Improving Uncertainty Quantification and Visualization for Spatiotemporal Earthquake Rate Models for the Pacific Northwest

Max Schneider

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Peter Guttorp, Chair

Yen-Chi Chen, Chair

Andrew Michael

Program Authorized to Offer Degree:

Statistics

University of Washington

Abstract

Improving Uncertainty Quantification and Visualization for Spatiotemporal Earthquake Rate Models for the Pacific Northwest

Max Schneider

Co-Chairs of the Supervisory Committee:
Professor Emeritus Peter Guttorp
Department of Statistics

Assistant Professor Yen-Chi Chen
Department of Statistics

The Pacific Northwest (PNW) has substantial earthquake risk, both due to the offshore Cascadia megathrust fault but also other fault systems that produce earthquakes under the region's population centers. Forecasts of aftershocks following large earthquakes are thus highly desirable and require statistical models of a catalog of the PNW's past earthquakes and aftershock sequences. This is complicated by the fact that the PNW contains multiple tectonic regimes hypothesized to have different aftershock dynamics as well as two types of earthquake clustering (aftershock sequences and swarms). The Epidemic-Type Aftershock Sequence (ETAS) model is a top-performing spatiotemporal point process model which describes the dynamics of earthquakes and aftershocks in a seismic region using a set of parameters. Typically, maximum likelihood estimation is used to fit ETAS to an earthquake catalog; however, the ETAS likelihood suffers from flatness near its optima, parameter correlation and numerical instability, making likelihood-based estimates less reliable. We present a Bayesian procedure for ETAS estimation, such that parameter estimates and uncertainty can be robustly quantified, even for small and complex catalogs like the PNW. The proce-

dure is conditional on knowing which earthquakes triggered which aftershocks; this latent structure and the ETAS parameters are estimated iteratively. The procedure uses a Gibbs sampler to conditionally estimate the posterior distributions of each part of the model. We simulate several synthetic catalogs and test the modelling procedure, showing well-mixed posterior distributions centered on true parameter values. We also use the procedure to model the continental PNW, using a new catalog formed by algorthmically combining US and Canadian data sources and then, identifying and removing earthquake swarms. While MLEs are unstable and depend on both the optimization procedure and its initial values, Bayesian estimates are insensitive to these choices. Bayesian estimates also fit the catalog better than do MLEs. We use the Bayesian method to quantify the uncertainty in ETAS estimates when including swarms in the model or modelling across different tectonic regimes, as well as from catalog measurement error.

Seismicity rate estimates and the earthquake forecasts they yield vary spatially and are usually represented as heat maps. While the visualization literature suggests that displaying forecast uncertainty improves understanding in users of forecast maps, research on uncertainty visualization (UV) is missing from earthquake science. In a pre-registered online experiment, we test the effectiveness of three UV techniques for displaying uncertainty in aftershock forecasts. Participants completed two map-reading tasks and a comparative judgment task, which demonstrated how successful a visualization was in reaching two key communication goals: indicating where many aftershocks and no aftershocks are likely (sure bets) and where the forecast is low but the uncertainty is high enough to imply potential risk (surprises). All visualizations performed equally well in the goal of communicating sure bet situations. But the visualization mapping the lower and upper bounds of an uncertainty interval was substantially better than the other map designs at communicating potential surprises. We discuss the implications of these experimental results for the communication

of uncertainty in aftershock forecast maps.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I would be remiss to not acknowledge the great deal of support I have benefited from on this PhD journey. Peter Guttorp has sharpened my instincts and greatly expanded my toolkit; he has advised me, believed in me and inspired me since my first day at the University of Washington. Yen-Chi Chen has been a consistent source of encouragement and support; I am grateful for his willingness to directly help solve research problems. Andy Michael has been incredibly generous in his guidance (on problems small and large) and graciously opened doors of huge opportunity. Joan Gomberg has provided critical suggestions and direction for my work and I thank Adrian Dobra for his support at a critical time. I am grateful for Harold Tobin's service as GSR as well as his providing of resources at the Pacific Northwest Seismic Network.

All three of my dissertation projects have relied on interdisciplinary collaborations and I am grateful to my collaborators for shaping my research lens and trajectory. I have been very lucky to work with the USGS Operational Aftershock Forecasting team. In additional to Andy Michael, Jeanne Hardebeck, Nicholas van der Elst, Morgan Page, Michael Barall, Andrea Llenos, and Sara McBride have made critical inputs, helping me think through problems and keeping me motivated. I can't think of a team I would more want to join and can't wait for our next years of working together. A heartfelt thank you to Nadine Fleischhut and Michelle McDowell for their commitment to our collaboration, their opening up of their research field and their patience and comraderie with me as I developed the skills necessary for this work. I am sincerely grateful for the research visits they organized, and their support outside these stays. Ashley Steel has been a constructive advisor, guiding me through this

interdisciplinary journey. I have gotten tremendous assistance from the Pacific Northwest Seismic Network and am grateful to Amy Wright, Paul Bodin, Renate Hartog, Steve Malone and Camille Brillon (Geological Survey of Canada) who have provided essential inputs for this dissertation. I extend great credit to Hank Flury whose tireless assistance made possible the meticulous work in Chapter 2.

The two primary sources of funding for my PhD work were the M9 project at UW (National Science Foundation) and an External Grant from the USGS Earthquake Hazards Program (Award G20AP00061). I am deeply grateful for this support.

I would also like to extend sincere gratitude to Alison Duvall, Erin Wirth, Aila Särkkä, Thordis Thorarinsdottir, Claudio Heinrich, Fabrice Cotton, Christian Molkenthin, and Sebastian Hainzl for their engaged collaboration or help with research problem-solving or funding. I would not have made it through this PhD program without my cohort and colleagues at UW Statistics; thank you for creating an environment conducive to learning, growth, and plenty of laughs.

Finally, I would not be writing this dissertation without my family and friends, whose patience, motivation and trust in me have buoyed me while navigating completing a dissertation across several timezones during a pandemic. I hope to be able to repay my gratitude for the rest of my years.

DEDICATION

To my parents,

Svetlana and Dmitry.

25 years ago, they moved their young family

Across an ocean and two continents to

Earthquake Country.

They couldn't foresee the events that this would trigger,

But I wouldn't be standing without them.

Я всегда буду благодарен вам обоим.

Chapter 1

INTRODUCTION

Oh, these little earthquakes;
Here we go again.
These little earthquakes,
Doesn't take much to rip us into pieces.

"Little Earthquakes," Tori Amos (1992)

Earthquakes are capable of resulting in incredible societal costs in the matter of minutes (Jaiswal et al., 2017; Bilham, 2009), but they stand out from most natural disasters in another way. In the time soon after an earthquake occurs, more earthquakes are likely to happen. While the initial earthquake (the mainshock) is not currently predictable by reproducible scientific means, the subsequent earthquakes (the aftershocks) follow seismological laws dictating how many should occur, as well as how their times and locations decay relative to the mainshock. These laws are governed by a set of parameters whose values are specific to each seismic region. For decades, these laws have been incorporated into statistical models of earthquake datasets. These models are now being used by both researchers and government authorities for various applications related to emergency response and long-term planning. While no statistical model can accomplish short-term earthquake prediction, these models can successfully project earthquake patterns over longer periods (on the years to decades scale) as well as offer short-term probabilistic forecasts for aftershocks following large earthquakes. The statistical estimation of the model parameters is thus of utmost import, as these estimates guide emergency response during aftershock sequences as well as long-term

earthquake planning.

The Pacific Northwest (PNW) of North America is a seismic region of particular importance because it has fault lines and seismic zones under urban areas. Aftershocks can put people at additional risk of harm, for example, by destroying buildings already destabilized by the mainshock. The last major earthquake (as of the time of this writing) in the PNW occurred in 2001, near Nisqually, WA, with a magnitude of 6.8. While this was felt widely (Chang and Falit-Baiamonte, 2002), killing 1 person and injuring around 400 people, it triggered less than five aftershocks, very few for an earthquake of its size. Meanwhile, five years prior, a magnitude 5.6 earthquake struck near Duvall, WA, about 100.5 km away. This earthquake triggered dozens of aftershocks, an order of magnitude more at least. Understanding whether a difference in aftershock dynamics exists between different regions of the PNW, and whether it changes parameters in earthquake models, can contribute to reducing the PNW's seismic risk.

Statistical models of earthquakes have been developed for over 100 years (Omori, 1894) and are successful at explaining the patterns in earthquake data over long-term scales (Schorlemmer et al., 2018). Like all statistical models, they come with aleatory uncertainty, due to the inherent stochasticity and complexity of the phenomenon. Challenges in earthquake measurement (see Section 1.2) contribute epistemic uncertainty, as does the capacity of earthquake models to describe earthquake dynamics across seismic regions. Furthermore, a single seismic region may contain substantial variability. For example, it may be composed of different tectonic *regimes*, or tectonic environments that host disparate faulting types, or different forms of clustering, which the specified model may not be built to accommodate. As described in Section 1.2, this is particularly true in the PNW. Disentangling these sources of variability and how they affect parameter estimates and their uncertainty is critical for reliable model results that can lead, for example, to trustworthy aftershock forecasts.

Aftershock forecasts have been used for decades by many professional communities, from

emergency managers to civil engineers. Emergency managers have expressed in focus groups that aftershock forecasts are useful in their decison-making processes (Gomberg and Jakobitz, 2013). In recent years, government agencies such as the United States Geological Survey (Michael et al., 2020), Italy's National Institute of Geophysics and Volcanology (Marzocchi et al., 2014), New Zealand's GNS Science (Becker et al., 2020), as well as the Japanese Meteorological Agency have released public aftershock forecasts following large earthquakes. Media coverage has dispersed aftershock forecasts to a wide and general audience. Forecasts have been communicated with tables showing how many aftershocks are expected for different time periods and magnitude intervals; these usually show intervals of expected aftershock counts, thus conveying uncertainty, though no spatial information is conveyed. Forecast maps provide the expected number of aftershocks at each location, for example, but no aftershock forecast maps have yet communicated the uncertainty in these estimates. Uncertainty visualization for aftershock forecast maps remains missing, both from the research literature and from common practice, without any clear guidelines on what is effective.

In this introduction, we provide some more detail on the challenges involved in modelling earthquakes in the PNW and visualizing uncertainty in their corresponding aftershock forecast maps. This lays the groundwork for the methdological and applied contributions we make to these research topics in the subsequent chapters.

1.1 Background

We explain here some key basics about earthquake measurement and modeling. Earthquakes occur due to abrupt motions of two neighboring tectonic plates at their border (*fault*); when the movement of the plates overcomes the frictional resistance holding the plates together, (much of) the energy is released through the plate, which finally causes the surface of the Earth to shake. A variety of different fault zones exist that have different patterns of earthquake and aftershock occurrence. The point at which the rupture of the plates past each

other begins is called the *hypocenter* of the earthquake. Each earthquake consists of several distinct seismic *phases*, known as the P- and S-waves.

Earthquakes are detected and measured by a seismic *network*, which operates a number of seismic *stations*. Networks are assigned a geographic area for which they are *authoritative*, or their measurements are usually preferred over other networks, though they may also measure earthquakes that occurred outside their authoritative zones. Observations on earthquakes typically include their *origin times* (the time at which the rupture began), their *epicentral locations* (the location of their epicenter, or point on the earth's surface vertically above the hypocenter), their *depths* (the distance from the earth's surface at which the hypocenter occurs), and their *magnitudes* (which can be measured on multiple scales, usually logarithmically). Many other variables exist to characterize an earthquake, but these are the only four used in this dissertation. An *earthquake catalog* is the collection of data for these variables for all earthquakes observed by a seismic network. A catalog is *complete* for a specific time period, spatial zone, and magnitude, when it contains all the earthquakes that occurred within that period in that zone, and above that magnitude level.

We refer to *seismicity* to mean the set of earthquakes that are observed over a period of time, which can be characterized by the region's authoritative network's earthquake catalog. As with the observation of any natural process, catalog data comes with limitations. A seismic station contains a *seismograph*, an instrument that measures the time and wave characteristics of any shaking of the ground. When an earthquake rupture occurs, this sends seismic waves through the tectonic plate which are eventually recorded by the seismograph. In order to obtain an estimate for the catalog variables for any earthquake, at least three stations must observe it. By comparing the origin times of different seismic phases across several stations and to predictions from a model of seismic velocities, an earthquake's time, epicentral location and depth can be triangulated. The number of stations contributing to this operation and their relative locations to the expected hypocenter of the earthquake de-

termines the level of *measurement error* in these variables. Though time errors are generally tiny (on the order of seconds), location errors can be high when stations are located at unfavorable geometries to the earthquake, or only few stations contribute to the triangulation. Earthquake magnitudes are estimated individually by each station, through formulas related to the time differences between different seismic phases; these are then summarized statistically across stations that reported the earthquake. Measurement error for magnitudes is dictated by the ability of algorithms and human analysts to detect seismic phases in the data as well as the number of stations involved (Husen and Hardebeck, 2010). A network will evolve over time, installing and de-installing stations across its authoritative region. We refer to a zone's *station coverage* as the presence of sufficient stations there to detect and measure earthquakes, which can vary over time and space.

Due to the urgency in earthquake information, networks release earthquake data as soon as they have preliminary solutions to the measurement problems described above. However, subsequent analysis continues, especially for large earthquakes.[1] Furthermore, seismic networks may change their operational procedures or algorithms, which may require a reanalysis of previous earthquakes to maintain consistency. Earthquake catalogs are used for various applications, but we focus on modelling them using statistical seismicity models. As mentioned in the previous section, these models are used for both long-term needs (e.g., seismic hazard assessment for the setting of a region's seismic building codes) and short-term needs (e.g., forecasting aftershocks following a large earthquake). Forecasts can be communicated to the public and different user group through *forecast products*, which can be textual, tabular, or graphical (or some combination thereof). Graphical visualizations can be especially useful in communicating about natural hazards across different user groups (Pang, 2008); examples include line plots that show how many aftershocks are expected over time and maps that show where aftershocks are expected for a given period.

[1]This is why earthquake magnitudes may change in the days following an earthquake (especially larger ones).

1.2 Data challenges in modelling PNW seismicity

Statistical seismicity modelling for the Pacific Northwest is complicated by its complex tectonic environment. The Juan de Fuca plate subducts under the North American continental plate and the resulting fault zone is capable of generating massive megathrust events, the last of which occurred in 1700. The subducting slab itself produces deep, intraslab earthquakes which are hypothesized to produce few aftershocks. For example, the 2001 Nisqually earthquake was at a depth of 51.8 km, placing it in the *deep regime*. Numerous shallow faults also exist in the crust of the North American plate which produce crustal earthquakes often smaller in magnitude. These earthquakes are also thought to have more productive aftershocks than intraslab earthquakes. For example, the 1996 Duvall earthquake was at a depth of 3.8 km, and thus from the *crustal regime*. Though there is a lore about how aftershock dynamics differ between the two regimes (Gomberg and Bodin, 2021), it has not yet been rigorously quantified.

Furthermore, the region is known to produce earthquake swarms, which are spatiotemporal clusters of earthquakes that have different dynamics from aftershock sequences. Most statistical models for aftershocks cannot accommodate swarm data and so detection of swarms in the PNW is important, not just for model assumptions, but for understanding this cluster type as a source of uncertainty in any aftershock model built with it. However, there are no automated or rigorous methods for detecting swarms in catalogs; this posed a challenge, which we met with several algorithmic approaches.

Another data challenge for the PNW is the fact that it is split between two countries, the United States of America and Canada. Authoritative earthquake catalogs for the two countries are run by different seismic networks, with different operational procedures. Each network also has their own history of station coverage, which leads to changes in earthquake detection and measurement error over space and time. Merging these catalogs is a necessary challenge to overcome to be able to model PNW seismicity.

1.3 Statistical estimation challenges in modelling PNW seismicity

A widespread statistical seismicity model is the Epidemic Type Afteshock Sequence (ETAS) model, which is built on three fundamental seimsological laws (described in detail in Chapter 3). The ETAS model considers the catalog as a combination of earthquakes that were triggered by previous earthquakes (aftershocks) and those that were not (background events). The parameters that govern aftershock dynamics are then estimated from the data. As mentioned above, these parameters are used for seismic risk reduction, thus making it critical to both quantify them and their uncertainty for each seismic region.

The traditional method of inference for ETAS parameters is by maximizing the ETAS likelihood function. As is common for multi-dimensional likelihood functions (depending on the model parameterization, there can be seven or more parameters that the likelihood is a function of), numerical optimization procedures are necessary. Numerical maximization of the ETAS likelihood comes with many challenges. The likelihood function is known to be quite flat in the vicinity of the values that maximize it for several parameters. It is furthermore multimodal due to the correlation between parameters, both within each seismological law (each law has two parameters), but also across laws. Numerical procedures are often unstable in these settings and depend on the chosen procedure and the initial values that it is started off with. The uncertainty around ETAS maximum likelihood estimates is usually estimated based on the inverse Hessian of the likelihood function, by asymptotic reasoning, which smaller catalogs cannot satisfy. This also requires further numerical operations and can also be unstable. This is is a well-documented problem and numerous authors have commented or studied the biases and other difficulties that likelihood-based estimation presents for the ETAS model (Seif et al., 2017; Harte, 2018; Wang et al., 2010; Schoenberg, 2013; Veen and Schoenberg, 2008).

We address these challenges by building a new estimation method for the spatiotemporal ETAS model. This method relies on a different likelihood function, conditional on additional

information about how the catalog can be split into background and aftershock clusters. This partitioning is unknown but can be estimated from the data, given a set of ETAS parameters. We thus alternate between estimating this partitioning and estimating the ETAS parameters. Inference with this likelihood is conducted with a Bayesian paradigm that removes the need to numerically optimize it. Bayesian parameter estimates are stable and the uncertainty can be easily quantified directly from the resulting posterior distributions for each parameter. We show in Chapter 3 that we can estimate ETAS parameters for the PNW, together with their uncertainty.

The Bayesian approach also offers several benefits missing from likelihood-based estimation. We can set prior distributions corresponding to specific beliefs about the PNW's aftershock dynamics. We can also incorporate the measurement error we have in the data, as well as other sources of uncertainty (e.g., presence of swarms, different tectonic regimes). Quantifying how different sources of uncertainty affect ETAS parameters for the PNW contributes both to robust inference about the region's aftershock dynamics and how this inference changes with different catalog choices.

1.4 Statistical visualization challenges in aftershock forecasts

Given the various sources of epistemic uncertainty and underlying high aleatory uncertainty of seismicity modelling, forecasts made from these models will have non-trivial levels of uncertainty. Importantly, we can expect uncertainty to not be identical at all points in space; there is likely more that aftershock forecasts can say with certainty near the mainshock than further away from it. While it can be challenging to effectively communicate uncertainty on a forecast map, repeated cognitive studies have found that non-specialist users will infer the level of uncertainty if it is not depicted (Morss et al., 2010; Ash et al., 2014; Mulder et al., 2017). Forecast maps without uncertainty can lead to worse judgment and decision-making than if the model-estimated uncertainty was optimally conveyed.

Forecast maps are often displayed as heat maps and uncertainty can be visualized on these in numerous ways. Previous authors have used transparency, symbols, shading or other patterns to visualize uncertainty on heat maps. However, no previous authors have proposed uncertainty visualizations for aftershock forecast maps. We developed three visualizations, using empirically-validated approaches from the visualization literature as well as recent developments for other operational forecasts.

Despite a plethora of available visualization options, there has never been research done on which are effective for aftershock forecast maps. The first challenge here was identifying what it means to be effective. We began by interviewing emergency managers, a key user group of aftershock forecasts, to understand how they use forecasts for natural hazards – and how forecast uncertainty influences this. The interviews spotlighted two key communication goals that an effective uncertainty visualization should satisfy.

Task-based experiments are often used to offer evidence on the effectiveness of risk communications about natural hazards. But previous work for spatial uncertainty visualization relied on tasks and experimental designs where effects of uncertainty visualizations could not be disentangled from other features of the maps or the task. Furthermore, evaluation tasks were often disconnected from a specific communication goal that could be broadly relevant, focusing instead on specialist needs that may not be translatable to a public audience. Devising targeted and broadly applicable tasks for such an experiment was a challenge; we focused our tasks on communication goals informed by emergency managers, but with a broader applicability.

Another challenge is to design a rigorous experiment in which the effect of visualizations can be carved out by the task responses. We developed a controlled online experiment, whose results can inform forecast product designers on which uncertainty visualizations are effective and for which goals.

1.5 Outline

The rest of this dissertation is outlined as follows. In Chapter 2, we collect a new catalog for the PNW from three data sources, resolving discrepancies and removing duplicates between them; we identify complete catalogs for the northern and southern parts of the PNW. We also identify clusters in the data with a window-based approach (showing later that we can get similar results with a semi-parameteric method), for labelling as swarms or sequences. In Chapter 3, we build Bayesian inference for the ETAS model, showing that it ameliorates the oft-reported issues with maximum likelihood estimation for this model. We use it to model the PNW catalog, identifying differences in aftershock dynamics between the north and south regions, the deep and crustal tectonic regimes, and the catalog with and without swarms. In Chapter 4, we develop three uncertainty visualizations for aftershock forecast maps and test them in a task-based experiment. Unlike previous experiments, our task is built around user-generated communication goals, and our results show that only one of the visualizations can accomplish both goals. In Chapter 5, we make overall conclusions and pave the way for future research on these topics.

Chapter 2

EARTHQUAKE CATALOG PROCESSING AND SWARM IDENTIFICATION FOR THE PACIFIC NORTHWEST

A fault runs through the valley,
It's long as it is mean,
It starts to tear into the earth,
And gets all in between.
It shakes up through the mountains,
It shakes down to the sea,
It shakes up to the volcano,
And then it starts to scream.
And Lord, it makes me tremble.

"Shakin' Shakin' Shakes," Los Lobos (1987)

2.1 Introduction

The Pacific Northwest (PNW), a region of North America, encompasses diverse tectonic settings, which produce damaging earthquakes and thus create seismic risk. Characterizing the seismicity of the PNW requires assembling an earthquake *catalog*, or the locations, times, depths, magnitudes and other measurements of instrumentally-observed earthquakes in the region. Our goal is a catalog harmonized across multiple sources for which we can build statistical models of earthquake and aftershock dynamics for different tectonic regimes within the cross-border PNW region.

The tectonic environment of the PNW hosts several distinct tectonic regimes. The Juan

de Fuca plate under the Pacific Ocean is subducting (sliding) under the North American continental plate, which creates a fault over 1000 km long. The stresses at the interface of the continental plate and the subducting slab of rock are capable of generating massive *megathrust* events, the last of which occurred in 1700. The geological processes and conditions within the depths of the slab itself also produce large and damaging *intraslab* earthquakes. Furthermore, there are also numerous shallow faults located in the crust of the North American plate that produce *crustal* earthquakes that are usually (though not always) smaller in magnitude. Due to the subduction process mentioned above, there are also numerous volcanoes in the PNW, which generate *volcanic* earthquakes, which follow patterns distinct from earthquake caused by tectonic stresses. The PNW's earthquakes are distributed inhomogenuously across the US states of Washington and Oregon and the Canadian province of British Columbia, often, but not always, following distinguishable faults or seismic zones (Bostock et al., 2019; Gomberg and Bodin, 2021).

This chapter describes algorithmic procedures for combining earthquake catalog data from three sources for the PNW. We combine catalogs from the Pacific Northwest Seismic Network (PNSN) and Natural Resources Canada's Geological Survey of Canada (GSC), for the US and Canadian portions of the region, respectively (see Figure 2.10). We also merge data quality variables for PNSN events from the Comprehensive Earthquake Catalog (ComCat) of the Advanced National Seismic System. Previous studies characterizing the seismicity of the PNW has also combined US and Canadian data sources to get a single catalog (Bostock et al., 2019) and Gomberg and Bodin (2021). Indeed, the presence of considerable seismicity at the national border (particularly near the Puget Sound and Vancouver Island, see Figure 2.10) necessitates combining data from both countries into a single catalog. The literature on statistical seismicity modelling (e.g., (Zhuang, 2011; Lombardi, 2017; Schoenberg, 2013)) rarely explicates procedures for processing catalogs prior to their modelling. We thus provide sufficient detail for future researchers to reproduce our procedures on catalogs

for other seismic regions, especially those that require merging multiple catalogs.

We also detect spatiotemporal earthquake clusters in the catalog using algorithmic and manual procedures. Subsequent seismicity modelling (see Chapter 3) will be based on modelling these clusters as aftershock sequences; however, the PNW is also known to host earthquake swarms, a different type of earthquake cluster. Swarms are not caused by the same tectonic sources as aftershock sequences and thus, do not follow the same patterns that sequences do. Swarms can complicate aftershock models built on catalogs (Llenos et al., 2009). We thus develop an algorithmic procedure to detect spatiotemporal earthquake clusters and apply this to the PNW. The identified clusters are then hand-labelled by colleagues as swarms or sequences. We expand on this by proposing another cluster detection procedure, which is more robust and adaptive to the catalog.

We have automated three kinds of catalog processing procedures:

1. **Finding matches and mismatches** in earthquake data between PNSN and ComCat, which should but often do not have the exact same values for all earthquakes.

2. **Finding and resolving duplicates** within PNSN, ComCat and GSC individually, and between the catalogs. Each catalog must contain mutually exclusive and deduplicated events before they are combined.

3. **Algorithmically detecting earthquake clusters** in the catalog, for subsequent classification as potential swarms or aftershock sequences. This is part of a multipronged strategy to identify earthquakes occurring within swarms.

In this chapter, we describe our procedures to accomplish the above three steps and their application to the PNW catalog. We then visually summarize the final PNW catalog and study what parts of it are complete.

