

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Developing Stochastic Models as Inputs for High-Frequency Ground Motion Simulations

Permalink

<https://escholarship.org/uc/item/6km954jv>

Author

Savran, William Harvey

Publication Date

2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

SAN DIEGO STATE UNIVERSITY

Developing Stochastic Models as Inputs for High-Frequency Ground Motion Simulations

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Geophysics

by

William Harvey Savran

Committee in charge:

University of California, San Diego

Professor Yifeng Cui
Professor Enrique Luco
Professor Peter Shearer

San Diego State University

Professor Kim Olsen, Chair
Professor George Christakos
Professor Steve Day

2017

The Dissertation of William Harvey Savran is approved, and it is acceptable
in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

San Diego State University

2017

DEDICATION

To: Elisabeth Adele Savran

TABLE OF CONTENTS

Signature Page	iii
Dedication	iv
Table of Contents	v
List of Tables	vi
List of Figures	vii
Acknowledgements	viii
Vita	xi
Abstract of the Dissertation	xii
 Chapter 1 Introduction	 1
 Chapter 2 Model for Small-Scale Crustal Heterogeneity in Los Angeles Basin Based on Sonic Log Data	 10
2.1 Introduction	10
2.2 Methods	13
2.2.1 Classical Variogram	14
2.2.2 Moving-Window Variogram	15
2.2.3 Monte-Carlo Inversion	16
2.3 Statistics of Los Angeles Basin	17
2.3.1 V_{S30} Measurements	17
2.3.2 Sonic Log Measurements	18
2.4 Conclusions	20
2.5 Acknowledgements	21
References	22
 Chapter 3 Ground Motion Simulation and Validation of the M_w 5.4 Chino Hills Earthquake in Scattering Media	 32
3.1 Introduction	32
3.2 Ground Motion Simulations	41
3.3 Discussion of Results	53
3.4 Conclusions	57
References	58
 Chapter 4 Kinematic Rupture Generator Based on 3D Spontaneous Rupture Simulations with Geometrically Rough Faults	 92
4.1 Introduction	93
4.1.1 Statistical Framework for Proposed KRG	94
4.2 Dynamic Rupture Simulations	103
4.3 Extracting Statistics for Kinematic Source Generation.	109
4.4 Validating the KRG	119
4.5 Discussion	123
4.6 Conclusions	125
References	126

LIST OF TABLES

Table 3.1: Simulations parameters for 2008, M_w 5.4 Chino Hills earthquake	66
Table 3.2: List of von Karman correlation function parameters corresponding to simulations considered for sensitivity testing of models of small-scale velocity fluctuations	67
Table 4.1: Coefficients for short-range coregionalization matrix	131
Table 4.2: Coefficients for long-range coregionalization matrix	132
Table 4.3: Zero-offset correlation coefficients based on best-fitting linear-model of coregionalization	133

LIST OF FIGURES

Figure 2.1: (Top Left) Apall borehole showing raw sonic log, long-period median filter, and smoothed log. (Top Right) Fluctuation profile used for analysis. (Bottom) Classical and moving-window variogram estimates showing estimated Hurst exponent and correlation length.	26
Figure 2.2: (Left) Locations of 371 V_{S30} measurements used in variogram analysis. (Right) Classical variogram estimate of V_{S30} measurements.	27
Figure 2.3: Map showing 35 deep borehole locations within the Los Angeles basin used for the inversion.	28
Figure 2.4: Inversion results showing acceptance rate of von Karman parameter pair (ν, a_z) for all boreholes in Los Angeles basin	29
Figure 2.5: Theoretical variograms plotted showing best-fitting parameters of von Karman autocorrelation functions. Variograms are compared against the estimated variogram from data.	30
Figure 2.6: Depth-dependent variogram computed over 500 m intervals using sonic log data set	31
Figure 3.1: Simulation domain of 2008, Chino Hills, CA validation	68
Figure 3.2: (Top Left) Map view of V_s from CVM-S4.26 including stochastic velocity fluctuations. (Bottom Left) Cross-section showing deep portion of Los Angeles basin. (Right) 1D profile extracted showing inclusion of small-scale heterogeneities	69
Figure 3.3: (Top) Slip-rate functions from finite-fault sourced used for ground motion simulations. (Bottom) Fourier transform of slip-rate functions.	70
Figure 3.4: (Left) Shear-wave quality factor (Q_s) plotted against shear-wave velocity for several attenuation models. (Right) Resulting Q_s due to frequency-dependent attenuation	71
Figure 3.5: Map showing station locations used for validation of ground motion simulations	72
Figure 3.6: Map view of V_s showing stochastic models used for sensitivity testing of von Karman parameters	73
Figure 3.7: Qualitative comparisons shown for stations RUS, LTP and SRN. The left three panels show velocity seismograms (cm/s) and the large right panel shows cumulated energy (cm ² /s)	74
Figure 3.8: Same as Figure 3.7 but for stations STS, PSR, and FUL	75

Figure 3.9: Qualitative comparisons between simulations including small-scale heterogeneities, simulations without heterogeneities and data. We show one-sided envelopes, cumulative kinetic energy, normalized Fourier velocity, and the response spectrum.	76
Figure 3.10: Same as Figure 3.9	77
Figure 3.11: Simulation-wide effect of small-scale heterogeneities on PGV amplitudes and comparison against data	78
Figure 3.12: Same as Figure 3.11, but for PGA	79
Figure 3.13: Same as Figure 3.11, but for cumulative energy	80
Figure 3.14: Same as Figure 3.11, but for Arias intensity	81
Figure 3.15: Same as Figure 3.11, but for duration	82
Figure 3.16: PGV, PGA, CKE, ENER, and DUR plotted as a function of epicentral distance for $Q_{s0}/V_s = 100$	83
Figure 3.17: Same as Figure 3.16, but considering an M_w of 5.5 instead of M_w 5.4	84
Figure 3.18: Same as Figure 3.16, but considering $Q_{s0}/V_s = 50$	85
Figure 3.19: Comparisons against recorded data for simulations using M_w 5.5 $Q_s/V_s=50$ for PGV, ENER, and DUR for 0.15 – 1.0 Hz and 0.15 – 2.5 Hz	86
Figure 3.20: Same as Figure 3.16	87
Figure 3.21: Comparisons between simulation assuming M_w 5.5 and $Q_{s0}/V_s = 100$ with simulation using M_w 5.4 and $Q_{s0}/V_s = 150$ for PGV, ENER, and DUR plotted between 0.15 – 1.0 Hz	88
Figure 3.22: Bias plot showing $\log(\text{data/synthetic})$ assessing M_w and Q_{s0}/V_s interdependency	89
Figure 3.23: Percent difference in GOF scores between $Q_{s0}/V_s = 100$ and M_w 5.5 with $Q_{s0}/V_s = 150$ and M_w 5.4 for metrics PGV, ENER, and DUR.	90
Figure 3.24: Percent difference in GOF scores for models with and without small-scale heterogeneities for metrics PGV, PGA, ENER, AI and DUR	91
Figure 4.1: Regularized Yoffe function in addition to useful definitions describing parameters	134
Figure 4.2: Illustrates fitting of kinematic slip-rate function to dynamic slip-rate Functions	135
Figure 4.3: Moment vs area scaling relationships from dynamic rupture simulations plotted against empirically derived models for inter-plate strike-slip earthquakes	136
Figure 4.4: Stacked source spectra computed from finite-fault slip-rate functions from	

different ensembles of dynamic rupture simulations	137
Figure 4.5: Marginal probability density functions and quantile-quantile plots shown for dynamic rupture fields	138
Figure 4.6: Depth-dependence of one-point statistics for the dynamic rupture fields . .	139
Figure 4.7: Preferred linear-model of coregionalization based on database of dynamic rupture simulations	140
Figure 4.8: Outline of two-step procedure to compute τ_s from the physical source fields derived from dynamic rupture simulations	141
Figure 4.9: Multiple linear model defined to estimate τ_r given the final slip and effective rupture duration	142
Figure 4.10: Example kinematic source model generated using kinematic rupture generator developed in this chapter	143
Figure 4.11: (Left) Material model used in ground motion validations based on 1D profile extracted from the CVM-S4.26 velocity model. (Right) Free-surface of simulation domain showing fault-profile and station locations used for validation	144
Figure 4.12: Final slip distributions for five finite-fault source models produced using the kinematic rupture generator used for validation.	145
Figure 4.13: Velocity seismograms plotted for five source models at station T3	146
Figure 4.14: Same as Figure 4.13, but for station B3	147
Figure 4.15: Normalized far-field displacement spectra computed for Source 3	148
Figure 4.16: Far-field acceleration spectra computed for Source 3	149
Figure 4.17: Comparisons of spectral accelerations for Source 1 (left), Source 2 (center), and Source 3 (right) against NGA West2 GMPEs	150
Figure 4.18: Bias plots comparing simulation median spectral accelerations for all five kinematic source models against GMPE medians.	151

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor Professor Kim Olsen. Without his support and guidance I would not have been able to finish my dissertation. His door was constantly open and he would always provide thoughtful insight into any issue that I was having with my research. He also taught me many soft-skills, which made me a more well-rounded individual and researcher. Most especially, he taught me what it means to be a good advisor and mentor. In addition, I would like to offer a special acknowledgement to Professor Steve Day. I have never had the opportunity of working with someone as knowledgeable, insightful, thoughtful, and kind. Steve offers everything he can to help others succeed, and has an uncanny ability to explain the most difficult problems in the simplest of terms. He taught me to never shy away from difficult problems, and how to approach them with elegant and clever solutions. I am truly grateful for having the opportunity to work this closely with both Professor Kim Olsen and Professor Steve Day at San Diego State University. I would also like to thank my committee members for their thoughtful comments on my dissertation and assistance throughout my Ph.D.

Throughout the course of my Ph.D., I have made several lifelong friendships and relationships. I owe a very special thanks to my girlfriend Jenna. She has been by my side for the entire journey, and was the rock necessary for me to stay grounded enough to withstand the long and arduous process of earning a doctorate degree. It is difficult to express my gratitude, but I can say these short words: Jenna I love you, and I appreciate everything you have done for me, both large and small. I would also like to acknowledge Charles Bradford Larson for always being around to support my radical hobbies and decompressing on the weekends at live music shows. Finally, I would like to acknowledge the music of Andre Hicks for helping make it through long nights debugging research codes.

Chapter 2, in full, is a reformatted version of the material as it appears in *Geophysical Journal International*: Savran, W. H., and K. B. Olsen (2016), Model for small-scale crustal heterogeneity in Los Angeles basin based on inversion of sonic log data, *Geophysical Journal International*, 205(2), 856–863, doi:10.1093/gji/ggw050. William Harvey Savran was the primary investigator and author of this paper.

Chapter 3, in full, is a reformatted version of a paper currently being prepared for submission for publication to *Bull. Seism. Soc. Am.*: Savran, W. H., and K. B. Olsen (2017), Ground Motion Simulation and Validation of the M_w 5.4 Chino Hills Earthquake in Scattering Media. William Harvey Savran was the primary investigator and author of this paper.

Chapter 4, in full, is a reformatted version of a paper currently being prepared for submission for publication to *Geophysical Journal International*: Savran, W. H., K. B. Olsen, and S. M. Day (2017), Kinematic Rupture Generator Based on 3D Spontaneous Rupture Simulations with Geometrically Rough Faults. William Harvey Savran was the primary investigator and author of this paper.

VITA

- 2012 B.S. in Geophysics, University of Nevada, Reno
- 2017 Ph.D. in Geophysics, University of California, San Diego and San Diego State University

PUBLICATIONS

Savran, W. H., and K. B. Olsen (2016), Model for small-scale crustal heterogeneity in Los Angeles basin based on inversion of sonic log data, *Geophysical Journal International*, 205(2), 856–863, doi:10.1093/gji/ggw050

W. H., and K. B. Olsen (2017), Ground Motion Simulation and Validation of the M_w 5.4 Chino Hills Earthquake in Scattering Media (Currently being prepared for submission to *Bull. Seism. Soc. Am.*).

Savran, W. H., K. B. Olsen, and S. M. Day (2017), Kinematic Rupture Generator Based on 3D Spontaneous Rupture Simulations with Geometrically Rough Faults (Currently being prepared for submission to *Geophysical Journal International*).

ABSTRACT OF THE DISSERTATION

Developing Stochastic Models as Inputs for High-Frequency Ground Motion Simulations

by

William Harvey Savran

Doctor of Philosophy in Geophysics

University of California, San Diego, 2017

San Diego State University, 2017

Professor Kim Olsen, Chair

High-frequency (~ 10 Hz) deterministic ground motion simulations are challenged by our understanding of the small-scale structure of the earth's crust and the rupture process during an earthquake. We will likely never obtain deterministic models that can accurately describe these processes down to the meter scale length required for broadband wave propagation. Instead, we can attempt to explain the behavior, in a statistical sense, by including stochastic models defined by correlations observed in the natural earth and through physics based simulations of the earthquake rupture process. Toward this goal, we develop stochastic models to address both of the primary considerations for deterministic ground motion simulations: namely, the description of the material properties in the crust, and broadband earthquake source descriptions.

Using borehole sonic log data recorded in Los Angeles basin, we estimate the spatial correlation structure of the small-scale fluctuations in P-wave velocities by determining the best-fitting parameters of a von Karman correlation function. We find that Hurst exponents, ν , between 0.0-0.2, vertical correlation lengths, a_z , of 15-150m, and

standard deviation, σ of about 5% characterize the variability in the borehole data. Using these parameters, we generated a stochastic model of velocity and density perturbations and combined with leading seismic velocity models to perform a validation exercise for the 2008, Chino Hills, CA using heterogeneous media. We find that models of velocity and density perturbations can have significant effects on the wavefield at frequencies as low as 0.3 Hz, with ensemble median values of various ground motion metrics varying up to $\pm 50\%$, at certain stations, compared to those computed solely from the CVM. Finally, we develop a kinematic rupture generator based on dynamic rupture simulations on geometrically complex faults. We analyze ~ 100 dynamic rupture simulations on strike-slip faults ranging from M_w 6.4 - 7.2. We find that our dynamic simulations follow empirical scaling relationships for inter-plate strike-slip events, and provide source spectra comparable with an ω^{-2} model. Our rupture generator reproduces GMPE medians and intra-event standard deviations spectral accelerations for an ensemble of 10 Hz fully-deterministic ground motion simulations, as compared to NGA West2 GMPE relationships up to 0.2 seconds.

Chapter 1

Introduction

High-frequency deterministic earthquake simulations pose a unique set of challenges due to the high-resolution of input parameters required to accurately describe the short scale-lengths present in the wavefield. In order to provide seismic hazard assessments for all buildings, structural engineers typically require ground motions containing frequency content up to at least 10 Hz. This is related to the approximate resonance frequency of single-story buildings, such as offices or family residences. Fortunately for citizens, there have been very few major ($M_w > 7.0$) strike-slip earthquakes recorded in recent history on the San Andreas fault system. However, this results in a dearth of recorded data making seismic hazard assessments reliant upon models constructed using data predominately recorded from other regions. Ground motion simulations can help augment sparse data sets at missing magnitudes and distances in close proximity to the fault. In addition, these simulations can be used to forecast ground motions from future earthquakes given sufficient capabilities to model the physical properties of the earth.

Current state-of-the-art techniques to simulate broadband synthetic seismograms incorporate a hybrid approach that merges low-frequency deterministic simulations with stochastic representations of the high-frequency wavefield (e.g., [Graves and Pitarka, 2015; Olsen and Takedatsu, 2015]). These methods show very promising results in validation exercises [Goulet et al., 2015] and for their predictive capabilities for future earthquakes. However, they do not explicitly consider the physics of wave-propagation for the stochastically generated high-frequencies, which prevents these methods from providing physical insight into the behavior of high-frequency wave propagation.

Deterministic ground motion simulations provide an appealing alternative to generating high-frequency synthetic seismograms due to their ability to provide insight into various wave propagation phenomena (e.g., Li et al., [1994]; Olsen, [2000]; Spudich and Olsen, [2001]; Aagaard and Heaton, [2004]; Day et al., [2008]). However, until recently, deterministic ground motion studies were typically applied at lower frequencies (< 1 Hz) especially when attempting to predict actual ground motion amplitudes (e.g., Olsen et al.,

[2009]). In the context of validation, *Taborda and Bielak*, [2013] simulated 4 Hz ground motions from the 2008 M_w 5.4 Chino Hills earthquake and concluded that the maximum frequency content of deterministic ground motion simulations could be extended past 1 Hz using current numerical models describing the material properties of the crust (e.g., *Lee et al.*, [2011]).

It has been long known from interpretations of seismic data that small-scale velocity heterogeneities exist in the earth's crust (e.g., *Aki and Chouet*, [1975]). Also, they have been observed directly in borehole sonic logs (e.g., *Holliger*, [1996]) and other various datasets (e.g., *Levander*, [1992]; *Mela and Louie*, [2001]). Typically, these small-scale heterogeneities have been ignored when modeling longer-period ground motions due to the general consensus that seismic scattering was typically a phenomenon affecting higher frequencies (>1 Hz). However, *Hartzell et al.*, [2010] showed that significant deviations exist in peak motions at frequencies as low as 0.5 Hz challenging this understanding, and indicating that small-scale heterogeneities could play an important role in deterministic ground motions computations at low frequencies.

Typically, small-scale heterogeneities are modeled using a random field with its autocorrelation representative of the autocorrelations observed in velocity and density fluctuations after removing some long-period trend (e.g., *Holliger*, [1996]). Using 2D finite-difference simulations, *Frankel and Clayton*, [1986] concluded that von Karman autocorrelation functions best describe observed features of the seismic wavefield. The von Karman autocorrelation function displays self-affine behavior past some corner wavenumber (k_c) and uncorrelated behavior between k_{min} and k_c and can be described using the parameters: ν , the Hurst exponent which controls the relative amplitudes of the small-scale fluctuations; a , the correlation length which relates to k_c ; and σ , the strength of the fluctuations. However, these parameters have not been well constrained in the literature and display significant variance, especially the correlation length. Chapter 2 of this dissertation focuses on the estimation of the von Karman parameters in southern California as constrained by digitized borehole sonic logs recorded during oil exploration efforts in the 1960s. These parameters can then be used to incorporate a representative

model of small-scale heterogeneities in deterministic ground motion simulations for southern California.

Taborda and Bielak, [2013] and a subsequent study *Taborda and Bielak*, [2014] did not consider the effects of small-scale heterogeneities in their validation exercise of the 2008 M_w 5.4 Chino Hills earthquake. This is likely due to their focus on demonstrating the viability of deterministic ground motion simulations past 1 Hz and understanding shortcomings of current 3D deterministic velocity models. Based on their results, and those of *Hartzell et al.*, [2010], we investigate the effects of small-scale heterogeneities on ground motion metrics commonly used in seismic hazard analysis. In addition, we are curious whether small-scale heterogeneities can potentially explain some of the observed misfits seen by *Taborda and Bielak*, [2013]. Chapter 3 of this dissertation largely extends these two studies. In particular, we focus on the effects of small-scale heterogeneities and refining the empirical models for anelastic attenuation.

Aside from modeling the small-scale variability present in the earth's crust, another primary challenge facing high-frequency deterministic earthquake simulations involves describing the finite-fault source down to the scale lengths required to generate the high-frequency wavefield. For several years, dynamic rupture simulations (e.g., *Day et al.*, [2005]) have been used successfully to describe finite-fault sources for ground motion simulations (e.g., *Dalguer et al.*, [2008]; *Olsen et al.*, [2009]; *Roten et al.*, [2011]). In addition, *Baumann and Dalguer*, [2014] showed that ensembles of dynamic rupture simulations produce ground motions comparable with leading ground motion prediction equations up to 1 Hz. Despite promising results at lower frequencies, there was no physical mechanism available to generate realistic high-frequency waves from the dynamic rupture simulations.

A pioneering study by *Dunham et al.*, [2011] showed, for the 2D case, that including self-similar fault topography provides a physical mechanism in dynamic rupture simulations that can generate high-frequency waves consistent with observations. As a follow up to this study, *Shi and Day*, [2013] confirmed this result in 3D using dynamic rupture simulations on strike-slip faults. Using the source models computed by [*Shi and Day*, 2013], the Ph.D. work of *Withers*, [2016] found promising comparisons between

ground motions simulations and GMPEs for both strike-slip and buried thrust faults. They found that dynamic sources on rough faults can be used to generate ground motions consistent with GMPE median spectral-acceleration amplitudes up to frequencies of engineering interest. Despite the favorable comparisons shown by *Withers*, [2016], dynamic rupture models require significant computational resources even for a single simulation, and, by nature, cannot be used with methods requiring a kinematic description of the earthquake source. These issues render them less effective than kinematic approaches for large-scale probabilistic physics-based seismic hazard evaluations given the currently available computational resources.

In addition to using statistical models to describe the material properties of the crust, seismic sources have also been characterized within a statistical framework (e.g., *Graves and Pitarka*, [2010]; *Schmedes et al.*, [2013]; *Song et al.*, [2013]) using so-called kinematic rupture generators (KRGs) or pseudo-dynamic source models. These stochastic source models generate kinematic source descriptions meaning that the spatiotemporal evolution of slip on the fault interface is defined a priori, in contrast with dynamic simulations that explicitly compute the slip evolution as part of the numerical solution. Generally speaking, KRGs operate by defining some spatial correlation structure between the parameters of a source time function. In addition, they might include information about the marginal probability distributions of these parameters (e.g., *Song et al.*, [2013]). Kinematic source descriptions allow for rapid generation of synthetic seismograms using methods involving Green's functions to describe the impulse response of the elastic medium. Due to the computational savings as compared with finite-difference or finite-element solutions to the elastodynamic equations, this technique has seen widespread adoption in probabilistic physics-based seismic hazard efforts such as the Southern California Earthquake Center (SCEC) Broadband Platform [*Maechling et al.*, 2015] or CyberShake [*Graves et al.*, 2011].

Current KRGs make use of dynamic rupture simulations on flat faults to inform the spatial dependency of source parameters, but they do not consider the effects of rough faults in their rupture generators. Most recently, *Graves and Pitarka*, [2016] showed that incorporating fault roughness and small-scale media heterogeneities could help reduce

directivity effects at high-frequencies, but their KRG does not explicitly consider the influence of the rough fault on the resulting rupture process. For her Ph.D., *Yao*, [2017] showed that rough faults display rupture characteristics that differ from ruptures on flat faults indicating that these considerations should be included into current kinematic rupture generators. In addition, KRGs have not successfully been shown to generate realistic ground motions to 10 Hz, which are desirable for broadband seismic hazard evaluations and to study high-frequency behavior of the seismic wavefield. In Chapter 4 of this dissertation, we develop a new KRG based on the statistics derived from an ensemble of 3D dynamic rupture simulations on rough faults with the primary goal of generating realistic ground motion time-series up to 10 Hz. As deterministic simulations move toward higher frequencies, we will increasingly rely on statistical relationships to generate useful comparisons with recorded data, as qualitative “wobble-for-wobble” comparisons will likely become impossible.

Generally speaking, the collection of projects presented in this dissertation constitute the necessary groundwork to implement a deterministic counterpart to the SCEC Broadband Platform. This new platform would incorporate stochastic input parameters, but compute the high-frequency wavefield using only physics-based methods. In this context, both the source and the medium description are realizations of a random variable, meaning that this approach should be incorporated into a Monte Carlo framework. This framework would allow researchers to quantify uncertainty by explicitly considering important physical aspects of wave propagation, such as scattering and non-linear material behavior. Ground motions computed using this Monte Carlo framework would generate probability distributions of ground motions as opposed to a single estimate, as with current methods, which could be compared against ground motions recorded from a single earthquake or compiled databases of ground motions (e.g. *Ancheta et al.*, [2014]). For example, this methodology could help confirm the observed interdependency between scattering attenuation and intrinsic attenuation which have previously relied on theoretical approximations of seismic scattering (e.g., *Wang and Shearer*, [2017]). Additionally, we could assess ground motion variability associated with a particular set of natural faults in southern California (e.g., *Aagaard et al.*, [2010]) or help refine the parameterizations of

ground motion prediction equations at short-periods similar to the approach that *Day et al.*, [2008] incorporated for long-period basin amplification effects.

References

- A. R. Levander (1992), A Stochastic View of Lower Crustal Fabric Based on Evidence From the Ivrea Zone, *Geophys. Res. Lett.*, 19(11), 1153–1156.
- Aagaard, B. T., and T. H. Heaton (2004), Near-Source Ground Motions from Simulations of Sustained Intersonic and Supersonic Fault Ruptures, *Bulletin of the Seismological Society of America*, 94(6), 2064–2078.
- Aagaard, B. T., R. W. Graves, A. Rodgers, T. M. Brocher, R. W. Simpson, D. Dreger, N. A. Petersson, S. C. Larsen, S. Ma, and R. C. Jachens (2010), Ground-Motion Modeling of Hayward Fault Scenario Earthquakes, Part II: Simulation of Long-Period and Broadband Ground Motions, *Bulletin of the Seismological Society of America*, 100(6), 2945–2977, doi:10.1785/0120090379.
- Aki, K., and B. Chouet (1975), Origin of coda waves: Source, attenuation, and scattering effects, *J. Geophys. Res.*, 80(23), 3322–3342, doi:10.1029/JB080i023p03322.
- Ancheta, T. D. et al. (2014), NGA-West2 Database, *Earthquake Spectra*, 30(3), 989–1005, doi:10.1193/070913eqs197m.
- Baumann, C., and L. A. Dalguer (2014), Evaluating the Compatibility of Dynamic Rupture-Based Synthetic Ground Motion with Empirical Ground-Motion Prediction Equation, *Bulletin of the Seismological Society of America*, 104(2), 634–652, doi:10.1785/0120130077.
- Dalguer, L. A., H. Miyake, S. M. Day, and K. Irikura (2008), Surface Rupturing and Buried Dynamic-Rupture Models Calibrated with Statistical Observations of Past Earthquakes, *Bulletin of the Seismological Society of America*, 98(3), 1147–1161, doi:10.1785/0120070134.
- Day, S. M., L. A. Dalguer, and N. Lapusta (2005), Comparison of finite difference and boundary integral solutions to three-dimensional spontaneous rupture, *Journal of Geophysical Research*, 110(B12), B12307, doi:10.1029/2005JB003813.
- Day, S. M., R. W. Graves, J. Bielak, D. Dreger, S. C. Larsen, K. B. Olsen, A. Pitarka, and L. Ramirez-Guzman (2008), Model for Basin Effects on Long-Period Response Spectra in Southern California, <http://dx.doi.org/10.1193/1.2857545>, 24(1), 257–277, doi:10.1193/1.2857545.
- Dunham, E. M., D. Belanger, L. Cong, and J. E. Kozdon (2011), Earthquake Ruptures with Strongly Rate-Weakening Friction and Off-Fault Plasticity, Part 1: Planar Faults,

- Bulletin of the Seismological Society of America*, 101(5), 2296–2307, doi:10.1785/0120100075.
- Frankel, A., and R. Clayton (1986), Finite difference simulations of seismic scattering: Implications for the propagation of short-period seismic waves in the crust and models of crustal heterogeneity, *Journal of Geophysical Research: Solid Earth*, 91(B6), 6465–6489.
- Goulet, C. A., N. A. Abrahamson, P. G. Somerville, and K. E. Wooddell (2015), The SCEC Broadband Platform Validation Exercise: Methodology for Code Validation in the Context of Seismic-Hazard Analyses, *Seismological Research Letters*, 86(1), 17–26, doi:10.1785/0220140104.
- Graves, R. W. et al. (2011), CyberShake: A Physics-Based Seismic Hazard Model for Southern California, *PAGEOPH*, 168(3-4), 367–381, doi:10.1007/s00024-010-0161-6.
- Graves, R. W., and A. Pitarka (2010), Broadband Ground-Motion Simulation Using a Hybrid Approach, *Bulletin of the Seismological Society of America*, 100(5A), 2095–2123, doi:10.1785/0120100057.
- Graves, R. W., and A. Pitarka (2015), Refinements to the Graves and Pitarka (2010) Broadband Ground-Motion Simulation Method, *Seismological Research Letters*, 86(1), 75–80, doi:10.1785/0220140101.
- Graves, R. W., and A. Pitarka (2016), Kinematic Ground-Motion Simulations on Rough Faults Including Effects of 3D Stochastic Velocity Perturbations, *Bulletin of the Seismological Society of America*, 106(5), 2136–2153, doi:10.1785/0120160088.
- Hartzell, S., S. Harmsen, and A. Frankel (2010), Effects of 3D random correlated velocity perturbations on predicted ground motions, *Bulletin of the Seismological Society of America*, 100(4), 1415–1426, doi:10.1785/0120090060.
- Holliger, K. (1996), Upper-crustal seismic velocity heterogeneity as derived from a variety of P-wave sonic logs, *Geophysical Journal International*, 125(3), 813–829, doi:10.1111/j.1365-246X.1996.tb06025.x.
- Lee, E.-J., P. Chen, T. H. Jordan, and L. Wang (2011), Rapid full-wave centroid moment tensor (CMT) inversion in a three-dimensional earth structure model for earthquakes in Southern California, *Geophysical Journal International*, 186(1), 311–330, doi:10.1111/j.1365-246X.2011.05031.x.
- Li, Y.-G., K. Aki, D. Adams, A. Hasemi, and W. H. K. Lee (1994), Seismic Guided-Waves Trapped in the Fault Zone of the Landers, California, Earthquake of 1992, *Journal of Geophysical Research: Solid Earth*, 99(B6), 11705–11722.

- Maechling, P. J., F. Silva, S. Callaghan, and T. H. Jordan (2015), SCEC Broadband Platform: System Architecture and Software Implementation, *Seismological Research Letters*, 86(1), 27–38, doi:10.1785/0220140125.
- Mela, K., and J. N. Louie (2001), Correlation length and fractal dimension interpretation from seismic data using variograms and power spectra, *GEOPHYSICS*, 66(5), 1372–1378, doi:10.1190/1.1487083.
- Olsen, K. B. (2000), Site Amplification in the Los Angeles Basin from Three-Dimensional Modeling of Ground Motion, *Bulletin of the Seismological Society of America*.
- Olsen, K. B. et al. (2009), ShakeOut-D: Ground motion estimates using an ensemble of large earthquakes on the southern San Andreas fault with spontaneous rupture propagation, *Geophys. Res. Lett.*, 36(4), L04303, doi:10.1029/2008GL036832.
- Olsen, K. B., and R. Takedatsu (2015), The SDSU Broadband Ground-Motion Generation Module BBtoolbox Version 1.5, *Seismological Research Letters*, 86(1), 81–88, doi:10.1785/0220140102.
- Roten, D., K. B. Olsen, H. Magistrale, J. C. Pechmann, and V. M. Cruz-Atienza (2011), 3D Simulations of M 7 Earthquakes on the Wasatch Fault, Utah, Part I: Long-Period (0-1 Hz) Ground Motion, *Bulletin of the Seismological Society of America*, 101(5), 2045–2063, doi:10.1785/0120110031.
- Schmedes, J., R. J. Archuleta, and D. Lavallee (2013), A kinematic rupture model generator incorporating spatial interdependency of earthquake source parameters, *Geophysical Journal International*, 192(3), 1116–1131, doi:10.1093/gji/ggs021.
- Shi, Z., and S. M. Day (2013), Rupture dynamics and ground motion from 3-D rough-fault simulations, *Journal of Geophysical Research: Solid Earth*, 118(3), 1122–1141, doi:10.1002/jgrb.50094.
- Song, S. G., L. A. Dalguer, and P. M. Mai (2013), Pseudo-dynamic source modelling with 1-point and 2-point statistics of earthquake source parameters, *Geophysical Journal International*, 196(3), ggt479–1786, doi:10.1093/gji/ggt479.
- Spudich, P., and K. B. Olsen (2001), Fault zone amplified waves as a possible seismic hazard along the Calaveras fault in central California, *Geophys. Res. Lett.*, 28(13), 2533–2536.
- Taborda, R., and J. Bielak (2013), Ground-Motion Simulation and Validation of the 2008 Chino Hills, California, Earthquake, *Bulletin of the Seismological Society of America*, 103(1), 131–156, doi:10.1785/0120110325.
- Taborda, R., and J. Bielak (2014), Ground-Motion Simulation and Validation of the 2008 Chino Hills, California, Earthquake Using Different Velocity Models, *Bulletin of the*

Seismological Society of America, 104(4), 1876–1898, doi:10.1785/0120130266.

Wang, W., and P. M. Shearer (2017), Using direct and coda wave envelopes to resolve the scattering and intrinsic attenuation structure of Southern California, *Journal of Geophysical Research: Solid Earth*, 100(B12), 24,015, doi:10.1002/2016JB013810.

Withers, K. B. (2016), Ground motion and variability from 3-D deterministic broadband simulations, 31 August.

Yao, Q. (2017), Dynamic Modeling of Earthquake Sources on Rough Faults, Ph.D. Thesis. SDSU/UCSD Joint Ph.D. Program in Geophysics, 30 August.

Chapter 2

Model for Small-Scale Crustal Heterogeneity in Los Angeles Basin Based on Sonic Log Data

High-frequency seismic ground motion (10+ Hz), as needed for earthquake engineering design purposes, is largely controlled by the meter-scale structure of the earth's crust. However, state-of-the-art velocity models poorly resolve small-scale features of the subsurface velocity and density variation. We invert 35 sonic logs (up to 3000 m in depth) in and near Los Angeles basin, CA, to obtain a statistical description of the small-scale heterogeneities of the basin. Assuming a von Karman autocorrelation function, our analysis finds Hurst numbers, ν , between 0.0-0.2, vertical correlation lengths, a_z of 15-150 m, and standard deviations of about 5% characterize the variability in the borehole data. We report average parameters for Los Angeles basin of $\nu = 0.064 (0.058, 0.069) \pm 0.01, (0.006, 0.012)$ and $a_z = 54 (51.1, 57.6) \text{ m}, \pm 5.9 (1.79, 9.53) \text{ m}$ with 95% confidence intervals listed in the parentheses. Despite the large depth range of the logs, there is no significant variation of the statistical parameters with depth. Our analysis of 371 depth-averaged shear wave velocities in the upper 30 m, V_{s30} , provides only an upper bound of basin scale-length estimates due to the coarse sampling distance, with a Hurst number of about 0.3 and lateral correlation lengths, a_x , of 5-10 km.

2.1 Introduction

Gaining a solid understanding of the expected range of ground motions from future large earthquakes is imperative so structures can be properly engineered to withstand this shaking. For example, crucial lifelines such as railroads, bridges, and pipelines span large horizontal distances across the ground where they experience significant variability in seismic velocities down to the smallest scale-lengths. Local site effects can drastically

amplify the ground motion from an earthquake, with prominent examples from the 1989 Loma Prieta earthquake in the Marina District of San Francisco and the 1985 Michoacan earthquake in Mexico City. Understanding this small-scale variability is necessary as simulations begin to resolve seismic wavelengths on the scale length of several meters. Until recently, fully 3D physics-based simulation of ground motions has been limited to around 1 Hz (e.g., *Olsen et al.*, [2006]; *Graves et al.*, [2008]; *Olsen et al.*, [2008; 2009]; *Roten et al.*, [2011]), while higher frequencies are modeled by stochastic or semi-stochastic approaches (e.g., *Frankel*, [2009]; *Crempien and Archuleta*, [2015]; *Graves and Pitarka*, [2015]; *Olsen and Takedatsu*, [2015]). Due to a surge in available powerful supercomputing resources and parallel/GPU programming, fully deterministic seismic wave propagation to frequencies of 10 Hz+ is now computationally within reach [*Cui et al.*, 2013; *Taborda and Bielak*, 2013]. Along with better understanding source complexity [*Shi and Day*, 2013] and non-linearity [*Roten et al.*, 2014], another limiting factor in realistic high-frequency ground motion simulations remains in characterizing the small-scale variation (from meters to tens of meters) of crustal seismic velocities. Current velocity models (e.g. the Southern California Earthquake Center (SCEC) Community Velocity Model (CVM)-S 4.0 [*Magistrale et al.*, 2000; *Kohler et al.*, 2003], CVM-H 11.9 [*Süss and Shaw*, 2003] and CVM-SI 4.26 [*Lee et al.*, 2011]) insufficiently resolve the small-scale variability known to exist in the earth's crust.

Due to large data acquisition costs at meter-scale resolution, it is not feasible to resolve the small-scale structure using direct measurements. Instead, the small-scale heterogeneities may be described by their two-point statistics (i.e., their spatial variability). While statistical representations of small-scale heterogeneities naturally do not include the actual physical locations of the short-wavelength velocity and density variations, ensembles of ground motion simulations computed using such distributions should reflect the correct average and variability observed in strong-motion recordings. A requirement for the simulated ground motion level and variability to be realistic is to use constraints from the statistics of actual small-scale variation observed in the earth's crust.

Frankel and Clayton, [1986] showed that the von Karman autocovariance function is well suited for characterizing the statistical variation of crustal heterogeneities. The functional form of the von Karman autocovariance function

$$\Phi_{\nu,a}(r) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{r}{a}\right)^\nu K_\nu\left(\frac{r}{a}\right), \quad 2.1$$

where ν is the Hurst exponent, a is the correlation length, K_ν is the modified Bessel function of order ν , $\Gamma(\nu)$ is the gamma function, and σ^2 is the variance with Fourier transform

$$P(k) = \frac{\sigma^2 (2\sqrt{\pi}a)^E \Gamma(\nu + E/2)}{\Gamma(\nu) (1 + k^2 a^2)^{\nu + E/2}}, \quad 2.2$$

where E is the Euclidean dimension. Previous works have constrained the values of the von Karman parameters - the correlation length a , Hurst exponent ν , and standard deviation σ from different data sets, such as sonic logs, digitized geological maps, and seismic reflection data (e.g., *Levander*, [1992; 1994]; *De et al.*, [1994]; *Wu et al.*, [1994]; *Holliger*, [1996; 1997]; *Dolan et al.*, [1997]; *Pullammanappallil et al.*, [1997]; *Nakata and Beroza*, [2015]). These studies report a wide range of Hurst exponents, 0.0-0.5, and correlation lengths between 30-300 m in the vertical direction, and horizontal to vertical anisotropy (a_x/a_z) between about 2 and 5. We attribute the wide range of parameters to the type of data from which the von Karman parameters were derived, and their analysis. We find upper crustal sonic logs tend to produce lower Hurst exponent estimations [*Holliger*, 1996; *Dolan et al.*, 1997] as compared with lower crustal data [*Levander*, 1992]. There remains some disagreement between correlation lengths observed in the data, which appears to be related to whether the authors modeled residuals or velocities [*Dolan et al.*, 1997; *Pullammanappallil et al.*, 1997] or used cross-correlation versus spectral methods [*Wu et al.*, 1994; *Dolan et al.*, 1997; *Pullammanappallil et al.*, 1997; *Nakata and Beroza*, 2015]. In this study, we aim to quantify the stochastic small-scale variability observed after removing a depth-dependent mean. For our purposes, we assume the CVMs to contain the deterministic or background portion of the velocity structure.

Several studies have also assessed the statistical parameters based on wave propagation simulations. *Hartzell et al.*, [2010] used $\nu=0.0$ with $a_x=5-10$ km and $\sigma=5-10\%$

based on results from *Frankel and Clayton*, [1986]. *Jacobsen and Olsen*, [2011] included statistical distributions with $\nu=(-0.5, 0, 0.5)$, $\alpha=250$ m and $\sigma=5-10\%$. *Imperator and Mai*, [2012] and *Bydlon and Dunham*, [2015] used $\nu=0.3$ with horizontal and vertical correlation lengths on the order of hundreds of meters to kilometers, also including multi-scale models created by the superposition of autocorrelation functions with multiple correlation lengths. *Frankel and Clayton*, [1986] generated small-scale heterogeneities with three distinct autocorrelation functions, Gaussian, exponential (corresponding to $\nu=0.5$), and self-similar (corresponding to $\nu=0.0$). By testing end-member cases (i.e. $\nu=0.0, 0.5$), *Frankel and Clayton*, [1986] generated von Karman models of small-scale heterogeneities with a preferred value of $\nu=0.0$ and correlation lengths between 5 and 10 km. The considerable variation in the preferred parameters of the autocorrelation found in the studies reflects the poor constraints from the underlying data. For example, the correlation lengths suggested from *Frankel and Clayton*, [1986] (>10 km) were obtained using teleseismic sources, which provide limited constraints at near-field stations. Other studies included limited or no data constraints in their modeling.

The primary goal of this study is to understand the spatial statistics of the small-scale variation in the seismic velocity structure in order to provide the community with a data-derived model of small-scale heterogeneities in Los Angeles basin. Toward this goal, we use variogram analysis on two separate data sets consisting of V_{S30} measurements and borehole sonic logs to estimate the parameters needed to model the small-scale variability for von Karman autocorrelation functions (Hurst exponent, standard deviations, and correlation length), so we perform our statistical analysis on the sonic log slowness residuals.

2.2 Methods

To estimate the parameters necessary to characterize the lateral variability observed in V_{S30} and the vertical variability observed in sonic log measurements we introduce two geostatistical methods: the classical variogram [*Matheron*, 1963] and the moving-window variogram [*Li and Lake*, 1994]. In the following we briefly summarize these two methods,

along with the Monte Carlo inversion that we use to estimate the von Karman parameters from the sonic logs.

2.2.1 Classical Variogram

We analyze the Vs30 data in the time-domain using the classical estimate of semi-variance given by *Matheron*, [1963] in (2.3).

$$\tilde{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2, \quad 2.3$$

where $N(h)$ is the number of data points at each lag distance, h is the lag distance, and $Z(x_i)$ is spatial random field at location x_i . This variogram estimator has a vetted history in geostatistics since its derivation and is frequently used in kriging algorithms where developing a model of spatial correlation is paramount. The factor of 2 appears in the denominator accounting for the double representation of field values (x_1, x_2) and (x_2, x_1) in the summation. The estimate requires binning of the data to specify lag distances, and we find this estimate performs best when there are near-equal representations of the data in each bin. We can directly interpret the correlation length of the velocity structure given an estimate of the variogram, while we must fit a model to the variogram to determine the Hurst exponent. Here, we model the variogram (2.4) using a fractional Brownian motion (fBm) form [*Li and Lake*, 1994; *Mela and Louie*, 2001], based on the results from previous studies [*Frankel and Clayton*, 1986; *Wu et al.*, 1994; *Holliger*, 1996; *Pullammanappallil et al.*, 1997; *Nakata and Beroza*, 2015]. Following (2.4) the Hurst exponent can be estimated as the half-slope of the linear regression in logarithmic space.

$$\begin{aligned} \tilde{\gamma}(h) &= \frac{\sigma^2}{a^{2\nu}} h^{2\nu} & h < a \\ \tilde{\gamma}(h) &= \sigma^2 & h \geq a \end{aligned} \quad 2.4$$

where $\gamma(h)$ is the semi-variance and a is the correlation length. This model follows a power-law for lags smaller than the correlation length and tends to a semi-variance equal

to σ^2 after the correlation length. The constant $\sigma^2/a^{2\nu}$ implies some interdependence of the two parameters on the resulting semi-variance estimate.

2.2.2 Moving–Window Variogram

We introduce an alternative variogram estimator (moving-window formulation) derived by *Li and Lake*, [1994] to analyze the sonic log measurements:

$$\hat{\gamma}_{N1}(h) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{1}{2m} \sum_{j \in D_{i,h}} [Z(x_i) - Z(x_j)]^2 \right\}. \quad 2.5$$

Here, m is the number of data points in $D_{i,h}$ which is the index set of the data points contained within the moving window, $\Delta_{i,h}$, which has size h . n is the total number of data points in the random field, Z . $Z(x_i)$ represents the midpoint of $\Delta_{i,h}$, and $Z(x_j)$ represents each data point contained within $D_{i,h}$ excluding $Z(x_i)$. The inner summation is over the moving-window while the outer sum is over the entire random field. For a more detailed explanation and derivation of the moving-window variogram, please refer to *Li and Lake*, [1994]. The moving-window formulation (2.5) provides an estimate utilizing all available data pairs at all lag distances, whereas other variogram estimates (such as the *Matheron*, [1963] method, discussed above) provide fewer data pairs as the observation distance increases resulting in larger variance as the lag distance increases (for example, see Figure 2.1).

We choose to use the moving-window variogram method for the borehole logs as it reduces the variance at large lag distances which aids our inversion, because the moving-window estimate provides more consistent mean-squared residual values. The choice of variogram method does not change the results of the inversion. Also, the moving-window variogram does not always provide a more accurate direct estimate as compared to the *Matheron*, [1963] method. For example, we find the classical variogram provides a more stable result for the V_{S30} data set.

2.2.3 Monte–Carlo Inversion

We incorporate the moving-window variogram method into an inversion routine which performs a grid search on the parameter space $\nu \in [0.0, 0.3]$ and $a_z \in [15 \text{ m}, 150 \text{ m}]$, while minimizing the mean-squared residual between each sonic log variogram and variogram estimated from a synthetic borehole with pre-specified, ν and a_z . Here, we assume that each sonic log is a 1D realization of a single random field of velocity fluctuations representing the entire Los Angeles basin, and the statistics are stationary. We invert each borehole independently resulting in an estimate of the von Karman parameters for each borehole. This allows us to quantitatively assign uncertainty to our estimation.

