for the Set 1.

The $V_S(z)$ profile comprising the third set represents a thicker soil site, still with $V_{S30} = 300\ m/s$ (Figure 4.8). For this case, the V_S in the shallow layers down to 40 m depth was reduced by fixed amount so that the $V_{S30} = 300\ m/s$. As with the profile of Set 2, the V_S for the next layer below 40m depth (layer from 52 m to 80 m depth) was reduced by one-half of the shift in the shallower layers. In this case, the softer soils extend to depths below the embedment depth, so the V_{S30_x} is changed for both one-half the embedment depth and the full embedment depth. This will lead to larger differences between the amplification using the GMM with the V_{S30_x} measured at either the full embedment depth or at one-half of the embedment depth.

4.7 Results

For the site response analyses with the GMM profiles, the response spectra were computed at the surface. For the site response analyses with the Seattle profiles, the response spectra were computed as the outcropping motion at the embedment depth. The ratio of the outcropping response spectra from the Seattle profile to surface response spectra for the GMM profile

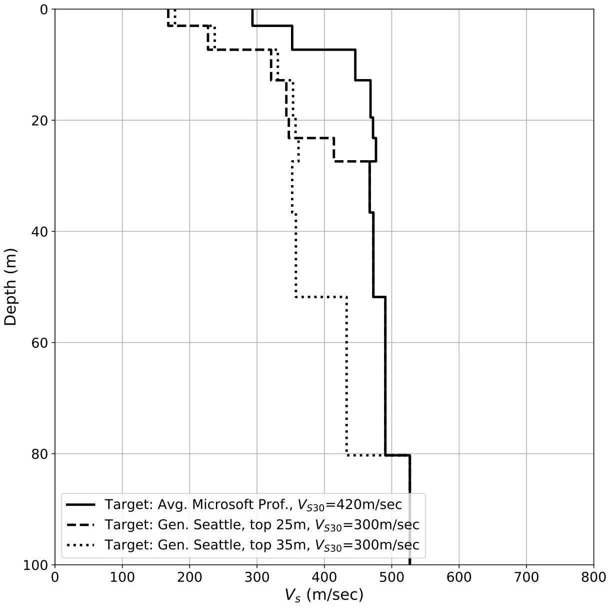


Figure 4.8: Three example velocity profile used for Seattle.

Table 4.2: Seattle Profiles Considered in this Analysis

Parameter	Set 1	Set 2	Set 3
V_{S30} at surface	$420m/s$	$300m/s$	$300m/s$
V_{S30_x} at 25m	$470m/s$	$470m/s$	$360m/s$
V_{S30_x} at 50m	$490m/s$	$490m/s$	$430m/s$
$Z_{2.5}$	$7km$	$7km$	$7km$

scaled by the basin factors (eq. 4.1 and eq. 4.2) were computed for the three alternative approaches for the V_{S30_x} used in the GMMs. The ratio is computed for the three Seattle sets using an embedment depth of 50 m.

4.7.1 Crustal GMM

For Set 1 (surface $V_{S30} = 420\ m/s$), the ratios shown in Figure 4.9 are similar using the V_{S30} measured from the surface or measured at the embedment depth. This lack of sensitivity occurs because the the V_S does not increase rapidly in the top 50 m for this profile (Figure 4.8).

For Set 2 (thin soil with surface $V_{S30} = 420\ m/s$), the ratios in Figure 4.10 show that using the surface V_{S30} leads to ratios much greater than unity for $T < 6$ sec compared to the using either the V_{S30_x} at the full embedment depth or one-half of the embedment depth.

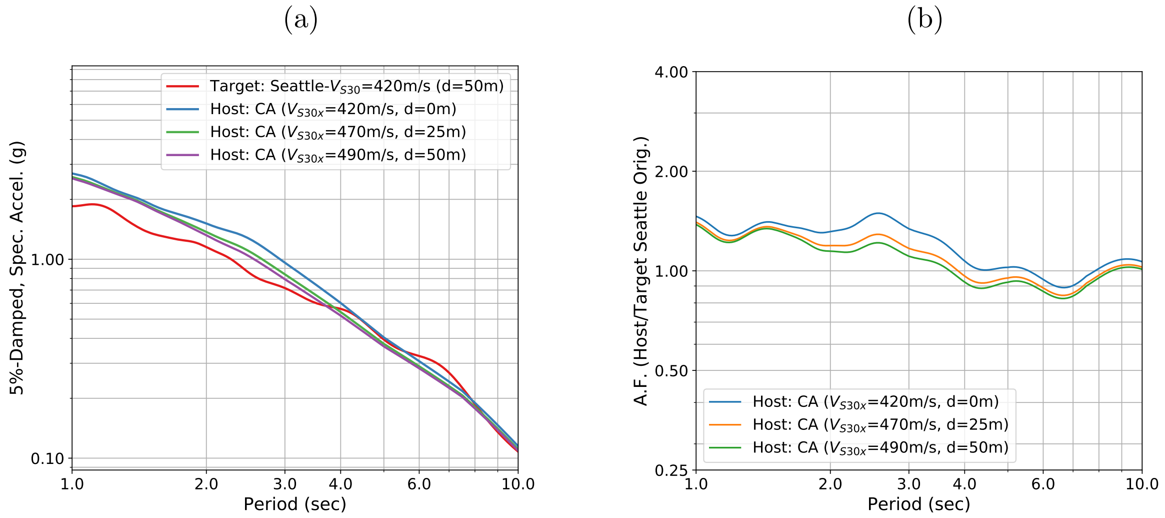


Figure 4.9: Comparison of spectra for the crustal GMM for the Seattle Profiles - Set 1 ($V_{S30} = 420\ m/sec$)

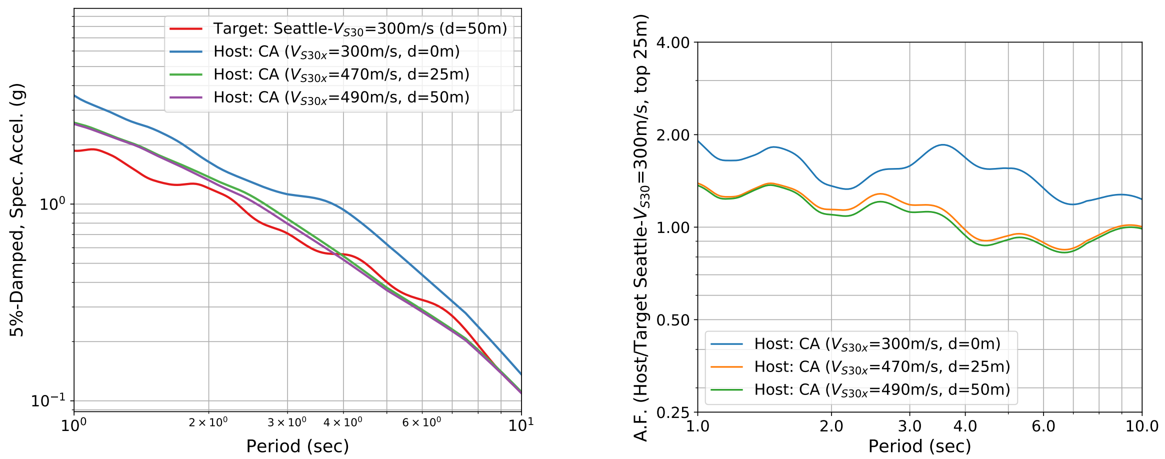
This indicates that using the surface V_{S30} will overestimate the site effects for $T < 6$ sec in the Seattle region. At periods greater than 6 seconds, using the V_{S30_x} measured at either embedment depth underestimates the site amplification by 10 to 20%. For this profile, the ratios for V_{S30_x} measured at one-half the embedment depth or the full embedment depth are similar because the V_{S30_x} did not change much if measured from 25 m depth or from 50 m depth. Recall that for Set 2, the main change to the $V_S(z)$ was in the top 20 m which is above the one-half embedment depth (Figure 4.8). Once the V_{S30} is measured below the softer soil, there is not a large effect on the results.

For the third set, we also used $V_{S30} = 300\ m/s$ but with the main change in the $V_s(z)$ profile extended to 40 m depth. The resulting spectra and ratios for Set 3 are shown in Figure 4.11. For this case, the V_{S30} measured from the full embedment depth is closer to unity than for the V_{S30} measured from one-half the embedment depth, except at periods of 6 to 8 seconds.

4.7.2 Subduction GMM

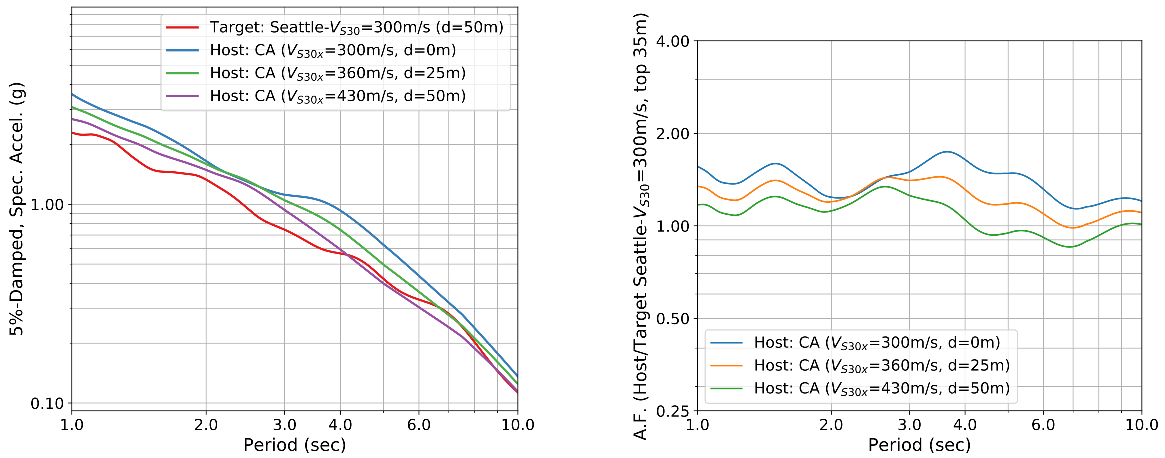
The same analysis was conducted for the subduction GMM using the Japanese profiles for the GMM. The subduction GMM results are similar to the crustal GMM results for Set 1. The ratios for Set 1 (Fig 4.12) are similar using the V_{S30} measured from the surface or measured at the embedment depth.

For Set 2, the ratio using the surface V_{S30} is larger than unity for $T < 4$ sec (Figure 4.13). Using either the V_{S30_x} at the full embedment depth or one-half of the embedment depth works well for $T < 4$ sec. For $T > 4$ sec, using the surface V_{S30} in the subduction GMM give



(a) (b)

Figure 4.10: Comparison of spectra for the crustal GMM for the Seattle Profile - Set 2 ($V_{S30} = 300\ m/sec$, thin soil)



(a) (b)

Figure 4.11: Comparison of spectra for the crustal GMM for the Seattle Profile - Set 3 ($V_{S30} = 300\ m/sec$, thicker soil)

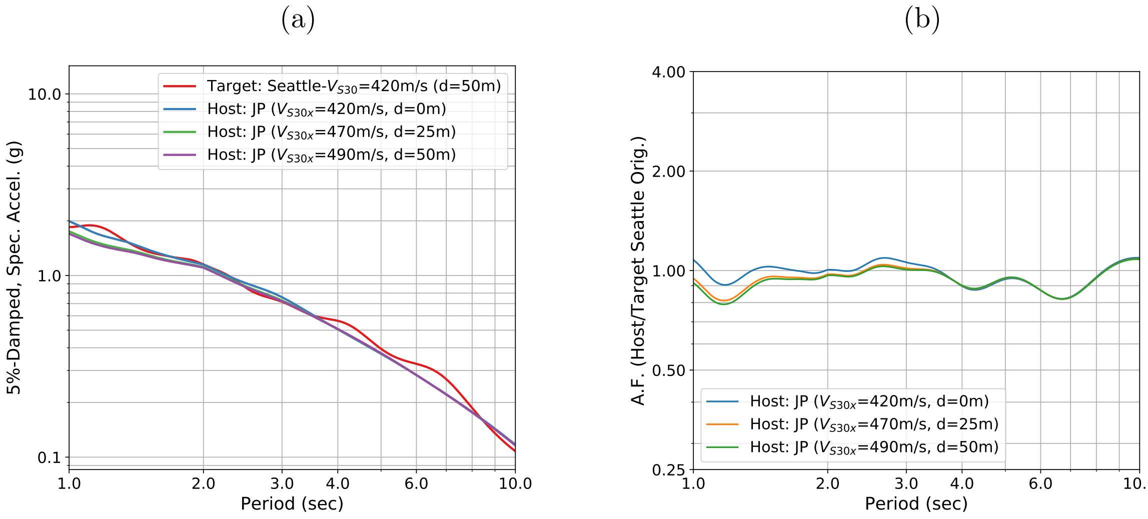


Figure 4.12: Comparison of spectra for the subduction GMM for the Seattle Profile - Set 1 ($V_{S30} = 420\ m/sec$)

a ratio closer to unity.

For Set 3, with the thicker soil layer, the results are similar to Set 2. The spectra and ratios for the three V_{S30} values used in the subduction GMM are shown in Figure 4.14. For $4 < T < 8$ sec, the ratio using the surface V_{S30} in the GMM overestimates the site amplification whereas the ratio using the embedded V_{S30_x} gives a ratio near unity. For $T > 4$ sec, the ratio using the surface V_{S30} in the GMM leads to ratios closer to unity than using the embedded V_{S30_x} in the subduction GMM.

4.8 Conclusions

The objective of this study was to provide a recommendation on the depth at which the V_{S30} should be measured for embedded tall buildings in the Seattle region for use in building code applications. The key issue is the empirical V_{S30} scaling models in both the crustal and subduction GMMs used for Seattle are not applicable to the Seattle region. In particular, the deep $V_S(z)$ profiles in Seattle have very different velocity gradients than in the representative profiles for the GMMs (Figure 4.1).

From the examples shown in this paper, for spectral periods less than 4 sec, using the V_{S30} measured at the full embedment depth as input to the GMM with the basin factors shown in Figure 4.2 is a better approximation of the ground motion in the Seattle region than using the V_{S30} measured at the surface or measured from the one-half embedment depth. For spectral periods of 4 to 8 sec, using the V_{S30} measured at the surface gives an overall better estimate of the site amplification than using the V_{S30_x} measured at either embedment depth.

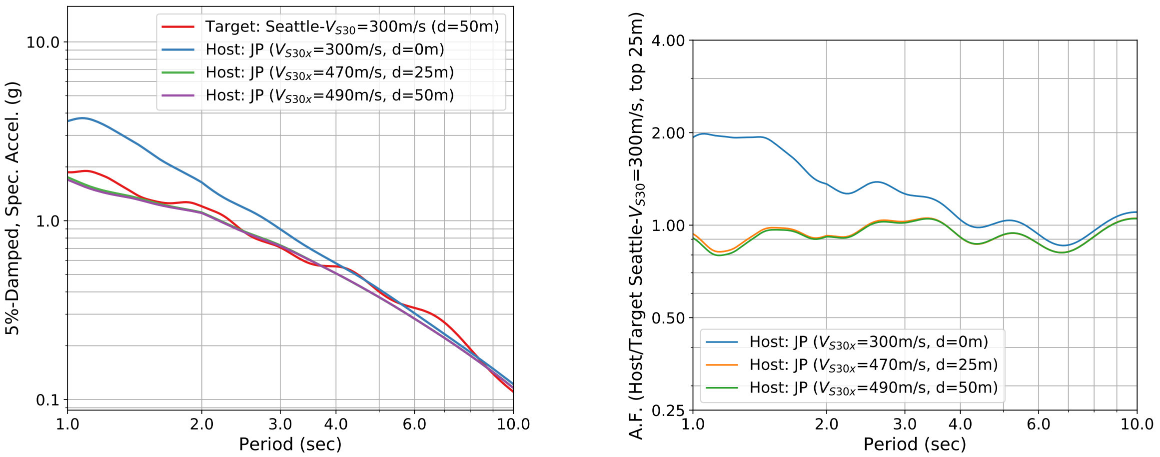

(a) (b)

Figure 4.13: Comparison of spectra for the subduction GMM for the Seattle Profile - Set 2 ($V_{S30} = 300\ m/sec$, thin soil)

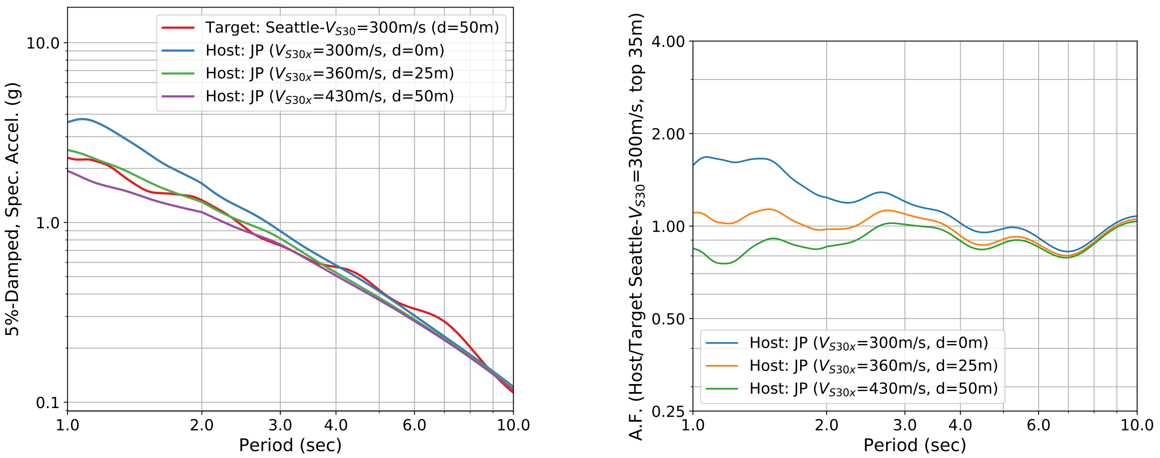

(a) (b)

Figure 4.14: Comparison of spectra for the subduction GMM for the Seattle Profile - Set 3 ($V_{S30} = 300\ m/sec$, thicker soil)

Based on these site response calculations, the under-prediction of the site amplification for $4 < T < 8$ sec using the V_{S30_x} measured at the embedment depth is not large (between 10 and 20%).

There is not a single depth for measuring V_{S30_x} that does a good job capturing the site amplification for all periods from 1 to 10 sec. To keep the approach simple, one option would be to use the V_{S30_x} measured at the full embedment depth, but increase the basin factors for $T > 4$ sec by 10 to 20 %.

4.8.1 Limitations

There are two main limitations of these results. The first limitation is the small number of profiles considered for the Seattle region. Using Set 1 as the base case, we assumed that profiles with lower V_{S30} values would only differ from the base profile in the shallow depths ($< 50\ m$). If the deep part of a site-specific Seattle velocity profiles differs significantly from the base profile used in this study at depths greater than 50 m, then the conclusions from this study need to be reevaluated before being applied to such a site.

A second limitation is that the differences in the site amplification between the Seattle region and the GMMs is based on differences in the amplification using 1-D wave propagation. There are 3-D effects in both the Seattle region and in the empirical data used to develop the GMM. By only accounting for the differences in the 1-D site amplification, there is an implicit assumption that the 3-D effects in the Seattle region are similar to the 3-D effects in the empirical GMMs. If there are much stronger 3-D effects in the Seattle region as compared to California or Japan, then the conclusions from this study need to be reevaluated.

4.8.2 Alternative Simplified Approach

While we were asked to address selection of V_{S30} for embedded tall buildings, this is not just an issue of embedment, but rather an issue of the very different deep $V_S(z)$ profiles in the Seattle region as compared to the $V_S(z)$ profile that is implied by the V_{S30} scaling in the GMMs. In support of this argument, Figure 4.15 presents the mean outcrop response spectra extracted at 0, 25 and 50 m depths of the profiles in Seattle Set 1 which shows that the response spectra are unaffected by the embedment depth for periods greater than 1 *sec*. The same comparison was repeated with the other two sets which lead to the same conclusion but for the sake of space these results are not presented here. The difference between the deep $V_S(z)$ profiles for Seattle and the deep $V_S(z)$ profiles implied for the GMMs will affect both embedded and non-embedded structures. That is, for a structure that is not embedded, using the V_{S30} measured from the surface will also not be appropriate for the site conditions in Seattle. Two alternative simplified approaches that account for the differences in $V_S(z)$ and are applicable to both embedded and non-embedded structures are: i) measuring the V_{S30} at some fixed depth for all structures in the Seattle region ii) using a fixed V_{S30} for structural periods greater than 1 *sec*. For the first approach, Figures 4.9 to 4.14 show that measuring the V_{S30} at a fixed depth of 25 m is a reasonable approach because

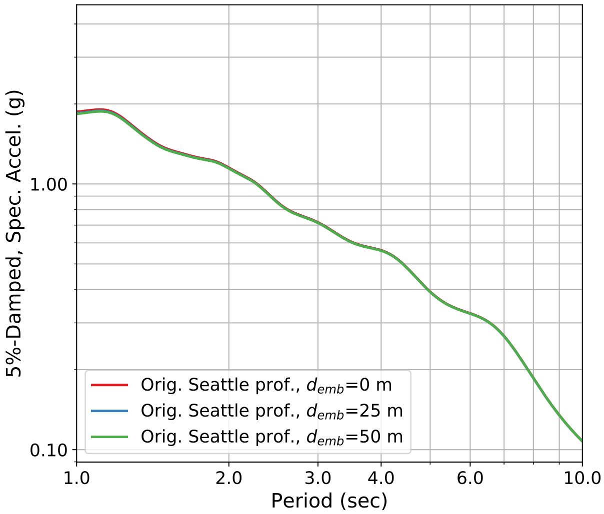


Figure 4.15: Outcrop spectra for the Seattle Profile - Set 1 extracted at 0, 25 and 50m depth.

it gives amplification ratios closest to unity for periods greater than 1 *sec*. For the second approach, Figures 4.16 to 4.21 compare the Seattle target outcrop spectra at the embedment depth with the surface spectra of the crustal and subduction profiles for V_{S30} equal to 350, 450 and 550 m/s. These results suggest that using a fixed V_{s30} of 450 m/sec is reasonable as the corresponding amplification ratios are closest to unity regardless of the target Seattle profiles and tectonic environment of the GMMs. In both simplified approaches the point is that for shallow embedment depths (up to 50 m), the long-period ground motion will not be very sensitive to the building-specific embedment depth, so a single depth or V_{S30} may be a reasonable simplified approach to account for the differences in the deep part of $V_S(z)$ profiles. The second option is more advantageous over the first one as it does not require the estimation of the V_{s30_x} at a depth of 25 to 55 m.

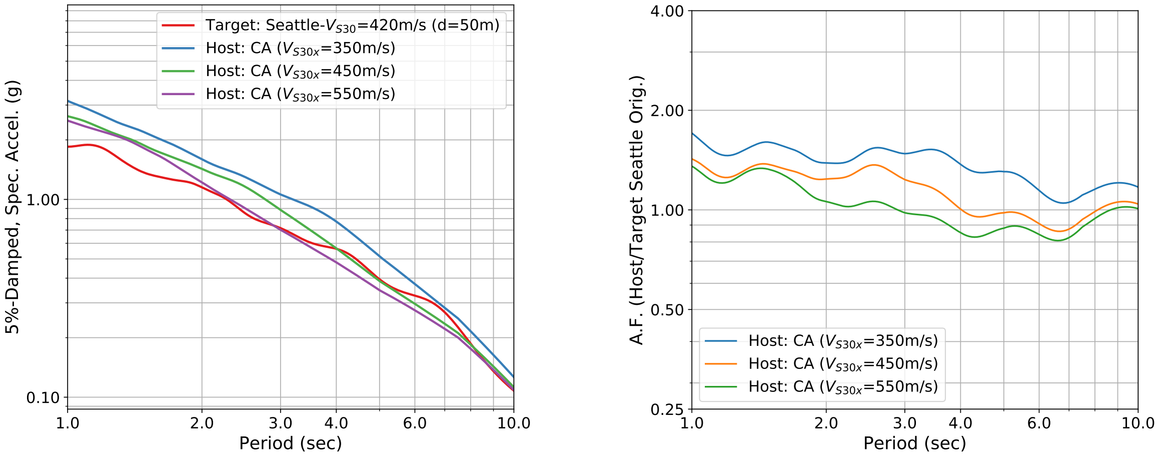

(a) (b)

Figure 4.16: Comparison of spectra for the crustal GMM for $V_{S30} = 350$, 450, and 550 m/sec with spectrum for the Seattle Profile - Set 1 ($V_{S30} = 420\ m/sec$)

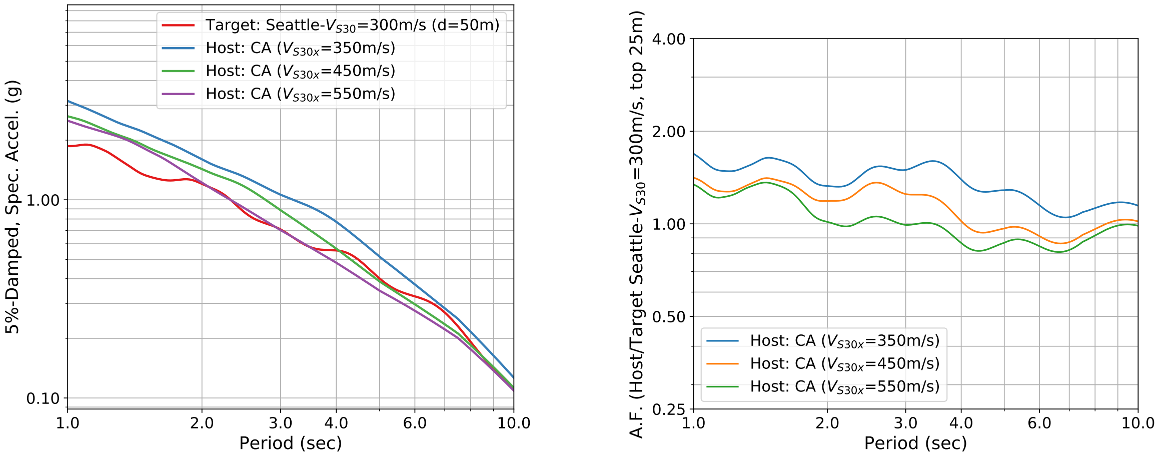

(a) (b)

Figure 4.17: Comparison of spectra for the crustal GMM for $V_{S30} = 350$, 450, and 550 m/sec with spectrum for the Seattle Profile - Set 2 ($V_{S30} = 300\ m/sec$, thin soil)

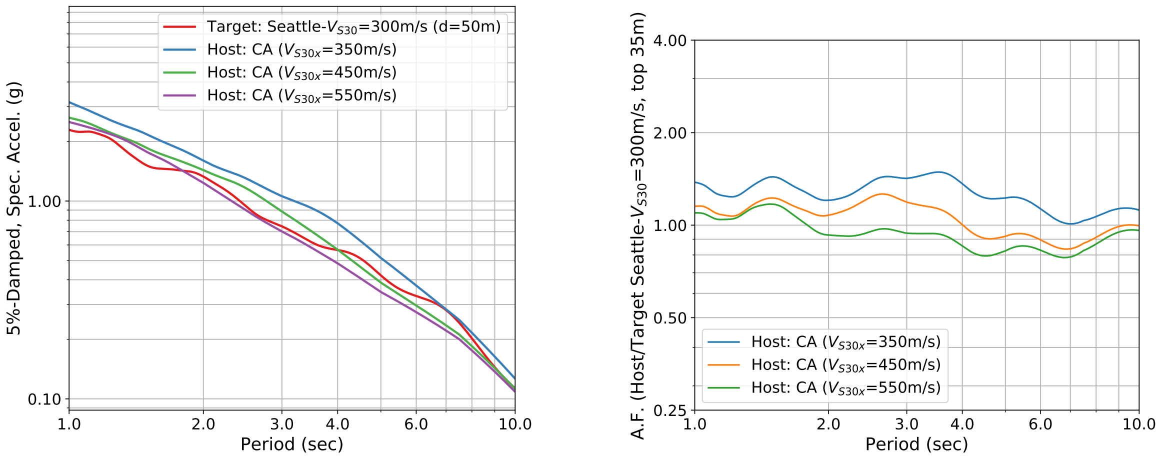

(a) (b)