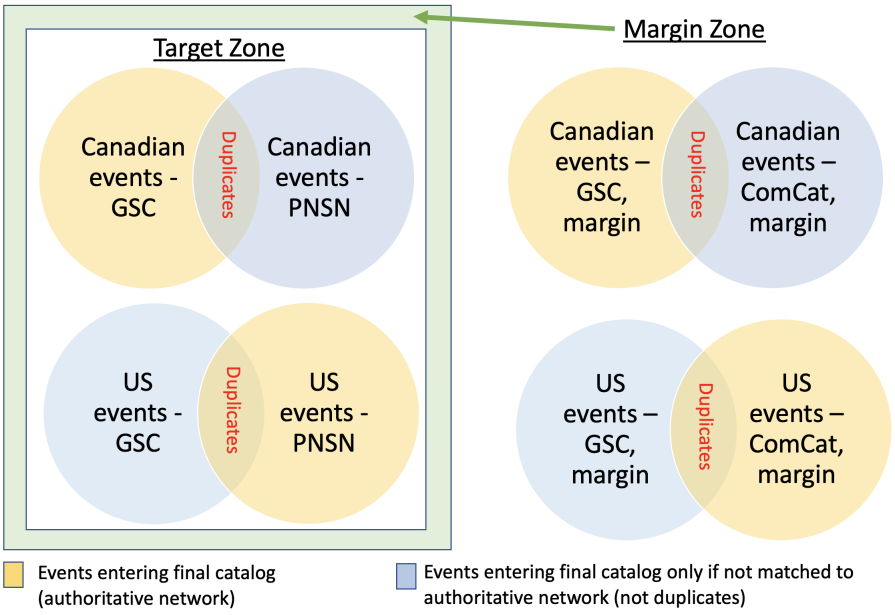

Figure 2.1: Diagram showing component catalogs that comprise the final catalog. We include all events in the territory of the authoritative network (removing duplicates). We then add events from other networks that could not be matched to an event from the authoritative network. Figure is not drawn to scale.

Catalog	Latitude	Longitude	Magnitude	Time Window	Size (Raw)	Size (without MSH)
PNSN	42°× 49°	-125°× -116.5°	M2.0+	1970-01-01 to 2019-01-01	14176	7815
ComCat	42°× 49°	-125°× -116.5°	M2.0+	1970-01-01 to 2019-01-01	14629	8049
GSC	42°× 49°	-125°× -116.5°	M2.0+	1985-01-01 to 2019-01-01	1118	1117

Table 2.1: Target catalog parameters and number of events for raw catalogs. The latitude and longitude correspond to a rectangular box around the US states of Washington and Oregon, which includes a portion of British Columbia. The time window begins in 1970, when the instrumental record of the PNSN began. For GSC, the catalog begins in 1985, as location and magnitude calculations are inconsistent before and after 1985 (Brillon, 2020). We take magnitudes above 2.0 on whichever scale the earthquake was measured in (see Section 2.2.1), the commonly used magnitude of completeness for the PNSN (Gomberg and Bodin, 2021). For all catalogs, we included only "earthquake" or "quake" event types, excluding explosions and low frequency events. After downloading the raw catalogs, we omitted events near the active volcano Mt. St. Helens (MSH in last column; see Section 2.5.1).

Catalog	Latitude	Longitude	Magnitude	Time Window	Size (Raw)	Size (without MSH)
ComCat	41°× 50°	-126°× -115.5°	M1.3+	1970-01-01 to 2019-01-01	49806	37554
GSC	41°× 50°	-126°× -115.5°	M1.8+	1985-01-01 to 2019-01-01	2110	2109

Table 2.2: Auxiliary catalog parameters and number of events for raw catalogs. We use a lower magnitude threshold for the ComCat catalog in order to be able to match to corresponding records with PNSN or GSC events with different magnitudes (either with re-analyzed magnitudes or different magnitude scales).

2.2 Catalog specifications

As described in Chapter 1, seismic networks measure earthquake activity and collect them into earthquake catalogs. We describe here the catalogs collected for the PNW region. A catalog is specified by the space-time-magnitude space for which all recorded earthquakes are collected, as described in Section 1.1. The spatial limits commonly correspond to the space for which the network providing the catalog is authoritative. A catalog typically begins when the network started collecting instrumental data or had sufficient station coverage to ensure consistent earthquake detection. The magnitude threshold is usually set to the network's magnitude of completeness, or the minimum magnitude for which it can expect to capture earthquakes throughout the spatiotemporal space defined above.

These boundaries are based on network operations rather than seismic sources and can bias statistical models made on catalog data (e.g., Zhuang (2011)). The spatial restriction is of particular concern in a region with frequent seismicity on its borders, like the PNW. A common tactic in statistical seismicity modelling is to fix the temporal range and consider a *target* spatial zone of interest and an *auxiliary* zone, which consists of the target zone plus some margin zone around it. We define the target and auxiliary zones for the PNW in Tables 3.6 and 2.2, respectively.

We collect earthquakes from three sources: the PNSN (authoritative for the US states of Washington and Oregon), GSC (authoritative for the Canadian portion of the region) and ComCat, which is supposed to contain authoritative earthquake data from all US global and regional seismic networks. ComCat combines data from different networks, prioritizing the authoritative network for each earthquake's location (Malone et al., 1996). As catalogs are dynamic, earthquake data can be added or revised as more information comes in to each network (Ammon et al., 2020). ComCat also contains data on catalog measurement error (see Section 1.1), which is missing from the available PNSN catalog, and which will be necessary for subsequent modelling. We thus need to merge the PNSN and ComCat datasets.

In May 2020, we downloaded raw catalog data from https://pnsn.org/events?custom_search=true (PNSN; Ludwin (2004)), https://earthquake.usgs.gov/earthquakes/search/ (ComCat; Malone et al. (1996)) and https://earthquakescanada.nrcan.gc.ca/stndon/NEDB-BNDS/bulletin-en.php (GSC; Earthquakes Canada (2020)), using the catalog parameters in Tables 3.6 and 2.2. Furthermore, the following catalog-specific search parameters were used:

Extra PNSN Catalog Parameters

- **Exclude Mt. St. Helens Events:** no

Extra GSC Catalog Parameters

- **Depth Maximum:** 100km

Extra ComCat Catalog Parameters

- **Depth/Azimuthal Gap:** none specified
- **Review Status, Impact, Catalog, Contributor, Product Type:** any

In order to use the ComCat measurement error metrics for the target zone, we must merge this data source with the PNSN target catalog. Each earthquake has an ID assigned to it by its network and so the IDs in ComCat target events should be the same as those in PNSN. Furthermore, the catalog data in ComCat should match that of the PNSN for every earthquake in the target zone, since it is the authoritative network for this area. However, as described in Section 2.3, there are numerous discrepancies between the two catalogs for the PNW target zone such that not all event IDs and event variables matched between the PNSN and ComCat. There are two main reasons for discrepancies between PNSN and ComCat:

1. the PNSN data was reanalyzed or re-listed under a different event ID and not yet updated in ComCat, or

2. ComCat had event values from a different network for this event, e.g., the Northern California network.

The first step in our processing is to resolve any data mismatches and harmonize event IDs between PNSN and ComCat to allow for data merging by ID. After this, we search for duplicates between the US sources (PNSN and ComCat) and the GSC catalog. Once duplicates have been resolved, we can simply combine the processed US and Canadian catalogs to obtain an international PNW catalog.

2.2.1 Magnitude scales between catalogs

Earthquake magnitudes are measured by several scales and this can vary both between networks and within a single network, over time. The PNSN almost exclusively used the local magnitude scale (M_l, Richter (1935)) prior to 2011 and the duration magnitude scale (M_d, Lee et al. (1972)) afterwards. GSC uses several magnitude scales throughout the catalog period but M_l is the dominant scale, accounting for over 98% of earthquakes in this catalog.

We examined magnitude frequencies for 10-year periods of the PNSN catalog for a signal in the change in scale (see Appendix, Figure A.1). No difference could be detected following the magnitude scale change, aside from a smaller number of events, which has been previously reported (Malone, 2019). We thus conclude that the change in magnitude scales in the PNSN should not affect full-catalog statistics.

2.3 Identifying and handling catalog discrepancies

We identify common data issues when merging catalogs, such as duplicate and unpartnered events and mismatches in catalog data. Though this is an essential pre-requisite for seis-

micity modelling, the description of this procedure is often neglected in published statistical analyses of multi-source catalogs like the PNW (e.g., Bostock et al. (2019) writes that they "[took] care to remove duplicate events" without explaining how). We algorithmically identify discrepancies and duplicates between catalogs, using similar procedures to those used by the United States Geological Survey for combining catalogs for seismic hazard analysis (Mueller, 2019).

There are three distinct steps in this processing pipeline:

1. Identify duplicate records within each of the three catalogs.

2. Find the correct match for each PNSN event within the ComCat catalog. Not all PNSN events are linked by ID to a partner in ComCat, requiring us to search for matches by catalog variables.

3. Process duplicates and discrepancies between GSC and the US catalogs by identifying all events within a catalog outside their authoritative region (e.g., Canadian events in the PNSN catalog). Determine whether these are duplicates for a record within the authoritative catalog (in which case, flag for removal) or new events that the authoritative catalog did not capture (in which case, keep), see Figure 2.1.

2.3.1 Terminology

The following terminology will be used to identify and handle catalog discrepancies.

- **Event Variables**: The five *event variables* of a (catalogued earthquake) event are its time, epicentral location (in degrees latitude and longitude), magnitude (in the given magnitude scale, see Section 2.2.1) and hypocentral depth (in km).

- **Strict Match**: Consider two events E_1 and E_2. We want to check if E_1 and E_2 correspond to the same event (that is, if the events match). A *strict match* for an

event variable occurs when E_1 and E_2 have the same value for that variable, to the level of precision of the less precise measurement. A *strict match between two events* occurs when all five of their variables are a strict match.

- **Loose Match**: A *loose match* for an *event variable* occurs when two events E_1 and E_2 have values for that variable within a given tolerance (e.g., Table 2.3). A *loose match between two events* occurs when there is at least one loose match between their five event variables and no mismatches.

- **Mismatch**: A *mismatch* for an *event variable* occurs when events E_1 and E_2 have values for that variable that are not within a given tolerance (e.g., Table 2.3). A *mismatch between two events* occurs when the two events have at least one event variable mismatch.

- **Unpartnered**: Consider two catalogs C_1 and C_2. We call an event *unpartnered* if it exists in C_1 with an event ID but this ID is not in C_2. That is, there is no partner by ID in C_2.

- **Duplicate by ID**: A *duplicate by ID* occurs when two events E_1 and E_2 share the same ID. This may occur within the same catalog or across two catalogs.

- **Duplicate by Value**: A *duplicate by value* occurs when two events E_1 and E_2 have the same or similar event variable values, for all variables. We may search for duplicates within the same catalog or across two catalogs. We may find duplicates by value or by either strictly or loosely matching event variables.

We set tolerances by which to find duplicates within a given catalog (see Table 2.3). We use values from previous studies that search for duplicates across catalogs (Mueller, 2019),

doubling the tolerances before 1990 when station density was lower (see Figures A.2-A.7 in the Appendix).

2.3.2 PNSN and ComCat

Duplicates within each catalog

There were no duplicate events in the PNSN nor the ComCat catalog, when we strictly matched by value. We found eight pairs of PNSN earthquakes whose values loosely matched each other (potential duplicates by value) but these were ruled out as duplicates by Amy Wright (PNSN), largely because of sizable differences in their magnitudes and times.

We furthermore found 18 pairs of earthquakes in the ComCat catalog that loosely matched each other. One of these pairs was confirmed as duplicates by Wright (PNSN), but had magnitudes below M1.8 so could be ignored.

PNSN had no duplicates by ID, and neither did ComCat, when we included in the matching the alphabetical prefix corresponding to which network provided the event's data.

Loosely matched and mismatched events

Variable	**Before 1990**	**1990 and After**
Epicenter Location	± 50 km	± 25 km
Origin Time	± 20 seconds	± 10 seconds
Magnitude	± 0.5	± 0.5

Table 2.3: Tolerances for each event variables used for finding loose matches.

We searched for mismatches, unpartnered events, or other discrepancies between the PNSN and ComCat catalogs. There are four distinct outcomes when matching a PNSN

event to ComCat (see Section 2.3.1 for definitions):

1. the event could not be matched by ID to ComCat (the PNSN event is unpartnered).

2. the event could be matched by ID to ComCat.

   (a) If all variables match exactly, this event is a strict match.

   (b) If all variables are within a designated tolerance (shown in Table 2.3), this event is a loose match.

   (c) If a single variable is outside the designated tolerance, this event is a mismatch.

Table 2.4 shows the results of matching PNSN events to ComCat.

Event Type	**# of Events**
All variables match strictly	6778
All variables match at least loosely (no mismatches)	847
At least one variable has a mismatch	102
Unpartnered events	88
Total	7815

Table 2.4: Accounting for the number of strictly, loosely and mismatched events between PNSN and ComCat.

Renate Hartog and Amy Wright, seismologists at the PNSN, examined all events whose events variables did not strictly match and confirmed that we should take the value from PNSN. The discrepancy was usually due to PNSN containing re-analyzed event variables (e.g., magnitude) that were not yet updated in ComCat. This means that these earthquakes will have re-analyzed PNSN values for event variables and will have the older values for

data quality variables from ComCat. However, when discrepancies exist between PNSN and ComCat's event variables, they are usually not large. 1.2% of all PNSN events had a magnitude mismatch with ComCat (the most common mismatch type), of which all but three were between 0.1 and 0.7 units.

Unpartnered events

There were 88 events in PNSN that did not have a partner (a match by ID) in ComCat (see Table 2.4). For each unpartnered event in the PNSN, we find all ComCat events that loosely match it using the tolerances from Table 2.3. There are three possible outcomes when loosely matching a PNSN event to ComCat:

1. There are no matches in ComCat

2. There is exactly one match in ComCat

3. There are multiple matches in ComCat

The results of the matching are given in Table 2.5. Of these 88 events, 31 loosely matched to a single event in ComCat and none matched to multiple events. That is, although the IDs were different, all the values of the PNSN event's variables were within the tolerance of a single ComCat event.

We attempted to *manually* match the remaining 57 (88-31) PNSN events to an event in ComCat. To accomplish this, we searched through ComCat to find events with similar values for all variables. Most often, times were within three seconds of the matching PNSN event. This combined with latitude and longitude were the most important variables in manually determining matches. Magnitude and especially depth tended to be quite variable, and only aided in differentiating between two close matches. We managed to manually match

25 unpartnered events, each to a single best-matching event in ComCat. Wright (PNSN) confirmed 21 of these matches and rejected the other four.

Of the 32 (57-25) unmatched events, Wright flagged a single event that was likely a sonic blast and should be removed from the dataset.

For confirmed loose/manual matches, the event in ComCat had its event ID changed to the corresponding match from the PNSN catalog. Most of these unpartnered events were at a state border and ComCat had used another seismic network's event values. For the 35 (31+4) PNSN events that were unmatchable to a record in ComCat, these events will be kept in the catalog, by PNSN advice, and simply not be merged with a corresponding entry to ComCat.

Condition on Unpartnered Events	**# of Events**
Total unpartnered events	88
Loosely matched	31
Manual match (and confirmed by PNSN)	21
Unmatched but kept in catalog by PNSN advice	35
Unmatched but removed from catalog by PNSN advice	1

Table 2.5: There were a total of 88 unpartnered events in PNSN, of which 62 could be matched to a ComCat event, 1 was to be removed and the remaining 35 could not be matched and will thus not contain variables from ComCat.

2.3.3 GSC

To merge catalogs across national borders, we need to find the earthquakes in each national catalog that occurred outside its authoritative zone (e.g., US events that were captured by

GSC). If such an event has a match in the authoritative catalog, then it must be removed, as it duplicates an event already captured; however, if it was not captured by the authoritative catalog, we include it in our final combined catalog. We split the GSC catalog into the target and margin zones (the difference between the auxiliary and target zone, so a 1° margin around the target zone). The former is compared to the PNSN catalog and the latter to the ComCat catalog (see Figure 2.1). Events needed to be matched by value as the networks used different ID systems.

Duplicates

We began by removing all events in GSC within 10 km of Mt. St. Helens (see Section 2.5.1), which amounted to only one event. We then searched for duplicates by loosely matching all earthquakes to each other and found 10 pairs of duplicate events. These were confirmed by Camille Brillon (seismologist at GSC) to be duplicates and we were instructed which event of each pair to remove.

Target zone (comparison to PNSN)

We compared the GSC target zone to the PNSN catalog (target zone). Given the different magnitude scales between catalogs, we consider GSC events within the spatial bounds but below the M2.0 threshold for comparison to PNSN. We further split the catalog into US and Canadian events (events occuring on either size of the national border, as given by the NOAA Office of Coastal Survey Maritime Limits and Boundary dataset (NOAA Office of Coast SurveyNOAA Office of Coast Survey, 2020)). Within the target zone, there were 1051 GSC events located in the US, 887 of which could be loosely matched to a PNSN event using tolerances in Table 2.3. We attempted to match the remaining 164 (1051-887) events and could manually match 27 of them to a PNSN event.

This left us with 137 (164-27) GSC events located in the US (target zone) that could

not be found in PNSN. Of these, 10 had been previously manually matched by the PNSN in an earlier version of this procedure, and were therefore removed from further comparison, leaving 127 (137-10) events. We attempted to match these to a lower-magnitude version of the PNSN catalog and loosely matched 74 events (using the tolerances in Table 2.3) and manually matched four more events. This meant that these events were captured by the authoritative network (PNSN) but with magnitudes under the M2.0 threshold for the target zone.

This left 49 (127-74-4) events that Wright (PNSN) assisted with matching. 20 of these GSC events in the US were matched to PNSN events with magnitudes below M1.8 or locations outside the target spatial window. Eight events were found to be probable blasts, also excluding them from analysis. Another seven events were matched to PNSN target events (which we were not able to find) or ComCat margin events (see below), meaning they duplicated an event in the authoritative sources. The remaining 14 (49-20-8-7) US GSC events had no equivalent in the PNSN, and therefore will remain in the catalog.

We also searched for Canadian PNSN events in the GSC catalog. Of the 89 Canadian events in the PNSN, only 30 could not be matched to an GSC event, neither strictly nor loosely. We attempted to match these events manually to GSC, but no matches could be made. Of these 30 events, 19 took place before 1985, which was before any available data for GSC, and we left the PNSN events in the final catalog. The remaining 11 (30-19) events were sent to Brillon (GSC). Of these 11 events, nine were matched to an event below the M2.0 threshold for the target zone and removed. The last two events could not be matched to an event in GSC so 21 (19+2) Canadian events in PNSN will remain in the final catalog.

A summary of the US events removed from GSC in the target zone can be found in Table 2.6.

Margin zone (comparison to ComCat)

We compare the GSC margin zone's catalog to the ComCat catalog for the margin zone. Within the margin, there were 467 GSC events in the US. We loosely and then manually matched these events to the ComCat catalog, yielding 362 loose and 21 manual matches. We also found two loose matches in the PNSN catalog (target region).

The unmatched US GSC margin events were investigated by Wright (PNSN) to find a reason for no match in the ComCat catalog. Of these 82 (467-362-21-2) events, 13 were flagged as potential blasts. 30 of these events were matched to a ComCat event outside the margin zone, either with a magnitude below M1.8 (20 events) or outside the spatial window (10 events). Finally, three GSC events were found by Wright to indeed match to ComCat events, which we were not able to find. This left 38 GSC margin events (82-13-30-3+2) in the US to remain in the final catalog, where the final two are due to overlapping reasons the events could not be matched.

We also searched for Canadian ComCat margin events in the GSC margin catalog. Of the 684 ComCat margin events in Canada, we could match 252 loosely to an GSC event. Of the remaining 432 (684-252) events, we manually found 24 matching GSC events. After excluding the 227 events that were outside the time or magnitude range for the GSC catalog (see Table 2.2), we had 181 (432-24-227) potential Canadian events in the margin zone that were not captured by the GSC catalog.

Brillon (GSC) investigated these and found that there were four blasts, and one event was associated with an existing GSC event that was not found in our manual match search, meaning these five events should be removed from the ComCat margin catalog. 89 events were also found by Brillon in the official GSC catalog but did not appear in the margin zone of our downloaded catalog, likely due to re-analysis by the GSC after our time of download. Thus, the remaining 176 (181-5) ComCat events in Canada could not be matched to a GSC partner. Finally, of the 227 Canadian ComCat events initially dropped from comparison

to GSC, we later realized that 33 of them occurred before 1985 and so we added them to the final catalog. In total, 209 (176+33) Canadian events in the ComCat margin catalog remained in the final catalog.

A summary of the events removed from GSC in the margin zone can be found in Table 2.7.

Condition	**Number of Target Events**
Total number of Target GSC events for matching to PNSN	**1051**
Number of GSC events in the US loosely matched to a PNSN event	887
Number of GSC events in the US manually matched to a PNSN event	27
Number of GSC events flagged for removal from previous analysis	10
Number of GSC events in the US loosely or manually matched to an event in a lower-magnitude PNSN catalog with magnitude outside of margin zone	78
Number of GSC events in the US unmatchable to neither PNSN nor ComCat but flagged as potential blasts	8
Number of GSC events matched by Wright to event outside spatiomagnitude margin zone	20
Number of GSC events matched to a ComCat event inside the margin zone	7
Number of GSC events in the US with no match to a US catalog event, to remain in final catalog	**14**

Table 2.6: Summary of events flagged for removal from GSC's target catalog.

Condition	**Number of Margin Events**
Number of GSC events in the US manually matched to a ComCat event	**467**
Number of GSC events in the US loosely matched to a ComCat event	362
Number of GSC events in the US manually matched to a ComCat event	21
Number of GSC events in the US manually matched by Wright to a ComCat event in the margin zone	3
Number of GSC events in the US matched by Wright to a ComCat event outside the margin zone	30
Number of events flagged as potential blasts	13
Number of events with no match to a US catalog event, to remain in final catalog	**38**

Table 2.7: Summary of events removed flagged for removal from GSC's margin catalog.

2.4 Identifying and handling potential swarms in earthquake catalogs

The statistical modelling of seismicity often requires considering which catalog events are background earthquakes and which were triggered by a previous event within an aftershock sequence. For example, in the Epidemic-Type Aftershock Sequence (ETAS) model (Ogata, 1998), the total seismicity rate λ at any given time and location (t, x, y), given the history of recorded earthquakes $\mathcal{H}_t$, is modelled as:

$$\lambda(t, x, y|\mathcal{H}_t) = \mu(x, y) + \sum_{j}^{t_j<t} a(t - t_j, x - x_j, y - y_j, M_j),$$

where $\mu(x, y)$ is the spatiotemporal background rate (number of events/(day $\cdot$ km^2)) at a given location (x, y) and $a(\cdot)$ is the spatiotemporal aftershock rate, collected over all previous events, and governed by several seismological laws. The ETAS model thus assumes that a catalog can be (probabilistically) split into a set of background events and sets of aftershock sequences. While the background seismicity rate $\mu(x, y)$ can be allowed to vary over space, it is assumed to be stationary and homogenuous in time.

Swarms are spatiotemporal clusters of earthquakes that complicate such a split of the catalog. Events in swarms are not triggered by a previous earthquake as in aftershock sequences; rather, they occur because of aseismic sources, such as from underground fluids (volcanic magma or groundwater) (Roland and McGuire, 2009) or anthropogenic sources (such as wastewater injection) (Llenos and Michael, 2013). Swarms manifest in a catalog as brief surges in the seismicity rate for a subregion, beyond what is typical for its background seismicity. The many volcanoes of the PNW contribute to the presence of swarms in this catalog (Bostock et al., 2019).

Including swarms in catalogs used for statistical seismicity modelling is problematic for two reasons. First, swarms temporarily raise the occurrence of earthquakes not triggered by previous earthquakes, which breaks the assumption of stationarity and homogeneity in time

of the background seismicity rate. Second, they follow spatial, temporal, and magnitude-frequency patterns that differ from the seismological laws that describe aftershocks; for instance, they start and end abruptly with no decay of events over time or space, as in aftershock sequences. Thus, they may bias ETAS parameter estimates because though the model may regard them as triggered seismicity due to their clustering, their patterns do not seem to comply with the seismological laws describing triggered seismicity (Llenos et al., 2009).

Exact definitions of swarms based on catalog variables do not exist. Previous methods to identify swarms require analysis of very low-magnitude events (e.g., Farrell et al. (2009)) or each event's waveforms (e.g., Vidale and Shearer (2006); Skoumal et al. (2016)). Other authors have used the ETAS model itself to diagnose a period of seismicity as following a pattern of elevated background rate or sharp deviations in triggering parameters, which can then be tied to swarm-like seismicity (Llenos and Michael, 2013; Okutani and Ide, 2011). These approaches are not suitable for our case, where we seek to detect swarms within a complete catalog, without appealing to additional event data nor yet attempting to fit a sophisticated model on a complex catalog. The existing classical declustering algorithms (e.g., Gardner and Knopoff (1974); Reasenberg (1985)) are also not suitable, as we seek to detect and study the characteristics of distinct clusters, rather than remove all clustered seismicity from the catalog. We thus require a different method to identify and flag (potential) swarms in the PNW catalog.

2.4.1 Algorithmically detecting spatiotemporal earthquake clusters

We propose the following simple and scaleable procedure to identify earthquakes in potential swarms. The first step is to detect spatiotemporal clusters in the data and we offer a deterministic approach based on fixed windows and rules. The second step is to hand-label the detected clusters as either swarm or aftershock sequence (when possible). Our cluster

detection step is similar in design to some previous work on cluster identification. While some authors have taken a purely visual approach (Holtkamp and Brudzinski, 2011), this can be more rigorously formalized using pre-specified spatiotemporal windows to collect earthquakes into clusters, as in classical declustering procedures (e.g., Reasenberg (1985)). Our approach here is closest to that of the CURATE method (Jacobs et al., 2013), where pre-specified spatial and temporal rules are applied to a catalog, followed by boundary checking.

We performed the following procedure on the PNSN target catalog only.

1. Split the region into five areas, as is common in seismicity analysis across large regions with disparate seismicity patterns, e.g., Veen and Schoenberg (2008)'s Figure 6 (see our Figure 2.2). We delineated these areas to minimize separating regions of concentrated seismicity.

2. Find temporal clusters. Find all earthquake clusters in which three or more earthquakes occurred within an area within a seven-day period. After this criterion is met for some seven-day period, use a moving time window to continue checking for the criterion in subsequent seven-day windows from the following earthquake in the cluster. All events meeting this criterion will belong to the same cluster, until the criterion is no longer met.

3. Find spatiotemporal clusters. Restrict the temporal clusters to those that have at least three events within an area 50km^2 of one another. This is done to further eliminate clusters of events that took place close together in time, but not space.

4. Find spatiotemporal-depth clusters. [1] Categorize clusters into, deep, shallow or mixed clusters using a simple depth cutoff. *Shallow* clusters are defined as clusters where two

[1]Note: this work was done before understanding that triggered seismicity have been known to occur across tectonic regimes in other subduction zones (Bilek and Lay, 2018), but is described as originally conducted.

or fewer events took place above 30 km. *Deep* clusters are defined as clusters where all but at a maximum two events took place below 30 km. A *mixed* cluster is a cluster that is neither a shallow nor deep cluster. For shallow and deep clusters, remove events that were not within the majority depth categorization (which by definition, is at most two events) from the cluster. For mixed clusters, keep all events.

