UNIVERSITY OF CALIFORNIA, SAN DIEGO

# Deviant Earthquakes: Data-driven Constraints on the Variability in Earthquake Source Properties and Seismic Hazard

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

in

Earth Sciences

by

Daniel Taylor Trugman

Committee in charge:

Professor Peter M. Shearer, Chair
Professor Adrian A. Borsa
Professor Yuri Fialko
Professor Peter Gerstoft
Professor Tara C. Hutchinson

2017

The dissertation of Daniel Taylor Trugman is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2017

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

# ACKNOWLEDGEMENTS

I view the world from a probabilistic perspective, and as a general rule avoid dealing in absolutes. But if there is one thing I am absolutely certain of, it is that this thesis never would have been possible without all of the people that have helped me along the way. While I can never truly express the depth of my gratitude with words on a printed piece of paper, I'll give it my best shot below.

First, I want to thank my wonderful advisor Peter Shearer and my doctoral committee – Adrian Borsa, Yuri Fialko, Peter Gerstoft, and Tara Hutchinson – for their guidance over the past few years. Peter S. is the platonic ideal of a graduate advisor. He is kind and caring, yet brilliant and always excited about to uncover new things about the Earth. His door is always open, and I made it a point to try to chat with him almost every day, whether about research or life in general, because he always seems to have something insightful to say. I've learned so much from him in my time at Scripps, and hope to make him proud with my research moving forward. Adrian has been an incredible mentor and empathetic friend. He has a unique perspective on life and on science, and his "two cents" is usually worth its weight in gold. Yuri is gifted and enthusiastic teacher with the uncanny ability to translate the difficult aspects of earthquake mechanics into concepts that poor students like me can actually identify with and understand. Peter G. is a remarkable and creative scientist, and has taught me a diverse set of skills in data analysis, parameter estimation, and machine learning that will serve me well in the

years to come. And while I know Tara the least well of my committee members, her keen advice and suggestions during committee meetings have really helped me understand the practical implications of my research.

In addition to my committee, I also want to thank the other teachers, mentors, and collaborators, both here at SIO and elsewhere, that have helped shape me as the scientist I am today. Research projects with Dave Sandwell, Yuri Fialko, and Adrian Borsa have expanded my technical skillset and capacity to think critically about important problems in the earth sciences. Collaborations with Sara Dougherty, Elizabeth Cochran, Rachel Hatch, Kenneth Smith, Rachel Abercrombie, Egill Hauksson, Zach Ross, and Keith Koper have helped augment these skills and provided unique research opportunities that are documented in the following chapters. Classes with Dave Sandwell, Yuri Fialko, Cathy Constable, Peter Shearer, Dave Stegman, Guy Masters, Duncan Agnew, Peter Gerstoft, and Donna Blackman have laid a solid foundational knowledge base in geophysics from which I hope to build in the years to come. Guy Masters and Steve Constable were nice enough to let me to serve as a TA for their Introduction to Geophysics class, which was a rewarding experience that I won't soon forget. Likewise, I treasured the week I spent with Debi Kilb teaching a remarkably precocious set of tweens a unified theory of Music and Earthquakes at the Sally Ride Science Camp. This thesis would also not have been possible without the dedicated behind-the-scenes work of Gilbert Bretado, Maria Rivas, Megan Strachan, and all of the SIO and IGPP administrators I've interacted with over the years. Outside of Scripps, I

am thankful to my undergraduate teachers and advisors at Stanford University, especially Eric Dunham and Greg Beroza, who instilled a love for geophysics in general and seismology in particular that drives me to this day. Paul Johnson then took me under his wing at Los Alamos National Laboratory and help foster this nascent passion for research, and I am incredibly excited to return there to work alongside Paul, Andrew Delorey, and Robert Guyer.

My peers and friends have played an equally important role supporting me during my graduate career. While the female quartet of the 2014 IGPP cohort may now have gone their separate ways, I'll never forget the good times and long hours spent in the Keller with Julia, May, Mara, and Georgie during our first year. Yuval is still around, and his boundless optimism shines like a beacon from SDSU that is visible from my windowless office on the 2nd floor of the Munk building. Wes, you are my finest friend at SIO (and favorite Texan), as is perhaps best evidenced by your willingness to regularly climb with me at 6am (so long as donuts were involved), despite never really having any interest in climbing before meeting me. Likewise, I always looked forward to seeing the shining faces of my Munk 220 officemates Georgie, Shunguo, and Weixing, and I admire their resilience in putting up with me each and every day. IGPP is such a friendly place that it makes me nervous to try to list off all of the other great friends I've made here, but here it goes. Katia, Wenyuan, Eric, Soli, John, Susheel, Adrian, Jesse, Dara, Matt, Dallas, Kang, Wei, Shawn, Janine, Junle, Adina, Drake, Maya, Chappy, Valeria, Bei, Ishita, Te-Yang, Nick: you guys are the best, and you are truly what

makes IGPP such an incredible place to be. And a special acknowledgement goes to my San Diego climbing buddies – Jess, Boyle, Paul, and Jens – who I'm sure will never read this but whose encouragement on and off belay have meant a lot to me these past few years.

Last but certainly not least, my family. I would never be where I am today without you. You have always been there for me, in the best of times and the worst. You have never lost confidence in my ability to push through the rough stretches, even when I have. I call home as often as possible, not because I think you are worried, but because hearing from you brightens my day. Mom, you are the wisest and most compassionate person I know, and I still can't fathom how you can balance such a high-powered and stressful job with being an outstanding mom, scientist, and athlete. Dad, I've learn so much from you over the years, from how to ski to how to do quantum mechanics. But the most important thing I've learned from you is how to think and to think critically, and you are the finest scientist I know in this regard. Anna, you are the rare diamond whose brilliance is matched only by your passion to make the world a better place. You have certainly made me a better person, and Earth's future would be secure if only their were more people like you. Curt, you are the newest member of my family, but I couldn't imagine a better partner for my sister. You are without a doubt the coolest person I know, and I look up to you like the older brother I never had. I love you all, and if there is one thing I could aspire to in this life, it would be to make you proud.

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program (NSFGRFP) under grant number DGE-1144086. Additional support was provided by the Paul G. Silver Young Scholar Research Enhancement Award, the Southern California Earthquake Center (SCEC) under grant numbers 14031, 14073, and 16020, the United States Geological Survey - National Earthquake Hazards Reduction Program (USGS-NEHRP) under grant number G15AS00037-2016-0072, and the NSF EarthScope program (EAR-1147427). I also wish to acknowledge support for my graduate studies from the San Diego Chapter of the Achievement Reward for Career Scientists (ARCS) Foundation.

Chapter 2, in full, is a reformatted version of material as it appears in Geophysical Research Letters: Trugman, D. T., A. A. Borsa, and D. T. Sandwell (2014), Did stresses from the Cerro Prieto Geothermal Field influence the El Mayor-Cucapah rupture sequence?, *Geophysical Research Letters*, 41(24), 8767–8774, doi: 10.1002/2014GL061959. I was the primary investigator and author of this paper.

Chapter 3, in full, is a reformatted version of material as it appears in Journal of Geophysical – Solid Earth: Trugman, D. T., P. M. Shearer, A. A. Borsa and Y. Fialko (2016), A comparison of long-term changes in seismicity at The Geysers, Salton Sea, and Coso geothermal fields, *Journal of Geophysical Research: Solid*

*Earth*, 121(1), 225–247, doi:10.1002/2015JB012510. I was the primary investigator and author of this paper.

Chapter 4, in full, is a reformatted version of material as it appears in Seismological Research Letters: Trugman, D. T. and P. M. Shearer (2017), GrowClust: A hierarchical clustering algorithm for relative earthquake relocation, with application to the Spanish Springs and Sheldon, Nevada, earthquake sequences, *Seismological Research Letters*, 88(2A), doi:10.1785/0220160188. I was the primary investigator and author of this paper.

Chapter 5, in full, is a reformatted version of material as it appears in Journal of Geophysical – Solid Earth: Trugman, D. T. and P. M. Shearer (2017), Application of an improved spectral decomposition method to examine earthquake source scaling in southern California, *Journal of Geophysical Research: Solid Earth*, 122(4), doi:10.1002/2017JB013971. I was the primary investigator and author of this paper.

Chapter 6, in full, is a reformatted version of material as it appears in Journal of Geophysical – Solid Earth: Trugman, D. T. and P. M. Shearer (2017), Source spectral properties of small-to-moderate earthquakes in southern Kansas, *Journal of Geophysical Research: Solid Earth*, 122(10), doi:10.1002/2017JB014649. I was the primary investigator and author of this paper.

Chapter 7, in full, is a reformatted version of material as it appears in Bulletin of the Seismological Society of America: Trugman, D. T. and P. M. Shearer (2017), Strong correlation between stress drop and peak ground acceleration for

recent seismicity in the San Francisco Bay Area, *Bulletin of the Seismological Society of America*, submitted. I was the primary investigator and author of this paper.

## VITA

| | |
|---|---|
| 2017 | Ph. D. in Earth Sciences, University of California, San Diego |
| 2015 | M. S. in Geophysics, University of California, San Diego |
| 2013 | B. S. with Honors and Distinction in Geophysics, Stanford University |

## PUBLICATIONS

**Trugman, D. T.** and P. M. Shearer (2017), Strong correlation between stress drop and peak ground acceleration for recent seismicity in the San Francisco Bay Area, *Bulletin of the Seismological Society of America*, submitted.

**Trugman, D. T.** and P. M. Shearer (2017), Source spectral properties of small-to-moderate earthquakes in southern Kansas, *Journal of Geophysical Research: Solid Earth* 122(10), doi:10.1002/2017JB014649.

**Trugman, D. T.** and P. M. Shearer (2017), Application of an improved spectral decomposition method to examine earthquake source scaling in southern California, *Journal of Geophysical Research: Solid Earth* 122(4), doi:10.1002/2017JB013971.

**Trugman, D. T.** and P. M. Shearer (2017), GrowClust: A hierarchical clustering algorithm for relative earthquake relocation, with application to the Spanish Springs and Sheldon, Nevada, earthquake sequences, *Seismological Research Letters* 88(2A), doi:10.1785/0220160188.

**Trugman, D. T.**, P. M. Shearer, A. A. Borsa and Y. Fialko (2016), A comparison of long-term changes in seismicity at The Geysers, Salton Sea, and Coso geothermal fields, *Journal of Geophysical Research: Solid Earth* 121(1), 225–247, doi:10.1002/2015JB012510.

Daub, E. G., **D. T. Trugman**, and P. A. Johnson (2015), Statistical tests on clustered global earthquake synthetic data sets, *Journal of Geophysical Research: Solid Earth* 120(8), 5693–5716, doi:10.1002/2014JB011777.

**Trugman, D. T.**, C. Wu, R. A. Guyer, and P. A. Johnson (2015), Synchronous low frequency earthquakes and implications for deep San Andreas fault slip, *Earth and Planetary Science Letters* 424, 132–139, doi:10.1016/j.epsl.2015.05.029.

Wu, C., R. Guyer, D. Shelly, **D. T. Trugman**, W. Frank, J. Gomberg, and P. Johnson (2015), Spatial-temporal variation of low-frequency earthquake bursts near Parkfield, California, *Geophysical Journal International* 202(2), 914–919, doi:10.1093/gji/ggv194.

**Trugman, D. T.**, A. A. Borsa, and D. T. Sandwell (2014), Did stresses from the Cerro Prieto Geothermal Field influence the El Mayor-Cucapah rupture sequence?, *Geophysical Research Letters* 41(24), 8767–8774, doi:10.1002/2014GL061959.

**Trugman, D. T.**, and E. M. Dunham (2014), A 2D pseudodynamic rupture model generator for earthquakes on geometrically complex faults, *Bulletin of the Seismological Society of America* 104(1), 95–112, doi:10.1785/0120130138.

**Trugman, D. T.**, E. G. Daub, R. A. Guyer, and P. A. Johnson (2013), Modeling dynamic triggering of tectonic tremor using a brittle-ductile friction model, *Geophysical Research Letters* 40(19), 5075–5079, doi:10.1002/grl.50981.

Johnson, P. A., B. Ferdowsi, B. M. Kaproth, M. Scuderi, M. Griffa, J. Carmeliet, R. A. Guyer, P.-Y. Le Bas, **D. T. Trugman**, and C. Marone (2013), Acoustic emission and microslip precursors to stick-slip failure in sheared granular material, *Geophysical Research Letters* 40(21), 5627–5631, doi:10.1002/2013GL057848.

Perkins, B. G., H. Y. Hwang, N. K. Grady, L. Yan, **D. T. Trugman**, Q. Jia, H. T. Chen, A. J. Taylor, and K. A. Nelson (2013), Nonlinear ultrafast dynamics of high temperature YBa2Cu3O7delta superconductors probed with THz pump/THz probe spectroscopy, *EPJ Web of Conferences* 41, 03010, doi:10.1051/epjconf/20134103010.

Grady, N. K., B. G. Perkins, H. Y. Hwang, N. C. Brandt, D. Torchinsky, R. Singh, L. Yan, **D. T. Trugman**, S. A. Trugman, Q. X. Jia, A. J. Taylor, K. A. Nelson, and H. T. Chen (2013), Nonlinear high-temperature superconducting terahertz metamaterials, *New Journal of Physics* 15(10), 105016, doi:10.1088/1367-2630/15/10/105016.

Budiman, A. S., N. Li, Q. Wei, J. K. Baldwin, J. Xiong, H. Luo, **D. T. Trugman**, Q. X. Jia, N. Tamura, M. Kunz, K. Chen, A Misra (2011), Growth and structural characterization of epitaxial Cu/Nb multilayers, *Thin Solid Films* 519(13), 4137-4143, doi:10.1016/j.tsf.2010.12.077.

Xiong, J., V. Matias, H. Wang, J. Y. Zhai, B. Maiorov, **D. T. Trugman**, B. W. Tao, Y. R. Li, Q. X. Jia (2010), Much simplified ion-beam assisted deposition-TiN template for high-performance coated conductors, *Journal of Applied Physics* 108(8), 083903-083903-4, doi:10.1063/1.3499270.

ABSTRACT OF THE DISSERTATION

**Deviant Earthquakes: Data-driven Constraints on the Variability in Earthquake Source Properties and Seismic Hazard**

by

Daniel Taylor Trugman

Doctor of Philosophy in Earth Sciences

University of California, San Diego, 2017

Professor Peter M. Shearer, Chair

The complexity of the earthquake rupture process makes earthquakes inherently unpredictable. Seismic hazard forecasts often presume that the rate of earthquake occurrence can be adequately modeled as a space-time homogenenous or stationary Poisson process and that the relation between the dynamical source properties of small and large earthquakes obey self-similar scaling relations. While these simplified models provide useful approximations and encapsulate the first-order statistical features of the historical seismic record, they are inconsistent with

the complexity underlying earthquake occurrence and can lead to misleading assessments of seismic hazard when applied in practice. The six principle chapters of this thesis explore the extent to which the behavior of real earthquakes deviates from these simplified models, and the implications that the observed deviations have for our understanding of earthquake rupture processes and seismic hazard.

Chapter 1 provides a brief thematic overview and introduction to the scope of this thesis. Chapter 2 examines the complexity of the 2010 M7.2 El Mayor-Cucapah earthquake, focusing on the relation between its unexpected and unprecedented occurrence and anthropogenic stresses from the nearby Cerro Prieto Geothermal Field. Chapter 3 compares long-term changes in seismicity within California's three largest geothermal fields in an effort to characterize the relative influence of natural and anthropogenic stress transients on local seismic hazard. Chapter 4 describes a hybrid, hierarchical clustering algorithm that can be used to relocate earthquakes using waveform cross-correlation, and applies the new algorithm to study the spatiotemporal evolution of two recent seismic swarms in western Nevada. Chapter 5 describes a new spectral decomposition technique that can be used to analyze the dynamic source properties of large datasets of earthquakes, and applies this approach to revisit the question of self-similar scaling of southern California seismicity. Chapter 6 builds upon these results and applies the same spectral decomposition technique to examine the source properties of several thousand recent earthquakes in southern Kansas that are likely human-induced by massive oil and gas operations in the region. Chapter 7 studies the

connection between source spectral properties and earthquake hazard, focusing on spatial variations in dynamic stress drop and its influence on ground motion amplitudes. Finally, Chapter 8 provides a summary of the key findings of and relations between these studies, and outlines potential avenues of future research.

# Chapter 1

# Introduction

## 1.1 Motivation

*When is the Big One coming?* It is a question I have been asked hundreds of times over my years studying as a seismologist. It is a pertinent question here in Earthquake Country, where the San Andreas Fault knifes with a northwesterly strike through the center of southern California, poised like the sword of Damocles over major population centers in Los Angeles and San Diego. It is a relevant question, given that the average recurrence period between major earthquakes on the southern San Andreas Fault is of order 100 years and the last such event occurred in 1857 (*Fialko and Simons*, 2000; *Field et al.*, 2014; *Smith-Konter and Sandwell*, 2009; *Hauksson et al.*, 2017), with only a handful of people around to record it. It is a question that I, like so many of my peers, can only wish to know the answer.

The ability to accurately predict earthquakes is the ultimate goal for many seismologists, myself included. Yet realizing this goal has proven difficult despite many decades of scientific effort, and is likely to remain elusive for many years hence (*Geller et al.*, 1997). On a fundamental level, earthquakes occurrence is driven by stress accumulation within Earth's crust (*Reid*, 1911; *Scholz*, 2002). The concentration of major earthquakes near the boundaries of tectonic plates is therefore no accident, and in general it is more straightforward to reliably predict *where* earthquakes are likely to occur than *when*. The reason is that the nucleation process that precedes an earthquake is highly complex and chaotic, sensitive to minute details in the stress field at seismogenic depth for which we have poor observational constraints (*Rice and Ruina*, 1983; *Rice*, 2006; *Ruina*, 1983; *Brodsky and Kanamori*, 2001; *Noda and Lapusta*, 2013).

While the possibility of accurately *predicting* earthquakes is remote, improving our ability to reliably *forecast* earthquakes is more tractable. The distinction between a deterministic prediction and a forecast, which is inherently probabilistic, is subtle but essential (*Vere-Jones*, 2003; *Jordan and Jones*, 2010; *Jordan et al.*, 2011). A prediction would imply that the physics of earthquake rupture and the relevant physical conditions of each fault zone are both perfectly known, while a probabilistic forecast allows for a degree of scientific uncertainty. The forecasting of earthquake hazard is a model-driven task, relying on physical or statistical constraints from studies of previous earthquakes. As our understanding of the relevant physics of earthquake nucleation, rupture, and triggering improves, so will

our ability to accurately characterize seismic hazard.

Most seismic hazard forecasts used in recent years are based on simple statistical models of earthquake occurrence that incorporate only the most well-established features of the earthquake record to date (*Baker*, 2013; *Mulargia et al.*, 2017). Often, earthquake occurrence is treated as a stochastic point process (*Daley and Vere-Jones*, 2003), where each event is characterized only in terms of its location and magnitude. In this framework, the rate of earthquake occurrence is often modeled statistically as a homogeneous Poisson process that is stationary in both space and time (*Snyder and Miller*, 1991). Meanwhile the distribution of earthquake magnitudes is canonically assumed to follow a Gutenberg-Richter power law (*Gutenberg and Richter*, 1944) with a slope of order 1 that is again constant in space and time. This modeling paradigm can be useful in the sense that it provides a computationally feasible means for probabilistic hazard assessment that not only accounts for the inherent randomness in earthquake occurrence, but imposes it.

Unfortunately, the real world is rarely so simple. Earthquakes cluster in space and time, often in an unpredictable but decidedly nonstationary manner (*van der Elst and Brodsky*, 2010; *Hainzl et al.*, 2013; *Kumazawa and Ogata*, 2013). The physical processes associated with earthquake triggering and stress transfer can cause large changes in the rate of earthquake occurrence (*Stein*, 1999; *Lin and Stein*, 2004; *Mallman and Parsons*, 2008; *Parsons and Velasco*, 2009; *Page et al.*, 2016). In recent years, more sophisticated hazard forecasts have begun to

encapsulate the first-order influence of earthquake clustering and stress-transfer within fault systems (*Field et al.*, 2014; *Gerstenberger et al.*, 2016). The improvement in the resulting hazard forecasts demonstrate the potential for advances in scientific understanding to have a quantifiable effect on matters of general public interest. However there remains a great deal of epistemic uncertainty in these forecasts (*Budnitz et al.*, 1997; *Anderson and Brune*, 1999; *Mulargia et al.*, 2017), particularly with respect to how point process models, which may be statistically viable for smaller earthquakes, can be applied to larger earthquakes in which the finite size of the earthquake source can no longer be ignored (*Field et al.*, 2017).

Moreover, earthquakes are complex phenomena with real variations in source properties related to dynamic rupture processes that are difficult to encapsulate within simple statistical models (*Meier et al.*, 2017). Classically, dynamic source properties have been viewed as self-similar with respect to earthquake size, with large earthquakes simply scaled-up versions of small earthquakes (*Aki*, 1967). Even the most sophisticated modern hazard forecasts presume self-similarity, as observational constraints on the behavior of large earthquakes within a given fault system are thankfully rare. In recent years, however, detailed analyses of earthquake source properties have challenged the assumption of self-similarity (*Mayeda and Walter*, 1996; *Izutani and Kanamori*, 2001; *Mori et al.*, 2003; *Mayeda et al.*, 2005; *Takahashi et al.*, 2005; *Mayeda et al.*, 2007; *Calderoni et al.*, 2013; *Pacor et al.*, 2016; *Lin et al.*, 2016; *Poli and Prieto*, 2016), though much controversy still remains (*Choy and Boatwright*, 1995; *Ide and Beroza*, 2001; *Ide*, 2003; *Prieto*, 2004;

*Allmann and Shearer*, 2009; *Baltay et al.*, 2010, 2011; *Abercrombie et al.*, 2017). It seems plausible that the inherent complexity of earthquake nucleation and rupture may cause a breakdown in self-similar scaling between small and large earthquakes (*Kanamori and Rivera*, 2004; *Abercrombie and Rice*, 2005; *Noda and Lapusta*, 2010), but more research is needed to test the extent to which this hypothesis is true. Likewise, the connection between ground motion amplitudes and dynamic source properties (*Boore*, 1983; *Baltay et al.*, 2013; *Lior and Ziv*, 2017) motivates further research into characterizing variations in earthquake source dynamics across different fault systems and tectonic settings.

The various chapters of this thesis explore the true variability in earthquake source properties and spatiotemporal patterns of occurrence. By providing robust data-driven constraints, my objective is to better understand the degree to which real earthquakes deviate from simplified, stationary and self-similar statistical models. The influence of human activity will play a prominent role in this thesis, as the proliferation of earthquakes triggered by anthropogenic stresses from geothermal energy and oil production have caused rapid and widespread changes in seismicity from previous historical norms (*Convertito et al.*, 2012; *Ellsworth*, 2013; *Ellsworth et al.*, 2015; *McNamara et al.*, 2015; *Petersen et al.*, 2016, 2017). But just as importantly, transient or nonstationary variations in earthquake occurrence and source properties can occur for entirely natural reasons (*Allmann and Shearer*, 2009; *Oth and Kaiser*, 2014; *Abercrombie et al.*, 2017; *Meier et al.*, 2017). Real earthquakes are not simply points in magnitude-space-time, but real

and complex phenomena with distinct spectral signatures. The outliers matter, as the rogue, deviant earthquakes that occur unexpectedly or with unusual source properties often prove to be the most hazardous, and can provide new insights into the physical processes underlying earthquake occurrence.

## 1.2 Thesis Structure

This thesis is structured as follows. The six principal Chapters (2 through 7) were originally written for individual publication and hence can be read in isolation. They are, however, listed in a logical order such that the ideas introduced in preceding chapters are built upon in the subsequent chapters. Chapter 2 (*Trugman et al.*, 2014) and Chapter 3 (*Trugman et al.*, 2016) examine the influence of anthropogenic stresses from geothermal energy production on seismicity within Baja and southern California. Chapter 4 (*Trugman and Shearer*, 2017a) describes a method for providing high-resolution earthquake relocations that is then applied to examine the spatiotemporal evolution of earthquake sequences in Nevada. The final three chapters focus on dynamic source properties derived from $P$-wave spectral estimates. In Chapter 5 (*Trugman and Shearer*, 2017b), I develop an improved spectral inversion technique to estimate dynamic source parameters for large waveform datasets, and I apply this technique to study deviations from self-similar scaling in southern California. Building on these results, in Chapter 6 (*Trugman et al.*, 2017) I study the source properties of human-induced earth-

quakes in southern Kansas, comparing the spectral characteristics of these events to those of naturally occurring earthquakes in southern California. In Chapter 7 (*Trugman and Shearer*, 2017c), I explore the relationship between dynamic stress drop and measured ground motion amplitudes, focusing on the implications that spatial variations in source properties have for seismic hazard. A brief overview of each of these individual chapters is provided below for the reader's convenience.

Chapter 2, in full, is a reformatted version of material as it appears in Geophysical Research Letters: Trugman, D. T., A. A. Borsa, and D. T. Sandwell (2014), Did stresses from the Cerro Prieto Geothermal Field influence the El Mayor-Cucapah rupture sequence?, *Geophysical Research Letters*, 41(24), 8767–8774, doi: 10.1002/2014GL061959. I was the primary investigator and author of this paper. Here we investigate the relation between geothermal energy production at the Cerro Prieto Geothermal field in northern Baja California, Mexico and the nearby 2010 M7.2 El Mayor-Cucapah earthquake. We use interferometric radar (InSAR) data from the ALOS satellite to measure surface subsidence in the vicinity of the geothermal field. We then use the measured surface displacement to invert for a model of volumetric contraction at depth within the geothermal reservoir. We find that Coulomb stresses generated from net fluid extraction at the geothermal field pushed the major fault planes that comprise the El Mayor-Cucapah earthquake rupture sequence toward failure. Anthropogenic activity may therefore have contributed to the nucleation of the El Mayor-Cucapah earthquake, whose occurrence was unexpected given that the rupture plane is poorly oriented

for failure within the regional tectonic stress field.

Chapter 3, in full, is a reformatted version of material as it appears in Journal of Geophysical – Solid Earth: Trugman, D. T., P. M. Shearer, A. A. Borsa and Y. Fialko (2016), A comparison of long-term changes in seismicity at The Geysers, Salton Sea, and Coso geothermal fields, *Journal of Geophysical Research – Solid Earth*, 121(1), 225–247, doi:10.1002/2015JB012510. I was the primary investigator and author of this paper. Here we analyze changes in the occurrence patterns of seismicity over decadal timescales within California's three largest geothermal fields: (1) The Geysers, located to the northeast of San Francisco, (2) Salton Sea Geothermal Field, located near California's southern border with Mexico and adjacent to the southern tip of the San Andreas fault, and (3) Coso Geothermal Field, located within the structurally complex Eastern California Shear Zone. Each geothermal field exists within a distinct tectonic setting and has experienced a different history of energy production. Although we observe measurable increases in background seismicity rate and changes to the earthquake magnitude and depth distributions within each field, both the detailed features of these changes and the relative influence of anthropogenic stresses vary substantially between the three fields. The differing responses suggest that changes in seismicity in California's geothermal fields are controlled by a complex interplay of stresses from regional tectonic and local anthropogenic sources.

Chapter 4, in full, is a reformatted version of material as it appears in Seismological Research Letters: Trugman, D. T. and P. M. Shearer (2017), Grow-

Clust: A hierarchical clustering algorithm for relative earthquake relocation, with application to the Spanish Springs and Sheldon, Nevada, earthquake sequences, *Seismological Research Letters*, 88(2A), doi:10.1785/0220160188. I was the primary investigator and author of this paper. Here we develop a new algorithm called GrowClust that can be used for the relocation of earthquake hypocenters based on differential travel time data obtained through waveform cross-correlation. The algorithm, which we present as a publicly available integrated software package, is robust to data outliers and is both computationally efficient and numerically stable when applied to large-scale datasets. We apply GrowClust to examine the spatiotemporal evolution of two recent earthquake swarms that occurred in western Nevada: the 2012–2015 Spanish Springs and the 2014–2016 Sheldon sequences. The results demonstrate the potential for high-precision relocations to uncover the detailed spatial and temporal features of unusual earthquakes sequences that give insight into physical mechanisms that drive them.

Chapter 5, in full, is a reformatted version of material as it appears in Journal of Geophysical – Solid Earth: Trugman, D. T. and P. M. Shearer (2017), Application of an improved spectral decomposition method to examine earthquake source scaling in southern California, *Journal of Geophysical Research – Solid Earth*, 122(4), doi:10.1002/2017JB013971. I was the primary investigator and author of this paper. Here we describe an improved $P$-wave spectral decomposition technique that can be used to analyze the dynamic source properties of large datasets of earthquakes. We apply this technique to revisit the question of

the self-similar scaling of earthquakes within five regions of dense, recent (2002–2016) seismicity in southern California: the Yuha Desert, the Trifurcation Zone of the San Jacinto Fault, and the Big Bear, Landers, and Hector Mine regions of the Mojave Desert. We observe an increase in median stress drop with seismic moment within each of the five analyzed regions, in direct contradiction to the classical self-similar, constant stress drop scaling that forms the basis of routine seismic hazard assessment. Moreover, we find evidence for nonstationary spatial variations in source properties on both regional and local length scales that provide important observational constraints for the distribution of crustal stresses and earthquake hazard in southern California.

Chapter 6, in full, is a reformatted version of material as it appears in Journal of Geophysical – Solid Earth: Trugman, D. T. and P. M. Shearer (2017), Source spectral properties of small-to-moderate earthquakes in southern Kansas, *Journal of Geophysical Research – Solid Earth*, 122(10), doi:10.1002/2017JB014649. I was the primary investigator and author of this paper. Here we apply the improved spectral decomposition approach described in Chapter 5 to analyze the source properties of earthquakes within southern Kansas from 2014–2016. This new dataset comprises more than 4000 events within a region of heavy oil and gas production, and most of these events are thought to be induced by stresses associated with industrial wastewater disposal. We find that earthquakes in southern Kansas exhibit lower median stress drop values compared to naturally occurring earthquakes in southern California. We also observe coherent spatial and temporal changes in

median stress drop within southern Kansas that are not easily explained by localized fluid injection but may be related to more widespread, regional wastewater disposal. These results have significant implications for our understanding of the rupture processes and the seismic hazard associated with human-induced seismicity in the central United States.

Chapter 7, in full, is a reformatted version of material as it appears in Bulletin of the Seismological Society of America: Trugman, D. T. and P. M. Shearer (2017), Strong correlation between stress drop and peak ground acceleration for recent seismicity in the San Francisco Bay Area, *Bulletin of the Seismological Society of America*, submitted. I was the primary investigator and author of this paper. Here we investigate the relation between the source spectral properties and the measured ground motion amplitudes of earthquakes occurring in the vicinity of the San Francisco Bay Area, CA. We apply the $P$-wave spectral decomposition method described in Chapter 5 to estimate dynamic stress drop for the more than 5000 events occurring in the Bay Area from 2002–2016. We then measure horizontal peak ground acceleration (PGA) for these same events recorded at the same set of stations. We develop a novel, data-driven approach based on a machine learning algorithm known as a Random Forest to model PGA as a function of magnitude, hypocentral distance, and site, and use this model to isolate the residual PGA associated with each event. We observe a strong correlation between dynamic stress drop and the residual PGA of each event, with the events with higher-than-expected PGA associated with higher values of stress drop. We find

that mainshock events have systematically higher median stress drop and residual PGA values than aftershocks of equivalent magnitude. We further observe coherent spatial variations in the distribution of dynamic stress drop that can be used as the basis for improved future seismic hazard assessments specific to the Bay Area.

Lastly, Chapter 8 provides a synthesis of the major finding presented in Chapters 2–7. The limitations and the broader implications of each study are discussed. The principal results presented in this thesis bring into focus several important unanswered questions concerning earthquake rupture, and I conclude by outlining potential future research directions that could be used to address these questions.

# References

Abercrombie, R. E., and J. R. Rice (2005), Can observations of earthquake scaling constrain slip weakening?, *Geophysical Journal International*, *162*(2), 406–424, doi:10.1111/j.1365-246X.2005.02579.x.

Abercrombie, R. E., S. Bannister, J. Ristau, and D. Doser (2017), Variability of earthquake stress drop in a subduction setting, the Hikurangi Margin, New Zealand, *Geophysical Journal International*, *208*(1), 306–320, doi:10.1093/gji/ggw393.

Aki, K. (1967), Scaling law of seismic spectrum, *Journal of Geophysical Research*, *72*(4), 1217–1231, doi:10.1029/JZ072i004p01217.

Allmann, B. P., and P. M. Shearer (2009), Global variations of stress drop for moderate to large earthquakes, *Journal of Geophysical Research: Solid Earth*, *114*(B1), B01,310, doi:10.1029/2008JB005821.

Anderson, J. G., and J. N. Brune (1999), Probabilistic Seismic Hazard Analysis without the Ergodic Assumption, *Seismological Research Letters*, *70*(1), 19–28, doi:10.1785/gssrl.70.1.19.

Baker, J. W. (2013), An introduction to probabilistic seismic hazard analysis, *White Paper Version 2.0*, pp. 1–79.

Baltay, A., G. Prieto, and G. C. Beroza (2010), Radiated seismic energy from coda measurements and no scaling in apparent stress with seismic moment, *Journal of Geophysical Research*, *115*(B8), doi:10.1029/2009JB006736.

Baltay, A., S. Ide, G. Prieto, and G. Beroza (2011), Variability in earthquake stress drop and apparent stress, *Geophysical Research Letters*, *38*(L06303), doi:10.1029/2011GL046698.

Baltay, A. S., T. C. Hanks, and G. C. Beroza (2013), Stable Stress-Drop Measurements and their Variability: Implications for Ground-Motion Prediction, *Bulletin of the Seismological Society of America*, *103*(1), 211–222, doi:10.1785/0120120161.

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

Brodsky, E. E., and H. Kanamori (2001), Elastohydrodynamic lubrication of faults, *Journal of Geophysical Research: Solid Earth*, *106*(B8), 16,357–16,374, doi:10.1029/2001JB000430.

Budnitz, R. J., G. Apostolakis, and D. M. Boore (1997), Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts, *Tech. Rep. NUREG/CR–6372-Vol.1; UCRL-ID–122160*, Nuclear Regulatory Commission, Washington, DC (United States). Div. of Engineering Technology; Lawrence Livermore National Lab., CA (United States); Electric Power Research Inst., Palo Alto, CA (United States); USDOE, Washington, DC (United States), 10.2172/479072.

Calderoni, G., A. Rovelli, and S. K. Singh (2013), Stress drop and source scaling of the 2009 April L'Aquila earthquakes, *Geophysical Journal International*, *192*(1), 260–274, doi:10.1093/gji/ggs011.

Choy, G. L., and J. L. Boatwright (1995), Global patterns of radiated seismic energy and apparent stress, *Journal of Geophysical Research: Solid Earth*, *100*(B9), 18,205–18,228, doi:10.1029/95JB01969.

Convertito, V., N. Maercklin, N. Sharma, and A. Zollo (2012), From Induced Seismicity to Direct Time-Dependent Seismic Hazard, *Bulletin of the Seismological Society of America*, *102*(6), 2563–2573, doi:10.1785/0120120036.

Daley, D. J., and D. Vere-Jones (2003), *An Introduction to the Theory of Point Processes*, 2nd ed., Springer, New York.

Ellsworth, W., A. Llenos, A. McGarr, A. Michael, J. Rubinstein, C. Mueller, M. Petersen, and E. Calais (2015), Increasing seismicity in the U. S. midcontinent: Implications for earthquake hazard, *The Leading Edge*, *34*(6), 618–626, doi:10.1190/tle34060618.1.

Ellsworth, W. L. (2013), Injection-Induced Earthquakes, *Science*, *341*(6142), 1225,942–1225,942, doi:10.1126/science.1225942.

Fialko, Y., and M. Simons (2000), Deformation and seismicity in the Coso geothermal area, Inyo County, California: Observations and modeling using satellite radar interferometry, *Journal of Geophysical Research*, *105*(B9), 21,781, doi:10.1029/2000JB900169.

Field, E. H., R. J. Arrowsmith, G. P. Biasi, P. Bird, T. E. Dawson, K. R. Felzer, D. D. Jackson, K. M. Johnson, T. H. Jordan, C. Madden, A. J. Michael, K. R. Milner, M. T. Page, T. Parsons, P. M. Powers, B. E. Shaw, W. R. Thatcher, R. J. Weldon, and Y. Zeng (2014), Uniform California Earthquake Rupture Forecast, Version 3 (UCERF3)–The Time-Independent Model, *Bulletin of the Seismological Society of America*, *104*(3), 1122–1180, doi:10.1785/0120130164.

Field, E. H., K. R. Milner, J. L. Hardebeck, M. T. Page, N. v. d. Elst, T. H. Jordan, A. J. Michael, B. E. Shaw, and M. J. Werner (2017), A Spatiotemporal Clustering Model for the Third Uniform California Earthquake Rupture Forecast

(UCERF3-ETAS): Toward an Operational Earthquake Forecast, *Bulletin of the Seismological Society of America*, *107*(3), doi:10.1785/0120160173.

Geller, R. J., D. D. Jackson, Y. Y. Kagan, and F. Mulargia (1997), Earthquakes Cannot Be Predicted, *Science*, *275*(5306), 1616–1616, doi:10.1126/science.275.5306.1616.

Gerstenberger, M. C., D. A. Rhoades, and G. H. McVerry (2016), A Hybrid Time-Dependent Probabilistic Seismic-Hazard Model for Canterbury, New Zealand, *Seismological Research Letters*, doi:10.1785/0220160084.

Gutenberg, B., and C. F. Richter (1944), Frequency of earthquakes in California, *Bulletin of the Seismological Society of America*, *34*(4), 185–188.

Hainzl, S., O. Zakharova, and D. Marsan (2013), Impact of Aseismic Transients on the Estimation of Aftershock Productivity Parameters, *Bulletin of the Seismological Society of America*, *103*(3), 1723–1732, doi:10.1785/0120120247.

Hauksson, E., M.-A. Meier, Z. E. Ross, and L. M. Jones (2017), Evolution of seismicity near the southernmost terminus of the San Andreas Fault: Implications of recent earthquake clusters for earthquake risk in southern California, *Geophysical Research Letters*, p. 2016GL072026, doi:10.1002/2016GL072026.

Ide, S. (2003), Apparent break in earthquake scaling due to path and site effects on deep borehole recordings, *Journal of Geophysical Research*, *108*(B5), doi:10.1029/2001JB001617.

Ide, S., and G. C. Beroza (2001), Does apparent stress vary with earthquake size?, *Geophysical Research Letters*, *28*(17), 3349–3352, doi:10.1029/2001GL013106.

Izutani, Y., and H. Kanamori (2001), Scale-dependence of seismic energy-to-moment ratio for strike-slip earthquakes in Japan, *Geophysical Research Letters*, *28*(20), 4007–4010, doi:10.1029/2001GL013402.

Jordan, T., Y.-T. Chen, P. Gasparini, R. Madariaga, I. Main, W. Marzocchi, G. Papadopoulos, K. Yamaoka, and J. Zschau (2011), Operational Earthquake Forecasting: State of Knowledge and Guidelines for Implementation., doi:10.4401/ag-5350.

Jordan, T. H., and L. M. Jones (2010), Operational Earthquake Forecasting: Some Thoughts on Why and How, *Seismological Research Letters*, *81*(4), 571–574, doi:10.1785/gssrl.81.4.571.

Kanamori, H., and L. Rivera (2004), Static and Dynamic Scaling Relations for Earthquakes and Their Implications for Rupture Speed and Stress Drop, *Bulletin of the Seismological Society of America*, *94*(1), 314–319, doi:10.1785/0120030159.

Kumazawa, T., and Y. Ogata (2013), Quantitative description of induced seismic activity before and after the 2011 Tohoku-Oki earthquake by nonstationary ETAS models, *Journal of Geophysical Research: Solid Earth*, *118*(12), 6165–6182, doi:10.1002/2013JB010259.

Lin, J., and R. S. Stein (2004), Stress triggering in thrust and subduction earthquakes and stress interaction between the southern San Andreas and nearby thrust and strike-slip faults, *Journal of Geophysical Research: Solid Earth*, *109*(B2), doi:10.1029/2003JB002607.

Lin, Y.-Y., K.-F. Ma, H. Kanamori, T.-R. A. Song, N. Lapusta, and V. C. Tsai (2016), Evidence for non-self-similarity of microearthquakes recorded at a Taiwan borehole seismometer array, *Geophysical Journal International*, *206*(2), 757–773, doi:10.1093/gji/ggw172.

Lior, I., and A. Ziv (2017), The Relation between Ground Acceleration and Earthquake Source Parameters: Theory and Observations, *Bulletin of the Seismological Society of America*, *107*(2), doi:10.1785/0120160251.

Mallman, E. P., and T. Parsons (2008), A global search for stress shadows, *Journal of Geophysical Research*, *113*(B12), doi:10.1029/2007JB005336.

Mayeda, K., and W. R. Walter (1996), Moment, energy, stress drop, and source spectra of western United States earthquakes from regional coda envelopes, *Journal of Geophysical Research: Solid Earth*, *101*(B5), 11,195–11,208, doi:10.1029/96JB00112.

Mayeda, K., R. Gok, W. R. Walter, and A. Hofstetter (2005), Evidence for nonconstant energy/moment scaling from coda-derived source spectra, *Geophysical Research Letters*, *32*(10), L10,306, doi:10.1029/2005GL022405.

Mayeda, K., L. Malagnini, and W. R. Walter (2007), A new spectral ratio method using narrow band coda envelopes: Evidence for non-self-similarity in the Hector Mine sequence, *Geophysical Research Letters*, *34*(11), L11,303, doi:10.1029/2007GL030041.

McNamara, D. E., J. L. Rubinstein, E. Myers, G. Smoczyk, H. M. Benz, R. A. Williams, G. Hayes, D. Wilson, R. Herrmann, N. D. McMahon, R. C. Aster, E. Bergman, A. Holland, and P. Earle (2015), Efforts to monitor and characterize the recent increasing seismicity in central Oklahoma, *The Leading Edge*, *34*(6), 628–639, doi:10.1190/tle34060628.1.

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

Mori, J., R. E. Abercrombie, and H. Kanamori (2003), Stress drops and radiated energies of aftershocks of the 1994 Northridge, California, earthquake, *Journal of Geophysical Research: Solid Earth*, *108*(B11), 2545, doi:10.1029/2001JB000474.

Mulargia, F., P. B. Stark, and R. J. Geller (2017), Why is Probabilistic Seismic Hazard Analysis (PSHA) still used?, *Physics of the Earth and Planetary Interiors*, *264*, 63–75, doi:10.1016/j.pepi.2016.12.002.

Noda, H., and N. Lapusta (2010), Three-dimensional earthquake sequence simulations with evolving temperature and pore pressure due to shear heating: Effect of heterogeneous hydraulic diffusivity, *Journal of Geophysical Research*, *115*(B12), doi:10.1029/2010JB007780.

Noda, H., and N. Lapusta (2013), Stable creeping fault segments can become destructive as a result of dynamic weakening, *Nature*, *493*(7433), 518–521, doi:10.1038/nature11703.

Oth, A., and A. E. Kaiser (2014), Stress Release and Source Scaling of the 2010-2011 Canterbury, New Zealand Earthquake Sequence from Spectral Inversion of Ground Motion Data, *Pure and Applied Geophysics*, *171*(10), 2767–2782, doi:10.1007/s00024-013-0751-1.

Pacor, F., D. Spallarossa, A. Oth, L. Luzi, R. Puglia, L. Cantore, A. Mercuri, M. D'Amico, and D. Bindi (2016), Spectral models for ground motion prediction in the L'Aquila region (central Italy): evidence for stress-drop dependence on magnitude and depth, *Geophysical Journal International*, *204*(2), 697–718, doi:10.1093/gji/ggv448.

Page, M. T., N. van der Elst, J. Hardebeck, K. Felzer, and A. J. Michael (2016), Three Ingredients for Improved Global Aftershock Forecasts: Tectonic Region, Time-Dependent Catalog Incompleteness, and Intersequence Variability, *Bulletin of the Seismological Society of America*, *106*(5), 2290–2301, doi:10.1785/0120160073.

Parsons, T., and A. A. Velasco (2009), On near-source earthquake triggering, *Journal of Geophysical Research*, *114*(B10), doi:10.1029/2008JB006277.

Petersen, M. D., C. S. Mueller, M. P. Moschetti, S. M. Hoover, A. L. Llenos, W. L. Ellsworth, A. J. Michael, J. L. Rubinstein, A. F. McGarr, and K. S. Rukstales (2016), Seismic-Hazard Forecast for 2016 Including Induced and Natural Earthquakes in the Central and Eastern United States, *Seismological Research Letters*, *87*(6), 1327–1341, doi:10.1785/0220160072.

Petersen, M. D., C. S. Mueller, M. P. Moschetti, S. M. Hoover, A. M. Shumway, D. E. McNamara, R. A. Williams, A. L. Llenos, W. L. Ellsworth, A. J. Michael, J. L. Rubinstein, A. F. McGarr, and K. S. Rukstales (2017), 2017 One-Year

Seismic-Hazard Forecast for the Central and Eastern United States from Induced and Natural Earthquakes, *Seismological Research Letters*, *88*(3), 772–783, doi: 10.1785/0220170005.

Poli, P., and G. A. Prieto (2016), Global rupture parameters for deep and intermediate-depth earthquakes, *Journal of Geophysical Research: Solid Earth*, *121*(12), 8871–8887, doi:10.1002/2016JB013521.

Prieto, G. A. (2004), Earthquake source scaling and self-similarity estimation from stacking P and S spectra, *Journal of Geophysical Research*, *109*(B8), doi:10. 1029/2004JB003084.

Reid, H. F. (1911), *The elastic-rebound theory of earthquakes*, vol. 6, 19, University of California Publications, Bulletin of the Department of Geology, Berkeley, CA.

Rice, J. R. (2006), Heating and weakening of faults during earthquake slip, *Journal of Geophysical Research*, *111*(B5), doi:10.1029/2005JB004006.

Rice, J. R., and A. L. Ruina (1983), Stability of Steady Frictional Slipping, *Journal of Applied Mechanics*, *50*(2), 343, doi:10.1115/1.3167042.

Ruina, A. (1983), Slip instability and state variable friction laws, *Journal of Geophysical Research*, *88*(B12), 10,359, doi:10.1029/JB088iB12p10359.

Scholz, C. H. (2002), *The Mechanics of Earthquakes and Faulting*, Cambridge University Press.

Smith-Konter, B., and D. Sandwell (2009), Stress evolution of the San Andreas fault system: Recurrence interval versus locking depth, *Geophysical Research Letters*, *36*(13).

Snyder, D. L., and M. I. Miller (1991), *Random Point Processes in Time and Space*, Springer New York, New York, NY.

Stein, R. S. (1999), The role of stress transfer in earthquake occurrence, *Nature*, *402*(6762), 605–609, doi:10.1038/45144.

Takahashi, T., H. Sato, M. Ohtake, and K. Obara (2005), Scale Dependence of Apparent Stress for Earthquakes along the Subducting Pacific Plate in Northeastern Honshu, Japan, *Bulletin of the Seismological Society of America*, *95*(4), 1334–1345, doi:10.1785/0120040075.

Trugman, D. T., and P. M. Shearer (2017a), GrowClust: A Hierarchical Clustering Algorithm for Relative Earthquake Relocation, with Application to the Spanish Springs and Sheldon, Nevada, Earthquake Sequences, *Seismological Research Letters*, *88*(2A), 379–391, doi:10.1785/0220160188.

Trugman, D. T., and P. M. Shearer (2017b), Application of an improved spectral decomposition method to examine earthquake source scaling in southern California, *Journal of Geophysical Research: Solid Earth*, *122*(4), 2017JB013,971, doi:10.1002/2017JB013971.

Trugman, D. T., and P. M. Shearer (2017c), Strong correlation between stress drop and peak ground acceleration for recent seismicity in the San Francisco Bay Area, *submitted to Bulletin of the Seismological Society of America*.

Trugman, D. T., A. A. Borsa, and D. T. Sandwell (2014), Did stresses from the Cerro Prieto Geothermal Field influence the El Mayor-Cucapah rupture sequence?, *Geophysical Research Letters*, *41*(24), 8767–8774, doi:10.1002/2014GL061959.

Trugman, D. T., P. M. Shearer, A. A. Borsa, and Y. Fialko (2016), A comparison of long-term changes in seismicity at The Geysers, Salton Sea, and Coso geothermal fields, *Journal of Geophysical Research: Solid Earth*, *121*(1), 225–247, doi:10.1002/2015JB012510.

Trugman, D. T., S. L. Dougherty, E. S. Cochran, and P. M. Shearer (2017), Source spectral properties of small-to-moderate earthquakes in southern Kansas, *Journal of Geophysical Research: Solid Earth*, *122*(10), JGRB52,338, doi:10.1002/2017JB014649.

van der Elst, N. J., and E. E. Brodsky (2010), Connecting near-field and far-field earthquake triggering to dynamic strain, *Journal of Geophysical Research*, *115*(B7), doi:10.1029/2009JB006681.

Vere-Jones, D. (2003), Probabilities and Information Gain for Earthquake Forecasting, in *Selected Papers From Volume 30 of Vychislitel'naya Seysmologiya*, edited by D. K. Chowdhury, J.-C. D. Bremaecker, K. Lashgari, E. Nyland, R. Odom, M. Sen, M. M. Vishik, V. I. Keilis-Borok, A. L. Levshin, G. M. Molchan, and B. M. Naimark, pp. 104–114, American Geophysical Union, dOI:10.1029/CS005p0104.

# Chapter 2

# Did stresses from the Cerro Prieto Geothermal Field influence the El Mayor-Cucapah rupture sequence?

## Abstract

The $M_W$7.2 El Mayor-Cucapah (EMC) earthquake ruptured a complex fault system in northern Baja California that was previously considered inactive. The Cerro Prieto Geothermal Field (CPGF), site of the world's second-largest geothermal power plant, is located approximately 15 km to the northeast of the EMC hypocenter. We investigate whether anthropogenic fluid extraction at the CPGF

caused a significant perturbation to the stress field in the EMC rupture zone. We use ALOS-satellite InSAR data to develop a laterally heterogeneous model of fluid extraction at the CPGF and estimate that this extraction generates positive Coulomb stressing rates of order 15 kPa/year near the EMC hypocenter, a value which exceeds the local tectonic stressing rate. Although we cannot definitively conclude that production at the CPGF triggered the EMC earthquake, its influence on the local stress field is substantial, and should not be neglected in local seismic hazard assessments.

## 2.1 Introduction

The southern edge of the Pacific-North America plate boundary cuts directly through the Valle de Mexicali in northwestern Baja California. This region is part of a broad zone of tectonic deformation characterized by subparallel dextral faults that connect the spreading centers of the Gulf of California to the south with the San Andreas and San Jacinto Faults to the north. Much of the right-lateral plate motion in the region is accommodated on the nearby Cerro Prieto and Imperial faults, each slipping at an estimated rate of 40 mm/year (*Bennett et al.*, 1996; *Atwater and Stock*, 1998). The 2010 El Mayor-Cucapah (EMC) earthquake itself occurred to the west of these main plate-boundary faults (Figure 2.1), but to the east of the Laguna Salada Fault, which hosted an $M_W$7.1 earthquake in 1892 (*Fletcher and Spelz*, 2009). Instead, the EMC event ruptured the Borrego

and Pescadores Faults in the Sierra Cucapah to the north of its hypocenter, and the Indiviso fault to its south (*Wei et al.*, 2011). The northern faults had been mapped prior to EMC event (*Barnard*, 1968; *Fletcher and Spelz*, 2009) but were presumed to slip at much lower rates than the adjacent plate boundary faults, and show little evidence of Holocene faulting (*Fletcher et al.*, 2014). The Indiviso Fault was previously unmapped, as it is buried beneath Colorado River Delta sediments.

The EMC earthquake was the largest in the region since the 1992 Landers earthquake (*Sieh et al.*, 1993). It resulted in the deaths of four people, injured hundreds of others, and triggered widespread landslides and liquefaction. Yet, the faults responsible for this damage were previously thought to contribute negligibly to the regional seismic hazard (*Bennett et al.*, 1996; *Fletcher et al.*, 2014). In this study, we consider whether the extraction of fluids and heat related to energy production at the nearby Cerro Prieto Geothermal Field (CPGF) could have played a role in stressing these faults and initiating the EMC event, or sustaining its prolonged rupture.

The CPGF is the second-largest geothermal power plant in the world (*Ocampo-Díaz et al.*, 2005; *CFE.*, 2006; *Sarychikhina et al.*, 2011), and is located approximately 15 km to the northeast of the EMC hypocenter, in the extensional stepover between the Cerro Prieto and Imperial Faults (Figure 2.1). Geothermal energy production, and the associated fluid extraction, has caused surface subsidence in the vicinity of the CPGF at previously reported rates of 10-15 cm/year (*Glowacka et al.*, 2005, 2010; *Sarychikhina et al.*, 2011).

The extraction process, in turn, perturbs the regional crustal stress field, and has been investigated for its possible role in triggering large, nearby earthquakes in the recent past (*Glowacka and Nava*, 1996). The central aim of this study is to estimate the stressing rate on the faults that ruptured in the EMC earthquake due to extraction at the CPGF. We begin by using InSAR data from ALOS satellite tracks recorded from 2006-2009 to constrain the magnitude and spatial extent of anthropogenic subsidence at the CPGF. We then model the complex pattern of fluid extraction and recharge at the CPGF as a distribution of Mogi-source spherical pressure cavities, using the observed InSAR surface deformation data in a regularized inversion scheme to estimate the source intensities. With this extraction model in hand, we compute the Coulomb stressing rate on the sequence of faults that ruptured during the EMC event, and compare this stressing rate to the stressing rate caused by deep, interseismic fault slip on the major regional faults. We find that extraction at the CPGF imparts positive Coulomb stresses of order +15 kPa/year on the faults involved in the EMC rupture sequence. Conversely, tectonic loading from other regional faults imparts a negative Coulomb stress of around -8 kPa/year in this same location.

## 2.2 Subsidence Measurements at the Cerro Prieto Geothermal Field

The Cerro Prieto Geothermal Field (CPGF) is located in the Mexicali Valley of northern Baja California. It is one of four major geothermal fields in the Salton Trough tectonic province (the others being the Salton Sea, East Mesa, and Heber geothermal fields, all in the United States). The CPGF has been in continuous operation by the Comisión Federal de Electricidad since 1973, and with a current installed capacity of 820 MW, is the worlds second largest geothermal energy source (*Ocampo-Díaz et al.*, 2005; *Sarychikhina et al.*, 2011).

Energy production at the CPGF requires extraction of hot water and steam from production wells with a mean depth of 2.7 km (*Gutiérrez-Negrín et al.*, 2010). In 2008, a total fluid volume of $6.3 \times 10^7$ $m^3$ was extracted at the CPGF, a value typical of the published extraction rates since 1994 (*Glowacka et al.*, 2005; *Gutiérrez-Negrín et al.*, 2010). While approximately 30% of the extracted fluid is reinjected on site (*Gutiérrez-Negrín et al.*, 2010), and nearly twice this amount is naturally recharged from the surrounding aquifers (*Glowacka et al.*, 2005), there is a net loss of fluids in the production zone that causes the earthś surface to subside. Surface subsidence at the CPGF is well-established: first with leveling surveys dating back to 1977, and more recently using Interferometric Synthetic Aperture Radar (InSAR) techniques (*Carnec and Fabriol*, 1999; *Hanssen*, 2001; *Glowacka et al.*, 2005). These studies estimated a vertical subsidence rate at Cerro Prieto of

12 cm/year, 90-95% of which is directly caused by fluid extraction at the CPGF (*Sarychikhina et al.*, 2011) .

We augment these previous observations with Advanced Land Observing Satellite (ALOS) InSAR data acquired from 2006.8-2009.1 to study in detail the surface subsidence immediately preceding the EMC event. After initial processing of 23 InSAR images of the study region (see Supplementary Text for details), we stacked the two highest quality, phase-unwrapped (*Chen and Zebker*, 2001) interferograms to estimate the line-of-site (LOS) surface velocity field near the Cerro Prieto Geothermal Field (Figure 2.2a).

We estimate a maximum vertical subsidence rate of 14 cm/year in the CPGF production zone, comparable to the findings of previous studies (*Sarychikhina et al.*, 2011). This subsidence is almost entirely anthropogenic in origin, as local tectonics (i.e., deformation due to the position of the CPGF in an extensional stepover) can account for at most 5% of the observed subsidence rate (Supplementary Figure 2.S1, see *Glowacka et al.*, 2005 for a similar assessment). The spatial distribution of subsidence is characterized by two prominent lobes of deformation: one primary lobe situated directly above the energy production site, and a secondary lobe offset to the northeast that has been attributed to recharge of the main production area from a deep aquifer adjacent to the Imperial Fault (*Glowacka et al.*, 2005). While this secondary lobe is still prominent in our ALOS InSAR data, it is somewhat diminished in magnitude relative to observations in earlier studies, perhaps indicating a decrease in recharge rate with time.

## 2.3 Fluid Extraction Model

The pronounced surface subsidence at the Cerro Prieto Geothermal Field is a product of a volume change at depth that is primarily caused by a net extraction of fluids, as thermoelastic effects are thought to contribute negligibly (*Mossop and Segall*, 1997). In early studies linking surface deformation to geothermal fluid extraction (*Mossop and Segall*, 1997; *Carnec and Fabriol*, 1999; *Mossop and Segall*, 1999), volumetric contraction was typically modeled as one or more point pressure cavity, or "Mogi" sources (*Mogi*, 1958). More recent studies (*Sarychikhina*, 2003; *Glowacka et al.*, 2005; *Sarychikhina et al.*, 2011) have used site-specific geologic constraints like fault boundaries to construct deformation models based on the superposition of large, rectangular, tensile (closing) cracks, while others (*Vasco et al.*, 2002) subdivide the model domain into volume elements extending laterally and vertically to allow the model to assume any arbitrary shape. We apply a hybrid approach in which we model the surface subsidence at the CPGF as the superposition of finely-spaced Mogi pressure sources embedded in an elastic half-space, applying constraints from independent information about the location and extent of the source region. For this study, we position the center of the source distribution in the production zone of the CPGF (-115.20° E, 32.1° N) and at the mean extraction depth of 2.7 km (*Lippmann et al.*, 1991; *Gutiérrez-Negrín et al.*, 2010). We discretize the source distribution using a single horizontal layer with a grid-spacing of 0.5 km in both lateral directions for a total of 436 evenly-spaced

sources. Although our approach does not allow us to distinguish between volumetric changes at different depths, it does allow the model to represent the horizontal variations in source intensity required to capture the complex spatial patterns of extraction, injection, and recharge beneath the CPGF. For a single Mogi source at position ($x_0$, $y_0$, $z_0 < 0$) within an elastic halfspace (Poissons ratio $\nu$), the displacement vector ($u_x$, $u_y$ , $u_z$) at surface position ($x$, $y$, $z = 0$), can be written as

$$(u_x, u_y, u_z) = \frac{(1-\nu)\Delta V}{\pi R^3}(x - x_0, y - y_0, -z_0) \tag{2.1}$$

where $R = \sqrt{(x-x_0)^2 + (y-y_0)^2 + (-z_0)^2}$ is the distance from the source to the surface observation point and $\Delta V$ is the associated volume change (i.e., the source intensity) (*Segall*, 2010). Subsidence requires $\Delta V < 0$. The observed surface deformation is the superposition of the deformation caused by each individual Mogi source in the distribution. We use the InSAR-derived LOS surface velocity field to perform a regularized, least-squares inversion for the source volumes (Figure 2.2b, see Supplementary Text for details). Our preferred model's displacement field (Figure 2.2c) provides a 69% variance reduction (Figure 2.2d) on the observed InSAR data. We estimate a net rate of volumetric contraction of -9.0x$10^6$ $m^3$/year. This estimated rate is insensitive to the details of the modeling assumptions (e.g., grid spacing and assumed depth) and inversion approach (e.g., choice of smoothing parameter), and is consistent both with the independent estimates of previous studies (*Sarychikhina et al.*, 2011) and with available production data (*Gutiérrez-*

*Negrín et al.*, 2010).

## 2.4 Anthropogenic Stressing Rates Near the EMC Hypocenter

The EMC rupture sequence was complex (*Wei et al.*, 2011), with slip likely initiating in a normal faulting subevent on a shallowly dipping fault plane (labeled F1 in Figures 2.1 and 2.3) striking almost due N. After a brief pause in moment release, the rupture jumped from the F1 fault plane onto NW-striking right-lateral faults (see Figures 2.1 and 2.3), and ruptured bilaterally northward to the California border (fault plane F2) and southward to Gulf of California (fault plane F3). The $M_W$7.2 EMC event was preceded by a vigorous foreshock sequence near the F1 hypocenter, culminating in a left-lateral, $M_W$4.3 event less than 24 hours prior to the mainshock (*Hauksson et al.*, 2011). We use our model of CPGF fluid extraction to compute the anthropogenic stressing rate in the El Mayor-Cucapah rupture zone. Given a fault orientation and location, one can resolve the local stress tensor on that fault plane to compute the change in Coulomb stress:

$$\Delta\sigma_C = \Delta\tau + \mu'\Delta\sigma_N \tag{2.2}$$

where $\Delta\tau$ is the change in shear stress in the direction of slip, $\Delta\sigma_N$ is the change in normal stress (assumed positive in extension), and $\mu'$ is the effective coefficient of friction. Positive Coulomb stress changes on a given fault are presumed to

push that fault toward failure (*King et al.*, 1994; *Stein*, 1999). To compute the Coulomb stresses associated with our fluid extraction model, we first generalize the Fourier-domain approach of (*Steketee*, 1958) (see Supplementary Text for details) to obtain the 3D stress tensor due to an arbitrary distribution of radial point sources embedded in an elastic halfspace. We then decompose the local stress tensor into shear and normal components for the fault geometries and slip directions of the subevents that comprise the EMC rupture sequence, and compute the local Coulomb stressing rate. We assume typical values for the shear modulus (32 GPa) and effective coefficient of friction of (0.4) (*Lin and Stein*, 2004). From our model, we estimate that energy production at the CPGF generates a Coulomb stressing rate of $\sim +11$ kPa/year at the $M_W$4.3 foreshock hypocenter (Figure 2.3a). This stressing rate is considerable, since Coulomb stress changes as low as 10 kPa have been known to trigger earthquake activity (*Stein*, 1999). While the foreshock was itself a relatively small event, it occurred in close proximity in both space and time to the hypocenter of the initial F1 mainshock subevent, so its role in initiating the mainshock event sequence cannot be discounted. Extraction at the CPGF also caused positive Coulomb stress changes on the faults that ruptured in the EMC mainshock. We estimate stressing rates of $+12$ kPa/year at the hypocenter of the initial F1 plane (Figure 2.3b). In contrast to the right-lateral subevents on F2 and F3, the F1 subevent ($M_W$6.3) was characterized by mostly normal slip, with the majority of the moment release occurring in the first 10 seconds of rupture (*Wei et al.*, 2011), the rupture jumped from the F1 fault plane to the steeply-dipping

F2 and F3 faults striking to the NW. Though the F2 and F3 fault planes extend more than 50 km to the north and south, respectively, their hypocenters were close enough to the CPGF to be influenced by its stress field (Figures 2.3c and 2.3d). In fact, the estimated stressing rate of $\sim +15$ kPa/year at the F2 hypocenter slightly exceeds the estimate for the stressing rate on F1. We further note that the modeled Coulomb stressing rates on the EMC subevents are caused primarily by increases in extensional stress (Supplementary Figure 2.S2), which unclamp the fault plane and allow for failure at lower levels of shear stress. Extraction at the CPGF therefore created a favorable environment for the EMC rupture to jump from F1 to F2 and F3, and continue to propagate bilaterally, rather than simply terminate as a smaller event with the cessation of slip on F1.

## 2.5 Tectonic Stressing Rates Near the EMC

Our results indicate that stresses from the CPGF may have played a role in stressing the previously latent faults that ruptured in the EMC earthquake. These faults, however, are also subjected to the tectonic stress field generated by interseismic slip on the deep extents of the nearby regional faults. To assess the effect of the magnitude of the tectonic stress field on the EMC rupture zone, we apply a regional stress-accumulation model (*Smith-Konter and Sandwell*, 2009) in which the southern San Andreas Fault system is subdivided into 18 different fault strands, each with its own slip rate and locking depth that are constrained by

geologic and geodetic observations. The tectonic stress field in the EMC rupture zone is dominated by the effects of the Imperial (40 mm/year slip rate and 5.9 km locking depth) and Cerro Prieto (40 mm/year slip rate and 10 km locking depth) faults and secondarily by the Laguna Salada (5 mm/year slip rate and 10 km locking depth) fault. We note that this model only provides an estimate of the tectonic stress during the interseismic period, so the long-term loading (i.e., including coseismic effects) of the EMC fault zone by these faults will be somewhat different. Using this model, we compute the tectonic Coulomb stressing rate on faults striking N48W, parallel to the EMC rupture trace (Figure 2.4). The tectonic stressing rate of $\sim$ -8 kPa/year at the EMC hypocenter is negative, primarily due to fault-normal compression (Supplementary Figure 2.S5), and is smaller in magnitude than the anthropogenic stressing rate from the CPGF. Thus, while regional right-lateral motion must have stressed the EMC faults to near failure over thousands of years prior to the EMC rupture, current interseismic tectonic stressing inhibits the initiation of rupture at the EMC hypocenter. Without the loading from the CPGF over the past 40 years, tectonic forces would have been changing the stress environment on the EMC faults over this time period to be less conducive to rupture.

## 2.6 Discussion

Our estimates of positive Coulomb stressing rates in the EMC rupture zone are driven primarily by the increased extensional stresses caused by volumetric contraction in the CPGF production zone. We note, however, that the a priori depth constraint on our model and the inherent uncertainties in the hypocentral depths of the EMC subevents (which are deeper than the production zone) are important sources of uncertainty in our estimates, as is our choice for the effective coefficient of friction on the fault interface, which maps extensional stress changes to Coulomb stress changes (Supplementary Figures 2.S3 and 2.S4). Furthermore, the Coulomb failure hypothesis is itself an oversimplification of the complex process of earthquake rupture initiation, a fact which precludes the establishment of a causal link between positive Coulomb stressing and the occurrence of any individual earthquake. For these reasons, our estimates of Coulomb stressing rates on the EMC rupture zone should only be interpreted as first-order approximations, and we hesitate to draw definitive conclusions about the role that the CPGF may have played in the EMC rupture sequence. Our results do, however, demonstrate that fluid extraction at the CPGF causes a substantial perturbation to the regional stress field. The magnitude of the Coulomb stressing rate within the EMC rupture zone depends mainly upon its proximity to the CPGF and the total rate of volume loss within the production zone, and is relatively insensitive to the fine spatial details of the fluid extraction model. It is notable that the magnitude of

the estimated anthropogenic stressing in the EMC rupture zone actually exceeds that of the background tectonic stressing rate (and is in fact of opposite sign), in contrast to stressing from ground water pumping in the San Joaquin Valley of California, which is 15 to 150 times smaller in magnitude, yet has been implicated in changes in seismicity on the San Andreas Fault (*Amos et al.*, 2014). Anthropogenic stresses from human activities, including those from reservoir impoundment (*Ge et al.*, 2009), wastewater injection at conventional oilfields (*Keranen et al.*, 2013), and geothermal energy production (*Deichmann and Giardini*, 2009) have all been directly linked to increased seismicity. Recent studies have found a direct correlation between net fluid extraction and local seismicity at the Salton Sea Geothermal Field (*Brodsky and Lajoie*, 2013), where the rate of net fluid extraction is almost an order of magnitude smaller than that of the CPGF. The National Resource Council 2013 concluded that maintaining a balance of extracted and injected fluids is essential to limiting the potential for energy production-related induced seismicity. Only 30% of the extracted fluid is reinjected at the CPGF (*Gutiérrez-Negrín et al.*, 2010), which stands in contrast to the more balanced reinjection practices at the other geothermal fields in the region (*Conservation*, 2014). The CPGF is a vital economic resource, but the influence of its anthropogenic stress field should not be ignored in future seismic hazard assessments of the Valle de Mexicali, home to more than a million people.

# Acknowledgements

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 2014176918. Additional support was provided by the NSF EarthScope program (EAR-1147427) and by the Southern California Earthquake (SCEC) Grant No. 14031. The ALOS PALSAR L1.0 data were provided by JAXA through the Alaska Satellite Facility (ASF). This research is also based on data provided by the Plate Boundary Observatory operated by UNAVCO for Earthscope (www.earthscope.org). The figures for this work were generated using the Generic Mapping Toolbox (GMT) (*Wessel et al.*, 2013). We thank P. Shearer for his helpful comments and suggestions, and are grateful for the thoughtful reviews provided by R. Burgmann, an anonymous reviewer, and the Associate Editor, all of which greatly improved the manuscript.

Chapter 2, in full, is a reformatted version of material as it appears in Geophysical Research Letters: Trugman, D. T., A. A. Borsa, and D. T. Sandwell (2014), Did stresses from the Cerro Prieto Geothermal Field influence the El Mayor-Cucapah rupture sequence?, *Geophysical Research Letters*, 41(24), 8767–8774, doi: 10.1002/2014GL061959. I was the primary investigator and author of this paper.

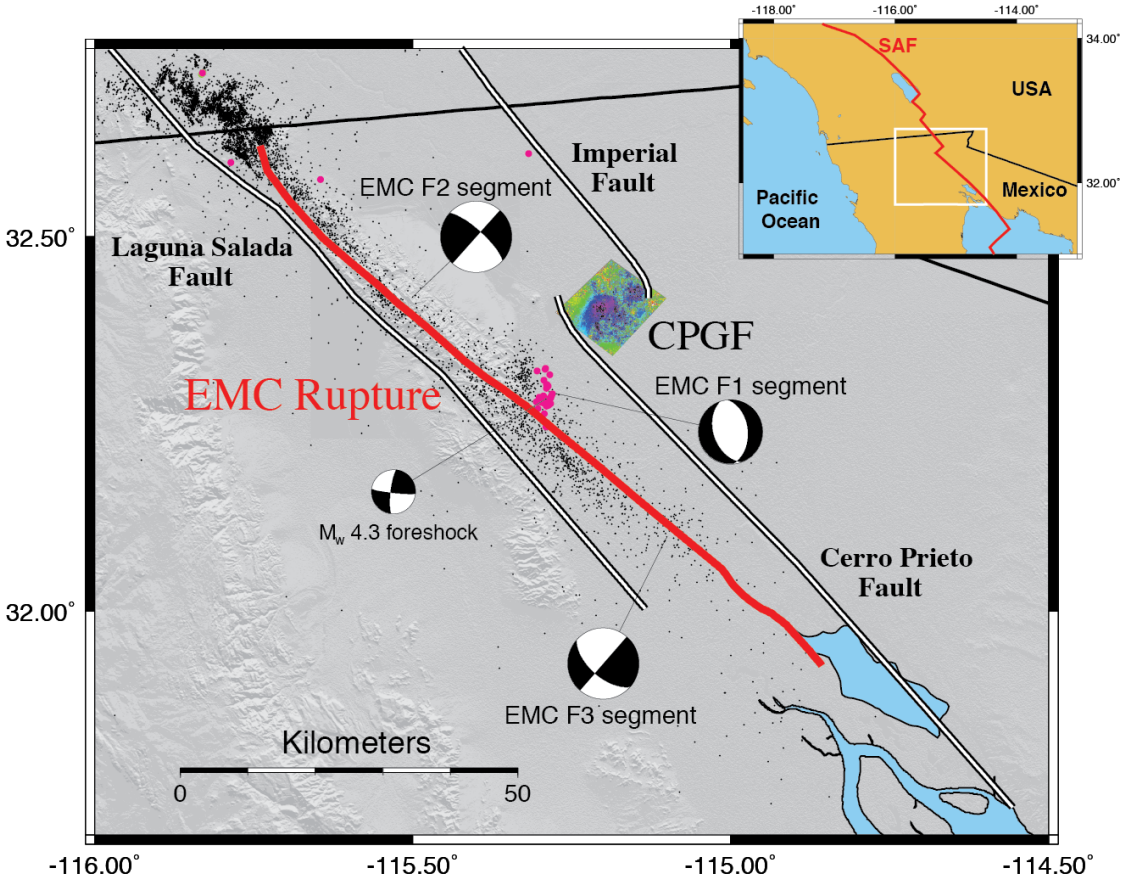


**Figure 2.1**: Map of the El Mayor-Cucapah (EMC) study region. The major regional faults (Laguna Salada, Cerro Prieto, and Imperial) are outlined in white, with the EMC earthquake rupture trace (*Wei et al.*, 2011) outlined in red. Pink circles show epicenters of the events in the foreshock sequence preceding the EMC rupture, with moment tensors for the $M_W$4.3 foreshock, and F1, F2 and F3 EMC subevents, shown for reference. Black dots indicate aftershock seismicity in the 30 days following the EMC event. All seismic events in the figure are derived from the Hauksson-Shearer waveform relocated earthquake catalog (*Hauksson et al.*, 2012). Inset: location of the study region (white box) within northern Mexico and southern California. The San Andreas Fault (SAF) boundary between the Pacific and North American Plates is marked in red.

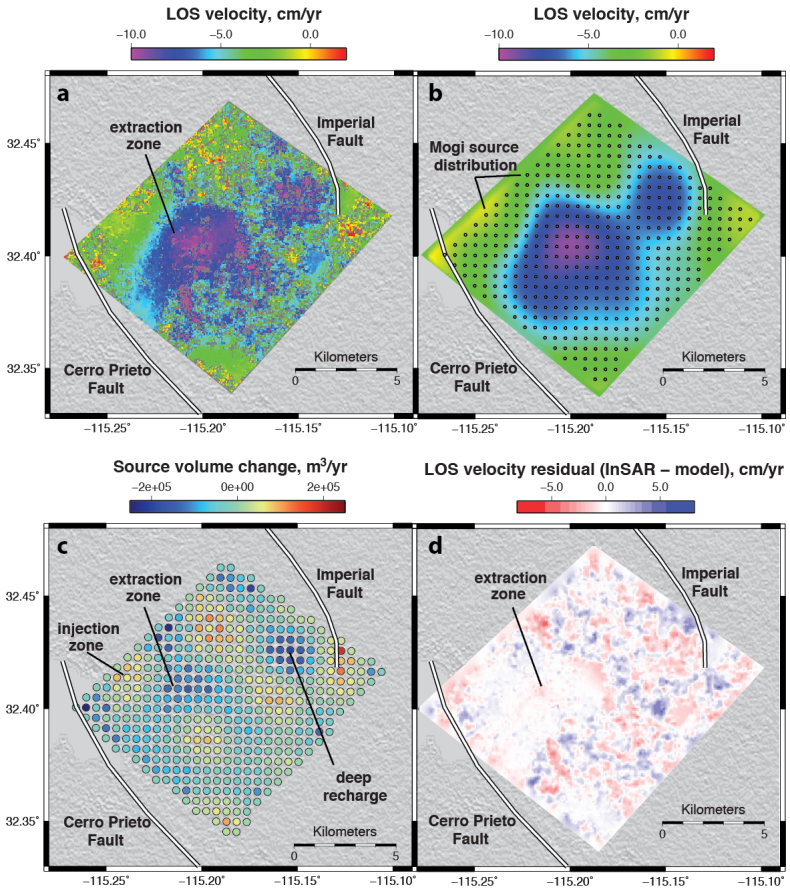


**Figure 2.2**: Subsidence at the Cerro Prieto Geothermal Field (CPGF). (a) InSAR image of the line-of-site (LOS) surface velocity field, made from stacking unwrapped ALOS interferograms from 2006-2009. Pixels with correlation $< 0.1$ are masked in the image. Negative velocities (blue) imply subsidence. The maximum vertical subsidence rate in the CPGF energy production site is $\sim 14$ cm/year. (b) Preferred Mogi source distribution fluid extraction model derived from a regularized, least-squares inversion of the InSAR LOS velocity field. Each Mogi source is color-coded by its rate of volume change ($m^3$/year). The integrated volume change over the source distribution is -9.0 x $10^6$ $m^3$/year. (c) LOS surface velocity field from our preferred Mogi source distribution fluid extraction model. Source locations are marked as open circles. The colorscale and region is identical to panel a). (d) Residual velocity field (InSAR  model). Our preferred model provides a variance reduction of 69% on the observed InSAR data.

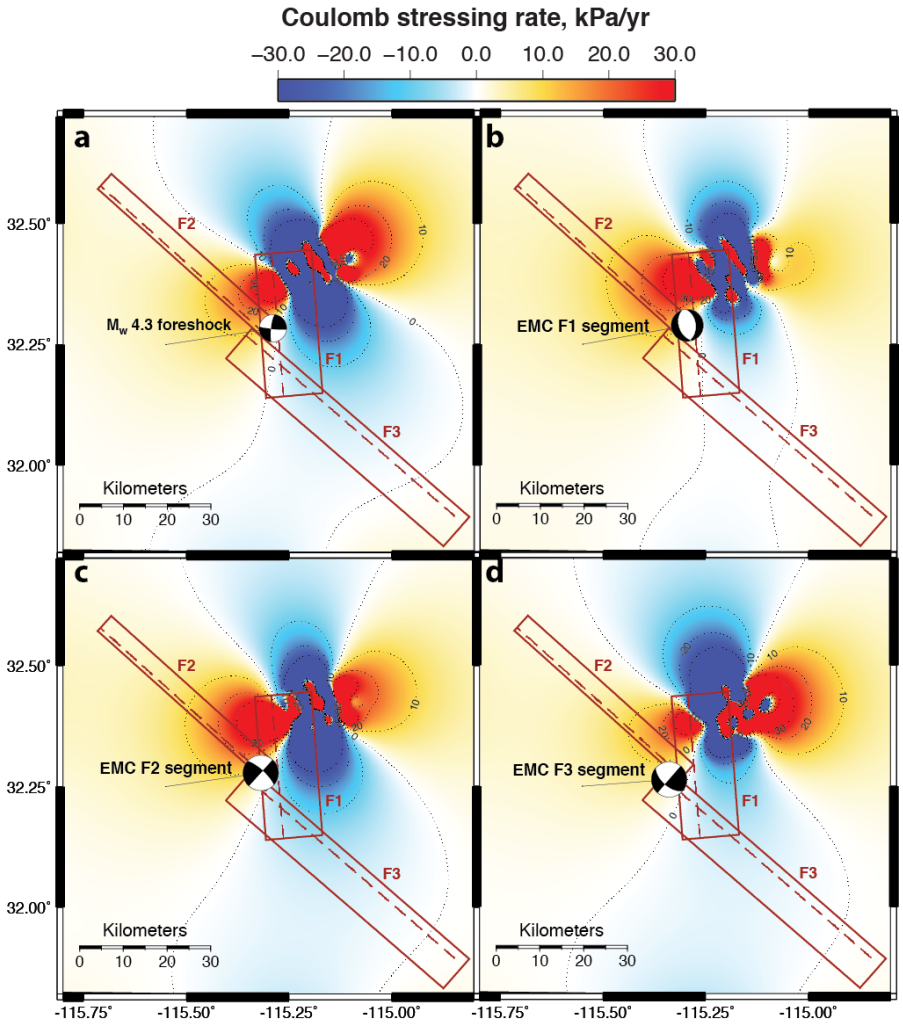


**Figure 2.3**: Coulomb stressing rates in the El Mayor-Cucapah (EMC) rupture zone due to fluid extraction at the Cerro Prieto Geothermal Field (CPGF). The F1, F2 and F3 subevent fault planes (*Wei et al.*, 2011) are shown for reference, with the assumed hypocentral depth of 5 km marked with a red dashed line. The color scale ranges from -30 to 30 kPa/year, with contours displayed in increments of 10 kPa/year. Coulomb stressing rate at 5 km depth are plotted for the fault geometries of: (a) the $M_W$4.3 foreshock (strike = 187° , dip = 79° , rake = 5° ), (b) the F1 EMC subevent (strike = 355° , dip = 45° , rake = -80° ), (c) the F2 EMC subevent (strike = 312° , dip = 75° , rake = -180° ), and (d) the F3 EMC subevent (strike = 131° , dip = 60° , rake = -180° ).

