How do earthquakes start? A seismological investigation into earthquake nucleation and determinism

R. L. Colquhoun

University College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2025

For Mummy and Daddy—
thank you for everything.

Acknowledgements

This work would not have been possible without the support of my supervisor, Jessica Hawthorne. From when I walked into her office and announced I wanted to do a DPhil on how earthquakes start, she has allowed me endless academic freedom, and for that I am particularly grateful. I brought Tom Garth in as second supervisor for his advice on source time functions, and he patiently waited for many years until I got there. Mike Kendall has provided prudent advice on careers and writing, and has helped me see the bigger picture when I felt like I may never escape statistics. He also taught me the final thing I learnt as an undergraduate, and perhaps the most useful thing for writing a geophysics thesis: how to punctuate equations!

In many senses of the phrase, I grew up in Oxford. Yes, I was born here, and had my first exposure to questions being answered with more questions as a school-child here (particular thanks to Mr Stephenson for that). But this department has played a significant role in my growth as a person and a scientist, from when I was 18 until today. In particular, Conall Mac Niocaill has always provided sage advice and encouragement, and I would not be the Earth Scientist I am today without him. Andrew Walker showed endless patience with my programming questions that ranged from the basic to the ridiculous. Our discussions on pedagogy and teaching have provided entertainment and education in equal measure and his measured nature is something I always wish to emulate. Laura Stevens was fortuitously assigned to be my college advisor as I began my DPhil, and they have provided wisdom and a listening ear throughout the years.

Between my third and fourth years, I did perhaps the best thing I could have done for my DPhil: took a year out! My time at Imperial allowed me to return refreshed and with a new perspective. My thanks to Gary Hampson for taking a chance on me, and to everyone I worked with there, especially the students. In particular, Ian Bastow who over the course of a year of lunches helped me see the wood for the trees and always had helpful advice for the classroom, marking and thesis writing! Adrian Muxworthy provided measured perspective on research and teaching and is responsible for teaching me that table captions always go above the table.

I have sat on many committees in my time at Oxford which have provided a fascinating (and sometimes alarming) insight into what makes this ancient institution tick. I hope that the department is an even better place thanks to the work we've

done. To my co-conspirators, particularly Ben Fernando, Gawain Antell and Hannah Sanderson—thanks for making change with me!

When I started my DPhil, I was randomly assigned to live at 18 Merton Street with 10 strangers. Looking back, it was good grounds for a reality TV show. I had enormous luck in meeting Fran, Nick and Gwion in that house, who provided fresh perspectives, endless fun facts and great chats. My friends from undergraduate (particularly Katie, Julia, Sofie, Sam, Georgie and Catherine) and school (Claudia, Alice, Jenny, Jenny and Alex) have provided welcome reminders of the world outside academia (though many of them have followed me back into it...).

My favourite thing about the department has been the friends I've made, in no particular order: Hannah, Lewis, Lot, Ally, Harri, Bei Bei, Nikki, Vicky, Vikki, Toby, Sarah, Sofia, Emily, Sophie, Odysseus, Rhiannon, Biao, Matthew, Miriam, Adewumni, Bram and Justin. (Ok, I admit there was some order—I just couldn't resist putting all the rhyming names together...) I was fortunate to meet Nikki when we did our PhD interviews and to then end up at the same college. Toby and Bei Bei completed the cross-over between Univ and Earth Sciences. The whole seismology group has asked probing questions and given new perspectives which have only made me a better scientist.

Dan, Alice and Ben deserve their own paragraph. Their support and advice has been invaluable, solving questions I had pondered for days, providing motivation when it was needed most, and entertaining me when nothing was going quite right. I look up to all of you.

My thanks to everyone I have met at conferences, and who I have visited at their home institutions. Determinism in particular is a controversial topic, and your thoughtful, thorough and good-faith questioning contributed to the rigour of this thesis. In particular, Daniel Trugman, Greg McLaskey and Camilla Cattania stand out for their helpful comments and patient discussions with me.

I have been fortunate to have support from many people in my home community. The patrons and staff of the Half Moon have been endlessly supportive and interested in my progress. Particular thanks must go to Gjorgji, who has made even the most difficult weeks easier with pizza and pats on the back.

Finally, most of all, my thanks to my family. My grandparents have supported me in every endeavour, even when I can only imagine how strange it seemed to an outsider. They have shown patience when Christmas and Easter had to be balanced with revision, and conferences interfered with birthdays—I promise it will be (a bit) more normal now! My parents held education in the highest regard and I am grateful for your endless encouragement and help that has let me reach all my dreams. From the days of Mummy doing my RE colouring homework and helping me learn spellings in four languages when I could barely manage one, to Daddy helping me build a seismometer, and finally to you both reading every word of this thesis, I couldn't have done any of this without you. I love you all.

Funding

This work was primarily funded by UK Research and Innovation through the Natural Environment Research Council's Doctoral Training Partnership at Oxford [NE/S007474/1], joint with an Oxford-Radcliffe Scholarship from University College. The donors of the Oxford-Radcliffe Scholarships remain anonymous, but I am indebted to their generosity.

I have been fortunate to travel the world as part of my DPhil, which I could not have done without supplementing my RTSG. My thanks to the University College Old Members' Trust travel grants and the College's Research support fund. I am also grateful to the Royal Astronomical Society, the British Geophysical Association Gray-Milne Travel Bursary, the Seismological Society of America, the UK Space Agency, CNRS and NASA. I was delighted to attend the Second National Convention on Geoscience and thank AGU and the National Science Foundation for funding me to attend such an important meeting into the legacy of geoscience and anti-racism efforts today.

Statement of Authorship

My thesis contains research that was completed in collaboration, which I give details of here.

Chapter 2: Correlations between earthquake magnitude and early features of rupture in a global dataset

This work is in prep for resubmission to JGR: Solid Earth as *Colquhoun, R. L. and Hawthorne, J. C. Correlations between earthquake magnitude and early features of rupture in a global dataset.*

Jessica Hawthorne supervised the project. I authored the paper and figures, wrote the code, and conducted the analysis.

Chapter 3: Patterns in source-time functions reflect variations in earthquake magnitude and inversion approach

Jessica Hawthorne and Tom Garth supervised the project. I authored the chapter and figures, wrote the code, and conducted the analysis.

Chapter 4: Using foreshocks to constrain earthquake nucleation: comparing foreshock to aftershock ratios in the Hikurangi subduction zone to results from ETAS

This work will be resubmitted to EPSL as *Colquhoun, R. L., Hawthorne, J. C. and Kamranzad, F. Using foreshocks to constrain earthquake nucleation: comparing foreshock to aftershock ratios in the Hikurangi subduction zone to results from ETAS.*

Jessica Hawthorne supervised the project and provided the original phase-coherence code. Farnaz Kamranzad ran the ETAS models. I authored the paper and figures, modified the phase coherence code, wrote supplementary code, and conducted the analysis.

Abstract

Scientists continue to debate many questions about the beginnings of earthquakes. At the single event scale, what processes go on in the earliest stages of rupture, and do they depend on magnitude? At a large scale, what processes cause earthquakes to occur when they do, and can foreshocks give insights into earthquake nucleation?

I begin by using short windows of seismic waveforms to examine the earliest parts of earthquakes around the world. I find that the predominant period, average period, and peak ground displacement scale with final magnitude even when calculated in windows just one-third of the earthquake duration. Meanwhile, the integral of the velocity squared scales with magnitude when calculated in windows just one-tenth of the earthquake duration. I propose physical origins for these relationships. These results suggest that small and large earthquakes begin differently and that earthquakes are deterministic.

Next, I investigate the differences between small and large earthquakes using source time functions (STFs), which show the time evolution of moment release. I find that larger earthquakes release more moment even in very short, early windows of time. I also find that STFs are generally complex with multiple peaks and have several clear phases of rupture. These STF properties have important implications for earthquake source modelling.

Finally, I investigate the extent to which earthquake-earthquake triggering can explain foreshock occurrence on the Hikurangi subduction zone, New Zealand. I compare new observations of seismicity to what is expected from an epidemic-type aftershock sequence (ETAS) model. I find that ETAS models explain the foreshock-to-aftershock ratio observed, but that the model underestimates the absolute number of foreshocks and aftershocks we observe and does not reproduce the power-law increase and decrease in seismicity we observe. Therefore, earthquake nucleation must be more complex, with stochastic triggering mediated by fluid flow or aseismic slip.

Contents

List of Figures

List of Tables

1 Introduction

Contents

1.1 Motivation

The processes that cause the ground to shake during earthquakes have been of fascination for centuries. The proposed explanations generally provide insights into

the prevailing rationales of their time, as well as highlighting the insidious nature of earthquakes in popular culture, from Shakespeare to manga.

For most of history, explanations for earthquakes centred on the divine. Many cultures believed earthquakes resulted from mystical animals supporting or brushing against the earth: weary *Ashtadiggaja* elephants in Hindu legend, twitching frogs in the stories of Mongolian lamas, itchy jaguars in the tales of the Tzotzil people of Mexico and attempts of a captured catfish to escape in Japan. Other cultures instead look to God(s) as explanations: for the ancient Greeks an irritated Poseidon was implicated, whilst Catholic Europe blamed heretics and adulterers for incurring the wrath of God (Robinson, 2012). The Buddhist *Taikyoku jishinki* described earthquakes as resulting from a blocking of the *qi* essence which balances the five elements of fire, water, wind, space and earth; with this imbalance resulting from people offending the *kami* (willful cosmic forces) (Smits, 2013). Others attributed earthquake occurrence to the behaviour of other people: colonial theft of tribal artefacts from the Yurok tribe, Korean immigration to Japan, and dynastic decline in China (Agnew, 2002; Robinson, 2012; Needham and Wang, 1959).

Nowadays, most people accept a physical explanation for earthquakes, and you would be forgiven for thinking this a modern development. In fact, within the frameworks of the time, physical theories for earthquakes arose as early as the sixth century BCE. Aristotle proposed earthquakes were essentially wind storms inside the earth (following the ideas of classical Greek elements; Oeser, 1992). He even developed the first classification scheme for earthquakes, based on the direction of shaking. Since the 18th century, the ideas which are now used to explain earthquakes have developed.

All these explanations speak to an intrinsic desire of humans from which we, as modern seismologists, are not exempt: the desire to understand the world around us. In this thesis, I seek to better understand the processes which occur before and during earthquakes.

1.2 Scientific approaches to how earthquakes start

I approach the titular question of 'how do earthquakes start' at two main scales. Considering individual earthquakes, I want to understand what the initial stages of earthquakes look like. More broadly, I want to understand what makes an earthquake happen at a given time. Chapters 3 and 4 are dedicated to the former, whilst Chapter 5 seeks to understand the latter.

1.2.1 Key terms

In the introduction that follows, I take many seismological terms as read. In Chapter 2, I outline some of the key theory that underpins the observations and modelling outlined here and in the rest of my thesis, which many readers may already be aware of. For now, I define two key terms.

Determinism When I ask what the initial stages of earthquakes look like, I am particularly interested in whether early stages of rupture are characteristic of the earthquake's final magnitude. If small and large magnitude earthquakes are fundamentally different in their earliest stages, then it can be said that earthquakes are deterministic.

Nucleation I use nucleation to refer to the initial stages of rupture, and the preparatory phases which may precede rupture.

1.2.2 Key questions

This thesis centres on three main questions:

1. Are the beginnings of large and small earthquakes different?
2. What controls slip evolution during earthquakes and are patterns in slip evolution magnitude dependent?
3. Do foreshocks result solely from earthquake-earthquake triggering or do they reflect a complex earthquake nucleation process?

In the past, many different approaches have been used to understand the processes of earthquake nucleation and determinism. I will now attempt to motivate

the central questions of this thesis by reviewing a majority of the work on these topics since the 1970s. By approaching this section roughly chronologically, I hope to illustrate the background and controversies which surround earthquake nucleation and determinism and that motivate my current work. In Section 1.3, I will outline the aims of each chapter of my thesis, in the context of the previous work discussed.

1.2.3 1960s

Significant research on seismicity happened in the 1960s and earlier, including the paradigm shift that was plate tectonics. Many of these ideas like the Gutenberg-Richter relation are now so fundamental that we take them as fact in introductory seismology classes, and thus I do not review them here. There are two aspects of research that I will briefly discuss here. Firstly, throughout the 1960s and earlier, there was a significant body of statistical work undertaken which now underpins much of our modelling of earthquake-earthquake triggering. However, at the time, it had not yet found its home in the modelling of seismicity. Secondly, and of particular relevance for the ideas I will come to, was the work of Haskell (1964), who found that the energy radiated by earthquakes could be explained with relatively short dislocation rise times. This implies that the slip at any given point on the fault occurs in a short time frame compared to the total duration of the earthquake. I will return to this idea over the coming decades of research.

1.2.4 1970s

Observing earthquakes is becoming ever easier, with improvements in instrumentation and our computational ability to process the large volumes of data inherent in seismology. But in the 1970s (and indeed for several decades thereafter) most studies focused on one or a few events, often spatially grouped. One of the earliest observational analyses of the early stages of rupture occurred in 1974, when Dziewonski and Gilbert (1974b) identified that two deep earthquakes experienced compression before their abrupt onset. This was controversial (Geller, 1974; Dziewonski and Gilbert, 1974a) but spurred a series of investigations over the

following decades into the potential existence of a slow nucleation phase such as Sacks et al. (1978). Kanamori and Anderson (1975) found that dislocation rise times are relatively short compared to earthquake durations, agreeing with the models of Haskell (1964) and Brune (1970). Brune (1970) had qualitatively proposed a self-healing slip pulse model to explain the differences between apparent stress drops calculated from seismic waves and static stress drops which measured the overall stress change on the fault.

Concurrently with these observations, theoretical modelling efforts were more conclusive that failure should begin with a period of slow, stable sliding in a small area of the fault (Dieterich, 1978), though it was noted that anelastic attenuation could make observed displacements more gradual than they actually are (Kanamori and Anderson, 1977). Laboratory studies also began, allowing the simulation of shear failure on real rocks (Dieterich, 1978), finding that seismic rupture may be preceded by two or more phases of aseismic slip. By modelling rupture with a constant rupture velocity and stress drop, Sato and Hirasawa (1973) found that peak ground displacement should scale as the square of time. They also found the spectral decay to be ω^{-2} at high frequencies.

By this point, the prevalence of foreshock activity before earthquakes had been observed, providing a different insight into the earthquake nucleation process. Jones and Molnar (1976) finding that 44% of M7+ events had teleseismically recorded foreshocks. This matched short-time-scale small-spatial-scale observations of earthquake precursors in laboratory experiments (Scholz, 1968).

1.2.5 1980s

In the 1980s, the efforts of the 1970s to search for an aseismic nucleation phase continued. This included further modelling efforts (Das and Scholz, 1981; Dieterich, 1986) and laboratory studies (Ohnaka et al., 1986), which continued to predict a slow nucleation phase.

At the same time, the development of the first early warning systems for earthquakes had begun in Japan. Nakamura and Tucker (1988) proposed that the

frequency content of the earliest portions of the waveform was indicative of the final earthquake magnitude, and in doing so, proposed the predominant frequency calculation, which would be fundamental to much of the work of the 2000s.

In 1988, Ogata reviewed several classes of statistical models for earthquake occurrence and found that epidemic-type models provide the best fit to observed seismicity. In this case, no extended nucleation is required to explain the observations. Epidemic models (a specific case of which is denoted ETAS, see Chapter 5) are essentially statistical earthquake-earthquake triggering models, where earthquakes are modelled as point processes (Hawkes, 1971) with a magnitude dependent on the triggering capability of earthquakes. Generally mainshocks are modelled as Poisson processes (that is, independent and randomly located at a given rate in time and space).

1.2.6 1990s

Heaton (1990) calculated time-dislocation histories of seven earthquakes and found that rupture tends to be more pulse-like with short dislocation rise-times relative to the duration of the earthquake. He proposed that large amplitude slip pulses are difficult to stop and so ruptures that start more strongly are more likely to be able to overcome stress barriers and thus to finish as larger magnitude events.

In 1995, Ellsworth and Beroza proposed three models for the nucleation phase of earthquakes: self similar, pre-slip and cascade (Figure 1.1). In the self-similar model, the area of the fault grows in time and with earthquake size. In the unbounded and homogeneous case shown in Figure 1.1, this means that the rupture radius grows linearly with time and leads to linear growth of the far-field velocity seismogram. In this, no nucleation phase is expected. In the cascade model, the earthquake is composed of several sub-events. The spontaneous occurrence of sub-event 1 eventually leads to the failure of area 2 (potentially with some delay), and the rest of the fault then fails at time 3. All slip has the potential to be seismically observed. In this, there is no difference between the early stages of large and small earthquakes; a large earthquake occurs by chance in the cases that a small

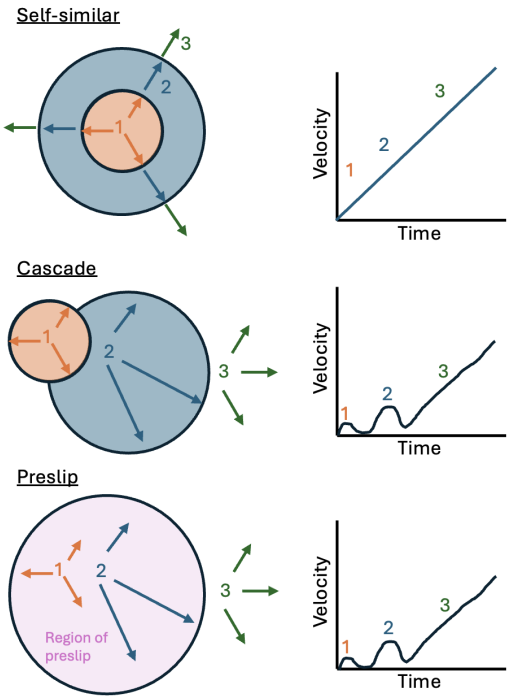


Figure 1.1: The three nucleation models described by Ellsworth and Beroza (1995).

earthquake triggers a cascade of increasingly larger slip events. On the other hand in the pre-slip model, some portion of the fault is 'prepared' through an episode of aseismic slip (pink region). Subevents 1 and 2 occur within this zone but at time 3 the rupture breaks out of the preparation zone. In this, large and small earthquakes may be distinguishable if their nucleation zones are of different sizes (i.e. based on the size of subevents 1 and 2). Differentiating between the cascade and pre-slip model is difficult using far-field seismograms.

Throughout the decade, a whole host of observations were made of the early stages of earthquakes, looking for two main attributes: whether features of the onsets of earthquakes scaled with magnitude, and for any evidence of a nucleation phase. Umeda (1990) observed a two phase onset to M5+ magnitude earthquakes, where the duration of the low-amplitude phase scaled with earthquake magnitude. This was one of the first observations that suggested earthquake nucleation might be deterministic. Conversely, Abercrombie and Mori (1994) found that the onsets of the M7.3 Landers earthquake and one of its M4.4 aftershocks were similar, whilst Anderson and Chen

(1995) found that M3–8 earthquakes in Mexico all had similar initiations.

Iio (1995) looked at microseismicity and observed that the far-field P wave velocity pulse had a slow initial phase before undergoing more rapid acceleration. These results were similar to Iio (1992) who found that the slow initial rise cannot be explained by models with constant rupture velocity or kinematic friction, but is instead best explained by a slip-weakening crack model. Ellsworth and Beroza (1995) observed a delay between the initial P wave and the sudden increase in growth of near-field velocity seismograms. On moment-acceleration curves, this appears as a slow growth portion, which ruled out the self-similar case (Ellsworth and Beroza, 1995). Incorporating the results of Umeda (1990) and Iio (1995), Ellsworth and Beroza (1995) observed a scaling between the radius of the nucleation zone and final event magnitude, leading them to favour the pre-slip model. Umeda et al. (1996) found a scaling between final earthquake moment and the time delay between the initiation of rupture and the rupture of a 'bright spot' (Umeda, 1992). Dodge et al. (1996) found that complex nucleation of the type observed by Ellsworth and Beroza (1995) might arise from aseismic nucleation processes. However, Mori and Kanamori (1996) again found that the initiation of the velocity waveforms was independent of the magnitude of the earthquake and suggested that the curvature (slow initial phase) observed by Ellsworth and Beroza (1995) could be explained by anelastic attenuation. They also suggested that any seismically observable 'precursory' radiation must result from a source no larger than that which would cause a M0.5 earthquake and thus would not be observable with the instrumentation of the time. Kilb and Gomberg (1999) found that the initial subevent of the Northridge earthquake had similar waveforms to nearby small earthquakes and that these did not scale with mainshock magnitude. Ellsworth and Beroza (1998) examined 5 events near Ridgecrest and found evidence for a seismic nucleation phase with a magnitude independent rapid onset (agreeing with Mori and Kanamori (1996) and disagreeing with Iio (1995) and Ellsworth and Beroza (1995)).

Beroza and Ellsworth (1996) examined source time functions and found a low-moment acceleration nucleation phase, the duration of which scales with final event

moment. Houston et al. (1998) also looked at source time functions and as expected found that overall source duration scaled with magnitude. They also found that shallow events tended to be longer than deep events, and that intermediate depth events tended to be more asymmetric and unexpectedly complex compared to deep events. Tanioka and Ruff (1997) produced a catalogue of source time functions of 255 moderate to great earthquakes from around the world.

During the 1990s, a number of groups also made creative efforts to computationally model the nucleation of earthquakes. Singh et al. (1998) used a composite source model which essentially represents the cascade model of Ellsworth and Beroza (1995). They found that a model of this type was able to capture the evolution of the earthquake source, including a nucleation phase which scales with final earthquake moment similar to the observations of Umeda et al. (1996), Ellsworth and Beroza (1995) and Beroza and Ellsworth (1996). They suggested that the various nucleation phases observed may represent the rupture of the first subevent, whilst the rapid increase in velocity amplitude may represent the rupture of the largest subevent. Meanwhile, Steacy and McCloskey (1998) used a 'heterogenous cellular automaton' to find that the final magnitude of the earthquake depends on local stress heterogeneities on the fault, rather than intrinsic differences between large and small earthquakes. Finally Sato and Kanamori (1999) modelled rupture arrest using the Griffiths' failure criterion and found this did not support a magnitude dependent nucleation phase, similar to the findings of Deichmann (1997).

Foreshocks continued to be observed throughout the 1990s and were used to probe the existence of a nucleation phase. Foreshocks were found to be expected from a rate-and-state model framework (Dieterich, 1994) tuned using laboratory observations (Marone, 1998), and their statistics continued to be analysed (Abercrombie and Mori, 1996). Studies of individual foreshock sequences found that foreshocks seem to cluster near structural discontinuities and may directly result from aseismic nucleation processes (Dodge et al., 1996). Defining the nucleation region using the foreshock sequences, Dodge et al. (1996) also found that the size of the nucleation region scaled with final earthquake magnitude. However, this nucleation region

continued to be difficult to detect aseismically, such as by using strainmeters (Johnston et al., 1990, 1994; Abercrombie et al., 1995).

1.2.7 2000s

Kanamori and Mori (2000) modelled cascade nucleation using percolation processes and suggested that it might be possible to quantify how close a region of fault is to its critical state. They proposed that fluid flow in the crust might be associated with earthquake precursors like foreshocks. Brodsky and Kanamori (2001) suggested that the large variation in scaled rupture energy between big and small earthquakes might result from a thin layer of high pressure viscous fluid on the fault. This elastohydrodynamic lubrication could also lead to reduced high-frequency radiation content when measured more than a few meters away from the fault.

Meanwhile, the observational focus had largely turned to whether measurements of the early portions of waveforms were indicative of the final magnitude of earthquakes. These parameters had potential for use in earthquake early warning which was by this time operation in Japan, Taiwan and Mexico. Nakatani et al. (2000) found that the P wave amplitude scales with final earthquake moment even in very short windows where the P wave is still accelerating for all events. They used synthetic seismograms to see if this result could be reproduced if the source time function grew magnitude-independently but anelastic attenuation was accounted for (following Mori and Kanamori, 1996), and could not produce results consistent with all observations. However, a model with size-dependent growth rate was able to match the observations. Allen and Kanamori (2003) investigated 53 earthquakes in California (M3–7). They built on the work of Nakamura and Tucker (1988) to define the predominant period parameter ($\tau_p^{\max}$). The predominant period is expected to scale with magnitude because small magnitude events are expected to slip on a small patch and thus to release higher-frequency energy compared to larger magnitude events that slip on larger patches. As predicted, they observed a positive relationship between final magnitude and $\tau_p^{\max}$ measured in windows of 2 and 4 s after the P wave arrival. Allen (2003) found a similar relationship and began to

develop attenuation corrections to allow the application of τ_p^{max} in early warning. Kanamori (2004) formulated a non-recursive measure of the average period of a waveform, τ_c, and found that this too scaled with magnitude when calculated in windows of 3 s, and showed no sign of saturation with $M > 7$. Dynamic rupture modelling by Lapusta and Rice (2003) found that moment acceleration in the early stages of seismic propagation varied, similar to the observation of Ellsworth and Beroza (1995), but that at least part of this variation is explained by heterogeneous stress on the fault. This heterogeneity arises from previous small events and by the concentration of stress by fault creep.

In 2005, Olson and Allen incited debate when they presented their analysis of 71 earthquakes from around the world using τ_p^{max}. They found a clear relationship between τ_p^{max} and magnitude when the parameter was calculated in a 4 s window. They also found a relationship between magnitude and the time taken to reach the maximum τ_p value, although this was weaker. This showed that measurements taken in windows less than one-tenth of the earthquake duration were still predictive of final magnitude (Olson and Allen, 2005, Figure 3b). Lockman and Allen (2005) found similar results using single station measurements of Californian earthquakes. Meanwhile, Wu and Kanamori (2005a) and Wu and Kanamori (2005b) extended the work of Kanamori (2004) using τ_c and a new parameter, the peak ground displacement (P_d), on earthquakes in Taiwan. They found a relationship between magnitude and both τ_c and P_d but with a significant degree of scatter.

So far, there had generally been consensus that these early seismogram parameters scale with final event magnitude. However, Rydelek and Horiuchi (2006) investigated M6+ events in Japan and found no relationship between final magnitude and τ_p^{max}. They claimed that there is a break in scaling around M5.5 in the results of Olson and Allen (2005), and at higher magnitudes the relationship is statistically dubious. Using a sample containing more large earthquakes, they concluded that τ_p^{max} does not scale with magnitude. There were also some concerns about the stability of the τ_p^{max} parameter (Wolfe, 2006; Shieh et al., 2008). Wolfe (2006) found that the parameter gives increased weighting to higher amplitudes and

frequencies, and so averaging τ_p over stations does not give a true average of the frequency content of an earthquake.

Meanwhile Wu and coauthors continued their previous work by analysing earthquakes in Southern California and found scalings between magnitude and τ_c, P_d and $\tau_p^{\max}$ (Wu et al., 2006; Wu and Zhao, 2006). Zollo et al. (2006) found that P_d in a 2 s window after the P wave scales with magnitude without saturation for earthquakes in the Mediterranean. However, Rydelek et al. (2007) disagreed and used accelerometer records from Japan to argue for a plateau in P_d at M5.5. In their response, Zollo et al. (2007) used the same database as Rydelek et al. (2007) to argue that P_d in a 2 second window does saturate, but not until around M6.5, and measurements in a 4 s window show no saturation. Lockman and Allen (2007), Tsang et al. (2007), and Wurman et al. (2007) all continued to lend credence to the use of $\tau_p^{\max}$ as a predictor for final event magnitude in both subduction zone and transform fault environments. At the same time, Wesnousky (2006) used geological observations to suggest that the underlying fault structure is the predominant control on final earthquake magnitude. He found that fault steps often serve to stop earthquake rupture with fault steps accounting for the end of two-thirds of strike-slip earthquakes.

Perhaps spurred on by the debate between Zollo et al. (2006) and Rydelek et al. (2007), in 2008 attention primarily moved to Japan. Shieh et al. (2008), Wu and Kanamori (2008a) and Wu and Kanamori (2008b) found relationships between τ_c, $\tau_p^{\max}$ and P_d with no evidence of saturation, whilst Lancieri and Zollo (2008) found a saturation at M6.5 when P_d was calculated in a 2 s window. Festa et al. (2008) were perhaps the first to explore the integral of the velocity squared (IV2) and found that this too scaled with magnitude, as did the ratio $P_d^2/IV2$. They found some evidence of IV2 calculated in a 4 s window saturating at around M5.8. IV2 is related to the energy radiated by an earthquake. Relatedly, Nielsen (2007) proposed based on observations that the ultimate size of an earthquake is determined by its initial energy budget, agreeing with Heaton (1990). τ_c was used in real-time earthquake early warning tests in California (Bose et al., 2009) and

Taiwan (Hsiao et al., 2009). Shieh et al. (2008) proposed combining τ_c with $\tau_p^{\max}$ to reduce uncertainty on estimated magnitude and add robustness against filtering choices, particularly in comparison to using only $\tau_p^{\max}$.

In 2009, Allen et al. attempted to use $\tau_p^{\max}$ and P_d in a real-time simulation for earthquake early warning in California and were able to successfully detect and determine the magnitude of two M5.4 earthquakes. The algorithm used was called ElarmS. Since California (thankfully) does not have many large magnitude events, Brown et al. (2009) tested the ElarmS algorithm in Japan and continued to find scalings between $\tau_p^{\max}$ and P_d and magnitude, although the relationship between P_d and magnitude was stronger than between $\tau_p^{\max}$ and magnitude. P_d was also used in real-time earthquake early warning testing in California (Bose et al., 2009), Taiwan (Hsiao et al., 2009), Turkey (Fleming et al., 2009) and Italy (Lancieri and Zollo, 2008). Through numerical simulations, Murphy and Nielsen (2009) found that peak ground displacement scales with magnitude for ruptures less than four times the size of the initial asperity, but plateaued for larger events.

Following Houston et al. (1998), work in the 2000s also continued on source time function features. Houston (2001), Persh and Houston (2004) and Tocheport et al. (2007) all analysed the geometries and statistical features of source time functions, such as the kurtosis, but did not investigate how these features scaled with magnitude (only with depth). In 2006, Sigloch and Nolet proposed a new method to produce source time functions for tomography based on matched-filter deconvolution.

Thinking about earthquake nucleation, Helmstetter and Sornette (2003) and Felzer et al. (2004) were both able to explain foreshock occurrence solely through earthquake-earthquake triggering. On the other hand, Liu and Rice (2007) proposed that earthquake nucleation and consequent foreshocks result from the interaction of pore-fluid pressure changes with the accelerating fault. McGuire et al. (2005) found elevated seismic activity (∼foreshocks) before large earthquakes on the East Pacific Rise.

1.2.8 2010s

As with most fields, there was an explosion of research on earthquake nucleation and determinism in the 2010s. Thus I consider four subtopics in turn: determinism, nucleation, source time functions and lab seismicity.

1.2.8.1 Studies of determinism

As the 2010s began, development of the ElarmS earthquake early warning algorithm continued with real-time testing in California and offline testing in Japan using $\tau_p^{\max}$ and P_d (Brown et al., 2011). Lancieri et al. (2011) analysed events in Chile using IV2, P_d, τ_c and $\tau_p^{\max}$ and found that these parameters scaled with magnitude but that the relationship was influenced by the type of seismometer used (accelerometer vs velocity). In contrast to Lancieri and Zollo (2008), who found that P_d saturated at $M > 5.8$, Lancieri et al. (2011) found a relationship between P_d and magnitude even though they only considered M6+ earthquakes. Peng et al. (2013) similarly found scalings between magnitude and all four of the parameters I have discussed thus far: τ_c, IV2, P_d, and $\tau_p^{\max}$. Kuyuk and Allen (2013) explored the best distance normalisation for P_d to correct for geometric spreading and found this was $R^{1.38}$ rather than the R^2 we might usually use for geometric spreading. They also looked at the ratio between peak ground displacement and peak ground acceleration (P_d/P_v) and found this scaled with magnitude. For earthquakes shorter than the window duration, they found that $\tau_p^{\max}$ does not scale with the final magnitude.

Following Abercrombie and Mori (1994), it was found that the foreshocks of the 2011 Tohoku earthquake looked similar to the mainshock (Hoshiba and Iwakiri, 2011; Colombelli et al., 2012). Similar results were found for the 2008 Wenchuan earthquake in a paper that was retracted for plagiarism but for which the results still stand (Peng et al., 2014). Colombelli et al. (2012) attempted to use P_d and τ_c measurements to estimate the final magnitude of the Tohoku earthquake and arrived at an estimate of M8.4. However, this estimation required ~ 50 s of waveform.

Gabriel et al. (2012) used dynamic rupture simulations to show that small changes in background stress or the earthquake nucleation size can result in earthquakes with

very different rupture styles (e.g. rise time, rupture speed, complexity and stability), and thus that these cannot be an intrinsic indicator of final event magnitude. Goldsby and Tullis (2011) showed that flash heating (and thus weakening) play an important role in the pulse-like model of earthquake rupture. Noda and Lapusta (2013) found that earthquake ruptures can propagate through creeping segments meaning that these do not always present a barrier to propagation. Whilst this result does not directly impact determinism, it lends credence to the ideas of Heaton (1990) and Nielsen (2007) in that final earthquake size is controlled by the initial energy budget. Relatedly, Bizzarri (2011) modelled earthquake rupture using two different friction laws (idealized, linear slip-weakening (SW) and a rate-and-state-dependent friction law) and found the key factor that influences observable metrics (e.g. stress drop, peak slip velocity) was the fracture energy budget. This suggests that seismically distinguishing different models of rupture is difficult if they are energetically identical.

Colombelli et al. (2014) found that P_d measured in short time windows scales with final magnitude, but that this scaling is more rapid for small events than large. This suggests that there may be some intermediary between the pro- and anti- saturation camps.

Most earthquake early warning systems rely on determining source properties (e.g. magnitude) using early portions of the seismogram. Hoshiba and Aoki (2015) suggested that early warning alerts could instead solely be based on ground motion.

Meier et al. (2016) found that rupture onset, as measured by P_d, is magnitude independent by using very near-source records (hypocentral distance less than 25 km). They attribute previous workers' results to using more distant recordings, where the onsets no longer correspond to direct P waves (Aki and Richards, 2009), and to averaging across stations. They also found that there is a distinct change in P_d growth rate around 0.1–0.2 s with a switch from $P_d \propto t^{\sim 3}$ to $P_d \propto t^{\sim 3/2}$. Neither of these are explained by a simple self-similar crack model with constant rupture velocity (e.g. Sato and Hirasawa, 1973). They suggest that these may reflect the non-instantaneous acceleration of rupture velocity to its steady state value (Deichmann,

1997; Sato and Kanamori, 1999) or may simply result from anelastic attenuation, and that the scaling at the source may always follow $P_d \propto t^{\sim 3/2}$ but sometimes appears to be $P_d \propto t^{\sim 3}$. This latter idea is similar to that proposed by Mori and Kanamori (1996). However, the change in scaling may also be a true observation, in line with observations by Umeda (1990), Iio (1995) and Ellsworth and Beroza (1995). This idea of a slow onset was further explored by Noda and Ellsworth (2016) who measured the time at which the (P wave) amplitude transitions from being similar (≈self-similar), to dependent on magnitude. They call the time at which this transition takes place relative to the P wave arrival, the delay-time (T_{dp}). They see that the absolute displacement, and the average absolute displacement for magnitude-hypocentral distance bins increase systematically after the P wave onset, until a point, at which they begin to slowly decrease. The time between the P wave arrival and the time at which these parameters start to decrease is called T_{dp}.

Calculating T_{dp} on the average absolute displacement, rather than individual records, removes the effect of radiation pattern and directivity, which would otherwise affect the value of T_{dp} differently at each station, and thus makes the resulting relationship more robust. This essentially accounts for much of the concern of Meier et al. (2016) whilst still allowing averaging, since different hypocentral distance recordings are kept separate.

As delay time defines some change in the rupture, Noda and Ellsworth (2016) therefore suggest that it is important to consider the delay time in selecting the time window to use in our other calculations (e.g. τ_p). If the time window used to calculate τ_p is less than T_{dp}, τ_p saturates at higher magnitudes (e.g. Lancieri et al., 2011, in Chile). This is because the full period of the 'fast' growth of the rupture is not sampled for the largest events. Conversely, if the time window used is longer than T_{dp}, τ_p appears to 'bottom out', as the slower growing section begins to be sampled (e.g. Kuyuk and Allen, 2013).

In an argument against determinism, but remaining somewhat positive for early warning, Leyton et al. (2018) found that using longer windows of measurement results in better magnitude estimates for earthquake early warning. Trugman et al.

(2019) measured peak ground displacement and found that peak ground displacement scales with magnitude when the measurement window is more than $\sim 1/3$ of the total earthquake duration. In shorter windows, P_d is found to be magnitude independent.

1.2.8.2 Studies of earthquake nucleation and foreshocks

In 2018, Cattania et al. compared static coulomb-stress-based physics models for earthquake forecasting to empirical statistical models (ETAS) during the 2010–2012 Canterbury, New Zealand, earthquake sequence. They found that Coulomb-based models give comparable forecasts compared to ETAS, but that a combined physics-statistical model performs better. They also find that a locally tailored ETAS model with aftershock zones based on the local fault geometry provides an improvement over traditional ETAS models. Cattania et al. (2018) did not model the influence of fluid flow. Cochran et al. (2018) looked at injection induced seismicity and found that pore-fluid pressure changes were important close to the wells but that earthquake-earthquake triggering was important further from the injection source.

Yoon et al. (2019) found that foreshocks of the Hector Mine earthquake can be explained through earthquake-earthquake triggering, whilst van den Ende and Ampuero (2020) found that very few earthquakes have foreshocks at all, with just 18% of foreshock sequences in Southern California being unexplained by background seismicity. This was in contrast with Trugman and Ross (2019) who looked in the same region and found that 72% of mainshocks have foreshocks. Bouchon et al. (2013) found that large magnitude events in the North Pacific were preceded by an increase in seismic activity, whilst Kato et al. (2016) found the same in Kyushu, Japan. Chen and Shearer (2016) found that foreshocks largely result in response to regional tectonic stresses and fault zone properties. Considering seismicity clusters of different magnitudes, they found that 66% of large magnitude clusters have foreshocks, whilst 30% of small magnitude clusters go on to lead to large clusters.

Bouchon et al. (2011) found that the İzmit earthquake was preceded by a 44-minute long seismic signal emanating from the hypocentre, with repetitive seismic bursts (foreshocks) and increased low-frequency seismic noise. The slip accelerated

further just before the mainshock. In contrast, Ellsworth and Bulut (2018) found that no aseismic slip is needed to explain the 44-minute long foreshock sequence observed. They argue that the foreshock geometry lends credence to earthquake-earthquake triggering being at play, since the events form a systematic sequence from west to east towards the mainshock hypocentre. However, this linear progression could also result from fluid flow or aseismic slip propagating in a single direction. Tape et al. (2018) found foreshock sequences in Alaska which may have resulted from a propagating aseismic front. Chen and Shearer (2013) analysed foreshock sequences of three $M > 7$ mainshocks in southern California. They found that the sequences were similar to swarms in their magnitude-time patterns and their spatial migration of seismicity. The foreshock sequences also have lower stress drops than the corresponding aftershock sequences. Chen and Shearer (2013) interpret these observations to suggest that foreshock sequences are a result of aseismic transients close to the mainshock hypocentres.

Brodsky (2011) found that the spatial density of foreshocks in Southern California can be explained solely by earthquake-earthquake triggering. They did observe a much lower foreshock-aftershock ratio than predicted by earthquake-earthquake triggering encapsulated in an ETAS model, but suggest that this could have arisen from catalogue incompleteness. Shearer (2012) modelled foreshock and aftershock occurrence assuming that the increased triggering power of larger magnitude events is exactly compensated for by their decreased numbers ($\alpha = b$). They found a foreshock-aftershock ratio which is too high to be explained by earthquake-earthquake triggering following Båth's Law (on average the difference between the magnitude of the mainshock and the largest aftershock should be 1.2 magnitude units) and a higher absolute number of aftershocks per mainshock than expected. This suggests that earthquake-earthquake triggering is not a primary control on space-time clustering of seismicity or that triggering self-similarity breaks down at small magnitudes. They also observe a lower b-value than expected for aftershocks, but this could arise from catalogue incompleteness (Hainzl, 2013; Shearer, 2013). Seif et al. (2018) performed a statistically rigorous analysis of foreshock occurrence

in California and Italy, and found more foreshocks than expected by ETAS model which fit the regional aftershock observations well ($\alpha = 0.05$).

1.2.8.3 Studies of source time functions

The 2010s had a rush of different source time function inversion methods developed. Vallée et al. (2011) developed an automated deconvolutional approach to determine seismic moment magnitude and source time functions of large earthquakes without imposing assumptions about the rupture process, a dataset known as SCARDEC. Meanwhile Stähler and Sigloch (2014) derived empirical orthogonal functions from the source time functions produced by Sigloch and Nolet (2006) and used these as the basis functions for source time function production. Ye et al. (2016b) and Hayes (2017) both produced finite fault models using teleseismic observations from which they extracted source time functions.

Meier et al. (2017) looked at the average source time function for different magnitude bins and observed that the median source time function, and the median normalised source time function both have a fairly regular (similar) behaviour in all magnitude groups. This suggests that $M \geq 7.25$ earthquakes all evolve similarly, and that there are no intrinsic differences between major and great earthquakes. However, Meier et al. (2017) note that the smallest magnitude bin ($M7$) shows a slower initial growth rate than the other bins ($M \geq 7.25$), which they suggest may result from lower certainty of source inversions in the first 10 s. Meier et al. (2020b) defined four conceptual models for rupture predictability: strong, moderate, weak and none. 'Strong' rupture predictability, allows final magnitude determination at times less than 10% of the earthquake duration (e.g. Nakatani et al., 2000; Ellsworth and Beroza, 1995; Beroza and Ellsworth, 1996; Olson and Allen, 2005)). 'Moderate' predictability allows for magnitude determination between one-tenth and one-third of the full rupture duration (Melgar and Hayes, 2017, 2019; Danré et al., 2019). In 'weak' rupture predictability, magnitude estimates are only possible between one-third and half of the full rupture duration (e.g. Meier et al., 2017; Trugman et al., 2019).

Melgar and Hayes (2019) looked for a 'growth phase' of the source time function. They calculated the average moment-acceleration over the first 1–20 s of source time functions, linearly interpolating between the moment-rate at zero-time and the moment-rate at the time in question. They found that at very short times (1–2 s) there is only a weak correlation between moment-acceleration and magnitude. However, once 5 s of rupture has passed, they find a clear log-linear relationship between moment-acceleration and magnitude. They propose that this is a representation of a two-stage rupture-initiation process in which the rupture initially nucleates and events of different final magnitudes are indistinguishable. At some later point, the rupture organises itself into a slip pulse, and this slip pulse has different properties depending on the final magnitude of the event. This suggests that earthquakes are moderately deterministic. Meier et al. (2020b) disagreed with the findings of Melgar and Hayes (2019), for two main reasons. The first is that the linear interpolation approach to understand the growth of the source time function is only valid where the time of interest is growth phase, which is unlikely for the longest time window (20 s) and the smallest magnitude events. The second is that the set of source time functions used by Melgar and Hayes (2019) is biased towards unusually long source time functions. Meier et al. (2020b) therefore repeat the analysis of Melgar and Hayes (2019) using only source time functions that have durations close to the expected duration and find that there is a qualitative scaling between the final magnitude and the interpolated moment-acceleration before the peak moment-rate is reached. They repeat this with synthetic source time functions based on a simple self-similar crack model (quadratic growth, maximum moment-rate at half-duration), and find that this scaling is no longer evident. They therefore propose that earthquakes are, at most, weakly deterministic but that source time functions may not be able to reliably resolve event onsets.

Danré et al. (2019) quantified the complexity of SCARDEC source time functions and found that larger earthquakes tend to have more complex source time functions. In general, they find that $M < 7$ earthquakes have fewer sub-events than major $M \geq 7$ events, suggesting that earthquakes are moderately deterministic. Larger

earthquakes have more and larger sub-events, opposing earthquake self-similarity. They also found that strike slip earthquakes have a stronger relationship between source time function complexity and magnitude, perhaps due to the constraint of the seismogenic width. Liu et al. (2023) would go on to find that strike-slip events tend to be more complex than dip-slip events.

1.2.8.4 Studies of lab seismicity

Throughout the 2010s, there were a number of advances in the generation and measurement of so-called labquakes, primarily through the analysis of acoustic emissions (AEs). Johnson et al. (2013) identified AEs in a sheared granular material, and found that AE precursors to larger stick-slip events occurred when the material was still dilating but the macroscopic frictional stress was no longer increasing. Meanwhile Goebel et al. (2013) simulated many seismic cycles on a lab fault and found that b-values varied in a way that correlated with stress on the fault. McLaskey and Kilgore (2013) found a localised aseismic slip zone in the centre of their lab fault, whilst the ends of the fault remained locked. High frequency foreshocks occurred in this slowly-slipping zone and the number of foreshocks depended on the shear-stressing rate, with more foreshocks occurring when loading was rapid. McLaskey and Lockner (2014) compared precursory AEs (foreshocks) to stick-slip events (mainshocks). They found that the first few microseconds look the same, but then stick-slip events rapidly grow until they rupture the whole fault.

A number of workers have attempted to use supervised and unsupervised machine learning to predict the timing and magnitude of labquakes. This builds on the observations of Latour et al. (2013) and Harbord et al. (2017) who found evidence for complex, two-stage nucleation in labquakes. Rouet-Leduc et al. (2017) predicted the time to failure based on a low amplitude signal that was previously thought to be noise but was in fact some kind of precursor. Bolton et al. (2019) and Bolton et al. (2020) were able to predict the timing and magnitude of lab quakes using the statistics of AEs. However, it is unclear to what degree these laboratory results translate to natural earthquakes.

1.2.9 2020s

Many of the early seismogram parameters were developed and tuned for earthquake early warning (e.g. Serdar Kuyuk et al., 2014; Kohler et al., 2018; Chung et al., 2019), but an important aspect of early warning is testing how well the warning zones correlate with regions of high intensity ground shaking. Minson et al. (2021) found that most ground motion results from earthquakes up to two magnitude units smaller than expected, even considering cases of very strong shaking. This suggests that $M < 7$ earthquakes are an important component of providing adequate warning of ground motion. Meanwhile, Meier et al. (2020a) conducted a mock-test of early warning in Japan and found that only 40–60% of areas that experienced strong to extreme ground motion would have received at least 5 s of early warning.

So far, every study looking at the relationship between early parameters of seismograms and magnitude has either found a positive relationship (higher values of parameters at larger magnitudes, e.g. Olson and Allen, 2005), a magnitude independent relationship (e.g. Rydelek and Horiuchi, 2006), or some combination (saturation at high magnitudes (e.g. Trugman et al., 2019) or bottoming out at lower magnitudes (e.g. Kuyuk and Allen, 2013)). However, Colombelli et al. (2020) found that larger events have lower values of P_d in the first second. Colombelli et al. (2020) attribute the difference between their results and Trugman et al. (2019) to several factors including variation in pick times and the spacing of P_d calculations, but do not provide explanations of discrepancies with other previous work (e.g. Colombelli et al., 2014).

Cochran et al. (2020) looked at clusters of seismicity induced by fluid injection. They found that the number of foreshocks was comparable to the number of aftershocks, but that the foreshock sequence was longer in duration than the aftershock sequence. Verdecchia et al. (2021) found that the change in pore-fluid pressure required to trigger earthquakes was less than 0.1 MPa. Slightly differently to Cochran et al. (2018), they found that a combination of pore-fluid and earthquake-earthquake triggering explained seismicity close to wells, whilst at far distances seismicity was explained by static Coulomb stress transfer. Cochran et al. (2023)

looked at a natural fluid-induced swarm and found that the largest earthquake occurred where there was significant fault complexity and high roughness. This is in line with the proposal of Wesnousky (2006) who suggested that the underlying fault structure is the predominant control on final earthquake magnitude.

Differentiating between the pre-slip and cascade models for foreshock occurrence remains a challenge (Ellsworth and Beroza, 1995; Gomberg, 2018). Cattania and Segall (2021) ran 2-D quasi-dynamic, elastic simulations of frictionally uniform faults following rate-state friction. They found that foreshocks and mainshocks both result from the failure of locked asperities, but mainshocks preferentially start on stronger asperities. In their simulations, foreshocks can be explained by a combination of static stress changes and pre-slip (creep), caused by episodic seismicity occurs on neighbouring asperities. This then causes aseismic slip to accelerate, loading other asperities and leading to further foreshocks. They did not incorporate fluids into their model.

Dresen et al. (2020) found that the rupture preparation phase of labquakes was detectable over long timescales and involved both AEs (foreshocks) and aseismic deformation (pre-slip). Bolton et al. (2021) found that the b-value of lab seismicity decreased as the fault approached failure, with this potentially resulting from fault slip acceleration. They also found that fault dilation leading to increased porosity resulted in larger foreshocks and a lower b-value. Song and McLaskey (2024) found that labquakes that were stopped or slowed by velocity-strengthening fault sections are closer to natural earthquakes than standard complete rupture events. This would suggest a low degree of determinism, since it would imply that most earthquakes are artificially stopped and thus cannot have intrinsically known their final magnitude early on.

1.3 Aims and Contribution

What is most apparent from reviewing so much of the earthquake-nucleation literature from the past half-century is the remaining level of disagreement on almost every topic. In particular, it remains unclear the extent to which earthquakes

are deterministic (that is, how different the earliest stages of big and small earthquakes are) and to what extent foreshock occurrences can be explained by earthquake-earthquake triggering. These discrepancies are difficult to disentangle from differences between studies. Many studies use small cherry picked datasets from a few regions of the world. Where studies have focussed on determinism, they have generally been focussed on the potential for application of the results for earthquake early warning, but have lacked curiosity about the physical processes that underpin these observations. Meanwhile, studies of foreshocks have often focussed on the relatively simple surrounds of southern California. Whilst this can be a good region to consider the potential impact of aseismic slip, it does not pose a natural testing ground for the role of fluids in earthquake nucleation.

I will now outline the structure of the rest of my thesis and highlight how I build on the findings and weaknesses of previous analyses to contribute to our understanding of earthquake nucleation and determinism. In each chapter I seek to address a few specific aims, which I outline before providing more detail on the novelty of my approach.

1.3.1 Chapter 3: Correlations between earthquake magnitude and early features of rupture in a global dataset

Aim 1 Collate a dataset of $M \geq 3$ earthquakes from around the world using an internally consistent approach. Calculate four parameters (predominant period, average period, peak ground displacement, integral of the velocity squared) on short windows of the seismograms.

Aim 2 Test whether these early features of seismograms are predictive of final earthquake magnitude, and if so, the relative duration of the window in which these features are required to be measured.

Aim 3 Pose physical explanations for our observations of early seismogram parameters which can be input into earthquake models.

The work in this chapter builds upon parameters which have previously been discussed in the context of determinism. By using a large, global dataset of earthquakes, I will be able to rigorously statistically evaluate the ability of these parameters to determine earthquake magnitude in windows shorter than the earthquake duration. Much of the previous work using these parameters has used windows of calculation which are long relative to the duration of moderate magnitude earthquakes. For work which seeks to establish differences between major and great earthquakes, this is not a problem. However, I seek to probe physical differences in the earliest stages of $M \geq 3$ earthquakes, and thus must use much shorter windows, and understand the relative durations of these windows. This will allow me to quantify the strength of any deterministic relationships, that is, the proportion of the earthquake that needs to have been measured before magnitude can be determined.

1.3.2 Chapter 4: Patterns in source time functions reflect variations in earthquake magnitude and inversion approach

Aim 1 Collate source time functions from five different databases of source time functions that are based on different inversion approaches.

Aim 2 Identify and quantify source time function features and test whether these scale with earthquake magnitude.

Aim 3 Examine the impact of dataset (and thus inversion method and assumptions) on source time function characteristics.

Compared to waveform-based determinism studies, there have been fewer studies of how source time function features scale with magnitude. However, this does not mean that these observations are without contention. Much of the criticism surrounds whether the source time functions used in the studies are a biased subset of the dataset, and indeed whether source time functions can reliably resolve the early stages of rupture. I therefore pose a broader question, whether any features of source time functions, not just early ones, are indicative of earthquake magnitude,

other than the obvious candidates of absolute moment release and total source duration. To provide an insight into the robustness of these features, I will use five datasets found via different inversion approaches. Features which appear in the results from multiple datasets are more likely to be a true representation of the earthquake source processes.

1.3.3 Chapter 5: Using foreshocks to constrain earthquake nucleation: comparing foreshock to aftershock ratios in the Hikurangi subduction zone to results from ETAS

Aim 1 Make new observations of foreshocks and aftershocks around $M \geq 3$ earthquakes on the Hikurangi subduction zone, New Zealand.

Aim 2 Analyse temporal patterns in the detected foreshocks and aftershocks in the hour around $M \geq 3$ mainshocks.

Aim 3 Compare the foreshock-aftershock ratio and absolute numbers of foreshocks and aftershocks between the observations and ETAS simulations.

Whether earthquake occurrence is best explained by simple earthquake-earthquake triggering driven nucleation or more complex extended external processes driven nucleation remains disputed. There are two main candidates for this potential physical process: aseismic slip or fluid flow. So far, most of our observations of earthquake nucleation have occurred in regions with infrequent-to-no fluid flow. The Hikurangi subduction zone is known to have both episodic slow slip events that release significant moment ($\approx M_W 7$), and spatially and temporally variable fluid flow. This makes it an excellent testing ground for more complex nucleation processes. A main limitation of our ability to study earthquake nucleation is having high-resolution detections of low magnitude earthquakes. I will use a relatively new method to detect small earthquakes around $M \geq 3$ mainshocks, and apply a novel statistical approach to correcting for background seismicity and false detections, to allow me to detect smaller magnitude events. I can then compare the observations to an ETAS simulation to understand whether more complex nucleation is at play.

2

A seismological primer

Contents

2.1 Wider definitions and context

A number of fundamental seismological theories and observations underpin much of the work in this thesis.

2.2 Seismological observations

For most of history, records of earthquakes rely on (1) the earthquake having been felt by people and (2) an enduring record being made of these observations. Chinese writings record every moderate and large earthquakes which occurred in China since 780 BCE, whilst Japanese records stretch back to 416 CE and are complete since 1600 CE. Records of earthquakes in the Mediterranean and Aegean exist

since about 300 CE, with some allusions in classical Greek, biblical and Arabic writings dating back to 1700 BCE (Bolt, 1999).

The first advance towards modern seismology came when, in 132 CE, Chinese scientist Zhang Heng developed a seismoscope which was able to record shaking when it was imperceptible to humans (Agnew, 2002; Needham and Wang, 1959). The seismoscope eventually developed into the seismometers we know today, which have been deployed in over 90% of countries.

The second advance, and arguably the inception of modern seismology occurred after the 1755 Lisbon earthquake, which provided amongst other things, evidence of ground motion at great distances (Agnew, 2002).

During the course of this thesis, we will make use of both of these advances: in trying to understand the physics of earthquakes, we will use recordings of earthquakes made at distance.

2.2.1 Elements of the seismic signal

When we record a signal at a seismometer, the signal is not a simple, direct representation of the earthquake. Instead, it has incorporated additional features in the journey it has taken to arrive at the seismometer, and in the process of being recorded. We call the first of these the 'path effect' and the second the 'instrument response'. In this section, we will discuss these aspects which interfere with our ability to directly image the source, and we will return to the the source itself in the following section.

We denote the earthquake source signal $x(t)$ which is then modified by path effects, $e(t)$ and $q(t)$, and the instrument response, $i(t)$. We can then write an equation for the displacement seismogram at time (t) as a convolution,

$$u(t) = x(t) * e(t) * q(t) * i(t). \tag{2.1}$$

It is often more helpful to work in the frequency domain, and so 2.1 can equivalently be written as the product of Fourier transforms of the four factors

$$U(\omega) = X(\omega)E(\omega)Q(\omega)I(\omega) \tag{2.2}$$

(Stein, 2005).

2.2.1.1 The Seismometer

Modern seismometers work through a process of electromagnetic induction: a magnet is attached to a mass which moves proportional to the ground motion. When the magnet moves through a coil of wire an electrical current is induced which is again proportional to the motion it experiences. Seismometers are better at recording signals at some frequencies than others, and this is encapsulated in the instrument response. Generally, instruments are designed to have a flat response over a certain frequency range, but by applying an instrument correction, it is possible to use data outside of this range.

To better understand the origins of the frequency-dependent instrument response, we begin by denoting the vertical Earth displacement as $u(t)$ and the displacement of the mass (m) as $z(t)$, both relative to their rest positions. We assume that the seismometer case does not move in relation to the Earth's surface and thus that the absolute displacement of the mass is given by the sum of $u(t)$ and $z(t)$. We henceforth drop the explicit mention of the dependence of u and z on t.

The motion of the mass is opposed by a spring, and by Hooke's law we can write that the force opposing the displacement of the mass (F_s) is given by

$$F_s = -kz \tag{2.3}$$

where k is the spring constant.

Most seismometers are damped using a dashpot to avoid excessive oscillations near the resonant frequency of the instrument. A dashpot is essentially a piston in a viscous fluid and the damping force F_d which results is given by

$$F_d = -D\dot{z}, \tag{2.4}$$

where the damping constant, D, depends on the specifics of the dashpot.

From Newton's second law and substitution,

$$F_s + F_d = ma \tag{2.5}$$

$$-kz + -D\dot{z} = m\frac{d^2}{dt^2}\left[u(t) + z(t)\right], \tag{2.6}$$

and rearranging we find

$$\ddot{z} + \frac{D}{m}\dot{z} + \frac{k}{m}z = -\ddot{u}. \tag{2.7}$$

If we define the resonant frequency of the un-damped ($D = 0$) system as w_0, we can write $w_0^2 = k/m$. We also define a damping parameter, ϵ, such that $2\epsilon = D/m$. We can now re-write equation 2.7 as

$$\ddot{z} + 2\epsilon\dot{z} + \omega_0^2 z = -\ddot{u}, \tag{2.8}$$

which may seem unnecessary at the moment, but it is helpful to build intuition for ϵ and ω_0 as we move into the frequency domain. This equation shows that the Earth's acceleration can be found from the mass displacement.

We now move into the frequency domain, since we would like to know how seismometers report different frequencies of ground motion. We can write $u(t)$ as a harmonic using complex numbers

$$u(t) = U(\omega)e^{-i\omega t}, \tag{2.9}$$

and we can do similarly for $z(t)$. We substitute these frequency-domain relations into equation 2.8 and divide by the common factor of $e^{-i\omega t}$, and find

$$-\omega^2 Z(\omega) + 2\epsilon i\omega Z(\omega) + \omega_0^2 Z(\omega) = \omega^2 U(\omega). \tag{2.10}$$

Finally, we can rearrange to find the factor $\mathcal{Z}(\omega)$ which relates $Z(\omega)$ to $U(\omega)$,

$$Z(\omega) = \frac{\omega^2}{\omega_0^2 - 2\epsilon i\omega - \omega^2}U(\omega) \tag{2.11}$$

$$= \mathcal{Z}(\omega)U(\omega), \tag{2.12}$$

where $\mathcal{Z}(\omega)$ is the frequency-response function of the sensor, and it is this that controls how different frequencies of signal are recorded by the seismometer. We can further subdivide $\mathcal{Z}(\omega)$ into amplitude and phase response components,

$$\mathcal{Z}(\omega) = A(\omega)e^{i\phi(\omega)}. \tag{2.13}$$

By controlling the relative strength of the damping compared to the stiffness of the spring (i.e. ϵ vs ω_0), we can design seismometers which are tuned for different frequencies.

One complication in understanding the frequency response arises from the fact that seismometers can measure displacement, velocity or acceleration of the sensor mass, and equally we can be interested in the displacement, velocity or acceleration of the Earth's surface. These need not directly map (i.e. an acceleration sensor can be used to look at displacement). Each time derivative serves to enrich the signal in high-frequency energy (since it introduces a factor of $-i\omega$). The frequency response is flat over a wide frequency band in two cases: (1) when the instrument and parameter of interest are the same (e.g. a velocity seismometer being used to measure velocity), and the frequency is high ($\omega \gg \omega_0$) or (2) when a displacement seismometer is used to look at ground acceleration and the frequency is low.

Where I directly use seismic waveforms, these are measured at short-period and broadband seismometers which are suited to recording high-frequency body wave arrivals. Broadband seismometers are designed to have a very flat frequency response over a broad range of frequencies, whilst short-period instruments are only suited for high-frequency measurements (above the microseism). Some of the source time functions that I use in Chapter 4 are calculated using long-period instruments which are good for low frequencies but are more expensive to produce.

In the course of processing seismic data, we must apply various filters to remove unwanted 'noise' signals. In particular, we often want to remove the noise resulting from water waves interacting, which we call the microseism. The microseism, and the similar phenomenon of the 'hum', are found between 0.003 and 1.0 Hz. For most of our analysis then, we high-pass filter above 1 Hz. In some analyses we also want

to remove the highest frequencies using a low-pass filter. The highest frequencies signals we can observe are given by the Nyquist frequency, which is half the sampling rate. Thus for 100 Hz data, the highest frequency signals that are recorded are 50 Hz.

2.2.1.2 The Path

The two components of the path effects $e(t)$ and $q(t)$ represent two different physical processes, elastic and anelastic respectively.

The $e(t)$ incorporates two main aspects: (1) geometric spreading and (2) the 'scattering' attenuation effect of reflections and conversions. Geometric spreading arises from the conservation of energy: consider a spherical (body-wave) wavefront moving away from a deep earthquake (to avoid the need to consider interactions with the free surface) in a homogeneous medium. Energy is conserved on the spherical wavefront and the area of a sphere is $A = 4\pi r^2$. Thus as the waves move further away from the source, and r increases, the energy per unit area decreases as $1/r^2$ and the amplitude decays as $1/r$. In reality the earth is not homogeneous, and the amplitude depends on focussing of the rays by the velocity structure. Surface waves are not particularly dealt with in this thesis, but the geometric spreading term follows $1/r$ for energy in this case.

Conversions and reflections also impact the amplitude of the wavefront. When a P wave is incident on a boundary, mode conversions occur which results in reflected and transmitted (refracted) P and S waves. The relative amplitudes of these four waves are given by the Zoeppritz equations:

$$\begin{bmatrix} R_{\mathrm{P}} \\ R_{\mathrm{S}} \\ T_{\mathrm{P}} \\ T_{\mathrm{S}} \end{bmatrix} = \begin{bmatrix} -\sin\theta_1 & -\cos\phi_1 & \sin\theta_2 & \cos\phi_2 \\ \cos\theta_1 & -\sin\phi_1 & \cos\theta_2 & -\sin\phi_2 \\ \sin 2\theta_1 & \frac{V_{\mathrm{P1}}}{V_{\mathrm{S1}}}\cos 2\phi_1 & \frac{\rho_2 V_{\mathrm{S2}}^2 V_{\mathrm{P1}}}{\rho_1 V_{\mathrm{S1}}^2 V_{\mathrm{P2}}}\sin 2\theta_2 & \frac{\rho_2 V_{\mathrm{S2}} V_{\mathrm{P1}}}{\rho_1 V_{\mathrm{S1}}^2}\cos 2\phi_2 \\ -\cos 2\phi_1 & \frac{V_{\mathrm{S1}}}{V_{\mathrm{P1}}}\sin 2\phi_1 & \frac{\rho_2 V_{P2}}{\rho_1 V_{P1}}\cos 2\phi_2 & -\frac{\rho_2 V_{S2}}{\rho_1 V_{P1}}\sin 2\phi_2 \end{bmatrix}^{-1} \begin{bmatrix} \sin\theta_1 \\ \cos\theta_1 \\ \sin 2\theta_1 \\ \cos 2\phi_1 \end{bmatrix}, \tag{2.14}$$

where R_P, R_S, T_P, and T_S, are the reflected and transmitted P and S wave amplitude coefficients respectively, and where θ_1 is the angle of incidence, θ_2 is the angle of the transmitted P wave, ϕ_1 is the angle of reflected S wave and ϕ_2 and angle of the transmitted S wave (Zoeppritz, 1919).

Meanwhile $q(t)$ parametrises anelastic attenuation, in which some of the mechanical energy of the wave is converted and lost. Primarily, this energy is converted to heat energy. This is what stops the Earth from oscillating constantly due to all past earthquakes. A range of mechanisms are required to explain anelastic attenuation at different length scales. At the crystal scale, these include point defects, line defects, dislocations and inclusions. In polycrystalline materials, grain boundaries become important, and at larger scales, cracks and fluids are important (Pujol, 2003). We quantify attenuation in a given location using the inverse of the quality factor, $Q = \frac{1}{\text{attenuation}}$. Q is (weakly) frequency dependent, particularly above 1 Hz (e.g. Anderson et al., 1977). Attenuation is spatially and temporally variable, making it difficult to correct for. Work which has mapped attenuation features is generally at a regional or smaller scale, and can sometimes, but not always, disentangle the effects of intrinsic (anelastic) attenuation from scattering (elastic) attenuation. Maps of Q^{-1} can be produced using perturbation theory, coda wave tomography or ambient noise tomography (Ranjan and Konstantinou, 2020; Cabrera-Pérez et al., 2024).

The path effect is unique for each station-receiver pair. However much we like to model the earth as a homogeneous half-space, it is in fact not homogeneous, and anisotropy has a significant impact on the seismic recordings we make. In this, wave speeds vary as a function of azimuth. At short time-scales, directivity is also important, as faults do not rupture instantaneously. This is comparable to the Doppler effect. It is therefore important to record earthquakes from a range of azimuths where possible to quantify and minimise directional effects.

2.2.2 Source Models

2.2.2.1 Moment-geometry scaling

Much of this thesis centres on whether there are differences between large and small earthquakes. For small ruptures, earthquakes can be modelled as growing circularly, and thus

$$M_0 \sim \mu \Delta\sigma r^3. \tag{2.15}$$

At some point the rupture grows large enough that it is constrained in one dimension and the rupture begins to grow roughly rectangularly. In this case, the slip can be determined by either L or W (e.g. Scholz, 1982). In the former case,

$$M_0 \sim \mu \Delta\sigma L W^2, \tag{2.16}$$

and in the latter,

$$M_0 \sim \mu \Delta\sigma W L^2. \tag{2.17}$$

2.2.2.2 Modelling the source

For small sources, far from the observer, we can use a simple point source model. In this case, the moment-tensor source time function hides the details of the rupture process. However, for larger earthquakes or at near source distances, it is important to account for the rupture propagation and fault geometry. Therefore we turn to finite fault models. These look to explain the timing, speed and amplitude of the slip on each part of the fault.

Dislocation models treat the earthquake source as a instantaneous slip on the fault. An example is the Haskell (1964) model, which defines the slip s at time t as

$$s(t) = \begin{cases} 0 & t < 0 \\ \frac{Dt}{T} & 0 < t < T \\ D & T > t, \end{cases} \tag{2.18}$$

where D is the final slip and T is the rise time. The physical realism of these models depends on the slip distribution specified. For instance, the Haskell model results in a spatial slip discontinuity at the edge of the fracture zone. At high-frequencies, the Haskell model is no longer physical (Madariaga, 1978).

A commonly used model is the Brune model (Brune, 1970). This is actually a point source model. It treats the rupture as a circular crack with instantaneous slip initiation and an exponential slip decay. This makes it suitable for fitting far-field displacement observations. It is also possible to model more complex earthquakes as the sum of many Brune displacements.

Crack-like models consider the rupture as an expanding crack. The rupture starts at a point and expands as rupture continues. Slip continues on all parts of the fault which have ruptured. The Madariaga (1976) model uses a non-infinite rupture velocity (and thus non-instantaneous slip) which means it is more physically realistic than the Brune model.

It is often informative to look at different models in the frequency domain (ω). The Brune model predicts a flat displacement spectrum (related to the moment) at low frequencies and a falloff as ω^{-2} at high frequencies. This ω^{-2} decay is the nucleation phase. The point at which the ω^{-2} decay begins is the corner frequency. ω^{-2} decay can also result from other models that have a finite slip duration, constant rupture velocity and smooth STF.

However, it does not make sense for the fault rupture to stop abruptly. The compact, far-field results of Sato and Hirasawa (1973) arise from considering a circularly growing fault which suddenly stops when the radius reaches the final radius. This violates causality because the centre of the fault cannot instantaneously know to stop.

Therefore other models introduce a stopping phase. For example, Molnar et al. (1973) model a fault where the rupture nucleates at the centre and grows rapidly in all directions at a constant velocity. Once the rupture reaches the maximum size, the rupture then contracts back to the centre at the same velocity. This model uses a ramp function to model slip at all locations, whilst Boatwright (1980) finds that the ramp model is only appropriate at the crack centre and the slip model should follow a square-root rise at other points. The stopping phase causes a sudden change in far-field displacement proportional to square-root time, and the spectral density for displacement has a high-frequency fall off asymptotic limit as $\omega^{-3/2}$ (and as ω^{-1} in the fault normal direction). This stopping signal dominates the fall off at high frequencies. However, the model of Molnar et al. (1973) decays as ω^{-2} to ω^{-3} depending on the direction. Theoretically, a stopping phase of this type should follow $\omega^{-5/2}$ as there is an additional linear dependence of the slip function on the distance from the crack tip. Similarly, Madariaga models produce a decay which is closer to ω^{-3} than ω^{-2}.

Some models even produce ω^{-3}. This steeper fall off may result from shorter slip durations (e.g. pulse like rupture) or the rupture stopping more abruptly and thus producing a stopping phase.

In a pulse-like rupture, as the rupture propagates, each part of the fault only slips for a short time. This can be considered as the fault healing behind the rupture front. This model matches many real earthquake observations better than crack models. It is often associated with rate-and-state friction or thermal pressurization effects that cause rapid weakening and healing.

2.2.3 Stress drop

Describing the earthquake size through the seismic moment, and equivalently magnitude (M_W), does not allow us to disentangle the amount of slip from the size of the fault area that slips. An earthquake in which a large amount of slip occurs on a small fault, looks the same in this framework as an earthquake which slips a large fault by a small amount. However, the former will cause much more stress to the surrounding fault. Thus we turn to quantifying the change in shear stress on a fault due to a given earthquake, which we term the stress drop. The stress drop, $\Delta\sigma$ is given by the surface integral of the difference in stress on the fault (S) between t_1 before the earthquake and t_2 after the earthquake, normalised by the area, A:

$$\Delta\sigma = \frac{1}{A}\int_S [\sigma(t_2) - \sigma(t_1)]dS. \tag{2.19}$$

In the ideal case of a circular fault with radius r in a homogenous half-space, Eshelby (1957) found an analytical solution such that

$$\Delta\sigma = \frac{7\pi\mu\bar{\mathbf{s}}}{16r} = \frac{7M_0}{16r^3}, \tag{2.20}$$

where $\bar{\mathbf{s}}$ is the average slip. For strike-slip motion on a shallow rectangular fault of length L and width w where $w \ll L$,

$$\Delta\sigma = \frac{2\mu\bar{\mathbf{s}}}{\pi w} = \frac{2M_0}{\pi w^2 L}, \tag{2.21}$$

(Knopoff, 1958). In a general case, we can write that

$$\Delta\sigma = c\frac{M_0}{L^3} = \frac{cM_0}{A^{3/2}}, \tag{2.22}$$

where c varies depending on fault shape and rupture direction (Brune, 1970). We choose to write the equation for stress drop in terms of the seismic moment, since seismological observations alone are most able to determine M_0. But we also need to know the fault dimensions in order to estimate the stress drop. This is particularly a problem for small earthquakes, since for large (and shallow) earthquakes we can use the surface rupture and aftershock studies.

How then to determine the fault geometry for smaller earthquakes or those with no surface rupture and few aftershocks? In this case, we generally estimate fault dimensions from far-field seismological observations, in particular the frequency spectra. This approach works because even very complex waveforms tend to have relatively simple spectra (Abercrombie, 2021). We previously discussed different models of fault rupture, and their resulting spectra in Section 2.2.2.

Brune (1970) proposed the first quantitative approach to estimating the stress drop using the ω^{-2} model. Applying this model, Brune showed that the corner frequency f_c is inversely proportional to the source radius. For a circular fault

$$r = \frac{k\beta}{f_c}, \tag{2.23}$$

where r is the source radius, β is the shear wave velocity and k is a factor which depends on the source model, fault geometry and rupture velocity used (e.g. Kaneko and Shearer, 2014, 2015). Considering the two earthquakes with the same moment that we previously discussed, the one which has a smaller area but higher slip will have a higher corner frequency and assuming the propagation factors are unaltered, higher ground velocities and ground accelerations will result (and thus the earthquake is likely to be more damaging; Abercrombie, 2021).

For a circular fault, we can then arrive at an equation for the stress-drop which does not require direct parametrisation of the fault geometry,

$$\Delta\sigma = \frac{7}{16}\left(\frac{f_c}{k\beta}\right)^3 M_0, \tag{2.24}$$

which comes from the substitution of equation 2.23 into equation 2.20. This relationship means that stress drop can be solely found from the earthquake spectrum: M_0 from the low-frequency portion and $\frac{f_c}{k\beta}$ from the high frequency portion. We call this approach a 'Brune-style' stress drop, regardless of any deviations from Brune's original approach. Whilst we no longer rely on our ability to estimate fault-geometry, our estimates of $\Delta\sigma$ are now very sensitive to changes in the assumed theoretical model (which affects k and the assumed rupture velocity), since $\Delta\sigma$ depends on the cube of the $\frac{f_c}{k\beta}$ term. For example, using the Brune (1970) model leads to stress drops five times larger than those found using a Madariaga (1976) model. A key difference between these two models is the rupture velocity, with Brune assuming $v_r = \infty$ whilst Madariaga assumed $v_r = 0.9\beta$.

2.2.4 Energy partitioning

The total energy released by an earthquake is not directly measurable. Some energy is radiated as seismic waves, which we measure as the radiated seismic energy, E_R. However, some of the energy is used to fracture the rock (the fracture energy, E_G) and yet more is used to overcome the friction on the fault (frictional energy, E_F). The radiated and fracture energy can be determined seismically, but the frictional energy depends on the absolute stress on the fault and thus is very difficult to measure.

The total energy released in an earthquake is given by

$$E = \frac{1}{2}\overline{(\sigma_1 + \sigma_2)\mathbf{s}}A, \tag{2.25}$$

where σ_1 is the initial stress and σ_2 is the final stress. As before, $\mathbf{s}$ is the slip on the fault and A is the fault area. This is usually approximated as

$$E = \bar{\sigma}\bar{\mathbf{s}}A = \frac{1}{2}\Delta\sigma\bar{\mathbf{s}}A + \sigma_2\bar{\mathbf{s}}A, \tag{2.26}$$

where the first term is equal to the sum of the radiated and fracture energies whilst the second term is equal to the frictional energy.

A simple model of an earthquake is given by the Orowan model, where the stress on the fault instantaneously drops from σ_1 to σ_2. In this case, $E_G = 0$,

and the stress drop is given by,

$$\Delta\sigma_{\text{Orowan}} = \frac{2E_R}{DA} = \frac{2\mu E_R}{M_0} = 2\mu\tilde{e}, \tag{2.27}$$

where $\tilde{e} = E_R/M_0$ is the scaled energy, which relates the radiated energy to the moment. The Orowan stress drop is a lower bound on the true stress drop.

In reality, we do not expect the stress on the fault to drop instantaneously. One model of the evolution of stress on the fault is the slip-weakening model, where the stress drops from σ_1 to σ_2 over a critical slip distance D_C. In this case,

$$\frac{1}{2}\Delta\sigma \mathbf{s} A = E_G + E_R = \frac{\Delta\sigma}{2\mu}. \tag{2.28}$$

The ratio of E_R to E_G is quantified by the radiation efficiency, η_R

$$\eta_R = \frac{E_R}{E_R + E_G}. \tag{2.29}$$

The stress on the fault may follow a complex trajectory, including rising above σ_1 early on in rupture (the yield stress), or for the stress state of the fault ending at a value above or below σ_2 (undershoot and overshoot). In crack-like ruptures, overshoot can result from a high-rupture velocity rupture arresting (no longer propagating) whilst other parts of the fault continue to slip (e.g. Madariaga, 1976). Higher rupture velocities lead to lower values of η_R, following

$$\eta_R = 1 - g(v_r), \tag{2.30}$$

where $g(v_r)$ is a model specific function that depends on the specific crack model used and the ratio of v_r to the Rayleigh or shear-wave velocity. The radiation efficiency is distinct from the seismic efficiency (η) which describes the fraction of total energy which is radiated into seismic waves.

2.2.5 Self-similarity

Earthquake self-similarity refers to the idea that earthquakes of different magnitudes are scaled versions of the same process. In this case, the physical processes that govern them would be scale-invariant. Self-similarity may be evident in a range of

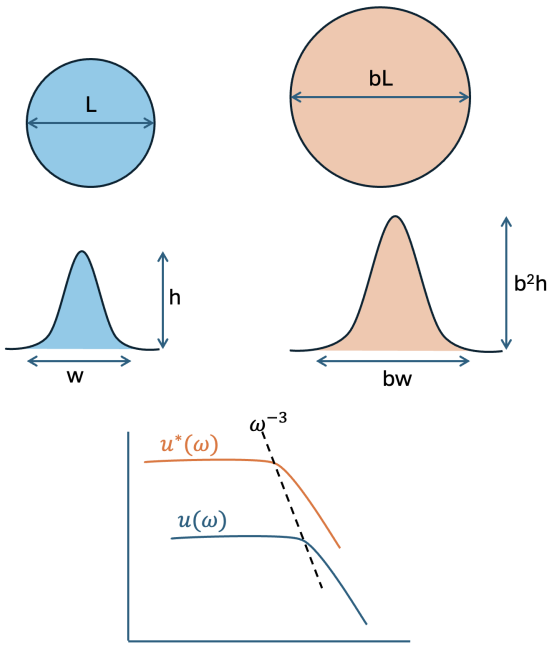


Figure 2.1: Earthquake self-similarity can be apparent in a number of ways. The top panels show two earthquakes where the larger, orange rupture has is scaled by b compared to the smaller blue rupture. This results in a scaling of the far-field displacement pulse (middle panels) and the spectrum (bottom panel). See text for more details. After Prieto et al. (2004). Not to scale.

observations. In Figure 2.1, we consider a rupture with length-scale L and area A. We also consider a larger rupture, which is scaled by b, such that $L^{\text{large}} = bL^{\text{small}}$. The area of this larger rupture is then $A^{\text{large}} = b^2 A^{\text{small}}$. The moment of the larger rupture is $M_0^{\text{large}} = b^3 M_0^{\text{small}}$. In the far field, the displacement pulse which results from the larger rupture will be b-times wider and the amplitude of the pulse will be increased by b^2. The recorded displacement pulse of the larger earthquake, u^*, is given by

$$u^*(t) = b^2 u(t/b), \tag{2.31}$$

where $u(t)$ is the displacement pulse of the smaller earthquake.

In the frequency domain, self-similarity means that the earthquake spectra are scaled relative to each other. The corner-frequency is magnitude dependent but the spectral fall-off follows the same scaling at all magnitudes (e.g. ω^{-2}). By the similarity theorem for Fourier transforms, the spectrum of the larger earthquake is identical to that of the smaller earthquake, but shifted along a line of ω^{-3}. The

spectrum of the larger earthquake is given by

$$u^* = b^3 u(b\omega), \tag{2.32}$$

where $u(\omega)$ is the spectrum of the first earthquake (Prieto et al., 2004).

If earthquakes are self similar, the scaled energy $\tilde{e} = E_R/M_0$ should be constant as a function of moment. This has been observed by some, but not all studies. Studies using only local records find a $M_0^{1/4}$ trend, whilst those using teleseismic records are more likely to find a moment-invariant relationship. Similarly, the stress drop should be constant with moment if earthquakes are self similar. Whether observations of stress drop support self-similarity is difficult to disentangle from variation in stress drop found by different methods. Discrepancies between whether stress-drop and scaled energy agree on self-similarity may also arise from errors in the energy calculation.

If earthquakes are self-similar, we would expect sources to be equally complex at all magnitudes. One way to model this may be through fractal faults, where the fault is equally complex at every scale.

Note that earthquake source models also consider a fracture mechanics definition of self-similarity. In this, the crack grows symmetrically at constant velocity from a zero initial length. This results in mechanical fields that are invariant with respect to an observer moving steadily away from the nucleation point. This is in contrast with steady state crack growth (Xia et al., 2006).

3

Correlations between earthquake magnitude and early features of rupture in a global dataset

Contents

Key Points

- We examine the beginnings of seismograms from 4000+ earthquakes from around the world.
- We identify statistically significant relationships between four features of the seismograms and the final earthquake magnitudes.
- These relationships provide insight into the physical processes underpinning rupture initiation.

Abstract

It remains uncertain whether an earthquake 'knows' its final magnitude part-way through its rupture. Several studies have found that early features of rupture can constrain final earthquake magnitude, but many of these have considered datasets with fewer than 100 earthquakes in order to reduce extraneous sources of variability. We consider a larger, but less carefully curated dataset. We calculate four parameters of the early portions of seismograms using >4000 earthquakes from around the world; we simply accept the scatter inherent in such a range of events. Despite this scatter, we find that on average larger earthquakes have distinct features in their early stages: they have longer predominant and average periods, release more energy, and reach higher maximum moment rates. The differences between small and large earthquakes persist even when we consider short measurement windows less than 1/3 of the earthquake duration. These results further evidence that there are physical differences between the early stages of large and small earthquakes. These, and further observations, could help us better understand how local stresses and frictional weakening drive earthquake rupture.

Plain Language Summary

Once an earthquake has finished, seismologists can easily calculate its magnitude. We are interested in whether the final magnitude is decided before the earthquake is over: do big earthquakes start differently than small earthquakes, or do they just happen to be able to grow for a longer time? This distinction is important as we develop earthquake early warning systems and as we improve our understanding of earthquake physics.

Several parameters have been used to quantify the early parts of earthquakes. We calculate 4 of these parameters on very short segments of the recordings of over 4000 earthquakes from around the world. We find that these parameters largely scale with the final earthquake magnitude. For instance, we find that large earthquakes tend to release more energy than small earthquakes even when just looking at the first 10% of the event. This gives us insights into the physics that underpins the early stages of earthquakes.

3.1 Introduction

After an earthquake ends, we can observe its features and calculate its magnitude. However, it is unclear if we can calculate an earthquake's final magnitude before rupture ends. If characteristics of nucleation and early rupture allow us to determine earthquake magnitude, we say earthquakes are deterministic. The deterministic nature of earthquake rupture, or lack thereof, is of interest for fundamental science—for understanding how earthquakes grow—and for society; if initial features of an earthquake determine the ultimate magnitude, then this information may be able to be incorporated into rapid earthquake early warning systems.

In the past, many different approaches have been used to understand the extent to which earthquakes are deterministic. One approach is to probe particular features of observed seismograms. Some studies have identified seismic nucleation phases that scale with the eventual size of the earthquake (Ellsworth and Beroza, 1995), whilst other studies have identified scalings between magnitude and various attributes, such as peak P wave displacement (Colombelli et al., 2013) or predominant period (Olson and Allen, 2005). For example, Lancieri et al. (2011), investigated four parameters of seismograms in short windows after the M_W7.8 Tocopilla earthquake and its aftershocks. They found that whilst some parameters scale with magnitude, other parameters scale with magnitude only for small earthquakes and that values saturate for larger earthquakes. Indeed, other studies have found no useful relationships between the early earthquake properties and final magnitude (Rydelek and Horiuchi, 2006; Meier et al., 2016) over a much wider magnitude range. Trugman et al. (2019) find that peak ground displacement saturates exactly when expected from modelling rupture with a simple, non-deterministic model. These observations of non-deterministic rupture would make sense if the underlying fault structure is the predominant control on final earthquake magnitude (Wesnousky, 2006). Some of the differences between observational studies come from the physical processes of the earthquakes but others may result from subtleties of the processing approaches.

Source time functions have also been used to probe the deterministic nature of large earthquakes. Some of these studies have found that early rupture processes

are universal: that ruptures are not deterministic (Meier et al., 2017, 2020b), whilst others have concluded that there are measurable differences between the onsets of large and small earthquakes (Melgar and Hayes, 2019).

A major uncertainty in determinism results centres around the potential biases in the data selection, processing, and analysis (as discussed by Meier et al., 2016). So, in this work, we aim to add a relatively large, global dataset to the available observations. We examine the early seismograms of several thousand earthquakes and calculate four parameters: predominant period, average period, the integral of the velocity squared, and peak ground displacement. We outline some of the theory and past work behind these four parameters in section 3.2. In section 3.3, we discuss how the length of our windows of calculation compare to earthquake durations, and in section 3.4 we outline our dataset and workflow. In section 3.5, we present our results including relationships between each of our four parameters and magnitude. Work on determinism often stops here, with a relationship between a parameter and magnitude, particularly where a deterministic relationship useful for early warning has been found. But there are many more questions we can ask about what these parameters actually represent, and what they tell us about earthquake source processes. We therefore finish section 3.6 by discussing the potential physical meanings of our parameters.

3.2 Methodological background

We calculate four parameters from the beginning of earthquake seismograms: predominant period, average period, the integral of the velocity squared, and the peak ground displacement. Here we outline some of the theory and equations behind these parameters and discuss previous work that made use of them.

3.2.1 Predominant Period

As the name suggests, the predominant period extracts the dominant period (or equivalently, the dominant frequency) from a signal of interest. We calculate the

predominant period, τ_p, from the ratio between observed velocities and accelerations, following Allen and Kanamori (2003).

We first compute the bandpass filtered ground velocity v_i with a 0.1–19 Hz, 4-corner Butterworth filter. We then use a recursive algorithm to compute the smoothed ground velocity squared:

$$V_i = \alpha V_{i-1} + v_i^2. \tag{3.1}$$

Here α is a smoothing constant, the value of which depends on the sampling frequency of the data. For our 100 Hz data, $\alpha = 0.99$. We also compute the smoothed acceleration squared:

$$A_i = \alpha A_{i-1} + \left(\frac{dv}{dt}\right)_i^2. \tag{3.2}$$

The predominant period, τ_p, is then

$$\tau_p = 2\pi\sqrt{\frac{V_i}{A_i}}. \tag{3.3}$$

Finally, we are interested in the maximum predominant period in our time window of interest:

$$\tau_p^{\max,\Delta t_i} = \max_{t\in[0,\Delta t_i]} \tau_p(t). \tag{3.4}$$

3.2.1.1 Previous results

A number of researchers have investigated the $\tau_p^{\max}$ of earthquakes (Nakamura, 2004; Olson and Allen, 2005; Lockman and Allen, 2007; Tsang et al., 2007; Wurman et al., 2007; Brown et al., 2009). Olson and Allen (2005) calculated the relationship between $\tau_p^{\max}$ and magnitude for 71 earthquakes in the United States of America, Taiwan and Japan. They found a roughly log-linear relationship ($M = 7.14 \log \tau_p^{\max} + 5.93$) with a linear correlation coefficient of 0.9.

$\tau_p^{\max}$ has been incorporated into the EPIC and ElarmS earthquake early warning algorithms for ShakeAlert, used on the US West Coast (Chung et al., 2019; Kohler

et al., 2020). Indeed, at one point $\tau_p^{\max}$ was used for magnitude estimates (Brown et al., 2011), but now it is used only to confirm that detections are truly earthquakes (Angela Lux, Personal Communication).

However, there are some concerns about the stability of the $\tau_p^{\max}$ parameter (Wolfe, 2006; Shieh et al., 2008). For example, Wolfe (2006) found that the parameter gives increased weighting to higher amplitudes and frequencies, and so averaging τ_p over stations does not give a true average of the frequency content of an earthquake.

The recursive nature of the predominant period can lead to inaccurate measurements in the earliest portion of a measurement window (e.g., Wolfe, 2006). For this reason, Olson and Allen (2005) began analysing the computed $\tau_p^{\max}$ values 0.05 s after the P arrival rather than upon the P arrival. We will examine results with several different 'blanking window' durations: 0, 0.05, 0.1, 0.25, and 0.5 s but find that the value minimally affects our results.

3.2.2 Average Period

The average period, τ_c, also measures a characteristic period of the ground motion within a given time window. However, we calculate τ_c using integrals of squared velocity and displacement rather than using recursive calculations. τ_c is given by,

$$\tau_c^{\Delta t_i} = \frac{2\pi}{\sqrt{\frac{\int_{t_i}^{t_i+\Delta t_i} \dot{u}^2(t)dt}{\int_{t_i}^{t_i+\Delta t_i} u^2(t)dt}}}, \tag{3.5}$$

where u is the observed displacement, Δt_i is our time window of interest, and $\dot{u}$ is the velocity. Both the velocity and displacement are highpass filtered from 0.078 Hz, using a 3-corner Butterworth filter.

3.2.2.1 Historical Background

The average period was derived by Kanamori (2005). To understand why the value in equation 3.5 is a sensible estimate of average period, note that we can rewrite the quantities inside the square root as integrals over frequency using Parseval's

theorem. The value inside the square root then becomes

$$\frac{4\pi^2 \int_0^\infty f^2 \left|\hat{u}(f)\right|^2 df}{\int_0^\infty \left|\hat{u}(f)\right|^2 df} = 4\pi^2 \left\langle f^2 \right\rangle, \tag{3.6}$$

where f is frequency, and $\hat{u}(f)$ is the displacement spectrum. The obtained value is thus proportional to $\langle f^2 \rangle$: the average of f^2, as weighted by $\left|\hat{u}(f)\right|^2$ (Kanamori, 2005). The average period τ_c is the inverse square root of this average-frequency-squared.

The parameter τ_c represents the effective period for multichromatic waves (and the true period for monochromatic, single-frequency waves). Since $f = 1/\tau$, we find

$$\tau_c^{\Delta t_i} = \frac{2\pi}{\sqrt{\frac{\int_{t_i}^{t_i+\Delta t_i} \dot{u}^2(t)dt}{\int_{t_i}^{t_i+\Delta t_i} u^2(t)dt}}} = \frac{2\pi}{\sqrt{4\pi^2 \left\langle f^2 \right\rangle}} = \frac{1}{\sqrt{\left\langle f^2 \right\rangle}}, \tag{3.7}$$

which is the same as equation 3.5.

3.2.2.2 Previous Results

Wu et al. (2006) calculated the relationship between τ_c and magnitude for 46 earthquakes in Taiwan. They found a roughly log-linear relationship ($M = 3.088 \log(\tau_c) + 5.300$) with a correlation coefficient of 0.77 and a standard deviation of 0.57.

Similar exercises have also been conducted in Japan, Taiwan and Southern California (Kanamori, 2004; Wu and Kanamori, 2005a,b; Wu and Zhao, 2006; Wu et al., 2007; Wu and Kanamori, 2008b). Wu and Kanamori (2008a) collate 54 events from these analyses which have data from at least four stations. They find that $M = 3.373 \log(\tau_c) + 5.787 \pm 0.412$ with a correlation coefficient of 0.933.

τ_c has been used in real-time EEW tests in California (Bose et al., 2009) and Taiwan (Hsiao et al., 2009). τ_c is more robust to filtering choices than $\tau_p^{\max}$, so combining τ_c and $\tau_p^{\max}$ can lead to a lower uncertainty on estimated magnitude than either value can provide alone (Shieh et al., 2008).

3.2.3 IV2

The integral of velocity squared (IV2) parameter (Festa et al., 2008) includes information about the energy radiated by the advancing rupture (e.g. Lancieri et al., 2011).

IV2 is given by

$$IV2^{\Delta t_i} = \int_{t}^{t+\Delta t_i} v^2(t)dt, \tag{3.8}$$

where t is the time of the first arrival, and v is the particle velocity measured on the seismograms. Δt_i is the time window considered in the analysis. We normalise each IV2 measurement to a reference epicentral distance of 1 km by multiplying by the geometric spreading term ($(R^2/1\ \mathrm{km}^2)$) as further discussed in Section 3.4.3.

Kanamori et al. (1993) showed that IV2 should be proportional to the energy radiated by an earthquake:

$$E_\phi = kR^2 c_\phi IV2_\phi, \tag{3.9}$$

where c_ϕ is the wave speed of phase ϕ, k incorporates several constants, and R is the hypocentral distance.

3.2.3.1 Previous Results

Festa et al. (2008), Lancieri et al. (2011), and Peng et al. (2013) found linear scalings between IV2 and moment in Japan, Chile, and China respectively. However, Festa et al. (2008) found that IV2 saturated at around $M5.8$ using a 4 s window of the P wave or a 2 s window of the S wave.

3.2.4 Peak Ground Displacement

Peak ground displacement, P_d, measures the maximum ground displacement in the time window of interest, Δt_i. To calculate P_d, we first highpass filter our velocity seismograms at 0.075 Hz to remove long-period noise. Then we integrate to displacement u. Finally, we take the maximum value of the displacement record in our window of interest and normalise to a reference distance (R_{ref}; see section 3.4.3), so that

$$P_d^{\Delta t_i} = \max_{t\in[0,\Delta t_i]} u(t) * \left(\frac{R}{R_{\mathrm{ref}}}\right)^2. \tag{3.10}$$

3.2.4.1 Previous Results

A number of studies have found relationships between P_d and magnitude without any obvious saturation. P_d has been used in a number of real-time EEW tests, often in combination with another parameter like τ_c, in California (Allen et al., 2009; Bose et al., 2009), Taiwan (Hsiao et al., 2009), Turkey (Fleming et al., 2009), Italy (Lancieri and Zollo, 2008) and Japan (Colombelli et al., 2020). It formed part of the VirtualSeismologist algorithm considered for use in earthquake early warning in California and Japan (Cua et al., 2009; Kohler et al., 2009). All of these studies find a positive relationship between P_d and magnitude except for Colombelli et al. (2020) who find that larger events have lower values of P_d in the first second.

P_d has been studied in a post-hoc sense in combination with IV2 in Chile (Lancieri et al., 2011) and China (Peng et al., 2013). Festa et al. (2008) proposed the use of $P_d^2/IV2$ in probabilistic EEW applications in Japan.

Trugman et al. (2019) predict and observe a saturation in P_d values at $\sim M5.8$ when calculated in a 1 s window, and similar behaviour at higher magnitudes in longer windows. This magnitude marks the midpoint of the transition from linear to magnitude-independent scaling, and thus there is still a weak relationship between P_d and magnitude up to $\sim M6$. This corresponds to an average source duration of $\sim$ 3 s. Where more than one-third of the earthquake duration has passed, they observe a linear relationship between P_d and magnitude in a study of 2409 earthquakes in Japan. Lancieri et al. (2011) also observe a saturation in P_d at $\sim M6.5$ for calculations in a 2 s window, but do not observe this in a 4 s window.

3.3 Earthquake Durations

We wish to know whether the beginning of an earthquake can tell us about the final magnitude. To do this, we must focus on windows of calculation (Δt_i) which are shorter than the earthquake duration.

We follow Trugman et al. (2019) in using a simple model of earthquake rupture to calculate average source durations. In a simple case, earthquakes begin by

growing roughly circularly, spreading out on the 2-dimensional fault. In this regime, the released moment $M_0 \sim \Delta\sigma L^3$, where L is the along-strike length and $\Delta\sigma$ is the stress drop. If the rupture velocity (V) is constant through time, the rupture length will grow linearly with time. If the rupture stops after a duration T, $L = VT$, and moment $M_0 \sim T^3$.

At some point in the growth of large earthquakes, the width (W) exceeds the width of the seismogenic zone ($W = W_{\max}$), and the rupture begins to grow in 1-dimension. Once this happens, $M_0 \sim \Delta\sigma W_{\max}^2 L$ and moment starts to increase linearly with time. Ruptures that stop well into this regime have moments that scale linearly with duration T.

Given this assumed growth, it is possible to calculate earthquake duration for a range of magnitudes (following e.g, Gomberg et al., 2016). We assume relatively common estimates of stress drop $\Delta\sigma = 5$ MPa, maximum width $W_{\max} = 60$ km, and rupture velocity $V = 2500$ ms^{-1}. Using these values, we calculate earthquake durations and compare them to our analysis window lengths (0.3, 0.5, 1, and 4 s). For instance, $M3$ earthquakes are on average 0.1 s long, shorter than our shortest window, whilst $M4$ earthquakes are on average 0.3 s long, comparable to our shortest window. $M5$ and $M6$ earthquakes are on average 1 and 4 s long, 3 and 12 times longer than the 0.3 s window, respectively. Changing the stress drop by a factor of 3 changes the duration by a factor of $\sqrt[3]{3}$ or about 1.44 (Figure A.1). The maximum width falls in the lower-middle of the range expected by Gomberg et al. (2016) (10–200 km) but reducing this value increases the duration of $M >\sim 6$ events as they become bounded earlier in their rupture and increasing the value of $W_{\max}$ has no impact over our magnitude range of interest. We thus proceed with $W_{\max} = 60$ as it is preferable to use a lower bound for earthquake duration (i.e. it is better for us to overestimate the proportion of rupture in a given time window than it is to underestimate it).

In what follows, we will calculate results for the full set of windows but will focus our interpretation on those earthquakes that are significantly longer than

the measurement window (the earthquake duration is two, three or ten times the measurement window, see Table A.2).

Our methods calculate a parameter in some time window of observation, Δt_i, but at near source distances, this window does not correspond to an equivalent time of rupture. A Δt_i s window of P waves images a smaller area of the fault than an equivalent Δt_i s window of S waves, due to their different propagation velocities, and both image an area smaller than the area ruptured Δt_i s after the rupture has begun.

3.4 Implementation

We calculate our four parameters of interest ($\tau_p^{\max}$, τ_c, P_d and IV2) from observations of the beginnings of thousands of earthquakes.

3.4.1 Data

We begin by selecting all $M \geq 3$ events in 2019 that have data hosted on the IRIS Data Center. To this dataset, we add $M \geq 5$ events between 2005 and 2021. These events are shown in Figure 3.1. In our analysis, we use data from all stations with HH* or BH* channels within 1° of the epicentre, regardless of network. We download the data from the IRIS Data Center using the ObsPy mass-downloader utility.

We require the signal-to-noise amplitude ratio of the unfiltered data to be at least 20, which leaves us with 4143 earthquakes, including 2245 $M \geq 5$ events. These data come from 173 networks: NZ, ZW, CI, AZ, Z3, 1B, GE, 8F, CX, ZA, TJ, MU, YM, NU, 6E, ZB, Y9, HP, ZY, NM, ZO, 3A, HA, HT, 2B, 7B, ZD, Y2, XY, IV, MN, RO, YC, XJ, 3H, CB, IQ, 4B, NO, RM, YL, 6D, WI, HL, CL, 6C, XS, C, BC, CN, Z8, XI, CK, KR, 4A, ZE, PL, GO, XD, ES, KO, XP, XM, OK, GU, XF, G, BK, TA, 6A, VU, ND, KZ, 5C, BA, Z7, X9, KC, 8E, NC, NN, Z6, IN, PA, TU, YZ, HV, PT, SV, EC, IU, Z5, WA, C1, NA, MT, FR, AC, TM, HC, GS, NX, 5L, OV, MQ, XV, AK, 7D, KN, N4, YQ, GL, IO, XL, TC, 6K, 3J, 9H, C8, 5K, CQ, 5F, XQ, 8G, 8H, MM, YR, Z1, YU, ZR, WC, AF, 7C, NS, CZ, SN, TX, HU, PM, VI, BR, XO, O2, GM, DE, PY, YB, UM, C0, YX, TD, RV, AE, KY, IE,

XA, CA, NL, UW, UO, ZC, YA, PE, LD, OH, 3L, CW, PQ, AG, GR, OE, PR, and JM. Further information about the networks is listed in the Open Research Section.

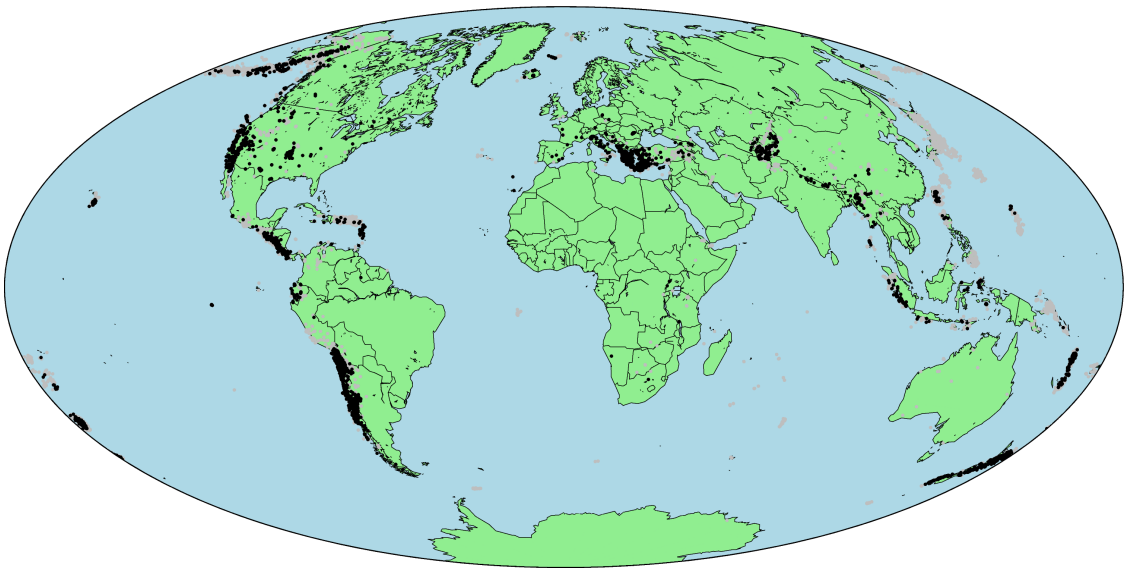


Figure 3.1: Map of the earthquakes used in this study. Grey dots show the complete IRIS catalogue: all $M \geq 3$ earthquakes in 2019, and all $M \geq 5$ earthquakes between 2005 and 2021. Black dots show the earthquakes that have appropriate data for analysis: those with a station within 100 km and a signal-to-noise ratio greater than 20.

3.4.2 Phase Picking

In order to analyse the beginning of earthquake ruptures, it is important to have accurate picks of when the seismic waves arrive at each station. We use a machine learning approach to identify the P wave arrival of each earthquake, specifically the EQTransformer model (Mousavi et al., 2020) implemented in SeisBench (Woollam et al., 2022). EQTransformer is an ideal model to use: it was trained on a global training dataset and it has high precision and recall, particularly in the epicentral distances we are interested in (Mousavi et al., 2020; Münchmeyer et al., 2022).

3.4.3 Distance corrections for amplitude

Different features of waveforms are affected by changing source-receiver distance in different ways. There are two main physical mechanisms which affect seismic waves as they travel through the earth: geometric spreading and attenuation.

3.4.3.1 Geometric spreading

Geometric spreading is relatively straightforward to correct. To a first order approximation, P waves radiate spherically from earthquake sources. Energy density thus decays following an inverse-square law ($1/R^2$), where R is the distance the wave has propagated from the source. Wave amplitude decays as $1/R$. Geometric spreading is important to consider for parameters based on amplitude, such as IV2 and P_d, but has no effect on the frequency content of the waves. For IV2, we analytically remove the geometrical spreading term by normalising measurements to a reference distance (1 km) by multiplying by $(R/1\text{ km})^2$.

To correct P_d, we use a more empirical approach. Kuyuk and Allen (2013) compared the magnitude predicted from P_d to the true magnitudes for 2166 globally distributed earthquakes. They found that individual P_d estimates could best retrieve magnitude if they were normalised by distance$^{1.38}$. We follow this approach in our work and so we multiply each P_d value by $(R/1\text{ km})^{1.38}$.

3.4.3.2 Attenuation

It is more difficult to correct for distance-dependent attenuation. Intrinsic and scattering attenuation are larger at higher frequencies (Anderson et al., 1977), so we expect a bias toward lower frequencies in more distant observations. However, given the complexity of frequency-dependent attenuation, we choose not to correct for attenuation in our global dataset. Instead, we will simply return to assess the potential influences of attenuation on our results in section 3.5. In that section, we reanalyse the results using only observations in particular distance ranges.

It is also important to note the magnitude dependence of the attenuation effect. Large earthquakes are less affected by attenuation, as they are less enriched in higher frequency energy.

3.4.4 Azimuthal coverage and number of stations

Looking at the azimuthal distribution of stations (Figure 3.2), several features stand out. $M3-4$ earthquakes have a lack of recordings in the South-East. This may

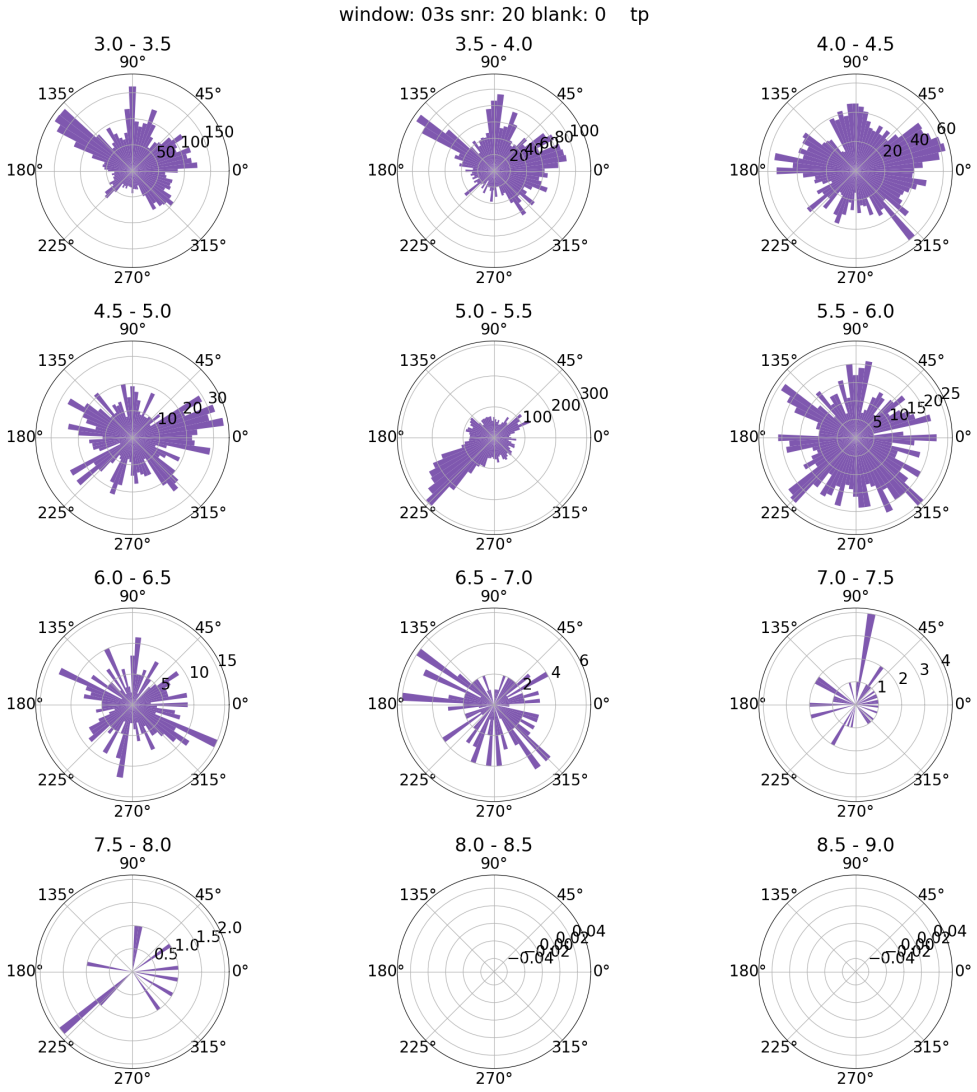


Figure 3.2: Rose diagram of azimuthal coverage by magnitude for $\tau_p^{\max}$.

result from a large portion of the seismicity forming part of the Ridgecrest sequence.