Figure 4.18: Comparison of spectra for the crustal GMM for V_{S30} = 350, 450, and 550 m/sec with spectrum for the Seattle Profile - Set 3 ($V_{S30} = 300\,m/sec$, thicker soil)

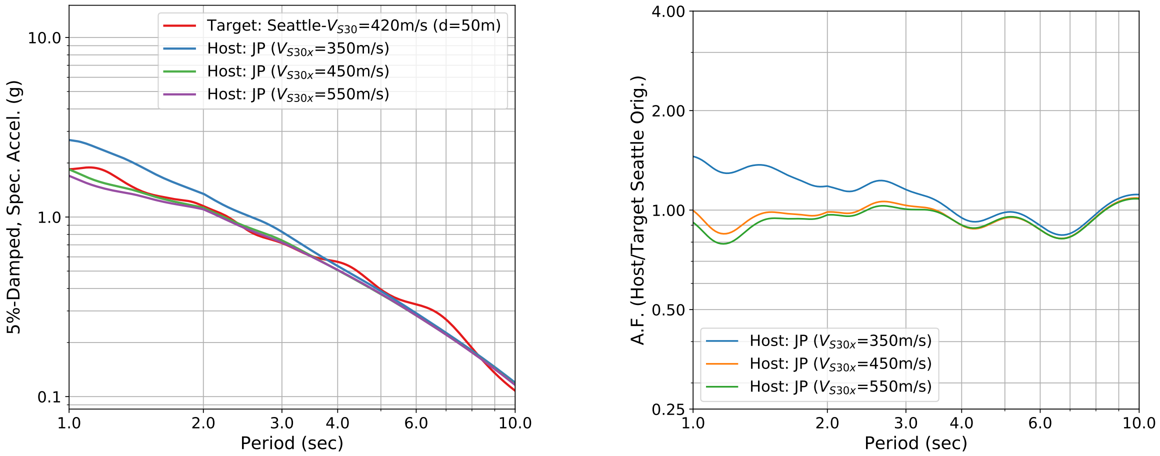

(a) (b)

Figure 4.19: Comparison of spectra for the subduction GMM for V_{S30} = 350, 450, and 550 m/sec with spectrum for the Seattle Profile - Set 1 (V_{S30} = 420 m/sec)

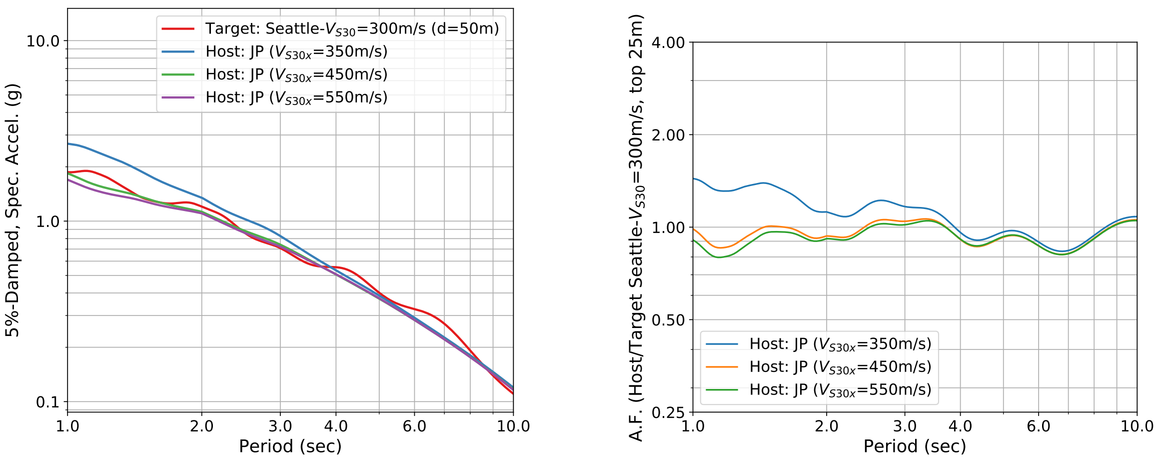

(a) (b)

Figure 4.20: Comparison of spectra for the subduction GMM for V_{S30} = 350, 450, and 550 m/sec with spectrum for the Seattle Profile - Set 2 ($V_{S30} = 300$ m/sec, thin soil)

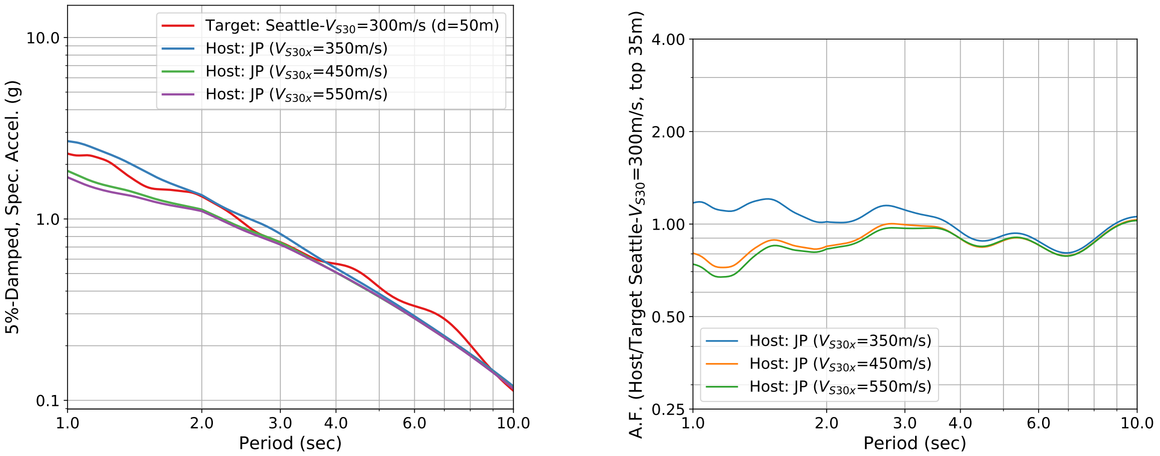

(a) (b)

Figure 4.21: Comparison of spectra for the subduction GMM for V_{S30} = 350, 450, and 550 m/sec with spectrum for the Seattle Profile - Set 3 ($V_{S30} = 300$ m/sec, thicker soil)

Chapter 5

Generation of Surface Slip Profiles in the Wavenumber Domain

The contents of this chapter are primarily from a journal article submitted to the Bulletin of the Seismological Society of America by Lavrentiadis, G., and Abrahamson, N.A. entitled: "Generation of Surface Slip Profiles in the Wavenumber Domain", which has been published.

5.1 Introduction

Most seismic hazard studies evaluate the ground motion at a site through a deterministic or probabilistic seismic hazard assessment (PSHA). For sites located closest to faults, in addition to ground-motion hazard, there is also surface-rupture hazard. The common approach to mitigate the surface-rupture hazard is relocation; however, for types of structures such as lifelines, railroads and bridges, crossing an active fault may be unavoidable, and so the estimation of rupture hazard necessary (ALA, 2005; CALTRANS, 2013). Over the last decade, there has been much less effort toward improving models and methods for surface-rupture hazard as compared to the efforts to improve models and methods for ground-motion hazard. To address this disparity, an international workshop entitled "Fault Displacement Hazard Analysis Workshop" was held in December 2016 at the U.S. Geological Survey (USGS) in Menlo Park, California. The workshop was sponsored by the California Geological Survey (CGS), the Institut de Radioprotection et de Surete Nucleaire (IRSN), the Instituto National di Geofisica e Vulcanologia (INGV), the Pacific Earthquake Engineering Research Center (PEER), and the USGS. In addition to reviewing the available data sets, the current state of fault rupture models, and the methodology used for probabilistic fault displacement hazard analysis (PFDHA), the workshop started the planning of a coordinated multi-year project that would significantly move the PFDHA practice forward, similar to the approaches used for coordinated studies for ground-motion hazard (e.g. Bozorgnia et al. 2014).

Following the recommendations from the 2016 workshop, a PFDHA project was initiated at the B. John Garrick Institute for the Risk Sciences located at UCLA. The UCLA PFDHA project includes five main tasks: (1) develop a framework for describing the spatial distributions of primary and secondary surface ruptures and folding for use in PFDHA; (2) develop an updated peer-reviewed data base of tectonic ground deformation from past earthquakes including primary and secondary surface ruptures and folding; (3) develop new methods to characterize the along-strike variability of surface slip profiles; (4) apply these new statistical methods to develop new probabilistic models of distributions of amplitudes and locations of surface ruptures and tectonic surface deformation as a function of the rupture geometry, style of faulting, and complexity of the fault, and (5) develop a new open-source PFDHA computer program that incorporates these new models.

Parallel to the UCLA PFDHA project, IRSN is conducting the SURE project (Baize et al., 2016) to develop a community-sourced, worldwide, unified database of surface-rupturing earthquakes from a wide range of tectonic environments. The SURE project began in 2015 and parts of the this data set are being incorporated into the UCLA PFDHA project data set.

In this paper, we address Task 3 in the UCLA PFDHA project and develop a model for the along-strike variability of the slip profile that has an underlying physical basis related to the shape of the wavenumber spectrum of surface slip profiles, rather than simple statistical modeling of sparse empirical data used in the currently available models. We still need to estimate the parameters of the wavenumber spectrum using empirical data, but we optimize the approach for the non-uniform and potentially biased sampling of surface fault ruptures that are typical of the available empirical data sets from past earthquakes.

5.2 PFDHA Methodology

In its simplest form, the PFDHA methodology (Youngs et al., 2003) is similar to the traditional PSHA methodology used for ground-motion hazard (McGuire, 2004) with two main differences: an additional term for the probability of surface rupture at the site is included and the ground-motion model is replaced with the surface-displacement model:

$$P(D > z|x) = N(M_{min}) \int_{M_{min}}^{M_{max}} f_m(M) P(SR_{\ at\ site}|M) P(D > z|SR\, at\, site, M, x) dM \quad (5.1)$$

where $N(M_{min})$ is the rate of earthquakes above M_{min} on the controlling fault, $f_m(M)$ is the probability density function for earthquake magnitudes on the fault, $P(SR_{\ at\ site}|M)$ is the conditional probability of surface rupture (SR) occurring at the site given the earthquake magnitude, $P(D > z|SR_{\ at\ site}, M, x)$ is the conditional probability of the surface fault displacement, D, exceeding a test value z, given that there is surface rupture at the site from a magnitude M earthquake, and x is the distance from the site to the end of the rupture.

The $P(D > z|SR_{\ at\ site}, M, x)$ term is typically separated into the distribution for the average displacement over the full rupture length (AD) and the distribution for the normalized displacement along strike (D_{norm}):

$$\begin{aligned} P(D > z|SR\,at\,site, M, x) = \int_{AD}\int_{D_{norm}} & f_{AD}\left(AD|M\right) f_{ND}(D_{norm}|x) \\ & P(D > z|AD, D_{norm})\; dD_{norm}\; dAD \end{aligned} \tag{5.2}$$

where f_{AD} is the probability density function of the average displacement, and f_{ND} is the probability density function of the normalized surface displacement. These two terms are described below.

The average displacement is typically estimated using empirical models that are a function of earthquake magnitude and source type. For example, Wells and Coppersmith 1994 provide empirical relations for the magnitude scaling of $AD(M)$ for different styles of faulting. The aleatory variability of AD is typically assumed to be lognormally distributed.

The normalized displacement is given by:

$$D_{norm}(x/SRL) = \frac{D(x)}{AD \times\ T(x/SRL)} \tag{5.3}$$

where SRL is the surface rupture length, x is the distance from the end of the rupture, $D(x)$ is the displacement at location x along the rupture, and $T(x/SRL)$ is a function that describes how the mean slip is tapered toward the ends of the rupture.

Multiple studies have found that the mean slip profiles are reduced near the ends of the rupture, but different models have been proposed for the shape of $T(x/SRL)$. The models can be categorized in terms of their assumptions about symmetry at the half rupture length and their generic shape (e.g. elliptic or triangular). Hemphill-Haley and Weldon (1999) developed a symmetrical elliptical model for mean shape of the slip profiles. Manighetti et al. (2005) found that most slip profiles are in better agreement with asymmetric triangular distributions than with symmetric elliptical models. Wesnousky (2008) found that asymmetric models matched the slip profiles better than symmetric distributions. Petersen et al. (2011) proposed a suite of alternative models for the along-strike taper using elliptical, piece-wise billinear, and quadratic forms for $T(x/SRL)$. These alternative models for the shape of $T(x/SRL)$ are based on analyses of empirical observations but without a underlying physical model to constrain the extrapolation of the model to larger magnitude earthquakes.

An important issue is the use of the normalized distance along strike, x/SRL. To combine the surface rupture data from earthquakes with different SRL into a single regression analysis, previous studies have homogenized the surface slip profiles by normalizing the along-strike distance by the surface-rupture length (x/SLR). This assumption implies that the shape of the taper at the end of the rupture scales linearly with SRL, which may not be appropriate. While using a normalized distance along strike is convenient, the physical basis for this normalization has not been evaluated.

The proposed wavenumber spectrum model for the slip profile developed in this study provides a model that is consistent with seismological modeling of slip on the rupture plane both from slip inversions and from dynamic rupture models. As an example, using dynamic simulations, Dieterich and Smith (2009) and Dunham et al. (2011) found that the slip variability on the fault plane is related to the fault roughness which translates to the slope of the wavenumber spectrum at the high-wavenumbers. In addition, the proposed wavenumber model does not use the normalized distance along strike and it does not require choosing a form for the taper along strike. These two features lead to an improved model for the shape of the mean slip along strike as compared to the purely empirical models currently used in PFDHA practice.

5.2.1 Normalized Displacement Models

The along-strike variability of the surface slip profile about the mean slip is given by the model used for $f_{ND}(D_{norm}|x)$. Alternative statistical distribution for $f_{ND}(D_{norm}|x)$ have been proposed. McCalpin (1998) found that the D_{norm} values followed a gamma distribution. In contrast, Petersen et al. (2011) assumed that the D_{norm} values followed a log-normal distribution. Moss and Ross (2011) compared the goodness-of-fit of normal, log-normal, gamma, and Weibull distributions to the D_{norm} values and showed that the latter two distributions are in better agreement with the observations.

The choice of the form of the probability distribution for $f_{ND}(D_{norm}|x)$ can have a significant impact on the slope of a hazard curve. The lognormal distribution has much fatter tails for large displacements which leads to much larger hazard at large return periods. The sensitivity of hazard to the choice of the form of $f_{ND}(D_{norm}|x)$ is shown later in the example application section (see Figure 5.15). The proposed wavenumber-spectrum method doesn't require assumptions about the distribution of D_{norm} in the space domain. Therefore, it can provide an independent evaluation of the distribution of D_{norm}.

5.2.2 Spatial Correlation of Along Strike Variability

Typically, the effect of spatial correlation of the measurements of slip along strike has not been formally considered in the normalized slip models, partly due to the widely separated measurements that are common for data sets from older earthquakes. Modern technologies can resolve fault displacement profiles at much higher spatial resolution and provide data that can used to evaluate the spatial correlation of the along-strike variability of slip; examples of such high-resolution surveys are: Oskin et al. (2012), Nissen et al. (2014) and Wedmore et al. (2019), who used LIDAR to measure the co-seismic deformations of the 2010 El Mayor-Cucapah, 2008 Iwate-Miyagi, 2011 Fukushima-Hamadori and 2016 Norcia earthquakes; and Milliner et al. (2016), who measured the total surface deformations of 1992 Landers and 1999 Hector Mine earthquakes using subpixel image correlation. The study by Rockwell and Klinger (2013) illustrates the differences in spatial correlation between field-based measurements and slip estimates measured with hight-resolution techniques: using historic areal photography, they augmented the number of slip measurements for the 1940 Imperial Valley, CA rupture profile and observed that the new dense measurements suggested higher variability of slip along strike than the original field measurements.

This higher variability observed by these recent studies would be missed if simple linear interpolation was used to estimate the slip between widely separated points along historic ruptures. The proposed wavenumber spectrum model implicitly addresses the spatial correlation length of the variability of D_{norm} along strike. As will be shown later, the decay of the amplitude of the wavenumber spectrum at high wavenumbers is related to the correlation length.

5.3 Data Base

Because the UCLA PFDHA project rupture data base is still under development, we used the existing fault-rupture data set of Wesnousky (2008) for this application of the proposed method based on the wavenumber spectrum model. We will apply the new method to the UCLA PFDHA project data base after the development and peer review of the data set is completed.

From the Wesnousky (2008) data set, we selected a subset of 32 earthquakes with magnitude M range from 6.1 to 7.9 (Table 5.1). We defined the amplitude of the surface slip as the total amplitude of the vector sum of the horizontal and vertical slip at each location. Earthquakes with any sense of slip are included in the selected subset: of the 32 earthquakes, seven are normal, 20 are strike-slip and five are reverse.

The data set includes both single and multi-segment ruptures (Table 5.1). The definition of segments is a controversial topic that is being addressed as part of the framework for distributed fault rupture in the UCLA PFDHA project. For this study, we adopted the segmentation as defined in Wesnousky (2008). As an example, Figure 5.1 illustrates the classification as single or multi-segment ruptures. In the case of multi-segment ruptures, each segment is analyzed separately, and the SRL is defined by the length of the individual segments. The treatment of segments will be reevaluated once the framework for the UCLA PFDHA project is complete.

5.4 Wavenumber Spectral Analysis

In this study, our focus is on the model for the $f_{ND}(D_{norm}|x)$ term in eq (5.2). The surface-slip model is formulated in the wavenumber domain which accounts for the along-strike correlation in the data, as well as avoiding the normalization of the distance along strike by the SRL. The use of the wavenumber domain builds on the seismological approach to model the two-dimensional slip distribution on the rupture plane at depth (Somerville et al., 1999). Using the wavenumber spectrum provides a physical basis for extrapolating the model beyond the empirical data used to develop the model that is missing from the purely statistical models currently used. Because the wavenumber spectrum has not been commonly used for fault rupture modeling, a brief description of the wavenumber spectrum is given below.

Table 5.1: Analyzed Earthquakes (Wesnousky, 2008)

Date (mm/dd/yyyy)	Location	Type	M	Length (km)	Reported A.D. (m)	Reported M.D. (m)	Segment. Flag[1]
01/09/1857	Fort Tejon, CA	SS	7.9	360	4.7	9.1	S
05/03/1887	Sonora, MX	N	7.2	70	2.2	4.1	S
10/28/1891	Neo-Dani, Japan	SS	7.3	80	3.1	7.9	S
8/31/1896	Rikuu, Japan	R	7.2	37	2.5	6.5	M
10/2/1915	Pleasant Valley, CA	N	7.3	61	1.8	5.8	M
11/2/1930	Kita-Izu, Japan	SS	6.7	35	1.1	3.5	M
12/25/1939	Erzincan, Turkey	SS	7.7	300	4.2	7.4	S
5/19/1940	Imperial, CA	SS	6.9	60	1.6	3.3	S
12/20/1942	Erbaa-Niksar, Turkey	SS	6.8	28	1.66	1.9	S
11/26/1943	Tosya, Turkey	SS	7.6	275	2.5	4.4	S
9/10/1943	Tottori, Japan	SS	6.3	10.5	0.6	1.5	M
2/1/1944	Gerede-Bolu, Turkey	SS	7.35	155	2.1	3.5	S
1/31/1945	Mikawa, Japan	R	6.2	4	1.3	2.1	S
12/16/1954	Fairview Peak, NV	N	7	62	1.1	5.2	M
12/16/1954	Dixie Valley, NV	N	6.8	47	0.8	3.1	M
8/18/1959	Hebgen Lake, MT	N	7	25	2.5	5.4	M
7/22/1967	Mudurnu, Turkey	SS	6.7	60	0.9	2	S
4/8/1968	Borrego Mtn, CA	SS	6.1	31	0.13	0.4	M
2/9/1971	San Fernando, CA	R	6.7	15	0.95	2.5	S
6/2/1979	Cadoux, Australia	R	6.1	10	0.6	1.2	S
7/29/1981	Sirch, Iran	SS	6.4	64	0.13	0.5	S
10/28/1983	Borah Peak, ID	N	6.9	34	0.94	2.8	M
3/2/1987	Edgecumbe, NZ	N	6.3	15.5	0.6	2.6	M
11/23/1987	Super. Hills, CA	SS	6.2	25	0.45	0.8	M
7/16/1990	Luzon, Philippines	SS	7.6	112	3.5	6.2	S
6/28/1992	Landers, CA	SS	7.2	77	2.3	6.7	S
3/14/1998	Fandoqa, Iran	SS	6.6	25	1.1	3.1	S
9/21/1999	Chi-Chi, Taiwan	R	7.4	72	3.5	12.7	S
10/16/1999	Hector Mine, CA	SS	6.9	44	1.56	5.2	S
11/12/1999	Duzce, Turkey	SS	7	40	2.1	5	S
11/14/2001	Kunlun, China	SS	7.8	424.5	2.4	8.3	S
11/3/2002	Denali, AK	SS	7.7	302	3.6	8.9	S

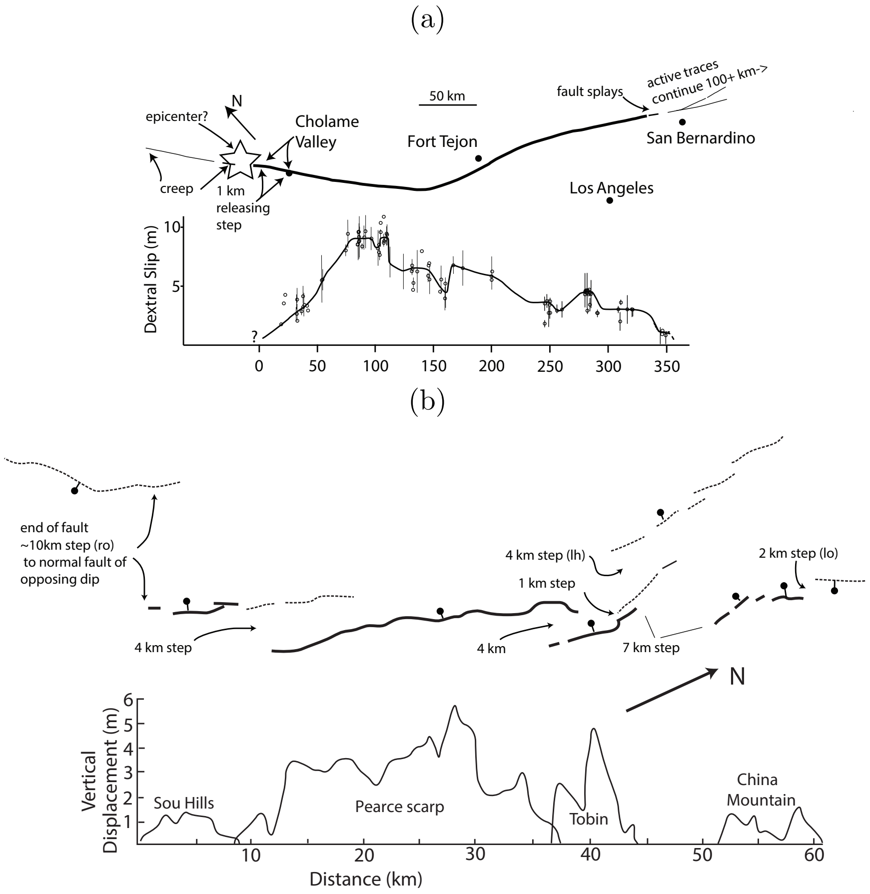


Figure 5.1: Examples of single and multi-segmented event classification, after Wesnousky (2008) electronic supplement. (a) 1857 Fort Tejon, CA surface slip profile, classified as single rupture event; (b) 1915 Pleasant Vallely, NV surface slip profile, classified as multi-segment rupture event

The Fourier Transform (eq. 5.4) allows us to decompose a displacement profile into a series of sinusoidal waveforms of different wavelengths. In eq (5.4), $y(x)$ is the surface displacement and $Y(k)$ is the Fourier transform of $y(x)$. Each sinusoidal function can be fully defined by its wavelength, amplitude (A), and phase angle (ϕ). The wavelength defines how broad the fluctuations of slip are over the surface rupture length, the amplitude is proportional to the amplitude of the sinusoid, and the phase angle quantifies the offset of the peak of the sinusoid from the origin. The amplitude and phase spectra are expressed as a functions of the wavenumber, k, which is the reciprocal of the wavelength. The sampling interval along the rupture, dx, defines the largest wavenumber that can be resolved (Nyquist wavenumber) and the SRL controls the wavenumber interval, dk.

$$Y(k) = \int_0^{SRL} y(x)e^{-i2\pi kx}dx = A(k)e^{i\phi(k)} \tag{5.4}$$

Similarly, the inverse Fourier Transform (eq. 5.5) can be used to compute the displacement profile if the amplitude and phase spectra are known.

$$y(x) = \sum_{i=n/2}^{n} Y(k_i)e^{i2\pi xk_i}dk = \sum_{i=n/2}^{n} A(k_i)e^{i\phi(k_i)}e^{i2\pi xk_i}dk \tag{5.5}$$

A Fourier series defines a periodic function that repeats itself. Discontinuities at the ends of the ruptures, that could create noise in the Fourier domain due to periodic assumptions, are avoided by setting the first and last data point of the profiles to be equal to zero.

Developing empirical models for the amplitude spectrum is straight forward, but developing empirical models for the phase is more difficult due to phase wrapping about 2π. Instead of working directly with phase angles, previous studies have developed models for the phase-difference distribution. For example, ground-motion studies (Matsukawa et al., 1987; Ohsaki, 1979; Yokoyama et al., 1988) have used phase-difference distributions and have shown that the width of the phase-difference distribution is related to the envelope of the time series. For surface rupture studies, the width of the phase-difference distribution is related to the envelope of the displacement profile. In this study, we use phase derivatives rather than phase differences because it allows the phase data from earthquakes with different SRL and different dk to be combined in a single regression analysis for developing empirical phase models.

The phase derivative ($\partial\phi/\partial k$) describes the relative location of peaks along strike from different wavenumbers. If the phase derivatives are from a narrow distribution, the peaks from the different wavenumbers will constructively interfere over a portion of the rupture and destructively interfere in other areas, leading to a more peaked slip profile along strike. In contrast, if the phase derivatives are from a broad distribution, then the peaks of the sinusoids will be randomly separated along strike, leading to a more uniform distribution of slip along strike.

When computing the phase derivatives, phase unwrapping needs to be considered. Phase wrapping is an numerical artifact when calculating the phase angles: if the phase exceeds π or is less than $-\pi$, it is wrapped to stay in the $[-\pi, \pi]$ range. Phase wrapping can create problems when calculating the phase-derivative numerically with a finite-difference approximation for the derivative: in the case in which there are two adjacent wavenumbers with one of them having the phase wrapped by 2π, there will be a discontinuity in the phase derivative.

