

Modeling of Earthquake Ground Motion: From the Source to the Site.

by

Tara Ashley Nye

A dissertation accepted and approved in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in Earth Sciences

Dissertation Committee:

Valerie Sahakian, Chair

Diego Melgar, Core Member

Amanda Thomas, Core Member

Ben Farr, Institutional Representative

University of Oregon

Winter 2024

© Winter 2024 Tara Ashley Nye
This work is licensed under a Creative
Commons Attribution License



DISSERTATION ABSTRACT

Tara Ashley Nye

Doctor of Philosophy in Earth Sciences

Title: Modeling of Earthquake Ground Motion: From the Source to the Site.

This dissertation investigates different approaches for improving modeling of earthquake ground motion by focusing efforts on the different physical properties which are traditionally thought to contribute to shaking—the source, the path, and the site. The robustness and informativeness of earthquake simulations and empirically-derived ground motion models rely on accurate modeling of these different components. Methods to improve ground motion modeling are highly studied; however, due to the complexity of processes that contribute to shaking intensities and uncertainty, much is still left poorly understood, especially when modeling rare or atypical events. This dissertation utilizes both simulated and real earthquake data to better constrain source, path, and site parameters.

In this dissertation, we use a one-dimensional semistochastic model to simulate data for the 2010 **M**7.8 Mentawai tsunami earthquake, as well as for 112 moderate-to-large arbitrary Cascadia Subduction Zone (CSZ) events. Tsunami earthquakes and CSZ ruptures are scarce, which results in minimal data to analyze and thus a poorer understanding of the rupture physics. However, a future occurrence of either scenario is likely to result in significant damage. With the paucity of real data, simulations are necessary to train and test earthquake and tsunami early warning algorithms. We use near-field observed data from the Mentawai tsunami earthquake to ascertain source parameter values needed to simulate tsunami earthquake-type events. With the absence of recordings from a CSZ interface event, we use global ground motion and scaling models to validate and tune modeling of the source and path in the simulation model. Through our analyses, we find the simulations to be well-modeled and representative of these uncommon events.

Finally, we constrain site effects in the seismically-active San Francisco Bay Area (SFBA) by using real recordings of small-to-moderate earthquakes to estimate the high frequency attenuation parameter (κ) and its site contribution (κ_0). We develop an automated algorithm for selecting the frequency bounds used to estimate κ , and we find spatial trends in κ_0 to be consistent with regional anelastic attenuation models and shallow geology. Through ground motion analyses, we find κ_0 to reasonably model regional ground motion and is thus a valuable contribution to future ground motion studies in the SFBA.

This dissertation includes previously published and unpublished co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Tara Ashley Nye

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR, USA
Brigham Young University, Provo, UT, USA

DEGREES AWARDED:

Doctor of Philosophy, Earth Sciences, 2024, University of Oregon
Bachelor of Science, Geology, 2019, Brigham Young University

AREAS OF SPECIAL INTEREST:

Seismology
Ground Motion Modeling
Earthquake Simulations

PROFESSIONAL EXPERIENCE:

Pathways Intern, USGS Earthquake Science Center, 2022

Graduate Employee, Department of Earth Sciences, University of Oregon,
2019–2024

Research Assistant, Department of Geological Sciences, Brigham Young
University, 2018–2019

Incorporated Research Institutions for Seismology (IRIS) Intern, USGS Geologic
Hazards Science Center, 2018

Research Assistant, Department of Geological Sciences, Brigham Young
University, 2016–2018

GRANTS, AWARDS AND HONORS:

Marthe E. Smith Memorial Science Scholarship, University of Oregon, 2023

Finalist, Three Minute Thesis (3MT), University of Oregon, 2023

FN-Johnston Geology Scholarship, University of Oregon, 2022

Mathilda Giles Scholarship, University of Oregon, 2021

Warren DuPre Smith Scholarship, University of Oregon, 2020

Promising Scholar Award, University of Oregon, 2019

Field Camp Scholarship, Brigham Young University, 2019

Honorable Mention, Undergraduate Student Paper Competition, Earthquake Engineering Research Institute (EERI), 2018

Kathryn L. Simpson Women in Science Scholarship, Brigham Young University, 2018

Section Winner, Student Research Conference, Brigham Young University, 2018

J. Keith Rigby Scholarship, Brigham Young University, 2016, 2018

PUBLICATIONS:

Nye, T. A., Sahakian, V. J., Schlesinger, A., Melgar, D., Williamson, A., Ferguson, E., Mahani, A. B., Lux, A., & Pirenne, B. (2023). *Validation framework for semistochastic simulations in cross-border Cascadia earthquake early warning.* Manuscript in preparation.

Nye, T. A., Sahakian, V. J., & Melgar, D. (2023). *Modeling ground motions and crustal deformation from tsunami earthquakes: Rupture parameter constraints from the 2010 Mentawai event.* Manuscript submitted for publication.

Nye, T. A., Sahakian, V. J., Baltay, A., King, E., & Klimasewski, A. (2023). Estimates of κ_0 and effects on ground motions in the San Francisco Bay Area. *Bulletin of the Seismological Society of America*, 113(2), doi: 10.1785/0120220046

***Nye, T. A., Sahakian, V. J., King, E., Baltay, A. S., & Klimasewski, A. R.** (2022). Estimates of kappa in the San Francisco Bay Area, *12th National Conference on Earthquake Engineering (12NCEE)*, Salt Lake City, UT, 27 June – 1 July.

*Peer reviewed conference paper

ACKNOWLEDGMENTS

My journey to this finish line has not been alone, and I would not be where I am today without the friendship and support of so many wonderful people. First and foremost I would like to thank my advisor, Valerie Sahakian. You are a brilliant scientist, and it's been an honor to learn from you over the years. You have taught me how see all pieces of the puzzle and how to craft a story to share my science with others. Thank you for your patience, kindness, and support as I've learned to become an independent scientist. I would also like to thank Diego Melgar, who's been an honorary advisor to me. Thank you for your endless encouragement and for answering my many questions about *FakeQuakes*. Your enthusiasm for science is contagious. Thank you to Amanda Thomas and Ben Farr for your mentorship on my committee. You've both provided valuable insight on this work and have been a wonderful support.

I've had many other teachers and mentors who've played a pivotal role in my search for knowledge throughout the years. To my elementary teachers, thank you for fostering my inquisitive nature. To Professor Bouker, thank you for sparking my love for Earth science. To my undergraduate research advisors John McBride and Ron Harris, thank you for helping me dip my toes into the world of research. To my IRIS and USGS internship mentors Kate Allstadt, Eric Thompson, Bruce Worden, and Alan Yong, thank you for pushing me to grow as a scientist and for giving me grace to figure things out on my own.

Thank you to all the collaborators who have been part of the different chapters of this work. This dissertation would not exist without each of your efforts, and I hope we can continue to collaborate on future work. I am also grateful to all of the professional colleagues and friends I've had the privilege of sharing enlightening conversations with over the years.

The unity of the graduate students is in part what drew me to this program, and I'm thankful for everyone I've had the pleasure of meeting along the way. To my current and former labmates Roey, Charity, Oluwaseun, Avigyan, and Lexi, thank you for both

the friendship and scientific discussions we've shared. To my fellow *Geo Gals* — Sydney, Becca, Lissie, Nicole, Annika, Angela, and Renee — I will forever treasure all of the laughter, conference shenanigans, game nights, and binging of trashy tv shows. As sad as I am for this chapter to come to a close, I look forward to seeing where life takes you, and I can't wait for our paths to cross again. I am also grateful to my long distance friends Patricia, Amber, Anja, Abby, and Lanier for continuing to support me, even thousands of miles away.

To my dearest family, I owe everything that I am to you. Thank you for encouraging all of my dreams, no matter how absurd (like pursuing a career in seismology simply because I love the movie *San Andreas*). I've taken the road less traveled, but you've made me believe that anything is possible, and you've cheered me on every step of the way. Mom, you've sacrificed so much for me to have the opportunities I've had, and for that, I am eternally grateful. Thank you for nurturing my curiosity and for instilling in me the belief that I can figure out any problem I put my mind to. Rodney, thank you for welcoming me into your life with open arms. Dad, thank you for always believing in me and for reminding me to laugh when things got tough. Kate, my partner in crime, thank you for inspiring me to find the joy in life. Your love and light are felt by everyone. I hope you continue to pursue your passions, whatever they may be.

Marcus, my loving partner, you have been my foundation throughout these chaotic years. I feel safe to leap into the unknown because I know you'll be there to catch me if I fall. Thank you for showing me unconditional love and patience and for embarking on this journey with me. I can't wait to see where life takes us without long distance getting in the way. I am also deeply grateful to your family for welcoming me into their lives and for providing me a home away from home.

Pursuing a graduate degree is like climbing a mountain, and I've had my fair share of rough patches along the way. However, I can truly say that this has been one of the best experiences of my life. Thank you to everyone who has climbed this mountain with me. I hope to see you all on the other side.

DEDICATION

To my mom, for never discouraging my relentless questions about the world and for encouraging me to seek out answers for myself— "Did you Google it?".

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	19
1.1. Background	19
1.1.1. Earthquake Simulations	19
1.1.2. Empirical ground motion models	21
1.1.3. The ground motion model puzzle	22
1.2. Outline	25
II. MODELING GROUND MOTIONS AND CRUSTAL DEFORMATION FROM TSUNAMI EARTHQUAKES: RUPTURE PARAMETER CONSTRAINTS FROM THE 2010 MENTAWAI EVENT	27
2.1. Introduction	27
2.2. Background	29
2.3. Data and Methods	34
2.3.1. Geophysical Dataset	35
2.3.2. Varying Rupture Parameters	39
2.3.3. Constraining Rupture Parameters	42
2.4. Results	43
2.4.1. Isolated Parameters	43
2.4.2. Covaried Parameters	45
2.4.3. Tsunami Earthquake Parameter Regression	47
2.5. Discussion	51
2.5.1. Simulation Limitations	51
2.5.2. Comparison with Previous Inversion Studies	53
2.5.3. Real-time Analysis	54

Chapter	Page
2.6. Conclusions	57
2.7. Data and Code Availability	58
2.8. Acknowledgments	59
2.9. Bridge	59
2.10. Supporting Information	60
III. VALIDATION FRAMEWORK FOR SEMISTOCHASTIC SIMULATIONS IN CROSS-BORDER CASCADIA EARTHQUAKE EARLY WARNING	68
3.1. Introduction	68
3.2. EEW in the Pacific Northwest	70
3.3. Stochastic Rupture Simulation	73
3.3.1. Kinematic Rupture Models	74
3.3.2. Simulated Waveforms	75
3.3.3. Post Processing	76
3.4. Earthquake Detection	78
3.5. Magnitude Estimation	82
3.6. Ground-Motion Intensity	85
3.6.1. Peak Ground Acceleration and Peak Ground Velocity	86
3.6.2. Modified Mercalli Intensity	90
3.6.3. Peak Ground Displacement	91
3.6.4. Ground-Motion Summary	97
3.7. EPIC and ONC-EW Performance Analysis	97
3.8. Conclusions	101
3.9. Data and Code Availability	104
3.10. Acknowledgments	104
3.11. Bridge	105

Chapter	Page
3.12. Supporting Information	106
3.12.1. Model Files	106
3.12.2. Offshore Noise	107
3.12.3. STA/LTA Triggering	108
3.12.4. Model Attenuation	111
IV. ESTIMATES OF κ_0 AND EFFECTS ON GROUND MOTIONS IN THE SAN FRANCISCO BAY AREA	112
4.1. Introduction	112
4.2. Background	114
4.3. Data and Methods	116
4.3.1. Dataset	116
4.3.2. Estimating Site Kappa	118
4.4. Results	123
4.4.1. Site Kappa	123
4.4.2. Producing a Regional Kappa Map	126
4.4.3. Site Parameter Evaluation	129
4.5. Discussion	136
4.5.1. Kappa Distance Dependence	136
4.5.2. Relationship Between κ_0 and Geology	137
4.5.3. Comparison with other Regional Studies	140
4.5.4. κ_0 and V_{S30} as Site Parameters	140
4.6. Conclusions	142
4.7. Data and Resources	143
4.8. Acknowledgments	144
4.9. Supporting Information	146
V. CONCLUSIONS AND FUTURE DIRECTIONS	152

Chapter	Page
. REFERENCES CITED	158

LIST OF FIGURES

Figure	Page
1.1. Ground-motion puzzle	24
2.1. Global map of tsunami earthquakes and Mentawai, Indonesia geographic region with the seismic and HR-GNSS stations	30
2.2. Finely discretized slip model and stochastic variations of the slip model	37
2.3. Waveform comparisons between the observed data and synthetic data and Fourier amplitude spectra comparisons for the waveforms	40
2.4. Effects of varying the rise time	44
2.5. Effects of varying the rupture velocity	45
2.6. Effects of varying the stress parameter	46
2.7. Gaussian process regression of residuals	50
2.8. Intensity measure residual analyses using standard parameters, mean values from the TsE parameter regression, and values within one standard deviation of the mean from the TsE parameter regression, and average STFs.	52
2.9. $\mathbf{M}_{PGA} - \mathbf{M}_{PGD}$ for observed events and simulations of the 2010 Mentawai tsunami earthquake	56
2.S1. Slip models for all 30 initial stochastic ruptures	60
2.S2. Coarsely discretized and finely discretized slip model	63
2.S3. Artificial notch in the broadband spectrum	64
2.S4. Smoothed spectrum absent of an artificial notch	65
2.S5. Original velocity model and three variations of the velocity model using softer layers	66
2.S6. Example solvation for \mathbf{M}_{PGA}	67

Figure	Page
3.1. Cascadia Subduction Zone study region and example stochastic rupture . . .	71
3.2. Example components of broadband waveform simulation	77
3.3. STA/LTA earthquake detection results for the Cascadia simulations	80
3.4. Simulation P_d measurements	84
3.5. Residuals between ground-motion model predictions and simulated ground motions for PGA and PGV	89
3.6. MMI residuals between ground-motion model predictions and simulated ground motions	91
3.7. PGD Residuals between ground-motion model predictions and simulated ground motions	94
3.8. Evaluation of synthetic GNSS noise and its effect on apparent PGD	96
3.9. Slip patterns of the six scenario events tested with EPIC and ONC-EW	99
3.10. Magnitude estimate evolution using both EPIC's and ONC- EW's offline system	100
3.11. Epicenter location error evolution using both EPIC's and ONC- EW's offline system	102
3.S1. Mesh file of the fault model and velocity model	106
3.S2. Station noise comparison	107
3.S3. Example late trigger of the STA/LTA algorithm	108
3.S4. Example trigger of the STA/LTA algorithm on the S-wave	109
3.S5. Absolute values the arrival time residuals	110
3.S6. PGA and PGV residuals generated using original attenuation parameters . . .	111
4.1. Study region of the San Francisco Bay area	113
4.2. Example application of the algorithm to select the frequency ranges, f_e and f_x , used to solve for κ	120
4.3. Example station spectra and κ estimates	123
4.4. Station-based κ_0 estimates and corresponding standard deviations	125

Figure	Page
4.5. Example semivariogram and binned empirical smivariogram	127
4.6. Median interpolated κ_0 and the standard deviation for the kriged κ_0 predictions	129
4.7. Site parameters κ_0 and V_{S30} versus GMM site residuals	133
4.8. Correlation between κ_0 estimates and both measured and proxy V_{S30}	135
4.S1. Example eHVSr curves	147
4.S2. Full interpolated κ_0 map and standard deviation maps	148
4.S3. Comparison of estimates of κ_0 with interpolated predictions	149
4.S4. Correlation between κ_0 estimates and V_{S30}	150

LIST OF TABLES

Table	Page
2.1. Tsunami earthquake simulation rupture parameters.	53
3.S1.Offshore noise variation from AL2H	107
4.S1.Stations removed due to unreliable κ_0 regression	151

CHAPTER I

INTRODUCTION

*If you have built castles in the air, your work need not be lost; that is where they should be.
Now put foundations under them.*

Henry David Thoreau, Walden

1.1 Background

How can we build a better model of earthquake hazards? Or in other words, how can we better model expected shaking (i.e., ground motion) from a likely earthquake in regions with high seismic risk? This question frames the work of this dissertation, but to be answered, it is necessary to understand what it means to model earthquake ground motion and what information such models provide. The field of ground motion modeling is broad and diverse, but it can primarily be divided into two end-member approaches: earthquake simulations and empirically-derived ground motion models.

1.1.1 Earthquake Simulations

Earthquake simulation models are used to generate realistic time evolutions of ground shaking (i.e. time-series waveforms) at locations of interest. These models simulate the types of data seismic and geodetic instruments would be recording in real time in the event of an earthquake. Physics-informed methods for simulating waveforms include stochastic models of high frequency ground motion (e.g., Boore et al., 2003), deterministic full waveform propagation through a two-dimensional or three-dimensional earth structure (e.g., Komatitsch and Tromp, 1999; Faccioli et al., 1997; Mazzieri et al.,

2013), and hybrid semi-stochastic models (e.g., Graves and Pitarka 2005, 2010; Melgar et al., 2016). A deterministic model propagates seismic waves through a known, typically three-dimensional, velocity structure, and it can be used to generate all frequencies of ground motion; however, it becomes quite computationally expensive at higher frequencies (Carrington et al., 2008). The cost benefit may not always be worth such computation costs if the velocity structure is not well-resolved at the higher frequencies, which is often the case. Instead, a stochastic approach can be used which models the ground motion in the frequency domain using well established models of earthquake sources and path and site effects. Semistochastic models use deterministic approaches to model low frequencies, which are combined with stochastic high frequencies using a matched filter (e.g., Graves and Pitarka, 2010, 2015). Some more recent studies have even developed machine learning algorithms to generate realistic synthetic waveforms (e.g., Wang et al., 2021).

Oftentimes, the simulated time-series are generated as an accompaniment to dynamic or kinematic fault ruptures. Dynamic ruptures are those where background stress and frictional conditions are prescribed, which drive the simulated rupture of an earthquake (e.g., Harris et al., 2018; Ulrich et al., 2019; Wollherr et al., 2019). Kinematic ruptures do not rely on pre-existing stress conditions, but rather are controlled by pre-defined static and time-dependent source parameters, such as the target magnitude, rupture area, rise time–magnitude scaling relations, and rupture velocity (e.g., Mai and Beroza, 2002; Melgar et al, 2016).

Earthquake simulations are desirable to better understand seismic and tsunamigenic hazards when real data are limited. This could be due to regional seismic quiescence, such as for the Cascadia Subduction Zone (CSZ) (e.g., Marafi et al., 2019; Melgar et al., 2016; Roten et al., 2020), which has not had a major earthquake in over 300 years. The newly funded Cascadia Region Earthquake Science Center (CRESCENT) has efforts in place to improve modeling of a large Cascadia megathrust earthquake through generation of regional velocity models and incorporation of fault coupling. Simulations resulting from these efforts will be critical for constraining expected hazards for communities along the

CSZ, a region with high seismic potential and highly vulnerable infrastructure. Simulations are also a necessary supplement when there's a global scarcity of certain event types, such as large or atypical earthquakes (e.g., Lin et al., 2021). Even when abundant data exist, the seismic record does not necessarily capture every possible future earthquake scenario. Thus, earthquake simulation models allow for generation of infinite earthquake scenarios with varying source properties, as well as exploration of biases resulting from poorly resolved or non-unique coupling and velocity structure (e.g., Small and Melgar, 2021). This is especially useful for geotechnical engineers who are interested in a building's unique response to shaking given site-specific near-surface properties (e.g., Dashti and Bray, 2012; Kusanovic et al., 2023; McCallen et al., 2021).

1.1.2 Empirical ground motion models

Although earthquake simulations are a form of ground motion modeling, the term “ground motion model” is reserved for an empirically-derived function which estimates some measure of ground motion intensity, such as peak ground motions or spectral accelerations, based on a set of inferred parameters. The simplest ground motion models consider only earthquake magnitude and distance from the rupture, whereas increasingly more complex models consider a multitude of parameters, including faulting style, time-averaged S-wave velocity in the upper 30 meters (V_{S30}), and depth to the S-wave 1.0 and 2.0 km/s isosurfaces ($Z_{1.0}$ and $Z_{2.0}$).

Hundreds of ground motion models exist, many of which can be accessed via the open source *OpenQuake* Engine (Pagani et al., 2014). Some are intended to be regional, so they are derived using only regional events. These models better capture regional effects but have high epistemic uncertainty (i.e. model uncertainty). Global models are derived using larger datasets but do not capture regional or path-specific effects, thus having higher aleatoric uncertainty (inherent data randomness or unmodeled effects). The current focus in this field is the move towards non-ergodic ground motion models, which capture all the regional heterogeneities but require significantly larger amounts of data to resolve path and site effects for each unique raypath (e.g., Dawood and Rodriguez-Marek, 2013; Landwehr

et al., 2016; Lavrentiadis et al., 2023; Sung et al., 2023).

The decision of model complexity and size of dataset have a significant effect on the model uncertainties. The generation of global models regressed on large datasets yields low epistemic uncertainties, but high aleatoric uncertainties, which means that most of the variability in the ground motion is attributed to unknown effects. The move towards more descriptive models calibrated for individual regions reduces aleatoric uncertainty as the model is better able to model complex features and captures regional crustal properties. Consequently, this also increases epistemic uncertainty as fewer data are available for small regions. As seismic datasets are exponentially increasing, the epistemic uncertainty can be reduced. Ground motion model estimates have wide applicability, and are often used to inform seismic hazard models and seismic risk assessments. They are heavily used in the earthquake engineering community to inform building codes (FEMA, 2022), and they are the basis of seismic hazard curves (e.g., Bommer et al., 2010; Lanzano et al., 2016) which underpin the National Seismic Hazard Model (Petersen et al., 2018).

1.1.3 The ground motion model puzzle

Modeling earthquake ground motion, whether in the form of an earthquake time-series simulation or estimation of a ground motion intensity measure, can be likened to solving a puzzle with three primary pieces representing the earthquake source, path, and site. The main pieces can be further divided into smaller pieces with information describing different physical properties of the three components, each of which has a unique influence on the resulting ground motion.

The source refers to the physics governing the earthquake rupture and is perhaps the most intuitive when understanding resulting shaking. Source parameters include those relating to the area of the rupture, amount of energy released, faulting mechanism, rupture directivity, rupture frequency content, amongst others. Naturally, a larger rupture will release more energy resulting in stronger shaking, and cities in the direction of rupture propagation will experience a focusing of seismic energy.

The seismic waves and frequency content become altered as the waves propagate

through the Earth. Such processes are considered path effects because they refer to changes in wave properties along the path between the earthquake source and where the shaking is felt. Just as sound waves in the air spread out and attenuate with distance, seismic waves attenuate as they travel greater distances, meaning the farther away you are from the earthquake, the less shaking you will likely feel. This path phenomenon is due to two main processes—geometric spreading and anelastic attenuation. If we assume the earthquake is a point source, the waves spread out concentrically around the hypocenter, and the energy is dispersed over larger area, like ripples in a pond after dropping a pebble in the water. In this case, energy is not lost, but the energy density decreases, resulting in an apparent decrease in energy with distance. Some energy is however lost due to anelastic attenuation. The Earth acts like a damping operator, and as the waves propagate and interact with the earth, some energy is lost in the form of heat. This attenuation, inversely related to the seismic quality factor (Q), refers to the amount of energy preserved with each cycle, so lower values of Q indicate greater attenuation. High frequencies attenuate faster than low frequencies as high frequency waves undergo more cycles in the same amount of time.

The seismic waves are further altered as they interact with near-surface conditions, referred to as site effects. Two seismic instruments could be located the same distance from an earthquake, but if situated above varying geologic conditions, they could record substantially different ground motions. Unlike path effects which primarily attenuate energy and decrease ground motion, near-surface conditions can both attenuate energy and amplify shaking. The source of both the amplification and attenuation is the weaker properties of the shallow crust, resulting from lower confining pressures, increased heterogeneity, and the presence of unconsolidated sediments. Seismic waves propagate more slowly through weak material and are subsequently amplified, which generally occurs at lower frequencies. The weaker subsurface also attenuates energy at the higher frequencies due to both anelastic attenuation and elastic scattering of energy from heterogenous crustal structure and sediments overlying an impedance boundary. The nature of the frequency content of ground motion is of great significance to earthquake

engineers because structures with greater heights and lengths are more sensitive to low frequency ground motions, and vice versa.

In a perfect world, we would know all the puzzle pieces required to perfectly replicate earthquake ground motion. Not only do we not know of every required piece, we also do not know what shape each piece should be. However, through continued data acquisition and analysis, we can better resolve the pieces of the puzzle and progress towards models that more closely match the ideal (Figure 1.1). With this dissertation, I have sought to better refine models of earthquake ground motion by focusing on aspects of each of the pieces of the ground-motion puzzle.

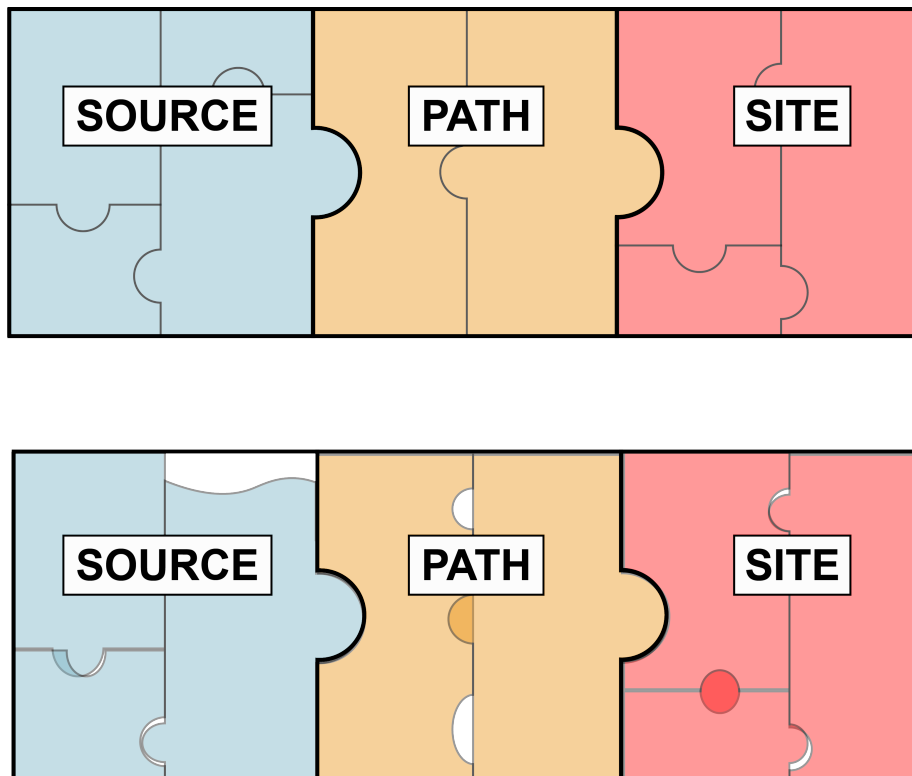


Figure 1.1. (top) Ideal ground-motion puzzle. (bottom) Example model of ground motion which closely approximates the ideal ground motion.

1.2 Outline

The journey of this dissertation begins with the source in Chapter II, where we will constrain the rupture parameters of the 2010 Mentawai tsunami earthquake. Tsunami earthquakes are rare, yet highly destructive events, which generate tsunamis around an order of magnitude higher than expected due to their shallow domain rupture. To improve tsunami early warning of these events, we need to understand how near-field recordings of tsunami earthquakes compare to typical earthquakes, which is challenging as near-field geophysical data currently only exist for the 2010 Mentawai tsunami earthquake. We use one-dimensional semistochastic simulations (Melgar et al., 2016) to infer which values of distinct rupture parameters of tsunami earthquakes are needed to model such an event. The work in this chapter was done in collaboration with Valerie Sahakian and Diego Melgar and is under review for publication in *Seismica*.

We continue with the emphasis on earthquake simulations, but transition from Chapter II's single-earthquake focus to a region-wide focus in Chapter III with the Cascadia Subduction Zone (CSZ). Similar to the notion that tsunami earthquakes are rare and poorly recorded in the near-field, Cascadia region earthquakes are infrequent, and most recent events have been small-to-moderate crustal or intraslab earthquakes. Although, the CSZ is capable of generating $M9$ events, to date, no ruptures along the main CSZ megathrust have been recorded on seismic instruments. We use the same modeling codes presented in Chapter II to simulate CSZ events of various magnitudes for use in testing regional earthquake early warning (EEW) systems. We will perform a series of analyses on these simulations to validate the effectiveness of the model at capturing certain features critical to EEW frameworks. These analyses broaden to include analyses of the source and the path. The work in this chapter was done in collaboration with multiple co-authors, namely Valerie Sahakian, Angela Schlesinger, Diego Melgar, Amy Williamson, Eli Ferguson, Alireza Babaei Mahani, Angie Lux, and Benoit Pirenne. It will soon be submitted for publication in *Seismica*.

In Chapter IV, we shift away from earthquake simulations and explore site

characterization to refine ground motion model (GMM) estimates. Currently, most GMMs model site response through site amplification proxies. However, we know that attenuation is also an important factor in ground motions. Some studies have sought to incorporate site-specific high-frequency attenuation (κ_0) in site response models, but κ_0 estimates are lacking in many regions as it is empirically derived and therefore dependent on a large suite of data. We will estimate κ_0 for over 200 stations in the seismically active San Francisco Bay Area and evaluate its effectiveness as a GMM parameter. We will also produce a spatially-continuous map of κ_0 to support efforts towards non-ergodic ground motion modeling. The work in this chapter was done in collaboration with Valerie Sahakian, Elias King, Annemarie Baltay, and Alexis Klimasewski, and it was published in the *Bulletin of the Seismological Society of America*.

Chapter V presents concluding remarks and thoughts on future directions in the field of ground motion modeling.

CHAPTER II

MODELING GROUND MOTIONS AND CRUSTAL DEFORMATION FROM TSUNAMI EARTHQUAKES: RUPTURE PARAMETER CONSTRAINTS FROM THE 2010 MENTAWAI EVENT

From Nye, T. A., Sahakian, V. J., & Melgar, D. (2023). *Modeling ground motions and crustal deformation from tsunami earthquakes: Rupture parameter constraints from the 2010 Mentawai event*. Manuscript submitted for publication. The writing of this chapter was done by me with co-authors providing editorial assistance. Valerie Sahakian and I conceptualized the work presented in this chapter. Diego Melgar provided the simulation codes, with some modifications by me. I performed all analyses presented in this chapter.

2.1 Introduction

Tsunami earthquakes are rare, end-member earthquakes that generate tsunamis much larger in height than expected for the magnitude of the event (Kanamori, 1972). These are typically a moment magnitude $M7-8$, but the runup is on the order of 10x greater than that of a typical megathrust earthquake of similar magnitude (Hill et al., 2012; Sahakian et al., 2019; Satake et al., 1993). This tsunami amplification, or *tsunami efficiency* (Lotto et al., 2017), is a result of the shallow rupture of tsunami earthquakes (hereafter referred to as TsEs). TsEs rupture the shallowest segment of the subducting

slab (i.e., domain A; Lay et al., 2012), a region generally considered aseismic due to the low confining pressures and high pore fluid pressures. These conditions yield velocity-strengthening frictional properties, which result in predominantly stable sliding that does not allow for elastic strain accumulation (Lay and Bilek, 2007; Lay et al., 2012). However, at times seismic slip is observed, such as when shallow slow slip events are recorded, and even less often, when the shallow megathrust ruptures seismically as a TsE.

Many fatalities are common as the result of TsEs because the tsunamis are unexpected in size, the shaking is milder than expected, and the moderate magnitude of these events is challenging for magnitude driven tsunami early warning algorithms. For instance, the 1896 **M**8.0 Sanriku, 1992 **M**7.6 Nicaragua, 2006 **M**7.8 Java, and 2010 **M**7.8 Mentawai events had tsunami or run-up heights reaching 25 m, 9.9 m, 8 m, and 16.9 m, and resulted in around 22,000, 170, 600, and 509 fatalities, respectively (Bilek et al., 2016; Hill et al., 2012; Mori et al., 2007; Satake, 1994; Tanioka and Satake, 1996; Yue et al., 2014). Early discrimination of these earthquakes is critical to provide ample warning to people within the inundation zone and reduce the number of casualties and deaths. Energy to moment ratios from teleseismic data are an effective discrimination tool (Newman and Okal, 1998); however, they are not often early enough to provide warning due to the travel time of global wave phases compared to tsunami inundation times (Newman et al., 2011).

Recent work (Sahakian et al., 2019) shows that early discrimination may be possible using a combination of near-field high-rate real-time Global Navigation Satellite Systems (HR-GNSS) and seismic data, but fully integrating this observational discrimination with local tsunami warning requires a basis in the physics of TsE rupture. Furthermore, it is difficult to evaluate the effectiveness of such an integration without data with which to test. As near-field geophysical data currently only exist for the 2010 Mentawai TsE, simulated TsE data would be a valuable supplement to both developing and validating local warning models (Tsushima and Ohta, 2014; Williamson et al., 2020), but it is first necessary to establish both the model parameters that control the behavior of TsEs and their expected values.

In this project, we determine these key TsE rupture parameters and obtain their expected means and standard deviations. The rupture parameters rise time, rupture velocity (V_{rupt}), and high-frequency stress parameter are related to shallow megathrust properties and are known to be characteristically extreme for TsEs. TsEs are energy deficient, long duration events and thus have shown to exhibit long rise times, slow rupture velocities, and low stress parameters. Using a combination of near-field geophysical data and modeled synthetic data, we aim to constrain the values of these parameters for TsEs, using the 2010 Mentawai TsE as a case study, to assist early tsunami warning algorithms in detecting these insidious events.

2.2 Background

TsEs are poorly understood, not only because of their infrequent occurrence and rupture of the typically aseismic domain, but also because they do not appear to be biased towards any set of interface conditions, meaning assigning likelihoods to their future occurrence, locales, and other characteristics is challenging. To date, we only have record of ~ 12 TsEs (some historical TsE classifications are debated): 1896 **M8.0** Sanriku, 1946 **M8.2** Aleutian Islands, 1947 **M7.0** Gisborne, 1947 **M7.0** Tolga Bay, 1960 **M7.6** Peru, 1963 **M7.8** Kuriles, 1975 **M7.5** Kuriles, 1992 **M7.6** Nicaragua, 1994 **M7.6** Java, 1996 **M7.4** Peru, 2006 **M7.8** Java, and the 2010 **M7.8** Mentawai events (Figure 2.1a). Of these, eight occurred along erosional margins, and four occurred along accretionary margins with defined accretionary prisms, but this erosional margin bias is likely due to the greater proportion of erosional margins ($\sim 75\%$) globally (Lotto et al., 2017). These events have also occurred in settings with diverse plate bathymetry, varying convergence rate, and varying plate age (Bilek and Lay, 2002). The only seeming universality among TsEs appears to be their rupture of the shallow domain A, which results in both slower and enhanced slip, regardless of the tectonic setting.

In erosional settings, sediment erodes the upper plate near the trench, and the downgoing plate develops a horst and graben structure which traps sediments (Ruff, 1989). Although lacking a defined prism, the presence of some sediment in erosional margins still

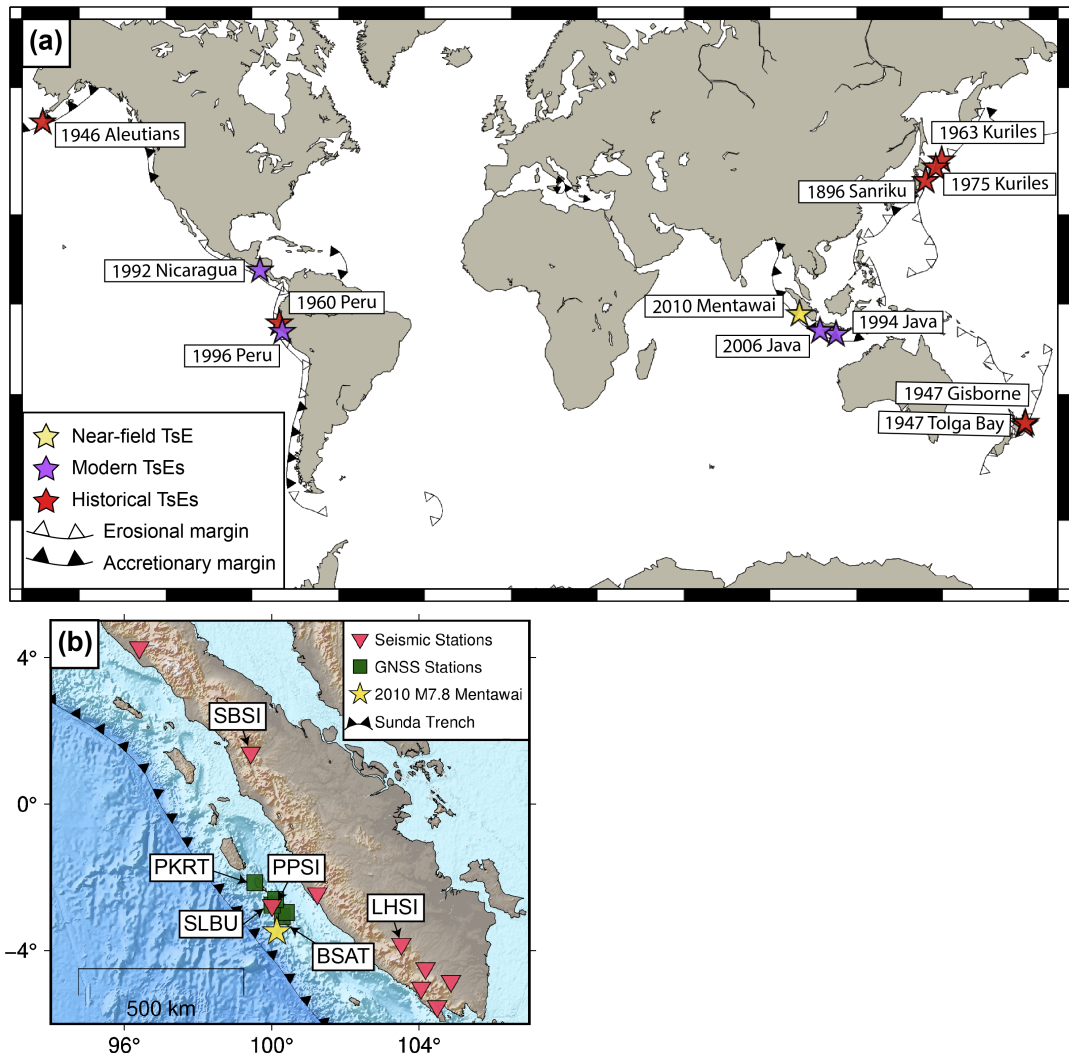


Figure 2.1. (a) Global map of tsunami earthquakes (TsEs). Red stars indicate historical TsEs that occurred before the development of modern electronic inertial seismic instruments, purple stars indicate TsEs that occurred since the development of modern seismic instruments, and the yellow star indicates the only near-field recorded TsE to date. Subduction classifications are taken from Clift and Vannucchi (2004), where erosional margins are indicated by trench lines with open triangles, and accretionary margins are indicated by trench lines with filled triangles. (b) Mentawai, Indonesia geographic region with the seismic stations (pink triangles) and HR-GNSS stations (green squares) used for this study. Stations referenced in the text are labeled on the map. The epicenter of the 2010 M7.8 Mentawai tsunami (yellow star), and trench (triangle line) are also indicated.

alters the frictional properties, resulting in greater but slow and jerky slip, and possessing the potential to break ocean floor and efficiently generate tsunamis (Kanamori and Kikuchi, 1993; Lay and Bilek, 2007; Lotto et al., 2017; Tanioka et al., 1997; Weinstein, 2005). In accretionary settings, a compliant prism of accreted sediment accumulates. The prism is $\sim 1/10$ th the rigidity of deeper segments of the slab that rupture during other subduction zone earthquakes, thus slip propagating into this medium is enhanced, which is described by the rearranged seismic moment (M_0) equation below,

$$D = \frac{M_0}{\mu A} \quad (2.1)$$

where D is the amount of slip on the fault, μ is the rigidity, and A is the fault area. Large-volume prisms increase tsunami height and tsunami efficiency by increasing the b -rate-and-state parameter (Lotto et al., 2017) (i.e., more velocity weakening and unstable; Scholz, 1998). For example, the largest tsunami waves for the great 2011 M9.0 Tohoku-Oki earthquake were observed ~ 100 km north of where the greatest slip was recorded because the wedge is much thicker to the north (Ma and Nie, 2019). The lower rigidity of the prism also decreases the shear wave velocity (V_S), which in turn is believed to lead to a slower V_{rupt} and subsequently longer rupture duration of the fault segment (Lay et al., 2011; Newman et al., 2011).

It is still unclear how the shallow aseismic domain on occasion hosts seismic events, but several theories have been proposed. Some studies suggest TsEs are generated by rupture of large asperities in conditionally stable regions just downdip of the aseismic zone, which then propagates into the aseismic domain if large enough (e.g., Bilek and Lay, 2002; Lay and Bilek, 2007; Lay et al., 2012). The 2011 Tohoku-Oki earthquake, although not a TsE, exemplifies this with its significant slip that ruptured through the shallow megathrust, possibly breaking the ocean floor (Jeppson et al., 2018). Although it seems any subduction zone could still host a TsE (Bilek and Lay, 2002), the historical record is too short to conclude this. Shallow aftershocks of a 2007 Mentawai, traditionally-rupturing (i.e., not TsE), event suggest frictionally unstable patches are in fact present all the way to the trench (Collings et al., 2012). Hill et al. (2012) suggest that the rupture does

not break the seafloor, but instead occurs along a blind thrust fault propagating along the base of the prism and extending to about 1.5 km below the surface. Because steeply dipping faults produce higher seafloor uplift, they could be a mechanism for TsEs (Kanamori and Kikuchi, 1993); the 1975 **M**7.0 Kuriles event is an example of this (Fukao, 1979). Although the mechanisms for seismicity along the shallow megathrust are debated, coseismic rupture is still possible along décollements (such as in the accretionary prism) regardless of tectonic setting, plate age, or fault type (Hubbard et al., 2015). Despite the range of conditions and environments which have hosted TsEs, their resulting ground motion data exhibit the same characteristics: long moment-rate-function and decreased high-frequency content.

Although tsunami early warning (TEW) systems are active in every major ocean basin (Hirshorn et al., 2020), their earthquake magnitude-driven algorithms are not currently capable of accurately relaying the hazard level of TsEs fast enough to issue a useful local warning (~ 5 – 10 min) (Hill et al., 2012). TsEs have a more moderate magnitude than typical tsunamigenic events ($\sim \mathbf{M}7$ – 8 as opposed to $\sim \mathbf{M}8$ – 9), so they do not register as a threat. Furthermore, TsEs are depleted in high frequencies due to the slower V_{rupt} (Yao et al., 2013), which means compared to a typical earthquake of the same magnitude, a TsE would produce unexpectedly low shaking in the near-field (Sahakian et al., 2019; Wirth et al., 2022), and it would appear even smaller in magnitude when looking strictly at strong motion data. Simply put, a TsE shakes like a **M**6 but produces a tsunami like a **M**9. In some regions where residents are encouraged to immediately seek higher ground if strong shaking is felt, the low shaking of a TsE would not necessarily suggest an impending large tsunami hazard.

Historically, methods to classify TsEs have monitored parameters relating radiated energy (E_R) to the time domain because the combination of slow rupture and energy deficiency of TsEs results in a low energy-to-moment (E_R/M_0) ratio, indicative of the “slowness” of rupture, or Θ (Newman and Okal, 1998). TsEs were determined to have Θ values ~ 1 log unit lower than those for a standard earthquake (Weinstein and Okal, 2005). Although determinable in real-time, there can be an issue of under or overestimating Θ ,

leading to false TsE identifications. This requires a detailed investigation of the source to confidently characterize an event as a TsE, which cannot be accomplished in real-time. Others, such as Newman et al. (2011), have explored the high-frequency energy-to-duration-cubed (E_{hf}/T_R^3) ratio. Because TsEs have longer rupture durations and radiate seismic waves inefficiently, low values of E_{hf}/T_R^3 are a strong indication of such an event. E_{hf} and T_R^3 can both be calculated in real-time, but because the algorithms use teleseismic data, they do not stabilize quickly enough to detect the TsE before inundation.

HR-GNSS and seismic data can, however, be used as near real-time proxies for local warning, if both are available in the near-field (Sahakian et al., 2019). Peak ground displacement (PGD) from HR-GNSS recordings is representative of the overall magnitude of the earthquake and is relatively insensitive to the shallow rupture processes. Conversely, peak ground acceleration (PGA) from seismic recordings is more sensitive to slow rupture of the shallow megathrust and reflects the energy deficiency of TsEs (Sahakian et al., 2019). Thus, the magnitude estimated using PGA would be significantly lower than the magnitude estimated using PGD, and acts as a proxy E_R/M_0 . However, with only 5 notable events in times of modern instrumentation (Ye et al., 2016) and near-field recordings from only the 2010 Mentawai event, we require models and synthetic data to learn how near-field sensors will respond in the event, and further understanding of the rupture physics that produce such observations is critical input to developing machine-learning algorithms for TEW.

Using synthetic data is necessary to fill in the gaps and allow us to fully explore the parameter space. The paucity of TsE recordings leaves a lot of unknowns when it comes to the physics and rupture parameters that govern TsE ruptures, and thus synthetic earthquakes that are currently used to design, test, and support recent early warning algorithms (e.g., Lin et al., 2021; Williamson et al., 2020) may not accurately model them. The rupture parameters of TsEs need to be constrained to generate TsE-type synthetic scenarios to incorporate into early warning algorithm development and testing.

In this work, we explore the rise time, V_{rupt} , and stress parameter, which are parameters known to be influenced by shallow megathrust properties and have been

observed as characteristically different for TsEs. The rise time is the amount of time it takes for slip to initiate on a patch of the fault and for the healing front to move through. The presence of accreted sediments in the shallow megathrust lends to slower rise times. Melgar and Hayes (2017) include a figure of M_0 versus mean rise time for a synthesized database of recordings from > 150 M7–9 events, four of which are TsEs (1994 and 2006 Java; 1994 Nicaragua; 2010 Mentawai) with mean rise times ranging from ~ 8 –20 s compared to ~ 2 –8 s for typical interplate events. Rupture velocities for standard earthquakes are typically around 1.5–2.5 km/s, but TsEs have shown to have slightly slower rupture velocities ranging from 1–2 km/s (Riquelme et al., 2019) because of the low rigidity rupture domain. Lastly, as TsEs are biased towards low-frequency energy, their recordings tend to have lower corner frequencies, which are proportional to the stress parameter. The stress parameter is a difficult value to constrain, and values span orders of magnitude. However, an acceptable range for typical interface events is 3–12 MPa (Atkinson and Macias, 2009), whereas TsEs often have stress parameters less than 1 MPa (Bilek et al., 2016; Ide et al., 2003; Kikuchi and Kanamori, 2005; Yue et al., 2014). Although previous studies have constrained values of these parameters for several tsunami earthquakes, it is unclear whether their prescription in a simplified model would necessarily generate data with the expected characteristics of TsEs, as the physics governing these models are informed by more typical events. With this study, we determine parameter values needed to simulate TsEs.

2.3 Data and Methods

Our methodology is to (1) produce semistochastic waveforms for 30 random realizations of source models around a mean slip model for the 2010 Mentawai earthquake by varying the rise time, V_{rupt} , and stress parameter, (2) compare intensity measures (IMs: PGA; PGD; time to reach PGD, tPGD; and Fourier amplitude spectra, FAS) from the simulated waveforms to observed strong motion and HR-GNSS time-dependent crustal deformation waveforms, and (3) use the residuals between observed and predicted IMs to determine which combinations of rupture parameters provide the best fit to the observed

data for this event. In this way, we are not producing a single non-unique slip model with a single set of rupture parameters that could describe the data, but rather searching for the average parameters that produce the most definable characteristics of TsEs (long tPGD, average PGD, low PGA, and depleted high frequencies in the FAS). Below, we describe the data and methods we use in the study in more detail.

2.3.1 Geophysical Dataset

Observed Dataset: The 2010 Mentawai TsE was recorded in the near-field on 13 HR-GNSS stations operated by Badan Informasi Geospasial (BIG) and 17 strong motion seismic stations operated by the Agency for Meteorology, Climatology, and Geophysics of Indonesia (BMKG). The number of stations is moderate; however, they have good azimuthal land coverage, and the number of stations is comparable to other megathrust event recordings (Sahakian et al., 2019). Both the seismic and HR-GNSS data come from the Ruhl et al. (2019) database. For this study, we eliminated eight seismic stations that had epicentral distances > 600 km to avoid surface wave bias of PGA, as well as eight HR-GNSS stations due to high noise levels. Our final dataset is shown in Figure 2.1b.

Simulated Dataset: We use a set of semistochastic forward modeling codes, referred to as *FakeQuakes*, to generate our synthetic dataset (Melgar et al., 2016). *FakeQuakes* uses a one-dimensional (1D) velocity model to generate stochastic kinematic rupture models patterned after the 2010 Mentawai event, as well as the associated HR-GNSS and seismic waveforms. We use the Yue et al. (2014) velocity model for this study; however, we add additional columns for P- and S-wave intrinsic attenuation (Q_P and Q_S , respectively). Q is not well resolved in this region, so we constructed a crude 1D attenuation model using relationships between V_P and Q_P for the Hikurangi subduction zone (Eberhart-Phillips et al., 2014; Eberhart-Phillips and Banister, 2015) and V_P for this region. Q_S was approximated as $0.5Q_P$. A detailed explanation of the simulation methodology is provided in the following sections.

Rupture Models: Various slip inversion models exist for the 2010 Mentawai TsE (e.g., Hill et al., 2012; Newman et al., 2011; Satake et al., 2013; US Geological Survey,

2018; Yue et al., 2014; Zhang et al., 2015), but as inversions are non-unique, we have generated a suite of 30 stochastic rupture models using a modified version of the Yue et al. (2014) slip inversion as a mean slip model. This is a special characteristic of *FakeQuakes* compared to other stochastic rupture modeling codes (e.g., Graves and Pitarka, 2010, 2015; LeVeque et al., 2017) that was first introduced by Goldberg and Melgar (2020). *FakeQuakes* allows one to generate rupture models that are perturbations of a background mean model. In this case the mean model is a log-frequency derived slip inversion. This approach has been validated by Small and Melgar (2023) which found simulations to produce realistic rupture patterns when compared with large historical events. Figure 2.2 shows a comparison between the mean slip model and four example stochastic models. (See supporting Figure 2.S1 for all 30 stochastic slip models). The Yue et al. (2014) model is preferred because their inversion used the most comprehensive dataset (near-field HR-GNSS recordings, teleseismic body waves, and tsunami recordings from deepwater buoys and tide gauges) and best resolves all observational data. This particular slip inversion is similar to the other inversions in geometry and slip directivity; however, it finds peak slip near the trench to be higher (> 20 m as opposed to 6–12 m), and it places the hypocenter at a shallower depth (~ 12 m as opposed to 13–20 m). Yue et al. (2014) provides a more in-depth comparison of the models.

The original Yue et al. (2014) model was coarsely discretized, characterized with only 72 subfaults with strike-slip and dip-slip dimensions of $14.25 \text{ km} \times 15 \text{ km}$. Using this as an initial model, we made a finer-discretized slip model containing 2886 subfaults with strike-slip and dip-slip dimensions of $2 \text{ km} \times 2 \text{ km}$. (Fig. 2.S2). We further modified the mean slip model by shifting the depth of the subfaults 3 km shallower. The Yue et al. (2014) study used their 1D velocity profile and slip inversion to generate tsunami scenarios, thus a 3 km water layer (i.e. $V_S = 0 \text{ km/s}$) is assumed in their models. *FakeQuakes* cannot accept a V_S layer equal to 0 km/s, so this layer was removed from the velocity model, and 3 km were subtracted from the fault model depth to maintain consistency between the fault and velocity models.

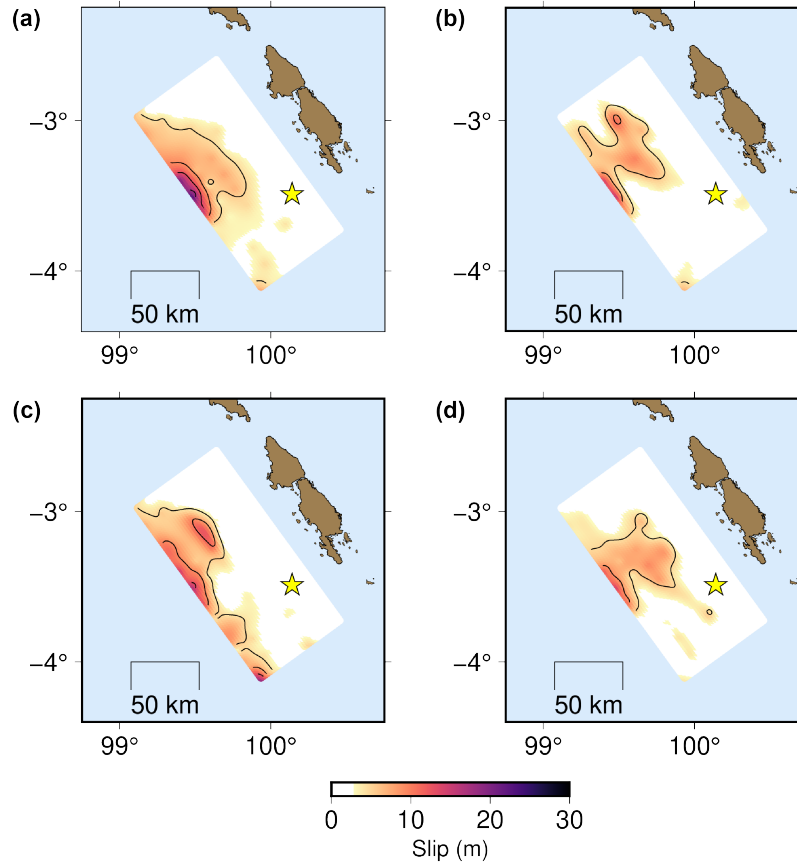


Figure 2.2. (a) Finely discretized slip model of the 2010 Mentawai tsunami earthquake from Yue et al. (2014) and (b–d) stochastic variations of the slip model in (a) generated using *FakeQuakes*. 5 m contours from the slip patterns are shown.