For the $M5-5.5$ group there is a significant bias towards events recorded at azimuths of around 45°. These result from the Bárðarbunga eruption which began in August 2014. In response to increasing seismic activity, a seismic array was deployed, primarily to the north-east of the seismicity. Thus the azimuthal bias results. We see a similar azimuthal distribution for the other parameters. (Figures A.2–A.4)

We also calculate the maximum azimuthal gap between stations recording any given earthquake (Figure 3.3). Events that are only recorded at one station have an azimuthal gap of 360°. There is a broad distribution of azimuthal gaps, with the

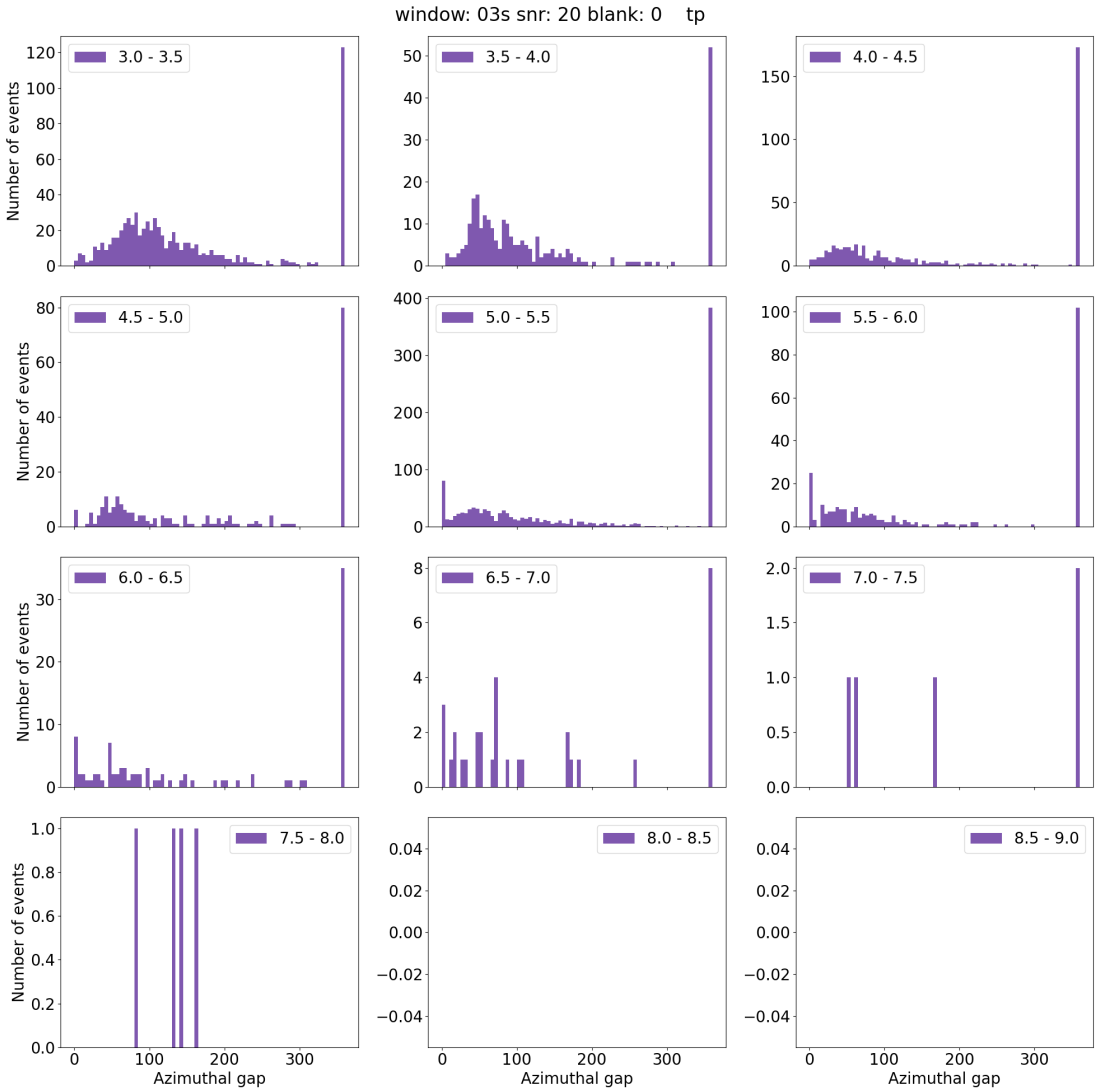


Figure 3.3: Histograms of the maximum azimuthal gap between stations used in calculation of $\tau_p^{\max}$ for given earthquakes. Events that are only recorded at one station have an azimuthal gap of 360°.

distribution peaking between 45 and 90°. Smaller events tend to have a slightly larger maximum azimuthal gap, but all events show a significant range. The maximum azimuthal gap shows similar distributions for the other parameters (Figures A.5–A.7).

Larger events tend to be recorded at fewer stations (Figures 3.4 and A.8–A.22). This makes sense: the catalogue is only complete to $M3$ in the US, where stations are closely spaced, and there may be many within 1° of the epicentre. Larger events are globally distributed where station spacing may be over 1° and so they

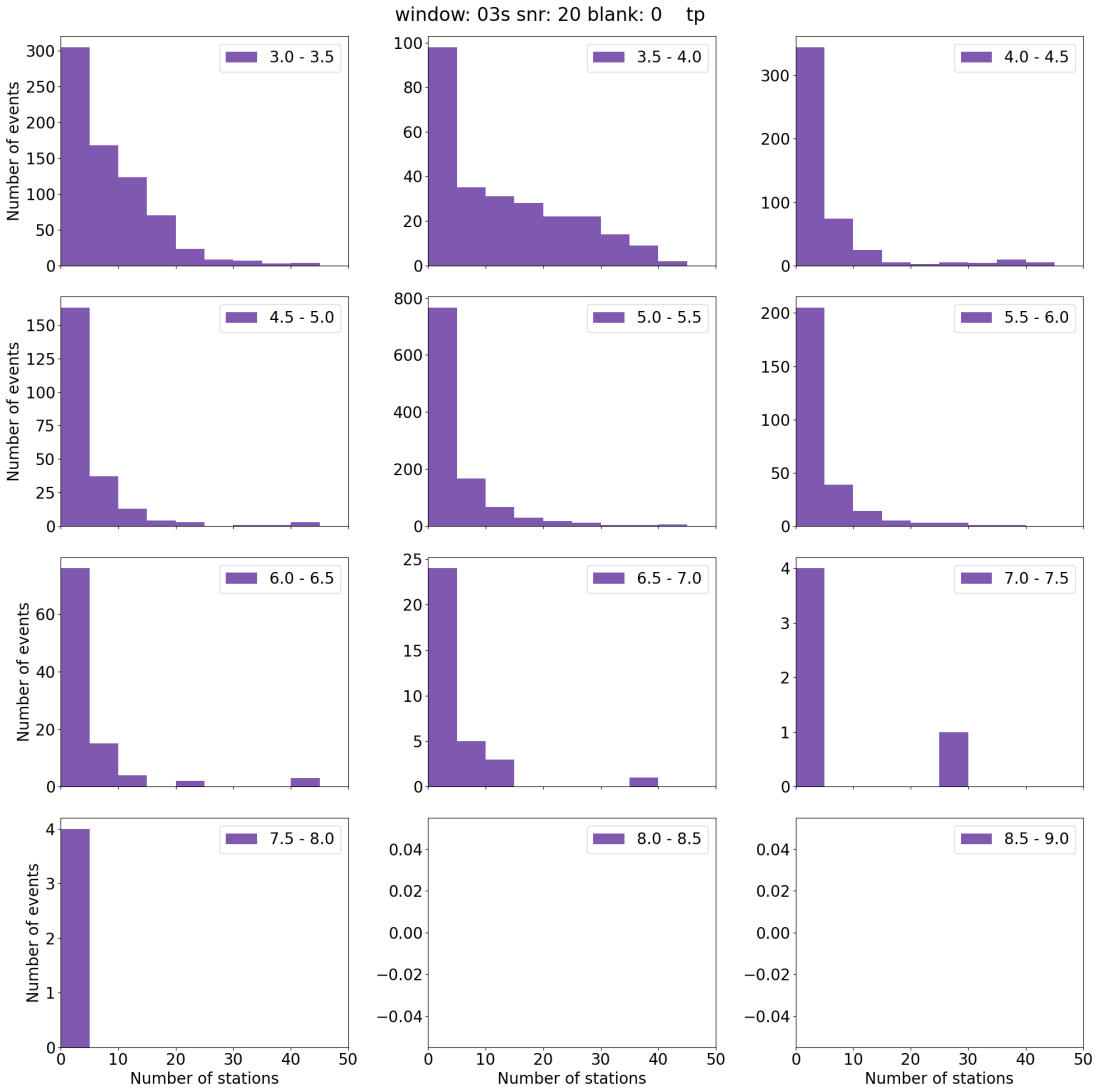


Figure 3.4: Number of stations used in the calculation of $\tau_p^{\max}$ by magnitude.

are recorded at fewer stations.

3.4.5 A note on our approach

The objective of our study is to look for broad-scale, average, relationships between four parameters of early rupture and final magnitude. In such a global approach, we expect a significant scatter. This scatter originates from the intrinsic differences between earthquakes which are independent of magnitude. The events have a range of focal mechanisms, tectonic settings, geological settings, depths, and instrumentation. However, none of these differences are likely to have strong

correlations with earthquake magnitude, so we expect them to add uncertainty to any obtained scaling rather than adding a large bias. This approach is in contrast to most previous studies which generally consider fewer than 100 earthquakes, albeit often recorded at large numbers of stations, in an attempt to isolate similar types of earthquakes and reduce variability due to tectonic setting (Table A.1).

3.5 Results

Once we have filtered and picked the data, calculated the four parameters of interest, and corrected for distance, we obtain values of $\tau_p^{\max}$, τ_c, P_d and IV2 for each recording of our 4143 earthquakes. Figure 3.5 shows the 4 parameters estimated for each earthquake as a function of earthquake magnitude (coloured crosses), along with the median value (grey dot) for each 0.1 magnitude-unit bin. Here we use a 0.3 s time window; we show results for longer time windows in Figures A.23–A.25). Following previous workers, we present $\tau_p^{\max,\Delta t_i}$ and $\tau_c^{\Delta t_i}$ values after averaging over stations, whilst $P_d^{\Delta t_i}$ and $IV2^{\Delta t_i}$ values are presented without averaging.

All four of the datasets show trends where the parameter values increase with increasing earthquake magnitude. Figure 3.6 shows these data again, but now with two best-fitting linear trends: one trend calculated over the full magnitude range and one trend calculated for earthquakes larger than $M4$, whose durations are, on average, longer than the 0.3 s time window. The trends are all positive, with slopes between 0.1 and 0.9. Longer windows are shown in Figures A.26–A.28.

The trends do change with magnitude for $\tau_p^{\max,0.3}$ and $\tau_c^{0.3}$. These parameters clearly increase with magnitude for small earthquakes but plateau slightly above $M6$. Further, all four parameters display significant scatter. In the plots, the scatter appears to increase at $M5$, but that increase is an artifact of the increase in data density at $M5$; There is no statistical increase in variance (see Section A.7, Figures A.29–A.31).

We use four measures to quantify the relationship between our observations of rupture initiation and the final earthquake magnitude (Figures 3.7 and 3.8). These measures are: the gradient of a straight line fit (panel a); the proportion of

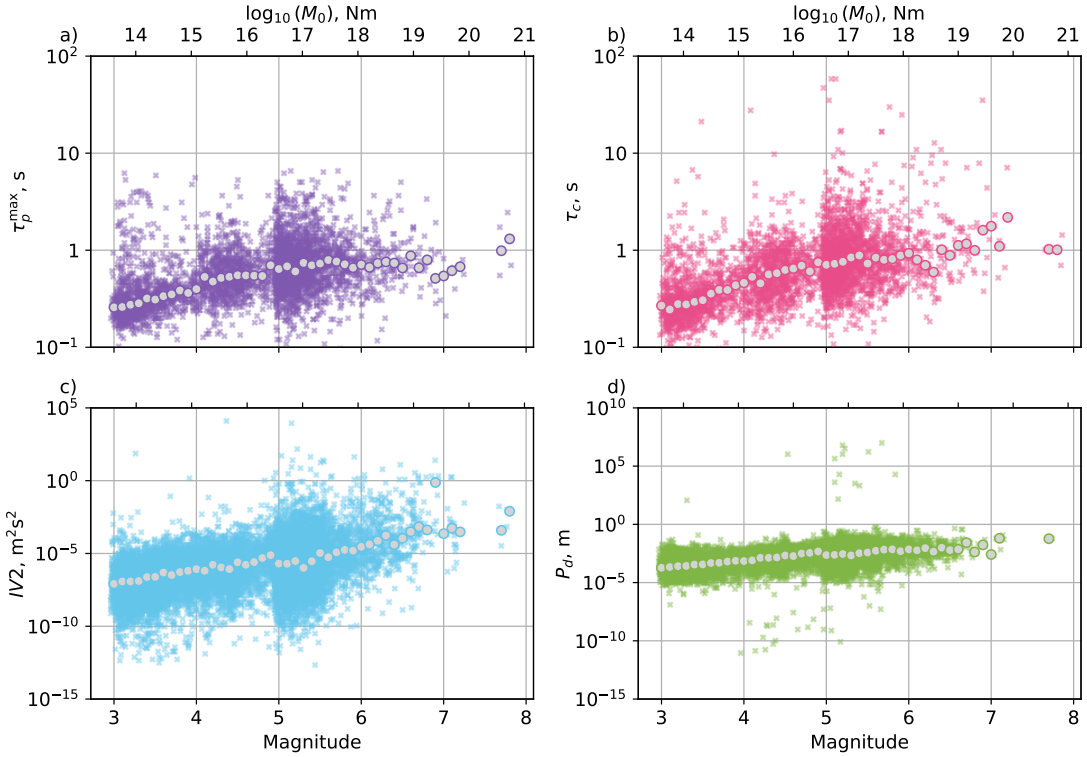


Figure 3.5: Parameters estimated from the early seismograms vs magnitude using a 0.3 s window. The plotted parameters are a) predominant period, b) average period, c) integral of velocity squared and d) peak ground displacement. In a) and b), each cross represents the mean value for a given earthquake, averaged over stations. In c) and d), each cross represents the value at one station corrected to 1 km epicentral distance (see text). Grey circles show the median value for each magnitude bin. Magnitude is preferred magnitude from the USGS (M_W where available).

the variance this straight line explains (panel b); the Spearman's rank correlation coefficient of the dataset (panel c); and the probability of getting this Spearman's rank correlation coefficient from ordinally uncorrelated data (panel d). The Spearman's rank correlation coefficient ($r \in [-1, 1]$) is particularly useful because it tests whether the values in the observed parameters are ordered in the same way as the magnitudes. It does not require data to be normally distributed.

We calculate each of these measures across various magnitude ranges so that we can assess whether the relationships persist if we only consider earthquakes with a duration longer than the observation window. The relationships do persist. For instance all four parameters have a significant relationship ($p < 0.05$) with magnitude, even if we consider 0.3 s windows and $M \geq 5.2$ earthquakes, which are

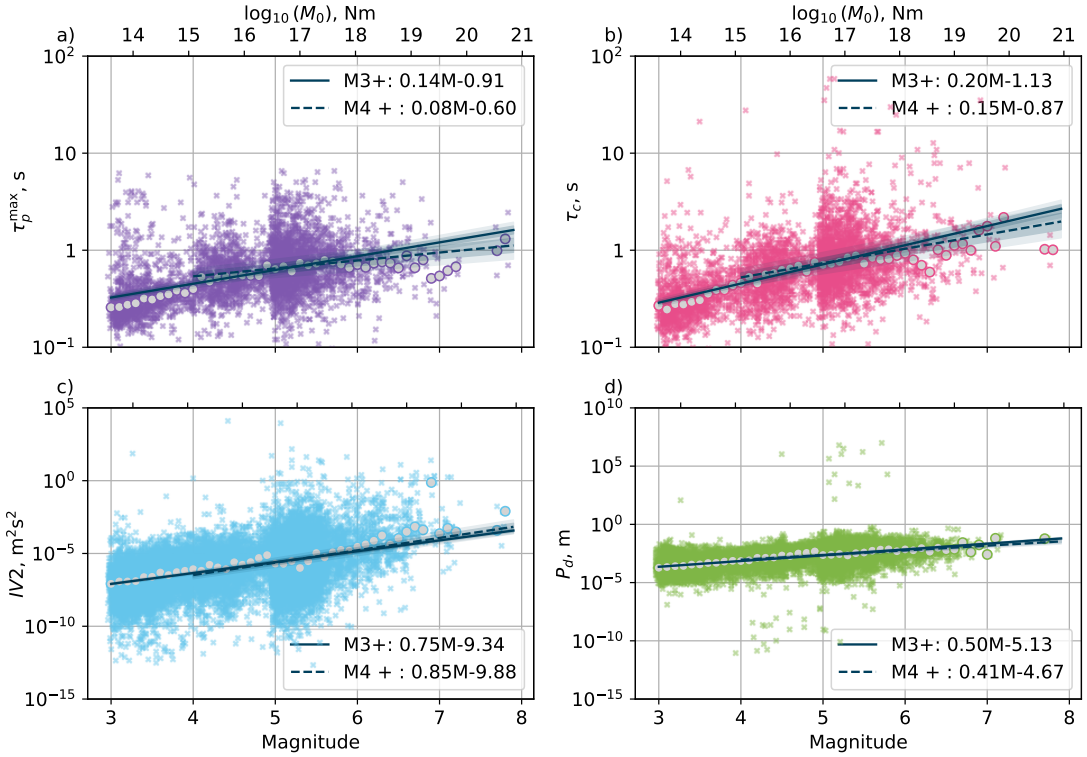


Figure 3.6: The same data as in Figure 3.5 with least squares line of best fit added (shading shows 1 and 2 standard deviations around this line). We calculate the line of best fit for two magnitude ranges: for the whole $M \geq 3$ dataset (solid line), and one where for events where the window is at least as long as the average earthquake duration (dashed line).

typically more than 4 times longer than the 0.3 s window (Figures 3.7 and 3.8).

In the following sections, we summarise the scalings obtained for each parameter, the dependence of those scalings and correlations on the magnitude range considered, and the effects of our parameter choices.

3.5.1 Predominant period

3.5.1.1 Correlation in different magnitude ranges

Given a 0.3 s window, and including only $M > 4$ earthquakes, which are on average longer than 0.3 s, the best fitting line for predominant period is $\log_{10}\left(\tau_p^{(\max,0.3)}\right) = 0.08M - 0.60 (\sigma_{\text{gradt}} = 0.0084)$. This implies that $\tau_p^{(\max,0.3)}$ scales as $M_0^{0.053}$. However, the linear trend is not a perfect match to the data. M3-4 earthquakes prefer a steeper trend and are offset below the line of best fit of the whole dataset. When

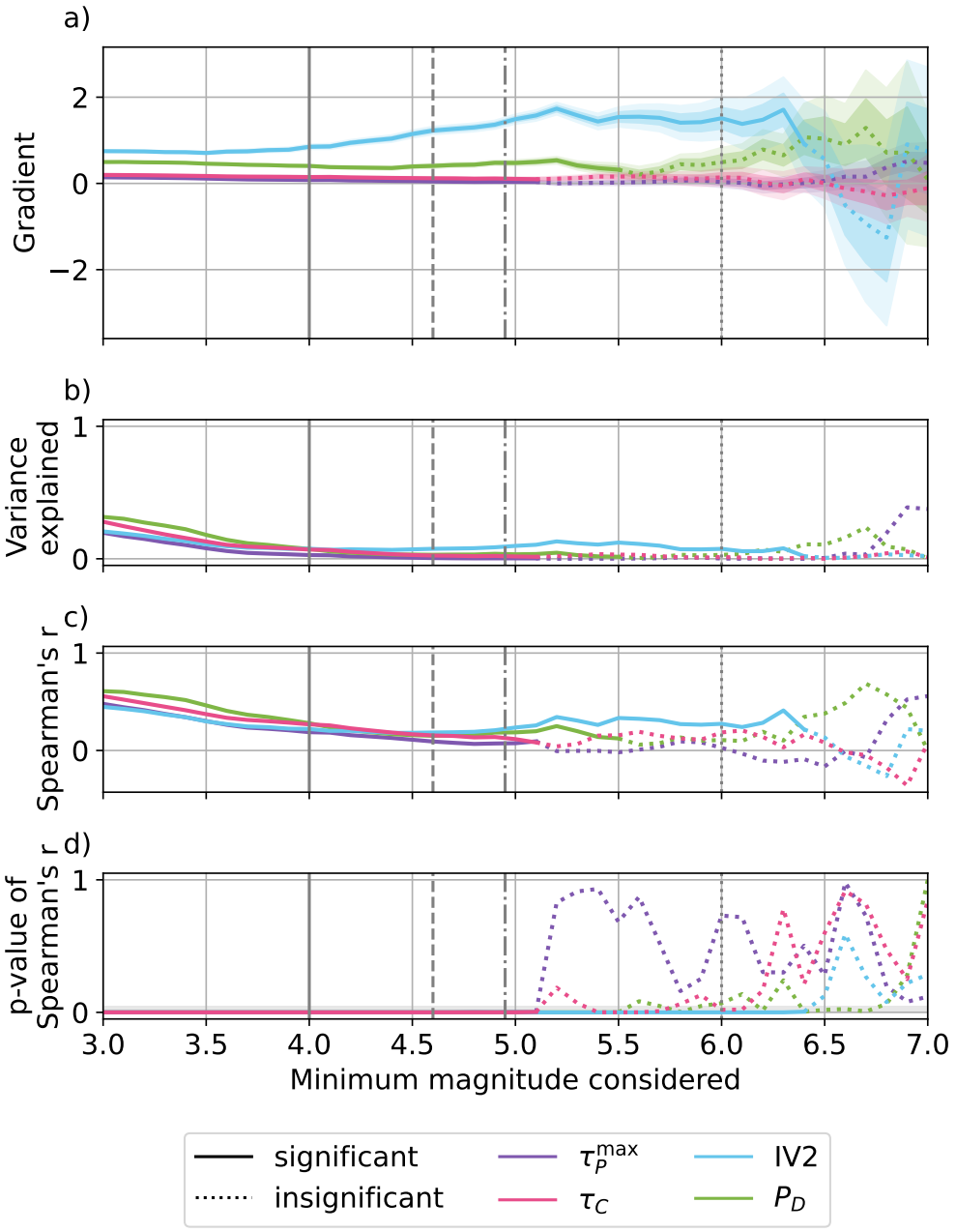


Figure 3.7: Statistics for analysis using a 0.3 s window. The value plotted at each magnitude is the value we obtain when we fit the data using only earthquakes with that magnitude or larger. Panel a) shows the gradient of a linear fit, panel b) shows the proportion of the variance in the data explained by that straight line, panel c) shows the Spearman's rank correlation coefficient and panel d) shows whether this correlation is statistically significantly different from a linearly uncorrelated dataset. Grey shaded bar in panel d shows the region where $p < 0.05$. Vertical lines denote average magnitudes of earthquakes with durations equal to the window length (solid), twice the window length (dashed), three times the window length (dash-dot) and ten times the window length (dotted).

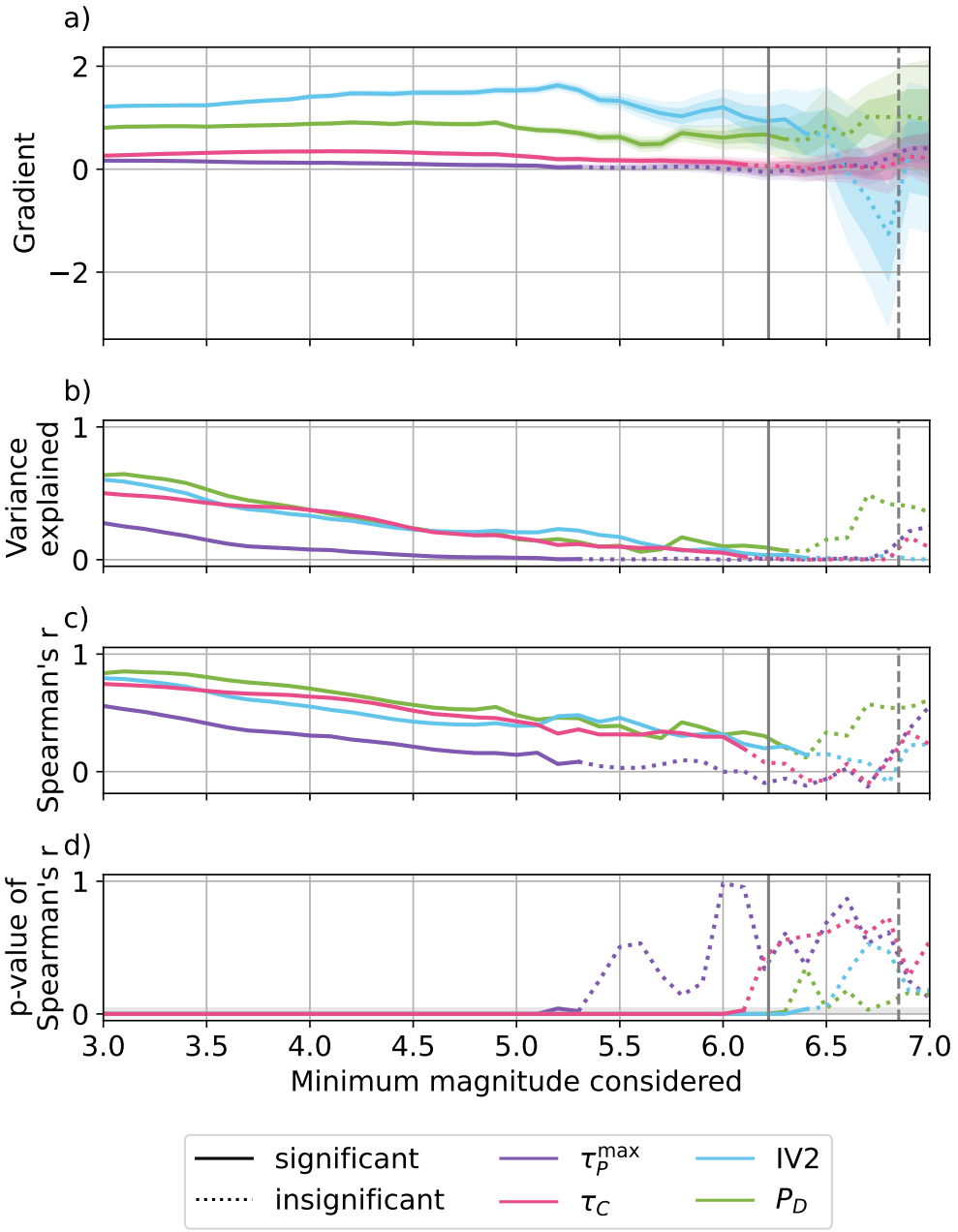


Figure 3.8: As Figure 3.7 but using a 4 s analysis window.

M3-4 earthquakes are included in the fit, $\tau_p^{\max,0.3} = 0.14\text{M-}0.91$ or equivalently, $\tau_p^{\max} \propto M_0^{0.093}$.

There is still a significant correlation when we consider only $M \geq 5$ earthquakes. The correlation only ceases to be significant when the minimum magnitude exceeds 5.1—5.4 (solid to dashed transition; Figures 3.7, 3.8, A.32, A.33).

3.5.1.2 Window length and magnitude range

Perhaps surprisingly, the magnitude at which significance is lost is unaffected by changing the calculation window length, Δt_i. This independence may give us insight into the processes that determine $\tau_p^{\max}$ (see section 3.6.1). The window length also only modestly affects the variance explained by a linear fit (Tables A.4). Increasing Δt_i changes the fraction of the variance explained from 44% to 53%. Similarly, increasing Δt_i does not significantly increase the Spearman's correlation coefficient (Table A.6).

3.5.1.3 Additional effects: Source-receiver distance

As we discussed in Section 3.4.3, the Earth acts as a lowpass filter on the seismic waves which propagate through it. We might expect that data recorded at stations which are further from the source would have undergone more attenuation, and so the predominant period values we find would be longer. Such distant observations may be more likely to record larger earthquakes and thus could introduce bias.

We now explore how our results change if we exclude observations close to the source. Specifically, we examine how a minimum station-receiver distance affects the magnitude at which the significance is lost (the point at which the line becomes dashed on figures like Figure 3.7). Figure A.34 shows that the distance range does not appear to bias the $\tau_p^{\max}$ calculation. For minimum source-station distances of less than 60 km, the minimum distance has very little effect on the magnitude at which significance is lost. Above this, there is a decrease in the magnitude which is largely a result of fewer earthquakes being recorded at stations this far away.

3.5.2 Average period

3.5.2.1 Correlation in different magnitude ranges

The scaling between average period and moment for $M > 4$ earthquakes is stronger than the predominant period-moment scaling: $\tau_c^{0.3} \propto M_0^{0.13}$, while $\tau_p^{\max,0.3} \propto M_0^{0.093}$.

The median values generally closely follow the line of best fit. However, for longer time windows $\Delta t_i = 1$ or 4 s (panel b of Figures A.25, A.24, A.28, A.27), τ_c

plateaus at a value around 3 s ($\log_{10}(\tau_c) = 0.5$) for the largest magnitudes, and the τ_c-magnitude slope is slightly larger when $M3$-4 earthquakes are included.

3.5.2.2 Window length and magnitude range

As before, we can consider the effect of gradually increasing the minimum magnitude of the dataset. Unlike for $\tau_p^{\max}$, when we use a longer window length, a significant relationship persists to larger minimum magnitudes (solid to dashed transition, Figures 3.7, 3.8, A.32, A.33).

Table A.4 shows the variance explained by linear fits of results for various window length (Figures 3.7, 3.8, A.32, A.33). Using a longer time window allows us to explain more of the variance in the data, increasing from 53% of the variance with $\Delta t_i = 0.3$ s to 71% with $\Delta t_i = 4$ s. However, if we look only at earthquakes with durations on average as long as the window, the variance does not increase with window length; 26% of the variance is explained by a fit on $M \geq 4$ earthquakes in a 0.3 s window, and 24% is explained by a fit on $M \geq 5.2$ earthquakes in a 1 s window. As we might expect, the correlation coefficient of average period based on all data increases as we increase Δt_i (Table A.6).

3.5.2.3 Additional effects: Source-receiver distance and number of stations

Average period is less sensitive to the number of recording stations than predominant period. The gradient of the line of best fit is statistically the same whether we consider earthquakes measured at one, two or three stations. This suggests the average period is a more robust measure, and is more likely to be able to be coherently calculated at multiple stations. Again, above a minimum source-station distance of 60 km, the magnitude at which the results become insignificant decreases significantly, due to smaller numbers of earthquakes fitting these criteria (Figure A.34).

3.5.3 IV2

3.5.3.1 Correlation in different magnitude ranges

IV2 shows the strongest magnitude scaling of all the parameters, with a gradient of 0.75 which implies that $IV2 \propto M_0^{0.5}$. IV2 is the only parameter where the gradient gets steeper when we consider only $M \geq 4$ events (0.85). Similar trends exist for the 0.5 and 1 s windows.

Figure 3.5c shows a step in the IV2 data at $M5$, with a steadily linear increase between $M3-5$ and an offset, less perfectly linear trend from $M \geq 5$. There is also still some variability in the medians, particularly when we include the longer dataset above $M5$.

The IV2 values appear more scattered than the results for average or predominant period. The scatter is partly due to the change in what is plotted; for $\tau_p^{\max}$ and τ_c (panels a and b), we plot the median value over all recordings of an earthquake. On the other hand, we plot the IV2 value (panel c) for every recording following previous workers.

The correlation between IV2 and magnitude remains significant until $\sim M6.4$, regardless of window length (solid to dashed transition, panel b in Figures 3.7, 3.8, A.32, A.33). For a 0.3 s window, this is a significantly higher cut-off than for the other parameters.

3.5.3.2 Window length and magnitude range

With IV2, we see a significant increase in the Spearman's correlation coefficient as we consider longer Δt_i, increasing from 0.45 at $\Delta t_i = 0.3$ s, to 0.79 with $\Delta t_i = 4$ s.

The gradient also increases with Δt_i and has the highest gradient in the 4 s window (1.21).

3.5.3.3 Additional effects: Source-receiver distance and number of stations

In Figure A.34, we can see that IV2 is consistently significant to $M6.4-6.6$, and that this depends only on the number of stations considered (colour), rather than the minimum source-receiver distance (x-axis). This suggests our correction for

geometric spreading is appropriate and that the effect of attenuation is negligible. For events recorded at 5 or more stations, the significance is lost 0.1 magnitude units earlier, probably because there are fewer earthquakes to analyse.

3.5.4 Peak Ground Displacement

3.5.4.1 Correlation in different magnitude ranges

For $\Delta t_i = 0.3$ and using the whole dataset, $\log_{10}(P_d^{0.3}) = 0.50M - 5.13$, implying that $P_d \propto M_0^{0.33}$. This relationship is stronger than for $\tau_c^{0.3}$ and $\tau_p^{(\max,0.3)}$ but weaker than that of IV2.

The individual P_d results are consistently clustered around a trend, and the increase in data density at $M \geq 5$ is the least apparent compared to other parameters. Whilst there are some significant outliers (as with IV2), these do not appear to be systematic.

In a 0.3 s window, the correlation between $P_d^{0.3}$ and magnitude is significantly distinguishable from linearly uncorrelated data until M_{min}=5.5 (average duration = 1.7 s; solid to dashed transition, Figure 3.7).

3.5.4.2 Window length and magnitude range

Changing the window length changes the distribution of points on the P_d-magnitude plot. For short Δt_i, two trends are clear in the median values: an almost perfect linear trend between $M3$ and $M5$, and a more varied, shallower slope trend from $M5$ upwards (panel d of Figures 3.5, A.23, A.24, A.25). As we look at longer values of Δt_i, this break in scaling disappears and the medians appear to fall on a single linear trend. As Δt_i increases, the gradient of the best fit line gets steeper (panel d of Figures A.25, A.27, A.24, A.27) and the break in significance increases from M_{min} =5.5 to 6.3 (average duration = 4.4 s, Figures 3.7, 3.8, A.32, A.33).

3.5.4.3 Additional effects: Source-receiver distance and number of stations

The P_d results are minimally affected by changing the minimum source-station distance or by the number of recording stations (Figure A.34). The geometric

spreading correction appears suitable, leaving no obvious effect of attenuation.

3.5.5 Impact of pick uncertainty on results

We are perhaps the first to use machine-learning based picks for analysis of the early windows of waveforms. Our ability to define our windows depends on the temporal accuracy of the ML picks. Münchmeyer et al. (2022) test a range of ML pickers on several datasets, and investigate the accuracy of the generated picks compared to manual P wave picks. The most applicable datasets to our use-case are the ETHZ, SCEDC and STEAD datasets, which contain local-to-regional waveforms and do not show any systematically higher residuals (see Section 3.3 of Münchmeyer et al., 2022, for further analysis of all the different datasets). The mean absolute errors (MAE) are ETHZ: 0.11, SCEDC: 0.12 and STEAD: 0.07. The STEAD dataset contains global events whereas the SCEDC and ETHZ are both region specific. The ETHZ dataset contains $\sim$ 200 times fewer events than the SCEDC and STEAD datasets. The distributions are centred on zero, but visual inspection of the pick residuals (Münchmeyer et al., 2022, Figure 2) shows that the mean and median residual are slightly negative and thus the predicted picks are more likely to be early than late. We explore the influence of the pick times by varying the pick by 0.1 seconds in both directions and look at the impact on the magnitude at which there is no longer a significant correlation (Table 3.1), and on the gradient of the relationship between the parameter and magnitude (Table 3.2).

A shift of -0.1 moves the pick 0.1 seconds earlier, essentially correcting for systematically late picks. Equivalently, a shift of $+0.1$ s moves the pick 0.1 seconds earlier and corrects for systematically early picks. In the 0.3 second window, we find that a pick shift of -0.1 has no impact on the magnitude at which significance is lost for τ_c, P_d and IV2, though the gradients of the relationships of these parameters with magnitude are slightly shallower with the shifted pick. However τ_c no longer shows a relationship with magnitude for any earthquakes with duration longer than the window. A shift of $+0.1$ s increases the magnitude at which significance is lost for τ_c and has no impact on the other parameters. We find similar results

Table 3.1: Impact of shifting the pick on the magnitude at which significance is lost. Window = 0.3 s (Average duration of $M4$ is 0.3 s).

Pick shift	τ_p	τ_c	P_d	IV2
-0.1	3.6	5.1	5.5	6.4
0	5.1	5.1	5.5	6.4
+0.1	5.1	5.6	5.5	6.4

Table 3.2: Impact of shifting the pick on the gradient between the parameter and earthquake magnitude. Window = 0.3 s (M=4).

Pick shift	τ_p	τ_c	P_d	IV2
-0.1	-0.01	0.14	0.31	0.89
0	0.10	0.15	0.41	0.86
+0.1	0.12	0.19	0.50	0.91

for the other windows (Tables A.7–A.12) with the exception of τ_c with a shift of -0.1, where the magnitude at which significance is lost is just 5.1 compared to 5.6 for shifts of 0 and $+0.1$.

These results suggest that our results for IV2 and P_d are likely insensitive to uncertainty in the picks. The τ_c parameter maintains a relationship with shifts of $+0.1$ and -0.1 when considering events longer than the calculation window, but these relationships become weaker in some circumstances. The $\tau_p^{\max}$ parameter is the most sensitive to shifting the pick; it no longer shows a significant relationship with magnitude when the pick is shifted by -0.1 seconds. This is not unexpected, since if the picks are correct then the window will now include a significant portion of noise before the earthquake has arrived at the station. Olson and Allen (2005) do not use the first 0.05 s after the P wave pick due to the recursive nature calculation and potential 'leakage of the frequency content of background noise'. Looking at the equivalent of Figure 3.6, the $\tau_p^{\max}$ points are now much more scattered with a higher proportion having high values of $\tau_p^{\max}$ (Figure A.35). This suggests that the events with high values of $\tau_p^{\max}$ may partly result from leakage from the time before the earthquake's arrival, suggesting that these picks were slightly early compared to a true pick. We also note that shifting the pick by -0.1 exacerbates the jump in $\tau_p^{\max}$ values at $\sim M4$ that we observed in the non-shifted pick results (Section

3.5.1) with a much clearer high gradient segment for $M < 4$ and a flat segment for $M > 4$ (Figure A.35). Shifting the pick by +0.1 reduces this feature (Figure A.36) compared to the zero-shift results. Results for shifting the pick in other windows are shown in Figures A.37–A.42.

3.6 Discussion

We have determined four properties of earthquake initiation records using a dataset of 4143 earthquakes from around the world. In our efforts to image the early portions of rupture, we calculated the parameters over four time windows. We find a statistically significant relationship between all four parameters and magnitude when calculated in a 0.3 s window, and these relationships persist even when we only analyse earthquakes which are on average at least three times as long as the window.

An obvious question to ask is why the beginning of the rupture should be correlated with its final size. In what follows, we speculate about what physical processes these parameters measure.

3.6.1 Predominant Period

We, like previous investigators (e.g. Olson and Allen, 2005; Tsang et al., 2007; Wurman et al., 2007), find that the predominant period in the first portion of the seismogram is correlated with the final earthquake magnitude. However, what predominant period actually measures remains unclear.

One idea is that the predominant period could measure the time to rupture an initial asperity. Larger, wider early asperities could lead to stronger earthquakes and higher magnitudes. The idea of a dominant early asperity rupturing within the first 0.3 s is consistent with our observations in different time windows; increasing the measurement window to 1, 2, or 4 s only modestly changes observed $\tau_p^{\max}$ values.

For $M5$ earthquakes, the median value of $\tau_p^{\max}$ is 0.63. If we assume the earthquake grows at a constant 2500 ms^{-1} (something which may not be the case, e.g. Wang et al., 2016), and take a slip duration of half the predominant period, we find a length scale of $\sim$ 790 m which is a plausible asperity size. Inferred

asperity sizes get closer to the final rupture length for small earthquakes: for $M3$ earthquakes, the median value of $\tau_p^{\max} = 0.25$ giving a length scale of ~ 300 m.

Since we have an estimate of the $\tau_p^{\max}$-moment scaling ($\tau_p^{\max} \propto M_0^{0.093}$), we can calculate a potential scaling between r_0, the radius of an initial asperity, and r_f, the final rupture radius. If we assume that rupture initially grows at a constant rate V_r, and that earthquakes have magnitude-independent stress drops $\Delta\sigma_s$, so that $r_f \propto M_0^{1/3}$ (Scholz, 1982), we arrive at $r_f \propto r_0^{3.57}$. This scaling, if true, implies that as we consider bigger and bigger earthquakes, the total rupture size begins to grow much faster than the initial asperity size, which changes minimally.

The initial asperity size, as inferred from the τ_p values, may even plateau for large-magnitude earthquakes. Visual inspection of Figures 3.5–3.7 suggests that the $\tau_p^{\max} - M_0$ gradient is roughly constant for $\sim M4$–$M6$ earthquakes, but for the largest earthquakes ($M \geq 6$), the gradient approaches zero. Indeed, when we consider only earthquakes larger than $M5.1$, predominant period has no significant correlation with final earthquake magnitude. The transition to insignificant correlation occurs around $M5.1$ regardless of the window length Δt_i (Figures 3.7, 3.8, A.32, A.33, A.43). Through numerical simulations, Murphy and Nielsen (2009) found that a scaling between the peak ground displacement and magnitude occurred for ruptures up to four times the initial asperity, which might suggest a similar saturation to our observation of predominant period.

There is, however, a potential window length effect when we examine the full range of earthquakes, including many $M3-4$ earthquakes that are shorter than our window length of interest. The $\tau_p - M$ gradient is stronger when we consider such short earthquakes (Figures A.43 and A.44), but in that case, τ_p may measure the full rupture duration, and we may simply see a relationship between duration and size.

We note that the sensitivity of $\tau_p^{\max}$ to uncertainty in the picks makes these results the least robust compared to the other parameters.

3.6.1.1 Comparison with past results

At first glance our results generally agree with previous work (e.g. Olson and Allen, 2005; Lockman and Allen, 2007; Tsang et al., 2007; Wurman et al., 2007; Brown et al., 2009) who found that the gradient of the relationship between $\log_{10} \tau_p^{\max}$ and magnitude to be between 0.14 and 0.21. We find a gradient based on $M > 3$ events of 0.14 in a 0.3 s window and 0.17 in a 4 s window. However, when we consider only larger magnitude events, we find that the gradient of the relationship decreases and eventually becomes indistinguishable from 0, unlike Lockman and Allen (2007) who found a higher gradient with $M > 5$ events compared to $M < 5$ events. We find no statistically significant relationship for $M > 6$ events unlike Shieh et al. (2008), but in agreement with Rydelek and Horiuchi (2006). Both Wolfe (2006) and Shieh et al. (2008) raise concerns about the stability of the parameter, and our results found from varying the pick timings do not alleviate this.

3.6.2 Average period

We also find a relationship between average period in a short initial window and final earthquake magnitude, and like previous investigators find that it is stronger than the relationship between predominant period and magnitude (e.g. Shieh et al., 2008). Like predominant period, estimates of average period, τ_c, could also be interpreted as a reflection of some initial asperity or asperities. However, the scalings are slightly different: $\tau_c \propto M_0^{0.13}$, while $\tau_p^{\max} \propto M_0^{0.09}$, so the asperity sizes estimated from this average period increase more rapidly with earthquake magnitude. For instance, we calculate a characteristic length scale for $M5$ earthquakes of 890 m. When we derive a scaling between the 'asperity' size r_0 and the final rupture size, r_f, we obtain $r_f \propto r_0^{2.56}$.

In a broad sense, we can make similar inferences about asperity effects here as we did with predominant period; earthquakes with larger initial asperities tend to grow bigger. However, while predominant period could highlight a dominant individual asperity, average period likely accounts for a number of complexities in

the earthquake rupture recovering an average period from all asperities ruptured in the first 0.3 s.

The ability of average period to incorporate rupture complexities are consistent with trends in its gradient when we change the window length. Both the gradient and the significance of the $\tau_c - M_0$ trend increase as we increase the duration of the measurement window relative to the earthquake duration (Figure A.45 and A.46).

Alternatively, one could choose to consider average period without considering asperity sizes. It is possible to frame the calculation as a ratio of different types of energy or as a normalised measure of frequency, but the physical intuition for those approaches is not obvious to us.

3.6.2.1 Comparison with past results

Other authors observe gradients of the relationship between $\log_{10} \tau_c$ and magnitude to be between 0.22 and 0.32 (Kanamori, 2004; Wu and Kanamori, 2005a; Wu and Zhao, 2006; Wu et al., 2006, 2007; Wu and Kanamori, 2008a,b). Most papers use a 3 s window and thus their results are most comparable to our 1 s (gradient = 0.29) and 4 s (gradient = 0.26) windows. Some papers (e.g. Wu and Kanamori, 2008b) find a significant relationship even with $M > 6$ events, but we find no significant relationship at these high magnitudes. We again see much more scatter in our results than other studies, which most likely is due to the large numbers of events and global focus. Wu et al. (2007) and Wu and Kanamori (2005a) both find a high scatter in the results which Wu et al. (2007) attribute to low SNR in the windows under consideration due to the seismometers used and the low overall amplitudes for some small events in the 3 s window they use.

3.6.3 IV2

In contrast to predominant and average period, IV2—the integral of ground velocity squared—is intuitively a measure of energy. As previously discussed, IV2 scales linearly with radiated seismic energy; this scaling follows from considering the integral of the energy flux through a shell (see eq 3–8 of Kanamori et al., 1993).

Since radiated energy is observed to increase with earthquake moment, often quasi-linearly (Abercrombie, 2021), it is unsurprising that IV2 increases with moment when the measurement window Δt_i is longer than the earthquake duration, and we have measured the whole earthquake (similar to e.g. Lancieri et al., 2011). Indeed, in this case, we would expect $IV2 \propto M_0$: not much above the power-law we obtain with our longest (4 s) windows: $IV2 \propto M_0^{0.81}$.

It is more surprising to see how IV2 scales with moment when we consider observations with different window lengths relative to the earthquake durations (Figures 3.7, 3.8, A.32, A.33, A.47 and A.48). Specifically, when we reduce the window length or consider longer earthquakes, so that the IV2 analysis considers a smaller fraction of the earthquake duration, the IV2 $-M_0$ scaling exponent *increases*. The exponent reaches a value of 1, for a linear scaling, only when we consider earthquakes that are 10 times longer than the 0.3 s window length.

This leads to several seemingly conflicting results. We obtain a linear IV2$-M_0$ scaling when we consider small windows of the earthquakes' beginnings; others observe a linear scaling between radiated energy and magnitude when they consider the whole earthquake (e.g. Abercrombie, 2021); yet when we consider datasets where window lengths range from small-to-large fractions of the earthquake duration, the scaling relationship is weaker.

These results could be reconciled if there are two physical processes occurring: one process that determines the whole-earthquake energy-moment scaling and another process that determines the relationship between the early energy and the final rupture. Analyses with a range of durations could be averaging the two; they may give a shallower gradient by averaging values from two parallel trends.

Here we are interested in the relationship between the early energy and the final earthquake magnitude. As noted earlier, we find a statistically significant trend between IV2 and magnitude ($IV2 \propto M_0^{0.98}$) even when the measurement window is just 10% of the average duration of the smallest earthquakes included (dotted vertical line on Figure 3.7). This trend suggests that ruptures which start more energetically tend to rupture farther and thus become higher magnitude

earthquakes. Perhaps these energetic ruptures are more able to overcome barriers, similar to the ideas of Heaton (1990).

It is relatively intuitive to think that earthquakes that start with higher energies can rupture further. However, the magnitudes of the early energy implied by the quasi-linear IV2 $-M_0$ scaling is surprisingly dramatic. For instance, our measurements imply that a typical $M6$ radiates 30 times more energy in the first 0.3 seconds than a typical $M5$. That higher energy from the $M6$ could come from larger initial asperities, as suggested by the τ_p scaling. Or it could come from larger initial stress drops or rupture heterogeneities.

3.6.3.1 Comparison with past results

Most previous work has looked at IV2 values in 2 and 4 s windows. We find a similar gradient between $\log_{10}$IV2 and magnitude to Lancieri et al. (2011) across the whole magnitude range, and a higher gradient than that found by Peng et al. (2013).

Festa et al. (2008) see a saturation in IV2 at around $M5.8$. We do not see a systematic reduction in the gradient as we consider larger magnitude events. The only observation which could hint at saturation is that, in the 4 s window, the gradient when we only consider events longer than the window ($M \geq 6.25$) is reduced compared to when we use all events. This is the only window where this occurs. Festa et al. (2008) attribute the break in scaling that they observe to the time-window Δt_ϕ entirely containing the direct waves. This suggests that any scaling over a wider magnitude range (valid for $M > 5.8$) must reflect the slip. However, in their calculations for $M > 5.8$ events the rupture area was fixed to be the same as a $M5.8$, which may also impact the scalings found.

3.6.4 Peak Ground Displacement

Our observed correlation between early peak ground displacement and final earthquake magnitude is also somewhat intuitive, and aligns with some (e.g. Wu et al., 2007; Wurman et al., 2007; Brown et al., 2009; Noda and Ellsworth, 2017) but not all previous results (e.g. Meier et al., 2016; Colombelli et al., 2020). Peak ground

displacement is generally considered to be a measure of the rate of moment release in an earthquake (Aki and Richards, 2009, section 4.3). This $P_d - M_0$ correlation thus suggests that larger earthquakes tend to have a higher rate of moment release early on. As with IV2, this intuitively makes sense—earthquakes which are more energetic (or release moment more quickly) at the beginning of rupture may be better able to overcome stress or frictional barriers and become higher-magnitude earthquakes.

The gradient of the relationship between P_d and M_0 clusters around 0.5, equivalent to $P_d \propto M_0^{1/3}$. This scaling is weaker than that observed in studies that examine the full rupture duration, which often find that $P_d \propto M_0$. There is no inconsistency here, however; models and observations of earthquakes' source time functions suggest that maximum moment rates typically occur about a third to halfway through the rupture duration (Houston et al., 1998; Persh and Houston, 2004; Juhel et al., 2018), and we focus on shorter time intervals at the beginning of rupture.

We find no systematic variations in the $P_d - M_0$ scaling exponent as we change the size of the window considered relative to the beginning of rupture (Figures A.49 and A.50). The lack of trend suggests that very few of our observations occur during the short nucleation phase observed by Ellsworth and Beroza (1995), where observed moment release was lower than in the rest of the earthquake. However, the P_d values may not be very easily interpretable, as values can be modified by filtering as part of the processing or naturally arising from the path effect.

Nevertheless, it is interesting to note how much P_d changes with magnitude when interpreted at face value. The changes in inferred early moment rate are smaller than the changes in inferred radiated energy. For instance, we find that the early moment rate of a typical $M6$ should be 3 times larger than the early moment rate of a typical $M5$.

3.6.4.1 Comparison with past results

Not all previous studies which have used P_d provide quantitative information on the relationship between P_d and magnitude and those that do show significant variation in their results.

Depending on the window used, we find a gradient of between 0.50 and 0.80 between $\log_{10} P_d$ and magnitude for $M \geq 3$ earthquakes. This relationship has a slightly lower gradient (0.41–0.67) when we consider only earthquakes longer than the calculation window. These gradients are similar to but slightly lower than what other authors have found in 2 and 4 s windows (Lancieri and Zollo, 2008; Lancieri et al., 2011; Peng et al., 2013). Lancieri et al. (2011) finding slightly higher gradients for results calculated on broadband records compared to strong-motion records. Peng et al. (2013) attribute the lower gradient that they find compared to previous work done in Taiwan and California (Zollo et al., 2006, 2007; Wu and Zhao, 2006) to this difference between strong-motion and broadband results, since low-frequency noise may be more dominant for the twice-integrated strong-motion records. However we find a gradient which is more comparable to studies using strong-motion records. This may imply that our filtering parameters or attenuation correction may need further tuning, or may simply result from regional or tectonic setting variations in the scaling between P_d and magnitude.

Whilst we see a similar trend to previous authors, the absolute values of our results are lower, even calculated in comparable window lengths. This likely results from the varying approaches to distance corrections that different studies use.

However, Colombelli et al. (2020) find the opposite relationship to us, Colombelli et al. (2014) and all the previously mentioned studies: larger events have lower peak displacements in the first second. Colombelli et al. (2020) ascribe the difference between their results and Trugman et al. (2019) to several factors including variation in pick times and the spacing of P_d calculations. Given the variation in human-analyst picks (e.g. Chai et al., 2020), and ML picks, we investigated the impact of varying our picks by ± 0.1 s, and find no impact on our results.

One explanation could be that Colombelli et al. (2013) use a large number of waveforms but a small number of events (48) from Japan and we do not include Japanese seismicity in our study. However, there is no physical reason to think that Japanese seismicity may be unique (given that we include other subduction zones) and Brown et al. (2011) find a similar scaling to us using (slightly more)

earthquakes from Japan. The scaling found by Brown et al. (2011) in Japan of $\log_{10} P_d \sim 0.66 M_L$ is slightly weaker than that found in a comparable study of California (Wurman et al., 2007, gradient = 0.73), but is crucially still positive.

Lancieri et al. (2011) find a significant relationship for only $M \geq 6$ events in both 2 and 4 s windows, in contrast to Lancieri and Zollo (2008) who find a saturation in P_d at $M6.5$ for the 2 s window. Peng et al. (2013) identified that the scaling between P_d and magnitude is weaker and shows more scatter for $M > 6$ earthquakes, though this may result from fewer events going into this estimate. Rydelek et al. (2007) use a 2 s window and similarly find a reduction in the gradient of the relationship when all magnitudes are considered compared to only considering $M < 5.5$, whilst Wu and Zhao (2006) find a relationship for events $M < 6.5$.

We do not conclusively see any evidence of saturation, unlike Trugman et al. (2019) and Lancieri and Zollo (2008). Trugman et al. (2019) expect P_d to saturate at $M5.7$ for a 1 s window. We find a statistically significant relationship ($\alpha = 0.05$) between P_d and magnitude with minimum dataset magnitudes up to $M6$ (Figure A.33); however, we do see a gentle increase in the p-value of the relationship being explained by ordinally uncorrelated data from $M5.7$ (bottom panel of Figure A.33). These results therefore do not confirm or deny the findings of Trugman et al. (2019). However, we also calculate P_d in shorter time windows, which better places us to probe the potential saturation of P_d. We find that the relationship between P_d and magnitude still persists up to minimum dataset magnitudes of $M6$, even though these windows are less than one-third of the average duration of a $M6$.

3.7 Conclusions

In this study we calculated four parameters (predominant period, average period, integral of the velocity squared, peak ground displacement) in short windows (0.3, 0.5, 1 and 4 s) at the beginning of over 4000 $M \geq 3$ earthquakes from around the world. We investigated their relationships with the earthquakes' final magnitude.

These four parameters have all previously been found to correlate with final magnitude in regional studies, often with a limited set of earthquakes (Table A.1).

Our results show that these correlations persist in a larger global dataset, even with limited processing. The observed scalings imply that, on average, larger earthquakes have distinct features in their early stages. They have longer predominant and average periods and higher values of P_d and IV2. These features can be measured and significantly correlated with final magnitude even when the measurement window is less than one-third to one-tenth of the typical duration of the smallest earthquake considered.

As our results inherently have more scatter than results from smaller, local studies, we do not find relationships that can be used as diagnostic for any particular earthquake (e.g. Wurman et al., 2007; Shieh et al., 2008; Lancieri et al., 2011). We can, however, consider physical processes that could match our observations. For instance, it may be that larger earthquakes tend to have larger initial asperities and therefore radiate more energy, and release more moment in their earliest stages.

Open Research Section

Code is available on GitHub and is released on Zenodo as Colquhoun (2024).

Results files from the processing are also available on Zenodo.

Processing was primarily in Python 3.10 and was undertaken using SciPy (Virtanen et al., 2020), NumPy (Harris et al., 2020), pandas (pandas development team, 2024; McKinney, 2010), and ObsPy (Beyreuther et al., 2010). Plotting used Matplotlib (Hunter, 2007), Generic Mapping Tools (Wessel et al., 2019) and Crameri colormaps (Crameri, 2023; Rollo, 2024). Phase picking used SeisBench (Woollam et al., 2022) and the EQTransformer model (Mousavi et al., 2020).

Raw data were collected from open-access repositories via IRIS. Networks used were: Federal Institute for Geosciences and Natural Resources (1976); Institut de physique du globe de Paris (IPGP) and École et Observatoire des Sciences de la Terre de Strasbourg (EOST) (1982); MedNet Project Partner Institutions (1990); KNMI (1993); Istituto Nazionale di Geofisica e Vulcanologia (INGV) (2005); KNMI (2006); Institut De Physique Du Globe De Paris (IPGP) (2008); Observatoire Volcanologique Et Sismologique De Martinique (OVSM) (2020); Observatoire

Volcanologique Et Sismologique De Guadeloupe (OVSG) (2020); University of Puerto Rico (1986); ZAMG - Zentralanstalt für Meterologie und Geodynamik (1987); Geological Survey of Canada (2013); National Centre for Seismological Research (CENAIS Cuba) (1998); Ohio Geological Survey (1999); Lamont Doherty Earth Observatory (LDEO), Columbia University (1970); Penn State University (2004); University of Oregon (2002); University of Washington (1963); Institut Cartogràfic i Geològic de Catalunya (1984); Idaho National Laboratory (1972); Kentucky Geological Survey/Univ. of Kentucky (1982); Arizona Geological Survey (2007); Alberta Geological Survey / Alberta Energy Regulator (2013); Martens and University Of Montana (2017); Colorado Geological Survey (2016); U.S. Geological Survey (2016); Oklahoma Geological Survey (2018); Icelandic Meteorological Office (1992); Instituto Português do Mar e da Atmosfera, I.P. (2006); Kandilli Observatory And Earthquake Research Institute (1971); Bureau of Economic Geology, The University of Texas at Austin (2016); University of Nevada, Reno (1992); Charles University in Prague (Czech) et al. (1973); University of Bergen (1982); Penn State University (2009); Department of Meteorology and Hydrology - National Earthquake Data Center (2016); Geological Survey Department Cyprus (2013); Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (2015); Albuquerque Seismological Laboratory/USGS (2013); Kyrgyz Institute of Seismology, IVTAN/KIS and University of California, San Diego (1991); Alaska Earthquake Center, Univ. of Alaska Fairbanks (1987); Obsercatorio Vulcanológico y Sismológico de Costa Rica, Universidad Nacional (1984); Nanometrics Seismological Instruments (2013); Albuquerque Seismological Laboratory (ASL)/USGS (1980); Technological Educational Institute of Crete (2006); Institute of GeoSciences (IGEO), Polytechnic University of Tirana (PUT) (2002); RESIF (1995); French Landslide Observatory – Seismological Datacenter / RESIF (2006); Universidad de Chile (2012); Albuquerque Seismological Laboratory/USGS (2014); Pacific Tsunami Warning Center (1965); USGS Hawaiian Volcano Observatory (HVO) (1956); Disaster and Emergency Management Authority (1990); Red Sismica Volcan Baru (2000); University of Nevada, Reno (1971); USGS Menlo Park (1966); Central Asian Institute for

Applied Geosciences (2008); KNDC/Institute of Geophysical Research (Kazakhstan) (1994); Centre IRD de Noumea, Nouvelle-Caledonie (2010); IRIS Transportable Array (2003); Northern California Earthquake Data Center (2014); University of Genoa (1967); Oklahoma Geological Survey (1978); Kandilli Observatory And Earthquake Research Institute (1971); Instituto Geografico Nacional, Spain (1999); Kyrgyz Institute of Seismology, KIS (2007); Natural Resources Canada (1975); Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE), Ensenada (1980); Corinth Rift Laboratory Team And RESIF Datacenter (2013); National Observatory of Athens, Institute of Geodynamics, Athens (1975); Regional Integrated Multi-Hazard Early Warning System (RIMES Thailand) (2008); Norsar (1971); Cesca et al. (2009); Institute of Geophysics China Earthquake Administration (IGPCEA) (2000); National Institute for Earth Physics (NIEP Romania) (1994); Aristotle University of Thessaloniki (1981); University of Athens (2008); Instituto Nicaraguense de Estudios Territoriales (INETER) (1975); University of Patras (2000); Geophysical Survey of the National Academy of sciences of Tajikistan (2009); GFZ German Research Centre For Geosciences and Institut Des Sciences De L'Univers-Centre National De La Recherche CNRS-INSU (2006); GEOFON Data Centre (1993); Vernon (1982); California Institute of Technology and United States Geological Survey Pasadena (1926); GNS Science (2021); de Costa Rica (2017).

4

Patterns in source time functions reflect variations in earthquake magnitude and inversion approach

Contents

Key Points

- We compile and analyse more than 5000 source time functions from five datasets.
- We find that large earthquakes release more moment in short initial windows.
- Many source time functions are complex and show several phases of rupture.

Abstract

Source time functions describe the evolution in earthquake moment release through time, and can be found through a range of inversion approaches. It is unclear to what extent source time functions show patterns with earthquake magnitude or depth, particularly in their earliest stages, and whether the inversion approach chosen affects any relationships which may exist. We compile a catalogue of over 5000 source time functions from five catalogues which use different inversion approaches. We find that larger earthquakes release more moment even in short, early, windows of time, and that a significant proportion of STFs are not well fit by a single peak, instead showing several phases of rupture which could be considered as sub-events. Identifying patterns in source time functions has important implications for earthquake source models and ground motion predictions, and for our understanding of earthquake determinism.

Plain Language Summary

Source time functions show how the moment is released throughout an earthquake. We can calculate STFs using different methods that make different assumptions and use different observations. It is not clear if factors like the size or depth of an earthquake, or how an STF was calculated, affect how STFs look, especially in the early stages. To explore this, we looked at over 5000 source time functions calculated in 5 different ways. We found that bigger earthquakes tend to release more moment in the early stages of rupture. We also noticed that many earthquakes have multiple phases of rupture, rather than just one simple peak. Understanding these patterns is important because it can help improve how we model earthquakes and can clarify how different STF calculation methods can influence the research they are used in.

4.1 Introduction

Earthquake determinism asks whether there are intrinsic differences between earthquakes of different magnitudes in their earliest stages. In Chapter 3, we considered whether parameters of the early windows of earthquake waveforms scale with final earthquake magnitude. Here, we approach determinism, and more broadly whether there are differences between earthquakes of different magnitudes at any stage in rupture, using source time functions (STFs). This paper uses multiple datasets of STFs to look at the temporal development of earthquakes, and to look for systematic relationships between STF features, the dataset used and the final magnitude of the earthquake.

STFs (also referred to as moment-rate-functions), describe the evolution in earthquake moment release through time (Figure 4.1 shows some examples of STFs). STFs provide one of the most robust characterisations of the source process. STFs are closely related to the observed seismograms, and require less modelling and interference compared to other approaches used to describe the source process. STFs also better resolve higher frequency components ($f < 1$Hz) compared to slip models (0.1Hz), which allows more detail of the source process to be resolved. If we consider

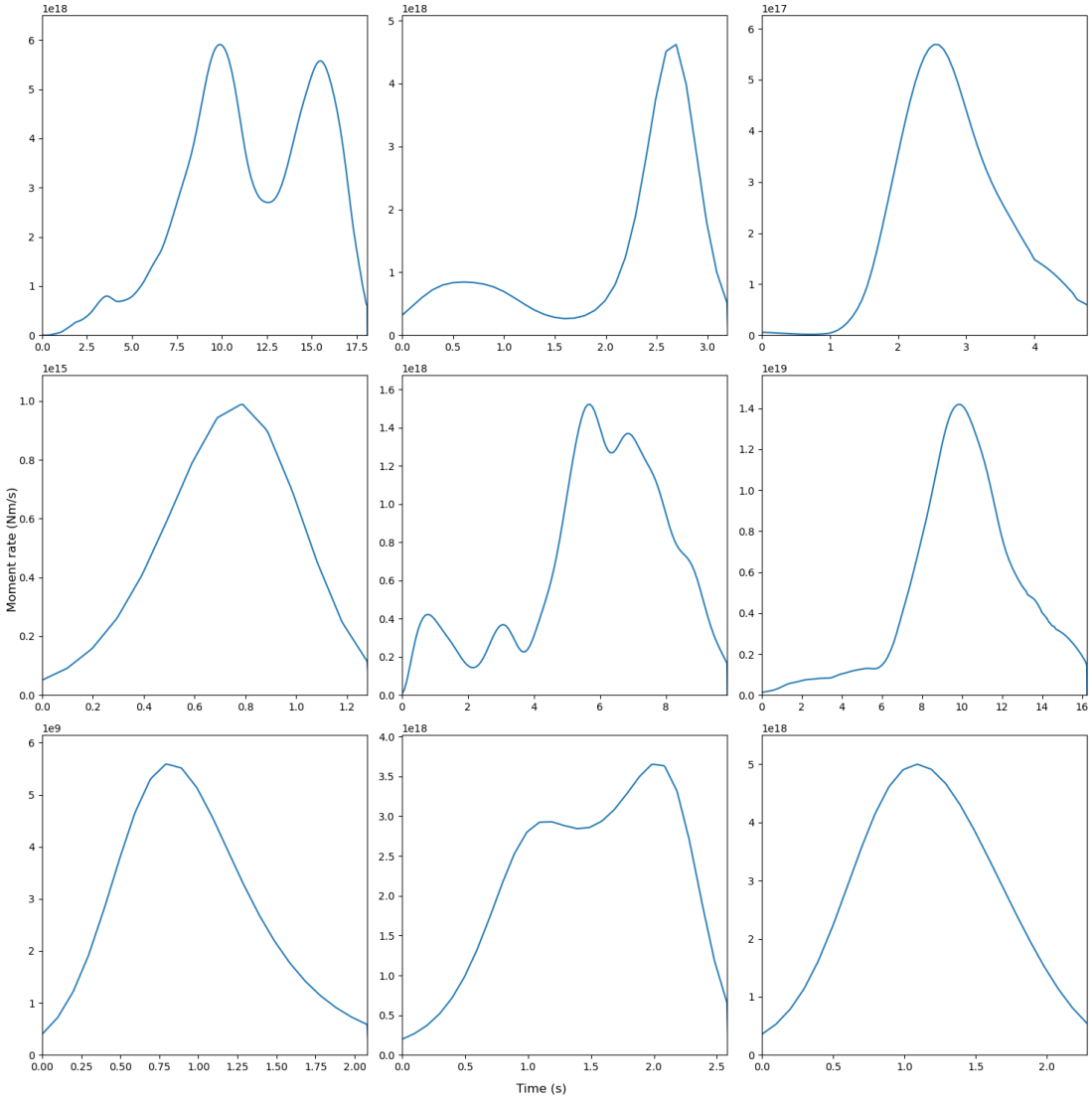


Figure 4.1: Nine randomly selected example STFs from the catalogue we compile (See Section 4.2 and 4.2.6).

the theoretical case of a point-source in an infinite non-attenuating medium, the STF can be found by simply scaling the P or S waveforms. This scaling factor accounts for the radiation pattern, the distance and the elastic properties of the medium (Vallée and Douet, 2016). In reality, the inversion process is more complex and there are a number of approaches to calculate STFs (Section 4.2) which all have different assumptions, and have been applied to different subsets of global seismicity (with different magnitude ranges).

In Chapter 3, we found that there are statistical differences between waveform

parameters for earthquakes of different magnitudes ($M \geq 3$). Peak ground displacement is a waveform metric which is known to scale with maximum moment rate. On average, we find that larger earthquakes have higher values of peak ground displacement, even in windows just 1/3 of the earthquake duration. This leads us to now investigate what information STFs give us about the evolution of earthquakes, both in their earliest stages and throughout the rupture process.

4.1.1 STF variation with magnitude

Perhaps the most ground-breaking work on STFs looked at the average STF for different magnitude bins (Meier et al., 2017), and it is from this work that many of the questions in this paper arose. In their work, Meier et al. (2017) take different magnitudes of interest (M_W 7, 7.25, 7.5,...,8.5) and choose the 20 events (from the Ye et al. (2016a) STF dataset) with closest magnitude. This means that events may occur in more than one group, and that the median magnitude of each bin may not equal the target magnitude particularly for larger target magnitudes (e.g. the median M_W is 7.98 for both the M_W8 and $M_W8.25$ target groups). They then calculate the median STF for each group of events.

Meier et al. (2017) observe that the median STF has a fairly regular behaviour in all magnitude groups, with 'near-triangular and almost symmetric shapes' (Meier et al., 2017, p. 1277 and Figure 2A). However, for larger magnitude bins, it is clear that the peak moment rate occurs before the centroid time. They also calculate the median of the normalised STFs in each group and these also have a high degree of similarity, with peak moment rates occurring between 35% and 55% of the duration. This suggests that $M \geq 7.25$ earthquakes all evolve similarly, and that there are no intrinsic differences between major and great earthquakes. However, Meier et al. (2017) note that the smallest magnitude bin ($M7$) shows a slower initial growth rate than the other bins, which they suggest may result from lower certainty of source inversions in the first 10s. One question we therefore investigate in this work, is whether this similarity is preserved at smaller ($\sim M6$) magnitudes.

Meier et al. (2017) primarily use the Ye et al. (2016a) STF dataset. They do look at the SCARDEC (Vallée et al., 2011) and Hayes (2017) datasets to confirm their results, and do see some differences. The normalised STFs calculated from the Hayes (2017) dataset are not as similar as those found from Ye et al. (2016a) and Vallée et al. (2011) datasets. In particular, the smaller magnitude groups in the Hayes (2017) dataset tend to have slightly later and higher amplitude peaks, and a faster decline in moment-rate after the peak. They also show some evidence of a low-moment-rate tail for the last 20% of the normalised duration. We will use STFs from the three datasets used by Meier et al. (2017) as well as those produced by the Sigloch group (following the method of Sigloch and Nolet, 2006) and the ISC-PPSM STFs. We discuss these datasets further in Section 4.2.

4.1.2 The growth phase of STFs

Previous work using STFs to understand earthquake determinism has attempted to look for a 'growth phase' of the STF. Melgar and Hayes (2019) calculated the average moment-acceleration over the first 1–20 s of STFs, by linearly interpolating between the moment-rate at zero-time and the moment-rate at the time in question. They found that in the first 1–2 seconds, there is only a weak correlation between moment-acceleration and magnitude. However, once 5 s of rupture has passed, they find a clear log-linear relationship between moment-acceleration and magnitude. They propose that this is a representation of a two stage rupture-initiation process in which the rupture initially nucleates and ruptures of different final magnitudes are indistinguishable. At some point, the rupture then organises itself into a slip pulse, and this slip pulse has different properties depending on the final magnitude of the event. This suggests that earthquakes are moderately to strongly deterministic (See Section 1.2.8 in overall intro and Meier et al. (2020b) for discussion of this definition). Meier et al. (2020b) disagree with the findings of Melgar and Hayes (2019), for two main reasons. The first is that the linear interpolation approach to understand the growth of the STF is only valid where the portion of rupture that is of interest is the growth phase, which is unlikely for the longest time window (20 s)

and the smallest magnitude events. The second is that whilst the dataset contains STFs from the SCARDEC, Ye and USGS datasets (see Section 4.2.6), the events chosen from these databases are biased towards unusually long STFs. Meier et al. (2020b) therefore repeat the analysis of Melgar and Hayes (2019). They find that with STFs that have durations close to the expected duration, there is a qualitative scaling between the final magnitude and the interpolated moment-acceleration taken before the peak moment-rate is reached (events of close to typical duration are defined as events with duration of 80–120% of the average duration of the SCARDEC STFs of that magnitude). They repeat this with synthetic STFs based on a simple self-similar crack model (quadratic growth, maximum moment-rate at half-duration), and find that this scaling is no longer evident. They therefore propose that earthquakes are, at most, weakly deterministic but that more work is required to understand the ability of STFs to reliably resolve event onsets.

4.1.3 STF complexity

Alongside trying to understand the early moment-acceleration of different earthquakes, another area of interest is the varying complexity of STFs. Danré et al. (2019) quantified the complexity of SCARDEC STFs and found that larger earthquakes tend to have more complex STFs. To do this, they decompose the STF into discrete Gaussian sub-events, by fitting every major peak in the STF in chronological order, and each peak is a 'sub-event'. In general, they find that $M < 7$ earthquakes have fewer sub-events (mean 3.07) compared to major $M \geq 7$ events (mean 5.84), and find similar trends for the USGS dataset too, suggesting that earthquakes are moderately deterministic. Generally, they find that strike slip earthquakes have a stronger relationship between STF complexity and magnitude, perhaps due to the constraint of the seismogenic width. Generally they found that sub-event moment scales with earthquake moment: larger earthquakes have more and larger detectable sub-events, opposing earthquake self-similarity.

We begin by outlining the different datasets we use and their approaches to inverting for the STFs. We then investigate different approaches to determine

the end of STFs. We can then finally arrive at the *raison d'être* of the chapter: interrogating the STFs to understand patterns in moment release across datasets and magnitude ranges. We frame this investigation around a number of questions:

- Are the median STFs of different magnitude groups similar?
- Are there qualitative patterns in the normalised median STFs with dataset or magnitude?
- How are the early portions of moment released?
- Does the timing of the maximum moment release vary systematically?
- How do statistical measures of STFs vary?
- Are there patterns in STF complexity?
- What geometry best fits a STF?
- How are increments of moment released throughout an earthquake?

Finally, we attempt to tie our observations into a physical framework.

4.2 Different STF databases and their methods

In this work we use 5 different datasets of STFs, which are all calculated differently. We aim to establish which STF characteristics are robust across inversion schemes and thus are most likely to represent physical processes of moderate-to-large magnitude earthquakes. To do this, we must first understand how each dataset of STFs is calculated. Each subsection is primarily based on one or two papers that describe the methods used, and these are cited at the start of each method description.

4.2.1 SCARDEC

This subsection is based on Vallée et al. (2011) and Vallée and Douet (2016).

The SCARDEC method (Seismic source ChAracteristics Retrieved from DEConvolving teleseismic body waves) aims to establish the double couple moment tensors (strike, dip, rake) and depths of earthquakes in order to find the moment magnitude, and uses teleseismic body-wave records. In doing so, it produces apparent source time functions for the recording of an earthquake at each station and also reports an 'average' STF for each event.

4.2.1.1 Data

SCARDEC uses both compressional and transverse teleseismic ($30° < \Delta < 90 - 95°$) waves of $M > 5.8$ events. Five phases are observed and modelled: P, PcP, PP, S, ScS. Hereafter we follow Vallée et al. (2011) in referring to the P, PcP and PP phases as 'compressive waves' and the S and ScS phases as 'transverse waves'. In large earthquakes, at least one of the core- or surface-reflected phases interferes with the direct P or SH wave, making STF production difficult if only direct phases are used. By modelling both direct and core- and surface-reflected waves, and their interference, SCARDEC is able to produce STFs for events where these arrivals cannot be deconvolved, such as large or very long duration events. Modelled waveforms are generated using the method of Bouchon (1976).

4.2.1.2 Inversion method

The SCARDEC method for producing STFs is fundamentally a deconvolutional approach, and makes no assumptions about rupture process. It has four main steps. (1) Source duration is independently constrained from the high frequency waveforms (1–3 Hz) for large earthquakes ($M \geq 7$) and from the GCMT equation for duration ($T_d = 4.5 \times 10^{-6}(M_0)^{1/3}$) for earthquakes $5.8 \leq M < 7$ (2 x half duration given in equation 1 of Ekström et al., 2012, after converting from dyne-cm to Nm). (2) Synthetic double-couple point source signals are deconvolved from the low-pass filtered observed transverse waveforms, whilst varying strike, dip, rake and depth of the source. This gives (many) apparent STFs produced for each high-frequency (non-low-pass-filtered) compressive waveform that has an appropriately clear signal and for each plausible moment tensor-depth pair produced. (3) Search for the optimal parameters which produce STFs which produce physically reasonable STFs. There are three main properties of an acceptable STF at this stage: positivity, causality and maximum duration equal to that calculated in step 1. The requirement for causality arises from the directivity term always being shorter than the rupture time for body waves. (4) Integrate the optimal STF to find the moment magnitude. Steps 2 and 3 are then repeated using both compressive and transverse waves, to

find the final apparent STFs. In doing this, an additional constraint is applied on the STFs, that the moment is close to the median moment calculated in step 4.

SCARDEC uses a neighbourhood algorithm to optimize the strike, dip, rake and depth of the earthquake, and to minimize the misfit between the data and synthetics, and (with lower weighting) to minimize the differences between the apparent STFs at each station. Transverse STFs are expected to vary more than compressive STFs and thus are down-weighted in the misfit function. Acceptable final parameters are those which lead to misfit values within 10% of the optimal misfit.

4.2.1.2.1 The SCARDEC product The SCARDEC method produces two STFs, *moySTF* and *optSTF*. The so-called *moySTF* is the average of all 'good' apparent STFs, and represents the common behaviour. The *optSTF* is the apparent STF that most closely resembles the *moySTF*. How 'good' each STF is determined by the agreement between the observed waveforms and the synthetic waveforms generated from the convolution of the apparent STF and the point source synthetics. Good STFs are also required to be similar to the averaged STF.

SCARDEC routinely produces STFs for $M > 5.8$ earthquakes globally, and releases the resulting STFs on a website.

4.2.1.3 Limitations

Whilst SCARDEC attempts to produce an STF for all earthquakes with $M > 5.8$, there are some cases in which this is not possible. Between 1992 and 2014, it was able to produce STFs suitable for release for about 50% of $M > 5.8$ earthquakes. Two groups of factors prevent the calculation of a reliable STF: data availability and source complexity.

To produce a reliable STF, there must be a sufficient number of stations with high quality data. This can be a challenge with old earthquakes or events which are preceded by a large earthquake (not necessarily nearby), where the data may be low quality or contaminated. Similarly, earthquakes without enough P wave broadband signals cannot have reliable apparent STFs estimated. This is particularly an

issue for moderate-magnitude strike-slip earthquakes, due to the low radiation coefficients (see e.g. Liu et al., 2023).

Complex earthquakes or earthquakes occurring in a complex structure are also difficult to work with. Earthquakes that undergo a significant mechanism change part way through (or even a relatively small mechanism change for strike-slip earthquakes) have periods with very different radiation patterns. Earthquakes under deep-water result in P wave broadband signals which are difficult to deconvolve from point-source synthetics. Even when deep ocean events can be used to produce STFs, they often have obvious water multiples, which we will return to in Section 4.3. It is also difficult to work with events that are anomalously long ($>$ 200 s for $M \geq 7$) and thus these are also excluded. Finally, it is hard to produce reliable STFs for moderate magnitude events ($5.8 \leq M < 7$) which are composed of several sub-events (with a slightly broader definition of sub-event than Danré et al. (2019)). This leads to a potential bias towards more simple moderate events, which we will investigate further in section 4.4.6.

These factors compound to mean that older, moderate strike-slip earthquakes are most likely to be removed from the analysis and thus are likely under-represented in the catalogue.

Another limitation results from the alignment of the STFs. The zero time is fixed by the first significant moment release of the *optSTF*, and thus the *moySTF* may begin at negative times. Indeed, because the STFs going into the calculation of the *moySTF* may not have been perfectly aligned, the early stages of the *moySTF* are therefore likely to be less accurate.

4.2.2 Sigloch

The 'Sigloch' database comprises $\sim$ 4000 STFs generated by the Princeton, Munich and Oxford Seismology Group working under the stewardship of Karin Sigloch (Mohammadzaheri, 2019; Totten, 2021; Kemp, 2025) following the procedure described in Sigloch and Nolet (2006). The STFs are a by-product of work to measure finite frequency amplitude and traveltime anomalies for use in seismic tomography.

4.2.2.1 Data

The Sigloch and Nolet (2006) method uses vertical-component, broadband, teleseismic P wave recordings. Initially the method was applied to earthquakes between $M5.8$ and $M6.9$, although the upper limit has been pushed to $\sim M7.2$. The data have the instrument response removed and are low-pass filtered at 3.5 Hz. Records with epicentral distances between 35° and 80° are used.

Rather than computing a single global STF, data are divided into geographically distinct groups (g) by clustering P waveforms based on their correlation. This also enables the removal of anomalous traces. The seismograms within each group are aligned in time using the cross correlation method of VanDecar and Crosson (1990) allowing a single pick to be made to align a correlated group of waveforms.

4.2.2.2 Inversion method

The Sigloch and Nolet (2006) method is fundamentally a matched filter technique. The best predictions for the observed broad-band wave shapes, $\tilde{\boldsymbol{s}}_k$, are found from

$$\tilde{\boldsymbol{s}}_k = a_k \boldsymbol{G}_k \boldsymbol{f} = a_k \boldsymbol{u}_k, \tag{4.1}$$

where $\boldsymbol{f}$ is the STF, a_k is the amplitude anomaly (a scalar energy correction factor for datum k) and $\boldsymbol{G}_k$ is an $N \times N$ convolution matrix where its rows contain time-shifted and reversed copies of the Green's function $\boldsymbol{g}_k^T$. The Sigloch method uses WKBJ synthetics (Chapman, 1978) which allow for the modelling of direct-P and depth phases.

Often, a_k is taken to be 1, meaning that $\tilde{\boldsymbol{s}}_k = \boldsymbol{u}_k$. This predicts waveforms (S) with shapes that match the observed $\boldsymbol{s}_k$ well, but with energy contents that differ from station to station (Ruff, 1989), since we are more likely to get the amplitudes of the waveforms wrong than the shape. Varying a_k allows for corrections to the amplitudes. Thus, Sigloch and Nolet (2006) invert jointly for $\boldsymbol{f}$ and a_k from the observed seismograms ($\boldsymbol{s}$ with dimensions $k \times 1$) which allows a significant reduction in overall misfit. The minimisation is

$$e = \min \left((\boldsymbol{s} - \tilde{\boldsymbol{s}})^T (\boldsymbol{s} - \tilde{\boldsymbol{s}}) \right), \tag{4.2}$$

where

$$\tilde{\boldsymbol{s}} = \begin{pmatrix} \tilde{\boldsymbol{s}}_1 \\ \tilde{\boldsymbol{s}}_2 \\ \vdots \\ \tilde{\boldsymbol{s}}_k \end{pmatrix} = \begin{pmatrix} a_1\boldsymbol{G}_1 \\ a_2\boldsymbol{G}_2 \\ \vdots \\ a_k\boldsymbol{G}_k \end{pmatrix} \boldsymbol{f} = \boldsymbol{A}\boldsymbol{f}. \tag{4.3}$$

This is a non-standard linear least squares problem because $\boldsymbol{f}$ and a_k multiply each other. Sigloch and Nolet (2006) avoid this additional constraint by iteratively solving equation 4.3 by 'bouncing' between finding $\boldsymbol{f}$ and a_k, which avoids imposing an additional normalisation constraint on either $\boldsymbol{f}$ or a_k (Ruff, 1989). It is also possible to regard the moment tensor as an additional unknown and invert for a correction to the published solution as part of this process. The process is followed separately for each group of seismograms, g.

First, solve equation 4.3 for $\boldsymbol{f}$, assuming that $a_k = 1$ for all k. Minimising e (equation 4.2) is then a standard least squares problem which Sigloch and Nolet (2006) solve using the LSQR algorithm (Paige and Saunders, 1982; Nolet, 1987). This gives the damped solution

$$\boldsymbol{f} = \left(\boldsymbol{A}^T\boldsymbol{A} + \epsilon\boldsymbol{J}\right)^{-1}\boldsymbol{A}^T\boldsymbol{s}, \tag{4.4}$$

where ϵ is the damping parameter and $\boldsymbol{J}$ is an $N \times N$ diagonal matrix where the values increase linearly from 0 to the average diagonal value of $\boldsymbol{A}^T\boldsymbol{A}$. $\epsilon\boldsymbol{J}$ imposes a penalty on later samples to favour short solutions and avoid oscillatory tails.

The estimates for $\boldsymbol{f}$ are now input into equation 4.1, and minimizing e (equation 4.2) again becomes a least squares problem, this time with a single unknown, a_k,

$$a_k = \frac{\boldsymbol{u}_k^T\tilde{\boldsymbol{s}}_k}{\boldsymbol{u}_k^T\boldsymbol{u}_k}. \tag{4.5}$$

This is the same RMS definition of amplitude used by Dahlen and Baig (2002).

The inversion process now iterates, and these estimates are used as a_k in equation 4.3 to invert for the new $\boldsymbol{f}$. Should the inversion also want to find a correction for the published moment tensor, this is done between inverting for $\boldsymbol{f}$ and inverting for a_k.

The process is repeated until the results converge or the STF develops a negative tail whilst the RMS error decreases only minimally.

4.2.2.3 Limitations

The inversion procedure requires $N \leq 256$, and thus the resulting STF can be no more than 25.6 s. This means no events longer than 25.6s can be modelled excluding larger magnitude events from the dataset (since duration scales with magnitude, see Section 3.3 and A.2) and also excludes unusually long smaller magnitude events.

There are no intrinsic motivations for the inversion to produce physically acceptable STFs. Indeed whilst we expect the true STF to be non-negative, since fault rupture is not thought to reverse direction, the inversion maps some source related effects into the STF ($\boldsymbol{f}$), so we expect some negative aspects. This is despite the fact that clustering into geographically distinct groups accounts for some near source effects, azimuthal dependencies and regional travel time anomalies.

STFs are produced for a range of source depths, to reduce reliance on external parameters. Intially the depth associated with the absolute minimum RMS is preferred, although this is evaluated by hand. As well as being as non-negative as possible, the preferred solution is ideally as simple as possible (favouring unimodal and short STFs over bimodal, oscillatory functions if the RMS is similar) and ideally the STFs for different groups look similar. The desire for simple STFs means that the dataset may be biased towards characterising events as a single event, rather than multiple sub-events.

4.2.3 USGS

The dataset which we call 'USGS' is that available through the USGS-NEIC combined slip-model catalogue. It is based on an approach first developed by Hayes (2017) and extended and developed by Goldberg et al. (2022).

The USGS produces STFs as a by product of finite fault modelling, in which they aim to match observed seismograms by modelling fault geometry and the spatio-temporal evolution of fault slip. The USGS dataset initially relied on inversion of body-wave and surface-wave records. This allowed the production of a uniform catalogue of finite fault models (and thus STFs) for most $M \geq 7.5$ shallow and intermediate depth earthquakes since 1990 (Hayes, 2017), and could be

extended down to $M7$. Goldberg et al. (2022) updated the modelling to now also incorporate Global Navigation Satellite System (GNSS) records, Interferometric Synthetic Aperture Radar (InSAR) observations, and local-to-regional accelerometer records. This allows the production of finite fault models down to $M6$. Note that in Danré et al. (2019), the USGS data only contained the original Hayes (2017) teleseismic STFs.

4.2.3.1 Data and weightings

The USGS inversions use a number of data-sources, which are briefly described here, based on Goldberg et al. (2022). The data-sources are not weighted equally in the inversion, and in addition to the weightings detailed below (Table 4.1), individual records can be down-weighted or removed based on signal-to-noise ratio or inconsistency with nearby observations.

Teleseismic body waves P and SH waves are used, bandpass filtered between 1 and 200 s (0.005–1 Hz). Data are chosen to produce an azimuthally balanced dataset and to avoid low signal-to-noise data. Teleseismic P waves are given twice the weighting of teleseismic SH-waves, due to SH waves often having lower SNR and more uncertain arrival times.

Teleseismic surface waves Both Rayleigh and Love wave records are used, bandpass filtered between 200 and 500 s (0.002 and 0.005 Hz). These are given twice the weight of teleseismic P waves, as their Green's functions are more reliable than for body waves (Shao et al., 2011).

Local-to-regional strong motion accelerometers To avoid needing to remove coseismic offsets, these are processed as velocity data rather than displacement data. They are bandpass filtered between 8–50 s (0.02 and 0.125 Hz), and are given the same weight as teleseismic surface waves in the inversion.

GNSS Two types of GNSS data are used: static and high-rate, and both are initially processed to displacement, the vertical and horizontal components of each are weighted differently. Both horizontal GNSS datasets are weighted equally to local-to-regional accelerometer data and teleseismic surface waves. Vertical high-rate data are given one-third of this weighting (see e.g. Bock et al., 2011, for vertical GNSS accuracy discussions) whilst vertical static data are given one-half of the weighting.

InSAR Any available InSAR (ascending or descending) can also be used in the inversion. Processed InSAR can have spurious features in the far-field, and these are removed as part of the inversion. InSAR observations are again given the same weighting as teleseismic surface waves.

Table 4.1: Weighting of different data sources in USGS inversion.

Method	Weighting
Teleseismic P waves	2
Teleseismic SH-waves	1
Teleseismic surface waves	4
Local strong motion accelerometers	4
Horizontal static GNSS	4
Vertical static GNSS	2
Horizontal high-rate GNSS	4
Vertical high-rate GNSS	4/3
InSAR	4

4.2.3.2 Inversion method

The inversion method uses a non-linear inversion in the wavelet domain, based on that described by Ji et al. (2002). This first divides the observed waveform into a sum of wavelets, which allows for both frequency and time domain information to be included in the inversion. The minimization problem then attempts to minimize the difference between the observed and synthetic waveforms and wavelets, with a higher weighting given to higher frequencies (since low and high frequency wavelets have different amplitudes, and it is the high frequencies which highlight the differences in rise-time).

To produce the synthetic waveforms, the approach of Hayes (2017) and Goldberg et al. (2022) models slip amplitude, rake, rupture time and rise time on a discretised fault plane made up of smaller sub-faults. The heat-bath simulated annealing method (Rothman, 1986) is used to randomly perturb the sub-fault parameters until the global minimum of the misfit is found.

Initial Parameters The code initially takes a moment-tensor defining the two nodal planes and the centroid location. Both nodal planes are tested. In most cases, a preferred solution is found from one of the nodal planes, but occasionally both are presented. The centroid location identifies the centre of the modelled fault plane, following Koch et al. (2019). The modelled fault plane defines the area over which slip can occur, as opposed to being an estimate of the ruptured area, which is initially estimated from the scaling relationships of Blaser et al. (2010). The modelled fault plane is initially divided into 225 sub-faults of equal size, though this can be varied during the inversion to allow better resolution of slip asperities. Each sub-fault's STF is modelled as an asymmetric cosine function. The rupture velocity can be constant or variable, and a range of velocities (and velocity patterns) are modelled.

4.2.3.3 Comparisons and Limitations

Goldberg et al. (2022) compare their resulting STFs (which use all of the outlined datasets) to those which are calculated only using teleseismic records (following Hayes (2017)), and find that the STFs are relatively similar. This is good news, since the other datasets we consider in this paper use only seismological data. The most difference arises when the teleseismic data is poor. Indeed it is possible for joint geodetic-seismological inversions to produce STFs for events where other datasets cannot. As discussed in 4.2.1 it is difficult to produce an STF for earthquakes where there is another large event in the hours preceeding, due to the contamination of the teleseismic data. The introduction of geodetic and local-to-regional seismic data allows the slip of these events to be modelled.

Whilst the modelling allows for significant deviation of the modelled parameters from their initial values, some bias is always possible particularly when fault

geometry is over-simplified. Similarly, whilst objective weightings of the different datasets are used in the inversion, the quality of the inversion depends on how well these weightings work for a particular earthquake. Given the varied instruments used worldwide, and the differences in instrument spacing, it is possible that the weightings used in Goldberg et al. (2022) inadvertently introduce more uncertainty for earthquakes in specific regions.

In May 2021, the USGS switched from reporting STFs in dyne-cm to Nm. Two earlier events are also posted in Nm (2004 Sumatra and 1995 Gisborne, New Zealand).

4.2.4 Ye et al. (2022)

Ye et al. (2016a) developed a dataset of 114 STFs to analyse source parameter scaling relationships (Ye et al., 2016a) and depth dependence of rupture characteristics (Ye et al., 2016b) across a uniformly inverted, global dataset.

4.2.4.1 Data

The Ye et al. (2016a) approach uses primarily teleseismic P waves, with occasional use of teleseismic SH wave data. The data are bandpass filtered between 0.005 and 0.9 Hz.

4.2.4.2 Inversion method

The inversion uses a least squares kinematic inversion (Kikuchi and Kanamori, 1991), which is similar to that of Hartzell and Heaton (1983). The method of Kikuchi and Kanamori (1991) considers the seismic source to be a series of point sources. By minimizing the difference between the observed waveforms and those modelled from a sum of point sources, the solution in the space-time domain is found for all sub-events (i.e. what the slip looks like for each location and each time). The focal mechanism for each source is determined as part of the inversion. Ye et al. (2016a) constrain the STF and slip model to be positive and for the rupture front to expand at a constant velocity. They model the fault as a planar fault divided into sub-faults and the STF of each sub-fault is parameterised by several overlapping triangles.

The total moment is found to minimize the difference between the inverted total seismic moment, and the GCMT moment estimate which is calculated from long period seismic waves (40–350 s). This moment constraint stabilises the inversion and allows for the use of sub-faults with smaller grid spacing.

Unless modelling results present a case for deviation, the hypocentral depth is initially based on USGS-NEIC estimates and the fault geometry is based on the best GCMT double-couple solution. Where rupture velocity estimates are available, these are used, and if not modelling is conducted with a range of velocities ($V_r = 2.0, 2.5, 3.0$ km s^{-1}), although this makes little difference to the final STF.

The final STF is determined from combining the sub-fault STFs, and is fault perpendicular.

4.2.4.3 Limitations

The constraints on the inverted moment generally lead to a final inverted moment within 20% of the GCMT moment. For earthquakes $M < 7.4$, the inverted moment is systematically higher than GCMT. This trend likely results from the different source regional crustal models used: GCMT uses PREM (Dziewonski and Anderson, 1981) whilst Ye et al. (2016a) uses Crust 2.0 (Bassin et al., 2000) and 1-D layered Green's functions found using the propagator matrix method (Bouchon, 1981). The moment estimate for these earthquakes is also influenced by the more compact source than for larger earthquakes, which means that variations in source centroid depth may be more important and due to low moment regions being more impacted by the positivity constraint. In our calculations we therefore avoid using magnitudes calculated directly from the inverted moment, instead using either the GCMT magnitude or the absolute value of moment released (and considering portions of moment release relative to this).

Ye et al. (2016a) note that the STFs produced have a bias towards being longer in duration, thanks to the positivity constraint. Some STFs also have long, low-amplitude tails due to inaccuracies in modelling the coda which arises from water multiples and late scattered waves.

The Ye et al. (2016a) dataset produces STFs for 114 of the 390 total $M \geq 7$ events 1990-01-01 to 2015-07-31. There are several reasons for earthquakes to be removed which are predominantly data based. Events with limited azimuthal coverage of data are unable to be modelled, as are earthquakes where the teleseismic recordings have low signal-to-noise, primarily due to being preceeded by a large event. Events which have unusually strong water and sediment multiples are also removed, where the authors were 'unable to model [these] satisfactorily'.

The STFs are more likely to be a robust representation of the true STF for events on relatively simple faults, due to the planar fault constraint and thus we might expect a bias away from earthquakes in more complex structures. Similarly, more complex ruptures are likely not as well modelled by a constant rupture velocity, and so we might expect to see fewer clear sub-events, and a trend towards simpler STFs.

4.2.5 ISC-PPSM

The International Seismological Centre Probabilistic Point Source Model (ISC-PPSM) approach solves for earthquake moment tensor, STF and source depth for global Mw 5.8 – 7.2 earthquakes. The inversion methodology solves for the earthquake point source, and its uncertainty, using a Bayesian direct search approach. As the method solves directly for the STF, this reduces the trade-off between STF length and earthquake depth. For ease, we refer to this method as the 'ISC' method here-after.

4.2.5.1 Data

The method uses teleseismic broadband P and SH observations.

4.2.5.2 Inversion Method

The ISC inversion describes the earthquake as a point source, which is fully described by 18 parameters. The parameters describe the depth (1), the moment tensor (5) and the discretised STF (12). The earthquake is constrained to have a symmetrical moment tensor and is parametrised as described in Tape and Tape (2012) and Tape and Tape (2015). This reduces the number of parameters required to describe

the symmetric moment tensor from six to five. (The requirement for symmetry means that the earthquake exerts no net torque (Julian et al., 1998)). The STF is discretised either as the sum of 12 harmonic basis functions (following Stähler and Sigloch, 2014) or into a sum of 4 Gaussians and each Gaussian can vary in amplitude, width and relative timing. Therefore 12 or 10 parameters are required to describe the STF respectively. The basis functions of Stähler and Sigloch (2014) are found from the 1000 of the most confident Sigloch STFs. The zeroth wavelet (s_0) is found as the mean of these 1000 STFs at each time. This was then subtracted and principal component analysis used to find the function s_1 orthonormal to s_0 that explains as much of the variance as possible. The optimally weighted s_1 is then subtracted from each STF and s_2 is found such that it is orthonormal to s_0 and s_1, and can explain as much of the remaining variance as possible. This is repeated until there are as many components as time-points (256). The top twelve form the basis used to define new STFs, which results in a root-mean-square misfit in total signal variance of less than 10% (Stähler and Sigloch, 2014).

The 16 or 18-dimension parameter space is explored using the neighbourhood algorithm adaptive grid search (Sambridge, 1999) to find the parameters which most closely reproduce the observations. The resulting sample set is then assessed using an MCMC approach (Goodman and Weare, 2010; Foreman-Mackey et al., 2013). A priori constraints are imposed that ensure the earthquake source model is physically realistic and encourages a double-couple moment-tensor. This means that any non-double couple solutions are likely important: the waveforms may be very noisy, the 1-D velocity structure may be inadequate or there may truly be an implosive or explosive component.

The posterior likelihood is based on correlation between observed broadband P and SH waveforms and synthetic seismograms. The synthetics are calculated using WKBJ methodology for P waves and Instaseis for both low-pass P and SH waves (Chapman, 1978; Van Driel et al., 2015).

The Wentzel–Kramers–Brillouin-Jeffreys (WKBJ) method for solving non-linear second-order differential equations allows the simulation of synthetic P waves

(Chapman, 1978). If we have an equation

$$\frac{d^2\phi}{dx^2} + \omega^2 s^2 \phi = 0, \tag{4.6}$$

where $s = s(x)$ is a complex function such that s^2 is monotonically increasing with x, and where ω is large and positive, we can begin by assuming that the solution is of the form $\phi = e^{i\omega\tau(x)}$, where $\tau(x)$ is an unknown complex function. By substituting this into equation 4.6, this equation can be solved through successive approximation. The seismological case is certainly more complicated to look at, since the analogous equation to equation 4.6 arises from the substitution of Legrendre polynomials into the wave equation, but a similar principle is applied (Box 9.6 Aki and Richards, 2009). This method allows us to propagate turning waves.

Meanwhile, Instaseis calculates broadband seismograms and allows for rapid simulation of seismograms thanks to the production of Green's function databases using AxiSEM and the reciprocity of the Green's functions. The Green's functions are generated at high-frequencies and are 4th order accurate in space and the simulations can be conducted with a 2D velocity model (Van Driel et al., 2015)

The inversion returns an ensemble of potential earthquake point source models, from which the average of the likely solutions and a measure of the reliability of the result are derived.

4.2.5.3 Limitations

Like with the Sigloch dataset, the ISC STFs are limited to durations less than 25.6 s. The ISC STFs found using the Stähler and Sigloch (2014) method inherit some of the assumptions and tendencies of the Sigloch method, since Sigloch STFs are used to find the basis functions. STFs found using this method tend to be very unstable.

The approach jointly inverts for the earthquake depth and STF, as previous work has shown that depth trades off with STF complexity (Houston et al., 1998; Houston, 2001; Persh and Houston, 2004; Tocheport et al., 2007). The ISC method is designed to add depth resolution by addressing this trade-off directly in a Bayesian inversion. The depth resolution comes from the depth phases in the waveforms.

This trade-off is mainly a feature of relatively shallow (< 40 km) earthquakes. We return to this in Section 4.4.6.4.

4.2.6 Dataset statistics

Across all these datasets, we have 5123 events from around the world which have one or more usable STFs. Most of our STFs come from the SCARDEC catalogue, with 4083 events having STFs in this dataset. Also providing a significant number of STFs is the Sigloch catalogue with 2565 STFs. The USGS (n=246), ISC ($n = 147$) and Ye ($n = 109$) catalogues do not provide as many STFs. Many events occur in multiple catalogues. Figure 4.2 shows how many events occur in different combinations of STF databases. The databases have different magnitude ranges (Figure 4.3), which contributes to the different numbers of events they provide STFs for (Table 4.2).

Table 4.2: Number of events and individual STFs in different magnitude bins.

Magnitude	Events	STFs
5.00–5.25	5	5
5.25–5.50	33	33
5.50–5.75	115	204
5.75–6.00	1506	2977
6.00–6.25	1242	3029
6.25–6.50	837	2129
6.50–6.75	479	1227
6.75–7.00	309	824
7.00–7.25	158	434
7.25–7.50	87	242
7.50–7.75	57	175
7.75–8.00	35	111
8.00–8.25	15	48
8.25–8.50	7	24
8.50–8.75	1	3
8.75–9.00	1	3
9.00–9.25	1	3

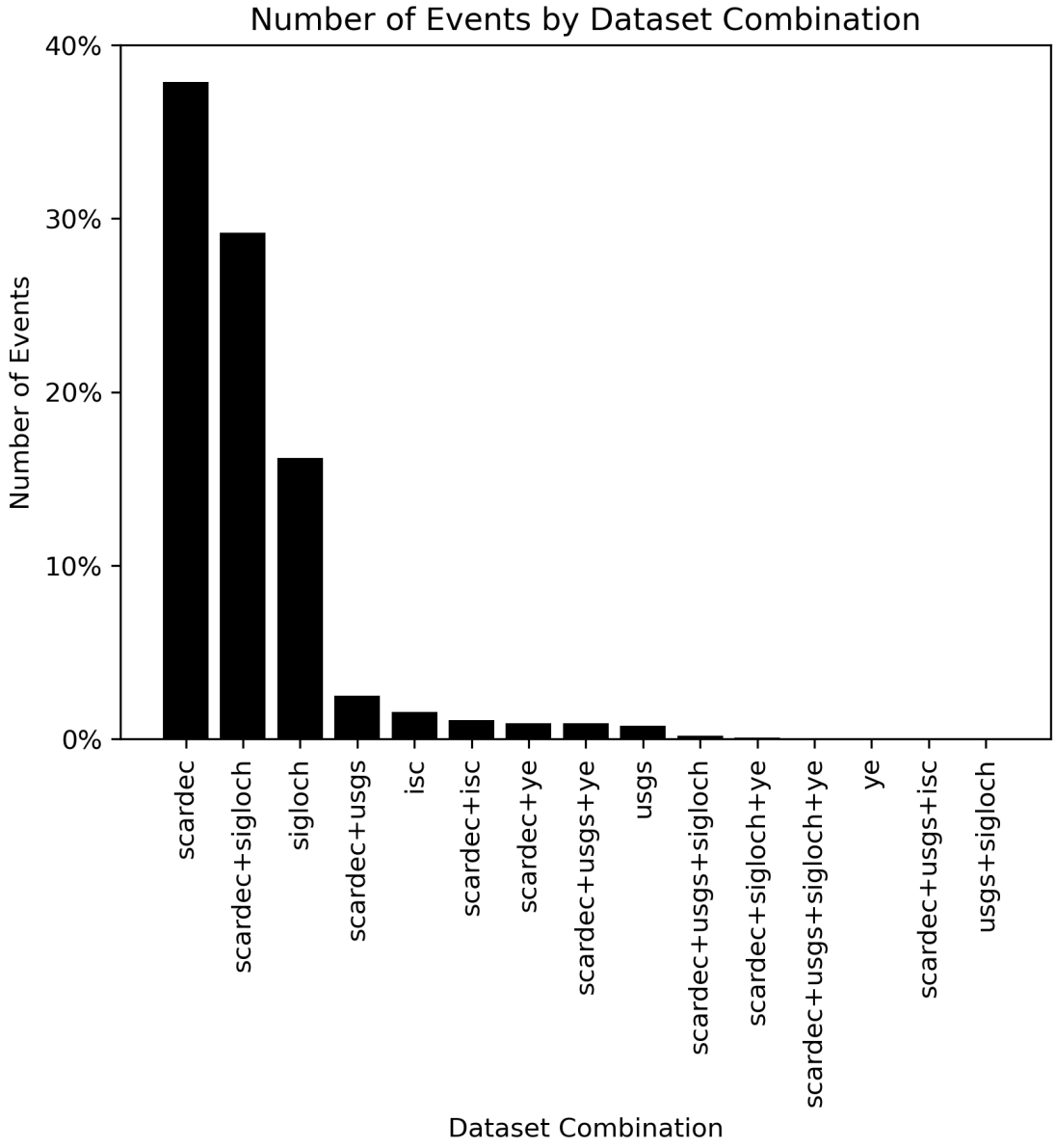


Figure 4.2: Proportion of events occurring in different combinations of STF catalogue.

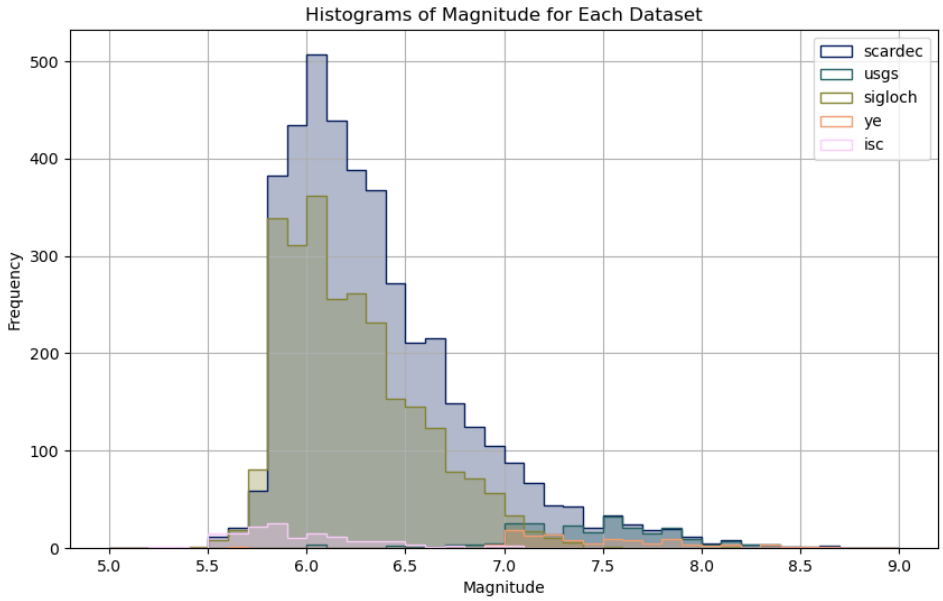


Figure 4.3: Magnitude distribution of each dataset.

4.2.7 Summary of advantages and disadvantages of each method

The different methods to invert for STF have different advantages and disadvantages, which might lead us to have different levels of trust in the various analyses we undertake.