The phase derivatives can be computed analytically (eq. 5.6) without any assumptions about phase unwrapping (Boore, 2003; Stoffa et al., 1974):

$$\frac{\partial\phi(k)}{\partial k} = 2\pi\frac{\Re(Y(k))\Re(G(k)) - \Im(Y(k))\Im(G(k)))}{A(k)^2} \tag{5.6}$$

In eq (5.6), $G(k)$ is the Fourier transform of the product of the slip profile and the along strike distance, $x \times y(x)$. The $\Re$ and $\Im$ functions are the operators for taking the real and imaginary parts of a complex number.

As examples of the representation of a slip profile in the wavenumber domain, we use the surface-slip profiles from the 1971 San Fernando and 1983 Borah Peak earthquakes (Figure 5.2). The slip profiles for the two earthquakes, as compiled in Wesnousky (2008) database, are shown in Figures 5.2a and 5.2b. These two rupture profiles have similar SRL, so the comparison of amplitude spectra and phase-derivative distributions for these two earthquakes are not influenced by differences in the SRL. The slip profile for 1971 San Fernando earthquake has a wider peak compared to the 1983 Borah Peak earthquake, which results in a wider phase-derivative distribution for 1971 San Fernando (Figure 5.2c) compared to that for the 1983 Borah Peak earthquake (Figure 5.2d). The amplitude spectra for these two earthquakes are compared in Figures 5.2e and 5.2f. By definition, the spectral ordinate at the zero wavenumber is equal to $AD \times SRL$. The AD values for the 1971 San Fernando and 1893 Borah Peak earthquakes are 0.87 m and 0.28 m, respectively. This corresponds to amplitudes in the wavenumber spectrum at $k = 0$ of 14500 m^2 and 4500 m^2 as shown in Figures 5.2e and 5.2f. Additionally, the narrower peaks in the 1893 Borah Peak slip profile, compared to those in the 1971 San Fernando slip profile, results in a smaller rate of decay in the amplitude spectrum of the 1893 Borah Peak earthquake because there is greater amplitude at the shorter wavelengths needed to create the narrower peaks.

5.4.1 Non-uniform Fourier Transform

A key challenge for using wavenumber spectra for surface displacement data is the estimation of the Fourier amplitude and phase spectra of non-uniformly sampled slip measurements. The standard discrete Fourier transform (DFT) algorithms are developed for signals that are sampled at a constant interval. Due to constraints in the field, such as inaccessibility of sites, vegetation, and limitations in the measuring techniques, many slip profiles have been sampled at non-uniform intervals. Modern remote sensing techniques, such as pixel matching (Milliner et al., 2016), Lidar (Gori et al., 2018) and structure for motion (Kayen et al., 2018), make the uniform sampling of displacement profiles feasible; however, there is still a need for a robust method that can be used for older data sets with uneven sampling.

The Fourier transform can be written in matrix notation:

$$\vec{Y} = \boldsymbol{F}\ \vec{y} \tag{5.7}$$

For surface-slip data sets, the ratio of the largest sample interval (dx_{max}) to the smallest sample interval (dx_{min}) is often large, leading to the Fourier Matrix ($\boldsymbol{F}$) in eq (5.7) being ill-conditioned (i.e. the solution is highly sensitive to small perturbations in the slip profile). The matrix $\boldsymbol{F}$ is given by

$$\boldsymbol{F} = exp(-i2\pi \vec{x}^{\intercal} \vec{k})\ dx \tag{5.8}$$

where $\vec{x}$ is the vector of along strike distances, $\vec{k}$ is the vector of wavenumbers for which the Fourier coefficients $\vec{Y}$ are estimated, and dx is the mean distance interval between samples.

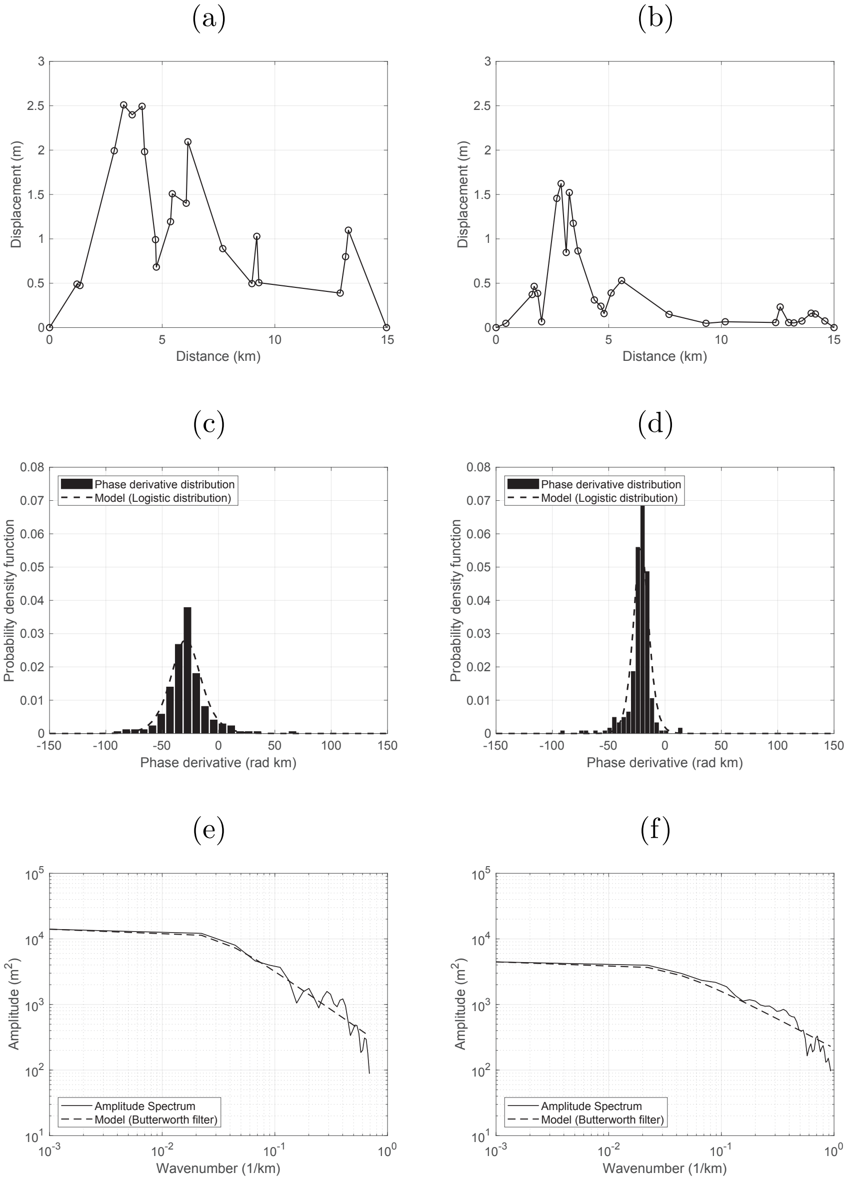

(a) (b) (c) (d) (e) (f)

Figure 5.2: Representation of 1971, San Fernando, CA and 1983, Borah Peak, ID, second segment, slip profiles in space and wavenumber domain. (a) 1971, San Fernando, CA earthquake surface-slip profile; (b) 1983, Borah Peak, ID earthquake, second segment surface-slip profile; (c) Phase-derivative distribution of the 1917, San Fernando, CA earthquake; (d) Phase-derivative distribution of the 1983, Borah Peak, ID earthquake; (e) Amplitude spectrum of the 1971, San Fernando, CA surface-slip profile; (f) Amplitude spectrum of the 1983, Borah Peak, ID surface-slip profile.

In many standard Fourier Transform computer programs, the dx is assumed to be unity without dimensions. In eq (5.8), we included dx to have physical units for the amplitude spectrum in the wavenumber domain, which in this case are m^2. This allows us to relate the amplitude at zero wavenumber to the physical parameter $AD \times SRL$.

One approach to circumvent issue of the ill-conditioned $\boldsymbol{F}$ matrix is to linearly interpolate the displacement profiles ($\vec{y}$) at a constant interval, and then perform a DFT; however, linear interpolation will imply a smoother profile as it connects the locations with measured displacements with line segments that will affect the Fourier spectra at high wavenumbers. Our approach uses a regularized least-squares regression, hereafter, called the regularized Fourier transform (RFT), given by:

$$\vec{Y} = \left(\boldsymbol{F}_{\boldsymbol{inv}}^{\intercal}\boldsymbol{F}_{\boldsymbol{inv}} + \boldsymbol{\Gamma}^{\intercal}\boldsymbol{\Gamma}\right)^{-1} \boldsymbol{F}_{\boldsymbol{inv}}^{\intercal}\vec{y} \tag{5.9}$$

where $\boldsymbol{F}_{\boldsymbol{inv}}$ is the inverse Fourier matrix given by

$$\boldsymbol{F}_{\boldsymbol{inv}} = exp(i2\pi\vec{k}^{\intercal}\vec{x})\ dk \tag{5.10}$$

and $\boldsymbol{\Gamma}$ is the regularization matrix. This approach does not require any interpolation of a slip profile prior to computing its Fourier spectra. We find $\vec{Y}$ that minimize the misfit between the slip profile and the inverse Fourier transform ($\vec{y}_{fwrd}$), with the additional constraint of minimizing the magnitude of $\vec{Y}$. This way, the solution fits the slip profile but also avoids over-fitting at the high wavenumbers. Using the inverse transform, the computed profile, $\vec{y}_{fwrd}$ is given by:

$$\vec{y}_{fwrd} = \boldsymbol{F}_{\boldsymbol{inv}}\ \vec{Y} \tag{5.11}$$

Based on validation tests presented later, it was found that the best results where achieved by a diagonal regularization matrix ($\boldsymbol{\Gamma}$), given by:

$$\boldsymbol{\Gamma} = \alpha\ diag(\vec{k}/dk)^n\ dk \tag{5.12}$$

The $\boldsymbol{\Gamma}$ matrix includes a proportionality factor, α, and is scaled by dk to have the proper units in the wavenumber domain. The element in $\boldsymbol{\Gamma}$ that corresponds to the zero wavenumber was set to $\alpha \times dk$ to avoid dividing by zero. This wavenumber-dependent $\boldsymbol{\Gamma}$ matrix, which applies a greater regularization to the high wavenumbers compared to the small ones, ensures that the bulk of the slip profile is fitted by the small wavenumbers, and the high wavenumbers are limited to fitting the narrow variations of slip.

The parameter α controls the relative weight between the size of residuals of $\vec{y} - \vec{y}_{fwrd}$ and the magnitude of $\vec{Y}$. If α is too large, $\boldsymbol{\Gamma}$ over-constrains the regression leading to an overly smooth slip profile and $\vec{y}_{fwrd}$ will not match the original profile $\vec{y}$. Alternatively, if α is too small, $\boldsymbol{\Gamma}$ doesn't impose enough regularization and results in the same $\vec{Y}$ as the ordinary least-squares solutions with the issues mentioned previously. An additional benefit of this approach is that it allows for a straight-forward consideration of the measurement uncertainty in the slip estimates. Estimates of uncertainty were not considered in this study because they were not available the current database, but this information will be part of the UCLA PFDHA data base that is currently being developed.

The optimum values of α and n are determined through validation tests that evaluated the ability of the proposed method to recover the wavenumber spectrum parameters of artificial data sets with known parameter values. The validation tests are described in detail in a subsequent section. In addition to the quantitative validation tests, using visual inspection, we checked that the inverse slip profiles ($\vec{y}_{fwrd}$) appear reasonable and are in agreement with the observed ones for past earthquakes in the database.

To implement the validation tests, we had to assume a model for the wavenumber spectrum to use as a test. Therefore, we used initial empirical models to set the wavenumber spectrum parameters for testing, then, we developed new empirical models using the selected approach and repeated the validation. Rather than stepping through this iterative process, we show the modeling using the selected final model parameters and then show the validation using the parameters from this final model.

5.5 Amplitude Model

We use the functional form of a scaled Butterworth filter (eq 5.13) to describe the shape of the amplitude spectrum.

$$\vec{B}_{amp} = \frac{B_0}{\sqrt{1+\left(\frac{\vec{k}}{KC}\right)^{2Np}}} = \frac{AD \times SRL}{\sqrt{1+\left(\frac{\vec{k}}{KC}\right)^{2Np}}} \tag{5.13}$$

Our selection of this form is based on the Somerville et al. (1999) study, who showed that a 2D Butterworth filter is a good approximation to the 2-D amplitude spectrum of the slip variation on the fault plane at depth.

The Butterworth filter is defined with two parameters: the number of poles (Np) and the corner wavenumber (KC). A schematic figure of the functional form for amplitude model is shown in Figure 5.3. The KC is the wavenumber at which the amplitude of the Butterworth filter is $1/\sqrt{2}$ of B_0. It is related to the SRL and the shape of the mean slip along strike. A high KC corresponds to slip profiles in which the slip goes to zero over small distance intervals along strike, while a low KC implies larger distances along strike between the zero displacements. The Np parameter controls the rate at which the high wavenumbers decay. It is related to the amount of variability along strike about the mean slip profile. A smaller value of Np leads to greater variability along strike.

For our slip model, there is a a third parameter, B_0, which is the amplitude at zero wavenumber and it scales the Butterworth filter. As noted earlier, the value of B_0 corresponds to the product of $AD \times SRL$.

The correlation length in the space domain is related to the amplitude parameters KC and Np in the wavenumber domain. For a SRL of $100km$ and a triangular shape of $T(x/SRL)$, a

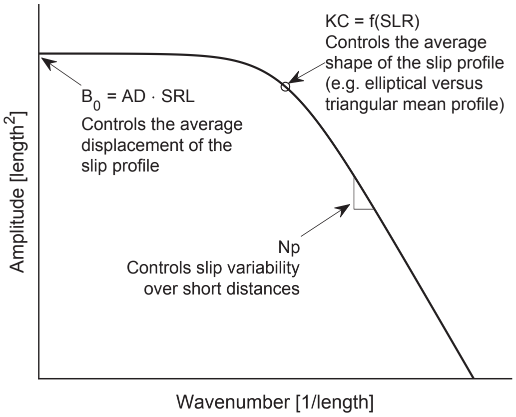


Figure 5.3: Schematic of the amplitude model. The functional form of the amplitude model corresponds to equation 5.13. The scaling relationships that control the corner wavenumber and number of poles are defined in eq. (5.15) and (5.16), respectively.

correlation length of 2 km leads to mean values of $KC = 0.0037/km$ and $Np = 1.07$; similarly a correlation length of 10 km, leads to mean values of $KC = 0.0044/km$ and $Np = 1.27$. Additional information on the relation between the correlation length and the wavenumber parameters is given in Figures B.1 and B.2 in Appendix B.

For each earthquake, we estimate the three parameters B_0, KC, and Np. The B_0 parameter is set equal to the computed amplitude spectrum at k=0. The KC and Np parameters are estimated by fitting the amplitude spectra of the displacement profiles to eq (5.14):

$$log_{10}Bamp_i(k_j) = log_{10}(B_{0,i}) - 0.5 \left[log_{10} \left(1 + \frac{k_j}{KC(SRL_i) + \eta_{KC,i}} \right)^{2(Np+\eta_{Np,i})} \right] + \epsilon_{ij} \quad (5.14)$$

where i is the index for the earthquake and j is the index for the wavenumber. The residuals are assumed to be log-normally distributed. The regression analysis is performed in a mixed-effects maximum-likelihood framework, following the Abrahamson and Youngs (1992) methodology, with the KC and NP modelled as random effects.

The functional form of the model used for KC is given by:

$$\log_{10}(KC) = C_1 + C_2(\log_{10}(SRL) - C_3) + \eta_{KC}, \quad \eta_{KC} \sim N(0, \tau_{KC}) \quad (5.15)$$

The KC is modeled as a function of the SRL (eq. 5.15) rather than earthquake magnitude because KC is related to the shape of the mean displacement profile which depends directly on SRL, but more weakely on M. Additionally, using SRL allows us to have the same scaling relationship for both single and multi-segment ruptures, with KC as a function of the length of each segment. The C_3 term is included in (eq. 5.15) to center the model on the weighted average SRL. For this data set, it is set at $C_3 = 1.6$. Preliminary analyses found Np to be independent of SRL. Therefore, the mean Np is modeled as a constant, independent of the SRL (eq. 5.16).

$$N_p = C_4 + \eta_{Np}, \quad \eta_{Np} \sim N(0, \tau_{Np}^2) \tag{5.16}$$

The likelihood function used to estimate the regression coefficients (θ) is given by:

$$ln\left(L(\theta)\right) = \frac{N}{2} ln(2\pi) - \frac{1}{2} ln\left(|\boldsymbol{\Sigma_{tot}}|\right) - \frac{1}{2}\left(\vec{f}_{amp} - \vec{B}_{amp}(\theta)\right)^{\intercal} \boldsymbol{\Sigma_{tot}}^{-1} \left(\vec{f}_{amp} - \vec{B}_{amp}(\theta)\right) \tag{5.17}$$

where N is the number of data points, f_{amp} the vector of Fourier amplitudes at all wavenumbers for all earthquakes, and $\boldsymbol{\Sigma_{tot}}$ is the covariance matrix described below. Because the functional form of the Butterworth filter is non-linear (eq 5.13), the likelihood function is maximized using the nonlinear maximum likelihood package, NLMER (Bates et al., 2015), given in the statistical software R (R Core Team, 2018).

The covariance matrix ($\boldsymbol{\Sigma_{tot}}$) is a block diagonal matrix with each block corresponding to the covariance matrix for one earthquake. The block for the i^{th} earthquake, $\boldsymbol{\Sigma_{tot,i}}$, is a square matrix with its size equal to the number of wavenumbers of the i^{th} earthquake. Each block consists of three terms: the intra-event covariance matrix $\boldsymbol{\Sigma_{int,i}}$ and the two inter-event covariance matrices ($\boldsymbol{\Sigma_{KC,i}}$ and $\boldsymbol{\Sigma_{Np,i}}$) for the random-effects terms. The covariance matrix is given by

$$\boldsymbol{\Sigma_{tot,i}} = \boldsymbol{\Sigma_{int,i}} + \boldsymbol{\Sigma_{KC,i}} + \boldsymbol{\Sigma_{Np,i}} \tag{5.18}$$

Weights are included in the intra-event covariance matrix to put less weight on the higher wavenumbers. The weights account for two effects: (1) there are larger uncertainties in the amplitude at high wavenumbers due to the uneven sampling, and (2) the evaluation of the amplitude fit is on log scale, but the sampling of the wavenumbers is linear, leading denser sampling of high wavenumbers for the same wavenumber interval in log space. We selected weights that are inversely proportional to their wavenumbers normalized by wavenumber interval (dk_i) of the i^{th} earthquake:

$$\boldsymbol{\Sigma_{int,i}} = \phi^2 diag(\vec{k}/dk) \tag{5.19}$$

To avoid dividing by zero, the weight corresponding to the zero wavenumber is set equal to the weight of the smallest non-zero wavenumber. The validation tests, presented in a subsequent section, confirmed that using the proposed weights for the $\boldsymbol{\Sigma_{int,i}}$ matrix performs better than using equal weights.

The $\boldsymbol{\Sigma_{KC,i}}$ and $\boldsymbol{\Sigma_{Np,i}}$ covariance matrices were computed by linearizing the Butterworth filter functional form with a first-order Taylor series expansion (Bevington and Robinson, 1993; Stafford, 2015; Abrahamson et al., 2016). With this approximation, the total residual

($\epsilon_{total,ij}$) is expressed as a linear combination of the intra-event ($\epsilon_{i,j}$) and inter-event ($\eta_{KC,i}$ and $\eta_{Np,i}$) residuals:

$$\epsilon_{total,ij} = \sqrt{dk_i/k_j}\ \epsilon_{i,j} + \frac{\partial Bamp_{i,j}}{\partial \eta_{KC}} \eta_{KC,i} + \frac{\partial Bamp_{i,j}}{\partial \eta_{Np}} \eta_{Np,i} \tag{5.20}$$

Assuming that ϵ, η_{KC}, and η_{Np} are uncorrelated, the total variance is equal to:

$$\sigma^2 = \frac{dk}{k}\phi^2 + \left(\frac{\partial Bamp}{\partial\ \eta_{KC}}\right)^2 \tau_{KC}^2 + \left(\frac{\partial Bamp}{\partial\ \eta_{Np}}\right)^2 \tau_{Np}^2, \tag{5.21}$$

where ϕ^2 is the intra-event variance and τ_{KC}^2 and τ_{Np}^2 are the inter-event variances for KC and Np, respectively. The structure of $\boldsymbol{\Sigma_{KC,i}}$ and $\boldsymbol{\Sigma_{Np,i}}$ covariance matrices is shown in below:

$$\boldsymbol{\Sigma_{KC,i}} = \tau_{KC}^2 \left(\frac{\partial \vec{Bamp_i}}{\partial \eta_{KC}}\right) \left(\frac{\partial \vec{Bamp_i}}{\partial \eta_{KC}}\right)^{\mathsf{T}} \tag{5.22}$$

$$\boldsymbol{\Sigma_{Np,i}} = \tau_{Np}^2 \left(\frac{\partial \vec{Bamp_i}}{\partial \eta_{Np}}\right) \left(\frac{\partial \vec{Bamp_i}}{\partial \eta_{Np}}\right)^{\mathsf{T}} \tag{5.23}$$

5.5.1 Regression Results for Amplitude Model

Before computing the RFT, the slip profiles were padded with zeros at the end of the rupture equal to twenty times their rupture length to provide a more robust estimate of the phase derivatives by reducing dk. The amplitude spectrum is modeled using eq (5.13), (5.15), and (5.16) with the η_{KC}, η_{Np}, and ϵ being normally distributed variates. A maximum-likelihood regression is conducted using the likelihood function given in eq (5.17) and the covariance given in eq (5.18). Table 5.2 summarizes the estimates and standard errors of the coefficients of the global (ergodic) amplitude model. A value of $C_2 = 1$ corresponds to a self similar scaling in terms of the SRL. From the regression, the estimate of the C_2 coefficient was 1.009 with a standard error of 0.02. This is not significantly different from a self-similar model, so we constrained $C_2 = 1$ in the final model shown in Table 5.2.

The reported standard errors in Table 5.2 represent the epistemic uncertainty due to the limited data set assuming the Butterworth filter is the correct functional form. To capture the full epistemic uncertainty, alternative functional forms of the amplitude spectrum should to be considered in addition to the statistical uncertainties.

Figures 5.4a and 5.4b compare the KC and Np of the individual earthquake regressions with the global relationships, respectively. The random variates η_{KC} and η_{Np} are in general agreement with the assumption that they are normally distributed (see Figure B.3b of Appendix B).

From the regression analysis, $Np = 1.24 \pm 0.21$. For the slip over the rupture plane at depth, Somerville et al. (1999) found that a Butterworth filter with $Np = 2$ provided a good fit to the wavenumber spectrum. A lower Np for the slip at the surface than the Np for the

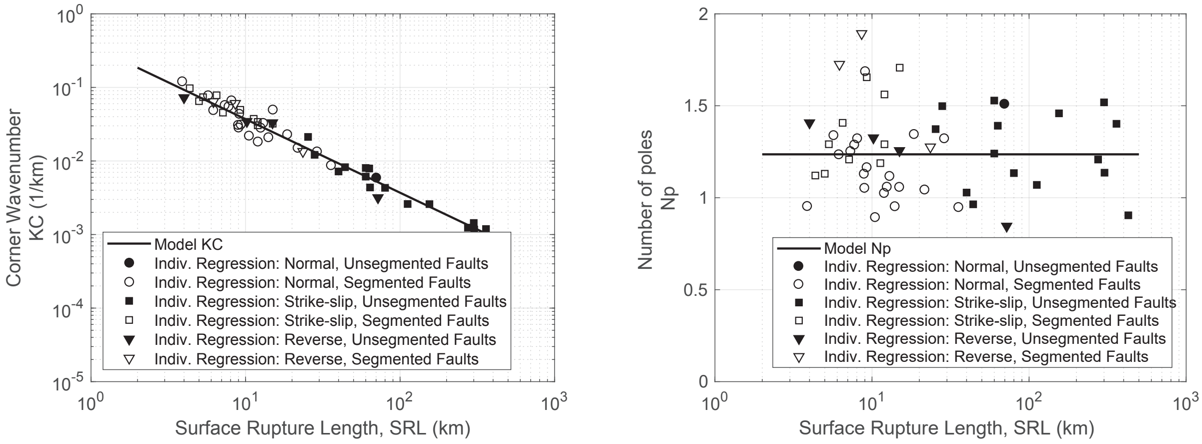

(a) (b)

Figure 5.4: Scaling relationships of the parameters of the amplitude model versus surface rupture length. (a) Scaling relationship of corner wavenumber (b) Scaling relationship of the number of poles. The different markers represent the individual regressions grouped by fault and segmentation type, the solid line corresponds to the global model.

slip at depth implied by source inversions may reflect greater along-strike variability at the surface than at depth or it may simply reflect the effect of constraints imposed on slip-model inversions that lead to smoothed slip distributions at depth. Milliner et al. (2015) found that the power spectrum of the total surface displacement profile of the 1992 Landers, CA earthquake was decaying with a rate of 1.87 at the high wavenumbers. This slope on the power spectrum corresponds to Np of 0.93 for the amplitude spectrum, and is within the $\pm 2\sigma$ range of the Np from our regression analysis.

The style of faulting is not included as a predictor variable because the estimated KC and Np values do not show clear trends with the style of faulting. Although the sense of faulting may have an impact in the average displacement (Wells and Coppersmith, 1994), our analysis suggests that the distribution of the normalized displacements along strike is not sensitive to the style of faulting, but it is based on a small sample size.

The dependence of the KC and Np scaling on the single and multi-segment rupture classification was also evaluated. A preliminary regression that included an additional term for the rupture classification showed that the single-rupture versus multiple-rupture classification had insignificant effects on the scaling for the KC and the Np. Figures 5.4a and 5.4b compare the estimates of the individual regressions with the mean global relationships for KC and Np. In these figures, it can be seen that neither the style of faulting nor segmentation grouping show systematic trends relative to the mean relationships.

