Improving Earthquake Rapid Response and Early Warning Performance with Geodesy and Machine Learning

by

Sydney N. Dybing

A dissertation accepted and approved in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in Earth Sciences

Dissertation Committee:

Diego Melgar, Chair

Amanda Thomas, Core Member

Valerie Sahakian, Core Member

Dare Baldwin, Core Member

Ben Farr, Institutional Representative

University of Oregon

Fall 2024

DISSERTATION ABSTRACT

Sydney N. Dybing

Doctor of Philosophy in Earth Sciences

Title: Improving Earthquake Rapid Response and Early Warning Performance with Geodesy and Machine Learning

This dissertation focuses on the work I have undertaken to investigate how quickly and accurately we can characterize earthquakes while and immediately after they occur, and potential ways to improve the systems we use for this. I focus on how we can use the modern technology we have for Earth monitoring as well as modern data processing and analysis techniques such as machine learning to improve the capabilities of the systems we rely on for earthquake disaster response. In this dissertation I present an exploration into the question of how early earthquakes are distinguishable by magnitude using borehole strainmeters, where we found that earthquakes do not appear to be strongly deterministic (i.e., large earthquakes are not inherently different from small earthquakes in their beginning stages). This has implications for how long it takes to accurately determine the magnitude of an earthquake, particularly for large events which rupture over longer periods of time. I then discuss our development of a machine learning algorithm for improving the performance of earthquake early warning systems for large earthquakes. This algorithm allows for discrimination between noisy GNSS waveforms which do actually contain seismic waves from earthquakes and those which do not, which enables us to reduce the amount of high-noise/low quality data that enters algorithms which determine the magnitude of such earthquakes. While earthquake early warning systems tend to operate only over specific regions such as the U.S. West Coast, organizations such as the USGS's National Earthquake Information Center also must rapidly publish information such as magnitude about worldwide earthquakes to aid in response efforts. However, magnitude estimation across a large range of earthquake sizes and tectonic settings is technically difficult. To help streamline this process, we developed another machine learning model which allows for the estimation of earthquake magnitudes uniformly for all locations and tectonic settings worldwide, which is also presented in this dissertation. Six multi-frame animations with more figures from this chapter are included as supplemental video files.

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

CURRICULUM VITAE

NAME OF AUTHOR: Sydney N. Dybing

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR, USA
Washington University in St. Louis, St. Louis, MO, USA.

DEGREES AWARDED:

Doctor of Philosophy in Earth Sciences, 2024, University of Oregon
Bachelor of Arts in Geophysics, 2019, Washington University in St. Louis

AREAS OF SPECIAL INTEREST:

Seismology
Geodesy
Machine Learning
Earthquake Early Warning

PROFESSIONAL EXPERIENCE:

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

Pathways Intern, U.S. Geological Survey Geologic Hazards Science Center, 2022

Undergraduate Research Assistant, Washington University Department of Earth and Planetary Sciences, 2017-2019

Incorporated Research Institutions for Seismology Internship, U.S. Geological Survey Albuquerque Seismological Laboratory, 2018

PUBLICATIONS:

Dybing, S. N., Yeck, W. L., Cole, H. M., & Melgar, D. (2024). Rapid estimation of single-station earthquake magnitudes with machine learning on a global scale. *Bulletin of the Seismological Society of America, 114*(3), 1523–1538. https://doi.org/10.1785/0120230171

Thomas, A. M., Melgar, D., **Dybing, S. N.**, & Searcy, J. R. (2023). Deep learning for denoising high-rate global navigation satellite system data. *Seismica, 2*(1). https://doi.org/10.26443/seismica.v2i1.240

Pasteris, J. D., Bae, Y., Giammar, D. E., **Dybing, S. N.**, Yoder, C. H., Zhao, J., & Hu, Y. (2021). Worth a closer look: Raman spectra of lead-pipe scale. *Minerals, 11*(10), 1047. https://doi.org/10.3390/min11101047

Dannemann Dugick, F. K., van der Lee, S., Prieto, G. A., **Dybing, S. N.**, Toney, L., & Cole, H. M. (2021). ROSES: Remote online sessions for emerging seismologists. *Seismological Research Letters, 92*(4), 2657–2667. https://doi.org/10.1785/0220200421

Dybing, S. N., Ringler, A. T., Wilson, D. C., & Anthony, R. E. (2019). Characteristics and spatial variability of wind noise on near-surface broadband seismometers. *Bulletin of the Seismological Society of America, 109*(3), 1082–1098. https://doi.org/10.1785/0120180227

GRANTS, AWARDS, AND HONORS:

Warren DuPre Smith Scholarship, University of Oregon Department of Earth Sciences, 2022

Outstanding Student Presentation Award, American Geophysical Union Fall Meeting, 2021

Honorable Mention, National Science Foundation Graduate Research Fellowship Program, 2021

Raymund Fellowship, University of Oregon Graduate School, 2019

Johnston Scholarship, University of Oregon Department of Earth Sciences, 2019

Courtney Werner Memorial Prize for Outstanding Academic Achievement in Earth Sciences, Washington University Department of Earth and Planetary Sciences, 2019

Edmund F. Maxwell Foundation Scholarship, The Edmund F. Maxwell Foundation, 2015

Ray Berg Memorial Scholarship, Sons of Norway Bothell Lodge, 2015

ACKNOWLEDGMENTS

The last five years have been both incredibly rewarding and also some of the most unexpectedly challenging years of my life. I am incredibly grateful to my wonderful advisor, Diego Melgar, for his steadfast support and mentorship. Thank you for always having an open door, and for giving me every opportunity to succeed as a scientist and student. All grad students should be lucky enough to have an advisor like Diego in their corner. I also greatly appreciate the advice and guidance that Valerie Sahakian, Amanda Thomas, Dare Baldwin, and Ben Farr have given me. Thank you to everyone on my committee for sharing your wealth of knowledge with me through conversations and classes over the years.

My entire academic experience has also greatly benefitted from working with people outside of the university setting. First, dating back to my IRIS internship in 2018 I want to thank all of the folks at the Albuquerque Seismological Laboratory for making my first introduction into what a career in seismology can look like a fantastic one, which really solidified my decision to go to grad school. Thank you particularly to Adam Ringler for sitting next to me for days on end to teach me Python from scratch, and for continuing to provide advice (and reviews!) far beyond my time at the ASL, and to Rob Anthony – I look forward to continuing to catch up with you both at conferences! Additionally, thank you to Will Yeck and Hank Cole for working with me during my 2022 internship with the National Earthquake Information Center, and for your mentorship and help to improve my research abilities.

As much as I love science and doing research, I also know that maintaining my friendships and hobbies outside of work has been key to my success in finishing this degree. Thank you to Marla Trox for not only being a wonderful support to me and everyone in the department, but also for recruiting me to play soccer with the Doodles during my time in Eugene! I'm also incredibly grateful to my figure skating community here in Eugene – it's a rare day when a trip to the rink doesn't have me feeling better about life in general. I would also not have made it to grad school without my brilliant, hilarious, and incredible Selskap friends from WashU, and I would not have made it through grad school without the support and encouragement from my amazing friends in my cohort, office, lab group, and the grad student body at large.

In particular, to Tara and Becca – I still can't believe you guys are coming all this way to watch my defense (actually, I can because that's the kind of people that you are), and I'm so

thankful for all the time we spent together over the last five years. Whether it was watching trashy reality TV, cinematic art like Hannibal (I refuse to add Twilight to this category), or just talking until the wee hours of the morning – you've kept me sane and also dabbled with me in insanity when necessary, and I love you both a lot. The grind of the past few months of finishing this dissertation weren't the same without you here, and I can't wait for all of the future times when we'll cross paths and pick up right where we left off!

And finally, my family. Thank you to my parents for always encouraging and enabling me to pursue my interests – for dragging me and my sister to Mount St. Helens three times as kids so that we could finally see it without clouds, for tolerating the rocks, shells, and plants I kept bringing into your house, and for spending weeks camped out with us so we could run around in the woods and look at the sky through telescopes. I doubt that without all of the time we spent exploring the natural world as a family that I would have found myself pursuing a career in the Earth Sciences. My entire life I never doubted that there was a place for me in science thanks to you. Mom, thank you for all of the tandem kayaking and for fostering our shared love of musical theater, and for helping me find jobs and internships that would allow me to achieve my goals throughout high school and college. Thank you for letting me borrow your car so that I could do my internship in Albuquerque, and I'm sorry that that resulted in it being totaled by hail (even though I think it worked out alright in the end!). Dad, thank you for helping me with all of my non-Mac related tech and car/garbage disposal mechanical issues, for enabling my pursuit of sports fandom and bobbleheads, and for always lifting up both me and my heavy stuff. Kaitlyn, thank you for sending me cute photos and videos of Chucky and commiserating about grad school with me. To know that I always have a loving home to come back to regardless of what's happening in my life is an enormous privilege and one that I am endlessly grateful for. I hope that I've made you proud!

DEDICATION

To Mom and Dad, for supporting me in everything I've ever done and never letting me believe that anything less than my wildest dreams were possible.

TABLE OF CONTENTS

Chapter Page

CHAPTER I: INTRODUCTION 18

1.1. How Earthquakes Begin 19

1.2. Rapid Earthquake Analysis Challenges 20

1.2.1. Earthquake Early Warning 20

1.2.2. NEIC Event Publication 21

1.3. Machine Learning Applications in Seismology 22

CHAPTER II: MAGNITUDE-INDEPENDENT RUPTURE NUCLEATION AS OBSERVED IN STRAIN DATA 24

2.1. Introduction and Background 24

2.2. Real Data Analysis 26

2.2.1. Methods 26

2.2.1.1. North America 26

2.2.1.2. Asia 29

2.2.2. Results 31

2.3. Modeling 37

2.3.1. Heterogeneous Rupture Model: The **M**7.1 2019 Ridgecrest Earthquake 38

2.3.1.1. Methods 38

2.3.1.2. Results 40

2.3.2. Deterministic Homogeneous Rupture Models 42

2.3.2.1. Methods 42

2.3.2.2. Results 45

2.4. Discussion 48

2.5. Conclusions 53

2.6. Data and Resources 53

2.7. Acknowledgements 54

2.8. Bridge 54

2.9. Supplemental Figures 55

CHAPTER III: DETECTING EARTHQUAKES IN NOISY REAL-TIME GNSS DATA WITH MACHINE LEARNING FOR IMPROVED PGD MAGNITUDE ESTIMATION 70

3.1. Introduction and Background 70

3.2. Data and Methods 74

3.2.1. Real Data Processing 74

3.2.2. Producing the FakeQuakes Displacement Data 76

3.2.3. Data Generation for the CNN 79

3.2.4. CNN Training Process 81

3.3. Results 81

3.3.1. FakeQuakes Testing Data 81

3.3.2. Real Testing Data 87

3.4. Discussion 94

3.5. Conclusions 102

3.6. Data and Resources 104

3.7. Acknowledgements 104

3.8. Bridge 104

3.9. Supplemental Figures 105

CHAPTER IV: RAPID ESTIMATION OF SINGLE-STATION EARTHQUAKE MAGNITUDES WITH MACHINE LEARNING ON A GLOBAL SCALE 108

4.1. Introduction 108

4.1.1. Machine Learning for Earthquake Source Parameter Estimation 109

4.2. Methods 111

4.2.1. MLAAPDE Data Access 111

4.2.2. Training Data Processing and Augmentation 112

4.3. Results and Discussion 116

4.3.1. Model Performance by Magnitude 116

4.3.2. Model Performance by Waveform and Earthquake Characteristics 121

4.3.3. Model Performance by Window Length 127

4.3.4. Model Performance on Edge Cases 131

4.4. Conclusions 133

4.5. Data and Resources 134
4.6. Acknowledgements 135
4.7. Supplemental Figures 136
CHAPTER V: CONCLUSIONS AND FUTURE DIRECTIONS 142
APPENDIX: FDSN NETWORK REFERENCES 145
REFERENCES CITED 155
SUPPLEMENTAL FILES
VIDEO: VIDEO 4.S1
VIDEO: VIDEO 4.S2
VIDEO: VIDEO 4.S3
VIDEO: VIDEO 4.S4
VIDEO: VIDEO 4.S5
VIDEO: VIDEO 4.S6

LIST OF FIGURES

Figure Page

2.1. Maps of the earthquakes and strainmeters used in this study 27

2.2. An example of how the root mean square strain timeseries were used to calculate the peak strain 30

2.3. Examples of peak strain waveforms from the North American earthquakes and Gladwin Tensor Borehole Strainmeters 32

2.4. Examples of peak strain waveforms from the Asian earthquakes 33

2.5. Results of the Markov chain Monte Carlo inversions 35

2.6. MCMC scatter plot results with the addition of event-average mean transition times and linear regression results 36

2.7. Views of the **M**7.1 Ridgecrest earthquake fault that we used to simulate the rupture in SW4 and the SW4 model domain and receiver stations 39

2.8. Plots of the real peak strain and peak strain from the SW4 simulations of the **M**7.1 Ridgecrest earthquake 41

2.9. Illustration of a homogeneous rupture model, a view of the planar fault used for the deterministic SW4 simulations, and a view of the SW4 model domain and receiver stations 44

2.10. Examples of peak strain waveforms from the deterministic-designed earthquakes simulated in SW4 46

2.11. Results of the MCMC inversions on the deterministic-designed earthquakes simulated in SW4 47

2.12. Results of the MCMC inversions on the high-pass filtered deterministic-designed earthquakes simulated in SW4 51

2.S1. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6 on April 13, 2023 55

2.S2. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.4 on December 20, 2022 55

2.S3. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.2 on November 22, 2022 56

2.S4. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.2 on December 20, 2021 56

2.S5. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6 on July 8, 2021 57

2.S6. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.5 on May 15, 2020 57

2.S7. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.5 on March 31, 2020 58

2.S8. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.3 on December 25, 2019 58

2.S9. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6 on December 23, 2019 at UTC hour 20 59

2.S10. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6 on December 23, 2019 at UTC hour 19 59

2.S11. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.3 on August 29, 2019 60

2.S12. Examples of peak strain waveforms from the North American earthquake dataset for an **M**7.1 on July 6, 2019 60

2.S13. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.4 on July 4, 2019 61

2.S14. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.2 on July 4, 2019 61

2.S15. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.8 on October 22, 2018 62

2.S16. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.5 on October 22, 2018 at UTC hour 5 62

2.S17. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.2 on August 22, 2018 63

2.S18. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.6 on December 8, 2016 63

2.S19. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.02 on August 24, 2014 64

2.S20. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.5 on April 24, 2014 64

2.S21. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6 on September 4, 2013 65

2.S22. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.1 on September 3, 2013 65

2.S23. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.3 on December 14, 2012 66

2.S24. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.1 on November 8, 2012 66

2.S25. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6 on April 11, 2012 67

2.S26. Examples of peak strain waveforms from the North American earthquake dataset for an **M**7.2 on April 4, 2010 67

2.S27. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.5 on January 10, 2010 68

2.S28. Results of the MCMC inversions on the noisy deterministic-designed earthquakes simulated in SW4 69

3.1. Map of the study region 74

3.2. Example of the rupture pattern for one of the FakeQuakes ruptures used to generate synthetic displacement waveforms for the training dataset 78

3.3. Example of the process of combining the synthetic waveforms with real noise to build realistically noisy synthetic waveforms, and examples of the waveforms, Gaussian labels, and targets 80

3.4. Results of the testing of the CNN model on the remaining FakeQuakes waveforms which it never saw in training 83

3.5. Confusion matrices of the results of the model predictions on the FakeQuakes testing data compared to the true sample labels............................... 84

3.6. Plots showing the accuracy of the CNN's predictions on the FakeQuakes testing data compared to source/waveform characteristics.................................. 85

3.7. A histogram of the differences between the actual P-wave arrival time and the CNN predicted arrival time for the FakeQuakes testing dataset............... 86

3.8. Results of the testing of the CNN model on the real testing dataset.................... 87

3.9. A plot showing the number of true positive, true negative, false positive, and false negative cases in the real testing data results 89

3.10. Confusion matrices of the results of the model predictions on the real testing data compared to the true sample labels... 90

3.11. Plots showing the outcomes of predictions for the 2123 individual waveforms that included P-wave arrivals in the real testing dataset 92

3.12. Histograms of the differences between the actual P-wave arrival time and the CNN predicted arrival time for the real testing dataset............................ 93

3.13. Plots of the event-averaged PGD magnitudes for the FakeQuakes earthquakes in the testing dataset using several different methods to restrict the dataset 96

3.14. Plots of the event-averaged PGD magnitudes for the earthquakes in the real testing dataset using several different methods to restrict the dataset 99

3.S1. Earthquake waveform magnitude distributions for the real earthquake testing data and the full set of FakeQuakes synthetic data 105

3.S2. Convolutional neural network architecture used in this study........................... 106

3.S3. Training history for the CNN model developed in this study 107

3.S4. Distribution of waveform characteristics in the two testing datasets 107

4.1. Maps of the earthquakes and seismometers that recorded them used in this study... 113

4.2. Timeseries plots demonstrating the different types of augmentation implemented in our data generator 113

4.3. Boxplots showing the model's estimated earthquake magnitudes for the testing dataset 116

4.4. A boxplot showing the error in the magnitude estimation made by the model plotted against the source depth of the earthquake 122

4.5. Boxplots showing the error in the magnitude estimation made by the model plotted against the hypocentral distance 123

4.6. Boxplots showing the error in the magnitude estimation made by the model plotted against the waveform signal to noise ratio........ 125

4.7. The median and standard deviation of the model's estimation error on the testing dataset plotted for each of the tested time windows........ 128

4.8. Plot showing the absolute error of the median of the estimations made by the model for each catalog magnitude bin 129

4.9. Plot of the model's estimated earthquake magnitudes for the testing dataset including the seven selected edge case earthquakes 133

4.S1. Logarithmic histograms showing the distributions of magnitudes associated with the MLAAPDE waveforms selected in the training, validation, and testing datasets 136

4.S2. The median and standard deviation of the model's estimation error on the testing dataset for the original and weighted training schemes, and the absolute error of the median of the estimations made by the original and weighted training schemes........ 136

4.S3. The standard deviation of the model's estimation error on the testing dataset for the single-station method and the event-averaged method 138

4.S4. Boxplots showing the error in the magnitude estimation made by the model plotted against the waveform signal to noise ratio for the historical and MLAAPDE data separately........ 139

4.S5. Plot showing the absolute error of the 95th percentile of the estimations made by the model for each catalog magnitude bin 140

4.S6. Plot of the model's estimated earthquake magnitudes for the testing dataset including two additional edge case earthquakes........ 141

LIST OF TABLES

Table Page

2.1. Fault and rupture parameters for the SW4 deterministic homogeneous earthquake simulations.......... 42

CHAPTER I

INTRODUCTION

Earthquakes are a fascinating and intimidating natural disaster for many people due to our inability to predict exactly when and where they will happen, as well as how strong they will be. Nevertheless, the shaking and other earthquake-associated hazards such as tsunamis that they produce has and will continue to cause massive loss of life and property, so a substantial amount of research has focused on how we can better understand earthquakes to mitigate these risks and also respond to them quickly and in the most effective ways. In the United States (U. S.), for example, the U. S. Geological Survey's (USGS) National Earthquake Information Center (NEIC) has for decades monitored earthquakes across the globe, today running a 24/7 rapid response system that publishes earthquake reports within minutes and hours of occurrence to provide emergency responders, governments, media, and the public information about such events. The last few decades have also seen the advent of public earthquake early warning (EEW) systems, which quickly detect and characterize earthquakes in order to alert the public and utilities to take protective action before strong shaking reaches them. In the U. S., these alerts are currently provided in the states of Washington, Oregon, and California by the USGS's ShakeAlert program.

Though ShakeAlert and the NEIC operate on slightly different timescales and with different amounts of automation, they both fundamentally must be able to quickly detect earthquakes in real-time data transmitted from networks of instruments that observe the earth, and from that data they must be able to determine characteristics of the earthquakes such as their size (magnitude) and location. Accomplishing these goals relies on an understanding of how earthquakes initiate and evolve and how to extract useful information from the ground motions they produce as they begin. This dissertation focuses both on the basic science question of how earthquakes start and how soon in the rupture process their final magnitude is determined, as well as how modern instrumentation and data analysis methods like machine learning can contribute to improvements in our ability to rapidly characterize them.

1.1. How Earthquakes Begin

An ongoing debate within a segment of the seismological community for years has been the question of whether earthquakes are deterministic or not. Essentially, when an earthquake begins, is its final magnitude already predetermined, or do all earthquakes start the same way and then some grow to be larger than others? The answer to this question has significant implications for our ability to rapidly analyze and appropriately respond to earthquakes. If earthquakes do begin differently depending on their final size, i.e., they are strongly deterministic, we should be able to use the earliest arriving data in the P-waves to accurately determine that final magnitude and have a good gauge for the destruction potential of the earthquake, whether it could cause a tsunami, etc. If they all begin the same, we must instead wait for potentially a substantial amount of the full rupture process to conclude (which can take minutes for the largest earthquakes) before we know the ultimate size of the earthquake.

A number of studies have proposed that earthquakes are strongly deterministic because certain data analyses showed differences between small and large events within the first few seconds of the P-wave data (e.g., Ellsworth and Beroza, 1995; Olson and Allen, 2005; Colombelli et al., 2014). Other work has disputed these results, instead finding no evidence for early systematic differences between earthquakes of different magnitudes, suggesting that small, handpicked datasets and attenuation, instrument effects, and artifacts could be responsible for the patterns observed in other studies (e.g. Mori and Kanamori, 1996; Rydelek and Horiuchi, 2006; Meier et al., 2016, 2017, 2020). These studies were primarily done using seismic data because the high sensitivities of such instruments allow for careful examination of P-waves. However, seismometers are not the only highly sensitive instruments used for earthquake research. A lesser-used type of geodetic instrument, the borehole strainmeter, had been used to study dynamic earthquake waveforms before (e.g., Barbour and Crowell, 2017), but no one had attempted to use strainmeter data to answer the question of whether earthquakes are strongly deterministic or not.

This literature gap is addressed in Chapter II of this dissertation. This chapter was coauthored with my advisor Diego Melgar, as well as Andrew J. Barbour, Alexandre Canitano, Suguru Yabe, and Dara E. Goldberg, and will be submitted for publication soon. We analyzed peak strain waveforms from instruments across North America and in Japan and Taiwan and found that there did not appear to be a significant relationship between the characteristics of the growth of peak strain and earthquake magnitude. Instead, hypocentral distance appeared to enforce stronger controls over the appearance of the peak strain waveforms. We built earthquake rupture models to produce strain waveforms using a finite-difference wave propagation software called SW4 (Petersson and Sjogreen, 2017) to show that our simulated strain waveforms were realistic compared to the real data, and then produced a series of earthquakes designed to be strongly deterministic. In this case, we found that we could observe a relationship between the peak strain waveforms and earthquake magnitude in this situation. These results suggested that if strong determinism did occur in nature, we should have observed it in our real peak strain analysis, but we did not.

1.2. Rapid Earthquake Analysis Challenges

1.2.1. Earthquake Early Warning

As previously discussed, how quickly earthquakes differentiate themselves by size strongly affects how long it takes us to estimate that final magnitude accurately. Arguably, in no setting is this more important than in EEW, where in order to send accurate and useful warnings to users of the EEW system, it must be able to quickly determine how large an earthquake is and where it is to estimate expected shaking. EEW systems have been in use in several countries worldwide for a number of years, with the U. S.'s system ShakeAlert covering the West Coast states of Washington, Oregon, and California currently. ShakeAlert works by using a dense seismometer network to quickly detect the P-waves from an earthquake soon after it initiates, using those P-wave observations to calculate a magnitude and location, and quickly sending alerts to users before the slower and more damaging S-waves arrive at their location and cause more significant shaking (Given et al., 2018).

However, these seismic-data-based algorithms tend to underestimate the true magnitudes of earthquakes larger than ~M7 (Trugman et al., 2019), a phenomenon known as magnitude saturation. This occurs for a few reasons related to both the instruments used and processing required for their data, as well as source effects for some magnitude types, which will be covered in more detail in the chapters of this dissertation. Magnitude saturation can have major implications for our ability to accurately determine the magnitude of a large earthquake in real-time. For example, the 2011 **M**9.1 Tōhoku earthquake in Japan was interpreted to be only an **M**8.1 even minutes after the earthquake began, resulting in underpredictions of the size of the tsunami it produced (Hoshiba and Ozaki, 2014).

One way to avoid this issue is by using non-seismic data to determine earthquake magnitude. For example, geodetic Global Navigation Satellite Systems (GNSS) data do not saturate because they measure displacement relative to an absolute reference frame (Melgar et al., 2015). This is why ShakeAlert was recently upgraded to include a GNSS-based algorithm called GFAST into its processing workflow, allowing for accurate magnitude estimations for large earthquakes such as the megathrust event expected along the Cascadia Subduction Zone at some point in the future. However, the noise levels of GNSS data are very high compared to traditional seismic data (Geng et al., 2018), which obscures P-wave arrivals. As a result, GFAST must be triggered by the seismic algorithms to begin processing. Additionally, it has been demonstrated for another GNSS-based earthquake characterization algorithm that inputting noisy data results in less accurate magnitude estimations (Lin et al., 2023). In Chapter III of this dissertation, we demonstrate how machine learning can be used to develop a model which detects earthquakes in noisy GNSS data, allowing for the construction of a “filter” which reduces the amount of low-quality data that enters an algorithm like GFAST.

1.2.2. NEIC Event Publication

In addition to the USGS’s responsibilities in operating ShakeAlert, they also publish information about earthquakes on a slightly longer timescale for earthquakes that occur everywhere in the world. Reports produced by the USGS’s NEIC are created through a combination of automatic processing and human intervention by analysts working 24/7 (Benz, 2017). They include information about the location and depth of earthquakes, as well as magnitude, which is of highest interest for audiences like the public to make a quick judgment about how severe the impacts of an earthquake will be. Magnitude estimation, however, is not a particularly simple process. Depending on factors such as the size of an earthquake, the tectonic environment, the amount of data available, etc., any one or several of a number of different magnitude types may be calculated that are only appropriate for certain earthquakes and situations, and some of these magnitude types are prone to higher errors than others. This offered an opportunity for the development of a system which can streamline this process and estimate earthquake magnitudes uniformly for all locations and tectonic settings across the globe, not relying on locating the earthquake first, which can then rapidly alert analysts of potentially significant larger events. The results of this work are presented in Chapter IV. Because of the NEIC’s long history of processing and publishing a global earthquake catalog, substantial amounts of data were available to draw from to accomplish this goal, and we exploited this dataset with a machine learning approach.

1.3. Machine Learning Applications in Seismology

The growth in availability and sophistication of machine learning resources in recent years has offered opportunities for the seismology discipline to explore new methods of data processing and analysis, which we utilize in Chapters III and IV in this dissertation. Some of the applications that have been published before include methods for improved detection and location of earthquakes, allowing for the expansion of earthquake catalogs, as well as approaches for determining the magnitudes of earthquakes (as reviewed in Mousavi and Beroza, 2023).

Machine learning typically requires great volumes of data for training in order to perform adequately. Conveniently, the seismological community has traditionally invested great time and energy into making seismic data publicly available through resources like the NEIC's catalog or the EarthScope waveform data archives. However, the amount of data available for GNSS stations instead of seismometers is substantially smaller, with fewer stations and fewer years of real-time data coverage available. For Chapter III, where we developed a machine learning model for detecting earthquakes in noisy real-time GNSS data to improve EEW system performance, we wanted to hold the limited real data we had access to in reserve for testing our model after it was trained. To produce our training dataset, we therefore followed an approach which had been shown to work well for several other GNSS machine learning studies (e.g., Lin et al., 2021, 2023) and generated a dataset comprised of >700,000 realistic synthetic displacement waveforms using a set of codes called FakeQuakes (Melgar et al., 2016). We combined these waveforms with real noise from the real data we had access to make sure that the training waveforms were realistic. We were then able to test our model both on additional synthetic data and the real data which we held back. In Chapter III, we discuss the performance of our trained model on both the unseen synthetic data and real data. We also demonstrate how the model can be used to selectively filter only high-quality data where earthquake signal is seen for input into an algorithm like GFAST, and how this has the capability to reduce the error in its earthquake magnitude estimations. Chapter III was coauthored with my advisor Diego Melgar, as well as Amanda M. Thomas, Dara E. Goldberg, David Mencin, and Brendan Crowell, and will also be submitted for publication soon.

In Chapter IV, we were able to use the more traditional seismic data from the NEIC's Preliminary Determination of Epicenters catalog to train our global model for earthquake magnitude characterization. In in this chapter we present AIMag, a model for rapid estimation of single-station earthquake magnitudes using raw three-component P-waveforms observed at local to teleseismic distances, independent of prior size or location information. This chapter has been previously published, and was coauthored with William L. Yeck, Hank M. Cole, and my advisor, Diego Melgar. We trained our model using ~2.4 million P-phase arrivals labeled by the authoritative magnitude assigned by the USGS, and tested input data parameters (e.g., window length) that could affect the performance of our model in near-real-time monitoring applications. We found that while we did observe underestimation of magnitudes for larger earthquakes, which we attribute to magnitude saturation and lack of events in the training data at these magnitudes, AIMag can accurately determine earthquake magnitudes from ~M2.3-M7.6 using individual stations' waveforms without instrument response correction or knowledge of an earthquake's source-station distance. This work may enable monitoring agencies to more rapidly recognize large, potentially tsunamigenic global earthquakes from few stations, allowing for faster event processing and reporting. This is critical for timely warnings for seismic-related hazards.

CHAPTER II

MAGNITUDE-INDEPENDENT RUPTURE NUCLEATION AS OBSERVED IN STRAIN DATA

This chapter was coauthored with my advisor Diego Melgar, who helped conceptualize and plan the study and provided editorial feedback, Andrew J. Barbour, who helped with information about processing the North American strain data, Alexandre Canitano, who provided the strain data from Taiwan and information about processing it, Suguru Yabe, who provided the strain and seismic data from Japan and information about processing it, and Dara E. Goldberg, who provided the fault models for the Ridgecrest earthquake. I performed all data analysis, produced all figures, and wrote all of the text in the manuscript.

2.1. Introduction and Background

Despite continuing improvements in seismic instrumentation, many questions still abound about the basic science governing the behavior of large (M>6) earthquakes, and notably about the very beginning of the earthquake rupture process. Of particular interest and debate over decades of research has been the question of how early we can tell how large an earthquake will be – how deterministic the rupture process is. If it were possible to determine the final magnitude of an earthquake from early arriving data rather than needing to wait for the entire rupture process to be complete, we would be able to deliver more accurate early warnings more quickly.

A number of studies have argued that earthquakes are strongly deterministic/predictable, where the final magnitude can be determined with just a few seconds of P-wave seismic data (e.g., Ellsworth and Beroza, 1995; Olson and Allen, 2005; Colombelli et al., 2014). Other work has suggested that the slope of growth of moment rate functions for earthquakes scale with earthquake magnitude (Ishihara et al., 1992; Melgar and Hayes, 2019). These studies generally support a model of earthquake rupture where properties of the initial nucleation phase exert at least some control over the final earthquake magnitude. For example, a rupture that begins with higher amplitude slip and more energy may be more likely to grow into a larger magnitude earthquake because the rupture is more capable of continuing to grow through fault heterogeneities that would otherwise impede it (Olson and Allen, 2005). This contrasts the "cascade model" of earthquake rupture, where there is no difference between the beginnings of small and large earthquakes. In this model, slip that initiates on one fault patch simply spreads to others until the state of stress on a new patch is unfavorable for continued propagation.

However, subsequent studies have disputed many of these results, arguing that strong determinism does not exist. For example, Mori and Kanamori (1996) did not find evidence for the systematic differences between the beginnings of large and small events as observed by Ishihara et al. (1992) or Ellsworth and Beroza (1995). They noted instead that attenuation and instrument effects could explain the differences in the beginnings of the waveforms, and that if a true nucleation signal did exist, it would not be resolvable by instrumentation in California at that time. Rydelek and Horiuchi (2006) used the same method as Olson and Allen (2005) for a different earthquake dataset and did not observe the same magnitude scaling trend as the original study. Meier et al. (2016) noted that previous studies often used datasets that were small and handpicked, had narrow magnitude ranges, or used data from large hypocentral distances that were not corrected for, and found with their own dataset that small and large ruptures were not distinguishable early on. Meier et al. (2017) followed this with a study of moment rate functions that found that for large earthquakes, the moment rate functions were scalable with magnitude, but were statistically indistinguishable until the peak moment rate was reached. Meier et al. (2020) found that the moment rate acceleration scaling with magnitude observed in Melgar and Hayes (2019) was at least in part an artifact of the data.

Out of this debate emerged a third regime between strong and no determinism, which has been referred to as weak determinism or weak/moderate predictability. In this regime, earthquake magnitude may not be distinguishable by mere seconds into the earthquake, but it is still possible to determine the magnitude before the entire rupture process is complete. Depending on the study, the time that it takes for enough information to have been received to estimate the magnitude varies. For example, peak moment rate can be used to determine earthquake magnitude, and this is reached within 35-55% of the total rupture duration time (Meier et al., 2017). In another model of earthquake rupture where earthquakes begin similarly but then organize into self-healing slip pulses as described by Heaton (1990), an earthquake's average rise time, or the amount of time that any particular fault patch is slipping, also scales with magnitude, and this time can be as little as 8-25% of the rupture duration (Melgar and Hayes, 2017).

In this study, we aimed to continue to work towards a better understanding of the earthquake rupture process by using data from a type of instrument that has not been used for this purpose before – the borehole strainmeter. We were interested in using strain data to analyze the early stages of earthquakes because these instruments do not carry some of the inherent issues that seismic and Global Navigation Satellite Systems (GNSS) data do in a setting where we are trying to interpret earthquake source properties. Seismic data suffer from magnitude saturation at high magnitudes, where large earthquakes become indistinguishable from even larger ones. This can occur for a few reasons, but one example is that magnitudes are often calculated using seismic data that is double-integrated from acceleration to displacement. This requires high-pass filtering the data to remove baseline drift effects, which removes low frequencies that may change the relationship between displacement and magnitude (Colombelli et al., 2012). GNSS data do not suffer from magnitude saturation, as ground displacement is directly measured with respect to an absolute reference frame (Melgar et al., 2015). However, they have a very high noise floor, which obscures P-wave arrivals (Geng et al., 2018; Melgar et al., 2020).

In contrast, strainmeters can record both small and large perturbations in strain with broadband fidelity to the static strain offset, so we can both observe the P-wave arrivals and avoid saturation at higher magnitudes (Barbour and Crowell, 2017). The goal of this study is therefore to understand what new insight into the earthquake rupture process strain data may offer, and explore whether or not any strength of magnitude determinism is evident in the data. We have taken both an *in situ* and modeling approach to answering these questions using earthquake recordings from borehole strainmeters emplaced in Western North America, Japan, and Taiwan, as well as modeling additional earthquake scenarios.

2.2. Real Data Analysis

2.2.1. Methods

2.2.1.1. North America

With the focus on large earthquakes, we chose to analyze strain data from 31 earthquakes ≥**M**6 that occurred in and offshore of Western North America between 2008 and 2023. We compiled a dataset of 20 Hz sampling rate borehole strainmeter (BSM) earthquake records from 71 Gladwin Tensor BSMs (Gladwin, 1984) in Western North America, operated by the EarthScope Consortium's Network of the Americas (NOTA). A map of the earthquakes and BSM stations used for analysis are included in Figure 2.1a. These data were downloaded in five-minute windows starting from the origin time of each earthquake from the IRIS Web Service using ObsPy (Beyreuther et al., 2010) in MiniSEED format.

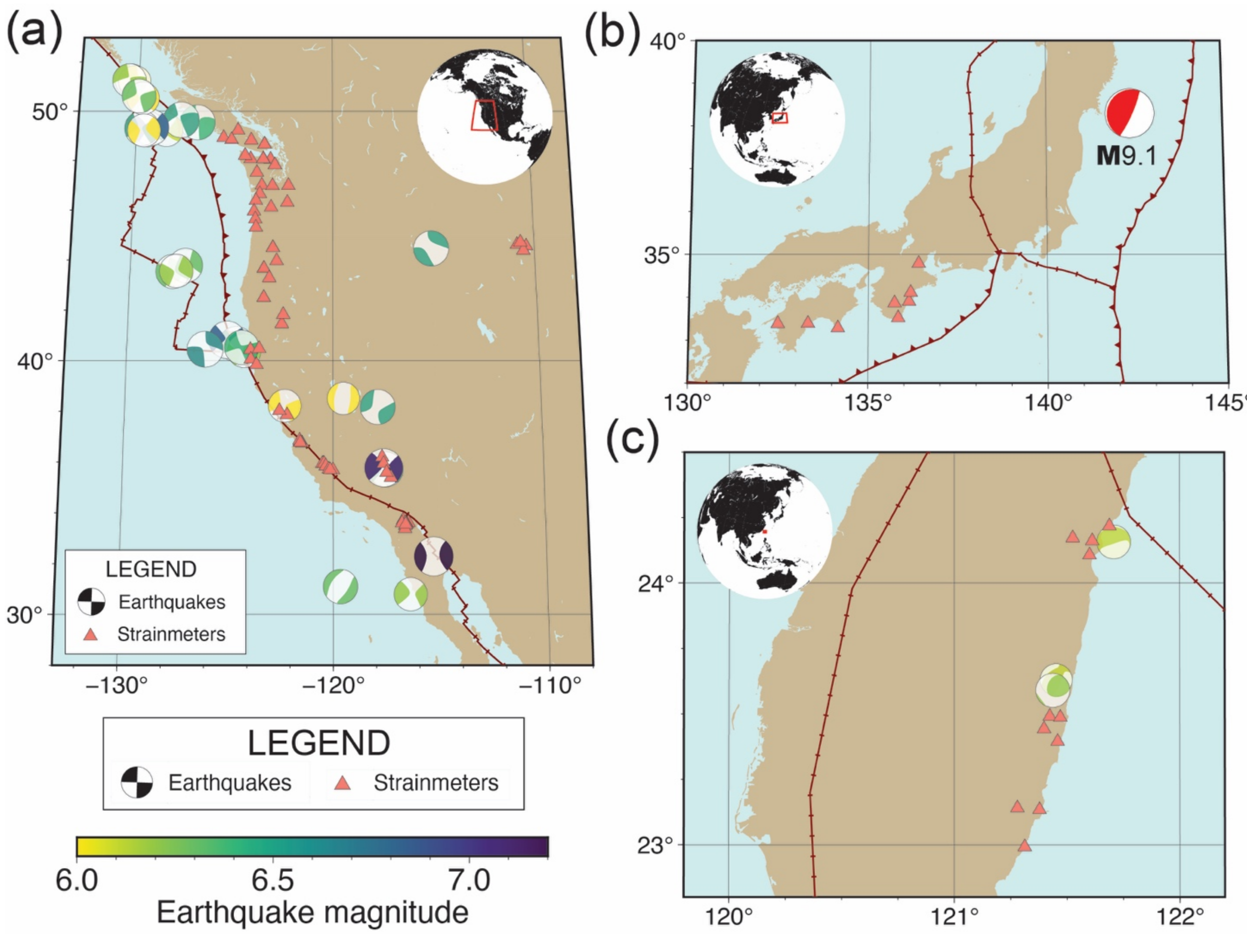


Figure 2.1. Maps of the earthquakes (colored focal mechanisms, accessed using EarthScope's Searchable Product Depository (Trabant et al., 2012) from the Global Centroid Moment Tensor catalog (Dziewonski et al., 1981; Ekström et al., 2012)) and strainmeters (orange triangles) from which data were used in this study. **(a)** The 31 earthquakes (**M**6.0-7.2) from 2008-2023 and 71 Gladwin Tensor borehole strainmeters used in Western North America. **(b)** The 2011 **M**9.1 Tōhoku earthquake and 8 Ishii-type borehole strainmeters used in Japan. **(c)** The 3 earthquakes (**M**6.1-6.2) from 2009-2018 and 11 Sacks-Everston borehole strainmeters (8 one-component dilatometers, 3 three-component) used in Taiwan. These maps were constructed using PyGMT (Uieda et al., 2023) with the plate boundaries (dark red) from Bird (2003).

Gladwin Tensor BSMs record horizontal strain with four pairs of capacitors oriented in different directions within the instrument (see Barbour and Crowell (2017), their Figure 1a). As seismic waves move through a borehole and deform it and the instrument, these sensor pairs record changes in capacitance. Three of the sensor pairs are oriented at 60 degree angles to each other, and a fourth redundant sensor pair is offset by only 30 degrees. In order to best compare the data from the Gladwin Tensor BSMs with data from other types of strainmeters, we chose to only analyze data from the three primary sensor pairs. We did test all of the analysis that will be described throughout the paper using all four components as well, and our findings remain unchanged.

To analyze the data as strain, we began by processing the three non-redundant components of the raw BSM capacitance data according to the methods in Barbour and Crowell (2017). This first involves a conversion from capacitance to linear extensional strain $\boldsymbol{e}$ using the following formula:

$$\boldsymbol{e} = R\,\frac{\boldsymbol{g}/C}{1-\boldsymbol{g}/C}\ , \tag{2.1}$$

where R is a ratio that depends on the specific instrument configuration (either 100 µm / 87 mm or 200 µm / 87 mm for the NOTA BSMs), $\boldsymbol{g}$ is the time series of capacitance, and C is an analog-to-digital conversion factor (10^8 for the NOTA BSMs). During this process, we also detrended the data and identified and removed any anomalous samples, filling the gaps via simple linear interpolation. As the Gladwin Tensor BSMs have a flat response from DC/0 Hz (the static strain offset) to 10 Hz (Gladwin, 1984), no instrument response removal was required in the processing of these data.

Because there have been questions raised about whether the Gladwin Tensor BSMs reliably measure static strain offsets (Barbour et al., 2015), we applied a causal high-pass filter with a corner frequency of 0.001Hz to the three linear extensional strain components. This removed static strain offsets while maintaining the wave phase and timing of the data, as we were interested in analyzing characteristics of the strain immediately following the P-wave arrivals.

We then calculated the root mean square (RMS) strain E:

$$E = \sqrt{(\boldsymbol{e}_1^2 + \boldsymbol{e}_2^2 + \boldsymbol{e}_3^2)/3}, \tag{2.2}$$

where $\boldsymbol{e}_{1\text{-}3}$ are the three filtered linear extensional strain components.

Our primary focus was analysis of earthquake onset on strain recordings, so we wanted to be able to compare the appearance of the P-waves across different earthquakes and stations. Unfortunately, P-wave picking was not easy to accomplish with data from these particular BSMs. While noise levels were low and sensitivity was high, arrivals were emergent and difficult to pinpoint. This is likely because of the nature of the instruments. Gladwin Tensor BSMs only record horizontal strain, whereas P-waves are compressive and have the highest amplitudes when recorded on vertical components. However, the NOTA BSM stations also include collocated seismometers fitted above the strainmeters in the boreholes, so P-wave picking was accomplished on the vertical component of these instruments using Pyrocko's Snuffler software (Heimann et al., 2017). We then trimmed the RMS strain waveforms by these picked P-wave arrivals so that each waveform was 2 minutes long, with the P-wave arrival at 10 seconds. We calculated signal to noise ratios (SNRs) for each RMS strain waveform using 8 seconds of noise before the P-wave arrival (or as many seconds as possible for waveforms with small hypocentral distances) and 10 seconds of signal after the P-wave arrival. We decided to focus our analysis on data with SNRs >2 and hypocentral distances ≤500 km.

Finally, we calculated timeseries of peak strain for each waveform in the focus group by cycling through each RMS strain waveform from the arrival of the P-wave and tracking each new largest strain value. Peak strain values for the 10 seconds before the P-wave were simply set to the value of the strain at the P-wave arrival. This was done to avoid any pre-arrival noise obscuring signals at the onset of the P-wave. A visual example of this process can be found in Figure 2.2.

2.2.1.2. Asia

To explore whether any of the features we observed in the North American strain data were visible in different regions and with different types of BSMs, we also analyzed strain data from the 2011 Tōhoku earthquake in Japan (Figure 2.1b). The BSMs that recorded this earthquake are Ishii-type instruments (Ishii et al., 2002), and record four components of horizontal strain with a 20 Hz sampling rate. The four sensors in these instruments are all oriented at 45 degree angles to each other, making one redundant, so we chose to analyze only data from the first three analogously to our decision for the data from the Gladwin Tensor BSMs. We applied the same high-pass filter as used with the North American data and calculated the RMS strain. These stations

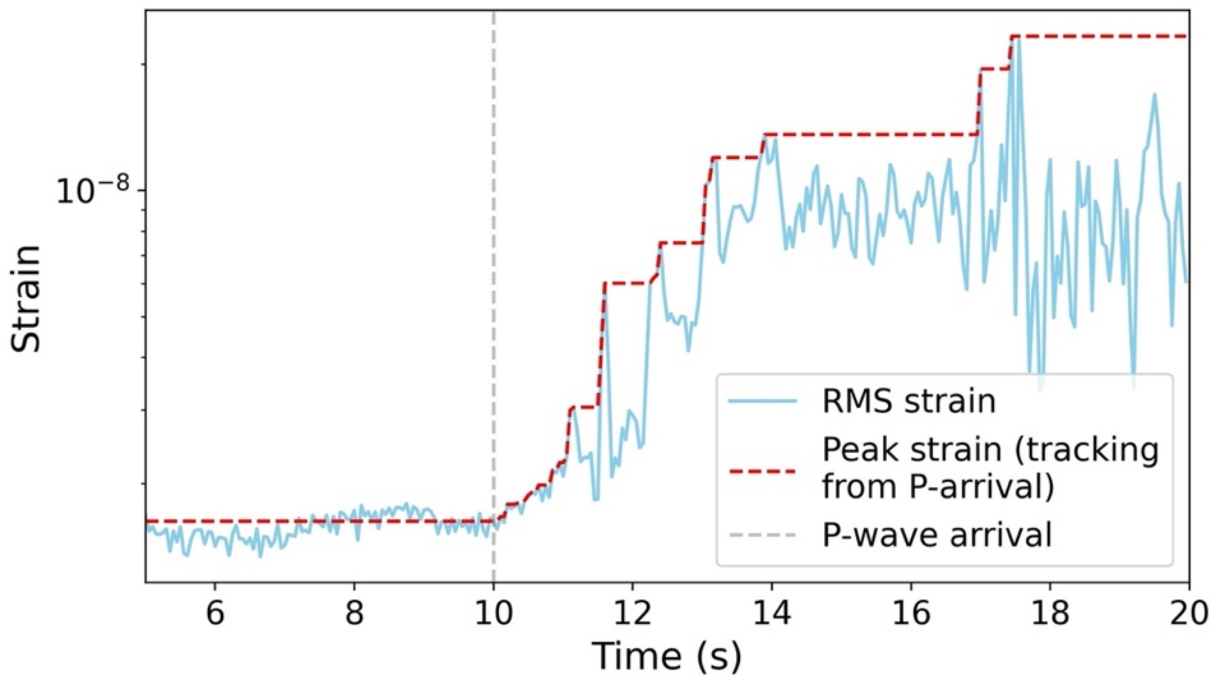


Figure 2.2. An example of how the root mean square (RMS) strain timeseries (blue) for each waveform were used to calculate the peak strain (red), where the value of the peak strain before the P-wave arrival was set to the value of the RMS strain at the P-wave arrival.

also had collocated seismometers, so we again picked P-waves on the vertical component of those stations and used these to trim the RMS strain waveforms down to two minutes with the P-wave arrival at 10 s, and from these calculated the peak strain waveforms. We did not implement any SNR or hypocentral distance focus filters for this dataset, as we did not want to eliminate any of the limited number of stations.

We also analyzed data from a set of Sacks-Evertson borehole strainmeters (Sacks et al, 1971; Hsu et al., 2015) in Taiwan for three earthquakes: the October 3, 2009 **M**6.1, October 31, 2013 **M**6.2 Ruisui earthquake (Canitano et al., 2015), and the February 4, 2018 **M**6.1 (Figure 1.1c). Of the 11 instruments in Taiwan, 8 are single-component dilatometers, and 3 are three-component strainmeters. We decided to only focus on analyzing the raw (uncalibrated) dilatational strain component from these instruments, as that was most comparable to the horizontal RMS strain calculated for the North American and Japanese data. Unlike with the horizontal component strainmeters, it was possible to pick P-waves on these data alone, likely because dilatation does incorporate the third dimension of vertical data. We implemented the same high-pass filter to these data, trimmed the waveforms by the P-wave arrival, and took the absolute value of the dilatation, as by nature the RMS calculations for the horizontal instruments also does this. The peak strain timeseries were then calculated from the absolute value of the dilatation.

2.2.2. Results

When we began visualizing the peak strain records for the 31 earthquakes in our North American dataset, we noticed that when the peak strain waveforms were plotted with the y-axis on a logarithmic scale, many of the waveforms exhibited a multi-stage growth pattern with the peak strain initially growing very quickly, and then transitioning to a more moderate growth rate.

Examples of this phenomenon for four different earthquakes are shown in Figure 2.3. For ease of comparison between the waveforms from different stations, for this figure we shifted each of the waveforms up or down so that they were aligned at the same initial value before the P-wave arrivals. The choice of 10^{-9} as the initial value best allowed for visualization of all peak strain waveforms without collapsing the smallest ones down to straight lines along the x-axis.

Figure 2.3a and 2.3b show an **M**6.8 earthquake off the coast of Cape Mendocino, California on March 10, 2014, and an **M**6.4 earthquake off the coast of Vancouver Island, British Columbia on September 9, 2011, respectively. In these two plots, each line represents the relative peak strain waveform for a single station, where the line color corresponds to the hypocentral distance for that station.

Figures 2.3c and 2.3d show the **M**7.1 Ridgecrest mainshock on July 6, 2019, and an **M**6.5 earthquake off the coast of Vancouver Island, British Columbia on April 24, 2014, respectively. In these two plots, each line still represents a single station, but color in this case corresponds to the signal to noise ratio (SNR) for that waveform. In Figures 2.3c and 2.3d, the red lines indicate waveforms with SNRs of >50 beyond the end of the colorbar. Those SNRs can be found in the legends for the three stations for which this applies.

From Figures 2.3a-d, as well as Figures 2.S1-2.S27 in section 2.9 of this chapter (Supplemental Figures) where the remainder of the earthquakes with waveforms of SNR >2 can be found (with the hypocentral distance coloring), we observed that most waveforms showed an initial fast stage of growth immediately following the P-wave arrival, and that at some point, that fast growth would slow to a more moderate rate. In Figures 2.3a and 2.3b, we observed that the height of this initial fast growth phase seemed to correlate inversely with hypocentral distance,

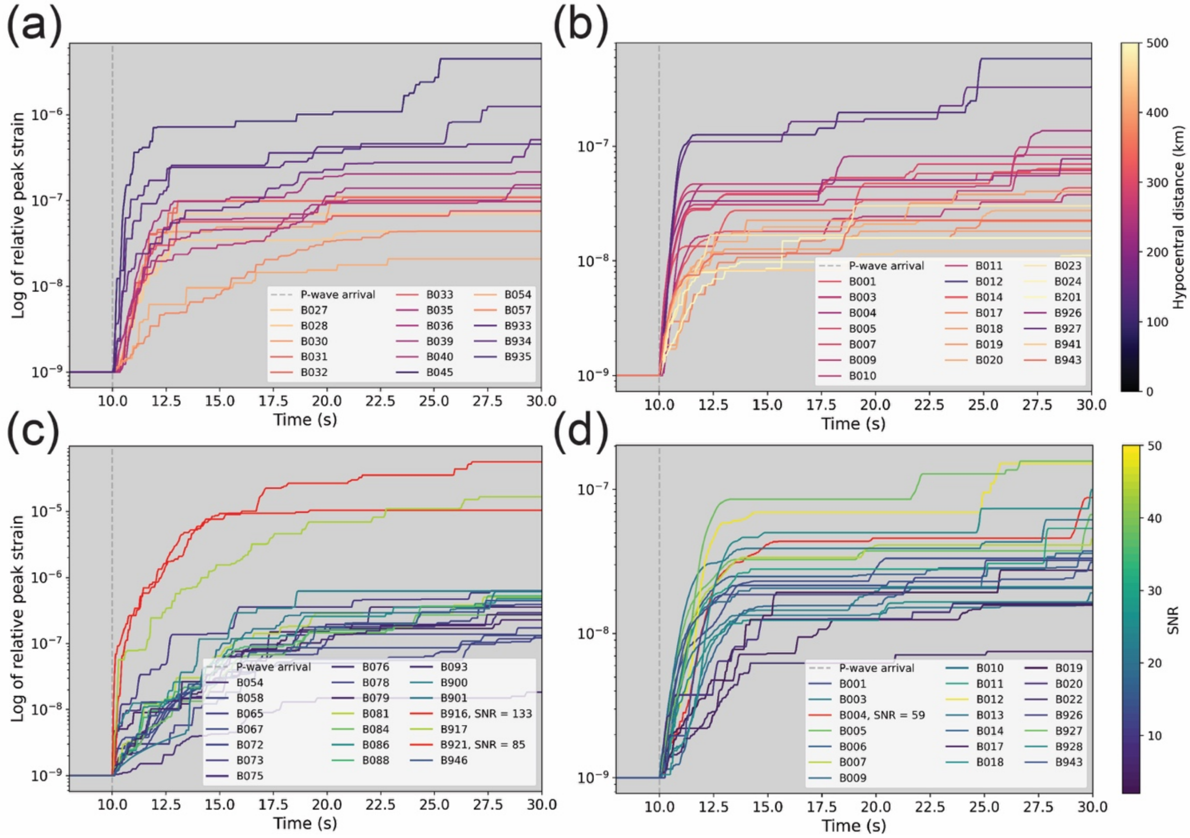


Figure 2.3. Examples of peak strain waveforms from the North American earthquakes and Gladwin Tensor Borehole Strainmeters, where each colored line represents a peak strain waveform from one station. Individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. **(a)** An **M**6.8 earthquake off the coast of Cape Mendocino, California on March 10, 2014, where waveforms are colored by hypocentral distance. **(b)** An **M**6.4 earthquake off the coast of Vancouver Island, British Columbia on September 9, 2011, where waveforms are colored by hypocentral distance. **(c)** The **M**7.1 Ridgecrest, California mainshock on July 6, 2019. Waveforms are colored by their signal to noise ratio (SNR), with the red waveforms indicating those with SNRs of >50 beyond the end of the colorbar. The SNRs for those waveforms can be found in the legend for the stations for which this applies. **(d)** An **M**6.5 earthquake off the coast of Vancouver Island, British Columbia on April 24, 2014, where waveforms are colored by their SNR.

with the stations closest to each earthquake reaching higher peak strain levels. In Figures 2.3c and 2.3d we observed that there was a positive correlation between SNR and the height of the initial fast growth phase, with the stations with the highest SNR reaching higher peak strain levels. These trends were consistent across all of the North American earthquakes (Figures 2.S1-2.S27).