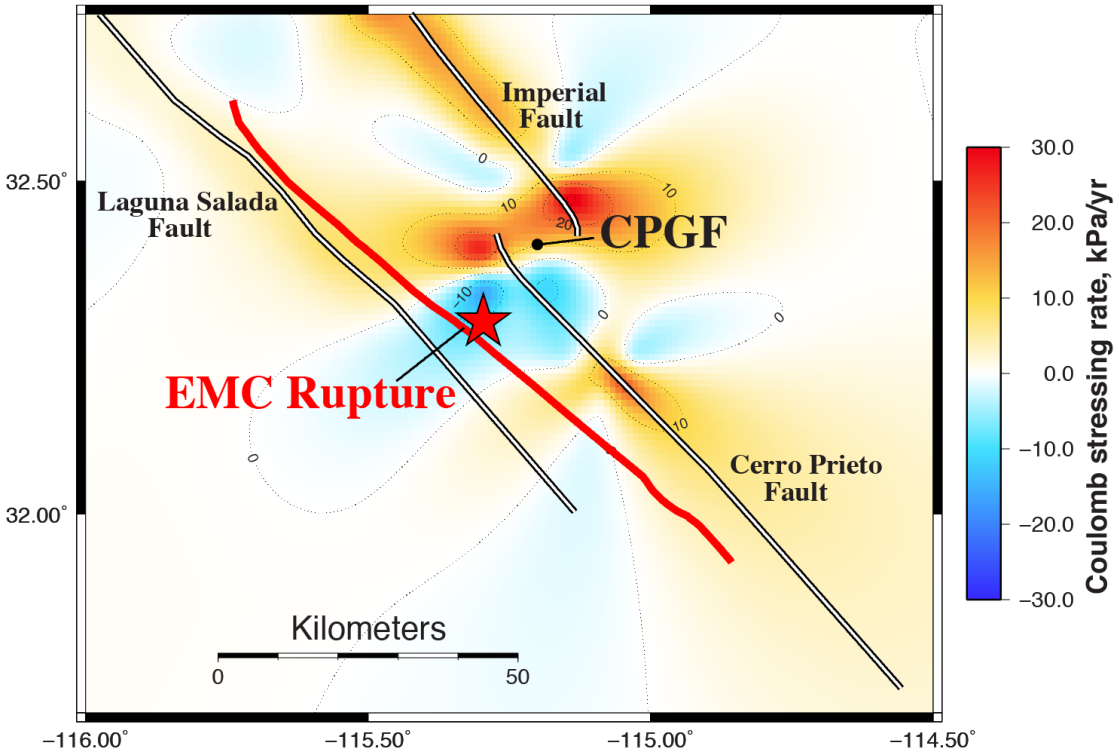


**Figure 2.4**: Coulomb stressing rate in the El Mayor-Cucapah (EMC) rupture zone due to tectonic stresses from regional faults (*Smith-Konter and Sandwell*, 2009). The stress field shown is for right-lateral slip at 5 km depth on vertical faults oriented parallel to the F2 and F3 EMC subevent fault planes (N48W). The color scale is identical to that of Figure 2.3.

# 2.7 Supplementary Materials

## Introduction

This Supplementary Material includes contains two primary components: Supplementary Text (with associated references) and Supplementary Figures 2.S1 through 2.S5 (with associated captions). The Supplementary Text details the methodology in this study related to: (1) InSAR observations, (2) our fluid extraction model, and (3) Coulomb stress computations. Figures 2.S1 through 2.S5 detail the vertical subsidence rate near the Cerro Prieto Geothermal Field (2.S1), decompose the Coulomb stress into shear and normal stress components (2.S2 for the fluid extraction model and 2.S5 for the regional tectonic model), and display the results of our sensitivity analyses for hypocentral depth (2.S3) and effective coefficient of friction (2.S4).

## Supplementary Text

### InSAR Observations

We use Advanced Land Observing Satellite (ALOS) InSAR data acquired from 2006-2009 to study in detail the surface subsidence at the Cerro Prieto Geothermal Field (CPGF) immediately preceding the El Mayor-Cucapah (EMC) event. The ALOS L-band radar achieves better temporal coherence in vegetated areas than does the C-band radar on satellites such as Envisat (*Sandwell et al.*,

2008), allowing for improved phase estimates in the vegetated regions near the CPGF (*Glowacka et al.*, 2005). All InSAR data were processed using the GMT-SAR software package (*Sandwell et al.*, 2011), using the digital elevation model from the Shuttle Radar Topography Mission (SRTM) to remove the topographic phase (*Farr et al.*, 2007). We use the SNAPHU algorithm (*Chen and Zebker*, 2001) to unwrap the phase of individual interferograms. The InSAR image of the line-of-sight (LOS) surface velocity (Figure 2.2) was generated by stacking two unwrapped ALOS interferograms. We initially processed 16 images along descending Track 211 and 7 images along ascending Track 532, but only used 4 of the T532 images where the phase unwrapping over the geothermal area was complete. These were 2-year interferograms (year 2006, day 309 to year 2008, day 315; year 2007, day 036 to year 2009, day 041). Because the subsidence rate in the area is very large, our selection process was based on phase continuity. The ascending and descending interferograms showed similar patterns, suggesting mainly vertical deformation in the region, in agreement with previous studies (*Sarychikhina*, 2003; *Glowacka et al.*, 2005; *Sarychikhina et al.*, 2011).

**Fluid Extraction Model**

We model fluid extraction at the CPGF as a lateral distribution of finely-spaced Mogi pressure sources at constant depth. This approach allows us to vary the source intensity on a fine spatial scale to better capture the complex spatial patterns of extraction, injection, and recharge beneath the CPGF. We position

the center of the source distribution in the production zone of the CPGF (-115.20° E, 32.1° N) and at the mean production depth of 2.7 km (*Lippmann et al.*, 1991; *Gutiérrez-Negrín et al.*, 2010). We discretize the source distribution using a single horizontal layer with a grid-spacing of 0.5 km in both lateral directions for a total of 436 evenly-spaced sources.

We invert for the source intensities using the InSAR observations of LOS surface velocities. For a single Mogi source at position ($x_0$, $y_0$, $z_0 < 0$) within an elastic halfspace (Poissons ratio $\nu$), the displacement vector ($u_x$, $u_y$ , $u_z$) at surface position ($x$, $y$, $z = 0$), can be written as

$$(u_x, u_y, u_z) = \frac{(1-\nu)\Delta V}{\pi R^3}(x - x_0, y - y_0, -z_0) \tag{2.3}$$

where $R = \sqrt{(x-x_0)^2 + (y-y_0)^2 + (-z_0)^2}$ is the distance from the source to the surface observation point and $\Delta V$ is the associated volume change (i.e., the source intensity) (*Segall*, 2010). Subsidence requires $\Delta V < 0$.

The observed surface deformation is the superposition of the deformation caused by each individual Mogi source in the distribution. If the source positions are known, then the forward computation of the LOS velocity field can be written in the form:

$$\boldsymbol{d} = \underline{\boldsymbol{G}}\boldsymbol{m} \tag{2.4}$$

where $\boldsymbol{d}$ is the data vector of observed LOS velocities (the dot product of the displacement velocity vector and satellite look vector), $\boldsymbol{m}$ is the model vector of source volumes, and $\underline{\boldsymbol{G}}$ is the matrix of LOS displacement Green's functions derived

from equation (2.3). We invert for the source volumes by performing a regularized least-squares inversion (e.g., *Parker*, 1994):

$$\boldsymbol{m_0} = \text{argmin}_{\boldsymbol{m}} \left\{ ||\boldsymbol{d} - \underline{\boldsymbol{G}}\boldsymbol{m}||^2 + \lambda^2 ||\underline{\boldsymbol{D}}\boldsymbol{m}||^2 \right\} \tag{2.5}$$

to obtain the model $\boldsymbol{m_0}$ that minimizes a linear combination of: (i) the residual norm between the observed (InSAR-derived) and model-predicted LOS velocity, and (ii) a model norm parameterized by a first-order Tikhonov smoothing operator $\underline{\boldsymbol{D}}$. We also performed analogous inversions with 3 and 5 horizontal layers, and with grid spacings ranging from 0.25 km to 2.5 km, and observed no appreciable change to the data misfit or modeled integrated volume loss. The addition of multiple layers tends to destabilize the inversion process, so our preferred model contains a single horizontal layer.

**Coulomb Stress Computations**

Given a fault plane orientation (parameterized in terms of a normal and slip vector), we define the Coulomb stress change as

$$\Delta\sigma_C = \Delta\tau + \mu' \Delta\sigma_N \tag{2.6}$$

where $\Delta\tau$ is the change in shear stress in the direction of slip, $\Delta\sigma_N$ is the change in normal stress (assumed positive in extension), and $\mu'$ is the effective coefficient of friction. Positive Coulomb stress changes on a given fault are presumed to push that fault toward failure (*King et al.*, 1994; *Stein*, 1999), and are caused by increases in shear stress or extensional normal stress (i.e., unclamping of the fault

plane). For the stress computations in this study, we assume a homogenous elastic medium with a shear modulus of 32 GPa, Poissons ratio of 0.25, and an effective coefficient of friction of 0.4 (*Lin and Stein*, 2004).

To compute the Coulomb stresses associated with our fluid extraction model, we first generalize the Fourier-domain approach of *Steketee* (1958) to obtain the Greens function for a radial point source in an elastic half-space. In this formulation, the half-space Green's function is a semi-analytic function of the horizontal wavenumbers ($k_x$ and $k_y$) and vertical position (z). This half-space Green's function is composed of the superposition of three terms: (i) the full-space Green's function for a source at depth $z = -a$, (ii) an image full-space Green's function a $z = +a$, and (iii) a Boussinesq correction to ensure zero traction at the free surface ($z = 0$).

With the half-space Green's function in hand, the full 3D strain tensor for an arbitrary distribution of radial point sources is easily obtained through convolution (multiplication in the wavenumber domain) of the Green's function and source distribution. To compute Coulomb stresses, we apply assume isotropic, linear elasticity to convert strains into stresses, and then resolve the local stress tensor on the fault plane geometries (and slip directions) of each of the individual subevents that comprise the EMC rupture sequence. For EMC subevents F1 through F3, we use the fault plane parameterization (strike, dip, and rake) presented in *Wei et al.* (2011) to obtain the slip and normal vectors. For the $M_W$4.3 foreshock, we use the fault plane and slip orientation described by *Hauksson et al.* (2011).

To assess the effect of the magnitude of the tectonic stress field on the EMC rupture zone, we apply the regional stress-accumulation model (*Smith-Konter and Sandwell*, 2009) described in the main text. We then compute the normal, shear and Coulomb tectonic stressing rates at 5 km depth on faults striking N48W, parallel to the EMC rupture trace (Figures 2.4 and 2.S5).

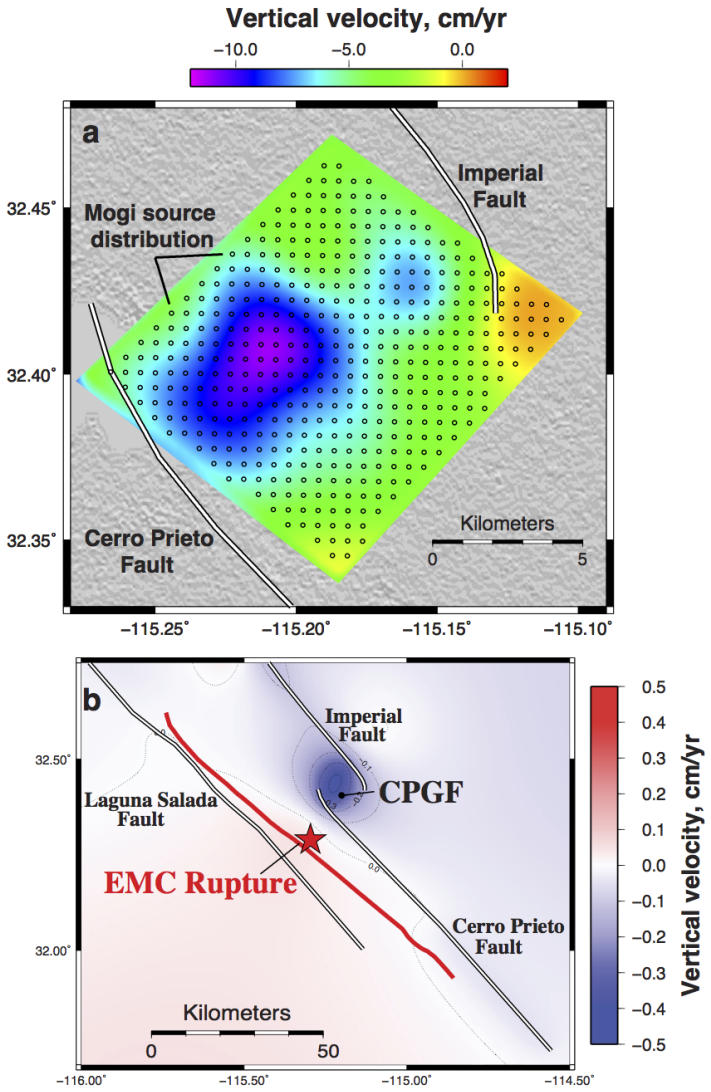


**Figure 2.S1**: Modeled and tectonic subsidence near the Cerro Prieto Geothermal Field (CPGF). (a) Rate of vertical surface subsidence near the CPGF predicted by our preferred Mogi source distribution fluid extraction model. The maximum rate of vertical surface subsidence is $\sim$ 14 cm/year. (b) Estimated rate of vertical surface subsidence in the El Mayor-Cucapah (EMC) rupture zone due to natural, interseismic deformation from regional faults. The tectonic subsidence rate was computed using the stress accumulation model of *Smith-Konter and Sandwell* (2009). The estimated tectonic subsidence rate of 0.4 cm/year within the CPGF is $\sim$ 3% of the observed rate, which is dominantly anthropogenic in origin.

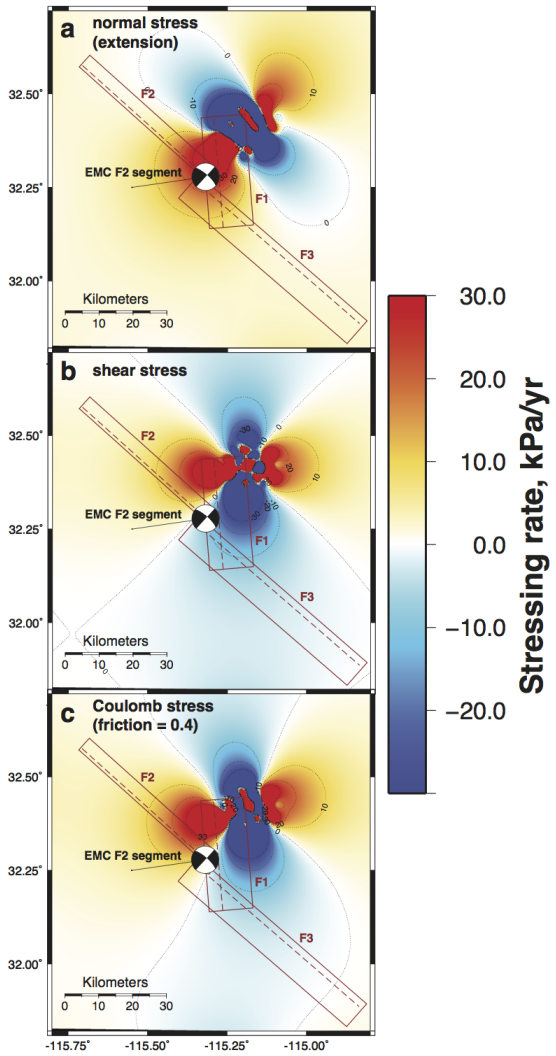


**Figure 2.S2**: Comparison of (a) normal, (b) shear, and (c) Coulomb stressing rates on El Mayor-Cucapah (EMC) fault plane F2, due to fluid extraction at the Cerro Prieto Geothermal Field. The stressing rates are computed at 5 km depth for the fault geometry of the EMC F2 subevent (strike = 312° , dip = 75° , rake = -180° ). An effective coefficient of friction of 0.4 is assumed for panel c). The stress decomposition for fault plane F3 is nearly identical (as the fault geometry is close to the same). The stress decomposition for F1 is also quite similar, but the extensional stress is slightly reduced (but still positive) at the F1 hypocenter, and the shear stress is also positive at the hypocenter. This difference is due to the difference in fault geometry and slip orientation, as F1 is a normal faulting event with slightly different strike and dip than F2 (a strike-slip event).

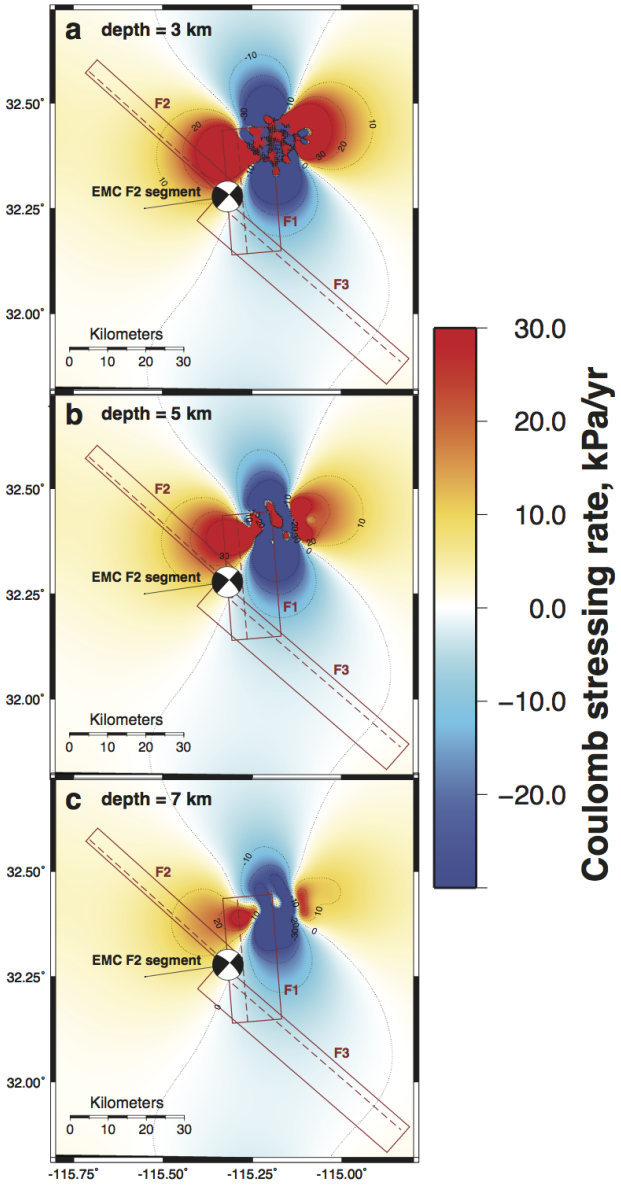


**Figure 2.S3**: Sensitivity analysis of hypocentral depth for Coulomb stressing rate on El Mayor-Cucapah (EMC) fault plane F2 due to fluid extraction at the Cerro Prieto Geothermal Field. An effective coefficient of friction of 0.4 is used for all panels. Modeled Coulomb stressing rate for the fault geometry of the EMC F2 subevent (strike = 312° , dip = 75° , rake = -180° ) at: (a) 3 km depth, (b) 5 km depth, and (c) 7 km depth. The results of this study assume hypocentral depths of 5 km (*Wei et al.*, 2011), and are similar for fault planes F1 and F3 (not shown here).

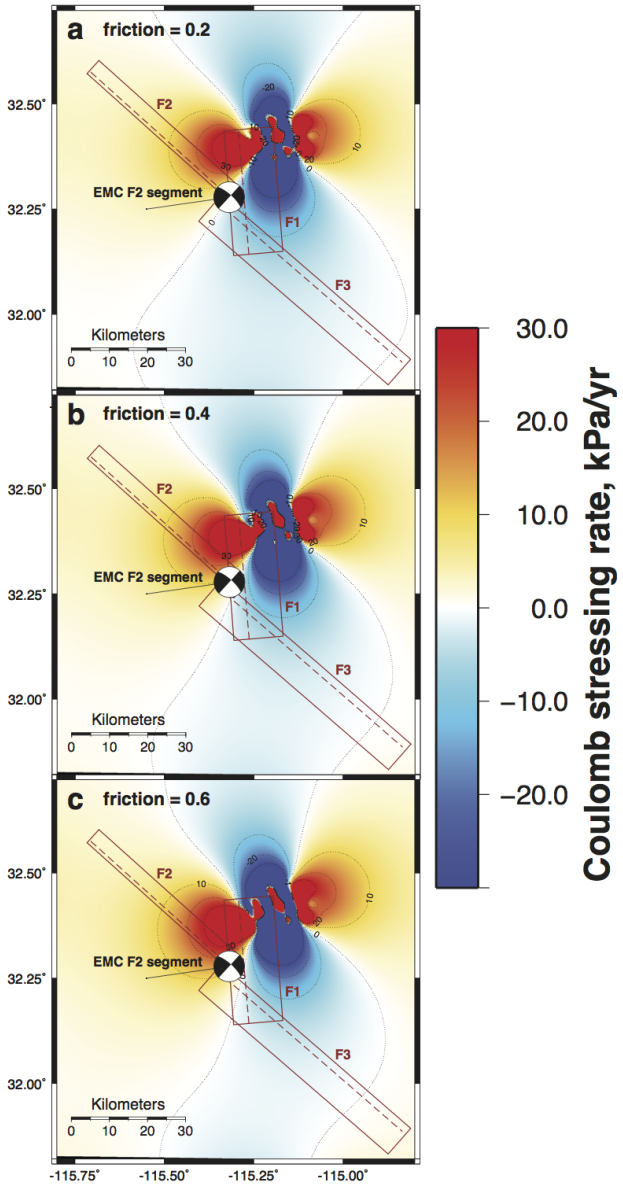


**Figure 2.S4**: Sensitivity analysis of the effective coefficient of friction for the Coulomb stressing rate on El Mayor-Cucapah (EMC) fault plane F2 due to fluid extraction at the Cerro Prieto Geothermal Field. Modeled Coulomb stressing rate at 5 km depth for the fault geometry of the EMC F2 subevent (strike = 312° , dip = 75° , rake = -180° ) assuming: (a) an effective coefficient of friction of 0.2, (b) an effective coefficient of friction of 0.4, and (c) an effective coefficient of friction of 0.6. The results of this study assume an effective coefficient of friction of 0.4, and are similar for fault planes F1 and F3 (not shown here).

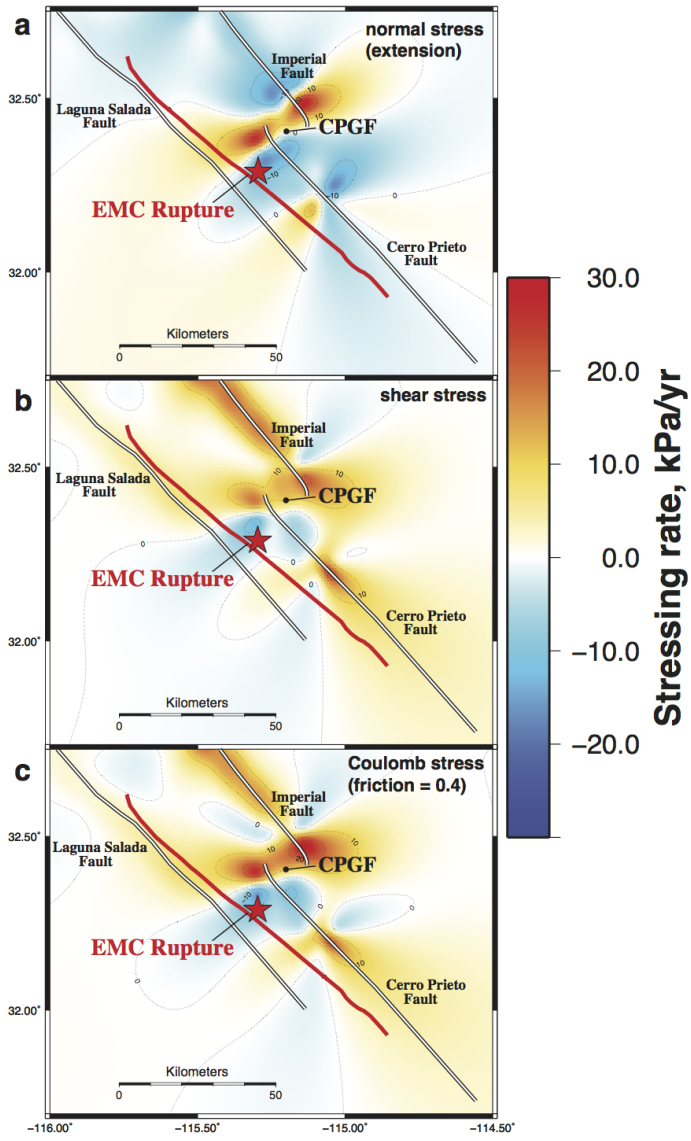


**Figure 2.S5**: Comparison of (a) normal, (b) shear, and (c) Coulomb stressing rates in the El Mayor-Cucapah (EMC) rupture zone due to tectonic stresses from regional faults (*Smith-Konter and Sandwell*, 2009). The stress fields shown are for right-lateral slip at 5 km depth on vertical faults oriented parallel to the F2 and F3 EMC subevent fault planes (N48W). An effective coefficient of friction of 0.4 is assumed for panel c).

# References


Amos, C. B., P. Audet, W. C. Hammond, R. Burgmann, I. A. Johanson, and G. Blewitt (2014), Uplift and seismicity driven by groundwater depletion in central California, *Nature*, *509*(7501), 483–486.

Atwater, T., and J. Stock (1998), Pacific North America plate tectonics of the Neogene southwestern United States: An update, *International Geology Review*, *40*, 375–402, 10.1080/00206819809465216.

Barnard, F. (1968), *Structural Geology of the Sierra de Los Cucapas, Northeastern Baja California, Mexico, and Imperial County, California*, University of Colorado, Boulder, CO.

Bennett, R. A., W. Rodi, and R. E. Reilinger (1996), Global Positioning System constraints on fault slip rates in southern California and northern Baja, Mexico, *Journal of Geophysical Research: Solid Earth*, *101*(B10), 21,943–21,960.

Brodsky, E. E., and L. J. Lajoie (2013), Anthropogenic Seismicity Rates and Operational Parameters at the Salton Sea Geothermal Field, *Science*, *341*(6145), 543–546, doi:10.1126/science.1239213.

Carnec, C., and H. Fabriol (1999), Monitoring and modeling land subsidence at the Cerro Prieto Geothermal Field, Baja California, Mexico, using SAR interferometry, *Geophysical Research Letters*, *26*(9), 1211–1214.

CFE. (2006), Cerro Prieto Geothermal Field (brochure), *Tech. rep.*

Chen, C. W., and H. A. Zebker (2001), Two-dimensional phase unwrapping with use of statistical models for cost functions in nonlinear optimization, *JOSA A*, *18*(2), 338–351.

Conservation, C. D. o. (2014), *GeoSteamQuery Geothermal Well Records, Production and Injection Data*, vol. 2014.

Deichmann, N., and D. Giardini (2009), Earthquakes Induced by the Stimulation of an Enhanced Geothermal System below Basel (Switzerland), *Seismological Research Letters*, *80*(5), 784–798, doi:10.1785/gssrl.80.5.784.

Farr, T. G., P. A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth, D. Seal, S. Shaffer, J. Shimada, J. Umland, M. Werner, M. Oskin, D. Burbank, and D. Alsdorf (2007), The Shuttle Radar Topography Mission, *Reviews of Geophysics*, *45*(2), – RG2004.

Fletcher, J. M., and R. M. Spelz (2009), Patterns of Quaternary deformation and rupture propagation associated with an active low-angle normal fault, Laguna Salada, Mexico: Evidence of a rolling hinge?, *Geosphere*, *5*, 385–407, 10.1130/GES00206.1.

Fletcher, J. M., O. J. Teran, T. K. Rockwell, M. E. Oskin, K. W. Hudnut, K. J. Mueller, R. M. Spelz, S. O. Akciz, E. Masana, G. Faneros, E. J. Fielding, S. Leprince, A. E. Morelan, J. Stock, D. K. Lynch, A. J. Elliott, P. Gold, J. Liu-Zeng, A. González-Ortega, A. Hinojosa-Corona, and J. González-García (2014), Assembly of a large earthquake from a complex fault system: Surface rupture kinematics of the 4 April 2010 El MayorCucapah (Mexico) Mw 7.2 earthquake, *Geosphere*, *10*(4), 797–827, doi:10.1130/GES00933.1.

Ge, S., M. Liu, N. Lu, J. W. Godt, and G. Luo (2009), Did the Zipingpu Reservoir trigger the 2008 Wenchuan earthquake?, *Geophysical Research Letters*, *36*(20), L20,315.

Glowacka, E., and F. A. Nava (1996), *Major earthquakes in Mexicali Valley, Mexico, and fluid extraction at Cerro Prieto Geothermal Field*, vol. 86.

Glowacka, E., O. Sarychikhina, and F. A. Nava (2005), Subsidence and stress change in the Cerro Prieto geothermal field, BC, Mexico, *Pure and Applied Geophysics*, *162*(11), 2095–2110.

Glowacka, E., O. Sarychikhina, F. Suárez, F. A. Nava, and R. Mellors (2010), Anthropogenic subsidence in the Mexicali Valley, Baja California, Mexico, and slip on the Saltillo fault, *Environmental Earth Sciences*, *59*(7), 1515–1524.

Gutiérrez-Negrín, L. C. A., R. Maya-González, and J. L. Quijano-León (2010), Current Status of Geothermics in Mexico, in *Proceedings World Geothermal Congress*, Bali, Indonesia.

Hanssen, R. F. (2001), *Radar Interferometry: Data Interpretation and Error Analysis*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Hauksson, E., J. Stock, K. Hutton, W. Yang, J. A. Vidal-Villegas, and H. Kanamori (2011), The 2010 M w 7.2 El Mayor-Cucapah Earthquake Sequence, Baja California, Mexico and Southernmost California, USA: Active Seismotectonics along the Mexican Pacific Margin, *Pure and Applied Geophysics*, *168*(8-9), 1255–1277, doi:10.1007/s00024-010-0209-7.

Hauksson, E., W. Yang, and P. M. Shearer (2012), Waveform Relocated Earthquake Catalog for Southern California (1981 to June 2011), *Bulletin of the Seismological Society of America*, *102*(5), 2239–2244, doi:10.1785/0120120010.

Keranen, K. M., H. M. Savage, G. A. Abers, and E. S. Cochran (2013), Potentially induced earthquakes in Oklahoma, USA: Links between wastewater injection and the 2011 Mw 5.7 earthquake sequence, *Geology*, *41*(6), 699–702, doi:10.1130/G34045.1.

King, G. C. P., R. S. Stein, and J. Lin (1994), Static stress changes and the triggering of earthquakes, *Bulletin of the Seismological Society of America*, *84*(3), 935–953.

Lin, J., and R. S. Stein (2004), Stress triggering in thrust and subduction earthquakes and stress interaction between the southern San Andreas and nearby thrust and strike-slip faults, *Journal of Geophysical Research: Solid Earth*, *109*(B2), doi:10.1029/2003JB002607.

Lippmann, M. J., A. H. Truesdell, S. E. Halfman-Dooley, and M. A (1991), A review of the hydrogeologic-geochemical model for Cerro Prieto, *Geothermics*, *20*(12), 39–52.

Mogi, K. (1958), Relations between the eruptions of various volcanoes and the deformations of the ground surfaces around them, *Bulletin of the Earthquake Research Institute - University of Tokyo*, *36*, 99–134.

Mossop, A., and P. Segall (1997), Subsidence at The Geysers Geothermal Field, N. California from a comparison of GPS and leveling surveys, *Geophysical Research Letters*, *24*(14), 1839–1842, doi:10.1029/97GL51792.

Mossop, A., and P. Segall (1999), Volume strain within The Geysers geothermal field, *Journal of Geophysical Research*, *104*(B12), 29,113, doi:10.1029/1999JB900284.

National Resource Council (2013), *Induced Seismicity Potential in Energy Technologies*, The National Academies Press, Washington, DC.

Ocampo-Díaz, J. D. D., B. Valdez-Salaz, M. Shorr, M. I. Sauceda, and N. Rosas-González (2005), Review of corrosion and scaling problems in Cerro Prieto geothermal field over 31 years of commercial operations, in *Proceedings of World Geothermal Congress, International Geothermal Association (IGA), Antalya, Turkey*, pp. 1–5.

Parker, R. L. (1994), *Geophysical Inverse Theory*, Princeton University Press, Princeton, N.J.

Sandwell, D., R. Mellors, X. Tong, M. Wei, and P. Wessel (2011), Open radar interferometry software for mapping surface deformation, *Eos, Transactions American Geophysical Union*, *92*(28), 234–234.

Sandwell, D. T., D. Myer, R. Mellors, M. Shimada, B. Brooks, and J. Foster (2008), Accuracy and Resolution of ALOS Interferometry: Vector Deformation Maps of the Father's Day Intrusion at Kilauea, *Geoscience and Remote Sensing, IEEE Transactions on*, *46*(11), 3524–3534, iD: 1.

Sarychikhina, O. (2003), Modelación de subsidencia en el campo geotérmico Cerro Prieto, M.s. thesis, CICESE, Ensenada, Mexico.

Sarychikhina, O., E. Glowacka, R. Mellors, and F. S. Vidal (2011), Land subsidence in the Cerro Prieto Geothermal Field, Baja California, Mexico, from 1994 to 2005: An integrated analysis of DInSAR, leveling and geological data, *Journal of Volcanology and Geothermal Research*, *204*(1), 76–90.

Segall, P. (2010), *Earthquake and Volcano Deformation*, Princeton University Press, Princeton, N.J.

Sieh, K., L. Jones, E. Hauksson, K. Hudnut, D. Eberhart-Phillips, T. Heaton, S. Hough, K. Hutton, H. Kanamori, A. Lilje, S. Lindvall, S. F. McGill, J. Mori, C. Rubin, J. A. Spotila, J. Stock, H. K. Thio, J. Treiman, B. Wernicke, and J. Zachariasen (1993), Near-Field Investigations of the Landers Earthquake Sequence, April to July 1992, *Science*, *260*(5105), 171–176, doi: 10.1126/science.260.5105.171.

Smith-Konter, B., and D. Sandwell (2009), Stress evolution of the San Andreas fault system: Recurrence interval versus locking depth, *Geophysical Research Letters*, *36*(13).

Stein, R. S. (1999), The role of stress transfer in earthquake occurrence, *Nature*, *402*(6762), 605–609, doi:10.1038/45144.

Steketee, J. A. (1958), On Volterra's dislocations in a semi-infinite elastic medium, *Canadian Journal of Physics*, *36*(2), 192–205, doi:10.1139/p58-024.

Vasco, D. W., C. Wicks, K. Karasaki, and O. Marques (2002), Geodetic imaging: reservoir monitoring using satellite interferometry, *Geophysical Journal International*, *149*(3), 555–571.

Wei, S., E. Fielding, S. Leprince, A. Sladen, J.-P. Avouac, D. Helmberger, E. Hauksson, R. Chu, M. Simons, K. Hudnut, T. Herring, and R. Briggs (2011), Superficial simplicity of the 2010 El MayorCucapah earthquake of Baja California in Mexico, *Nature Geoscience*, *4*(9), 615–618, doi:10.1038/ngeo1213.

Wessel, P., W. H. F. Smith, R. Scharroo, J. Luis, and F. Wobbe (2013), Generic Mapping Tools: Improved Version Released, *Eos, Transactions American Geophysical Union*, *94*(45), 409–410, doi:10.1002/2013EO450001.

# Chapter 3

# A comparison of long-term changes in seismicity at the Geysers, Salton Sea, and Coso geothermal fields

## Abstract

Geothermal energy is an important source of renewable energy, yet its production is known to induce seismicity. Here we analyze seismicity at the three largest geothermal fields in California: The Geysers, Salton Sea, and Coso. We focus on resolving the temporal evolution of seismicity rates, which provides important observational constraints on how geothermal fields respond to natural and

anthropogenic loading. We develop an iterative, regularized inversion procedure to partition the observed seismicity rate into two components: (1) the interaction rate due to earthquake-earthquake triggering, and (2) the smoothly-varying background rate controlled by other time-dependent stresses, including anthropogenic forcing. We apply our methodology to compare long-term changes in seismicity to monthly records of fluid injection and withdrawal. At The Geysers, we find that the background seismicity rate is highly correlated with fluid injection, with the mean rate increasing by approximately 50 percent and exhibiting strong seasonal fluctuations following construction of the Santa Rosa pipeline in 2003. In contrast, at both Salton Sea and Coso, the background seismicity rate has remained relatively stable since 1990, though both experience short-term rate fluctuations that are not obviously modulated by geothermal plant operation. We also observe significant temporal variations in Gutenberg-Richter $b$-value, earthquake magnitude distribution, and earthquake depth distribution, providing further evidence for the dynamic evolution of stresses within these fields. The differing field-wide responses to fluid injection and withdrawal may reflect differences in in-situ reservoir conditions and local tectonics, suggesting that a complex interplay of natural and anthropogenic stressing controls seismicity within California's geothermal fields.

## 3.1 Introduction

Recent technological developments have vastly expanded our capacity to exploit new sources of energy from beneath Earth's surface. At the same time, there is a rising level of public and scientific concern over seismicity induced by the anthropogenic stresses from energy production (*National Resource Council*, 2013; *Ellsworth*, 2013; *Weingarten et al.*, 2015). In recent years, fluid injection and withdrawal at oil and gas production facilities are thought to have triggered earthquakes in a number of states within the central and eastern United States, including Oklahoma (*Keranen et al.*, 2013, 2014), Texas (*Frohlich et al.*, 2011; *Frohlich and Brunt*, 2013), Ohio (*Kim*, 2013), Colorado (*Ake et al.*, 2005), Arkansas (*Horton*, 2012), and New Mexico (*Rubinstein et al.*, 2014). Two primary physical mechanisms related to fluid injection and withdrawal can account for this induced seismicity. Fluid fluxes from injection and withdrawal can directly increase the local pore pressure, which reduces the effective normal stress on nearby faults and hence pushes them closer to failure (*Hubbert and Rubey*, 1959; *Healy et al.*, 1968; *Raleigh et al.*, 1976). Fluid injection and withdrawal can also modify the local stress state in the solid host rock through poroelastic coupling (*Biot*, 1941), which under certain conditions can trigger earthquakes (*Segall*, 1989; *Suckale*, 2009; *Segall and Lu*, 2015), and more generally can create poroelastic stresses at considerable distances from production and injection wells (*Segall et al.*, 1994). Through these mechanisms — direct pore pressure increase and poroelastic stressing — both fluid injection and

withdrawal can lead to induced seismicity in regions subject to intense oil and gas production, such as the central and eastern United States.

California has a high ambient rate of natural (tectonic) earthquakes, which makes distinguishing between natural and induced seismicity a less straightforward task than in the central and eastern United States, where the natural seismicity rate is low (*Goebel*, 2015). In addition to hosting the extensive conventional oil and gas production fields of California's Kern County, San Joaquin Valley, and Los Angeles Basin, the state also accounts for more than eighty percent of geothermal energy production within the United States (*Geothermal Energy Association*, 2013). Geothermal fields like The Geysers, Salton Sea, and Coso are among the world's largest (*Matek and Gawell*, 2014). As in conventional oil and gas production, geothermal energy production requires the extraction of fluids from the subsurface, and the resulting local changes in pore pressure and poroelastic stresses can trigger earthquakes (*Deichmann and Giardini*, 2009; *Brodsky and Lajoie*, 2013). In geothermal fields, however, a third mechanism — thermoelastic stresses from fluid fluxes within hot geothermal reservoirs (*Segall and Fitzgerald*, 1998; *Majer and Peterson*, 2007) — can also induce seismicity. And while it is common practice to replenish over-exploited geothermal reservoirs through fluid injection, failure to maintain a net fluid balance between injection and withdrawal can result in volumetric contraction within the reservoir, causing the surface to subside locally (*Mossop and Segall*, 1997; *Fialko and Simons*, 2000) and generating substantial poroelastic stresses at remote distances (*Trugman et al.*, 2014). These differences

make it uncertain whether insights gained from studies of induced seismicity within the central and eastern United States will generalize to the geothermal fields of California.

This study focuses on long-term changes in seismicity within the three largest geothermal fields in California: (1) The Geysers, located in Sonoma and Lake Counties to the north of San Francisco, (2) Salton Sea Geothermal Field, located in Imperial county in southeastern California, and (3) Coso Geothermal Field, located in Inyo county in eastern California. Induced seismicity has been previously studied at The Geysers (*Mossop and Segall*, 1997; *Majer and Peterson*, 2007; *Martinez-Garzon et al.*, 2014), Salton Sea (*Brodsky and Lajoie*, 2013) and Coso (*Feng and Lees*, 1998; *Fialko and Simons*, 2000; *Schoenball et al.*, 2015) geothermal fields, each in isolation. Our aim in this study is to provide a quantitative comparison of the temporal evolution of seismicity within these three prominent California geothermal fields, each with distinct local tectonics, reservoir conditions, and histories of energy production. In so doing, we hope to better understand the extent to which the production of geothermal energy, a nominally renewable energy source with tremendous potential for expansion in California (*Monastero*, 2002; *Adams*, 2011; *Matek and Gawell*, 2014), influences local seismicity, and by extension, local seismic hazard.

This manuscript is organized as follows. We begin by describing our methodology for estimating the time-varying background seismicity rate in each geothermal field. This is an essential step in our analyses, as the total observed seismicity

rate is highly sensitive to coseismic stress changes from earthquake-earthquake interactions (most notably, aftershock sequences) that are unrelated to energy production. In brief, we use an Epidemic Type Aftershock Sequence (ETAS) framework (*Ogata*, 1988) to separate the total seismicity rate into two components: (1) the expected seismicity rate due to earthquake interaction and coseismic stresses (parameterized using the ETAS model), and (2) the time-varying background seismicity rate driven by fluctuations in anthropogenic and tectonic stresses. The generality of the method allows for its application to study seismicity rate changes in a variety of contexts, and hence is described in some detail. We next apply our methodology to each of the three geothermal fields in turn, comparing the temporal evolution of seismicity rates, as well as earthquake magnitude and depth distributions, to monthly records of fluid injection and withdrawal. At The Geysers, we find that the background seismicity rate is strongly correlated with the monthly rate of fluid injection, and we observe significant temporal changes in Gutenberg-Richter $b$-value, magnitude distribution, and depth distribution of earthquakes within the field. At both the Salton Sea and Coso geothermal fields, however, we find that the background seismicity rate is strongly correlated with fluid injection and withdrawal only during the early years of field operation (before 1990). We further observe temporal variations in $b$-value and in the magnitude and depth distributions of earthquakes at the Salton Sea and Coso that may in part be the consequence of geothermal field operation, but are also likely controlled by natural tectonic processes. We conclude our study with a comparison of the tem-

poral evolution of seismicity in each field, and discuss how our results contribute to the current understanding of induced seismicity in a broader context.

## 3.2 Methods and Description of Seismicity Rate Model

### 3.2.1 Background on ETAS Modeling

Our methodology for computing time variations in seismicity rate is based upon the Epidemic Type Aftershock Sequence (ETAS) model (*Ogata*, 1988). In the ETAS model, earthquake occurrence within a region is modeled as a non-homogeneous Poisson process with a total seismicity rate function, $\lambda(t)$, that is the sum of two terms: the background seismicity rate $\mu$, assumed constant in time, and an interaction term $\nu(t)$ that represents the expected seismicity rate increase and subsequent decay characteristic of aftershock sequences (*Omori*, 1894; *Utsu*, 1961). The latter term is typically defined using a point process model (*Daley and Vere-Jones*, 2003) in which each observed earthquake generates a step increase in seismicity rate that decays with time following a modified Omori law (*Utsu et al.*, 1995). With this parameterization, the ETAS model for seismicity rate then takes the form:

$$\lambda(t) = \mu + \nu(t) = \mu + \sum_{i;\, t>t_i} \frac{K\, 10^{\alpha(M_i - M_c)}}{(t - t_i + c)^p}, \tag{3.1}$$

where $t_i$ and $M_i$ are the occurrence time and magnitude of the $i$-th earthquake, $K$ and $\alpha$ are aftershock productivity parameters, $p$ and $c$ are Omori decay parameters, and $M_c$ is the magnitude of completion (above which the seismicity rate is computed). Thus, for an observed catalog of earthquakes $(t_i, M_i)$ and assumed magnitude of completion $M_c$, the standard ETAS model of seismicity rate is fully parameterized by five constants: $\mu$, $K$, $\alpha$, $p$, and $c$. Alternative forms of the ETAS model include additional free parameters related to the expected spatial dependence of aftershock earthquake clustering (*Ogata*, 2004). We use the space-independent formulation presented above because it is more appropriate for studying field-wide changes in seismicity, especially in the absence of reliable estimates for the spatial anisotropy in the coseismic stresses that trigger aftershock activity based on earthquake catalog information alone (*Marsan and Lengliné*, 2010).

The optimal set of ETAS parameters $\theta = \{\mu, K, \alpha, p, c\}$ is region-specific, and the parameter values are typically estimated by maximizing the likelihood of the observed earthquake data (*Ogata*, 1983, 1992). This is equivalent to minimizing the negative log-likelihood (*Daley and Vere-Jones*, 2003), which for the ETAS model can be written as:

$$NLL(\theta) = -\left[\sum_{i;\, T_0<t_i<T_1} \log \lambda(t_i \,|\, \theta)\right] + \int_{T_0}^{T_1} \lambda(t \,|\, \theta) dt, \tag{3.2}$$

where $T_0$ and $T_1$ denote the temporal limits of the observation interval for the earthquake catalog. In practice, the minimization of (3.2) can be implemented by any of a number of numerical optimization routines (*Press*, 2007). However,

since the rate function $\lambda(t)$ at time $t$ is conditionally dependent on the history of earthquake occurrence (all events with $t_i < t$), the starting time of the observation interval $T_0$ should be chosen to be some time well after the first observed events in order to obtain unbiased parameter estimates (*Ogata*, 1992).

### 3.2.2 Temporal Variations in the Background Seismicity Rate

In the standard form of the ETAS model (described in the preceding subsection), the background seismicity rate $\mu$ is a constant that is independent of time. Thus, in this model the entire time-dependence in the seismicity rate function $\lambda(t)$ (Equation 3.1) is due to earthquake-earthquake triggering, or equivalently, coseismic stress changes. However, transient stresses from processes unrelated to earthquake interaction, including anthropogenic forcing (*Nicholson and Wesson*, 1992; *Keranen et al.*, 2014; *Weingarten et al.*, 2015) and aseismic fault slip (*Toda et al.*, 2002; *Lohman and McGuire*, 2007; *Llenos et al.*, 2009), can cause time-dependent changes in seismicity rate. These changes in seismicity can be encompassed in a generalized ETAS framework by allowing the background rate to vary with time: $\mu = \mu(t)$ (*Marsan et al.*, 2013; *Kumazawa and Ogata*, 2013).

One problem with this more generalized formulation of ETAS is that there is no straightforward way to estimate $\mu(t)$ without making additional assumptions. Recall that for $\mu$ constant, the best-fitting ETAS parameters can be estimated by

minimizing the negative log-likelihood of the observed data $(t_i, M_i)$, as in Equation 3.2. While the negative log-likelihood (3.2) is a highly nonlinear function of $(t_i, M_i)$, given enough data ($N \gg 5$), the five ETAS parameters can be stably estimated. However, if $\mu$ is allowed to vary with time, the estimation procedure becomes highly unstable, as a truly continuous $\mu(t)$ has an infinite number of free parameters (*Snyder and Miller*, 1991). In essence, the data can be overfit by allowing $\mu(t)$ to become an arbitrarily rough function of time.

Our algorithm for computing the time-varying background seismicity rate $\mu(t)$ makes the physically plausible assumption that rate changes due to aseismic stresses and field-wide, anthropogenic forcing occur smoothly with time compared to the intense, step-like changes in seismicity rate characteristic of aftershock sequences. We thus impose a smoothness criterion, or regularization, on the background rate term $\mu(t)$, and seek to obtain the smoothest estimate of the time-varying background rate $\mu(t)$ that fits the data acceptably well. In doing so, we may lose resolution on very rough (short-term) fluctuations in background rate, but we ensure that any fluctuations that do appear in our inferred background rate are truly required by the observed data.

### 3.2.3 Regularized Inversion Algorithm

The seismicity rate model we employ is parameterized in terms of a background rate function $\mu(t)$ (nominally a continuous function of time) and a set $\theta_{ETAS}$ of ETAS parameters $[K, \alpha, p, c]$. To estimate the model parameters for a

given catalog of events $(t_i, M_i)$, we proceed as follows. We discretize $\mu(t)$ into $N_\mu$ entries, uniformly spaced in time with spacing $\Delta t$. Denoting this discretized rate function $\boldsymbol{\mu}$, we now wish to infer parameter values for the $N_\mu + 4$ entries in the parameter vector $\boldsymbol{\theta} = [\,\boldsymbol{\mu},\, K,\, \alpha,\, p,\, c\,]$. To do so, rather than minimize the negative log-likelihood function as in (3.2), we minimize a penalized negative log-likelihood function (*Good and Gaskins*, 1971; *Kumazawa and Ogata*, 2013) of the form:

$$\Psi(\boldsymbol{\theta}) = NLL(\boldsymbol{\theta}) + \gamma \cdot \Phi(\boldsymbol{\mu}), \tag{3.3}$$

where $NLL(\boldsymbol{\theta})$ refers to the negative log-likelihood defined in (3.2), $\Phi(\boldsymbol{\mu})$ is a roughness penalty function, and $\gamma$ is a regularization parameter that weights the relative importance between minimizing data misfit ($NLL$) and model roughness ($\Phi$). The regularization stabilizes the parameter estimation procedure, and prevents the inferred background rate $\boldsymbol{\mu}$ from becoming overly rough. For the purposes of this study, we use a first-derivative roughness norm as our penalty function:

$$\Phi(\boldsymbol{\mu}) = \sum_{j=1}^{N_\mu - 1} \left( \frac{\mu_{j+1} - \mu_j}{\Delta t} \right)^2 \Delta t. \tag{3.4}$$

We choose the regularization parameter $\gamma$ by finding the value of $\gamma$ that minimizes the effective Akaike Information Criteria (AIC) (*Akaike*, 1974) of the model:

$$\mathrm{AIC}(\boldsymbol{\theta}) = 2\,\kappa(\boldsymbol{\theta}) + 2\,NLL(\boldsymbol{\theta}), \tag{3.5}$$

where the effective model degrees of freedom $\kappa$ is estimated from the model roughness $\Phi$ (Equation 3.4). Conceptually, the AIC criterion provides a quantitative

assessment of model fit, with the minimum value corresponding to a theoretically optimal balance between effective model degrees of freedom ($\kappa$) and data misfit ($NLL$) (*Akaike*, 1974). However, the results presented here are not highly sensitive to the exact choice of $\gamma$, which can be chosen by a number of other standard approaches in inverse theory, including L-curve optimization or application of the discrepancy principle (*Snyder and Miller*, 1991; *Parker*, 1994; *Aster and Thurber*, 2013). Alternatively, the penalized negative log-likelihood can be interpreted in a fully Bayesian framework (*Akaike*, 1980) with a smooth prior distribution for $\boldsymbol{\mu}$, and $\gamma$ can be chosen by maximizing the posterior probability of the model parameters $\boldsymbol{\theta}$, given the observed data $(t_i, M_i)$ (*Tarantola*, 2005).

In our implementation of the regularized inversion algorithm, we minimize (3.3) using an iterative procedure. For a given data set, we first minimize the unpenalized negative log-likelihood function (3.2) to obtain an initial guess at the mean background rate ($\mu$), and the four other ETAS parameters ($\theta_{ETAS} = [K, \alpha, p, c\,]$). We then form an initial estimate for the time-varying background rate $\boldsymbol{\mu}$ by minimizing (3.3) with the four ETAS parameters held fixed. We then re-estimate the $\theta_{ETAS}$ by minimizing (3.3), now with $\boldsymbol{\mu}$ held fixed. We repeat this procedure — alternately re-estimating and fixing $\boldsymbol{\mu}$ and $\theta_{ETAS}$, respectively — until convergence (i.e., negligible change in estimated parameter values by additional iteration). We adopt this iterative approach rather than minimize (3.3) directly in order to mitigate the effect of strong parameter tradeoffs between $\boldsymbol{\mu}$ and $\theta_{ETAS}$ (*Hainzl et al.*, 2013), which can bias numerical optimization algorithms. For the

data sets used in this study, we find that the iterative algorithm converges rapidly ($\sim$ 3 iterations).

### 3.2.4 Uncertainty Estimates For Model Parameters

The nonlinearity in the inversion process precludes a straightforward mapping between data and model uncertainties. We therefore use a form of Monte Carlo error propagation (*Ogilvie*, 1984) to compute uncertainty estimates in the model parameters. For a given set of earthquake catalog data, we first apply the methods described above to obtain a seismicity rate model for that data set. We then create an ensemble of 50 synthetic earthquake catalogs through stochastic simulation of an ETAS-like Poisson process (*Daley and Vere-Jones*, 2003) with the time-varying background rate function and ETAS parameters derived from the seismicity rate model. Finally, by treating each synthetic earthquake catalog as "data" and applying the seismicity rate inversion scheme described in the previous subsections, we obtain a distribution of seismicity rate models (i.e., one for each synthetic catalog) that approximate the uncertainty and covariance structure of the seismicity rate model of the true (i.e., observed) earthquake catalog data.

### 3.2.5 Testing for Changes in $b$-value and Magnitude Distribution

We use two distinct methods to examine temporal changes in the magnitude distributions of earthquakes within each geothermal field. For the first method, we assume that the cumulative magnitude distribution of earthquakes in each field can be approximated by a Gutenberg-Richter distribution (*Gutenberg and Richter*, 1944) of the form $N \geq M = 10^{a-bM}$. We then compute the maximum likelihood estimate for the $b$-value, $b_{ml}$ (*Bender*, 1983; *Marzocchi and Sandri*, 2009):

$$b_{ml}(t_i) = \frac{1}{\log(10)\left[< M >_i - (M_c - \Delta M/2)\right]}, \tag{3.6}$$

at each earthquake occurrence time $t_i$ by selecting all events with $M \geq M_c$ within a moving window of 401 total events, centered on event $i$. Here $<M>_i$ denotes the mean magnitude of all events within the window centered on event $i$, while $\Delta M$ denotes the magnitude increment within the earthquake catalog. This first approach has the advantage of being able to resolve the timing of significant changes in the first-order features of the magnitude distribution, but its efficacy depends strongly on the validity of the underlying assumption that the magnitude distribution does in fact follow a Gutenberg-Richter power law.

To mitigate this concern, our second method examines the full frequency-magnitude distributions of earthquakes within each field, and allows for deviations from the standard power law at high magnitudes. For this method, we first divide the earthquake catalogs in each field into four 7-year, non-overlapping time

intervals. We then use a maximum likelihood technique to fit the magnitude distributions in each time interval with both standard Gutenberg-Richter (GR) and tapered Gutenberg-Richter model distributions (*Utsu*, 1999; *Kagan*, 2002). The standard GR model is characterized by a single scalar parameter (the $b$-value), while the tapered GR model is characterized both by a $b$-value and upper corner magnitude $M_x$ above which the magnitude distribution tapers to zero. We compare the fit of the two models to the observed earthquake data using an AIC test (Equation 3.5) that accounts both for the relative likelihood of the models and the difference in the number of free parameters (one for the standard GR model, and two for the tapered GR model). This second method is more powerful in resolving changes in the magnitude distribution at high magnitudes, but lacks the finer temporal resolution of the first method (which uses a moving-window approach).

## 3.3 Data and Resources

### 3.3.1 Earthquake Catalog Data

In this study, we use earthquake catalog data from two sources. For The Geysers, we use the waveform-relocated catalog of *Waldhauser and Schaff* (2008), which contains events in northern California from 1984 through 2011. For the Salton Sea and Coso geothermal fields, we use the waveform-relocated catalog of *Hauksson et al.* (2012), which contains events in southern California from 1981 through 2013. We first select all events within a rectangular box surrounding each

geothermal field (Figure 3.1). We then estimate the minimum magnitude of completion $M_c$ within each field for the years 1981–2013 (for Salton Sea and Coso) and 1984–2011 (for The Geysers) by applying the method of maximum curvature (*Woessner and Wiemer*, 2005) to five-year, non-overlapping windows of catalog data. Using this approach, we conservatively estimate that the catalog data for the Geysers, Salton Sea, and Coso are complete above magnitude 2.0, 2.0, and 1.5, respectively (Table 3.1) over their entire duration. As an additional test, we used a moving window approach to search for short-term transients in $M_c$. We found that $M_c$ increases during swarms and other active earthquake sequences, but does not exceed our cutoff values. While $M_c$ for each field is typically lower than these thresholds in more recent years, we use the more conservative $M_c$ values listed above to avoid introducing temporal biases into our seismicity analyses. Thus, we can interpret the seismicity rates we estimate as the rate of earthquakes of magnitude $M \geq M_c$ within the geothermal field boundary. We note here that the magnitudes from the Waldhouser-Schaff catalog for The Geysers are local magnitudes ($M_L$) as listed by the Northern California Seismic Network (NCSN) catalog, while the magnitudes from the Haukkson-Shearer relocated catalog for the Salton Sea and Coso are the Southern California Seismic Network (SCSN) preferred magnitudes. These are typically moment magnitude $M_w$ when available, and local or helicorder magnitude otherwise.

In choosing to model seismicity only within the rectangular boundaries surrounding each geothermal field, we make the implicit assumption that aftershock

triggering from earthquakes located outside each boundary has a relatively small impact on seismicity rates within the boundary itself. In cases where this assumption is violated, the inferred background seismicity rate will increase within the field (*Sornette*, 2005), so it is important to consider the influence of tectonic stresses from regional seismicity when interpreting transient increases in field-wide seismicity rates. As we will show below, regional seismicity near The Geysers is predominantly comprised of events within the geothermal field boundary, while there is evidence at both Salton Sea and Coso that regional seismicity influences local seismicity within the fields themselves.

### 3.3.2 Geothermal Operational Data

California state law requires geothermal fields within the state to release field-wide production (i.e., extracted fluid) volume and injection volume data to the California Department of Conservation's Department of Oil, Gas and Geothermal Resources (DOGGR) on a monthly basis. We use the monthly production and injection data from the DOGGR online database (available at http://www.conservation.ca.gov/dog/geothermal) for comparison with our seismicity rate models. For The Geysers, production and injection data extend back to year 1969, well before the start of the northern California relocated earthquake catalog (1984). For the Salton Sea and Coso geothermal fields, production and injection data extend back to years 1982 and 1986, respectively, both of which are after the start of the southern California relocated earthquake catalog (1981). Production

and injection data are available for each field through 2015, after the cessation of both relocated earthquake catalogs.

## 3.4 Results

### 3.4.1 The Geysers

#### Background

The Geysers, located approximately 120 km north of San Francisco (Figure 3.1), is California's largest geothermal field (*Geothermal Energy Association*, 2013; *Matek and Gawell*, 2014). Producing an average of 666 MW of electric power in 2014 (*Calpine*, 2014; *Boyd et al.*, 2015), The Geysers generates enough energy to supply nearly one million residents in northern California. Energy production at The Geysers began in 1960s and peaked in 1987 (*Adams*, 2011), after which production declined during the 1990s due to fluid depletion within the geothermal reservoir. Initially, the only form of fluid injection to take place was the disposal of wastewater from steam production at the field itself (*Adams*, 2011). In 1997, however, a pipeline was constructed to transport treated water from the Lake County Sanitation Plant for the purpose of fluid injection in the southeast sector of The Geysers (*Adams*, 2011; *Calpine*, 2014). This new pipeline both removed excess wastewater from the Lake County sewage system (which was over capacity in the wet winter months) while simultaneously helping to replenish reservoir vol-

umes at The Geysers through increased fluid injection. In 2003, a second pipeline to the northwest sector of The Geysers was created as part of the Santa Rosa Recharge Project (*Majer and Peterson*, 2007; *Vasco et al.*, 2013; *Martinez-Garzon et al.*, 2014), and the Santa Rosa pipeline now accounts for most of the injected wastewater volume at The Geysers (*Calpine*, 2014).

The geothermal reservoir at The Geysers is vapor-dominated (i.e., gas is the dominant phase within the reservoir), rather than liquid-dominated (as at Salton Sea and Coso), making The Geysers unique among California geothermal fields (*Matek and Gawell*, 2014). Electric power is produced from hot steam that is extracted from wells within the reservoir at depths of 1–3 km (*Calpine*, 2014). The 14 producing plants at The Geysers are concentrated within its northwest and southeast sectors. While the Northwest Geysers in particular has seen increased development in the years since the completion of the Santa Rosa pipeline (*Majer and Peterson*, 2007), total power production at The Geysers as a whole has declined since its peak in the late 1980s due to more concerted efforts to achieve a net fluid balance within the reservoir (*Boyd et al.*, 2015).

### Changes in Background Seismicity Rate

It is well-established that seismicity within The Geysers is modulated by geothermal plant operation (*Majer*, 1978; *Stark*, 1992; *Gomberg and Davis*, 1996; *Mossop and Segall*, 1999), and there was little historical seismicity in the region prior to the start of energy production in 1960 (*Oppenheimer*, 1986). Previous

studies have suggested that the total seismicity rate within the field has remained nearly constant since the 1980s, despite declining production volumes and changes to injection patterns (*Majer and Peterson*, 2007). More recently, changes in injection patterns after the completion of the Santa Rosa pipeline have motivated a number of studies to investigate the effects of isolated injection events using microseismic (*Martinez-Garzon et al.*, 2014) and surface deformation (*Vasco et al.*, 2013) data. Here we apply the methodology developed in Section 3.2 to analyze field-wide changes in seismicity from 1984–2011. Our seismicity rate model partitions the total seismicity rate $\lambda(t)$ into a time-varying background seismicity rate $\mu(t)$, and an interaction seismicity rate $\nu(t)$ that captures expected seismicity due to aftershock activity. The latter is parameterized in terms of a set of ETAS parameters $\theta_{ETAS} = [K, \alpha, p, c]$ that describe the intensity and temporal decay of typical aftershock sequences within the field.

The results of our seismicity rate inversion for The Geysers are summarized in Figure 3.2 and Table 3.2. The inferred background seismicity rate $\mu(t)$ in particular exhibits a number of interesting features. When averaged over timescales greater than a year or so, the background rate stays approximately constant at $\sim$ 0.7 earthquakes/day from 1984 to 2003, which is consistent with the widely held notion of a steady seismicity rate at The Geysers since the 1980s (*Majer and Peterson*, 2007). However, using our methodology we find that the background seismicity rate increases notably to an average of $\sim$ 1.0 earthquakes/day after 2003, when injection volumes increased due to the installation of the Santa Rosa

pipeline. Furthermore, we find that the background rate exhibits seasonal oscillations, with peaks in seismicity that are aligned with peaks in the monthly fluid injection volumes of the field (Figure 3.2d). Fluid injection at The Geysers is itself highly seasonal (*Majer et al.*, 2007; *Adams*, 2011), with higher injection volumes during the winter months when more wastewater is available. Seismicity follows this same seasonal pattern, with higher background seismicity rates during the winter than during the summer. These seasonal peaks in seismicity are particularly pronounced in 1997–98 (after the Lake County pipeline installation), and in the years since 2003 (after the Santa Rosa pipeline installation).

We quantify these observations more formally using step-wise linear regression analysis (*Mendenhall and Sincich*, 2007). Applying this technique to The Geysers over the full catalog duration (1984–2011), we find that the correlation coefficient between monthly injection and background seismicity rate is 0.61, with a vanishingly small p-value for the regression model ($1.5 \times 10^{-34}$, a value which represents the inferred probability that the correlation could be zero, given the data). These statistical tests strongly corroborate the visual correlation apparent in Figure 3.2. Notice that while production rates at The Geysers followed the same seasonal pattern as injection during the 1980s and 1990s, in more recent years production has been nearly constant, with only injection following a seasonal pattern (Figure 3.2d). Seasonality in background seismicity rate becomes even more pronounced in these more recent years, indicating that stresses from injection, and not production, drive the observed seasonal patterns in seismicity at The Geysers.

**Changes in $b$-value, Magnitude Distribution, and Depth Distribution**

Temporal changes in the magnitude and depth distributions of earthquakes can provide additional information about the evolution of stresses within The Geysers beyond that provided by analysis of seismicity rate changes alone. As a preliminary test for changes in magnitude distribution at The Geysers, we apply the first method detailed in Section 3.2.5 to compute maximum-likelihood $b$-value estimates over 401-event moving windows. We observe an increase in $b$-value beginning in late 2003 (Figure 3.3a), coinciding with the construction of the Santa Rosa pipeline. It is important to note, however, that this approach maps any potential change in the overall magnitude distribution into a change in a single scalar parameter, $b$. As such, if one looks only for changes in $b$-value, one may miss more subtle changes in the magnitude distribution as a whole. Indeed, when we divide the 28-year catalog for The Geysers into four 7-year intervals and directly compare the magnitude distributions of each interval, the most obvious visual change is not a change in $b$-value, but rather a noticeable increase in large ($M \geq 3.5$) events in recent years (Figure 3.3b).

To quantify the significance of this observation, we use the second method described in Section 3.2.5 to fit the full magnitude distributions in each time interval with both standard Gutenberg-Richter (GR) and tapered Gutenberg-Richter model distributions (*Utsu*, 1999; *Kagan*, 2002). The standard GR model is characterized by one parameter: the $b$-value, while the tapered GR model is characterized

by two: the $b$-value and a upper corner magnitude $M_x$ above which the magnitude distribution tapers to zero. At The Geysers, we find that the $b$-value is relatively constant during the first three time intervals, but increases from 1.13 ($2\sigma$ confidence interval of 1.08–1.18) during the third (1998–2004) to 1.26 (1.21–1.31) during the fourth (2005–2011). We also find that the corner magnitude of the tapered GR model increases with time, from 3.64 (3.50–3.78) during 1984–1991 to 4.63 (4.00–5.26) during 2005–2011. This latter estimate should be interpreted with some caution, however, as the tapered GR distribution provides a better model fit (lower AIC) only during the first three time intervals. Nevertheless, it appears that while the aforementioned increase in $b$-value is significant, but is an incomplete description of the change in the magnitude distribution within The Geysers. Indeed, from a seismic hazard perspective, the increased frequency of large earthquakes in The Geysers is the more important temporal change to consider. We explore possible mechanisms responsible for these changes in Section 3.5.

Like the magnitude distribution, the depth distribution of earthquakes at The Geysers also changes significantly with time. To show this, we divide the catalog into the same four equal time intervals as before, and compare the probability density and cumulative density functions of earthquake depths during each interval (Figure 3.4). Since the error distribution for absolute hypocentral depth is not well-constrained in the relocated earthquake catalogs, and may evolve with time, we use a bootstrap technique (*Efron and Tibshirani*, 1994) to compute 95-percent confidence intervals for each depth distribution and analyze only the most

robust features of the depth distributions. During all time intervals, the earthquake depth distribution at The Geysers is bimodal, with seismicity clustered around both shallower (1–3 km) and deeper (3–5 km) peaks in the depth distribution. The shallower peak corresponds to seismicity within and around the geothermal reservoir itself. The deeper cluster in seismicity consists almost exclusively of events within the Northwest Geysers. Previous studies (*Stark*, 2003) have suggested that these deeper events may correspond to earthquakes induced by thermoelastic stresses from injected fluid that has diffused from the geothermal reservoir (which is composed predominantly of greywacke rocks) into the metamorphosed, hornfelsic metagreywacke rocks in the high-temperature zone (HTZ) beneath the reservoir (*Jeanne et al.*, 2014; *Boyle and Zoback*, 2014).

While earthquakes at The Geysers exhibit this characteristic bimodal depth distribution during all time periods, the overall shape of the distribution (i.e., the location, width, and intensity of the peaks in the distribution) has evolved over time. Most notably, the shallow peak in seismicity has become both sharper and deeper with time (Figure 3.4), with the shallow peak shifting from its initial position of $1.45 \pm 0.3$ km depth (1984–1990) to $2.45 \pm 0.1$ km depth in the most recent years (2005–2011). This transformation may be caused by changes in geothermal plant operation at The Geysers. In particular, the increase in injection volumes in the Northwest Geysers from the Santa Rosa pipeline likely caused the shallower seismicity to cluster around the mean injection well depth of 2.5 km in that sector. Moreover, while the deeper cluster in seismicity (3–5 km) does not exhibit a shift

in peak location with time, it is most prominent in these later years, when injection at the Northwest Geysers was at its most prevalent.

### 3.4.2 Salton Sea Geothermal Field

#### Background

The Salton Sea Geothermal Field (SSGF) is located in southeastern California, on the southern shore of the Salton Sea. The surrounding region, known as the Salton Trough, is tectonically active and encompasses the Pacific-North America plate boundary as it transitions from an extensional spreading center in the Gulf of California to a predominantly strike-slip boundary in central and northern California. The SSGF itself lies within the extensional stepover between the Imperial and Southern San Andreas Faults (*Muffler and White*, 1969; *Younker et al.*, 1982; *Crowell et al.*, 2013), with the San Jacinto and Superstition Hills Fault Zones located to the west of the SSGF (Figure 3.1). Earthquake swarms — spatially and temporally clustered earthquake sequences with no clear mainshock — are relatively common occurrences in the region, with the 1981 Westmorland Swarm, 2005 Obsidian Buttes Swarm, and 2012 Brawley Swarm (each with multiple M5+ events) being the most notable examples (*Vidale and Shearer*, 2006; *Lohman and McGuire*, 2007; *Hauksson et al.*, 2013). Regional fault systems also produce large earthquakes that follow more typical mainshock-aftershock sequences, such as the 1987 M6.6 Superstition Hills (*Hudnut et al.*, 1989) and the 2010 M7.2 El Mayor-

Cucapah (*Wei et al.*, 2011) earthquakes. The El Mayor-Cucapah event ruptured a set of faults within the Cucapah Mountains in Baja California (*Fletcher et al.*, 2014; *Gonzalez-Ortega et al.*, 2014), but its extensive aftershock sequence extended northward into the Salton Trough (*Hauksson et al.*, 2011).

The SSGF reservoir is liquid-dominated, and the hot brine that exists in-situ in the reservoir must therefore be flashed to steam to produce electricity. Energy production at the SSGF did not begin until the 1980s (more than twenty years after the initial development of The Geysers), in part due to the highly saline nature of the geothermal brine in the SSGF reservoir (*Adams*, 2011), which rapidly corroded much of the initial production well and pipeline infrastructure. Technological advances helped mitigate these early problems, making energy production more cost-efficient and allowing for further development of the field. Construction of the geothermal plant was mostly complete by the early 1990s, though additional wells have been regularly constructed in the years since. The current energy production capacity of the SSGF is 437 MW (*Matek and Gawell*, 2014), but there is still significant potential for future growth (*Adams*, 2011), as the SSGF's vast geothermal reservoir is estimated to be capable of producing up to 1800 MW of power (*Matek and Gawell*, 2014). However, the high up-front costs associated with energy production in the Salton Trough, along with the inherent difficulties of energy transport to more populous regions of California, may prevent further expansion of the SSGF.

**Changes in Background Seismicity Rate**

In Figure 3.5, we show the results of applying our method to compute seismicity rate changes (Section 3.2) within the SSGF. We use the relocated earthquake catalog of *Hauksson et al.* (2012), which runs from 1981 through 2013 and thus contains events that occurred both before and during geothermal plant operation. We find that the background seismicity rate increases from a mean value of 0.01 earthquakes/day in 1981–1985, prior to geothermal plant operation, to a mean value of $\sim$ 0.04 earthquakes/day in 1987, roughly a year after production began (Figure 3.5c). From year 1990 onward, the background seismicity rate remains relatively steady about this mean value, with occasional, but transient, increases or decreases in seismicity rate. These short term fluctuations do not appear to be correlated with changes in operation at the SSGF field (Figure 3.5d). Instead, these fluctuations (such as the short-term rate increase observed in 2000–2001), seem more consistent with swarm-like earthquake sequences (*Lohman and McGuire*, 2007) driven by transient, but natural, tectonic stresses (*Llenos et al.*, 2009). We discuss this hypothesis further in Section 3.5.

We confirmed the apparent lack of correlation in the later years more formally by performing step-wise linear regression analysis (*Mendenhall and Sincich*, 2007). In so doing, we found no evidence for any systematic relationship between background seismicity rate and any linear combination of monthly production or injection volumes at the SSGF in the years since 1990. Applying this same tech-

nique to the years preceding 1990, however, we find that the background seismicity rate is correlated with a linear model comprised of both injection and net production (production minus injection) volumes, with a model correlation coefficient of 0.74 (and p-value for the regression model of $1.27 \times 10^{-18}$). This correlation in these early years of plant operation is consistent with our previous observation of a coincident rise in background rate and energy production during the late 1980s, but it is important to note that the relationship between the two may well be nonlinear.

*Brodsky and Lajoie* (2013) used a similar approach based on the ETAS model to compare temporal changes in seismicity rate to geothermal production and injection data from the SSGF. Although their methodology differs somewhat from ours, their background seismicity rate model also exhibits a robust rate increase that begins shortly after initial energy production at the Salton Sea. Moreover, *Brodsky and Lajoie* (2013) show that both net production and injection data, rather injection data alone, are necessary to explain changes in seismicity rate during this time. *Brodsky and Lajoie* (2013) also find that the relationship between seismicity rate and geothermal production and injection is less straightforward after 1990, although they proceed to construct a linear model of injection, net production, and seismicity rates by allowing for time-variations in the regression coefficients. We do not attempt to recreate a complex model of this type here because of the inherent instabilities in fitting a large number of free parameters.

**Changes in $b$-value, Magnitude Distribution, and Depth Distribution**

We also find evidence for significant time-variations in the magnitude and depth distributions of earthquakes within the SSGF. Using the approach detailed in Section 3.2.5, we first test for temporal changes in $b$-value by computing maximum likelihood estimates over 401-event moving windows. The $b$-value appears to slowly increase after production ramps up in the late 1980s, before experiencing a sharp decline in 2000–2001 (Figure 3.6a). This decline in $b$-value coincides with the earthquakes swarms that occur within the field during 2000 and 2001, with another sharp drop in $b$-value occurring during the prominent 2005 Obsidian Buttes Swarm. The slow increase and subsequent sharp decrease in $b$-value are also apparent when one examines the overall magnitude distributions for different time intervals (Figure 3.6b). In fitting both standard and tapered GR model distributions to each of the four intervals (Section 3.4.1), we observe a decline in $b$-value from 1.06 ($2\sigma$ confidence interval of 0.97–1.14) during the third interval (2000–2006) to 0.91 (0.79–1.03) during the fourth (2007–2013). Interestingly, the standard GR model outperforms the tapered GR model (as measured by model AIC values) in all but the latest interval, which has an apparent corner magnitude $M_x$ of 3.67 (3.42–3.92) that is significantly lower than any previous interval.

Changes in the depth distribution of earthquakes likewise suggest that both tectonic and anthropogenic stresses cause time-dependent changes in seismicity within the SSGF. In Figure 3.7, we compare earthquake depths in the SSGF over

the same four time intervals. The depth distribution transitions from unimodal with a median value of $5.38 \pm 0.13$ km during the first interval (1986–1992), to bimodal, centered around that same depth, during the second interval (1993–1999). It is notable that the dominant fraction of these events occurs at depths greater than 4 km, which is far beneath typical reservoir depths of 1.0–2.5 km (*Younker et al.*, 1982). During the third interval (2000–2006), the depth distribution includes a greater proportion of deep events (6–8km) than in prior years, possibly a result of tectonic stresses associated with the 2000–2001 and 2005 swarms. During the final interval (2007–2013), however, the depth distribution both becomes notably shallower (median depth of $3.52 \pm 0.15$ km) and is strongly bimodal. These most recent years correspond to the highest levels of production and injection at the SSGF, and much of the seismicity is concentrated just beneath the reservoir.

### 3.4.3 Coso Geothermal Field

#### Background

Coso Geothermal Field (CGF) is located in eastern California, between the Southern Sierra Nevada and the Argus Range (Figure 3.1). This region is part of a broad transition zone between the strike-slip tectonics of the San Andreas Fault and the extensional tectonics of the Basin and Range province (*Duffield et al.*, 1980). The CGF itself lies within an extensional stepover between the Little Lake Fault to the southwest and a set of unnamed faults to the northeast, with the Airport Lake

Fault striking northward through the field (*Monastero et al.*, 2005). These faults, while active, accommodate only 6.5 mm/year of dextral offset between the Argus Range and the Sierra Nevada (*McClusky et al.*, 2001). Despite this relatively low tectonic deformation rate, the CGF and surrounding region is seismically active, and like the Salton Trough, prone to earthquake swarms (*Bhattacharyya and Lees*, 2002). The most notable swarm in recent times occurred during the summer of 2001, culminating in a magnitude 5.2 earthquake just west of the geothermal field.

Coso Geothermal Field is California's third largest, with an installed capacity of 302 MW in 2014 and potential for more than 200 MW of expansion (*Matek and Gawell*, 2014). A shallow magmatic chamber (5–8 km depth) acts as a heat source for the CGF's fluid-dominated geothermal reservoir, which ranges in depth from 0.5 to 3.5 km (*Duffield et al.*, 1980; *Manley*, 2000). While privately owned, the CGF is located on the premises of the Naval Air Weapons Station at China Lake, and is collaboratively managed by the Navy's Geothermal Program Office. Energy production at the CGF began in 1987 after a thorough review of the Coso area's potential as a geothermal resource (*Combs*, 1980), with further plant construction taking place until 1990 (*Monastero*, 2002). Fluid production and injection volumes peaked in 1990 and have slowly declined in the years since, although technological improvements have allowed for consistent levels of energy production (*Matek and Gawell*, 2014). Net fluid extraction at the CGF causes the surface to subside locally, creating poroelastic stresses within and near the reservoir (*Fialko and Simons*, 2000), and depleting the total fluid volume available for

future energy production.

### Changes in Background Seismicity Rate

Both tectonic and anthropogenic stresses are thought to contribute to observed seismicity at the CGF in the years since the plant's initial construction in 1987 (*Feng and Lees*, 1998; *Monastero et al.*, 2005; *Schoenball et al.*, 2015). We use the relocated earthquake catalog of *Hauksson et al.* (2012) from 1981–2013 and the methods described in Section 3.2 to construct a seismicity rate model for the CGF (Figure 3.8). As was observed at the Salton Sea Geothermal Field, the background seismicity rate within the CGF increases during the initial phase of energy production, with the background rate rising from a mean value of 0.02 earthquakes/day prior to 1987 to 0.05 earthquakes/day by 1990. After this time, the background seismicity rate fluctuates about this mean value, with short-term rate transients that do not appear to be modulated by geothermal injection or production (Figure 3.8d).

Application of a step-wise linear regression procedure (*Mendenhall and Sincich*, 2007) supports these observations. For 1987–1989, the optimal linear regression model obtained from this technique suggests that the background seismicity rate depends strongly on injection and production volumes during these years (correlation coefficient: 0.64, regression model p-value: $2.9 \times 10^{-6}$). This result is consistent with our observation of a persistent increase in background rate starting during this time period. From 1990 on, however, the correlation is much weaker

(correlation coefficient: 0.18), though not entirely insignificant, at least from a statistical perspective (regression model p-value = 1.7 $\times 10^{-3}$). Note that because injection and production volumes track each other so closely at the CGF (Figure 3.8d), it is impossible to statistically distinguish whether one variable contributes more significantly than the other to potentially induced seismicity at the CGF.

While energy production does appear to have elevated the mean background seismicity rate at the CGF, the most prominent transient peaks in seismicity (e.g., during late 1994, 1999, and 2001) correspond to earthquake swarms and contain events both inside and far outside the field boundaries. The most prominent of these swarms took place during the summer of 2001, with a extended sequence of $M \geq 3$ earthquakes that occurred within and to the west of the CGF (*Hauksson and Unruh*, 2007). The largest event in this swarm was a M5.2 earthquake that took place on July 17th, 2001, and was located to the west of the field. The presence of these and other similar, far-reaching earthquake sequences are a reminder that the Coso area was seismically active long before the construction of the CGF (*Walter and Weaver*, 1980).

**Changes in *b*-value, Magnitude Distribution, and Depth Distribution**

The 2001 swarm also coincides with a fundamental change in the magnitude distribution of earthquakes within the CGF. Applying the moving-window approach detailed in Section 3.2.5, we find that the *b*-value for earthquakes at the CGF has declined significantly since 2001 (Figure 3.9a). This change is also ap-

parent in comparing the overall magnitude distributions of 7-year, non-overlapping time intervals (Figure 3.9b). If we proceed as before and fit both standard and tapered GR model magnitude distributions to each of the four time intervals (Section 3.4.1), the inferred $b$-value of 0.79 (and $2\sigma$ confidence interval of 0.69–0.89) during 2007–2013 is much less than any of the previous intervals, which have $b$-values in the range of 1.2–1.3. Further, during the most recent interval (2007–2013), larger earthquakes ($M \geq 3$) become more frequent both in relative and absolute terms, despite no appreciable change in seismicity rate since 1990. Both of the two recorded $M \geq 4$ earthquakes to occur within the field also took place during this time interval (a M4.1 event in 2007, and a M4.4 event in 2010), and are the only two $M \geq 4$ events to occur within the field boundary in the past 30 years (*Kaven et al.*, 2013). These observations can also be quantified in terms of an increase in apparent upper corner magnitude $M_x$, which rises from 2.71 (2.41–3.00) during the first interval to 4.37 (3.71–5.04) during the final interval. We emphasize, however, that these corner magnitude estimates do not necessarily indicate a true deviation of the magnitude distribution from a standard Gutenberg-Richter power law, as AIC tests suggest that the data are actually better explained by the standard GR model (which has no corner magnitude) than the tapered GR model (which does).