To pre-processes the logs before inversion, we first detrend the data using a long-period median filter providing a zero-mean profile. We find the choice of demeaning filter can significantly affect the estimation of the correlation length. For example, choosing too short of a long-period median filter can force the correlation length to be proportional with the window size of the median filter. Due to large variability in the data, we find choosing long period median filters proportional to the borehole depth extent provide more stable estimates of the correlation length than low-order polynomial regression (see Figure. 2.1c). Next, we apply a short-period median filter to remove artifacts resulting from the downhole logging procedure using a median filter with window size of 1.5 m, corresponding to the approximate resolution of the sonic logging tool. Also, we performed a sensitivity analysis on the effect of short-period median filtering the logs and find the results are robust with filter lengths $< 5 \text{ m}$. We assume that the large spikes seen in the logs are due to measurement errors resulting from either decoupling with the logging tool and the host rock or cycle skips, a phenomenon resulting in anomalously high transit times or low velocities. Figure 2.1a,b shows the preprocessing method performed on one borehole log, namely apall located approximately 5 km north of Long Beach, CA, along with the classical and moving-window variogram (Figure. 2.1c) estimates for this single borehole. We perform all statistical analysis in the slowness domain. Notice the larger variance associated with the classical estimate of semi-variance. For the remainder of the manuscript we refer to the pre-processed zero-mean logs as fluctuation profiles.

For the inversion, (1) we estimate the semi-variance using (2.5) for each borehole. (2) We simulate 500 realizations for each combination of parameters (ν, a) and estimate the semi-variance for each realization using (2.5). (3) We compute the mean-squared residual between the borehole variogram and each synthetic variogram. (4) We average the residuals over all realizations to produce a single expected mean-squared residual value for each von Karman parameter pair. (5) We accept any value with a mean-squared residual less than or equal to a absolute tolerance value of 0.01, visually selected from all the inversion results as the value above which the mean-squared residual increases significantly. This methodology is similar to the inversion routine used by *Holliger*, [1996].

2.3 Statistics of Los Angeles Basin

First, we analyze the lateral variation of near-surface time-averaged shear wave velocity measurements in greater Los Angeles. We compile 639 V_{s30} measurements consisting of 350 measurements incorporated in the SCEC CVM-S 4.0 [*Magistrale et al.*, 2000], 102 measurements from *Louie*, [2005; 2007] and 187 measurements from *Yong et al.*, [2013]. Next, we show our inversion results for the von Karman parameters derived from the 35 V_p sonic logs in Los Angeles basin.

2.3.1 V_{s30} Measurements

V_{s30} is defined as

$$V_{s30} = \frac{30 \text{ m}}{\sum \frac{d}{V_s}} \quad 2.6,$$

where d is the layer thickness (m) from the 1D profile below the site of interest, and V_s is the depth-dependent S-wave velocity (m/s). We choose a subset of the measurements located near Los Angeles basin resulting in 371 values used in the analysis (see Figure 2.2a) to maintain consistency with the spatial extent of the sonic log analysis (see the following section). Since the V_{s30} measurements are sparse and irregularly located, we

assign a relatively large bin spacing for the variogram of 500 m. Also, to obtain a more uniform distribution of observation pairs at all lag distances, we restrict the maximum lag to 30 km.

Figure 2.2b shows the classical variogram estimated from the 371 V_{S30} data points. We interpret the correlation length of the heterogeneity in the V_{S30} values to be approximately 5-10 km (where the variogram approaches a constant value), and the Hurst exponent to be $\nu = 0.33 \pm 0.09$. This analysis provides an understanding of the statistical parameters of larger-scale basin features only, as the V_{S30} measurements lack the resolution to constrain smaller-scale variation. Incidentally, the relatively large Hurst exponent and correlation lengths estimated from the V_{S30} data are consistent with the results by *Frankel and Clayton*, [1986], *Imperator and Mai*, [2012] and *Levander*, [1994]. We assume the SCEC CVMs have already accounted for these larger-scale variations, and we plan to constrain the parameters of the small-scale variations (on the order of tens of meters) using the densely sampled sonic logs. We note that current state-of-the-art CVMs may artificially taper V_{S30} values from the near-surface to deeper sections of the models, possibly biasing velocity variability in the deeper sections.

2.3.2 Sonic Log Measurements

In order to estimate the small-scale structure in Los Angeles basin, we require much more densely sampled data than that provided by V_{S30} measurements, as discussed above. Velocity logs provide regularly and densely sampled direct measurements of in situ seismic velocities. We note that our analysis only provides constraints regarding heterogeneity observed in the vertical direction.

We have acquired 38 V_p sonic logs (*Andreas Plesch, pers. comm.*, [2012]) located throughout the Los Angeles basin, extending to more than 3 km depth and containing more than 300,000 measurements shown in Figure 2.3. Some logs are missing sections of data, and these values are omitted from the analysis as variograms do not require the data to be evenly spaced. We also discard three boreholes due to significant sections of poor data quality or short sample lengths from which accurate correlation length estimates within our

parameter space are not possible. Due to the spatial extent of the boreholes, we assume the boreholes provide a representative sample of the spatial statistics in Los Angeles basin. Assuming a dense enough network of sonic log measurements, one could also compute horizontal statistics of depth-averaged sections of adjacent boreholes allowing for a much higher resolution lateral study than available from the V_{S30} data similar to the analysis by *Wu et al.*, [1994].

The final inversion result for the 35 boreholes is shown in Figure 2.4 which indicates the percentage of boreholes that can be modeled by a given von Karman parameter pair within a mean-squared residual of 0.01. We find that parameters ranging from $\nu = (0.0 - 0.2)$ and $a_z = (15 \text{ m} - 150 \text{ m})$ represent the variability to an agreement greater than approximately 90% of all 35 Los Angeles Basin sonic logs. Figure 2.4 shows a trade-off between ν and a_z meaning that while large a_z are acceptable with low ν and vice-versa, large ν fail to represent the Los Angeles sonic logs even at $a_z = 45 \text{ m}$. On the other hand, at $\nu = 0.0$, large correlation lengths can provide seemingly reasonable fits to the data. This is because the semi-variance at the nugget (or first lag distance) is approximately equal to the total variance in the sonic logs making the transition from correlated to uncorrelated behavior hard to distinguish in these variogram estimates. For example, in Figure 2.5 we show that a model with $\nu = 0.0$ and $a_z = 500 \text{ m}$ represents the statistics of the sonic logs, where the gray region represents the range of all Los Angeles basin sonic log variograms. This highlights the interdependence between a_z and ν in estimates of semi-variance. Also, we show variograms for several combinations of ν and a_z that could represent reasonable values for Los Angeles Basin. Recent work by *Nakata and Beroza*, [2015] shows that near Long Beach in Los Angeles basin vertical correlation lengths are on the order of 100 m with an anisotropy factor, a_x/a_z of approximately 5.

Based on a bootstrapped analysis of the average values calculated for each borehole, we find average values of $\nu = 0.064 (0.058, 0.069) \pm 0.01, (0.006, 0.012)$ and $a_z = 54 (51.1, 57.6) \text{ m}, \pm 5.9 (1.79, 9.53) \text{ m}$, where the 95% confidence intervals for the mean and standard deviation are shown in the parentheses. We compute the average values by applying a weighting function to the inversion residuals such that models with lower residual values are weighted higher. The electronic supplement shows the mean-squared

residuals computed for each borehole. We also show the raw and processed versions of the boreholes used in the inversion along with moving-window and classical semi-variance estimates.

Finally, we analyze the data set to determine if there are any depth-dependent effects. Toward this goal, we window the data set in 500 m segments and apply traditional variogram analysis to the windowed data, meaning no inversion is performed on the depth-windowed data. For this analysis, we pre-process the logs as described in the Methods section and stack the resulting moving-window variogram estimates to produce a single variogram representative of a given depth-interval. Figure 2.6 shows the results from our analysis. This figure suggests the lack of any significant depth dependency for the von Karman parameters. We also find no depth-dependence for σ . Even though there might be depth-dependency for some of the parameters at individual boreholes (e.g. $\sigma(z)$ in `apal1` shown in Figure 2.1), on average, our analysis suggests limited depth-dependence of the von Karman parameters in Los Angeles basin.

2.4 Conclusions

We have used 371 V_{S30} measurements for southern California and 35 borehole sonic logs from the Los Angeles basin to constrain the parameters describing von Karman distributions of small-scale heterogeneities. The borehole logs indicate that the ranges of ν and a most representative of the velocity variability are ν , between 0.0-0.2, vertical correlation lengths, a_z of 15-150 m, and standard deviations of about $5 \pm 2.5\%$. We find average values of $\nu = 0.064$ (0.058, 0.069) ± 0.01 , (0.006, 0.012) and $a_z = 54$ (51.1, 57.6) m, ± 5.9 (1.79, 9.53) m with 95% confidence intervals listed in the parentheses. The V_{S30} data provide only an upper bound on the parameters ($\nu = 0.33 \pm 0.09$ and $a_x = 5 - 10$ km) due to the relatively coarse sampling. We find insignificant depth dependency of the von Karman parameters from 500 m to 2500 m depth. These statistical parameters constitute a data-driven model that can be used for high-frequency simulations of seismic wave propagation in Southern California. The sonic logs and V_{S30} data sets constrain the vertical and lateral features, respectively. We hesitate to infer an anisotropy ratio between

the two data sets due to the large differences in resolution between the V_{S30} and borehole data. We suspect that higher resolution V_{S30} data, on say a 50 m grid, may help refine the results presented here.

The parameters estimated from the sonic log analysis section of the manuscript only provide information regarding the P-wave heterogeneity in Los Angeles basin. Directly applying these parameters to the S-wave data implies some correlation between these two measurements. Based on simple elastic theory and analysis of V_p and V_s sonic logs [Dolan *et al.*, 1997] we expect some correlation between the two sets of parameters as they are both dependent on the shear modulus and the density of the material. However, without access to full-sonic scanner measurements (i.e. P-wave and S-wave logs) at the same locations, we cannot definitively comment on the correlation between P-wave and S-wave heterogeneity for this data set. The level of correlation does have implications for modeling of the small-scale heterogeneities, as uncorrelated or non-perfectly correlated heterogeneity models must simulate separate realizations of the stochastic field for both P-wave and S-wave velocities.

Our inversion procedure provides limited resolution of a_z for $\nu \approx 0.0$. This means larger correlation lengths may accurately model the spatial statistics observed in Los Angeles basin. Also, we find an interdependence between ν and a_z for all combinations of von Karman parameters. Care must be taken to choose appropriate parameters of ν and a_z that fall within the accepted range for Los Angeles basin. The black line in Figure 2.4 shows the values of ν and a_z that represent accepted von Karman parameters for the majority of borehole sonic logs in Los Angeles basin. We recommend choosing parameters that fall within this region, however one could argue for choosing large values of a_z for $\nu = 0.0$.

2.5 Acknowledgements

We thank Andreas Plesch and John Shaw for providing the deep borehole logs, and to John Louie, Harold Magistrale, and Alan Yong for providing us with V_{S30} measurements throughout Los Angeles basin. We are grateful for stimulating comments from Robert

Graves and an anonymous reviewer that helped strengthen the manuscript. This research was supported by NSF awards OCI-1148493 (SI2-SSI), EAR-1135455 (FESD) and EAR-1226343 EAR-1349180, the Southern California Earthquake Center (SCEC), and U.S. Geological Survey award G12AP20015. SCEC is funded by NSF Cooperative Agreement EAR-1033462 and USGS Cooperative Agreement G12AC20038. This is SCEC contribution #6178. Both the V_{s30} and borehole data used in this publication can be found at <http://www-rohan.sdsu.edu/~kbolsen/vonkarman.html>. The code used for this report is available upon request from authors.

References

- Bydlon, S. A., and E. M. Dunham (2015), Rupture dynamics and ground motions from earthquakes in 2-D heterogeneous media, *Geophys. Res. Lett.*, *42*(6), 1701–1709, doi:10.1002/2014GL062982.
- Crempien, J. G. F., and R. J. Archuleta (2015), UCSB Method for Simulation of Broadband Ground Motion from Kinematic Earthquake Sources, *Seismological Research Letters*, *86*(1), 61–67, doi:10.1785/0220140103.
- Cui, Y. et al. (2013), *Physics-based seismic hazard analysis on petascale heterogeneous supercomputers*, ACM, New York, New York, USA.
- De, G. S., D. F. Winterstein, and M. A. Meadows (1994), Comparison of P- and S-wave velocities and Q's from VSP and sonic log data, *GEOPHYSICS*, *59*(10), 1512–1529, doi:10.1190/1.1443541.
- Dolan, S., C. Bean, and B. Riollot (1997), The broad-band fractal nature of heterogeneity in the upper crust from petrophysical logs, *Geophysical Journal International*, *132*(3), 489–507, doi:10.1046/j.1365-246X.1998.00410.x.
- Frankel, A. (2009), A Constant Stress-Drop Model for Producing Broadband Synthetic Seismograms: Comparison with the Next Generation Attenuation Relations, *Bulletin of the Seismological Society of America*, *99*(2A), 664–680, doi:10.1785/0120080079.
- Frankel, A., and R. Clayton (1986), Finite difference simulations of seismic scattering: Implications for the propagation of short-period seismic waves in the crust and models of crustal heterogeneity, *Journal of Geophysical Research: Solid Earth*, *91*(B6), 6465–6489.

- Graves, R. W., and A. Pitarka (2015), Refinements to the Graves and Pitarka (2010) Broadband Ground-Motion Simulation Method, *Seismological Research Letters*, 86(1), 75–80, doi:10.1785/0220140101.
- Graves, R. W., B. T. Aagaard, K. W. Hudnut, K. W. Hudnut, L. M. Star, J. P. Stewart, and T. H. Jordan (2008), Broadband simulations for Mw 7.8 southern San Andreas earthquakes: Ground motion sensitivity to rupture speed, *Geophys. Res. Lett.*, 35(22), L22302, doi:10.1029/2008GL035750.
- Hartzell, S., S. Harmsen, and A. Frankel (2010), Effects of 3D random correlated velocity perturbations on predicted ground motions, *Bulletin of the Seismological Society of America*, 100(4), 1415–1426, doi:10.1785/0120090060.
- Holliger, K. (1996), Upper-crustal seismic velocity heterogeneity as derived from a variety of P-wave sonic logs, *Geophysical Journal International*, 125(3), 813–829, doi:10.1111/j.1365-246X.1996.tb06025.x.
- Holliger, K. (1997), Seismic scattering in the upper crystalline crust based on evidence from sonic logs, *Geophysical Journal International*, 128(1), 65–72.
- Imperator, W., and P. M. Mai (2012), Broad-band near-field ground motion simulations in 3-dimensional scattering media, *Geophysical Journal International*, 192(2), 725–744, doi:10.1093/gji/ggs041.
- Jacobsen, B. H., and K. B. Olsen (2011), Spatial variability of ground motion amplification from low-velocity sediments including fractal inhomogeneities with special reference to the Southern California basins, *EGU General ...*.
- Kohler, M. D., H. Magistrale, and R. Clayton (2003), Mantle Heterogeneities and the SCEC Reference Three-Dimensional Seismic Velocity Model Version 3, *Bulletin of the Seismological Society of America*, 93(2), 757–774, doi:10.1785/0120020017.
- Lee, E.-J., P. Chen, T. H. Jordan, and L. Wang (2011), Rapid full-wave centroid moment tensor (CMT) inversion in a three-dimensional earth structure model for earthquakes in Southern California, *Geophysical Journal International*, 186(1), 311–330, doi:10.1111/j.1365-246X.2011.05031.x.
- Levander, A. R. (1992), A Stochastic View of Lower Crustal Fabric Based on Evidence From the Ivrea Zone, *Geophys. Res. Lett.*, 19(11), 1153–1156.
- Levander, A. R. (1994), Structure and seismic response of extended continental crust: Stochastic analysis of the Strona-Ceneri and Ivrea zones, Italy, *Geology*.
- Li, D. C., and L. W. Lake (1994), A Moving Window Semivariance Estimator, *Water Resources Research*, 30(5), 1479–1489, doi:10.1029/94WR00319.
- Louie, J. N. (2005), Improving next-generation attenuation models with shear-velocity

- measurements at all TriNet and strong-motion stations in LA, *NEHRP Final Technical Report*.
- Louie, J. N. (2007), Shear-wave velocity map for California: Collaborative research with CGS and UNR, *NEHRP Final Technical Report*.
- Magistrale, H., S. M. Day, R. Clayton, and R. W. Graves (2000), The SCEC Southern California Reference Three-Dimensional Seismic Velocity Model Version 2, *Bulletin of the Seismological Society of America*, 90(6B), S65–S76, doi:10.1785/0120000510.
- Matheron, G. (1963), Principles of geostatistics, *Economic Geology*, 58(8), 1246–1266, doi:10.2113/gsecongeo.58.8.1246.
- Mela, K., and J. N. Louie (2001), Correlation length and fractal dimension interpretation from seismic data using variograms and power spectra, *GEOPHYSICS*, 66(5), 1372–1378, doi:10.1190/1.1487083.
- Nakata, N., and G. C. Beroza (2015), Stochastic characterization of mesoscale seismic velocity heterogeneity in Long Beach, California, *Geophysical Journal International*, 203(3), 2049–2054, doi:10.1093/gji/ggv421.
- Olsen, K. B. et al. (2009), ShakeOut-D: Ground motion estimates using an ensemble of large earthquakes on the southern San Andreas fault with spontaneous rupture propagation, *Geophys. Res. Lett.*, 36(4), L04303, doi:10.1029/2008GL036832.
- Olsen, K. B., and R. Takedatsu (2015), The SDSU Broadband Ground-Motion Generation Module BBtoolbox Version 1.5, *Seismological Research Letters*, 86(1), 81–88, doi:10.1785/0220140102.
- Olsen, K. B., S. M. Day, J. B. Minster, Y. Cui, A. Chourasia, D. Okaya, P. Maechling, and T. Jordan (2008), TeraShake2: Spontaneous Rupture Simulations of Mw 7.7 Earthquakes on the Southern San Andreas Fault, *Bulletin of the Seismological Society of America*, 98(3), 1162–1185, doi:10.1785/0120070148.
- Olsen, K. B., S. M. Day, J. B. Minster, Y. Cui, A. Chourasia, M. Faerman, R. Moore, P. Maechling, and T. Jordan (2006), Strong shaking in Los Angeles expected from southern San Andreas earthquake, *Geophys. Res. Lett.*, 33(7), L07305, doi:10.1029/2005GL025472.
- Pullammanappallil, S., A. R. Levander, and S. P. Larkin (1997), Estimation of crustal stochastic parameters from seismic exploration data, *Journal of Geophysical Research: Solid Earth*, 102(B7), 15269–15286, doi:10.1029/97JB01144.
- Roten, D., K. B. Olsen, H. Magistrale, J. C. Pechmann, and V. M. Cruz-Atienza (2011), 3D Simulations of M 7 Earthquakes on the Wasatch Fault, Utah, Part I: Long-Period (0-1 Hz) Ground Motion, *Bulletin of the Seismological Society of America*, 101(5),

2045–2063, doi:10.1785/0120110031.

- Roten, D., K. B. Olsen, S. M. Day, Y. Cui, and D. Faeh (2014), Expected seismic shaking in Los Angeles reduced by San Andreas fault zone plasticity, *Geophys. Res. Lett.*, *41*(8), 2769–2777, doi:10.1002/2014GL059411.
- Shi, Z., and S. M. Day (2013), Rupture dynamics and ground motion from 3-D rough-fault simulations, *Journal of Geophysical Research: Solid Earth*, *118*(3), 1122–1141, doi:10.1002/jgrb.50094.
- Süss, M. P., and J. H. Shaw (2003), P wave seismic velocity structure derived from sonic logs and industry reflection data in the Los Angeles basin, California, *Journal of Geophysical Research*, *108*(B3), 2170, doi:10.1029/2001JB001628.
- Taborda, R., and J. Bielak (2013), Ground-Motion Simulation and Validation of the 2008 Chino Hills, California, Earthquake, *Bulletin of the Seismological Society of America*, *103*(1), 131–156, doi:10.1785/0120110325.
- Wu, R. S., Wu, R.-S., Z. Xu, and X.-P. Li (1994), Heterogeneity spectrum and scale-anisotropy in the upper crust revealed by the German Continental Deep-Drilling (KTB) Holes, *Geophys. Res. Lett.*, *21*(10), 911–914, doi:10.1029/94GL00772.
- Yong, A., A. Martin, K. Stokoe, and J. Diehl (2013), ARRA-funded Vs30 measurements using multi-technique approach at strong-motion stations in California and central-eastern United States, *U.S. Geological Survey Open-File Report*.

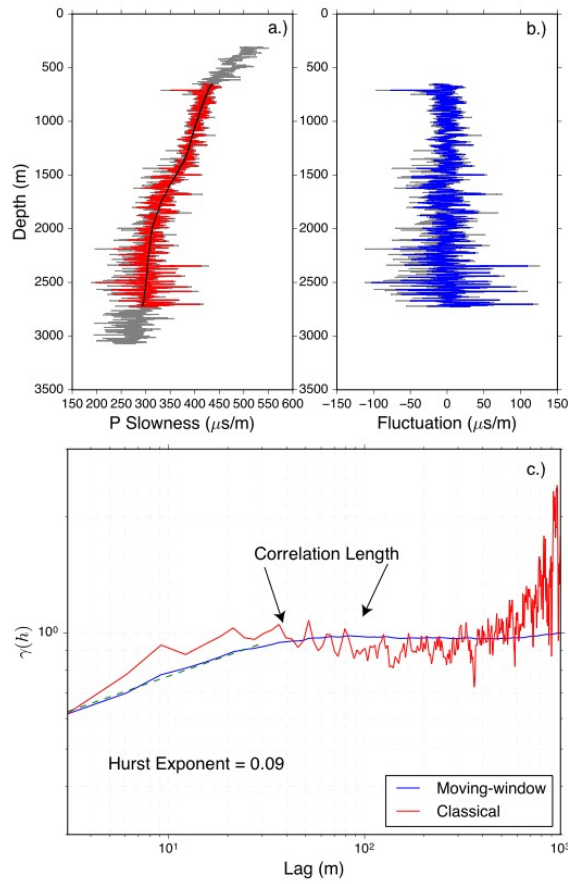


Figure 2.1: a) Apall Vp sonic log (gray) with long-period median filter (black), and 1.5 m smoothed log (red). b) De-trended fluctuation profile (blue) generated by subtracting the long-period median filter (black) from the red and gray logs in a), and used to estimate variograms. The fluctuation profile (blue) becomes shorter due to edge effects from the median filter. We show logs in terms of slowness, because all data analysis is performed in the slowness domain. c) Variogram estimate from the fluctuation profile (blue) shown in b). The dashed green line is the logarithmic regression used to estimate the Hurst exponent. We estimate the correlation length to be between 40 and 90 m, and $\nu=0.09$. Notice the much larger variance associated with the classical estimate of *Matheron* [1963] as opposed with the moving-window estimate of *Li & Lake* [1994].

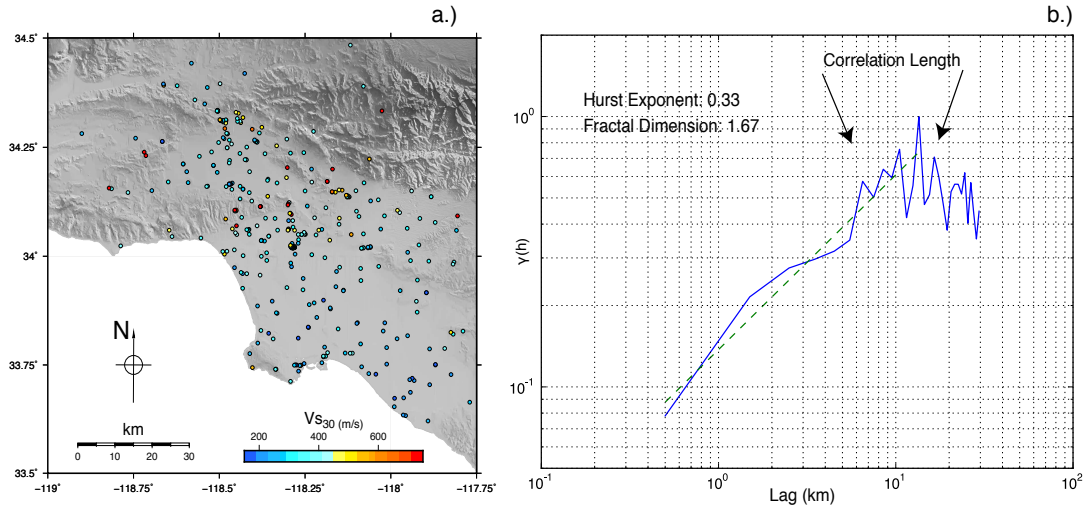


Figure 2.2: a) Map showing the 371 V_{s30} measurements used in our variogram analysis. We color the V_{s30} values to provide a visual representation of the spatial variability. b) Variogram estimated for the V_{s30} measurements in the Los Angeles basin. The dashed green line shows the logarithmic regression used to estimate the Hurst exponent, $\nu = 0.33$. We estimate lateral correlation lengths between 5 and 10 km.

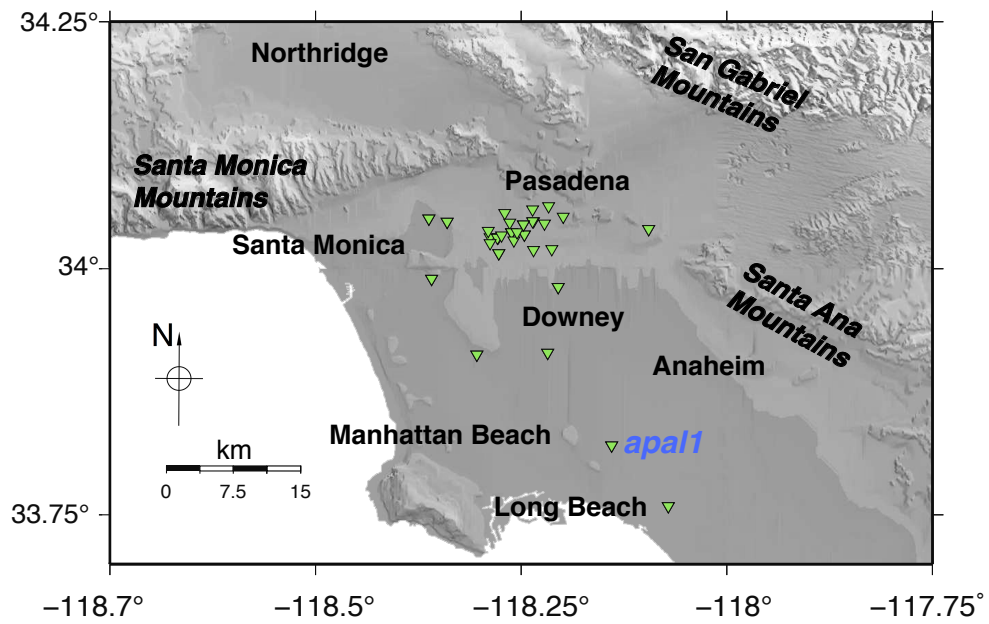


Figure 2.3: Map showing 35 deep borehole locations within the Los Angeles basin that are used for the inversion. Borehole apal1 is highlighted.

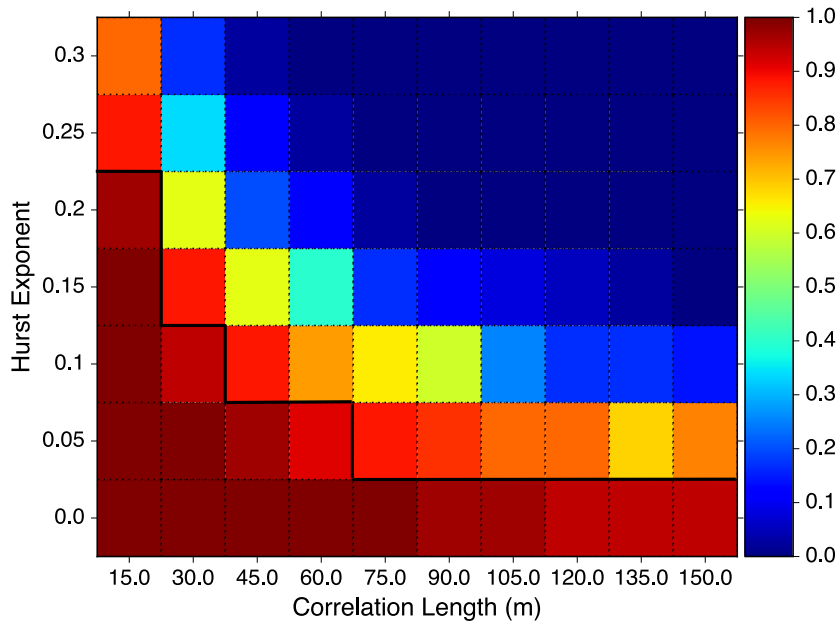


Figure 2.4: Results shown for the parameter spaced analyzed by our inversion routine. Here, we bin all accepted von Karman pairs and normalize by the total number of boreholes. A value of 1.0 indicates that a particular (v, a_z) pair satisfies the acceptance criteria for every sonic log. Likewise, a value of 0.0 has failed to accept the criterion at every borehole. The von Karman parameters that agree between more than approximately 90% of the borehole sonic logs are defined by (v, a_z) pairs below the solid line.

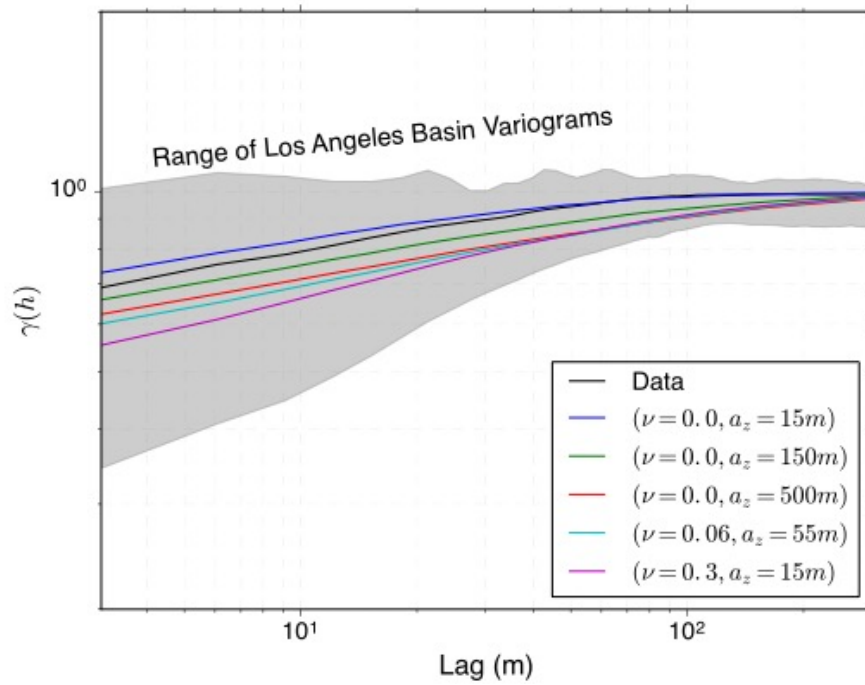


Figure 2.5: Selected variograms plotted for different combinations of von Karman parameters. The shaded gray region represents the range spanned by all data variograms. We show the variograms representing 100% acceptance as well as the variogram representing the average von Karman parameters for Los Angeles basin. Finally, we show the stacked data variogram computed from an average of each borehole variogram in black. Notice, the model with $\nu=0.0$ and $a_z=500$ m lies within the data variogram range in Los Angeles basin.

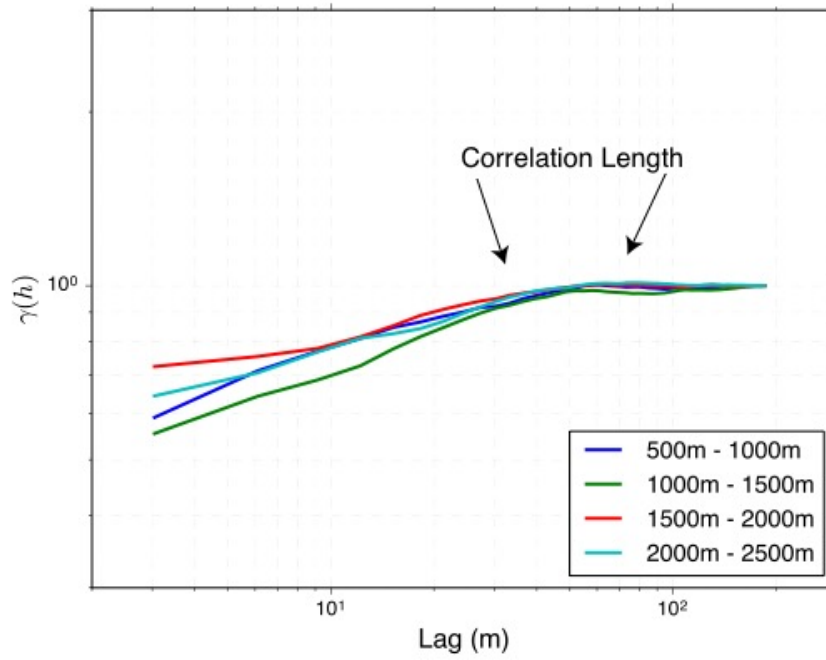


Figure 2.6: Depth dependent variogram computed for 500 m windows of the sonic log data set. We find no significant depth-dependence of the von Karman parameters. Each curve represents the stacked variogram over each specified depth interval.

Chapter 3

Ground Motion Simulation and Validation of the M_w 5.4 Chino Hills Earthquake in Scattering Media

We simulate 0 – 2.5 Hz deterministic wave propagation in 3D velocity models for the M_w 5.4 2008 Chino Hills, CA, earthquake using an inverted finite-fault source model and frequency-dependent anelastic attenuation. Small-scale heterogeneities are modeled as 3D random fields defined using an elliptically anisotropic von Karman autocorrelation function with its parameters constrained using Los Angeles basin borehole data. We superimpose the heterogeneity models on a leading deterministic Community Velocity Model (CVM) of southern California. We find that models of velocity and density perturbations can have significant effects on the wavefield at frequencies as low as 0.3 Hz, with ensemble median values of various ground motion metrics varying up to $\pm 50\%$ compared to those computed using only deterministic CVM. In addition, we show that frequency-independent values of the shear-wave quality factor (Q_{s0}) parameterized as $Q_{s0} = 150V_s$ (V_s in km/s) provides the best agreement with data when assuming the published M_w 5.4 ($M_0 = 1.6 \times 10^{17}$ Nm) of the inverted finite-fault source model. This model for Q_{s0} trades off with $Q_{s0} = 100V_s$ assuming M_w 5.5 ($M_0 = 2.2 \times 10^{17}$ Nm), which represents an upper bound of the M_w estimates for this event. We find the addition of small-scale heterogeneities provides limited overall improvement to the misfit between simulations and data, for the considered ground motion metrics, due to the primary sources of misfit originating from the deterministic CVM or the finite-fault source description.

3.1 Introduction

Gaining an understanding of the expected range of ground motions from future earthquakes remains a principal goal of seismologists and earthquake engineers. Due largely to continual developments in high-performance computing, deterministic

earthquake simulations have shown potential for their ability to accurately characterize the broadband seismic wavefield. For the task of predicting ground motions from future earthquakes, physics-based deterministic earthquake simulations have advantages over purely data-based approaches (e.g., *Abrahamson et al.*, [2014]; *Boore et al.*, [2014]; *Campbell and Bozorgnia*, [2014]; *Chiou and Youngs*, [2014]), because they do not inherently assume that future earthquakes behave like previous earthquakes. With that being said, deterministic simulations are not without their own limitations; namely, the difficulty in characterizing the problem in terms of physical parameters in addition to the computational expense associated with broadband regional scale simulations. Despite these limitations, deterministic simulations have proven their usefulness for unraveling questions regarding seismic hazard and wave propagation phenomena (e.g., *Li et al.*, [1994]; *Kawase*, [1996]; *Olsen*, [2000]; *Spudich and Olsen*, [2001]; *Aagaard and Heaton*, [2004]; *Day et al.*, [2008]; *Pitarka et al.*, [2009]; *Duan and Day*, [2010]; *Aagaard et al.*, [2010b]; *Roten et al.*, [2014]).

Deterministic wave propagation simulations require modelers to define the material properties of the earth (i.e., elastic moduli, viscoelastic, and potentially plastic parameters) along with a description of the earthquake source. Generally, these parameters fields are poorly constrained for small scale-lengths (i.e., $< \sim 100$ m), leading to a general consensus in the community that the numerical models describing these parameters are insufficient to support the complexities associated with high-frequency wave propagation. In spite of this assumption, two recent studies *Taborda and Bielak*, [2013] and *Taborda and Bielak*, [2014], hereafter TB13 and TB14, respectively), thoughtfully investigated the limitations of physics-based earthquake simulations with a specific focus on the July 29, 2008 Chino Hills earthquake in Southern California. In both studies, the authors simulate 4 Hz ground motions using Hercules [*Bielak et al.*, 2011], a parallel finite-element code, used to solve the elastodynamic equations governing seismic wave propagation. Specifically, the authors validate their simulations against strong motion records using different state-of-the-art community velocity models (CVMs). The validation process quantitatively evaluates whether a physical model accurately represents the system it intends to study. Based on their analysis, the authors find good comparisons below 1 Hz, but note discrepancies

between the wave propagation model and ground motion records at higher frequencies; however, they conclude that extending the frequency resolution past 1 Hz is a worthwhile effort.

Current state-of-the-art CVMs (e.g., *Magistrale et al.*, [2000]; *Kohler et al.*, [2003]; *Süss and Shaw*, [2003]; *Lee et al.*, [2014]) insufficiently resolve the structural complexity known to exist in the earth due predominately to the low cut-off frequency ($f_{max} \approx 0.5$ Hz) of the data used for 3D tomographic inversions or the limited data available to describe geologic structures at depth. Due to the cost and complexity of acquiring meter-scale seismic velocity models, a favorable approach to include velocity heterogeneity has been to superimpose statistical descriptions of the small-scale velocity fluctuations onto a deterministic velocity model (e.g., *Frankel and Clayton*, [1986]; *Hartzell et al.*, [2010]; *Imperator and Mai*, [2012]; *Bydlon and Dunham*, [2015]). Assuming these statistical distributions accurately represent the variability present in the earth, we can expect, at least in a statistical sense, the scattered wave field will be comparable with those originating from natural earthquakes. However, a caveat to this approach is that it requires ensembles of ground motions need to be considered to understand the expected ground motions caused by scattering media (e.g., *Imperator and Mai*, [2012]).

Largely, this manuscript extends the work done by TB13 and TB14 to include effects of small-scale velocity and density fluctuations. We study the effects of small-scale heterogeneity in seismic velocities and densities and their trade-offs with various attenuation models in the context of seismic hazard analysis. Toward this goal, we compare simulated seismograms against strong-motion data recorded for the 2008 M_w 5.4 Chino Hills earthquake. Building on TB13 and TB14 we investigate two other factors that influence simulated ground motions; namely, anelastic attenuation and statistical distributions of small-scale velocity and density fluctuations. A recent study by *Withers et al.*, [2015] demonstrated the necessity of including frequency-dependent attenuation for simulations of the 2008 Chino Hills earthquake, so we focus our efforts on refining the empirical relationships used to define the attenuation model. Several studies have investigated the effects of random media on scattering attenuation and coda waves (e.g., *Wu and Aki*, [1985]; *Frankel and Clayton*, [1986]; *Korn*, [1993]; *Imperator and Mai*,

[2012] and more recently *Bydlon and Dunham*, [2015; *Wang and Shearer*, [2017]]. Here, we address the effects of scattering attenuation with a focus on individual ground motion metrics relevant to seismic hazard. With the progress made by physics-based probabilistic seismic hazard techniques, such as the Southern California Earthquake Center (SCEC) CyberShake [*Graves et al.*, 2011] or the SCEC Broadband Platform [*Maechling et al.*, 2015], understanding the effects of scattering can help inform these large-scale efforts.

The July 29, 2008 Chino Hills Earthquake

At 11:42 a.m. PDT on July 29, 2008 the greater Los Angeles area experienced the largest magnitude earthquake since the 1994 M_w 6.7 Northridge event. The preliminary analysis of the event [*Hauksson et al.*, 2008] indicated oblique faulting (thrust and strike-slip) shown by the focal mechanism in Figure 3.1. Based on their analysis, the epicenter was found to be located between two known structures - the Whitter and Chino Hills faults, and was the fifth main shock-aftershock sequence to occur in this region since 1987. Despite its proximity to the Los Angeles basin there were no reported fatalities or significant damage from this earthquake [*Hauksson et al.*, 2008]. With much fortune to seismologists, this event was also recorded by over 450 stations in the greater Los Angeles area providing an excellent opportunity to validate deterministic ground motion simulations and community velocity models.

A subsequent inversion study done by [*Shao et al.*, 2012] finds that a northeast dipping plane (strike = 289° and dip = 62°) provides the best fits with the recorded data. The inversion results indicate a hypocentral depth of 14.6 km near the base of the seismogenic zone. In addition, they find that approximately 50% of the moment release occurs from a 1.8 km^2 asperity with a maximum stress drop of 80 MPa and a maximum slip of 1.8 m. The average slip is approximately 0.5 m with an average stress drop between 19 to 38 MPa, much larger than the average stress drop (~ 3 MPa) observed for southern California earthquakes [*Shearer et al.*, 2006].

Numerical Method for Simulating Ground Motions

To simulate 0-2.5 Hz ground motions from the M_w 5.4 Chino Hills event we solve the Navier-Cauchy equations for elastodynamics using the 4th order space 2nd order time staggered-grid finite difference code AWP-ODC-CPU [Cui *et al.*, 2010; 2013] including frequency-dependent attenuation [Withers *et al.*, 2015]. We mitigate reflected energy propagating into the simulation domain using damping boundary conditions described by Cerjan *et al.*, [1985]. Even though we do not use the preferable perfectly-matched layers [Marcinkovich and Olsen, 2003], we find that using a sufficient number of damping elements (30-50 nodes) prevents any significant spurious energy from contaminating our numerical solution. For an in-depth explanation of the numerical method itself and its implementation for high-performance computers, we refer the reader to Cui *et al.*, [2010; 2013]. In the following sections, we explain the components necessary to simulate deterministic ground motions; namely, the finite-fault source model, the heterogeneous velocity model, and the parameterization of anelastic attenuation considered in this study.

Velocity Model with Small-scale Velocity Fluctuations

We discretize a 56 km x 40 km x 24 km region with a regular spacing of 16 m encompassing the majority of Los Angeles basin. Due to the number of simulations required for ensembles we use a smaller region than TB13 and TB14. We show the surface projection of our simulation domain in Figure 3.1 along with the epicenter and focal mechanism for the 2008 M_w 5.4 Chino Hills earthquake.

From the elastic theory of wave propagation, the parameters λ , μ , and ρ are necessary and sufficient to completely describe an isotropic elastic medium. We obtain a model of these elastic parameters from the SCEC CVM-S4.26 [Lee *et al.*, 2011] which is an improvement to CVM-S4 [Magistrale *et al.*, 2000; Kohler *et al.*, 2003] including results from a 3D tomographic study in Southern California. The inversion was performed using a resolution of 500 m and the results are tri-linearly interpolated to smaller resolutions, if necessary. CVM-S4.26 utilizes a software interface that returns (V_p , V_s , and ρ) for a queried

latitude, longitude, and depth. Seismic velocities are converted to elastic parameters using $\mu = V_s^2 \rho$ and $\lambda = \rho(V_p^2 - 2V_s^2)$. The CVM-S4.26 model is developed and supported by SCEC and can be acquired through the SCEC UCVM package [Small et al., 2017]. We used UCVM version 14.3.0 to generate the deterministic velocity model used in this study.

We include small-scale velocity and density fluctuations, known to exist in the crust, by superimposing a statistical model of the fluctuations onto a deterministic background model. Here, CVM-S4.26 provides the deterministic velocity model. We define velocity and density fluctuations with a normally-distributed random variable following a von Karman autocorrelation function (3.1)

$$\Phi_{v,a}(r) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{r}{a}\right)^\nu K_\nu\left(\frac{r}{a}\right), \quad 3.1$$

which has Fourier transform (3.2),

$$P(k) = \frac{\sigma^2 (2\sqrt{\pi}a)^E \Gamma(\nu + E/2)}{\Gamma(\nu)(1 + k^2 a^2)^{\nu+E/2}} \quad 3.2$$

[Klimeš, 2002]. The parameters of the von Karman autocorrelation function are: correlation length, a , Hurst exponent, ν , and standard deviation, σ . Also, $\Gamma(\nu)$ is the gamma function, K_ν is the modified Bessel function of the second kind with order ν , and E represents the Euclidean dimension. We generate the velocity fluctuations using a spectral approach described in detail by Pardo-Igúzquiza and Chica-Olmo, [1993] and Klimeš, [2002]. Using this approach, we simulate the fluctuation profile in the wavenumber domain using the spectral amplitudes defined in (3.2) while leaving the phases random. This approach unconditionally simulates a zero-mean stochastic process with the autocorrelation function defined by (3.1).