Some methods invert for source depth as part of the STF inversion (ISC, Sigloch) or allow for some variation of the source depth away from published (e.g GCMT) estimates (Ye). These might lead us to have more trust in any observed patterns with depth from these methods.

Some methods use a rupture velocity as part of the inversion. The USGS method allows for this to vary in time in contrast to the Ye method which only considers constant rupture velocities. Therefore we might trust the timing of features more in the USGS results. Other methods require the assumption of the local velocity structure (ISC, Sigloch, SCARDEC). By clustering into geographic groups, the Sigloch method allows for regional travel time anomalies which may just be averaged out in the other methods. It also potentially accounts for or adds first order constraints on directivity. This also may make the timings more reliable in the Sigloch results. The Sigloch method also explicitly allows for source directivity.

Different methods have different reliance on the point source assumption. Both Ye and USGS model the fault as a series of small subfaults, but the Ye method imposes a relatively simple and planar fault geometry, which may reduce overall STF complexity. SCARDEC defines 'good' STFs as those that when they are convolved with point source synthetics, give a result which is close to the observed waveforms. This may serve to reduce STF complexity, as more complex STFs may not match the observations as closely when convolved with point-source synthetics compared to if they were convolved with more complex synthetics.

Complexity can also be impacted by other aspects of the inversion process. The Sigloch STFs are preferred to be simple and smooth and thus likely under-represent true source complexity. On the other hand, ISC STFs are composed of the sum of four Gaussians and thus may add unecessary complexity to simple events. There is

no obvious impact of the SCARDEC method on STF complexity, since the method explicitly allows for several episodes of moment release, but there is a bias against complex, moderate magnitude events being included in the dataset.

Finally, different methods take different approaches to producing the 'final STF'. Some (average SCARDEC, ISC) require averaging either across stations or the ensemble, which could impact the STF. In particular this likely impacts early times of the average SCARDEC STF, where the STFs were not aligned perfectly in time. Meanwhile the Sigloch dataset simply produces STFs for each of the geographic groups it considers, without averaging. This means that the STFs are likely a truer representation of the fault processes as viewed from that region, but may not represent the overall earthquake evolution. At the extreme of this, the optimal SCARDEC STF is the best fitting STF from a single station. This makes the early portion of the STF more reliable but the STF may not be representative of the whole rupture.

We will return to some of these advantages and disadvantages in the discussion, where we will reflect on the reliability of our answers to the questions we posed in the introduction (Section 4.1).

4.3 Can we automatically define the end of an STF?

Now that we have defined datasets, we are nearly ready to probe the patterns in moment release across STF datasets and magnitude ranges. But in order to understand STF characteristics, we must first determine when they end. This will allow us to divide the STFs into different temporal segments and to look at features of the beginning, middle and end of the STFs.

Given a theoretical STF, we might define the start of the STF to be the first point where the $\dot{M}$ ceases to be 0, and similarly define the end as the time at which $\dot{M}$ returns to 0. Unfortunately, with real data it is not this simple.

As discussed previously, STFs can have artificially long tails which for a whole range of reasons resulting from both the data (e.g. complex reverberations due

to sediment and water) and due to the inversion process (e.g. due to positivity constraints, failure to fully remove near source effects). These tails are often low amplitude, but sometimes have a higher amplitude peak, or peaks (see Figure 4.4 and Figures B.1–B.10 for example STFs). Where there are multiple peaks, these are noticeably evenly spaced in time. These are water (or sediment) multiples, and we want to remove these as they do not reflect source processes. These factors mean that durations which rely on $\dot{M}$ reaching zero will tend to be longer than the true STF duration, and we do not expect this variation to be necessarily uniform or to scale with magnitude. Therefore we want to explore how else to determine the end of the STF. Yin et al. (2021) classify approaches to determining duration into three groups: moment-based (Houston, 2001), threshold-based (Vallée, 2013; Denolle, 2019) and centroid-based (Meier et al., 2017).

To negotiate the difficulty in ascertaining the end of an STF, Meier et al. (2017) choose to normalise by twice the centroid time, where the centroid time is the time of maximum moment release, given by

$$t_{\text{centroid}} = \frac{\int_0^\infty t\dot{m}dt}{\int_0^\infty \dot{m}dt}. \tag{4.7}$$

As Meier et al. (2017) note, the centroid times are much less affected by poorly constrained tails. The centroid time can also be considered as the first statistical moment, or mean, of the distribution. However, whilst this potentially works well for smooth STF determinations, for very skewed or complex distributions, this may not be the best or most accurate metric. There is also a difference between defining a duration for normalisation (as Meier et al. (2017) do), and defining a duration to cut off spurious tails and multiples. It is in the latter that we are most interested, and thus we must look to other approaches to define the 'end' of the STF.

We define three potential threshold-based 'ends' and compare these to our manual interpretation of the STF:

1. moment-rate reaches 0,
2. moment-rate goes below 0.01 times the maximum moment-rate,

3. moment-rate goes below 0.1 times the maximum moment-rate, and over 50% of the total area (moment) has been released.

In Figure 4.4, we show the STFs generated from a single earthquake, and the automatically interpreted ends. Figures B.1–B.4 show these for a number of other events. In the discussions that follow, we refer to the 'interpreted moment' to be the moment found from the integral of the STF, whilst the 'total interpreted moment' is that found from the STF without any trimming. The 'true moment' is that determined from other methods and scales with M_W. Most methods constrain the STF to have a 'total interpreted moment' which is close to the 'true moment' (see Section 4.2 for more detail on each method's approach). For events where the end interpreted through all three approaches are the same, we expect the interpreted moment and total interpreted moment to be the same and this should ideally equal the true moment. For events with significant multiples, the interpreted moment will be much less than the total interpreted moment, and the STF may need rescaling such that the interpreted moment equals the true moment. For the ISC and Sigloch data, we set any negative moment-rates to zero before integrating.

Interpreting the end through approach (1) has two main flaws. One which occurs with all datasets is that any spurious tails and water multiples are included, which makes the calculated duration longer than the true duration, as we expected. The other is that the Sigloch data never reaches zero. Therefore the end interpreted by (1) is always at the maximum time of 25.6 s (note this often occurs at times greater than the maximum limit of the x-axis of the figures). This is obviously incorrect. The Sigloch data suffers similarly when determining the start of the STF. Unlike all the other datasets, where we can define the start as when the moment-rate deviates from zero, for Sigloch we must define a threshold of when the moment-rate exceeds 5% of the maximum moment-rate, as all values in the STF file are greater than zero.

If we use a slightly more lenient threshold (2), the durations become slightly more sensible. The worst offenders of the long, low moment-rate tails are now removed, and some of the water multiples are removed. However, placing the threshold at 1% of the maximum moment-rate means some events with multiples

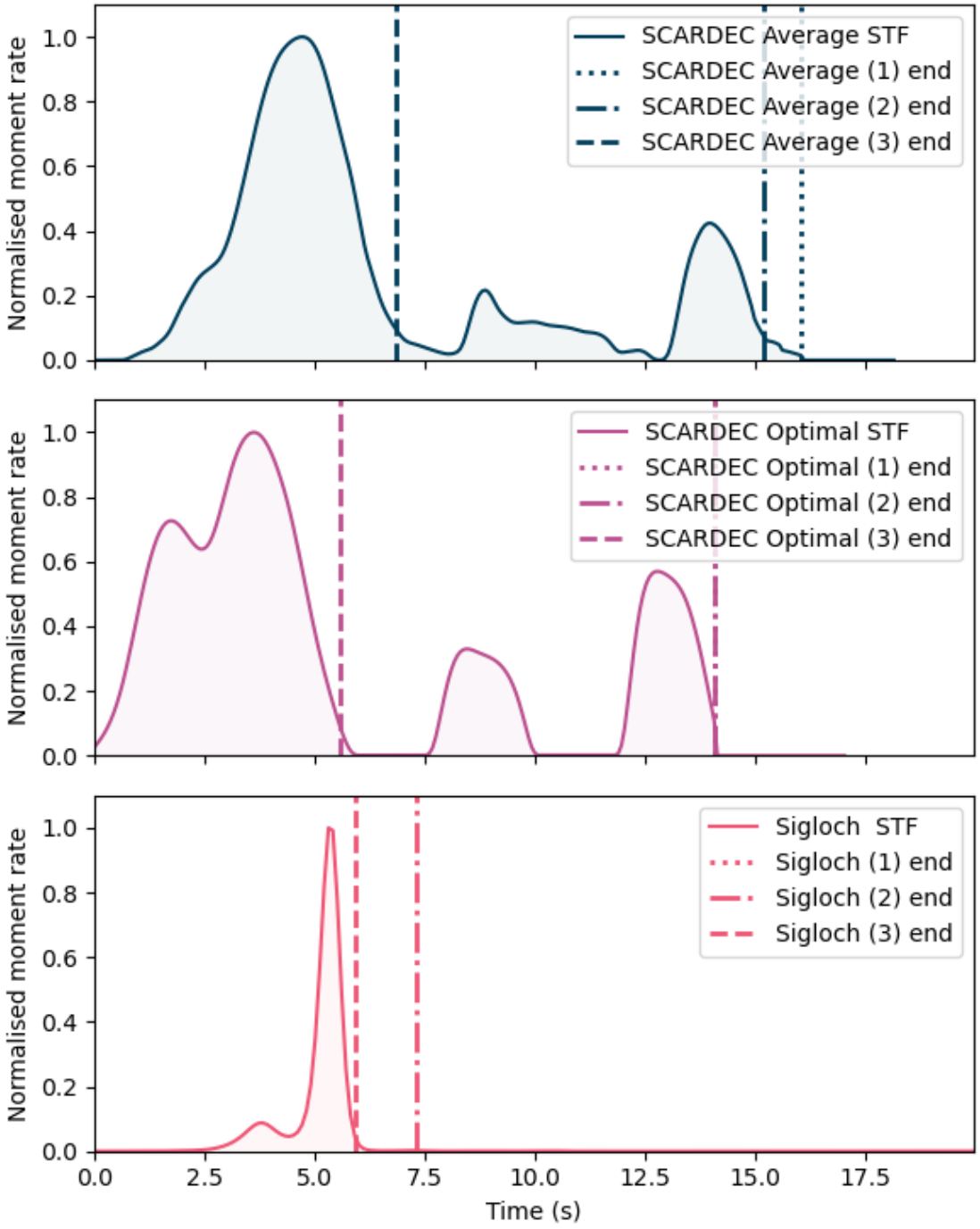


Figure 4.4: Normalised STFs of the 2010-02-05 southeast Indian Ridge M6.2 earthquake. The SCARDEC STFs show clear multiples which we do not want to be included in our duration estimates. The end from approach 1 is dotted, approach 2 in a dot-dash line and approach 3 in a dashed line. Approaches 2 and 3 are always shown on the figures, but may appear hidden by another line (line 1 is hidden by line 2 in the second panel, but is off the limit of the axis in the third panel).

are not appropriately cut, as the moment-rate does not drop this low before the multiple 'arrives'. Indeed, Denolle (2019) define the end as the last time that the moment rate dips below 10% of the maximum and even this includes multiples.

Our final approach is based on the USGS interpreted ends (United States Geological Survey, 2025), and discussion with Gavin Hayes (personal communication, 2024). In this, we wait for the moment-rate to drop below 10% of the peak moment-rate. However, with this condition alone, events which start slowly, or make a bit of a 'false start' before getting going are interpreted to have very short durations. Thus, we add an additional constraint that over 50% of the area under the curve must have occurred before the interpreted end. This threshold was determined based on the observation that even events with several, high-amplitude, multiples have less than 50% of their area contained in these.

We find that 82% of events have an 'interpreted moment' which is at least 90% of the 'total interpreted moment'. Figure B.5 shows that events have on average released 96% percent (median; 92% mean) of their total moment (found from the area) by the interpreted end time, and that there is a weak correlation ($\alpha = 0.05$) between the ratio of the interpreted and total interpreted moment and magnitude ($\frac{\text{interpreted moment}}{\text{total interpreted moment}} = 0.011M + 0.852$, Spearman R: 0.119, p-value: 1.73×10^{-32}). That we tend to 'cut-off' more of smaller earthquakes is not unexpected; for smaller, shorter earthquakes the multiples are more likely to occur after the true end of the STF, and thus we can cut these off. For larger, longer earthquakes, multiples of the first P wave arrivals may arrive before or during the S wave arrival and coda, and so are hidden inside the STF. We note that durations determined with this approach are often *slightly* shorter than those we would manually interpret, but are much closer than those we find from approaches 1 or 2. The events which are cut with a very high proportion of their area remaining tend to be highly complex, and so it is difficult to manually interpret where the STF ends and any multiples or artefacts begin (Figures B.6–B.8). On balance, we therefore continue with approach 3.

4.4 Interrogating our STFs

We now begin to address the questions we posed in Section 4.1.

In Section 4.4.1 we investigate whether the median STFs of different magnitude groups are similar, as a test of earthquake self-similarity with varying magnitude. In Section 4.4.2 we normalise the STFs to look for patterns in the STFs independent of maximum moment release or duration, to further test self-similarity and the role of dataset on STF shape. In Section 4.4.3 we ask whether the early portions of moment release vary with magnitude to understand the extent to which STFs can provide insights into earthquake determinism. In Section 4.4.4 we look for patterns in the timing of maximum moment release with earthquake magnitude and depth, to understand the suitability of centroid time as a measure of the STF. We continue this line of thinking in Section 4.4.5 where we calculate the skew and kurtosis of the STFs with magnitude and depth. In Section 4.4.6 we quantify the complexity and number of peaks in STFs and look for patterns with magnitude and dataset. This gives insight into the suitability of simple Gaussian STFs for rupture modelling, and into the robustness of STF complexity as a measure of earthquake complexity. We follow this by fitting geometric models to the STFs in Section 4.4.7 to define simple stages in the STF evolution. Finally, in Section 4.4.8 we quantify these fits by analysing how quickly different increments of the total moment are released.

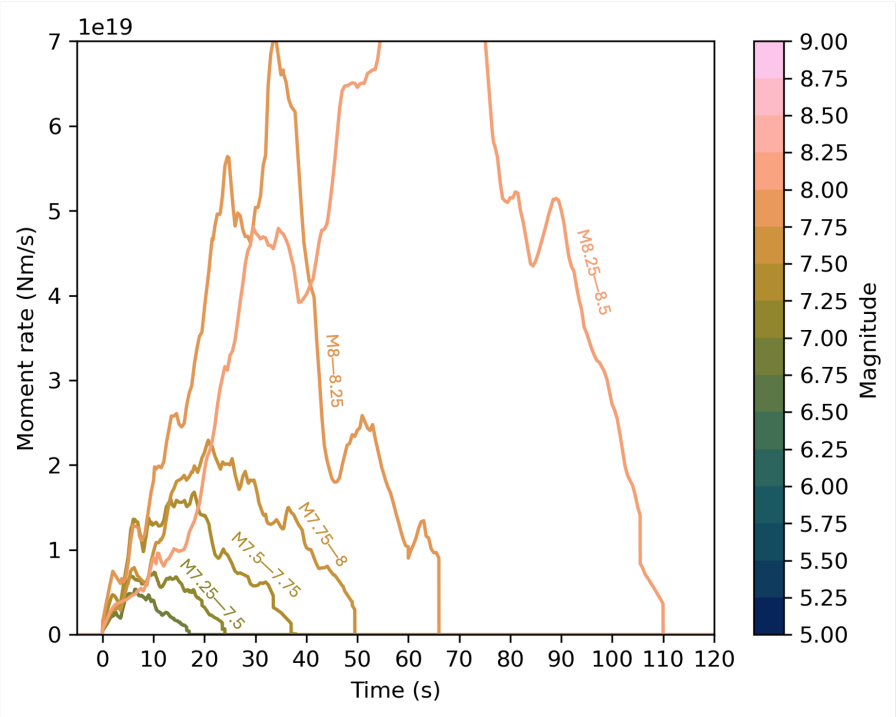
4.4.1 Are the median STFs of different magnitude groups similar?

Meier et al. (2017) observe that the average STFs all have a similar with 'near-triangular and almost symmetric shapes' (p. 1277 and Figure 2A). This is based solely on the $M \geq 7$ Ye dataset, and so we would like to know if the same is true for 1) other datasets in the same magnitude range and 2) smaller magnitude earthquakes. In Figure 4.5a, we show a modified version of Figure 2A in Meier et al. (2017), where events are binned into M0.25 bins rather than looking for the 25 events closest in magnitude to different target values. This means that each event is included in just one group. The effect of this change in approach is most

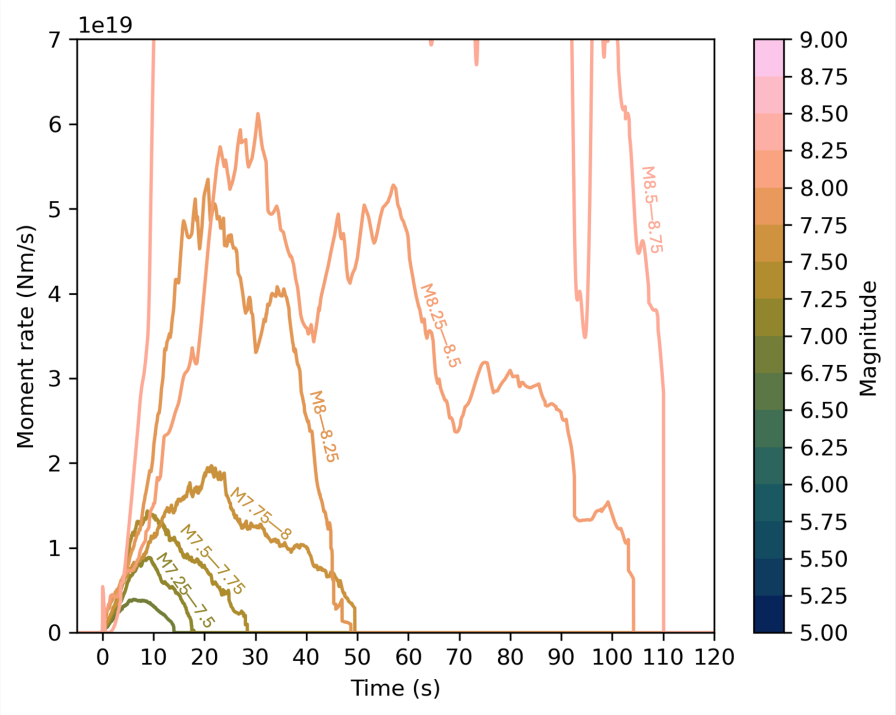
obvious for $M \geq 8$ events, which have much higher amplitudes than the Meier et al. (2017) figure would suggest. We apply our definition of the end to the STFs plotted, and this produces the more abrupt cut-offs than shown in Meier et al. (2017).

Just like Meier et al. (2017), we see a general trend of relatively constant early moment-acceleration, both using solely the Ye dataset (Figure 4.5a) and using all datasets (Figure 4.5b). There are a few deviations from this behaviour. Firstly, the M8.25–8.5 bin has a much lower early moment-acceleration than the other bins. The low number of events in Figure 4.5a makes this difficult to interpret, but in Figure 4.5b, the M8.25–8.5 bin now has $n = 24$ but still shows a lower moment acceleration than the other groups, suggesting this may be a robust feature. We also note an unusual feature where the M7 to M8.25 bins all have similar shapes except for the M7.5–7.75 bin which has a much later peak, and a slower moment-acceleration rate. Both of these features persist when we use all 5 datasets, however, one feature does change: the duration of the median STF for the M8–8.25 bin is much shorter when using all 5 datasets compared to using just the Ye dataset, and indeed is the same duration as the M7.75–8 bin, which does not change when changing the underlying data.

Given the vastly different moment-rates and durations involved in M5 vs M9 ruptures, it is difficult to include all magnitude bins on one plot. Therefore in Figure 4.6 we plot just the M5–7.25 events. Generally, the median STFs still have a similar shape to those of the $M \geq 7$ events, but are smoother and simpler. There is also clearly more variation in the early moment-acceleration, with larger events accelerating much faster. $M < 6.5$ groups also generally show a kink in the median STF, where moment-rate seems to plateau for a fraction of a second before accelerating once again. We return to this kink in Section 4.5. In Figure 4.7, we show the first 3 seconds of the median STFs for all magnitude bins, and we can see that smaller events have lower moment accelerations in their initial phases. The trend in the early moment acceleration is likely obscured in Figure 4.6a due to the long duration of the STFs relative to the 3 s window in Figure 4.7.



(a) Ye et al. (2016a) dataset.



(b) All 5 datasets.

Figure 4.5: Median STF for $M \geq 7$ events calculated in M0.25 bins. Panel a is comparable to Figure 2A of Meier et al. (2017). See Figure 4.7 for a zoom of the first 3 seconds of panel b.

Summary

- Median STFs have broadly similar shapes (until $M \geq 8.25$ where n is small).
- STFs of smaller earthquakes start more slowly.

4.4.2 Are there qualitative patterns in the normalised median STFs with dataset or magnitude?

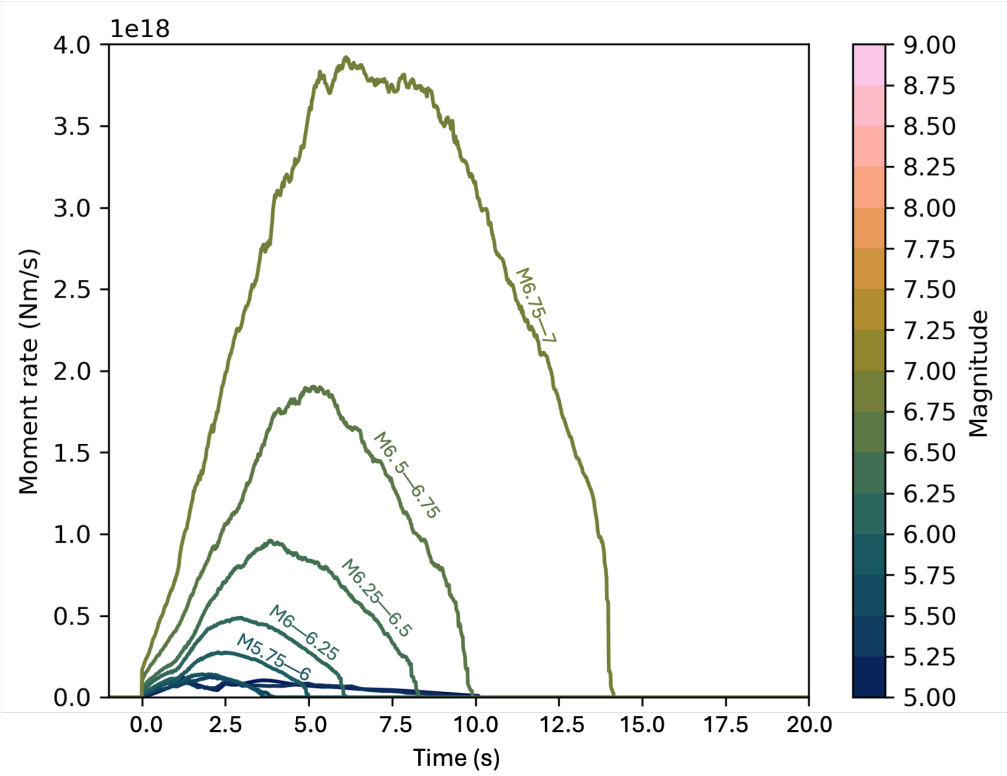
Manually examining the shapes of the $\sim$ 12000 STFs from the 5 datasets, we see significant variation, including features like the number, amplitude and timing of peaks. However, there are too many events to individually look at. A common approach is to average STFs in different magnitude bins, to allow us to compare behaviour between earthquakes of different magnitudes. In this work, we also want to look at variations across datasets, and this is where we begin.

4.4.2.1 Average STF for each dataset

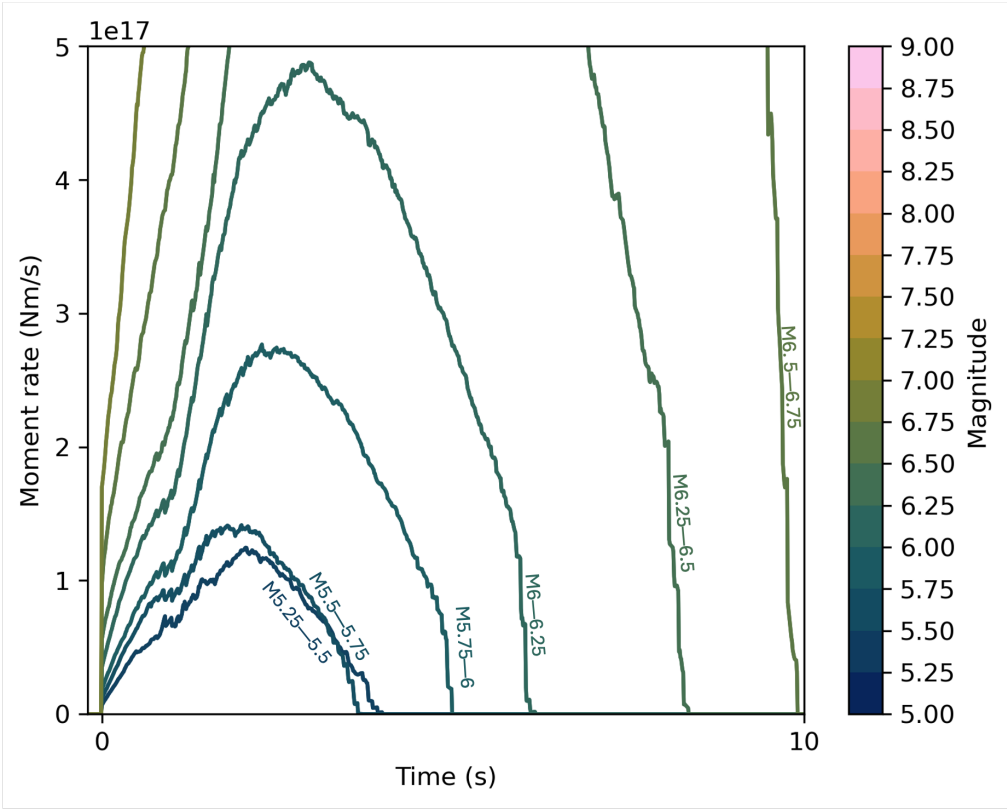
In Figure 4.8 we show the ensemble of the normalised STFs from each of our datasets (in grey), and the median calculated from each of these (blue). With the exception of the ISC dataset, the results seem to have several commonalities. This is more obvious in Figure 4.9 where we plot the median STFs for each dataset on a single plot. Other than the ISC dataset, the median STFs have a single peak, which occurs between 45% and 60% of the normalised duration. This result is similar to the findings of Meier et al. (2017), who found that the peak moment rate for median normalised STFs of M7-8.5 earthquakes fell between 35% and 55% of the total duration.

One interesting comparison is between the optimal and average STFs from SCARDEC (scardec_opt vs scardec_moy). The median of the average SCARDEC STFs has a later peak, and has very slow moment-rate for the first 10–20% of the normalised time. The median optimal STF meanwhile is broader, and looks more similar to the median Ye STF. The average SCARDEC median STF also has much more of a kink in the gradient at early stages than the optimal SCARDEC median STF, which is more linear.

Before taking the median, we choose normalise so that the peak moment-rate of each STF is 1. Therefore the peak value of the median STF tells us about



(a) M5–7.25 events.



(b) Zoom of the early stages of panel a.

Figure 4.6: Median STFs calculated using all 5 datasets for M5–7.25 events.

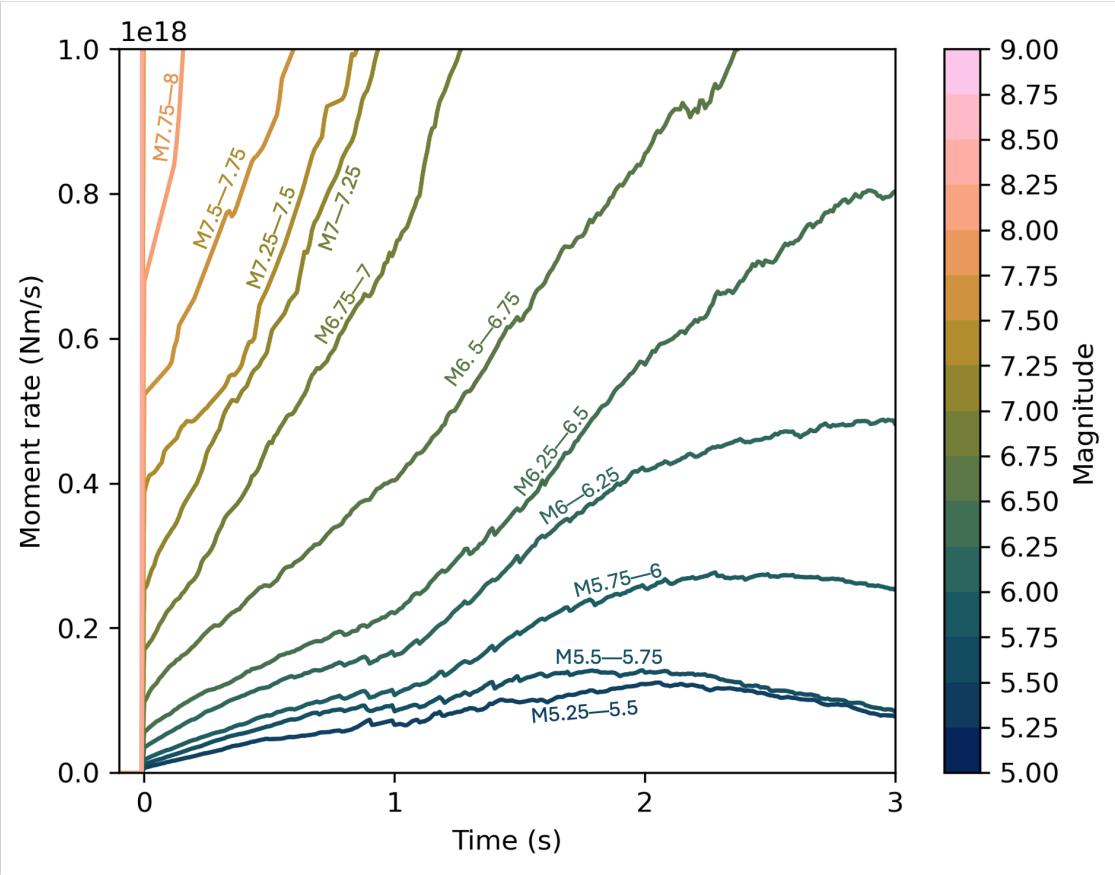


Figure 4.7: The first 3 s of moment-rate for the median STF of each magnitude bin. Cooler colours are smaller magnitudes.
Numbers of events in each bin: $M5.0$–5.25: $N = 5$; $M5.25$–5.5: $N = 33$; $M5.5$–5.75: $N = 204$; $M5.75$–6.0: $N = 2977$; $M6.0$–6.25: $N = 3029$; $M6.25$–6.5: $N = 2129$; $M6.5$–6.75: $N = 1227$; $M6.75$–7.0: $N = 824$; $M7.0$–7.25: $N = 434$; $M7.25$–7.5: $N = 242$; $M7.5$–7.75: $N = 175$; $M7.75$–8.0: $N = 111$; $M8.0$–8.25: $N = 48$; $M8.25$–8.5: $N = 24$; $M8.5$–8.75: $N = 3$; $M8.75$–9.0: $N = 3$; $M9.0$–9.25: $N = 3$.

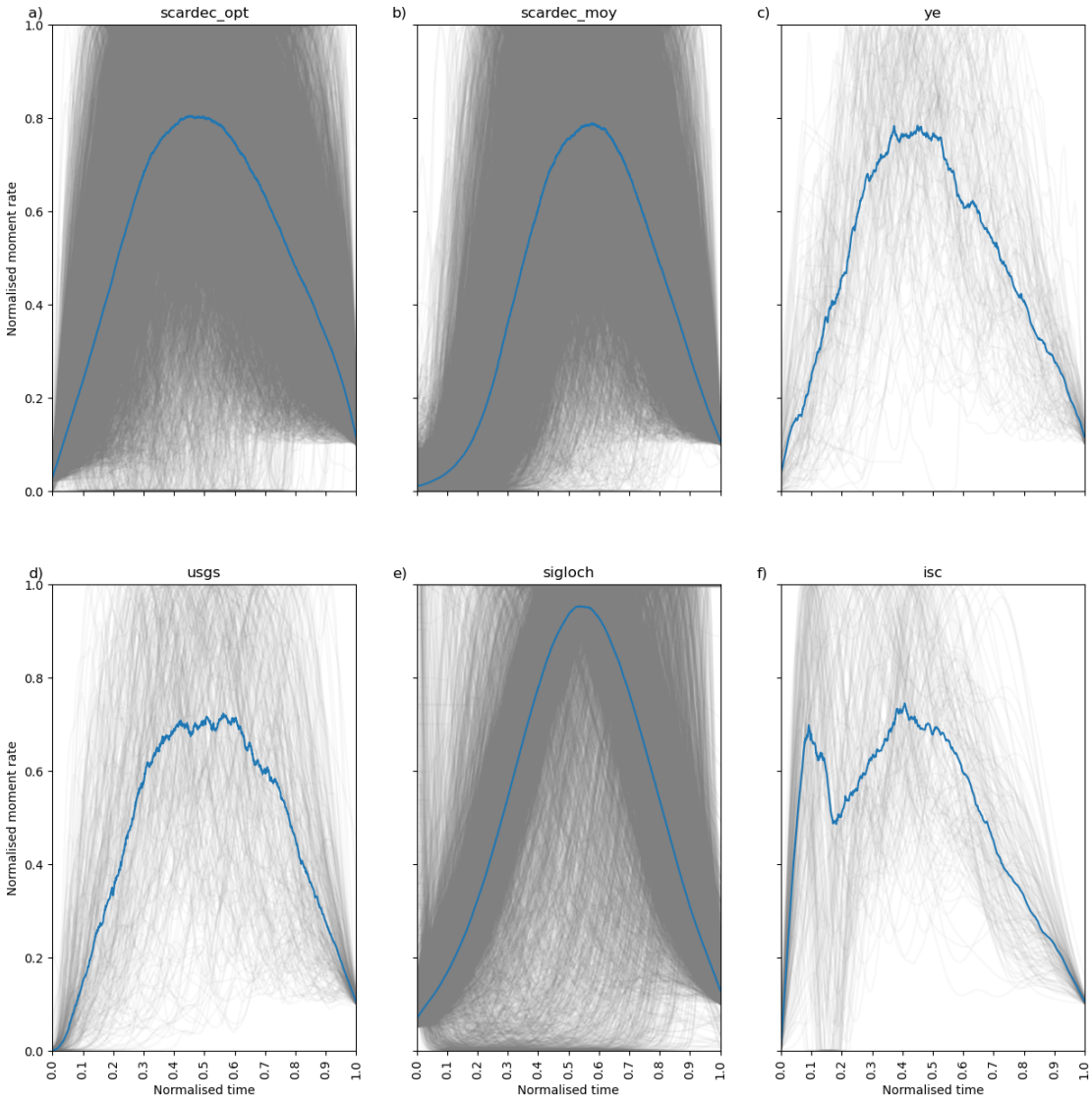


Figure 4.8: The median normalised STF for each dataset's STFs (blue). In grey, we show all STFs from that dataset.

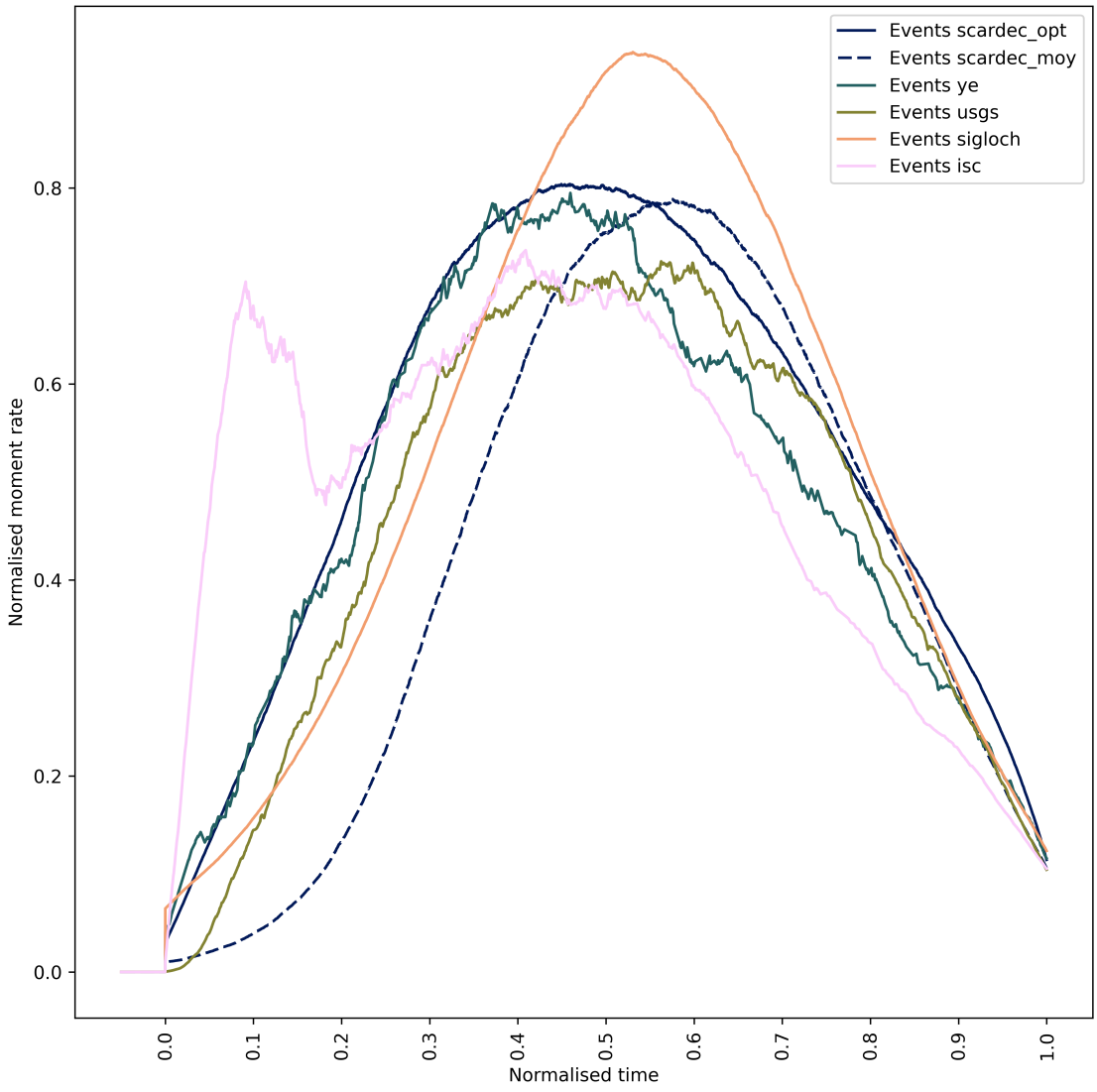


Figure 4.9: Compiled median normalised STFs.

how varied the positions of the peaks are in the normalised dataset. A theoretical dataset where the peak moment rates all occur at the same normalised time would have a median normalised STF peak with amplitude 1. More variation in the peak position leads to lower peak moment-rates of the median STF. The Sigloch dataset has a significantly higher peak value (Figure 4.9). This suggests that the STFs are more similar to each other compared to other datasets, and their peak timings are more closely aligned. This is also evident in Figure 4.8e, where there is a clear tract where most Sigloch STFs fall. The USGS median STF shows more of a plateau at its peak compared to the other datasets, which makes sense given the relatively large distribution of the observed STFs in Figure 4.8d. This may partly result from the larger magnitude events in the USGS dataset and the paucity of moderate magnitude events compared to the other datasets (apart from Ye). The ISC, Sigloch and SCARDEC datasets are dominated by moderate magnitude (M5.8–7) events, since these occur more frequently.

We have discussed the similarities between the SCARDEC, Ye, USGS and Sigloch datasets, however the ISC is a notable outlier. The median STF (pink, Figure 4.9) has two clear peaks, one at $\sim 10\%$ duration, and the second aligned with the other datasets at $\sim 50\%$ duration. This early and abrupt moment-acceleration is also evident in Figure 4.8f, where there is a relatively high density of STFs with very high initial moment acceleration.

The bimodal shape of the median ISC STF could arise from a number of factors. At least in part, it likely arises due to the STFs being formed from sums of Gaussians or harmonic basis functions which often have relatively abrupt onsets and lead to STFs with periods of low moment-rate. The STFs found using harmonic basis functions are often unstable (Garth, personal communication, 2025) and have periods of negative moment rate. The inversion also incorporates features that the Green's function does not resolve into the STF, particularly since the Green's function is determined using a 1-dimensional velocity model. As the ISC dataset is a beta-dataset, incorporating both STFs composed of Gaussians and those composed of harmonic basis functions, it is difficult to constrain the

relative impact of these different factors. As a result of these different factors we are hesitant to make any physical interpretation of the rapid moment-rate onset and bimodal shape of the STF.

4.4.2.2 Effect of magnitude on the median STF

As well as looking at variations between datasets, we also consider the effect of magnitude on the average STF shape. Figure 4.10 shows the median STF for 0.1 magnitude unit bins. Whilst there is a relatively high degree of scatter, the variation is among bins with fewer than 100 STFs (dashed). The solid lines form a clear, relatively uniform shape. The smaller magnitude events (cool colours), have a narrower but higher peak, suggesting there is less variation in the timing of the peak moment-rate or that STFs of small earthquakes tend to have a single peak and that large earthquakes have more peaks. This might suggest that smaller events tend to be simpler and have less variation in their evolution. For all the groups containing more than 100 STFs, the peak of the median STF occurs at about 0.54 (mean) of the duration (range 0.45–0.61, standard deviation 0.037). Figures B.11–B.16 show the median STFs by magnitude for individual datasets (blue), and the STFs which went into the calculation of each of the medians (grey). These figures highlight that the features we see in Figure 4.9 are not driven by events of particular magnitudes and instead are driven by differences between the datasets.

Summary

- SCARDEC, Sigloch, USGS and Ye STFs all have a single peak in their median STF between 45–60% of the normalised duration.
- SCARDEC average STF starts more slowly and peaks later than the optimal STF.
- Maximum median normalised moment rate is higher for the Sigloch dataset than the other datasets.
- Smaller magnitude events have a narrower but higher peak moment rate in the median normalised STF

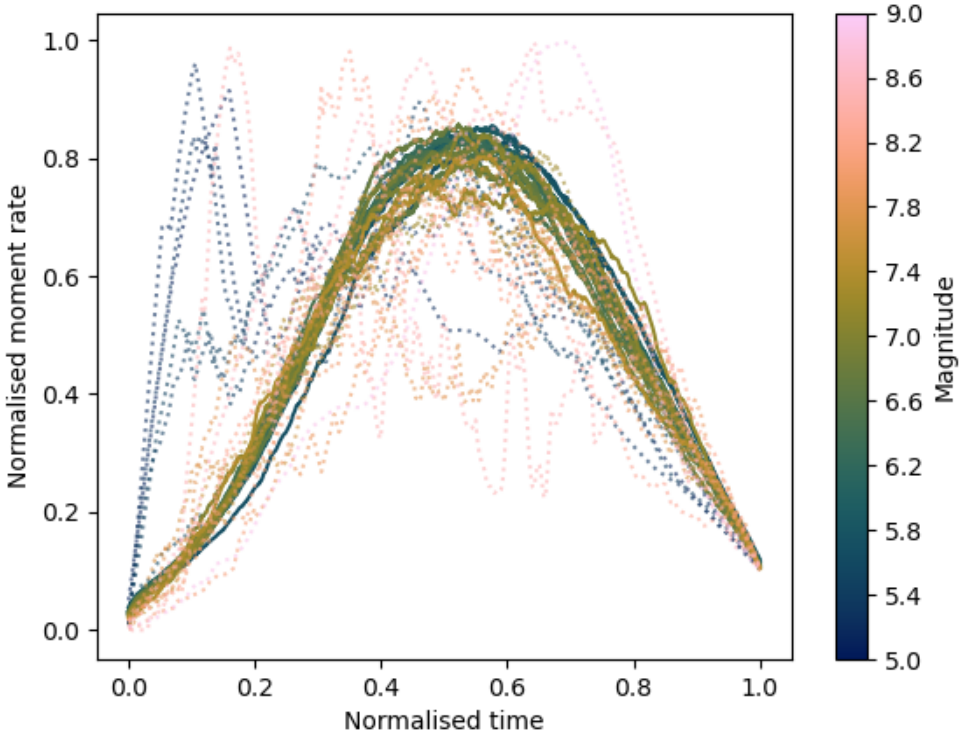


Figure 4.10: The median STF shape for $M0.1$ bins. Cool colours are lower magnitudes and warmer, lighter colours are higher magnitudes. Dashed lines show bins with fewer than 100 earthquakes.

4.4.3 How are the early portions of moment released?

4.4.3.1 Moment released in absolute windows of time

Danré et al. (2019) identify that measurements of the first sub-event allow the statistical inference of the 'future number and size of sub-events' and thus the final earthquake magnitude. In a similar vein, we would like to know whether the moment released in an absolute window of time (e.g. 1 s) has a statistical scaling with the final magnitude of the earthquake. In Figures 4.11–4.14, we plot the absolute moment released in one and three seconds respectively against the total moment (see Figures B.17–B.24 for 2, 4, 5 and 10 s windows). In both windows, all datasets except USGS have a statistically robust relationship with magnitude (Table 4.3).

Our results for the 1 second window are perhaps most interesting. Just 3.3 % of events have released all their moment in this window (Figures B.25 and B.26, these are predominantly in the Sigloch dataset), and 14% of events have released over one-third of their moment, and thus might be slowing down. However, we still see a statistical relationship between the final moment released and the moment released in this window. This suggests that larger earthquakes tend to release moment more quickly.

Table 4.3: Scaling relationships between the absolute moment in the first 1 and 3 s and the final moment.

Dataset	1 s			3 s		
	Grad.	Spearman	p-value	Grad.	Spearman	p-value
opt. SCARDEC	0.45	0.53	0.00	0.52	0.73	0.00
ave. SCARDEC	0.71	0.63	0.00	0.47	0.57	0.00
USGS	-0.29	-0.11	0.09	0.02	0.11	0.09
Sigloch	1.67	0.92	0.00	1.26	0.97	0.00
Ye	0.28	0.36	0.00	0.22	0.37	0.00
ISC	0.95	0.88	0.00	0.97	0.88	0.00
Overall	0.65	0.52	0.00	0.56	0.69	0.00

Notes: *Grad.* is the gradient (m) of the line of best fit $\log_{10}(M_0^{\text{final}}) = m \log_{10}(M_0^{\text{window}}) + c$. *Spearman* is the spearman correlation coefficient between $\log_{10}(M_0^{\text{final}})$ and $\log_{10}(M_0^{\text{window}})$. *p-value* is the p-value for a hypothesis test where H0 is that the data are ordinally uncorrelated.

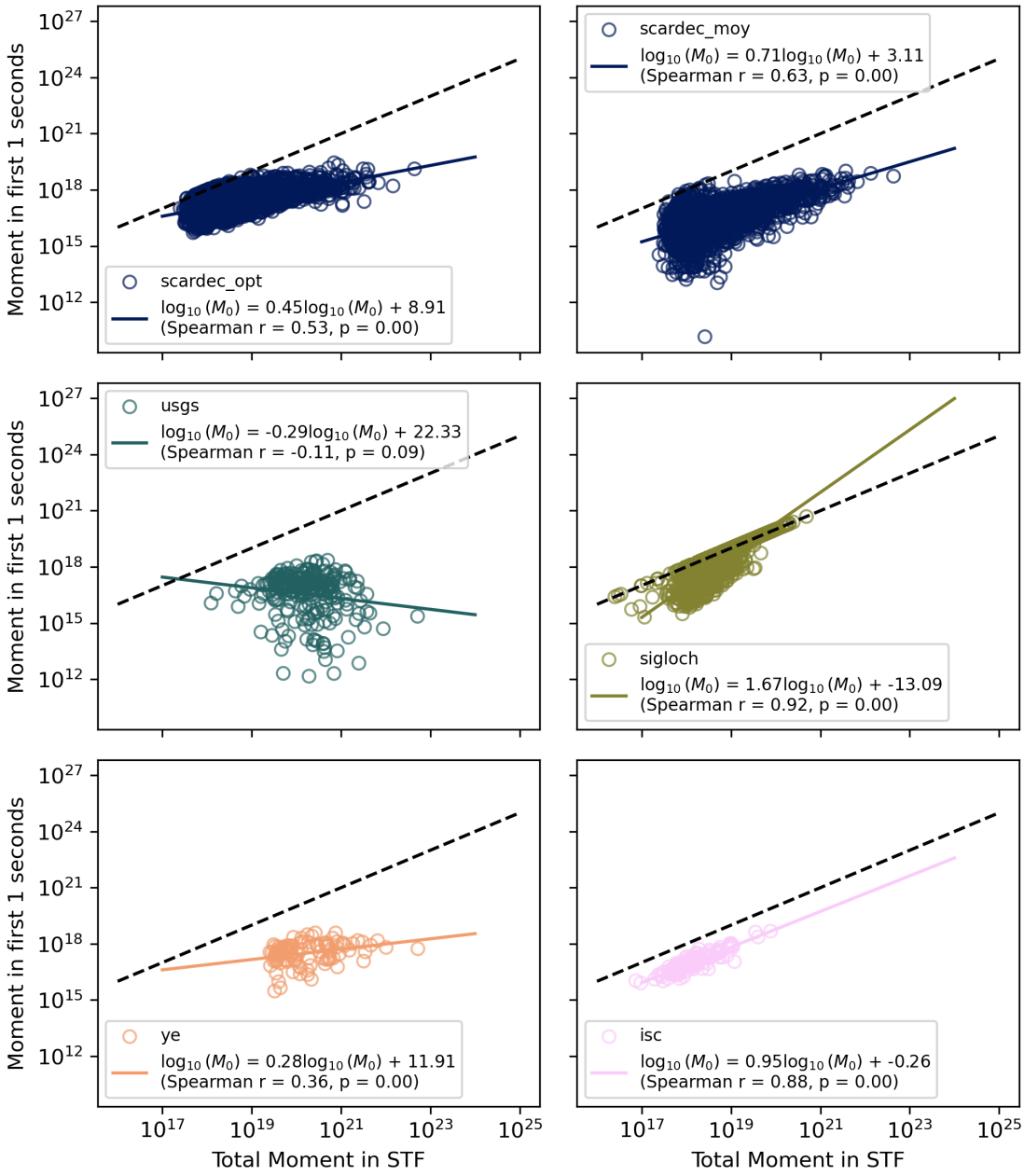


Figure 4.11: Absolute moment (Nm) released in first second of earthquake. Black dashed line shows a 1:1 scaling, that is events on this line release all of their moment in the first second.

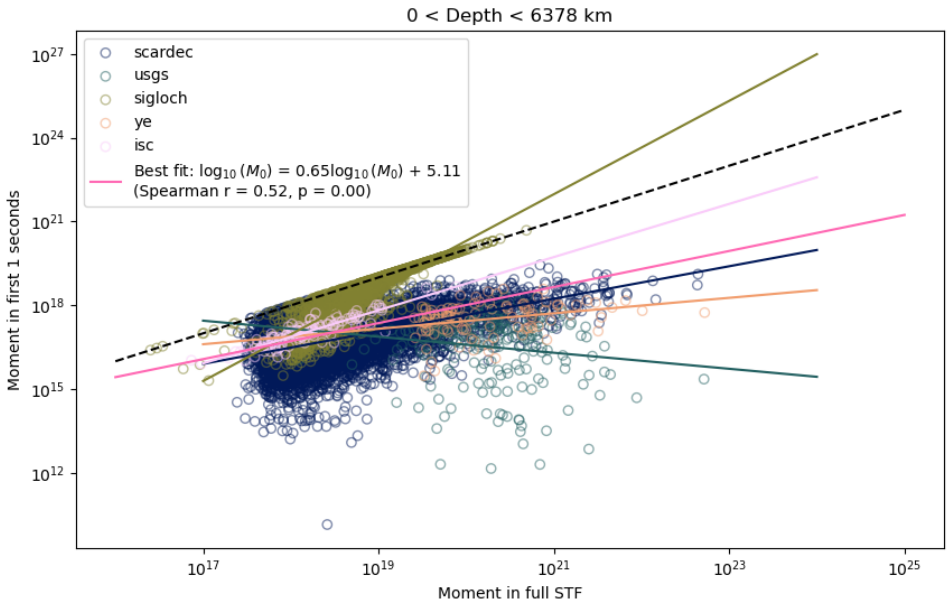


Figure 4.12: Absolute moment (Nm) released in first second of earthquake, with each colour denoting a different dataset. Black dashed line shows a 1:1 scaling, that is events on this line release all of their moment in the first second.

Looking at the proportion of moment released in the first seconds (Figures B.25–B.28), it is striking that the Sigloch data has a positive relationship (larger earthquakes are more likely to have released a high proportion of their moment early on) whereas for the other datasets it is smaller earthquakes which are likely to have released the majority of their moment in the first second. This is perhaps an artefact of the inversion process. Visually, some Sigloch STFs are very high amplitude and short in duration, particularly for large earthquakes which are at the upper bound of the method's capabilities.

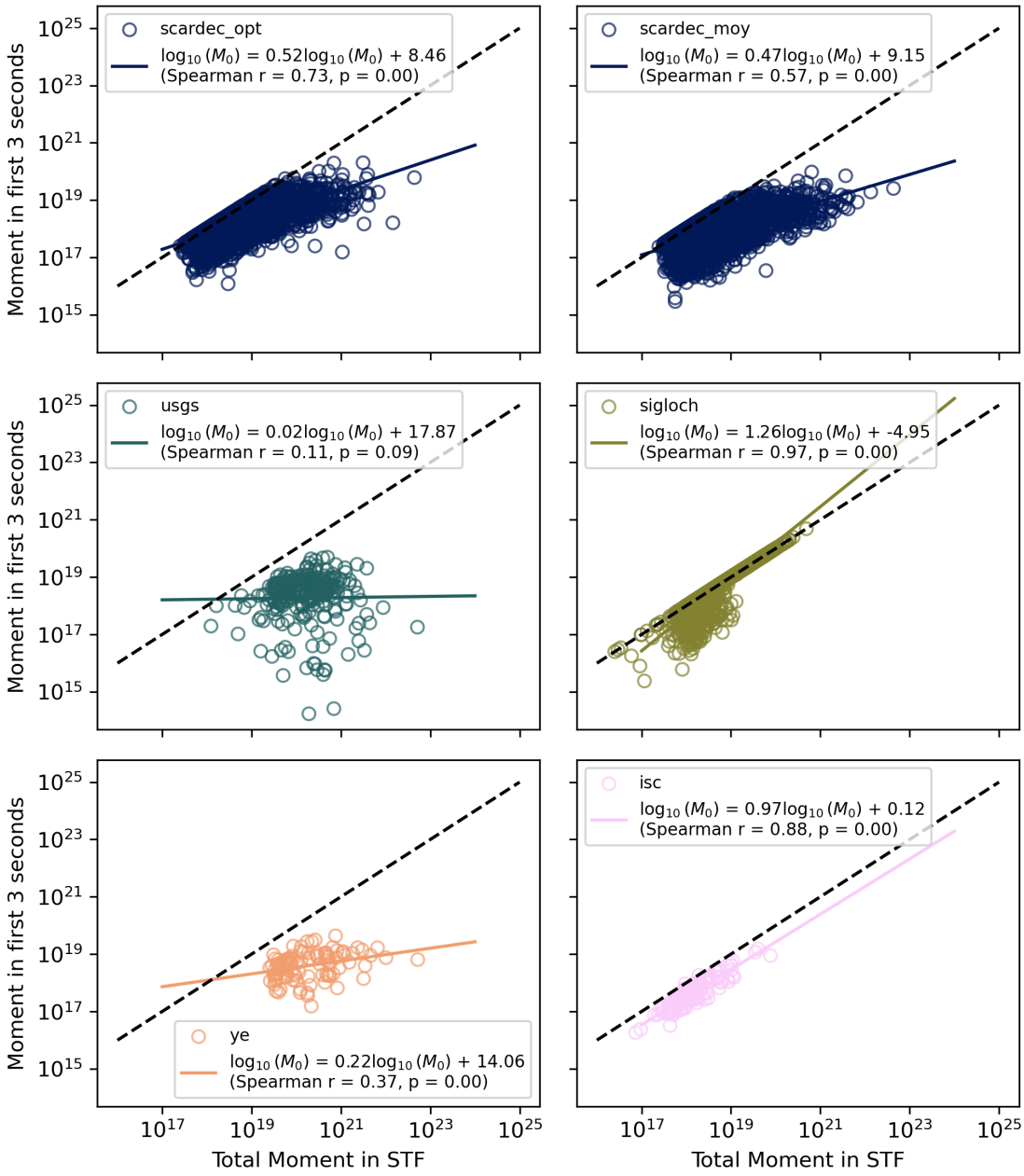


Figure 4.13: Absolute moment (Nm) released in first three seconds of earthquake. Black dashed line shows a 1:1 scaling, that is events on this line release all of their moment in the first three seconds.

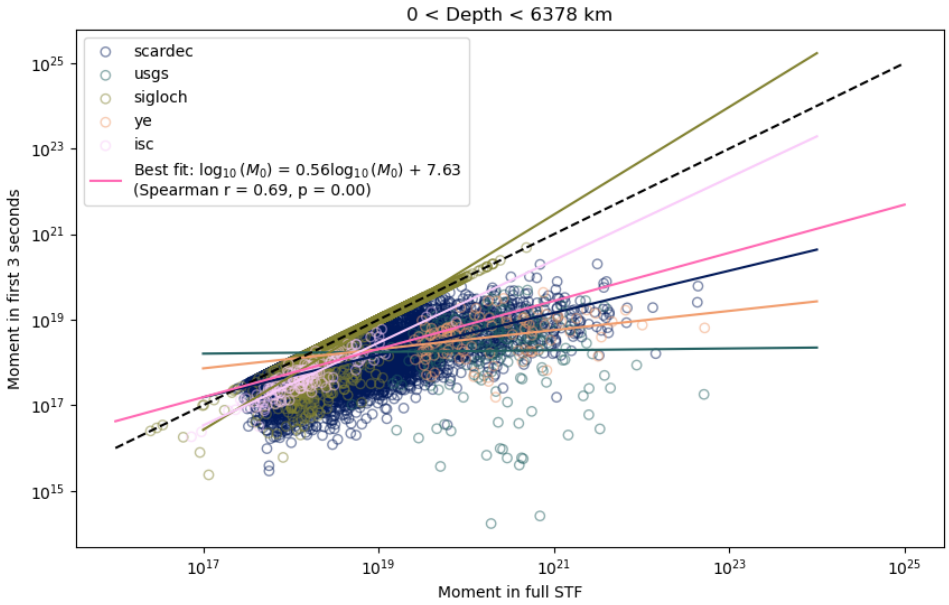


Figure 4.14: Absolute moment (Nm) released in first three seconds of earthquake, with each colour denoting a different dataset. Black dashed line shows a 1:1 scaling, that is events on this line release all of their moment in the first three seconds.

4.4.3.2 Time to release absolute amounts of moment

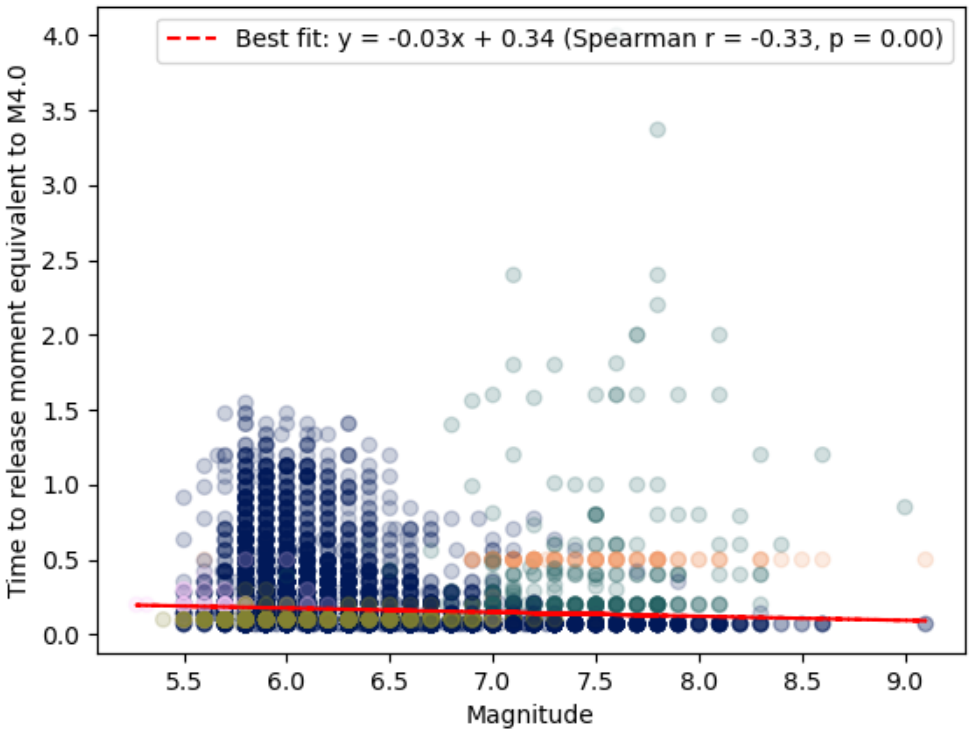


Figure 4.15: Time to release moment equivalent to a M4 earthquake vs final earthquake magnitude. Colours show datasets following previous figures.

If earthquakes are self-similar, the beginning of a large earthquake should be identical to that of a smaller earthquake. To test this, we look at how long (in absolute time) it takes for earthquakes of different magnitudes to release some quantity of moment. This is another way of thinking about the beginnings of earthquakes, rather than looking at the moment released in an absolute time window, we look at the time to release an absolute increment of moment. First we look at the relationship between final earthquake magnitude and the time taken to release the equivalent of a M4 earthquake (Figure 4.15). We see a very weak, but statistically significant, negative relationship suggesting that large earthquakes take *slightly* less time to release this very initial portion of their moment (gradient = -0.03, Spearman r = -0.33, p = 0.00). We also look at the time taken to release

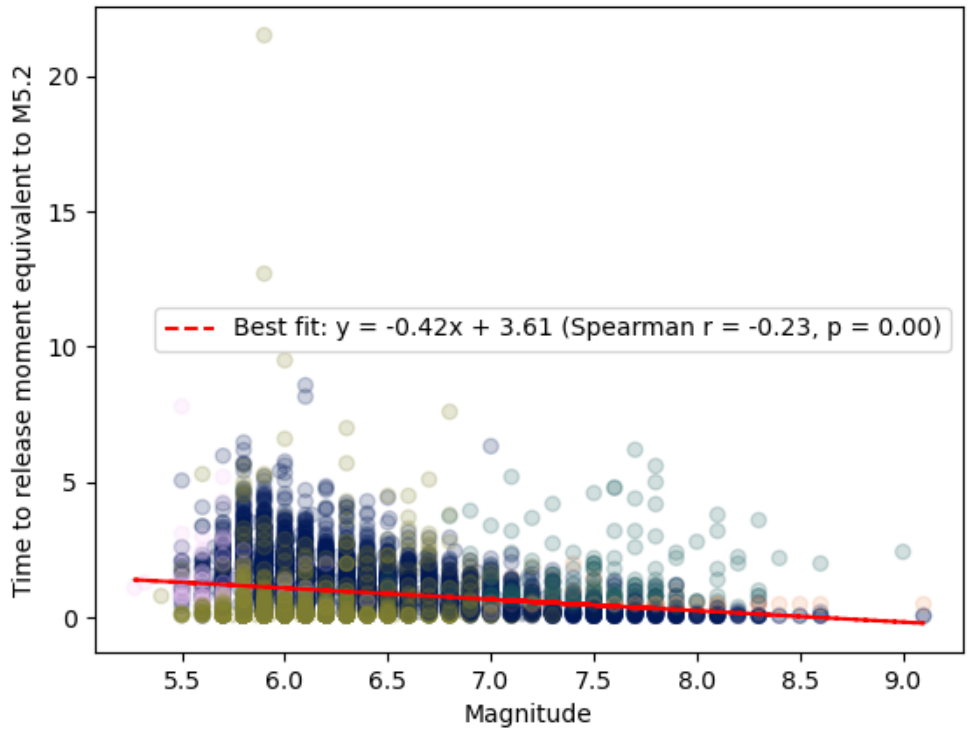


Figure 4.16: As Figure 4.15 but for releasing moment equivalent to 10% of a M6 (=M5.2).

moment equivalent to the first 10% of a M6 earthquake (Figure 4.16). This has a much stronger relationship (gradient = 0.48, Spearman r = -0.23).

In Tables B.1–B.4, we show the strength of the relationship between the time taken to release various amounts of moment (equal to M4–7 earthquakes) and the final magnitude of the events (see also Figure B.29). Fits only consider events which have a total moment more than the moment of interest. Looking at the relationship for each individual dataset, the SCARDEC and ISC datasets show a strong negative relationship between the time to release even relatively small portions of moment, and the final magnitude of the earthquake. The Ye and USGS datasets contain much larger earthquakes on average, and show much less of a trend with magnitude. The USGS data only shows a significant trend once the equivalent of a M6.5 earthquake has been released. The Ye data is slightly more useful, only requiring the equivalent of a M5.5 earthquake to have been released

before a significant relationship with magnitude is uncovered.

As with the proportion of moment released in absolute time windows (Figures B.25–B.28), the Sigloch data has the opposite trend with magnitude to the other datasets. Here, larger events in the Sigloch data tend take longer to release the moment associated with a smaller event. This seems almost contradictory with the observations of the proportion of moment released in an absolute time window, where larger events tended to have released proportionally more of their moment. This may result from the relative simplicity of the Sigloch STFs, of which over 90% can be adequately described by a single Gaussian, and the inversion-imposed relatively short duration of the Sigloch STFs (noting that most are considerably shorter than the maximum allowable 25.6 s).

Summary

- All but the USGS dataset show a statistically significant relationship between the moment released in the first second and the total moment released.
- Only 14% of events have released more than 1/3 of their moment in the first second, and just 3% have released all their moment.
- Larger earthquakes take less time to release the same absolute quantity of moment.

4.4.4 Does the timing of the maximum moment release vary systematically?

Another first order attribute of STFs is the relative timing of the peak moment rate. We find that there is no systematic variation of the timing of peak moment rate with magnitude (mean = 0.53, Figure 4.17) or with earthquake depth (Figure 4.18). The scatter in the timing of the maximum moment rate primarily results from intra-event variation (see grey lines of Figure 4.8 and Section 4.4.2.2), other than ISC STFS tending to have a very early peak (mean values of peak timing are: optimal SCARDEC 0.49, average SCARDEC 0.56, Sigloch 0.54, ISC 0.30, USGS 0.51, Ye 0.44).

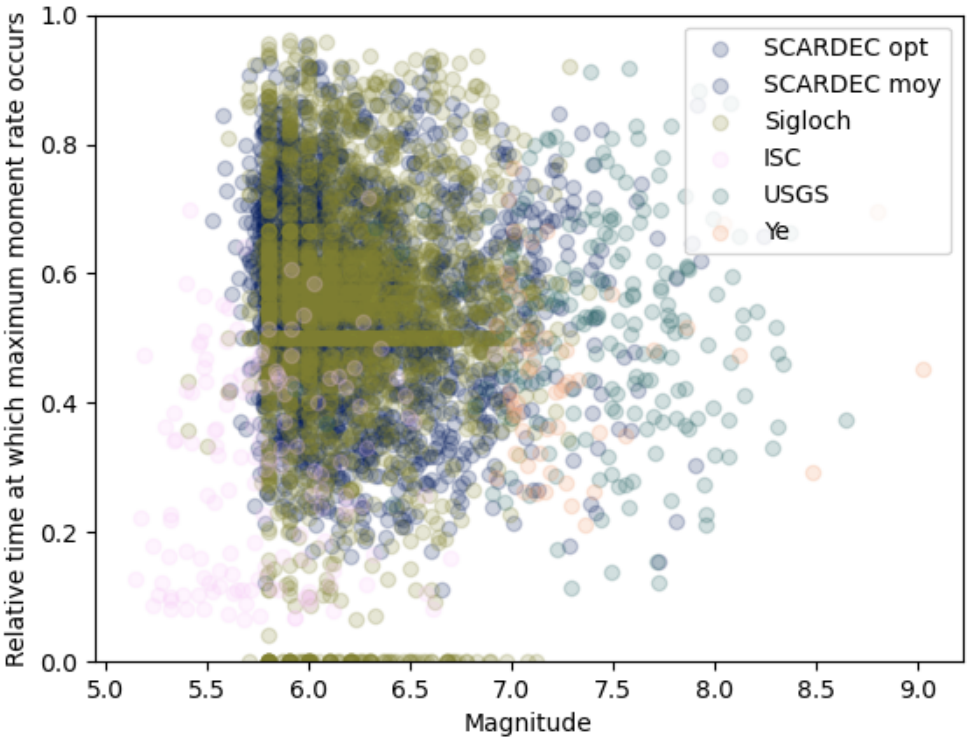


Figure 4.17: The relative timing of the maximum moment rate for each earthquake. Colours represent different datasets.

4.4.4.1 Comparing the maximum moment rate vs average moment rate

We also compare the value of the maximum moment rate to the average moment rate (Figure B.30). In this we see that magnitude has no effect on the ratio of the maximum to average moment rate (Spearman r: 0.0091, p-value: 0.32). This might suggest that there is some degree of similarity between STFs of different magnitudes.

Summary

- We find no systematic variation in the normalised timing of peak moment-rate with magnitude or depth.
- We find no trend in the ratio of average moment rate to maximum moment rate with magnitude.

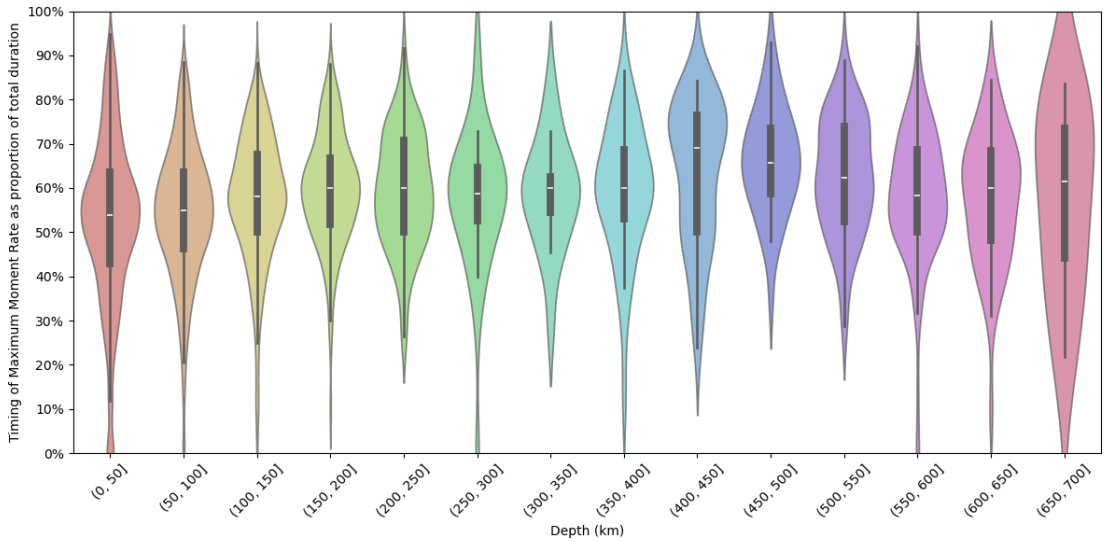


Figure 4.18: Violin plot of the timing of the maximum moment rate for different depth bins of events.

4.4.5 How do statistical measures of the STF vary?

We can use statistical measures of the STF such as skew and kurtosis to look for and quantify trends in the STF with depth, magntiude and dataset.

4.4.5.1 Quantifying the shape of the STF: skew and kurtosis

The skew and kurtosis quantify the symmetry and peakedness of the STFs respectively. The mean skew is 0.39 and the mean kurtosis is -0.91. This suggests that the STFs tend to be slightly right tailed, and much flatter than a Gaussian. Physically, this suggests that moment-rate accelerates at a slightly higher rate than it decelerates. Sigloch STFs make up the majority of events with very high values of skew and/or kurtosis (Figures B.33, B.36). In these few cases, the long tail perhaps implies our adjusted beginning choice may be too late, and that the STF is very peaked compared to the other datasets (and indeed a Gaussian). Just 1% of STFs have a kurtosis greater than 3 (the kurtosis of a standard Gaussian) and 63/68 of these are from the Sigloch dataset.

4.4.5.2 Trends in skew and kurtosis with depth and magnitude

We see no trend in skew or kurtosis with magnitude or depth (Figures B.31, B.32, B.34, B.35), unlike Houston et al. (1998). Our skew values may be an underestimate due to the definition we use of the end of the STF (a trend of early terminating STFs impacting skewness is noted by Houston (2001) for another catalog). Our kurtosis values are generally lower than previous studies (e.g. Tocheport et al., 2007) and are lower than we expect from a Gaussian distribution (~ 2.7), aligning with our geometric fits of the median STFs.

Summary

- The mean skew is positive suggesting moment rate increases more rapidly than it decreases.
- All datasets have a negative kurtosis which is much lower than expected from a standard Gaussian.

- No systematic variations in skew or kurtosis with magnitude or depth

4.4.6 Are there patterns in STF complexity?

Visual inspection of STFs shows that some STFs are very simple and are composed of a single peak, whilst others are very complex. Some of this complexity is likely real and results from variation in earthquake source complexity. This may have a relationship with earthquake magnitude or depth, as many authors have observed. For example, Neely et al. (2024) suggest that magnitude–stress-drop trends may result from changing STF complexity, rather than trends in actual stress drop and Houston et al. (1998) observe that the STFs of very deep earthquakes are less complex than those of intermediate and shallow depths. Yin et al. (2021) analyse groups of STFs through dynamic time warping, and find that shallow SCARDEC STFs tend to be more complex than deeper events, and that there is no relationship between STF complexity and magnitude, whereas Danré et al. (2019) find that larger events are more complex and composed of more sub-events. Sato and Mori (2006) use waveform complexity to suggest that small earthquakes are a mixture of simple and complex events, whilst large earthquakes are always complex.

However, some of the complexity apparent in STFs may be an artefact of the inversion process, and this may vary between datasets. In Section 4.2, we posed some questions about the effect of different method choices on the resulting STFs. Based on the approaches used to produce STFs, we expected that Sigloch STFs might tend to be more simple and more often have a single peak, since that is what is preferred in their method. We also suggested that SCARDEC might have a bias in moderate events towards STFs that are more simple. On the other hand, the ISC STFs may be expected to be more complex since they are produced by the addition of several Gaussians.

In this section we therefore want to quantify STF complexity (primarily through fitting of Gaussians) and look at trends in complexity between datasets and as a function of earthquake magnitude and depth.

4.4.6.1 Quantifying complexity by fitting Gaussians

We want to calculate the number of Gaussians required to fit each STF with $R^2 \geq 95\%$. We first smooth the STF using a moving average box by convolution. The length of the box is defined as 5% of the duration of the STF. We then attempt to fit a single Gaussian to the STF, with the initial guess having a mean equal to the peak moment-rate, and calculate the R^2 value between the observed STF and the optimized Gaussian. If this is above our threshold of 0.95, we consider a single Gaussian to be a good fit to the data. Otherwise, we attempt to fit ever increasing numbers of Gaussians until the $R^2 > 0.95$, with initial proposed Gaussians having evenly spaced means, although this is allowed to vary in the course of the optimization.

In past work, every peak has been considered as a sub-event with only minor filtering (0.5 Hz for Houston et al. (1998) or amplitude requirements ($4\sigma > 1$s, peak over 10% of STF maximum for Danré et al. (2019)). These approaches all lead to some events having relatively large numbers of STFs. Instead we smooth the STF with a boxcar that has a length of 5% of the duration and we also limit the number of Gaussians that are allowed to be used to 10, which prevents very large numbers of sub-events being fit, since these are more likely to be small perturbations. We must also note that using the number of Gaussians that result from our fitting approach does not directly allow us to infer a number of sub-events. Peaks in the STF which are not Gaussian in shape may require more than one Gaussian to appropriately fit them. However, there is a correlation between the number of Gaussians required and the complexity of an event (as determined from the number of peaks of the STF).

Considering each dataset as a whole, there is significant variation in the mean number of peaks required to fit the STFs (Column '5–9' of Table 4.5). The Sigloch data requires a mean of 1.36 Gaussians, whilst the ISC requires over twice that many Gaussians on average, with a mean of 3.49. This suggests that the complexity of the STF does differ depending on the inversion process used and so in Figure 4.19, we compare the number of Gaussians required to fit the various STFs of earthquakes which appear in more than one dataset. There are some systematic trends obvious

Table 4.4: Number of events in each magnitude bin.

	Magnitude bins									
Dataset	5-6	5-7	5-8	5-9	6-7	6-8	6-9	7-8	7-9	8-9
opt. SCARDEC	1042	3561	3875	3898	2519	2833	2856	314	337	23
ave. SCARDEC	966	3494	3822	3846	2528	2856	2880	328	352	24
Sigloch	348	1043	1067	1068	695	719	720	24	25	1
ISC	100	137	137	137	37	37	37	0	0	0
USGS	1	25	187	206	24	186	205	162	181	19
Ye	0	12	97	107	12	97	107	85	95	10

Table 4.5: Mean number of Gaussians to fit STF.

	Magnitude bins									
Dataset	5-6	5-7	5-8	5-9	6-7	6-8	6-9	7-8	7-9	8-9
opt. SCARDEC	1.80	2.02	2.06	2.06	2.11	2.15	2.16	2.51	2.53	2.74
ave. SCARDEC	1.61	1.76	1.79	1.80	1.82	1.85	1.86	2.13	2.16	2.50
Sigloch	1.42	1.37	1.36	1.36	1.34	1.33	1.33	1.12	1.12	1.00
ISC	3.21	3.26	3.26	3.26	3.38	3.38	3.38	-	-	-
USGS	3.00	2.32	2.27	2.26	2.29	2.26	2.25	2.26	2.25	2.16
Ye	-	2.75	2.49	2.47	2.75	2.49	2.47	2.46	2.43	2.20

Note: categories with no data are denoted with a dash.

in these comparisons: (1) the average SCARDEC STF tends to have more peaks than the optimal SCARDEC STF; (2) the number of peaks in the Sigloch STF is relatively independent of the number of peaks in any other database; (3) the number of peaks in the Ye STFs is better correlated with the optimal SCARDEC STF than the average SCARDEC STF; (4) the ISC STFs almost always have more peaks than corresponding STFs in other datasets, with no particularly strong correlation in how many more peaks there are. The lack of correlation between some datasets (points 2 and 4) give an important perspective: the implications of method choice do not simply scale the complexity of the STF, but actually alter or remove it.

4.4.6.2 Events which require a single Gaussian to fit them

Another metric for average dataset complexity is the number of events which require only a single peak to fit them. Liu et al. (2023) decompose average SCARDEC STFs

using a Brune model and find that 51% are composed of a single sub-event. The Brune model, even when adapted to have two corner frequencies, is fundamentally derived for simple (and relatively small) earthquakes where the source complexity does not cause deviation from a frequency-squared spectral decay (e.g. Luco, 1985; Beresnev and Atkinson, 2001; Denolle and Shearer, 2016) and may not be well suited to more complex, larger events. The Gaussian fits of our work and Danré et al. (2019) have an additional free parameter compared to a Brune fit, and Danré et al. (2019) therefore find that just 19% of events are composed of a single sub-event. Our results vary depending on the length-scale of the smoothing used. With smoothing at 5% of the total duration, we find that 46% of events can be fit by a single Gaussian (optimal SCARDEC: 38%, average SCARDEC: 46%, Ye: 3%, USGS: 13%, Sigloch: 91%, ISC: 5%). At a longer-term smoothing (20% of total duration), which we use to consider larger scale features, 67% of events are fit by a single peak (optimal SCARDEC: 63%, average SCARDEC: 70%, Ye: 46%, USGS: 49%, Sigloch: 95%, ISC: 34%).

4.4.6.3 Patterns in complexity with magnitude

In Danré et al.'s (2019) analysis of the SCARDEC data, they observe that larger magnitude earthquakes are more complex and contain more sub-events, each represented by a peak in the STF. Given how different methods can lead to vastly different estimates of complexity, do we see a similar trend of increasing complexity with magnitude?

Despite our different approach to counting peaks, for the SCARDEC data we see a similar trend to Danré et al. (2019) (top two rows of Figure 4.20; blue). As we consider groups of events with larger magnitudes, the mean number of Gaussians required to fit the optimal SCARDEC STF increases from 1.8 (M5–6) and $\sim 2.04 \pm 0.02$ (M5-7, M5-8, M5–9) to 2.52 ± 0.01 (M7–8, 7–9) (Tables 4.4 and 4.5). The ISC STFs also show a trend of increasing complexity with magnitude (increase from 3.21 (M5–6) to 3.38 ($M \geq 6$), Figure 4.20 and Tables 4.4 and 4.5).

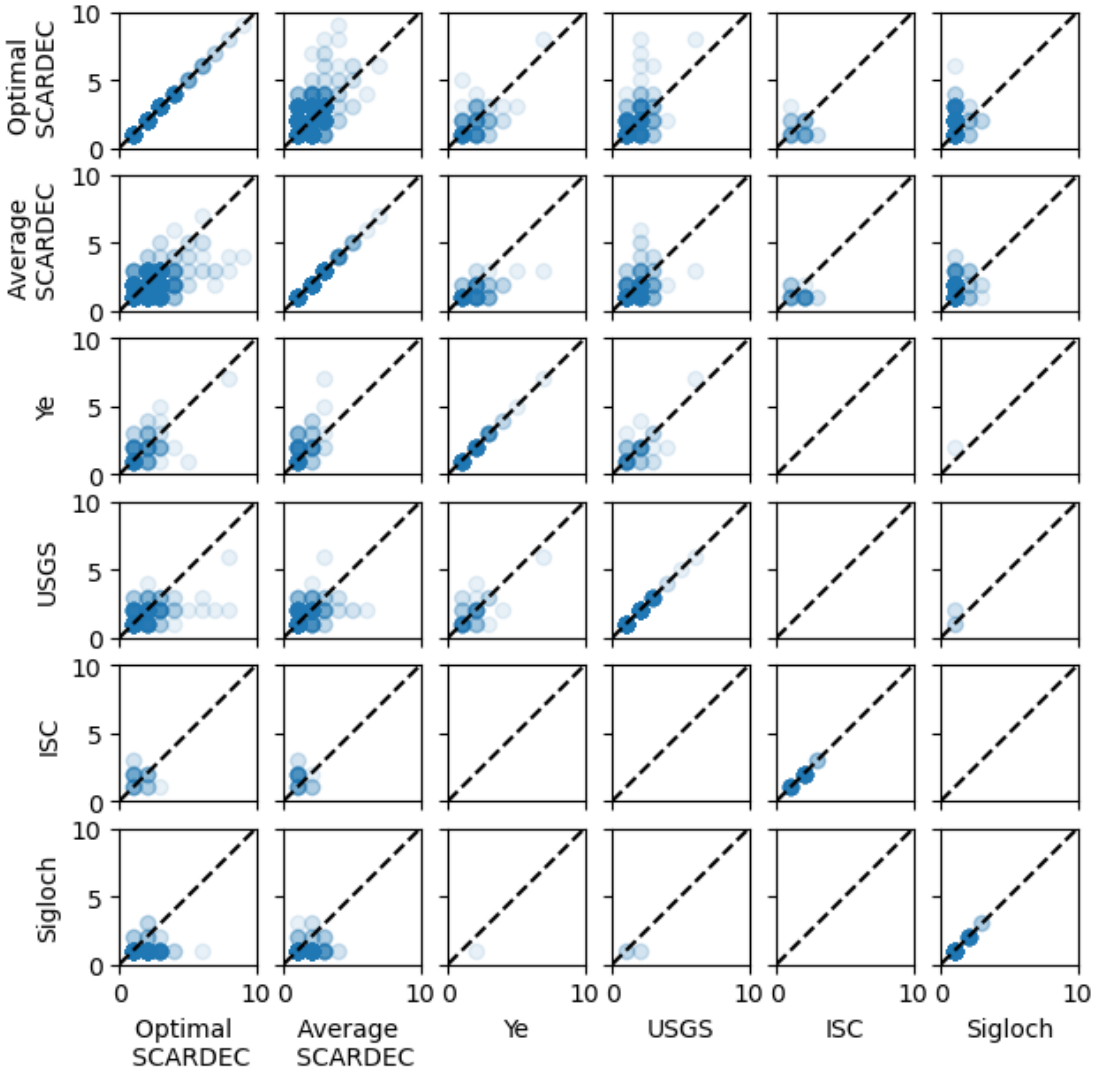


Figure 4.19: Comparison of the number of Gaussians required to fit a corresponding earthquake that appears in more than one dataset.

However the other datasets we examine have either no real trend between complexity and magnitude, or a trend of decreasing complexity with magnitude. The USGS dataset shows a much smaller variation in the number of Gaussians required to fit the STF. Except for the smallest and largest magnitude bins ($M5$–6 and $M8$–9), all bins have a mean number of Gaussians between 2.25 and 2.32. The $M5$–6 bin has $n = 1$ and so can be ignored (Table 4.4). The $M8$–9 bin has $n = 19$ and requires a mean of 2.16 Gaussians to fit the STF. This trend is also present with the Ye dataset. The small numbers of events in the $M5$–7 and $M6$–7 bins ($n = 12$) require a mean of 2.75 Gaussians to fit their STFs. As we consider larger events, the

mean number of Gaussians required falls very slightly from 2.49 ($M5$–8, $M6$–8) to 2.43 ($M7$–9). Only looking at the largest earthquakes ($M8$–9), the mean number of Gaussians is only 2.20, although there are few events in this group ($n = 10$). In the larger Sigloch dataset, we again see something similar. Except for the $M5$–6 group (1.42), the $M \geq 5$ bins have a mean number of Gaussians of ~ 1.36, the $M \geq 6$ bins have a mean of ~ 1.33 and the $M \geq 7$ bins have a mean of 1.12, though these $M \geq 7$ groups should be taken with caution, since the inversion method is only intended for earthquakes up to $M7.2$ (also note n is 24 and 25 for the two $M \geq 7$ bins). See Figure 4.20 and Tables 4.4 and 4.5 for detail of the mean number of Gaussians required to fit different datasets and earthquake magnitudes.

By looking at broader scale features of the STF by first smoothing them, we mitigate some of the impact of attenuation on our results. Nonetheless, smaller STFs are more likely to be smoothed by attenuation, and thus for their complexity to be underestimated.

That the Ye STFs tend to be simpler is not unexpected, since the method tends to oversimplify fault geometry and imposes a constant rupture velocity (Section 4.2.4.3). However it is more surprising that we do not see a trend of increasing complexity with magnitude for the USGS data, where there are fewer imposed constraints on the fault geometry and rupture velocities. This may simply be a coincidence: the SCARDEC dataset has far more earthquakes and so could be considered to be more reliable.

When initially discussing the SCARDEC method (Section 4.2.1), we suggested that the SCARDEC dataset might have a bias at moderate magnitudes towards more simple events. Our results cannot confirm whether this is the case, since we cannot fit Gaussians to STFs which do not exist. However, we would expect our observation that SCARDEC STFs exhibit increasing complexity with magnitude to stand, even if the STF complexities we have are a lower bound for moderate magnitudes.

4.4.6.4 Patterns in complexity with depth

In addition to trends between STF complexity and magnitude, previous work has also investigated how STF complexity varies with earthquake source depth. Houston et al. (1998) define complexity as the number of zero crossings in the moment-acceleration curve, after lowpass filtering at 0.5 Hz. Analysis of complexity using this definition has found that very deep earthquakes have the simplest STFs, and that the most complex STFs occur between ~ 250 and ~ 550 km, depending on the dataset used (Houston et al., 1998; Houston, 2001; Persh and Houston, 2004; Tocheport et al., 2007). For the Sigloch, USGS and both SCARDEC datasets, we see a subtle trend of decreasing complexity with depth (Figure 4.21), statistically distinguishable from no ordinal correlation at $\alpha = 0.05$. The Ye dataset only contains earthquakes with depths less than 100 km, and so it makes sense there is no clear trend with magnitude. The ISC data is predominantly depth < 250 km, except for a single deep earthquake. For both datasets, despite the line of best fit having a positive gradient, $p > 0.05$ and thus we accept the null hypothesis that the data are ordinally uncorrelated. We do not see evidence that earthquakes between 250 and 550 km depth are more complex.

Summary

- We quantify complexity by asking how many Gaussians are required to adequately fit a given STF.
- Many STFs are not well fit by a single Gaussian, and have multiple peaks.
- SCARDEC STFs have an increasing complexity with magnitude.
- Sigloch and Ye STFs have a more subtle decrease in complexity with magnitude.
- USGS and ISC STFs show little trend in complexity with magnitude.
- Some datasets have consistently higher complexities than others; ISC STFs tend to be complex whilst Sigloch STFs tend to be simple.
- A M8 earthquake is likely to have a similar complexity of USGS STF as a M6 earthquake, but an earthquake with a complicated STF in the USGS dataset is likely to also have a complicated SCARDEC STF.

- For SCARDEC, Sigloch and ISC STFs, there is a very weak but statistically significant negative trend between complexity and earthquake depth. USGS and Ye datasets have few deep events.

4.4.7 What geometry best fits an STF?

4.4.7.1 Geometric fits to the overall median STF

We would still like to be able to compare STFs to a generic geometric model of an average STF. Despite the differences apparent between the STFs from the different datasets, as an initial effort we calculate the median STF of all the STFs in our dataset (Figure B.37). We propose five geometric models: a Gaussian, a triangle, a boxcar, a boxcar tapered with sine functions (in the limit this is just two independently inverted sine functions), and a trapezium (Figure B.38). We systematically vary the parameters which describe each geometry, aiming to reduce the L2-norm difference between the geometric model and our observed median STF. We propose a Gaussian since the median STF appears by eye to be relatively Gaussian, and the ISC STFs are built on sums of Gaussians. We propose the boxcar as a potential geometric fit as this is the shape of the STF used in GCMT inversions before 2004 (Ekström et al., 2012). We also propose a sine taper on either end, since the STFs do not appear to suddenly start and stop. Since 2004, GCMT have modelled the STF as an isosceles triangle; however, visual analysis of our median STFs and the observations of Meier et al. (2017) prompt us to remove the requirement of the triangle being isosceles, since they observe the peak of the median STF to have some variation around the half-duration (the peak occurs at 35%–55% of the normalised duration). Finally, we also propose a trapezium, based on the physical intuition that an earthquake might accelerate, followed by the moment-rate plateauing for the main phase of rupture and then, as the earthquake dies, the moment rate decreasing again. Additionally, we allow the trapezium and triangle to have a linear ramp at the start, as the average STF seems to have a change of gradient around 20% duration and some previous workers have proposed a low moment-rate early stage of rupture (nucleation phase, Ellsworth and Beroza (1995)).

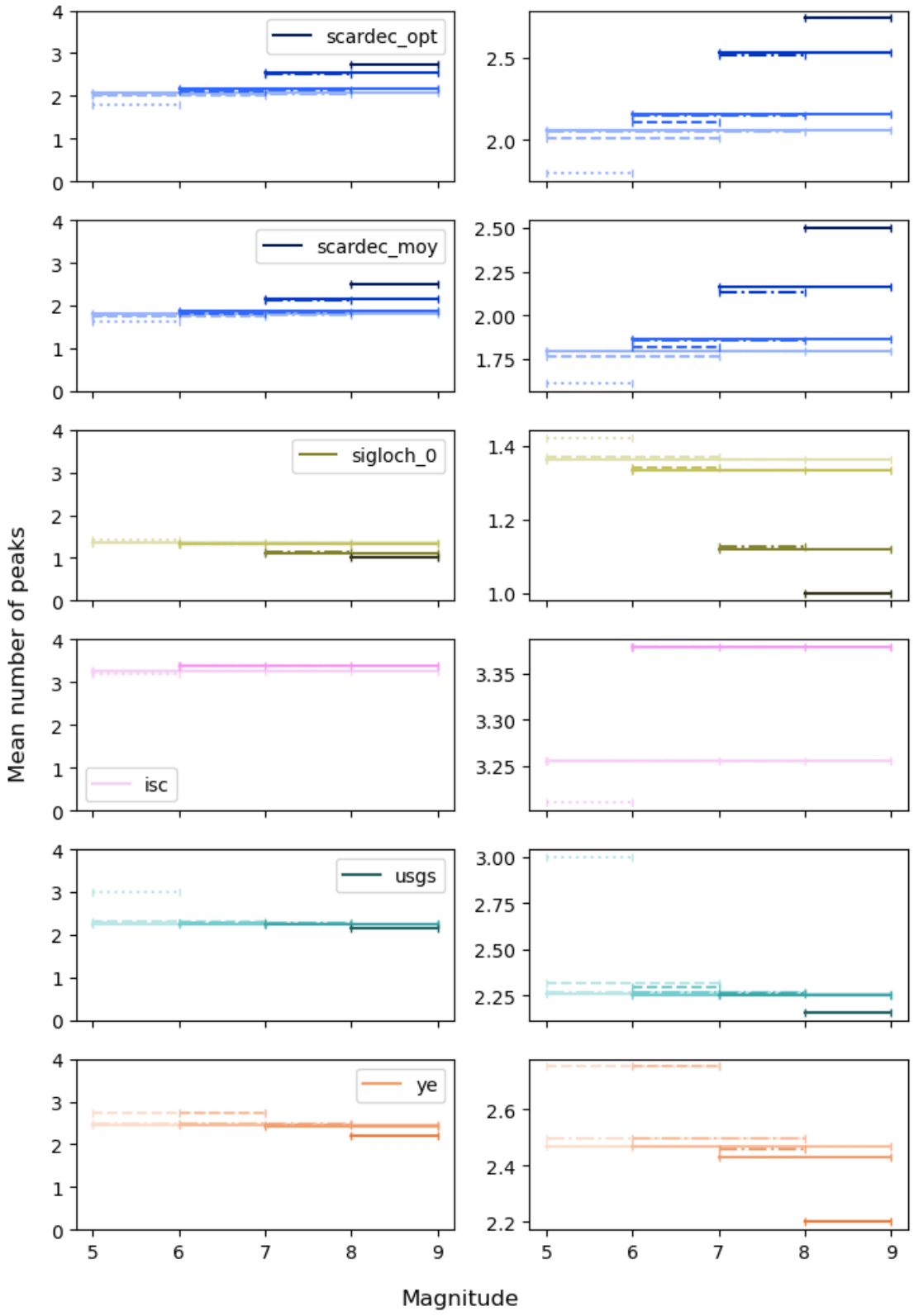


Figure 4.20: Mean number of Gaussians required to fit STFs from different datasets in different magnitude bins. Each dataset is shown in a different colour and the shade of the colour shows the minimum magnitude. Different maximum magnitudes have different line styles. Right hand column is a zoom of the left hand column to aid interpretation.

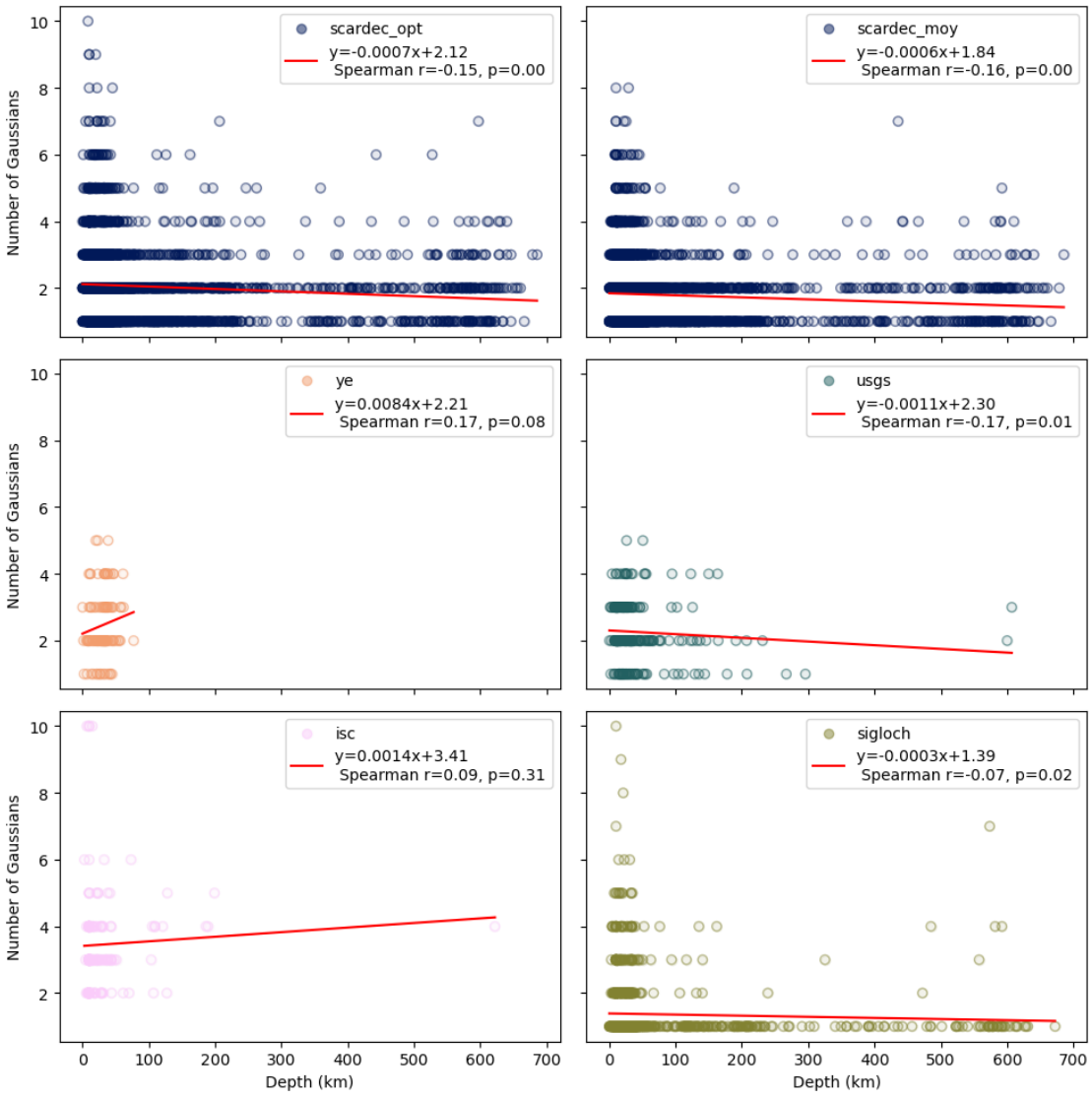


Figure 4.21: Plot of the number of Gaussians required to fit each STF against depth. Lines of best fit (red) are drawn over the range of depths present in the dataset.

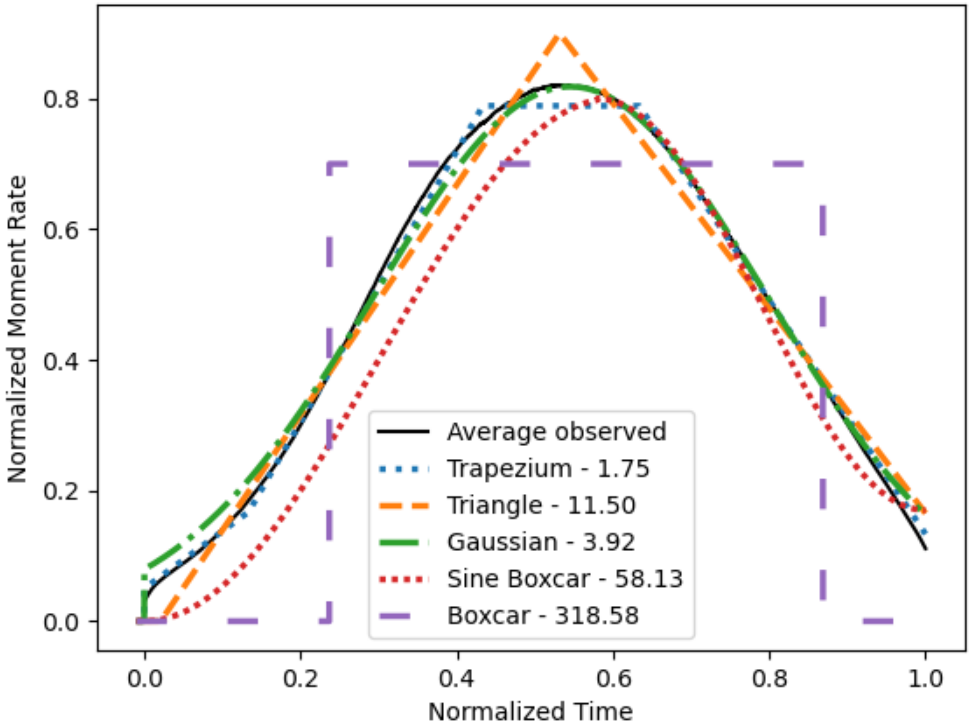


Figure 4.22: The median STF (black) calculated from all magnitudes and datasets and the optimal fit of the five geometric models. Numbers after labels in the legend are the L2-Norm for each geometric fit, see Section 4.4.7.1 for additional discussion of this metric.

Figure 4.22 shows the optimized fit of each of the 5 geometric models (colours) and the overall median STF (black). From this, it is clear that it is possible to reach a good fit using a ramp plus trapezium model, or a Gaussian (purple, red and blue lines). Since our definition of the 'end' of the STF does not require the moment-rate to reach zero, we allow the moment-rate at the end of the model to vary, along with the amplitude, centre and any other points required to define the shape. Since the different geometries have different degrees of freedom, it is improper to directly compare the L2-norms which result. However, since the degrees of freedom for all models are 5 ± 1, an order of magnitude difference in the L2-norm is likely significant, and we use this as part of interpreting the best fit models.

In Figures B.39–B.44, we fit our geometric models to the median STF calculated from each dataset individually. The different numbers of events in each dataset means that the overall average STF is biased by the SCARDEC STFs. We find that the trapezium model provides the best fit to all the datasets, but that there is variation in the parameters which describe the trapezium, hence why a gaussian provides a similar or better fit to the complete dataset. Some datasets require a ramp to provide a good fit. The Ye (Figure B.42) and optimal SCARDEC data (Figure B.39) are best fit with a simple trapezium. The USGS data (Figure B.41) requires the trapezium to begin with a slight delay of 0.03 of the duration. The average SCARDEC (Figure B.40) and Sigloch (Figure B.43) datasets both require a significant ramp, with a lower gradient than the left-hand leg of the main trapezium. The ISC dataset (Figure B.44) is an outlier once again, with the ramp having a much steeper gradient than the left-hand leg. This geometry would perhaps more accurately be called an irregular convex pentagon.

For the non-ISC datasets, we can divide the trapezium fit into three stages (Figures B.39– B.43). The earthquake begins with a period of moment-acceleration, parametrised by the ramp (if present) and left-hand leg of the trapezium. Where a ramp is present, this suggests that the earthquake growth began slowly, for roughly the first 20% of the duration, before accelerating to the maximum moment rate. The earthquake then enters a phase of constant moment-rate, lasting anywhere from 13%

of the duration (Ye) to 28% (USGS). The optimal SCARDEC STF spends more time in this phase than the average SCARDEC STF. Finally the moment-rate reduces, and the earthquake decelerates to an end. This deceleration phase begins at 60–70% of the total duration, except for the Ye data where it begins at 50% duration.

The ISC data are poorly fit by the geometric models. The median STF has two clear peaks. The moment-rate initially increases rapidly. It then undergoes a period of decline, before increasing again, but more gradually. As with the Ye dataset, the final deceleration of the earthquake begins at about 50% of the total duration.

4.4.7.2 Effect of magnitude on STF geometry

The optimal trapezium fit for different magnitude bins have no clear trends between magnitude and either the duration of the ramp (Figure B.45) or the duration of the top (constant moment rate) portion (Figure B.46). We also compare the gradient of the ramp with the gradient of the left leg of the trapezium, and find no systematic relationship with magnitude (Figure B.47). We do not constrain the ramp to have a gradient lower than the left leg of the trapezium. The gradient of the 'ramp' of the ISC data is generally very high, and thus the left leg of the trapezium essentially forms a low moment-acceleration sector before the constant moment-rate portion. For all the other datasets there are no clear trends with magnitude of variation in either of the two gradients, although some datasets tend to be closer to the 1:1 line than others (in which case the ramp and left leg are essentially forming part of the same moment-acceleration phase).

Summary

- Using the complete dataset, Gaussian model and trapezium model both provide a good fit to the median STF.
- Considering each dataset individually, the trapezium model gives the best fit. This suggests there are three stages to the STF and earthquake rupture process: a beginning characterised by rapidly increasing moment-rate, a constant moment-rate middle and the end characterised by moment-rate decrease

- ISC STFs are not well fit by any of the proposed models.
- The best geometric fit has no clear trends with magnitude.

4.4.8 How are increments of moment released throughout an earthquake?

In Section 4.4.7.1, we identified that the median STFs are best fit by a trapezium. This suggests that there may be different phases to rupture, with an early acceleration, a period of relatively constant moment release rate, and then a deceleration. In this section, we further investigate this by looking at the rate of moment release in different windows of the STF (Figure 4.23).

There are three clear phases of moment release visible with the complete dataset in Figure 4.23 (black line). The first 10% of moment is released slowly, taking 20–30% of the total earthquake duration. Each additional 10% increment of moment takes less than 10% of the total duration, except for the final 10% which takes $\sim$ 15% of the duration. Within the central 80% of moment, the 40–50% increment of moment is the fastest to release, with a gentle increase in time taken for the increments on either side. The coloured lines show the average STF for each dataset, with shading showing the variation around this average. The grey lines show the fit from each of the geometric models shown in Figure 4.22.

All the models overestimate how long it takes to release moment deciles between 10 and 30%. Between 40–60% moment, the trapezium model (solid grey line) provides the best fit for the data, as we expect from fitting the average STF. The trapezium and Gaussian (dotted line) models are both able to fit the final 30% of moment release well. The triangle model (dashed line) appears to be a perfect fit at estimating how long it takes to release the 40–50% and 50-60% deciles. This is because the model equally over and underestimates the time taken to release smaller increments of moment within each 10% bin.

It is interesting that the models provide a better fit to the later portions of moment-release than the early stages. Geometrically, the final 30% of the STF can be fit well by a straight line, whilst there is more variation in the moment-acceleration in the early stages.