The depth distribution of earthquakes within CGF also changes systematically with time. The median earthquake depth increases during each successive time interval (Figure 3.10), from a low of $1.09 \pm 0.10$ km during the early years of energy production (1986–1992) to a high of $2.31 \pm 0.08$ km during the final

interval (2007–2013). This progressive shift in earthquake depth is puzzling, and may in part be related to changes in production and injection volumes, both of which have been steadily declining since 1990 (Figure 3.8d). However, the deeper seismicity in recent years is comprised predominantly of the earthquake sequences that are spatially clustered around the locations of the 2007 M4.1 and 2010 M4.4 events. Changes to the depth distribution may therefore reflect an altered stress state within the field in the aftermath of the 2001 swarm and the subsequent M4 events, though anthropogenic forcing may play a secondary role.

## 3.5 Discussion and Comparison of Geothermal Fields

### 3.5.1 Changes in Seismicity Rate

In the preceding section, we presented the results of our analysis of changes in background seismicity rates at The Geysers, Salton Sea, and Coso geothermal fields. The background seismicity rate at The Geysers is strongly correlated with fluid injection volume, and we observe significant rate increases associated with the installation of the Lake County and Santa Rosa pipelines. On shorter timescales, the seismicity rate exhibits a seasonal pattern that tracks the seasonality in available injection volumes. At the Salton Sea and Coso geothermal fields, the start of the earthquake catalog we use for our analyses predates initial plant operation,

and we observe a significant increase in the background seismicity rate as operation commences. In later years, both the Salton Sea and Coso fields exhibit short-term fluctuations in background rate, as at The Geysers. However, in contrast to The Geysers, these fluctuations at Salton Sea and Coso are not strongly correlated with any discernible feature in monthly fluid injection or production volumes. We explore possible explanations for these differences below.

Analyses of focal mechanisms (*Oppenheimer*, 1986; *Boyle and Zoback*, 2014) indicate that earthquakes within The Geysers occur on preexisting structures that are well-aligned with the regional stress field. This suggests that regional tectonic stresses help supply much of the necessary background stress (as well as the in-situ fault structures) for earthquakes to occur within The Geysers. However, as noted by *Oppenheimer* (1986) and others, there was little seismicity in and around The Geysers prior to the operation of the geothermal field. The intense cyclical pattern of fluid injection at The Geysers (which can exceed production during the winter months) may therefore provide the necessary stress perturbation to directly induce shear slip on faults that are near-critically stressed by regional tectonics. Further, the extensive and transient fluid injection volumes, when combined with high in-situ reservoir temperatures (*Stark*, 2003; *Jeanne et al.*, 2014) create localized thermoelastic stresses that greatly exceed tectonic stresses (*Altmann et al.*, 2013). Indeed, *Martinez-Garzon et al.* (2013) observed a local temporal correlation between stress tensor orientation and injection rate for clusters of seismicity around selected injection wells in the Northwest sector of The Geysers. Our results

further demonstrate that at The Geysers, these transient anthropogenic stresses can plausibly modulate seismicity, even on a field-wide basis.

Nearly all of the seismicity in the region surrounding The Geysers is concentrated within the field itself (Figure 3.1). There is little observational evidence therefore that tectonic stress transients could account for the observed temporal changes in seismicity at The Geysers, though regional tectonic stresses undoubted contribute to slow, steady-state loading of the in-situ fault structures. In contrast, both Salton Sea and Coso are located in dynamic tectonic regions with active fault systems that surround the geothermal fields (*Younker et al.*, 1982; *Walter and Weaver*, 1980). As a result, seismicity within the two geothermal fields is strongly influenced by natural processes of time-dependent tectonic stress transfer from regional earthquake activity, including the aseismic stress transients associated with the earthquake swarms characteristic of the Salton Trough and Coso Range. The elevated background seismicity rates observed at the Salton Sea (e.g., during 2000–2001) and Coso (e.g., during 1999 and 2001) correspond to earthquake swarms that are regional in scale and extend well beyond the boundaries of the geothermal field. These fluctuations in seismicity rate due to natural (tectonic) stress transients tend to mask subtle changes in seismicity rate caused by anthropogenic stresses, especially since fluid injection and production volumes have remained relatively stable at both fields since their initial construction (unlike at The Geysers).

While our study has focused primarily on the background seismicity rate,

it is also instructive to compare the set of inferred ETAS parameters (Table 3.2) that describe typical aftershock activity within each field. The Geysers in particular is notable for its anomalously low aftershock productivity parameter $K$, indicating that most earthquakes within The Geysers do not appear to trigger extensive aftershock sequences. We can formalize this observation by computing the branching ratio $R_b$, defined as the ratio of the expected number of triggered events (aftershocks) to total events, for each field. The inferred branching ratio of 0.09—0.11 at the Geysers is well below the 95% confidence interval for typical California seismicity (*Hardebeck*, 2014). The low aftershock productivity may indicate that most earthquakes at The Geysers are directly triggered by highly localized, anthropogenic stresses, rather than the more broadly distributed coseismic stresses from other earthquakes within the field. It has also been noted previously that earthquakes in northern California tend to have have lower aftershock productivity than earthquakes in southern California (*Llenos*, 2014; *Brocher et al.*, 2015), so the very low aftershock productivity at The Geysers may be the consequence of both regional and local (i.e., energy production-related) factors. The inferred aftershock productivity (and hence, branching ratio) is higher at both the Salton Sea ($R_b = 0.69$–$0.79$) and Coso ($R_b = 0.42$–$0.52$) and is more typical of California seismicity. Both the Salton Sea and Coso, have low inferred ETAS $\alpha$ parameters, as is common in regions prone to earthquake swarms (*Hainzl*, 2002; *Kumazawa et al.*, 2010; *Hainzl et al.*, 2013).

### 3.5.2 Changes in $b$-value, Magnitude Distribution, and Depth Distribution

At The Geysers, we observed a fundamental change in the earthquake magnitude distribution that coincided with the installation of the Santa Rosa pipeline in 2003. The significant increase in $b$-value at this time is consistent with observational (*Shapiro et al.*, 2011; *Bachmann et al.*, 2011) and theoretical (*Bachmann et al.*, 2012; *Segall and Lu*, 2015) studies that suggest that induced earthquakes tend to have higher $b$-values than natural earthquakes. However, the change in the magnitude distribution is more subtle than a simple change in $b$-value, as the rate of large ($M \geq 4$) earthquakes also increases during this time. This latter phenomenon may be explained in part by the increased concentration of events in the deeper northwest sector of The Geysers after 2003, which appears to be characterized by a different magnitude distribution than the southeast sector (*Convertito et al.*, 2012). *Kwiatek et al.* (2015) observed no significant variation in $b$-value from 2007–2014 within certain clusters of seismicity within the Northwest sector of the Geysers, which suggests that any field-wide changes in $b$-value that may occur during these years are primarily due to an increased frequency of events occurring in the Northwest sector, relative to the Southeast sector, rather than changes within the Northwest sector itself. We note, however, that most of the observed field-wide increase in $b$-value occurs from 2003–2006 (Figure 3.3a), prior to the study of (*Kwiatek et al.*, 2015). In addition to these longer-term trends, there is some

evidence (Figure 3.3a) that the $b$-value of earthquakes at The Geysers evolves on shorter timescales (perhaps even seasonally), a topic examined in more detail by *Convertito et al.* (2012) and *Martinez-Garzon et al.* (2014).

Earthquakes at the Salton Sea Geothermal Field, in contrast, appear to exhibit a slow increase in $b$-value before year 2000, followed by steep declines in $b$-value following earthquake swarms in 2000–2001 and 2005. Although the initial increase in $b$-value following the beginning of energy production is consistent with previous studies of induced seismicity within geothermal fields (*Bachmann et al.*, 2011), the subsequent decline coincident with earthquake swarm activity suggests that natural tectonic processes (like those underlying swarm activity) strongly influence seismicity at the SSGF. Indeed, one would expect just such a decrease in $b$-value during the earthquake swarms that characterize the Salton Trough, which tend to consist of multiple large earthquakes (with no clear mainshock) and a relative deficiency of smaller earthquakes (*Lohman and McGuire*, 2007; *Chen and Shearer*, 2011).

Similarly, at Coso Geothermal Field, the $b$-value remained relatively steady until 2001, when a prominent earthquake swarm occurred within and surrounding the field. The 2001 swarm included a $M5.2$ earthquake located to the west of the field, as well as a number of other $M \geq 3$ events both within and around the field. The apparent change in the magnitude distribution following this sequence is striking (Figures 3.9), and suggests that the swarm fundamentally altered the local stress state. It is notable that the largest earthquakes to occur within the

field (including the only two $M \geq 4$ events) occurred either during or after the 2001 sequence.

Temporal changes to the depth distribution of earthquakes must be treated with some caution, as absolute errors in the hypocentral depths of the smaller earthquakes can be substantial. Nevertheless, while the overall depth distributions of each field are likely controlled by local geology and tectonics, the more robust features of the temporal changes in the depth distribution may reflect many of the same physical processes that control temporal changes in seismicity rates or magnitude distributions. For example, the deepening of the shallow peak of the bimodal distribution that characterizes The Geysers (Figure 3.4) is likely related to the increased concentration of events in the deeper northwest sector of The Geysers following the construction of Santa Rosa pipeline. The changes to the depth distributions at Salton Sea and Coso geothermal fields are more complex, and undoubtedly modulated by both tectonic and anthropogenic stresses.

## 3.6 Concluding Remarks

The results we have presented here for The Geysers, Salton Sea, and Coso geothermal fields are concerned with long-term and field-wide changes in seismicity. It is well-established that individual episodes of fluid injection or production can trigger microseismicity over short spatial and temporal scales (*Ellsworth*, 2013). Here our central focus is not microseismicity, but rather the changing patterns of

seismicity that can be observed over larger spatial scales (tens of kilometers) and longer time periods (months to decades). In seeking the smoothest model of background seismicity rate that adequately explains the observed earthquake catalog, our method is tailored to study robust changes in seismicity that occur over these same spatial and temporal scales. While we lack the resolution to explore highly transient or spatially isolated changes in seismicity, a uniform application of our methodology allows us to quantitatively compare temporal changes in seismicity at each geothermal field. In so doing, we can better assess the range of plausible mechanisms that may account for the differing responses of each field to tectonic and anthropogenic stressing.

At The Geysers, field-wide seismicity appears to be strongly controlled by anthropogenic stressing due to fluid injection and production. The mean background seismicity rate at The Geysers increased by 50 percent following the construction of the Santa Rosa pipeline in 2003 (and the associated increase in injection volume). Likewise, background seismicity rate increased after the installation of the Lake County pipeline in 1997, though this rate increase decayed with time as injection volumes returned to pre-1997 levels. Seasonal injection patterns at The Geysers — a consequence of more available fluid volume during the winter months — result in seasonal fluctuations in seismicity rate, which have become more intense in the years since the installation of the Santa Rosa pipeline. Fluid injection patterns also seem to influence the $b$-value, magnitude distribution, and depth distribution of earthquakes within The Geysers, all of which show significant

temporal changes that coincide with installation of the Santa Rosa pipeline.

The relationship between observed seismicity and geothermal energy production is less straightforward at the Salton Sea and Coso geothermal fields. At both the Salton Sea and Coso, the mean background seismicity rate increases during the initial phase of geothermal energy production, indicating anthropogenic stresses have an observable influence on field-wide seismicity. However, the most notable rate increases at both the Salton Sea and Coso are related to earthquake swarms that are regional in scale and extend far beyond the boundaries of the geothermal fields themselves. Earthquake swarms were prevalent in the Salton Sea and Coso regions before energy production began (*Lohman and McGuire*, 2007; *Bhattacharyya and Lees*, 2002), which suggests that the current swarms, which are not temporally correlated with changes to fluid injection or production, are caused by natural, rather than anthropogenic, stress transients. In any case, the presence of these swarms influences not only the seismicity rate, but also the inferred $b$-values, magnitude distributions, and depth distributions of earthquakes at both the Salton Sea and Coso geothermal fields.

Viewed holistically, our observations indicate that anthropogenic stresses are the primary control on transient changes in seismicity at The Geysers, while they play a less prominent role in modulating seismicity at Salton Sea and Coso. This is not to suggest that induced seismicity is of negligible concern at Salton Sea and Coso. Indeed, the proximity of the Salton Sea and Coso geothermal fields to active faults that are capable of producing large, damaging earthquakes amplifies

the importance of even small changes in seismicity due to anthropogenic stresses (*Convertito et al.*, 2012; *Brodsky and Lajoie*, 2013). Rather, these observations highlight the complicated process of differentiating induced from natural seismicity in California's geothermal fields, and demonstrate that differences in tectonic setting, reservoir conditions, and history of energy production all likely contribute to the differing patterns of seismicity observed within each geothermal field.

# Acknowledgements

The relocated earthquake catalog for northern California (including The Geysers) is available from the NCEDC website (http://www.ncedc.org/ncedc/catalog-search.html), while the relocated earthquake catalog for southern California (including the Salton Sea and Coso) is available from the SCEDC website (http://scedc.caltech.edu/research-tools/alt-2011-dd-hauksson-yang-shearer.html). Fluid projection and injection data for The Geysers, Salton Sea, and Coso are available from the California Department of Conservation DOGGR website (http://www.conservation.ca.gov/dog/geothermal).

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program (NSFGRFP) under grant number DGE-1144086. Additional support was provided by the Southern California Earthquake (SCEC) grant number 14073. Map-based figures for this work were generated using the Generic Mapping Toolbox (GMT) (*Wessel et al.*, 2013). We thank

the Associate Editor and reviewers for their insightful comments, which greatly improved the manuscript, both in content and clarity. We also thank W. Wang, W. Neely, and Q. Zhang for illuminating internal discussions regarding early versions of this work.

Chapter 3, in full, is a reformatted version of material as it appears in Journal of Geophysical – Solid Earth: Trugman, D. T., P. M. Shearer, A. A. Borsa and Y. Fialko (2016), A comparison of long-term changes in seismicity at The Geysers, Salton Sea, and Coso geothermal fields, *Journal of Geophysical Research: Solid Earth*, 121(1), 225–247, doi:10.1002/2015JB012510. I was the primary investigator and author of this paper.

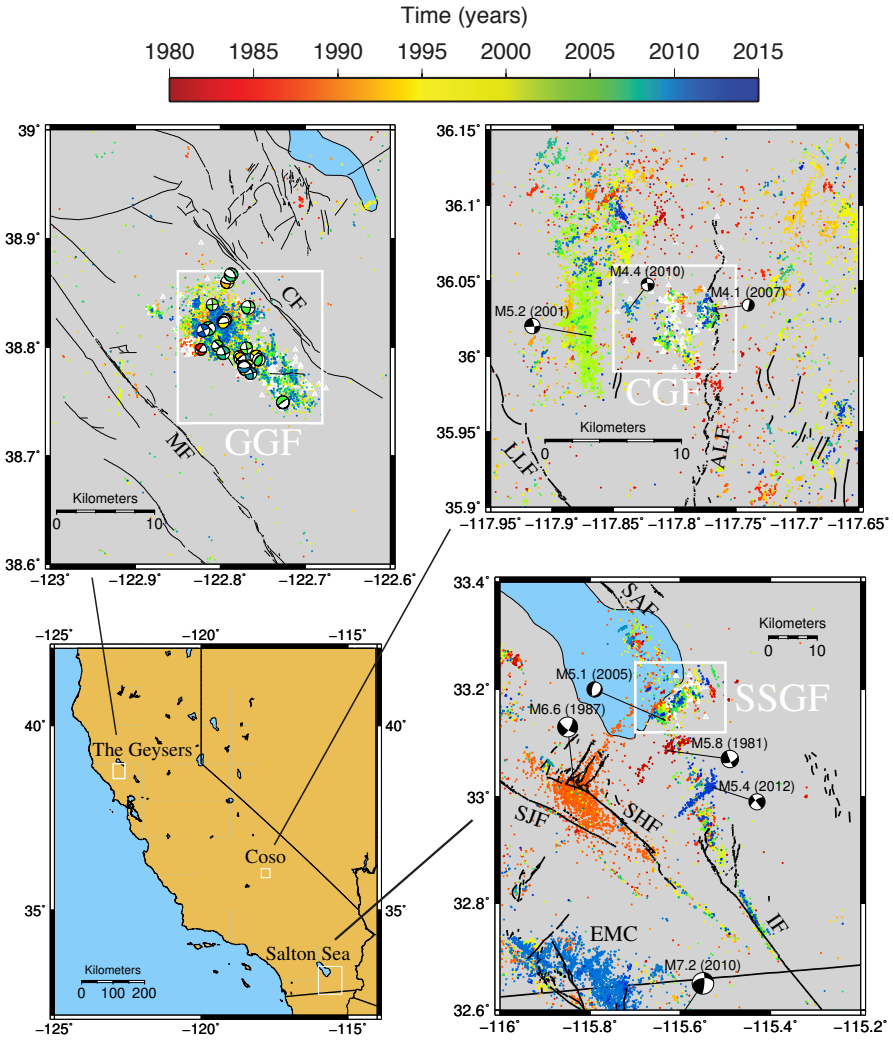


**Figure 3.1**: Map view of seismicity at the Geysers, Coso and Salton Sea geothermal fields. Earthquake locations are from the catalogs of *Waldhauser and Schaff* (2008) and *Hauksson et al.* (2012), with each event color-coded by its occurrence time (years). The study area boundaries for the Geysers (GGF), Coso (CGF), and Salton Sea (SSGF) geothermal fields are outlined in white, with well locations from the DOGGR database (http://www.conservation.ca.gov/dog/geothermal) denoted with white triangles. Quaternary faults (http://earthquake.usgs.gov/hazards/qfaults) are outlined in black. Focal mechanisms for $M \geq 4$ within GGF are color-coded by year.

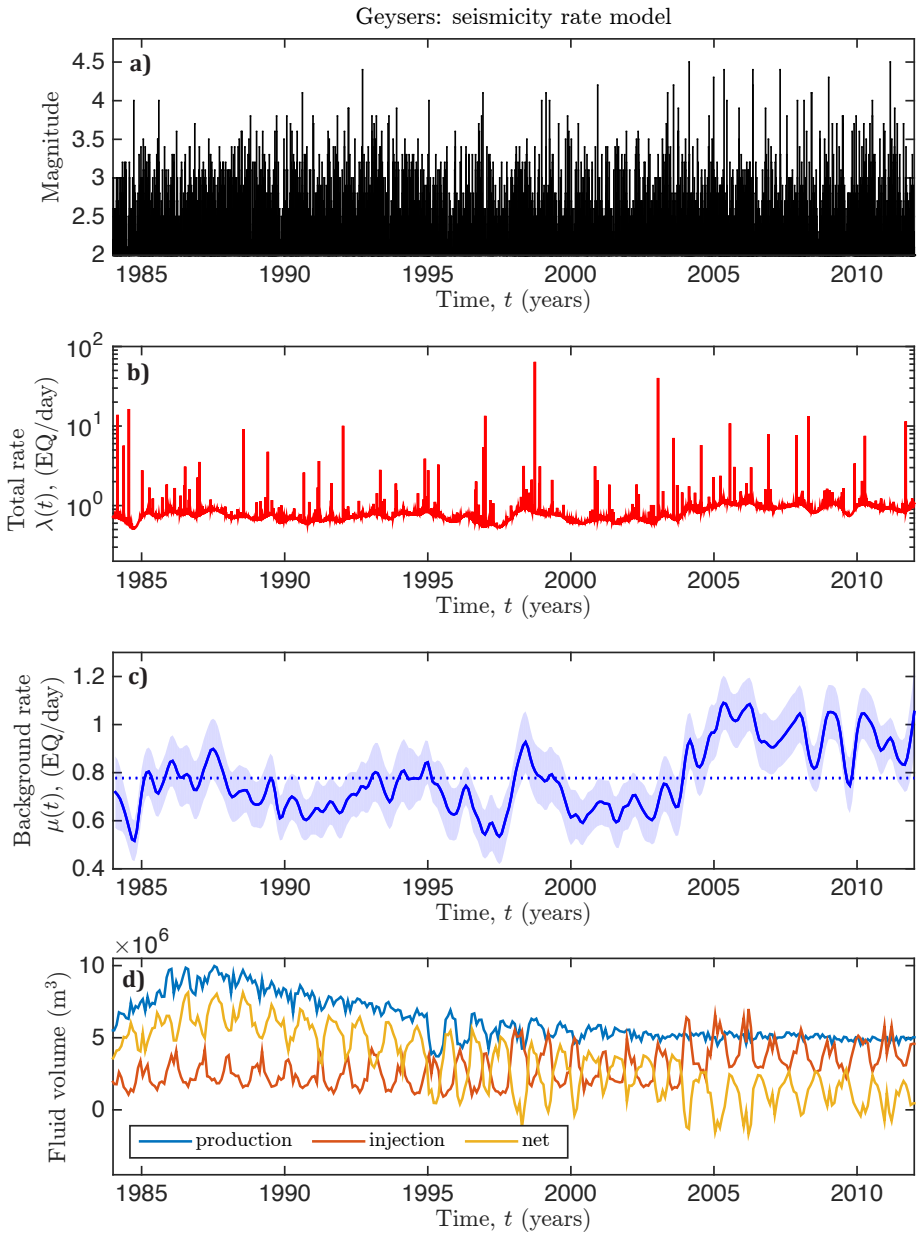


**Figure 3.2**: Seismicity rate model for The Geysers. a) Magnitude versus time for earthquakes within the field. b) Total seismicity rate, $\lambda(t) = \mu(t) + \nu(t)$. c) Background seismicity rate $\mu(t)$, with two-sigma errorbars (shaded) and mean rate (dotted line) plotted for reference. d) Monthly rate of fluid production (blue), injection (red) and net production (production minus injection, yellow) at The Geysers (data available at http://www.conservation.ca.gov/dog/geothermal).

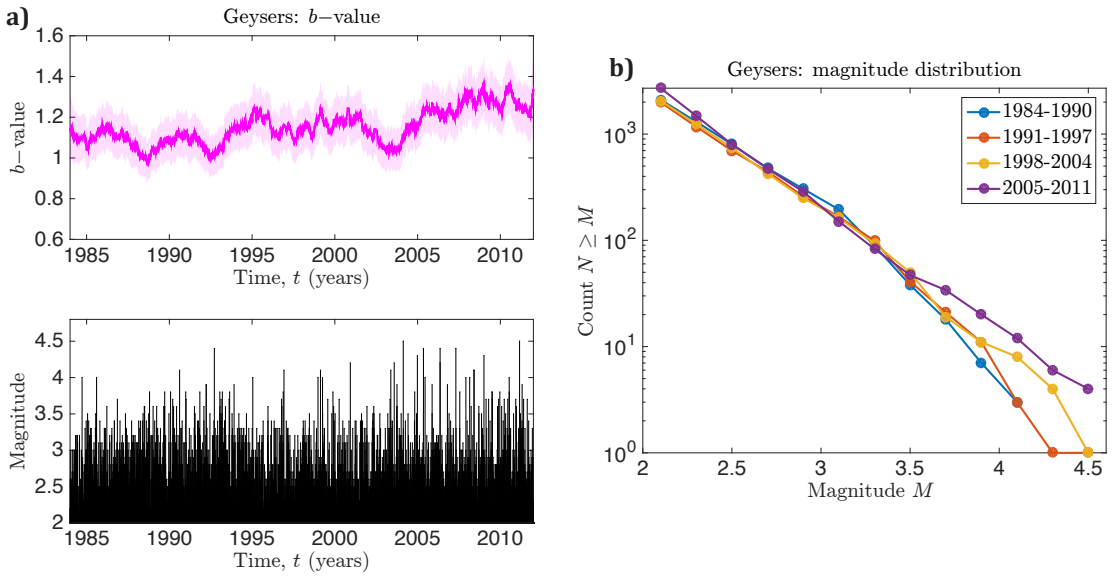


**Figure 3.3**: Magnitude distribution of earthquakes at The Geysers. a) Top: Maximum-likelihood estimates for the Gutenberg-Richter $b$-value (Equation 3.6) plotted as a function of time using a 401-event moving window. Two-sigma error-bars are shaded. Bottom: Magnitude versus time for earthquakes within the field, plotted for reference. b) Magnitude distributions for earthquakes at The Geysers during four, non-overlapping time intervals (1984-1990, 1991-1997, 1998-2004, 2005-2011).

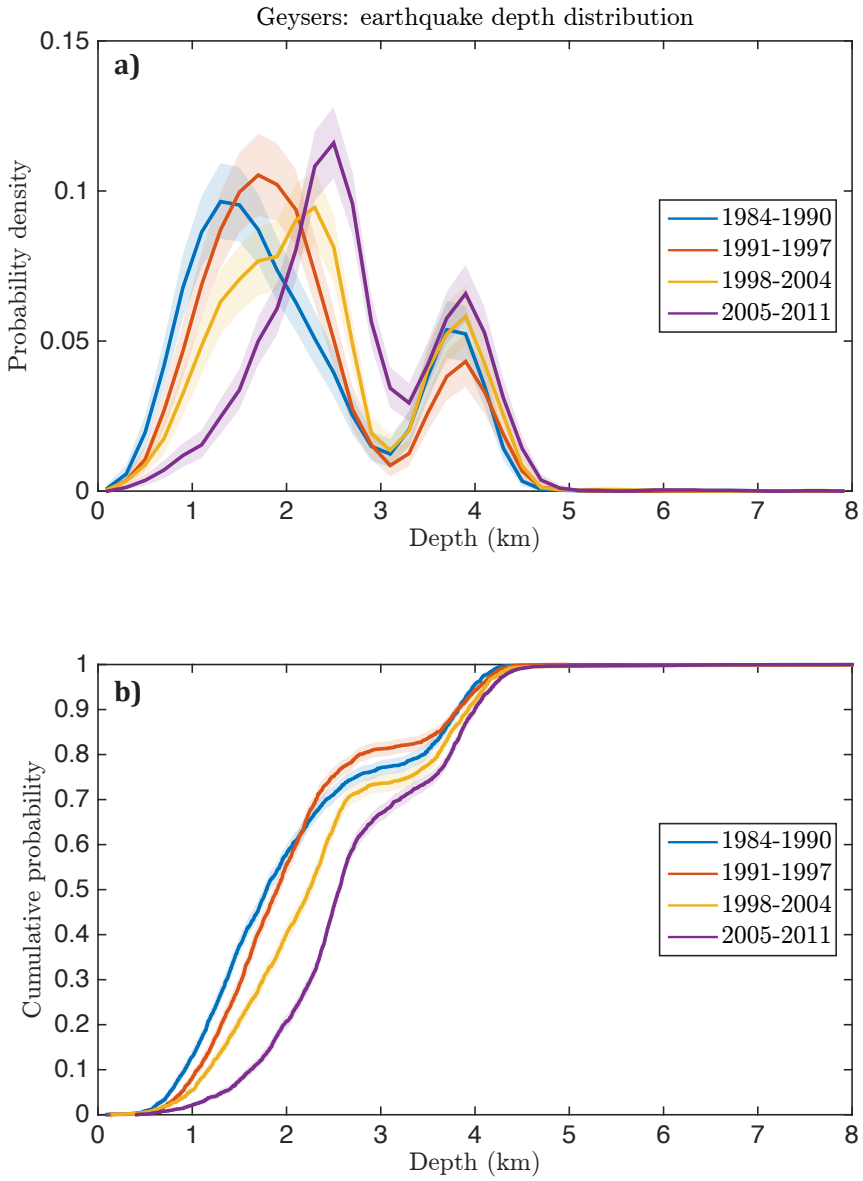


**Figure 3.4**: Depth distribution of earthquakes at The Geysers. Probability density functions (a) and cumulative density functions (b) for the depth distributions of earthquakes at The Geysers during the same four time intervals shown in Figure 3.3. 95-percent confidence intervals for the depth distributions are shaded.

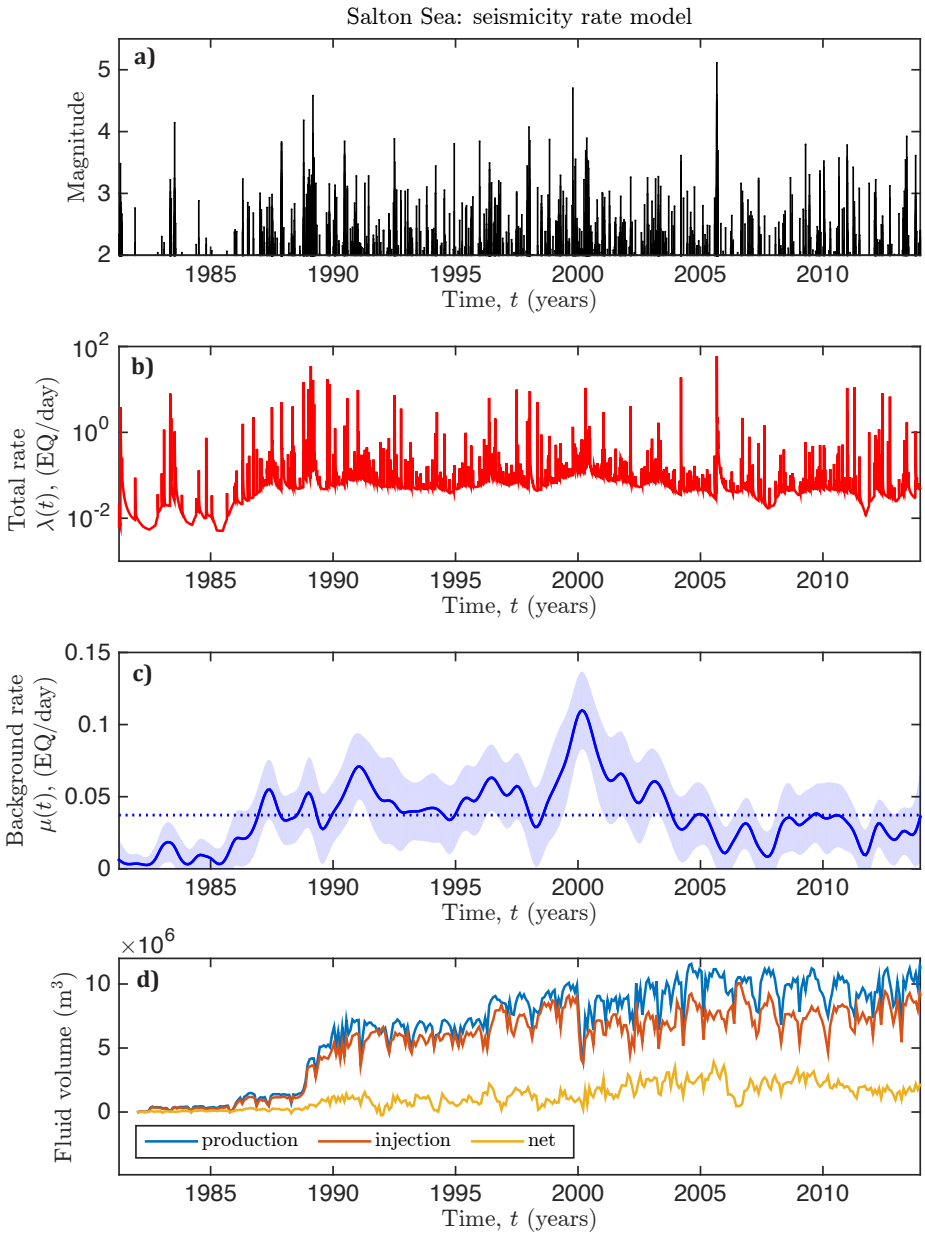


**Figure 3.5**: Seismicity rate model for Salton Sea. a) Magnitude versus time for earthquakes within the field. b) Total seismicity rate, $\lambda(t) = \mu(t) + \nu(t)$. c) Background seismicity rate $\mu(t)$, with two-sigma errorbars (shaded) and mean rate (dotted line) plotted for reference. d) Monthly rate of fluid production (blue), injection (red) and net production (production minus injection, yellow) at Salton Sea (data available at http://www.conservation.ca.gov/dog/geothermal).

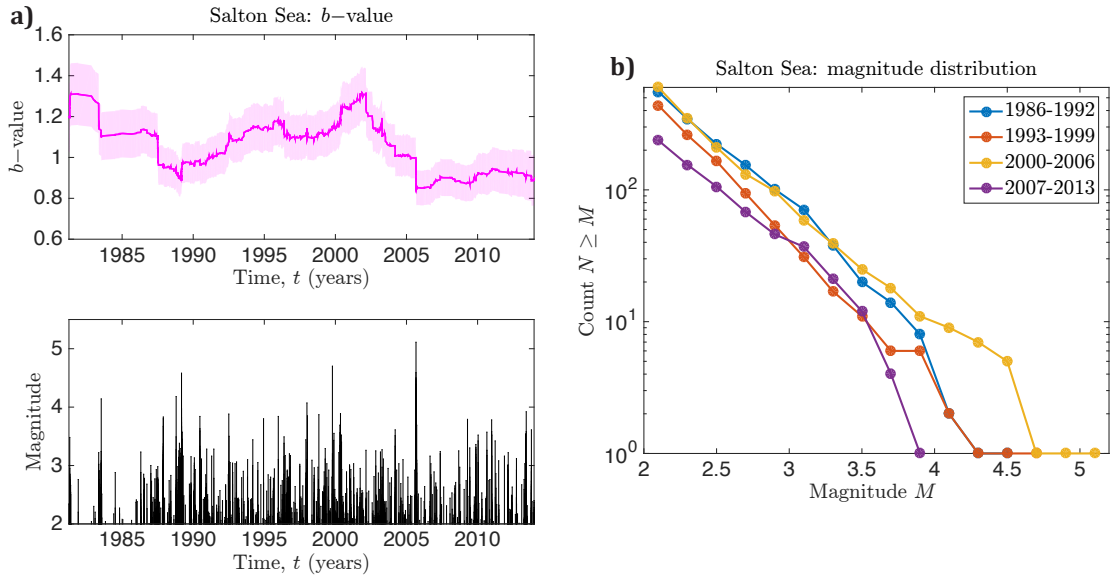


**Figure 3.6**: Magnitude distribution of earthquakes at Salton Sea. a) Top: Maximum-likelihood estimates for the Gutenberg-Richter $b$-value (Equation 3.6) plotted as a function of time using a 401-event moving window. Two-sigma errorbars are shaded. Bottom: Magnitude versus time for earthquakes within the field, plotted for reference. b) Magnitude distributions for earthquakes at Salton Sea during four, non-overlapping time intervals (1986-1992, 1993-1999, 2000-2006, 2007-2013).

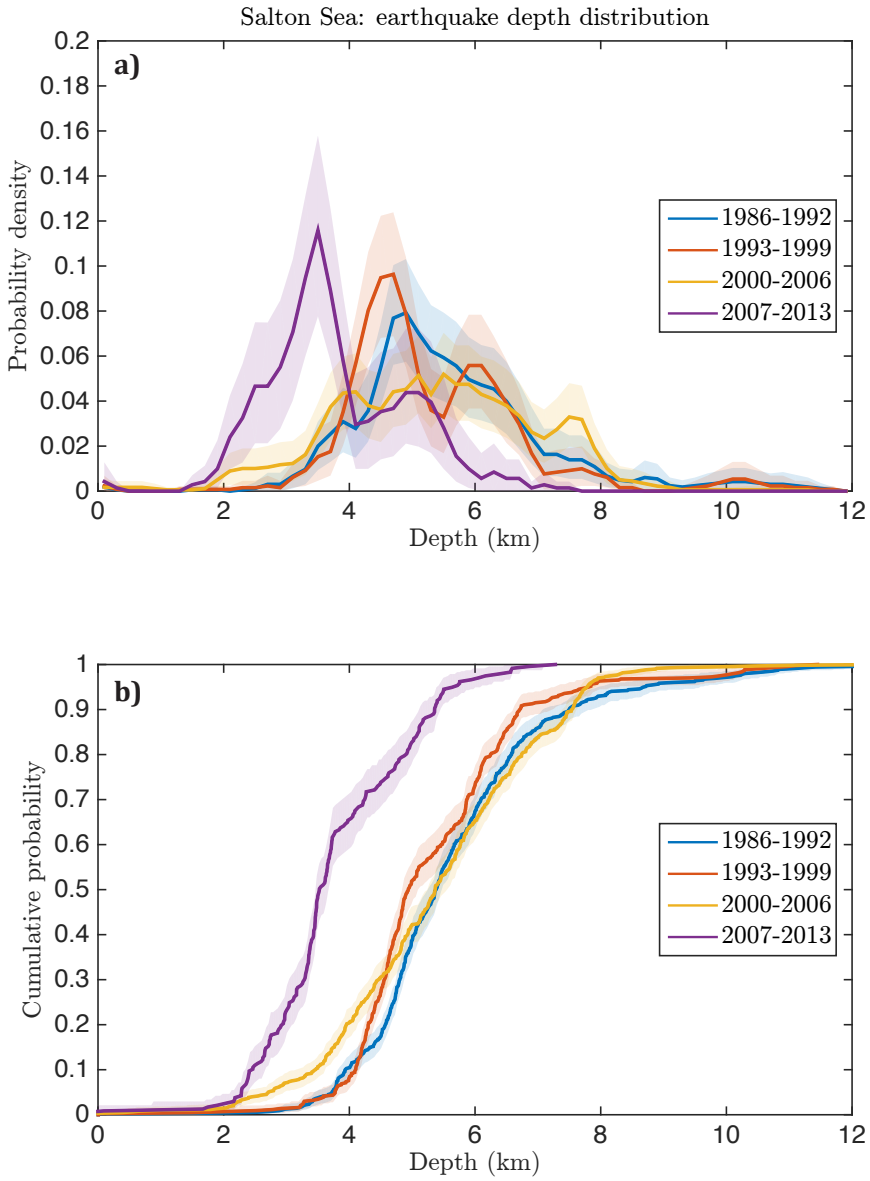


**Figure 3.7**: Depth distribution of earthquakes at Salton Sea. Probability density functions (a) and cumulative density functions (b) for the depth distributions of earthquakes at Salton Sea during the same four time intervals shown in Figure 3.6. 95-percent confidence intervals for the depth distributions are shaded.

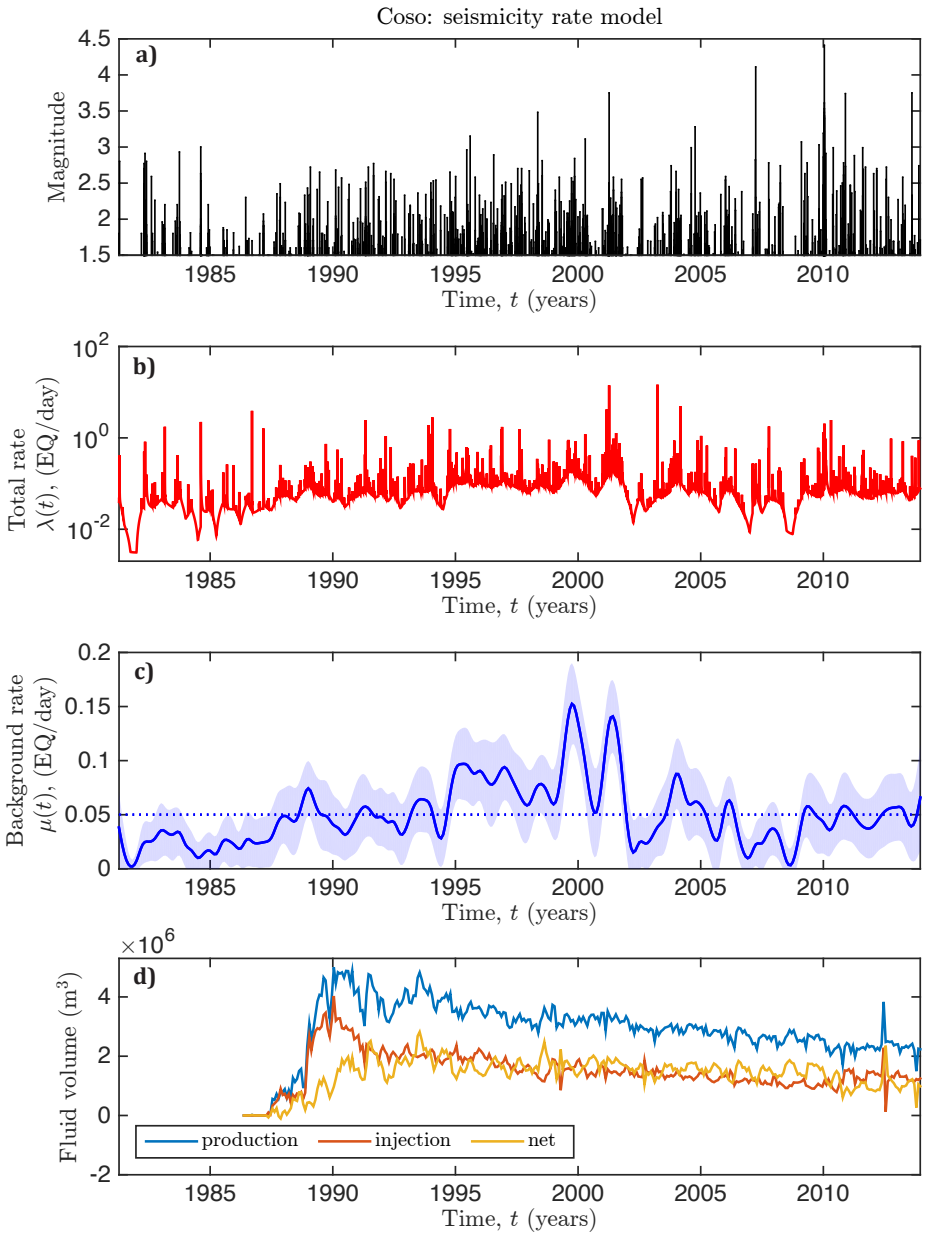


**Figure 3.8**: Seismicity rate model for Coso. a) Magnitude versus time for earthquakes within the field. b) Total seismicity rate, $\lambda(t) = \mu(t) + \nu(t)$. c) Background seismicity rate $\mu(t)$, with two-sigma errorbars (shaded) and mean rate (dotted line) plotted for reference. d) Monthly rate of fluid production (blue), injection (red) and net production (production minus injection, yellow) at Coso (data available at http://www.conservation.ca.gov/dog/geothermal).

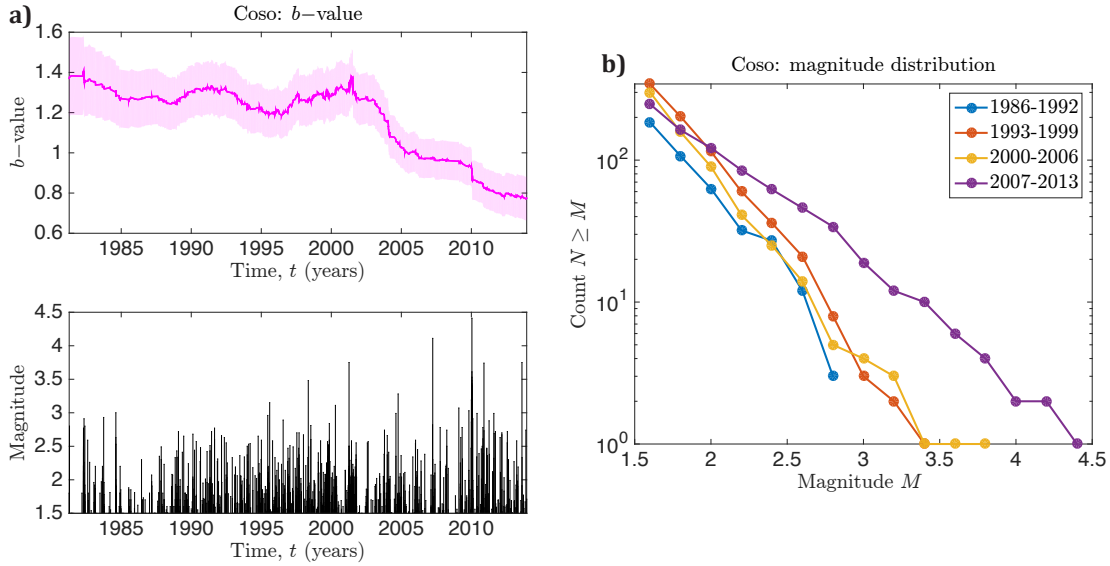


**Figure 3.9**: Magnitude distribution of earthquakes at Coso. a) Top: Maximum-likelihood estimates for the Gutenberg-Richter $b$-value (Equation 3.6) plotted as a function of time using a 401-event moving window. Two-sigma errorbars are shaded. Bottom: Magnitude versus time for earthquakes within the field, plotted for reference. b) Magnitude distributions for earthquakes at Coso during four, non-overlapping time intervals (1986-1992, 1993-1999, 2000-2006, 2007-2013).

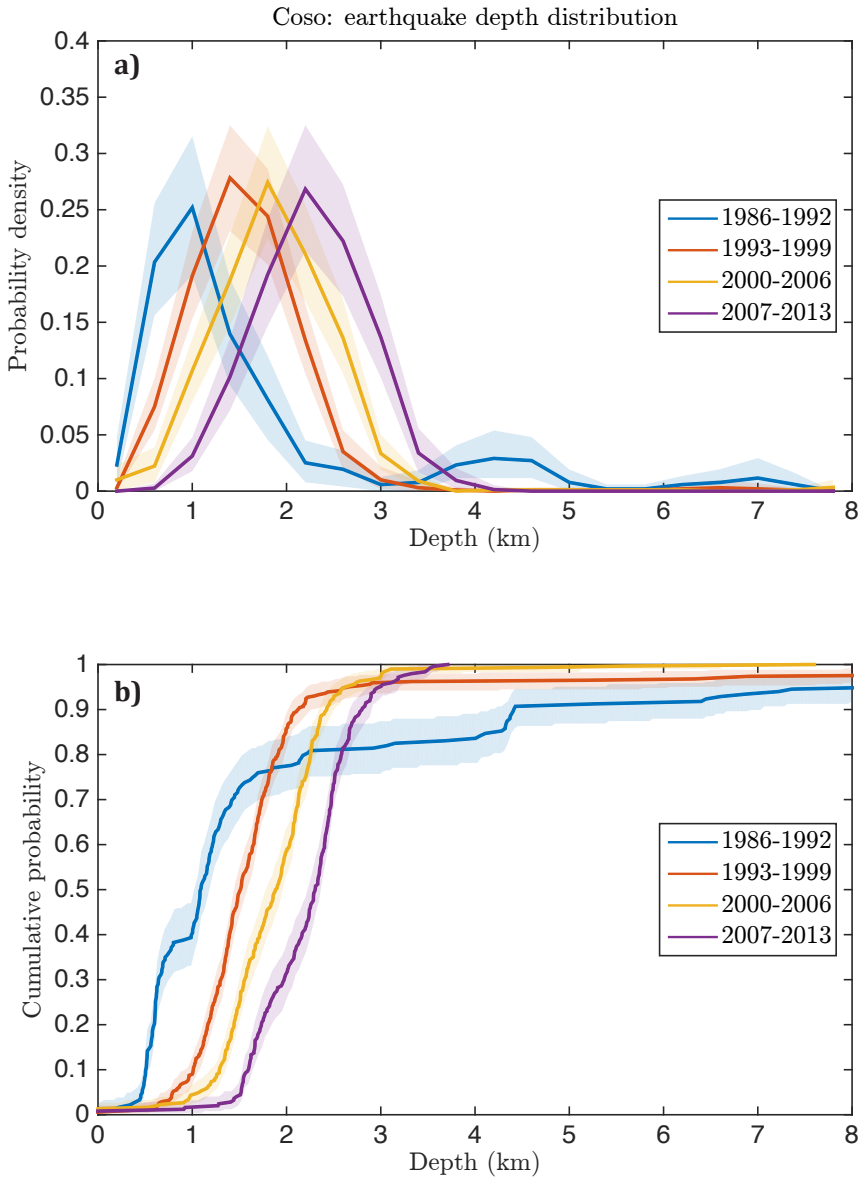


**Figure 3.10**: Depth distribution of earthquakes at Coso. Probability density functions (a) and cumulative density functions (b) for the depth distributions of earthquakes at Coso during the same four time intervals shown in Figure 3.9. 95-percent confidence intervals for the depth distributions are shaded.

**Table 3.1**: Earthquake Catalog Data for The Geysers, Salton Sea, and Coso.

| **Field** | **Catalog duration** (years) | **Lon. bounds** (degrees E) | **Lat. bounds** (degrees N) | $M_c$ | $N_Q$ |
|---|---|---|---|---|---|
| TG | 1984–2011 | [-122.85, -122.68] | [38.73, 38.87] | 2.0 | 8857 |
| SSGF | 1981–2013 | [-115.70, -115.50] | [32.12, 33.25] | 2.0 | 1992 |
| CGF | 1981–2013 | [-117.85, -117.75] | [35.99, 36.05] | 1.5 | 1201 |

**Table 3.2**: ETAS Parameter Estimates (and 95% Confidence Intervals).

| **Field** | $K$ (95% C.I.) | $\alpha$ (95% C.I.) | $p$ (95% C.I.) | $c$ (days) (95% C.I.) | $R_b$ (95% C.I.) |
|---|---|---|---|---|---|
| TG | 0.0005 (0.0002,0.0011) | 0.77 (0.71,0.84) | 1.56 (1.41,1.77) | 0.0010 (0.0007,0.0015) | 0.10 (0.09,0.11) |
| SSGF | 0.026 (0.021,0.028) | 0.41 (0.37,0.47) | 1.21 (1.19,1.27) | 0.0018 (0.0016,0.0026) | 0.74 (0.69,0.79) |
| CGF | 0.016 (0.011,0.017) | 0.50 (0.41,0.56) | 1.12 (1.10,1.23) | 0.0008 (0.0005,0.0017) | 0.48 (0.42,0.53) |

# References

Adams, J. W. (2011), The Geysers and Salton Sea Geothermal Fields, *Tech. rep.*, California State Lands Commision - Mineral Resources Management Division, Long Beach, CA.

Akaike, H. (1974), A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, *19*(6), 716–723, doi:10.1109/TAC.1974.1100705.

Akaike, H. (1980), Likelihood and the Bayes Procedure, *Trabajos de Estadistica Y de Investigacion Operativa*, *31*(1), 143–166, doi:10.1007/BF02888350.

Ake, J., K. Mahrer, D. OConnell, and L. Block (2005), Deep-Injection and Closely Monitored Induced Seismicity at Paradox Valley, Colorado, *Bulletin of the Seismological Society of America*, *95*(2), 664–683, doi:10.1785/0120040072.

Altmann, J. B., O. Heidbach, and R. Gritto (2013), Relative importance of processes leading to stress changes in The Geysers geothermal area, in *Thirty-Eighth Workshop on Geothermal Reservoir Engineering*, Stanford University, Stanford, CA.

Aster, R. C., and C. H. Thurber (2013), *Parameter Estimation and Inverse Problems*, 2nd ed ed., Academic Press, Waltham, MA.

Bachmann, C. E., S. Wiemer, J. Woessner, and S. Hainzl (2011), Statistical analysis of the induced Basel 2006 earthquake sequence: introducing a probability-based monitoring approach for Enhanced Geothermal Systems: Probability-based monitoring approach for EGS, *Geophysical Journal International*, *186*(2), 793–807, doi:10.1111/j.1365-246X.2011.05068.x.

Bachmann, C. E., S. Wiemer, B. P. Goertz-Allmann, and J. Woessner (2012), Influence of pore-pressure on the event-size distribution of induced earthquakes, *Geophysical Research Letters*, *39*(9), doi:10.1029/2012GL051480.

Bender, B. (1983), Maximum likelihood estimation of b values for magnitude grouped data, *Bulletin of the Seismological Society of America*, *73*(3), 831–851.

Bhattacharyya, J., and J. M. Lees (2002), Seismicity and seismic stress in the Coso Range, Coso geothermal field, and Indian Wells Valley region, southeast-central California, in *Memoir 195: Geologic Evolution of the Mojave Desert and Southwestern Basin and Range*, vol. 195, pp. 243–257, Geological Society of America.

Biot, M. A. (1941), General Theory of Three-Dimensional Consolidation, *Journal of Applied Physics*, *12*(2), 155, doi:10.1063/1.1712886.

Boyd, T. L., A. Sifford, and J. W. Lund (2015), The United States of America Country Update 2015, in *Proceedings World Geothermal Congress 2015*, Melbourne, Australia.

Boyle, K., and M. Zoback (2014), The Stress State of the Northwest Geysers, California Geothermal Field, and Implications for Fault-Controlled Fluid Flow, *Bulletin of the Seismological Society of America*, *104*(5), 2303–2312, doi:10.1785/0120130284.

Brocher, T. M., A. S. Baltay, J. L. Hardebeck, F. F. Pollitz, J. R. Murray, A. L. Llenos, D. P. Schwartz, J. L. Blair, D. J. Ponti, J. J. Lienkaemper, V. E. Langenheim, T. E. Dawson, K. W. Hudnut, D. R. Shelly, D. S. Dreger, J. Boatwright, B. T. Aagaard, D. J. Wald, R. M. Allen, W. D. Barnhart, K. L. Knudsen, B. A. Brooks, and K. M. Scharer (2015), The Mw 6.0 24 August 2014 South Napa Earthquake, *Seismological Research Letters*, *86*(2A), 309–326, doi:10.1785/0220150004.

Brodsky, E. E., and L. J. Lajoie (2013), Anthropogenic Seismicity Rates and Operational Parameters at the Salton Sea Geothermal Field, *Science*, *341*(6145), 543–546, doi:10.1126/science.1239213.

Calpine (2014), The Geysers Geothermal Field 2014 Statistics, *Tech. rep.*, Calpine Corporation, Middletown, CA.

Chen, X., and P. M. Shearer (2011), Comprehensive analysis of earthquake source spectra and swarms in the Salton Trough, California, *Journal of Geophysical Research*, *116*(B9), doi:10.1029/2011JB008263.

Combs, J. (1980), Heat flow in the Coso Geothermal Area, Inyo County, California, *Journal of Geophysical Research*, *85*(B5), 2411, doi:10.1029/JB085iB05p02411.

Convertito, V., N. Maercklin, N. Sharma, and A. Zollo (2012), From Induced Seismicity to Direct Time-Dependent Seismic Hazard, *Bulletin of the Seismological Society of America*, *102*(6), 2563–2573, doi:10.1785/0120120036.

Crowell, B. W., Y. Bock, D. T. Sandwell, and Y. Fialko (2013), Geodetic investigation into the deformation of the Salton Trough, *Journal of Geophysical Research: Solid Earth*, *118*(9), 5030–5039, doi:10.1002/jgrb.50347.

Daley, D. J., and D. Vere-Jones (2003), *An Introduction to the Theory of Point Processes*, 2nd ed., Springer, New York.

Deichmann, N., and D. Giardini (2009), Earthquakes Induced by the Stimulation of an Enhanced Geothermal System below Basel (Switzerland), *Seismological Research Letters*, *80*(5), 784–798, doi:10.1785/gssrl.80.5.784.

Duffield, W. A., C. R. Bacon, and G. B. Dalrymple (1980), Late Cenozoic volcanism, geochronology, and structure of the Coso Range, Inyo County, California, *Journal of Geophysical Research*, *85*(B5), 2381, doi:10.1029/JB085iB05p02381.

Efron, B., and R. J. Tibshirani (1994), *An Introduction to the Bootstrap*, CRC press, Boca Raton, FL.

Ellsworth, W. L. (2013), Injection-Induced Earthquakes, *Science*, *341*(6142), 1225,942–1225,942, doi:10.1126/science.1225942.

Feng, Q., and J. M. Lees (1998), Microseismicity, stress, and fracture in the Coso geothermal field, California, *Tectonophysics*, *289*(1-3), 221–238, doi:10.1016/S0040-1951(97)00317-X.

Fialko, Y., and M. Simons (2000), Deformation and seismicity in the Coso geothermal area, Inyo County, California: Observations and modeling using satellite radar interferometry, *Journal of Geophysical Research*, *105*(B9), 21,781, doi:10.1029/2000JB900169.

Fletcher, J. M., O. J. Teran, T. K. Rockwell, M. E. Oskin, K. W. Hudnut, K. J. Mueller, R. M. Spelz, S. O. Akciz, E. Masana, G. Faneros, E. J. Fielding, S. Leprince, A. E. Morelan, J. Stock, D. K. Lynch, A. J. Elliott, P. Gold, J. Liu-Zeng, A. González-Ortega, A. Hinojosa-Corona, and J. González-García (2014), Assembly of a large earthquake from a complex fault system: Surface rupture kinematics of the 4 April 2010 El MayorCucapah (Mexico) Mw 7.2 earthquake, *Geosphere*, *10*(4), 797–827, doi:10.1130/GES00933.1.

Frohlich, C., and M. Brunt (2013), Two-year survey of earthquakes and injection/production wells in the Eagle Ford Shale, Texas, prior to the October 2011 earthquake, *Earth and Planetary Science Letters*, *379*, 56–63, doi:10.1016/j.epsl.2013.07.025.

Frohlich, C., C. Hayward, B. Stump, and E. Potter (2011), The Dallas-Fort Worth Earthquake Sequence: October 2008 through May 2009, *Bulletin of the Seismological Society of America*, *101*(1), 327–340, doi:10.1785/0120100131.

Geothermal Energy Association (2013), 2013 Annual US Geothermal Power Production and Development Report, *Tech. rep.*, Washington D.C., USA.

Goebel, T. (2015), A comparison of seismicity rates and fluid-injection operations in Oklahoma and California: Implications for crustal stresses, *The Leading Edge*, *34*(6), 640–648, doi:10.1190/tle34060640.1.

Gomberg, J., and S. Davis (1996), Stress/strain changes and triggered seismicity at The Geysers, California, *Journal of Geophysical Research*, *101*(B1), 733, doi:10.1029/95JB03250.

Gonzalez-Ortega, A., Y. Fialko, D. Sandwell, F. Alejandro Nava-Pichardo, J. Fletcher, J. Gonzalez-Garcia, B. Lipovsky, M. Floyd, and G. Funning (2014), El Mayor-Cucapah (Mw 7.2) earthquake: Early near-field postseismic deformation from InSAR and GPS observations, *Journal of Geophysical Research: Solid Earth*, *119*(2), 1482–1497, doi:10.1002/2013JB010193.

Good, I. J., and R. A. Gaskins (1971), Nonparametric Roughness Penalties for Probability Densities, *Biometrika*, *58*(2), 255, doi:10.2307/2334515.

Gutenberg, B., and C. F. Richter (1944), Frequency of earthquakes in California, *Bulletin of the Seismological Society of America*, *34*(4), 185–188.

Hainzl, S. (2002), Indications for a successively triggered rupture growth underlying the 2000 earthquake swarm in Vogtland/NW Bohemia, *Journal of Geophysical Research*, *107*(B12), doi:10.1029/2002JB001865.

Hainzl, S., O. Zakharova, and D. Marsan (2013), Impact of Aseismic Transients on the Estimation of Aftershock Productivity Parameters, *Bulletin of the Seismological Society of America*, *103*(3), 1723–1732, doi:10.1785/0120120247.

Hardebeck, J. L. (2014), Appendix S: Constraining Epidemic Type Aftershock Sequence (ETAS) Parameters from the Uniform California Earthquake Rupture Forecast, Version 3: Catalog and Validating the ETAS Model for Magnitude 6.5 or Greater Earthquakes.

Hauksson, E., and J. Unruh (2007), Regional tectonics of the Coso geothermal area along the intracontinental plate boundary in central eastern California: Three-dimensional Vp and Vp/Vs models, spatial-temporal seismicity patterns, and seismogenic deformation, *Journal of Geophysical Research*, *112*(B6), doi:10.1029/2006JB004721.

Hauksson, E., J. Stock, K. Hutton, W. Yang, J. A. Vidal-Villegas, and H. Kanamori (2011), The 2010 M w 7.2 El Mayor-Cucapah Earthquake Sequence, Baja California, Mexico and Southernmost California, USA: Active Seismotectonics along the Mexican Pacific Margin, *Pure and Applied Geophysics*, *168*(8-9), 1255–1277, doi:10.1007/s00024-010-0209-7.

Hauksson, E., W. Yang, and P. M. Shearer (2012), Waveform Relocated Earthquake Catalog for Southern California (1981 to June 2011), *Bulletin of the Seismological Society of America*, *102*(5), 2239–2244, doi:10.1785/0120120010.

Hauksson, E., J. Stock, R. Bilham, M. Boese, X. Chen, E. J. Fielding, J. Galetzka, K. W. Hudnut, K. Hutton, L. M. Jones, H. Kanamori, P. M. Shearer, J. Steidl, J. Treiman, S. Wei, and W. Yang (2013), Report on the August 2012 Brawley Earthquake Swarm in Imperial Valley, Southern California, *Seismological Research Letters*, *84*(2), 177–189, doi:10.1785/0220120169.

Healy, J. H., W. W. Rubey, D. T. Griggs, and C. B. Raleigh (1968), The Denver Earthquakes, *Science*, *161*(3848), 1301–1310.

Horton, S. (2012), Disposal of Hydrofracking Waste Fluid by Injection into Subsurface Aquifers Triggers Earthquake Swarm in Central Arkansas with Potential for Damaging Earthquake, *Seismological Research Letters*, *83*(2), 250–260, doi: 10.1785/gssrl.83.2.250.

Hubbert, M. K., and W. W. Rubey (1959), Role of fluid pressure in mechanics of overthrust faulting, *Geological Society of America Bulletin*, *70*(2), 115, doi: 10.1130/0016-7606(1959)70[115:ROFPIM]2.0.CO;2.

Hudnut, K. W., L. Seeber, and J. Pacheco (1989), Cross-fault triggering in the November 1987 Superstition Hills Earthquake Sequence, southern California, *Geophysical Research Letters*, *16*(2), 199–202, doi:10.1029/GL016i002p00199.

Jeanne, P., J. Rutqvist, D. Vasco, J. Garcia, P. F. Dobson, M. Walters, C. Hartline, and A. Borgia (2014), A 3d hydrogeological and geomechanical model of an Enhanced Geothermal System at The Geysers, California, *Geothermics*, *51*, 240–252, doi:10.1016/j.geothermics.2014.01.013.

Kagan, Y. Y. (2002), Seismic moment distribution revisited: I. Statistical results: Seismic moment distribution: I, *Geophysical Journal International*, *148*(3), 520–541, doi:10.1046/j.1365-246x.2002.01594.x.

Kaven, J. O., S. H. Hickman, and N. C. Davatzes (2013), Micro-seismicity within the Coso Geothermal Field, California, from 1996-2012, in *Thirty-Eighth Workshop on Geothermal Reservoir Engineering*, Stanford Geothermal Workshop, Stanford University, Stanford, CA.

Keranen, K. M., H. M. Savage, G. A. Abers, and E. S. Cochran (2013), Potentially induced earthquakes in Oklahoma, USA: Links between wastewater injection and the 2011 Mw 5.7 earthquake sequence, *Geology*, *41*(6), 699–702, doi:10.1130/G34045.1.

Keranen, K. M., M. Weingarten, G. A. Abers, B. A. Bekins, and S. Ge (2014), Sharp increase in central Oklahoma seismicity since 2008 induced by massive wastewater injection, *Science*, *345*(6195), 448–451, doi:10.1126/science.1255802.

Kim, W.-Y. (2013), Induced seismicity associated with fluid injection into a deep well in Youngstown, Ohio, *Journal of Geophysical Research: Solid Earth*, *118*(7), 3506–3518, doi:10.1002/jgrb.50247.

Kumazawa, T., and Y. Ogata (2013), Quantitative description of induced seismic activity before and after the 2011 Tohoku-Oki earthquake by nonstationary ETAS models, *Journal of Geophysical Research: Solid Earth*, *118*(12), 6165–6182, doi:10.1002/2013JB010259.

Kumazawa, T., Y. Ogata, and S. Toda (2010), Precursory seismic anomalies and transient crustal deformation prior to the 2008 Mw 6.9 Iwate-Miyagi Nairiku, Japan, earthquake, *Journal of Geophysical Research*, *115*(B10), doi:10.1029/2010JB007567.

Kwiatek, G., P. Martinez-Garzon, G. Dresen, M. Bohnhoff, H. Sone, and C. Hartline (2015), Effects of long-term fluid injection on induced seismicity parameters and maximum magnitude in northwestern part of The Geysers geothermal field, *Journal of Geophysical Research: Solid Earth*, doi:10.1002/2015JB012362.

Llenos, A. L. (2014), The (Un) Productivity of the 2014 M6.0 South Napa Aftershock Sequence, *AGU Fall Meeting Abstracts*, *1*, 4928.

Llenos, A. L., J. J. McGuire, and Y. Ogata (2009), Modeling seismic swarms triggered by aseismic transients, *Earth and Planetary Science Letters*, *281*(1-2), 59–69, doi:10.1016/j.epsl.2009.02.011.

Lohman, R. B., and J. J. McGuire (2007), Earthquake swarms driven by aseismic creep in the Salton Trough, California, *Journal of Geophysical Research*, *112*(B4), doi:10.1029/2006JB004596.

Majer, E. L. (1978), Seismological investigation in geothermal regions, PhD Thesis, UC Berkeley, Berkeley, CA.

Majer, E. L., and J. E. Peterson (2007), The impact of injection on seismicity at The Geysers, California Geothermal Field, *International Journal of Rock Mechanics and Mining Sciences*, *44*(8), 1079–1090, doi:10.1016/j.ijrmms.2007.07.023.

Majer, E. L., R. Baria, M. Stark, S. Oates, J. Bommer, B. Smith, and H. Asanuma (2007), Induced seismicity associated with Enhanced Geothermal Systems, *Geothermics*, *36*(3), 185–222, doi:10.1016/j.geothermics.2007.03.003.

Manley, C. R. (2000), Rhyolite Thermobarometry and the Shallowing of the Magma Reservoir, Coso Volcanic Field, California, *Journal of Petrology*, *41*(1), 149–174, doi:10.1093/petrology/41.1.149.

Marsan, D., and O. Lengliné (2010), A new estimation of the decay of aftershock density with distance to the mainshock, *Journal of Geophysical Research*, *115*(B9), doi:10.1029/2009JB007119.

Marsan, D., E. Prono, and A. Helmstetter (2013), Monitoring Aseismic Forcing in Fault Zones Using Earthquake Time Series, *Bulletin of the Seismological Society of America*, *103*(1), 169–179, doi:10.1785/0120110304.

Martinez-Garzon, P., M. Bohnhoff, G. Kwiatek, and G. Dresen (2013), Stress tensor changes related to fluid injection at The Geysers geothermal field, California, *Geophysical Research Letters*, *40*(11), 2596–2601, doi:10.1002/grl.50438.

Martinez-Garzon, P., G. Kwiatek, H. Sone, M. Bohnhoff, G. Dresen, and C. Hartline (2014), Spatiotemporal changes, faulting regimes, and source parameters of induced seismicity: A case study from The Geysers geothermal field: Characterization of induced seismicity, *Journal of Geophysical Research: Solid Earth*, *119*(11), 8378–8396, doi:10.1002/2014JB011385.

Marzocchi, W., and L. Sandri (2009), A review and new insights on the estimation of the b-value and its uncertainty, *Annals of Geophysics*, (6), doi:10.4401/ag-3472.

Matek, B., and K. Gawell (2014), Report on the state of geothermal energy in California, *Tech. rep.*, Geothermal Energy Association, Washington D.C., USA.

McClusky, S. C., S. C. Bjornstad, B. H. Hager, R. W. King, B. J. Meade, M. M. Miller, F. C. Monastero, and B. J. Souter (2001), Present day kinematics of the Eastern California Shear Zone from a geodetically constrained block model, *Geophysical Research Letters*, *28*(17), 3369–3372, doi:10.1029/2001GL013091.

Mendenhall, W., and T. Sincich (2007), *Statistics for Engineering and the Sciences*, Pearson Prentice-Hall, Upper Saddle River, N.J.

Monastero, F., A. Katzenstein, J. Miller, J. Unruh, M. Adams, and K. Richards-Dinger (2005), The Coso geothermal field: A nascent metamorphic core complex, *Geological Society of America Bulletin*, *117*(11), 1534, doi:10.1130/B25600.1.

Monastero, F. C. (2002), An overview of industry  military cooperation in the development of power operations at the Coso geothermal field in southern California, *Geothermal Resources Council Bulletin*, *31*(5), 188–195.

Mossop, A., and P. Segall (1997), Subsidence at The Geysers Geothermal Field, N. California from a comparison of GPS and leveling surveys, *Geophysical Research Letters*, *24*(14), 1839–1842, doi:10.1029/97GL51792.

Mossop, A., and P. Segall (1999), Volume strain within The Geysers geothermal field, *Journal of Geophysical Research*, *104*(B12), 29,113, doi:10.1029/1999JB900284.

Muffler, L. J., and D. E. White (1969), Active Metamorphism of Upper Cenozoic Sediments in the Salton Sea Geothermal Field and the Salton Trough, Southeastern California, *Geological Society of America Bulletin*, *80*(2), 157, doi:10.1130/0016-7606(1969)80[157:AMOUCS]2.0.CO;2.

National Resource Council (2013), *Induced Seismicity Potential in Energy Technologies*, The National Academies Press, Washington, DC.

Nicholson, C., and R. L. Wesson (1992), Triggered earthquakes and deep well activities, *Pure and Applied Geophysics*, *139*(3-4), 561–578, doi:10.1007/BF00879951.

Ogata, Y. (1983), Estimation of the parameters in the modified Omori formula for aftershock frequencies by the maximum likelihood procedure., *Journal of Physics of the Earth*, *31*(2), 115–124, doi:10.4294/jpe1952.31.115.

Ogata, Y. (1988), Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes, *Journal of the American Statistical Association*, *83*(401), 9–27, doi:10.1080/01621459.1988.10478560.

Ogata, Y. (1992), Detection of precursory relative quiescence before great earthquakes through a statistical model, *Journal of Geophysical Research*, *97*(B13), 19,845, doi:10.1029/92JB00708.

Ogata, Y. (2004), Space-time model for regional seismicity and detection of crustal stress changes, *Journal of Geophysical Research*, *109*(B3), doi:10.1029/2003JB002621.

Ogilvie, J. (1984), A monte-carlo approach to error propagation, *Computers & Chemistry*, *8*(3), 205–207, doi:10.1016/0097-8485(84)80007-8.

Omori, F. (1894), Investigation of aftershocks, *Rep. Earthquake Inv. Comm*, *2*, 103–139.

Oppenheimer, D. H. (1986), Extensional tectonics at The Geysers Geothermal Area, California, *Journal of Geophysical Research*, *91*(B11), 11,463, doi:10.1029/JB091iB11p11463.

Parker, R. L. (1994), *Geophysical Inverse Theory*, Princeton University Press, Princeton, N.J.

Press, W. H. (Ed.) (2007), *Numerical Recipes: The Art of Scientific Computing*, 3rd ed ed., Cambridge University Press, Cambridge, UK ; New York.

Raleigh, C. B., J. H. Healy, and J. D. Bredehoeft (1976), An Experiment in Earthquake Control at Rangely, Colorado, *Science*, *191*(4233), 1230–1237, doi:10.1126/science.191.4233.1230.

Rubinstein, J. L., W. L. Ellsworth, A. McGarr, and H. M. Benz (2014), The 2001-Present Induced Earthquake Sequence in the Raton Basin of Northern New Mexico and Southern Colorado, *Bulletin of the Seismological Society of America*, *104*(5), 2162–2181, doi:10.1785/0120140009.

Schoenball, M., N. C. Davatzes, and J. M. G. Glen (2015), Differentiating induced and natural seismicity using space-time-magnitude statistics applied to the Coso Geothermal field, *Geophysical Research Letters*, *42*(15), 2015GL064,772, doi: 10.1002/2015GL064772.

Segall, P. (1989), Earthquakes triggered by fluid extraction, *Geology*, *17*(10), 942, doi:10.1130/0091-7613(1989)017⟨0942:ETBFE⟩2.3.CO;2.

Segall, P., and S. D. Fitzgerald (1998), A note on induced stress changes in hydrocarbon and geothermal reservoirs, *Tectonophysics*, *289*(1-3), 117–128, doi: 10.1016/S0040-1951(97)00311-9.

Segall, P., and S. Lu (2015), Injection-induced seismicity: Poroelastic and earthquake nucleation effects, *Journal of Geophysical Research: Solid Earth*, *120*(7), 2015JB012,060, doi:10.1002/2015JB012060.

Segall, P., J.-R. Grasso, and A. Mossop (1994), Poroelastic stressing and induced seismicity near the Lacq gas field, southwestern France, *Journal of Geophysical Research*, *99*(B8), 15,423, doi:10.1029/94JB00989.

Shapiro, S. A., O. S. Krüger, C. Dinske, and C. Langenbruch (2011), Magnitudes of induced earthquakes and geometric scales of fluid-stimulated rock volumes, *Geophysics*, *76*(6), WC55–WC63, doi:10.1190/geo2010-0349.1.

Snyder, D. L., and M. I. Miller (1991), *Random Point Processes in Time and Space*, Springer New York, New York, NY.

Sornette, D. (2005), Apparent clustering and apparent background earthquakes biased by undetected seismicity, *Journal of Geophysical Research*, *110*(B9), doi: 10.1029/2005JB003621.

Stark, M. (2003), Seismic evidence for a long-lived Enhanced Geothermal System (EGS) in The Northern Geysers reservoir, *Geothermal Resources Council Transactions*, *27*, 727–731.

Stark, M. A. (1992), Microearthquakes: a tool to track injected water in The Geysers reservoir, in *Monograph on The Geysers Geothermal Field, Special Report*, vol. 17, pp. 111–117, Geothermal Resources Council, Davis, CA.

Suckale, J. (2009), Induced Seismicity in Hydrocarbon Fields, in *Advances in Geophysics*, vol. 51, pp. 55–106, Elsevier.

Tarantola, A. (2005), *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM.

Toda, S., R. S. Stein, and T. Sagiya (2002), Evidence from the AD 2000 Izu islands earthquake swarm that stressing rate governs seismicity, *Nature*, *419*(6902), 58–61, doi:10.1038/nature00997.

Trugman, D. T., A. A. Borsa, and D. T. Sandwell (2014), Did stresses from the Cerro Prieto Geothermal Field influence the El Mayor-Cucapah rupture sequence?, *Geophysical Research Letters*, *41*(24), 8767–8774, doi:10.1002/2014GL061959.

Utsu, T. (1961), A statistical study on the occurrence of aftershocks., *Geophysical Magazine*, *30*(4), 521–605.

Utsu, T. (1999), Representation and Analysis of the Earthquake Size Distribution: A Historical Review and Some New Approaches, *Pure and Applied Geophysics*, *155*(2-4), 509–535, doi:10.1007/s000240050276.

Utsu, T., Y. Ogata, R. S, and Matsu'ura (1995), The Centenary of the Omori Formula for a Decay Law of Aftershock Activity, *Journal of Physics of the Earth*, *43*(1), 1–33, doi:10.4294/jpe1952.43.1.

Vasco, D. W., J. Rutqvist, A. Ferretti, A. Rucci, F. Bellotti, P. Dobson, C. Oldenburg, J. Garcia, M. Walters, and C. Hartline (2013), Monitoring deformation at the Geysers Geothermal Field, California using C-band and X-band interferometric synthetic aperture radar, *Geophysical Research Letters*, *40*(11), 2567–2572, doi:10.1002/grl.50314.

Vidale, J. E., and P. M. Shearer (2006), A survey of 71 earthquake bursts across southern California: Exploring the role of pore fluid pressure fluctuations and aseismic slip as drivers, *Journal of Geophysical Research*, *111*(B5), doi:10.1029/2005JB004034.

Waldhauser, F., and D. P. Schaff (2008), Large-scale relocation of two decades of Northern California seismicity using cross-correlation and double-difference methods, *Journal of Geophysical Research*, *113*(B8), doi:10.1029/2007JB005479.

Walter, A. W., and C. S. Weaver (1980), Seismicity of the Coso Range, California, *Journal of Geophysical Research*, *85*(B5), 2441, doi:10.1029/JB085iB05p02441.

Wei, S., E. Fielding, S. Leprince, A. Sladen, J.-P. Avouac, D. Helmberger, E. Hauksson, R. Chu, M. Simons, K. Hudnut, T. Herring, and R. Briggs (2011), Superficial simplicity of the 2010 El MayorCucapah earthquake of Baja California in Mexico, *Nature Geoscience*, *4*(9), 615–618, doi:10.1038/ngeo1213.

Weingarten, M., S. Ge, J. W. Godt, B. A. Bekins, and J. L. Rubinstein (2015), High-rate injection is associated with the increase in U.S. mid-continent seismicity, *Science*, *348*(6241), 1336–1340, doi:10.1126/science.aab1345.

Wessel, P., W. H. F. Smith, R. Scharroo, J. Luis, and F. Wobbe (2013), Generic Mapping Tools: Improved Version Released, *Eos, Transactions American Geophysical Union*, *94*(45), 409–410, doi:10.1002/2013EO450001.

Woessner, J., and S. Wiemer (2005), Assessing the Quality of Earthquake Catalogues: Estimating the Magnitude of Completeness and Its Uncertainty, *Bulletin of the Seismological Society of America*, *95*(2), 684–698, doi:10.1785/0120040007.

Younker, L. W., P. W. Kasameyer, and J. D. Tewhey (1982), Geological, geophysical, and thermal characteristics of the Salton Sea Geothermal Field, California, *Journal of Volcanology and Geothermal Research*, *12*(3-4), 221–258, doi:10.1016/0377-0273(82)90028-2.

# Chapter 4

# GrowClust: A hierarchical clustering algorithm for relative earthquake relocation, with application to Nevada earthquake sequences

## Abstract

Accurate earthquake locations are essential for providing reliable hazard assessments, understanding the physical mechanisms driving extended earthquake

sequences, and interpreting fault structure. Techniques based on waveform cross-correlation can significantly improve the precision of the relative locations of event pairs observed at a set of common stations. Here we describe GrowClust, an open-source, relative relocation algorithm that can provide robust relocation results for earthquake sequences over a wide range of spatial and temporal scales. The method uses input differential travel times, cross-correlation values, and reference starting locations, and applies a hybrid, hierarchical clustering algorithm to simultaneously group and relocate events within similar event clusters. The method is computationally efficient and numerically stable in its capacity to process large data sets, and naturally applies greater weight to more similar event pairs. Additionally, it outputs location error estimates that can be used to help interpret the reliability and resolution of relocation results. As an example, we apply the GrowClust method to the recent Spanish Springs and Sheldon, Nevada, earthquake swarms. These sequences highlight the future potential for applying the GrowClust relocation method on a much larger scale within the region, where existing relocation results are sparse but vital for understanding the seismotectonics and seismic hazard of Nevada and eastern California.

## 4.1 Introduction

The location of earthquake hypocenters using observations of seismic phase arrival times is a classic inverse problem in geophysics, with a rich history of

conceptual and methodological advancements dating back more than a century. During this time, catalogs of earthquake locations have become one of the most important and widely-used forms of seismological data. They provide fundamental constraints on a number of important seismotectonic problems, from the resolution and imaging of fault zone structure, to the understanding of the physical mechanisms underlying earthquake triggering and interaction, to the improvement of seismic hazard assessments. In turn, the degree to which these and other outstanding problems can be resolved depends intrinsically on the quality and reliability of earthquake location methodology.

The travel time of a seismic phase observed at a given station depends nonlinearly on both the earthquake hypocentral coordinates and the subsurface velocity structure (*Geiger*, 1910; *Buland*, 1976; *Thurber*, 1985). Because of this, improvements in absolute earthquake location accuracy will always be limited by imperfect knowledge of 3D variations in Earth structure (*Thurber*, 1983, 1992; *Thurber and Eberhart-Phillips*, 1999). However, in recent years, numerous methods have been developed that yield significant improvements to relative earthquake location accuracy through the joint relocation of pairs or clusters of linked events (*Douglas*, 1967; *Frohlich*, 1979; *Jordan and Sverdrup*, 1981; *Got et al.*, 1994; *Shearer*, 1998; *Waldhauser and Ellsworth*, 2000; *Richards-Dinger and Shearer*, 2000; *Lin et al.*, 2007). While initial catalog locations are routinely determined from the noisy, often emergent phase picks and associated travel times of each event in isolation, relative relocation techniques are based primarily on differen-

tial travel times of pairs of events observed at common stations. This formulation helps mitigate common-mode errors introduced by the biasing effects of unmodeled velocity structure. Furthermore, waveform cross-correlation techniques (e.g., *Poupinet et al.*, 1984; *Ito*, 1985; *Fremont and Malone*, 1987; *Nadeau et al.*, 1995; *Phillips et al.*, 1997; *Rowe et al.*, 2002) can be used to extract differential travel times with considerably greater precision than is possible from the absolute travel times alone. Application of these relative relocation techniques can therefore result in a dramatic sharpening in seismicity and resolution of fine-scale fault structure (e.g., *Rubin et al.*, 1999; *Waldhauser et al.*, 1999; *Astiz and Shearer*, 2000; *Shearer*, 2002) compared to standard catalog locations.