We also observed this trend in the relative peak strain data from Japan and Taiwan, indicating that the observations are not likely just a result of some characteristic of the Gladwin Tensor BSM data. Figure 2.4a shows the March 3, 2011 **M**9.1 Tōhoku earthquake in Japan, Figure 2.4b shows the October 3, 2009 **M**6.1 earthquake in Taiwan, Figure 2.4c shows the October 31, 2013 **M**6.2 Ruisui earthquake in Taiwan, and Figure 2.4d shows the February 4, 2018 **M**6.1 earthquake in Taiwan. All lines on these plots again represent an individual station's peak strain waveform, and are colored by hypocentral distance. Notably, all of the Tōhoku waveforms shown in Figure 2.4a were recorded at hypocentral distances longer than the 500 km cutoff we used for the data from North America, and the colorbar is different than in Figures 2.4b-d.

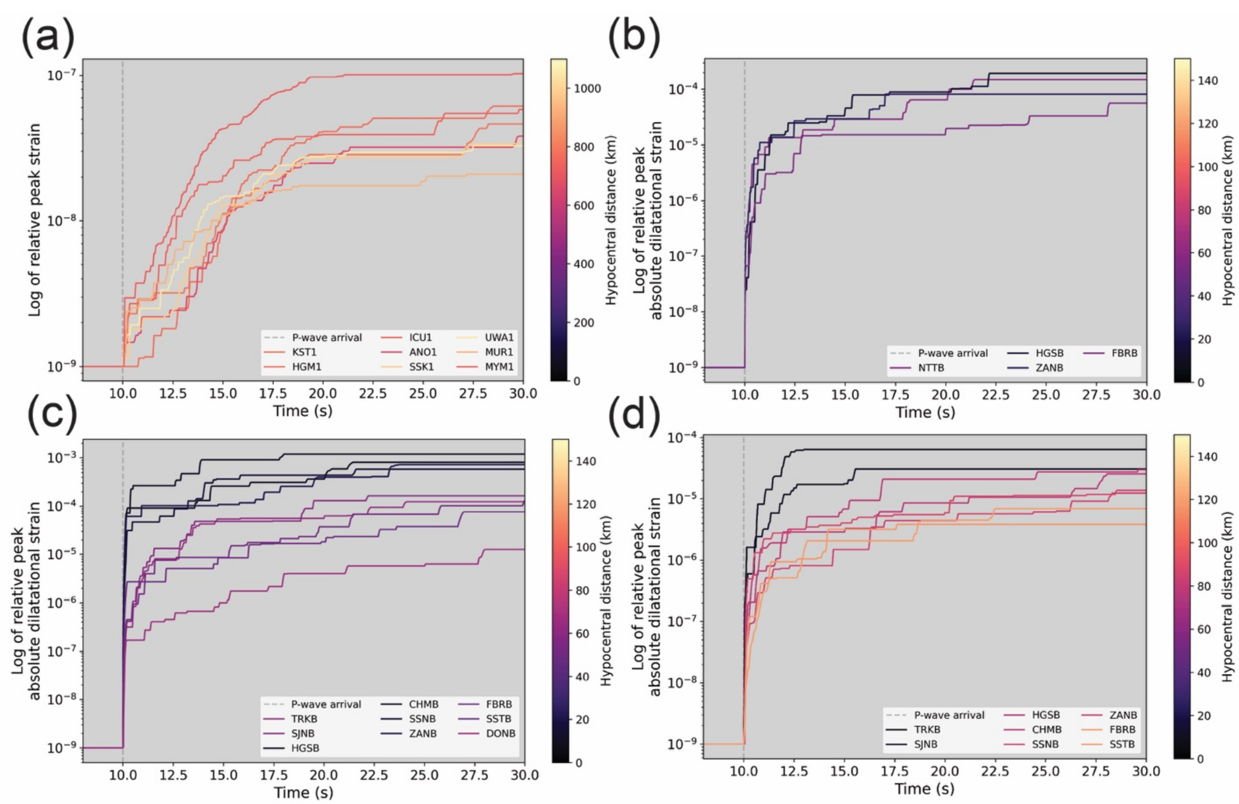


Figure 2.4. Examples of peak strain waveforms from the Asian earthquakes, where each colored line represents a peak strain waveform from one station. Individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. Waveforms in each panel are colored by hypocentral distance, but note that the colorbar scale for (a) is not the same as the scale in (b)-(d). **(a)** The **M**9.1 Tōhoku earthquake off the coast of Japan on March 3, 2011. **(b)** The October 3, 2009 **M**6.1 earthquake in Taiwan. **(c)** The October 31, 2013 **M**6.2 Ruisui earthquake (Canitano et al., 2015) in Taiwan. **(d)** The February 4, 2018 **M**6.1 earthquake in Taiwan.

After making these observations, we were interested in quantifying whether the characteristics of this multi-stage growth pattern, such as the slope of each phase and the time at which the transition between phases occurred, correlated with simple earthquake source or waveform characteristics such as the magnitude and hypocentral distance to offer a possible explanation for the pattern's appearance. We will refer to the time elapsed between the P-wave arrival and the transition between growth phases as τ. We used a Markov chain Monte Carlo (MCMC) approach facilitated by the PyMC programming library (Abril-Pla et al., 2023) where we defined a function with two straight lines connected at an intersection point and solved for the optimal slopes and location of the intersection.

Following completion of the MCMC modeling, we sorted the waveforms into categories based on the error between each line of the modeled two-straight-line fit and the corresponding section of the peak strain waveform, calculated with

$$error = \sqrt{\sum(y_{model} - y_{observed})^2}. \tag{2.3}$$

For a waveform to be considered "low error" compared to the fit by the two-straight-line model, we required an error for the first line segment to be <3.5, and for the second line segment <6.5. Because our primary interest was in the characteristics of the initial fast growth phase and when it transitioned, we chose to be stricter about the error required for the first line.

We then plotted the results of the MCMC modeling for each waveform against the source and waveform characteristics we described before, as shown in Figure 2.5. In this figure, low error waveforms are plotted with circles colored by which dataset they come from, and waveforms with high error are plotted with 'x' symbols colored similarly.

Figure 2.5a shows the relationship between earthquake magnitude and the time to transition τ, where we observed no clear correlation between the two variables. Small magnitude earthquakes have a number of low-error waveforms with long transition times as well as short ones, creating substantial scatter in the results. We plotted earthquake magnitudes versus event-averaged mean transition times and also performed simple linear regressions on several subsets of the data (Figure 2.6). Figure 2.6a shows linear regression results for the event-averaged mean transition time data (lime green), the low-error individual waveform data (black), and the data from all waveforms regardless of error (red). We found correlation coefficients for the low-error and all waveform datasets of 0.24 and 0.23, and a correlation coefficient for the event-averaged mean data of 0.39,

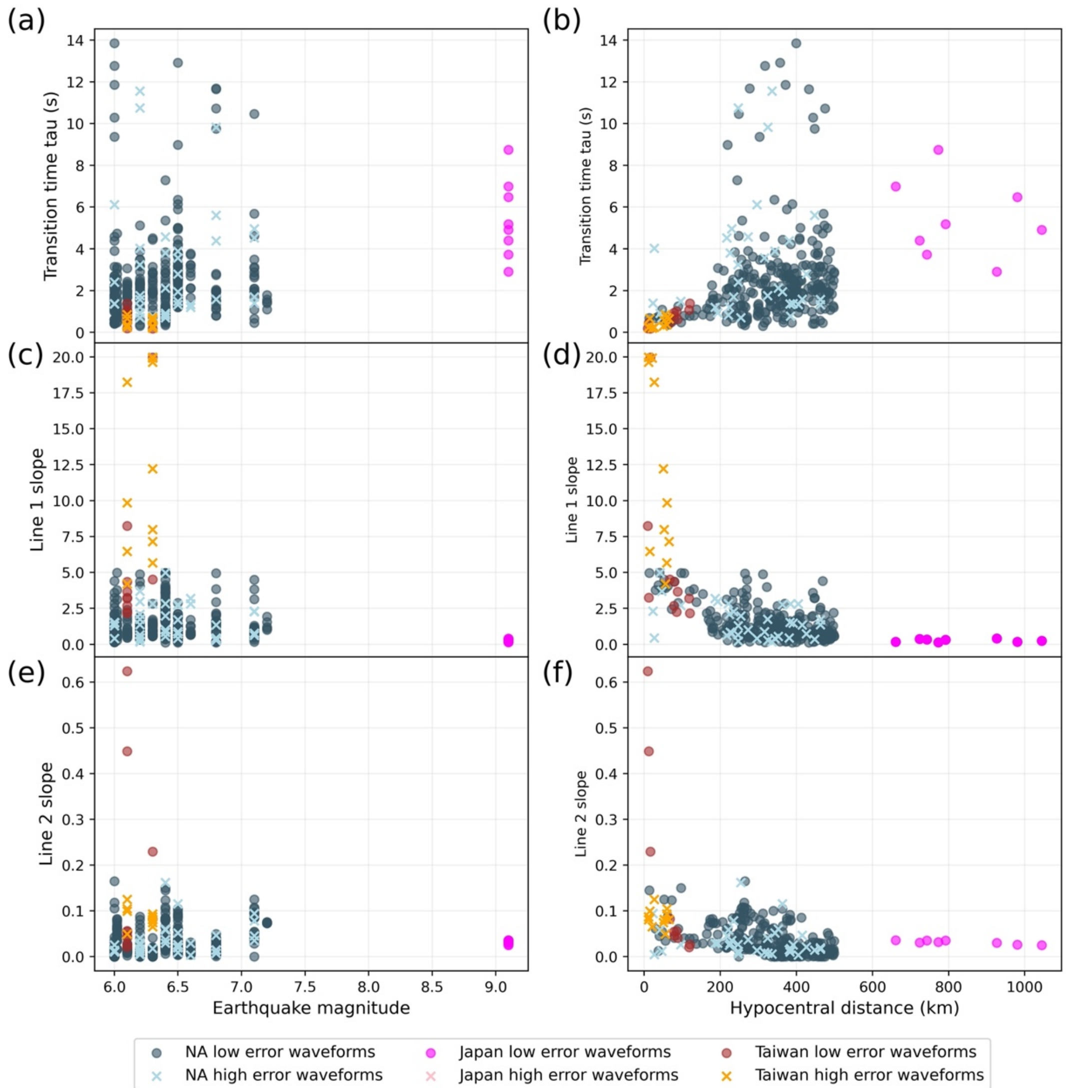


Figure 2.5. Results of the Markov chain Monte Carlo (MCMC) inversions for the transition point from the fast initial growth phase of the peak strain waveforms and the slower growth phase and the slopes of the two growth stages plotted against the earthquake magnitudes and station hypocentral distances. For each of the subplots, the results for the North American (NA) waveforms with low error between the simulated data and the MCMC fit are the dark blue circles, NA waveforms with high error are the blue 'x' symbols, Tōhoku waveforms with low error are the pink circles, Tōhoku waveforms with high error are the pink 'x' symbols, Taiwan waveforms with low error are the brown circles, and Taiwan waveforms with high error are orange 'x' symbols. **(a)** Earthquake magnitude versus the time (s) it takes to transition to slower growth after the arrival of the P-waves (τ). **(b)** Hypocentral distance (km) versus τ. **(c)** Earthquake magnitude versus the slope of the first line (fast growth phase). **(d)** Hypocentral distance (km) versus the slope of the first line (fast growth phase). **(e)** Earthquake magnitude versus the slope of the second line (slower growth phase). **(f)** Hypocentral distance (km) versus the slope of the second line (slower growth phase).

suggesting a weak to moderate correlation at best. However, we suspected that the MCMC results from the Tōhoku waveforms were not fairly comparable to those from the North American and Taiwanese data as the hypocentral distances were much longer (>700 km), which may increase the influence of the seismic wave path on the peak strain waveform characteristics. When the Tōhoku waveforms were removed from the dataset and only the North American and Taiwanese transition time results were used in linear regression (Figure 2.6b), the correlation coefficients decreased to 0.11 and 0.12 for the low-error and all waveform datasets and 0.29 for the event-averaged mean data, weakening the strength of any potential correlation further. This analysis should be extended to a lower magnitude range to solidify this interpretation.

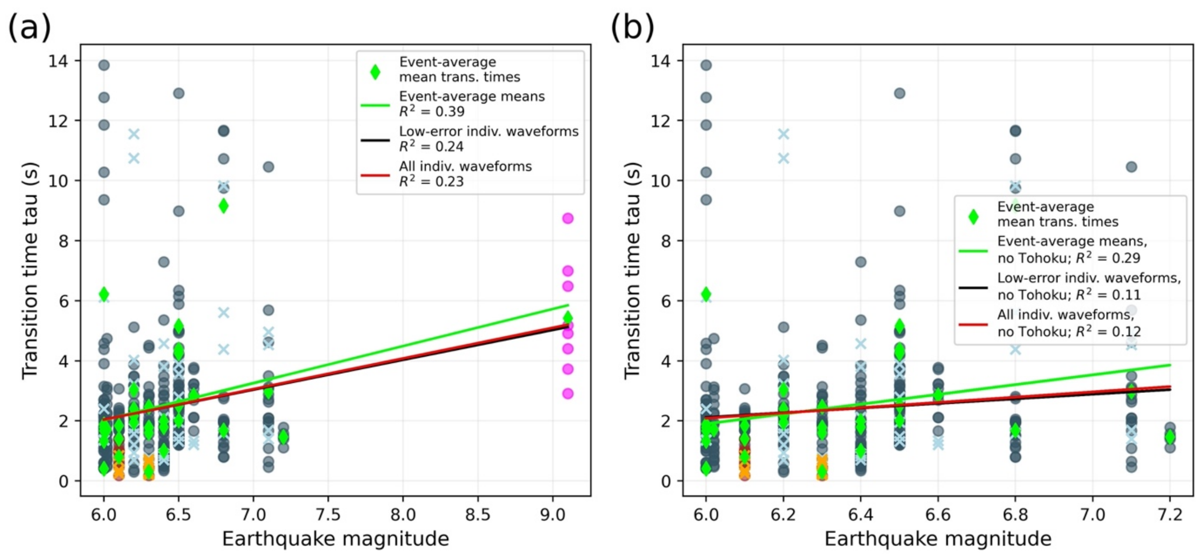


Figure 2.6. The same MCMC scatter plot results for earthquake magnitude versus time to transition between peak strain growth phases as shown in Figure 2.5a, with the addition of event-average mean transition times shown by the lime green diamond symbols and linear regression results. **(a)** Earthquake magnitude versus the time (s) it takes to transition to slower growth after the arrival of the P-waves (τ) for all earthquakes in the study between **M**6 and **M**9.1. The green trendline represents the results of a linear regression using only the event-average mean transition times, the black trendline represents the results of a linear regression using the individual results for only the low-error waveforms, and the red trendline represents the results of a linear regression using all individual waveforms, regardless of error. **(b)** Earthquake magnitude versus the time (s) it takes to transition to slower growth after the arrival of the P-waves (τ) for all earthquakes in the study except the **M**9.1 Tōhoku earthquake. The trendlines are the same as in (a), but calculated without inclusion of the Tōhoku earthquake.

Figure 2.5b shows the relationship between the waveform hypocentral distance and τ. A positive correlation is observed with this data, where waveforms recorded on stations with small hypocentral distances tend to have shorter τs with an approximately linear relationship. These data are less scattered at short hypocentral distances than the results shown in Figure 2.5a for smaller magnitude earthquakes, making the trend more apparent, but there is some scatter in the results for the North American data at intermediate (~200-500 km) hypocentral distances.

Figure 2.5c plots the earthquake magnitude versus the slope of the first line of the MCMC inversion, which represents the first phase of growth. There appears to be no correlation between these variables. The Taiwanese data (brown and orange) tend to have larger slopes than the other data, which may be attributed to the small hypocentral distances available in that dataset. This is shown in Figure 2.5d, which plots the waveform hypocentral distance versus the slope of the first phase of growth. In this case, we observed that the slope of the first phase is inversely proportional to the hypocentral distance, where at shorter hypocentral distances the first growth phase slope is steeper than at longer hypocentral distances. This relationship appears to be nonlinear and more representative of exponential decay.

Figure 2.5e shows the relationship between earthquake magnitude and the slope of the second, slower phase of growth, and there does not appear to be a correlation between these variables. Figure 2.5f shows the relationship between hypocentral distance and the slope of the second phase, and it appears that the variables are inversely related. Similar to the relationship between hypocentral distance and the slope of the first phase of growth shown in Figure 2.5d, waveforms from short hypocentral distances also tend to have steeper second phases as well.

The lack of strong correlation between earthquake source properties like magnitude and the slope of the lines or transition time does not allow us to conclude that strong determinism is present. To explore whether or not we should have expected to find it, we turned to modeling experiments.

2.3. Modeling

As strainmeter data are not commonly used to study earthquake source processes, to understand what our real strain recordings showed we also simulated earthquake ruptures using the SW4 finite-difference wave propagation code (Petersson and Sjogreen, 2017) to forward-model for strain data for several different scenarios. Our first goal was to verify whether the strain waveforms produced by SW4 simulations were similar in character to those that we observed in the real data by simulating a complex heterogeneous rupture, the 2019 **M**7.1 Ridgecrest earthquake. We then simulated a series of simple homogeneous ruptures designed to be strongly deterministic to explore whether the analysis methodology we used would in fact capture these characteristics.

2.3.1. Heterogeneous Rupture Model: The M7.1 2019 Ridgecrest Earthquake

2.3.1.1. Methods

The **M**7.1 2019 Ridgecrest, California mainshock made for a useful case study for SW4 simulation because its rupture process has been studied in detail. We made use of the complex rupture model for the mainshock of the Ridgecrest sequence from Goldberg et al. (2020) for this simulation, which was developed through joint inversion of multiple types of data from the earthquake. We converted Goldberg et al.'s rupture inversion, which consisted of 624 subfaults across 4 fault segments (Figures 2.7a-b), to Standard Rupture Format (SRF), which can be used in SW4 as a rupture input.

We defined a rectangular model domain for the SW4 simulation which extended across Southern California from approximately 33° to 37°N latitude and 116° to 121°W longitude, and contained a total of 162 receiver stations (Figure 2.7c). These stations were placed at the locations of the 8 real BSM stations and 154 real GNSS stations within the model domain. The BSM-like receiver stations captured waveforms of the 6 components of the symmetric strain tensor and 3 components of displacement, and the GNSS-like receiver stations only captured the 3 components of displacement.

We began with the Mojave area 1D velocity model from the Southern California Earthquake Center's (SCEC) Broadband Platform v15.3.0 (Maechling et al., 2015), but because of the 5 thin surface layers of thickness 2, 4, 6, 8, and 10 m, the computation time for the SW4 simulation was unreasonable. We instead simplified the top 5 layers into one 30 m thick layer whose properties were an average of the original 5. The final grid spacing of the model domain was 810 m. The simulation was run on Talapas, the University of Oregon's high-performance computing (HPC) cluster.

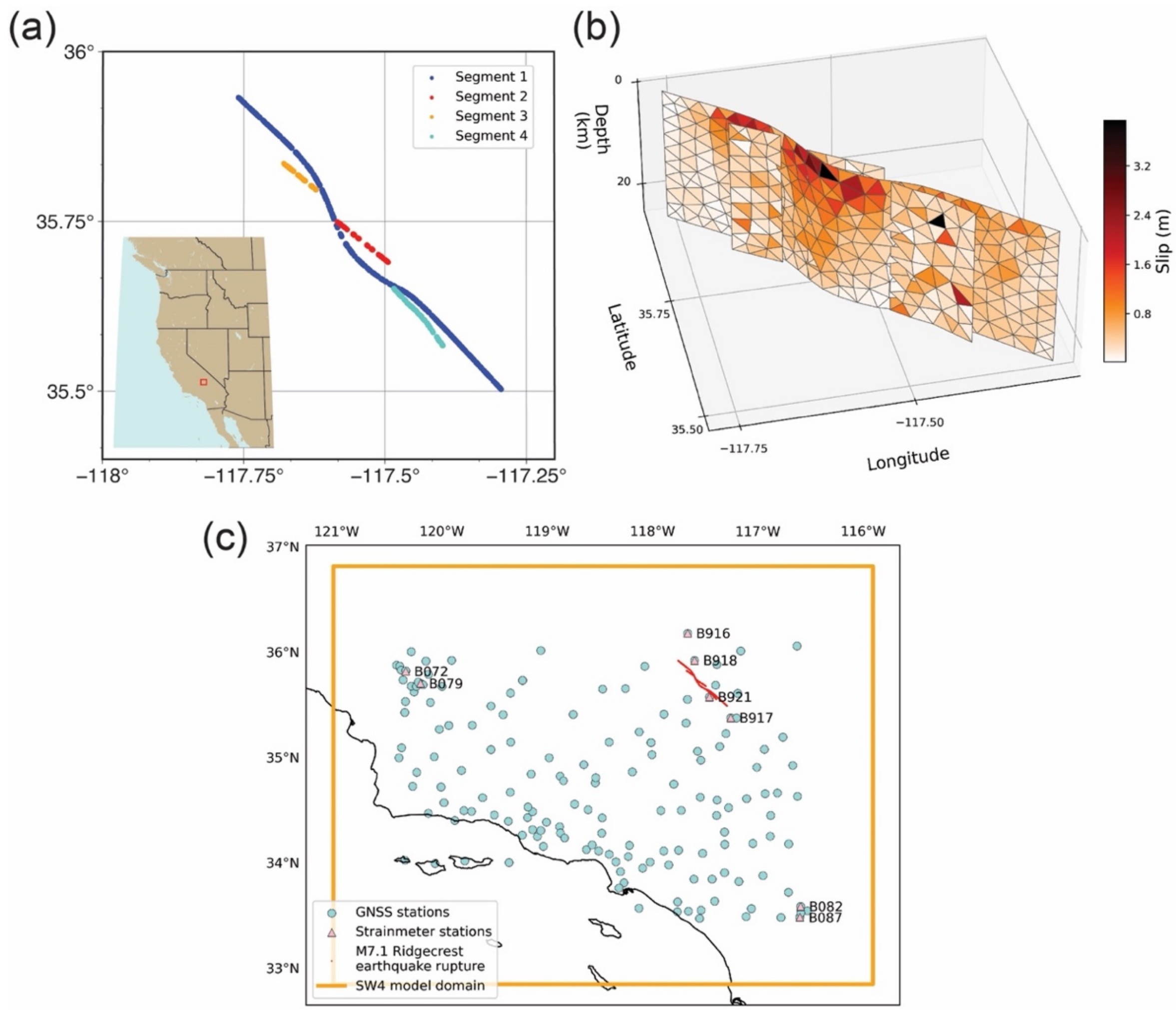


Figure 2.7. (a) Map-view of the four segments of the fault that ruptured during the **M**7.1 Ridgecrest earthquake from the inversion model of Goldberg et al. (2020), which we used to simulate the rupture in SW4. The points mark each individual subfault and are colored by the segment of the fault that they belong to. **(b)** 3-D view of the same fault where each triangle represents an individual subfault and is colored by the amount of slip on that subfault in the rupture model. **(c)** Map-view of the entire model domain (bounded by the orange box) used to simulate the **M**7.1 Ridgecrest earthquake in SW4. The fault model (red), locations of strainmeter receivers (pink triangles), and locations of GNSS receivers (blue circles) are shown. These receiver locations match the locations of true NOTA BSMs and GNSS stations.

After computation, the strain waveforms were processed similarly to the North American real data that we analyzed, with a few differences. No capacitance to strain conversion was required because the SW4 output was strain already, but the same causal high-pass filter was applied. The SW4 RMS strain $\boldsymbol{E}_{SW4}$ was calculated with the $\boldsymbol{e}_{xx}$, $\boldsymbol{e}_{xy}$, and $\boldsymbol{e}_{yy}$ components of the symmetric strain tensor as

$$E_{SW4} = \sqrt{(\boldsymbol{e}_{xx}^2 + \boldsymbol{e}_{xy}^2 + \boldsymbol{e}_{yy}^2)/3}. \tag{2.4}$$

To facilitate direct comparison of the simulated Ridgecrest earthquake waveforms to the real ones, we also needed to add noise to the data that SW4 produced. To do this, we downloaded additional real NOTA BSM data from the stations within the SW4 model domain – specifically, from 2 minutes before the **M**7.1 earthquake origin time until the origin time. This captures for each station a time window with the same background noise that each strainmeter would have seen before the **M**7.1. The only exception was for one station, B918, which was not operational during the **M**7.1 Ridgecrest earthquake. Instead, its noise timeseries came from the time period before another earthquake in the dataset on August 29, 2019, as the station was operational at that time. These noise timeseries were then processed the same way the other Gladwin Tensor BSM data were with the conversion to linear extensional strain, high-pass filtration, and calculation of RMS strain. These RMS strain noise timeseries were then added to the RMS strain waveforms produced in the SW4 simulations, and then the peak strain was calculated as with the real data.

Because we defined a number of stations which captured displacement data to simulate GNSS stations, we also computed RMS and peak horizontal displacement timeseries. The SW4 RMS displacement $\boldsymbol{D}_{SW4}$ was calculated using the two horizontal displacement components $\boldsymbol{d}_x$ and $\boldsymbol{d}_y$ as

$$D_{SW4} = \sqrt{(\boldsymbol{d}_x^2 + \boldsymbol{d}_y^2)/2}. \tag{2.5}$$

2.3.1.2. Results

We plotted the SW4 peak strain waveforms with noise for the simulated stations in the real locations of the Gladwin Tensor BSMs, as well as the real data from those stations from the actual Ridgecrest earthquake (Figure 2.8a-e). The panels in Figure 2.8 each show the real peak strain data (red) and SW4 simulated peak strain data (blue) for the five stations for which real data were available, and the simulated stations at those same locations. The waveforms have been shifted laterally so that the P-wave arrivals both align at 10 seconds (gray dashed line) like the previous real data analysis. From these plots we observed that the real and simulated waveforms are of comparable order of magnitude and shape for each station, suggesting that the earthquake simulation produces results that are comparable to real data.

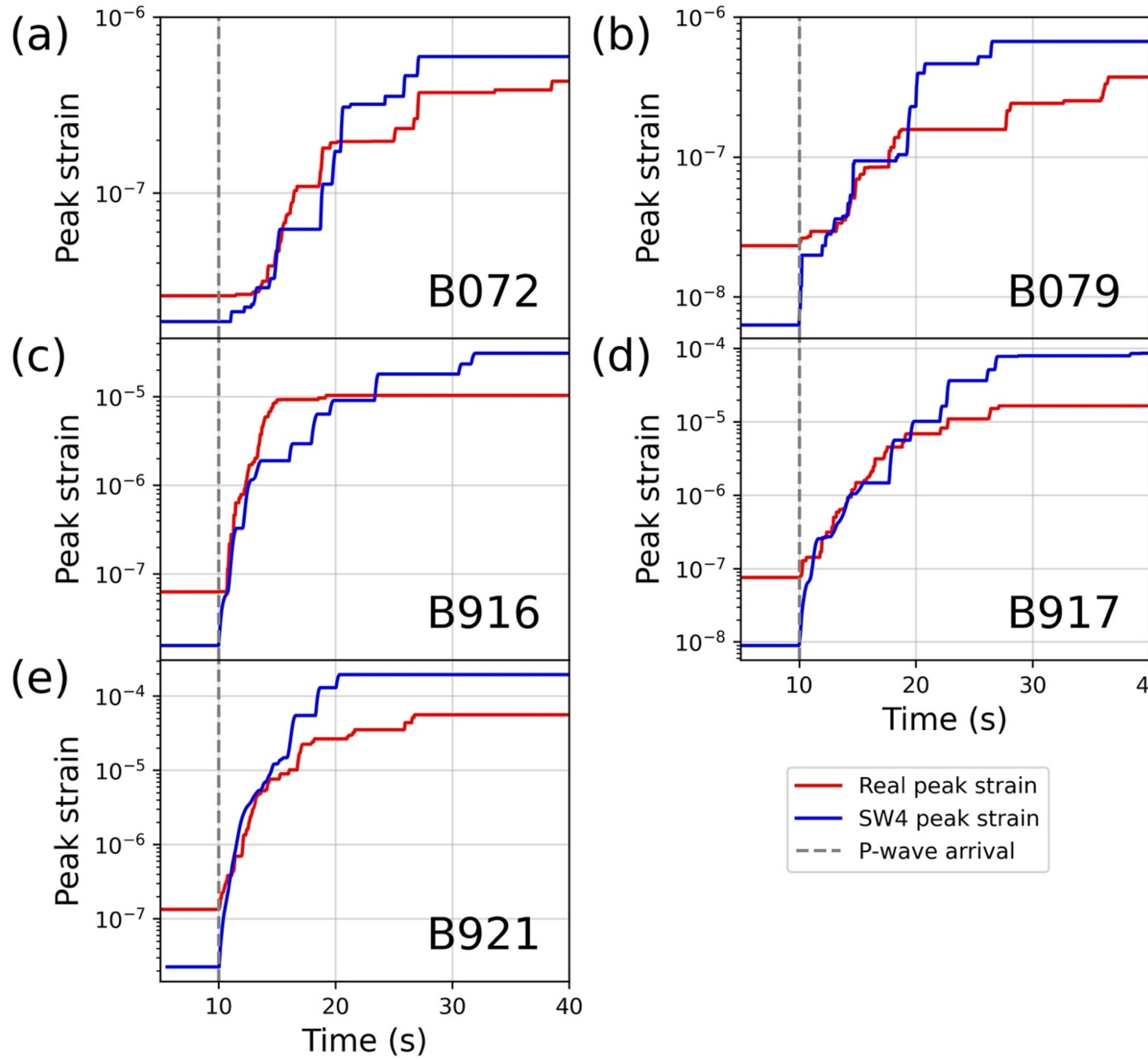


Figure 2.8. Plots of the real peak strain and peak strain from the SW4 simulations of the **M**7.1 Ridgecrest earthquake for the 5 nearby borehole strainmeter stations that recorded data during the real earthquake. Panels **(a)**-**(e)** show the waveforms from stations B072, B079, B916, B917, and B921, respectively. The real peak strain data are shown in red, and the peak strain data from the SW4 simulations are shown in blue. The waveforms are aligned to a common P-wave arrival time, shown by the gray dashed line.

Given these results, we wanted to explore whether the lack of strong determinism found in the real data was because earthquakes of different magnitudes truly do not start differently, or if these features might still be obscured by noise or an instrument effect.

2.3.2. Deterministic Homogeneous Rupture Models

2.3.2.1. Methods

We simulated several additional earthquakes in SW4 that we designed to be strongly deterministic, where earthquakes of different magnitudes would be able to be distinguished very early (within seconds) of P-wave data arrival. Early studies which suggested that earthquakes were strongly deterministic proposed a physical model where earthquakes that initiate with higher amounts of slip would be more likely to evolve into a large earthquake because there was more energy available to "push" additional parts of the fault into rupture and continue its propagation (Ellsworth and Beroza, 1995; Olson and Allen, 2005). We also know from more recent studies that the average rise time of an earthquake, or the average amount of time that any particular patch of a fault is actively slipping, scales with magnitude (Melgar and Hayes, 2017).

We built 4 homogeneous rupture models, where for each earthquake with a magnitude of 5, 6, 7, or 8, the slip amount and rise time (τ_r) at each subfault were identical across the fault, including at the hypocenter where rupture initiated. We scaled these characteristics (as well as rupture length) with the magnitude to create strongly deterministic earthquakes (Table 2.1). For these experiments, we utilized a simplified single-plane strike-slip fault with a strike of 139°, matching the strike of the Ridgecrest **M**7.1 fault from the USGS finite fault solution of the earthquake. We used a dip of 90° for perfectly vertical strike-slip faults, and a fault depth of 15 kilometers corresponding to the average seismogenic depth in Southern California (Nazareth and Hauksson, 2004). We continued to use the same SCEC 1D Mojave velocity model that we used for the heterogeneous Ridgecrest simulation.

Table 2.1. Fault and rupture parameters for the SW4 deterministic homogeneous earthquake simulations. The rupture length and average slip calculations were made using the scaling relationships for strike-slip faults from Wells and Coppersmith (1994). The rise time calculation was made using the scaling relationship from Melgar and Hayes (2017).

Moment magnitude	Rupture length (km)	Average slip (m)	Rise time (s)
5	3.39	0.015	0.336
6	14.13	0.120	0.924
7	58.88	0.955	2.542
8	245.47	7.586	6.993

Slip on fault patches occurs between the arrival of the rupture front expanding outward from the hypocenter and the healing front that follows it (Figure 2.9a), and as previously mentioned the duration of this slip is the rise time. In physical space, this correlates to the width of an earthquake's slip pulse. Figure 2.9b shows the geometry of our **M**7 homogeneous model in 3-D as an example, where the subfault color indicates the rupture onset time in seconds. We calculated the expected rise times for moment magnitudes 5, 6, 7, and 8 earthquakes following the relation from Melgar and Hayes (2017), where

$$log_{10}(\tau_R) = -5.323 + 0.293\, log_{10}(M_0), \tag{2.6}$$

with τ_R in seconds and M_0, the source moment, in Newton-meters. We calculated M_0 (N-m) with

$$M_0 = 10^{(1.5\boldsymbol{M}+9.05)} \text{ (Hanks and Kanamori, 1979)}, \tag{2.7}$$

where $\boldsymbol{M}$ is the moment magnitude.

Subsurface rupture length and average slip amount also scale with magnitude, so we calculated the expected values for our earthquakes using the scaling relationships from Wells and Coppersmith (1994) for strike-slip earthquakes. Subsurface rupture length (*RLD*) in kilometers was calculated with

$$log_{10}(RLD) = -2.57 + 0.62 * \boldsymbol{M}, \tag{2.8}$$

where $\boldsymbol{M}$ is the moment magnitude, and average slip amount (*AD*) in meters was calculated with

$$log_{10}(AD) = -6.32 + 0.9 * \boldsymbol{M}, \tag{2.9}$$

where $\boldsymbol{M}$ is still the moment magnitude. A summary of each of the rupture parameters used can be found in Table 2.1.

The faults all shared a common hypocenter geographically in the center of the faults, with a depth of 7.5 km (half of the seismogenic depth). We set up the faults with 100 subfaults along-strike (so the spacing varied by earthquake magnitude, and therefore fault length), and used 30 subfaults along-dip spaced every 0.5 km for all four earthquakes. We then generated a synthetic rupture for each of the earthquake magnitudes beginning at the common hypocenter using FakeQuakes/MudPy, a kinematic rupture forward modeling code (Melgar et al., 2016). This allowed us to compute the subfault slip onset times while using the prescribed fault length and rise time. From these rupture models we made the slip homogeneous across each fault, with each subfault slipping the same amount according to the Wells and Coppersmith (1994) scaling relation, and then converted the rupture models to SRF format to be used in SW4.

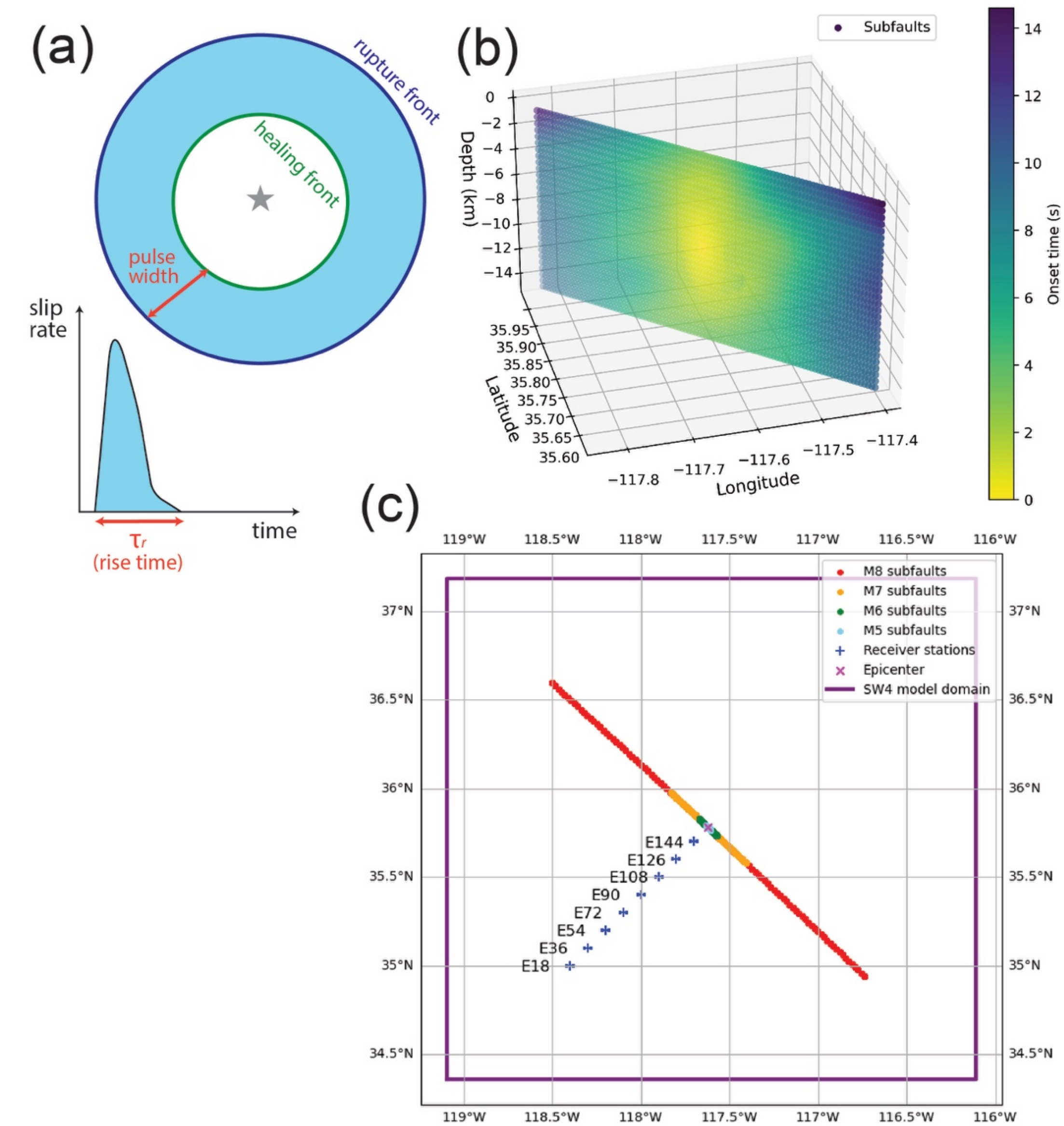


Figure 2.9. **(a)** Illustration of a homogeneous earthquake rupture model, where the rupture begins at the hypocenter (gray star). The rupture front expands outward in a uniform ring, followed by the healing front. The rise time (τ_r) is the amount of time that any patch of the fault is actively slipping between the arrivals of the rupture and healing fronts, which in physical space correlates to the width of the slip pulse. **(b)** 3-D view of the planar fault used for the **M**7 deterministic homogeneous rupture model experiment as an example, with the subfaults colored by the time of onset of slip on each subfault. Subfault rupture start times are modulated by the velocity model used rather than being a perfect uniform ring. **(c)** Map-view of the entire model domain (bounded by the purple box) used to run the deterministic homogeneous rupture model experiments in SW4. The fault models for the **M**5-8 earthquakes are indicated by colored dots plotted at the location of each subfault within each fault model are shown, as well as the locations of the selected strainmeter receivers (blue crosses) from which data were analyzed and the common hypocenter (epicenter in this map view; magenta 'x').

The shared SW4 model domain for all four ruptures was appropriately sized to encapsulate the longest fault for the **M**8 earthquake, extending from approximately 34.4° to 37.2°N latitude and 116.1° to 119.1°W longitude. We established a grid of receiver stations across the domain capturing strain waveforms spaced 0.1° apart, and picked several in a line perpendicular to the fault planes beginning near the epicenter from which to analyze data (Figure 2.9c). We used the same modified 1D velocity model as used in the heterogeneous complex **M**7.1 Ridgecrest simulation, with an 810 m simulation grid spacing. The models were then also run with SW4 on the Talapas HPC cluster.

After computation, the strain waveforms were processed the same to calculate RMS and peak strain as they were for the heterogeneous simulation, but initially without the addition of the real BSM noise or the high-pass filter because we were more interested in comparing these simulation results to each other than to the real data. RMS and peak displacement data were computed similarly. We then ran the same two-straight-line MCMC modeling on the peak strain data to quantify the slopes of the two growth stages as well as the transition time, and the same error sorting procedure as described with the real strainmeter data.

2.3.2.2. Results

Figures 2.10a-d show the peak strain waveforms for four of the eight stations in the line perpendicular to the center of the faults, with the hypocentral distances of the stations being 11 km, 40 km, 68 km, and 97 km respectively. As with Figures 2.3 and 2.4 for the real data, the waveforms were all shifted vertically so that the pre-P-arrival strain level was 10^{-9} for ease of comparison. On each of the plots, the **M**8 earthquake is in red, the **M**7 is in orange, the **M**6 is in green, and the **M**5 is in blue.

We observed that the strain waveforms are similar in general shape for each of the earthquake magnitudes, and that the larger magnitude earthquakes grew to higher peak strain levels than the smaller magnitude earthquakes with time. Upon close inspection of the first few tenths of seconds past the P-wave arrivals, it appeared that the slope of the initial growth phase of the peak strain was steeper for the higher magnitude earthquakes, and we also saw that the transition corner appeared to be later for the higher magnitude earthquakes. To quantify these observations, we ran the same MCMC inversion procedure as we used for the real data, and sorted the waveforms by high and low error the same way as well.

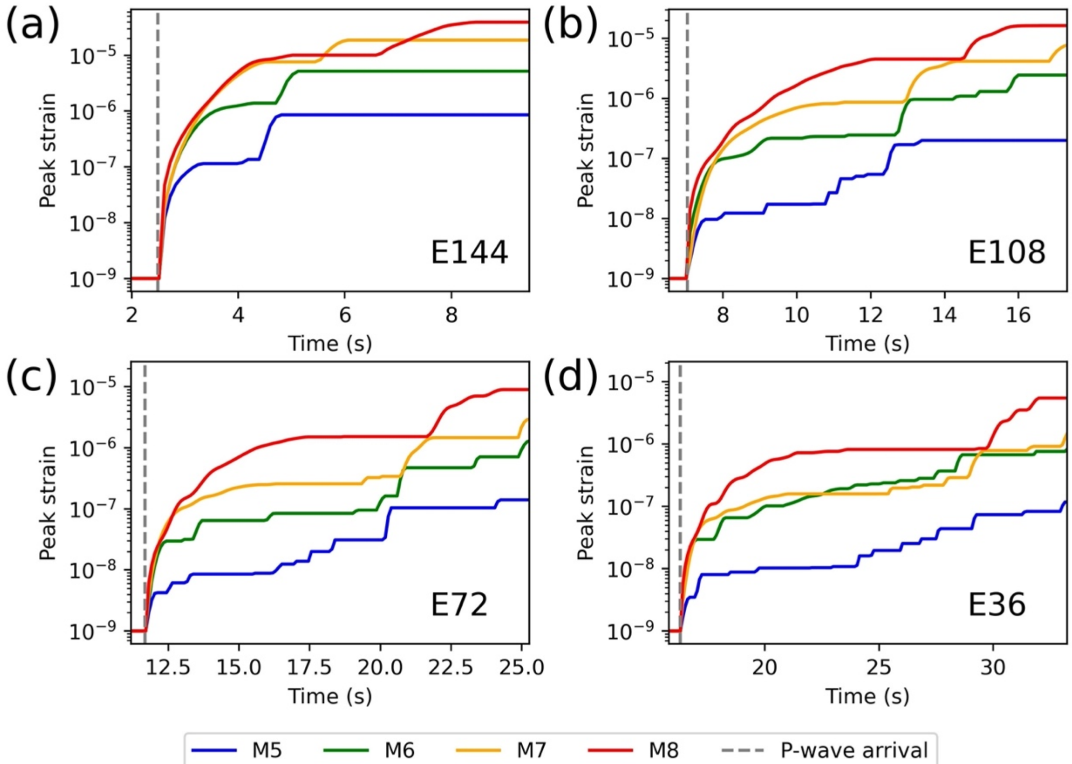


Figure 2.10. Examples of peak strain waveforms from the deterministic-designed earthquakes simulated in SW4, where each colored line represents the peak strain waveform from one station colored by the magnitude of the earthquake that waveform came from. Individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. **(a)** Waveforms from station E144, which has a hypocentral distance of 11 km. **(b)** Waveforms from station E108, which has a hypocentral distance of 40 km. **(c)** Waveforms from station E72, which has a hypocentral distance of 68 km. **(d)** Waveforms from station E36, which has a hypocentral distance of 97 km.

The results of these inversion are shown in Figure 2.11, which is analogous to Figure 2.5 in format. Black circles show the results for waveforms with low error between the data and the MCMC fit, and red 'x' symbols represent waveforms with high error. Figure 2.11a shows the time to the transition point between slopes versus the earthquake magnitude. In this figure, as opposed to in Figure 2.5a for the real data, particularly if the points for waveforms with high error are disregarded we do observe an exponential positive correlation between earthquake magnitude and transition time, where the larger the magnitude is the longer it takes for the fast growth period to end. This is expected given that we designed the earthquake ruptures to be strongly deterministic. We also found that the transition times for even the **M**8 earthquake in our deterministic experimental results stayed shorter than ~3 s, whereas for earthquakes as small as **M**6 in the real data, transition times sometimes exceeded 10 s. Noise and the differences in hypocentral distance ranges between these two tests may account for this difference.

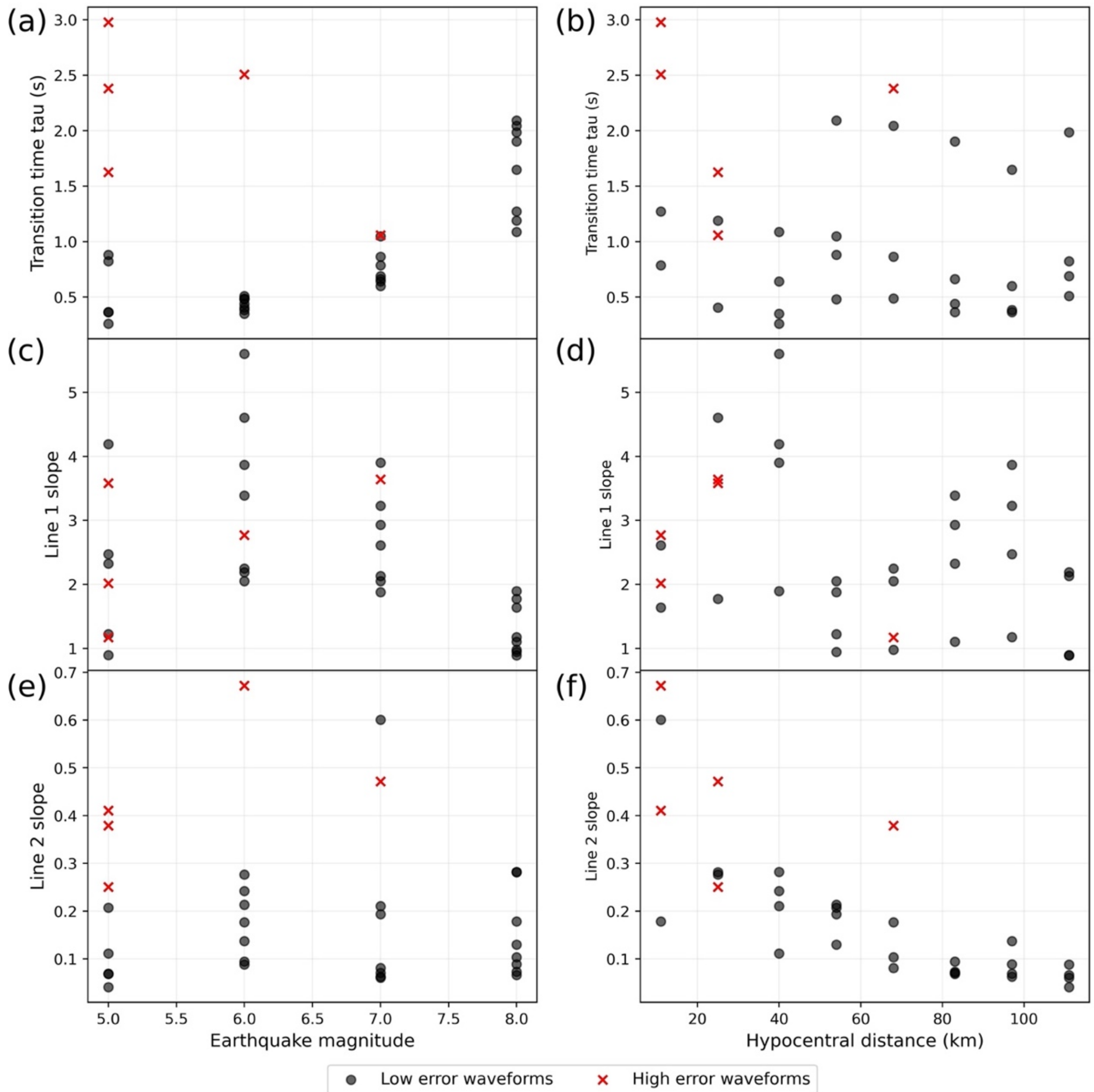


Figure 2.11. Results of the MCMC inversions on the deterministic-designed earthquakes simulated in SW4, where we inverted for the transition point from the fast initial growth phase of the peak strain waveforms and the slower growth phase and the slopes of the two growth stages plotted against the earthquake magnitudes and station hypocentral distances. For each of the subplots, the results for the waveforms with low error between the simulated data and the MCMC fit are the black circles, and the results for the waveforms with high error are the red 'x' symbols. **(a)** Earthquake magnitude versus the time (s) it takes to transition to slower growth after the arrival of the P-waves (τ). **(b)** Hypocentral distance (km) versus τ. **(c)** Earthquake magnitude versus the slope of the first line (fast growth phase). **(d)** Hypocentral distance (km) versus the slope of the first line (fast growth phase). **(e)** Earthquake magnitude versus the slope of the second line (slower growth phase). **(f)** Hypocentral distance (km) versus the slope of the second line (slower growth phase).

In Figure 2.11b, which shows the waveform hypocentral distance versus time to transition, we observe that there is no apparent correlation, in opposition to the positive correlation seen in Figure 2.5b. However, it should be noted that the hypocentral distance range in this figure is much smaller than the range shown for the real data in Figure 2.5b.

Figure 2.11c plots the slope of the fast growth phase against earthquake magnitude, and in this figure we observe a pattern that appears to follow an inverted quadratic form, where the slope is smaller for the **M**5 and **M**8 earthquakes than the **M**6 and **M**7 earthquakes. We did not observe this same pattern in the real data, as shown in Figure 2.5c. In Figure 2.11d where the fast growth slope is plotted versus the hypocentral distance, we again do not observe a correlation, in contrast to the negative correlation seen with the real data in Figure 2.5d.

In Figure 2.11e we plotted the slope of the second, slower phase of growth against the earthquake magnitude, and do not observe a correlation, similar to the results for the real data as shown in Figure 2.5e. In Figure 2.11f, however, where we plotted the slope of the second phase versus the hypocentral distance, we do observe a negative correlation between slope and distance, where the stations farther from the hypocenter exhibit shallower slopes for the second phase. This matches the observation with the real data as shown in Figure 2.5f, despite the lack of correlation between the results for the first slope versus hypocentral distance discussed previously.