The stochastic models follow a similar methodology to Mai and Beroza (2002), which assume a von Karman correlation function for slip correlation amongst subfaults. The stochastic slip vector fields are determined using Karhunen Loève (K–L) expansion (LeVeque et al., 2016), where the range of possible slip on a subfault is taken from a probability density function defined by a mean value equal to the slip on the equivalent Yue et al. (2014) model subfault. Details of this process are in Melgar et al. (2016) and Melgar and Goldberg (2020).

HR-GNSS displacement waveforms are generated deterministically with a 2 Hz sampling frequency by combining Green’s functions computed following the matrix

propagator method of Zhu and Rivera (2002) with slip on the fault. Real HR-GNSS data generally have higher noise arising from uncertainty in the position inversion during processing, electromagnetic signals in the ionosphere and troposphere, and multi-pathing (Melgar et al., 2020). Unlike seismic noise, HR-GNSS noise is often in the frequency range of the signal and is thus difficult to filter out. It is reasonable then for observed PGD to have some noise bias, so we add real noise from the observed HR-GNSS recordings to the simulated HR-GNSS data for equal comparison. For each simulated recording, 512 s of noise are taken from the corresponding station observed time series, ending approximately 2.5 minutes before the origin time. The observed HR-GNSS recordings have a sampling rate of 1 Hz, so we first decimate the simulated data for time-domain consistency when combining the two sets of data.

Strong motion acceleration waveforms are generated semistochastically with a 100 Hz sampling rate in a three-step process, closely following the broadband simulation methodology from Graves and Pitarka (2010, 2015). The low-frequency (< 1 Hz) portion of the waveforms are first generated deterministically using the same methodology as with the HR-GNSS waveforms. The high-frequency (> 1 Hz) portion of the waveforms are then generated in the frequency domain using the following equation,

$$A_i(f) = \sum_{j=1,M} C_{ij} S_i(f) G_{ij}(f) P(f) \quad (2.2)$$

where $A_i(f)$ is the amplitude spectrum for subfault i , $j = 1, M$ is the summation of different rays, C_{ij} is the radiation scale factor, $S_i(f)$ is the source spectrum, $G_{ij}(f)$ is the path attenuation term, and $P(f)$ is the site-specific high-frequency attenuation. The individual subfault spectra are summed, and stochastic white noise is added to represent the uncertainty in fine-scale heterogeneity of the subsurface. Finally, the low-frequency waveforms are double differentiated from displacement to acceleration and combined with the acceleration high-frequency waveforms using an acausal matched 4th order Butterworth filter. We deviate from Graves and Pitarka (2010) by employing a separate filter corner frequency (f_c) for the low-frequency and high-frequency data, rather than using a common filter f_c of 1 Hz, because we find that the latter results in an artificial notch in the broadband

spectra between $\sim 0.1\text{--}1$ Hz. The notch appears to be a result of early roll-off of the low-frequency spectra, which is not observed in the high-frequency spectra, rather than a result of the use of a common filter f_c . We find that using a 1 Hz lowpass filter of the low-frequency data and 0.1 Hz highpass filter of the high-frequency data minimizes the notch (Figs. 2.S3–2.S4).

Figure 2.3 shows example simulated waveforms and their Fourier amplitude spectra (FAS) compared with observed for a subset of stations. Note that these simulated data were generated using typical megathrust parameters and are not expected to closely match the observed.

All the observed data (seismic and geodetic) contain significant low-frequency ringing, but the source of the ringing is unclear. We attempted to account for this by including some shallow low-velocity layers in the velocity model. However, this only added low-frequency energy at certain peaks, rather than over the spectrum of missing frequencies (Fig. 2.S5). Because we see this effect across both types of stations and at all locations, it is likely a source effect, and one we are unable to recreate using a simple 1D semistochastic model. To avoid this biasing our results, we only evaluate acceleration FAS at frequencies above 0.1 Hz. We have also excluded displacement FAS from our analyses as these are dominated by the range of frequencies characterizing the ringing and are minimally affected by the variation of parameters.

2.3.2 *Varying Rupture Parameters*

Rise Time: The subfault slip duration, or rise time (T_i), is proportional to the square root of the subfault slip (s_i) (Melgar et al., 2016). Graves and Pitarka (2010) suggest a doubling of rise time at shallow depths to represent a reduction in slip speed observed with surface rupturing events. Following Melgar et al. (2016), we use a shallow depth of 10 km, rather than the 5 km proposed by Graves and Pitarka (2010) as their study focused

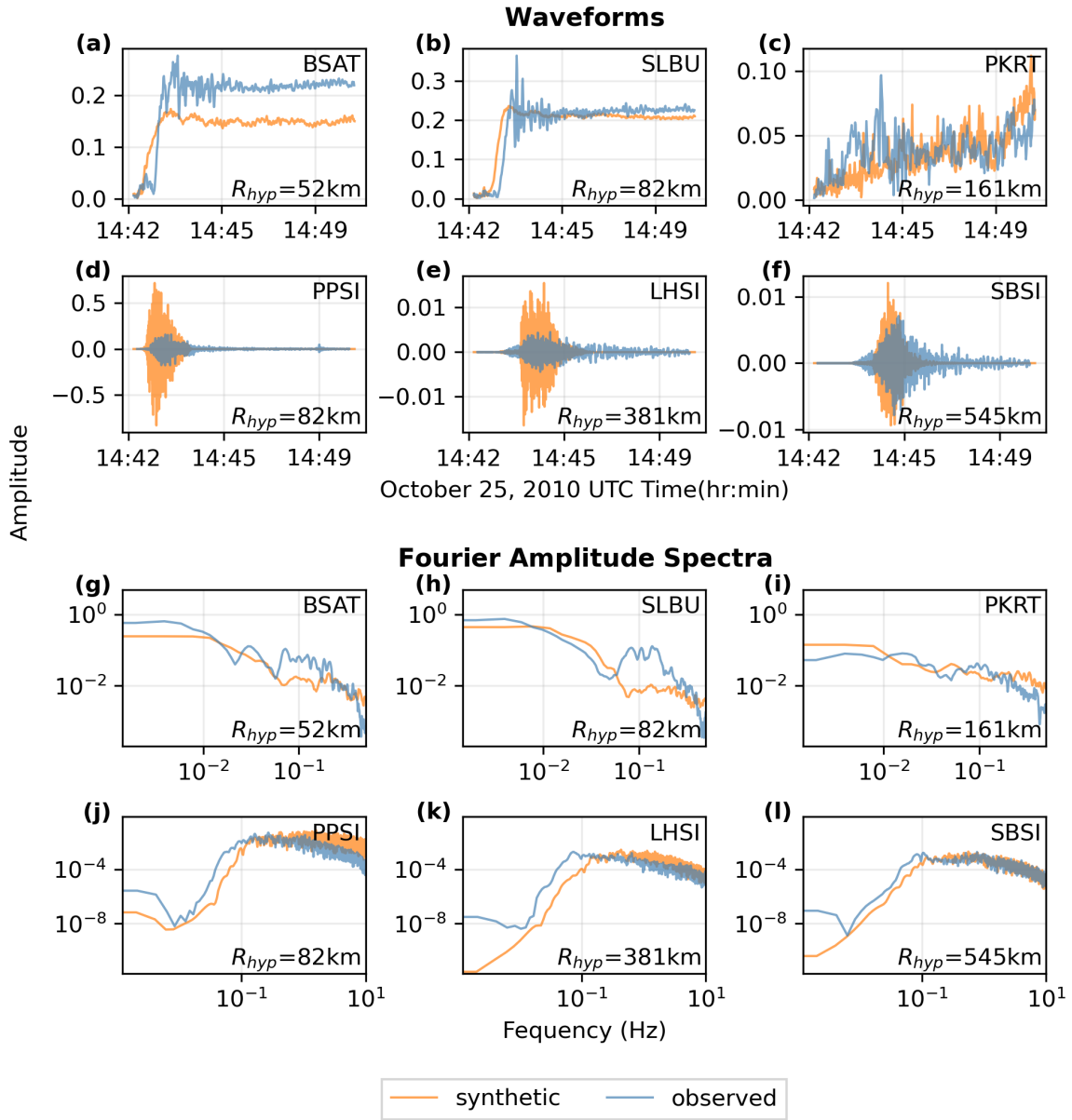


Figure 2.3. (a–f) Waveform comparisons between the observed (blue) data and synthetic (orange) data using typical megathrust parameters. (a–c) Displacement waveforms for strong motion seismic stations PPSI, LHSI, and SBSI, respectively. (d–f) Acceleration waveforms for HR-GNSS stations BSAT, SLBU, and NGNG, respectively. (g–l) Fourier amplitude spectra (FAS) comparisons for the waveforms in (a–f). (g–i) Displacement FAS for seismic stations PPSI, LHSI, and SBSI, respectively. (j–l) Acceleration FAS for HR-GNSS stations BSAT, SLBU, and NGNG, respectively.

on crustal ruptures rather than subduction interface events.

$$T_i = \begin{cases} 2ks_i^{1/2}, & d < 10km \\ ks_i^{1/2}, & d > 15km \end{cases} \quad (2.3)$$

The scaling factor, k , is applied to ensure the total mean rise time (T_a) fits a rise time– M_0 scaling relation. The Melgar and Hayes (2017) dataset give the following relation,

$$T_a = 4.226 \times 10^{-8} \times M_0^{0.293}. \quad (2.4)$$

Existing rise time– M_0 relationships are biased towards typical megathrust ruptures, but it is evident from Figure 1 of Melgar and Hayes (2017) that TsEs deviate from the norm, having significantly longer rise times compared to their M_0 . As there are only four TsEs in their dataset, it is difficult to constrain a unique relationship for these events, but we can modify k to reflect variations in the scaling by multiplying the individual subfault rise times by a “ k -factor”. In this study, we evaluate k -factors ranging from 1–3, and hence rise times ranging from 5.4–16.2 s.

Rupture Velocity: The rupture propagation speed has been found to be proportional to V_S , where $V_{rupt} = SF \times V_S$. Here, SF is the shear wave fraction. *FakeQuakes* uses this value as the background V_{rupt} , but spatial heterogeneity in rupture propagation is accounted for by perturbing the onset times to be earlier where slip is large and later where slip is small (Graves and Pitarka, 2010; Melgar et al., 2016). A high-resolution velocity profile does not exist in the Mentawai, Indonesia region, so our velocity model does not adequately capture the thick, weak accretionary wedge. Even fine scale resolution would not yield reliable results because 1D semistochastic models inherently assume a single velocity model is representative of the entire region (source to site). The weak accretionary wedge would inevitably be included in site-specific components of the model as well. Instead, we find a better method to be to modify the shear wave fraction, as this only affects the rupture.

FakeQuakes follows a bimodal shear wave fraction framework to account for the presence of a weak shallow megathrust (Graves and Pitarka, 2010). An SF value of 0.8 is generally assumed for deeper depths (> 15 km for our model), whereas *FakeQuakes* uses

a value of 0.49 for shallower depths (< 10 km for our model). The shear wave fraction at depths between 10–15 km is a linear interpolation between the deep and shallow shear wave fractions. In this study, we only modify the shallow shear wave fraction (SSF). To determine an average SSF value per rupture, we find that multiplying the SSF by 3.25 results in approximately the mean V_{rupt} for any value of SSF . Although the mean V_{rupt} , and thus SSF , of the stochastic simulations varies slightly amongst themselves, they are still usually within 0.05 km/s of each other. Thus, we determine the range of SSF 's needed to have the desired range in V_{rupt} by dividing these V_{rupt} values by 3.25. In this study, we evaluate SSF ranging from 0.3–0.49, and hence V_{rupt} ranging from 1–1.6 km/s.

High-Frequency Stress Parameter: The stress parameter, (Brune stress; Brune, 1970) is the simplest parameter to modify because it can be directly defined in the parameter file. To simulate lower stress parameter events, we evaluate stress parameters ranging from 0.1–2.0 MPa, with 5 MPa used as the standard value for comparison.

2.3.3 *Constraining Rupture Parameters*

Through varying the rise time, V_{rupt} , and stress parameter, we aim to constrain the parameter space that best recreates the 2010 Mentawai TsE by minimizing the residuals of a suite of characteristic IMs. We want to capture the slow slip, long duration, and energy deficient nature of TsEs, whilst preserving the seismic moment. Thus, we evaluate the PGD, tPGD, PGA, and acceleration FAS bin averages. For PGD, PGA, and the binned FAS, we compute the residuals as $\ln(\text{observed}/\text{simulated})$, and for tPGD we compute the residuals as $\text{observed} - \text{simulated}$. The acceleration FAS are binned into 10 log-spaced bins ranging 0.1–10 Hz. The upper bound of 10 Hz is a limitation of seismic data processing prior to our acquisition of the data.

We hypothesize that the PGD residuals will have a zero mean and will be invariant to variations in the rupture parameters as it is dependent primarily on the static slip and subsurface structure, rather than the kinematics. We anticipate tPGD residuals to be positive and decrease in magnitude with both increasing rise time and decreasing V_{rupt} . Both these parameters should slow down the rupture, thus increasing the time it takes for

the waves to reach the stations. Because we do not include any complex site amplification, PGA should be derived from higher frequencies, thus we expect PGA and the acceleration FAS bin averages to be negative and decrease in amplitude by decreasing the stress parameter.

2.4 Results

2.4.1 Isolated Parameters

The parameters are varied initially through isolated tests, where only one parameter is varied at a time, and compared with scenarios generated using standard megathrust parameters. For these tests, only a few values of each parameter are evaluated to gauge how the IMs are affected by the varying rupture parameter, and to get a coarse idea which parameter values are ideal for recreating the Mentawai scenario. The standard stochastic simulations have a mean rise time of 5.4 s and mean V_{rupt} of 1.6 km/s.

For varying rise time, we evaluate simulations using a scaling factor of $1k$, $2k$ and $3k$ (i.e., mean rise times of 5.4, 10.8 s and 16.2 s) (Fig. 2.4). Modifying the rise time primarily influences the timing of shaking (i.e., tPGD). As expected, using the standard $1k$ results in positive tPGD residuals because the simulated data have significantly shorter moment-rates. With increased mean rise time, the mean amplitude of the residuals decreases, and we find that if only rise time is modified, a value close to 11 seconds is needed to yield a tPGD residual of 0.

For varying V_{rupt} , we evaluate simulations using SSF values of 0.49, 0.4, and 0.3 (i.e., mean V_{rupt} of 1.6, 1.3, and 1.0 km/s) (Fig. 2.5). As with rise time, tPGD primarily affected when varying V_{rupt} , and using the standard SSF of 0.49 results in positive tPGD residuals. The residual mean amplitude is reduced using slower rupture propagation speeds, and we find a V_{rupt} close to 1.3 km/s is needed to yield a tPGD residual of 0 if only V_{rupt} is modified. These values for mean rise time and V_{rupt} are our endmember values of the parameter space. Because we will be varying the rise time and V_{rupt} together, we expect the ideal values to fall between a rise time of 5.4–11 s and a V_{rupt} of 1.3–1.6 km/s.

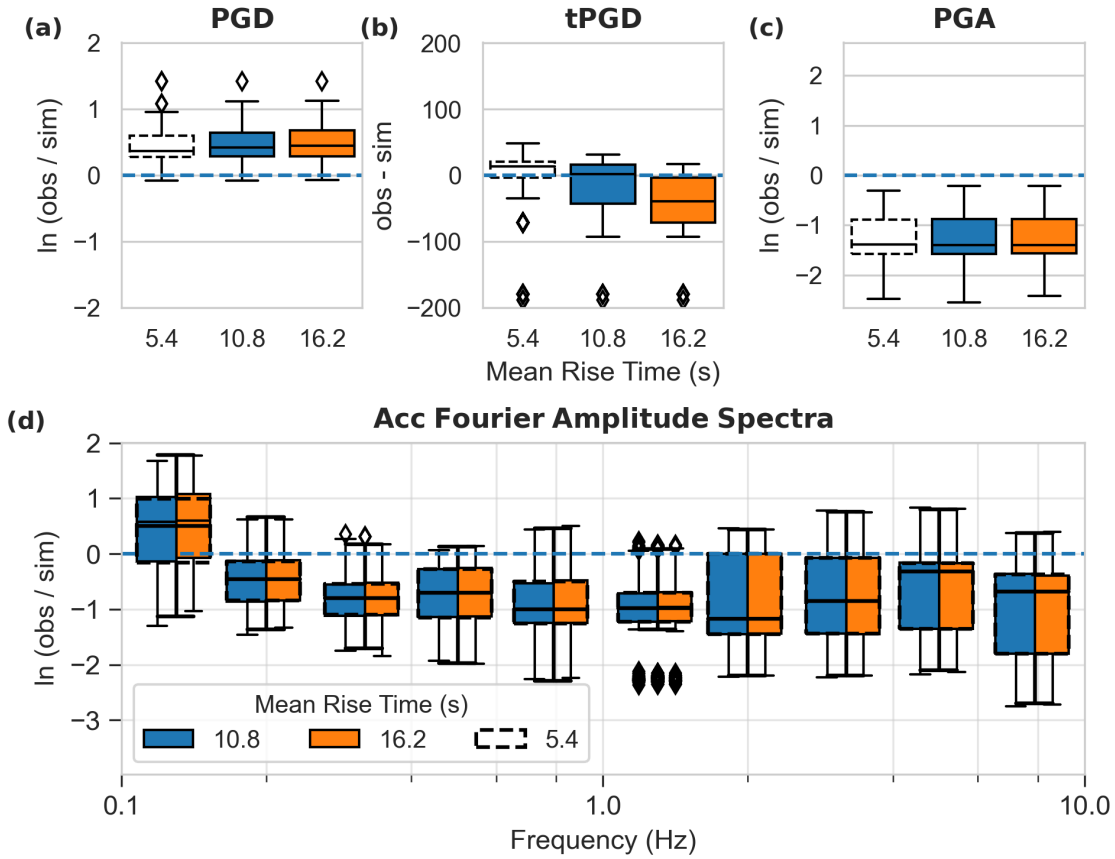


Figure 2.4. Effects of the varying rise time on (a) PGD, (b) tPGD, (c) PGA, and (d) acceleration FAS. The dashed boxes represent residuals using standard rise time– M_0 scaling relations (k from the Melgar and Hayes 2017 study), resulting in a mean rise time of 5.4 s. The orange boxes represent rise times computed using $2k$, resulting in a mean rise time of 10.8 s, and the blue boxes represent rise times computed using $3k$, resulting in a mean rise time of 16.2 s.

For varying stress parameter, we evaluate simulations using stress parameters of 0.1, 1, and 2 MPa with 5 MPa being the standard value for comparison. We find that unlike rise time and V_{rupt} , the stress parameter primarily influences the high-frequency intensity measures (i.e., PGA and acceleration FAS) (Fig. 2.6). Using a typical value of 5 MPa results in negative residuals for all the high-frequency IMs, suggesting the simulated data have more high-frequency energy than the observed. However, decreasing the stress parameter decreases the negative bias of the residuals, and we find an ideal value to be

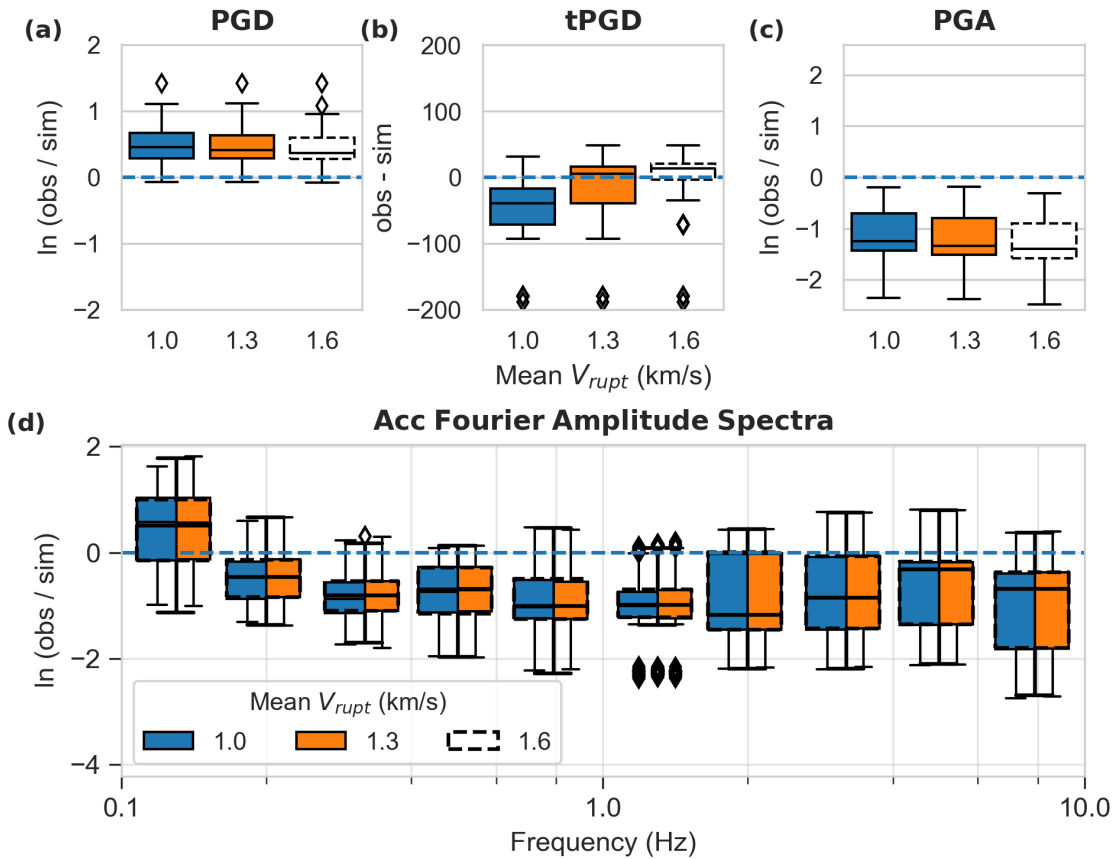


Figure 2.5. Effects of varying the rupture velocity on (a) PGD, (b) tPGD, (c) PGA, and (d) acceleration FAS. The dashed boxes represent residuals using a standard shallow shear wave fraction of 0.49, resulting in a mean V_{rupt} of 1.6 km/s. The orange boxes represent V_{rupt} computed using a shallow shear wave fraction of 0.4, resulting in a mean V_{rupt} of 1.3 km/s, and the blue boxes represent V_{rupt} computed using a shallow shear wave fraction of 0.3, resulting in a mean V_{rupt} of 1.0 km/s.

close to 1 MPa when only considering PGA., or close to 2 MPa when also considering the acceleration FAS bin residuals.

2.4.2 Covaried Parameters

The IM residual results from the previous section illustrate overall trends in the data resulting from variations in the rise time, rupture velocity, and stress parameter. We use these results to guide the design of a larger dataset, which provides a finer sampling of the three-dimensional (3D) parameter space and a better constraint on the ideal values for

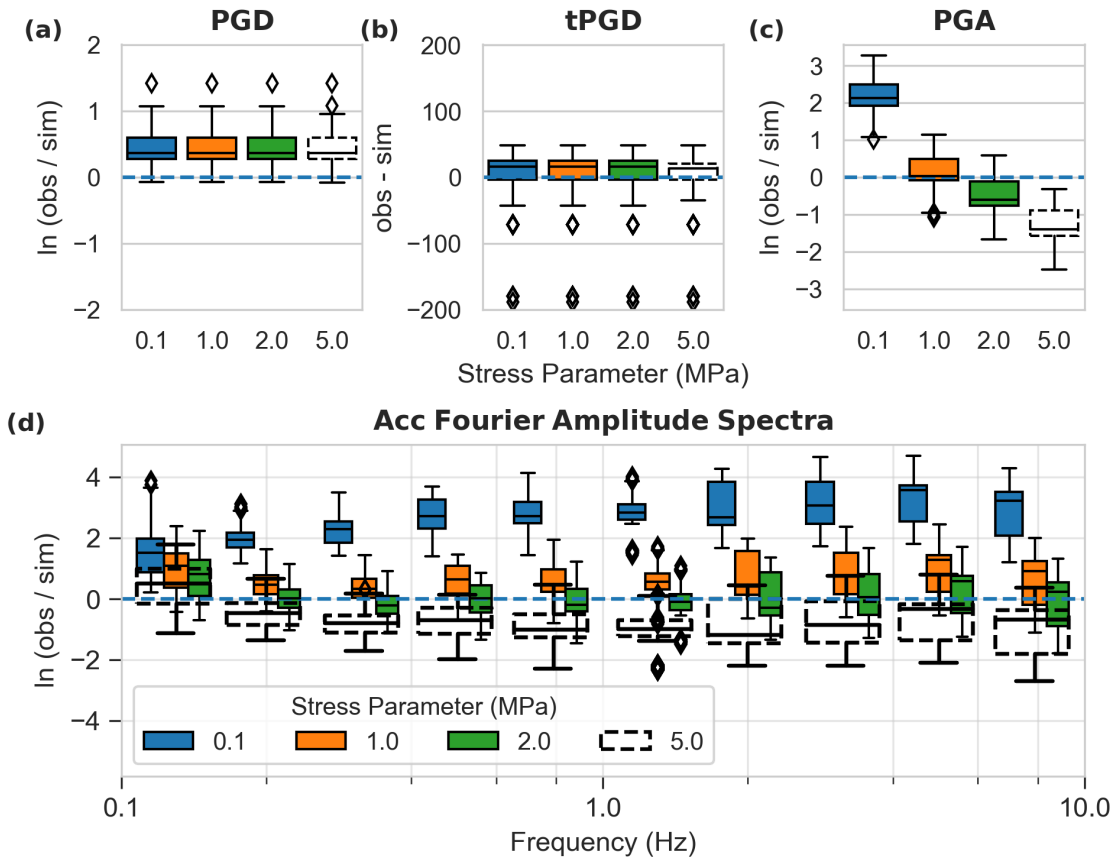


Figure 2.6. Effects of varying the stress parameter on (a) PGD, (b) tPGD, (c) PGA, and (d) acceleration FAS. The dashed boxes represent residuals using a standard stress parameter of 5.0 MPa. The green boxes represent residuals using a stress parameter of 2.0 MPa, the orange boxes represent residuals using a stress parameter of 1.0 MPa, and the blue boxes represent residuals using a stress parameter of 0.1 MPa.

simulating TsE scenarios.

We extend the full parameter space surveyed beyond the approximated endmember values to ensure resolution at the edges of the ideal parameter space. Considering the endmember rise time value is around 11 s, we run simulations spanning the rise time parameter space of 5.4 s to 12.4 s by using multiplication factors ranging from 1–2.3 at intervals of 0.1 (parameter space resolution of 0.54 s). Considering the endmember V_{rupt} is around 1.3 km/s, we run simulations spanning the V_{rupt} parameter space from 1.2 to 1.6 km/s by using SSF values ranging from 0.37–0.49 at intervals of 0.03 (parameter space

resolution of 0.1 km/s). Because stress parameter can vary orders of magnitude, we do not find it advantageous to run large suites of simulations with a fine parameter sampling. Instead, we run simulations using the following values: 0.1, 0.5, 1.0, 1.5, 2.0, and 3.0 MPa.

Initially, we intended to survey the full 3D parameter space by simulating data for each possible combination of rise time, rupture velocity, and stress parameter. Findings from the isolated parameter tests (Figs. 2.4–2.6) reveal, however, that the effects of varying the rise time and rupture velocity are essentially separate from effects of varying the stress parameter. We instead split the parameter space into a two-dimensional (2D) survey of the rise time and rupture velocity and a 1D survey of the stress parameter, which significantly decreases computation by reducing the number of parameter combinations, and allows for simpler regression of the ideal parameter space. Our final regression dataset consists of simulated data for 76 combinations of parameters with 30 stochastic scenarios per combination, resulting in a total of 2280 simulations.

2.4.3 Tsunami Earthquake Parameter Regression

First evaluating the effects of individually varying the rise time, rupture velocity, and stress parameter, we find that varying these rupture parameters in a simulation model produces the expected IM residual trends, which validates their classification as key rupture parameters for TsEs. Our ultimate objective of this study is to present the mean and standard deviation of these parameters to govern simulation of TsE scenarios. Because we discretely sampled our parameter space, we obtain these values by performing Gaussian process regression (GPR) of the IM residuals.

GPR is a supervised statistical learning method which computes nonparametric probabilistic models to fit a dataset. In simple terms, we assume an infinite number of functions can fit an observed set of data, and GPR computes the most probable, or mean, function along with its uncertainty. For this study, we perform a 2D GPR of the rise time–rupture velocity parameter space and a 1D GPR of the stress parameter parameter space using the Scikit-learn Python library (Pedregosa et al., 2011). The simulation IM residual results of the regression dataset are used as a prior for the GPR, along with a

kernel to describe the covariance of the random process. We assume the commonly used radial basis function (RBF) kernel with a fixed length scale of 1.0, due to its universality against most functions.

We compute a summary residual parameter for tPGD (δ_{tPGD}) for each event i , k -factor k , and SSF l , which is a median (*med*) of the j station tPGD residuals, given the specific stochastic rupture and rupture parameter combination. The median is preferred over the mean to reduce bias from a possible poorly modeled station. This results in 2100 summary tPGD residual parameters (5 values of SSF \times 14 values of k -factor \times 30 stochastic scenarios).

$$\delta_{tPGD}(k\text{-factor}_k, SSF_l)_i = \text{med} \begin{bmatrix} \delta_{tPGD}(k\text{-factor}_k, SSF_l)_{j=1} \\ \vdots \\ \delta_{tPGD}(k\text{-factor}_k, SSF_l)_{j=5} \end{bmatrix} \quad (2.5)$$

We also compute a summary residual parameter for the high-frequency IMs (δ_{HF}) for each event i and stress parameter (σ) k . Each row in the δ_{HF} vector represents a single station residual and is computed as the average of the PGA residual (δ_{PGA}) and the mean FAS residual ($\overline{\delta_{FAS}}$). The FAS residuals are averaged prior to being averaged with δ_{PGA} to ensure equal weighting of IMs. This results in 180 summary high-frequency residual parameters (5 values of σ \times 30 stochastic scenarios).

$$\delta_{HF}(\sigma_k)_i = \text{med} \begin{bmatrix} 0.5(\delta_{PGA}(\sigma_j) + \overline{\delta_{FAS}(\sigma_j)})_{j=1} \\ \vdots \\ 0.5(\delta_{PGA}(\sigma_j) + \overline{\delta_{FAS}(\sigma_j)})_{j=9} \end{bmatrix} \quad (2.6)$$

The results of the 2D GPR of rise time and rupture velocity and the 1D GPR of stress parameter are shown in Figure 2.7. In Figure 2.7a, the red regions indicate parameter combinations resulting in a tPGD that is too early, which is consistent with shorter rise times and slower rupture velocities, whereas the blue regions indicate parameter

combinations with a tPGD that is too late. The white band, which highlights the ideal range of parameter combinations mostly likely to result in a δ_{tPGD} of 0 s, is nearly linear and extends from a rise time–rupture velocity combination of about 5.4 s–1.23 km/s to 12 s–1.6 km/s. The standard deviations from the 2D regression highlighted in Figure 2.7b are relatively uniform because they are conditioned upon the spatial sampling of the prior, which was evenly sampled by our dataset. Although uncertainty estimates are a desirable feature of GPR, these provide a measure of uncertainty in the values of the function, rather than uncertainties in parameter coordinates resulting in a single function value. Therefore, we present standard deviations in rise time and rupture velocity as the standard deviation of all tPGD residuals from the GPR mean predictions, which are represented by the dashed black lines. Fitting a line to the linear segment of the 0-residual line, we obtain the following equation which can be used to select reasonable values of rise time (T_a) and V_{rupt} for simulating TsEs:

$$V_{rupt} = 0.056(T_a) + 0.85 \pm 1.5 \text{ km/s.} \quad (2.7)$$

The uncertainty of 1.5 is taken as the approximate vertical distance between the 1-standard deviation lines and the 0-residual line.

In Figure 2.7c, positive residuals indicate stress parameters resulting in strong motion waveforms lacking high-frequency energy compared to the observed 2010 Mentawai event, whereas negative residuals indicate stress parameters biased towards high-frequency energy. The Nelder-Mead method of the SciPy optimization Python library (Virtanen et al., 2020) is used to solve for the ideal stress parameter resulting in a mean δ_{HF} of 0, which is found to be 1.43 MPa. As with the 2D regression, we present the standard deviation in stress parameter as the standard deviation of the total high-frequency residuals from the GPR mean predictions. Stress parameter values within approximately +/- 1 MPa result in residuals within the 1-standard deviation lines; thus, we present the following equation for selecting a reasonable TsE stress parameter:

$$\sigma = 1.43 \pm 1.25 \text{ MPa.} \quad (2.8)$$

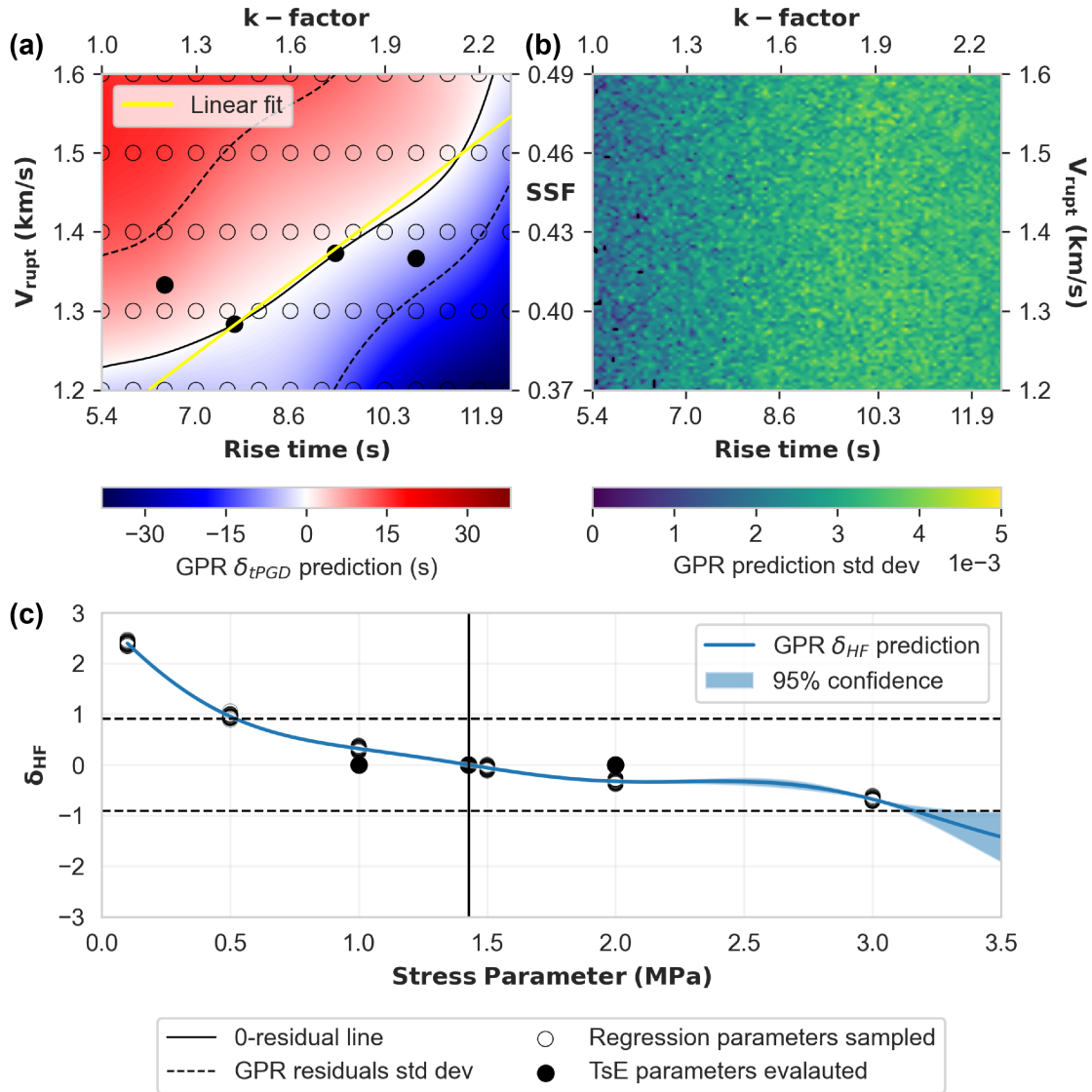


Figure 2.7. (a) 2D Gaussian process regression (GPR) of tPGD residuals (δ_{tPGD}) of simulated data using varying rise time and V_{rupt} simulation parameters (k -factor and SSF , respectively). (b) Standard deviation of the GPR mean function in (a). (c) 1D GPR of high-frequency intensity measure residuals (i.e., mean of PGA and acceleration FAS residuals) for the simulation stress parameter. The mean function and 95% confidence are indicated by the solid blue line and shaded regions, respectively. For all subplots, the solid black line represents the parameter(s) where the mean function has a residual of 0, and the dashed black lines represent \pm one standard deviation of the GPR prediction residuals. The open circles indicate the parameter space surveyed by the simulation dataset, where each discrete datapoint contains mean residuals for each of the 30 stochastic events. The filled circles indicate the tsunami earthquake (TsE) parameter space surveyed in a final dataset, but were not used for the GPR.

In theory, we expect simulations run with parameters falling within one standard deviation of the mean model to be reasonably representative of a real TsE. To validate this, the results of the two regressions are used to inform the parameters of a final dataset representing likely TsE scenarios. We generate a new set of 30 stochastic slip models for this dataset following similar standard practices of machine learning where separate data are used for testing and training. This TsE dataset consists of six parameter configurations: two using mean values from the 0-residual GPR predictions, and four using values within the one standard deviation bounds. Table 2.1 lists the values of parameters used for these simulations, and the parameter space coordinates of the rise times, rupture velocities, and stress parameters are indicated on the regression plots in Figure 2.7. We find that the simulated TsE scenarios adequately model the observed IM's of the Mentawai event, as indicated by the near-zero residuals of the IM's in Figure 2.8. The source time functions (*STFs*) of the simulations are also evaluated, and we find reasonable similarity with the observed *STF*.

2.5 Discussion

2.5.1 Simulation Limitations

Through various permutations of rise time, rupture velocity, and stress parameter, we capture key features of the observed data in the simulations, namely slow moment-rate (i.e., large tPGD) and high-frequency deficiency (i.e., low PGA and high-frequency amplitudes of acceleration FAS). There are, however, aspects of the data we are unable to reconstruct. The observed displacement and acceleration data both exhibit significant ringing, which is likely a source effect due to its presence in recordings of all HR-GNSS and strong motion stations. It is unclear whether this is a unique property of all TsEs, a result of rupturing the compliant wedge in an accretionary setting, or an isolated occurrence due to the complex 3D crustal structure of the Mentawai region. Until more near-field TsE recordings exist or a high-resolution and high-fidelity velocity model is available in the region, we cannot definitively answer this question. Regardless, this appears to have little

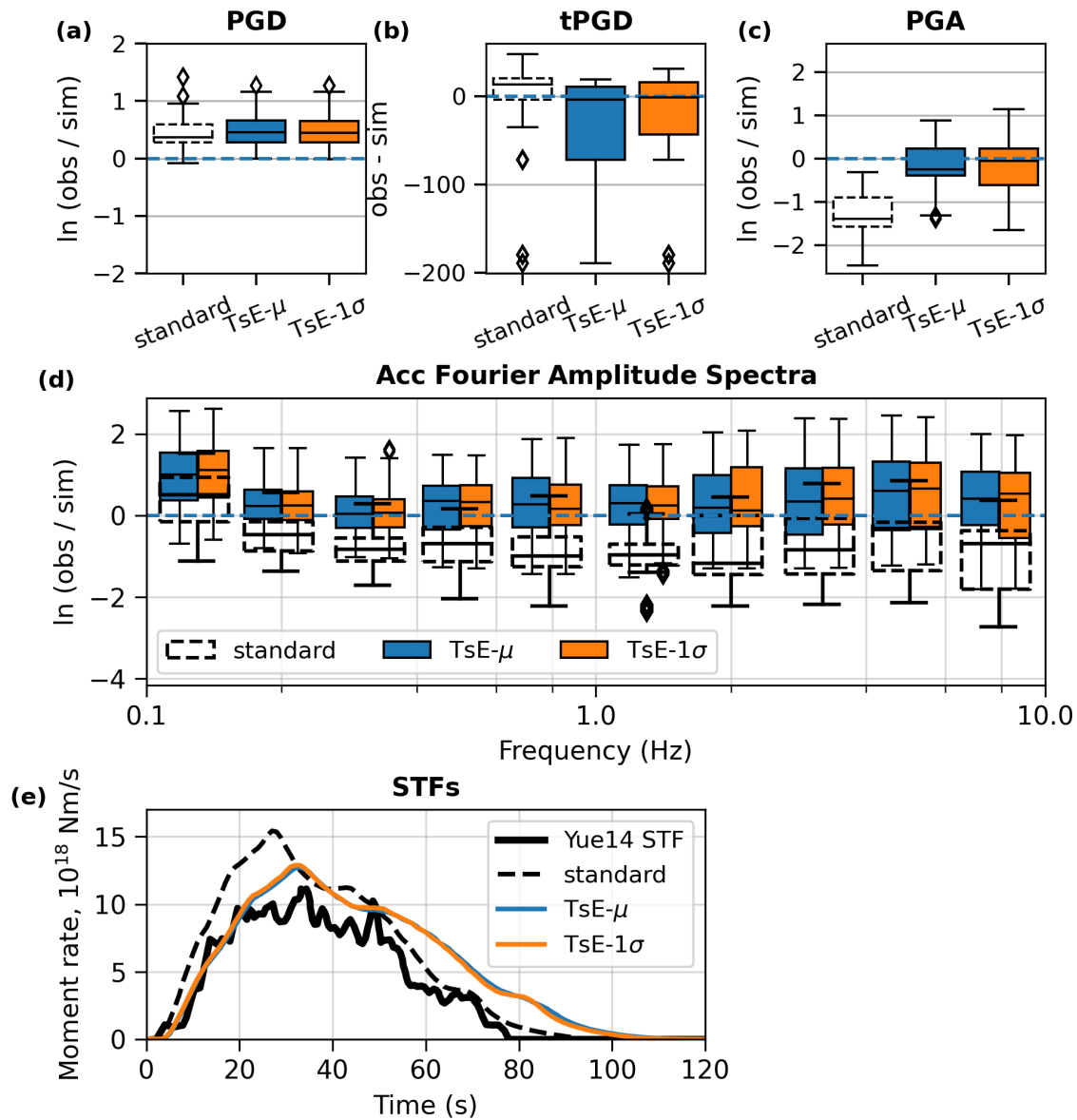


Figure 2.8. Intensity measure residual analyses comparing observed and simulated data generated using standard parameters, mean values from the TsE parameter regression (TsE- μ), and values within one standard deviation of the mean from the TsE parameter regression (TsE- 1σ) for (a) PGD, (b) tPGD, (c) PGA, and (d) acceleration FAS. (e) We also compare the average STFs for the simulated data using TsE parameters and standard parameters with a STF computed using observed data (Yue et al., 2014). All IMs and the STFs better fit observed data when using recommended TsE parameters.

effect on the IMs necessary for real-time analysis.

Interestingly, a systematic under-generation of PGD is still observed for all HR-

Table 2.1. Tsunami earthquake simulation rupture parameters.

Parameters ^a	k -factor	Rise time (s)	SSF	V_{rupt} (km/s)	Stress Parameter (MPa)
μ -a	1.42	7.67	0.40	1.30	1.43
μ -b	1.74	9.40	0.42	1.37	1.43
1σ -a	1.2	6.48	0.41	1.33	1.0
1σ -b	1.2	6.48	0.41	1.33	2.0
1σ -c	2.0	10.80	0.42	1.37	1.0
1σ -d	2.0	10.80	0.42	1.37	2.0

^aParameters μ -a and μ -b refer to two sets of parameter combinations using mean values of the rupture parameters from the TsE parameter regression. Parameters 1σ -a – 1σ -d refer to four sets of parameter combinations using values of rupture parameters within one standard deviation from the TsE parameter regression.

GNSS simulated data, regardless of rupture parameter configuration. Albeit, the scale of the discrepancy is minute, with the residuals averaging less than 0.5 natural log unit. It is possible that the source of the ringing in the observed HR-GNSS data also amplified observed PGD. However, this phenomenon has been observed in other studies employing similar 1D forward model methodology (e.g., Fadugba et al., 2023, Preprint; Melgar et al., 2016). Fadugba et al. (2023, Preprint) find that using 3D rather than 1D velocity structure significantly improves PGD residuals, as complex 3D effects are not fully captured by the simple 1D profiles. Furthermore, the 1D velocity profile is considered uniform across the entire region. We should expect the structure to differ between the subduction interface (characterized by cold, dense oceanic crust and weak accretionary material) from that of younger crust where the stations are located. 3D models can better capture these lateral variations, but they are more computationally expensive to run. Thus, their implementation reduces the size of the resulting dataset— a non-insignificant loss in the field of early warning where large datasets are desired for robust testing. Given that the amplitudes of the PGD residuals are small, we maintain that the simulation results from this study are a valuable input for future tsunami early warning testing.

2.5.2 Comparison with Previous Inversion Studies

Through stochastic forward simulations exploring a range of rupture parameters, we have characterized source features of the 2010 Mentawai event, including the rupture

parameters rise time, rupture velocity, and stress parameter, as well as the source duration. We find that our results are consistent with source inversion studies of the Mentawai event. Although we did not solve for a single rise time and rupture velocity, we obtained a range of values capable of replicating the observed tPGD. Melgar and Hayes (2017) report the mean rise time to be ~ 8 s, and other studies have found the rupture velocity to range from ~ 1.25 – 2 km/s (e.g., Lay et al., 2011; Newman et al., 2011; Yue et al., 2014; Zhang et al., 2015). These parameter configurations fall within one standard deviation of the mean prediction for zero residuals from Figure 2.7. We find that a simulation stress parameter of 1.43 MPa best captures the PGA and acceleration FAS amplitudes, which is slightly overestimated compared to the 0.9 MPa estimate from Yue et al. (2014). However, strong motion FAS were not considered in their inversion. If we were to perform the 1D GPR solely on PGA, we would likely obtain a similar result. The average source durations of the TsE simulations from the final dataset are about 110 s, which is well within the 80–125 s range of other studies (e.g., Lay et al., 2011; Newman et al., 2011; Yue et al., 2014; Zhang et al., 2015).

2.5.3 *Real-time Analysis*

A real-time analysis algorithm using PGD- and PGA-derived magnitudes was proposed by Sahakian et al. (2019) for distinguishing TsEs. This algorithm builds upon the premise that TsEs are energy deficient, thus having characteristically low PGA compared to PGD. Therefore, an event magnitude estimate using PGA as a proxy (\mathbf{M}_{PGA}) would underestimate the true magnitude, which can be approximated using PGD (\mathbf{M}_{PGD}). In their study, they computed $\mathbf{M}_{PGA} - \mathbf{M}_{PGD}$ for a handful of moderate-to-large events, including the 2010 Mentawai TsE, and found that typical events had an average $\mathbf{M}_{PGA} - \mathbf{M}_{PGD}$ of 0, whereas the Mentawai event had a $\mathbf{M}_{PGA} - \mathbf{M}_{PGD}$ close to -1. Here, we perform the analysis on the final TsE dataset to ensure a similar result with the simulations.

\mathbf{M}_{PGD} is computed by minimizing a least squares problem governed by a PGD ground-motion model (GMM). Rather than predicting PGD given some magnitude and distance, we invert for the magnitude using known PGD and distance parameters for the set

of HR-GNSS stations. We have chosen to use the Goldberg et al. (2021) model because this model employs the generalized mean rupture distance (R_p) as the distance metric, which better describes a station's relation to the primary slip of an event compared to hypocentral distance and is better suited for close distances to a finite rupture compared to minimum distance metrics. This model was also regressed on a dataset consisting of both observed data and data simulated using *FakeQuakes*.

Models for PGA are functionally more complex than for PGD, and the parameters used are not as well constrained as magnitude or distance, especially in a real-time application. We instead determine \mathbf{M}_{PGA} by finding the event magnitude whose GMM mean PGA and variance best fit the simulated PGA. We do this by first predicting PGA using the Zhao et al. (2006) GMM for a wide range of earthquake magnitudes ($\mathbf{M}5.5$ – 8.5 sampled every 0.1 magnitude unit). We next compute PGA residuals (simulated–predicted) for each of the trial magnitudes. Finally, we perform a series of two-sample Kolmogorov–Smirnov (K–S) tests to determine which magnitude best represents the simulated PGA. The two datasets used consist of the PGA residuals and a random distribution with a mean of zero and a standard deviation equal to the GMM standard deviation. If the simulated PGA are well represented by a GMM given a certain trial magnitude, we should expect the residuals to have a zero-average. The K–S test returns a p-value, which indicates the likelihood the two datasets come from different distributions. Once the p-value reaches a significance value of at least 0.05 for a trial magnitude, we consider that magnitude to be \mathbf{M}_{PGA} (Fig 2.S6).

We obtain \mathbf{M}_{PGD} values slightly lower than that computed by Sahakian et al. (2019) (Fig. 2.9a), and \mathbf{M}_{PGD} for both studies underestimate the true moment magnitude ($\sim\mathbf{M}_{PGD}7.2$ – 7.6 compared to $\mathbf{M}7.8$), but that is likely due to limitations of the point source assumption in GMMs. As anticipated, \mathbf{M}_{PGA} values for the simulations are appreciably lower than \mathbf{M}_{PGD} , although they are also slightly lower than that computed by Sahakian et al. (2019) (Fig. 2.9b). Despite the simulated magnitude estimates being lower than the observed estimates, the resulting \mathbf{M}_{PGA} – \mathbf{M}_{PGD} values for the simulations average around

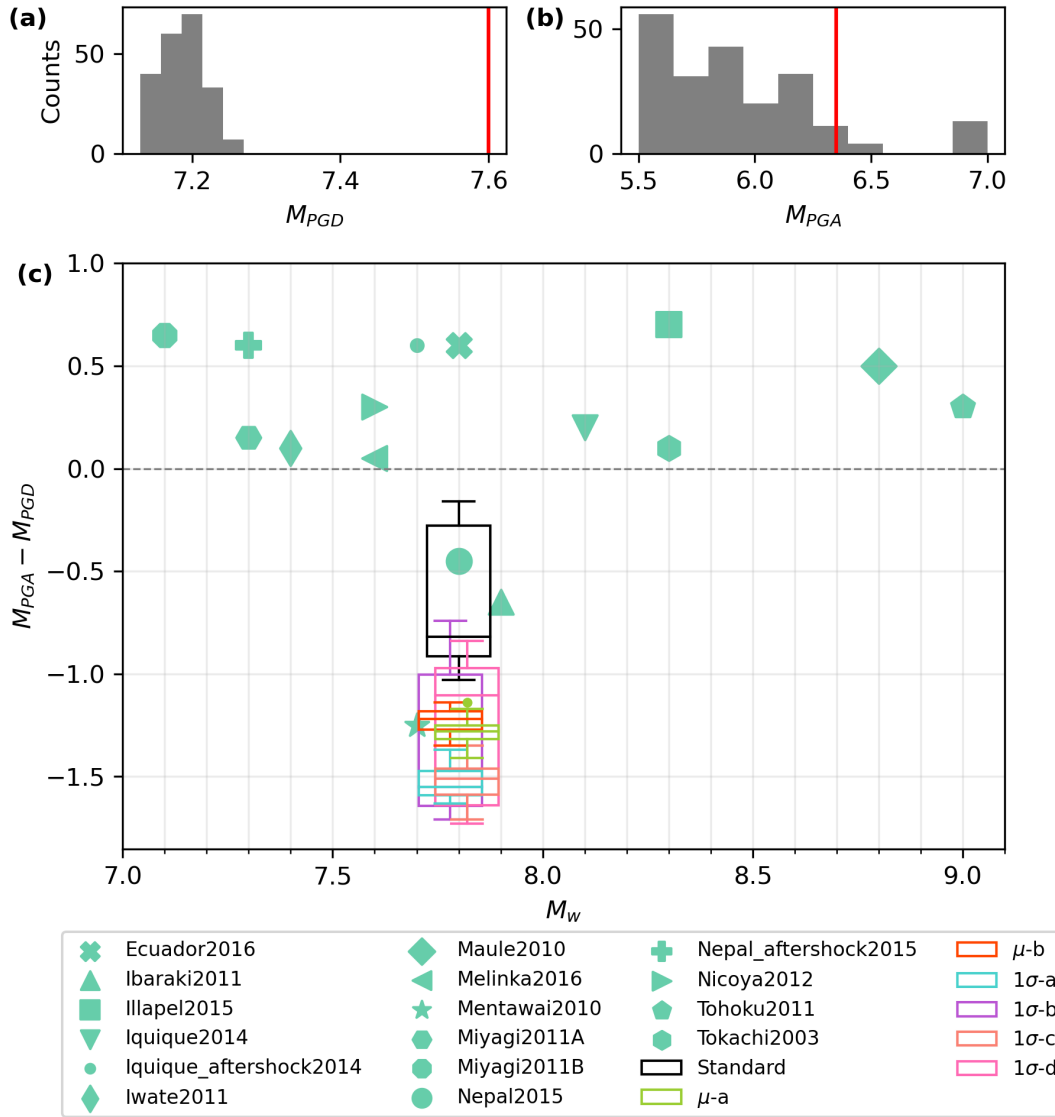


Figure 2.9. (a–b) Histograms of M_{PGD} and M_{PGA} , respectively, for simulations of the 2010 Mentawai tsunami earthquake. Red lines indicate the 2010 Mentawai magnitude value calculated by Sahakian et al. (2019) using observed data. (c) $M_{PGA} - M_{PGD}$ for 15 observed events and simulations of the 2010 Mentawai tsunami earthquake using six different rupture parameter combinations. Each boxplot represents values for 30 stochastic events. μ -a and μ -b are parameter combinations using values of the mean Gaussian process regression resulting in tPGD and HF residuals of 0 (i.e., along the solid black lines in Figure 2.7). 1σ -a, 1σ -b, 1σ -c, and 1σ -d are parameter combinations using values within one standard deviation of the total residuals (i.e., within the dashed black lines in Figure 2.7). The specific values for each parameter combination are given in Table 2.1.