Note, clusters of size five (which totaled 49 of our 322 clusters up to this step) are not included in this analysis, since they have a unique susceptibility to incorrect categorization due to the presence of spatial outliers. Consider a five-member cluster in which:

- three events constitute a real mixed cluster (say, an actual swarm)
- two events are spatial outliers (that is, not actually part of the swarm) but take place within the same depth categorization as the minority of the real swarm events.

This cluster would not get categorized as a mixed event (it has a majority depth class). Thus, two real swarm events in the minority depth class would be removed.

Clusters of any other size either contain too many events for this to occur, or too few events to have a true mixed swarm and contain spatial outliers that could misclassify the cluster. Clusters of size five are thus simply skipped in this step.

5. Map each spatiotemporal-depth cluster to visually determine that all events are still spatially co-located. Discard events that are visibly far from where the seismicity is concentrated.

6. Visually inspect the border regions of the five areas for any visible spatial clusters and check whether any combination of them meets the three-events-in-seven-days threshold. This is to check for any clusters that the procedure in steps 2-5 would have missed.

7. Visually inspect all identified clusters and remove obvious spatial outliers that were not found by the algorithm in step 3. Dissolve any of the remaining clusters where the three-events-in-seven-days criterion is no longer met.

The resulting spatiotemporal clusters were sent to colleagues at the PNSN for classification. We produced summary plots including plots of events' magnitudes vs. time and a map of event locations with varying dimensions depending on the cluster spatial extent (see Figures 2.3-2.5).

2.4.2 Classifying spatiotemporal clusters as swarms or sequences

The classification of clusters was done by expert opinion by Hartog and Wright (PNSN) using the following six-point scale:

- **0: Neither a swarm nor an aftershock sequence.** These earthquakes do not appear to be related enough to be a cluster or possess neither the characteristics of swarms, nor sequences.

- **1: Definitely a swarm.** The cluster has the features of a swarm, as far as times, locations, magnitudes and depths.

- **2: Likely/maybe a swarm.** The cluster exhibits more features of a swarm than an aftershock sequence, but it is not definite.

- **3: Don't know / too hard to tell.** The cluster could either be a swarm or an aftershock sequence.

- **4: Likely/maybe a sequence.** The cluster exhibits more features of an aftershock sequence than a swarm, but it is not definite.

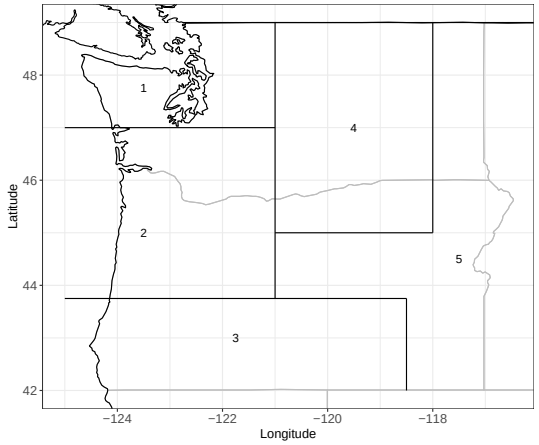

Figure 2.2: Map of the PNW (target zone) split into areas for swarm identification.

- **5: Definitely a sequence.** The cluster has the features of an aftershock sequence, as far as times, locations, magnitudes and depths.

Wright and Hartog (PNSN) also provided comments for each cluster, indicating whether it contained subclusters that should be classified separately, or individual events that did not belong to the bulk of the cluster. For any such catalog, we made the appropriate splits or removal of events, ensuring that the resulting (sub)clusters still maintained the criteria in steps 2-4 (e.g., still had its first 3+ events in a consecutive seven-day period). Any (sub)cluster that did not meet such a criteria had its scale value relabelled to 0. Otherwise, we assigned the scale value given to a cluster to each cluster member.

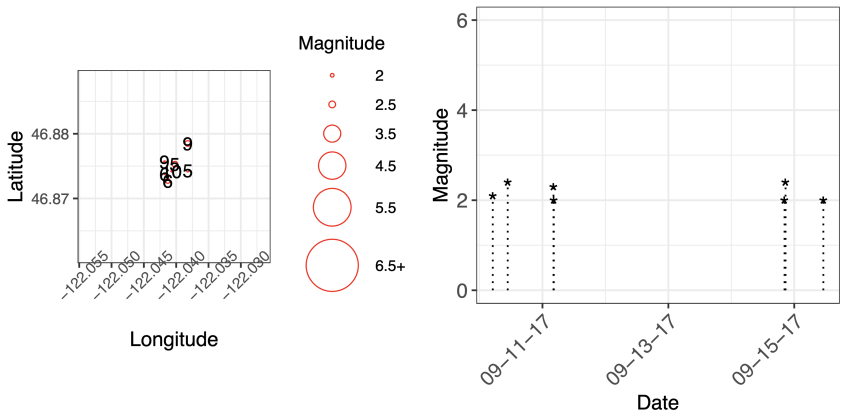

Figure 2.3: An identified cluster without the features of an aftershock sequence.

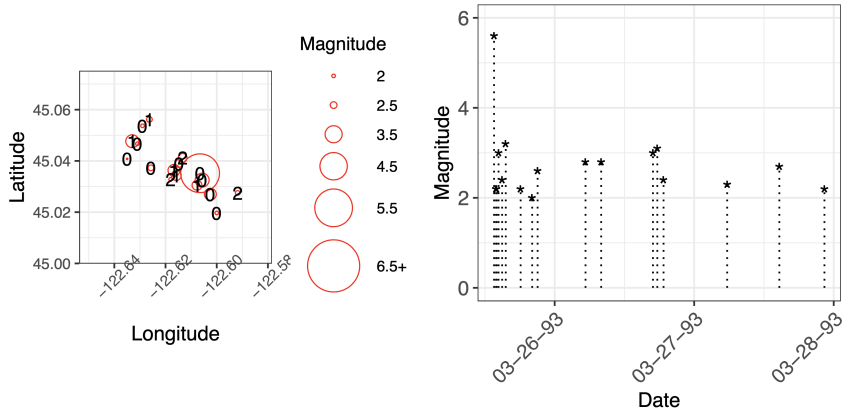

Figure 2.4: An identified cluster with a power-law temporal decay, difference in magnitudes between the first and subsequent events, and other features of an aftershock sequence.

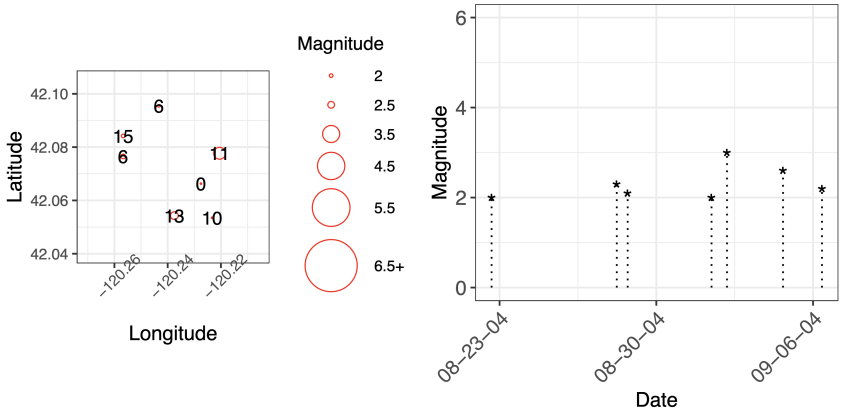

Figure 2.5: An identified cluster that is harder to classify. It does not begin with its highest-magnitude event and also lacks the typical decay of earthquakes with time, but still may be an sequence with aftershocks that have magnitudes unusually similar to the mainshock.

2.5 Definite and potential swarms in the PNW

2.5.1 Removing Mt. St. Helens

Mt. St. Helens is an active volcano in Washington State and also lies near several faults. The volcano generated several prolonged periods of swarm activity during our catalog period. Swarms are often indistinguishable from tectonic events in active volcanic zones, which requires excluding events near Mt. St. Helens from our catalog. As the very first step of our analysis, we removed any events within a circular area of radius 10 km from the volcanic center at longitude -122.1956 and latitude 46.1914. This is a conservative exclusion area in line with other studies of PNW seismicity (Malone (2019, 2020)).

2.5.2 Results of algorithmic swarm search

Through our algorithmic procedure, we detected 322 distinct clusters, involving 2620 events in the target catalog. Hartog and Wright (PNSN) classified each detected cluster as either a potential or confirmed swarm or sequence, impossible to distinguish, or not actually a cluster. For each cluster, they primarily used the spatial clustering, magnitude-time distribution, the range of depth values, and the difference between the highest and second-highest magnitudes.

The results of the classification are given in Figure 2.6.

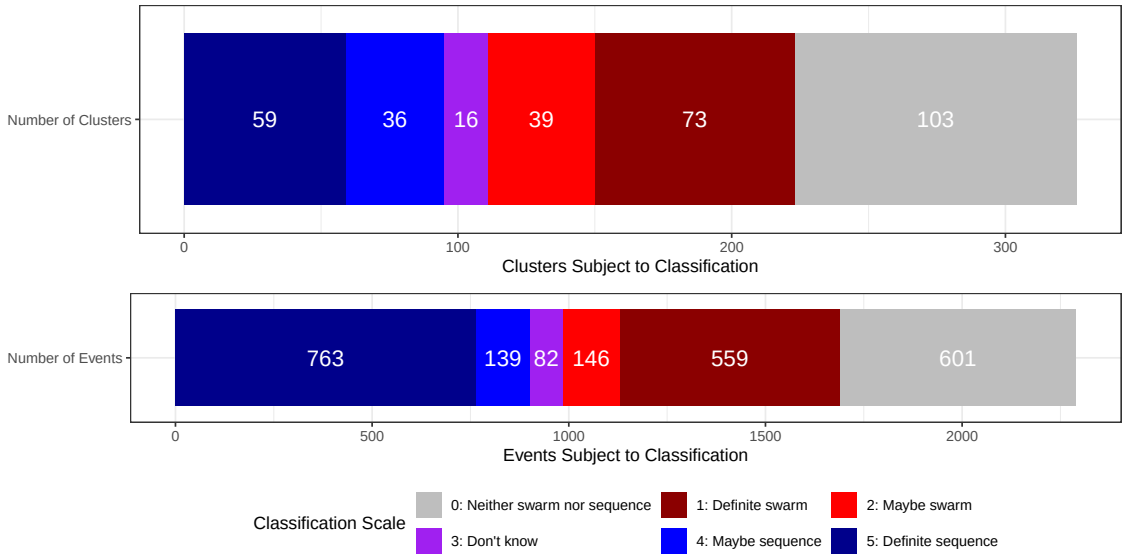

Figure 2.6: Number of clusters and events that were classified into six classes representing definite/possible swarms and aftershock sequences.

Event Type	**Number of Events**
Total PNSN events	7815
Events in confirmed swarms (68 clusters)	418
Events flagged in potential swarms (42 clusters)	147
Total PNSN events not flagged in maybe/confirmed swarms after this step	7250

Table 2.8: Summary of the algorithmic search for swarms.

2.5.3 Searching for documented swarms

Earthquakes can also be part of known, documented swarms; however, because of the time duration or magnitude levels of the swarms, there may not have been three events in any seven-day window to be flagged in our complete catalog. We must thus also search for catalog events that are part of documented swarms.

Swarm ID	Event Name	Latitude, Longitude	Period of Activity	Relevant Link
D1	Wooded Island 1969-71	46.43° – 119.23°	1969-1971	https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2011JB008433
D2	Wahluke	46.74° – 119.9°	Dec. 2-17, 1972	https://pubs.geoscienceworld.org/ssa/bssa/article/65/4/855/117419/details-of-microearthquake-swarms-in-the-columbia
D3	Eltopia	46.46° – 119.01°	June 4-July 31, 1973	https://pubs.geoscienceworld.org/ssa/bssa/article/65/4/855/117419/details-of-microearthquake-swarms-in-the-columbia
D4	Royal	46.9° – 119.64°	Dec 1973-Feb 1974	https://pubs.geoscienceworld.org/ssa/bssa/article/65/4/855/117419/details-of-microearthquake-swarms-in-the-columbia
D5	Wooded Island 1975	46.43° – 119.23°	Apr-Aug 1975	https://pubs.geoscienceworld.org/ssa/bssa/article/102/1/309/349695/Contemporary-Seismicity-in-and-around-the-Yakima
D6	Maupin 1976	45.18° – 121.08°	1976	https://assets.pnsn.org/legacy_news/PRESS_RELEASES/maupin/
D7	Elk Lake, WA	47.85° – 123.64°	May 18-July 25, 1980	https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/grl.50354
D8	Spokane	47.66° – 117.43°	June 2001, Nov 2001	https://pnsn.org/blog/2012/02/08/the-spokane-swarm-about-10-years-ago
D9	Three Sisters, Oregon	44.14° – 121.77°	March 23, 2004 - March 24, 2004	https://assets.pnsn.org/legacy_news/PRESS_RELEASES/March242004Info.html
D10	Maupin 2000s	45.18° – 121.08°	Jan 2007 - Nov 2011	https://assets.pnsn.org/legacy_news/PRESS_RELEASES/maupin/
D11	Wooded Island 2009	46.43° – 119.23°	Jan 4 2009 - Jul 2009	https://pubs.geoscienceworld.org/ssa/bssa/article/102/4/1786/325491/Tectonic-Setting-of-the-Wooded-Island-Earthquake https://assets.pnsn.org/notable/WEBDIR_09010401314e/summary.html
D12	Mt. Rainier	46.85° – 121.76°	September 2009	https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/grl.50354
D13	Mt. Hood	45.37° – 121.7°	Feb 23, 2012	https://pnsn.org/blog/2012/2/23/mount-hood-earthquake-swarm-of-feb-23-2012
D14	Mt Mclaughlin	42.44° – 122.31°	March 23 2013	https://pnsn.org/blog/2013/03/24/small-swarm-near-mount-mcloughlin-last-night
D15	20 km N of MSH	46.37° – 122.2°	August 24 2013	https://pnsn.org/blog/2013/08/24/peppy-seismic-swarm-20-km-nw-of-mount-st-helens
D16	Three Sisters	44.14° – 121.77°	March 2014	https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2011GC003826
D17	Bend	44.06° – 121.31°	Oct 20 - Oct 26 2015	https://pnsn.org/blog/2015/10/23/unusual-earthquake-swarm-south-of-bend-or
D18	NE Bremerton/Whidbey Island	48° – 122.44°	May 10 - May 11, 2017	https://pnsn.org/blog/2017/05/11/earthquake-swarm-ne-of-bremerton

Table 2.9: Documented PNSN swarm events.

Table 2.10 is the result of a search for any earthquakes documented by PNSN or other sources as being a swarm. The search consisted of two parts:

1. We searched through PNSN's 2002-2011 press releases (https://assets.pnsn.org/legacy_news/) and more recent blogposts (using the search feature at https://pnsn.org). We used search terms: "swarm", "swarms" and "aseismic". For each distinct swarm, we identified the location (place name), year and associated time length of the swarm. We geo-located the swarm by its associated place and assigned that latitude/longitude to the swarm.

2. We did a literature search using Google Scholar and Web of Science, for any published articles or reports on swarms in this region. For each search engine, we used the following procedure:

    - Search [washington "earthquake swarm"] in search engine. For any relevant article:
        - Search for "swarm" in the article and identify the location (associated place name), year of the swarm and approximate time length of the swarm from the article.
        - Add to Table 2.9 if an entry did not exist for it from step 1. If an entry existed from a PNSN page, update the information based on the article, assuming this was more conclusive than a blogpost/press release.
    - Repeat for search terms: [oregon "earthquake swarm"]; [cascadia "earthquake swarm"]; ["pacific northwest" "earthquake swarm"].

Using the location of the event as the center of the swarm, we created a square with side lengths of 0.3 degrees latitude/longitude to search for events in the catalog. We searched for events that happened within two weeks of the start and end date of the documented swarm. These were mapped and we removed any clear spatial outliers from where the swarm

events were concentrated. We gave all events that were found in these documented swarms a value of 1 in our swarm/sequence scale. Of the 177 found earthquakes in these swarms, 25 events were in confirmed swarms already in the previous step, and 22 events were in potential swarms and thus reclassified to confirmed swarms.

Event Type	**Number of Events**
Total PNSN events	7815
Events in confirmed swarms (73 clusters)	559
Events flagged in potential swarms (39 clusters)	146
Total PNSN events not flagged in maybe/confirmed swarms after this step	7110

Table 2.10: Summary of swarms in catalog after the combined search for swarms (algorithmic search and processing documented swarms).

We summarize the results of the combined search for swarms and sequences in Figure 2.6. All confirmed and potential swarms and sequences are mapped in Figure 2.7.

2.6 Semiparametric cluster detection

The procedure used to identify clusters described in Section 2.4.1 relies on arbitrary windows in time and space (three events in seven days in an area of 50 km^2). It also requires several manual steps, including searching for spatial outliers by eye. It is important for cluster detection to not be based on visual inspection but also to be adaptive to the different levels of seismicity and clustering across the areas of the PNW. It is furthermore desirable for this method to be less parametric, or less reliant on arbitrary spatiotemporal windows or rules to find clusters. Window-based methods typically use fixed thresholds based on rules of thumb, or use magnitude-spatial zone relationships that require additional information,

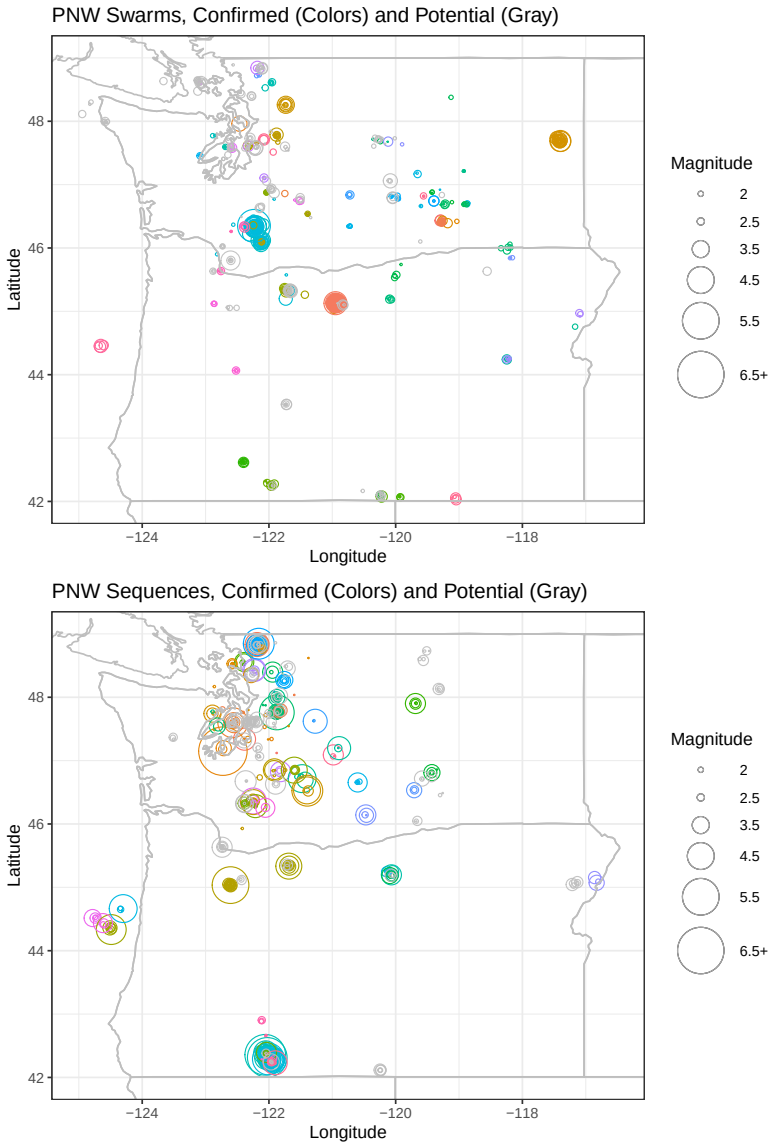

Figure 2.7: All swarms (top) and sequences (bottom) mapped, with size scaled according to magnitude. All confirmed swarms/sequences are plotted in colors and all potential swarms/sequences in grayscale.

e.g., on fault lengths (Wells and Coppersmith, 1994). The existing literature on earthquake cluster detection that is not based on spatiotemporal windows (e.g., Jacobs et al. (2013) and Zaliapin and Ben-Zion (2013a)) also require either estimating or user-inputting parameters that describe clustering in a catalog. We thus require a new cluster detection method that is simple, adaptive to the catalog, and based solely on catalog variables.

In this section, we propose a semi-parametric approach which uses statistical cluster analysis on empirical distributions of catalog earthquakes' time and spatial distances. We start by ordering the catalog in chronological order and calculating *consecutive interevent times*, $\Delta t_{i,i-1}$, the difference in time between any earthquake i and the previous earthquake, $i-1, (i = 2, \ldots, n)$. We also calculate *interevent epicentral distances* (Δs_{ij}) between *all* pairs of catalog earthquakes i and j as there is no natural ordering for space. We do not use the depth variable for clustering in this procedure because there is evidence of triggering across tectonic regimes in other subduction zones (Bilek and Lay, 2018). Also, unlike epicentral locations and times, we do not expect earthquakes within a swarm or a sequence to necessarily be of a very similar depth.

We again use the five-area split of the PNW region (see Figure 2.2) and follow the procedure below, separately for each area.

1. **Find temporal earthquake clusters using $\Delta t_{i,i-1}$.** If two consecutive earthquakes have a time difference less than the q_tth percentile of the empirical distribution of $\Delta t_{i,i-1}$, then they form a cluster. Once a cluster is established, add the next earthquake to it if this earthquake's $\Delta t_{i,i-1}$ is also under the threshold. This will continue until an earthquake happens whose $\Delta t_{i,i-1}$ is above the threshold. Thus, the minimum cluster size is two.

2. **Search for spatiotemporal clusters within the temporal clusters.** Each temporal cluster may consist of one or multiple distinct spatiotemporal clusters. We form

these clusters using statistical *agglomerative hierarchical clustering*, which calculates the pairwise distances between all points and step-wise combines them into clusters based on a linkage rule. We use Euclidean distance and *single linkage* (minimum distance between points) to link points into clusters. We use single linkage because it is possible for some clusters to have a directionality in space; thus, distance to the nearest cluster member should be the criteria for inclusion into the cluster, rather than from the furthest cluster member (complete linkage) or average cluster member (average linkage).

Within each temporal cluster, we link together those earthquakes/earthquake groups whose (single linkage) distances are below the q_sth percentile of the empirical distribution of Δs_{ij}.

- We also ensure that the temporal condition still holds for the spatiotemporal cluster(s) found in Step 2. Repeat Step 1 on each identified spatiotemporal cluster in Step 2 and remove any events that are not sufficiently close together in time. Any clusters that no longer have two or more events return to the set of unclustered events.

3. **Combine clusters that are reasonably close in time and space.** There can be larger clusters that appear in our complete catalog as multiple distinct clusters, because in between them, the seismicity temporarily dropped below the minimum magnitude, so in the complete catalog, they appear as distinct clusters. We need to combine these *split clusters*. Combine a cluster c_k with the previous cluster in time c_{k-1} ($k = 2, \ldots, n_{cluster}$) if:

   (a) c_k has a start date within $\tilde{t}$ days of c_{k-1}'s end date AND

   (b) c_k has a cluster center within $\tilde{s}$ km of c_{k-1}'s cluster center.

Here, $\tilde{t}$, $\tilde{s}$ can again be defined using percentiles of the empirical distributions of $\Delta t_{i,i-1}$ and Δs_{ij} respectively.

We consider this method *semi-parametric* because it solely relies on pre-specified percentiles (q_t, q_s, $\tilde{t}$, $\tilde{s}$) of the empirical inter-event time differences and distances. These may be adjusted to study clustering under different time- and space-thresholds, as called for in the literature (Zaliapin and Ben-Zion, 2021); because these thresholds are percentiles of empirical catalog distributions, they are also adaptive to the seismicity being studied, rather than pre-specified based on global patterns or other rules of thumb.

2.6.1 Demonstration of procedure

We will demonstrate the efficacy of this method by comparing it to the more parametric and manual method previously used. The estimated clustering in the catalog will depend on the parameter values described above (q_t, q_s, $\tilde{t}$, $\tilde{s}$). For each area of the PNW, we ran the method several times, using combinations of the following percentiles for q_t and q_s: $\{10, 20, 25, 30, 35, 40, 50\}$ (as pilot runs indicated that lower percentiles were not appropriate for cluster detection). In Tables A.1 and A.2 in Appendix Section A.3, we give the times and distances that these percentiles correspond to, for each area. For the cluster combination percentiles, $\tilde{t}$ and $\tilde{s}$, we check results when $\{\tilde{t} = q_t, \tilde{s} = q_s\}$ and $\{\tilde{t} = 2q_t, \tilde{s} = 2q_s\}$.

For each combination of percentiles, we will compare the detected clusters with the clusters confirmed by PNSN seismologists ("true clusters", regardless if they were labelled as swarms or sequences; that is, all clusters whose scale value was 1, 2, 3, 4 or 5). We will thus only consider clusters as detected if they had at least three members, sending all clusters of size two to the background set. Since the aim is to replicate the true clusters, we assess our method's performance at the *cluster level*, using the following metrics:

- **Sensitivity rate at cluster level** (true positive rate): #{True clusters whose members are all in the same detected cluster} / #{True clusters }

- **Δ number of clusters:** Number of detected clusters - Number of true clusters

We complement these metrics with performance metrics at the *earthquake level*:

- **Specificity rate at earthquake level** (true negative rate): #{Earthquakes NOT in true clusters that were also NOT in detected clusters} / #{ Earthquakes NOT in true clusters }

- **Δ background proportion (earthquake-level):** Detected background proportion - True background (non-clustered) proportion

- **Sensitivity rate at earthquake level** (true positive rate): #{Earthquakes in true clusters that were also in any detected clusters} / #{ Earthquakes in true clusters }

2.6.2 Results and outlook

The choice of the time threshold dominates the performance of the cluster detection, with very low (cluster-level) sensitivity around $q_t = 10$ rising monotonically to higher sensitivity at $q_t = 50$ (see Figure 2.8, top). There is a corresponding but much smaller drop in the specificity with rising q_t (all areas had specificity of at least 0.72 across all time and space cutoffs). Changes in the spatial threshold, q_s, make very minor differences in sensitivity and specificity. Results vary by area; Areas 2, 4 and 5 have the highest (cluster-level) sensitivities (over 0.75), while the other two areas do worse at detecting cluster structure exactly right, with Area 3 having the lowest sensitivity (0.22), when $q_t = 50$ and $q_s = 10$. This area has a far larger proportion of events lying in true clusters (see Table 2.11) as well as fairly small number of events. Area 1 with a sensitivity of 0.56 has the largest number of events and clusters, when $q_t = 50$ and $q_s = 10$.