In this modeling experiment where we designed the earthquakes to be strongly deterministic, with different magnitudes distinguishable early on in the rupture process, as we expected we did observe a difference particularly in the time to transition from fast to slow growth by magnitude, in contrast to the results from the real data analysis.

2.4. Discussion

Our study aimed to explore whether strain data might provide new insight into the question of whether earthquakes are strongly deterministic, and therefore whether strainmeter data may provide a novel method for rapid earthquake characterization, which would have implications for earthquake early warning and rapid response. When we simulated earthquakes designed to be strongly deterministic and analyzed the resulting waveforms, we observed that the timing of a change in regime of strain growth strongly depended on magnitude (Figure 2.11). However, when this same analysis was performed on real strain waveforms from three different types of borehole strainmeters and across many different type of earthquakes, we did not observe the same pattern

(Figure 2.5). While a weak correlation may have been observed when time to peak strain phase transition was plotted against magnitude (Figure 2.6), there was significant scatter in the transition time data from waveforms from the same magnitude earthquakes, particularly at the low end of our magnitude range. Additionally, the high end of our magnitude range consisted of data from only one earthquake, and waveforms only from very far hypocentral distances (>700 km), which makes them harder to fairly compare to the other events in our dataset. We also have no data for a large magnitude gap in between the single large earthquake and those smaller than it. From these observations, we ultimately cannot conclude that earthquakes are strongly deterministic. As with seismic data, we do not appear to be able to use strain data to clearly distinguish the magnitudes of different earthquakes with only early P-wave data.

In the real data, it appears that hypocentral distance is the main control on the features of the peak strain waveforms rather than magnitude, as the only qualitatively strong correlations seen in Figure 2.5 occur when the waveform characteristics are plotted against hypocentral distance. However, our dataset is skewed towards the lower end of this study's magnitude range (**M**6-9.1) and smaller hypocentral distance earthquakes. The only dynamic strain data we had access to above **M**7.2 was the **M**9.1 Tōhoku earthquake, and those stations were no closer than ~700 km to the hypocenter. It is possible that data from close hypocentral distances for other large earthquakes in our magnitude gap would provide additional information that we cannot currently see. In future work, the analysis we performed should also be extended to a lower magnitude range than **M**6 to see if the same patterns (or lack thereof) hold. The weak correlation demonstrated in Figure 2.6 may become much more or less clear if smaller magnitude earthquakes are included in a dataset, and utilizing smaller earthquakes may also give us access to more onshore events closer to strainmeters, expanding data availability at short hypocentral distances.

There are two other significant differences between our deterministic SW4 simulations and the real data: the presence of noise in the real strain data, and the lack of the high-pass filter that was implemented with the real data. The real P-wave data characteristics that we analyzed should not be substantially obscured by noise considering that, for example, the North American Gladwin-Tensor BSMs in Southern California have lower or similar SNRs to broadband seismometers at the frequencies where seismic waves are recorded (Barbour and Angew, 2012). However, we decided to test adding the same real strain noise to the deterministic earthquake waveforms as used in the heterogeneous **M**7.1 Ridgecrest earthquake simulation to see whether the patterns found in the MCMC results still held. Because we do not have real stations in the locations of the simulated receiver stations used for the deterministic ruptures, we chose randomly from the available stations near Ridgecrest which station to add as the noise for each of the simulation stations with each of the rupture magnitudes. After running the MCMC inversions again for this noisy simulated peak strain data for the deterministic ruptures, we found that the observations were unchanged – strong determinism was still visible, with higher magnitudes having longer times to transition from the fast growth phase. The MCMC results scatter plots for this test can be found in Figure 2.S28. While simple, this test suggests that the addition of noise in the real data would not necessarily hide evidence for strong determinism if it did exist with real earthquakes. If it did exist, our analysis should still have been able to detect it.

We also tested implementing the same high-pass filter to the deterministic SW4 waveforms that we used for the real data. After adding the filter and recomputing the RMS and peak strain waveforms, we ran the MCMC inversions again, the results of which can be found in Figure 2.12. In this case, it appears that the implementation of the filter did cause more substantial changes to the results. In Figure 2.12a, where the earthquake magnitude is plotted against the transition time, the clear exponential growth trend observed in Figure 2.11a with the unfiltered data is obscured, and the trend appears to be more linear. However, there qualitatively is still a positive correlation between the variables. The other plot that showed some differences was Figure 2.12c, where the inverted quadratic shape is less easily visualized because of scatter in the slopes for the **M**5 earthquake. The remaining plots in Figure 2.12b and 2.12d-f do not show any substantial changes from those in Figure 2.11 where the unfiltered data was used. Ultimately, this still supports the conclusion that if earthquakes were strongly deterministic, our analysis of the real data should have been able to demonstrate that. If this was the case, however, the differences between Figures 2.11a and 2.12a suggest that the high-pass filtering process may disguise a potential exponential trend between magnitude and transition time and make it appear linear instead.

Outstanding questions remain about what exactly the multi-stage growth pattern and time to transition mean. Meier et al. (2016) used near source (≤25 km) vertical component seismic data integrated to displacement to examine how peak ground displacement evolved over time with the P-wave arrival, and found that for larger magnitude events where the initial displacement pulse did not encompass the entire rupture process, there was no correlation between the initial pulse and final magnitude. However, they also observed that there was a change in growth rate of the

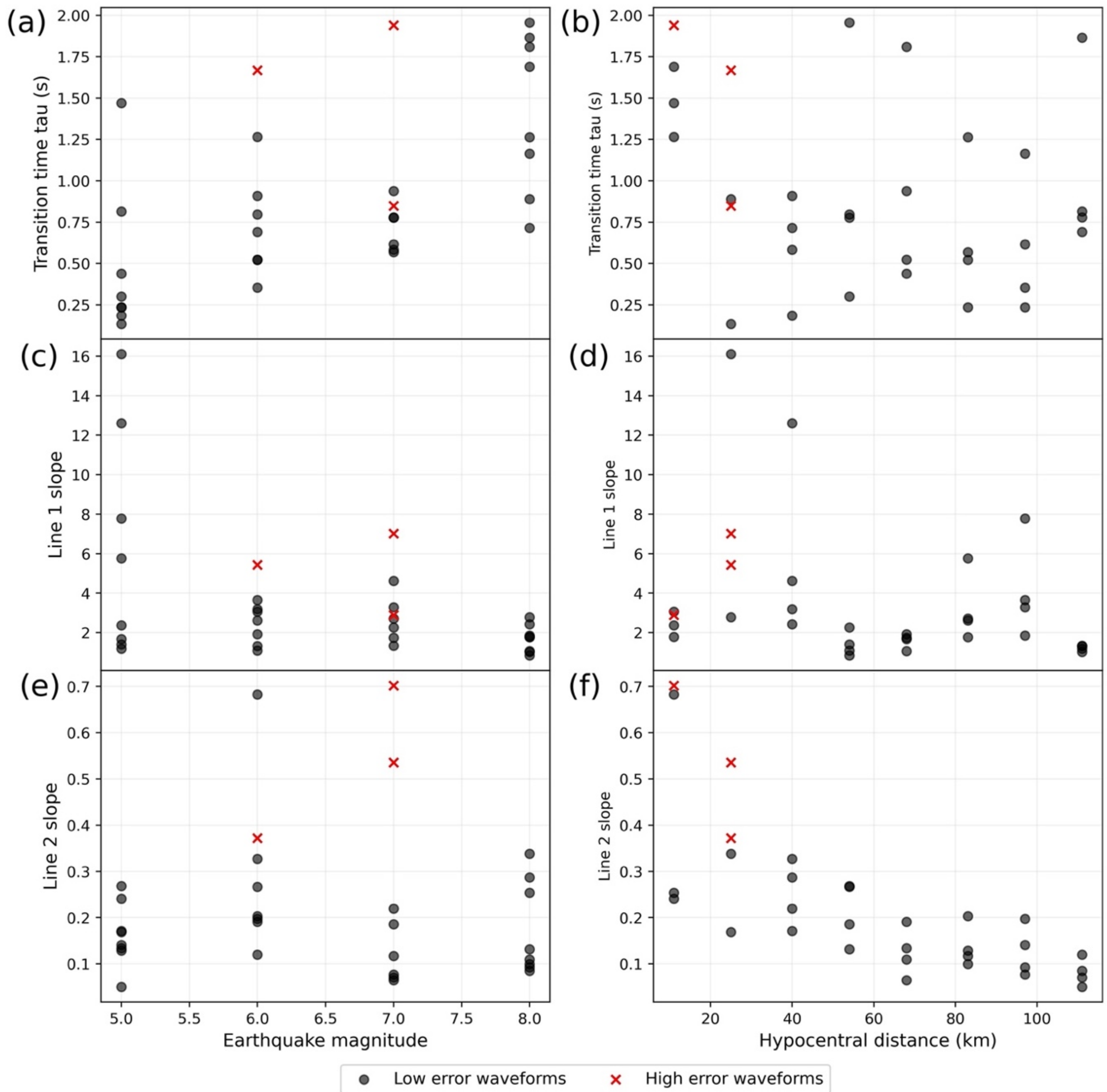


Figure 2.12. The same as Figure 2.11, but adding the high-pass filter that was applied to the real strain data to the deterministic-designed SW4 simulated earthquake waveforms. This plot shows the results of the MCMC inversions on the filtered data, where we inverted for the transition point from the fast initial growth phase of the peak strain waveforms and the slower growth phase and the slopes of the two growth stages plotted against the earthquake magnitudes and station hypocentral distances. For each of the subplots, the results for the waveforms with low error between the simulated data and the MCMC fit are the black circles, and the results for the waveforms with high error are the red 'x' symbols. **(a)** Earthquake magnitude versus the time (s) it takes to transition to slower growth after the arrival of the P-waves (τ). **(b)** Hypocentral distance (km) versus τ. **(c)** Earthquake magnitude versus the slope of the first line (fast growth phase). **(d)** Hypocentral distance (km) versus the slope of the first line (fast growth phase). **(e)** Earthquake magnitude versus the slope of the second line (slower growth phase). **(f)** Hypocentral distance (km) versus the slope of the second line (slower growth phase).

median peak ground displacement curves around 0.1-0.2 s, which they suggest could possibly be related to a transition in rupture mechanism from a crack-like nucleation phase to pulse-like rupture. Unfortunately, because strainmeters are not as common of an instrument installment as seismometers and we only included earthquakes **M**6 or larger in this study, we have limited records within 25 km in our dataset from which to compare. From the North American data, we have only two waveforms recorded at 13 and 23 km from their earthquake epicenters, which had peak strain growth phase transition times of 0.63 and 4.37 s, respectively. Several of the Taiwanese strainmeters were closer to the earthquakes, and we have 7 waveforms with hypocentral distances of 9-21 km recorded, with transition times from 0.17 to 0.89 s. These transition times are longer than those reported by Meier et al. (2017), though they are of a similar order of magnitude. Our dataset of near-source records is quite small to be able to draw meaningful conclusions from, and we also did not bin waveforms by magnitude or take the median of the peak strain waveforms the way that Meier et al. (2017) did. It is possible that a crack-like to pulse-like rupture transition could also explain the pattern found in our peak strain data.

Given the lack of clear relationship between earthquake source properties and characteristics of the peak strain waveforms observed in the real data in this study, it is likely that path effects are more influential on the final appearance of the peak strain waveforms than the source. Another outstanding question is why some stations for particular earthquakes display the multi-stage growth pattern much more clearly than others, which should be explored in future work. Examples of this can be found in Figure 2.3c, where the stations with lower SNRs do not exhibit a clear initial fast growth phase like the stations with high SNRs, as well as in several of the other earthquakes found in section 2.9 (e.g., Figures 2.S4, 2.S19, and 2.S20). From the plots which color waveforms by hypocentral distance it appears that stations farther from the hypocenters may tend to be more likely to lack a strong initial growth phase, but this is not universal and there are some examples where closer stations also lack it (e.g., Figures 2.S6, 2.S15, and 2.S19). One potential consideration is the effect that P-wave radiation patterns may have on the strength of the signals arriving at different strainmeter stations depending on the azimuth of the station from the event with respect to the radiation pattern. Theoretical radiation patterns for strain data have been published by Rodriguez and Wuestefeld (2020). In future work, we plan to compare the spatial distribution of stations which do and do not display a strong initial fast growth phase to these theoretical radiation patterns to see if they can offer insight into the cause of this variation.

2.5. Conclusions

In this study we explored whether borehole strainmeter data could provide new insight into the question of how earthquakes begin, and whether they are distinguishable by magnitude early in the rupture process or not. We analyzed the growth pattern of peak strain waveforms from three types of borehole strainmeters for instruments in North America, Japan, and Taiwan, and observed that most waveforms exhibited an early rapid growth/doubling phase before transitioning to a more gradual growth phase. We used MCMC modeling to invert for the slopes of these two phases of growth for each individual waveform, as well as the point of transition, and plotted these against earthquake magnitude and hypocentral distance. There did not appear to be any strong relationship between the characteristics of the peak strain growth and earthquake magnitude, so we cannot conclude that earthquakes are strongly deterministic. Instead, hypocentral distance had stronger correlations to the characteristics of the peak strain waveforms.

We built models of several earthquakes using the seismic wave propagation software SW4. First, we modeled the complex **M**7.1 Ridgecrest earthquake to demonstrate that the strain waveforms produced by SW4 were similar to the real strain waveforms. We then produced a series of earthquakes from **M**5-8 designed to be strongly deterministic and performed the same MCMC analysis on them, finding that in this case, the time of transition between growth phases was correlated with earthquake magnitude for these models. This demonstrated that if earthquakes were strongly deterministic, the analysis we performed should have been able to detect that. Future work should focus on understanding what causes the variability in appearance of the multi-stage growth pattern between stations for the same earthquake, and solidifying what physical mechanism may be responsible for this pattern.

2.6. Data and Resources

Codes used to process and analyze data and make the figures for this study are available at https://github.com/UO-Geophysics/strain and https://github.com/UO-Geophysics/sw4_strain.

North American seismic and strainmeter data were downloaded from the Incorporated Research Institutions for Seismology Data Management Center (IRIS-DMC) under network code PB. Japanese seismic and strainmeter data were provided by Suguru Yabe from the Geological Survey of Japan of the National Institute of Advanced Industrial Science and Technology (AIST). Taiwanese strainmeter data were provided by Alexandre Canitano from the Institute of Earth Sciences at Academia Sinica. Moment tensors used to plot the earthquakes in Figure 2.1 were accessed using EarthScope's Searchable Product Depository (Trabant et al., 2012) from the Global Centroid Moment Tensor catalog (Dziewonski et al., 1981; Ekström et al., 2012). The complex rupture geometry used for the SW4 model of the 2019 **M**7.1 Ridgecrest earthquake can be found in Goldberg et al. (2020); the USGS finite fault model for the 2019 **M**7.1 Ridgecrest earthquake can be found at https://earthquake.usgs.gov/earthquakes/eventpage/ci38457511/executive.

2.7. Acknowledgements

This material is based on services provided by the NSF GAGE Facility, operated by EarthScope Consortium, with support from the National Science Foundation, the National Aeronautics and Space Administration, and the U.S. Geological Survey under NSF Cooperative Agreement EAR-1724794.

We would like to acknowledge use of the Broadband Platform software provided by the Southern California Earthquake Center (http://scec.org) which is funded by NSF Cooperative Agreement EAR-1600087 and USGS Cooperative Agreement G17AC00047.

2.8. Bridge

In this chapter, we covered how a lesser-used geodetic instrument called the borehole strainmeter can offer insight into how earthquakes begin. We found that borehole strainmeter data from earthquakes **M**6+ did not show evidence for strong determinism, or the idea that small and large earthquakes begin differently and are distinguishable by features of the earliest-arriving P-wave data. In the following chapters, we will focus on the systems that we use to quickly characterize earthquakes as data are received that do not rely on the existence of strong determinism to function. The next chapter will present the results of our work to improve the performance of earthquake early warning systems using another type of geodetic data – GNSS.

2.9. Supplemental Figures

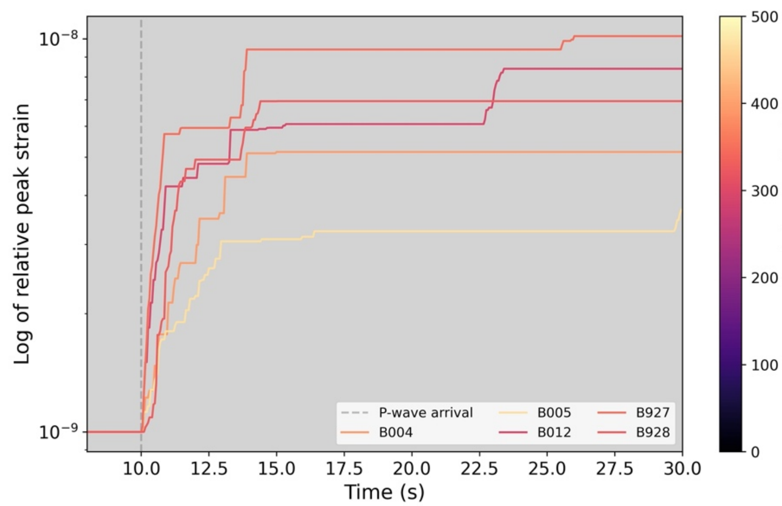


Figure 2.S1. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6 on April 13, 2023. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

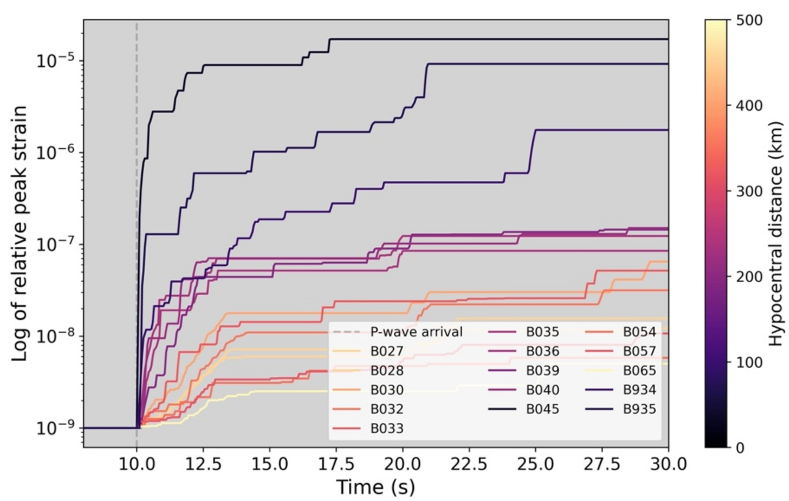


Figure 2.S2. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.4 on December 20, 2022. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

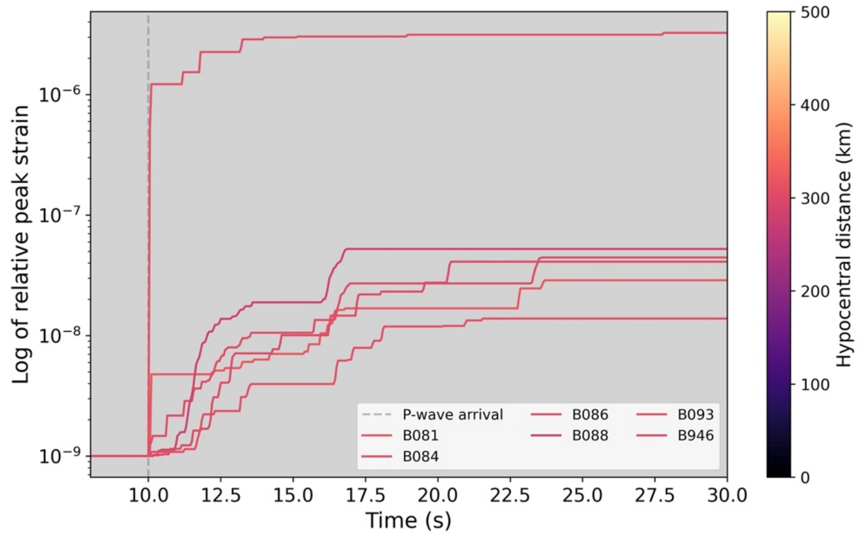


Figure 2.S3. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.2 on November 22, 2022. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

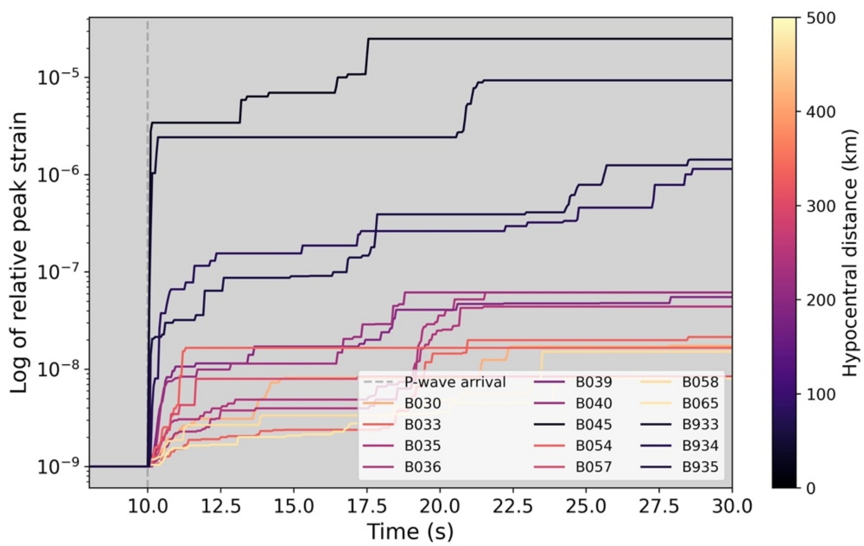


Figure 2.S4. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.2 on December 20, 2021. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

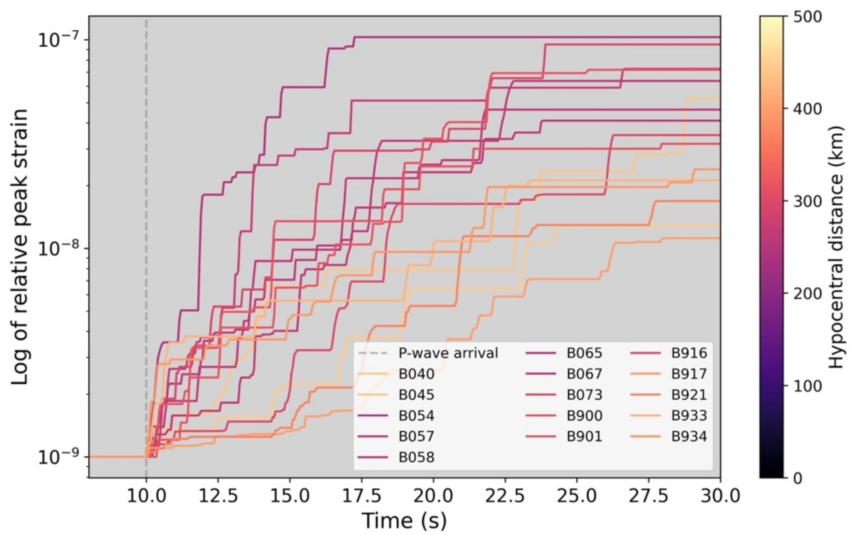


Figure 2.S5. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6 on July 8, 2021. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

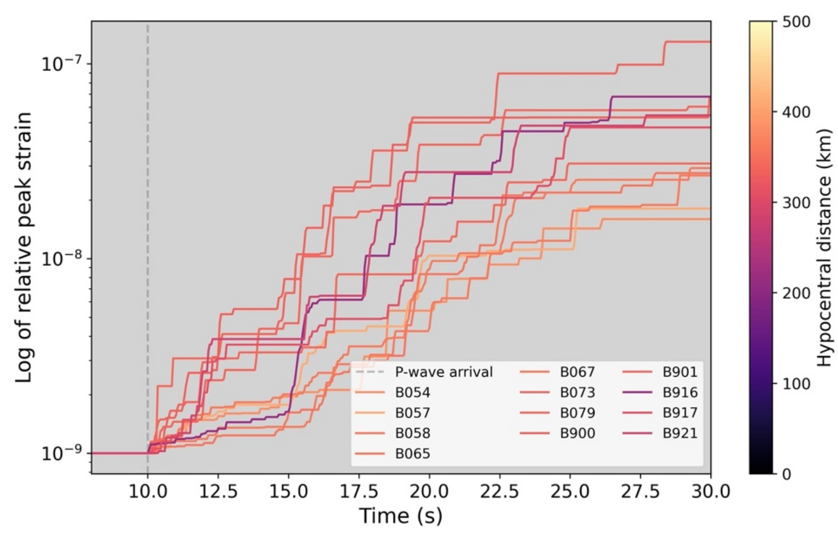


Figure 2.S6. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.5 on May 15, 2020. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

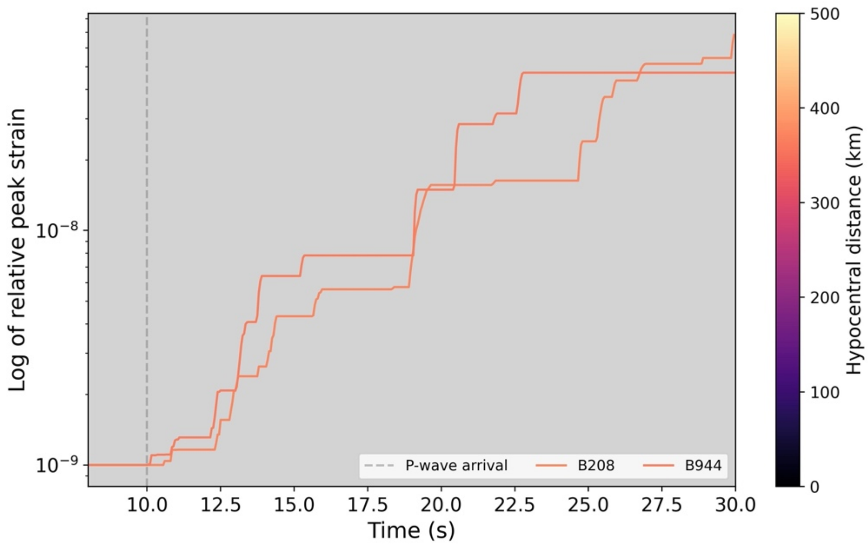


Figure 2.S7. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.5 on March 31, 2020. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

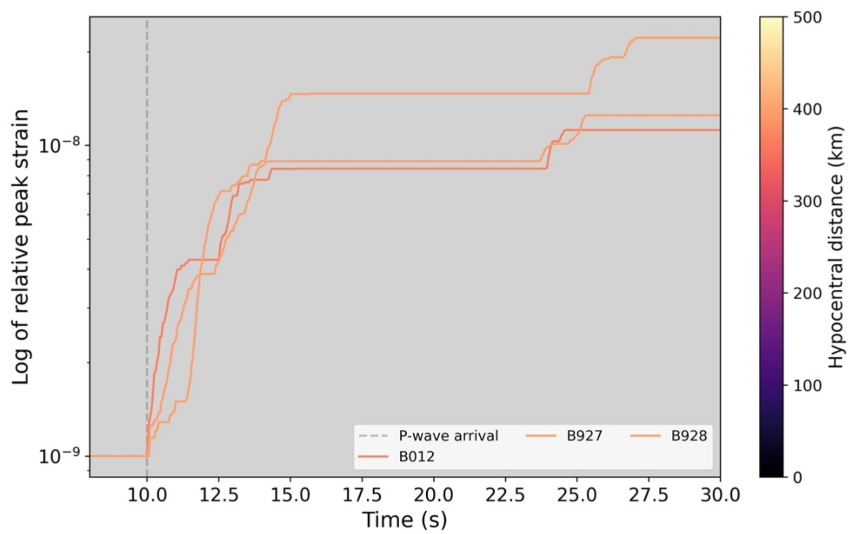


Figure 2.S8. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.3 on December 25, 2019. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

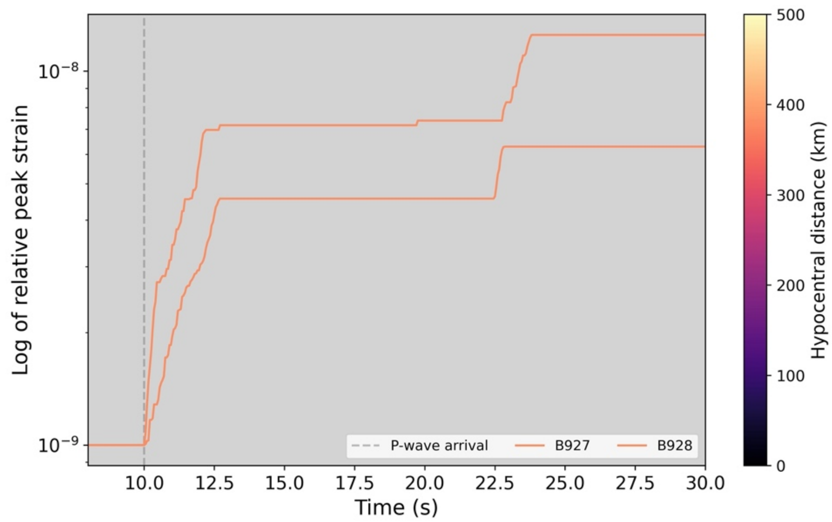


Figure 2.S9. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6 on December 23, 2019 at UTC hour 20. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

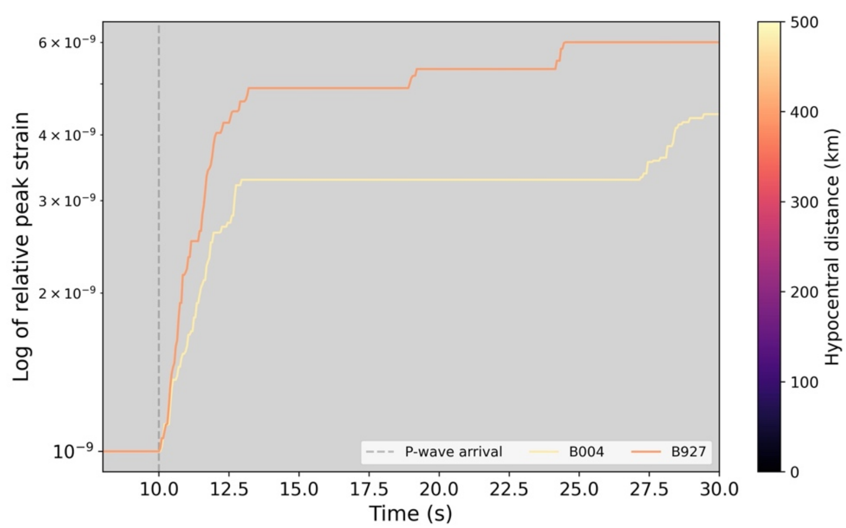


Figure 2.S10. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6 on December 23, 2019 at UTC hour 19. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

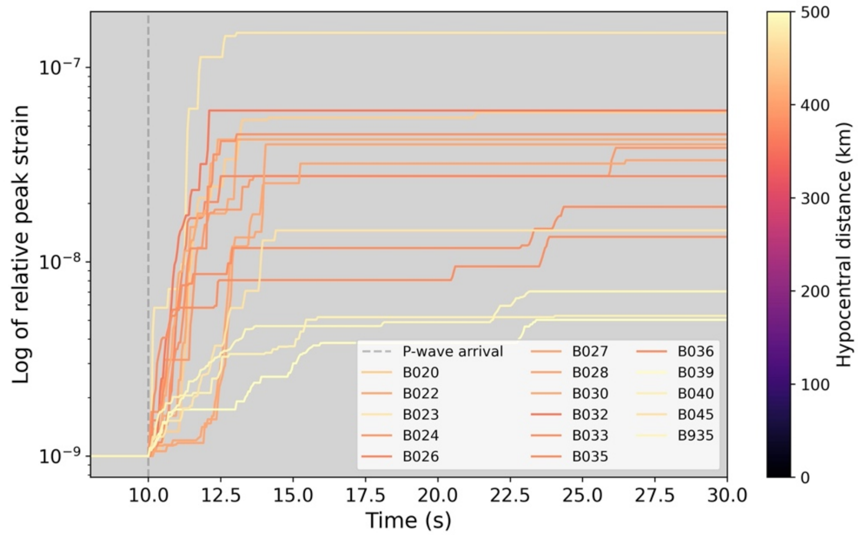


Figure 2.S11. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.3 on August 29, 2019. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

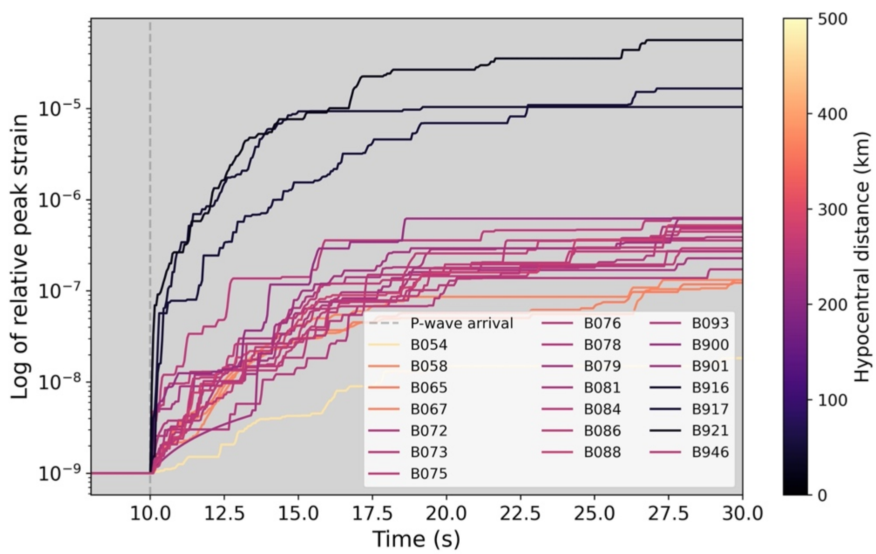


Figure 2.S12. Examples of peak strain waveforms from the North American earthquake dataset for an **M**7.1 on July 6, 2019. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

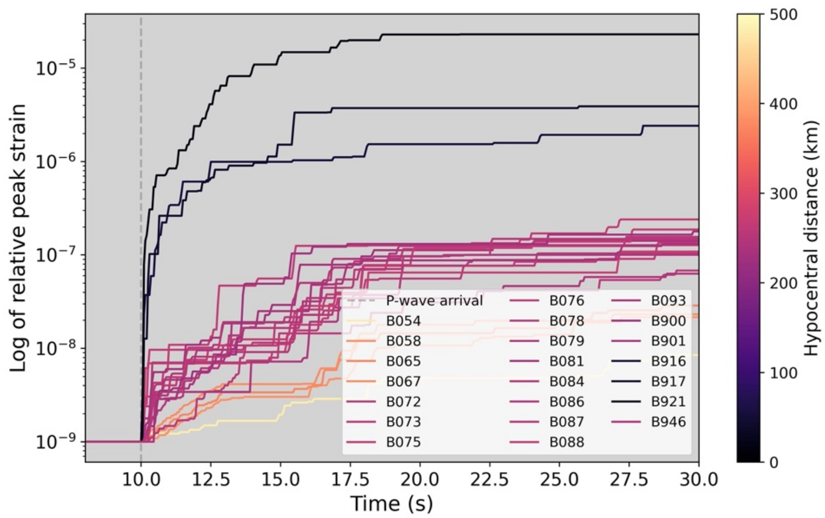


Figure 2.S13. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.4 on July 4, 2019. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

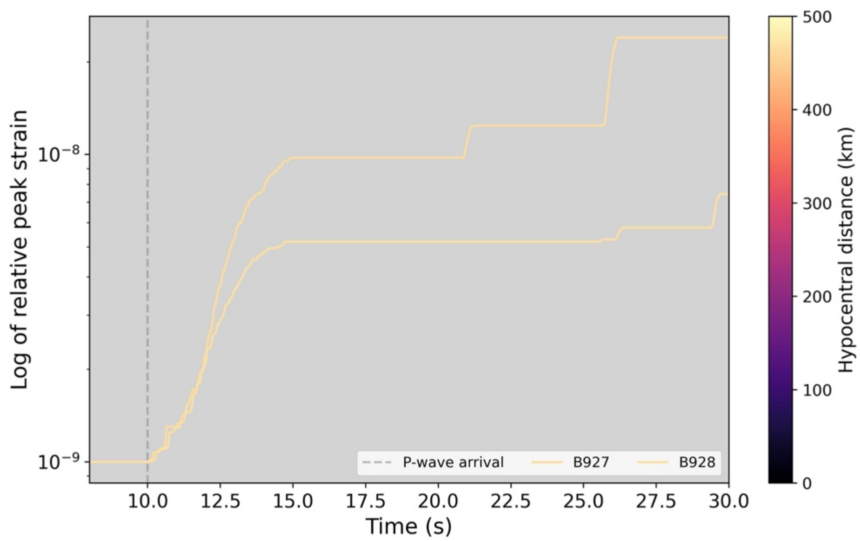


Figure 2.S14. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.2 on July 4, 2019. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

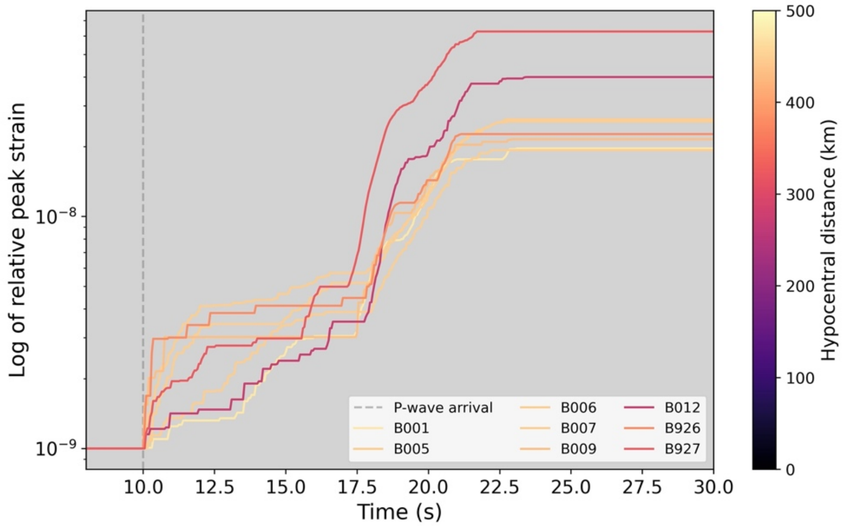


Figure 2.S15. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.8 on October 22, 2018. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

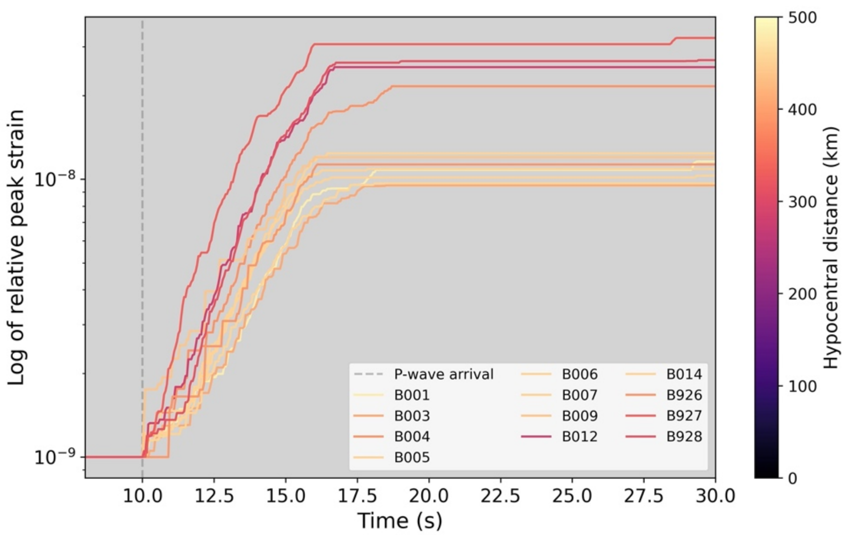


Figure 2.S16. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.5 on October 22, 2018 at UTC hour 5. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

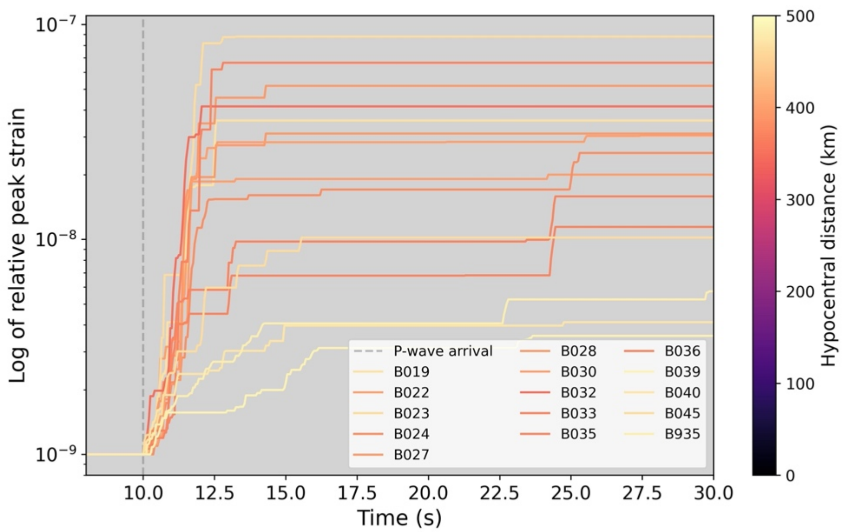


Figure 2.S17. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.2 on August 22, 2018. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

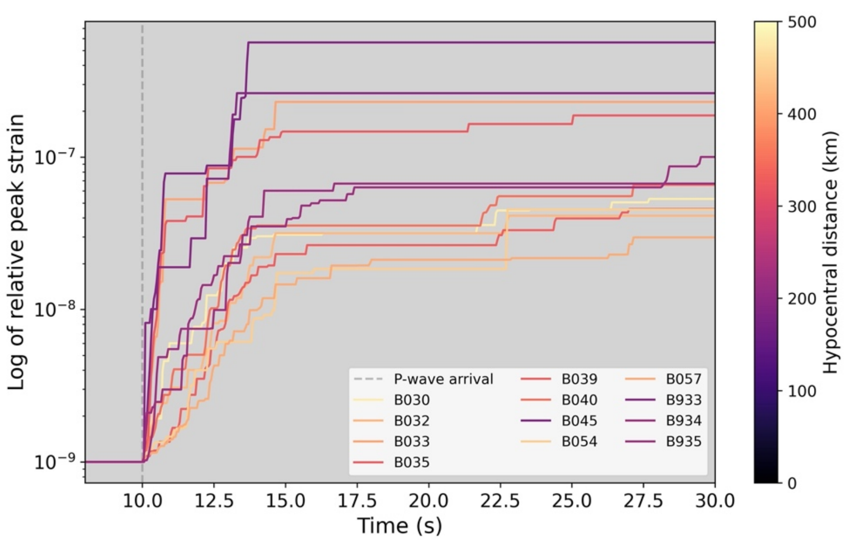


Figure 2.S18. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.6 on December 8, 2016. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

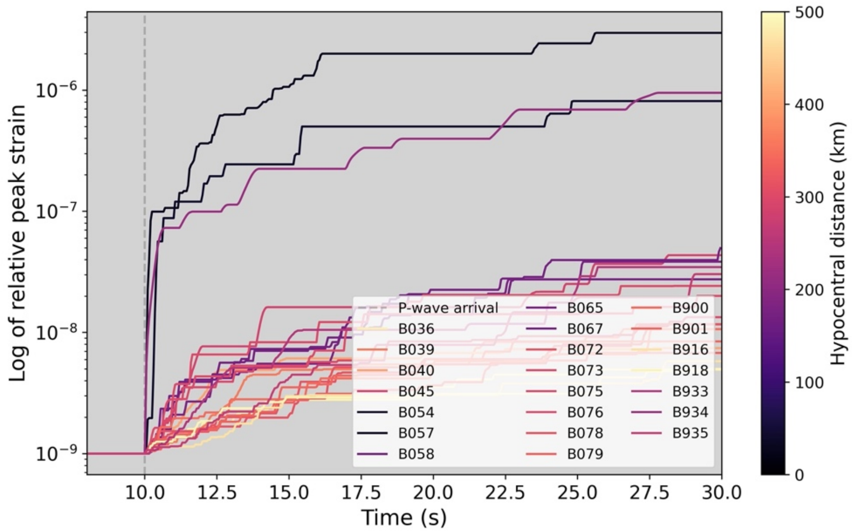


Figure 2.S19. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.02 on August 24, 2014. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

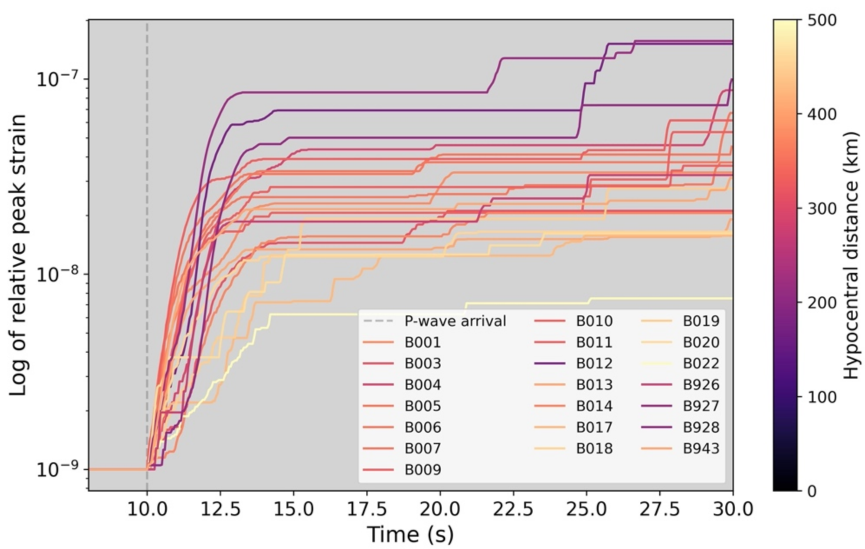


Figure 2.S20. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.5 on April 24, 2014. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

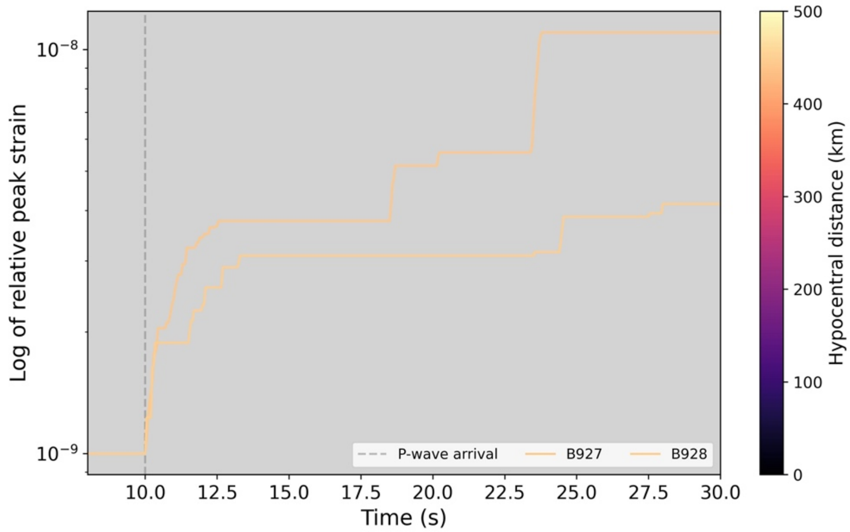


Figure 2.S21. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6 on September 4, 2013. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

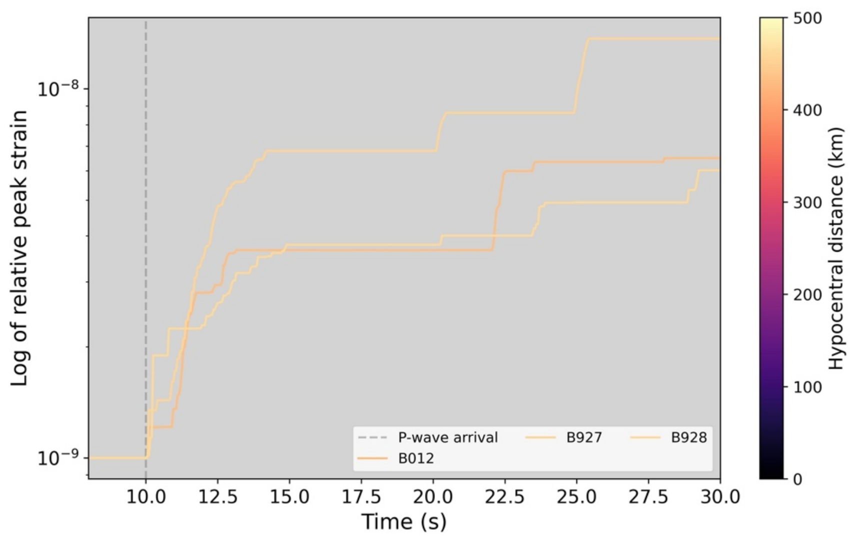


Figure 2.S22. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.1 on September 3, 2013. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

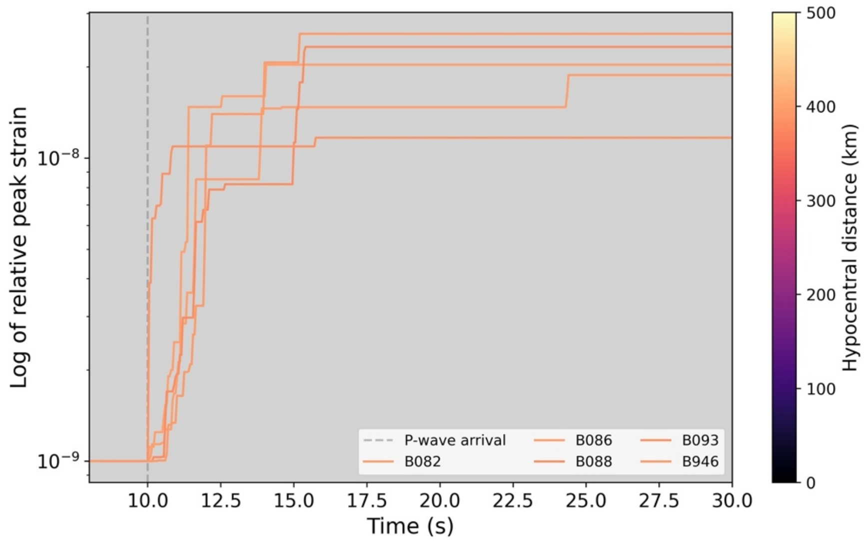


Figure 2.S23. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.3 on December 14, 2012. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

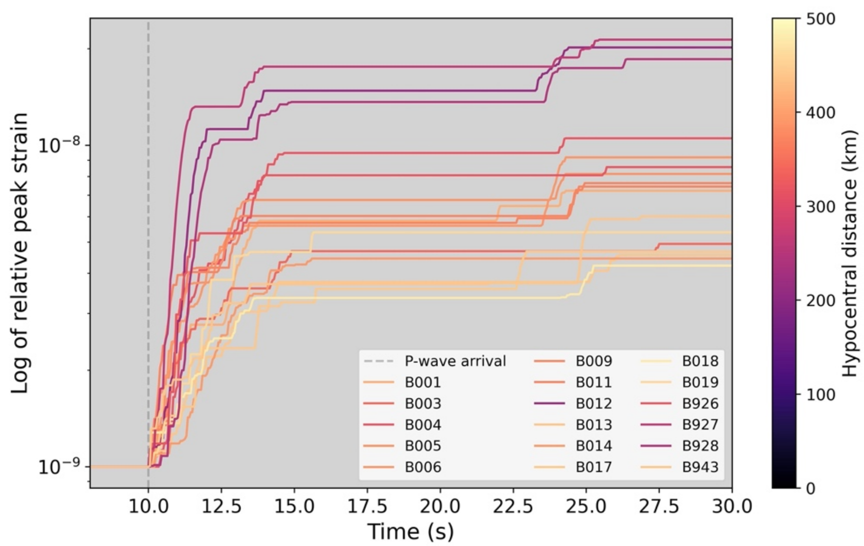


Figure 2.S24. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.1 on November 8, 2012. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

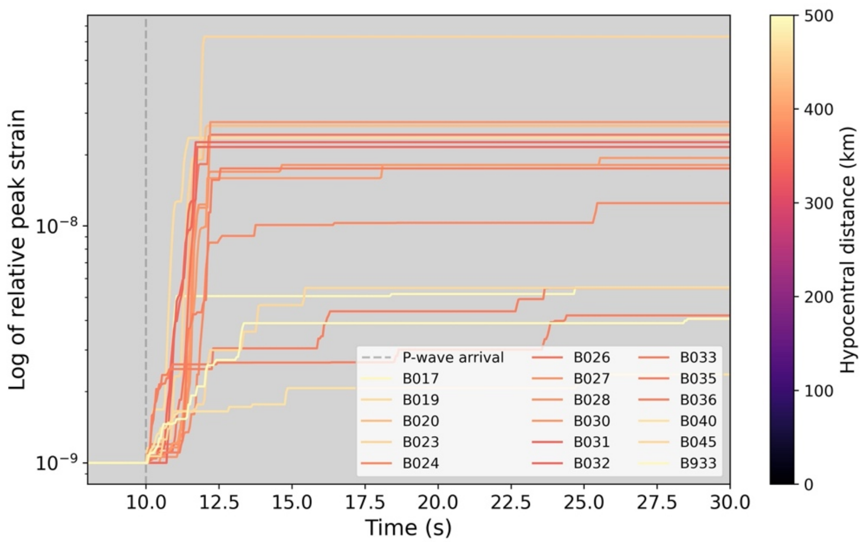


Figure 2.S25. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6 on April 11, 2012. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

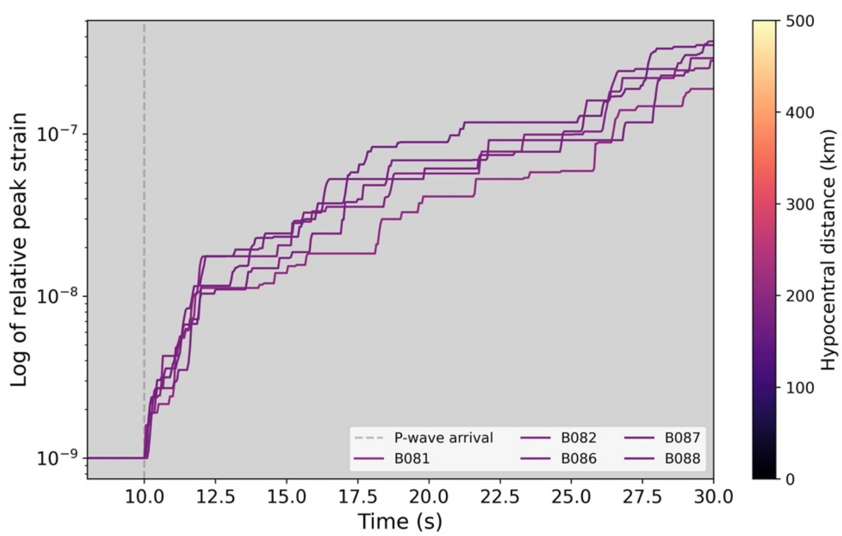


Figure 2.S26. Examples of peak strain waveforms from the North American earthquake dataset for an **M**7.2 on April 4, 2010. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

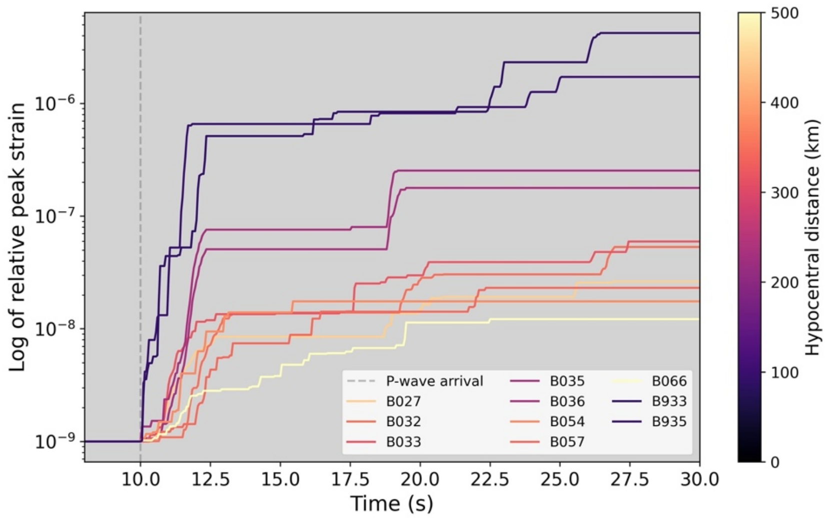


Figure 2.S27. Examples of peak strain waveforms from the North American earthquake dataset for an **M**6.5 on January 10, 2010. Each colored line represents a peak strain waveform from one of the Gladwin Tensor Borehole Strainmeters, and individual waveforms have been shifted up or down so that they all align at the same initial value before the P-wave arrivals. The waveforms are colored by hypocentral distance.