Implementation of these methods is straightforward in the case of single, compact clusters of events, for which the common mode errors can be approximated as identical, but becomes more complicated for distributed seismicity where the errors associated with unmodeled 3D velocity structure vary considerably with event location. The most widely used approach to relocate distributed seismicity is the double-difference (DD) method (*Waldhauser and Ellsworth*, 2000), which sets up a linear system of equations for the differences in event locations as a function of the input 1D velocity model and differences in differential times (as well as absolute times, if desired) and then applies matrix inversion. The DD technique has been applied to problems across a wide range of scales, including relocating over 500,000 events in the Northern California earthquake catalog (*Waldhauser and Schaff*, 2008).

Here we describe an alternative to DD for relocating seismicity using differential times, which we term the GrowClust algorithm, and are making available to the community as an integrated software package (see Data and Resources). Given the success of the DD algorithm and its growing user base, why did we develop a different method? One motivation was to permit applying a more robust misfit criteria than the L2-norm (least squares) than is used in standard matrix inversion. GrowClust uses the L1-norm, which is less sensitive to outliers (bad data points) in the input times and has been shown in some cases (e.g., *Shearer*, 1997) to yield improved results compared to L2-norm solutions. Another concern is that for many problems the matrix representing the complete solution in DD is so ill-conditioned that inversions become unstable. For example, this can occur when event clusters are linked by differential times from a single cross-correlation pair, or a chain of event pairs, and hence the stability issue becomes more pronounced for large-scale relocation problems containing multiple discrete clusters of seismicity. The DD algorithm does provide various algorithm control parameters to help stabilize the inversion by, for example, downweighting data from more distant or more poorly correlated event pairs. However, the optimal parameter choices and iteration-by-iteration sequential weights required to obtain the best locations are not always obvious to the user, and even these optimal parameter choices may not guarantee stability for certain problems. Our approach is to apply cluster analysis ideas to decide which groups of events are linked and relocated, and which cluster pairs retain independent locations. While our method has its own set of parameter

choices, we have designed them to be as few and as straightforward as possible. Finally, we wanted to implement a method that runs fast enough that bootstrap resampling to estimate location errors is practical on small computer systems.

The GrowClust algorithm unifies a series of programs that were developed to relocate seismicity in southern California using waveform cross-correlation (*Shearer*, 2005; *Lin et al.*, 2007; *Hauksson et al.*, 2012). GrowClust combines cluster analysis and relocation of events within each cluster, which were previously performed as two separate steps. This provides a significant boost in computational efficiency and convergence stability for large data sets, which typically contain multiple large clusters. GrowClust uses differential travel times and cross-correlation results in a hybrid, hierarchical clustering algorithm that both groups events into clusters based on waveform similarity, and relocates each event with respect to the linked events within its unique cluster. The method is computationally efficient and multiscale in its applicability to both small and large sequences. Unlike many other related relocation techniques, the GrowClust algorithm intrinsically gives greater weight to more highly correlated event pairs and does not require explicit matrix inversion.

A preliminary version of the GrowClust algorithm was first applied to perform high-precision relocation of a large data set of more than 100,000 events recorded on Hawaii Island from 1992–2009 (*Matoza et al.*, 2013). We present here a more complete description of the improved, newly open-source algorithm, which now implements a nonparametric statistical resampling technique to esti-

mate errors in the relocated event positions, which prove useful in assessing and interpreting the relocation results. As a demonstration of the new method, we use GrowClust to relocate two prominent, recent earthquake sequences in western Nevada: (1) the 2012–2015 Spanish Springs swarm (mainshock $M_W$4.2), which caused non-structural damage and significant shaking in Reno, and (2) the ongoing, 2014–present Sheldon swarm, an extensive sequence (dozens of M4 and hundreds of M3 events have occurred to date) occurring in the northwest corner of the state, where station coverage is sparse and catalog locations are particularly scattered. Large-scale, systematic event relocations in northern (*Waldhauser and Schaff*, 2008) and southern (*Hauksson et al.*, 2012) California have significantly advanced our understanding of seismicity, seismic hazard, and fault structure within that state. Our initial results suggest that, with methodological improvements and the recent modernization of the Nevada Seismic Network (*Kent et al.*, 2015), the same potential for scientific advancement exists within Nevada and eastern California.

## 4.2 Methods and Algorithm Description

In this section, we detail the methodology and computational details underlying the current implementation of GrowClust. As we intend this code to be open-source (see Data and Resources), future releases of the code may include minor modifications to the algorithm presented here. We begin this section with a

brief outline of the data pre-processing steps required to obtain the necessary inputs for GrowClust. Next, we describe the basic algorithm used by GrowClust, which simultaneously groups individual earthquakes into similar event clusters based on the input cross-correlation data, and relocates each earthquake with respect to its cluster neighbors. Finally, we describe the resampling approach that GrowClust (optionally) performs to assess uncertainties of the relocated event positions. The basic GrowClust workflow is summarized in Figure 4.1.

### 4.2.1 Data pre-processing

The fundamental input data for the GrowClust algorithm are differential travel times and cross-correlation values, obtained through waveform cross-correlation of sets of earthquake event pairs observed at common stations. Both are needed, as the differential times are used to relocate the events, while the cross-correlation values are used to weight the data by quality and group events by waveform similarity. However, the algorithm is sensitive only to the relative values of cross-correlation coefficients, as the absolute values depend on both the data quality and auxiliary factors such as the length of the cross-correlation window and filter type, if any is used.

GrowClust is flexible in its accommodation of cross-correlation results obtained using any method, whether time-domain or frequency-domain, and can use any combination of $P$ and/or $S$ phase cross-correlation results. For the results presented in this paper, we use a time-domain cross-correlation approach

that uses spline interpolation to achieve millisecond precision in differential times. Frequency-domain techniques that achieve similar precision (e.g. *Poupinet et al.*, 1984) should work equally well. For our approach, we find that applying a zero-phase bandpass filter from 1-10 Hz to the raw waveform data is useful in isolating the relevant phase arrivals and mitigating noise that may cause spurious cross-correlation results.

In addition to cross-correlation times, GrowClust requires a velocity model to compute predicted differential travel times. Though not requisite for the GrowClust method, we use travel-time tables derived from the velocity model to select cross-correlation windows around predicted $P$ and $S$ phase arrival times, a practice which can vastly expand the cross-correlation data set to include waveform pairs devoid of operator $P$ and $S$ phase picks. Finally, GrowClust requires input event and station lists that uniquely identify each earthquake (e.g., with an EVID) and each station (e.g., with a station name). The various inputs to GrowClust, along with algorithm control parameters described in the following section, are combined in an input file read by GrowClust upon initial computation (Figure 4.1).

### 4.2.2 The GrowClust algorithm

In this section, we provide a conceptual outline of the hybrid clustering algorithm used by GrowClust to perform relative event relocation. Upon program initiation, GrowClust reads the cross-correlation data, parameters from the algorithm control file, and input station and event lists into memory (Figure 4.1). The

latter contains the initial (catalog or other reference) hypocentral positions from which events are relocated. The program also constructs travel time tables for both $P$ and $S$ body wave phases based on the input velocity model, which are later used to compute the predicted differential times necessary for event relocation.

Following this initial input and data organization stage, GrowClust begins its hybrid clustering and relocation algorithm, which works as follows (see Figure 4.2 for a simplified conceptual example):

1. Assign each of the $N$ events to a distinct starting cluster number. Though many of these clusters will later be merged as part of the relocation process, the $N$ initial clusters each begin as a single event.

2. For each event pair $(i, j)$, compute a similarity coefficient $Z_{ij}$, that serves as a metric to measure the data quality and waveform similarity of each distinct event pair. Here, we take $Z_{ij}$ for an event pair to be a sum over the cross-correlation values $r_{ij;k}$ observed at the $k$ common stations within a maximum station distance $\Delta_{max}$ and that exceed a minimum value, $r_{min}$:

$$Z_{ij} = \sum_k r_{ij;k} \quad \forall \quad r_{ij;k} \geq r_{min} \text{ and } \Delta_k \leq \Delta_{max}. \tag{4.1}$$

   The control parameters $\Delta_{max}$ and $r_{min}$ are chosen by the user in the input control file based on the data set at hand, and cross-correlation data for $P$ and $S$ phases are treated equally unless otherwise specified.

3. Sort the event pairs $(i, j)$ by similarity coefficient $Z_{ij}$, and process each event pair in turn, starting with the most similar. As the algorithm proceeds,

there are three situations it may encounter when considering a new event pair (Figure 4.2):

(A) Both event $i$ and $j$ are members of single-event clusters. In this situation, the events are merged into a new cluster (now with two events), and both are relocated with respect to the new cluster centroid. The relative relocation algorithm uses a grid-search approach to find the relative locations that minimize the L1-norm of the residual between observed and predicted differential travel times (*Shearer*, 2005):

$$\text{minimize} \quad ||dR||_{L1} = \sum_k |dtt_{ij,k} - d\hat{t}t_{ij,k}|, \tag{4.2}$$

Here $dtt_{ij,k}$ and $d\hat{t}t_{ij,k}$ denote the observed and predicted differential travel times for the event pair $(i, j)$ at station $k$, where the predicted times depend on the relative event locations and the velocity model. Note that the grid search is performed over the relative event locations, and the L1 norm of the residuals is computed over the set of common stations with waveform observations of the event pair.

(B) Either of events $i$ or $j$ are members of multi-event clusters. In this situation, the algorithm performs a series of tests to decide whether to merge the clusters and relocate all events within both clusters with respect to one another. First, the algorithm searches for all other event pairs that link the two clusters. If the ratio of observed links to total possible links fails to exceed a specified threshold (e.g., 0.01), the cluster merger is

rejected, and the event pair is skipped. Otherwise, the algorithm performs a test relocation of the two clusters with respect to one another using only the ten strongest links (i.e., linking event pairs with the ten highest $Z_{ij}$ values) for robustness. The cluster relocation uses a multi-event generalization of the L1-norm approach applied to single-event clusters in situation (A). The relative positions of the events within each of the two distinct clusters are held fixed, while the two cluster centroids are adjusted about the centroid of the combined cluster. Once this trial relocation is performed, the algorithm checks to see whether the differential travel-time residuals of the newly merged cluster or the centroid shifts of the initial clusters exceed tolerance values specified by the user. These tolerance criteria help to ensure the stability of the algorithm by preventing mergers that are not required by the data or that involve unreasonably large location adjustments that would violate the assumption of small location shifts implicit in relative relocation methods (*Geiger*, 1912). If neither criterion is violated, the two clusters are merged into a single, combined cluster that contains all of events at their new, relocated positions.

(C) Both event $i$ and $j$ belong to the same cluster. In this case, the algorithm simply skips the relocation of event pair, as both events have already been relocated further up the algorithm.

4. The algorithm continues processing clusters in this way until no more of them can be merged and relocated, given the algorithm control parameters. At this stage, GrowClust then computes the final run statistics for user assessment, and saves these along with the relocation and clustering results for later output (Figure 4.1). It is important to note that GrowClust is a purely relative relocation algorithm: events within each cluster are relocated with respect to one another, with the cluster centroids held fixed at its initial reference position to ensure stability. As such, the events that comprise single-event clusters due to lack of waveform similarity are not relocated by the algorithm, and hence remain at their initial reference positions.

The overall strategy of GrowClust is to begin by relocating the highest quality event pairs and most similar event clusters, and then hold the relative locations of these events fixed when computing additional relative locations. An advantage of this approach is that it permits a grid search relocation method to be applied to cluster pairs in which there are only four free parameters at each relocation step: the $dx$, $dy$ and $dz$ offsets of one of the clusters from a reference position, plus any origin time shift ($dt$). The algorithm is computationally efficient because clusters are treated as coherent event sets, which vastly reduces the effective degrees of freedom during each relocation step, as linking event pairs are not relocated independently. A further advantage of this approach is that it can be easily modified to optimize against any desired misfit norm, allowing for improved

robustness through the use of the L1 norm rather than conventional least-squares (L2).

It is worth noting that the internal architecture of GrowClust is well-suited to some additional applications. Because the method naturally builds its results by performing new locations relative to a set of previously determined locations, it should be possible to adapt the method to compute near-real-time locations, in which new events are relocated with respect to previous events, rather than having to relocate the entire data set when new data become available. In addition, the grid search location algorithm reads from a set of travel-time tables, which are currently computed for a 1D velocity model, but, with only a slight increase in complexity, could be altered to work with travel-time tables based on a 3D velocity model. Finally, the procedure described below that is used to assess location uncertainties is fully parallelizable, and future implementations could be adapted to take advantage of this fact and further improve computational efficiency for large-scale problems.

### 4.2.3 Relative location uncertainties

The complexity of the GrowClust algorithm precludes simple, parametric techniques for assessing formal uncertainties in the relocated hypocentral positions. However, since a measure of location uncertainty is often fundamental in the interpretation of relocation results, we have developed and incorporated a non-parametric procedure within the GrowClust program that users can use to estimate

location uncertainties for all relocated events. The method implements a modified bootstrap approach based on statistical resampling theory (*Efron and Tibshirani*, 1994).

Upon initial input to GrowClust, cross-correlation data (differential times, cross-correlation values, and associated station metadata) are organized into arrays of length $N_{ph}$, where $N_{ph}$ denotes the total number of combined $P$ and $S$ phase observations. For each bootstrap iteration, the algorithm procedure generates resampling vectors (*Efron and Tibshirani*, 1994) to perform efficient, random resampling of these input arrays. The bootstrap-resampled cross-correlation data are then input to the main GrowClust algorithm, resulting in a perturbed set of event locations specific to that bootstrap iteration. Repeating the resampling procedure $B$ times, a bootstrap distribution of hypocentral positions (longitude, latitude, depth, and origin time) is constructed for each relocated event.

Though a complete analysis of the full sampling distribution using this method may require the number of boostrap resamples, $B$, to be of order 1000 or greater, estimates of standard errors in hypocentral parameters typically stabilize much faster ($B \sim 50$ to 100). The nonparametric error estimates output by GrowClust are obtained from the median absolute deviations (*Leys et al.*, 2013),

$$MAD(X) = median\left(|X_i - median(X)|\right) \tag{4.3}$$

of the bootstrap distribution of hypocentral coordinates. This provides a more robust characterization of location uncertainty than the raw bootstrap standard

errors, which may be biased if the underlying bootstrap distribution is skewed (*Hesterberg et al.*, 2008; *Mammen*, 2012).

## 4.3 Results: Relocation of Nevada Seismicity

Moderate-to-large, damaging earthquakes (M $\geq$ 5) occur more frequently in Nevada than in any state within the continental US except California. Three M $\geq$ 7 events shook the western half of state during the 20th century, and from the historical record M6 events have occurred on average every six years (*VanWormer and Ryall*, 1980; *Smith et al.*, 2008) in the Nevada-California border region. The strike-slip faults of the Walker Lane Belt in western Nevada and eastern California are capable of hosting M $\geq$ 7 earthquakes (e.g., the 1932 M7.1 Cedar Mountain earthquake), and the range-bounding normal faults of the eastern Sierra frontal fault system have a history of large, mid-M7, events (*Ramelli et al.*, 1999; *Dingler et al.*, 2009; *Wesnousky et al.*, 2012). Though these fault systems pose a significant hazard to the population centers of Reno and Carson City, as well as rural communities throughout the region, a comprehensive study of earthquake occurrence within the Walker Lane and the state as a whole has yet to be undertaken.

To demonstrate the use and efficacy of the GrowClust method, we apply it to two recent earthquake sequences in western Nevada: the 2012–2015 Spanish Springs and 2014–present Sheldon swarms (for a regional map and station locations, see Figure 4.S1 in the Supplementary Material). Both sequences are

prominent within Nevada's contemporary seismic record, and are worthy of scientific investigation in their own right. The Spanish Springs sequence is spatially compact, with more than 1600 events occurring over a length scale of several kilometers. The sequence was well-recorded by near-source stations operated by the Nevada Seismological Laboratory, and its $M_W$4.2 mainshock was widely felt in the Reno area. In contrast, the Sheldon sequence occurred within the remote northwestern corner of the state, where station coverage is sparse and hence the initial catalog locations are poorly constrained. The Sheldon sequence was larger in scale than Spanish Springs and is of particular scientific interest due to its persistent, swarm-like seismicity, replete with several discrete clusters containing multiple M $\geq$ 4 events. In the following, we focus primarily on the interpretation of the GrowClust-relocated event positions, while deferring a more complete assessment of the tectonics and source mechanisms of the two sequences to future studies.

### 4.3.1 Waveform data and cross-correlation

For our analysis of the Spanish Springs and Sheldon sequences, we use waveform data archived by the Nevada Seismic Laboratory (NSL). The NSL database is organized using an Antelope/DataScope software system (see Data and Resources), and earthquakes are routinely located by the NSL using the Antelope dblocsat2 and USGS HYPOINVERSE algorithms (*Klein*, 2002), assuming a reference velocity model listed in Table 4.S1 (available in the Supplementary Material). These event locations are then forwarded to the USGS/ANSS Comprehensive Earthquake

Catalog (ComCat; see Data and Resources Section), and we take these locations to be the initial event positions prior to performing the GrowClust relocations. We use the Antelope relational database to extract the waveforms from the events associated with each sequence.

Prior to performing waveform cross-correlation, we proceed as described in Methods, filtering all traces from 1-10 Hz using a bandpass filter with a gentle roll-off that retains some energy up to 15 Hz. For event pairs in the two sequences, we compute cross-correlation functions separately for $P$ and $S$ phases on all available channels of all common stations. We use cross-correlation windows of -1.0s to 1.5s and -1.0s to 2.5s for $P$ and $S$ phases, relative to the predicted arrival times of the respective phases. The use of predicted arrival times in lieu of operator picks can significantly expand the cross-correlation data set. Differential travel times and cross-correlation coefficients are derived from the raw cross-correlation functions using a spline interpolation technique that provides millisecond sampling precision. We use the vertical-component channel for $P$-phases (if available), and we select the horizontal channel with the highest cross-correlation coefficient for $S$-phases (and $P$-phases with no available vertical channel). We ensure the data quality of our cross-correlation database by further considering only those event pairs with an average cross-correlation coefficient of 0.45 across all phases, and a minimum of 8 phases with cross-correlation coefficients greater than 0.6.

### 4.3.2 The 2012–2015 Spanish Springs sequence

The 2012–2015 Spanish Springs sequence was quite active but spatially compact, with thousands of events occurring over a length scale of several kilometers. The sequence is named for its close spatial proximity to the North Reno suburb of Spanish Springs, and its largest event ($M_W$4.2) was widely felt through the city and resulted in non-structural damage to select local buildings. The sequence was well-recorded by the dense distribution of Nevada Seismic Network (NSN) stations in the Reno area, where the local magnitude of completion is $\sim$ 0.0 (*Kent et al.*, 2015), with many recorded events of even smaller magnitude.

We use the NSL/ANSS event locations as the initial catalog positions for our GrowClust relocation analysis of this sequence. We take advantage of the dense station coverage and only consider stations within 80 km, resulting in 51 unique short-period, broadband, and strong motion stations. We do not implement a magnitude cutoff for this sequence, and hence do not expect to provide relocated positions for all events (the majority of which are M0.5 or less and were initially located by the NSL using as few as three stations). Even so, we are able to extract adequate waveform similarity to relocate 793 of the 1616 recorded events in the sequence, including all events M1.0 and greater.

GrowClust relocation results for these 793 events are presented in Figure 4.3, with the initial catalog positions shown for reference. Despite the dense station coverage, the initial catalog positions are highly scattered both in map view

and in cross section (panels a/c), with little hint of local faulting structure. In contrast, the relocated seismicity is noticeably sharper, and the relocated event positions clearly outline several distinct faulting structures (panels b/d). The primary fault structure strikes to the northeast and is nearly vertical (cross-section $A \rightarrow A'$), while a secondary strand with a more northerly strike branches off of the primary structure's northeast end. Fault-perpendicular cross-sections ($B \rightarrow B'$) provide evidence that seismicity within a secondary cluster to the northwest of the primary structure occurs on a distinct, subparallel fault strand. The GrowClust relocations likewise aid in the interpretation of the spatiotemporal migration of seismicity during the Spanish Springs sequence (Figure 4.4). The sequence began as a short burst of M1.0 and lower seismicity in late 2012, with all events occuring on the southeasterly branch of the main fault structure. The $M_W$4.2 mainshock occurred on August 27, 2013, and was preceded by a vigorous foreshock sequence that migrated linearly from the southwest toward the mainshock hypocenter along the main fault strand. The subsequent aftershock sequence was extended in duration, with several distinct swarms occurring on spatially isolated sections of the different structures. Intriguingly, the relocation results reveal a hole (i.e., area free of seimisicity) on the mainshock fault plane, adjacent to its hypocenter. This hole may outline the region of largest slip or near complete stress release within the mainshock rupture zone, with aftershocks localized at the stress concentrations around its perimeter.

We further use GrowClust's nonparametric error estimation procedure to

examine the lateral and vertical location uncertainties of relocated events within the sequence. Overall, the structural features of the sequence are quite well-resolved, with median lateral and vertical location errors of 11 and 62 m, respectively (Figure 4.4c and Figure 4.S2 in the Supplementary Material). These low nominal uncertainties, a consequence of the dense seismicity, completeness of the catalog, and good azimuthal coverage of near-source stations, lend confidence to the interpretation of the salient structural and spatiotemporal features of the Spanish Springs sequence.

### 4.3.3 The 2014–present Sheldon sequence

The Sheldon earthquake sequence began in July of 2014, with swarms of seismicity occurring within and near the Sheldon Wildlife Refuge in the northwestern corner of Nevada. In contrast to the well-recorded Spanish Springs sequence, station coverage in this region is quite sparse, particularly before November 2014, when the near-field station COLR was installed at $\sim 15$ km distance. The sequence is of considerable scientific interest due to the large moment release (28 M $\geq$ 4.0 and 263 M $\geq$ 3.0 events to date, with the largest being $M_W$4.8) and its persistent, swarm-like seismicity that defies conventional earthquake triggering models. Though distant from major population centers, the larger events have been felt strongly by local residents and farming communities near Vya, Nevada.

As with Spanish Springs, we use the NSL/ANSS event locations as the initial catalog positions for our GrowClust relocation analysis. Because of the

inherent limitation in near-field station coverage, we include in our analysis all nine recording stations within 250 km. To ensure adequate signal-to-noise, we implement the previously described quality-control criteria for our cross-correlation data (see Methods), and restrict our analysis to events of local magnitude 1.8 or greater occurring on or after November 18, 2014, when the COLR station was installed. This cutoff is slightly more conservative than the local magnitude of completion, which is approximately M1.5. Using these criteria, our input event list consists of 1369 total events (through August 28, 2016), 1232 of which we are able to relocate.

The sparseness of station coverage in the Sheldon region is reflected in the highly scattered initial catalog positions (Figure 4.5). Despite these limitations, the GrowClust relocations provide a noticeable improvement, resolving distinct fault structures from the cloud of initial positions. Seismicity is concentrated within two subparallel clusters that strike north-northeast. The western cluster is more active, containing the dominant portion of events within the sequence, and is comprised of two distinct fault strands (striking north-northeast and north-northwest). Though the location uncertainties are larger for the Sheldon sequence (median horizontal and vertical errors of 117 and 91 m, Figure 4.6c), they are not so large as to inhibit the resolution of in-situ structure. This is particularly true for the western cluster, which with its denser seismicity and hence more available similar-event pairs is better resolved than the eastern cluster (Figure 4.S3 in the Supplementary Material). Analyses of fault-parallel and fault-perpendicular cross-sections reveal that

both clusters dip steeply to the east, with deeper seismicity in the western cluster (10–13 km) than the eastern cluster (9–11 km). This structural interpretation is consistent with moment tensor analyses of the larger events, which predominantly show normal faulting on steeply dipping normal faults striking north-northeast (*Ruhl et al.*, 2016).

The improved resolution provided by the GrowClust relocations allows us to examine the spatiotemporal evolution of the Sheldon sequence in more detail. The sequence is overall characterized by relatively continuous seismicity, but exhibits distinctive patterns of spatial migration. Early in 2015, seismicity migrates from north to south within the northwesterly branch of the major (western) cluster (Figure 4.6). After a brief quiescent period, the sequence is reactivated during the second half of the year, marked by the occurrence of three events in excess of magnitude 4.5, and dozens more in excess of magnitude 3.5. The seismicity is swarm-like, with no clear mainshock, and exhibits a tendency to migrate from the southwest to the northeast, again predominantly within the western cluster. It is not until late 2015 and early 2016 that the minor (eastern) cluster becomes active, though seismicity continues unabated within the northern portion of the major (western) cluster during this time.

## 4.4 Discussion

### 4.4.1 Comparison of GrowClust and HypoDD relocation results

It is instructive to compare the relocation results obtaining using GrowClust for the Spanish Springs and Sheldon sequences to those obtained using the HypoDD double-difference technique (*Waldhauser and Ellsworth* 2000, see Data and Resources). Doing so not only provides a benchmark of sorts for GrowClust (as HypoDD is well-established and justifiably popular), but also helps elucidate the more relevant differences in methodology. For the comparison, we used identical data inputs (initial hypocentral locations, velocity model, and cross-correlation data) as were used for the GrowClust results presented above. We performed sensitivity analysis on the iterative weighting scheme, algorithm damping, and other control parameters found in the HypoDD input file to obtain the best relocation results possible, given our limited experience in the technique's use. We note that both sequences are too large in scale and too poorly-conditioned to obtain singular value decomposition (SVD) solutions to the double difference equations, so the HypoDD results presented here are those obtained using the damped least-squares conjugate gradient (LSQR) solution (*Waldhauser and Ellsworth*, 2000), and we have taken care to ensure reasonable condition numbers for each algorithm iteration. As discussed in *Waldhauser and Ellsworth* (2000), the LSQR solution does

not produce accurate location uncertainty estimates, precluding a formal comparison with those obtained using GrowClust. Both the GrowClust and HypoDD methods relocate a comparable fraction of the total number of events in each sequence (793 versus 764 of the 1616 total Spanish Springs events, 1232 versus 1282 of the 1369 total Sheldon events, for GrowClust and HypoDD respectively). The runtime on a standard laptop computer is slightly faster for GrowClust (14.9 and 34.2 seconds for the two sequences) than for HypoDD (65.0 and 95.3 seconds).

Overall, the GrowClust and HypoDD results are in good visual agreement. This is particularly true for the well-recorded Spanish Springs sequence (Figure 4.7), with both methods providing good resolution of vertically-dipping fault structures and partitioning events into a single dominant cluster. The similarity in results for Spanish Springs is not surprising, given that the dense set of differential times ensures strong linkages between events (and hence, a well-conditioned system for the double-difference algorithm). For this reason, it is also unsurprising that there is greater disparity in the results for the Sheldon sequence (Figure 4.8), in which event pairs are not as robustly linked due to the sparse station coverage. In this case, the differences in the underlying computational framework — hierarchical clustering versus matrix inversion — become more apparent. Of particular interest is the location of the smaller, eastern cluster of seismicity, which differs by more than a kilometer from method to method. For the GrowClust relocations, the centroid of this smaller cluster is fixed to its initial reference position, while no such constraint is imposed in HypoDD, permitting a centroid shift of greater than

1 km. The centroid of larger, western cluster also undergoes a shift of 133 m to the northeast in the HypoDD relocations, while remaining fixed in the GrowClust relocations. Note that GrowClust's clustering algorithm considers the western and eastern clusters as independent units (i.e., it does not find a sufficient number of linking event pairs to join the clusters), while HypoDD relocates both clusters as one coherent unit, finding one or more linking chains of event pairs between the two clusters.

### 4.4.2 Implications of relocation results for the understanding of seismotectonics in the Nevada-California border region

Both conventional mainshock-aftershock sequences and extended swarm-like sequences are common within the state of Nevada (*Ichinose et al.*, 1998; *Smith et al.*, 2008; *Ruhl*, 2015), and relocation results such as those presented here help to study and differentiate the active physical mechanisms driving these sequences. The Spanish Springs and Sheldon sequences provide important test cases in this regard, and demonstrate the subtlety of the task at hand. Spanish Springs may well fall into the former classification, as its moment release is dominated by the $M_W$4.2 mainshock. However, complexity is pervasive even within such a compact sequence, and standard triggering models cannot adequately account for the extended duration or the spatial progression of the foreshock and aftershock se-

quences. The foreshock sequence is of particular interest, as its systematic linear migration toward the mainshock hypocenter lends insight into the nucleation process. Likewise, Sheldon's distinctive spatial migration pattern, combined with its unusually persistent and swarm-like seismicity, suggest that aseismic or fluid-driven processes may be the dominant physical mechanisms driving the sequence, rather than the coseismic stress changes from isolated, individual mainshocks (*Hainzl*, 2002; *Lohman and McGuire*, 2007; *Shearer*, 2012; *Shelly et al.*, 2016).

The GrowClust results presented here for Spanish Springs and Sheldon establish the potential for large-scale relocation efforts, analogous to those undertaken in recent years in California (*Waldhauser and Schaff*, 2008; *Hauksson et al.*, 2012), to elucidate subtle features of Nevada seismicity. Such efforts may prove to be particularly valuable in Nevada, where the western half of the state comprises a transition zone that accommodates approximately 20% of the Pacific-North America plate boundary deformation, and is characterized by a spatially complex distribution of intersecting normal, sinistral, and dextral faults (*Ichinose et al.*, 2003; *Faulds et al.*, 2005; *Wesnousky et al.*, 2012). The relocated seismicity provides a means to image these structures in high resolution, supplying important observational constraints on the seismotectonic evolution of the Walker Lane and Central Nevada Seismic Belts. The extensive historical record of large, damaging earthquakes in Nevada (*VanWormer and Ryall*, 1980) makes apparent the importance of an improved and more rigorously quantitative understanding of seismic hazard. We plan in future studies to expand our use of GrowClust beyond in-

dividual sequences, and apply it on a much wider scale within the state and the Nevada-California border region.

## 4.5 Summary

GrowClust is a new and open-source earthquake relocation algorithm. It uses waveform cross-correlation input data – differential times and cross-correlation values – in a hybrid, hierarchical clustering algorithm that simultaneously groups and relocates earthquakes in similar event clusters. The method is fast, flexible, and robust with respect data outliers. GrowClust also includes a built-in mechanism for performing uncertainty analysis that gives users a more complete assessment of the resolving power of the relocation results. We apply the GrowClust method to two prominent Nevada earthquake sequences: the 2012–2015 Spanish Springs and the 2014–present Sheldon swarms. The encouraging results for these examples demonstrate the scientific potential for large-scale relocation efforts within the region using the GrowClust algorithm.

## Data and Resources

The waveform data used in this study are archived locally by the Nevada Seismological Laboratory and is publicly available from the IRIS Data Center (http://ds.iris.edu/ds/nodes/dmc/, last accessed 9/29/2016). Initial catalog positions are consistent with the ANSS Comprehensive Earthquake Catalog ( http:

//earthquake.usgs.gov/data/comcat/, last accessed 10/04/2016). We use Antelope/Datascope software (http://www.brtt.com/, last accessed 09/01/2016) to extract waveforms for each event. We provide catalogs that list both initial and relocated positions for the Spanish Springs and Sheldon sequences analyzed in this paper in the electronic supplement to the article.

The GrowClust relocation codes described in this paper comprise an open-source software package under the GNU General Public License v.3. The GrowClust source distribution, which includes source codes, a user guide, and an example data set, is publicly available for download at https://github.com/dttrugman/GrowClust. We use in this paper the HypoDD implementation of the double-difference technique (available at http://www.ldeo.columbia.edu/~felixw/hypoDD.html, last accessed 10/06/2016) for comparison to GrowClust relocations.

# Acknowledgements

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program (NSFGRFP) under grant number DGE-1144086. Additional support was provided by the United States Geological Survey - National Earthquake Hazards Reduction Program (USGS-NEHRP) grant number G15AS00037-2016-0072. We thank K. Smith, R. Hatch, and C. Ruhl for their guidance to the application of GrowClust to Nevada seismicity, and R. Matoza, R. Abercrombie, and A. Borsa for stimulating discussions that helped

improve the algorithm design. We also thank editor Z. Peng and two anonymous reviewers for their useful suggestions to improve this manuscript.

Chapter 4, in full, is a reformatted version of material as it appears in Seismological Research Letters: Trugman, D. T. and P. M. Shearer (2017), GrowClust: A hierarchical clustering algorithm for relative earthquake relocation, with application to the Spanish Springs and Sheldon, Nevada, earthquake sequences, *Seismological Research Letters*, 88(2A), doi:10.1785/0220160188. I was the primary investigator and author of this paper.

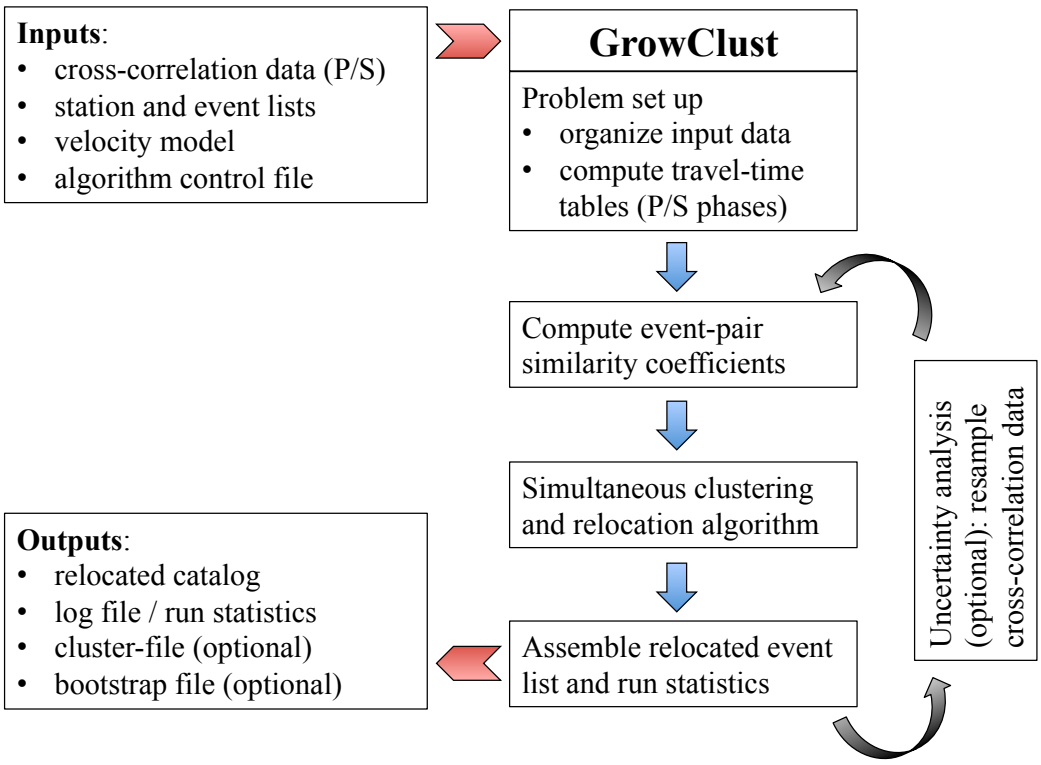


**Figure 4.1**: Schematic chart illustrating the basic GrowClust workflow. Users input cross-correlation data, station and event lists, a velocity model, and an algorithm control file into the GrowClust source code. GrowClust then organizes the input data and creates travel-time tables for $P$ and $S$ body wave phases. After computing event-pair similarity coefficients, GrowClust initiates its simultaneous clustering and relocation algorithm. Finally, after obtaining the final, relocated event positions, the program (optionally) resamples the input data for bootstrap uncertainty analysis, or proceeds directly to output of catalog, cluster, and log files.

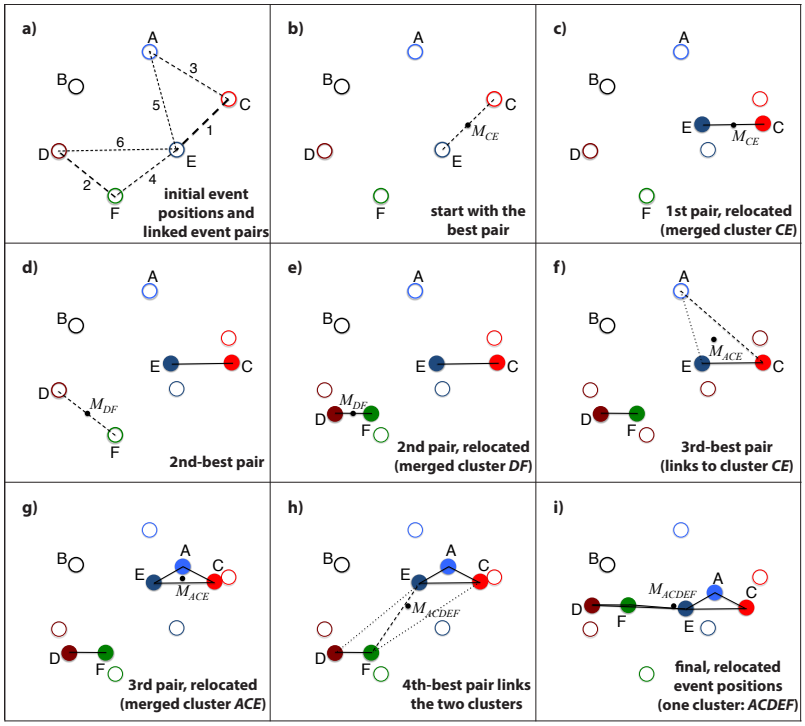


**Figure 4.2**: Simplified example of GrowClust's hybrid clustering and relocation algorithm. In this example, there are 6 events (labeled with capital letters) and 6 linked event pairs with quality cross-correlation data (labeled with numbers, in sequential order of waveform similarity). Event B has no links, and hence its location is held fixed throughout the algorithm. a) Initial catalog location of events A through F, with event pairs linked and numbered in order of decreasing waveform similarity. Note that all events begin as single-event clusters. b) The most-similar event pair, $C$-$E$, and its centroid $M_{CE}$. c) The relocation algorithm begins by relocating the most-similar event pair about its centroid. The relocated pair now belong to a single cluster, $CE$. d) The second event pair, $D$-$F$, and its centroid $M_{DF}$. e) Relocation of $D$-$F$ about its centroid. The relocated events are merged into a single cluster, $DF$. f) The third link connects event A to cluster $CE$. g) Event A and cluster $CE$ are now relocated about the total centroid $M_{ACE}$, and merged into a single cluster, $ACE$. Note that the relative positions of events C and E are held fixed. h) The fourth link connects clusters $DF$ and $ACE$. i) Clusters $DF$ and $ACE$ are relocated about their total centroid $M_{ACDEF}$, and merged into a single cluster. All events are now in their final, relocated positions. Note that links 5 and 6 are used in support of the final proposed cluster merger.

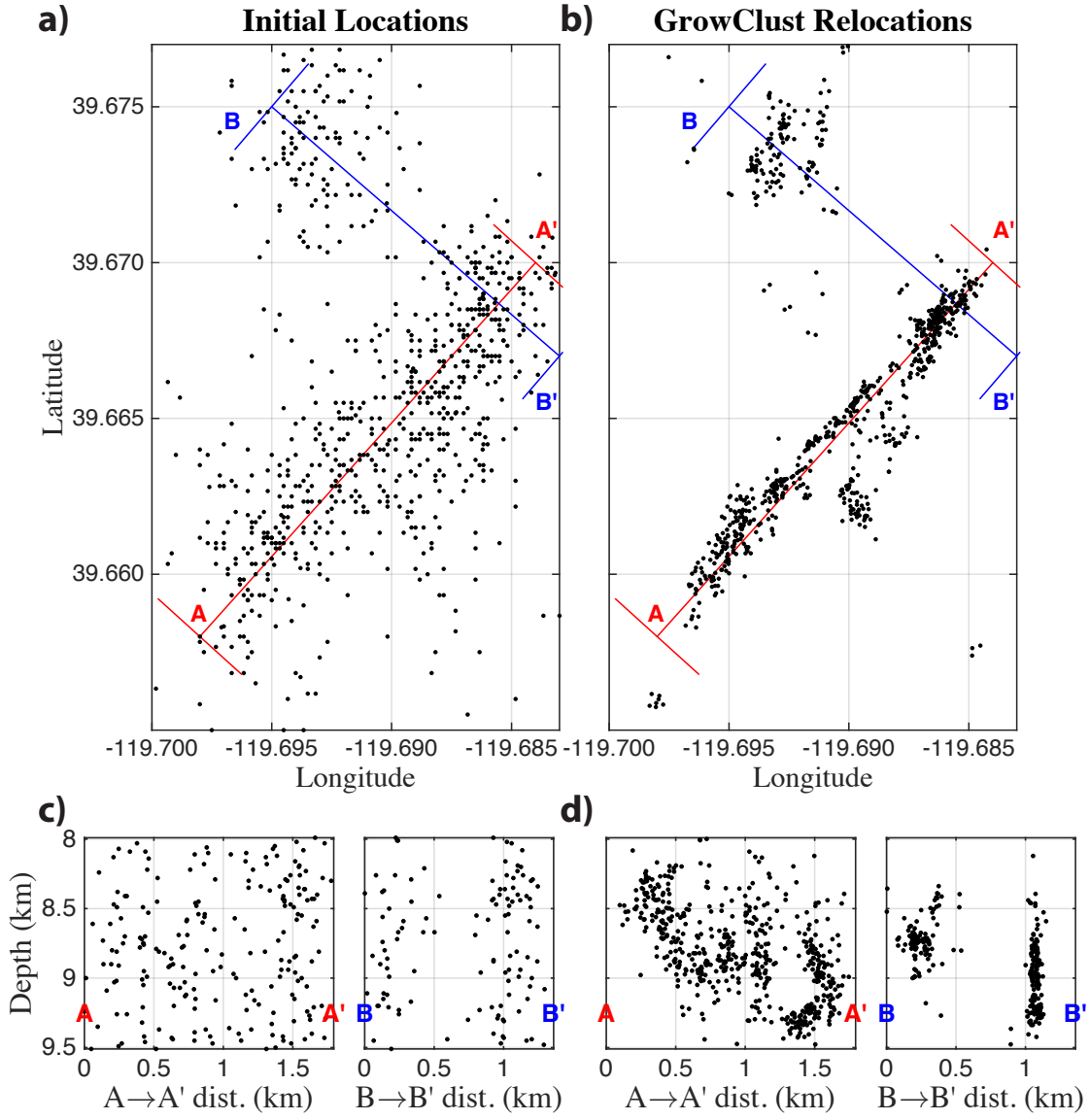


**Figure 4.3**: Comparison of initial ANSS catalog locations and GrowClust relocations for the 2012–2015 Spanish Springs sequence. Panels (a, catalog) and (b, GrowClust) provide a map view comparison of the initial and relocated event positions. Panels (c, catalog) and (d, GrowClust) provide a comparison of fault-parallel ($A \rightarrow A'$, red) and fault-perpendicular cross-sections ($B \rightarrow B'$, blue).

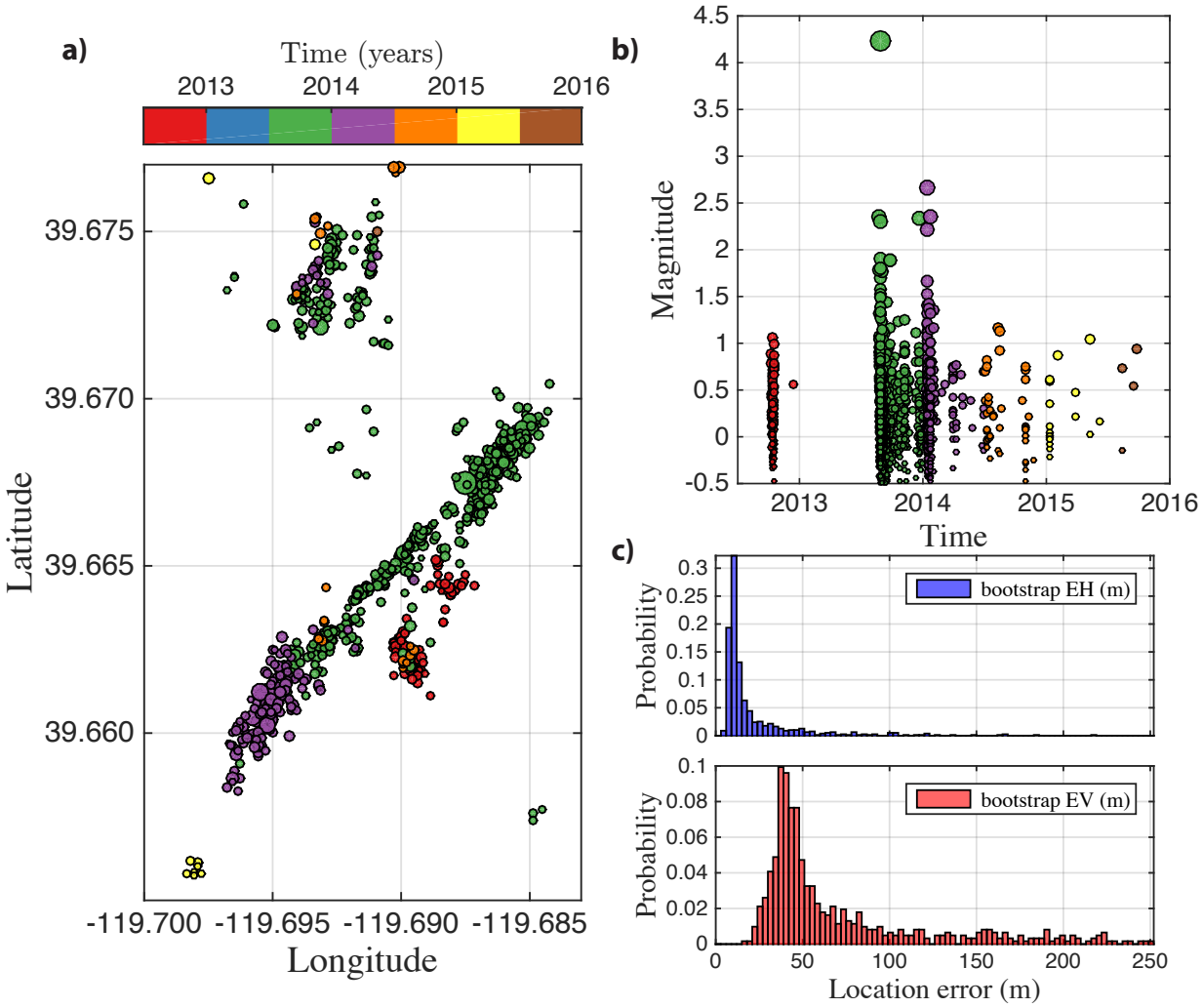


**Figure 4.4**: Space-time evolution of the 2012–2015 Spanish springs sequence. a) Map view of the Spanish Springs sequence, with events color-coded by occurrence time (colors are discretized in time bins of 0.5 years from 2012.5 through 2016). b) Magnitude as a function of time for these events (colors consistent with panel a). c) Histograms of horizontal (top, blue) and vertical (bottom, light red) location uncertainties for these events (obtained via 100 bootstrap resamplings of the input cross-correlation data, see text).

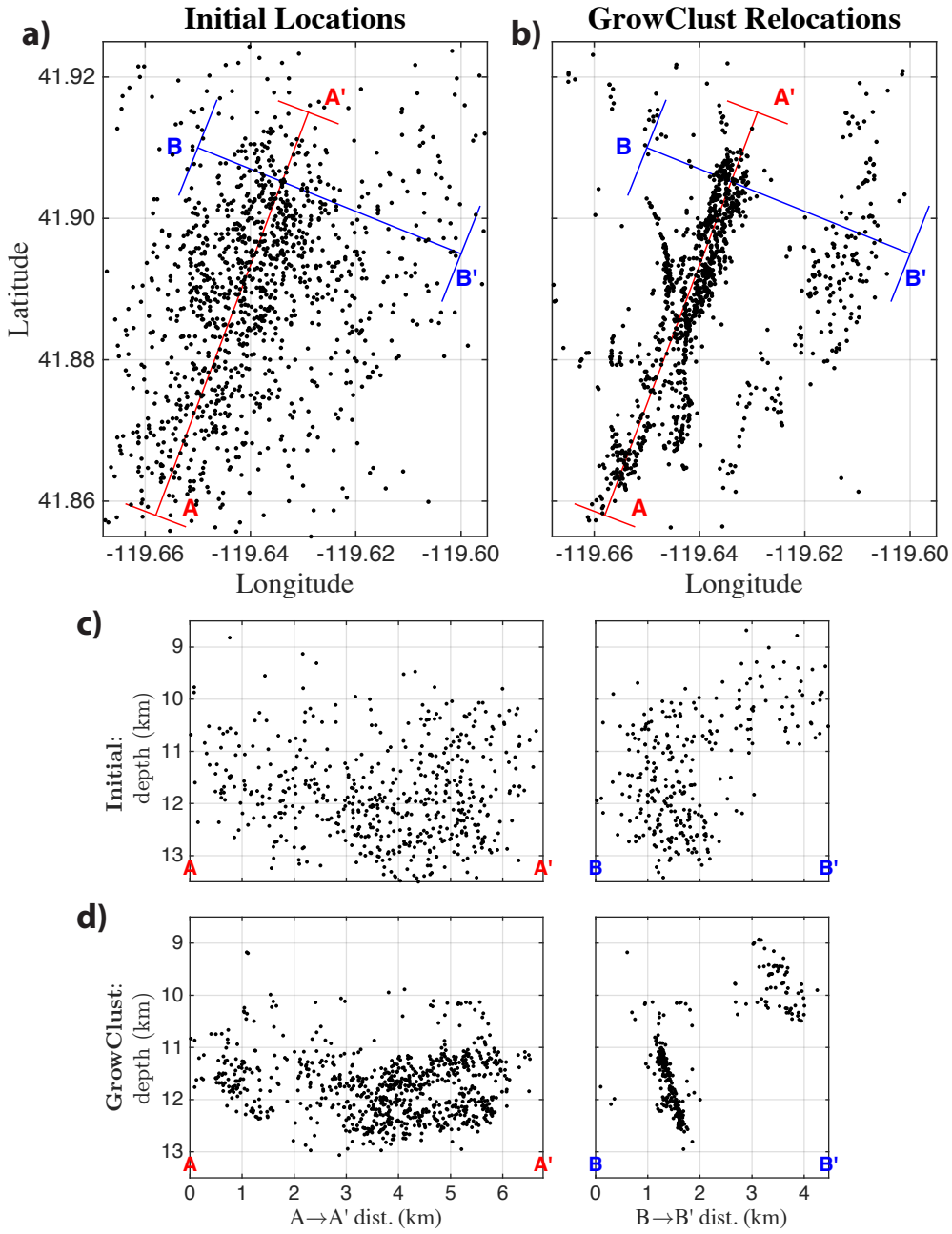


**Figure 4.5**: Comparison of initial ANSS catalog locations and GrowClust relocations for the 2014–present Sheldon sequence. Panels (a, catalog) and (b, GrowClust) provide a map view comparison of the initial and relocated event positions. Panels (c, catalog) and (d, GrowClust) provide a comparison of fault-parallel ($A \rightarrow A'$, red) and fault-perpendicular cross-sections ($B \rightarrow B'$, blue).

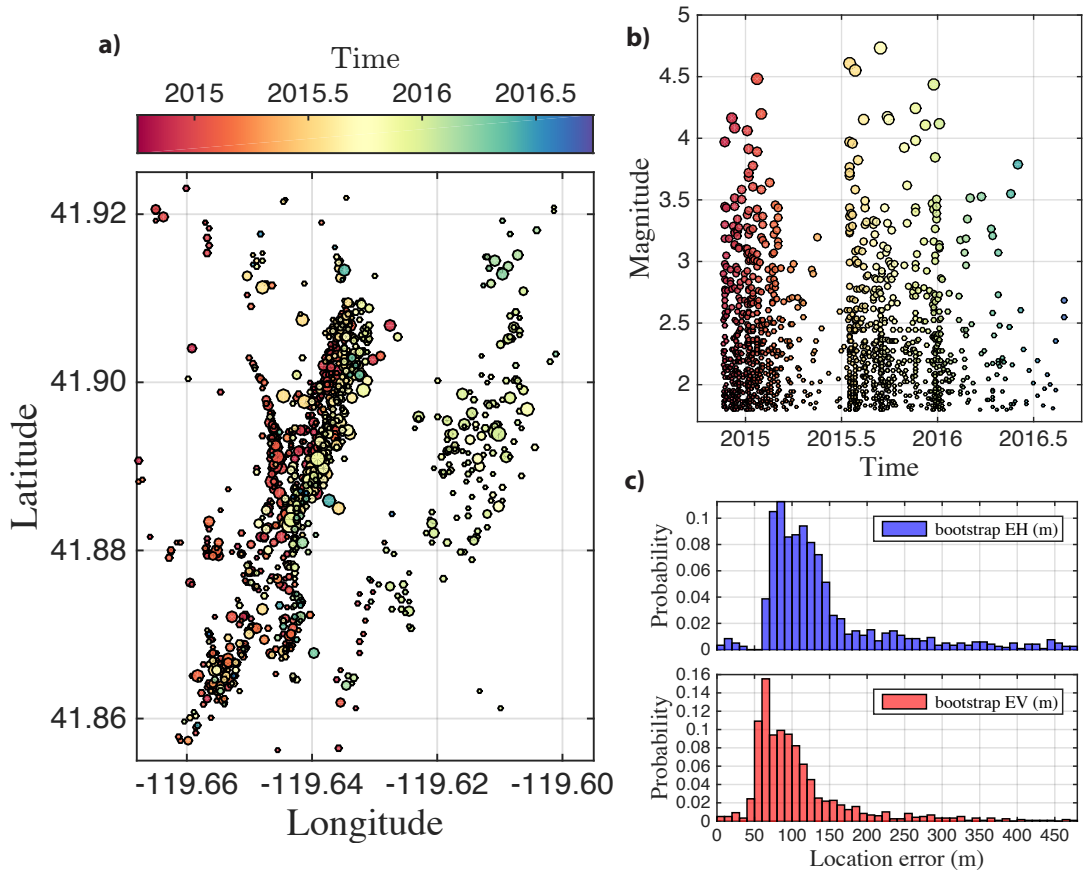


**Figure 4.6**: Space-time evolution of the 2014–present Sheldon sequence. a) Map view of the Sheldon sequence, with events color-coded by occurrence time (colors scale varies continuously from 2014.75 to 2016.75). b) Magnitude as a function of time for these events (colors consistent with panel a). c) Histograms of horizontal (top, blue) and vertical (bottom, light red) location uncertainties for these events (obtained via 100 bootstrap resamplings of the input cross-correlation data, see text).

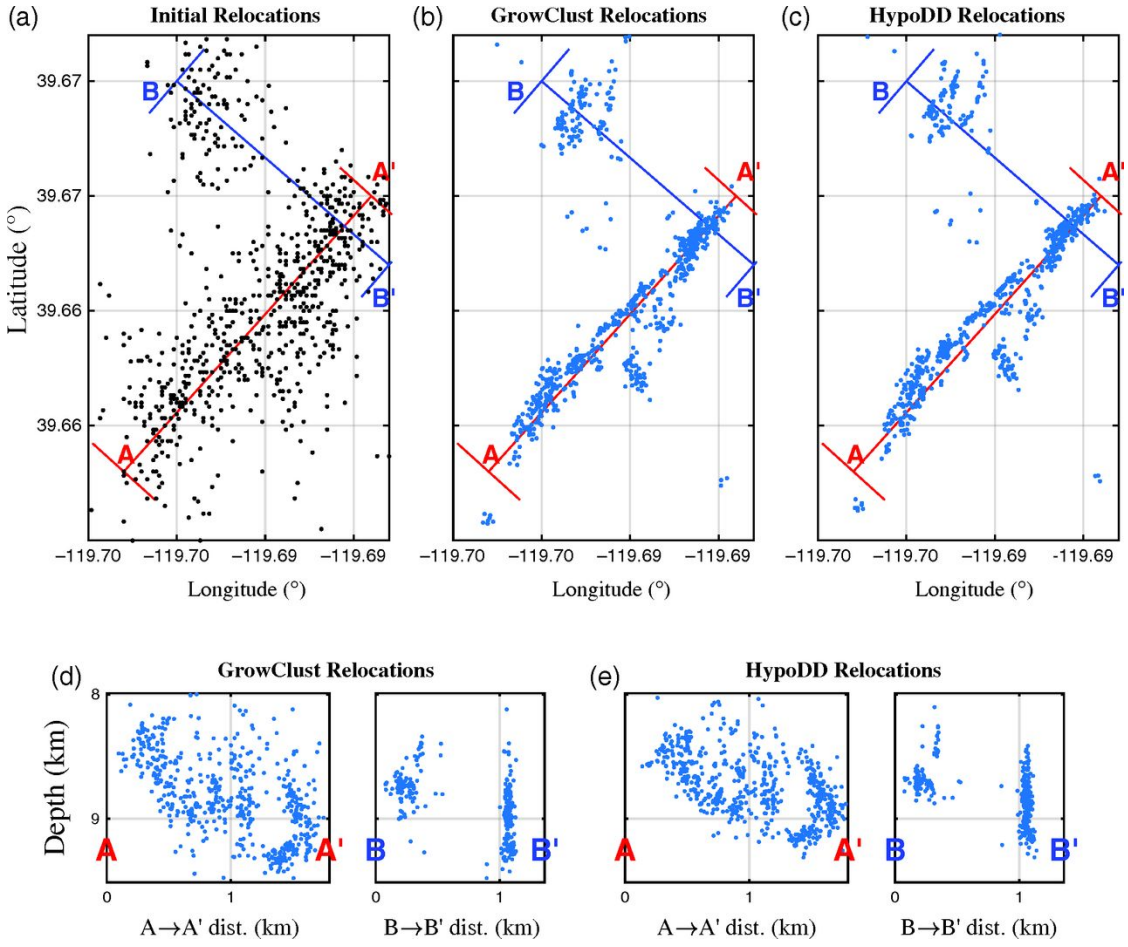


**Figure 4.7**: Comparison of relocation results for the Spanish Springs sequence using the GrowClust and HypoDD methods. Panels b) and c) show the map view improvement of GrowClust and HypoDD relocations over the initial catalog locations (panel a), while panels d) and e) compare GrowClust and HypoDD relocations in cross-section. Only those events that are successfully relocated by both methods are shown, and events are color-coded by the cluster ID number of each method. For the Spanish Springs sequence, both GrowClust and HypoDD partition events into one dominant cluster.

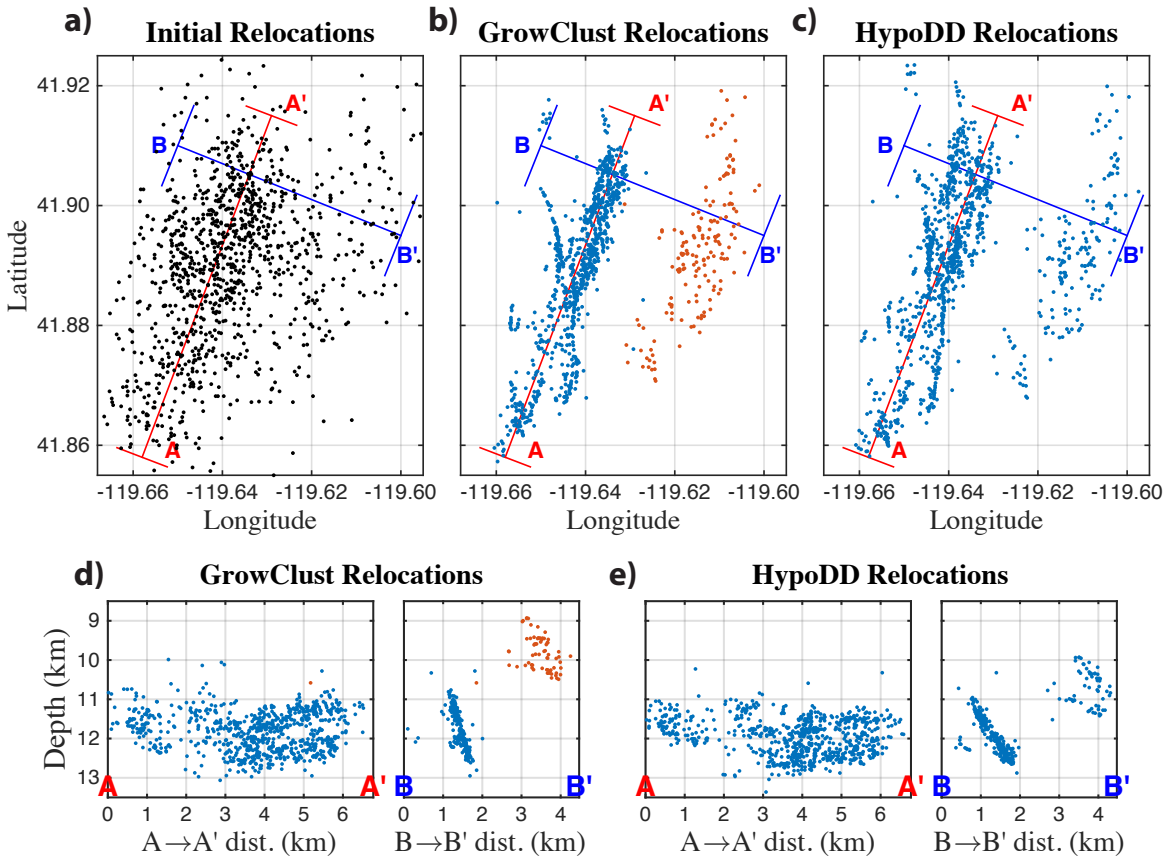


**Figure 4.8**: Comparison of relocation results for the Sheldon sequence using the GrowClust and HypoDD methods. Panels b) and c) show the map view improvement of GrowClust and HypoDD relocations over the initial catalog locations (panel a), while panels d) and e) compare GrowClust and HypoDD relocations in cross-section. Only those events that are successfully relocated by both methods are shown, and events are color-coded by the cluster ID number of each method. For the Sheldon sequence, GrowClust partitions events into two major clusters, while HypoDD uses only one.

## 4.6 Supplementary Materials

### Overview

Figure 4.S1 provides an overview map of the western Nevada study region, including the locations of the seismic stations used in this analysis. Figures 4.S2 and 4.S3 show the spatial distribution of relative location uncertainties for the Spanish Springs and Sheldon sequences, obtained through GrowClust's nonparametric error analysis subroutines described in the main text. Table 4.S1 lists the velocity model assumed in the relocation results presented in this study.

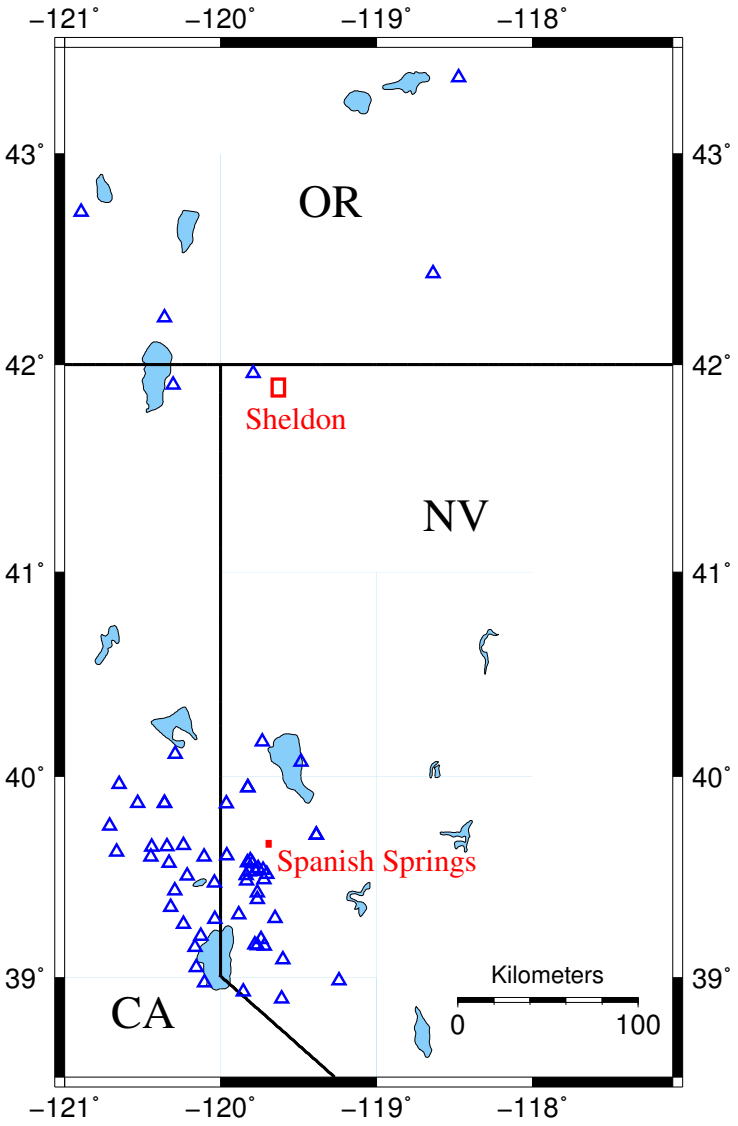


**Figure 4.S1**: Regional map for the Spanish Springs and Sheldon sequences. The rectangular study regions for the two sequences are outlined with red rectangles, and the seismic stations associated with the waveform data used in this study are marked with blue triangles.

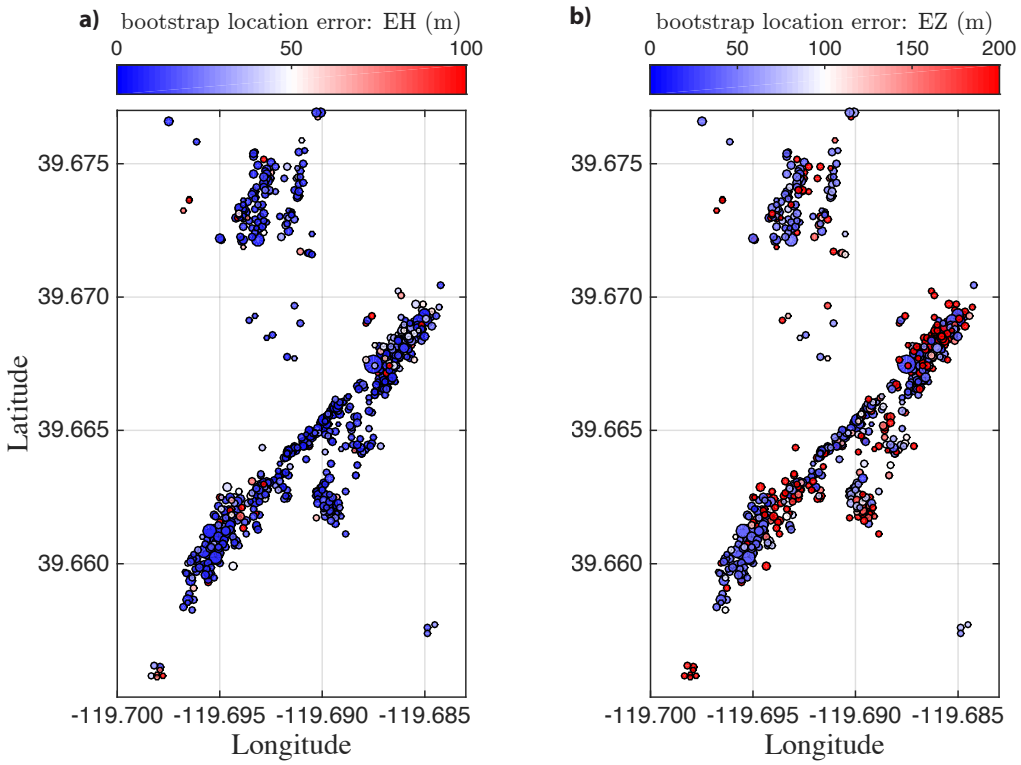


**Figure 4.S2**: Map view of the Spanish Springs sequence, with events color-coded by a) horizontal and b) vertical location error. Uncertainty estimates are obtained via 100 bootstrap resamplings of the input cross-correlation data, see main text for details.

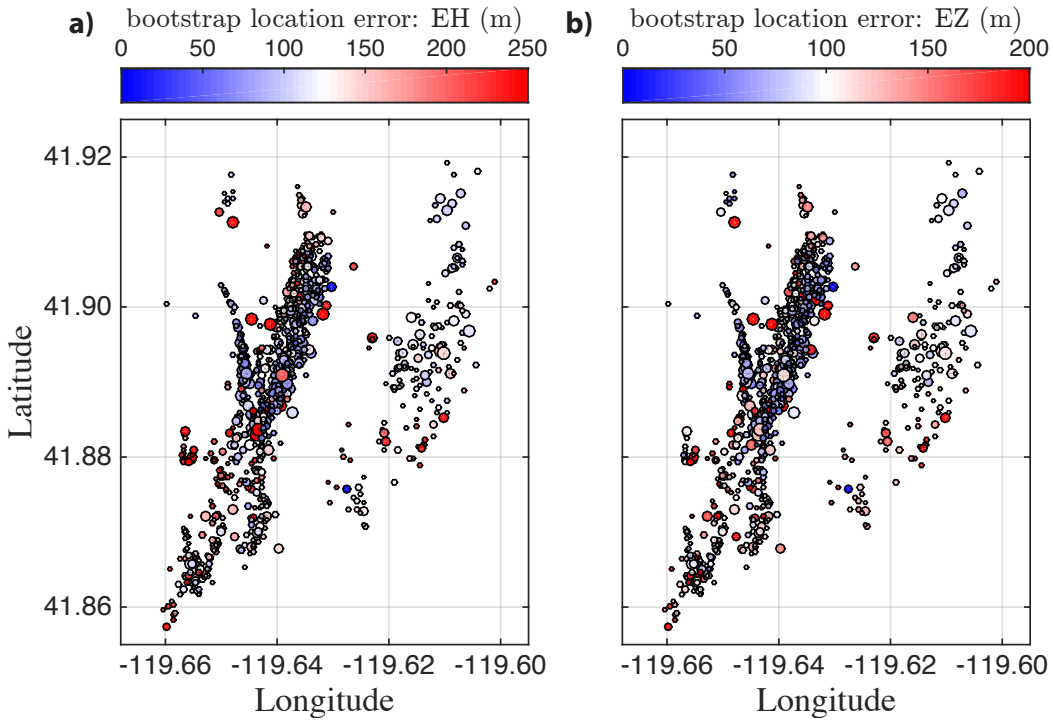


**Figure 4.S3**: Map view of the Sheldon sequence, with events color-coded by a) horizontal and b) vertical location error. Uncertainty estimates are obtained via 100 bootstrap resamplings of the input cross-correlation data, see main text for details.

**Table 4.S1**: Velocity Model Used for GrowClust Relocations of the Spanish Springs and Sheldon earthquake sequences. Column 1 lists the depth to the top of each constant velocity layer (km), and column 2 lists the corresponding P wave velocity for that layer (km/s). The velocity model assumes a layered structure with constant $P/S$ velocity ratio $V_p/V_s = \sqrt{3}$.

| **depth (km)** | $V_P$ **(km/s)** |
|---|---|
| 0.0 | 3.0 |
| 1.0 | 4.5 |
| 2.0 | 5.5 |
| 4.0 | 6.0 |
| 7.0 | 6.1 |
| 12.0 | 6.2 |
| 18.0 | 6.4 |
| 28.0 | 6.8 |
| 38.0 | 7.8 |
| 50.0 | 8.0 |

# References

Astiz, L., and P. M. Shearer (2000), Earthquake Locations in the Inner Continental Borderland, Offshore Southern California, *Bulletin of the Seismological Society of America*, *90*(2), 425–449, doi:10.1785/0119990022.

Buland, R. (1976), The mechanics of locating earthquakes, *Bulletin of the Seismological Society of America*, *66*(1), 173–187.

Dingler, J., G. Kent, N. Driscoll, J. Babcock, A. Harding, G. Seitz, B. Karlin, and C. Goldman (2009), A high-resolution seismic CHIRP investigation of active normal faulting across Lake Tahoe Basin, California-Nevada, *Geological Society of America Bulletin*, *121*(7-8), 1089–1107, doi:10.1130/B26244.1.

Douglas, A. (1967), Joint Epicentre Determination, *Nature*, *215*(5096), 47–48, doi:10.1038/215047a0.

Efron, B., and R. J. Tibshirani (1994), *An Introduction to the Bootstrap*, CRC press, Boca Raton, FL.

Faulds, J. E., C. D. Henry, and N. H. Hinz (2005), Kinematics of the northern Walker Lane: An incipient transform fault along the PacificNorth American plate boundary, *Geology*, *33*(6), 505, doi:10.1130/G21274.1.

Fremont, M.-J., and S. D. Malone (1987), High precision relative locations of earthquakes at Mount St. Helens, Washington, *Journal of Geophysical Research: Solid Earth*, *92*(B10), 10,223–10,236, doi:10.1029/JB092iB10p10223.

Frohlich, C. (1979), An efficient method for joint hypocenter determination for large groups of earthquakes, *Computers & Geosciences*, *5*(3), 387–389, doi:10.1016/0098-3004(79)90034-7.

Geiger, L. (1910), Herdbestimmung bei Erdbeben aus den Ankunftzeiten, *K. Gessel. Wiss. Goett*, *4*, 331–349.

Geiger, L. (1912), Probability method for the determination of earthquake epicenters from the arrival time only, *Bull. St. Louis Univ*, *8*, 60–71.

Got, J.-L., J. Fréchet, and F. W. Klein (1994), Deep fault plane geometry inferred from multiplet relative relocation beneath the south flank of Kilauea, *Journal of Geophysical Research: Solid Earth*, *99*(B8), 15,375–15,386, doi:10.1029/94JB00577.

Hainzl, S. (2002), Indications for a successively triggered rupture growth underlying the 2000 earthquake swarm in Vogtland/NW Bohemia, *Journal of Geophysical Research*, *107*(B12), doi:10.1029/2002JB001865.

Hauksson, E., W. Yang, and P. M. Shearer (2012), Waveform Relocated Earthquake Catalog for Southern California (1981 to June 2011), *Bulletin of the Seismological Society of America*, *102*(5), 2239–2244, doi:10.1785/0120120010.

Hesterberg, T., S. Monaghan, D. S. Moore, A. Clipson, and R. Epstein (2008), Bootstrap Methods and Permutation Tests, in *The Practice of Business Statistics: Using Data for Decisions*, 2nd ed., pp. 169–183, W. H. Freeman, New York, NY.

Ichinose, G. A., K. D. Smith, and J. G. Anderson (1998), Moment tensor solutions of the 1994 to 1996 Double Spring Flat, Nevada, earthquake sequence and implications for local tectonic models, *Bulletin of the Seismological Society of America*, *88*(6), 1363–1378.

Ichinose, G. A., J. G. Anderson, K. D. Smith, and Y. Zeng (2003), Source Parameters of Eastern California and Western Nevada Earthquakes from Regional Moment Tensor Inversion, *Bulletin of the Seismological Society of America*, *93*(1), 61–84, doi:10.1785/0120020063.

Ito, A. (1985), High Resolution Relative Hypocenters of Similar Earthquakes by Cross-Spectral Analysis Method, *Journal of Physics of the Earth*, *33*(4), 279–294, doi:10.4294/jpe1952.33.279.

Jordan, T. H., and K. A. Sverdrup (1981), Teleseismic location techniques and their application to earthquake clusters in the South-Central Pacific, *Bulletin of the Seismological Society of America*, *71*(4), 1105–1130.

Kent, G., K. D. Smith, and G. Biasi (2015), Final Report: Nevada Seismic Network, *Tech. Rep. G10AC00090*, University of Nevada Reno, Reno, Nevada.

Klein, F. W. (2002), User's guide to HYPOINVERSE-2000, a Fortran program to solve for earthquake locations and magnitudes, *Report 2002-171*, U.S Dept. of the Interior, Geological Survey, Reston, VA.

Leys, C., C. Ley, O. Klein, P. Bernard, and L. Licata (2013), Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median, *Journal of Experimental Social Psychology*, *49*(4), 764–766, doi:10.1016/j.jesp.2013.03.013.

Lin, G., P. M. Shearer, and E. Hauksson (2007), Applying a three-dimensional velocity model, waveform cross correlation, and cluster analysis to locate southern California seismicity from 1981 to 2005, *Journal of Geophysical Research*, *112*(B12), doi:10.1029/2007JB004986.

Lohman, R. B., and J. J. McGuire (2007), Earthquake swarms driven by aseismic creep in the Salton Trough, California, *Journal of Geophysical Research*, *112*(B4), doi:10.1029/2006JB004596.

Mammen, E. (2012), *When Does Bootstrap Work?: Asymptotic Results and Simulations*, Springer Science & Business Media.

Matoza, R. S., P. M. Shearer, G. Lin, C. J. Wolfe, and P. G. Okubo (2013), Systematic relocation of seismicity on Hawaii Island from 1992 to 2009 using waveform cross correlation and cluster analysis, *Journal of Geophysical Research: Solid Earth*, *118*(5), 2275–2288, doi:10.1002/jgrb.50189.

Nadeau, R. M., W. Foxall, and T. V. McEvilly (1995), Clustering and Periodic Recurrence of Microearthquakes on the San Andreas Fault at Parkfield, California, *Science*, *267*(5197), 503–507, doi:10.1126/science.267.5197.503.

Phillips, W. S., L. S. House, and M. C. Fehler (1997), Detailed joint structure in a geothermal reservoir from studies of induced microearthquake clusters, *Journal of Geophysical Research: Solid Earth*, *102*(B6), 11,745–11,763, doi:10.1029/97JB00762.

Poupinet, G., W. L. Ellsworth, and J. Frechet (1984), Monitoring velocity variations in the crust using earthquake doublets: An application to the Calaveras Fault, California, *Journal of Geophysical Research: Solid Earth*, *89*(B7), 5719–5731, doi:10.1029/JB089iB07p05719.

Ramelli, A. R., J. W. Bell, C. M. dePolo, and J. C. Yount (1999), Large-magnitude, late Holocene earthquakes on the Genoa fault, west-central Nevada and eastern California, *Bulletin of the Seismological Society of America*, *89*(6), 1458–1472.

Richards-Dinger, K. B., and P. M. Shearer (2000), Earthquake locations in southern California obtained using source-specific station terms, *Journal of Geophysical Research: Solid Earth*, *105*(B5), 10,939–10,960, doi:10.1029/2000JB900014.

Rowe, C. A., R. C. Aster, W. S. Phillips, R. H. Jones, B. Borchers, and M. C. Fehler (2002), Using Automated, High-precision Repicking to Improve Delineation of Microseismic Structures at the Soultz Geothermal Reservoir, *Pure and Applied Geophysics*, *159*(1-3), 563–596, doi:10.1007/PL00001265.

Rubin, A. M., D. Gillard, and J.-L. Got (1999), Streaks of microearthquakes along creeping faults, *Nature*, *400*(6745), 635–641, doi:10.1038/23196.

Ruhl, C. (2015), Inside an Earthquake Swarm: Objective Identification and Analysis of Spatiotemporal Subclusters of the Mogul 2008 Earthquake Swarm in Reno, NV, in *AGU Fall Meeting Abstracts*, San Francisco, CA.

Ruhl, C. J., K. D. Smith, and R. E. Abercrombie (2016), Moderate-to-small magnitude seismicity clusters persist in Northwest Nevada for over 19 months, in *SSA Annual Meeting Abstracts*, Reno, Nevada.

Shearer, P. (2005), Southern California Hypocenter Relocation with Waveform Cross-Correlation, Part 2: Results Using Source-Specific Station Terms and Cluster Analysis, *Bulletin of the Seismological Society of America*, *95*(3), 904–915, doi:10.1785/0120040168.

Shearer, P. M. (1997), Improving local earthquake locations using the L1 norm and waveform cross correlation: Application to the Whittier Narrows, California, aftershock sequence, *Journal of Geophysical Research: Solid Earth*, *102*(B4), 8269–8283, doi:10.1029/96JB03228.

Shearer, P. M. (1998), Evidence from a cluster of small earthquakes for a fault at 18 km depth beneath Oak Ridge, southern California, *Bulletin of the Seismological Society of America*, *88*(6), 1327–1336.