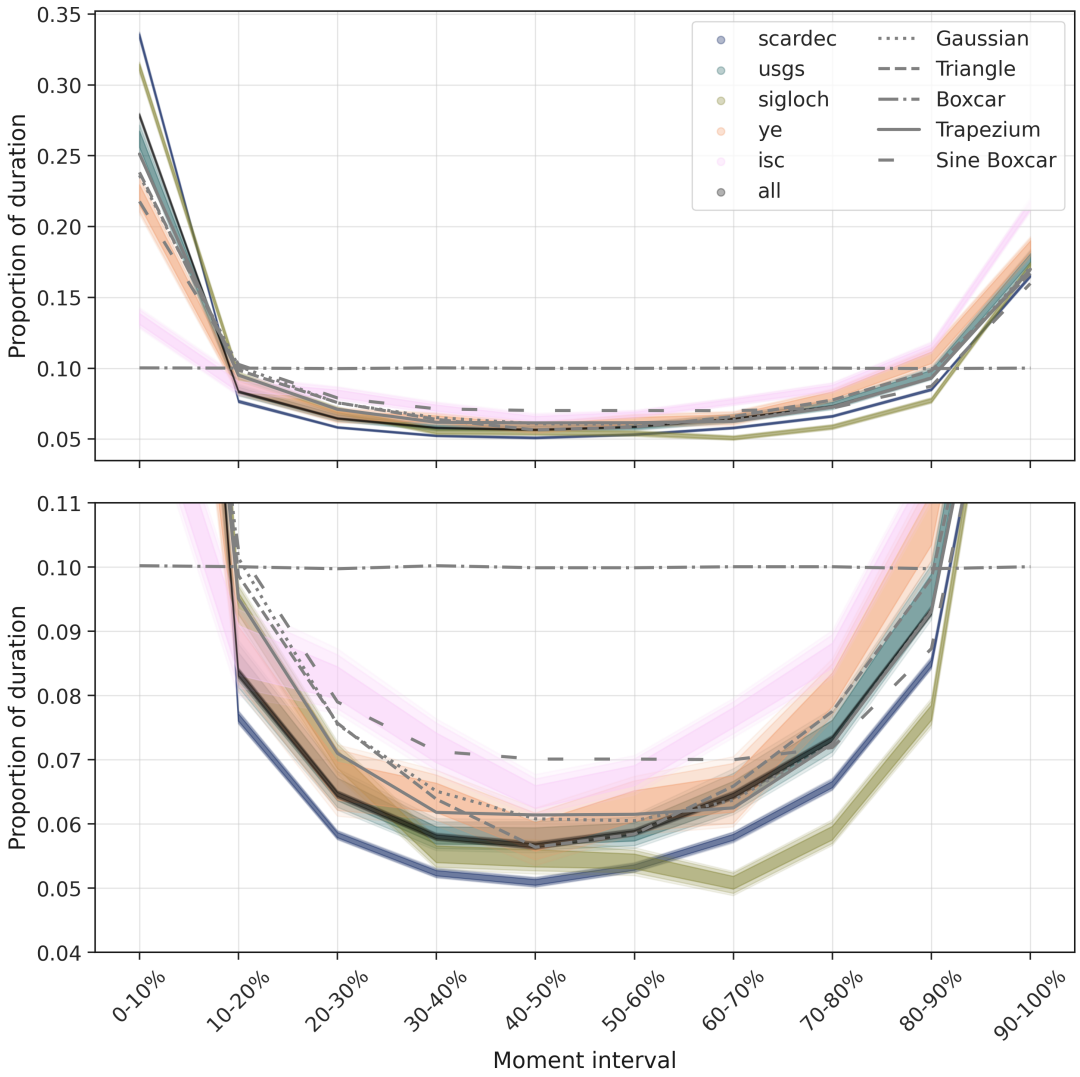


Figure 4.23: The proportion of duration taken to release 10% increments of moment. In black, the median result using all STFs. Different datasets are shown in different colours, whilst different geometric models are shown in different styles of grey line. Bottom panel is a zoom of the top panel.

The time taken for the final 10–20% of moment released is most sensitive to the definition of the end which is chosen. With more conservative definitions (for example definitions 1 and 2 proposed in Section 4.3), these portions take an increased proportion of the duration to release.

Summary

- There are three clear phases of moment release: start, middle and end.
- The first 10% of moment takes 20–30% of the total duration to be released.
- The central 80% of moment takes $\sim 60\%$ of the duration to be released.
- The final 10% of moment takes $\sim 15\%$ of the duration to be released. This is most sensitive to the definition of the end used (later ends mean this segment takes proportionally longer).

4.5 Synthesising our observations to gain a physical insight

We have made three key observations, which we now attempt to provide some physical reasoning for.

4.5.1 STFs are complex and poorly fit with a single Gaussian

Looking at source time functions individually, it is obvious that many STFs are complex. Our observations of the proportion of events which are well fit by a single Gaussian (46% at 5% smoothing, 67% at 20% smoothing) suggests that a single sub-event is not a good fit for a significant number of earthquakes, and this will also lead to variation in the timing of peak moment-rate. This has important implications for the rupture models, where the STFs used often have one major peak. This would imply that earthquakes do not generally wane and then re-accelerate if, for example, they are propagating through a low stress region and then rupture into a high-stress region which 're-invigorates' them. However, our observations show that a significant proportion of earthquakes do decelerate and then re-accelerate. Different datasets show different relationships between STF

complexity and earthquake magnitude and depth. We agree with Danré et al. (2019) that SCARDEC STFs show increasing complexity with magnitude, despite using a subtly different definition of complexity. However, we caution that this trend does not apply to all datasets. We also agree with Yin et al. (2021) that shallow SCARDEC STFs tend to be more complex, though this trend is minor (but statistically significant). In particular, deep earthquakes are significantly less likely to be complex (Figures 4.20 and 4.21 and Houston et al., 1998; Houston, 2001; Persh and Houston, 2004; Tocheport et al., 2007). Unlike Houston et al. (1998), we do not see a clear group of more complex STFs at intermediate depths.

Another contributor to variation in the peak moment rate is uncertainty in the duration estimates of the STF durations (Meier et al., 2017). In a median sense, this is likely a key factor as a 10% variation in the duration could lead to up to a 5% variation in the timing of the peak. However, the spread between individual STFs visible in Figures 4.8 and B.11–B.16 suggest that there is also intrinsic variation in the timing of the peak moment rate between earthquakes. We see no systematic variation in the timing of the maximum moment rate with depth or magnitude (Figures 4.17 and 4.18). This is in contrast with Houston et al. (1998) who observed that deeper events tend to release their moment later. Whilst Houston et al. (1998) use a subtly different metric, we also do not see patterns in skew (or kurtosis) with depth and thus these measures are likely correlated.

4.5.2 Early STF characteristics vary with magnitude

We identify scalings between the final event magnitude and three features: the initial moment-acceleration (Sections 4.4.1 and 4.4.8), the moment released in the first seconds of the STF (Section 4.4.3.1), and the time taken to release increments of moment that are less than the final moment (Section 4.4.3.2). We find that in a statistical sense, larger magnitude events release their early moment more quickly than small events. This is in contrast with Meier et al. (2020a) whilst still avoiding the interpolation that impacts the results of Melgar and Hayes (2019).

Our observation that the early STF features vary with magnitude aligns with the observations in Chapter 3, where we see a correlation between peak ground displacement and final moment/magnitude. One way to physically conceptualise this, is that earthquakes that increase their rate of moment release more quickly at the beginning of rupture are better able to overcome stress or frictional barriers, and therefore become higher-magnitude earthquakes. This is similar to the ideas of Heaton (1990) and Nielsen (2007)

We see more variation in the early portion of the STF, and this is more difficult to fit geometrically. This may be a result of the early moment-acceleration relying on overcoming energy and stress barriers and the rupture jumping from patch-to-patch. The final deceleration of the earthquake may be more simple, as the earthquake slowly fades.

4.5.3 Median STFs are best fit with a trapezium model

We have identified that a trapezium geometric model provides the best overall fit to our observed median STF, and this suggests that there are three main phases of rupture. The earthquake begins with a period of moment-acceleration, parametrised by the ramp (if present) and left-hand leg of the trapezium. The earthquake then enters a phase of constant moment-rate, lasting anywhere from 13% of the duration (Ye) to 28% (USGS). Finally the moment-rate reduces, and the earthquake decelerates to an end. This deceleration phase begins at 60–70% of the total duration, except for the Ye data where it begins at 50% duration. Our results contrast with Meier et al. (2017) who found that the median STF shape is near triangular. We do consider much smaller events than they did, but we see no systematic trend in the shape of the trapezium with earthquake magnitude and so this is unlikely to explain the discrepancy. We also find a significant positive skew whereas Meier et al. (2017) suggest that the median STFs are near-symmetric.

4.5.3.1 Early moment-acceleration

The early phase of moment-acceleration sometimes occurs in two phases (i.e. a 'ramp' is required) and sometimes only one. This varies across the datasets.

That we find a low-moment acceleration portion at the start of rupture is not unexpected, and a number of explanations have been proposed. Ellsworth and Beroza (1995) and Beroza and Ellsworth (1996) proposed a 'nucleation phase', where the moment-rate grows more slowly. They expected the duration of this to scale with the total duration of the event (or equivalently with $M_0^{1/3}$), which would mean we would expect a roughly constant duration for the ramp in a scaled STF.

Modelling the case that an earthquake nucleates close to the edge of the stressed region, Ke et al. (2022) observe an initial quadratic phase of growth, as the circular crack grows unbounded, before a transition to linear growth of moment-rate with time, as the growth becomes bounded in at least one dimension. The time of the transition from quadratic to linear growth gives information on how much the rupture can grow before it becomes bounded. In this case, we do not expect the duration of the bounded-growth portion to scale with magnitude, because the transition to bounded growth is defined by the width of the seismogenic zone and the timing of this is controlled by the rupture velocity, which shows no systematic trends with magnitude (e.g. Table S1 of Chounet et al., 2018). For small (M3–5.8) earthquakes, STFs can modelled as quadratic until the peak, and linear after this time (Archuleta and Ji, 2016), as the rupture is unbounded for the whole rupture. This quadratic growth until the peak, followed by linear decay of moment-rate is consistent with an asperity breaking and propagating into a region of low stress drop (Boatwright, 1988; Archuleta and Ji, 2016). Looking only at the first 20% of the median STF, we can fit a quadratic curve (L2-norm = 0.12). If we consider the 20–35% portion of the median STF, there is a clear linear trend (L2-norm = 0.067), as we discovered in our efforts to fit a trapezium geometry. For $M \geq 7$ earthquakes, the STF would be dominated by the long central periods of high moment-rate, where the rupture growth is likely to be bounded. This explains the linear moment-rate growth observed by Meier et al. (2017), and explains why our trapezium-based fits

of the USGS and Ye datasets, which only contain higher magnitude earthquakes, do not require any significant 'ramp' (length of ramp is 3-6% total duration).

We do not see a clear relationship between the scaled duration of the ramp portion and the final magnitude of the earthquake, suggesting the duration of the lower moment acceleration phase may scale with magnitude. However, there is a trade off between ramp duration and ramp gradient (with the ramp gradient sometimes being very close to the left-leg gradient), making it hard to make concrete conclusions.

4.5.3.2 Constant moment rate

The physical interpretation of this phase depends on if we consider a crack-like or pulse-like rupture. In the pulse-like case, we can consider that the rupture area is constant during this phase, and each patch is slipping by a similar amount. On average the moment-rate is constant because the increases and decreases in moment-rate through time cancel each other out. In the crack-like case, we can suggest that perhaps rupture is already beginning to die – the slipping area is still growing, but the slip rate has already begun to decrease. Meier et al. (2017) do not observe a comparable phase.

4.5.4 Limitations and Future Work

In Section 4.2.7, we outlined some of the advantages and disadvantages of the different methods to invert for STFs. We now return to these in light of our analyses, and look at the impact on our confidence in our results.

A key limitation of our work remains exactly how well STFs constrain the early portions of rupture. That we see scaling relationships between the early portions of STFs and magnitude for all datasets lends credence to the STFs ability to resolve the early features of rupture. However, we might have expected the USGS STFs to be better able to resolve the early STF features, since they use geodetic and local seismic data-sources in addition to teleseismic body-waves. That we see weaker relationships between early features of the USGS STFs and magnitude could suggest that STFs are not particularly good at resolving early features of rupture.

Alternatively, this may suggest that early features of rupture are less apparent in the relatively low frequency waveforms used in the USGS catalogue. However, it is hard to disentangle these ideas from the fact that the USGS dataset is primarily composed of larger magnitude events, and that the STFs tend to be longer in duration (compare Figure B.19 between datasets: comparable magnitudes of other datasets have at least some earthquakes which have released their moment but USGS does not). We expect the trends in median STF with magnitude to be particularly robust for the SCARDEC and Sigloch datasets, where we have a large number of events spanning moderate and large magnitude events. As Meier et al. (2017) said, we still need more work on how well STFs resolve the beginning of earthquakes.

In outlining each method, we posed several possible biases in STF complexity. We thought that SCARDEC STFs might be biased at moderate magnitudes to be more simple, that Sigloch STFs would be simple compared to other methods and that ISC STFs would be more complex. We found that these predictions were generally correct. SCARDEC STFs tend to be slightly simpler than USGS STFs. There is little cross-over between the USGS dataset and either the ISC or Sigloch datasets, but we find that the ISC STFs are generally much more complex than the SCARDEC STFs and the SCARDEC STFs are in turn more complex than the Sigloch STFs. Given that the USGS complexity is perhaps the most reliable, since it results from a realistic representation of the fault and has no obvious biases, it makes sense to use this as a benchmark. Crucially, our results highlight the importance in not making direct absolute comparisons between complexity found from different STF datasets. Whilst we find that events which are more complex in one dataset will tend to also be more complex in another, there is a high degree of scatter in this relationship. Future work could further investigate the drivers of variation in STF complexity across different datasets.

The ISC inversion returns an ensemble of potential earthquake point source models, from which the average of the likely solutions and a measure of the reliability of the result are derived. This means that in the future, further investigation could be done on the reliability of the results we outline here. In particular, since the

ISC invert jointly for depth and STF complexity, further analysis of patterns in STF complexity could be undertaken.

Future work could investigate biases in the STFs available in each dataset, to understand to what extent patterns in the STFs might represent global trends. In particular, we might seek to understand regional biases in the USGS STFs thanks to different weightings of datasets, and whether the SCARDEC STFs actually show a bias towards simpler events at moderate magnitudes.

Finally, we should seek to understand the impact of the averaging implicit in some STF inversion methods on our results. One way to approach this would be to use synthetic events in each of the inversion methods to see any biases which result.

4.6 Conclusion

In this study we compiled a dataset of source time functions of over 5000 $M \geq 5$ earthquakes from around the world, from five data-sources. We investigated how STF features scale with the final earthquake magnitude.

We have a number of key findings, but of most interest to our understanding of determinism is the fact that larger earthquakes release more moment even in short windows of time. In making this observation, we use all available STFs from the five catalogues, with no selection based on duration.

We see that a significant proportion of STFs are not well fit by a single peak, although we observe no significant scaling between earthquake magnitude and STF complexity. This is important for earthquake modelling efforts as it suggests that events of all magnitudes can wane during the course of the event and then re-accelerate.

For some events and datasets we see evidence of an early low moment-rate sector. The presence of this feature varies between datasets, heeding caution to any concrete interpretation of a nucleation phase.

4.7 Open Research Section

Processing was primarily in Python 3.10 and was undertaken using Scipy (Virtanen et al., 2020), Numpy (Harris et al., 2020), and pandas (pandas development team, 2024; McKinney, 2010). Plotting used Matplotlib (Hunter, 2007) and Crameri colormaps (Crameri, 2023; Rollo, 2024). SCARDEC STFs are available via http://scardec.projects.sismo.ipgp.fr. USGS STFs are available via https://earthquake.usgs.gov/data/finitefault/. ISC STFs are available via https://www.isc.ac.uk/iscbulletin/search/catalogue/. Ye STFs are available via email to Lingling Ye (SUSTech). Sigloch STFs are not routinely available. Contact Karin Sigloch (GeoAzur) for more information. Processing and analysis code is available on Github (https://github.com/RebeccaColquhoun/stf).

5

Using foreshocks to constrain earthquake nucleation: comparing foreshock to aftershock ratios in the Hikurangi subduction zone to results from ETAS

Contents

Key Points

- We make new observations of foreshocks and aftershocks using phase coherence.
- An Epidemic-Type Aftershock Sequence model predicts foreshock and aftershock numbers based on earthquake-earthquake triggering.
- Our observed ratio of foreshocks-to-aftershocks is in line with expectations from ETAS, but we observe higher absolute numbers of foreshocks and aftershocks per mainshock.
- We see an exponential increase in foreshocks and decrease in aftershocks, which is not apparent in the ETAS model results.
- Therefore, earthquake nucleation requires a combination of stochastic triggering and external process mediation.

Abstract

The complexity and duration of earthquake nucleation is an open question. Our understanding of this process is limited by a lack of high-quality observations of foreshocks and aftershocks, and we therefore apply a coherence-based template matching approach to search for more foreshocks and aftershocks. We examine the hour around thousands of $M \geq 3$ mainshocks on the Hikurangi subduction zone in New Zealand, and make new detections of foreshocks and aftershocks which are close in space and time to the mainshocks. The foreshock-to-aftershock ratio we observe is inline with that expected from Epidemic-Type Aftershock Sequence models, but we observe higher numbers of foreshocks and aftershocks per mainshock than expected from ETAS. This suggests that foreshock occurrence is more complex

and that stochastic triggering occurs in an external process mediated environment such as through changes in pore-fluid pressure or aseismic slip.

Plain language summary

We do not fully understand how earthquakes start, especially the physics that determines whether earthquakes are preceded by smaller earthquakes (foreshocks). Some models suggest that earthquakes cause other earthquakes through transfer of stress (related to force). To investigate this, we used a new method to search for foreshocks and aftershocks around earthquakes in New Zealand's Hikurangi subduction zone. We found thousands of foreshocks and aftershocks occurring close in space and time to our main earthquakes. We calculated the ratio of foreshocks to aftershocks around our earthquakes and found this is in line with the models based on earthquakes causing other earthquakes. However, each large earthquake had more foreshocks and aftershocks than these models suggest. This suggests that the occurrence of foreshocks is more complicated than previously thought, and that additional factors like fluid movement or slow slip along the fault are needed to explain foreshock occurrence.

5.1 Introduction

There have been varied observations of foreshocks before earthquakes. For example, Trugman and Ross (2019) suggest that 72% of earthquakes in Southern California are preceded by increases in earthquake activity. Other authors find the foreshock rate to be around 40-55% (Jones and Molnar, 1976; Abercrombie and Mori, 1996; Chen and Shearer, 2016), and van den Ende and Ampuero (2020) find that as few as 18% of earthquakes in southern California have increases in seismicity which cannot simply be explained by variations in background seismicity rate. Significant foreshock activity is expected from rate and state models (Dieterich, 1994; Marone, 1998) and is observed on short time-scales in laboratory experiments (e.g. Scholz, 1968; Johnson

et al., 2013; Goebel et al., 2013; McLaskey and Kilgore, 2013; McLaskey and Lockner, 2014; Rouet-Leduc et al., 2017; Bolton et al., 2019, 2020; Dresen et al., 2020).

A number of researchers have investigated foreshocks because these events can provide insights into earthquake nucleation. Some researchers have found that nucleation could be short and simple and that all clustering of earthquakes can be explained by earthquake-earthquake triggering (Helmstetter and Sornette, 2003; Felzer et al., 2004; Yoon et al., 2019). For example, Felzer et al. (2004) were able to explain foreshock occurrences solely through the same earthquake-earthquake triggering which is responsible for aftershocks. Ellsworth and Bulut (2018) suggest that no aseismic processes were involved in the foreshock sequence preceding the 1999 İzmit earthquake (cf. Bouchon et al., 2011).

On the other hand, nucleation may be a complex, extended process that lasts minutes to months, perhaps driven by slowly accelerating aseismic slip or by changes in pore-fluid pressure. Seismologists have observed significant increases in seismicity in the hours to days before large earthquakes in Southern California (Dodge et al., 1996; Chen and Shearer, 2016), in the North Pacific (Bouchon et al., 2013), in Kyushu, SW Japan (Kato et al., 2016), and on the East Pacific Rise (McGuire et al., 2005). Some statistical analyses suggest that these foreshocks, mainshocks, and aftershocks nucleate or are triggered via different processes (Brodsky, 2011; Shearer, 2012; Seif et al., 2018). Some of these processes could result in complex, long-duration nucleation, as is sometimes inferred from observations. For instance, Tape et al. (2018) identified foreshock sequences before earthquakes in Alaska that lasted tens of seconds. These sequences may represent earthquake nucleation triggered by a propagating aseismic front. Similarly, Bouchon et al. (2011) identified an extended seismic signal before the 1999 İzmit earthquake. Complex, two-stage, nucleation has also been observed in laboratory studies (Latour et al., 2013; Harbord et al., 2017) and over very short durations at the beginning of earthquakes in California (Ellsworth and Beroza, 1995; Beroza and Ellsworth, 1996). Complex nucleation could last months, or occur intensely for minutes, and may result from aseismic

nucleation processes (Dodge et al., 1996) or from the interaction of pore-fluid pressure changes on the accelerating fault (Liu and Rice, 2007).

It has been difficult to constrain the process of earthquake nucleation due to the limited availability of high-quality observations of foreshocks and aftershocks, particularly on short timescales. In this work, we seek to add one more observation of foreshock rates. We look for foreshocks, aftershocks, and sequences of foreshocks around thousands of earthquakes in New Zealand. We use a phase coherence-based technique to detect small events that are located spatially close to, and within an hour-long window centred on each mainshock. We then compare the observed foreshock:aftershock ratio to the ratio expected from ETAS models.

5.1.1 ETAS models

The Epidemic-Type Aftershock Sequence (ETAS) model is widely regarded as one of the most reliable tools for forecasting earthquakes. ETAS models (Ogata, 1988) often model foreshock and aftershock behaviour well, using relatively simple earthquake-earthquake triggering. They are based on the concept of self-exciting point processes (introduced by Hawkes, 1971), where earthquakes above a specific magnitude threshold (M_0) can trigger subsequent earthquakes, often referred to as 'offspring' or 'daughters' (Molkenthin et al., 2022). These triggered events can, in turn, trigger further events, creating a cascading, branching pattern that effectively describes how seismicity clusters over time and space (Reinhart, 2018).

The ETAS model is represented by a conditional intensity function

$$\lambda(t \mid \mathcal{H}_t) = \mu + \sum_{(t_i,m_i)\in\mathcal{H}_t} K e^{\alpha(m_i - M_0)} \left(\frac{t - t_i}{c} + 1\right)^{-p}, \tag{5.1}$$

where λ represents the expected number of earthquakes at a given time t, conditioned on the history of all previous events up to that point denoted as $\mathcal{H}_t$. This history includes all earlier earthquakes represented as (t_i,m_i), where t_i is the time and m_i is the magnitude of the i^{th} earthquake.

The ETAS model relies on five key parameters (μ, K, α, c and p) which are estimated during the modelling process. The background seismicity rate, μ, accounts

for independent, spontaneous earthquakes that are not triggered by previous events. The baseline productivity parameter, K, determines the average number of direct aftershocks triggered by an earthquake with a reference magnitude M_0. The magnitude-scaling productivity parameter, α, describes how the productivity of aftershocks increases with the magnitude of the triggering earthquake. A short term temporal offset, c, prevents singularities in the model by accounting for delays immediately following an earthquake. Finally, p is the Omori-law decay exponent, which characterises how the rate of aftershocks decreases over time.

Physical explanations for earthquake-earthquake triggering

The earthquake-earthquake triggering inherent in ETAS models is purely statistical and does not require any physical process to underpin it. It simply arises from every earthquake changing the probability of future events occurring, and thus earthquake occurrence being related to the history, $\mathcal{H}_t$, of past events. However, there are several physical processes that might underpin an observed cascade of earthquakes, such as static stress transfer or aseismic afterslip. Some, but not all, observations and models require that aftershock occurrence and deep afterslip follow the same temporal evolution (e.g. Perfettini and Avouac, 2004).

An integral part of ETAS is Omori's Law, which is an empirically derived relationship that describes the temporal decay of aftershock activity as $N \sim t^{-1}$ (Utsu et al., 1995). Theoretical interpretations of this relationship have since been derived, including models based on rate-and-state friction (e.g. Dieterich, 1994).

5.2 Mainshock and data selection

We begin by identifying a collection of mainshocks. We will search for foreshocks and aftershocks around these events.

5.2.1 Seismic data and earthquake catalogue

We investigate earthquakes on the Hikurangi subduction zone beneath the North Island of New Zealand (Figure 5.1) which is an ideal place to test whether external

processes are involved in earthquake nucleation. The region hosts significant slow slip events ($\approx M_W 7$) as well as spatially variable pore-fluid pressures (Wallace et al., 2012, 2016; Naif and Key, 2018): two phenomena that could encourage extended earthquake nucleation.

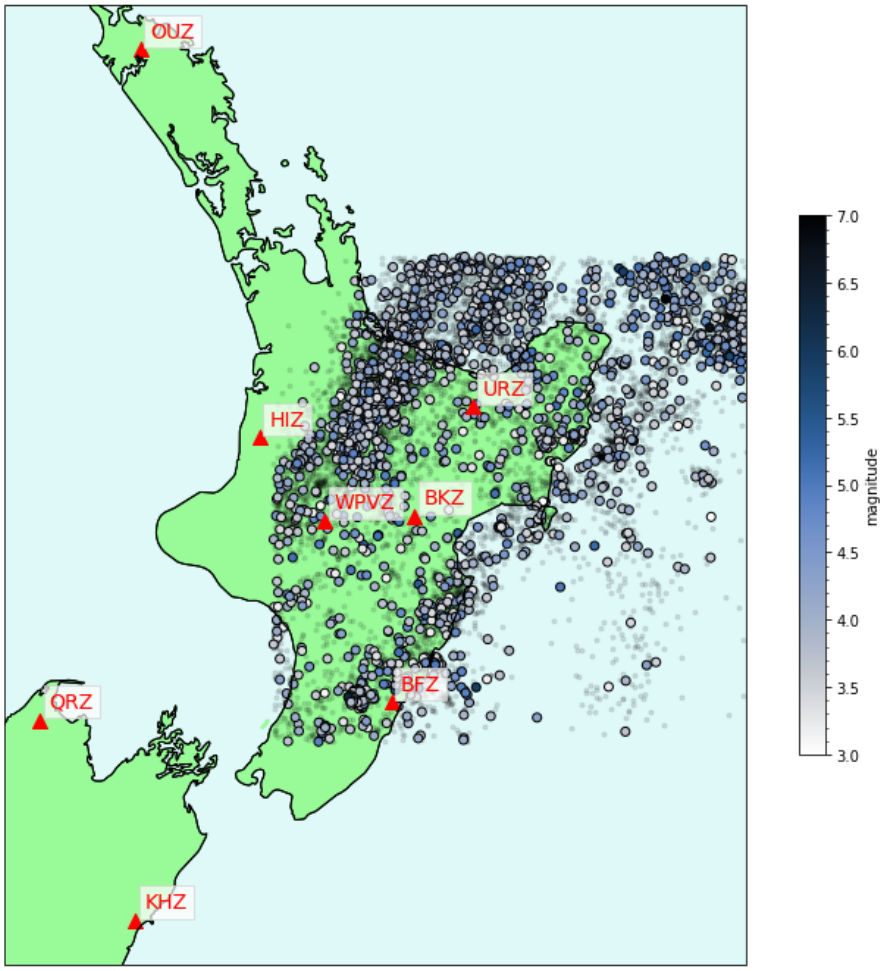


Figure 5.1: Map of New Zealand with earthquakes and stations used. Grey dots represent all earthquakes, whilst coloured dots represent mainshocks considered under isolation at 24 hours (colour = magnitude). Stations are all part of the GNS network (FDSN network code NZ) and are labelled with their station codes.

We use events listed in the GNS catalogue: the catalogue created by New Zealand's Te Pū Ao, or Institute of Geological and Nuclear Sciences. We gather seismograms from GNS seismic stations stored in the IRIS data centre, using the obspyDMT software (Hosseini and Sigloch, 2017). Only stations within 4° of an earthquake's epicentre are included.

5.2.2 Identifying mainshocks

The initial earthquake catalogue consists of 12,769 $M \geq 3$ events between 2005-01-01 and 2020-01-01, in a region between 175° and 180°W and 37° and 40°S, with no constraint on focal mechanism. Of these earthquakes, 11,236 have usable data: there is at least one station where all 3 components have data for an hour before and after these earthquakes.

Each of the identified $M \geq 3$ earthquakes is a candidate mainshock. However, even $M \geq 3$ earthquakes are clustered, and we want to consider our mainshock earthquakes independently, without considering the effects of nearby seismicity. We also want the background seismicity to be roughly constant throughout the 30 minutes before and after each of our mainshocks. Therefore, we do not want our mainshocks to be too close in time to previous large earthquakes, which often trigger a rapidly decaying seismicity rate and so we accept an earthquake to be a mainshock only if it is larger than all other earthquakes within a certain time interval and within a distance radius of 0.3°. All results presented consider the largest event within 24 hours (giving $n = 1947$).

5.3 Detection Methods

Once we have identified our mainshocks, we can search for foreshocks and aftershocks.

5.3.1 Mainshock templates

To begin, we identify and extract each mainshock's P waves to use as templates. We make a preliminary estimation of the P wave pick by calculating the travel time of the earthquake waves to each station using obspy TauP (Crotwell et al., 1999) and the AK135 model (Kennett et al., 1995). This calculation predicts the P pick to within ±10 s. To improve the pick accuracy, we then apply a STA-LTA algorithm to a high-passed version of the data in this 20 second window, using a corner frequency of 1.5 Hz. The earliest STA-LTA trigger (Withers et al., 1998) on any component is taken as the P wave arrival. After we identify the arrival, we

apply a bandpass filter at 1.5 to 10 Hz and extract an interval from 1 second before to 2 seconds after the P arrival pick. This part of the seismogram is our template.

5.3.2 Phase coherence calculation: Theory

We use this template to search for earthquakes with similar Green's functions within a one-hour window of the continuous data, 30 minutes either side of the mainshock. We use the phase coherence method outlined by Hawthorne and Ampuero (2017) and summarised in Figure 5.2, which identifies co-located seismic sources by comparing the seismograms recorded at multiple stations or on multiple components at a single station.

Consider two co-located earthquakes: one recorded in the continuous data and one recorded in the template data. Their seismograms are a convolution of the source-time function, s, and the Green's function, g. Thus their seismograms at station or component k can be written as $d_{ck} = s_c * g_k$ and $d_{tk} = s_t * g_k$ respectively. The Green's function, g_k, is common between them as they are co-located and measured at the same station.

To calculate the phase coherence, we first cross-correlate the two signals measured at station k,

$$\hat{x}_k = (\hat{s}_c \hat{g}_k)\,(\hat{s}_t \hat{g}_k)^*\,, \tag{5.2}$$

where $\hat{s}_t$ is the Fourier transform of s_t and * denotes the complex conjugate.

The co-located earthquakes were also measured at another station, l. The cross-correlation of both waveforms at that station is $\hat{x}_l = (\hat{s}_c \hat{g}_l)\,(\hat{s}_t \hat{g}_l)^*$.

We can now cross-correlate between these two stations,

$$C_p = \mathrm{Re}\left[\frac{((\hat{s}_c \hat{g}_k)\,(\hat{s}_t \hat{g}_k)^*)\,((\hat{s}_c \hat{g}_l)\,(\hat{s}_t \hat{g}_l)^*)^*}{\left|((\hat{s}_c \hat{g}_k)(\hat{s}_t \hat{g}_k)^*)\,((\hat{s}_c \hat{g}_l)\,(\hat{s}_t \hat{g}_l)^*)^*\right|}\right], \tag{5.3}$$

and if the two sources were indeed co-located, then the normalised value of this inter-station cross correlation would be 1. As we discuss in Section 5.3.3, it is also possible to cross-correlate the signals recorded on different components at

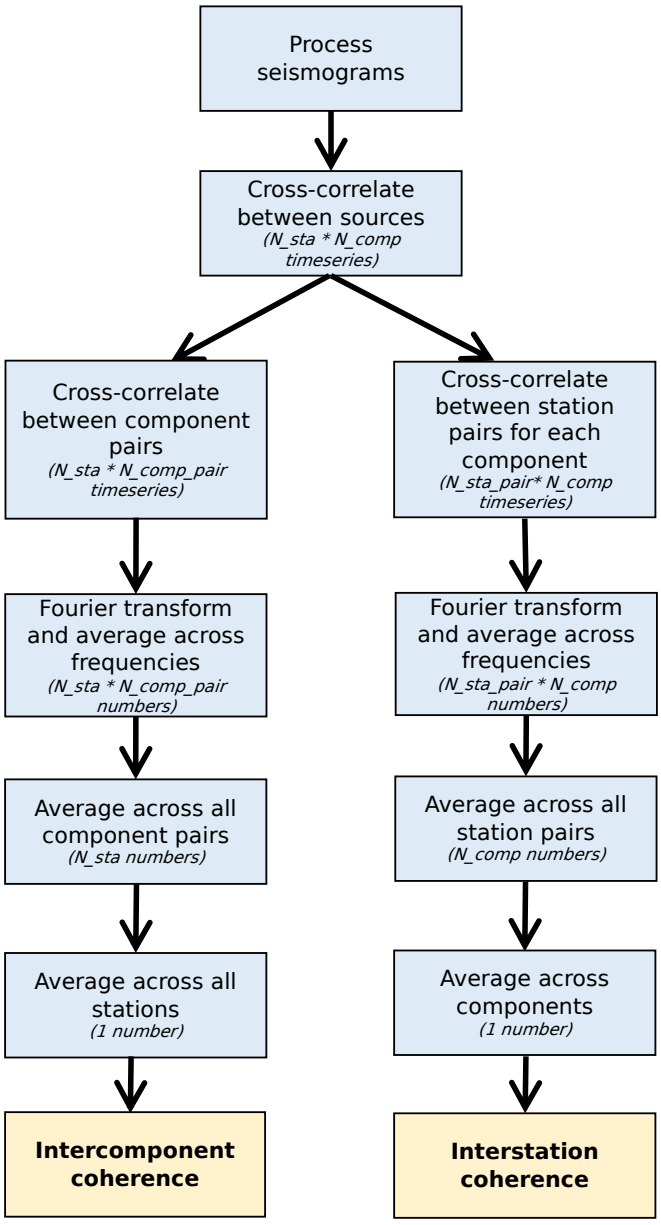


Figure 5.2: A flowchart illustrating the phase coherence calculation. In italics, the number of pieces of data (timeseries or single numbers) after each step. N_sta represents the number of stations, N_comp, the number of components, N_sta_pair the number of station pairs, and N_comp_pair the number of component pairs.

a single station, rather than between two different stations. We call this inter-component phase coherence.

When we calculate the phase coherence

$$C_p = \mathrm{Re}\left[\frac{\hat{x}_k \hat{x}_l^*}{|\hat{x}_k \hat{x}_l^*|}\right], \tag{5.4}$$

we calculate it at each frequency between 1.5 and 10 Hz. We then average C_p over our frequency range of interest,

$$C_p = \mathrm{Re}\left(\frac{1}{\sum_{m=1}^{m=N} w_m} \sum_{m=1}^{m=N} w_m \left[\frac{\hat{x}_{1m}^* \hat{x}_{2m}}{|\hat{x}_{1m}^* \hat{x}_{2m}|}\right]\right). \tag{5.5}$$

The weighting w at frequency m is 1 when m is in the centre of the frequency interval and tapered to zero with a cosine taper at the edges of the frequency interval.

As in the case of the co-located earthquakes we discussed previously, the C_p value should be high if the continuous data contains a source co-located with the template. In this case, the inter-source correlations implicit in the $\hat{x}_k = \hat{d}_{ck}\hat{d}_{tk}^*$ calculations turn out to eliminate the phases of the Green's function. The subsequent inter-station or inter-component correlations are denoted by $\hat{x}_k \hat{x}_l^*$ and are a frequency domain version of an unnormalised cross-correlation in the time domain, which eliminates the phases of the source time functions. So if the continuous data segment contains a source co-located with the template, all the phases that result from inter-source cross-correlation (x_k and x_l) are eliminated. Both the numerator and denominator should be real and positive, and C_p should equal 1.

When looking at real earthquakes, of course, C_p never reaches 1 because there is always some noise, and filtering modifies equation 5.3 so it is no longer exact. The data are also modified when windows of the seismograms are extracted for calculation. We therefore search for significantly positive values of C_p, and we follow the windowing and tapering approach used by Hawthorne and Ampuero (2017) to mitigate the effects of truncation. Specifically, we first cross-correlate our templates with the entire continuous data record. We then extract 1-second windows of the data and calculate $\hat{x}$.

This phase coherence method allows us to search for a variety of signal types. It can identify nearby seismic sources even if they have complex, extended source time functions. We can detect foreshocks with source time functions similar to the mainshock, foreshocks with shorter source time functions, and any tremor-like foreshock sequences.

5.3.3 Phase Coherence Types and Detections

We compute two types of phase coherence (Equation 5.5): C_{p-stat}, the inter-station phase coherence; and C_{p-comp}, the inter-component phase coherence. We show the difference in the two types schematically in Figure 5.2. We calculate both in 1-second windows, separated by 0.2 s, for 1800 s before and after each mainshock and plot the results in Figure 5.3.

We cross-correlate the signal from different stations to find the inter-station coherence (C_{p-stat})—k and l index different stations in equation 5.5—and we then average across station pairs. Note that inter-station coherence can only detect earthquakes which are within a fraction of a seismic wavelength of the mainshock (Geller and Mueller, 1980; Gombert and Hawthorne, 2023). Any shift in the earthquake location shifts the station arrival times. The time shifts make the Green's functions (g in equation 5.3) appear different between the mainshock and foreshock and thus reduce the phase coherence between the two signals.

Inter-component phase coherence quantifies coherence between the different components (E, N, Z) at the same station—k and l index different components in equation 5.5—and we then average across component pairs and station pairs. The limited number of channels makes the output noisier. However, this approach also allows us to detect foreshocks and aftershocks that are some distance from the mainshock. With inter-component coherence, shifted earthquake locations still change the station arrival times, but the time shifts are the same across all three components at a given station, and those time-shifts are eliminated when we compute the inter-component coherence. C_{p-comp} thus measures the similarity in the shape of the Green's functions between the mainshock template and a

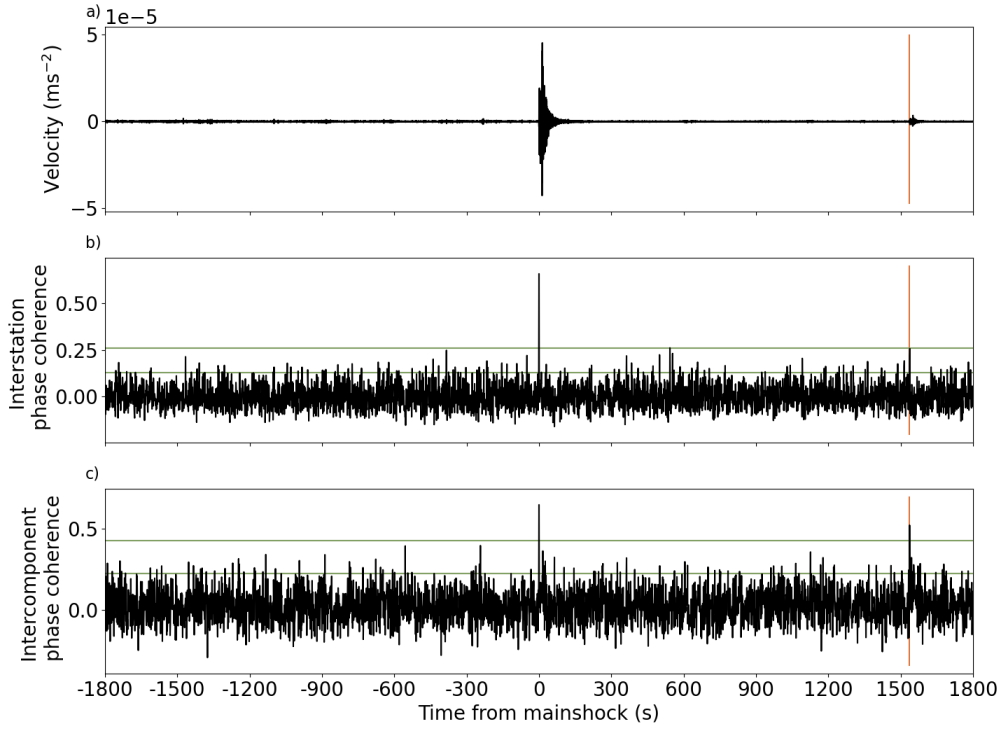


Figure 5.3: a): Velocity seismogram at one station (NZ.BFZ) for 1 hour around an earthquake (2019-10-19 at 17:28:31.26; 37.919°S, 176.426°E; 4.3 mb).
b): Inter-station phase coherence.
c): Inter-component phase coherence.
In b) and c), time is centred on the mainshock and horizontal lines denote 2 and 4 s.d. above the mean phase coherence values. The red vertical line shows the signal in the phase coherence values of an aftershock clearly visible in the velocity seismogram. 6 stations are used in the calculation (BFZ, BKZ, HIZ, KHZ, QRZ, URZ).

window of the continuous signal. We are therefore able to make detections of foreshocks and aftershocks in a region up to around 10 km from the mainshock (Gombert and Hawthorne, 2023).

5.3.4 Detection thresholds and uncertainty

We set thresholds to define detections within the continuous phase coherence records. First, we take the mean of the phase coherence over the full 3600 seconds. We then define a detection as when the phase coherence exceeds 2, 3, or 4 times the standard deviation from the mean. We plot histograms of the number of detections in C_{p-stat} and C_{p-comp} through time in Figure 5.4.

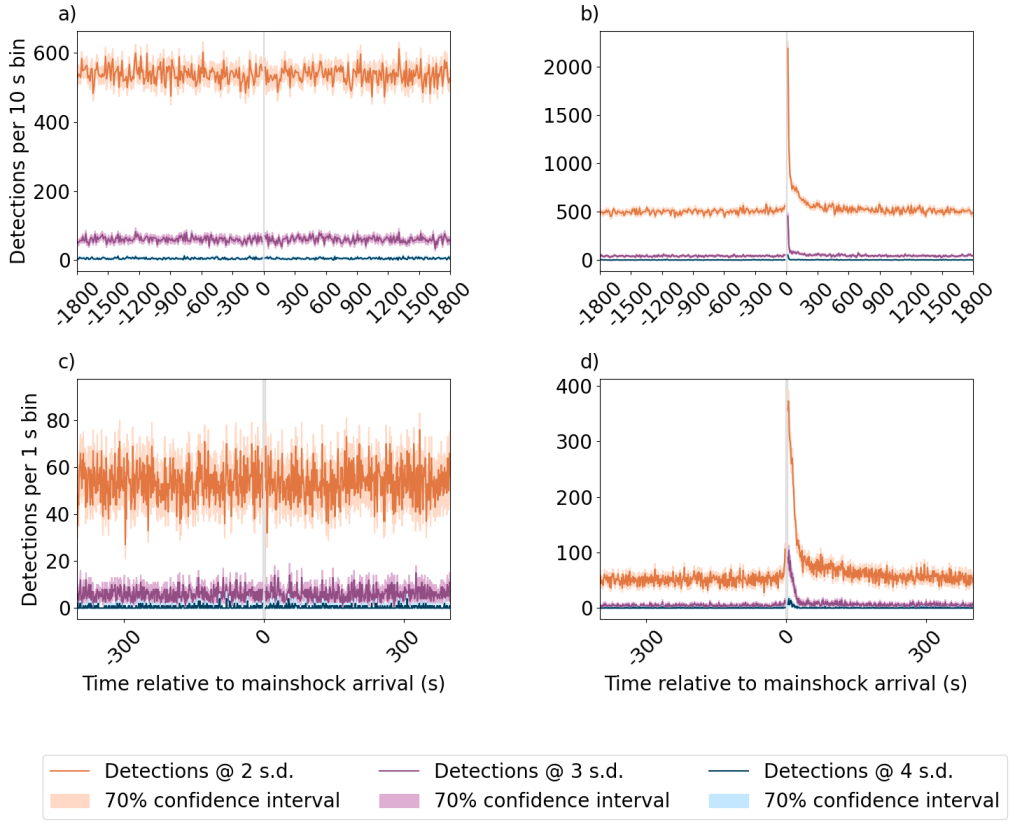


Figure 5.4: Plots of detections in phase coherence through time. The lines join the midpoint of the top of each histogram bar, showing the distribution in the number of detections through time. Panels a) and c) show inter-station phase coherence whilst panels b) and d) show inter-component phase coherence. Panels a) and b) show the full time around the earthquake and bin detections into 10 s bins. Panels c) and d) show 400s before and after the mainshock and use 1 s bins. Shading shows the 70% confidence interval. Orange line and shading is for detections at 2 s.d., purple for 3 s.d., and blue for 4 s.d.

Many of our detections are false detections, due to the low detection thresholds we use. However, we are interested in deviations from this stationary rate, which shows changes from the background rate of seismicity.

To assess the uncertainty in the detection rate through time, we use bootstrap resampling to recompute the number of detections using different subsets of the mainshock population. To create each subset, we resample the mainshock population randomly, with replacement, until the resampled population is the same size as the original population. We then calculate the detection rate again. We repeat this process 100 times to estimate the uncertainty on the detection rate, as illustrated

with the shading in Figure 5.4.

In principle the signal to noise ratio (SNR) should affect the detection threshold. In practice, though, we find that the SNR has minimal effect on our detection capability. We calculate the SNR normalized by moment for each earthquake-component pair and found very little difference ($\sim 1\%$) between the number of detections made for the 1000 mainshocks with the lowest SNR compared to the 1000 with the highest SNR. This is likely due to the averaging across stations which is inherent in the method (see Figure 5.2), and thus most earthquakes have recordings at some stations which have high SNR and others which have low.

5.3.5 Magnitude resolution

We want to compare our detections to expectations from ETAS models which requires knowledge of our detection capability. Here, then, we estimate the magnitude of completeness of our detections.

Our analysis leaves us with 1,165,952 detections including background earthquakes, false detections, foreshocks, and aftershocks. All of the detected earthquakes should be smaller than $M3$, as we considered all $M \geq 3$ earthquakes in the GNS catalogue as potential mainshocks. But some of the detected earthquakes are between $M2.5$ and 3. These earthquakes should be in the GNS catalogue, as that catalogue is complete to $M2.5$. So we search the GNS catalogue for $M2.5 - 3$ earthquakes that occur at the same time as our detections. We identify 918 such earthquakes distributed at a range of times before and after the mainshocks.

However, many of these earthquakes are part of a background seismicity rate, as are many of our other detections, and we cannot distinguish a constant background seismicity rate from a constant rate of false detections. So we subtract an estimated background rate from our detections: we remove the average number of events in a 400-s window at the start of the 30-minute interval. This subtraction leaves 11,233 foreshock and aftershock detections in total, 100 of which are between $M2.5$ and 3 and are present in GNS catalogue.

We use the number of $M2.5-3$ foreshocks and aftershocks ($N_{2.5>M>3}$) to find the parameter a of a Gutenberg-Richter distribution, where the number of earthquakes between $M2.5$ and 3 is $10^{a-2.5b} - 10^{a-3b}$. Here a is a measure of the total seismicity in the region, and we estimate it to be 4.665. We assume $b = 1$.

Given the a-value, we can calculate the number of events above any given minimum magnitude M_{min}:

$$N(M > M_{min}) = 10^{a}(10^{-bM_{min}}), \tag{5.6}$$

We set this number equal to 11,233, the number of foreshocks and aftershocks we detect, and solve for the minimum magnitude M_{min}, obtaining

$$M_{min} = -\log_{10}\left(\frac{11233}{10^{4.665}}\right) = 0.61. \tag{5.7}$$

These calculations suggest that we have detected earthquakes down to around $M0.6$.

In the above, we consider a perfectly defined magnitude of completeness, where we detect all earthquakes above M_c, and no earthquakes below M_c (Figure C.2a). In reality this is an over-simplification as there is a gradual taper from measuring all earthquakes, to measuring none, over some magnitude window. We discuss this more in Section C.2 and find that with a cosine taper that the magnitude of completeness is between $M0$ and $M1$ (Figure C.3).

5.4 Patterns in phase coherence through time

Now that we have our earthquake detections and an estimate of the range of earthquake magnitudes we can detect, we examine how the number of detections varies with time from the mainshock. Throughout our calculations, we ignore detections between -1 s and 2 s of the mainshock, as that interval is contaminated by the mainshock.

If earthquake-earthquake triggering controls all earthquake clustering, we expect the foreshock and aftershock rate to follow Omori's law, with the earthquake rate decaying as t^{-1} with time before or after the mainshock (Parsons, 2002; Helmstetter et al., 2003; Felzer et al., 2004). However, if nucleation is more complex, and

external processes influence slip acceleration, the earthquake rate may or may not follow this characteristic power-law decay.

5.4.1 Inter-station Coherence

Figure 5.4a shows the inter-station detections through time, averaged across all three components. We do not see any patterns in the detection rate. The detection rate is constant within error, with a rate of 544 detections per 10-second bin, outside of the window around the mainshock. We also see no variation in detection rate on a shorter timescale: in the 400 s before and after the mainshock, using a histogram bin width of 1 s (Figure 5.4c).

5.4.2 Inter-component

Inter-component phase coherence, averaged across different stations has a background detection rate of about 495 per 10-s time bin (Figure 5.4b). Many of these are false detections, where noise in the 3-component calculation happens to be slightly coherent with the template, but that false detection rate is constant in time and cannot be distinguished from the background seismicity rate. On top of the constant, we see a variation in detection rate which appears to come from foreshocks and aftershocks. The number of detections increases just before the mainshock and then gradually decreases after the mainshock. Even after the detection rate has decreased and starts to level out, the number of detections remains slightly elevated; we consider only 30 minutes after the mainshock, and seismicity has not yet returned to regular background levels.

As one check on our results, we calculate the ratio of the peak detections to the average background detection rate using various detection thresholds. We find that the peak detection rate is 2.44× the background for detections at 2 s.d., 4.70× the background for detections at 3 s.d. at 13.6× the background for detections at 4 s.d.. This lends credence to the idea that our detections above the background rate are real; as we increase the strength of our detection criterion we detect a higher proportion of high quality (real and above the background rate) detections.

Shorter-timescale variations may be better seen in Figure 5.4d, where we plot the detection rate in the 400 s before and after the mainshock, using 1-s bins. The detection rate increases abruptly in the seconds before the mainshock. After the mainshock, detections decrease steadily following a power-law distribution.

5.4.3 Temporal patterns in foreshock and aftershock activity

To better examine the distribution in time of our detections (at 2 s.d.), we plot them in log-log space, where we see a clear power-law decay in detections (Figure 5.5a). In an ETAS model, the seismicity rate before or after a mainshock decays as time^{-p}, following Omori's law (Utsu et al., 1995). Here p is a decay parameter which is typically around 1. If foreshocks and aftershocks both result from inter-earthquake triggering, we expect the same p value to describe both Omori fits.

Here we attempt to fit our foreshock and aftershock rate as

$$N(t) = C_1 + C_2 t^{-p}, \tag{5.8}$$

where C_1 is a constant representing the background rate, including false detections, and C_2 is a constant representing the number of foreshocks or aftershocks. The aftershock distribution is fit well by this Omori's law scaling, using $p = 1$ but the numbers of foreshocks are underestimated at times close to the mainshock. Figure 5.5a shows that the observed aftershock rate, denoted by the pink curve (shading showing 70% confidence interval), is relatively close to the best-fit Omori curve (purple curve) at all times from the mainshock. In fitting the Omori law curve, we fix C_1 as the background seismicity rate, calculated over the first 400 seconds, and look to optimise C_2.

In Figure 5.5b, we optimise for the exponent, p, as well as for C_2. Whilst the Omori law with $p = 1$ gives a reasonable fit by eye, the optimised value of $p = 0.56$ shows that a better fit to the foreshock observations is achieved by varying the p value away from 1, especially at times close to the mainshock. We can quantify the goodness of fit using the R^2 metric. We find $R^2 = 0.265$ for $p = 1$ and $R^2 = 0.309$

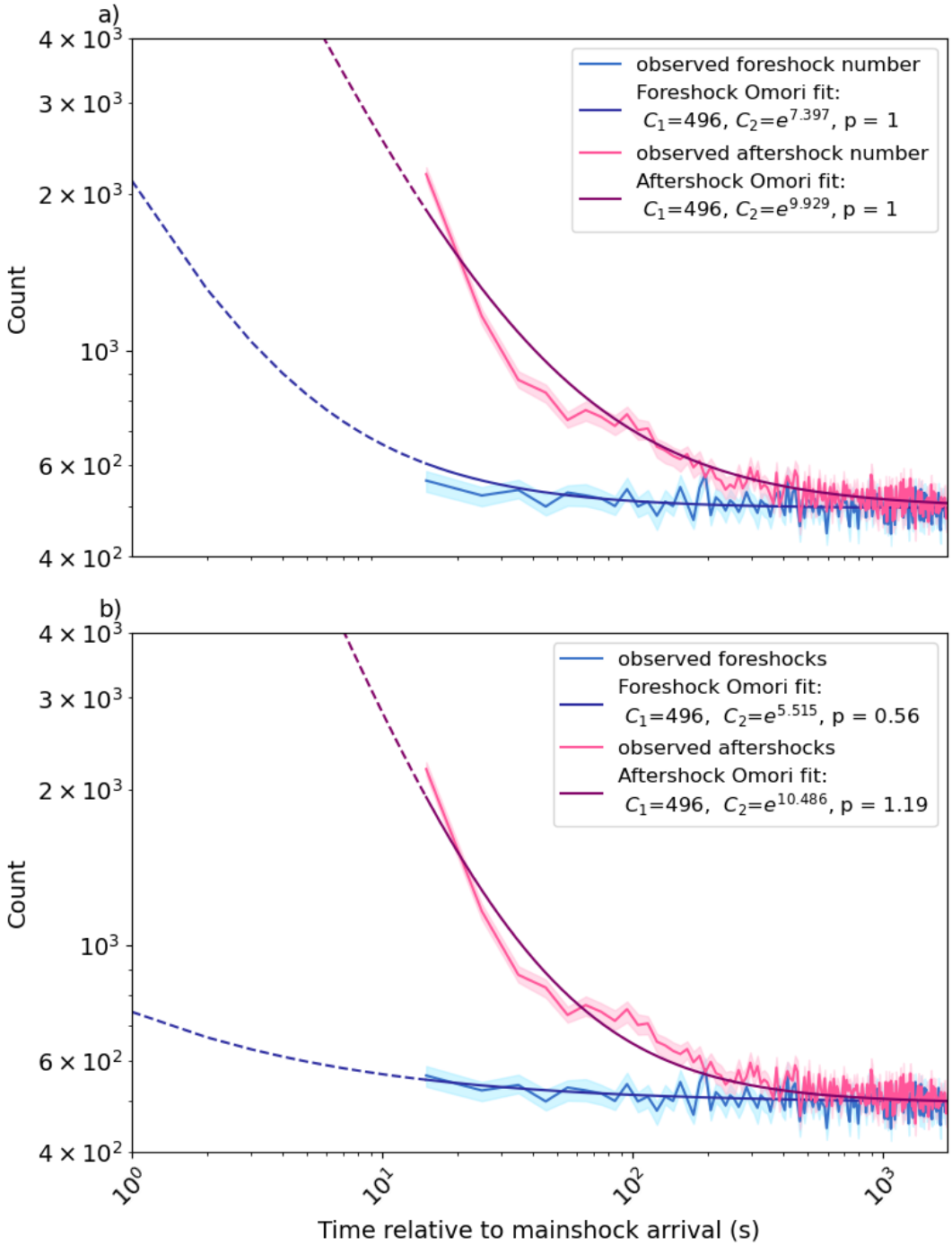


Figure 5.5: Log-log plots of foreshock (blue) and aftershock (pink) detections at two standard deviations time relative to mainshock arrival. Power law relations, following Omori's law, are plotted: in a) with the p-exponent fixed to 1, and in b) where p is optimised independently for foreshocks and aftershocks. The asymptote for both is fixed at the background detection value over the first 500 s (496 detections/10 s bin). The point at 5 s (corresponding to 0-10 s for aftershocks and -10 –0 s for the foreshocks) is not plotted, as we remove all detections within ±2 s of the mainshock due the peak spreading.

for $p = 0.56$. Varying p has no real effect on the R^2 value for aftershocks (0.947 for $p = 1$ and 0.961 for $p = 1.118$. Despite the small improvement in the R^2 metric, $p = 0.56$ is statistically significantly different than $p = 1$. In Figure C.4, we bootstrap our detections in the same way as Section 5.3.4 and then fit Omori decays to each of these bootstraps, to understand the uncertainty in the power-law exponent. We plot the distribution of the exponent in Figure C.5, and see that both the foreshocks and the aftershocks have an optimal fit exponent which is not 1.

5.5 ETAS model

We would also like to estimate the numbers of foreshocks and aftershocks expected from an ETAS model. To do this, we must first estimate the ETAS parameters (see Section 5.1.1) for our study area. To invert for the ETAS parameters, we use the same GNS earthquake catalogue ($M \geq 3$, 2005–2020) that we used to identify the mainshock events for phase coherence.

While the ETAS model parameter estimates can show some sensitivity to the chosen minimum triggering threshold (Bantidi et al., 2024), we select $M_0 = 3$ based on two key considerations: (1) using the same dataset with same minimum magnitude ensures consistency and comparability between the ETAS results and our observational findings, and (2) this threshold aligns with established practices in regional studies, which typically use $2.5 \leq M_0 \leq 4$. This range ensures that all earthquakes capable of contributing to the triggering process are included, particularly for regions with strong seismic network coverage (e.g. Guo et al., 2015; Bantidi et al., 2024; Li et al., 2024)

The time series of our dataset includes various periods of temporal incompleteness between 2005–2020 (Figure C.6). Studies have shown that such gaps in earthquake catalogues, ranging from short-term to long-term incompleteness, can introduce significant biases in the ETAS model inversions and parameter estimations (Hainzl, 2021; Naylor et al., 2023). To address this issue, we applied a modified version of the ETAS model proposed by Kamranzad (2024), which incorporates a time-dependent censorship function specifically designed to account for temporal incompleteness in

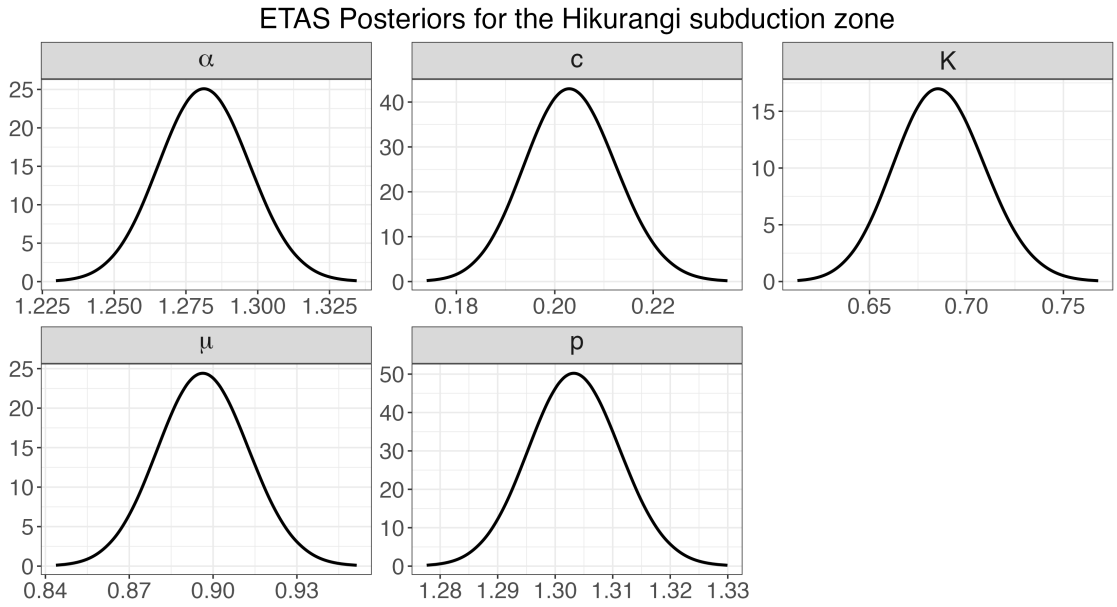


Figure 5.6: Posteriors of the ETAS parameters. μ has units of events per day; c has units of days; and K, α and p have no units.

the data. By adjusting the ETAS conditional intensity and likelihood functions, it compensates for missing events, reducing the impact of incompleteness and improving the accuracy of model parameter estimates. The modified ETAS model uses a fast Bayesian approximation called the Integrated Nested Laplace Approximation (INLA) method (Rue et al., 2009; Bachl et al., 2019). This method, applied in the ETAS modelling (Naylor et al., 2023; Kamranzad, 2024), produces posterior distributions for the model parameters instead of single point estimates, enabling the quantification of uncertainty associated with each parameter estimate. The resulting posterior distributions are illustrated in Figure 5.6, with the vector of posterior modes given as ($\mu = 0.89$ per day, $K = 0.68$, $\alpha = 1.28$, $c = 0.2$ days, $p = 1.3$).

To account for parameter variability and incorporate uncertainties, we sample from the ETAS posterior distributions. Using different sets of the sampled parameters, we simulate 40 synthetic earthquake catalogues over a 15-year period, including events with magnitudes down to $M = 0$.

Table 5.1: Observed and bootstrapped foreshock-aftershock ratios.

Mag. range	Obs. ratio	Mean bootstrap	St.dev. bootstrap
$M \geq 3$	0.113	0.103	0.048
$M \geq 4$	0.096	0.092	0.048
$M3 - 4$	0.159	0.156	0.117

Mag. range gives the magnitude range under consideration. *Obs. ratio* gives the observed ratio. *Mean bootstrap* gives the mean ratio found from bootstrapping, whilst *St.Dev. bootstrap* gives the standard deviation on this ratio.

5.6 Foreshock:aftershock ratio

We compare the observed foreshock:aftershock ratio to that expected from our ETAS model, to provide a more rigorous assessment of the earthquake-earthquake triggering model.

5.6.1 Observations

To compute the number of foreshocks and aftershocks, we first subtract the background detection rate: the average rate in the -1800 to -1400 seconds before the mainshock. We assume that each remaining detection represents a single earthquake, and we sum the number of detections before and after the mainshocks to get the number of foreshocks and aftershocks, respectively. We then compute the foreshock:aftershock ratio for groups of mainshocks with different magnitudes.

In Figure 5.7, we plot the observed ratio (navy line) and its bootstrapped distribution and confidence intervals (blue bars) for all mainshocks (also see Table 5.1). Figures C.7 and C.8 show the result for the other mainshock magnitude groups.

The declustering window has more of an impact on the observed ratio than the background window (Table C.1). However, since we use the same background and declustering windows for analysis of both our observations and the ETAS results, this should not impact our conclusions.

For some magnitude ranges, the foreshock:aftershock ratios estimated from bootstrapping include negative values, which are not physically interpretable. The

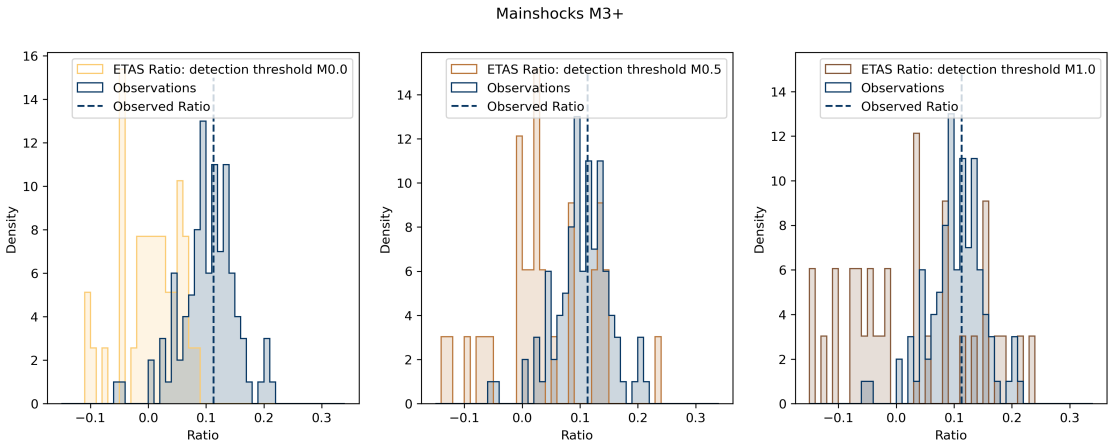


Figure 5.7: The foreshock-aftershock ratios found from our observations (blue) and from ETAS models (yellow/brown) for all mainshocks. The vertical dashed line shows the observed ratio, and the blue distribution shows the uncertainty on this ratio found from bootstrapping. The yellow/brown histograms show the distribution of the ratio found from the 40 ETAS simulations. The different panels have a different detection threshold of foreshocks and aftershocks in the ETAS simulation ($M0$, $M0.5$ and $M1$ from left to right).

negative values in the distribution simply tells us that the uncertainty includes zero wherein there are zero detectable foreshocks above the background seismicity rate.

5.6.2 Comparing observations and predictions

5.6.2.1 Foreshock-aftershock ratio

We compare the distribution of observed foreshock-aftershock ratios against the distribution of ratios found from the ETAS model (Figure 5.8), in three mainshock magnitude ranges: all the mainshocks ($M \geq 3$), $M3-4$ and $M \geq 4$. For the ETAS models, we compute the foreshock-to-aftershock ratio for each of our 40 simulations. The smallest magnitude of event we can detect is a major source of uncertainty. We estimated it to be ≈ 0.5 in Section 5.3.5, but we additionally do these calculations for detection limits of both $M0$ and $M1$ (Section C.2).

We compare the distribution of observed ratios against the distribution of ratios found from the ETAS model for all mainshocks ($M \geq 3$) in Figure 5.7 (Figures C.7 and C.8 for the $M \geq 4$ and $M3-4$ subsets respectively), and see that the observed ratio seems to generally be higher than predicted from ETAS, but that there is a relatively large range within the results. The distribution of the ratio found from

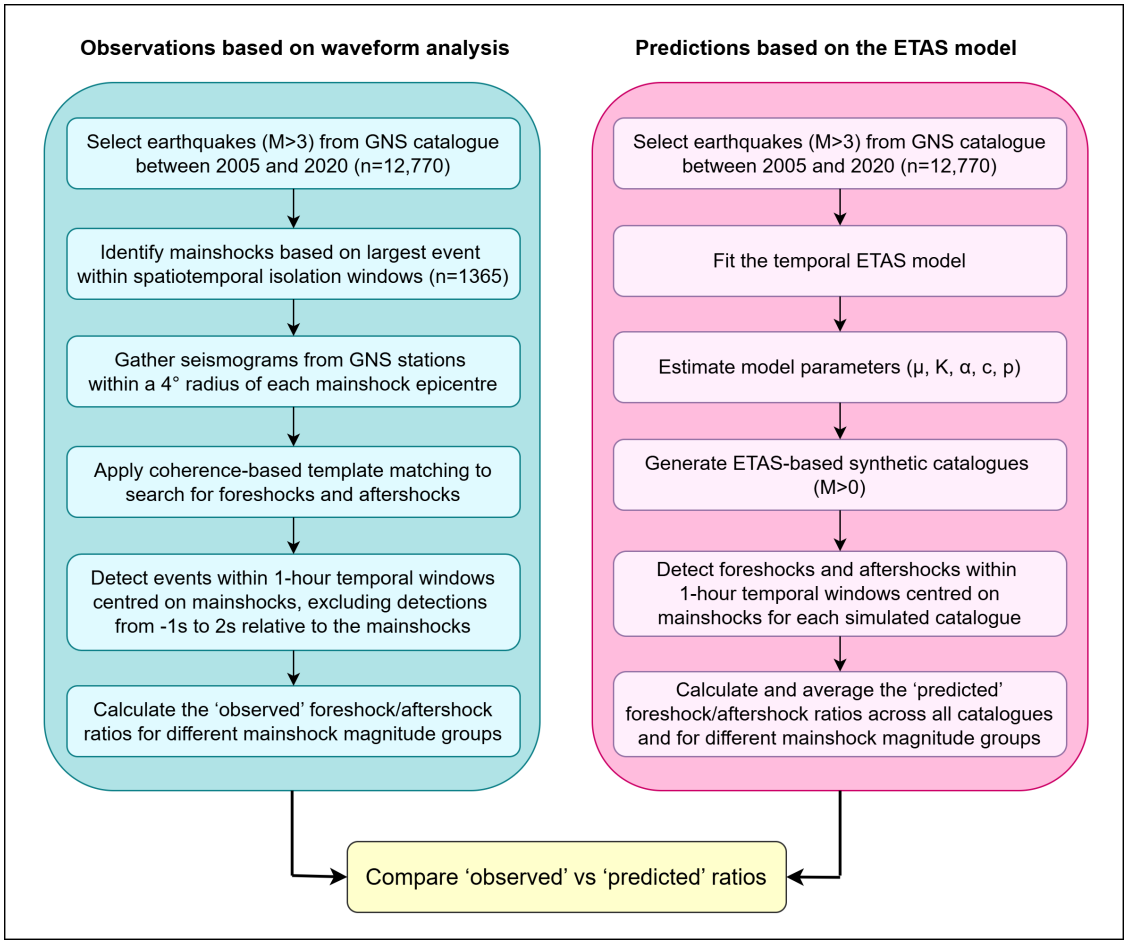


Figure 5.8: Workflow for investigating 'observed' and 'predicted (expected)' foreshock/aftershock ratios.

ETAS, and the distribution of the ratio found from our observations, reflect subtly different uncertainties. The distribution of the ratio found from ETAS reflects how uncertainty in the ETAS parameters (Figure 5.6) impacts the foreshock-aftershock ratio we find. Meanwhile the distribution of the ratio found from our observations reflects the uncertainty in our observed ratio. This means we cannot determine whether ETAS adequately reflects our observations simply by testing whether the two distributions are statistically likely to arise from the same parent distribution (i.e. by conducting a two-sample Kolmogorov-Smirnov test).

Instead we randomly sample from both distributions and ask whether the ratio from the ETAS distribution is larger or smaller than the ratio from the distribution of our observations. By repeating this 1000 times, we can generate a distribution of

this difference. If ETAS explains the observations, then we expect this distribution to be centred on 0. For $M \geq 3$ mainshocks, we find that the mean difference is 0.10 ($\sigma = 0.07$) and the median difference is 0.10. So whilst the ETAS ratio is smaller than the observed ratio 93% of the time (Table 5.2), the distribution of the difference is indistinguishable from zero at two standard deviations.

Table 5.2: Statistics of the distribution of the difference between the foreshock-aftershock ratio calculated from ETAS and from observations.

Min M	Max. M	Det. lim.	Mean diff.	StDev diff.	Median diff.	$\% > 0$
3.0	10.0	0.0	0.10	0.07	0.10	93.25
		0.5	0.11	0.12	0.09	80.50
		1.0	0.12	0.16	0.12	75.52
4.0	10.0	0.0	0.09	0.06	0.08	91.38
		0.5	0.08	0.11	0.07	78.70
		1.0	0.11	0.16	0.09	72.58
3.0	4.0	0.0	0.22	0.31	0.17	76.78
		0.5	0.26	0.46	0.19	73.45
		1.0	-1.30	10.93	0.24	69.50

Notes: *Min. M* is the minimum mainshock magnitude considered; *Max. M* is the maximum mainshock magnitude considered (10 is larger than the largest event in the catalogue); *Det. lim.* is the smallest foreshock or aftershock we can detect; *Mean diff.* is mean difference between the ETAS and observed FS:AS ratios after sampling as described in the main text; *Median diff.* is median difference; *StDev diff.* is the standard deviation of the difference; $\% > 0$ is the proportion of pairwise comparisons where the ratio observed is greater than the ratio predicted from ETAS.

As we look at the distribution of the difference between the foreshock-aftershock ratio found from ETAS and that from our observations, we see all the mainshock magnitude and detection limit combinations form a vaguely Gaussian distribution (Figure 5.9). The $M \geq 3$ and $M \geq 4$ mainshocks with a detection limit of $M0$ form a very clear normal distribution. The ratio found from our observations is greater than that found from ETAS 93% and 91% of the time respectively. As we look at higher detection limits, the distribution becomes broader, less Gaussian and develops a positive skew (long tail at high difference values). The $M3-4$ mainshock group results are much broader again and less smooth, which is at least

partly attributable to the smaller number of mainshocks going into this estimate. The $M1$ detection limit results for the $M3-4$ mainshocks have a very different mean difference value, of -1.30 and a high standard deviation (10.93) (Table 5.2).

The $M \geq 3$ and $M \geq 4$ results are the most robust and thus we can conclude that our observed foreshock-aftershock ratio could result from an ETAS model at 2-sigma confidence.

The ETAS models produce a much higher probability of a negative foreshock-aftershock ratio than our observations (Figures 5.7, C.7 and C.8). In all cases, the negative foreshock-aftershock ratio is attributable to a lack of increase in foreshock activity compared to the background in the 30 minutes before the mainshock.

5.6.2.2 Temporal patterns

In our observations, there is a clear exponential increase in activity before the mainshock, and an exponential decay after the mainshock (see Section 5.4 and Figure 5.5). In the results from the ETAS models (Figures 5.10 and C.9) we do not see this pattern. Indeed, a linear fit (bold dotted lines) appears to provide a similar goodness-of-fit. Quantifying this by calculating the R^2 value we find that the linear fit produces a higher R^2 for both foreshocks and aftershocks (Foreshocks: linear $R^2 = 0.00137$, Omori $R^2 = 0.000$; Aftershocks: linear $R^2 = 0.0644$, Omori $R^2 = 0.00530$). However, all the R^2 values are very low (recall $R^2 = [0, 1]$), suggesting that neither line is a particularly good fit. Looking at Figure 5.10, this makes sense, since the variation in the ratio between different data-points is so significant compared to the trend of the line.

It is also notable that the ETAS models produce fewer foreshocks and aftershocks at background compared to our observations.

5.6.3 Foreshocks and aftershocks per mainshock

Comparing the number of aftershocks per mainshock, we see that ETAS produces fewer aftershocks per mainshock in all mainshock magnitude ranges and detection limits (Figures 5.11, C.10 and C.11). Higher detection limits result in fewer

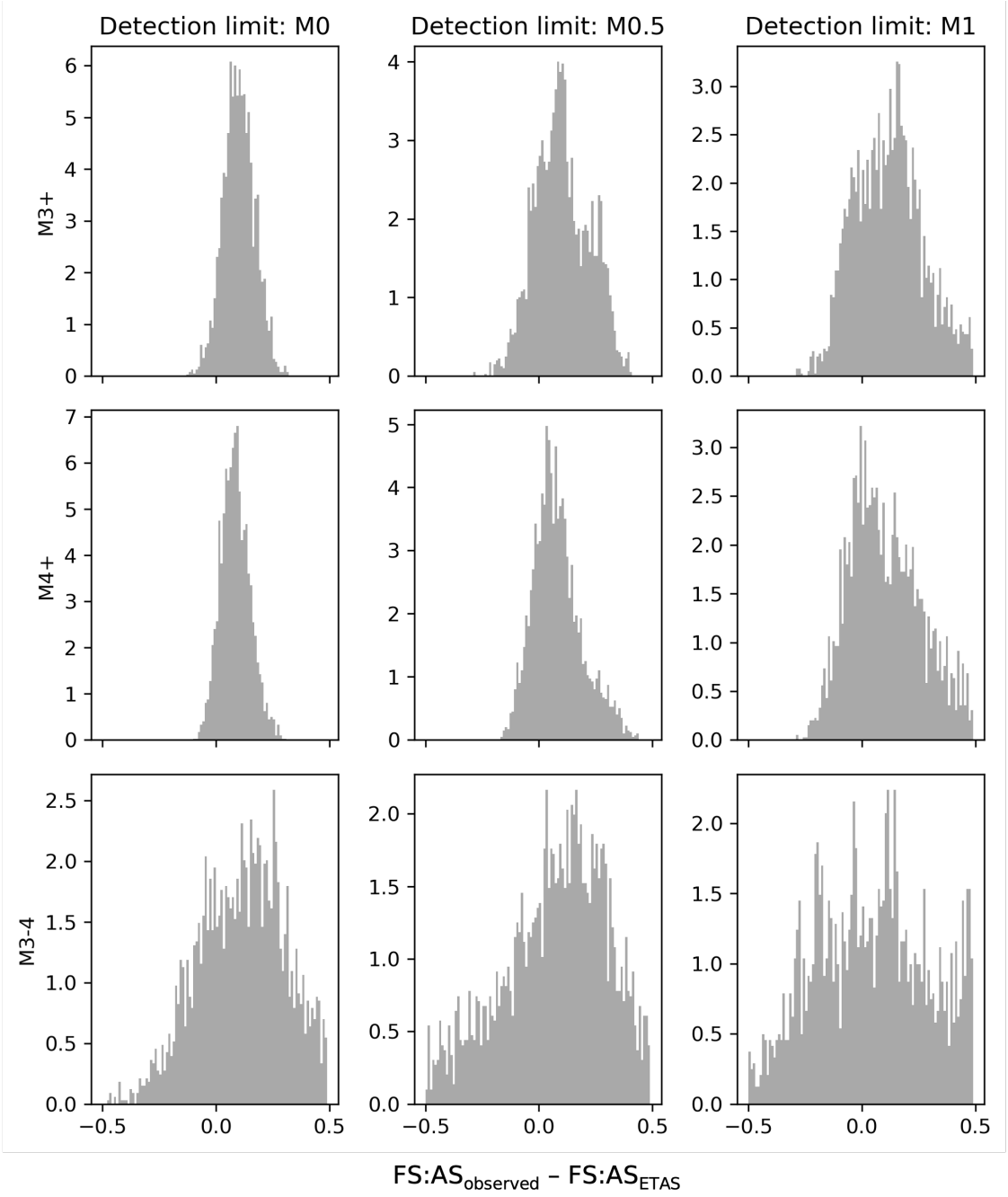


Figure 5.9: Histograms of the difference in foreshock-aftershock ratio between randomly sampled bootstraps of the observations and randomly sampled ETAS model runs.

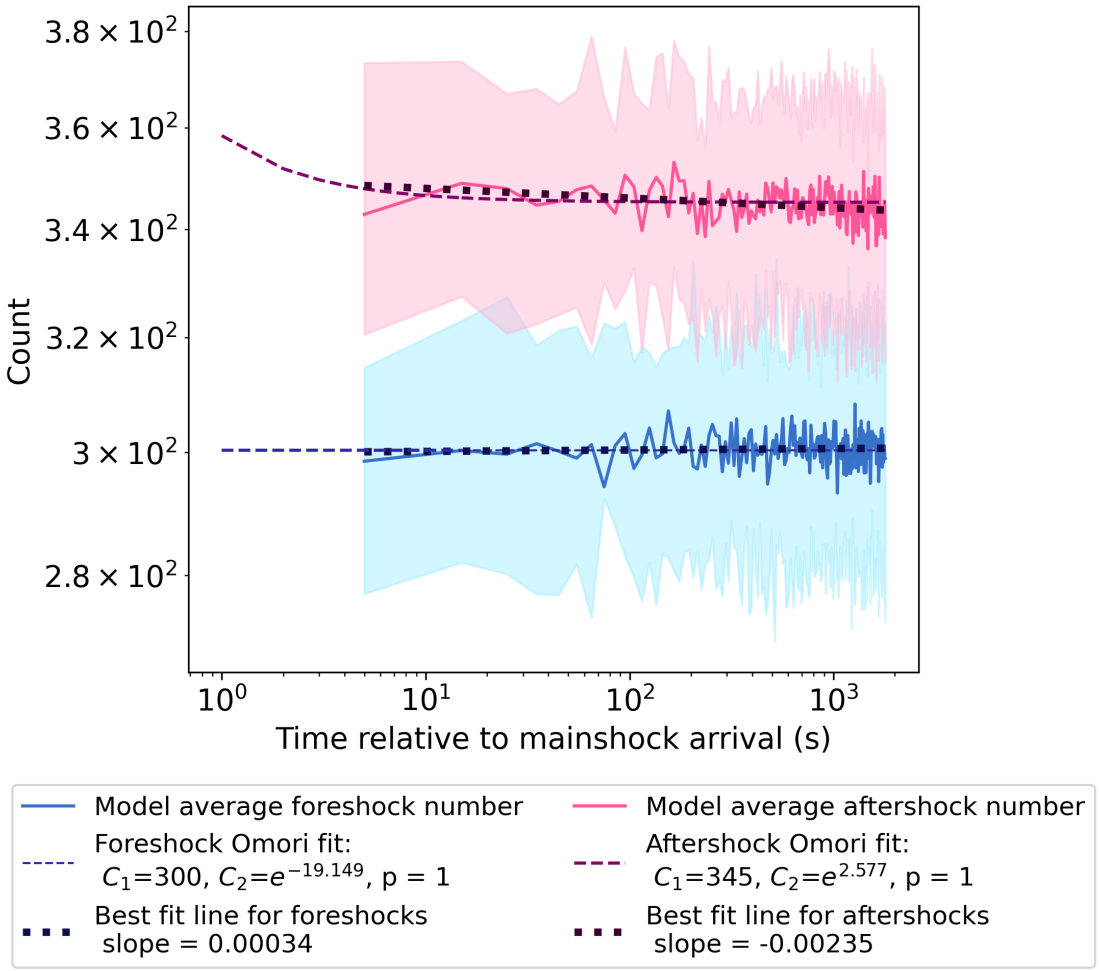


Figure 5.10: Log-log plot of foreshock (blue) and aftershock (pink) detections found from the ETAS models with time relative to mainshock arrival. Power law relations, following Omori's law, are plotted as dashed lines. Linear relationship is plotted as a dotted line.

aftershocks per mainshock, which is to be expected. We observe that larger events have more aftershocks per mainshock (mean = 6.05 for $M \geq 3$ mainshocks, 9.22 for $M \geq 4$ mainshocks and 3.34 for $M3 - 4$ mainshocks).

Moving onto the number of foreshocks per mainshock, all three detection limits produce fewer foreshocks per mainshock than we observe (Figures 5.11, C.10 and C.11). We see that larger mainshocks have more foreshocks per mainshock. The distribution of foreshocks per mainshock is much tighter for the $M \geq 4$ mainshock group than for the $M \geq 3$ and $M3 - 4$ groups.

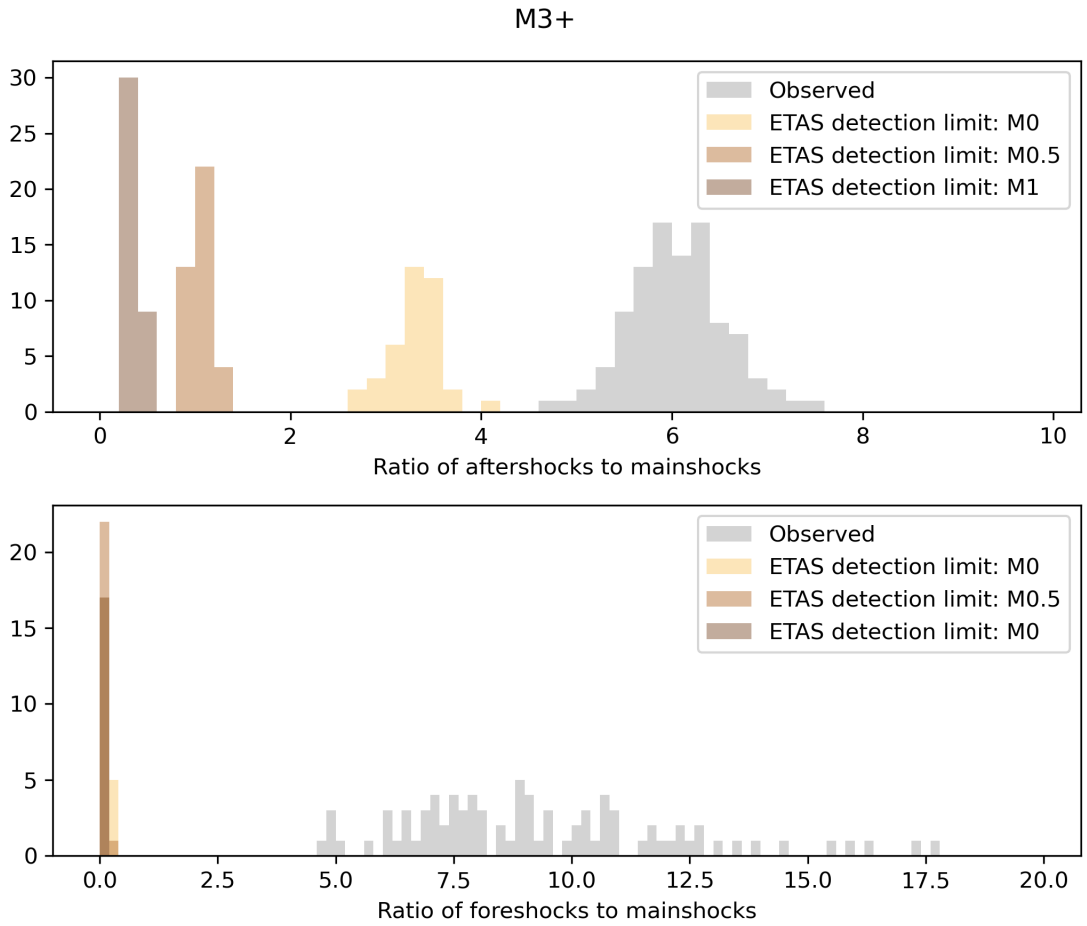


Figure 5.11: Histograms of the number of aftershocks per mainshock (top panel) and foreshocks per mainshock (bottom panel) using all mainshocks. The observations are in grey and the ratios found from ETAS in yellow-brown (darker colours show increased detection limit).

5.7 Sequences

The improved Omori fit for foreshocks when $p \neq 1$ suggests that nucleation is not entirely explained by earthquake-earthquake triggering, but it is a relatively subtle indication. We therefore look for something which would more obviously indicate slip acceleration: foreshock sequences. Tape et al. (2018) identified intense, minute-long sequences of foreshocks before mainshocks in Alaska. We thus systematically look for sequences of detections before and after our mainshocks.

We look for sequences in windows of different lengths, from 5 to 20 s. For each window length, we compute the fraction of the 1-s bins which contain detections. A

high fraction indicates a number of detections in short succession, which we consider a foreshock sequence. We then compare this fraction to a range of thresholds, between 10% and 100%, to determine if the window contains a sequence. With these thresholds, we identify a large number of sequences, particularly at times close to the mainshock and for short window lengths (Figure 5.12).

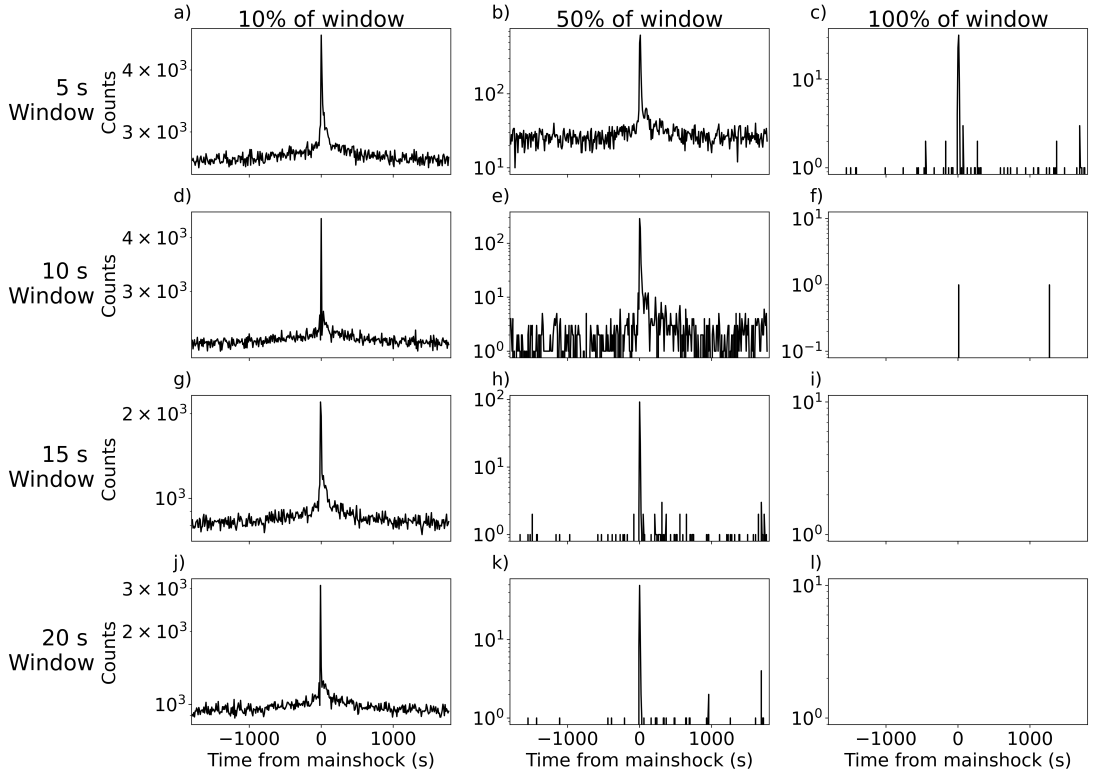


Figure 5.12: Histograms of detections of sequences for different time windows and proportion of window filled. The mainshock occurs at time 0, but to avoid any double counting, we ignore detections for the 2 seconds before and after it. Blank boxes show no sequences were detected.

We examine a number of the apparent sequences visually. In Figure 5.13, we show the phase coherence record and a seismogram for one of these sequences, shown by the blue box. This sequence has a detection in 100% of the 5 second window: the signal that originates in this time window is coherent with that of the mainshock.

We compare the ratio of sequences to detections in different time windows to see if there is a statistically significant increase in the number of sequences as we approach the mainshock.

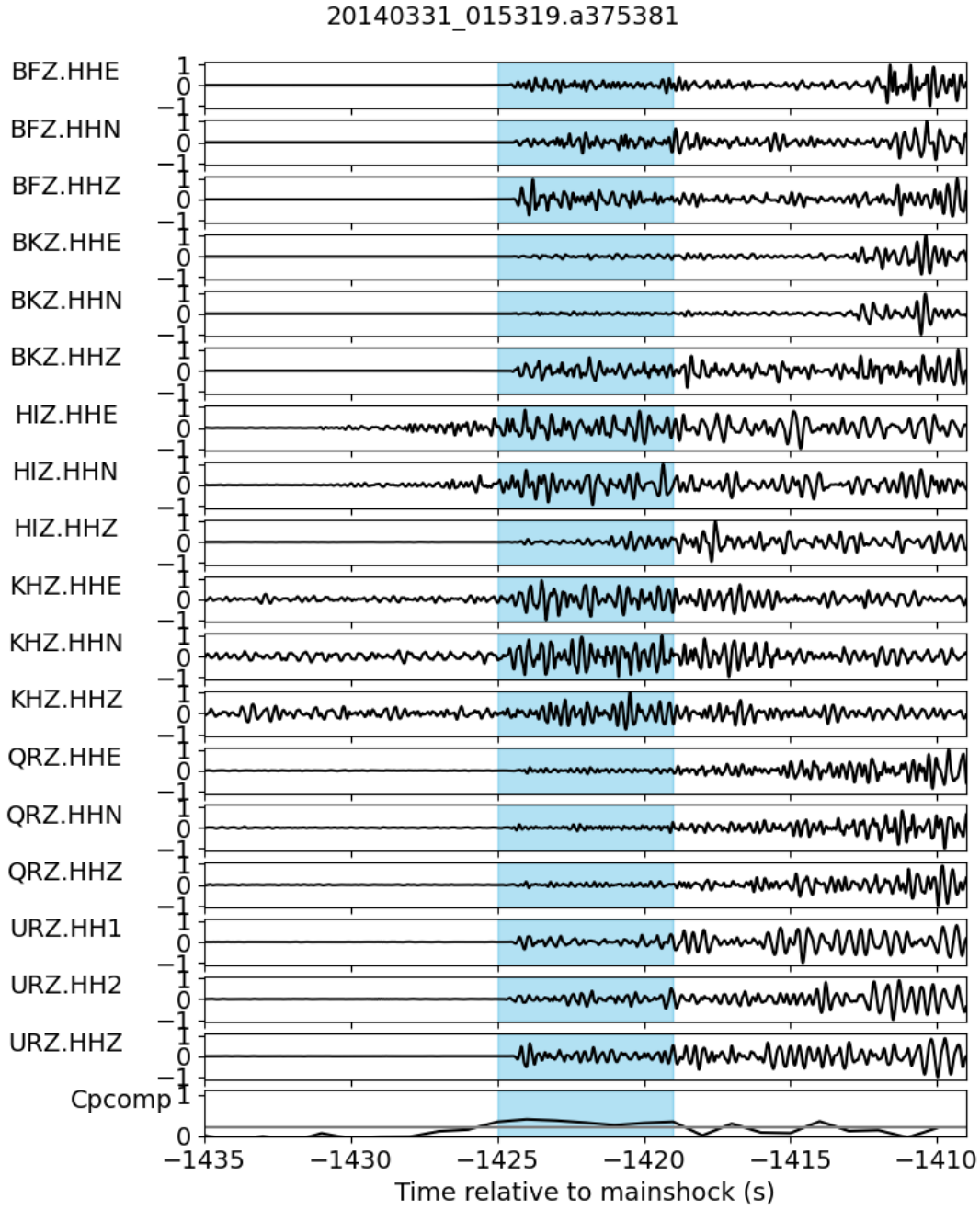


Figure 5.13: A sequence detection. The top panels show normalised velocity seismograms from 6 stations with the detected sequence highlighted by the blue box. The bottom panel shows the phase coherence value, with the horizontal grey line being 2 s.d. above the mean value. Note that the C_P value for this window is above the detection limit. For comparison, the average C_p value of the full 3600 s record shown in Figure 5.3 is 0.001.

Over the first 200 seconds of the record (−1800 – −1600 s, considered to be representative of the background), there are 0.0083 sequences per detection (considering a sequence to be 60%+ of a 5 second window). The 95% confidence interval on this number is 0.0075 – 0.0090. In the window 1750–1797 (the 50 seconds before the mainshock, removing the blanked window around the mainshock), the ratio is 0.0113. In other words the sequence rate has increased by 17% whilst the detection rate (including false detections) has increased by 10%. The more dramatic increase in sequence rate suggests the increased sequence rate comes from detection clustering, not just an increased number of detections, though we have not robustly analysed the statistics.

As we consider larger proportions of the window, and longer windows, the sequence rate appears constant through time, but this may just result from an increase in the uncertainty, as very few sequences are identified in any one time interval.

5.8 Discussion

In this work, we have used a coherence-based approach to detect numerous foreshocks and aftershocks in the 30 minutes before and after mainshocks. Whilst the inter-component phase coherence makes many detections of foreshocks and aftershocks, the inter-station phase coherence, on the other hand, detects few to no foreshocks and aftershocks. The lack of inter-station detections could imply that 1) the foreshocks and aftershocks are not perfectly co-located with the mainshock or 2) the foreshocks and aftershocks are too small to be detected on more than one station.

We have chosen to analyse the detections made with inter-component phase coherence. We find that:

1. The foreshock:aftershock ratio is in-line with that expected from ETAS models at 2 standard deviations.

2. We observe more foreshocks and aftershocks per mainshock than predicted by an ETAS model.

3. The observed foreshock and aftershock rates are well fit by a power-law.

4. The observed foreshock rate is better fit by an Omori power law decay with $p = 0.56$ than one with $p = 1$.

5. The ETAS models do not produce a significant power-law increase or decay on the short time windows under consideration.

6. There is a statistically significant increase in the number of foreshock sequences before the mainshocks, but there are no obvious tremor-like precursors.

That the uncertainty in the difference between our observed foreshock-aftershock ratio and the ratio we find from ETAS includes zero, suggests that ETAS explains at least some of our observations.

However, some of our observations are not well explained by ETAS: we see far more foreshocks and aftershocks per mainshock than modelled by ETAS, and we observe power-law decays in our observations of both foreshocks and aftershocks which are not apparent in the ETAS simulations.

We therefore propose that on the short time scales we consider, a combination of stochastic triggering and external processes such as aseismic slip or pore-fluid pressure changes are involved in earthquake nucleation. These processes would help explain the elevated foreshock and aftershock activity we observe as well as explaining why we see a much stronger exponential decay in our observations than predicted by ETAS. This proposal is similar to the findings of Yao et al. (2020) who required a combination of aseismic slip and cascade triggering to best explain foreshock observations in the 2010 El Mayor-Cucapeh sequence. The external processes acting on the fault may even cause fault conditions to change between foreshocks and aftershocks (unlike Brodsky, 2011), so that external-process-mediated earthquake-earthquake triggering occurs in different conditions (e.g. Helmstetter et al., 2003). Changing conditions on the fault, such as through fault-valve action during the mainshock, might also explain why the optimised Omori fits for both foreshocks and aftershocks are found by varying the p exponent away from 1, but that foreshocks require an exponent of 0.56, whilst aftershocks require an exponent of 1.188 (Figures 5.5b and C.4).

If foreshocks are clustered due to an external physical process rather than solely resulting from earthquake-earthquake triggering, they may form an earthquake swarm. A number of earthquake swarms have been recorded on the Hikurangi subduction zone, thought to be triggered by both high-pressure fluid flow and slow slip. These swarms are thought to have triggered some, but not all, larger earthquakes (Evison and Rhoades, 1999; Reyners and Bannister, 2007). We might expect swarms to have high rates of both foreshocks and aftershocks, since the largest event (which we consider a mainshock) is unlikely to occur right at the beginning or end of the swarm. On the timescales on which we look for foreshocks and aftershocks (30 minutes), we may not see any increase in activity, since the period we use to calculate the background rate is part of the swarm itself. This means that we would correct for any increase in background rate and thus would not record an increase in foreshock or aftershock activity.

We are unlikely to be seeing the full foreshock or aftershock sequence given our short (30 minute) search window. Cochran et al. (2020) looked at clusters of seismicity induced by fluid injection. They found that for larger clusters (more than 10 events), the number of foreshocks and aftershocks was comparable, but that the foreshock sequence was longer in duration than the aftershock sequence. In windows shorter than the foreshock sequence, this would serve to reduce the observed foreshock-aftershock ratio. In this case, the role of earthquake-earthquake triggering may be reduced or removed compared to the role of external processes, since our ratio would likely become different at $\alpha = 0.05$.

Tectonic setting may also play a role in our observations. Observations of foreshocks suggest that between 18% (van den Ende and Ampuero, 2020) and 72% (Trugman and Ross, 2019) of mainshocks in Southern California exhibit elevated earthquake activity in the preceeding hours and days. Reasenberg (1999) find that subduction zones have more foreshocks per mainshock compared to continental settings like the San Andreas. This may therefore lead to a higher foreshock:aftershock ratio if all settings have a similar propensity for aftershocks. This might explain why whilst our observations are statistically similar to ETAS

results, our observations do tend to have a higher foreshock-aftershock ratio than expected (as described by the difference between randomly selected bootstraps of our observations and randomly selected ETAS simulations).

It is strange that we do not see a power-law decay in the ETAS model foreshock and aftershock rates. The strength of the power-law decay is primarily driven by the parameter c that seems to be well constrained by the ETAS inversion that is, itself, well justified. One explanation for this may be that c has some magnitude dependence. If we consider that c may represent the nucleation time of nearby earthquakes, or the timescale over which afterslip accumulates and loads a patch before that patch ruptures, then it could make sense that smaller earthquakes can be triggered at shorter timescales than the $M \geq 3$ events used in the inversion. We suggest that the magnitude-sensitivity of c, and subsequent impacts on the Omori decay observed at short timescales, should be further investigated. Variation of c may also impact the foreshock-aftershock ratio (Sornette and Werner, 2005; McGuire et al., 2005; Brodsky, 2011). Unlike Brodsky (2011), we correct for short-timescale catalogue incompleteness in our inversion for c, and run our calculations for a number of M_{min} cases which should alleviate some concerns around catalogue incompleteness.

Our observations lead us to propose that the nucleation of earthquakes occurs via stochastic triggering in an external process mediated environment. A next step would be to probe the spatial relationships of foreshocks, mainshocks and aftershocks on fault planes to understand the patterns of stress transfer and interaction, to better understand the relative importance of these two processes.

5.9 Conclusions

The nature of earthquake nucleation remains unclear. It is difficult to constrain the processes involved, be they simple or complex, because there are limited high-quality observations of foreshocks and aftershocks. Here we have made new observations of foreshocks and aftershocks around $M \geq 3$ mainshocks on the Hikurangi subduction zone. We used a template-based coherence approach to detect these small earthquakes. We have used a robust statistical approach to account for

false detections by correcting for a background seismicity rate which incorporates a temporally-stationary false detection rate.

We have found that the foreshock:aftershock ratio on the Hikurangi margin can be explained by ETAS models, but that the models underestimate the absolute numbers of foreshocks and aftershocks per mainshock. The seismicity rate decays as a power law, better fit by Omori's law with $p = 0.56$ than by $p = 1$, and the ETAS models do not produce a clear power-law increase or decay in the short windows around the mainshock that we consider. These observations lead us to suggest that earthquake nucleation is explained by stochastic triggering in an external process mediated environment such as aseismic slip, changes in pore fluid pressure for example due to fluid migration, or a combination.

5.10 Open Research Section

Processing was primarily in Python 3.7 and was undertaken using Scipy (Virtanen et al., 2020), Numpy (Harris et al., 2020), and Obspy (Beyreuther et al., 2010). Plotting used Matplotlib (Hunter, 2007). Data was downloaded using obspyDMT (Hosseini and Sigloch, 2017) and collected on the New Zealand seismic network, NZ. Code is available here: `https://github.com/RebeccaColquhoun/Colquhoun_and_Hawthorne_earthquake_precursors` and data was downloaded from the IRIS data repository.

6 Discussion and Conclusions

6.1 Aims and Contribution

In the introduction, I outlined three aims for each chapter. I begin by reflecting on these aims before posing future work which would directly build on the work of each chapter.

6.1.1 Chapter 3: Correlations between earthquake magnitude and early features of rupture in a global dataset

6.1.1.1 Reflection and review

Aim 1 Collate a dataset of $M \geq 3$ earthquakes from around the world using an internally consistent approach. Calculate four parameters (predominant period, average period, peak ground displacement, integral of the velocity squared) on short windows of the seismograms.

I collated a dataset of 4143 $M \geq 3$ earthquakes from around the world, between 2005 and 2020. I used EQTransformer, a machine learning picker, to pick the P wave arrivals on high signal-to-noise recordings of these earthquakes, with source-receiver distances of less than 100 km. This is the first time that an ML picker has been used in a determinism study, and this facilitates the use of such a large, global dataset. I calculated the predominant period, average period, peak ground

displacement and integral of the velocity squared in windows of 0.3, 0.5, 1 and 4 s after the P wave arrival. I also shifted these windows by ± 0.1 s to understand the sensitivity of our results to the P wave pick time.

Aim 2 Test whether these early features of seismograms are predictive of final earthquake magnitude, and if so, the relative duration of the window in which these features are required to be measured.

I found that all four parameters have a relationship with magnitude when calculated in windows less than one-third of the earthquake duration. I calculated statistics of the relationship between the parameters and magnitude as I gradually increased the minimum magnitude of the dataset. I describe the minimum subset magnitude at which the relationship between the parameter and magnitude is no longer correlated as the 'loss of significance magnitude', as distinct from just identifying the point at which the data deviates from a regression line as is usually done. The weakest relationship is between predominant period and magnitude. There is no relationship between predominant period and magnitude for earthquakes over M5.1 in a 0.3 s window, and this does not significantly change as the window lengthens (M5.3 for 4 s window), where the window is no longer significantly longer than the events under consideration. As the window of calculation is extended, average period and peak ground displacement both have an increasing magnitude at which significance is lost, which corresponds to earthquakes with durations more than three times the window length. The integral of the velocity squared has a significant relationship with magnitude even in windows just one-tenth of the earthquake duration. The loss of significance could be a real feature but may simply result from a reduction in data density at high magnitudes. The loss of significance magnitude is not impacted by the window length.

Aim 3 Pose physical explanations for our observations of early seismogram parameters which can be input into earthquake models.

In response to our observations, I propose for the first time that predominant period may be a measure of the time to rupture the initial asperity. For small

earthquakes, this plays an important role in determining the final magnitude of the earthquake, but becomes increasingly less important at higher magnitudes. Relatedly, I propose that the average period may be a measure of all the asperities ruptured during the window of measurement which could better account for complexity in the earthquake rupture. Therefore, it maintains some relationship with magnitude for larger earthquakes and longer windows. As has been previously identified, I suggest that IV2 is related to the energy radiated by the earthquake, and thus that larger earthquakes tend to start more energetically because they are more likely to be able to overcome stress barriers. Finally, I propose that the peak ground displacement is likely a measure of the early moment release in the earthquake, and thus can make similar energetic arguments as I did for IV2. I do not observe any short, low-moment rate nucleation phases.

6.1.1.2 Future Work

I have shown that early parameters of earthquakes scale with magnitude using a large global dataset. However there is significant scatter in my results, which likely originates from intrinsic differences between earthquakes due to things like tectonic setting. As my next step, I would subdivide my dataset by tectonic setting and see if the variance of the data is reduced. I would also conduct some regional studies using the same unified approach that I have used for the global dataset. This would allow for a better attenuation correction including regional scale maps of Q.

Finally, I would look for multivariate relationships. I avoided these during this study because they provide fewer physical insights and I do not know whether the parameters are independent, or just different measures of the same physical process. If it is possible to account for auto-correlation, multi-variate regressions may make my results more useful for hazard analysis and earthquake early warning. However, this requires us to first test the understanding of the physical meanings I propose for each parameter using source models.

It would also be interesting to model how well IV2, which is understood to be related to the radiated seismic energy, gives insights into the total energy budget of

an earthquake since, as I reviewed in Chapter 2, radiated energy is not the only form of work done during an earthquake. It would be interesting to simulate various source models to understand how well IV2 (and thus the radiated energy) represent the total energy budget of the earthquake, and whether this can give insights into if IV2 is better suited as a metric for some earthquakes.

6.1.2 Chapter 4: Patterns in source time functions reflect variations in earthquake magnitude and inversion approach

6.1.2.1 Reflection and review

Aim 1 Collate source time functions from five different databases of source time functions that are based on different inversion approaches.

I collated a dataset of 11471 STFs from 4888 events and 5 databases. These methods used a range of approaches to invert for the STF and thus impose different assumptions on the resulting STF. I explored different ways to automatically define the end of an STF in order to exclude water and sediment multiples and artificially long tails.

Aim 2 Identify and quantify source time function features and test whether these scale with earthquake magnitude.

I identified a range of features of STFs and find that many of these scale with earthquake magnitude. Particularly, I found that the moment released in the first seconds of the STF and the initial moment-acceleration rate both scale with the final earthquake magnitude. However, it remains uncertain the degree to which earthquakes have a slow nucleation phase, as there is a lot of variation between events and datasets in the very earliest portion of the STF. I identified that a trapezium geometry best fits the median STF, with a period of early moment acceleration followed by a period of releasing moment at a constant rate and finally a reduction in the moment-release rate as the earthquake ends. The relative durations of each of these stages does not depend on magnitude. Finally, I quantified the complexity of the STF, and found that a majority of STFs were not well fit by a

single peak Gaussian function, which has important implications for earthquake source modelling. Any patterns of STF complexity with magnitude were outweighed by variations between datasets.

Aim 3 Examine the impact of dataset (and thus inversion method and assumptions) on source time function characteristics.