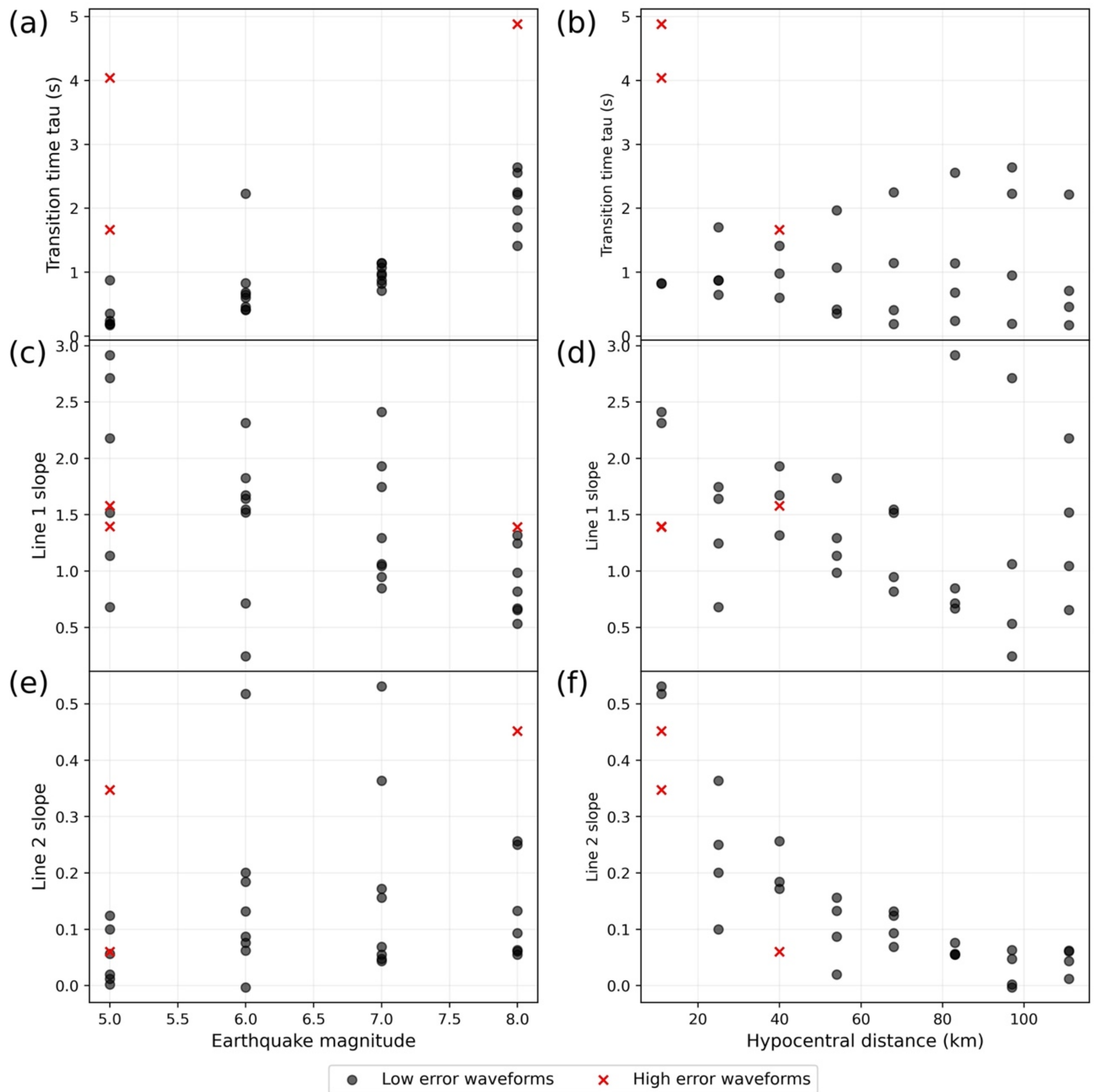


Figure 2.S28. The same as Figure 2.11, but adding real strain noise to the deterministic-designed SW4 simulated earthquake waveforms. This plot shows the results of the MCMC inversions on the noisy data, where we inverted for the transition point from the fast initial growth phase of the peak strain waveforms and the slower growth phase and the slopes of the two growth stages plotted against the earthquake magnitudes and station hypocentral distances. For each of the subplots, the results for the waveforms with low error between the simulated data and the MCMC fit are the black circles, and the results for the waveforms with high error are the red 'x' symbols. **(a)** Earthquake magnitude versus the time (s) it takes to transition to slower growth after the arrival of the P-waves (τ). **(b)** Hypocentral distance (km) versus τ. **(c)** Earthquake magnitude versus the slope of the first line (fast growth phase). **(d)** Hypocentral distance (km) versus the slope of the first line (fast growth phase). **(e)** Earthquake magnitude versus the slope of the second line (slower growth phase). **(f)** Hypocentral distance (km) versus the slope of the second line (slower growth phase).

CHAPTER III

DETECTING EARTHQUAKES IN NOISY REAL-TIME GNSS DATA WITH MACHINE LEARNING FOR IMPROVED PGD MAGNITUDE ESTIMATION

This chapter was coauthored with my advisor Diego Melgar, who helped conceptualize and plan the study and provided editorial feedback, Amanda M. Thomas, who developed the basis for the CNN model architecture I used, Dara E. Goldberg, who provided the Ridgecrest fault model, David Mencin, who helped us access the archived real-time GNSS data from EarthScope, and Brendan Crowell, who provided information about GFAST. I performed all data analysis, produced all figures, and wrote all of the text in the manuscript.

3.1. Introduction and Background

The recent incorporation of Global Navigation Satellite Systems (GNSS) data into the U.S. Geological Survey's ShakeAlert earthquake early warning (EEW) system through the GFAST (Crowell et al., 2013, 2016) algorithm represents a first of its kind advancement in preparing EEW systems to handle very large earthquakes (Murray et al., 2023). In Western North America, particularly in the Pacific Northwest region, the Cascadia Subduction Zone poses a great (**M**8-9+) earthquake hazard. Most EEW systems rely only on seismic data to detect and analyze earthquakes, but for great earthquakes, seismic-only systems will generally fail to register the true size of these events because of a phenomenon called magnitude saturation, where large earthquakes are indistinguishable from even larger ones.

Magnitudes calculated using seismic P-wave waveforms saturate at the high end of the scale for a few reasons, including instrument and source effects. Traditionally, to calculate magnitudes seismic data are double integrated from acceleration to displacement, but this requires the implementation of high-pass filters to remove baseline drift effects that appear during integration. This removes low frequency content from the data, changing the relationship between observed displacement and the calculated magnitude (Colombelli et al., 2012). While the relationship between the log of peak ground displacement and magnitude is linear for moderate magnitude earthquakes, for larger magnitudes this relationship does not hold (Trugman et al., 2019). Additionally, the source spectra of earthquakes of different magnitudes can affect the accuracy of a magnitude computed from those seismic waves. As earthquake magnitude increases, the source duration and rise time of the earthquake also increase (Melgar and Hayes, 2017). This results in the corner frequency of the source spectrum of the earthquake decreasing as the earthquake magnitude increases (Geller, 1976). For some seismically derived magnitude types such as short-period body wave (m_b) or surface wave (m_S), the magnitude calculation is made from data of a specific period (e.g., 1 s and 20 s respectively for those two examples). If the corner frequency for an earthquake is smaller than the measuring frequency, the source spectra is not flat at that frequency, resulting in a calculated magnitude that is too small and not proportional to the earthquake moment.

This limitation of seismic data for large magnitude earthquakes has major implications for the dissemination of accurate earthquake and tsunami warnings for systems which only use seismic data. A notable example of this occurred during the 2011 Tōhoku earthquake in Japan, when even the fifteenth EEW "forecast" 116.8 seconds after the system was triggered still interpreted the **M**9.1 as an **M**8.1, as the system saturated for events over **M**8 (Hoshiba and Ozaki, 2014). In contrast, GNSS data do not suffer from magnitude saturation because they measure displacement of the ground relative to an absolute reference frame (Melgar et al., 2015), making it possible to discriminate the largest earthquakes from each other. In anticipation of large earthquakes within the West Coast alerting region, this is why the ShakeAlert system has been upgraded to include GNSS data analysis with the GFAST algorithm. GNSS position solutions are derived from the time delay and phase change of microwave signals emitted by orbiting satellites and collected by receivers on the ground (Melgar et al., 2020). These calculations require knowledge of how much delay time the troposphere and ionosphere introduce as well as the orbits of the satellites and timing of their clocks, which can be difficult to ascertain with precision in real-time. Additionally, the physical setting of a receiver station can also affect the position solutions because of reflections of the microwaves off of surrounding structures and terrain. Combined, these factors result in real-time GNSS data which have a very high noise floor on the order of a few to 10 cm (Geng et al., 2018), which obscures small signals such as P-wave arrivals. This noise can also be highly irregular, containing large spikes with unidentified origins that can be mistaken for large ground motions caused by earthquakes (Kawamoto et al., 2017). This means that in order to be reliably operational in a real-time EEW environment, GNSS algorithms like GFAST require a seismic trigger to begin analysis for earthquakes in progress.

Although it would be useful to have a standalone GNSS-based EEW system that could both accurately identify when earthquakes are occurring, as well as estimate their magnitudes without saturation, the former goal is much more difficult to achieve because of the aforementioned noise levels that would hide seismic wave arrivals until long after they had been detected by seismometers. However, it has been demonstrated for methods such as the Machine-Learning Assessed Rapid Geodetic Earthquake (M-LARGE) algorithm, which uses 3-component GNSS data to ascertain earthquake source properties such as magnitude, centroid location, and rupture length and width, that reductions in the noise levels of data fed into the algorithm improve the accuracy of the model's predictions of the source properties (Lin et al., 2023). Our goal for this study was to develop an algorithm which can quickly detect earthquakes in noisy real-time GNSS data so that it may act as a "filter" for high-noise data, ensuring that only waveforms where earthquakes are actually detected and the seismic signal is strong are fed into algorithms that use them to determine earthquake source properties.

To accomplish this, we took a machine learning approach. Applications of machine learning in seismology and earth sciences have expanded in recent years, with a review of the history, current use, and future directions available from Mousavi and Beroza (2023). These applications include earthquake detection and location (e.g., Perol et al., 2018; Zhang et al., 2020; Mousavi and Beroza, 2020a), low-frequency earthquake detection (Thomas et al., 2021), picking of P- and S-phase arrivals for improved locations and velocity models and earthquake catalog expansion (e.g., Ross et al., 2018; Zhu and Beroza, 2019; Mousavi et al., 2020), and also determination of source parameters such as magnitude (e.g., Lomax et al., 2019; Mousavi and Beroza, 2020b; Ristea and Radoi, 2021; van den Ende and Ampuero, 2020; Münchmeyer et al., 2021; Lin et al., 2021, 2023; Licciardi et al., 2022).

Machine learning methods typically require large datasets with which to train the models in order to perform well. While for seismic-only models the seismological community tradition of maintaining large public datasets works well (e.g., the catalog from the National Earthquake Information Center (NEIC) or the EarthScope waveform data archives), for GNSS data there is less availability. GNSS instruments and data analysis techniques were introduced into the study of earthquakes much more recently than seismometers, so records of earthquakes are less plentiful. Furthermore, as previously discussed GNSS earthquake and tsunami early warning algorithms are primarily useful during large earthquakes, since smaller earthquakes can be handled by seismic algorithms without magnitude saturation. However, as described by the Gutenberg-Richter law (Gutenberg and Richter, 1949) there are many more small magnitude earthquakes from which data are available than large ones. Therefore, in order to train a model on a more balanced dataset with increased availability of waveforms with large ground displacements, we generated a dataset of synthetic noisy displacement waveforms using FakeQuakes/MudPy, a set of codes described in Melgar et al. (2016) which have subsequently been used to generate synthetic GNSS data for several machine learning (ML) studies (e.g., Lin et al., 2021, 2023).

In this study, we will present a Convolutional Neural Network (CNN) method for the detection of earthquakes in noisy GNSS data. This CNN architecture was based on U-net (Ronneberger et al., 2015), a model developed for biomedical image segmentation. This structure was also used by Zhu and Beroza (2019) to develop PhaseNet, a model for identifying P- and S-phase arrivals in seismic data. U-net features both a downsampling branch to localize and distill the most important features of the training data, as well as an upsampling branch to increase the resolution of the output and reconnect features from the downsampling branch to this output. In the case of detection of an earthquake, the model is being trained to determine whether a waveform contains an earthquake of not, and if it does, to identify the location in the waveform where the earthquake begins to appear (i.e., the P-wave arrival).

The variation of U-net that we used was built for identifying low frequency earthquakes in seismic data (Thomas et al., 2021), which are more emergent than impulsive. This method also seemed appropriate for the detection of P-waves in noisy GNSS data, as for some time after arrival the waves are likely to be obscured by noise. Our model was trained with synthetic 3-component displacement data contaminated with real noise, as well as pure real noise alone, and each waveform sample was labeled with the timing of the P-wave arrival for earthquake samples, or just as noise for the noise-only samples. We then tested the performance of the model on unseen additional synthetic data, as well as a real real-time GNSS dataset provided by EarthScope. Finally, we applied the trained model to a simplified version of the magnitude determination method used in GFAST to determine whether using only GNSS data where an earthquake was detected would improve the accuracy of the real-time magnitude calculation.

3.2. Data and Methods

3.2.1. Real Data Processing

We chose the 2019 Ridgecrest earthquake sequence as our focus for this study, as there were many nearby GNSS stations from which to collect real earthquake waveforms, as well as availability of a fault model and local velocity model to facilitate simulation of additional earthquakes for training data. We received a dataset provided by EarthScope comprised of day-long three-component real-time GNSS displacement recordings spanning from June 1, 2019 to July 31, 2020. The data came from all available GNSS stations within 300 km of the Ridgecrest, CA area (Figure 3.1). We parsed these data from the provided Real Time eXtended (RTX) format into day-long miniSEED files.

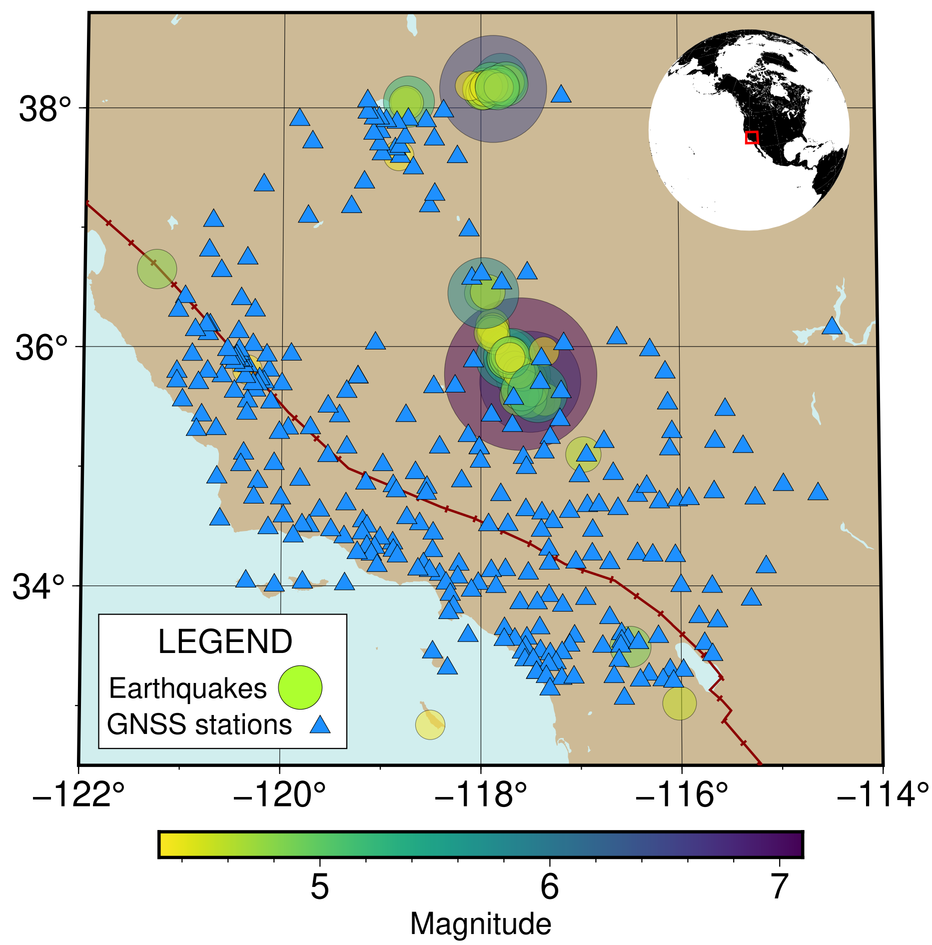


Figure 3.1. Map of the study region. GNSS stations from which we were given access to archived real-time position solutions in a 300 km radius around Ridgecrest, California are shown with blue triangles. Earthquakes ≥**M**4.3 from the Southern California Earthquake Data Center catalog from which P-wave picks were available are shown by the circles, which vary in size and color by magnitude.

This real real-time data served two purposes for this study. First, it provided a source of real real-time GNSS noise that could be added to the clean synthetic displacement waveforms we generated for the training of the ML model, which will be described in the next section. Second, it provided real examples of how earthquakes of a range of magnitudes appear in noisy GNSS data across our study region at a range of hypocentral distances. These samples were used to test the performance of our ML algorithm after it was trained, validated, and tested on synthetic data.

We built a large array of noise-only data samples by selecting random stations and random days from the daily real-time miniSEED dataset, skipping any days for which there were **M**4.3 or greater earthquakes in the catalog, and selecting random starting times from the valid daily samples. We trimmed the noise samples to 256 seconds, demeaned them, and concatenated the three components of displacement in North (N), East (E), and Vertical (Z) order so that the ML algorithm would receive information from all three directions of motion. This culminated in a noise dataset consisting of over 1.8 million noise samples. These noise samples were then randomly combined with the synthetic displacement waveforms we generated in a process that trimmed them down to 128 seconds, which will be described later.

To create the earthquake-containing samples that would be used in the final stage of testing of the ML model, we split all of the daily miniSEED files into separate 128 second samples. Traces were merged with an interpolation method if there were gaps and demeaned, and the data were also concatenated in NEZ order. This process resulted in a real data dataset with nearly 1.1 million samples. We then cleaned the dataset by removing any samples where there was a data gap of at least 60 seconds, reducing the dataset to ~994,000.

To be able to interpret the results of the ML testing with the real data, we needed to know which of these 1.1 million samples actually contained earthquakes. We downloaded a pick catalog for earthquakes **M**4.3 or greater (circles shown in Figure 3.1) during the June 1, 2019 to July 31, 2020 study window using the SCEDC catalog (SCEDC, 2013). For a total of 362 events (including two distant earthquakes not shown in Figure 3.1), we collected P-wave picks from all available seismic stations in the study region from network codes CI (California Institute of Technology and United States Geological Survey Pasadena, 1926) and PB for all of the event origin times in the earthquake catalog. It is rare for seismic stations and GNSS stations to be collocated, so we needed to be able to approximate P-wave arrival times at the locations of our GNSS stations using the seismic pick catalog. To do this, we calculated the travel times from each event origin time to each of the seismic stations for which we had P-wave picks, and then interpolated those seismic travel times to the locations of the GNSS stations using a piecewise linear interpolator. We then converted these back to P-wave arrival times using the event origin times.

Because our ~994,000 data samples still had the original timing information associated with them, we were able to use these interpolated P-wave arrival times to determine which samples contained a P-wave arrival and where it was in the waveform, and which did not. This left us with 2123 samples from the full real data dataset which contained earthquakes, or 0.214% of the dataset. A histogram of the magnitude distribution of the waveform samples can be found in section 3.9 (Supplemental Figures) as Figure 3.S1a.

3.2.2. Producing the FakeQuakes Displacement Data

As discussed in the Introduction and Background, ML methods require a substantial amount of training data to perform well in testing. The real data we had available was very skewed, with less than 1% of the dataset containing the P-wave arrival from an earthquake. To produce a more balanced training dataset, we generated a set of realistic earthquake ruptures and displacement waveforms using FakeQuakes/MudPy (Melgar et al., 2016). These waveforms were used in the training, validation, and initial testing process of our ML algorithm, and after testing on this simulated data we also tested the algorithm's proficiency with the real GNSS data described previously.

FakeQuakes generates synthetic earthquakes by defining slip distributions and kinematic rupture parameters for a rupture, then synthesizing waveforms from that rupture. This process is explained in greater detail in Melgar et al. (2016) but will be summarized here. Slip distributions are generated by calculating the effective length and width of a fault with the target magnitude defined by the user, which are based on previously published scaling laws using a stochastic approach that allows for some variability in rupture size (Blaser et al., 2010). The mean slip across the entire fault is prescribed to be adequate to meet the target magnitude, and the actual slip on each individual subfault varies and is controlled by a Von Karman correlation function, which describes real observed earthquake slip patterns well (Graves and Pitarka, 2010, 2015; Mena et al., 2010). This allows for construction of a covariance matrix scaled to the total number of subfaults, where the vector of slip at each subfault is expressed by the Karhunen-Loève expansion and is a function of the mean slip and the eigenvalues and eigenvectors of the covariance matrix. This allows for the generation of many slip distributions, which are sampled from a probability density function with the right covariance and therefore have the right spatial statistics of slip (LeVeque et al., 2016).

Next, kinematic rupture parameters are defined, including the velocity of rupture which depends on the local shear-wave speed in the velocity model and the depth of each subfault as it joins the rupture. The rise time at each subfault, or the amount of time that slip is occurring on that particular subfault, is scaled by the square root of the total slip on that subfault. This rise time is then used to compute the slip rate function for each subfault, based on the Dreger slip rate function (Mena et al., 2010). To generate waveforms, a matrix of Green's functions is constructed by computing impulse responses for every subfault-station pair following the frequency-wavenumber algorithm from Zhu and Rivera (2002). The subfault slip rate functions are then convolved with the Green's function matrix and multiplied by the slip vector, producing three-component waveforms.

For this study we elected to generate FakeQuakes ruptures along the largest of the four fault planes involved in the 2019 **M**7.1 Ridgecrest mainshock from the inversion of Goldberg et al. (2020), which was comprised of 426 subfaults. This fault plane was set in the 1D Mojave area velocity model from the Southern California Earthquake Center's (SCEC) Broadband Platform v15.3.0 (Maechling et al., 2015). We placed receiver stations for the simulation at the locations of the real GNSS stations within 300 km of Ridgecrest, since we also received real data for those stations.

We generated ruptures with requested magnitudes between 4.3 and 7.5, with 100 ruptures per magnitude bin (increasing by 0.1 each time) for a total of 3300 ruptures. The earthquake hypocenters were assigned randomly with the generation of each rupture, and we allowed the earthquake magnitudes to vary slightly with each rupture as well. An example visualization of one particular rupture plotted by slip amount on the fault geometry can be found in Figure 3.2. A histogram of the final magnitude distribution of all ruptures (**M**3.6-7.8) can be found in Figure 3.S1b. The displacement waveforms generated were 250 seconds long at a 1 Hz sampling rate to match the real real-time GNSS data we received.

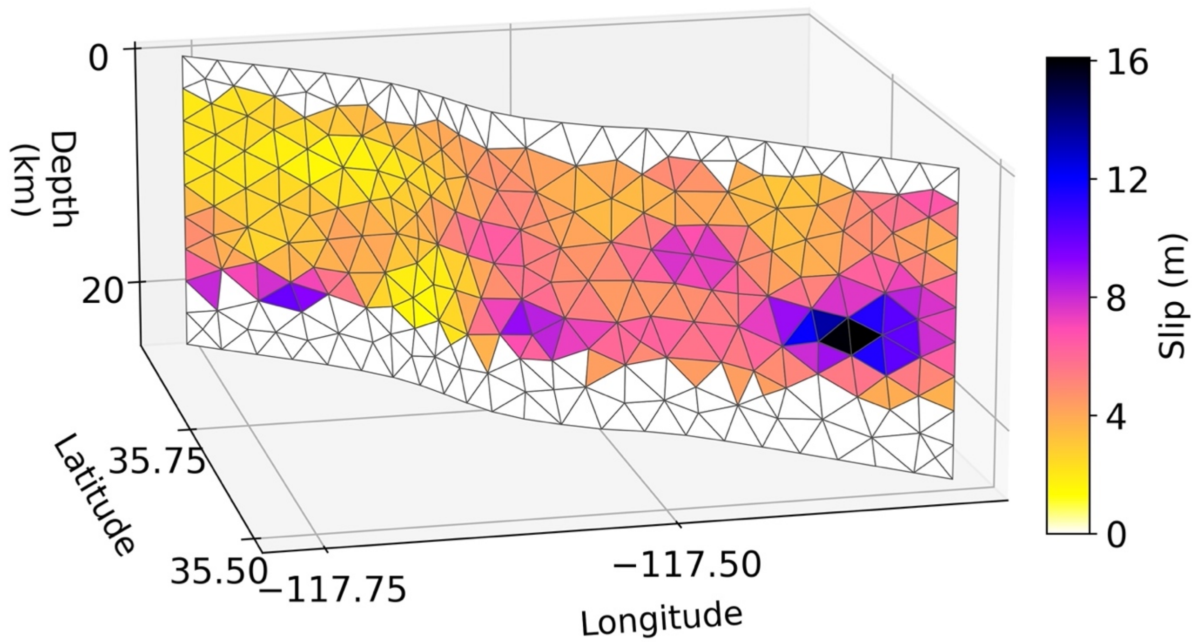


Figure 3.2. Example of the rupture pattern for one of the FakeQuakes ruptures used to generate synthetic displacement waveforms for the training dataset. The triangular shapes represent individual subfaults. This earthquake is an **M**7.45.

After the ruptures and waveforms were generated, we calculated P-wave arrival times using ObsPy's (Beyreuther et al., 2010) TauP function and the same 1D Mojave velocity model. The data was then transformed into the format in which it would be introduced to the ML model. The 250-second-long waveforms were zero-padded and trimmed to 256 seconds so that the P-wave arrival was placed in the center of the waveform for each component. The three components were then concatenated in NEZ order to match the format of the real data as described previously. This provided a total of ~916,000 displacement waveforms.

Because FakeQuakes produces ruptures systematically in order across the range of magnitudes requested, the waveforms were shuffled to ensure that the ML model could not learn by the order of the data. Finally, the waveforms were divided into three groups: 80% of the data (~733,000 waveforms) were set aside for training, 10% (~92,000) were used for the validation dataset, and the remaining 10% were saved for the first stage of the testing process.

3.2.3. Data Generation for the CNN

The FakeQuakes displacement waveforms we produced were noise free, which was not a good analogue to real real-time GNSS data that would be utilized in an earthquake early warning setting. While MudPy has the capability to generate realistic synthetic noise as well based on a real-time noise model from Melgar et al. (2020), we instead chose to use real noise from the real-time GNSS data around Ridgecrest that we received.

We constructed a data generator similar to that of Thomas et al. (2021) to feed the data samples into the ML model for training. This data generator first took a random subset of the real real-time noise data to be used as non-earthquake samples in training. It then took an identically sized random subset of the FakeQuakes displacement waveforms and a different random subset of the real real-time noise and combined them. An example of this process and the resulting noisy waveforms is shown in Figure 3.3a-c, where Figure 3.3a shows a pure displacement waveform produced by FakeQuakes, Figure 3.3b shows a random timeseries of pure noise, and Figure 3.3c shows the resulting noisy waveform from the combination of the two. This resulted in a training dataset where 50% of the waveforms contain earthquakes, and 50% contain only noise. Recall that the FakeQuakes waveforms were 256 seconds long and padded as necessary so that the P-wave arrival was directly in the center of the waveform. The data generator randomly offset each of the waveforms and trimmed them down to 128 seconds so that the P-wave arrival was no longer in the center and instead varied throughout the dataset.

The data generator also created "target" arrays during this process to act as the "labels" for the ML model to learn from. The target arrays for non-earthquake samples of pure noise consisted of an array of zeros. The target arrays for samples with earthquakes consisted of a Gaussian with a user-defined standard deviation (we chose 3 seconds) centered on the same sample as the P-wave arrival in the waveform so that the model could learn features of the waveform that indicated where the P-wave arrival was. Figure 3.3e shows an example of an earthquake waveform (blue) with the Gaussian target array (black), and Figure 3.3f shows a noise waveform (blue) with a target array of zeros (black).

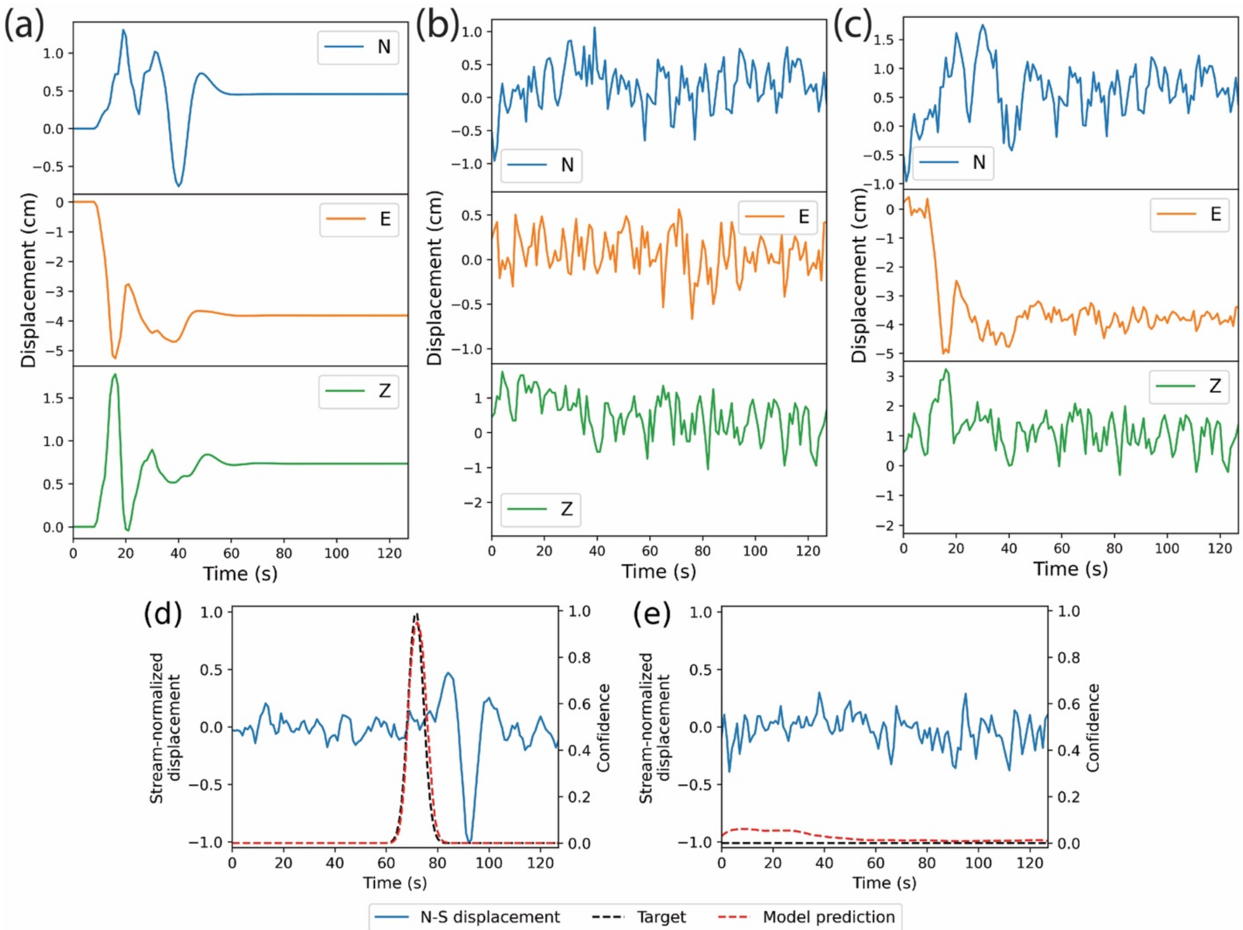


Figure 3.3. Example of the process of combining the **(a)** synthetic displacement waveforms generated with FakeQuakes with **(b)** real noise from the GNSS stations near Ridgecrest to build **(c)** realistically noisy synthetic displacement waveforms for the CNN training process. **(d)** An example of a sample from the FakeQuakes synthetic testing dataset which contains an earthquake, where the North-South component of displacement is shown in blue, the target Gaussian label indicating where the P-wave arrival is shown in black, and the prediction made by the trained CNN is shown in red. **(e)** An example of a sample from the synthetic testing dataset which contains only noise, where the North-South component of displacement is shown in blue, the target Gaussian label indicating where the P-wave arrival is shown in black, and the prediction made by the trained CNN is shown in red.

The earthquake and non-earthquake samples were then shuffled together. The target and metadata arrays were shuffled the same way to ensure that the labels stayed aligned with the data and that properties of the waveform and earthquake were still trackable after data generation. Finally, all earthquake and non-earthquake (noise) samples were stream-normalized by the maximum data value across all three components.

The data generation procedure using the real earthquake data (used in the testing process) was very similar, just leaving aside the unnecessary procedures to combine the waveforms with noise. For this data, the data generation process only involved creating the Gaussian target array for the samples which did contain P-wave arrivals, and stream-normalizing all samples.

3.2.4. CNN Training Process

As described in the Introduction and Background, we utilized a variation on the U-net (Ronneberger et al., 2015) CNN architecture built for identifying low frequency earthquakes in seismic data (Thomas et al., 2021) to train our model. The model was constructed in and run using Tensorflow (Abadi et al., 2016). The contracting side of its structure consists of an input layer followed by three pairs of one-dimensional (1D) convolutional layers with 1D MaxPooling (downsampling) layers. The resulting output is flattened and fed into a dense fully connected layer which forms the base of the network. The dense layer output is reshaped, then enters the upsampling side of the architecture with three sets of 1D convolutional, 1D upsampling, and concatenation layers (which connect the convolutional layers from the contracting side of the network to the upsampling side). One final 1D convolutional layer completes the network, and the output is then flattened again. This architecture can be visualized in Figure 3.S2.

During training we used the Adam optimizer (Kingma and Ba, 2015) and a binary cross-entropy loss function. An updated trained model was saved if the validation loss had improved at the end of each epoch. Training was initiated with a learning rate of 10^{-4}, and we implemented a learning rate plateau function where the learning rate would decrease when the model proceeded through 4 training epochs without an improvement in validation loss. We also implemented an early stopping function where the model would quit training after 10 epochs with no validation loss improvement once the learning rate had already reached its lowest allowed value to avoid overfitting. Training stopped automatically because of these functions after 66 epochs, and the training history is shown in Figure 3.S3.

3.3. Results

3.3.1. FakeQuakes Testing Data

After completion of training, we first tested the CNN with the remaining 10% of the FakeQuakes waveforms (combined with noise) which it had not seen in training or validation. We began with a set of binary classification tests. As with the distribution of the training dataset, the testing dataset consisted of 50% earthquake samples containing P-wave arrivals and 50% noise samples, so we wanted to determine how often the model was correct or incorrect about whether a sample contained an earthquake or not. However, our model did not simply output a classification of whether or not it predicted which group a sample fell into; instead, the output was a vector with the same structure as the target arrays used in training. Where Figure 3.3d shows an earthquake-containing sample's target array with the black dashed line with the Gaussian at the P-wave arrival, it also shows the output prediction array with the red dashed line. Figure 3.3e is similar, but for a sample that does not contain an earthquake. To classify whether the model predicted any particular sample to contain a P-wave arrival, we needed to choose a confidence level threshold. If the red prediction array's maximum value exceeded that of the confidence threshold, for that sample the model predicted an earthquake; if the maximum value of the prediction array was lower than the threshold, the model predicted only noise. This binary classification system resulted in four possible outcomes: 1) true positive (TP) cases, where the sample contained an earthquake/P-wave arrival, and the model predicted that the sample contained an earthquake/P-wave arrival, 2) true negative (TN) cases, where the sample contained only noise, and the model predicted that the sample contained only noise, 3) false positive (FP) cases, where the sample contained only noise, but the model predicted that the sample contained an earthquake, and 4) false negative (FN) cases, where the sample contained an earthquake, but the model predicted that the sample contained only noise.

We calculated classification statistics for all of the predictions made on the samples in the FakeQuakes testing dataset using a range of confidence threshold values from 0 to 1. Accuracies were computed with (TP + TN)/(TP + TN + FP + FN), and the results are shown in Figure 3.4a. At a threshold of 0, any value of a prediction array higher than 0 resulted in the prediction that a sample did contain an earthquake, so the model predicted that every sample contained an earthquake. Because the FakeQuakes testing dataset was 50% earthquakes and 50% noise, the accuracy at this threshold was 50%. Similarly, at a threshold of 1, the model would only predict that a sample contained an earthquake if the confidence value was 1, so it predicted that every sample contained only noise, resulting in a 50% accuracy as well. We achieved the highest overall accuracy of 59.66% at an intermediate threshold of 0.135.

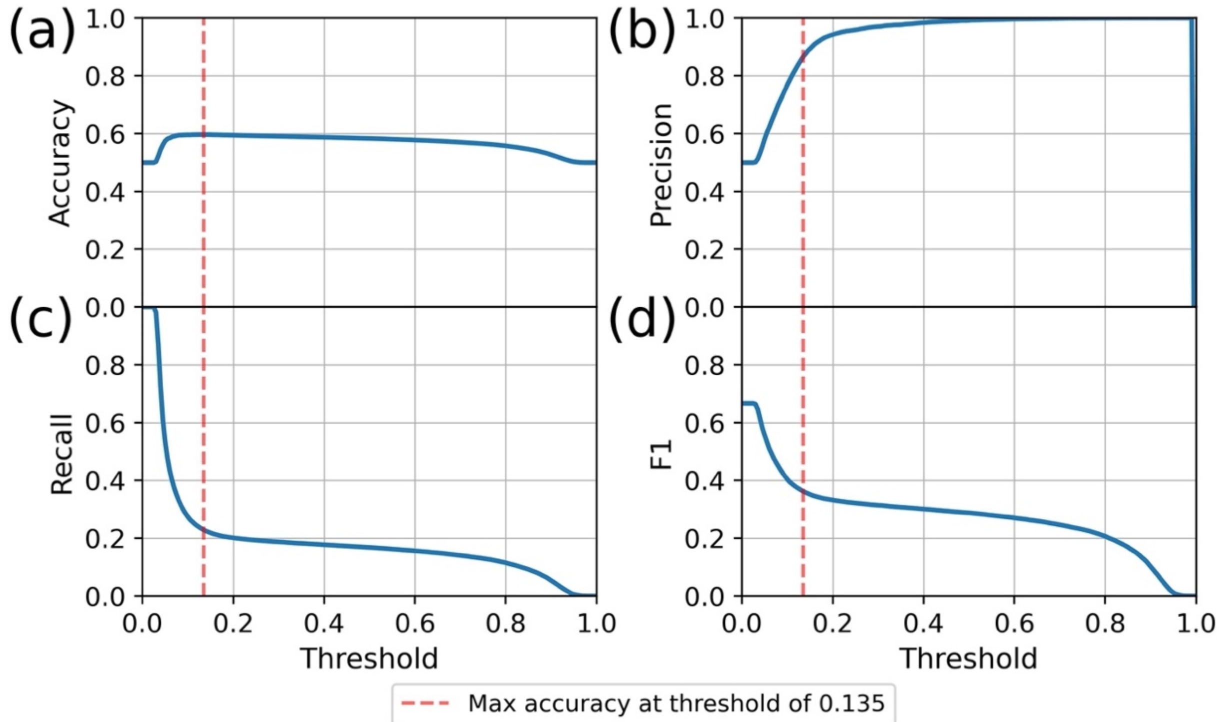


Figure 3.4. Results of the testing of the CNN model on the remaining 10% of the FakeQuakes synthetic displacement waveforms which it never saw in training. **(a)** Accuracy, **(b)** precision, **(c)** recall, and **(d)** F1 score were calculated across a range of confidence thresholds from 0 to 1, where the threshold determined the cutoff of the maximum confidence of the predicted P-wave arrival Gaussian (Figure 3.3d-3.3e). A maximum confidence above the threshold meant a prediction that the sample contained a P-wave arrival from an earthquake, and a maximum confidence below the threshold meant a prediction that the sample contained only noise. The threshold at which maximum accuracy was achieved was 0.135, which is indicated on all four panels with the vertical red dashed line.

We also computed the precision and recall of the model for the same range of thresholds. Precision was computed with TP/(TP + FP), representing how often the model was correct about the samples that did actually contain earthquakes (Figure 3.4b). The highest precisions were found from thresholds of 0.835-0.99. Using a threshold this high in practice would result in limited false positives, but it would also increase the number of false negatives/missed earthquakes. Recall was computed with TP/(TP+FN), and represented how successful the model was at finding all of the earthquakes in the dataset (Figure 3.4c). Recall was highest at very low thresholds from 0-0.025. However, using a threshold this low would result in very high numbers of false positives, as the model would predict nearly every sample to contain an earthquake. We also computed F1 scores for the threshold range using 2 * (precision * recall)/(precision + recall) to explore a metric that balances the contrasting precision and recall (Figure 3.4d), but the highest F1 score was also reached at a low threshold of 0.025, which would lead to a preponderance of false positives.

Figure 3.5 shows confusion matrices plotted for the three threshold values that optimized recall/F1 score (0.025) (Figure 3.5a), accuracy (0.135) (Figure 3.5b), and precision (0.835) (Figure 3.5c). This demonstrated how using a low threshold to optimize recall or F1 score resulted in the model correctly identifying all of the earthquake samples, but incorrectly identifying nearly all of the noise samples, leading to a very high FP rate. Increasing the threshold to 0.135 where accuracy was optimized reduced the number of FPs and led to the model correctly identifying the majority of the noise samples (TN), but the FN rate of missed earthquakes increased as the TP rate decreased. Finally, choosing a high threshold that optimized precision eliminated the FPs, but further reduced the number of TP earthquakes that the model correctly identified. For the following analyses, we focused on the accuracy-optimized threshold of 0.135 for a more balanced view of the model's performance.

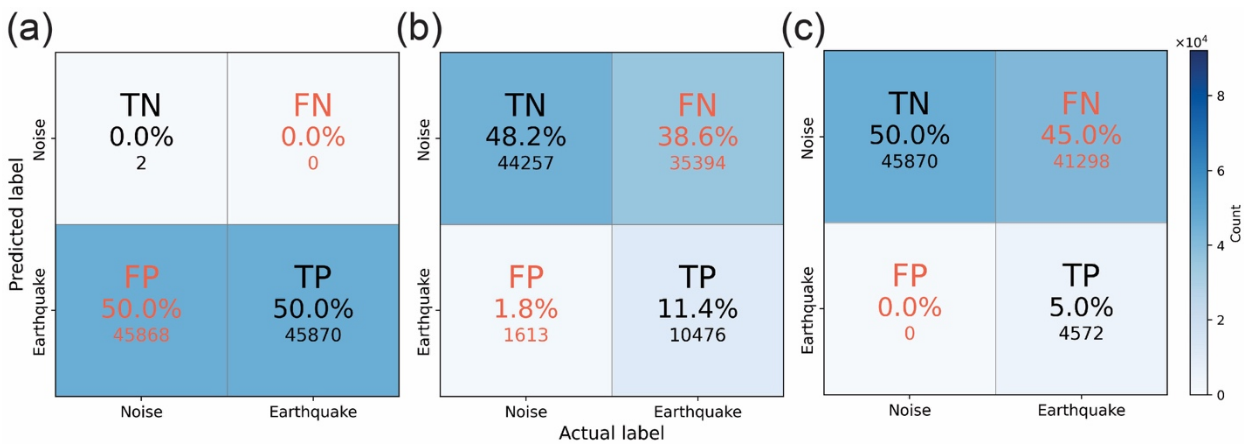


Figure 3.5. Confusion matrices of the results of the model predictions on the FakeQuakes testing data compared to the true sample labels for **(a)** a best recall and F1 score confidence threshold of 0.025, **(b)** the best accuracy threshold of 0.135, and **(c)** a best precision threshold of 0.835. 'TN' indicates "true negative" cases where the sample contained only noise and the model predicted only noise, 'FN' indicates "false negative" cases where the sample contained a P-wave arrival from an earthquake but the model predicted only noise, 'FP' indicates "false positive" where the sample contained only noise but the model predicted there was an earthquake, and 'TP' indicates "true positive" where the sample contained an earthquake and the model predicted that there was an earthquake. The colorbar indicates how many samples of each case are in a particular box in the matrix, the percentages indicate the percentage of the total number of predictions that each of the four categories contain, and the numbers under the percentages indicate the number of predictions in each category. The FP and FN boxes that represent incorrect predictions by the model have red colored text.

For the samples that did contain earthquakes, we wanted to explore how well the model performed based on the characteristics of those waveforms or the earthquake sources associated with them. Figure 3.6a shows the accuracy of the model's predictions across the range of hypocentral distances in the FakeQuakes testing dataset, and we observed that while the model tended to be slightly more accurate for stations closer to the hypocenters of the earthquakes than farther ones, the overall accuracy in each bin was low, indicating that hypocentral distance was not a primary control on the accuracy of the model's predictions.

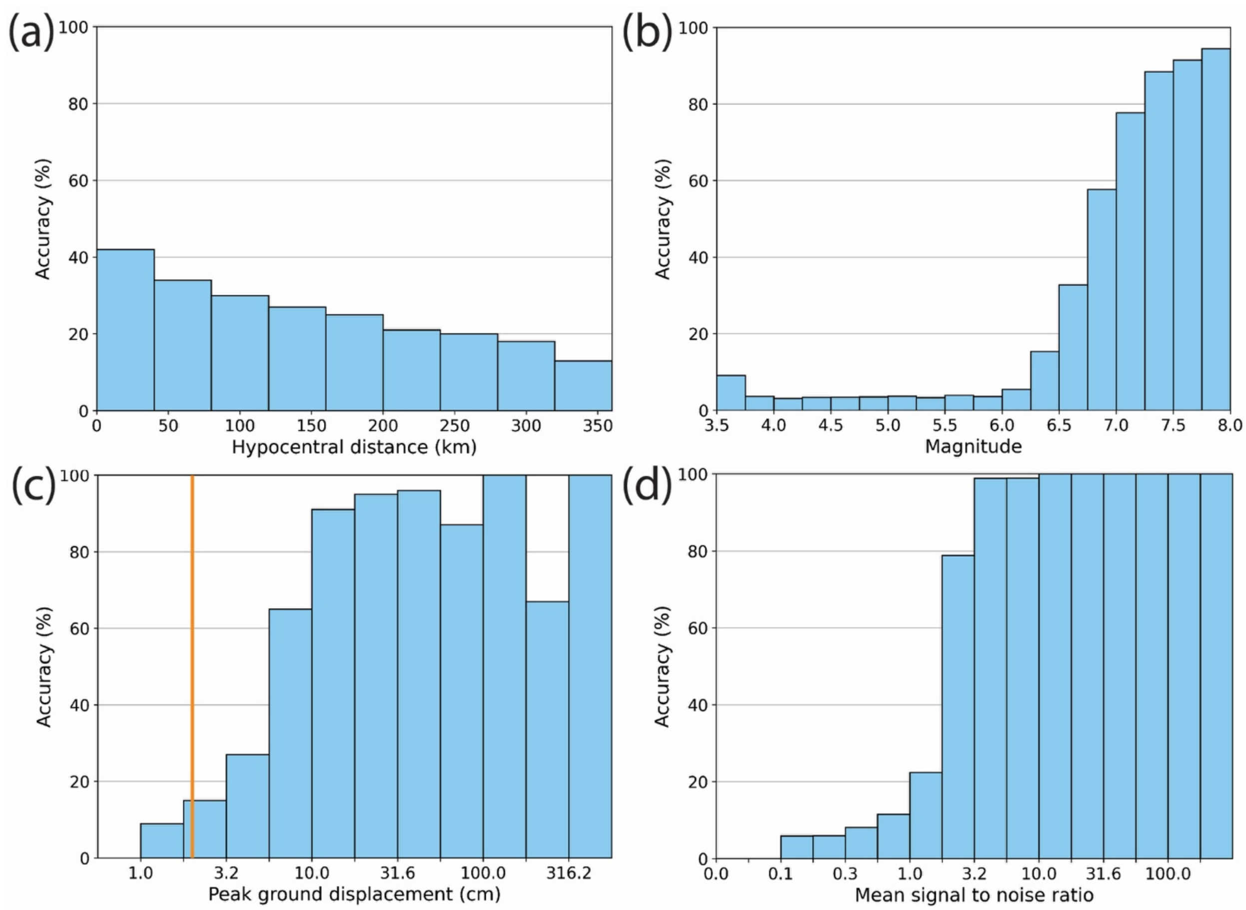


Figure 3.6. Plots showing the accuracy of the CNN's model predictions on the FakeQuakes testing data compared to binned source and waveform characteristics such as **(a)** hypocentral distance in kilometers, **(b)** earthquake magnitude, **(c)** waveform peak ground displacement in centimeters, and **(d)** the mean signal to noise ratio (SNR) (computed as the mean of the SNR for the N, E, and Z components). All plots show results for the confidence threshold of 0.135. Note in (c) and (d) that the x-axis scales logarithmically, not linearly. (c) also has a vertical orange line indicating the approximate noise level of GNSS data (~2 cm).