We adopt parameters representative of the Los Angeles area from a study by Savran and Olsen, [2016] that investigated a suite of 39 borehole sonic logs recorded in Los Angeles basin. The authors found that $\nu = 0.05$, $a_z = 50 - 150$ m, and $\sigma = 5.0\%$ best characterize the fluctuation profiles observed in the borehole logs. We also incorporate elliptical anisotropy in the fluctuations with a ratio of horizontal-vertical correlation lengths

$(a_x/a_z) = 5$. These parameters are comparable with an independent study of the same borehole data (*Pers. Comm., Andreas Plesch*, [2012]), as well as a spectral analysis of a 3D velocity model computed from ambient noise data [*Nakata et al.*, 2015] generated from a high-resolution seismic dataset deployed in the Long Beach area [*Nakata and Beroza*, 2015]. The study area of *Nakata and Beroza*, [2015] spans a different region of Los Angeles basin than the data locations of *Savran and Olsen*, [2016], but provided similar parameters of the von Karman autocorrelation function, suggesting that these parameters constitute a representative model for Los Angeles basin.

Figure 3.2a shows a horizontal slice of CVM-S4.26 superimposed with small-scale velocity fluctuations at a depth of 160 m (10 grid points). We show a cross-sectional view of the velocity model (Figure 3.2b) illustrating the anisotropy of the velocity fluctuations, and a vertical profile extracted along this transect (Figure 3.3b) in the deeper part of Los Angeles basin. We truncate V_p to 7600 m/s and update V_s preserving the V_p/V_s ratio to eliminate unphysically large velocities in the upper crust and due to computational limitations. TB14 investigated the effects of different velocity models including models for the shallow sediments [*Shaw et al.*, 2015], and they found that ground motions display significant sensitivity to the velocity model considered. In this article, we focus on the effects of small-scale velocity fluctuations as opposed to comparing the effectiveness of an individual velocity model. As such, our analysis only includes simulations from a single deterministic velocity model. However, we suspect that our results showing the effects of velocity fluctuations are generalizable to other leading 3D velocity models for the greater Los Angeles region.

Finite Fault Source Model

Shao et al., [2012] performed a finite-fault inversion using 26 strong motion stations located at distances less than 50 km from the Chino Hills fault to eliminate contamination from Moho reflections. Before inverting for the finite-fault model, their data were band-pass filtered from 0.16 to 2.5 Hz, and P and SH waveforms were chosen based on having desirable signal-to-noise ratios. Their inversion employed a simulated annealing

approach that estimates the slip, rake, rupture-time, and parameters of the asymmetric cosine slip-rate function [Ji *et al.*, 2003]. Figure 3.3 shows the inverted slip-rate functions and Fourier amplitude spectra for all 140 subfaults.

The top of the fault starts at a depth of 13.2 km and has dimensions 5.6 km along strike and 4.0 km along dip. The finite fault is comprised of 140 subfaults, each with an area of 0.16 km² (400m x 400m). Each subfault has the same strike = 289° and dip=62° with a mean rake = 133.5°. The rake varies between 115° and 160° across the fault plane. Due to the lack of data resolution above 2.5 Hz, we low-pass filter our slip-rate functions to a maximum frequency (f_{max}) of 2.5 Hz before scaling the slip-rate functions to the desired $M_0 = 1.6 \times 10^{17}$ Nm for our simulations. We use the inversion results corresponding to “Model 1” from [Shao *et al.*, 2012].

Instead of using $f_{max} = 4.0$ Hz following TB13 and TB14, we choose to consider the band-limited region up to $f_{max} = 2.5$ Hz for two main reasons: 1) the data used in the inversion were bandpass filtered between (0.15 and 2.5 Hz) and 2) TB13 and TB14 found significant misfits between their simulations and recorded data in addition to uncertainty about the velocity model that we attribute to the lack of data resolution above $f_{max} = 2.5$ Hz. For a detailed analysis of the inversion method used to obtain the source model we refer the reader to Shao *et al.*, [2012].

An accurate estimate of the size of the earthquake is critical for assessing the effects of features of the crustal model, including anelastic attenuation and scattering. Shao *et al.* (2012) found a moment of 1.60e17Nm, corresponding to Mw5.4, which is used for the reference in our study. However, other source inversion studies have found higher moments, such as USGS moment tensor (TMTS) and Centroid moment tensor (M_{wc}) estimates of 1.83×10^{17} Nm and 1.96×10^{17} Nm, respectively, corresponding to $M_w = 5.5$. We will explore the effects of a possibly larger moment on the fit between data and simulations.

Frequency-Dependent Attenuation

Anelastic attenuation is necessary to accurately model the amplitudes of seismic waves that propagate for distances much larger than their dominant wavelength, as is the case with regional-scale deterministic ground motion simulations. Typically, numerical studies (e.g., *Olsen et al.*, [2003]; *Graves and Pitarka*, [2010]; *Taborda and Bielak*, [2013]; [2014]) incorporate a frequency-independent attenuation model for the quality factor, Q , meaning that intrinsic energy losses per cycle are equivalent across all frequencies. This model is sufficient for wave propagation with $f_{max} < 1.0$ Hz, but tends to disagree with seismic observations at higher frequencies (e.g., *Aki*, [1980]; *Raoof et al.*, [1999]; *Erickson et al.*, [2004]; *Phillips et al.*, [2013]; *Wang and Shearer*, [2017]), which show that a frequency-dependent attenuation model produces better fits to the recorded data. Typically, empirical relationships for attenuation are based on the local shear-wave velocity due to our understanding that regions with hard rock tends to attenuate the wave-field less than areas with broken and fractured rock or soft-sediments [*Hauksson and Shearer*, 2006].

Withers et al., [2015] developed a memory-efficient technique to incorporate frequency-dependent attenuation using a coarse-grained memory variable approach following *Day*, [1998]; *Day and Bradley*, [2001]. In the frequency-dependent case the weights of the viscoelastic relaxation times are modified so that the target Q spectrum follows a power-law (2.3) above a reference frequency, f_0 , with power-law exponent γ .

$$Q(f) = \begin{cases} Q_0 \left(\frac{f}{f_0}\right)^\gamma, & f \geq f_0, \\ Q_0, & f < f_0 \end{cases} \quad 3.3$$

In this study, we focus on empirical models for the reference Q_0 , due largely to our simulation $f_{max} = 2.5$ Hz, as $Q(f)$ has limited effects at 2.5 Hz. However, we use $f_0 = 1.0$ Hz and $\gamma = 0.6$. Because shear waves attenuate faster than compressional waves, most empirical models define separate values Q_p and Q_s for the P- and S-wave quality factors, respectively. Additionally, most models incorporate Q_s relationships of the form $Q_s = a_i (V_s)^{b_i}$, where a_i and b_i are scalar coefficients, and V_s (m/s) is the shear-wave velocity. In

this study, we choose $a_i = (50, 100, 150)$ and $b_i = 1$. Following *Brocher*, [2008], the simulations performed by TB13 and TB14 incorporate higher order terms in their empirical relationship. In this study, we attempt to find the best linear model for the empirical relationship defining Q_s . We assume that $Q_p = 2Q_s$; while this relation is slightly different than the Q_p attenuation model chosen by TB13 and TB14, *Olsen et al.*, [2003] finds that ground motions are relatively insensitive to Q_p .

Figure 3.4a compares several Q models proposed in the literature [*Olsen et al.*, 2003; *Brocher*, 2008; *Graves and Pitarka*, 2010; *Taborda and Bielak*, 2013; 2014] and two other models used in our simulations, namely, $Q_s = 100V_s$ and $Q_s = 150V_s$. We note that the *Graves and Pitarka*, [2010] model is equivalent to the $Q_s = 50V_s$ model also investigated here. The inset of Figure 3.4a shows the attenuation models for $V_s < 0.8$ km/s. Figure 3.4b shows the effects of the frequency-dependence of Q_s for the model where $Q_0 = 100V_s$ and $\gamma = 0.6$. While the different Q models vary significantly, an important point to mention is that ground motions are influenced by the value of Q itself and not specifically by the coefficients of the empirical relationship used to define the intrinsic attenuation model. In other words, we suspect that the choice of $\min(V_s)$ can explain the different attenuation models proposed by various simulation studies for the Southern California area: as simulations with larger $\min(V_s)$ (e.g., *Olsen et al.*, [2003; *Graves and Pitarka*, [2010]) tend to choose lower values of a_i , but ultimately result in approximately the same $\min(Q_s) \approx 20$.

3.2 Ground Motion Simulations

We simulate 100 seconds of wave propagation with $f_{max} = 2.5$ Hz, to ensure that the wavefield propagates throughout the entire simulation with the chosen $\min(V_s) = 200$ m/s. We discretize the simulation domain with $dx = 16$ m to resolve the smallest wavelength ($\lambda_{min} = \min(V_s)/f_{max}$) with 5 grid points satisfying both the stability and spatial resolution requirements of a 4th order staggered-grid finite-difference solution [*Olsen et al.*, 2003]. The resulting mesh has 13.125 billion elements, and each simulation

conducted for this study requires approximately 6.5 hours of wall-time using 13,125 processors on NCSA Blue Waters. Table 1 shows the simulation parameters used for the models presented in this study.

Validating Ground Motion Simulations

Verification of a numerical method focuses on the accuracy of a particular solution with respect to the conceptual problem underlying the numerical method. For deterministic ground motion simulations, the verification process typically involves comparing the numerical solution against analytical solutions (e.g., *Levander*, [1988]), semi-analytical frequency-wavenumber solutions (e.g., *Zhu and Rivera*, [2002]) or against equivalent numerical methods (e.g., *Day et al.*, [2005]; *Bielak et al.*, [2010]), in the case of problems with no closed-form solution.

Validation is needed to estimate the efficacy in predicting ground motions for future events. The validation of ground motion simulations requires comparing synthesized ground motions against ground motion records to assess the accuracy of the model with respect to the entire physical system under study. Until TB13, TB14 and the SCEC Broadband Platform validation (e.g., *Goulet et al.*, [2015]), most validation exercises were performed using qualitative comparisons between seismograms at select stations or using simple quantitative comparisons (e.g., *Olsen et al.*, [2003]; *Graves and Pitarka*, [2010]; *Aagaard et al.*, [2010a]). However, recent developments in model wide validation techniques using multiple goodness-of-fit metrics [*Anderson*, [2004]; *Olsen and Mayhew*, 2010] or strict time-frequency misfit criteria *Kristekova et al.*, [2009] have encouraged seismologists to focus on rigorous validation and verification of their models.

In the case of the Chino Hills simulations presented here the physical system encompasses: 1) the CVM-S4.26 deterministic velocity model, 2) the stochastic model of velocity and density fluctuations, 3) the empirical model for anelastic attenuation, and 4) the finite-fault model used to describe the Chino Hills event. Due to the complexity of the physical system being modeled a perfect validation of the physical system is likely impossible [*Oreskes et al.*, 1994]. Despite the seemingly impossible task of validating

deterministic ground motion models, the exercise remains informative to highlight strengths and weaknesses of different aspects of the physical system. TB13 and TB14 investigated many of these aspects with respect to the Chino Hills earthquake under study here. As such, we focus our efforts on validating statistical models of velocity and density fluctuations and empirical relationships for intrinsic attenuation rather than the finite-fault source or velocity models.

Goodness of Fit Criteria

To perform our quantitative validation, we use a simplified version of the goodness-of-fit (GOF) criteria proposed by *Olsen and Mayhew*, [2010]. In this section, we summarize the methodology of Olsen and Mayhew, and explain the motivation for the GOF metrics used in this study. As such, we refer the reader to *Olsen and Mayhew*, [2010] for a detailed explanation of their methodology. Olsen and Mayhew incorporate 10 GOF metrics into their method: peak velocity (PGV); peak acceleration (PGA); peak displacement (PGD); response spectral acceleration averaged between 0.1 and 10s (RS); spectral acceleration estimated by NGA relationships for 16 periods (SA16); smoothed Fourier amplitude spectrum (FAS); energy duration (DUR); cumulative energy (ENER); cross-correlation (XCOR); and the inelastic to elastic deformation ratio (IE). In general, the method computes a GOF score between 0 – 100 for each station, where 100 indicates a perfect match, and subsequently aggregates them for a global GOF score.

The Olsen and Mayhew method was designed specifically for use with broadband (0-10+ Hz simulations) to assess multiple aspects of the underlying physical model. Here, we select four of their metrics that we believe to provide the clearest interpretation of our results, namely, PGV, PGA, ENER, DUR, and one additional metric considered by *Anderson*, [2004] and *Taborda and Bielak*, [2013; 2014], the Arias intensity (AI). AI is defined as $\frac{\pi}{2g} \int_0^T a(t)^2 dt$ and describes the progress in time of the acceleration record, where $a(t)$ is the acceleration time record, T is the duration of the signal, and g is the acceleration due to gravity. This metric describes the cumulative energy-per-weight from a uniformly distributed set of single degree-of-freedom oscillators on $[0, \infty]$ [*Arias*, 1970].

Also, we define DUR for each component as $DUR = \int_0^t v(\tau)^2 d\tau$, where t is the time when 75% of the cumulated energy has arrived at the station. We assume the signal starts when 5% of the total energy has arrived at the station. In effect, DUR gives the time required to for a station obtain 5% - 75% of the total energy from an event.

Following Olsen and Mayhew, we compute the GOF score for each metric (G_{met}) using a mapping of the normalized residual to the complementary error function ($erfc$) using (3.4)

$$G_{met} = 100 \operatorname{erfc}\left(\frac{2|x - y|}{x + y}\right), \quad 3.4$$

where x and y are two positive scalar metrics, and the result of the mapping is scaled by 100 resulting in GOF scores that lie between 0 – 100. Next, we compute a single GOF score for the time-series (G_{ts}) by combining all metrics using an equal-weighted average (3.5)

$$G_{ts} = \frac{1}{N} \sum_N G_{met}. \quad 3.5$$

We then compute a GOF score for a particular station (G_{sta}) by averaging G_{ts} assuming equal weights for each component of the three-component ground motion record using $G_{sta} = \frac{1}{3} \sum_3 G_{ts}$. Finally, we assign a GOF score to the entire simulation (G_{sim}) by averaging the GOF scores across all stations. We favor the *erfc* implementation over the exponential function used by *Anderson*, [2004] due to its ability to provide discrimination at high GOF scores. In practice, we find that the GOF approach presented here is similar to that used by previous studies.

Broadband Data Processing

Due to the proximity of the epicenter to the Los Angeles area, more than 450 strong-motion stations from different networks recorded the event, providing some of the best data coverage of any Southern California earthquake. From the 450 stations recording the event, we include 118 stations that lie within our domain for validating our simulations. In addition to excluding stations outside of our computational domain, we also remove 9 stations before comparison due to having incomplete records, visibly erroneous data, or being located within the sponge zones. Figure 3.5 shows the locations of the 109 stations included in our validation procedure, as well as the locations of 14 stations used to show ground motion recordings.

We obtained processed strong-motion recordings from *R. Taborda, Pers. Comm.* [2013], thus our records are processed following the approach of TB13 and TB14, summarized below. First, TB13 performed baseline and gain corrections and integrated the acceleration time-series to velocity and displacement with a high-pass cutoff frequency of 0.1 Hz. All recorded seismograms not aligned with the east-west and north-south directions were rotated into this coordinate system. We note that our computational domain aligns with east-west and north-south (+x indicates east and +y indicates north). Next, the recordings were synchronized temporally using the trigger time reported in the header files along with the rupture time of the event based on the USGS records. We apply a systematic time-shift of 0.9 s to all records corresponding to the average rise-time in our source description. Finally, in contrast with TB13, we bandpass filter the records to 0.15 - 2.5 Hz to allow for comparisons with our synthetics. As with TB13 all filters used in the data processing are Chebyshev Type I filters.

3.3.1 Results of Ground Motion Simulations

First, we perform a sensitivity test of the von Karman parameters (ν , α , σ , H/V) to understand their effects on simulated ground motions and the variability introduced by including different statistical models of velocity fluctuations. Next, we make qualitative

comparisons at select stations to provide some insight into the behavior of small-scale heterogeneities for metrics relevant to seismic hazard, such as the Fourier velocity spectra, response spectra, and cumulated velocity energy. We also show the effects of small-scale heterogeneities on individual metrics aggregated across the simulation domain. In addition to analyzing the effects of the small-scale heterogeneities, we investigate the distance dependence of the five selected metrics (PGA, PGV, DUR, ENER, AI) to showcase the behavior of different attenuation models.

Sensitivity Testing of Small-Scale Heterogeneities

We make five comparisons between different choices of parameters for the von Karman autocorrelation function representing end member values described in the literature. Even though *Nakata and Beroza*, [2015] and *Savran and Olsen*, [2016] obtained agreeable results for the von Karman parameters estimated for Southern California, there are discrepancies in correlation length estimates between different studies that can be explained by inconsistencies during the data processing stage of data analysis (e.g., *Dolan and Bean*, [1997]) or different data types (i.e., seismic, geological, borehole, simulation studies). Correlation length estimates range in orders of magnitude from meters to several kilometers. We find that Hurst exponent estimates typically range from 0.0 – 0.3 (e.g., *Frankel and Clayton*, [1986]; *Levander*, [1992]). Next, we compare radial anisotropy ratios (H/V) of two and five indicating that radial correlation lengths are two or five times larger than vertical correlation lengths, respectively. We compare standard deviations of $\sigma = 5$ and $\sigma = 10\%$ as constrained by the analysis shown in *Nakata and Beroza*, [2015]; *Savran and Olsen*, [2016]. Finally, we vary the spatial locations of the velocity fluctuations while keeping the von Karman parameters identical. Figure 6 shows a horizontal slice of V_s at depth $z = 160$ m for each velocity model considered in the sensitivity study. Table 2 shows the different parameters for the models considered. Qualitative comparisons are made at six stations, STS, LTP, SRN, FUL, PSR, and RUS.

In general, we find that for model C1 (Figure 3.7C1, 3.8C1), which considers the Hurst exponent $\nu = 0.0$ and $\nu = 0.3$, stations exhibit only minor differences between

cumulated energy. Most stations display up to 10% difference between simulations. Some stations, such as SRN and FUL, show no obvious differences in arriving energy or waveforms shapes.

Model C2 (Figure 3.7C2, 3.8C2) considers end-members cases having vertical correlation lengths of 150 – 5000m, with all other parameters unchanged. We find the wavefield displays much more sensitivity to the correlation length than to the Hurst exponent, which is unfortunate considering the lack of agreement of correlation length estimates among studies. Stations such as STS display approximately twice the cumulative energy with a model using larger correlation lengths, and can be explained by the relatively larger size of fluctuations (Figure 3.6d) interacting with a larger portion of the wavefield. The nondimensional parameter ak where a is the correlation length and k is the wave number largely controls the behavior of scattering. In the case where $ak \gg 1$ or in the strong scattering regime, most of the energy is scattered forward, contributing more to the first arrival. In contrast, when $ak \ll 1$, the energy is predominantly scattered in the backward direction. With $a = 5000$ m the modeled wavefield results in $ak \gg 1$ for most of the domain [Przybilla *et al.*, 2009]. The large peak amplitudes obtained at stations LTP and STS likely originate from forward scattering causing geometric focusing effects.

Model C3 (Figure 3.7C3, 3.8C3) tests differences in ground motions between different horizontal-vertical anisotropy ratios. In these tests the vertical correlations lengths are fixed. In general, these comparisons are similar for the tests conducted for differences in correlation length. We find that synthetics typically exhibit larger cumulated velocity with models containing larger horizontal correlation lengths. Synthetics at stations such as PSR exhibit up to approximately 30% differences in cumulated energy with most stations having between 10-20% variability in the ground motions. Most stations exhibit larger synthetic peak amplitudes with larger H/V ratios. However, at some stations such as LTP, synthetics show smaller cumulative velocities.

Model C4 (Figure 3.7C4, 3.8C4) looks at differences in simulated ground motions caused by end-member cases of σ , the strength or magnitude of the velocity fluctuations. Station PSR experiences approximately 50% larger ground motions by increasing σ from 5% to 10%; however, STS experiences 20% lower ground motions. The focusing effects

at PSR are apparent with the large observed first arrival for $\sigma = 10\%$ as compared with $\sigma = 5\%$. Additionally, stations such as FUL and SRN show nearly no difference in synthetic ground motions between the two models.

Finally, model C5 (Figures 3.7C5, 3.8C5) compares the effects of varying the locations of the scatterers. In particular, the locations of the fluctuations are controlled by the distribution of random numbers generated to produce the fluctuations. The distributions of random numbers are defined using a seed number which is subsequently fed into a pseudo-random number generator. In practice, this results in velocity models having the same von Karman parameters, but different along-path scattering effects. In general, we observe the largest variability between synthetic ground motions by altering the seed number. Without knowing the exact locations of the velocity fluctuations, we cannot recommend a single model that produces the best comparisons with data. From these chosen seed numbers, we find that no single model performs systematically better than another. For example, seed = 5 produces the best comparisons at FUL, but the worst comparison at PSR. This indicates that ensembles of simulations need to be considered to capture the systematic effects of scattering.

Qualitative Validation at Select Stations

We begin our validation by showing qualitative plots of ground motion metrics at 14 stations that are evenly distributed around our simulation domain; namely, the waveform envelope (one-sided Hilbert transform), the cumulated energy, Fourier velocity spectrum, and acceleration response spectra (Figures 3.9 and 3.10). In addition to the qualitative plots of different metrics, we also show G_{sta} for both models with and without velocity fluctuations. We selected these stations for several reasons: they are relatively evenly distributed across our simulation domain, and they belong to the same seismic network. In addition, these stations represent a fair distribution of the 109 stations included in our analysis and include some of the best and worst performing stations. These stations also provide a representative sample of the simulation-wide results presented in later sections.

As pointed out by TB13, without any prior knowledge of the recorded Chino Hills ground motions, the simulations could be mistakable for data. This is also the case with the synthetic seismograms shown from the sensitivity test in Figures 3.7 and 3.8. In other words, even though there is disagreement between the data and simulations, the synthetics appear to have originated from a real earthquake. In particular, the progress-in-time metric, the cumulative energy (ENER), provides the most agreeable fits with the observed data. At some stations, (e.g. DLA, LTP, BRE, WTT, FUL) all sets of synthetics are underpredicting the amplitude of the S-wave arrivals causing large discrepancies in the resulting metrics. We attribute this mainly to the velocity model, as the distributions of velocity and density fluctuations appear to only slightly alter the peak arrivals.

The most noteworthy feature of the comparisons is that the ensemble means of models including small-scale heterogeneities are not equal to models that do not consider velocity fluctuations. This is especially apparent in the plots of acceleration response spectrum, kinetic energy, and Fourier velocity. This result indicates that velocity fluctuations affect expected ground motion metrics at a single station and do not simply average out when considering an ensemble of realizations. Like the comparisons made during the sensitivity test, there is no systematic way in which the velocity fluctuations affect these metrics. For example, BRE and DLA have larger and smaller ensemble means than models with no fluctuations.

Seismic scattering disturbs the rupture front and redistributes energy that is recorded later in the seismogram. This mechanism provides an explanation for differences between ensemble means and models not considering small-scale heterogeneities. Even though there is seemingly more energy arriving at the stations, it is important to remember that velocity and density fluctuations do not increase the overall energy present in the wavefield; they simply redistribute energy toward or away from the receivers. Additionally, scattering seems to play an appreciable role above 0.5 Hz at most stations. Looking at the Fourier velocity spectrum, the difference between the ensemble mean and the model with no fluctuations diverges above ~ 0.5 Hz, where they are equal at frequencies < 0.5 Hz. This result is consistent with *Hartzell et al.*, [2010] that shows significant variations in PGV at $f_{max} \approx 0.5$ Hz.

We find particularly good comparisons at station PSR, which is relatively close to the source location and situated over a shallow basin ($Z_{1.0} = 150$ m). In contrast, we find that stations in the deep basin (e.g. LTP, WTT) have poor GOF scores indicating there are discrepancies between the underlying velocity model and the real earth. This could also be related to local site-effects that are not correctly captured by the CVM-S4.26 used for this study. TB14 found that including geotechnical models cannot explain all of the observed differences between the data and synthetics. In addition, most stations with poor GOF scores (e.g. WLT, RUS, SRN) seem to have much larger durations than considered by our simulations. This could be due to strong scattering in the upper crust as noted by *Wang and Shearer*, [2017] or topographic scattering effects that are not considered in our models.

Based on the qualitative comparisons made in Figures 3.9 and 3.10, we suspect that the first-order discrepancies between the synthetics and the data arise from inadequacies in the CVM at scale-lengths typically considered to be somewhat constrained observationally, the attenuation model, or the source model, with the models of velocity fluctuations holding a second-order role.

Effects of Including Statistical Descriptions of Small-scale Heterogeneities

In this section, we investigate area-wide effects of small-scale heterogeneities with a focus on differences in the amplitudes of ground motion metrics between ensembles of models with and without heterogeneities. We do not consider GOF between simulations and data in this portion of the analysis. The simulations use the model for anelastic attenuation where $Q_{s0}/V_s = 100$. We compute ensemble averages over five realizations of ground motions simulated using random fields of velocity and density fluctuations generated using seeds 1 – 5.

First, we show how each individual metric (PGV, PGA, ENER, AI, and DUR) is affected by stochastic models of velocity and density fluctuations. Figure 3.11 shows interpolated PGV from both the data and synthetics at each station in the simulation domain, treating the simulations and data identically. The amplitudes values for PGV and PGA (Figure 3.12) are changed by less than 1%, on average, due to ensembles with velocity

fluctuations. However, at individual stations up to a $\pm 25\%$ difference in metric amplitudes can occur (Figures 3.11d, 3.12d) when compared with models not considering small-scale heterogeneities. The magnitude of this effect increases as a function of distance from the source, which indicates that small-scale heterogeneities will play an increasingly important role at larger propagation distances. In the case of the largest de/amplifications, PGV and PGA are systematically lower for all realizations within the ensemble, indicating a significant effect due to the velocity fluctuations.

It is apparent that the simulations cannot reproduce basin-amplification effects or local site effects resulting in the larger PGV or PGA observed in the data over the deeper portion of Los Angeles basin at larger distances from the source. We find that the discrepancies are larger than the effects of including velocity fluctuations, so the misfit between simulations and data are likely related to inadequacies with the underlying velocity model or source description. There are no apparent relative differences in rock or basin sites or any correlation with the underlying basin depth for PGV and PGA.

The next three comparisons (ENER, AI, and DUR) can be considered ‘progress-in-time’ metrics or metrics that involve the entire waveform where, as expected, the effects of considering velocity and density fluctuations become much more apparent. ENER (Figure 3.13) increases 7% when considering velocity and density fluctuations in addition to the deterministic velocity model, as shown by the median value of the histogram plotted in Figure 3.13d. Also, the majority of stations in the simulation domain experience increased ENER, in contrast with both peak-metrics. AI (Figure 3.14) displays the largest effect from velocity and density fluctuations with a 13% increase model-wide (Figure 3.14d). Like ENER, the increase is observed at the majority of stations in the domain. Finally, DUR (Figure 3.15) shows that stations experience 4% longer shaking when considering models of velocity fluctuations.

In contrast with the peak-metrics (PGV and PGA), models including velocity and density fluctuations affect progress-in-time metrics simulation wide. Additionally, these metrics do not appear to possess the same distance dependence as PGA and PGV. If all metrics were unchanged when considering ensembles of models including velocity fluctuations, it would be an acceptable decision to not include them in similar simulations

considered here. However, progress-in-time metrics show a significant change in values predicted simulation wide. Thus, these metrics represent an important consideration, especially when looking to determine trade-offs with other simulation parameters, such as anelastic attenuation.

An interesting feature in all of these comparisons is that ensemble averages of models considering velocity fluctuations do not significantly affect the spatial distribution of metrics across the simulation domain. Plots comparing CVM-S4.26 with CVM-S4.26 + velocity fluctuations (Figures 3.11-3.15b,c) look virtually identical only with slightly different amplitudes at the affected stations. This indicates that discrepancies in peak-metrics between simulations and data cannot be remedied by simply including models of velocity fluctuations. While heterogeneities do not largely influence averages for peak-ground metrics at $f_{max} = 2.5 \text{ Hz}$, they show systematic effects for progress-in-time metrics.

We find the ground motion metrics appear correlated amongst one another in both the simulations and the data (Figures 3.11-3.15a,b,c) with exception of the DUR metric showing anti-correlations with PGV, PGA, ENER, and AI. Stations experiencing larger than average PGV also experience larger than average PGA, ENER, and AI, but smaller DUR. This agrees with recent work by *Bradley*, [2015] investigating correlations in similar metrics from earthquake records present in the NGA-West2 database [*Ancheta et al.*, 2014]. In addition to predicting realistic amplitudes for ground motions, understanding the intra-event variability and spatial coherence of ground motions is an import feature to capture in realistic ground motion simulations. In this paper, we do not address the spatial correlations of ground motions or correlations amongst metrics (e.g., *Loth and Baker*, 2013; *Bradley*, [2015]); however, we suspect that velocity fluctuations will play an increasingly important role in the analysis of ground motion variability as small-scale heterogeneities effectively reduce the intra-event standard deviations in ground motions through a spatial redistribution of energy [*Withers*, 2016].

Figure 3.16a-e shows the selected metrics plotted as a function of distance (r_{jb}) for CVM-S4.26 + velocity fluctuations, CVM-S4.26, and recorded data. We observe good fits at close stations with the amplitude misfit increasing as a function of distance. The results

indicate that the attenuation model $Q_{s0}/V_s = 100$ seems provide too much damping. The overly attenuating model is apparent by the difference in slopes between the data (black line) and the simulations (red and blue lines). However, this difference could potentially be explained by an uncertainty in moment estimation, which we considered in the Discussion section. We note that if the moment was underestimated, the resulting increase in moment would cause a translation of the distance relationships plotted in Figure 16, and would. TB13 also observed a similar trend in their comparisons with PGV as a function of distance when considering their non-linear Q_s/V_s relationship (Figure 3.6). The underprediction around 30 – 40 km from the source originates above the deeper portion of Los Angeles basin, are likely due to uncertainty in the CVM.

The distance dependence of the metrics (PGV, PGA, ENER, and AI) all have nearly identical forms implying a strong correlation between individual metrics. However, DUR is anti-correlated with the other four metrics considered. This agrees with recent work by *Bradley*, [2015] investigating the correlations in similar metrics from earthquake records present in the NGA database. In general, DUR increases with increasing distance from the source, which can be explained primarily by wave propagation effects along the path, such as basin-amplification and local site-effects. Despite the simulations under predicting peak-motions in the deeper basin sites, DUR is represented well by our simulations (and thus by the underlying velocity model as shown by Figure 3.16e).

3.3 Discussion of Results

Our simulations, as well as TB13 and TB14, show promising results toward accurately modeling deterministic wave propagation to higher-frequencies. However, we find some obvious sources of misfit between the simulations and the data. First, we find that our simulations improperly address the distance decay of metrics observed in the data. Secondly, we find that seismograms extracted at stations located over deeper portions of the basin (e.g. station WTT) drastically under predict the peak-value metrics (PGA, PGV) leading to poor GOF scores at those stations. TB13 pointed out regions in the velocity model that likely contribute to these misfits, so instead we investigate model-wide

discrepancies not related to the underlying velocity model; namely, the attenuation model in addition to possible uncertainty in the moment estimates.

Effects of Moment Estimations of Source Model

According to the several moment tensor inversions provided by the USGS, the magnitude of the 2008 Chino Hills Earthquake was reported between M_w 5.3 and M_w 5.5, with their chosen inversion having a magnitude of M_w 5.4. This estimate is consistent with the publications by *Hauksson et al.*, [2008] and *Shao et al.*, [2012]. Figure 3.17 shows the comparisons between our simulations and data for the attenuation model $Q_{s0}/V_s=100$ when considering M_w 5.5 ($M_0 = 2.2 \times 10^{17}$ Nm). We find that increasing the moment provides better comparisons for all metrics considered across the simulated bandwidth of 0.15 – 2.5 Hz. In particular, for the metrics including coda arrivals (e.g., Figure 3.17c,d), we find that both the distance decay and the general amplitudes show good agreement with the data. Even with an increased M_w , we find the same discrepancies observed with peak-metrics (Figure 3.17a,b) seen with the M_w 5.4 source (Figure 3.16a,b), namely the under-prediction of peak amplitudes at certain stations with large distances from the fault. As with the M_w 5.4 simulation, we suspect that these discrepancies are related either basin effects or local site effects that are not properly accounted for in CVM-S4.26. Based on these comparisons, we believe that the moment of the 2008 Chino Hills earthquake is greater than M_w 5.4 proposed by *Shao et al.*, [2012], closer to M_w 5.5.

Anelastic Attenuation

When determining the appropriate attenuation model, we consider results from both M_w 5.4 and M_w 5.5 source magnitudes to assess the trade-off between the chosen anelastic attenuation model and uncertainty in the source estimation.

First, we show the simulation results considering M_w 5.4 and $Q_s/V_s = 50$ (Figure 3.18) to compare our results against previously published relationships between Q_{s0} and V_s (e.g., *Graves and Pitarka*, [2010]). The simulation results show a much steeper slope in

predicted amplitudes as a function of distance for $Q_{s0}/V_s=50$ as compared with $Q_{s0}/V_s=100$ and data. This result indicates that $Q_{s0}/V_s=50$ is likely too attenuating for southern California when considering simulations with $\min(V_s) = 200$ m/s. We verify that this result is not biased by assuming M_w 5.4 by recomputing the metrics from a simulation with M_w 5.5 and $Q_{s0}/V_s=50$ (Figure 3.19). We only show metrics involving the velocity time-series; namely, PGV, ENER, and DUR as they provide sufficient insight into the performance of the attenuation model. We find that considering M_w 5.5 cannot explain the misfit between the simulations and data for $Q_{s0}/V_s=50$, especially at $f > 1.0$ Hz, thus strengthening our argument that $Q_{s0}/V_s=50$ likely provides too much attenuation for southern California.

Figure 3.20 shows the simulation results considering M_w 5.4 and $Q_{s0}/V_s = 150$. According to the global GOF score, this model produces the best representation of data without requiring any scaling of the source moment. Even so, we still observe regions in the deeper portion of Los Angeles basin that provide poor GOF scores with the data. Because we obtained similar GOF scores for the M_w 5.5 simulation using $Q_{s0}/V_s = 100$, we compare 0.15 – 1.0 Hz metrics (PGV, ENER, and DUR) between these two simulation to assess trade-offs between the attenuation model and uncertainty in moment estimations. The two sets of simulations perform very similarly when comparing against data. Visually, the model using M_w 5.5 and $Q_{s0}/V_s = 100$ appears to provide better fits with data, especially at farther distances from the source. This is particularly noticeable for PGV (Figure 3.21a,b) and ENER (Figure 3.21c,d). On the other hand, DUR (Figure 3.21e,f) seems to be better represented with the M_w 5.4 model and $Q_{s0}/V_s = 150$. The M_w 5.5 model shows larger misfit at stations close to the source, as compared with the M_w 5.4 model. In addition to the M_w 5.5 simulation using $Q_{s0}/V_s = 100$, we also simulated a M_w 5.5 model with $Q_{s0}/V_s = 150$ (not shown here). This comparison shows systematic over-predictions of the data, especially at $f_{max} < 1.0$ Hz. Figure 22 shows the bias based on the results presented in Figure 21. Aside from PGV (Figure 3.22a,b), we do not observe any significant trend in the bias, indicating that both models reproduce the general distance decay observed in the data equally well. Additionally, the plots depicting the two models (Figure 3.21): $Q_{s0}/V_s=100$ with M_w 5.5 and $Q_{s0}/V_s=150$ with M_w 5.4 are visually indistinguishable with the $Q_{s0}/V_s=100$ model producing better fits with ENER (Figure 3.22c,d), but worse fits

with PGV (Figure 3.22a,b). This plot illustrates an example of the interdependence between simulation parameters, and the challenges present when attempting to obtain best fitting model parameters from a single earthquake scenario.

Figure 3.23 shows area-wide GOF scores for the individual metrics (PGV, CKE and DUR). Here, we chose to ignore the acceleration-based metrics (PGA, AI) because we are dealing with data with $f_{max} = 1.0 \text{ Hz}$. Figure 3.23 shows that metrics PGV (Figure 23a) and ENER (Figure 3.23b) with $Q_{s0}/V_s=150$ and the M_w 5.4 source provides better GOF at closer stations, and worse fits at larger distance. Whereas the $Q_{s0}/V_s=100$ model provides better fits with data in the deeper basin, but worse fits at close distances to the source. However, DUR seems to systematically be better represented by the simulation using $Q_{s0}/V_s = 150$. In addition, both simulations produce G_{sim} scores of 54 (for $f_{max} = 1.0 \text{ Hz}$) when considering metrics PGV, CKE, and DUR across the entire domain.

Assessing the best-fitting relationship for Q_{s0}/V_s provides significant challenges when considering the uncertainty present in the simulation input parameters. As shown from the results in this section, it is difficult to attribute the misfit observed between synthetics and data to uncertainties in the moment of the source model or the chosen anelastic attenuation model. However, based on these results we can conclude that when considering simulations with $\min(V_s) = 200 \text{ m/s}$ and $f_{max} = 2.5 \text{ Hz}$ the previously recommended relationship of $Q_{s0}/V_s = 50$ provides too much viscoelastic damping. These results also indicate that a systematic study following an analogous approach to the SCEC Broadband platform validation exercise (e.g. *Goulet et al.*, [2015]) should be conducted using available earthquake data and GMPE calculations, before a confident assessment can be made about the correct anelastic attenuation model.

Effects of Velocity Heterogeneities on Goodness of Fit

Figure 3.24 shows GOF scores for individual metrics and computed percent differences between the GOF scores computed for CVM-S4.26 and CVM-S4.26 + velocity fluctuations, using M_w 5.5 with attenuation model $Q_{s0}/V_s=100$, as this model provides near optimal GOF values. The median of model ensembles of velocity and density fluctuations

can affect the GOF scores up to 50% for some metrics (ENER, AI, and DUR) with others such as PGV and PGA much less influenced. Velocity fluctuations likely play an increasingly important role in modeling high-frequency ground motions to larger distances than considered here. Recent work by *Wang and Shearer*, [2017] shows the interdependence of anelastic attenuation and scattering attenuation, which implies that scattering should be included for high-frequency simulations when determining the parameters for frequency-dependent attenuation; namely the power-law exponent γ . Future work should focus on properly addressing the trade-off between intrinsic attenuation and scattering attenuation in high-frequency simulations including $Q(f)$ and velocity fluctuations.

3.4 Conclusions

We have simulated 0-2.5 Hz ground motions for the 2008 M_w 5.4 Chino Hills earthquake using frequency-dependent attenuation and ensembles ground motions computed in realizations of stochastic models of small-scale crustal velocity and density fluctuations described by a von Karman autocorrelation function. We find that the most influential parameter controlling the ground motions is the seed number of the random distribution used to define the velocity fluctuations. In effect, the seed number changes the spatial locations of the fluctuations causing significantly different path and site effects. We find that five simulation ensembles start to divert from simulations that do not consider small-scale heterogeneities, at frequencies as low as 0.3 Hz at certain stations. Some stations display both amplifications and deamplifications with respect to the unperturbed model within the same ensemble. The most significant effects occur at frequencies above 1 Hz.

The ensemble median values for ground motion metrics (PGV, PGA, ENER, AI, and DUR) show differences of up to $\pm 50\%$ compared to results without small-scale heterogeneities at various locations in the model. In contrast, the progress-in-time metrics display systematic effects from the velocity and density fluctuations. These metrics show significant differences in that nearly all stations experience increased amplitudes. The

largest effect is observed with AI showing a difference of 13%, on average, between models with and without velocity and density fluctuations.

Our results suggest that the first-order sources of misfit associated with the Chino Hills simulation are related to the underlying velocity model the source model, and cannot be remedied only using ensemble median values of small-scale heterogeneities for the models involved here. However, we suspect that small-scale heterogeneities will play a much more important role when investigating the intraevent variability, spatial correlation structure of observed ground motions, and when considering higher-frequency waves propagated to regional-scale distances.

We find that our simulations drastically under predict ground motions with an increase in misfit as a function of distance using the chosen attenuation model for other Southern California studies ($Q_{s0}/V_s=50$). Our preferred model is obtained using a moment magnitude of M_w 5.5 and $Q_{s0}/V_s=100$ using the ensemble of small-scale heterogeneities. This model trades off with smaller magnitudes and less anelastic attenuation (e.g., M_w 5.4 and $Q_{s0}/V_s=150$ produces almost equally good GOF scores). We note that the proposed Q models may not generalize to wave propagation studies conducted in different regions.

References

- A. R. Levander (1992), A Stochastic View of Lower Crustal Fabric Based on Evidence From the Ivrea Zone, *Geophys. Res. Lett.*, 19(11), 1153–1156.
- Aagaard, B. T., and T. H. Heaton (2004), Near-Source Ground Motions from Simulations of Sustained Intersonic and Supersonic Fault Ruptures, *Bulletin of the Seismological Society of America*, 94(6), 2064–2078.
- Aagaard, B. T., R. W. Graves, A. Rodgers, T. M. Brocher, R. W. Simpson, D. Dreger, N. A. Petersson, S. C. Larsen, S. Ma, and R. C. Jachens (2010a), Ground-Motion Modeling of Hayward Fault Scenario Earthquakes, Part II: Simulation of Long-Period and Broadband Ground Motions, *Bulletin of the Seismological Society of America*, 100(6), 2945–2977, doi:10.1785/0120090379.
- Aagaard, B. T., R. W. Graves, D. P. Schwartz, D. A. Ponce, and R. W. Graymer (2010b), Ground-Motion Modeling of Hayward Fault Scenario Earthquakes, Part I: Construction of the Suite of Scenarios, *Bulletin of the Seismological Society of*

- America*, 100(6), 2927–2944, doi:10.1785/0120090324.
- 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, doi:10.1193/070913EQS198M.
- Aki, K. (1980), Attenuation of shear-waves in the lithosphere for frequencies from 0.05 to 25 Hz, *Physics of the Earth and Planetary Interiors*, 21(1), 50–60, doi:10.1016/0031-9201(80)90019-9.
- Ancheta, T. D. et al. (2014), NGA-West2 Database, *Earthquake Spectra*, 30(3), 989–1005, doi:10.1193/070913eqs197m.
- Anderson, J. G. (2004), Quantitative measure of the goodness-of-fit of synthetic seismograms.
- Arias, A. (1970), Measure of Earthquake Intensity, in *Seismic Design for Nuclear Power Plants*, pp. 438–483, Massachusetts Inst. of Tech., Cambridge. Univ. of Chile, Santiago de Chile.
- Bielak, J. et al. (2010), The ShakeOut earthquake scenario: Verification of three simulation sets, *Geophysical Journal International*, 180(1), 375–404, doi:10.1111/j.1365-246X.2009.04417.x.
- Bielak, J., H. Karaoglu, and R. Taborda (2011), Memory-efficient displacement-based internal friction for wave propagation simulation, *GEOPHYSICS*, 76(6), T131–T145, doi:10.1190/geo2011-0019.1.
- 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 Engineering Research Institute.
- Bradley, B. (2015), Correlation of Arias intensity with amplitude, duration and cumulative intensity measures, *Soil Dynamics and Earthquake Engineering*, 78, 89–98, doi:10.1016/j.soildyn.2015.07.009.
- Brocher, T. M. (2008), Compressional and Shear-Wave Velocity versus Depth Relations for Common Rock Types in Northern California, *Bulletin of the Seismological Society of America*, 98(2), 950–968, doi:10.1785/0120060403.
- Bydlon, S. A., and E. M. Dunham (2015), Rupture dynamics and ground motions from earthquakes in 2-D heterogeneous media, *Geophys. Res. Lett.*, 42(6), 1701–1709, doi:10.1002/2014GL062982.
- 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 Engineering Research Institute.