Shearer, P. M. (2002), Parallel fault strands at 9-km depth resolved on the Imperial Fault, Southern California, *Geophysical Research Letters*, *29*(14), 19–1–19–4, doi:10.1029/2002GL015302.

Shearer, P. M. (2012), Space-time clustering of seismicity in California and the distance dependence of earthquake triggering, *Journal of Geophysical Research: Solid Earth*, *117*(B10), doi:10.1029/2012JB009471.

Shelly, D. R., W. L. Ellsworth, and D. P. Hill (2016), Fluid-faulting evolution in high definition: Connecting fault structure and frequency-magnitude variations during the 2014 Long Valley Caldera, California, earthquake swarm, *Journal of Geophysical Research: Solid Earth*, *121*(3), 1776–1795, doi:10.1002/2015JB012719.

Smith, K., D. H. von Seggern, D. Depolo, J. G. Anderson, G. P. Biasi, and R. Anooshehpoor (2008), Seismicity of the 2008 Mogul-Somersett West Reno, Nevada Earthquake Sequence, in *AGU Fall Meeting Abstracts (2008)*, vol. 53.

Thurber, C., and D. Eberhart-Phillips (1999), Local earthquake tomography with flexible gridding, *Computers & Geosciences*, *25*(7), 809–818, doi:10.1016/S0098-3004(99)00007-2.

Thurber, C. H. (1983), Earthquake locations and three-dimensional crustal structure in the Coyote Lake Area, central California, *Journal of Geophysical Research: Solid Earth*, *88*(B10), 8226–8236, doi:10.1029/JB088iB10p08226.

Thurber, C. H. (1985), Nonlinear earthquake location: Theory and examples, *Bulletin of the Seismological Society of America*, *75*(3), 779–790.

Thurber, C. H. (1992), Hypocenter-velocity structure coupling in local earthquake tomography, *Physics of the Earth and Planetary Interiors*, *75*(1), 55–62, doi:10.1016/0031-9201(92)90117-E.

VanWormer, J. D., and A. S. Ryall (1980), Sierra Nevada-Great Basin boundary zone: Earthquake hazard related to structure, active tectonic processes, and anomalous patterns of earthquake occurrence, *Bulletin of the Seismological Society of America*, *70*(5), 1557–1572.

Waldhauser, F., and W. L. Ellsworth (2000), A Double-Difference Earthquake Location Algorithm: Method and Application to the Northern Hayward Fault, California, *Bulletin of the Seismological Society of America*, *90*(6), 1353–1368, doi:10.1785/0120000006.

Waldhauser, F., and D. P. Schaff (2008), Large-scale relocation of two decades of Northern California seismicity using cross-correlation and double-difference methods, *Journal of Geophysical Research*, *113*(B8), doi:10.1029/2007JB005479.

Waldhauser, F., W. L. Ellsworth, and A. Cole (1999), Slip-parallel seismic lineations on the Northern Hayward Fault, California, *Geophysical Research Letters*, *26*(23), 3525–3528, doi:10.1029/1999GL010462.

Wesnousky, S. G., J. M. Bormann, C. Kreemer, W. C. Hammond, and J. N. Brune (2012), Neotectonics, geodesy, and seismic hazard in the Northern Walker Lane of Western North America: Thirty kilometers of crustal shear and no strike-slip?, *Earth and Planetary Science Letters*, *329-330*, 133–140, doi:10.1016/j.epsl.2012.02.018.

# Chapter 5

# Application of an improved spectral decomposition method to examine earthquake source scaling in southern California

## Abstract

Earthquake source spectra contain fundamental information about the dynamics of earthquake rupture. However, the inherent tradeoffs in separating source and path effects, when combined with limitations in recorded signal bandwidth, make it challenging to obtain reliable source spectral estimates for large earthquake data sets. We present here a stable and statistically robust spectral decomposi-

tion method that iteratively partitions the observed waveform spectra into source, receiver, and path terms. Unlike previous methods of its kind, our new approach provides formal uncertainty estimates and does not assume self-similar scaling in earthquake source properties. Its computational efficiency allows us to examine large data sets (tens of thousands of earthquakes) that would be impractical to analyze using standard empirical Green's function-based approaches. We apply the spectral decomposition technique to $P$-wave spectra from five areas of active contemporary seismicity in southern California: the Yuha Desert, the San Jacinto Fault, and the Big Bear, Landers, and Hector Mine regions of the Mojave Desert. We show that the source spectra are generally consistent with an increase in median Brune-type stress drop with seismic moment, but that this observed deviation from self-similar scaling is both model-dependent and varies in strength from region to region. We also present evidence for significant variations in median stress drop and stress drop variability on regional and local length scales. These results both contribute to our current understanding of earthquake source physics and have practical implications for the next generation of ground-motion prediction assessments.

## 5.1 Introduction

Analyses of waveforms recorded by arrays of seismic stations provide the most fundamental observational constraints on earthquake occurrence. While the

arrival times of different phases of seismic energy are routinely used to determine the hypocentral location, the detailed features of the spectra of recorded waveforms give a snapshot of the earthquake source, and can yield insight into the complex physical processes underlying earthquake nucleation, rupture, and arrest. Of particular interest is the earthquake source spectrum, the frequency-domain analog of the moment-rate (or source-time) function that describes the temporal evolution of slip on the fault interface. The amplitude and shape of source spectra can be used to infer key earthquake source parameters like seismic moment, radiated energy, and stress drop that are essential for seismic hazard assessment.

The waveform spectrum recorded by a seismometer is a convolution of source, path, and receiver terms. To isolate source spectra, one must somehow correct for these effects. There exist a number of established approaches for attempting to do so, including empirical Green's function methods tailored for detailed study of individual earthquakes (e.g., *Mori and Frankel* 1990; *Hough* 1997; *Prieto et al.* 2006; *Abercrombie* 2013; *Huang et al.* 2016), and larger-scale stacking and generalized inversion approaches (e.g., *Castro et al.* 1990; *Shearer et al.* 2006; *Oth et al.* 2011) that are more readily applicable to larger data sets of earthquakes. While each technique has its own strengths, none can perfectly resolve the contribution of the earthquake source to the observed waveform spectra, even in idealized scenarios. This tradeoff between source and path effects, when combined with noise and the limited signal bandwidth of raw waveform data, implies that earthquake source parameter estimates, while critical to our understanding of earthquake haz-

ard, are subject to considerable uncertainty (e.g., *Hough* 1996; *Kane et al.* 2011; *Cotton et al.* 2013). Thus, any technique designed to study earthquake source spectra and source parameters should include realistic uncertainty estimates.

Here we describe a spectral decomposition technique that is capable of providing reliable source spectral estimates for large ensembles of earthquakes. The method builds upon the basic approach used by *Shearer et al.* (2006) to analyze $P$-wave spectra of more than 60,000 $M_L$ 1.5–3.1 earthquakes occurring in southern California from 1989 to 2001. The technique uses an iterative, robust-least-squares algorithm to partition the observed spectra into source, station, and travel time (path) terms, with L1-norm weights applied to large misfit residuals to mitigate the influence of the outliers pervasive in seismic data sets. We develop a nonparametric resampling approach to estimate source parameter uncertainties, and apply an automated algorithm to detect and discard clipped waveforms that are common for larger events.

A crucial improvement of our new technique over that of *Shearer et al.* (2006) is that our approach does not presume the self-similar scaling of earthquake source properties first proposed by *Aki* (1967), in which stress drop is constant with moment. While self-similarity is intuitively appealing and has been supported in many studies (e.g., *Choy and Boatwright* 1995; *Ide and Beroza* 2001; *Ide* 2003; *Prieto* 2004; *Allmann and Shearer* 2009; *Baltay et al.* 2010, 2011; *Abercrombie et al.* 2017), other studies have found evidence for a slight increase in stress drop or scaled energy with moment (e.g., *Mayeda and Walter* 1996; *Izutani and Kanamori* 2001;

*Mori et al.* 2003; *Mayeda et al.* 2005; *Takahashi et al.* 2005; *Mayeda et al.* 2007; *Calderoni et al.* 2013; *Pacor et al.* 2016a; *Lin et al.* 2016; *Poli and Prieto* 2016). This controversy over self-similarity has important theoretical implications for our understanding of the scale dependence of earthquake rupture processes (*Kanamori and Rivera*, 2004; *Abercrombie and Rice*, 2005; *Walter et al.*, 2006; *Cocco et al.*, 2016), as well as practical ramifications for hazard analyses (e.g., *Field et al.* 2014; *Yenier and Atkinson* 2015; *Petersen et al.* 2016) that implicitly assume self-similar scaling.

We apply the improved spectral decomposition technique to revisit the question of earthquake scaling in southern California. We analyze source parameters and their scaling with moment for more than ten thousand **M**1 to **M**5 earthquakes in five regions of recent seismic activity (Figure 5.1) recorded by the modern (2002–2016) Southern California Seismic Network (SCSN) (*Hutton et al.*, 2010): (1) the Yuha Desert, including thousands of aftershocks of the 2010 $M_W$ 7.2 El Mayor-Cucapah earthquake, (2) the trifurcation zone of the San Jacinto Fault, in which the recent 2016 $M_W$ 5.2 Borrego Springs earthquake was located, (3) the Big Bear region on the western boundary of the Mojave Desert, and the rupture zones of the prominent (4) 1992 $M_W$ 7.3 Landers and (5) 1999 $M_W$ 7.1 Hector Mine earthquakes. We present evidence for an increase in median stress drop with moment for these regions, but also demonstrate that the observed deviation from self-similarity is sensitive to modeling assumptions and in particular on the assumed high-frequency falloff rate. We further show that there are significant regional and

local variations in median stress drop, stress drop variability, and source parameter scaling. This study outlines the computational framework for future studies to extend these preliminary findings to larger regions to obtain a updated catalog of source parameter estimates (*Shearer et al.*, 2006) that would provide key constraints on earthquake source physics and inform the next generation of ground motion prediction equations in southern California and worldwide.

## 5.2 Data and Methods

Our spectral decomposition technique to analyze earthquake source spectra has four main steps: (1) computation of $P$-wave spectra from the vertical-component waveform time series records of each earthquake, (2) application of an iterative, robust-least-squares inversion procedure to decompose the observed (data) spectra into event, station, and travel time terms, (3) stacking of relative event spectra in bins of spectral moment to estimate an empirical correction (or EGF) for average near-source attenuation and other path effects common to all events, and (4) estimation of corner frequency, seismic moment, and stress drop for individual events, with estimates of parameter uncertainties obtained using a nonparametric resampling method. Steps (1) and (2) are largely the same as the method described by (*Shearer et al.*, 2006), with minor modifications designed to improve the algorithm's robustness and capability of processing datasets over a wider magnitude range than the $M_L$ 1.5–3.1 considered for their study. Steps (3)

and (4), while conceptually similar, contain major modifications and improvements that allow for a more rigorous examination of self-similarity.

### 5.2.1 Waveform data and spectral computation

We use the Seismic Transfer Protocol (STP) tool of the Southern California Earthquake Data Center (SCEDC, http://scedc.caltech.edu/, last accessed 12/4/2016) to obtain raw waveform data and phase pick information from $M_L$ 1.1 and greater events occurring during the time period from January 2002 through September 2016 within the five regions listed in Figure 5.1. For each event, we consider waveforms from vertical-component, short-period and high-broadband channels (EHZ or HHZ) on stations within 80km. Most such records have 100Hz sampling frequencies, and we exclude all records with a lower sampling rate, while subsampling any 200Hz records to 100Hz to maximize data availability.

For each event, we select a magnitude-dependent window length for spectral computations ranging from 1.5 to 4.5 seconds, with longer windows used for events of higher magnitude to ensure adequate resolution of the spectral corner frequency (*Ross and Ben-Zion*, 2016; *Abercrombie et al.*, 2017). We define the signal window to begin 0.05s before the cataloged $P$ arrival to account for potential errors in the pick time, and define a noise window of equal length (used for signal-to-noise computations) immediately preceding the signal window. Clipped waveforms are common for near-source stations recording $\mathbf{M} \geq 3$ events (*Shearer et al.*, 2006; *Yang and Ben-Zion*, 2010), and we detect and exclude all waveforms flagged as

clipped using a classification algorithm that compares the probability distribution of observed waveform amplitudes to that of known, clipped waveforms. For the non-clipped waveforms, we then compute velocity amplitude spectra for the signal and noise windows using a multitaper algorithm (*Park et al.*, 1987; *Prieto et al.*, 2009). We convert these velocity spectra to displacement, resample all spectra to the frequency points corresponding to the shortest time windows (1.5s, which have the coarsest frequency sampling), and only further consider spectra with average signal-to-noise amplitude greater than 5 in each of five frequency bands (2.5–6, 6–10, 10–15, 15–20, and 20–25Hz). Note that these are somewhat more stringent quality control criteria than those used by *Shearer et al.* (2006), who used a fixed spectral window length, tested signal-to-noise from 5–20Hz, and did not explicitly check for clipped waveforms. We implemented the above pre-processing steps and selection criteria to ensure that we could reliably assess source scaling over a wide range of event magnitudes (both smaller and larger). The $P$-wave spectra meeting these quality control criteria comprise the input data to the spectral decomposition inversion technique described below.

### 5.2.2 Spectral decomposition and relative source spectra

The observed displacement spectra are a convolution of source, path, and station effects, the latter of which includes both site effects and the instrument response. In the frequency domain, the recorded spectra $d(f)$ can thus be written

as a product of event (source), path, and station spectra:

$$d(f) = e(f)\ p(f)\ st(f), \tag{5.1}$$

which can be linearized by working in the log-$f$ domain,

$$\log d(f) = \log e(f) + \log p(f) + \log st(f). \tag{5.2}$$

The basic idea behind the spectral decomposition technique is that because each earthquake will be recorded by many stations, each station will record many earthquakes, and each approximate source-receiver path will be traversed many times, it should be possible to solve for each contribution as part of an overdetermined system of equations defined by the input data spectra.

More precisely, consider the observed displacement spectra $d_{ij}$ corresponding to event $i$, recorded by station $j$. Approximating the path term $p_{ij}$ as a travel-time dependent term $tt_{k(i,j)}$ that depends primarily on the distance from the source to the station, we can write a linear equation of the form

$$d_{ij} = e_i + tt_{k(i,j)} + st_j + r_{ij}, \tag{5.3}$$

for each observed spectra, where $r_{ij}$ is a residual error term. Equation (5.3) defines an over-determined inverse problem in which the number of observations (recorded spectra, $d$) outnumber the desired model parameters $e, tt, st$. We discretize the travel-time terms (indexed by $k$) in bins of 1s of observed source-receiver travel time, and only further consider spectra from events that are observed by at least five stations, and stations that observe at least 20 events.

For stability and to mitigate the influence of data outliers, we use an iterative, robust-least-squares method to solve sequentially for model parameters $tt_k$, $st_j$, and $e_i$, where $L_1$-norm weights are applied to large misfit residuals at each iteration. The inversion is performed independently for each frequency point. The focus of this study is on the source terms $e_i$, so the travel-time and station terms $tt_k$ and $st_j$ are essentially nuisance parameters, but do contain useful information about midcrustal attenuation and site effects (e.g. *Castro et al.* 1990; *Shearer et al.* 2006; *Oth et al.* 2011) that may be of interest for ground motion and seismic hazard analyses. Note that strong near-surface attenuation and its lateral variability will largely be absorbed into the station (site) terms and the empirical Green's function correction described below.

### 5.2.3 Spectral stacking and EGF-corrected source spectra

One limitation of the above spectral decomposition is that it can only resolve relative differences between the source spectra of each event. Thus, it is necessary to further correct for propagation effects that are common to all paths (e.g., near-source attenuation and common site or instrument effects) in order to interpret the source spectra within the framework of theoretical earthquake source models. *Shearer et al.* (2006) resolved this ambiguity by first computing averaged, stacked spectra in bins corresponding to a range of spectral moments (i.e., long-period spectral amplitudes). They then solved for the correction term (i.e., the empirical Green's function or EGF) that minimized the root-mean-square (RMS)

misfit between the EGF-corrected stacked spectra, and a Brune-type (*Brune*, 1970) circular crack source model of the form

$$u(f) = \frac{\Omega_0}{1 + (f/f_c)^n}, \tag{5.4}$$

where $\Omega_0, f_c, n$ are the spectral moment (proportional to seismic moment, $M_0$), corner frequency, and high-frequency falloff rate, respectively. For simplicity, *Shearer et al.* (2006) fixed $n = 2$ (as in the canonical $\omega^{-2}$ model), and assumed that the stacked spectra were well described by a self-similar, constant stress drop model in which stress drop $\Delta\sigma$ is related to corner frequency and seismic moment by:

$$\Delta\sigma = \frac{7}{16} M_0 \left( \frac{f_c}{k\beta} \right)^3 \tag{5.5}$$

for shear wave speed $\beta$ and a constant $k$ (here 0.32 following *Madariaga* 1976, derived for a circular crack with constant stress drop, elliptical slip, and a constant rupture velocity $v_r = 0.9\beta$). *Shearer et al.* (2006) performed a grid search over trial values of stress drop $\Delta\sigma$, and determined the EGF correction from the average residual (across all stacks) between the observed and theoretical Brune-type spectra (equation 5.4), given $\Delta\sigma$ and the relative moment of each stack. They showed that the resulting EGF and self-similar Brune model produced a reasonable overall fit to the moment-binned spectral stacks averaged over their entire data set of ~60,000 earthquakes, and proceeded to show that median stress drops estimated for individual events using this EGF were nearly constant with moment. However, this approach may have biased the results toward self-similarity because it did not

test whether non-self-similar models could produce equal or even better fits to the data.

Here, we generalize the stacking approach *Shearer et al.* (2006) used to obtain the EGF, while relaxing the assumption of self-similar scaling between the stacked spectra. Following *Shearer et al.* (2006), we stack source spectra in bins of spectral moment, $\Omega_0$. Each of the stacked spectra thus represents the average relative source spectra of events with comparable seismic moment $M_0$ (which is proportional to $\Omega_0$). Note that we define $\Omega_0$ using the mean amplitude over the 2.5–4 Hz band, as the signal-to-noise worsens at lower frequencies.

Rather than assume a constant stress drop $\Delta\sigma$ for each stack, we allow for the possibility of a variation in mean stress drop with moment by comparing generalized linear models of the form $\log_{10}\Delta\sigma\left(\Omega_0|\boldsymbol{\epsilon}\right) = \epsilon_0 + \epsilon_1\,\Omega_0 + \epsilon_2\,\Omega_0^2 + ... + \epsilon_m\,\Omega_0^m$, where $\boldsymbol{\epsilon} = [\epsilon_0, \epsilon_1, \epsilon_2, ..., \epsilon_m]$ is a model vector whose parameters specify a degree-$m$ polynomial-type scaling of stress drop and moment (both measured in logarithmic units). The degree-0 model corresponds to a constant stress drop model that is functionally equivalent to the approach used by *Shearer et al.* (2006). The results presented in this study compare the self-similar (degree-0) parameterization to models with linear (degree-one) scaling (e.g., *Kanamori and Rivera* 2004; *Mayeda et al.* 2007; *Oth* 2013):

$$\log_{10}\Delta\sigma\left(\Omega_0|\boldsymbol{\epsilon}\right) = \epsilon_0 + \epsilon_1\,\Omega_0. \tag{5.6}$$

We also tested higher-order models, as well as those with unconstrained variation

of stress drop with moment, but found the marginal improvement in data misfit provided by such models to be insufficient to warrant the increase in model complexity (*Akaike*, 1974) and tendency for instability (symptomatic of overfitting noise in the stacked spectra). The basic steps in our new, iterative algorithm can be summarized as follows (see Figure 5.2 for a representative example using data from the Yuha Desert region):

1. Bin events by spectral moment $\Omega_0$ and compute stacked source spectra for each bin, requiring at least 20 events per bin to ensure the stacked spectra are well-resolved.

2. Initialize the parameter vector $\boldsymbol{\epsilon_i} = [0, 0]$ to a starting value implying constant stress drop for each bin (here the subscript $i$ denotes the current iteration of the optimization algorithm).

3. Compute theoretical source spectra for each stack, given its stress drop value defined by the current parameter vector $\boldsymbol{\epsilon_i}$ and mean spectral moment $\Omega_0$ (proportional to $M_0$).

4. Estimate the EGF correction from the residual between the observed and theoretical stacked spectra, at all frequency points, averaged across all stacks. Note that while each stack has its own unique stress drop value, the EGF correction is common to all of the stacked spectra.

5. Compute the weighted RMS misfit between the EGF-corrected, observed

spectra and the theoretical spectra in the frequency band from 2.5–25Hz. Here we apply weights inversely proportional to log-$f$ to prevent random fluctuations in the high-frequency portion of the spectra from dominating the fit.

6. Update $\boldsymbol{\epsilon_i} \rightarrow \boldsymbol{\epsilon_{i+1}}$ using a conjugate gradient algorithm, and repeat steps (3)–(6) until convergence to the minimum of the misfit function. Note that while 1st-order models can be solved in analogous ways using a grid search over the model parameters, the iterative conjugate gradient approach provides a unified computational framework for efficiently comparing higher-order or otherwise more complex models.

The final EGF obtained at the convergence point is used to correct the shape of the relative source spectra of each individual event for propagation effects that are common to all paths, making them directly comparable to the theoretical source model of interest (Figures 5.2 and 5.3). For the main results presented in this paper, we assume the widely-used *Brune* (1970) source model with $\omega^{-2}$ spectral falloff (equation 5.4 with $n = 2$), but discuss in detail the influence of this choice in Section 5.4.

The use of stacked spectra in the EGF estimation procedure has certain advantages, most notably that it relies solely on the relative shape of well-constrained, averaged spectra within the high signal-to-noise band, and not on resolving the corner frequencies of smaller individual events, some of which may be beyond the

usable signal bandwidth. However, it does neglect the lateral and depth-dependent variations in average attenuation properties that are not accounted for in the station terms, which may be important for studies over larger length scales than within the five regions we consider here. In these cases, nearest-neighbor or spatial interpolation techniques could be incorporated into the stacking procedure in order to include lateral and depth-dependent attenuation corrections into the spectral decomposition algorithm (e.g., *Shearer et al.* 2006).

### 5.2.4 Source parameter estimates: corner frequency, moment, and stress drop

With the EGF-corrected source spectra in hand, we are now able to estimate source parameters for each event. To compute stress drop $\Delta\sigma$ as defined by (5.5), we need estimates of corner frequency $f_c$, seismic moment $M_0$, and shear wave velocity $\beta$. We first estimate the corner frequency $f_c$ using a bounded optimization algorithm that minimizes the weighted RMS residual between the observed, EGF-corrected source spectra $s_i(f)$ and theoretical spectra $\hat{s}_i(f|f_c)$ in the 2.5–25Hz band (where again, inverse log-$f$ weights are applied to prevent the high-frequency portion of the spectra from dominating the fit).

Although some of the larger events ($M_L \geq 3.5$) have independent estimates of $M_0$ from regional analyses of long-period waveforms, the vast majority of the events we consider are listed by local magnitude $M_L$ rather than moment magni-

tude $M_W$. We therefore need a self-consistent way of computing $M_0$ for all events, large and small, as numerous studies (e.g., *Hutton et al.* 2010; *Ross et al.* 2016; *Munafo et al.* 2016) have demonstrated that $M_W \neq M_L$ for smaller events. We follow the same basic approach used by *Shearer et al.* (2006) and assume the $M_0$ is proportional to the spectral moment $\Omega_0$, but apply a correction term to the $\Omega_0$ estimates obtained from the 2.5–4Hz band (Section 5.2.3) to zero frequency using the observed corner frequency $f_c$ (this correction is more important for the larger events with $f_c < 4$Hz, where the spectral decay significantly reduces the observed 2.5–4Hz amplitude). We derive the proportionality constant between $M_0$ and $\Omega_0$ in two steps (*Shearer et al.*, 2006). We first perform regression analysis to calibrate a linear relationship between $M_L$ and $\Omega_0$: $\hat{M}_L = a_0 + a_1\,\Omega_0$. The amplitude-adjusted local magnitudes $\hat{M}_L$ are then converted to moment magnitudes $M_w$ and absolute moment $M_0$ by assuming that $M_W \approx \hat{M}_L$ at $\mathbf{M} = 3.50$ for earthquakes within southern California (*Hutton et al.*, 2010; *Ross et al.*, 2016). This regression analysis is performed separately for each region, as the regression coefficients depend slightly on local attenuation.

Finally, we use a smoothed version of the 1D velocity model of *Hadley and Kanamori* (1977) and the waveform-relocated event depths from the catalog of *Hauksson et al.* (2012) to compute depth-dependent shear wave velocities $\beta$ for all events, and use our estimates of $f_c$, $M_0$, and $\beta$ to compute stress drop $\Delta\sigma$ for each event (equation 5.5). We use the revised $P$-wave value of the constant $k = 0.38$ from the numerical study of *Kaneko and Shearer* (2014), though encourage

caution in interpreting the absolute values of the stress drop estimates, which can be highly model-dependent. More robust are the relative variations in stress drop estimated using a uniform and consistent set of modeling assumptions and processing procedures.

### 5.2.5 Source parameter uncertainties: corner frequency, moment, and stress drop

Each source spectra estimate is in some sense an average over all stations recording the event, and we utilize this fact in order to characterize the parameter uncertainties associated with our $f_c$ estimates. To do so, we first define the apparent source spectra $\tilde{s}_{ij}$ (e.g., *Pacor et al.* 2016b) for each station $j$ recording event $i$:

$$\tilde{s}_{ij} = d_{ij} - tt_{k(i,j)} - st_j - \bar{r}_{ij} - EGF_{corr}, \tag{5.7}$$

where $d$, $tt$, $st$, $EGF_{corr}$, and $\bar{r}$ are the data spectra, travel time spectra, and station spectra, EGF correction, and mean residual spectra, respectively. Note that $\bar{r}$ represents an average residual across all stations recording the event and is usually quite small, but may be nonzero because of the robust L1-norm weighting scheme in the inversion algorithm. The apparent spectra defined by equation (5.7) have the useful property that their mean value is equal to the inferred EGF-corrected source spectra: $\frac{1}{N}\sum_{j=1}^{N}\tilde{s}_{ij} = s_i$. This fact allows us to use a bootstrap resampling approach to assess the uncertainties with each corner frequency estimate as follows.

For each event, we first obtain a corner frequency estimate $\hat{f}_c$ using the EGF-corrected source spectra $s_i(f)$ as described in Section 5.2.4. We next synthesize a set of $B = 100$ bootstrap-resampled source spectra by resampling with replacement from the $N$ apparent spectra associated with the event and taking the mean. We then estimate the corner frequency of each of the $B$ resampled source spectra to obtain a bootstrap distribution of $\hat{f}_c$ estimates for each event, and use the bias-corrected, accelerated percentile technique (*Efron and Tibshirani*, 1994; *Carpenter and Bithell*, 2000) to derive confidence intervals for $\hat{f}_c$ at the 50% (i.e., the inter-quartile range) and 90% levels from the raw bootstrap distribution.

We use a similar approach to obtain uncertainties in $M_0$ by examining the variability in spectral moment $\tilde{\Omega}_{0(ij)}$ across all stations recording an event. In this case, the bootstrapping procedure is unnecessary, as the $\Omega_0$ estimate is a linear function of the apparent spectra (a mean over the 2.4–4Hz frequency band). We use the median absolute deviation in the apparent spectral moments of each event (i.e., the $\tilde{\Omega}_{0(ij)}$ of each event $i$) to quantify uncertainties in moment, which are typically of order 0.1 (in $\log_{10}$ units). Error estimates for $\hat{f}_c$ and $\hat{M}_0$ can then be used to derive confidence intervals for $\Delta\hat{\sigma}$ through propagation of errors (using equation 5.5), though we caution that these measures of uncertainty are likely lower bounds, as they do not account for uncertainty in rupture velocity, and also do not address any of the epistemic or modeling uncertainties associated with the assumed source model.

## 5.3 Results

We apply the spectral decomposition technique (Section 5.2) to each of the five regions shown in Figure 5.1, performing a separate inversion for each region. Below, we present the salient features of the source parameter estimates for each region individually, and compare the similarities and differences in the results across all regions. We focus in particular on the question of self-similarity (the scaling of $\Delta\sigma$ with $M_0$) and the observed spatiotemporal variations in stress drop on both regional and local scales.

### 5.3.1 Yuha Desert

The 2010 $M_W$ 7.2 El Mayor-Cucapah earthquake is the largest event within the footprint of the SCSN during our study period (January 2002 through September 2016). The complex, bilateral faulting of the El Mayor-Cucapah earthquake (*Wei et al.*, 2011; *Fletcher et al.*, 2014) triggered a dense cloud of aftershocks north of the USA-Mexico border in the Yuha Desert (*Kroll et al.*, 2013; *Hauksson et al.*, 2011). Our Yuha Desert study region is comprised primarily of these more northerly aftershocks, as the SCSN does not have an adequate azimuthal distribution of station coverage to make reliable spectral estimates for earthquakes to the south of the USA-Mexico border (Figure 5.1). The detailed relocation analysis of the first two months of aftershocks performed by *Kroll et al.* (2013) imaged the structural complexity of this region before the June 2010 $M_W$ 5.7 Ocotillo earth-

quake, which triggered a second wave of seismicity in mid-to-late 2010 (*Hauksson et al.*, 2011).

In applying the improved spectral decomposition technique (Section 5.2) to the Yuha Desert region, we obtain source parameter estimates for 3706 events that meet our minimum quality control criteria. Normalized uncertainties in corner frequency ($\Delta f_c / f_c$) tend to be higher for the lowest and highest magnitude events, which have corner frequencies at the limits of the signal bandwidth, but overall the source spectra are well-resolved. Corner frequency and stress drop show a moderate but consistent deviation from self-similarity, with median stress drop tending to increase as a function of seismic moment (Figure 5.4a). To quantify this more precisely, we define bins of width 0.4 in $\log_{10} M_0$ and compute the median stress drop for events in each bin. We then perform a weighted least-squares regression analysis on the binned data to determine the best-fitting scaling parameter $\epsilon_1$ in a linear relation of the form

$$\log_{10} \Delta\sigma = \epsilon_0 + \epsilon_1 \log_{10} M_0. \tag{5.8}$$

Here the weights account for both the number of events in each bin and the median uncertainty in stress drop, where the latter depends primarily upon the uncertainty in corner frequency and secondarily on the uncertainty in moment (equation 5.5). The scaling parameter $\epsilon_1 = 0.18$ ($\pm$ 0.039) inferred in this manner for the Yuha Desert is statistically positive, which is true for all five regions considered in this study (Figure 5.4). Median stress drop for this region is comparable to other

regions (2.29 MPa), but of lower variability (Figure 5.5a, left), with nearly all events having stress drops in the range 0.5 to 20 MPa. Seismicity rates in the Yuha Desert region during our study period are greatest in 2010, following the April 2010 $M_W$ 7.2 El Mayor-Cucapah and June 2010 $M_W$ 5.7 Ocotillo events, and the $\mathbf{M} \geq 3$ events of this time period tend to exhibit high stress drops (Figure 5.5a, right). However, we do not observe any systematic time-evolution in the stress drop, as events of comparable magnitude occurring at different times exhibit comparable stress drop. In other words, the conditional probability distribution of stress drop, given seismic moment, is stationary with time, but 2010 has an unusually large number of $\mathbf{M} \geq 3$ events due to its high seismicity rate.

We do observe significant spatial variations in median stress drop, with a tendency for events within the eastern portion of the Yuha Desert study region to have higher median stress drop than those in the the west (Figure 5.6a). To check whether this observation is truly due to differences in source properties, and not a result of spatial variation in attenuation, we implemented a spatial interpolation-based EGF technique analogous to the (nearest-neighbor) spatial EGF correction used by *Shearer et al.* (2006), finding the inferred spatial patterns in source parameters largely unchanged by the laterally-varying EGF. This is not surprising, given the relatively small length-scales (tens of kilometers) over which the region is defined. We further observe an increase in median corner frequency as a function of depth, larger than can be explained by the assumed increase in mid-crustal shear velocity (Figure 5.7a).

### 5.3.2 San Jacinto Fault: Trifurcation Zone

The San Jacinto Fault (SJF) is the most seismically actively fault system in California (*Sanders et al.*, 1986; *Kagan et al.*, 2006; *Wdowinski*, 2009), and, along with the southern portion of the San Andreas Fault located to its east, accommodates the dominant portion of the plate boundary motion in southern California (*Rockwell et al.*, 1990; *Fialko*, 2006; *Lindsey et al.*, 2014). While the Anza section of the central SJF is notable for its lack of microseismicity and major earthquake sequences during historical times (*Sanders and Kanamori*, 1984; *Zöller and Ben-Zion*, 2014; *Rockwell et al.*, 2015; *Jiang and Fialko*, 2016), the intersection of the Coyote Creek, Clark, and Buck Ridge faults to the southeast of the Anza gap is responsible for approximately ten percent of all earthquake production in southern California during our study period (2002–2016). This region of particularly dense seismicity has produced multiple distinct sequences of $\mathbf{M} \geq 4$ earthquakes distributed across the three subparallel faults that comprise the trifurcation zone (*Allam et al.*, 2014), including the recent (June 2016) $M_W$ 5.2 Borrego Springs event. Though the mainshocks within these sequences show predominantly strike-slip mechanisms, the individual faulting structures within the trifurcation zone are particularly complex, with high resolution tomography (*Allam et al.*, 2014; *Lin et al.*, 2016) and studies of fault zone trapped waves (*Li and Vernon*, 2001; *Lewis et al.*, 2005; *Ross and Ben-Zion*, 2015) indicating the prevalence of damage zones.

We use the spectral decomposition technique (Section 5.2) to obtain source

parameter estimates for 4564 earthquakes in the SJF trifurcation zone region. The median stress drop (2.27 MPa) is comparable to that of the Yuha region. However, the inferred stress drops for events within the SJF trifurcation zone exhibit greater heterogeneity, as evidenced by the significantly larger standard deviation in the stress drop distribution than is observed for Yuha (Figure 5.5b, left). This heterogeneity may reflect the tectonic complexity of this region, which is characterized by a hierarchical network of structures that accommodate a diverse set of faulting mechanisms and varying levels of recent seismic activity (*Sanders and Kanamori*, 1984; *Li and Vernon*, 2001; *Lewis et al.*, 2005).

Seismicity within the SJF trifurcation zone is also notable for its strong deviation from classical self-similarity, with median stress drop increasing by nearly a factor of ten across the magnitude range considered in our data set (Figure 5.4b). The scaling parameter $\epsilon_1 = 0.35$ ($\pm$ 0.064) inferred using the weighted regression procedure described in Section 5.3.1 is the highest of any of the five regions in our study, and would be on the higher end of previously reported values, which typically are in the range of 0.1–0.4 (*Mayeda and Walter*, 1996; *Izutani and Kanamori*, 2001; *Kanamori and Rivera*, 2004; *Takahashi et al.*, 2005; *Mayeda et al.*, 2005; *Venkataraman et al.*, 2006; *Pacor et al.*, 2016a). This trend in scaling is persistent over the duration of our study period (Figure 5.5b, right), and may in part be related to fact that the larger ($\mathbf{M} \geq 3.5$) events in the trifurcation zone tend to occur along the three major fault strands, while microseismicity preferentially occurs in the off-fault and intra-fault regions. A full exploration into the causative

mechanisms of this strong regional trend in scaling is however beyond the scope of this study.

Despite the lack of clear time dependence in scaling, there does appear to be some variability in the distribution of stress drops within individual earthquake sequences. Of particular interest is the June 2016 $M_W$ 5.2 Borrego Springs event, which is the most recent prominent mainshock in our data set. Both the Borrego Spring event and its aftershocks exhibit unusually high stress drop values compared to other events with equivalent moment (Figure 5.5b, right). The Borrego Springs event occurred on the Clark Fault, near the site of two $M_W \geq 4$ events that occurred in 2008 but were of much lower stress drop. Most of the Borrego Springs aftershocks occurred between the Clark and Buck Ridge faults, at depths of $\sim$ 12 km and on previously inactive structures that are almost orthogonal to the primary strands. These events are visually prominent in the map view shown in Figure 5.6b due to their high median stress drop.

### 5.3.3 Mojave Desert Regions: Big Bear, Landers and Hector Mine

In contrast to the Yuha Desert and San Jacinto Fault regions examined above, seismicity within the Big Bear, Landers, and Hector Mine regions of the Mojave Desert occurs to the east of the main Pacific-North American plate boundary. These regions comprise the southern portion of the Eastern California Shear

Zone (ECSZ), a diffuse deformational belt characterized by a complex distribution of incipient faulting structures (*VanWormer and Ryall*, 1980; *Faulds and Henry*, 2008; *Wesnousky et al.*, 2012). Studies of earthquake hazard in the ECSZ became more exigent with the occurrence of the 1992 Landers ($M_W$ 7.3), 1992 Big Bear ($M_W$ 6.5), and 1999 Hector Mine ($M_W$ 7.1) earthquakes, three of the largest events in southern California in the past century (*Hauksson et al.*, 1993; *Cohee and Beroza*, 1994; *Wald and Heaton*, 1994; *Fialko*, 2004). Stress changes from these events had a significant influence on the local stress field and seismicity during their immediate aftermath (e.g., *Hauksson* 1994; *King et al.* 1994; *Hardebeck et al.* 1998; *Lin and Stein* 2004). Our study period begins in 2002 and thus considers only the longer-term aftershocks of these events and present-day background seismicity within each region. In our study, we perform the spectral decomposition for each region independently in order to mitigate the effects of differences in local attenuation on source spectral estimates, but present the results for all three regions here, as they form a coherent tectonic zone within the southern ECSZ.

We obtain source parameter estimates for 1723, 1215, and 810 events in the Big Bear, Landers, and Hector Mine regions, respectively. Despite their proximity, we observe quantifiable differences in the source properties of earthquakes in each region. Earthquakes in the Landers rupture zone are characterized by significantly higher median stress drop (4.36 MPa) than those in the Hector Mine rupture zone (1.02 MPa), with Big Bear (2.68 MPa) falling in between (Figure 5.5cde). The distribution of stress drop is also more variable in the Landers and Big Bear

regions than in Hector Mine. Applying the binning procedure and the weighted regression analysis described in Section 5.3.1, we also observe a moderate increase in median stress drop with moment in each of the Big Bear ($\epsilon_1 = 0.17 \pm 0.027$), Landers ($\epsilon_1 = 0.28 \pm 0.064$) and Hector Mine ($\epsilon_1 = 0.13 \pm 0.066$) regions. We note however that the inferred scaling parameters ($\epsilon_1$) are more poorly constrained for these three regions due to the relative scarcity of $\mathbf{M} \geq 3$ events, as compared to the dense seismicity of the Yuha or SJF regions.

Although localized clusters of higher and lower median stress drops can be discerned in each of the three regions (Figure 5.8), the spatial variations are most apparent within the Landers rupture zone. Specifically, we observe significant along-strike variations in median stress drop, with higher values in the northernmost and southernmost segments, and lower values in the central portion of the Landers rupture (Figure 5.8b), a pattern similar to that found in the stress drop study of *Shearer et al.* (2006) for Landers aftershocks from 1992 through 2001. Overall, there appears to be a rough anticorrelation between the stress drops of Landers aftershocks (both early and late) and the magnitude of fault slip during the 1992 rupture, during which peak values of fault slip were observed along the central portion of the rupture and lower slip at the northern and southern tips and auxiliary fault segments (*Fialko*, 2004). The lack of significant pre-mainshock seismicity in this region makes it hard to assess whether this stress drop pattern is a long-standing feature (perhaps caused by local variations in fault strength, with the stronger zones both inhibiting slip during the Landers rupture and producing

ambient seismicity with higher stress drops), or whether it is reflects stress changes caused by the mainshock rupture, with lowered stress in the high-slip regions and increased stress near the fault tips. Unfortunately, the large uncertainties in finite slip models for the Landers mainshock (as evidenced by the substantial differences among the published models, e.g. *Mai and Thingbaijam* 2014), complicate more detailed comparisons between aftershock stress drops and mainshock slip at finer spatial scales.

Some evidence for long-lived spatial variations of small earthquake stress drops was provided by a stress drop study at Parkfield (*Allmann and Shearer*, 2007), in which the overall pattern of high and low stress drop regions was unaffected by the 2004 **M**6.0 Parkfield earthquake. Although the Parkfield section of the San Andreas Fault is distinctly different from our study regions (having a single well-defined fault, and many areas of fault creep), it is nonetheless interesting that the spatial patterns we observe in stress drop appear to persist over long time periods. For example, the spatial heterogeneity in source parameters observed within the Mojave Desert regions remains largely unchanged from the study of *Shearer et al.* (2006), who considered a completely independent source parameter data set ending in year 2001. The long-term persistence of these trends — lower stress drop in the Hector Mine rupture zone, higher stress drop in the Landers rupture zone, with significant along-strike variations — supports the notion that there exist real spatial variations in median stress drop that are both quantifiable and somewhat predictable due to their stationarity in time. If so, this has important

implications for hazard assessments that use seismically-determined stress drop or a related proxy stress parameter as input for ground motion prediction equations (*Yenier and Atkinson*, 2014; *Atkinson et al.*, 2015). Also of interest from a hazard perspective is that the shallowest seismicity (observed primarily in the Big Bear region) is characterized by particularly low stress drops (Figure 5.7c). At more intermediate depths, corner frequency increases with depth at roughly the same rate as the shear wave speed (and hence, mean rupture velocity), such that the inferred stress drop estimates remain approximately depth-invariant.

## 5.4 Discussion

The observed trends in earthquake source properties within the five regions have important implications for our understanding of earthquake rupture and seismic hazard. As such, it is worth examining in closer detail the limitations of the methods applied to obtain these source parameter estimates, and their potential influence on the results presented here. These limitations fall into two main classes: (1) limitations in the ability of the spectral decomposition method to isolate the source spectrum from the raw waveform data, and (2) limitations in the source spectral model to adequately characterize the salient properties of earthquake rupture.

Addressing (1) first, a key advantage of the spectral decomposition method is that it provides an empirical and completely nonparametric framework for par-

titioning the observed spectra into source, path, and station terms, with no need to explicitly evaluate the instrument response or model attenuation along the ray-path. This framework, while designed to minimize the tradeoffs between source and path effects, may still fail to adequately resolve the source term, especially in circumstances where the ratio of the number of input, observed spectra to desired, output source spectra is small. The EGF correction that is inferred from the shape of stacked source spectra (Section 5.2.3) is an essential step in the spectral decomposition algorithm, as it removes non-source effects that are common to all travel paths. This includes both near-source and near-receiver attenuation, which if left uncorrected, can introduce an artificial $f_{max}$ into the spectra that may bias source parameter estimates (*Hanks*, 1982; *Anderson and Hough*, 1984; *Anderson*, 1986). We do not believe that uncorrected high-frequency attenuation is causing problems in our analysis for several reasons: (1) we explicitly require that all of the individual spectra we analyze have signal-to-noise of five or greater out to the 25 Hz upper bandwidth, and (2) we see no evidence for any sharp change in our travel-time terms or in our EGFs as frequency approaches 25 Hz. Certainly attenuation will limit the ability to resolve corner frequencies above 25 Hz, but accurately determining these corner frequencies is not crucial for our analysis.

However, the spectral stacking procedure requires a reasonable distribution of events at different magnitudes over a relatively small length scale (tens of kilometers). This does not present a significant obstacle for the five regions considered here, which have relatively dense seismicity sampling a wide range of magnitudes

(**M** 1 to 5) and are well-recorded by modern broadband and short-period SCSN stations. It is important, however, to keep these considerations in mind in future applications of the spectral decomposition technique to scenarios with sparse seismicity or station coverage.

The second limitation — the possible inadequacy of the commonly-applied Brune-type, omega-squared ($\omega^{-2}$) source model to describe the earthquake rupture process — presents a more serious problem for the events considered here. The observed deviation from self-similarity, in which stress drop appears to increase as a function of seismic moment ($\epsilon_1 > 0$), is based on the assumption that the high-frequency spectral falloff rate ($n$ in equation 5.4) is 2, as in the classical $\omega^{-2}$ model (*Aki*, 1967). There is however a fundamental tradeoff between $\epsilon_1$ and $n$, as shown schematically in Figure 5.9 for the Yuha Desert region (the other four regions considered in this study exhibit a similar effect). If we instead allow for $n < 2$ (gentler spectral falloff), the data once again become consistent with self-similarity; the difference in misfit between the $\epsilon_1 > 0, n = 2$ and $\epsilon_1 = 0, n < 2$ models is only 1.2%. This makes intuitive sense if we consider that our fundamental observation is that the source spectra of larger events contain more high-frequency content than would be predicted by a self-similar Brune model. The tradeoff between $\epsilon_1$ and $n$ occurs because the high-frequency falloff rate substantially affects the results only well above the corner frequency, a portion of the spectra that is observable only for the larger events in our data set. There are two different end-member ways of explaining the high frequencies we observe for the larger events: we can

fix the spectral falloff and relax the assumption of self-similarity (Figure 5.9a), or we can assume that earthquakes are (on average) self-similar, and allow for gentler spectral falloff (Figure 5.9b).

We focus on the former approach for the purposes of this study because it makes our source parameter estimates (corner frequency $f_c$ and stress drop $\Delta\sigma$) more directly comparable to those of other studies, which are typically derived under the assumption of an $\omega^{-2}$ model. The discussion above, however, demonstrates that the absolute values of these source parameter estimates depend strongly on modeling assumptions. This makes it difficult to interpret $f_c$ and $\Delta\sigma$ directly in terms of actual physical properties of the earthquake source, such as the source dimension or average static stress drop. Moreover, the tradeoff complicates studies of earthquake source scaling, as the inferred scaling parameter will in turn depend on the assumed falloff rate (again, typically fixed to 2 in such studies). Though the high-frequency falloff $n$ must be greater than 1.5 to bound the radiated energy (*Walter and Brune*, 1993), this limitation only applies in the high-frequency limit, i.e., not to intermediate fall-off rates that may occur in double-corner-frequency models. For single corner frequency models, if the true falloff rate falls anywhere in the range $1.5 < n < 2$, the scaling will be less intense than that inferred with $n = 2$, as is typically assumed for such studies (*Kanamori and Rivera*, 2004). Finally, we note that the observed scaling relationships are pertinent only to the magnitude range that comprises the bulk of our data set ($1 \leq \mathbf{M} \leq 4$), and should not be extrapolated without further study. Independent of these issues, we can

still pose the following questions: what are the underlying causes of the observed deviations from classical source models, and what are the broader implications for our understanding of earthquake rupture processes?

*Brune* (1970) developed an instantaneous rupture model for a circular crack that related the source radius of an earthquake to the inverse of the corner frequency of the calculated source spectra. This model was later extended by *Madariaga* (1976) and others who considered more realistic circular crack models in which the rupture propagates at a constant fraction of the shear wave velocity. However, real source spectra, especially those of pulse-like ruptures, can be much more complex, and may be characterized by several timescales (and hence corner frequencies), notably those related to the total rupture duration and to the duration of slip at a given point on the fault (e.g., *Haskell* 1969; *Luco* 1985; *Lin et al.* 2016). Resolution of these features in real source spectra is challenging due to the inherent noise and bandwidth limitations of the data, but a number of studies of larger earthquakes (e.g., *Papageorgiou and Aki* 1983; *Joyner* 1984; *Atkinson* 1990; *Denolle and Shearer* 2016) have provided evidence for source spectra with multiple corners and intermediate spectral falloffs ($n < 2$), the details of which can likewise be scale-dependent (*Walter et al.*, 2006).

We do not attempt to resolve multiple corners within the source spectra in our study, as lower magnitude events have poor signal-to-noise within the higher-frequency bands in which the secondary corner would be expected to occur. However, the presence of multiple corners, even if they cannot be independently resolved

in the spectra, could account for the observed tradeoff between high-frequency falloff rate $n$ and self-similarity $\epsilon_1$. Further complexity in the rupture process, including directivity effects (e.g., *Boatwright* 1980; *Pacor et al.* 2016b; *Ross et al.* 2016), fault-roughness (e.g., *Madariaga et al.* 2006; *Dunham et al.* 2011; *Trugman and Dunham* 2014), and scale-dependent frictional properties or dynamic weakening mechanisms (*Brodsky and Kanamori*, 2001; *Abercrombie and Rice*, 2005; *Cocco et al.*, 2016) may also produce the same effect. Though it is difficult to discriminate among these plausible causal mechanisms, their unifying feature — source complexity beyond that of a circular crack — makes it worth exploring alternative or nonparametric means for directly comparing source spectra, rather than comparing source parameters derivative of an assumed model (e.g., *Uchide and Imanishi*, 2016).

The results presented here also bear significance for seismic hazard assessment. Ground motion prediction equations (GMPEs) are fundamental to hazard assessments, and source spectral studies that characterize regional variations in stress drop provide valuable constraints for the source terms of modern GMPEs, which are becoming increasingly regionalized (*Bozorgnia et al.*, 2014; *Yenier and Atkinson*, 2015; *Douglas and Edwards*, 2016). Previous work (e.g., *Allmann and Shearer* 2009; *Oth* 2013; *Uchide et al.* 2014) has provided compelling evidence for regional variations in median stress drop. Here we find in addition that both median stress drop and the intensity of the apparent deviation from self-similarity may vary regionally, with the San Jacinto Fault trifurcation zone being of par-

ticular note in this regard (Figure 5.4). Accurately characterizing such variations may help reconcile the well-known difference in scatter between the $\Delta\sigma$ of GMPEs (stress parameter) and the $\Delta\sigma$ of source spectral estimates (stress drop) (*Atkinson and Beresnev*, 1997; *Cotton et al.*, 2013; *Atkinson et al.*, 2015), although it is worth again cautioning that the scaling relationships inferred here likely do not extrapolate to the higher magnitude ($\mathbf{M} > 4$) events that cause the most damaging ground motions. Future studies should examine the robustness of these preliminary observations over wider magnitude ranges and larger spatial scales, and explore how the modeling limitations of source parameter estimates may influence the relationship between seismically-inferred stress drop and observed ground motion.

## 5.5 Conclusions

We describe an improved spectral decomposition approach to compute earthquake source parameters that is suitable for analyzing large data sets containing thousands or more events. The technique uses an iterative, robust least-squares algorithm to partition the observed waveform spectra into source, site, and travel-time dependent path terms. Unlike previous methods of its kind, this technique requires no assumption about self-similarity in earthquake source parameters. We leverage this improvement to explore variations in source parameters and source scaling within five regions of active, contemporary (2002–2016) seismicity in southern California. In each region, we find that if one assumes the classical,

$\omega^{-2}$ high-frequency falloff rate, the data are most consistent with an increase of stress drop with moment, in direct contrast to the self-similar paradigm. We show, however, that this conclusion is model-dependent, with the inferred deviation from self-similarity trading off with the assumed high-frequency falloff rate. Independent of these parametric limitations, the source spectra of the larger magnitude earthquakes in this study contain greater high-frequency content than would be predicted by a self-similar model with $\omega^{-2}$ falloff. This observation, along with the evidence we present for regional, local, and depth-dependent variations in earthquake source parameters, may provide important constraints for seismic hazard assessments and for our understanding of earthquake rupture processes.

# Acknowledgements

The waveform data used in this study is archived by and is publicly available from the Southern California Earthquake Data Center (http://scedc.caltech.edu/, last accessed 12/4/2016). The source parameter catalogs associated with this study are available in the electronic supplement to this manuscript (Supplementary Datasets D1 through D5). This manuscript is based upon work supported by the National Science Foundation Graduate Research Fellowship Program (NSFGRFP) under grant number DGE-1144086. Additional support was provided by the Southern California Earthquake Center (SCEC) under grant number 16020. We wish to thank Annemarie Baltay and Bill Walter for their thoughtful and con-

structive reviews, as well as Rachel Abercrombie, Zachary Ross, Egill Hauksson, Yuri Fialko, and Adrian Borsa for lively intellectual discussions that contributed to the content of this manuscript.

Chapter 5, in full, is a reformatted version of material as it appears in Journal of Geophysical – Solid Earth: Trugman, D. T. and P. M. Shearer (2017), Application of an improved spectral decomposition method to examine earthquake source scaling in southern California, *Journal of Geophysical Research: Solid Earth*, 122(4), doi:10.1002/2017JB013971. I was the primary investigator and author of this paper.

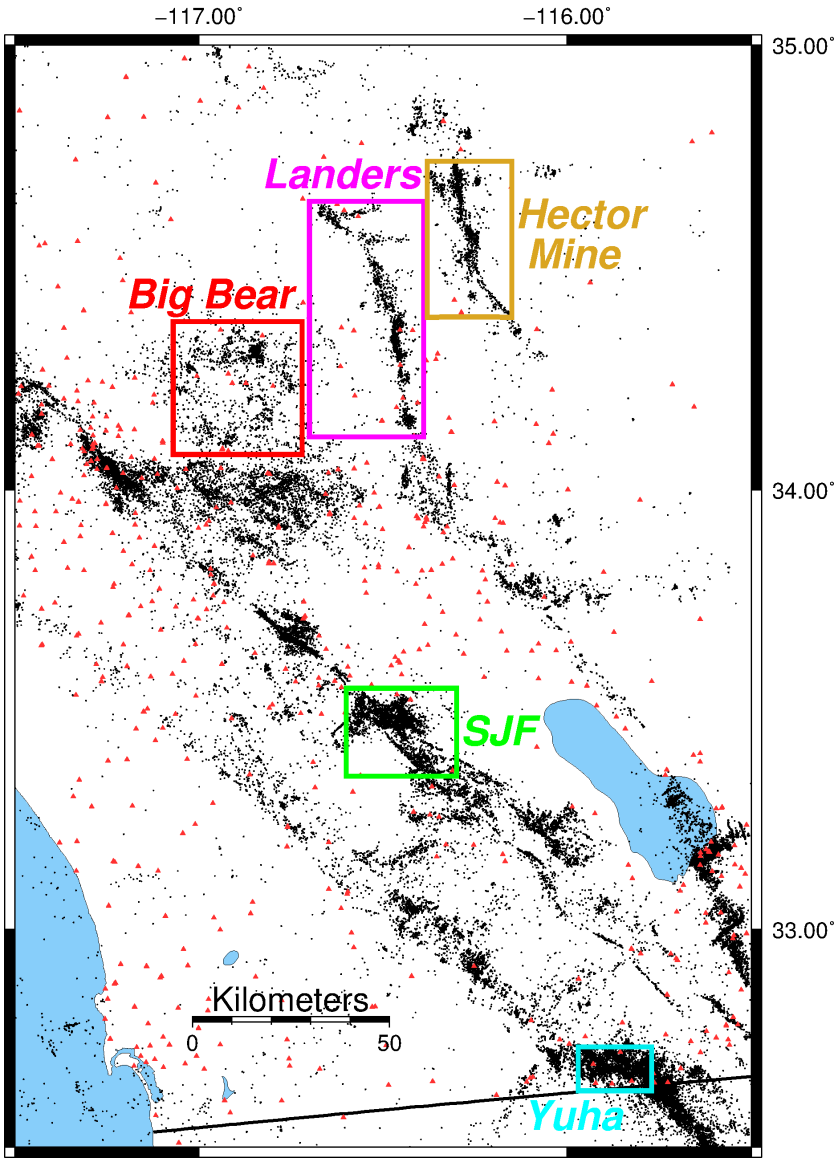


**Figure 5.1**: Map view of southern California seismicity, showing geographic bounds for the five study regions — Yuha Desert (Yuha), San Jacinto Fault Trifurcation Zone (SJF), Big Bear, Landers, and Hector Mine — considered in this study. Seismicity (black dots) is derived from the relocated catalog of (*Hauksson et al.*, 2012), with only $M_L \geq 1.1$ events occurring from January 2002 through September 2016 shown (consistent with the magnitude range and time period of this study). Vertical-component stations from the Southern California Seismic Network are marked (red triangles) for reference.

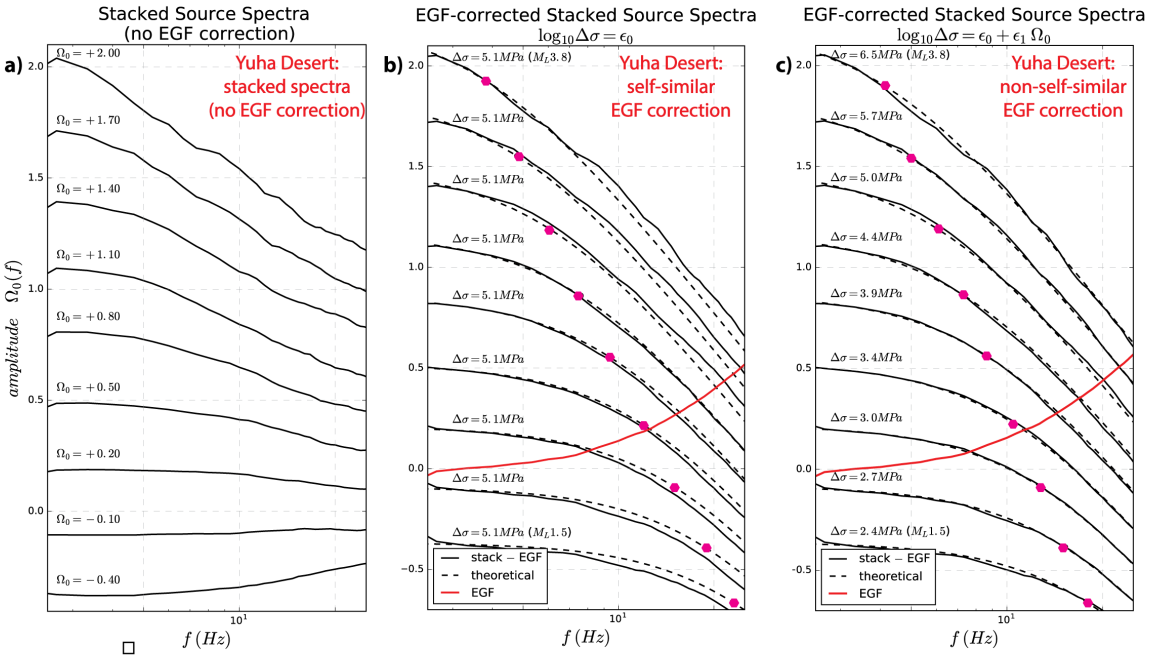


**Figure 5.2**: Example of stacked source spectra from the Yuha Desert region, before and after the EGF correction for common path effects (see text for details). (a) Stacked source spectra (binned by spectral moment $\Omega_0$), prior to the EGF correction. (b) EGF-corrected source spectra (solid black lines), assuming a self-similar, constant stress drop source model with $f^{-2}$ falloff at high frequencies. Theoretical source spectra corresponding to each stack are shown with dashed black lines, with the implied corner frequency of the fit marked for reference. The self-similar model fits poorly for both lower and higher values of $\Omega_0$. (c) EGF-corrected source spectra (solid black lines), assuming a source model in which stress drop increases linearly with moment (both measured in logarithmic units). Theoretical source spectra corresponding to each stack are shown with dashed black lines, with the implied corner frequency of the fit marked for reference. This model provides a significantly better fit than the self-similar model.

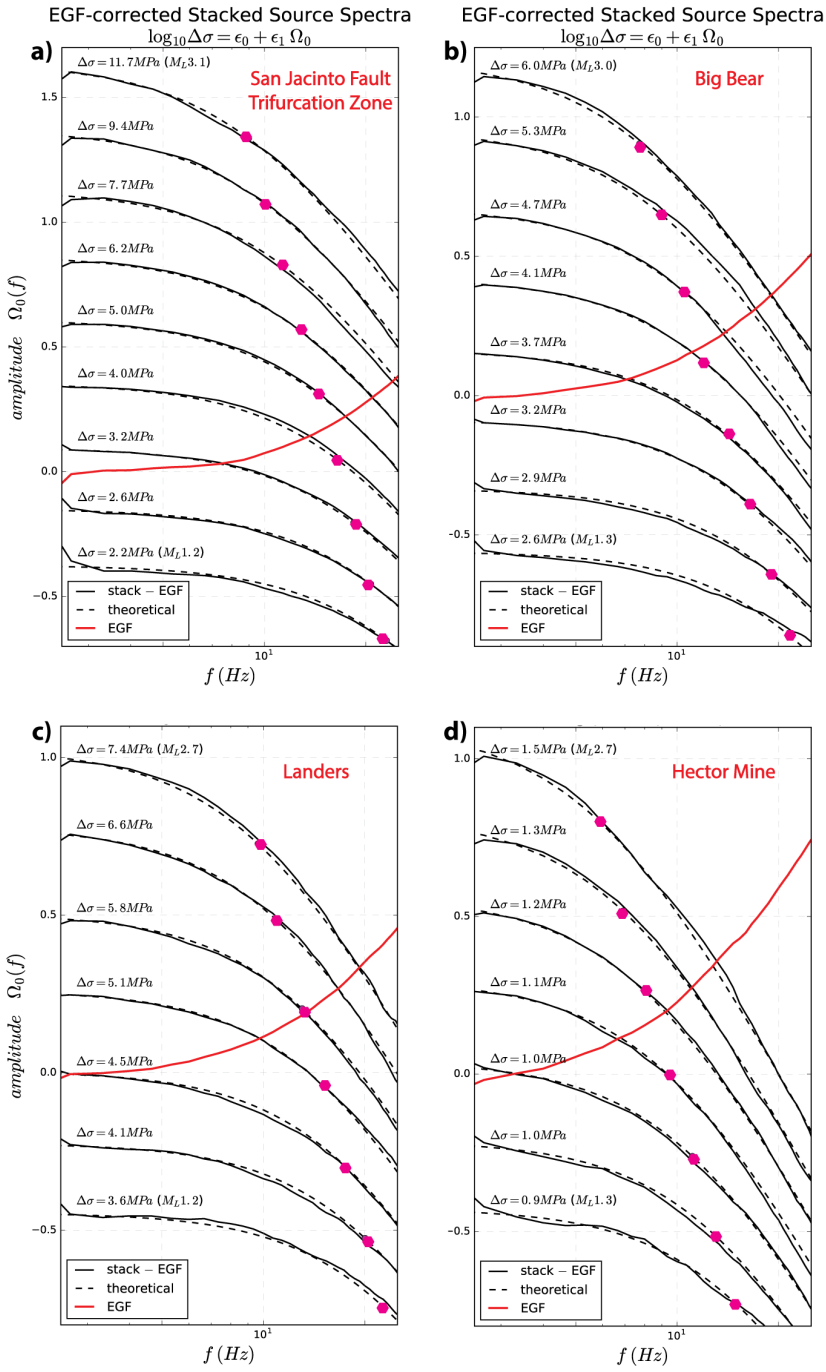


**Figure 5.3**: EGF-corrected, stacked source spectra from the (a) San Jacinto Fault Trifurcation Zone, (b) Big Bear, (c) Landers, and (d) Hector Mine regions. Each EGF-corrected source spectra (solid black lines, binned by spectral moment $\Omega_0$) assumes a source model in which stress drop increases with moment (as in Figure 5.2c for the Yuha region). Theoretical source spectra corresponding to each stack are shown with dashed black lines, with the implied corner frequency of the fit marked for reference.

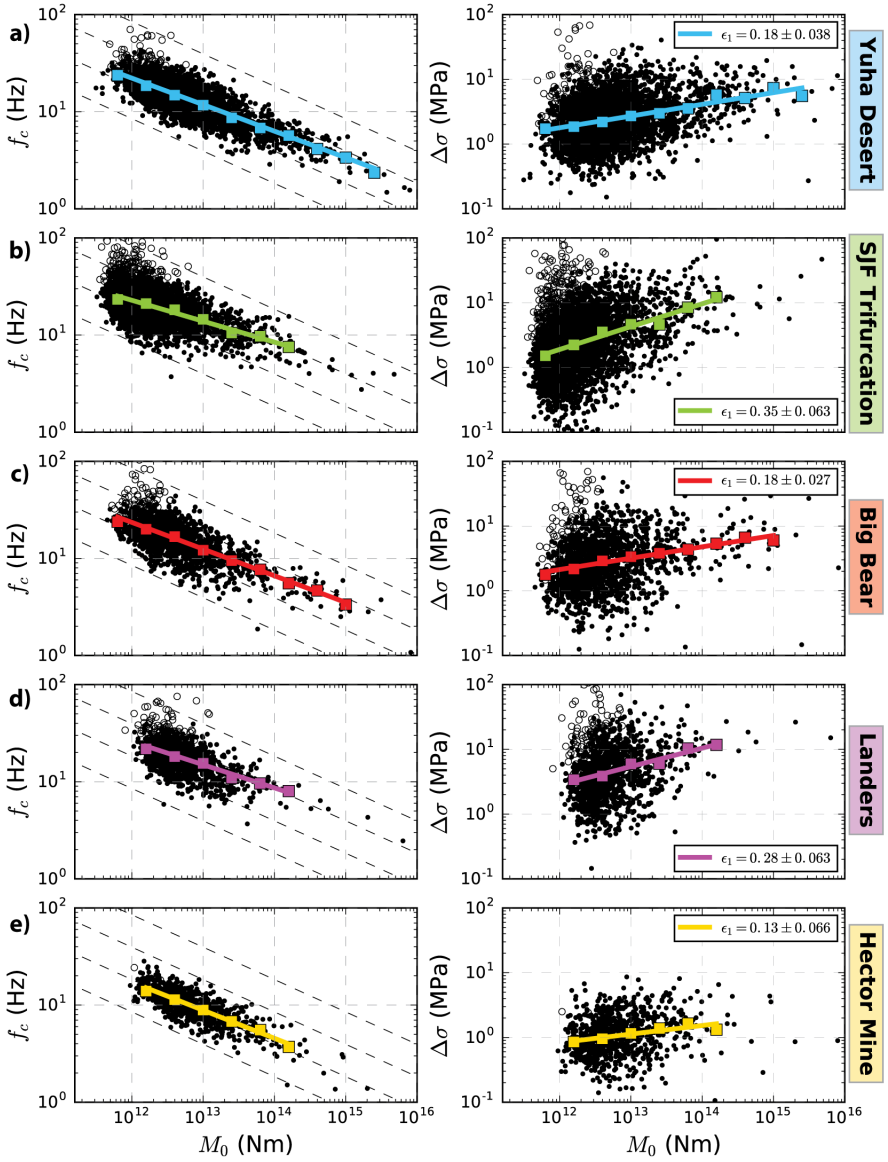


**Figure 5.4**: Scaling of corner frequency ($f_c$, left) and stress drop ($\Delta\sigma$, right) with seismic moment $M_0$. Each panel corresponds to one of the five study regions: (a) Yuha Desert, (b) San Jacinto Fault Trifurcation Zone, (c) Big Bear, (d) Landers, and (e) Hector Mine. In each panel, the black dots correspond to measurements of source properties for individual events, and the median $f_c$ and $\Delta\sigma$ in $M_0$ bins of 0.4 ($\log_{10}$ N-m units) are marked with square symbols. The best-fitting scaling parameter $\epsilon_1$ and two-sigma uncertainty for the binned data (obtained from weighted regression analysis, see text for details) is denoted in each right inset and plotted with a solid line. Events with poorly resolved corner frequencies due to bandwidth limitations are marked with open circles.

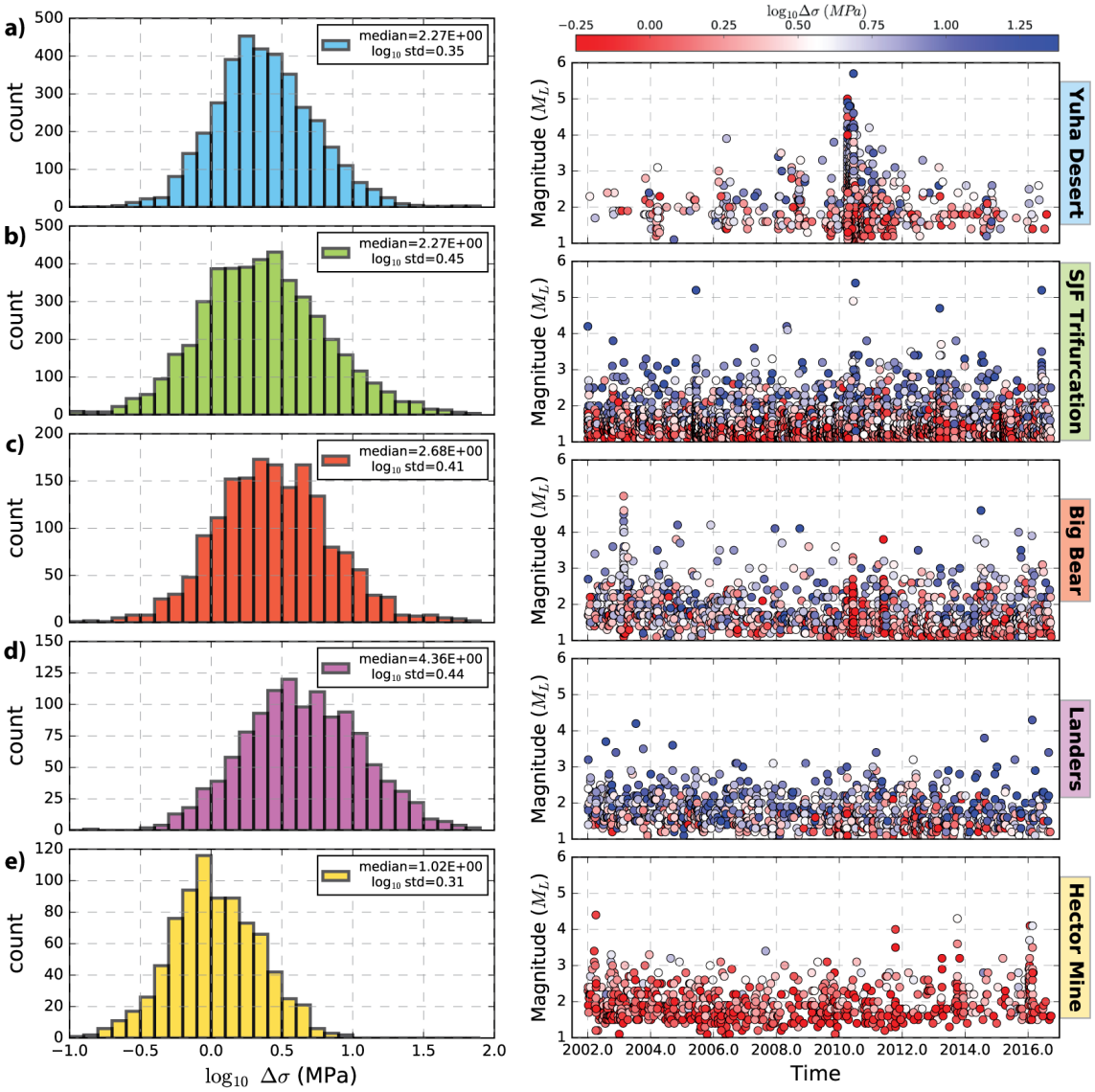


**Figure 5.5**: Histograms of stress drop (left) and magnitude-time plots (right) for each of the five study regions: (a) Yuha Desert, (b) San Jacinto Fault Trifurcation Zone, (c) Big Bear, (d) Landers, and (e) Hector Mine. Median and standard deviation values ($\log_{10}$ MPa) of the stress drop distributions are marked in the histogram insets (left). Events are color-coded by stress drop in each magnitude-time plot (right), with bluer colors indicating higher stress drop (more high-frequency energy).

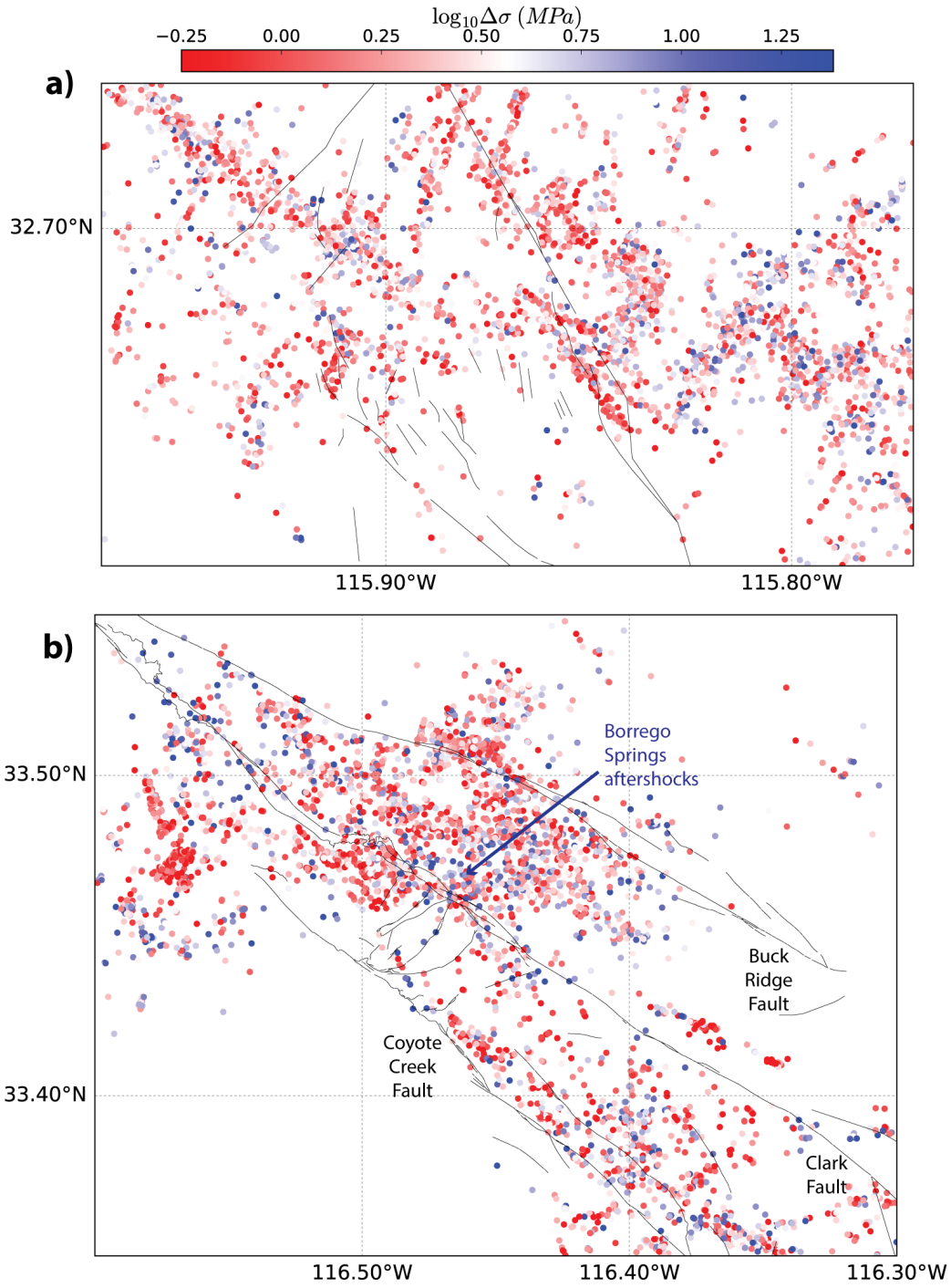


**Figure 5.6**: Map view of source parameter estimates for (a) Yuha Desert and (b) San Jacinto Fault Trifurcation Zone regions. Events are color-coded by stress drop in each map, with bluer colors indicating higher stress drop (more high-frequency energy).

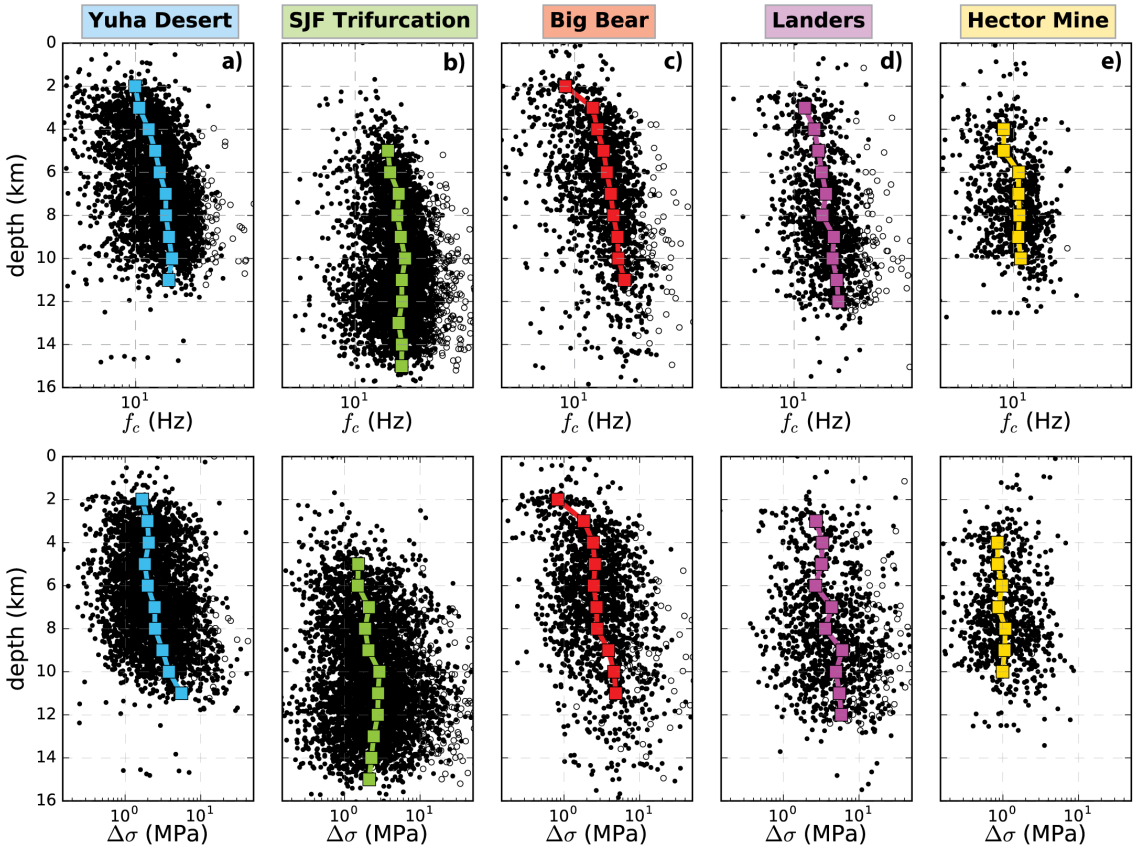


**Figure 5.7**: Depth-dependence of corner frequency ($f_c$, top) and stress drop ($\Delta\sigma$, bottom) for each of the five study regions: (a) Yuha Desert, (b) San Jacinto Fault Trifurcation Zone, (c) Big Bear, (d) Landers, and (e) Hector Mine. In each plot, the black dots correspond to measurements of source properties for individual events, and the median $f_c$ and $\Delta\sigma$ in depth bins of 1 km are marked with solid lines. Events with poorly resolved corner frequencies due to bandwidth limitations are marked with open circles.

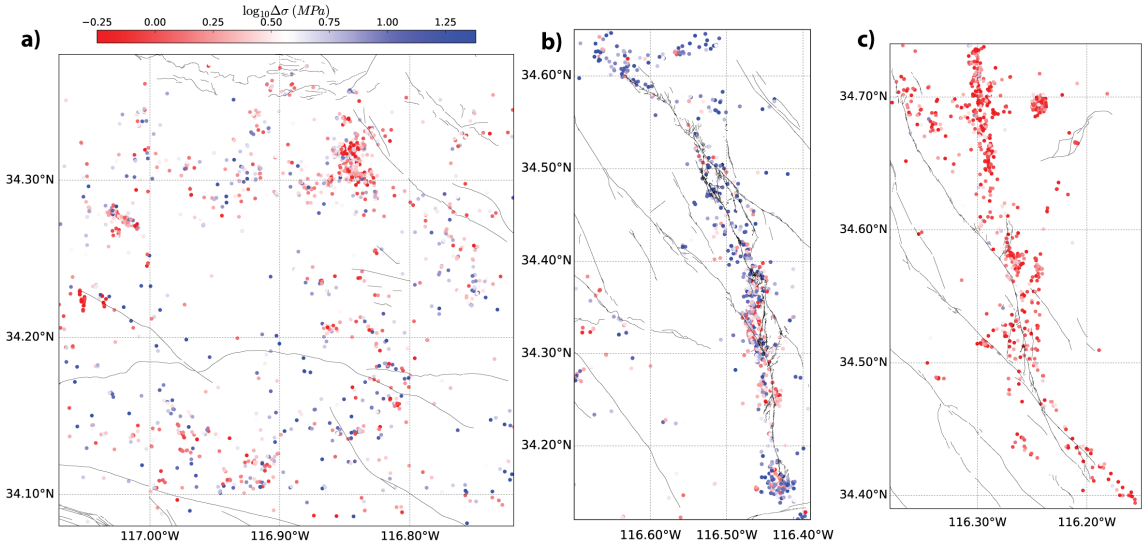


**Figure 5.8**: Map view of source parameter estimates for the (a) Big Bear, (b) Landers, and (c) Hector Mine regions. Events are color-coded by stress drop in each map, with bluer colors indicating higher stress drop (more high-frequency energy).