Figure 5.5 presents the intra-event residuals and the binned mean residual versus the

Table 5.2: Amplitude model coefficients

	C_1	C_2	C_3	C_4	τ_{KC}	τ_{Np}	ϕ_0
Estimate	-2.031	-1.00	1.6	1.236	0.071	0.211	0.265
Std. Error	0.0145	-	-	0.0317	-	-	-

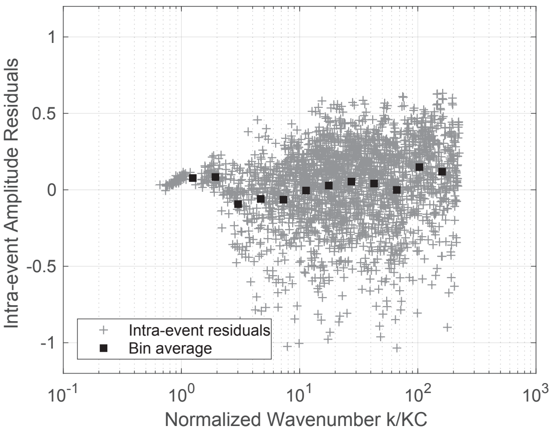


Figure 5.5: Distribution of intra-event residuals. Crosses represent the intra-event residuals. Solid squares correspond to the mean of the residuals in each bin

wavenumber normalized by the event-specific KC. It shows that the mean of the amplitude model is unbiased as the mean residuals of all bins are close to zero and also that the variance of the intra-event residuals is approximately constant for $k/KC_i > 5$ (the binned standard deviation values are shown in Figure B.4 in Appendix B). Figure 5.5 shows a reduction in variability for $KC < 5$. In test applications, both a constant value of ϕ_0 and a k/KC-dependent value of ϕ_0 that reduces ϕ_0 for small $K/KC < 5$ were evaluated. The differences in the generated profiles for these two alternative models for ϕ_0 were small, so a constant ϕ_0 model (eq. 5.24) was selected for simplicity.

$$var(\epsilon) = \phi_0^2 \tag{5.24}$$

5.6 Phase-Derivative Model

The phase derivatives are modeled using a logistic distribution rather than a normal distribution because the latter implied fatter tails than observed. This has also been observed for the phase-derivative distributions for ground motions (Baglio et al., 2017). The logistic distribution is defined with two parameters: the location (μ) and the scale (s), which

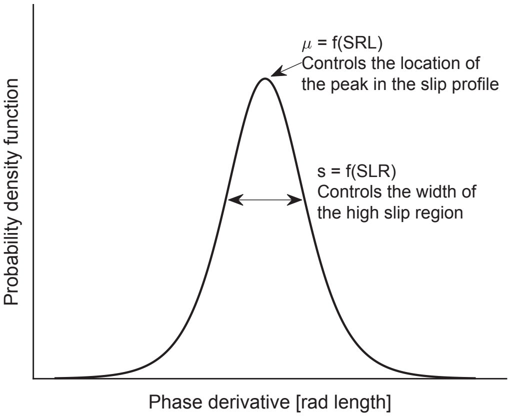


Figure 5.6: Schematic of the phase-derivative model. The scaling relationships that control the center and with of the distribution are defined in eq. (5.25) and (5.26), respectively.

control the center and width of the distribution, respectively. A schematic figure of the phase-derivative distribution model is shown in figure 5.6. A preliminary analysis showed that both parameters are functions of the SRL and they are modeled by:

$$\mu_i/SRL_i = -\pi + \epsilon_{\mu_i},\ \epsilon_\mu \sim N(0, \sigma_\mu^2) \tag{5.25}$$

$$log_{10}(s_i) = D_1 + D_2(log_{10}(SRL_i) - D_3) + \epsilon_s, \quad \epsilon_s \sim N(0, \sigma_s^2) \tag{5.26}$$

The starting edge of the rupture in the data file is arbitrary, so we included each event twice in the phase-derivative regressions: once with the original order of surface slips and once with reverse order. For the reverse order case, the slip profile is mirrored on its center axis. The phase angles of the reversed profile are equal to:

$$\vec{\phi}_{rev} = -\vec{\phi}_{orig} - 2\pi\ SRL_i\ \vec{k} \tag{5.27}$$

where $\vec{\phi}_{orig}$ and $\vec{\phi}_{rev}$ are the phase angles of the profile with the original and reversed polarity respectively (see derivation in Section B.2 of Appendix B). Following from eq. (5.27), the phase derivatives of the reversed profile are equal to:

$$\frac{\partial \vec{\phi}_{rev}}{\partial k} = -\frac{\partial \vec{\phi}_{orig}}{\partial k} - 2\pi\ SRL_i \tag{5.28}$$

Thus, the mean of phase derivatives of the original and reversed profiles is $-\pi\ SRL$ which explains the functional form for μ_i eq. (5.25).

As an example, Figure 5.7 shows the phase-derivative distributions of the 1939 Erzincan, Turkey earthquake with the original and reversed polarity. The displacement profiles of this

Table 5.3: Phase-derivative model coefficients

	D_1	D_2	D_3	σ_μ	σ_s
Estimate	1.493	1.0	1.6	0.917	0.159
Std. Error	0.0172	-	-	-	-

earthquake for the original and reverse order cases are presented in Figure 5.7a. These two slip profiles have identical amplitude spectra and their phase-derivative distributions are mirrored (Figure 5.7b). For each earthquake, the mean value of μ passes though the point of symmetry of the phase-derivative distributions between the original and reverse order cases.

The regressions for the global phase-derivative relationships are performed in two steps. First, the μ and s parameters are estimated for the phase-derivative distribution of each individual event, and second, the coefficients of the global relationships (eq. 5.25 and 5.26) are estimated by fitting the event-specific μ_i and s_i using least-squares. The estimate of D_2 is 0.996 with a standard error of 0.03. which is consistent with $D_2 = 1.0$ for self-similar scaling. Therefore, the D_2 term is set to unity.

The estimate and standard error of the coefficients for the phase-derivative model are listed in Table 5.3. The D_3 term is the same parameter as C_3 in the amplitude model and it was used to center the SRL data to reduce the correlation between the intercept and slope of the phase-derivative model. The scaling relationships for μ and s are compared with the event-specific estimates in Figures 5.8a and 5.8b, respectively. Only μ is affected by the polarity of the slip profiles. As found for the amplitude model, the style of faulting and the segmentation of the slip profiles do not have noticeable effects on the distribution of the phase derivatives.

5.7 Validation of the RFT Method with Synthetic Data Sets

As described earlier, the data sets for surface rupture typically have uneven sampling that makes it difficult to compute a Fourier transform. In addition, the locations at which the geologists measured the slip may be biased to regions with larger slips that are easier to measure or are of more interest. An objective of the method is to provide stable and accurate estimation of the amplitude and phase spectrum of a displacement profile based on limited number of unequally-spaced observations. To test that the proposed RFT method can be successfully used to compute unbiased estimates of the wavenumber spectra of slip profiles with typical (unevent and biased) sampling, we performed a series of validation tests. In these tests, target profiles are generated with known amplitude and phase-derivative parameters and then the profiles are down-sampled to approximate the limited number of slip measurements typically available in field surveys. We then check if the method can successfully retrieve the

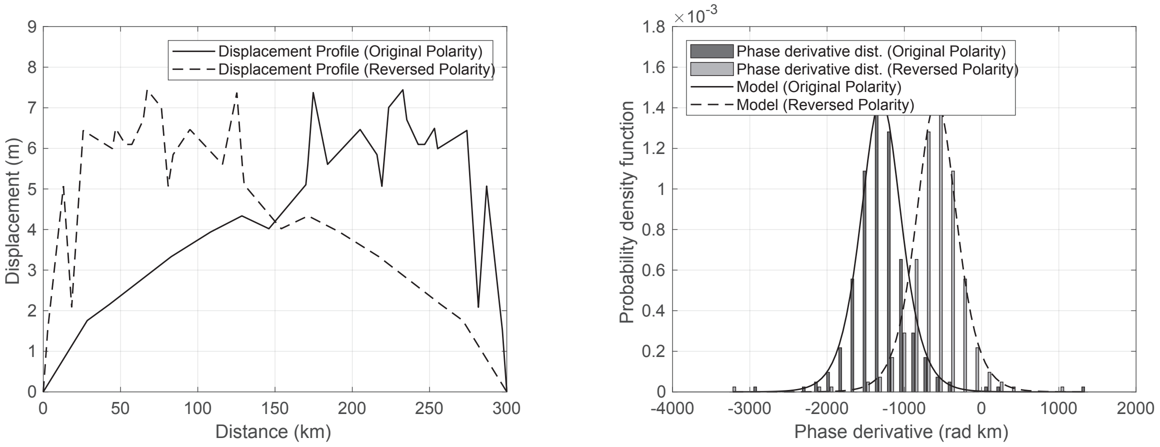

(a) (b)

Figure 5.7: Connection between the polarity of the 1939, Erzincan, Turkey slip profile (i.e. location of the peaks) and the phase-derivative distribution in the wavenumber domain. (a) Slip profiles with the two polarities. The solid line represent the slip profile in the original order. The dashed line shows the profile in the reversed order. (b) Phase-derivative distributions of the slip profiles in subfigure (a). The dark bars bars and solid line are associated with the phase-derivative distribution and model of the profile in the original order. The light bars and dashed lines are related to the phase-derivative distribution and model of the profile in the reversed order

parameters that were used to create target profiles.

The first step in this process is to generate dense, uniformly sampled, target profiles for given sets of amplitude parameters (AD, KC, Np) and phase-derivative distribution parameters (μ and s) for a given SRL. The five sets of parameters used to generate the target profiles are presented in Table 5.4.

For each group of parameters, the amplitude and phase-derivative models were sampled to develop a pair amplitude and phase target spectra. Through an inverse Fourier Transform (eq. 5.5), a candidate slip profile was generated from each pair of target spectra, then it was checked to ensure that the slip goes to zero at the two edges and that the area under any negative slip is less than 0.1% of the total area underneath the slip profile. If a candidate profile didn't satisfy these constraints, it was rejected, new target amplitude and phase spectra were developed and a new candidate profile was created until 25 target profiles were generated that satisfied these constrains. The space-domain sampling (dx) of these profiles was equal to $0.005 \times SRL$ (i.e. 200 samples, equally spaced along the surface rupture).

The sample data sets are created by down-sampling the target profiles to represent the types of sampling in the empirical data set. Two down-sampling schemes are used: random

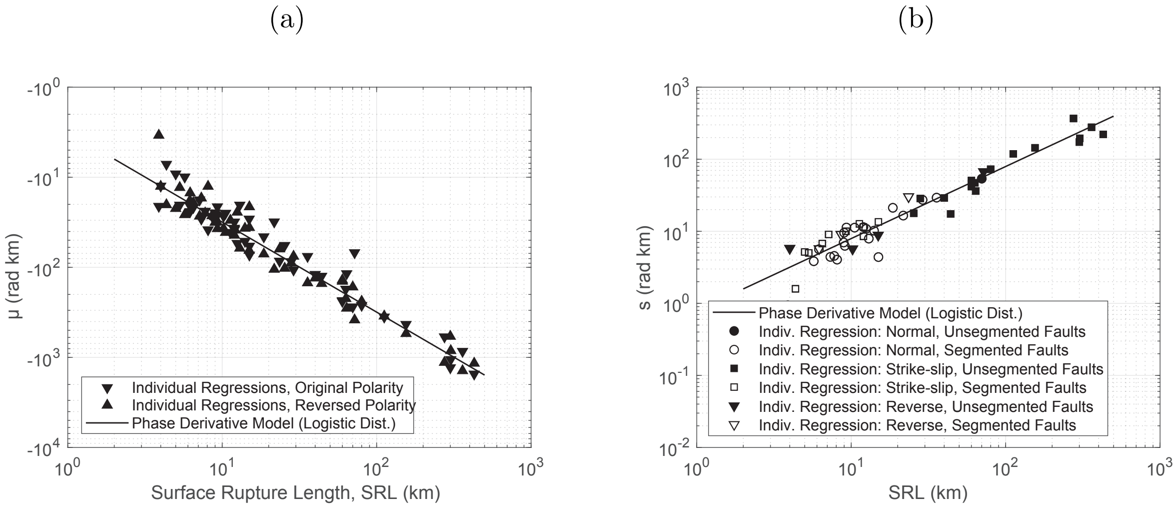


Figure 5.8: Scaling relationship between the phase derivative parameters and Surface Rupture Length. (a) Scaling relationship of location of the logistic distribution (μ). The down-pointing triangles represent the slip profiles in the original order, the up-pointing triangles correspond to the slip profiles in the reversed order and the solid line are the global scaling relationships. (b) scaling relationship of the scale of the logistic distribution (s). The different markers represent the individual regressions grouped by fault and segmentation type, the solid line corresponds to the global model.

sampling (scheme 1), and biased sampling towards peaks (scheme 2). In scheme 1, a down-sampled profile is generated by randomly choosing, without replacement, a portion of the along strike locations where the target profile is evaluated. The number of the sampled locations is defined in terms of the down-sampling ratio of the number of points in the target profiles to the number of points in the down-sampled profiles. In scheme 2, a target profile is separated into eight equal size segments, and, for each segment, the locations with the largest slip values is selected until the number of selected points correspond to the desired down-sampling ratio. The second option represents typical surface rupture data sets in which areas with high slip are more likely to be sampled.

Multiple down-sampling ratios were tested; however, for the sake of space, only the result with 5% sampling of the points in the target profiles are presented here. This down-sampling ratio is representative of the historical earthquakes in the database. Figure 5.9a illustrates the two down-sampling procedures for one of the target profiles. The amplitude spectra of these slip profiles are presented in Figure 5.9b.

For each of the 25 target slip profiles, 100 test profiles generated for each down-sampling technique, resulting in 2500 random and 2500 biased down-sampled profiles. The spectral parameters of the down-sampled profiles were estimated using the alternative methods for

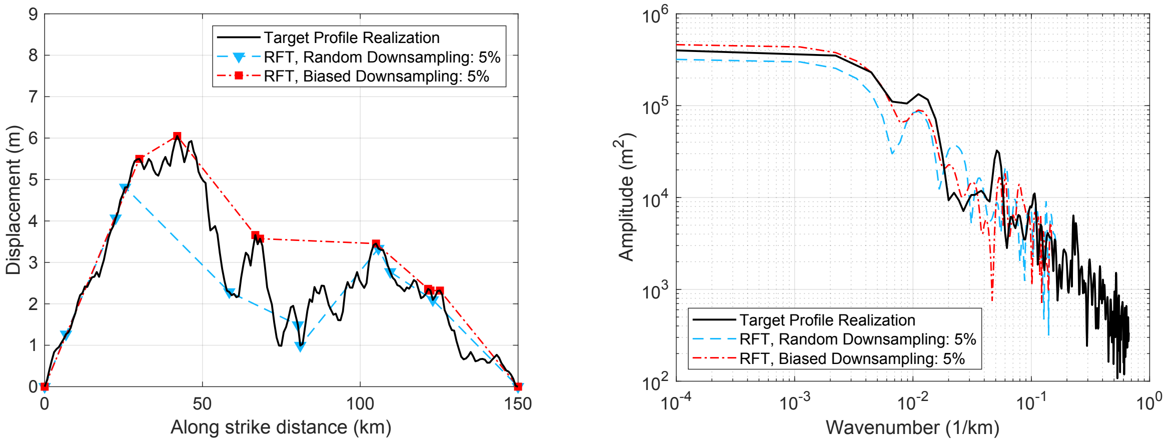

(a) (b)

Figure 5.9: Comparison of random and biased down-sampling methods. (a) Space domain representation, solid line corresponds to the target profile, triangles represent the randomly sampled locations and squares the biased sampling locations towards the peaks of the target profile. (b) Wavenumber domain representation, solid line corresponds to the amplitude spectrum of the target profile realization. The dashed lines and dashed-dotted lines correspond to the amplitude spectra of the randomly and biased sampled profiles shown in figure 5.9a.

computing the Fourier transform. The accuracy of the alternative approaches and alternative model parameter values is evaluated by comparing the amplitude spectrum parameters (B_0, KC, and Np) estimated from the down-sampled sets with respect to parameters used to generate the target profiles. The B_0 and KC errors are computed as the difference between the logarithms of the estimated values from the down-sampled sets and values for the target profiles. The Np errors were calculated using the differences in the arithmetic values. If the error of a test profile is equal to zero, it means that the approach used to estimate its spectral parameters retrieved the true parameters that were used to create it. Large positive or negative errors imply that the approach did not recover the true parameters for the test profiles.

We evaluated the RFT method with the exponent n values of 0.2, 0.5 and 1.0. For the DFT method, linear interpolation at 1 km intervals is used. The mean errors of these methods for all the target realizations, for both the random and biased down-sampling schemes, are presented in Tables 5.5 and 5.6, respectively. The range of the errors is shown in Figure B.12 of Appendix B. Overall, the RFT approach with $n = 0.5$ outperformed the alternative approaches considered for all five SRL values, both in terms of bias and efficiency. In the $n = 0.2$ case, the slip profiles are over fitted by the high wavenumbers, resulting in a negative bias in the estimated Np value. In the $n = 1$ case, the effect of the regularization

Table 5.4: Validation tests amplitude and phase-derivative parameters

SRL (km)	AD (m)	KC (1/km)	Np	μ (rad/km)	s (rad/km)
10	0.38	0.0375	1.24	-31.4	7.9
20	0.73	0.0186	1.24	-62.8	15.7
60	1.39	0.0062	1.24	-188.5	46.8
150	2.66	0.0024	1.24	-471.2	116.7
400	5.09	0.00091	1.24	-1256.6	310.0

on the spectrum at the high wavenumbers is too strong, leading to underestimation of the high-wavenumber content and, subsequently, a positive bias in Np. Based on these tests, we set $n = 0.5$

The DFT approach did not do well in these tests. Using the DFT approach, the Np values are significantly biased to higher values for both down-sampling schemes. As a representative example, Figure 5.10 compares the amplitude spectra computed by RFT with $n = 0.5$ and by DFT with linear interpolation of the same randomly down-sampled profile. For reference, the amplitude spectrum of the target profile from which the down-sampled profile was generated is also presented on the same figure. It can be seen that the linear interpolation method introduces artificial noise in the amplitude spectrum that affects the high-wavenumbers content and results in biased estimates for Np. This happens because the linear interpolation connects the sampled points with straight lines, implying a smoother profile than the target. These tests show that the linear interpolation method is not suitable in approximating a slip distribution when the sampling interval is highly variable.

A similar test was performed for alternative values of α between 0.01 and 0.2, with a fixed value of $n = 0.5$. The results of these tests are presented in B.10 of Appendix B. It was found that the computed Fourier spectra were not sensitive to α between 0.01 and 0.2. Therefore, $\alpha = 0.1$ was used.

The synthetic data-set can also be used to provide insights for the selection of the weights for the covariance matrix. Table 5.7 compares the mean misfit of all target realizations of the KC and Np estimates with a unweighted and $1/k$ weighted covariance matrix. The range of the errors for the different SRL values is presented in B.11 of Appendix B. The B_0 parameter is not affected by the covariance matrix selection, so it is not presented. Both parameters show greater variability and bias when an unweighted covariance matrix is used as it places more weight in the high wavenumbers that are more sensitive to the uneven sampling. In this example, the synthetic profiles were created by random down-sampling, but we found that the biased down-sampling case gave similar results. These results justify our choice to use the weighted covariance matrix, with the weights inversely proportional to the wavenumber.

Table 5.5: Mean errors of regularized and discrete Fourier transforms, using random down sampling

Mean Error	RFT n=0.5	RFT n=0.2	RFT, n=1.0	DFT dx=1km
$log_{10}(AD)$	-0.071	-0.222	-0.023	-0.024
$log_{10}(KC)$	-0.014	-0.364	0.211	0.205
Np	-0.015	-0.517	0.635	0.595

Table 5.6: Mean errors of regularized and discrete Fourier transforms, using biased down sampling

Mean Error	RFT n=0.5	RFT n=0.2	RFT, n=1.0	DFT dx=1km
$log_{10}(AD)$	0.022	-0.145	0.076	0.075
$log_{10}(KC)$	-0.022	-0.316	0.187	0.170
Np	0.018	-0.507	0.748	0.692

Table 5.7: Mean errors of regressions with $1/k$ and constant weighted covariance

Mean Error	$1/k$ weighted	unweighted
$log_{10}(AD)$	-0.0706	-0.0706
$log_{10}(KC)$	-0.0143	-0.1946
Np	-0.0147	-0.1664

5.8 Average Displacement

The estimated AD computed from the observed surface slips depends on the weighting applied to the observations. The simplest approach is to compute the mean of the observed values ignoring the uneven sampling. Another approach is to divide the rupture into intervals of equal length and compute the average slip given equal weight to each interval. This weighted approach typically uses linear interpolation between the slip observations along strike. In contrast, using the wavenumber spectrum, the AD is given by B_0/SRL. The effect of correlation of the slip variability along strike is captured by the shape of the wavenumber spectrum.

As a check, we compare the estimates of the AD for these three approaches. The ratio of the AD from the wavenumber model to AD based on a simple average of the observed surface slips ranges from 0.8 to 1.6 with a mean value of 1.03 and a standard deviation of 0.15. Similarly, the ratio of the AD from the wavenumber model to the AD based on equal weight to each interval ranges from 0.8 to 1.3 with a mean value of 1.02 and a standard deviation of 0.07. (See Figure B.12 of Appendix B for more details on the differences in the

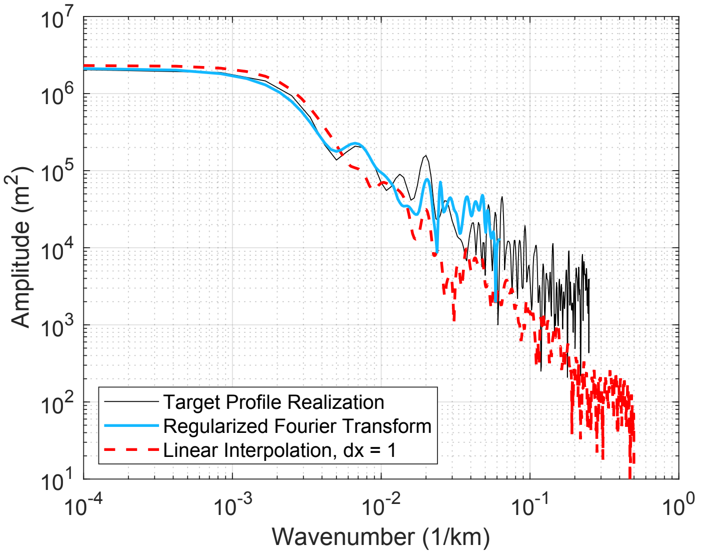


Figure 5.10: Comparison of amplitude spectra computed by RFT and by DFT with linear interpolation. The thin solid line represents the amplitude spectrum of a target profile realization. The thick solid line corresponds to the amplitude spectrum of a down-sampled profile computed with RFT and the dashed line indicates the amplitude spectrum of the same down-sampled profile computed with DFT.

AD estimates). These comparisons show that, on average, the *AD* from the wavenumber model is consistent with the *AD* from current approaches based on averaging in the space domain.

5.9 Along-Strike Variability

The sensitivity of the along-strike variability to the parameters in the wavenumber model is not intuitive. To provide insights into how the along-strike variability of the rupture at short separation distances depends on the wavenumber spectrum parameters, we show results of a sensitivity study for the differences in slip between sites that are separated by 1 km. The along-strike variability is mainly controlled by the *Np* parameter.

Table 5.8 summarizes the sensitivity of the slip variability to the *Np* for three rupture scenarios with M6, M7, and M7.5. The *SRL* and *AD* of these scenarios were estimated with Wells and Coppersmith (1994) empirical scaling relationships for all fault types. For these three magnitudes, the *SRL* values are 8 km, 40 km, and 90 km and the *AD* values are 0.22 m, 1.07 m, and 2.37 m, respectively. The slip variability in Table 5.8 is defined as the standard deviation, in linear space, of the slip difference between all combinations of along strike locations separated by 1km. For example, for the M 7 earthquake with an average

Table 5.8: Representative slip variability between sites at 1 km apart for M equal to 6.0, 7.0 and 7.5.

M		6	7	7.5
SRL (km)		8	40	90
AD (m)		0.22	1.07	2.37
Np	η_{Np}	Slip variability (m)		
0.81	-2.0	0.24	0.81	1.41
1.02	-1.0	0.20	0.48	0.72
1.24	0.0	0.17	0.28	0.35
1.44	1.0	0.16	0.19	0.21
1.66	2.0	0.15	0.15	0.16

$Np = 1.24$, there is 68% probability that two slips 1km apart will be within $\pm 0.28m$ of each other. For smaller values of Np, there is larger variability between nearby sites.

Table 5.8 shows that for high values of Np, the slip variability becomes independent of magnitude. This may seem counter-intuitive. The reason for this behavior is that, for high value of Np, the amplitude of the high wavenumbers is small and the generated profiles start approaching the median shape, (see Figure 5.11d for representative profiles with high and small Np). Therefore, for high value of Np, the slip difference at 1 km interval is controlled by the slope of the median shape. Because the magnitude scaling for SRL and AD is the same in Wells and Coppersmith (1994), the slope of the median shapes is independent of magnitude, so the slip variability over short distances is also magnitude independent.

5.9.1 Example profiles for sensitivity on model parameters

Figure 5.11 presents the results of the sensitivity analysis on different parameters of the model. The profiles shown here are just for illustration purposes to show the effect of the model parameters in the shape of individual profiles, they are not intended to be used in any hazard calculations. In each sensitivity analysis, the parameter under investigation was varied based on their fractile distributions while the remaining parameters were kept fixed at their median values. The profiles were developed in the same fashion as those used to examine the slip variability. The benchmark case corresponds to profiles with a SRL of 100 km and AD equal to 2 m.

Figure 5.11a shows displacement profiles for 16th, median and 84th percentile of μ. This parameter controls the location of the peak of the profiles. For values of μ less than the median, the generated profiles are skewed to right and for values of μ larger than the median, the generated profiles are skewed to the left. Similarly, Figure 5.11b shows example slip profiles for 16th and 84th percentile of s. The s parameter affects the width of the generated profiles. As s increases, the profiles become broader. As s decreases, the profile becomes narrower and the slip at the center of the profiles has to increase to have the same AD. Figure 5.11c shows representative slip profiles for AD equal to 2m and 4m. To generate the slip profiles for these two cases, B_0 was fixed to 210^5 and 410^5 m^2, respectively. It can be seen that as AD increases, the slip increases at all locations.

The amplitude parameters KC and Np do not control the distribution of displacements at each location individually but they control the spatial correlation between them. Figure 5.11d shows two representative examples for the 16^{th} and 84^{th} percentile of Np. As the value of Np decreases, the high-wavenumber content in the amplitude spectrum increases, which leads to a reduction in the correlation length and the peaks and troughs become narrower. The KC has the opposite effect on the correlation length, but because its variability in the model is much smaller, it is not presented here.

5.10 Example Application

In this section, we show an example application of a PFDHA using the proposed wavenumber spectrum method for computing the pdf of the fault displacement and compare the results with the current models of McCalpin (1998) and Petersen et al. (2011) for the displacement pdf. In addition, we show results using an empirical displacement pdf computed from the Wesnousky (2008) data-base. For this example, we used a strike-slip fault with $SRL = 100km$ and a average displacement of $2.0m$. For the wavenumber spectrum method, we create a suite of 60,000 slip profiles by monte carlo sampling of the random terms in the amplitude and phase-derivative models (η_{KC}, η_{Np}, η_{μ}, η_s, and ϵ).

To ensure that the generated profiles are reasonable at the ends of the ruptures, they were constrained so that they go to zero at the edges and that the area under any negative slip to be less than 0.5% of the total area underneath the slip profile. Generated profiles that do not meet these constraints are rejected. A post-evaluation of accepted profiles showed negligible changes in the statistical properties of amplitude and phase-derivative parameters due to the imposed constrains.

The suite of accepted profiles were normalized by AD and SRL to have the same normalization as the published models. Figure 5.12a presents as an example five representative profiles generated for this comparison. Figure 5.12b shows the 16^{th}, 50^{th}, and 84^{th} percentile distributions of all the normalized profiles. These distributions were created by computing and plotting the percentiles independently at every along-strike location. The distributions are symmetric around $x/SRL = 0.5$ even though each individual displacement profile is not symmetric, this happens because the distribution of μ around the global model (eq. 5.25) is symmetric in linear-space which implies that there is equal probability in a generated profile being skewed to the left or right. This figure shows that there is large variability in the displacement profiles for the different realizations and that the shape of the median profile is more similar to a triangular shape for $T(x/SLR)$ as found by Manighetti et al. (2005) than to an ellipical shape as proposed by Wesnousky (2008).