For the first time, I was able to systematically compare STFs between different databases and inversion schemes. I found that the complexity of the STF is heavily influenced by the inversion approach used. Some databases (e.g. Sigloch) produce mostly very simple STFs, whilst others (e.g. SCARDEC) produce both simple and complex STFs. Some databases also show a trend of increasing complexity with magnitude, whilst others show a complex or decreasing complexity with magnitude. I also see that STFs from some databases (e.g. Sigloch) release their moment much more quickly than others (e.g. USGS). The median STF for the ISC dataset looks manifestly different than all the other databases.

6.1.2.2 Future work

The design of this chapter aimed to maximise the number of STFs available, in order to make our conclusions as robust as possible. However, as with Chapter 3, one future development could be to look at variation in STFs with magnitude and dataset using a subset of events in a specific region or tectonic setting (e.g. subduction zones). This may help clarify the drivers of STF variability. I also propose that further work should be undertaken on the causes of inter-dataset variation. One way to approach this would be to use the different STF inversion methods on synthetic waveform data, which would allow us to assess the fidelity of the resulting STFs from different inversion schemes. That early moment release scales with final magnitude suggests that large earthquakes may begin more energetically, and thus are more able to overcome stress barriers. It would be interesting to use source models to understand whether this theory is true, and how different the early stages of rupture need to be in order to be distinguishable in the STF. Finally, I propose further work on the reliability of the early stages of STFs through making use of the

uncertainties which inherently arise from the ISC inversion approaches, but which are not routinely published. It may be possible to add some measure of uncertainty to the other inversion schemes or at least to examine the limits of the parameter space that produce reasonable STF estimates for different earthquakes.

6.1.3 Chapter 5: Using foreshocks to constrain earthquake nucleation: comparing foreshock to aftershock ratios in the Hikurangi subduction zone to results from ETAS

6.1.3.1 Reflection and review

Aim 1 Make new observations of foreshocks and aftershocks around $M \geq 3$ earthquakes on the Hikurangi subduction zone, New Zealand.

Using phase coherence, I made over 1 million new detections of earthquakes in the hour around $M \geq 3$ mainshocks on the Hikurangi subduction zone. I approached correcting for background seismicity and false detections in a novel manner which results in a lower magnitude of completeness (∼M0.6). After this correction, I made 11,233 detections of foreshocks and aftershocks.

Aim 2 Analyse temporal patterns in the detected foreshocks and aftershocks in the hour around $M \geq 3$ mainshocks.

I stacked our observations and saw a power-law increase in seismicity in the seconds before the mainshocks and a power-law decay after the mainshock. This was not reproduced by an ETAS simulation that had its parameters tuned using the $M \geq 3$ seismicity in the region.

Aim 3 Compare the foreshock-aftershock ratio and absolute numbers of foreshocks and aftershocks between the observations and ETAS simulations.

At $\alpha = 0.05$, the ETAS model can explain our observed foreshock-aftershock ratio, but I see far more foreshocks and aftershocks per mainshock than the ETAS model predicts. This could suggest that c has some magnitude dependence, since the posterior distribution for c is well constrained and is found using a well-justified

approach. I propose that the earthquake nucleation I observe is best explained by stochastic earthquake-earthquake triggering occurring in an external process mediated environment. I do not constrain whether this external process is aseismic slip, fluid flow or a combination.

6.1.3.2 Future work

Our approach to false detections and background seismicity allows us to detect seismicity down to ∼M0.6. However, it means that I cannot determine the location or magnitude of specific foreshocks or aftershocks. Future work could use different methods to identify seismicity and therefore be able to track the spatial progression of seismicity before and after mainshocks. This might elucidate the relative role of fluid flow and aseismic slip on earthquake nucleation. I also suggest that the magnitude sensitivity of c be further investigated, as this may impact the agreement of the observed foreshock-aftershock ratios and those arising from an ETAS simulation. Finally, it would be interesting to use synthetic waveforms to understand the impact of earthquake determinism (or lack thereof) on our ability to use phase coherence to detect foreshocks and aftershocks. One avenue to exploring this would be to vary the window lengths in which the phase coherence is calculated.

6.2 What is next for earthquake mechanics?

I have discussed the natural progression for the research presented in each chapter. In this section I ask more broadly what is next for earthquake mechanics and what, as a field, do we need to make it happen?

A major limitation in our ability to measure the earliest stages of earthquake rupture is the sampling rate of our seismometers. The waveform based analysis in this thesis used at most 100 Hz data. This means that I can only measure features of up to 50 Hz. In reality, this was very difficult to do, and once I required a period of waveform to analyse (Chapter 3), I could not analyse windows shorter than 0.3 s or very small earthquakes ($M < 3$). Many seismometers are 20 or 40 Hz, and thus were unsuitable for my analysis. I propose densely instrumenting

a few particularly seismically active regions with more 100 Hz seismometers and much higher frequency (up to 1000 Hz) borehole seismometers. This would allow us to gain better insights into the early phases of earthquakes and to test whether high-sample rate instrumentation allows for better early warning capability.

A recent advance in seismic instrumentation comes from seismic nodes: small, easily deployed instruments which allow incredibly dense deployment and deployment in difficult conditions. "Large-N" seismology in seismically active areas would facilitate very near source observations of seismicity potentially with inter-station spacing that is less than the rupture dimensions. For small earthquakes, which are normally considered as a point source this would allow us to understand the evolution of rupture on the fault. By associating on/near fault observations with early waveform parameters, it might be possible to observationally test my proposed explanations for the different waveform parameters I calculated.

Dense seismic networks also allow the reduction in the minimum magnitude of seismic detections, and an increase in accuracy of earthquake locations. This would facilitate the spatial tracking of seismicity before large earthquakes, since I expect different patterns of seismicity between fluid mediated seismicity and aseismic slip induced seismicity.

Another development in instrumentation has been the advent of Distributed Acoustic Sensing (DAS). Yin et al. (2023) has already tested the DAS for early warning in offshore seismicity and discovered that it can increase warning times by up to 3 s. DAS allows measurements of strain rate at very high spatial resolution whilst still being practical over large distances. Whilst DAS has a lot of potential for giving accurate earthquake location estimates, at the moment most locations are ambiguous due to the linear geometries involved. One solution for this is time-domain beam-forming and phase arrival-time difference on much shorter optical fibre lengths, which can give much more accurate epicentre locations and is viable for real-time EEW (Ben-Zeev and Lior, 2025). Increasingly, I think we will also see the integration of geodetic data with seismic data for earthquake monitoring and mitigation.

So far, I've discussed some ideas for what the bigger and better instrumentation looks like. However, there is also significant value in long-term datasets. Being able to make statistically robust findings relies on having lots of data. There was a step change in instrumentation in 2005, and so every year that goes by, there are more earthquakes to analyse and our understanding can only increase. Rolling out our current instrumentation levels to more earthquake-prone regions of the world will help facilitate understanding of the regional differences in seismicity.

I have discussed what advances in natural earthquake observation I hope to see. However, I think that advances in lab seismicity will also play a significant role in our understanding of earthquake nucleation. For example, Song and McLaskey (2024) showed that interrupted events look closer to natural earthquakes than complete ruptures. Therefore I might ask how different types of 'interruption' impact our observations of seismicity and how parallels can be drawn to natural features. It will also be interesting to explore how different lab-fault materials impact our observations of nucleation phases (I have seen conference talks using foam as the fault surface and even rice!).

Lab studies also form part of asking how our observations scale to smaller earthquakes, alongside induced seismicity. Fluids play an important role in microseismicity nucleation, and have been observed to mediate earthquakes induced by fluid injection (Ellsworth, 2013; Foulger et al., 2018; Keranen and Weingarten, 2018). Can observations of earthquakes where fluids are known to be involved be used to better understand if earthquakes look different when fluids are involved? It may be that there is some interplay between the degree to which earthquakes are deterministic and the processes involved in their nucleation. Icequakes are another natural setting to look at this: in glacial settings, water is prevalent and the temporal and spatial variations in water content are better resolved. Thus far, most research that has investigated the interplay between fluid flow and glacial seismicity has focused on icequake rates and locations (Podolskiy and Walter, 2016). But glacial seismicity can also act as a natural laboratory to understand whether the intrinsic nature of the seismicity is changed by fluid involvement.

Whilst understanding earthquake processes is integral to building earthquake early warning systems, I believe that the key to reducing the risk that earthquakes present is in preparedness and resilience building. This includes having an earthquake building code and ensuring buildings are up to code, as well as public awareness of emergency response procedures and mitigating actions they can take. Even in California, where ShakeAlert sends notifications to everyone's mobile phones, early warning can't help if people do not respond to the alert.

6.3 Concluding remarks

I started my PhD thinking that it would be really helpful for earthquake early warning. It turns out that doing statistics well often means that the grand, revolutionary results found by previous authors are much more subtle than expected, and with much more variation. However, I have robustly shown that windows of the waveform less than one-third of the earthquake duration contain information on the final magnitude of the earthquake. Indeed the integral of the velocity squared parameter, which is related to radiated energy, can help constrain the magnitude when only 10% of the earthquake has occurred. This suggests that earthquakes are deterministic and that large and small earthquakes develop differently in their early stages. I then proposed physical mechanisms to explain our observations.

I have shown that large and small earthquakes release their moment differently using both waveforms and source time functions. Both of these show that large earthquakes release early portions of their moment more quickly. Using source time functions, I also found that STFs are generally complex with multiple peaks, and have several clear phases of rupture. This has important implications for earthquake source modelling.

All of this work requires earthquakes to have occurred. I therefore closed by investigating whether earthquake-earthquake triggering could explain foreshock occurrence on the Hikurangi subduction zone, New Zealand. I found that the foreshock-to-aftershock ratio observed is explained by earthquake-earthquake triggering models, but that the model underestimates the number of foreshocks and

aftershocks I observe per mainshock and does not reproduce the power-law increase and decrease in seismicity I observe around the $M \geq 3$ mainshocks. Therefore earthquake nucleation must be more complex, with stochastic triggering mediated by fluid flow or aseismic slip.

Appendices

A

Supplementary Information for Chapter 3

Contents

A.1 Literature review

In Table A.1, we present a literature review of papers which investigate the relationship between our parameters and magnitude. We highlight the number of earthquakes used in their analysis.

Table A.1: Review of previous studies. Unless otherwise specified N refers to number of earthquakes included.

Citation	Region	N	Magnitude range	Methods
Allen and Kanamori (2003)	Southern California	53	3–7.3	Predominant period
Allen (2003)	California	53	3.0–7.5	Predominant period
Kanamori (2004)	California, Japan and Taiwan	23	2.5–8	Average period
Olson and Allen (2005)	California, Japan, Taiwan, Alaska	71	3–8.3	Predominant Period
Lockman and Allen (2005)	California	50	3.0–7.3	Predominant period
Wu and Kanamori (2005a)	Taiwan	26	5.0–7.6	Average period, Peak ground displacement
Wu and Kanamori (2005b)	Taiwan	26	5.0–7.6	Average period, Peak ground displacement
Rydelek and Horiuchi (2006)	Japan	50+	6+	Predominant period (negative)
Wu et al. (2006)	Taiwan	46	4.0–7.6	Average period, Peak ground displacement
Wu and Zhao (2006)	Southern California	25	4.0–7.1	Peak ground displacement
Lockman and Allen (2007)	Pacific Northwest, Japan	86	3.5–6.8	Predominant period
Rydelek et al. (2007)	Japan	1363	3.5+	Peak ground displacement
Tsang et al. (2007)	Southern California	59	3.0–7.3	Predominant period, Peak ground displacement
Wu et al. (2007)	Southern California	27	4.0–7.3	Average period, Peak ground displacement
Wurman et al. (2007)	Northern California	43	3–7.1	Predominant period, Peak ground displacement

Citation	Region	N	Magnitude range	Methods
Festa et al. (2008)	Japan	2000 records	4.0–7.0	IV2, $P_d^2/IV2$
Shieh et al. (2008)	Japan	16	6.0–8.3	Predominant period, average period
Wu and Kanamori (2008a)	Southern California, Japan and Taiwan	54	4.0–8.3	Average period, Peak ground displacement
Wu and Kanamori (2008b)	Japan	21	6.0–8.3	Average period, Peak ground displacement
Brown et al. (2009)	Japan	84	4–8	Predominant period, Max amplitude
Allen et al. (2009)	California	2 (realtime)	5.4	Predominant period, peak ground displacement
Brown et al. (2009)	Japan	84	4.0–8.0	Predominant period, peak ground displacement
Brown et al. (2011)	California and Japan	87	4–8	Predominant period, Peak ground displacement
Lancieri et al. (2011)	Chile	69	4.0–7.8	IV2, predominant period, average period
Kuyuk and Allen (2013)	California, Japan	174	3.0–8.0	Predominant period, Peak ground displacement
Meier et al. (2016)	Japan, China, Southern California, Western USA	3075 records	4+	Peak ground displacement (negative)

A.2 Magnitude-Duration Scalings

As outlined in Section 3, we can calculate the duration of average earthquakes compared to our window lengths. We give comparisons between average duration and magnitude in Table A.2 and Figure A.2.

Table A.2: Magnitude-duration comparisons.

	0.3 s	0.5 s	1 s	4 s
Whole earthquake	4.0	4.45	5.05	6.25
2 x earthquake	4.6	5.05	5.65	6.85
3 x earthquake	4.95	5.4	6.0	7.2
10 x earthquake	6.0	6.45	7.05	7.95

Note: 2×, 3× and 10× give the magnitude of the earthquakes which are at least 2, 3 or 10 times the length of the window (0.3, 0.5, 1, 4 s).

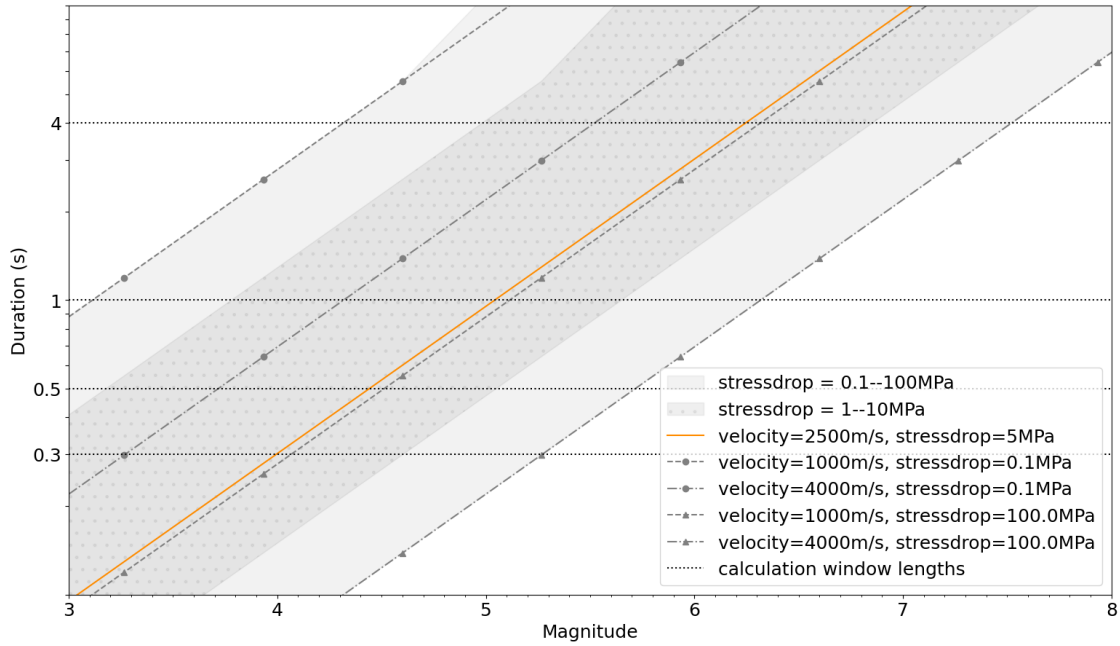


Figure A.1: Duration-magnitude scalings based on different values of rupture velocity (shape) and stress drop (linestyle). The orange line shows the scaling used in the main text. Lighter shading shows the variation possible within stress drop bounds of 0.1–100 MPa, and darker, dotted shading shows the variation expected within stress drop bounds of 1–10 MPa.

A.3 Data filtering

We calculate the signal-to-noise ratio of the data before filtering, and discard data with $SNR < 20$ (or where specified $SNR < 5$). We then filter the data as in Table A.3 before doing our computations.

Table A.3: Filtering parameters used

Method	Type	Frequency Range (Hz)	Corners
Predominant Period	Bandpass	0.1–19	4
Average Period	Highpass	0.078	3
IV2	-	-	-
P_d	Highpass	0.075	4

These filtering parameters are determined efforts to reproduce results from the literature. The highpass filter with three-corners at 0.078 Hz is similar to a four-corner filter at 0.1 Hz, but manual analysis showed that the 0.078 Hz filter more acccurately reproduced previously published results (see Table A.1).

A.4 Azimuthal Coverage

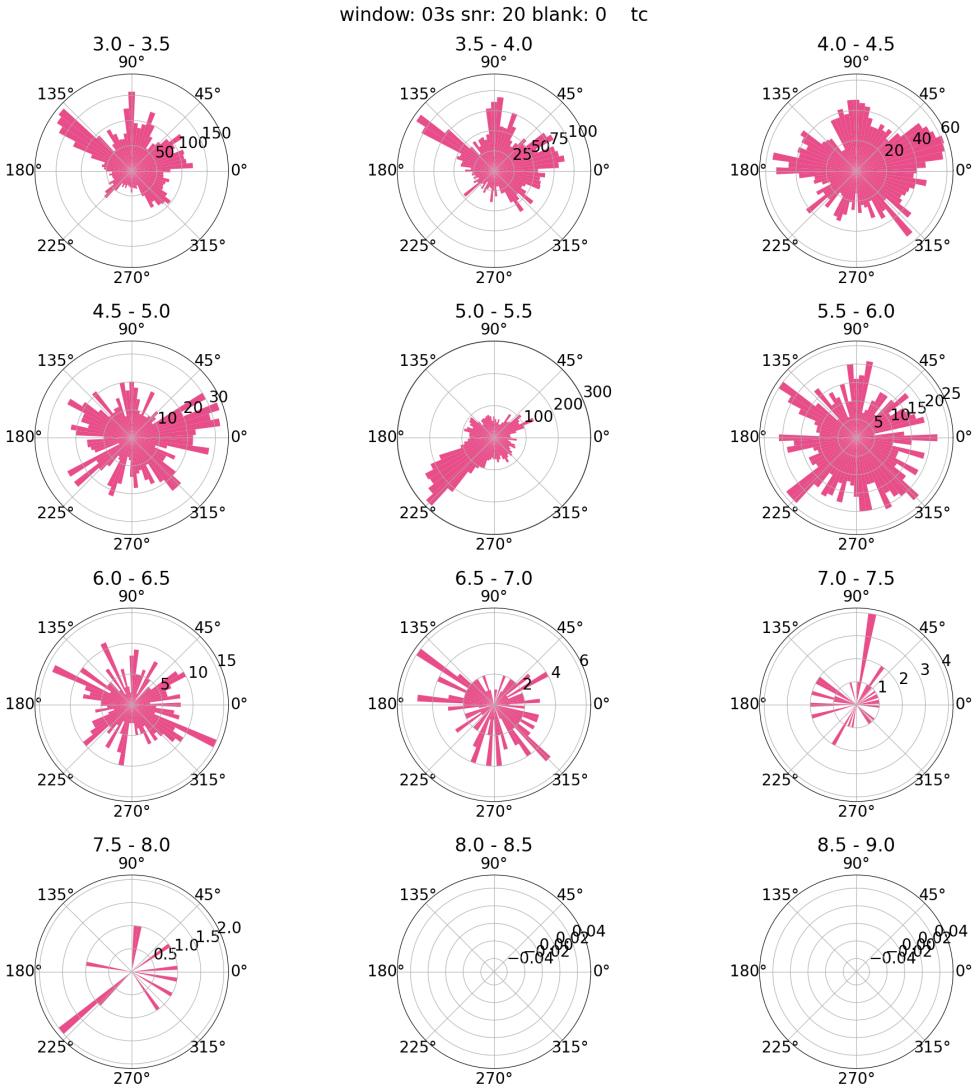


Figure A.2: As Figure 3.2 but for τ_c.

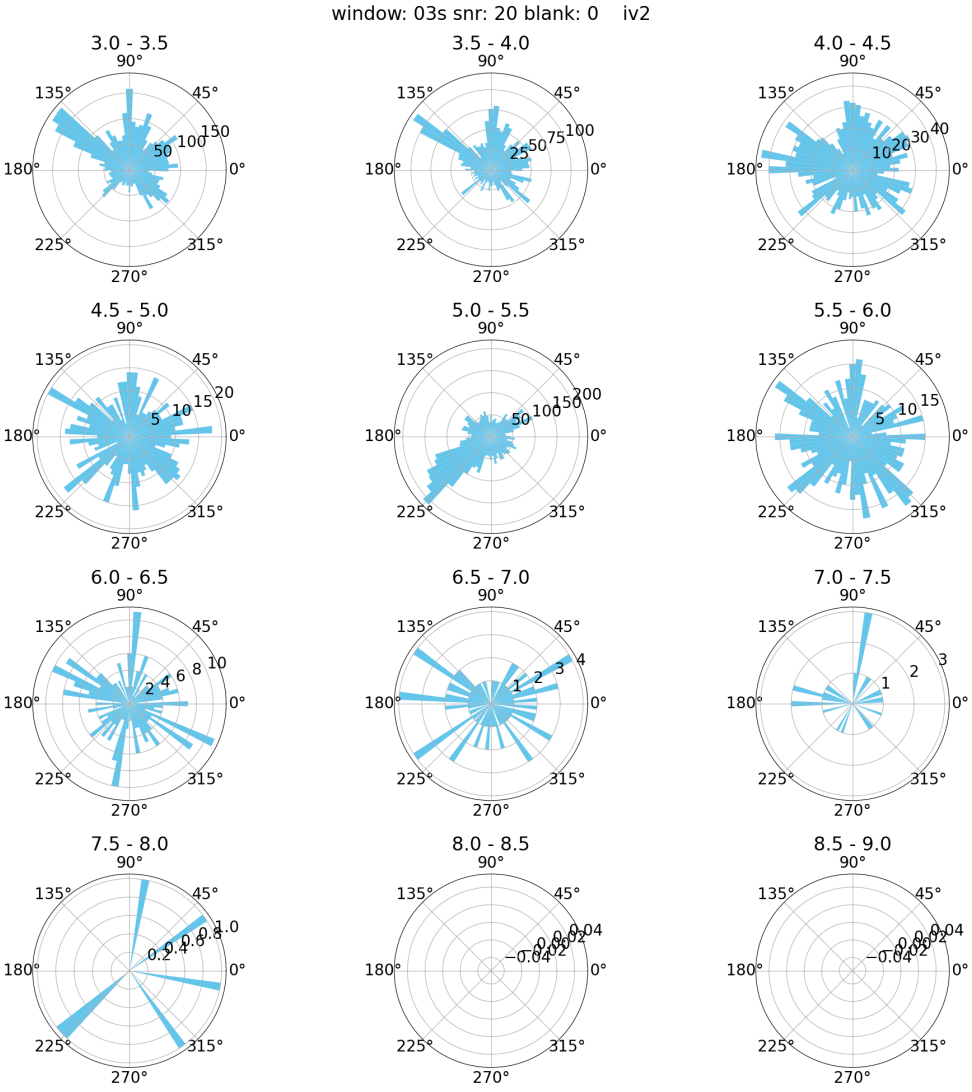


Figure A.3: As Figure 3.2 but for IV2.

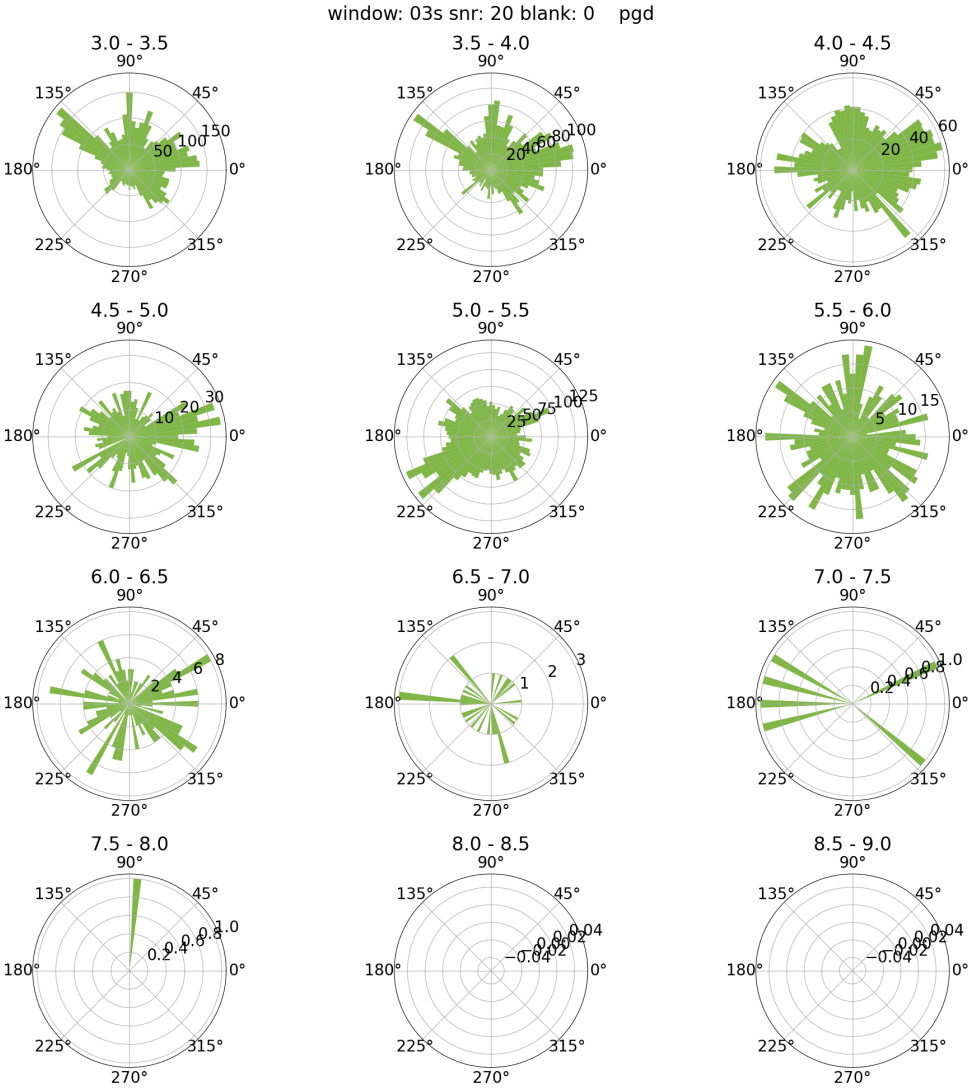


Figure A.4: As Figure 3.2 but for P_d.

A.4.1 Azimuthal gaps

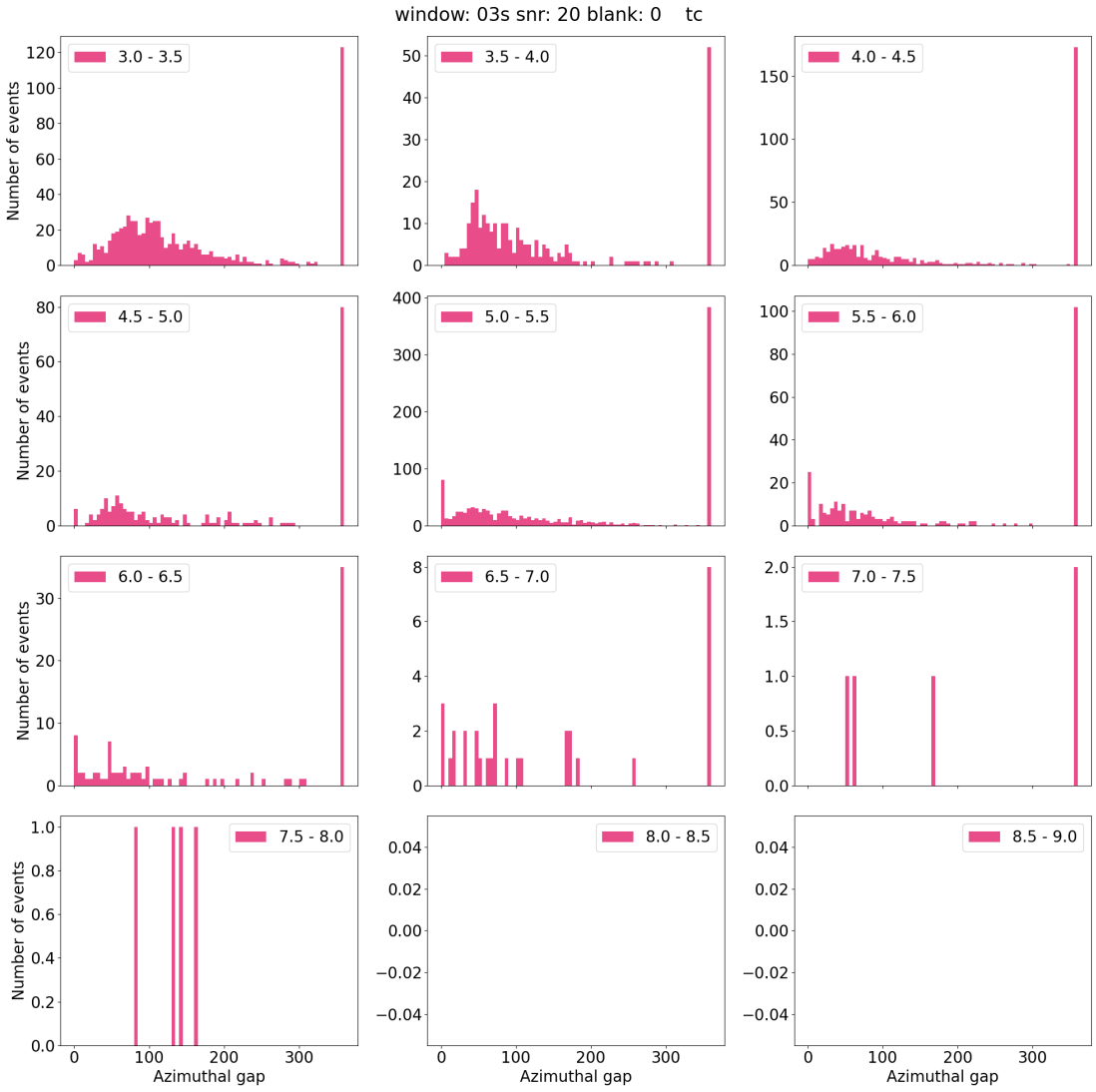


Figure A.5: As Figure 3.3 but for τ_c.

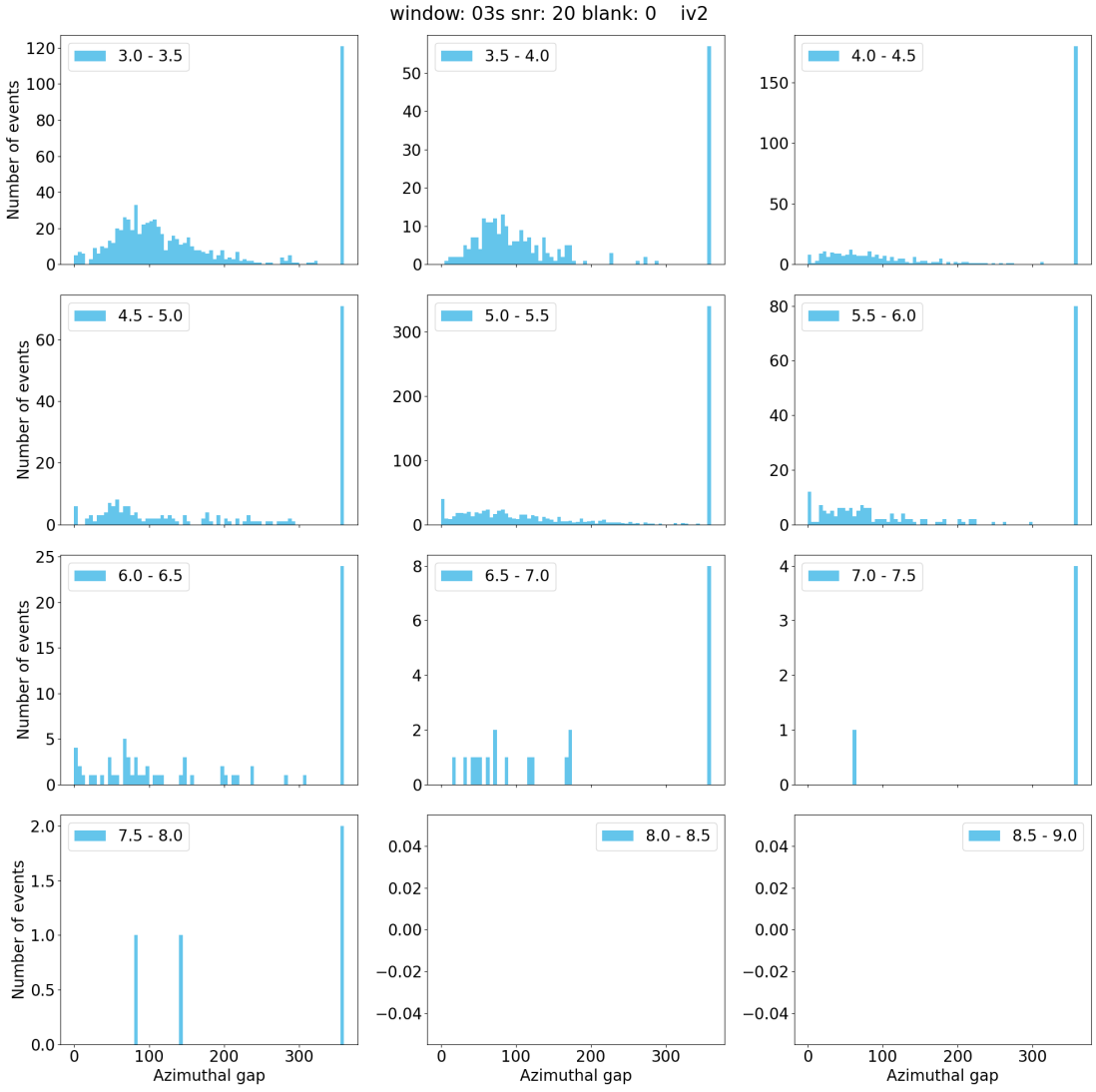


Figure A.6: As Figure 3.3 but for IV2.

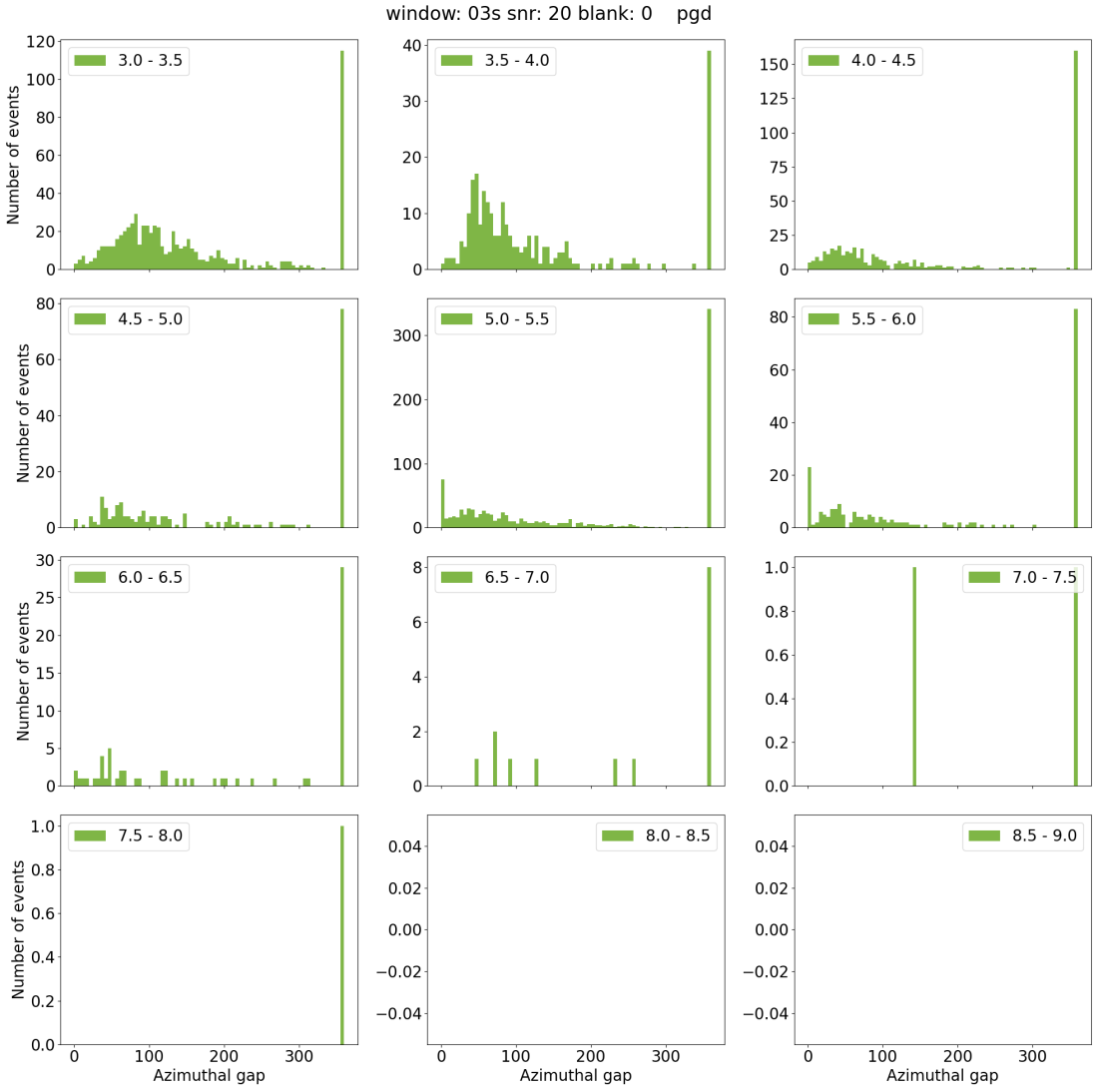


Figure A.7: As Figure 3.3 but for P_d.

A.5 Number of Stations used

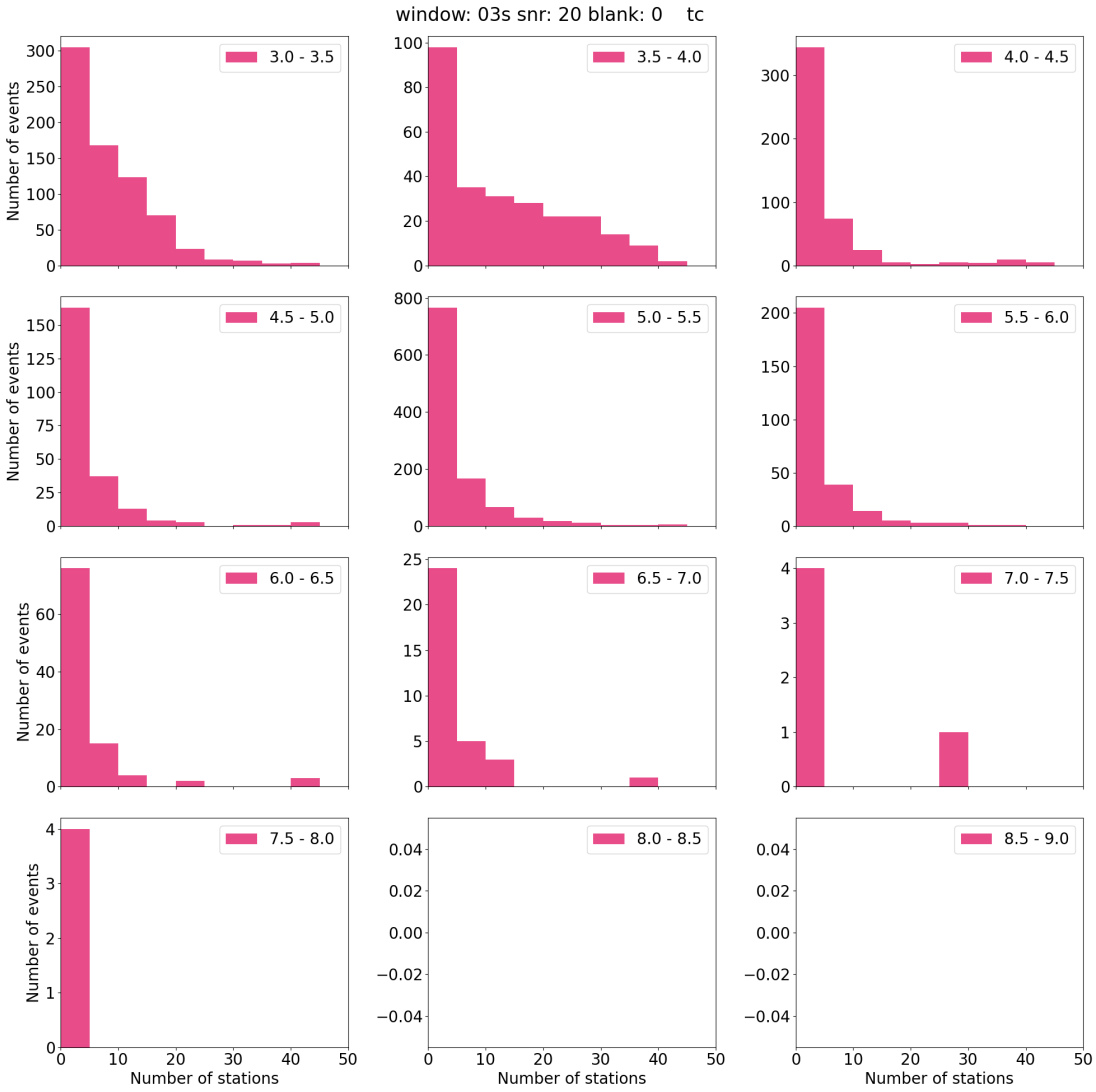


Figure A.8: As Figure 3.4 but for τ_c.

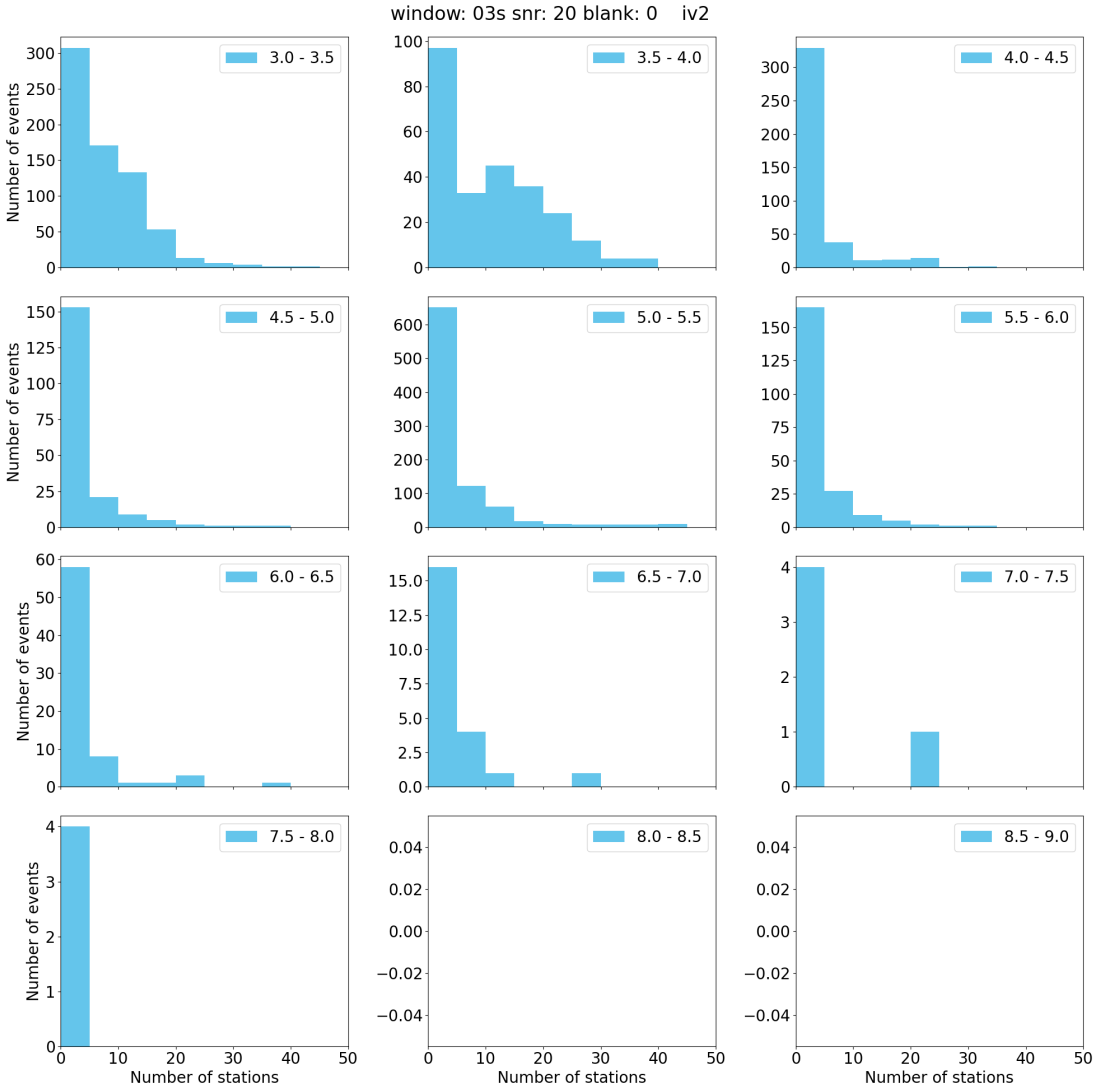


Figure A.9: As Figure 3.4 but for τ_c.

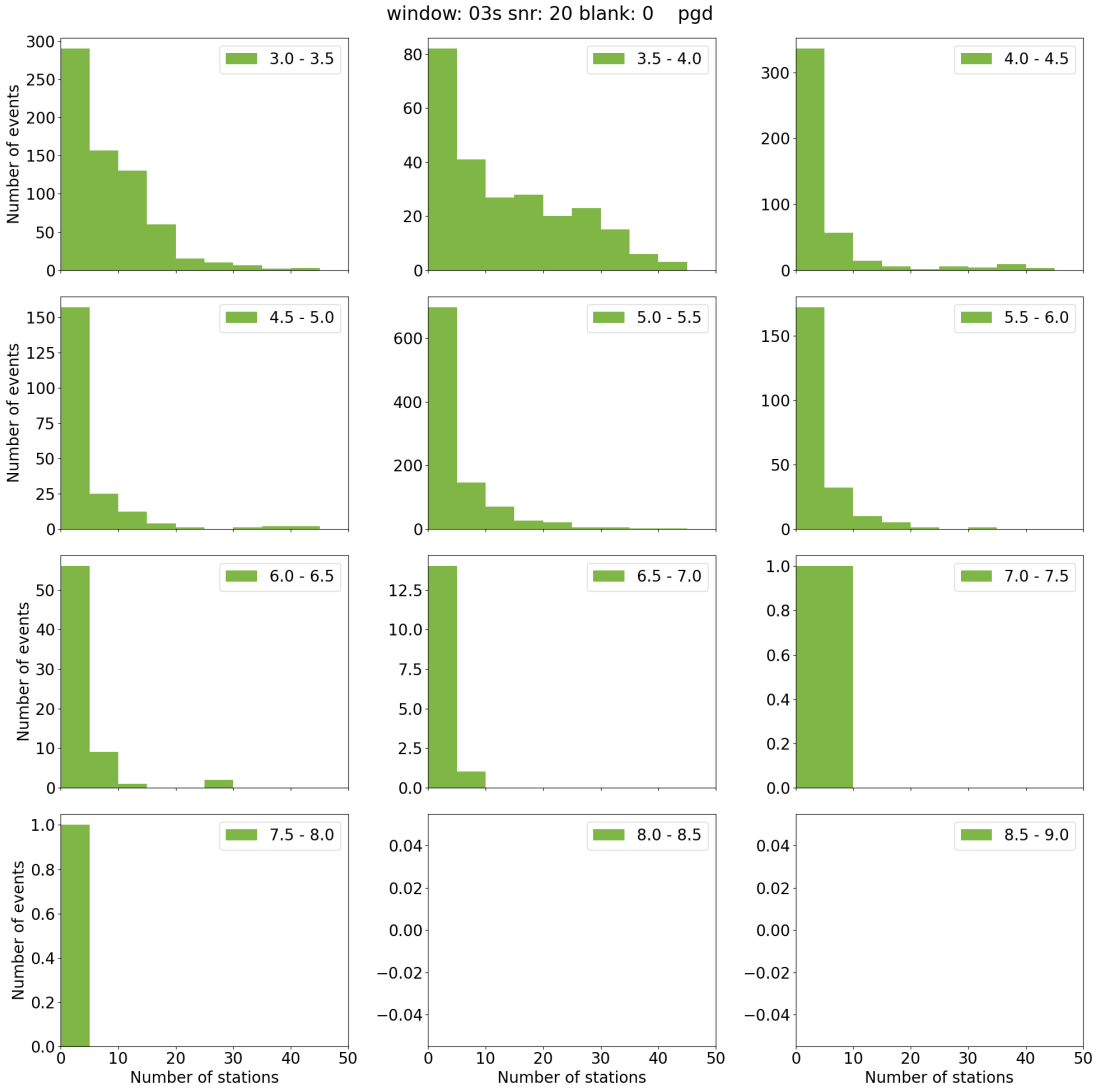


Figure A.10: As Figure 3.4 but for P_d.

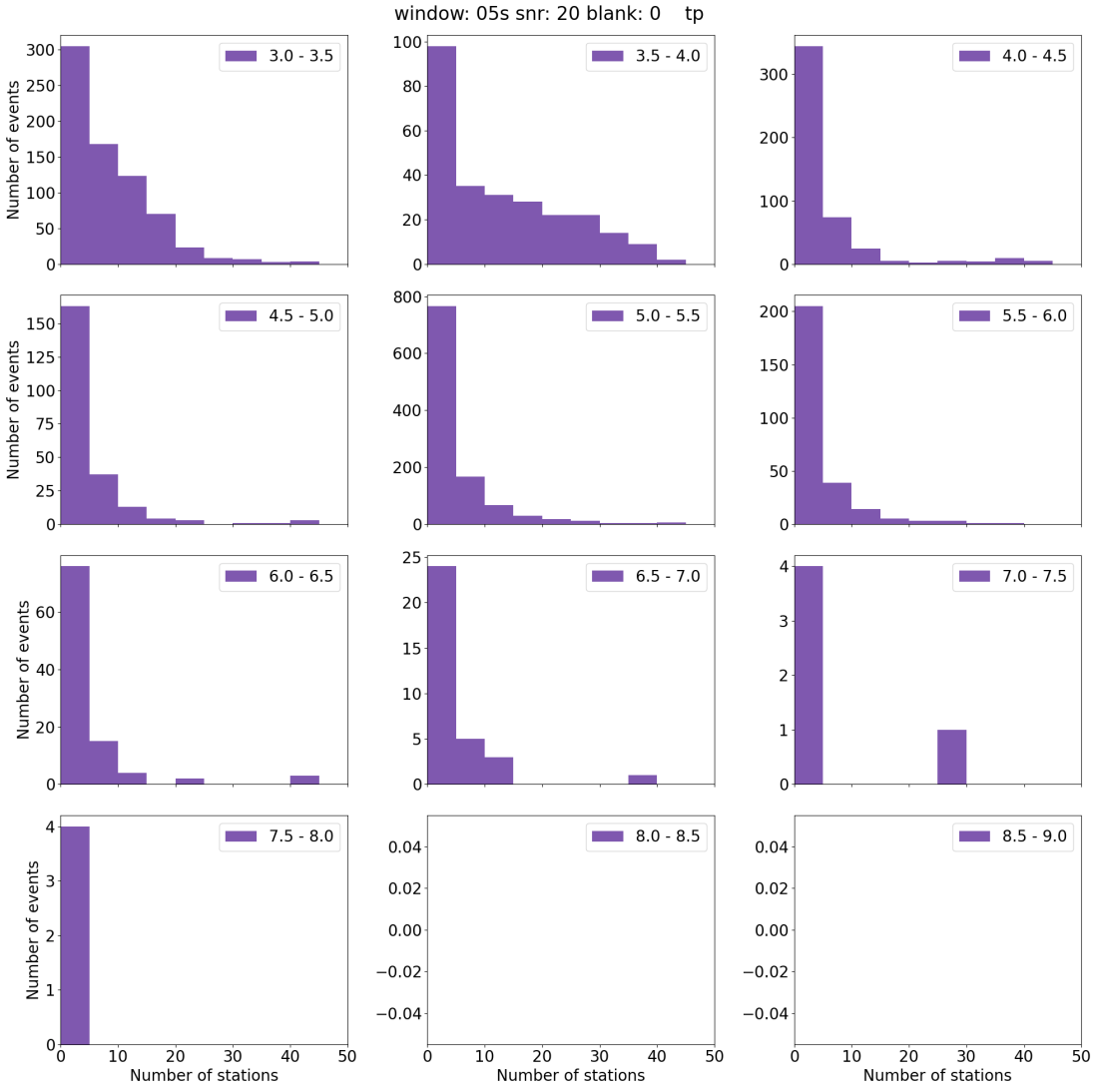


Figure A.11: As Figure 3.4 but calculated in a 0.5 s window.

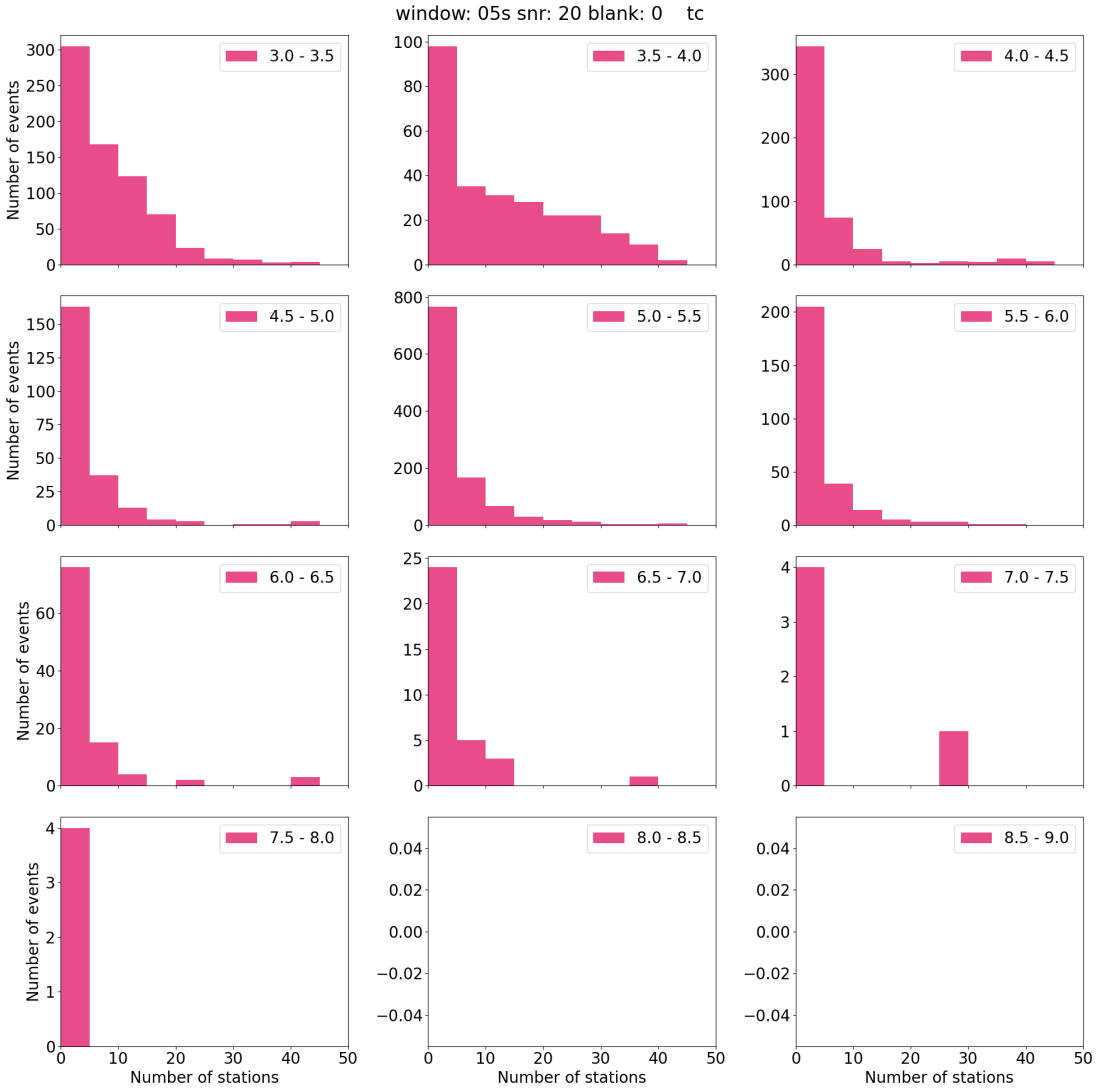


Figure A.12: As Figure 3.4 but for τ_c calculated in a 0.5 s window.

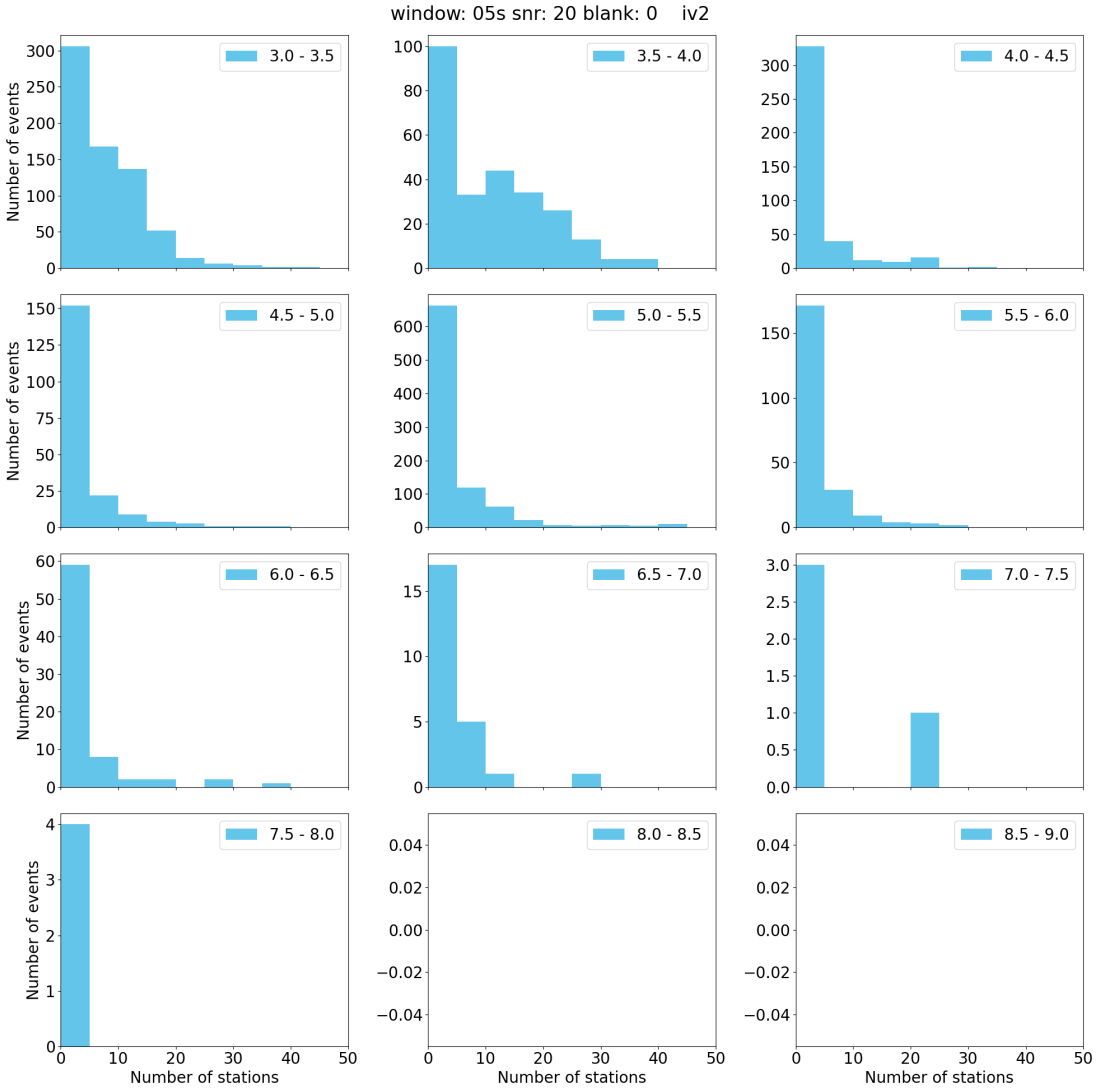


Figure A.13: As Figure 3.4 but for IV2 calculated in a 0.5 s window.

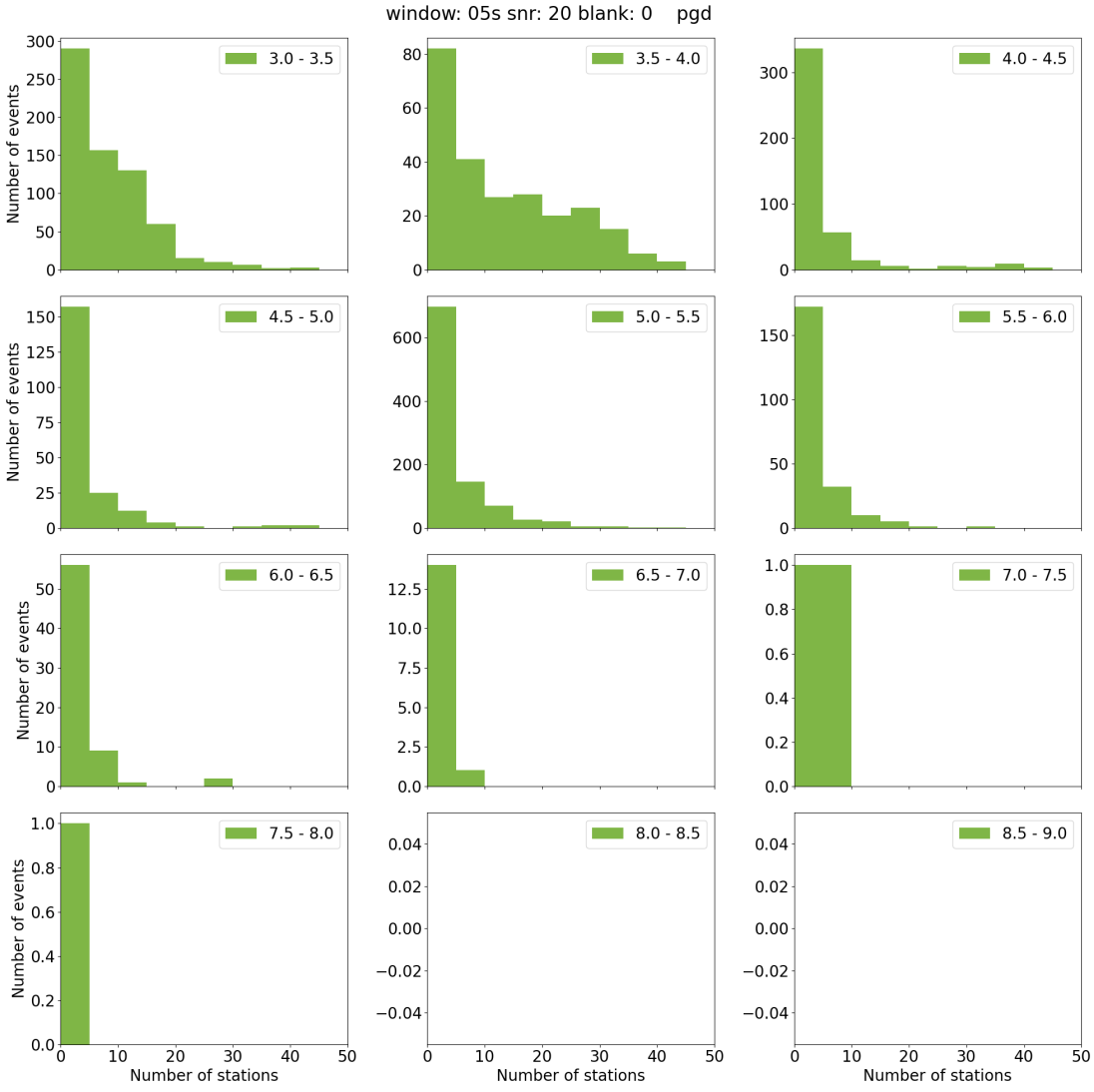


Figure A.14: As Figure 3.4 but for P_d calculated in a 0.5 s window.

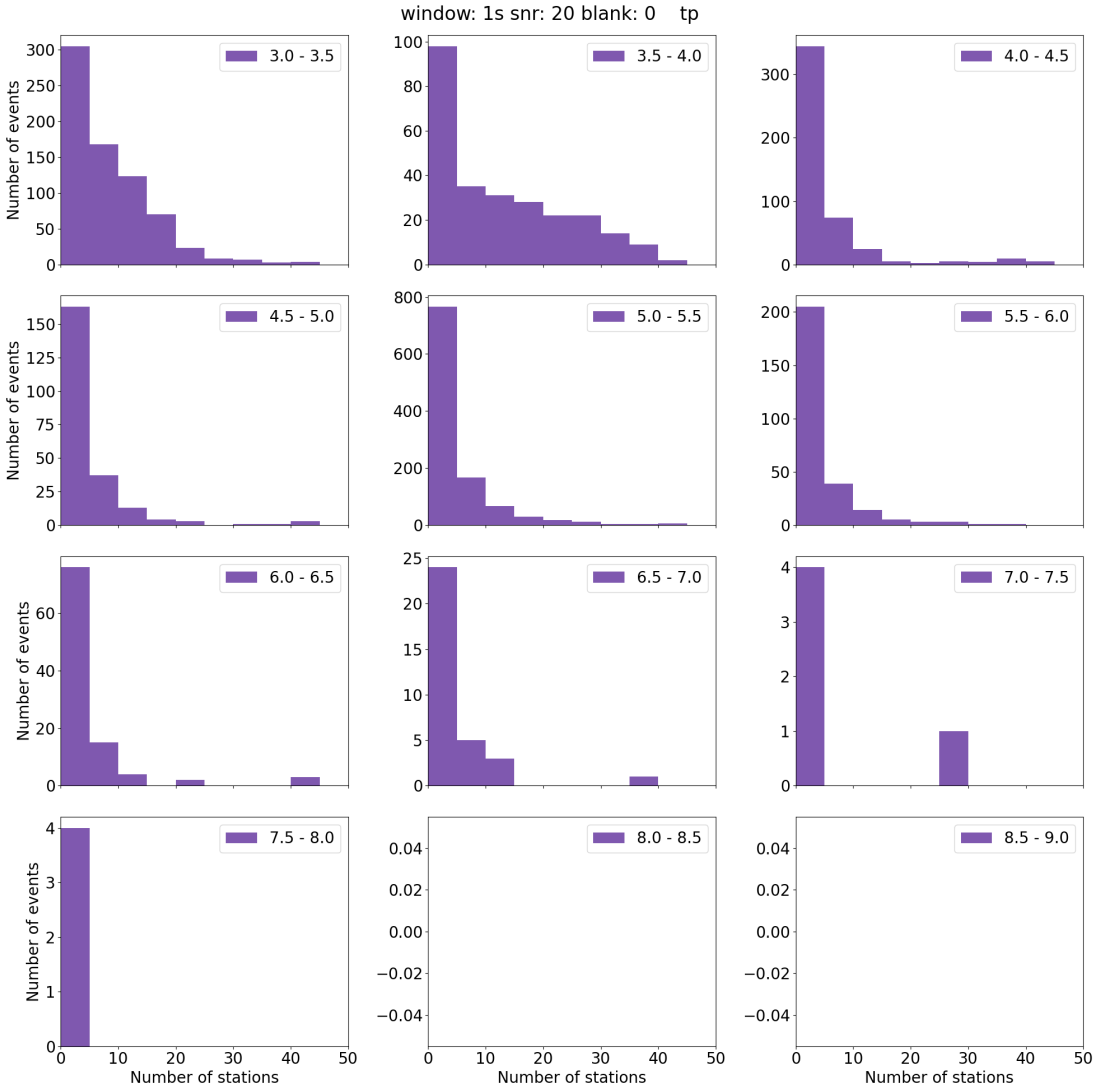


Figure A.15: As Figure 3.4 but for a 1 s window.

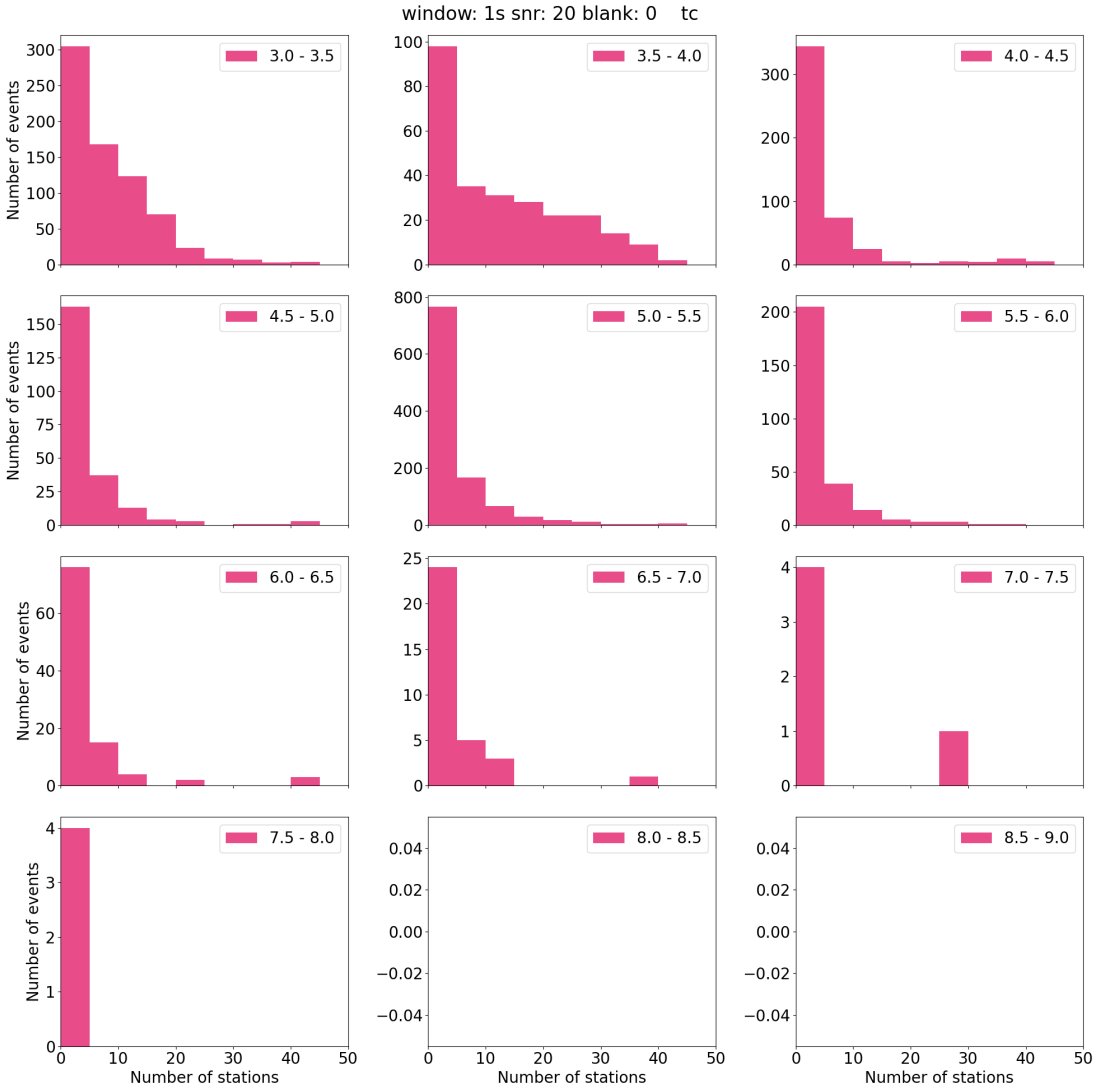


Figure A.16: As Figure 3.4 but for τ_c calculated in a 1 s window.

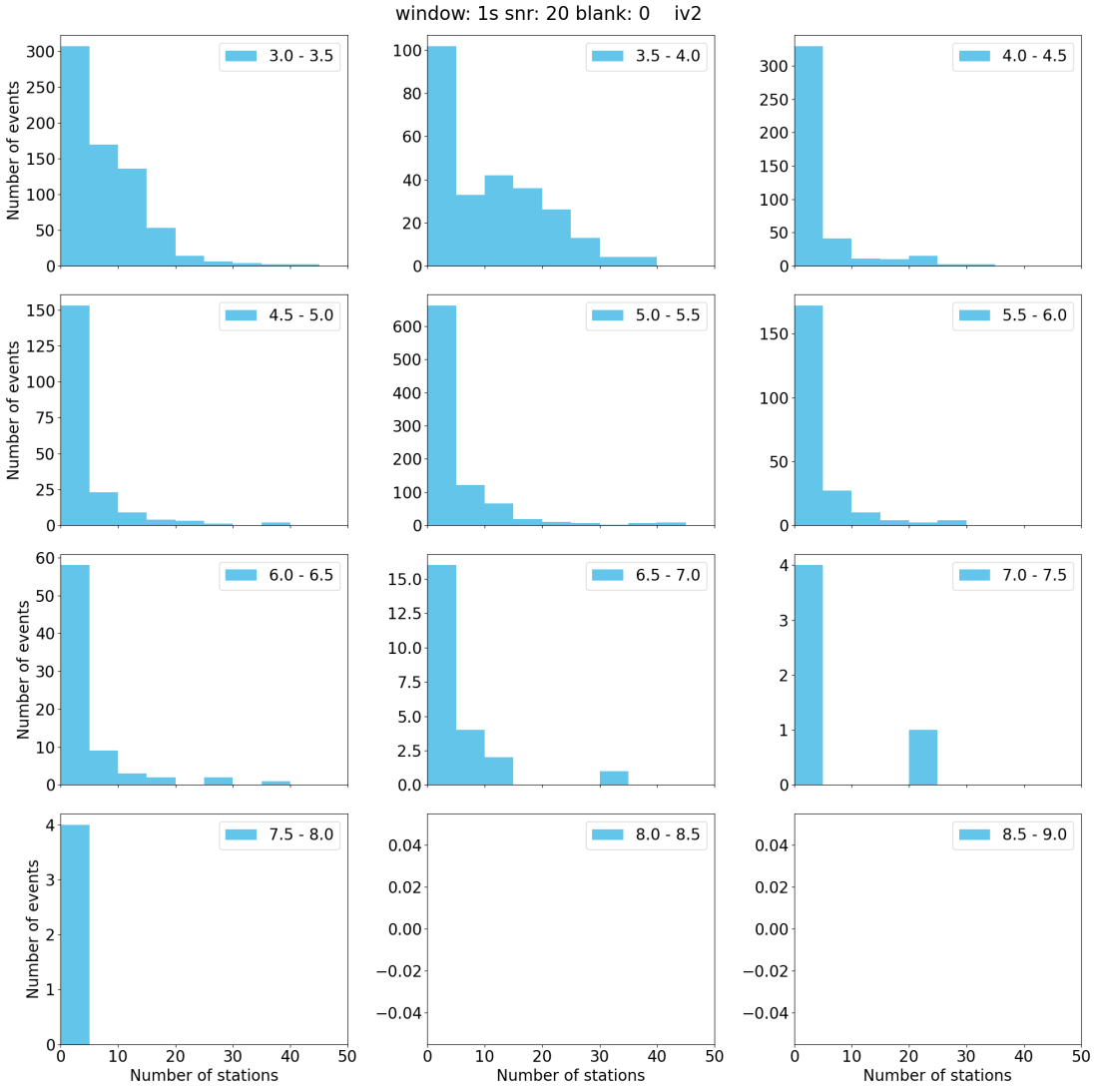


Figure A.17: As Figure 3.4 but for IV2 calculated in a 1 s window.

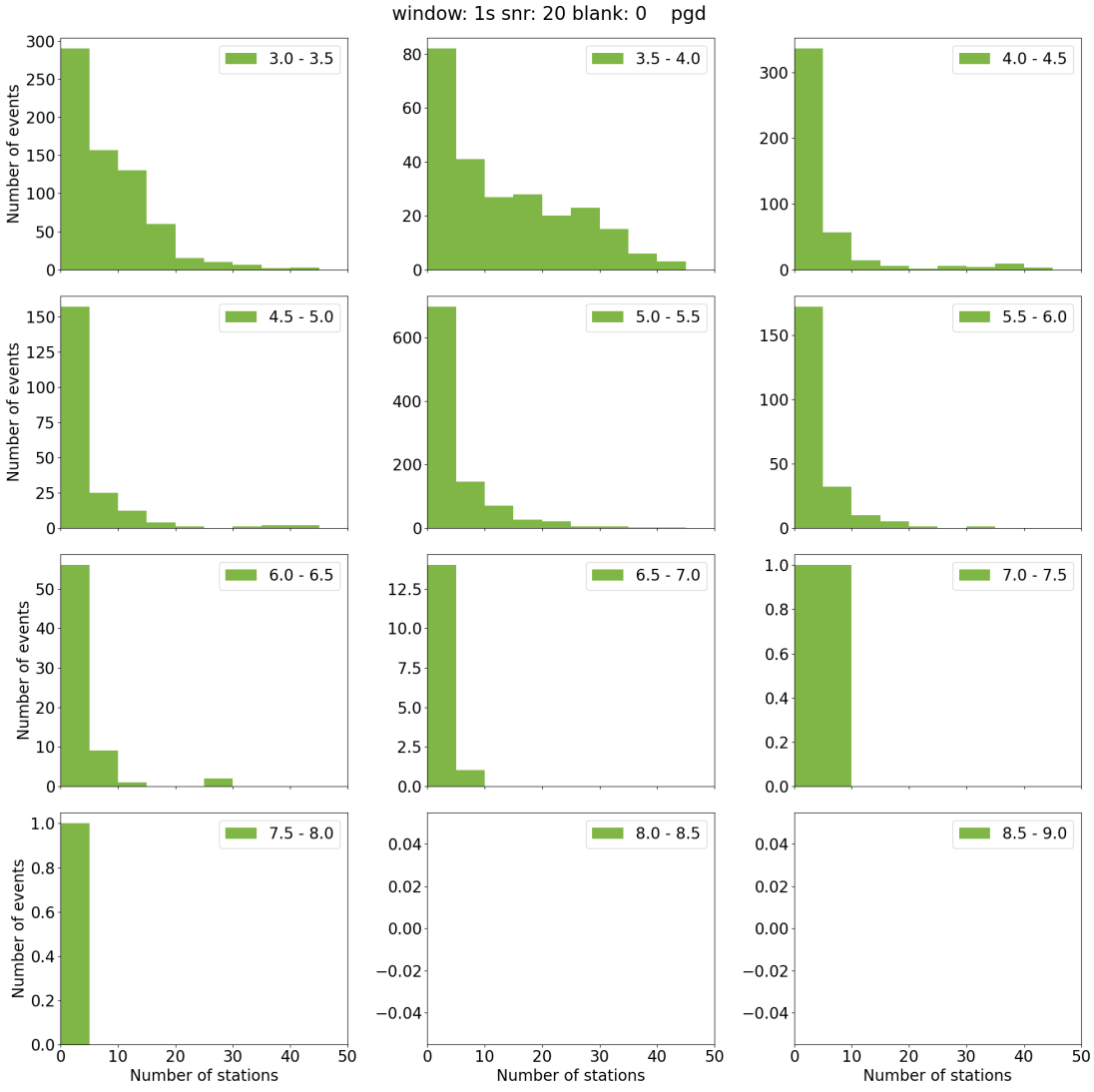


Figure A.18: As Figure 3.4 but for P_d calculated in a 1 s window.

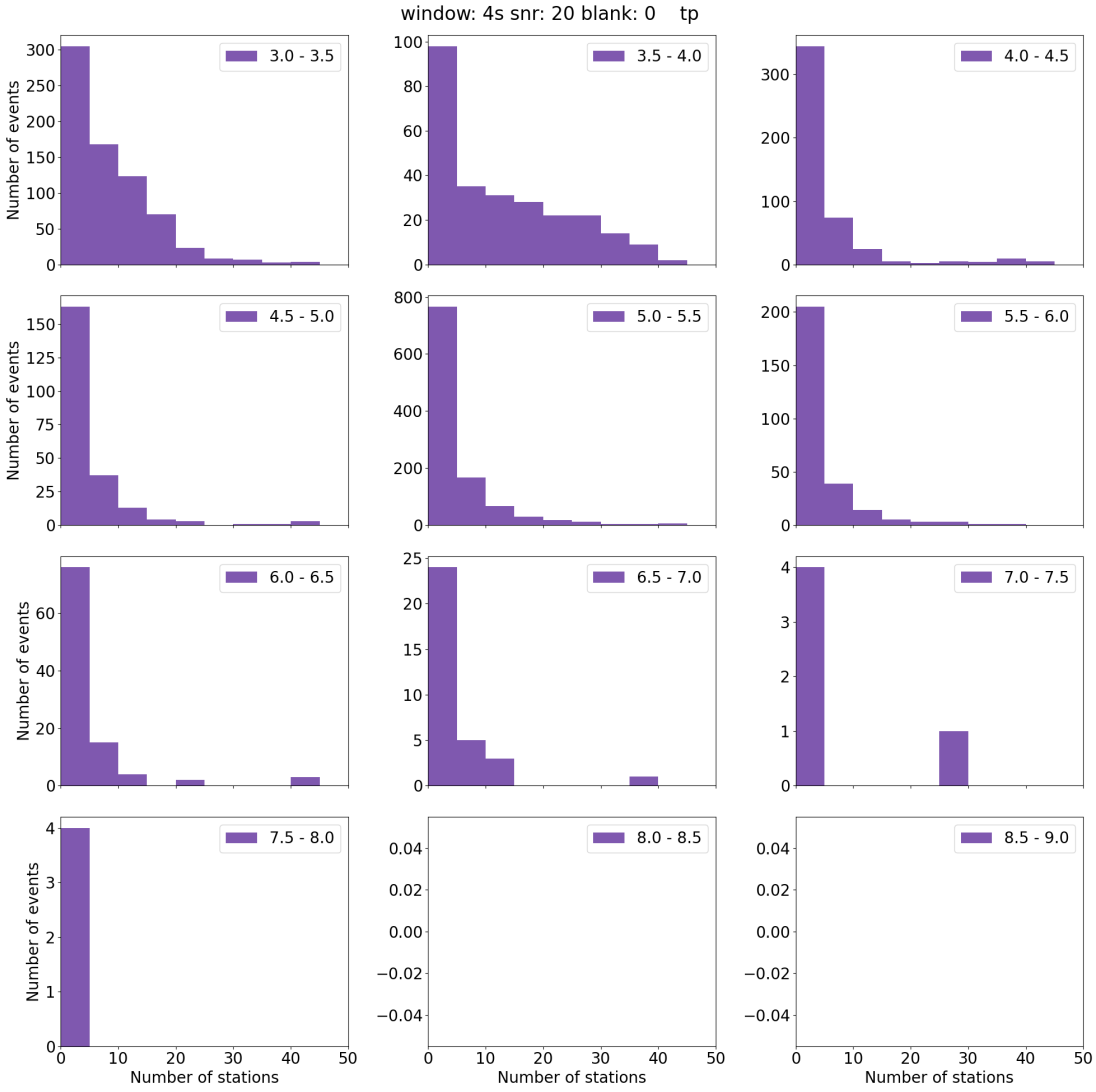


Figure A.19: As Figure 3.4 but for a 4 s window.

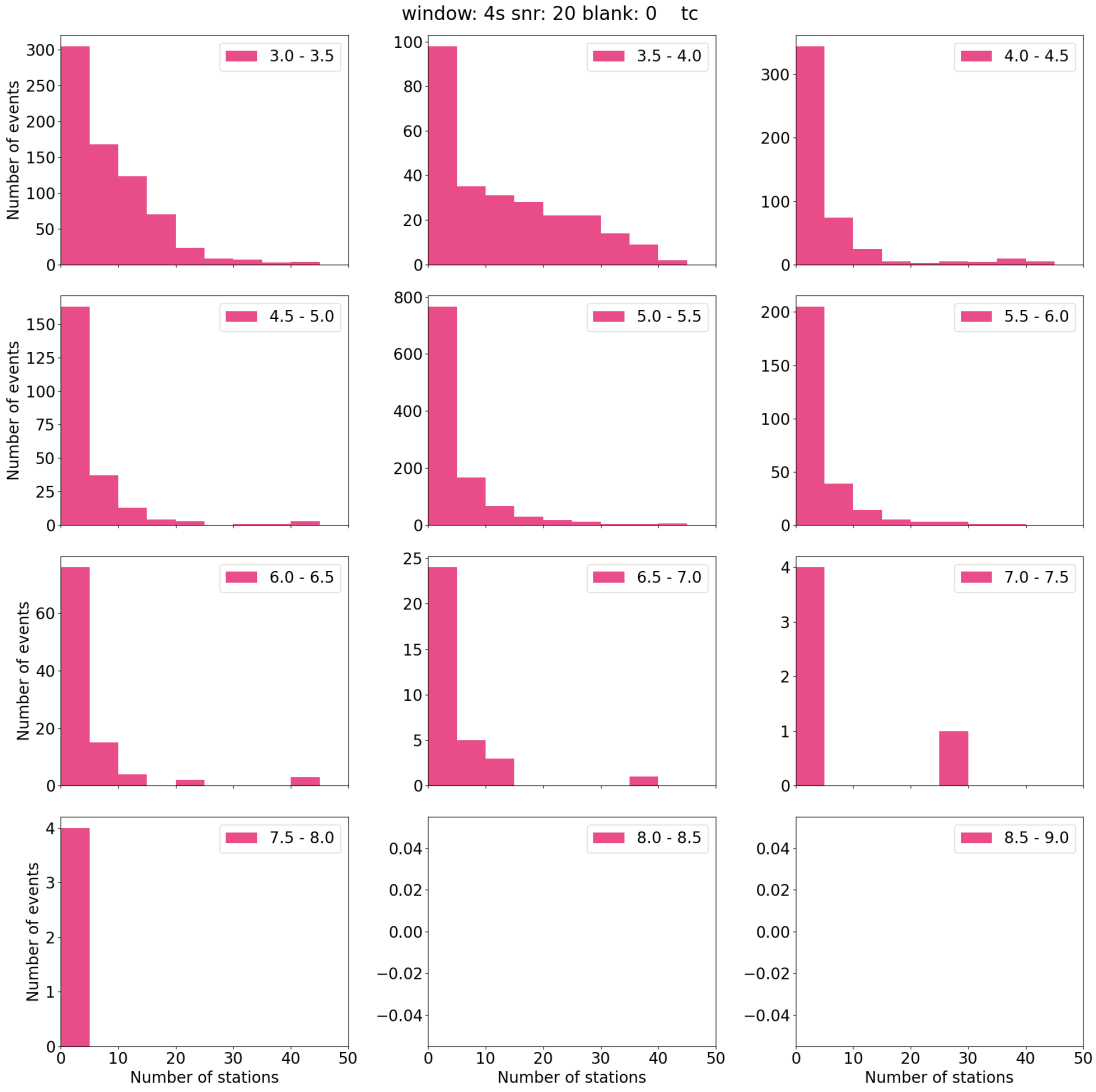


Figure A.20: As Figure 3.4 but for τ_c calculated in a 4 s window.

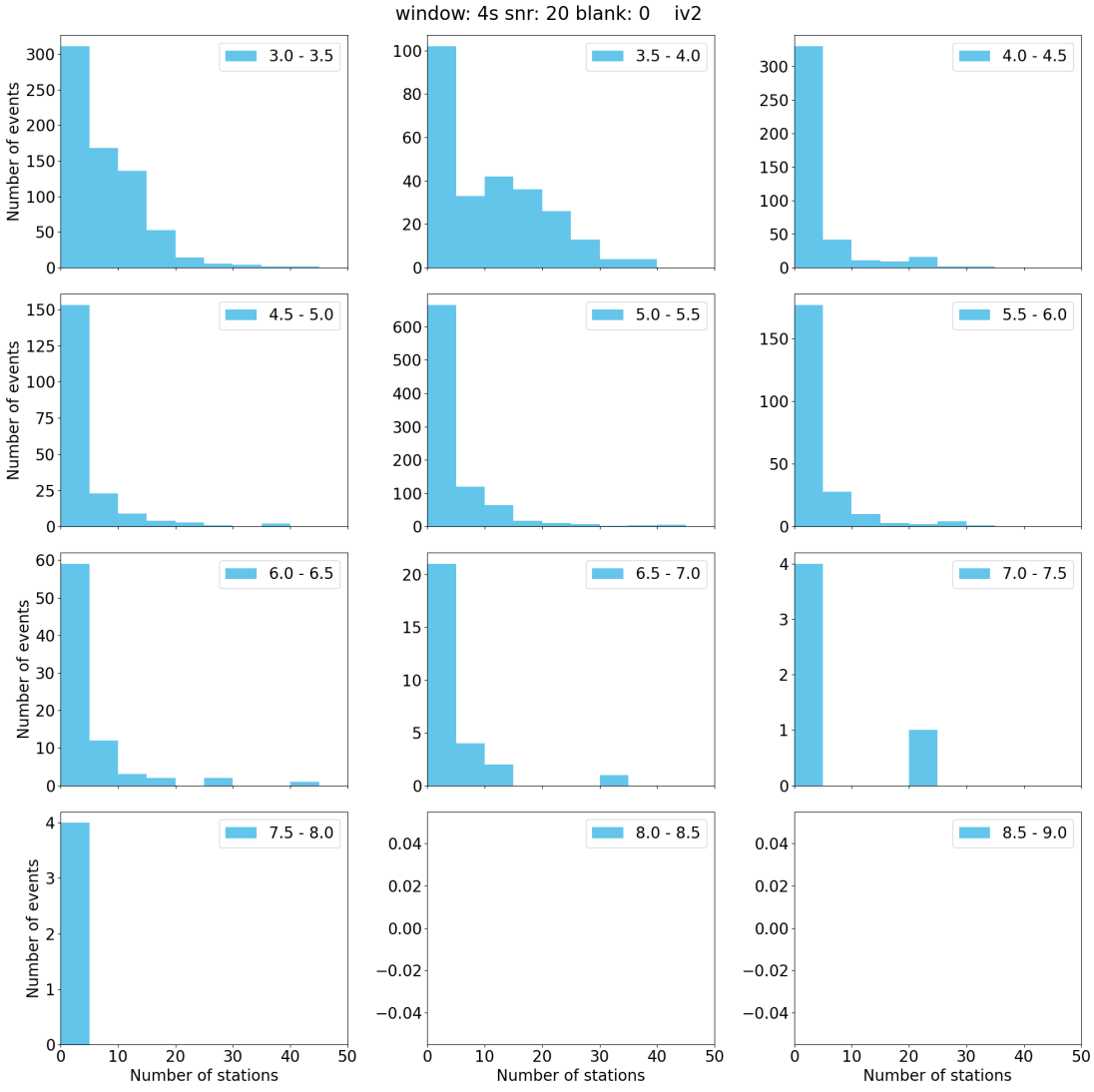


Figure A.21: As Figure 3.4 but for IV2 calculated in a 4 s window.

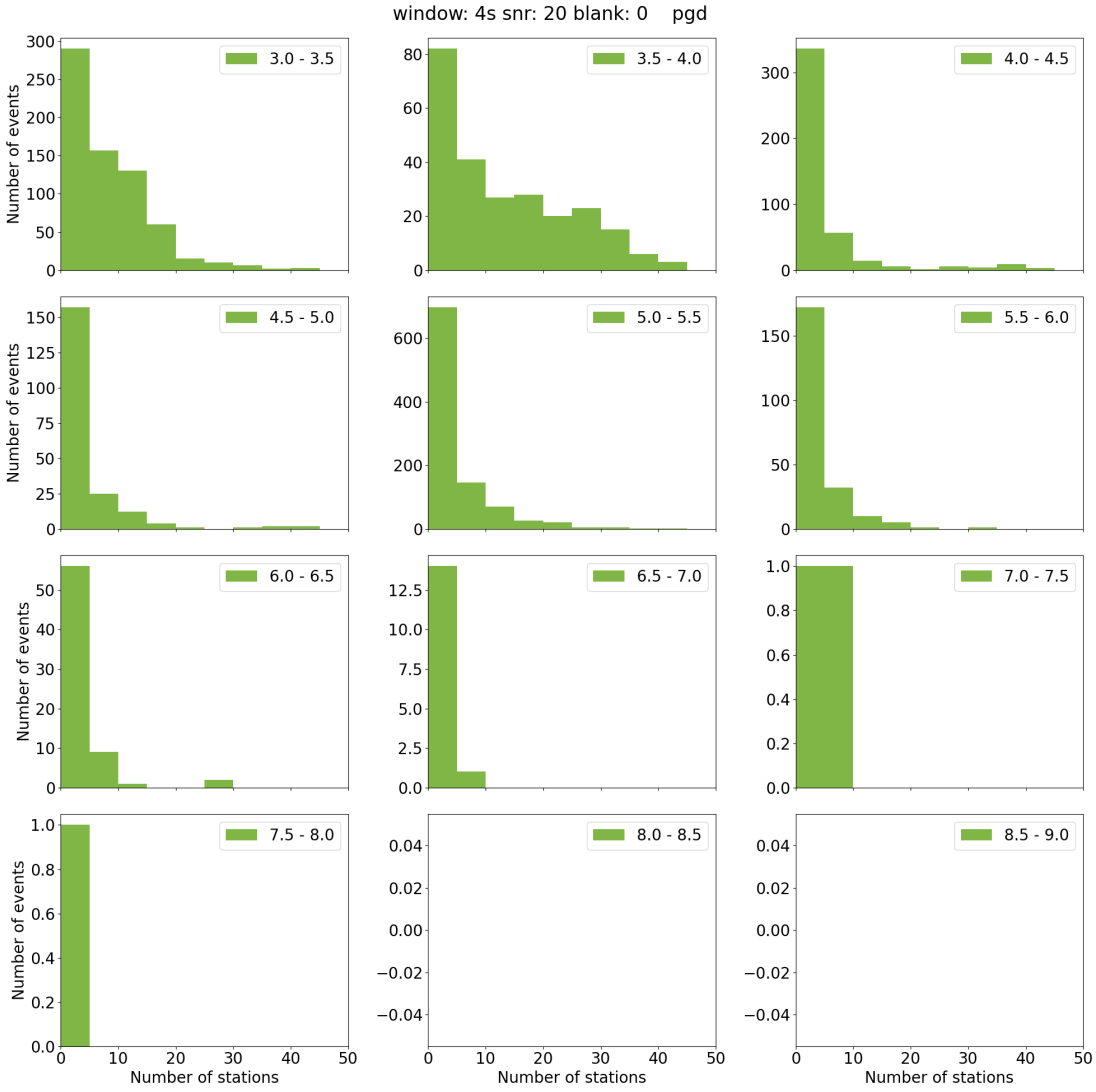


Figure A.22: As Figure 3.4 but for P_d calculated in a 4 s window.

A.6 Additional results

We calculate parameters for a range of time windows, and could not include all the results in the main text. Additional results tables and figures are presented here. See the main text for details.

Table A.4: Percentage of variance explained by a linear least squares fit. SNR = 20.

Window	Parameter	$M \geq 3$	$M \geq 4$	$M \geq 4.4$	$M \geq 5$	$M \geq 6.2$
0.3	$\tau_p^{\max}$	0.44	0.18	0.11	0.06	-0.07
	τ_c	0.53	0.26	0.19	0.13	0.01
	IV2	0.45	0.28	0.26	0.31	0.24
	P_d	0.56	0.27	0.17	0.18	0.26
0.5	$\tau_p^{\max}$	0.49	0.23	0.16	0.09	-0.09
	τ_c	0.67	0.43	0.31	0.19	0.02
	IV2	0.55	0.33	0.29	0.33	0.23
	P_d	0.65	0.38	0.26	0.23	0.20
1.0	$\tau_p^{\max}$	0.52	0.27	0.19	0.11	-0.08
	τ_c	0.73	0.59	0.48	0.35	-0.01
	IV2	0.66	0.41	0.34	0.35	0.15
	P_d	0.74	0.51	0.38	0.28	0.15
4.0	$\tau_p^{\max}$	0.53	0.28	0.21	0.13	-0.09
	τ_c	0.71	0.61	0.52	0.41	0.09
	IV2	0.78	0.58	0.49	0.45	0.19
	P_d	0.79	0.61	0.50	0.39	0.30

Table A.5: Percentage of variance explained by a linear least squares fit. SNR = 5.

Window	Parameter	$M \geq 3$	$M \geq 4$	$M \geq 4.4$	$M \geq 5$	M6+
0.3	$\tau_p^{\max}$	0.36	0.15	0.11	0.00	0.03
	τ_c	0.41	0.21	0.15	0.11	0.13
	IV2	0.52	0.29	0.26	0.37	0.27
	P_d	0.58	0.28	0.18	0.22	0.17
0.5	$\tau_p^{\max}$	0.41	0.21	0.15	0.02	-0.02
	τ_c	0.54	0.37	0.27	0.12	0.08
	IV2	0.60	0.34	0.29	0.39	0.28
	P_d	0.67	0.40	0.27	0.27	0.16
1.0	$\tau_p^{\max}$	0.43	0.24	0.18	0.04	0.00
	τ_c	0.62	0.53	0.44	0.23	0.06
	IV2	0.69	0.42	0.34	0.40	0.24
	P_d	0.75	0.52	0.39	0.31	0.16
4.0	$\tau_p^{\max}$	0.44	0.25	0.19	0.05	0.01
	τ_c	0.58	0.55	0.48	0.31	0.20
	IV2	0.80	0.58	0.48	0.48	0.27
	P_d	0.80	0.62	0.51	0.40	0.33

Table A.6: The Spearman's rank correlation coefficients for each of the parameters, calculated for four Δt_i values: 0.3, 0.5, 1 and 4 seconds.

Δt_i	0.3 s	0.5 s	1 s	4 s
τ_p	0.48	0.52	0.55	0.56
τ_c	0.55	0.68	0.76	0.75
IV2	0.45	0.55	0.66	0.79
P_d	0.60	0.70	0.78	0.83

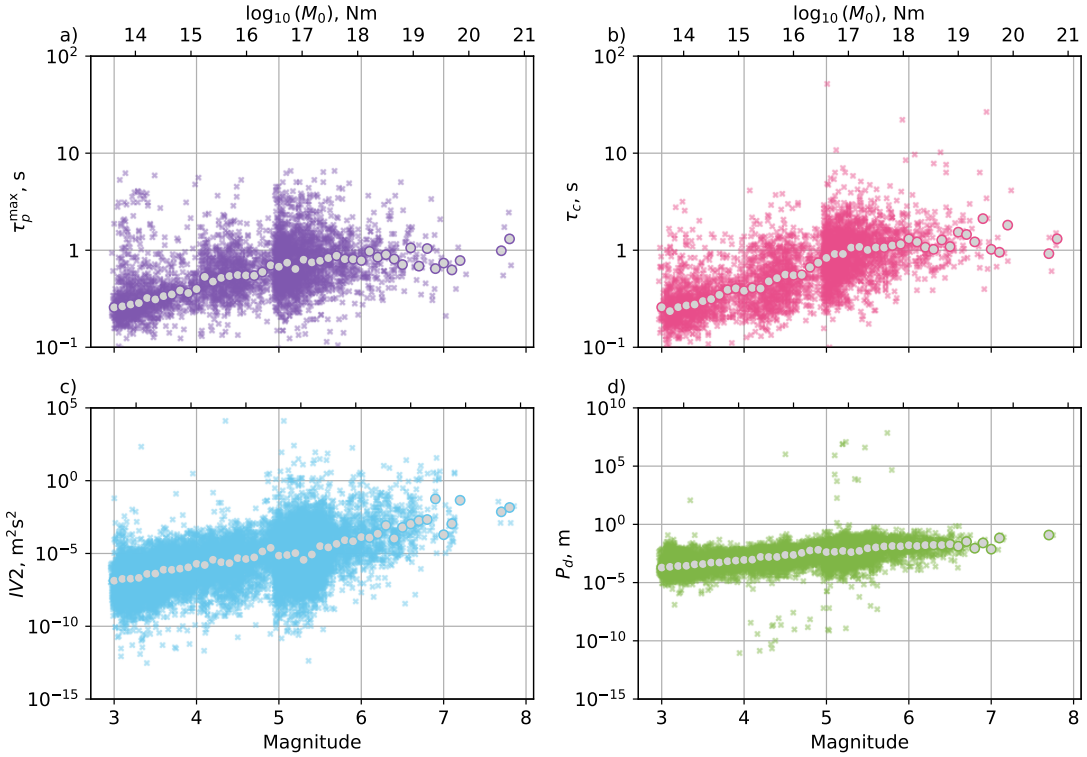


Figure A.23: As in Figure 3.5, but for a 0.5 s window.

A.7 Is the increase in scatter at M5 real?

In Figures 3.5, 3.6 and A.23–A.28, we see a large increase in scatter at M5. This is due to the increase in data density at M5.

First we plot the same figure as Figure 3.5 but for only earthquakes in 2019 (Figure A.29). Here we see that the large increase in scatter is no longer apparent.

We can also compare the distributions of the data between M4.9–5.0 and M5.0–5.1. We find the standard deviation of the data is 0.23 and 0.26 respectively. Plotting the distribution of the data in 2019, we see that the two subsets are similarly distributed (Figure A.30).

Plotting the distributions of only M5 plus earthquakes by year, we see that there is natural variation in the distributions, but that the median values remain similar through time (Figure A.31).

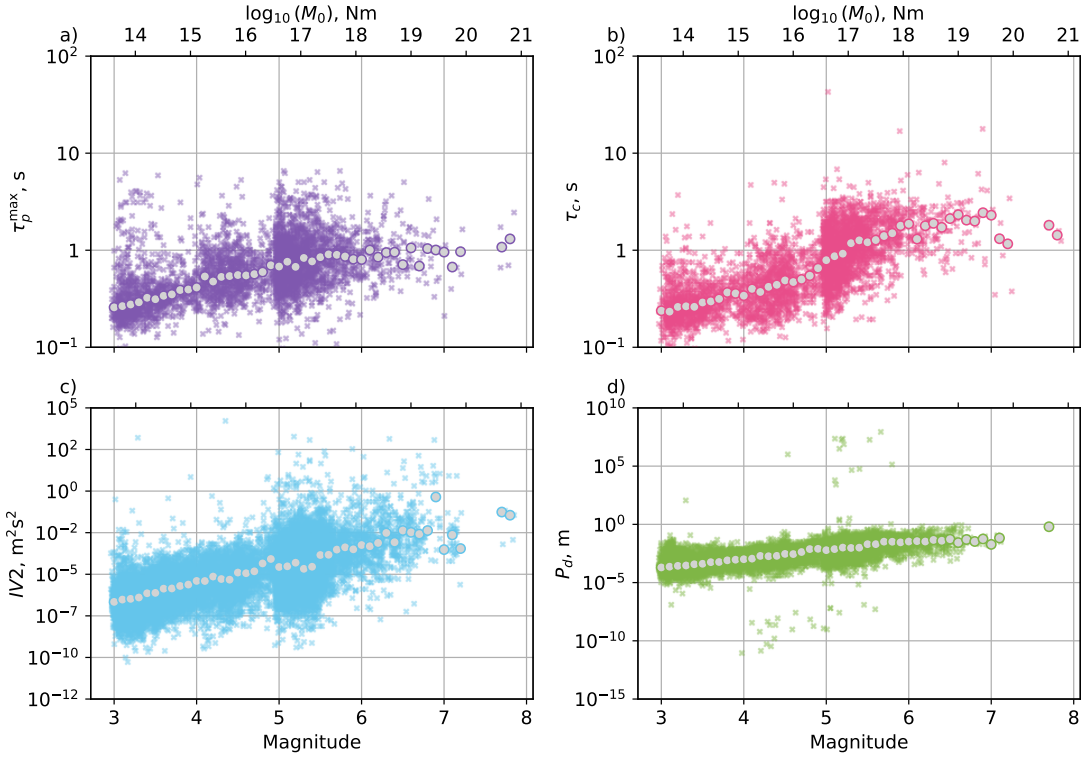


Figure A.24: As in Figure 3.5, but for a 1 s window.

A.8 Correlations in different time windows

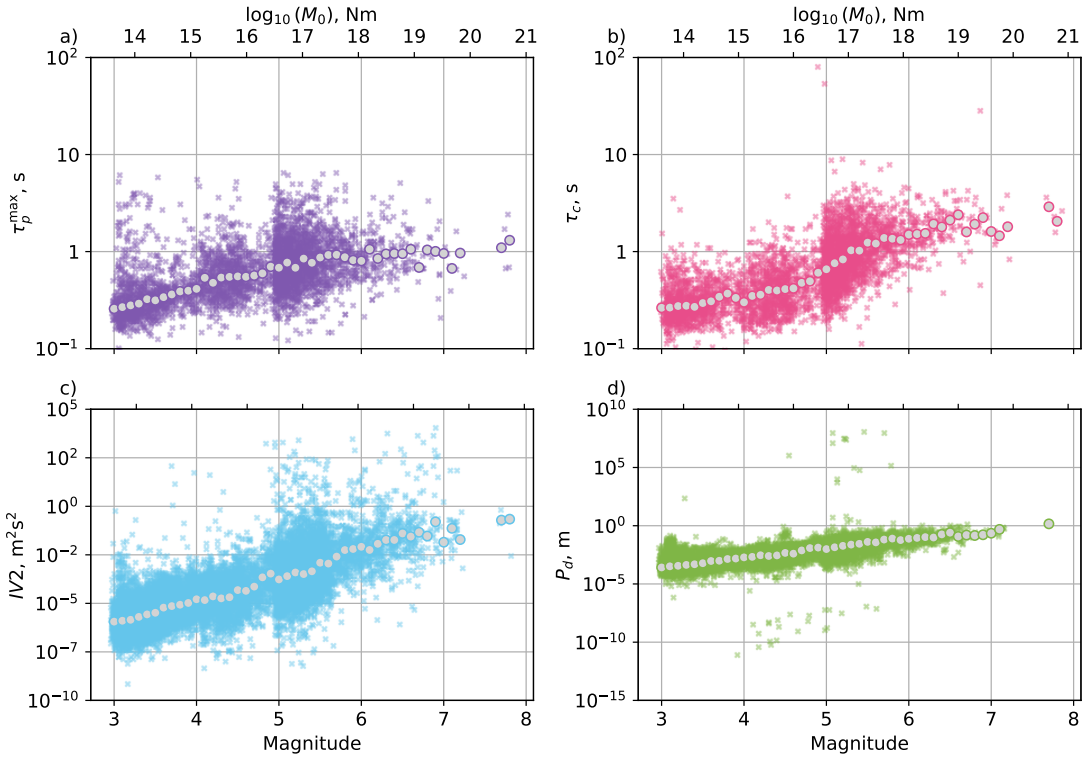


Figure A.25: As in Figure 3.5, but for a 4 s window.