-1.5 (Fig. 2.9c), similar to that computed by Sahakian et al., 2019. A two-sample K–S test between the TsE simulation $\mathbf{M}_{PGA}-\mathbf{M}_{PGD}$ values and $\mathbf{M}_{PGA}-\mathbf{M}_{PGD}$ for the observed typical events has a p-value on the order of 10^{-22} , and a two-sample K–S test between the TsE simulation $\mathbf{M}_{PGA}-\mathbf{M}_{PGD}$ values and standard simulation $\mathbf{M}_{PGA}-\mathbf{M}_{PGD}$ values has a p-value on the order of 10^{-18} . The extremely small p-values for both tests signify the simulated TsE scenarios are statistically different from standard earthquakes. Conversely, two-sample K–S tests between simulated and observed $\mathbf{M}_{PGA}-\mathbf{M}_{PGD}$ for similar data types (e.g., standard simulations and observed typical earthquakes; TsE simulations and a distribution centered on the observed Mentawai value) have a p-value on the order of 10^{-7} . Although these p-values are still small, they are significantly higher than the p-values for K–S tests between TsE simulations and standard earthquakes, which suggests they are more statistically similar. Thus, we find simulations generated using recommended TsE parameters successfully capture the IM-derived magnitude disparity expected in real-time analysis of TsEs, which further validates their use for improving tsunami early warning capabilities.

2.6 Conclusions

TsEs are a uniquely challenging problem for both earthquake and tsunami early warning as they rupture typically aseismic regions, result in unusually weak shaking even in the near-field, and yet produce tsunami amplitudes on par with some of the largest earthquake-generated tsunamis in history. Unassuming to seismic-driven earthquake early warning systems and magnitude-driven tsunami early warning systems, previous TsEs have not allowed for ample warning prior to tsunami inundation and are highly destructive to communities in their wake. The rarity of these events further complicates the problem, as minimal data exist to help better understand the underlying physics governing the ruptures and develop algorithms capable of discriminating TsEs in real time. Using near-field observations and simulated seismogeodetic data, we aimed to establish constraints on acceptable rupture parameter values for use in simulating TsE scenarios as a supplement for real data.

Using *FakeQuakes* forward modeling, we constrain ideal rupture parameter configurations, which successfully recreate the near-field ensemble of ground motions for the 2010 Mentawai TsE. Our intent is for this configuration (Figure 2.7; equations 2.7–2.8) to be used to generate TsE scenarios in other regions around the world to aid in the assessment of tsunami early warning systems in the case of an unexpected tsunami earthquake. We are aware that our results are based solely on one event. However, as these events are so rare, very few recordings exist for TsEs, especially in the critical near field. Nevertheless, tsunami earthquakes still have characteristic features that differ strongly from typical earthquakes, irrespective of the geometry and physical properties of the subduction zone. Such features include a low stress parameter, long rise time, and slow rupture velocity, as well as low energy-to-moment ratios. This is also the first step in making progress towards improving our understanding of such rare events, and the ability of tsunami early warning systems to provide effective alerts. As more events are expectedly recorded in the future, we will be able to further constrain these values.

2.7 Data and Code Availability

Observed data for the 2010 Mentawai tsunami earthquake were obtained from Muzli Muzli of the BMKG and from the Rhul et al. (2019) database. These data were recorded by HR-GNSS instruments operated by the BIG and strong motion stations operated by the BMKG. Simulated data were generated using *FakeQuakes*, and the codes are available online at <https://github.com/dmelgarm/MudPy>. Inversion data used for simulating the Mentawai event were obtained from Yue et al. (2014). Maps were produced using the Generic Mapping Tools software (Wessel et al., 2019) with a digital elevation model downloaded from GMTSAR (Sandwell et al., 2011), and all analyses were completed with in-house code, available online at <https://github.com/taranye96/tsuquakes>. This includes a parameter configuration file for simulating data with *FakeQuakes*, and all major codes used for processing and analysis of the data. The supporting information contains more figures illustrating the slip models, an analysis on

the matched filter technique implemented, example effects of varying the shallow velocity structure, and an example of M_{PGA} estimation.

2.8 Acknowledgments

This work was supported by the NASA ROSES grant 80NSSC21K0841, and U.S. Geological Survey grants G19AP00071 and G23AP00048. We thank Badan Informasi Geospasial (BIG) and the Agency for Meteorology, Climatology, and Geophysics of Indonesia (BMKG) for collecting and sharing geophysical data for the 2010 Mentawai earthquake. We are grateful to Christine Ruhl and Dara Goldberg for providing and processing the HR-GNSS data, and Muzli Muzli for providing processed strong motion data for the Mentawai event. We also express special thanks to Ben Farr for providing valuable insight on statistical analyses, including Gaussian process regression.

2.9 Bridge

This chapter has focused on the source component of ground motion by using a combination of observed recordings and 1D semistochastic simulations to constrain source characteristics of the 2010 Mentawai tsunami earthquake. In particular, the equations presented in this chapter can be used to establish reasonable rupture parameter combinations for simulating tsunami earthquake type scenarios. Such simulations can be used to test implementation of real-time algorithms for discerning tsunami earthquakes. The next two chapters will broaden to include studies of the path and site components of ground motion. The following chapter will continue with earthquake simulations but will focus on the Cascadia Subduction Zone, a region with high seismic potential but infrequent seismicity. Simulations generated for this region are validated against global models to ensure the source and path are reasonably modeled. Chapter IV will transition to focus solely on the site component for ground-motion model applications.

2.10 Supporting Information

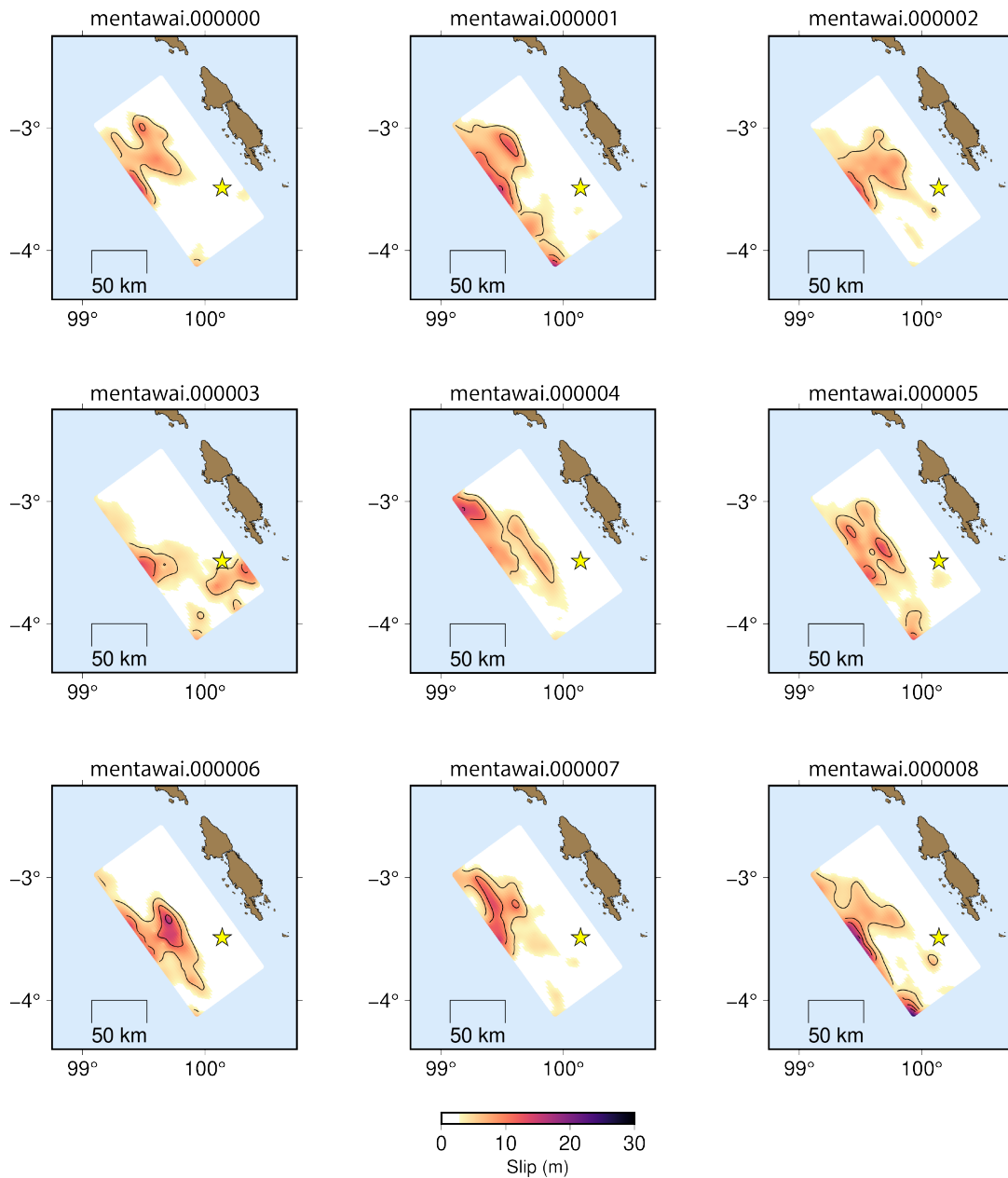


Figure 2.S1. Slip models for all 30 initial stochastic ruptures (i.e., mentawai.000000, mentawai.000001, . . . , mentawai.000029, respectively). 5 m contours from the stochastic slip patterns are shown.

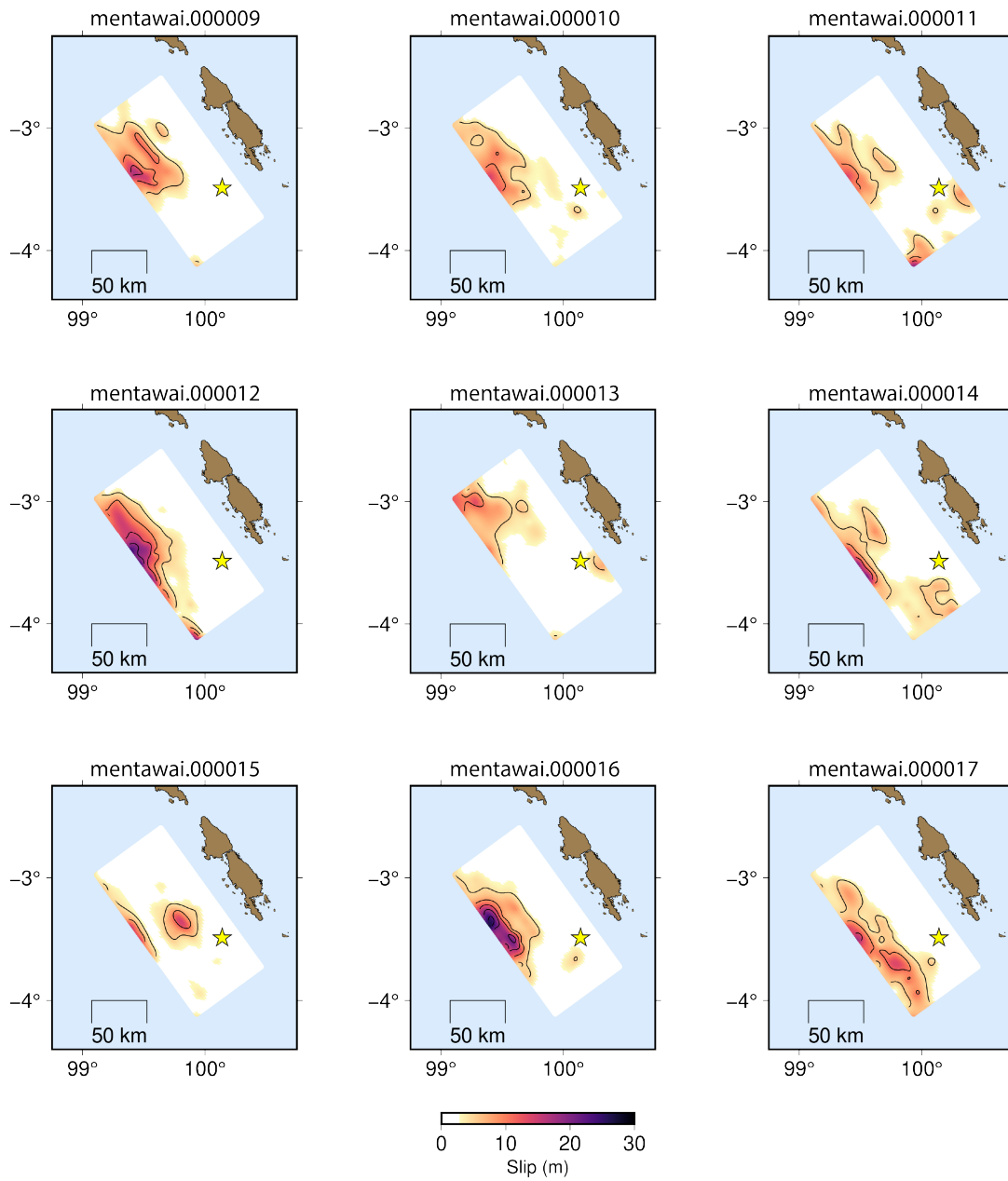


Figure 2.S1. (continued)

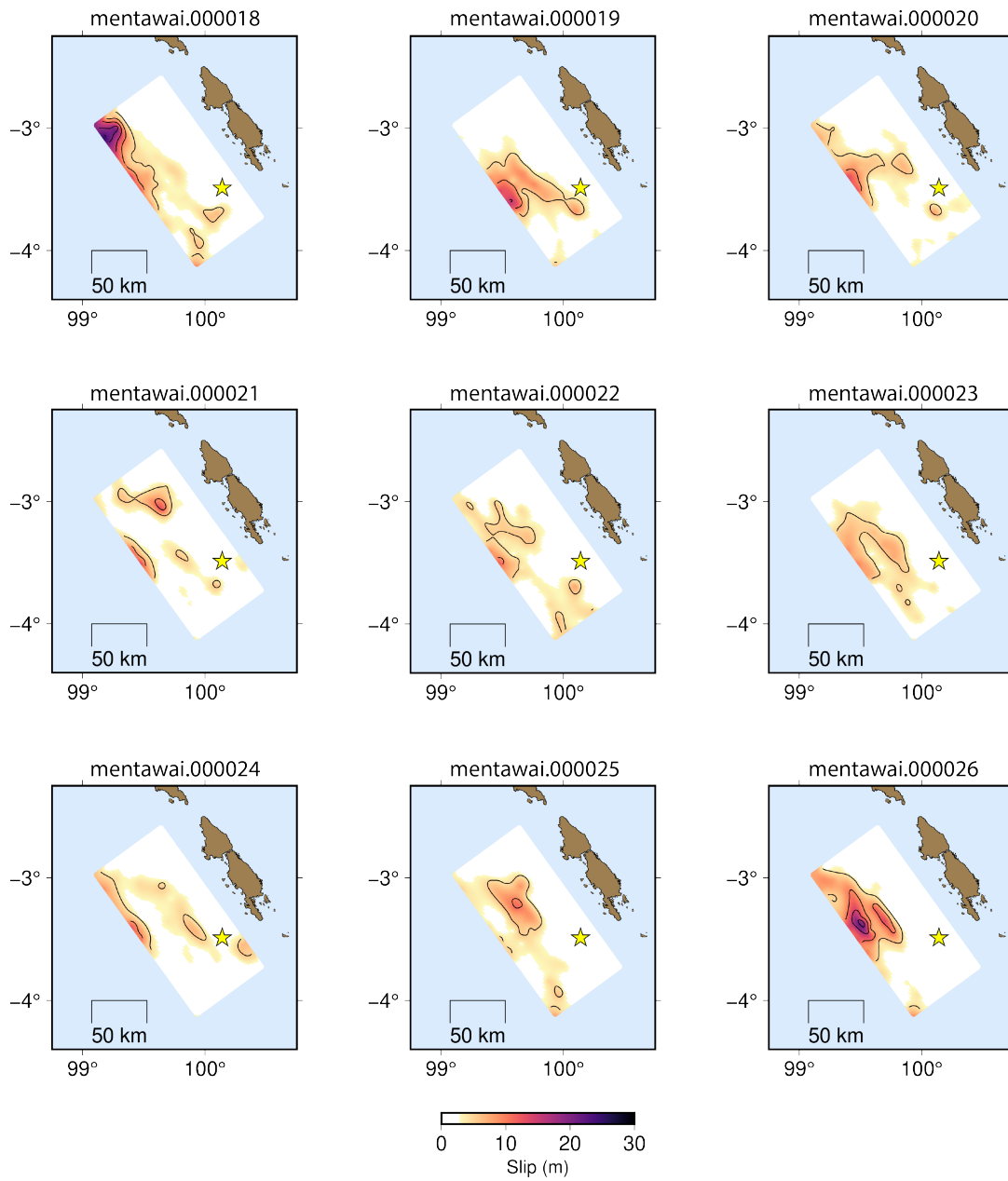


Figure 2.S1. (continued)

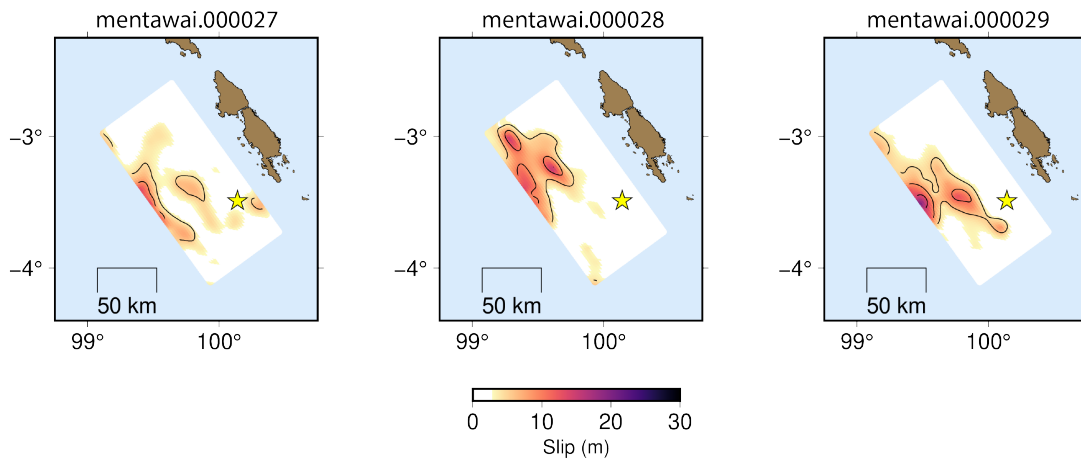


Figure 2.S1. (continued)

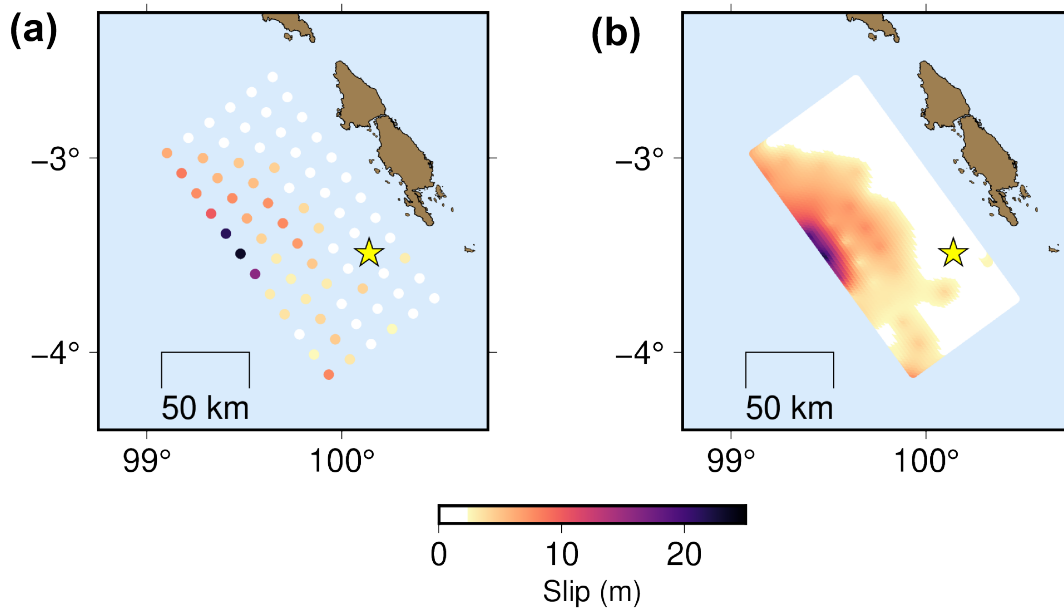


Figure 2.S2. (a) Coarsely discretized slip model from Yue et al. (2014). (b) Finely discretized version of the slip model from (a) used as the mean slip model for this study.

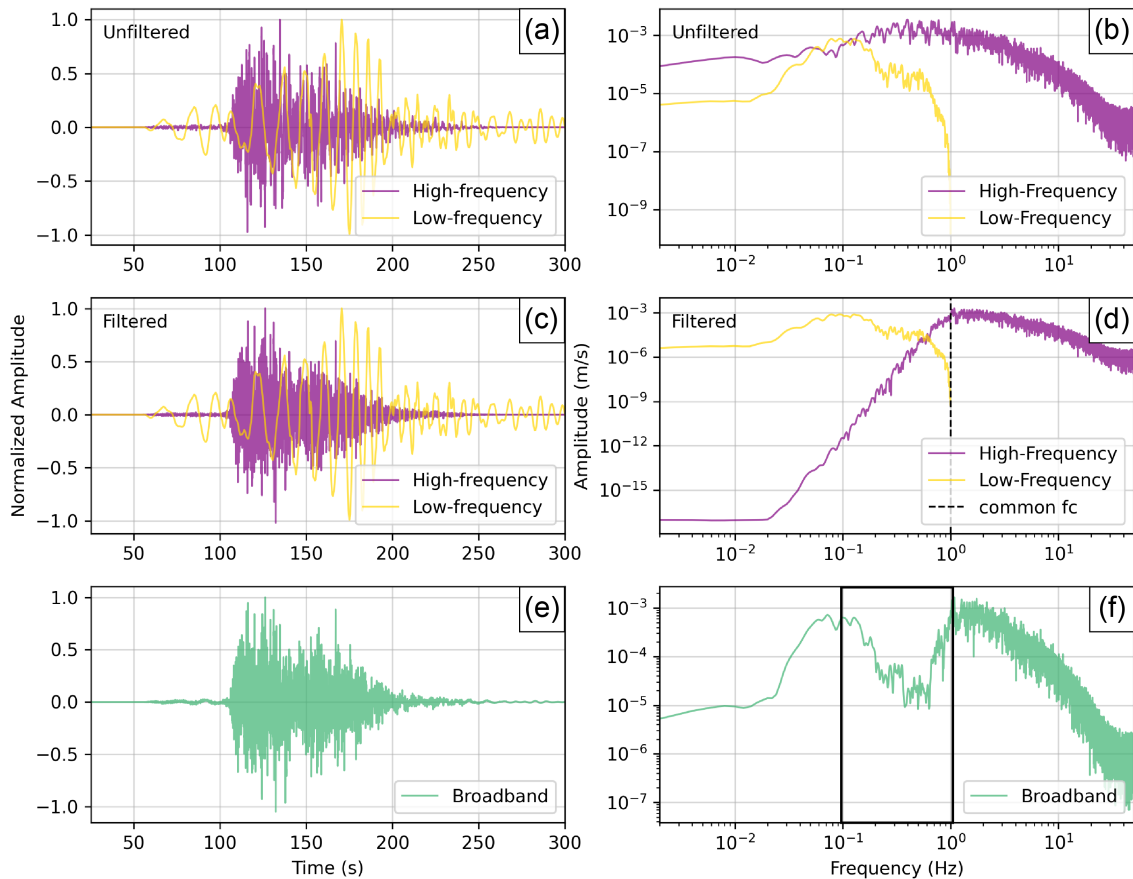


Figure 2.S3. (a) Normalized amplitudes of the individual low-frequency (yellow) and high-frequency (purple) waveforms prior to filtering. (c) Normalized amplitudes of the individual lowpass filtered low-frequency (yellow) and highpass filtered high-frequency (purple) waveforms, both with a filter f_c of 0.998 Hz. (e) Normalized amplitude of the resulting broadband waveform after combining the waveforms in (c). (b,d,f) Fourier amplitude spectra of the waveforms in (a,c,e). The dashed line in (d) indicates the filter f_c , and the black box in (f) highlights the artificial notch in the broadband spectrum.

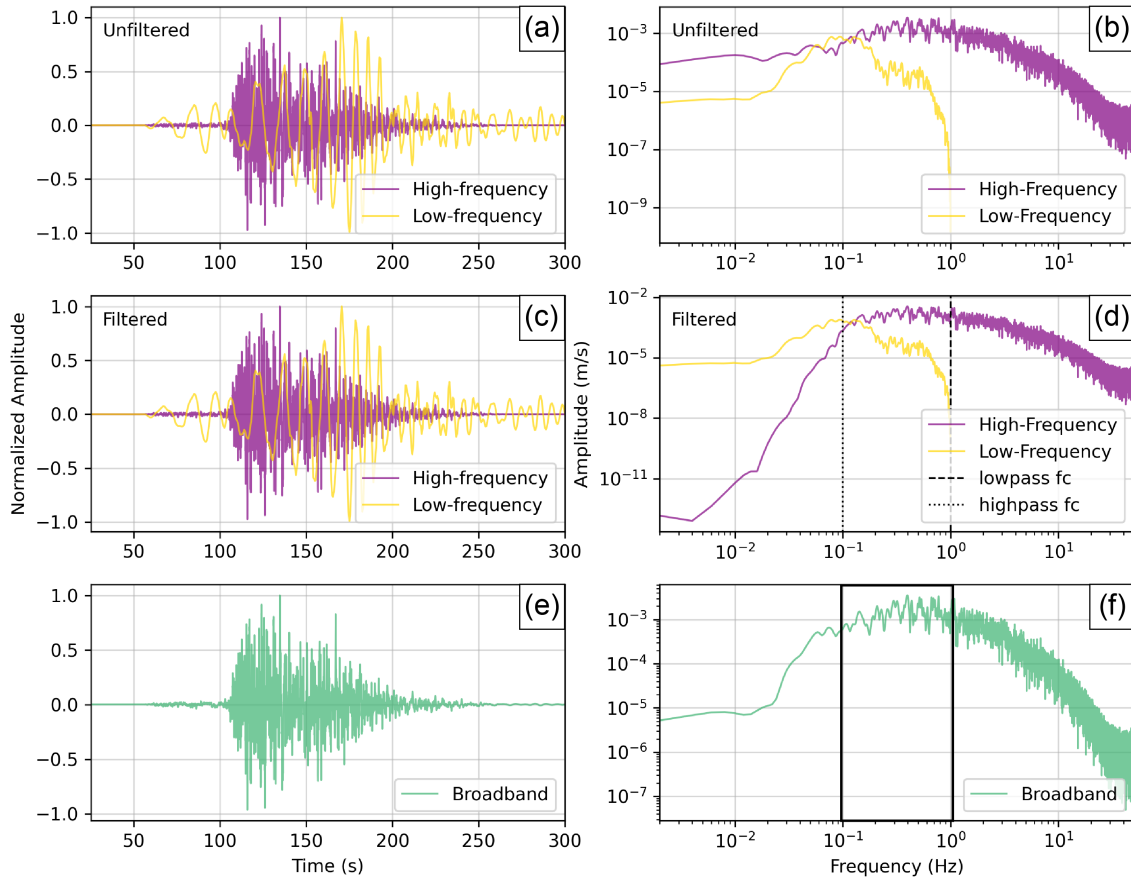


Figure 2.S4. Figure S4. (a) Normalized amplitudes of the individual low-frequency (yellow) and high-frequency (purple) waveforms prior to filtering. (c) Normalized amplitudes of the individual low-frequency (yellow) waveform lowpass filtered with a filter f_c of 0.998 Hz, and the high-frequency (purple) waveform highpass filtered with a filter f_c of 0.1 Hz. (e) Normalized amplitude of the resulting broadband waveform after combing the waveforms in (c). (b,d,f) Fourier amplitude spectra of the waveforms in (a,c,e). The dashed and dotted lines in (d) indicate the lowpass and highpass filter f_c , respectively, and the black box in (f) highlights the smoothed spectrum absent of an artificial notch.

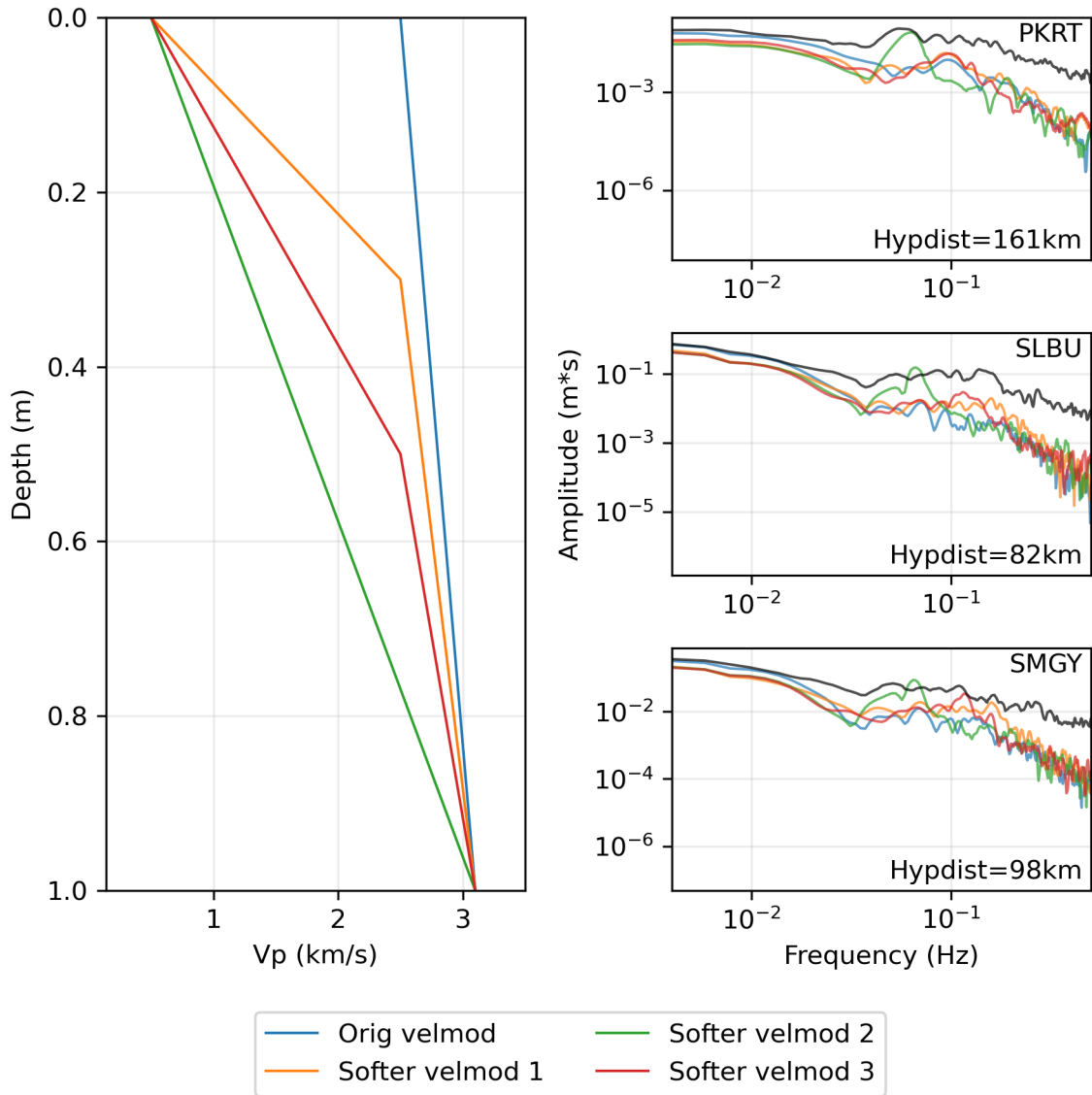


Figure 2.S5. (left) Comparison between the original velocity model used for this study and three variations of the velocity model using softer layers in the shallow subsurface. (right) Binned displacement Fourier amplitude spectra for three example HR-GNSS stations using the different velocity models.

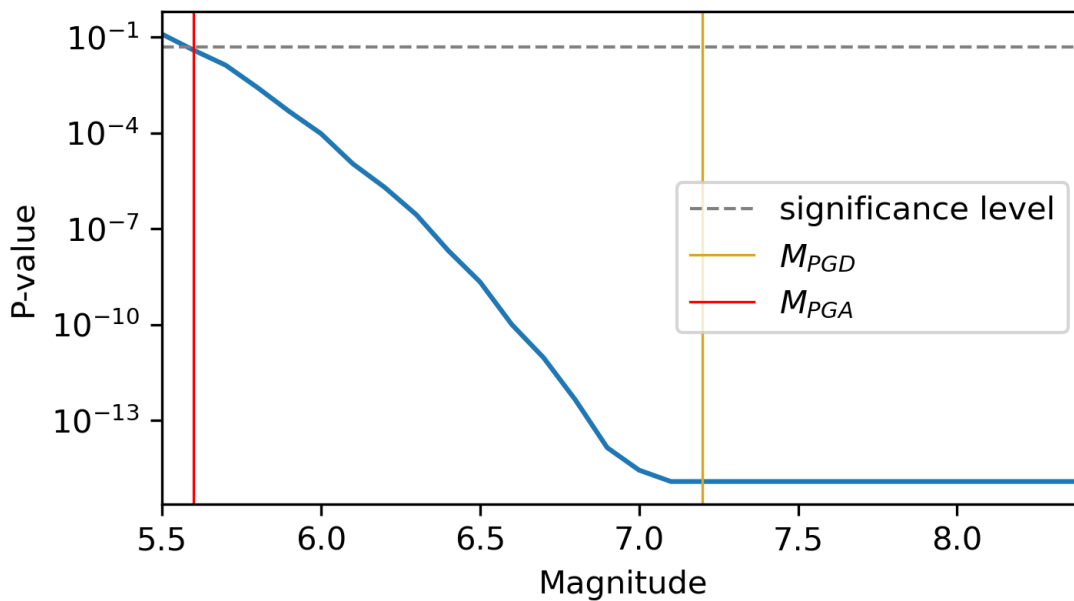


Figure 2.S6. Example solvation for M_{PGA} for one stochastic event with one set of parameters. The p-values result from Kolmogorov–Smirnov (K–S) tests performed between simulated PGA values and GMM PGA predictions for trial magnitudes ranging $M_{5.5}$ – $M_{8.5}$. M_{PGD} and M_{PGA} for this event are indicated by the gold and red lines, respectively.

CHAPTER III

VALIDATION FRAMEWORK FOR SEMISTOCHASTIC SIMULATIONS IN CROSS-BORDER CASCADIA EARTHQUAKE EARLY WARNING

This chapter contains co-authored material and is in preparation to be submitted for publication. The writing of this chapter was done solely by me with co-authors providing editorial assistance. Valerie Sahakian, Angela Schlesinger, Benoit Pirenne, and I conceptualized the work presented in this chapter. Diego Melgar provided the simulation codes, which I used to generate the synthetic dataset. Alireza Babaie Mahani and Eli Ferguson ran the simulations through Ocean Networks Canada’s offline early warning system, and Amy Williamson and Angie Lux ran the simulations through EPIC’s offline early warning system, each of whom contributed to analysis of the early warning system performance. I performed all additional analyses.

3.1 Introduction

Earthquake early warning (EEW) has been at the forefront of seismological research for several decades. Initially conceptualized in the late 1800s, advancements in seismic instrumentation and telecommunication prompted implementation of local EEW systems for specific users in Japan in the 1960s–70s, with the first public roll out of an EEW system in Mexico in 1991 (Suarez et al., 2018). As of today, public EEW systems are currently active or under development in Mexico, Japan, South Korea, Taiwan, China, Italy, Canada,

India, Turkey, Romania, and the United States (U.S.) West Coast (Chamoli et al., 2021; Chung et al., 2019; Clinton et al., 2016; Cuéllar et al., 2018; Dong-Hoon et al., 2017; Kohler et al., 2020; Nakamura, 1988; Schlesinger et al., 2021; Wu et al., 2021; Zhang et al., 2016; Zollo et al., 2009). Although varying in technical details, the first alert issued by most modern EEW systems is from information recorded from the initial small P-waves. This can be leveraged to forecast future shaking and provide seconds to minutes of warning time to either the public or special groups of interest. Ideally, these alerts provide enough time for individuals to take cover and for critical operations to be protected, such as electric isolation of computer systems and power grids, shutting down of nuclear power plants and rail systems, and protection of hospitals and fire stations (Allen and Melgar, 2019).

Because EEW systems are fully automated and have no human interaction at any point in the processing chain, and because the consequences of missed or false alerts are significant, testing and validating them so that their performance is understood completely is critical. Oftentimes, time-series data from real regional earthquakes are used, as these capture the unique crustal properties inherent to the region and are likely representative of future ground motions. However, there are instances where observed data may be insufficient, such as in regions experiencing high seismicity but lacking seismic or geodetic instruments (e.g., Central and Southeast Asia, Eastern Africa, and some small island nations), or in regions with high seismic potential but less frequent seismicity (e.g., Central Eastern and Northwestern United States). Or simply on a global scale, the desire to build systems capable of alerting large **M**8.5+ events, which contribute significantly to seismic hazard and result in considerable shaking, structural damage, and injuries, and yet, minimal data exist for such large events in the near-field compared to more moderate events due to their long recurrence intervals of several centuries. For such circumstances, simulated earthquake data are a necessary supplement (e.g., Lin et al., 2021; Ruhl et al., 2017; Thompson et al., 2023; Zollo et al., 2009).

Although earthquake simulations have been widely used in various applications, validation of the models used to generate the data have focused primarily on rupture

physics, and to some extent ground motions (e.g., Chapter II of this dissertation; Lin et al., 2021). A validation framework is yet to exist for the generation of simulated data for use as a training mechanism for earthquake early warning, especially for shaking-related intensity measures in the Cascadia region. In this study, we (1) generate a comprehensive dataset of kinematic slip models and associated waveforms for use in testing EEW, and we (2) evaluate these data based on the key components of EEW systems (i.e., earthquake detection, magnitude estimation, and ground-motion forecasting) using the CSZ as a case study.

3.2 EEW in the Pacific Northwest

The CSZ is the interface boundary between the Explorer, Juan de Fuca, and Gorda plates and the North American plate, extending 1100 km from Southwestern British Columbia to Northern California (Fig. 3.1a). As is the case for nearly all subduction zones, the CSZ has hosted large **M**8–9 earthquakes, with the last major rupture event occurring in 1700 C.E. (Satake et al., 2003; Walton et al., 2021). The CSZ has accumulated enough strain to generate a **M**9 earthquake if the full margin were to rupture, which has a calculated odds of 7–15% in the next 50 years (Wang et al., 2013). The odds of a large **M**7+ rupture along the southern margin is around 37%, which would still result in strong shaking for nearby cities.

Although historically the CSZ has hosted large events, it is anomalous in that it experiences comparatively very little seismicity during its main interseismic cycle (~500 years; Bostock et al., 2019; Wirth et al., 2022). Conversely, other major subduction zones generally record **M**7+ events every couple of years. A possible societal consequence of such earthquake quiescence is that few people consider the impending seismic hazard in their short time-scale frame of reference, thus resulting in non-resilient infrastructure and a community mentally unprepared to take action in the event of a major earthquake (Butler, 2018; Cutts et al., 2017). Most of the population in the Pacific Northwest (PNW) lives within 300 km of the trench and would likely experience shaking of a modified Mercalli intensity (MMI) \geq VII in the event of a **M**9 rupture (Madin and Burns, 2013; Wirth

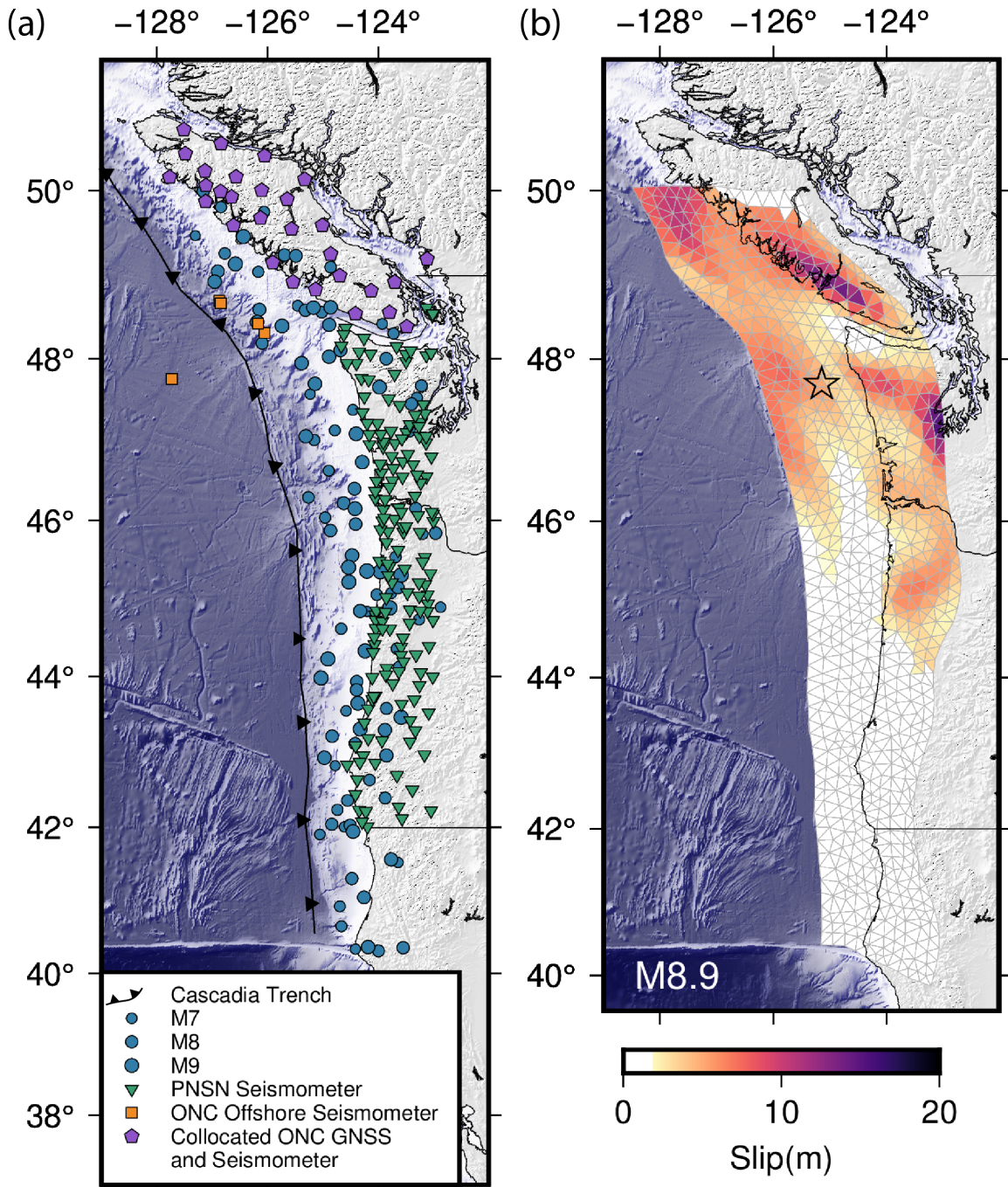


Figure 3.1. (a) Cascadia Subduction Zone study region with simulated event epicenters (blue circles) and the set of High-Rate Global Navigation Satellite System (HR-GNSS) and seismic stations (colored by station network and type) used for waveform simulation. PNSN: Pacific Northwest Seismic Network, ONC: Ocean Networks Canada. (b) Example stochastic rupture scenario generated using *FakeQuakes*.

et al., 2021). However, alert times from an EEW system can provide more warning to communities in the PNW than in concentrated strike-slip settings such as California due to the greater distance from the fault to population centers. The CSZ is a prime candidate for this study, as it is a region of high seismic hazard and risk, and yet lacks the desired real-time observed data for EEW testing for large magnitudes.

Evaluating EEW performance is particularly challenging for a major CSZ rupture, not only because of the dearth of data to test such algorithms, but also because the megathrust extends along two countries (Canada and the United States). Hence, different EEW algorithms, seismic networks, detection boundaries, and alerting policies are in place.

ShakeAlert is the operational EEW system covering California, Oregon, and Washington of the U.S. West Coast (Kohler et al., 2020). Currently, it is supported by two algorithms, EPIC and FinDer. EPIC is a modified version of the ElarmS point source algorithm originating from U.C. Berkeley. Initially developed for Southern California, ElarmS was later extended to Northern California (Wurman et al., 2007) and eventually the entire west coast (Kohler et al., 2018; Kohler et al., 2020). The latest version of ElarmS, ElarmS-3 (Chung et al., 2019), was merged with slight modifications into ShakeAlert as the new algorithm EPIC (Kohler et al., 2018). FinDer is a finite fault algorithm which estimates the linear fault extent and strike using real-time observations of ground motion (Allen and Melgar, 2019). It is slower than EPIC but provides more accurate shaking estimates and is used to update alerts as the rupture progresses.

In 2015, Ocean Networks Canada (ONC), an initiative hosted and owned by the University of Victoria, developed an EEW system for Southwestern British Columbia, Canada (hereafter referred to as ONC-EW) in a collaborative effort with Natural Resources Canada (NRCAN). ONC-EW's network configuration is leading edge as it is comprised of collocated High-Rate Global Navigation Satellite System (HR-GNSS) and strong motion sensors, which only exists at a handful of sites around the world (Schlesinger et al., 2021). This allows for onsite combination of high-frequency information provided by seismic instruments, which inform MMI estimates for alerts, with low-frequency

geodetic information unbiased by magnitude, without the typical time latency transferring information between stations. Another unique feature of the network configuration is the addition of five subsea strong motion sensors connected to an underwater cabled network, NEPTUNE (North-East Pacific Time-series Undersea Networked Experiments), which drapes over the CSZ and is the world’s first multi-node cabled ocean observatory.

EEW system performance updates highlight successful parts of the algorithms and illuminate features requiring modification for more efficient and accurate alerts. ShakeAlert algorithms have undergone numerous performance evaluations. Böse et al. (2022) provided performance updates on EPIC and FinDer. Their evaluations showed EPIC to be the fastest algorithm with the most accurate source location estimates but underestimates large magnitude events. FinDer presents unsaturated magnitude estimates but sometimes misses offshore events. Their analyses did not use any subduction intraslab or interface events. Hartog et al. (2016) focused on performance of events offshore Oregon and Washington, but only $M < 7$ events were considered as real time-series data were used, most of which did not originate from a megathrust event. Rosenberger et al. (2019) tested the performance of ONC-EW’s epicenter location accuracy on over 2000 $M7$ simulations, and several $3 < M < 6.5$ offshore events were detected between 2018 and 2020 as the system was still undergoing development. However, none of these events occurred along the CSZ interface (Schlesinger et al., 2021). As the CSZ will inevitably host another $M8-9$ event, it is critical to test both ShakeAlert and ONC’s performance in the event of a large megathrust rupture. To support such efforts, we generate a dataset of moderate to large CSZ earthquakes and validate them to ensure robust performance for EEW testing. Namely, we evaluate the triggering success of simulated time-series data, magnitude estimation from P-wave amplitudes, and ground-motion intensities.

3.3 Stochastic Rupture Simulation

We curate a simulated dataset for the CSZ consisting of 112 kinematic rupture scenarios and associated waveforms using *FakeQuakes*, a one-dimensional (1D) semistochastic model (Melgar et al., 2016). Our objective is to assess how well these

waveforms perform at replicating reality with respect to key metrics needed for an EEW system, and subsequently test them in the system itself.

Various simulation methodologies exist, ranging from simple stochastic models of high-frequency ground motion (e.g., Boore, 2003) to complex deterministic models involving numerical simulations of full waveform propagation through a three-dimensional (3D) velocity structure (e.g., Faccioli et al., 1997; Komatitsch and Tromp, 1999; Mazzieri et al., 2013). *FakeQuakes* is a 1D semistochastic broadband simulation model, which provides more realistic ground motions compared to a simple stochastic model, yet is efficient enough to allow for large data generation. *FakeQuakes* is not the only existing 1D semistochastic model; however, it is preferred for this study, as opposed to other similar models (e.g., SCEC Broadband platform, Maechling et al., 2015), as it allows for P-wave generation (necessary for EEW) and has previously been tested in the Cascadia region (Melgar et al., 2016; Goldberg and Melgar, 2020; Small and Melgar, 2023).

3.3.1 Kinematic Rupture Models

The CSZ skeletal framework (based on Slab2.0 geometry; Hayes, 2018) and velocity structure are taken from Melgar et al. (2016) (Fig. 3.S1). Using this framework, we generate rupture scenarios with target magnitudes ranging from **M6.8–9.5**, where four ruptures are generated for each 0.1 target magnitude unit. The stochastic nature of the kinematic rupture generation results in final magnitudes similar, but not always equal to, the target magnitude, and our final dataset consists of events ranging from **M6.6–9.4**. The finite length and width of each simulated rupture is determined using stochastic rupture dimension–magnitude scaling laws (Blaser et al., 2010), the slip distribution along the active part of the fault is constructed using a Von Karman autocorrelation function (Mai and Beroza, 2002) and Karhunen Loève expansion (LeVeque et al., 2016), and the total slip (μ) is the amount of slip required to produce the target magnitude. We truncate the series of eigenvalues and eigenvectors (equal in length to the number of subfaults) to 72 modes. Higher modes describe higher-frequency components of the slip pattern, which are desirable for seismic wave generation. Although we do not use the full eigenseries, we find

72 modes to be suitable for this purpose, which is significantly higher than the number of modes used for geodetic purposes (~ 3) (LeVeque et al., 2016). The kinematics of the event are controlled by the velocity structure and the static slip amplitudes.

3.3.2 *Simulated Waveforms*

For each of the simulated events, we generate waveforms for 29 collocated HR-GNSS and strong motion sites onshore Vancouver Island, 5 cabled array ocean bottom seismometers (OBSs) offshore Vancouver Island, and 157 broadband and strong motion sites along coastal Washington and Oregon. We use a subset of stations comprising the network configuration for ShakeAlert as ShakeAlert’s primary coverage is crustal events in California rather than slab-originating events. The stations selected for ShakeAlert cover the major cities landward of the CSZ and are located within a few hundred km of the trench. We also use a subset of stations operated by ONC as these cover the Canadian areas of the CSZ.

HR-GNSS Waveforms: The waveforms for the HR-GNSS stations and the low-frequency (< 1 Hz) portion of the broadband waveforms for the seismic stations are generated following equivalent deterministic (i.e., physics-governed) methodology. Green’s functions are first computed with a 2 Hz sampling rate for each subfault station pair following the frequency–waveform matrix propagator method of Zhu and Rivera (2002), after which they are convolved with a “Dreger” source time function (Melgar et al., 2016) and corrected for a moment tensor with appropriate strike and dip for the prescribed slab structure. These synthetics are then multiplied by slip on each subfault and summed to form a single waveform for each event–station pair.

Broadband Waveforms: A stochastic approach is used instead for the modeling of high frequencies (> 1 Hz) (Graves and Pitarka, 2010, 2015) because computational cost significantly increases with higher sampling rates (e.g. Rodgers et al., 2020). Even with greater computational resources, a stochastic approach is still desired because current resolution of the CSZ velocity structure is not fine enough to capture small-scale features necessary for realistic high-frequency simulation. It is also far more challenging to resolve

physical properties of the subsurface offshore where there are fewer instruments. Thus, high-frequency waveforms generated using such methods would prove uninformative.

The stress parameter (equivalent to the Brune high-frequency stress parameter; Brune, 1970) is a user-defined parameter which controls subfault corner frequency and seismic moment scaling in the source term of the stochastic model. Stress parameters range orders of magnitude and do not appear to be magnitude-dependent (Baltay et al., 2011), and stress parameter–depth dependence studies have had varied conclusions (Abercrombie et al., 2011; Baltay et al., 2011). For simplicity, we use a single value of 50 bars for all events, which is within a reasonable range of stress parameters for interface events (30–120 bars; Atkinson and Macias, 2009). For the site-specific high-frequency attenuation (i.e., κ_0 ; Anderson and Hough, 1984), we use a uniform value of 0.04 s for all sites, as site-specific response is not well constrained for all stations in this dataset.

The low- and high-frequency components are combined to form a broadband (i.e., full spectrum) waveform (generated for both broadband and strong motion stations) by first double differentiating the low-frequency displacements to acceleration, after which the two components undergo a two-frequency matched 4th order causal Butterworth filter (Fig. 3.2).