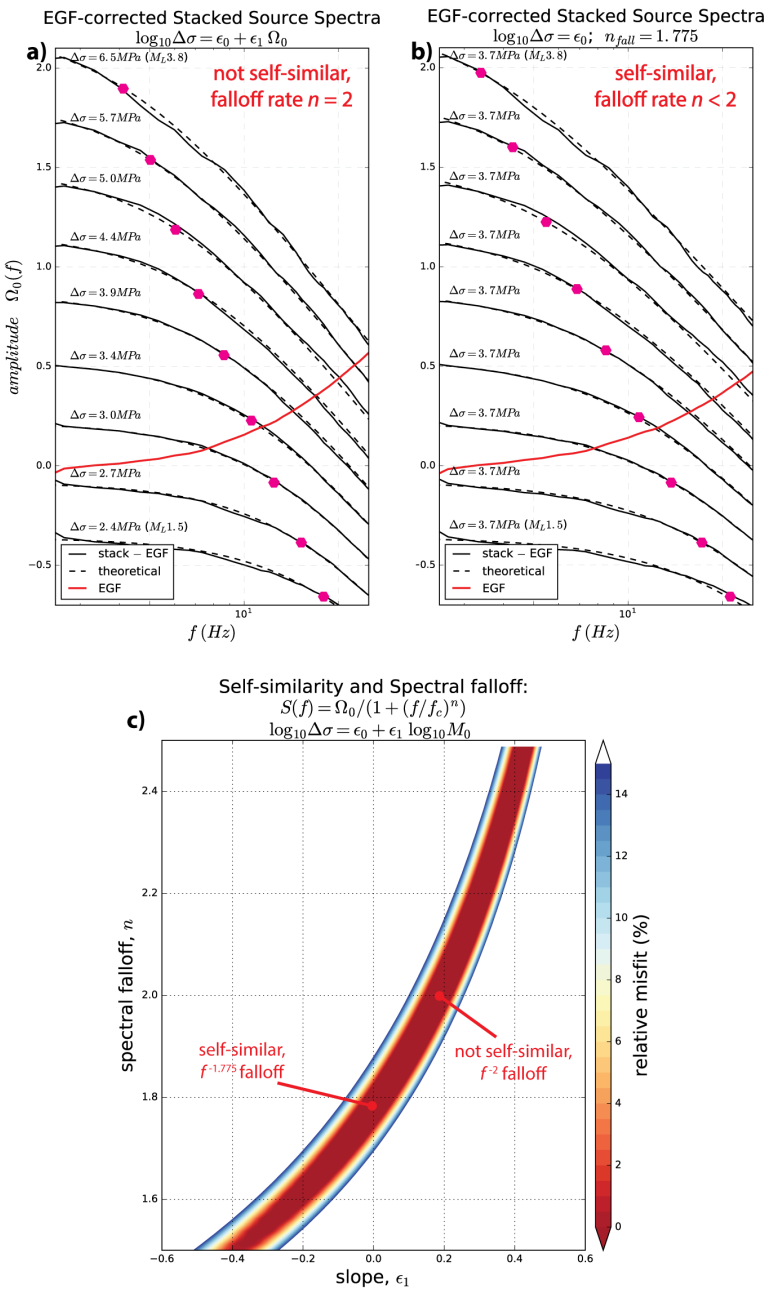


**Figure 5.9**: Example showing the tradeoff between scaling parameter $\epsilon_1$ and spectral falloff rate $n$. (a) EGF-corrected stacked source spectra (solid black lines) assuming a source model with $\omega^{-2}$ spectral falloff rate ($n = 2$) and scaling parameter $\epsilon_1 > 0$. $\log_{10}\Delta\sigma$ is thus permitted to vary linearly with $\log_{10} M_0$. (b) EGF-corrected source spectra (solid black lines), now assuming a lower spectral falloff rate ($n < 2$) and fixed scaling parameter $\epsilon_1 = 0$. $\Delta\sigma$ is thus constrained to be invariant with $M_0$. (c) Contour plot of the relative misfit between the EGF-corrected observed and theoretical spectra, plotted as a function of $\epsilon_1$ (x-axis) and $n$ (y-axis). Source models that assume $n = 2$ (an $\omega^{-2}$ model) require $\epsilon_1 > 0$, while source models that assume $\epsilon_1 = 0$ (a self-similar model) require $n < 2$.

# References

Abercrombie, R. E. (2013), Comparison of direct and coda wave stress drop measurements for the Wells, Nevada, earthquake sequence, *Journal of Geophysical Research: Solid Earth*, *118*(4), 1458–1470, doi:10.1029/2012JB009638.

Abercrombie, R. E., and J. R. Rice (2005), Can observations of earthquake scaling constrain slip weakening?, *Geophysical Journal International*, *162*(2), 406–424, doi:10.1111/j.1365-246X.2005.02579.x.

Abercrombie, R. E., S. Bannister, J. Ristau, and D. Doser (2017), Variability of earthquake stress drop in a subduction setting, the Hikurangi Margin, New Zealand, *Geophysical Journal International*, *208*(1), 306–320, doi:10.1093/gji/ggw393.

Akaike, H. (1974), A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, *19*(6), 716–723, doi:10.1109/TAC.1974.1100705.

Aki, K. (1967), Scaling law of seismic spectrum, *Journal of Geophysical Research*, *72*(4), 1217–1231, doi:10.1029/JZ072i004p01217.

Allam, A. A., Y. Ben-Zion, I. Kurzon, and F. Vernon (2014), Seismic velocity structure in the Hot Springs and Trifurcation areas of the San Jacinto fault zone, California, from double-difference tomography, *Geophysical Journal International*, *198*(2), 978–999, doi:10.1093/gji/ggu176.

Allmann, B. P., and P. M. Shearer (2007), Spatial and temporal stress drop variations in small earthquakes near Parkfield, California, *Journal of Geophysical Research: Solid Earth*, *112*(B4), B04,305, doi:10.1029/2006JB004395.

Allmann, B. P., and P. M. Shearer (2009), Global variations of stress drop for moderate to large earthquakes, *Journal of Geophysical Research: Solid Earth*, *114*(B1), B01,310, doi:10.1029/2008JB005821.

Anderson, J. G. (1986), Implication of Attenuation for Studies of the Earthquake Source, in *Earthquake Source Mechanics*, edited by S. Das, J. Boatwright, and C. H. Scholz, pp. 311–318, American Geophysical Union, dOI: 10.1029/GM037p0311.

Anderson, J. G., and S. E. Hough (1984), A model for the shape of the fourier amplitude spectrum of acceleration at high frequencies, *Bulletin of the Seismological Society of America*, *74*(5), 1969–1993.

Atkinson, G., K. Assatourians, B. Cheadle, and W. Greig (2015), Ground Motions from Three Recent Earthquakes in Western Alberta and Northeastern British Columbia and Their Implications for Induced-Seismicity Hazard

in Eastern Regions, *Seismological Research Letters*, *86*(3), 1022–1031, doi: 10.1785/0220140195.

Atkinson, G. M. (1990), A Comparison of Eastern North American ground Motion Observations with Theoretical Predictions, *Seismological Research Letters*, *61*(3-4), 171–180, doi:10.1785/gssrl.61.3-4.171.

Atkinson, G. M., and I. Beresnev (1997), Don't Call it Stress Drop, *Seismological Research Letters*, *68*(1), 3–4, doi:10.1785/gssrl.68.1.3.

Baltay, A., G. Prieto, and G. C. Beroza (2010), Radiated seismic energy from coda measurements and no scaling in apparent stress with seismic moment, *Journal of Geophysical Research*, *115*(B8), doi:10.1029/2009JB006736.

Baltay, A., S. Ide, G. Prieto, and G. Beroza (2011), Variability in earthquake stress drop and apparent stress, *Geophysical Research Letters*, *38*(L06303), doi: 10.1029/2011GL046698.

Boatwright, J. (1980), A spectral theory for circular seismic sources; simple estimates of source dimension, dynamic stress drop, and radiated seismic energy, *Bulletin of the Seismological Society of America*, *70*(1), 1–27.

Bozorgnia, Y., N. A. Abrahamson, L. A. Atik, T. D. Ancheta, G. M. Atkinson, J. W. Baker, A. Baltay, D. M. Boore, K. W. Campbell, B. S.-J. Chiou, R. Darragh, S. Day, J. Donahue, R. W. Graves, N. Gregor, T. Hanks, I. M. Idriss, R. Kamai, T. Kishida, A. Kottke, S. A. Mahin, S. Rezaeian, B. Rowshandel, E. Seyhan, S. Shahi, T. Shantz, W. Silva, P. Spudich, J. P. Stewart, J. Watson-Lamprey, K. Wooddell, and R. Youngs (2014), NGA-West2 Research Project, *Earthquake Spectra*, *30*(3), 973–987, doi:10.1193/072113EQS209M.

Brodsky, E. E., and H. Kanamori (2001), Elastohydrodynamic lubrication of faults, *Journal of Geophysical Research: Solid Earth*, *106*(B8), 16,357–16,374, doi:10.1029/2001JB000430.

Brune, J. N. (1970), Tectonic stress and the spectra of seismic shear waves from earthquakes, *Journal of Geophysical Research*, *75*(26), 4997–5009, doi:10.1029/JB075i026p04997.

Calderoni, G., A. Rovelli, and S. K. Singh (2013), Stress drop and source scaling of the 2009 April L'Aquila earthquakes, *Geophysical Journal International*, *192*(1), 260–274, doi:10.1093/gji/ggs011.

Carpenter, J., and J. Bithell (2000), Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians, *Statistics in Medicine*, *19*(9), 1141–1164, doi:10.1002/(SICI)1097-0258(20000515)19:9⟨1141::AID-SIM479⟩3.0.CO;2-F.

Castro, R. R., J. G. Anderson, and S. K. Singh (1990), Site response, attenuation and source spectra of S waves along the Guerrero, Mexico, subduction zone, *Bulletin of the Seismological Society of America*, *80*(6A), 1481–1503.

Choy, G. L., and J. L. Boatwright (1995), Global patterns of radiated seismic energy and apparent stress, *Journal of Geophysical Research: Solid Earth*, *100*(B9), 18,205–18,228, doi:10.1029/95JB01969.

Cocco, M., E. Tinti, and A. Cirella (2016), On the scale dependence of earthquake stress drop, *Journal of Seismology*, pp. 1–20, doi:10.1007/s10950-016-9594-4.

Cohee, B. P., and G. C. Beroza (1994), Slip distribution of the 1992 Landers earthquake and its implications for earthquake source mechanics, *Bulletin of the Seismological Society of America*, *84*(3), 692–712.

Cotton, F., R. Archuleta, and M. Causse (2013), What is Sigma of the Stress Drop?, *Seismological Research Letters*, *84*(1), 42–48, doi:10.1785/0220120087.

Denolle, M. A., and P. M. Shearer (2016), New perspectives on self-similarity for shallow thrust earthquakes, *Journal of Geophysical Research: Solid Earth*, p. 2016JB013105, doi:10.1002/2016JB013105.

Douglas, J., and B. Edwards (2016), Recent and future developments in earthquake ground motion estimation, *Earth-Science Reviews*, *160*, 203–219, doi:10.1016/j.earscirev.2016.07.005.

Dunham, E. M., D. Belanger, L. Cong, and J. E. Kozdon (2011), Earthquake Ruptures with Strongly Rate-Weakening Friction and Off-Fault Plasticity, Part 2: Nonplanar Faults, *Bulletin of the Seismological Society of America*, *101*(5), 2308–2322, doi:10.1785/0120100076.

Efron, B., and R. J. Tibshirani (1994), *An Introduction to the Bootstrap*, CRC press, Boca Raton, FL.

Faulds, J. E., and C. D. Henry (2008), Tectonic influences on the spatial and temporal evolution of the Walker Lane: An incipient transform fault along the evolving Pacific - North American plate boundary, *Arizona Geological Society Digest*, *22*, 437–470.

Fialko, Y. (2004), Probing the mechanical properties of seismically active crust with space geodesy: Study of the coseismic deformation due to the 1992 $M_w$ 7.3 Landers (southern California) earthquake, *Journal of Geophysical Research*, *109*(B3), doi:10.1029/2003JB002756.

Fialko, Y. (2006), Interseismic strain accumulation and the earthquake potential on the southern San Andreas fault system, *Nature*, *441*(7096), 968–971, doi:10.1038/nature04797.

Field, E. H., R. J. Arrowsmith, G. P. Biasi, P. Bird, T. E. Dawson, K. R. Felzer, D. D. Jackson, K. M. Johnson, T. H. Jordan, C. Madden, A. J. Michael, K. R. Milner, M. T. Page, T. Parsons, P. M. Powers, B. E. Shaw, W. R. Thatcher, R. J. Weldon, and Y. Zeng (2014), Uniform California Earthquake Rupture Forecast, Version 3 (UCERF3)–The Time-Independent Model, *Bulletin of the Seismological Society of America*, *104*(3), 1122–1180, doi:10.1785/0120130164.

Fletcher, J. M., O. J. Teran, T. K. Rockwell, M. E. Oskin, K. W. Hudnut, K. J. Mueller, R. M. Spelz, S. O. Akciz, E. Masana, G. Faneros, E. J. Fielding, S. Leprince, A. E. Morelan, J. Stock, D. K. Lynch, A. J. Elliott, P. Gold, J. Liu-Zeng, A. González-Ortega, A. Hinojosa-Corona, and J. González-García (2014), Assembly of a large earthquake from a complex fault system: Surface rupture kinematics of the 4 April 2010 El MayorCucapah (Mexico) Mw 7.2 earthquake, *Geosphere*, *10*(4), 797–827, doi:10.1130/GES00933.1.

Hadley, D., and H. Kanamori (1977), Seismic structure of the Transverse Ranges, California, *Geological Society of America Bulletin*, *88*(10), 1469–1478.

Hanks, T. C. (1982), fmax, *Bulletin of the Seismological Society of America*, *72*(6A), 1867–1879.

Hardebeck, J. L., J. J. Nazareth, and E. Hauksson (1998), The static stress change triggering model: Constraints from two southern California aftershock sequences, *Journal of Geophysical Research*, *103*(B10), 24,427, doi:10.1029/98JB00573.

Haskell, N. A. (1969), Elastic displacements in the near-field of a propagating fault, *Bulletin of the Seismological Society of America*, *59*(2), 865–908.

Hauksson, E. (1994), State of stress from focal mechanisms before and after the 1992 landers earthquake sequence, *Bulletin of the Seismological Society of America*, *84*(3), 917–934.

Hauksson, E., L. M. Jones, K. Hutton, and D. Eberhart-Phillips (1993), The 1992 Landers Earthquake Sequence: Seismological observations, *Journal of Geophysical Research*, *98*(B11), 19,835, doi:10.1029/93JB02384.

Hauksson, E., J. Stock, K. Hutton, W. Yang, J. A. Vidal-Villegas, and H. Kanamori (2011), The 2010 M w 7.2 El Mayor-Cucapah Earthquake Sequence, Baja California, Mexico and Southernmost California, USA: Active Seismotectonics along the Mexican Pacific Margin, *Pure and Applied Geophysics*, *168*(8-9), 1255–1277, doi:10.1007/s00024-010-0209-7.

Hauksson, E., W. Yang, and P. M. Shearer (2012), Waveform Relocated Earthquake Catalog for Southern California (1981 to June 2011), *Bulletin of the Seismological Society of America*, *102*(5), 2239–2244, doi:10.1785/0120120010.

Hough, S. E. (1996), Observational constraints on earthquake source scaling: understanding the limits in resolution, *Tectonophysics*, *261*(1), 83–95, doi: 10.1016/0040-1951(96)00058-3.

Hough, S. E. (1997), Empirical Green's function analysis: Taking the next step, *Journal of Geophysical Research: Solid Earth*, *102*(B3), 5369–5384, doi:10.1029/96JB03488.

Huang, Y., G. C. Beroza, and W. L. Ellsworth (2016), Stress drop estimates of potentially induced earthquakes in the Guy-Greenbrier sequence, *Journal of Geophysical Research: Solid Earth*, *121*(9), 6597–6607, doi:10.1002/2016JB013067.

Hutton, K., J. Woessner, and E. Hauksson (2010), Earthquake Monitoring in Southern California for Seventy-Seven Years (1932-2008), *Bulletin of the Seismological Society of America*, *100*(2), 423–446, doi:10.1785/0120090130.

Ide, S. (2003), Apparent break in earthquake scaling due to path and site effects on deep borehole recordings, *Journal of Geophysical Research*, *108*(B5), doi: 10.1029/2001JB001617.

Ide, S., and G. C. Beroza (2001), Does apparent stress vary with earthquake size?, *Geophysical Research Letters*, *28*(17), 3349–3352, doi:10.1029/2001GL013106.

Izutani, Y., and H. Kanamori (2001), Scale-dependence of seismic energy-to-moment ratio for strike-slip earthquakes in Japan, *Geophysical Research Letters*, *28*(20), 4007–4010, doi:10.1029/2001GL013402.

Jiang, J., and Y. Fialko (2016), Reconciling seismicity and geodetic locking depths on the Anza section of the San Jacinto fault, *Geophysical Research Letters*, *43*, 10,663–10,671, doi:10.1002/2016GL071113.

Joyner, W. B. (1984), A scaling law for the spectra of large earthquakes, *Bulletin of the Seismological Society of America*, *74*(4), 1167–1188.

Kagan, Y. Y., D. D. Jackson, and Y. Rong (2006), A New Catalog of Southern California Earthquakes, 18002005, *Seismological Research Letters*, *77*(1), 30–38, doi:10.1785/gssrl.77.1.30.

Kanamori, H., and L. Rivera (2004), Static and Dynamic Scaling Relations for Earthquakes and Their Implications for Rupture Speed and Stress Drop, *Bulletin of the Seismological Society of America*, *94*(1), 314–319, doi:10.1785/0120030159.

Kane, D. L., G. A. Prieto, F. L. Vernon, and P. M. Shearer (2011), Quantifying Seismic Source Parameter Uncertainties, *Bulletin of the Seismological Society of America*, *101*(2), 535–543, doi:10.1785/0120100166.

Kaneko, Y., and P. M. Shearer (2014), Seismic source spectra and estimated stress drop derived from cohesive-zone models of circular subshear rupture, *Geophysical Journal International*, *197*(2), 1002–1015, doi:10.1093/gji/ggu030.

King, G. C. P., R. S. Stein, and J. Lin (1994), Static stress changes and the triggering of earthquakes, *Bulletin of the Seismological Society of America*, *84*(3), 935–953.

Kroll, K. A., E. S. Cochran, K. B. Richards-Dinger, and D. F. Sumy (2013), Aftershocks of the 2010 Mw 7.2 El Mayor-Cucapah earthquake reveal complex faulting in the Yuha Desert, California, *Journal of Geophysical Research: Solid Earth*, *118*(12), 6146–6164, doi:10.1002/2013JB010529.

Lewis, M. A., Z. Peng, Y. Ben-Zion, and F. L. Vernon (2005), Shallow seismic trapping structure in the San Jacinto fault zone near Anza, California, *Geophysical Journal International*, *162*(3), 867–881, doi:10.1111/j.1365-246X.2005.02684.x.

Li, Y.-G., and F. L. Vernon (2001), Characterization of the San Jacinto fault zone near Anza, California, by fault zone trapped waves, *Journal of Geophysical Research: Solid Earth*, *106*(B12), 30,671–30,688, doi:10.1029/2000JB000107.

Lin, J., and R. S. Stein (2004), Stress triggering in thrust and subduction earthquakes and stress interaction between the southern San Andreas and nearby thrust and strike-slip faults, *Journal of Geophysical Research: Solid Earth*, *109*(B2), doi:10.1029/2003JB002607.

Lin, Y.-Y., K.-F. Ma, H. Kanamori, T.-R. A. Song, N. Lapusta, and V. C. Tsai (2016), Evidence for non-self-similarity of microearthquakes recorded at a Taiwan borehole seismometer array, *Geophysical Journal International*, *206*(2), 757–773, doi:10.1093/gji/ggw172.

Lindsey, E. O., V. J. Sahakian, Y. Fialko, Y. Bock, S. Barbot, and T. K. Rockwell (2014), Interseismic Strain Localization in the San Jacinto Fault Zone, *Pure and Applied Geophysics*, *171*(11), 2937–2954, doi:10.1007/s00024-013-0753-z.

Luco, J. E. (1985), On strong ground motion estimates based on models of the radiated spectrum, *Bulletin of the Seismological Society of America*, *75*(3), 641–649.

Madariaga, R. (1976), Dynamics of an expanding circular fault, *Bulletin of the Seismological Society of America*, *66*(3), 639–666.

Madariaga, R., J. P. Ampuero, and M. Adda-Bedia (2006), Seismic Radiation from Simple Models of Earthquakes, in *Earthquakes: Radiated Energy and the Physics of Faulting*, edited by R. Abercrombie, A. McGarr, G. D. Toro, and H. Kanamori, pp. 223–236, American Geophysical Union.

Mai, P. M., and K. K. S. Thingbaijam (2014), SRCMOD: An Online Database of Finite-Fault Rupture Models, *Seismological Research Letters*, *85*(6), 1348–1357, doi:10.1785/0220140077.

Mayeda, K., and W. R. Walter (1996), Moment, energy, stress drop, and source spectra of western United States earthquakes from regional coda envelopes, *Journal of Geophysical Research: Solid Earth*, *101*(B5), 11,195–11,208, doi:10.1029/96JB00112.

Mayeda, K., R. Gok, W. R. Walter, and A. Hofstetter (2005), Evidence for non-constant energy/moment scaling from coda-derived source spectra, *Geophysical Research Letters*, *32*(10), L10,306, doi:10.1029/2005GL022405.

Mayeda, K., L. Malagnini, and W. R. Walter (2007), A new spectral ratio method using narrow band coda envelopes: Evidence for non-self-similarity in the Hector Mine sequence, *Geophysical Research Letters*, *34*(11), L11,303, doi:10.1029/2007GL030041.

Mori, J., and A. Frankel (1990), Source parameters for small events associated with the 1986 North Palm Springs, California, earthquake determined using empirical Green functions, *Bulletin of the Seismological Society of America*, *80*(2), 278–295.

Mori, J., R. E. Abercrombie, and H. Kanamori (2003), Stress drops and radiated energies of aftershocks of the 1994 Northridge, California, earthquake, *Journal of Geophysical Research: Solid Earth*, *108*(B11), 2545, doi:10.1029/2001JB000474.

Munafo, I., L. Malagnini, and L. Chiaraluce (2016), On the Relationship between Mw and ML for Small Earthquakes, *Bulletin of the Seismological Society of America*, *106*(5), 2402–2408, doi:10.1785/0120160130.

Oth, A. (2013), On the characteristics of earthquake stress release variations in Japan, *Earth and Planetary Science Letters*, *377-378*, 132–141, doi:10.1016/j.epsl.2013.06.037.

Oth, A., D. Bindi, S. Parolai, and D. Di Giacomo (2011), Spectral Analysis of K-NET and KiK-net Data in Japan, Part II: On Attenuation Characteristics, Source Spectra, and Site Response of Borehole and Surface Stations, *Bulletin of the Seismological Society of America*, *101*(2), 667–687, doi:10.1785/0120100135.

Pacor, F., D. Spallarossa, A. Oth, L. Luzi, R. Puglia, L. Cantore, A. Mercuri, M. D'Amico, and D. Bindi (2016a), Spectral models for ground motion prediction in the L'Aquila region (central Italy): evidence for stress-drop dependence on magnitude and depth, *Geophysical Journal International*, *204*(2), 697–718, doi:10.1093/gji/ggv448.

Pacor, F., F. Gallovic, R. Puglia, L. Luzi, and M. D'Amico (2016b), Diminishing high-frequency directivity due to a source effect: Empirical evidence from small earthquakes in the Abruzzo region, Italy, *Geophysical Research Letters*, *43*(10), 5000–5008, doi:10.1002/2016GL068546.

Papageorgiou, A. S., and K. Aki (1983), A specific barrier model for the quantitative description of inhomogeneous faulting and the prediction of strong ground motion. Part II. Applications of the model, *Bulletin of the Seismological Society of America*, *73*(4), 953–978.

Park, J., C. R. Lindberg, and F. L. Vernon (1987), Multitaper spectral analysis of high-frequency seismograms, *Journal of Geophysical Research*, *92*(B12), 12,675, doi:10.1029/JB092iB12p12675.

Petersen, M. D., C. S. Mueller, M. P. Moschetti, S. M. Hoover, A. L. Llenos, W. L. Ellsworth, A. J. Michael, J. L. Rubinstein, A. F. McGarr, and K. S. Rukstales (2016), Seismic-Hazard Forecast for 2016 Including Induced and Natural Earthquakes in the Central and Eastern United States, *Seismological Research Letters*, *87*(6), 1327–1341, doi:10.1785/0220160072.

Poli, P., and G. A. Prieto (2016), Global rupture parameters for deep and intermediate-depth earthquakes, *Journal of Geophysical Research: Solid Earth*, *121*(12), 8871–8887, doi:10.1002/2016JB013521.

Prieto, G., R. Parker, and F. Vernon III (2009), A Fortran 90 library for multitaper spectrum analysis, *Computers & Geosciences*, *35*(8), 1701–1710, doi:10.1016/j.cageo.2008.06.007.

Prieto, G. A. (2004), Earthquake source scaling and self-similarity estimation from stacking P and S spectra, *Journal of Geophysical Research*, *109*(B8), doi:10.1029/2004JB003084.

Prieto, G. A., R. L. Parker, F. L. Vernon, P. M. Shearer, and D. J. Thomson (2006), Uncertainties in earthquake source spectrum estimation using empirical Green functions, in *Geophysical Monograph Series*, vol. 170, edited by R. Abercrombie, A. McGarr, H. Kanamori, and G. Di Toro, pp. 69–74, American Geophysical Union, Washington, D. C.

Rockwell, T., C. Loughman, and P. Merifield (1990), Late Quaternary rate of slip along the San Jacinto Fault Zone near Anza, southern California, *Journal of Geophysical Research: Solid Earth*, *95*(B6), 8593–8605, doi:10.1029/JB095iB06p08593.

Rockwell, T. K., T. E. Dawson, J. Y. Ben-Horin, and G. Seitz (2015), A 21-Event, 4,000-Year History of Surface Ruptures in the Anza Seismic Gap, San Jacinto Fault, and Implications for Long-term Earthquake Production on a Major Plate

Boundary Fault, *Pure and Applied Geophysics*, *172*(5), 1143–1165, doi:10.1007/s00024-014-0955-z.

Ross, Z. E., and Y. Ben-Zion (2015), An algorithm for automated identification of fault zone trapped waves, *Geophysical Journal International*, *202*(2), 933–942, doi:10.1093/gji/ggv197.

Ross, Z. E., and Y. Ben-Zion (2016), Toward reliable automated estimates of earthquake source properties from body wave spectra, *Journal of Geophysical Research: Solid Earth*, *121*(6), 2016JB013,003, doi:10.1002/2016JB013003.

Ross, Z. E., Y. Ben-Zion, M. C. White, and F. L. Vernon (2016), Analysis of earthquake body wave spectra for potency and magnitude values: implications for magnitude scaling relations, *Geophysical Journal International*, *207*(2), 1158–1164, doi:10.1093/gji/ggw327.

Sanders, C., H. Magistrale, and H. Kanamori (1986), Rupture patterns and preshocks of large earthquakes in the southern San Jacinto fault zone, *Bulletin of the Seismological Society of America*, *76*(5), 1187–1206.

Sanders, C. O., and H. Kanamori (1984), A seismotectonic analysis of the Anza Seismic Gap, San Jacinto Fault Zone, southern California, *Journal of Geophysical Research: Solid Earth*, *89*(B7), 5873–5890, doi:10.1029/JB089iB07p05873.

Shearer, P. M., G. A. Prieto, and E. Hauksson (2006), Comprehensive analysis of earthquake source spectra in southern California, *Journal of Geophysical Research*, *111*(B6), doi:10.1029/2005JB003979.

Takahashi, T., H. Sato, M. Ohtake, and K. Obara (2005), Scale Dependence of Apparent Stress for Earthquakes along the Subducting Pacific Plate in Northeastern Honshu, Japan, *Bulletin of the Seismological Society of America*, *95*(4), 1334–1345, doi:10.1785/0120040075.

Trugman, D. T., and E. M. Dunham (2014), A 2d Pseudodynamic Rupture Model Generator for Earthquakes on Geometrically Complex Faults, *Bulletin of the Seismological Society of America*, *104*(1), 95–112, doi:10.1785/0120130138.

Uchide, T., and K. Imanishi (2016), Small Earthquakes Deviate from the Omega-Square Model as Revealed by Multiple Spectral Ratio Analysis, *Bulletin of the Seismological Society of America*, doi:10.1785/0120150322.

Uchide, T., P. M. Shearer, and K. Imanishi (2014), Stress drop variations among small earthquakes before the 2011 Tohoku-oki, Japan, earthquake and implications for the main shock, *Journal of Geophysical Research: Solid Earth*, *119*(9), 2014JB010,943, doi:10.1002/2014JB010943.

VanWormer, J. D., and A. S. Ryall (1980), Sierra Nevada-Great Basin boundary zone: Earthquake hazard related to structure, active tectonic processes, and anomalous patterns of earthquake occurrence, *Bulletin of the Seismological Society of America*, *70*(5), 1557–1572.

Venkataraman, A., G. C. Beroza, S. Ide, K. Imanishi, H. Ito, and Y. Iio (2006), Measurements of spectral similarity for microearthquakes in western Nagano, Japan, *Journal of Geophysical Research: Solid Earth*, *111*(B3), B03,303, doi: 10.1029/2005JB003834.

Wald, D. J., and T. H. Heaton (1994), Spatial and temporal distribution of slip for the 1992 Landers, California, earthquake, *Bulletin of the Seismological Society of America*, *84*(3), 668–691.

Walter, W. R., and J. N. Brune (1993), Spectra of seismic radiation from a tensile crack, *Journal of Geophysical Research*, *98*(B3), 4449, doi:10.1029/92JB02414.

Walter, W. R., K. Mayeda, R. Gok, and A. Hofstetter (2006), The Scaling of Seismic Energy With Moment: Simple Models Compared With Observations, in *Earthquakes: Radiated Energy and the Physics of Faulting*, edited by R. Abercrombie, A. McGarr, G. D. Toro, and H. Kanamori, pp. 25–41, American Geophysical Union.

Wdowinski, S. (2009), Deep creep as a cause for the excess seismicity along the San Jacinto fault, *Nature Geoscience*, *2*(12), 882–885, doi:10.1038/ngeo684.

Wei, S., E. Fielding, S. Leprince, A. Sladen, J.-P. Avouac, D. Helmberger, E. Hauksson, R. Chu, M. Simons, K. Hudnut, T. Herring, and R. Briggs (2011), Superficial simplicity of the 2010 El MayorCucapah earthquake of Baja California in Mexico, *Nature Geoscience*, *4*(9), 615–618, doi:10.1038/ngeo1213.

Wesnousky, S. G., J. M. Bormann, C. Kreemer, W. C. Hammond, and J. N. Brune (2012), Neotectonics, geodesy, and seismic hazard in the Northern Walker Lane of Western North America: Thirty kilometers of crustal shear and no strike-slip?, *Earth and Planetary Science Letters*, *329-330*, 133–140, doi:10.1016/j.epsl.2012.02.018.

Yang, W., and Y. Ben-Zion (2010), An Algorithm for Detecting Clipped Waveforms and Suggested Correction Procedures, *Seismological Research Letters*, *81*(1), 53–62, doi:10.1785/gssrl.81.1.53.

Yenier, E., and G. M. Atkinson (2014), Equivalent Point-Source Modeling of Moderate-to-Large Magnitude Earthquakes and Associated Ground-Motion Saturation Effects, *Bulletin of the Seismological Society of America*, *104*(3), 1458–1478, doi:10.1785/0120130147.

Yenier, E., and G. M. Atkinson (2015), Regionally Adjustable Generic Ground Motion Prediction Equation Based on Equivalent Point Source Simulations: Application to Central and Eastern North America, *Bulletin of the Seismological Society of America*, *105*(4), 1989–2009, doi:10.1785/0120140332.

Zöller, G., and Y. Ben-Zion (2014), Large Earthquake Hazard of the San Jacinto Fault Zone, CA, from Long Record of Simulated Seismicity Assimilating the Available Instrumental and Paleoseismic Data, *Pure and Applied Geophysics*, *171*(11), 2955–2965, doi:10.1007/s00024-014-0783-1.

# Chapter 6

# Source spectral properties of small-to-moderate earthquakes in southern Kansas

## Abstract

The source spectral properties of injection-induced earthquakes give insight into their nucleation, rupture processes, and influence on ground motion. Here we apply a spectral decomposition approach to analyze $P$-wave spectra and estimate Brune-type stress drop for more than 2000 $M_L$1.5–5.2 earthquakes occurring in southern Kansas from 2014 to 2016. We find that these earthquakes are characterized by low stress drop values (median $\sim$0.4MPa) compared to natural seismicity in California. We observe a significant increase in stress drop as a function of

depth, but the shallow depth distribution of these events is not by itself sufficient to explain their lower stress drop. Stress drop increases with magnitude from M1.5–M3.5, but this scaling trend may weaken above M4 and also depends on the assumed source model. Although we observe a nonstationary, sequence-specific temporal evolution in stress drop, we find no clear systematic relation with the activity of nearby injection wells.

## 6.1 Introduction

Over the past ten years, seismicity rates have risen to historically unprecedented levels within the oil-producing regions of Oklahoma and southern Kansas. Scientific consensus has attributed much of the elevated seismicity rate to anthropogenic activity, and in particular to the injection of wastewater from the oil production process into the Arbuckle Group that is stratigraphically above the granitic basement (e.g., *Ellsworth* 2013; *Buchanan* 2015; *Ellsworth et al.* 2015; *Rubinstein and Mahani* 2015; *Walsh and Zoback* 2015; *Yeck et al.* 2017). The abrupt increase in seismic hazard within this region (*Petersen et al.*, 2016, 2017) has spurred numerous observational studies focused on connections between fluid injection and seismicity rates (*Keranen et al.*, 2013, 2014; *Weingarten et al.*, 2015; *Goebel*, 2015; *Choy et al.*, 2016; *Barbour et al.*, 2017), the source properties of a subset of the larger events (*Choy et al.*, 2016; *Boyd et al.*, 2017; *Cramer*, 2017; *Walter et al.*, 2017; *Sumy et al.*, 2017), and observed ground motion amplitudes (*Hough*,

2014; *Atkinson et al.*, 2016; *Atkinson and Assatourians*, 2017; *Yenier et al.*, 2017). Although these studies have rapidly advanced scientific understanding of these earthquakes, there is still much that remains unanswered, and the nonstationary nature of the seismicity warrants continued monitoring.

One of the key unresolved questions is whether the dynamic source properties of these likely-induced events, such as corner frequency and stress drop, differ from events within active tectonic regions like California (*Huang et al.*, 2016). This question is of particular importance for ground motion estimation due to the positive correlation of ground motion amplitude and stress drop (e.g. *Boore*, 2003; *Atkinson and Morrison*, 2009; *Baltay et al.*, 2017). Several recent studies have used various parametric spectral fitting methods (*Cramer*, 2017; *Sumy et al.*, 2017) or empirical Green's function spectral ratio approaches (*Boyd et al.*, 2017; *Walter et al.*, 2017; *Huang et al.*, 2017) to analyze source parameters of a subset of the largest of these events, including the 2011 Prague, Oklahoma sequence and the 2014 Milan, Kansas earthquakes. In this study, we apply a spectral decomposition technique (*Trugman and Shearer*, 2017a) to perform a comprehensive analysis of the source spectra and source parameters of seismicity occurring in southern Kansas from 2014 through 2016. We use $P$-wave spectra to derive source parameter estimates – seismic moment, corner frequency, and stress drop – for more than 2000 events with local magnitudes ranging from 1.5 to 5.2 that we have relocated using waveform cross-correlation based techniques. These source parameter estimates present an opportunity for a quantitative comparison to the

source properties of naturally occurring events in California that were analyzed using identical methodology.

We focus our analysis on the distribution and variability in source parameter estimates of the southern Kansas dataset as whole, rather than on individual target events or earthquake sequences. We begin by providing an overview of the southern Kansas study region and the associated waveform data. We then briefly describe the methodology we use to derive relocated event positions, source parameter estimates, and parameter uncertainties for each event in our data set. We next examine the most robust statistical features of our source parameter estimates, including the depth-dependence of corner frequency and stress drop, the scaling of stress drop and seismic moment, and the nonstationary temporal evolution of stress drop during our study period (2014–2016). We compare our source parameter observations to those of natural (tectonic) earthquakes in southern California, and discuss the physical and practical implications of our results for the scientific understanding of earthquake rupture processes and occurrence in southern Kansas, and for the probabilistic assessment of ground motion amplitudes and seismic hazard in oil-producing regions of the central United States.

## 6.2 Data and Study Region

Seismicity rates in southern Kansas began to sharply increase in 2013 compared to historical norms (*Hildebrand et al.*, 1988; *Buchanan*, 2015; *Choy et al.*,

2016), with this increase attributed in part to the proliferation of wastewater injection related to oil and gas production within the Mississippian limestone play that underlies southern Kansas and northern Oklahoma (*Ellsworth*, 2013; *Ellsworth et al.*, 2015; *Buchanan et al.*, 2014; *Weingarten et al.*, 2015). The Precambrian basement formation begins at a depth of 1.7 to 2.0 km and is cross-cut by numerous in-situ fault systems, the most notable of which is the northeast striking Nemaha fault (*Steeples et al.*, 1979; *Baars and Watney*, 1991; *McBee*, 2003; *Niemi*, 2004). Oil production wells tap into the shallower sedimentary strata, with wastewater disposal typically occurring in the permeable Arbuckle Group that directly overlies the granitic basement (*Buchanan et al.*, 2014; *Kroll et al.*, 2017). We use in this study the wastewater injection and enhanced oil recovery well locations publicly archived by the Kansas Corporation Commission (http://kcc.ks.gov, last accessed April 2017).

Here we analyze seismicity occurring within southern Kansas from March 21, 2014 through December 31, 2016 (Figure 6.1). The start date for our study period was chosen based on the installation date (March 19–21, 2014) of the U. S. Geological Survey (USGS) Induced Seismicity Menlo Park Project (ISMP) network that was established to monitor seismicity within this region (*Rubinstein et al.*, 2014). Azimuthal station coverage in this area is generally good once the ISMP network was fully installed (late summer 2014), and as such we have few source parameter estimates prior to September 2014. We take initial locations and magnitudes for earthquakes in our dataset from the ISMP catalog, which is

a subsidiary of the Advanced National Seismic Systems (ANSS) Comprehensive Earthquake Catalog (ComCat, https://earthquake.usgs.gov/earthquakes/search/, last accessed May 2017) that lists events by local magnitude. For our study, we convert waveform data with sampling rates of 100Hz and 200Hz from the USGS, Central and Eastern US, NetQuakes, Oklahoma Seismic, and US National Seismic networks (network codes GS, N4, NQ, OK, US) into multiplexed event-based files for later analysis and processing. In total, we consider 5269 events occurring within our study region during this time period, though only a well-recorded subset of 2069 of these events met the quality control criteria for our source parameter estimates (Section 6.3).

## 6.3 Methods: Relocations and Source Parameter Estimates

The primary focus of this article is the analysis and interpretation of earthquake source spectra and source parameters. Although highly accurate earthquake locations are not essential for the spectral analysis, they are useful in the interpretation of spatial variations in source parameters such as systematic trends with depth or with distance from injection wells. Because of this, as a preliminary step in our analysis of source properties, we apply the GrowClust algorithm (*Trugman and Shearer*, 2017b) to obtain relocated event positions and location uncertainties for the southern Kansas earthquakes in our dataset. For these relocations,

we perform time-domain waveform cross-correlation of pairs of events within 4km catalog distance of each other, saving cross-correlation results for all event pairs with a minimum of 8 differential times with cross-correlation values greater than 0.7. We input the resulting approximately one million differential times and cross-correlation values into the GrowClust program, which uses a hybrid hierarchical clustering and relocation algorithm that provides stable relocation results for large-scale catalogs with multiple discrete clusters. The relocated seismicity has median horizontal and vertical location errors of 131m and 281m, respectively.

We next obtain source parameter estimates for a subset of the relocated seismicity using the spectral decomposition method. We follow closely the algorithm described in detail by *Trugman and Shearer* (2017a), and summarize only the main points here. The central idea underlying this approach is that for large and well-recorded seismicity datasets, each earthquake is recorded by many stations, each station records many earthquakes, and each approximate source-station travel path is traversed many times. If this assumption is valid, then it is possible to decompose the waveform data spectra from event $i$ recorded at station $j$ into relative source, station, and path-dependent terms, plus a residual error term ($r_{ij}$) for each trace. Working in the log-frequency domain, the relative contributions add linearly, and the spectral decomposition at each frequency can be written in the form

$$d_{ij} = e_i + st_j + tt_{k(i,j)} + r_{ij}, \tag{6.1}$$

where $d_{ij}$ is the recorded waveform data spectra, $e_i$ is the relative source term, $st_j$ is the relative station term, and $tt_{k(i,j)}$ is a relative path-dependent term that is typically assumed to be isotropic and discretized by source-receiver travel time such that there is a single relative path term for all traces in the $k^{th}$ travel time bin.

The spectral decomposition method consists of four basic steps:

1. Compute the amplitude spectra $d_{ij}(f)$ from the waveform data of each trace.

2. Decompose the amplitude data spectrum of all traces into relative source, station, path, and residual terms by solving Equation (6.1) at each frequency point using an iterative, robust least-squares inversion algorithm with outlier suppression.

3. Infer the empirical Green's function correction spectrum (EGF) that best captures path effects common to all sources, such as average near-source and near-receiver attenuation.

4. Subtract the EGF from each source spectra: $s_i(f) = e_i(f) - EGF(f)$, and use the corrected source spectra $s_i$ to obtain source parameter estimates and uncertainties for seismic moment $M_0$, corner frequency $f_c$, stress drop $\Delta\sigma$.

For steps (1) and (2), we consider $P$-wave spectra of earthquakes with local magnitude $M_L$1.5 and greater, recorded on vertical-component, high-broadband and short-period channels (HHZ, HNZ, EHZ), at stations within 150km distance.

For the spectral estimates, we use a magnitude-dependent window length ranging from a minimum length of 1.5s to a maximum length of 4.5s, where longer windows correspond to larger events in order to permit adequate corner frequency resolution (*Ross and Ben-Zion*, 2016; *Abercrombie et al.*, 2017; *Trugman and Shearer*, 2017a). We define the signal window to begin 0.05s before the catalog-listed $P$-phase arrival time, truncate each window before the catalog-listed $S$-phase arrival when necessary, and define a noise window that immediately precedes the signal window and is of equal length. We discard clipped waveforms using an automated detection algorithm (*Trugman and Shearer*, 2017a) and resample the spectra obtained from longer window lengths to the frequency points corresponding to the minimum window length (1.5s). We compute the average signal-to-noise amplitude in each of five frequency bands (2.5–6, 6–10, 10–15, 15–20, and 20–25Hz), and only further consider events that are recorded at a minimum of six stations with a signal-to-noise ratio greater than 3 in each frequency band.

To estimate the EGF correction term (step 3), we use the technique described by *Trugman and Shearer* (2017a) that fits stacked relative source spectra, averaged in bins of spectral moment $\Omega_0$ to a Brune-type theoretical spectrum of the form:

$$\hat{s}(\, f \mid \Omega_0, f_c, n) = \frac{\Omega_0}{1 + (f/f_c)^n}, \tag{6.2}$$

where $f_c$ and $\Omega_0$ are the corner frequency and spectral moment of each stacked spectra, and the high-frequency falloff rate $n$ is fixed to 2 per the widely-used $\omega^{-2}$

model (*Aki*, 1967; *Brune*, 1970). In contrast to previous implementations of the spectral decomposition method (e.g., *Shearer et al.* 2006), our technique does not require an assumption of self-similar, or constant stress drop scaling, and instead infers the optimal scaling directly from the shape of the stacked spectra. Here we find that the optimal fit requires an EGF with non-self-similar scaling such that stress drop increases with spectral moment (Figure 6.2), a result in agreement with a recent analysis of earthquakes in California (*Trugman and Shearer*, 2017a). Inference of the EGF is an essential part of the spectral decomposition technique because the source terms $e_i$ produced by the solution to Equation (6.1) are relative (median amplitude zero), and thus must be corrected for propagation effects that are common to all sources. This includes the spatially averaged near-source attenuation that is not removed by the nearest travel time terms $tt_{k=1}$, as well as average near-station attenuation, since the station terms $st_j$ isolate only relative differences in near-station and instrument effects on the observed spectra.

We account for the possibility of lateral variations in attenuation by applying a modified spectral stacking technique that uses a distance weighting to allow for spatial variations in the EGF. This technique is similar to the nearest-neighbors EGF approach first used by *Shearer et al.* (2006), but in practice tends to be more stable. In brief, the modified technique uses cluster analysis to define a non-uniform set of grid points that conform to the contours of the observed seismicity. For each grid point (we use six in this study), an EGF is inferred from stacks of relative source spectra that are weighted by inverse distance to the event

locations in the study region, such that events that are closer to the grid point assume more weight. The EGF correction to the source spectrum of each event is then computed as a linear combination of the set of EGFs, again weighted by inverse distance. We note that applying this distance-weighted stacking algorithm to account for lateral variations in attenuation does not significantly influence the results presented in this study, but may be an important consideration for study regions that extend over larger length scales.

It is also important to recognize that the increase in median stress drop with moment for individual events that we present in Section 6.4 is a direct consequence of applying an EGF based on non-self-similar scaling. The evidence for such scaling is the markedly superior fit that we obtain to the stacked spectra (Figure 6.2c) compared to the fit for a self-similar model (Figure 6.2b: requiring self-similarity increases the overall misfit by more than a factor of 3). Note that this result is based only on observations within the 2.5 to 25 Hz band where we have good signal-to-noise, and does not require resolving corner frequencies outside of this band. However, as discussed in *Trugman and Shearer* (2017a), the case for an increase in average stress drop with moment does depend upon the assumption of the Brune spectral model and its $f^{-2}$ high-frequency falloff rate, as reasonable fits to the stacked spectra are possible for self-similar models with high-frequency falloff rates less than 2.

Lastly (step 4), we use the EGF-corrected spectra $s_i(f)$ to estimate source parameters and associated uncertainties. To estimate seismic moment $M_0$, we

assume that on average the observed spectral moment $\Omega_0$ is proportional to the seismic moment $M_0$, and perform a regression analysis between $\Omega_0$ and $M_L$ to calibrate the appropriate scale factor (*Shearer et al.*, 2006). This analysis can be used to obtain a linear relationship between $M_W$ and $M_L$ that is valid for the smaller earthquakes for which $M_W$ is not routinely estimated. The smallest events in our dataset ($M_L$1.5) correspond to $M_W$1.9, and the inferred $M_W$-$M_L$ slope of 0.72 (Figure 6.3) is comparable to the slope of 0.75 obtained for southern California by *Ross et al.* (2016). Although the relationship between $M_W$ and $M_L$ may be slightly nonlinear for larger events (*Ben-Zion and Zhu*, 2002; *Goertz-Allmann et al.*, 2011; *Edwards and Douglas*, 2014; *Munafo et al.*, 2016), we do not observe a significant bias between our $M_W$ estimates and those obtained through moment tensor analysis and listed by ANSS ComCat (https://earthquake.usgs.gov/earthquakes/search/, last accessed May 2017).

We then estimate the corner frequency $f_c$ using a bounded optimization algorithm that minimizes the root-mean-square residual between the observed, EGF-corrected source spectrum $s_i(f)$ and the Brune theoretical spectrum $\hat{s}_i(f|f_c)$ in the 2.5–25Hz band in which we have measured adequate signal-to-noise, and that is insensitive to site resonances and spectral contamination from leaky-mode surface waves that are prevalent for earthquakes in this region (*Cramer*, 2017). Given $f_c$ and $M_0$, we compute the Brune-type stress drop

$$\Delta\sigma = \frac{7}{16} M_0 \left( \frac{f_c}{k\beta} \right)^3, \tag{6.3}$$

appropriate for a simplified circular crack model of the earthquake source with constant stress drop and elliptical distribution of slip (*Brune*, 1970; *Madariaga*, 1976). We obtain a depth-dependent estimate of the shear wave speed $\beta$ from the velocity model used by the ISMP network to locate the events (*Rubinstein et al.*, 2015), and set the numerical constant $k = 0.38$ following the recent numerical analysis of *Kaneko and Shearer* (2014). We derive uncertainty estimates for $M_0$, $f_c$, and $\Delta\sigma$ using the statistical resampling techniques detailed in *Trugman and Shearer* (2017a) that are based upon the variability in the apparent source spectra recorded at each station. Normalized uncertainties in corner frequency ($\Delta f_c / f_c$) tend to be higher for the lowest and highest magnitude events, which have $f_c$ that approach the 2.5–25Hz limits of the spectral bandwidth. However, the spectral decomposition results for the dataset in aggregate are insensitive to resolution of source parameters of these individual events (which are few in number), but are instead controlled primarily by the relative shape of the stacked spectra (see *Trugman and Shearer*, 2017a for a complete discussion).

## 6.4 Results

We apply the spectral decomposition method (Section 6.3) to analyze the earthquake source parameters of 2069 well-recorded earthquakes in our study region. The southern Kansas earthquakes in our dataset have relatively low stress drop values, with an overall distribution that is approximately log-normal with a

median stress drop of 0.41MPa and $\log_{10}$ standard deviation of 0.35 (Figure 6.4a). The relative simplicity of this total marginal histogram of $\Delta\sigma$ does however obscure several notable trends. As a concrete example, in Figure 6.4 we also show conditional histograms of $\Delta\sigma$ for four distinct magnitude ranges (1.5–2.0, 2.0–2.5, 2.5–3.0, and 3.0–3.5), and plot the distribution of corner frequency $f_c$ and stress drop $\Delta\sigma$ as a function of seismic moment $M_0$. From this perspective it is apparent that median stress drop tends to increase as function of moment, a result that is consistent with the inferred scaling of the stacked spectra (Figure 6.2), but is in direct violation of the classical self-similar model first proposed by *Aki* (1967).

We can quantify this scaling trend by performing a weighted regression analysis of stress drop and moment, fitting a linear model of the form:

$$\log_{10} \Delta\sigma = \epsilon_0 + \epsilon_1 \log_{10} M_0, \tag{6.4}$$

where the parameter $\epsilon_1$ measures the slope of the increase of $\log_{10} \Delta\sigma$ with $\log_{10} M_0$. For the results presented here, we compute median stress drop in bins of width 0.4 in $\log_{10} M_0$, and apply weights based on the median uncertainty in $\Delta\sigma$ and number of observations in each bin, but obtain comparable results for both unbinned and unweighted regression. The scaling parameter $\epsilon_1 = 0.25$ (2-$\sigma$ uncertainty $\pm 0.035$) is clearly positive and therefore inconsistent with the null hypothesis of self-similar, constant-$\Delta\sigma$ data, which would have $\epsilon_1 = 0$ to within the uncertainties. These results for the southern Kansas events are within the $\epsilon_1 \sim 0.1$–$0.4$ range of scaling results obtained for California earthquakes by *Trugman and Shearer* (2017a) and

by other studies that have quantified an increase in stress drop or scaled energy with moment (e.g., *Mayeda and Walter* 1996; *Izutani and Kanamori* 2001; *Mori et al.* 2003; *Mayeda et al.* 2005; *Takahashi et al.* 2005; *Mayeda et al.* 2007; *Calderoni et al.* 2013; *Pacor et al.* 2016; *Agurto-Detzel et al.* 2017). However, our Kansas dataset contains few higher-magnitude events (20 $M_L > 3.5$ events, which is less than 1% of the total count), and as such we have poor resolution of the distribution and scaling of $\Delta\sigma$ above M $\sim$ 3.5. Based on the data we do have, it would be reasonable to expect earthquakes with $\Delta\sigma$ in the 1–10MPa range for the M $\geq$ 4 that are of fundamental interest to hazard calculations.

Another point of interest from a hazard perspective is the systematic increase in median $f_c$ and $\Delta\sigma$ with hypocentral depth (Figure 6.5ab). In contrast, we do not observe a comparable depth-dependent trend in $M_0$ that could potentially account for the scaling results presented above (Figure 6.5c). The depth-dependence in both $f_c$ and $\Delta\sigma$ (but not $M_0$) implies an increase in rupture velocity $v_r$ that outpaces the expected increase in shear wave speed $\beta$ with depth (i.e., an increase in the ratio $v_r/\beta$ with depth). Because earthquakes in southern Kansas are characterized by a shallower depth-distribution (2–8km) compared to seismicity in the western and eastern United States (which typically extends to a significantly greater maximum depth), it is plausible that the lower median $\Delta\sigma$ values we observe are due in part to the shallowness of the seismicity (*Hardebeck and Aron*, 2009; *Pacor et al.*, 2016; *Agurto-Detzel et al.*, 2017; *Boyd et al.*, 2017; *Sumy et al.*, 2017). However, the shallower depth distribution of the events in our

dataset does not fully explain their anomalously low $\Delta\sigma$, as can be seen quantitatively by comparing median $\Delta\sigma$ for Kansas and California seismicity within a fixed depth range. For example, in Kansas we observe a median $\Delta\sigma$ of 0.4MPa at 5km depth (Figure 6.5), compared to the 1–4MPa range observed at this depth in five different regions of California (*Trugman and Shearer*, 2017a).

We next turn our attention to the temporal variability of $\Delta\sigma$ for earthquakes in our study region. It is useful at this stage to introduce a normalized, magnitude-adjusted $\Delta\sigma$:

$$Z_{\Delta\sigma} = \frac{\log_{10}\Delta\sigma - E[\log_{10}\Delta\sigma \mid M_0]}{STD\left\{\log_{10}\Delta\sigma - E[\log_{10}\Delta\sigma \mid M_0]\right\}}, \tag{6.5}$$

where $E[\log_{10}\Delta\sigma|M_0]$ is expected $\Delta\sigma$, given $M_0$ (Equation 6.4), and $STD\{\cdot\}$ refers to the standard deviation. Thus the metric $Z_{\Delta\sigma}$ is a normalized measure of the deviation from the expected $\Delta\sigma$ value of each event, conditioned on the observed $M_0$. This framework allows us to better isolate significant temporal variations in $\Delta\sigma$ by accounting for the magnitude-scaling trend observed in Figure 6.4. In Figure 6.6, we plot local magnitude as a function of time, with events color-coded by $Z_{\Delta\sigma}$ such that bluer colors correspond to events with higher than expected $\Delta\sigma$ (i.e., events enriched in high-frequency energy). Prior to September 2014, we have few well-resolved source parameter estimates due to the sparsity in local station coverage before the ISMP network was fully installed. Following this, in the weeks preceding the 12 November 2014 Milan earthquake ($M_L$5.2, $M_W$4.9) (*Choy et al.*, 2016), we observe elevated levels of $\Delta\sigma$. The Milan event itself has a slightly lower

estimated $\Delta\sigma$ (3.6 MPa) than would be expected by extrapolating the scaling trend of Equation (6.4), though still nearly an order of magnitude greater than the median value of the dataset as a whole. The early aftershocks of Milan have lower than expected $\Delta\sigma$ values that appear to increase with time, consistent with observations of the Prague, OK sequence (*Sumy et al.*, 2017; *Yenier et al.*, 2017). In early 2015, we observe another temporal cluster of high $\Delta\sigma$ events, with several other sequences later in 2015 and 2016 exhibiting analogous behavior.

To examine the variability in $\Delta\sigma$ within and between prominent earthquake sequences, we use the method described by *Zaliapin and Ben-Zion* (2013) that partitions events into individual sequences based upon nearest-neighbor space-time distances. The nearest-neighbors method has been shown to be effective in characterizing the space-time clustering statistics of both tectonic and induced earthquake sequences (*Schoenball et al.*, 2015; *Zaliapin and Ben-Zion*, 2016). It defines the distance $\eta_{ij}$ between an event pair (parent $i$, daughter $j$) to be the product of a rescaled time $T_{ij} = dt_{ij}10^{-M_i/2.}$ and rescaled distance $R_{ij} = dr_{ij}{}^{d}10^{-M_i/2.}$, where $dt_{ij}$ is the difference in time in years, $dr_{ij}$ is the spatial distance in km, $M_i$ is the magnitude of the parent event, and $d = 1.6$ is the assumed fractal dimension. Events are then linked to their nearest neighbors, and individual sequences are defined by selecting a threshold distance $\eta_{ij}$ such that the sequences are sufficiently clustered in space and time (Figure 6.7a). With sequences thus defined, we then compare the variability in $\Delta\sigma$ both within and between the most prominent sequences in our dataset, each of which contains at least one $M_L > 3.0$ event

(the largest of which we classify as the mainshock of the sequence). Although $\Delta\sigma$ typically varies by slightly more than an order of magnitude within each sequence, median values can vary by as much as a factor of three between sequences (Figure 6.7b). Mainshock values of $\Delta\sigma$ tend to be higher than the median value of their respective sequences, as expected given the observed magnitude scaling. However, we do not observe comparable systematic differences for the magnitude-corrected values of $Z_{\Delta\sigma}$ (Figure 6.7c).

These temporal and sequence-specific variations in $\Delta\sigma$ could be caused by a number of factors, including local variations in geologic properties, crustal stress heterogeneity and its redistribution during and between individual earthquake sequences, and time-dependent changes in anthropogenic stressing from oil production and fluid injection. To gain insight into the latter, in Figure 6.8, we plot $Z_{\Delta\sigma}$ in map view for earthquakes within our study region and compare to the locations of active wastewater injection and enhanced oil recovery wells (Section 6.2), both of which are thought to influence seismicity rates (e.g. *Rubinstein and Mahani* 2015). We do not observe a significant correlation between $Z_{\Delta\sigma}$ and radial distance to the nearest active well (Figure 6.9), although it is interesting that clusters of events with the highest $Z_{\Delta\sigma}$ values tend to be near active wells. This weak or nonexistent dependence suggests that the presence of temporally and spatially coherent clusters of events with similar stress drop are caused primarily by factors unrelated to a localized influence injection activity of the nearest wells, such as local differences in fault strength, the distribution of geological or geometric

asperities, or lithology. However, we note that because we do not have access to daily injection records for each well, it is difficult to perform a truly quantitative analysis in this regard.

## 6.5 Discussion

In this study, we use $P$-wave spectra from earthquakes in southern Kansas to provide source parameter and uncertainty estimates for seismic moment ($M_0$), corner frequency ($f_c$), and Brune-type stress drop ($\Delta\sigma$). However, we emphasize that the absolute values of the source parameters are valid only under the assumption of the assumed source model, which in this case is a Brune-type spectrum (*Brune*, 1970) with $\omega^{-2}$ high-frequency falloff (Equation 6.2 with $n = 2$). Our uncertainty estimates are therefore lower bounds because they neglect the epistemic uncertainties associated with this parameterization of the source spectral model and with the assumption that a circular crack rupture (Equation 6.3) adequately describes the relevant source physics. As discussed in *Trugman and Shearer* (2017a), the strength of the inferred scaling ($\epsilon_1$) of $\Delta\sigma$ with $M_0$ is correlated with the assumed high-frequency falloff rate $n$. If for example, the true average $n$ for these earthquakes is less than the canonical $\omega^{-2}$ value of 2 (*Brune*, 1970), then the scaling parameter $\epsilon_1$ will be lower, and more generally if $n$ varies on an event-to-event basis, this will bias estimates of $f_c$ and $\Delta\sigma$ for events in which $n$ differs markedly from 2. The spectra do not contain adequate signal bandwidth

or precision to provide independent estimates of $f_c$ and $n$, only their joint influence on the spectral shape.

Despite these concerns, we can still draw useful conclusions by (i) focusing on relative variations in the source parameters, which tend to be robust with respect to the model parameterization, and (ii) by comparing to other datasets of earthquakes analyzed using the same methodological assumptions. Even from this more cautionary perspective, we can conclude that the southern Kansas earthquakes are characterized by relatively low $\Delta\sigma$ values compared to naturally occurring seismicity within tectonically active regions of California. The observed increase in $\Delta\sigma$ with hypocentral depth cannot fully account for this discrepancy, which suggests that the nucleation or rupture processes of these events may differ in some more fundamental way. While these results are consistent with several recent studies of likely-induced earthquakes (*Agurto-Detzel et al.*, 2017; *Boyd et al.*, 2017; *Sumy et al.*, 2017), others (*Zhang et al.*, 2016; *Huang et al.*, 2016, 2017) have suggested that tectonic and induced events have comparable source parameters. These disparate conclusions may indicate that differences in local faulting conditions, tectonic stress regime, or history of anthropogenic activity may all play an important role. For example, the 25 earthquakes in the Guy-Greenbrier, Arkansas earthquakes analyzed by *Huang et al.* (2016) differ substantially in both their tectonic setting and exposure to widespread regional injection compared to the southern Kansas earthquakes. *Boyd et al.* (2017) and *Huang et al.* (2017) both suggest that variations in $\Delta\sigma$ within the US can be understood in the context of

Mohr-Coulomb theory, in which the failure stress depends on hypocentral depth, fault style, and coefficient of friction (a proxy for fault strength). The results we present here are broadly consistent with this framework, as the shallower hypocentral depths and transtensional tectonic regime of southern Kansas would portend lower median $\Delta\sigma$ than in the deeper, transpressional regime of southern California. Local variations in fault strength and its dependence on fluid injection may have an additional modulating effect that could be explored further in future studies.

We also observe an increase in $\Delta\sigma$ with magnitude for the smaller events that comprise the majority of our dataset ($M < 3.5$). Although the inferred scaling is comparable to that observed in California (*Trugman and Shearer*, 2017a), it is sensitive to the modeling assumptions as discussed above. Bandwidth limitations can in some circumstances hinder the resolution of $f_c$ for smaller magnitude events (*Ide*, 2003; *Abercrombie*, 2015; *Huang et al.*, 2016), but we believe this effect does not significantly influence the results presented in this study. The low median stress drop of the southern Kansas earthquakes, when combined with the low regional attenuation and wide available signal bandwidth, provides a near-optimal setting for source spectral analyses, and indeed only a small fraction of our dataset is poorly resolved (Figure 6.4). Further, the observed scaling is controlled fundamentally by the shape of the stacked relative source spectra (*Trugman and Shearer*, 2017a) not the inferred source parameter values of individual events. It is however possible that this scaling trend may not extrapolate linearly to the larger magnitudes (M4 and M5) that are poorly sampled by our dataset. Still, the larger events in our

dataset, including the $M_L$5.2 Milan earthquake, typically have $\Delta\sigma$ in the 1–10 MPa range, which is 2.5–25x greater than the median value of the dataset as a whole.

We observe coherent spatial and temporal variations and clustering of stress drop within our study region, but these variations do not appear to have a clear relation with distance to nearby injection and enhanced oil recovery wells. This is perhaps not surprising, as we lack adequate spatial and temporal resolution within the publicly available data to disentangle its effect from the other features controlling source properties. Further, while the first-order influence of injection on seismicity rate is apparent based on comparison to the historical record (*Buchanan*, 2015; *Buchanan et al.*, 2014; *Weingarten et al.*, 2015; *Choy et al.*, 2016), its immediate influence on source properties is more nebulous both from an observational perspective due to the lack of historical precedent, and from a geophysical perspective due to the complex, nonlinear interactions between anthropogenic stressing and the rupture dynamics of triggered events. *Sumy et al.* (2017) likewise observe both significant temporal changes in $\Delta\sigma$ and a lack of correlation with injection well distance for aftershocks of the 2011 Prague, Oklahoma sequence, and attribute the low observed $\Delta\sigma$ values to the more widespread effects of regional injection patterns that weakened basement fault structures on regional rather than local length scales. Fluid injection on the spatial scale of that observed in southern Kansas during this time period may generate significant stress perturbations at distances of tens of kilometers or more through a combination

of pore pressure increase and poroelastic stressing (*Segall and Lu*, 2015; *Goebel et al.*, 2017). It is also notable that *Boyd et al.* (2017) observe both a mild increase of $\Delta\sigma$ with $M_0$ for their data set comprised entirely of M $>$ 3 events, as well as anomalously low $\Delta\sigma$ for select aftershock sequences in the central United States like that of the 2014 Milan event. In addition to their utility in understanding source dynamics, source parameter estimates may also provide observational constraints for seismic hazard assessment. Because ground motion intensities at high frequencies are controlled primarily by stress drop (*Boore*, 2003; *Baltay et al.*, 2013; *Yenier and Atkinson*, 2014; *Douglas and Edwards*, 2016; *Baltay et al.*, 2017), its characterization is of fundamental interest to studies that aim to develop ground motion prediction equations (GMPEs) for induced events (*Atkinson et al.*, 2016; *Atkinson and Assatourians*, 2017; *Yenier et al.*, 2017). In this study, we observe quantifiable time-dependent and depth-dependent variations in stress drop, both of which are in accord with the conclusions of *Yenier et al.* (2017) and *Atkinson and Assatourians* (2017) for ground motions of recent seismicity in Oklahoma. This consistency suggests that the results we present could potentially serve as a basis for future studies focused on quantifying the influence that spatiotemporal and depth-dependent variations in earthquake source properties may have on the observed ground motion amplitudes of induced earthquakes in the central United States.

## 6.6 Summary

We estimate seismic moment, corner frequency, and Brune-type stress drop for 2069 $M_L$1.5–5.2 earthquakes occurring from 2014–2016 in an active area of wastewater injection and oil and gas production in southern Kansas. We find that these earthquakes have relatively low stress drop values that increase with hypocentral depth. We observe an increase in median stress drop as a function of magnitude for the M1.5–3.5 earthquakes that comprise the majority of our dataset. However, this scaling trend may partially slow or saturate at higher magnitudes, and its strength is sensitive to the parameterization of the assumed source model. We find coherent temporal and spatial variations in the source parameters of earthquakes in southern Kansas, but these variations are not systematically related to the activity of nearby wastewater injection and enhanced oil recovery wells.

## Acknowledgements

The waveform data used in this study is archived by and is publicly available from the IRIS Data Management Center (http://ds.iris.edu/ds/, last accessed May 2017). The source parameter catalog associated with this study is available in the electronic supplement to this manuscript (Supplementary Dataset D1). This manuscript is based upon work supported by the National Science Foundation Graduate Research Fellowship Program (NSFGRFP) under grant number DGE-1144086. Additional support was provided by the Paul G. Silver Young Scholar

Research Enhancement Award and the Southern California Earthquake Center (SCEC) under grant number 16020. We wish to thank Yihe Huang, Jeanne Hardebeck, Valerie Sahakian, Editor Yehuda Ben-Zion, and an anonymous reviewer for their constructive comments and reviews, as well as Justin Rubinstein, Annemarie Baltay, and Rachel Abercrombie for their helpful advice and for stimulating discussion that contributed to the content of this manuscript.

Chapter 6, in full, is a reformatted version of material as it appears in Journal of Geophysical – Solid Earth: Trugman, D. T. and P. M. Shearer (2017), Source spectral properties of small-to-moderate earthquakes in southern Kansas, *Journal of Geophysical Research: Solid Earth*, 122(10), doi:10.1002/2017JB014649. I was the primary investigator and author of this paper.

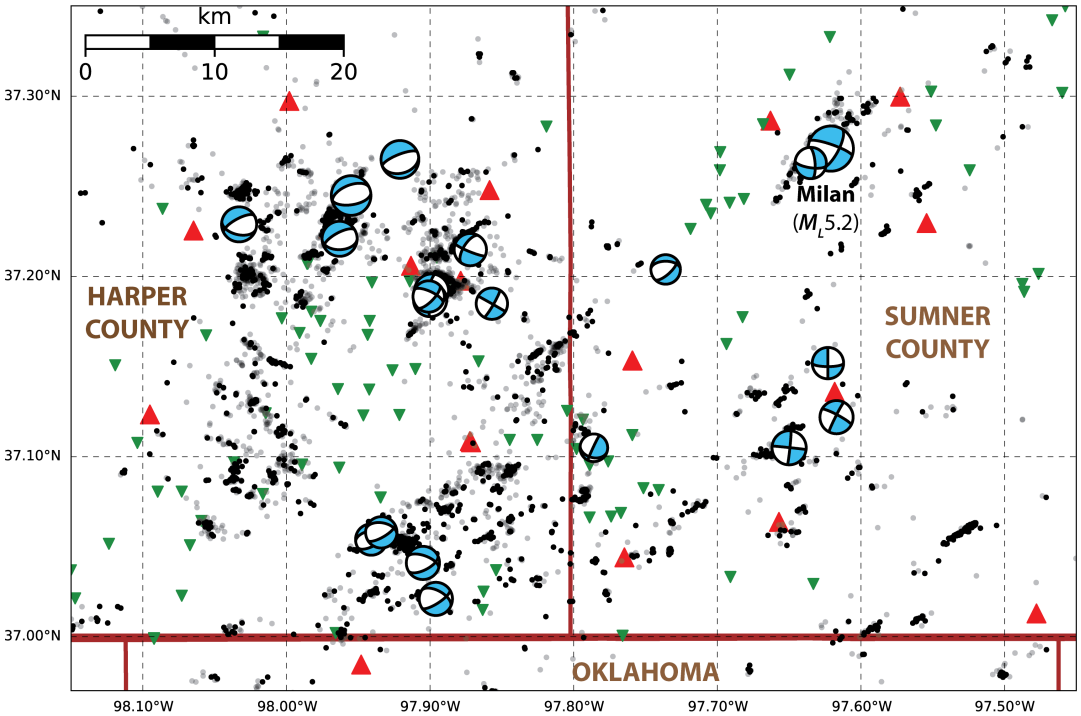


**Figure 6.1**: Map view of 2014–2016 southern Kansas seismicity, with the relocated epicenters of events considered in this study shown in gray, and the subset of these events for which we obtain source parameter estimates shown in black. Local station coverage (red triangles) and locations of active wastewater disposal and enhanced oil recovery wells (green inverted triangles) obtained from the Kansas Corporation Commission (http://kcc.ks.gov, last accessed April 2017), and ANSS ComCat focal mechanism estimates (https://earthquake.usgs.gov/earthquakes/search/, last accessed May 2017) for M3 and greater events are shown for reference. Thin and thick brown lines correspond to county and State boundaries, respectively.

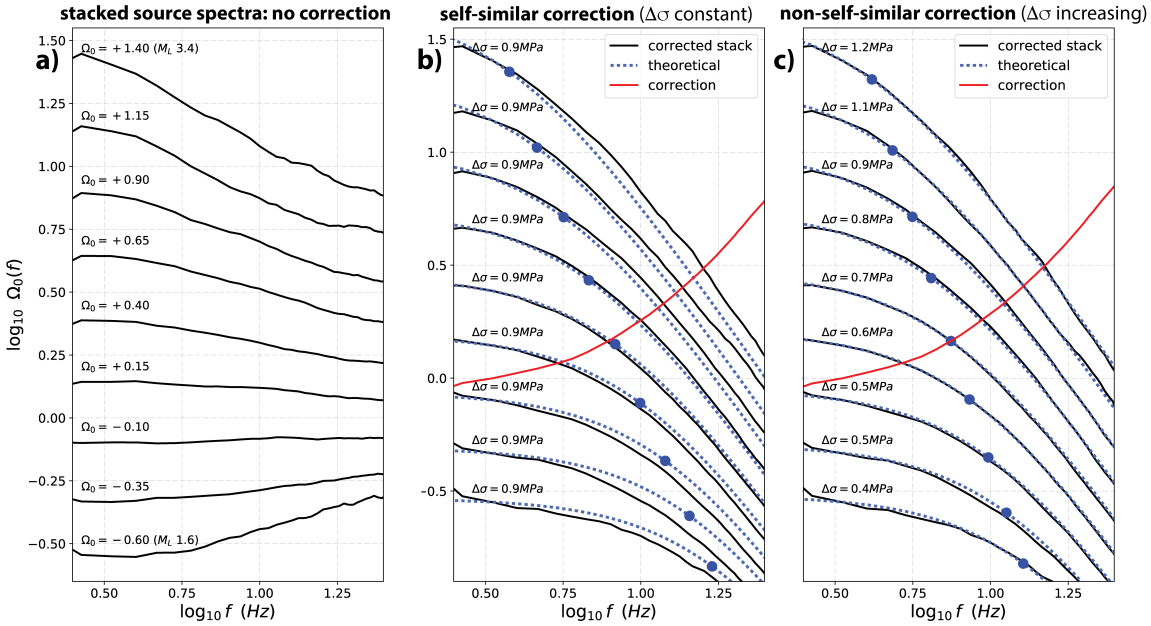


**Figure 6.2**: Example of stacked source spectra from earthquakes in southern Kansas, before and after the EGF correction for common path effects (see text for details). Panel (a) shows stack-averaged relative source spectra, binned by spectral moment $\Omega_0$ (logarithmic scale), prior to the EGF correction. Panels (b) and (c) both show EGF-corrected stacked source spectra (solid black lines) and a comparison to theoretical source spectra (dashed blue lines), with the implied corner frequency of each stack (blue dots) and the EGF spectrum (red line) marked for reference. In panel (b), the EGF is estimated with a self-similarity (constant stress drop) constraint, and provides a significantly worse fit than the EGF in panel (c) with no such constraint.

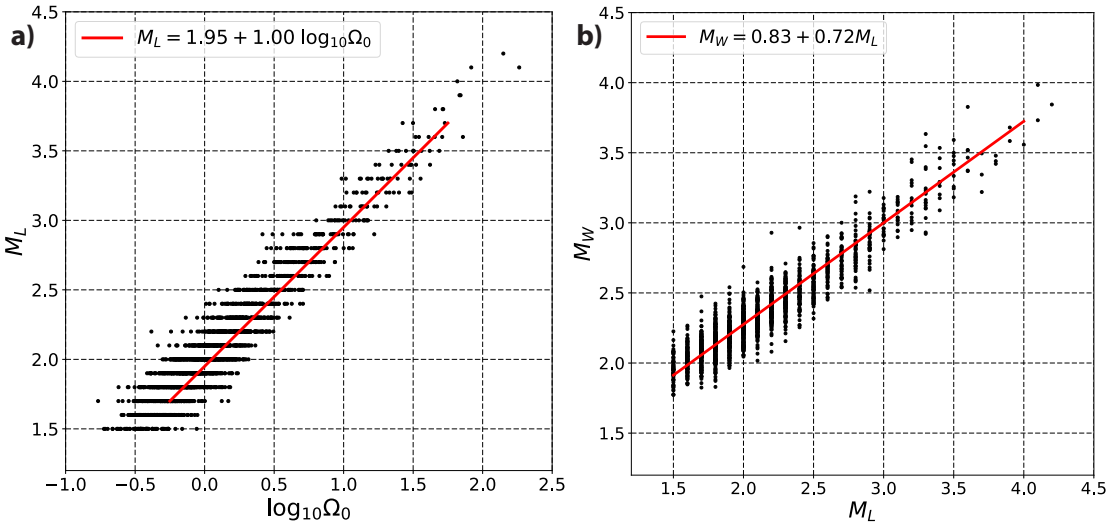


**Figure 6.3**: Scatterplots showing the relation between (a) $M_L$ and $\log_{10}\Omega_0$ and (b) $M_W$ and $M_L$ for the southern Kansas earthquakes. The best-fitting regression line in each panel is marked with a solid red line and labeled in the upper left inset.

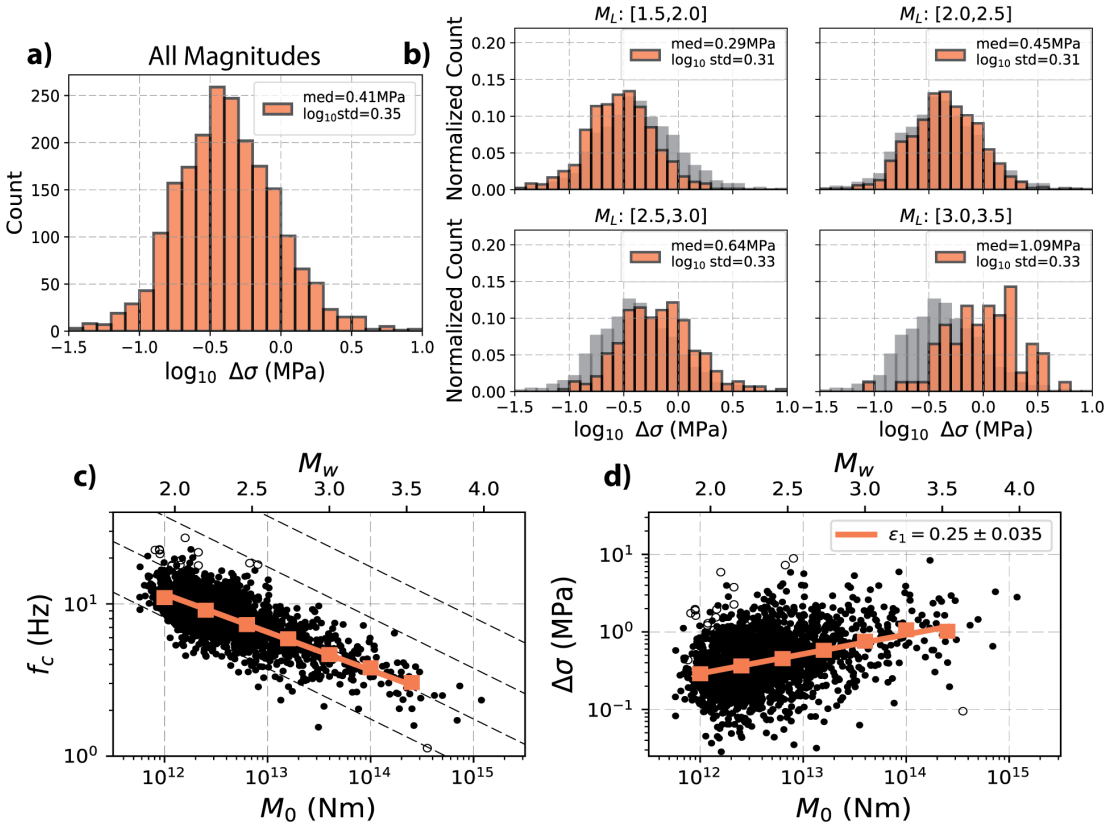


**Figure 6.4**: The distribution and magnitude-scaling of corner frequency ($f_c$) and stress drop ($\Delta\sigma$) for southern Kansas earthquakes. The top panels show histograms of (a) the total $\Delta\sigma$ distribution, marginalized over all magnitudes and (b) distributions corresponding to different magnitude ranges. The gray bars in panel (b) correspond to the total histogram (for all magnitudes) shown in panel (a). The bottom panels show the scaling of (c) $f_c$ and (d) $\Delta\sigma$ with seismic moment $M_0$. Black dots correspond to measurements of source properties for individual events, and the median $f_c$ and $\Delta\sigma$ in $M_0$ bins of width 0.4 ($\log_{10}$ N-m units) are marked with orange squares. The best-fitting scaling parameter $\epsilon_1$ for the binned data (obtained from weighted regression analysis, see text for details) is plotted with a solid orange line, and its numerical value and two-sigma uncertainty is listed in the panel (d) inset. The dashed black lines in panel (c) correspond to constant-$\Delta\sigma$ contours of 0.1, 1, 10, and 100 MPa. Open circles denote events with poorly resolved $f_c$ due to bandwidth limitations ($f_c > 25$Hz or bootstrap inter-quartile uncertainty $> 5$ Hz).

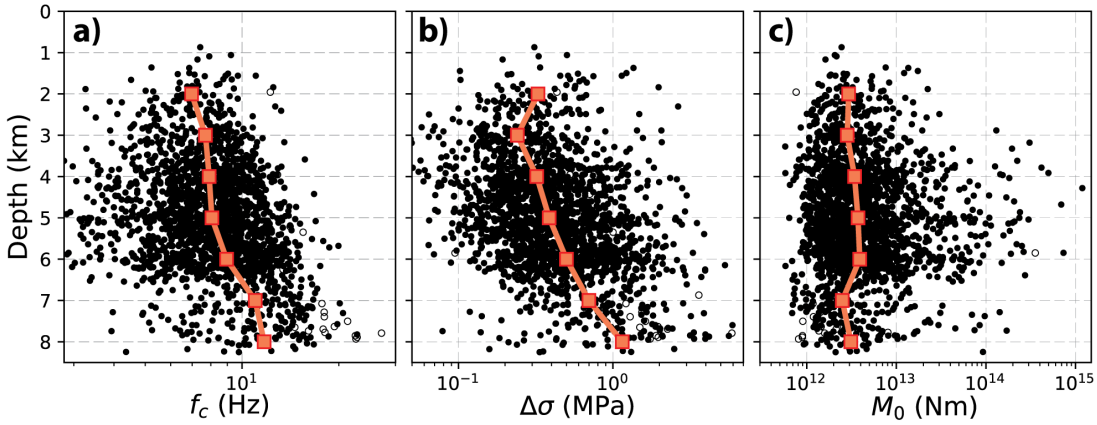


**Figure 6.5**: Depth-dependence of (a) corner frequency $f_c$, (b) stress drop $\Delta\sigma$, and (c) seismic moment $M_0$. Black dots correspond to measurements of source properties for individual events, and the median $f_c$, $\Delta\sigma$, $M_0$ values in depth bins of 1 km are marked with orange squares. Open circles denote events with poorly resolved corner frequencies due to bandwidth limitations ($f_c > 25$Hz or bootstrap inter-quartile uncertainty $> 5$ Hz).