Figure 3.6b shows the accuracy of the model's predictions across the range of earthquake magnitudes in the testing data, and there we observed a much stronger correlation for this parameter, with high accuracies exceeding 75% only for earthquakes ≥**M**7. This aligns with the observation from Figure 3.6c where accuracy was plotted versus peak ground displacement (PGD) and showed a similar correlation, as PGD does increase with earthquake magnitude (Crowell et al, 2013; Ruhl et al., 2019). Note that for this figure, the scaling of the x-axis is logarithmic. Waveforms containing earthquakes with higher PGDs would be expected to be more easily distinguishable from the noise, likely increasing the model's ability to identify them. Further, Figure 3.6d shows the accuracy of the model's predictions compared to the mean signal to noise ratio (SNR) computed as the average of the SNR from each of the three components of data. The scaling of the x-axis in this figure is logarithmic as well, and we observed that by the time the SNR exceeds ~2, the model is ~80% accurate, and by an SNR of ~3, it is nearly 100% accurate.

We next explored how well the model was able to identify the correct position of the P-wave arrival in the TP cases with earthquakes, again using the intermediate accuracy-optimizing threshold of 0.135. Figure 3.7 shows a histogram of the distribution of differences between the actual P-wave arrival time and the model-predicted arrival time for the FakeQuakes testing dataset. Because 95% of the picks fell within a 30 second difference, we only show that range in the plot. 84% of the picks fell within 5 seconds of the actual P-wave arrival time, and 50% fell within 1 second. These differences are larger than those typically observed for seismic arrival time data from machine learning studies (e.g., Zhu and Beroza, 2019), but this was not unexpected given the noise levels of the GNSS data, as in many cases the arrivals were not plainly visible.

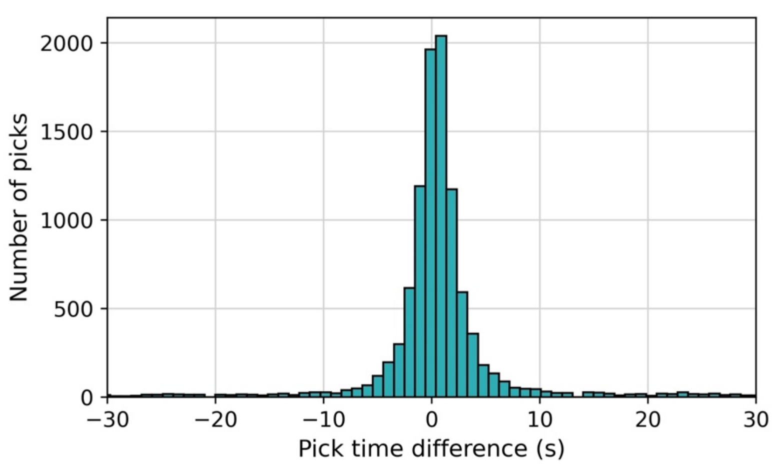


Figure 3.7. A histogram of the distribution of differences between the actual P-wave arrival time and the CNN predicted arrival time for the FakeQuakes testing dataset using the confidence threshold of 0.135. 95% of the samples fall within a 30 second difference, 84% within 5 seconds, and 50% within 1 second.

3.3.2. Real Testing Data

After testing the trained model on the remaining FakeQuakes data, we then tested it with the real data from the stations and earthquakes around Ridgecrest, CA shown in Figure 3.1. Figure 3.8 shows the same results of this testing for the real data as Figure 3.4 for the FakeQuakes data, where for a range of thresholds from 0 to 1 we calculated the accuracy (Figure 3.8a), precision (Figure 3.8b), recall (Figure 3.8c), and F1 score (Figure 3.8d) for the model's predictions.

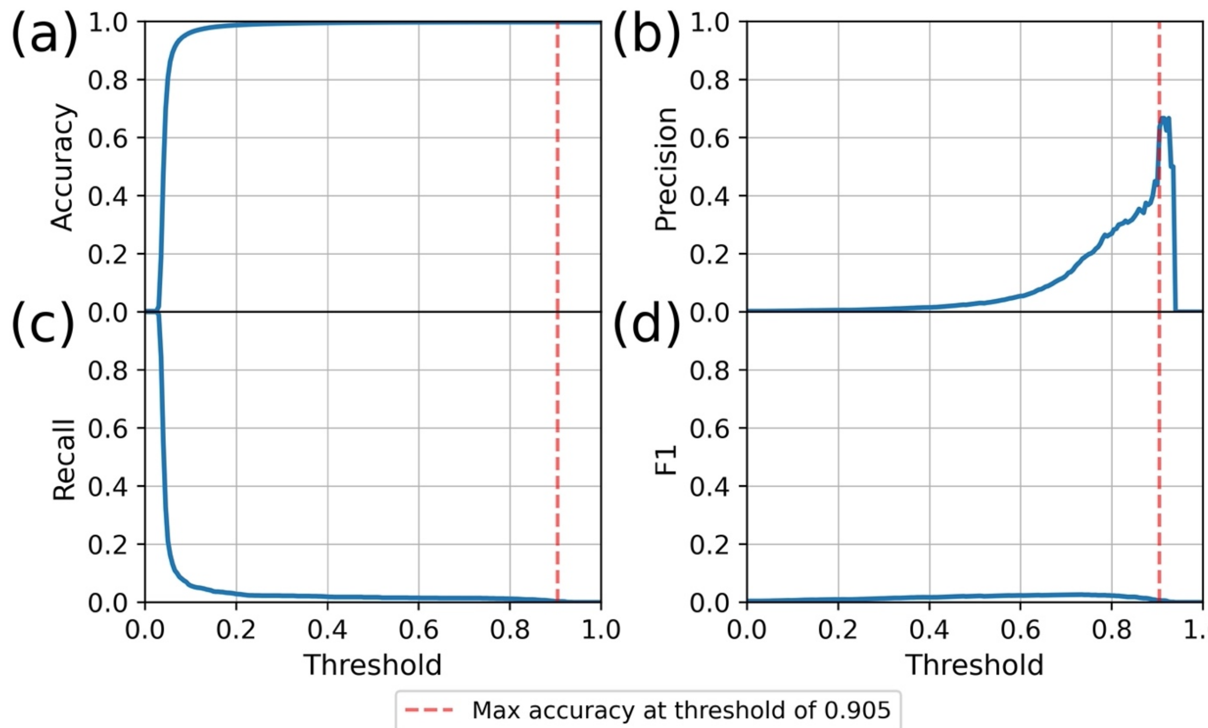


Figure 3.8. Results of the testing of the CNN model on the real testing dataset. **(a)** Accuracy, **(b)** precision, **(c)** recall, and **(d)** F1 score were calculated across a range of confidence thresholds from 0 to 1, where the threshold determined the cutoff of the maximum confidence of the predicted P-wave arrival Gaussian (Figure 3.3d-3.3e). A maximum confidence above the threshold meant a prediction that the sample contained a P-wave arrival from an earthquake, and a maximum confidence below the threshold meant a prediction that the sample contained only noise. The threshold at which maximum accuracy was achieved was 0.905, which is indicated on all four panels with the vertical red dashed line.

The results for the real testing data looked very different from the results for the FakeQuakes testing data. Whereas the FakeQuakes data was split evenly to be half earthquakes and half noise, in the real testing dataset only 2123 of the ~994,000 samples actually contained earthquake P-wave arrivals (Figure 3.S1a). Therefore, as Figure 3.8a shows, with a very high confidence threshold the model predicted that almost every sample did not contain an earthquake, resulting in a very high accuracy because so few of the samples did actually contain earthquakes. The highest accuracy for the real testing dataset was 99.79% at a threshold of 0.905. Precision was highest at thresholds of 0.91-0.925 (Figure 3.8b), and recall was highest at very low thresholds of 0-0.025 (Figure 3.8c). F1 scores for this dataset remained very low across the entire threshold range (Figure 3.8d).

Because of the low number of true earthquake samples in the dataset from which TP or TN results could be produced, we also plotted the number of each of the TP (green), TN (blue), FP (red), and FN (orange) cases across the full threshold range, which can be seen in Figure 3.9. The most dramatic change in outcomes occurred at very low thresholds around 0.04, which can be visualized more clearly in the cropped inset plot in Figure 3.9. Lower than this threshold, the TP and FP cases were maximized because at low thresholds, the model predicted that every sample contained an earthquake. The number of TPs and FPs quickly decreased asymptotically to zero higher than this threshold. In contrast, lower than this threshold there were no FN or TN cases, but higher than the threshold they quickly rose in opposition to the decrease in TPs and FPs. This made it difficult to choose an appropriate threshold for analysis, as ideally we would want to maximize TP and TN cases and minimize FP and FN cases, and there was no threshold where both goals could be simultaneously achieved.

For this reason, the next analyses we present we performed at several different thresholds: 1) 0.025, where recall was maximized and all 2123 earthquakes were found, 2) 0.04, where the crossover occurred in Figure 3.9, 3) 0.135, which was the highest accuracy threshold found with the FakeQuakes testing dataset, and 4) 0.905, which maximized the accuracy for the real testing dataset and was just slightly below where precision was optimized as well.

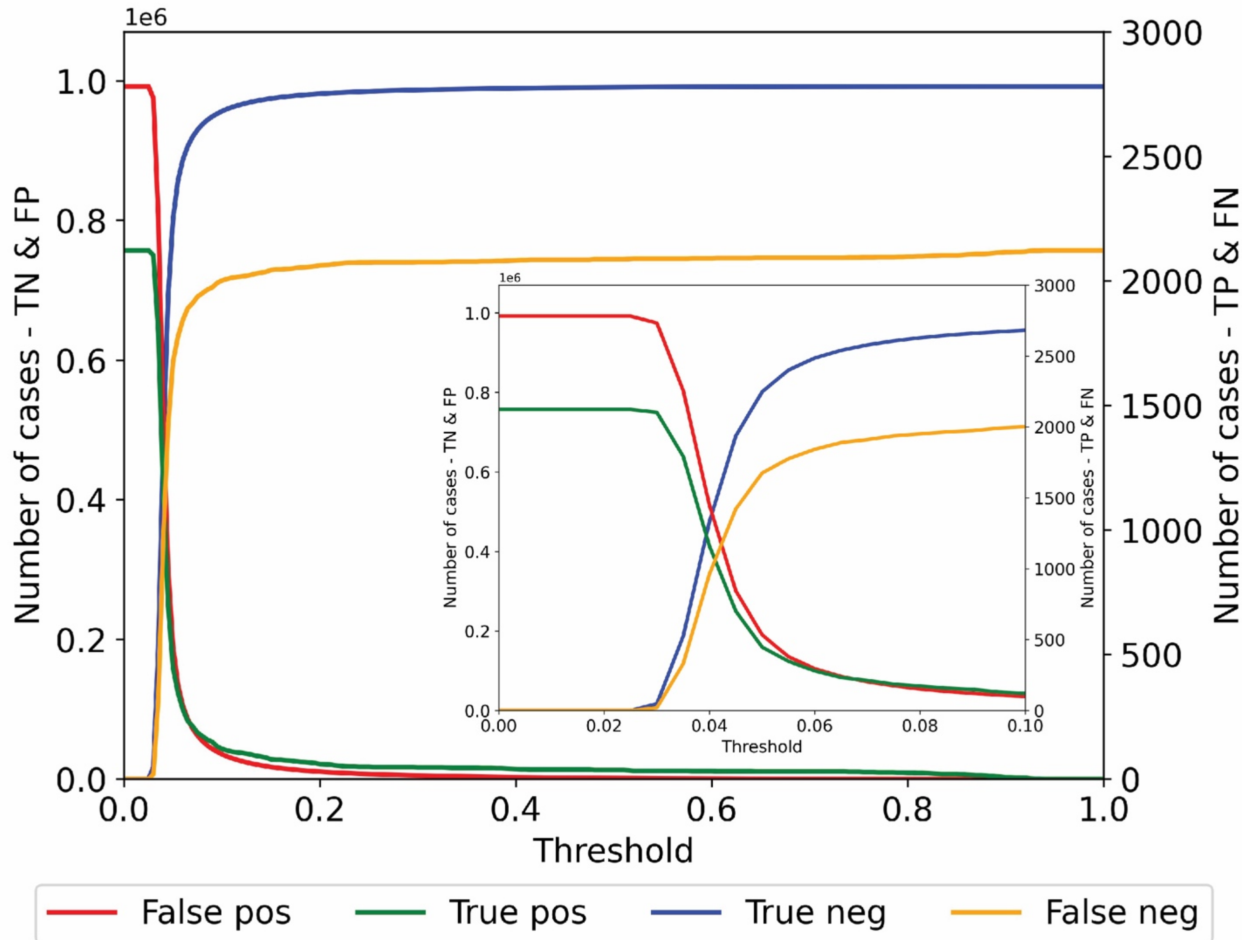


Figure 3.9. A plot showing how the number of true positive (green), true negative (blue), false positive (red), and false negative (orange) cases in the real testing data results changes across the range of confidence thresholds. The inset plot shows a closer-cropped look at the region of the larger graph where the lines converge and cross over around a threshold of 0.04.

We plotted confusion matrices for each of these four thresholds in Figure 3.10. Figure 3.10a, where we used the recall-optimized threshold of 0.025, showed that there were 2123 TP cases, capturing all of the earthquakes in the real testing dataset, but the vast majority of the results for this threshold were FPs. In Figure 3.10b we used the threshold of 0.04 from the crossover in Figure 3.9, and we observed that with this threshold, approximately 55% of the real earthquakes were captured as TP cases, with the remainder moving to the FN category. The number of FP cases had also decreased to approximately half the size of that in Figure 3.10a, instead increasing the number of TN cases.

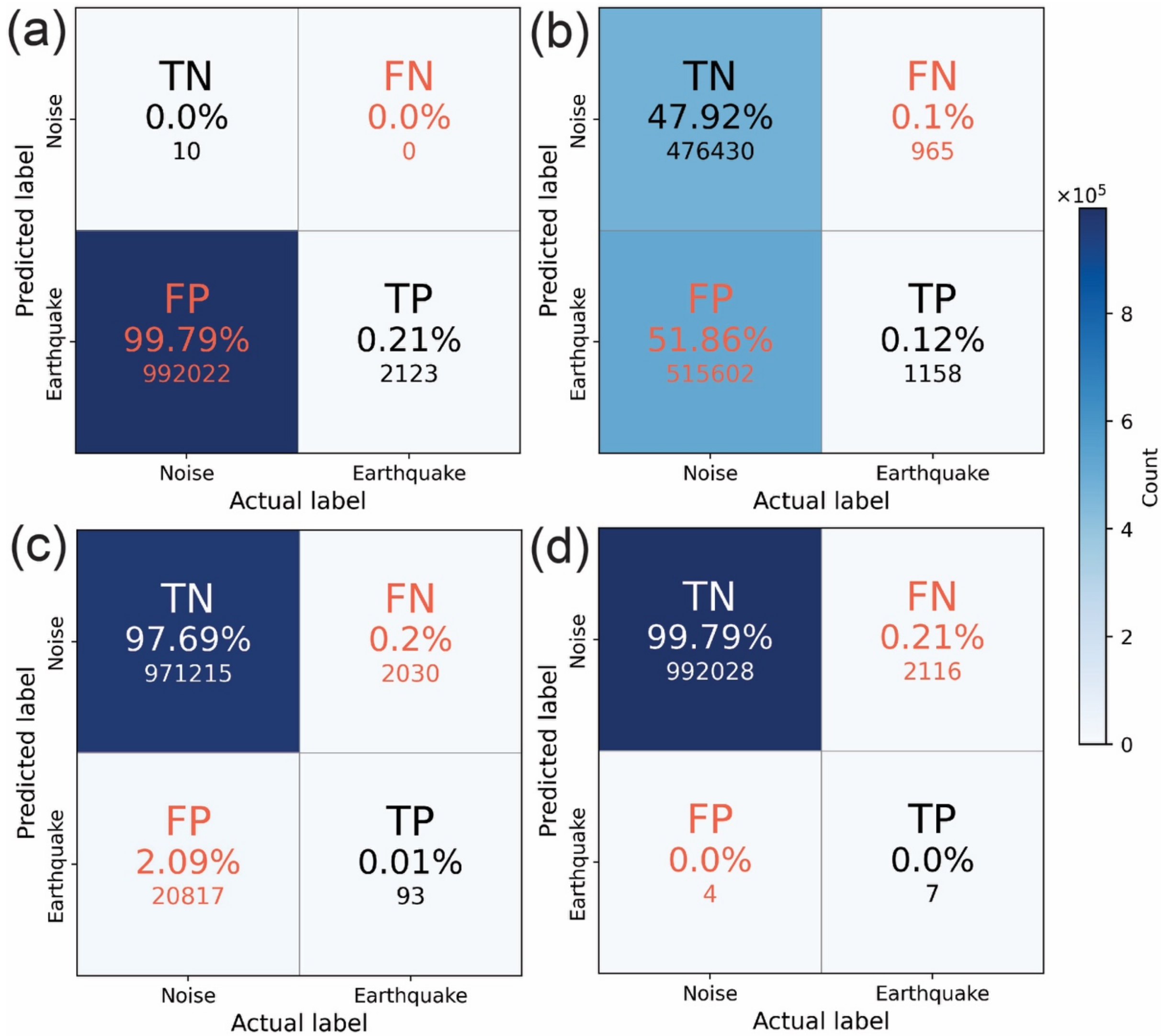


Figure 3.10. Confusion matrices of the results of the model predictions on the real testing data compared to the true sample labels for **(a)** a threshold of 0.025, where the model captures all of the earthquakes (and recall is maximized), **(b)** the threshold of 0.04 at which the crossover in Figure 3.9 occurs, **(c)** the best accuracy threshold from the FakeQuakes testing results of 0.135, and **(d)** the best accuracy threshold from the real data testing results of 0.905, which is just shy of where precision is maximized for this dataset as well. 'TN' indicates "true negative" cases where the sample contained only noise and the model predicted only noise, 'FN' indicates "false negative" cases where the sample contained a P-wave arrival from an earthquake but the model predicted only noise, 'FP' indicates "false positive" where the sample contained only noise but the model predicted there was an earthquake, and 'TP' indicates "true positive" where the sample contained an earthquake and the model predicted that there was an earthquake. The colorbar indicates how many samples of each case are in a particular box in the matrix, the percentages indicate the percentage of the total number of predictions that each of the four categories contain, and the numbers under the percentages indicate the number of predictions in each category. The FP and FN boxes that represent incorrect predictions by the model have red colored text.

In Figure 3.10c, using the threshold of 0.135 from the best accuracy results of the FakeQuakes testing data reduced the number of actual earthquakes captured (TPs) to only ~4% of the total number of earthquakes, but the number of FP cases significantly improved. Finally, for Figure 3.10d we used the threshold of 0.905 which maximized the accuracy in the real testing data and was close to the precision-optimized threshold range as well. In this case, the number of TPs and FPs were both very low, as the model was predicting that almost no samples contained P-wave arrivals from earthquakes. But because the number of TN cases was so high, and the vast majority of the samples in the real testing dataset were just noise, the overall accuracy was very high.

In Figure 3.11, we took a closer look at the model's performance on the 2123 samples that did contain earthquake P-wave arrivals. These figures show scatter plots of the samples by their associated magnitude, PGD, and mean SNR, where black points represent those samples for which the model correctly identified that there was a P-wave arrival in the sample, and red dots represent those samples that the model missed identifying. The top row of plots (Figure 3.11a-c) shows the results for the confidence threshold of 0.135, and the bottom row (Figure 3.11d-f) shows the results for the threshold of 0.905. The two lower thresholds shown in Figure 3.10a-b were omitted because they were essentially entirely black dots. The blue dashed horizontal lines on each plot show the average magnitude, PGD, and mean SNR for all 2123 earthquakes for the relevant parameter for each panel. The average magnitude in the real testing dataset was 4.82, the average PGD was 2.93 centimeters, and the average mean SNR was 1.18. Figure 3.11a shows that the model correctly identified the samples consisting of waveforms from the largest earthquake in the dataset (the **M**7.1 Ridgecrest mainshock) and many of the samples from the **M**6.4 Ridgecrest foreshock, but the results were more scattered for magnitudes smaller than this. In Figure 3.11b we observed that there was not a particularly strong correlation between PGD and whether the model identified the samples correctly; there are several very high PGD waveforms which the model missed. Figure 3.11c plots the waveform SNRs, and we observed that in this case, the majority of the anomalously high SNR samples were correctly identified by the model. For this threshold, the average magnitude associated with the earthquake waveforms the model correctly identified was 5.67, and the average magnitude associated with the earthquake waveforms the model failed to identify was 4.78. The average PGD for the correctly identified samples was 4.74 cm, and the average PGD for the missed samples was 2.85 cm. The average mean SNR for the correctly identified samples was 3.48, and the average mean SNR for the missed samples was 1.06.

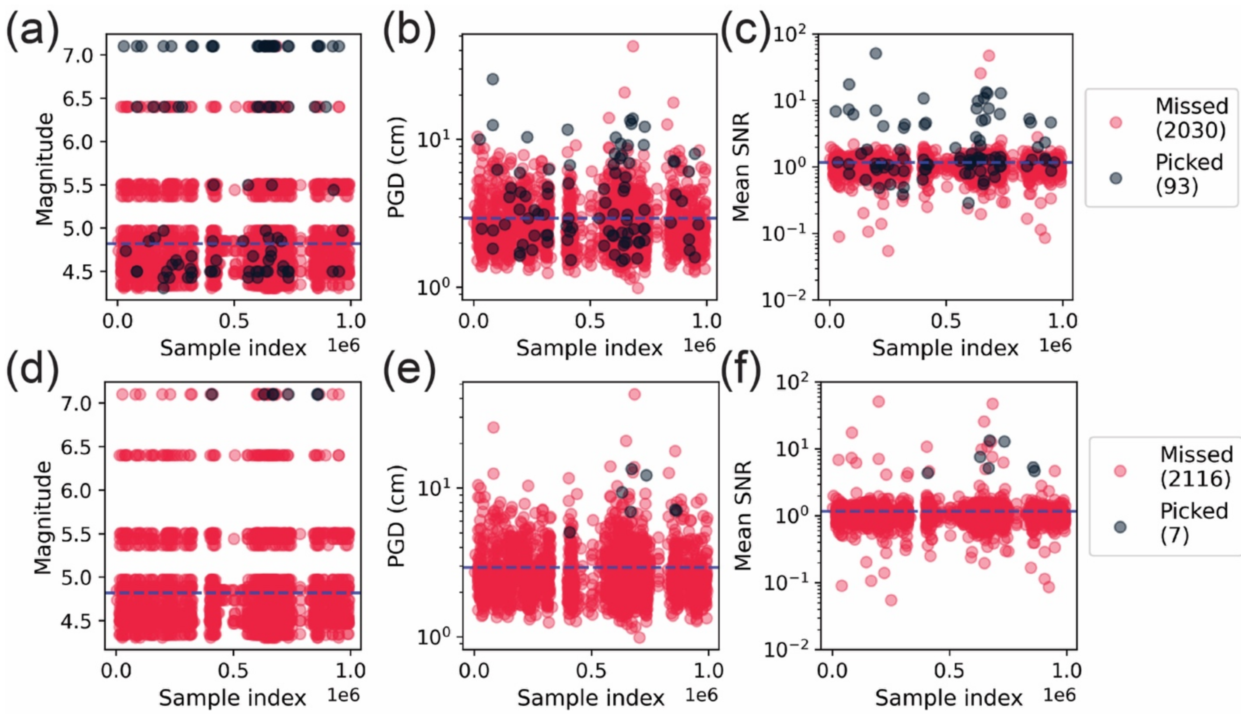


Figure 3.11. Plots showing the outcomes of predictions for the 2123 individual waveforms that included P-wave arrivals in the real testing dataset. Waveforms where the model successfully identified the earthquakes are colored black, and waveforms where the model did not identify the earthquakes are colored red. For the confidence threshold of 0.135, **(a)** shows the results compared to earthquake magnitude, **(b)** shows the results compared to waveform peak ground displacement (PGD) in centimeters, and **(c)** shows the results compared to waveform mean signal to noise ratio (SNR). For the confidence threshold of 0.905, **(d)** shows the results compared to earthquake magnitude, **(e)** shows the results compared to waveform PGD in centimeters, and **(f)** shows the results compared to waveform mean SNR. The mean values of the parameter shown in each subplot for the dataset are plotted with horizontal blue dashed lines.

When we increased the confidence threshold to 0.905, the model was only able to correctly pick samples from the **M**7.1 Ridgecrest earthquake (Figure 3.11d). Figures 3.11e and 3.11f also show that the few samples the model did correctly identify as containing earthquakes had substantially higher PGDs and SNRs than the means of the dataset, respectively. For this threshold, the average magnitude associated with the earthquake waveforms the model correctly identified was 7.1 (only the Ridgecrest mainshock), and the average magnitude associated with the earthquake waveforms the model failed to identify was 4.81. The average PGD for the correctly identified samples was 8.75 cm, and the average PGD for the missed samples was 2.91 cm. The average mean SNR for the correctly identified samples was 7.66, and the average mean SNR for the missed samples was 1.15.

Finally, Figure 3.12 shows the model's performance at identifying the correct position of the P-wave arrival in the TP cases with earthquakes for the real testing dataset using all four previously discussed thresholds (0.025 for Figure 3.12a, 0.04 for Figure 3.12b, 0.135 for Figure 3.12c, and 0.905 for Figure 3.12d). We generally observed that the model's performance for the real data was degraded compared to that of the FakeQuakes data as shown in Figure 3.7, with a much wider distribution of large differences between the true and predicted arrival times. For the real testing dataset and the threshold of 0.135 (the same threshold as used with the FakeQuakes data to produce Figure 3.7), only 63% of the picks fell within a 30 second difference as opposed to 95% of the FakeQuakes data, 38% fell within 5 seconds as opposed to 84% with the FakeQuakes data, and 16% fell within 1 second as opposed to 50% with the FakeQuakes data. This difference may be accounted for by the difference in the distribution of magnitudes, PGDs, and SNRs in the FakeQuakes testing dataset compared to the real testing dataset, histograms of which can be found in Figure 3.S4. For example, in the FakeQuakes dataset 30% of the waveforms came from earthquakes over **M**6.5, 10.7% of the waveforms had PGDs over 5 cm, and 5.3% of the waveforms had SNRs over 2, whereas the real testing dataset only had 1.5%, 7.1%, and 2% of the waveforms in the same categories, respectively. As shown previously, the model performed better with data from larger earthquakes with higher PGDs and SNRs, and the poorer performance with respect to the timing of the picks in the real testing dataset could reflect the relative lack of higher-quality waveforms in it.

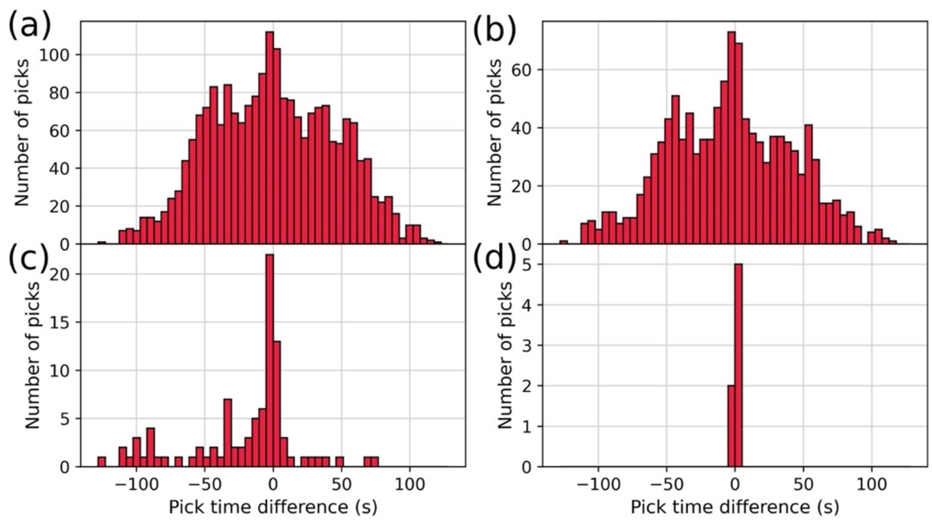


Figure 3.12. Histograms of the distributions of differences between the actual P-wave arrival time and the CNN predicted arrival time for the real testing dataset using confidence thresholds of **(a)** 0.025, **(b)** 0.04. **(c)** 0.135, and **(d)** 0.905.

3.4. Discussion

The results presented show that the performance of the model heavily depended on what confidence threshold was chosen to distinguish between samples which were determined to contain earthquakes, and which ones did not. The proper choice of threshold would likely vary depending on use case. Based on the high numbers of false positive results observed when lower confidence thresholds were used in this study, which would be required if the goal was to capture the majority of true earthquakes in the data, as discussed in the Introduction and Background it seems that a more appropriate use case of this model would be to act as a "filter" to restrict the data used in GNSS magnitude calculation algorithms to only those waveforms which are high enough quality (i.e., low enough noise) to provide useful information about the earthquake, rather than to construct a standalone GNSS EEW system that can trigger itself by detecting earthquakes.

For this study, we explored the CNN model's potential utility in a realistic scenario where the model contributes to the GNSS side of an EEW system such as ShakeAlert, where the GNSS analysis algorithm (GFAST) is triggered by earthquake detections made using seismic data. GFAST computes earthquake magnitudes in real-time by calculating PGD for all stations with available data once it is triggered by the seismic algorithm, and then using the estimated station hypocentral distance to compute magnitude using the scaling relationship

$$\log_{10}(PGD) = A + BM_w + CM_w \log_{10}\left(r_{hyp}\right), \quad (3.1)$$

where *PGD* is the peak ground displacement, *A*, *B*, and *C* are coefficients, M_w is the moment magnitude, and r_{hyp} is the hypocentral distance in kilometers (Crowell et al., 2013, 2016; Murray et al., 2023). Crowell et al. (2016) found $A = -6.687$, $B = 1.5$, and $C = -0.214$ using PGD in centimeters. GFAST also enforces a travel-time mask where only stations within a 3 km/s radius from the earthquake origin time are included in the analysis, as this would conservatively ensure that the larger S-waves where the PGD would occur had already arrived at those stations. An exponential hypocentral distance weighting is also implemented where near-source stations are emphasized more strongly in the inversion for event magnitude which includes all station estimates.

We wanted to test how our model could potentially improve the magnitude estimates made by an algorithm like GFAST by excluding any station waveforms where an earthquake was not detected. Because our model is not currently constructed to operate in real time, only working with

128-second-long waveforms, it is not a perfect recreation of how GFAST works; instead, we present a first-order demonstration of its possible utility. For this test, we re-formatted the 45,870 earthquake-containing waveforms from the FakeQuakes testing dataset so that instead of the P-wave arrival being randomly positioned in the waveform, each 128-second-long waveform began at the origin time of the earthquake. This reduced the number of waveforms where the P-wave arrived at the end of the sample, therefore reducing the number of waveforms where no S-waves where the PGD would occur were in the sample. We still combined each of these waveforms with a random sample of the GNSS real-time noise, but we did not include the noise-only waveforms in this test, as a seismic-triggered system would remove the possibility that any waveforms contain only noise (even if a waveform is so buried by the GNSS noise that it cannot be visualized).

For a 128-second-long waveform, a 3 km/s travel time mask like the one used by GFAST would permit stations within 384 km of the hypocenter to contribute to a magnitude estimation. However, the largest hypocentral distance in our FakeQuakes testing dataset was 350 km, so a 3 km/s mask would not actually exclude any waveforms from the magnitude computation. We did, however, implement the distance weighting used by Crowell et al. (2016), where the weight for any particular station is

$$w_i = \exp\left(-\frac{r_{epi,i}^2}{8r_{epi,min}^2}\right), \quad (3.2)$$

where $r_{epi,i}$ is the epicentral distance for the ith station, and $r_{epi,min}$ is the epicentral distance for the station closest to the epicenter. We also included the requirement that an earthquake must have 4 stations reporting individual PGD magnitudes to compute the weighted event-average magnitude.

As the control to compare to, we computed the event-averaged PGD magnitudes for the entire earthquake-containing FakeQuakes testing dataset without any kind of GFAST-like modifications or additional filtering. The mean squared error (MSE) between the true earthquake magnitudes and the calculated event-averaged PGD magnitudes for this unfiltered dataset was 2.557. Then, in addition to the GFAST-like magnitude computation with hypocentral distance weighting described above, we also computed event-averaged PGD magnitudes using an SNR filter (where we only allowed waveforms with SNRs > 1, 2, and 3 to contribute), as well as event-averaged PGD magnitudes using only waveforms which our CNN model detected as "true positives" (TPs) using both the accuracy-optimized threshold of 0.135 and the precision-optimized threshold of 0.835. The results of this test are shown in Figure 3.13. In each of the panels, the unfiltered control data are shown in red, whereas the results after whatever filtering operation was implemented are shown in blue. Each panel also has a black dashed line that represents a one-to-one relationship between true magnitude and the event-averaged magnitude calculated from the PGD scaling law, with the gray dashed lines indicating half a magnitude unit above and below the one-to-one line.

Figure 3.13a shows the results for the GFAST-like test where the individual station PGD magnitudes were weighted by hypocentral distance in the event-averaged magnitude computation, and the four station minimum requirement for each event was implemented. Because the 3 km/s travel time mask was irrelevant for this test due to the 350 km maximum hypocentral distance in the dataset, and there were more than four stations available for each event, no events were excluded from the final plot. While the implementation of the distance weighting did bring some of the smaller magnitude earthquakes down in calculated PGD magnitude closer to the one-to-one line, there were still many earthquakes whose calculated event-averaged magnitude were far higher than the true magnitude. As a result, the MSE for the GFAST-like test was 1.938.

Figures 3.13b-d show the results when SNR filters were used. Figure 3.13b only included waveforms with SNR > 1 in the event-averaged magnitude computation, which still allowed 99.97% of the earthquakes through. We observed that this does not improve the performance over the unfiltered dataset; the MSE for this test was even slightly worse than for the unfiltered data at 2.625.

Figure 3.13 (next page)**.** Plots of the event-averaged PGD magnitudes for the FakeQuakes earthquakes in the testing dataset using several different methods to restrict the dataset. Red points indicate event-averaged PGD magnitudes calculated using all of the data with no filters, and blue points indicate event-averaged PGD magnitudes calculated with a restriction on the waveforms permitted to contribute. The black dashed lines represent the one-to-one relationship between true magnitude and the event-averaged magnitude calculated from the PGD scaling law, with the gray dashed lines indicating half a magnitude unit above and below the one-to-one line. **(a)** Contributing station magnitudes computed using a GFAST-like exponential hypocentral distance weighting scheme and requiring a minimum of 4 stations per event. **(b)** Contributing stations limited to those with waveform SNRs > 1. **(c)** Contributing stations limited to those with waveform SNRs > 2. **(d)** Contributing stations limited to those with waveform SNRs > 3. **(e)** Only stations with "true positive" results as picked by the CNN developed in this study were allowed to contribute, using the accuracy-optimized confidence threshold of 0.135. **(f)** Only stations with "true positive" results as picked by the CNN developed in this study were allowed to contribute, using the precision-optimized confidence threshold of 0.835.

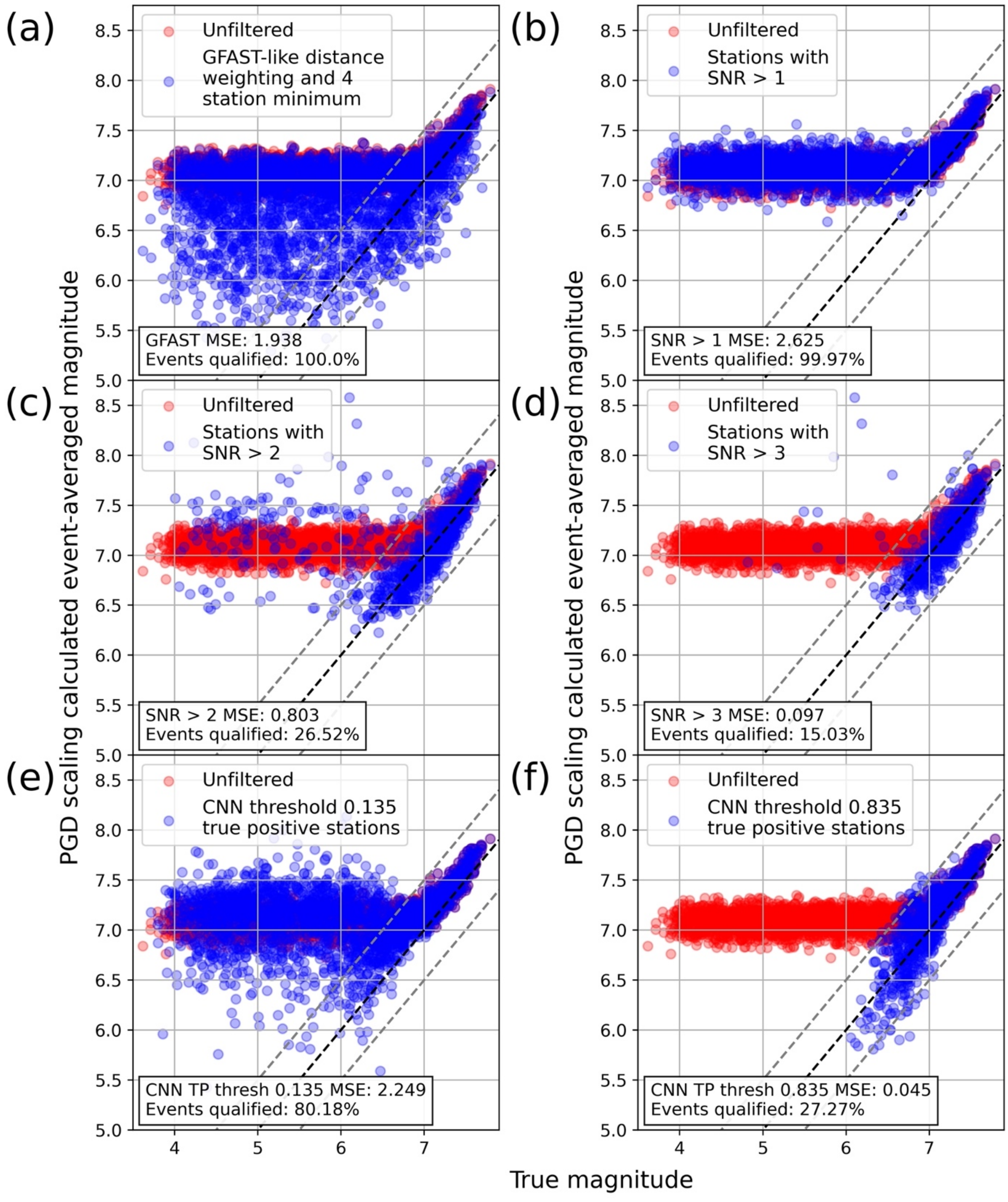


Figure 3.13c only permitted use of the earthquake waveforms with SNR > 2, which was approximately where in the FakeQuakes testing dataset the accuracy of the predictions began to show a dramatic improvement (Figure 3.6d). With this limit, only 26.52% of the earthquakes had

event-averaged magnitudes computed for them, which removed the majority of the smaller-magnitude events where the difference between the true and computed magnitude was large. At magnitudes below ~**M**7, though, the magnitudes of what events remained still were not calculated to be close to the true magnitudes, resulting in an MSE of 0.803. Figure 3.13d further restricted the permitted waveforms to only those with SNR > 3, which reduced the number of events that had computed magnitudes to 15.03%. While some outliers remained, the magnitudes computed for the remaining events above ~**M**6.5 fell close to the one-to-one line, resulting in an MSE of 0.097.

Figures 3.13e and 3.13f show the results of the tests when our CNN model was used to filter out waveforms, in this case by only using those with TP earthquake detections for the event-averaged PGD magnitude calculations. In Figure 3.13e we used the confidence threshold of 0.135 that maximized the model's performance in terms of accuracy on the FakeQuakes testing dataset as discussed in the Results section. This permitted magnitude estimations for 80.18% of the earthquakes, but included many of the smaller magnitude earthquakes that had high errors compared to the true magnitudes. This resulted in an MSE of 2.249, which is better than that of the unfiltered data and the SNR > 1 filter but worse than the GFAST-like distance weighting approach. When we instead used the confidence threshold of 0.835 that maximized precision (Figure 3.13f), the percentage of permitted events was reduced to 27.27%, but the MSE improved to 0.045. This was the lowest MSE of any of the filters we used, and as Figure 3.13f shows, the vast majority of the magnitude estimates for earthquakes over **M**6 fell close to the one-to-one line and between the gray lines that indicate half a magnitude unit of difference. The scatter in the results around the one-to-one line also improved with increased earthquake magnitude.

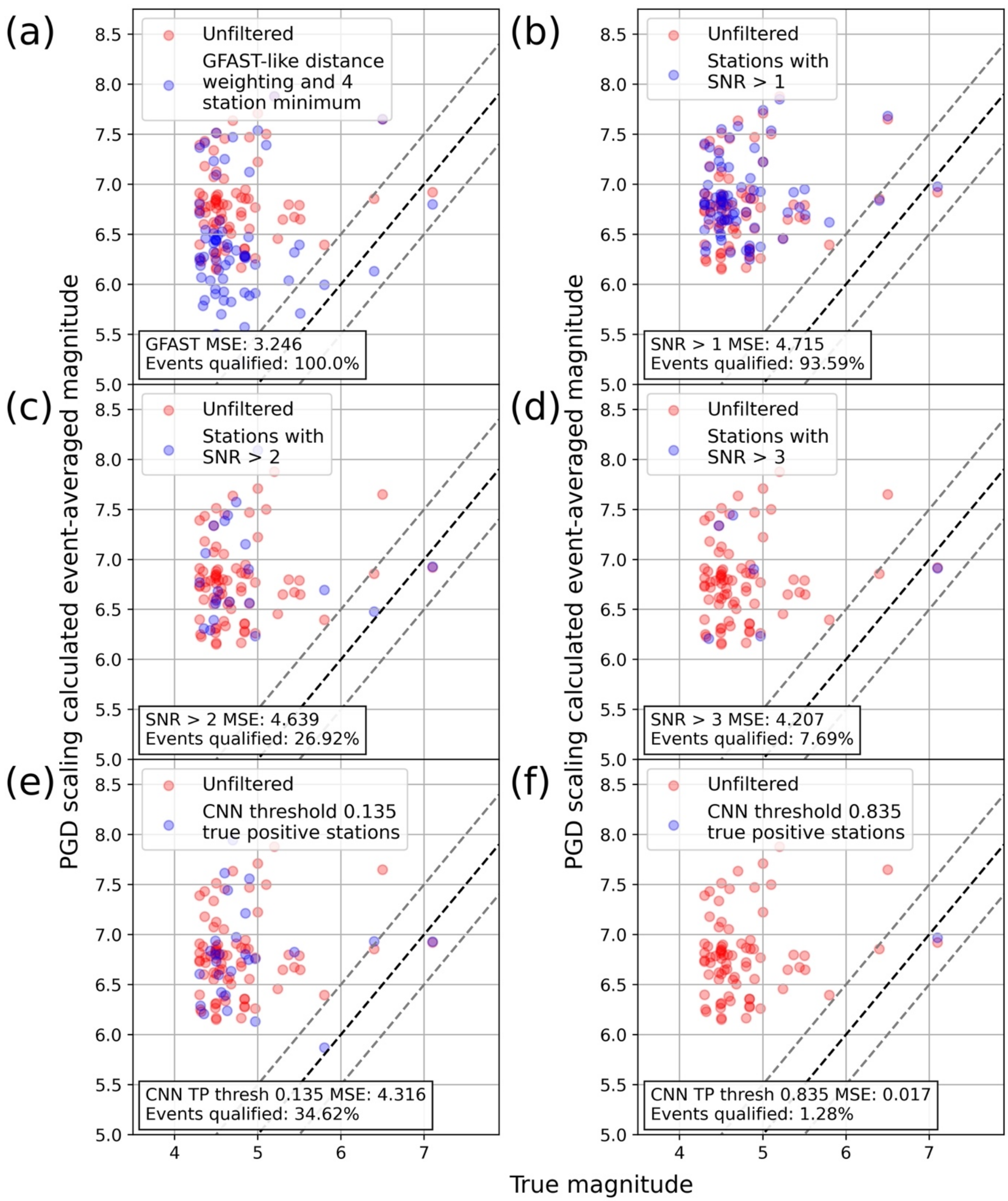
We also performed these tests with the real testing dataset even though the sample size of earthquakes was very limited (Figure 3.14). For the unfiltered real data, the MSE between the true magnitudes and event-averaged PGD magnitudes was 4.467, which is higher than the MSE for the unfiltered FakeQuakes data. For simplicity and ease of comparison, we used the same two thresholds of 0.135 and 0.835 for this test on the real data as we used for the FakeQuakes data.

Figure 3.14a shows the results for the GFAST-like test where the individual station PGD magnitudes were weighted by hypocentral distance in the event-averaged magnitude computation. As in Figure 3.13a, no events were excluded from the final plot due to their hypocentral distance, and all events had at least 4 stations included in the computation. Figure 3.14a shows more clearly the difference that the weighting makes, as the majority of the smaller earthquake magnitude estimates were drawn down toward the one-to-one line using the weighting scheme. However, the error was still relatively high with an MSE of 3.246, which was higher than the GFAST test MSE for the FakeQuakes data.

Figures 3.14b-d show the results when the same SNR filters were used. Figure 3.14b shows the results when only waveforms with SNR > 1 were used in the magnitude computation, which permitted 93.59% of events through the filter. Similar to the results shown in Figure 3.13b, this filter resulted in an even higher MSE than that of the unfiltered data at 4.715. Restricting the data further using a filter of SNR > 2 reduced the qualified events to 26.92%, but only marginally improved the MSE to 4.639 (Figure 3.14c), and restricting the data to SNR > 3 reduced the qualified events further to 7.69%, but still only improved the MSE to 4.207 (Figure 3.14c). As Figures 3.14c and 3.14d show, for the real data using SNR restrictions did not seem to leave only the higher magnitude earthquakes behind as observed in Figure 3.13c and 3.13d.

Figures 3.14e and 3.14f show the results of the tests when we allowed only the TP cases detected by the CNN to contribute to the event-averaged PGD magnitude. In Figure 3.14e where we used the threshold of 0.135, similar to the results shown in Figure 3.13e with the FakeQuakes data, many of the smaller magnitude earthquakes that have large differences between true and computed magnitude are still included, with 34.62% of events passing through the filter. The MSE is still high for this threshold at 4.316, which is better than the SNR > 1 and > 2 filters but worse than the SNR > 3 filter and the GFAST-like hypocentral distance weighting scheme.

Figure 3.14 (next page)**.** Plots of the event-averaged PGD magnitudes for the earthquakes in the real testing dataset using several different methods to restrict the dataset. Red points indicate event-averaged PGD magnitudes calculated using all of the data with no filters, and blue points indicate event-averaged PGD magnitudes calculated with a restriction on the waveforms permitted to contribute. The black dashed lines represent the one-to-one relationship between true magnitude and the event-averaged magnitude calculated from the PGD scaling law, with the gray dashed lines indicating half a magnitude unit above and below the one-to-one line. **(a)** Contributing station magnitudes computed using a GFAST-like exponential hypocentral distance weighting scheme and requiring a minimum of 4 stations per event. **(b)** Contributing stations limited to those with waveform SNRs > 1. **(c)** Contributing stations limited to those with waveform SNRs > 2. **(d)** Contributing stations limited to those with waveform SNRs > 3. **(e)** Only stations with "true positive" results as picked by the CNN developed in this study were allowed to contribute, using the accuracy-optimized confidence threshold of 0.135 from the FakeQuakes test. **(f)** Only stations with "true positive" results as picked by the CNN developed in this study were allowed to contribute, using the precision-optimized confidence threshold of 0.835 from the FakeQuakes test.



Increasing the threshold to 0.835 results in only one earthquake, the **M**7.1 Ridgecrest mainshock, passing through the filter (Figure 3.14f). Given the results shown in Figure 3.11d-f for the high confidence threshold where only that earthquake was successfully picked, this was not necessarily unexpected. Accordingly, the MSE for this event was 0.017, much lower than for the other filtering tests. The PGD magnitude computed for the M7.1 Ridgecrest mainshock using this 0.835 TP threshold was 6.97, but the computed magnitude for the unfiltered dataset was also 6.92 – well within the 0.5 magnitude unit lines on the plot. Using the CNN model provides a marginal improvement at best for this earthquake.

We also would have liked to see that the other two events >**M**6 in this dataset had their magnitude estimates improved by using the CNN filter with a threshold of 0.835 as was observed for the FakeQuakes data in Figure 3.13f, but the CNN failed to identify any stations from those events as TPs for this threshold. This is explainable when examining the data. The **M**6.5 in the dataset occurred in Nevada (Figure 3.1), and there were only 6 GNSS stations from which we were able to interpolate P-wave arrivals for that event. The closest of these stations to the hypocenter was over 400 km away, and the SNRs of the earthquake at those stations ranged from 0.8-1.7, which was lower than where we observed the model's best performance in this study. The **M**6.4 was the Ridgecrest foreshock, so there were 66 stations from which we had waveforms containing those interpolated P-wave arrival times and some were quite close to the hypocenter. However, of the 66 waveforms only two had SNRs >1.5 of 1.71 and 2.28, which were still quite low. Particularly for a higher threshold like 0.835, the model not having high enough confidence that a P-wave arrival existed in the sample with an SNR that low was not an unexpected result.

The results of these tests suggest that using a CNN like the one trained in this study to filter out waveforms in which earthquakes are not detected would improve the results of PGD magnitude estimations for the larger earthquakes for which GNSS data must be used to avoid magnitude saturation issues. This follows the results of the Lin et al. (2023) study which found that the performance of the M-LARGE algorithm that ascertained earthquake source properties such as magnitude, centroid location, and rupture extent was improved by reducing the amount of noise fed into the model. It should be noted, however, that the CNN model we presented in this study would require significant re-configuration and likely retraining to allow it to operate in real-time, and additional tests would be required to show that it still performed as expected. The scope of this study was also limited regionally and temporally to Southern California and a timespan around the 2019 Ridgecrest earthquake sequence, as well as to the real-time GNSS position solutions produced at that time by UNAVCO. Future work should include the training and testing of this type of model to other regions, particularly the rest of the West Coast where ShakeAlert and GFAST are active, as well as to other types of real-time position solutions such as those processed with the Fastlane software (Santillan et al., 2013; Melbourne et al., 2021) in real-time by Central Washington University.

Another potential future avenue for improvement would be the implementation of GNSS denoising into the workflow for computing earthquake magnitudes from GNSS data. Thomas et al. (2023) presented a deep learning model for this application similarly trained on synthetic data which showed promise when applied to real data as well. This type of model could potentially be used before a model trained to detect P-wave arrivals in GNSS data like the one presented in this study to improve the performance of such a model in terms of the binary classification results, as well as the actual position of the predicted P-wave arrivals. If successful, this combination may provide more hope for the development of a standalone GNSS EEW system where data is denoised in real-time, earthquakes are detected/low-quality data is filtered out, and then PGD magnitudes are computed.

3.5. Conclusions

In this study we presented the construction, training, and testing of a machine learning model for detecting earthquakes in noisy real-time GNSS data. We trained this model using a suite of synthetic strike-slip earthquakes combined with real real-time noise, and tested it on additional unseen synthetic data as well as real earthquake data from Southern California over approximately a year around the 2019 Ridgecrest earthquake sequence. When testing the performance of the trained model, we observed that it was generally difficult to pick an appropriate confidence threshold to maximize true positives and true negatives while also minimizing false positives and false negatives. The synthetic testing dataset was comprised of equal numbers of noise waveforms and earthquake-containing waveforms. Using a confidence threshold which optimized the accuracy of the model, we found that the model's accuracy was highest (>75%) at detecting earthquakes for events >**M**7 and waveforms with PGDs > 10 cm and SNRs above ~2. For the synthetic data with this threshold, the model's predicted P-wave arrival times fell within 5 seconds of the actual arrival time for 84% of the picks. In contrast, the real testing dataset was highly asymmetrical, with only 0.214% of the tested samples actually containing earthquakes. This made choosing an appropriate threshold even more difficult. Using the same threshold that maximized accuracy for the synthetic testing data, we observed that the model's capability for real earthquake detection also depended substantially on the PGD and SNR of the sample as well, and for the real testing data, only 38% of the predicted P-wave arrivals were within 5 seconds of the actual arrival time.