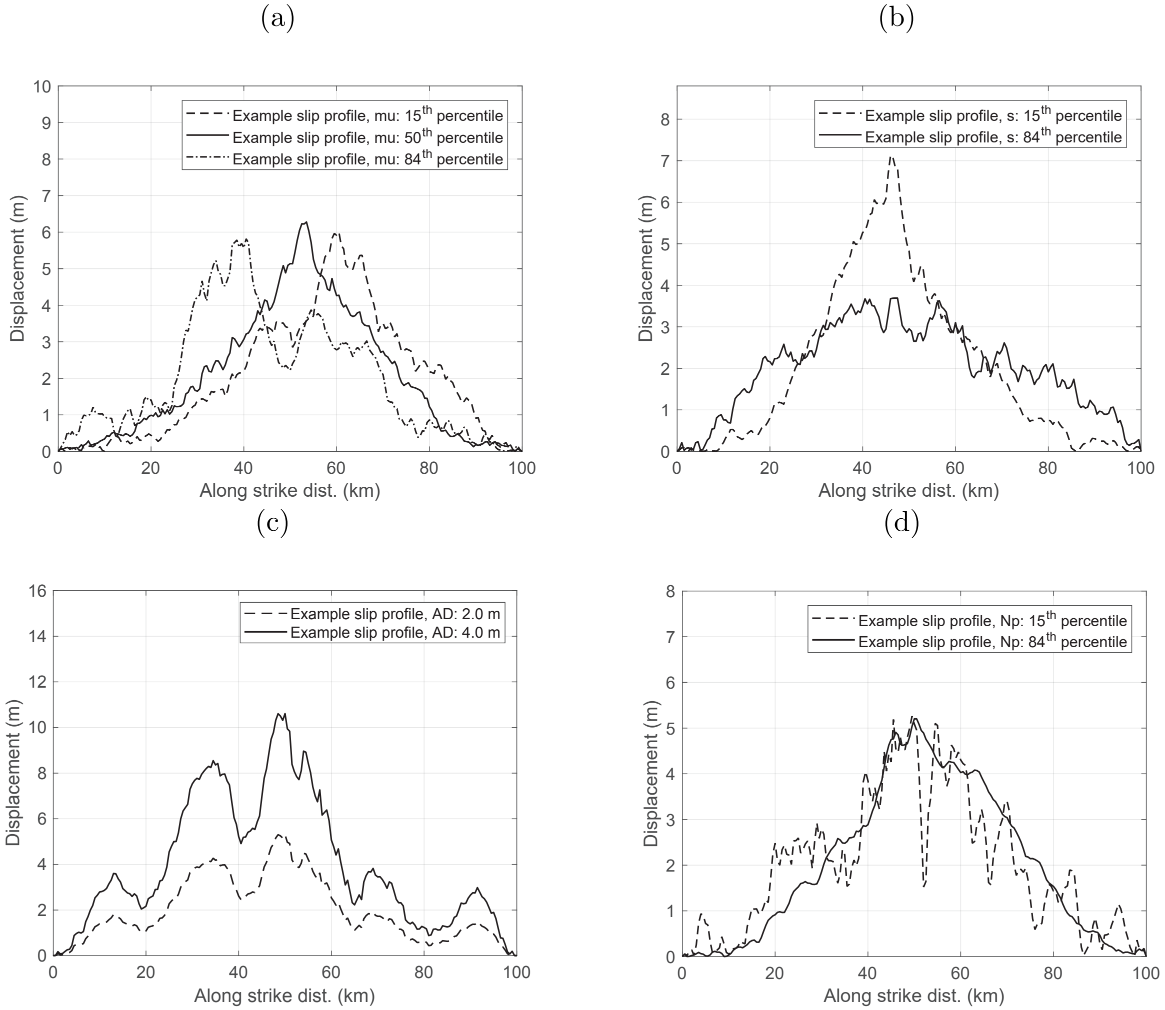

(a) (b) (c) (d)

Figure 5.11: Example applications of proposed model. (a) Representative profiles for the 15^{th}, 50^{th} and 84^{th} fractile on mu; (b) Representative profiles for the 15^{th} and 84^{th} fractile on s; (c) Representative profiles for AD equal to 2 and 4m; (d) Representative profiles for the 16^{th} and 84^{th} fractile on Np

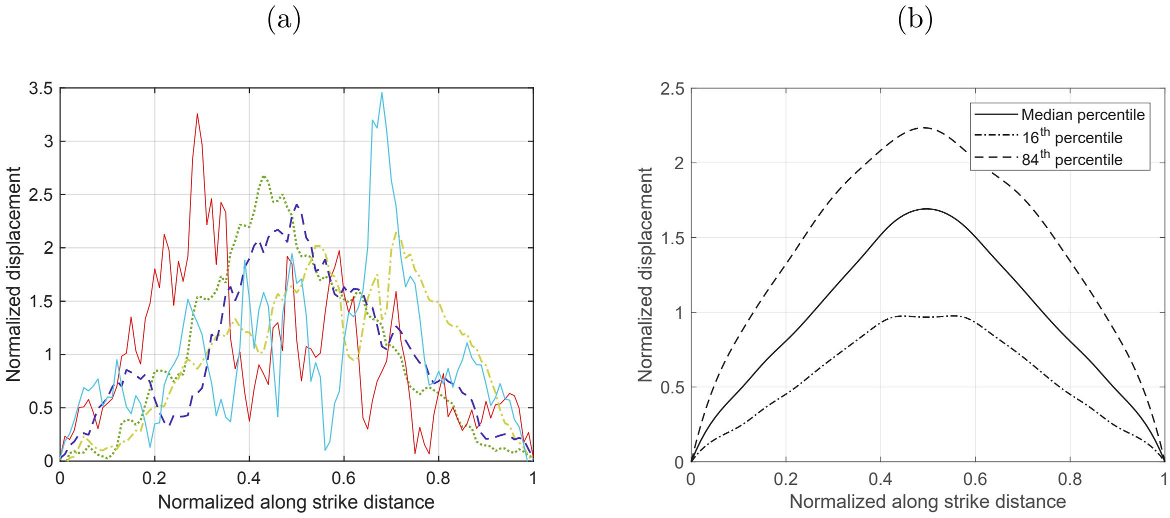


Figure 5.12: Generation of slip profiles by the proposed model. Subfigure (a) presents representative normalized slip profiles generated by the wavenumber model. Subfigure (b) shows the along-strike fractile distribution of all the profiles generated to evaluate the slip variability of the proposed model.

The normalized slip distribution was further evaluated by partitioning the D/AD estimates into different bins based on x/SLR and calculating the cumulative distribution of D/AD for each of these bins. The slip distribution was assumed symmetric along the x-axis. Therefore, the normalized along-strike distance, x/SRL, was folded at 0.5 to limit the range to 0.0-0.5. The same procedure was implemented for the slip profiles in the Wesnousky (2008) data-base to develop an empirical pdf for the normalized slip.

Figure 5.13 compares the distribution of D/AD of the model over the entire rupture with the empirical distribution from the Wesnousky (2008) data-base over the same range and also with the cumulative distribution of all the faults presented in Hemphill-Haley and Weldon (1999). This figure shows that there is a good agreement between the wavenumber model and the data. The maximum likelihood fit of a Log-normal, Gamma, and Weibull distribution is also shown on the same figure; the difference of the log-likelihood (LL) between the Weibull and Normal distributions ($\mathcal{L}_W(\theta|D/AD) - \mathcal{L}_{logN}(\theta|D/AD)$) is 186 and the difference of the LL between the Weibull and Gamma distributions ($\mathcal{L}_W(\theta|D/AD) - \mathcal{L}_\Gamma(\theta|D/AD)$) is 23. This example illustrates that the choice of a log-normal distribution to describe the displacement variability in the space domain may not be ideal because the fit is mainly controlled by the small displacements and it overestimates the D/AD in the upper tail compared to the empirical distributions from the Hemphill-Haley and Weldon (1999) and Wesnousky (2008) data sets. The Gamma and Weibull distributions provide a good fit to the distributions from the empirical data with the Weibull distribution being slightly better.

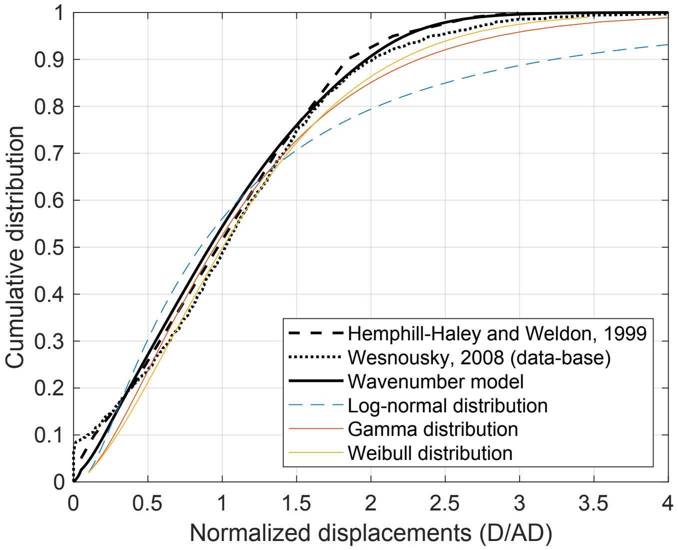
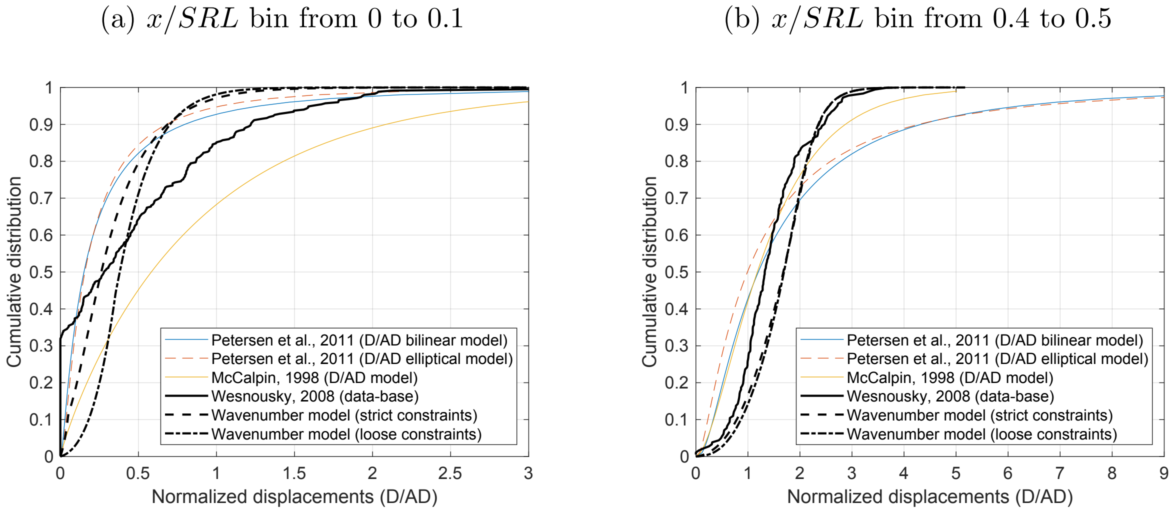


Figure 5.13: Comparison of normalized displacement distribution over the entire fault. The thick solid black lines depict the displacement distribution of the proposed model. The thick dashed and dotted lines represent the distributions of the normalized displacements in Hemphill-Haley and Weldon (1999) and Wesnousky (2008) respectively. Thin lines show the log-normal gamma and Weibull distribution fits to the Wesnousky (2008) normalized displacements.

Figure 5.14 compares the D/AD cumulative distribution based on the wavenumber spectrum model with the models of McCalpin (1998) and Petersen et al. (2011), and with the empirical data from Wesnousky (2008) for two bins of x/SRL: [0,0.1] and [0.45, 0.5]. At the edge of the rupture (e.g. $x/SLR = [0, 0.1]$), the wavenumber spectrum model leads to less variability than the empirical model, but it is in good agreement with the empirical pdf at the center of the rupture (e.g. $x/SLR = [0.4, 0.5]$). The misfit near the edges could be due to limitations of the data base: the end of the ruptures may not have been accurately identified for all earthquakes and there may be small displacements at the ends that haven't been mapped. Alternatively, the phase-derivative distribution at the small wavenumbers may differ from the single distribution that was assumed to be applicable to all wavenumbers in the current model.

As a sensitivity test of the wavenumber spectrum method, we repeated the generation of the slip profiles but with a looser constraint on the slip at the ends of the rupture. In this case, we accepted slip profiles whose displacement at the two ends of the rupture are between 0 and 0.3 of AD. This change increases the median value of the D/AD in the 0 to 0.1 x/SRL bin, but the overall shape of the cumulative distribution did not change significantly because the slip at all locations increases by approximately the same amount. The distribution of


(a) x/SRL bin from 0 to 0.1

(b) x/SRL bin from 0.4 to 0.5

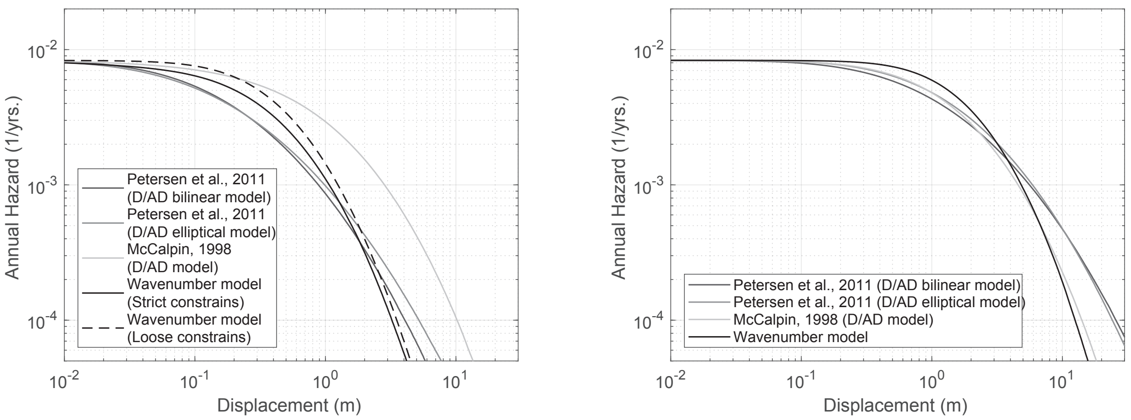
Figure 5.14: Comparison of normalized displacement distribution at different x/SRL bins; subfigure (a) presents the normalized displacements at x/SRL bin from 0 to 0.1 and subifigure (b) the normalized displacements at 0.4 to 0.5 x/SRL bin. Thick dashed black lines depict the displacement distribution of the proposed model and the thick solid black line represent the displacement distribution of the actual slip profiles in Wesnousky (2008) database. Thin lines correspond to displacement distributions of published models.

D/AD at bins away from the ends of the rupture wasn't affected by the looser constraint. Currently, we do not have a method to improve the model in terms of the underestimation of the width of the distribution of slip near the ends of the rupture. This misfit will be revisited in an updated version of the model with better characterized slip data sets, including the sampling near the edges of the ruptures, that are being developed by the UCLA PFDHA project.

5.10.1 Hazard Example

As an illustrative example of the influence of the D/AD distribution in the surface rupture hazard, we perform a simplified PFDHA for the Hayward fault using a simplified source characterization with a single M 7 earthquake with a mean recurrence interval of 120 years. The average slip was estimated with Wells and Coppersmith (1994) relationship: mean slip = 2.3m and standard deviation of the mean slip of 0.36 log10 units. The along-strike variability is estimated using four alternative models: Petersen et al. (2011) billinear and elliptic model, McCalpin (1998) model, and the wavenumber spectrum model.

The hazard calculation is performed for two sites: one at the edge of the fault ($x/SLR = 0.05$) and the other at the center of the rupture ($x/SLR = 0.5$). The distribution of the


(a)

(b)

Figure 5.15: Comparison of surface rupture hazard at x/SLR equal to 0.05 and 0.5, for a M 7 earthquake with 120-year recurrence interval. (a) PFDHA comparison at $x/SRL = 0.05$; (b) PFDHA comparison at $x/SRL = 0.5$

wavenumber spectrum model was defined by piece-wise linear interpolation of the cumulative distributions in the corresponding bins (Figures 5.14a and 5.14a). The total hazard curves are computed using eq. 5.2 and are shown in (Figure 5.15). At the center of the rupture, the hazard curves using the wavenumber spectrum and the McCalpin models are similar and have much steeper slopes than the Petersen models. This is due to the use of a lognormal distribution in the Petersen models. Near the end of the rupture, the hazard from the McCalpin model is shifted to the right compared to the wavenumber model as the taper function in McCalpin is not constrained to go to zero at the ends of the rupture. The median normalized slip at the ends of the rupture in McCalpin is approximately equal to 0.5. The hazard from the Petersen models have flatter slopes, again due to the use of a lognormal distribution.

Using the loose constraint at the ends of the slip profiles leads to larger slips at the ends of the rupture, shifting the hazard to the right. At long return periods, the hazard curves for the strict and loose constraints are similar due to the slightly narrower distribution for the losse constraint case. As the distribution of D/AD becomes wider, the hazard curve becomes flatter, which significantly increases the estimate of displacement at long return periods (Figure 5.15b).

5.11 Conclusions - Discussion

The proposed approach to compute the Fourier spectra, using the regularized-least squares method, is applicable to slip profiles with uneven and potentially biased sampling towards the peaks. The regression results show that the distribution of the normalized displacements is not sensitive to the style of faulting; however this observation needs to be confirmed with a larger data-set. The displacement variability implied by the model is in good agreement with the variability from empirical data in Wesnousky (2008) and Hemphill-Haley and Weldon (1999) except at the ends of the rupture. Currently, the model underestimates the variability of the D/AD distribution near the edges of the rupture. Before modifying the model parameters, new data sets with improved sampling at the ends of the ruptures should be evaluated to determine if this discrepancy is a results of incomplete data in the current data sets or if it is the result of a deficiency in the model.

In the space domain, the results of forward modeling indicate that the D/AD distribution is closer to a Gamma or Weibull distribution rather than log-normal. The D/AD distribution has a large impact on the slope of a PFDHA hazard curve. In particular, use of a log-normal distribution will lead to a much larger slip values for the large return periods compared to the case where Gamma or Weibull distributions is used.

The proposed framework can accommodate a non-ergodic amplitude and phase-derivative model which will be a key issue for PFDHA. Information from past earthquakes along the same fault can be used to limit the range of displacement that is expected in future events and, in combination the site-specific coefficient of variation of slip from Hecker et al. (2013), can be used to reject realizations of slip that are inconsistent with the slip constraints from past earthquakes. A non-ergodic model will further reduce the aleatory variability leading to steeper hazard curves, but the full range of the epistemic uncertainty initial estimated slip from past earthquakes must be considered in a non-ergodic approach.

Currently there is an on-going effort to augment the available surface displacement database with recent events which have been surveyed with high quality techniques. More densely sampled profiles will helps us to better evaluate the high wavenumbers which will affect the extent of spatial correlation. Updated models for the amplitude and phase-derivative scaling will be developed from these expanded data sets as they become available. An additional benefit of this approach it that it allows a straight-forward consideration of the effect of measurement uncertainty in the wavenumber domain. Estimates of uncertainty of the measured slips are not available in the database used in this study, but this information will be part of the UCLA PFDHA data base that is currently being developed.

The main limitations of the current model are that it is regressed on data complied with traditional methodologies that are known to miss off-fault displacements (broad warping and displacement on secondary traces), as shown by Milliner et al. (2016), and does not consider important parameters such as soil conditions. Surface slip profiles that are measured with new techniques (Gold et al., 2015; Milliner et al., 2015; Scott et al., 2018) are likely to include the diffused off-fault displacements that could affect the slip variability and correlation lengths. Furthermore, numerous numerical and laboratory experiments (Bray et al., 1994;

Anastasopoulos et al., 2007; Oettle and Bray, 2017; Garcia and Bray, 2018), as well as field studies (Rockwell et al., 2002; Milliner et al., 2016), have shown that soil effects play an important role in the ratio of horizontal to vertical slip and in the ratio of principal to total displacements at the surface. Future updates of the model will incorporate these effects to better quantify the rupture hazard and develop recommendations on the implementation of a non-ergodic PFDHA analysis.

5.12 Data and Resources

The data used in this study come from Wesnousky (2008) Electronic Supplement (`http://www.seismosoc.org/publications/BSSA_html/bssa_98-4/2007111-esupp/index.html`, last accessed September 2018)

Chapter 6

Conclusions

6.1 Summary

6.1.1 Non-ergodic ground-motion models

This research develops new non-ergodic EAS and PSa GMMs for the state of California. The non-ergodic EAS GMM is based on Bayless and Abrahamson (2019b), BA18, ergodic EAS GMM. The non-ergodic effects are modeled as a Bayesian hierarchical model: the source and site non-ergodic terms are modeled as spatially varying coefficients that are functions of the earthquake and site locations respectively, and the non-ergodic path effects are captured by the cell-specific anelastic attenuation. The main benefits of using an ergodic GMM as a backbone for deriving a non-ergodic GMM are: i) the local data may non be adequate to estimate the scaling of complex terms, and ii) the adoption of the constraints built into the ergodic GMM ensure the proper extrapolation of the ground-motion outside the range of data. The non-ergodic EAS terms are modeled as magnitude independent, as the Fourier transform is linear operator.

The non-ergodic EAS GMM has a 30 to 40% smaller total aleatory standard deviation than BA18. Close to stations and past events, the mean values of the non-ergodic terms deviate from zero to capture the systematic effects and their epistemic uncertainty is small. In areas with sparse data, the epistemic uncertainty of the non-ergodic terms is large, as the systematic effects cannot be determined. Furthermore, the cross-validation test shows that the non-ergodic GMM performs better than BA18 in predicting the ground motion for events that were not part of the regression dataset.

Two non-ergodic PSa GMM are developed: GMM_1 uses the ASK14 (Abrahamson et al., 2014) GMM for the average scaling and is applicable to periods $T_0 = 0.01 - 10sec$, and GMM_2 uses the CY14 (Chiou and Youngs, 2014) GMM for the average scaling and is applicable to periods $T_0 = 0.01 - 5sec$ In both non-ergodic GMMs, the non-ergodic effects are expressed as non-ergodic ratios, that is the difference between log of PSa estimated with RVT and the non-ergodic EAS and the log of PSa estimated with RVT and the ergodic EAS. The magnitude dependence of the non-ergodic PSa ratios is introduced in the RVT step. The

RVT calculations are performed with the V75 PF, the median estimate of D_{a5-85} from AS96 for the ground-motion duration, and the BT15 for the oscillator duration. The RVT components were chosen based on a thorough evaluation of alternative models for the peak factors, ground-motion duration and oscillator duration.

The aleatory variability of the two non-ergodic PSa GMMs is approximately 30 to 35% smaller than the aleatory variability of the corresponding ergodic PSa GMMs. In remote areas with no data, the reduction in aleatory variability is accompanied by an increase of epistemic uncertainty. The smaller aleatory variability has a significant impact on hazard calculations, especially at large return periods.

6.1.2 Embedment factors for the Seattle Region

An approach to adjust a ground-motion model for the site conditions in Pacific North West presented in Chapter 4. The key issue is that the empirical V_{S30} scaling in both shallow crustal and subduction GMMs is not applicable to the Seattle region, that is because, the velocity gradient at the deep part of the Seattle $V_S(z)$ profiles is different to the velocity gradient of representative $V_S(z)$ profiles from both tectonic environments. The site adjustments are expressed in terms of GMM-to-site scale factors; that is the amplification ratios between the host and target profiles. The host profiles are representative of the site conditions in CA and JP, while the target profiles are representative of the site conditions in Seattle. $1D$ site response was used to estimate the amplification of each of the profiles.

6.1.3 Wavenumber fault displacement model

A new methodology to model surface slip profiles and generate potential displacement profiles for use in probabilistic fault rupture hazard analysis is developed in Chapter 5. In this approach, the fault displacement model is formulated in the wavenumber domain. The benefits of creating a model in the wavenumber domain are that it captures the along-strike slip correlation and that it avoids the surface-rupture-length normalization. A regularized Fourier Transform (RFT) is used to compute the Fourier spectra of the displacement profiles. Validation tests showed that RFT successfully estimates the Fourier spectra of profiles with uneven and potentially biased sampling towards the peaks.

The displacement variability implied by the model is, overall, in agreement with with the variability of the empirical data in Wesnousky (2008) and Hemphill-Haley and Weldon (1999). Currently, the model underestimates the D/AD variability near the edges of the rupture.

Due to the along strike slip correlation, the proposed methodology can be used in non-ergodic PFDHA studies. Information from past earthquakes along the same fault can be used to limit the range of displacement that is expected in future events and, in combination the site-specific coefficient of variation of slip from Hecker et al. (2013), can be used to reject realizations of slip that are inconsistent with the slip constraints from past earthquakes. A non-ergodic model will further reduce the aleatory variability leading to steeper hazard curves, but with a shift in the median (either positive or negative). The full range of the epistemic uncertainty of median slip from past earthquakes must be considered in a non-ergodic approach.

The main shortcomings of the current model are that it is regressed on data complied with traditional methodologies that are known to miss off-fault displacements (broad warping and displacement on secondary traces), as shown by Milliner et al. (2016), and does not consider important parameters such as soil conditions. Surface slip profiles that are measured with new techniques (Gold et al., 2015; Milliner et al., 2015; Scott et al., 2018) are likely to include the diffused off-fault displacements that could affect the slip variability and correlation lengths.

6.2 Future Research

This research provides tools for the estimation of the seismic hazard such as, non-ergodic ground-motion models, fault displacement models, and GMM-to-site adjustments. Some limitation and research opportunities related to this work to have been identified, including:

- The non-ergodic GMMs of this study are based on the California subset of the NGAWest2 data-base. As more data become available, the proposed models should be evaluated and potentially expanded with additional non-ergodic terms. Additional non-ergodic terms, such as spatially varying geometrical spreading, are expected to further reduce the epistemic uncertainty of the non-ergodic GMM and result in a more accurate estimation of the ground motion.

- Augmenting the empirical dataset with ground-motion data from broadband earthquake simulations has also the potential to further reduce the epistemic uncertainty, especially in remote areas with sparse data. In addition, numerical simulations would be useful for testing various GMM assumptions. For instance, the path for the cell-specific anelastic attenuation connects the site with the closest point on the rupture. This path was selected because its length is equal to R_{rup}, which is the distance metric used in the anelastic attenuation of ergodic GMMs. However, with the aid of earthquake simulations, other paths connecting a point on the rupture with the site can be tested to find which one is th most appropriate for the cell-specific anelastic attenuation.

- At $T_0 = 1 - 4sec$, the PSa values which were estimated with RVT and the records' EAS exhibit a small positive bias compared to the records' PSa. The origin of this positive bias should be further examined and potential changes to the PF and duration models should be proposed so that the RVT predictions are consistent with the records' PSa.

- The GMM-to-site adjustment factors for the Seattle region have been estimated with a limited number of velocity profiles and a site response analysis that assumed $1-D$ wave propagation. If there are much stronger $3-D$ effects in the Seattle region as compared to California or Japan, this study should be repeated with site response analyses that capture these effects. Furthermore, similar GMM-to-site adjustment factors can be developed for other regions where the depositional environment is not consistent with the V_{S30} scaling of the GMMs.

- The current fault displacement model is based on the Wesnousky (2008) dataset, which is comprised of 32 events. As more surface rupture earthquakes, which are more densely mapped, become available, the displacement model should be reexamined and updated.

- Additionally, the displacement model is developed for the aggregate fault displacements, that is the sum of displacements from sub-parallel ruptures. A model to distribute the aggregate displacements to the individual ruptures needs to developed if the displacement hazard for a specific rupture is of interest.

- The current displacement model underestimates the slip variability near the ends of the fault segments; however, at this point it is not clear if it is due to limitations of the model or if the ends for the fault segments have not been accurately mapped. Because the displacements tapper off at the ends of the segments, it is possible that conventional mapping techniques may prematurely terminate the surface ruptures. Using surface rupture profiles which have been mapped with techniques that are more likely to capture small displacements, such as satellite imaging (Milliner et al., 2015, 2016), LiDAR (Gori et al., 2018), etc., this discrepancy in slip variability should be reevaluated and if deemed necessary the fault displacement model should be updated.

- Numerous laboratory experiments and numerical simulations have shown that the soil conditions affect the magnitude and horizontal-to-vertical ratio of surface displacements. The current model does not capture these effects because the regressed dataset did not contain any description of the soil conditions, but soil effects should be incorporated in future updates of the model to better quantify the rupture hazard.