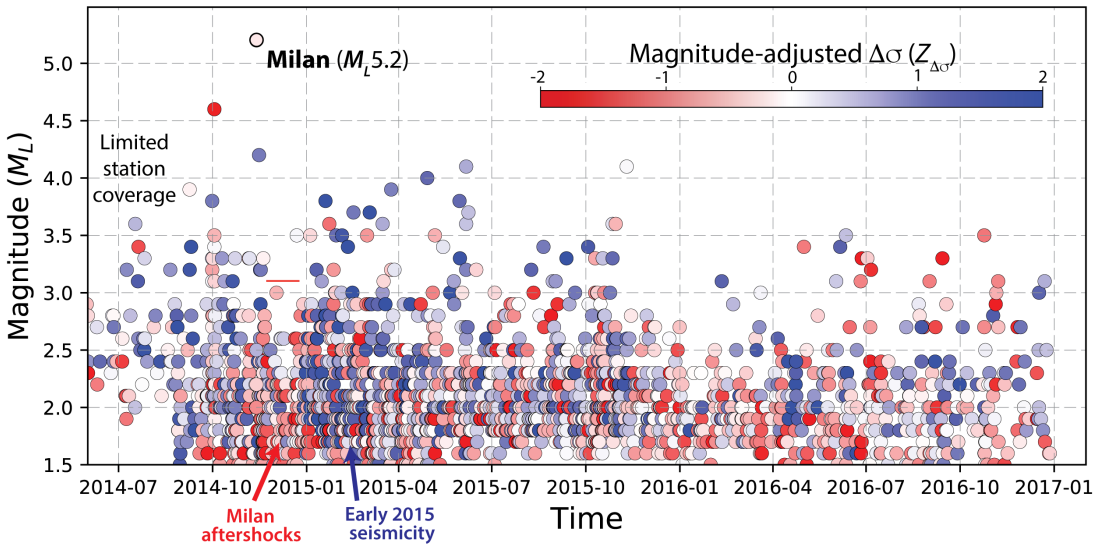


**Figure 6.6**: Temporal evolution of source parameters in southern Kansas. Magnitude ($M_L$) versus time, with events color-coded by magnitude-adjusted, normalized stress drop $Z_{\Delta\sigma}$ (Equation 6.5), with bluer colors indicating higher normalized stress drop. Local station coverage is sparse before September 2014, resulting in fewer events with resolvable source parameter estimates during this time.

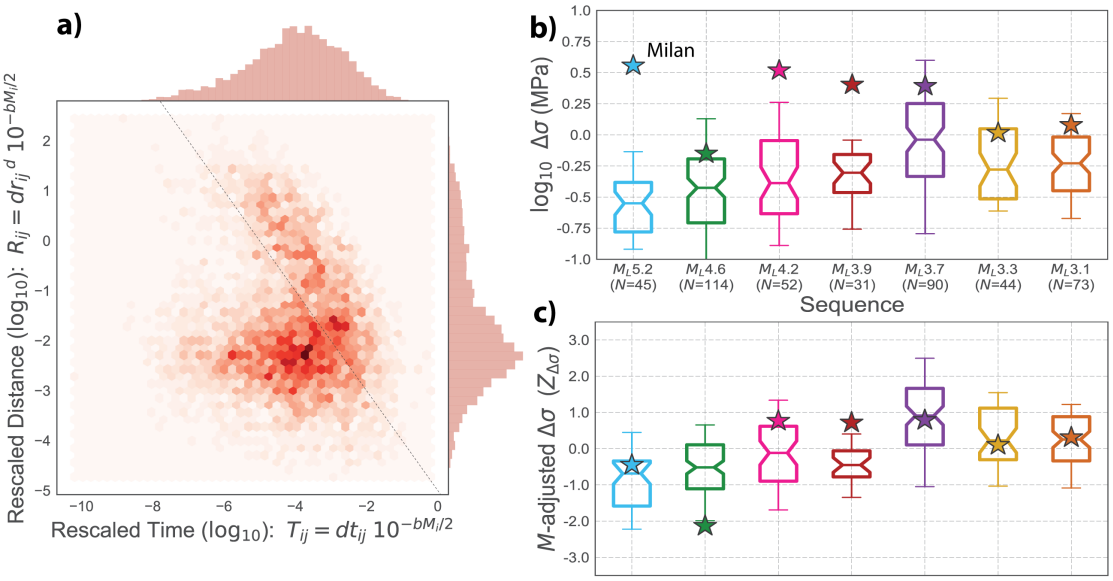


**Figure 6.7**: Variability of source parameters within and between different earthquake sequences in southern Kansas. (a) Joint probability density of rescaled time $T_{ij}$ and rescaled distance $R_{ij}$ (log-log scale) for the nearest-neighbor cluster identification method (*Zaliapin and Ben-Zion*, 2013). The black dashed line corresponds to the threshold distance $\eta_{ij}$ used to partition events into individual sequences. (b) Boxplot distribution of $\Delta\sigma$ (MPa) for prominent earthquake sequences in southern Kansas. Sequences medians are denoted with a solid horizontal line, while the box and whiskers denote the inter-quartile range (50% confidence interval) and 90% confidence interval, respectively. $\Delta\sigma$ values for the largest event in each sequence (mainshocks) are marked with stars, and the corresponding magnitudes are listed along the x-axis. (c) Similar to panel (b), but for the distribution of magnitude-adjusted, normalized stress drop $Z_{\Delta\sigma}$ (Equation 6.5).

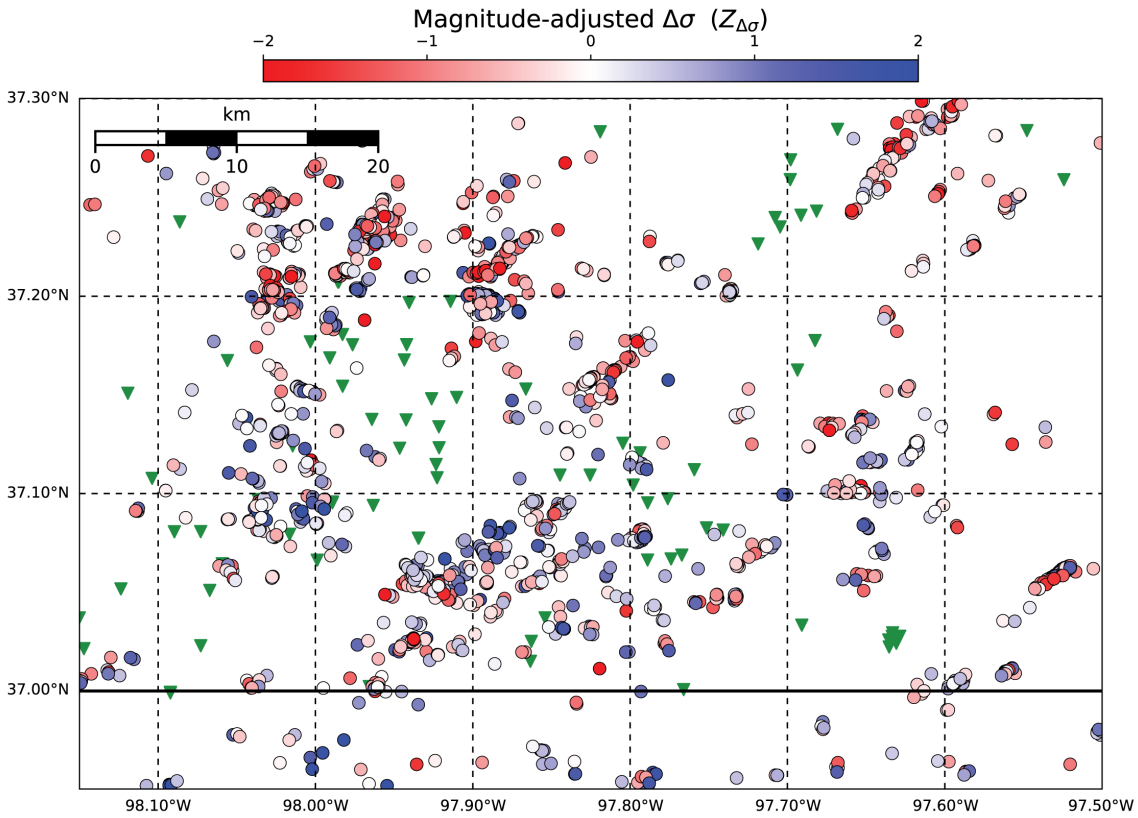


**Figure 6.8**: Map view of southern Kansas source parameter estimates, with events color-coded by magnitude-adjusted, normalized stress drop $Z_{\Delta\sigma}$ (Equation 6.5), with bluer colors indicating higher stress drop. Locations of active wastewater disposal and enhanced oil recovery wells are shown for reference (green triangles).

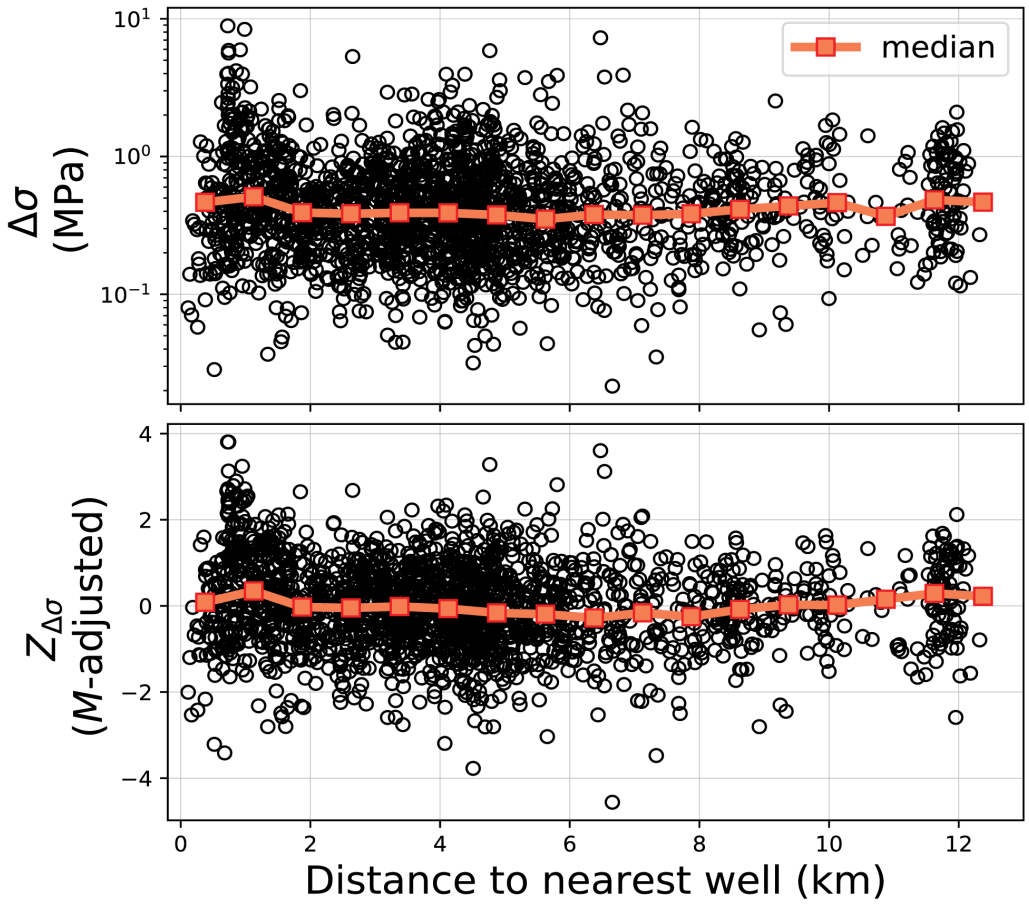


**Figure 6.9**: Stress drop $\Delta\sigma$ (top, MPa) and magnitude-adjusted, normalized stress drop $Z_{\Delta\sigma}$ (Equation 6.5, bottom) plotted as a function of distance to the nearest active injection well. Median values in 0.75km bins marked with orange squares.

# References

Abercrombie, R. E. (2015), Investigating uncertainties in empirical Green's function analysis of earthquake source parameters, *Journal of Geophysical Research: Solid Earth*, *120*(6), 4263–4277, doi:10.1002/2015JB011984.

Abercrombie, R. E., S. Bannister, J. Ristau, and D. Doser (2017), Variability of earthquake stress drop in a subduction setting, the Hikurangi Margin, New Zealand, *Geophysical Journal International*, *208*(1), 306–320, doi:10.1093/gji/ggw393.

Agurto-Detzel, H., M. Bianchi, G. A. Prieto, and M. Assumpcao (2017), Earthquake source properties of a shallow induced seismic sequence in SE Brazil, *Journal of Geophysical Research: Solid Earth*, *122*(4), 2016JB013,623, doi:10.1002/2016JB013623.

Aki, K. (1967), Scaling law of seismic spectrum, *Journal of Geophysical Research*, *72*(4), 1217–1231, doi:10.1029/JZ072i004p01217.

Atkinson, G. M., and K. Assatourians (2017), Are Ground-Motion Models Derived from Natural Events Applicable to the Estimation of Expected Motions for Induced Earthquakes?, *Seismological Research Letters*, *88*(2A), 430–441, doi:10.1785/0220160153.

Atkinson, G. M., and M. Morrison (2009), Observations on Regional Variability in Ground-Motion Amplitudes for Small-to-Moderate Earthquakes in North America, *Bulletin of the Seismological Society of America*, *99*(4), 2393–2409, doi:10.1785/0120080223.

Atkinson, G. M., E. Yenier, N. Sharma, and V. Convertito (2016), Constraints on the Near-Distance Saturation of Ground-Motion Amplitudes for Small-to-Moderate Induced Earthquakes, *Bulletin of the Seismological Society of America*, *106*(5), 2104–2111, doi:10.1785/0120160075.

Baars, D., and W. Watney (1991), Paleotectonic control of reservoir facies, in *Sedimentary Modeling: Computer Simulations and Methods for Improved Parameter Definition*, vol. 0097-4471, pp. 253–262, Kansas Geological Survey Bulletin, Lawrence, KS.

Baltay, A. S., T. C. Hanks, and G. C. Beroza (2013), Stable Stress-Drop Measurements and their Variability: Implications for Ground-Motion Prediction, *Bulletin of the Seismological Society of America*, *103*(1), 211–222, doi:10.1785/0120120161.

Baltay, A. S., T. C. Hanks, and N. A. Abrahamson (2017), Uncertainty, Variability, and Earthquake Physics in Ground-Motion Prediction Equations, *Bulletin of the Seismological Society of America*, *107*(4), 1754–1772, doi:10.1785/0120160164.

Barbour, A. J., J. H. Norbeck, and J. L. Rubinstein (2017), The Effects of Varying Injection Rates in Osage County, Oklahoma, on the 2016 Mw 5.8 Pawnee Earthquake, *Seismological Research Letters*, *88*(4), 1040–1053, doi: 10.1785/0220170003.

Ben-Zion, Y., and L. Zhu (2002), Potency-magnitude scaling relations for southern California earthquakes with 1.0<ML<7.0, *Geophysical Journal International*, *148*(3), F1–F5, doi:10.1046/j.1365-246X.2002.01637.x.

Boore, D. M. (2003), Simulation of Ground Motion Using the Stochastic Method, in *Seismic Motion, Lithospheric Structures, Earthquake and Volcanic Sources: The Keiiti Aki Volume*, edited by Y. Ben-Zion, Pageoph Topical Volumes, pp. 635–676, Birkhauser Basel, doi:10.1007/978-3-0348-8010-7_10.

Boyd, O. S., D. E. McNamara, S. Hartzell, and G. Choy (2017), Influence of Lithostatic Stress on Earthquake Stress Drops in North America, *Bulletin of the Seismological Society of America*, *107*(2), 856–868, doi:10.1785/0120160219.

Brune, J. N. (1970), Tectonic stress and the spectra of seismic shear waves from earthquakes, *Journal of Geophysical Research*, *75*(26), 4997–5009, doi:10.1029/JB075i026p04997.

Buchanan, R. C. (2015), Increased seismicity in Kansas, *The Leading Edge*, *34*(6), 614–617, doi:10.1190/tle34060614.1.

Buchanan, R. C., K. D. Newell, C. S. Evans, and R. D. Miller (2014), Induced Seismicity: The Potential for Triggered Earthquakes in Kansas, *Tech. Rep. 36*, Kansas Geological Survey, Lawrence, KS.

Calderoni, G., A. Rovelli, and S. K. Singh (2013), Stress drop and source scaling of the 2009 April L'Aquila earthquakes, *Geophysical Journal International*, *192*(1), 260–274, doi:10.1093/gji/ggs011.

Choy, G. L., J. L. Rubinstein, W. L. Yeck, D. E. McNamara, C. S. Mueller, and O. S. Boyd (2016), A Rare Moderate-Sized (Mw 4.9) Earthquake in Kansas: Rupture Process of the Milan, Kansas, Earthquake of 12 November 2014 and Its Relationship to Fluid Injection, *Seismological Research Letters*, *87*(6), 1433–1441, doi:10.1785/0220160100.

Cramer, C. H. (2017), Brune Stress Parameter Estimates for the 2016 Mw 5.8 Pawnee and Other Oklahoma Earthquakes, *Seismological Research Letters*, *88*(4), 1005–1016, doi:10.1785/0220160224.

Douglas, J., and B. Edwards (2016), Recent and future developments in earthquake ground motion estimation, *Earth-Science Reviews*, *160*, 203–219, doi:10.1016/j.earscirev.2016.07.005.

Edwards, B., and J. Douglas (2014), Magnitude scaling of induced earthquakes, *Geothermics*, *52*, 132–139, doi:10.1016/j.geothermics.2013.09.012.

Ellsworth, W., A. Llenos, A. McGarr, A. Michael, J. Rubinstein, C. Mueller, M. Petersen, and E. Calais (2015), Increasing seismicity in the U. S. midcontinent: Implications for earthquake hazard, *The Leading Edge*, *34*(6), 618–626, doi:10.1190/tle34060618.1.

Ellsworth, W. L. (2013), Injection-Induced Earthquakes, *Science*, *341*(6142), 1225,942–1225,942, doi:10.1126/science.1225942.

Goebel, T. (2015), A comparison of seismicity rates and fluid-injection operations in Oklahoma and California: Implications for crustal stresses, *The Leading Edge*, *34*(6), 640–648, doi:10.1190/tle34060640.1.

Goebel, T. H. W., M. Weingarten, X. Chen, J. Haffener, and E. E. Brodsky (2017), The 2016 Mw5.1 Fairview, Oklahoma earthquakes: Evidence for long-range poroelastic triggering at >40 km from fluid disposal wells, *Earth and Planetary Science Letters*, *472*, 50–61, doi:10.1016/j.epsl.2017.05.011.

Goertz-Allmann, B. P., A. Goertz, and S. Wiemer (2011), Stress drop variations of induced earthquakes at the Basel geothermal site, *Geophysical Research Letters*, *38*(9), doi:10.1029/2011GL047498.

Hardebeck, J. L., and A. Aron (2009), Earthquake Stress Drops and Inferred Fault Strength on the Hayward Fault, East San Francisco Bay, California, *Bulletin of the Seismological Society of America*, *99*(3), 1801–1814, doi:10.1785/0120080242.

Hildebrand, G. M., D. W. Steeples, R. W. Knapp, R. D. Miller, and B. C. Bennett (1988), Microearthquakes in Kansas and Nebraska 1977-87, *Seismological Research Letters*, *59*(4), 159–163, doi:10.1785/gssrl.59.4.159.

Hough, S. E. (2014), Shaking from Injection-Induced Earthquakes in the Central and Eastern United States, *Bulletin of the Seismological Society of America*, *104*(5), 2619–2626, doi:10.1785/0120140099.

Huang, Y., G. C. Beroza, and W. L. Ellsworth (2016), Stress drop estimates of potentially induced earthquakes in the Guy-Greenbrier sequence, *Journal of Geophysical Research: Solid Earth*, *121*(9), 6597–6607, doi:10.1002/2016JB013067.

Huang, Y., W. L. Ellsworth, and G. C. Beroza (2017), Stress drops of induced and tectonic earthquakes in the central United States are indistinguishable, *Science Advances*, *3*(8), e1700,772, doi:10.1126/sciadv.1700772.

Ide, S. (2003), Apparent break in earthquake scaling due to path and site effects on deep borehole recordings, *Journal of Geophysical Research*, *108*(B5), doi: 10.1029/2001JB001617.

Izutani, Y., and H. Kanamori (2001), Scale-dependence of seismic energy-to-moment ratio for strike-slip earthquakes in Japan, *Geophysical Research Letters*, *28*(20), 4007–4010, doi:10.1029/2001GL013402.

Kaneko, Y., and P. M. Shearer (2014), Seismic source spectra and estimated stress drop derived from cohesive-zone models of circular subshear rupture, *Geophysical Journal International*, *197*(2), 1002–1015, doi:10.1093/gji/ggu030.

Keranen, K. M., H. M. Savage, G. A. Abers, and E. S. Cochran (2013), Potentially induced earthquakes in Oklahoma, USA: Links between wastewater injection and the 2011 Mw 5.7 earthquake sequence, *Geology*, *41*(6), 699–702, doi:10. 1130/G34045.1.

Keranen, K. M., M. Weingarten, G. A. Abers, B. A. Bekins, and S. Ge (2014), Sharp increase in central Oklahoma seismicity since 2008 induced by massive wastewater injection, *Science*, *345*(6195), 448–451, doi:10.1126/science.1255802.

Kroll, K. A., E. S. Cochran, and K. E. Murray (2017), Poroelastic Properties of the Arbuckle Group in Oklahoma Derived from Well Fluid Level Response to the 3 September 2016 Mw 5.8 Pawnee and 7 November 2016 Mw 5.0 Cushing Earthquakes, *Seismological Research Letters*, *88*(4), 963–970, doi:10.1785/0220160228.

Madariaga, R. (1976), Dynamics of an expanding circular fault, *Bulletin of the Seismological Society of America*, *66*(3), 639–666.

Mayeda, K., and W. R. Walter (1996), Moment, energy, stress drop, and source spectra of western United States earthquakes from regional coda envelopes, *Journal of Geophysical Research: Solid Earth*, *101*(B5), 11,195–11,208, doi: 10.1029/96JB00112.

Mayeda, K., R. Gok, W. R. Walter, and A. Hofstetter (2005), Evidence for non-constant energy/moment scaling from coda-derived source spectra, *Geophysical Research Letters*, *32*(10), L10,306, doi:10.1029/2005GL022405.

Mayeda, K., L. Malagnini, and W. R. Walter (2007), A new spectral ratio method using narrow band coda envelopes: Evidence for non-self-similarity in the Hector Mine sequence, *Geophysical Research Letters*, *34*(11), L11,303, doi: 10.1029/2007GL030041.

McBee, W. (2003), The Nemaha and other strike-slip faults in the midcontinent USA, in *AAPG Mid-Continent Section Meeting Proceedings*, pp. 1–23, Tulsa, OK.

Mori, J., R. E. Abercrombie, and H. Kanamori (2003), Stress drops and radiated energies of aftershocks of the 1994 Northridge, California, earthquake, *Journal of Geophysical Research: Solid Earth*, *108*(B11), 2545, doi:10.1029/2001JB000474.

Munafo, I., L. Malagnini, and L. Chiaraluce (2016), On the Relationship between Mw and ML for Small Earthquakes, *Bulletin of the Seismological Society of America*, *106*(5), 2402–2408, doi:10.1785/0120160130.

Niemi, T. M. (2004), Investigation of Microearthquakes, Macroseismic Data, and Liquefaction Associated with the 1867 Wamego Earthquake in Eastern Kansas, *Bulletin of the Seismological Society of America*, *94*(6), 2317–2329, doi:10.1785/0120030101.

Pacor, F., D. Spallarossa, A. Oth, L. Luzi, R. Puglia, L. Cantore, A. Mercuri, M. D'Amico, and D. Bindi (2016), Spectral models for ground motion prediction in the L'Aquila region (central Italy): evidence for stress-drop dependence on magnitude and depth, *Geophysical Journal International*, *204*(2), 697–718, doi:10.1093/gji/ggv448.

Petersen, M. D., C. S. Mueller, M. P. Moschetti, S. M. Hoover, A. L. Llenos, W. L. Ellsworth, A. J. Michael, J. L. Rubinstein, A. F. McGarr, and K. S. Rukstales (2016), Seismic-Hazard Forecast for 2016 Including Induced and Natural Earthquakes in the Central and Eastern United States, *Seismological Research Letters*, *87*(6), 1327–1341, doi:10.1785/0220160072.

Petersen, M. D., C. S. Mueller, M. P. Moschetti, S. M. Hoover, A. M. Shumway, D. E. McNamara, R. A. Williams, A. L. Llenos, W. L. Ellsworth, A. J. Michael, J. L. Rubinstein, A. F. McGarr, and K. S. Rukstales (2017), 2017 One-Year Seismic-Hazard Forecast for the Central and Eastern United States from Induced and Natural Earthquakes, *Seismological Research Letters*, *88*(3), 772–783, doi:10.1785/0220170005.

Ross, Z. E., and Y. Ben-Zion (2016), Toward reliable automated estimates of earthquake source properties from body wave spectra, *Journal of Geophysical Research: Solid Earth*, *121*(6), 2016JB013,003, doi:10.1002/2016JB013003.

Ross, Z. E., Y. Ben-Zion, M. C. White, and F. L. Vernon (2016), Analysis of earthquake body wave spectra for potency and magnitude values: implications for magnitude scaling relations, *Geophysical Journal International*, *207*(2), 1158–1164, doi:10.1093/gji/ggw327.

Rubinstein, J., W. Ellsworth, A. Llenos, and S. Walter (2014), Is the recent increase in seismicity in southern Kansas natural?, in *AGU Fall Meeting Abstracts*, pp. S53E–08.

Rubinstein, J. L., and A. B. Mahani (2015), Myths and Facts on Wastewater Injection, Hydraulic Fracturing, Enhanced Oil Recovery, and Induced Seismicity, *Seismological Research Letters*, *86*(4), 1060–1067, doi:10.1785/0220150067.

Rubinstein, J. L., F. M. Terra, and W. L. Ellsworth (2015), Wastewater Disposal, Hydraulic Fracturing, and Seismicity in Southern Kansas, in *AGU Fall Meeting Abstracts*, pp. S22A–01.

Schoenball, M., N. C. Davatzes, and J. M. G. Glen (2015), Differentiating induced and natural seismicity using space-time-magnitude statistics applied to the Coso Geothermal field, *Geophysical Research Letters*, *42*(15), 2015GL064,772, doi:10.1002/2015GL064772.

Segall, P., and S. Lu (2015), Injection-induced seismicity: Poroelastic and earthquake nucleation effects, *Journal of Geophysical Research: Solid Earth*, *120*(7), 2015JB012,060, doi:10.1002/2015JB012060.

Shearer, P. M., G. A. Prieto, and E. Hauksson (2006), Comprehensive analysis of earthquake source spectra in southern California, *Journal of Geophysical Research*, *111*(B6), doi:10.1029/2005JB003979.

Steeples, D. W., S. M. DuBois, and F. W. Wilson (1979), Seismicity, faulting, and geophysical anomalies in Nemaha County, Kansas: Relationship to regional structures, *Geology*, *7*(3), 134–138.

Sumy, D. F., C. J. Neighbors, E. S. Cochran, and K. M. Keranen (2017), Low stress drops observed for aftershocks of the 2011 Mw 5.7 Prague, Oklahoma, earthquake, *Journal of Geophysical Research: Solid Earth*, *122*(5), 2016JB013,153, doi:10.1002/2016JB013153.

Takahashi, T., H. Sato, M. Ohtake, and K. Obara (2005), Scale Dependence of Apparent Stress for Earthquakes along the Subducting Pacific Plate in Northeastern Honshu, Japan, *Bulletin of the Seismological Society of America*, *95*(4), 1334–1345, doi:10.1785/0120040075.

Trugman, D. T., and P. M. Shearer (2017a), Application of an improved spectral decomposition method to examine earthquake source scaling in southern California, *Journal of Geophysical Research: Solid Earth*, *122*(4), 2017JB013,971, doi:10.1002/2017JB013971.

Trugman, D. T., and P. M. Shearer (2017b), GrowClust: A Hierarchical Clustering Algorithm for Relative Earthquake Relocation, with Application to the Spanish Springs and Sheldon, Nevada, Earthquake Sequences, *Seismological Research Letters*, *88*(2A), 379–391, doi:10.1785/0220160188.

Walsh, F. R., and M. D. Zoback (2015), Oklahoma's recent earthquakes and saltwater disposal, *Science Advances*, *1*(5), e1500,195–e1500,195, doi:10.1126/sciadv.1500195.

Walter, W. R., S.-H. Yoo, K. Mayeda, and R. Gok (2017), Earthquake stress via event ratio levels: Application to the 2011 and 2016 Oklahoma seismic sequences, *Geophysical Research Letters*, *44*(7), 3147–3155, doi:10.1002/2016GL072348.

Weingarten, M., S. Ge, J. W. Godt, B. A. Bekins, and J. L. Rubinstein (2015), High-rate injection is associated with the increase in U.S. mid-continent seismicity, *Science*, *348*(6241), 1336–1340, doi:10.1126/science.aab1345.

Yeck, W. L., G. P. Hayes, D. E. McNamara, J. L. Rubinstein, W. D. Barnhart, P. S. Earle, and H. M. Benz (2017), Oklahoma experiences largest earthquake during ongoing regional wastewater injection hazard mitigation efforts, *Geophysical Research Letters*, *44*(2), 711–717, doi:10.1002/2016GL071685.

Yenier, E., and G. M. Atkinson (2014), Equivalent Point-Source Modeling of Moderate-to-Large Magnitude Earthquakes and Associated Ground-Motion Saturation Effects, *Bulletin of the Seismological Society of America*, *104*(3), 1458–1478, doi:10.1785/0120130147.

Yenier, E., G. M. Atkinson, and D. F. Sumy (2017), Ground Motions for Induced Earthquakes in Oklahoma, *Bulletin of the Seismological Society of America*, *107*(1), 198–215, doi:10.1785/0120160114.

Zaliapin, I., and Y. Ben-Zion (2013), Earthquake clusters in southern California I: Identification and stability, *Journal of Geophysical Research: Solid Earth*, *118*(6), 2847–2864, doi:10.1002/jgrb.50179.

Zaliapin, I., and Y. Ben-Zion (2016), Discriminating Characteristics of Tectonic and Human-Induced Seismicity, *Bulletin of the Seismological Society of America*, *106*(3), 846–859, doi:10.1785/0120150211.

Zhang, H., D. W. Eaton, G. Li, Y. Liu, and R. M. Harrington (2016), Discriminating induced seismicity from natural earthquakes using moment tensors and source spectra, *Journal of Geophysical Research: Solid Earth*, *121*(2), 972–993, doi:10.1002/2015JB012603.

# Chapter 7

# Strong correlation between stress drop and peak ground acceleration for recent seismicity in the San Francisco Bay Area

## Abstract

Theoretical and observational studies have suggested that between-event variability in the median ground motions of larger ($M \geq 5$) earthquakes is controlled primarily by the dynamic properties of the earthquake source, such as Brune-type stress drop. Analogous results remain equivocal for smaller events due to the lack of comprehensive and overlapping ground motion and source parameter

datasets in this regime. Here we investigate the relationship between peak ground acceleration (PGA) and dynamic stress drop for a new dataset of 5297 $M \geq 1.5$ events occurring in the San Francisco Bay Area from 2002 through 2016. For each event, we measure peak ground acceleration on horizontal-component channels of stations within 100km distance and estimate stress drop from $P$-wave spectra recorded on vertical-component channels from the same stations. We then develop a nonparametric ground motion prediction equation using a mixed-effects generalization of the Random Forest algorithm that we use to model the joint influence of magnitude, distance, and near-site effects on observed PGA. We find a strong correlation between dynamic stress drop and the residual PGA of each event, with the events with higher-than-expected PGA associated with higher values of stress drop. The strength of this correlation increases as a function of magnitude but remains significant even for smaller magnitude events with corner frequencies that approach the observable bandwidth of the acceleration records. Mainshock events are characterized by systematically higher stress drop and PGA than aftershocks of equivalent magnitude. Coherent local variations in the distribution of dynamic stress drop provide observational constraints to support the future development of nonergodic ground motion prediction equations that account for variations in median stress drop at different source locations.

## 7.1 Introduction

The intensity of earthquake-generated ground motion depends on a complex interaction of source, path, and site effects. Ground motion prediction equations (GMPEs), in which observed ground motion amplitudes are statistically modeled as a function of magnitude, source-site distance, and other auxiliary factors, have long served as an empirical basis for understanding the various features that most strongly influence ground motion intensity. Although modern GMPEs have been quite successful in modeling the first-order influence of magnitude scaling and distance-dependent geometrical spreading and attenuation (*Douglas*, 2003), the modulating influence of source complexity on earthquake ground motion remains an area of active research that is of fundamental importance to the scientific understanding of earthquake rupture and the practical implementation of probabilistic seismic hazard assessment (*Bozorgnia et al.*, 2014; *Douglas and Edwards*, 2016; *Baltay et al.*, 2017). Probabilistic hazard forecasts are particularly sensitive to the between-event variability in the predicted ground motion, as the hazard curve for the time horizons of interest to the development of building codes is dominated by the occurrence of rare events with ground motions in the tails of the inferred probability distribution (*Anderson and Brune*, 1999; *Bommer and Abrahamson*, 2006; *Cotton et al.*, 2013; *D'Amico et al.*, 2017).

From a theoretical perspective, one of the most important features driving between-event variability in ground motion is variability in the dynamic stress drop

(or stress parameter) of the events in question (e.g., *Atkinson*, 1990; *Atkinson and Morrison*, 2009; *Baltay et al.*, 2013; *Baltay and Hanks*, 2014; *Yenier and Atkinson*, 2014). The dynamic stress drop as formulated by *Brune* (1970, 1971) can be estimated entirely from the seismically-observed spectra of waveforms recording an earthquake, and is typically used as proxy for the relative proportion of high-frequency energy radiated by the earthquake source during rupture (*Atkinson and Beresnev*, 1997). For larger ($M \geq 5$) earthquakes, finite fault effects cause ground motion amplitudes to saturate with increasing magnitude, and thus the near-source ground motion amplitude is controlled primarily by stress drop (*Baltay and Hanks*, 2014).

Numerous observational and theoretical studies have validated this basic correlation between stress drop and ground motion (e.g., *Hanks* 1979; *Hanks and McGuire* 1981; *Boatwright* 1982; *Boore* 1983, 2003; *Baltay et al.* 2013; *Yenier and Atkinson* 2015; *Lior and Ziv* 2017; *Oth et al.* 2017). However, the precise functional form of this relation remains poorly understood, and in particular the ways in which the relative influence of stress drop varies in response to other interacting factors such as magnitude, distance, depth, and source region. Progress towards resolving these questions may provide a considerable step forward in reducing the epistemic uncertainties associated with ground motion prediction (*Anderson and Brune*, 1999). However, exploring the influence of stress drop requires careful analyses of extensive joint datasets of ground motion recordings and source parameter estimates, which historically have been produced independently and for sparsely

overlapping earthquake sequences with few events in common.

In this study, we focus on the relationship between peak ground acceleration (PGA) and dynamic stress drop for a new dataset of more the 5000 earthquakes occurring in the vicinity of the San Francisco Bay Area, CA, from 2002 through 2016 (Figure 7.1). We use the spectral decomposition technique of *Trugman and Shearer* (2017) to derive dynamic stress drop estimates from $P$-wave spectra recorded on vertical component channels of broadband and short-period stations. We then measure PGA using the full waveform, horizontal component records of these events at the same set of stations. Analysis of between-event variability in ground motion requires a reference GMPE that models the scaling of ground motion intensity as a function of magnitude and external features unrelated to the source, such as distance and local site effects. The set of existing GMPEs designed for earthquakes in California were derived from records of earthquakes that sample different source regions and magnitude ranges than those that comprise our dataset, and hence would be inappropriate to extrapolate for this purpose.

We instead apply a novel, data-driven approach based on the Random Forest algorithm (*Breiman*, 2001) to derive a nonparametric GMPE that can be used to correct the observed ground motion amplitudes for the interacting effects of geometrical spreading, attenuation, magnitude scaling, and near-site effects. This approach has the advantage over techniques based on linear regression in that it can account for complex interactions between these features without risk of overfitting observational noise within the data or introducing systematic trends with

magnitude or distance into the residuals from the model prediction (e.g. *Bindi* 2017). We use this framework to define a between-event residual term that quantifies the empirical contribution of each event to the observed ground motion. The unbiased precision of the Random Forest GMPE is particularly useful in this context as it permits a detailed examination of the physical correspondence between PGA and the dynamic properties of the earthquake source.

## 7.2 Study Region and Waveform Data

We examine earthquakes occurring from 2002 through 2016 within a region encompassing the San Francisco Bay Area, California (latitude and longitude bounds of [37.0, 38.5] and [-123.0, -121.5], respectively). This region is of particular interest for earthquake hazard analysis due to its active seismicity and high population density. Major fault systems within this region include the San Andreas Fault near the San Jose and San Francisco metro areas and the Hayward and Calaveras faults in the Oakland/East Bay metro area (*Waldhauser and Ellsworth*, 2002; *Hardebeck et al.*, 2007; *Hardebeck and Aron*, 2009). Numerous other subparallel faults of varying seismicity rate also strike through this region (*Field et al.*, 2014). The 24 August 2014 $M_W$6.0 South Napa earthquake is the largest earthquake occurring near the Bay Area during our study period (*Brocher et al.*, 2015; *Dreger et al.*, 2015; *Ji et al.*, 2015; *Wei et al.*, 2015), and it triggered hundreds of aftershocks throughout the West Napa fault system and surrounding region over

the subsequent months (*Hardebeck and Shelly*, 2016; *Llenos and Michael*, 2017).

Earthquakes occurring in the San Francisco Bay Area during our study period are well-recorded by the Northern California Seismic Network (NCSN). We obtain NCSN waveforms archived by the Northern California Earthquake Data Center (NCEDC) using the eventdata query system (see Data and Resources Section). We restrict our analysis to the subset of earthquakes with catalog magnitude (typically $M_D$) of 1.5 and greater that have been relocated using waveform cross correlation based techniques (*Waldhauser and Schaff*, 2008). *P*-wave spectral estimates are derived from the vertical component HN, EH, and HH channels at stations with an epicentral distance from the source of less than 100km. PGA estimates are derived from the geometric mean of the horizontal components of the same set of stations. The eventdata query system provides waveform data in miniSEED format, which we then convert to SAC format in order to use the SAC waveform processing subroutines (see Data and Resources Section).

## 7.3 Methods

### 7.3.1 Source spectral analysis and dynamic stress drop

Measurements of the dynamic properties of the earthquake source, including the dynamic stress drop, $\Delta\sigma$, are derived from spectra of earthquake waveforms. We obtain estimates of $\Delta\sigma$ for the earthquakes in our study region using the spectral decomposition approach described in detail by *Trugman and Shearer* (2017),

and summarize only the essential points of the algorithm here. As discussed by *Atkinson and Beresnev* (1997) and others, it is important to understand that the dynamically-measured $\Delta\sigma$ based on the *Brune* (1970) model of earthquake source spectra may differ from the true, static stress release on the fault plane, and hence the dynamic $\Delta\sigma$ is better conceptualized as a measure of the frequency content of the earthquake source. For this reason, $\Delta\sigma$ is often referred to as the stress parameter, rather than stress drop, in the ground motion and earthquake hazard literature, but we continue to use the term stress drop for consistency with previous earthquake source studies.

The conceptual framework underlying the spectral decomposition method can be understood by noting that for sufficiently large and well-recorded waveform datasets such as that of our Bay Area study region, each earthquake is recorded by many stations, each station records many earthquakes, and each source-receiver raypath is sampled (to good approximation) many times. Thus, each of the three main contributions to the observed waveform spectra – the source, the path, and the site – have sufficient observational constraints to be resolved as part of an overdetermined inverse problem. Consider the waveform spectra $d_{ij}(f)$ of event $i$ recorded at station $j$, which in general is a convolutional product of source, site, and path effects. In the log frequency domain, these three contributions are additive, and thus the $d_{ij}$ at a given frequency $f$ can be decomposed into a linear equation of the form

$$d_{ij} = e_i + st_j + tt_{k(i,j)} + r_{ij}, \tag{7.1}$$

where $e_i$ and $st_j$ denote the relative contribution of the source and station, $tt_{k(i,j)}$ is a travel-time dependent path term that we approximate as isotropic and parameterize into $k$ discrete travel time bins of 0.5s width, and $r_{ij}$ is a residual error term.

We apply spectral decomposition to the earthquakes in our Bay Area dataset as follows. For each event, we compute $P$-wave spectra from the vertical component of high broadband and short period channels (HN, HH, and EH, with a preference for HN when available) of all NCSN stations within 100km distance. These spectral estimates are obtained using a multitaper algorithm (*Park et al.*, 1987; *Prieto et al.*, 2009) and a temporal window that begins 0.05s before the catalog-listed $P$-phase arrival time. Each temporal window bracketing the $P$-phase arrival has a window length that ranges from a minimum of 1.5s to a maximum of 4.5s, with longer windows applied to larger magnitude events (*Ross et al.*, 2016; *Abercrombie et al.*, 2017). We discard clipped waveforms and those with obvious noise spikes using an automatic detection algorithm based on the observed amplitude distribution (*Trugman and Shearer*, 2017). We ensure adequate data quality and station coverage by only considering events recorded by at least six stations in which the average signal-to-noise amplitude is greater than 3 in each of five frequency bands: 2.5–6, 6–10, 10–15, 15–20, and 20–25Hz. We also limit our analyses to NCSN stations that have recorded at least 50 different events.

We convert the remaining, quality-controlled spectra to units of displacement and resample to a uniform frequency grid between 2.5 and 25Hz. We then

solve equation (7.1) using an iterative, robust least-squares inversion algorithm with outlier suppression. This spectral decomposition is effective in isolating the relative contributions of source, path, and site at each frequency point, not their absolute values. In order to yield source parameter estimates compatible with the *Brune* [1970] model, it is therefore necessary to apply an empirical correction term that represents path and site effects that are common to all sources, including the average near-source and near-receiver attenuation (relative differences in average site attenuation are incorporated into the station terms, $st_j$, themselves). To do so, we use the approach described by *Trugman and Shearer* (2017) that infers the optimal empirical correction by fitting stacked relative source spectra (Figure 7.2a), averaged in bins of spectral moment, to a Brune-type theoretical spectrum of the form:

$$\hat{s}(\, f \mid \Omega_0, f_c, n) = \frac{\Omega_0}{1 + (f/f_c)^n}, \tag{7.2}$$

where $\Omega_0$ is the long-period spectral moment of each stack (and is proportional to seismic moment), $f_c$ is the best-fitting corner frequency, and $n$ is the high-frequency falloff rate that is fixed to 2 per the canonical $\omega^{-2}$ spectral model (*Aki*, 1967; *Brune*, 1970). This technique differs from the stacking approach of *Shearer et al.* (2006) in that it does not presume self-similar scaling of the stacked spectra (Figure 7.2b). A notable advantage of this technique is that the inference of the empirical correction term is based entirely upon the shape of the stacked spectra within the 2.5–25 Hz frequency band in which we have ensured adequate signal-to-noise, and

does not require resolving corner frequencies of individual events. We assume a spatially uniform correction term for the results presented in this study, which is justified based on sensitivity tests in which we found that permitting smooth lateral variations in the inferred correction term produces statistically insignificant differences in the final results.

The corrected source spectra are used to estimate two source parameters – seismic moment $M_0$ and Brune corner frequency $f_c$ – which are required to compute the dynamic stress drop (*Brune*, 1970; *Madariaga*, 1976),

$$\Delta\sigma = \frac{7}{16} M_0 \left( \frac{f_c}{k\beta} \right)^3, \tag{7.3}$$

where $\beta$ is the shear wave speed and $k$ is a numerical factor that we set to 0.38 following *Kaneko and Shearer* (2014). $M_0$ cannot in general be obtained directly from the NCSN catalog, as listed magnitudes $M$ are of mixed type but are typically duration magnitude $M_D$ for the smaller ($M < 3.5$) events for which moment tensors are not routinely estimated. To obtain a uniform set of $M_0$ (and hence $M_W$) estimates for our dataset, we follow the approach of *Shearer et al.* (2006), which uses a regression analysis between the observed spectral moment $\Omega_0$ and the catalog-listed $M$ under the assumption that on average $M_0$ should be proportional to $\Omega_0$. Our estimates of $M_0$ and $M_W$ are thus constrained primarily by the spectral amplitude at long periods, rather than the catalog-listed $M$ values. We do however note that this approach can be used to obtain an average linear relationship between $M_W$ and $M_D$ with a slope that is identical to within uncertainties to that

inferred by *Hawthorne et al.* (2016), though with significant differences from the best-fit line for individual events (Figure 7.S1 in the supplement to this article).

We obtain $f_c$ estimates for each event using a bounded optimization algorithm that minimizes the root-mean-square residual between the corrected relative source spectrum $s_i(f)$ and the Brune theoretical spectrum $\hat{s}_i(f|f_c)$ in the 2.5–25Hz band in which we have measured adequate signal-to-noise. To compute $\Delta\sigma$ given $M_0$ and $f_c$, we use a fixed shear wave speed $\beta$ in equation (7.3) for consistency with previous studies of the influence of $\Delta\sigma$ on ground motion (*Atkinson and Morrison*, 2009; *Bozorgnia et al.*, 2014; *Baltay et al.*, 2015), and as such $\Delta\sigma$ should be interpreted as a measure of relative frequency content rather than an approximation of static stress release. Lastly, we derive uncertainty estimates for $M_0$, $f_c$, and $\Delta\sigma$ using a bootstrap resampling procedure (*Trugman and Shearer*, 2017) that assesses the variability in the measured source spectra at each station.

### 7.3.2 Peak ground acceleration and the Random Forest GMPE

As our objective is to analyze the extent to which $\Delta\sigma$ influences ground motion intensity (PGA), we require: (i) measurements of PGA at the stations recording the events comprising our dataset, and (ii) a model of how median PGA values should be expected to vary as a function of distance, magnitude, site, and other effects unrelated to the dynamic properties of the earthquake source. In

this section, we describe both our procedure to obtain PGA measurements for this dataset, and our approach to model PGA using a data-driven GMPE based on a Random Forest regression algorithm.

Although the Next Generation Attenuation West 2 (NGA-W2) database (*Bozorgnia et al.*, 2014) includes high-quality ground motion records of a small subset of the events in our dataset, most of the events we consider are unlisted. To compile a complete and self-consistent PGA database, we compute PGA as follows. For each event, we consider horizontal-component records of stations within 100km source-station distance (consistent with the data selection for the $\Delta\sigma$ computations described in the previous subsection). Using the SAC waveform analysis software, we then demean and detrend each record, remove the instrument response, and convert to units of acceleration where necessary. This requires application of a bandpass filter, which we choose to implement as a relatively gentle filter with a lowpass transition band between 0.5 and 1.5 Hz, and highpass transition band between 30 and 40 Hz. The latter corresponds to a significantly higher upper corner than is used for the NGA-W2 database (*Bozorgnia et al.*, 2014), but this modification is necessary to analyze the lower magnitude events that comprise the majority of our dataset. We then use an automated procedure to discard records that have high-amplitude noise spikes or have low signal-to-noise RMS amplitudes relative to a pre-event time window. We also exclude a small subset of events with origin times that are nearly-overlapping (within 45s spacing of each other), as in these cases the attribution of peak ground motions to one event or another

becomes ambiguous. For the remaining quality-controlled records, we compute PGA from the geometric mean of both horizontal components. For most events, the peak amplitude is from the $S$-wave or surface wave, whereas our dynamic $\Delta\sigma$ estimates are derived from $P$-waves recorded on the vertical component. Our PGA estimates are in good agreement with respect to records that are cross-listed in the NGA-W2 database (Figure 7.S2 in the supplement to this article).

Ground motion intensity measures such as PGA are typically modeled using GMPEs based on linear regression. Such GMPEs predict the expected level of ground motion $\log \hat{y}$ as a linear combination ($w$) of input features ($X$) that are thought to influence ground motion amplitudes:

$$\log \hat{y} = Xw \tag{7.4}$$

The feature matrix $X$ typically includes (but is not limited to) the magnitude (usually $M_W$) of the earthquake and a measure of the distance $R$ from the station to the source (usually epicentral distance, hypocentral distance, or finite fault generalizations such as the Joyner-Boore distance). In recent years, GMPEs based on linear regression have tended to become more complex in order to accommodate a larger selection of potential input features, higher-order (e.g., quadratic) terms, and the possibility of interactions as function of magnitude and distance (e.g. *Campbell and Bozorgnia* 2008; *Boore et al.* 2013; *Yenier and Atkinson* 2015).

Typically, the performance of the GMPE is measured in terms of its uncertainty, often denoted $\sigma$ and defined as the standard deviation of the distribution

of residuals between the observed and GMPE-predicted ground motion intensity (both measured in logarithmic units). Reduction in the uncertainty of the GMPE is appealing because it leads to improved accuracy in seismic hazard forecasts, especially over long time horizons. However, because $\sigma$ is measured with respect to the dataset used to fit the model, it is not a measure of the uncertainty with respect to future predictions (e.g., *Hastie et al.* 2009; *Murphy* 2012). Thus, a reduction in the total $\sigma$ through increased model complexity is only beneficial in cases in which the performance of a model that is developed on the input (or training) data set is validated by a corresponding decrease in $\sigma$ with respect to an independent testing data set (*Bindi*, 2017; *Mak et al.*, 2017). Further, GMPEs based on standard linear regression techniques may produce systematic residuals as a function of $M$ and $R$ if the true interaction between features deviates from the assumed linear model (*Bishop*, 2006), or if the number of observations varies significantly with $M$ or $R$, as is often the case in circumstances with limited data availability.

To study the relative influence of $\Delta\sigma$ on PGA for our study, we require a GMPE with the following properties:

(i) Is applicable to the ground motions of $M \geq 1.5$ earthquakes within the Bay Area and is valid to (at least) 100km hypocentral distance.

(ii) Can be used to correct the observed PGA for the potentially nonlinear influence of and interactions between magnitude, distance, and site, which we postulate are the primary features controlling ground motion within this regime

that are unrelated to dynamic earthquake source properties.

(iii) Does not produce systematic misfits as a function of magnitude or distance that may bias our interpretation of residual PGA in relation to $\Delta\sigma$.

(iv) Provides a comparable or better GMPE uncertainty $\sigma$ as compared to established linear regression techniques, as well as a means to assess whether this improved performance is generalizable to an independent dataset.

(v) Is robust with respect to outlier data points, which are more common for the recordings of the lower magnitude events that are an important part of our dataset.

With these requirements in mind, we develop a data-driven GMPE based on a statistical modeling technique known as a Random Forest (*Breiman*, 2001). While application of Random Forests to regression problems is well-established within the pattern recognition and machine learning community (*Bishop*, 2006; *Geurts et al.*, 2006; *Hastie et al.*, 2009), to our knowledge this is the first application to the modeling of earthquake ground motions. Random Forest regression models are particularly suitable for our purposes because of their capability to handle nonlinear feature influence and interaction, their robustness with respect to data outliers and codependent features, and their ability to automatically assess the predictive performance of the model through out-of-bag sampling (which we will describe in detail below) (*Louppe*, 2014; *Fernandez-Delgado et al.*, 2014).

Random Forests use an ensemble learning approach in which a set of randomized and decorrelated decision tree regressors are trained independently to predict ground motion amplitudes, with the final prediction of the Random Forest based on an ensemble average of the individual trees. The randomization and decorrelation of individual tree-based regressors helps reduce the variance in the ensemble prediction through averaging, and mitigates the tendency of individual trees to overfit their training dataset and hence provide poor generalization to independent testing data (*Breiman*, 2001; *Murphy*, 2012). In a Random Forest, each tree takes as input a bootstrap-resampling of the original training data set, and at each node in the decision tree, a random subset of the features is considered for partitioning. This multiscale randomization decorrelates the individual trees and contributes to their robustness, both of which are advantageous to the final ensemble model prediction.

Our implementation of the Random Forest GMPE makes use of the publicly available scikit-learn Python programming package (*Pedregosa et al.* 2011, see Data and Resources Section) with some GMPE-specific algorithm modifications that we describe below. Specifically, our input dataset consists of the measured PGA for the subset of well-recorded events for which we obtained estimates of $M_W$ and $\Delta\sigma$ (see Methods Section). We use a Random Forest to model the measured PGA values as a nonparametric function $G^{rf}$ of moment magnitude $M_W$ and hypocentral distance $\log R_{hyp}$, and treat the contribution of each event $\Delta PGA_i$ and station $\Delta S_j$

as random effect terms (e.g., *Abrahamson and Youngs* 1992; *Stafford* 2014):

$$\log \hat{y} = \frac{1}{B} \sum_{b=1}^{B} G_b^{rf}(M_W, \log R_{hyp}) + \Delta PGA_i + \Delta S_j. \tag{7.5}$$

The summation in equation (7.5) is an ensemble average of $B$ randomized regression trees that are fit to bootstrap-resampled datasets. We apply an iterative procedure to estimate the event and station terms $\Delta PGA_i$ and $\Delta S_j$ from the mean model residuals, and in practice the algorithm converges after only 2–3 iterations. Defined in this way, the $\Delta PGA_i$ values measure the empirical contribution of each event $i$ to the observed ground motion amplitude (after controlling for magnitude, source-receiver distance, and site), and the distribution of $\Delta PGA$ is a measure of between-event variability. It is important to note that to apply the Random Forest GMPE to predict ground motions for a new set of earthquakes, the station terms $\Delta S_j$ will be known but in general the event terms $\Delta PGA_i$ will be unknown a priori. In this case, the expected total variability in the ground motion from the model prediction,

$$\sigma^2 = E\left[(\log y - \log \hat{y})^2\right] = \tau^2 + \phi^2 \tag{7.6}$$

can be partitioned into between-event variability (variance $\tau^2$), defined by the distribution of $\Delta PGA$, and within-event variability (variance $\phi^2$), defined by the distribution of GMPE model residuals (e.g., *Atik et al.* 2010; *Baltay et al.* 2017).

The fitting procedure of each tree involves repeatedly partitioning the input data in feature space. We use scikit-learn's ExtraTreesRegressor algorithm (*Geurts et al.*, 2006) for this purpose to provide an additional layer of randomiza-

tion in the partitioning level that helps decorrelate the individual trees. At each node in a given tree, the algorithm randomly selects one of the two features ($M_W$ or $\log R_{hyp}$) for partitioning. Rather than optimize performance or misfit to the training data itself, we can use the Random Forest to synthesize an independent testing dataset without explicitly withholding a subset of input data. To do this, we take advantage of the fact that for each tree in the Random Forest, bootstrap resampling leaves out approximately one third of the input data. These excluded data points are known as out-of-bag samples, and each tree has a different subset of in-bag and out-of-bag samples. For a sufficiently large number of trees (we use $B = 200$ in this study), each data point will be left out-of-bag by multiple trees, where it can then be leveraged as part of an independent validation dataset with minimal computational effort. We can use this property to optimize predictive performance by constraining the maximum depth of the individual trees that comprise the Random Forest. Minimization of the misfit with respect to the set out-of-bag samples yields an optimal tree depth of 18 (Figure 7.S3 in the supplement to this article), which is itself a conservative choice that errs on the side of underfitting rather than overfitting due to the slight pessimistic bias in the out-of-bag misfit error compared to the true prediction error (*Breiman*, 2001; *Louppe*, 2014).

## 7.4 Results

In total, 5297 events within our study region are sufficiently well-recorded to meet our quality control criteria for measurements of $\Delta\sigma$ and PGA. Overall, we find the $\Delta\sigma$ estimates of these events closely follow a lognormal distribution with median value of 2.95 MPa and $\log_{10}$ standard deviation of 0.40. We observe a mild increase of $\Delta\sigma$ with $M_0$ (Figure 7.3), which we can quantify in terms of a scaling parameter $\epsilon_1$ that measures the slope of the increase in $\log_{10}\Delta\sigma$ with $\log_{10} M_0$:

$$\log_{10}\Delta\sigma = \epsilon_0 + \epsilon_1 \log_{10} M_0. \tag{7.7}$$

A weighted regression analysis based on median values in bins of width 0.4 in $\log_{10} M_0$ yields a scaling parameter $\epsilon_1 = 0.06$, which is consistent with the scaling inferred from the shape of the stacked spectra (Figure 7.2), but shallower in slope and hence closer to the self-similar value of $\epsilon_1 = 0$ than any of the five southern California study regions analyzed by *Trugman and Shearer* (2017).

Both $f_c$ and $\Delta\sigma$ increase slightly as a function of depth within the upper 8km of the crust (Figure 7.3cd), but we do not observe an analogous depth-dependence in $M_0$ (Figure 7.3e). These trends would be consistent with an increase in average rupture velocity with depth, although other systematic variations in rupture characteristics or fault geometry may also contribute. If we examine the spatial patterns of $\Delta\sigma$ in map view (Figure 7.4), we observe variations in median $\Delta\sigma$ on regional length scales, but significant coherence in $\Delta\sigma$ within more localized and spatially compact clusters of events. For example, $\Delta\sigma$ values are relatively high

along the peninsular San Andreas Fault and the southern portion of the Calaveras Fault to the south of its junction with the Hayward Fault. Median $\Delta\sigma$ values are markedly lower along the northern segment of the Calaveras Fault near the Mt. Diablo Thrust and the stepover with the Concord-Green Valley Fault. Seismicity along the Hayward and Rodgers Creek faults tends to exhibit intermediate $\Delta\sigma$ values, though with some systematic along-strike variations.

The GMPE modeling procedure described in Section 7.3.2 provides estimates of between-event residual ground motions ($\Delta PGA$) that measure the deviation between the observed and model-predicted ground motion amplitude. We observe a strong correlation between $\Delta\sigma$ and $\Delta PGA$ (Figure 7.5), which suggests that variations in dynamic source properties have a significant influence on the between-event variability in ground motion. The strength of this correlation increases as a function of magnitude, with a correlation coefficient of 0.46 for $M_W < 2.25$, compared to 0.62 for $2.25 < M_W < 2.50$ and 0.74 for $M_W > 2.50$. The magnitude-dependence of this correlation is expected, as the measured ground motion amplitude is influenced by attenuation along the raypath from source to site, which tends to mask or dampen the high-frequency energy typical of smaller earthquake sources (e.g, *Baltay and Hanks*, 2014). It is also notable that the variability in $\Delta PGA$ increases significantly with magnitude, with a $\log_{10}$ standard deviation in $\Delta PGA$ of 0.11 for $M_W < 2.25$ and 0.13 for $2.25 < M_W < 2.50$, compared to 0.18 for $2.50 < M_W < 3.00$ and 0.26 for $3.00 < M_W < 4.00$. Meanwhile, the variability in $\Delta\sigma$ is nearly constant with magnitude. We discuss this effect further in the

following section.

In map view, the spatial patterns of $\Delta PGA$ mirror those of $\Delta\sigma$ (Figure 7.4), with spatial coherence on local length scales, but significant variations in median values across the study region. The strength of the correlation remains unchanged whether one considers the absolute value of $\Delta\sigma$, or a magnitude-adjusted, relative stress drop (*Trugman and Shearer*, 2017) that corrects for the mild increase in $\Delta\sigma$ with $M_0$ (Figure 7.S4b in the supplement to this article). In contrast, the event term $\Delta PGA$ exhibits a rather weak relationship with source depth (Figure 7.S4c in the supplement to this article), except perhaps for the shallowest of source depths (0–3km) where rupture of weaker lithology may tend to reduce ground motion amplitudes. Likewise, the strong observed correlation between $\Delta\sigma$ and $\Delta PGA$ is not simply a result of the observed depth-dependence of $\Delta\sigma$, as the correlation remains nearly as strong (Figure 7.S4d in the supplement to this article) if we compute a depth-adjustment to our $\Delta\sigma$ estimates based on the median trend with depth shown in Figure 7.3d.

The Random Forest GMPE provides a consistently good fit to the observed distribution of PGA, and does not exhibit any significant biases with magnitude, distance, or site (Figure 7.6a). To get a sense for the performance of this model relative to linear GMPEs, we compare the Random Forest GMPE to two analogous, mixed-effects linear regression models (e.g., *Campbell and Bozorgnia* 2008; *Bindi*

*et al.* 2011; *Kurzon et al.* 2014) of the form:

$$\log \hat{y} = a_0 + a_1 M + a_2 \log R + \Delta PGA_i + \Delta S_j \quad (7.8)$$

$$\log \hat{y} = b_0 + (b_1 + b_2 M) \log R + b_3 R + b_4 M + b_5 M^2 + \Delta PGA_i + \Delta S_j \quad (7.9)$$

where $\{a_k\}$ and $\{b_k\}$ are linear regression model coefficients, and $\Delta PGA_i$ and $\Delta S_j$ are event and station random effect terms analogous to those in the Random Forest GMPE. Equation (7.8) is a relatively simple 1st-order linear regression model (Figure 7.6b) that accounts for the first-order influence of magnitude, geometrical spreading, and site, while the model described by equation (7.9) allows for more complexity (Figure 7.6c) through higher-order features, interaction terms, and a linear term in $R$ to account for attenuation. In both cases, the Random Forest GMPE provides a better fit to the data, both in terms of the total misfit $\sigma$ (0.634 for the Random Forest compared to 0.698 and 0.679 for the regression models), and its lack of systematic trends in misfit as a function of $M$ and $R$ (Figure 7.6). For this comparison, it is important to note that we have guarded against the potential for the Random Forest GMPE to overfit the data using the out-of-bag sampling procedure described in the Methods Section. While it may be possible to devise a more complex linear regression model to match its performance, this simple experiment is sufficient to motivate the use of the Random Forest GMPE for our purposes, and we discuss its potential application to future GMPE development in the following section.

From a practical perspective, one question of interest is whether on not

there are systematic differences in the distributions of $\Delta\sigma$ or $\Delta PGA$ between mainshock and aftershock events. Many established GMPE models either exclude aftershock events entirely from the model fit or explicitly include an adjustment factor that quantifies the difference in expected ground motion intensity of the two classes of events (*Douglas and Edwards*, 2016). However, distinguishing between mainshock events and their associated aftershocks is often difficult in practice, as exemplified by the wide range of plausible declustering algorithms designed to accomplish this task (*van Stiphout et al.*, 2012). The classical space-time windowing method of *Gardner and Knopoff* (1974) is widely used for this purpose in studies of large-magnitude mainshock-aftershock sequences, in part due to its simplicity and insensitivity to the absence of smaller events in a given ground motion database (*Wooddell and Abrahamson*, 2014). However, the *Gardner and Knopoff* (1974) declustering algorithm is not optimal for use in this study for two reasons. First, the formulas for the space-time windows are designed for larger magnitudes than those that comprise much of our dataset, and an extrapolation leads to poorly defined mainshock-aftershock classifications. Second, the *Gardner and Knopoff* (1974) cannot be used directly to group events into individual mainshock-aftershock sequences due to the nonuniqueness that occurs in cases of overlapping space-time windows.

We instead apply the magnitude-space-time nearest-neighbors method of (*Zaliapin and Ben-Zion*, 2013), which has been used effectively to characterize clustered seismicity in numerous previous studies. The nearest-neighbor algorithm

is more complicated and has more free parameters to consider than *Gardner and Knopoff* (1974) declustering, but is appropriate for the magnitude of our dataset and has the additional advantage that it can be used to define individual sequences of events in which the nearest neighbors are closely linked in distance-time-magnitude space. For the purposes of this study, we denote the largest event in each sequence as a mainshock (including those events within singleton sequences), and denote the remaining events as aftershocks. In applying the *Zaliapin and Ben-Zion* (2013) algorithm, we use the full set of $M \geq 1.1$ events within the relocated catalog of *Waldhauser and Schaff* (2008), rather than just the 5297 that comprise our database. Doing so helps mitigate potential clustering artifacts related to the edge effects of events outside but adjacent to our study region in space and time, as well as those within our study region but beneath the nominal completeness of our database.

If one compares the distribution of $\Delta\sigma$ for events classified as mainshocks and aftershocks (Figure 7.7a), it is apparent that the mainshocks exhibit systematically higher values of $\Delta\sigma$ than do aftershocks of equivalent moment. The difference between the median $\Delta\sigma$ values of mainshocks and aftershocks is slightly larger for high $M_0$ events than for low $M_0$ events, and indeed the two classes are virtually indistinguishable for the smallest events in our dataset. An analogous trend is observed for the between-event residual $\Delta PGA$, which is to be expected given the strong correspondence between $\Delta\sigma$ and $\Delta PGA$ (Figure 7.7b). These results are broadly consistent with previous studies examining the short-period ground motion

amplitudes of larger magnitude mainshocks and aftershocks (*Campbell and Bozorgnia*, 2008; *Boore et al.*, 2013; *Wooddell and Abrahamson*, 2014; *Yenier et al.*, 2017). The existing literature on possible mainshock versus aftershock differences in $\Delta\sigma$ is more equivocal, with various studies finding similar results to those presented here (e.g., *Izutani*, 2005; *Mayeda et al.*, 2007; *Sumy et al.*, 2017; *Cramer*, 2017; *Boyd et al.*, 2017), while others observe no such trend (e.g. *Allmann and Shearer*, 2009; *Viegas et al.*, 2010; *Baltay et al.*, 2013; *Abercrombie et al.*, 2017).

As noted above, the nearest-neighbors method of *Zaliapin and Ben-Zion* (2013) can be used to group events into individual earthquake sequences that are clustered in space and time. We use this framework to examine $\Delta\sigma$ and $\Delta PGA$ for the most prominent sequences in our dataset. We limit our analysis to sequences in which the largest event (the mainshock) has $M_W \geq 3.0$ and that contain at least ten other (aftershock) events with measured values of $\Delta\sigma$ and $\Delta PGA$. Mainshock $\Delta\sigma$ estimates are higher than the median value for all ten sequences (Figure 7.7c) that meet this criteria, while the same is true for $\Delta PGA$ for nine of the ten sequences (Figure 7.7d). Median sequence values of $\Delta\sigma$ and $\Delta PGA$ are strongly correlated, while the correlation for within-sequence variability is somewhat weaker but still significant. These observations suggest that analysis of the relation between source parameters and sequence-specific ground-motion data may warrant further consideration for Operational Earthquake Forecasting and real-time hazard assessment (e.g., *Jordan et al.* 2011; *Page et al.* 2016).

## 7.5 Discussion

In this study, we present evidence for a strong correlation between dynamic stress drop ($\Delta\sigma$) and residual ground motion ($\Delta PGA$). Although this correspondence has been previously documented for large earthquakes, our study brings to light several new aspects of this relation that have important implications for our understanding of how variability in earthquake source properties can influence earthquake hazard. First, we augment the existing literature with a new, jointly derived dataset of source parameter and ground motion estimates for contemporary seismicity in the San Francisco Bay Area (Table S1 in the electronic supplement to this article). The events included within this dataset sample a wide magnitude range and include smaller-magnitude events than is typical for ground motion studies, allowing us to examine how the relationship between $\Delta\sigma$ and between-event variability in PGA varies as a function of magnitude. In so doing, we find that the observed correlation between $\Delta\sigma$ and PGA weakens with decreasing magnitude, but is still apparent for the smallest events in our dataset.

The relation between $\Delta\sigma$ and PGA has been studied extensively using theoretical point source models and random vibration theory (*Boore*, 1983, 2003; *Baltay and Hanks*, 2014; *Yenier and Atkinson*, 2014) in which the acceleration spectra are typically assumed to be Brune-type and thus flat within the frequency band $f_c < f < f_{max}$. In this model, the parameter $f_{max}$ is used as a proxy for the effects of attenuation, especially in the shallow subsurface near the site (*Hanks*,

1982; *Anderson and Hough*, 1984). The observed increase in the strength of the correlation between $\Delta\sigma$ and $\Delta PGA$ with increasing $M$ is consistent with the predictions of this model, as smaller $M$ events have higher $f_c$ on average and thus less bandwidth between $f_c$ and $f_{max}$. However, the fact that we still observe a significant correlation between $\Delta\sigma$ and $\Delta PGA$ even for the smallest events in our dataset is somewhat unexpected, and is perhaps suggestive of the limitations in validity of these theoretical models of acceleration spectra within this regime.

Studies of smaller magnitude events are also useful in the sense that they can provide a sufficient density of events to examine spatial variations in source properties or ground motion amplitudes. Earthquakes in our study region exhibit both local coherence in median $\Delta\sigma$ and $\Delta PGA$ as well as systematic variations in these median values over larger, regional length scales throughout the Bay Area. If this finding that the distribution of $\Delta\sigma$ and hence $\Delta PGA$ varies as a function of source region is robust, it would provide an important observational constraint for the reduction of epistemic uncertainty in ground motion models through the removal of the ergodic assumption (*Anderson and Brune*, 1999; *Atik et al.*, 2010; *Lin et al.*, 2011; *Stafford*, 2014; *Baltay et al.*, 2017). One relatively simple implementation of nonergodicity would be to allow predicted ground motions to vary as a function of source region, guided by the spatial statistics and correlation length scales of the $\Delta\sigma$ estimates. This model would effectively assume time stationarity in average source properties — that the distribution of $\Delta\sigma$ recorded to date is representative of the expected values of $\Delta\sigma$ of future events — which is a topic

that warrants further study.

The between-event variability in $\Delta PGA$ for events in our dataset is significantly less ($\log_{10}$ standard deviation of 0.16) than would be implied by the between-event variability in $\Delta\sigma$ ($\log_{10}$ standard deviation of 0.40). This paradox has been discussed in many previous studies (e.g., *Cotton et al.*, 2013; *Baltay et al.*, 2013; *Oth et al.*, 2017), and we do not presume to offer a satisfactory resolution, only additional observational constraints. Measurements of ground motion amplitudes are relatively simple and require fewer modeling assumptions than analogous source spectral estimates. It is therefore possible that the combined effects of measurement uncertainties (e.g., precision in the resolution of the corner frequency) and uncertainties associated with the model parameterization (e.g., the deviation of the spectra of real earthquakes from the idealized Brune model) may account for a significant fraction of this disparity. The fact that the nominal measurement uncertainties in $\Delta\sigma$ as computed from a bootstrap analysis of the apparent source spectra (*Trugman and Shearer*, 2017) are comparable to the within-event variability in PGA (median $\log_{10}$ values of 0.219 and 0.214, respectively), suggest that the latter may be an important consideration. Another key difference is that measurements of $\Delta\sigma$ that are derived from source spectra must explicitly account for raypath and near-site attenuation, while measurements of ground motion intensities are effectively dampened because they apply no such correction for attenuation, and are further band-limited through the application of bandpass filtering. The damping of PGA variability from attenuation effects should be stronger for

smaller earthquakes because of their higher frequency content. Our results support the importance of this damping effect, as the variability in $\Delta PGA$ is observed to increase significantly with magnitude (Figure 7.5), while the variability in $\Delta\sigma$ does not. For $M_W$ 3 to 4 earthquakes, the ratio of log $\Delta PGA$ to $\Delta\sigma$ variability is 0.58, only 30% less than the theoretically expected value of 5/6 (*Cotton et al.*, 2013) in the absence of attenuation.

One further issue we can address using the Random Forest GMPE is the possibility that the measured between-event variability in PGA is biased low relative to the true value. This situation could potentially arise if true dynamic source effects are absorbed into the modeled ground motion through correlations of source properties with explicitly modeled features or data selection artifacts. To test for this possibility, we generate a synthetic PGA dataset $\delta y_{ij}$ in which we replace the measured between-event residual $\Delta PGA_i$ with a synthetic residual $\delta PGA_i$ equal to the deviation in PGA predicted by the deviation $\delta\sigma_i$ in stress drop from the average value (*Cotton et al.*, 2013):

$$\delta PGA_i = \frac{5}{6}\,\delta\sigma_i, \tag{7.10}$$

(where $\delta PGA$ and $\delta\sigma$ are both measured in consistent logarithmic units). We then apply the algorithm described in the Methods Section to fit the synthetic dataset with a Random Forest GMPE and hence derive model estimates of the between-event residuals of the synthetic dataset. The estimated synthetic residuals closely match the input synthetic distribution and are thus consistent with the observed

$\Delta\sigma$ variability (Figure 7.8), which suggests that we can rule out GMPE modeling artifacts as the cause of the disparity between the variability in ground motion and stress drop.

A final point worthy of discussion is our new approach to GMPE modeling based upon a mixed-effects generalization of the Random Forest algorithm. There are advantages and disadvantages inherent to using this technique, but we argue that the former outweigh the latter for the purposes of this study, and perhaps for several other potential applications of ground-motion modeling and estimation. The most appealing aspect of Random Forest in the context of GMPEs is its ability use an arbitrarily complex set of features to make robust predictions without the need to specify a parametric form for the relationship between these predictive features and the target response: a ground motion intensity measure like PGA. The Random Forest GMPE described in this study is simplistic in that it only uses two such features – magnitude and source-site distance – to predict PGA, but the same basic framework could easily be extended to include a much wider range of potential features and to make predictions for other intensity measures such as peak ground velocity or spectral acceleration at different periods. This would be particularly useful for GMPEs designed to predict ground motion intensity of large-magnitude events, in which predictive features related to finite fault rupture such as hanging wall effects and directivity, or those related to nonlinear site response, could be included in the model without defining a functional form for their influence on ground motion intensity. One could then examine the various metrics

of feature importance derived from the Random Forest model fit (*Breiman*, 2001) to evaluate which of these input features provide the most fundamental constraints on predicted ground motions.

The primary disadvantage to the Random Forest approach is that its lack of a parametric form makes it more difficult to export for application by external users. However, this drawback is easily overcome with only modest knowledge of computer programming and statistical theory, as we demonstrate in the form of a Python script available in the electronic supplement to this article. The nonparametric form of the Random Forest is in fact critical to its performance, as it mitigates the tendency for parametric models to introduce biased predictions or model residuals that vary systematically as a function of magnitude and distance. Instead, the algorithm used to train each tree in the Random Forest involves iteratively searching for and refining an optimal partitioning in feature space that minimizes the predictive misfit. This procedure, when combined with ensemble averaging of randomized and hence decorrelated individual trees, helps ensure that the Random Forest predictions are both robust to data outliers and are valid locally within feature space. Lastly, Random Forest regressors automatically provide a means of predictive validation through the evaluation of the the out-of-bag $\sigma$, and in this way allows the user to assess its expected performance and uncertainty with respect to future predictions.

## 7.6 Summary

We examine the relation between Brune-type stress drop ($\Delta\sigma$) and ground motion amplitudes (PGA) for $M \geq 1.5$ earthquakes occurring near the San Francisco Bay Area, California from 2002 through 2016. We estimate $\Delta\sigma$ for each event using a spectral decomposition method applied to vertical component $P$-wave spectra recorded by NCSN stations within 100km. We then measure PGA from horizontal component channels of the same set of stations, and develop a nonparametric GMPE using a mixed-effects implementation of the Random Forest algorithm. We use this GMPE to examine between-event variability in ground motion by defining event terms, $\Delta PGA$, as the average residual between the observed and model-predicted ground motion, given $M_W$, $R_{hyp}$, and station. The between-event residual $\Delta PGA$ exhibits a strong correlation with $\Delta\sigma$, especially for earthquakes with $M_W \geq 2.5$. Estimated values of $\Delta\sigma$ and $\Delta PGA$ are slightly higher for mainshocks than for aftershocks for the dataset as whole and for individual earthquake sequences in particular. Local coherence in the spatial patterns of $\Delta\sigma$ and $\Delta PGA$ supports future research into nonergodic GMPEs in which the distribution of expected ground motion depends on the location of the source region.

## Data and Resources

Waveform data, earthquake catalogs, and station metadata for this study were accessed through the Northern California Earthquake Data Center (NCEDC) and are publicly available from http://service.ncedc.org/ (last accessed August 2017). We use waveform analysis software for data in miniSEED and SAC format that is publicly available as part of the Incorporated Research Institutions for Seismology (IRIS) consortium (http://ds.iris.edu/ds/nodes/dmc/software/, last accessed July 2017). The mapped faulting structures shown in Figure 7.1 were obtained from the United States Geological Survey Quaternary Fault and Fold Database for the United States (http//earthquakes.usgs.gov/hazards/qfaults/, last accessed June 2017). The ground motion data associated with this study is available as Table S1 in the electronic supplement to this article. We use functions from the scikit-learn Python programming package (*Pedregosa et al.*, 2011) for our analyses in this study, and we provide an example script of our implementation of the Random Forest GMPE in the electronic supplement to this article.

## Acknowledgements

This manuscript is based upon work supported by the National Science Foundation Graduate Research Fellowship Program (NSFGRFP) under grant number DGE-1144086. Additional support was provided by the Southern California Earthquake Center (SCEC) under grant number 16020.

Chapter 7, in full, is a reformatted version of material as it appears in Bulletin of the Seismological Society of America: Trugman, D. T. and P. M. Shearer (2017), Strong correlation between stress drop and peak ground acceleration for recent seismicity in the San Francisco Bay Area, *Bulletin of the Seismological Society of America*, submitted. I was the primary investigator and author of this paper.

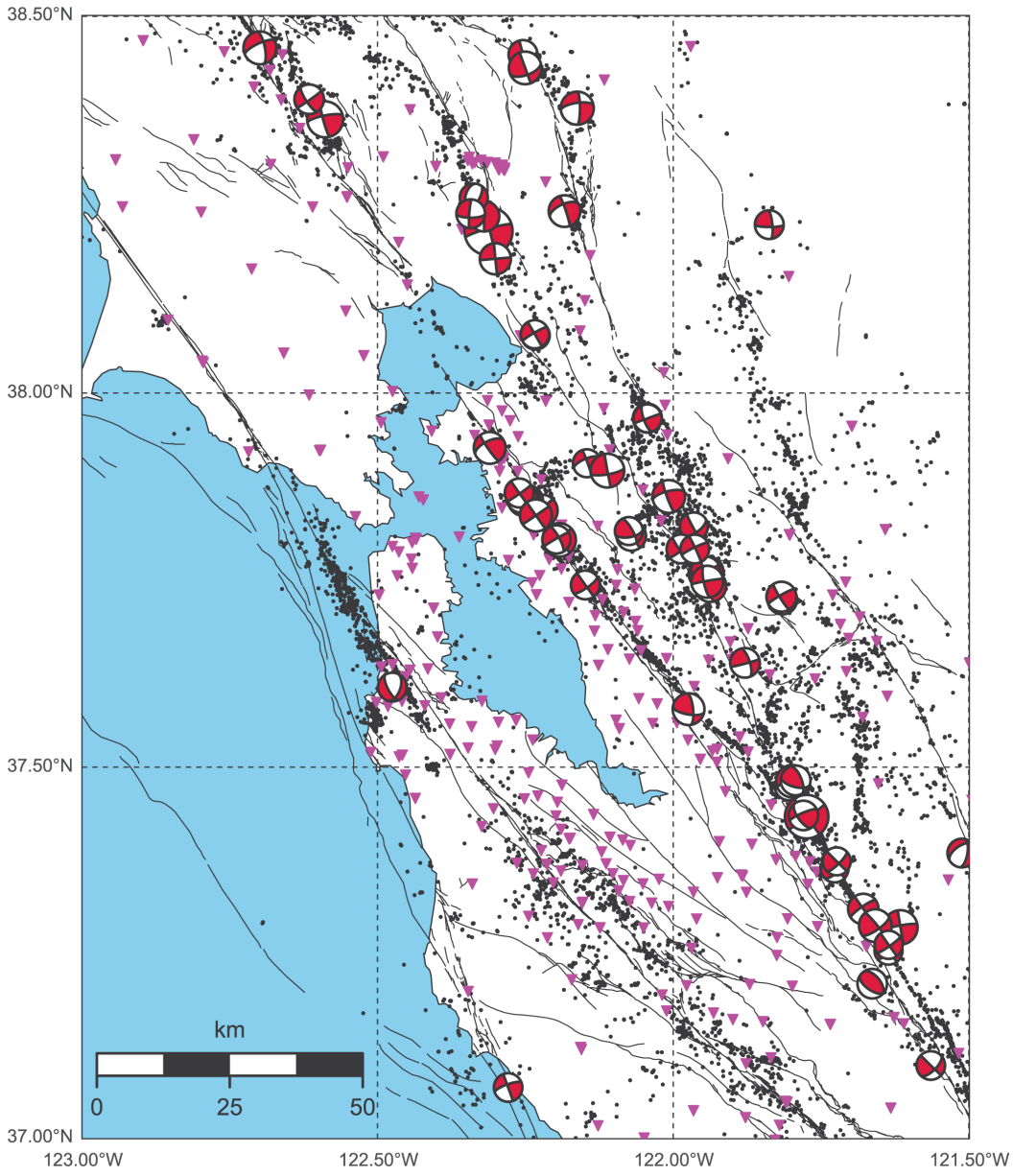


**Figure 7.1**: Map view of the San Francisco Bay Area study region. $M \geq 1.0$ seismicity from the relocated catalog of *Waldhauser and Schaff* (2008). Focal mechanisms for $M \geq 3.5$ events, NCSN station coverage, and mapped fault structures (see Data and Resources) are shown for reference.