3.3.3 Post Processing

We apply minor post-processing to the raw output waveforms from *FakeQuakes* to simulate what an early warning algorithm would be analyzing in real time. First, two minutes of zero-padding are prepended to each waveform, so as not to bias the algorithm with data containing event arrivals near the start of the record. Second, each simulated recording is combined with noise. For strong motion recordings of large earthquakes at close distances, the signal amplitude is generally large enough so that the peak intensity measures (IMs) are unbiased by the presence of noise. This is not always the case for small-to-moderate events or greater distances. HR-GNSS instruments contain significantly higher levels of noise resulting from uncertainty in post processing and sensitivity to electromagnetic signals (Melgar et al., 2020), and it is generally more difficult to filter

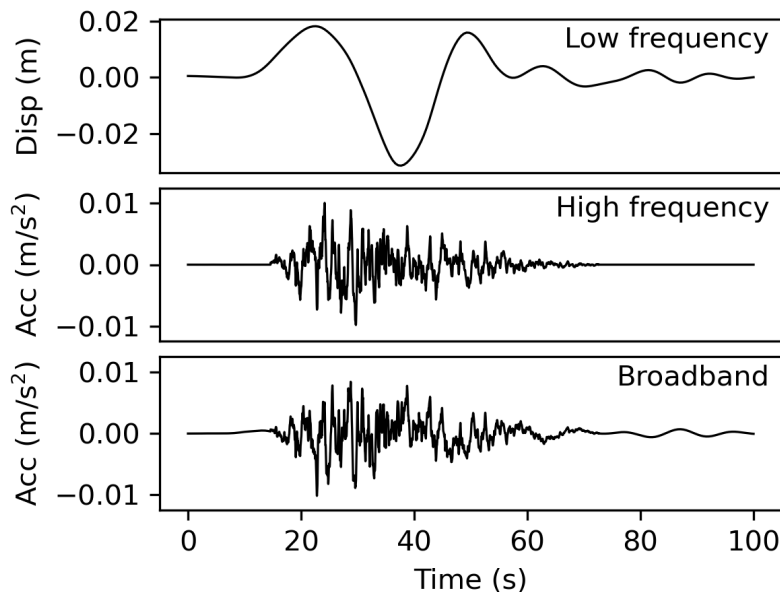


Figure 3.2. Example components of broadband waveform simulation. (top) Low-frequency displacement waveforms are combined with (middle) high-frequency acceleration waveforms using a matched filter to produce (bottom) broadband (i.e., full spectrum) waveforms.

out as it is often characterized by the same frequencies as the earthquake signal. For such reasons, it is necessary to include noise in the simulated data, ensuring robust analysis of these data as a supplement in EEW testing.

For waveforms simulated at ONC- and PNSN-operated onshore broadband and strong motion sites, segments of real noise are applied. One day of acceleration data (02/20/2022, 02/02/2022, and 02/26/2022), was downloaded from IRIS for the Ocean Networks Canada Onshore Network (OW) stations AL2H, CMBR, and PHRD, respectively. The noise streams were screened to remove potential earthquake signals or unusual responses. Subsets of noise are randomly selected and demeaned prior to being combined with the onshore broadband waveforms.

OBSs generally have higher levels of environmental noise when compared with land sensors, resulting from ocean currents, sediment gas bubbles, and biological disturbances (Hilmo and Wilcock, 2020). This poses a challenge when incorporating such offshore sensors into an EEW network, as the noise can easily obscure the initial P-waves.

However, the cabled array of OBSs installed by ONC is completely buried and relatively quiet (Table 3.S1 and Fig. 3.S2). These sensors could provide vital information about the initial seismic waves from an interface event, thus increasing the warning time before the destructive waves reach onshore. As oceanic conditions can be highly variable in different environments, one day of acceleration data (10/31/2018) was downloaded from IRIS for each of the five offshore sites. Segments of the noise are randomly selected, demeaned, and added to their respective station's event waveforms. Unlike the broadband waveforms, we choose to use synthetic noise for the simulated HR-GNSS waveforms (Melgar et al., 2020).

HR-GNSS noise is much more variable over time compared to that of inertial instruments, thus one day of noise would not be an adequate representation of conditions over a long period of time. Moreover, the synthetic noise model from Melgar et al. (2020) was derived using data from the Cascadia region, rendering it a better average compared to a single day of real HR-GNSS data.

The additional processing steps described in the following sections are applied individually for each analysis.

3.4 Earthquake Detection

Successful detection of an earthquake signal in real-time geophysical recordings is vital, as it initiates the workflow of an EEW algorithm and influences the amount of warning time provided to the public. Events are detected via a short-term average/long-term average (STA/LTA) trigger, which is a ratio of averages of the signal amplitude computed over two user defined time periods. If this function surpasses a predetermined threshold, the algorithm records a trigger. The STA, LTA, and threshold parameters used are subject to individual discernment (Trnkoczy, 2012). As the CSZ is monitored by both ShakeAlert and ONC-EW, we evaluate the success of the simulated data applying parameters used by both algorithms. EPIC, the algorithm umbrellaed under ShakeAlert, uses STA/LTA/threshold parameters of 0.05/5/20 on velocity time-series data. ONC-EW uses parameters of 1/5/4.9 on acceleration time-series data.

For the STA/LTA analysis using EPIC parameters, we filter the vertical component of the resulting noisy acceleration broadband waveforms using a 0.075 Hz 4th order causal highpass Butterworth filter (Wu and Zhao, 2006). After integration to velocity, an additional highpass filter is applied. For the analysis using ONC-EW parameters, we highpass filter all three components of the noisy acceleration broadband waveforms. Additionally, ONC-EW utilizes a polarization filter to isolate the P-wave from the full signal (Rosenberger, 2010, 2019; Schlesinger et al., 2021). Polarization is commonly used for detecting P-waves and is based upon recursive singular value decomposition (SVD) of the signal (e.g., Baillard et al., 2014; Cua et al., 2009; Flinn, 1965; Ross and Ben-Zion, 2014; Ross et al., 2016;). The SVD returns parameters describing the linearity and incidence of the signal, and a nearly linear vertical signal is assumed to be the P-wave. After applying the polarization filter, the summed square of the three filtered components is run through the detection algorithm. We use the *classic_sta_lta* function from the Obspy Python module (Beyreuther et al., 2010) for the earthquake detection analyses.

The STA/LTA parameters are designed to be quite sensitive to ensure high detection success. However, this can lead to the algorithm triggering on noise spikes, which can happen even during seismically-quiet periods. In a real-time framework, the EEW workflow would not progress until a certain number of stations triggered within a reasonable geographical range. A minimum of four station triggers are required to send an alert for both EPIC and ONC-EW (Chung et al., 2019; Schlesinger et al., 2021). For this analysis, we did not track the spatial- or time-variance of the triggers. Thus, we ignored any triggers occurring before the P-wave arrival time and selected the first trigger following the origin time as the STA/LTA pick time (Fig. 3.3).

Using ONC-EW parameters, the STA/LTA algorithm triggered (i.e., detected an event) for 99.03% of the recordings, whereas the algorithm triggered for 84.55% of the recordings using the EPIC parameters. The STA/LTA algorithm is sensitive to the signal-to-noise ratio (SNR), where a weak signal would dampen the STA/LTA ratio; thus it may not detect an event for smaller magnitudes or greater distances. We evaluate

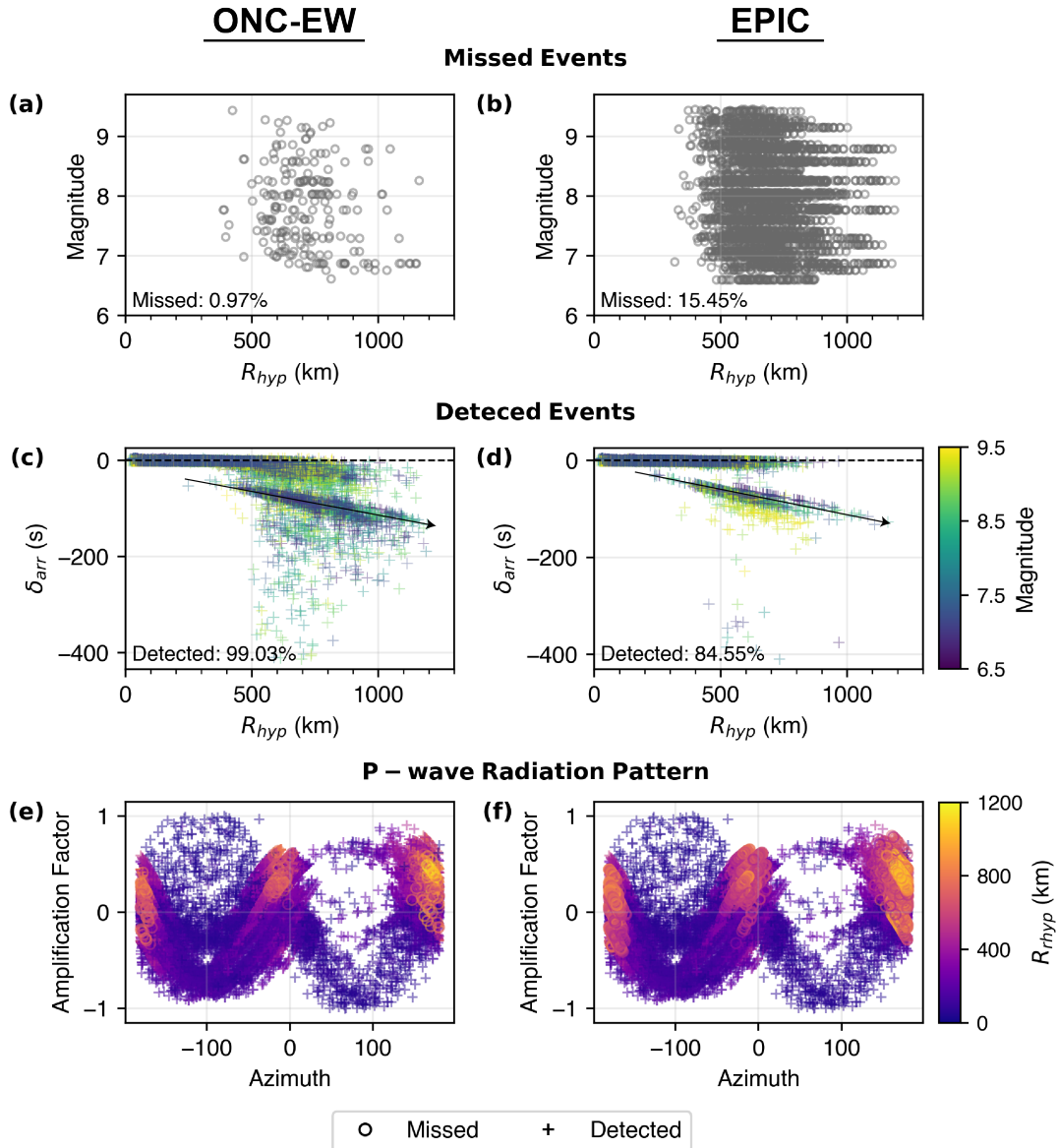


Figure 3.3. STA/LTA earthquake detection results for the Cascadia simulations using (left) ONC-EW and (right) EPIC STA/LTA configuration. (a–b) Distance and magnitude distribution of missed events. (c–d) Detection time residuals (δ_{arr} ; true P-arrival–STA/LTA pick) for each simulation that triggered the STA/LTA algorithm. Residuals are plotted as a function of distance and colored by magnitude, and the zero-residuals are indicated by a dashed line. The linear bias with distance (marked by the arrow) represents records that triggered the algorithm on the S-wave, rather than the P-wave. (e–f) Records plotted as a function of azimuth and P-wave amplification factor (i.e., radiation pattern) colored by distance, with marker shapes delineating between missed and detected events.

the hypocentral distance (R_{hyp})–magnitude distribution of missed events for both sets of STA/LTA parameters (Fig. 3.3a–b), and we observe relatively no bias toward magnitude. However, most of the missed events occurred for records > 500 km, which is the case for both sets of STA/LTA parameters. The missed detections also likely result from the more emergent nature of P-waves at non-ideal radiation azimuths (Fig. 3.3e–f), which is also expected for real earthquakes.

We next analyze the recordings that detected events to determine the time difference between the true P-arrival and the event detection time. It is crucial for effective alerts that this latency be as small as possible. Otherwise, the early warning algorithms would forecast a later arrival of the more destructive waves, thus overestimating alert times. Direct P-waves are computed using the *obs.py.taup* Python module (Krischer et al., 2015) with the Cascadia velocity model. We calculate arrival time residuals (δ_{arr}) as the true P-wave arrival time minus the STA/LTA pick time. A subset of recordings exhibits a negative δ_{arr} bias with distance, indicating recordings that triggered late because the P-wave was too weak compared to the background noise for the STA/LTA parameter configuration. For these cases, the algorithm is sometimes triggered by a noise feature rather than the earthquake signal (Fig. 3.S3). The linear trend in the residuals (indicated by the arrows in Fig. 3.3c–d) is a result of S–P travel time difference for recordings that triggered on the S-wave arrival instead of the P-wave arrival (Fig. 3.S4). The more moderate magnitude events in our dataset ($M \sim 6.5–7.5$) tend to have slightly larger residuals compared to the larger events, which is expected as smaller events have a weaker signal. A significant percentage of the records (73.71% for ONC-EW parameters and 85.74% for EPIC parameters) have a detection residual < 1 s, which both speaks to the success of the simulations for use in earthquake detection, as well as the success of the parameter configurations themselves (Fig. 3.S5). Since the true P-wave arrival is not known for real earthquake recordings, it is difficult to compare this to the performance of EEW algorithms. However, if we assume an analyst’s pick to be accurate, ShakeAlert picked arrival times within 0.5 s of the analyst pick times for 89% of their real recordings (Hartog et al., 2016). Our simulated recordings

have a slightly lower trigger success, but still perform well in comparison.

The application of a polarization filter in the ONC-EW analysis results in a significant improvement in number of recordings triggered, yet also results in greater latencies between the true arrival time and the pick time. Most of the recordings with greater δ_{arr} using ONC-EW parameters are at distances > 500 km. At these greater distances, it would take around 1.5+ minutes for these stations to trigger, meaning they would likely contribute less to the source characterization and alert levels. Allen and Melgar (2019) do caution that this type of P-detector may only be useful for borehole sites as soil response at free-field stations could reduce the quality of the filtered data. As near-surface soil response is not included in our simulations, we cannot evaluate this claim.

It is also informative to note that EPIC system operators have previously been unsuccessful running simulations through offline tests due to false triggers arising from the boundary where the noise and signal are combined. Our methodology of applying continuous noise to the padded signal eliminated these issues.

3.5 Magnitude Estimation

Following event detection, an initial magnitude is estimated; most often, this is accomplished using properties of the evolving P-wave, which provide information about the growing source (Kodera et al., 2018). The predominant period (τ_p^{max}) and displacement amplitude (P_d) of P-waves are commonly used metrics as they have been found to scale with magnitude for small-to-moderate events (e.g., Kuyuk and Allen, 2013; Zhang et al., 2022). Although they have not been found to be deterministic of earthquake magnitude, they provide *weak deterministic* information (Goldberg et al., 2018; Goldberg and Melgar 2020; Melgar and Hayes 2017). Or in other words, moderate to very large earthquakes cannot be distinguished within the first few seconds of rupture, but as the rupture organizes into a slip pulse, the kinematic properties scale with magnitude.

As both EPIC and ONC-EW use P_d as a magnitude proxy, we examine the time-evolution of P_d to evaluate the success of magnitude estimation with our simulations. Note that in the initial report for the ONC-EW (Schlesinger et al., 2021), P_d and τ_p^{max} are used

for magnitude estimation, but their algorithm has since been modified to only consider P_d . ONC-EW also uses peak ground displacement (PGD)– M scaling relations to update their magnitude estimates for larger events, but PGD is addressed in section 3.6.4.

P_d is measured as the maximum amplitude of the displacement P-wave over some time interval of several seconds. The $\log_{10}(P_d)$ has shown to scale linearly with event magnitude and $\log_{10}(\text{distance})$ for events with magnitudes $\leq M$ 7–8, depending on the time window used to calculate P_d (e.g., Crowell et al., 2013; Goldberg and Melgar, 2020; Kuyuk and Allen, 2013; Trugman et al., 2019; Wurman et al., 2007). It is thus critical that the simulated P-wave growth over the initial few seconds matches that generally observed by real earthquakes.

P-wave generation capability was added to *FakeQuakes* by Goldberg and Melgar (2020) by extending the radiation scale factor to include conically averaged P-wave radiation, which considers P-wave velocity (V_P) and attenuation (Q_P) at each subfault. Goldberg and Melgar (2020) generated a large Cascadia dataset, which was used to better constrain the empirical M – P_d scaling from Trugman et al. (2019) for magnitudes $M \geq 7$, as observations are limited at these greater magnitudes. Although the magnitude scaling of simulated P_d has already been evaluated using the same simulation methodology and region as this study, Goldberg and Melgar (2020) did not include noise in their simulations when performing the analyses. We find it valuable to re-perform the analyses using noisy simulations to ensure equivalent comparison with observed data.

We process the simulated waveforms following the methodology of Trugman et al. (2019) and Goldberg and Melgar (2020), which includes demeaning the recordings, applying a 0.075 Hz 4th order causal highpass Butterworth filter, double integrating from acceleration to displacement, and applying an additional bandpass filter between 0.075–3 Hz. We also truncate the recordings at 95% the S-wave arrival time to avoid S-wave contamination. The records are normalized to 10 km epicentral distance (R_{epi}) (Kuyuk and Allen, 2013), and P_d is calculated on the vertical components as the maximum amplitude at intervals of 1, 2, 4, 6, 12, and 20 s following the P-wave arrival.

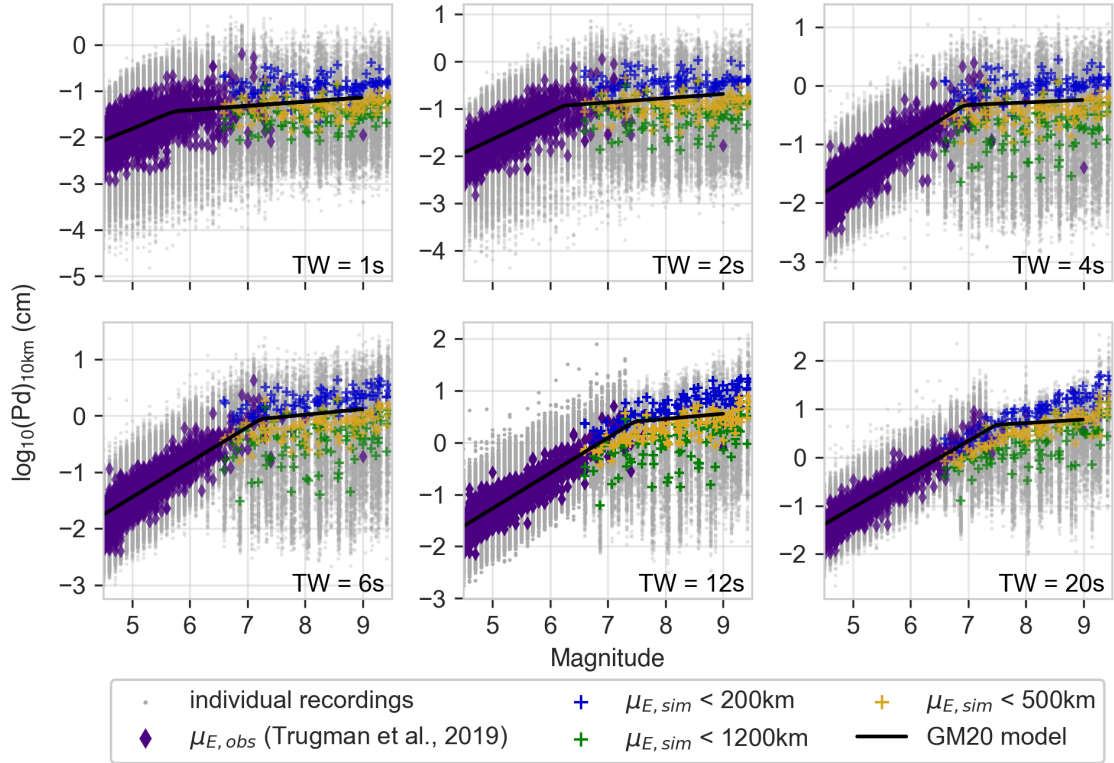


Figure 3.4. Simulation P_d measurements measured over time windows of 1, 2, 4, 6, 12, and 20 s. P_d is corrected to 10 km R_{hyp} . Gray markers indicate P_d measured on individual recordings, whereas colored markers indicate event mean (μ_E) P_d . The purple markers are μ_E 's from observed recordings, and the blue, yellow, and green markers are μ_E 's using only simulations within an R_{epi} of 200, 500, and 1200 km, respectively. The black line represents the model from Goldberg and Melgar, 2020.

Trugman et al. (2019) only considered event recordings with $R_{epi} \leq 200$ km for quality control. Applying this same distance criteria, the event mean (μ_E) P_d measurements from records in our dataset are generally up to $0.5 \log_{10}$ unit higher than the Goldberg and Melgar (2020) model (GM20; Fig. 3.4). The fit is improved for time windows (TW) of 12 and 20 s up to $M \sim 7.5$. Goldberg and Melgar (2020) did not impose the same distance criteria as Trugman et al. (2019), so their dataset consists of records with R_{epi} up to ~ 1000 km. Extending our dataset to include all records results in up to $1.5 \log_{10}$ unit underestimation compared to the model, and we find the best fit to be with records $R_{epi} \leq 500$ km (Fig. 3.4). The variance in the simulated P_d values appears to be much larger

compared to observed P_d values, especially for time windows > 2 s. This is heavily biased by the larger distance recordings in our dataset though. The variance of simulated P_d from recordings within 200 km R_{epi} is within about 1% of the observed variance for all time windows.

With all distance criteria evaluations, $\mu_E P_d$ from our simulations do not exhibit the same level of magnitude saturation observed by the GM20 model. A key deviation in our methodology compared to Goldberg and Melgar (2020) was the use of a two-frequency matched filter for the broadband simulation. The matched filter process involves applying a lowpass filter to the double-integrated low-frequency displacement waveforms and a highpass filter to the high-frequency acceleration waveforms prior to summation. They used a single filter corner frequency (f_c) of 1 Hz for both the lowpass and highpass filters, whereas we chose to use a lowpass f_c of 1 Hz and a highpass f_c of 0.1 Hz. The use of a two-frequency approach minimizes the notch in the broadband spectra where the low- and high-frequency components are combined (see Chapter II), resulting in increased energy in the simulated recordings. This likely contributes to the decreased magnitude saturation. We recognize that this method also adds additional energy outside the bounds of the notch due to filter roll off, primarily for the smaller events in our dataset. Both methods have their limitations, and it is difficult to establish which lends to more accurate results as larger magnitude observations are limited. Regardless, we find that the simulations exhibit the expected trends in P_d evolution, where P_d increases with increased measurement time window, larger magnitudes are better resolved when longer P-wave windows are available, and some level of magnitude saturation is observed due to limitations of seismic processing. Furthermore, both the Kuyuk and Allen (2013) and Crowell et al. (2013) relationships only consider up to the first four seconds of the P-wave, and our simulations are well fit by the model for this short time window.

3.6 Ground-Motion Intensity

Most earthquake early warning studies report performance levels based on the timeliness and accuracy of epicenter location and source characterization. These features,

although crucial for effective earthquake alerts, miss a key component of early warning—the ground-motion predictions themselves. Meier (2017) proposed the best early warning metric to be ground-motion prediction error as a function of warning time which has been suggested for the ShakeAlert alert region (Saunders et al., 2022a). If simulated earthquake data are included in EEW testing, it is necessary that they adequately capture the expected distribution of ground shaking. To that extent, we evaluate the key ground-motion and intensity parameters used in EEW for each of the simulated waveforms and compare these with expected ground motion for earthquakes of equal magnitude. Ideal circumstances would allow for direct comparison of simulated recordings with observed events in the region of interest, but limitations imposed by the dearth of interface events in the CSZ require comparison of the simulated ground motions with predictions from empirical global subduction zone models.

3.6.1 Peak Ground Acceleration and Peak Ground Velocity

Observed peak ground acceleration (PGA) and peak ground velocity (PGV) do not directly inform earthquake alerts for point-source based EEW algorithms, such as EPIC and ONC-EW, but rather real-time magnitude and location estimates are used in ground-motion models (GMMs) to estimate PGA and PGV, which are then used to obtain MMI estimates through scaling relations (Worden et al., 2012). FinDer, the secondary algorithm supporting ShakeAlert, does use two-dimensional spatial template matching to infer a line-source model based on the spatial distribution of observed PGA (Bosë et al., 2018). It is thus valuable to evaluate how the simulated high-frequency ground motions compare with expected ground motions.

The broadband acceleration waveforms are filtered using a 0.075 Hz 2nd order acausal highpass Butterworth filter (Wu and Zhao, 2006), and PGA is computed on the RotD50 of the horizontal components. The simulated PGA values are compared with predictions from both the BCHydro (Abrahamson et al., 2016, 2018) and NGA-Sub (Parker et al., 2020) GMMs. Filtered acceleration waveforms are integrated and highpass filtered a second time, and PGV is computed on the RotD50 of the horizontal components. For PGV

analysis, we only consider NGA-Sub predictions as BCHydro does not have regression coefficients for PGV. These models were derived using global interface and intraslab recordings, including recordings from Cascadia. However, due to the few recordings available from Cascadia, these were not included in the interface model for BCHydro and were mainly used to develop regional adjustment factors for V_{S30} scaling and basin amplification terms for NGA-Sub. These models are considered applicable for magnitudes up to **M**9.5 and closest rupture distances (R_{rup}) out to 1000 km, thus applicable for our simulated data ranges.

We evaluate the simulated PGA by computing residuals as the $\ln(\text{GMM PGA prediction/simulation PGA})$. Well-modeled ground motions should exhibit relatively no bias with either magnitude or distance. Initial analyses showed a substantial negative bias with distance, indicating simulated PGA values were much larger than predicted (Fig. 3.S6). Until now, *FakeQuakes* has predominantly been used to generate low-frequency geodetic waveforms, and peak ground accelerations have been minimally validated. Goldberg and Melgar (2020) compared simulated PGA with observed values for the 2014 **M**8.1 Iquique, Chile event, and found the mean PGA residuals to be within 0.5 \ln unit. However, they presented the entire ensemble of PGA residuals as a single relation to the mean, so it is unclear whether a distance bias was also observed in their dataset. Furthermore, their 2014 Iquique simulations were primarily within 100–200 km and would be less subject to attenuation variations.

To improve the PGA–attenuation relation in *FakeQuakes* to be more suitable for future EEW validation and training applications, the following adjustments were made to the source code. Previously, the direct arrivals used to generate the S-wave packets were taken as the first arriving ray, which was nearly always the ray turning below the Moho. The deep propagation of these rays results in less attenuation compared to crustal rays, so now only crustal rays are considered for the direct P- and S-wave arrivals. The geometric spreading term was changed from $1/r$ to $1/r^{1.2}$ (Sahakian et al., 2018). Lastly, an elastic attenuation term (Q_C) was added with the form $Q_{C,base}f^n$ (Farahbod et al., 2016; Pujades

et al., 1996), and we found a Q_C of $150f^0$ to best capture the missing attenuation effects. Total attenuation (Q_{total}) is now modeled as a function of Q_C and shear-wave attenuation (Q_S) (Pujades et al., 1997):

$$Q_{total} = \frac{1}{1/Q_S + 1/Q_C} \quad (3.1)$$

Additionally, the frequency factor of Q_S was changed from 0.6 to 0.8 to improve attenuation modeling. Atkinson and Macias (2009) determined the appropriate frequency factor for the Cascadia region to be nearly half that at 0.45, but this was solved for using only four events and three stations. To account for the addition of another attenuation term, Q_P and Q_S were reduced in the velocity model. The results in the previous sections are based on the updated source code and velocity model.

With these modifications, we observe a substantial reduction in PGA distance bias (Fig. 3.5a–b). Overall, the PGA residuals average close to zero. We do observe a slight positive residual trend with magnitude where recordings from events $< \mathbf{M7.75}$ average higher than predicted, whereas recordings from events $> \mathbf{M7.75}$ average lower than predicted (Fig. 3.5a). One might interpret this observation as an issue with finite rupture characterization or source scaling in the high-frequency generation. However, this is more likely due to attenuation modeling at large distances. When plotted as a function of R_{rup} , PGA residuals show relatively no bias out to ~ 600 km (Fig. 3.5b), beyond which there is a strong negative trend where simulations have a higher PGA on average than predicted. Recordings with the largest R_{rup} come from the smaller magnitude events in our dataset. This attenuation bias is likely due to several factors. From a modeling perspective, a 1D velocity model simply cannot capture complex 3D structure and attenuation effects present in real recordings (Fadugba et al., 2023). The various attenuation parameters may also need to be adjusted to better model attenuation at these larger distances. Additionally, the GMMs have a magnitude–attenuation relationship which is absent in *FakeQuakes*. It is also important to note that, although BCHydro and NGA-Sub are considered applicable for distances up to 1000 km, the models are not well constrained at the upper end of this range. The BCHydro and NGA-Sub datasets are primarily comprised of recordings within

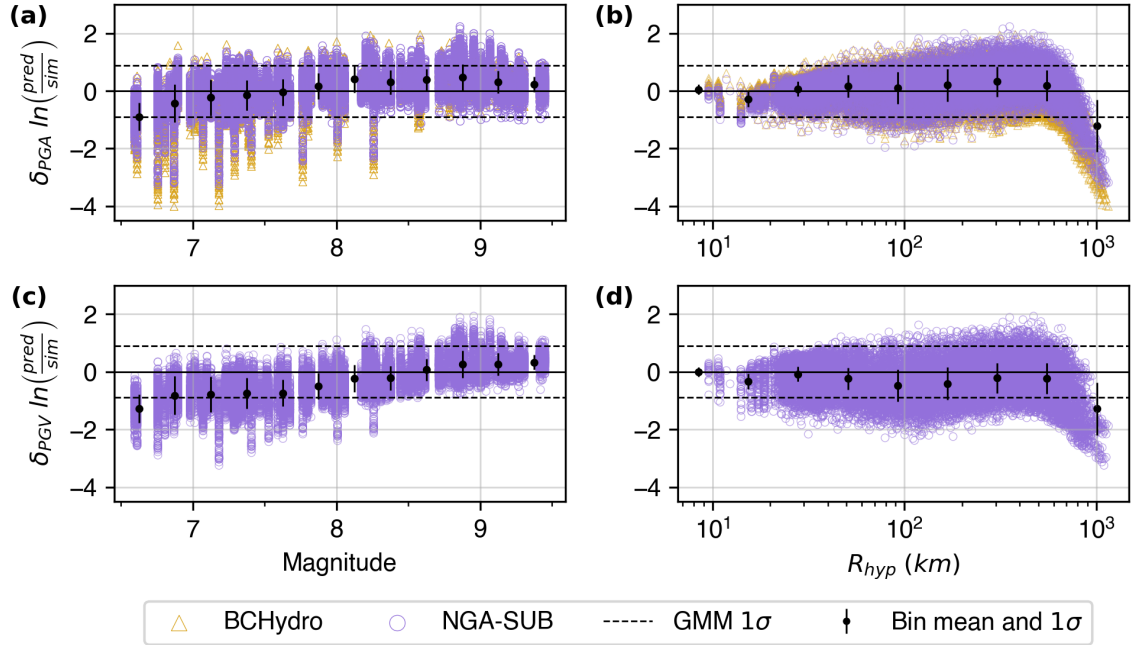


Figure 3.5. Residuals between ground-motion model predictions (pred) and simulated ground motions (sim) for (a–b) PGA (δ_{PGA}) and (c–d) PGV (δ_{PGV}) as a function of magnitude (left) and distance (right). Black circles indicate binned mean residuals, and the dashed black lines represent the GMM standard deviation. For δ_{PGA} (a–b), the standard deviation is the maximum standard deviation between BCHydro and NGA-Sub. For δ_{PGV} (c–d), the standard deviation is the NGA-Sub standard deviation.

300 km and 600 km R_{rup} , respectively. Regardless of the source of this bias, expected PGA at these distances (whether predicted or modeled) would likely be too small to be felt.

We evaluate the simulated PGV by computing residuals as the $\ln(\text{GMM PGV prediction/simulation PGV})$, and we find PGV residual trends to be similar to PGA (Fig. 3.5c–d). The residuals exhibit a slight positive residual trend with magnitude, but are relatively consistent for distances out to 600 km. Unlike PGA, PGV residual averages are systematically lower than predicted for distances between 20–600 km. The mean and standard deviations of the PGV residuals still fall within the NGA-Sub standard deviation for distances out to 600 km.

3.6.2 *Modified Mercalli Intensity*

Following source characterization, alerts are issued based on expected MMI levels. Although qualitative compared to PGA and PGV, MMI is the parameter used for alerts as it is more intuitive, indicating the level of shaking and damage one will likely experience (e.g., MMI IV suggests light shaking will be felt by many people indoors and some people outdoors, and that household objects may be disturbed). The MMI level for which alerts are issued has been a topic of discussion for years and holds significant societal consequences. Too high of a threshold will result in possible missed alerts, whereas too low of a threshold will result in more frequent over-alerting (Kohler et al., 2020; Saunders et al., 2022b). Thus, the desired alert level is subject to vary depending on the intended receivers and their sensitivity to missed or false alerts. We evaluate the simulation MMI distributions to ensure they adequately capture the expected shaking intensities alerted by the EEW systems.

Although a final reported MMI is obtained from damage assessments and felt reports, MMI is obtained through intensity measure scaling relations for EEW alerts. There has been discussion on whether MMI should be obtained from PGA or PGV, or some combination of the two depending on event characteristics (Atkinson and Kaka, 2003; Mooney and Wange, 2014). For this analysis, we present MMI results obtained through both the PGA and PGV scaling models from Worden et al. (2012). For MMI, residuals are calculated as GMM MMI prediction – simulation MMI. Again, we observe a slight trend with magnitude and no distance bias for distances out to 600 km for both MMI_{PGA} and MMI_{PGV} (Fig. 3.6). Although averaging close to zero, there is a bulge in the residuals around 20–200 km, where the variance is greater compared to PGA and PGV. Specifically looking at MMI_{PGV} , residuals are systematically lower than predicted, as with PGV (Fig. 3.6d). The mean and standard deviations still fall mostly within the standard deviation of the Worden et al. (2012) model, and the negative bias at distances greater than 600 km is less severe for MMI compared to PGA and PGV.

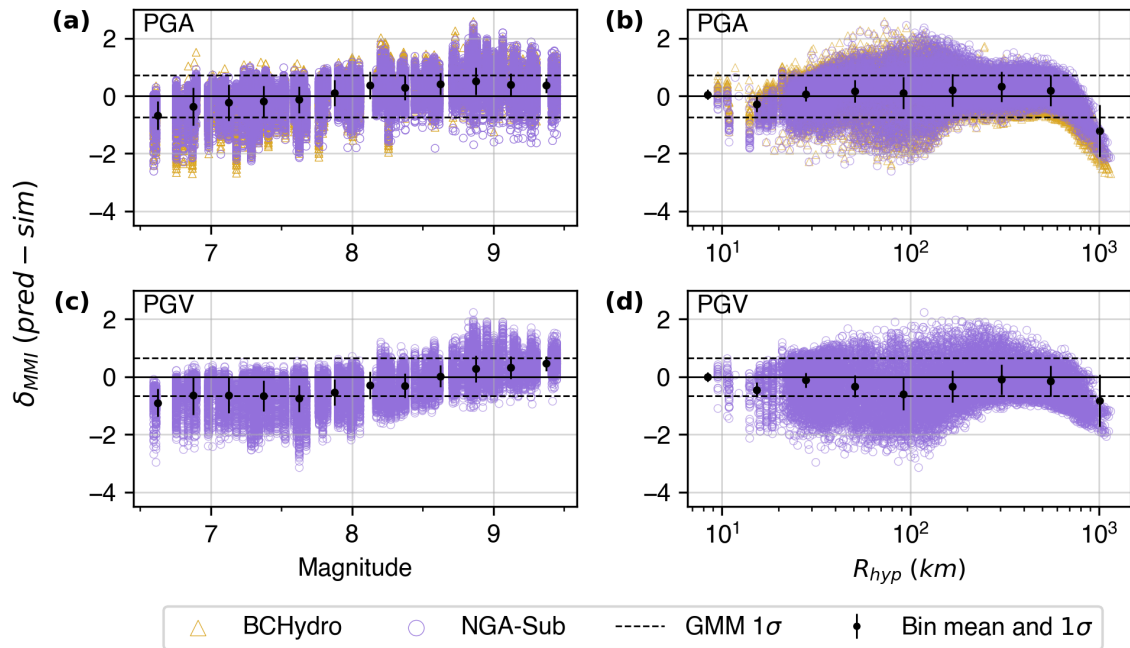


Figure 3.6. MMI residuals (δ_{MMI}) between ground-motion model predictions (pred) and simulated ground motions (sim) for (a–b) MMI computed using PGA scaling and (c–d) MMI computed using PGV scaling as a function of magnitude (left) and distance (right). Black circles indicate binned mean residuals, and the dashed black lines represent the GMM standard deviation from Worden et al. (2012).

3.6.3 Peak Ground Displacement

Magnitude estimation from seismic recordings is subject to saturation for larger events due to limitations in inertial instruments and seismic processing, which presents challenges for effective alerting of a major megathrust event. PGD from geodetic recordings, on the other hand, does not suffer from magnitude saturation and can thus help better constrain ground-motion estimates for large events. Currently, PGD is not actively used in ShakeAlert; however, three geodetic early warning algorithms have previously undergone testing— G-larmS (Grapenthin et al., 2014a, 2014b), G-FAST (Crowell et al., 2016), and BEFORES (Minson et al., 2014). G-FAST, which includes PGD-magnitude scaling as a key element, has been selected for implementation into ShakeAlert (Murray et al., 2023). Because ONC-EW’s network consists of collocated HR-GNSS and strong

motion instruments, HR-GNSS data are already being incorporated into their public alerting notification framework. ONC-EW obtains seismogeodetic magnitude estimates for moderate-to-large events from the Crowell et al. (2013) PGD– M scaling relation. For such algorithms, it is necessary that PGD be well-modeled by our simulations.

We use HR-GNSS displacement waveforms from the collocated ONC sensors for PGD analysis. We evaluate the effects of different noise percentiles (10th, 50th, and 90th) on simulated PGD and how these values compare to GMM predictions. PGD is computed on the three-component seismogram (e.g., Melgar et al., 2015; Goldberg et al., 2021)

$$PGD = \max \left(\sqrt{[N(t)^2 + E(t)^2 + Z(t)^2]} \right), \quad (3.2)$$

and we compare simulated PGD with the Goldberg et al., (2021; GA21) model. This model contains the most up to date PGD scaling coefficients and considers the generalized mean rupture distance (R_p), which better accounts for fault finiteness of large ruptures and slip heterogeneity compared to the more commonly used hypocentral distance, and it has shown to better represent unilateral ruptures.

$$R_p = \left(\sum_{i=1}^n w_i R_i^p \right)^{1/p}. \quad (3.3)$$

Here, n is the number of subfaults, w_i is the weight of the i th subfault, R_i is the distance to the i th subfault, and p is the power of the mean. Goldberg et al. (2021) derived three sets of coefficients for datasets consisting of only observed events, only simulated events, and a joint observed–simulated dataset. Recordings from large events are limited, so to better constrain their GMM for larger endmember events, they added Cascadia simulations; however their simulations did not contain any noise. We compare our simulated PGD with predictions using both their observed GMM and joint GMM, which use the distance measures $R_p^{-4.5}$ and $R_p^{-2.3}$, respectively.

$$\log_{10}(PGD) = -3.841 + 0.919(M) - 0.122(M)\log_{10}R^{-4.5} \quad (3.4)$$

$$\log_{10}(PGD) = -5.902 + 1.303(M) - 0.168(M)\log_{10}R^{-2.3} \quad (3.5)$$

Simulated HR-GNSS recordings lacking the addition of noise present systematically lower PGD than predicted by GA21, but this deficiency does not appear to be biased by magnitude or hypocentral distance (R_{hyp}) (Fig. 3.7a–b). The means of these residuals are mostly between 0.5–1 ln unit, but individual recording residuals range above 2 ln units. Melgar et al. (2016) observed a bias in PGD from *FakeQuakes* simulations; however, their recordings were primarily larger than predicted from the Melgar et al. (2015) PGD GMM model, directly opposing residual trends from this study. Note that they compute residuals as $ln(\text{simulation}/\text{prediction})$, so although their residuals are mostly positive, they have a different implication from the positive residuals we observe. They also observed a magnitude bias, where larger magnitude events had a greater misfit. This discrepancy between our study and Melgar et al. (2016) could be due to the difference in PGD GMMs used for validation. *FakeQuakes* has also undergone many modifications since its initial publication, which could further lend to the difference in residual trends. A more recent study (Fadugba et al., 2023) compared simulations generated with *FakeQuakes* (1D velocity structure) and SW4 (3D velocity structure) with observations from several Japanese events. They observed PGD residuals (calculated as $ln(\text{observation}/\text{simulation})$) in line with those in this study, where they are primarily positive without magnitude or distance bias. They attributed this to the simplification of the earth structure using a 1D velocity model as the 3D simulations had greatly improved residuals, although still systematically positive.

Another source of this bias could be the presence of noise. The similar frequency content between the HR-GNSS signal and noise hinders the ability to cleanly filter out the noise. Thus, it is reasonable to assume observed PGD measurements, and subsequently the GMM models regressed on such datasets, are biased by HR-GNSS noise, especially for smaller magnitude events (Fig. 3.8a–b). Figure 3.7c–h highlights the effects of adding 10th, 50th, and 90th percentile noise on PGD residuals. With the addition of noise, we do start to observe both a positive magnitude bias and a negative distance bias. With the addition of 10th percentile noise, PGD from events with magnitudes $> M7.75$ are primarily

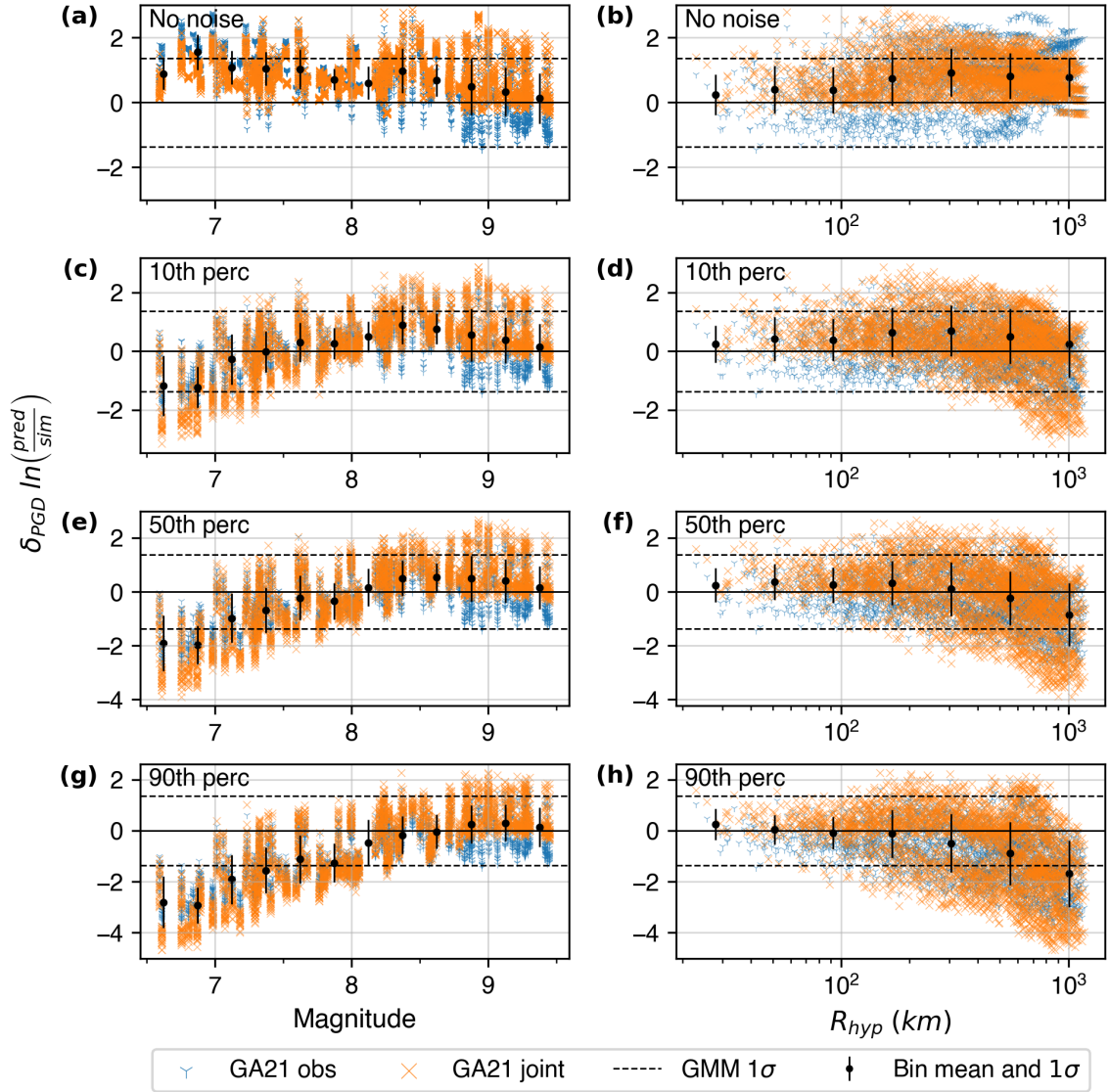


Figure 3.7. PGD Residuals (δ_{PGD}) between ground-motion model predictions (pred) and simulated ground motions (sim) as a function of magnitude (left) and distance (right) with (a–b) no noise added, (c–d) 10th percentile noise added, (e–f) 50th percentile noise added, and (g–h) 90th percentile noise added. Black circles indicate binned mean residuals, and the dashed black lines represent the maximum standard deviation between MA15 and GA21.

unaffected, and events with magnitudes $< M7.75$ are fairly well-modeled (i.e., bin means are within the model uncertainty) compared to GA21 (Fig. 3.7c). For $R_{hyp} > 500$ km, PGD residuals become increasingly negative; however, all bin means fall well within the GMM uncertainty for the conditions (Fig. 3.7d). With the addition of 50th percentile noise,

we start to observe an effect for magnitudes $< \mathbf{M}8.5$, but records from events $> \mathbf{M}7$ are still well-modeled (Fig. 3.7e). The distance bias is not as strongly affected by the addition of noise, and residuals from records at all distances are still within the model uncertainty (Fig. 3.7f). Simulated PGD calculations are most strongly influenced by the addition of 90th percentile noise (Fig. 3.7g–h). With this level of HR-GNSS noise, simulations with magnitudes $> \mathbf{M}7.5$ and $R_{hyp} < 500$ km are well-modeled.

The δ_{PGD} bias with increased noise levels arises from a decrease in the SNR. Larger magnitude events have a stronger signal and are more easily distinguishable from the background noise (Fig. 3.8). For example, Figure 3.8a–b compares simulated HR-GNSS waveforms with various noise levels from a $\mathbf{M}7.3$ and a $\mathbf{M}9.0$ event. The $\mathbf{M}7.3$ main signal is indistinguishable with the addition of 90th percentile noise and is barely discernable even with the addition of lower noise levels. Conversely, the $\mathbf{M}9.0$ event signal is still discernable with all noise levels, and PGD is only minorly affected. Figure 3.7c–d highlights the relationship between event magnitude, SNR, and the resulting δ_{PGD} for the different noise levels. Intuitively, lower levels of noise result in increased SNR, and larger magnitude events have higher SNRs overall (Fig. 3.8c). Relating this to δ_{PGD} , we find that residuals are large and highly scattered for recordings with a $\text{SNR} < 4$, which occurs for nearly all events with $\mathbf{M} < 7$, regardless of noise levels. δ_{PGD} trends towards zero with increased SNR, and records with a $\text{SNR} > 50$ are consistently within the GMM uncertainty, which occurs for events with $\mathbf{M} > 8$.

PGD is challenging to work with, both for modeling and real-time analyses. It is highly sensitive to complex 3D structure, and the high noise levels in the data render quality analysis an arduous task for even the most experienced geodesists. Even though data used for the GA21 model underwent a manual inspection to reduce bias from poor quality records, a more in-depth quality control, such as one involving SNR, was not used. Furthermore, PGD scaling laws are based on a point source assumption, which breaks down for events greater than $\mathbf{M} \sim 6$, well below the magnitudes of the events in our dataset. Considering these factors, we find the simulated HR-GNSS data to be well-

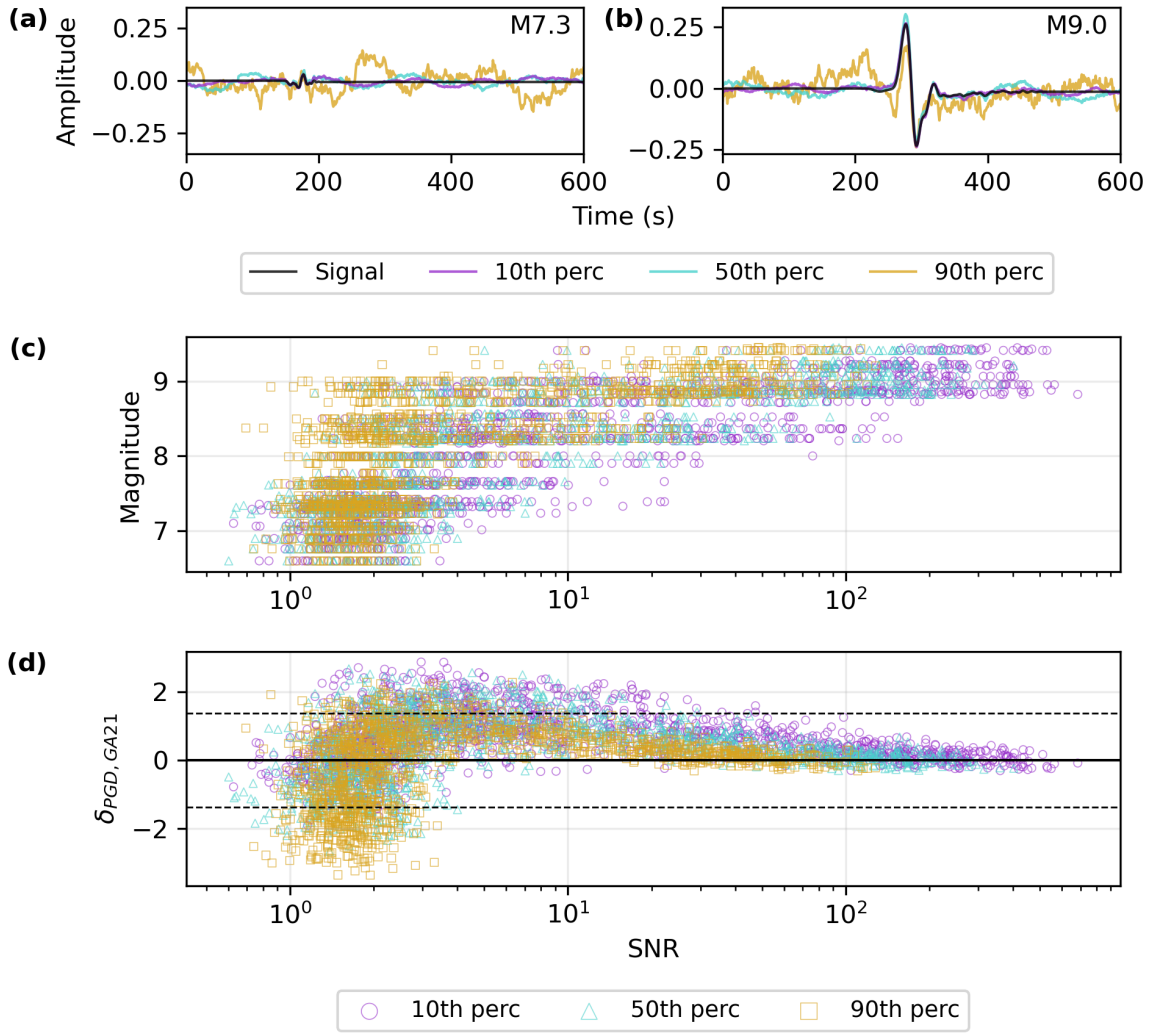


Figure 3.8. Evaluation of synthetic GNSS noise and its effect on apparent PGD. (a) The raw signal waveform and the signal with additions of synthetic noise at 10th, 50th, and 90th percentiles for a M7.3 event. (b) Same as (a) but for a M9.0 event. Relationship of SNR with (c) event magnitude and (d) PGD residuals (δ_{PGD}) using the Goldberg et al. (2021) joint model for records with $R_{rup} \leq 400$ km. Marker symbols denote the percentile of synthetic noise added to the GNSS records. The dashed black lines represent the GMM standard deviation from Figure 3.7.

modeled overall. Although, we do caution the use of recordings with a $SNR < 10$ for real-time testing.

3.6.4 *Ground-Motion Summary*

Considering the infrequency of large subduction zone earthquakes, ground-motion models for such events are regressed on a sparse dataset of global events. Thus, predictions from such models have both high epistemic uncertainty (poor resolution due to small number of events) and high aleatory uncertainty (regional variations not accounted for in the models). We recognize that ensuring our simulated data are reasonably modeled by such a GMM may result in poor representation of individual study regions. For example, Atkinson and Boore (2003) find that ground motions at geologically similar sites from similar events in Cascadia and Japan result in ground motion amplitude differences of more than a factor of two. Nevertheless, we believe these generalized ground motions are sufficient for EEW testing purposes, and they can be further improved upon as more large earthquakes are recorded in the future.