Bibliography

Abrahamson, C., H.-J. M. Shi, and B. Yang (2016). *Fround-Motion Prediction Equations for Arias Intensity Consistent with the NGA-West2 Ground-Motion Models*. PEER.

Abrahamson, N. A., L. Al-Atik, J. Bayless, A. Dinsick, D. S. Dreger, N. Gregor, N. Kuehn, M. Walling, J. Watson-Lamprey, K. Wooddell, and R. R. Youngs (2015). Southwestern united states ground motion characterization sshac level 3. Technical report, GeoPentech. rev. 2.

Abrahamson, N. A., G. Atkinson, D. M. Boore, Y. Bozorgnia, K. Campbell, B. S. Chiou, I. M. Idriss, W. J. Silva, and R. R. Youngs (2008). Comparisons of the NGA ground-motion relations. *Earthquake Spectra 24*(1), 45–66.

Abrahamson, N. A., N. M. Kuehn, M. Walling, and N. Landwehr (2019, 7). Probabilistic Seismic Hazard Analysis in California Using Nonergodic Ground Motion Models. *Bulletin of the Seismological Society of America 109*(4), 1235–1249.

Abrahamson, N. A. and W. J. Silva (1996). Apendix A: Empirical Ground Motion Models, Description and Validation of the Stochastic Ground Motion Model. Technical report, Brookhaven National Laboratory, New York, NY.

Abrahamson, N. A., W. J. Silva, and R. Kamai (2014). Summary of the ASK14 ground motion relation for active crustal regions. *Earthquake Spectra 30*(3), 1025–1055.

Abrahamson, N. A. and R. R. Youngs (1992). A stable algorithm for regression analyses using the random effects model. *Bulletin of the Seismological Society of America 82*(1), 505–510.

Afshari, K. and J. P. Stewart (2016, 11). Physically Parameterized Prediction Equations for Significant Duration in Active Crustal Regions. *Earthquake Spectra 32*(4), 2057–2081.

Akkar, S. and Z. Çağnan (2010). A local ground-motion predictive model for Turkey, and its comparison with other regional and global ground-motion models. *Bulletin of the Seismological Society of America 100*(6), 2978–2995.

Al Atik, L., N. A. Abrahamson, J. J. Bommer, F. Scherbaum, F. Cotton, and N. M. Kuehn (2010). The Variability of Ground-Motion Prediction Models and Its Components. *Seismological Research Letters 81*(5), 794–801.

ALA (2005). *Guideline for Assessing the Performance of Oil and Natural Gas Pipeline Systems in Natural Hazard and Human Threat Event.* American Lifelines Alliance.

Anastasopoulos, I., G. Gazetas, M. F. Bransby, M. C. R. Davies, and A. El Nahas (2007). Fault Rupture Propagation through Sand: Finite-Element Analysis and Validation through Centrifuge Experiments. *Journal of Geotechnical and Geoenvironmental Engineering 133*(8), 943–958.

Ancheta, T. D., R. B. Darragh, J. P. Stewart, E. Seyhan, W. J. Silva, B. S. Chiou, K. E. Wooddell, R. W. Graves, A. R. Kottke, D. M. Boore, T. Kishida, and J. L. Donahue (2013). PEER NGA-West2 database. Technical report, PEER, Berkeley, CA.

Ancheta, T. D., R. B. Darragh, J. P. Stewart, E. Seyhan, W. J. Silva, B. S. Chiou, K. E. Wooddell, R. W. Graves, A. R. Kottke, D. M. Boore, T. Kishida, and J. L. Donahue (2014). NGA-West2 database. *Earthquake Spectra 30*(3), 989–1005.

Anderson, J. G. and J. N. Brune (1999). Probabilistic seismic hazard analysis without the ergodic assumption. *Seismological Research Letters 70*(1), 19–28.

Anderson, J. G. and S. E. Hough (1984). Spectrum of Acceleration At High Frequencies. *Bulletin of the Seismological Society of America 74*(5), 1969–1993.

Atkinson, G. M. and D. M. Boore (2011). Modifications to existing ground-motion prediction equations in light of new data. *Bulletin of the Seismological Society of America 101*(3), 1121–1135.

Baglio, M. G., N. A. Abrahamson, and G. P. Cimellaro (2017). *Stochastic ground motion method combining a Fourier amplitude spectrum model from a response spectrum with application of phase derivatives distribution prediction.* Ph. D. thesis, Politecnico di Torino.

Baize, S., F. Cinti, C. Costa, T. Dawson, A. Elliott, L. Guerrieri, J. McCalpin, K. Okumura, O. Scotti, M. Takao, P. Villamor, and R. Walker (2016, September). Surface rupture database for seismic hazard assessment - 2015 Report of the kick-off meeting of the SURE database working group.

Bates, D., M. Mächler, B. Bolker, and S. Walker (2015). Fitting linear mixed-effects models using lme4.

Bates, D., M. Mächler, B. Bolker, and S. Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software 67*(1), 251–264.

Bayless, J. and N. A. Abrahamson (2018). Evaluation of the interperiod correlation of ground-motion simulations. *Bulletin of the Seismological Society of America 108*(6), 3413–3430.

Bayless, J. and N. A. Abrahamson (2019a). An Empirical Model for the Interfrequency Correlation of Epsilon for Fourier Amplitude Spectra. *Bulletin of the Seismological Society of America 109*(3), 1058–1070.

Bayless, J. and N. A. Abrahamson (2019b, 10). Summary of the BA18 Ground-Motion Model for Fourier Amplitude Spectra for Crustal Earthquakes in California. *Bulletin of the Seismological Society of America 109*(5), 2088–2105.

Bevington, P. R. and D. K. Robinson (1993). *Data reduction and error analysis for the physical sciences.* McGraw-Hill, New York.

Bindi, D., F. Pacor, L. Luzi, R. Puglia, M. Massa, G. Ameri, and R. Paolucci (2011). Ground motion prediction equations derived from the Italian strong motion database. *Bulletin of Earthquake Engineering 9*(6), 1899–1920.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning.* New York, NY: Springer.

Bommer, J. J., K. J. Coppersmith, R. T. Coppersmith, K. L. Hanson, A. Mangongolo, J. Neveling, E. M. Rathje, A. Rodriguez-Marek, F. Scherbaum, R. Shelembe, P. J. Stafford, and F. O. Strasser (2015). A SSHAC level 3 probabilistic seismic hazard analysis for a new-build nuclear site in South Africa. *Earthquake Spectra 31*(2), 661–698.

Boore, D. M. (1983). Stochastic simulation of high-frequency ground motions based on seismological models of the radiated spectra. *Bulletin of the Seismological Society of America 73*(6), 1865–1894.

Boore, D. M. (2003, 6). Phase Derivatives and Simulation of Strong Ground Motions. *Bulletin of the Seismological Society of America 93*(3), 1132–1143.

Boore, D. M. and W. B. Joyner (1984). A note on the use of random vibration theory to predict peak amplitudes of transient signals. *Bulletin of the Seismological Society of America 74*(5), 2035–2039.

Boore, D. M., J. P. Stewart, E. Seyhan, and G. M. Atkinson (2014). NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes. *Earthquake Spectra 30*(3), 1057–1085.

Boore, D. M. and E. M. Thompson (2012). Empirical improvements for Estimating earthquake response spectra with random-vibration theory. *Bulletin of the Seismological Society of America 102*(2), 761–772.

Boore, D. M. and E. M. Thompson (2015). Revisions to some parameters used in stochastic-method simulations of ground motion. *Bulletin of the Seismological Society of America 105*(2), 1029–1041.

Bora, S. S., F. Cotton, and F. Scherbaum (2019). NGA-West2 empirical fourier and duration models to generate adjustable response spectra. *Earthquake Spectra 55*(1), 61–93.

Bora, S. S., F. Scherbaum, N. M. Kuehn, P. J. Stafford, and B. Edwards (2015). Development of a response spectral ground-motion prediction equation (GMPE) for seismic-hazard analysis from empirical fourier spectral and duration models. *Bulletin of the Seismological Society of America 105*(4), 2192–2218.

Bozorgnia, Y., N. A. Abrahamson, L. Al Atik, T. D. Ancheta, G. M. Atkinson, J. W. Baker, A. Baltay, D. M. Boore, K. W. Campbell, B. S. Chiou, R. Darragh, S. Day, J. Donahue, R. W. Graves, N. Gregor, T. C. Hanks, I. M. Idriss, R. Kamai, T. Kishida, A. R. Kottke, S. A. Mahin, S. Rezaeian, B. Rowshandel, E. Seyhan, S. Shahi, T. Shantz, W. J. Silva, P. Spudich, J. P. Stewart, J. Watson-Lamprey, K. Wooddell, and R. R. Youngs (2014). NGA-West2 research project. *Earthquake Spectra 30*(3), 973–987.

Bragato, P. L. and D. Slejko (2005). Empirical ground-motion attenuation relations for the eastern Alps in the magnitude range 2.5-6.3. *Bulletin of the Seismological Society of America 95*(1), 252–276.

Bray, J. D., R. B. Seed, L. S. Cluff, and H. Bolton Seed (1994). Earthquake fault rupture propagation through soil. *Journal of Geotechnical Engineering 120*(3), 543–561.

Brune, J. N. (1970). Tectonic stress and the spectra of seismic shear waves from earthquakes. *Journal of Geophysical Research 75*(26), 4997–5009.

Brune, J. N. (1971). Correction to Tectonic stress and the spectra of seismic shear waves from earthquakes. *Journal of Geophysical Research 76*(20), 1971.

Bussas, M., C. Sawade, N. M. Kuehn, T. Scheffer, and N. Landwehr (2017). Varying-coefficient models for geospatial transfer learning. *Machine Learning 106*(9-10), 1419–1440.

CALTRANS (2013). *Memo to Designers 20-10, Fault Rupture.* California Dept. of Transportation.

Campbell, K. W. (2003). Prediction of strong ground motion using the hybrid empirical method and its use in the development of ground-motion (attenuation) relations in Eastern North America. *Bulletin of the Seismological Society of America 93*(3), 1012–1033.

Campbell, K. W. and Y. Bozorgnia (2014). NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra. *Earthquake Spectra 30*(3), 1087–1114.

Cartwright, D. E. (1956). The statistical distribution of the maxima of a random function. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences 237*(1209), 212–232.

Chiou, B. S., R. Youngs, N. A. Abrahamson, and K. Addo (2010). Ground-motion attenuation model for small-to-moderate shallow crustal earthquakes in California and its implications on regionalization of ground-motion prediction models. *Earthquake Spectra 26*(4), 907–926.

Chiou, B. S. and R. R. Youngs (2014, 8). Update of the Chiou and Youngs NGA Model for the Average Horizontal Component of Peak Ground Motion and Response Spectra. *Earthquake Spectra 30*(3), 1117–1153.

Coppersmith, K., J. Bommer, K. Hanson, J. Unruh, R. Coppersmith, L. Wolf, R. Youngs, A. Rodriguez-Marek, L. Al Atik, G. Toro, et al. (2014). *Hanford sitewide probabilistic seismic hazard analysis.* Richland, Washington: PNNL-23361 Pacific Northwest National Laboratory.

Danciu, L. and G.-A. Tselentis (2007, 2). Engineering Ground-Motion Parameters Attenuation Relationships for Greece. *Bulletin of the Seismological Society of America 97*(1B), 162–183.

Davenport, A. G. (1964). Note on the Distribution of the Largest Value of a Random Function With Application To Gust Loading. *Proceedings of the Institution of Civil Engineers 28*(2), 187–196.

Dawood, H. M. and A. Rodriguez-Marek (2013). A Method for including path effects in ground-motion prediction equations: An example using the Mw 9.0 Tohoku earthquake aftershocks. *Bulletin of the Seismological Society of America 103*(2 B), 1360–1372.

Der Kiureghian, A. (1980). Structural response to stationary excitation. *Journal of the Engineering Mechanics Division 106*(6), 1195–1213.

Dieterich, J. H. and D. E. Smith (2009). Nonplanar faults: Mechanics of slip and off-fault damage. *Pure and Applied Geophysics 166*(10-11), 1799–1815.

Douglas, J., S. Akkar, G. Ameri, P. Y. Bard, D. Bindi, J. J. Bommer, S. S. Bora, F. Cotton, B. Derras, M. Hermkes, N. M. Kuehn, L. Luzi, M. Massa, F. Pacor, C. Riggelsen, M. A. Sandıkkaya, F. Scherbaum, P. J. Stafford, and P. Traversa (2014). Comparisons among the five ground-motion models developed using RESORCE for the prediction of response spectral accelerations due to earthquakes in Europe and the Middle East. *Bulletin of Earthquake Engineering 12*(1), 341–358.

Dunham, E., D. Belanger, L. Cong, and J. E. Kozdon (2011). Earthquake ruptures with strongly rate-weakening friction and off-fault plasticity, part 2: Nonplanar faults. *Bull. Seismol. Soc. Am. 101*(5), 2308–2322.

Eberhart-Phillips, D. (2016). Northern California seismic attenuation: 3D Qp and Qs models. *Bulletin of the Seismological Society of America 106*(6), 2558–2573.

Elzhov, T. V., K. M. Mullen, A.-N. Spiess, and B. Bolker (2016). minpack.lm: R interface to the levenberg-marquardt nonlinear least-squares algorithm found in minpack, plus support for bounds. R package version 1.2-1.

Garcia, F. E. and J. D. Bray (2018). Distinct element simulations of earthquake fault rupture through materials of varying density. *Soils and foundations 58*(4), 986–1000.

Gold, R. D., N. G. Reitman, R. W. Briggs, W. D. Barnhart, G. P. Hayes, and E. Wilson (2015). On- and off-fault deformation associated with the September 2013 Mw7.7 Balochistan earthquake: Implications for geologic slip rate measurements. *Tectonophysics 660*(September 2013), 65–78.

Gori, S., E. Falcucci, F. Galadini, P. Zimmaro, A. Pizzi, R. E. Kayen, B. N. Lingwall, M. Moro, M. Saroli, G. Fubelli, A. Di Domenica, and J. P. Stewart (2018). Surface faulting caused by the 2016 central Italy seismic sequence: Field mapping and LiDAR/UAV imaging. *Earthquake Spectra 34*(4), 1585–1610.

Goulet, C., A. Kottke, D. Boore, Y. Bozorgnia, J. Hollenback, T. Kishida, A. Der Kiureghian, O. Ktenidou, N. Kuehn, E. Rathje, et al. (2018). Effective amplitude spectrum (eas) as a metric for ground motion modeling using fourier amplitudes. In *2018 Seismology of the Americas Meeting.*

Gregor, N., N. A. Abrahamson, G. M. Atkinson, D. M. Boore, Y. Bozorgnia, K. W. Campbell, B. S. Chiou, I. M. Idriss, R. Kamai, E. Seyhan, W. J. Silva, J. P. Stewart, and R. R. Youngs (2014). Comparison of NGA-West2 GMPEs. *Earthquake Spectra 30*(3), 1179–1197.

Hanks, T. C. and R. K. McGuire (1981). The character of high-frequency strong ground motion. *Bull. Seism. Soc. Amer. 71*, 1897–1919.

Hecker, S., N. A. Abrahamson, and K. E. Wooddell (2013). Variability of displacement at a point: Implications for earthquake-size distribution and rupture hazard on faults. *Bulletin of the Seismological Society of America 103*(2 A), 651–674.

Hemphill-Haley, M. A. and R. J. Weldon (1999). Estimating prehistoric earthquake magnitude from point measurements of surface rupture. *Bulletin of the Seismological Society of America 89*(5), 1264–1279.

Hiemer, S., F. Scherbaum, D. Roessler, and N. M. Kuehn (2011, 5). Determination of 0 and Rock Site from Records of the 2008/2009 Earthquake Swarm in Western Bohemia. *Seismological Research Letters 82*(3), 387–393.

Kamai, R., N. A. Abrahamson, and W. J. Silva (2013). PACIFIC EARTHQUAKE ENGINEERING Nonlinear Horizontal Site Response for the NGA-West2 Project San Francisco , California. Technical Report May, PEER, Berkeley, CA.

Kamai, R., N. A. Abrahamson, and W. J. Silva (2014). Nonlinear horizontal site amplification for constraining the NGA-West2 GMPEs. *Earthquake Spectra 30*(3), 1223–1240.

Kayen, R., S. Gori, B. Lingwall, F. Galadini, E. Falcucci, K. Franke, J. Stewart, and P. Zimmaro (2018). Mt. vettore fault zone rupture: Lidar-and uas-based structure-from-motion computational imaging. In *16th European Conference on Earthquake Engineering.* EAEE and HSEE.

Kempton, J. J. and J. P. Stewart (2006). Prediction equations for significant duration of earthquake ground motions considering site and near-source effects. *Earthquake Spectra 22*(4), 985–1013.

Kottke, A. (2019). pySRA Site Response Analysis Toolkit for Python.

Kottke, A. (2020, January). arkottke/pyrvt v0.7.2.

Kottke, A. R., N. A. Abrahamson, D. M. Boore, Y. Bozorgnia, C. Goulet, J. Hollenback, T. Kishida, A. Der Kiureghian, O.-J. Ktenidou, N. M. Kuehn, E. M. Rathje, W. J. Silva, E. M. Thompson, and X. Wang (2018). Selection of Random Vibration Procedures for the NGA-East Project. Technical report, Pacific Earthquake Engineering Center, Berkeley, CA.

Ktenidou, O.-J., F. Cotton, N. A. Abrahamson, and J. G. Anderson (2014). Taxonomy of κ: A review of definitions and estimation approaches targeted to applications. *Seismological Research Letters 85*(1), 135–146.

Kuehn, N. M., N. A. Abrahamson, and M. A. Walling (2019). Incorporating Nonergodic Path Effects into the NGA-West2 Ground-Motion Prediction Equations. *Bulletin of the Seismological Society of America 109*(2), 575–585.

Kuehn, N. M., Y. Bozorgnia, K. W. Campbell, and N. Gregor (2020). Partially Non-Ergodic Ground-Motion Model for Subduction Regions using the NGA-Subduction Database. Technical Report September, PEER.

Kutner, M. H., C. J. Nachtsheim, J. Neter, W. Li, et al. (2005). *Applied linear statistical models*, Volume 5. McGraw-Hill Irwin New York.

Landwehr, N., N. M. Kuehn, T. Scheffer, and N. A. Abrahamson (2016). A nonergodic ground-motion model for California with spatially varying coefficients. *Bulletin of the Seismological Society of America 106*(6), 2574–2583.

Lin, P. S., B. S. Chiou, N. A. Abrahamson, M. Walling, C. T. Lee, and C. T. Cheng (2011). Repeatable source, site, and path effects on the standard deviation for empirical ground-motion prediction models. *Bulletin of the Seismological Society of America 101*(5), 2281–2295.

Liu, L. and S. Pezeshk (1999). An improvement on the estimation of pseudoresponse spectral velocity using RVT method. *Bulletin of the Seismological Society of America 89*(5), 1384–1389.

Manighetti, I., M. Campillo, C. G. Sammis, P. M. Mai, and G. King (2005). Evidence for self-similar, triangular slip distributions on earthquakes: Implications for earthquake and fault mechanics. *Journal of Geophysical Research: Solid Earth 110*(5), 1–25.

Matsukawa, K., M. Watabe, N. Theofanopulos, and M. Tohdo (1987). Phase characteristics of earthquake ground motions and those applications to synthetic ones. In *Transactions of the 9th international conference on structural mechanics in reactor technology. Vol. K1.*

McCalpin, J. P. (1998, March). *Statistics of Paleoseismic Data, Final Technical Report.* U.S. Geological Survey. Contract 1434- HQ-96-GR-02752.

McGuire, R. K. (2004). *Seismic hazard and risk analysis.* Earthquake Engineering Research Institute.

Milliner, C. W. D., J. F. Dolan, J. Hollingsworth, S. Leprince, and F. Ayoub (2016). Comparison of coseismic near-field and off-fault surface deformation patterns of the 1992 Mw7.3 Landers and 1999 Mw7.1 Hector Mine earthquakes: Implications for controls on the distribution of surface strain. *Geophysical Research Letters 43*(19), 115–10.

Milliner, C. W. D., J. F. Dolan, J. Hollingsworth, S. Leprince, F. Ayoub, and C. G. Sammis (2015, 5). Quantifying near-field and off-fault deformation patterns of the 1992 Mw 7.3 Landers earthquake. *Geochemistry, Geophysics, Geosystems 16*(5), 1577–1598.

Moss, R. E. S. and Z. E. Ross (2011). Probabilistic fault displacement hazard analysis for reverse faults. *Bulletin of the Seismological Society of America 101*(4), 1542–1553.

Nissen, E., T. Maruyama, J. Ramón Arrowsmith, J. R. Elliott, A. K. Krishnan, M. E. Oskin, and S. Saripalli (2014, 11). Coseismic fault zone deformation revealed with differential lidar: Examples from Japanese Mw 7 intraplate earthquakes. *Earth and Planetary Science Letters 405*, 244–256.

Oettle, N. K. and J. D. Bray (2017). Numerical Procedures for Simulating Earthquake Fault Rupture Propagation. *International Journal of Geomechanics 17*(1), 04016025.

Ohsaki, Y. (1979, 9). On the significance of phase content in earthquake ground motions. *Earthquake Engineering & Structural Dynamics 7*(5), 427–439.

Oskin, M. E., J. R. Arrowsmith, A. H. Corona, A. J. Elliott, J. M. Fletcher, E. J. Fielding, P. O. Gold, J. J. G. Garcia, K. W. Hudnut, J. Liu-Zeng, and O. J. Teran (2012, feb). Near-Field Deformation from the El Mayor-Cucapah Earthquake Revealed by Differential LIDAR. *Science (80-.). 335*(6069), 702–705.

Petersen, M. D., T. E. Dawson, R. Chen, T. Cao, C. J. Wills, D. P. Schwartz, and A. D. Frankel (2011). Fault displacement hazard for strike-slip faults. *Bulletin of the Seismological Society of America 101*(2), 805–825.

Phillips, W. S., K. M. Mayeda, and L. Malagnini (2014). How to Invert Multi-Band, Regional Phase Amplitudes for 2-D Attenuation and Source Parameters: Tests Using the USArray. *Pure and Applied Geophysics 171*(3-5), 469–484.

Power, M., B. S. Chiou, N. A. Abrahamson, Y. Bozorgnia, T. Shantz, and C. Roblee (2008). An overview of the NGA project. *Earthquake Spectra 24*(1), 3–21.

R Core Team (2018). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

R Core Team (2020). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Rasmussen, C. E. and C. K. I. Williams (2006). Gaussian processes for machine learning. *The MIT Press, Cambridge, MA, USA 38*, 715–719.

Rockwell, T. K. and Y. Klinger (2013, 4). Surface Rupture and Slip Distribution of the 1940 Imperial Valley Earthquake, Imperial Fault, Southern California: Implications for Rupture Segmentation and Dynamics. *Bulletin of the Seismological Society of America 103*(2A), 629–640.

Rockwell, T. K., S. Lindvall, T. Dawson, R. Langridge, W. Lettis, and Y. Klinger (2002). Lateral offsets on surveyed cultural features resulting from the 1999 İzmit and Düzce earthquakes, Turkey. *Bull. Seism. Soc. Am. 92*(1), 79–94.

Scott, C. P., J. R. Arrowsmith, E. Nissen, L. Lajoie, T. Maruyama, and T. Chiba (2018). The M7 2016 Kumamoto, Japan, Earthquake: 3-D Deformation Along the Fault and Within the Damage Zone Constrained From Differential Lidar Topography. *Journal of Geophysical Research: Solid Earth 123*(7), 6138–6155.

Simpson, D., H. Rue, A. Riebler, T. G. Martins, and S. H. Sørbye (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science 32*(1), 1–28.

Somerville, P., K. Irikura, R. Graves, S. Sawada, D. Wald, N. A. Abrahamson, Y. Iwasaki, T. Kagawa, N. Smith, and A. Kowada (1999). Characterizing Crustal Earthquake Slip

Models for the Prediction of Strong Ground Motion. *Seismological Research Letters 70*(1), 59–80.

Stafford, P. J. (2015, 12). Extension of the Random-Effects Regression Algorithm to Account for the Effects of Nonlinear Site Response. *Bulletin of the Seismological Society of America 105*(6), 3196–3202.

Stan Development Team (2019). The Stan Core Library. Version 2.25.0.

Stephenson, W. J., N. G. Reitman, and S. J. Angster (2017). P- and S-wave velocity models incorporating the Cascadia subduction zone for 3D earthquake ground motion simulations—Update for Open-File Report 2007–1348. Technical Report Version 1.6, USGS.

Stoffa, P. L., P. Buhl, and G. M. Bryan (1974, 8). The Application of Homomorphic Deconvolution To Shallow-Water Marine Seismology—Part I: Models. *Geophysics 39*(4), 401–416.

Toro, G. R. (1995). Probabilistic models of site velocity profiles for generic and site-specific ground-motion amplification studies. Technical Report No. 779574, Brookhaven National Laboratory, Upton, N.Y.

Tromans, I. J., G. Aldama-Bustos, J. Douglas, A. Lessi-Cheimariou, S. Hunt, M. Daví, R. M. W. Musson, G. Garrard, F. O. Strasser, and C. Robertson (2019, 1). Probabilistic seismic hazard assessment for a new-build nuclear power plant site in the UK. *Bulletin of Earthquake Engineering 17*(1), 1–36.

Van Rossum, G. and F. L. Drake (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

Vanmarcke, E. H. (1972). Properties of spectral moments with applications to random vibration. *ASCE Journal of the Engineering Mechanics Division 98*(EM2), 425–446.

Vanmarcke, E. H. (1975, 3). On the Distribution of the First-Passage Time for Normal Stationary Random Processes. *Journal of Applied Mechanics 42*(1), 215–220.

Vanmarcke, E. H. (1976). Structural Response to Earthquakes. In *Developments in Geotechnical Engineering*, pp. 287–337. Elsevier Inc.

Wedmore, L. N. J., L. C. Gregory, K. J. W. McCaffrey, H. Goodall, and R. J. Walters (2019, 3). Partitioned Off-Fault Deformation in the 2016 Norcia Earthquake Captured by Differential Terrestrial Laser Scanning. *Geophysical Research Letters 46*(6), 3199–3205.

Wells, D. L. and K. J. Coppersmith (1994). New Empirical Relationships among Magnitude, Rupture Length, Rupture Width, Rupture Area, and Surface Displacement. *Bulletin of the Seismological Society of America 84*(4), 974–1002.

Wesnousky, S. G. (2008). Displacement and geometrical characteristics of earthquake surface ruptures: Issues and implications for seismic-hazard analysis and the process of earthquake rupture. *Bulletin of the Seismological Society of America 98*(4), 1609–1632.

Wirth, E., S. Chang, and A. Frankel (2018, September). 2018 report on incorporating sedimentary basin response into the design of tall buildings in seattle, washington. Technical Report Open-File Report 2018–1149, United States Geologic Survey.

Yagoda-Biran, G., J. G. Anderson, H. Miyake, and K. Koketsu (2015). Between-event variance for large repeating earthquakes. *Bulletin of the Seismological Society of America 105*(4), 2023–2040.

Yokoyama, T., N. Theofanopoulos, and M. Watabe (1988). Distribution of phase differences in relation to the earthquake magnitude, distance to the fault and local soil conditions. In *Proc. 9th WCEE, Abstr.*, Volume 2, pp. 55—-55.