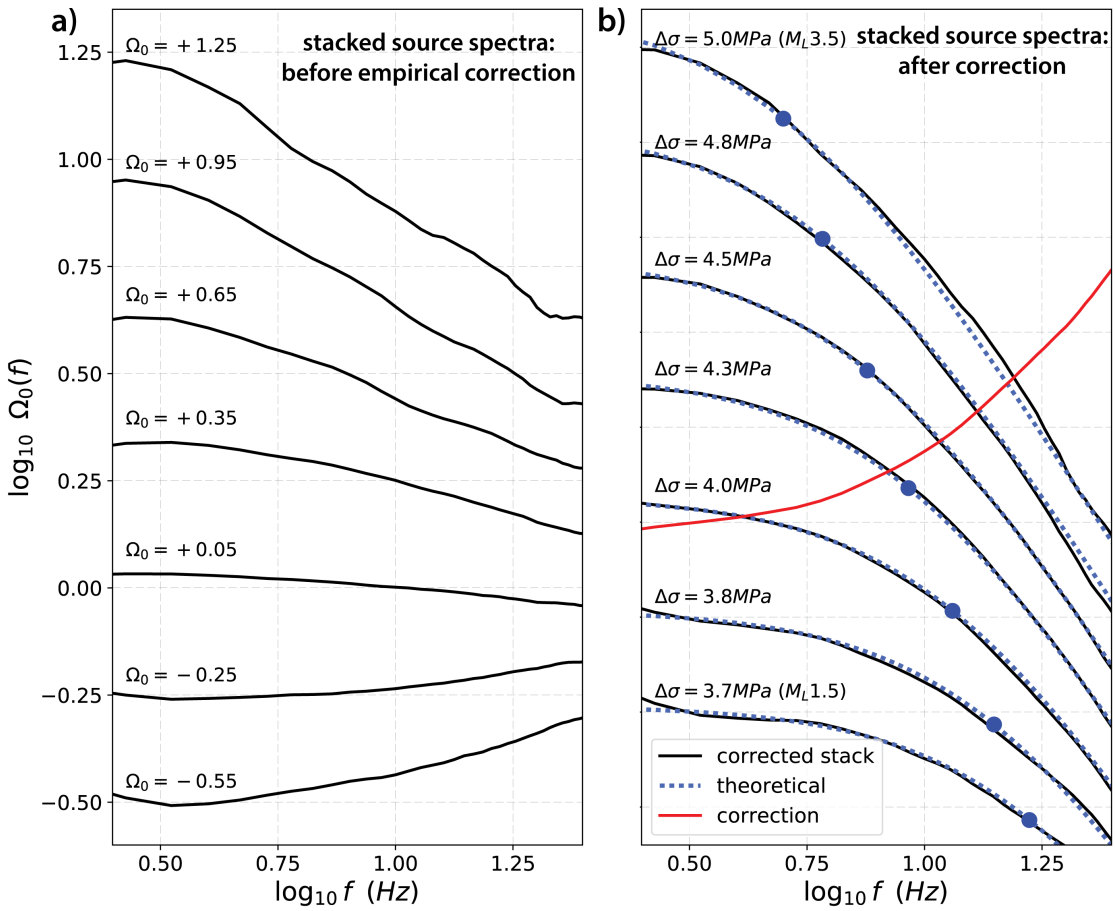


**Figure 7.2**: Stacked relative and corrected source spectra from earthquakes in the San Francisco Bay Area. (a) Stacked relative source spectra, binned by spectral moment $\Omega_0$ (logarithmic scale), prior to the application of the empirical correction term that accounts for common path effects. (b) Corrected source spectra (solid black lines), and comparison to Brune-type theoretical spectra (dashed blue lines). The implied corner frequency of each stack (blue dots) and the correction spectrum (red line) are marked for reference.

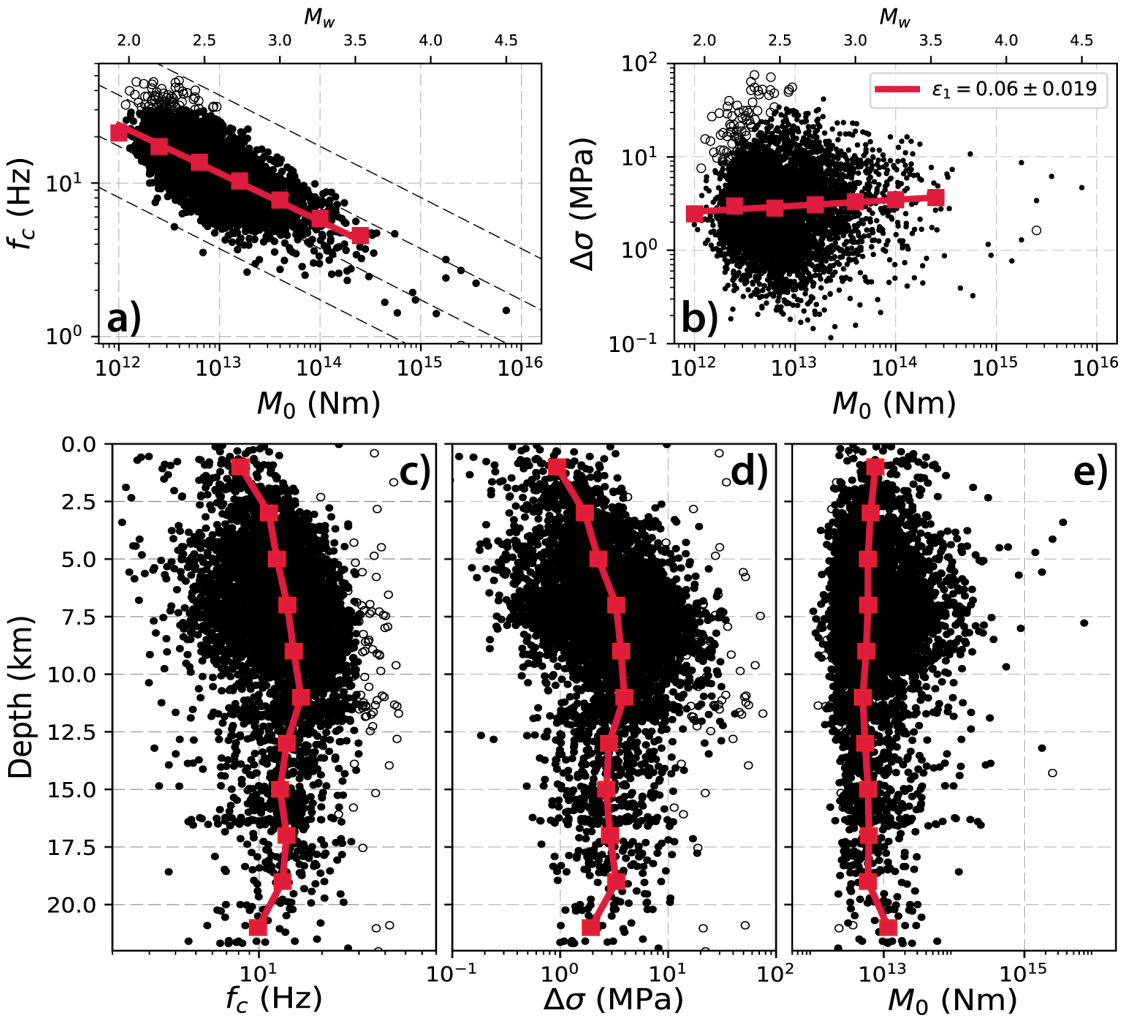


**Figure 7.3**: Distribution, scaling, and depth-dependence of $M_0$, $f_c$, and $\Delta\sigma$. In panels (a) and (b), corner frequency $f_c$ and stress drop $\Delta\sigma$ are plotted versus seismic moment $M_0$. Black dots in panels correspond to measurements of source properties for individual events, and the median $f_c$ and $\Delta\sigma$ in $M_0$ bins of width 0.4 ($\log_{10}$ N-m units) are marked with red squares. The best-fitting scaling parameter $\epsilon_1$ for the binned data is plotted with a red line, with its numerical value and two-sigma uncertainty listed in the panel (d) inset. The dashed black lines in panel (b) correspond to constant-$\Delta\sigma$ contours of 0.1, 1, 10, and 100 MPa. Events with poorly resolved corner frequencies due to bandwidth limitations are marked with open circles. Panels (c), (d), and (e) show the depth-dependence of $f_c$, $\Delta\sigma$, $M_0$, with the trend of median values in bins of 2km width marked with solid red lines.

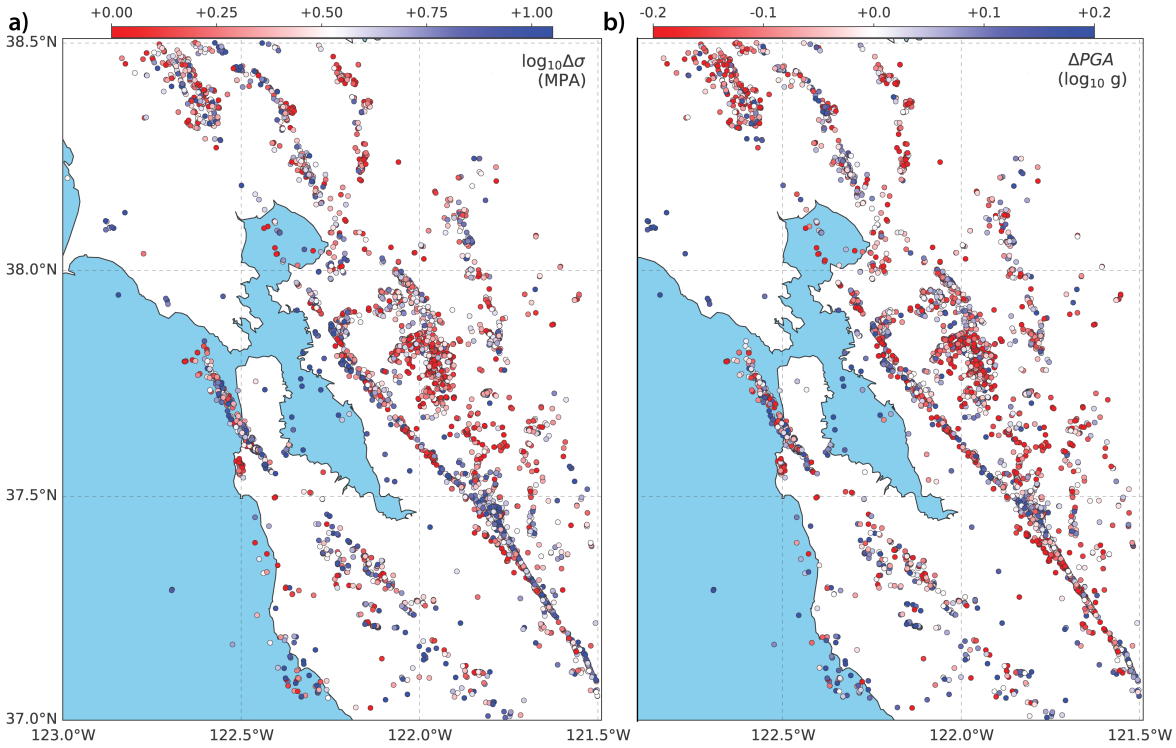


**Figure 7.4**: Spatial variations in stress drop and between-event ground motion residual for events within the San Francisco Bay Area study region.

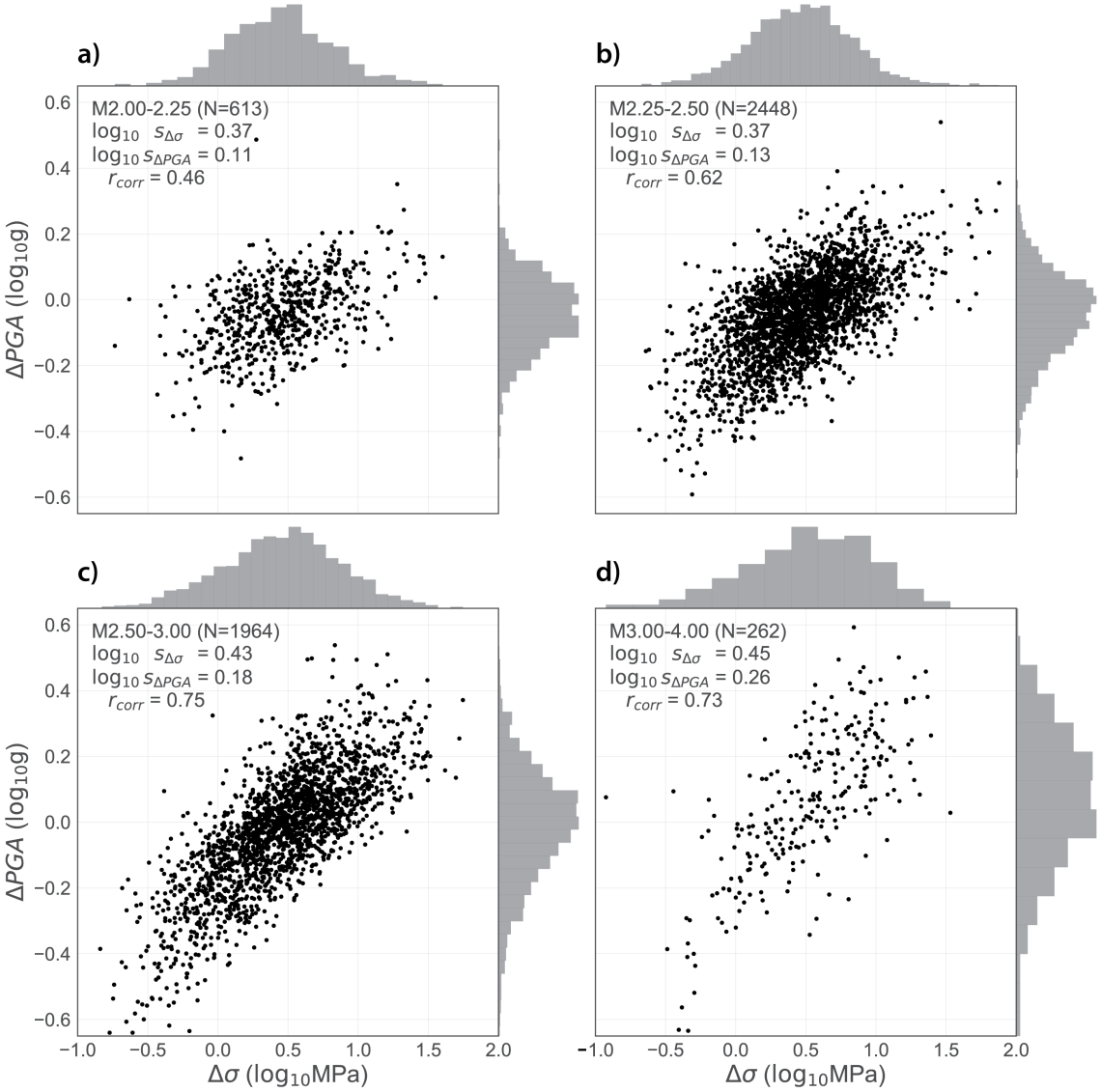


**Figure 7.5**: Correlation between stress drop, $\Delta\sigma$, and between-event residual, $\Delta PGA$, for events within the San Francisco Bay Area study region. Each panel corresponds to events within four distinct magnitude ranges: (a) $2.00 \leq M_W \leq 2.25$, (b) $2.25 \leq M_W \leq 2.50$, (c) $2.50 \leq M_W \leq 3.00$, (d) $3.00 \leq M_W \leq 4.00$. The number of events $N$, the $\log_{10}$ standard deviations $s_{\Delta\sigma}$ and $s_{\Delta PGA}$, and the correlation coefficient $r_{corr}$ are labeled in the inset of each panel.

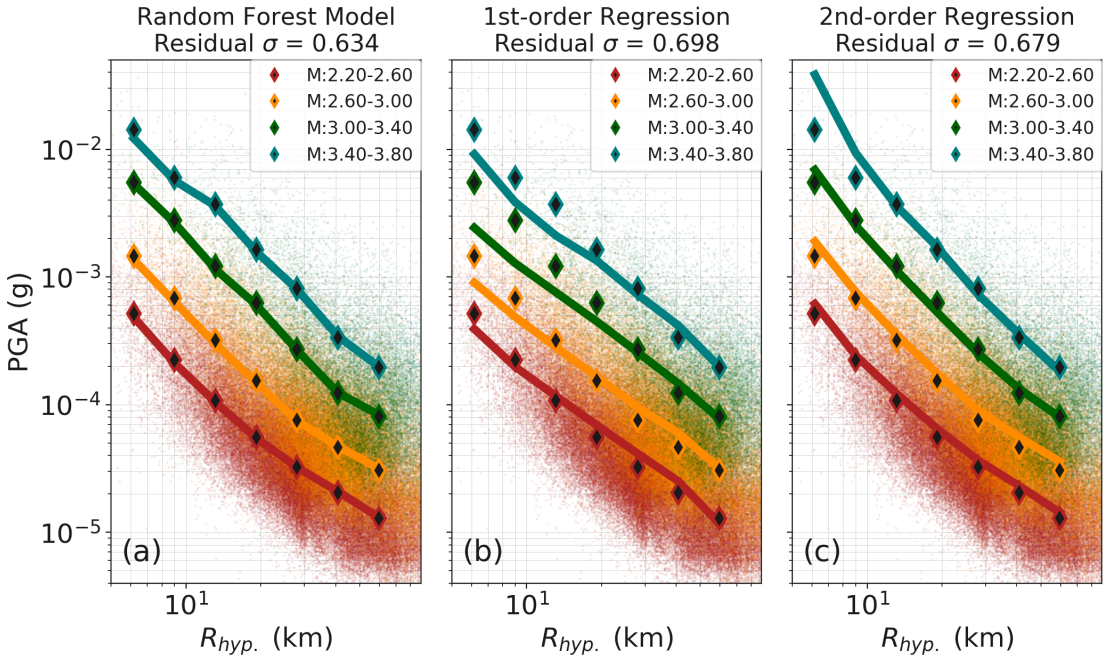


**Figure 7.6**: Random Forest GMPE model fit and comparison to mixed-effects linear regression. (a) Random Forest model prediction for PGA as a function of hypocentral distance ($R_{hyp}$, $x$-axis) and moment magnitude ($M_W$, colored lines). Median values of observed PGA, binned by $R_{hyp}$ and color-coded by magnitude range, are shown with diamond markers. Both the observations and model predictions have been corrected for station effects based upon the station terms inferred using the iterative procedure described in the text. Panels (b) and (c) show model fits for the first-order and second-order linear regression models defined by equations 7.8 and 7.9.

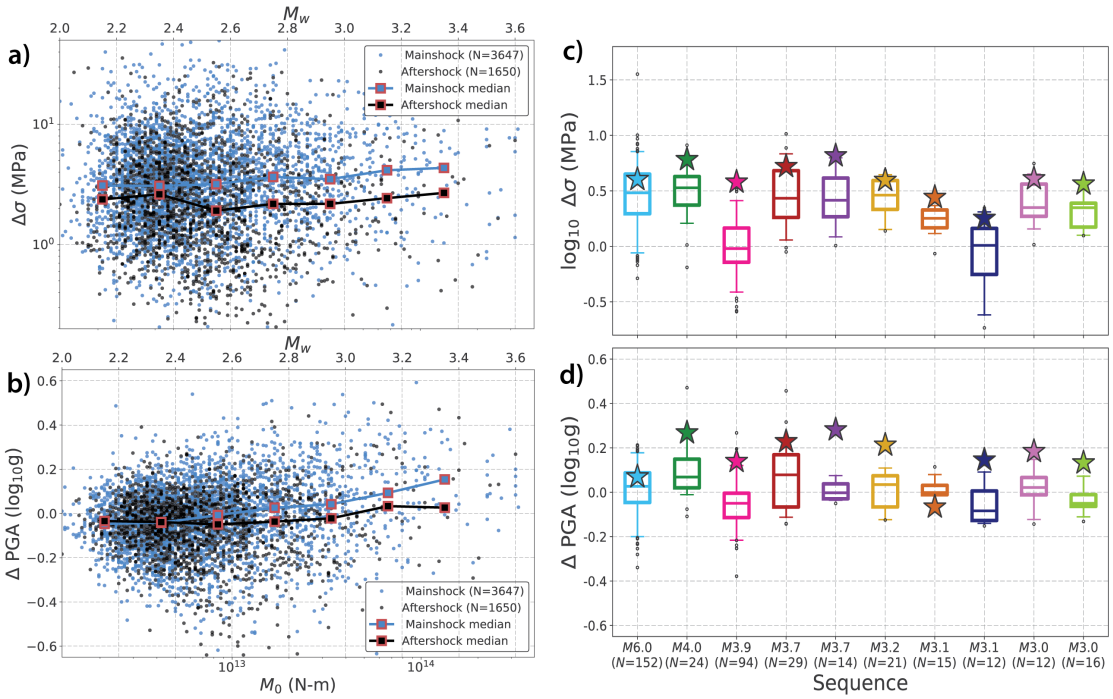


**Figure 7.7**: Differences in the distributions of stress drop and ground motion amplitudes of mainshock and aftershock events. (a) Stress drop $\Delta\sigma$ and (b) between-event residual $\Delta PGA$ vs seismic moment $M_0$. As detailed in the text, the neighbors clustering algorithm of *Zaliapin and Ben-Zion* (2013) is used to classify events as mainshocks (events that are the largest within their respective cluster, plotted in blue), and aftershocks (non-mainshock events, plotted in black). Median values of $\Delta\sigma$, binned by $M_0$, for mainshock and aftershock events are plotted as solid blue and black lines, respectively. Variability in (c) $\Delta\sigma$ and (d) $\Delta PGA$ for sequences with mainshock magnitude $M_W > 3.0$ with at least ten events. Sequence medians are denoted with a solid horizontal line, while the box and whiskers denote the inter-quartile range (50% confidence interval) and 90% confidence interval, respectively. Mainshock values of (c) $\Delta\sigma$ and (d) $\Delta PGA$ are marked with stars, and their corresponding magnitudes are listed along the $x$-axis.

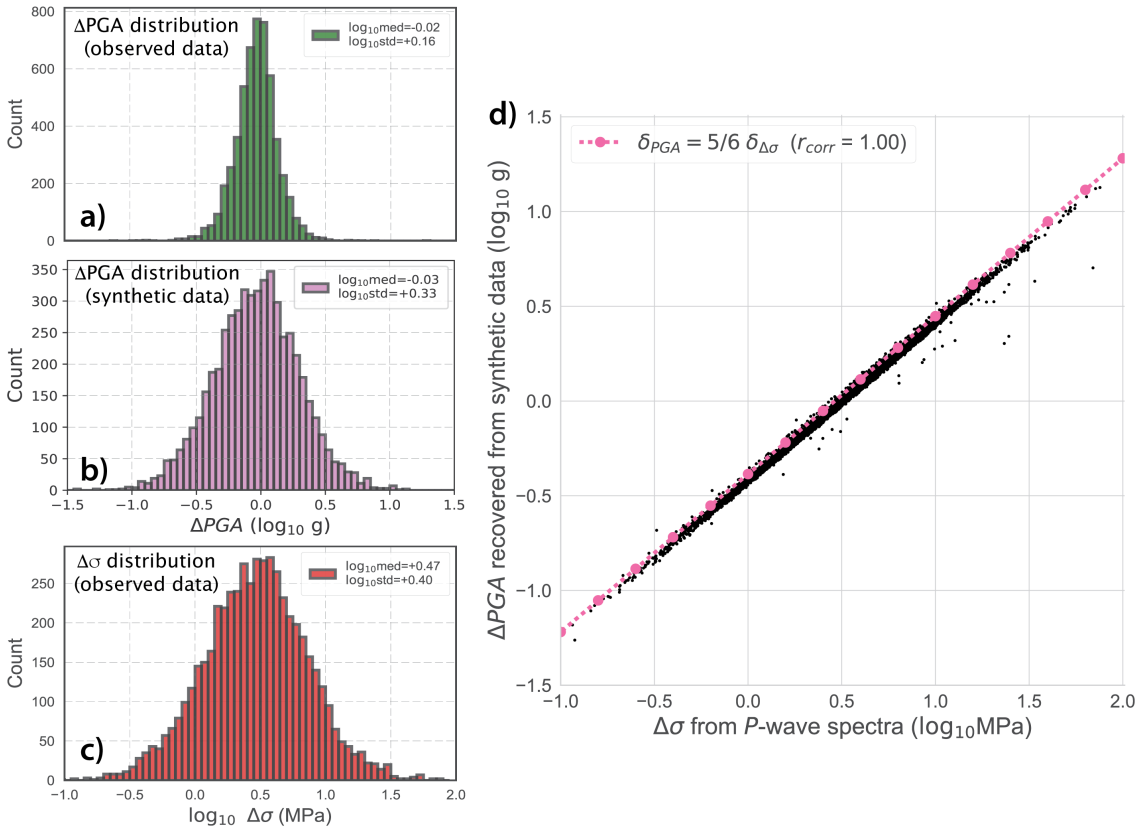


**Figure 7.8**: Variability in the distributions of stress drop $\Delta\sigma$ and between-event ground motion residual $\Delta PGA$. (a) Histogram of $\Delta PGA$ derived from the Bay Area region ground motion dataset (this study). Median and standard deviation values ($\log_{10}$ g) are labeled in the inset. (b) Histogram of Random Forest GMPE model-estimated $\Delta PGA$ values recovered from the synthetic ground motion dataset generated from the measured $\Delta\sigma$ values as described in the Discussion Section. (c) Histogram of $\Delta\sigma$ estimates from this study, with median and standard deviation values ($\log_{10}$ MPA) labeled in the inset. (d) Random Forest GMPE model-estimated $\Delta PGA$ values derived from the synthetic ground motion data plotted as a function of $\Delta\sigma$. The model-estimated $\Delta PGA$ values for the synthetic dataset closely match the input theoretical prediction of $\delta PGA = \frac{5}{6}\delta\sigma$ (*Cotton et al.*, 2013).

# 7.7 Supplementary Materials

## Overview

Figure 7.S1 shows the relation between $M_W$ and $M_L$ derived from analysis of long-period spectral amplitudes. Figure 7.S2 compares the ground motion measurements from this study with the cross-listed records in the Next Generation Attenuation ground motion database. Figure 7.S3 shows the validation curve used to select the optimal tree depth for the Random Forest GMPE used in this study. Figure 7.S4 plots the between-event ground motion residual versus: stress drop, magnitude-adjusted stress drop, depth, and depth-adjusted stress drop.

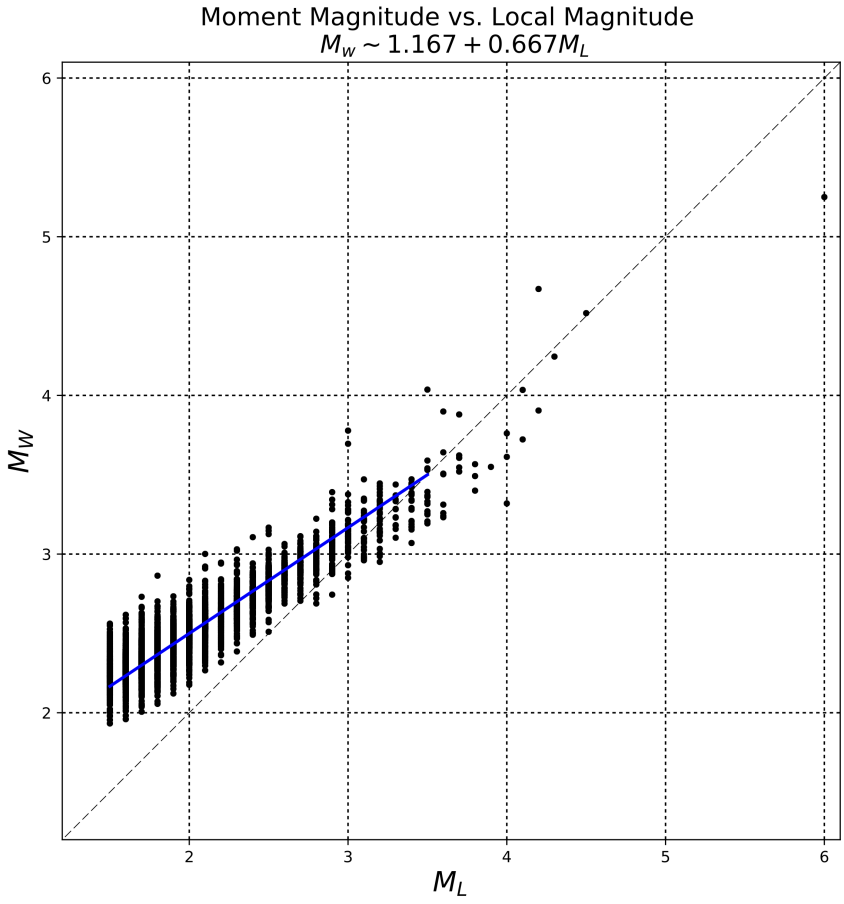


**Figure 7.S1**: Moment magnitude $M_W$ estimated from $P$-wave spectral decomposition versus catalog-listed magnitude $M_L$ for the San Francisco Bay Area earthquakes considered in study. The best-fit line inferred from linear regression is $M_W = 1.167 + 0.667 M_L$.

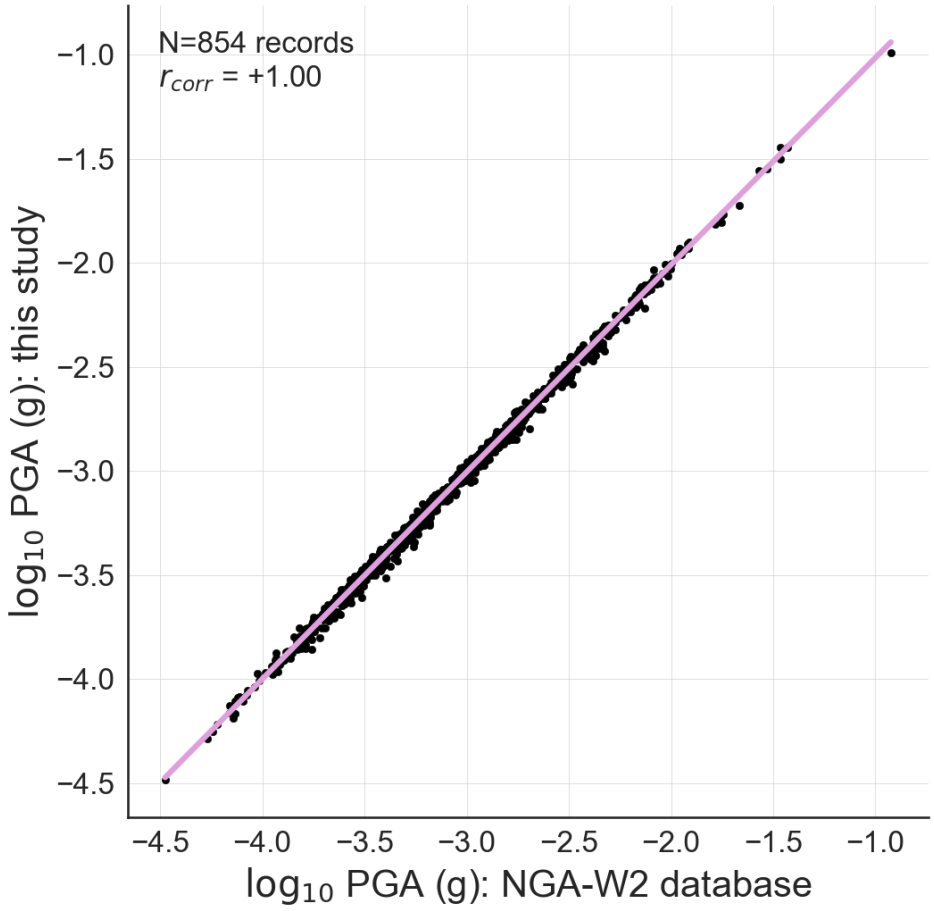


**Figure 7.S2**: Comparison of the peak ground acceleration (PGA) measurements from this study versus the corresponding PGA measurements listed in the Next Generation Attenuation Relationships for the Western US (NGA-W2). In total, there are 854 records analyzed in this study that are listed in the NGA-W2 database (see Data and Resources).

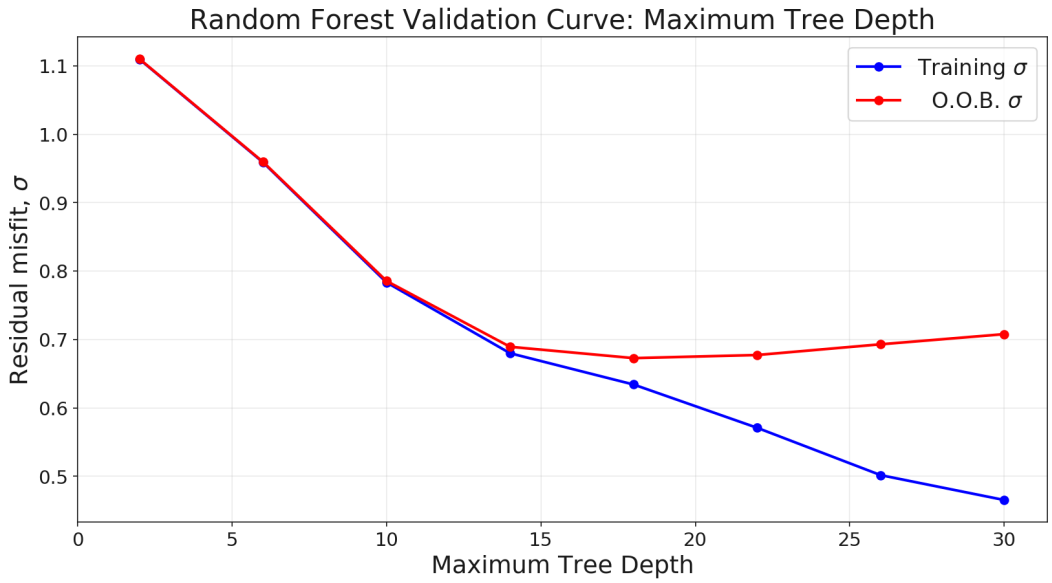


**Figure 7.S3**: Validation curve for the Random Forest GMPE developed in this study. Model misfit (defined as standard deviation of the residual distribution) is plotted as a function of the maximum tree depth of the individual decision trees in the Random Forest. The blue curve corresponds to the misfit measured with respect to the full input dataset. The red curve corresponds to the misfit measured with respect to the data points left out-of bag as part of the Random Forests bootstrapping of each tree. In contrast to the blue curve, which decreases monotonically with increasing tree depth, the red curve achieves its minimal value at a tree depth of 18 before gradually increasing due to model overfit. This validation curve is used to select the optimal tree depth for the Random Forest GMPE applied in this study.

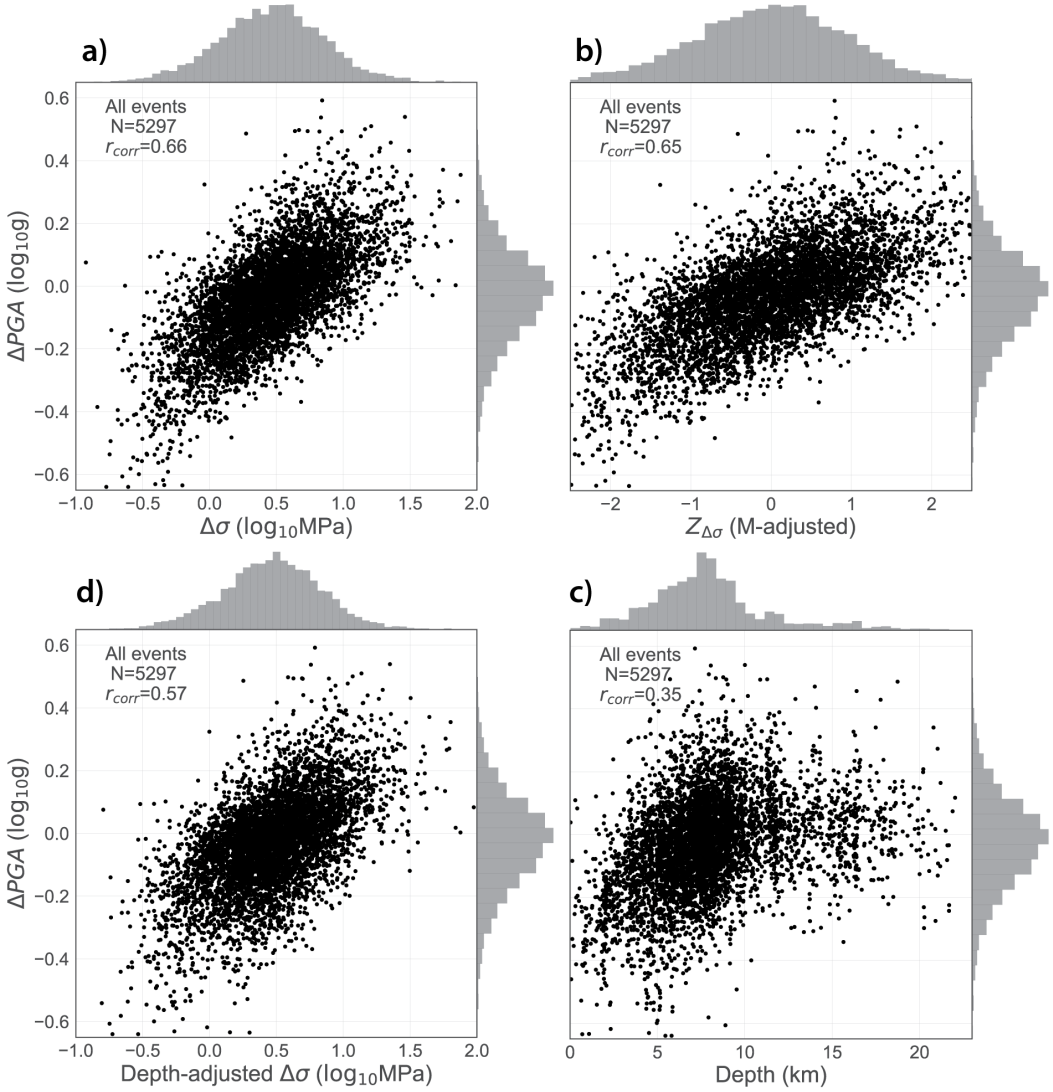


**Figure 7.S4**: Correlation between the between-event peak ground acceleration residual, $\Delta PGA$ and (a) log10 stress drop, $\Delta\sigma$ (MPa), (b) normalized, magnitude-adjusted $\Delta\sigma$, $Z_{\Delta\sigma}$, (c) hypocentral depth (km), and (d) log10 depth-adjusted $\Delta\sigma$. The correlation coefficient between $\Delta PGA$ and the associated feature ($\Delta\sigma$, magnitude-adjusted $\Delta\sigma$, depth, depth-adjusted $\Delta\sigma$) for the 5297 events analyzed in this study is listed in the inset of each panel.

# References

Abercrombie, R. E., S. Bannister, J. Ristau, and D. Doser (2017), Variability of earthquake stress drop in a subduction setting, the Hikurangi Margin, New Zealand, *Geophysical Journal International*, *208*(1), 306–320, doi:10.1093/gji/ggw393.

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

Aki, K. (1967), Scaling law of seismic spectrum, *Journal of Geophysical Research*, *72*(4), 1217–1231, doi:10.1029/JZ072i004p01217.

Allmann, B. P., and P. M. Shearer (2009), Global variations of stress drop for moderate to large earthquakes, *Journal of Geophysical Research: Solid Earth*, *114*(B1), B01,310, doi:10.1029/2008JB005821.

Anderson, J. G., and J. N. Brune (1999), Probabilistic Seismic Hazard Analysis without the Ergodic Assumption, *Seismological Research Letters*, *70*(1), 19–28, doi:10.1785/gssrl.70.1.19.

Anderson, J. G., and S. E. Hough (1984), A model for the shape of the fourier amplitude spectrum of acceleration at high frequencies, *Bulletin of the Seismological Society of America*, *74*(5), 1969–1993.

Atik, L. A., N. Abrahamson, J. J. Bommer, F. Scherbaum, F. Cotton, and N. Kuehn (2010), The Variability of Ground-Motion Prediction Models and Its Components, *Seismological Research Letters*, *81*(5), 794–801, doi:10.1785/gssrl.81.5.794.

Atkinson, G. M. (1990), A Comparison of Eastern North American ground Motion Observations with Theoretical Predictions, *Seismological Research Letters*, *61*(3-4), 171–180, doi:10.1785/gssrl.61.3-4.171.

Atkinson, G. M., and I. Beresnev (1997), Don't Call it Stress Drop, *Seismological Research Letters*, *68*(1), 3–4, doi:10.1785/gssrl.68.1.3.

Atkinson, G. M., and M. Morrison (2009), Observations on Regional Variability in Ground-Motion Amplitudes for Small-to-Moderate Earthquakes in North America, *Bulletin of the Seismological Society of America*, *99*(4), 2393–2409, doi:10.1785/0120080223.

Baltay, A., J. L. Rubinstein, F. M. Terra, T. C. Hanks, and R. B. Herrmann (2015), Stress Drop and Depth Controls on Ground Motion From Induced Earthquakes, *AGU Fall Meeting Abstracts*.

Baltay, A. S., and T. C. Hanks (2014), Understanding the Magnitude Dependence of PGA and PGV in NGA-West 2 Data, *Bulletin of the Seismological Society of America*, *104*(6), 2851–2865, doi:10.1785/0120130283.

Baltay, A. S., T. C. Hanks, and G. C. Beroza (2013), Stable Stress-Drop Measurements and their Variability: Implications for Ground-Motion Prediction, *Bulletin of the Seismological Society of America*, *103*(1), 211–222, doi:10.1785/0120120161.

Baltay, A. S., T. C. Hanks, and N. A. Abrahamson (2017), Uncertainty, Variability, and Earthquake Physics in Ground-Motion Prediction Equations, *Bulletin of the Seismological Society of America*, *107*(4), 1754–1772, doi:10.1785/0120160164.

Bindi, D. (2017), The Predictive Power of Ground-Motion Prediction Equations, *Bulletin of the Seismological Society of America*, doi:10.1785/0120160224.

Bindi, D., F. Pacor, L. Luzi, R. Puglia, M. Massa, G. Ameri, and R. Paolucci (2011), Ground motion prediction equations derived from the Italian strong motion database, *Bulletin of Earthquake Engineering*, *9*(6), 1899–1920, doi:10.1007/s10518-011-9313-z.

Bishop, C. M. (2006), *Pattern Recognition and Machine Learning*, Information Science and Statistics, Springer, New York.

Boatwright, J. (1982), A dynamic model for far-field acceleration, *Bulletin of the Seismological Society of America*, *72*(4), 1049–1068.

Bommer, J. J., and N. A. Abrahamson (2006), Why Do Modern Probabilistic Seismic-Hazard Analyses Often Lead to Increased Hazard Estimates?, *Bulletin of the Seismological Society of America*, *96*(6), 1967–1977, doi:10.1785/0120060043.

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

Boore, D. M. (2003), Simulation of Ground Motion Using the Stochastic Method, in *Seismic Motion, Lithospheric Structures, Earthquake and Volcanic Sources: The Keiiti Aki Volume*, edited by Y. Ben-Zion, Pageoph Topical Volumes, pp. 635–676, Birkhauser Basel, doi:10.1007/978-3-0348-8010-7_10.

Boore, D. M., J. P. Stewart, E. Seyhan, and G. M. Atkinson (2013), NGA-West2 Equations for Predicting PGA, PGV, and 5% Damped PSA for Shallow Crustal Earthquakes, *Earthquake Spectra*, *30*(3), 1057–1085, doi:10.1193/070113EQS184M.

Boyd, O. S., D. E. McNamara, S. Hartzell, and G. Choy (2017), Influence of Lithostatic Stress on Earthquake Stress Drops in North America, *Bulletin of the Seismological Society of America*, *107*(2), 856–868, doi:10.1785/0120160219.

Bozorgnia, Y., N. A. Abrahamson, L. A. Atik, T. D. Ancheta, G. M. Atkinson, J. W. Baker, A. Baltay, D. M. Boore, K. W. Campbell, B. S.-J. Chiou, R. Darragh, S. Day, J. Donahue, R. W. Graves, N. Gregor, T. Hanks, I. M. Idriss, R. Kamai, T. Kishida, A. Kottke, S. A. Mahin, S. Rezaeian, B. Rowshandel, E. Seyhan, S. Shahi, T. Shantz, W. Silva, P. Spudich, J. P. Stewart, J. Watson-Lamprey, K. Wooddell, and R. Youngs (2014), NGA-West2 Research Project, *Earthquake Spectra*, *30*(3), 973–987, doi:10.1193/072113EQS209M.

Breiman, L. (2001), Random Forests, *Machine Learning*, *45*(1), 5–32, doi:10.1023/A:1010933404324.

Brocher, T. M., A. S. Baltay, J. L. Hardebeck, F. F. Pollitz, J. R. Murray, A. L. Llenos, D. P. Schwartz, J. L. Blair, D. J. Ponti, J. J. Lienkaemper, V. E. Langenheim, T. E. Dawson, K. W. Hudnut, D. R. Shelly, D. S. Dreger, J. Boatwright, B. T. Aagaard, D. J. Wald, R. M. Allen, W. D. Barnhart, K. L. Knudsen, B. A. Brooks, and K. M. Scharer (2015), The Mw 6.0 24 August 2014 South Napa Earthquake, *Seismological Research Letters*, *86*(2A), 309–326, doi:10.1785/0220150004.

Brune, J. N. (1970), Tectonic stress and the spectra of seismic shear waves from earthquakes, *Journal of Geophysical Research*, *75*(26), 4997–5009, doi:10.1029/JB075i026p04997.

Brune, J. N. (1971), Seismic sources, fault plane studies and tectonics, *Eos, Transactions American Geophysical Union*, *52*(5), IUGG 178, doi:10.1029/EO052i005pIU178.

Campbell, K. W., and Y. Bozorgnia (2008), NGA Ground Motion Model for the Geometric Mean Horizontal Component of PGA, PGV, PGD and 5% Damped Linear Elastic Response Spectra for Periods Ranging from 0.01 to 10 s, *Earthquake Spectra*, *24*(1), 139–171, doi:10.1193/1.2857546.

Cotton, F., R. Archuleta, and M. Causse (2013), What is Sigma of the Stress Drop?, *Seismological Research Letters*, *84*(1), 42–48, doi:10.1785/0220120087.

Cramer, C. H. (2017), Brune Stress Parameter Estimates for the 2016 Mw 5.8 Pawnee and Other Oklahoma Earthquakes, *Seismological Research Letters*, *88*(4), 1005–1016, doi:10.1785/0220160224.

D'Amico, M., M. M. Tiberti, E. Russo, F. Pacor, and R. Basili (2017), Ground-Motion Variability for Single Site and Single Source through Deterministic Stochastic Method Simulations: Implications for PSHA, *Bulletin of the Seismological Society of America*, doi:10.1785/0120150377.

Douglas, J. (2003), Earthquake ground motion estimation using strong-motion records: a review of equations for the estimation of peak ground acceleration and response spectral ordinates, *Earth-Science Reviews*, *61*(1), 43–104, doi:10.1016/S0012-8252(02)00112-5.

Douglas, J., and B. Edwards (2016), Recent and future developments in earthquake ground motion estimation, *Earth-Science Reviews*, *160*, 203–219, doi:10.1016/j.earscirev.2016.07.005.

Dreger, D. S., M.-H. Huang, A. Rodgers, T. Taira, and K. Wooddell (2015), Kinematic Finite-Source Model for the 24 August 2014 South Napa, California, Earthquake from Joint Inversion of Seismic, GPS, and InSAR Data, *Seismological Research Letters*, *86*(2A), 327–334, doi:10.1785/0220140244.

Fernandez-Delgado, M., E. Cernadas, S. Barro, and D. Amorim (2014), Do We Need Hundreds of Classifiers to Solve Real World Classification Problems?, *Journal of Machine Learning Research*, *15*, 3133–3181.

Field, E. H., R. J. Arrowsmith, G. P. Biasi, P. Bird, T. E. Dawson, K. R. Felzer, D. D. Jackson, K. M. Johnson, T. H. Jordan, C. Madden, A. J. Michael, K. R. Milner, M. T. Page, T. Parsons, P. M. Powers, B. E. Shaw, W. R. Thatcher, R. J. Weldon, and Y. Zeng (2014), Uniform California Earthquake Rupture Forecast, Version 3 (UCERF3)–The Time-Independent Model, *Bulletin of the Seismological Society of America*, *104*(3), 1122–1180, doi:10.1785/0120130164.

Gardner, J. K., and L. Knopoff (1974), Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian?, *Bulletin of the Seismological Society of America*, *64*(5), 1363–1367.

Geurts, P., D. Ernst, and L. Wehenkel (2006), Extremely randomized trees, *Machine Learning*, *63*(1), 3–42, doi:10.1007/s10994-006-6226-1.

Hanks, T. C. (1979), b values and omega-gamma seismic source models: Implications for tectonic stress variations along active crustal fault zones and the estimation of high-frequency strong ground motion, *Journal of Geophysical Research: Solid Earth*, *84*(B5), 2235–2242, doi:10.1029/JB084iB05p02235.

Hanks, T. C. (1982), fmax, *Bulletin of the Seismological Society of America*, *72*(6A), 1867–1879.

Hanks, T. C., and R. K. McGuire (1981), The character of high-frequency strong ground motion, *Bulletin of the Seismological Society of America*, *71*(6), 2071–2095.

Hardebeck, J. L., and A. Aron (2009), Earthquake Stress Drops and Inferred Fault Strength on the Hayward Fault, East San Francisco Bay, California, *Bulletin of the Seismological Society of America*, *99*(3), 1801–1814, doi:10.1785/0120080242.

Hardebeck, J. L., and D. R. Shelly (2016), Aftershocks of the 2014 South Napa, California, Earthquake: Complex Faulting on Secondary Faults, *Bulletin of the Seismological Society of America*, doi:10.1785/0120150169.

Hardebeck, J. L., A. J. Michael, and T. M. Brocher (2007), Seismic Velocity Structure and Seismotectonics of the Eastern San Francisco Bay Region, California, *Bulletin of the Seismological Society of America*, *97*(3), 826–842, doi:10.1785/0120060032.

Hastie, T., R. Tibshirani, and J. Friedman (2009), *The Elements of Statistical Learning*, Springer Series in Statistics, Springer Publishing, New York, NY, doi:10.1007/978-0-387-84858-7.

Hawthorne, J. C., J.-P. Ampuero, and M. Simons (2016), A Method for Calibration of the Local Magnitude Scale Based on Relative Spectral Amplitudes, and Application to the San Juan Bautista, California, Area, *Bulletin of the Seismological Society of America*, doi:10.1785/0120160141.

Izutani, Y. (2005), Radiated energy from the mid Niigata, Japan, earthquake of October 23, 2004, and its aftershocks, *Geophysical Research Letters*, *32*(21), doi:10.1029/2005GL024116.

Ji, C., R. J. Archuleta, and C. Twardzik (2015), Rupture history of 2014 Mw 6.0 South Napa earthquake inferred from near-fault strong motion data and its impact to the practice of ground strong motion prediction, *Geophysical Research Letters*, *42*(7), 2015GL063,335, doi:10.1002/2015GL063335.

Jordan, T., Y.-T. Chen, P. Gasparini, R. Madariaga, I. Main, W. Marzocchi, G. Papadopoulos, K. Yamaoka, and J. Zschau (2011), Operational Earthquake Forecasting: State of Knowledge and Guidelines for Implementation., doi:10.4401/ag-5350.

Kaneko, Y., and P. M. Shearer (2014), Seismic source spectra and estimated stress drop derived from cohesive-zone models of circular subshear rupture, *Geophysical Journal International*, *197*(2), 1002–1015, doi:10.1093/gji/ggu030.

Kurzon, I., F. L. Vernon, Y. Ben-Zion, and G. Atkinson (2014), Ground Motion Prediction Equations in the San Jacinto Fault Zone: Significant Effects of Rupture Directivity and Fault Zone Amplification, *Pure and Applied Geophysics*, *171*(11), 3045–3081, doi:10.1007/s00024-014-0855-2.

Lin, P.-S., B. Chiou, N. Abrahamson, M. Walling, C.-T. Lee, and C.-T. Cheng (2011), Repeatable Source, Site, and Path Effects on the Standard Deviation for Empirical Ground-Motion Prediction Models, *Bulletin of the Seismological Society of America*, *101*(5), 2281–2295, doi:10.1785/0120090312.

Lior, I., and A. Ziv (2017), The Relation between Ground Acceleration and Earthquake Source Parameters: Theory and Observations, *Bulletin of the Seismological Society of America*, *107*(2), doi:10.1785/0120160251.

Llenos, A. L., and A. J. Michael (2017), Forecasting the (Un)Productivity of the 2014 M6.0 South Napa Aftershock Sequence, *Seismological Research Letters*, doi:10.1785/0220170050.

Louppe, G. (2014), Understanding Random Forests: From Theory to Practice, Ph.D. thesis, Universite de Liege, Liege, Belgium.

Madariaga, R. (1976), Dynamics of an expanding circular fault, *Bulletin of the Seismological Society of America*, *66*(3), 639–666.

Mak, S., F. Cotton, and D. Schorlemmer (2017), Measuring the Performance of Ground-Motion Models: The Importance of Being Independent, *Seismological Research Letters*, doi:10.1785/0220170097.

Mayeda, K., L. Malagnini, and W. R. Walter (2007), A new spectral ratio method using narrow band coda envelopes: Evidence for non-self-similarity in the Hector Mine sequence, *Geophysical Research Letters*, *34*(11), L11,303, doi:10.1029/2007GL030041.

Murphy, K. P. (2012), *Machine Learning: A Probabilistic Perspective*, MIT Press.

Oth, A., H. Miyake, and D. Bindi (2017), On the relation of earthquake stress drop and ground motion variability, *Journal of Geophysical Research: Solid Earth*, p. 2017JB014026, doi:10.1002/2017JB014026.

Page, M. T., N. van der Elst, J. Hardebeck, K. Felzer, and A. J. Michael (2016), Three Ingredients for Improved Global Aftershock Forecasts: Tectonic Region, Time-Dependent Catalog Incompleteness, and Intersequence Variability, *Bulletin of the Seismological Society of America*, *106*(5), 2290–2301, doi:10.1785/0120160073.

Park, J., C. R. Lindberg, and F. L. Vernon (1987), Multitaper spectral analysis of high-frequency seismograms, *Journal of Geophysical Research*, *92*(B12), 12,675, doi:10.1029/JB092iB12p12675.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos,

D. Cournapeau, M. Brucher, M. Perrot, and e. Duchesnay (2011), Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research*, *12*, 2825–2830.

Prieto, G., R. Parker, and F. Vernon III (2009), A Fortran 90 library for multitaper spectrum analysis, *Computers & Geosciences*, *35*(8), 1701–1710, doi:10.1016/j.cageo.2008.06.007.

Ross, Z. E., Y. Ben-Zion, M. C. White, and F. L. Vernon (2016), Analysis of earthquake body wave spectra for potency and magnitude values: implications for magnitude scaling relations, *Geophysical Journal International*, *207*(2), 1158–1164, doi:10.1093/gji/ggw327.

Shearer, P. M., G. A. Prieto, and E. Hauksson (2006), Comprehensive analysis of earthquake source spectra in southern California, *Journal of Geophysical Research*, *111*(B6), doi:10.1029/2005JB003979.

Stafford, P. J. (2014), Crossed and Nested Mixed-Effects Approaches for Enhanced Model Development and Removal of the Ergodic Assumption in Empirical Ground-Motion Models, *Bulletin of the Seismological Society of America*, *104*(2), 702–719, doi:10.1785/0120130145.

Sumy, D. F., C. J. Neighbors, E. S. Cochran, and K. M. Keranen (2017), Low stress drops observed for aftershocks of the 2011 Mw 5.7 Prague, Oklahoma, earthquake, *Journal of Geophysical Research: Solid Earth*, *122*(5), 2016JB013,153, doi:10.1002/2016JB013153.

Trugman, D. T., and P. M. Shearer (2017), Application of an improved spectral decomposition method to examine earthquake source scaling in southern California, *Journal of Geophysical Research: Solid Earth*, *122*(4), 2017JB013,971, doi:10.1002/2017JB013971.

van Stiphout, T., J. Zhuang, and D. Marsan (2012), Seismicity declustering, *Community Online Resource for Statistical Seismicity Analysis*, *10*, 1.

Viegas, G., R. E. Abercrombie, and W.-Y. Kim (2010), The 2002 M5 Au Sable Forks, NY, earthquake sequence: Source scaling relationships and energy budget, *Journal of Geophysical Research: Solid Earth*, *115*(B7), B07,310, doi:10.1029/2009JB006799.

Waldhauser, F., and W. L. Ellsworth (2002), Fault structure and mechanics of the Hayward Fault, California, from double-difference earthquake locations, *Journal of Geophysical Research: Solid Earth*, *107*(B3), ESE 3–1, doi:10.1029/2000JB000084.

Waldhauser, F., and D. P. Schaff (2008), Large-scale relocation of two decades of Northern California seismicity using cross-correlation and double-difference methods, *Journal of Geophysical Research*, *113*(B8), doi:10.1029/2007JB005479.

Wei, S., S. Barbot, R. Graves, J. J. Lienkaemper, T. Wang, K. Hudnut, Y. Fu, and D. Helmberger (2015), The 2014 Mw6.1 South Napa Earthquake: A Unilateral Rupture with Shallow Asperity and Rapid Afterslip, *Seismological Research Letters*, *86*(2A), 344–354, doi:10.1785/0220140249.

Wooddell, K. E., and N. A. Abrahamson (2014), Classification of Main Shocks and Aftershocks in the NGA-West2 Database, *Earthquake Spectra*, *30*(3), 1257–1267, doi:10.1193/071913EQS208M.

Yenier, E., and G. M. Atkinson (2014), Equivalent Point-Source Modeling of Moderate-to-Large Magnitude Earthquakes and Associated Ground-Motion Saturation Effects, *Bulletin of the Seismological Society of America*, *104*(3), 1458–1478, doi:10.1785/0120130147.

Yenier, E., and G. M. Atkinson (2015), Regionally Adjustable Generic Ground Motion Prediction Equation Based on Equivalent Point Source Simulations: Application to Central and Eastern North America, *Bulletin of the Seismological Society of America*, *105*(4), 1989–2009, doi:10.1785/0120140332.

Yenier, E., G. M. Atkinson, and D. F. Sumy (2017), Ground Motions for Induced Earthquakes in Oklahoma, *Bulletin of the Seismological Society of America*, *107*(1), 198–215, doi:10.1785/0120160114.

Zaliapin, I., and Y. Ben-Zion (2013), Earthquake clusters in southern California I: Identification and stability, *Journal of Geophysical Research: Solid Earth*, *118*(6), 2847–2864, doi:10.1002/jgrb.50179.

# Chapter 8

# Conclusions

## 8.1 Summary

The six studies that comprise this thesis consider different ways in which real earthquakes deviate from the canonical behavior implied by simple statistical models. Earthquake nucleation and rupture processes are complex physical phenomena, and as such it should be unsurprising that the observed variability in earthquake occurrence patterns and source properties exceeds that which would be predicted by these models. But in providing new observational constraints and in developing new techniques to quantitatively characterize this variability, we can augment our current scientific understanding of earthquake source processes. In turn, we can apply this new scientific knowledge to generalize our physical and statistical models of earthquake occurrence and improve the seismic hazard forecasts that rely upon them. The prominent statistician George Box is noted for saying

"All models are wrong but some are useful." By understanding how and why even our most useful models are wrong, we can make them more useful while learning something fundamental about the Earth.

In Chapter 2, we examined the complexity of the 2010 M7.2 El Mayor-Cucapah earthquake and its relation to energy production at the nearby Cerro Prieto Geothermal Field. The El Mayor-Cucapah earthquake was the largest event in the region of northern Mexico and southern California in decades, yet its occurrence was unexpected in that the primary faults that ruptured were previously thought to be inactive and of low hazard. We demonstrated that crustal stresses from net fluid extraction at the Cerro Prieto Geothermal Field may have primed these faults for failure, and thus provide a plausible explanation for the El Mayor-Cucapah earthquake's unexpected occurrence. However, the anthropogenic stress perturbation is only one of several possible mechanisms in this regard, and should be viewed with some skepticism because of the large epistemic uncertainties associated with Coulomb stress calculations (*Woessner et al.*, 2012; *Cattania et al.*, 2014; *Mildon et al.*, 2016). The Cerro Prieto Geothermal Field is a vital resource for northern Baja California that provides power for hundreds of thousands of local inhabitants. It would be unwise to restrict Cerro Prieto's energy production without further study, but the results presented in this study make clear that its potential influence on local hazard should not be neglected.

In Chapter 3, we turned our attention to seismicity within California's three largest geothermal fields: The Geysers, located to the north of the San Francisco,

the Salton Sea Geothermal Field, in southern California's Imperial Valley, and the Coso Geothermal field, east of the Sierra Nevada in central California. Collectively, these fields provide a important source of renewable energy for California's energy budget. However, the location of these fields in regions of historical earthquake activity makes it important to understand the extent to which stresses from energy production influence local seismicity and earthquake hazard. To this end, we studied temporal changes in earthquake occurrence patterns within each field over the preceding 25 years. At The Geysers, we observed clear temporal changes in background seismicity rates that are driven by seasonal and secular changes in fluid injection within the geothermal reservoir. At Salton Sea and Coso, we likewise observed a permanent elevation in mean background rates following initial energy production, but later transient changes in seismicity are less clearly associated with localized anthropogenic stresses and may instead be driven by regional earthquake swarm activity. These results suggest that temporal changes in earthquake hazard associated with geothermal activity are region-specific and jointly controlled by tectonic and anthropogenic stress transients, rather than just the latter in isolation.

In Chapter 4, we developed a new algorithm to obtain high-resolution earthquake locations using differential travel time data. The GrowClust method uses a hybrid, hierarchical clustering algorithm to simultaneously group events into clusters based on waveform similarity and relocate each event with respect to its cluster neighbors based on the observed distribution of differential travel times. GrowClust is fast, stable, and robust to the data outliers that are commonly found

in seismic waveform datasets. We applied the GrowClust algorithm to examine two recent earthquake sequences in Nevada: the 2012-2015 Spanish Springs swarm near the population center of Reno, and the 2014–2016 Sheldon sequence in the remote northwest corner of the state. The GrowClust relocations of each sequence allow us to image the spatiotemporal evolution of each sequence along distinct fault structures that were previously obscured by the location uncertainty of the initial catalog. The results demonstrate the utility of high-precision relocations in uncovering the underlying physical mechanisms driving complex earthquake sequences. The GrowClust algorithm is publicly available as a open source software package to help facilitate future work in this regard (e.g., *Ross and Hauksson* 2017).

In Chapter 5, we shifted our attention to earthquake source properties, focusing in particular on dynamic (or Brune-type) stress drop, a source parameter which provides a scalar metric of the relative frequency content of each event. We developed an improved spectral decomposition algorithm that can be used to analyze the earthquake source spectra of large waveform datasets and provide robust estimates of the associated dynamic source parameters. We then applied this algorithm to characterize the scaling of stress drop with seismic moment for 2002–2016 seismicity in the Yuha Desert, San Jacinto Fault, Big Bear, Landers, and Hector Mine regions of southern California. In each region, we observed a tendency for stress drop to increase with moment within the magnitude range spanned by the data, in contrast to the classical model of self-similar, constant stress drop scaling first proposed by *Aki* (1967). The variability in the scaling

and stress drop distributions between and within each region give insight into the heterogeneity of crustal stresses within California's active fault systems and may provide useful observational constraints for future earthquake hazard assessments.

In Chapter 6, we built on these results by applying the spectral decomposition method to study the source properties of earthquakes in Harper and Sumner counties in southern Kansas. Seismicity rates in southern Kansas increased sharply over historical norms beginning in mid-2013. This change in seismicity is largely attributed to the proliferation of oil and gas production activity, and in particular the injection of industrial wastewater into the Arbuckle layer overlying in-situ faulting structures in the granitic basement formation. We observed that the recent (2014–2016) likely-induced earthquakes occurring in Harper and Sumner counties are characterized by low stress drop values compared to the naturally-occurring earthquakes in southern California. Though stress drop was observed to increase with hypocentral depth, this trend alone is insufficient to explain the discrepancy with California seismicity. Earthquakes in southern Kansas do however exhibit a scaling of stress drop with seismic moment that deviates from classical self-similarity and is comparable in intensity to the scaling trends observed in southern California. Characterization of the earthquake source properties of these and other likely-induced events in the oil-producing regions of the central United States will provide important observational constraints for hazard assessments in regions where rapid, nonstationary changes in seismicity are driven by human activity (*Ellsworth et al.*, 2015; *Petersen et al.*, 2016).

In Chapter 7, we examined in detail the relation between dynamic source properties and measured ground motion amplitudes. Regional hazard forecasts for long time horizons depend strongly on the assumed between-event variability in earthquake ground motions (*Bommer and Abrahamson*, 2006). Previous observational and theoretical studies have attributed much of the variability in ground motion for large events to differences in stress drop (*Hanks and McGuire*, 1981; *Boore*, 2003), but analogous results for smaller earthquakes remain more uncertain. In this study, we used the spectral decomposition algorithm described in Chapter 5 to compute estimates of dynamic stress drop for a new dataset of more than 5000 $M1.5$ and greater earthquakes occurring in the San Francisco Bay Area from 2002–2016. We then measured the peak ground acceleration from the full waveform records of the same set of earthquakes, and used a mixed-effects generalization of the Random Forest machine learning algorithm to correct the observed ground motion amplitudes for the joint influences of magnitude, source-station distance, and near-site effects. We observed a significant correlation between dynamic stress drop and the between-event residual in peak ground acceleration, confirming that the variability in dynamic source properties strongly controls the between-event variability in ground motion amplitudes. Spatial variations in earthquake stress drop are likewise correlated with residual ground motions, and thus could form the basis of future hazard forecasts that relax the ergodic assumption through source-specific ground motions (*Baltay et al.*, 2017). The observed elevation in the stress drop and residual ground motions of mainshock events compared to aftershock

events within individual earthquake sequences is intriguing and may prove useful in developing sequence-specific short-term operational hazard forecasts (*Jordan and Jones*, 2010; *Page et al.*, 2016).

## 8.2 Unresolved Questions and Future Directions

The results presented within this thesis bring to light several fundamental but as-of-yet unresolved questions that deserve further attention in future research studies. Principal among these are the physical mechanisms that generate the marked variability in earthquake clustering patterns documented both here and elsewhere. In some instances, the time-evolution of earthquake sequences is well-described by an *Omori* (1894) temporal decay in aftershock activity following the mainshock event that releases the dominant portion of seismic energy. Other sequences are more swarm-like: extended in duration and lacking a dominant mainshock (*Hainzl*, 2002; *Vidale and Shearer*, 2006). There is a general consensus that both of these end-member clustering styles commonly occur in nature, but the physical conditions that distinguish them are less well understood. Individual earthquake swarms have been attributed to fluid transfer processes (*Chen et al.*, 2012; *Shelly et al.*, 2016) or aseismic fault slip (*Lohman and McGuire*, 2007; *Llenos et al.*, 2009). However, it is unclear whether either or both of these mechanisms provide necessary or sufficient conditions to produce earthquake swarm activity. Likewise, the role of anthropogenic stresses from geothermal energy (*Chen and*

*Shearer*, 2011) or oil and gas production (*Goebel*, 2015; *Goebel et al.*, 2016) remains a open but increasingly exigent question. The application of methods such as GrowClust (*Trugman and Shearer*, 2017a) may prove useful in better characterizing the spatial and temporal progression of these sequences to provide insight into the underlying triggering mechanisms.

The evidence presented in this thesis for nonstationary variability of earthquake source properties and deviations from self similar scaling may also serve as a starting point for other promising avenues of research. Spatial variation in median stress drop on regional and local length scales appears pervasive (e.g. *Shearer et al.* 2006; *Allmann and Shearer* 2009; *Oth* 2013; *Poli and Prieto* 2016; *Trugman and Shearer* 2017b; *Abercrombie et al.* 2017). Whether these spatial patterns are persistent over time is less well constrained but equally important to resolve. Further, the inferred scaling relations and spatial variability in source properties are controlled primarily by the statistical characteristics of the small and moderate ($M \leq 4$) earthquakes that comprise the bulk of the datasets studied here. Larger earthquakes are thankfully rare, but their infrequency makes it difficult to assess how well the inferences from the behavior of smaller earthquakes generalize to the largest and most damaging of earthquakes. Observations of ground motion amplitudes suggest that large earthquakes have smaller variability in ground motions than smaller events (*Youngs et al.*, 1995; *Bozorgnia et al.*, 2014; *Douglas and Edwards*, 2016). If the same heteroskedasticity applies to dynamic source parameters like earthquake stress drop, it would help elucidate the scale-dependence of earth-

quake rupture and put physical bounds on the distribution of stress heterogeneity in Earth's crust.

Ultimately, the major findings from these and other studies of dynamic source parameters can and should be incorporated to improve the accuracy and reliability of the next generation of seismic hazard forecasts (*Jordan and Jones*, 2010; *Field et al.*, 2017; *Mulargia et al.*, 2017). The connection between stress drop and between-event variability in ground motion motivates the development of non-ergodic ground motion prediction equations (*Anderson and Brune*, 1999; *Landwehr et al.*, 2016; *Baltay et al.*, 2017). Earthquake sources located within regions of relatively high median stress drop would be expected to produce stronger shaking, but this is likely a frequency-dependent and possibly a magnitude-dependent effect that warrants more comprehensive study and rigorous testing. Earthquakes are complex phenomena that are inherently difficult to characterize and understand. We as seismologists must always stay humble and be willing to challenge our own models and assumptions in light of new data.

# References

Abercrombie, R. E., S. Bannister, J. Ristau, and D. Doser (2017), Variability of earthquake stress drop in a subduction setting, the Hikurangi Margin, New Zealand, *Geophysical Journal International*, *208*(1), 306–320, doi:10.1093/gji/ggw393.

Aki, K. (1967), Scaling law of seismic spectrum, *Journal of Geophysical Research*, *72*(4), 1217–1231, doi:10.1029/JZ072i004p01217.

Allmann, B. P., and P. M. Shearer (2009), Global variations of stress drop for moderate to large earthquakes, *Journal of Geophysical Research: Solid Earth*, *114*(B1), B01,310, doi:10.1029/2008JB005821.

Anderson, J. G., and J. N. Brune (1999), Probabilistic Seismic Hazard Analysis without the Ergodic Assumption, *Seismological Research Letters*, *70*(1), 19–28, doi:10.1785/gssrl.70.1.19.

Baltay, A. S., T. C. Hanks, and N. A. Abrahamson (2017), Uncertainty, Variability, and Earthquake Physics in Ground-Motion Prediction Equations, *Bulletin of the Seismological Society of America*, *107*(4), 1754–1772, doi:10.1785/0120160164.

Bommer, J. J., and N. A. Abrahamson (2006), Why Do Modern Probabilistic Seismic-Hazard Analyses Often Lead to Increased Hazard Estimates?, *Bulletin of the Seismological Society of America*, *96*(6), 1967–1977, doi:10.1785/0120060043.

Boore, D. M. (2003), Simulation of Ground Motion Using the Stochastic Method, in *Seismic Motion, Lithospheric Structures, Earthquake and Volcanic Sources: The Keiiti Aki Volume*, edited by Y. Ben-Zion, Pageoph Topical Volumes, pp. 635–676, Birkhauser Basel, doi:10.1007/978-3-0348-8010-7_10.

Bozorgnia, Y., N. A. Abrahamson, L. A. Atik, T. D. Ancheta, G. M. Atkinson, J. W. Baker, A. Baltay, D. M. Boore, K. W. Campbell, B. S.-J. Chiou, R. Darragh, S. Day, J. Donahue, R. W. Graves, N. Gregor, T. Hanks, I. M. Idriss, R. Kamai, T. Kishida, A. Kottke, S. A. Mahin, S. Rezaeian, B. Rowshandel, E. Seyhan, S. Shahi, T. Shantz, W. Silva, P. Spudich, J. P. Stewart, J. Watson-Lamprey, K. Wooddell, and R. Youngs (2014), NGA-West2 Research Project, *Earthquake Spectra*, *30*(3), 973–987, doi:10.1193/072113EQS209M.

Cattania, C., S. Hainzl, L. Wang, F. Roth, and B. Enescu (2014), Propagation of Coulomb stress uncertainties in physics-based aftershock models, *Journal of Geophysical Research: Solid Earth*, *119*(10), 7846–7864, doi:10.1002/2014JB011183.

Chen, X., and P. M. Shearer (2011), Comprehensive analysis of earthquake source spectra and swarms in the Salton Trough, California, *Journal of Geophysical Research*, *116*(B9), doi:10.1029/2011JB008263.

Chen, X., P. M. Shearer, and R. E. Abercrombie (2012), Spatial migration of earthquakes within seismic clusters in Southern California: Evidence for fluid diffusion, *Journal of Geophysical Research*, *117*(B4), doi:10.1029/2011JB008973.

Douglas, J., and B. Edwards (2016), Recent and future developments in earthquake ground motion estimation, *Earth-Science Reviews*, *160*, 203–219, doi:10.1016/j.earscirev.2016.07.005.

Ellsworth, W., A. Llenos, A. McGarr, A. Michael, J. Rubinstein, C. Mueller, M. Petersen, and E. Calais (2015), Increasing seismicity in the U. S. midcontinent: Implications for earthquake hazard, *The Leading Edge*, *34*(6), 618–626, doi:10.1190/tle34060618.1.

Field, E. H., K. R. Milner, J. L. Hardebeck, M. T. Page, N. v. d. Elst, T. H. Jordan, A. J. Michael, B. E. Shaw, and M. J. Werner (2017), A Spatiotemporal Clustering Model for the Third Uniform California Earthquake Rupture Forecast (UCERF3-ETAS): Toward an Operational Earthquake Forecast, *Bulletin of the Seismological Society of America*, *107*(3), doi:10.1785/0120160173.

Goebel, T. (2015), A comparison of seismicity rates and fluid-injection operations in Oklahoma and California: Implications for crustal stresses, *The Leading Edge*, *34*(6), 640–648, doi:10.1190/tle34060640.1.

Goebel, T. H. W., S. M. Hosseini, F. Cappa, E. Hauksson, J. P. Ampuero, F. Aminzadeh, and J. B. Saleeby (2016), Wastewater disposal and earthquake swarm activity at the southern end of the Central Valley, California, *Geophysical Research Letters*, doi:10.1002/2015GL066948.

Hainzl, S. (2002), Indications for a successively triggered rupture growth underlying the 2000 earthquake swarm in Vogtland/NW Bohemia, *Journal of Geophysical Research*, *107*(B12), doi:10.1029/2002JB001865.

Hanks, T. C., and R. K. McGuire (1981), The character of high-frequency strong ground motion, *Bulletin of the Seismological Society of America*, *71*(6), 2071–2095.

Jordan, T. H., and L. M. Jones (2010), Operational Earthquake Forecasting: Some Thoughts on Why and How, *Seismological Research Letters*, *81*(4), 571–574, doi:10.1785/gssrl.81.4.571.

Landwehr, N., N. M. Kuehn, T. Scheffer, and N. Abrahamson (2016), A Nonergodic Ground-Motion Model for California with Spatially Varying Coefficients, *Bulletin of the Seismological Society of America*, *106*(6), 2574–2583, doi: 10.1785/0120160118.

Llenos, A. L., J. J. McGuire, and Y. Ogata (2009), Modeling seismic swarms triggered by aseismic transients, *Earth and Planetary Science Letters*, *281*(1-2), 59–69, doi:10.1016/j.epsl.2009.02.011.

Lohman, R. B., and J. J. McGuire (2007), Earthquake swarms driven by aseismic creep in the Salton Trough, California, *Journal of Geophysical Research*, *112*(B4), doi:10.1029/2006JB004596.

Mildon, Z. K., S. Toda, J. P. Faure Walker, and G. P. Roberts (2016), Evaluating models of Coulomb stress transfer: Is variable fault geometry important?, *Geophysical Research Letters*, doi:10.1002/2016GL071128.

Mulargia, F., P. B. Stark, and R. J. Geller (2017), Why is Probabilistic Seismic Hazard Analysis (PSHA) still used?, *Physics of the Earth and Planetary Interiors*, *264*, 63–75, doi:10.1016/j.pepi.2016.12.002.

Omori, F. (1894), Investigation of aftershocks, *Rep. Earthquake Inv. Comm*, *2*, 103–139.

Oth, A. (2013), On the characteristics of earthquake stress release variations in Japan, *Earth and Planetary Science Letters*, *377-378*, 132–141, doi:10.1016/j.epsl.2013.06.037.

Page, M. T., N. van der Elst, J. Hardebeck, K. Felzer, and A. J. Michael (2016), Three Ingredients for Improved Global Aftershock Forecasts: Tectonic Region, Time-Dependent Catalog Incompleteness, and Intersequence Variability, *Bulletin of the Seismological Society of America*, *106*(5), 2290–2301, doi: 10.1785/0120160073.

Petersen, M. D., C. S. Mueller, M. P. Moschetti, S. M. Hoover, A. L. Llenos, W. L. Ellsworth, A. J. Michael, J. L. Rubinstein, A. F. McGarr, and K. S. Rukstales (2016), Seismic-Hazard Forecast for 2016 Including Induced and Natural Earthquakes in the Central and Eastern United States, *Seismological Research Letters*, *87*(6), 1327–1341, doi:10.1785/0220160072.

Poli, P., and G. A. Prieto (2016), Global rupture parameters for deep and intermediate-depth earthquakes, *Journal of Geophysical Research: Solid Earth*, *121*(12), 8871–8887, doi:10.1002/2016JB013521.

Ross, Z. E., and E. Hauksson (2017), Orthogonal faults and bimodal seismicity rates from trifurcation of the San Jacinto fault, *Science Advances (submitted)*.

Shearer, P. M., G. A. Prieto, and E. Hauksson (2006), Comprehensive analysis of earthquake source spectra in southern California, *Journal of Geophysical Research*, *111*(B6), doi:10.1029/2005JB003979.

Shelly, D. R., W. L. Ellsworth, and D. P. Hill (2016), Fluid-faulting evolution in high definition: Connecting fault structure and frequency-magnitude variations during the 2014 Long Valley Caldera, California, earthquake swarm, *Journal of Geophysical Research: Solid Earth*, *121*(3), 1776–1795, doi:10.1002/2015JB012719.

Trugman, D. T., and P. M. Shearer (2017a), GrowClust: A Hierarchical Clustering Algorithm for Relative Earthquake Relocation, with Application to the Spanish Springs and Sheldon, Nevada, Earthquake Sequences, *Seismological Research Letters*, *88*(2A), 379–391, doi:10.1785/0220160188.

Trugman, D. T., and P. M. Shearer (2017b), Application of an improved spectral decomposition method to examine earthquake source scaling in southern California, *Journal of Geophysical Research: Solid Earth*, *122*(4), 2017JB013,971, doi:10.1002/2017JB013971.

Vidale, J. E., and P. M. Shearer (2006), A survey of 71 earthquake bursts across southern California: Exploring the role of pore fluid pressure fluctuations and aseismic slip as drivers, *Journal of Geophysical Research*, *111*(B5), doi:10.1029/2005JB004034.

Woessner, J., S. Jónsson, H. Sudhaus, and C. Baumann (2012), Reliability of Coulomb stress changes inferred from correlated uncertainties of finite-fault source models, *Journal of Geophysical Research*, *117*(B7), doi:10.1029/2011JB009121.

Youngs, R. R., N. Abrahamson, F. I. Makdisi, and K. Sadigh (1995), Magnitude-dependent variance of peak ground acceleration, *Bulletin of the Seismological Society of America*, *85*(4), 1161–1176.