The number of false positive results observed in both the synthetic and real testing data was high when a confidence threshold low enough to detect the majority of the actual earthquakes in the data was used. Current EEW systems that incorporate GNSS data, such as ShakeAlert in the Western United States with the GFAST algorithm, require triggering by detections made with seismic data for the GNSS algorithm to begin analysis. Our study results suggest that at least using this particular model, it may not be realistic to expect a GNSS system to detect earthquakes and trigger itself alone due to the high noise levels in the data. However, we also explored how our model performed as a "filter" to remove waveforms from stations where an earthquake was not observed through the noise from incorporation into an algorithm like GFAST that calculates the magnitude of an earthquake from the PGDs of GNSS data. We tested the computation of event-averaged PGD magnitudes using all available data with no alterations, as well as a hypocentral distance weighting system like the one used in GFAST, where stations closest to the hypocenter are more strongly weighted in the magnitude computation. We then also tried filtering waveforms by their SNR, so that only high-SNR waveforms could contribute to the magnitude estimation, and also using this study's CNN model to filter stations by whether the model detected an earthquake in them or not, so that only the ones with positive detections were used in the magnitude computation. We found that for both the synthetic and real testing datasets, the mean squared error of the event-averaged magnitude estimates compared to the true magnitudes was lowest when the CNN model was used compared to using all of the unaltered data, the GFAST weighting method, or an SNR filter.

This suggests that using a CNN model like the one presented in this study to only allow GNSS waveforms where earthquakes are visible above the noise to contribute to the magnitude determination process could improve the accuracy of the magnitude estimates. In future work, this model should be adapted and tested for real-time performance, and data from additional regions, time periods, and GNSS position solution types should be incorporated to explore its applicability to a wider range of scenarios.

3.6. Data and Resources

Codes used to process and analyze data and make the figures for this study are available at https://github.com/UO-Geophysics/gnss-picker. The seismic data used for P-wave picking in this study were downloaded from the Incorporated Research Institutions for Seismology Data Management Center (IRIS-DMC) under network codes CI and PB. The real-time GNSS position solutions we used were provided by Kathleen Hodgkinson, formerly with UNAVCO (now EarthScope) and at the time of writing now with Sandia National Laboratories.

3.7. Acknowledgements

This material is based on services provided by the NSF GAGE Facility, operated by EarthScope Consortium, with support from the National Science Foundation, the National Aeronautics and Space Administration, and the U.S. Geological Survey under NSF Cooperative Agreement EAR-1724794.

We would like to acknowledge use of the Broadband Platform software provided by the Southern California Earthquake Center (http://scec.org) which is funded by NSF Cooperative Agreement EAR-1600087 and USGS Cooperative Agreement G17AC00047.

We make use of data from the networks CI (California Institute of Technology and United States Geological Survey Pasadena, 1926) and PB. We appreciate the hard work of the engineers and network operators who make this data available in near real-time.

3.8. Bridge

In this chapter, we detailed the development of a machine learning model which can detect earthquakes in noisy real-time GNSS data. This model can be used to improve the performance of GNSS-based earthquake early warning magnitude determination algorithms by filtering out noisy/low quality data which increases the error in the estimated magnitudes. In the next chapter, we will expand our focus from earthquake early warning systems alone to the rapid event publication process that the NEIC undertakes on a 24/7 basis. While early warning systems tend to be regional, the NEIC must analyze earthquakes across the globe and in many tectonic settings, making the process of determining earthquake magnitudes more complex. We will present another machine learning model in the following chapter, this time designed to allow for a uniform process for rapid magnitude computation worldwide.

3.9. Supplemental Figures

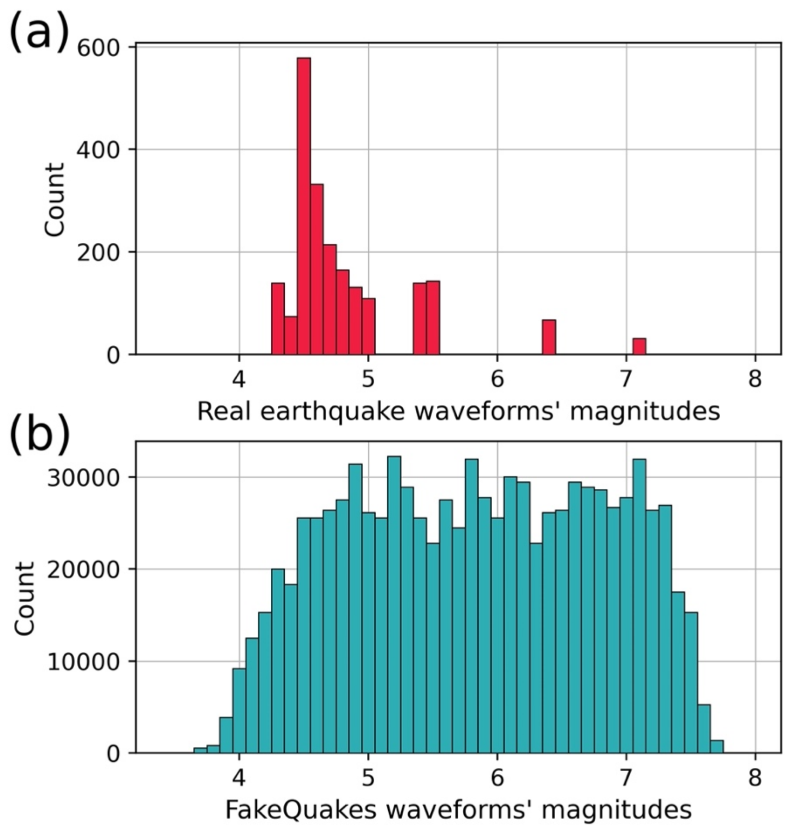


Figure 3.S1. Earthquake waveform magnitude distributions for **(a)** the real earthquake testing data and **(b)** the full set of FakeQuakes synthetic data.

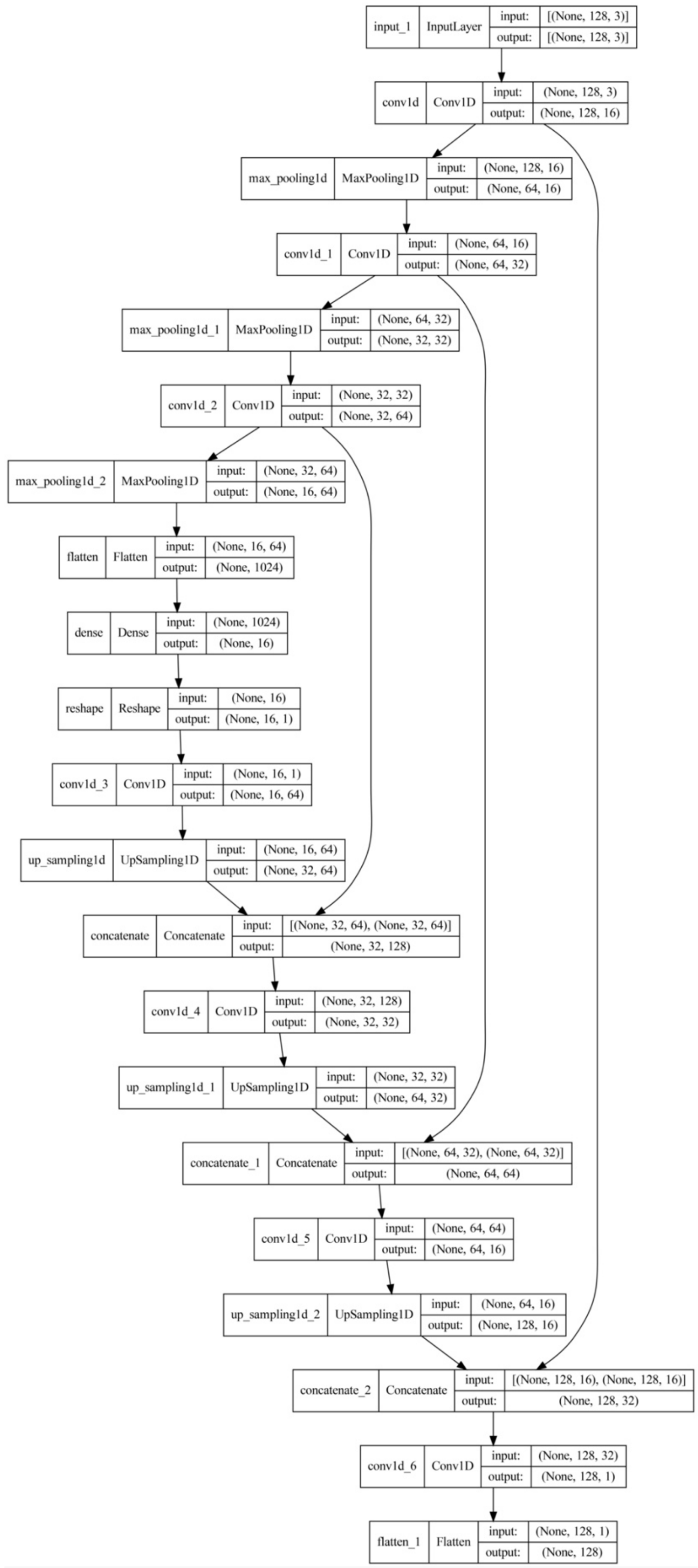


Figure 3.S2. Convolutional neural network (CNN) architecture used in this study.

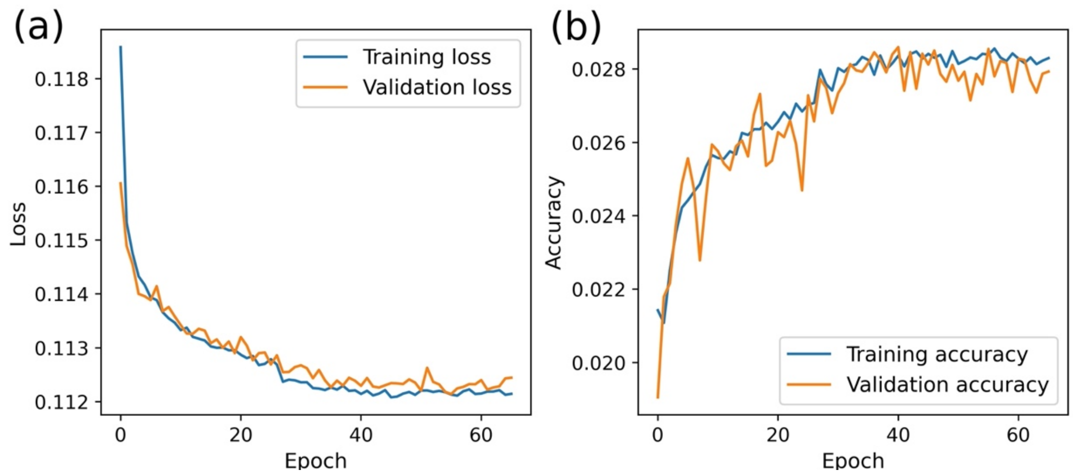


Figure 3.S3. Training history for the CNN model developed in this study. **(a)** Training (blue) and validation (orange) loss over time during training. **(b)** Training (blue) and validation (orange) accuracy over time during training.

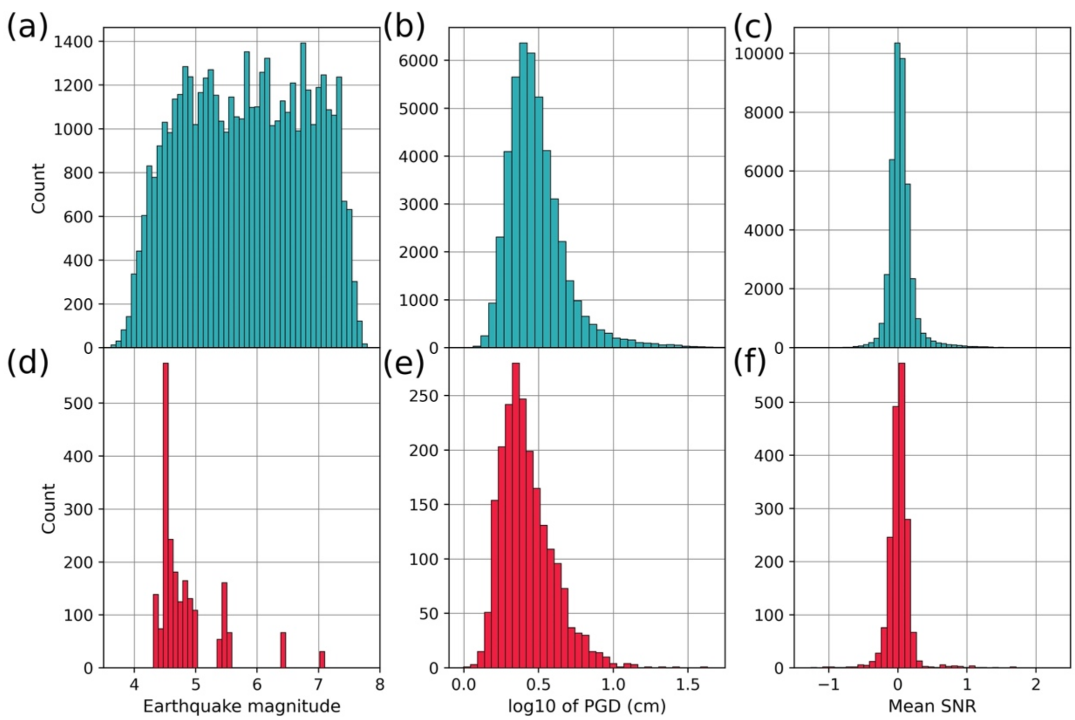


Figure 3.S4. Distribution of waveform characteristics in the two testing datasets. **(a)** Magnitude, **(b)** peak ground displacement (PGD) in centimeters, and **(c)** mean signal to noise ratio (SNR) distribution in the FakeQuakes testing dataset. **(d)** Magnitude, **(e)** PGD in cm, and **(f)** mean SNR distribution in the real data testing dataset.

CHAPTER IV

RAPID ESTIMATION OF SINGLE-STATION EARTHQUAKE MAGNITUDES WITH MACHINE LEARNING ON A GLOBAL SCALE

From Dybing, S. N., Yeck, W. L., Cole, H. M., & Melgar, D. (2024). Rapid estimation of single-station earthquake magnitudes with machine learning on a global scale. *Bulletin of the Seismological Society of America*, *114*(3), 1523–1538. https://doi.org/10.1785/0120230171

4.1. Introduction

Monitoring agencies worldwide ensure that earthquake information is shared in near real-time as quickly and as accurately as possible. Earthquake information products are key for organizing emergency response efforts after substantial shaking, as well as anticipating and modeling earthquake-associated hazards that can also have catastrophic effects on people and property. For example, the U.S. Tsunami Warning Centers monitor the globe for large earthquakes that could cause tsunamis, and the U.S. Geological Survey (USGS) analyzes and publishes information about earthquake economic and loss of life effects, as well as landslide and liquefaction risk. Additionally, national and international security groups (e.g., the Air Force Technical Applications Center in the United States) conduct global monitoring to detect explosive seismic sources that are not earthquakes, which may indicate violations of the Comprehensive Test Ban Treaty, which bans nuclear explosions worldwide (Ringler et al., 2022). Foundational to all of these missions is the ability to rapidly detect, locate, and characterize seismic sources.

In the United States, the USGS National Earthquake Information Center (NEIC) is tasked with 24/7 rapid response to earthquakes both nationally and across the globe. Automatic processing software and human analysts at the NEIC work to accurately assess the character of earthquakes within minutes to aid in estimates of economic and life loss, and provide information to emergency responders, the media, governments, and the public (Benz, 2017). Key earthquake characteristics include location, magnitude, and rupture extent and pattern, all of which can inform of the potential for associated hazards such as strong shaking and tsunamis. Of these, magnitude is a particularly important source parameter because it is the foundation of subsequent hazard estimates (e.g., shaking intensities) and is the fundamental parameter used by the public to interpret the severity of an earthquake. However, magnitude estimation can be a complicated process, and more simplistic methods may fail to adequately characterize an event. This results in variability of the magnitude based on how it is calculated, and it is prone to errors propagated from location and station metadata inaccuracies.

Magnitude estimation is particularly complex at the NEIC because it monitors earthquakes both domestically and globally in all tectonic environments. For small domestic earthquakes, the NEIC relies on amplitude-based measurements, preferring either M_L (local magnitude) or m_{b_lg} (short period surface wave magnitude) depending on the tectonic environment, or measurements derived from waveform modeling (M_{wr} – regional magnitude) when sufficient data are available. For larger global earthquakes, the NEIC relies on different amplitude-based magnitudes (primarily m_b – short period body wave magnitude) or magnitudes calculated with waveform modeling (primarily M_{ww} – moment W-phase magnitude or M_{wb} – body wave magnitude).

Each of these magnitude types is only appropriate for specific magnitude ranges, and the appropriateness of each can depend on the properties of the waveforms. As an example, while the NEIC uses M_{ww} as its preferred magnitude, M_{ww} can perform poorly after larger earthquakes due to increased long-period noise levels. Amplitude-based magnitudes are often the first magnitude produced internally in the NEIC's processing system, but these are prone to substantial errors because of mislocation due to their heavy reliance on attenuation relationships. Therefore, magnitude processing that can span all settings and is independent of location information has the potential to aid internal processing systems, as well as rapidly alert analysts of potential significant events.

4.1.1. Machine Learning for Earthquake Source Parameter Estimation

In recent years, research into applications of machine learning in seismology and geophysics has grown substantially, with several studies demonstrating that deep learning models improve upon traditional data processing methods. A review of the history of machine learning use in the seismology field and its current applications and future directions can be found in Mousavi and Beroza (2023). Deep learning typically requires large amounts of data to perform well, and the standard of maintaining publicly accessible data within the geophysics community (e.g., the NEIC's earthquake catalog, the EarthScope waveform data archives) lends itself well to this methodology. The utility of deep learning techniques has been established for many seismology applications, including detection and location of earthquakes and expansion of earthquake catalogs through accurate P- and S-phase picking for improved locations and velocity models (e.g., Ross et al., 2018; Perol et al., 2018; Zhu and Beroza, 2019; Mousavi et al., 2020; Zhang et al., 2020; Mousavi and Beroza, 2020a).

Another application, the estimation of earthquake magnitudes, has been approached with a variety of deep learning techniques. As reviewed by Mousavi and Beroza (2023), these include seismic single-station classification and regression approaches (Lomax et al., 2019; Mousavi and Beroza, 2020b; Ristea and Radoi, 2021), as well as multi-station methods, which in addition to seismic data (van den Ende and Ampuero, 2020; Münchmeyer et al., 2021) can include the use of Global Navigation Satellite Systems (GNSS) (Lin et al., 2021) and elastogravity data (Licciardi et al., 2022). Mousavi and Beroza (2020b) demonstrated that a regression model built with the machine learning framework TensorFlow (Abadi et al., 2015) as a combination of a convolutional neural network and a recurrent neural network could accurately estimate local magnitudes from single waveforms with no instrument response removal. The Mousavi and Beroza (2020b) model, known as MagNet, was trained and tested on data from global earthquakes (primarily of <M5) from the STanford EArthquake Dataset (STEAD) (Mousavi et al., 2019) with epicentral distances of less than one degree. The MagNet model shows great promise, but its applicability as currently trained is limited to small scales and therefore is not appropriate for global earthquake monitoring.

Cole et al. (2023) constructed the Machine Learning Asset Aggregation of the Preliminary Determination of Epicenters (MLAAPDE, pronounced "millipede") to aid in the development of machine learning tools at the NEIC. MLAAPDE is a dataset and query tool built from the Preliminary Determination of Epicenters (PDE) bulletin published by the NEIC (Sipkin et al., 2000). MLAAPDE is formatted such that users will readily be able to use the >5.1 million waveforms and associated labels in model development, training, and testing. As such, it is this dataset that we utilize for this study. Detailed information about MLAAPDE and how to access the code and default dataset is available in Cole et al. (2023). In this study, we leverage the previous work on the MagNet model with the MLAAPDE dataset to test the applicability of the approach to global earthquake monitoring.

4.2. Methods

4.2.1. MLAAPDE Data Access

The default MLAAPDE catalog contains 120-second-long waveforms, centered on an analyst-verified phase pick and in units of raw counts with a 40-Hz sampling rate. Our goal was to create a model that can be applied in rapid response settings, so we chose to only sample from the available first arrival P-phase picks in the catalog (either P, Pn, or Pg). We stream-normalized input waveforms by dividing each observation by the maximum value of all three channel traces for each station. We did not apply instrument response correction or otherwise filter the waveform data. However, the waveforms were decimated to a sampling rate of 20 Hz to reduce computation time and memory required for the code to run.

A common practice in machine learning is to split datasets into training, validation, and testing partitions with a convenient ratio (e.g., 80% samples for training, 20% for testing). Our training dataset was composed of earthquakes with origin times spanning five years and two months from August 1, 2013, to September 30, 2018, and earthquake magnitudes between M0.0 and M8.3, containing 2,431,341 waveforms from 2,121 stations in the MLAAPDE dataset. The validation dataset was composed of earthquakes with origin times spanning one year and three months from October 1, 2018, to December 31, 2019, and earthquake magnitudes between M0.7 and M8.0, containing 489,268 waveforms from 1,615 stations in the MLAAPDE dataset. The testing dataset was composed primarily of 324,365 MLAAPDE waveforms from 1,536 stations, with earthquake origin times spanning one year from January 1, 2020, to December 31, 2020, and earthquake magnitudes between M0.7 and M7.8. The full dataset contains a variety of different magnitude types, but the majority are either m_b, M_L, or M_{ww} (refer to Cole et al., 2023, their Figure 2, for the full list and distribution of magnitude types in MLAAPDE). We ensured that no earthquakes in the training dataset were in the validation or testing datasets to reduce leakage and the likelihood that the model would learn the earthquake magnitudes from the waveform similarity of a single event. The separation between validation and testing datasets was made to ensure that the final model would not be biased by the testing process, as the best model is chosen in training based on the lowest validation loss achieved.

We also manually expanded the testing dataset beyond MLAAPDE in order to test the model on larger earthquakes, even though it was not trained on these events. MLAAPDE currently only draws from the most modern version of the PDE, which has been used since mid-2013. As a result, the training dataset used in this study does not contain any earthquakes above M8.3, as this was the largest event that occurred between August 1, 2013, and September 30, 2018. To supplement the testing dataset, we used Obspy (Beyreuther et al., 2010) to download 13,151 waveforms from the EarthScope Data Management Center (DMC) from an additional 78 earthquakes >M7.5 from January 1, 2000, to July 31, 2013. This dataset includes several great earthquakes, such as the 2004 M9.1 Sumatra earthquake, 2010 M8.8 Maule earthquake, and 2011 M9.1 Tōhoku earthquake. Waveforms for the earthquakes in this extended testing dataset came from 1,736 stations.

In total, the full dataset was split to be approximately 75% training, 15% validation, and 10% testing (Figure 4.S1, found in section 4.7 (Supplemental Figures)). Maps of the earthquakes and stations from which waveforms were drawn for this study can be found in Figure 4.1.

4.2.2. Training Data Processing and Augmentation

The speed at which an earthquake magnitude can be determined in a real-time environment depends on the arrival-time and telemetry delay of waveform data, the latency to load the waveform into monitoring applications, the window of waveform data required, and the speed of the magnitude calculation. For this reason, it is important to find the minimum effective window for magnitude estimation. We tested our model across a range of 1 second to 114 seconds of waveform data. MLAAPDE stores waveforms at a length of 120 seconds (Figure 4.2a), which we shortened further in our data augmentation process (Figures 4.2b-f).

Because MLAAPDE is a well-curated catalog of human-reviewed picks, we manually augmented the data to simulate the challenges of real-time processing, as we wanted our model to be resilient to issues that can manifest in seismic monitoring. Data augmentation involves synthetically expanding the number of samples in a dataset, as well as their variety, to prevent overfitting in machine learning models and improve generalization (Zhu et al., 2020).

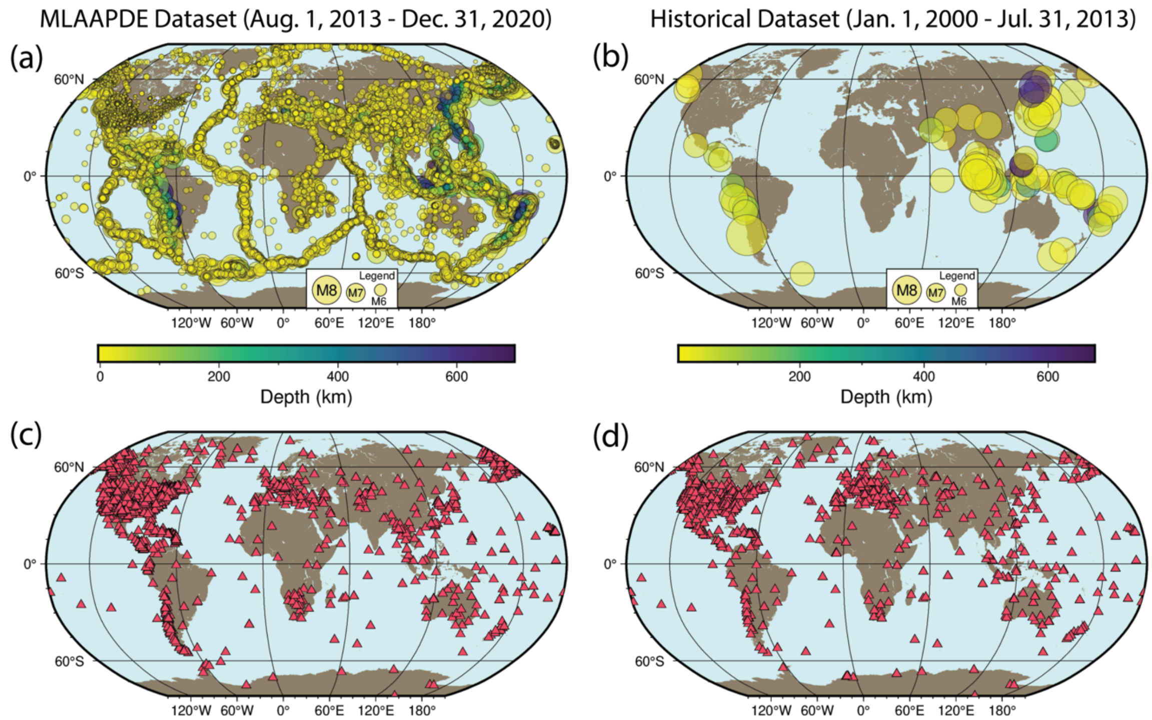
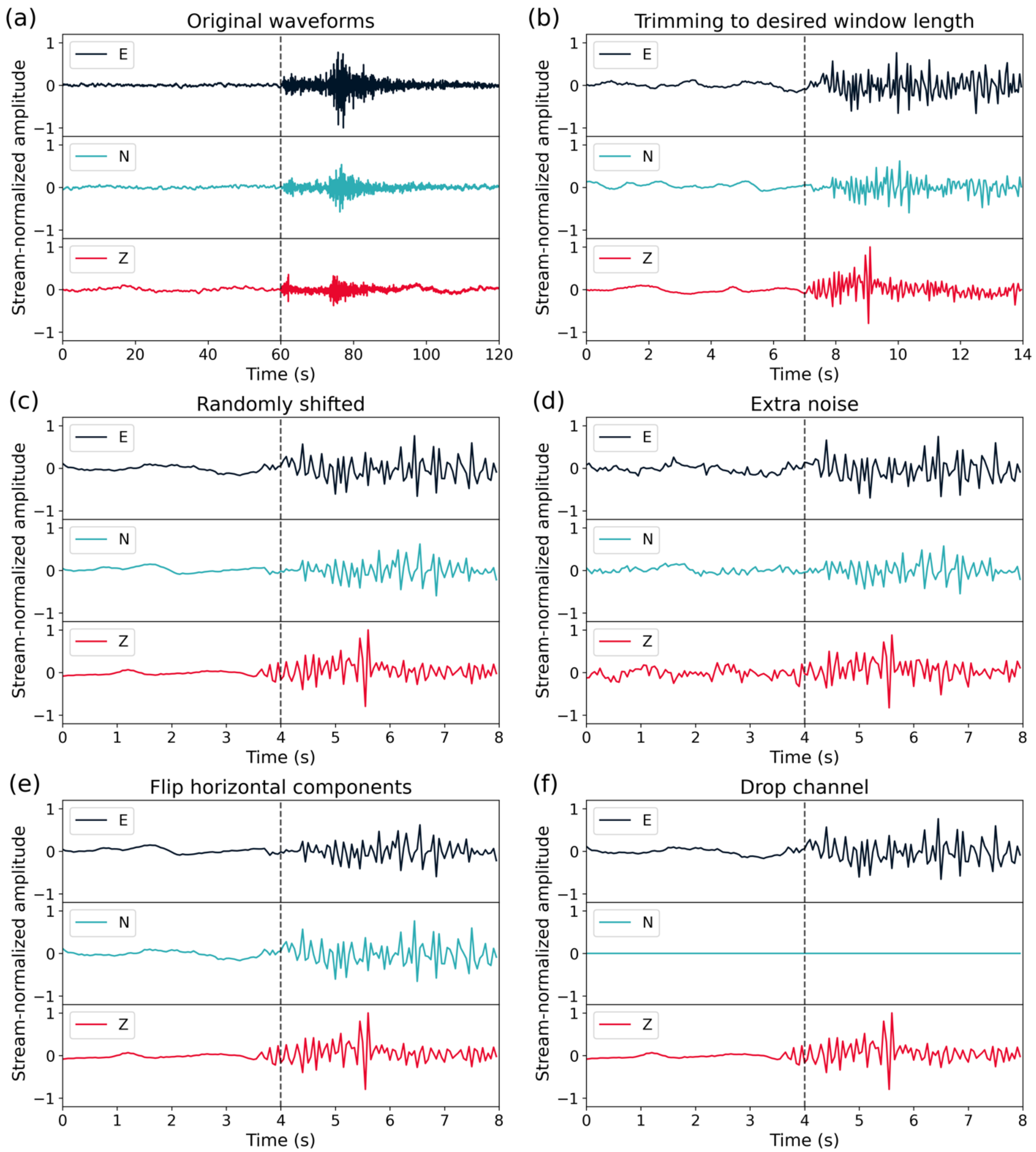


Figure 4.1. Maps of the earthquakes and seismometers that recorded them used in this study. **(a)** The 189,475 earthquakes in the combined training, validation, and testing datasets from the MLAAPDE catalog with origin times from August 1, 2013, to December 31, 2020. **(b)** The 78 earthquakes >M7.5 in the expanded historical testing dataset with origin times from January 1, 2000, to July 31, 2013. **(c)** The 2422 stations from which waveforms were drawn for the MLAAPDE earthquake catalog shown in panel (a). (**d**) The 1,736 stations from which waveforms were drawn for the historical earthquake catalog shown in panel (b). These maps were constructed using PyGMT (Uieda et al., 2023).

Figure 4.2 (next page)**.** Timeseries plots demonstrating the different types of augmentation implemented in our data generator. **(a)** The original three-component 120-second waveforms loaded from the MLAAPDE dataset, stream-normalized with no instrument response removed. The black trace indicates the east-west component, the blue trace the north-south component, and the red trace the vertical component. The phase pick is in the center of the figure at 60 seconds. The center of the figure is indicated by the black dashed line. **(b)** The same waveforms as in panel (a), cut down to a smaller window length (14 seconds in this example). **(c)** The same waveforms as in panel (b), but the window length has been cut down an additional 6 seconds to accommodate an applied random shift of the location of the phase pick by up to 3 seconds forward or backward. The center of the figure (which is no longer aligned with the phase pick) is still indicated by the black dashed line. **(d)** The same waveforms as in panel (c) with additional noise applied. For each point in the waveform, a value drawn from a normal distribution was added, with the standard deviation of each of those normal distributions being drawn from a uniform distribution between 0.01 and 0.15. The amount of added noise is not consistent between components. **(e)** The same waveforms as in panel (c), but the east-west and north-south components (black and blue) have been switched with each other. **(f)** The same waveforms as in panel (c), but in this instance the north-south (blue) component has been "dropped" by zeroing-out the trace.



We built our data augmentation using a TensorFlow Keras Sequence data generator (Abadi et al., 2015) that prepared the arrays to be fed into the model for training. By implementing the augmentation "on the fly," the data were shuffled and augmented differently with every training epoch, therefore increasing variability.

We augmented the data from MLAAPDE in four different ways following methods tested in Zhu et al. (2020). We first implemented a random shift in the location of the phase pick so that it was no longer in the exact center of each waveform. The phase picks in MLAAPDE are all human-reviewed, and in many cases, repicked for accuracy. In real-time monitoring settings, phase picks are performed automatically, often using short-term average (STA) over long-term average (LTA) methods (Lee and Stewart, 1981). Automatic picks are often not as accurate as manual picks from analysts. Therefore, to allow our model to learn to estimate earthquake magnitudes on real-time data that may not have perfectly accurate phase arrival times, we implemented a random shift of up to 3 seconds (forward or backward) into each waveform from MLAAPDE (Figure 4.2c), which further shortened the window length by 6 seconds. This random shift was applied to all waveforms in the training, validation, and testing datasets.

Next, noise was added to some of the waveforms only in the training dataset. As the data generator looped through the waveforms, approximately 50 percent of the time, additional noise was introduced. For each point in the waveform, a value drawn from a normal distribution was added, with the standard deviation of each of those normal distributions being drawn from a uniform distribution between 0.01 and 0.15 (Figure 4.2d). Many of the waveforms in the MLAAPDE dataset already have low signal to noise ratios, so this augmentation simply increased the number of "different" waveforms that the model could learn about in training.

The third augmentation involved swapping the two horizontal components approximately 50 percent of the time in the training data (Figure 4.2e). This simulated potential station orientation issues and changed the source-station path, again introducing more data variety.

The fourth and final augmentation dropped each channel approximately 5 percent of the time, with a limitation to prevent all three from being dropped simultaneously (in this case, the third or vertical channel was kept). An example of this is shown in Figure 4.2f. In a real-time setting, it is very possible that some instruments or channels may not be recording or transmitting data due to malfunctions or telemetry problems. Introducing dropped channels into the training data allowed for the machine learning model to be resilient to these issues when making magnitude estimations.

In the training process, we used the same Adam optimizer (Kingma and Ba, 2015) as Mousavi and Beroza (2020b) as well as a mean squared error loss function, and we utilized the same early stopping and learning rate reduction to prevent overfitting and stop the training when the loss rate plateaued. Training was repeated for each different window length of data, resulting in a total of 26 trained models.

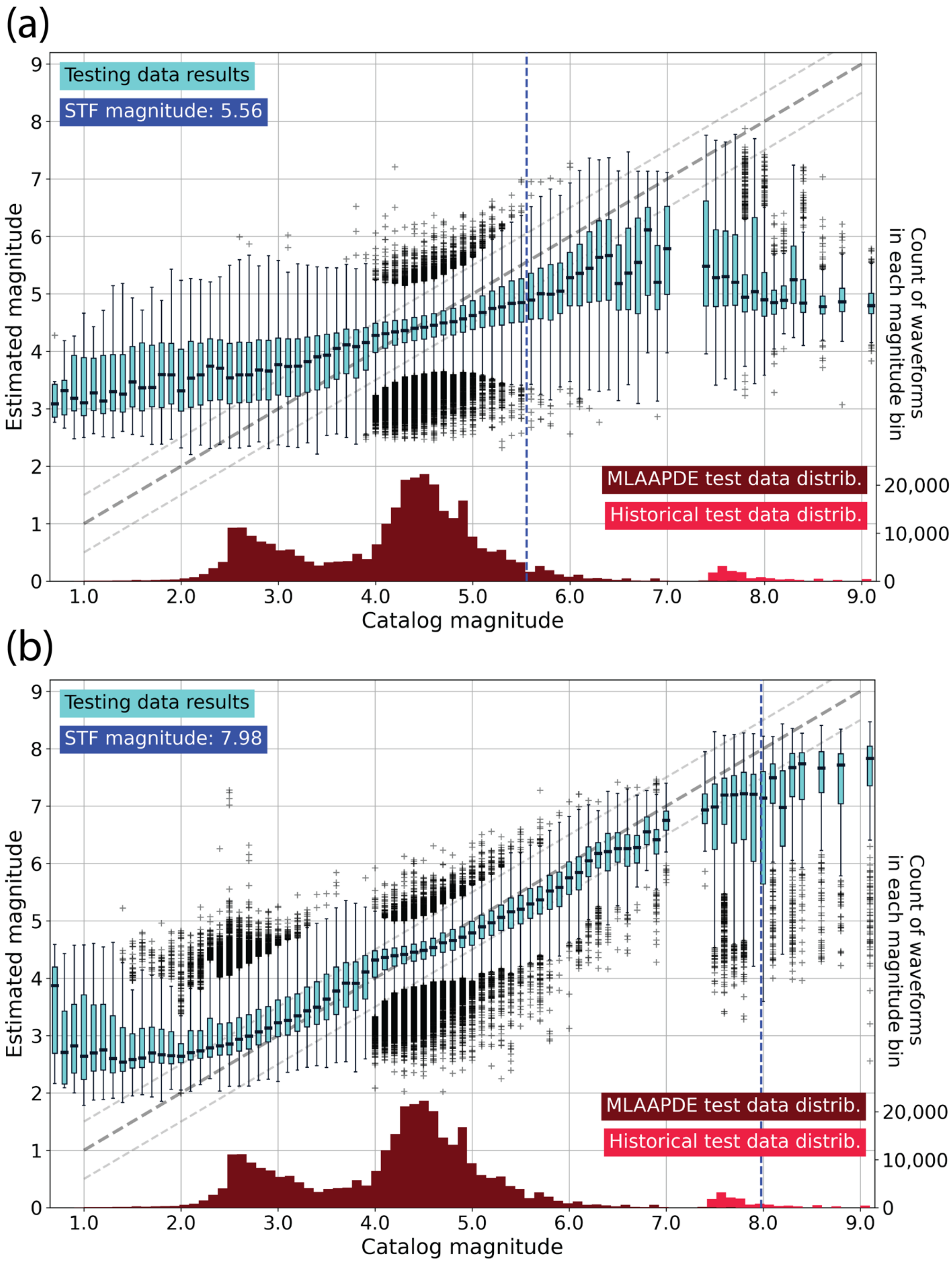
4.3. Results and Discussion

4.3.1. Model Performance by Magnitude

After training, our machine learning model (which will be referred to as AIMag – “Artificial Intelligence Magnitude”) was used to make estimations on the previously unseen testing dataset. The testing process was, like the training, repeated with each of the datasets of different window length for a total of 26 iterations.

Because most of the earthquakes in the dataset have observations from multiple stations, and AIMag makes magnitude estimations for individual stations, there are multiple estimated magnitudes for each earthquake in the catalog, and the statistics for these estimations can be more easily visualized using a boxplot plotting method. Results shown as boxplots are included for two selected iterations in Figure 4.3, where Figure 4.3a shows the results for the data with a shorter window length of 7 seconds, and Figure 4.3b shows the results for the longest window length of 114 seconds.

Figure 4.3 (next page)**.** Boxplots plotted using Matplotlib (Hunter, 2007) showing the model’s estimated earthquake magnitudes for the testing dataset (both MLAAPDE and historical data) for two example time windows, with the catalog magnitude for the earthquakes on the x-axis, and the model’s estimated magnitudes for those same earthquakes on the left y-axis. An animation showing the boxplots for all time windows can be found in the supplemental files included with this dissertation as Video 4.S1. **(a)** Model magnitude estimations for data with a waveform window length of 7 seconds. **(b)** Model magnitude estimations for data with the longest tested waveform window length of 114 seconds. In each of these panels, the median magnitude estimation for each catalog magnitude bin on the x-axis is shown by the horizontal black line in the center of each light blue box. The lower and upper bounds of the light blue boxes indicate the first and third quartiles of the estimations in each bin, respectively, and the whiskers extend to 1.5 times the inter-quartile range. The individual points beyond the whiskers indicate outliers. The diagonal bold dashed gray line represents a one-to-one magnitude estimation line, with the lighter gray dashed lines above and below it +/- half a magnitude unit. The vertical blue dashed line indicates the magnitude of a theoretical earthquake whose source-time function (STF) would entirely fit within half of the window length for that particular plot, as the P-wave should arrive at approximately halfway through the window. Finally, the red histogram at the bottom of each panel shows the magnitude distribution for the testing dataset, with the testing data from the MLAAPDE dataset in the year 2020 in dark red, and the historical testing dataset in light red. A scale for the counts for each of the histogram bins is shown on the right y-axis.



(a)
Testing data results
STF magnitude: 5.56
MLAAPDE test data distrib.
Historical test data distrib.
Estimated magnitude
Count of waveforms in each magnitude bin
Catalog magnitude
(b)
Testing data results
STF magnitude: 7.98
MLAAPDE test data distrib.
Historical test data distrib.
Estimated magnitude
Count of waveforms in each magnitude bin
Catalog magnitude

The full set of figures for all window lengths can be found in the supplemental files included with this dissertation as Video 4.S1, along with all other videos that will be referenced in this chapter. As with the default waveforms, the phase arrivals are approximately in the center of these waveforms with the random shift, so approximately half of the window length of these waveforms are noise and half are signal.

By examining the changes between Figures 4.3a and 4.3b, it is possible to observe that as the window length increases, the median estimated magnitudes and their corresponding blue boxes converge towards a one-to-one relationship with the catalog magnitudes, as indicated by the diagonal bold gray dashed line. The lighter gray dashed lines above and below the bold gray line show half a magnitude above and below the one-to-one relationship.

At the longest window length of 114 seconds as shown in Figure 4.3b, the smallest earthquake magnitudes are overestimated by AIMag, and the largest earthquake magnitudes are underestimated. Some of this effect is likely attributable to limited data availability at the lowest and highest magnitudes, as shown in the histograms in each panel of Figure 4.3. The most populated magnitude bins in the catalog are around M2.5 and M4.5. This is because the NEIC's PDE catalog publication criteria include earthquakes M2.5 and larger or felt in the United States (excluding California, which is published for M3.0 and larger), and M4.0 and larger or felt in the rest of the world, resulting in a catalog that is complete to about M4.5 globally. Therefore, fewer earthquakes <M2.5 are in the catalog, and naturally fewer high magnitude earthquakes are included because of the Gutenberg-Richter law (Gutenberg and Richter, 1949). This data paucity in training likely affects the model's ability to accurately assess the magnitudes of the smallest and largest earthquakes. Where data are sparse, the model tends to hedge the magnitude estimates to be more toward the mean of the magnitudes in the dataset (4.4 for the training dataset, 4.2 for the validation dataset, and 4.3 for the testing dataset). For example, at the 1-second time window (Video 4.S1) the median estimated magnitudes are all close to 4.3. As the time window grows, the estimates shift to fall more along the one-to-one line, but they never move low enough for the magnitudes in the lowest ranges – likely due to the undersampling.

Another feature of these plots is the vertical darker blue dashed line, which moves to the right as the window length increases from Figure 4.3a to 4.3b. This line indicates the magnitude of a theoretical earthquake for which the source-time function would entirely fit within half of the window length for that particular plot (i.e., the length of earthquake signal in the window). This theoretical magnitude was calculated by dividing the window length by two and using that as the total rupture duration (τ_t) in the following equation:

$$M_0 = \tau_t^3 \times 0.625 \times 10^{23}, \tag{4.1}$$

where M_0 gives the seismic moment in dyne-cm and τ_t is the total rupture duration. Singh et al. (2000) gave a range of 0.25×10^{23} to 1.0×10^{23} dyne-cm/s^3 for the value of M_0/τ_t^3 for shallow earthquakes, so the midpoint of this range was used for the approximation of earthquake moment in this study. Using that calculated M_0, we then converted this to the moment magnitude M_w following Kanamori (1977) using the following equation:

$$M_w = \frac{2}{3}\log_{10}(M_0) - 10.73, \tag{4.2}$$

where M_0 again gives the seismic moment in dyne-cm. This M_w is what was plotted as the vertical dark blue dashed line in Figure 4.3. We expected that AIMag would be more accurately able to estimate the magnitude of an earthquake that can be entirely seen within the window of data provided (i.e., for which the magnitude bin is to the left of the blue dashed line), and that it would struggle with earthquakes larger than this (where the magnitude bin is to the right of the line). This effect can be seen most clearly in Video 4.S1, in which all tested time windows are shown.

At the longest time window of 114 seconds (Figure 4.3b), the theoretical earthquake magnitude is ~8, and it is true that the earthquakes above this magnitude are underestimated by AIMag. However, some of the larger earthquakes still to the left of the line are also underestimated. In this higher magnitude range, the model is likely limited by the small number of waveforms from large magnitude earthquakes in the training dataset. The largest earthquake in the training dataset is M8.3, and the entire training dataset only contains 31 earthquakes M7.5 and greater for a total of 4,486 waveforms. Compared to the 170,553 training waveforms available for just M4.5 alone, for example, this is a very limited sample from which AIMag can learn. Future work could explore the possibility of including synthetic large magnitude earthquakes in the training dataset to improve the model's exposure to these larger events during training. This could be achieved using codes such as FakeQuakes (Melgar et al., 2016), which is capable of generating synthetic teleseismic data. Another potential avenue for increasing the proportion of large earthquakes in the dataset would be to upsample the existing ones, as was done in Münchmeyer et al. (2021). We also explored the possibility of improving the performance of the models at the extreme ends of the magnitude scale by implementing weighting into the training based on the relative abundance of samples in the dataset by magnitude. We split the dataset into bins by 0.5 magnitude intervals and defined the weight for each bin as 1 – (samples in bin divided by the total number of samples in the training dataset). This way, a magnitude bin with many samples would have a small weight in the training process, and a bin with few samples would have a large weight. These weights were then applied to the waveform samples as they were processed through the data generator in the training process. We found slight improvements for some time windows with this method (Video 4.S2; Figure 4.S2), which indicates that for future work exploring a more sophisticated weighting scheme may lead to improved results.

We included the historical catalog of larger earthquakes in the testing dataset to determine whether AIMag was capable of accurately estimating their magnitudes, despite not being trained on events >M8.3. Although the magnitudes are underestimated, the model can at least determine that they are greater than M7 in many cases at the longest window length of data, shown by the positions of the medians for those magnitude bins in Figure 4.3b. This result makes AIMag useful for flagging large magnitude earthquake detections in monitoring operations.

Another potential reason that the highest magnitude earthquakes are underestimated by AIMag is because of the complications of magnitude saturation in seismic data. Magnitudes calculated using the P-waves in seismic waveforms tend to saturate at the high end of the scale (i.e., large earthquakes become indistinguishable from even larger ones). The relationship between log peak ground displacement and magnitude is linear for moderate magnitude earthquakes, but transitions to a scaling that is independent of magnitude at larger magnitudes (with this transition point moving to higher magnitudes as the data window length increases) (refer to Trugman et al, 2019, their Figure 4). Additionally, traditional seismic data processing requires high pass filtering to remove baseline drift effects that appear when data are double integrated to displacement, and this filtering removes low frequency content that may change the relationship between measured displacement and earthquake magnitude (Colombelli et al., 2012). In this study, although we do not remove the instrument response, integrate, or filter any data before training AIMag on it, by virtue of the diversity of the NEIC's contributor networks we are using a variety of different types of instruments, some of which likely have a smaller range of periods at which they have a flat response. This means that some longer period information may be suppressed, which could contribute to the model's underestimations at high magnitudes. In future studies, performance could potentially be improved for large earthquakes by the use or inclusion of displacement data as an input in the training process. Machine learning methods have previously been shown to be effective for real-time earthquake magnitude estimation using GNSS displacement data (e.g., Lin et al., 2021).

Although this paper focuses on AIMag's single-station magnitude estimation results, in operational practice, magnitudes are typically determined and reported by averaging the magnitudes calculated from many different stations. We calculated event-averaged magnitudes from the single-station results described previously (Video 4.S3) and observed that this approach had the expected effect of suppressing many of the outlier single-station magnitudes. This effect could also be seen when the standard deviations of the magnitude estimation errors for the event-averaged magnitude method were compared to the standard deviations of the errors for the single-station method (Figure 4.S3), as the standard deviations for the event-averaged magnitude method were lower.

4.3.2. Model Performance by Waveform and Earthquake Characteristics

We were interested in determining whether AIMag had any biases toward particular characteristics of earthquakes or waveforms. We analyzed the performance of the model by earthquake source depth, hypocentral distance (the distance between the hypocenter and the receiver station), and the signal to noise ratio (SNR) of the waveforms. For Figures 4.4, 4.5, and 4.6 as discussed in the following section, we calculated the magnitude estimation error by subtracting the model's estimated magnitude from the true catalog magnitude.

In Figure 4.4, we plotted the magnitude estimation errors by the source depth of the earthquakes in the testing dataset. Only the plot for the 114-second window is shown, as the results do not change substantially across time windows for this parameter. The remaining plots can be seen in an animation in Video 4.S4. For this figure we have also plotted a histogram showing the distribution of earthquake source depths corresponding to the waveforms in the testing dataset. The very populous second bin is as such because it encompasses the 10-kilometer depth range, and 10 kilometers is the default depth assigned by the NEIC when a shallow earthquake's depth cannot be precisely constrained. In general, Figure 4.4 shows that AIMag is not biased in its estimations by the depth of the source earthquake, with the median error for most catalog magnitude bins being near zero. It only shows more variability in error at the far end of the depth range, as very few 650 kilometer and greater deep earthquakes are available either for training or testing.

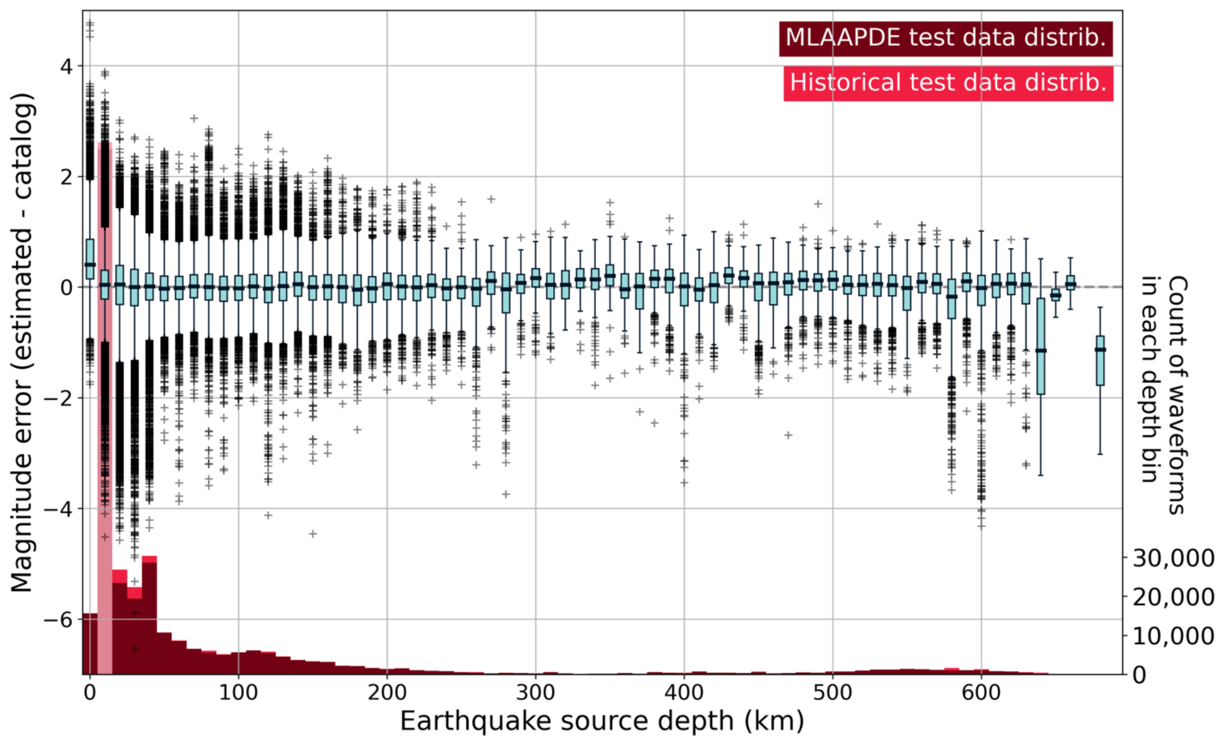
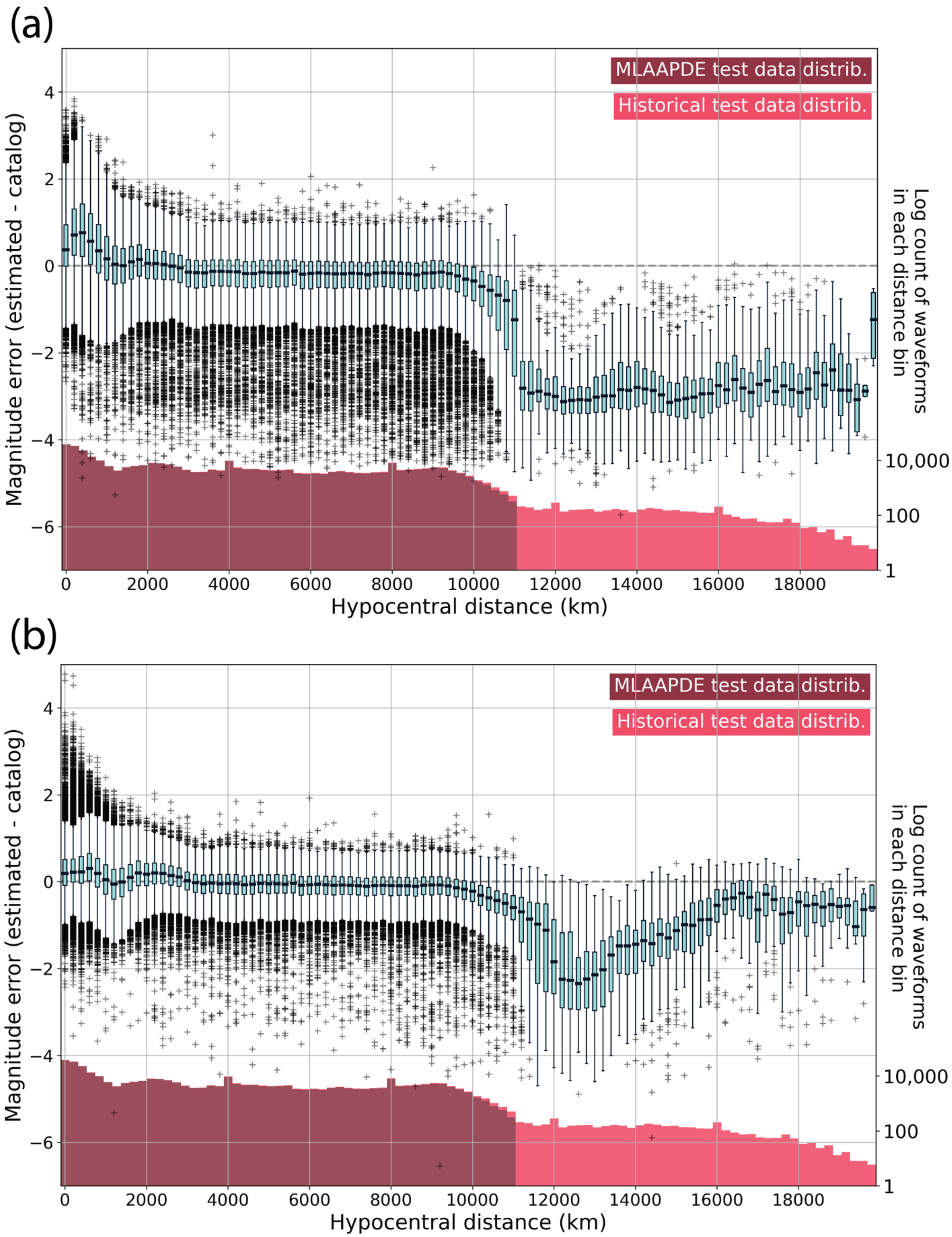


Figure 4.4. A boxplot showing the error in the magnitude estimation (estimated magnitude – catalog magnitude) made by the model on the left y-axis plotted against the source depth of the earthquake in kilometers on the x-axis, corresponding to each estimation for the 114-second time window. An animation showing the boxplots for all time windows can be found in the supplemental files included with this dissertation as Video 4.S4. The characteristics of the boxplot in this figure are the same as those for Figure 4.3, with the median magnitude estimation error for each earthquake source depth bin shown by the horizontal black line in the center of each light blue box. On the right y-axis, we have plotted a histogram showing the distribution of earthquake source depths corresponding to the waveforms in the testing dataset. The distribution of source depths in the MLAAPDE dataset is shown in dark red, with the distribution of source depths in the historical dataset stacked on top of those in light red. The transparency of the second (10 km) source depth histogram bin was reduced to ease visualization of the overlaid boxplot data.