3.7 EPIC and ONC-EW Performance Analysis

The previous validation analyses demonstrate the success of using a 1D semistochastic model to simulate waveforms that capture necessary features for EEW testing. We now present a first-order performance analysis of Cascadia region EEW in the event of a large megathrust rupture by running six of the scenarios through EPIC and ONC-EW's offline systems. EPIC's event detection boundaries extend from California up through most of Vancouver Island, but ONC-EW's detection grid is focused primarily north of Oregon (Kohler et al., 2020; Schlesinger et al., 2021). We have thus selected two $M \sim 7, 8, \text{ and } 9$ events (six total) with either hypocenters or main slip patches occurring above a latitude of 46°N to replicate scenarios where both EEW systems would be activated (Fig. 3.9). For the EPIC tests, only recordings for the land-based PNSN and ONC strong motion sensors are incorporated, whereas all available recordings are incorporated for the ONC-EW tests, including offshore stations. For the HR-GNSS data, recordings with 50th percentile noise are used. Although cross-border sensors are not currently used in either EEW system, there are efforts being made to incorporate them in the future. Because

PNSN stations are used in the ONC-EW analysis, their detection grid is extended down to northern California. Additionally, a P_d /PGD SNR flag is disabled in the ONC-EW analysis, which was preventing many of the HR-GNSS recordings from being incorporated. Considering the simulation characteristics are in line with expected values based on our previous analyses, this is likely a limitation of the error threshold defined by ONC-EW and should be re-evaluated.

EPIC and ONC-EW both detected each of the trial events and would have issued an alert in a real-time scenario (Fig. 3.11). Chung et al. (2017) reported that ElarmS-3 (the parent algorithm of EPIC) reported several missed or false events located offshore Oregon and Washington due to the sparse networks, compared to California. As these were real, recent earthquakes, they were smaller in magnitude compared to the scenario events in this study. The Cascadia region EEW shows promise for detecting and alerting a large megathrust event compared to smaller offshore events. Another known challenge is that for larger events, ShakeAlert sometimes reports them as two smaller events, rather than one large event (Hartog et al., 2016), which can significantly affect the issued alerts. None of the six trial events in this study were recorded as split events. Recently, EPIC has modified their algorithm to disallow triggers of recordings > 200 km from being used in any new event associations while the current one is active, which proves to successfully reduce split events.

The evolution of magnitude estimation for the events is shown in Figure 3.10. For the $M \sim 7$ events (i.e., scenario events 15 and 18), the maximum magnitudes estimated by EPIC are within 0.1 magnitude unit of the true event magnitudes. For the $M \sim 8, 9$ events (i.e., scenario events 51, 65, 96, and 108), the maximum reported magnitudes by EPIC are within 2 magnitude units of the true magnitudes. For each event, EPIC underestimates the true magnitude and generally saturates between $M 6.5-7$ depending on the event. This saturation is expected because the P_d scaling relation used by EPIC is based upon the first 4 seconds of the P-wave (Kuyuk and Allen, 2013), which can only resolve magnitudes up to $M \sim 7$ (Fig. 3.4). This is a common limitation for most early warning systems dependent

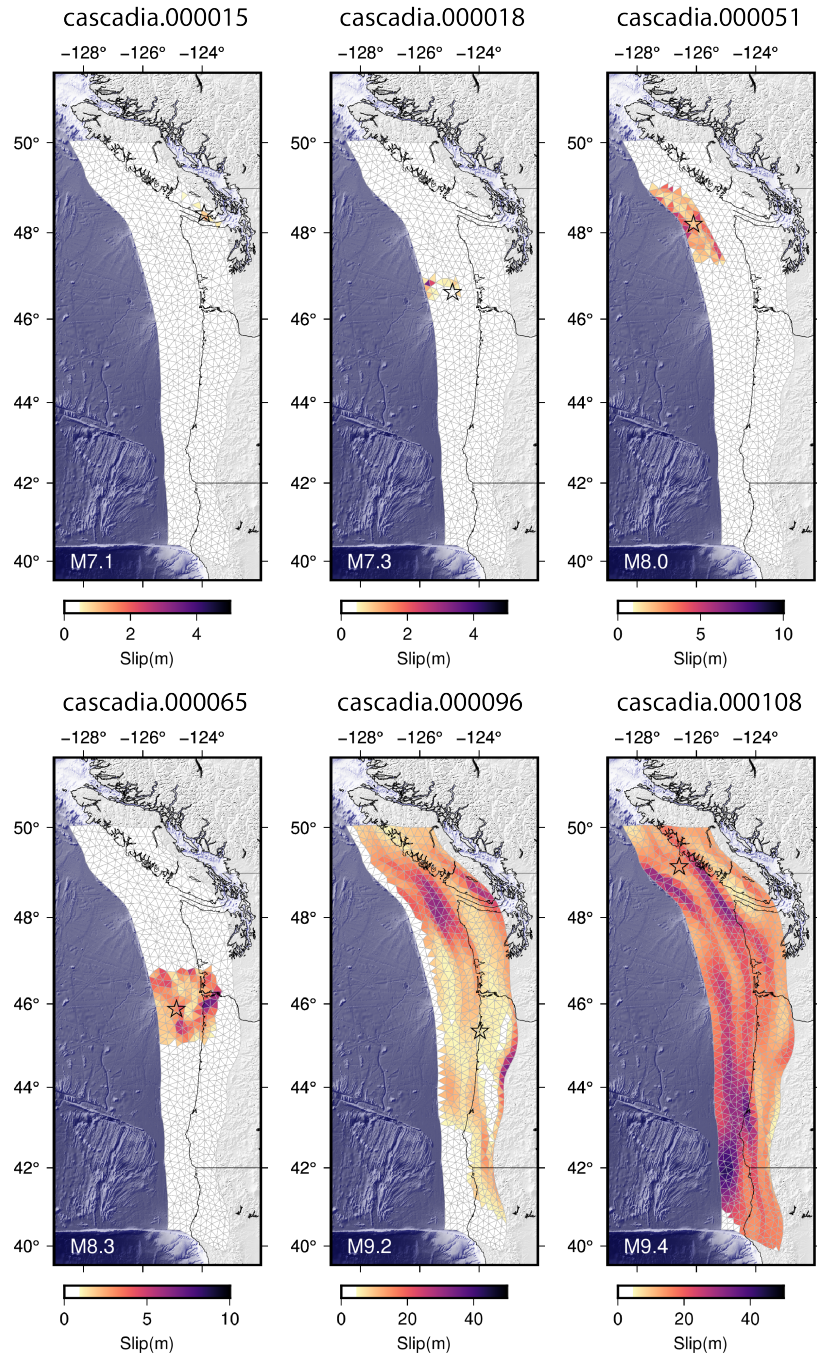


Figure 3.9. Slip patterns of the six scenario events tested with EPIC and ONC-EW. The hypocenters are indicated by the stars on each plot.

solely on seismic data, but ONC-EW's magnitude estimates show a sizeable improvement with their incorporation of geodetic data. Note that the magnitude evolution in Figure 3.10

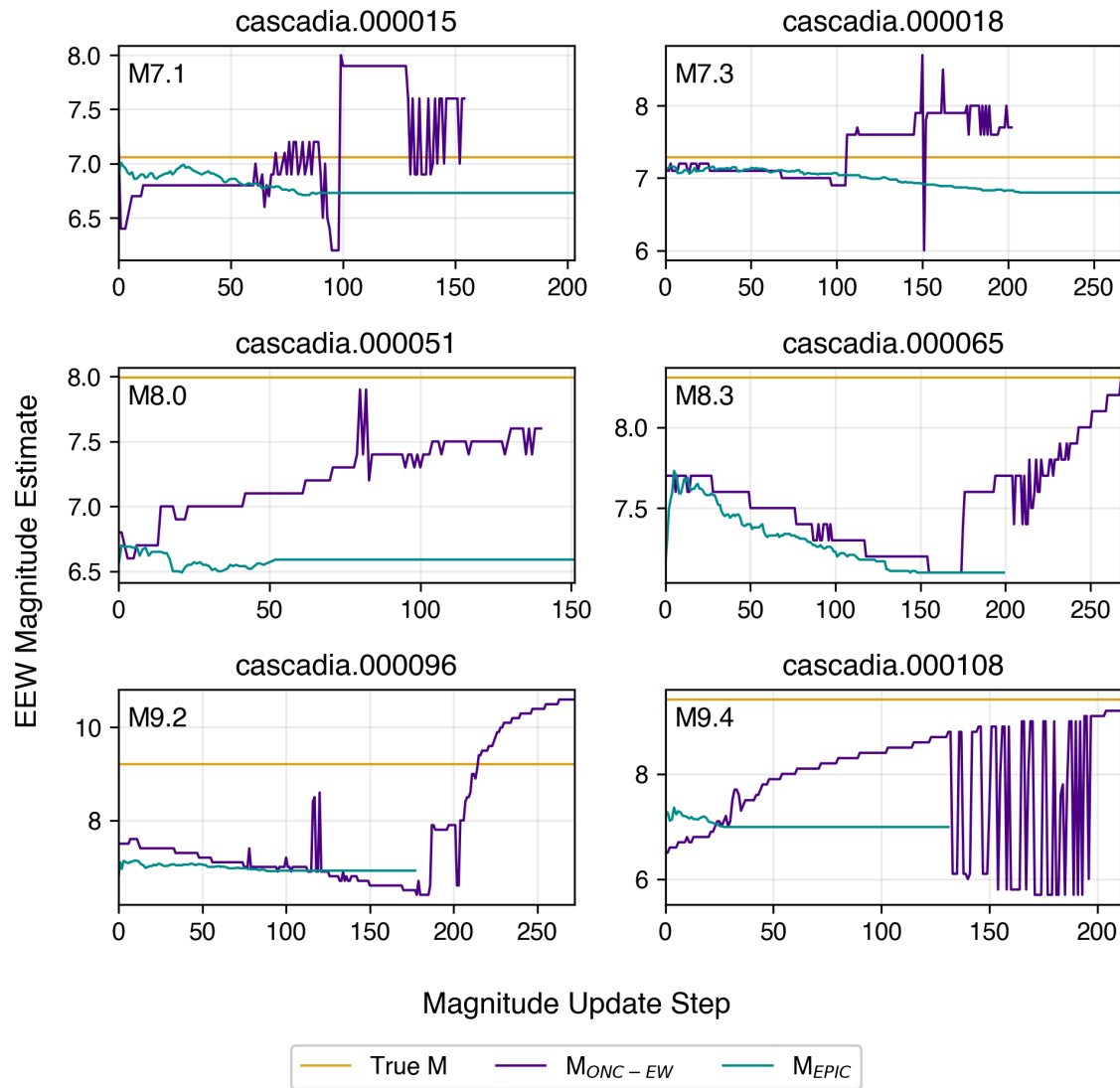


Figure 3.10. Magnitude estimate evolution for six scenario events using both EPIC’s (blue) and ONC-EW’s (purple) offline system. The true magnitude is indicated by the yellow line. The magnitude estimates are updated at irregular intervals as new stations are added, so these figures do not directly show the time-evolution of the magnitude estimates, and update steps do not necessarily correspond to the same time between EPIC and ONC-EW.

is shown as a function of update step, which varies between EPIC and ONC-EW as well as within an individual system as it is related to available real-time computational resources; thus, times are not equivalent between systems in these plots.

The unique network configuration of ONC-EW allows for the formation of hybrid

waveforms which combine the high-frequency data from the strong motion instruments with unsaturated displacement data from the HR-GNSS instruments using a Kalman filter (Bock et al., 2011; Kalman, 1960; Melgar et al., 2013; Rosenberger et al., 2018; Smyth and Wu, 2007). The resulting un-biased displacement and velocity time-series are used for all steps in the EEW system workflow. The initial magnitude estimate reported by ONC-EW is obtained using the Crowell et al. (2013) scaling with P_d measured on the un-biased displacement data and is later updated using the Crowell et al. (2013) PGD relation for events with $M > 6.5$. PGD measured on geodetic data does not suffer from magnitude saturation, and allows for more accurate magnitude estimates, especially for the largest events. The maximum magnitude estimated by ONC-EW for the $M \sim 7$ events is a full magnitude higher than the true magnitude but drops within 0.5 magnitude unit for the final estimates. This is likely due to the presence of noise in the HR-GNSS data. With the addition of 50th percentile noise, PGD is on average $2 \ln$ units larger than expected for a $M7$ event (Fig. 3.7), thus biasing magnitude estimates. This no longer appears to be an issue for the $M \sim 8, 9$ events, except for scenario event 96, for which ONC-EW reports a final magnitude of 10.6. An event of such magnitude is physically impossible, but this overestimation is a limitation of point source PGD GMMs when finite rupture presents large slip patches far from the hypocenter, as is the case for event 96 (Williamson et al., 2020).

We do not include earthquake location validation tests for this study as they are highly dependent on network configuration, and we are not using the full network of stations utilized by EPIC and ONC-EW. However, via the offline tests, we find that EPIC and ONC-EW resolve the epicenters within several kilometers (Fig. 3.11).

3.8 Conclusions

The Pacific Northwest is a region of high seismic risk, and when a $M9$ earthquake inevitably ruptures, it will leave thousands of unreinforced buildings, bridges, and other critical infrastructure inoperative. In such a scenario, efficient and reliable EEW alerts will be critical for minimizing injuries and disruption from the event. Although the U.S. West

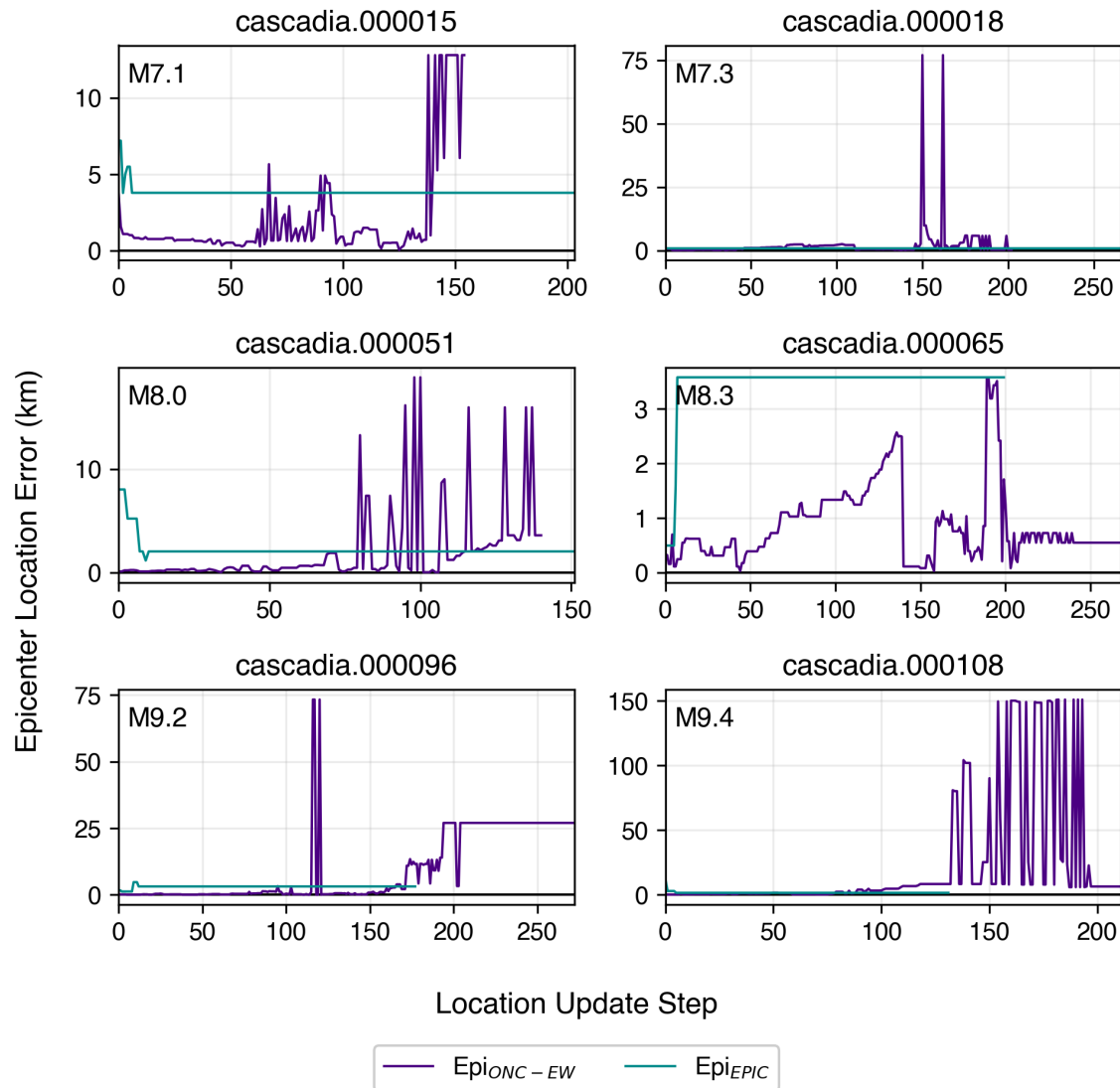


Figure 3.11. Epicenter location error evolution for six scenario events using both EPIC’s (blue) and ONC-EW’s (purple) offline system. The epicenter locations are updated at irregular intervals as new stations are added, so these figures do not directly show the time-evolution of the location error, and update steps do not necessarily correspond to the same time between EPIC and ONC-EW.

Coast and Western Canada have operational EEW systems, it is currently unclear how these systems would perform in the event of a major megathrust event. ShakeAlert in the U.S. has been operational for years but has primarily been tested using real-time or replay data from small to moderate crustal events. ONC-EW in Canada has only become operational

within the past couple years and has recorded only a handful of moderate regional events in that time. It is becoming clear that in the absence of real seismic data, simulated data are the way forward for continued testing and training of these EEW systems.

We have generated a dataset of moderate to large CSZ ruptures and waveforms using *FakeQuakes*, which can be used to evaluate the performance of both ShakeAlert and ONC-EW and identify components requiring refinement for early warning of a large megathrust event. We have performed several validation tests on the simulations to ensure they capture the key features important for EEW, including their triggering capability for earthquake detection, P-wave growth for magnitude estimation, and ground-motion intensity distributions. We find that the simulations are well-modeled and can be a useful testing dataset. Cascadia region simulations could be further improved with the addition of a coupling model (Small and Melgar, 2021), improved velocity resolution from the recent Cascadia2021 initiative (Ashraf et al., 20223) and upcoming efforts from the newly funded Cascadia Region Earthquake Science Center (CRESCENT), as well as waveforms generated using realistic 3D Earth structure (Fadugba et al., 2023).

EPIC and ONC-EW both use a direct grid search algorithm, whereas ONC-EW additionally uses a linear least squares algorithm for epicenter location (Chung et al., 2019; Schlesinger et al., 2021). For the scenario events, EPIC's maximum location error is within 12 km, and it quickly resolves a final epicenter estimate within 4 km of the true epicenter. ONC-EW estimates show greater variance with maximum location errors ranging over 100 km, but it is able to resolve a final epicenter within 4 km for the scenario events 18, 51, and 65. The location errors for these events are slightly higher than the 2 km median error reported by ElarmS-3 for crustal events in California (Chung et al., 2019). Considering the scenario events are offshore and more complex than those in California, EPIC and ONC-EW perform well.

For this preliminary analysis of Cascadia EEW system performance, we did not evaluate ground-motion accuracy. EPIC itself does not report MMI, but rather the ShakeAlert decision module determines MMI contours based on information provided by

both EPIC and FinDer. It would be valuable to run these simulations through FinDer in a future study, especially considering FinDer is better designed to characterize large events compared to the point source algorithm EPIC. Six of the scenarios were run through EPIC and ONC-EW, and with this first-order analysis, we find that EPIC and ONC-EW are successful at locating the epicenter, which is key for accurate alert times. EPIC's magnitude estimates saturate as expected, but with the addition of HR-GNSS data, ONC-EW's magnitude estimates are closer to the true magnitudes. ONC-EW does, however, overestimate with smaller events due to the lower SNR and with larger events if large slip patches occur at far distances from the epicenter. We recommend that operators of ShakeAlert and ONC-EW perform more in-depth performance analyses using all simulations in the dataset to more comprehensively identify possible challenges that may arise in a large complex rupture of the CSZ.

3.9 Data and Code Availability

Noise data streams for ONC-operated onshore and offshore sites were downloaded from <http://ds.iris.edu/mda/OW> and <http://ds.iris.edu/mda/NV>, respectively. Maps were produced using the Generic Mapping Tools software (Wessel et al., 2019), with a digital elevation model downloaded from GMTSAR (Sandwell et al., 2011). Simulations were generated using codes accessed from <https://github.com/dmelgarm/MudPy>, and all analyses were completed with in-house code, available online at https://github.com/taranye96/cascadia_EEW. The supporting information contains figures illustrating the Cascadia fault and velocity structure used for simulations, offshore noise levels, late or missed STA/LTA triggers, and poor attenuation modeling.

3.10 Acknowledgments

This work was partially supported by a contract from Ocean Networks Canada and NASA ROSES grant 80NSSC21K0841.

3.11 Bridge

This chapter has focused on Cascadia Subduction Zone simulations and ensuring they capture the necessary characteristics to be an effective training set for regional earthquake early warning. The validation tests performed in this chapter primarily focused on the source and path components of the 1D semistochastic simulation model, and the simulations are found to be well-modeled compared to global empirical models. These simulations can be used to evaluate the performance of ShakeAlert and ONC-EW in the event of a large megathrust event. The final chapter addresses the site component of ground motion by estimating site-specific high-frequency attenuation in the San Francisco Bay Area. These estimates can be used to improve estimates of ground-motion intensity in the region.

3.12 Supporting Information

3.12.1 Model Files

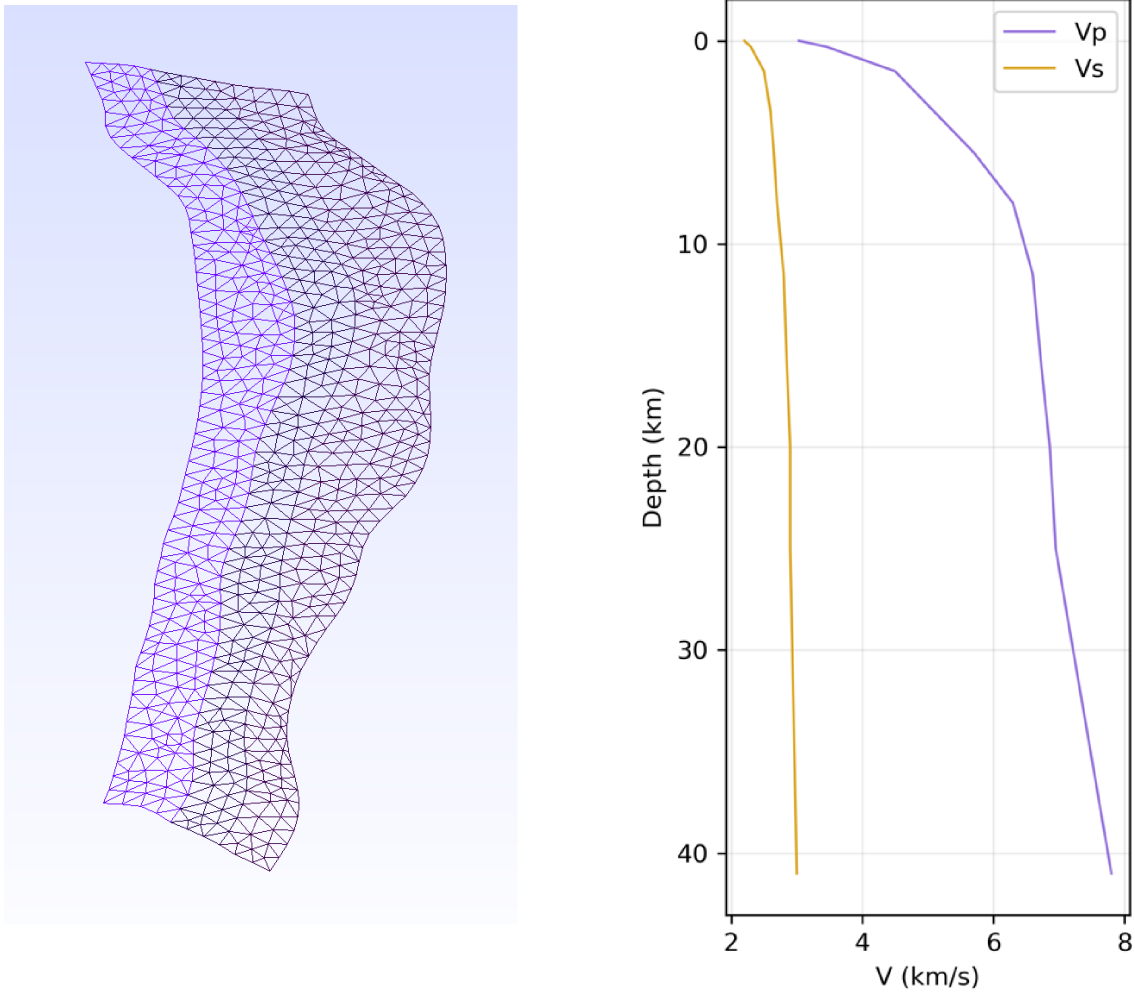


Figure 3.S1. (left) Mesh file of the fault model and (right) velocity model used to generate Cascadia rupture scenarios and waveforms. These were taken from (Melgar et al., 2016).

3.12.2 Offshore Noise

The following equation is used to calculate the decibel difference in noise (db) between each offshore site and one coastal onshore site (AL2H), which is listed in Table 3.S1:

$$db = 20 \log_{10} \left(\frac{|\text{offshore noise}|}{|\text{AL2H noise}|} \right). \quad (3.S1)$$

Table 3.S1. Offshore noise variation from AL2H in decibels (db).

Station*	db from AL2H Noise
BACME	-16.63
CBC27	-21.413
NC89	-6.12
CQS64	-19.7

*Note: Offshore station NCBC is not listed because it was dug out prior to 02-20-2022. An earlier date was not used for comparison because land site data are not archived or sent continuously to IRIS, so I only have access to very recent data for those sites.

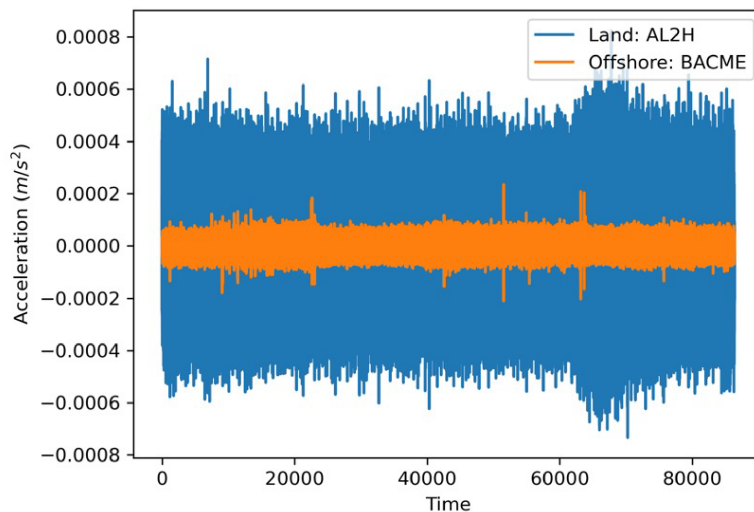


Figure 3.S2. Station noise comparison for the land site AL2H and the offshore site BACME from 02-20-2022.

3.12.3 STA/LTA Triggering

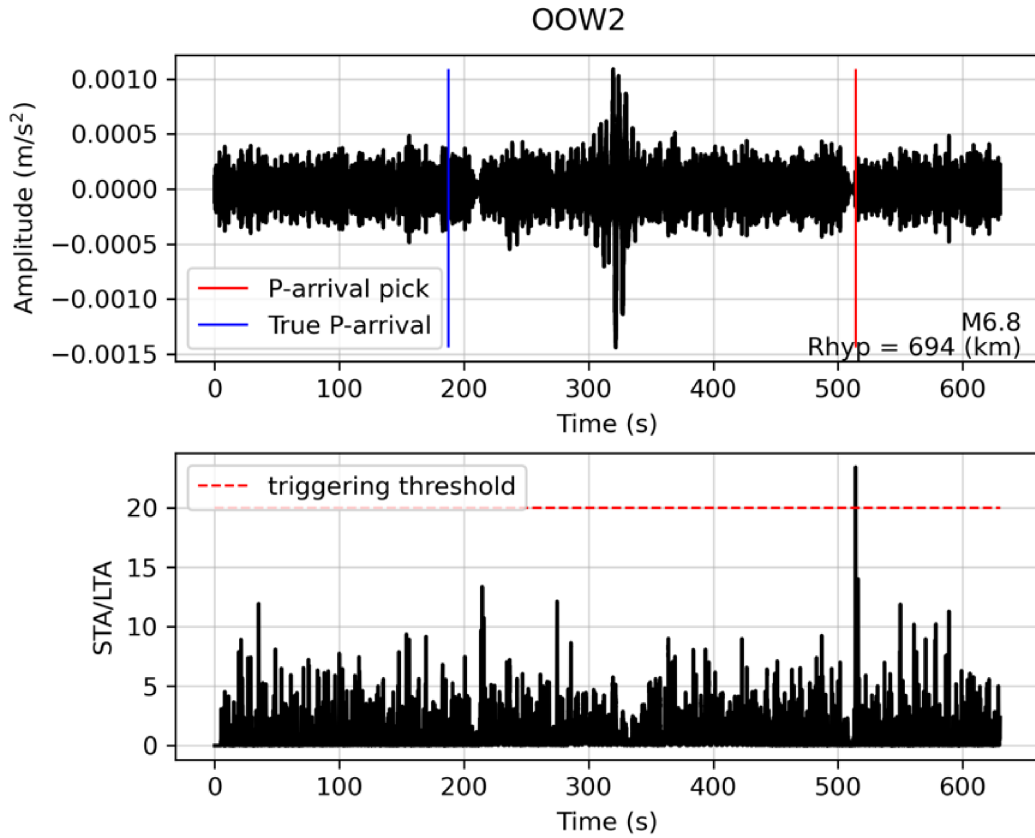


Figure 3.S3. Example late trigger of the STA/LTA algorithm using EPIC parameter configuration for scenario event 2 at station OOW2. The algorithm was triggered by a noise feature rather than the earthquake signal. (top) Time-series with the true P-wave arrival and STA/LTA pick times indicated. (bottom) STA/LTA function with the triggering threshold indicated.

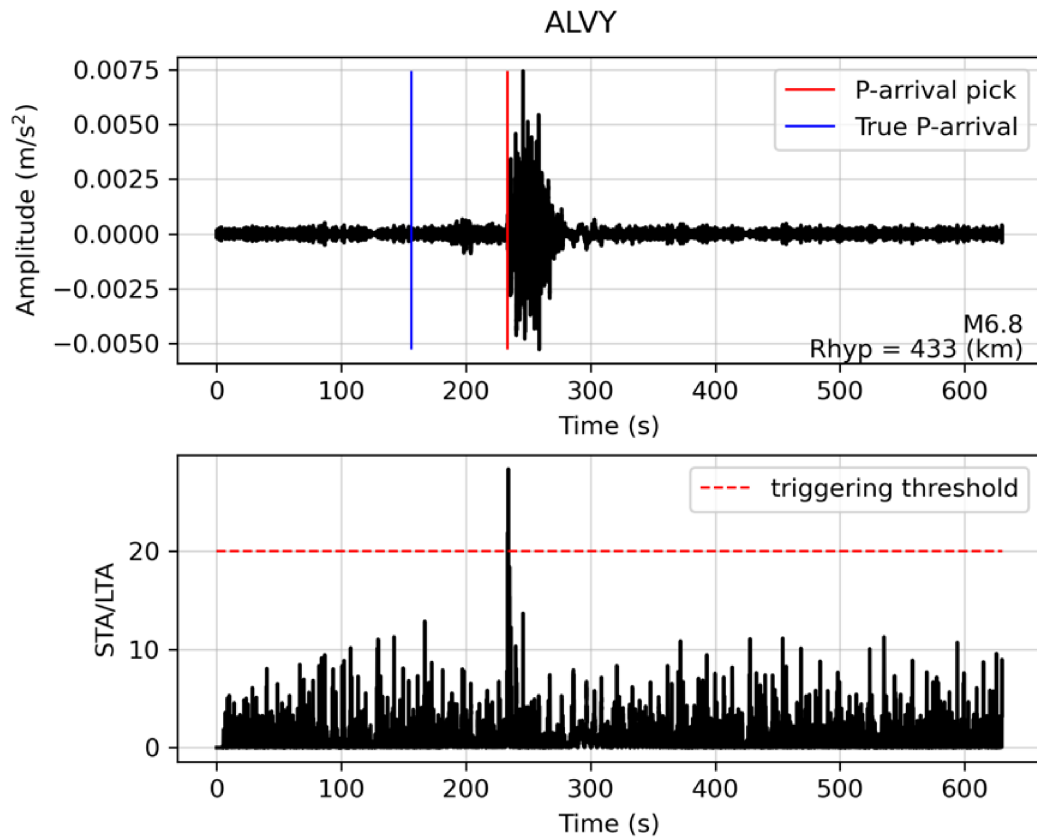


Figure 3.S4. Example trigger of the STA/LTA algorithm on the S-wave using EPIC parameter configuration for scenario event 1 at station ALVY. (top) Time-series with the true P-wave arrival and STA/LTA pick times indicated. (bottom) STA/LTA function with the triggering threshold indicated.

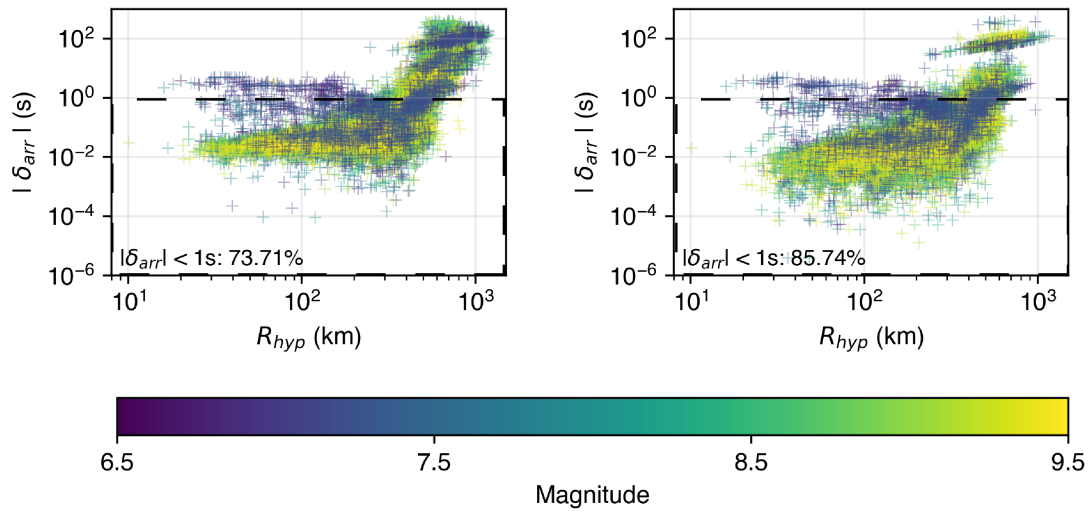


Figure 3.S5. The absolute values of the arrival time residuals in Figure 3.3c–d plotted in log–log space to highlight the smaller residuals. (left) 73.71% of the successful event detections using (left) ONC-EW STA/LTA configuration and 85.74% using (right) EPIC STA/LTA configuration had a δ_{arr} within 1 s of the true arrival, which is highlighted by the dashed boxes.

3.12.4 Model Attenuation

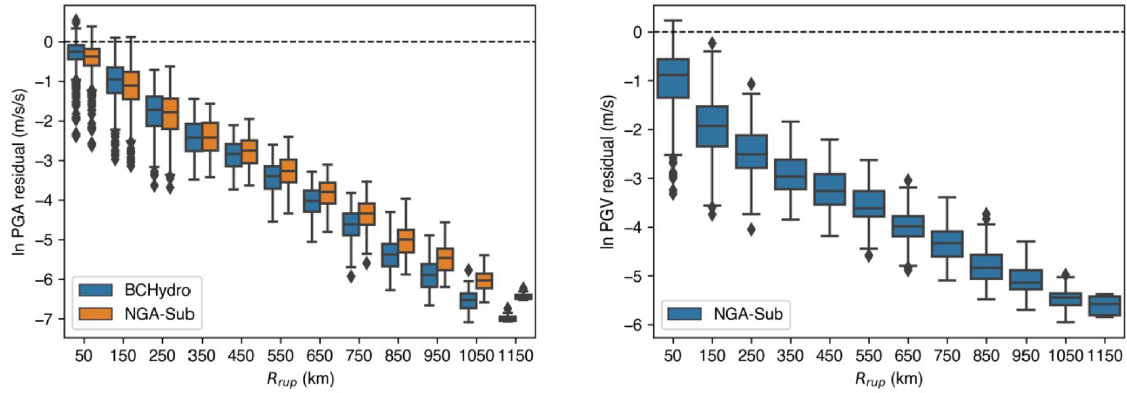


Figure 3.S6. (left) PGA and (right) PGV residuals (GMM prediction – simulation) as a function of closest rupture distance (R_{rup}). The color boxes indicate the ground-motion model used (BCHydro: Abrahamson et al., 2016, 2018; NGA-Sub: Parker et al., 2020). These simulations were generated using original attenuation parameters but have since been modified to reduce the linear attenuation bias.

CHAPTER IV

ESTIMATES OF κ_0 AND EFFECTS ON GROUND MOTIONS IN THE SAN FRANCISCO BAY AREA

From Nye, T. A., Sahakian, V. J., King, E., Baltay, A., & Klimasewski, A. (2022). Estimates of κ_0 and effects on ground motions in the San Francisco Bay Area. *Bulletin of the Seismological Society of America*, 113(2), 823–842. doi: 10.1785/0120220046. This chapter is a reprint of the material as it appears in *Bulletin of the Seismological Society of America*, with only minor grammatical and editorial modifications.

4.1 Introduction

Kappa (κ) is an observational spectral parameter that describes the high-frequency decay of S-wave acceleration spectra (Anderson and Hough, 1984). Considered to be controlled mainly by site characteristics, κ can be a valuable parameter to characterize high-frequency site conditions, which are critical to earthquake engineering and seismic hazards studies. In the past decade, some empirical ground-motion models (GMMs) have incorporated κ as a site parameter (Hassani and Atkinson, 2018; Laurendeau et al., 2013). Synthetic ground-motion models, a necessary supplement for regions lacking frequent events or seismic instrumentation, have applied κ as a parameter for high-frequency attenuation with various stochastic methods (e.g. Boore, 2003; Graves and Pitarka, 2004; Mai et al., 2010; Motazedian and Atkinson, 2005; Somerville et al., 2009). In addition, κ is often used for host-to-site adjustments in empirical ground-motion models which adjust

modeled ground motion using an average regional κ value for the study region (Biro and Renault, 2012; Douglas, 2006; Van Houtte et al., 2011). Although such studies have not been performed yet, κ could also be used to help constrain regional velocity models, as greater-attenuating media are likely correlated with weaker, low-velocity zones (Dalton et al., 2009; Gordon and Davis, 1968; Hough et al., 1999).

κ_0 can be empirically determined but does require many earthquakes recorded on each seismic station of interest. The San Francisco Bay area (SFBA) is a seismically active region with a dense seismic network (Fig. 4.1); however, no such κ_0 models exist here. The main objectives for this study are to (1) model κ_0 in the SFBA using a dense network of seismic stations and (2) evaluate how well the model represents ground motions and site effects in this region. We also aim to relate these modeled parameters with known geologic features to further constrain the subsurface geology of the SFBA.

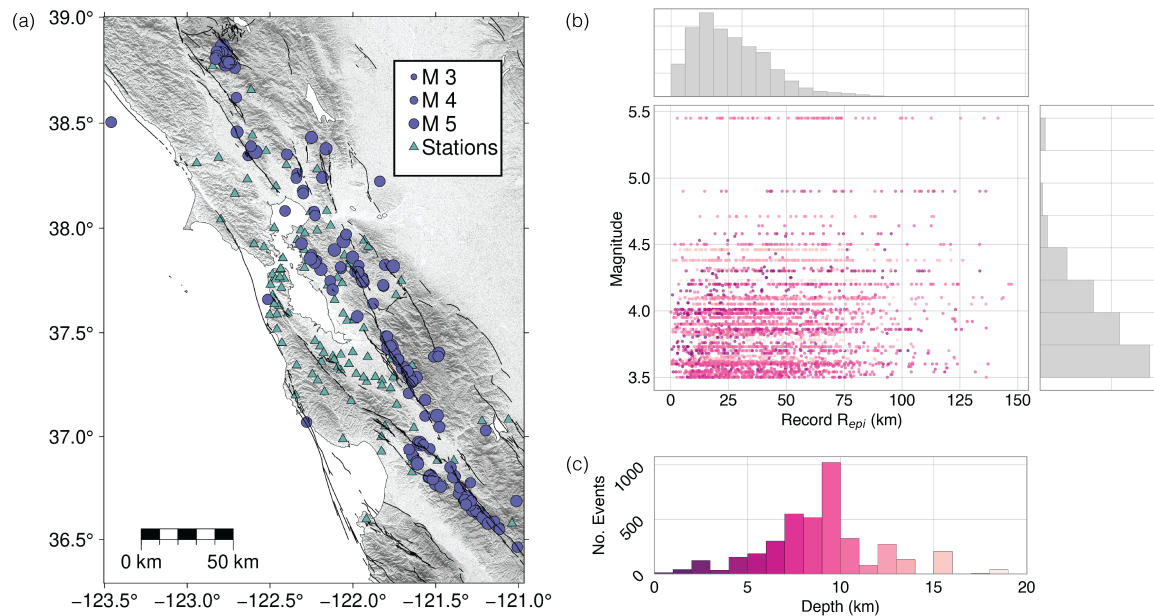


Figure 4.1. (a) Study region of the San Francisco Bay area (SFBA) with events (circles) and seismic stations (triangles) used to estimate κ_0 . Quaternary faults are represented by solid black lines. (b) Magnitude and epicentral distance (R_{epi}) distribution of records used in this study colored by depth. (c) Depth distribution of records used.

Lastly, because κ_0 is estimated from seismic records, its spatial measurement is

biased by the location of seismic stations. Subsequently, its utility for region-specific or spatially-varying (i.e., non-ergodic) GMMs (e.g. Landwehr et al., 2016) is limited. Producing a regional map of κ_0 will not only support the development of non-ergodic GMMs, but will also be useful for ground-motion seismology studies in regions lacking seismic instrumentation. To our knowledge, only two continuous maps of κ_0 have been developed at the time of this publication: New Zealand (Van Houtte et al., 2018) and Europe (Pilz et al., 2019). However, both maps are intended to represent hard rock conditions only. Our final product is a spatially-continuous map of κ_0 in the SFBA, representing a spectrum of site conditions (Nye et al., 2022b), which will be a valuable contribution to the U.S. Geological Survey (USGS) Bay area velocity model shallow layer (Aagaard and Hirakawa, 2021; Brocher et al., 2006; Jachens et al., 2006; Rodgers et al., 2020).

4.2 Background

The theoretical ω^2 earthquake source model assumes a flat spectral acceleration response at frequencies higher than the corner frequency (Brune, 1970). However, empirical observations demonstrate additional attenuation in the high-frequency range above the transitional frequency, f_{max} , typically above the corner frequency, f_c (Hanks, 1982). Assuming a flat site spectrum (i.e., no amplification), Anderson and Hough (1984) first defined κ as a way to quantify this deviation of observed spectra from the theoretical ω^2 source model by fitting the decay of the high-frequency portion of acceleration spectra,

$$A(f) = A_0 \exp(-\pi \kappa f) \quad (4.1)$$

where f denotes the frequency, A_0 is an amplitude parameter dependent on the source and path, and κ describes the slope of the decay in linear–log space.

As κ is estimated from seismic recordings, it describes an observation rather than a driving mechanism. Although κ can be related to physical parameters, there has been much debate on its origins, especially its association with source, path, and site effects. Many studies have examined the path-dependency of κ (e.g. Anderson, 1991; Anderson and Hough, 1984; Castro and Ávila-Barrientos, 2015; Castro et al., 2022; Fernández et

al., 2010; Kilb et al., 2012; Ktenidou et al., 2013). There has also been discussion over whether κ is a product of the source or site, as κ signifies a relative depletion of high frequencies, which could be due to attenuation or a source deficiency. A handful of studies address possible source mechanisms, such as non-elasticity of the fault (Papageorgiou and Aki, 1983), or a tectonic region dependency (Ktenidou et al., 2013). However, κ has been shown to have no magnitude-dependence (Anderson, 1986; Castro and Ávila-Barrientos, 2015; Fernández et al., 2010; Haendel et al., 2020; Ktenidou et al., 2013). Ultimately, it is largely accepted that κ is influenced mainly by near-site geologic conditions, and to a lesser degree by the path and source. κ typically increases with distance because it is influenced by the overall path attenuation (i.e., Q -profile), and scatter in κ values at a site from nearby events is likely due to varying source properties (Kilb et al., 2012).

Of particular interest is κ 's ability to predict ground-motion site characteristics, given its local control on the high-frequency content of seismic recordings. GMMs approximate shaking intensity from earthquakes based on a set of independent parameters describing the earthquake source, attenuation, and site effects (Douglas and Edwards, 2016). Site characterization is a key component of these GMMs, as site conditions are more easily measurable and can have a large effect on the local ground shaking. V_{S30} , the time-averaged shear wave velocity in the upper 30 m of crust, is a common site parameter in GMMs (e.g. Bayless and Abrahamson, 2018; Boore et al., 1993; Borchardt, 1994; Campbell and Bozorgnia, 2008) because of its large-scale proxy-based availability and its representation of local amplification. It is, however, a poorly constrained parameter as direct measurements of V_{S30} are less abundant, and thus V_{S30} is often inferred from local geology. Estimated V_{S30} data can vary up to 50% from true V_{S30} , which can have a significant effect on ground-motion studies (Ktenidou et al., 2014). κ , however, has shown to be a similar, or better predictor of certain high-frequency ground-motions due to site effects and should also be considered in GMMs (Klimasewski et al., 2019, 2021; Silva and Darragh, 1995;). Relationships between κ and V_{S30} are, however, desirable for instances where measured κ values may not exist but are still required for GMMs.

In ergodic GMMs, average regional conditions are assumed to be representative of all local conditions (Douglas and Edwards, 2016). While reasonable to a degree, there can still be vast variability over a small scale. Moreover, although modern ergodic GMMs are developed with site terms, application of these models requires site-specific parameters, which can be limited to locations of seismic instruments. To apply these models to any location, partially or fully non-ergodic models are needed. The use of non-ergodic, fully spatially-varying models provides more accurate predictions and could reduce aleatory variability (Al Atik et al., 2014; Douglas and Edwards, 2016; Stewart et al., 2019), even up to $\sim 40\%$ (Landwehr et al., 2016). However, such models require the availability of spatially-varying parameters, which motivates this study’s objective to produce a regional map of κ_0 .

4.3 Data and Methods

4.3.1 Dataset

To compute κ_0 values in the SFBA, we use a dataset of approximately 11,000 three-component earthquake acceleration recordings from 227 events in the SFBA region that occurred between 2000–2021 (Fig. 4.1a). Because this method requires a substantial frequency range above the corner frequency with which to solve for κ , we limit our dataset to include only earthquakes with a moment magnitude (\mathbf{M}) ≥ 3.5 . We also limit our dataset to records recorded at distances (R_{epi} , epicentral distance) ≤ 150 km to reduce bias from longer travel time in the subsurface and varying Q -structure at greater depths (Ji et al., 2022). Most of the events in our dataset are $3.5 < \mathbf{M} < 4$, with the largest being $\mathbf{M}5.5$ (Fig. 4.1b). The majority of earthquakes were reported in moment magnitude, with the exceptions of 12 events reported in local magnitude (M_L) and 2 in duration magnitude (M_d); the local magnitudes are assumed to be the same as moment magnitude in this range (Ross et al., 2016). The earthquakes were recorded on 228 stations— 15 broadband velocity stations from the Berkeley Digital Seismic (BK) network and 213 accelerometers from the BK, Berkeley Geysers (BG), USGS Northern California Regional (NC), National Strong

Motion Project (NP), and California Department of Water (WR) networks.

All data were high-pass filtered at 0.28 Hz to filter out a microseism, demeaned, baseline-corrected, and corrected for instrument response with output units in acceleration (m/s^2). To maintain sufficient quality data for our inversions, all recordings have a minimum signal-to-noise ratio (SNR) of 3 (Ktenidou et al., 2013) across the high-frequency range of 10–30 Hz, across the whole spectrum, and in the time domain. We not only retain records with a $\text{SNR} \geq 3$ for all scenarios to avoid high-frequency noise that could contaminate κ estimates, but we also use the whole spectrum in horizontal-to-vertical spectral ratio (HVSr) calculations. We also want to eliminate records that may have a significant SNR in the frequency domain but are noisy in the time domain.

Frequency domain SNR is calculated as the mean of the signal spectrum divided by the mean of the noise spectrum. We define the noise as the first 6–8 seconds of useable record (i.e., unaffected by the cosine taper in removal of instrument response), depending on how long the pre-P-arrival record is. If less than 6 seconds are available, the last 8 seconds of the coda are used (Ktenidou et al., 2016). We define the signal as an equivalent number of seconds following the theoretical S-wave arrival (assuming 3.5 km/s average S-wave velocity). Time domain SNR is calculated as the ratio of the standard deviation of signal to the standard deviation of noise, using the same time windows as previously described. The size of our dataset does not contain all available recordings from the SFBA that meet these criteria as we have culled our dataset to only use recordings from stations that have at least 8 useable records to solve for κ to improve the robustness of our estimates. This is further explained in the *Estimating Site Kappa* section.

κ is calculated from the S-wave contribution to acceleration spectra (Anderson and Hough, 1984). However, several studies show the signal window has little effect on the results of κ , provided that the energetic part of the S-wave is preserved (Anderson and Hough, 1984; Van Houtte et al., 2011). To ensure this, our seismograms are trimmed to 2 seconds before the theoretical S-wave arrival and are cut at variable times after, depending on the energy content of the records. The upper time limit is the maximum time at which

80% of the Arias intensity (representative of total energy) is met for the three components of each record (Fernández et al., 2010; Van Houtte et al., 2011). We obtain the Fourier amplitude spectra (FAS) for each horizontal component using *mtspec* (Prieto et al., 2009; Krischer, 2016), which also returns the jackknife 5% and 95% confidence intervals. We then obtain the orientation-independent vector sum (Eq. 4.2) of each recording’s horizontal components (Ktenidou et al., 2016), and the error from *mtspec* is propagated into the inversion for κ on the averaged spectra. The equation for the vector sum is provided below, where $A_{VS}(f)$ denotes the vector sum amplitude as a function of frequency (f), and H_1 and H_2 are the horizontal components.

$$A_{VS}(f) = \sqrt{H_1(f)^2 + H_2(f)^2} \quad (4.2)$$

4.3.2 Estimating Site Kappa

Individual Record Kappa: Various methods have been used to measure κ , often taking a record-by-record approach, such as the widely used method of performing a linear least-squares fit to the high-frequency decay of S-wave acceleration spectra (e.g. Anderson and Hough, 1984; Askan et al., 2014; Castro and Ávila-Barrientos, 2014; Castro et al., 2022; Douglas et al., 2010; Fernández et al., 2010; Ji et al., 2021, 2022; Ktenidou et al., 2013, 2015; Ktenidou and Abrahamson, 2016; Palmer and Atkinson, 2020; Purvance and Anderson, 2003; Van Houtte et al., 2011, 2014, 2018; Xu et al., 2020). Other record-by-record methods include solving for M_0 , f_0 , and κ by minimizing the root-mean-square misfit between the observed and theoretical spectra (Anderson and Humphrey, 1991; Humphrey and Anderson, 1992), solving for κ on the low-frequency slope of small magnitude ($M_L < 1$) displacement spectra (Biasi and Smith, 2001; Kilb et al., 2012), and fixing the stress drop and using M_0 and f_0 to solve for κ (Kilb et al., 2012). Spectral decomposition, a method in which a large suite of records is used to deconvolve source and site spectra, has also been used to solve for κ on the site spectra (Klimasewski et al., 2019). Note, this is not the full list of methods. For a more complete summary, see Ktenidou et al. (2014).

We follow the traditional method of Anderson and Hough (1984) to perform a least-

squares fit on a high-frequency range ($f_e < f < f_x$) of the acceleration spectra in linear–log space, where f_e is some frequency well-above the corner frequency, and f_x is the noise floor of the spectrum. Anderson and Hough (1984) mainly applied this method to a dataset with events **M5+**. There has been some debate around applying this method to smaller magnitudes, as these events typically have a higher corner frequency, resulting in possible source effect contamination. However, Kilb et al. (2012) suggest this is mainly an issue for smaller events ($M < 3.5$) that exist below our magnitude range, and several studies have successfully applied this method with events in our magnitude range (e.g. Ktenidou et al., 2013, 2015; Van Houtte et al., 2011). The events in our study have an approximate expected corner frequency of 0.5–5 Hz, which is well below the high-frequency range typically used to solve for κ . We thus assume this method to be suitable for our dataset and have carefully chosen frequency ranges that are sufficiently above the earthquake corner frequencies.

Frequency Range Selection: Most κ studies have manually picked the frequency ranges for each record. However, considering the large number of records used in this study (over 10,000), we have developed an algorithm to select the lower and upper frequency bounds on each record that is two-stage. (1) Prior to selecting f_e and f_x on each record, we determine maximum limits for the frequency bounds that are influenced by the (a) corner frequency, (b) instrument response, and (c) possible site amplification. (2) The frequency bounds, f_e and f_x , are selected based on characteristics of the first derivative of a degree 15 polynomial fit to each record spectrum (Fig. 4.2).