Youngs, R. R., W. J. Arabasz, R. E. Anderson, A. R. Ramelli, J. P. Ake, D. B. Slemmons, J. P. McCalpin, D. I. Doser, C. J. Fridrich, F. H. Swan, A. M. Rogers, J. C. Yount, L. W. Anderson, K. D. Smith, R. L. Bruhn, P. L. K. Knuepfer, R. B. Smith, C. M. DePolo, D. W. O'Leary, K. J. Coppersmith, S. K. Pezzopane, D. P. Schwartz, J. W. Whitney, S. S. Olig, and G. R. Toro (2003). A methodology for probabilistic fault displacement hazard analysis (PFDHA). *Earthquake Spectra 19*(1), 191–219.

Appendix A

Supplemental Material for Chapter 2

A.1 Comparison of different significant duration intervals for D_{gm}

Figures A.1 to A.5 show the residuals between the records' PSa and the PSa estimated with RVT. The RVT PSa were estimated with $V75$ peak factors, records' actual duration for D_{gm}, and $BT15$ for D_{rms}. D_{gm} is equal to: D_{a5-75} in Figure A.1, D_{a5-80} in Figure A.2, D_{a5-85} in Figure A.3, D_{a5-90} in Figure A.4, and D_{a5-95} in Figure A.5.

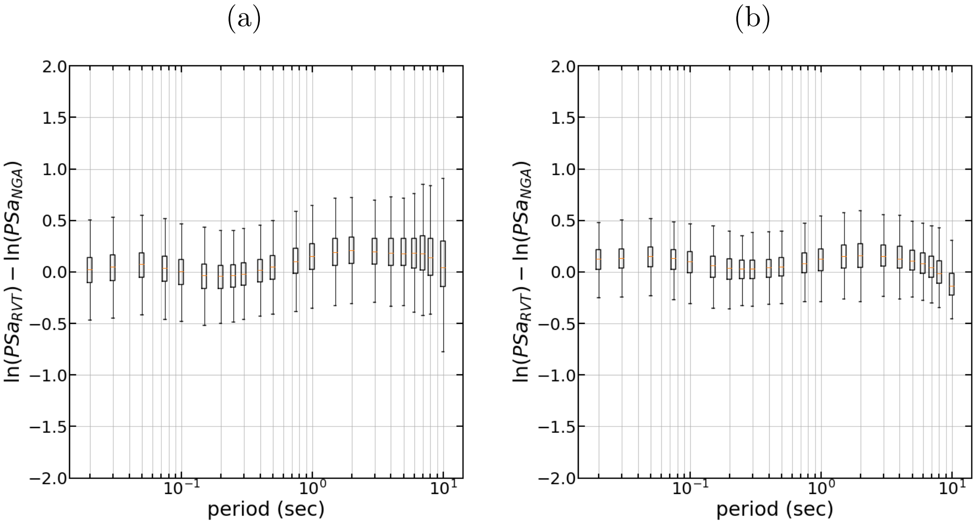

(a)
(b)

Figure A.1: Residuals between the records' PSa and the RVT PSa. V75 PF, records' $D_{a0.05-0.75}$ as D_{gm}, and BT15 D_{rms}. (a) residuals for all M, (b) residuals for $M > 5$

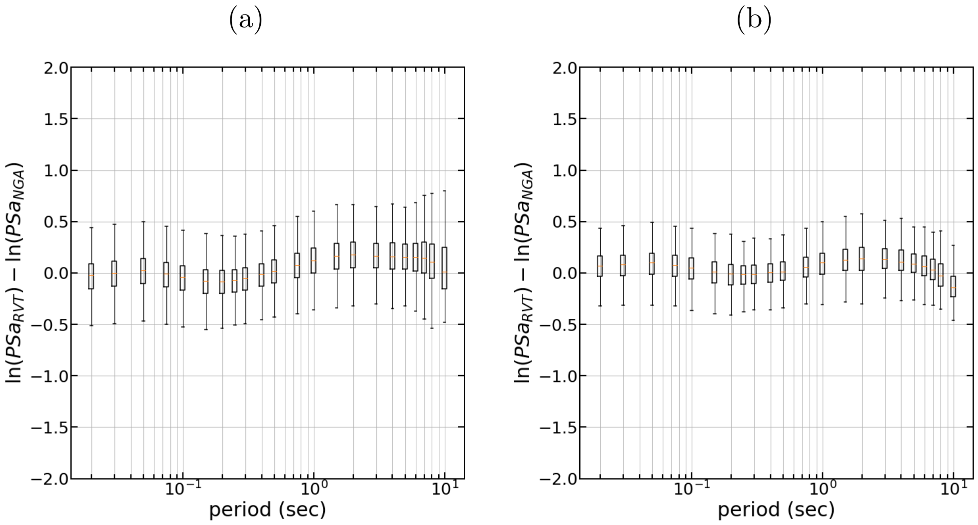

(a) (b)

Figure A.2: Residuals between the records' PSa and the RVT PSa. V75 PF, records' $D_{a0.05-0.80}$ as D_{gm}, and BT15 D_{rms}. (a) residuals for all M, (b) residuals for $M > 5$

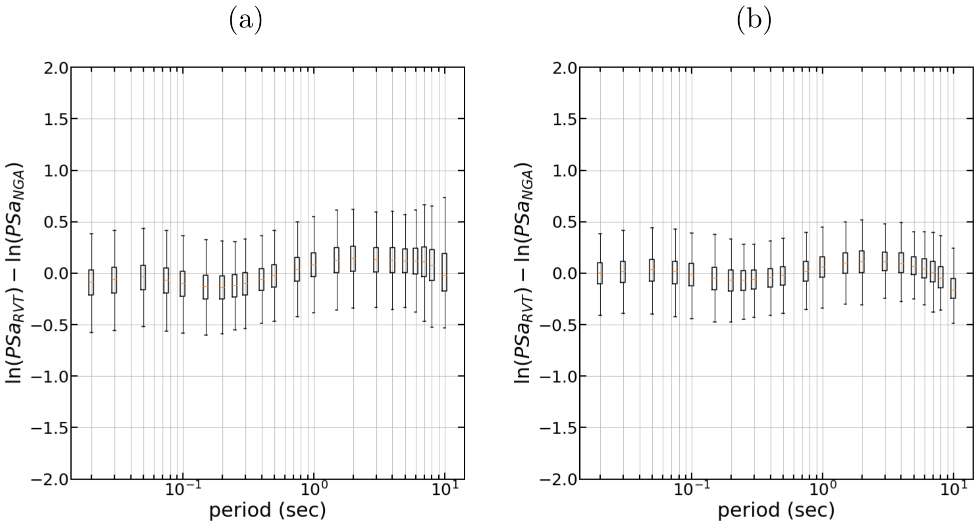

(a) (b)

Figure A.3: Residuals between the records' PSa and the RVT PSa. V75 PF, records' $D_{a0.05-0.85}$ as D_{gm}, and BT15 D_{rms}. (a) residuals for all M, (b) residuals for $M > 5$

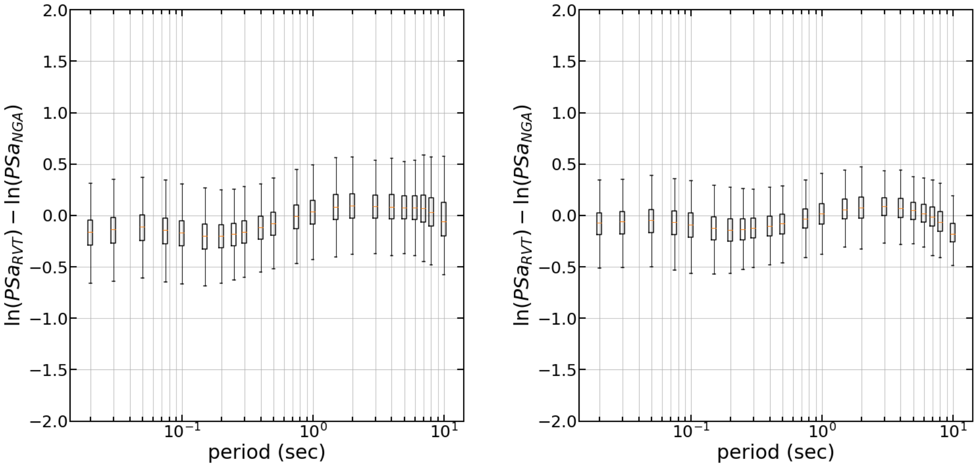

(a) (b)

Figure A.4: Residuals between the records' PSa and the RVT PSa. V75 PF, records' $D_{a0.05-0.90}$ as D_{gm}, and BT15 D_{rms}. (a) residuals for all M, (b) residuals for $M > 5$

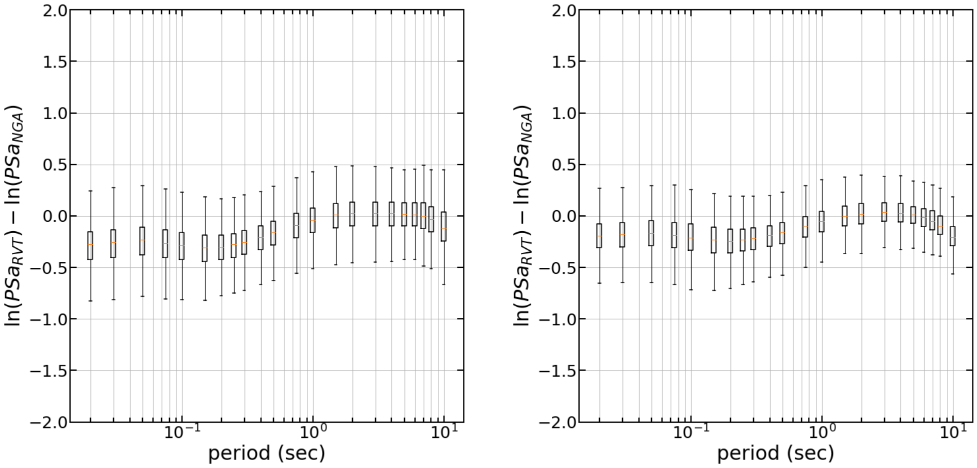

(a) (b)

Figure A.5: Residuals between the records' PSa and the RVT PSa. V75 PF, records' $D_{a0.05-0.95}$ as D_{gm}, and BT15 D_{rms}. (a) residuals for all M, (b) residuals for $M > 5$

A.2 Comparison of different duration models for D_{gm}

Figures A.6 to A.12 show the residuals between the records' PSa and the PSa estimated with RVT. The RVT PSa were estimated with $V75$ peak factors, the D_{gm} estimated with a duration GMM, and $BT15$ for D_{rms}. D_{gm} is based on: KS06 D_{a5-75} in Figure A.6, KS06 D_{a5-95} in Figure A.7, KS06 D_{v5-75} in Figure A.8, the KS06 D_{v5-95} in Figure A.9, AS16 D_{a5-75} in Figure A.10, AS16 D_{a5-95} in Figure A.11, and AS16 2 D_{a20-80} in Figure A.12.

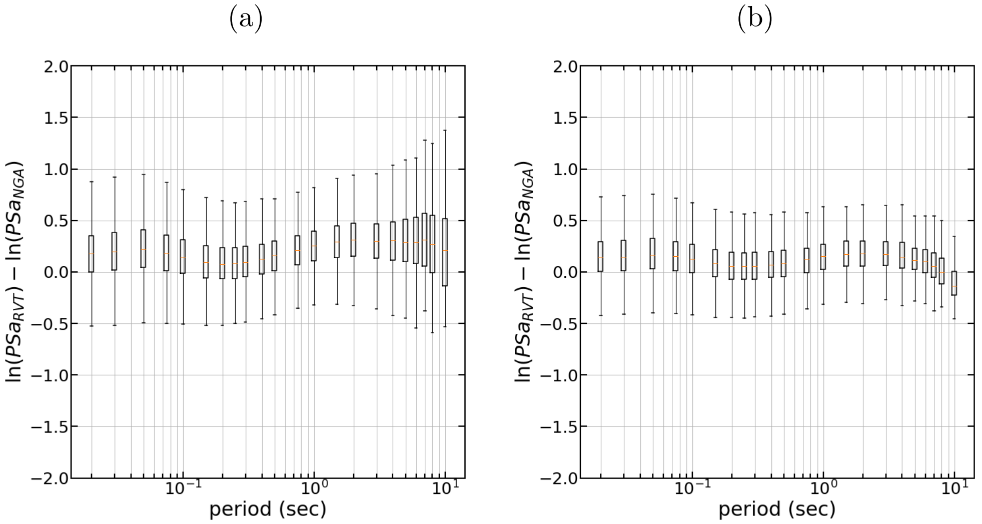


Figure A.6: Residuals between the records' PSa and the RVT PSa. V75 PF, $D_{a0.05-0.75}$ from KS06 as D_{gm}, and BT15 D_{rms}. (a) residuals for all M, (b) residuals for $M > 5$

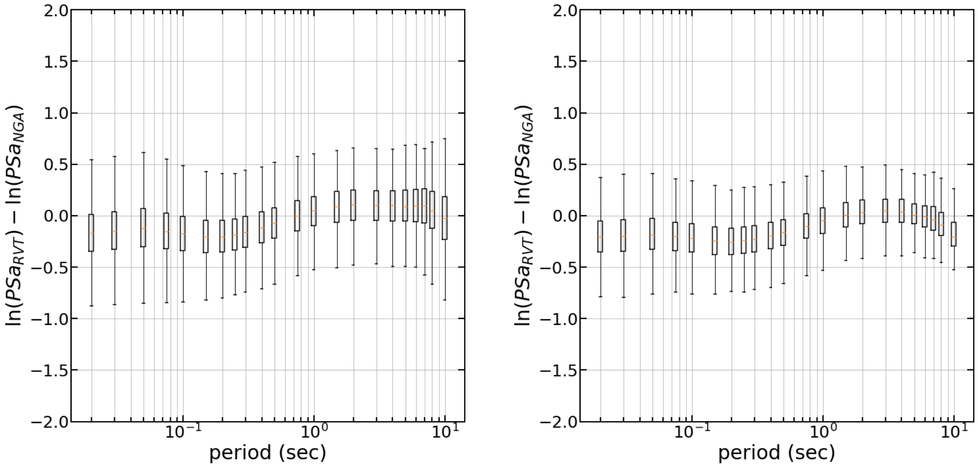

(a) (b)

Figure A.7: Residuals between the records' PSa and the RVT PSa. V75 PF, $D_{a0.05-0.95}$ from KS06 as D_{gm}, and BT15 D_{rms}. (a) residuals for all M, (b) residuals for $M > 5$

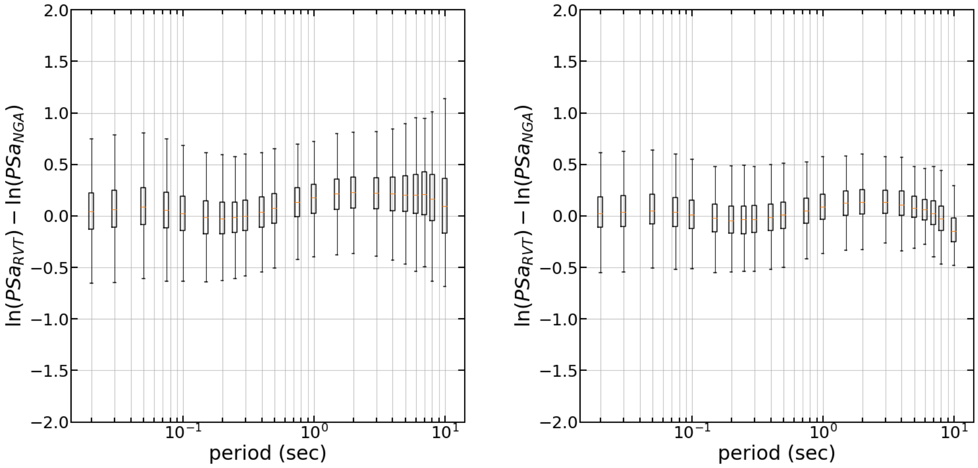

(a) (b)

Figure A.8: Residuals between the records' PSa and the RVT PSa. V75 PF, $D_{v0.05-0.75}$ from KS06 as D_{gm}, and BT15 D_{rms}. (a) residuals for all M, (b) residuals for $M > 5$

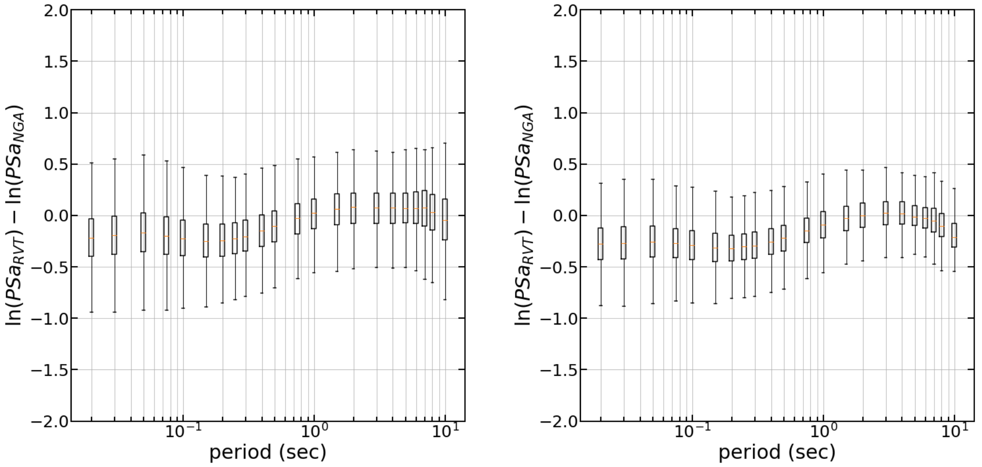

(a) (b)

Figure A.9: Residuals between the records' PSa and the RVT PSa. V75 PF, $D_{v0.05-0.95}$ from KS06 as D_{gm}, and BT15 D_{rms}. (a) residuals for all M, (b) residuals for $M > 5$

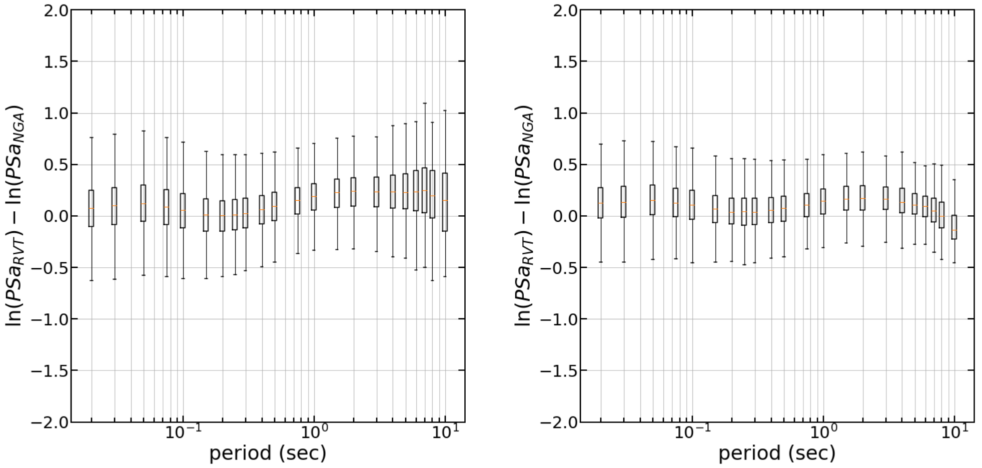

(a) (b)

Figure A.10: Residuals between the records' PSa and the RVT PSa. V75 PF, $D_{a0.05-0.75}$ from KS06 as D_{gm}, and BT15 D_{rms}. (a) residuals for all M, (b) residuals for $M > 5$

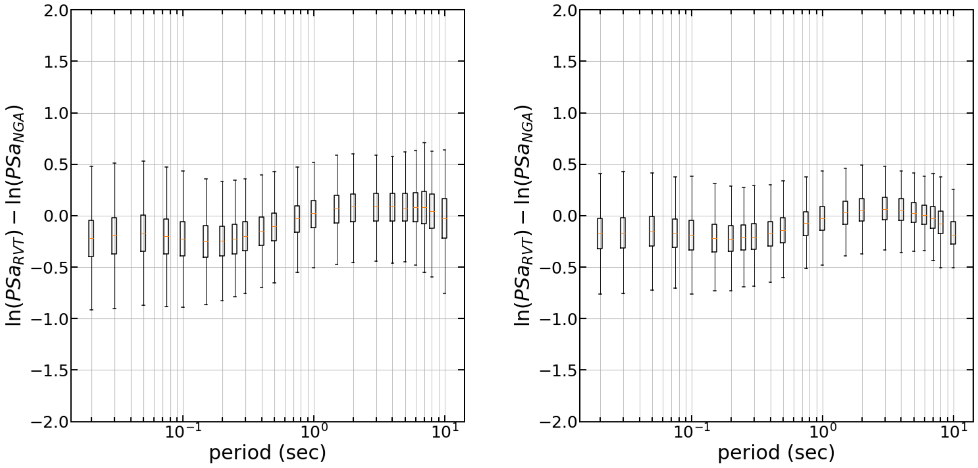

(a) (b)

Figure A.11: Residuals between the records' PSa and the RVT PSa. V75 PF, $D_{a0.05-0.95}$ from KS06 as D_{gm}, and BT15 D_{rms}. (a) residuals for all M, (b) residuals for $M > 5$

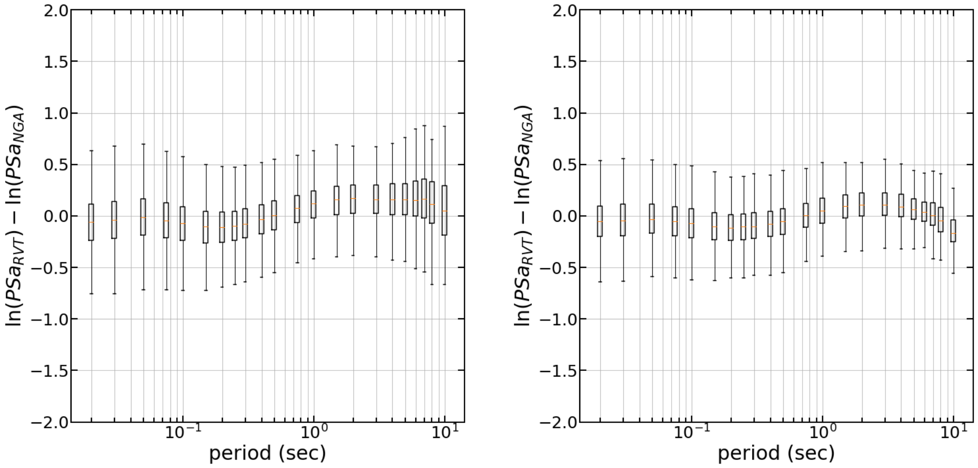

(a) (b)

Figure A.12: Residuals between the records' PSa and the RVT PSa. V75 PF, 2 $D_{a0.20-0.95}$ from KS06 as D_{gm}, and BT15 D_{rms}. (a) residuals for all M, (b) residuals for $M > 5$

Appendix B

Supplemental Material for Chapter 5

B.1 Impact of regularization parameters α and n in computing the Fourier Spectra

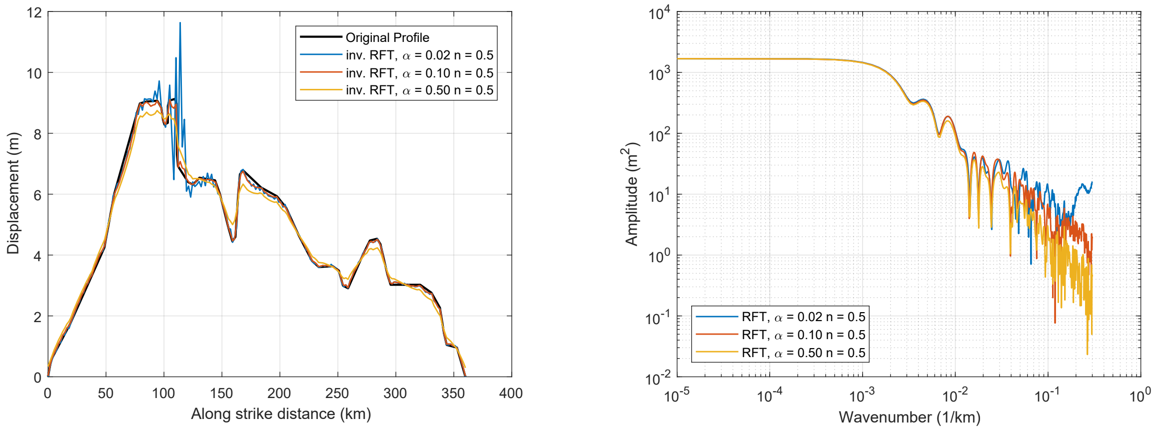
The goal of this section is to support the discussion in the Non-uniform Fourier Transform section regarding the selection of the parameters α and n for the Regularized Fourier transform. The slip profiles of 1857 Fort Tejon, CA and 1954 Fairview Peak, NV (2nd Main Segment) are used here as representative examples of a smooth profile and a profile with high variability.

As mentioned in the manuscript α controls the relative weight between the magnitude of the Fourier spectra ($\vec{Y}$) and the size of residuals between the original profile ($\vec{y}$) and the profile implied by RFT ($\vec{y}_{fwrd}$). Figure B.1 presents the amplitude spectra and the implied slip profiles for the 1857 Fort Tejon for a range of values for α. Figure B.1a shows that for small values of α the high wavenumbers are not well constrained, the penalty for minimizing $\vec{Y}$ is small, resulting in artificial oscillations in locations with sudden changes of slip. The large content of high wavenumbers can also be seen in figure B.1b where for $\alpha = 0.02$ the amplitude spectrum starts increasing after $k = 0.1$ $1/km$. Selecting a large value for α has the opposite effect in the implied slip profile and amplitude spectrum. It over-constrains the problem, it puts a high penalty in minimizing $\vec{Y}$ and because of that the narrow variations of slip cannot be fitted. This also shows in amplitude spectrum as it is bellow the spectra with lower α

Figure B.2 makes a similar comparison for 2nd Segment of the 1954 Fairview Peak rupture profile. Here the surveyed slip profile exhibits narrow variations of slip and because the data points are spaced close enough, the high wavenumbers are well constrained even when α is small. We can see that the $\vec{y}_{fwrd}$ profiles and amplitude spectra for α equal to 0.02 and 0.10 are similar meaning that a value of $\alpha = 0.1$ does not over-constrain the fit even for slip profiles with high variability. For $\alpha = 0.5$ the regularization over-constrains the fit and due to that $\vec{y}_{fwrd}$ cannot fit the original profile.

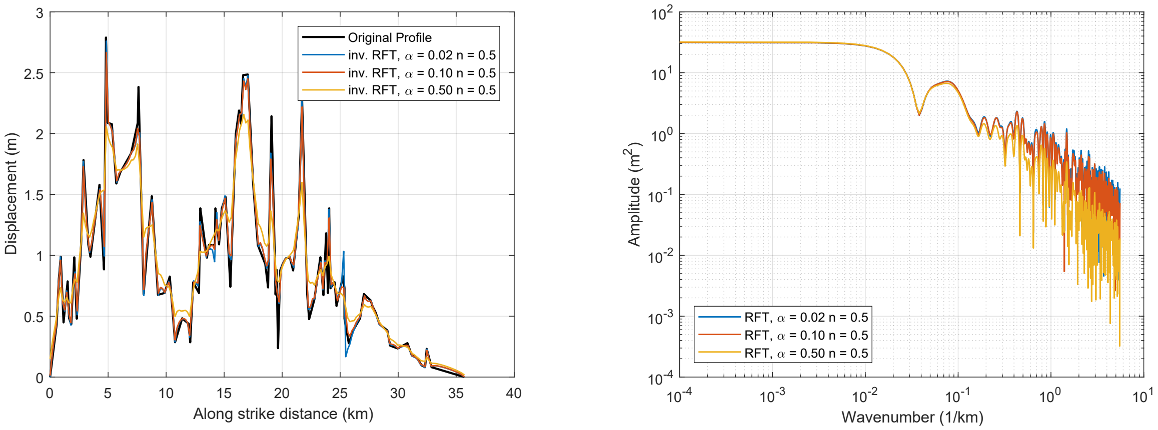
The parameter n controls the relative weight of the low and high wavenumbers, the larger the value of n the more weight is applied at the high wavenumbers. Figures B.3 and B.4

compare the impact of different values of n to the 1857 Fort Tejon, CA and 1954 Fairview Peak, NV (2nd Main Segment) profiles. In both cases, when n is small the high wavenumbers are under-constrained which leads to artificial oscillations in the space domain. When n is large the high wavenumbers are over-constrained and so $\vec{y}_{fwrd}$ cannot fit the narrow variations.