The time and spatial thresholds also influence how well the method detects the number of clusters and the background proportion (see Figure 2.9). Here the optimal value is 0,

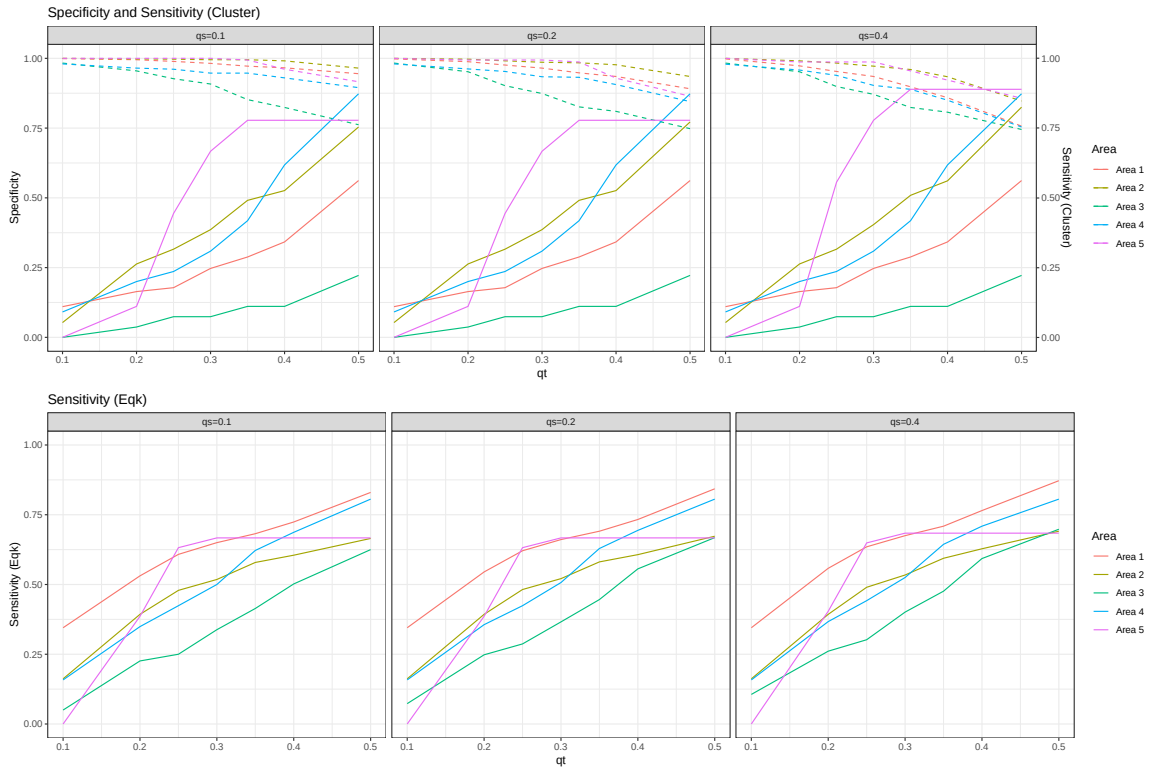

Figure 2.8: Top: Cluster detection sensitivity (cluster level) in solid line and specificity in dotted line, compared to the true cluster. Bottom: Cluster detection sensitivity at the earthquake level.

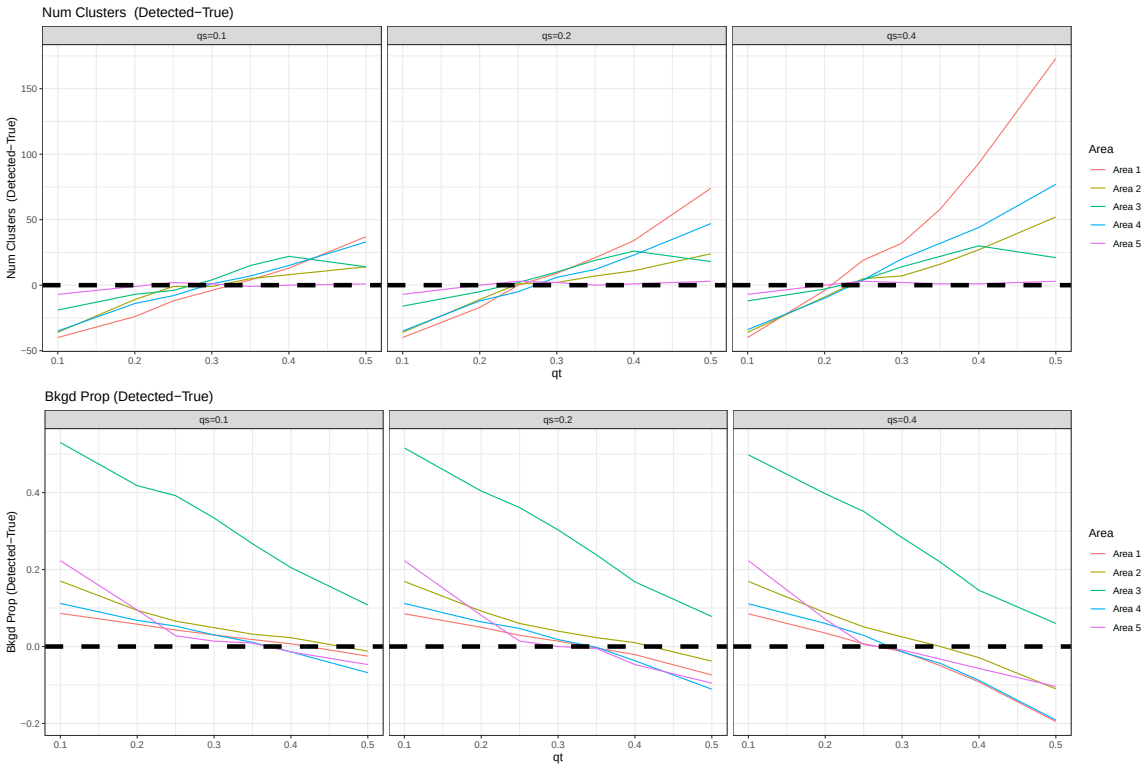

Figure 2.9: Cluster detection performance metrics for number of clusters detected and proportion of background (non-clustered) events detected.

	Prop. Clustered	Number Events	Number True Clusters
Area 1	0.13	3359	72
Area 2	0.17	1666	56
Area 3	0.57	821	27
Area 4	0.13	1814	54
Area 5	0.18	211	8

Table 2.11: Information about observed clustering in each area.

which represents exactly correct detection. For $q_t = 10$, the method somewhat underdetects the number of clusters, hits an optimal detection around $q_t = 25 - 30$ and then rises again to overdetect the number of clusters thereafter. As q_s increases, the slope of these rises increases, leading to much larger overdetection when both q_s and q_t are around 50. The reverse is true for detection of the background proportion: here, the method overdetects the unclustered seimsicity for lower q_t, hitting an optimal detection around $q_t = 25 - 30$ and then underdetecting the background proportion for higher q_t; the only exception is Area 3, where the method overdetects the background proportion for all q_t and q_s values.

We finally examine the earthquake-level sensitivity in Figure 2.8 (bottom), which represents correct detection of earthquakes in a cluster, regardless if the cluster structure being correct. Here, Areas 1 and 3 performed much better (sensitivity of 0.83 and 0.63, respectively), suggesting that the method likely failed at correctly identifying their cluster structure, either due to very long sequences it did not pick up (e.g, the Klamath Falls sequence in Area 3) or shorter sequences that were not fully realized by the method (e.g., the 52 clusters in Area 1 with 5 earthquakes or less).

These results are for cluster combination percentiles $\tilde{t}$ and $\tilde{s}$ equal to q_t and q_s. When we double these, there is only a minimal difference in results (see Figures A.8 and A.9 in

Appendix Section A.3). It appears the choice of cluster combination thresholds does not affect the performance of this cluster detection method.

In our simple and semiparametric cluster detection method, we obtained reasonable performance when q_s was around 10 and q_t was between 25-50, for all areas outside Area 3. The tradeoff for higher q_t values was between getting a better sensitivity rate (at the cluster and earthquake levels) and overdetecting the number of clusters. As our method is adaptive to the specifics of the catalog, it had similar performance across areas of high and low seismicity. We suspect that the reason cluster detection performed worse for Area 3 is because its seismicity is concentrated into only a few explosive sequences.

2.7 Merging catalogs

2.7.1 Merging PNSN and ComCat

Section 2.3.2 describes the process of identifying partners for all PNSN events in the ComCat catalog, either by ID or by value. Of the 7815 PNSN events in the target catalog, 7646 of them had a partner found in ComCat. We merged the two catalogs by ID, which added ComCat variables to the matchable PNSN events. The remaining 169 events remained in the catalog but without ComCat variables.

We then added ComCat margin events to form an auxiliary PNW catalog. Of these margin events from ComCat, 116 were found to match to a PNSN target event; after removing these duplicates, 9462 margin events joined the catalog, leading to an auxiliary PNW catalog of 17208 events.

2.7.2 Combining PNSN and GSC

Following the processing described in Section 2.3.3, we obtain the component catalogs that comprise the full international PNW catalog (see Figure 2.1). Table 2.12 lists the components that are bound together to form the final catalog for modeling.

Catalog	Size
PNSN US events in the target zone	7725
GSC Canadian events in the target zone	64
PNSN Canadian events, not represented in GSC, in the target zone	21
GSC US events, not represented in PNSN, in the target zone	14
Total in target zone	7824
ComCat US events in the margin zone (not represented in PNSN target zone)	9255
GSC Canadian events in the margin zone	523
ComCat Canadian events, neither represented anywhere in GSC nor PNSN's target zone, in the margin zone	207
GSC US events, not represented in ComCat, in the margin zone	38
Total in margin zone	10023
Total in auxiliary (target+margin) zone	17847

Table 2.12: Component catalogs comprising the final international PNW catalog (target, margin, and auxiliary).

2.8 Final PNW catalog

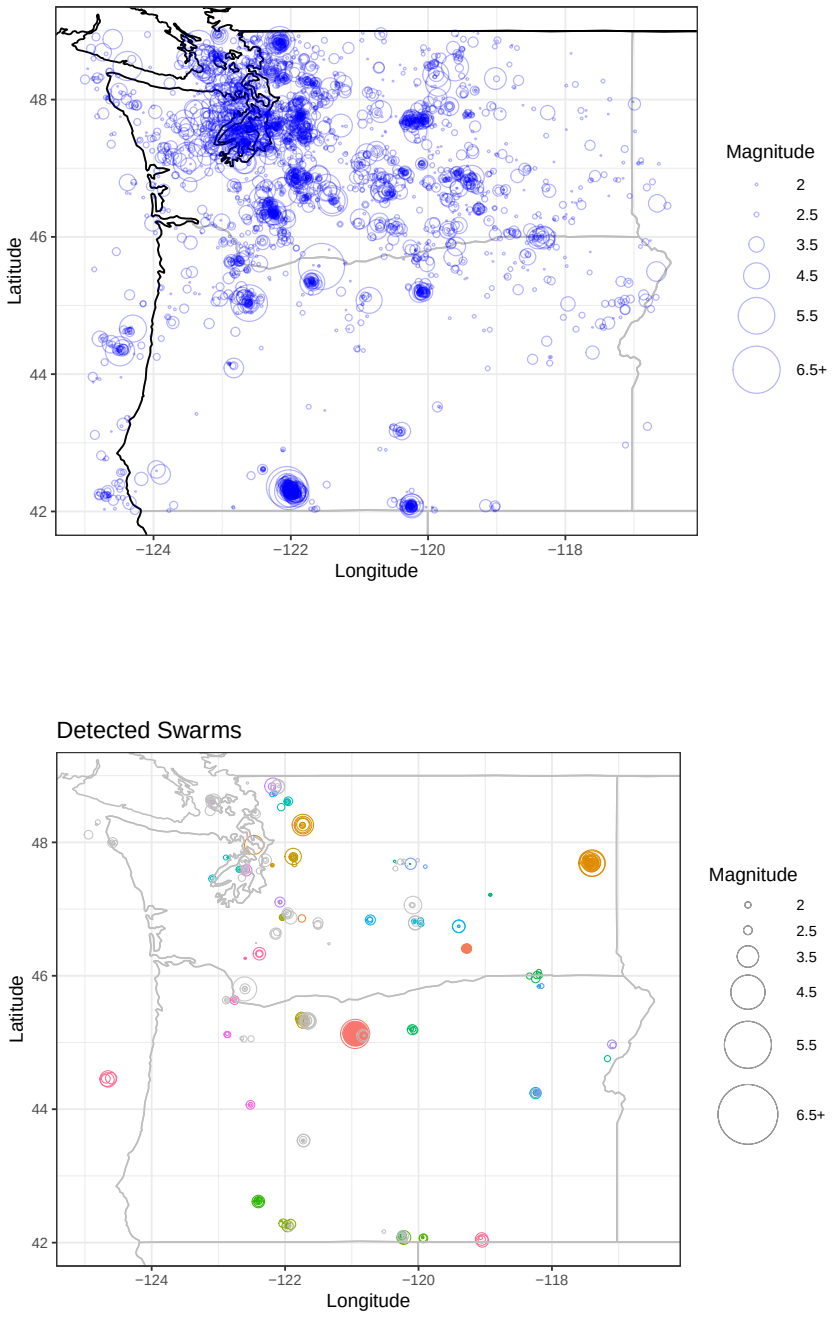

Figure 2.10: Target catalog with swarms removed (top) and the algorithmically identified or previously documented swarms (bottom). Swarms that were confirmed or documented are plotted in colors while potential swarms are plotted in gray.

The combined target PNW catalog, with swarms removed, is mapped in Figure 2.10 (left) with close-ups of each demarcated area in Figures 2.11-2.13 (target events in blue and margin events in gray). In Washington State, seismicity is clustered in the Puget Sound area, along with Vancouver Island (British Columbia), with smaller concentrations in central and southwestern Washington. In Oregon, seismicity is much sparser with smaller clusters occurring in the Willamette Valley and in several parts of southern and offshore central Oregon.

Confirmed and documented swarms were located throughout the region (see Figure 2.10, right. Swarms are concentrated both in areas of high seismicity (Puget Sound), medium seismicity (southwestern Washington) and low seismicity (southeastern Washington). In total, there were 559 events in 73 confirmed/documented swarms (7.14% of the target catalog) and 146 events in 39 possible swarms (1.87% of the target catalog).

Most areas of the PNW have somewhat time-stationary seismicity, which improves slightly when swarms are removed (compare Figures 2.14 and 2.15). This is evidenced by the fairly constant slope in cumulative number plots across the time period, in all areas but Area 3. This area experienced a large cluster of earthquakes (multiple connected aftershock sequences at Klamath Falls, OR) in autumn 1993. Areas 3 and especially 5 have a slower increase of observed seismicity in the early period, likely due to the lower network coverage then.

In fact, there is a discrepancy in network coverage between the northern part of the region (Washington State and northern Oregon) and the southern part (the rest of Oregon), for much of the period. Figures A.2 - A.7 in the Appendix show the locations of PNSN stations at the start of each year (1970-2018); there were no stations south of 45°N until 1980 and station coverage has remained much lower than in the north.

2.8.1 Complete PNW catalogs

Before one can statistically analyze the PNW catalog (especially using any point process methods), it is critical to assess catalog completeness. As defined in Section 1.1, a complete catalog (with regards to a spatial zone, time period and magnitude) is one that contains all of the earthquakes within that spatial zone, time period and above that magnitude. Point process methods typically assume no missing earthquakes, so catalogs must be complete to comply with this.

The difference in station coverage between the northern and southern regions of the PNW suggested that catalog completeness would be different between them. We divided the region by a latitude threshold for simplicity, rather than using the five areas specified above. To determine this latitude cutoff, we examined cumulative number plots for 1° latitude bands from 42-49°N (see Figure A.10 in the Appendix), looking for constant-slope linearity to indicate time stationarity. From 45°N and above, seismicity appeared to increase linearly, with the band from 45-46°N having less of a linear shape early in the period, which may be caused by fewer stations in this zone until about 1981. Thus, we split the region at 45°N. We did not consider a spatial split by longitude, as station coverage seemed similar in the eastern and western regions (above 45°N), likely due to population centers in western Washington and northwestern Oregon and a nuclear power plant in eastern Washington (Malone, 2019; Gomberg et al., 2012).

Given the difference in station coverage, each region will have its own start year and magnitude of completeness. We first assessed the Gutenberg-Richter relationships of the earthquakes in each region. The Gutenberg-Richter law describes the frequency N of earthquakes of magnitude M as

$$\log N(M) = a - bM,$$

with Gutenberg-Richter parameters a and b that must be estimated from a catalog.

This implies that magnitudes are exponentially distributed, which can be checked by comparing the empirical distribution of magnitudes to the theoretical exponential distribution. Consider the magnitude *tail frequency*, $F_n(M) = \#\{j : M_j < M\}/n$, where n is the catalog size. One way of assessing whether a set of earthquakes follows a Gutenberg-Richter relationship is by plotting the simultaneous confidence bands of $\log F_n(M)$ vs. M (see Guttorp (1987) for the formula of the confidence band). If a straight line can fit within the confidence bands, then the catalog can have the Gutenberg-Richter parameters estimated for it (Guttorp, 1987). We show in Figures A.11 and A.12 that the PNW North and South both can have a Gutenberg-Richter law fit to them at various start years.

To determine the time period of completeness, we examined plots of magnitudes over time. Earthquakes in the PNW North did not have sizable gaps in time for lower magnitudes, starting from around 1980 (Figure A.13, top two panels). We show the PNW North's magnitude-frequency distribution for different start years and magnitude cutoffs in Figures A.14-A.15 in the Appendix (filled triangles: cumulative distribution of magnitude frequencies and empty triangles: bin-wise distribution for 0.1 magnitude bins). We look at years corresponding to large increases in the number of stations in the PNW North (Michael, 2014). In each plot, we add the best-fit Gutenberg-Richter line with parameters estimated by MLE (using standard procedures for estimating the MLE (Aki, 1965) and their standard errors (Shi and Bolt, 1982)). We identify completeness by looking for the magnitude at which the catalog frequencies are below the ones predicted by the Gutenberg-Richter line. We also used the maximum curvature method (Wiemer and Wyss, 2000), where the magnitude of completeness is the magnitude bin with the greatest number of earthquakes. Both methods support the catalog being complete down to M=2.0, and potentially lower, perhaps as early as 1970. However, given that the Canadian catalog only begins in 1985 and for consistency with previous studies (Bostock et al., 2019; Gomberg and Bodin, 2021) and expert opinion (Malone, 2020), we use a start year of 1985 and a magnitude of 2.0 for the complete PNW

North.

Completeness is less clear for the PNW South. The magnitude time plot (Figure A.13) shows that earthquake detection was irregular until potentially 2004. Even though more PNSN stations were installed after the Klamath Falls sequence of 1993 (Malone, 2019), there are months-long gaps in earthquake detection in the late 1990s and early 2000s. Whether this is due to a lack of seismicity or just poor earthquake detection would require additional data sources to infer, and is outside the scope of this work. After the Goose Lake sequence of 2004, the distribution of earthquake magnitudes became arguably more regular, though seismicity remained sparse overall.

We examined magnitude frequency plots for the PNW South (Figures A.16-A.19) at different magnitude cutoffs and time periods chosen based on rises in seismicity levels (i.e., separating out the Klamath Falls and Goose Lake sequences). We omit the time before 1993 as station coverage and thus seismicity was very limited. Within each time period, there is greater fluctuation in $\hat{b}$ values based on the magnitude cutoff, than seen in the same plots for the PNW North. As such, it is difficult to settle on a single start year or magnitude of completeness. Based on the trends in Figure A.13 (bottom two panels) and discussions with colleagues (Michael, 2021), we use a start year of 2004 (right before the Goose Lake sequence) and suggest three different magnitudes of completeness (M_c):

1. $M_c = 2.0$: possibly incomplete
2. $M_c = 2.3$: arguably complete
3. $M_c = 2.5$: likely complete

The magnitude frequency distribution for the complete catalogs (Figure A.20) looks very similar between the catalog with swarms (top row) and without swarms (bottom row), suggesting that removing swarms may not greatly change any parameters that represent magnitude relationships.

2.9 Conclusions

In this chapter, we combined three data sources to collect a new earthquake catalog for the PNW, both for a target zone around Washington State and Oregon and for a larger auxiliary zone around it. We used authoritative sources for the US and Canada and merged in a third data source that contained measurement error variables. In order to accomplish this, we automated the detection of discrepancies between the three catalogs, carefully resolving mismatches and removing duplicates. We then performed the first-ever (to our knowledge) study of the completeness of the PNW catalog. We identified the time periods and magnitudes of completeness for two subregions, split based on differences in station coverage between the northern and southern zones.

We furthermore developed several algorithmic approaches to detect spatiotemporal earthquake clusters, in order to remove earthquake swarms prior to building aftershock models on this catalog. We first used a window-based and manual technique; each cluster was then hand-labelled as a likely or definite swarm or aftershock sequence by a seismologist. We also proposed a more rigorous semi-parametric and adaptive cluster identification procedure based on the empirical distributions of the time differences and epicentral distances between catalog events. The clusters detected by this procedure were adequately close to the ones hand-labelled by seismologists, if the right time and space thresholds were chosen. Adjusting these thresholds gives different views on the clustering in the catalog.

A variety of future work should be done on both aims of this chapter. The completeness of this PNW catalog should be more carefully studied, perhaps through a model of completeness (e.g, Page et al. (2016)). We saw evidence of time-varying completeness, especially in the south, which can be accommodated by a time-varying b-value (Guttorp and Hopkins, 1986; van der Elst, 2021). The hand-labelled swarms and sequences could be studied in closer detail to understand how well they conform to the assumptions of the ETAS model.

The cluster identification procedure described in Section 2.6 may further be improved by

imposing restrictions to make it more realistic. For example, the spatial threshold to form clusters should not exceed some value, which may be set by characteristics of the seismic region. It would also be useful to achieve an algorithmic classification of earthquake clusters as swarms or sequences. This may be based on characteristic properties of aftershocks (e.g., Båth's Law (Shearer, 2012), or the seismological laws on which the ETAS model is based, see Chapter 3) and swarms (e.g., their reported difference in b-values (Holtkamp and Brudzinski, 2011) or other statistics discussed in Zaliapin and Ben-Zion (2013b)).

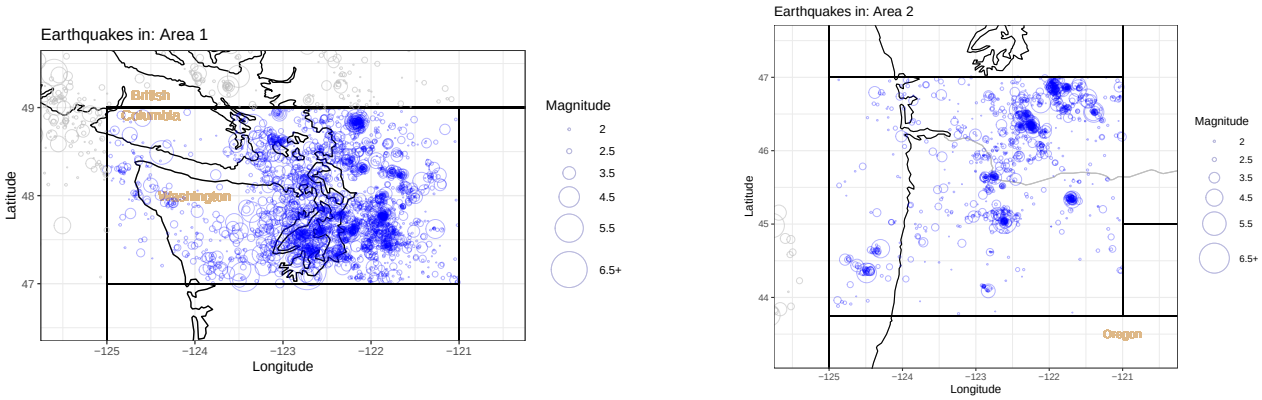

Figure 2.11: Catalog without definite/possible swarms, Areas 1 and 2.

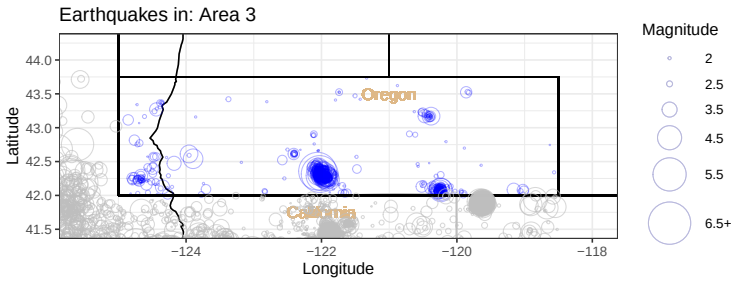

Figure 2.12: Catalog without definite/possible swarms, Area 3.

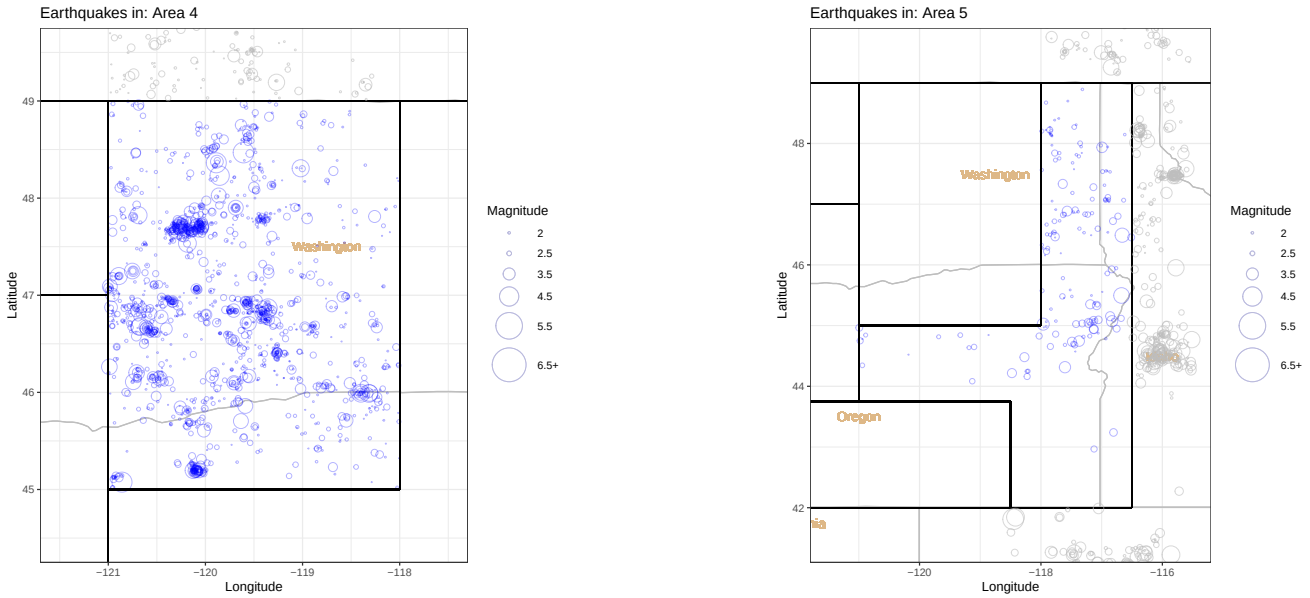

Figure 2.13: Catalog without definite/possible swarms, Areas 4 and 5.