- Cerjan, C., D. Kosloff, R. Kosloff, and M. Reshef (1985), A nonreflecting boundary condition for discrete acoustic and elastic wave equations, *GEOPHYSICS*, 50(4), 705–708, doi:10.1190/1.1441945.
- Chiou, B. S. J., and R. R. Youngs (2014), 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, doi:10.1193/072813EQS219M.
- Cui, Y. et al. (2010), Scalable Earthquake Simulation on Petascale Supercomputers, pp. 1–20, IEEE.
- Cui, Y. et al. (2013), *Physics-based seismic hazard analysis on petascale heterogeneous supercomputers*, ACM, New York, New York, USA.
- Day, S. M. (1998), Efficient simulation of constant Q using coarse-grained memory variables, *Bulletin of the Seismological Society of America*, 88(4), 1051–1062.
- Day, S. M., and C. R. Bradley (2001), Memory-Efficient Simulation of Anelastic Wave Propagation, *Bulletin of the Seismological Society of America*, 91(3), 520–531, doi:10.1785/0120000103.
- Day, S. M., L. A. Dalguer, and N. Lapusta (2005), Comparison of finite difference and boundary integral solutions to three-dimensional spontaneous rupture, *Journal of Geophysical Research*, 110(B12), B12307, doi:10.1029/2005JB003813.
- Day, S. M., R. W. Graves, J. Bielak, D. Dreger, S. C. Larsen, K. B. Olsen, A. Pitarka, and L. Ramirez-Guzman (2008), Model for Basin Effects on Long-Period Response Spectra in Southern California, <http://dx.doi.org/10.1193/1.2857545>, 24(1), 257–277, doi:10.1193/1.2857545.
- Dolan, S. S., and C. J. Bean (1997), Some remarks on the estimation of fractal scaling parameters from borehole wire-line logs, *Geophys. Res. Lett.*, 24(10), 1271–1274.
- Duan, B., and S. M. Day (2010), Sensitivity Study of Physical Limits on Ground Motion at Yucca Mountain, *Bulletin of the Seismological Society of America*, 100(6), 2996–3019, doi:10.1785/0120090372.
- Erickson, D., D. E. McNamara, and H. M. Benz (2004), Frequency-Dependent Lg Q within the Continental United States, *Bulletin of the Seismological Society of America*, 94(5), 1630–1643, doi:10.1785/012003218.
- Frankel, A., and R. Clayton (1986), Finite difference simulations of seismic scattering: Implications for the propagation of short-period seismic waves in the crust and models of crustal heterogeneity, *Journal of Geophysical Research: Solid Earth*, 91(B6), 6465–6489.
- Goulet, C. A., N. A. Abrahamson, P. G. Somerville, and K. E. Wooddell (2015), The SCEC

- Broadband Platform Validation Exercise: Methodology for Code Validation in the Context of Seismic-Hazard Analyses, *Seismological Research Letters*, 86(1), 17–26, doi:10.1785/0220140104.
- Graves, R. W. et al. (2011), CyberShake: A Physics-Based Seismic Hazard Model for Southern California, *PAGEOPH*, 168(3-4), 367–381, doi:10.1007/s00024-010-0161-6.
- Graves, R. W., and A. Pitarka (2010), Broadband Ground-Motion Simulation Using a Hybrid Approach, *Bulletin of the Seismological Society of America*, 100(5A), 2095–2123, doi:10.1785/0120100057.
- Hartzell, S., S. Harmsen, and A. Frankel (2010), Effects of 3D random correlated velocity perturbations on predicted ground motions, *Bulletin of the Seismological Society of America*, 100(4), 1415–1426, doi:10.1785/0120090060.
- Hauksson, E., and P. M. Shearer (2006), Attenuation models (QP and QS) in three dimensions of the southern California crust: Inferred fluid saturation at seismogenic depths, *J. Geophys. Res.*, 111(B5), B05302–n/a, doi:10.1029/2005JB003947.
- Hauksson, E., K. Felzer, D. Given, M. Giveon, S. Hough, K. Hutton, H. Kanamori, V. Sevilgen, S. Wei, and A. Yong (2008), Preliminary Report on the 29 July 2008 Mw 5.4 Chino Hills, Eastern Los Angeles Basin, California, Earthquake Sequence, *Seismological Research Letters*, 79(6), 855–866, doi:10.1785/gssrl.79.6.855.
- Imperator, W., and P. M. Mai (2012), Broad-band near-field ground motion simulations in 3-dimensional scattering media, *Geophysical Journal International*, 192(2), 725–744, doi:10.1093/gji/ggs041.
- Ji, C., D. V. Helmberger, and D. J. Wald (2003), Slip history and dynamic implications of the 1999 Chi-Chi, Taiwan, earthquake, *J. Geophys. Res.*, 108(B9), 2412, doi:10.1029/2002JB001764.
- Kawase, H. (1996), The Cause of the Damage Belt in Kobe: “The Basin-Edge Effect,” Constructive Interference of the Direct S-Wave with the Basin-Induced Diffracted/Rayleigh Waves, *Seismological Research Letters*, 67(5), 25–34, doi:10.1785/gssrl.67.5.25.
- Klimeš, L. (2002), Correlation functions of random media, *PAGEOPH*, 159(7-8), 1811–1831.
- Kohler, M. D., H. Magistrale, and R. Clayton (2003), Mantle Heterogeneities and the SCEC Reference Three-Dimensional Seismic Velocity Model Version 3, *Bulletin of the Seismological Society of America*, 93(2), 757–774, doi:10.1785/0120020017.

- Korn, M. (1993), Determination of Site-Dependent Scattering-Q From P-Wave Coda Analysis with an Energy-Flux Model, *Geophysical Journal International*, 113(1), 54–72.
- Kristekova, M., J. Kristek, and P. Moczo (2009), Time-frequency misfit and goodness-of-fit criteria for quantitative comparison of time signals, *Geophysical Journal International*, 178(2), 813–825, doi:10.1111/j.1365-246X.2009.04177.x.
- Lee, E. J., P. Chen, and T. H. Jordan (2014), Testing Waveform Predictions of 3D Velocity Models against Two Recent Los Angeles Earthquakes, *Seismological Research Letters*, 85(6), 1275–1284, doi:10.1785/0220140093.
- Lee, E.-J., P. Chen, T. H. Jordan, and L. Wang (2011), Rapid full-wave centroid moment tensor (CMT) inversion in a three-dimensional earth structure model for earthquakes in Southern California, *Geophysical Journal International*, 186(1), 311–330, doi:10.1111/j.1365-246X.2011.05031.x.
- Levander, A. R. (1988), 4th-Order Finite-Difference P-Sv Seismograms, *GEOPHYSICS*, 53(11), 1425–1436, doi:10.1190/1.1442422.
- Li, Y.-G., K. Aki, D. Adams, A. Hasemi, and W. H. K. Lee (1994), Seismic Guided-Waves Trapped in the Fault Zone of the Landers, California, Earthquake of 1992, *Journal of Geophysical Research: Solid Earth*, 99(B6), 11705–11722.
- Loth, C., and J. W. Baker (2013), A spatial cross-correlation model of spectral accelerations at multiple periods, *Earthquake Engineering & Structural Dynamics*, 42(3), 397–417, doi:10.1002/eqe.2212.
- Maechling, P. J., F. Silva, S. Callaghan, and T. H. Jordan (2015), SCEC Broadband Platform: System Architecture and Software Implementation, *Seismological Research Letters*, 86(1), 27–38, doi:10.1785/0220140125.
- Magistrale, H., S. M. Day, R. Clayton, and R. W. Graves (2000), The SCEC Southern California Reference Three-Dimensional Seismic Velocity Model Version 2, *Bulletin of the Seismological Society of America*, 90(6B), S65–S76, doi:10.1785/0120000510.
- Marcinkovich, C., and K. B. Olsen (2003), On the implementation of perfectly matched layers in a three-dimensional fourth-order velocity-stress finite difference scheme, *Journal of Geophysical Research*, 108(B5), 2276, doi:10.1029/2002JB002235.
- Nakata, N., and G. C. Beroza (2015), Stochastic characterization of mesoscale seismic velocity heterogeneity in Long Beach, California, *Geophysical Journal International*, 203(3), 2049–2054, doi:10.1093/gji/ggv421.
- Nakata, N., J. P. Chang, J. F. Lawrence, and P. Boué (2015), Body wave extraction and tomography at Long Beach, California, with ambient-noise interferometry, *Journal of*

- Geophysical Research: Solid Earth*, 120(2), 1159–1173, doi:10.1002/2015JB011870.
- Olsen, K. B. (2000), Site Amplification in the Los Angeles Basin from Three-Dimensional Modeling of Ground Motion, *Bulletin of the Seismological Society of America*.
- Olsen, K. B., and J. E. Mayhew (2010), Goodness-of-fit criteria for broadband synthetic seismograms, with application to the 2008 Mw 5.4 Chino Hills, California, earthquake, *Seismological Research Letters*, 81(5), 715–723, doi:10.1785/gssrl.81.5.715.
- Olsen, K. B., S. M. Day, and C. R. Bradley (2003), Estimation of Q for Long-Period (>2 sec) Waves in the Los Angeles Basin, *Bulletin of the Seismological Society of America*, 93(2), 627–638, doi:10.1785/0120020135.
- Oreskes, N., K. Shraderfrechette, and K. Belitz (1994), Verification, Validation, and Confirmation of Numerical-Models in the Earth-Sciences, *Science*, 263(5147), 641–646.
- Pardo-Igúzquiza, E., and M. Chica-Olmo (1993), The Fourier Integral Method: An efficient spectral method for simulation of random fields, *Math Geol*, 25(2), 177–217, doi:10.1007/BF00893272.
- Phillips, W. S., K. M. Mayeda, and L. Malagnini (2013), How to Invert Multi-Band, Regional Phase Amplitudes for 2-D Attenuation and Source Parameters: Tests Using the USArray, *PAGEOPH*, 171(3-5), 469–484, doi:10.1007/s00024-013-0646-1.
- Pitarka, A., L. A. Dalguer, S. M. Day, P. G. Somerville, and K. Dan (2009), Numerical Study of Ground-Motion Differences between Buried-Rupturing and Surface-Rupturing Earthquakes, *Bulletin of the Seismological Society of America*, 99(3), 1521–1537, doi:10.1785/0120080193.
- Przybilla, J., U. Wegler, and M. Korn (2009), Estimation of crustal scattering parameters with elastic radiative transfer theory, *Geophysical Journal International*, 178(2), 1105–1111, doi:10.1111/j.1365-246X.2009.04204.x.
- Raoof, M., R. B. Herrmann, and L. Malagini (1999), Attenuation and Excitation of Three-Component Ground Motion in Southern California, *Bulletin of the Seismological Society of America*, 89, 1–15.
- Roten, D., K. B. Olsen, S. M. Day, Y. Cui, and D. Faeh (2014), Expected seismic shaking in Los Angeles reduced by San Andreas fault zone plasticity, *Geophys. Res. Lett.*, 41(8), 2769–2777, doi:10.1002/2014GL059411.
- Savran, W. H., and K. B. Olsen (2016), Model for small-scale crustal heterogeneity in Los Angeles basin based on inversion of sonic log data, *Geophysical Journal International*, 205(2), 856–863, doi:10.1093/gji/ggw050.
- Shao, G., C. Ji, C. Ji, and E. Hauksson (2012), Rupture process and energy budget of the

- 29 July 2008 Mw 5.4 Chino Hills, California, earthquake - Shao - 2012 - Journal of Geophysical Research: Solid Earth (1978–2012) - Wiley Online Library, *J. Geophys. Res.*
- Shaw, J. H. et al. (2015), Unified Structural Representation of the southern California crust and upper mantle, *Earth and Planetary Science Letters*, 415(C), 1–15, doi:10.1016/j.epsl.2015.01.016.
- Shearer, P. M., G. A. Prieto, and E. Hauksson (2006), Comprehensive analysis of earthquake source spectra in southern California, *Journal of Geophysical Research*, 111(B6), B06303–n/a, doi:10.1029/2005JB003979.
- Small, P., D. Gill, P. J. Maechling, R. Taborda, K. B. Olsen, G. P. Ely, and C. A. Goulet (2017), The SCEC Unified Community Velocity Model Software Framework, *Seismological Research Letters*, 1–14, doi:10.1785/0220170082.
- Spudich, P., and K. B. Olsen (2001), Fault zone amplified waves as a possible seismic hazard along the Calaveras fault in central California, *Geophys. Res. Lett.*, 28(13), 2533–2536.
- Süss, M. P., and J. H. Shaw (2003), P wave seismic velocity structure derived from sonic logs and industry reflection data in the Los Angeles basin, California, *Journal of Geophysical Research*, 108(B3), 2170, doi:10.1029/2001JB001628.
- Taborda, R., and J. Bielak (2013), Ground-Motion Simulation and Validation of the 2008 Chino Hills, California, Earthquake, *Bulletin of the Seismological Society of America*, 103(1), 131–156, doi:10.1785/0120110325.
- Taborda, R., and J. Bielak (2014), Ground-Motion Simulation and Validation of the 2008 Chino Hills, California, Earthquake Using Different Velocity Models, *Bulletin of the Seismological Society of America*, 104(4), 1876–1898, doi:10.1785/0120130266.
- Wang, W., and P. M. Shearer (2017), Using direct and coda wave envelopes to resolve the scattering and intrinsic attenuation structure of Southern California, *Journal of Geophysical Research: Solid Earth*, 100(B12), 24,015, doi:10.1002/2016JB013810.
- Withers, K. B. (2016), Ground motion and variability from 3-D deterministic broadband simulations, 31 August.
- Withers, K. B., K. B. Olsen, and S. M. Day (2015), Memory-Efficient Simulation of Frequency-Dependent Q, *Bulletin of the Seismological Society of America*, 1–14, doi:10.1785/0120150020.
- Wu, R. S., and K. Aki (1985), The Fractal Nature of the Inhomogeneities in the Lithosphere Evidenced From Seismic-Wave Scattering, *PAGEOPH*, 123(6), 805–818.

Zhu, L. P., and L. A. Rivera (2002), A note on the dynamic and static displacements from a point source in multilayered media, *Geophysical Journal International*, 148(3), 619–627.

Table 3.1: Simulation parameters used for the deterministic ground motion simulations of the M_w 5.4 Chino Hills earthquake. We express the points per minimum wavelength assuming $f_{max} = 2.5 \text{ Hz}$ and $\min(V_s) = 200 \text{ m/s}$.

Domain	
Length	56 km (3500 nodes)
Width	40 km (2500 nodes)
Depth	24 km (1500 nodes)
Southwest Corner	33.7500, -118.3000
Spatial Resolution	
Maximum Frequency	2.5 Hz
Minimum V_s	200 m/s
Points per wavelength	5
Grid Discretization	16 m
Number of cells	13.125 billion
Wall-clock time	6.5 hours
Number of Processors	13,125
Temporal Resolution	
Time Discretization	0.001 s
Simulation Time	100.0 s
Number of Timesteps	100,000

Table 3.2: Simulations considered for the sensitivity testing different parameterizations of velocity fluctuations. We consider end-member cases for the Hurst exponent, ν , vertical correlation length, a_z , ratio of horizontal to vertical correlation lengths, H/V , standard deviation expressed as the coefficient of variation, σ , and the spatial locations of fluctuations determined by the seed number. The values inside the parentheses next to the model name indicate the velocity models corresponding to the subplots in Figure 3.6.

Model	ν	a_z (m)	H/V	σ (%)	seed
C1 (b,c)	0.0, 0.3	150	5	5	1
C2 (a,d)	0.05	150, 5000	5	5	1
C3 (a,f)	0.05	150	2, 5	5	1
C4 (a,e)	0.05	150	5	5, 10	1
C5 (a,g)	0.05	150	5	5	1 - 5

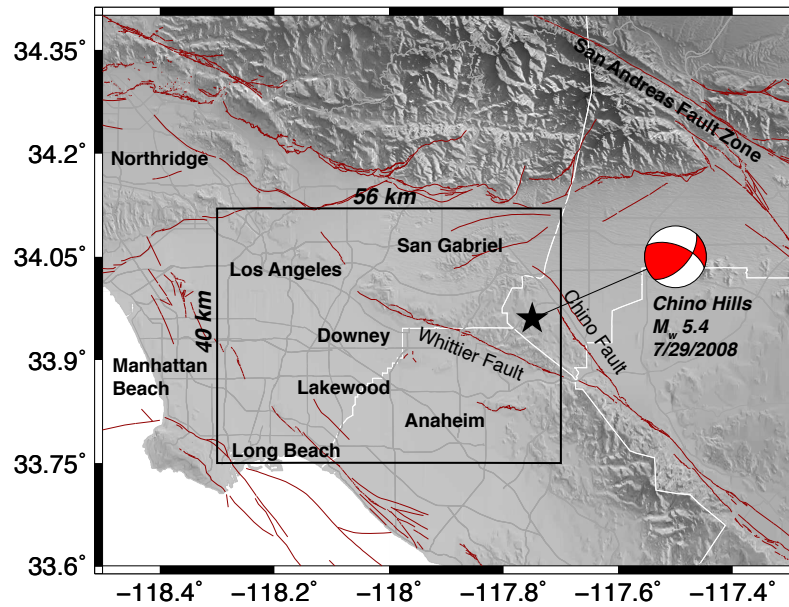


Figure 3.1: Simulation domain used for 2.5 Hz ground motions for the M_w 5.4 Chino Hills earthquake scenario validation study in scattering media. We show the focal mechanism for the oblique failure along with the epicenter (black star) located between the Chino Fault and the Whittier Fault.

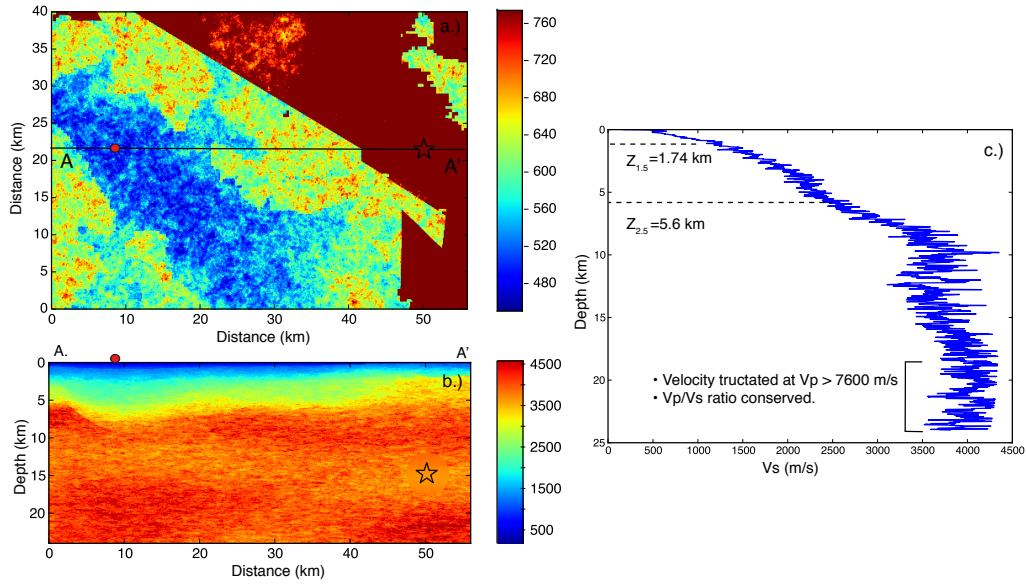


Figure 3.2: Velocity fluctuations from the von Karman correlation function with parameters $\nu = 0.05$, $a = 150$ m, $\sigma = 5\%$, $H/V = 5$ superimposed on SCEC CVM-S4.26. (a) Depth slice at 160m depth, (b) cross-section taken at A-A' showing the deeper area of Los Angeles basin, and (c) profile extracted at the red circle in (a, b). Velocities are listed in (m/s). Note in b) that the perturbations are excluded from the source region following *Frankel and Clayton*, [1986] to preserve the moment released during the earthquake. The star depicts the hypocenter for the Chino Hills event.

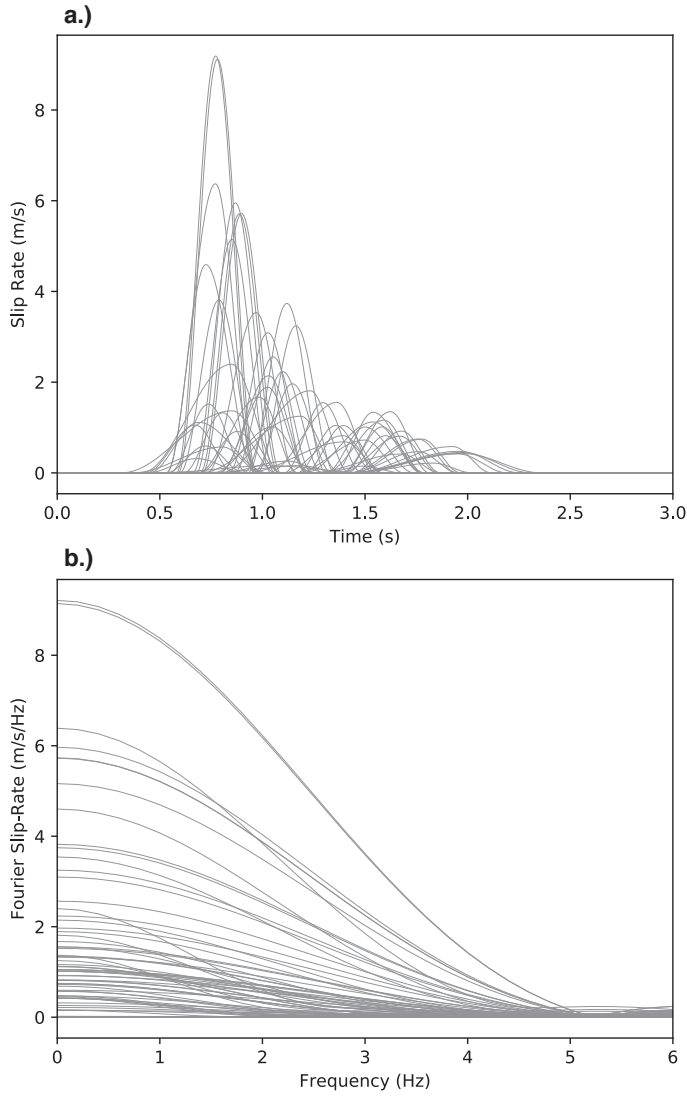


Figure 3.3: (a) Slip-rate functions from Model 1 of the *Shao et al.*, [2012] finite-fault inversion. (b) Fourier amplitude spectrum of the slip-rate functions. Despite showing frequency content above 2.5 Hz for some subfaults, we limit our simulation $f_{max} = 2.5$ Hz corresponding to the f_{max} of the data used for the finite-fault slip inversion.

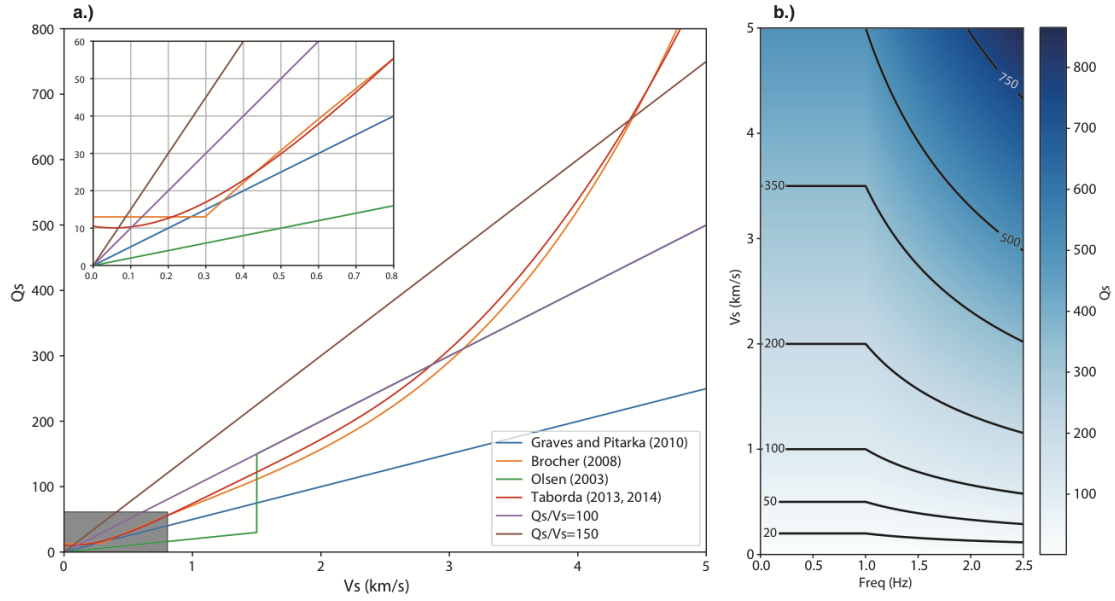


Figure 3.4: (a) Shear-wave quality factor (Q_s) plotted against V_s (km/s) for several attenuation models widely used in the literature [Olsen *et al.*, 2003; Brocher, 2008; Graves and Pitarka, 2010; Taborda and Bielak, 2013; 2014]. The inset figure in the upper left shows the different attenuation models for $V_s < 0.8$ km/s. These plots are valid for the frequency independent portion (Q_0) of our attenuation law (< 1 Hz), after which the Q_s values are modified by the equation $Q(f) = Q_0 f^\gamma$, where $\gamma = 0.6$. This plot is based on Taborda and Bielak, [2013]. (b) Plot showing frequency dependent attenuation for $Q(f) = Q_0 f^\gamma$, where $Q_0 = 100V_s$ and $\gamma = 0.6$. Contour lines show the effects of $Q(f)$, with the most prominent effects happening at large V_s . The contour labels show the value of Q_s .

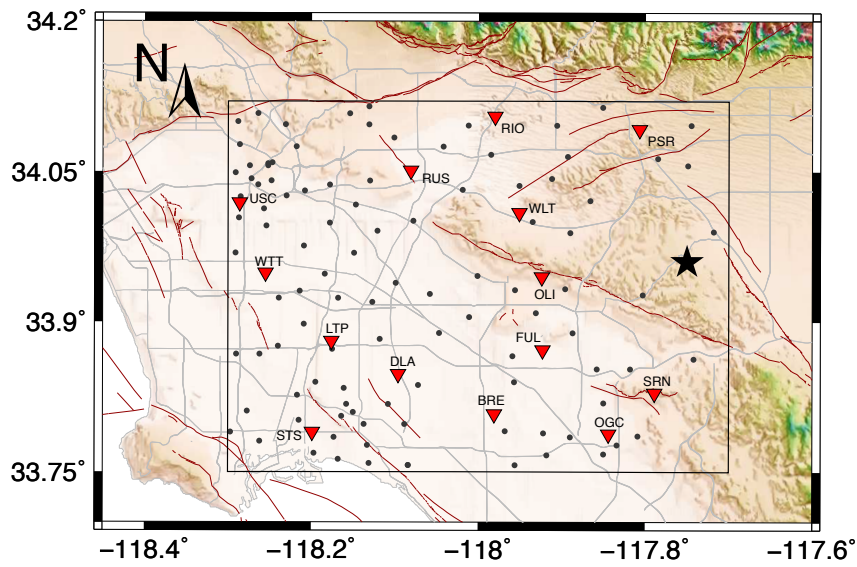


Figure 3.5: Station locations of the 118 strong-motions recording used during our validation. The star represents the epicenter of the M_w 5.4 Chino Hills event, and the black box denotes our simulation domain. Stations shown in red triangles are used for qualitative comparisons between the recordings and data.

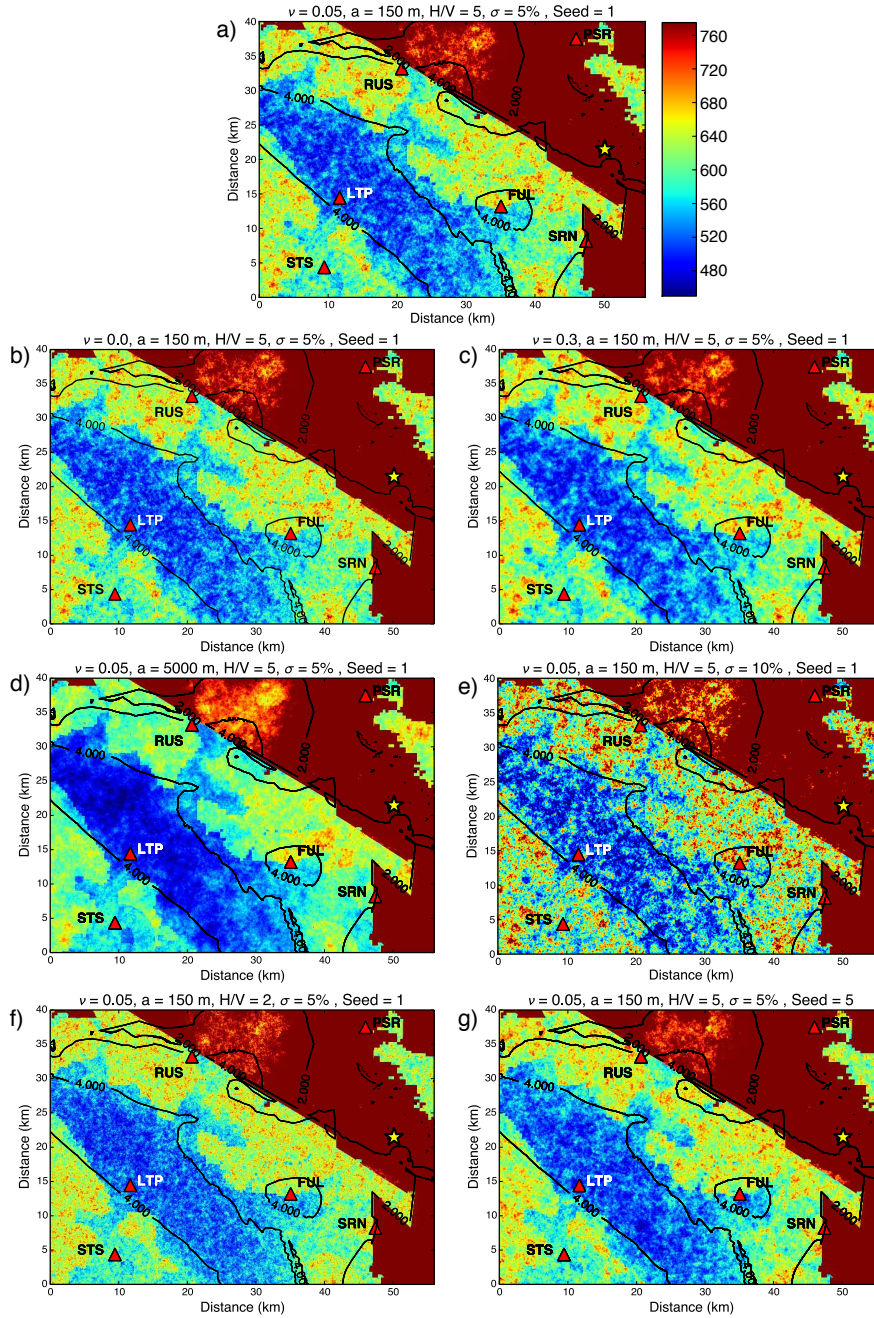


Figure 3.6: Horizontal slices shown at $z = 160 \text{ m}$ for different models of velocity fluctuations considered during sensitivity testing. The contours indicate depth in km to the $V_s = 2.5 \text{ km/s}$ isosurface ($Z_{2.5}$). We consider a base model (a) with parameters $\nu = 0.05, a = 150 \text{ m}, H/V = 5, \sigma = 5\%$ and random seed = 1. Sub-figures (b-g) show the effects of different parameterizations of the von Karman autocorrelation function. The parameters for each model are shown directly above the corresponding model. Table 3.2 shows the parameters of the velocity fluctuation models compared during the sensitivity test.

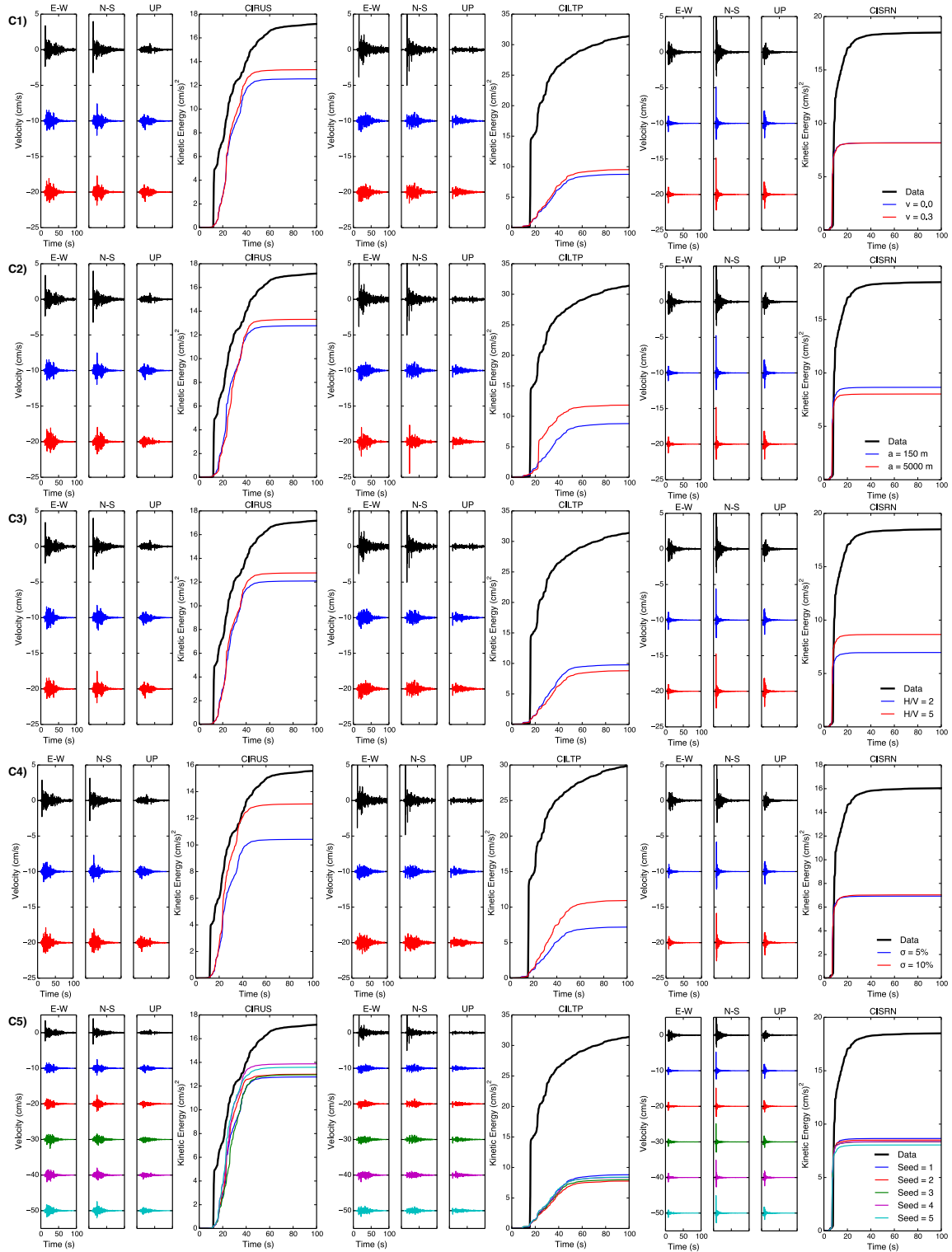


Figure 3.7: Qualitative comparisons shown for stations RUS, LTP and SRN. The left three panels show velocity seismograms (cm/s) and the large right panel shows cumulated energy (cm^2/s).

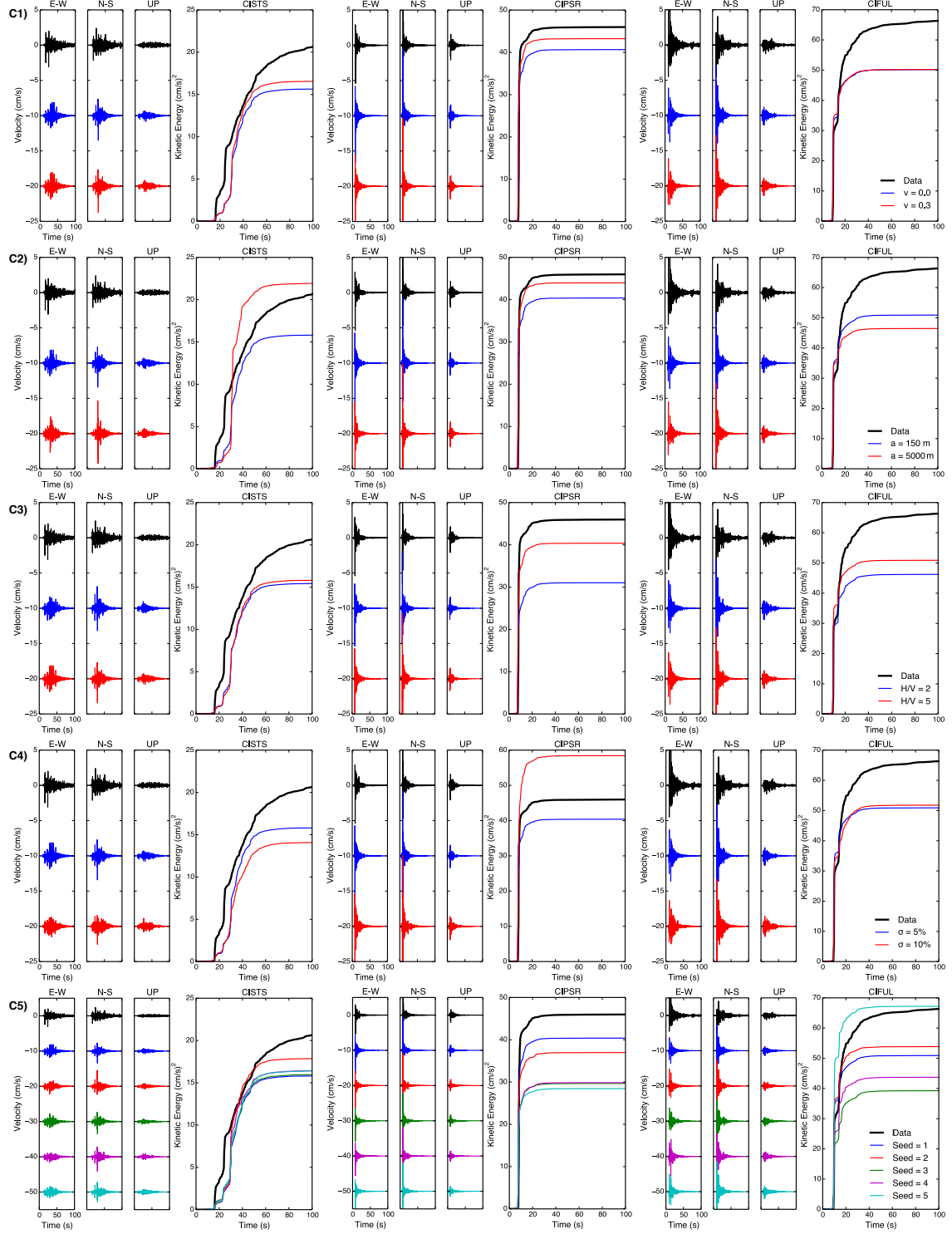


Figure 3.8: Same as Figure 3.7, but plotted for stations STS, PSR, and FUL.

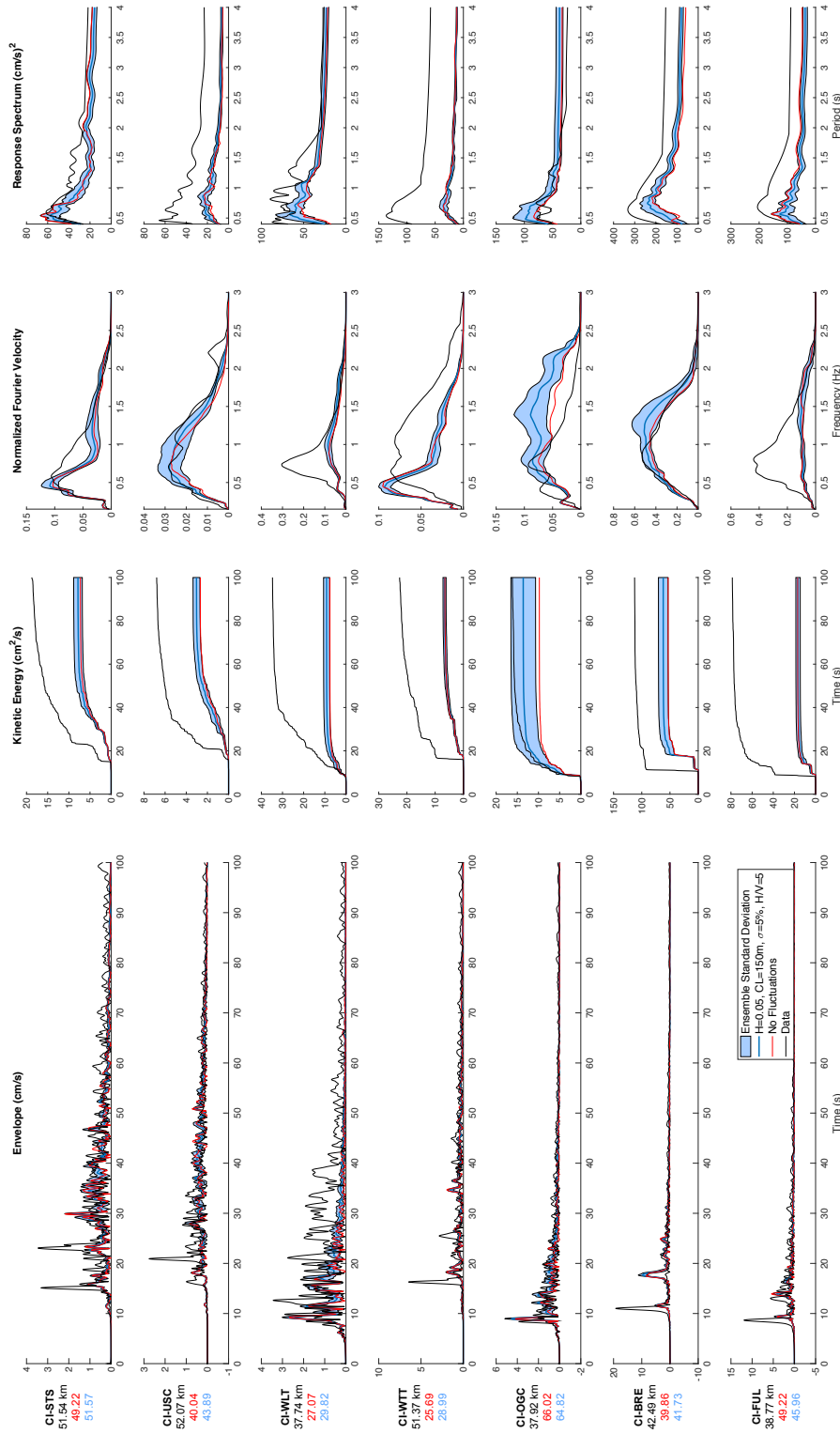


Figure 3.9: Comparisons between synthetics including velocity fluctuations (blue), synthetics without velocity fluctuations (red) and data (black). We plot one-side envelop functions, cumulative energy, Fourier velocity, and acceleration response spectra. Fourier velocity is normalized by the variance of the signal. The shaded blue region represents extrema from the five-simulation ensemble. The labels to the left of each figure provide the station name, r_{up} distance, and the value of the average GOF score for the station including fluctuations (blue) and without fluctuations (red). The GOF score for the ensemble (blue) represents the GOF averaged across all five simulations. The station locations of this analysis are shown in Figure 3.5.

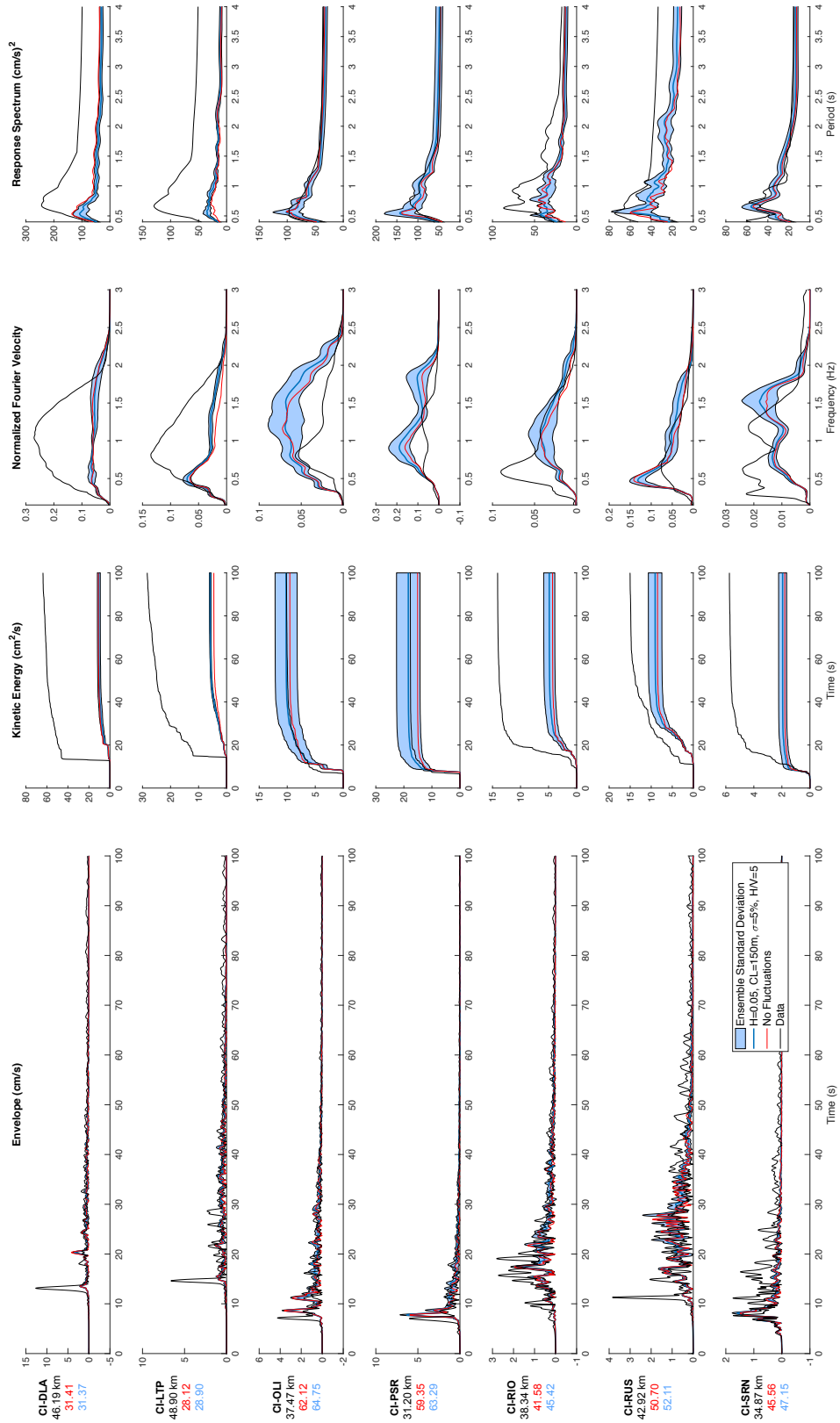


Figure 3.10: Same as Figure 3.9.