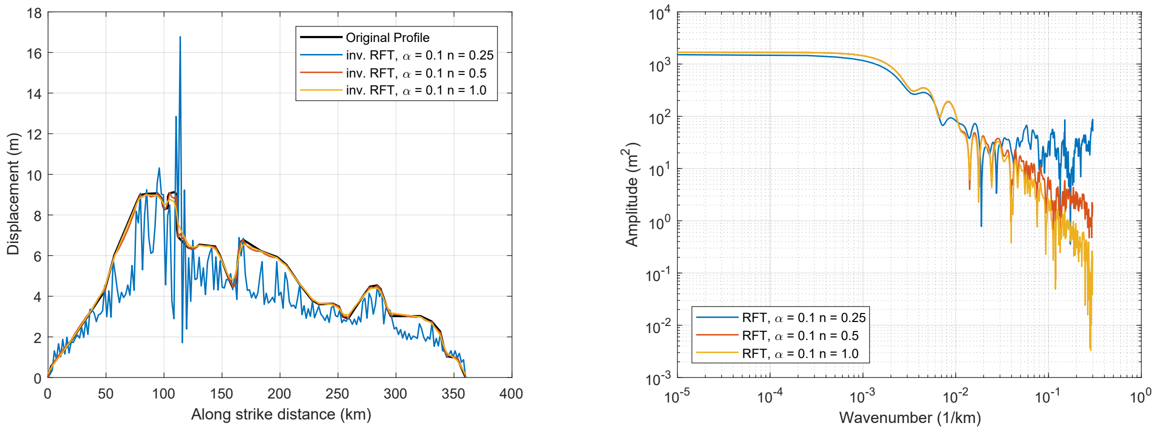
(a) (b)

Figure B.1: Implied slip profiles and Fourier amplitude spectra of the 1857 Fort Tejon, CA rupture profile for varying values of the α parameter. (a) The thick black line corresponds to the surveyed slip profile. The colored lines correspond to the profiles impled by the RFT transform for α equal to 0.02, 0.1 and 1.0. (b) Amplitude spectra of the 1857 Fort Tejon, CA rupture profile. The colored lines correspond to amplitude spectra with α equal to 0.02, 0.1 and 1.0



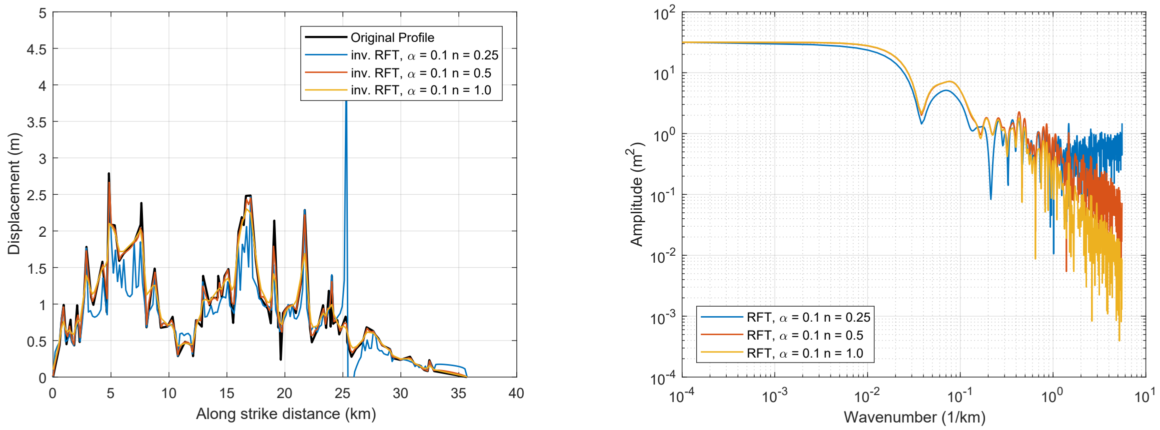
(a) (b)

Figure B.2: Implied slip profiles and Fourier amplitude spectra of the 1954 Fairview Peak, NV (2nd Main Segment) rupture profile for varying values of the α parameter. (a) The thick black line corresponds to the surveyed slip profile. The colored lines correspond to the profiles implied by the RFT transform for α equal to 0.02, 0.1 and 1.0. (b) Amplitude spectra of the 1954 Fairview Peak, NV (2nd Main Segment) rupture profile. The colored lines correspond to amplitude spectra with α equal to 0.02, 0.1 and 1.0



(a) (b)

Figure B.3: Implied slip profiles and Fourier amplitude spectra of the 1857 Fort Tejon, CA rupture profile for varying values of the n parameter. (a) The thick black line corresponds to the surveyed slip profile. The colored lines correspond to the profiles implied by the RFT transform for n equal to 0.25, 0.5 and 1.0. (b) Amplitude spectra of the 1857 Fort Tejon, CA rupture profile. The colored lines correspond to amplitude spectra with α equal to 0.25, 0.5 and 1.0



(a) (b)

Figure B.4: Implied slip profiles and Fourier amplitude spectra of the 1954 Fairview Peak, NV (2nd Main Segment) rupture profile for varying values of the n parameter. (a) The thick black line corresponds to the surveyed slip profile. The colored lines correspond to the profiles implied by the RFT transform for n equal to 0.25, 0.5 and 1.0. (b) Amplitude spectra of the 1954 Fairview Peak, NV (2nd Main Segment) rupture profile. The colored lines correspond to amplitude spectra with n equal to 0.25, 0.5 and 1.0

B.2 Derivation of phase angles of reversed polarity profiles

The reversed polarity slip profile is defined as:

$$y_r(x) = y_o(SLR - x) \tag{B.1}$$

where $y_o(x)$ is the slip profile in the original polarity and SLR is the surface rupture length. The Fourier transform of $y_o(x)$ is defined as:

$$\mathcal{F}[y_o(x)] = Y_o(k) = \int_0^{SLR} y_o(x) e^{-i2\pi kx} dx \tag{B.2}$$

and $Y(k)$ can be written in terms of the amplitude ($A(k)$) and phase angle ($\phi(k)$) as:

$$Y(k) = A_o(k) e^{i\phi_o(k)} \tag{B.3}$$

The Fourier transform of the reversed polarity profile is equal to:

$$\mathcal{F}[y_r(x)] = Y_r(k) = \int_0^{SLR} y_r(x) e^{-i2\pi kx} dx \tag{B.4}$$

Combining equations (B.1) and (B.4) and changing variables to $z = SLR - x$ gives:

$$Y_r(k) = \int_0^{SLR} y_o(z) e^{-i2\pi(-k)z} dz \; e^{-i2\pi SLR \; k} \tag{B.5}$$

The integral in equation (B.5) is the complex conjugate of the Fourier transform of the original profile and combining it with equation (B.3) gives:

$$Y_r(k) = A_o(k) e^{i(-\phi_o(k) - 2\pi SLR \; k)} \tag{B.6}$$

Based on equation (B.6) the phase angles of the reversed polarity slip profile are equal to:

$$\phi_r(k) = -\phi_o(k) - 2\pi SLR \; k \tag{B.7}$$

B.3 Supplementary figures

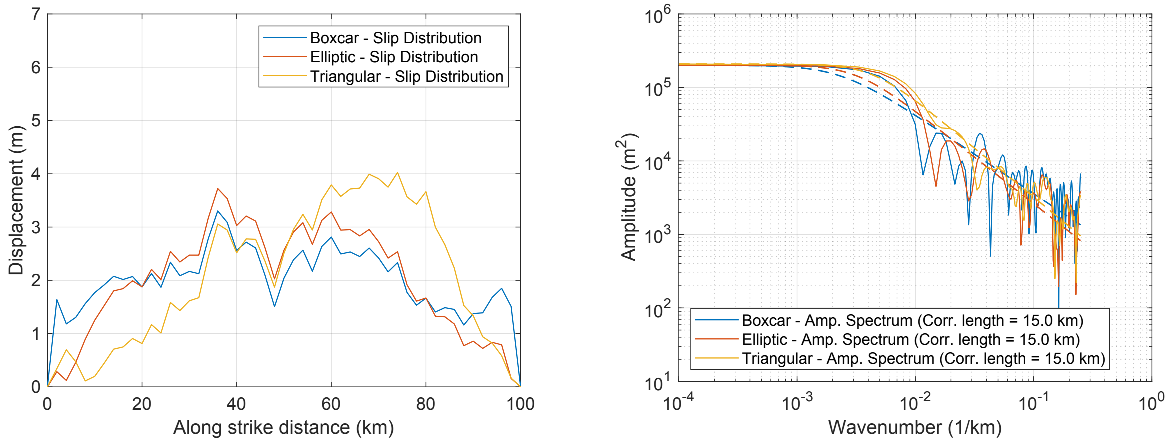

(a) (b)

Figure B.5: Comparison of slip profiles with different correlation lengths and mean displacement profiles. Subfigure (a) shows the representative slip profiles in the space domain with correlation length equal to 15km. Subfigure (b) presents with thin solid lines the amplitude spectra of the profiles shown in subfigure (a) and with thick dashed lines the amplitude models (eq: 14) fitted to the amplitude spectra.

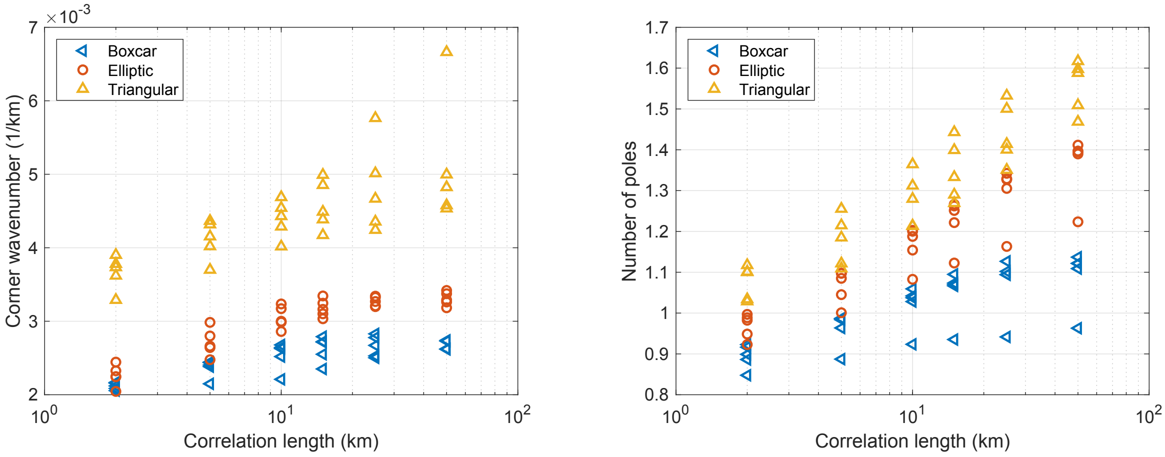

(a) (b)

Figure B.6: Trend of KC and Np with correlation length. Subfigure (a) presents how KC changes for multiple realizations of slip profiles for different correlation lengths and mean displacement profiles. Subfigure (b) shows the change in Np for multiple slip realizations with different correlation lengths and mean displacement profiles.

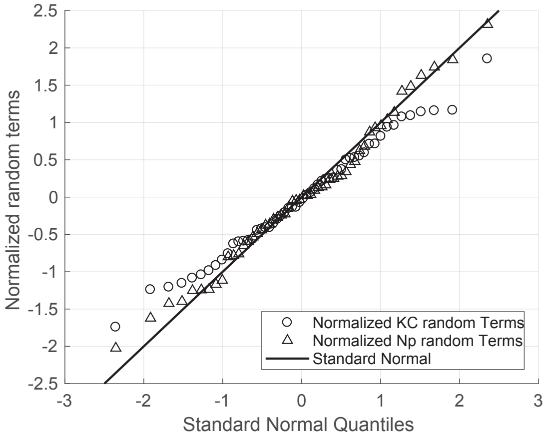


Figure B.7: Q–Q plot of the normalized η_{KC} and η_{Np}. Open markers represent the individual random terms, the line (slope 1:1) corresponds to the standard normal distribution

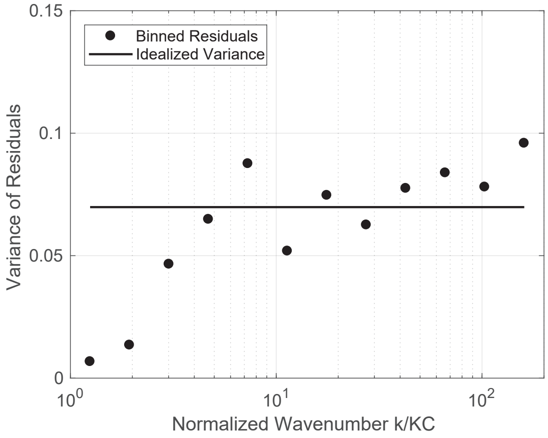


Figure B.8: Idealized variance model for forward predictions. Dots indicate the variance of each bin and the solid line represents the proposed variance model.

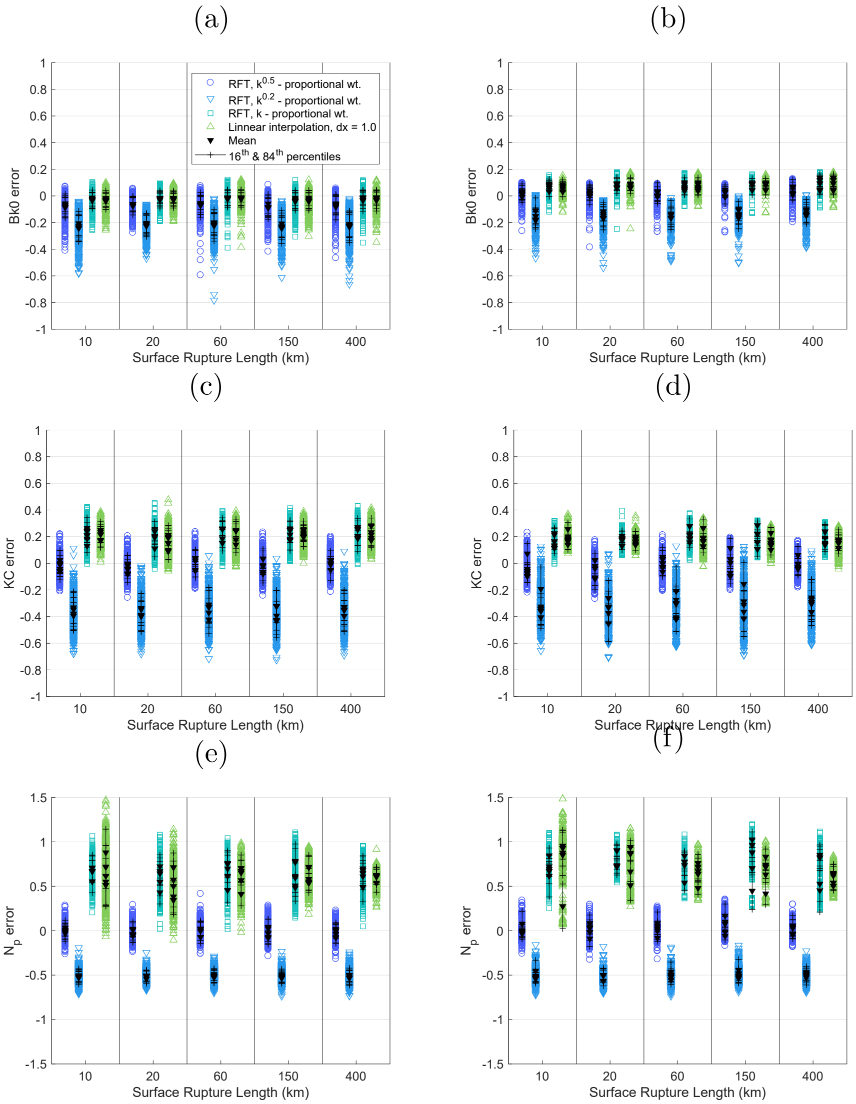


Figure B.9: Misfit of the amplitude parameters with different Fourier transform methods. The amplitude and phase derivative parameters of the different groups are summarized in Table 5 of the manuscript. The open markers are the errors of the individual estimates of the down-sampled profiles. The solid squares correspond to the mean error of each group and the crosses indicate the 16th and 84th percentile of the group errors. (a) Misfit of B_0 with random down-sampling (scheme 1); (b) Misfit of B_0 with biased down-sampling (scheme 2); (c) Misfit of KC with random down-sampling (scheme 1); (d) Misfit of KC with biased down-sampling (scheme 2); (e) Misfit of Np with random down-sampling (scheme 1); (f) Misfit of Np with biased down-sampling (scheme 2)

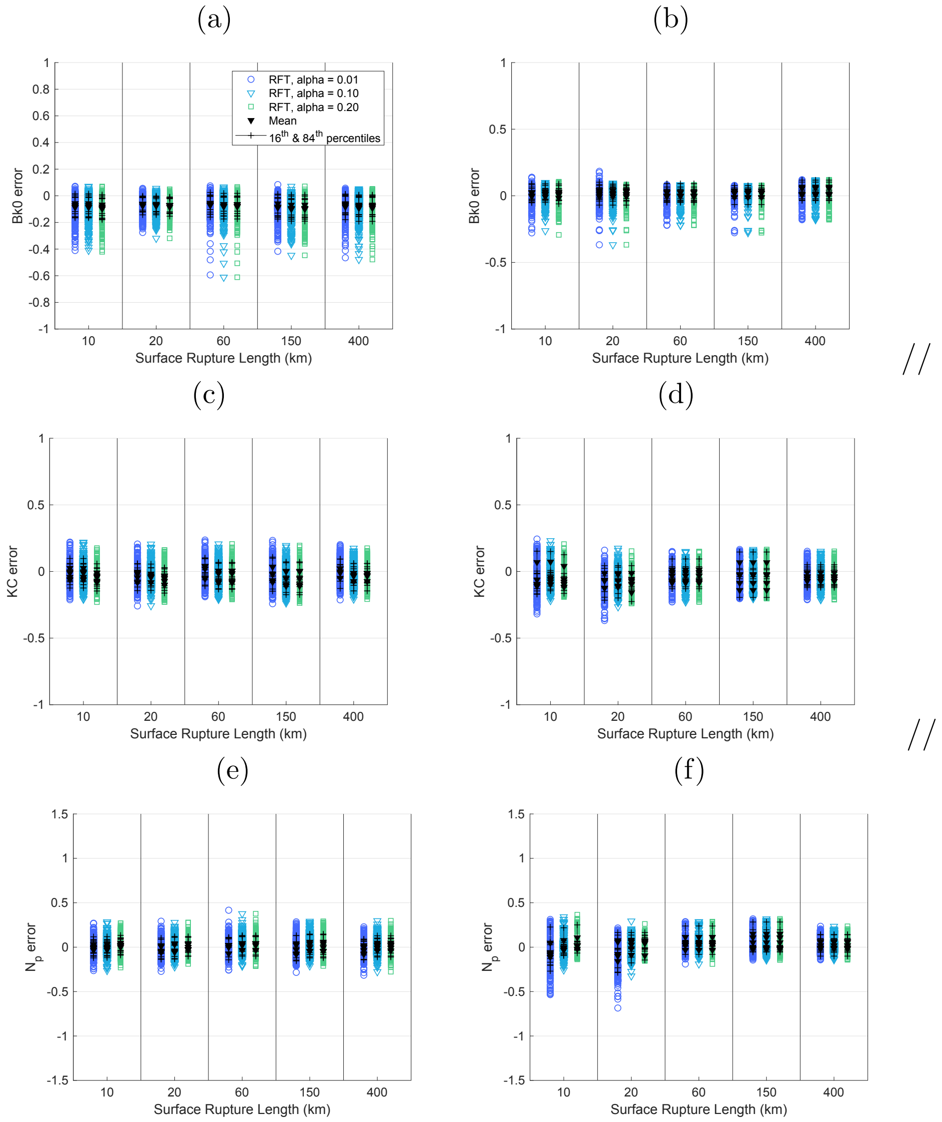
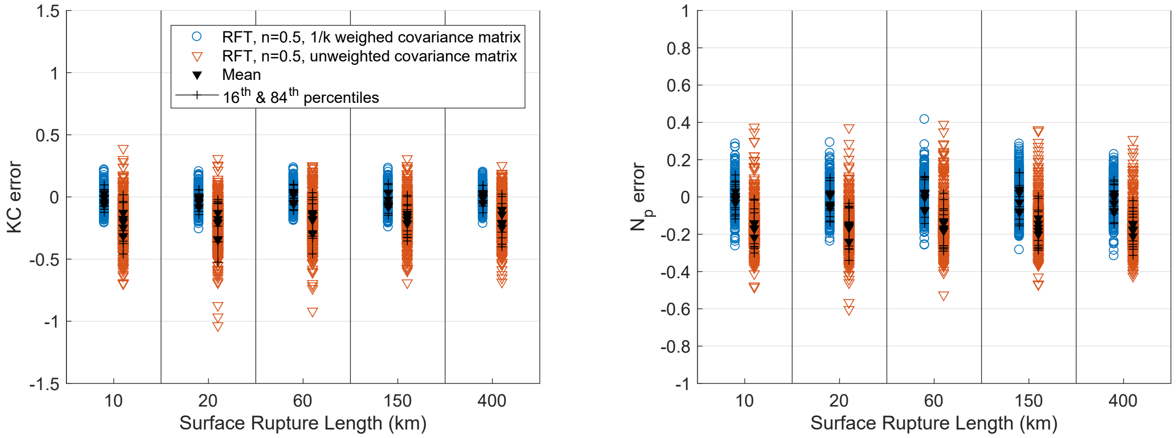

(a) (b) (c) (d) (e) (f)

Figure B.10: Misfit of the amplitude parameters with Fourier transforms with different proportionality factors (α) for the regularization weights. The amplitude and phase derivative parameters of the different groups are summarized in Table 5 of the manuscript. The open markers are the errors of the individual estimates of the down-sampled profiles. The solid squares correspond to the mean error of each group and the crosses indicate the 16th and 84th percentile of the group errors. (a) Misfit of B_0 with random down-sampling (scheme 1); (b) Misfit of B_0 with biased down-sampling (scheme 2); (c) Misfit of KC with random down-sampling (scheme 1); (d) Misfit of KC with biased down-sampling (scheme 2); (e) Misfit of Np with random down-sampling (scheme 1); (f) Misfit of Np with biased down-sampling (scheme 2)



(a) (b)

Figure B.11: Misfit of the amplitude parameters, with unweighted and $1/k$ weighted covariance matrices. The amplitude and phase derivative parameters of the different groups are summarized in Table 5 of the manuscript. The open markers are the errors of the individual estimates of the down-sampled profiles. The solid squares correspond to the mean error of each group and the crosses indicate the 16th and 84th percentile of the group errors. (a) KC estimation error; (b) Np estimation error

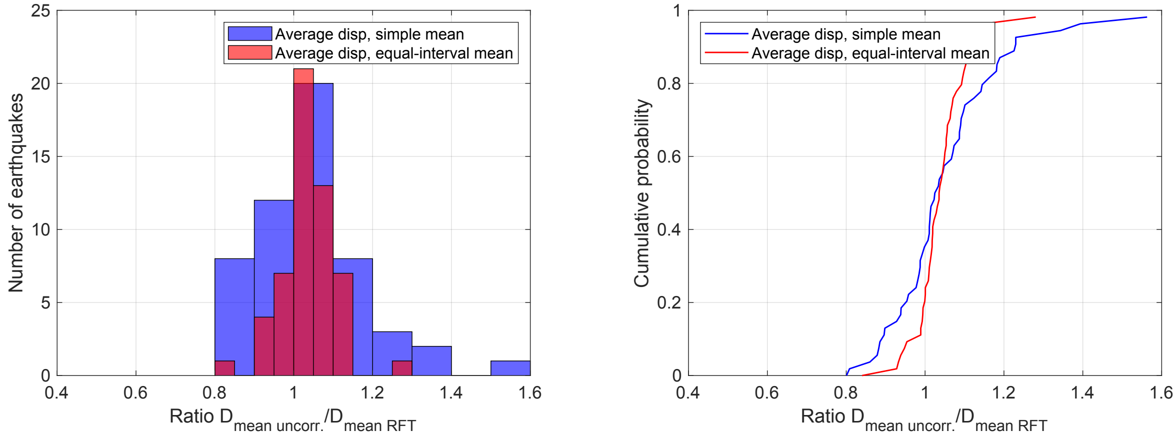

(a) (b)

Figure B.12: Comparison of D_{mean} estimate with and without the assumption of correlated slip measurements. Blue bars and line correspond to the ratio of D_{mean} computed as a simple mean of the displacement points to $D_{mean\ RFT}$ computed with the regularized Fourier transform approach. Red bars and line correspond to the ratio of D_{mean} computed as the mean of the interpolated points to $D_{meanRFT}$ computed with the regularized Fourier transform approach.

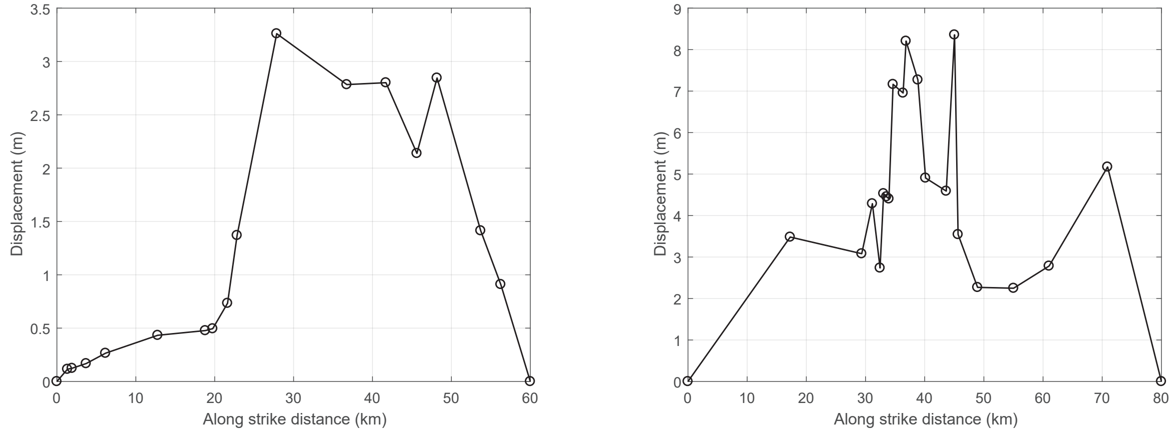

(a) (b)

Figure B.13: Representative profiles that lead to under and overestimation of D_{mean} when the assumption of uncorrelated displacements is used. (a) slip profile of 1940, Imperial, CA earthquake with $D_{simple\ mean}/D_{mean\ RFT} = 0.8$ (b) slip profile of 1891, Neo-Dani, Japan earthquake with $D_{simple\ mean}/D_{mean\ RFT} = 1.3$

B.4 Data and Resources

Future updates of the scripts can be found at: (`https://github.com/glavrentiadis/PFDHA_public`, last accessed February 2019)