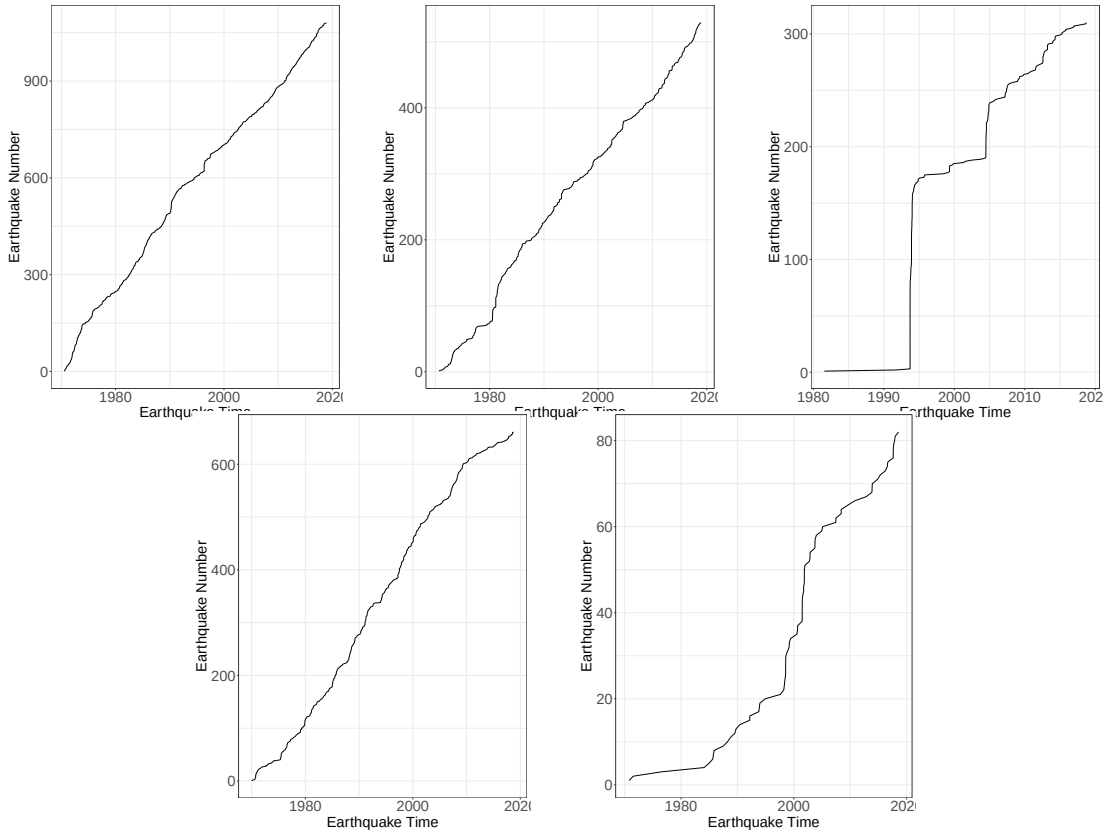

Figure 2.14: Cumulative earthquake counts (including swarms), split by area, in order of Area 1 to Area 5.

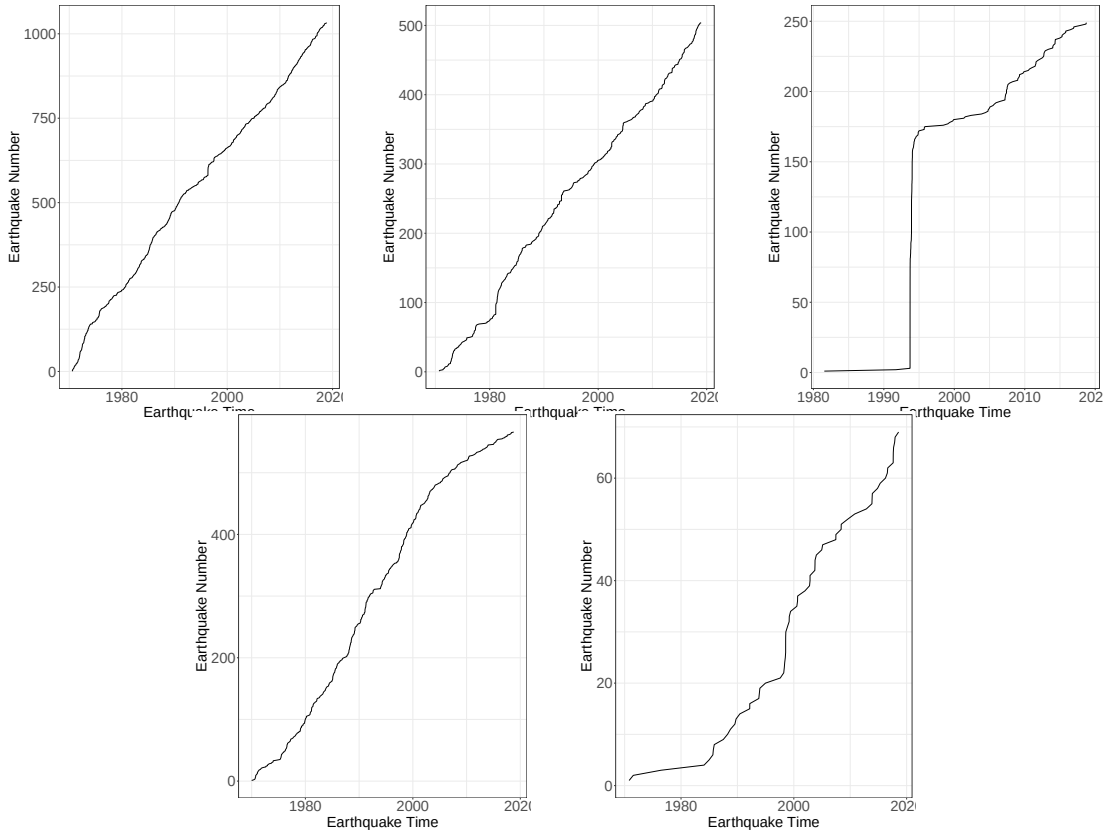

Figure 2.15: Cumulative earthquake counts (with definite and possible swarms removed) for the target catalog only, split by area, in order of Area 1 to Area 5 (target zone only).

Chapter 3

BAYESIAN ETAS: TOWARDS IMPROVED STATISTICAL SEISMICITY MODELS FOR THE PACIFIC NORTHWEST

Look at you -
You look just like a little old earthquake.
(Break it, break it,
You got to break it.)
Ain't no order in this courtroom.

"Dance Apocalyptic," Janelle Monáe (2013)

3.1 Introduction

The high seismic risk of the Pacific Northwest (PNW) can be better mitigated when assessed by an optimal statistical characterization of its earthquake and aftershock dynamics. After collecting a new catalog for the PNW and describing its completeness (Chapter 2), we consider how to best model it to address the specific issues in this region. In this chapter, we describe common statistical models of seismicity and issues in their parameter estimation and uncertainty quantification. We then build a novel solution using Bayesian inference that allows us to quantify how issues in the catalog affect paramter estimates and their uncertainties. We focus on general issues, like catalog measurement error, as well as regional issues, such as the PNW's diverse tectonics and the presence of swarms.

3.1.1 Statistical seismicity models

Statistical seismicity models are key inputs into other earthquake models, from aftershock forecasts to long-term seismic hazard assessment. These have societal importance as they inform emergency planning and response (Goltz, 2015) and seismic building codes (Petersen et al., 2012). Generally, statistical representations of seismicity are based on point process models, which are a type of stochastic models on a domain D; for earthquake models, D can be a time interval $[0, T] \subset \mathbb{R}$ (temporal point process), a spatial plane $S \subset \mathbb{R}^2$ (spatial point process), or space-time $[0, T] \times S \subset \mathbb{R} \times \mathbb{R}^2$ (spatiotemporal point process).

A temporal point process can be specified by its *conditional intensity function*, $\lambda(t|\mathcal{H}_t)$ which is the rate of occurrence of a point at some time t, conditional on $\mathcal{H}_t$, all points occurring up to time t (Guttorp, 1995). More formally, $\lambda(t|\mathcal{H}_t)$ may be defined (Zhuang et al., 2002) as

$$\lambda(t|\mathcal{H}_t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(\text{one event occurs in the interval } [t, t + \Delta t]|\mathcal{H}_t).$$

Conditional intensities may be defined analogously for spatiotemporal point process models. Spatial point processes, which lack a natural ordering like in time, can be specified by the Papangelou conditional intensity instead (Zhuang, 2006). Any point process with a random variable associated with it is called a *marked* point process.

A basic statistical seismicity model can be constructed solely based on the physical fact that earthquakes spatially cluster together near fault lines. A spatially non-homogenuous Poisson point process (NHPP) model is a spatial point process whose (Papangelou) conditional intensity is a function of location, $\lambda = \lambda(x, y)$. In the absence of a functional or parametric form of $\lambda(x, y)$, we may estimate it by taking kernel density estimates of the points over the spatial region (Frankel, 1995; Cucala, 2008).

A NHPP model ignores the fact that earthquakes have different magnitudes, which is of fundamental importance to characterizing seismicity. It also ignores that many earthquakes

are clustered together in time and space through previously-described aftershock sequences. Several seismological laws describing magnitude relationships and aftershock dynamics have been empirically established over the past century (Utsu et al., 1995) and can be incorporated into statistical seismicity models.

The number of aftershocks produced by an earthquake is referred to as its *productivity* and is related to the earthquake's magnitude. The productivity law (Utsu, 1972) states that the number of aftershocks increases exponentially with magnitude. Two parametrizations of this law are common in the literature:

$$N(M) = 10^{a+b(M-M_0)}; \tag{3.1}$$

$$N(M) = K \exp(\alpha(M - M_0)), \tag{3.2}$$

where M_0 is the minimum catalog magnitude, $K = 10^a$ is the average productivity for mainshocks of magnitude M_0, and $\alpha = b\ln(10)$ represents how an increase in magnitude increases productivity.

The Gutenberg-Richter law is related to the productivity law and describes the magnitude-frequency distribution of all earthquakes (aftershocks and non-aftershocks) (Gutenberg and Richter, 1944). It implies that magnitudes are exponentially distributed (Guttorp, 1987) with a probability density of:

$$f(M) = \beta(exp(\beta(M - M_0)),$$

where $\beta = \frac{b}{ln(10)}$.

The modified Omori law (Utsu, 1961) describes the decay of aftershocks in time after the mainshock j's time (t_j) using a power law:

$$h(t - t_j) = (t - t_j + c)^{-p},$$

where p is the temporal rate of decay and c is an offset parameter (usually in days) used to prevent a singularity at time $t = 0$. c is also interpreted as the characteristic time before the power-law decay of aftershocks begins. j indexes the earthquakes in a catalog of size n ($j = 1, \ldots, n$).

The Utsu-Seki law (Utsu and Seki, 1955) describes the decay of aftershocks in space around the mainshock j's location (x_j, y_j), again with a power law:

$$g(x - x_j, y - y_j) = \frac{q-1}{\pi d^{1-q}}((x - x_j)^2 + (y - y_j)^2 + d)^{-q},$$

where q is the spatial rate of decay and d is the spatial offset parameter (we use units km^2)[1]. d is also the characteristic distance before aftershock distances from the mainshock begin to decay. This spatial decay law assumes an *isotropic* (rotation-invariant) spatial distribution of aftershocks around the mainshock and was found to better describe aftershocks in Japan over two other spatial functions (Ogata, 1998).

Epidemic-Type Aftershock Sequence (ETAS) models

The above laws have been incorporated into statistical seismicity models, formulated as marked point process models, with the mark representing magnitude. The Reasenberg-Jones model (Reasenberg and Jones, 1989) is a temporal-only point process model that uses the productivity law in Equation 3.1 and the Omori law. In this model, every earthquake is regarded as a background event (which can trigger an aftershock), or an aftershock; that is, aftershocks are not able to trigger their own aftershocks. However, secondary triggering has been demonstrated to comprise a substantial, if not majority, portion of many catalogs (Felzer et al., 2003).

---

[1]We convert from spherical geographic coordinates (degrees longitude/latitude) to grid-based coordinates (km east-west/north-south) using the Universal Transverse Mercator system, as provided in the latlong2grid function in the R package SpatialEpi (Kim and Wakefield, 2010)

The spatiotemporal Epidemic-Type Aftershock Sequence (ETAS) model was originally proposed in Ogata (1998), though based on a temporal-only model of Ogata (1988). It describes the total seismicity rate $\lambda(t, x, y, M|_t)$, at a given time t and location (x, y), conditional on the history up to that time $(_t)$ as:

$$\lambda(t, x, y, M|_t) = f(M)(\mu_{ST} + \sum_{j}^{t_j<t} k(M_j)h(t - t_j)g(x - x_j, y - y_j)), \tag{3.3}$$

where $k(M_j)$ is the productivity law given in Equation 3.2, and magnitude-frequency distribution function f and decay functions h and g are as given in the preceding section. This is a Hawkes-style self-exciting point process model (Hawkes, 1971), meaning an aftershock may trigger its own aftershocks. To ensure stationarity, we assume $\int_D \int_{M_0}^{\infty} \int \sum_j^{t_j<t} k(M_j)h(t - t_j)g(x - x_j, y - y_j)dMdxdydt < 1$, called the *criticality condition* (Veen and Schoenberg, 2008). Most authors omit the $f(M)$ term in describing the conditional intensity to shorten notation and refer to the marked conditional intensity $\lambda(t, x, y, M|_t)$ simply as $\lambda(t, x, y|_t)$.

The ETAS model has been applied in hundreds of studies and has been extended in various ways (Reinhart, 2018). As earthquakes occur along fault lines, background seismicity is known to vary spatially and many models exist where $\mu_{ST} = \mu_{ST}(x, y)$ is estimated in a spatially-varying way. The offset parameter d in the Utsu-Seki law is also often assumed to vary with mainshock magnitude (the larger the mainshock, the larger the zone around it before Utsu-Seki aftershock decay begins).

Regardless of the model parameterization, there are several important model assumptions that the catalog has to satisfy. The ETAS model assumes that background seismicity is stationary with respect to time, meaning that sudden increases in seismicity are associated with aftershock sequences and not with natural variation in the processes that cause background events. In addition, catalogs are assumed to be *complete*, meaning that within their spatiotemporal range, all earthquakes above M_c, the magnitude of completeness are

included in the catalog (that is, have been detected by the seismic network). The standard ETAS model thus assumes no earthquake missingness.

The ETAS model has become the most frequently used statistical aftershock model. It has out-performed other statistical and physical seismicity models in earthquake forecast evaluations (Schorlemmer et al., 2018). It continues to be used frequently, not just by researchers, but also by government authorities such as Italy's National Institute of Geophysics and Volcanology (Marzocchi et al., 2014) and the United States Geological Survey (USGS) (Milner et al., 2020).

3.1.2 Statistical issues with the ETAS likelihood

The ETAS model considers seismicity as composed of two distinct components: background seismicity and all seismicity triggered by a previous earthquake (the sum term in Equation 3.3). Some authors use rules based on spatial and temporal windows to identify aftershock sequences in a catalog and apply the ETAS model individually to each sequence (Zhang et al., 2020; Guo and Ogata, 1997); such a window-based approach is common in the analysis of seismicity (Gomberg and Bodin, 2021; Page et al., 2016; Reasenberg, 1985). However, such deterministic splits of the catalog ignore the fact that there is never certainty about which events belong to a given sequence. Methods also exist for a probabilistic assignment of every earthquake into the background set or an aftershock sequence (Zhuang et al., 2002). These better reflect the inherent uncertainty in the classification of aftershocks in a catalog. This approach is more suitable to models of full catalogs, where the ability to disentangle the exact partition of the catalog into the background set and sets of sequences is limited (Seif et al., 2017; Utsu et al., 1995).

Given a deterministic or probabilistic split of the catalog between background and aftershock events, ETAS parameters $\theta = \{\mu, K, \alpha, c, p, d, q\}$ may be estimated. Estimates of their parameters (and their accompanying uncertanties) are key not just in describing a region's

aftershock patterns, but in forecasting aftershocks after large earthquakes (Michael, 2018; Llenos and Michael, 2019; Hardebeck et al., 2019; Michael et al., 2020). Parameters are typically estimated by maximizing the ETAS log-likelihood:

$$l(\theta|_T) = \sum_{i=1}^{n} \log(\lambda(t_i, x_i, y_i|_{t_i})) - \int_0^T \int \int_S \lambda(t, x, y|_t) dx dy dt.$$

The maximum (log-)likelihood estimator (MLE) for the ETAS model has been shown to be asymptotically unbiased, asymptotically normal and efficient (Ogata, 1978), even under some misspecification of the intensity function (Schoenberg, 2016). But several problems have been documented with using this likelihood to perform inference on real-world finite-sample catalogs. Simply evaluating the multi-dimensional likelihood function is non-trivial, particularly the integral term, for which no closed-form expression exists, forcing researchers to use either numerical solutions (Ogata, 1998; Lippiello et al., 2014) or other approximations (Schoenberg, 2013). Optimizing the likelihood function similarly requires iterative numerical procedures, with final maximum likelihood estimates (MLEs) being those that result in the likelihood iteratively converging to its maximum. But the likelihood is known to be flat near the optima for most ETAS parameters (Veen and Schoenberg, 2008). Furthermore, parameter estimates are correlated between several parameter pairs (Chu et al., 2011; Guo and Ogata, 1997), leading to a likelihood surface that is multimodal (Lombardi, 2015; Veen and Schoenberg, 2008). All these issues make finite-sample ETAS MLEs volatile but also biased in ways that are still unclear (Seif et al., 2017; Harte, 2013). MLEs are often sensitive to both the numerical optimization procedure used or the initial values used by it.

Furthermore, quantifying uncertainty around these parameter estimates is typically done with the inverse Hessian of the likelihood function. This is rooted in asymptotic statistical reasoning, which time-limited earthquake catalogs rarely satisfy. It also requires further numerical procedures or approximations for matrix inversion, which are again problematic even for large catalogs. As such, uncertainty quantification about ETAS parameter esti-

mates based on the traditional likelihood is unstable. It may even be unattainable for small catalogs with complex seismicity (Harte, 2018), as the numerical procedures involved may fail to converge (Veen and Schoenberg, 2008). Despite these limitations, the vast majority of published ETAS uncertainties are based on the Hessian (e.g, Zhuang (2011); Zhang et al. (2020); Seif et al. (2017); Schoenberg (2013), among many others).

These statistical challenges with the ETAS likelihood have been reported for decades and researchers have made different attempts to ameliorate them. In Veen and Schoenberg (2008), the authors utilize the (unknown) *branching structure* of the catalog, $B = \{B_i\}$, for catalog earthquakes $i = 1, \ldots, n$, where

$$B_i = \begin{cases} 0 \text{ if earthquake } i \text{ was produced by background process} \\ j \text{ if earthquake } i \text{ was triggered by previous earthquake } j. \end{cases}$$

Conditional on values for B_i, they derive a new likelihood function, which we term the *branching likelihood*. ETAS parameters can then be estimated in a iterative procedure: first, estimate some $\hat{B}$ based on initial guesses for the ETAS parameters θ. Then, iteratively maximize the branching likelihood to obtain ETAS parameter estimates $\hat{\theta}$ and then update $\hat{B}$ based on the new estimates $\hat{\theta}$. Iterate the previous step until estimates $\hat{\theta}$ converge to a final solution.

Veen and Schoenberg show that MLEs based on this estimation prcoedure are less biased and more robust to choice of initial values than traditional MLEs; however, maximizing even this simpler likelihood still requires a numerical procedure and still has similar issues with parameter flatness near optima. Other authors have proposed using different numerical procedures to maximize the traditional ETAS likelihood (Lombardi, 2015; Kasahara et al., 2016), but they do not address the fundamental challenges inherent in optimizing this function.

3.1.3 Bayesian inference for parameter estimation

Bayesian inference is a statistical paradigm for parameter estimation that eschews maximizing a model's likelihood in order to find parameter estimates (Reich and Ghosh, 2019). Rather, it considers the parameter itself as a random variable that possesses a probability distribution. Sampling from this *posterior* distribution requires specifying the prior knowledge on this parameter (also in the form of a distribution) and then scaling it by the model likelihood. Bayesian inference is available for spatiotemporal point process models and is often applied to studies of natural events (Guttorp and Thorarinsdottir, 2012). This framework has not been entirely ignored in previous research with the ETAS model. Several studies have investigated how aftershocks forecasts built using ETAS can be updated using the Bayesian approach during an ongoing sequence (Ebrahimian and Jalayer, 2017; Omi et al., 2015), but these require the initial computation of parameter MLEs.

Recent work has tackled the Bayesian estimation of ETAS parameters from full catalogs. Ross (2017) and Kolev and Ross (2019) work with the temporal-only ETAS model and, similar to Veen and Schoenberg (2008), derive a branching likelihood that does not have many of the statistical issues in the traditional likelihood. They use a Monte Carlo Markov Chain routine to sample from the parameter posterior distributions to more rigorously quantify uncertainty in the parameter estimates. Ross (2021) (published at the time of the writing of this dissertation) extends this to the spatiotemporal ETAS model, though uses non-informative priors that are unrealistic for real-world catalogs.

We expand on this foundation by customizing a Bayesian inference procedure to the specific needs of a new catalog for the Pacific Northwest region. We show how our method achieves robust and stable parameter estimates, compared to MLEs. We also demonstrate its ability to quantify how different sources of uncertainty in the PNW affect parameter estimates.

3.1.4 Pacific Northwest seismicity

The Pacific Northwest of North America has high seismic risk due to a diversity of fault systems, some of which lie under population centers (Bostock et al., 2019). Statistical seismicity modelling for the Pacific Northwest is complicated by the complex tectonics of the region. The Juan de Fuca plate subducts under the North American continental plate and the resulting fault zone is capable of generating massive but very infrequent megathrust events (Wirth and Frankel, 2019). The subducting slab itself produces deep, intraslab earthquakes which are hypothesized to produce few aftershocks. There are also numerous shallow faults located in the crust of the North American plate that produce crustal earthquakes that occur near densely populated areas. These crustal earthquakes are also thought to have more productive aftershocks than intraslab earthquakes.

Furthermore, the region is known to produce earthquake swarms, which are spatiotemporal clusters of earthquakes that are not triggered by a previous earthquake as in aftershock sequences; rather, they occur because of aseismic sources, such as from underground fluids (volcanic magma or groundwater) or anthropogenic sources (such as wastewater injection). Swarms are also known to following different patterns of triggering, i.e., not having Omori-type temporal decay. The many volcanoes of the PNW contribute to the presence of swarms in this catalog (Bostock et al., 2019).

Even though the region's seismicity is complex, the PNW does not have a large set of contemporary instrumentally-measured earthquakes (Malone, 2019), compared to other seismic regions. Due to sizable seismicity at the US/Canada border, any catalog for the PNW region must consider earthquakes recorded by both US and Canadian seismic networks. Combined international PNW catalogs have been quantitatively studied (e.g., Bostock et al. (2019)), though almost no previous studies have been made of aftershock patterns across the PNW. The one recent exception found that intraslab events do indeed appear to have lower productivities than crustal mainshocks (Gomberg and Bodin, 2021). This study was based

on the Reasenberg-Jones model, with sequence-specific parameters fit using MLE. Given the importance of aftershock forecasts to emergency managers and other potential users (Gomberg and Jakobitz, 2013), the spatiotemporal ETAS model may better characterize and thus forecast aftershock patterns in the PNW. But the heterogenuous and limited catalog problematizes any traditional MLE-based ETAS model for the region. Our methodology for Bayesian inference for the ETAS model will thus be pivotal to estimate aftershock parameters together with their uncertainties for the PNW.

3.2 Bayesian estimation for ETAS

This section describes the traditional and branching likelihoods for the spatiotemporal ETAS model and details challenges in using these for parameter estimation. We then present the Bayesian inference procedure to overcome these challenges and estimate parameters and quantify their uncertainty.

3.2.1 Spatiotemporal ETAS likelihood

Consider $l(\theta|Z)$, the traditional log-likelihood function for the spatiotemporal ETAS model for a catalog Z, as given in Ogata (1998):

$$l(\theta|Z) = \sum_{i=1}^{n} \log(\lambda(t_i, x_i, y_i|_{t_i})) - \int_0^T \int \int_S \lambda(t, x, y|_t) dx dy dt. \tag{3.4}$$

The integral term of equation 3.4 can be rewritten as:

$$\int_0^T \int\int_S \lambda(t,x,y|_t)dxdydt = \mu T|S| + \int_0^T \int\int_S \int_{M_0}^{\infty} N(ds,d\xi,d\eta,dM)\cdot \int_0^{T-s} \int\int_{S-(\xi,\eta)} a(t,x,y,M)dtdxdy$$

$$INT = INT_b + INT_{trig},$$

where $ds, d\xi, d\eta, dM$ refers to an infinitesimal interval in time-longitude-latitude-magnitude, the triggering function $a(t,x,y,M) = K\exp(\alpha(M_j - M_0))(t-t_j+c)^{-p}\frac{q-1}{\pi d^{1-q}}((x-x_j)^2+(y-y_j)^2+d)^{-q}$, T is the catalog period, S is the two-dimensional catalog spatial zone and $|S|$ is its area. In this model and for the rest of this chapter, we assume a constant background rate in space, $\mu_{ST}(x,y) = \mu\frac{1}{|S|}$.