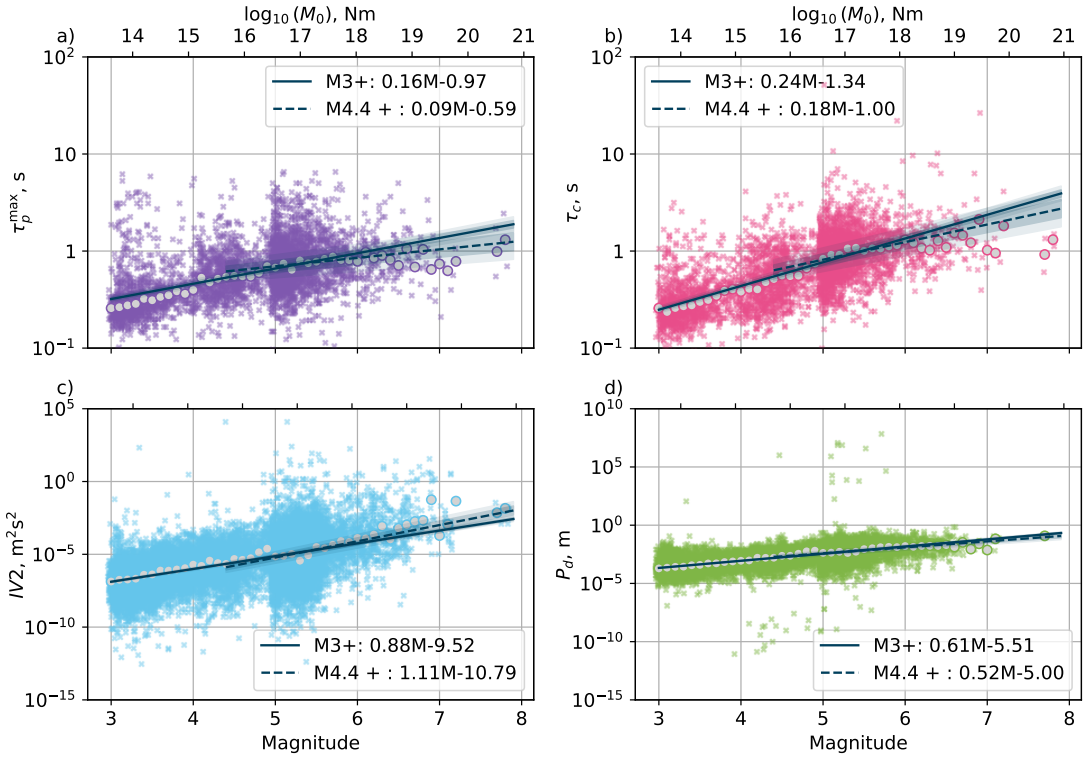


Figure A.26: As Figure 3.6, but for a 0.5 s window.

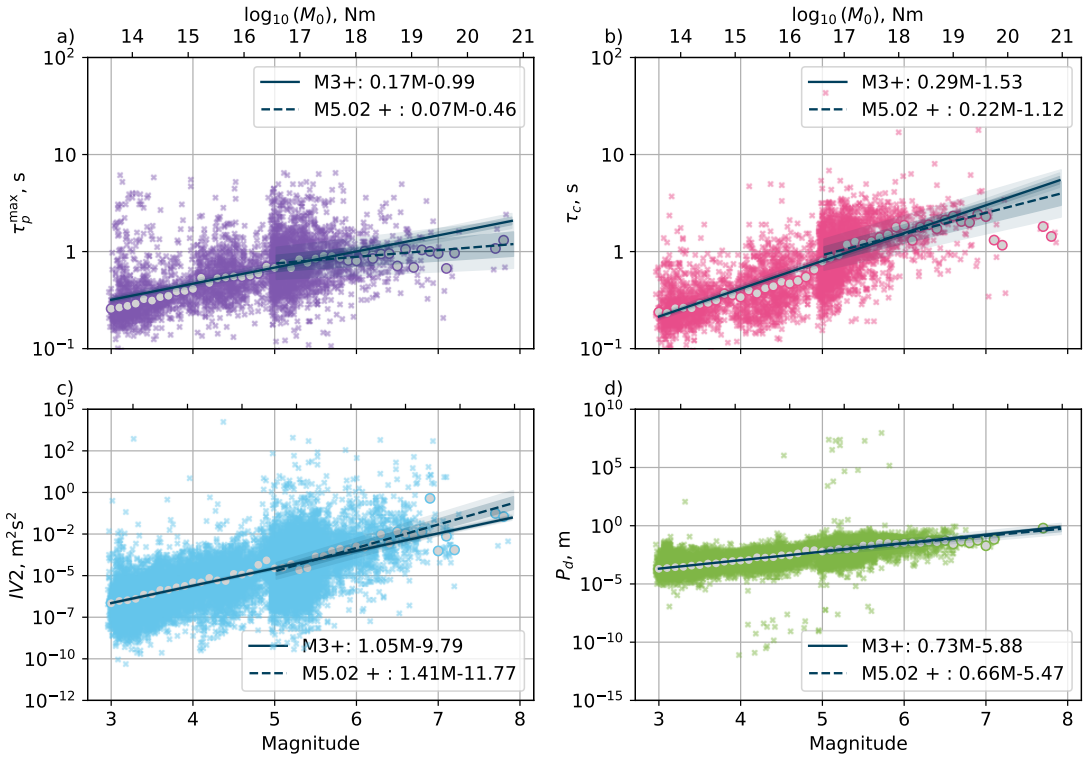


Figure A.27: As Figure 3.6, but for a 1 s window.

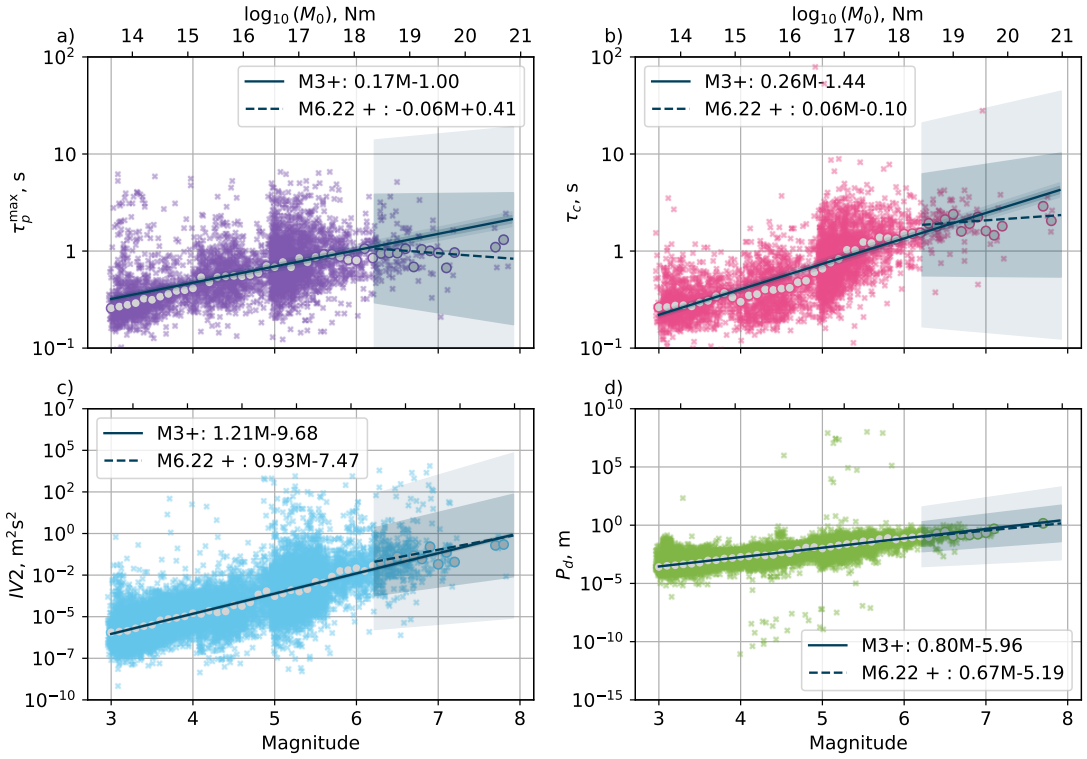


Figure A.28: As Figure 3.6, but for a 4 s window.

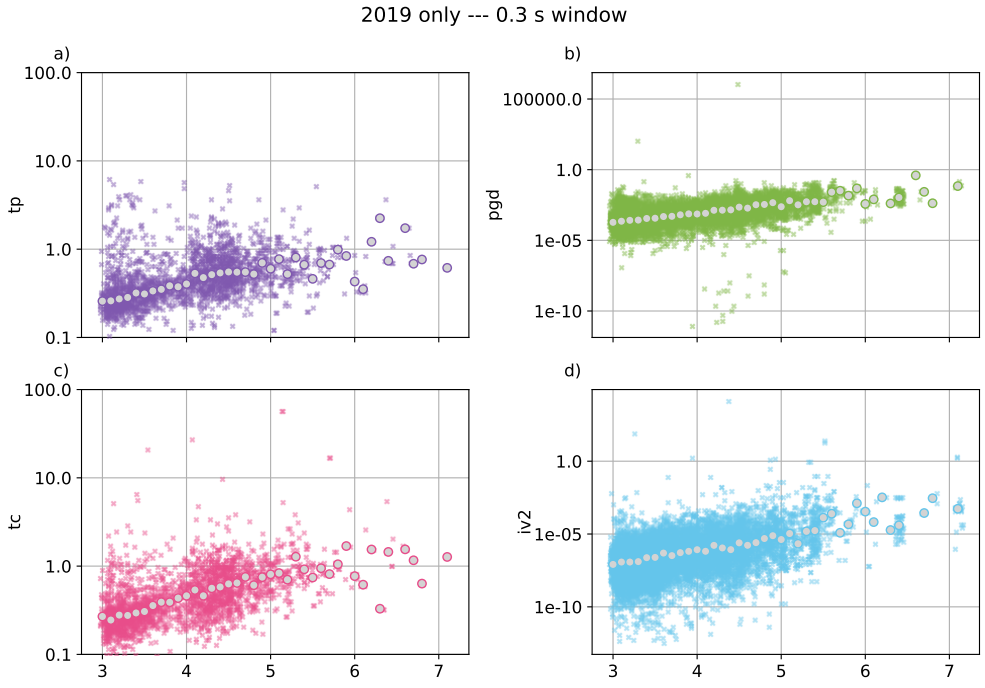


Figure A.29: As Figure 3.6 but with only data from 2019.

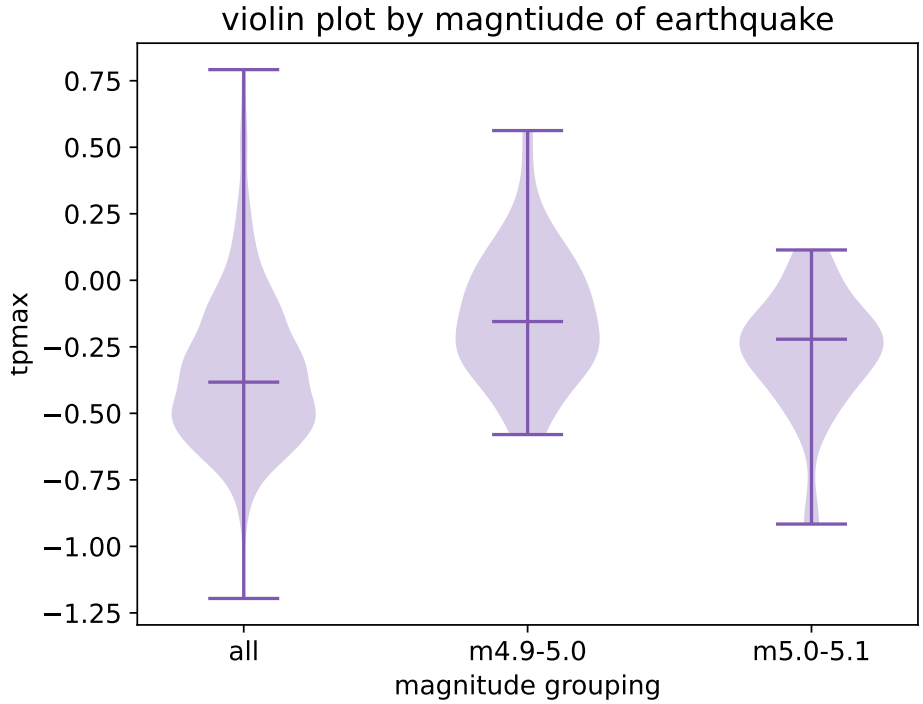


Figure A.30: Violin plot of the distribution of predominant period values in 3 different magnitude groups.

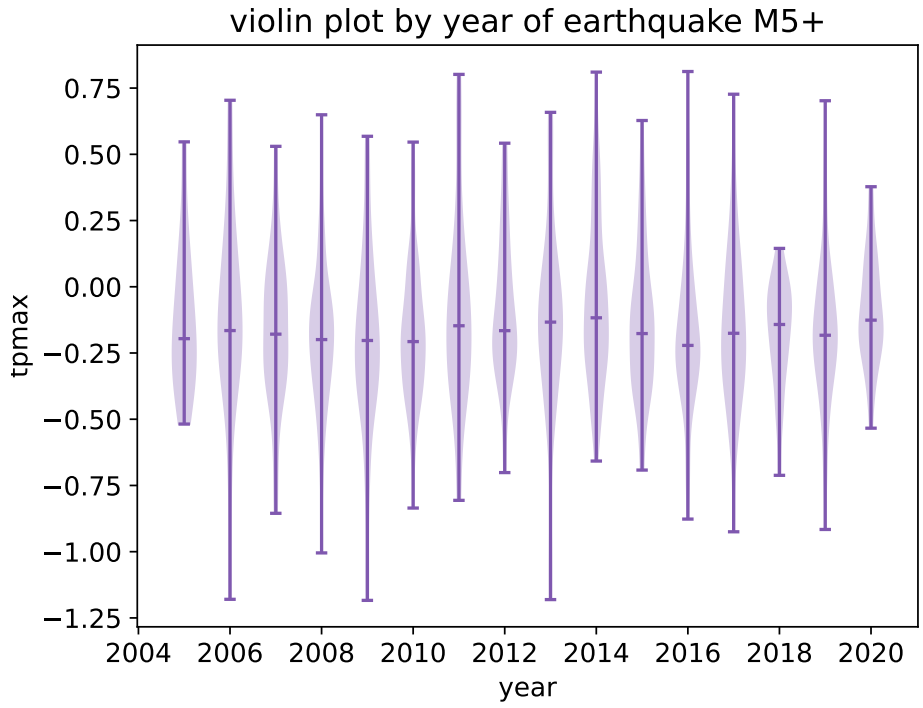


Figure A.31: Violin plot of the distribution of predominant period values for data from $M \geq 5$ earthquakes each year

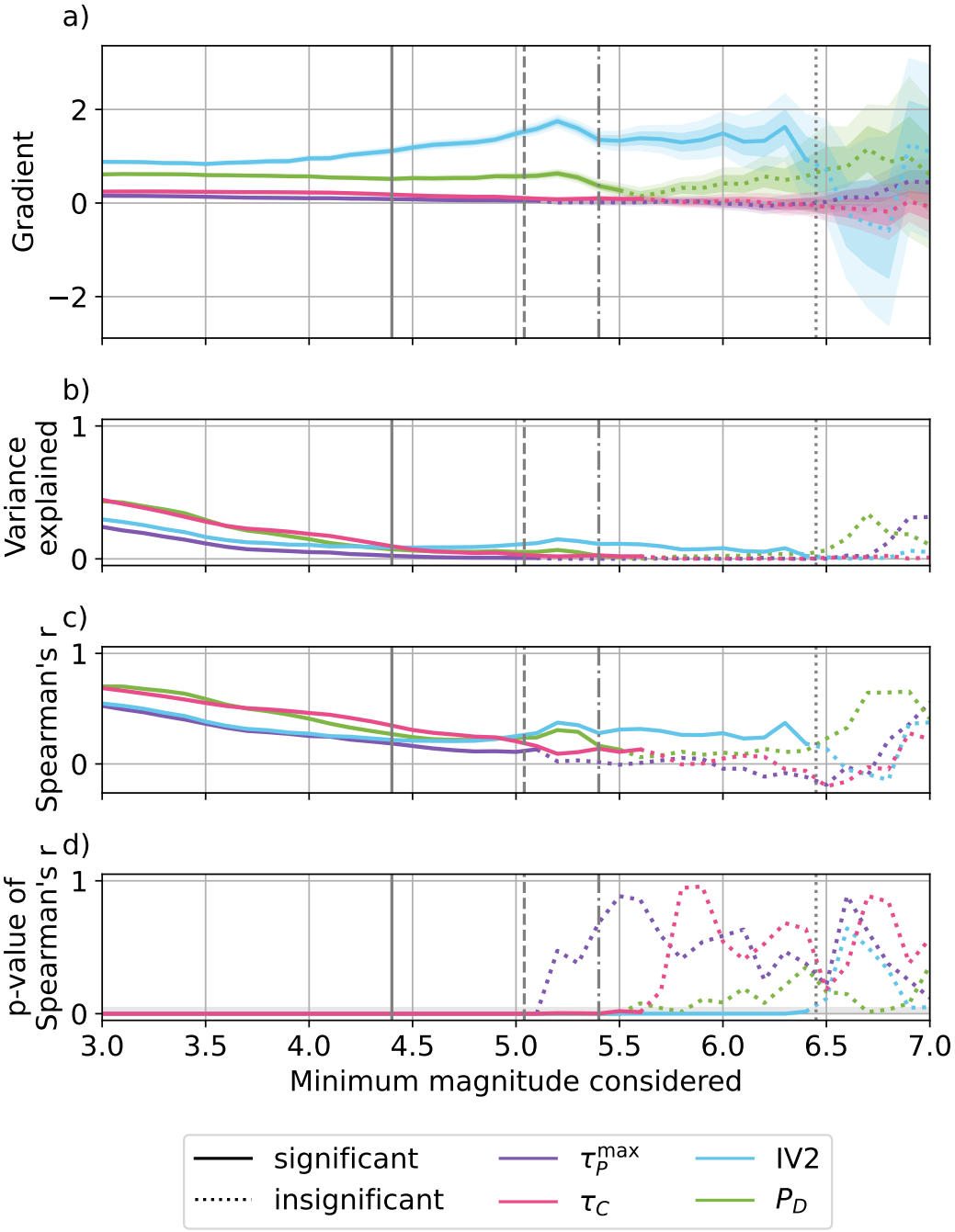


Figure A.32: As Figure 3.7, but for a 0.5 s window.

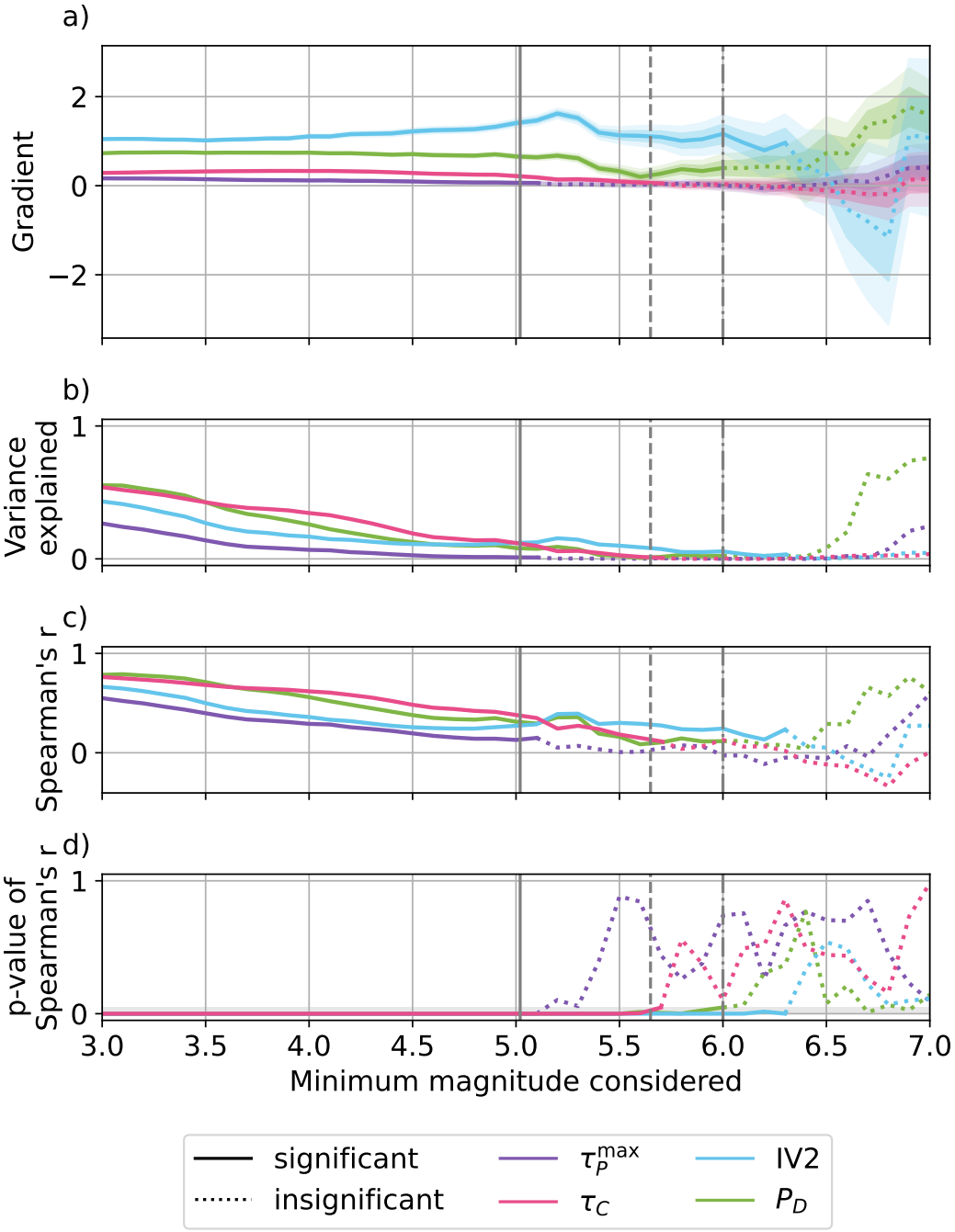


Figure A.33: As Figure 3.7, but for a 1 s window.

A.9 Impact of source-receiver distance and number of stations

In Figure A.34, we show the magnitude at which there is no longer a significant relationship between the parameter and magnitude, as we vary the minimum source-receiver distance and the minimum number of stations used in the calculation. Colours show the minimum number of stations the event was recorded at.

For predominant period the magnitude where significance is lost is reduced for larger numbers of stations. For the other parameters, it has little impact other than when only considering earthquakes recorded at more than five or six stations. This is a result of the reduced numbers of earthquakes that are in this subset and thus the increased uncertainty on the lines of best fit. Along the x-axis, we show the minimum source-receiver distance of the data used to fit the line of best fit. Predominant and average period show a trend of losing significance earlier when the minimum distance is larger. Meanwhile, peak ground displacement and IV2 are relatively insensitive to changing the minimum distance.

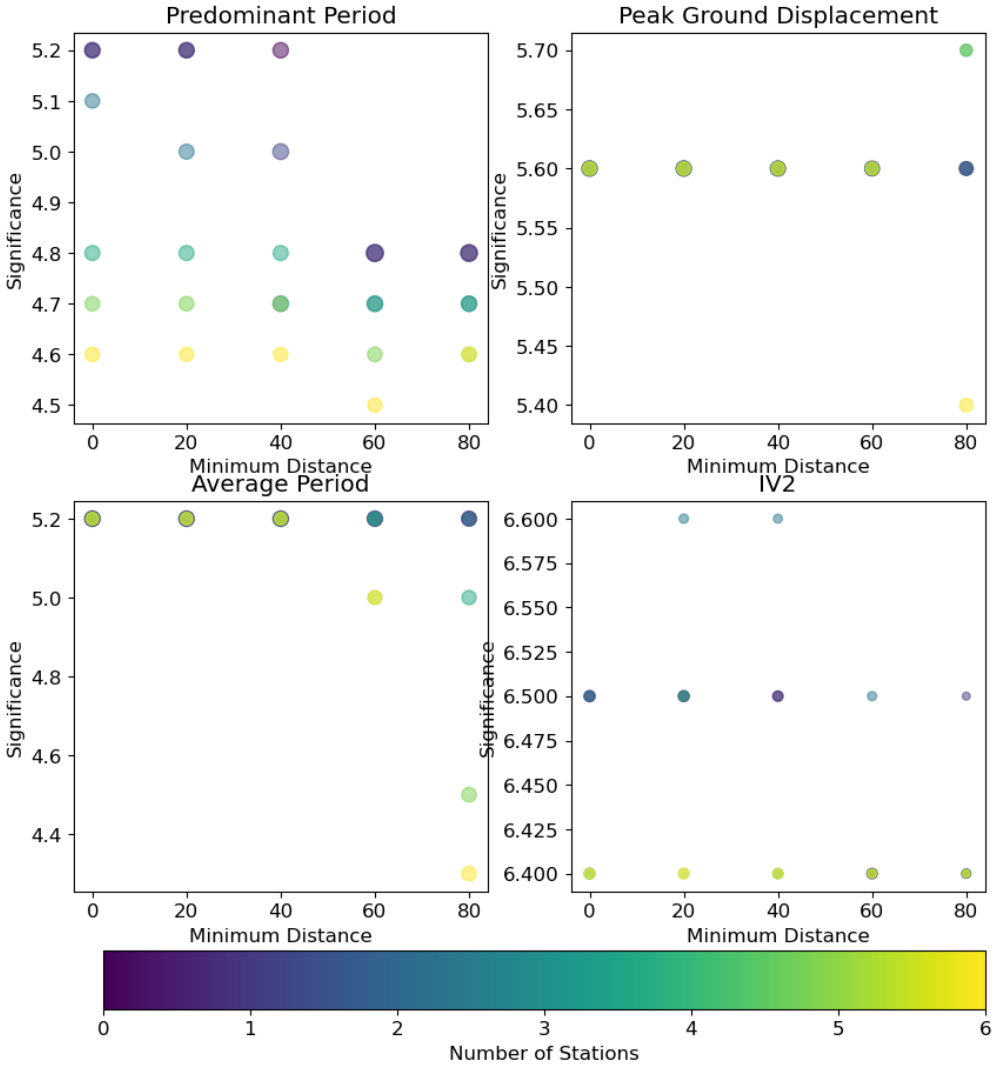


Figure A.34: A plot of the minimum magnitude at which the data set no longer has a statistically significant relationship between the parameter and magnitude against the minimum source-receiver distance considered and the minimum number of stations used in the calculation.

A.10 Impact of shifting the pick times

We shift the pick by 0.1 seconds in each direction and look at the impact on the magnitude at which there is no longer a significant relationship, and the gradient of the relationship between the parameter and magnitude.

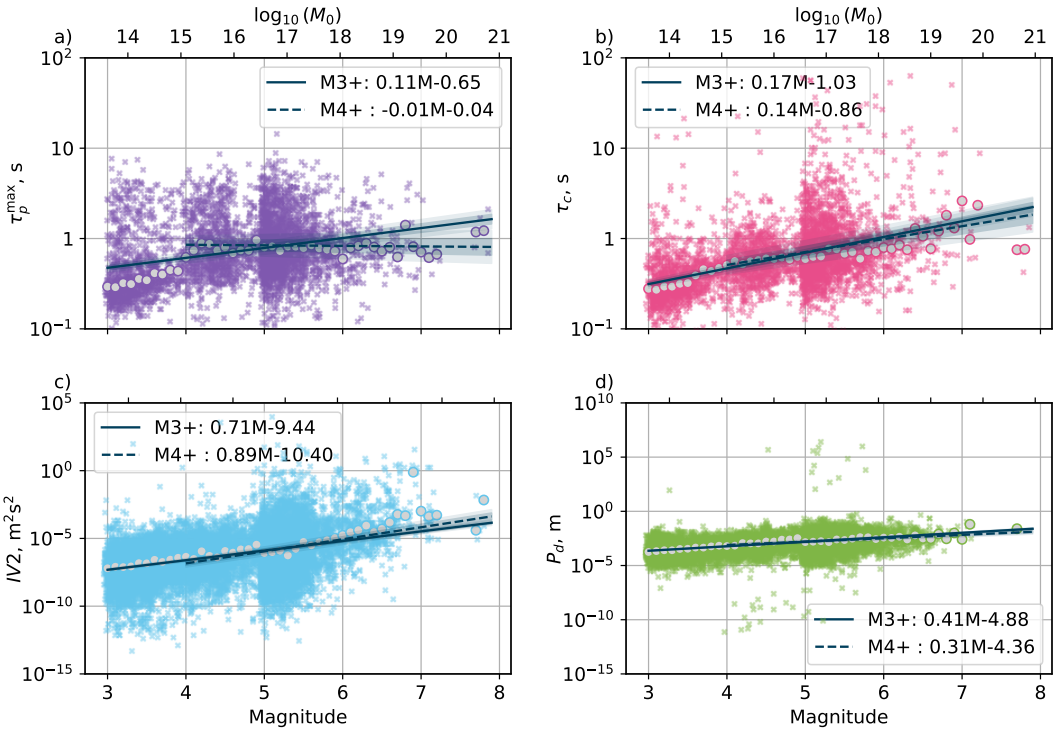


Figure A.35: As Figure 3.6, but for a 0.3 s window with the pick shifted by −0.1 s.

Table A.7: As Table 3.1 but for 0.5 s window (Average duration of M4.4 is 0.5 s).

Pick shift	$\tau_p^{\max}$	τ_c	P_d	IV2
-0.1	4.3	5.1	5.5	6.4
0	5.1	5.6	5.5	6.4
+0.1	5.1	5.6	5.5	6.4

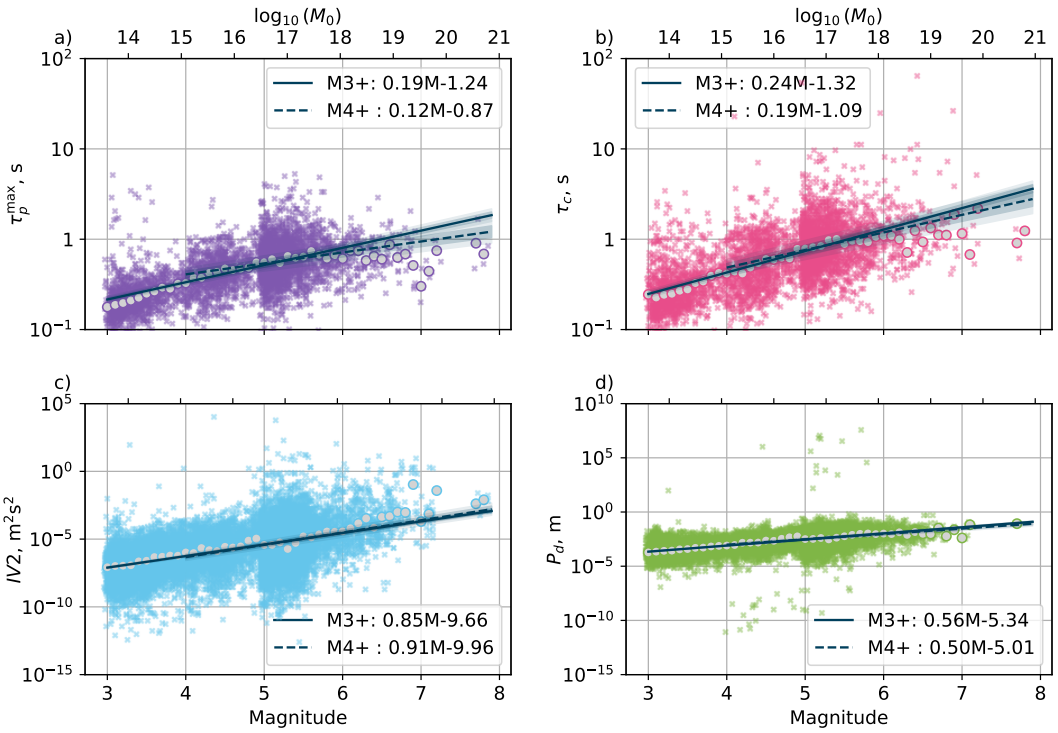


Figure A.36: As Figure 3.6, but for a 0.3 s window with the pick shifted by +0.1 s.

Table A.8: As Table 3.2 but for 0.5 s window. (M=4.4).

Pick shift	$\tau_p^{\max}$	τ_c	P_d	IV2
-0.1	0.01	0.15	0.45	1.10
0	0.09	0.18	0.52	1.11
+0.1	0.11	0.22	0.57	1.12

Table A.9: As Table 3.1 but for 1 s window (Average duration of M5.02 is 1 s).

Pick shift	$\tau_p^{\max}$	τ_c	P_d	IV2
-0.1	4.6	5.6	6.0	6.3
0	5.1	5.7	6.0	6.3
+0.1	5.3	5.6	6.0	6.3

Table A.10: As Table 3.2 but for 1 s window (M=5.02).

Pick shift	$\tau_p^{\max}$	τ_c	P_d	IV2
-0.1	0.02	0.20	0.64	1.41
0	0.07	0.22	0.66	1.41
+0.1	0.08	0.22	0.67	1.41

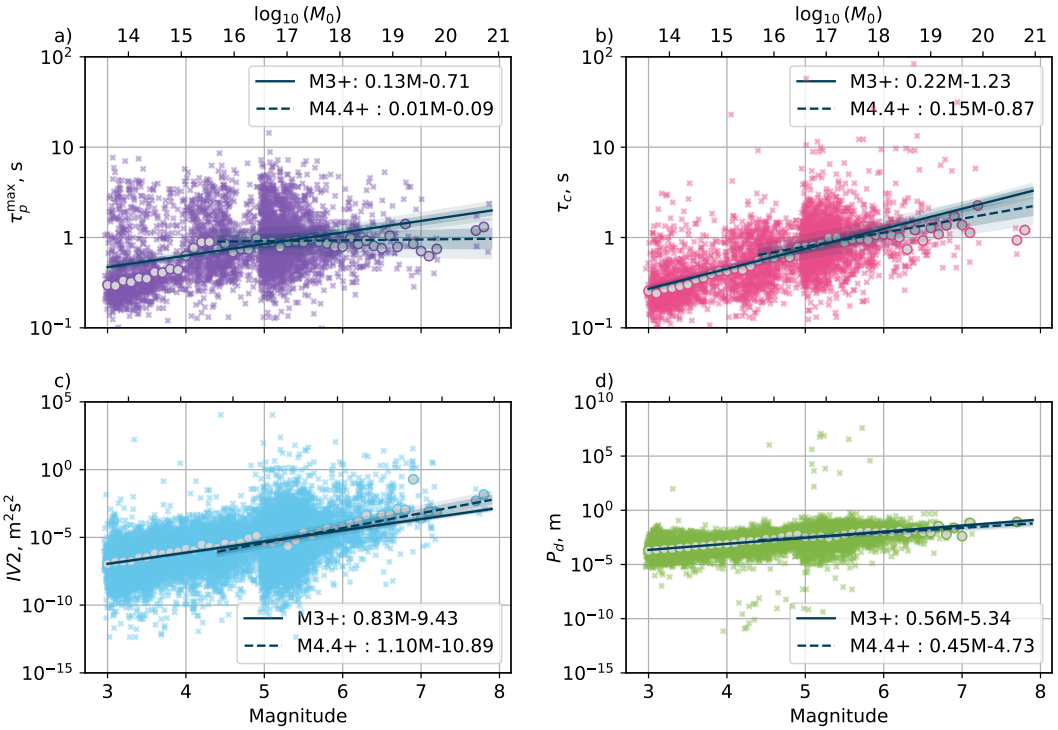


Figure A.37: As Figure 3.6, but for a 0.5 s window with the pick shifted by −0.1 s.

Table A.11: As Table 3.1 but for 4 s window (Average duration of M6.22 is 4 s).

Pick shift	$\tau_p^{\max}$	τ_c	P_d	IV2
-0.1	5.1	6.1	6.3	6.4
0	5.3	6.1	6.3	6.5
+0.1	5.3	6.1	6.3	6.4

Table A.12: As Table 3.2 but for 4 s window (M=6.22).

Pick shift	$\tau_p^{\max}$	τ_c	P_d	IV2
-0.1	-0.03	0.07	0.64	0.94
0	-0.06	0.06	0.67	0.93
+0.1	-0.04	0.06	0.70	0.94

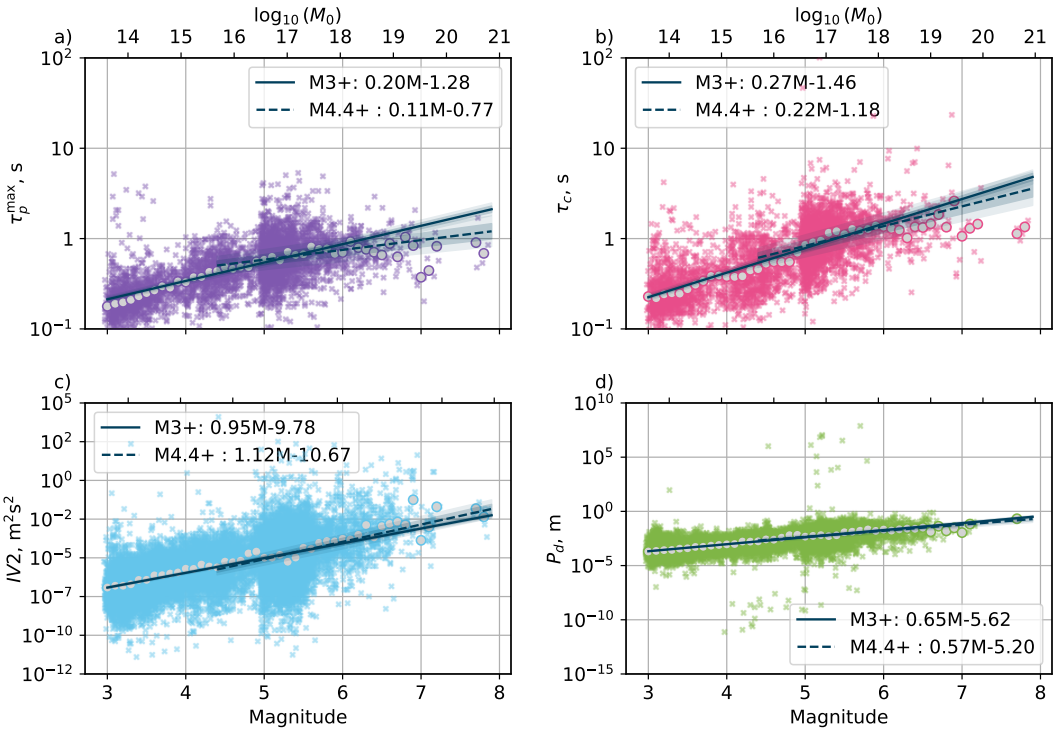


Figure A.38: As Figure 3.6, but for a 0.5 s window with the pick shifted by +0.1 s.

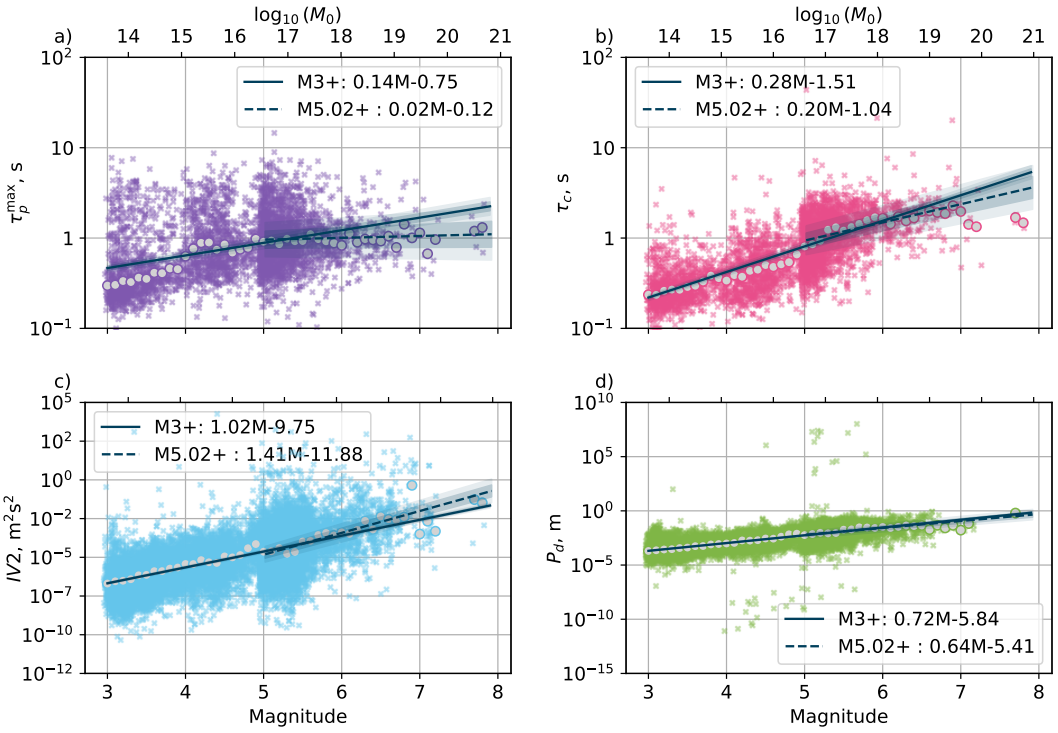


Figure A.39: As Figure 3.6, but for a 1 s window with the pick shifted by −0.1 s.

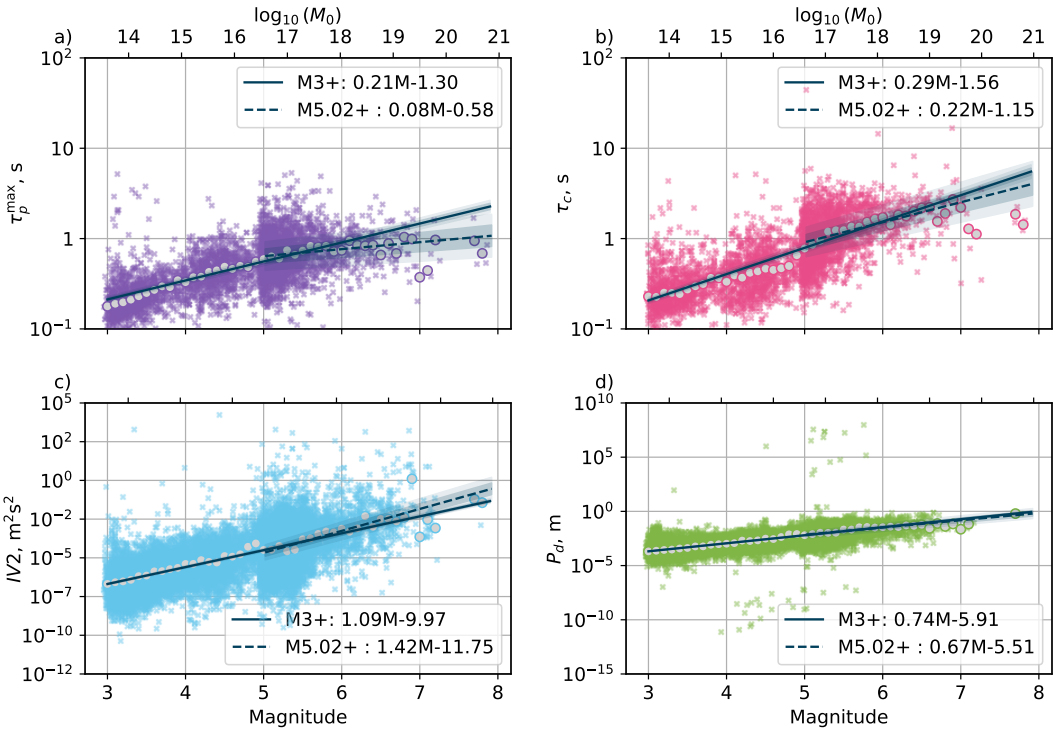


Figure A.40: As Figure 3.6, but for a 1 s window with the pick shifted by +0.1 s.

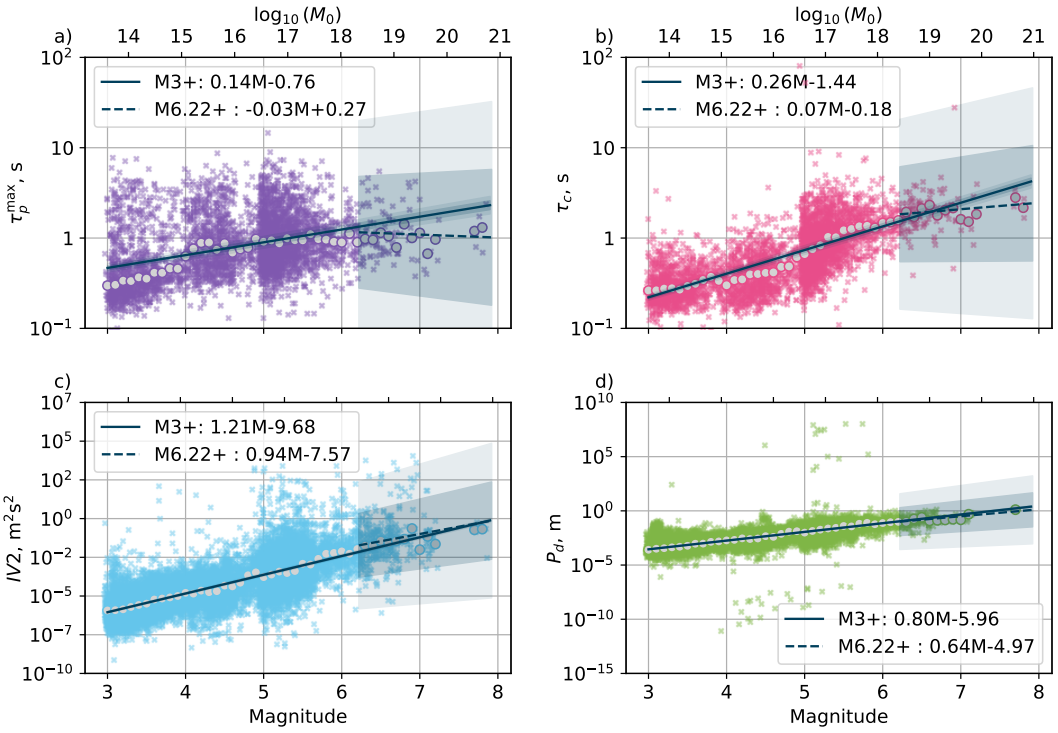


Figure A.41: As Figure 3.6, but for a 4 s window with the pick shifted by −0.1 s.

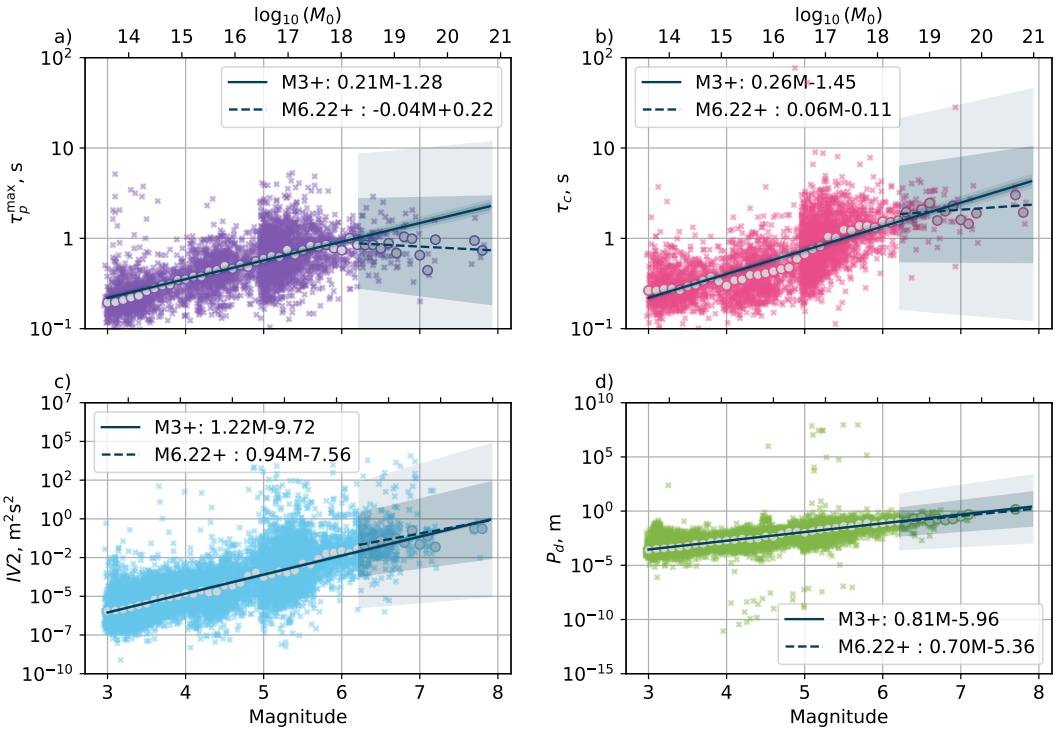


Figure A.42: As Figure 3.6, but for a 4 s window with the pick shifted by +0.1 s.

A.11 Keeping the relative duration of the window constant

We can compare the gradients calculated in different time windows of calculation based on earthquakes where the window measures a constant proportion of the total duration (Figures A.43–A.48).

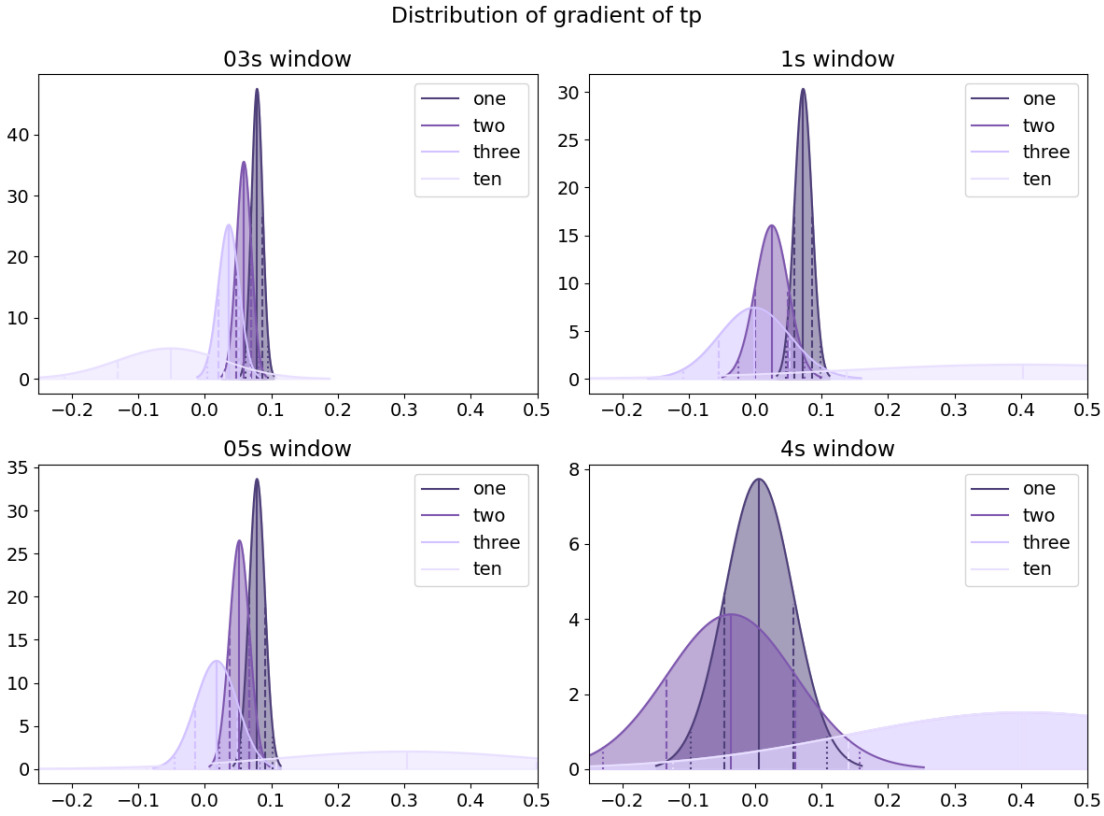


Figure A.43: Distribution of the gradient of the line of best fit found between predominant period and magnitude. Each panel is a different window length for the calculation (clockwise from top right: 0.3 s, 0.5 s, 1 s, 4 s) and each line shows the gradient calculated only considering earthquakes with average duration of one (darkest), two, three or ten (lightest) times the window duration.

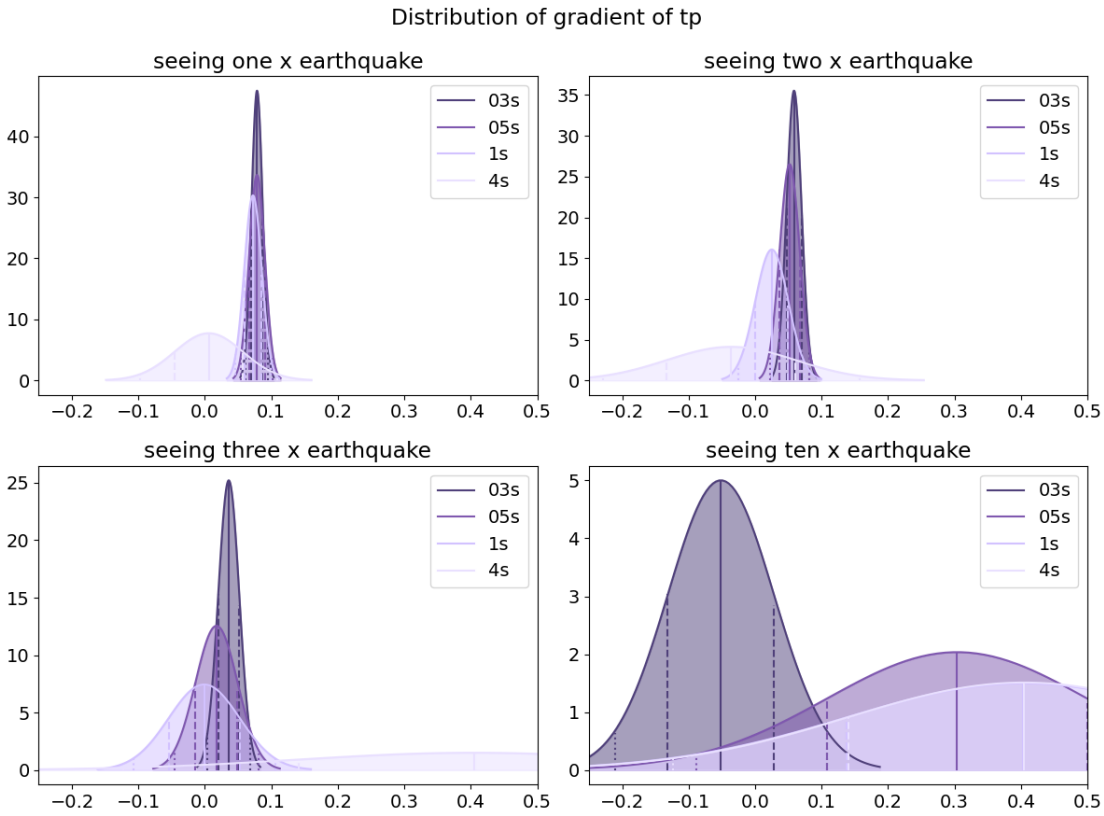


Figure A.44: Distribution of the gradient of the line of best fit found between predominant period and magnitude. Each panel shows the gradient calculated only considering earthquakes with average duration longer than of one, two, three or ten times the window duration. Each line is a different window length for the calculation from 0.3 s (darkest), 0.5 s, 1 s to 4 s (lightest).

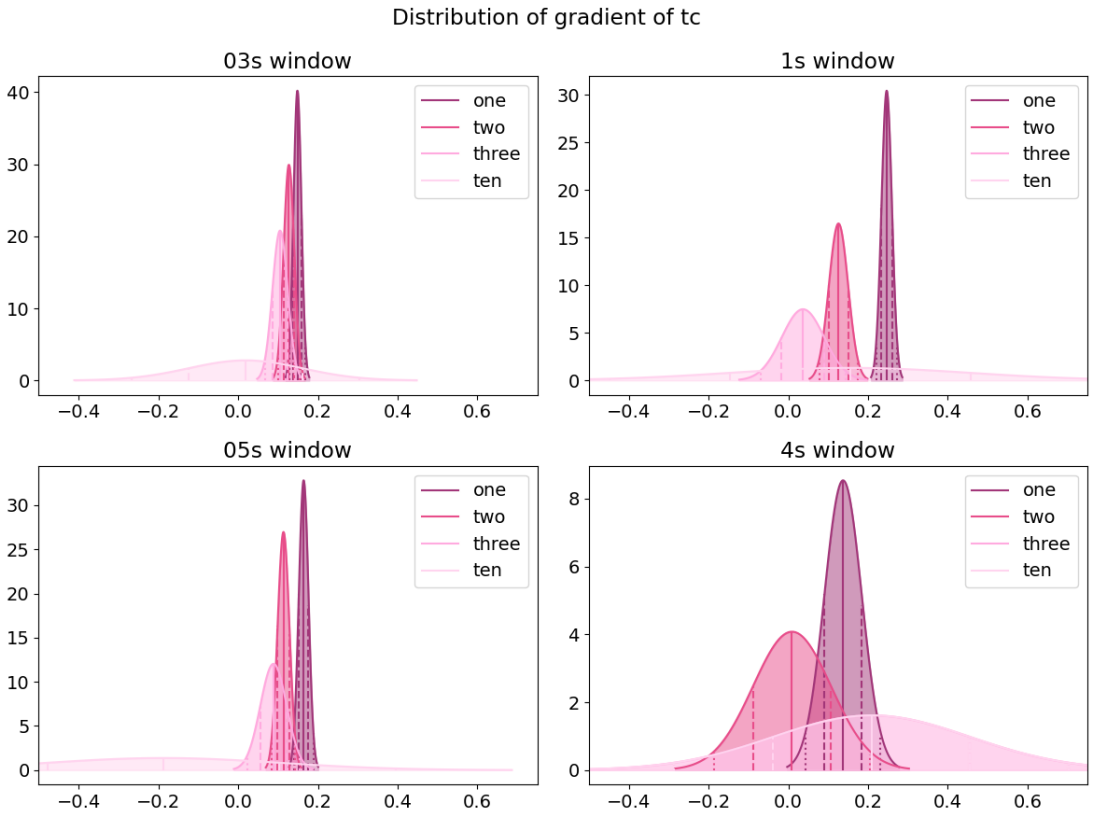


Figure A.45: Distribution of the gradient of the line of best fit found between average period and magnitude. Each panel is a different window length for the calculation (clockwise from top right: 0.3 s, 0.5 s, 1 s, 4 s) and each line shows the gradient calculated only considering earthquakes with average duration of one (darkest), two, three or ten (lightest) times the window duration.

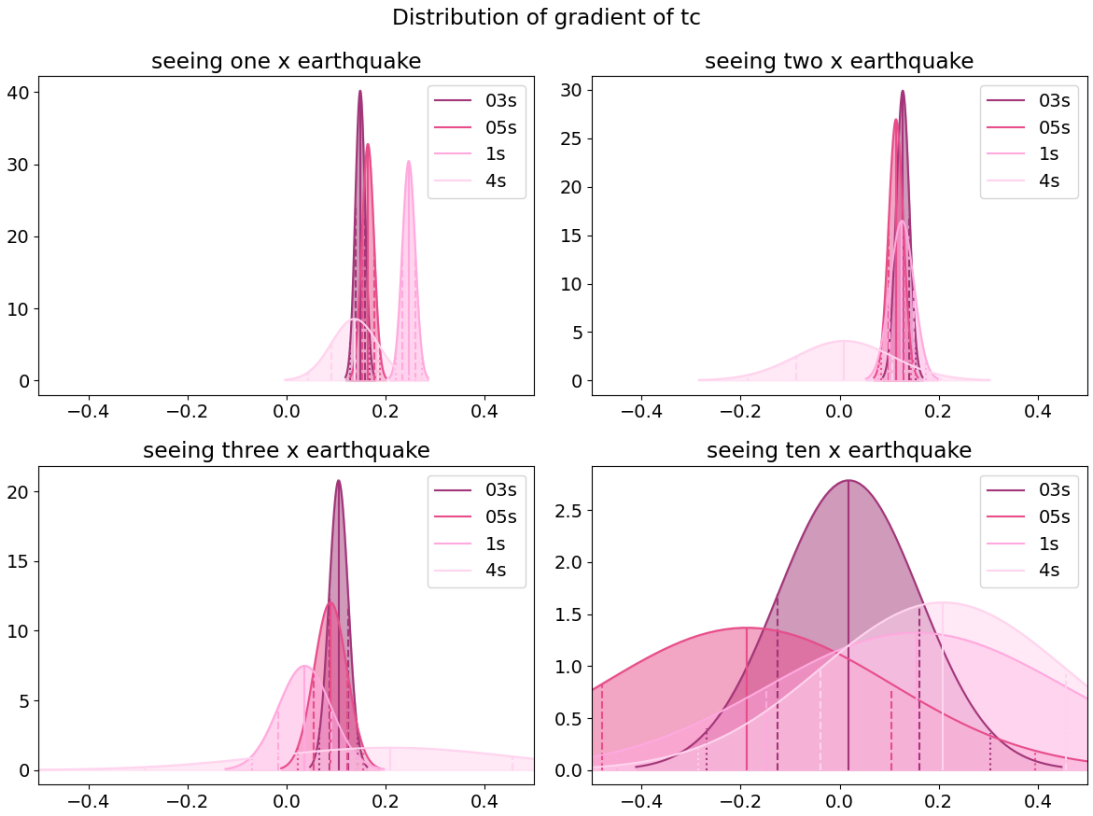


Figure A.46: Distribution of the gradient of the line of best fit found between average period and magnitude. Each panel shows the gradient calculated only considering earthquakes with average duration of one, two, three or ten times the window duration. Each line is a different window length for the calculation from 0.3 s (darkest), 0.5 s, 1 s to 4 s (lightest).

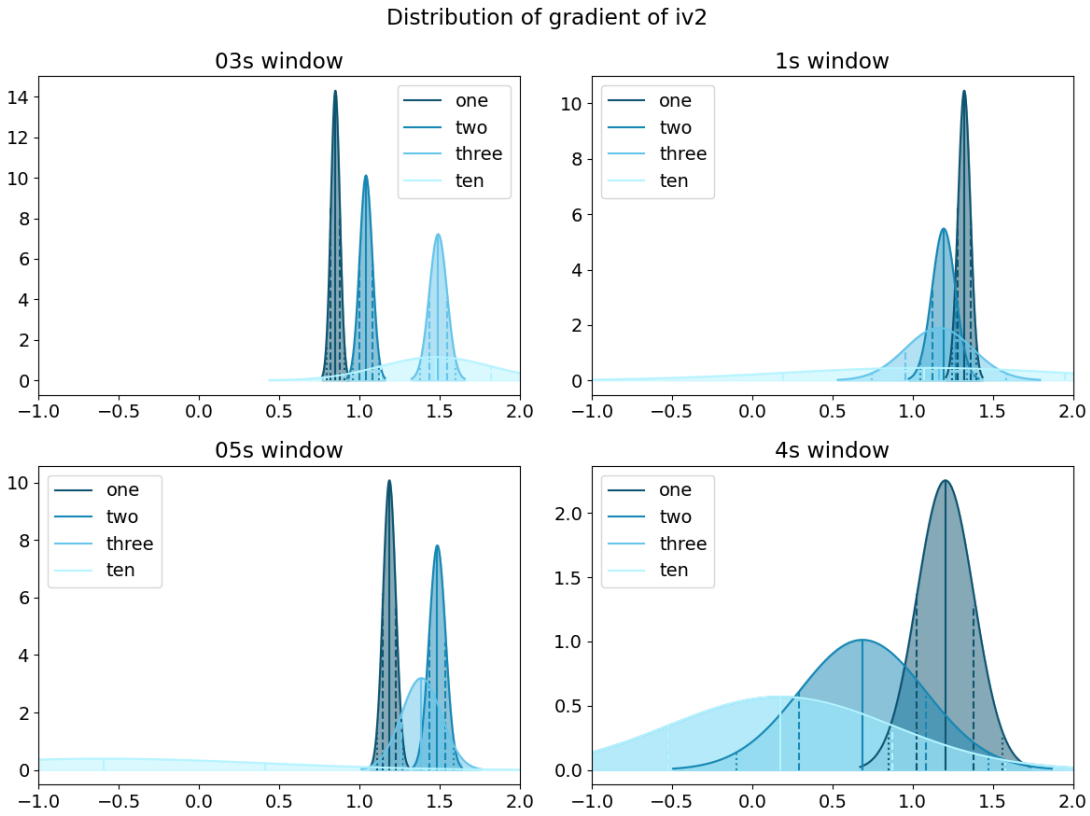


Figure A.47: Distribution of the gradient of the line of best fit found between IV2 and magnitude. Each panel is a different window length for the calculation (clockwise from top right: 0.3 s, 0.5 s, 1 s, 4 s) and each line shows the gradient calculated only considering earthquakes with average duration of one (darkest), two, three or ten (lightest) times the window duration.

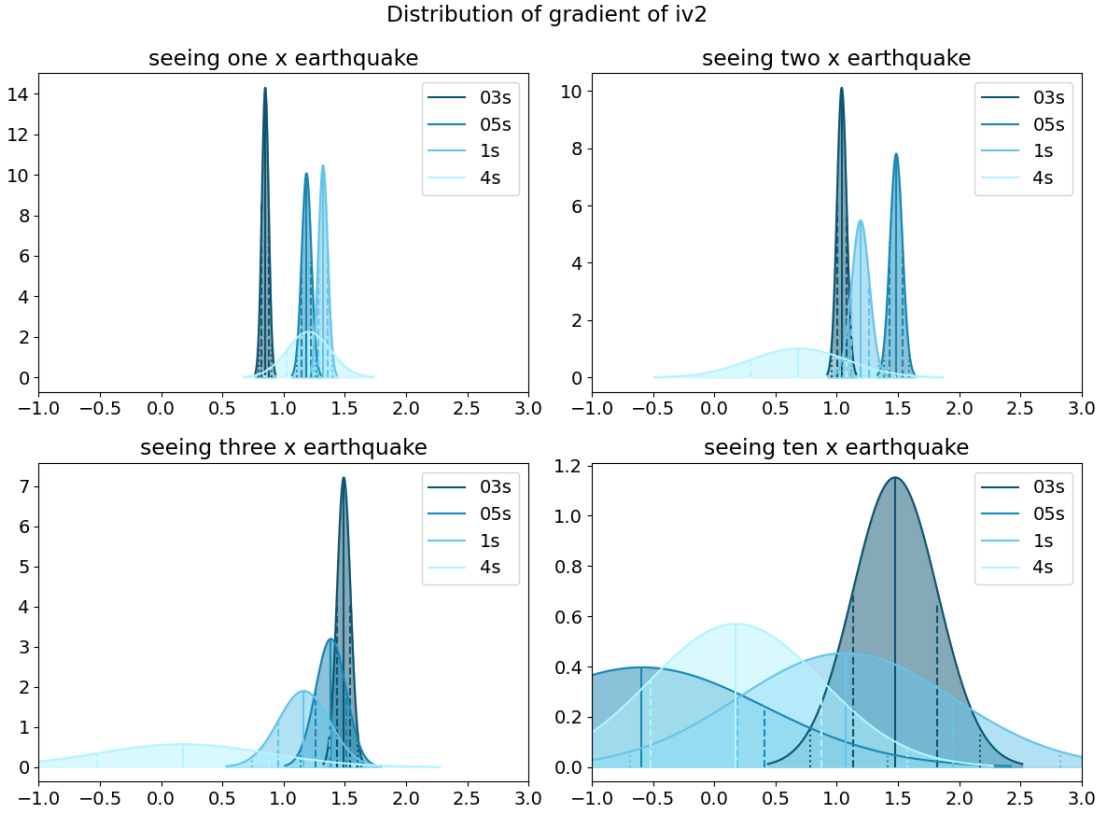


Figure A.48: Distribution of the gradient of the line of best fit found between IV2 and magnitude. Each panel shows the gradient calculated only considering earthquakes with average duration of one, two, three or ten times the window duration. Each line is a different window length for the calculation from 0.3 s (darkest), 0.5 s, 1 s to 4 s (lightest).

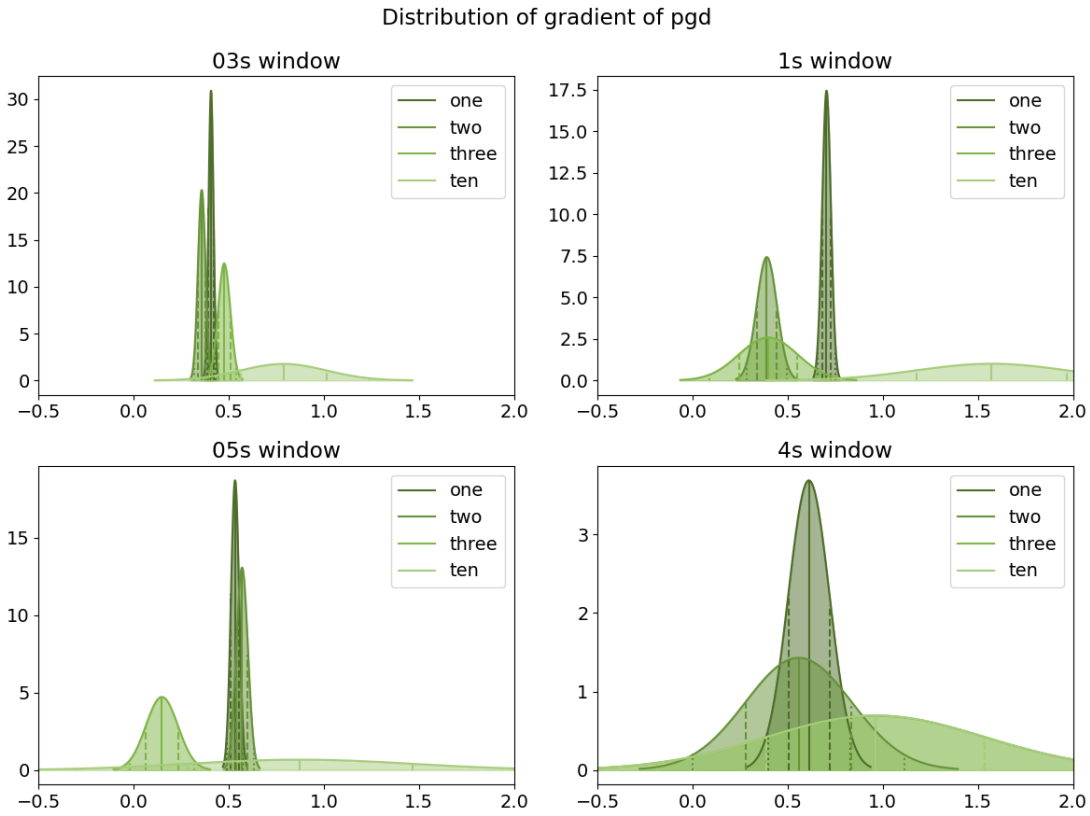


Figure A.49: Distribution of the gradient of the line of best fit found between peak ground displacement and magnitude. Each panel is a different window length for the calculation (clockwise from top right: 0.3 s, 0.5 s, 1 s, 4 s) and each line shows the gradient calculated only considering earthquakes with average duration of one (darkest), two, three or ten (lightest) times the window duration.

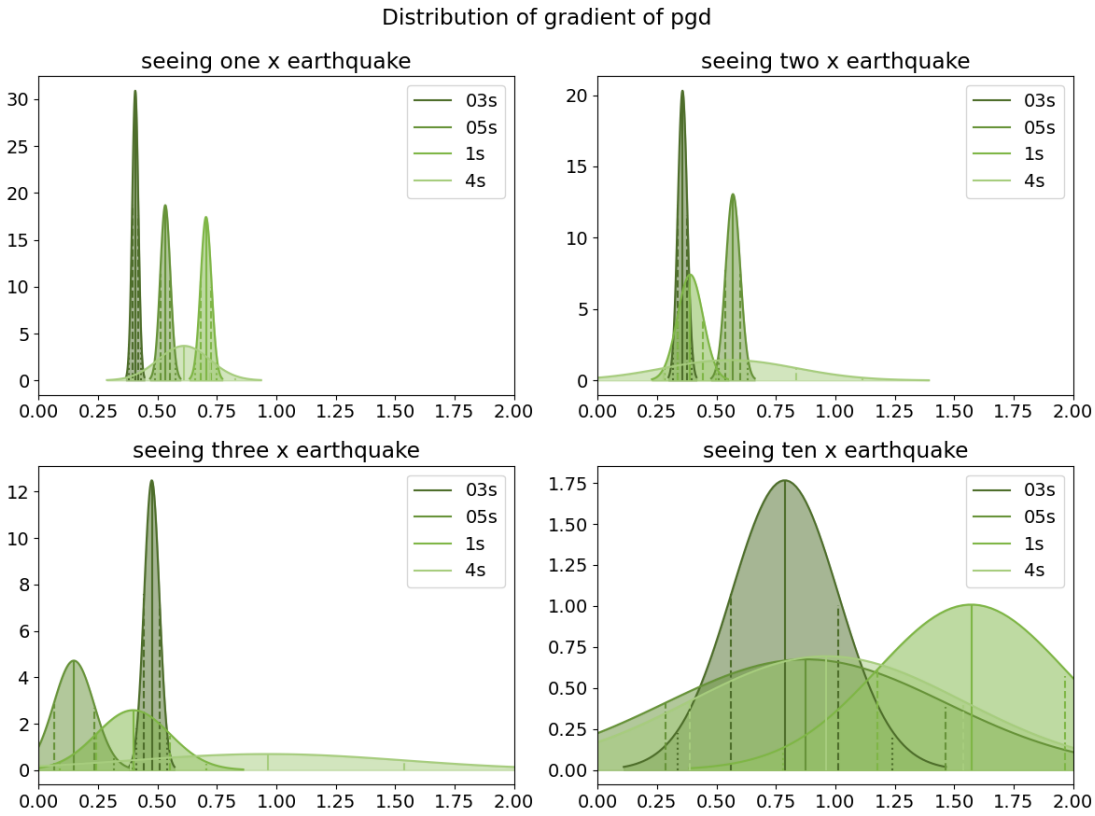


Figure A.50: Distribution of the gradient of the line of best fit found between peak ground displacement period and magnitude. Each panel shows the gradient calculated only considering earthquakes with average duration of one, two, three or ten times the window duration. Each line is a different window length for the calculation from 0.3 s (darkest), 0.5 s, 1 s to 4 s (lightest).

B

Supplementary Information for Chapter 4

Contents

B.1 Ends

Here we show some additional examples of the different ends found using approaches 1, 2, and 3 outlined in Section 4.3. To recap:

1. moment-rate reaches 0
2. moment-rate goes below 0.01 times the maximum moment-rate
3. moment-rate goes below 0.1 times the maximum moment-rate, and over 50% of the total area (moment) has been released

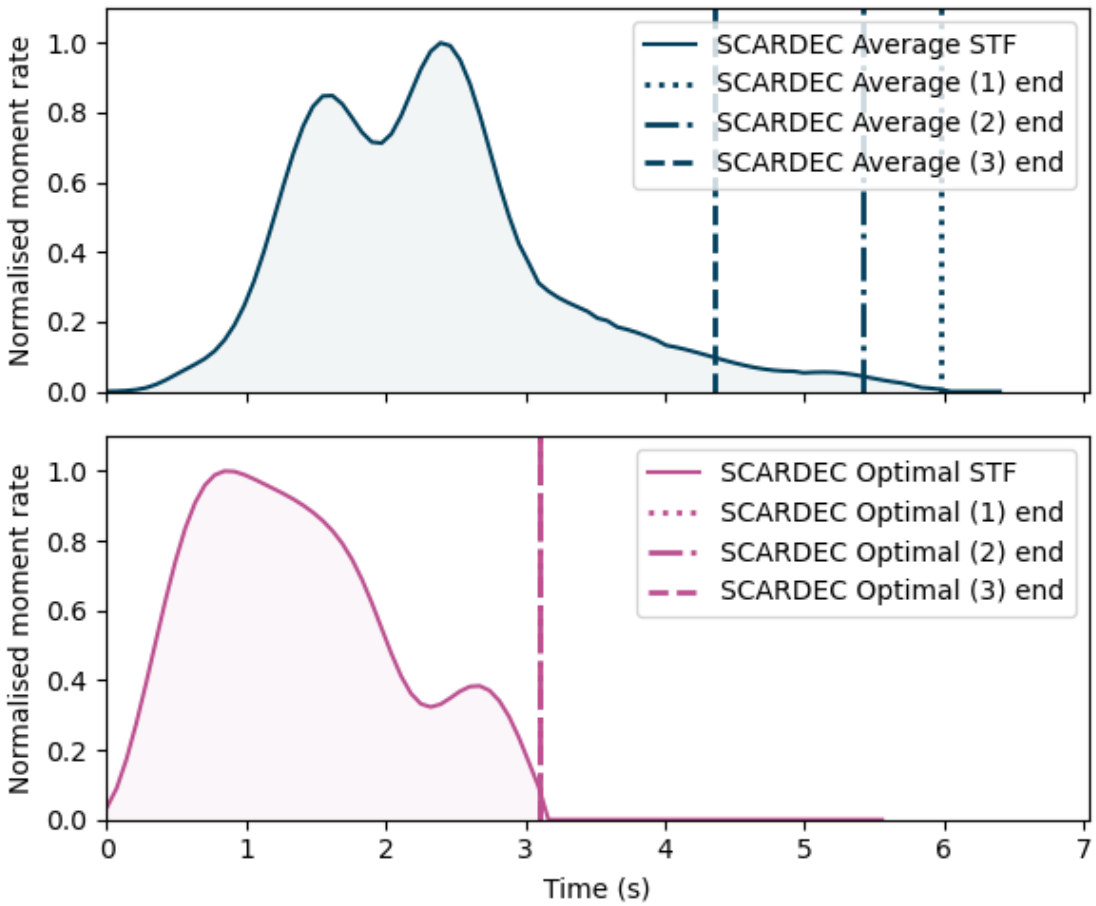


Figure B.1: Normalised STFs of the 2007-03-08 Izu Islands, Japan M6.1 earthquake. The SCARDEC average STF shows a long tail which we want to end at a sensible point.

We show ends from approach 1 in a dotted line (sometimes this is not within the bounds of the x-axis), approach 2 in a dot-dash line and approach 3 in a dashed line. Approaches 2 and 3 are always shown on the figures, but may appear hidden by each other.

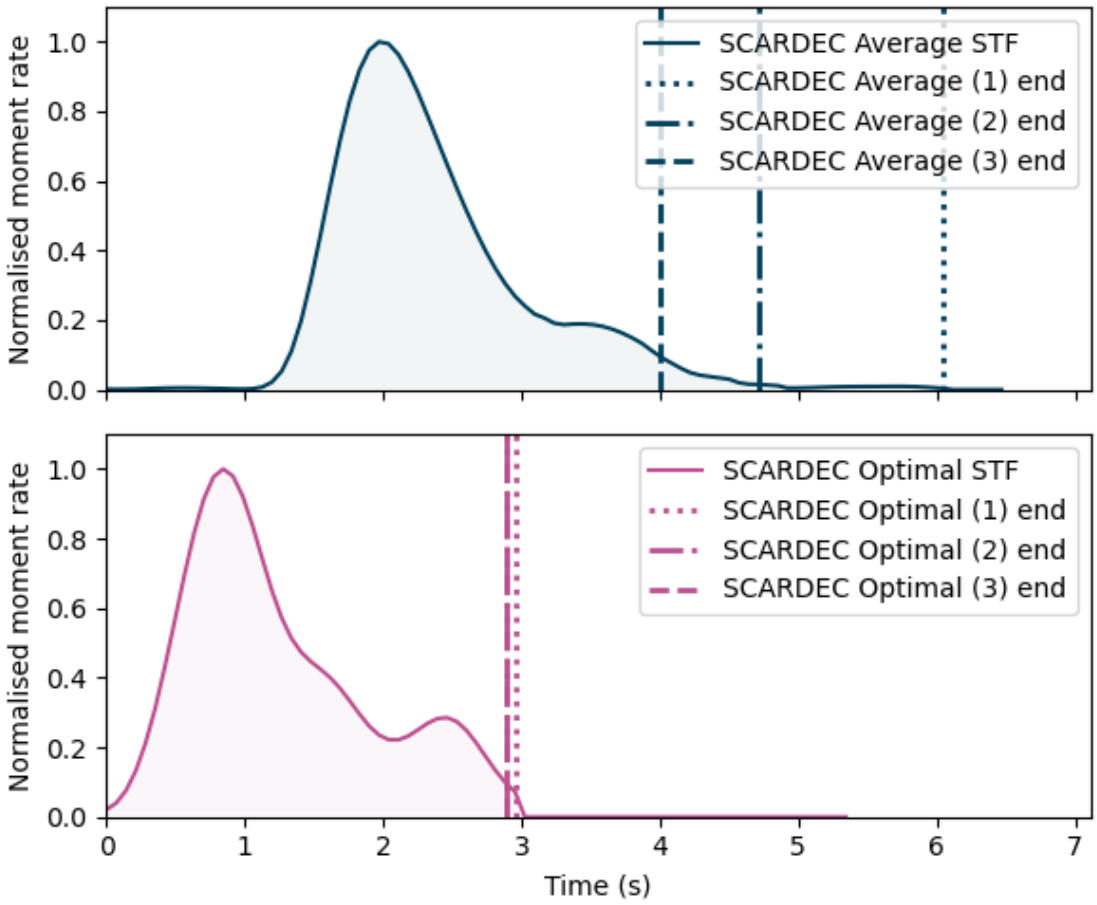


Figure B.2: Normalised STFs of the 2008-01-06 Leonídio, Greece M6.2 earthquake. The SCARDEC average STF shows a long low amplitude tail which we want to end at a sensible point.

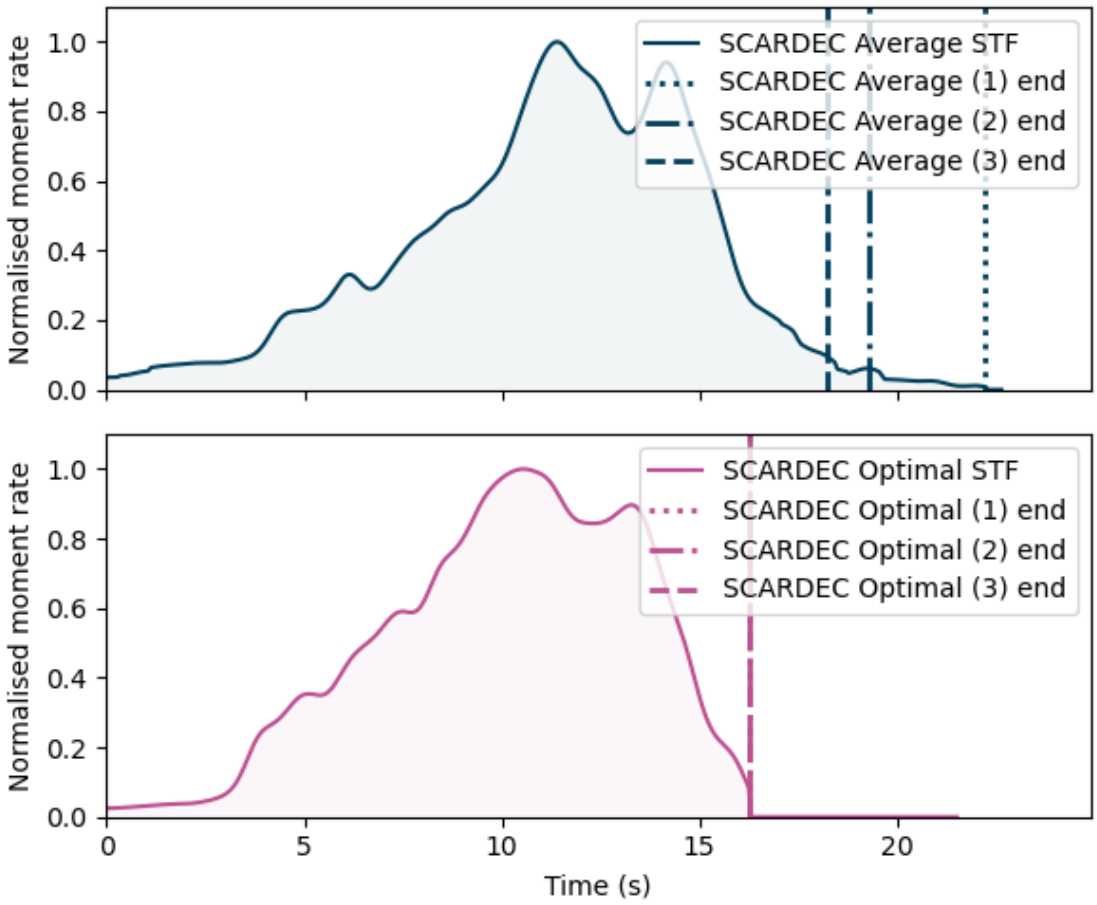


Figure B.3: Normalised STFs of the 2012-09-30 San Agustín, Colombia M7.3 earthquake. The SCARDEC average STF shows a long low amplitude tail which we want to end at a sensible point.

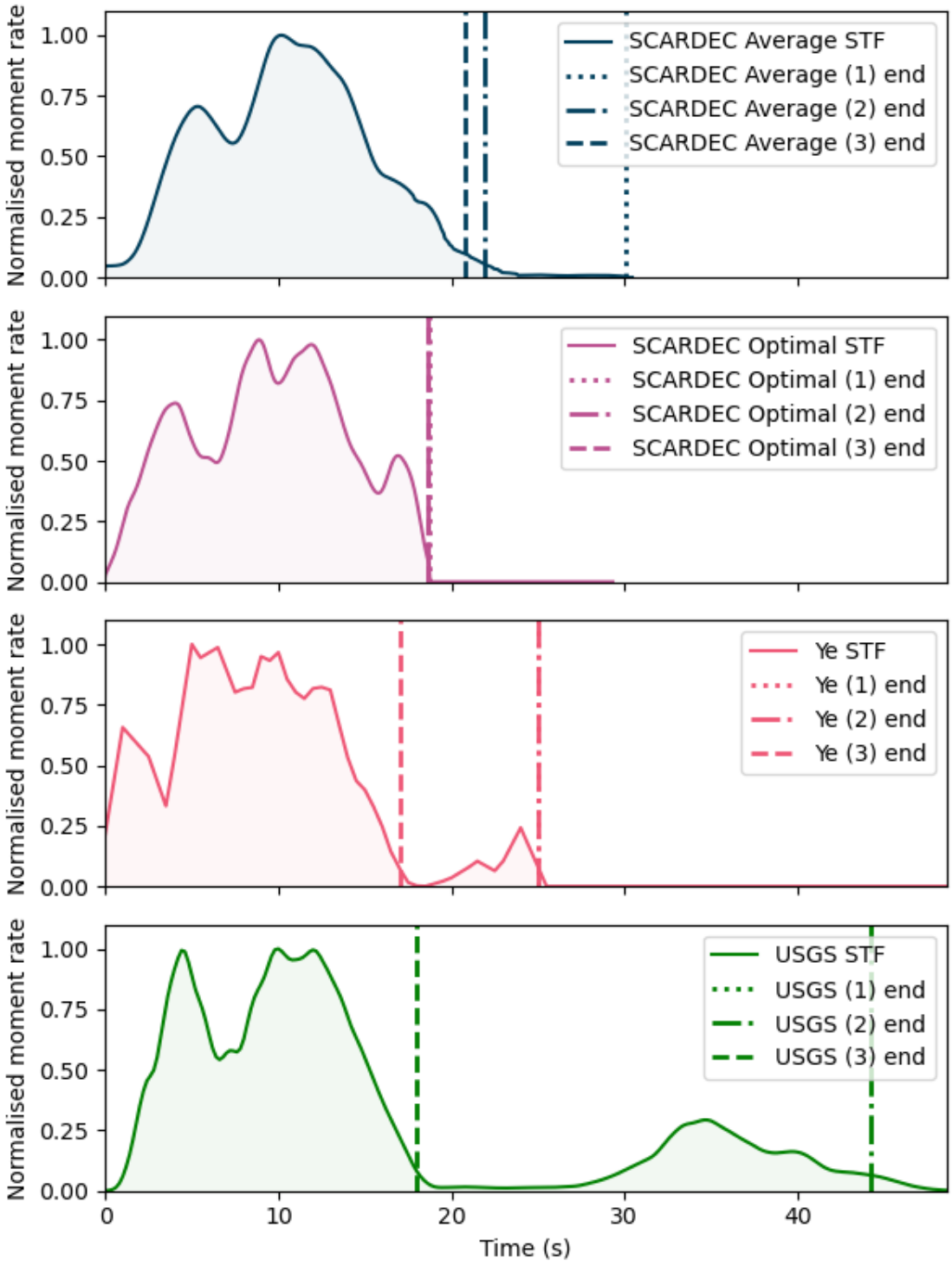


Figure B.4: Normalised STFs of the 2016-12-25 Quellón, Chile M7.6 earthquake. The Ye and USGS STFs have a late final peak which is unlikely to be a legitimate part of the source.

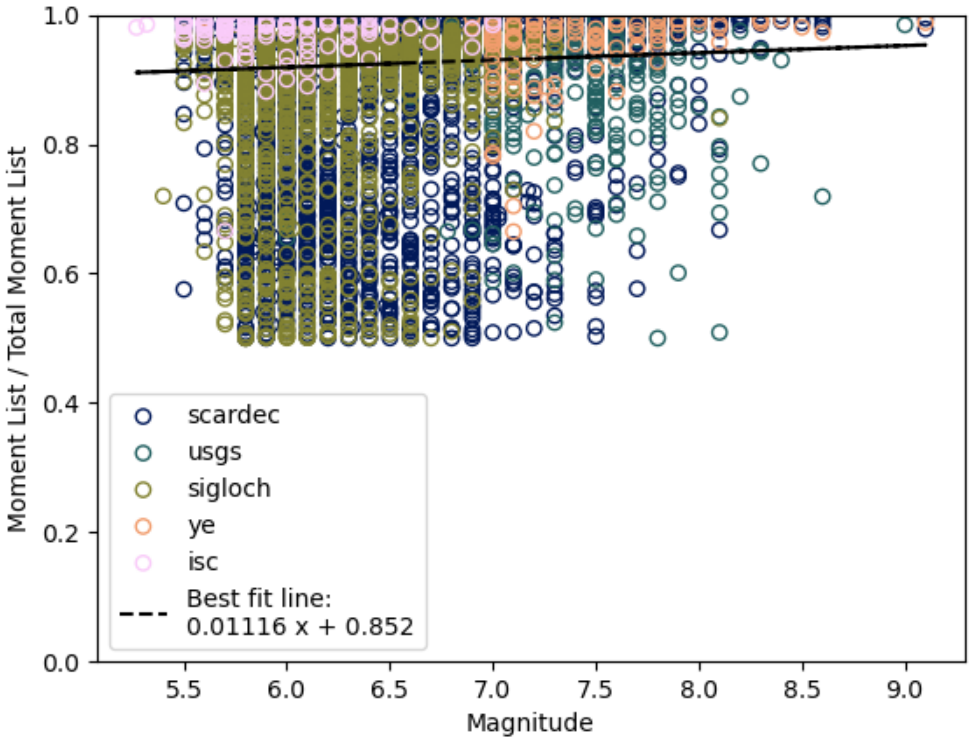


Figure B.5: The proportion of the total area (moment) which remains after the STF end has been detected via method 3. Colours show the different datasets of STFs.

B.1.1 Early cut-offs

As Figure B.5 shows, some events have an interpreted moment that is much smaller than the total interpreted moment. Here (Figures B.6–B.10) we show some example events where a large proportion of the total moment is removed due to an interpreted end being relatively early. In each figure caption, we discuss the validity of the end selected.

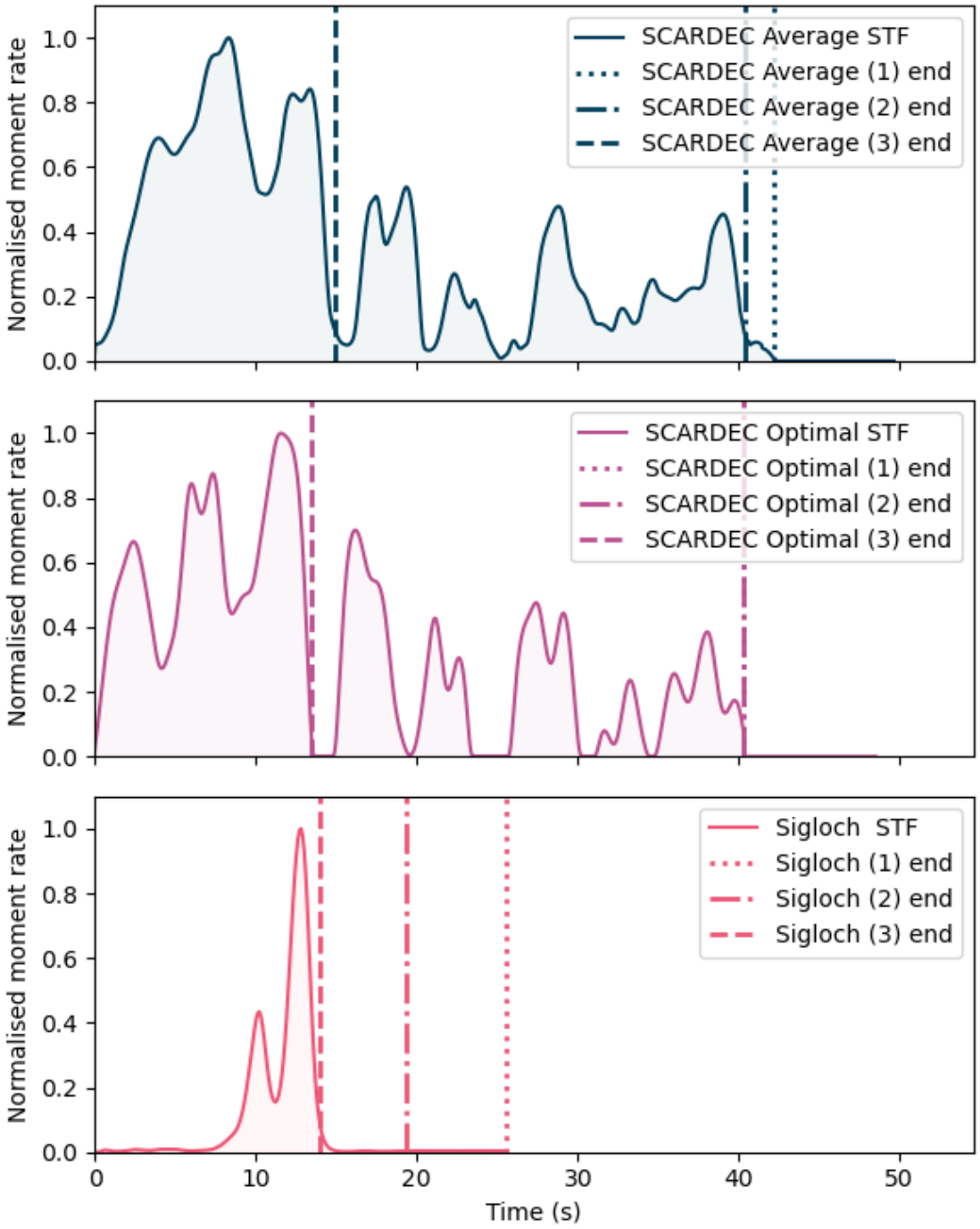


Figure B.6: Normalised STFs of the 1992-07-20 07:46 Svalbard M6.6 earthquake. Both SCARDEC STFs are long and complex, though the end interpreted by method (3) is reasonable, particularly for Optimal SCARDEC.

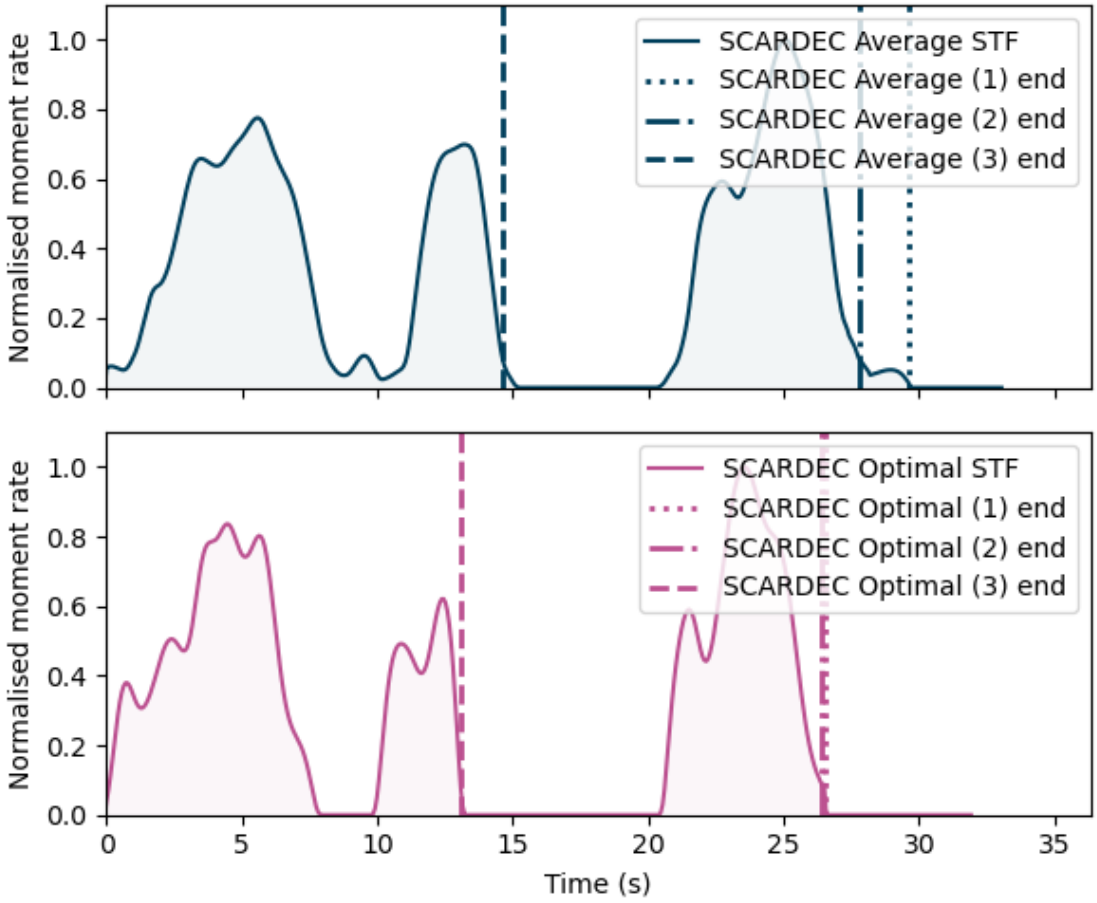


Figure B.7: Normalised STFs of the 1994-06-03 21:06 South of Java, Indonesia M6.3 earthquake. Both SCARDEC STFs have three main peaks, with periods of low or no moment release betwen them. Arguably, the method (3) end is too late, since the moment rate has already reached zero for a time.

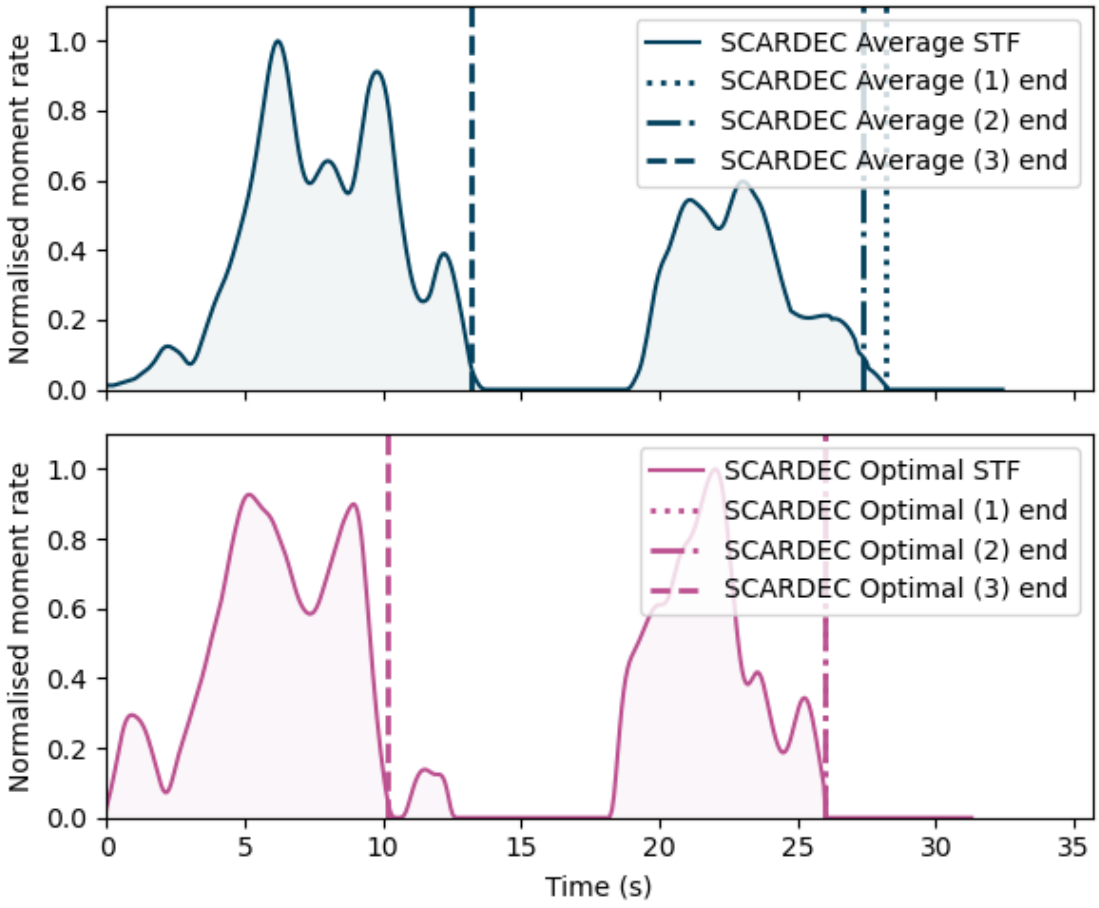


Figure B.8: Normalised STFs of the 1995-02-10 01:45 North Island, New Zealand M6.3 earthquake. This is an example where despite the interpreted moment being significantly less than the interpreted total moment, the end interpreted through method (3) is reasonable.

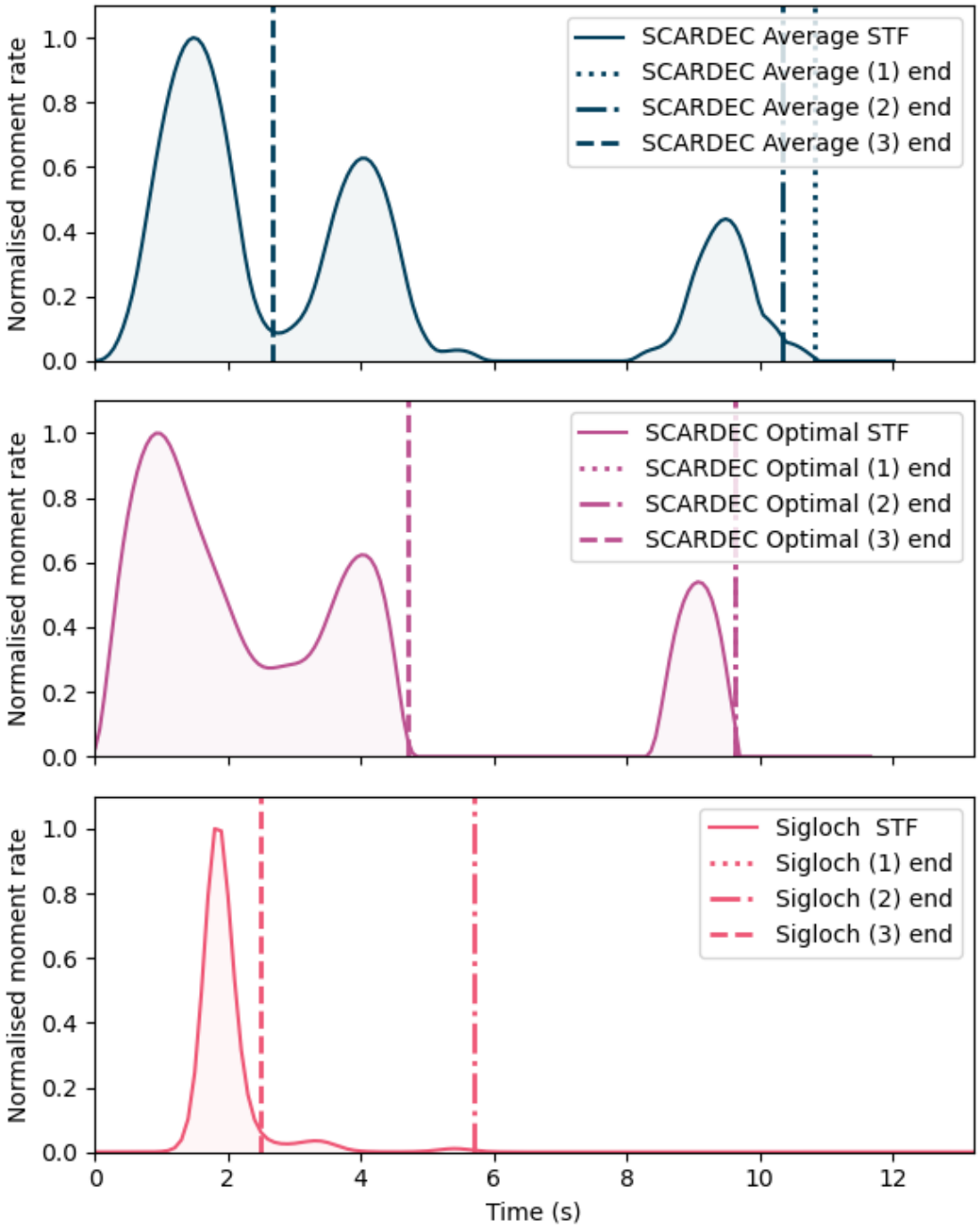


Figure B.9: Normalised STFs of the 2001-07-04 17:47 Honshu, Japan M5.8 earthquake. Here, the pick for the optimal SCARDEC STF seems reasonable, but the pick for the average SCARDEC STF is too early.