We next looked at the magnitude estimation error by the hypocentral distance for each waveform. These results can be found in Figure 4.5, in which boxplots for the 7-second window (Figure 4.5a) and 114-second window (Figure 4.5b) are shown. The remaining plots for all other time windows can be seen in the animation in Video 4.S5. We have also plotted in these panels a logarithmic histogram showing the distribution of hypocentral distances corresponding to the waveforms in the testing dataset. The MLAAPDE dataset only contains waveforms with maximum hypocentral distances of ~100 degrees (or ~11,110 kilometers), so the only waveforms with further hypocentral distances come from the historical dataset. This is why we elected to use a logarithmic histogram for this plot – the relatively smaller size of the historical dataset made the histogram bars nearly disappear with a non-logarithmic plot.

In comparing Figure 4.5a to 4.5b across the time window range, we observe that the main difference in performance is with the far hypocentral distances greater than 10,000 kilometers where only the historical data exist. In Figure 4.5a with a 7-second window, AIMag greatly underestimates the magnitudes of the earthquakes in the waveforms with large hypocentral distances, and in Figure 4.5b at the 114-second window length, the performance is improved (although still underestimated). We would expect AIMag to perform poorly at these far distances because the training dataset includes no waveforms with hypocentral distances over ~11,110 kilometers, and because only a small portion of the P-wave will exist in waveforms with such large hypocentral distances. Of note, however, the model with a longer window length of data was able to estimate the magnitudes more closely, and it performs better at the very longest distance range of 16,000 kilometers and greater compared to a more intermediate range of 11,000-16,000 kilometers. From Figure 4.5 we can generally conclude that AIMag, when tested with data with a similar hypocentral distance range to that on which it was trained, is not substantially biased by hypocentral distance. We therefore recommend using this model as currently trained only for hypocentral distances of ~100 degrees or less.

Figure 4.5 (next page)**.** Boxplots for **(a)** the 7-second window and **(b)** the 114-second window showing the error in the magnitude estimation (estimated magnitude – catalog magnitude) made by the model on the left y-axis plotted against the hypocentral distance in kilometers on the x-axis corresponding to each estimation. An animation showing the boxplots for all time windows can be found in the supplemental files included with this dissertation as Video 4.S5. The characteristics of the boxplot in this figure are the same as those for Figures 4.3 and 4.4, with the median magnitude estimation error for each hypocentral distance bin shown by the horizontal black line in the center of each light blue box. On the right y-axis, we have plotted a logarithmic histogram showing the distribution of hypocentral distances corresponding to the waveforms in the testing dataset. The distribution of hypocentral distances in the MLAAPDE dataset is shown in dark red, with the distribution of hypocentral distances in the historical dataset stacked on top of those in light red. The MLAAPDE dataset in the format used in this study only contains waveforms with hypocentral distances of ~100 degrees or less, so the only waveforms with further hypocentral distances come from the historical dataset.



(a)
MLAAPDE test data distrib.
Historical test data distrib.
Magnitude error (estimated - catalog)
Log count of waveforms in each distance bin
Hypocentral distance (km)
4
2
0
−2
−4
−6
10,000
100
1
0
2000
4000
6000
8000
10000
12000
14000
16000
18000
(b)
MLAAPDE test data distrib.
Historical test data distrib.
Magnitude error (estimated - catalog)
Log count of waveforms in each distance bin
Hypocentral distance (km)
4
2
0
−2
−4
−6
10,000
100
1
0
2000
4000
6000
8000
10000
12000
14000
16000
18000

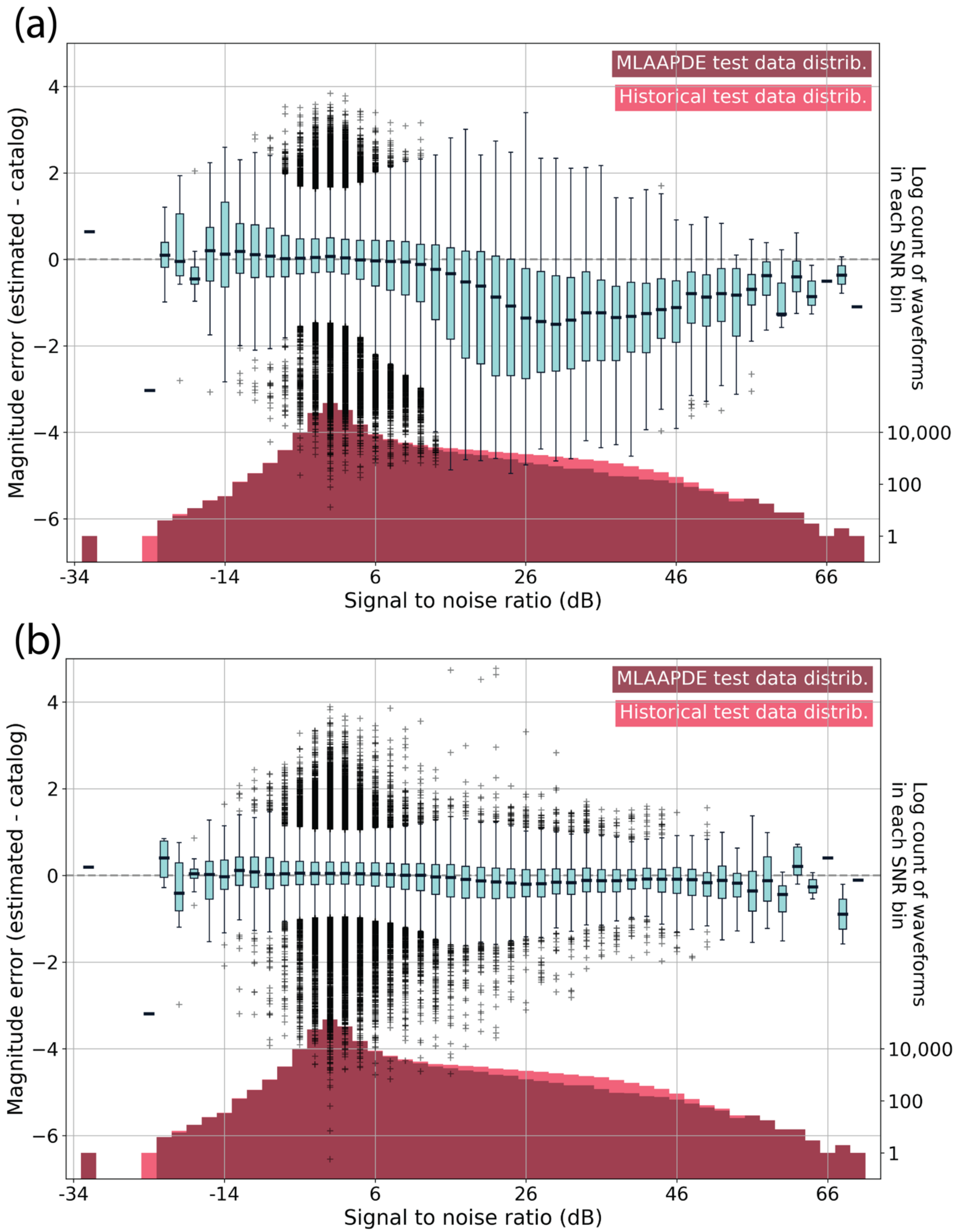
Finally, we examined the performance of AIMag by the SNR in deciBels (dB) of the waveform. We calculated waveform SNRs for the MLAAPDE and historical datasets according to the following formula:

$$SNR_{dB} = 20 \times log_{10}\left(\frac{S_{RMS}}{N_{RMS}}\right), \tag{4.3}$$

where S_{RMS} is the root mean square of the earthquake signal time series from the P-wave arrival to 30 seconds after, and N_{RMS} is the root mean square of the background noise from 60 seconds before the P-wave arrival up until 1 second before its arrival (for a total of 59 seconds of noise).

Figure 4.6 shows boxplots for the magnitude estimation error plotted against the waveform SNR corresponding to each estimation for the 7-second window (Figure 4.6a) and the 114-second window (Figure 4.6b). The remaining plots for all other time windows can be seen in the animation in Video 4.S6. We have again also included a logarithmic histogram showing the distribution of SNRs corresponding to the waveforms in the testing dataset. We observe that at the full waveform window length of 114 seconds in Figure 4.6b, the median errors have converged around zero, indicating that there is not a bias in AIMag's performance by SNR. However, with the 7-second window in Figure 4.6a, the model has more spread in the estimations and appears to underestimate the magnitudes of waveforms with a SNR above ~10 dB, but it performs better below this range.

Figure 4.6 (next page)**.** Boxplots for **(a)** the 7-second window and **(b)** the 114-second window showing the error in the magnitude estimation (estimated magnitude – catalog magnitude) made by the model on the left y-axis plotted against the waveform signal to noise ratio (SNR) in decibels (dB) on the x-axis corresponding to each estimation. An animation showing the boxplots for all time windows can be found in the supplemental files included with this dissertation as Video 4.S6. The characteristics of the boxplot in this figure are the same as those for Figures 4.3, 4.4, and 4.5, with the median magnitude estimation error for each SNR bin shown by the horizontal black line in the center of each light blue box. On the right y-axis, we have plotted a logarithmic histogram showing the distribution of SNRs corresponding to the waveforms in the testing dataset. The distribution of SNRs in the MLAAPDE dataset is shown in dark red, with the distribution of SNRs in the historical dataset stacked on top of those in light red.


(a)
MLAAPDE test data distrib.
Historical test data distrib.
Magnitude error (estimated - catalog)
Log count of waveforms in each SNR bin
4
2
0
−2
−4
−6
10,000
100
1
-34
-14
6
26
46
66
Signal to noise ratio (dB)
(b)
MLAAPDE test data distrib.
Historical test data distrib.
Magnitude error (estimated - catalog)
Log count of waveforms in each SNR bin
Signal to noise ratio (dB)

This range of underestimations aligns with where the historical data are present in greatest number, as seen in the histogram in Figure 4.6a. To verify that the data type was causing this bias above ~10 dB, we plotted the magnitude error versus SNR for the MLAAPDE and historical testing waveform datasets separately (Figure 4.S4). This showed that the underestimations seen in Figure 4.6 came primarily from the historical waveforms, as the MLAAPDE waveforms have median estimation errors of around zero even at 1-second window lengths. It is therefore likely that some properties of the historical waveforms that were not present in the MLAAPDE waveforms led to the magnitude underestimations by AIMag at short time windows, as the model was not trained on these types of waveforms. In Figure 4.6a at the short 7-second time window, the model appears to perform better below the ~10 dB range because there are many more MLAAPDE data samples in those bins, and the MLAAPDE data magnitudes do not tend to be underestimated by AIMag (Figure 4.S4b). In the higher SNR bins where underestimation occurs, a larger proportion of the samples come from the historical data that were not included in the training dataset. At longer time windows (Figure 4.6b), this gap closes, as the magnitude errors are lower for the historical data than at shorter time windows (Figure 4.S4c). However, AIMag still underestimates the magnitudes of earthquakes from the historical testing dataset from waveforms with lower SNRs, even at the longest 114-second time window.

4.3.3. Model Performance by Window Length

To assess the performance of AIMag by window length, for each time window we calculated the error (estimated magnitude – catalog magnitude) for each model estimation, followed by the median and standard deviation of those errors. We then plotted the median and standard deviation of the errors for each time window (Figure 4.7). Generally, the median error is close to zero across the entire range of tested time windows, but the standard deviation of the errors tends to decrease from a maximum of 1.18 at the 1-second time window to a minimum of 0.55 at the 114-second time window. The standard deviation decreases most quickly at the smaller time windows through 14 seconds, and decreases more slowly after that.

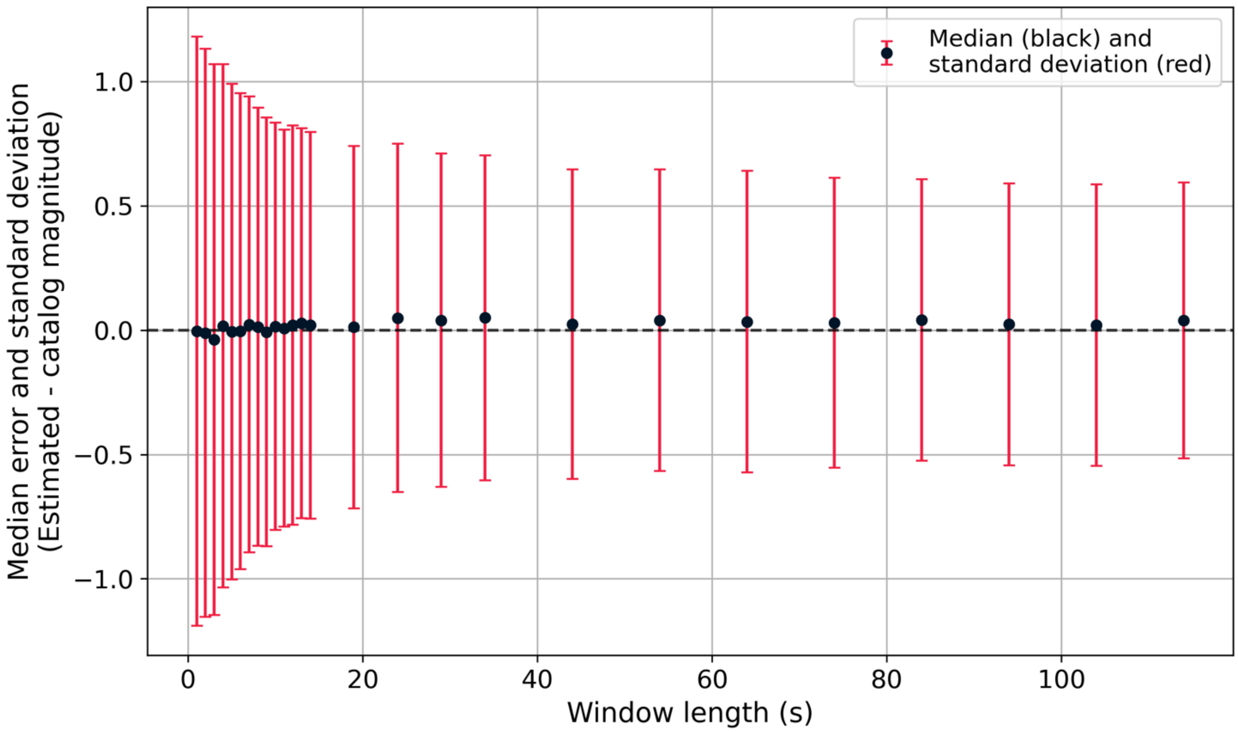


Figure 4.7. The median and standard deviation of the model's estimation error on the testing dataset (estimated magnitude – catalog magnitude) plotted for each of the tested time windows. The black dots show the median error for each time window, and the red bars extend to one standard deviation above the median and one standard deviation below.

As the potential utility of this study is for real-time earthquake analysis, we were interested in investigating the tradeoff between estimation accuracy and speed. The estimation speed will be limited by how long of a waveform AIMag needs to make the estimation, which is why we iterated our training and testing over a range of time windows. To assess this tradeoff, we took a closer look at the estimation errors within each time window using the median (50^{th} percentile) estimated magnitude from the boxplots for each time window as shown in Figure 4.3 and Video 4.S1. Because AIMag's performance varies greatly based on earthquake magnitude and our interest is primarily on larger events, we assessed the quality of our models by disaggregating their performance by event magnitude. We set our estimation error tolerance to 0.5 (i.e., if AIMag's median estimation for a particular catalog magnitude bin is within 0.5 of the correct catalog magnitude, we call this an acceptable estimation. We could then plot the median errors across every time window to determine what the acceptable estimation range was for each time window. Figure 4.8 shows that by the time the window length reaches 29 seconds, the high magnitude range by which we can achieve an acceptable estimation error performance is already essentially maximized, with limited improvements above this window. For the 29-second window, magnitude estimate median errors are acceptable between M2.7 and M7.6, excluding M7.5 where the errors are slightly too high. For the longest 114-second window, magnitude estimate median errors are acceptable between M2.3 and M7.6, also excluding M7.5 for the same reason.

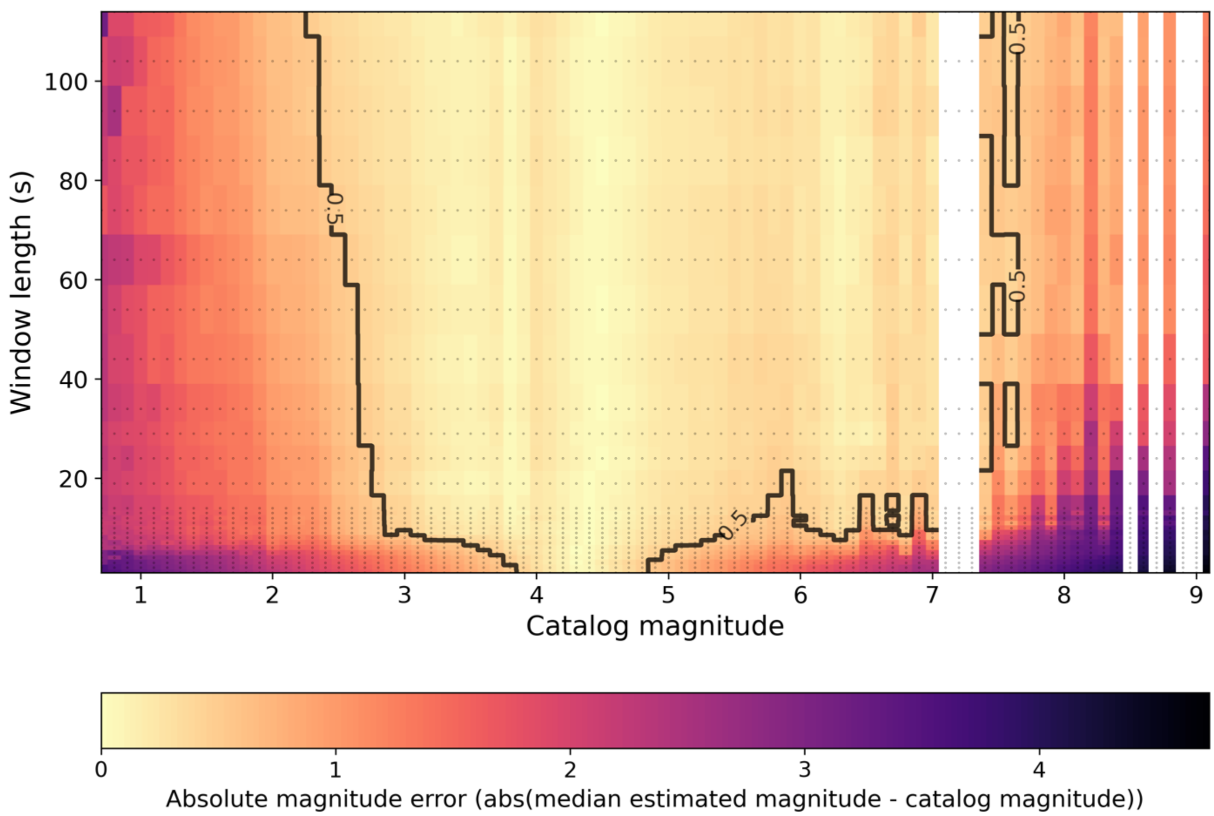


Figure 4.8. Plot showing the absolute error of the median of the estimations made by the model for each catalog magnitude bin (abs(median_$M_{estimated}$ – $M_{catalog}$)) across the range of window lengths tested (1 to 114 seconds) for the testing dataset. The catalog magnitude is on the x-axis and the waveform window lengths are on the y-axis. The color spectrum shows the absolute error, with lighter colors indicating a smaller error and darker colors indicating a larger error. White gaps indicate magnitudes for which there were no earthquakes in the testing dataset. The black line is the contour of an absolute error of 0.5 magnitude units.

We also calculated and plotted the absolute magnitude estimation errors for the 95th percentile of the data (Figure 4.S5) and found that the range over which the absolute error was less than or equal to 0.5 magnitude units extends from M4.4 to M8.6 for the 114-second window. For the 29-second window discussed in the previous paragraph, the 95th percentile absolute error is less than or equal to 0.5 for the magnitude range M4.4-M8.3, excluding M8.2 where the errors are slightly too high. Using the 95th percentile to choose our acceptable performance range instead of the median (50th percentile) would exclude the data below M4.4, as they tend to be overestimated more often. However, it would also result in including the ~M7.5-M8.6 data that we know tend to be underestimated. Which metric to use for determining an acceptable operating magnitude range over which to implement this model is likely best left as a user choice based on the monitoring goal.

As discussed in the Model Performance by Magnitude section, AIMag tends to underestimate the magnitudes of large earthquakes, likely primarily due to the short time windows we use. Traditionally, the magnitudes of large earthquakes are characterized with long-period observations. The W-phase is particularly important – it is a long-period (~100-1000 second) phase that arrives in displacement seismograms between the P- and S-waves and represents a superposition of other body-wave phases (Kanamori, 1993; Kanamori and Rivera, 2008). W-phase moment magnitudes (M_{ww}) provide accurate measures of earthquake size that work for moderate (Hayes et al., 2009) as well as great earthquakes (e.g., the 2011 Tōhoku earthquake as demonstrated in Duputel et al., 2011). M_{ww} can be calculated relatively rapidly and is the preferred magnitude used by the NEIC.

In this study, we used raw waveforms (stream-normalized in units of counts) rather than displacement waveforms and window lengths only up to 114 seconds (only half of which contains earthquake signal). The W-phase is therefore not fully captured in the observations because it will not have arrived in its entirety yet. It is possible that training and testing with longer waveforms that include the S-wave (and therefore the entire W-phase) or data scaling that enhances long period data would improve AIMag's performance on large earthquakes. The drawback of that approach would be an increase in the complexity of data pre-processing and potentially the amount of time required to obtain a good magnitude estimate. For comparison, how fast M_{ww} can be calculated depends on earthquake location and station density, and in a best-case scenario the NEIC can calculate M_{ww} in as few as ~7 minutes and typically within about 20 minutes (Hayes et al., 2009). This is much longer than the 29-second time window beyond which we saw limited returns in this study (Figure 4.8). Although that 29 seconds does not account for the time it takes for the signal to be picked up by the seismometer, as ours is a single-station method, the additional delay time only depends on the travel time to the single closest station to the hypocenter.

A closer analog to our magnitude estimation method is likely that of the broadband P-wave moment magnitude (M_{wp}). Similar to our model, M_{wp} only relies on the P-waves from earthquakes, but like M_{ww} it is also an inversion method that takes integrated displacement waveforms. It performs well on as little as 120 seconds of waveform data, but it is of limited use for earthquakes smaller than ~M5.5 when long period signals are weaker and SNR is low, as it may overestimate the magnitude (Tsuboi et al., 1999). Additionally, when data are used from seismometers that are less sensitive to longer periods, the M_{wp} method underestimates the magnitudes of large earthquakes (~M8 and greater). It is also less reliable for data from stations in the near field because the method was based on far field P-waveform displacement, and this leads to underestimations of the earthquake magnitude (Hirshorn et al., 2013).

We have shown that although the magnitudes of larger earthquakes may be underestimated by AIMag, similar to the M_{wp} method, our model can still tell that they are large (Figure 4.3). Additionally, it also works for much smaller magnitude earthquakes than M_{wp} (Figure 4.3), performs well in the near field (Figure 4.5b), and can provide estimates more quickly than M_{ww}. AIMag can feasibly be useful for flagging earthquakes that a monitoring agency would be concerned are large and potentially tsunamigenic within a much shorter time window than is currently possible with the M_{ww} inversion method and without some of the drawbacks of the M_{wp} method.

4.3.4. Model Performance on Edge Cases

With a model designed to generalize across the globe, we wanted to check AIMag's performance for specific earthquakes with several unusual characteristics. We browsed the testing data catalog for more atypical earthquakes, both from the historical data recorded between 2000 and 2013 and the MLAAPDE data recorded in the year 2020, and hand-picked seven events to examine more closely.

This group of seven included two tsunami earthquakes, which are earthquakes that produce tsunamis that are much larger than expected based on their magnitude (Kanamori, 1972). They are typically depleted in high frequency seismic radiation, and likely rupture all the way to the trench at slow rupture velocities (Lay et al., 2012). The two earthquakes we tested were the M7.7 2006

Pangandaran earthquake off the coast of Java, Indonesia (Ammon et al., 2006) and the M7.8 2010 Mentawai earthquake off the coast of Sumatra, Indonesia (Yue et al., 2014). We also looked at three unusually deep earthquakes, as most of the waveforms come from shallow events as seen in the histogram in Figure 4.4. These were a 598-kilometer-deep M8.3, a 624-kilometer-deep M6.9, and a 641-kilometer-deep M7.5 earthquake. The final two earthquakes in the group of seven were unusually "complex" events to determine how the model could handle this. We chose the M8.6 2012 Wharton Basin earthquake off the coast of Sumatra, which is the largest strike-slip earthquake ever recorded and ruptured multiple faults (Meng et al., 2012; Yue et al., 2012), and the M7.6 2020 Sand Point earthquake on the Aleutian Subduction Zone, which generated an unusually large tsunami for the reported strike-slip mechanism of the event (Herman and Furlong, 2021; Santellanes et al., 2022).

We again produced a boxplot of AIMag's estimated earthquake magnitudes versus the catalog magnitudes for the testing dataset for the 114-second window (Figure 4.9), but in this figure we plotted the seven selected edge case earthquakes in light red and all other earthquakes in light blue. All of the estimations for these edge case events, with the exception of the 641-kilometer-deep event (marked B in Figure 4.9), fall within the typical range of estimations for the other earthquakes in the dataset for their respective magnitude bins, indicating that in many cases the model performs well even with unusual earthquakes.

The underestimated 641-kilometer-deep M7.5 (B) is one of a series of three deep earthquakes that occurred west of the Philippines on July 23, 2010 (Hayes et al., 2017). With magnitudes of M7.3, M7.5, and M7.6, the first of this series was not included in the expanded historical testing dataset because we only downloaded data for earthquakes M7.5 and greater. To explore why AIMag may have underestimated the magnitude of the 641-kilometer-deep event, we also pulled out and plotted the estimations made for the M7.6 earthquake that occurred at a depth of 578 kilometers in this same area. These two earthquakes can be seen in Figure 4.S6. Although the M7.6 is not underestimated as much as the M7.5, it is still underestimated compared to the other edge cases visualized in Figure 4.9. This suggests that the underestimation for these specific events could be related to the tectonic setting in this area, or potentially other spatial effects such as the geometry of the seismic stations that recorded the waveforms from these events.

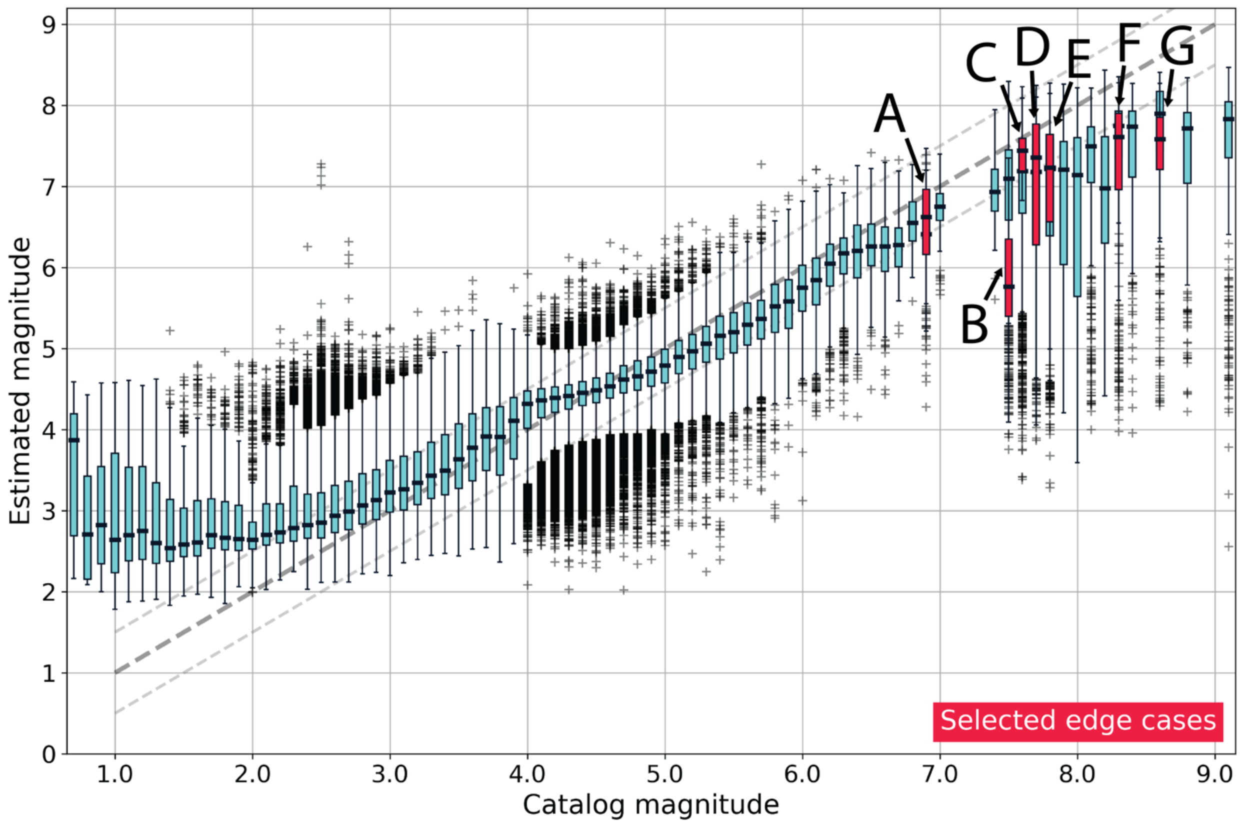


Figure 4.9. Plot of the model's estimated earthquake magnitudes for the testing dataset (both MLAAPDE data and historical) for the 114-second window (similar to Figure 4.3b), with the catalog magnitude for the earthquakes on the x-axis, and the model's estimated magnitudes for those same earthquakes on the left y-axis. The seven selected edge case earthquakes as described in the body of the paper are shown in light red, whereas all other earthquakes are plotted in light blue. The earthquake marked A on the plot is the 624-kilometer-deep event, B is the 641-kilometer-deep event, C is the Sand Point complex earthquake, D is the Pangandaran tsunami earthquake, E is the Mentawai tsunami earthquake, F is the 598-kilometer-deep event, and G is the Wharton Basin complex earthquake.

4.4. Conclusions

In this study, we trained a machine learning model for global single-station earthquake magnitude estimation based solely on the recorded waveform and requiring no additional information about the event or station involved. We demonstrate here the effectiveness and utility of the MLAAPDE dataset described in Cole et al. (2023), which was designed to make the USGS NEIC's PDE catalog more easily accessible for machine learning research. Our training process implements data augmentation to improve and generalize model performance.

At its best, our model, AIMag, can successfully determine earthquake magnitudes with an absolute error margin in median estimated magnitude of 0.5 magnitude units or less for a range of earthquake magnitudes from M2.3 to M7.6. It can achieve success across this magnitude range with a minimum window length of 29 seconds of data (with the phase arrival at halfway through this window). AIMag is not biased in its magnitude estimations by either earthquake source depth, hypocentral distance, or waveform signal to noise ratio, and it also performs well on most tested "edge case" earthquakes that are unique in one aspect or another (e.g., tsunami earthquakes, deep earthquakes, and complex ruptures).

AIMag can aid near real-time global earthquake monitoring operations by quickly and accurately assessing the magnitudes of earthquake recordings by avoiding reliance on earthquake location. Future work may include testing the performance of this model within the NEIC's real-time monitoring system and expanding its capability to handle variable-length input data in real-time, as well as combining it with other work on a machine-learning-based pick association method in development at the NEIC.

4.5. Data and Resources

Supplemental Material for this chapter contains 6 figures as well as 6 multi-frame figure animations to provide extra information about the datasets used in training and testing, as well as additional testing results for all time windows as described in this manuscript.

The codes used to train and test the models and construct the figures described in this manuscript are available at https://github.com/UO-Geophysics/AIMag. As of the time of publication, this repository was last updated in December 2023. The MLAAPDE module used in this work is available in Cole and Yeck (2022).

The facilities of EarthScope Consortium were used for access to waveforms, related metadata, and/or derived products used in this study. These services are funded through the Seismological Facility for the Advancement of Geoscience (SAGE) Award of the National Science Foundation under Cooperative Support Agreement EAR-1851048.

We make use of data from the following networks: AC, AE, AF, AG, AI, AK, AT, AU, AV, AY, AZ, BC, BE, BK, BL, BX, C, C0, C1, CA, CC, CH, CI, CM, CN, CO, CW, CY, CZ, DK, DR, EC, EO, EP, ET, G, GB, GE, GM, GO, GR, GS, GT, HK, HL, HT, HV, IE, II, IM, IN, IO, IU, JP, KC, KG, KN, KO, KR, KS, KY, KZ, LB, LD, LO, LX, MB, MG, MI, MM, MN, MP,

MU, MX, MY, N4, NA, NC, NI, NM, NN, NP, NQ, NU, NV, NY, NZ, O2, OE, OH, OK, OO, OV, OX, PA, PE, PL, PM, PO, PR, PS, PT, RM, RO, RV, SB, SC, SE, SN, SS, SV, TA, TC, TJ, TM, TW, TX, UO, US, UU, UW, VU, WC, WI, WM, WU, WY, YX, ZC, ZD, and ZW. We appreciate the hard work of the engineers and network operators who make this data available in near real-time. We have included the International Federation of Digital Seismograph Network (FDSN) references in the Appendix.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

4.6. Acknowledgements

This work was conducted as part of a USGS Pathways Internship at the NEIC in Golden, Colorado. It was also supported by National Science Foundation (NSF) grant OAC-1835661 and National Aeronautics and Space Administration (NASA) grants 80NSSC19K0360 and 80NSSC22K0458.

We thank two anonymous reviewers, Adam Ringler, Brian Shiro, and two additional internal USGS reviewers for their careful reading and feedback, which helped improve this manuscript.

4.7. Supplemental Figures

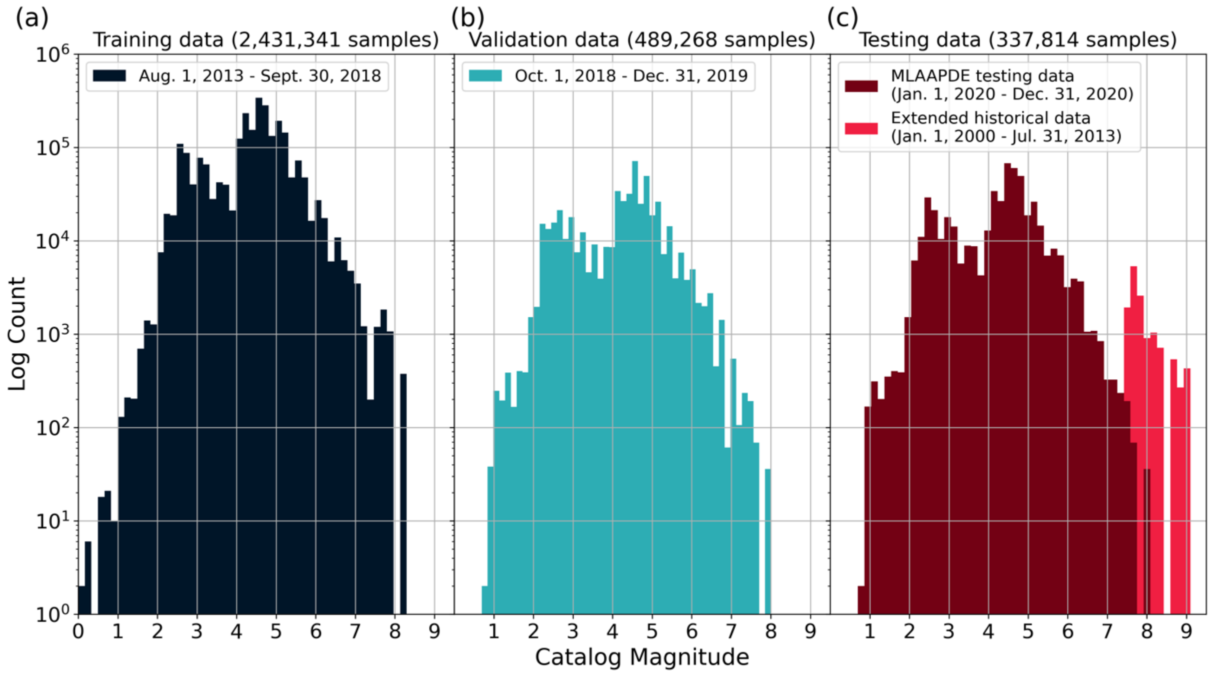
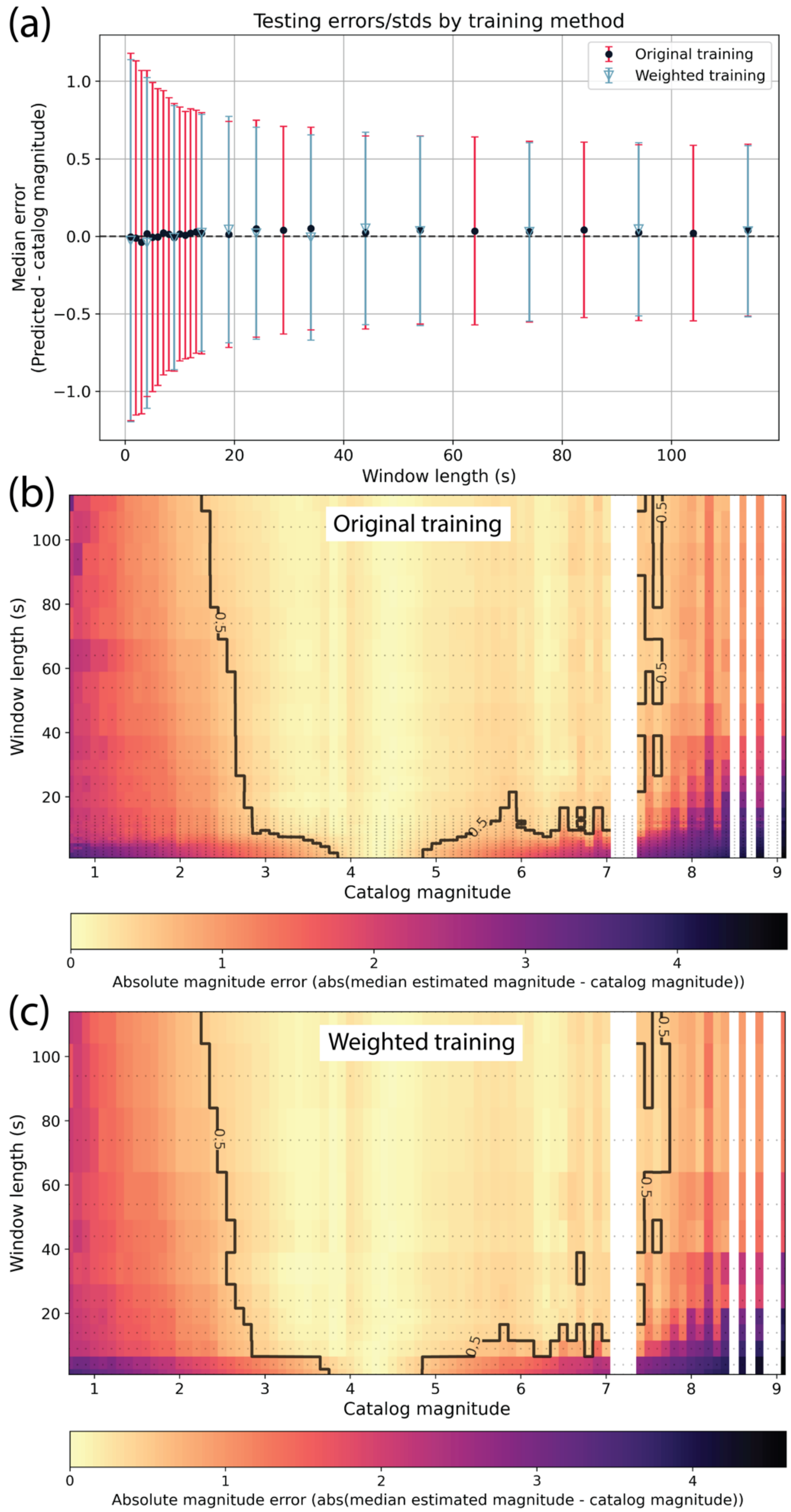


Figure 4.S1. Logarithmic histograms showing the distributions of magnitudes associated with the MLAAPDE waveforms selected in the **(a)** training (dark blue), **(b)** validation (light blue), and **(c)** testing (dark red) datasets. Panel (c) includes the expanded historical testing data from 2000 to mid-2013 (light red) that is not from the MLAAPDE dataset.

Figure 4.S2 (next page)**. (a)** The median and standard deviation of the model's estimation error on the testing dataset (estimated magnitude – catalog magnitude) plotted for each of the tested time windows. The black dots show the median error for each time window for the original training scheme described in the manuscript, with the corresponding red bars extending to one standard deviation above the median and one standard deviation below. The blue triangles show the median error for each time window for the weighted training scheme, with the corresponding blue bars extending to one standard deviation above the median and one standard deviation below. **(b)** Figure 4.8 from the manuscript, which has the results for the original training scheme described in the manuscript shown only for comparison to (c). **(c)** Analogous to (b)/Figure 4.8 for the weighted training scheme – a plot showing the absolute error of the median of the estimations made by the weighted training model for each catalog magnitude bin (absolute value of (median estimated magnitude – catalog magnitude)) across the range of window lengths tested (1 to 114 seconds) for the testing dataset. The catalog magnitude is on the x-axis, and the waveform window lengths are on the y-axis. The color spectrum shows the absolute error, with lighter colors indicating a smaller error and darker colors indicating a larger error. White gaps indicate magnitudes for which there were no earthquakes in the testing dataset. The black line is the contour of an absolute error of 0.5 magnitude units.



(a)
Testing errors/stds by training method
Original training
Weighted training
Median error
(Predicted - catalog magnitude)
Window length (s)
(b)
Original training
Window length (s)
Catalog magnitude
Absolute magnitude error (abs(median estimated magnitude - catalog magnitude))
(c)
Weighted training
Window length (s)
Catalog magnitude
Absolute magnitude error (abs(median estimated magnitude - catalog magnitude))

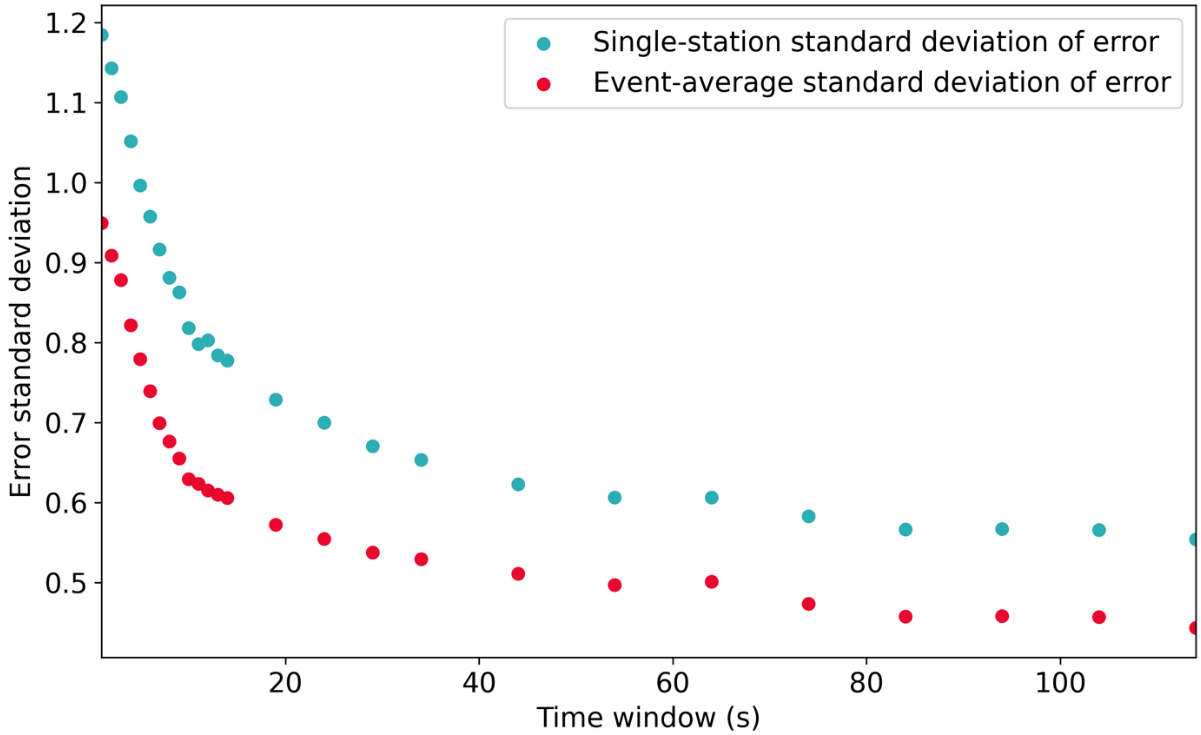


Figure 4.S3. The standard deviation of the model's estimation error on the testing dataset (estimated magnitude – catalog magnitude) plotted for each of the tested time windows for the single-station method (blue) and the event-averaged method (red).

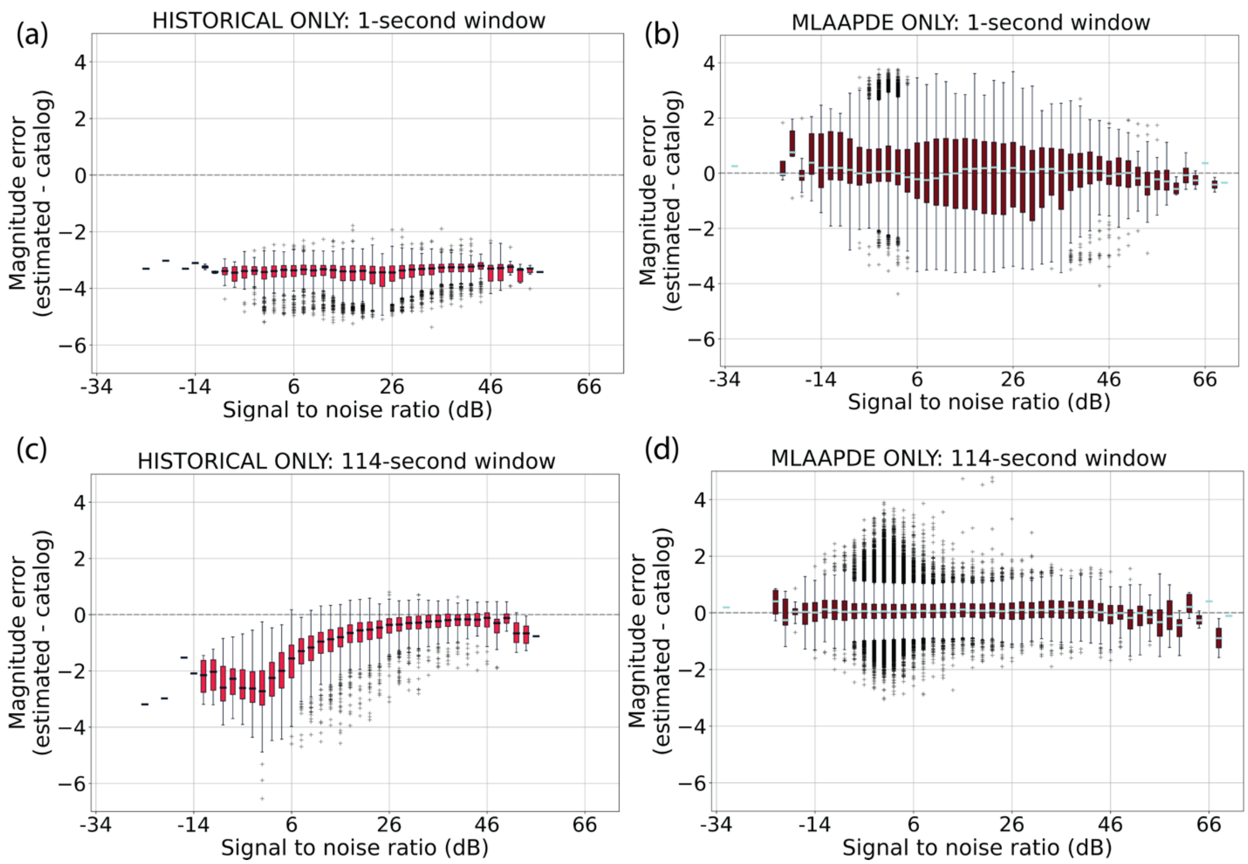


Figure 4.S4. Boxplots showing the error in the magnitude estimation (estimated magnitude – catalog magnitude) made by the model on the y-axis plotted against the waveform signal to noise ratio (SNR) in deciBels (dB) on the x-axis corresponding to each estimation. In each of these panels, the median magnitude estimation error for each SNR bin on the x-axis is shown by the horizontal black line in the center of each light blue box. The lower and upper bounds of the light blue boxes indicate the first and third quartiles of the estimations in each bin, respectively, and the whiskers extend to 1.5 times the inter-quartile range. The individual points beyond the whiskers indicate outliers. The panels show the magnitude error estimations for **(a)** the 1-second window for the historical testing dataset only in light red (medians in black), **(b)** the 1-second window for the MLAAPDE testing dataset only in dark red (medians in light blue), **(c)** the 114-second window for the historical testing dataset only in light red (medians in black), and **(d)** the 114-second window for the MLAAPDE testing dataset only in dark red (medians in light blue).