A key assumption in solving for κ_0 is that the high-frequency decay is isolated from source effects (Anderson and Hough, 1984). In step 1a of the algorithm, we account for this by imposing a lower limit of 1.5 times the theoretical corner frequency (f_c) (Ktenidou et al., 2016), which is derived using the following equation from Hanks and Thatcher (1972):

$$f_c = \beta \left(\frac{\Delta\sigma}{8.47M_0} \right)^{1/3} \quad (4.3)$$

The seismic moment (M_0) is obtained using the moment magnitude of each event and the Hanks and Kanamori (1979) relation. We assume an average regional S-wave velocity (β) of 3.5 km/s and an average stress drop (κ_0) of 5 MPa (Atkinson and Beresnev, 1997;

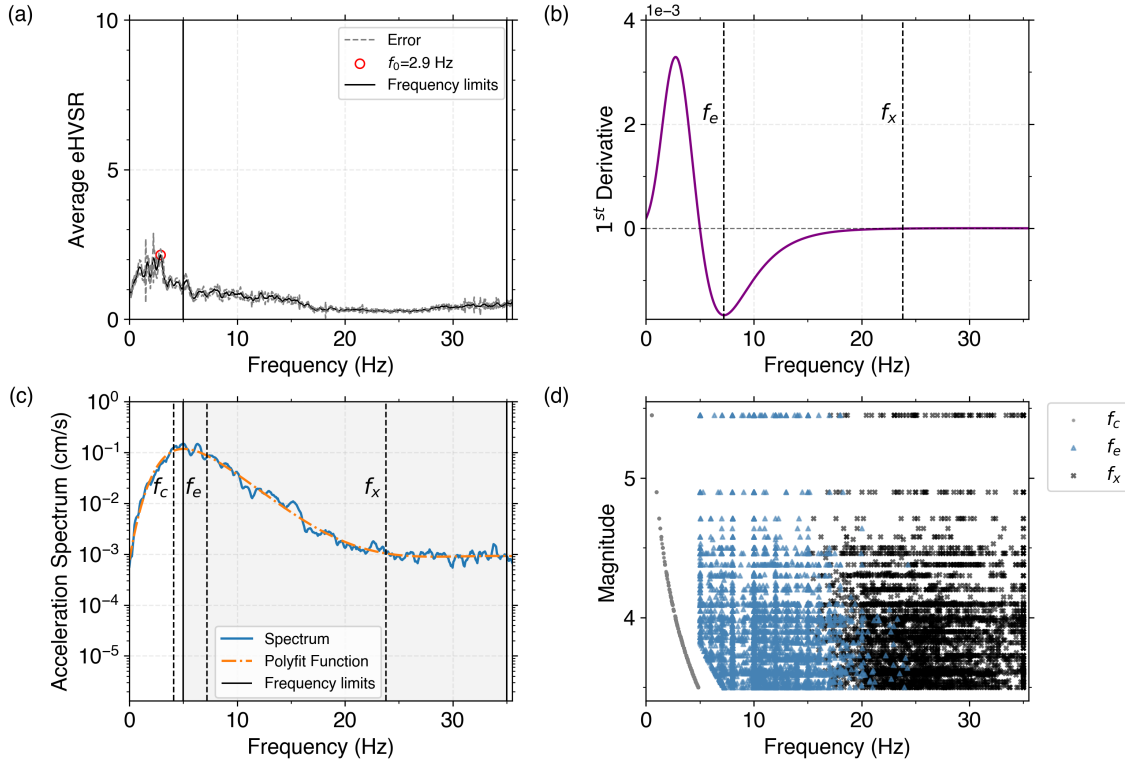


Figure 4.2. Example application of the algorithm to select the frequency ranges, f_e and f_x , used to solve for κ on an individual record. (a) Average eHVS for station J030, with the fundamental frequency (f_0) indicated by a red circle and the manually-selected frequency limits indicated solid black lines. (b) First derivative of the polyfit function in (c). f_e is the frequency at which the first derivative is lowest, and f_x is the frequency at which the first derivative reaches a threshold value of 8×10^{-5} . (c) Example record spectrum for station J030 with a degree 15 polyfit function fit to the spectrum. f_c indicates the theoretical corner frequency of this event, and f_e and f_x are the frequency bounds determined in (b). The shaded region indicates the maximum frequency limits bounded by the eHVS frequencies (solid black lines) from (a). (d) Distribution of f_c , f_e , and f_x for all records in the dataset as a function of earthquake magnitude.

Baltay and Hanks, 2014). To address the instrument response in step 1b of our algorithm, an upper limit of 80% of the Nyquist frequency is set for lower sampling rate (e.g. 50, 80 Hz) stations, but we impose a 35 Hz limit for higher sampling rate stations, as this is the smallest maximum value of flat instrument response for such stations (Van Houtte et al., 2014).

The presence of site amplification can bias κ estimates, especially if the

fundamental frequency falls within the frequency range of interest (Ktenidou et al., 2016; Parolai et al., 2004). As the SFBA is covered by considerable weak material in the shallow subsurface, this is a possible concern for our dataset. In step 1c of our algorithm, we minimize the effects of site amplification in solving for κ by computing the average earthquake-based HVSR (eHVSR) at each site (Lermo and Chavez-Garcia, 1993; Ktenidou et al., 2016; Ito et al., 2020). For each site eHVSR curve, we manually select frequency ranges that bound a generally-flat portion of the curve, regardless of the true amplitude of the eHVSR (Ktenidou et al., 2016). These frequency ranges serve as maximum limits to f_e and f_x (see Fig. 4.S1 for example eHVSR curves or the online Zenodo dataset, Nye et al., 2022b, for the full set of stations). We compute site eHVSR using three-component S-wave records by first smoothing the waveforms with a Hanning filter (Çelebi et al., 2018) and then calculating the spectra using *mtspec*— a Python wrapper for the Multitaper Spectrum Estimation Library (Prieto et al., 2009). The horizontal spectral components for each recording are geometrically-averaged and divided by the vertical component, resulting in an eHVSR for each recording. The eHVSR records for each site are averaged in logarithmic space to obtain an average eHVSR site spectrum (Ito et al., 2020). We recognize that the eHVSR is not a true site amplification transfer function, but rather a simplistic approach that provides some insight into the amplifying conditions and helps further constrain the part of the spectrum dominated by site attenuation.

After determining the maximum frequency limits, f_e and f_x are selected in step 2 of our algorithm based on the first derivative of a degree 15 polynomial fit to each spectrum. The lower bound f_e is the frequency of the minimum of the first derivative, as this indicates the steepest negative slope of the spectrum. The upper bound f_x is the frequency at which the first derivative reaches an arbitrary threshold of -8×10^{-5} (Fig. 4.2b), assuming data units of cm/s^2 . This indicates the relative frequency at which most of the spectral records in our dataset start to plateau. We retain only records with a $\Delta f \geq 10$ Hz (Van Houtte et al., 2014; Ktenidou et al., 2016) and stations that retain a minimum of 8 records.

Using our above-described algorithm, we find robust estimation of f_e and f_x

in solving for κ on the individual records. Earthquake spectra are complex, and the mathematical forms used to describe them (such as the first derivative) do not always match the ideal functions. Thus, the algorithm sometimes overestimates f_x for certain records. However, we still find the linear fit to model the shape of the decay reasonably well. The use of a weighted regression to solve for κ_0 , explained in the next section, also reduces bias from outlier record κ 's resulting from a poor frequency determination.

Kappa Regression: In the model for high-frequency decay (Eq. 4.1), κ represents total high-frequency attenuation, which means it is characterized by local site attenuation, as well as attenuation influenced by the wavefield paths. Anderson (1991) formed a model to represent both contributions to κ :

$$\kappa(R, S) = \kappa_0(S) + \tilde{\kappa}(R), \quad (4.4)$$

in which κ_0 describes attenuation from the shallow subsurface at the site, and $\tilde{\kappa}(R)$ describes attenuation from the path and is a function of the rupture distance and the overall path attenuation, modeled as quality factor Q . To isolate κ_0 , we plot the individual records' κ as a function of epicentral distance (Ktenidou et al., 2013) for each station and perform a weighted regression which considers the uncertainty of the κ estimates (Fig. 4.3b). This allows us to extrapolate κ at $R_{epi} = 0$ km, which we accept as κ_0 . Note that this method of using Anderson and Hough (1984) to estimate κ for each record at a station and then extrapolating for κ_0 is referred to as the $\kappa_{0,AS}$ (acceleration spectrum) method in Ktenidou et al. (2014). We perform this regression and obtain a κ_0 estimate for each site (Fig. 4.4a). The uncertainty in κ_0 estimates is obtained from each regression covariance matrix, which is weighted by the individual record κ 's and is a propagation of the individual spectra uncertainties.

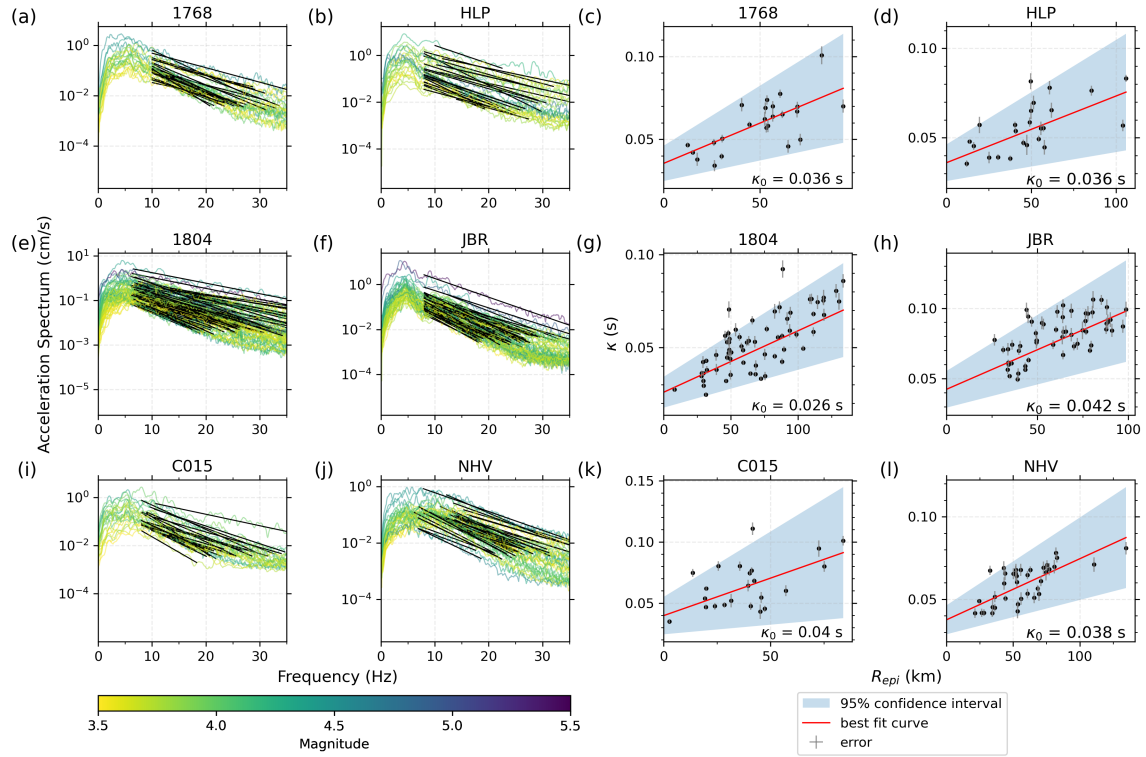


Figure 4.3. Example station spectra and κ estimates. (a–b, e–f, i–j) Acceleration spectra recorded at stations 1786, HLP, 1804, JBR, C015, and NHV colored by magnitude. Solid black lines represent the linear fit on each record using the selected frequency ranges. Amplitudes plateau around ~ 20 – 30 Hz at station C015 and JBR, demonstrating the effectiveness of our algorithm to avoid such frequency ranges when solving for κ . (c–d, g–h, k–l) Regression of record κ as a function of epicentral distance (R_{epi}) for stations 1786, HLP, 1804, JBR, C015, and NHV. Error bars are indicated by light gray lines, and the 95% confidence interval is represented by the shaded region. κ_0 at 1768, HLP, 1804, JBR, C015, and NHV are 0.036, 0.036, 0.026, 0.042, 0.040, and 0.038 s, respectively.

4.4 Results

4.4.1 Site Kappa

Site Kappa Estimates: We find κ_0 to vary substantially in the SFBA with reliable estimates for $\sim 89\%$ of the stations in our dataset. However, 24 of the stations presented unreliable estimates of κ_0 , either due to a negative correlation of κ with distance, a negative κ_0 , or obvious site response in the frequency of interest. A negative correlation contradicts the theoretical understanding that κ (i.e., attenuation) increases with ray path distance

traveled. The negative correlation may be due to a poor polynomial fit, and hence poor frequency range selection, or possible bias from unaccounted for site amplification. A negative κ_0 was observed at one station, J003, but this is also physically unrealistic as it suggests negative attenuation. We have removed the stations from further analysis, and the following results and figures do not include these stations. However, figures with spectra and their linear fits, as well as $R_{epi}-\kappa$ regression plots, can be found for all stations online in the Zenodo dataset (Nye et al., 2022b). For a full list of the stations removed, see Table 4.S1.

For the remaining 204 stations, κ_0 estimates range over an order of magnitude from 0.003–0.072 s (Fig. 4.4a), and standard deviations range from 0.002–0.028 s (Fig. 4.4b). Although the spread in values is large, the inter-quartile range of κ_0 falls between 0.028–0.045 s (Fig. 4.4a). κ_0 exhibits some spatial dependence in the SFBA which was not imposed a priori and is likely a result of the heterogenous geology of the Bay Area. With this spatial trend, we identify four characteristic zones in the region.

The first (zone 1) is a pocket of larger κ_0 estimates north of the bay around the Santa Rosa Plain (SRP) and Rodgers Creek Fault (RCF), and extending west to the San Andreas Fault (SAF). κ_0 in this zone is moderately high with 80% of the values falling within 0.034–0.046 s. The three stations in zone 1 with the largest κ_0 are MCCM, NSP, and 1835 (0.072, 0.049, 0.049 s, respectively). The generally larger κ_0 in this region likely correlates with the weak, lower velocity material in the plain and basin areas (Eberhart-Phillips, 2016; Sweetkind et al., 2010), as well as possible deformation associated with the San Andreas Fault (SAF).

The second (zone 2) includes the East Bay, and the Sacramento–San Joaquin Delta (SSD). This zone also contains moderately large κ_0 values, with 80% of the values falling within 0.034–0.053 s. The three largest estimates in zone 2 are observed at stations 1843, CUSLD, and CBP (all with an estimate of 0.060 s). The larger κ_0 values imply greater attenuation at the stations, which may be consistent with softer material at these sites, including east bay mud and SSD sediments (Eberhart-Phillips, 2014; Erdem et al., 2019;

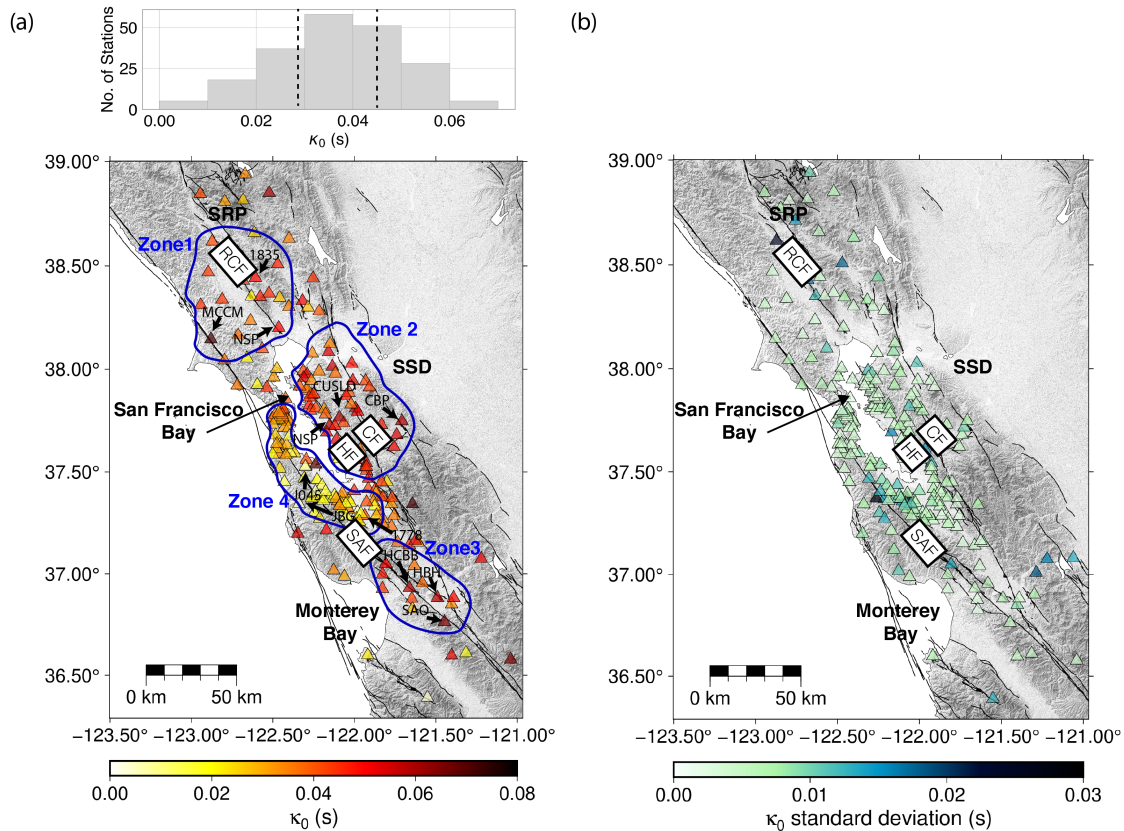


Figure 4.4. (a) Station-based κ_0 estimates and (b) corresponding standard deviations in our study region. κ_0 estimates range from 0.003 to 0.072 s but most fall between 0.028 to 0.045 s, as indicated by the histogram in (a), where the dashed black lines indicate the inter-quartile range. Four key zones are highlighted: zone 1: North Bay, zone 2: East Bay, zone 3: South Bay, and zone 4: West Bay. The Sacramento–San Joaquin Delta (SSD) and Santa Rosa Plain (SRP) are labeled, as well as the San Andreas Fault (SAF), Hayward Fault (HF), Calaveras Fault (CF) and Rodgers Creek Fault (RCF). Other Quaternary faults are also represented by solid black lines. Several key stations mentioned in the text are indicated by the arrows. Zone 1 has κ_0 estimates ranging from 0.021 to 0.072 s, zone 2 has κ_0 estimates ranging from 0.016 to 0.060 s, zone 3 has κ_0 estimates ranging from 0.028 to 0.065 s, and zone 4 has the smallest estimates ranging from 0.004 to 0.069 s.

Hough, 1990). The abundance of fault zones in the east bay region likely also increases attenuation, as higher fracture density is often associated with lower Q (Eberhart-Phillips, 2014, 2016).

The third (zone 3) includes the region around the portion of the SAF extending farther to the south of the bay, near Loma Prieta. Zone 3 includes the largest percentage of

higher κ_0 estimates, with 80% of values falling within 0.037–0.052 s. The three largest estimates in this zone are observed at SAO, HBH, and HCBB (0.065, 0.059, 0.057 s, respectively). This region contains an abundance of quaternary sediments and weaker rock (Lees and Lindley, 1994), as well as multiple fractured fault zones (the Zayante-Vergeles, San Andreas, and Sargent fault zones).

The fourth and final zone of interest (zone 4) surrounds the SAF along the western side of the bay. The κ_0 estimates at stations near and along this fault are lower than all surrounding regions, with 80% falling within 0.020–0.034 s. The three smallest estimates observed at stations J045, 1778, and JBG (0.004, 0.005, and 0.005 s, respectively), implying significantly less attenuation. This may be attributed to generally higher-velocity material and harder rock sites (Eberhart-Phillips, 2016).

4.4.2 Producing a Regional Kappa Map

Often, hazards studies require evaluating expected ground motion in regions lacking seismic instrumentation. The hazards community is also transitioning towards non-ergodic GMMs, which require the availability of spatially-varying parameters, including a continuous map of κ_0 . We use the method of ordinary kriging (OK) to produce a spatially-continuous map of κ_0 for the SFBA (Van Houtte et al., 2018) (Fig. 4.5a). Kriging is a spatial-interpolation method which seeks to minimize the prediction error (in our case the prediction of κ_0), and OK assumes a constant mean and variance across the region. Kriging is preferred over simpler spatial-interpolation methods, such as inverse distance weighted interpolation, linear regression, or Gaussian decays, as it assumes a spatial correlation among the data points. It also produces prediction uncertainties, which are often desirable in hazard estimates.

Semivariogram: The first step in spatial interpolation via kriging involves computing a semivariogram of the data. This is a visual representation of how spatially “related” or dependent the data are, and on what spatial scales they correlate (Fig. 4.5a). The x axis of a semivariogram is the distance between points, and the y axis is the semivariance. The semivariogram typically increases before eventually plateauing,

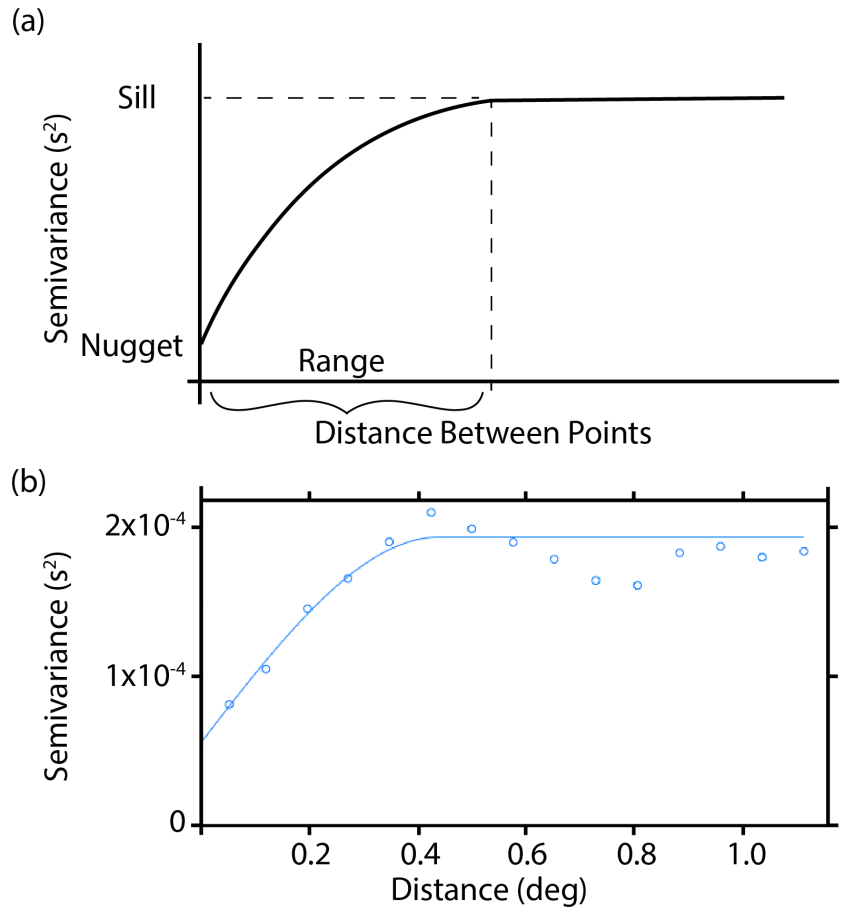


Figure 4.5. (a) Example semivariogram with the range, sill, and nugget indicated. (b) Binned empirical semivariogram computed for the κ_0 estimates assuming a spherical model. The oscillation of semivariance past the range could be due to significant variations in geology at moderate distances (such as boundaries between the four highlighted zones in the *Site-Kappa Estimates* section).

representing data that are no longer correlated. The x and y values at which the semivariogram plateaus are the range and sill, respectively, and the y -intercept is the nugget. The nugget variance (τ^2) represents the spatial variation at distances less than the inter-station distances of the dataset, or in other words, near-site randomness. The structure of the computed semivariogram is used to determine the weights of the nearby datapoints in the kriging prediction. Van Houtte et al. (2018) assume a semivariogram model with a Matérn order, but we find a spherical model better fits our data (Fig. 4.5b). We also

compute the semivariogram directly on the κ_0 estimates, whereas Van Houtte et al. (2018) assumed a lognormal distribution.

Spatial Interpolation: Using the semivariogram covariance model, we interpolate regional κ_0 from the site κ_0 estimates. The resolution of the kriged model is highly dependent on station density, thus regions with high station density can capture more fine-scale variations in κ_0 with a lower variance than regions with low station density. We cut our model to only retain regions with a variance less than the sill of the semivariogram ($1.95 \times 10^{-4} \text{ s}^2$ or 0.014 s), which represents the variance at which the stations are no longer correlated (Fig. 4.6a). We recommend using this model and refer to it as our “preferred model”; however, we include the full regional model in the supplement (see Fig. 4.S2). Both the refined model and full model κ_0 predictions range from 0.013 to 0.061 s, and we observe the same general spatial trend as we see with the initial κ_0 estimates, where κ_0 is largest to the north, east, and south of the bay, and lowest west of the bay. We observe a fine-scale resolution in the κ_0 estimates, especially near the San Francisco Bay where station density is highest. With this resolution, we observe several pockets of higher κ_0 , such as the coastline to the north of Monterey Bay, the northern part of the Calaveras Fault (CF), the northern part of the SAF, and the region where the Hayward Fault (HF) and CF join in the south. We also observe lower κ_0 zones, such as to the far north of the bay and along the peninsula of the San Francisco Bay. For a comparison of estimated versus predicted κ_0 , see Figure 4.S3.

The prediction standard deviations are obtained by taking the square root of the model variance. The standard deviations in our refined model range from 0.008 to 0.014 s (Fig. 4.6b). The standard deviation is generally lower where we have many stations and smoothly increases where station density is sparser. The full and cut .grd files for the kriged κ_0 and standard deviations are accessible on the online Zenodo dataset (Nye et al., 2022b).

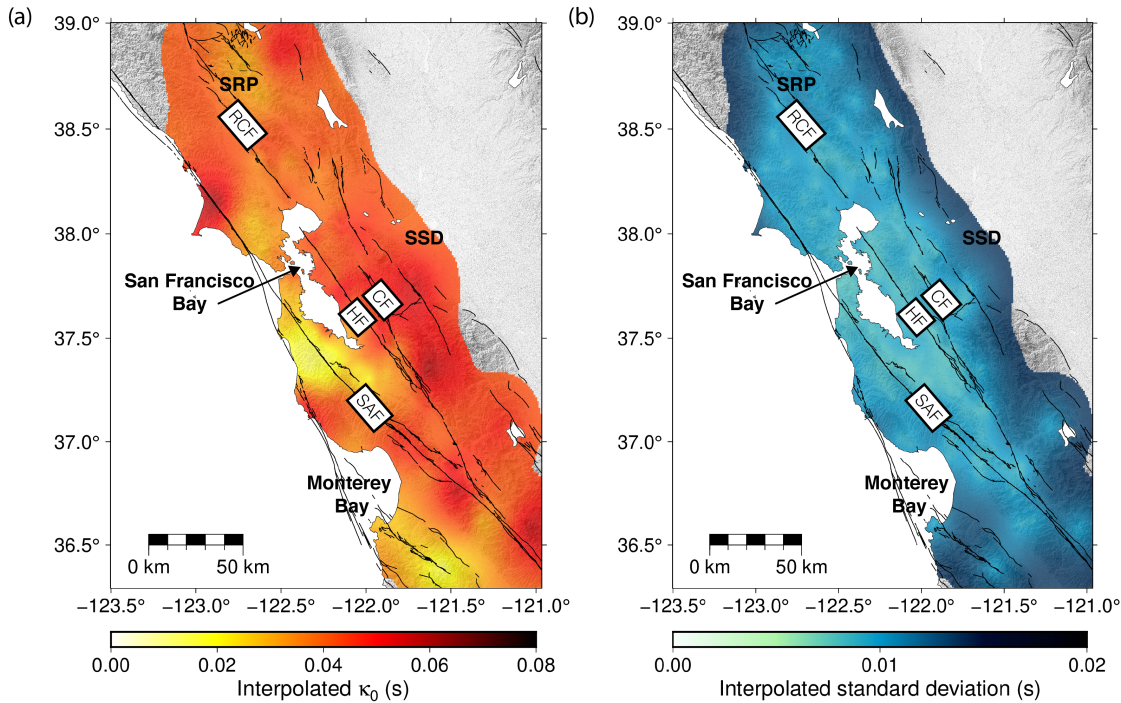


Figure 4.6. (a) Median interpolated κ_0 for the San Francisco Bay area and (b) the standard deviation for the kriged κ_0 predictions. Predictions of κ_0 range from 0.013 s west of the bay to 0.061 s east of the bay. The Sacramento–San Joaquin Delta (SSD) and Santa Rosa Plain (SRP) are labeled, as well as the San Andreas Fault (SAF), Hayward Fault (HF), Calaveras Fault (CF) and Rodgers Creek Fault (RCF). Other Quaternary faults are also represented by solid black lines.

4.4.3 Site Parameter Evaluation

One of the main motivations for understanding regional attenuating properties, such as κ_0 , is its applicability in GMMs. Most modern GMMs use V_{S30} to approximate site response. However as previously mentioned, it is not always the singular best representation for site effects. κ has been used in methods for adjusting GMMs to fit non-source regions (Atik et al., 2014; Biro and Renault, 2012; Douglas, 2006; Van Houtte et al., 2011), and it has been directly incorporated as a parameter in several modern GMMs (Hassani and Atkinson, 2018; Laurendeau et al., 2013). If shown to consistently be a more suitable site descriptor than or in addition to V_{S30} , κ_0 could be a key component in the development of future GMMs in this region.

GMM Implementation: We compare our κ_0 estimates, as well as V_{S30} , with predicted ground motion in the SFBA to evaluate and compare their effectiveness as site parameters in the region. Using the Abrahamson et al. (2014) (ASK14) and Boore et al. (2014) (BA14) GMMs via the *OpenQuake* engine (Pagani et al., 2014) and the events and sites in our dataset, we compute the predicted intensity measures (IMs) peak ground acceleration (PGA) and peak ground velocity (PGV). ASK14 and BA14 were developed using the NGA-West2 dataset (Ancheta et al., 2014) of shallow crustal events, many of which overlap with events in our dataset. We choose the ASK14 and BA14 models as they were regressed on a dataset of events with a $M > 3$, which is suitable for the magnitude range of our events. The inclusion of two GMMs was to further evaluate the relationship of these site parameters with GMMs of varying complexity.

BA14 requires as source parameters the earthquake magnitude, rupture rake, and the depth to the top of rupture (Z_{TOR}), and ASK14 also requires the rupture dip and width. Both models require the closest distance to the rupture plane (R_{rup}) as the main distance measure, as well as the Joyner-Boore distance (R_{JB}), but ASK14 also considers the R_x and R_{y0} distance measures, all of which can be obtained using R_{rup} and Z_{TOR} . (For further explanation of the distance measures, see Abrahamson et al. (2014)). Required site parameters include V_{S30} for both GMMs, and ASK14 also requires the depth to the 1 km/s S-wave isosurface ($Z_{1.0}$).

We approximate Z_{TOR} for our small, likely point-source, earthquakes as the hypocentral depth, and due to the short source-to-site distances of the records in our dataset, we approximate R_{hyp} as R_{rup} . To obtain the rupture rake and dip, we follow a hierarchy of sources. We first reference the NGA-West2 finite fault file (Ancheta et al., 2014; Pacific Earthquake Engineering Research Center, 2013). If the catalog does not contain the event, we reference a catalog obtained through *libcomcat* query (US Geological Survey, 2020a). Finally, if neither catalog contains the event, we use a default rake and dip of 0 and 90 degrees, respectively, representing a pure strike-slip fault. For the rupture width, we follow a similar hierarchy, first referencing the NGA-West2 finite fault file. If the event is not

found in the catalog, we estimate a rupture width using the Wells and Coppersmith (1984) scaling law for events $\mathbf{M} < 5$, and the Thingbaijam et al. (2017) scaling relation for events $\mathbf{M} > 5$. For V_{S30} , we use the GMM reference V_{S30} (760 m/s for both models), to correlate the site parameters with modeled ground motion without inherent amplification from incorporating site-specific V_{S30} into the models. Lastly, for $Z_{1.0}$, we estimate the depth from the USGS three-dimensional velocity model (US Geological Survey, 2020b).

Correlation with GMM Site Residuals: We calculate the RotD50 observed PGA and PGV and compare with predicted values using the ASK14 and BA14 GMMs by computing natural log residuals: $\delta IM = \ln \text{observed} - \ln \text{predicted}$. These residuals are a function of various source, path, and site effects that the GMM does not account for. Using a mixed-effects maximum-likelihood model (e.g. Parker et al., 2020; Sahakian et al., 2018), we decompose the residuals into their event and site contributions to isolate the ground-motion variability resulting solely from local site conditions:

$$\ln(\delta IM) = \delta E_i + \delta S_j + \delta W_{ij}^0, \quad (4.5)$$

where δE_i is the average event mean, δS_j is the site residual, and δW_{ij}^0 represents all other residual components (Baltay et al., 2017; Sahakian et al., 2019).

The site residual, δS_j , illustrates whether observed ground motion at a site is stronger or weaker than predicted by a GMM due to local geology. For example, a positive site residual for event i at site j would suggest the observed IM was larger than predicted due to either unaccounted for amplifying effects or less attenuating local conditions. Similarly, a negative site residual would suggest the observed IM was lower than predicted due to greater regional and local attenuation. We correlate our estimated κ_0 values and V_{S30} with the GMM site residuals to assess the robustness of these parameters to describe ground motion in the SFBA (Fig. 4.7).

For correlation with the site residuals, we use measured values where available ($V_{S30,meas}$) and proxy estimates ($V_{S30,proxy}$) elsewhere. The hierarchy for V_{S30} follows as: measurements from McPhillips et al. (2020), measurements and estimates from the NGA-West2 site database file (Ancheta et al., 2014; Pacific Earthquake Engineering Research

Center, 2013), and finally estimates from the USGS hybrid global V_{S30} database (Heath et al., 2020; Thompson et al., 2014), with the exception of station GDXB. Although a measured value exists in McPhillips et al. (2020), it is unusually high (~ 1200 m/s) and biased the regressions, so we use a proxy value from the NGA-West2 database. Measurements from McPhillips et al. (2020) were obtained from a variety of downhole and array-based methods. The NGA-West2 site database uses a combination of geology, geotechnical, geomorphology, and slope-based methods for V_{S30} estimates (Ancheta et al., 2014; Seyhan and Stewart, 2014). The USGS global database uses a slope-based approach to approximate V_{S30} but includes more detailed high-resolution regional maps where available.

We compute simple linear regression between each PGA and PGV site residuals (hereafter referred to as $\delta S_{j,PGA}$ and $\delta S_{j,PGV}$, respectively) and the site parameters κ_0 and V_{S30} (Fig. 4.7). We observe a weak negative correlation between κ_0 and $\delta S_{j,PGA}$, which indicates larger values of κ_0 typically result in weaker ground motion than predicted by the GMM (Fig. 4.7a). This relationship is consistent with our understanding that larger κ_0 describes a site which attenuates high-frequency energy more strongly than other sites. To quantify the statistical significance of the correlation, we examine the Pearson R value, the p -value, and the statistical power of the regression. The ASK14 and BA14 κ_0 - $\delta S_{j,PGA}$ regressions have a Pearson R of -0.17 and -0.18, respectively, indicating a weak correlation with reasonable scatter. However, because the p -value is low (< 0.05), there is a less than 5% chance that the correlation observed is fictional and influenced by the sample of data. This correlation also has moderate (but not statistically significant, > 0.8) statistical powers of 0.42 and 0.46, which indicate an approximate 40% chance a low p -value (< 0.05) will still be observed using a subsample of the data. With all of these factors considered, the κ_0 - $\delta S_{j,PGA}$ correlation is significant, implying κ_0 is a reasonable predictor of high-frequency ground motion in our small magnitude dataset.

We secondly examine the relationship between κ_0 and the PGV site residual ($\delta S_{j,PGV}$; Fig. 4.7b). We observe no correlation of κ_0 with $\delta S_{j,PGV}$, having a Pearson

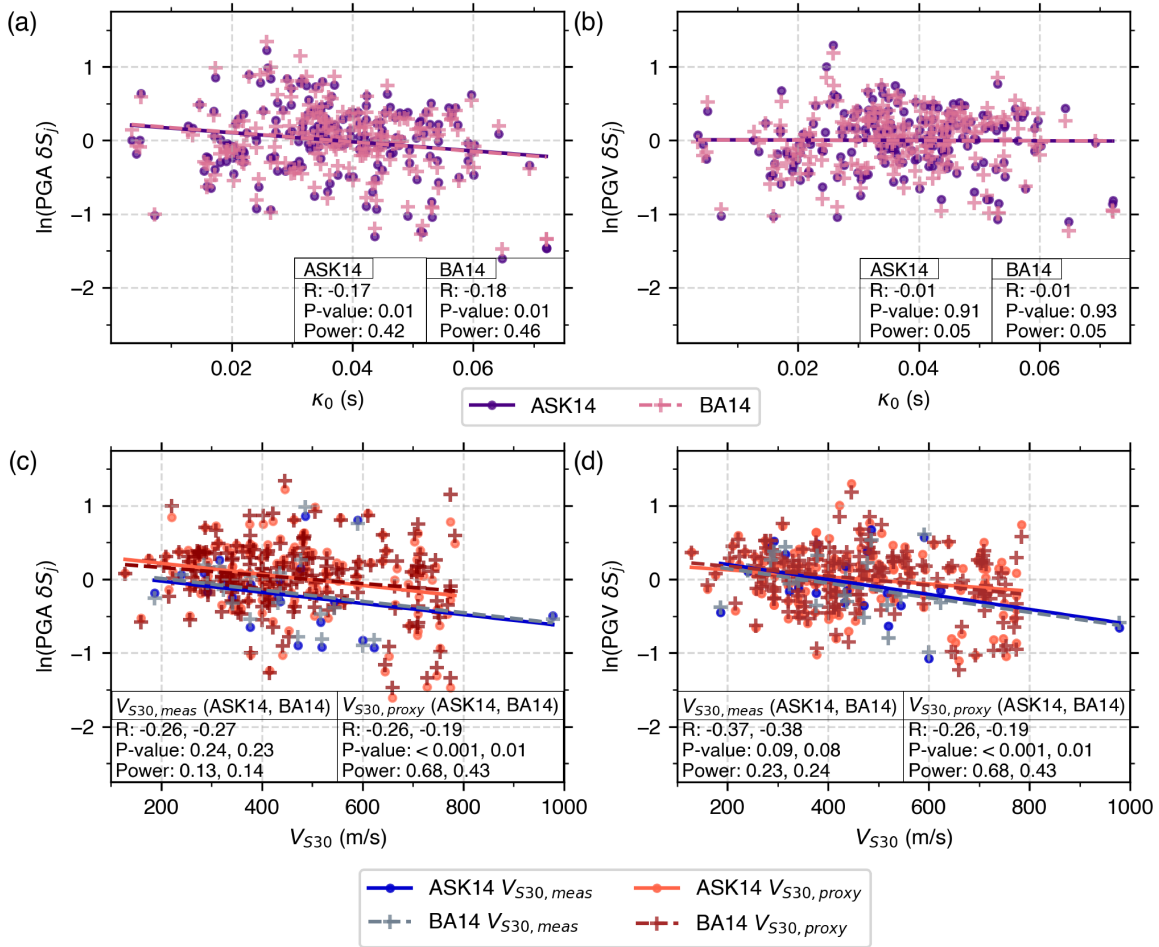


Figure 4.7. Site parameters κ_0 and V_{S30} versus GMM site residuals obtained using the ASK14 and BA14 GMMs. The statistical parameters Pearson R correlation coefficient (R), p -value, and statistical power (Power) are also indicated. (a) $\kappa_0 - \delta S_{j,PGA}$ relation shows a weak negative correlation, a statistically-significant p -value (< 0.05), and a moderate statistical power for both the ASK14 and BA14 GMMs. (b) $\kappa_0 - \delta S_{j,PGV}$ relation shows a weak, essentially non-existent correlation, a large p -value, and a low statistical power (< 0.8) for both GMMs. (c) $V_{S30,meas} - \delta S_{j,PGA}$ and $V_{S30,proxy} - \delta S_{j,PGA}$ relations show a weak negative correlation for both GMMs. However, correlation with $\delta S_{j,PGA}$ computed using ASK14 exhibits a more statistically-significant relationship compared to BA14 for $V_{S30,proxy}$. (d) $V_{S30,meas} - \delta S_{j,PGV}$ and $V_{S30,proxy} - \delta S_{j,PGV}$ relations also show a weak negative correlation. However, $V_{S30,proxy} - \delta S_{j,PGV}$ is more statistically robust than $V_{S30,meas} - \delta S_{j,PGV}$, having a low p -value and moderate statistical power for both GMMs.

R of -0.01 for both ASK14 and BA14, respectively. The p -value is higher as well (0.91 and 0.93 for ASK14 and BA14, respectively), and we observe a low statistical power (0.05 for both GMMs), suggesting essentially no relationship between κ_0 and PGV.

For V_{S30} , we compare both $V_{S30,meas}$ and $V_{S30,proxy}$ with ASK14 and BA14 PGA site residuals individually (Fig. 4.7c). As with κ_0 , we observe a weakly negative correlation with $\delta S_{j,PGA}$. This suggests regions with lower V_{S30} tend to result in higher ground accelerations. $V_{S30,meas}-\delta S_{j,PGA}$ has a Pearson R of -0.26 and -0.27 for ASK14 and BA14, respectively, whereas $V_{S30,proxy}-\delta S_{j,PGA}$ has a Pearson R of -0.26 and -0.19 . Neither the p -values nor the statistical powers for $V_{S30,meas}-\delta S_{j,PGA}$ are statistically-significant for $V_{S30,meas}-\delta S_{j,PGA}$. However, the correlation is stronger for $V_{S30,proxy}-\delta S_{j,PGA}$, especially when considering the ASK14 GMM. The p -value and statistical powers are < 0.001 and 0.68 (respectively) for ASK14, and 0.01 and 0.43 (respectively) for BA14.

We finally compare $V_{S30,meas}$ and $V_{S30,proxy}$ with PGV site residuals (Fig. 4.7d). Again, we observe a weakly negative correlation. $V_{S30,meas}$ has a stronger correlation with $\delta S_{j,PGV}$ than with $\delta S_{j,PGA}$. Here, the Pearson R is the highest for all correlations at -0.37 and -0.39 for ASK14 and BA14, respectively. The p -values are near the statistically-significant threshold at 0.09 and 0.07 for ASK14 and BA14, respectively. However, the statistical powers are still low at 0.23 and 0.24 . $V_{S30,proxy}$ exhibits the same level of correlation with $\delta S_{j,PGV}$ as it does with $\delta S_{j,PGA}$.

κ_0 - V_{S30} Relation: As determined in the previous section, κ_0 exhibits the expected correlation with observed ground motion, which supports the motivation of this study to broaden the community κ_0 dataset for seismic hazard applications. Since κ_0 is an observational parameter, seismic recordings are necessary to constrain such values, yet some regions are lacking adequate seismic instrumentation or activity. V_{S30} datasets are more abundant, so a κ_0 - V_{S30} relation would prove useful in such circumstances where κ_0 is required but unavailable. Previous studies have been varied in their conclusions of a κ_0 - V_{S30} relation (Chandler et al., 2006; Drouet et al., 2010; Edwards et al., 2011; Ktenidou et al., 2012, 2013, 2014, 2015; Ktenidou and Van Houtte, 2012; Klimasewski et al., 2019; Silva et al., 1998; Van Houtte et al., 2011). Most studies show a negative correlation between the parameters, but the regression models are variable with reasonable scatter.

The correlations seem to be influenced in part by the study region and inversion method (Ktenidou et al., 2014). One such study in Southern California (Klimasewski et al., 2019) noticed a discrepancy dependent on the method of obtaining V_{S30} , where measured V_{S30} demonstrated a negative correlation, and proxy V_{S30} demonstrated a very weak positive correlation. We investigate the nature of a relationship between κ_0 and V_{S30} for the SFBA, the only current study of this kind for the region (Fig. 4.8).

We compare κ_0 with both $V_{S30,meas}$ where available and $V_{S30,proxy}$ elsewhere (Fig.

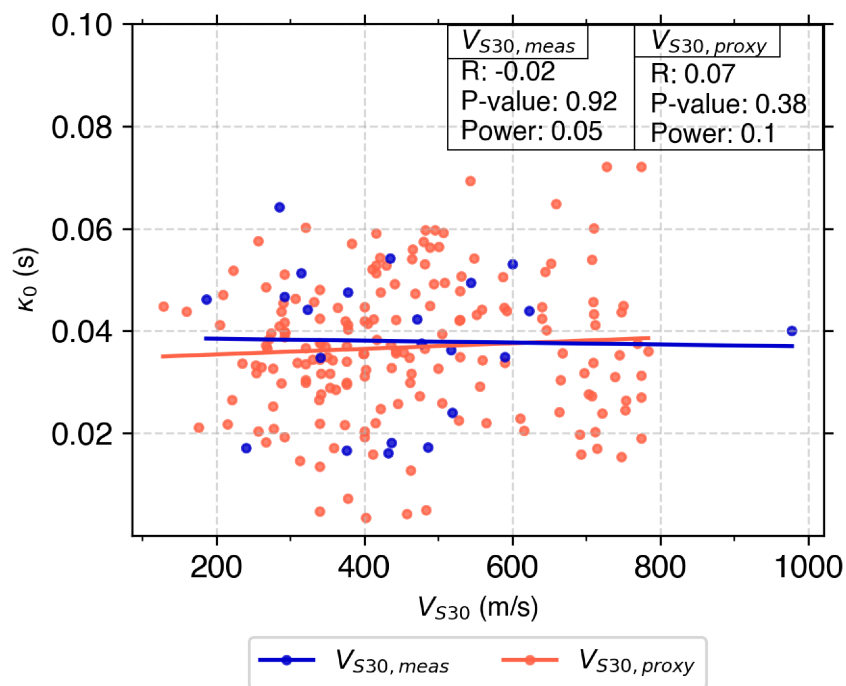


Figure 4.8. Correlation between κ_0 estimates from this study and both measured, and proxy V_{S30} ($V_{S30,meas}$ and $V_{S30,proxy}$, respectively). The statistical parameters Pearson R correlation coefficient (R), p -value, and statistical power (Power) are also indicated. There is a weak, slightly positive correlation between κ_0 and $V_{S30,proxy}$ and essentially no correlation with $V_{S30,meas}$. Both the κ_0 - $V_{S30,meas}$ and κ_0 - $V_{S30,proxy}$ relations show no statistical significance with p -values $\gg 0.05$ and statistical powers $\ll 0.8$.

4.8). The correlation is presented in linear space, but as some V_{S30} - κ_0 correlations have been presented in log-log space (Drouet et al., 2010; Ktenidou et al., 2012, 2014, 2015; Ktenidou and Van Houtte, 2012; Van Houtte et al., 2011), we include a log-log correlation in the supplement (see Fig. 4.S4). Other studies provide little justification for the choice of

correlation scaling, but we assume a linear relation, and the statistical parameters are not significant for either assumption. As with Klimasewski et al. (2019), we observe a weakly negative correlation for $V_{S30,meas}-\kappa_0$ and a weakly positive correlation for $V_{S30,proxy}-\kappa_0$. However, neither could be considered statistically significant, with both having a large p -value and low statistical power. The $\kappa_0-V_{S30,meas}$ and $\kappa_0-V_{S30,proxy}$ correlations have a Pearson R of -0.02 and 0.07 , a p -value of 0.92 and 0.38 , and a statistical power of 0.05 and 0.01 , respectively. We expect SFBA sites with larger V_{S30} values (\geq approximately 500 m/s, NEHRP class B/C; Holzer et al., 2005), such as soft-to-hard rock to generally be less attenuating, and we find that they are correlated with smaller values of κ_0 . Likewise, we expect lower V_{S30} sites (\leq approximately 500 m/s, National Earthquake Hazards Reduction Program [NEHRP] class C/D; Holzer et al., 2005), such as softer rock or loose sediment, to be more attenuating and correlate with larger values of κ_0 . We observe this trend with $V_{S30,meas}$, but the paucity of measured values likely contributes to the poor correlation.

4.5 Discussion

4.5.1 *Kappa Distance Dependence*

Most studies of κ have determined a distance dependence, because it is related to the Q -profile and generally increases with epicentral distance. However, the relationships between κ and distance from such studies are varied. Often, the relationship is assumed to be monotonically linear (Anderson and Hough, 1984; Ktenidou et al., 2013), but some studies have found the relationship to be nonlinear (Anderson, 1991) or varying with distance (Kilb et al., 2012). For example, Kilb et al. (2012) only observe a linear relationship for distances between 40–70 km. For this study, we assumed a linear distance dependence in the regression for κ_0 , which we find consistent with our κ versus distance plots. We still observe a linear relationship even though many of the stations in our dataset recorded events outside of the 40–70 km range (see κ versus distance plots in the online Zenodo dataset, Nye et al., 2022b).

4.5.2 *Relationship Between κ_0 and Geology*

The κ_0 estimates from this study suggest a spatial dependence in the region. We qualitatively investigate the relationships between κ_0 and local geology to evaluate whether this trend is influenced by surficial geology and/or proximity to regional faults.

Shallow Geology: The Santa Rosa Plain (SRP) region (zone 1) is covered by thick units of weak, loosely consolidated rocks which range in thickness up to several km in some basins, such as the Windsor Basin to the north and the Cotati Basin to the south (Sweetkind et al., 2010). A marine basin is also observed in zone 3, extending between the Zayante and Sargent fault zones (Eberhart-Phillips, 2016; Lees and Lindley, 1994). Weak and unconsolidated materials, such as the layers of silt, clays, sands, gravels, cobbles, and tuffaceous rocks found in these basins typically have a lower V_S and Q , which influence intrinsic attenuation of energy (Eberhart-Phillips, 2016). Assuming κ is related only to attenuation, it is mathematically equivalent to the full path attenuation parameter, t^* , and is inversely related to V_S and Q (Anderson and Hough, 1984; Fletcher and Boatwright, 2020). Thus, the higher κ_0 observed in these zones is consistent with expected attenuation trends and existing attenuation studies of the region (Eberhart-Phillips, 2016; Lees and Lindley, 1994).

Zone 1 also contains volcanic deposits to the southeast of the SRP (US Geological Survey, n.d.). Although these rocks are more compact than the weak basins, they are highly fractured from nearby faults (Sweetkind et al., 2010). Fractures are contributors of high-frequency attenuation, as they scatter energy from higher frequencies to lower frequencies (Abercrombie, 1997; Frankel, 1991; Kneib and Shapiro, 1993). Thus, it makes physical sense that we also estimate moderately high κ_0 values here. This likely also contributes to the larger κ_0 observed near segments of the SAF to the northwest (zone 1) and southeast (zone 3) of the bay, as well as near the HF and CF east of the bay (zone 2). The greater attenuation indicated in these regions is consistent with low- Q zones imaged in previous attenuation studies of northern-central California (Lees and Lindley, 1994; Eberhart-Phillips, 2016).

The east bay region (zone 2) is not only dominated by two major fault zones (HF and CF) and several more minor fault zones, but also mantled by weak material, such as bay mud, artificial fill, and alluvium (Eberhart-Phillips, 2016; Hough, 1990). This zone also encompasses unconsolidated sediment in the SSD, which is sourced from the Sacramento and San Joaquin rivers—two of the largest rivers in California. The thick units of weak material characterizing this region correlate with larger κ_0 from this study, as well as low Q from other attenuation studies of the SSD (Eberhart-Phillips et al., 2014; Erdem et al., 2019).

The shoreline along the western edge of the bay (zone 4) stands out as having some of the smallest values of κ_0 in the SFBA. Although zone 4 contains some alluvial units, the peninsula is more widely dominated by uplifted sedimentary, volcanic, and metamorphic rocks of the Franciscan and Great Valley complexes (US Geological Survey, n.d.). The absence of much surficial sediment west of the shoreline likely allows for greater preservation of high-frequency energy. Another consideration for the low κ_0 in this zone is that the SAF could act as a waveguide, thus amplifying high-frequency energy from trapped waves. Evidence of this effect has been observed in southern sections of the fault zone (Li and Teng, 1990; Li et al., 2004) Olsen et al., 2006). However, this effect appears to be localized to the peninsula. Other sections of the SAF exhibit higher values of κ_0 (zones 1 and 3). The greater preservation of high-frequency energy in this region is especially pertinent for seismic hazard, as the SAF is a source of large earthquakes.

Although much of the SFBA exhibits attenuating conditions which could reduce the seismic hazard associated with high-frequency ground motion, this same material can also result in strong amplification of low frequencies if a sharp impedance boundary exists below the sediment layer (Aki, 1988, 1993; Hough, 1990). As low-frequency amplification can still have a great effect on the resilience of structures enduring earthquake ground motion when considering resonant frequencies of structures, these high κ_0 regions should not be disregarded as regions of seismic hazard or risk.

Fault Zones: The SFBA is an active tectonic region and contains several major active faults previously mentioned, including the SAF, HF, CF, and RCF, as well as numerous other minor faults. It is well known that fault activity can deform the surrounding wall rock resulting in a fault damage zone characterized by microfractures, folds, and fractured rock (Choi et al., 2016). It is reasonable to assume such zones would influence seismic attenuation, so we investigate the relationship between κ_0 and proximity to faults.