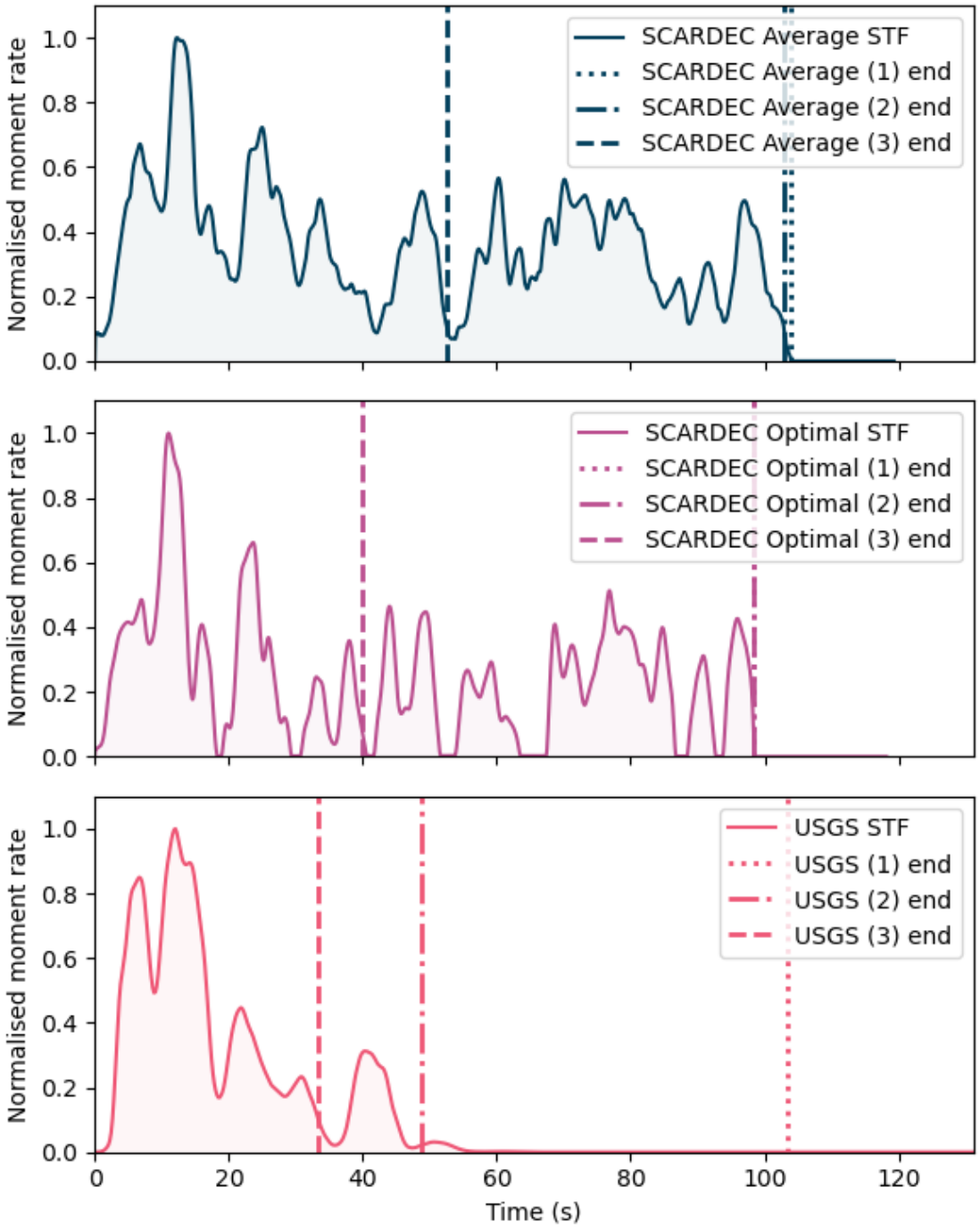


Figure B.10: Normalised STFs of the 2002-10-10 10:50 Irian Jaya Region, Indonesia M7.5 earthquake. Here, the pick for the optimal SCARDEC STF seems reasonable, but the pick for the average SCARDEC STF is too early.

B.2 Median STFs by magnitude and dataset

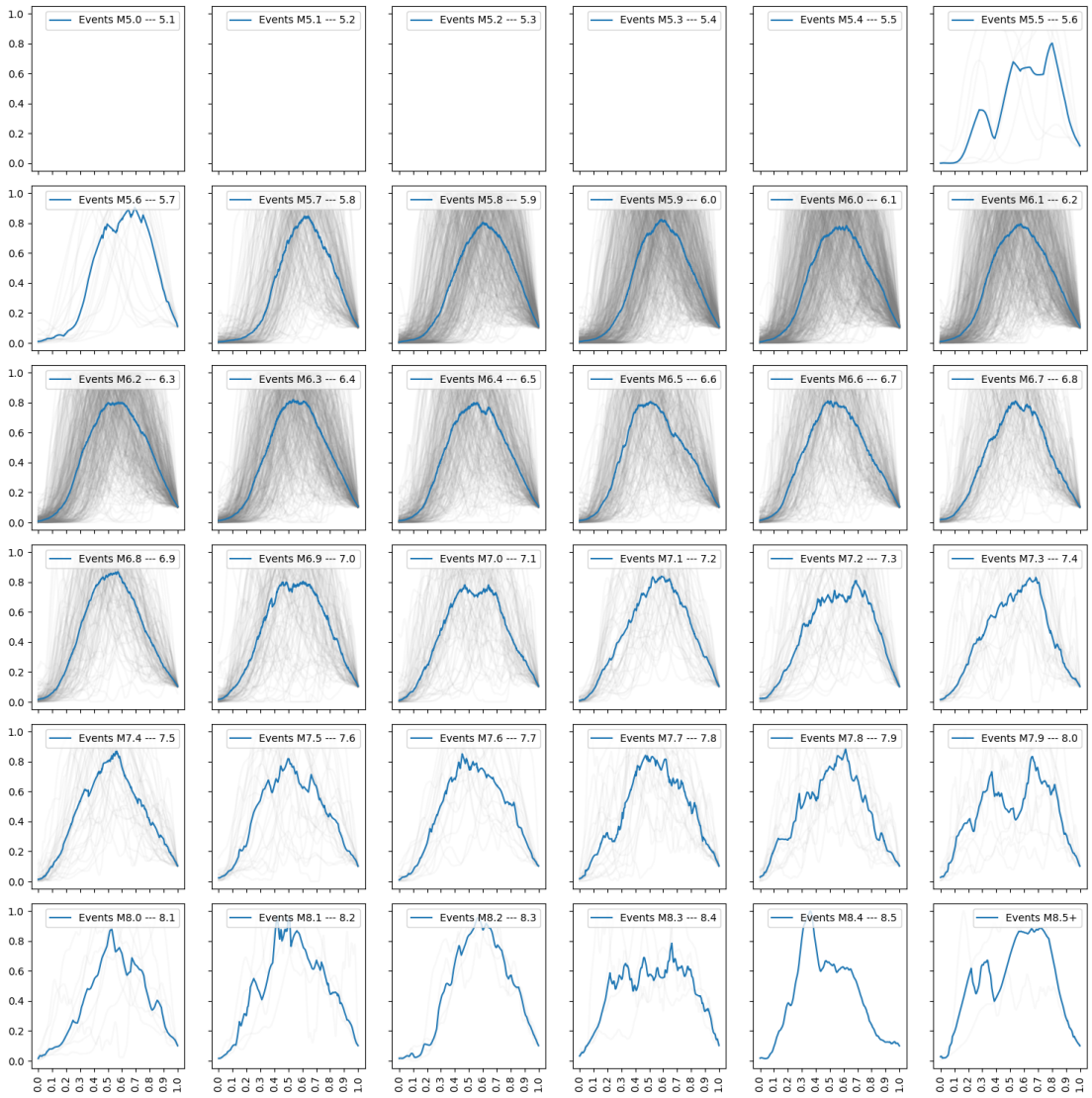


Figure B.11: The median STF shape for $M0.1$ bins. Each panel is a single bin, with all normalised STFs available from SCARDEC average STFs for that bin in grey, and the median in blue.

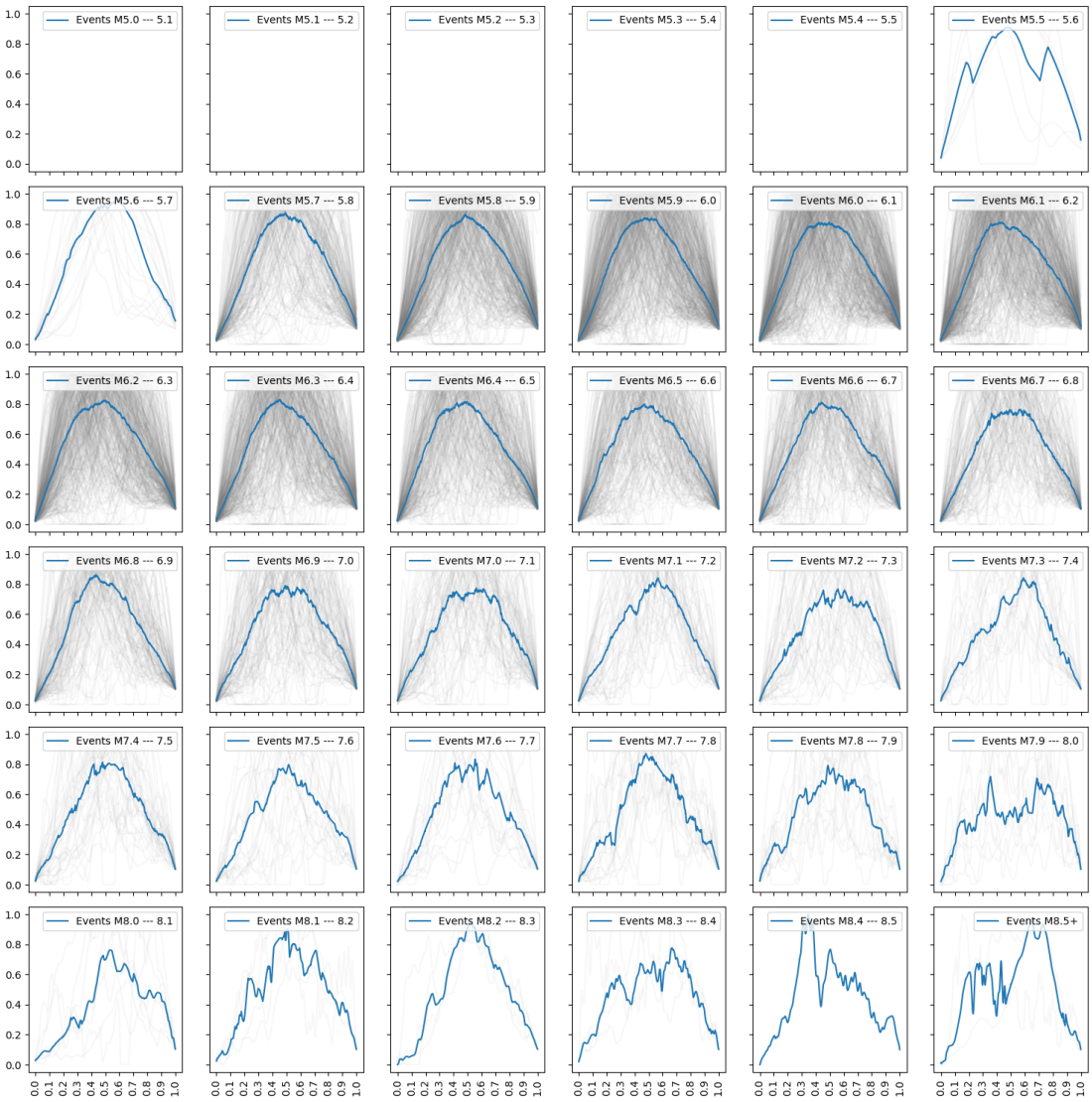


Figure B.12: As Figure B.11 but for the optimal SCARDEC catalog.

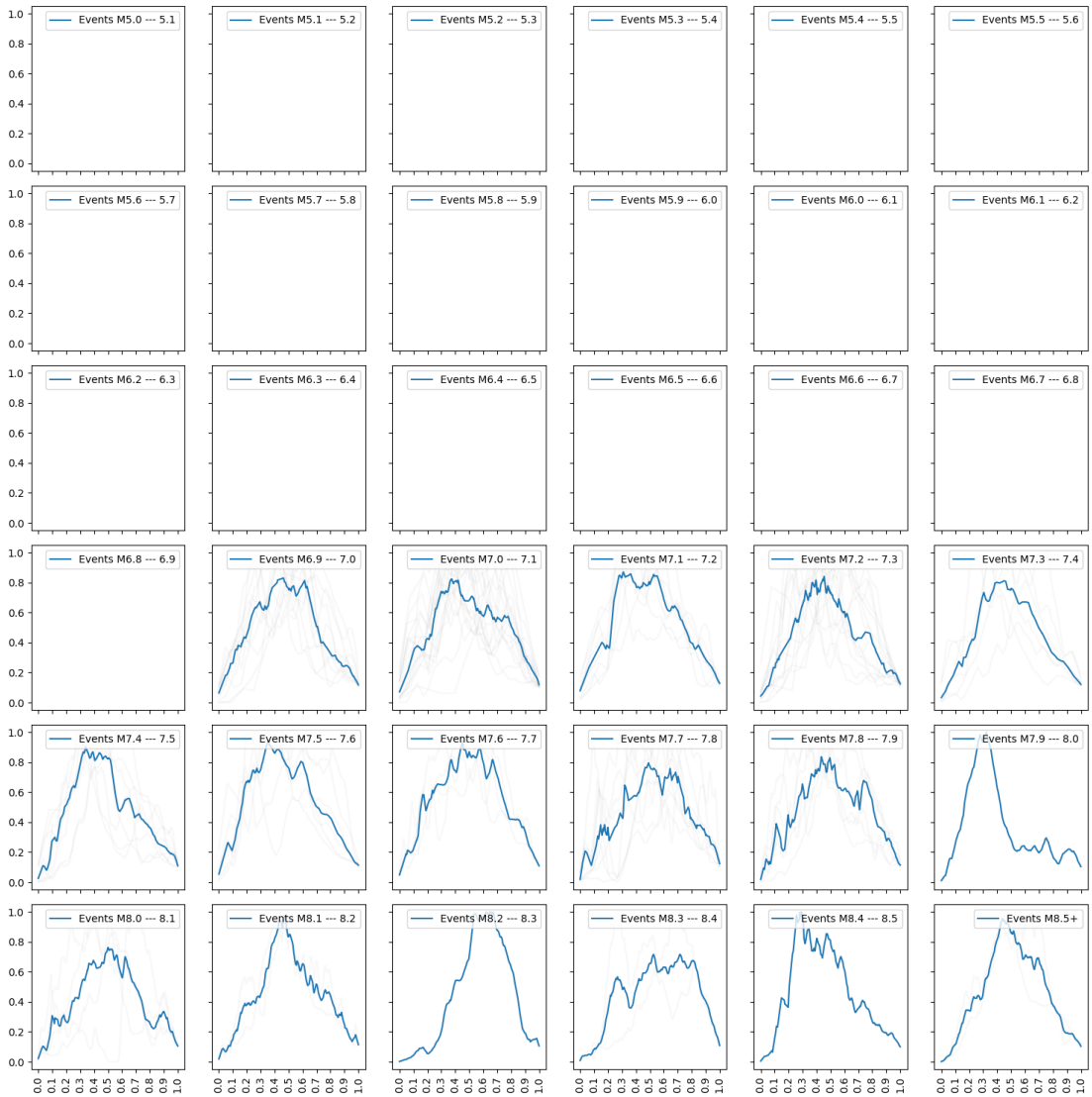


Figure B.13: As Figure B.11 but for the Ye catalog.

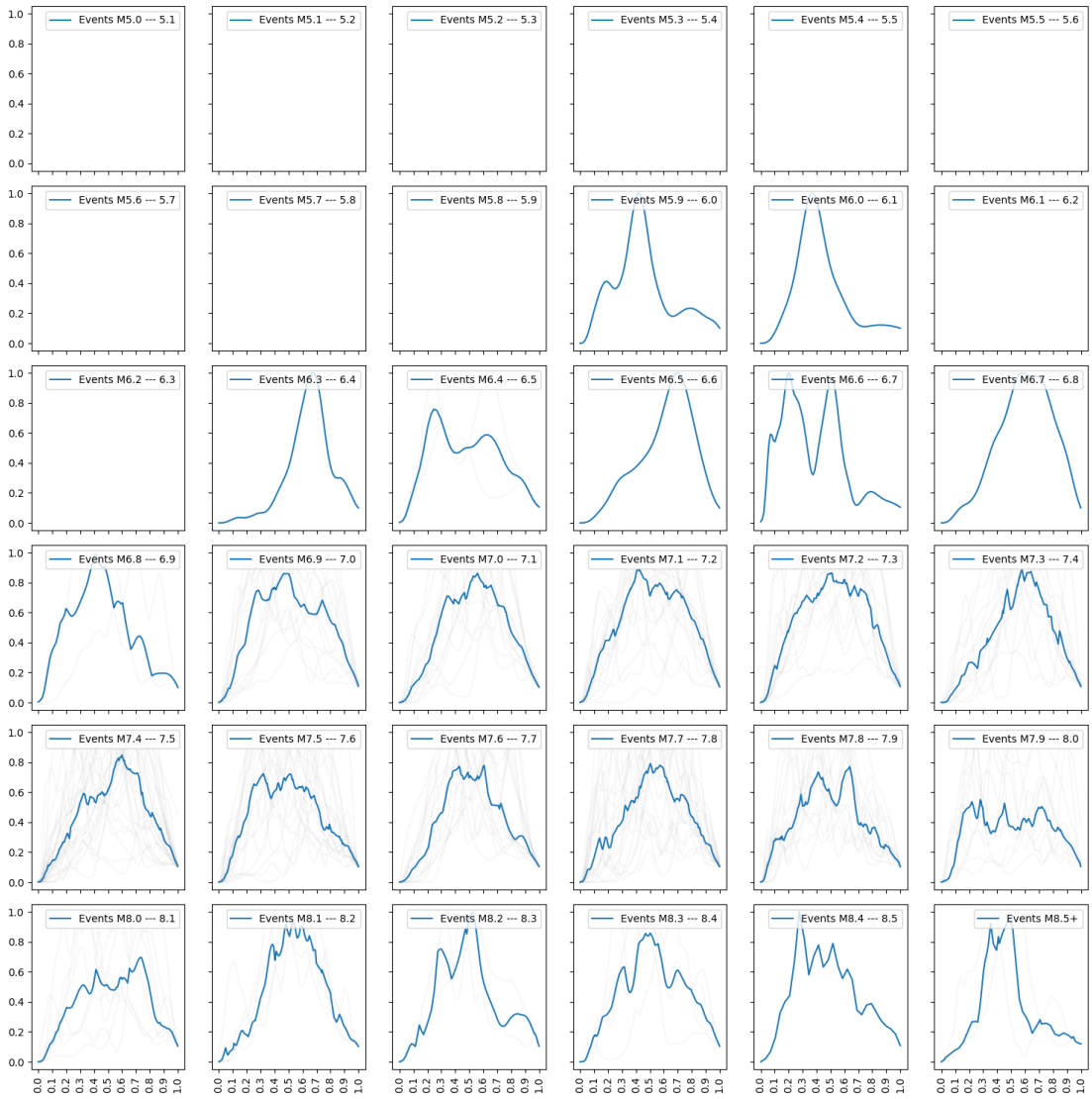


Figure B.14: As Figure B.11 but for the USGS catalog.

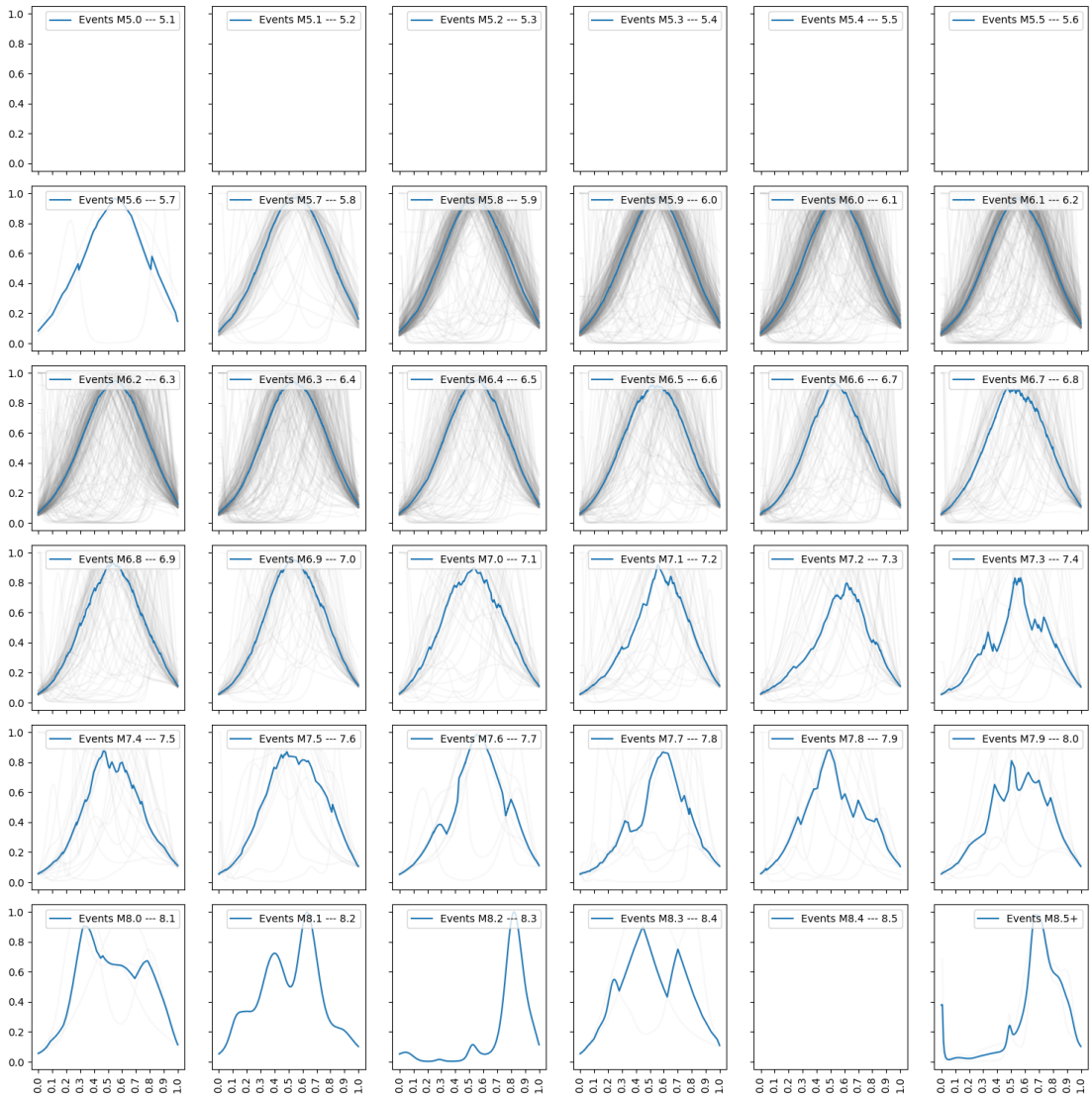


Figure B.15: As Figure B.11 but for the Sigloch catalog.

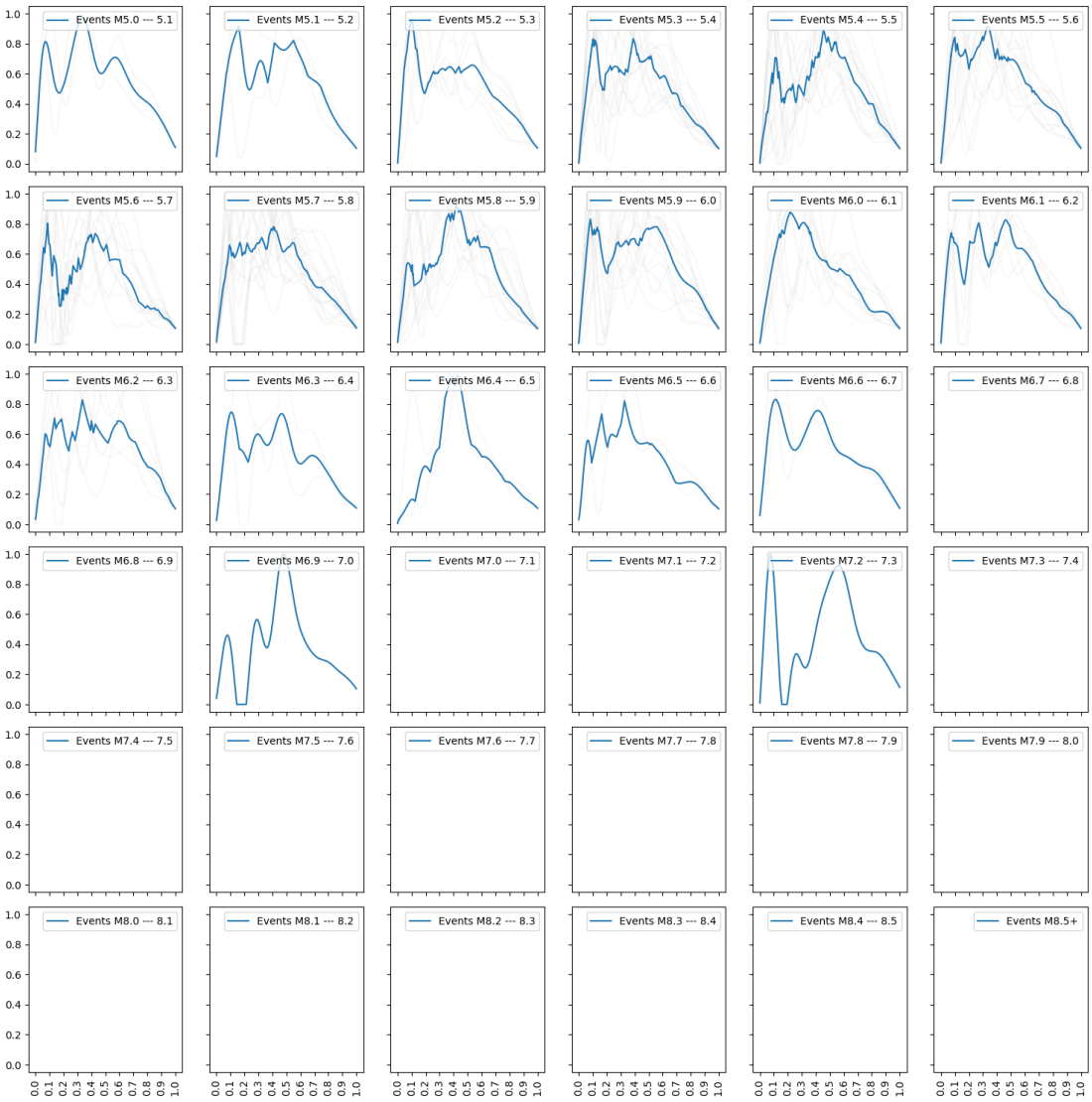


Figure B.16: As Figure B.11 but for the ISC catalog.

B.3 Absolute moment released in absolute time

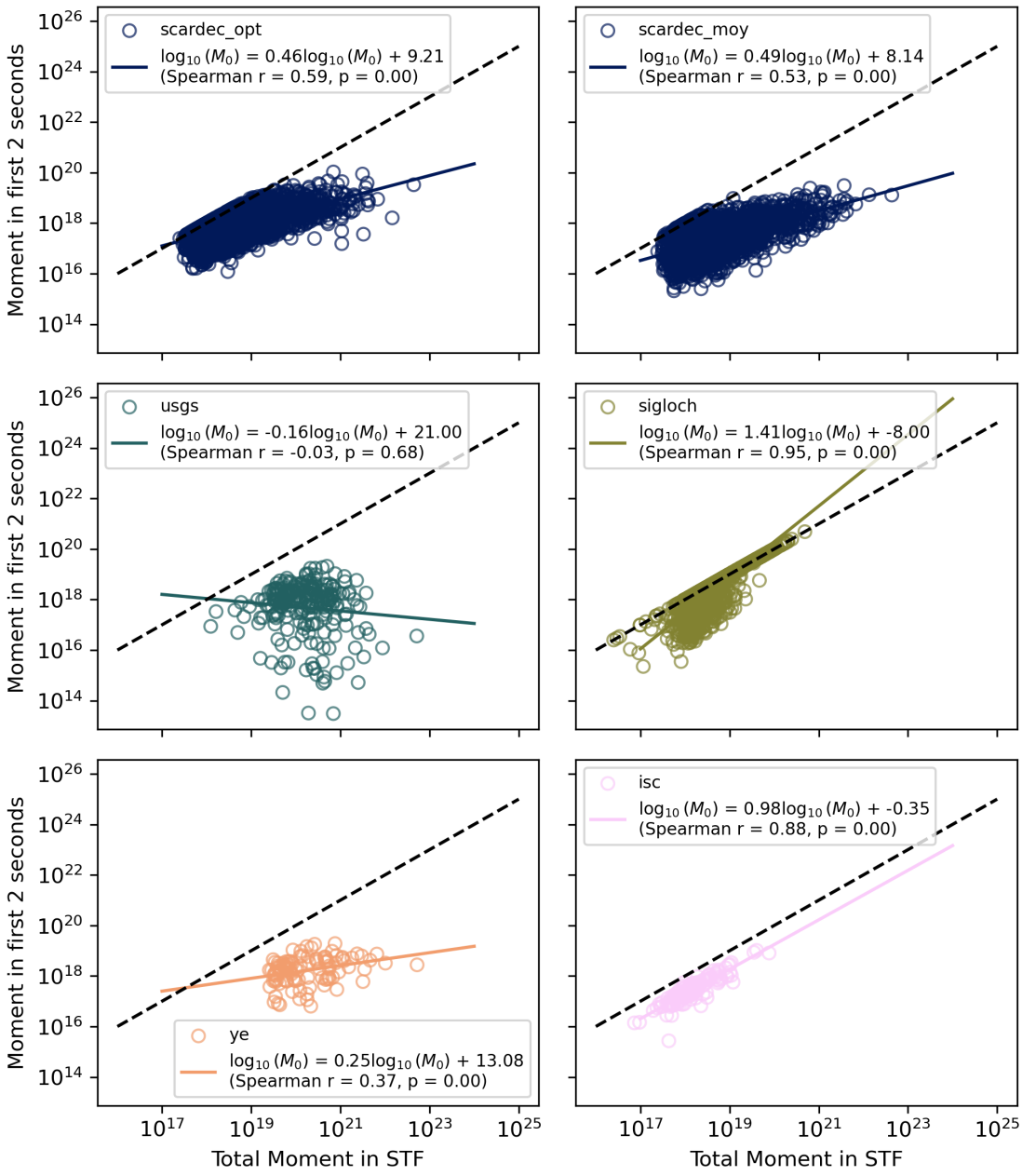


Figure B.17: Absolute moment (Nm) released in first two seconds of earthquake.

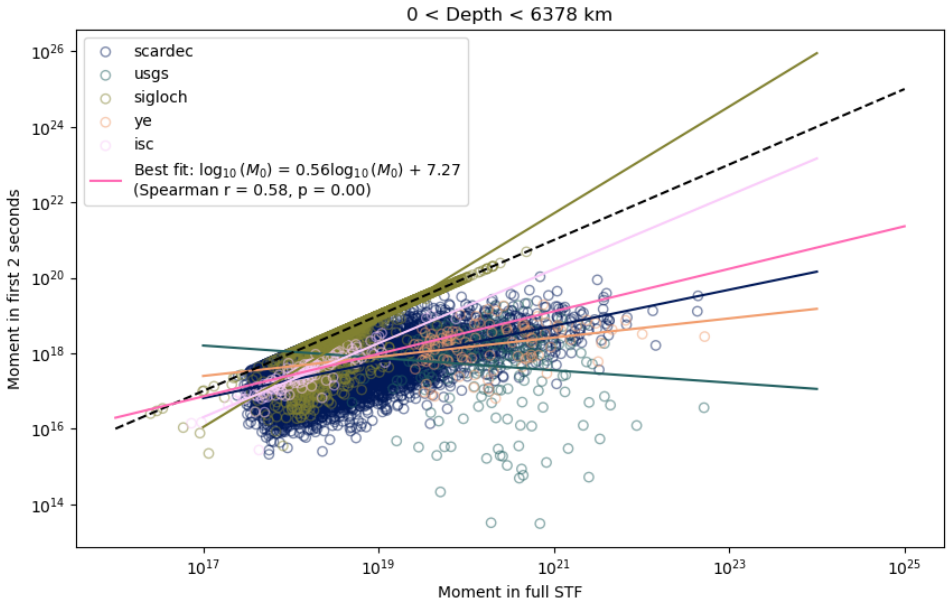


Figure B.18: Absolute moment (Nm) released in first two seconds of earthquake, with each color denoting a different dataset.

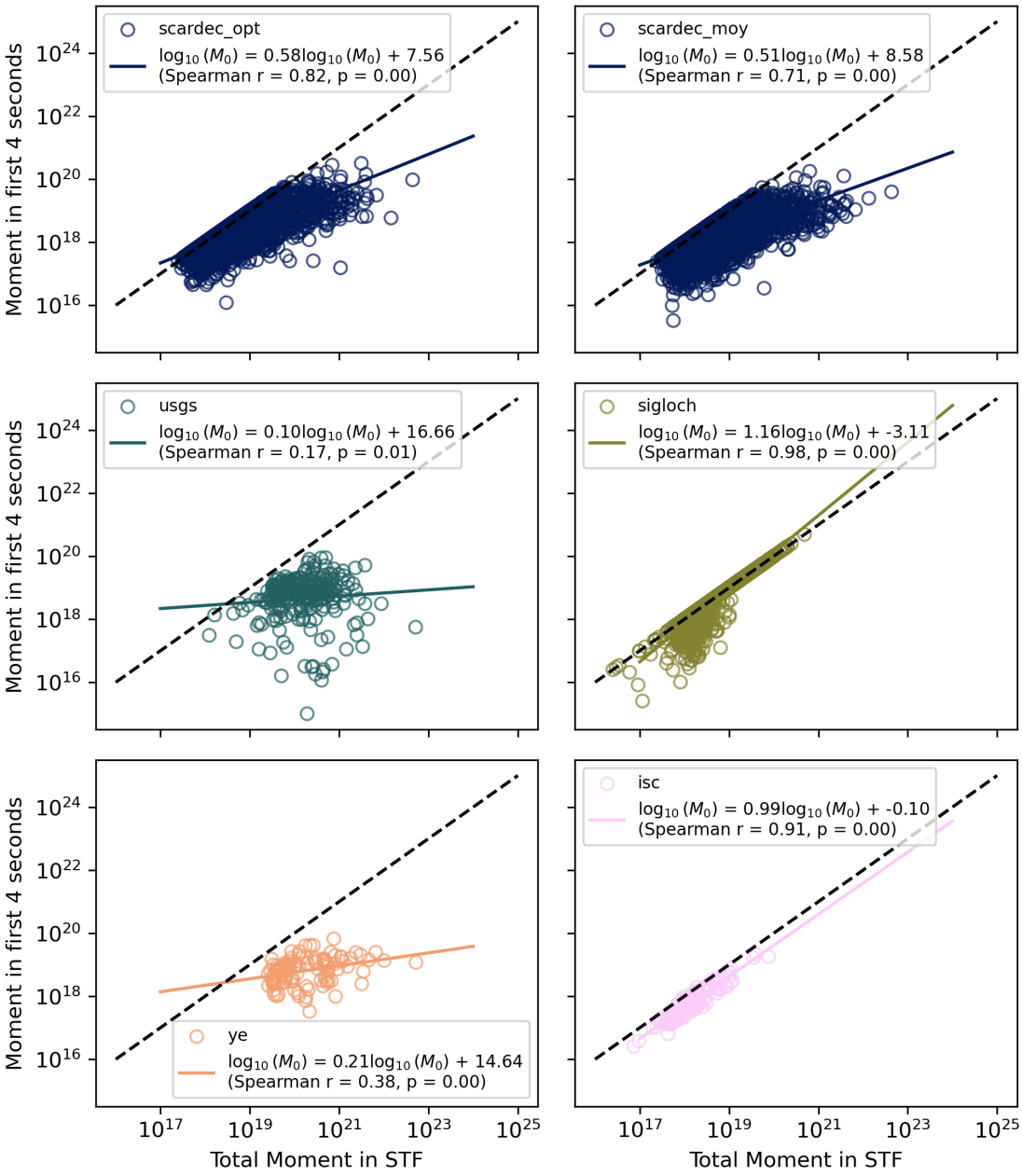


Figure B.19: Absolute moment (Nm) released in first four seconds of earthquake

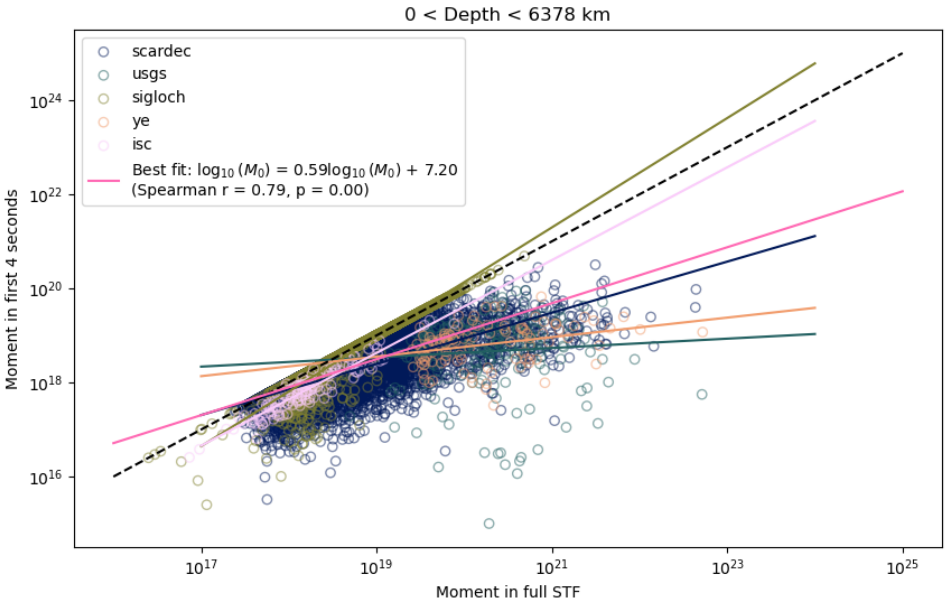


Figure B.20: Absolute moment (Nm) released in first four seconds of earthquake, with each color denoting a different dataset.

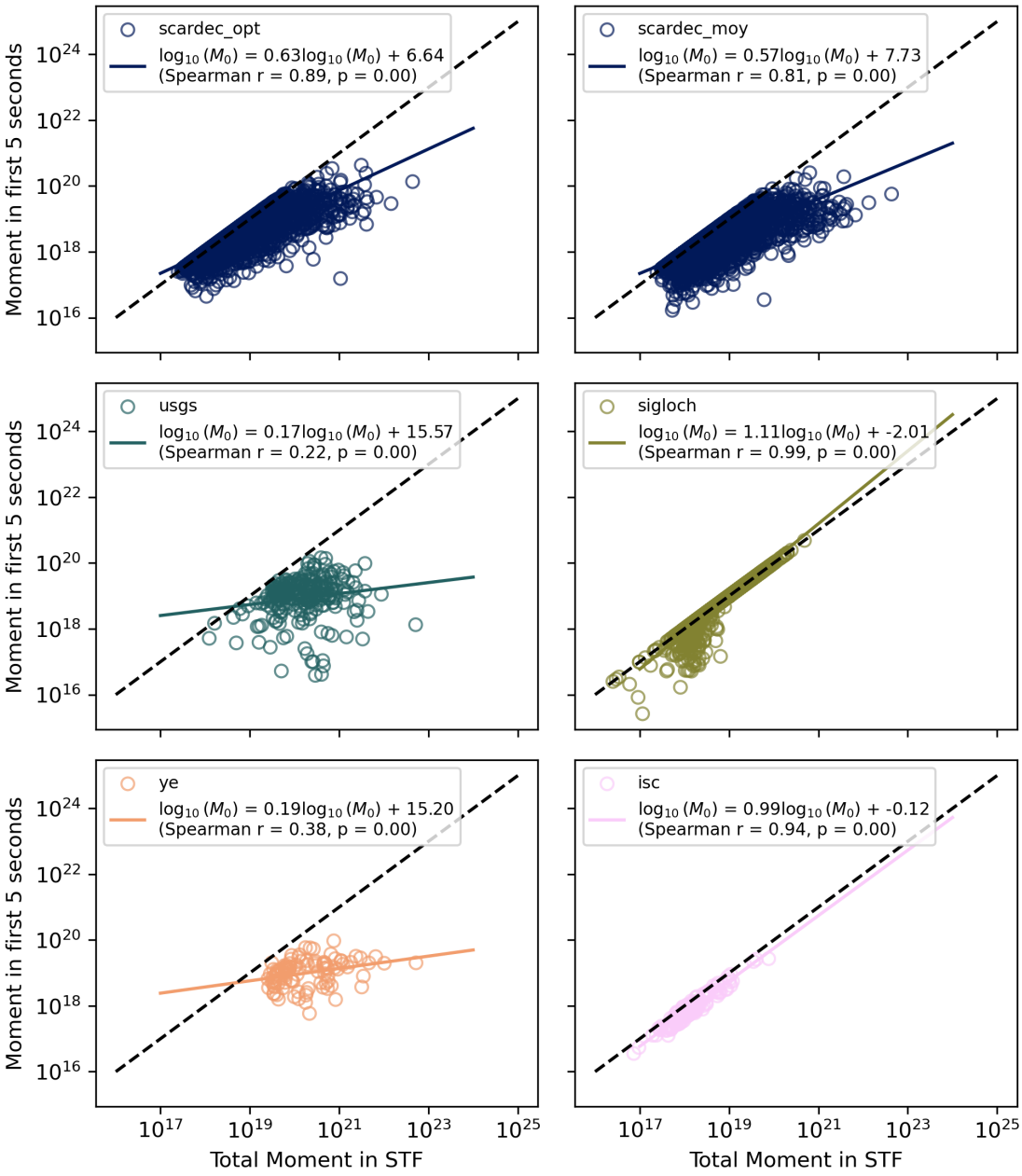


Figure B.21: Absolute moment (Nm) released in first five seconds of earthquake.

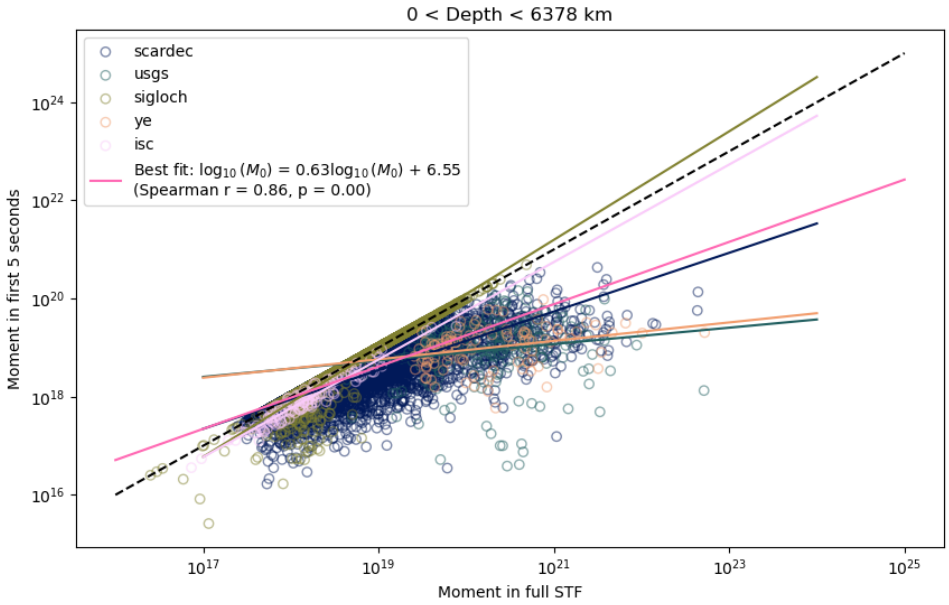


Figure B.22: Absolute moment (Nm) released in first five seconds of earthquake, with each color denoting a different dataset.

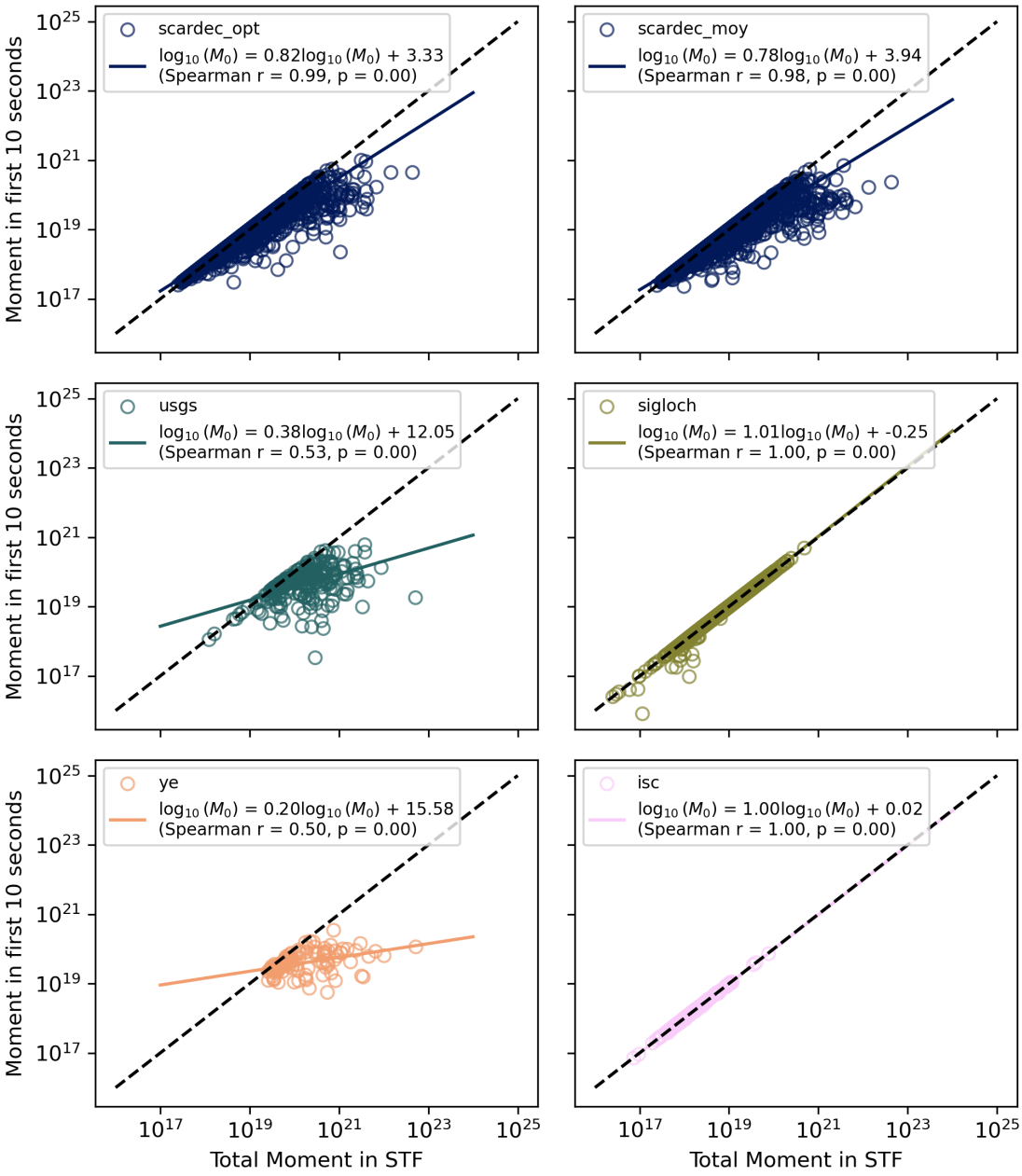


Figure B.23: Absolute moment (Nm) released in first ten seconds of earthquake

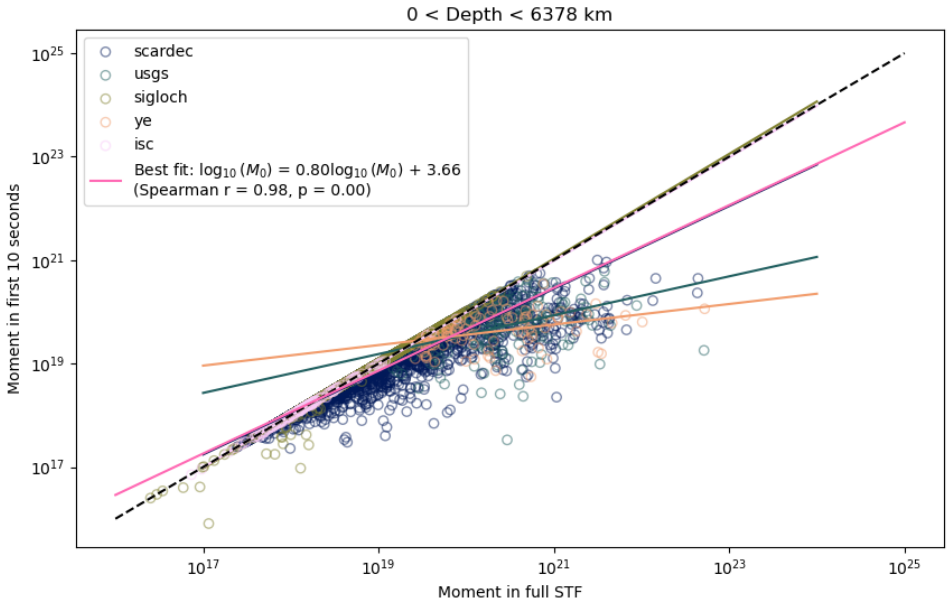


Figure B.24: Absolute moment (Nm) released in first ten seconds of earthquake, with each color denoting a different dataset.

B.4 Proportion of moment released in absolute time

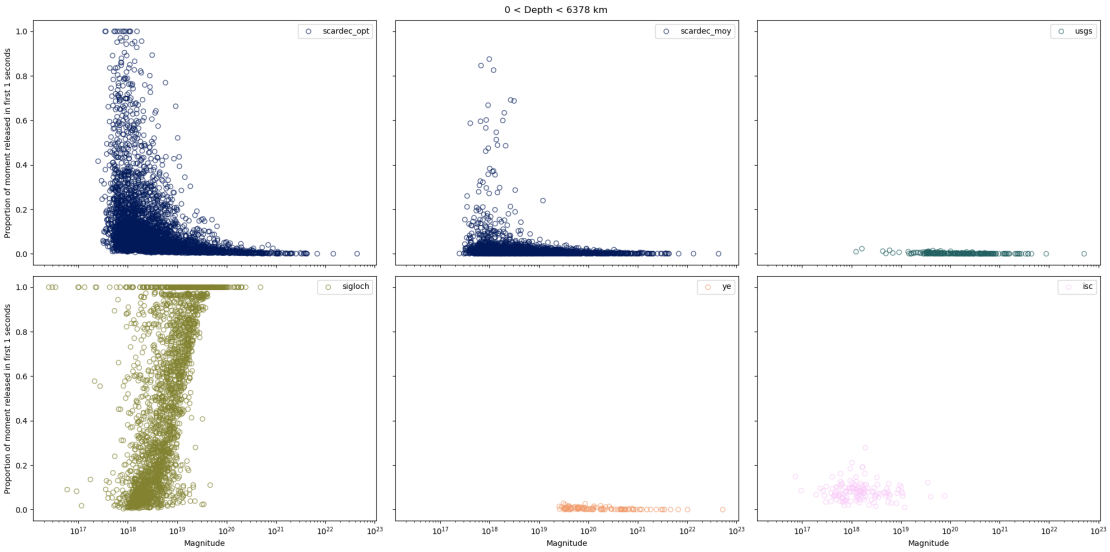


Figure B.25: Proportion of total moment (Nm) released in first second of earthquake.

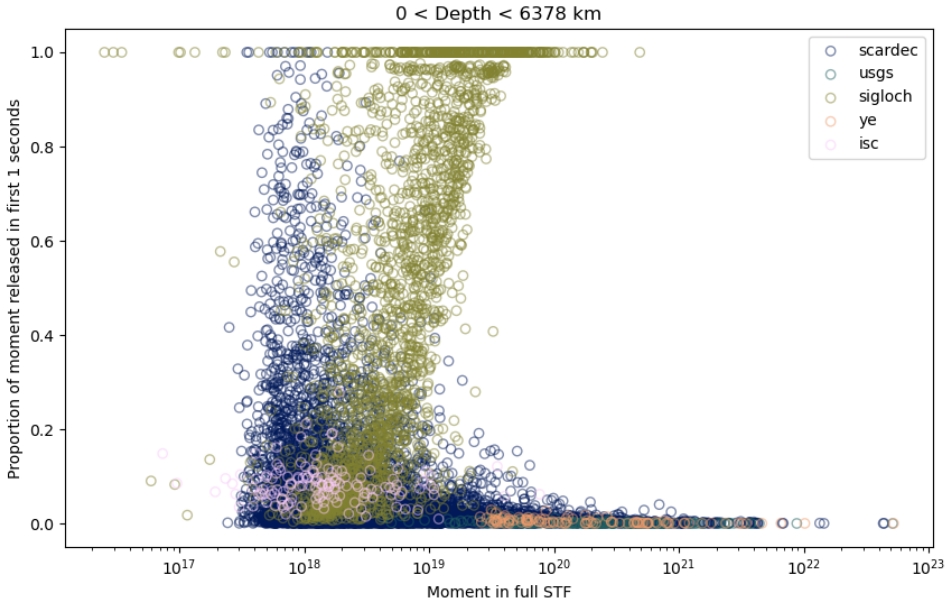


Figure B.26: Proportion of total moment (Nm) released in first second of earthquake, colors represent each dataset.

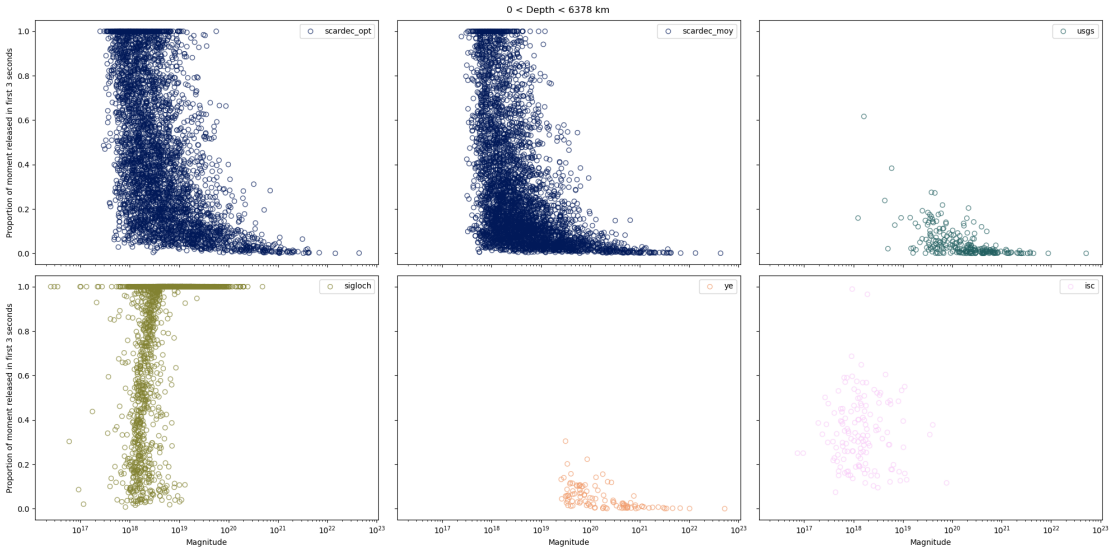


Figure B.27: Proportion of total moment (Nm) released in first three seconds of earthquake.

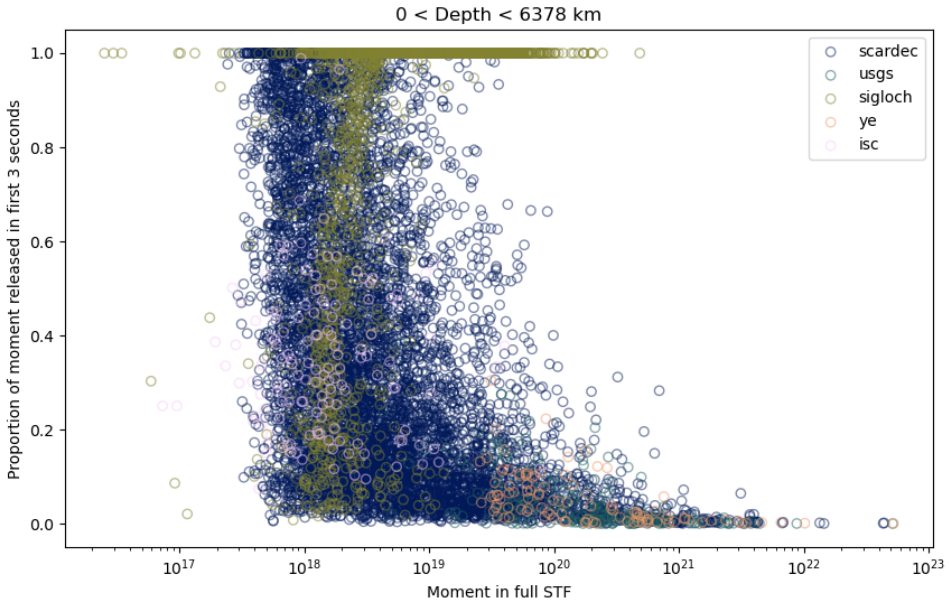


Figure B.28: Proportion of total moment (Nm) released in first three seconds of earthquake, colors denote dataset.

B.5 Absolute time to release absolute moment

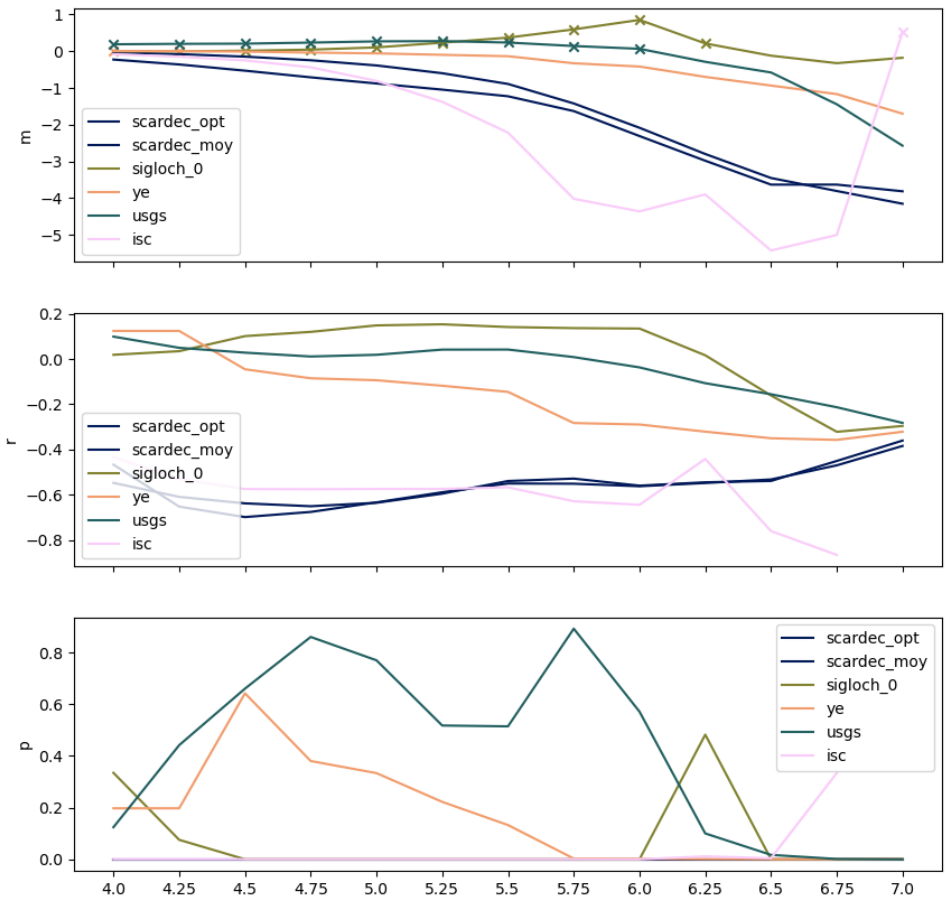


Figure B.29: Plots of the data shown in Tables B.1–B.4. In the top panel, crosses show where the gradient is positive. The top panel gives the gradient (m) of the relationship between the time to release moment equal to some magnitude of earthquake (x-axis) and the final earthquake magnitude. The middle panel gives the spearman correlation coefficient (r) of this relationship, and the bottom panel gives the probability (p) of this correlation coefficient arising from linearly uncorrelated data.

Table B.1: Gradient of the relationship between the time to release moment equal to some magnitude of earthquake and the final earthquake magnitude.

parameter	4.0	4.25	4.5	4.75	5.0	5.25	5.5	5.75	6.0	6.25	6.5	6.75	7.0
opt. SCARDEC	-0.03	-0.08	-0.15	-0.25	-0.39	-0.60	-0.89	-1.42	-2.08	-2.80	-3.45	-3.80	-4.15
ave. SCARDEC	-0.23	-0.36	-0.53	-0.71	-0.88	-1.04	-1.23	-1.63	-2.31	-2.98	-3.63	-3.63	-3.81
Sigloch	-0.00	-0.00	0.01	0.04	0.10	0.23	0.37	0.59	0.85	0.21	-0.12	-0.32	-0.18
Ye	0.00	0.00	-0.01	-0.04	-0.06	-0.10	-0.14	-0.33	-0.42	-0.70	-0.93	-1.17	-1.70
USGS	0.19	0.20	0.21	0.24	0.27	0.28	0.24	0.14	0.07	-0.29	-0.58	-1.44	-2.57
ISC	-0.08	-0.15	-0.25	-0.44	-0.81	-1.38	-2.21	-4.02	-4.36	-3.90	-5.42	-5.00	0.52

Table B.2: Y intercept of the relationship between the time to release moment equal to some magnitude of earthquake and the final earthquake magnitude.

parameter	4.0	4.25	4.5	4.75	5.0	5.25	5.5	5.75	6.0	6.25	6.5	6.75	7.0
opt. SCARDEC	0.28	0.62	1.12	1.86	2.94	4.57	6.88	11.00	16.36	22.56	28.88	33.78	39.43
ave. SCARDEC	1.69	2.67	3.93	5.34	6.74	8.22	9.91	13.27	18.82	24.81	31.18	33.57	38.07
Sigloch	0.10	0.12	0.04	-0.09	-0.37	-1.00	-1.58	-2.60	-3.86	-0.29	1.56	2.70	1.67
Ye	0.46	0.46	0.59	0.80	0.99	1.35	1.78	3.55	4.81	7.82	11.02	14.96	22.16
USGS	-0.93	-0.92	-0.83	-0.86	-0.90	-0.70	-0.09	1.02	2.12	5.56	8.89	17.25	28.51
ISC	0.64	1.10	1.78	3.06	5.53	9.38	15.31	27.49	31.13	29.44	40.45	38.50	3.65

Table B.3: Spearman r-value of the relationship between the time to release moment equal to some magnitude of earthquake and the final earthquake magnitude.

parameter	4.0	4.25	4.5	4.75	5.0	5.25	5.5	5.75	6.0	6.25	6.5	6.75	7.0
opt. SCARDEC	-0.47	-0.65	-0.70	-0.68	-0.63	-0.59	-0.55	-0.55	-0.56	-0.55	-0.53	-0.47	-0.38
ave. SCARDEC	-0.55	-0.61	-0.64	-0.65	-0.64	-0.60	-0.54	-0.53	-0.56	-0.54	-0.54	-0.45	-0.36
Sigloch	0.02	0.03	0.10	0.12	0.15	0.15	0.14	0.14	0.13	0.02	-0.16	-0.32	-0.30
Ye	0.12	0.12	-0.05	-0.08	-0.09	-0.12	-0.15	-0.28	-0.29	-0.32	-0.35	-0.36	-0.32
USGS	0.10	0.05	0.03	0.01	0.02	0.04	0.04	0.01	-0.04	-0.11	-0.16	-0.21	-0.28
ISC	-0.43	-0.53	-0.57	-0.58	-0.57	-0.57	-0.57	-0.63	-0.64	-0.44	-0.76	-0.87	NaN

Table B.4: Spearman p-value of the relationship between the time to release moment equal to some magnitude of earthquake and the final earthquake magnitude.

parameter	4.0	4.25	4.5	4.75	5.0	5.25	5.5	5.75	6.0	6.25	6.5	6.75	7.0
opt. SCARDEC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ave. SCARDEC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sigloch	0.33	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.48	0.00	0.00	0.00
Ye	0.20	0.20	0.64	0.38	0.33	0.22	0.13	0.00	0.00	0.00	0.00	0.00	0.00
USGS	0.12	0.44	0.66	0.86	0.77	0.52	0.51	0.89	0.57	0.10	0.02	0.00	0.00
ISC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.33	NaN

B.6 Statistical attributes of STFs

B.6.1 Maximum moment rate vs average moment rate

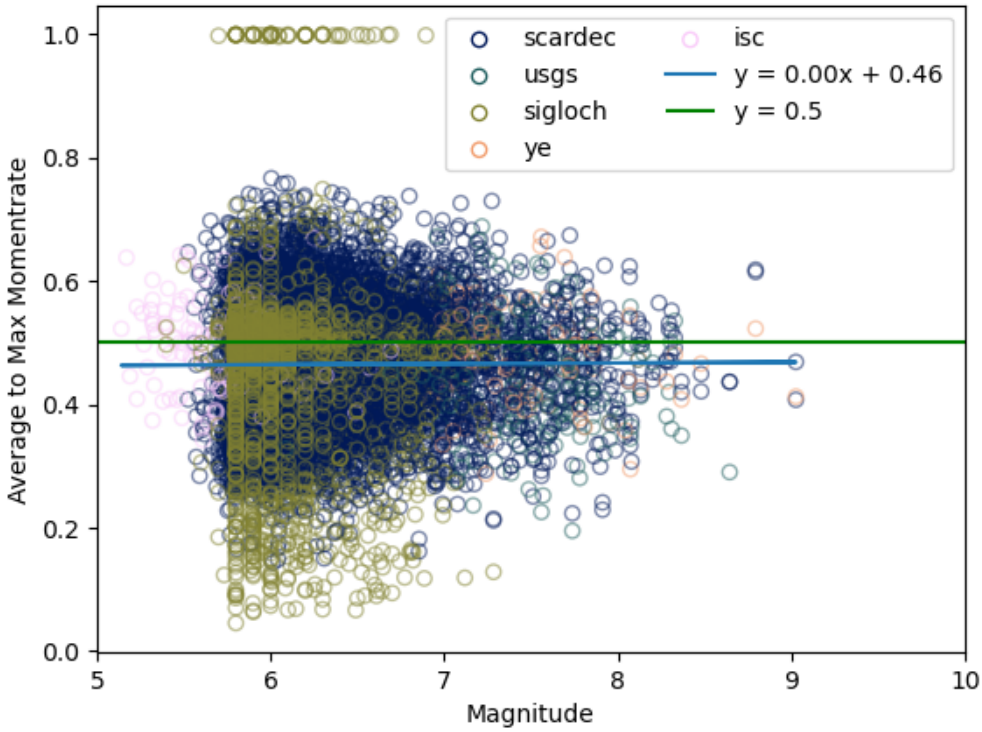


Figure B.30: Average moment-rate of each STF vs maximum moment-rate.

B.6.2 Skew and Kurtosis

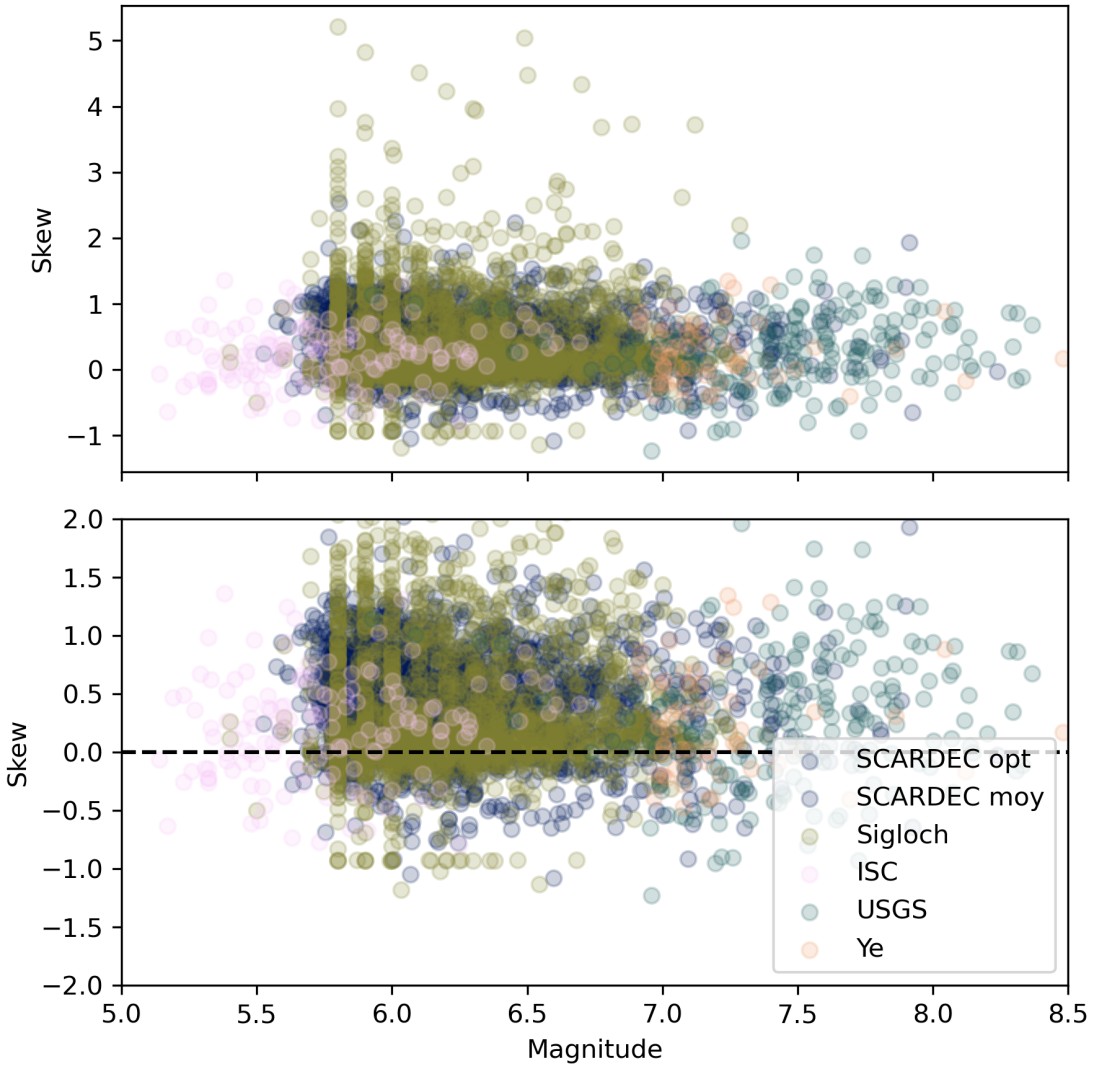


Figure B.31: Skew vs magnitude. Different colours denote different datasets.

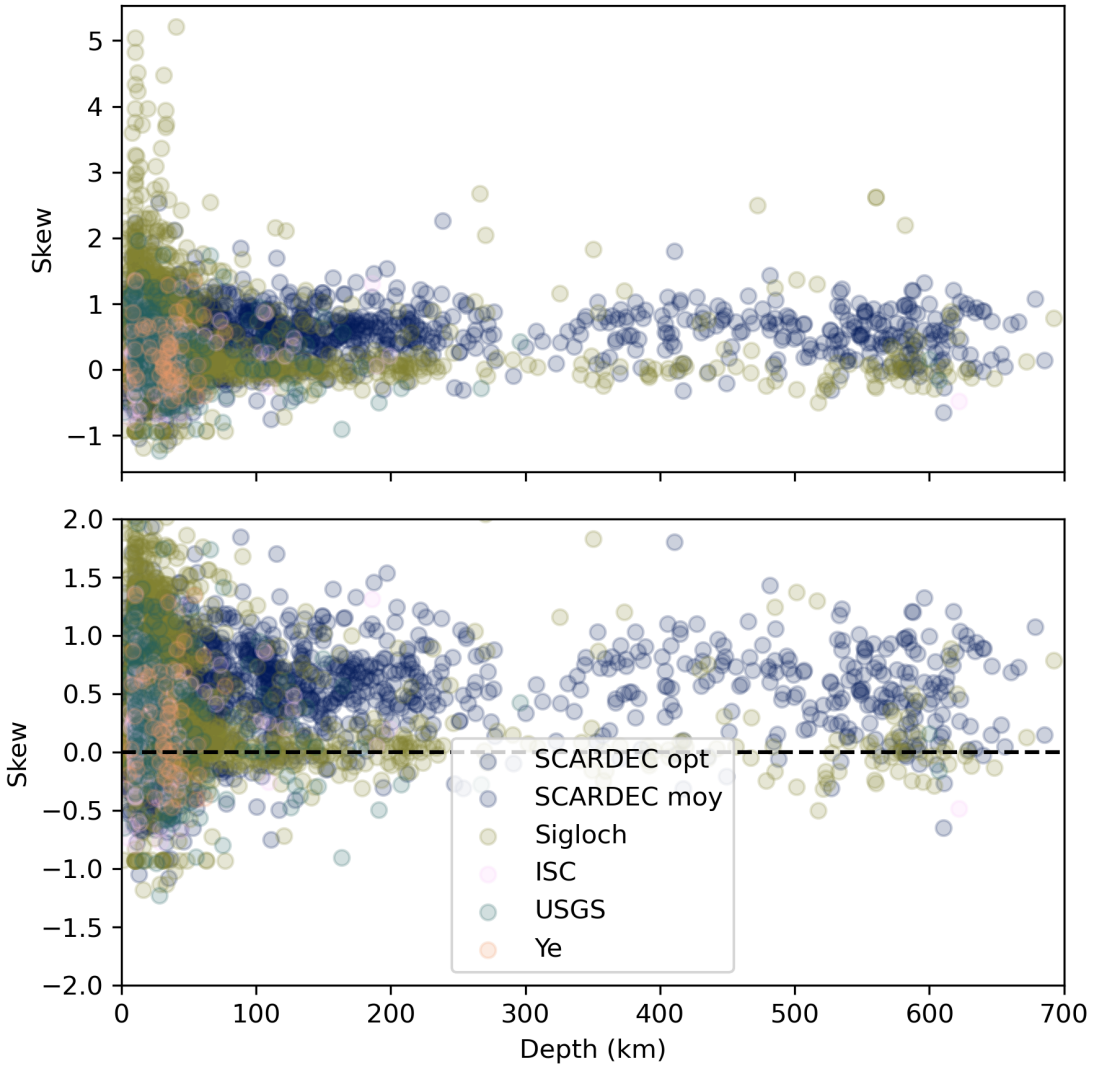


Figure B.32: Skew vs event depth. Different colors denote different datasets.

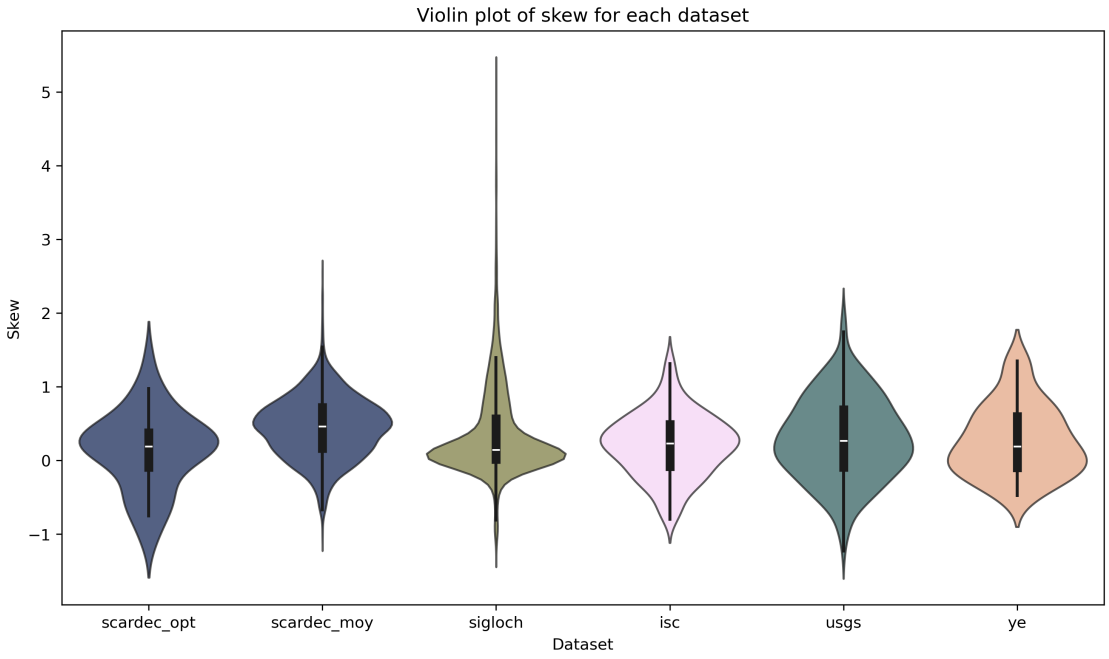


Figure B.33: Violin plot of skew for different datasets.

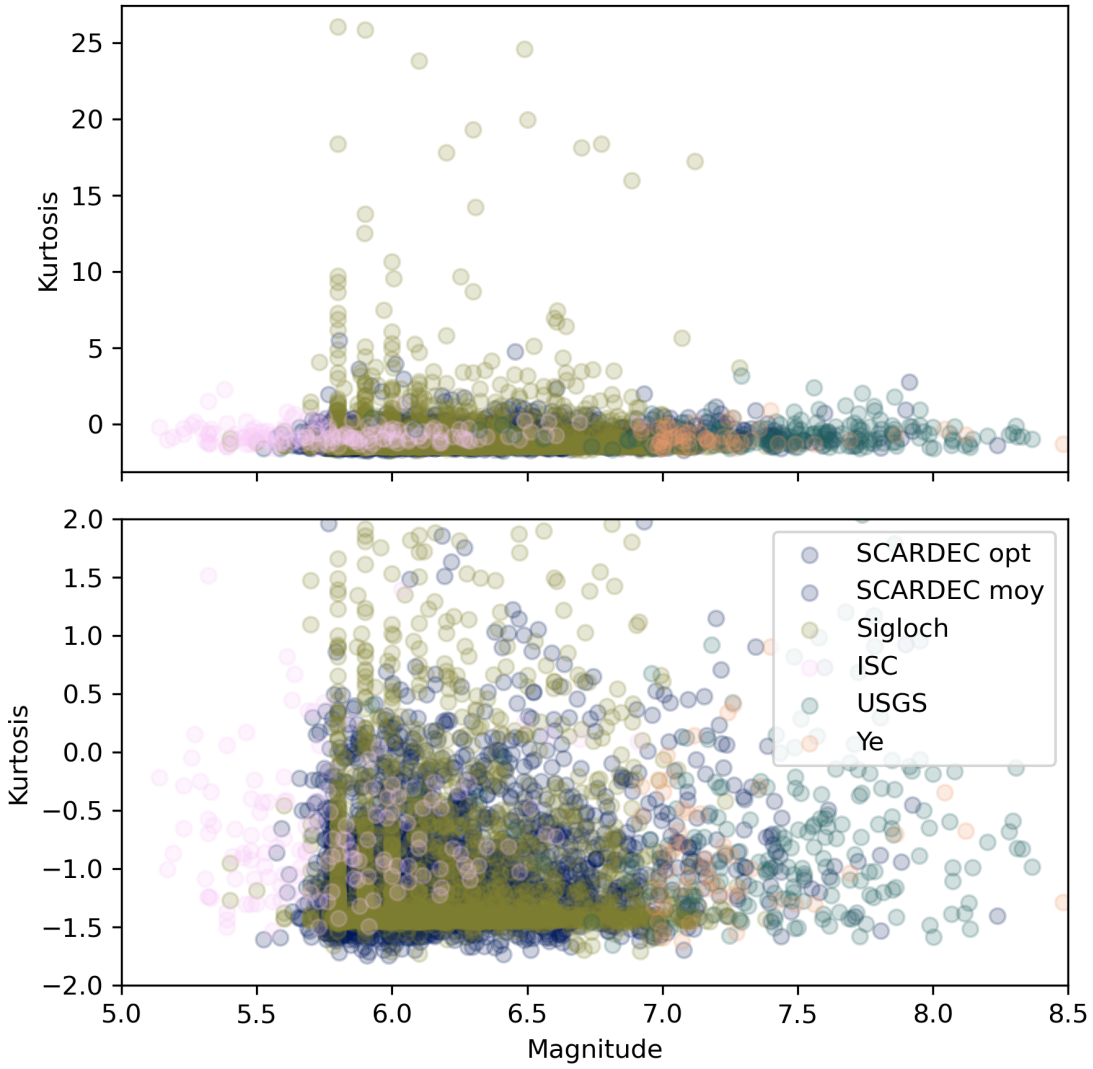


Figure B.34: Kurtosis vs magnitude. Different colors denote different datasets.

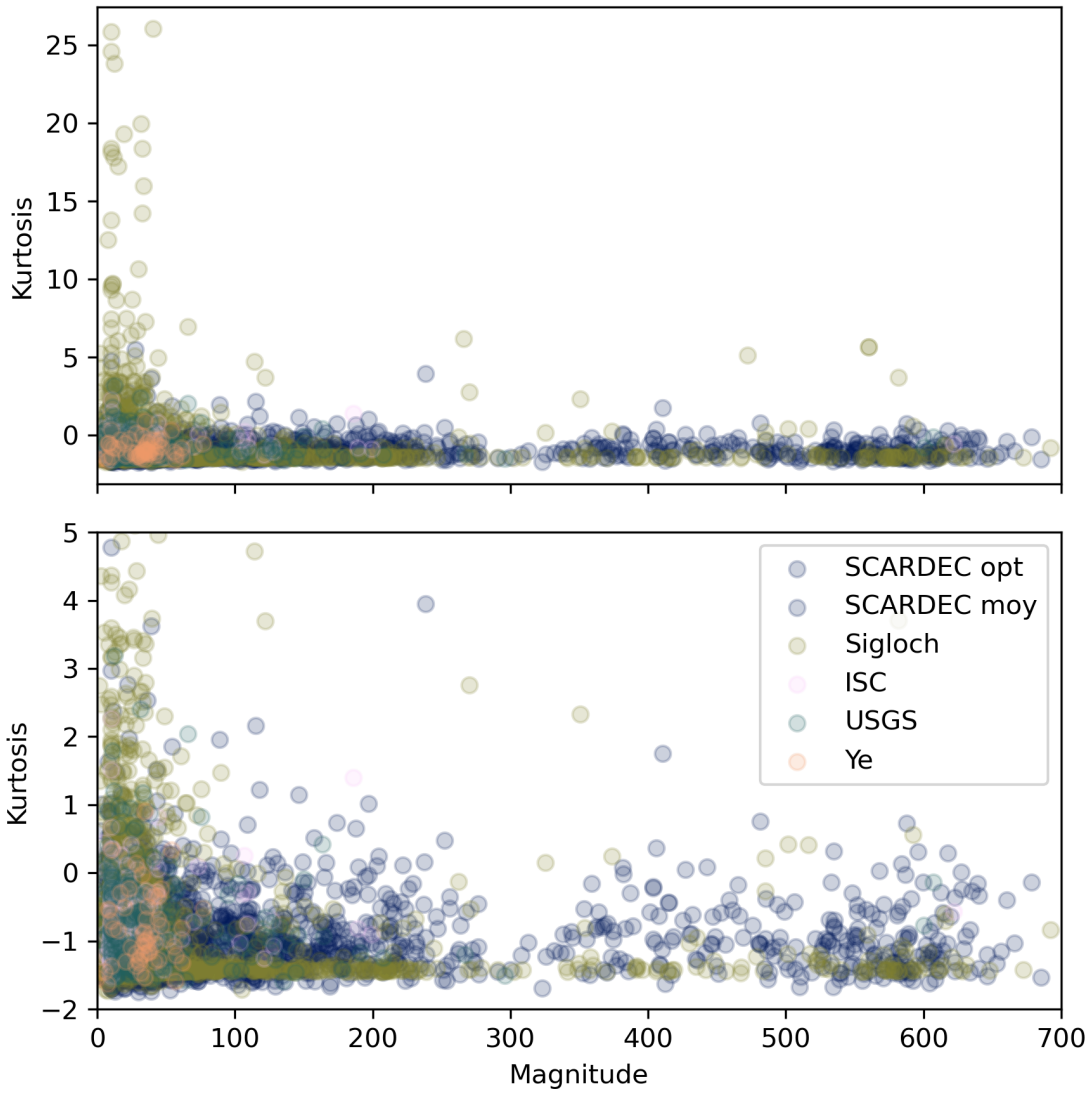


Figure B.35: Kurtosis vs depth. Different colors denote different datasets.

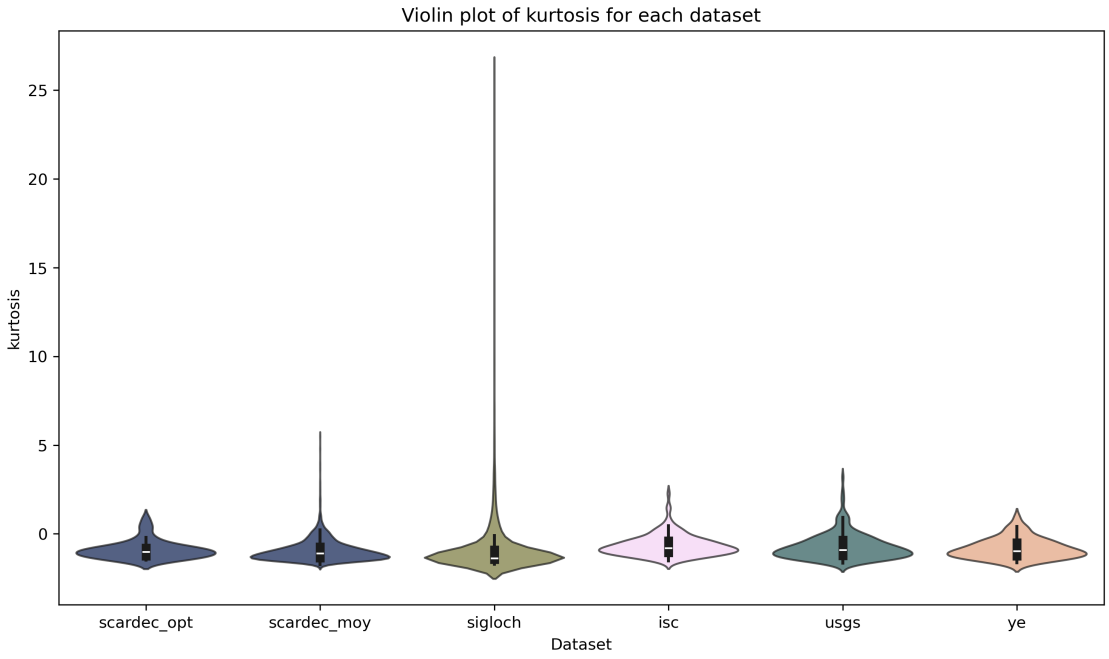


Figure B.36: Violin plot of kurtosis for different datasets.

B.7 Average STFs

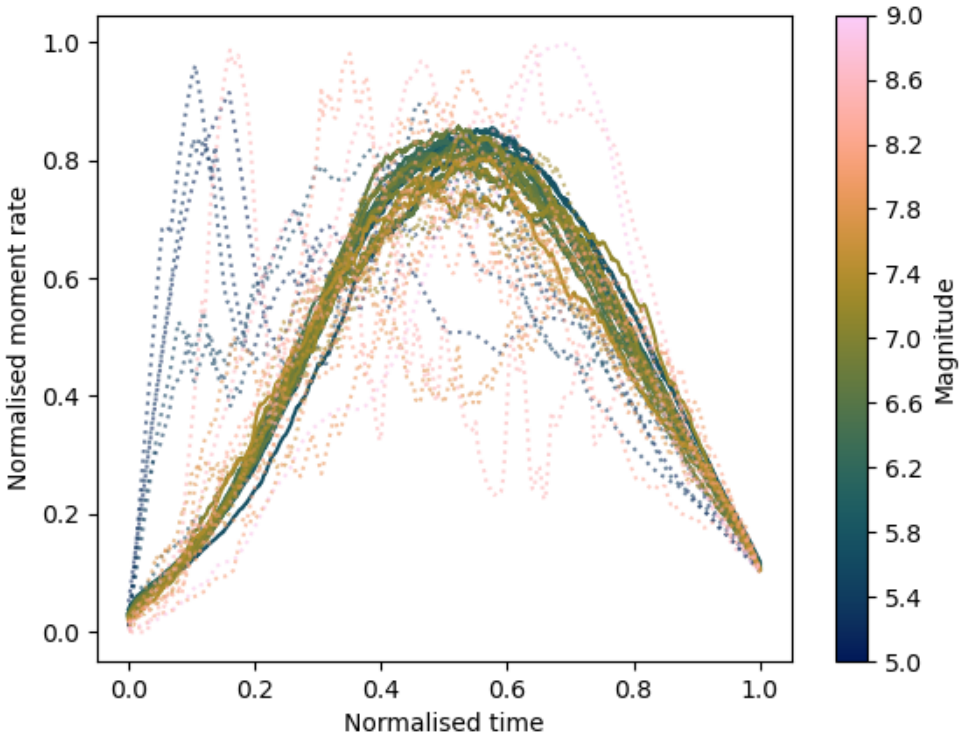
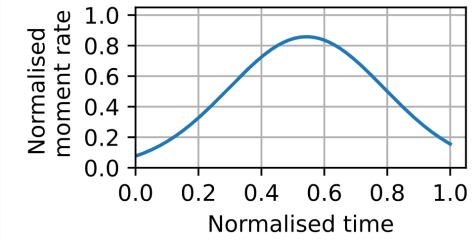
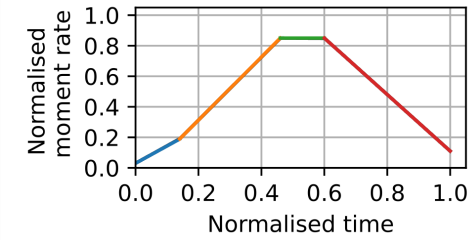


Figure B.37: The median STF shape for $M0.1$ bins. Cool colours are lower magnitudes and warmer, lighter colours are higher magnitudes. Dashed lines show bins with fewer than 100 earthquakes. Black line is the median STF based on the complete dataset.

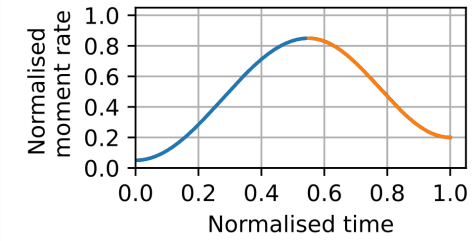
B.8 Trapezium fits with magnitude



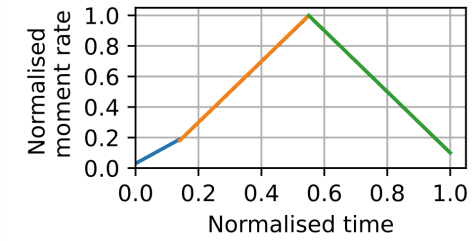
Gaussian
2-norm = 1.96
μ = 0.54
σ = 0.247
Amplitude = 0.85



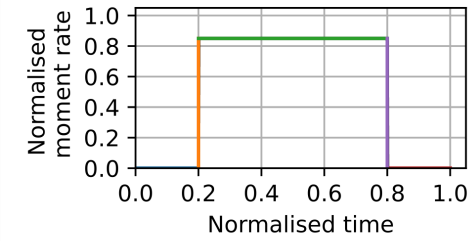
Ramp + Trapezium
2-norm = 3.21
$x < 0.14$: $y = 1.13x + 0.03$
$0.14 \leq x < 0.46$: $y = 2.065x - 0.1$
$0.46 \leq x < 0.6$: $y = 0.85$
$0.6 \leq x$: $y = 1.875x + 1.975$



Sine boxcar
2-norm = 7.38
$x < 0.55$: $f(\theta) = 0.4 \sin\theta + 0.45$
$x \geq 0.55$: $f(\theta) = 0.325 \sin\theta + 0.525$
$\theta \in \left[\frac{-\pi}{2}, \frac{\pi}{2}\right]$



Ramp + Triangle
2-norm = 17.1
$x < 0.14$: $y = 1.13x + 0.03$
$0.14 \leq x < 0.55$: $y = 2x - 0.1$
$0.55 \geq x$: $y = -2x + 2.1$



Boxcar
2-norm = 584.8
$x < 0.2$: $y = 0$
$0.2 < x < 0.8$: $y = 0.85$
$0.8 > x$: $y = 0$

Figure B.38: The optimized models and the parameters which describe them.

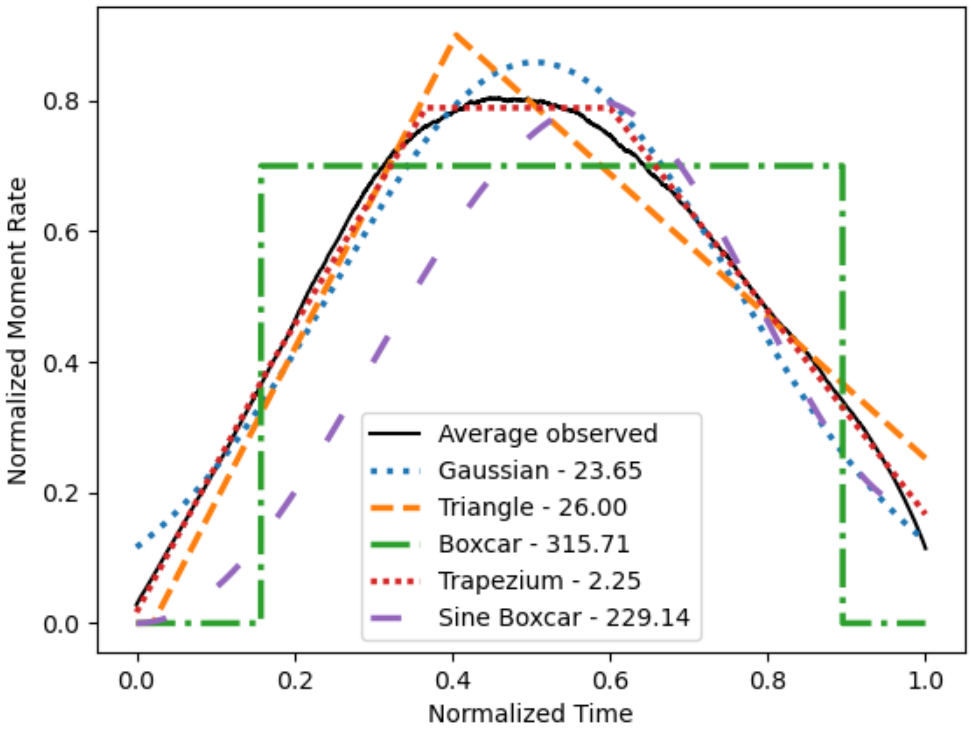


Figure B.39: The median STF for the SCARDEC_opt data, with the best fit geometric models. Numbers in the legend are the 2-norm.

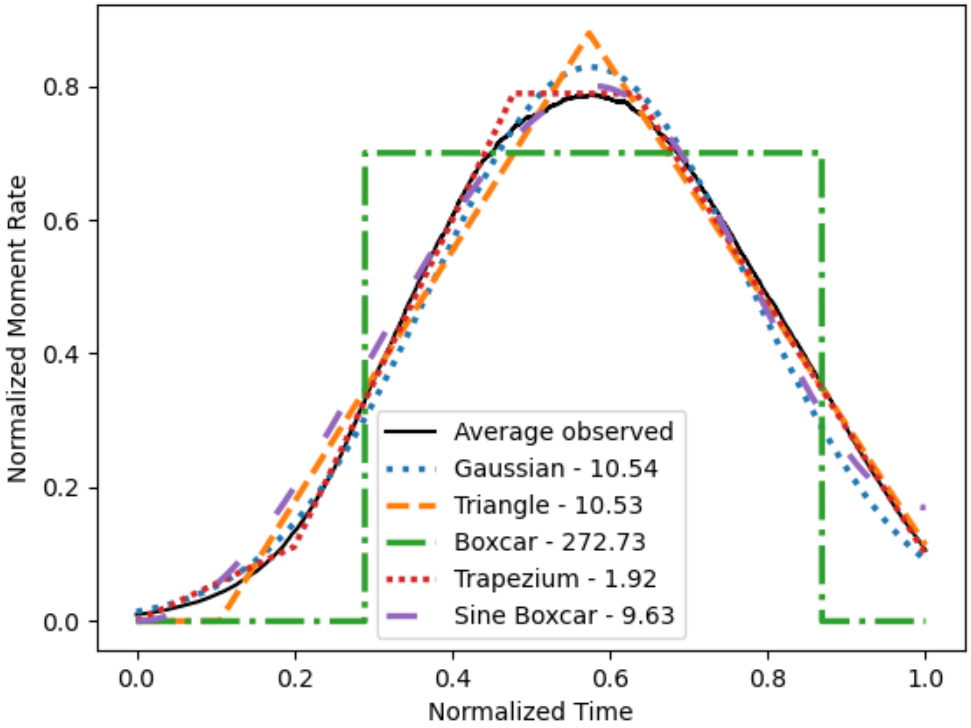


Figure B.40: As Figure B.39 but for the SCARDEC_moy data.

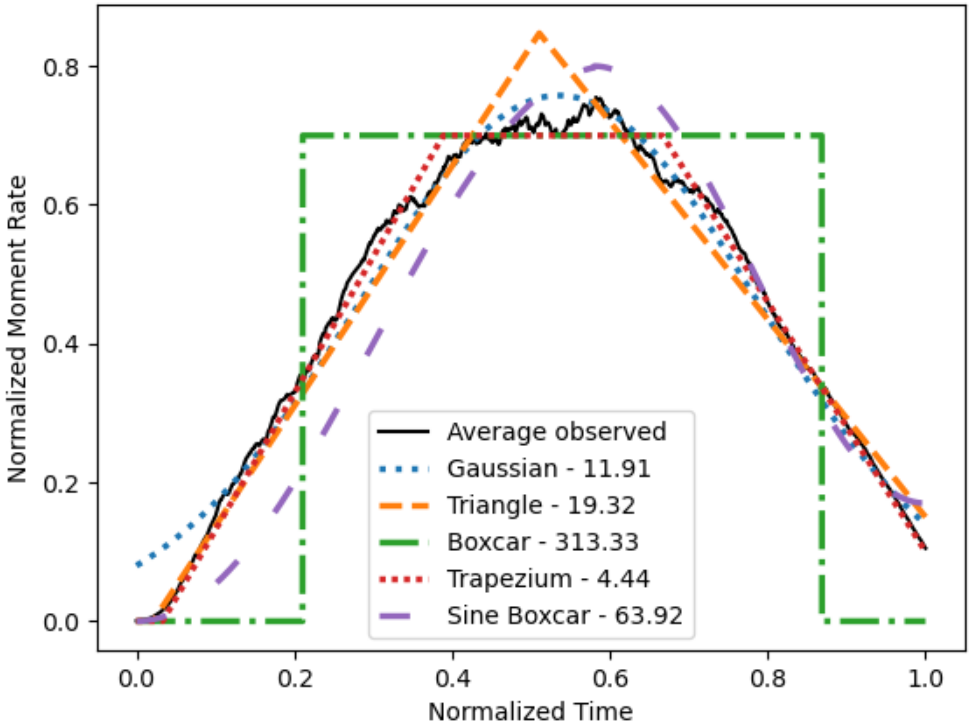


Figure B.41: As Figure B.39 but for the USGS data.

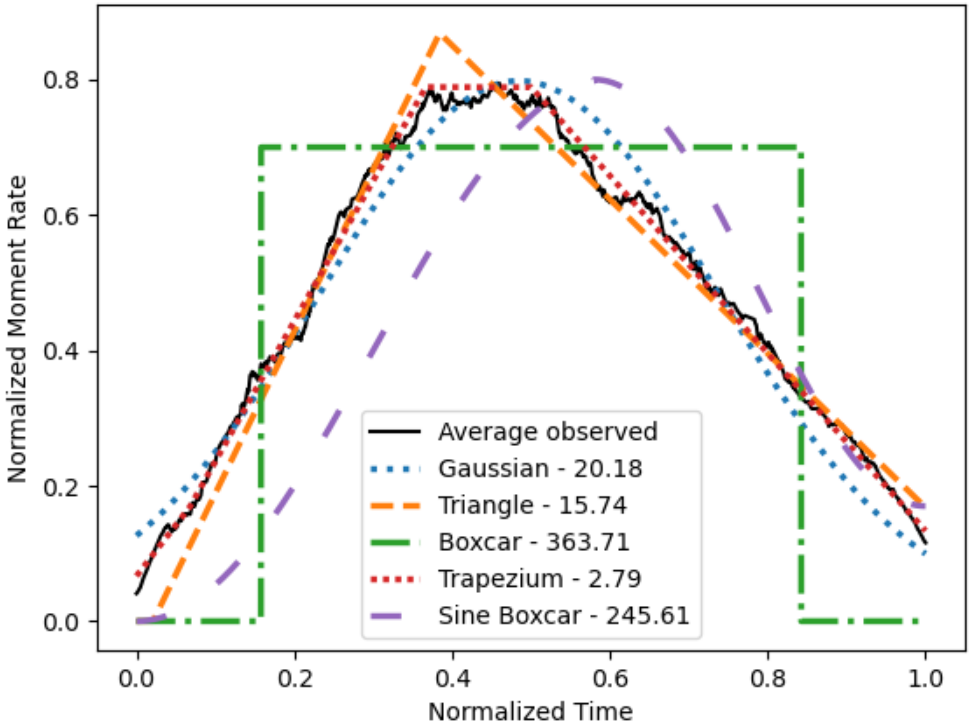


Figure B.42: As Figure B.39 but for the Ye data.

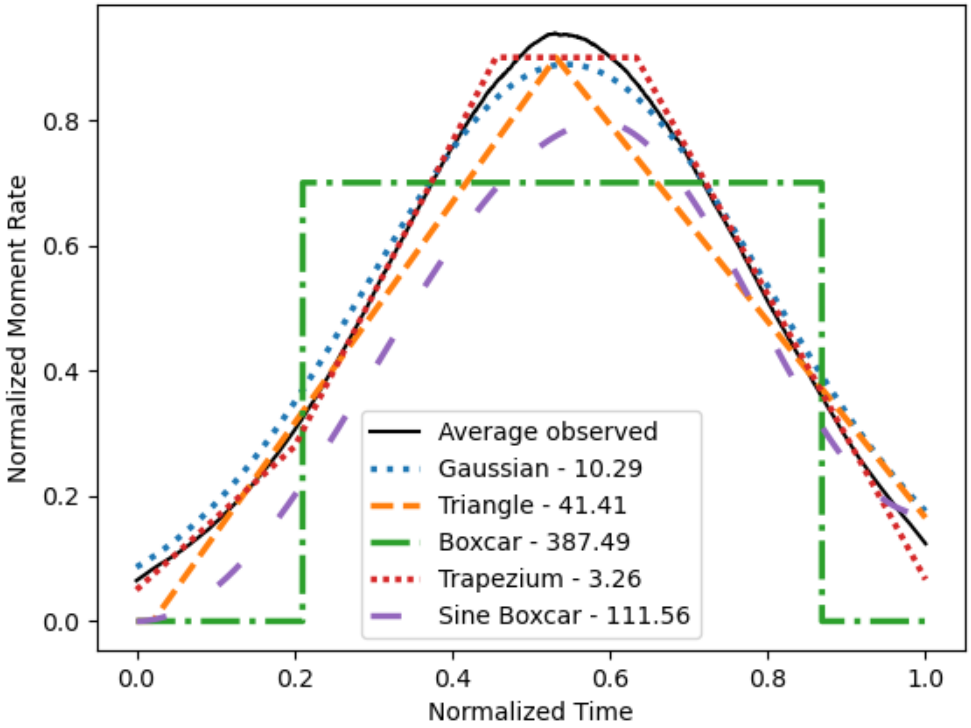


Figure B.43: As Figure B.39 but for the Sigloch data.

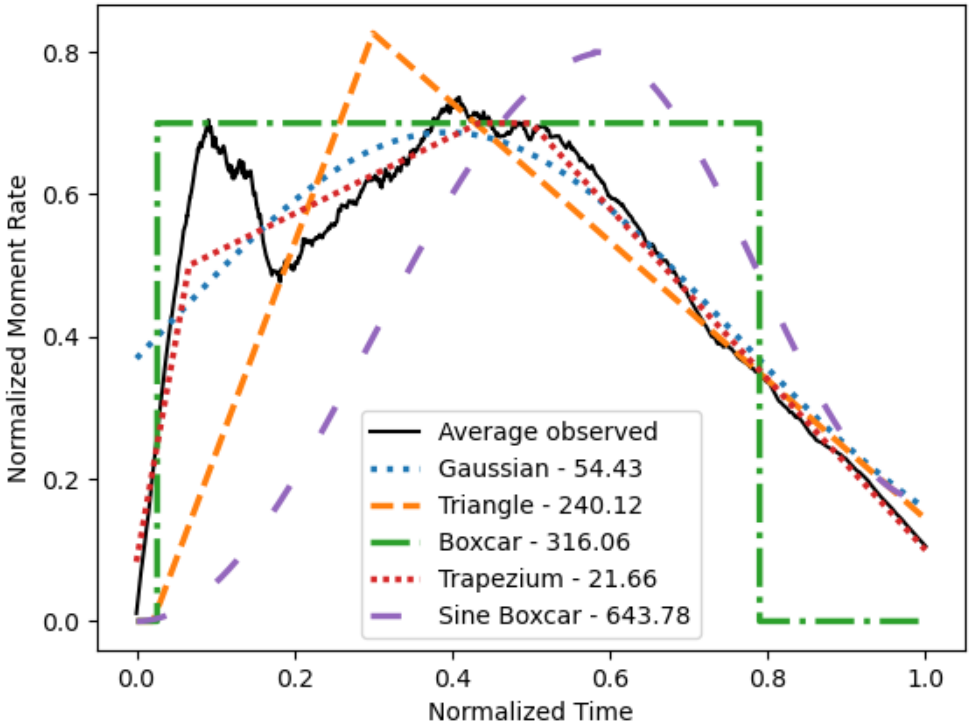


Figure B.44: As Figure B.39 but for the ISC data.