We can decompose INT into the background component (b) and triggered component ($trig$). In INT_{trig}, the first integral term can be interpreted as "how many triggered earthquakes occur in an infinitesimal interval, when integrated over all of time, space and magnitude?" and the second integral term as "what is the contribution to the total triggered intensity from these earthquakes?". The first integral in INT_{trig} thus translates to a sum over all earthquakes in the catalog as all earthquakes are regarded as being potentially triggered in this probabilistic assignment (Zhuang et al., 2002). We can thus rewrite INT_{trig} as

$$INT_{trig} = \sum_{j=1}^{n} \int_0^{T-t_j} \int\int_{S^{(j)}} a(t_j,x_j,y_j,M_j)dtdxdy,$$

with $S^{(j)} = \{(x-x_j, y-y_j) : (x,y) \in S\}$. This can be split into its temporal and spatial components:

$$INT_{trig} = \sum_{j=1}^{n} \int_0^{T-t_j} \frac{K\exp(\alpha(M_j-M_0))}{(t+c)^p}dt \int\int_{S^{(j)}} \frac{q-1}{\pi d^{1-q}}\frac{1}{(x^2+y^2+d)^q}dxdy.$$

The temporal integral, denoted $H(T-t_j)$, can be analytically derived (derivation provided in Appendix, Section B.6) as:

$$\begin{aligned} H(T-t_j) &= \int_0^{T-t_j} \frac{K\exp(\alpha(M_j-M_0))}{(t+c)^p}dt \\ &= K\exp(\alpha(M_j-M_0))\Big(\frac{1}{1-p}((T-t_j+c)^{1-p}-(c)^{1-p}\Big). \end{aligned}$$

In this derivation, as in the rest of this dissertation, we ignore the case when $p=1$, which requires a slightly different functional form (Ogata, 1983).

The spatial integral in INT_{trig}, denoted $G(S^{(j)})$ and abbreviated G, does not have a direct analytical solution and thus poses a particular challenge in the evaluation of the likelihood function. We address this in the following section.

3.2.2 Evaluating the spatial integral

Different authors have proposed different solutions to the spatial integral in the ETAS likelihood. Ogata (1998) originally suggested to evaluate G by radially approximating it across the catalog zone, and this was adopted by other authors (Jalilian, 2019; Chiodi and Adelfio, 2021). Section B.1 in the Appendix provides a comprehensive explanation of this radial approximation and some other approaches to handling this integral. The radial approximation relies on several arbitrary choices on how to split the region and is computationally expensive (Veen and Schoenberg, 2008). The most common approach in the literature has become simply to set $G=1$, allowing to ignore it (Schoenberg, 2013). While this is more computationally efficient and does not require arbitrary choices, it can bias parameter estimates for seismic regions with seismicity near its borders, which is the case for the PNW (see Figure 3.7). We required a different approach to evaluate G for ETAS modelling of the PNW.

It is possible to numerically approximate the solution to the two-dimensional integral (though this comes with challenges, see additional remarks in Appendix, Section B.1). An

easier approach starts by first analytically solving for G in one dimension (say, x). This yields a function that is an integral over another dimension (y). This latter integral does not have a closed form solution but can be numerically evaluated. Such an analytical/numerical approach to solving difficult integrals has been mentioned in previous earthquake modelling studies (e.g., Werner et al. (2011)). Specifically, solving the (indefinite) spatial integral, G, in one dimension yields:

$$\begin{aligned} G &= \int (x^2 + y^2 + d)^{-q} dx \\ &= x(d + x^2 + y^2)^{-q}(1 + \frac{x^2}{d + y^2})^q \cdot {}_2F_1(\frac{1}{2}, q; \frac{3}{2}; \frac{-x^2}{y^2 + d}), \end{aligned}$$

where ${}_2F_1$ is the hyper-geometric function that results from integrating a function with a negative power.

We can now rewrite the two-dimensional spatial integral as:

$$\begin{aligned} &\int_{S_y^{(j)}} \int_{S_x^{(j)}} (x^2 + y^2 + d)^{-q} dx dy \\ &= \int_{S_y^{(j)}} x(d + x^2 + y^2)^{-q}(1 + \frac{x^2}{d + y^2})^q \cdot {}_2F_1(\frac{1}{2}, q; \frac{3}{2}; \frac{-x^2}{y^2 + d})|_{S_x^{(j)}} dy, \end{aligned}$$

where $S_x^{(j)}$ and $S_y^{(j)}$ are the range of the spatial zone in each direction.

This expression can be evaluated by one-dimensional numerical integration and the standard approach is via some quadrature rule. We use Gaussian-Kronrod quadrature (Fausett (2008), in R's function "integrate") with a convergence tolerance of 10^{-6} for each earthquake.

Computing an updated G value for each earthquake for each proposed d/q combination would be computationally costly, especially within an iterative procedure requiring this computation many times (see Section 3.2.4). Our solution is to first compute G for a somewhat

Optimizer	Description	Reference
NLM	Newton-type gradient descent, with several algorithms to find step size and approximate Hessian	(Schnabel et al., 1985)
BFGS	Quasi-Newton gradient descent based on an initial approximation of the Hessian	(Fletcher, 1987)
Nelder-Mead (NM)	Downhill simplex method	(Nelder and Mead, 1965)
Conjugate Gradient (CG)	Gradient descent method that does not require the Hessian but requires finding "conjugate" orthogonal vectors over which to perform gradient descent	(Fletcher and Reeves, 1964)
L-BFGS-B	BFGS method but done with constraints and uses less memory	(Byrd et al., 1995)

Table 3.1: Optimization procedures commonly used for numerical maximization.

fine grid of d/q values. This is done for every earthquake in the catalog, since the product in the INT_{trig} term is evaluated earthquake-wise. The grids used in the synthetic catalog analysis are given in Table B.1 in Appendix Section B.1.3. We then interpolate from the grid of G values for a specific proposed d/q combination. We use bilinear interpolation because of observed smoothness in the G surfaces for several synthetic catalogs. Figure B.2 in Appendix Section shows G surfaces for two earthquakes, near the spatial boundary (left plot, lower G values) and in the middle of the region (right plot, higher G values, as there is more space to integrate over).

The traditional approach to ETAS inference is then based on maximizing the ETAS

likelihood, typically with a numerical gradient descent procedure. The modeller must choose their numerical optimizer and a set of initial values to initialize it with. A variety of optimizers exist and we describe the procedures readily available in the optim() and nlm() functions in R in Table 3.1.

Standard errors around parameter MLEs are estimated assuming asymptotic normality, which was established in Ogata (1978). We first compute $I(\theta)$, the Fisher information matrix, or the the Hessian of the likelihood function:

$$I(\theta) = -\mathbb{E}[\frac{\delta^2 l(\theta|Y)}{\delta\theta\delta\theta^T}].$$

Since $\text{var}(\hat{\theta}_{MLE}) = I(\hat{\theta}_{MLE})^{-1}$, standard errors are the square root of the diagonal elements of the $I(\hat{\theta}_{MLE})^{-1}$ matrix.

We can summarize the inference from a given MLE by computing the asymptotic 95% confidence interval for a parameter θ as

$$\hat{\theta}_{MLE} \pm 1.96 \cdot SE(\hat{\theta}_{MLE}).$$

Another approach for uncertainty quantification is to simulate synthetic catalogs from the ETAS model (see Section 3.5), using MLEs as the parameter values (Wang et al., 2010). The standard error is then taken as the root mean square of the differences between parameter estimates and the MLE, across the simulations. This estimator of the standard error has been shown to be less biased than the Hessian-based estimator when applied to finite-time catalogs; however, it still assumes that computed MLEs are the optimal set of parameters, which cannot be guaranteed for small catalogs like the PNW.

3.2.3 Branching likelihood for ETAS

As described in Section 3.1.2, maximizing the traditional ETAS likelihood can be numerically unstable due to likelihood flatness and multimodality. We introduce here a simplified

likelihood function, conditioned on a branching structure for the catalog, which does not have the same degree of parameter correlation.

Assume that we know B as given in Section 1.2 for our catalog of interest. Conditional on this B, a branching likelihood $L(\theta|Z, B)$ can be derived (we provide this derivation in Section B.2). This branching likelihood was originally proposed by Veen and Schoenberg (2008) and further developed in the Bayesian context by Ross (2017) and Ross (2021). For the spatiotemporal ETAS model, this branching likelihood is

$$\begin{aligned} L(Y|\theta, B) = &\exp(-\mu_{ST}T)\prod_{j=1}^{|A_0|}\mu_{ST} \\ &\prod_{j=1}^{n}\Big(\exp(-k(M_j)H(T-t_j)G((x-x_j, y-y_j))k(M_j)^{|A_j|} \\ &\prod_{t_i\in A_j} h(t_i-t_j)g(x_i-x_j, y_i-y_j)\Big), \end{aligned}$$

where

$$\begin{aligned} H(T-t_j) &= \int_0^{T-t_j} h(t)dt, \\ G(x-x_j, y-y_j) &= \int\int_{S^{(j)}} g(x-x_j, y-y_j)dxdy, \end{aligned}$$

A_j is the set of all earthquake triggered by earthquake j and A_0 is the set of all earthquake not triggered by a previous earthquake (background events) and the $|\cdot|$ notation represents the number of earthquakes in that set. In the branching likelihood, the first two terms are the contribution of the background earthquakes and the terms inside the $\prod_{j=1}^{n}$ are the aftershock contribution.

We will work with the log-likelihood $l(\theta|Z,B) = \log(L(\theta|Z,B))$:

$$l(\theta|Z,B) = |A_0|\log(\mu_{ST}) + \sum_{j=1}^{n} |A_j|\log(k(M_j)) + \sum_{t_i \in A_j} \log(h(t_i - t_j)) + \log(g(x_i - x_j, y_i - y_j)) \tag{3.5}$$

$$+ (-\mu_{ST} T|S|) + \sum_{j=1}^{n} -k(M_j)H(T - t_j)G(x - x_j, y - y_j).$$

There is only one expression in this function that combines different model components (the final sum term) and none that combines the background and triggered components. This is because when we condition on B, we are able to separate out the background and triggered events, and so the background and triggered components of the likelihood can be evaluated separately. Compare this to the traditional likelihood function (equation 3.4), in which all earthquakes are considered as probabilistically being in the background or triggered sets. In equation 3.4, both the sum and integral term include the background component as well as all three triggered components, which increases parameter correlation in the function, making optimization more difficult by increasing multimodality in the likeihood.

Figure 3.1 shows a surface for $l(\theta|Z,B)$ with varying K and α near their true values for a synthetic catalog (Catalog 2A, see Section 3.5). Figure 3.2 shows a surface for $l(\theta|Z,B)$ with varying d and q near their true values for a different synthetic catalog (Catalog 3D).

3.2.4 Improved ETAS inference with Bayesian statistics

The frequentist inference framework described in Section 3.2.2 is predicated on the statistical properties of the likelihood function, which are only valid asymptotically. Earthquake catalogs are by definition time-limited and authors have pointed out that MLEs and Hessian-based standard errors are not unbiased for finite-time catalogs (Wang et al., 2010; Seif et al.,

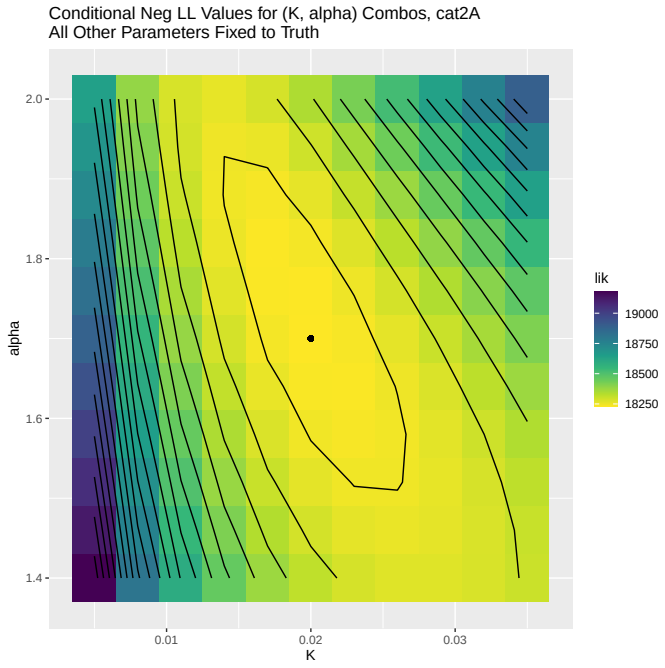

Figure 3.1: Branching likelihood surface for Catalog 2A, varying K and α near their true values and keeping other parameters fixed to their true values. Note that this shows the negative log-likelihood values (lowest values are the maximal values of the log-likelihood values). Contours for every 50 values of the likelihood. The point marks the true K and α for Catalog 2A.

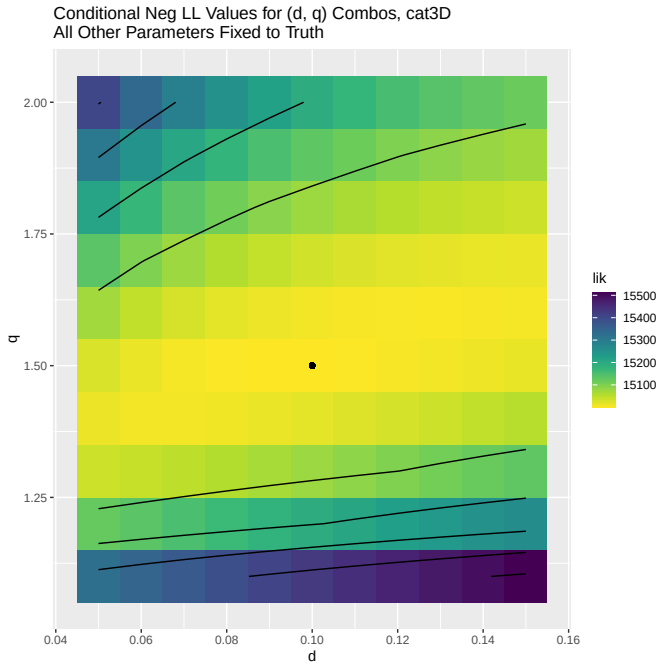

Figure 3.2: Branching likelihood surface for Catalog 3D, varying d and q near their true values and keeping other parameters fixed to their true values. Contours for every 100 values of the likelihood.

2017). Fitting an ETAS model to a small catalog by MLE has been documented to lead to both biased and unstable results (Harte, 2018; Wang et al., 2010; Schoenberg, 2013). Furthermore, authors have called for greater study of how sources of uncertainty in earthquake catalogs affect ETAS estimates (Zhuang et al., 2002; Seif et al., 2017), despite the known issues in quantifying uncertainty with likelihood Hessians.

Bayesian statistical inference is an alternative paradigm that can address these methodological issues. It is based on drawing samples from the posterior distribution for a parameter of interest, θ. The parameter's posterior $p(\theta)$ is proportional to its prior distribution $\pi(\theta)$ (our initial beliefs about the values the parameter can take before modeling the catalog) scaled by the likelihood function (which expresses how likely any given parameter value is, given the catalog and our model). That is, the catalog enters the model through the likelihood and only adjusts the prior to the extent that the catalog contains information about the parameter. We use the branching likelihood $L(\theta|Z, B)$ with the spatial integral term as described in Section 3.2.2.

We will work with the logged posteriors, so

$$p(\theta) \propto \pi(\theta)L(\theta|Z, B)$$
$$\log(p(\theta)) \propto \log(\pi(\theta)) + l(\theta|Z, B).$$

We can re-express $l(\theta|Z, B)$ from equation 3.5 as:

$$\begin{aligned} l(\theta|Z, B) =& |A_0| \log(\mu_{ST}) - \mu_{ST} T|S| \\ &+ \sum_{j=1}^{n} -k(M_j)H(T - t_j)G(x - x_j, y - y_j) \\ &+ \sum_{j=1}^{n} \Big(|A_j| \log(k(M_j)) + \sum_{t_i \in A_j} \log(h(t_i - t_j)) + \log(g(x_i - x_j, y_i - y_j)) \Big). \end{aligned}$$

We will pair the parameters belonging to each model component and sample from their conditional posteriors, conditional on fixed values for the other ETAS parameters (Reich and Ghosh, 2019). The (log-)posterior density functions for each parameter pair are derived in Appendix Section B.3. We use a Gibbs sampler procedure (Marin and Robert, 2007) to iteratively draw samples from each conditional posterior distribution, fixing all other parameters to their current posterior draw.

For most parameters, we take the approach of drawing samples from the parameter's conditional posterior distribution with the Metropolis-Hastings sampler (Robert and Casella, 1999), a type of Markov Chain Monte Carlo sampler. This works by starting with an initial value for the parameter and computing the posterior function for it. A proposed value for the parameter is drawn from a *proposal distribution* and the posterior function is computed for it. The proposal value is accepted randomly: if the ratio of the posterior values is above a random number from the Uniform(0, 1) distribution, then the proposed value is accepted as a draw from the posterior distribution; otherwise, it is rejected and the starting value is taken as a draw from the posterior. If the proposal distribution is chosen to be a Gaussian distribution centered at the previously drawn posterior value, this sampler amounts to a random walk (Reich and Ghosh, 2019).

We follow these steps to perform Bayesian inference on the parameters of the ETAS model:

1. Set initial values for B and ETAS parameters $\theta = \{\mu, K, \alpha, c, p, d, q\}$. Again, we assume a spatially constant background rate so $\mu_{ST}(x, y) = \mu\frac{1}{|S|}$.

2. Take a uniform prior distribution on each B_i (each previous earthquake has equal probability of triggering earthquake i, and it is as probable to be a background event). Sample from the posterior of each B_i by randomly drawing an integer from 0 to j

$(j < i)$ with the following probability weights:

$$P(B_i = j|Y,\theta) = \begin{cases} \frac{\mu_{ST}}{\mu_{ST}+\sum_{j=1}^{i-1} k(M_j)h(t_j)g(x-x_j,y-y_j)} & \text{if j} = 0 \\ \frac{k(M_j)h(t_j)g(x-x_j,y-y_j)}{\mu_{ST}+\sum_{j=1}^{i-1} k(M_j)h(t_j)g(x-x_j,y-y_j)} & \text{if } j \in 1,2,\ldots,i-1. \end{cases}$$

This is equivalent to sampling from a discrete Uniform distribution with probabilities given by $P(B_i = j|Y,\theta)$.

3. Using the information from B, sample from the posteriors of ETAS parameters:

   - For μ, use a Gamma prior with pre-specified hyperparameters. This allows for a conjugate Gamma posterior, which can be directly sampled.

   - For all other ETAS parameters, use a random walk Metropolis-Hastings algorithm to sample from their conditional posteriors, in the following pairs:

     – (K, α)

     – (c, p)

     – (d, q)

     We use a low standard deviation (σ=0.05) in our Gaussian proposal distribution, though posterior results were insensitive to this choice.

4. Repeat steps 2 and 3, updating B with new posterior draws for θ and then the estimation of θ using the updated B, for a fixed number of iterations (e.g., 10,000).

We implemented this computationally as a set of C++ functions that are used within R. Ross (2021) demonstrated this procedure to have reasonable computational cost for medium- or small-sized catalogs. We can summarize the posterior distributions for a given parameter θ with $100 \cdot w\%$ *credible intervals*, defined to be the interval C such that

Parameter	Uninformative Prior
B_i	Uniform on previous events
μ_{ST}	Gamma(0.1, 0.1)
K	Uniform(0, 10)
α	Uniform(0, 10)
c	Uniform(0, 8)
p	Uniform(0, 8)
d	Uniform(0, 8)
q	Uniform(1, 8)

Table 3.2: Priors for spatiotemporal ETAS parameters taken from Ross (2021).

$$\int_C p(\theta|Y)d\theta = w.$$

There are many intervals C that can satisfy this equation and we use *equal-tailed credible intervals*, which represent an interval spanning the middle $100 \cdot w\%$ of the posterior distribution density. These can be easily retrieved from the posterior sample, e.g., as the 2.5th and 97.5th percentiles of the posterior samples when $w = 0.95$. The Bayesian credible interval is philosphically different from the frequentist confidence interval in considering the parameter to be a random variable, rather than a fixed object. Thus, an equal-tailed credible interval can be compared to a frequentist confidence interval for the purposes of summarizing the underlying distribution's central tendency and shape.

3.3 Prior elicitation for ETAS

The Bayesian paradigm requires the setting of prior distributions for each parameter being estimated. This allows the modeller to express beliefs held about the parameter, prior to

fitting it to the catalog. Posterior estimates thus represent the extent to which the catalog (represented by the likelihood) has updated the beliefs held prior to the model. Previous applications of Bayesian estimation for the ETAS model have used broad and uninformative priors (Ross (2021); Kolev and Ross (2019), see Table 3.2), despite calls for the importance of regional variation in aftershock dynamics (e.g., Page et al. (2016)).

This is especially important for modelling PNW seismicity. As previously mentioned, the PNW does not have the amount of data of many other seismic regions (e.g., California), and has numerous physical processes (e.g., different tectonics, clustering types) reacting with a variety of unknowns (e.g., the degree to which seismicity from different tectonic regimes are physically linked). This makes it an open question to what degree a model of the data can expand on previously held views of the PNW's aftershock dynamics. Furthermore, there exist varying beliefs about aftershock dynamics in the PNW, which can be represented by different sets of priors. The Bayesian inference approach thus allows to explore the level to which beliefs about the PNW are consistent with the data, as expressed through the ETAS model.

3.3.1 Eliciting PNW-specific priors

Prior beliefs about parameter values can be elicited by interviewing experts about the values that the parameter might take. A range of methods exist, from asking directly for summary statistics of the prior distribution (such as, e.g., medians, modes, ranges) to having experts draw probability distribution or having them bet on the parameter value being in different intervals (Johnson et al., 2010). We used the bisection method (Oakley, 2010), wherein the expert is first asked to specify the probable lower and upper limits of the parameter range, l and u, respectively. The question to the expert is framed as "what values could be ruled out and what values could not?" to avoid anchoring on a particular value, per the literature on anchoring and heuristics (Garthwaite et al., 2005)). They are then asked to provide a

number m_{50} such that there is a 50% probability of the parameter being less than that value and the same probability that the parameter is above that value. They are then asked to split the interval from l to m_{50} by another value m_{25}, where there is an equal chance of the value being in $[l,\, m_{25})$ and in $[m_{25},\, m_{50})$. The next question yields the split into the intervals $[m_{50},\, m_{75})$ and $[m_{75},\, u]$. We can fit parametric distributions to the resulting five-number summary $\{l, m_{25}, m_{50}, m_{75}, u\}$ to use as prior distributions.

To elicit priors for ETAS parameters, we apply the bisection method with the following procedure.

1. **Elicitation phase.** We began by describing the components of the ETAS model and their parameters.

   We directly elicited values for the ETAS parameters for the μ, c and d parameters, since these have immediate practical interpretations. For the other parameters, we instead elicited values for the following aftershock characteristics: number of aftershocks (for parameters K and α), aftershock duration (for parameter p), and aftershock zones (for parameter q) for a reference M6 mainshock, separately for the two dominant tectonic regimes (deep intraslab events and shallow crustal events) of the PNW.

   For each parameter or aftershock characteristic, we attempted to elicit the five-number summary described above.

2. **Prior construction phase.** In general, when we had five-number summaries to represent prior beliefs, we specified priors using the Gamma distribution. The interviewed experts indicated that a right-skewed distribution, such as a Gamma, matches their prior beliefs about PNW aftershock patterns. For parameters where experts could only provide lower and upper limits and had no beliefs about which values in between were more probable, we specified Uniform priors. We developed priors for each parameter separately between the two tectonic regimes.

We constructed Gamma priors for c and d directly from an elicited five-number summary using the web tool of Morris et al. (2014).

We constructed Gamma priors for the K parameter, by first fixing α to prototypical values (α=1.7 (so $b < 1$) for deep events, and $\alpha = \log(10)$ (so $b = 1$) for crustal events). We calculated the expected productivity under the ETAS productivity law (equation 3.1) over a grid of K values and found the values that came closest to each of the five values summarizing the expected aftershock counts. We constructed a Gamma prior for K based on these five values, separately for each tectonic regime.

We constructed priors for the α parameter by examining grids of the expected productivity from different combinations of α and the span of possible Ks given above, for two reference earthquakes: $M = 5$ and $M = 6$. We found the minimum/maximum α values that resulted in the minimum/maximum aftershock counts under the range of K values specified in its prior for the two reference earthquakes. We took these as limits of a Uniform distribution, as there was no prior belief on which α values were most probable.

We constructed Uniform priors for p using prior beliefs (five-number summaries) on aftershock duration. These were based on defining an aftershock sequence's duration, t^*, as the time it takes for the aftershock rate to equal the background rate. We derived an equation for this from the ETAS model (derivation in Appendix, Section B.4)), which depends on p, c, K, α and μ. We computed t^* over combinations of minimum and maximum values for c, K and α, three different values for μ (to account for the spatial variation of background seismicity across the PNW) and a fine-spaced grid of p values. We then found the minimum/maximum p values that came closest to each of the five values summarizing the expected aftershock duration; these were the smallest and largest p values that could lead to the given aftershock durations, under realistic values for the other parameters. Extrema of these values were taken as the limits of a

Uniform prior on p.

We constructed Uniform priors for q in a similar way. Using prior beliefs elicited about aftershock zones, we could define the radius of the aftershock zone, r^*, by the boundary of the corresponding circle after which the aftershock rate equals the background rate; the derived equation depends on d, q, K, α and μ (see equation in Appendix, Section B.4). Again, we computed r^* over extrema combinations of d, K and α and three values for μ and a fine grid of q values. We found the minimum/maximum q value that came closest to each of the five values summarizing the expected aftershock zone; these spanned the range of values that could produce the given aftershock zones. Extrema of these values were then taken as limits of a Uniform prior on q.

Finally, we constructed Gamma priors for μ by eliciting beliefs about the background proportion for a long-term period of seismicity. In the PNW, 100 earthquakes occur roughly every seven years, on average. We thus phrased this question as: "For the next 100 earthquakes (seven years), what proportion of them do you expect to be triggered by previous seismicity?" A five-number summary for this number was converted to a background proportion, from which we constructed a Gamma prior.

3. **Feedback phase.** We sent the (tectonic regime-specific) priors to the experts for their approval and adjusted as needed.

We followed the protocol above for two experts on PNW seismicity. One provided beliefs and five-number summaries specific to the PNW and the resulting priors are given in Figure 3.3. The other expert could not give PNW-specific values, citing the uncertainty and insufficient data on aftershock dynamics in the region. They expressed that their prior beliefs on the aftershock dynamics in the PNW should be based on published ETAS estimates for other subduction zones.