Determining the boundaries of fault damage zones has varied between studies, resulting in damage zone widths that span orders of magnitude (Choi et al., 2016). In California, studies have determined fault zone widths ranging between tens of meters up to around 1.5 km (e.g. Cochran et al., 2009; Li et al., 2016; Ostermeijer et al., 2020; Powers and Jordan, 2010; Schulz and Evans, 2000). For the sake of simplicity, we assume an average fault zone width of 1 km and evaluate κ_0 of stations located within this boundary for the SAF, HF, CF, and RCF. For a more detailed analysis of the effect from fault damage zones on κ_0 estimates, one could perform a seismic velocity study to better determine the damage zone boundaries for each fault of interest (e.g. Cochran et al., 2009).

For the above criteria, we find an average κ_0 of 0.040 s for stations within the damage zones and 0.037 s for stations outside the damage zones. The higher attenuation observed within the fault zones could be due to the fractured, heterogeneous properties of the surrounding rock. We perform a Kolmogorov-Smirnov (K-S) hypothesis test on these datasets to evaluate the significance of the relationship, and we observe a D-statistic of 0.21 and a p -value of 0.31. The low D-statistic (< 0.31 : critical value for dataset sample size) and the large p -value both suggest the two datasets are statistically similar, thus suggesting little relationship between κ_0 and fault damage zones. This could be an indication of the scale at which regional geology influences κ_0 , but it could also be biased by the few stations within our designated fault zones (22). Increased instrumentation within fault deformation zones has been of recent interest in the geophysical community to better constrain rupture processes, and could also provide information for site response within a fault zone. The Community Near-Fault Observatory, a current IRIS instrumentation initiative (IRIS, n.d.),

has proposed installing instruments across major faults in Southern California, but a similar initiative in the SFBA could be an avenue for future work.

4.5.3 Comparison with other Regional Studies

This is the first regional κ_0 study of the SFBA; however, there have been several κ_0 studies in southern California near the San Jacinto Fault (SJF) zone (Anderson 1991; Johnson et al., 2020; Kilb et al., 2012; Klimasewski et al., 2019). Their estimates ranged from 0.002 to 0.023 s, 0.021 to 0.037 s, 0.022 to 0.047 s, and 0.017 to 0.059 s, respectively. A majority of estimates from this study fall within the range of values from studies in Southern California, but the SFBA exhibits greater high-frequency attenuation as indicated by the larger upper range (0.072 s). Geology around the SJF is more homogenous than in the SFBA and is characterized more by volcanic and metamorphic and other consolidated rocks, whereas the SFBA contains significantly more loose sedimentary cover. Thus, we find the larger estimates of this study to be in line with the trend of other studies.

4.5.4 κ_0 and V_{S30} as Site Parameters

Correlation with Observed Ground Motion: Our study is motivated by the assumption that κ_0 is a suitable descriptor of site characteristics and that seismic hazard studies can benefit from its inclusion as a site parameter. Although it has been studied for several decades, literature on κ is still limited, thus we aim to explore this assumption with our κ_0 estimates to evaluate its robustness. We find a modest correlation between κ_0 with site terms from small-to-moderate magnitude earthquake PGA in the SFBA; however, significant scatter is observed. Although PGA is generally understood to come from higher-frequency ground motion based on theoretical source spectra (Brune, 1970), some studies have found PGA to be influenced by a wider frequency band (Bora et al., 2016; Ktenidou et al., 2022). Acceleration amplitudes could also be controlled by low-frequency ground motion if amplification from near-site conditions overpowers high-frequency energy. Considering the SFBA is dominated by weak sites, this could also be a contributing factor in the scatter. Nevertheless, the trend observed supports our

understanding of κ_0 , where PGA site response shows a negative relationship with κ_0 . The negative relationship indicates sites with a larger κ_0 (greater high-frequency attenuation) experience weaker peak ground motion, which is consistent with results from Klimasewski et al. (2019). Although κ_0 correlates well with $\delta S_{j,PGA}$, V_{S30} correlates equally well when considering proxy values. Thus, both could be considered reasonable parameters for PGA studies in the SFBA. However, while one might think to disregard κ_0 considering its equivalent performance to V_{S30} , both provide uniquely important information about the ground motion and frequency content, and as such, κ_0 would still be a useful parameter in future ground motion studies of the region. In particular, κ_0 could be a useful parameter for modeling FAS at higher frequencies (Bayless and Abrahamson, 2018). We did not consider FAS GMM site residuals in this study, but this could be another avenue for future work.

There is essentially no relationship between κ_0 and $\delta S_{j,PGV}$, as indicated by the κ_0 - $\delta S_{j,PGV}$ results (Fig. 4.7b), which is unsurprising as PGV is understood to be derived from a lower frequency range than PGA, and κ_0 is considered a high-frequency parameter (Brune, 1970). V_{S30} , whether measured or proxy, would be a more robust site-parameter for PGV studies.

These analyses were performed using site residuals obtained from two different GMMs of varying complexity. Relevant to site effects, ASK14 includes $Z_{1.0}$ in addition to V_{S30} , which is also included in BA14. We find that κ_0 relates comparably to site residuals from both GMMs, suggesting the addition of $Z_{1.0}$ does not account for high-frequency attenuation. This is unsurprising as $Z_{1.0}$ is included in models to account for basin effects, which is usually an amplification of low-frequency energy. However, V_{S30} relates slightly better with site residuals from the more complex ASK14 GMM. This suggests the addition of $Z_{1.0}$ accounts for some site effects unaccounted for by V_{S30} . Thus, the remaining site effects after including basin effects a priori are better attributed to V_{S30} .

κ_0 - V_{S30} Relation: Although V_{S30} has shown to be limited in its application as a proxy for site conditions in some studies, it is often the only site measurement accessible. Because of this, we found it profitable to examine the relationship between V_{S30} and

κ_0 for when κ_0 measurements are not available. $V_{S30,meas}$ correlates negatively with κ_0 , which is consistent with most other κ_0 – V_{S30} studies (Chandler et al., 2006; Drouet et al., 2010; Edwards et al., 2011; Ktenidou et al., 2012, 2013; Ktenidou and Van Houtte, 2012; Klimasewski et al., 2019; Silva et al., 1998; Van Houtte et al., 2011). Lower V_{S30} units ($\leq \sim 500$ m/s), such as weak rock, soft soil, and loose sediments tend to attenuate high-frequency energy as V_{S30} and Q are inversely related to intrinsic attenuation (Anderson and Hough, 1984). Neither correlation with $V_{S30,meas}$ or $V_{S30,proxy}$ could be considered statistically significant, but $V_{S30,meas}$ exhibits a stronger correlation with κ_0 than $V_{S30,proxy}$. If measurements of V_{S30} are available, V_{S30} could be used as a proxy for κ_0 , but the use of V_{S30} estimates as a proxy would likely result in more unreliable estimates of κ_0 .

4.6 Conclusions

In this study, we estimate the high-frequency attenuation parameter, κ_0 , for 228 stations in the San Francisco Bay area following the method of Anderson and Hough (1984). Our results show a range in values from 0.003 to 0.072s, with most falling between 0.028 and 0.045 s. The spatial distribution of κ_0 suggests a spatial dependence that is controlled in part by the heterogeneous shallow geology of the region. The largest κ_0 estimates are observed to the north, east, and south of the bay, especially in the Sacramento–San Joaquin Delta, whereas the smallest are observed on the western peninsula around the San Andreas Fault.

We further evaluate κ_0 as a site parameter and compare it with the commonly used V_{S30} . κ_0 correlates well with PGA, which is promising for future seismic hazard applications. Correlations of ground motion with V_{S30} can be equally reliable but are sensitive to whether V_{S30} is measured or obtained by proxy. Interestingly, proxy values for V_{S30} correlate better with PGA, but this could be biased by the limited V_{S30} measurements.

The regional map of κ_0 from this study may be a useful contribution to future hazard studies in the SFBA. Ideally, κ_0 would be incorporated directly as a site-parameter in the development of new GMMs, especially in regions where sufficient data exist to develop fine-resolution models. Even if the development of new GMMs is unlikely in

the foreseeable future, these models are still valuable in host-to-site adjustments which have typically used an average value of κ to describe the entire region. Site-specific κ adjustments would better constrain expected ground motion and hazards, especially in regions with fine-scale heterogeneities in regional geology.

It is important to recognize that the results presented in this study are based on weaker ground motions. The relationship between κ_0 and soil response resulting from stronger ground motions has been investigated in several studies (Ji et al., 2021; Ktenidou et al., 2015), and a moderate correlation has been observed where stronger ground motion results in smaller κ_0 . However, the magnitude of the effects of soil nonlinearity is site-dependent and it seems to be a greater factor with softer sites (Ji et al., 2021). As the SFBA is dominated mostly by softer shallow geologic conditions and has the possibility of experiencing shaking from \sim M8 events, these effects are important to consider in ground-motion studies of the SFBA.

We also recognize the limitations of κ_0 in that it is solely a high-frequency parameter and does not account for low-frequency amplification, an important consideration especially in engineering of large structures. Thus, characterization of other site metrics, where available, could contribute important additional information in seismic hazard studies.

4.7 Data and Resources

Our earthquake catalog was accessed from the U.S. Geological Survey (USGS) on August 2019 using *ObsPy* (Beyreuther et al., 2010). Broadband and strong motion seismic records were downloaded from the Northern California Earthquake Data Center (NCEDC) using *ObsPy* (last accessed June 2022). Seismic records belong to the BK (Northern California Earthquake Data Center, 2014), BG (no doi), NC (USGS Menlo Park, 1967), NP (US Geological Survey, 1931), and WR (no doi) seismic networks. Fault geometries were obtained through the USGS Fault and Fold database (US Geological Survey, 2017, last accessed August 2017). Measured V_{S30} values were taken from the McPhillips et al. (2020) catalog. Estimated V_{S30} values were downloaded from the Next Generation

Attenuation-West2 Project (NGA-West2) site database file (Ancheta et al., 2014; Pacific Earthquake Engineering Research Center, 2013) and USGS hybrid global V_{S30} database (Heath et al., 2020; Thompson et al., 2014). The global database was last accessed on August 2021. The USGS 3D velocity model of the Bay Area was queried using CenCalVM (US Geological Survey, 2018, 2020b). Maps were produced using the Generic Mapping Tools (GMT) software (Wessel et al., 2019) with a digital elevation model downloaded from GMTSAR (Sandwell et al., 2011), and all the analyses were completed with in-house code, available online at https://github.com/taranye96/kappa_bayarea (last accessed November 2022). This includes the code to determine f_e and f_x , compute κ on individual spectra, complete the regressions, and spatially interpolate κ_0 . The supporting information contains figures showing example station earthquake-based horizontal-to-vertical spectral ratio (eHVSr) plots, a regional kappa map for the entire SFBA, a scatterplot of estimated versus interpolated κ_0 , a log-log V_{S30} - κ_0 plot, and a table listing stations with less reliable estimates of κ_0 . Other supporting material can be found in the Zenodo dataset (Nye et al., 2022b), which includes the eHVSr curves, spectra, and κ_0 regression plots for each station, .kml and .txt files of the κ_0 estimates, and .grd files of the kriged κ_0 predictions. Initial efforts for this project using different methodology were presented at the 2019 Southern California Earthquake Center (SCEC) Annual Meeting (King et al., 2019). Preliminary data processing and station κ_0 results using the updated methodology were presented at the 2022 12th National Conference on Earthquake Engineering (Nye et al., 2022a), but results differ slightly in this paper due to corrections in data processing and minor revisions to methodology.

4.8 Acknowledgments

The authors are grateful to the U.S. Geological Survey (USGS) for supporting this work under the USGS EHP Grant Number G19AP00071. The authors graciously thank Susan Hough as a USGS internal reviewer, as well as Samantha Palmer and a second anonymous reviewer, whose comments and revisions enhanced the quality and scientific rigor of this article. The authors also give special thanks to Stefano Parolai and Ashly

Cabas for providing their valuable insight on the relationships between site response and
κ. Any use of trade, firm, or product names is for descriptive purposes only and does not
imply endorsement by the U.S. Government.

4.9 Supporting Information

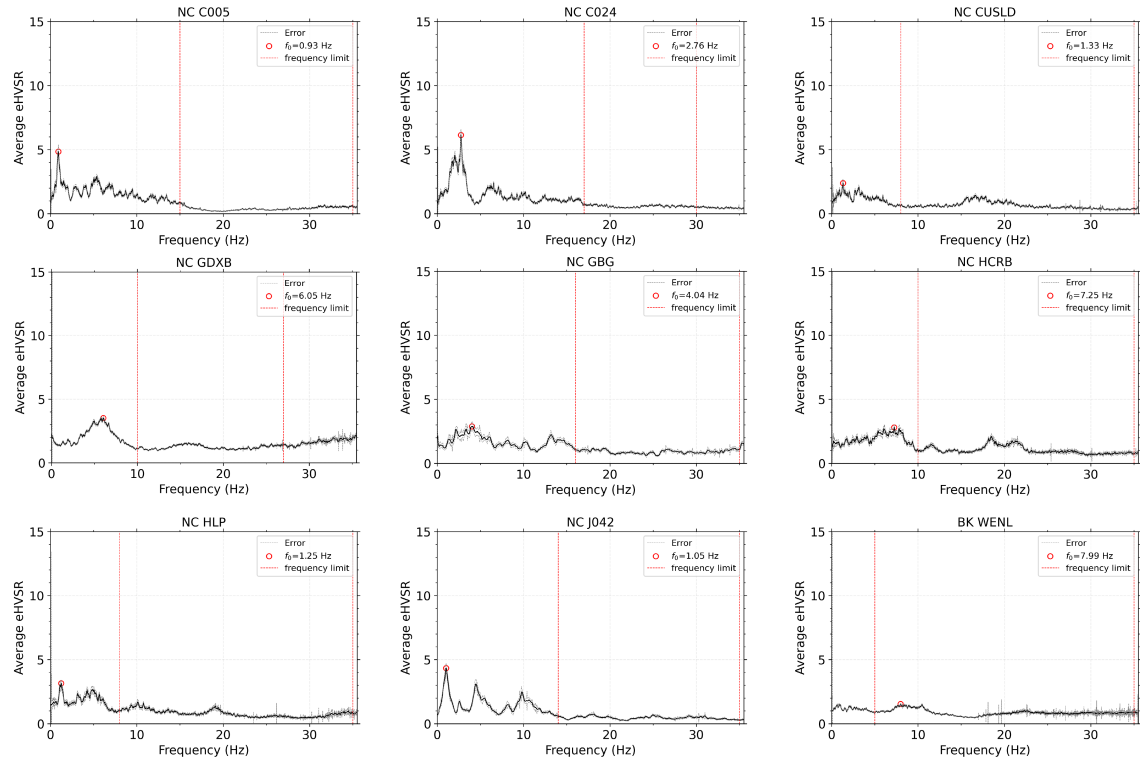


Figure 4.S1. Example eHVSr curves for nine stations. The average eHVSr curve is represented by a solid black line, the error is represented by light gray lines, the fundamental frequency is indicated by a red circle, and the manually-selected maximum frequency limits are indicated by two dashed red lines.

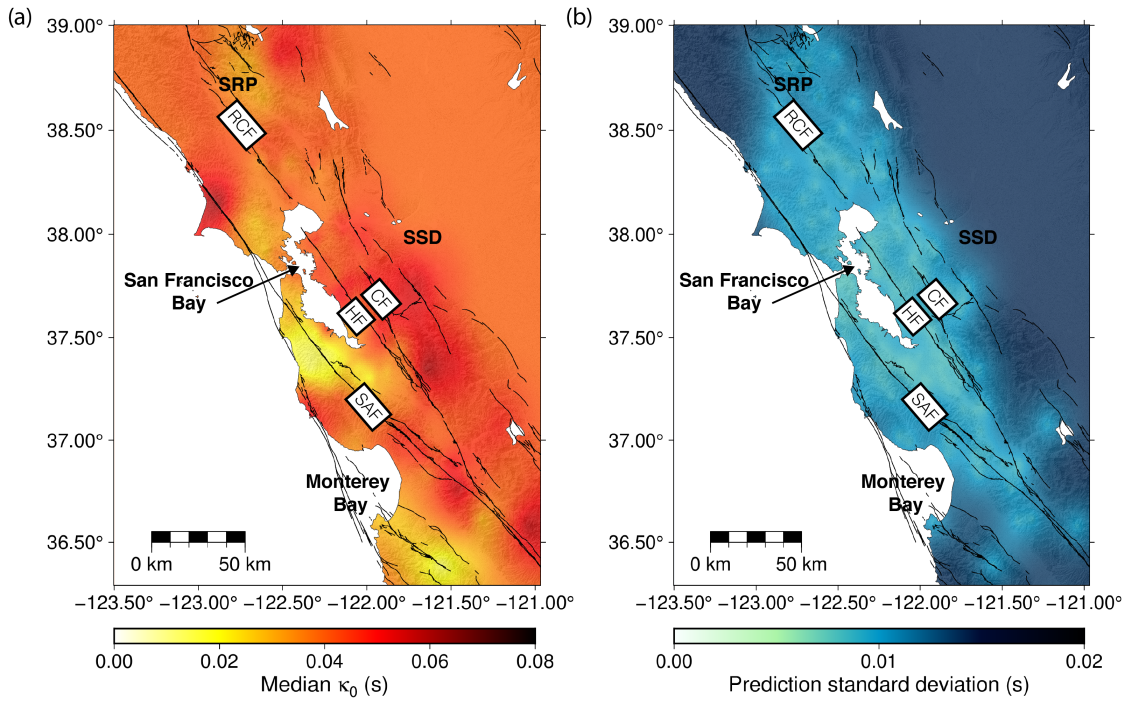


Figure 4.S2. Full interpolated (a) κ_0 map and (b) standard deviation maps. The Sacramento-San Joaquin Delta (SSD) and Santa Rosa Plain (SRP) are labeled, as well as the San Andreas Fault (SAF), Hayward Fault (HF), Calaveras Fault (CF) and Rodgers Creek Fault (RCF). Other Quaternary faults are also represented by solid black lines.

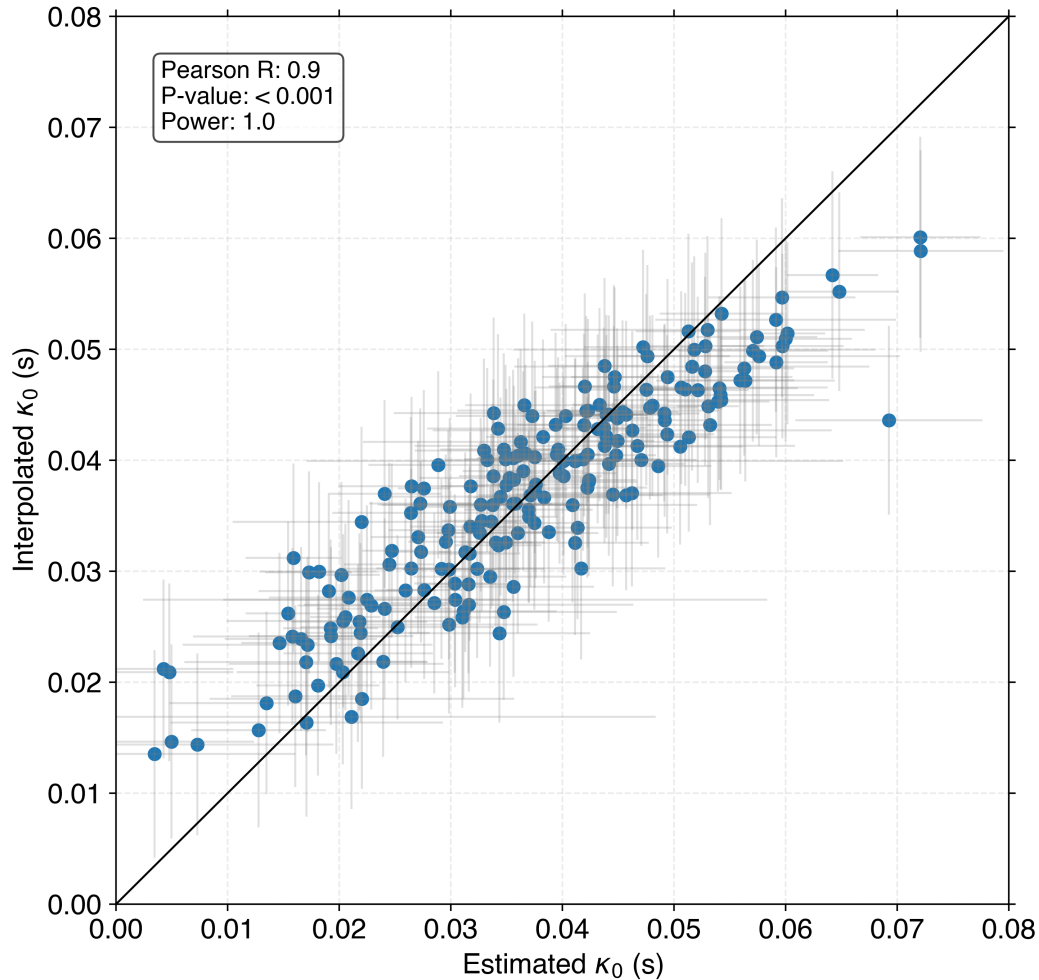


Figure 4.S3. Comparison of estimates of κ_0 with interpolated predictions for the 224 sites in our dataset. Error bars are represented by light gray lines, and the line of equality is represented by a black line. The statistical parameters Pearson R correlation coefficient (R), p -value, and statistical power (Power) are indicated on the figure. The kriged model fits the estimates well for estimates between ~ 0.02 and 0.06 s. The kriged model overpredicts for smaller estimates and underpredicts for larger estimates, suggesting the kriged model smooths extreme outliers in the estimates.

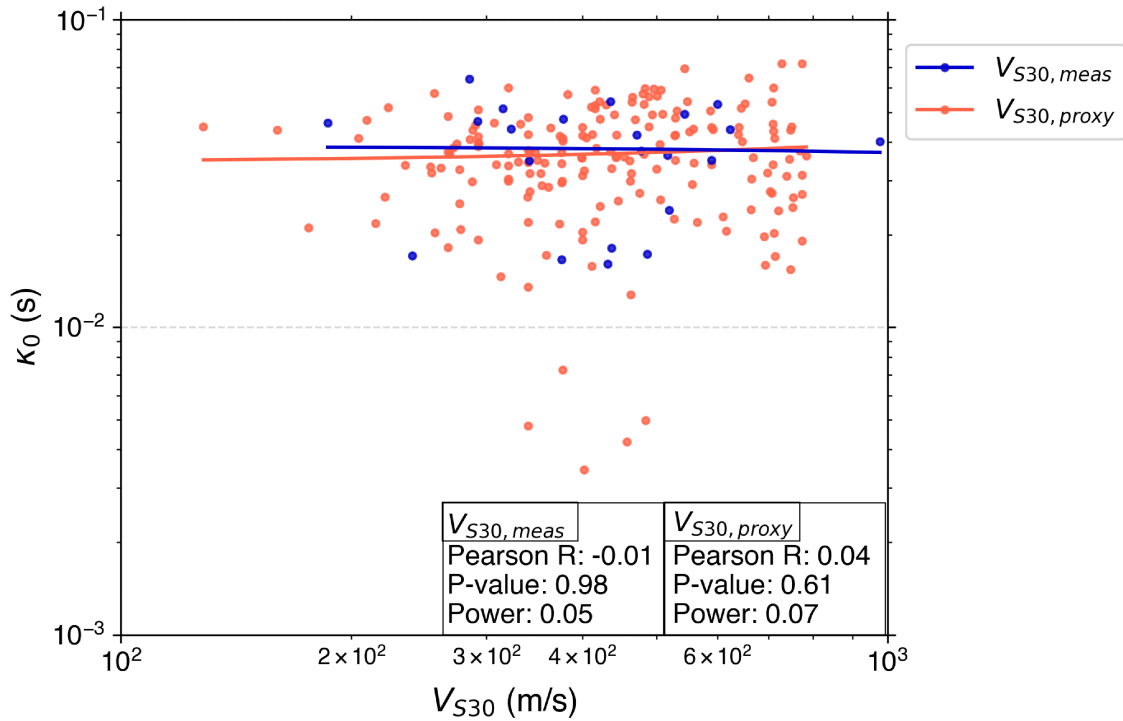


Figure 4.S4. Correlation between κ_0 estimates from this study and both measured and proxy V_{S30} ($V_{S30,meas}$ and $V_{S30,proxy}$, respectively) in log–log space. The statistical parameters Pearson R correlation coefficient (R), p -value, and statistical power (Power) are indicated on the figure. There is a weak, slightly positive correlation between κ_0 and $V_{S30,proxy}$, and essentially no correlation with $V_{S30,meas}$. Both the κ_0 – $V_{S30,meas}$ and κ_0 – $V_{S30,proxy}$ relations show no statistical significance with p -values $\gg 0.05$ and statistical powers $\ll 0.8$.

Table 4.S1. Stations removed due to unreliable κ_0 regression

Station	Reason
1780	Site response
1803	Negative κ_0 -epicentral distance correlation
1807	Negative κ_0 -epicentral distance correlation
1819	Site response
1828	Negative κ_0 -epicentral distance correlation
1838	Negative κ_0 -epicentral distance correlation
1863	Site response
BRIB	Negative κ_0 -epicentral distance correlation
C012	Negative κ_0 -epicentral distance correlation
C013	Negative κ_0 -epicentral distance correlation
C021	Negative κ_0 -epicentral distance correlation
C024	Negative κ_0 -epicentral distance correlation
C028	Negative κ_0 -epicentral distance correlation
CBR	Negative κ_0 -epicentral distance correlation
CPM	Non-linear κ_0 -epicentral distance correlation
GAXB	Non-linear κ_0 -epicentral distance correlation
GGPC	Negative κ_0 -epicentral distance correlation
HLO	Negative κ_0 -epicentral distance correlation
J003	Negative κ_0
J029	Negative κ_0 -epicentral distance correlation
JBNB	Negative κ_0 -epicentral distance correlation
JPR	Negative κ_0 -epicentral distance correlation
PACP	Negative κ_0 -epicentral distance correlation
SPR	Negative κ_0 -epicentral distance correlation

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

The central theme tying together each of the chapters of this dissertation is improving models of earthquake ground motion through the lens of earthquake sources, paths, and sites. Considering ground-motion modeling drives many seismic hazard and risk studies, well-constrained models are vital for development of national seismic hazard models, effective building codes, insurance estimates, and earthquake early warning performance analyses.

Chapter II began with addressing the source component of the ground-motion puzzle by constraining rise time, rupture velocity, and high frequency stress parameter values for modeling rare, yet destructive, “tsunami earthquakes”. Near-field recordings necessary for both analyzing the source physics of such events and testing early warning capabilities only exist for the most recent 2010 **M**7.8 Mentawai tsunami earthquake. Using available near-field HR-GNSS and strong motion data for this event, we performed various perturbations on one-dimensional semistochastic simulation rupture parameters to ascertain the values needed to model characteristic observed features of the tsunami earthquake. We found that a long rise time, slow rupture velocity, and low stress parameter reasonably capture the long duration and energy deficiency, and the derived values align well with values from inversion studies of the same event. The simulations also performed similarly to the observed data when run through an example real-time tsunami earthquake discrimination algorithm: $\mathbf{M}_{PGA}-\mathbf{M}_{PGD}$ (Sahakian et al., 2019). These results support the use of simulations in testing such an algorithm in tsunami early warning practices.

The results of Chapter II were based on data from one event; however, considering long rise time, slow rupture velocity, and energy deficiency are universally shared characteristics for tsunami earthquakes, we maintain that our results can be used to simulate realistic tsunami earthquake scenarios in other subduction zones globally. We presented methods for adjusting these parameters in the *FakeQuakes* model as only the stress parameter is directly prescribed in the parameter file, and we also presented equations for selecting reasonable rupture parameter values for simulating various tsunami earthquakes. With such information, a clear path forward includes extracting values from the tsunami earthquake parameter space to simulate events in regions with a known tsunami earthquake history.

Although the tsunami earthquake historical record is sparse, the Sunda, Peru–Chile, and Japan trenches have each hosted several tsunami earthquakes. These regions are thus statistically higher to experience another destructive tsunami earthquake in the future and would greatly benefit from improved tsunami early warning for such events. Simulations could be used to test tsunami earthquake discrimination algorithms in these regions. Because the historical record is sparse, near-trench coupling is poorly-resolved, and these events do not appear to favor any sort of tectonic conditions, we cannot be certain that other regions could not also host a tsunami earthquake at some point. It would be beneficial to evaluate how early warning systems would also perform in the event of a shallow tsunami earthquake in regions without a known historical tsunami earthquake. However, one might be able to deduce segments of a trench more likely to generate a tsunami earthquake by analyzing spatial distributions of low stress shallow events (Bilek et al., 2016).

The M_{PGA} – M_{PGD} algorithm for real-time tsunami earthquake discrimination requires the availability of both near-field seismic and geodetic data. Although installation of inertial seismic instruments has increased substantially over the past several decades, many regions still lack abundant HR-GNSS instruments. Global geodetic networks are being promoted and improved upon by the Global Geodetic Observing System (GGOS).

Chapter III also focused on ground-motion simulations but expanded to include

source and path effects of an entire region. We modeled moderate-to-large Cascadia Subduction Zone earthquakes and evaluated their effectiveness as an earthquake early warning training set by validating the earthquake detection, magnitude estimation, and ground-motion intensities of the simulations. P-wave generation was recently added to *FakeQuakes* (Goldberg and Melgar, 2020), but it was validated only at short distances. A full margin Cascadia rupture could include hypocentral distances up to around 1000 km, especially if it were to initiate along the southern margin, which has a higher probability. We found that most of the event recording P-waves were detected by short-term average/long-term average (STA/LTA) algorithms, including some at the greater distances. Furthermore, even with the addition of noise, the STA/LTA algorithms detected P-wave arrivals within 1 s of the true arrival for a majority of the records. We also found the simulated P-waves to exhibit expected growth patterns used for real-time magnitude estimation. Regarding the simulated ground motion, initial evaluations found a strong distance bias with high frequency intensity measures, which can be attributed to poor attenuation modeling. Adjustments and new attenuation parameters were subsequently added to *FakeQuakes*, which improved modeling of ground motion out to ~ 600 km for peak ground acceleration, peak ground velocity, and modified Mercalli intensity. Peak ground displacement did not exhibit a strong attenuation bias; however, it is highly sensitive to the percentile of noise added, especially for smaller magnitude events. Recordings from several of the simulated scenarios were run through both EPIC's and ONC-EW's offline systems, and both systems performed reasonably well considering they were developed based on point source assumptions rather than large finite ruptures.

With the successful validation of the simulations presented in Chapter III, the simulated dataset in its entirety can now be run through Cascadia region earthquake early warning systems to evaluate expected performance more thoroughly in the event of a large megathrust event. Our first order performance evaluations, although fairly successful, do highlight components that can be improved upon, such as the magnitude estimation used by ONC-EW. Currently, PGD–**M** scaling relations are used by ONC for larger magnitude

estimation. Such relations are based on point source assumptions and can result in significant overestimations if large patches of slip are observed at greater distances from the hypocenter. EPIC primarily underestimates the magnitude as it is estimated from the seismic data. Installation of collocated or nearly collocated geodetic instruments at many of the PNSN sites would allow for generation of unbiased displacements and minimize magnitude saturation. A more reliable approach for both early warning systems could be the incorporation of machine learning for magnitude estimation, which is not limited by simple scaling models (Lin et al., 2021).

Although the Cascadia dataset presented has been found suitable for regional earthquake early warning applications, the simulated data can be further improved as more detailed information arise about regional properties. The Pacific Northwest is seismically quiet. As such, the simulated ground motions were modeled to fit global ground-motion models. Several initiatives are currently underway in developing a higher resolution velocity model for the Cascadia region (Cascadia2021 and CRESCENT working groups).

Lastly, Chapter IV focused solely on improving modeling of site effects in empirical ground-motion models (GMMs) for the San Francisco Bay Area (SFBA). Using a large dataset of small-to-moderate earthquakes, we estimated high frequency attenuation (κ) and its site-specific component (κ_0) at over 200 sites in the SFBA. Because κ is a measurement of decay of in the Fourier spectrum, frequency bounds must be selected within which the decay is primarily observed. Some studies use uniform bounds for all recordings, which can introduce bias from varying source corner frequencies and site amplification effects. We presented an algorithm for selecting record-specific frequency bounds, which minimizes such biases and reduces manual labor and time. Using this algorithm, we identified four primary κ_0 zones in the SFBA. Zones 1–3 to the north, east, and south, exhibit higher levels of high frequency attenuation, whereas zone 4 exhibits less high frequency attenuation. These spatial trends are consistent with the regional anelastic models (Eberhart-Phillips et al., 2014, 2016), shallow geology, and fault geometry.

This work was performed with the intent that the resulting κ_0 map would be used

for site-specific corrections of existing ground-motion models, or as a regression parameter in future ground-motion models. We evaluated the effectiveness of κ_0 as a descriptor of ground motion by evaluating relationships between the parameter and GMM site residuals. We found that for peak ground acceleration, κ_0 performs equally well compared to the more commonly used V_{S30} . However, V_{S30} is a better descriptor of lower frequency peak ground velocities. Considering κ_0 relates to specific frequencies of ground motion, it will likely have a more significant effect on Fourier amplitude spectra GMMs than peak ground acceleration, which is not well constrained to a certain range of frequencies. It would be informative to perform a similar GMM site residual analysis with the Bayless and Abrahamson (2018) Fourier amplitude spectra model.

κ_0 has the potential to be a valuable contribution to ground-motion studies as it informs expected frequencies of ground motion critical for seismic building design, and it is not depth-constrained like V_{S30} . Unfortunately, unlike V_{S30} which can be directly measured through active source studies or estimated using various relationships with geological and topographical features, κ_0 must be empirically estimated, and its robustness relies on the abundance and quality of data available. To expand the contribution of κ_0 to ground-motion studies, existing datasets of κ_0 must be available. Such datasets already exist for the SFBA (Nye et al., 2022a), Southern California (Anderson, 1991; Kilb et al., 2012; Klimasewski et al., 2019), Midwestern United States (e.g., Biasi and Anderson, 1991; Darragh et al., 2019), Canada (e.g., Palmer and Atkinson, 2020, 2023), Chile (e.g., Neighbors et al, 2015; Pozo et al., 2022), Mexico (e.g., Castro and Ávila-Barrientos; Fernández et al., 2010; Lermo et al., 2016), Iceland (Sonnemann et al., 2019), the Azores (Carvalho et al., 2016), United Kingdom (Rietbrock and Edwards, 2017), France (Douglas et al., 2010; Perron et al., 2017), Italy (Castro et al., 2022), Switzerland (Edwards et al., 2011, 2015), the Alps (Gentili and Franceschina, 2011), Greece (Ktenidou et al., 2012), Croatia (e.g., Stanko et al., 2017, 2020), Romania (Pavel and Vacareanu, 2015), Turkey (e.g., Altindal and Askan, 2022; Askan et al., 2014; Sertcelik et al., 2022; Şişman et al., 2018), Iran (e.g., Samaei et al., 2016; Shafiee et al., 2021; Tafreshi et al., 2022), Japan (e.g., Ji et al., 2020, 2021; Tao

et al., 2020), China (e.g., Lei and Xiao-Jun, 2017; Sun et al., 2013), Taiwan (e.g., Chang et al., 2019; Huang et al., 2017; Lai et al., 2016), South Korea (Park et al., 2020), India (e.g., S. Kumar et al., 2018; V. Kumar et al., 2020; Mittal et al., 2021; Yadav et al., 2018), Morocco (Boulanouar et al., 2022), and New Zealand (Van Houtte et al., 2014, 2018). Similar studies should be carried out in other regions with high seismicity or high seismic risk, such as the Central-Eastern United States, Kyrgyzstan, Puerto Rico, Haiti, Indonesia, and the Philippines.

The results from each of the chapters in this dissertation can be leveraged to support better modeling of earthquake ground motion for three specific scenarios: rare tsunami earthquakes, large CSZ ruptures, and moderate crustal events in the SFBA. The field of ground-motion modeling, however, is vast, and there are a myriad of puzzle pieces still left undiscovered or poorly resolved. For example, the complex growth patterns and source scaling of large $M8.5+$ earthquakes, the intricacies of multi-fault interplay in crustal earthquake simulations, or the joint effects of site amplification and attenuation on ground motions.

Perhaps the most ambiguous piece of the puzzle is the path between the earthquake and site of interest. Numerous tomographic studies have developed three-dimensional models of velocity and attenuation, but their resolutions are limited by available data. Furthermore, the partitioning of Q attenuation into its different components is not always distinct. Even with perfect knowledge of such attenuation effects, their sources are not well known. What physical properties are contributing to the heterogeneities in attenuation? Is this something that can be better understood with improved imaging of both the shallow and deep subsurface? Is there a direct relationship between velocity and attenuation, and if not, where do those differences originate from? These questions are fundamental as the GMM community transitions to more fully non-ergodic modeling, which considers unique properties of each raypath. With continued improvements to models of earthquake ground motion and their uncertainties, we can improve our confidence in seismic hazard assessments and strive for a more seismically-resilient community.

REFERENCES CITED

- Aagaard, B. T., & Hirakawa, E. T. (2021). *San Francisco Bay region 3D seismic velocity model* (v21.1) [Data set]. U.S. Geological Survey data release.
- Abercrombie, R. E. (1997). Near-surface attenuation and site effects from comparison of surface and deep borehole recordings. *Bulletin of the Seismological Society of America*, 87(3), 731–744.
- Abercrombie, R. E., Trugman, D. T., Shearer, P. M., Chen, X., Zhang, J., Pennington, C. N., Hardebeck, J. L., Goebel, T. H. W., & Ruhl, C. J. (2021). Does earthquake stress drop increase with depth in the crust?. *Journal of Geophysical Research: Solid Earth*, 126(10), 1–22.
- Abrahamson N. A., Gregor N., & Addo K. (2016). BC Hydro ground motion prediction equations for subduction earthquakes. *Earthquake Spectra*, 32(1), 23–44.
- Abrahamson, N. A., Silva, W. J., & Kamai, R. (2014). Summary of the ASK14 ground motion relation for active crustal regions. *Earthquake Spectra*, 30(3), 1025–1055.
- Abrahamson, N., Kuehn, N., Gulerce, Z., Gregor, N., Bozorgnia, Y., Parker, G., Stewart, J., Chiou, B., Idriss, I.M., Campbell, K., & Youngs, R. (2018). *Update of the BC Hydro Subduction Ground-Motion Model using the NGA-Subduction Dataset*. PEER Report 2018/02. Pacific Earthquake Engineering Research Center, University of California, Berkeley, California.
- Aki, K. (1988). Local site effect on ground motion. *Earthquake Engineering and Soil Dynamics II*, 103–155.
- Aki, K. (1993). Local site effects on weak and strong ground motion. *Tectonophysics*, 218(1–3), 93–111.

- Al Atik, L., Kottke, A., Abrahamson, N., & Hollenback, J. (2014). Kappa (κ) scaling of ground-motion prediction equations using an inverse random vibration theory approach. *Bulletin of the Seismological Society of America*, *104*(1), 336–346.
- Allen, R. M., & Melgar, D. (2019). Earthquake early warning: Advances, scientific challenges, and societal needs. *Annual Reviews of Earth and Planetary Sciences*, *47*, 361–388.
- Allmann, B. P., & Shearer, P. M. (2007). Spatial and temporal stress drop variations in small earthquakes near Parkfield, California. *Journal of Geophysical Research*, *112*(B04305), 1–17.
- Altindal, A., & Askan, A. (2022). Predictive kappa (κ) models for Turkey: Regional effects and uncertainty analysis. *Earthquake Spectra*, *38*(4), 2479–2499.
- Ancheta, T. D., Darragh, R. B., Stewart, J. P., Seyhan, E., Silva, W. J., Chiou, B. S.-J., Wooddell, K. E., Graves, R. W., Kottke, A. R., Boore, D. M., Kishida, T., & Donahue, J. L. (2014). NGA-West2 database. *Earthquake Spectra*, *30*(3), 989–1005.
- Anderson, J. G. (1986). Implication of attenuation for studies of the earthquake source. *Earthquake Source Mechanics*, *37*, 311–318.
- Anderson, J. G. (1991). A preliminary descriptive model for the distance dependence of the spectral decay parameter in southern California. *Bulletin of the Seismological Society of America*, *81*(6), 2186–2193.
- Anderson, J. G., & Hough, S. E. (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.
- Anderson, J., & Humphrey, J. (1991). A least squares method for objective determination of earthquake source parameters. *Seismological Research Letters*, *62*(3–4), 201–209.
- Ashraf, A., Hooft, E. E. E., Trehu, A. M., Carbotte, S. M., Nolan, S., Wirth, E., & Ward, K. E. (2023). 3-D P-wave velocity model capturing the transition from Siletz to Franciscan Terrane in the South-Central Cascadia Subduction Zone. Abstract (T53E-0193) presented at 2023 AGU Fall Meeting, 11–15 December.

- Askan, A., Sisman, F. N., & Pekcan, O. (2014). A regional near-surface high frequency spectral attenuation (κ) model for northwestern Turkey. *Soil Dynamics and Earthquake Engineering*, *65*, 113–125.
- Atkinson, G. M., & Beresnev, I. (1997). Don't call it stress drop. *Seismological Research Letters*, *68*(1), 3–4.
- Atkinson, G. M., & Boore, D. M. (2003). Empirical ground-motion relations for subduction-zone earthquakes and their application to Cascadia and other regions. *Bulletin of the Seismological Society of America*, *93*(4), 1703–1729.
- Atkinson, G. M., & Kaka, S.L. I. (2007). Relationships between felt intensity and instrumental ground motion in the Central United States and California. *Bulletin of the Seismological Society of America*, *97*(2), 497–510.
- Atkinson, G. M., & Macias, M. (2009). Predicted ground motions for great interface earthquakes in the Cascadia Subduction Zone. *Bulletin of the Seismological Society of America*, *99*(3), 1552–1578.
- Baillard, C., Crawford, W. C., Ballu, V., Hibert, C., & Mangeney, A. (2014). An automatic kurtosis-based P and S phase picker designed for local seismic networks. *Bulletin of the Seismological Society of America*, *104*(1), 394–409.
- Baltay, A., Ide, S., Prieto, G., & Beroza, G. (2011). Variability in earthquake stress drop and apparent stress. *Geophysical Research Letters*, *38*(6), 1–6.
- Bayless, J., & Abrahamson, N. A. (2018). *An empirical model for Fourier amplitude spectra using the NGA-West2 database*. PEER Report 2018/07. Pacific Earthquake Engineering Research Center, University of California, Berkeley, California.
- Beyreuther, M., Barsch, R., Krischer, L., Megies, Behr, Y., & Wassermann, J. (2010). ObsPy: A Python toolbox for seismology. *Seismological Research Letters*, *81*(3), 530–533.
- Biasi, G., & Anderson, J. G. (2007). *Measurement of the parameter κ , and reevaluation of κ for small to moderate earthquakes at seismic stations in the vicinity of Yucca Mountain, Nevada*. Report TR-07-007. Nevada Seismological Laboratory.
- Bilek, S. L., & Lay, T. (1999). Rigidity variations with depth along interplate megathrust faults in subduction zones. *Nature*, *400*, 443–446.

- Bilek, S. L., & Lay, T. (2002). Tsunami earthquakes possibly widespread manifestations of frictional conditional stability. *Geophysical Research Letters*, 29(14), 18-1–18-4.
- Bilek, S. L., Rotman, H. M. M., & Phillips, W. S. (2016). Low stress drop earthquakes in the rupture zone of the 1992 Nicaragua tsunami earthquake. *Geophysical Research Letters*, 43(19), 10180–10188.
- Blaser, L., Krüger, F., Ohrnberger, M., & Scherbaum, F. (2010). Scaling relations of earthquake source parameter estimates with special focus on subduction environment. *Bulletin of the Seismological Society of America*, 100(6), 2914–2926.
- Bock, Y., Melgar, D., & Crowell, B. (2011). Real-time strong-motion broadband displacements from collocated GPS and accelerometers. *Bulletin of the Seismological Society of America*, 101(6), 2904–2925.
- Bommer, J. J., Douglas, J., Scherbaum, F., Cotton, F., Bungum, H., & Fäh, D. (2010). On the selection of ground-motion prediction equations for seismic hazard analysis. *Seismological Research Letters*, 81(5), 783–793.
- Boore, D. M. (2003). Simulation of ground motion using the stochastic method. *Pure and Applied Geophysics*, 160, 635–676.
- Böse M., Andrews, J., O'Rourke, C., Kilb, D., Lux, A., Bunn, J., & McGuire, J. (2022). Testing the ShakeAlert earthquake early warning system using synthesized earthquake sequences. *Seismological Research Letters*, 94(1), 243–259.
- Böse, M., Smith, D. E., Felizardo, C., Meier, M.-A., Heaton, T. H., & Clinton, J. F. (2018). FinDer v.2: Improved real-time ground-motion predictions for M2–M9 with seismic finite-source characterization. *Geophysical Journal International*, 212(1), 725–742.
- Boulanouar, A., Mittal, H., Rahmouni, A., Zian, A., Chourak, M., Géraud, Y., Harnafi, M., & Sebbani, J. (2022). Estimation of seismic kappa parameter and near-surface attenuation in Morocco. *Ecological Engineering & Environmental Technology*, 23(6), 128–139.
- Brune, J. N. (1970). Tectonic stress and the spectra of seismic shear waves from earthquakes. *Journal of Geophysical Research*, 75(26), 4997–5009.

- Butler, R. F. (2018). Cascadia earthquake science and hazards. In C. V. Fletcher & J. Lovejoy (Eds.), *Natural disasters and risk communication: Implications for the Cascadia subduction zone megaquake*.
- Carrington, L., Komatitsch, D., Laurenzano, M., Tikir, M. M., Michéa, D., Le Goff, N., Snively, A., & Tromp, J. (2008, November). High-frequency simulations of global seismic wave propagation using SPECFEM3D_GLOBE on 62K processors. In *Proceedings of the 2008 ACM/IEEE Conference on Supercomputing*. Austin, Texas USA, 27 June–1 July.
- Carvalho, A., Reis, C., & Vales, D. (2016). Source and high-frequency decay parameters for the Azores region for stochastic finite-fault ground motion simulations. *Bulletin of Earthquake Engineering*, *14*, 1885–1902.
- Castro, R.R., & Ávila-Barrientos, L. (2015). Estimation of the spectral parameter kappa in the region of the Gulf of California, Mexico. *Journal of Seismology*, *19*, 809–829.
- x
- Castro, R. R., Colavitti, L., Vidales-Basurto, C. A., Pacor, F., Sgobba, S., & Lanzano, G. (2022). Near-source attenuation and spatial variability of the spectral decay parameter kappa in Central Italy. *Seismological Research Letters*, *93*(4), 2299–2310.
- Chamoli, B. P., Kumar, A., Chen, D.-Y., Gairola, A., Jakka, R. S., Pandey, B., Kumar, P., & Rathore, G. (2021). A prototype earthquake early warning system for northern India. *Journal of Earthquake Engineering*, *25*(12), 2455–2473.
- Chandler, A. M., Lam, N. T. K., & Tsang, H. H. (2006). Near-surface attenuation modelling based on rock shear-wave velocity profile. *Soil Dynamics and Earthquake Engineering*, *26*(11), 1004–1014.
- Chang, S. C., Wen, K. L., Huang, M. W., Kuo, C. H., Lin, C. M., Chen, C. T., & Huang, J. Y. (2019). The high-frequency decay parameter (Kappa) in Taiwan. *Pure and Applied Geophysics*, *176*, 4861–4879.
- Choi, J. H., Edwards, P., Ko, K., & Kim, Y. S. (2016). Definition and classification of fault damage zones: A review and a new methodological approach. *Earth-Science Reviews*, *152*, 70–87.
- Chung, A. I., Henson, I., & Allen, R. M. (2019). Optimizing earthquake early warning performance: ElarmS-3. *Seismological Research Letters*, *90*(2A), 727–743.

- Clinton J., Zollo, A., Marmureanu, A., Zulfikar, C., & Parolai, S. (2016). State-of-the art and future of earthquake early warning in the European region. *Bulletin of Earthquake Engineering*, 14(9), 2441–2458.
- Cochran, E. S., Li, Y. G., Shearer, P. M., Barbot, S., Fialko, Y., & Vidale, J. E. (2009). Seismic and geodetic evidence for extensive, long-lived fault damage zones. *Geology*, 37(4), 315–318.
- Collings, R., Lange, D. A., Rietbrock, F., Tilmann, D., Natawidjaja, B., Suwargadi, B., et al. (2012). Structure and seismogenic properties of the Mentawai segment of the Sumatra subduction zone revealed by local earthquake traveltime tomography. *Journal of Geophysical Research: Solid Earth*, 117(B1), 1–23.
- Crowell, B. W., Melgar, D., Bock, Y., Haase, J. S., & Geng, J. (2013). Earthquake magnitude scaling using seismogeodetic data. *Geophysical Research Letters*, 40(23), 6089–6094.
- Crowell, B. W., Schmidt, D. A., Bodin, P., Vidale, J. E., Gomberg, J., Hartog, J. R., Kress, V., Melbourne, T., Santillan, M., Minson, S. E., & Jamison, D. (2016). Demonstration of the Cascadia G-FAST geodetic earthquake early warning system for the Nisqually, Washington, earthquake. *Seismological Research Letters*, 87(4), 930–943.
- Cua, G., Fischer, M., Heaton, T., & Wiemer, S. (2009). Real-time performance of the virtual seismologist earthquake early warning algorithm in Southern California. *Seismological Research Letters*, 80(5), 740–747.
- Cuéllar, A., Suarez, G., & Espinosa-Aranda, J. (2018). A fast earthquake early warning algorithm based on the first 3 s of the P-wave coda. *Bulletin of the Seismological Society of America*, 108(4), 2068–2079.
- Dalton, C. A., Ekström, G., & Dziewonski, A. M. (2009). Global seismological shear velocity and attenuation: A comparison with experimental observations. *Earth and Planetary Science Letters*, 284(1–2), 65–75.
- Darragh, R., Wong, I., Silva, W. (2019). Evaluating kappa, Q(f), and stress parameter in the Southern Rocky Mountains of Central Colorado. *Bulletin of the Seismological Society of America*, 109(2), 586–599.
- Dashti, S., & Bray, J. D. (2012). Numerical simulation of building response on liquefiable sand. *Journal of Geotechnical and Geoenvironmental Engineering*, 139(8), 1235–1249.

- Dawood, H. M., & Rodriguez-Marek, A. (2013). A method for including path effects in ground-motion prediction equations: An example using the Mw 9.0 Tohoku earthquake aftershocks. *Bulletin of the Seismological Society of America*, *103*(2B), 1360–1372.
- Dong-Hoon, S., Park, J.-H., Chi, H.-C., Hwang, E.-H., Lim, I.-S., Seong, Y. J., & Pak, J. (2017). The first stage of an earthquake early warning system in South Korea. *Seismological Research Letters*, *88*(6), 1491–1498.
- Douglas, J. (2006). Ground-motion prediction equations for southern Spain and Southern Norway obtained using the composite model perspective. *Journal of Earthquake Engineering*, *10*(1), 33–72.
- Douglas, J., & Edwards, B. (2016). Recent and future developments in earthquake ground motion estimation. *Earth-Science Reviews*, *160*, 203–219.
- Douglas, J., Gehl, P. L., Bonilla, F., & Gélis, C. (2010). A κ model for mainland France. *Pure and Applied Geophysics*, *167*(11), 1303–1315.
- Drouet, S., Cotton, F., & Guéguen, P. (2010). V_{S30} , κ , regional attenuation and Mw from accelerograms: Application to magnitude 3–5 French earthquakes. *Geophysical Journal International*, *182*(2), 880–898.
- Eberhart-Phillips, D. (2016). Northern California seismic attenuation: 3D QP and QS models. *Bulletin of the Seismological Society of America*, *106*(6), 2558–2573.
- Eberhart-Phillips, D., & Banister, S. (2015). 3-D imaging of the northern Hikurangi Subduction Zone, New Zealand: Variations in subducted sediment, slab fluids and slow slip. *Geophysical Journal International*, *201*(2), 838–855.
- Eberhart-Phillips, D., Thurber, C., & Fletcher, J. B. (2014). Imaging P and S attenuation in the Sacramento–San Joaquin Delta Region, Northern California. *Bulletin of the Seismological Society of America*, *104*(5), 2322–2336.
- Edwards, B., Fäh, D., & Giardini, D. (2011). Attenuation of seismic shear wave energy in Switzerland. *Geophysical Journal International*, *185*(2), 967–984.
- Edwards, B., Ktenidou, O. J., Cotton, F., Abrahamson, N., Van Houtte, C., & Fäh, D. (2015). Epistemic uncertainty and limitations of the κ_0 model for near-surface attenuation at hard rock sites. *Geophysical Journal International*, *202*(3), 1627–1645.