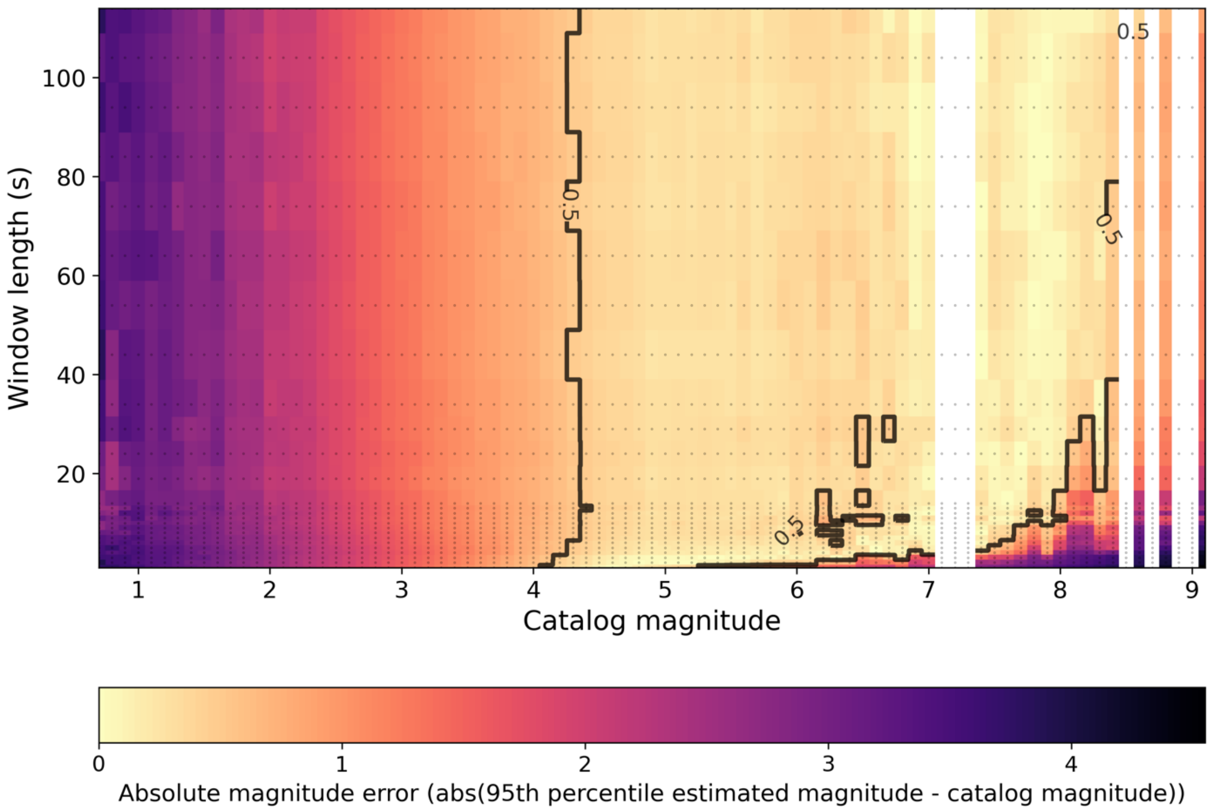


Figure 4.S5. Plot showing the absolute error of the 95th percentile of the estimations made by the model for each catalog magnitude bin (absolute value of (95th percentile estimated magnitude – catalog magnitude)) across the range of window lengths tested (1 to 114 seconds) for the testing dataset. The catalog magnitude is on the x-axis, and the waveform window lengths are on the y-axis. The color spectrum shows the absolute error, with lighter colors indicating a smaller error and darker colors indicating a larger error. White gaps indicate magnitudes for which there were no earthquakes in the testing dataset. The black line is the contour of an absolute error of 0.5 magnitude units.

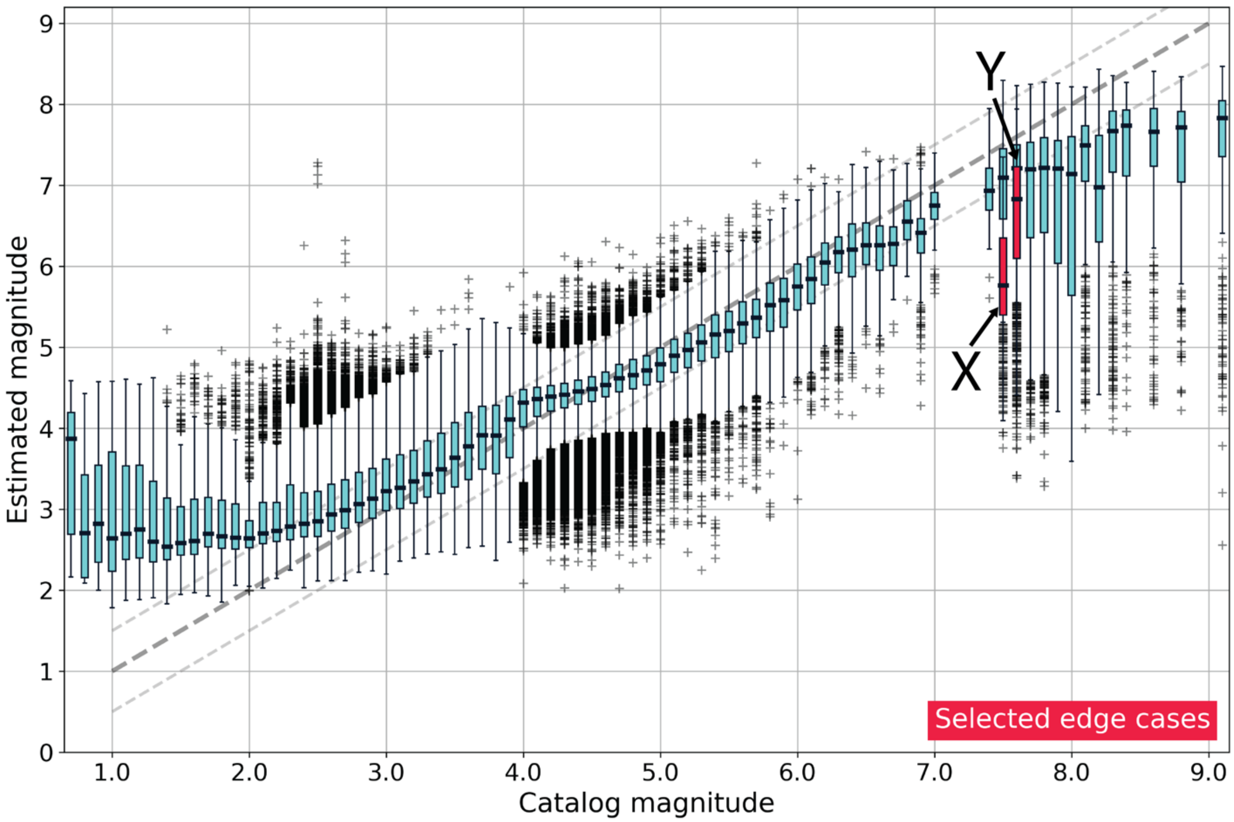


Figure 4.S6. Plot of the model's estimated earthquake magnitudes for the testing dataset (both MLAAPDE and historical data) for the 114-second window (similar to Figure 4.3c), with the catalog magnitude for the earthquakes on the x-axis, and the model's estimated magnitudes for those same earthquakes on the left y-axis. The two selected edge case earthquakes are shown in light red, whereas all other earthquakes are plotted in light blue. The earthquake marked X on the plot is the M7.5 641-kilometer-deep event, and Y is the M7.6 578-kilometer-deep event in the same series of three deep earthquakes in the Philippines on July 23, 2010 (Hayes et al., 2017).

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

In this dissertation, I have presented work addressing several opportunities we found in the realm of earthquake early warning and rapid response for improvements to be made and knowledge to be gained using geodetic instrumentation and machine learning.

In Chapter II, we focused on the basic science question of how earthquakes begin using borehole strainmeters. I explained the ongoing debate about whether strong determinism exists and how soon small and large earthquakes can be distinguished from one another, which crucially affects our ability to estimate their magnitudes in real-time for early warning applications. Although there are far fewer strainmeters in use around the world than seismometers or GNSS stations, we used data from earthquakes **M**6 or greater in North America, Taiwan, and Japan to explore whether strong determinism was visible in peak strain data, ultimately concluding that we did not observe it. We generated synthetic earthquake strain waveforms using SW4 for earthquakes that we designed to be deterministic and did observe characteristics of strong determinism in that data, suggesting that if it did exist in the real data, our analysis should have revealed it. However, there are several ways in which this work should be expanded to confirm the conclusion that we reached. Earthquakes smaller than **M**6 should be analyzed as well, along with additional events from outside of North America if possible. Furthermore, there are still some unexplained features in the peak strain waveforms that warrant more investigation. In particular, future work should be undertaken to discover what may be causing the station-to-station variability in the exact pattern of the peak strain growth for the same earthquakes. Looking into the relationship between the earthquake radiation patterns in strain and these features of the peak strain waveforms would be a worthwhile first approach to this issue.

In Chapter III, I covered the development of a machine learning model trained to detect and identify the timing of P-wave arrivals in noisy real-time GNSS data. We were motivated by the need to incorporate GNSS data into earthquake early warning systems in order to accurately determine the magnitudes of the largest earthquakes quickly, with the challenge being that GNSS data is very noisy, and including noisy data in a magnitude determination algorithm may degrade that algorithm's performance. We trained a model using synthetic displacement waveforms combined with real noise, and found that for both additional synthetic data and real data, our model performed well as a "filter" to remove waveforms where P-wave arrivals were not detected from inclusion in an event-average PGD magnitude estimation algorithm, decreasing the error in the estimations. However, as a pure "detector" of earthquakes we found that the model's performance in terms of true negative/positive and false negative/positive outcomes strongly depended on the choice of confidence threshold used for analysis, and the model also tended to perform best for the largest earthquakes where waveforms had large displacements and high signal to noise ratios. For the real data in particular, there was a lot of spread in the model's predictions of the P-wave arrival times in the waveforms. This was likely due in part to the heavily skewed real testing dataset, as many more of the waveforms contained only noise than actual earthquakes, and we also had very few large earthquakes in the dataset (which the model did perform better on).

The incorporation of GNSS data into ShakeAlert with the GFAST algorithm has taken many years to accomplish, and any additional improvements would also require extensive testing to prove that they are reliable and accurate enough to use. While we demonstrated that our model has the potential to improve the magnitude estimates made by GFAST, it would require significant reconfiguration to operate in real-time, and it should also be trained and tested with more GNSS data from other regions. Other machine learning models trained using GNSS data also show promise toward improving how our early warning systems work for large earthquakes. For example, Thomas et al. (2023) trained a model to denoise GNSS data, and Lin et al. (2021, 2023) developed a model that can make real-time predictions of earthquake source parameters and shaking intensity. With a combination of these models and ours as presented in Chapter III, one can envision a GNSS-based earthquake early warning system which denoises data, detects earthquakes, and then uses this higher-quality denoised data where seismic wave arrivals are detected to more accurately forecast shaking and issue alerts – potentially without requiring the involvement of seismic data at all.

Finally, in Chapter IV I expanded the scope of this dissertation from regional earthquake early warning to global rapid earthquake analysis. The USGS's NEIC must process and publish information about earthquakes worldwide across many tectonic settings, and as a result determining the magnitude of such a wide variety of earthquakes can be complex and involve many different types of magnitudes. Motivated by the vast availability of historical data published by the NEIC, we employed machine learning again to develop a model which could estimate earthquake magnitudes uniformly regardless of the location of an earthquake, and independently from station to station. Though the model tended to underestimate the magnitudes of the largest earthquakes, as expected from the magnitude saturation phenomenon in seismic data and the relative lack of large earthquakes in the training dataset, we showed that the model was accurate with a median estimated magnitude error of 0.5 magnitude units or less for earthquakes from M2.3 to M7.6. This kind of model could help to streamline real-time global earthquake monitoring operations; however, like the GNSS detection model presented in Chapter III, in future work it would need to be reconfigured to operate in real-time and evaluate its performance within the NEIC's monitoring system.

In this dissertation I have endeavored to demonstrate the potential benefits of continuing to expand what types of data and methods we use for rapid earthquake analysis. They have the ability to improve the systems that we rely on to evaluate the hazards associated with earthquakes as they occur, and the quicker and more accurate these systems are, the more effective they will be for risk mitigation to ensure that life and property can be protected.

APPENDIX

FDSN NETWORK REFERENCES

Alaska Earthquake Center, Univ. of Alaska Fairbanks. (1987). *Alaska Geophysical Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/AK

Alaska Volcano Observatory/USGS. (1988). *Alaska Volcano Observatory* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/AV

Alberta Geological Survey / Alberta Energy Regulator. (2013). *Regional Alberta Observatory for Earthquake Studies Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/RV

Albuquerque Seismological Laboratory (ASL)/USGS. (1980). *US Geological Survey Networks* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/GS

Albuquerque Seismological Laboratory (ASL)/USGS. (1990). *United States National Seismic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/US

Albuquerque Seismological Laboratory (ASL)/USGS. (1993). *Global Telemetered Seismograph Network (USAF/USGS)* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/GT

Albuquerque Seismological Laboratory/USGS. (2013). *Central and Eastern US Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/N4

Albuquerque Seismological Laboratory/USGS. (2014). *Global Seismograph Network (GSN - IRIS/USGS)* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/IU

Aristotle University of Thessaloniki. (1981). *Aristotle University of Thessaloniki Seismological Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/HT

Arizona Geological Survey. (2007). *Arizona Broadband Seismic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/AE

Aster, R., & Kyle, P. (2007). *Erebus Tomography and Source Studies* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/ZW_2007

Balling, N., Tilmann, F., & Kind, R. (2010). *The Jutland-to-Lower-Saxony (JULS) profile* [Data set]. GFZ Data Services. https://doi.org/10.14470/MJ7565224230

Botswana Geoscience Institute. (2001). *Botswana Seismological Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/BX

British Geological Survey. (1970). *Great Britain Seismograph Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/av8j-nc83

Bureau of Economic Geology, The University of Texas at Austin. (2016). *Texas Seismological Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/TX

California Institute of Technology and United States Geological Survey Pasadena. (1926). *Southern California Seismic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/CI

Cascades Volcano Observatory/USGS. (2001). *Cascade Chain Volcano Monitoring* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/CC

Central Asian Institute for Applied Geosciences. (2008). *Central Asian Seismic Network of CAIAG* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/KC

Centro de Geociencias, UNAM. (2003). *Permanent Seismic Network of Queretaro State, Mexico* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/MG

Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE), Ensenada. (1980). *Red Sísmica del Noroeste de México* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/BC

Charles University in Prague (Czech), Institute of Geonics, Institute of Geophysics, Academy of Sciences of the Czech Republic, Institute of Physics of the Earth Masaryk University (Czech), & Institute of Rock Structure and Mechanics. (1973). *Czech Regional Seismic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/CZ

Cochran, J., Li, Y.-G., & Steidl, J. (2008). *Seismology Rapid Response Test During the SoSAF Shakeout* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/YX_2008

Colorado Geological Survey. (2016). *Colorado Geological Survey Seismic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/C0

Darold, A. (2014). *4D Integrated Study Using Geology, Geophysics, Reservoir Modeling & Rock Mechanics to Develop Assessment Models for Potential In* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/ZD_2014

DellaGiustina, D. (2021). *SIIOS Flagstaff 2021 Mini Campaign* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/ZD_2021

Department of Meteorology and Hydrology - National Earthquake Data Center. (2016). *Myanmar National Seismic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/MM

DeShon, H., Hayward, C., Stump, B., Magnani, M. B., & Hornbach, M. (2013). *North Texas Earthquake Study: Azle and Irving/Dallas* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/ZW_2013

Federal Institute for Geosciences and Natural Resources. (1976). *German Regional Seismic Network (GRSN)*. Bundesanstalt für Geowissenschaften und Rohstoffe. https://doi.org/10.25928/mbx6-hr74

Fuchs, F., Novoselov, A., Gröschl, G., & Bokelmann, G. (2020). *ThunderSeis Seismo-Acoustic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/ZD_2020

GEOFON Data Centre. (1993). *GEOFON Seismic Network* [Data set]. Deutsches GeoForschungsZentrum GFZ. https://doi.org/10.14470/TR560404

Geophysical Survey of the National Academy of sciences of Tajikistan. (2009). *Tajikistan National Seismic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/TJ

Geoscience Australia. (2021). *Australian National Seismograph Network Data Collection* (Version 2.0, September 2018). Commonwealth of Australia (Geoscience Australia). https://doi.org/10.26186/144675

GNS Science. (2021). *GeoNet Aotearoa New Zealand Seismic Digital Waveform Dataset* [Data set]. GNS Science. https://doi.org/10.21420/G19Y-9D40

Idaho National Laboratory. (1972). *INL Seismic Monitoring Program* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/IE

Incorporated Research Institutions For Seismology. (1970). *SINGLE STATION* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/SS

Institut Cartogràfic i Geològic de Catalunya. (1984). *Catalan Seismic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/CA

Institut de physique du globe de Paris (IPGP), & École et Observatoire des Sciences de la Terre de Strasbourg (EOST). (1982). *GEOSCOPE, French Global Network of broad band seismic stations*. Institut de physique du globe de Paris (IPGP), Université de Paris. https://doi.org/10.18715/GEOSCOPE.G

Institut De Physique Du Globe De Paris (IPGP). (2008). *GNSS, seismic broadband and strong motion permanent networks in West Indies*. Institut de physique du globe de Paris (IPGP), Université de Paris. https://doi.org/10.18715/antilles.WI

Institute of Earth Sciences, Academia Sinica, Taiwan. (1996). *Broadband Array in Taiwan for Seismology* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/TW

Institute of GeoSciences (IGEO), Polytechnic University of Tirana (PUT). (2002). *Albanian Seismological Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/AC

Instituto Dom Luiz (IDL) - Faculdade de Ciências da Universidade de Lisboa. (2003). *Instituto Dom Luiz (IDL) - Faculdade de Ciencias Universidade de Lisboa* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/LX

Instituto Nicaraguense de Estudios Territoriales (INETER). (1975). *Nicaraguan Seismic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/NU

Instituto Politecnico Loyola. (2012). *Observatorio Sismológico Politécnico Loyola* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/LO

Instituto Português do Mar e da Atmosfera, I.P. (2006). *Portuguese National Seismic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/PM

IRIS Transportable Array. (2003). *USArray Transportable Array* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/TA

Istituto Nazionale di Oceanografia e di Geofisica Sperimentale - OGS. (2016). *North-East Italy Seismic Network* [Data set]. FDSN. https://doi.org/10.7914/SN/OX

Istituto Nazionale di Oceanografia e di Geofisica Sperimentale. (1992). *Antarctic Seismographic Argentinean Italian Network - ASAIN* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/AI

Istituto Nazionale di Oceanografia e di Geofisica Sperimentale. (2014). *International OGS Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/IO

Kandilli Observatory And Earthquake Research Institute, Boğaziçi University. (1971). *Kandilli Observatory And Earthquake Research Institute (KOERI)* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/KO

Keller, R. (2009). *SINOPROBE-2 Controlled Source* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/ZC_2009

Kentucky Geological Survey/Univ. of Kentucky. (1982). *Kentucky Seismic and Strong Motion Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/KY

Kenyon, T., & Thelen, W. (2023). *Seismic determination of bedload and morphodynamic trends in the Nisqually River of Mount Rainier National Park, WA, USA* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/7g14-9528

Klemperer, S. & Miller, K. (2010). *Flexarray 3D Passive Seismic Imaging of Core-Complex Extension in the Ruby Range Nevada* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/YX_2010

KNDC/Institute of Geophysical Research (Kazakhstan). (1994). *Kazakhstan Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/KZ

KNMI. (2006). *Caribbean Netherlands Seismic Network*. Royal Netherlands Meteorological Institute (KNMI). https://doi.org/10.21944/dffa7a3f-7e3a-3b33-a436-516a01b6af3f

Kyrgyz Institute of Seismology, IVTAN/KIS , & University of California, San Diego. (1991). *Kyrgyz Seismic Telemetry Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/KN

Kyrgyz Institute of Seismology, KIS . (2007). *Kyrgyz Digital Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/KR

Lamont Doherty Earth Observatory (LDEO), Columbia University. (1970). *Lamont-Doherty Cooperative Seismographic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/LD

Malone, S., & Bodin, P. (2005). *Mt. St. Helens Temporary Array 2005* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/ZC_2005

Marshall, H.-P., & Havens, S. (2010). *Seismic Detection of Avalanches* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/ZC_2010

Mechie, J., & Weber, M. (2006). *GEO-DESIRE Wide-Angle Reflection / Refraction experiment, Israel and Jordan 2006* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/ZW_2006

MedNet Project Partner Institutions. (1990). *Mediterranean Very Broadband Seismographic Network (MedNet)*. Istituto Nazionale di Geofisica e Vulcanologia (INGV). https://doi.org/10.13127/SD/fBBBtDtd6q

Montana Bureau of Mines and Geology/Montana Tech (MBMG, MT USA). (1982). *Montana Regional Seismic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/MB

Moretti, I., Cornet, F., & Virieux, J. (2017). *3F-Corinth Project temporary experiment (ENK6-CT-2000-00056)*. RESIF - Réseau Sismologique et géodésique Français. https://doi.org/10.15778/RESIF.YX2001

National Centre for Seismological Research (CENAIS Cuba). (1998). *Servicio Sismologico Nacional de Cuba* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/CW

National Institute for Earth Physics (NIEP Romania). (1994). *Romanian Seismic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/RO

National Observatory of Athens, Institute of Geodynamics, Athens. (1975). *National Observatory of Athens Seismic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/HL

National Seismological Centre. (1998). *Centro Nacional de Sismologia* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/DR

Natural Resources Canada. (1975). *Canadian National Seismograph Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/CN

NOAA National Oceanic and Atmospheric Administration (USA). (1967). *National Tsunami Warning Center Alaska Seismic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/AT

Northern California Earthquake Data Center. (2014). *Berkeley Digital Seismic Network (BDSN)* [Data set]. Northern California Earthquake Data Center. https://doi.org/10.7932/BDSN

Obsercatorio Vulcanológico y Sismológico de Costa Rica, Universidad Nacional. (1984). *Observatorio Vulcanológico y Sismológico de Costa Rica* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/OV

Ocean Networks Canada. (2009). *NEPTUNE seismic stations* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/NV

OGS (Istituto Nazionale di Oceanografia e di Geofisica Sperimentale) and University of Trieste. (2002). *North-East Italy Broadband Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/NI

Ohio Geological Survey. (1999). *Ohio Seismic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/OH

Oklahoma Geological Survey. (1978). *Oklahoma Seismic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/OK

Oklahoma Geological Survey. (2018). *Oklahoma Consolidated Temporary Seismic Networks* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/O2

Pacific Tsunami Warning Center. (1965). *Pacific Tsunami Warning Seismic System* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/PT

Penn State University. (2004). *AfricaArray* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/AF

Penn State University. (2004). *Pennsylvania State Seismic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/PE

Pulliam, J. (2013). *Greater Antilles Seismic Program* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/ZC_2013

Red Sismica Volcan Baru. (2000). *ChiriNet* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/PA

Red Sismológica Nacional de Costa Rica (2021). *The Costa Rica National Seismological Network Catalog during 1975-2020*. https://doi.org/10.15517/TC

Regional Integrated Multi-Hazard Early Warning System (RIMES Thailand). (2008). *Regional Integrated Multi-Hazard Early Warning System* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/RM

Rietbrock, A., Haberland, C., Bataille, K., Lange, D., & Dahm, T. (2004). *TIPTEQ-Temporary seismological network (North) (2004/2005)* [Data set]. GFZ Data Services. https://doi.org/10.14470/MJ7559637482

Royal Observatory of Belgium. (1985). *Belgian Seismic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/BE

Rutgers University. (2013). *Ocean Observatories Initiative* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/OO

San Fernando Royal Naval Observatory (ROA), Universidad Complutense De Madrid (UCM), Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum (GFZ), Universidade De Évora (UEVORA, Portugal), & Institute Scientifique Of Rabat (ISRABAT, Morocco). (1996). *The Western Mediterranean BB seismic Network* [Data set]. Deutsches GeoForschungsZentrum GFZ. https://doi.org/10.14470/JZ581150

Schmandt, B. (2016). *Raton Basin UNM Broadband Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/YX_2016

Scripps Institution of Oceanography. (1986). *Global Seismograph Network - IRIS/IDA* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/II

Seismological Laboratory of University of Basrah. (2014). *Iraqi Seismic Observatory* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/MP

Servicio Geológico Colombiano. (1993). *Red Sismologica Nacional de Colombia* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/CM

Servicio Sismológico Nacional, Instituto de Geofísica, Universidad Nacional Autónoma de México. (n.d.) *Red Sísmica Mexicana*. https://doi.org/10.21766/SSNMX/SN/MX

Stakova-Pursley, J., Aster, R., & Bilek, S. (2009). *Socorro Swarm 2009 RAMP* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/YX_2009

Swiss Seismological Service (SED) At ETH Zurich. (1983). *National Seismic Networks of Switzerland*. ETH Zürich. https://doi.org/10.12686/sed/networks/ch

U.S. Geological Survey. (1931). *United States National Strong-Motion Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/NP

U.S. Geological Survey. (1989). *NetQuakes* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/NQ

U.S. Geological Survey. (2016). *U.S. Geological Survey Networks* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/GM

UC Santa Barbara. (1989). *UC Santa Barbara Engineering Seismology Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/SB

Universidad de Chile. (2012). *Red Sismologica Nacional* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/C1

University of Calgary (U of C Canada). (2018). *EON-ROSE* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/EO

University of Nevada, Reno. (1971). *Nevada Seismic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/NN

University of Nevada, Reno. (1992). *Southern Great Basin Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/SN

University of Oregon. (1990). *Pacific Northwest Seismic Network - University of Oregon* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/UO

University of Ottawa (uOttawa Canada). (2013). *Yukon-Northwest Seismic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/NY

University of Puerto Rico. (1986). *Puerto Rico Seismic Network & Puerto Rico Strong Motion Program* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/PR

University of South Carolina. (1987). *South Carolina Seismic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/CO

University of Utah. (1962). *University of Utah Regional Seismic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/UU

University of Utah. (1983). *Yellowstone National Park Seismograph Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/WY

University of Washington. (1963). *Pacific Northwest Seismic Network - University of Washington* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/UW

USGS Alaska Anchorage. (2000). *USGS Northern Mariana Islands Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/MI

USGS Hawaiian Volcano Observatory (HVO). (1956). *Hawaiian Volcano Observatory Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/HV

USGS Menlo Park. (1966). *USGS Northern California Seismic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/NC

Various Institutions. (1965). *International Miscellaneous Stations* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/vefq-vh75

Vernon, F. (1982). *ANZA Regional Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/AZ

Wagner, L., Beck, S., & Long, M. (2010). *PerU Lithosphere and Slab Experiment* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/ZD_2010

Wiens, D. (2006). *Anatahan and submarine Mariana volcanos* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/ZC_2006

ZAMG - Zentralanstalt für Meterologie und Geodynamik. (1987). *Austrian Seismic Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/OE

REFERENCES CITED

Abadi, M., Ashish, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jozefowicz, R., Jia, Y., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Schuster, M., Monga, R., Moore, S., Murray, D., Olah, C., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., & Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems* [Software]. Zenodo. https://doi.org/10.5281/zenodo.4724125

Abril-Pla, O., Andreani, V., Carroll, C., Dong, L., Fonnesbeck, C. J., Kochurov, M., Kumar, R., Lao, J., Luhmann, C. C., Martin, O. A., Osthege, M., Vieira, R., Wiecki, T., & Zinkov, R. (2023). PyMC: A modern, and comprehensive probabilistic programming framework in Python. *PeerJ Computer Science, 9*, e1516. https://doi.org/10.7717/peerj-cs.1516

Ammon, C. J., Kanamori, H., Lay, T., & Velasco, A. A. (2006). The 17 July 2006 Java tsunami earthquake. *Geophysical Research Letters, 33*(24), L24308. https://doi.org/10.1029/2006GL028005

Barbour, A. J., Agnew, D. C., & Wyatt, F. K. (2015). Coseismic strains on Plate Boundary Observatory borehole strainmeters in Southern California. *Bulletin of the Seismological Society of America, 105*(1), 431–444. https://doi.org/10.1785/0120140199

Barbour, A. J., & Agnew, D. C. (2012). Detection of seismic signals using seismometers and strainmeters. *Bulletin of the Seismological Society of America, 102*(6), 2484–2490. https://doi.org/10.1785/0120110298

Barbour, A. J., & Crowell, B. W. (2017). Dynamic strains for earthquake source characterization. *Seismological Research Letters, 88*(2A), 354–370. https://doi.org/10.1785/0220160155

Benz, H. (2017). Building a national seismic monitoring center: NEIC from 2000 to the present. *Seismological Research Letters, 88*(2B), 457–461. https://doi.org/10.1785/0220170034

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., & Wassermann, J. (2010). ObsPy: A Python toolbox for seismology. *Seismological Research Letters, 81*(3), 530–533. https://doi.org/10.1785/gssrl.81.3.530

Bird, P. (2003). An updated digital model of plate boundaries. *Geochemistry, Geophysics, Geosystems, 4*, 1027–1078. https://doi.org/10.1029/2001GC000252

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*, 2914–2926. https://doi.org/10.1785/0120100111

California Institute of Technology, & United States Geological Survey Pasadena. (1926). Southern California Seismic Network [Data set]. *International Federation of Digital Seismograph Networks*. https://doi.org/10.7914/SN/CI

Canitano, A., Hsu, Y. J., Lee, H. M., Linde, A. T., & Sacks, S. (2015). Near-field strain observations of the October 2013 Ruisui, Taiwan, earthquake: Source parameters and limits of very short-term strain detection. *Earth, Planets and Space, 67*, 125. https://doi.org/10.1186/s40623-015-0284-1

Cole, H. M., Yeck, W. L., & Benz, H. M. (2023). MLAAPDE: A machine learning dataset for determining global earthquake source parameters. *Seismological Research Letters, 94*(5), 2489–2499. https://doi.org/10.1785/0220230021

Cole, H. M., & Yeck, W. L. (2022). Global Earthquake Machine Learning Dataset: Machine Learning Asset Aggregation of the PDE (MLAAPDE) [Data set]. *U.S. Geological Survey*. https://doi.org/10.5066/P96FABIB

Colombelli, S., Zollo, A., Festa, G., & Kanamori, H. (2012). Early magnitude and potential damage zone estimates for the great Mw 9 Tohoku-Oki earthquake. *Geophysical Research Letters, 39*, L22306. https://doi.org/10.1029/2012GL053923

Colombelli, S., Zollo, A., Festa, G., & Picozzi, M. (2014). Evidence for a difference in rupture initiation between small and large earthquakes. *Nature Communications, 5*, 4958. https://doi.org/10.1038/ncomms4958

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. https://doi.org/10.1002/2013gl058391

Crowell, B. W., Schmidt, D. A., Bodin, P., Vidale, J. E., Gomberg, J., Hartog, J. R., Kress, V. C., Melbourne, T. I., Santillan, M., Minson, S. E., & Jamison, D. G. (2016). Demonstration of the Cascadia G-FAST Geodetic Earthquake Early Warning System for the Nisqually, Washington, earthquake. *Seismological Research Letters, 87*(4), 930–943. https://doi.org/10.1785/0220150255

Duputel, Z., Rivera, L., Kanamori, H., Hayes, G. P., Hirshorn, B., & Weinstein, S. (2011). Real-time W phase inversion during the 2011 off the Pacific coast of Tohoku earthquake. *Earth, Planets and Space, 63*(5). https://doi.org/10.5047/eps.2011.05.032

Dybing, S. N., Yeck, W. L., Cole, H. M., & Melgar, D. (2024). Rapid estimation of single-station earthquake magnitudes with machine learning on a global scale. *Bulletin of the Seismological Society of America, 114*(3), 1523–1538. https://doi.org/10.1785/0120230171

Dziewonski, A. M., Chou, T.-A., & Woodhouse, J. H. (1981). Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *Journal of Geophysical Research, 86*(B4), 2825–2852. https://doi.org/10.1029/JB086iB04p02825

Ekström, G., Nettles, M., & Dziewonski, A. M. (2012). The global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes. *Physics of the Earth and Planetary Interiors, 200-201*, 1–9. https://doi.org/10.1016/j.pepi.2012.04.002

Ellsworth, W. L., & Beroza, G. C. (1995). Seismic evidence for an earthquake nucleation phase. *Science, 268*(5212), 851–855. https://doi.org/10.1126/science.268.5212.851

Geller, R. J. (1976). Scaling relations for earthquake source parameters and magnitudes. *Bulletin of the Seismological Society of America, 66*(5), 1501–1523. https://doi.org/10.1785/BSSA0660051501

Geng, J., Pan, Y., Li, X., Guo, J., Liu, J., Chen, X., & Zhang, Y. (2018). Noise characteristics of high-rate multi-GNSS for subdaily crustal deformation monitoring. *Journal of Geophysical Research: Solid Earth, 123*, 1987–2002. https://doi.org/10.1002/2018JB015527

Given, D. D., Allen, R. M., Baltay, A. S., Bodin, P., Cochran, E. S., Creager, K., de Groot, R. M., Gee, L. S., Hauksson, E., Heaton, T. H., Hellweg, M., Murray, J. R., Thomas, V. I., Toomey, D., & Yelin, T. S. (2018). *Revised technical implementation plan for the ShakeAlert system—An earthquake early warning system for the West Coast of the United States* (U.S. Geological Survey Open-File Report 2018–1155). U.S. Geological Survey. https://doi.org/10.3133/ofr20181155

Gladwin, M. T. (1984). High precision multi-component borehole deformation monitoring. *Review of Scientific Instruments, 55*(12), 2011. https://doi.org/10.1063/1.1137704

Goldberg, D. E., Melgar, D., Sahakian, V. J., Thomas, A. M., Xu, X., Crowell, B. W., & Geng, J. (2020). Complex rupture of an immature fault zone: A simultaneous kinematic model of the 2019 Ridgecrest, CA earthquakes. *Geophysical Research Letters, 47*. https://doi.org/10.1029/2019GL086382

Graves, R. W., & Pitarka, A. (2010). Broadband ground-motion simulation using a hybrid approach. *Bulletin of the Seismological Society of America, 100*, 2095–2123. https://doi.org/10.1785/0120100057

Graves, R., & Pitarka, A. (2015). Refinements to the Graves and Pitarka (2010) broadband ground-motion simulation method. *Seismological Research Letters, 86*(1), 75–80. https://doi.org/10.1785/0220140101

Gutenberg, B., & Richter, C. F. (1949). *Seismicity of the Earth and Associated Phenomena*. Princeton University Press.

Hanks, T. C., & Kanamori, H. (1979). A moment magnitude scale. *Journal of Geophysical Research, 84*, 2348–2350. https://doi.org/10.1029/JB084iB05p02348

Hayes, G. P., Myers, E. K., Dewey, J. W., Briggs, R. W., Earle, P. S., Benz, H. M., Smoczyk, G. M., Flamme, H. E., Barnhart, W. D., Gold, R. D., & Furlong, K. P. (2017). Tectonic summaries of magnitude 7 and greater earthquakes from 2000 to 2015 (U.S. Geological Survey Open-File Report No. 2016–1192, 148 pp.). https://doi.org/10.3133/ofr20161192

Hayes, G. P., Rivera, L., & Kanamori, H. (2009). Source inversion of the W-Phase: Real-time implementation and extension to low magnitudes. *Seismological Research Letters, 80*(5), 817–822. https://doi.org/10.1785/gssrl.80.5.817

Heaton, T. H. (1990). Evidence for and implications of self-healing pulses of slip in earthquake rupture. *Physics of the Earth and Planetary Interiors, 64*, 1–20. https://doi.org/10.1016/0031-9201(90)90002-F

Heimann, S., Kriegerowski, M., Isken, M., Cesca, S., Daout, S., Grigoli, F., Juretzek, C., Megies, T., Nooshiri, N., Steinberg, A., Sudhaus, H., Vasyura-Bathke, H., Willey, T., & Dahm, T. (2017). Pyrocko - An open-source seismology toolbox and library (Version 0.3). GFZ Data Services. https://doi.org/10.5880/GFZ.2.1.2017.001

Herman, M. W., & Furlong, K. P. (2021). Triggering an unexpected earthquake in an uncoupled subduction zone. *Science Advances, 7*(13), eabf7590. https://doi.org/10.1126/sciadv.abf7590

Hirshorn, B., Weinstein, S., & Tsuboi, S. (2013). On the application of M_{wp} in the near field and the March 11, 2011 Tohoku earthquake. *Pure and Applied Geophysics, 170*, 975–991. https://doi.org/10.1007/s00024-012-0495-3

Hoshiba, M., & Ozaki, T. (2014). Earthquake early warning and tsunami warning of the Japan Meteorological Agency, and their performance in the 2011 off the Pacific coast of Tohoku earthquake (*M*w 9.0). In F. Wenzel & J. Zschau (Eds.), *Early warning for geological disasters: Advanced technologies in earth sciences* (pp. 1–12). Springer. https://doi.org/10.1007/978-3-642-12233-0_1

Hsu, Y.-J., Chang, Y.-S., Liu, C.-C., Lee, H.-M., Linde, A. T., Sacks, S. I., Kitagawa, G., & Chen, Y.-G. (2015). Revisiting borehole strain, typhoons, and slow earthquakes using quantitative estimates of precipitation-induced strain changes. *Journal of Geophysical Research: Solid Earth, 120*, 4556–4571. https://doi.org/10.1002/2014JB011807

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering, 9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Ishihara, Y., Fukao, Y., Yamada, I., & Aoki, H. (1992). Rising slope of moment rate functions: The 1989 earthquake off the East Coast of Honshu. *Geophysical Research Letters, 19*, 873–876. https://doi.org/10.1029/92GL00367

Ishii, H., Yamauchi, T., Matsumoto, S., Hirata, Y., & Nakao, S. (2002). Development of multi-component borehole instruments for earthquake prediction study: Some observed examples of precursory and co-seismic phenomena relating to earthquake swarms and application of the instrument for rock mechanics in seismogenic process monitoring. In H. Ogasawara, T. Yanagidani, & M. Ando (Eds.), *Seismogenic Process Monitoring: Proceedings of a joint Japan-Poland Symposium on Mining and Experimental Seismology, Kyoto, Japan, November 1999* (1st ed., pp. 365-377). Routledge.

Kanamori, H. (1972). Mechanism of tsunami earthquakes. *Physics of the Earth and Planetary Interiors, 6*(5), 346–359. https://doi.org/10.1016/0031-9201(72)90058-1

Kanamori, H. (1977). The energy release in great earthquakes. *Journal of Geophysical Research: Solid Earth and Planets, 82*(20), 2981–2987. https://doi.org/10.1029/JB082i020p02981

Kanamori, H. (1993). W phase. *Geophysical Research Letters, 20*(16), 1691–1694. https://doi.org/10.1029/93GL01883

Kanamori, H., & Rivera, L. (2008). Source inversion of W phase: Speeding up seismic tsunami warning. *Geophysical Journal International, 175*(1), 222–238. https://doi.org/10.1111/j.1365-246X.2008.03887.x

Kawamoto, S., Ohta, Y., Hiyama, Y., Todoriki, M., Nishimura, T., Furuya, T., Sato, Y., Yahagi, T., & Miyagawa, K. (2017). REGARD: A new GNSS-based real-time finite fault modeling system for GEONET. *Journal of Geophysical Research: Solid Earth, 122*, 1324–1349. https://doi.org/10.1002/2016JB013485

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations*. San Diego, CA. https://doi.org/10.48550/arXiv.1412.6980

Lay, T., Kanamori, H., Ammon, C. J., Koper, K. D., Hutko, A. R., Ye, L., Yue, H., & Rushing, T. M. (2012). Depth-varying rupture properties of subduction zone megathrust faults. *Journal of Geophysical Research: Solid Earth, 117*(B4), B04311. https://doi.org/10.1029/2011JB009133

Lee, W. H. K., & Stewart, S. W. (1981). *Advances in Geophysics, Supplement 2—Principles and applications of microearthquake networks*. Academic Press.

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. https://doi.org/10.1007/s00024-016-1357-1

Licciardi, A., Bletery, Q., Rouet-Leduc, B., Ampuero, J.-P., & Juhel, K. (2022). Instantaneous tracking of earthquake growth with elastogravity signals. *Nature, 606*, 319–324. https://doi.org/10.1038/s41586-022-04672-7

Lin, J.-T., Melgar, D., Sahakian, V. J., Thomas, A. M., & Searcy, J. (2023). Real-time fault tracking and ground motion prediction for large earthquakes with HR-GNSS and deep learning. *Journal of Geophysical Research: Solid Earth, 128*, e2023JB027255. https://doi.org/10.1029/2023JB027255

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*, e2021JB022703. https://doi.org/10.1029/2021JB022703

Lomax, A., Michelini, A., & Jozinović, D. (2019). An investigation of rapid earthquake characterization using single-station waveforms and a convolutional neural network. *Seismological Research Letters, 90*(2A), 517–529. https://doi.org/10.1785/0220180311

Maechling, P. J., Silva, F., Callaghan, S., & Jordan, T. H. (2015). SCEC broadband platform: System architecture and software implementation. *Seismological Research Letters, 86*(1). https://doi.org/10.1785/0220140125

Meier, M.-A., Ampuero, J. P., & Heaton, T. H. (2017). The hidden simplicity of subduction megathrust earthquakes. *Science, 357*(6357), 1277–1281. https://doi.org/10.1126/science.aan5643

Meier, M.-A., Ampuero, J. P., Cochran, E., & Page, M. (2020). Apparent earthquake rupture predictability. *Geophysical Journal International, 225*(1), 657–663. https://doi.org/10.1093/gji/ggaa610

Meier, M.-A., Heaton, T., & Clinton, J. (2016). Evidence for universal earthquake rupture initiation behavior. *Geophysical Research Letters, 43*, 7991–7996. https://doi.org/10.1002/2016GL070081

Melbourne, T. I., Szeliga, W. M., Marcelo Santillan, V., & Scrivner, C. W. (2021). Global navigational satellite system seismic monitoring. *Bulletin of the Seismological Society of America, 111*, 1248–1262. https://doi.org/10.1785/0120200356

Melgar, D., & Hayes, G. P. (2019). Characterizing large earthquakes before rupture is complete. *Science Advances, 5*, eaav2032. https://doi.org/10.1126/sciadv.aav2032

Melgar, D., & Hayes, G. P. (2017). Systematic observations of the slip pulse properties of large earthquake ruptures. *Geophysical Research Letters, 44*, 9691–9698. https://doi.org/10.1002/2017GL074916

Melgar, D., Crowell, B. W., Geng, J., Allen, R. M., Bock, Y., Riquelme, S., Hill, E. M., Protti, M., & Ganas, A. (2015). Earthquake magnitude calculation without saturation from the scaling of peak ground displacement. *Geophysical Research Letters, 42*. https://doi.org/10.1002/2015GL064278

Melgar, D., Crowell, B. W., Melbourne, T. I., Szeliga, W., Santillan, M., & Scrivner, C. (2020). Noise characteristics of operational real-time high-rate GNSS positions in a large aperture network. *Journal of Geophysical Research: Solid Earth, 125*, e2019JB019197. https://doi.org/10.1029/2019JB019197

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. https://doi.org/10.1002/2016JB013314

Mena, B., Mai, P. M., Olsen, K. B., Purvance, M. D., & Brune, J. N. (2010). Hybrid broadband ground-motion simulation using scattering Green's functions: Application to large-magnitude events. *Bulletin of the Seismological Society of America, 100*(5A), 2143–2162. https://doi.org/10.1785/0120080318

Meng, L., Ampuero, J. P., Stock, J., Duputel, Z., Luo, Y., & Tsai, V. C. (2012). Earthquake in a maze: Compressional rupture branching during the 2012 M_w 8.6 Sumatra earthquake. *Science, 337*(6095), 724–726. https://doi.org/10.1126/science.1224030

Mori, J., & Kanamori, H. (1996). Initial rupture of earthquakes in the 1995 Ridgecrest, California sequence. *Geophysical Research Letters, 23*, 2437–2440. https://doi.org/10.1029/96GL02491

Mousavi, S. M., & Beroza, G. C. (2020a). Bayesian-deep-learning estimation of earthquake location from single-station observations. *IEEE Transactions on Geoscience and Remote Sensing, 58*(11), 8211–8224. https://doi.org/10.1109/TGRS.2020.2988770

Mousavi, S. M., & Beroza, G. C. (2020b). A machine-learning approach for earthquake magnitude estimation. *Geophysical Research Letters, 47*(1), e2019GL085976. https://doi.org/10.1029/2019GL085976

Mousavi, S. M., & Beroza, G. C. (2023). Machine learning in earthquake seismology. *Annual Review of Earth and Planetary Sciences, 51*(1), 105–129. https://doi.org/10.1146/annurev-earth-071822-100323

Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020). Earthquake transformer—An attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications, 11*, 3952. https://doi.org/10.1038/s41467-020-17591-w

Mousavi, S. M., Sheng, Y., Zhu, W., & Beroza, G. C. (2019). STanford EArthquake Dataset (STEAD): A global data set of seismic signals for AI. *IEEE Access, 7*, 179464–179476. https://doi.org/10.1109/ACCESS.2019.2947848

Münchmeyer, J., Bindi, D., Leser, U., & Tilmann, F. (2021). Earthquake magnitude and location estimation from real time seismic waveforms with a transformer network. *Geophysical Journal International, 226*(2), 1086–1104. https://doi.org/10.1093/gji/ggab139

Murray, J. R., Crowell, B. W., Murray, M. H., Ulberg, C. W., McGuire, J. J., Aranha, M. A., & Hagerty, M. T. (2023). Incorporation of real-time earthquake magnitudes estimated via peak ground displacement scaling in the ShakeAlert earthquake early warning system. *Bulletin of the Seismological Society of America, 113*(3), 1286–1310. https://doi.org/10.1785/0120220181

Nazareth, J. J., & Hauksson, E. (2004). The seismogenic thickness of the southern California crust. *Bulletin of the Seismological Society of America, 94*(3), 940–960. https://doi.org/10.1785/0120020129

Perol, T., Gharbi, M., & Denolle, M. (2018). Convolutional neural network for earthquake detection and location. *Science Advances, 4*(2), e1700578. https://doi.org/10.1126/sciadv.1700578

Petersson, N. A., & Sjogreen, B. (2017). SW4, version 2.01 [Software]. Computational Infrastructure of Geodynamics. https://doi.org/10.5281/zenodo.1063644

Ringler, A. T., Anthony, R. E., Aster, R. C., Ammon, C. J., Arrowsmith, S., Benz, H., Ebeling, C., Frassetto, A., Kim, W.-Y., Koelemeijer, P., Lau, H. C. P., Lekić, V., Montagner, J. P., Richards, P. G., Schaff, D. P., Vallée, M., & Yeck, W. (2022). Achievements and prospects of global broadband seismographic networks after 30 years of continuous geophysical observations. *Reviews of Geophysics, 60*(3), e2021RG000749. https://doi.org/10.1029/2021RG000749

Ristea, N.-C., & Radoi, A. (2021). Complex neural networks for estimating epicentral distance, depth, and magnitude of seismic waves. *IEEE Transactions on Geoscience and Remote Sensing, 19*, 7502305. https://doi.org/10.1109/LGRS.2021.3059422

Rodriguez, I. V., & Wuestefeld, A. (2020). Strain microseismics: Radiation patterns, synthetics, and moment tensor resolvability with distributed acoustic sensing in isotropic media. *Geophysics, 85*(3), KS101–KS114. https://doi.org/10.1190/geo2019-0373.1

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 234–241). https://doi.org/10.1007/978-3-319-24574-4_28

Ross, Z. E., Meier, M.-A., Hauksson, E., & Heaton, T. H. (2018). Generalized seismic phase detection with deep learning. *Bulletin of the Seismological Society of America, 108*(5A), 2894–2901. https://doi.org/10.1785/0120180080

Ruhl, C. J., Melgar, D., Geng, J., Goldberg, D. E., Crowell, B. W., Allen, R. M., Bock, Y., Barrientos, S., Riquelme, S., Baez, J. C., Cabral-Cano, E., Pérez-Campos, X., Hill, E. M., Protti, M., Ganas, A., Ruiz, M., Mothes, P., Jarrín, P., Nocquet, J.-M., Avouac, J.-P., & D'Anastasio, E. (2018). A global database of strong-motion displacement GNSS recordings and an example application to PGD scaling. *Seismological Research Letters, 90*(1), 271–279. https://doi.org/10.1785/0220180177

Rydelek, P., & Horiuchi, S. (2006). Earth science: Is earthquake rupture deterministic? *Nature, 442*(7100), E5–E6. https://doi.org/10.1038/nature04963

Sacks, I. S., Suyehiro, S., Yamagishi, Y., & Evertson, D. W. (1971). Sacks-Evertson strainmeter, its installation in Japan and some preliminary results concerning strain steps. *Papers in Meteorology and Geophysics, 22*, 195–207. https://doi.org/10.2467/mripapers1950.22.3-4_195

Santellanes, S. R., Melgar, D., Crowell, B. W., & Lin, J.-T. (2022). Potential megathrust co-seismic slip during the 2020 Sand Point, Alaska strike-slip earthquake. *ESS Open Archive*. https://doi.org/10.1002/essoar.10508002.2

Santillan, V. M., Melbourne, T. I., Szeliga, W. M., & Scrivner, C. W. (2013). A fast-convergence stream editor for real-time precise point positioning. In *Annual Meeting of the American Geophysical Union*, San Francisco, California, 9–13 December 2013, G53B–0930.

SCEDC (2013). Southern California Earthquake Center [Data set]. Caltech. https://doi.org/10.7909/C3WD3xH1

Singh, S. K., Pacheco, J., Ordaz, M., & Kostoglodov, V. (2000). Source time function and duration of Mexican earthquakes. *Bulletin of the Seismological Society of America, 90*(2), 468–482. https://doi.org/10.1785/0119990081

Sipkin, S. A., Person, W. J., & Presgrave, B. W. (2000). Earthquake bulletins and catalogs at the USGS NEIC. *IRIS Newsletter, 2000*(1), 2–4. Retrieved from http://ftp.iris.edu/news/IRISnewsletter/volume2000no1/page-02-04.htm (last accessed November 2023).

Thomas, A. M., Inbal, A., Searcy, J., Shelly, D. R., & Bürgmann, R. (2021). Identification of low-frequency earthquakes on the San Andreas fault with deep learning. *Geophysical Research Letters, 48*, e2021GL093157. https://doi.org/10.1029/2021GL093157

Thomas, A., Melgar, D., Dybing, S. N., & Searcy, J. R. (2023). Deep learning for denoising High-Rate Global Navigation Satellite System data. *Seismica, 2*(1). https://doi.org/10.26443/seismica.v2i1.240

Trabant, C., Hutko, A. R., Bahavar, M., Karstens, R., Ahern, T., & Aster, R. (2012). Data products at the IRIS DMC: Stepping stones for research and other applications. *Seismological Research Letters, 83*(5), 846–854. https://doi.org/10.1785/0220120032

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*, 4642–4653. https://doi.org/10.1029/2018JB017093

Tsuboi, S., Whitmore, P. M., & Sokolowski, T. J. (1999). Application of M_{wp} to deep and teleseismic earthquakes. *Bulletin of the Seismological Society of America, 89*(5), 1345–1351. https://doi.org/10.1785/BSSA0890051345

Uieda, L., Tian, D., Leong, W. J., Schlitzer, W., Grund, M., Jones, M., Fröhlich, Y., Toney, L., Yao, J., Magen, Y., Jing-Hui, T., Materna, K., Belem, A., Newton, T., Anant, A., Ziebarth, M., Quinn, J., & Wessel, P. (2023). PyGMT: A Python interface for the Generic Mapping Tools (v0.9.0). Zenodo. https://doi.org/10.5281/zenodo.3781524

van den Ende, M. P. A., & Ampuero, J.-P. (2020). Automated seismic source characterization using deep graph neural networks. *Geophysical Research Letters, 47*(17), e2020GL088690. https://doi.org/10.1029/2020GL088690

Wells, D. L., & Coppersmith, K. J. (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of the Seismological Society of America, 84*(4), 974–1002. https://doi.org/10.1785/BSSA0840040974

Yue, H., Lay, T., & Koper, K. D. (2012). En échelon and orthogonal fault ruptures of the 11 April 2012 great intraplate earthquakes. *Nature, 490*, 245–249. https://doi.org/10.1038/nature11492

Yue, H., Lay, T., Rivera, L., Bai, Y., Yamazaki, Y., Cheung, K. F., Hill, E. M., Sieh, K., Kongko, W., & Muhari, A. (2014). Rupture process of the 2010 M_w7.8 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. https://doi.org/10.1002/2014JB011082

Zhang, X., Zhang, J., Yuan, C., Liu, S., Chen, Z., & Li, W. (2020). Locating induced earthquakes with a network of seismic stations in Oklahoma via a deep learning method. *Scientific Reports, 10*, 1941. https://doi.org/10.1038/s41598-020-58908-5

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*, 619–627. https://doi.org/10.1046/j.1365-246X.2002.01610.x

Zhu, W., & Beroza, G. C. (2019). PhaseNet: A deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International, 216*(1), 261–273. https://doi.org/10.1093/gji/ggy423

Zhu, W., Mousavi, S. M., & Beroza, G. C. (2020). Seismic signal augmentation to improve generalization of deep neural networks. *Advances in Geophysics, 61*, 151–177. https://doi.org/10.1016/bs.agph.2020.07.003