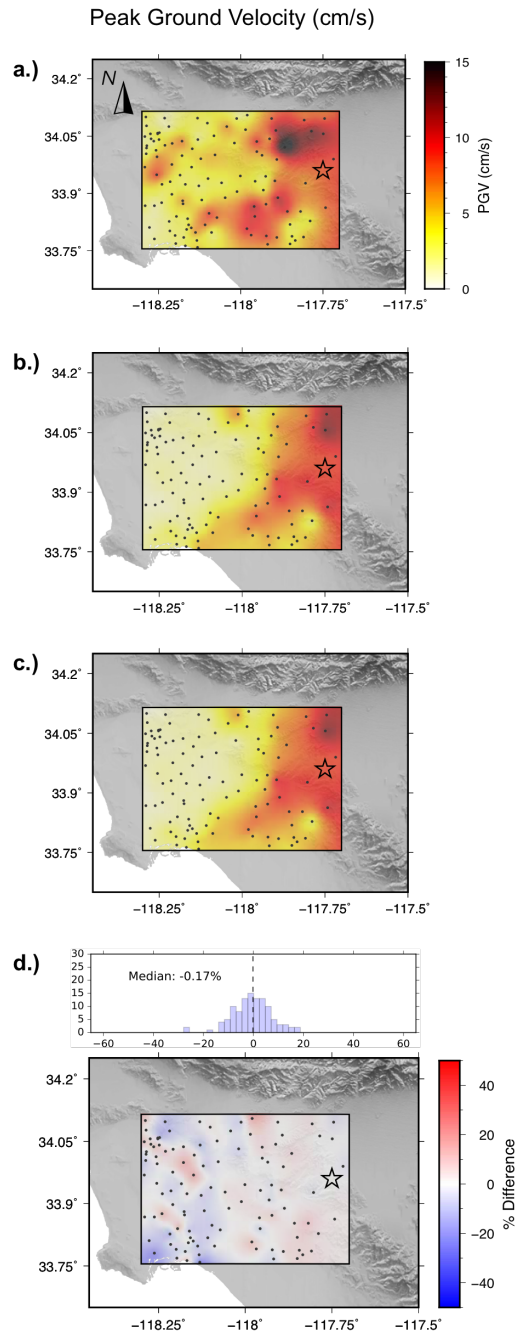


Figure 3.11: Model wide comparison of peak-ground velocity (PGV) between data and synthetics computed with and without models of velocity fluctuations. (a) Interpolated PGV recorded from the 2008 M_w 5.4 Chino Hills earthquake. (b) Interpolated PGV from simulation without models of velocity fluctuations. (c) Interpolated PGV of the ensemble average of five simulations considering models of velocity fluctuations with seeds 1 - 5. (d) Percent difference in PGV amplitude between ensemble of models considering velocity fluctuations against simulation that does not consider velocity fluctuations. The histogram shows the values represented in (d).

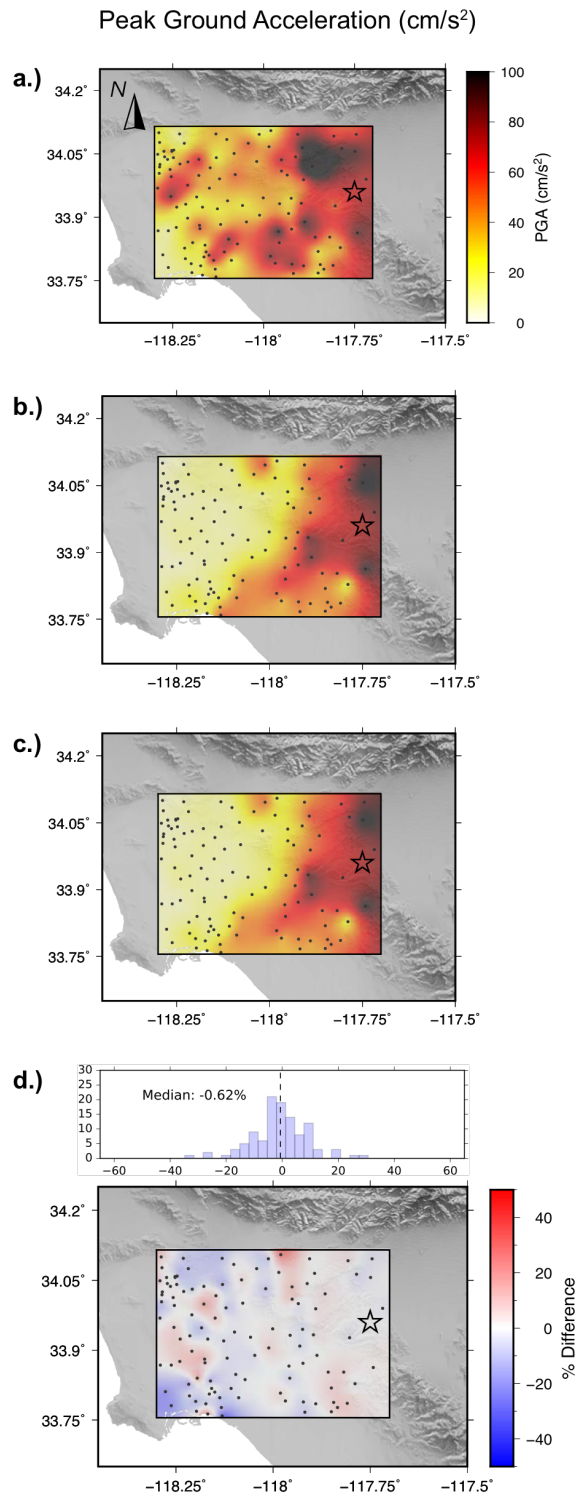


Figure 3.12: Same as Figure 3.11, but for peak-ground acceleration (PGA).

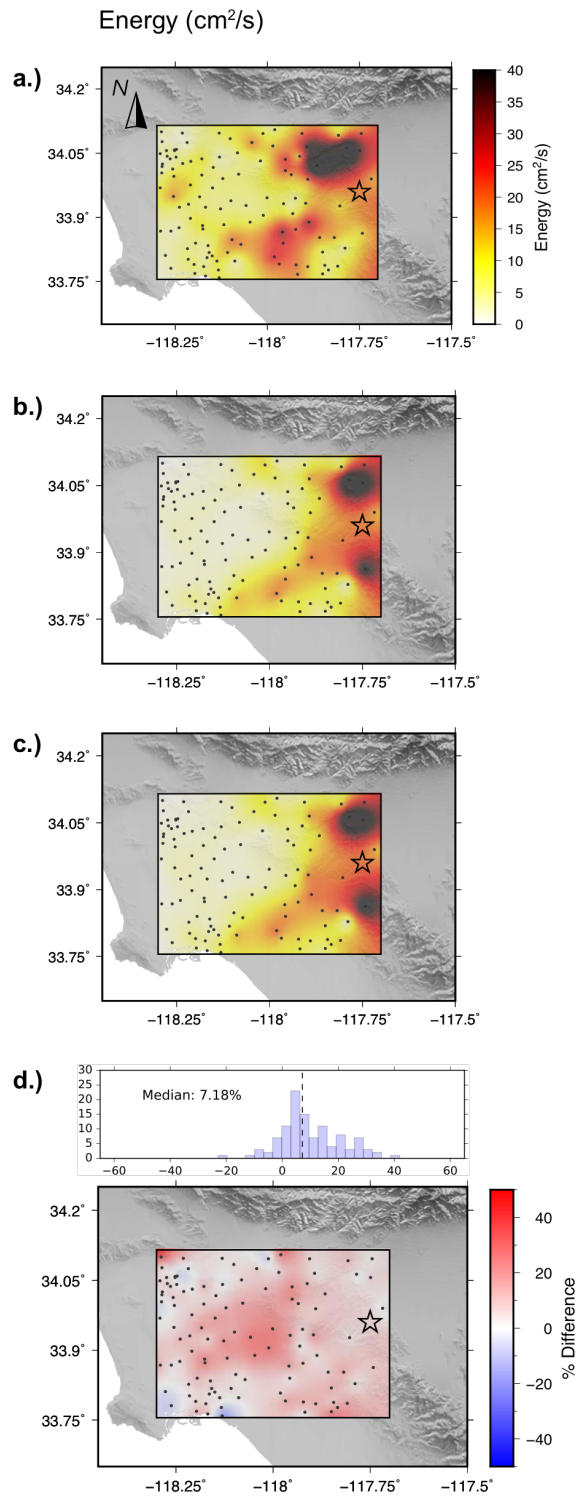


Figure 3.13: Same as Figure 3.11, but for cumulated energy (ENER).

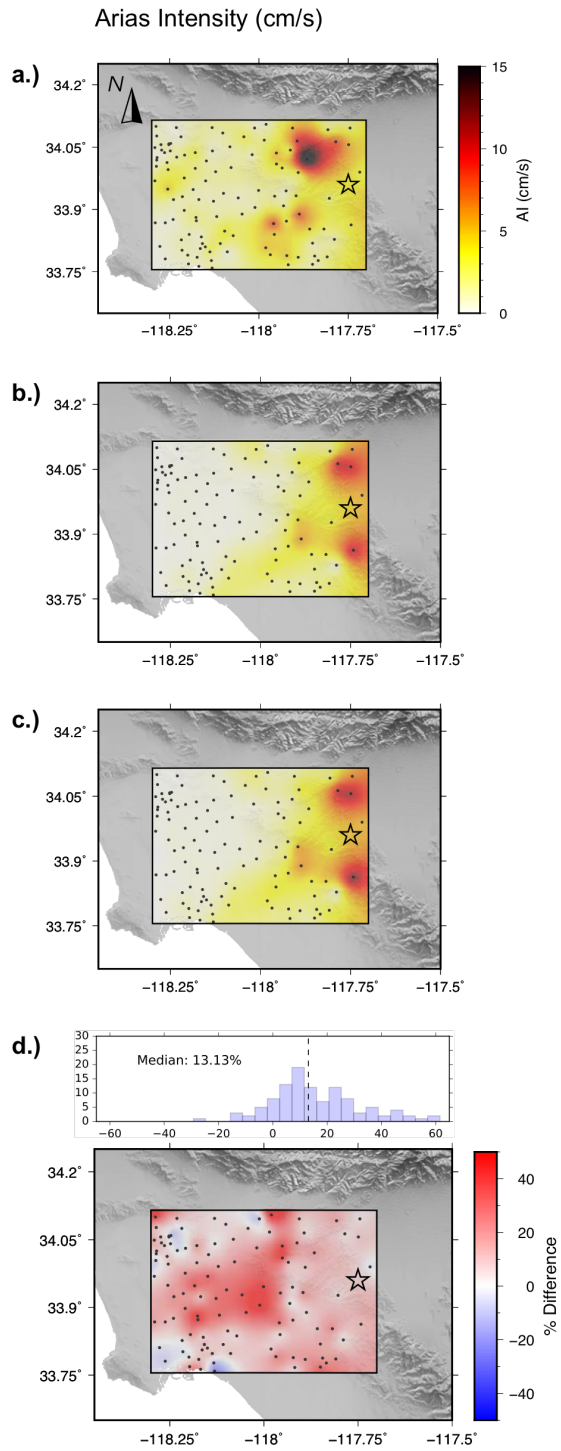


Figure 3.14: Same as Figure 3.11, but for Arias intensity (AI).

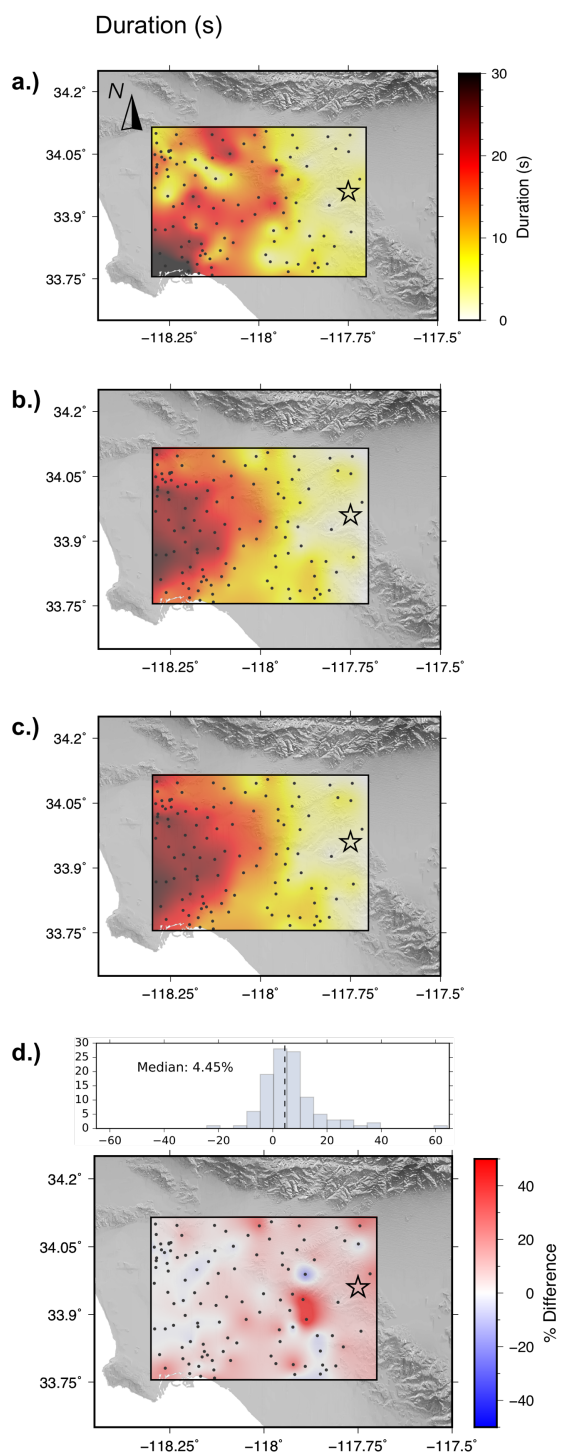


Figure 3.15: Same as Figure 3.11, but for duration (DUR).

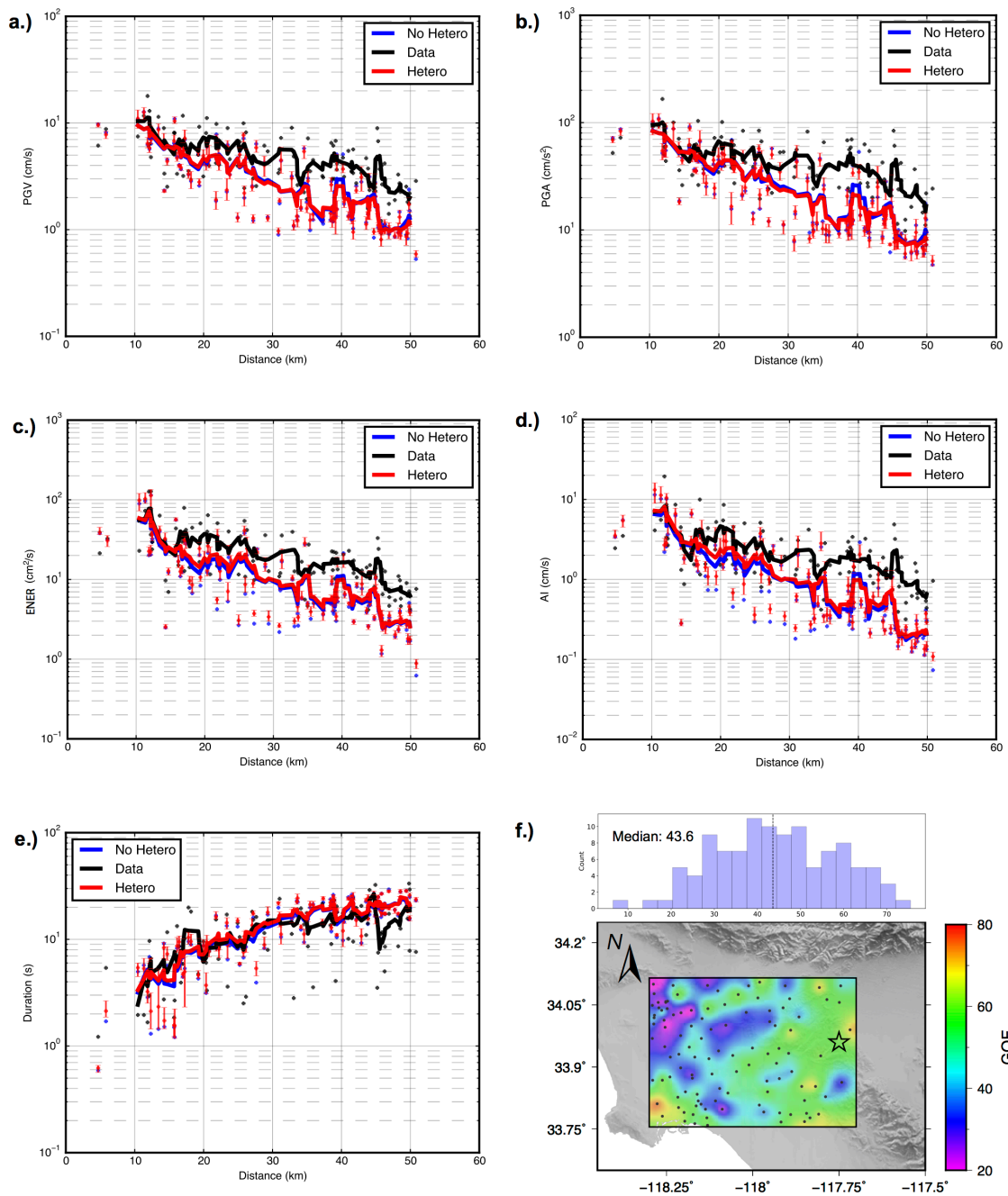


Figure 3.16: Five metrics PGV (a), PGA (b), ENER (c), AI (d), and DUR (e) plotted as a function of epicentral distance. Ensemble extrema are shown with red error bars, metric values not including heterogeneities are shown in blue, and data are shown using black. Solid lines represent five station simple moving-average. (f) Broadband (0.15-2.5 Hz) GOF shown for entire simulation domain. The histogram above the spatial plot shows the distribution of GOF scores in the simulation.

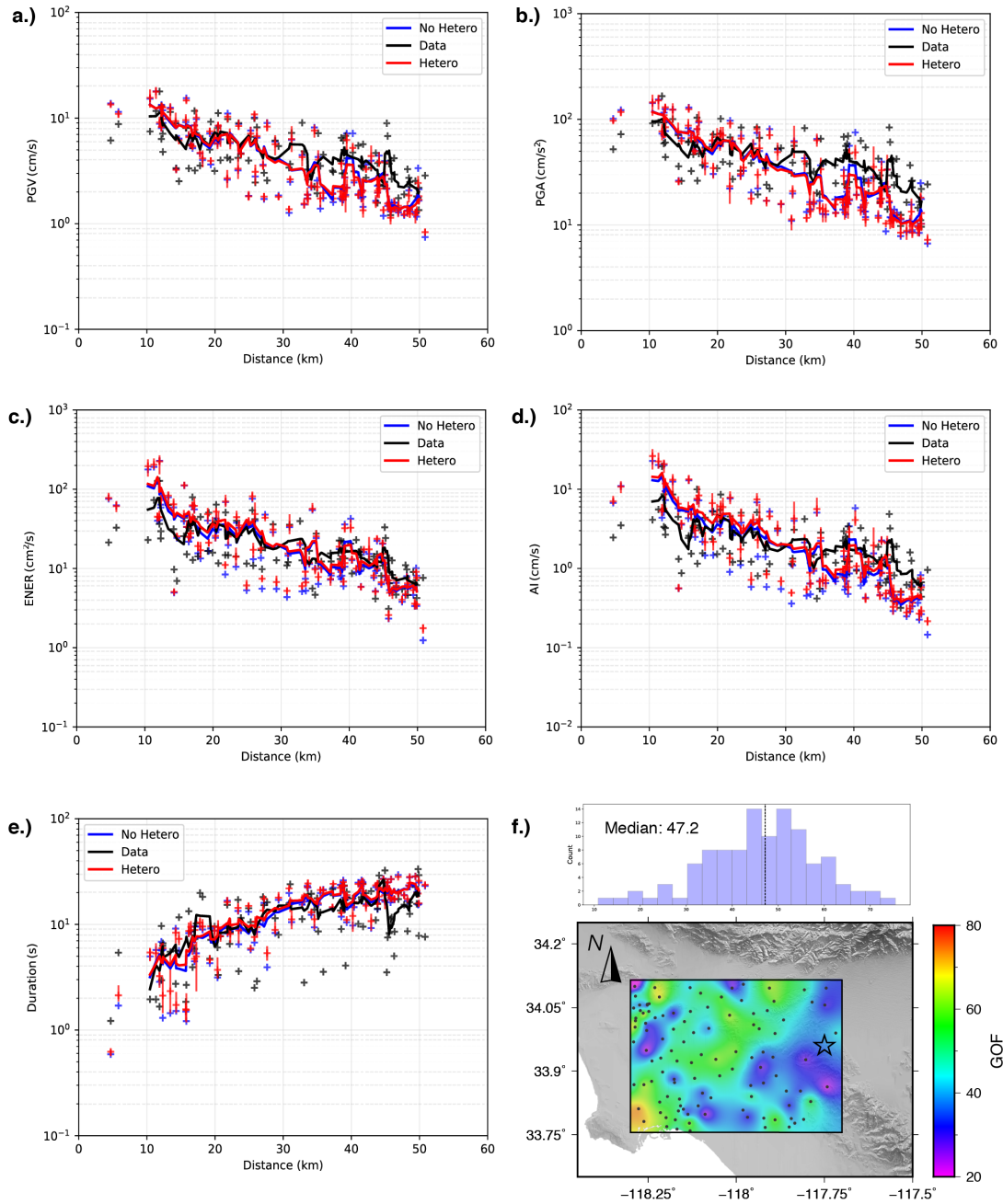


Figure 3.17: Same as Figure 3.16, but considering the Chino Hills earthquake having M_w 5.5 as opposed to M_w 5.4. This simulation uses the linear-attenuation relationship of $Q_s / V_s = 100$.

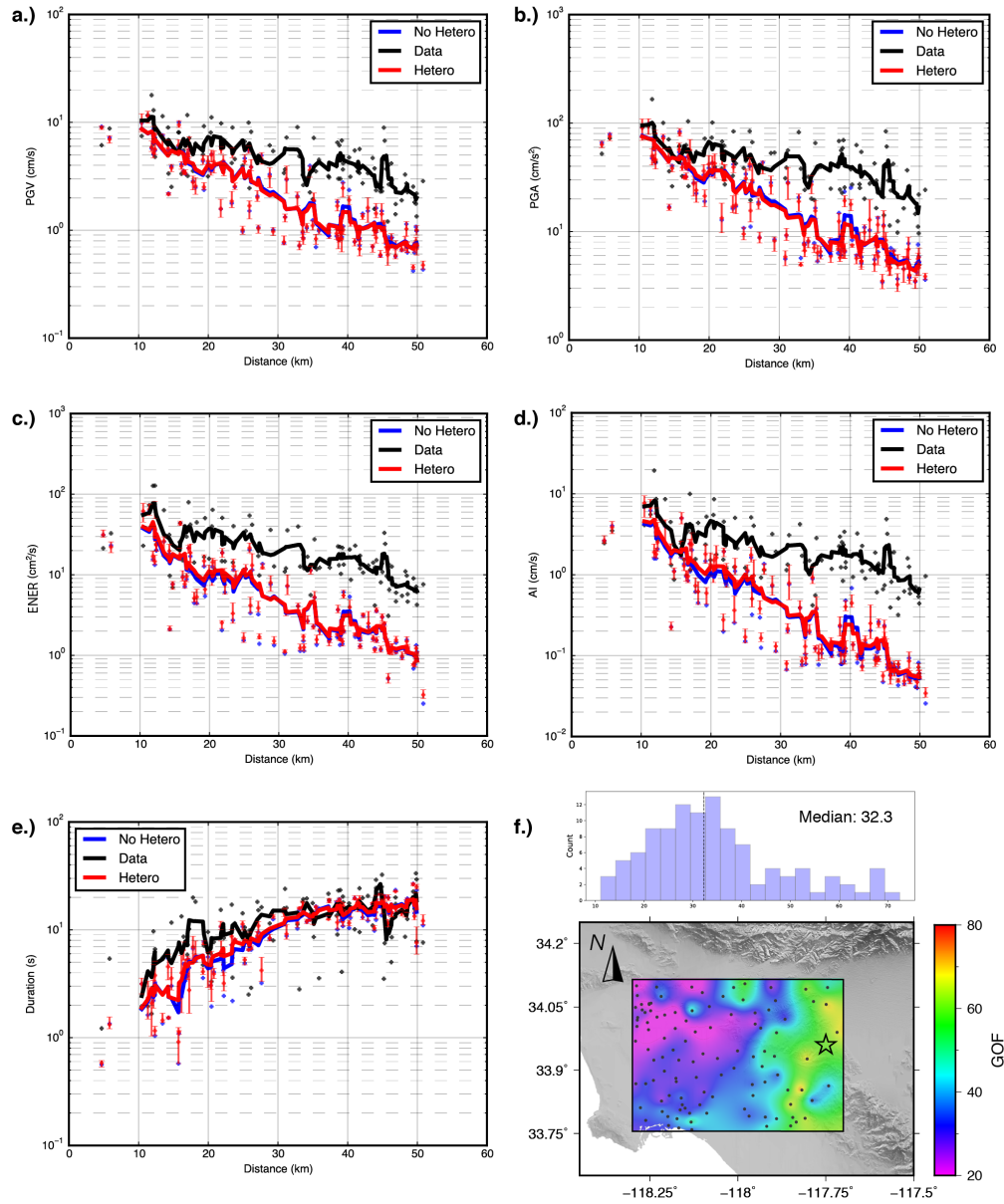


Figure 3.18: Same as Figure 3.16, but for the $Q_s / V_s = 50$ empirical relationship.

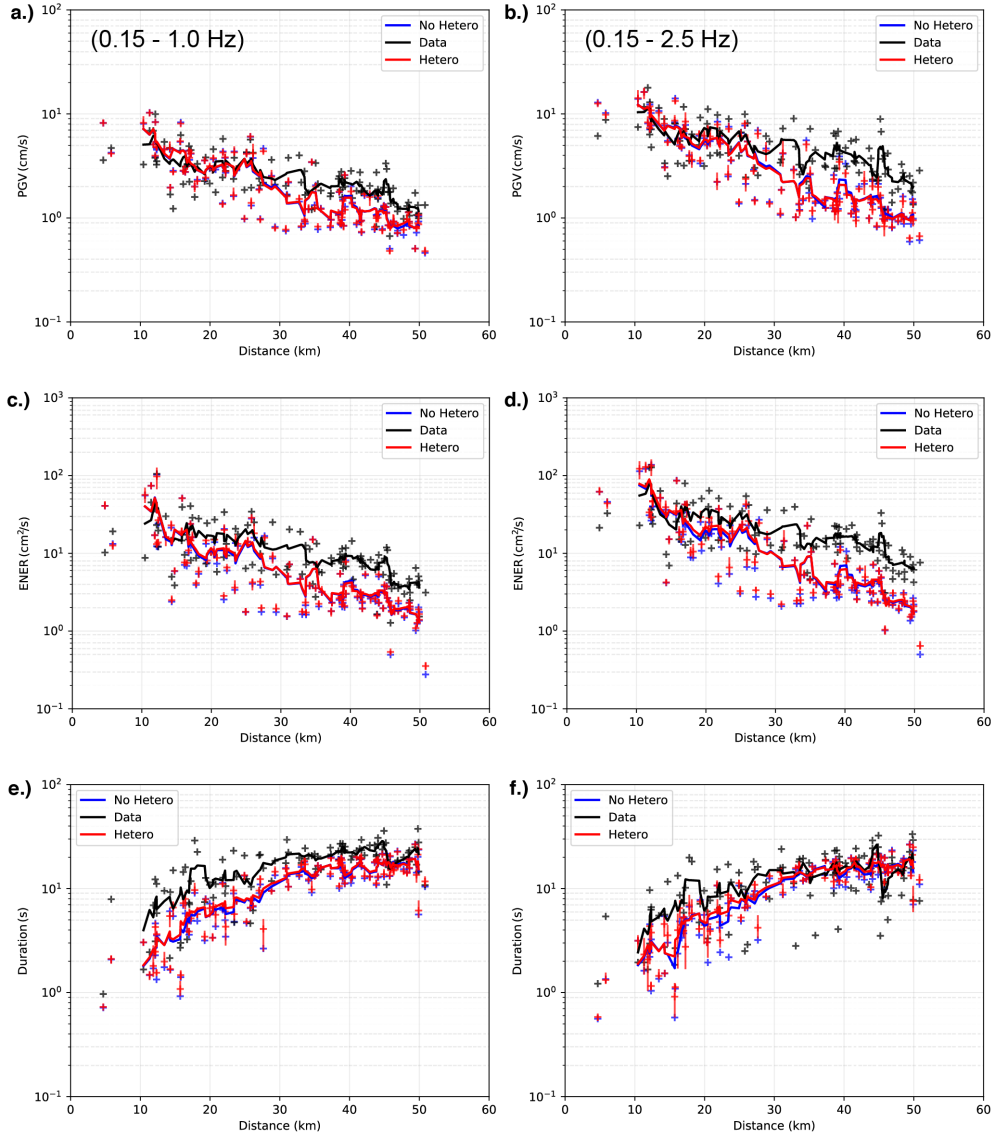


Figure 3.19: Comparisons against recorded data for simulations using source $M_w = 5.5$ and $Q_s/V_s = 50$ for metrics PGV (a), ENER (b), and DUR (c). We show results for bandwidth of 0.15 – 1.0 Hz (left column) and 0.15 – 2.5 Hz (right column) to discriminate any frequency-dependent effects of the attenuation model. Thick lines depict moving averages with window length = 5 for data (black), ensemble of simulations considering small-scale heterogeneities (red), and simulation considering no small-scale heterogeneities (blue). Ensemble extrema are represented by the red error bars.

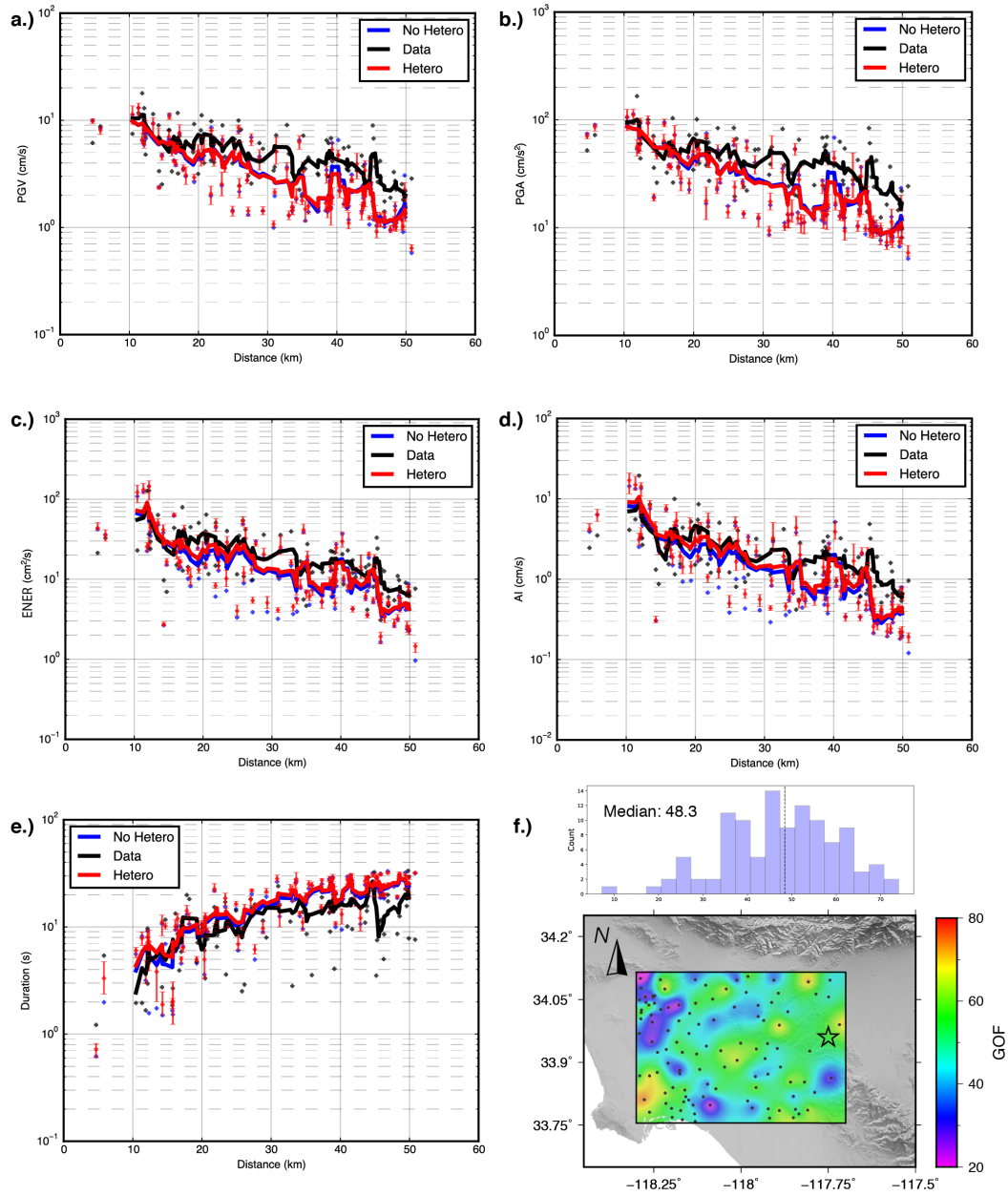


Figure 3.20: Same as Figure 3.16, but for $Q_s / V_s = 150$.

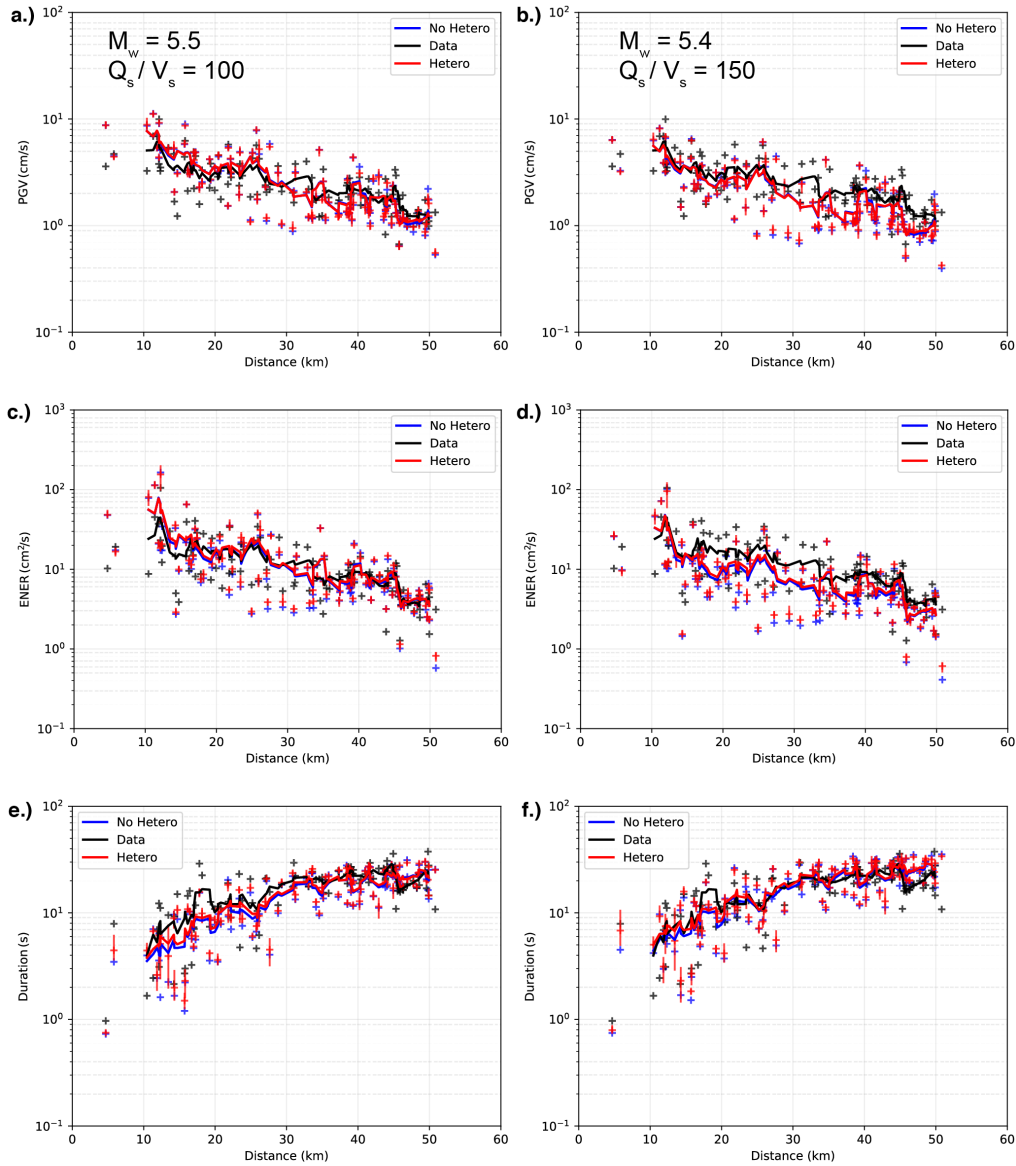


Figure 3.21: Comparisons between simulation assuming $M_w = 5.5$ and $Q_s / V_s = 100$ with simulation using $M_w = 5.4$ and $Q_s / V_s = 150$ for metrics PGV (a,b), PGA (c,d), and DUR (e,f). We bandpass filter the synthetics and data between 0.15 – 1.0 Hz before plotting. Thick lines depict moving averages with window length = 5 for data (black), ensemble of simulations considering small-scale heterogeneities (red), and simulation considering no small-scale heterogeneities (blue). Ensemble extrema are represented by the red error bars.

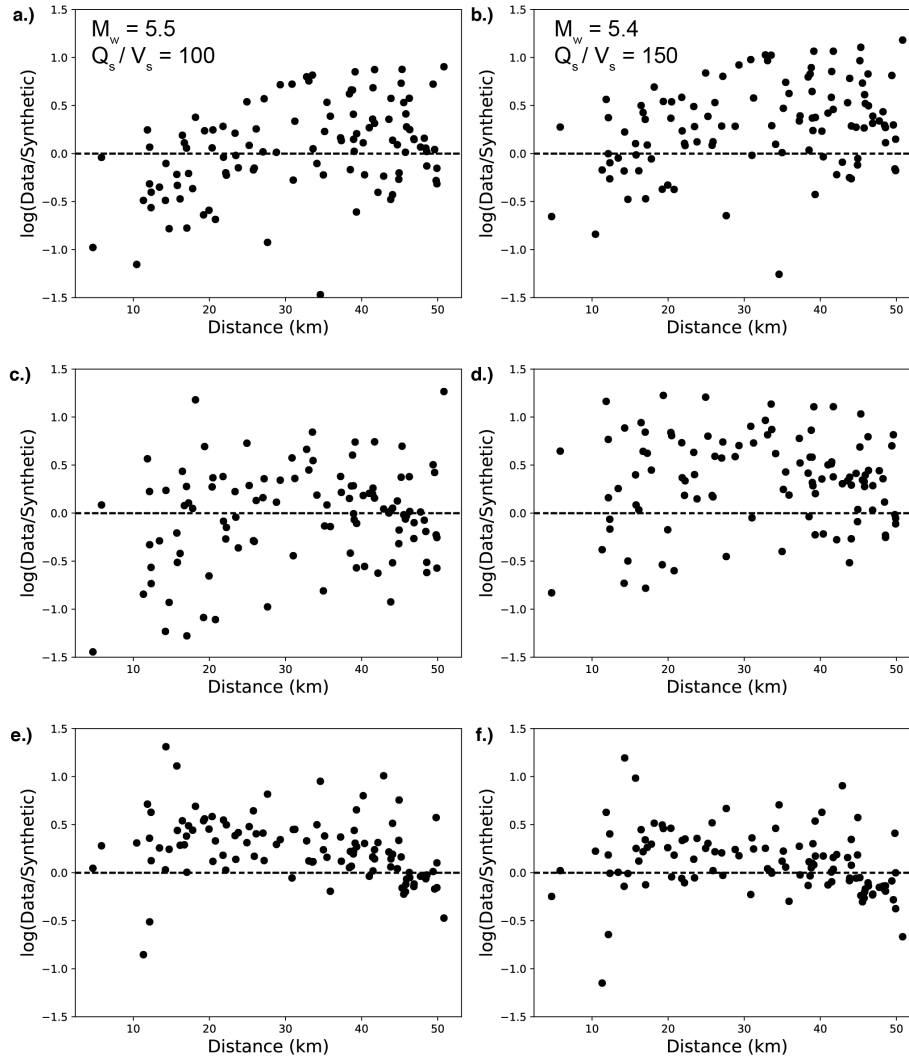


Figure 3.22. Plots showing the natural logarithm of the ratio between simulations and recorded data for metrics PGV (a,b), ENER (c,d) and DUR (e,f). We show biases for both the M_w 5.5 simulation using $Q_s / V_s = 100$ (left column) and the M_w 5.4 simulation using $Q_s / V_s = 150$ (right column). Bias plots are made using bandpass filtered data from 0.15 – 1.0 Hz.

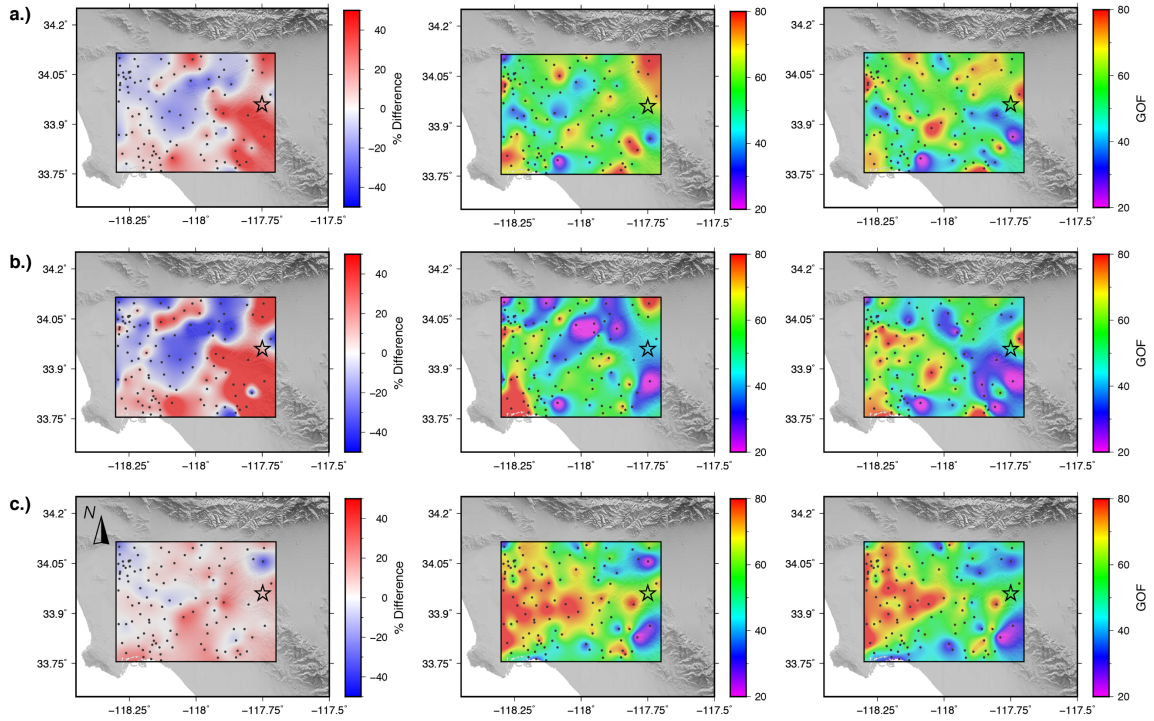


Figure 3.23: Percent difference of GOF (left column) computed between GOF scores for simulation considering M_w 5.4 source and $Q_s/V_s = 150$ attenuation relationship (center column) and the M_w 5.5 source with $Q_s/V_s=100$ (right column). We show PGV (a), ENER (b), and DUR (c) as individual metrics to be consistent with previous analyses. Percent difference plots are colored such that red areas indicate the $Q_s/V_s = 150$ model provides better GOF to data as compared with the $Q_s / V_s = 100$ model, and blue areas indicate the $Q_s/V_s = 100$ model with scaled source provides better GOF. Synthetics and data are bandpass filtered to 1.0 Hz before computing GOF scores and percent difference.