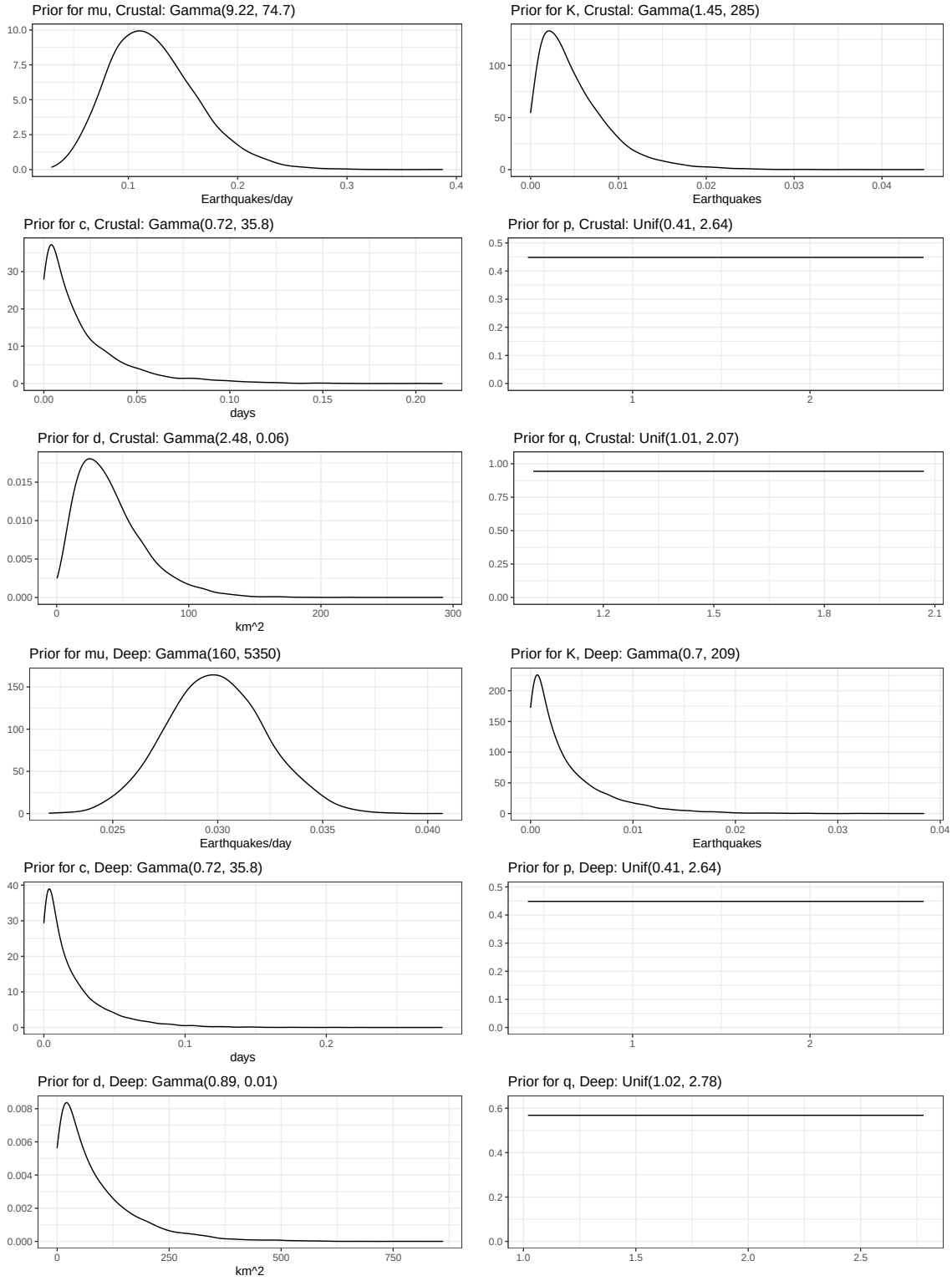

Figure 3.3: Prior distributions for ETAS parameters elicited specifically for the PNW region, separately for the crustal tectonic regime (top) and the deep tectonic regime (bottom).

Japan	Indonesia	Philippines
Papua New Guinea	New Zealand	Alaska
“western Pacific"	“eastern Pacific"	Mexico
Costa Rica	El Salvador	“Central America"
Peru	Ecuador	Chile
“South America"		

Table 3.3: Search terms for literature review of previous ETAS models for subduction zones.

3.3.2 Constructing general priors for subduction zones

To formulate priors pertaining to subduction zone aftershock dynamics, we surveyed the literature on ETAS models for known subduction zones. We restricted our search to English-language studies presenting full-catalog ETAS models for specific subduction zone regions, i.e., those described in Zhang et al. (2020)'s Figure 1 and Table 2. We ran the following searches in Google Scholar and Web of Science: [ETAS + "X"] or [Omori + "X"], where X was a search term given in Table 3.3.

We assumed that studies of any subregions inside these countries/large regions would at least mention the country/region name, thereby making them discoverable through our search. Most of the found literature discussed models only for specific sequences, not full-catalog (or even multiple-sequence) models; these were omitted. We acknowledge that in this list are large countries or seismic zones that may not have the same characteristics as the PNW; for example, though Japan contains a subduction zone, it also has much more seismicity at the subducting slab interface, which is largely missing from the PNW. However, since it was difficult to find subnational seismicity studies, we included studies at the national level.

We found eight studies with this search, three for Japan (Zhuang, 2011, 2015; Ogata,

2011), two for New Zealand (Harte, 2013; Console et al., 2010) and one each for Chile (Nicolis et al., 2015), Alaska (Wei et al., 2012), and the Mexican state of Guerrero (Yazdi et al., 2019). Nearly all studies used the spatiotemporal ETAS model with a magnitude-dependent spatial kernel for the triggered intensity, though two studies had slightly different formulations (Console et al., 2010; Wei et al., 2012). We attempted to harmonize these models to the spatiotemporal ETAS model with magnitude-dependent triggering spatial kernel and collected the reported parameter estimates from tables and figures; we assigned NAs to parameters that were not in the used model. Some studies reported ranges or standard errors around their parameter estimates (usually MLEs, or assumed as such when unspecified in the study); when standard errors were reported, lower/upper bounds for each parameter were taken by the reported estimate $\pm$ the standard error. We had to remove Ogata (2011) from this procedure as we could not extract values from their results plots, due to scale label overplotting. All information about the resulting seven studies, the catalogs used, and the reported ETAS parameter estimates are in Table B.2 in the Appendix.

We specified subduction-zone priors by taking Uniform distributions bounded by the minimum lower bound and maximum upper bound for each parameter estimate (after removing NAs), across all studies. These are shown in Figure 3.4. These priors were the same across crustal and deep earthquakes, reflecting the expert's belief that there is insufficient scientific knowledge to differentiate what is known from other subduction zones between the two tectonic regimes.

3.4 Evaluating ETAS model goodness of fit

Once we fit an ETAS model to a catalog, it is critical to quantify how well the estimated model parameters fit the catalog itself. We do this in two distinct steps.

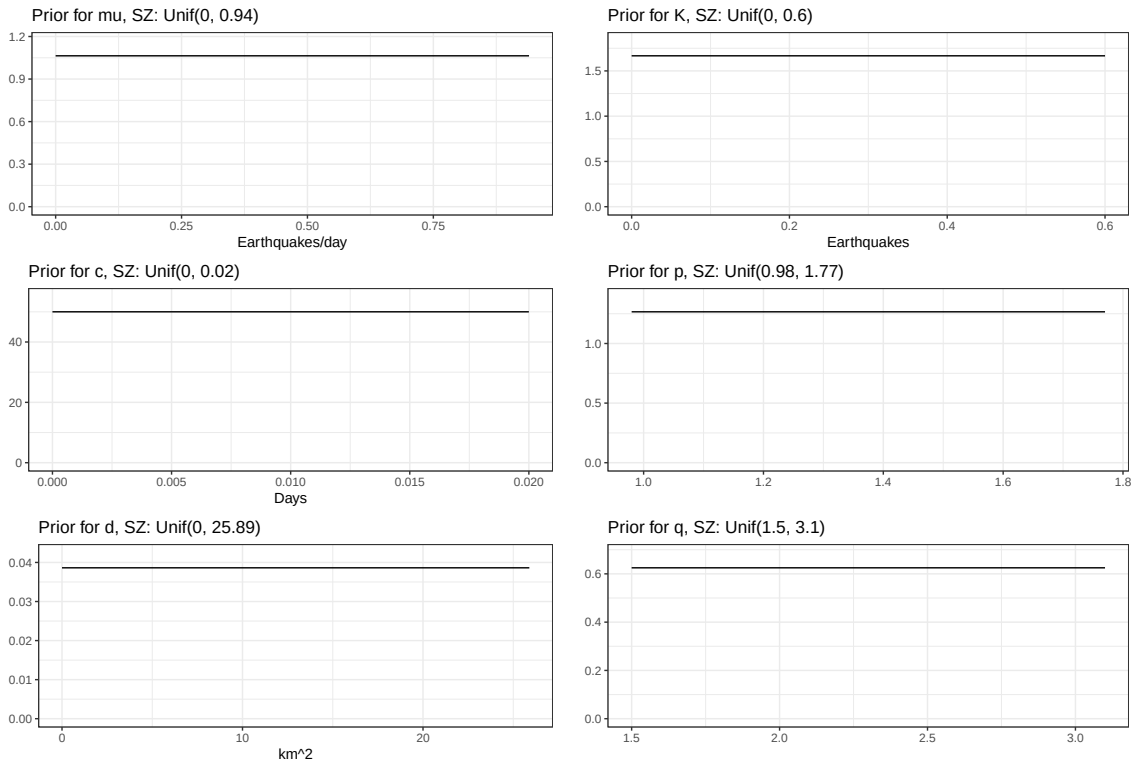

Figure 3.4: Prior distributions for ETAS parameters constructed from a literature search for all subduction zones.

3.4.1 Comparison to simpler model

We first check that the model structure imposed by ETAS is necessary to model the PNW. To do this, we build a spatial-only NHPP model of the catalog, which ignores the presence of aftershocks and earthquake magnitudes. This constitutes a null model against which to check the utility of the ETAS model. Note that our ETAS model assumes a spatially-constant μ_{ST} so in comparing the two models, we are comparing whether (1) explicitly modeling aftershocks and using earthquake magnitudes but ignoring the (2) spatial heterogeneity of background earthquakes with ETAS improves model fit to the PNW, compared to simply modelling point (2). In particular, we wish to compare the ability of the model to capture the spatial clustering evident in the catalog.

A common summary statistic to assess the spatial clustering of a spatial point process with conditional intensity λ on a spatial region of area $|S|$ is $K(r)$, Ripley's K-function (Ripley, 1976), defined as:

$$K(r) = \frac{1}{|S|}\frac{1}{\lambda}\sum_{i \neq j} I\{x_i - x_j\},$$

where i and j index points in the process and r is the distance in space for which clustering is assessed. The K-function may be estimated from a collection of points (Baddeley and Turner, 2005). We have developed a scoring rule (Heinrich-Mertsching et al., 2021) that compares the estimated K-functions between the observed catalog and a model fit to it, as described below. This scoring rule is *proper*, meaning that it will be optimized if the model exactly represents the data (catalog) generating mechanism.

Consider a realization from a fitted model y and two independent and identically distributed realizations from the data generating process X, X', for example, two disjoint periods from the PNW catalog. We define the proper *K-function score* for the model as:

$$\int_0^R \mathbb{E}[|\hat{K}(y,r) - \hat{K}(X,r)|]dr - \frac{1}{2}\int_0^R \mathbb{E}[|\hat{K}(X',r) - \hat{K}(X,r)|]dr,$$

where the value of R is chosen to be small relative to the scale of the spatial region S.

To evaluate the two models fit to the PNW catalog, we split it into two equal-length disjoint parts: a training catalog and a testing catalog. We train the ETAS and NHPP models on the training catalog and then simulate 100 realizations from each fitted model, for the length of the testing catalog. We compute the K-function score for each simulated realization against the testing catalog and compare the respective models by the distribution of their K-function scores. The lower the score, the better the model is at capturing the spatial clustering in the PNW catalog.

3.4.2 Transformed times

A point process model's goodness of fit can further be assessed with the sequence of *transformed times*, τ_i, defined as

$$\tau_i = \int_S \int_0^{t_i} \hat{\lambda}(t,x,y)dtdxdy,$$

where $\hat{\lambda}$ uses parameter estimates from an ETAS model fit to the catalog. These transformed times are related to the raw temporal residuals given in Zhuang (2006) and measure the level of misfit of the temporal part of a spatiotemporal point process model, based on the following reasoning. The integral of a point process' conditional intensity over some space-time interval represents the number of events expected for that interval. As such, we would expect the integral of the fitted conditional intensity up to t_i to be close to i, the number of catalog events until that time. To assess the quality of the parameter estimates, we can compare the sequence of τ_i to a Poisson process with a rate of 1 (a diagonal line of slope 1, or a perfect fit). If τ_i is below or above the unit-rate Poisson model, the parameter estimates over- or underpredict, respectively, the seismicity in the catalog. The derivation of τ_i is provided in

Appendix Section B.6. We specifically compare τ_n, the total number of events predicted at the final catalog time, with the actual number of catalog events, n.

3.5 Simulation of spatiotemporal ETAS catalogs

In order to evaluate the utility of the Bayesian ETAS procedure and our computational implementation, we estimate parameters for catalogs simulated from the spatiotemporal ETAS model. We used several fixed parameter sets and spatiotemporal square region $[0, T] \times [0, S]^2$. To account for known boundary issues (Harte, 2013), we simulate catalogs for a larger auxiliary space-time window and estimate parameters for a smaller space-time window $[0, T_{target}] \times [0, S_{target}]^2$, where $T = T_{target} + T_{margin}$ and $S = S_{target} + S_{margin}$. Other approaches exist for periodic boundary corrections (van der Elst, 2017), but we opt for this for simplicity.

Our simulation, based on Zhuang and Touati (2015), follows the following steps:

1. Draw a set of n_{bkgd} background earthquakes, where $n_{bkgd} \sim Pois(\mu \cdot T)$.

   (a) For each background event i, simulate a random time $t_i \sim Unif(0, T)$, a random location $x_i \sim Unif(0, S), y_i \sim Unif(0, S)$, and random magnitude $M_i \sim Exp(\beta)$, the Gutenberg-Richter law (Exponential distribution) with slope (rate parameter) β.

2. Set $T_{as} = 2 \cdot T$ as the maximum aftershock duration. For each background event (t_i, M_i), simulate a random number of triggered events $n_{trig,i} \sim Pois(\lambda_{trig,i})$, where

$$\lambda_{trig,i}(M_i) = K \exp(\alpha \cdot (M_i - M_0)) \cdot ((T_{as} + c)^{1-p} - c^{1-p}) \cdot \frac{1}{1-p}.$$

   This is the integral of the unnormalized Omori temporal density from $t = 0$ to $t = T_{as}$, scaled by the time-independent productivity law. This represents the expected number of triggered events within the finite time window $[0, T_{as}]$. Thus, the number of triggered events is a function solely of the mainshock's magnitude.

3. Each of these triggered events ($j = 1, \ldots, n_{trig,i}$) needs a random time, location, magnitude and number of triggered events.

   (a) Draw a random time using the inverse transform sampling technique. The inverse transform for the Omori temporal decay is derived in Section B.7. Simulate $r \sim Unif(0, 1)$ and draw a random time difference Δt_j from [0, T_{as}]:

   $$\Delta t_j = (r \cdot ((T_{as} + c)^{1-p} - c^{1-p}) + c^{1-p})^{1/1-p} - c.$$

   Set the triggered event's time $t_j = t_i + \Delta t_j$. Remove any times that fall outside T.

   (b) Draw a random location using inverse transform sampling from the Utsu-Seki spatial law. (The inverse transform is derived in Section B.7). This has two steps:

      i. Simulate a random distance by drawing $r \sim Unif(0, 1)$ and drawing a random distance Δs_j from [0, S]:

      $$\Delta s_j = \sqrt{[(r-1)(-d^{1-q})]^{1/1-q} - d}.$$

      ii. Draw a random angle in the circle, uniformly from $(0, 2\pi)$. The aftershock is located at the polar coordinate determined by this random distance and angle.

   (c) Draw a random magnitude $\sim Exp(\beta)$.

   (d) Draw a random number of triggered events again using $n_{trig,j} \sim Pois(\lambda_{trig,j})$, where

   $$\lambda_{trig,j}(t_j, M_j) = K \exp(\alpha \cdot (M_j - M_0)) \cdot ((T_{as} + c)^{1-p} - c^{1-p}) \cdot \frac{1}{1-p}.$$

4. Repeat the above step for all generations of triggered events until there are no more. The number of triggered events is guaranteed to converge to 0 for subcritical catalogs (van der Elst, 2017).

	Description	μ	K	α	c	p	β	T	n_{aux}	n_{tar}	$n_{M>4}$	M_{max}
Catalog 2	Low α, high K	0.1	0.02	1.7	0.05	1.08	ln(10)	20000	8121	1590	45	5.72
Catalog 3	High α	0.1	0.006	ln(10)	0.05	1.08	ln(10)	20000	5866	1384	46	5.84
Units		$\frac{\text{events}}{(\text{days}\cdot\text{km}^2)}$	events	$\frac{1}{\text{M}}$	days	-	-	days	events	events	events	M

Table 3.4: Synthetic test catalogs (M_{min}=2.5 for all). Catalogs were simulated for an auxiliary spatial window of (0,1000; $S = 1000km^2$) and values for the last four columns are for the target window of (250,750) only ($S_{target} = 500km^2$). To account for the temporal boundary issues, we used $T = 20000$ days and $T_{target} = 15000$ days.

	d	q
A	1	2
B	1	1.5
C	0.1	2
D	0.1	1.5
Units	km^2	-

Table 3.5: Simulation values for spatial parameters. Catalogs 2 and 3 had all four variations simulated.

The above procedure was performed for the "A" version of the catalog (d=1, q=2, see Table 3.5). We then used this catalog to reassign earthquakes to different locations based on the other d and q combinations. Thus, the total number of earthquakes is kept fixed for these

catalog versions and aftershocks are simply relocated based on different spatial parameters. Taking as input the auxiliary catalog (in our case, S=1000) and a d/q combination, we use the following procedure:

1. Identify all mainshocks as earthquakes that were not triggered by a previous earthquake. Reassign all mainshock locations uniformly, so $x_i \sim Unif(0, S), y_i \sim Unif(0, S)$ for all mainshocks i.

2. For each mainshock, find its triggered events. Assign to each one a random distance Δs from the mainshock and random angle as in Step 3b above. This provides the (x, y) of the relocated aftershock.

3.6 Pacific Northwest catalog

We analyze a newly collected catalog of contemporary seismicity for the international PNW region (see Chapter 2). The raw catalog contained all earthquakes from 1970-2018 instrumentally observed by the authoritative seismic network for a rectangular box around the states of Washington and Oregon (*target* region) and a larger *auxiliary* region, which comprises of the target zone and a margin zone around it. The authoritative networks are the Pacific Northwest Seismic Network (PNSN) for all earthquakes in Washington and Oregon, the Geological Survey of Canada (GSC) for all earthquakes in British Columbia, and other regional networks for the US states bordering Washington and Oregon, as included in the Advanced National Seismic System's Comprehensive Catalog (ComCat).

We initially collected all earthquakes above magnitude 2.0, as this is believed to be the magnitude of completeness (Malone, 2020) for the PNSN. We subsequently performed a completeness study (see Chapter 2, Section 2.8.1 and the following subsection) to identify complete subcatalogs to model.

Duplicate events between catalogs were algorithmically identified and confirmed by seismologists from the seismic networks; we used authoritative solutions when possible. More

Catalog	Latitude	Longitude	Magnitude	Time Window
Target	42°× 49°	-125°× -116.5°	M2.0+	1970-01-01 to 2019-01-01
Auxiliary	41°× 50°	-126°× -115.5°	M2.0+	1970-01-01 to 2019-01-01

Table 3.6: Target catalog parameters and number of events for raw catalogs. The latitude and longitude correspond to a rectangular box around the US states of Washington and Oregon, as well as a portion of British Columbia. The time window begins in 1970, when the instrumental record of the PNSN began. For GSC, the catalog begins in 1985, as location and magnitude calculations are inconsistent before and after 1985 (Brillon, 2020).

details on this procedure are in Chapter 2, Section 2.4.1. We removed blasts and attempted to identify earthquake swarms, which are common in the PNW (Bostock et al., 2019) and do not comply with the time-stationarity assumed by the ETAS model of the background rate. We first removed any seismicity within a 10 km radius circular area of Mt. St. Helens, which produced thousands of volcanic (non-tectonic) earthquakes. We then use a window-based deterministic procedure (similar to those used by Jacobs et al. (2013)) to identify earthquake clusters in horizontal space, time and depth (within the target zone only). These were classified by PNSN seismologists on a six-point scale between "definitely a swarm" and "definitely a sequence" (full details of this procedure are in Chapter 2, Section 2.4.1), which is mapped in Figure 3.5.

The presence of the subducting slab leads to different tectonic regimes in this region, as described in Section 3.1.4. We had collaborators use the Slab 2.0 model (Hayes et al., 2018) to calculate the distance between each event's hypocenter and the estimated location of the slab interface. As in Gomberg and Bodin (2021), we considered an event in the crustal regime if it was at least 10 km above the estimated slab interface, due to the uncertainty in the location of the interface (n=3935 in the target zone, see Figure 3.6, left). The Slab

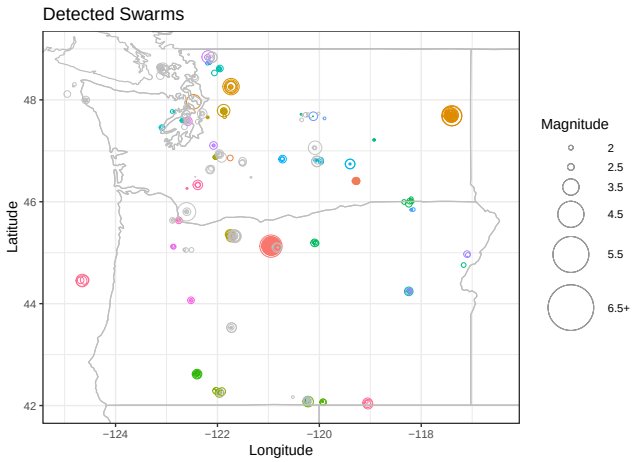

Figure 3.5: Earthquakes classified as definitely being in a swarm (in rainbow colors) and maybe being in a swarm (in grayscale colors).

2.0 model only goes out to longitude -121°; we assumed all events east of this were crustal events. All events that were below the estimated slab interface were considered intraplate events ($n = 546$ in the target zone, see Figure 3.6, right); further restricting to events that were at least 10 km below the estimated interface would reduce this to less than 150 events, so we did not do this.

We computed the spatial integral G for all earthquakes in the PNW target zone, using the d/q grid given in Table B.1 in Appendix Section B.1.

3.6.1 Complete PNW subcatalogs

As explained in Section 1.2, it is important for an earthquake catalog to be complete in order to model it with a point process model. Our previous study of completeness and

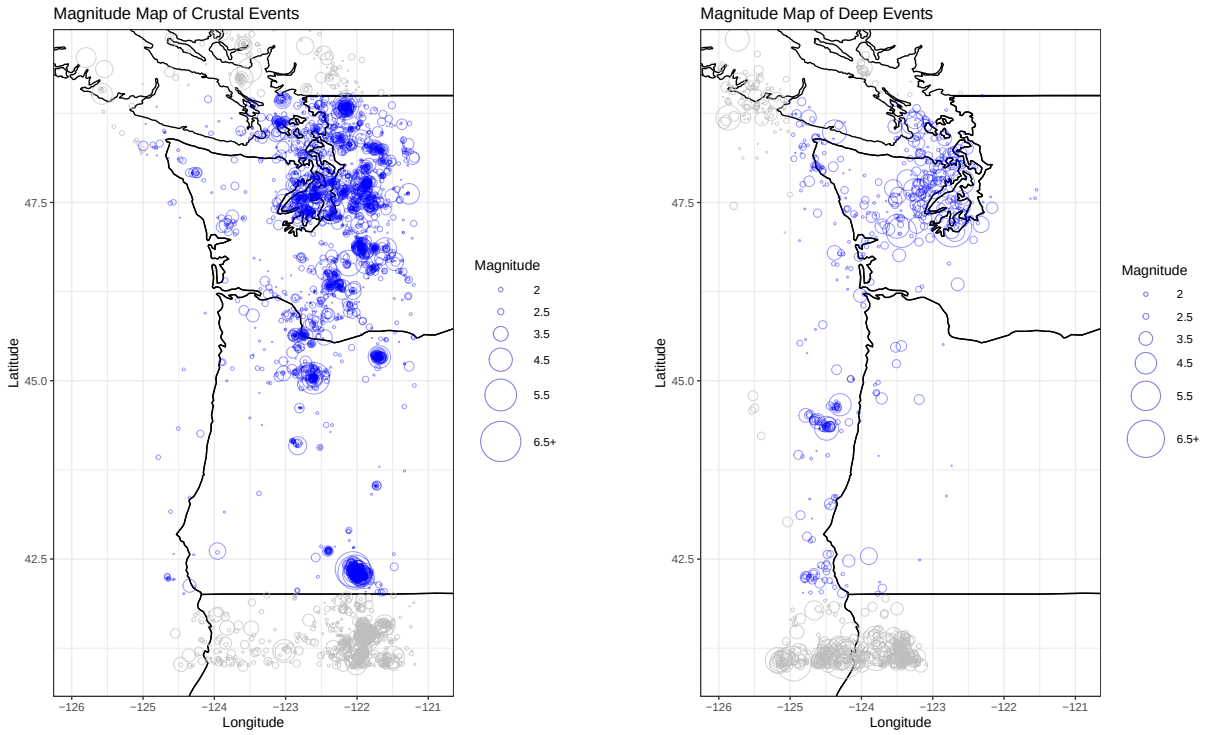

Figure 3.6: Earthquakes classified as below the slab (left) and earthquakes classified as above the slab (right), using the Slab 2.0 model (Hayes et al., 2018).

station coverage over time and space (see Chapter 2, Section 2.8.1) led us to split the PNW region into northern and southern areas at the latitude of 45°N (see Figure 3.7). This latitude break is based simply on where observed patterns of earthquake counts over time visually differed (see Figure 2.25 in Chapter 2).