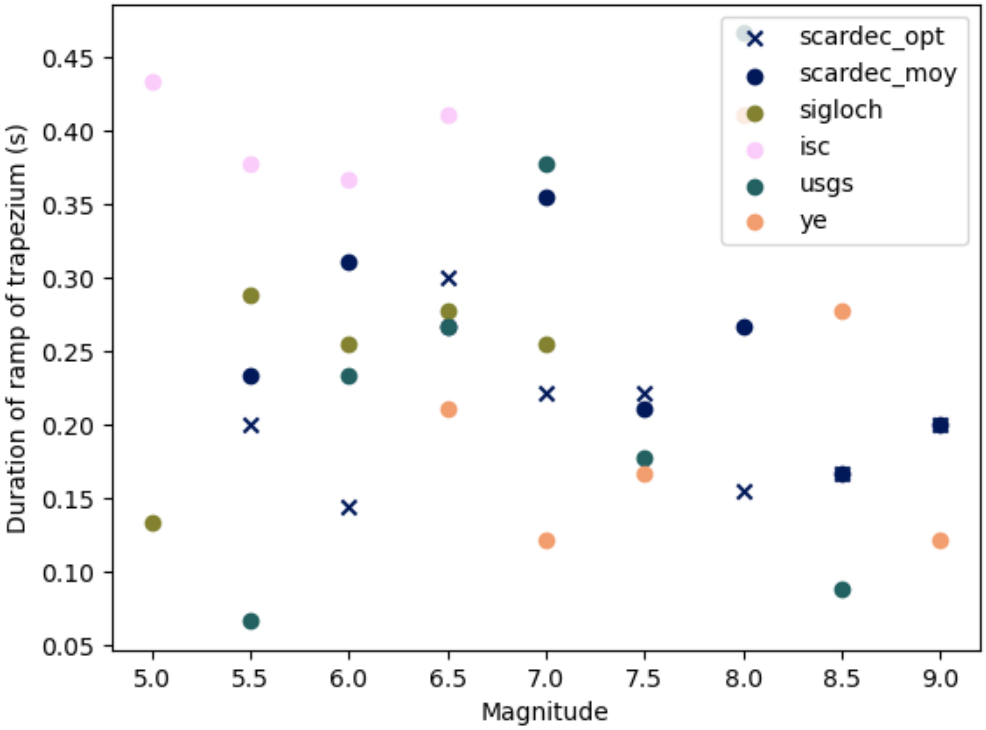


Figure B.45: Duration of the ramp before the best fit trapezium vs magnitude for different datasets.

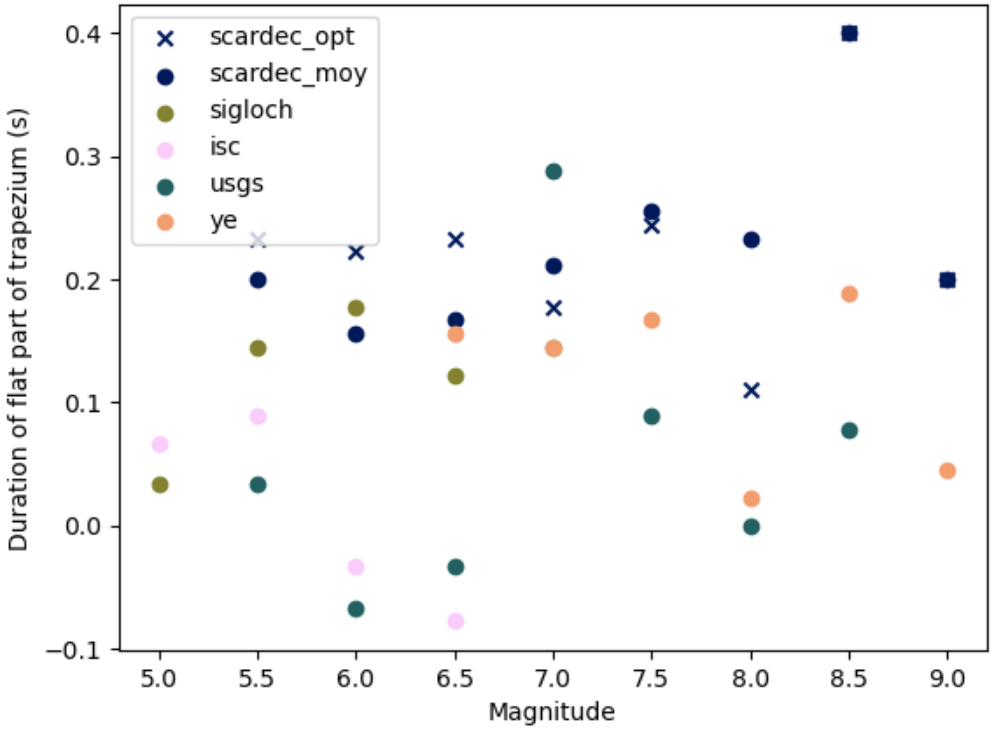


Figure B.46: Duration of the top of the best fit trapezium vs magnitude for different datasets.

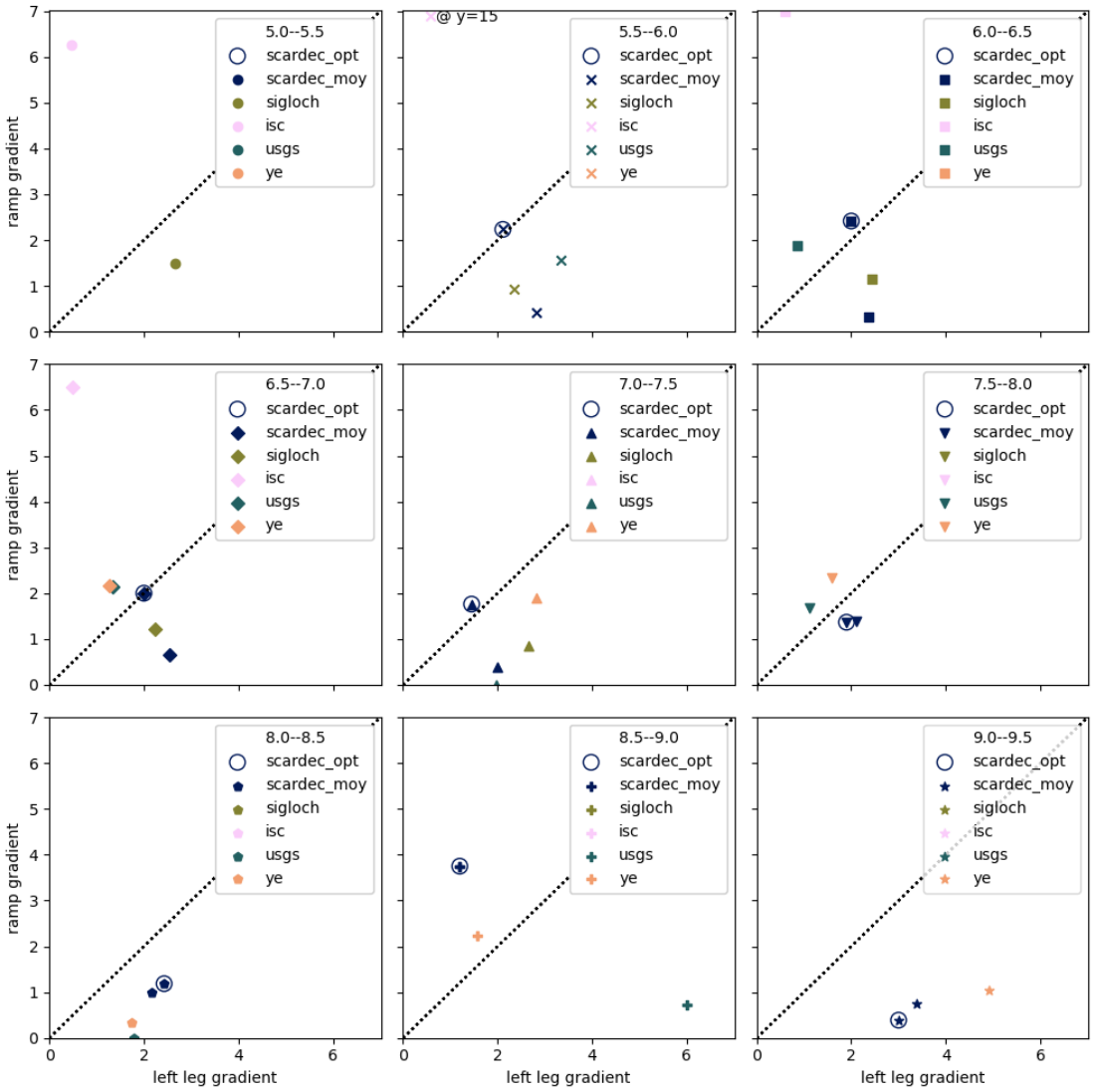


Figure B.47: Gradient of the ramp of the best fit trapezium vs gradient of the left leg of the trapezium for different datasets.

C

Supplementary Information for Chapter 5

Contents

C.1 Mainshock distribution

In Figure 5.1 we show the spatial distribution of mainshocks prior to declustering, coloured by magnitude. We select all $M \geq 3$ within the box defined by 175° and 180°W and 37° and 40°S. In Figure C.1 we colour these events by depth which shows that our region samples events at a range of depths of the subduction interface, and that there are very few non-subduction related earthquakes included in the dataset (e.g. shallow crustal earthquakes in the Taupo Volcanic Field).

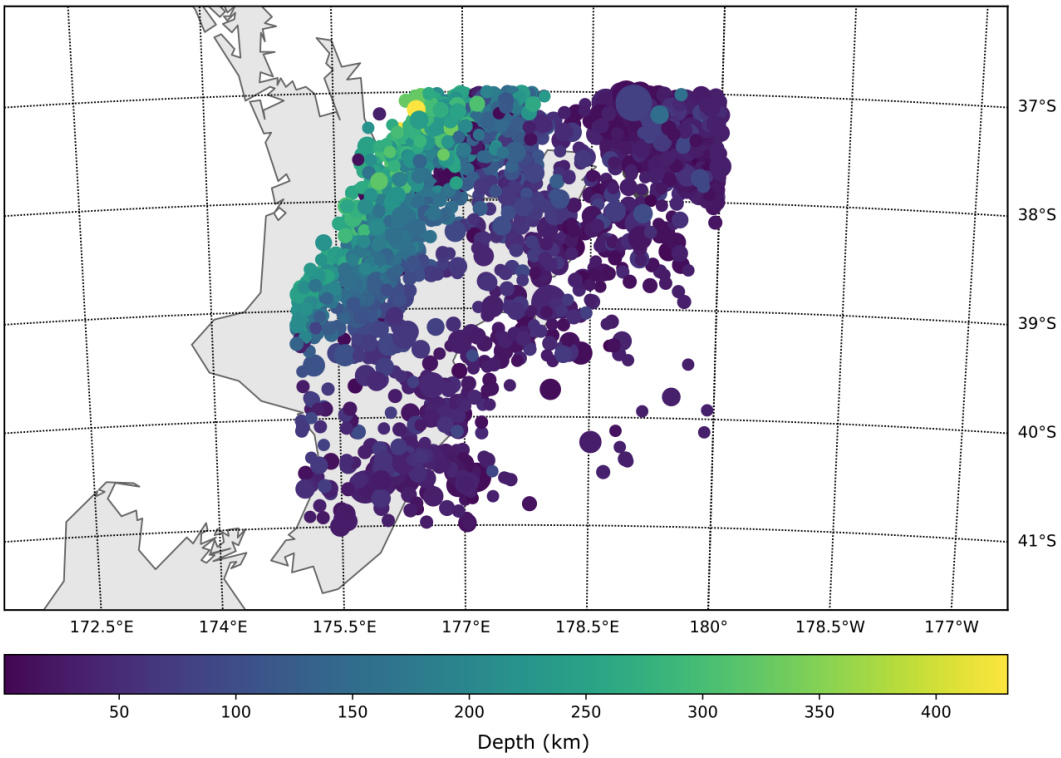


Figure C.1: Complete mainshock catalogue before declustering, coloured by depth. Size of marker indicates magnitude, but this is more clearly shown in Figure 5.1.

C.2 Magnitude of Completeness

As mentioned in Section 5.3.4, in our calculations of magnitude of completeness, we considered a perfectly defined magnitude of completeness M_c, where we detect all earthquakes above M_c, and no earthquakes below M_c (Figure C.2a). In reality this is an over-simplification. A more accurate representation would be to consider a cosine taper between the two straight line segments (Figure C.2b) We can write the observed number of earthquakes of some magnitude as the product of the true number of earthquakes of that magnitude and the probability of observing events of that magnitude.

$$n^m_{\text{obs}} = n^m_{\text{true}} \cdot P^m. \tag{C.1}$$

The total number of earthquakes observed is then:

$$N_{\text{obs}} = \int_{m=-\infty}^{m=\infty} n^m_{\text{obs}} dm = \int_{m=-\infty}^{m=\infty} n^m_{\text{true}} \cdot P^m dm \tag{C.2}$$

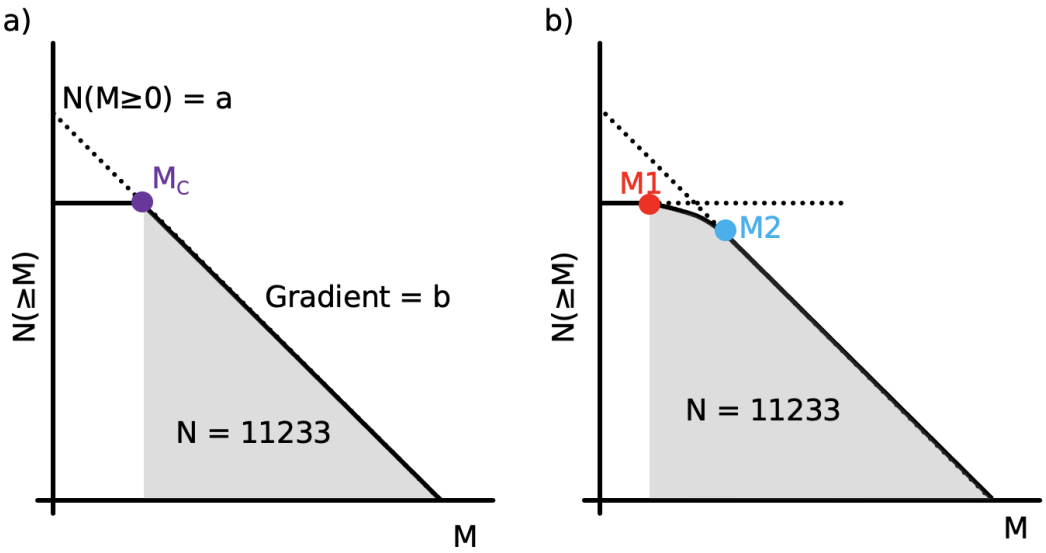


Figure C.2: Figures illustrating the idealised Gutenberg-Richter, and one which has a tapered magnitude of completeness.

where

$$P(m) = \begin{cases} 0 & \text{if } m < m_1, \\ 0.5 \cdot \left(1 - \cos\left(\pi \cdot \frac{m-m_1}{m_2-m_1}\right)\right) & \text{if } m_1 \geq m \geq m_2, \text{ and} \\ 1 & \text{if } m > m_2. \end{cases} \tag{C.3}$$

Noting that

$$N_{\text{obs}} = \int_{m=-\infty}^{m=\infty} n^m_{\text{true}} \cdot P^m dm = \int_{m=m_1}^{m=\infty} n^m_{\text{true}} \cdot P^m dm, \tag{C.4}$$

we find

$$\begin{aligned} N_{\text{obs}} = &\int_{m=m_1}^{m=m_2} 0.5 \cdot \left(1 - \cos\left(\pi \cdot \frac{m-m_1}{m_2-m_1}\right)\right) \cdot \left(-b \cdot \ln 10 \cdot 10^{a-bm}\right) dm \\ &+ \int_{m=m_2}^{m=\infty} \left(-b \cdot \ln 10 \cdot 10^{a-bm}\right) dm, \end{aligned} \tag{C.5}$$

and after some algebra

$$\begin{aligned} N_{\text{obs}} =& 0.5 \cdot \left(10^{a-bm_2} - 10^{a-bm_1}\right) \\ &+ \frac{b^2 \cdot (\ln 10)^2 \cdot (m_1 - m_2)^2 \cdot 10^{a-b(m_1-m_2)} \cdot \left(10^{-bm_1} + 10^{-bm_2}\right)}{b^2 \cdot (\ln 10)^2 \cdot (m_1 - m_2)^2 + \pi^2}. \end{aligned} \tag{C.6}$$

We can now estimate the trade off between M1 and M2 (Figure C.3).

This shows that M1 is somewhere between 0 and 1, and so our estimate in Section 5.3.4 of a magntiude of completeness of M0.6 is sensible, since with this approach it is more difficult to define exactly what the magnitude of completeness is.

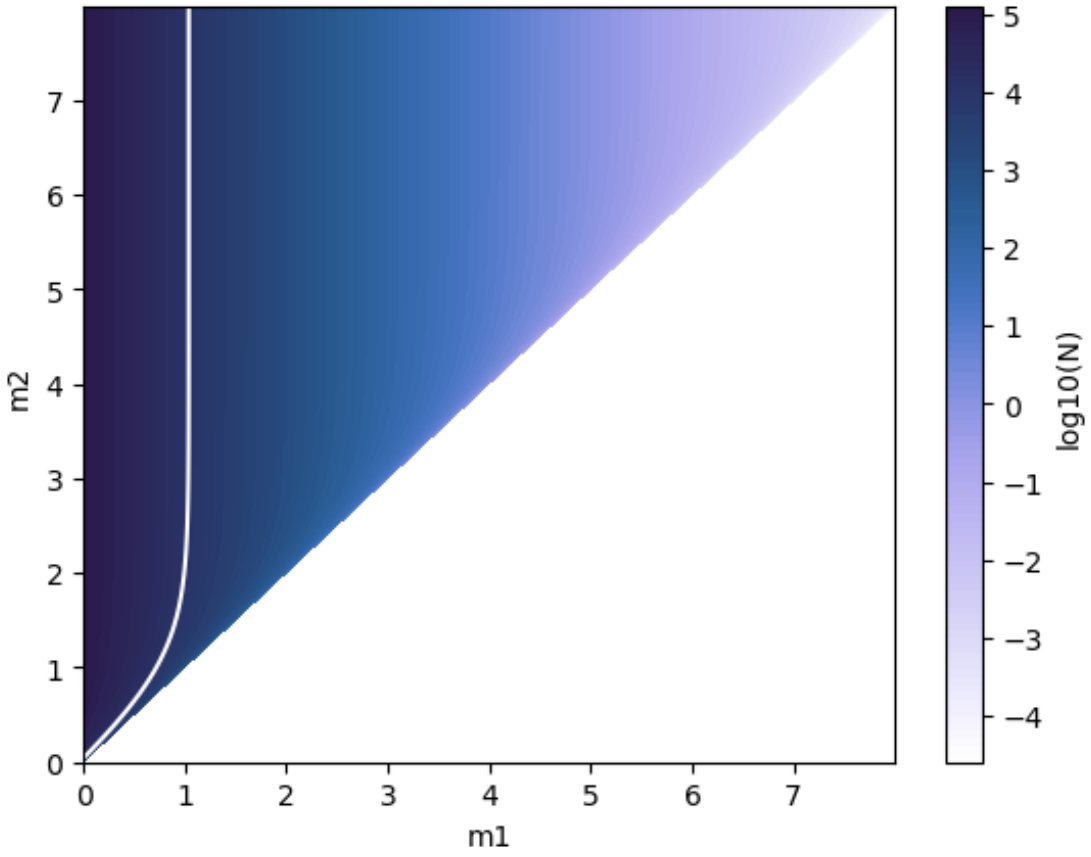


Figure C.3: A contour map of trade off between M1 and M2. Colour indicates $\log N$.

C.3 Uncertainty on Omori exponent

We can use bootstrapping to better understand the uncertainty of the Omori exponent found in Section 5.4.3.

We sample with replacement our detection data 100 times, and fit the omori curve in the same way as before. This gives us 100 lines, which we plot in Figure C.4. We then plot a histogram of the Omori exponent for the foreshocks and aftershocks to compare the distributions to p=1. In Figure C.5 we show that both the foreshocks and aftershocks have a p value significantly different than 1.

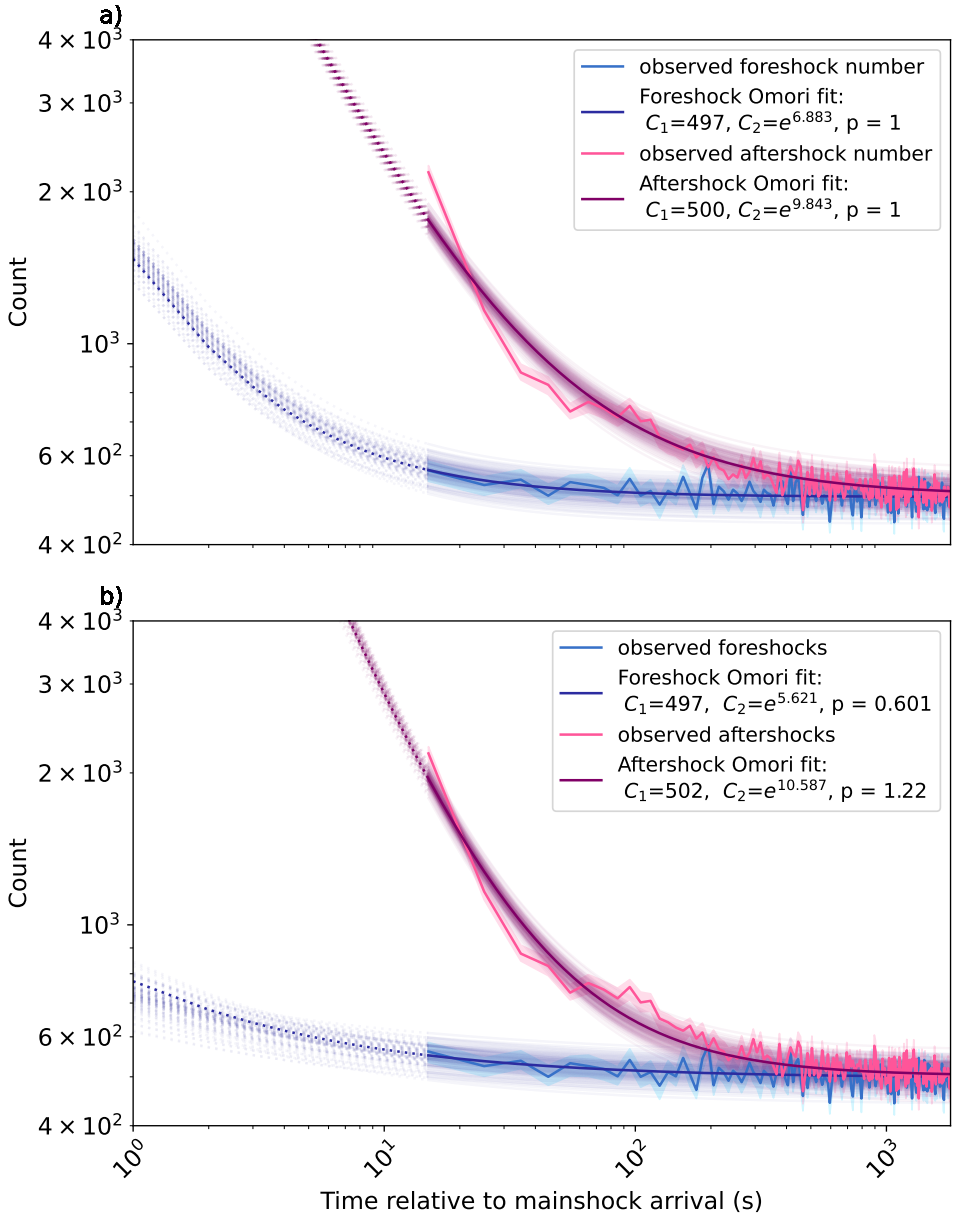


Figure C.4: Log-log plots of foreshock (blue) and aftershock (pink) detections at two standard deviations time relative to mainshock arrival for 100 bootstraps of the detection data. Power law relations, following Omori's law, are plotted: in a) with the p-exponent fixed to 1, and in b) where p is optimised independently for foreshocks and aftershocks. The asymptote for each bootstrapped line is fixed at the background detection value over the first 500 s. The point at 5 s (corresponding to 0-10 s for aftershocks and -10 –0 s for the foreshocks) is not plotted, as we remove all detections within ±2s of the mainshock due the peak spreading.

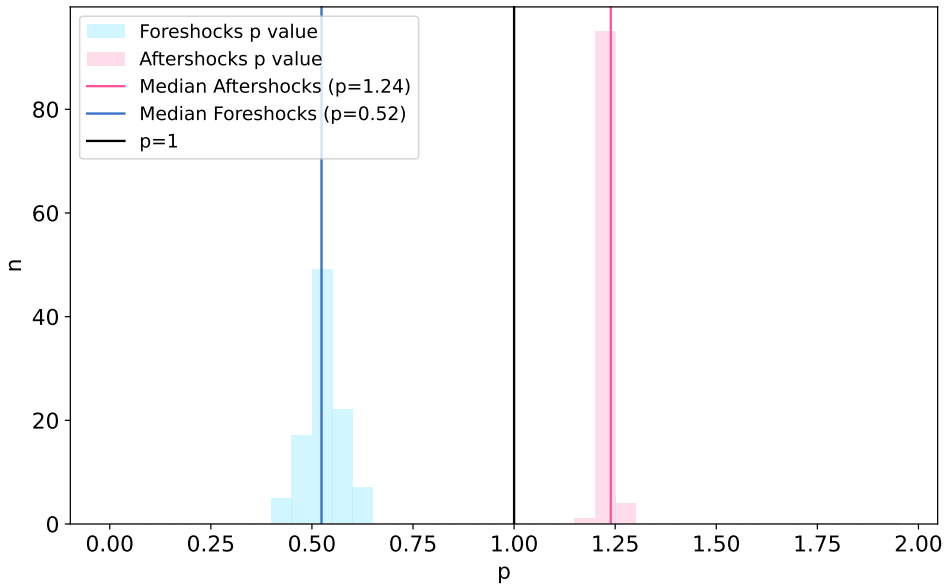


Figure C.5: Histogram of the bootstrapped p value for foreshocks (blue) and aftershocks (pink). Vertical coloured lines indicate the median values, and the black vertical line shows p=1.

C.4 ETAS Models

The time series of our dataset includes various periods of temporal incompleteness between 2005–2020. These are shown by the gaps in recordings of small magnitude earthquakes in the magnitude-time series plot in Figure C.6.

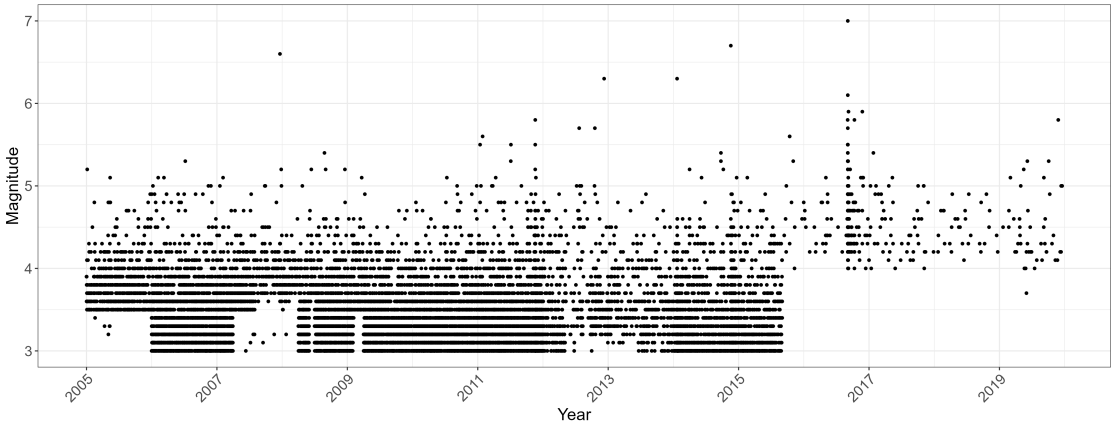


Figure C.6: Magnitude time series of the earthquakes between 2005-2020 in the Hikurangi subduction zone.

C.4.1 Observed ratio using different background windows and declustering windows

Table C.1: Foreshock:aftershock ratios for $M \geq 3$ events using different background windows for calculating background seismicity and declustering events at different windows.

Background Window (s)	Declustering window (hours)							
	4	8	12	24	36	48	72	100
100	0.328	0.218	0.165	0.104	0.127	0.112	-0.032	0.072
200	0.332	0.250	0.175	0.120	0.061	0.069	-0.027	0.050
400	0.284	0.207	0.161	0.113	0.072	0.087	0.059	0.102
600	0.278	0.202	0.156	0.097	0.078	0.075	0.038	0.041
800	0.265	0.198	0.152	0.103	0.095	0.089	0.058	0.058
1000	0.241	0.176	0.132	0.070	0.082	0.070	0.051	0.047

C.4.2 Comparisons between ETAS and observed ratios

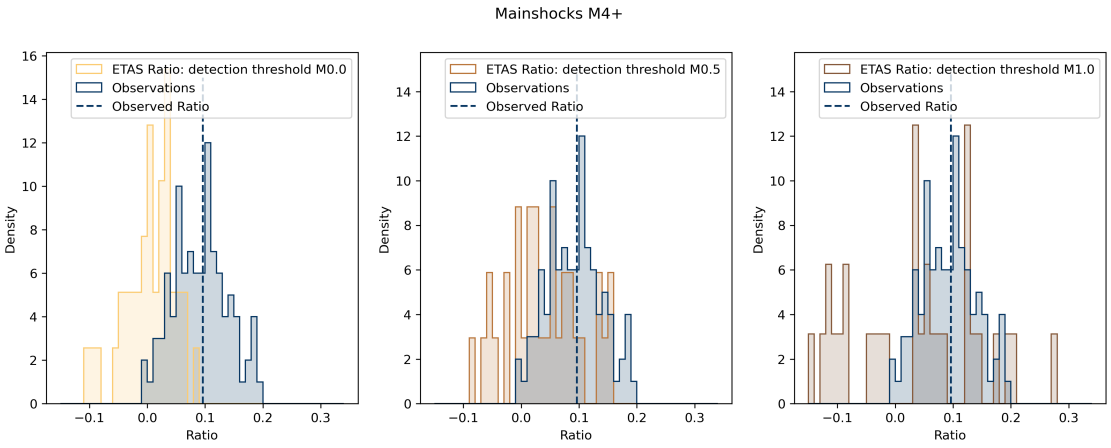


Figure C.7: As Figure 5.7 but for M4+ mainshocks.

C.4.3 Temporal patterns in ETAS foreshocks and aftershocks

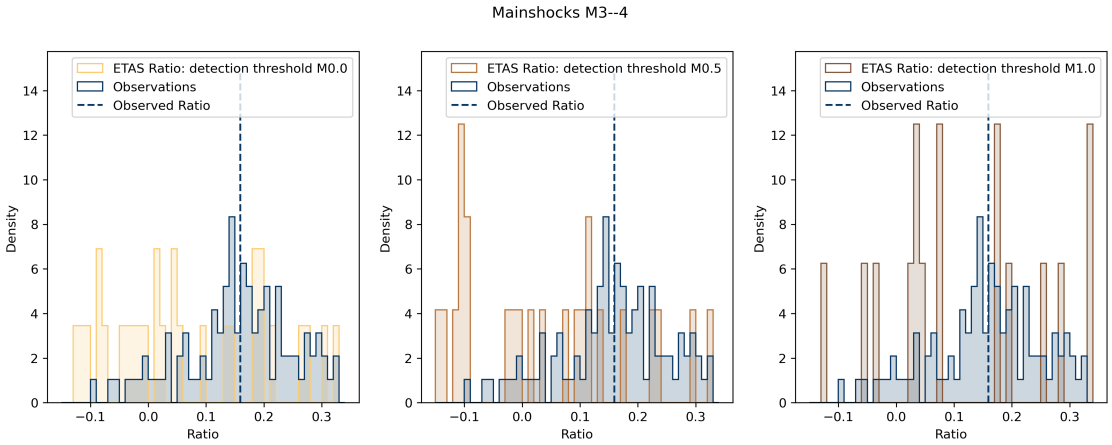


Figure C.8: As Figure 5.7 but for M3–4 mainshocks.

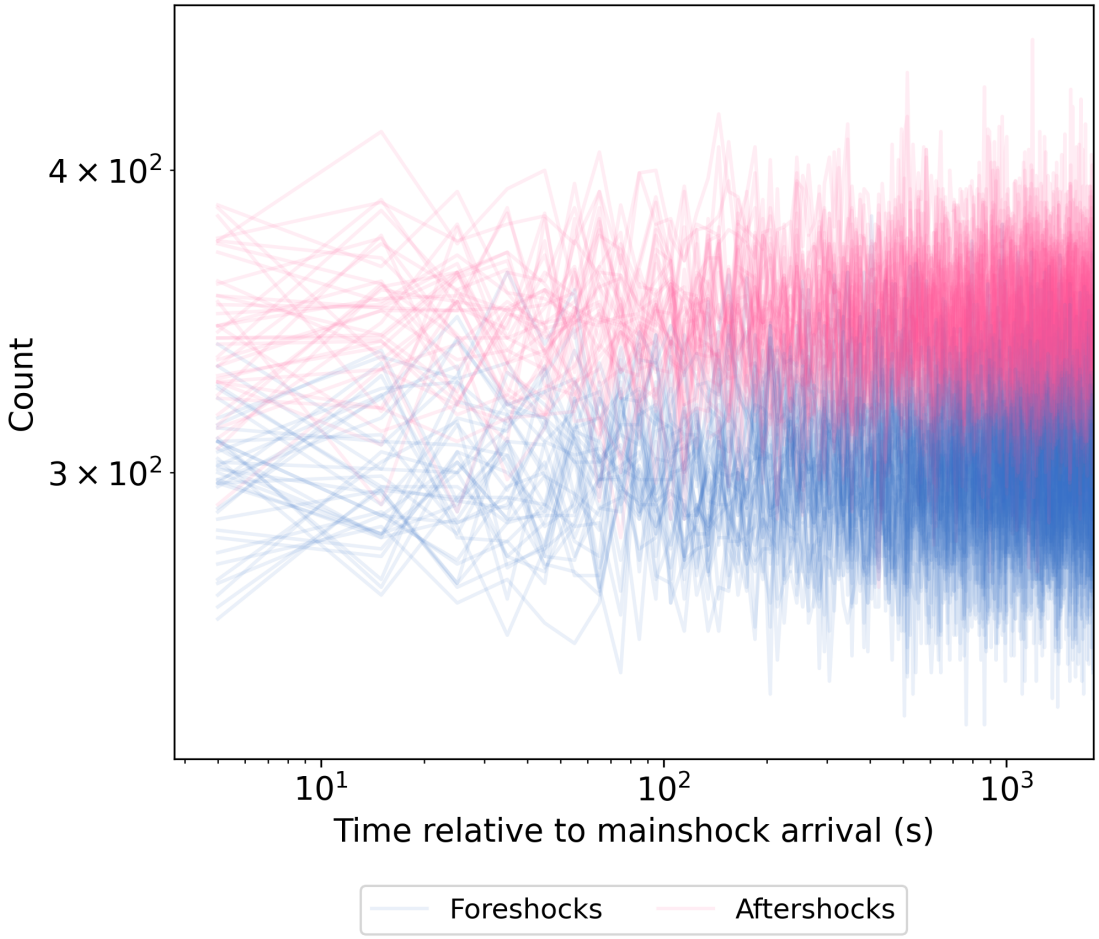


Figure C.9: Number of foreshocks and aftershocks from each run of the ETAS simulation.

C.4.4 Comparison between ETAS and observed numbers of foreshocks and aftershocks per mainshock

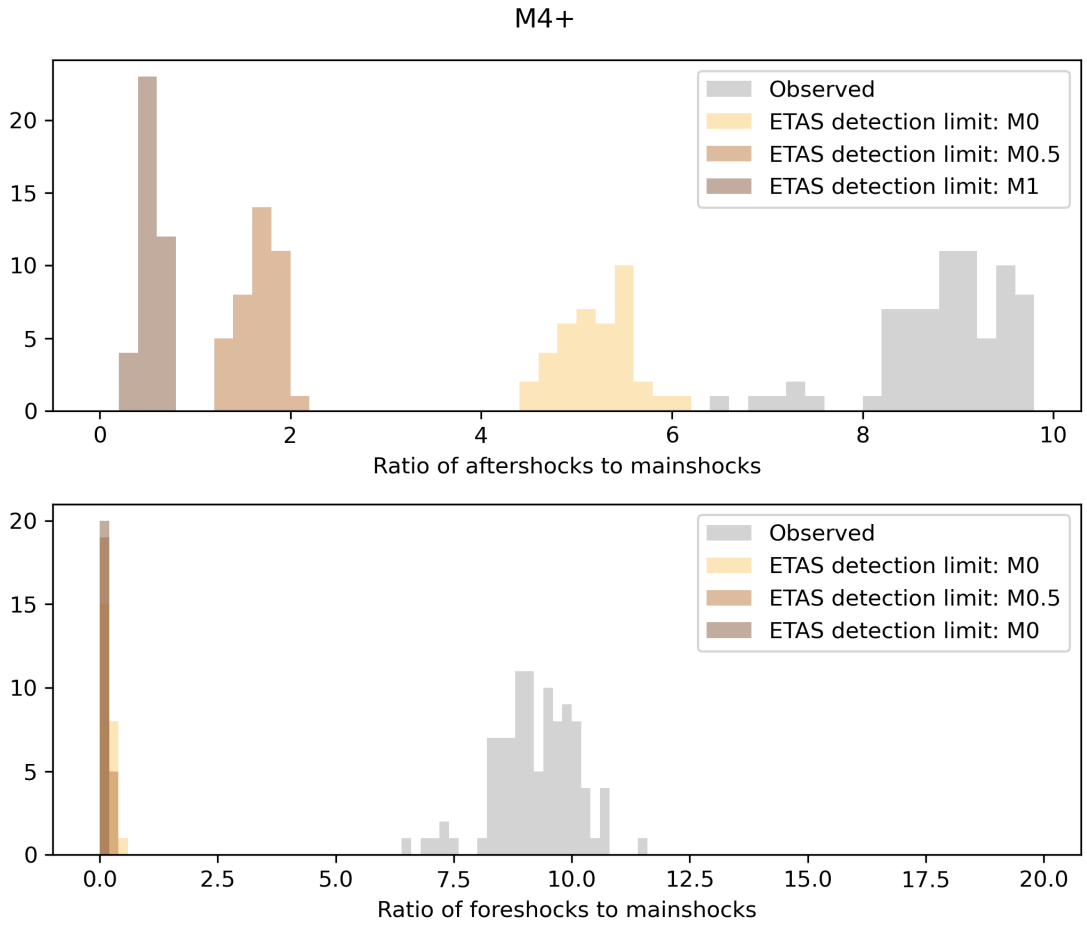


Figure C.10: As Figure 5.11 but for M4+ mainshocks.

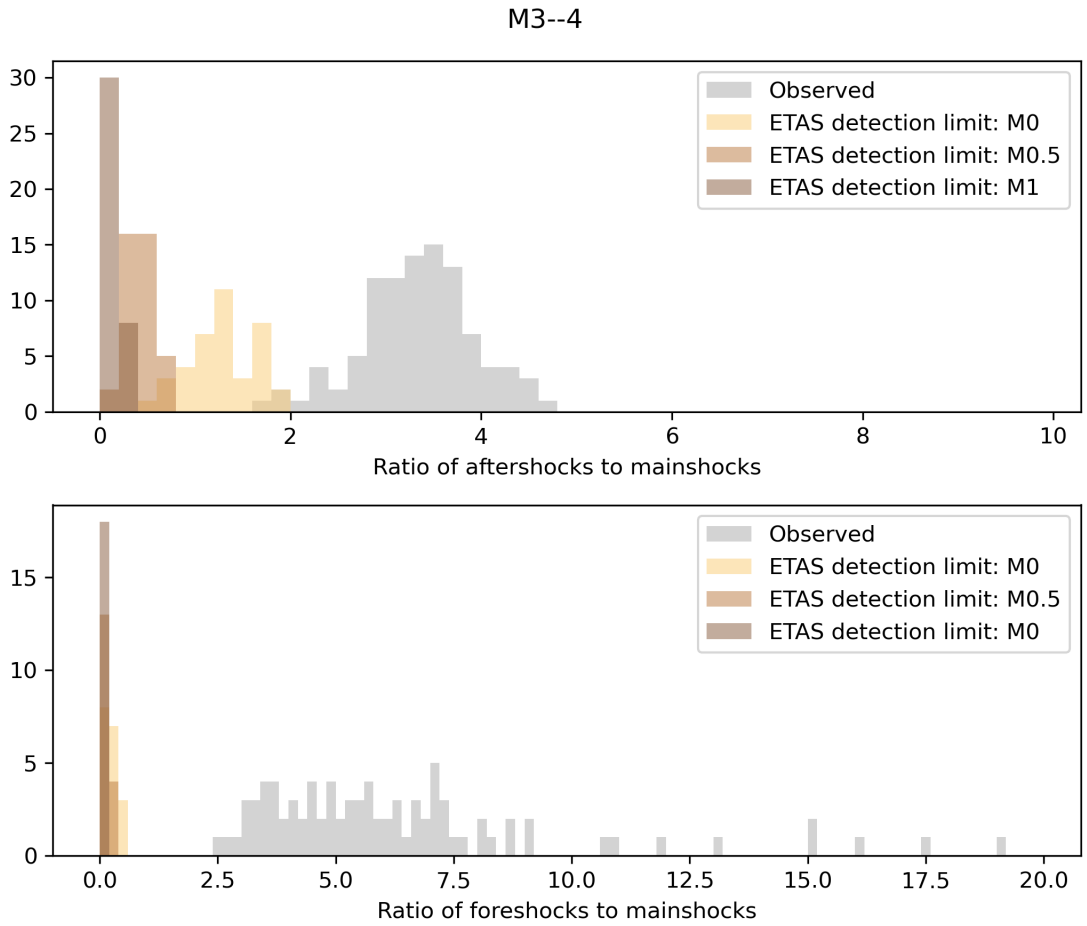


Figure C.11: As Figure 5.11 but for M3–4 mainshocks.

C.5 Phase coherence amplitudes vs magnitude

We find a very weak relationship between the amplitude of phase coherence detections and the magnitude of their associated events from the GNS catalogue (Figure C.12). That this relationship is weak is not unexpected, as many other factors influence the amplitude of the phase coherence result, including source-event distance and signal-to-noise ratio.

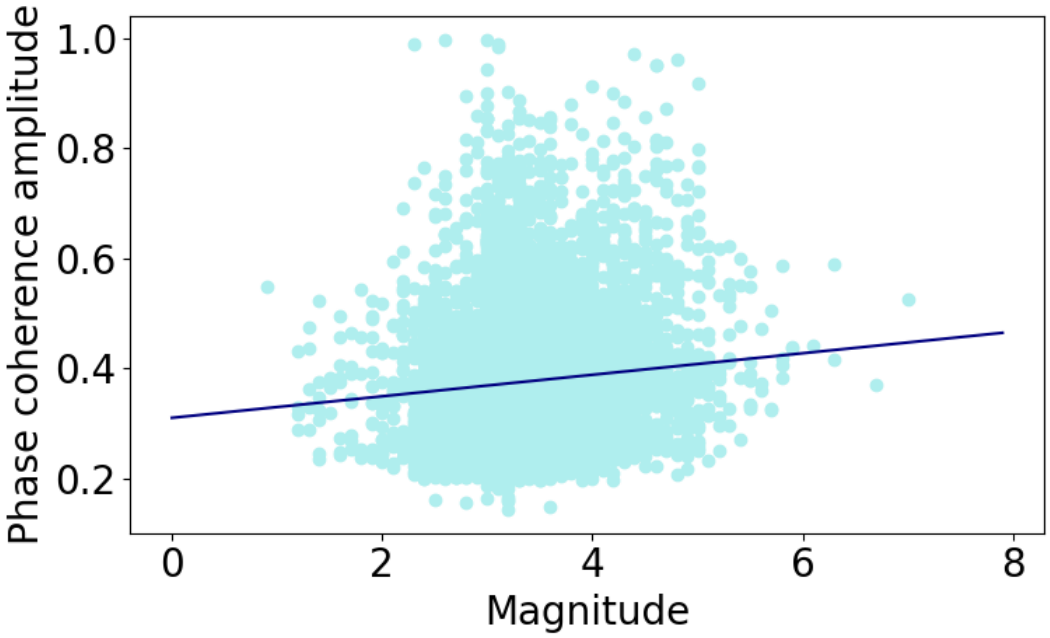


Figure C.12: Plot of phase coherence detection amplitudes vs magnitude of associated events. Linear regression line is shown in navy: slope=0.020, intercept=0.310, rvalue=0.092, pvalue=1.783e-24, stderr=0.002, intercept stderr=0.007)

Bibliography

Abercrombie, R. and Mori, J. (1994). Local observations of the onset of a large earthquake: 28 June 1992 Landers, California. *Bulletin of the Seismological Society of America*, 84(3):725–734.

Abercrombie, R. E. (2021). Resolution and uncertainties in estimates of earthquake stress drop and energy release. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2196):20200131.

Abercrombie, R. E., Agnew, D. C., and Wyatt, F. K. (1995). Testing a model of earthquake nucleation. *Bulletin of the Seismological Society of America*, 85(6):1873–1878.

Abercrombie, R. E. and Mori, J. (1996). Occurrence patterns of foreshocks to large earthquakes in the western United States. *Nature*, 381(6580):303–307.

Agnew, D. C. (2002). 1 History of seismology. In *International Geophysics*, volume 81, pages 3–11. Elsevier.

Aki, K. and Richards, P. G. (2009). *Quantitative Seismology*. Univ. Science Books, Mill Valley, Calif, 2. ed., corr. print edition.

Alaska Earthquake Center, Univ. of Alaska Fairbanks (1987). Alaska Geophysical Network.

Alberta Geological Survey / Alberta Energy Regulator (2013). Regional Alberta Observatory for Earthquake Studies Network.

Albuquerque Seismological Laboratory (ASL)/USGS (1980). US Geological Survey Networks.

Albuquerque Seismological Laboratory/USGS (2013). Central and Eastern US Network.

Albuquerque Seismological Laboratory/USGS (2014). Global Seismograph Network (GSN - IRIS/USGS).

Allen, R. M. (2003). Rapid magnitude determination for earthquake early warning. In *The Many Facets of Seismic Risk*.

Allen, R. M., Brown, H., Hellweg, M., Khainovski, O., Lombard, P., and Neuhauser, D. (2009). Real-time earthquake detection and hazard assessment by ElarmS across California. *Geophysical Research Letters*, 36(5):L00B08.

Allen, R. M. and Kanamori, H. (2003). The Potential for Earthquake Early Warning in Southern California. *Science*, 300(5620):786–789.

Anderson, D. L., Kanamori, H., Hart, R. S., and Liu, H.-P. (1977). The Earth as a Seismic Absorption Band. *Science*, 196(4294):1104–1106.

Anderson, J. G. and Chen, Q. (1995). Beginnings of earthquakes in the Mexican subduction zone on strong-motion accelerograms. *Bulletin of the Seismological Society of America*, 85(4):1107–1115.

Archuleta, R. J. and Ji, C. (2016). Moment rate scaling for earthquakes $3.3 \leq M \leq 5.3$ with implications for stress drop. *Geophysical Research Letters*, 43(23):12,004–12,011.

Aristotle University of Thessaloniki (1981). Aristotle University of Thessaloniki Seismological Network.

Arizona Geological Survey (2007). Arizona Broadband Seismic Network.

Bachl, F. E., Lindgren, F., Borchers, D. L., and Illian, J. B. (2019). inlabru: an R package for Bayesian spatial modelling from ecological survey data. *Methods in Ecology and Evolution*, 10(6):760–766.

Bantidi, T. M., Ishibe, T., Tuluka, G. M., and Enescu, B. (2024). Estimating spatio-temporal variable parameters of Epidemic Type Aftershock Sequence model in a region with limited seismic network coverage: a case study of the East African Rift System. *Geophysical Journal International*, 239(2):882–900.

Bassin, C., Laske, G., and Masters, G. (2000). Crust 2.0.

Ben-Zeev, S. and Lior, I. (2025). Earthquake Location with Distributed Acoustic Sensing Subarray Beamforming with Implications for Earthquake Early Warning. *Seismological Research Letters*.

Beresnev, I. and Atkinson, G. (2001). Subevent structure of large earthquakes—A ground-motion perspective. *Geophysical Research Letters*, 28(1):53–56.

Beroza, G. C. and Ellsworth, W. L. (1996). Properties of the seismic nucleation phase. *Tectonophysics*, 261(1-3 SPEC. ISS.):209–227.

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., and Wassermann, J. (2010). ObsPy: A Python Toolbox for Seismology. *Seismological Research Letters*, 81(3):530–533.

Bizzarri, A. (2011). On the Deterministic Description of Earthquakes. *Reviews of Geophysics*, 49(3).

Blaser, L., Krüger, F., Ohrnberger, M., and Scherbaum, F. (2010). Scaling relations of earthquake source parameter estimates with special focus on subduction environment. *Bulletin of the Seismological Society of America*, 100(6):2914–2926.

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.

Boatwright, J. (1988). The seismic radiation from composite models of faulting. *Bulletin of the Seismological Society of America*, 78(2):489–508.

Bock, Y., Melgar, D., and Crowell, B. W. (2011). Real-Time Strong-Motion Broadband Displacements from Collocated GPS and Accelerometers. *Bulletin of the Seismological Society of America*, 101(6):2904–2925.

Bolt, B. A. (1999). *Earthquakes*. W.H. Freeman, New York, 4. ed., 1. printing edition.

Bolton, D. C., Shokouhi, P., Rouet-Leduc, B., Hulbert, C., Rivière, J., Marone, C., and Johnson, P. A. (2019). Characterizing Acoustic Signals and Searching for Precursors during the Laboratory Seismic Cycle Using Unsupervised Machine Learning. *Seismological Research Letters*, 90(3):1088–1098.

Bolton, D. C., Shreedharan, S., Rivière, J., and Marone, C. (2020). Acoustic Energy Release During the Laboratory Seismic Cycle: Insights on Laboratory Earthquake Precursors and Prediction. *Journal of Geophysical Research: Solid Earth*, 125(8):e2019JB018975.

Bolton, D. C., Shreedharan, S., Rivière, J., and Marone, C. (2021). Frequency-Magnitude Statistics of Laboratory Foreshocks Vary With Shear Velocity, Fault Slip Rate, and Shear Stress. *Journal of Geophysical Research: Solid Earth*, 126(11):e2021JB022175.

Bose, M., Hauksson, E., Solanki, K., Kanamori, H., Wu, Y.-M., and Heaton, T. H. (2009). A New Trigger Criterion for Improved Real-Time Performance of Onsite Earthquake Early Warning in Southern California. *Bulletin of the Seismological Society of America*, 99(2A):897–905.

Bouchon, M. (1976). Teleseismic body wave radiation from a seismic source in a layered medium. *Geophysical Journal International*, 47(3):515–530.

Bouchon, M. (1981). A simple method to calculate Green's functions for elastic layered media. *Bulletin of the Seismological Society of America*, 71(4):959–971.

Bouchon, M., Durand, V., Marsan, D., Karabulut, H., and Schmittbuhl, J. (2013). The long precursory phase of most large interplate earthquakes. *Nature Geoscience*, 6(4):299–302.

Bouchon, M., Karabulut, H., Aktar, M., Özalaybey, S., Schmittbuhl, J., and Bouin, M. P. (2011). Extended nucleation of the 1999 M_W7.6 Izmit earthquake. *Science*, 331(6019):877–880.

Brodsky, E. E. (2011). The spatial density of foreshocks. *Geophysical Research Letters*, 38(10).

Brodsky, E. E. and Kanamori, H. (2001). Elastohydrodynamic lubrication of faults. *Journal of Geophysical Research: Solid Earth*, 106(B8):16357–16374.

Brown, H. M., Allen, R. M., and Grasso, V. F. (2009). Testing ElarmS in Japan. *Seismological Research Letters*, 80(5):727–739.

Brown, H. M., Allen, R. M., Hellweg, M., Khainovski, O., Neuhauser, D., and Souf, A. (2011). Development of the ElarmS methodology for earthquake early warning: Realtime application in California and offline testing in Japan. *Soil Dynamics and Earthquake Engineering*, 31(2):188–200.

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

Bureau of Economic Geology, The University of Texas at Austin (2016). Texas Seismological Network.

Cabrera-Pérez, I., D'Auria, L., Soubestre, J., del Pezzo, E., Prudencio, J., Ibáñez, J. M., Jiménez-Mejías, M., Padilla, G. D., Barrancos, J., and Pérez, N. M. (2024). 3-D intrinsic attenuation tomography using ambient seismic noise applied to La Palma Island (Canary Islands). *Scientific Reports*, 14(1):27354.

California Institute of Technology and United States Geological Survey Pasadena (1926). Southern California Seismic Network.

Cattania, C. and Segall, P. (2021). Precursory slow slip and foreshocks on rough faults. *Journal of Geophysical Research: Solid Earth*, 126(4):e2020JB020430.

Cattania, C., Werner, M. J., Marzocchi, W., Hainzl, S., Rhoades, D., Gerstenberger, M., Liukis, M., Savran, W., Christophersen, A., Helmstetter, A., Jimenez, A., Steacy, S., and Jordan, T. H. (2018). The forecasting skill of physics-based seismicity models during the 2010-2012 Canterbury, New Zealand, earthquake sequence. *Seismological Research Letters*, 89(4):1238–1250.

Central Asian Institute for Applied Geosciences (2008). Central Asian Seismic Network of CAIAG.

Centre IRD de Noumea, Nouvelle-Caledonie (2010). New CaleDonia Broadband Seismic Network.

Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE), Ensenada (1980). Red Sísmica del Noroeste de México.

Cesca, S., Sobiesiak, M., Tassara, A., Olcay, M., Günther, E., Mikulla, S., and Dahm, T. (2009). The Iquique Local Network and PicArray.

Chai, C., Maceira, M., Santos-Villalobos, H. J., Schoenball, M., and EGS Collab Team (2020). Automatic Seismic Phase Picking Using Deep Learning for the EGS Collab project. In *Proceedings, 45th Workshop on Geothermal Reservoir Engineering*, Stanford.

Chapman, C. H. (1978). A new method for computing synthetic seismograms. *Geophysical Journal International*, 54(3):481–518.

Charles University in Prague (Czech), Institute of Geonics, Institute of Geophysics, Academy of Sciences of the Czech Republic, Institute of Physics of the Earth Masaryk University (Czech), and Institute of Rock Structure and Mechanics (1973). Czech Regional Seismic Network.

Chen, X. and Shearer, P. M. (2013). California foreshock sequences suggest aseismic triggering process. *Geophysical Research Letters*, 40(11):2602–2607.

Chen, X. and Shearer, P. M. (2016). Analysis of Foreshock Sequences in California and Implications for Earthquake Triggering. *Pure and Applied Geophysics*, 173(1):133–152.

Chounet, A., Vallée, M., Causse, M., and Courboulex, F. (2018). Global catalog of earthquake rupture velocities shows anticorrelation between stress drop and rupture velocity. *Tectonophysics*, 733:148–158.

Chung, A. I., Henson, I., and Allen, R. M. (2019). Optimizing Earthquake Early Warning Performance: ElarmS-3. *Seismological Research Letters*, 90(2A):727–743.

Cochran, E. S., Page, M. T., Van Der Elst, N. J., Ross, Z. E., and Trugman, D. T. (2023). Fault Roughness at Seismogenic Depths and Links to Earthquake Behavior. *The Seismic Record*, 3(1):37–47.

Cochran, E. S., Ross, Z. E., Harrington, R. M., Dougherty, S. L., and Rubinstein, J. L. (2018). Induced Earthquake Families Reveal Distinctive Evolutionary Patterns Near Disposal Wells. *Journal of Geophysical Research: Solid Earth*, 123(9):8045–8055.

Cochran, E. S., Wickham-Piotrowski, A., Kemna, K. B., Harrington, R. M., Dougherty, S. L., and Peña Castro, A. F. (2020). Minimal Clustering of Injection-Induced Earthquakes Observed with a Large-n Seismic Array. *Bulletin of the Seismological Society of America*, 110(5):2005–2017.

Colombelli, S., Allen, R. M., and Zollo, A. (2013). Application of real-time GPS to earthquake early warning in subduction and strike-slip environments. *Journal of Geophysical Research: Solid Earth*, 118(7):3448–3461.

Colombelli, S., Festa, G., and Zollo, A. (2020). Early rupture signals predict the final earthquake size. *Geophysical Journal International*, 223(1):692–706.

Colombelli, S., Zollo, A., Festa, G., and Kanamori, H. (2012). Early magnitude and potential damage zone estimates for the great M_W 9 Tohoku-Oki earthquake. *Geophysical Research Letters*, 39(22).

Colombelli, S., Zollo, A., Festa, G., and Picozzi, M. (2014). Evidence for a difference in rupture initiation between small and large earthquakes. *Nature Communications*, 5:1–5.

Colorado Geological Survey (2016). Colorado Geological Survey Seismic Network.

Colquhoun, R. (2024). seismo_det: A toolbox to investigate earthquake determinism using early features of seismograms.

Corinth Rift Laboratory Team And RESIF Datacenter (2013). CL - Corinth Rift Laboratory Seismological Network (CRLNET).

Crameri, F. (2023). Scientific colour maps.

Crotwell, H. P., Owens, T. J., and Ritsema, J. (1999). The TauP Toolkit: Flexible Seismic Travel-time and Ray-path Utilities. *Seismological Research Letters*, 70(2):154–160.

Cua, G., Fischer, M., Heaton, T., and Wiemer, S. (2009). Real-time Performance of the Virtual Seismologist Earthquake Early Warning Algorithm in Southern California. *Seismological Research Letters*, 80(5):740–747.

Dahlen, F. A. and Baig, A. M. (2002). Fréchet kernels for body-wave amplitudes. *Geophysical Journal International*, 150(2):440–466.

Danré, P., Yin, J., Lipovsky, B. P., and Denolle, M. A. (2019). Earthquakes Within Earthquakes: Patterns in Rupture Complexity. *Geophysical Research Letters*, 46(13):7352–7360.

Das, S. and Scholz, C. H. (1981). Theory of time-dependent rupture in the Earth. *Journal of Geophysical Research: Solid Earth*, 86(B7):6039–6051.

de Costa Rica, U. (2017). Información de la Red Sismológica Nacional de Costa Rica.

Deichmann, N. (1997). Far-field pulse shapes from circular sources with variable rupture velocities. *Bulletin of the Seismological Society of America*, 87(5):1288–1296.

Denolle, M. A. (2019). Energetic Onset of Earthquakes. *Geophysical Research Letters*, 46(5):2458–2466.

Denolle, M. A. and Shearer, P. M. (2016). New perspectives on self-similarity for shallow thrust earthquakes. *Journal of Geophysical Research: Solid Earth*, 121(9):6533–6565.

Department of Meteorology and Hydrology - National Earthquake Data Center (2016). Myanmar National Seismic Network.

Dieterich, J. (1994). A constitutive law for rate of earthquake production and its application to earthquake clustering. *Journal of Geophysical Research: Solid Earth*, 99(B2):2601–2618.

Dieterich, J. H. (1978). Preseismic fault slip and earthquake prediction. *Journal of Geophysical Research: Solid Earth*, 83(8):3940–3948.

Dieterich, J. H. (1986). A Model for the Nucleation of Earthquake Slip. In Das, S., Boatwright, J., and Scholz, C. H., editors, *Geophysical Monograph Series*, pages 37–47. American Geophysical Union, Washington, D. C.

Disaster and Emergency Management Authority (1990). Turkish National Seismic Network.

Dodge, D. A., Beroza, G. C., and Ellsworth, W. L. (1996). Detailed observations of California foreshock sequences: Implications for the earthquake initiation process. *Journal of Geophysical Research: Solid Earth*, 101(B10):22371–22392.

Dresen, G., Kwiatek, G., Goebel, T., and Ben-Zion, Y. (2020). Seismic and Aseismic Preparatory Processes Before Large Stick–Slip Failure. *Pure and Applied Geophysics*, 177(12):5741–5760.

Dziewonski, A. M. and Anderson, D. L. (1981). Preliminary reference Earth model. *Physics of the Earth and Planetary Interiors*, 25(4):297–356.

Dziewonski, A. M. and Gilbert, F. (1974a). Evidence of precursive compression for two deep earthquakes (reply). *Nature*, 252(5478):29–29.

Dziewonski, A. M. and Gilbert, F. (1974b). Temporal Variation of the Seismic Moment Tensor and the Evidence of Precursive Compression for Two Deep Earthquakes. *Nature*, 247(5438):185–188.

Ekström, G., Nettles, M., and Dziewoński, A. M. (2012). The global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes. *Physics of the Earth and Planetary Interiors*, 200-201:1–9.

Ellsworth, W. L. (2013). Injection-Induced Earthquakes. *Science*, 341(6142):1225942.

Ellsworth, W. L. and Beroza, G. C. (1995). Seismic evidence for an earthquake nucleation phase. *Science*, 268(5212):851–855.

Ellsworth, W. L. and Beroza, G. C. (1998). Observation of the seismic nucleation phase in the Ridgecrest, California, Earthquake sequence. *Geophysical Research Letters*, 25(3):401–404.

Ellsworth, W. L. and Bulut, F. (2018). Nucleation of the 1999 Izmit earthquake by a triggered cascade of foreshocks. *Nature Geoscience*, 11(7):531–535.

Eshelby, J. D. (1957). The determination of the elastic field of an ellipsoidal inclusion, and related problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 241(1226):376–396.

Evison, F. F. and Rhoades, D. A. (1999). The precursory earthquake swarm and the inferred precursory quarm. *New Zealand Journal of Geology and Geophysics*, 42(2):229–236.

Federal Institute for Geosciences and Natural Resources (1976). German Regional Seismic Network (GRSN).

Felzer, K. R., Abercrombie, R. E., and Ekström, G. (2004). A Common Origin for Aftershocks, Foreshocks, and Multiplets. *Bulletin of the Seismological Society of America*, 94(1):88–98.

Festa, G., Zollo, A., and Lancieri, M. (2008). Earthquake magnitude estimation from early radiated energy. *Geophysical Research Letters*, 35(22):L22307.

Fleming, K., Picozzi, M., Milkereit, C., Kuhnlenz, F., Lichtblau, B., Fischer, J., Zulfikar, C., Ozel, O., and the SAFER and EDIM working groups (2009). The Self-organizing Seismic Early Warning Information Network (SOSEWIN). *Seismological Research Letters*, 80(5):755–771.

Foreman-Mackey, D., Hogg, D. W., Lang, D., and Goodman, J. (2013). emcee: The MCMC Hammer. *Publications of the Astronomical Society of the Pacific*, 125(925):306.

Foulger, G. R., Wilson, M. P., Gluyas, J. G., Julian, B. R., and Davies, R. J. (2018). Global review of human-induced earthquakes. *Earth-Science Reviews*, 178:438–514.

French Landslide Observatory – Seismological Datacenter / RESIF (2006). Observatoire Multi-disciplinaire des Instabilités de Versants (OMIV).

Gabriel, A., Ampuero, J., Dalguer, L. A., and Mai, P. M. (2012). The transition of dynamic rupture styles in elastic media under velocity-weakening friction. *Journal of Geophysical Research: Solid Earth*, 117(B9):2012JB009468.

Geller, R. and Mueller, C. (1980). Four similar earthquakes in central California. *Geophysical Research Letters*, 7(10):821–824.

Geller, R. J. (1974). Evidence of precursive compression for two deep earthquakes. *Nature*, 252(5478):28–29.

GEOFON Data Centre (1993). GEOFON Seismic Network.

Geological Survey Department Cyprus (2013). Cyprus Broadband Seismological Network.

Geological Survey of Canada (2013). Public Safety Geoscience Program Canadian Research Network.

Geophysical Survey of the National Academy of sciences of Tajikistan (2009). Tajikistan National Seismic Network.

GFZ German Research Centre For Geosciences and Institut Des Sciences De L'Univers-Centre National De La Recherche CNRS-INSU (2006). IPOC Seismic Network.

GNS Science (2021). GeoNet Aotearoa New Zealand Seismic Digital Waveform Dataset.

Goebel, T. H. W., Schorlemmer, D., Becker, T. W., Dresen, G., and Sammis, C. G. (2013). Acoustic emissions document stress changes over many seismic cycles in stick-slip experiments. *Geophysical Research Letters*, 40(10):2049–2054.

Goldberg, D. E., Koch, P., Melgar, D., Riquelme, S., and Yeck, W. L. (2022). Beyond the Teleseism: Introducing Regional Seismic and Geodetic Data into Routine USGS Finite-Fault Modeling. *Seismological Research Letters*, 93(6):3308–3323.

Goldsby, D. L. and Tullis, T. E. (2011). Flash Heating Leads to Low Frictional Strength of Crustal Rocks at Earthquake Slip Rates. *Science*, 334(6053):216–218.

Gomberg, J. (2018). Unsettled earthquake nucleation. *Nature Geoscience*, 11(7):463–464.

Gomberg, J., Wech, A., Creager, K., Obara, K., and Agnew, D. (2016). Reconsidering earthquake scaling. *Geophysical Research Letters*, 43(12):6243–6251.

Gombert, B. and Hawthorne, J. C. (2023). Rapid Tremor Migration During Few Minute-Long Slow Earthquakes in Cascadia. *Journal of Geophysical Research: Solid Earth*, 128(2):e2022JB025034.

Goodman, J. and Weare, J. (2010). Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, 5(1):65–80.

Guo, Y., Zhuang, J., and Zhou, S. (2015). A hypocentral version of the space–time ETAS model. *Geophysical Journal International*, 203(1):366–372.

Hainzl, S. (2013). Comment on "Self-similar earthquake triggering, Båth's law, and foreshock/aftershock magnitudes: Simulations, theory, and results for southern California" by P. M. Shearer. *Journal of Geophysical Research: Solid Earth*, 118(3):1188–1191.

Hainzl, S. (2021). ETAS-Approach Accounting for Short-Term Incompleteness of Earthquake Catalogs. *Bulletin of the Seismological Society of America*, 112(1):494–507.

Harbord, C. W., Nielsen, S. B., De Paola, N., and Holdsworth, R. E. (2017). Earthquake nucleation on rough faults. *Geology*, 45(10):931–934.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825):357–362.

Hartzell, S. H. and Heaton, T. H. (1983). Inversion of strong ground motion and teleseismic waveform data for the fault rupture history of the 1979 Imperial Valley, California, earthquake. *Bulletin of the Seismological Society of America*, 73(6A):1553–1583.

Haskell, N. A. (1964). Total energy and energy spectral density of elastic wave radiation from propagating faults. *Bulletin of the Seismological Society of America*, 54(6A):1811–1841.

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90.

Hawthorne, J. C. and Ampuero, J.-P. (2017). A phase coherence approach to identifying co-located earthquakes and tremor. *Geophysical Journal International*, page ggx012.

Hayes, G. P. (2017). The finite, kinematic rupture properties of great-sized earthquakes since 1990. *Earth and Planetary Science Letters*, 468:94–100.

Heaton, T. H. (1990). Evidence for and implications of self-healing pulses of slip in earthquake rupture. *Physics of the Earth and Planetary Interiors*, 64(1):1–20.

Helmstetter, A. and Sornette, D. (2003). Foreshocks explained by cascades of triggered seismicity. *Journal of Geophysical Research: Solid Earth*, 108(B10).

Helmstetter, A., Sornette, D., and Grasso, J.-R. (2003). Mainshocks are aftershocks of conditional foreshocks: How do foreshock statistical properties emerge from aftershock laws. *Journal of Geophysical Research: Solid Earth*, 108(B1).

Hoshiba, M. and Aoki, S. (2015). Numerical Shake Prediction for Earthquake Early Warning: Data Assimilation, Real-Time Shake Mapping, and Simulation of Wave Propagation. *Bulletin of the Seismological Society of America*, 105(3):1324–1338.

Hoshiba, M. and Iwakiri, K. (2011). Initial 30 seconds of the 2011 off the Pacific coast of Tohoku Earthquake (M_W 9.0)—amplitude and τ c for magnitude estimation for Earthquake Early Warning—. *Earth, Planets and Space*, 63(7):553–557.

Hosseini, K. and Sigloch, K. (2017). ObspyDMT: a Python toolbox for retrieving and processing large seismological data sets. *Solid Earth*, 8(5):1047–1070.

Houston, H. (2001). Influence of depth, focal mechanism, and tectonic setting on the shape and duration of earthquake source time functions. *Journal of Geophysical Research: Solid Earth*, 106(B6):11137–11150.

Houston, H., Benz, H. M., and Vidale, J. E. (1998). Time functions of deep earthquakes from broadband and short-period stacks. *Journal of Geophysical Research: Solid Earth*, 103(B12):29895–29913.

Hsiao, N.-C., Wu, Y.-M., Shin, T.-C., Zhao, L., and Teng, T.-L. (2009). Development of earthquake early warning system in Taiwan. *Geophysical Research Letters*, 36(2):L00B02.

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3):90–95.

Icelandic Meteorological Office (1992). Icelandic National Digital Seismograph Network.

Idaho National Laboratory (1972). INL Seismic Monitoring Program.

Iio, Y. (1992). Slow initial phase of the P-wave velocity pulse generated by microearthquakes. *Geophysical Research Letters*, 19(5):477–480.

Iio, Y. (1995). Observations of the slow initial phase generated by microearthquakes: Implications for earthquake nucleation and propagation. *Journal of Geophysical Research: Solid Earth*, 100(B8):15333–15349.

Institut Cartogràfic i Geològic de Catalunya (1984). Catalan Seismic Network.

Institut De Physique Du Globe De Paris (IPGP) (2008). GNSS, seismic broadband and strong motion permanent networks in West Indies.

Institut de physique du globe de Paris (IPGP) and École et Observatoire des Sciences de la Terre de Strasbourg (EOST) (1982). GEOSCOPE, French Global Network of broad band seismic stations.

Institute of Geophysics China Earthquake Administration (IGPCEA) (2000). China National Seismic Network, Data Management Centre of China National Seismic Network at Institute of Geophysics, CEA.

Institute of GeoSciences (IGEO), Polytechnic University of Tirana (PUT) (2002). Albanian Seismological Network.

Instituto Geografico Nacional, Spain (1999). Spanish Digital Seismic Network.

Instituto Nicaraguense de Estudios Territoriales (INETER) (1975). Nicaraguan Seismic Network.

Instituto Português do Mar e da Atmosfera, I.P. (2006). Portuguese National Seismic Network.

IRIS Transportable Array (2003). USArray Transportable Array.

Istituto Nazionale di Geofisica e Vulcanologia (INGV) (2005). Rete Sismica Nazionale (RSN).

Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (2015). International OGS Network.

Ji, C., Wald, D. J., and Helmberger, D. V. (2002). Source Description of the 1999 Hector Mine, California, Earthquake, Part I: Wavelet Domain Inversion Theory and Resolution Analysis. *Bulletin of the Seismological Society of America*, 92(4):1192–1207.

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

Johnston, M. J. S., Linde, A. T., and Agnew, D. C. (1994). Continuous borehole strain in the San Andreas fault zone before, during, and after the 28 June 1992, MW 7.3 Landers, California, earthquake. *Bulletin of the Seismological Society of America*, 84(3):799–805.

Johnston, M. J. S., Linde, A. T., and Gladwin, M. T. (1990). Near-field high resolution strain measurements prior to the October 18, 1989, Loma Prieta M $_s$ 7.1 Earthquake. *Geophysical Research Letters*, 17(10):1777–1780.

Jones, L. and Molnar, P. (1976). Frequency of foreshocks. *Nature*, 262(5570):677–679.

Juhel, K., Ampuero, J. P., Barsuglia, M., Bernard, P., Chassande-Mottin, E., Fiorucci, D., Harms, J., Montagner, J.-P., Vallée, M., and Whiting, B. F. (2018). Earthquake Early Warning Using Future Generation Gravity Strainmeters. *Journal of Geophysical Research: Solid Earth*, 123(12):10,889–10,902.

Julian, B. R., Miller, A. D., and Foulger, G. R. (1998). Non-double-couple earthquakes 1. Theory. *Reviews of Geophysics*, 36(4):525–549.

Kamranzad, F. (2024). *Resolving bias due to short- and long-term catalogue incompleteness and improving accuracy by optimal sampling in the epidemic-type aftershock sequence (ETAS) model*. PhD thesis, University of Edinburgh.

Kanamori, H. (2004). The diversity of the physics of earthquakes. *Proceedings of the Japan Academy, Series B*, 80(7):297–316.

Kanamori, H. (2005). Real-Time Seismology and Earthquake Damage Mitigation. *Annual Review of Earth and Planetary Sciences*, 33(1):195–214.

Kanamori, H. and Anderson, D. L. (1975). Theoretical basis of some empirical relations in seismology. *Bulletin of the Seismological Society of America*, 65(5):1073–1095.

Kanamori, H. and Anderson, D. L. (1977). Importance of physical dispersion in surface wave and free oscillation problems: Review. *Reviews of Geophysics*, 15(1):105–112.

Kanamori, H. and Mori, J. (2000). Microscopic processes on a fault plane and their implications for earthquake dynamics. In Boschi, E., Ekström, G., and Morelli, A., editors, *Problems in Geophysics for the New Millennium. A collection of papers in honor of Adam M. Dziewonski*. Compositori.

Kanamori, H., Mori, J., Hauksson, E., Heaton, T. H., Hutton, L. K., and Jones, L. M. (1993). Determination of energy release and m_l using TerraScope. *Bulletin of the Seismological Society of America*, 83(2):330–346.

Kandilli Observatory And Earthquake Research Institute (1971). Kandilli Observatory And Earthquake Research Institute (KOERI).

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

Kaneko, Y. and Shearer, P. M. (2015). Variability of seismic source spectra, estimated stress drop, and radiated energy, derived from cohesive-zone models of symmetrical and asymmetrical circular and elliptical ruptures. *Journal of Geophysical Research: Solid Earth*, 120(2):1053–1079.

Kato, A., Fukuda, J., Nakagawa, S., and Obara, K. (2016). Foreshock migration preceding the 2016 M_W 7.0 Kumamoto earthquake, Japan. *Geophysical Research Letters*, 43(17):8945–8953.

Ke, C.-Y., McLaskey, G. C., and Kammer, D. S. (2022). Earthquake breakdown energy scaling despite constant fracture energy. *Nature Communications*, 13(1):1005.

Kemp, M. (2025). *Alaska through time: reconstructing plate tectonics by linking the deep mantle to the surface*. PhD thesis, University of Oxford.

Kennett, B. L. N., Engdahl, E. R., and Buland, R. (1995). Constraints on seismic velocities in the Earth from traveltimes. *Geophysical Journal International*, 122(1):108–124.

Kentucky Geological Survey/Univ. of Kentucky (1982). Kentucky Seismic and Strong Motion Network.

Keranen, K. M. and Weingarten, M. (2018). Induced Seismicity. *Annual Review of Earth and Planetary Sciences*, 46(1):149–174.

Kikuchi, M. and Kanamori, H. (1991). Inversion of complex body waves—III. *Bulletin of the Seismological Society of America*, 81(6):2335–2350.

Kilb, D. and Gomberg, J. (1999). The initial subevent of the 1994 Northridge, California, earthquake: Is earthquake size predictable? *Journal of Seismology*, 3(4):409–420.

KNDC/Institute of Geophysical Research (Kazakhstan) (1994). Kazakhstan Network.

KNMI (1993). Netherlands Seismic and Acoustic Network.

KNMI (2006). Caribbean Netherlands Seismic Network.

Knopoff, L. (1958). Energy Release in Earthquakes. *Geophysical Journal International*, 1(1):44–52.

Koch, P., Bravo, F., Riquelme, S., and Crempien, J. G. F. (2019). Near-Real-Time Finite-Fault Inversions for Large Earthquakes in Chile Using Strong-Motion Data. *Seismological Research Letters*.

Kohler, M. D., Cochran, E. S., Given, D., Guiwits, S., Neuhauser, D., Henson, I., Hartog, R., Bodin, P., Kress, V., Thompson, S., Felizardo, C., Brody, J., Bhadha, R., and Schwarz, S. (2018). Earthquake Early Warning ShakeAlert System: West Coast Wide Production Prototype. *Seismological Research Letters*, 89(1):99–107.

Kohler, M. D., Smith, D. E., Andrews, J., Chung, A. I., Hartog, R., Henson, I., Given, D. D., De Groot, R., and Guiwits, S. (2020). Earthquake Early Warning ShakeAlert 2.0: Public Rollout. *Seismological Research Letters*, 91(3):1763–1775.

Kohler, N., Cua, G., Wenzel, F., and Bose, M. (2009). Rapid Source Parameter Estimations of Southern California Earthquakes Using PreSEIS. *Seismological Research Letters*, 80(5):748–754.

Kuyuk, H. S. and Allen, R. M. (2013). A global approach to provide magnitude estimates for earthquake early warning alerts. *Geophysical Research Letters*, 40(24):6329–6333.

Kyrgyz Institute of Seismology, IVTAN/KIS and University of California, San Diego (1991). Kyrgyz Seismic Telemetry Network.

Kyrgyz Institute of Seismology, KIS (2007). Kyrgyz Digital Network.

Lamont Doherty Earth Observatory (LDEO), Columbia University (1970). Lamont-Doherty Cooperative Seismographic Network.

Lancieri, M., Fuenzalida, A., Ruiz, S., and Madariaga, R. (2011). Magnitude Scaling of Early-Warning Parameters for the M_W 7.8 Tocopilla, Chile, Earthquake and Its Aftershocks. *Bulletin of the Seismological Society of America*, 101(2):447–463.

Lancieri, M. and Zollo, A. (2008). A Bayesian approach to the real-time estimation of magnitude from the early P and S wave displacement peaks. *Journal of Geophysical Research*, 113(B12):B12302.

Lapusta, N. and Rice, J. R. (2003). Nucleation and early seismic propagation of small and large events in a crustal earthquake model. *Journal of Geophysical Research: Solid Earth*, 108(B4).

Latour, S., Schubnel, A., Nielsen, S., Madariaga, R., and Vinciguerra, S. (2013). Characterization of nucleation during laboratory earthquakes. *Geophysical Research Letters*, 40(19):5064–5069.

Leyton, F., Ruiz, S., Baez, J. C., Meneses, G., and Madariaga, R. (2018). How Fast Can We Reliably Estimate the Magnitude of Subduction Earthquakes? *Geophysical Research Letters*, 45(18):9633–9641.

Li, J., Sornette, D., Wu, Z., Zhuang, J., and Jiang, C. (2024). Revisiting Seismicity Criticality: A New Framework for Bias Correction of Statistical Seismology Model Calibrations.

Liu, M., Huang, Y., and Ritsema, J. (2023). Characterizing Multisubevent Earthquakes Using the Brune Source Model. *Bulletin of the Seismological Society of America*, 113(2):577–591.

Liu, Y. and Rice, J. R. (2007). Spontaneous and triggered aseismic deformation transients in a subduction fault model. *Journal of Geophysical Research: Solid Earth*, 112(9):1–23.

Lockman, A. B. and Allen, R. M. (2005). Single-Station Earthquake Characterization for Early Warning. *Bulletin of the Seismological Society of America*, 95(6):2029–2039.

Lockman, A. B. and Allen, R. M. (2007). Magnitude-Period Scaling Relations for Japan and the Pacific Northwest: Implications for Earthquake Early Warning. *Bulletin of the Seismological Society of America*, 97(1B):140–150.

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. (1978). The dynamic field of Haskell's rectangular dislocation fault model. *Bulletin of the Seismological Society of America*, 68(4):869–887.

Marone, C. (1998). Laboratory-Derived Friction Laws and Their Application to Seismic Faulting. *Annual Review of Earth and Planetary Sciences*, 26(1):643–696.

Martens, H. R. and University Of Montana (2017). University of Montana Seismic Network.

McGuire, J., Boettcher, M., and Jordan, T. (2005). Foreshock sequences and short-term earthquake predictability on East Pacific Rise transform faults. *Nature*, 434:457–61.

McKinney, W. (2010). Data Structures for Statistical Computing in Python. pages 56–61, Austin, Texas.

McLaskey, G. C. and Kilgore, B. D. (2013). Foreshocks during the nucleation of stick-slip instability. *Journal of Geophysical Research: Solid Earth*, 118(6):2982–2997.

McLaskey, G. C. and Lockner, D. A. (2014). Preslip and cascade processes initiating laboratory stick slip. *Journal of Geophysical Research: Solid Earth*, 119(8):6323–6336.

MedNet Project Partner Institutions (1990). Mediterranean Very Broadband Seismographic Network (MedNet).

Meier, M., Kodera, Y., Böse, M., Chung, A., Hoshiba, M., Cochran, E., Minson, S., Hauksson, E., and Heaton, T. (2020a). How Often Can Earthquake Early Warning Systems Alert Sites With High-Intensity Ground Motion? *Journal of Geophysical Research: Solid Earth*, 125(2):e2019JB017718.

Meier, M.-A., Ampuero, J.-P., Cochran, E., and Page, M. (2020b). Apparent earthquake rupture predictability. *Geophysical Journal International*, 225(1):657–663.

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

Meier, M. A., Heaton, T., and Clinton, J. (2016). Evidence for universal earthquake rupture initiation behavior. *Geophysical Research Letters*, 43(15):7991–7996.

Melgar, D. and Hayes, G. P. (2017). Systematic Observations of the Slip Pulse Properties of Large Earthquake Ruptures. *Geophysical Research Letters*, 44(19):9691–9698.

Melgar, D. and Hayes, G. P. (2019). Characterizing large earthquakes before rupture is complete. *Science Advances*, 5(5):1–8.

Minson, S. E., Baltay, A. S., Cochran, E. S., McBride, S. K., and Milner, K. R. (2021). Shaking is Almost Always a Surprise: The Earthquakes That Produce Significant Ground Motion. *Seismological Research Letters*, 92(1):460–468.

Mohammadzaheri, A. (2019). *P-wave velocity structures under South America from multi-frequency tomography*. phd, University of Oxford / University of Oxford.

Molkenthin, C., Donner, C., Reich, S., Zöller, G., Hainzl, S., Holschneider, M., and Opper, M. (2022). GP-ETAS: semiparametric Bayesian inference for the spatio-temporal epidemic type aftershock sequence model. *Statistics and Computing*, 32(2):29.

Molnar, P., Tucker, B. E., and Brune, J. N. (1973). Corner frequencies of P and S waves and models of earthquake sources. *Bulletin of the Seismological Society of America*, 63(6-1):2091–2104.

Mori, J. and Kanamori, H. (1996). Initial rupture of earthquakes in the 1995 Ridgecrest, California Sequence. *Geophysical Research Letters*, 23(18):2437–2440.

Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., and Beroza, G. C. (2020). Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, 11(1):3952.

Murphy, S. and Nielsen, S. (2009). Estimating Earthquake Magnitude with Early Arrivals: A Test Using Dynamic and Kinematic Models. *Bulletin of the Seismological Society of America*, 99(1).

Münchmeyer, J., Woollam, J., Rietbrock, A., Tilmann, F., Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D., Michelini, A., Saul, J., and Soto, H. (2022). Which Picker Fits My Data? A Quantitative Evaluation of Deep Learning Based Seismic Pickers. *Journal of Geophysical Research: Solid Earth*, 127(1):e2021JB023499.

Naif, S. and Key, K. (2018). Hikurangi Trench Regional Electromagnetic Survey to Image the Subduction Thrust.

Nakamura, Y. (2004). UrEDAS, Urgent Earthquake Detection and Alarm System, Now and Future. page 9, Vancouver, B.C., Canada.

Nakamura, Y. and Tucker, B. (1988). Japan's earthquake early warning system: Should it be imported to California? *Calif. Geol.*, 41:33–40.

Nakatani, M., Kaneshima, S., and Fukao, Y. (2000). Size-dependent microearthquake initiation inferred from high-gain and low-noise observations at Nikko district, Japan. *Journal of Geophysical Research: Solid Earth*, 105(B12):28095–28109.

Nanometrics Seismological Instruments (2013). Nanometrics Research Network.

National Centre for Seismological Research (CENAIS Cuba) (1998). Servicio Sismologico Nacional de Cuba.

National Institute for Earth Physics (NIEP Romania) (1994). Romanian Seismic Network.

National Observatory of Athens, Institute of Geodynamics, Athens (1975). National Observatory of Athens Seismic Network.

Natural Resources Canada (1975). Canadian National Seismograph Network.

Naylor, M., Serafini, F., Lindgren, F., and Main, I. G. (2023). Bayesian modeling of the temporal evolution of seismicity using the ETAS.inlabru package. *Frontiers in Applied Mathematics and Statistics*, 9.

Needham, J. and Wang, L. (1959). *Science and civilisation in China. Vol. 3, Mathematics and the sciences of the heavens and the earth*. Cambridge University Press, Cambridge.

Neely, J. S., Park, S., and Baltay, A. (2024). The Impact of Source Time Function Complexity on Stress-Drop Estimates. *Bulletin of the Seismological Society of America*.

Nielsen, S. (2007). Can Earthquake Size be Controlled by the Initial Seconds of Rupture? In Gasparini, P., Manfredi, G., and Zschau, J., editors, *Earthquake Early Warning Systems*, pages 9–20. Springer Berlin Heidelberg, Berlin, Heidelberg.

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

Noda, S. and Ellsworth, W. L. (2016). Scaling relation between earthquake magnitude and the departure time from P wave similar growth. *Geophysical Research Letters*, 43(17):9053–9060.

Noda, S. and Ellsworth, W. L. (2017). Determination of Earthquake Magnitude for Early Warning from the Time Dependence of P-Wave Amplitudes. *Bulletin of the Seismological Society of America*, 107(4):1860–1867.

Nolet, G. (1987). Seismic wave propagation and seismic tomography. In Nolet, G., editor, *Seismic Tomography: With Applications in Global Seismology and Exploration Geophysics*, pages 1–23. Springer Netherlands, Dordrecht.

Norsar (1971). NORSAR Station Network.

Northern California Earthquake Data Center (2014). Berkeley Digital Seismic Network (BDSN).

Obsercatorio Vulcanológico y Sismológico de Costa Rica, Universidad Nacional (1984). Observatorio Vulcanológico y Sismológico de Costa Rica.

Observatoire Volcanologique Et Sismologique De Guadeloupe (OVSG) (2020). Seismic, tiltmeter, extensometer, tide , magnetic and weather permanent networks on La Soufrière volcano and Guadeloupe.

Observatoire Volcanologique Et Sismologique De Martinique (OVSM) (2020). Seismic, tiltmeter, groundwater, magnetic and weather permanent networks on Montagne Pelée volcano and Martinique.

Oeser, E. (1992). Historical Earthquake Theories from ARISTOTLE to KANT. In Gutdeutsch, R., Grünthal, G., and Musson, R., editors, *Historical earthquakes in Central Europe: monographs*, number Bd. 48 in Abhandlungen der Geologischen Bundesanstalt. Geologische Bundesanstalt, Wien.

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

Ohio Geological Survey (1999). Ohio Seismic Network.

Ohnaka, M., Kuwahara, Y., Yamamoto, K., and Hirasawa, T. (1986). Dynamic Breakdown Processes and the Generating Mechanism for High-Frequency Elastic Radiation During Stick-Slip Instabilities. In *Earthquake Source Mechanics*, pages 13–24. American Geophysical Union (AGU).

Oklahoma Geological Survey (1978). Oklahoma Seismic Network.

Oklahoma Geological Survey (2018). Oklahoma Consolidated Temporary Seismic Networks.

Olson, E. L. and Allen, R. M. (2005). The deterministic nature of earthquake rupture. *Nature*, 438(7065):212–215.

Pacific Tsunami Warning Center (1965). Pacific Tsunami Warning Seismic System.

Paige, C. C. and Saunders, M. A. (1982). LSQR: An Algorithm for Sparse Linear Equations and Sparse Least Squares. *ACM Trans. Math. Softw.*, 8(1):43–71.

pandas development team, T. (2024). pandas-dev/pandas: Pandas.

Parsons, T. (2002). Global Omori law decay of triggered earthquakes: Large aftershocks outside the classical aftershock zone. *Journal of Geophysical Research: Solid Earth*, 107(B9):ESE 9–1–ESE 9–20.

Peng, C., Yang, J., Zheng, Y., Xu, Z., and Jiang, X. (2014). Early magnitude estimation for the MW7.9 Wenchuan earthquake using progressively expanded P-wave time window. *Scientific Reports*, 4(1):6770.

Peng, C.-Y., Yang, J.-S., Xue, B., Chen, Y., and Zhu, X.-Y. (2013). Research on correlation between early-warning parameters and magnitude for the Wenchuan Earthquake and its aftershocks. *Chinese Journal of Geophysics*, 56(10):3404–3415.

Penn State University (2004). Pennsylvania State Seismic Network.

Penn State University (2009). AfricaArray.

Perfettini, H. and Avouac, J. (2004). Postseismic relaxation driven by brittle creep: A possible mechanism to reconcile geodetic measurements and the decay rate of aftershocks, application to the Chi-Chi earthquake, Taiwan. *Journal of Geophysical Research: Solid Earth*, 109(B2):2003JB002488.

Persh, S. E. and Houston, H. (2004). Deep earthquake rupture histories determined by global stacking of broadband P waveforms. *Journal of Geophysical Research: Solid Earth*, 109(B4):2003JB002762.

Podolskiy, E. A. and Walter, F. (2016). Cryoseismology. *Reviews of Geophysics*, 54(4):708–758.

Prieto, G. A., Shearer, P. M., Vernon, F. L., and Kilb, D. (2004). Earthquake source scaling and self-similarity estimation from stacking P and S spectra. *Journal of Geophysical Research: Solid Earth*, 109(B8).

Pujol, J. (2003). Anelastic attenuation. In *Elastic Wave Propagation and Generation in Seismology*, pages 357–390. Cambridge University Press.

Ranjan, P. and Konstantinou, K. I. (2020). Mapping Intrinsic and Scattering Attenuation in the Southern Aegean Crust Using S Wave Envelope Inversion and Sensitivity Kernels Derived From Perturbation Theory. *Journal of Geophysical Research: Solid Earth*, 125(10):e2020JB020821.

Reasenberg, P. A. (1999). Foreshock occurrence before large earthquakes. *Journal of Geophysical Research: Solid Earth*, 104(B3):4755–4768.

Red Sismica Volcan Baru (2000). ChiriNet.

Regional Integrated Multi-Hazard Early Warning System (RIMES Thailand) (2008). Regional Integrated Multi-Hazard Early Warning System.

Reinhart, A. (2018). A Review of Self-Exciting Spatio-Temporal Point Processes and Their Applications. *Statistical Science*, 33(3).

RESIF (1995). RESIF-RLBP French Broad-band network, RESIF-RAP strong motion network and other seismic stations in metropolitan France.

Reyners, M. and Bannister, S. (2007). Earthquakes triggered by slow slip at the plate interface in the Hikurangi subduction zone, New Zealand. *Geophysical Research Letters*, 34(14):2007GL030511.

Robinson, A. (2012). *Earthquake: Nature and Culture*. Reaktion Books, London, 1. ed edition.

Rollo, C. (2024). cmcrameri.

Rothman, D. H. (1986). Automatic estimation of large residual statics corrections. *GEOPHYSICS*, 51(2):332–346.

Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., and Johnson, P. A. (2017). Machine Learning Predicts Laboratory Earthquakes. *Geophysical Research Letters*, 44(18):9276–9282.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.

Ruff, L. J. (1989). Multi-trace deconvolution with unknown trace scale factors: Omnilinear inversion of P and S waves for source time functions. *Geophysical Research Letters*, 16(9):1043–1046.

Rydelek, P. and Horiuchi, S. (2006). Earth science: Is earthquake rupture deterministic? *Nature*, 442(7100):10–11.

Rydelek, P., Wu, C., and Horiuchi, S. (2007). Comment on "Earthquake magnitude estimation from peak amplitudes of very early seismic signals on strong motion records" by Aldo Zollo, Maria Lancieri, and Stefan Nielsen. *Geophysical Research Letters*, 34(20).

Sacks, I. S., Linde, A. T., Suyehiro, S., and Snoke, J. A. (1978). Slow earthquakes and stress redistribution. *Nature*, 275(5681):599–602.

Sambridge, M. (1999). Geophysical inversion with a neighbourhood algorithm—I. Searching a parameter space. *Geophysical Journal International*, 138(2):479–494.

Sato, K. and Mori, J. (2006). Relationship between rupture process complexity and earthquake size. *Journal of Geophysical Research: Solid Earth*, 111(B5).

Sato, T. and Hirasawa, T. (1973). Body wave spectra from propagating shear cracks. *Journal of Physics of the Earth*, 21(4):415–431.

Sato, T. and Kanamori, H. (1999). Beginning of earthquakes modeled with the Griffith's fracture criterion. *Bulletin of the Seismological Society of America*, 89(1):80–93.

Scholz, C. H. (1968). The frequency-magnitude relation of microfracturing in rock and its relation to earthquakes. *Bulletin of the Seismological Society of America*, 58(1):399–415.

Scholz, C. H. (1982). Scaling laws for large earthquakes: Consequences for physical models. *Bulletin of the Seismological Society of America*, 72(1):1–14.

Seif, S., Zechar, J. D., Mignan, A., Nandan, S., and Wiemer, S. (2018). Foreshocks and Their Potential Deviation from General Seismicity. *Bulletin of the Seismological Society of America*, 109(1):1–18.

Serdar Kuyuk, H., Allen, R. M., Brown, H., Hellweg, M., Henson, I., and Neuhauser, D. (2014). Designing a Network-Based Earthquake Early Warning Algorithm for California: ElarmS-2. *Bulletin of the Seismological Society of America*, 104(1):162–173.

Shao, G., Li, X., Ji, C., and Maeda, T. (2011). Focal mechanism and slip history of the 2011 M w 9.1 off the Pacific coast of Tohoku Earthquake, constrained with teleseismic body and surface waves. *Earth, Planets and Space*, 63(7):559–564.

Shearer, P. M. (2012). Self-similar earthquake triggering, Båth's law, and foreshock/aftershock magnitudes: Simulations, theory, and results for southern California. *Journal of Geophysical Research: Solid Earth*, 117(6):1–15.

Shearer, P. M. (2013). Reply to comment by S. Hainzl on "self-similar earthquake triggering, Båth's Law, and foreshock/aftershock magnitudes: Simulations, theory and results for southern California". *Journal of Geophysical Research: Solid Earth*, 118(3):1192.

Shieh, J.-T., Wu, Y.-M., and Allen, R. M. (2008). A comparison of \tau_c and \tau_p^{\max} for magnitude estimation in earthquake early warning. *Geophysical Research Letters*, 35(20):L20301.

Sigloch, K. and Nolet, G. (2006). Measuring finite-frequency body-wave amplitudes and traveltimes. *Geophysical Journal International*, 167(1):271–287.

Singh, S. K., Ordaz, M., Mikumo, T., Pacheco, J., Valdés, C., and Mandal, P. (1998). Implications of a composite source model and seismic-wave attenuation for the observed simplicity of small earthquakes and reported duration of earthquake initiation phase. *Bulletin of the Seismological Society of America*, 88(5):1171–1181.

Smits, G. (2013). Why the Earth Shakes. In Smits, G., editor, *Seismic Japan: The Long History and Continuing Legacy of the Ansei Edo Earthquake*, page 0. University of Hawai'i Press.

Song, J. Y. and McLaskey, G. C. (2024). Laboratory Earthquake Ruptures Contained by Velocity Strengthening Fault Patches. *Journal of Geophysical Research: Solid Earth*, 129(4):e2023JB028509.

Sornette, D. and Werner, M. J. (2005). Apparent clustering and apparent background earthquakes biased by undetected seismicity. *Journal of Geophysical Research: Solid Earth*, 110(B9).

Steacy, S. J. and McCloskey, J. J. (1998). What controls an earthquake's size? Results from a heterogeneous cellular automaton. *Geophysical Journal International*, 133(1):F11–F14.

Stein, R. S. (2005). Conversations. *Scientific American*.

Stähler, S. C. and Sigloch, K. (2014). Fully probabilistic seismic source inversion – Part 1: Efficient parameterisation. *Solid Earth*, 5(2):1055–1069.

Tanioka, Y. and Ruff, L. J. (1997). Source Time Functions. *Seismological Research Letters*, 68(3):386–400.

Tape, C., Holtkamp, S., Silwal, V., Hawthorne, J. C., Kaneko, Y., Ampuero, J. P., Ji, C., Ruppert, N., Smith, K., and West, M. E. (2018). Earthquake nucleation and fault slip complexity in the lower crust of central Alaska. *Nature Geoscience*, 11(7):536–541.

Tape, W. and Tape, C. (2012). A geometric setting for moment tensors: A geometric setting for moment tensors. *Geophysical Journal International*, 190(1):476–498.

Tape, W. and Tape, C. (2015). A uniform parametrization of moment tensors. *Geophysical Journal International*, 202(3):2074–2081.

Technological Educational Institute of Crete (2006). Seismological Network of Crete.

Tocheport, A., Rivera, L., and Chevrot, S. (2007). A systematic study of source time functions and moment tensors of intermediate and deep earthquakes. *Journal of Geophysical Research: Solid Earth*, 112(B7).

Totten, E. (2021). *Finite-frequency tomography using regional P-waves, with application to the western Pacific upper mantle*. phd, University of Oxford / University of Oxford.

Trugman, D. T., Page, M. T., Minson, S. E., and Cochran, E. S. (2019). Peak Ground Displacement Saturates Exactly When Expected: Implications for Earthquake Early Warning. *Journal of Geophysical Research: Solid Earth*, 124(5):4642–4653.

Trugman, D. T. and Ross, Z. E. (2019). Pervasive Foreshock Activity Across Southern California. *Geophysical Research Letters*, 46(15):8772–8781.

Tsang, L. L. H., Allen, R. M., and Wurman, G. (2007). Magnitude scaling relations from P-waves in southern California. *Geophysical Research Letters*, 34(19).

Umeda, Y. (1990). High-amplitude seismic waves radiated from the bright spot of an earthquake. *Tectonophysics*, 175(1-3):81–92.

Umeda, Y. (1992). The bright spot of an earthquake. *Tectonophysics*, 211(1-4):13–22.

Umeda, Y., Yamashita, T., Tada, T., and Kame, N. (1996). Possible mechanisms of dynamic nucleation and arresting of shallow earthquake faulting. *Tectonophysics*, 261(1):179–192.

United States Geological Survey (2025). USGS Finite Fault Models.

Universidad de Chile (2012). Red Sismologica Nacional.

University of Athens (2008). Hellenic Seismological Network, University of Athens, Seismological Laboratory.

University of Bergen (1982). University of Bergen Seismic Network.

University of Genoa (1967). Regional Seismic Network of North Western Italy.

University of Nevada, Reno (1971). Nevada Seismic Network.

University of Nevada, Reno (1992). Southern Great Basin Network.

University of Oregon (2002). Pacific Northwest Seismic Network - University of Oregon.

University of Patras (2000). University of Patras, Seismological Laboratory.

University of Puerto Rico (1986). Puerto Rico Seismic Network & Puerto Rico Strong Motion Program.

University of Washington (1963). Pacific Northwest Seismic Network - University of Washington.

U.S. Geological Survey (2016). U.S. Geological Survey Networks.

USGS Hawaiian Volcano Observatory (HVO) (1956). Hawaiian Volcano Observatory Network.

USGS Menlo Park (1966). USGS Northern California Seismic Network.

Utsu, T., Ogata, Y., S, R., 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.

Vallée, M. (2013). Source time function properties indicate a strain drop independent of earthquake depth and magnitude. *Nature Communications*, 4(1):2606.

Vallée, M., Charléty, J., Ferreira, A. M. G., Delouis, B., and Vergoz, J. (2011). SCARDEC: a new technique for the rapid determination of seismic moment magnitude, focal mechanism and source time functions for large earthquakes using body-wave deconvolution: Wave deconvolution and earthquake parameters. *Geophysical Journal International*, 184(1):338–358.

Vallée, M. and Douet, V. (2016). A new database of source time functions (STFs) extracted from the SCARDEC method. *Physics of the Earth and Planetary Interiors*, 257:149–157.

van den Ende, M. P. A. and Ampuero, J.-P. (2020). On the Statistical Significance of Foreshock Sequences in Southern California. *Geophysical Research Letters*, 47(3):e2019GL086224.

Van Driel, M., Krischer, L., Stähler, S. C., Hosseini, K., and Nissen-Meyer, T. (2015). Instaseis: instant global seismograms based on a broadband waveform database. *Solid Earth*, 6(2):701–717.

VanDecar, J. C. and Crosson, R. S. (1990). Determination of teleseismic relative phase arrival times using multi-channel cross-correlation and least squares. *Bulletin of the Seismological Society of America*, 80(1):150–169.

Verdecchia, A., Cochran, E. S., and Harrington, R. M. (2021). Fluid-Earthquake and Earthquake-Earthquake Interactions in Southern Kansas, USA. *Journal of Geophysical Research: Solid Earth*, 126(3):e2020JB020384.

Vernon, F. (1982). ANZA Regional Network.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, \., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors, Vijaykumar, A., Bardelli, A. P., Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C. N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D. A., Hagen, D. R., Pasechnik, D. V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G. A., Ingold, G.-L., Allen, G. E., Lee, G. R., Audren, H., Probst, I., Dietrich, J. P., Silterra, J., Webber, J. T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J. L., de Miranda Cardoso, J. V., Reimer, J., Harrington, J., Rodríguez, J. L. C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N. J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P. A., Lee, P., McGibbon, R. T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T. J., Robitaille, T. P., Spura, T., Jones, T. R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y. O., and Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272.

Wallace, L. M., Barnes, P., Beavan, J., Van Dissen, R., Litchfield, N., Mountjoy, J., Langridge, R., Lamarche, G., and Pondard, N. (2012). The kinematics of a transition from subduction to strike-slip: An example from the central New Zealand plate boundary. *Journal of Geophysical Research: Solid Earth*, 117(2).

Wallace, L. M., Webb, S. C., Ito, Y., Mochizuki, K., Hino, R., Henrys, S., Schwartz, S. Y., and Sheehan, A. F. (2016). Slow slip near the trench at the Hikurangi subduction zone, New Zealand. *Science*, 352(6286):701–704.

Wang, Z., Kato, T., Zhou, X., and Fukuda, J. (2016). Source process with heterogeneous rupture velocity for the 2011 Tohoku-Oki earthquake based on 1-Hz GPS data. *Earth, Planets and Space*, 68(1):193.

Wesnousky, S. G. (2006). Predicting the endpoints of earthquake ruptures. *Nature*, 444(7117):358–360.

Wessel, P., Luis, J. F., Uieda, L., Scharroo, R., Wobbe, F., Smith, W. H. F., and Tian, D. (2019). The Generic Mapping Tools Version 6. *Geochemistry, Geophysics, Geosystems*, 20(11):5556–5564.

Withers, M., Aster, R., Young, C., Beiriger, J., Harris, M., Moore, S., and Trujillo, J. (1998). A comparison of select trigger algorithms for automated global seismic phase and event detection. *Bulletin of the Seismological Society of America*, 88(1):95–106.

Wolfe, C. J. (2006). On the Properties of Predominant-Period Estimators for Earthquake Early Warning. *Bulletin of the Seismological Society of America*, 96(5):1961–1965.

Woollam, J., Münchmeyer, J., Tilmann, F., Rietbrock, A., Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D., Michelini, A., Saul, J., and Soto, H. (2022). SeisBench—A Toolbox for Machine Learning in Seismology. *Seismological Research Letters*, 93(3):1695–1709.

Wu, Y.-M. and Kanamori, H. (2005a). Experiment on an Onsite Early Warning Method for the Taiwan Early Warning System. *Bulletin of the Seismological Society of America*, 95(1):347–353.

Wu, Y.-M. and Kanamori, H. (2005b). Rapid Assessment of Damage Potential of Earthquakes in Taiwan from the Beginning of P Waves. *Bulletin of the Seismological Society of America*, 95(3):1181–1185.

Wu, Y.-M. and Kanamori, H. (2008a). Development of an Earthquake Early Warning System Using Real-Time Strong Motion Signals. *Sensors*, 8(1):1–9.

Wu, Y.-M. and Kanamori, H. (2008b). Exploring the feasibility of on-site earthquake early warning using close-in records of the 2007 Noto Hanto earthquake. *Earth, Planets and Space*, 60(2):155–160.

Wu, Y.-M., Kanamori, H., Allen, R. M., and Hauksson, E. (2007). Determination of earthquake early warning parameters, τ_c and p_d, for southern California. *Geophysical Journal International*, 170(2):711–717.

Wu, Y.-M., Yen, H.-Y., Zhao, L., Huang, B.-S., and Liang, W.-T. (2006). Magnitude determination using initial P waves: A single-station approach. *Geophysical Research Letters*, 33(5):L05306.

Wu, Y.-M. and Zhao, L. (2006). Magnitude estimation using the first three seconds P-wave amplitude in earthquake early warning. *Geophysical Research Letters*, 33(16).

Wurman, G., Allen, R. M., and Lombard, P. (2007). Toward earthquake early warning in northern California. *Journal of Geophysical Research*, 112(B8):B08311.

Xia, K., Chalivendra, V. B., and Rosakis, A. J. (2006). Observing ideal "self-similar" crack growth in experiments. *Engineering Fracture Mechanics*, 73(18):2748–2755.

Yao, D., Huang, Y., Peng, Z., and Castro, R. R. (2020). Detailed Investigation of the Foreshock Sequence of the 2010 M $_{w}$ 7.2 El Mayor-Cucapah Earthquake. *Journal of Geophysical Research: Solid Earth*, 125(6).

Ye, L., Lay, T., Kanamori, H., and Rivera, L. (2016a). Rupture characteristics of major and great (M_w7+) megathrust earthquakes from 1990 to 2015: 1. Source parameter scaling relationships. *Journal of Geophysical Research: Solid Earth*, 121(2):826–844.

Ye, L., Lay, T., Kanamori, H., and Rivera, L. (2016b). Rupture characteristics of major and great (M_w7+) megathrust earthquakes from 1990 to 2015: 2. Depth dependence. *Journal of Geophysical Research: Solid Earth*, 121(2):845–863.

Yin, J., Li, Z., and Denolle, M. A. (2021). Source Time Function Clustering Reveals Patterns in Earthquake Dynamics. *Seismological Research Letters*.

Yin, J., Soto, M. A., Ramírez, J., Kamalov, V., Zhu, W., Husker, A., and Zhan, Z. (2023). Real-Data Testing of Distributed Acoustic Sensing for Offshore Earthquake Early Warning. *The Seismic Record*, 3(4):269–277.

Yoon, C. E., Yoshimitsu, N., Ellsworth, W. L., and Beroza, G. C. (2019). Foreshocks and Mainshock Nucleation of the 1999 M_W 7.1 Hector Mine, California, Earthquake. *Journal of Geophysical Research: Solid Earth*, 124(2):1569–1582.

ZAMG - Zentralanstalt für Meterologie und Geodynamik (1987). Austrian Seismic Network.

Zoeppritz, K. (1919). VII b. Uber Reflexion und Durchgang seismischer Wellen durch Unstetigkeitsflaechen. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*.

Zollo, A., Lancieri, M., and Nielsen, S. (2006). Earthquake magnitude estimation from peak amplitudes of very early seismic signals on strong motion records. *Geophysical Research Letters*, 33(23):2006GL027795.

Zollo, A., Lancieri, M., and Nielsen, S. (2007). Reply to comment by P. Rydelek et al. on "Earthquake magnitude estimation from peak amplitudes of very early seismic signals on strong motion records". *Geophysical Research Letters*, 34(20):2007GL030560.