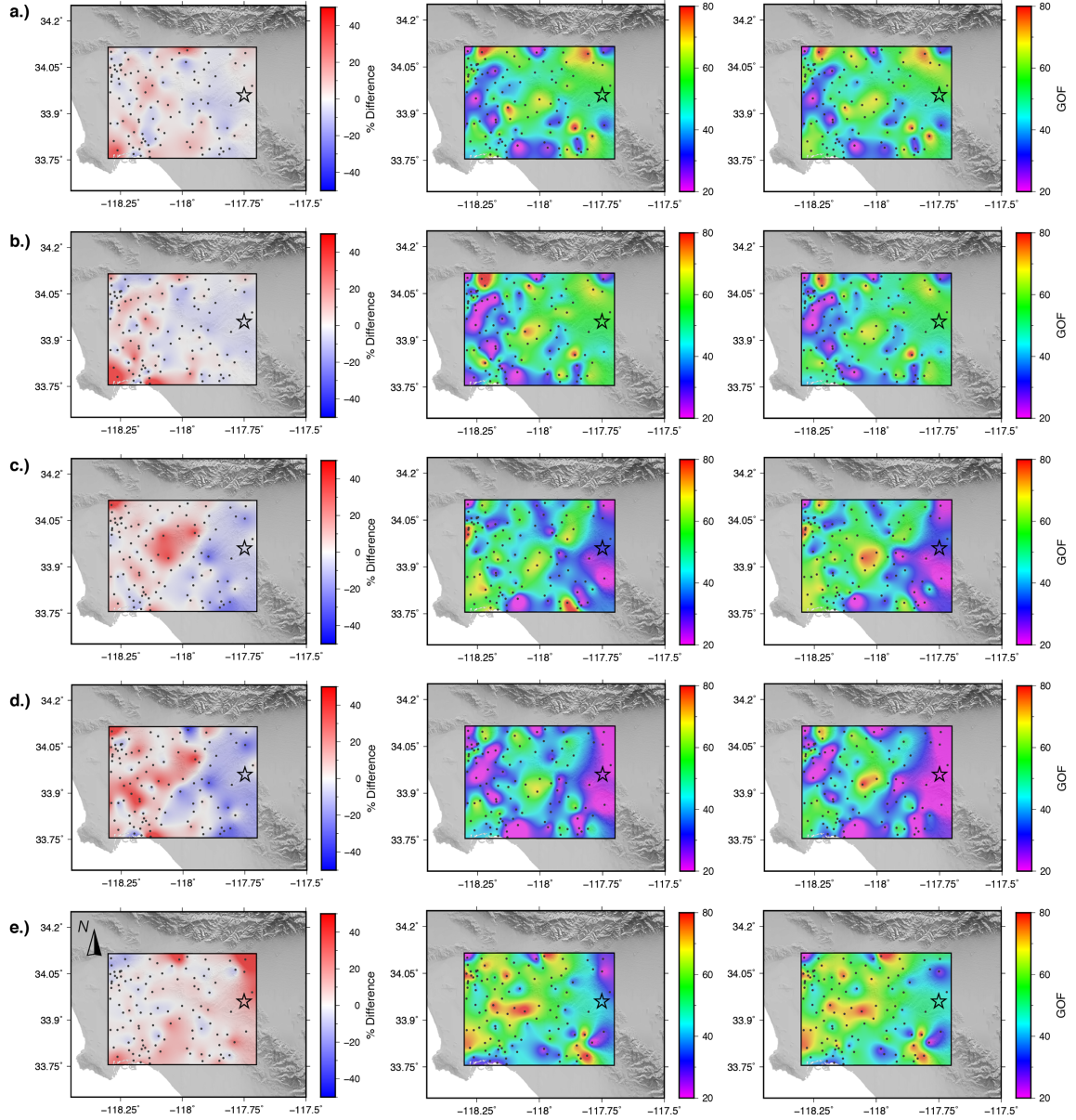


Figure 3.24: Percent difference of GOF (left column) computed between GOF scores for CVM-S4.26 (center column) and CVM-S4.26 + velocity fluctuations (right column). We show PGV (a), PGA (b), ENER (c), AI (d), and DUR (e) as individual metrics to be consistent with previous analyses. Percent difference plots are colored such that red areas indicate the CVM-S4.26 + velocity fluctuations provide better GOF to data as compared with CVM-S4.26, and blue areas indicate the CVM-S4.26 provides better GOF. Synthetics and data are lowpass filtered to 2.5 Hz before computing GOF scores and percent difference.

Chapter 4

Kinematic Rupture Generator Based on 3D Spontaneous Rupture Simulations with Geometrically Rough Faults

Spontaneous rupture simulations on geometrically-rough faults have been shown to produce realistic far-field spectra and comparable fits with GMPEs, but they are too computationally demanding for use with physics-based probabilistic seismic hazard analysis efforts. Here, we present our implementation of a kinematic rupture generator that mimics, at least in a statistical sense, the processes of rough-fault spontaneous rupture models in order to generate broadband source models. To this end, we analyze ~ 100 dynamic rupture simulations on strike-slip faults ranging from M_w 6.4 - 7.2. We find that our dynamic simulations follow empirical scaling relationships for inter-plate strike-slip events, and provide source spectra comparable with an ω^{-2} source model. To define our kinematic source model, we use a regularized Yoffe function parameterized in terms of slip, peak-time, rise-time, and rupture initiation time. These parameters are represented by random fields whose one- and two-point statistics are estimated from the dynamic rupture simulations. Our rupture generator reproduces GMPE medians and intra-event standard deviations for an ensemble of 10 Hz fully-deterministic ground motion simulations, as compared to NGA West2 GMPE relationships up to natural frequencies of 5 Hz. These results indicate that our rupture generator produces 10 Hz kinematic source models that can be used in broadband physics-based probabilistic seismic hazard efforts or to supplement limited data for the development of future GMPEs.

4.1 Introduction

Most buildings have resonance frequencies between 1 and 10 Hz, with single-family houses in the upper part of the range. Therefore, it is important to understand the behavior of these high-frequency waves to predict shaking from future earthquakes. Broadband physics-based ground motion simulations can provide insight into ground motions from future earthquakes; however, computing the high-frequency wave-field poses significant challenges. In addition to understanding the small-scale properties of the earth's crust and upper mantle, characterizing the high-frequency behavior (>5 Hz) of the earthquake's source still remains a largely unsolved problem.

One promising approach to broadband ground motion simulations uses dynamic rupture calculations on a geometrically rough fault. Geologists have observed that faults display roughness over several decades of scale lengths (e.g., *Candela et al.*, [2012]). Fault roughness has important implications for ground motions originating from rupture on these complex faults. The geometric complexity causes heterogeneous shear and normal tractions on the fault interface resulting in complex rupture behavior due to the variable strength along the fault interface. *Dunham et al.*, [2011]; *Shi and Day*, [2013] have shown that more complexity in the rupture process decreases the coherency of the wave front, and thus in the resulting ground motions. These studies show that dynamic rupture simulations on rough faults can produce desirable characteristics of observed ground motions such as flat acceleration spectra for a broad range of frequencies.

While this approach has been shown to capture a number of key features of recorded ground motion, its modeling advantages come at a high computational cost. In contrast with dynamically computed source models, kinematic earthquake sources provide a priori descriptions of the spatiotemporal evolution of slip along a fault interface. KRGs are typically constructed using information resulting from finite-fault inversions [*Mai and Beroza*, 2002], dynamic rupture simulations [*Schmedes et al.*, 2010; 2013; *Song et al.*, 2013; *Trugman and Dunham*, 2014; *Song*, 2015], or theoretical considerations regarding earthquake rupture [*Andrews*, 1980; *Guatteri et al.*, 2004].

Typically, the KRG is generated from a stochastic spatial distribution of slip and relating the remaining rupture parameters through direct correlations or empirical relationships derived from various earthquake source studies. The stochastic nature of the KRGs introduces heterogeneity in the rupture using spatially varying fields to define the rupture process. These fields constitute the parameters of a slip-rate function used to define the temporal evolution of slip on each subfault. The choices of slip-rate functions are by no means unique and differ based on the individual implementation of the kinematic rupture generator.

Currently both deterministic and so-called hybrid ground motion simulations use kinematic rupture generators (e.g. *Graves and Pitarka*, [2015]; *Olsen and Takedatsu*, [2015]) or directly incorporate the dynamic rupture description (e.g. *Olsen et al.*, [2009]; *Roten et al.*, [2014]) to define their input source models. Kinematic rupture generators provide a computational convenience, as large dynamic rupture simulations are too computationally expensive for physics-based probabilistic seismic hazard efforts such as the Southern California Earthquake Center (SCEC) CyberShake [*Graves et al.*, 2011] or Broadband Platform (e.g., *Goulet et al.*, [2015]) requiring large ensembles of rupture realizations.

In this manuscript, we present our implementation of a KRG based on 3D rough-fault dynamic rupture simulations. First, we provide an overview of our KRG. We follow with an explanation of the statistical relationships derived from the dynamic rupture calculations used to generate our kinematic sources. Finally, we show examples of 10 Hz fully-deterministic ground-motions originating from these KRG sources and comparisons against NGA-West2 GMPEs – effectively validating our rupture generator for generating realistic broadband ground motions up to 10 Hz.

4.1.1 Statistical Framework for the Proposed KRG

A finite-fault kinematic earthquake source can be defined as a collection of discrete subfaults each having an individually defined slip history. Within the finite-fault, each subfault has its evolution of slip prescribed through a chosen slip-rate function. We

use the regularized Yoffe slip-rate function [Tinti *et al.*, 2005] in our KRG formulation. The regularized Yoffe function provides an appealing choice due to the empirical connections made to dynamic rupture models (e.g. [Tinti *et al.*, 2005; 2009]) and the theoretical basis of the Yoffe function [Yoffe, 1951]. The Yoffe slip-rate can be expressed as

$$\dot{s}(t) = \Delta u \int_{-\infty}^{\infty} g(t - T) y(t) dT, \quad 4.1$$

where,

$$g(t) = \frac{1}{\tau_s^2} [tH(t)H(\tau_p - t) + (2\tau_s - t)H(t - \tau_s)H(2\tau_s - t)],$$

and

$$y(t) = \frac{2}{\pi\tau_r} H(t)H(\tau_r - t)\sqrt{\tau_r - t}.$$

The parameters: final slip (Δu), peak time (τ_s), rise time (τ_r) and rupture initiation time (t_0) are needed to completely define the regularized Yoffe function (Figure 4.1). Our kinematic rupture generator uses statistics computed from dynamic rupture simulations on geometrically rough faults to simulate spatial fields representing parameters of this source-time function. These spatial distributions of parameters define the source-time function at each subfault, which completely describes the kinematic finite-fault source in both space and time.

Our kinematic rupture generator operates conceptually in two different phases. First, we simulate the ‘physical’ spatial fields: Δu , peak slip velocity (v_{peak}), and normalized rupture velocity (v_{rup}/c_s) based on the one- and two-point statistics computed from dynamic rupture simulations on rough faults. Second, we compute the parameters of the regularized Yoffe function (t_0, τ_r, τ_s) using empirical relationships between the simulated ‘physical’ fields and these kinematic parameters.

Two recent studies by *Trugman and Dunham*, [2014] and *Yao*, [2017] investigated spatial correlations between rough fault topography and ‘physical’ spatial fields using dynamic ruptures on rough faults. In the 2D case, *Trugman and Dunham*, [2014] found that Δu , V_{peak} , and v_{rup}/c_s were strongly correlated with the local slope of the rough fault topography. However, for 3D, *Yao*, [2017] found that the ‘physical’ source fields are correlated with the initial friction (μ_0) or the ratio of shear traction to normal traction on the fault. The 2D and 3D results are consistent in the sense that μ_0 represents a proxy for the initial fault slope. These relationships constitute a physical basis for the relationship between the rough fault topography and the resulting rupture process. As such, we follow the general guidance of *Yao*, [2017] when constructing our KRG.

Linear Model of Coregionalization

To describe the spatial interdependency noticed by *Trugman and Dunham*, [2014] and *Yao*, [2017], we define a 4D Gaussian random variable consisting of Δu , V_{peak} , v_{rup}/c_s , and μ_0 having its spatial correlations defined by a linear model of coregionalization. In this section, we provide a brief background on linear models of coregionalization and list some advantages and disadvantages over using empirically defined correlation structures (e.g. *Song et al.*, [2013]).

Linear models of (co)regionalization are used to define a permissible model of spatial correlation (covariance matrix, $C(\mathbf{h})$, containing auto- and cross-correlations) between model features. A covariance matrix is said to be permissible if the matrix is positive-definite. The simplest way to ensure that permissible covariance structures exist is through building a set of intercorrelated random functions $Z_i(\mathbf{u})$ (Equation 4.2) defined as a linear combination of independent random functions $Y_k^l(\mathbf{u})$. We summarize the theoretical background explained in [*Goovaerts*, 1997] in the following.

Let $Z_i(\mathbf{u})$ be a set of intercorrelated random functions:

$$Z_i(\mathbf{u}) = \sum_{l=0}^L \sum_{k=1}^{n_l} a_{ik}^l Y_k^l(\mathbf{u}) + m_i \quad 4.2$$

The random function $Y_k^l(\mathbf{u})$ has zero mean and a basic covariance function called $c_l(\mathbf{h})$. In addition, $\text{Cov}\{Y_k^l(\mathbf{u}), Y_{k'}^{l'}(\mathbf{u})\} = c_l(\mathbf{h})\delta_{ll'}\delta_{kk'}$, which shows the independence of these functions. One can define the cross-covariance between any two random variables $Z_i(\mathbf{u})$ and $Z_j(\mathbf{u} + \mathbf{h})$ using a linear combination of cross covariances between two random variables $Y_k^l(\mathbf{u})$ and $Y_{k'}^{l'}(\mathbf{u} + \mathbf{h})$. This gives the covariance matrix shown in Equation 4.3.

$$\begin{aligned} C_{ij}(\mathbf{h}) &= \text{Cov}\{Z_i(\mathbf{u}), Z_j(\mathbf{u} + \mathbf{h})\} \\ &= \sum_{l=0}^L \sum_{l'=1}^L \sum_{k=1}^{n_l} \sum_{k'=1}^{n_{l'}} a_{ik}^l a_{jk'}^{l'} \text{Cov}\{Y_k^l(\mathbf{u}), Y_{k'}^{l'}(\mathbf{u} + \mathbf{h})\}. \end{aligned} \quad 4.3$$

We can simplify this expression based on the independence between random variables $Y_k^l(\mathbf{u})$ and $Y_{k'}^{l'}(\mathbf{u} + \mathbf{h})$. Therefore,

$$C_{ij}(\mathbf{h}) = \sum_{l=0}^L \sum_{k=1}^{n_l} a_{ik}^l a_{jk}^l c_l(\mathbf{h}). \quad 4.4$$

From Equation 4.4 we define the linear model of coregionalization as the set of N_v by N_v auto- and cross- covariance models $C_{ij}(\mathbf{h})$ defined as

$$C_{ij}(\mathbf{h}) = \sum_{l=0}^L b_{ij}^l c_l(\mathbf{h}). \quad 4.5$$

Looking at Equation 4.3, we see that the sill b_{ij}^l of $c_l(\mathbf{h})$ is $b_{ij}^l = \sum_{k=1}^{n_l} a_{ik}^l a_{jk}^l$. The matrix $\mathbf{B}_l = b_{ij}^l$ is referred to as a coregionalization matrix. The two required conditions for the linear model of coregionalization (Equation 4.5) to be permissible are that (1) the covariance functions $c_l(\mathbf{h})$ are permissible and (2) the $(L+1)$ coregionalization matrices $\mathbf{B}_l$ are positive semi-definite.

We showed the statistical analysis in terms of basic covariance functions $c_l(\mathbf{h})$, but the linear model of coregionalization can also be defined using semivariogram functions. We note that we use a semivariogram based approach in our KRG. If we assume that $g_l(\mathbf{h})$ represents a semivariogram of the random function $Y_k^l(\mathbf{u})$, we denote the cross-semivariogram between $Y_k^l(\mathbf{u})$ and $Y_{k'}^{l'}(\mathbf{u} + \mathbf{h})$ using Equation 4.6

$$\frac{1}{2} E\{[Y_k^l(\mathbf{u}) - Y_k^l(\mathbf{u} + \mathbf{h})][Y_{k'}^{l'}(\mathbf{u}) - Y_{k'}^{l'}(\mathbf{u} + \mathbf{h})]\} = g_l(\mathbf{h})\delta_{ll'}\delta_{kk'} \quad 4.6$$

We can then define the linear model of coregionalization incorporating semivariograms using

$$\gamma_{ij}(\mathbf{h}) = \sum_{l=0}^L b_{ij}^l g_l(\mathbf{h}), \quad 4.7$$

where the same two conditions apply to ensure a permissible linear model of coregionalization. Equation 4.7 states that the semivariogram γ_{ij} represents the weighted sum of permissible semivariogram models $g_l(\mathbf{h})$ having weighting coefficients b_{ij}^l . We refer to the semivariogram matrix as $\mathbf{\Gamma}(\mathbf{h})$.

By using linear combinations of negative semidefinite semivariogram models, $\gamma(\mathbf{h})$, we ensure that the linear model of coregionalization produces non-negative variance, assuming $\mathbf{B}_l$ are positive semi-definite. For completeness, we define permissibility for the multivariate case, let $Z_i(\mathbf{u})$ be a set of N_v intercorrelated random functions with $\{\mathbf{u}_\alpha, \alpha = 1, \dots, n\}$ a set of n data locations. We can write the variance of any finite linear combination Y of the variables $Z_i(\mathbf{u})$ as a linear combination of cross covariances.

$$\begin{aligned} \text{Var}\{Y\} &= \text{Var}\left\{\sum_{i=1}^{N_v} \sum_{\alpha=1}^n \lambda_{\alpha i} Z_i(\mathbf{u}_\alpha)\right\} \\ &= \sum_{i=1}^{N_v} \sum_{j=1}^{N_v} \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha i} \lambda_{\beta j} C_{ij}(\mathbf{u}_\alpha - \mathbf{u}_\beta) \geq 0. \end{aligned} \quad 4.8$$

If we rewrite Equation 4.7 in terms of semivariograms using the relationship $\mathbf{C}(\mathbf{h}) = \mathbf{C}(0) - \mathbf{\Gamma}(\mathbf{h})$ we find a similar definition. Using a matrix notation,

$$\text{Var}\{Y\} = \mathbf{C}(0) \sum_{\alpha=1}^n \lambda_{\alpha}^T \sum_{\beta=1}^n \lambda_{\beta} - \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha}^T \mathbf{\Gamma}(\mathbf{h}) \lambda_{\beta}^T \quad 4.9$$

Equation 4.8 assumes that the semivariogram models are bounded, therefore having a covariance counterpart. Equation 4.9 shows that in order for $\text{Var}[Y]$ to be non-negative the matrix of semivariogram models $\mathbf{\Gamma}(\mathbf{h})$ must be negative-semidefinite.

In practice, linear models of coregionalization are constructed using linear combinations of semivariogram functions that have been shown to be permissible. This eliminates the need to test a posteriori whether the covariance matrix is permissible [Christakos, 1984]. Here, we use the exponential semivariogram function (Equation 4.10) in our linear model of coregionalization.

$$g(\mathbf{h}) = 1 - \exp\left(-\frac{3\mathbf{h}}{a}\right), \quad 4.10$$

These models can be combined with different range parameters, a , to capture the effects of various scale lengths in the problem.

The primary advantage of using a linear model of coregionalization is that permissible models can be easily verified. Assuming that our $g_l(\mathbf{h})$ are permissible we only must verify the $(L+1)$ coregionalization matrices are positive semi-definite, rather than the entire covariance matrix $\mathbf{C}(\mathbf{h})$. This process can be quite computationally expensive for large covariance matrices as mentioned by Song *et al.*, [2013]. On the other hand, the linear model of coregionalization requires that all direct and cross semivariograms share the same set of $g_l(\mathbf{h})$, which prevents complex correlation structures within the model. Despite this limitation, we find the linear model of

coregionalization provides a worthwhile tradeoff between Fourier based approaches and using direct covariance-matrix based approaches.

Sequential Gaussian Simulation

Several techniques exist for simulating realizations of stochastic fields, with the most common technique for KRG applications being unconditional Fourier based methods (e.g., *Graves and Pitarka*, [2010]; *Schmedes et al.*, [2013]) which are described by *Klimeš*, [2002]. Additionally, one can simulate random fields for a multivariate Gaussian random variable using an approach called sequential Gaussian simulation. This approach outlines simulating a single field $Z_1(\mathbf{u}_\alpha)$, but joint simulation of multiple fields follows the same approach. See *Goovaerts*, [1997] for a complete explanation of this method.

The sequential Gaussian simulation process begins by applying a normal-score transformation to each set of z_i data $\{z_i(\mathbf{u}_{\alpha_i}), \alpha_i = 1, \dots, n_i\}$ into the normal-score y_i $\{y_i(\mathbf{u}_{\alpha_i}), \alpha_i = 1, \dots, n_i\}$ data using Equation 4.11, where $\mathbf{u}_\alpha$ are the n data points:

$$y_i(\mathbf{u}_\alpha) = G^{-1} \left[F_i^* \left(z_i(\mathbf{u}_{\alpha_i}) \right) \right] \quad 4.11$$

where G^{-1} represents the standard normal quantile function, and F_i^* is the marginal cdf of the random variable Z_i . Notice that y_i and z_i share the same cumulative probability, but have different values. Sampling from the multivariate Gaussian variable requires knowing both the mean and variance of the conditional cumulative distribution function (ccdf) for Y . The ccdf of Y can be written as a chain of one-point ccdfs:

$$\begin{aligned} F[Y(\mathbf{u}_1, \dots, \mathbf{u}_n)] &= F[Y(\mathbf{u}_N) | Y(\mathbf{u}_1, \dots, Y(\mathbf{u}_{N-1}))] \\ &\cdot F[Y(\mathbf{u}_{N-1}) | Y(\mathbf{u}_1, \dots, Y(\mathbf{u}_{N-2}))] \cdot \dots \\ &\cdot F[Y(\mathbf{u}_2) | Y(\mathbf{u}_1)] \cdot F[Y(\mathbf{u}_1)] \end{aligned} \quad 4.12$$

We note that Equation 4.12 contains data values provided prior to the sequential simulation (μ_0 in this case) in addition to values obtained during the simulation (i.e. Δu , V_{peak} , and v_{rup}/c_s). In other words, each simulated value conditions the ccdf used for

subsequently simulated values. We can estimate the mean and variance of the ccdf, using the simple cokriging (SCK) estimate $y_{SCK}^*(\mathbf{u})$ and simple cokriging variance $(\sigma_{SCK}(\mathbf{u}))^2$. Using the simple cokriging estimate and variance, we can write the ccdf

$$[G(\mathbf{u}; y_i | (n))]_{SCK}^* = G\left(\frac{y_i - y_{SCK}^{(i)*}(\mathbf{u})}{\sigma_{SCK}(\mathbf{u})}\right) \quad 4.13$$

where y_i represents a datum from previous iterations. The mean and variance are given by Equation 4.14 and Equation 4.15.

$$y_{SCK}^{(i)*}(\mathbf{u}) = \sum_{i=1}^{N_v} \sum_{\alpha_i=1}^{n_i(\mathbf{u})} \lambda_{\alpha_i}^{SCK}(\mathbf{u}) y_i(\mathbf{u}_{\alpha_i}) \quad 4.14$$

and

$$(\sigma_{SCK}(\mathbf{u}))^2 = C_{11}(0) - \sum_{i=1}^{N_v} \sum_{\alpha_i=1}^{n_i(\mathbf{u})} \lambda_{\alpha_i}^{SCK}(\mathbf{u}) C_{i1}(\mathbf{u}_{\alpha_i} - \mathbf{u}). \quad 4.15$$

Here G represents the standard normal distribution function. The kriging weights $\lambda_{\alpha_i}^{SCK}$ are computed by solving a simple cokriging system. We refer readers to *Goovaerts*, [1997] for an explanation of simple cokriging. Even though we incorporate simple kriging in our simulation method, more sophisticated kriging algorithms could be included within this framework.

With the basic theoretical background covered, we provide an overview of the general steps to simulate the ‘physical’ source fields used in our rupture generator: Δu , V_{peak} and v_{rup}/c_s , as follows [Goovaerts, 1997]:

- (1) Simulate rough fault model following approach outlined in [Shi and Day, 2013], and compute the initial friction using $\mu_0 = |\boldsymbol{\sigma}_s|/\sigma_n$. The scalar shear-traction σ_s is computed as the norm of the vector shear traction $\boldsymbol{\sigma}_s = (\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}) \cdot \boldsymbol{\sigma} \cdot \hat{\mathbf{n}}$ and the normal traction $\sigma_n = \hat{\mathbf{n}} \cdot \boldsymbol{\sigma} \cdot \hat{\mathbf{n}}$. Here, $\mathbf{I}$ is the identity matrix and $\hat{\mathbf{n}}$ is the fault normal vector. We note that in this paper we assume the same background stress

model used in the dynamic rupture computation to estimate μ_0 , although in theory any sensible model could be used.

- (2) Define a simulation area, A , along with a random path traversing A that visits each node u' only once.
- (3) At each u' in A , compute the mean and variance for the Gaussian ccdf $G(u'; y_i | (n))$ using Equations 4.14 and 4.15, respectively. The simple cokriging system is solved using the spatial correlation model defined by the linear model of coregionalization with variogram matrix $\mathbf{r}(\mathbf{h})$ described in the previous section. Here, the conditioning information (n) consists of μ_0 generated in step (1) and all previously simulated $y_i(u')$.
- (4) Sample $G(u'; y_i | (n))$ using the parameters obtained in previous step (Equation 4.12), and add $y_i(u')$ to the data set.
- (5) Advance to the next node and complete the previous two steps. Repeat until all nodes are simulated.
- (6) Apply transformation of the simulated normal-scores into the simulated values $z_i(u')$ using the reverse of the process shown in Equation 4.10. The simulated normal-score $y_i(u')$ and simulated field value $z_i(u')$ share the same cumulative probability.

This framework provides flexibility to improve future versions of the kinematic rupture generator. For example, more sophisticated kriging estimators can be applied to capture specific dynamic effects in addition to different models of the spatial correlation structure. In addition, the conditional simulation approach allows us to include other information, such as a model of Δu constrained by finite-fault inversions or dynamic rupture simulations. In the following sections, we explain how we incorporate this statistical framework to capture the statistics observed in dynamic rupture simulations on rough faults.

4.2 Dynamic Rupture Simulations

Modeling Overview

We used the Support Operator Rupture Dynamics (SORD) code to compute our dynamic rupture simulations. SORD is a 2nd order in space and time finite-difference solution of the elastodynamic equation including a split-node fault boundary condition to allow for dynamic rupture [Ely *et al.*, 2008; 2009]. SORD uses a structured non-regular mesh and incorporates Drucker-Prager plasticity along with both rate-and-state friction and slip-weakening friction models [Shi and Day, 2013]. In total, we computed 108 dynamic rupture simulations to analyze for constructing the KRG.

We computed all ruptures on right-lateral strike-slip faults with mean vertical fault planes. The models in the database range in M_w between 6.4 and 7.2. The variability in our database results from the rough fault geometry and choice of initial model parameters influencing the rupture models. We used the same mesh dimensions for all simulations, but the actual rupture length (and M_w) is determined a posteriori by the rupture process. We impose a maximum rupture length of 45 km, and include a velocity-strengthening zone at the base of the fault limiting the effective fault width to 15 – 17 km. Each simulation uses $dx = 25$ m and has dimensions of 65 km x 40 km x 20 km with 20 node wide perfectly-matched layers [Ely *et al.*, 2008] in order to eliminate any artifacts in the rupture due to boundary reflections.

We include both strongly rate-weakening and slip-weakening friction models in our rupture database along with models incorporating plasticity. The frictional parameters are based on previously published studies (e.g., Shi and Day, [2013] and the SCEC Rupture Validation exercises Harris *et al.*, [2009]). All simulations share the same 1D background velocity model with $\min(V_s) = 1250$ m/s, which is derived from a model representing a hard-rock site in Los Angeles basin. We assume two different models for the general behavior of the initial stress field. One model has linearly increasing normal-stress which is assumed to be lithostatic (i.e., $\tau_n = (\rho_b - \rho_w)gz$, where ρ_b is the bulk density given by the velocity model, ρ_w is the density of water assumed to be 1000 kg/m^3

and g represents acceleration due to gravity). The second model includes depth-dependent lithostatic normal stress to $z = 8 \text{ km}$, and constant normal stress where $z > 8 \text{ km}$; the water table is assumed to be at the surface. This formulation provides ruptures that have stress-drop distributions consistent with observations of earthquakes in Southern California [Shearer *et al.*, 2006]. A physical explanation for this model comes from an increase of pore-pressure with depth, creating a constant effective normal stress. Rupture models include different levels of pre-stress with most models having $\tau_s \approx 0.3$. In all ruptures the maximum compressive stress is set to 45° relative to the mean fault plane.

In the slip-weakening ensembles, we define two-different sets of frictional parameters, each resulting in the same average static stress-drop of $\Delta\sigma = 5 \text{ MPa}$. We use values of the initial friction μ_0 of 0.225 and 0.3 to be consistent with observations of the low magnitude of shear-stress on the San Andreas fault (e.g. [Zoback *et al.*, 1987]), as opposed to favoring Byerlee rock strength. Our choice of parameters in the slip-weakening model results in shear stress on the order of 60 MPa at the hypocentral depth, consistent with previous studies (e.g. [Schmedes *et al.*, 2010]). To prevent bias from this assumption, the parameters of the rate-and-state model incorporate failure consistent with Byerlee rock strength, resulting in large dynamic stress and strength drops.

We define the resulting slip-weakening frictional parameters μ_s and μ_d to obtain an average S ratio of 2.0 over the fault plane. In addition, we chose the D_c parameter to ensure that at least 3 - 5 nodes resolve the breakdown zone at the crack-tip. In all simulations (including rate-and-state), the frictional coefficients remain constant over the velocity-weakening portion of the fault. Even though the distributions of frictional coefficients are constant, the resulting distributions of material strength are heterogeneous due to fault roughness.

For each set of input parameters, we nucleate ruptures on 32 rough faults. We include stochastic roughness down to a minimum wavelength of $\lambda_{min} = 4 \text{ dx}$ or 100 m . This results in equivalent scale lengths for the cohesive zone and roughness wavelength, which is near the resolution limit where the rupture will be able to interact with the

shortest wavelengths. Fault roughness is described using a dimensionless parameter $\alpha = h_{rms}/L$, where L is the length of the fault, and h_{rms} is the RMS roughness of the 2D fault profile. We include faults using $\alpha = [0.005, 0.007, \text{ and } 0.009]$ in our database. The α describing fault roughness represents a different parameter than presented previously in the “Statistical Framework for the Proposed KRG” section.

We follow the approach of *Fang and Dunham*, [2013] to choose the hypocenter locations on each rough fault. For each fault, we smooth the initial friction $\mu_0 = ||\boldsymbol{\tau}_s||/\tau_n$ using a Gaussian kernel with a window-length of 2.5 km, roughly corresponding to the critical radius for spontaneous rupture. We define the hypocenter as the region having smoothed $\max(\mu_0)$, where the fault is closest to failure. Hypocenter locations are chosen between 5 and 10 km depth, and are restricted to being at least 5 km from the edge of the fault. Even though we attempt to simulate 32 separate models for each set of input parameters, the rough fault can prohibit nucleation and spontaneous rupture. In this case, we discard the model and do not include the rupture in our database and subsequent analysis. In addition, we consider only one set of hypocenters for each rough fault realization.

Studies have shown rough-fault simulations can require plasticity to eliminate large concentrations of stresses around geometric complexities on the fault (e.g. *Dunham et al.*, [2011]), especially when using strongly-rate weakening friction laws. For the rate-and-state models, we assume a material cohesion of 5 MPa and an internal friction coefficient $\tan(\phi) = 0.75$ following *Shi and Day*, [2013]. We also include plasticity in one ensemble of slip-weakening models. For these slip-weakening models we use a cohesionless material and internal friction coefficient $\tan(\phi) = 0.45$ to be consistent with the material strength defined by the static friction coefficient $\mu_s = 0.4$ in those models. The cohesionless yield surface prevents tensile stress state on the fault plane, which effectively prevents fault opening (e.g., *Dunham et al.*, [2011]). We note that a cohesionless yield surface likely produces over-estimation of plastic yielding and reduction in ‘physical’ source fields near the free-surface. At depth, this effect becomes minimal.

In the slip-weakening models including plasticity, we do not artificially emulate velocity-strengthening behavior near the free-surface. We allow the reduced elastic parameters from the 1D velocity model along with the plastic deformation to approximate the energy-absorbing behavior of rupture at the free-surface. However, rate-and-state friction allows us to control the rate-weakening behavior of the model. We adjust the a and b parameters to cause velocity-strengthening above $z = -4 \text{ km}$ following *Shi and Day*, [2013]; effectively, only $z > -1.5 \text{ km}$ has parameters $a - b < 0$. Finally, for the elastic slip-weakening models, we emulate velocity-strengthening by adjusting the dynamic coefficient of friction μ_d to be equal to μ_s following *Ma and Andrews*, [2010].

In all of the slip-weakening models, we use a forced-rupture technique to nucleate rupture on the rough faults. The fault is forced to rupture at a fixed velocity of $0.7 V_s$ over a critical distance which depends on the slip-weakening parameters and initial stress. This produces a gradual nucleation when sufficient stress intensity has accumulated at the crack-tip to promote spontaneous rupture. In the rate-and-state models, we cannot force the nucleation using the same approach. Instead, we apply a perturbation in shear stress at the hypocentral location, which is typically 5-10 MPa greater than the peak-strength, τ_p . We calibrate the initial stress perturbation to ensure that unnecessarily large nucleation phases did not contaminate the solution.

Fitting Kinematic Slip-Rate Functions

In addition to knowing the spatial structure of the ‘physical’ fields computed during the dynamic rupture (i.e. Δu , V_{peak} , and v_{rup}/c_s), we need to understand how these ‘physical’ fields relate to the kinematic parameters of the regularized Yoffe function (Equation 4.1). The parameters of the regularized Yoffe function are the peak-time, τ_s , the rise-time, τ_r , the kinematic slip, Δu_{kin} , and the rupture initiation-time, t_0 . We can recover the kinematic parameters at each subfault by fitting the dynamic slip-rate function with the function shown in Equation 4.1. The recovered kinematic parameters Δu_{kin} and $V_{peak_{kin}}$ are different than their dynamic counterparts. Hence, we make the distinction between the two sets of parameters when necessary by using this subscript notation.

Finding the best-fitting regularized Yoffe function for each dynamic slip-rate function requires a straightforward grid-search optimization on each subfault. In our database, every dynamic rupture model has ~ 2 million subfaults ($nx = 2601$, $nz = 801$) and 10001 time steps ($dt = 0.002$ s). Due to the high-frequency behavior of the simulations, that result in small values of τ_s , we have to fit slip-rate functions without downsampling the simulation output. Due to the size of the three-component slip rate function outputs, we determine the kinematic slip-rates for a single ensemble of slip-weakening simulations. In total, we analyzed approximately 4 TB (about 80 million subfaults) of slip-rate functions for determining the kinematic parameters.

We fit the parameters of the regularized Yoffe function in two separate steps. In the first step, we determine parameter values that can be directly computed from the dynamic slip-rate functions. Then, we use these values along with a grid-search optimization routine to estimate the remaining parameters. First, we determine t_0 directly from the dynamic slip-rate function. Following the approach of *Mai et al.*, [2017] and *Schmedes et al.*, [2013] we define a time t_{end} where the slip reaches zero for the first time. We then compute the kinematic slip using $\Delta u_{kin} = \int_0^{t_{end}} \dot{s}_{dyn}(t) dt$, where $\dot{s}_{dyn}(t)$ is the dynamic slip-rate function, which provides better recovery of the dynamic V_{peak} . Using the computed parameters t_0 and Δu_{kin} , we optimize a least-squares cost function to determine the best-fitting values of τ_r and τ_p on each subfault. We note that simply estimating the rise-time using $\tau_r = t_{end} - t_0$ produces an incorrect estimate of τ_r , since $\tau_{r_{eff}} = t_{end} - t_0 = 2\tau_s + \tau_r$ [*Tinti et al.*, 2005], which results from the convolution of the Yoffe function with parameter τ_r and the triangular function with half-length τ_s .

Figure 4.2 shows a typical dynamic slip-rate function with the best fitting regularized Yoffe function. We find that the regularized Yoffe function provides a good approximation to the dynamic slip-rate functions, and recovers the frequency content in the dynamic simulations sufficiently.

Ensemble Overview

We simulated 108 events on rough faults, including different initial parameterizations in an attempt to capture the epistemic uncertainty associated with the dynamic rupture problem. We chose end-member cases of initial conditions to produce rupture models that would, on average, reproduce the characteristics of models previously published in the literature. Also, unless specified otherwise we treat each rupture model as having an equal likelihood of occurrence. In effect, this means we compute a single set of one-point statistics and two-point statistics that are representative of all ruptures in our database and combine assuming equal weight to each simulation.

Our goal is a kinematic rupture generator that produces realistic broadband ground motions, consistent with observations. We compare our dynamic rupture simulations against empirically derived moment-area scaling relationships [*Wells and Coppersmith*, 1994; *Leonard*, 2010] to show that our database produces comparable rupture models with observed natural faults. Figure 4.3 shows the dynamic rupture simulations plotted against two moment-area scaling relationships for intra-plate strike-slip earthquakes. In this analysis, we computed the rupture area by circumscribing the subfaults with $\Delta u > 0$ m with a polygon. In this figure, we distinguish the rupture models by both fault roughness and friction law. Our ruptures produce moment-area relationships above empirically values, but produce agreeable comparisons with the overall scaling trend.

In addition to comparing our database against moment-scaling relationships, we expect source spectra computed from ruptures in our database to follow an approximate ω^{-2} high-frequency decay in spectral behavior described by *Brune*, [1970]. Figure 4.4 shows stacked spectra for the five different ensembles considered in this study. We compute the moment-rate function on the fault using $\dot{M}_0(t) = dx^2 \mu(\mathbf{x}) \int \dot{s}(\mathbf{x}, t) d\Sigma$, where $\dot{s}(x_1, x_2, t)$ is the slip-rate function at point $\mathbf{x}$, $\mu(\mathbf{x})$ is the shear-modulus at $\mathbf{x}$, and dx^2 is the area of the subfault. The integral is computed at each time-step for the duration of the rupture. We normalize all spectra to unit variance before stacking to account for ruptures with different moment magnitudes. We find that all ensembles generally follow

the ω^{-2} trend, implying that far-field acceleration spectra follow an ω^0 trend up to $f_{max} = 10.0 \text{ Hz}$, as expected from observations

4.3 Extracting Statistics for Kinematic Source Generation

Data Preparation

We apply a pre-processing routine to the computed dynamic rupture models before computing any statistics, consistent between all ensembles and applied to each simulation in the ensemble. This process involves eliminating undesirable points that should not be included in the analysis. We eliminate points based on Boolean criteria using a logical masking procedure. First, we exclude all subfaults at depths $z > 15 \text{ km}$ and $z < 4.0 \text{ km}$, representing points that rupture within velocity-strengthening regions on the fault. Next, we mask all points falling within a circle having radius = 4 km and its origin at the hypocenter. Due to the artificial nucleation of dynamic rupture simulations, the points directly affected by the nucleation technique should not be included in the analysis. We also exclude points with $V_{peak} < 0.1 \text{ m/s}$, because we are interested in computing statistics for subfaults that rupture spontaneously and contribute significantly to the moment, M_0 . Finally, we only consider points with sub-shear rupture velocities in this analysis ($v_{rup}/c_s < 1$).

One-point Statistics

The Sequential Gaussian simulation technique operates under the assumption of a multivariate Gaussian random variable. This means that simulated fields values are initially defined as quantiles of a standard normal variable. In order to recover the desired one-point statistics, a normal-score transform must be applied to the simulated data. Several studies have shown that one-point statistics of dynamic rupture models are non-Gaussian. For example, [Schmedes *et al.*, 2010; Song *et al.*, 2013] analyzed ruptures on flat faults and show depth-dependence in the marginal distributions of one-point parameters in addition to a dependency on hypocentral distance. However, a recent study

on the statistics of rough faults by [Yao, 2017] showed that sufficiently rough faults eliminate the dependency on the distance to the hypocenter. We find similar results to [Yao, 2017] in our dynamic rupture database, and therefore do not include any distance dependence of the marginal PDFs (mPDF) in our KRG.

Recently, *Thingbaijam and Mai*, [2016] analyzed a database of finite-fault inversions and suggested that the slip-distribution could be modeled using a truncated exponential distribution. In agreement with this study, we find that the mPDF estimates from our dynamic rupture models also follow truncated distributions. The truncated distributions provide representations of the rupture fields that are consistent with the physical limitations of the rupture process, namely, all rupture fields must be greater than zero. Figure 4.5a shows the marginal distributions estimated when considering all of the dynamic rupture simulations in our database. We find that the estimated density function is approximately symmetric, therefore we cannot account for the truncation at zero using a logarithmic transformation. In addition, we find that mPDFs are generally non-Gaussian, which is consistent with previous studies. Figure 4.5b shows QQ plots based on a standard normal variable. In both Δu and V_{peak} the truncated values are observed at approximated -2.5 quantiles from the mean value. In our KRG, we incorporate an empirical mPDF derived from the dynamic rupture simulations, as opposed to using a general distribution function. Incorporating this type of generalized marginal distribution would be a primary target for future improvements to the KRG.

While the normalized representations of the mPDFs provide insight into the nature of the distributional statistics from the dynamic ruptures, they are insufficient for transforming simulated values into their desired quantities. Figure 4.5c shows the transformation functions used to compute the desired field values from the simulated normal scores. These empirical mPDF functions include the observed truncation in the rupture fields (e.g., *Thingbaijam and Mai*, [2016]), and preserve the order statistics from the dynamic rupture database. After simulating the rupture fields using sequential Gaussian simulation, we transform the k^{th} largest value from the simulated standard normal distribution with the k^{th} largest value observed in the functions shown in Figure

4.5c. The mPDFs are subject to the limitations of the dynamic ruptures computed for our analysis. In other words, we cannot say that these mPDFs would represent the distributional statistics observed for arbitrary rupture mechanisms or magnitude, but they are representative of the $M_w 6.4 - M_w 7.2$ ruptures on strike-slip faults considered in this study.

Before computing the two-point statistics and determining the best-fitting linear model of coregionalization, we need to know the depth-dependency of the rupture parameters. Due to the limitations of simple kriging (i.e., assumption that the mean is known), we only consider regions on the fault that do not display a trend as a function of depth. This allows us to compute residuals using a scalar mean value, and use simple kriging as the mean value is known to be zero for the residual fields. Figure 4.6 shows that between $4 \text{ km} < z < 15 \text{ km}$ the rupture fields display reasonably constant mean as a function of depth. This motivates limiting the region of analysis for determining two-point statistics between $4 \text{ km} < z < 15 \text{ km}$. This result does not necessarily imply that the mPDFs are constant with depth; however, we do not incorporate depth-dependency into the mPDFs in our KRG. We account for the near-surface depth-dependency observed in final slip Δu (Figure 4.6a) by applying a linear tapering above 4 km based on these results and those obtained by *Xu et al.*, [2016] relating to shallow slip deficit. We take guidance from the results of *Xu et al.*, [2016], because incorporating a cohesionless yield surface in the dynamic rupture models likely leads to an over-estimate of the slip reduction at the surface due to plastic yielding. We note that these tapering parameters are adjustable in our KRG implementation.