- Erdem, J. E., Boatwright, J., & Fletcher, J. B. (2019). Ground-motion attenuation in the Sacramento–San Joaquin Delta, California, from 14 Bay Area Earthquakes, including the 2014 M 6.0 South Napa Earthquake. *Bulletin of the Seismological Society of America*, *109*(3), 1025–1033.
- Faccioli, E., Maggio, F., Paolucci, R., & Quarteroni, A. (1997). 2D and 3D elastic wave propagation by a pseudo-spectral domain decomposition method. *Journal of Seismology*, *1*, 237–251.
- Fadugba, O. I., Sahakian, V. J., Melgar, D., Rodgers, A., & Shimony, R. (2023). *The impact of the three-dimensional structure of a subduction zone on time-dependent crustal deformation measured by HR-GNSS*. EarthArXiv.
- Farahbod, A. M., Calvert, A. J., Cassidy, J. F., & Brillon, C. (2016). Coda Q in the Northern Cascadia Subduction Zone. *Bulletin of the Seismological Society of America*, *106*(5), 1939–1947.
- FEMA (2022). *Earthquake-resistant design concepts*. Report P-749. Federal Emergency Management Agency. https://www.fema.gov/sites/default/files/documents/fema-p-749-earthquake-resistant-design-concepts_112022.pdf
- Fernández, A. I., Castro, R. R., & Huerta, C. I. (2010). The spectral decay parameter kappa in Northeastern Sonora, Mexico. *Bulletin of the Seismological Society of America*, *100*(1), 196–206.
- Fletcher, J. B., & Boatwright, J. (2020). Peak ground motions and site response at Anza and Imperial Valley, California. *Pure and Applied Geophysics*, *177*, 2753–2769.
- Flinn, E. A. (1965). Signal analysis using rectilinearity and direction of particle motion. In *Proceedings of the IEEE*, *53*(12), 1874–1876.
- Frankel, A. (1991). Mechanisms of seismic attenuation in the crust: scattering and anelasticity in New York State, South Africa, and Southern California. *Journal of Geophysical Research*, *96*(B4), 6269–6289.
- Fu, L., & Li, X.-J. (2017). The kappa (κ_0) model of the Longmenshan region and its application to simulation of strong ground-motion by the Wenchuan M_S 8.0 earthquake. *Chinese Journal of Geophysics*, *60*(8), 2935–2947.
- Fukao, Y. (1979). Tsunami earthquakes and subduction processes near deep-sea trenches. *Journal of Geophysical Research*, *84*(B5), 2303–2314.

- Gentili, S., & Franceschina, G. (2011). High frequency attenuation of shear waves in the southeastern Alps and northern Dinarides. *Geophysical Journal International*, 185(3), 1393–1416. x
- Goldberg, D. E., & Melgar, D. (2020). Generation and validation of broadband synthetic P waves in semistochastic models of large earthquakes. *Bulletin of the Seismological Society of America*, 110(4), 1982–1995.
- Goldberg, D. E., Melgar, D., & Brock, Y. (2019). Seismogeodetic P-wave amplitude: No evidence for strong determinism. *Geophysical Research Letters*, 46(20), 11118–11126.
- Goldberg, D. E., Melgar, D., Bock, Y., & Allen, R. M. (2018). Geodetic observations of weak determinism in rupture evolution of large earthquakes. *Journal of Geophysical Research: Solid Earth*, 123, 9950–9962.
- Goldberg, D. E., Melgar, D., Hayes, G. P., Crowell, B. W., & Sahakian, V. J. (2021). A ground-motion model for GNSS peak ground displacement. *Bulletin of the Seismological Society of America*, 111(5), 2393–2407.
- Gordon, R. B., & Davis, L. A. (1968). Velocity and attenuation of seismic waves in imperfectly elastic rock. *Journal of Geophysical Research*, 73(12), 3917–3935.
- Grapenthin, R., Johanson, I., & Allen, R. M. (2014a). The 2014 MW 6.0 Napa earthquake, California: Observations from real-time GPS-enhanced earthquake early warning. *Geophysical Research Letters*, 41(23), 8269–8276.
- Grapenthin, R., Johanson, I. A., & Allen, R. M., (2014b). Operational real-time GPS-enhanced earthquake early warning. *Journal of Geophysical Research: Solid Earth*, 119(10), 7944–7965.
- Graves, R. W., & Pitarka, A. (2004). Broadband time history simulation using a hybrid approach. In *Proceedings of the 13th World Conference on Earthquake Engineering*. International Association for Earthquake Engineering. Vancouver, British Columbia, Canada, Paper 1098.
- Graves, R. W., & Pitarka, A. (2010). Broadband ground-motion simulation using a hybrid approach. *Bulletin of the Seismological Society of America*, 100(5A), 2095–2123.
- Graves, R., & Pitarka, A. (2015). Refinements to the Graves and Pitarka (2010) broadband ground-motion simulation method. *Seismological Research Letters*, 86(1), 75–80.

- Haendel, A., Anderson, J. G., Pilz, M., & Cotton, F. (2020). A frequency-dependent model for the shape of the Fourier amplitude spectrum of acceleration at high frequencies. *Bulletin of the Seismological Society of America*, 110(6), 2742–2754.
- Hanks, T. C. (1982). Fmax. *Bulletin of the Seismological Society of America*, 72(6), 1867–1879.
- Hanks, T. C., & Kanamori, H. (1979). A moment magnitude scale. *Journal of Geophysical Research: Solid Earth*, 84(B5), 2348–2350.
- Hanks, T. C., & Thatcher, W. (1972). A graphical representation of seismic source parameters. *Journal of Geophysical Research*, 77(23), 4393–4405.
- Harris, R. A., Barall, M., Aagaard, B., Ma, S., Roten, D., Olsen, K., Duan, B., Liu, D., Luo, B., Bai, K., Ampuero, J.-P., Kaneko, Y., Gabriel, A.-A., Duru, K., Ulrich, T., Wollherr, S., Shi, Z., Dunham, E., Bydlon, S., Zhang, Z., Chen, X., Nadh Somala, S., Pelties, C., Tago, J., Cruz-Atienza, V. M., Kozdon, J., Daub, E., Aslam, K., Kase, Y., Withers, K., & Dalguer, L. (2018). A suite of exercises for verifying dynamic earthquake rupture codes. *Seismological Research Letters*, 89(3), 1146–1162.
- Hartog, J. R., Kress, V. C., Malone, S. D., Bodin, P., Vidale, J. E., & Crowell, B. W. (2016). Earthquake early warning: ShakeAlert in the Pacific Northwest. *Bulletin of the Seismological Society of America*, 106(4), 1875–1886.
- Hassani, B., & Atkinson, G. M. (2018). Adjustable generic ground-motion prediction equation based on equivalent point-source simulations: Accounting for kappa effects. *Bulletin of the Seismological Society of America*, 108(2), 913–928.
- Hayes, G. (2018). Slab2 - A Comprehensive Subduction Zone Geometry Model: U.S. Geological Survey data release.
- Heath, D., Wald, D. J., Worden, C. B., Thompson, E. M., & Scmocyk, G. (2020). A Global hybrid V_{S30} map with a topographic-slope-based default and regional map insets. *Earthquake Spectra*, 36(3), 1570–1584.
- Heaton, T. H. (1985). A model for a seismic computerized alert network, *Science*, 228, 987–990.
- Hill, E. M., Borrero, J. C., Huang, Z., Qiu, Q., Banerjee, P., Natawidjaja, D. H., et al. (2012). The 2010 MW 7.8 Mentawai earthquake: Very shallow source of a rare

- tsunami earthquake determined from tsunami field survey and near-field GPS data. *Journal of Geophysical Research*, 117(B06402), 1–21.
- Hilmo, R., & Wilcock, W. S. D. (2020). Physical sources of high-frequency seismic noise on Cascadia Initiative ocean bottom seismometers. *Geochemistry, Geophysics, Geosystems*, 21(10), 1–20.
- Hirshorn, B. & Weinstein, S. (2011). Earthquake source parameters, rapid estimates for tsunami warning, In R. F. Myers (Ed.), *Extreme Environmental Events: Complexity in Forecasting and Early Warning* (pp. 447–461). New York City, NY: Springer.
- Holzer, T. L., Padovani, A. C., Bennett, M. J., Noce, T. E., & Tinsley, J. C. (2005). Mapping NEHRP V_{S30} site classes. *Earthquake Spectra*, 21(2), 353–370.
- Hough, S. E. (1990). Constraining sediment thickness in the San Francisco Bay Area using observed resonances and P-to-S conversions. *Geophysical Research Letters*, 17(9), 1469–1472.
- Hough, S. E., & Anderson, J. G. (1988). High-frequency spectra observed at Anza, California: Implications for Q structure (USA). *Bulletin of the Seismological Society of America*, 78(2), 692–707.
- Hough, S. E., Lees, J. M., & Monastero, F. (1999). Attenuation and source properties at the Coso Geothermal Area, California. *Bulletin of the Seismological Society of America*, 89(6), 1606–1619.
- Huang, M. W., Wen, K. L., Chang, S. C., Chang, C. L., Liu, S. Y., & Chen, K. P. (2017). The High-Cut Parameter (κ) for the Near-Surface Geology in and around the Taipei Basin, Taiwan. *Bulletin of the Seismological Society of America*, 107(3), 1254–1264.
- Hubbard, J., Barbot, S., Hill, E. M., & Tapponnier, P. (2015). Coseismic slip on shallow décollement megathrusts: Implications for seismic and tsunami hazard. *Earth-Science Reviews*, 141, 45–55.
- Humphrey, J. R., & Anderson, J. G. (1992). Shear-wave attenuation and site response in Guerrero, Mexico. *Bulletin of the Seismological Society of America*, 81(4), 1622–1645.
- IRIS. (n.d.). *Community near-fault observatory*.
https://www.iris.edu/hq/initiatives/near_fault (Accessed 9 November 2022).

- Ito, E., Nakano, K., Nagashima, F., & Kawase, H. (2020). A method to directly estimate S-wave site amplification factor from horizontal-to-vertical spectral ratio of earthquakes (eHVSRS). *Bulletin of the Seismological Society of America*, *110*(6), 2892–2911.
- Jachens, R. C., Simpson, R. W., Graymer, R. W., Wentworth, C. M., & Brocher, T. M. (2006). Three-dimensional geologic map of northern and central California: A basic model for supporting earthquake simulations and other predictive modeling (abstract). *Seismological Research Letters*, *77*, 270.
- Jeppson, T. N., Tobin, H. J., & Hashimoto, Y. (2018). Laboratory measurements quantifying elastic properties of accretionary wedge sediments: Implications for slip to the trench during the 2011 Mw 9.0 Tohoku-Oki earthquake. *Geosphere*, *14*(4), 1411–1424.
- Ji, C., Cabas, A., Cotton, F., Pilz, M., & Bindi, D. (2020). Within-station variability in kappa: Evidence of directionality effects. *Bulletin of the Seismological Society of America*, *110*(3), 1247–1259.
- Ji, C., Cabas, A., Bonilla, F., & Gélis, C. (2022). Variability in kappa (kr): Contributions from the computation procedure. In *Geo-Congress 2022*. Charlotte, North Carolina, 20–23 March.
- Ji, C., Cabas, A., Bonilla, L. F., & Gélis, C. (2021). Effects of nonlinear soil behavior on kappa (k): Observations from the KiK-Net database. *Bulletin of the Seismological Society of America*, *111*(4), 2138–2157.
- Johnson, C. W., Kilb, D., Baltay, A., & Vernon, F. (2020). Peak ground velocity spatial variability revealed by dense seismic array in southern California. *Journal of Geophysical Research: Solid Earth*, *125*(6), 1–17.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME—Journal of Basic Engineering*, *82*(D), 35–45.
- Kanamori H. (1972). Mechanism of tsunami earthquakes. *Physics of the Earth and Planetary Interiors*, *6*(5), 346–359.
- Kanamori, H., and M. Kikuchi (1993). The 1992 Nicaragua earthquake: A slow tsunami earthquake associated with subducted sediments. *Nature*, *361*, 714–716.
- Kilb, D., Biasi, G., Anderson, J., Brune, J., Peng, Z., & Vernon F. L. (2012). A comparison of spectral parameter kappa from small and moderate earthquakes

- using Southern California ANZA seismic network data. *Bulletin of the Seismological Society of America*, 102(1), 284–300.
- King, E., Klimasewski, A., Sahakian, V. J., & Baltay, A. S. (2019). Source and site spectral inversion for κ_0 computation in the Bay Area. Abstract (001) presented at 2019 SCEC Annual Meeting, 8–11 September.
- Klimasewski, A., Sahakian, V., & Thomas, A. (2021). Comparing artificial neural networks with traditional ground-motion models for small-magnitude earthquakes in southern California. *Bulletin of the Seismological Society of America*, 111(3), 1577–1589.
- Klimasewski, A., Sahakian, V., Baltay, A., Boatwright, J., Fletcher, J. B., & Baker, L. M. (2019). κ_0 and broadband site spectra in southern California from source model-constrained inversion. *Bulletin of the Seismological Society of America*, 109(5), 1878–1889.
- Kneib, G., & Shapiro, S. A. (1992). Attenuation by scattering: Theory and numerical results. In *SEG Technical Program Expanded Abstracts 1992*. Society of Exploration Geophysicists.
- Kodera, Y. (2018). Real-time detection of rupture development: Earthquake early warning using P waves from growing ruptures. *Geophysical Research Letters*, 45(1), 156–165.
- Kohler, M. D., Cochran, E. S., Given, D., Guiwits, S., Nauhauser, D., Henson, I., Hartog, R., Bodin, P., Kress, V., Thompson, S., Felizardo, C., Brody, J., Bhadha, R., & Schwarz, S. (2018). Earthquake early warning ShakeAlert system: West coast wide production prototype. *Seismological Research Letters*, 89(1), 99–107.
- Kohler, M. D., Smith, D. E., Andrews, J., Chung, A. I., Hartog, R., Henson, I., Given, D. D., de Groot, R., & Guiwits, S. (2020). Earthquake early warning ShakeAlert 2.0: Public rollout. *Seismological Research Letters*, 91(3), 1763–1775.
- Komatitsch, D., & Tromp, J. (1999). Introduction to the spectral-element method for 3-D seismic wave propagation. *Geophysical Journal International*, 139(3), 806–822.
- Krischer, L. (2016). *mtspec Python wrappers (v0.3.2)* [Data set]. Zenodo.
- Krischer, L., Megies, T., Barsch, R., Beyreuther, M., Lecocq, T., Caudron, C., & Wassermann, J. (2015). *Computational Science & Discovery*, 8(1), 1–17.

- Ktenidou O.-J., Drouet, S., Theodulidis, N., Chaljub, M., Arnaouti, S., & Cotton, F. (2012). Estimation of kappa (κ) for a sedimentary basin in Greece (EUROSEISTEST) - Correlation to site characterisation parameters. In *Proceedings of the 15th World Conference of Earthquake Engineering*. Lisbon, Portugal, 4–28 September.
- Ktenidou, O.-J. (2022). Hard as a Rock? Reconsidering Rock-Site Seismic Response and Reference Ground Motions. In Vacareanu, R., Ionescu, C. (Eds.), *Progresses in European Earthquake Engineering and Seismology*. ECEES 2022. Springer Proceedings in Earth and Environmental Sciences. Springer, Cham.
- Ktenidou, O.-J., & Abrahamson, N. A. (2016). *A methodology for the estimation of kappa (κ) from large datasets: Example application to rock sites in the NGA-East database and implications on design motions*, PEER Report 2016/01, Pacific Earthquake Engineering Research Center, University of California, Berkeley, California.
- Ktenidou, O.-J., & Van Houtte, C. (2012). *Empirical Estimation of kappa from Rock Velocity Profiles at the Swiss NPP Sites (19.02.2012)*. Report TP2-TB-1090. PEGASOS Refinement Project.
- Ktenidou, O.-J., Abrahamson, N. A., Drouet, S., & Cotton, F. (2015). Understanding the physics of kappa (κ): Insights from a downhole array. *Geophysical Journal International*, 203(1), 678–691.
- Ktenidou, O.-J., Cotton, F., Abrahamson, N. A., & Anderson, J. G. (2014). Taxonomy of κ : A review of definitions and estimation approaches targeted to applications. *Seismological Research Letters*, 85(1), 135–146.
- Ktenidou, O.-J., Gélis, C., & Bonilla, L. F. (2013). A study on the variability of kappa (κ) in a Borehole: Implications of the computation process. *Bulletin of the Seismological Society of America*, 103(2A), 1048–1068.
- Kumar, S., Kumar, D., Rastogi, B. K., & Singh, A. P. (2018). Kappa (κ) model for Kachchh region of Western India. *Geomatics, Natural Hazards and Risk*, 9(1), 442–455.
- Kumar, V., Chopra, S., Choudhury, P., & Kumar, D. (2020). Estimation of near surface attenuation parameter kappa (κ) in Northwest and Northeast Himalaya region. *Soil Dynamics and Earthquake Engineering*, 136, 106237.

- Kusanovic, D. S., Seylabi, E. E., Ayoubi, P., Nguyen, K. T., Garcia-Suarez, G., Kottke, A. R., & Asimaki, D. (2023). Seismo-VLAB: An Open-Source Software for Soil–Structure Interaction Analyses. *Mathematics*, *11*(21), 4530.
- Kuyuk, H. S., & Allen, R. M. (2013). A global approach to provide magnitude estimates for earthquake early warning alerts. *Geophysical Research Letters*, *40*(24), 6329–6333.
- Lai, T. S., Mittal, H., Chao, W. A., & Wu, Y. M. (2016). A study on kappa value in Taiwan using borehole and surface seismic array. *Bulletin of the Seismological Society of America*, *106*(4), 1509–1517.
- Landwehr, N., Kuehn, N. M., Scheffer, T., & Abrahamson, N. (2016). A nonergodic ground-motion model for California with spatially varying coefficients. *Bulletin of the Seismological Society of America*, *106*(6), 2574–2583.
- Lanzano, G., D’Amico, M., Felicetta, C., Puglia, R., Luzi, L., Pacor, F., & Bindi, D. (2016). Ground-motion prediction equations for region-specific probabilistic seismic-hazard analysis. *Bulletin of the Seismological Society of America*, *106*(1), 73–92.
- Laurendeau, A., Cotton, F., Ktenidou, O.-J., Bonilla, L. F., & Hollender, F. (2013). Rock and stiff-soil site amplification: Dependency on V_{S30} and kappa (κ_0). *Bulletin of the Seismological Society of America*, *103*(6), 3131–3148.
- Lavrentiadis, G., Abrahamson, N. A., Nicolas, K. M., Bozorgnia, Y., Goulet, C. A., Babič, A., Macedo, J., Dolšek, M., Gregor, M., Kottke, A. R., Lacour, M., Liu, C., Meng, X., Phung, V.-B., Sung, C.-H., & Walling, M. (2023). Overview and introduction to development of non-ergodic earthquake ground-motion models. *Bulletin of Earthquake Engineering*, *21*, 5121–5150.
- Lay, T., & Bilek, S. L. (2007). Anomalous earthquake ruptures at shallow depths on subduction zone megathrusts. In T. H. Dixon & C. Moore (Eds.), *The seismogenic zone of subduction thrust faults* (pp. 476–511). New York City, NY: Columbia University Press.
- Lay, T., Ammon, C. J., Kanamori, H., Yamazaki, Y., Cheung, K. F., & Hutko, A. R. (2011). The 25 October 2010 Mentawai tsunami earthquake (MW 7.8) and the tsunami hazard presented by shallow megathrust ruptures. *Geophysical Research Letters*, *38*(L06302), 1–5.

- Lay, T., Kanamori, H., Ammon, C. J., Koper, K. D., Hutko, A. R., Ye, L., et al. (2012). Depth-varying rupture properties of subduction zone megathrust faults. *Journal of Geophysical Research*, *117*(B04311), 1–21.
- Lees, J., & Lindley, G. (1994). Three-dimensional attenuation tomography at Loma Prieta: Inverting t^* for Q . *Journal of Geophysical Research*, *99*(B4), 6843–6863.
- Lermo, J., & Chávez-García, F. J. (1993). Site effect evaluation using spectral ratios with only one station. *Bulletin of the Seismological Society of America*, *83*(5), 1574–1594.
- Lermo, J., Santoyo, M. A., Jaimes, M. A., Antayhua, Y., & Chavacán, M. (2016). Local earthquakes of the Mexico Basin in Mexico City: κ , Q , source spectra, and stress drop. *Bulletin of the Seismological Society of America*, *106*(4), 1423–1437.
- LeVeque, R. J., Waagan, K., González, F. I., Rim, D., & Lin, G. (2016). Generating random earthquake events for probabilistic tsunami hazard assessment. *Pure and Applied Geophysics*, *173*, 3671–3692.
- Li, Y.-G., & Teng, T.-L. (1990). Observation and analysis of trapped modes in the San Andreas Fault Zone. In *SEG Technical Programs Extended Abstracts 1990*. Society of Exploration Geophysicists.
- Li, Y.-G., Catchings, R. D., & Goldman, M. R. (2016). Subsurface fault damage zone of the 2014 MW 6.0 south Napa, California, earthquake viewed from fault-zone trapped waves. *Bulletin of the Seismological Society of America*, *106*(6), 2747–2763.
- Li, Y.-G., Vidale, J. E., & Cochran, E. S. (2004). Low-velocity damaged structure of the San Andreas Fault at Parkfield from fault zone trapped waves. *Geophysical Research Letters*, *31*(12), 1–5.
- Lin, J.-T., Melgar, D., Thomas, A. M., & Searcy, J. (2021). Early warning for great earthquakes from characterization of crustal deformation patterns with deep learning. *Journal of Geophysical Research: Solid Earth*, *126*(10), 1–17.
- Lotto, G. C., Dunham, E. M., Jeppson, T. N., & Tobin, H. J. (2017). The effect of compliant prisms on subduction zone earthquakes and tsunamis. *Earth and Planetary Science Letters*, *458*, 213–222.

- Ma, S., & Nie, S. (2019). Dynamic wedge failure and along-arc variations of tsunamigenesis in the Japan Trench margin. *Geophysical Research Letters*, *46*(15), 8782–8790.
- Madin, I. P., & Burns, W. J. (2013). *Map of earthquake and tsunami damage potential for a simulated magnitude 9 Cascadia earthquake*. Report O-13-06. State of Oregon Department of Geology and Mineral Industries.
<https://www.wsdot.wa.gov/research/reports/fullreports/896-4.pdf>
- Maechling, P. J., Silva, F., Callaghan, S., & Jordan, T. H. (2015). SCEC Broadband Platform: System architecture and software implementation. *Seismological Research Letters*, *86*(1).
- Mai, P. M., Imperatori, W., & Olsen, K. B. (2010). Hybrid broadband ground-motion simulations: Combining long-period deterministic synthetics with high-frequency multiple S-to-S backscattering. *Bulletin of the Seismological Society of America*, *100*(5A), 2124–2142.
- Mai., P. M., & Beroza, G. C. (2002). A spatial random field model to characterize complexity in earthquake slip. *Journal of Geophysical Research*, *107*(B11), 10-1–10-21.
- Marafi, N. A., Eberhard M. O., Berman, J. W., Wirth, E. A., & Frankel, A. D. (2019). Impacts of Simulated M9 Cascadia Subduction Zone Motions on Idealized Systems. *Earthquake Spectra*, *35*(3), 1261–1287.
- Mazzieri, I., Stupazzini, M., Guidotti, R., & Smerzini, C. (2013). SPEED: SPectral Elements in Elastodynamics with Discontinuous Galerkin: a non-conforming approach for 3D multi-scale problems. *International Journal for Numerical Methods in Engineering*, *95*(12), 991–1010.
- McCallen, D., Petrone, F., Miah, M., Pitarka, A., Rodgers, A., & Abrahamson, N. (2021). EQSIM—A multidisciplinary framework for fault-to-structure earthquake simulations on exascale computers, part II: Regional simulations of building response. *Earthquake Spectra*, *37*(2), 736–761.
- McPhillips, D. F., Herrick, J. A., Ahdi, S., Yong, A. K., & Haefner, S. (2020). Updated compilation of VS30 data for the United States: U.S. Geological Survey data release.

- Meier, M.-A. (2017). How “good” are real- time ground motion predictions from earthquake early warning systems?. *Journal of Geophysical Research: Solid Earth*, 122(7), 5561–5577.
- Melgar, D. (2021). Was the January 26th, 1700 Cascadia earthquake part of a rupture sequence?. *Journal of Geophysical Research: Solid Earth*, 126(10), 1–22.
- Melgar, D., & Hayes, G. P. (2017). Systematic observations of the slip pulse properties of large earthquake ruptures. *Geophysical Research Letters*, 44(19), 9691–9698.
- Melgar, D., Bock, Y., Sanchez, D., & Crowell, B. W. (2013). On robust and reliable automated baseline corrections for strong motion seismology. *Journal of Geophysical Research: Solid Earth*, 118(3), 1177–1187.
- Melgar, D., Crowell, B. W., Geng, J., Allen, R. M., Bock, Y., Riquelme, S., et al. (2015). Earthquake magnitude calculation without saturation from the scaling of peak ground displacement. *Geophysical Research Letters*, 42(13), 5197–5205.
- Melgar, D., Crowell, B. W., Melbourne, T. I., Szeliga, W., Santillan, M., & Scriver, C. (2020). Noise characteristics of operational real-time high-rate GNSS positions in a large aperture network. *Journal of Geophysical Research: Solid Earth*, 125(7), 1–15.
- Melgar, D., Leveque, R. J., Dreger, D. S., & Allen, R. M. (2016). Kinematic rupture scenarios and synthetic displacement data: An example application to the Cascadia subduction zone. *Journal of Geophysical Research: Solid Earth*, 121(9), 6658–6674.
- Minson, S. E., Murray, J. R., Langbein, J. O., & Gomberg, J. S. (2014). Real-time inversions for finite fault slip models and rupture geometry based on high-rate GPS data. *Journal of Geophysical Research: Solid Earth*, 119(4), 3201–3231.
- Mittal, H., Sharma, B., Sandhu, M., & Kumar, D. (2021). Spatial distribution of high-frequency spectral decay factor kappa (κ) for Delhi, India. *Acta Geophysica*, 69, 2113–2127.
- Mooney, M. D., & Wang, H. (2014). Seismic intensities, PGA, and PGV for the 20 April 2013, Mw 6.6 Lushan, China, earthquake, and a comparison with North America. *Seismological Research Letters*, 85(5), 1034–1042.

- Mori, J., Mooney, W. D., Afnimar, Kurniawan, S., Anaya, A. I., & Widiyantoro, S. (2007). The 17 July 2006 tsunami earthquake in west Java, Indonesia. *Seismological Research Letters*, 78(2), 201–207.
- Motazedian, D., & Atkinson, G. M. (2005). Stochastic finite-fault modeling based on a dynamic corner frequency. *Bulletin of the Seismological Society of America*, 95(3), 995–1010.
- Nakamura, Y. (1988). On the urgent earthquake detection and alarm system (UrEDAS). In *Proceedings of the 9th World Conference on Earthquake Engineering*. Tokyo-Kyoto, Japan, 2–9 August.
- Neighbors, C., Liao, E. J., Cochran, E. S., Funning, G. J., Chung, A. I., Lawrence, J. F., Christensen, C., Miller, M., Belmonte, A., & Andrés Sepulveda, H. H. (2015). Investigation of the high-frequency attenuation parameter, κ (kappa), from aftershocks of the 2010 M_W 8.8 Maule, Chile earthquake. *Geophysical Journal International*, 200(1), 200–215.
- Newman A. V., & Okal, E. A. (1998). Teleseismic estimates of radiated seismic energy: The E/M0 discriminant for tsunami earthquakes. *Journal of Geophysical Research*, 103(B11), 26885–26898.
- Newman, A. V., Hayes, G., Wei, Y., & Convers, J. (2011). The 25 October 2010 Mentawai tsunami earthquake, from real-time discriminants, finite-fault rupture, and tsunami excitation. *Geophysical Research Letters*, 38(L05302), 1–7.
- Northern California Earthquake Data Center. (2014). *Berkeley Digital Seismic Network (BDSN)* [Data set]. Northern California Earthquake Data Center.
- Nye, T. A., Sahakian, V. J., King, E., Baltay, A., & Klimasewski, A. (2022a). Estimates of kappa in the San Francisco Bay Area. In *Proceedings of the 12th National Conference in Earthquake Engineering*. Salt Lake City, Utah USA, 27 June–1 July.
- Nye, T., Sahakian, V. J., King, E., Baltay, A., & Klimasewski, A. (2022b). *San Francisco Bay Area kappa* [Data set]. Zenodo.
- Olsen, K. B., Day, S. M., Minster, J. B., Cui, Y., Chourasia, A., Faerman, M., Moore, R., Maechling, P., & Jordan, T. (2006). Strong shaking in Los Angeles expected from southern San Andreas earthquake. *Geophysical Research Letters*, 33(7), 2–5.

- Ostermeijer, G. A., Mitchell, T. M., Aben, F. M., Dorsey, M. T., Browning, J., Rockwell, T. K., Fletcher, J. M., & Ostermeijer, F. (2020). Damage zone heterogeneity on seismogenic faults in crystalline rock; A field study of the Borrego Fault, Baja California. *Journal of Structural Geology*, *137*, 104016.
- Pacific Earthquake Engineering Research Center. (2013). *NGA-West2 database flatfile* [Data set]. UC Berkeley.
https://apps.peer.berkeley.edu/assets/NGA_West2_supporting_data_for_flatfile.zip
- Pagani, M., Monelli, D., Weatherill, G., Danciu, L., Crowley, H., Silva, V., Henshaw, P., Butler, L., Nastasi, M., Panzeri, L., Simionato, M., & Vigano D. (2014). OpenQuake Engine: An open hazard (and risk) software for the global earthquake model. *Seismological Research Letters*, *85*(3), 692–702.
- Palmer, S. M., & Atkinson, G. M. (2020). The high-frequency decay slope of spectra (κ) for $M \geq 3.5$ earthquakes on rock sites in eastern and western Canada. *Bulletin of the Seismological Society of America*, *110*(2), 471–488.
- Palmer, S. M., & Atkinson, G. M. (2023). Uncertainties in broadband determination of the high-frequency spectral decay, κ , in Eastern Canada. *Bulletin of the Seismological Society of America*, *113*(6), 2666–2688.
- Papageorgiou, A. S., & Aki, K. (1983). A specific barrier model for the quantitative description of inhomogeneous faulting and the prediction of strong ground motion. I. Description of the model. *Bulletin of the Seismological Society of America*, *73*(3), 693–722.
- Park, S. J., Lee, J. M., & Baag, C. E. (2020). Estimation of high-frequency spectral decay (κ) for the Gyeongju area in south Korea. *Bulletin of the Seismological Society of America*, *110*(1), 295–311.
- Parker, G. A., Stewart, J. P., Boore, D. M., Atkinson, G. A., & Hassani, B. (2020). *NGA-Subduction Global Ground-Motion Models with Regional Adjustment Factors*. PEER Report 2020/03. Pacific Earthquake Engineering Research Center, University of California, Berkeley, California.
- Parker, G. A., Baltay, A. S., Rekoske, J., & Thompson, E. M. (2020). Repeatable source, path, and site effects from the 2019 M 7.1 ridgecrest earthquake sequence. *Bulletin of the Seismological Society of America*, *110*(4), 1530–1548.
- Parolai, S., & Bindi, D. (2004). Influence of soil-layer properties on κ evaluation. *Bulletin of the Seismological Society of America*, *94*(1), 349–356.

- Pavel, F., & Vacareanu, R. (2015). Kappa and regional attenuation for Vrancea (Romania) earthquakes. *Journal of Seismology*, *19*, 791–799.
- Pedregosa, F., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830.
- Perron, V., Hollender, F., Bard, P. Y., Gélis, C., Guyonnet-Benaize, C., Hernandez, B., & Ktenidou, O. J. (2017). Robustness of kappa (κ) measurement in low-to-moderate seismicity areas: Insight from a site-specific study in Provence, France. *Bulletin of the Seismological Society of America*, *107*(5), 2272-2292.
- Petersen, M. D., Shumway, A. M., Powers, P. M., Mueller, S. C., Moschetti, M. P., Frankel, A. D., Rezaeian, S., McNamara, D. E., Luco, N., Boyd, O. S., Rukstales, K. S., Jaiswal, K. S., Thompson, E. M., Hoover, S. M., Clayton, B. S., Field, E. H., & Zeng, Y. (2020). The 2018 update of the US National Seismic Hazard Model: Overview of model and implications. *Earthquake Spectra*, *36*(1), 5–41.
- Pilz, M., Cotton, F., Zaccarelli, R., & Bindi, D. (2019). Capturing regional variations of hard-rock attenuation in Europe. *Bulletin of the Seismological Society of America*, *109*(4), 1401–1418.
- Powers, P. M., & Jordan, T. H. (2010). Distribution of seismicity across strike-slip faults in California. *Journal of Geophysical Research: Solid Earth*, *115*(B5).
- Pozo, I., Montalva, G., & Miller, M. (2022). Assessment of kappa values in the Chilean Subduction Zone for interface and in-slab events. *Seismological Research Letters*, *94*(1), 385–398.
- Prieto, G. A., Parket, R. L., & Vernon, F. L. (2009). A Fortran 90 library for multitaper spectrum analysis. *Computers and Geosciences*, *35*(8), 1701–1710.
- Pujades, L. G., Ugalde, U., Canas, J. A., Navarro, M., Badal, F. J., & Corchete, V. (1996). Intrinsic and scattering attenuation from observed seismic codas in the Almeria Basin (Southeastern Iberian Peninsula). *Geophysical Journal International*, *129*(2), 281–291.
- Rietbrock A., & Edwards B. (2017). *Wylfa Seismic Hazard Consultancy Project*, University of Liverpool, Leo Ref: 1619&2082, Version 2.3 (Revision June 2017).

- Riquelme, S., Schwarze, H., Fuentes, M., & Campos, J. (2020). Near-field effects of earthquake rupture velocity into tsunami runup heights. *Journal of Geophysical Research Solid Earth*, *125*(6), 1–18.
- Rodgers, A. J., Pitarka, A., Pankajakshan, R., Sjögreen, B., & Petersson, N. A. (2020). Regional-scale 3D ground-motion simulations of MW 7 earthquakes on the Hayward fault, Northern California resolving frequencies 0–10 Hz and including site-response corrections. *Bulletin of the Seismological Society of America*, *110*(6), 2862–2881.
- Rosenberger, A. (2010). Realtime ground-motion analysis: Distinguishing P and S arrivals in a noisy environment. *Bulletin of the Seismological Society of America*, *100*(3), 1252–1262.
- Rosenberger, A. (2019). *Three component accelerometer signal processing for WARN (0.6)* [Report]. Zenodo.
- Rosenberger, A., Banville, S., Collins, P., Henton, J., & Ferguson, E. (2018). *Kalman filter algorithm for the joint processing of GNSS PPP and accelerometer data, EEW parameters from the unbiased displacement time-series (0.9)* [Report]. Zenodo.
- Rosenberger, A., Crosby, B., Ferguson, E., Heesemann, M., Leech, B., MacArthur, M., Schlesinger, A., Scholte, L., Wang, L., & Zheng, Y. (2019). *Epicentre location and association of events from P-wave on-set times, magnitude estimation (2.1)* [Report]. Zenodo.
- Ross, Z. E., & Ben-Zion, Y. (2014). Automatic picking of direct P, S seismic phases and fault zone head waves. *Geophysical Journal International*, *199*(1), 368–381.
- Ross, Z. E., Ben-Zion, Y., White, M. C., & Vernon, F. L. (2016). Analysis of earthquake body wave spectra for potency and magnitude values: Implications for magnitude scaling relations. *Geophysical Journal International*, *207*(2), 1158–1164.
- Ross, Z. E., White, M. C., Vernon, F. L., & Ben-Zion, Y. (2016). An improved algorithm for real-time S wave picking with application to the (augmented) ANZA network in Southern California. *Bulletin of the Seismological Society of America*, *106*(5), 2013–2022.
- Roten, D., Olsen, K. B., & Takedatsu, R. (2020). Numerical Simulation of M9 Megathrust Earthquakes in the Cascadia Subduction Zone. *Pure and Applied Geophysics*, *177*, 2125–2141. <https://doi.org/10.1007/s00024-018-2085-5> Ruff, L. J. (1989). Do

- trench sediments affect great earthquake occurrence in subduction zones?. *Pure and Applied Geophysics*, 129(1/2), 263–282.
- Ruhl, C. J., Melgar, D., Chung, A. I., Grapenthin, R., & Allen, R. M. (2019). Quantifying the value of real-time geodetic constraints for earthquake early warning using a global seismic and geodetic data set. *Journal of Geophysical Research: Solid Earth*, 124(4), 3819–3837.
- Ruhl, C., J., Melgar, D., Grapenthin, R., & Allen, R. M. (2017). The value of real-time GNSS to earthquake early warning. *Geophysical Research Letters*, 44(16), 8311–8319.
- Sahakian, V. J., Baltay, A., Hanks, T. C., Buehler, J., Vernon, F. L., Kilb, D., & Abrahamson, N. A. (2019). Ground motion residuals, path effects, and crustal properties: A pilot study in Southern California. *Journal of Geophysical Research: Solid Earth*, 124(6), 5738–5753.
- Sahakian, V., Baltay, A., Hanks, T., Buehler, J., Vernon, F., Kilb, D., & Abrahamson, N. (2018). Decomposing leftovers: Event, path, and site residuals for a small-magnitude Anza region GMPE. *Bulletin of the Seismological Society of America*, 108(5), 2478–2492.
- Sahakian, V. J., Melgar, D., & Muzli, M. (2019). Weak near-field behavior of a tsunami earthquake: Toward real-time identification for local warning. *Geophysical Research Letters*, 46(16), 9519–9528.
- Samaei, M., Miyajima, M., Yazdani, A., & Jaafari, F. (2016). High frequency decay parameter (κ) for Ahar–Varzaghan double earthquakes, Iran (MW 6.5 & 6.3). *Journal of Earthquake and Tsunami*, 10(2), 1640006.
- Sandwell, D., Mellors, R., Tong, X., Wei, M., & Wessel, P. (2011). Open radar interferometry software for mapping surface deformation. *Eos, Transactions American Geophysical Union*, 92(28), 233–240.
- Satake, K. (1994). Mechanism of the 1992 Nicaragua tsunami earthquake. *Geophysical Research Letters*, 21(23), 2519–2522.
- Satake, K., Bourgeois, J., Abe, K., Abe, K., Tsuji, Y., Imamura, F., Lio, Y., Katao, H., Noguera, E., & Estrada, F. (1993). Tsunami field survey of the 1992 Nicaragua earthquake. *Eos, Transactions American Geophysical Union*, 74(13), 145–160.

- Satake, K., Nishimura, Y., Putra, P. S., Gusman, A. R., Sunendar, H., Fujii, Y., Tanioka, Y., Latief, H., & Yuliento, E. (2013). Tsunami source of the 2010 Mentawai, Indonesia earthquake inferred from tsunami field survey and waveform modeling. *Pure and Applied Geophysics*, *170*, 1567–1582.
- Saunders, J. K., Minson, S. E., & Baltay, A. S. (2022a). How low should we alert? Quantifying intensity threshold alerting strategies for earthquake early warning in the United States. *Earth's Future*, *10*(3), e2021EF002515.
- Saunders, J. K., Minson, S. E., Baltay, A. S., Bunn, J. J., Cochran, E. S., Kilb, D. L., O'Rourke, C. T., Hoshiba, M., & Kodera, Y. (2022b). Real-time earthquake detection and alerting behavior of PLUM ground-motion-based early warning in the United States. *Bulletin of the Seismological Society of America*, *112*(5), 2668–2688.
- Schlesinger, A., Kukovica, J., Rosenberger, A., Heesemann, M., Pirenne, B., Robinson, J., & Morley, M. (2021). An earthquake early warning system for Southwestern British Columbia, *Frontiers in Earth Science*, *9*, 1–14.
- Scholz, C. H. (1998). Earthquakes and friction laws. *Nature*, *391*, 37–42.
- Schulz, S. E., & Evans, J. P. (2000). Mesoscopic structure of the Punchbowl Fault, Southern California and the geologic and geophysical structure of active strike-slip faults. *Journal of Structural Geology*, *22*(7), 913–930.
- Sertcelik, F., Akçay, D., Livaoglu, H., & Gerdan, S. (2022). The spectral decay parameter kappa in Marmara Region, Turkey. *Arabian Journal of Geosciences*, *15*(271).
- Seyhan, E., & Stewart, J. P. (2014). Semi-empirical nonlinear site amplification from NGAWest2 data and simulations. *Earthquake Spectra*, *30*(3), 1241–1256.
- Shafiee, A. H., Zafarani, H., & Vajdi Bilesavar, M. (2021). Investigation of characteristics of decay parameters kappa based on Iranian dataset. *Journal of Seismology and Earthquake Engineering*, *23*(3), 1–9.
- Silva, W. J., & Darragh, R. B. (1995). *Engineering characterization of strong ground motion recorded flatrock sites*. Report TR-102262. Electric Power Research Institute, Palo Alto, California.
- Silva, W., Darragh, R. B., Gregor, N., Martin, G., Abrahamson, N., & Kircher, C. (1998). *Reassessment of Site Coefficients and Near-Fault Factors for Building Code*

- Provisions*. Technical Report Program Element II: 98-HQ-GR-1010. Pacific Engineering and Analysis, El Cerrito, California.
- Şişman, F. N., Kurtulmuş, T. Ö., & Askan Gündoğan, A. (2018). *High-Frequency Attenuation (Kappa, K) Estimations From The Recently-Compiled Turkey Strong Ground Motion Database For Western Turkey*. <https://hdl.handle.net/11511/76974>
- Small, D. T., & Melgar, D. (2023). Can stochastic slip rupture modeling produce realistic M9+ events?. *Journal of Geophysical Research: Solid Earth*, 128(3), 1–23.
- Small, D. T., & Melgar, D. (2021). Geodetic coupling models as constraints on stochastic earthquake ruptures: An example application to PTHA in Cascadia. *Journal of Geophysical Research: Solid Earth*, 126(7), 1–20.
- Smyth, A., & Wu, M. (2007). Multi-rate Kalman filtering for the data fusion of displacement and acceleration response measurements in dynamic system monitoring. *Mechanical Systems and Signal Processing*, 21(2), 706–723.
- Somerville, P., Graves, R., Collins, N., Song, S. G., Ni, S., & Cummins, P. (2009). Source and ground motion models for Australian earthquakes. In *Proceedings of the 2009 Annual Conference of the Australian Earthquake Engineering Society*. Newcastle, Australia, 11–13 December.
- Sonnemann, T., Halldorsson, B., & Jónsson, S. (2019). Automatic estimation of earthquake high-frequency strong-motion spectral decay in south Iceland. *Soil Dynamics and Earthquake Engineering*, 125, 105676.
- Stanko, D., Markušić, S., Ivančić, I., Mario, G., & Gülerce, Z. (2017). Preliminary estimation of kappa parameter in Croatia. *IOP Conference Series: Earth and Environmental Science*, 95.
- Stanko, D., Markušić, S., Korbar, T., & Ivančić, J. (2020). Estimation of the high-frequency attenuation parameter kappa for the Zagreb (Croatia) seismic stations. *Applied Sciences*, 10(24), 8974.
- Stewart, J. P., Wang, P., Teague, D. P., & Vecchiotti, A. (2019). Applications of non-ergodic site response in ground motion modeling. In *Proceedings of the 7th International Conference on Earthquake Geotechnical Engineering*. CRC Press.
- Sun, X., Tao, X., Duan, S., & Liu, C. (2013). Kappa (k) derived from accelerograms recorded in the 2008 Wenchuan mainshock, Sichuan, China. *Journal of Asian Earth Sciences*, 73, 306–316.

- Sung, C.-H., Abrahamson, N., & Lacour, M. (2023). Methodology for including path effects due to 3D velocity structure in nonergodic ground-motion models. *Bulletin of the Seismological Society of America*, 113(5), 2144–2163.
- Sweetkind, D. S., Taylor, E. M., McCabe, C. A., Langenheim, V. E., & McLaughlin, R. J. (2010). Three-dimensional geologic modeling of the Santa Rosa Plain, California. *Geosphere*, 6(3), 237–274.
- Tafreshi, M. D., Bora, S. S., Ghofrani, H., Mirzaei, N., & Kazemian, J. (2022). Region- and site-specific measurements of kappa (κ_0) and associated variabilities for Iran. *Bulletin of the Seismological Society of America*, 112(6), 3046–3062.
- Tanioka Y., & Ruff, L. J. (1997). Source time functions. *Seismological Research Letters*, 68(3), 386–400.
- Tanioka, Y., & Satake, K. (1996). Fault parameters of the 1896 Sanriku tsunami earthquake estimated from tsunami numerical modeling. *Geophysical Research Letters*, 23(13), 1549–1552.
- Tao, Z., Wang, X., Zhu, B., & Shang, T. (2020). Kappa (κ) in the 2011 Great East Japan earthquake. *Journal of Earthquake and Tsunami*, 14(6), 2050024.
- Thingbaijam, K., Mai, P., and Goda, K. (2017). New empirical earthquake source-scaling laws. *Bulletin of the Seismological Society of America*, 107(5), 2225–2246.
- Thompson, E. M., Wald, D. J., & Worden, C. B. (2014). A V_{S30} map for California with geologic and topographic constraints. *Bulletin of the Seismological Society of America*, 104(5), 2313–2321.
- Thompson, M., Hartog, J. R., & Wirth, E. A. (2023). A population-based performance evaluation of the ShakeAlert earthquake early warning system for M 9 megathrust earthquakes in the Pacific Northwest, U.S.A.. *Bulletin of the Seismological Society of America*, 1–21.
- Trnkoczy, A. (2009). Understanding and parameter setting of STA/LTA trigger algorithm. In *New Manual of Seismological Observatory Practice (NMSOP)* (pp. 1-20). Deutsches GeoForschungsZentrum GFZ.
- Trugman, D. T., Page, M. T., Minson, S. E., & Cochran, E. S. (2019). Peak ground displacement saturates exactly when expected: Implications for earthquake early warning. *Journal of Geophysical Research: Solid Earth*, 124(5), 4642–4653.

- Tsushima, H., & Ohta, Y. (2014). Review on near-field tsunami forecasting from offshore tsunami data and onshore GNSS data for tsunami early warning. *Journal of Disaster Research*, 9(3), 339-357.
- Ulrich, T., Gabriel, A.-A., Ampuero, J. P., & Xu, W. (2019). Dynamic viability of the 2016 Mw 7.8 Kaikōura earthquake cascade on weak crustal faults. *Nature Communications*, 10, 1–16.
- US Geological Survey. (1931). *United States National Strong-Motion Network* [Data set]. International Federation of Digital Seismograph Networks.
- US Geological Survey. (1967). *USGS Northern California Network* [Data set]. International Federation of Digital Seismograph Networks.
- US Geological Survey. (2017). *Quaternary Fault and Fold Database for the United States* [Data set]. <https://www.usgs.gov/natural-hazards/earthquake-hazards/faults> (Retrieved 1 August 2020).
- US Geological Survey. (2018). *CenValVM v1.1.0* [Github Repository]. <https://github.com/usgs/earthquake-cencalvm>
- US Geological Survey. (2018). *Finite Fault Model Maps* [Data set]. <https://earthquake.usgs.gov/earthquakes/eventpage/usp000hnj4/finite-fault> (retrieved July 2023)
- US Geological Survey. (2020a). *Libcomcat v2.0* [Github Repository].
- US Geological Survey. (2020b). *USGS Bay Area Velocity Model* [Data Set]. <ftp://ehzftp.wr.usgs.gov/baagaard/cencalvm/database>
- US Geological Survey. (n.d.). *Bay Area Geologic Map* [Data Set]. <https://earthquake.usgs.gov/education/geologicmaps/geology.php> (Retrieved 9 December 2021)
- Van Houtte, C., Drouet, S., & Cotton, F. (2011). Analysis of the origins of κ (kappa) to compute hard rock to rock adjustment factors for GMPEs. *Bulletin of the Seismological Society of America*, 101(6), 2926–2941.
- Van Houtte, C., Ktenidou, O.-J., Larkin, T., & Holden, C. (2014). Hard-site κ_0 (kappa) calculations for Christchurch, New Zealand, and comparison with local ground-motion prediction models. *Bulletin of the Seismological Society of America*, 104(4), 1899–1913.

- Van Houtte, C., Ktenidou, O.-J., Larkin, T., & Holden, C. (2018). A continuous map of near-surface S-wave attenuation in New Zealand. *Geophysical Journal International*, 213(1), 408–425.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., . . . , Vázquez-Baeza, Y. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17(3), 261–272.
- Wang Y., Raskin, J., Wolf, E. (2013). *The Oregon Resilience Plan: Reducing Risk and Improving Recovery for the Next Cascadia Earthquake and Tsunami*. Report to the 77th Legislative Assembly from Oregon Seismic Safety Policy Advisory Commission (OSSPAC).
https://www.oregon.gov/oem/documents/oregon_resilience_plan_final.pdf (last accessed December 2023).
- Wang, T., Trugman, D., & Lin, Y. (2021). SeismoGen: Seismic waveform synthesis using GAN with application to seismic data augmentation. *Journal of Geophysical Research: Solid Earth*, 126(4), 1–19.
- Weinstein, S. A., & Okal, E. A. (2005). The mantle magnitude M_m and the slowness parameter Θ : Five years of real-time use in the context of tsunami warning. *Bulletin of the Seismological Society of America*, 95(3), 779–799.
- Wells, D., & Coppersmith, K. (1984). 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.
- Wessel, P., Luis, J. F., Uieda, L., Scharroo, R., Wobbe, F., Smith, W. H. F., & Tian, D. (2019). The Generic Mapping Tools Version 6. *Geochemistry, Geophysics, Geosystems*, 20(11), 5556–5564.
- Williamson, A. L., Melgar, D., Crowell, B. W., Arcas, D., Melbourne, T. I., Wei, Y., & Kwong, K. (2020). Toward near-field tsunami forecasting along the Cascadia Subduction Zone using rapid GNSS source models. *Journal of Geophysical Research: Solid Earth*, 125(8), 1–19.
- Wirth, E. A., Sahakian, V. J., Wallace, L. M., & Melnick, D. (2022). The occurrence and hazards of great subduction zone earthquakes. *Nature Reviews Earth and Environment*, 3, 125–140.

- Wollherr, S., Gabriel, A.-A., & Mai, M. (2019). Landers 1992 “reloaded”: Integrative dynamic earthquake rupture modeling. *Journal of Geophysical Research: Solid Earth*, *124*(7), 6666–6702.
- Worden, C. B., Gerstenberger, M. C., Rhoades, D. A., & Wald, D. J. (2012). Probabilistic relationships between ground-motion parameters and modified Mercalli intensity in California. *Bulletin of the Seismological Society of America*, *102*(1), 204–221.
- Wu, Y.-M., Mittal, H., Chen, D.-Y., Hsu, T.-Y., & Lin, P.-Y. (2021). Earthquake early warning systems in Taiwan: Current status. *Journal of the Geological Society of India*, *97*, 1525–1532.
- Wu, Y., & Zhao, L. (2006). Magnitude estimation using the first three seconds P-wave amplitude in earthquake early warning. *Geophysical Research Letters*, *33*(16), 4–7.
- Wurman, G., Allen, R. M., & Lombard, P. (2007). Toward earthquake early warning in Northern California. *Journal of Geophysical Research: Solid Earth*, *112*(B08311), 1–19.
- Xu, B., Rathje, E. M., Hashash, Y., Stewart, J., Campbell, K., & Silva, W. J. (2020). k_0 for soil sites: Observations from KiK-net sites and their use in constraining small-strain damping profiles for site response analysis. *Earthquake Spectra*, *36*(1), 111–137.
- Yadav, R., Kumar, D., & Chopra, S. (2018). The high frequency decay parameter κ (kappa) in the region of North East India. *Open Journal of Earthquake Research*, *7*(2), 141–159.
- Yanagisawa, H., Goto, K., Sugawara, D., Kanamaru, K., Iwamoto, N., & Takamori, Y. (2016). Tsunami earthquake can occur elsewhere along the Japan Trench—Historical and geological evidence for the 1677 earthquake and tsunami. *Journal of Geophysical Research: Solid Earth*, *121*(5), 3504–3516.
- Yao, H., Shearer, P. M., & Gerstoft, P. (2013). Compressive sensing of frequency-dependent seismic radiation from subduction zone megathrust ruptures. In *Proceedings of the National Academy of Sciences*, *110*(12), 4512–4517.
- Ye, L., Lay, T., Kanamori, H., & Rivera, L. (2016). Rupture characteristics of major and great ($M_W \geq 7.0$) megathrust earthquakes from 1990 to 2015: 1. Source parameter scaling relationships. *Journal of Geophysical Research: Solid Earth*, *121*(2), 826–844.

- Yue, H., Lay, T., Rivera, L., Bai, Y., Yamazaki, Y., Cheung, K. F., et al. (2014). Rupture process of the 2010 MW 7.8 Mentawai tsunami earthquake from joint inversion of near-field HR-GPS and teleseismic body wave recordings constrained by tsunami observations. *Journal of Geophysical Research: Solid Earth*, *119*(7), 5574–5593.
- Zhang, G.-q., Li, D.-n., Gao, Y., & Yang, J.-q. (2022). A magnitude estimation approach and application to the Yunnan earthquake early warning network. *Seismological Research Letters*, *94*(1), 234–242.
- Zhang, H., Jin, X., Wei, Y., Li, J., Kang, L., Wang, S., Huang, L., & Yu, P. (2016). An earthquake early warning system in Fujian, China. *Bulletin of the Seismological Society of America*, *106*(2), 755–765.
- Zhang, L., Liao, W., Li, J., & Wang, Q. (2015). Estimation of the 2010 Mentawai tsunami earthquake rupture process from joint inversion of teleseismic and strong ground motion data. *Geodesy and Geodynamics*, *6*(3), 180–186.
- Zhao, J. X., Zhang, J., Asano, A., Ohno, Y., Oouchi, T., Takahashi, T., Ogawa, H., Irikura, K., Thio, H. K., Somerville, P. G., Fukushima, Y., & Fukushima, Y. (2006). Attenuation relations of strong ground motion in Japan using site classification based on predominant period. *Bulletin of the Seismological Society of America*, *96*(3), 898–913.
- Zhu, L. P., & Rivera, L. A. (2002). A note on the dynamic and static displacements from a point source in multilayered media. *Geophysical Journal International*, *148*(3), 619–627.
- Zollo, A., Iannaccone, G., Lancieri, M., Cantore, L., Convertito, C., Emolo, A., Festa, G., Gallovič, F., Vassallo, M., Martino, C., Satriano, C., & Gasparini, P. (2009). Earthquake early warning system in southern Italy: Methodologies and performance evaluation. *Geophysical Research Letters*, *36*(5), 1–6.