Two-Point Statistics

Estimating regionalization models poses minor challenges in the multivariate case. These challenges arise from the limitations inherent with linear models of coregionalization where both semivariograms and cross-semivariograms must be described by semivariogram models with equal range parameters. In practice, this results in a trade-off between the semivariogram estimates for direct and cross semivariogram estimates within the linear model of coregionalization, with better fits for certain

semivariograms and worse fits for others. In this section, we outline our approach for obtaining the linear-model of coregionalization used to represent our dynamic rupture database.

The first step in computing the linear-model of coregionalization is estimating the semivariograms and cross-semivariograms from the dynamic rupture fields of interest. The semivariance can be estimated using:

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z(\mathbf{u}_{\alpha} + \mathbf{h}) - z(\mathbf{u}_{\alpha})]^2 \quad 4.16$$

Here, $\mathbf{h} = \mathbf{u} - \mathbf{u}'$ represents a separation vector, $z(\mathbf{u}_{\alpha})$ an observation from the random variable Z at location $\mathbf{u}_{\alpha}$, and $N(\mathbf{h})$ represents the number of observed pairs at the lag distance $\mathbf{h}$. In practice, using Equation 4.16 to estimate the semivariance uses a tolerance called Δ , so that the semivariance estimate at lag distance h includes all points separated by $[h - \Delta, h + \Delta]$. In addition, the factor of two is present in the denominator to account for the double counting of points $z(\mathbf{u}_{\alpha} + \mathbf{h}) - z(\mathbf{u}_{\alpha})$ and $z(\mathbf{u}_{\alpha}) - z(\mathbf{u}_{\alpha} + \mathbf{h})$ in the summation. It is common practice to define a cutoff value for the separation vector $\mathbf{h}$ to remove data points for small $N(\mathbf{h})$ at large distances. In our analysis, we chose a tolerance $\Delta = 100 \text{ m}$ corresponding to 4 grid points in our dynamic rupture simulations, and a variogram cutoff of 6 km based on the distance along dip after eliminating subfaults that ruptured in the velocity-strengthening region.

To determine the variogram model representative of our dynamic rupture database, we first estimate individual semivariograms between all rupture fields for each simulation. We discard ruptures that do not produce stable variogram estimates, typically caused by models with insufficient rupture area. We combine the semivariogram estimates for each simulation by stacking with equal weight given to each simulation. This provides an estimate of the spatial correlation structure that, on average, represents the entire database without discriminating by a particular set of dynamic rupture input parameters. Although this provides a means to discriminate between different rupture

types, we intend that our approach prevents the KRG from becoming biased by a particular set of input parameters.

We estimate the preferred model of linear-coreionalization through an iterative trial and error procedure to determine the optimal range parameters for the long-range and short-range semivariogram models. We first fit the range parameters of the correlation model defined by:

$$\Gamma(\mathbf{h}) = B^1 \left(1 - \exp \left(-\frac{3\mathbf{h}}{a_s} \right) \right) + B^2 \left(1 - \exp \left(-\frac{3\mathbf{h}}{a_l} \right) \right) \quad 4.17$$

where a_s and a_l represent the short-scale and long-scale range parameters, respectively. This does not represent a linear model of coreionalization, but provides insight into the scale length present in the data. We find that length scales for a_s on the order of hundreds of meters and a_l is on the order of a few kilometers. These scale lengths are similar to those of the breakdown zone and the crack length of the rupture models. Through trial and error, we settled on Equation 4.18:

$$\Gamma(\mathbf{h}) = B^1 \left(1 - \exp \left(-\frac{3\mathbf{h}}{250} \right) \right) + B^2 \left(1 - \exp \left(-\frac{3\mathbf{h}}{5000} \right) \right) \quad 4.18$$

Figure 4.7 shows the best fitting linear model of coreionalization fit using range parameters $a_s = 250 \text{ m}$ and $a_l = 5000 \text{ m}$. We chose range parameters that best represent both the direct semivariograms in addition to the cross semivariogram with μ_0 . We define the nested exponential model shown using Equation 4.18 which provides the simplest model that represents our semivariogram estimates. The estimated spatial correlation structure from our dynamic ensemble is shown using the blue circles. The black line depicts our preferred linear model of coreionalization, which we explain below. For μ_0 and v_{rup} the semivariance seems to be dominated by short scale-length processes, and Δu and V_{peak} seem to be controlled by long-scale processes. We attribute the small-scale length process to behavior on the scale length of the breakdown zone, which can explain the small-scale lengths for v_{rup} . The longer correlation length of Δu

and V_{peak} is likely controlled by the crack-length or the pulse-width of the dynamic rupture simulations.

Tables 4.1 and 4.2 show the coefficients of the coregionalization matrices $\mathbf{B}^l$ for the short-range and long-range components, respectively. Notice that we do not include a nugget effect in this model. In nearly all the direct and cross semivariogram estimates the value of the semivariogram is near zero at zero lag. Only v_{rup} displays a significant effect at $h = 0$. We note that we do not consider anisotropy in our linear-model of coregionalization, however *Mai and Beroza, [2002]* find that large strike-slip events display anisotropy, likely due to the aspect ratio of long rectangular faults.

We can compute the zero-offset correlation matrix at $\mathbf{h} = 0$ based on our linear-model of coregionalization using the relationship $\mathbf{C}(0) = \lim_{h \rightarrow +\infty} \mathbf{\Gamma}(h) = \mathbf{B}^1 + \mathbf{B}^2$, where $\mathbf{B}^l$ are the coregionalization matrices. Table 4 shows the zero-offset correlation from our linear model of coregionalization. We present normalized coregionalization matrices so that the diagonal components of $\mathbf{C}(0)$ equal unity, allowing us to compare our spatial correlation model with the results from *Yao, [2017]*. We observe agreement between correlations between source parameters. In addition, our values all fall within the distributions of correlation coefficients found in that study. We believe that our preferred linear model of coregionalization reproduces the general features of our dynamic rupture simulations. *Trugman and Dunham, [2014]* found much larger correlations between the source fields. However, we believe the differences are due to the more complex rupture behavior in 3D as opposed to the 2D simulations used in their study. In 3D, the fault has the ability to rupture around an asperity whereas a 2D rupture cannot. This likely explains the lower correlations with μ_0 . In the following section, we explain the relationships used to derive the kinematic source parameters from the ‘physical’ source fields simulated using our model of spatial statistics.

Empirical Relationships for Kinematic Parameters

Based on our approach to simulate spatial fields Δu , V_{peak} , and v_{rup}/c_s conditioned using a distribution of μ_0 we now estimate the parameters of the regularized

Yoffe function, namely, the rupture initiation time, t_0 , the peak-time τ_s , and the rise-time τ_r , in order to produce a complete kinematic slip model.

Rupture Initiation Time. When computing t_0 , we first scale the normalized v_{rup}/c_s by the chosen shear-wave velocity model of the ground motion simulation to produce v_{rup} (m/s). After determining v_{rup} using the chosen shear-wave velocity model c_s , we define an arbitrary hypocenter location on the fault plane. Using a fast-marching solution of the Eikonal equation, we convert the simulated v_{rup} into rupture initiation times t_0 for each subfault. Next, we show how to derive the peak time τ_s using the simulated V_{peak} .

Peak Time. Using laboratory experiments, *Ohnaka and Yamashita*, [1989] proposed the relationship $V_{peak} \propto D_c f_{max}^s$, where D_c is the critical slip-weakening distance, and f_{max}^s represents the cutoff frequency of the slip-rate function. In addition, *Tinti et al.*, [2005; 2009] confirmed that this relationship holds true for the regularized Yoffe function. We can write this relationship in terms of the peak-time τ_s by knowing that f_{max}^s is inversely related to the time t_c required for the crack-tip to break down, which results in $f_{max}^s \approx 1/t_c$ [*Ohnaka and Yamashita*, 1989]. From the data in Table 1 of *Tinti et al.*, [2005], we computed that $t_{acc} \approx 0.85t_c$, and *Tinti et al.*, [2005] showed that $t_{acc} \approx 1.3\tau_s$; therefore, we can say that $t_c \approx 1.5\tau_s$. Combining these relationships, we can write:

$$\frac{V_{peak}}{d'_0} = \beta_0 f_{max}^s, \quad 4.19$$

where we include the factor of 1.5 in the coefficient β_0 . Using Equation 4.19 we can find the value of d'_0 providing, on average, the desired f_{max}^s of the kinematic source model:

$$\overline{d'_0} = \frac{\overline{V_{peak}}}{\beta_0 f_{max}^s}. \quad 4.20$$

Equation 4.20 provides the d'_0 that produces the average chosen cutoff frequency f_{max}^s for the kinematic source model, where the overbar denotes the expected value. This allows an estimate of τ_s by:

$$\tau_s = \beta_1 \frac{\overline{d'_0}}{V_{peak}}. \quad 4.21$$

Notice, we replaced D_c with d'_0 in Equations 4.19-4.21. We make the change in notation, because we implement the ability for the user to control the frequency content of the slip-rate function. Effectively, $\overline{d'_0}$ becomes the parameter used to control the maximum frequency content of the kinematic source model. Even though the D_c of our dynamic rupture simulations are constant across the fault plane, the differences in Δu and Δu_{kin} is described by the parameter d'_0 on the fault plane. This means in order to solve for β_0 and β_1 we must determine an estimate of d'_0 from our dynamic rupture simulations for Equations 4.19-20. In theory, the ratio of coefficients β_0/β_1 are related and we can express β_1 in terms of β_0 using $\beta_1 = \beta_0^{-1}/1.5$ because both Equation 4.19 and 4.21 are derived from the relationship $V_{peak} \propto D_c f_{max}^s$; however, we find that due to the approximations made in these relationships the estimated coefficients differ in our regression models.

Tinti et al., [2005] showed that we can relate the kinematic parameters τ_s , τ_r , and Δu to d'_0 using:

$$d'_0 \approx \Delta u \sqrt{\frac{1.3\tau_s}{\tau_r}} \quad 4.22$$

Using Equation 4.21 we estimate d'_0 for 16 simulations selected from our ensemble. The 16 chosen simulations share a dynamic $D_c = 0.16 \text{ m}$ and we recover an average value of $d'_0 = 0.13 \text{ m}$ using Equation 4.22. Using this estimate of d'_0 , we compute the coefficients $\beta_0 \approx 1.25$ and $\beta_1 \approx 0.6$ by performing a least-squares estimate on Equations 4.19 and 4.21, respectively. We do not include a y-intercept term in these regression models.

Here, we summarize the process to estimate the variable τ_s on the fault given the simulated spatial field V_{peak} : (1) assume a target f_{max}^s for the kinematic source model and compute the ratio of V_{peak}/d'_0 corresponding to the chosen f_{max}^s using Equation 4.19 (Figure 4.8a) with $\beta_0 = 1.25$; (2) using the average V_{peak} from the simulated spatial field, compute the corresponding value of d'_0 ; (3) compute the value of τ_s for each subfault using the value of the effective V_{peak} simulated on that subfault and d'_0 computed in step 2.

In order to prevent unrealistically large values of τ_s , we have implemented some empirical rules to define the effective V_{peak} derived from the dynamic rupture simulations. First, we limit the ratio between simulated Δu and V_{peak} to 2 to eliminate areas on the fault with large Δu , but low V_{peak} . Next, we limit the $\min(V_{peak})$ to 0.1 m/s to prevent unrealistically large values of τ_s . The choices of these rules are largely arbitrary, but produce distributions of τ_s consistent with those computed from the dynamic rupture simulations.

Rise Time. Finally, we need to calculate τ_r before we can generate our finite-fault kinematic slip model. [Gattereri et al., 2004] proposed that the effective rupture duration t_{dur} relates linearly to the total slip duration on a given subfault (τ_{reff}). In our dynamic rupture models, $\tau_s \ll 1$, indicating that $\tau_r \approx \tau_{reff}$, so the relationship proposed for the triangular source time function used by Gattereri et al., [2004] should apply here. In addition, we know that Δu correlates with τ_r from previous studies (e.g., Schmedes et al., [2010]; Yao, [2017]). We construct a multiple linear model using final-slip Δu and effective rupture duration t_{dur} as predictors to estimate τ_r . This relationship is defined using:

$$\tau_r = \beta_0 \Delta u + \beta_1 t_{dur} \quad 4.23$$

where the coefficients β_0 and β_1 are found to be 3.55 and 0.08, respectively. We also investigated the possibility of an interaction between Δu and t_{dur} , but do not obtain passable models for predicting τ_r , when including the interaction term in Equation 4.23.

Figure 4.9 shows the predictions of this model compared with the results from dynamic rupture simulations. We show scatter plots for each predictor (Δu and t_{dur}) plotted against τ_r by fixing the value of the non-plotted predictor to its mean value.

According to *Tinti et al.*, [2005] there exists an asymptotic relationship between the kinematic parameters and V_{peak} stating that:

$$V_{peak} \propto \frac{\Delta u}{\sqrt{t_{acc}}\sqrt{\tau_r}}. \quad 4.24$$

We find that when using Equation 4.24 to compute τ_r we can obtain unrealistically large values of τ_r due to the small values of τ_s associated with $f_{max}^s \approx 10 \text{ Hz}$. As this relationship does not incorporate information about rupture duration, the estimate of τ_r using Equation 4.24 results in a majority of subfaults having τ_r much greater than the finite-fault rupture duration. Because Equation 4.23 reproduces τ_r values observed from the dynamic rupture database, we do not implement Equation 4.24 in our KRG. In practice, this results in a distribution of V_{peak} that may be different than the V_{peak} simulated using the sequential simulation algorithm. We note that this modification only affects subfaults having large ratios of $\Delta u/V_{peak}$, and it seems sensible to increase V_{peak} in areas with large Δu due to the observed correlations between Δu and V_{peak} (e.g., *Yao*, [2017]).

Finally, t_{dur} must be specified in order to compute τ_r using Equation 4.23. We define t_{dur} for a simulation by using the average value of t_0 on the fault boundary. This parameter can be chosen automatically or chosen by the user. Subfault locations that have negative t_{dur} (i.e. subfaults that rupture after the chosen t_{dur}) are not permitted to rupture by setting $\Delta u = 0$.

Sample Kinematic Rupture Realization

Figure 4.10 shows an example kinematic source generated using the methodology described in this paper. We can see the correlation between τ_r and Δu in addition to the anti-correlation between V_{peak} and peak-time τ_s . In addition, the variability in v_{rup}/c_s

results in the rough distribution of rupture initiation times t_0 . In the next section, we present ground motions computed using finite-fault sources generated using this methodology.

4.4 Validating the KRG

The validation procedure can be considered an impossible task for natural systems [Oreskes *et al.*, 1994], but nonetheless it is imperative to display that models provide useful and sensible results. In this context, we are interested in understanding the usefulness of this KRG implementation to generate broadband seismic sources that produce reasonable ground motion amplitudes and variability. We constrain the validation of the KRG to ground motions within 25 km from the finite-fault where effects from anelastic attenuation or scattering are limited.

We validate the ground motions produced using sources from the kinematic rupture generator described in this paper by comparing 10 Hz deterministic simulations against leading GMPEs (e.g., *Abrahamson et al.*, [2014]; *Boore et al.*, [2014]; *Campbell and Bozorgnia*, [2014]; *Chiou and Youngs*, [2014]) up to a period of 0.2 seconds. In addition to GMPE comparisons, we show visualizations of far-field acceleration and displacement spectra in addition to seismograms extracted at select stations around the finite-fault.

We computed ground motions using the 4th order space 2nd order time staggered-grid finite difference code AWP-ODC-CPU [Cui *et al.*, 2010; 2013] including frequency-dependent attenuation [Withers *et al.*, 2015] and perfectly-matched layers [Marcinkovich and Olsen, 2003] using 20 nodes to prevent reflected energy from contaminating the solution. We refer the interested reader to Cui *et al.*, [2010; 2013] for an in-depth explanation of the numerical method.

We simulate 40.0 seconds of 0-10 Hz deterministic wave propagation in a 90 km x 50 km x 25 km domain discretized using $dx = 25\text{ m}$. The simulations run for approximately 3.5 hours using 7200 processors on ORNL Titan. We incorporate a 1D velocity profile with $\min(V_s) = 1250\text{ m/s}$ representative of a hard-rock site in southern California (Figure 4.11), resolving frequencies up to 10 Hz with a minimum wavelength $\lambda_{min} = 125\text{ m}$ using at least 5 points per wavelength. To model anelastic attenuation, we incorporate a linear-relationship between Q_s and V_s where $Q_{s0}/V_s = 100$ for V_s given in km/s for frequencies less than 1 Hz. For $f > 1\text{ Hz}$, we model frequency-dependent attenuation using a power-law with exponent $\gamma = 0.6$. We express this frequency-dependent attenuation relationship using:

$$Q(f) = \begin{cases} Q_0 \left(\frac{f}{f_0}\right)^\gamma, & f \geq f_0 \\ Q_0, & f < f_0 \end{cases} \quad 4.25$$

where $f_0 = 1.0\text{ Hz}$, and $Q_0 = 100 V_s$. Equation 4.25 provides the shear-wave quality factor Q_s , and we represent the P-wave quality factor as $Q_p = 2Q_s$.

Kinematic Source Models for AWP-ODC

We generate five finite-fault kinematic slip models using the rupture generator methodology described above, and insert into AWP-ODC via moment-rate components. The strike ϕ_s and dip δ can be computed directly from the rough fault model used to generate the distribution of initial friction μ_0 . Given the fault normal vector $\hat{\mathbf{n}}$ for a subfault, we compute the dip δ as the angle between $\hat{\mathbf{n}}$ and the surface projection ($\hat{\mathbf{n}}_{proj}$) of $\hat{\mathbf{n}}$ using $\delta = 90 - \cos^{-1}(\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}_{proj})$, with $\delta = 90^\circ$ corresponding to vertical. In addition, we compute strike using $\phi_s = \cos^{-1}(\hat{\mathbf{n}}_{proj} \cdot \hat{\mathbf{x}})$, where $\hat{\mathbf{x}}$ denotes the unit vector depicting the Cartesian axis. Lastly, the rake λ is defined using $\lambda = \cos^{-1}(\boldsymbol{\phi}_s \cdot \mathbf{u})$ where $\boldsymbol{\phi}_s$ denotes the strike-vector and $\mathbf{u}$ denotes the slip-vector on the fault. In other words, λ represents the angle between fault-strike and the slip vector.

Since the dip and strike can be directly computed from the rough fault model, we need to define the fault rake in order to generate moment-rates. *Mai et al.*, [2017] proposed a method to estimate λ using an anti-correlated random field with δ ; however, we could not recover their findings in our dynamic rupture database. This is likely due to the sensitivity of λ to the orientation of the background stress field to the fault plane in the dynamic rupture simulations. From our dynamic rupture simulations, we find that λ largely follows ϕ_s with only 10% of subfaults showing $|\lambda - \phi_s| > 1^\circ$, and thus assign $\lambda = \phi_s$.

Figure 4.12 shows the scalar slip on the fault plane for the five source models considered in the validation. The variation in M_w arises due to the fact that each individual realization from the sequential simulation technique does not necessarily have zero-mean, even though an ensemble of realizations would. We generate Δu using the one-point distribution functions shown in Figure 4.5c. We assign median values to our kinematic sources of $\Delta u = 0.81 \text{ m}$, $V_{peak} = 1.51 \text{ m/s}$, and $v_{rup}/c_s = 0.72$, based on dynamic rupture simulations providing average moment-scaling relationships. In all of the simulations we project the kinematic model onto a flat fault, but preserve the moment-tensor orientation of the rough fault. This follows the approach of *Mai et al.*, [2017] who showed that projecting moment-rates onto a flat fault produces consistent ground motions as long as the moment-tensor components reflect the geometry of the rough fault.

Ground Motion Validation

Figures 4.13 and 4.14 show three-component 0-10 Hz velocity seismograms computed for stations T3 and B3 (see Fig. 4.11b), respectively, 15 km from the epicenter in the along strike direction and 5 km from the fault plane. These figures show that seismograms originating from KRG sources produce ground motions that pass the ‘eyeball’ test. In other words, these synthetic records could be mistaken for recorded broadband ground motions without any other context.

Figure 4.15 shows normalized Fourier amplitude of displacement computed at the six stations shown in Figure 4.11b for Source 3. We show that our rupture generator produces far-field displacement spectra showing ω^{-2} high-frequency decay up to the simulated $f_{max} = 10 \text{ Hz}$, consistent with theoretical considerations for shear failure on faults (e.g. [Brune, 1970]). Figure 4.16 shows the corresponding far-field fault-normal acceleration, showing amplitudes that increase up to a corner frequency f_c and display flat spectra to $f_{max} = 10 \text{ Hz}$ (following the ω^{-2} source model). This behavior has been observed in natural earthquakes (e.g., records from the 1980 Mexicali earthquake shown by *Anderson and Hough*, [1984]) and published dynamic rupture simulations on rough faults (e.g., *Dunham et al.*, [2011]; *Shi and Day*, [2013]). We find identical behavior for the high-frequency decay in far-field spectra from the other four kinematic source models considered here.

We have shown that ground motions display qualitative characteristics representative of real earthquakes, such as flat acceleration spectra and sensible synthetic seismograms, but it is important to verify whether the simulated amplitudes and variability are comparable with spectral accelerations predicted by GMPEs. Figure 4.17 shows a comparison between 5%-damped GMRotD50 spectral accelerations [Boore, 2006] for source models 1, 2, and 3 and GMPEs derived from the NGA West2 database (i.e., *Abrahamson et al.*, [2014]; *Boore et al.*, [2014]; *Campbell and Bozorgnia*, [2014]; *Chiou and Youngs*, [2014]) for periods between 2.0 and 0.2 seconds at stations located on a 500 m x 500 m grid (see Figure 4.11b). We find that our source models produce comparable fits with the GMPE medians and display ground motions that fall within $\pm 1 \sigma$.

Finally, we compute GMPE median bias plots as a function of distance for all five kinematic source models. The bias is computed as the natural logarithm of the ratio between GMPE and median GMRotD50 spectral acceleration in 500 m distance bins between 0 and 25 km. Figure 4.18 shows the bias plots for the five periods between 0.2 and 2 seconds. We find that for all periods, our simulation medians fall within 1σ of the GMPE estimates. Additionally, a noteworthy feature from this comparison is that our

simulations generally match the observed variability provided by the GMPEs, indicating that our rupture generator can produce ground motions having variability consistent with observed earthquakes.

4.5 Discussion

The GMPE comparison by no means represents a systematic validation of the rupture generator; however, it provides a promising indication this model could become a useful tool to generate broadband (10 Hz) seismograms for seismic hazard evaluation. At high frequencies, effects from small-scale velocity heterogeneities (e.g., *Imperator and Mai*, [2012]; *Bydlon and Dunham*, [2015]) and non-linear effects (e.g., *Roten et al.*, [2017]) become increasingly important. We suspect this KRG could be used to understand the trade-offs between complexity in both the source and the medium, and the interactions of intrinsic and scattering parameters. For example, the choice of power-law exponent γ in the $Q(f)$ model could change when considering small-scale heterogeneities. In addition, this model could be compared directly against broadband calculations using hybrid methods (e.g., *Graves and Pitarka*, [2015]; *Olsen and Takedatsu*, [2015]).

The implementation of this rupture generator is modular and flexible in the sense that nearly every aspect of the method can be easily improved or modified when new information becomes available. For instance, our choice to separate the ‘physical’ source fields from estimating kinematic parameters allows for a straightforward incorporation of alternative source-time functions. In addition, the marginal distributions or two-point statistics could also be modified to reflect future analyses. Even the simulation technique itself could be replaced. This allows for a rupture generator that can evolve along with new information and understanding of the rupture process.

Despite promising comparisons with GMPEs there are some limitations of this KRG from a practical standpoint. The most urgent improvement of the KRG is the ability to generalize the empirical cumulative distribution function for the simulated Δu . In the KRG’s current form, the target M_w can be only chosen through specifying a fault

dimension, as the empirical distribution functions for Δu , V_{peak} , and v_{rup}/c_s are fixed. In practice, as displayed by the five simulations shown here, the target M_w will vary depending on the result of the sequential simulation. This is caused by the fact that the sequential simulation method does not guarantee that the fields of a single realization will have zero mean, even though the random variable has zero-mean. A recent study analyzing a database of finite-fault inversions [Thingbaijam and Mai, 2016] suggests that a truncated Exponential distribution best describes the distributional statistics of earthquake slip. Using this distribution function, we could map the standard normal quantiles to the values of an exponential distribution function. This would allow the user to specify the mean slip for a given rupture scenario, but would not fix the limitation that the simulated fields are not required to have zero-mean.

Previous rupture generators implement power-law or von Karman correlation functions to describe the autocorrelation structure. For example, Mai and Beroza, [2002] find that von Karman autocorrelation functions best represent that spatial statistics of observed slip models. In future work, one could introduce more complex spatial correlation models that can account for some of the misfit observed in Figure 4.7. With our KRG implementation it is straightforward to incorporate more complex linear models of coregionalization.

Since we only simulated dynamic rupture models on strike-slip faults in a relatively limited magnitude range (M_w 6.5 – M_w 7.2) our rupture generator should only be used to simulate strike-slip sources in this magnitude range. Based on the fault dimensions in our rupture database there are data representing faults from 10 km x 10 km up to 65 km x 15 km. Previous studies have shown that dipping faults behave like strike-slip faults (e.g., Schmedes *et al.*, [2013]), in a statistical sense, but we recommend to confirm that relationship on rough faults before applying this model in that context. Lastly, scaling relationships should be adopted to expand this framework to various fault lengths.

4.6 Conclusions

We have developed a new kinematic rupture generator based on the statistics of dynamic simulations on geometrically complex faults to produce broadband 10 Hz ground motions. We define a random variable model relating the initial friction on the rough fault μ_0 to spatial fields representing ‘physical’ source parameters (Δu , V_{peak} , and v_{rup}/c_s) from the dynamic rupture simulation. Using over 100 dynamic rupture simulations, we fit a linear model of coregionalization using two nested exponential variogram functions with range parameters of 250 m and 5000 m to capture the scale-lengths’ dominating behavior at the breakdown zone and along the slip-path, respectively (Equation 4.18). In addition, we define empirical functions relating standard normal quantiles to the one-point statistics seen in the dynamic rupture simulations. We use a sequential simulation technique to simulate realizations of this 4D random variable conditioned on an initial distribution of μ_0 .

The slip-evolution for each subfault is defined using a regularized Yoffe function [Tinti *et al.*, 2005], which requires the kinematic parameters τ_s , τ_r , t_0 , and Δu . Using the dynamic rupture models, we define empirical relationships to relate the ‘physical’ fields simulated from the 4D random variable with the kinematic parameters of the regularized Yoffe function. We estimate τ_s , which controls the time of the positive slip acceleration, using a two-step method. First, we introduce a parameter d'_0 which allows the modeler to define the upper cutoff frequency f_{max}^s of the individual slip-rate functions using the average V_{peak} of the kinematic source model (Equation 4.20). Using this value of d'_0 , we compute the value of τ_s on each subfault using Equation 4.21. To estimate τ_r we use a multiple linear model with predictors final slip Δu and effective rupture duration t_{dur} (Equation 4.22). Finally, the normalized rupture velocity v_{rup}/c_s is converted to v_{rup} using a model of shear-wave velocities. We solve the Eikonal equation using the fast-marching method to convert v_{rup} into rupture initiation times t_0 .

We computed 10 Hz deterministic ground motion simulations using five source models generated using the KRG implementation presented in this paper. We validate our

ground motions to 0.2 seconds against NGA West 2 GMPEs up to distances of 25 km. We find that our KRG produces ground motions that show promising comparisons with the GMPE medians as well as intra-event standard deviations from the GMPEs. In addition, our rupture generator produces flat acceleration across the desired model bandwidth and produces seismograms that qualitatively appear similar to recorded ground motions. Based on these comparisons, we believe that our rupture generator will help provide insight into the broadband behavior of future earthquakes. Future validation using this rupture generator should incorporate small-scale media heterogeneities and non-linear material response in order to capture other physical mechanisms involved in the wave propagation problem, and assess the interdependency between these parameters and the resulting ground motions.

References

- 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, doi:10.1193/070913EQS198M.
- Aki, K., and P. G. Richards (2002), *Quantitative Seismology*, 2nd ed., edited by J. Ellis.
- Anderson, J. G., and S. Hough (1984), A model for the shape of the Fourier amplitude spectrum of acceleration at high frequencies, *Bulletin of the Seismological Society of America*, 74(5), 1969–1993.
- Andrews, D. J. (1980), A Stochastic Fault Model 1. Static Case, *J. Geophys. Res.*, 85(NB7), 3867–3877.
- Boore, D. M. (2006), Orientation-Independent Measures of Ground Motion, *Bulletin of the Seismological Society of America*, 96(4A), 1502–1511, doi:10.1785/0120050209.
- 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 Engineering Research Institute.
- Brune, J. N. (1970), Tectonic stress and the spectra of seismic shear waves from earthquakes, *Journal of Geophysical Research: Solid Earth*, 75(26), 4997–5009.
- Bydlon, S. A., and E. M. Dunham (2015), Rupture dynamics and ground motions from earthquakes in 2-D heterogeneous media, *Geophys. Res. Lett.*, 42(6), 1701–1709, doi:10.1002/2014GL062982.

- 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 Engineering Research Institute.
- Candela, T., F. Renard, Y. Klinger, K. Mair, J. Schmittbuhl, and E. E. Brodsky (2012), Roughness of fault surfaces over nine decades of length scales, *Journal of Geophysical Research*, 117(B8), B08409–n/a, doi:10.1029/2011JB009041.
- Chiou, B. S. J., and R. R. Youngs (2014), 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, doi:10.1193/072813EQS219M.
- Christakos, G. (1984), On the Problem of Permissible Covariance and Variogram Models, *Water Resources Research*, 20(2), 251–265, doi:10.1029/WR020i002p00251.
- Cui, Y. et al. (2010), Scalable Earthquake Simulation on Petascale Supercomputers, pp. 1–20, IEEE.
- Cui, Y. et al. (2013), *Physics-based seismic hazard analysis on petascale heterogeneous supercomputers*, ACM, New York, New York, USA.
- Dunham, E. M., D. Belanger, L. Cong, and J. E. Kozdon (2011), Earthquake Ruptures with Strongly Rate-Weakening Friction and Off-Fault Plasticity, Part 2: Nonplanar Faults, *Bulletin of the Seismological Society of America*, 101(5), 2308–2322, doi:10.1785/0120100076.
- Ely, G. P., S. M. Day, and J.-B. Minster (2008), A support-operator method for viscoelastic wave modelling in 3-D heterogeneous media, *Geophysical Journal International*, 172(1), 331–344, doi:10.1111/j.1365-246X.2007.03633.x.
- Ely, G. P., S. M. Day, and J.-B. Minster (2009), A support-operator method for 3-D rupture dynamics, *Geophysical Journal International*, 177(3), 1140–1150, doi:10.1111/j.1365-246X.2009.04117.x.
- Fang, Z., and E. M. Dunham (2013), Additional shear resistance from fault roughness and stress levels on geometrically complex faults, *Journal of Geophysical Research: Solid Earth*, 118(7), 3642–3654, doi:10.1002/jgrb.50262.
- Goovaerts, P. (1997), *Geostatistics for Natural Resources Evaluation*, Oxford University Press, New York, New York.
- Graves, R. W. et al. (2011), CyberShake: A Physics-Based Seismic Hazard Model for Southern California, *PAGEOPH*, 168(3-4), 367–381, doi:10.1007/s00024-010-0161-6.
- Graves, R. W., and A. Pitarka (2010), Broadband Ground-Motion Simulation Using a

- Hybrid Approach, *Bulletin of the Seismological Society of America*, 100(5A), 2095–2123, doi:10.1785/0120100057.
- Graves, R. W., and A. Pitarka (2015), Refinements to the Graves and Pitarka (2010) Broadband Ground-Motion Simulation Method, *Seismological Research Letters*, 86(1), 75–80, doi:10.1785/0220140101.
- Guatteri, M., P. M. Mai, and G. C. Beroza (2004), A Pseudo-Dynamic Approximation to Dynamic Rupture Models for Strong Ground Motion Prediction, *Bulletin of the Seismological Society of America*, 94(6), 2051–2063, doi:10.1785/0120040037.
- Harris, R. A. et al. (2009), The SCEC/USGS Dynamic Earthquake Rupture Code Verification Exercise, *Seismological Research Letters*, 80(1), 119–126, doi:10.1785/gssrl.80.1.119.
- Imperator, W., and P. M. Mai (2012), Broad-band near-field ground motion simulations in 3-dimensional scattering media, *Geophysical Journal International*, 192(2), 725–744, doi:10.1093/gji/ggs041.
- Klimeš, L. (2002), Correlation functions of random media, *PAGEOPH*, 159(7-8), 1811–1831.
- Leonard, M. (2010), Earthquake Fault Scaling: Self-Consistent Relating of Rupture Length, Width, Average Displacement, and Moment Release, *Bulletin of the Seismological Society of America*, 100(5A), 1971–1988, doi:10.1785/0120090189.
- Ma, S., and D. J. Andrews (2010), Inelastic off-fault response and three-dimensional dynamics of earthquake rupture on a strike-slip fault, *Journal of Geophysical Research*, 115(B4), B04304, doi:10.1029/2009JB006382.
- Mai, P. M., and G. C. Beroza (2002), A spatial random field model to characterize complexity in earthquake slip, *Journal of Geophysical Research: Solid Earth*, 107(B11), ESE 10–1–ESE 10–21, doi:10.1029/2001JB000588.
- Mai, P. M., M. Galis, K. K. S. Thingbaijam, J. C. Vyas, and E. M. Dunham (2017), Accounting for Fault Roughness in Pseudo-Dynamic Ground-Motion Simulations, *PAGEOPH*, 12(5679), 1–32, doi:10.1007/s00024-017-1536-8.
- Marcinkovich, C., and K. B. Olsen (2003), On the implementation of perfectly matched layers in a three-dimensional fourth-order velocity-stress finite difference scheme, *Journal of Geophysical Research*, 108(B5), 2276, doi:10.1029/2002JB002235.
- Ohnaka, M., and T. Yamashita (1989), A Cohesive Zone Model for Dynamic Shear Faulting Based on Experimentally Inferred Constitutive Relation and Strong Motion Source Parameters, *J. Geophys. Res.*, 94(B4), 4089–4104.

- Olsen, K. B. et al. (2009), ShakeOut-D: Ground motion estimates using an ensemble of large earthquakes on the southern San Andreas fault with spontaneous rupture propagation, *Geophys. Res. Lett.*, 36(4), L04303, doi:10.1029/2008GL036832.
- Olsen, K. B., and R. Takedatsu (2015), The SDSU Broadband Ground-Motion Generation Module BBtoolbox Version 1.5, *Seismological Research Letters*, 86(1), 81–88, doi:10.1785/0220140102.
- Olsen, K. B., S. M. Day, and C. R. Bradley (2003), Estimation of Q for Long-Period (>2 sec) Waves in the Los Angeles Basin, *Bulletin of the Seismological Society of America*, 93(2), 627–638, doi:10.1785/0120020135.
- Olsen, K. B., S. M. Day, J. B. Minster, Y. Cui, A. Chourasia, D. Okaya, P. Maechling, and T. Jordan (2008), TeraShake2: Spontaneous Rupture Simulations of Mw 7.7 Earthquakes on the Southern San Andreas Fault, *Bulletin of the Seismological Society of America*, 98(3), 1162–1185, doi:10.1785/0120070148.
- Oreskes, N., K. Shraderfrechette, and K. Belitz (1994), Verification, Validation, and Confirmation of Numerical-Models in the Earth-Sciences, *Science*, 263(5147), 641–646.
- Roten, D., K. B. Olsen, and S. M. Day (2017), Off-fault deformations and shallow slip deficit from dynamic rupture simulations with fault zone plasticity, *Geophys. Res. Lett.*, 44(15), 7733–7742, doi:10.1002/2017GL074323.
- Roten, D., K. B. Olsen, S. M. Day, Y. Cui, and D. Faeh (2014), Expected seismic shaking in Los Angeles reduced by San Andreas fault zone plasticity, *Geophys. Res. Lett.*, 41(8), 2769–2777, doi:10.1002/2014GL059411.
- Schmedes, J., R. J. Archuleta, and D. Lavalley (2010), Correlation of earthquake source parameters inferred from dynamic rupture simulations, *Journal of Geophysical Research*, 115(B3), B03304, doi:10.1029/2009JB006689.
- Schmedes, J., R. J. Archuleta, and D. Lavalley (2013), A kinematic rupture model generator incorporating spatial interdependency of earthquake source parameters, *Geophysical Journal International*, 192(3), 1116–1131, doi:10.1093/gji/ggs021.
- Shearer, P. M., G. A. Prieto, and E. Hauksson (2006), Comprehensive analysis of earthquake source spectra in southern California, *Journal of Geophysical Research*, 111(B6), B06303–n/a, doi:10.1029/2005JB003979.
- Shi, Z., and S. M. Day (2013), Rupture dynamics and ground motion from 3-D rough-fault simulations, *Journal of Geophysical Research: Solid Earth*, 118(3), 1122–1141, doi:10.1002/jgrb.50094.
- Song, S. G. (2015), Developing a generalized pseudo-dynamic source model of Mw6.5–

- 7.0 to simulate strong ground motions, *Geophysical Journal International*, 204(2), 1254–1265, doi:10.1093/gji/ggv521.
- Song, S. G., L. A. Dalgner, and P. M. Mai (2013), Pseudo-dynamic source modelling with 1-point and 2-point statistics of earthquake source parameters, *Geophysical Journal International*, 196(3), ggt479–1786, doi:10.1093/gji/ggt479.
- Thingbaijam, K. K. S., and P. M. Mai (2016), Evidence for Truncated Exponential Probability Distribution of Earthquake Slip, *Bulletin of the Seismological Society of America*, 106(4), 1802–1816, doi:10.1785/0120150291.
- Tinti, E., E. Fukuyama, and A. Piatanesi (2005), A kinematic source-time function compatible with earthquake dynamics, *Bulletin of the Seismological Society of America*, 95(4), 1211–1223, doi:10.1785/0120040177.
- Tinti, E., M. Cocco, E. Fukuyama, and A. Piatanesi (2009), Dependence of slip weakening distance (D_c) on final slip during dynamic rupture of earthquakes, *Geophysical Journal International*, 177(3), 1205–1220, doi:10.1111/j.1365-246X.2009.04143.x.
- Trugman, D. T., and E. M. Dunham (2014), A 2D Pseudodynamic Rupture Model Generator for Earthquakes on Geometrically Complex Faults, *Bulletin of the Seismological Society of America*, 104(1), 95–112, doi:10.1785/0120130138.
- 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.
- Withers, K. B., K. B. Olsen, and S. M. Day (2015), Memory-Efficient Simulation of Frequency-Dependent Q , *Bulletin of the Seismological Society of America*, 1–14, doi:10.1785/0120150020.
- Xu, X., X. Tong, D. T. Sandwell, C. W. D. Milliner, J. F. Dolan, J. Hollingsworth, S. Leprince, and F. Ayoub (2016), Refining the shallow slip deficit, *Geophysical Journal International*, 204(3), 1843–1862, doi:10.1093/gji/ggv563.
- Yao, Q. (2017), Dynamic Modeling of Earthquake Sources on Rough Faults, Ph.D. Thesis. SDSU/UCSD Joint Ph.D. Program in Geophysics, 30 August.
- Yoffe, E. H. (1951), The moving griffith crack, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 42(330), 739–750, doi:10.1080/14786445108561302.
- Zoback, M. D. et al. (1987), New Evidence on the State of Stress of the San Andreas Fault System, *Science*, 238(4830), 1105–1111, doi:10.1126/science.238.4830.1105.

Table 4.1: Coefficients for short range coregionalization matrix, $\mathbf{B}^1$.

Field	Δu	V_{peak}	$vrup$	$\mu 0$
Δu	0.0282			
V_{peak}	0.0002	0.0403		
$vrup$	0.0164	0.0631	0.6917	
$\mu 0$	0.0783	0.1215	0.154	0.6049

Table 4.2: Coefficients for long range coregionalization matrix, $\mathbf{B}^2$.

Field	Δu	V_{peak}	$vrup$	$\mu 0$
Δu	0.9718			
V_{peak}	0.81	0.9597		
$vrup$	0.1504	0.1841	0.3083	
$\mu 0$	0.0946	0.15	0.0859	0.3951

Table 4.3: Zero-offset correlation coefficients, $C(0)$, between source parameters computed from linear-model of coregionalization.

Field	Δu	V_{peak}	$vrup$	μ_0
Δu	1			
V_{peak}	0.8102	1		
$vrup$	0.1667	0.2472	1	
μ_0	0.1729	0.2715	0.24	1

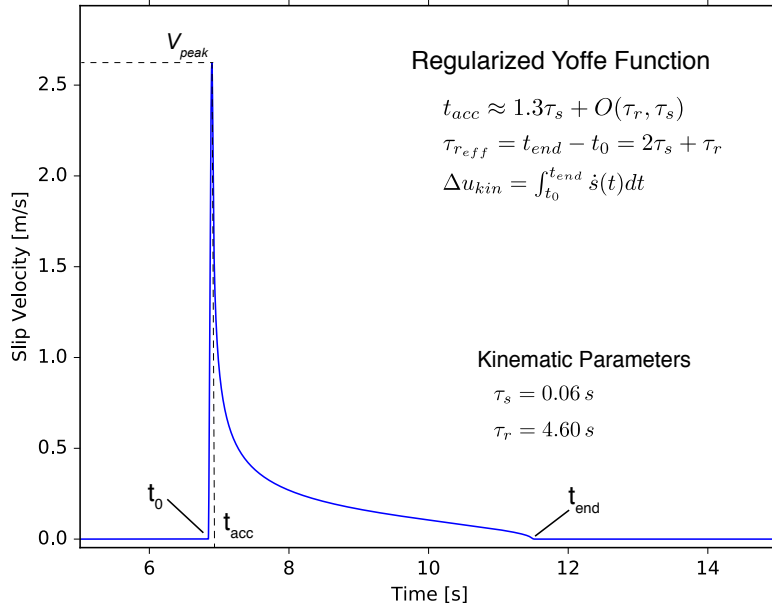


Figure 4.1: Regularized Yoffe function calculated using $\tau_s = 0.06 \text{ s}$ and $\tau_r = 4.60 \text{ s}$. In addition, we show useful definitions relating the kinematic parameters to characteristics of the source-time function.

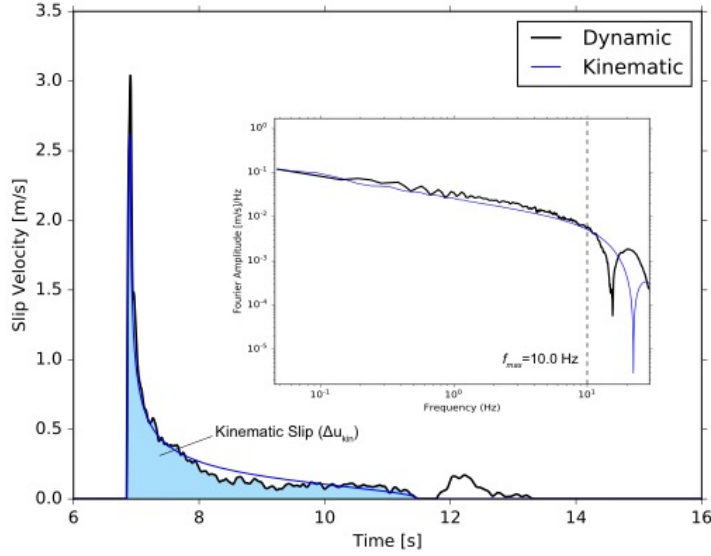


Figure 4.2: Dynamically computed slip-rate function (black) and best-fitting kinematic approximation (blue). The shaded region under the dynamic function depicts the kinematic slip, which does not include after-slip. The inset of the figure shows the comparison in the frequency domain. In the inset, the dashed line represents the target frequency of our KRG.

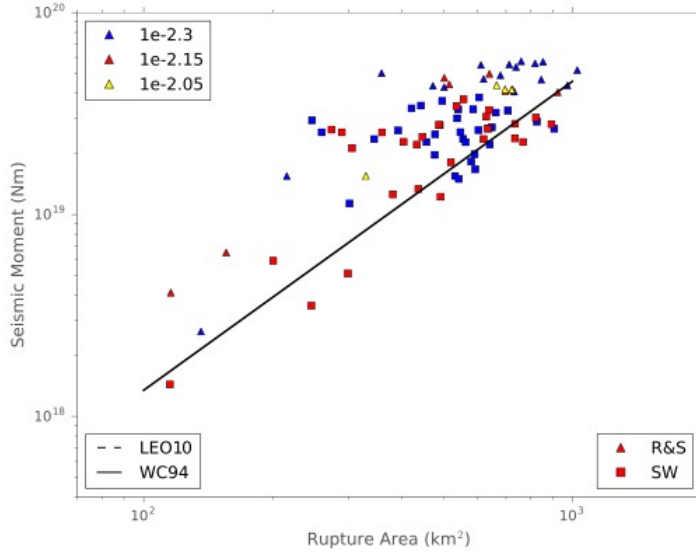


Figure 4.3: Moment vs. area scaling relationships plotted against empirically derived models for inter-plate strike-slip events [Leonard, 2010; Wells and Coppersmith, 1994]. We distinguish between fault roughness using color (blue: $\alpha = 0.005$; red: $\alpha = 0.007$; and yellow: $\alpha = 0.009$) and the friction law using shapes (triangle: rate-and-state; square = slip weakening), where α is the RMS of the rough fault profile normalized by fault length (h_{rms} / L).

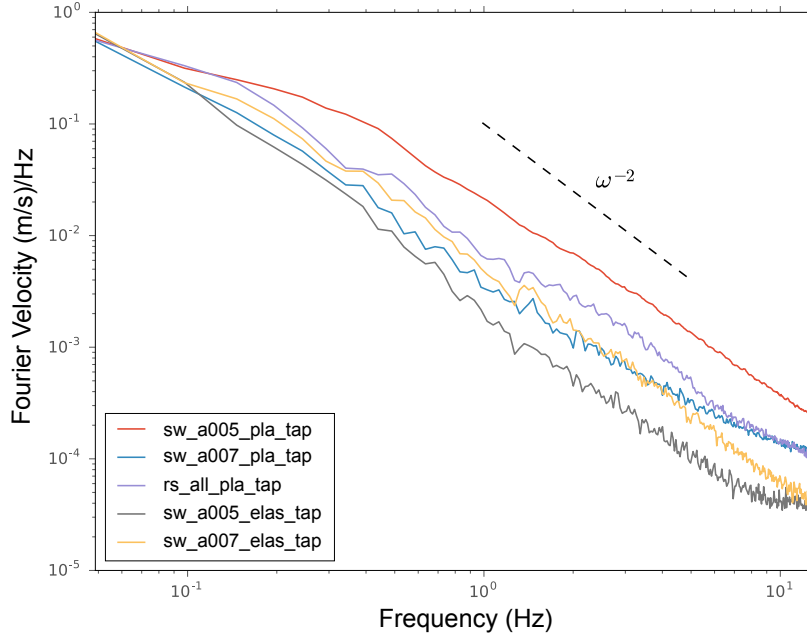


Figure 4.4: Stacked source spectra computed from finite-fault moment-rate functions for all five dynamic ensembles. Each colored line represents an ensemble average computed across all realizations in the ensemble. We normalize all spectra to unit variance before stacking. The dashed black line depicts the ω^{-2} high-frequency decay.

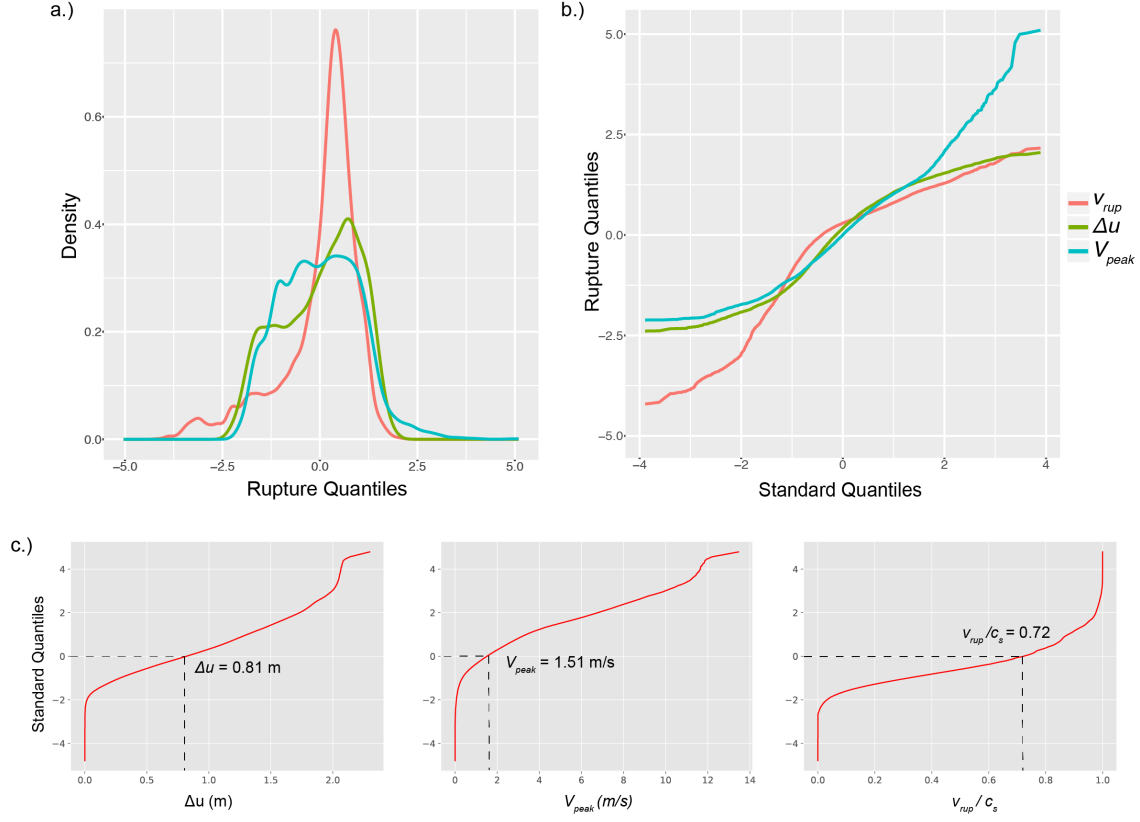


Figure 4.5: Density functions (a) and quantile-quantile plot (b) shown for normalized physical source fields. The one-point statistics are computed for all dynamic ruptures in our database. We normalize the physical fields by the simulation mean and standard deviation before computing the density function. (c) Mapping from simulated normal-score value to desired physical quantity. These functions preserve the order statistics observed in the dynamic rupture simulations by equating the k^{th} largest simulated value with the k^{th} largest value of the desired marginal probability density function. The median values are depicted on the figure with dashed lines.

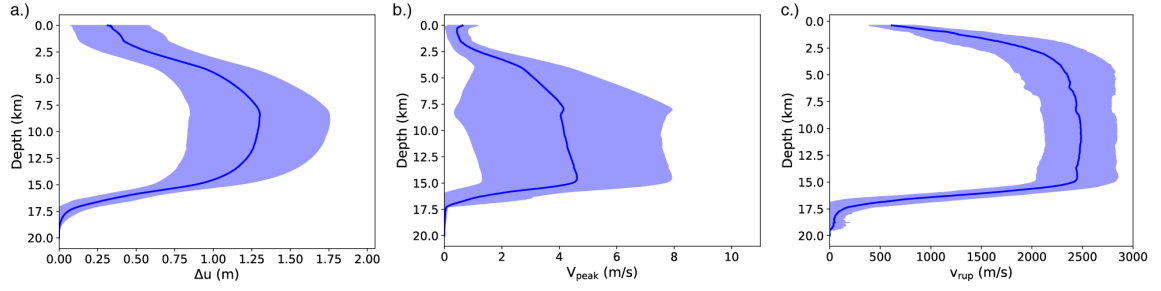


Figure 4.6: Depth-dependency of one-point statistics computed for (a) final slip Δu (b) peak slip velocity V_{peak} (c) and rupture velocity v_{rup} for all ruptures in our database. The thick blue line depicts the mean value computed at that depth for all ruptures, and the shaded region represents ± 1 standard deviation of the mean value. The near-surface effects of the velocity strengthening imposed in the rupture models are apparent in all of the physical fields. Notice the large variability in V_{peak} as compared with the other rupture fields.

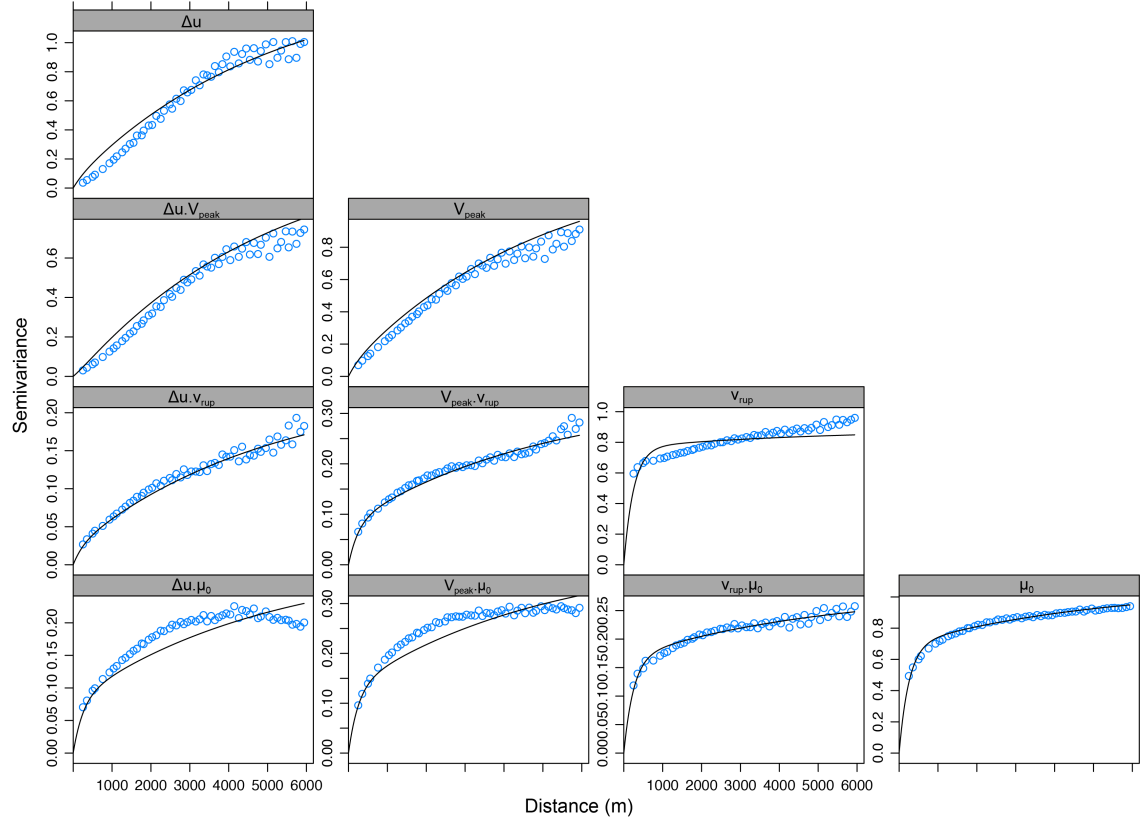


Figure 4.7: Preferred linear model of coregionalization for the dynamic rupture ensembles. Our model consists of two nested exponential semivariograms with range parameters 250 m and 5000 m. Expected semivariance estimates are shown using blue circles and the corresponding linear-model of coregionalization is depicted with a black line.

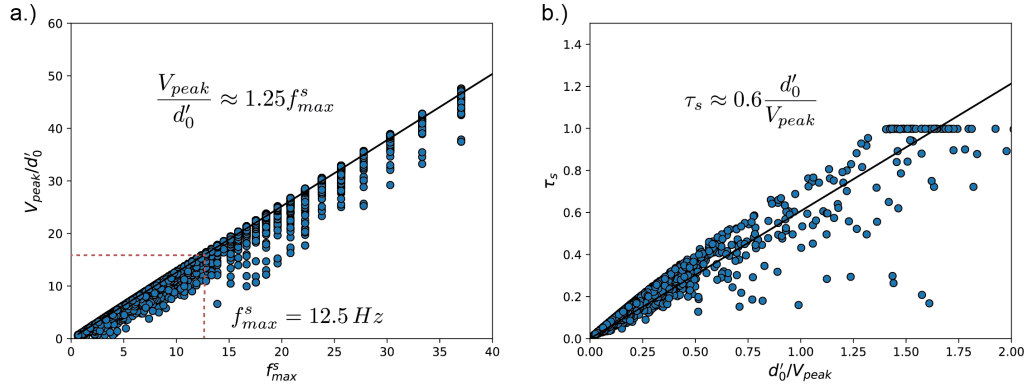


Figure 4.8: Outline of the two-step procedure to estimate τ_s from the physical source fields. (1) By choosing an arbitrary f_{max}^s , we can compute the d'_0 associated with this f_{max}^s by using the average V_{peak} for the simulation and applying the relationship shown in (a). (2) Using the estimate of d'_0 we calculate the corresponding value of τ_s assuming the subfault V_{peak} using the relationship shown in (b). This process produces spatially variable τ_s with average f_{max}^s chosen in step one.

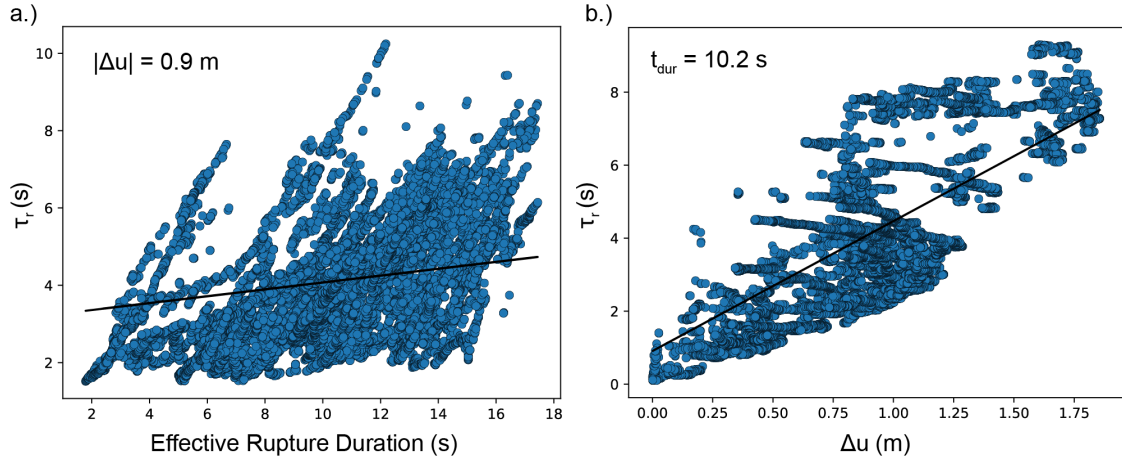


Figure 4.9: Multiple linear model defined to estimate τ_r given the final slip Δu and the effective rupture duration t_{dur} . (a) τ_r as a function of t_{dur} for subfaults where $\Delta u = 0.9 \text{ m} \pm 0.001 \text{ m}$. The black line shows the predicted values of τ_r using Equation 4.21. (b) τ_r as a function of Δu for subfaults where $t_{dur} = 10.2 \text{ s} \pm 0.01 \text{ s}$, the black line shows the predicted τ_r using Equation 4.21.

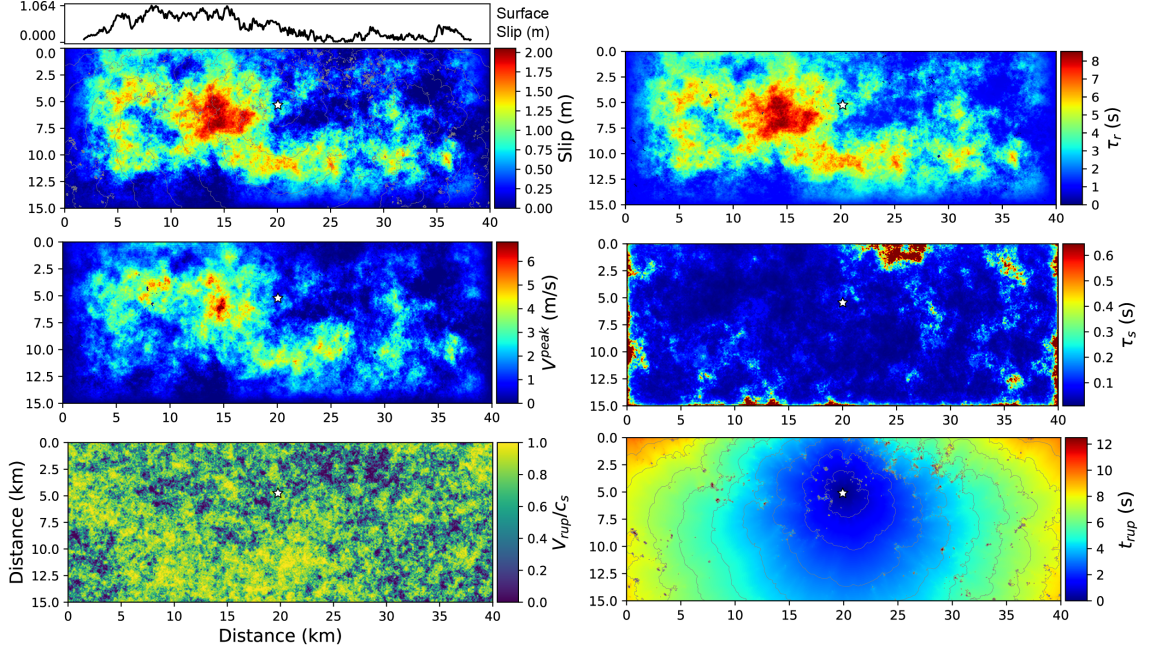


Figure 4.10: Example kinematic source model generated using methodology presented in this paper. The kinematic parameters Δu , τ_s , τ_r , and t_0 completely described the spatiotemporal evolution of slip on the fault using the regularized Yoffe function. We show the corresponding V_{peak} and v_{rup}/c_s associated with the kinematic parameters. The yellow star depicts the hypocenter location set at $\mathbf{x} = (20 \text{ km}, 5 \text{ km})$.

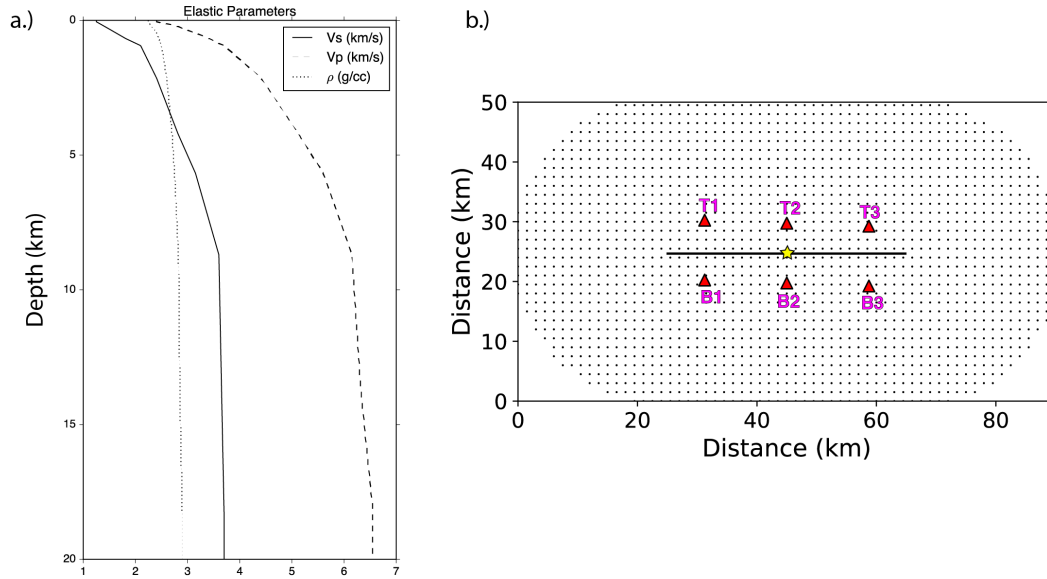


Figure 4.11: (a) Material model for shear-wave velocity (V_s), P-wave velocity (V_p), and density (ρ). This model represents a smooth 1D profile representative of a rock-site in southern California based on a 1D profile extracted from the SCEC CVM-S4.26 velocity model. (b) Free-surface of simulation domain showing six stations (red triangles) used for qualitative visualizations of seismograms and spectra. Black dots depict stations used to extract spectral accelerations for GMPE comparisons. The epicenter location is shown using the yellow star.

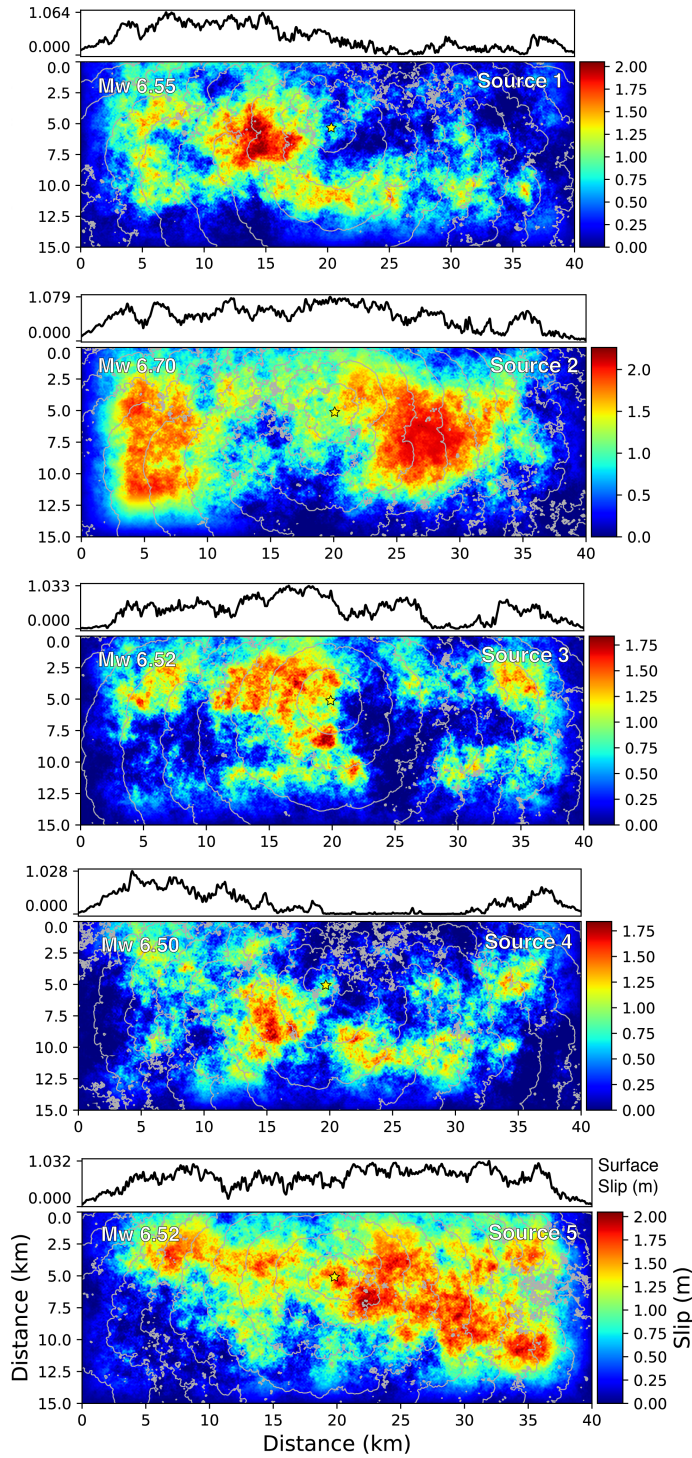


Figure 4.12: Final slip Δu distributions for five finite-fault source models produced using the kinematic rupture generator used for validation. We show the M_w of each earthquake in the upper left of each plot. Contour lines indicate 1 second intervals of rupture initiation times, and we depict the hypocenter location with the yellow star.

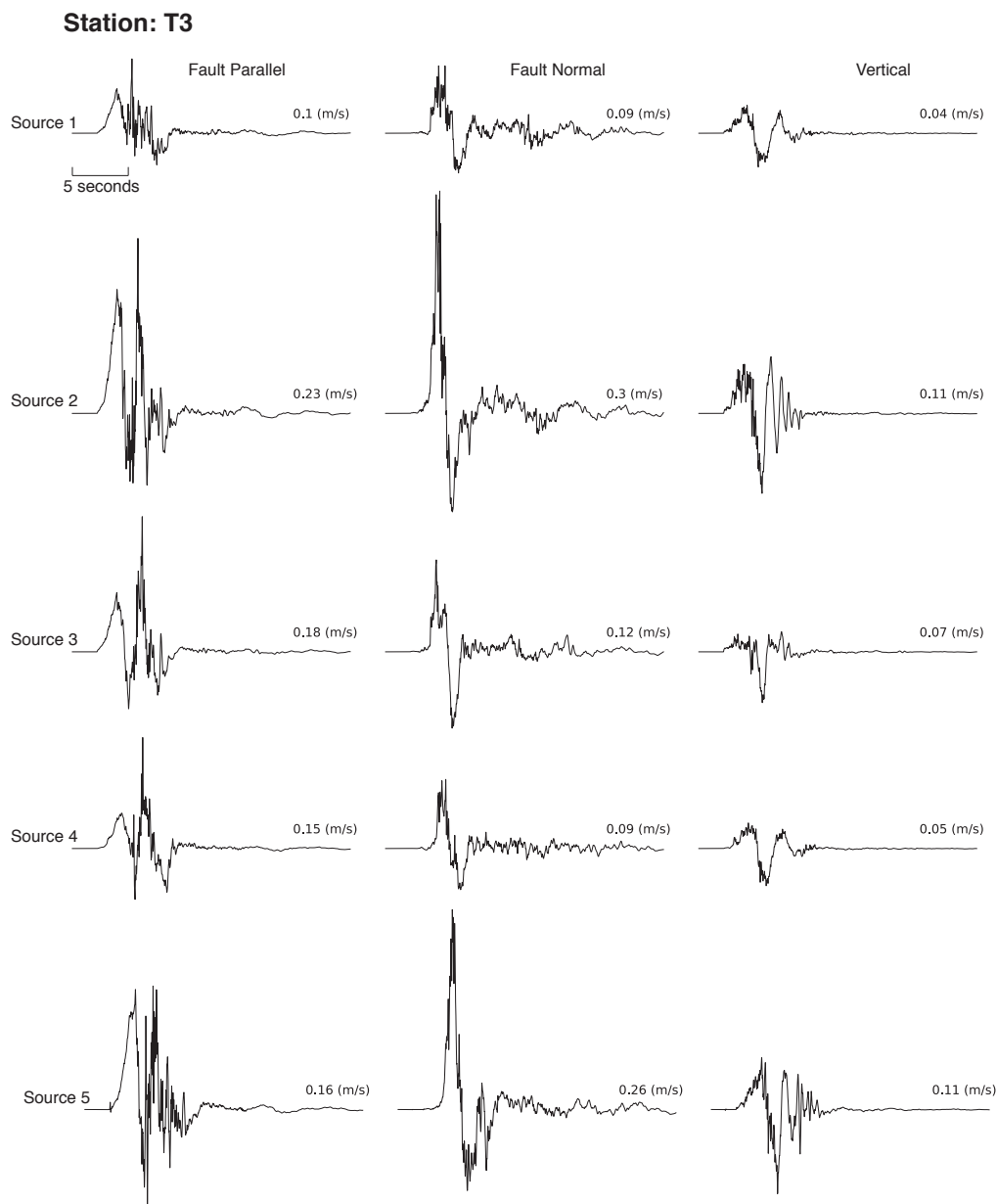


Figure 4.13: Velocity seismograms plotted for five source models at station T3 (see Figure 4.11b). The peak motions for each component are shown above the time-series. The synthetic seismograms are labeled by their respective source model (see Figure 4.12).

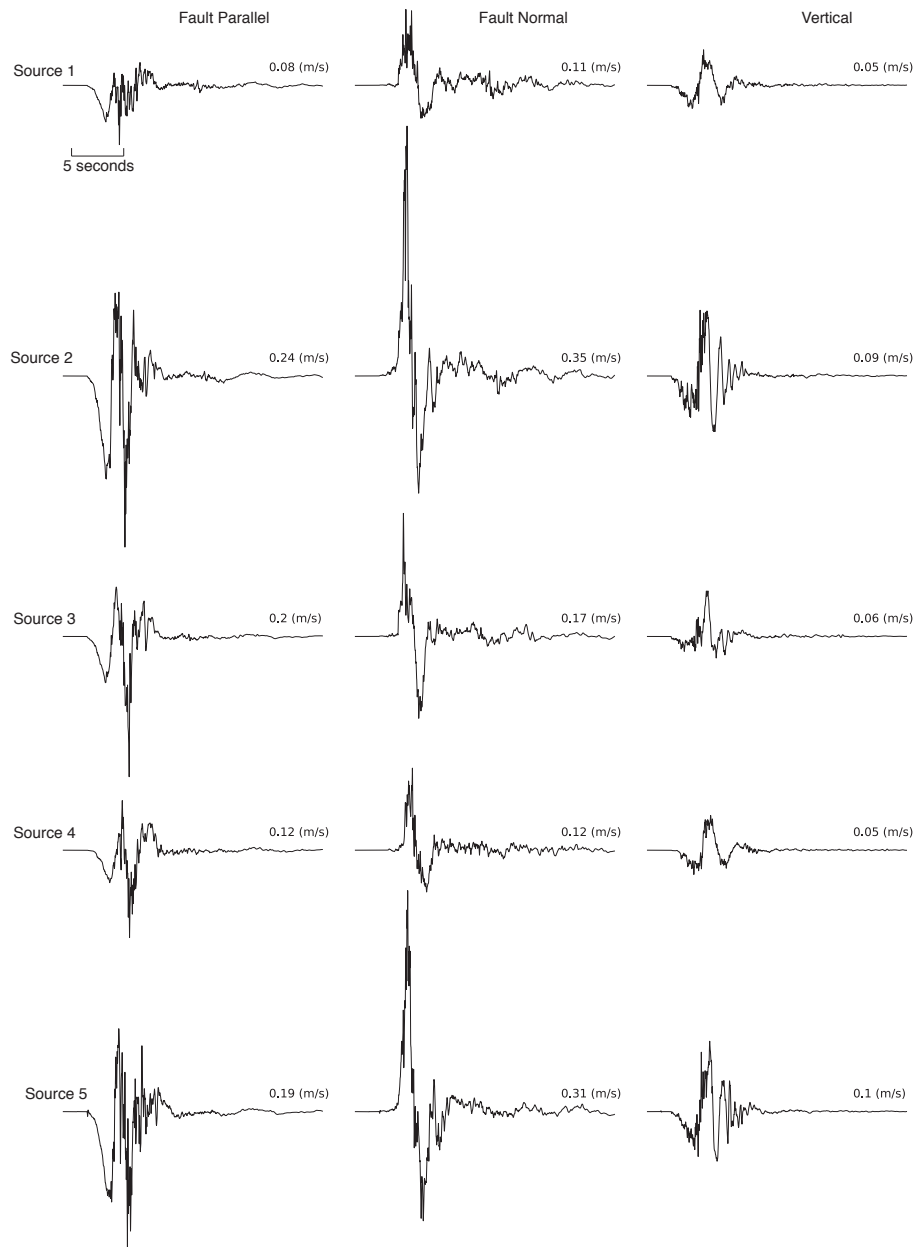
Station: B3

Figure 4.14: Same as Figure 4.13 but for station B3.

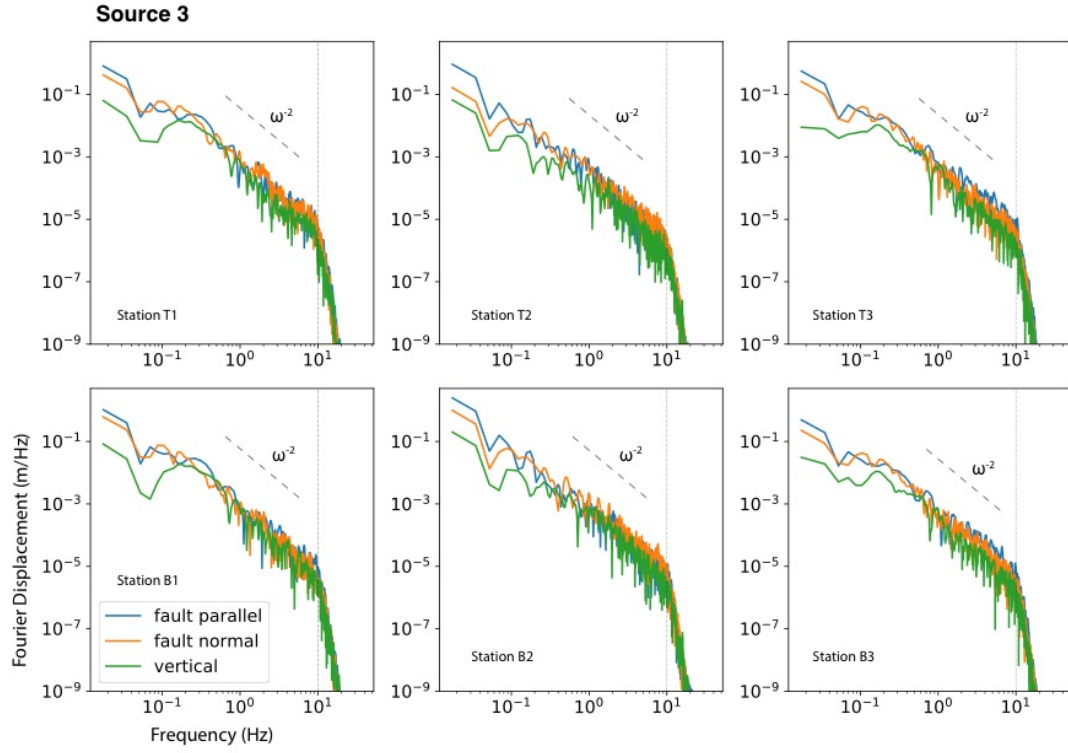


Figure 4.15: Normalized far-field displacement spectra computed for Source 3. The vertical dashed line depicts the $f_{max} = 10$ Hz of our deterministic simulations, and the thick dashed line shows the ω^{-2} high-frequency decay.

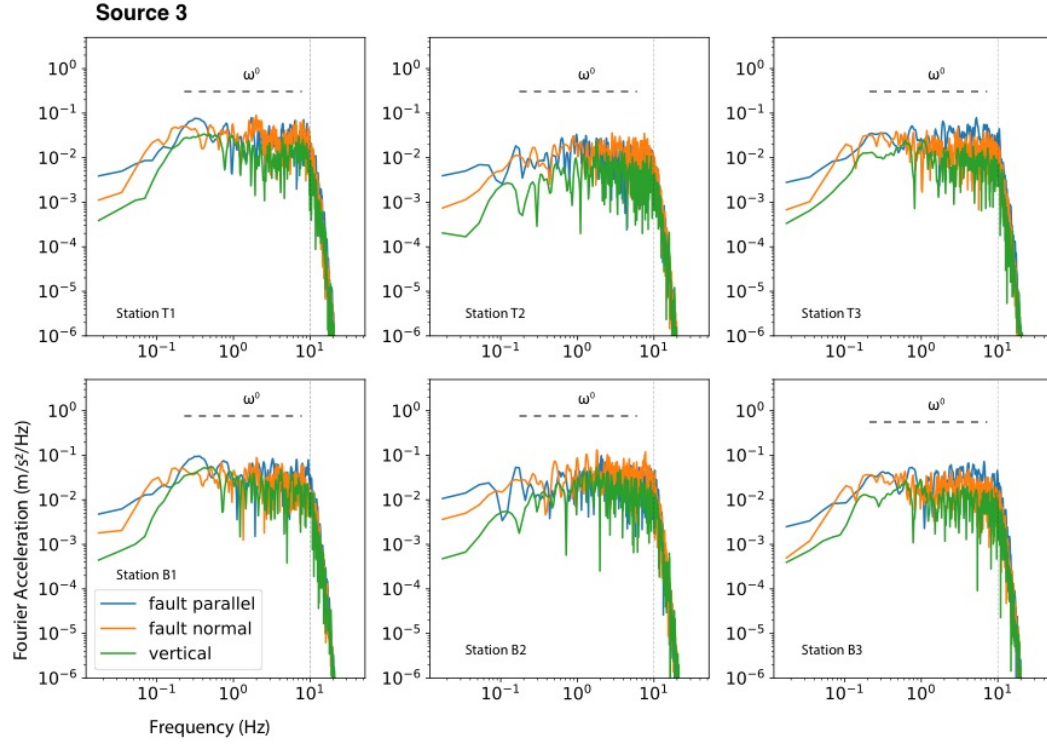


Figure 4.16: Far-field acceleration spectra computed for Source 3. The vertical dashed line depicts the $f_{\max} = 10 \text{ Hz}$ of our deterministic simulations, and the thick dashed line shows the ω^0 high-frequency decay.

Source 1

Source 2

Source 3

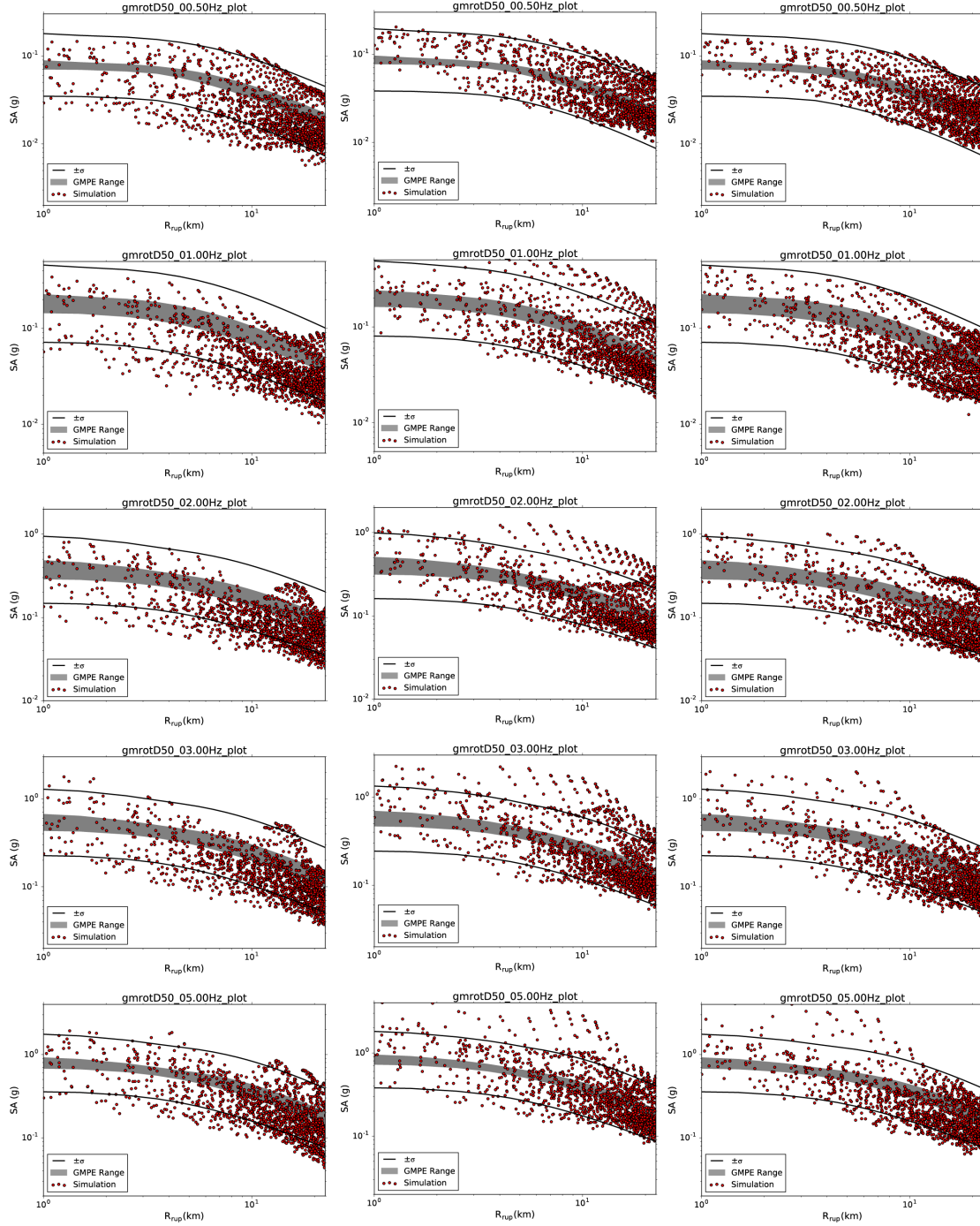


Figure 4.17: Comparisons of spectral accelerations for Source 1 (left), Source 2 (center), and Source 3 (right) against NGA West2 GMPE. Red dots represent spectral acceleration values computed at the stations depicted (see Figure 4.11a). The shaded gray region shows the range of GMPE median estimates and the black line indicates ± 1 total sigma.

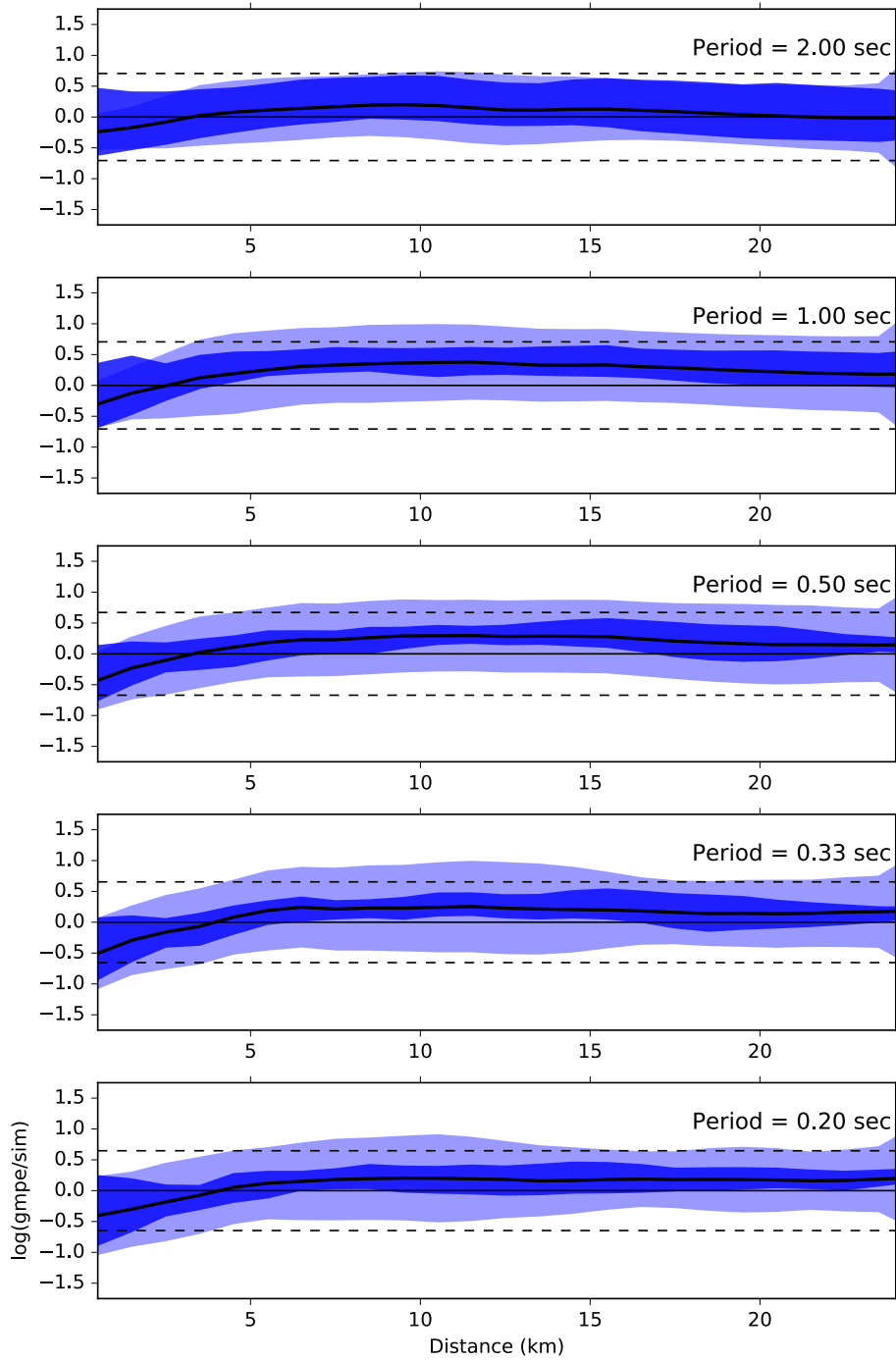


Figure 4.18: Bias plots comparing simulated median spectral accelerations for all five kinematic source models against GMPE medians. The thick black line depicts the ensemble average over all simulations, and the dark shaded blue region denotes the range of ensemble median values. The dashed black lines depict ± 1 intra-event σ (in log units) for each period predicted by the GMPE, and the light shaded blue region shows the standard deviation (in log units) in each distance bin from the simulated ensemble.