In the northern area, the PNSN had broad station coverage from early in its history. The GSC, however, did not start consistently instrumentally recording earthquakes until 1985 (Brillon, 2020). Thus, we start the complete PNW North catalog in 1985. Our completeness study in Chapter 2, Section 2.8.1 found that M2.0 is the magnitude of completeness for the PNW North. The b-value for this catalog was estimated by MLE (Aki, 1965) to be 0.935 with a standard error of 0.013, using the technique of Shi and Bolt (1982)

In the southern area, the PNSN had limited stations starting only in 1980 and these did not cover the entire area. In 1993, after the Klamath Falls sequence in southern Oregon, a few stations were added to the surrounding zone but seismicity detection did not appear regular until around 2004, when another large sequence occurred in the Goose Lake, Oregon area. The very low seismicity level in the area makes it difficult to assess the time and magnitude of completeness. We assume the PNW South to be complete from 2004 and consider results under three different M_c values with the following b-value MLEs (and standard errors):

1. $M_c = 2.0$, potentially incomplete ($\hat{b}_{MLE} = 0.882$ (0.036))
2. $M_c = 2.3$, arguably complete ($\hat{b}_{MLE} = 0.981$ (0.055))
3. $M_c = 2.5$, highly likely complete ($\hat{b}_{MLE} = 1.07$ (0.079))

3.6.2 Jittering catalogs based on measurement error

Earthquakes are measured by a network of seismic stations that are not located precisely at each earthquake epicenter; thus, all measurements of earthquakes come with error. In Section 1.1, we explain the main sources of error in the measurement of earthquake locations

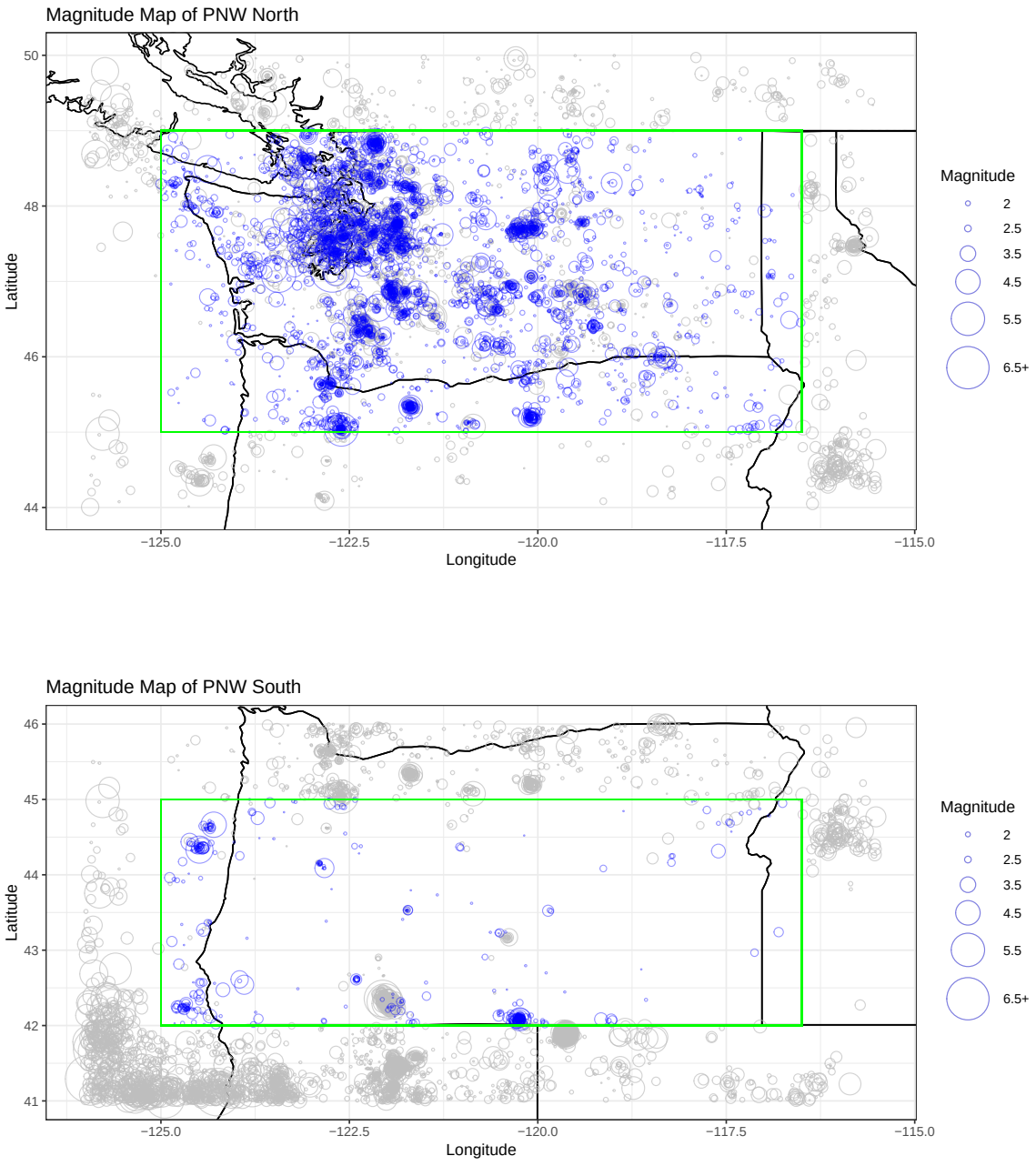

Figure 3.7: Instrumental catalog for the two PNW regions. Target events are in blue (with the target zone bounded by a green box) and margin events in gray. The southwest corner of the southern region is the Mendocino Triple Junction seismic zone, which is very seismically active.

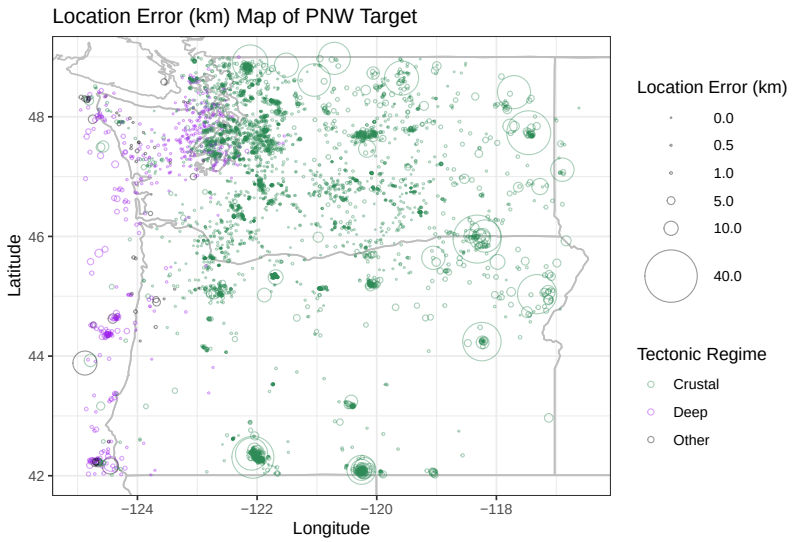

Figure 3.8: Location errors for the complete PNW, with the two main tectonic regimes in different colors.

and magnitudes. Given the importance of earthquake locations and magnitudes in both linking earthquakes that were triggered and in fitting their productivity and spatial decay parameters, measurement errors need to be considered when modelling catalogs.

Location errors are mapped in Figure 3.8 and magnitude errors in Figure 3.9. There are some general spatial patterns, with larger errors often occuring at the borders of the region, which is unsurprising considering that these border areas are distant from the majority of PNSN stations. There does not appear to be a big difference in error distributions between the crustal and deep tectonic regimes.

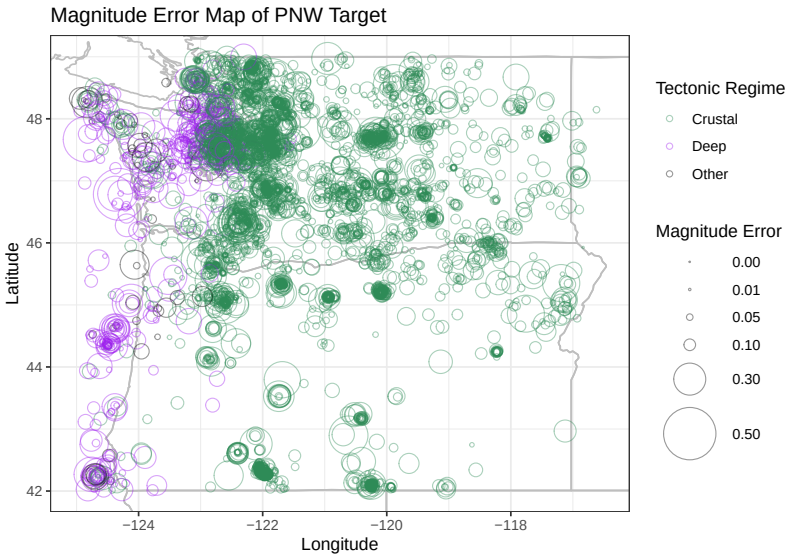

Figure 3.9: Magnitude errors for the complete PNW, with the two main tectonic regimes in different colors.

We incorporate measurement error into the model analysis using the procedure below:

1. Jitter new latitudes, longitudes and magnitudes for each catalog event i

   - $\tilde{x}_i, \tilde{y}_i \sim$ circle of radius horizontal error$_i$ and center (x_i, y_i)
   - $\tilde{m}_i \sim$ Normal(m_i, magnitude error$_i$),

   Here, the horizontal error (in km) is the length of largest projection of the error ellipsoid around earthquake i's location. We take this to approximate the radius of a circular error distribution around the reported location and draw a random value within this circle.

   The magnitude error is the network-provided standard error in estimating the earthquake's magnitude across reports from multiple stations. We follow the literature (Felzer, 2007) and colleague advice (Michael, 2021) and take this as the standard deviation of a Gaussian distribution of errors around the reported magnitude.

2. Subset the desired subcatalog based on $\{\tilde{x}, \tilde{y}, \tilde{m}\}$ (e.g., in the northern region, with or without swarms, etc). All of the information about swarms and tectonics is carried over from the original catalog; only the locations and magnitude change. Model the jittered subcatalog.

3. Repeat steps 1-2 100 times to get 100 jittered parameter estimates.

4. For each parameter, combine the posteriors for all jittered catalog into one grand posterior. This captures the uncertainty due to the (reported) catalog measurement error. Compare to the posterior under the reported catalog.

3.7 Results for simulated catalogs

We first show the efficacy of the Bayesian ETAS method on synthetic catalogs, simulated as described in Section 3.5. We do this by comparison to MLEs as well as Bayesian diagnostics.

3.7.1 MLE inference

Table B.8 has MLEs for spatiotemporal ETAS parameters using the branching likelihood function, for four end-member synthetic catalogs. MLEs use the non-linear maximization (NLM) optimizer, a Newton-style gradient descent method (see Table 3.1). In general, MLEs are close to true values, though there is a slight bias in overpredicting μ and underpredicting p. We also provide MLEs under the traditional likelihood function for synthetic catalogs in the Appendix, Section B.9.

	Value	μ	K	α	c	p	d	q
Cat 2A	True	0.1	0.02	1.7	0.05	1.08	1	2
Cat 2A	MLEs	0.113 (0.005)	0.019 (0.001)	1.715 (0.034)	0.043 (0.008)	1.063 (0.012)	1.008 (0.124)	2.009 (0.07)
Cat 2D	True	0.1	0.02	1.7	0.05	1.08	0.1	1.5
Cat 2D	MLEs	0.125 (0.005)	0.019 (0.001)	1.702 (0.019)	0.04 (0.005)	1.061 (0.009)	0.082 (0.007)	1.476 (0.018)
Cat 3A	True	0.1	0.006	ln(10)	0.05	1.08	1	2
Cat 3A	MLEs	0.108 (0.005)	0.007 (0.001)	2.252 (0.034)	0.052 (0.011)	1.083 (0.014)	0.983 (0.121)	1.986 (0.072)
Cat 3D	True	0.1	0.006	ln(10)	0.05	1.08	0.1	1.5
Cat 3D	MLEs	0.122 (0.005)	0.006 (0.001)	2.29 (0.036)	0.055 (0.012)	1.067 (0.015)	0.091 (0.011)	1.491 (0.028)

Table 3.7: MLEs for spatiotemporal ETAS parameters in several synthetic catalogs, using the branching log-likelihood $l(\theta|Z, B)$ and Hessian-based standard errors in parentheses (using the NLM optimizer). The optimizer is given initial values close to true parameters, though results are insensitive to different initial values. Other catalogs not shown did equally well.

3.7.2 Bayesian inference

We evaluated the goodness of fit of the posterior distributions on simulated catalogs using visual diagnostics. For each parameter, we inspected whether the true value was contained within the posterior density curve. We also inspected whether the true value was systematically below or above the posterior median to evaluate the extent of bias in the estimation (towards over- or under-prediction of the parameter). We further examined trace plots for each parameter, which show the posterior sample by iteration. We considered whether the sampling was well-mixed or whether the sampler took excursions to one part of the parameter space. We finally examined scatterplot matrices of the posterior samples across all parameter pairs to evaluate the level of parameter correlation and trend.

Figure 3.10 shows the posterior results for the ETAS parameters θ for two simulated catalogs (Catalog 3A (top) and 2D (bottom)) using the following modelling settings: parameters initialized close to the true values, branching structure initialized close to the true branching structure, and prior distributions taken from Ross (2021) (Table 3.2). Posteriors are usually well-centered around the true values (Figure 3.10) and trace plots show a well-mixed sampling of the posterior (Figure B.4 in the appendix). The posterior scatterplot matrix in Figure B.5 in the appendix shows large and relatively linear correlations between parameters in the same model component (K and α; c and p; d and q) but also across model components (e.g, K and p). Our posterior estimates allow to more closely investigate parameter correlation that has been reported in past literature (Lombardi, 2015; Wang et al., 2010).

We ran the following sensitivity checks to check the robustness of our Bayesian parameter estimation procedure and computational implementation.

Changes to model implementation:

- Initialized the ETAS parameters θ to values far from the true simulated values
- Initialized the branching structure B to assuming all events are background events

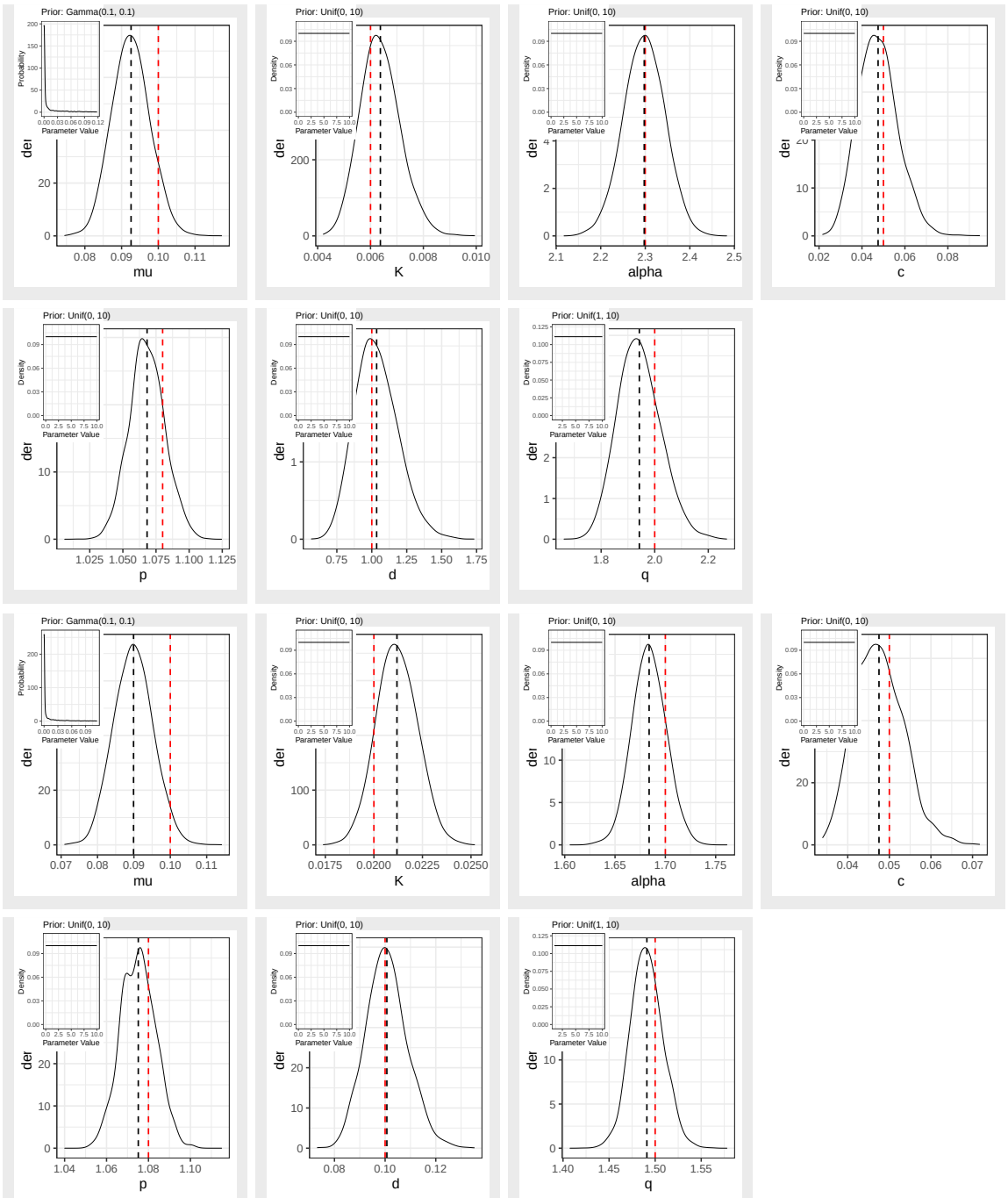

Figure 3.10: Sampled posteriors for ETAS parameters for synthetic Catalogs 3A (top) and 2D (bottom), with posterior medians in black and true parameter values in red. Prior distributions are given in the upper left corner. Results were similar for the other synthetic catalogs.

$(B_i = 0 \ \forall i)$

- Fixing the spatial integral for each earthquake $G_i = 1 \ \forall i$, as suggested by Schoenberg (2013) and commonly done by ETAS modellers, rather than calculating it using our numerical+analytical+interpolation procedure.

Changes to synthetic catalogs:

- Re-simulated the synthetic catalogs with a longer time window and repeated inference
- Repeated inference on 10 versions of each synthetic catalog

Posteriors were not sensitive to any of these changes, for all eight synthetic catalogs (Catalogs 2A-3D). These sensitivity checks suggested that not only is our computational implementation valid for modelling real catalogs, but that our posterior estimations are robust to different parameter initializations, which is a known problem for MLEs for the ETAS model (see Section 3.1.2).

3.8 Results for PNW catalogs

3.8.1 Modelling the PNW North

We sampled posterior distributions for the ETAS parameters for different subsets of the PNW catalog. We begin by showing results using non-informative prior distributions (Ross (2021); see Table 3.2). We first focus on the PNW North catalog, as described in Section 3.6.1. Figure 3.11 shows posterior densities for ETAS parameters for the *baseline catalog*: all earthquakes excluding confirmed swarms (that is, a score of 1 as assessed by PNSN seismologists, see Section 3.6) and solely for the crustal tectonic regime (n=3887). We use our novel solution to calculate and interpolate G_i for all earthquakes i, and initialize parameters with the Catalog 2 initial values (see Table 3.6). We also fix $\alpha = \hat{b}_{MLE} \cdot \ln(10)$, following other ETAS studies (van der Elst, 2017), due to the spatial extent of aftershock zones being magnitude dependent, which is not included in our model. Since α is fixed, we omit it from all results plots. Figure 3.12 shows good MCMC mixing as the conditional posteriors are sampled and Figure 3.13 shows scatterplot matrices for these posterior values.

We also calculated MLEs for this region using the traditional ETAS likelihood function (Ogata, 1998). We used several common numerical optimization procedures to maximize this likelihood, described in Table 3.1. We calculated asymptotic standard errors around these estimates using the likelihood Hessian as described in Section 3.2.2. As in the Bayesian approach, α was fixed to assume a self-similar catalog and the spatial integral was solved for. We used multiple convergence tolerances and found our results were insensitive to this; we present results using a tolerance of 10^{-6}.

We compare results between our Bayesian method and the estimation based on maximum likelihood with a standard numerical optimizer; we do this despite the known issues with MLEs for ETAS (see Section 3.1.2) because this is the most common estimation method currently used. Inference based on the Bayesian and likelihood approaches, under multiple

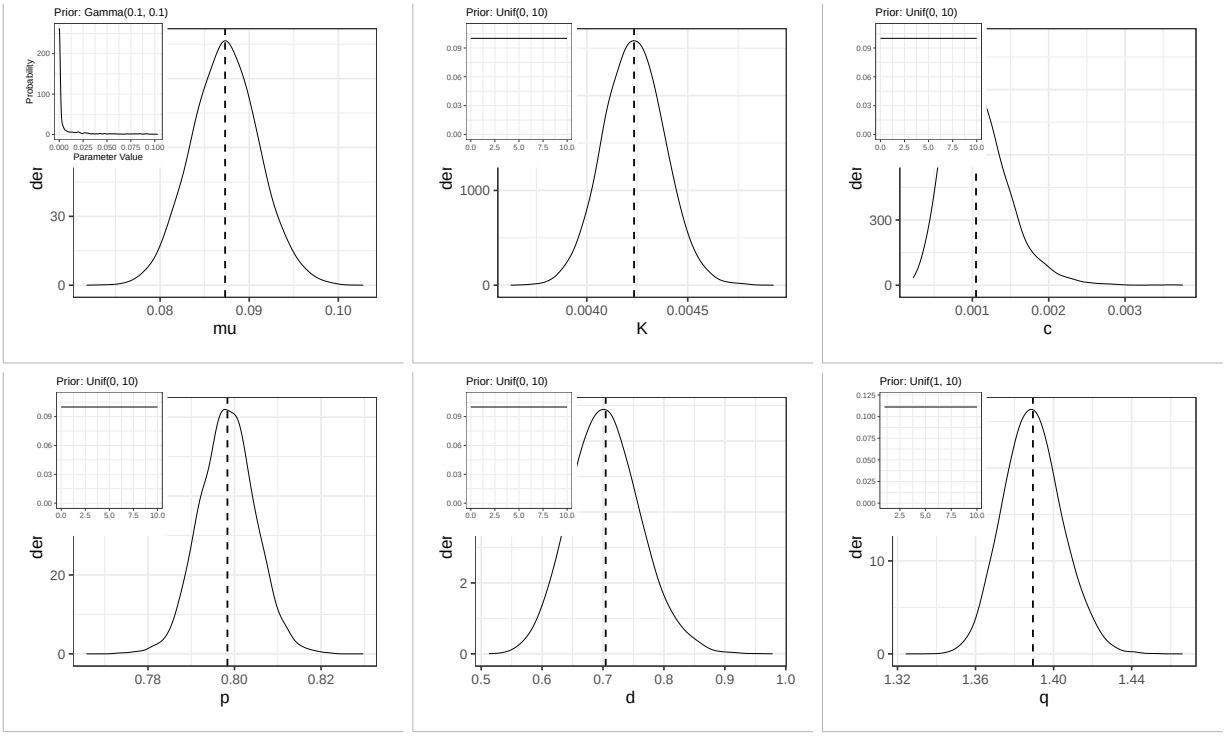

Figure 3.11: Sampled posteriors for ETAS parameters for PNW North crustal catalog, without maybe/confirmed swarms, with α fixed to $\hat{b}_{MLE} \ln(10)$. Posterior medians marked with black lines.

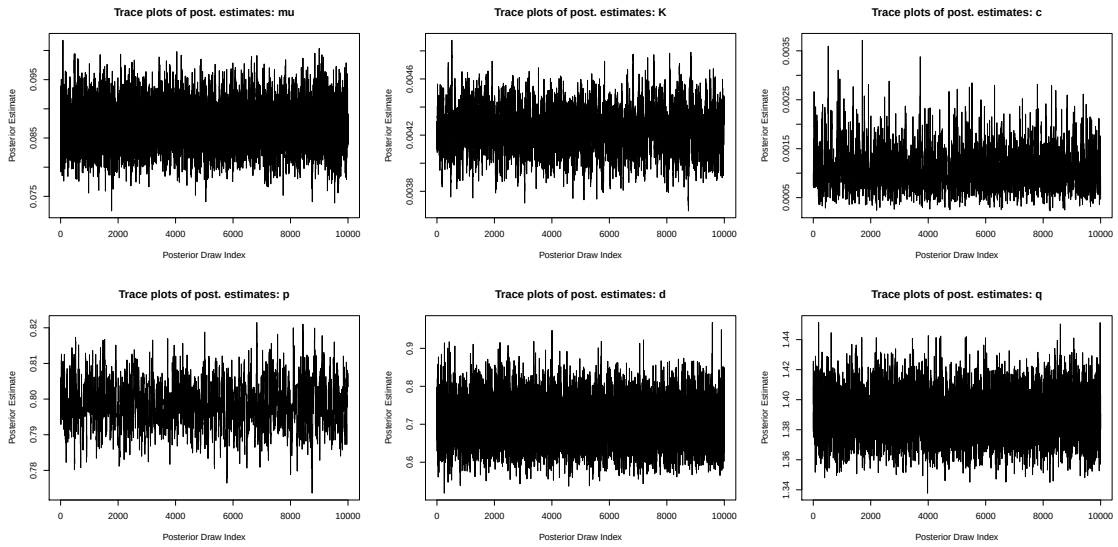

Figure 3.12: Trace plots for Bayesian inference for ETAS parameters for crustal catalog, PNW North.

optimizers, is shown in Figure 3.14, for a particular set of initial values (Cat 2 values, see Table 3.8). There is considerable variation in parameter estimates, particularly for K and c.

We ran the Bayesian and MLE methods under different initial value sets, given in Table 3.8, see Figures 3.15 to 3.17. There is considerable variation between the optimizers and between the MLEs they produce under different initial values, while the Bayesian solution is insensitive to initial values. When initial values are set to the posterior medians (Figure 3.17), MLEs are more similar to Bayesian results, but not across all optimizers for all parameters.

Model evaluation

Figure 3.14 shows that parameter estimates differ somewhat between the MLE and Bayesian approaches. This may be attributed to the approaches using different likelihood functions. We plot the (negative) traditional log-likelihood for varying K and varying d and q (Figure

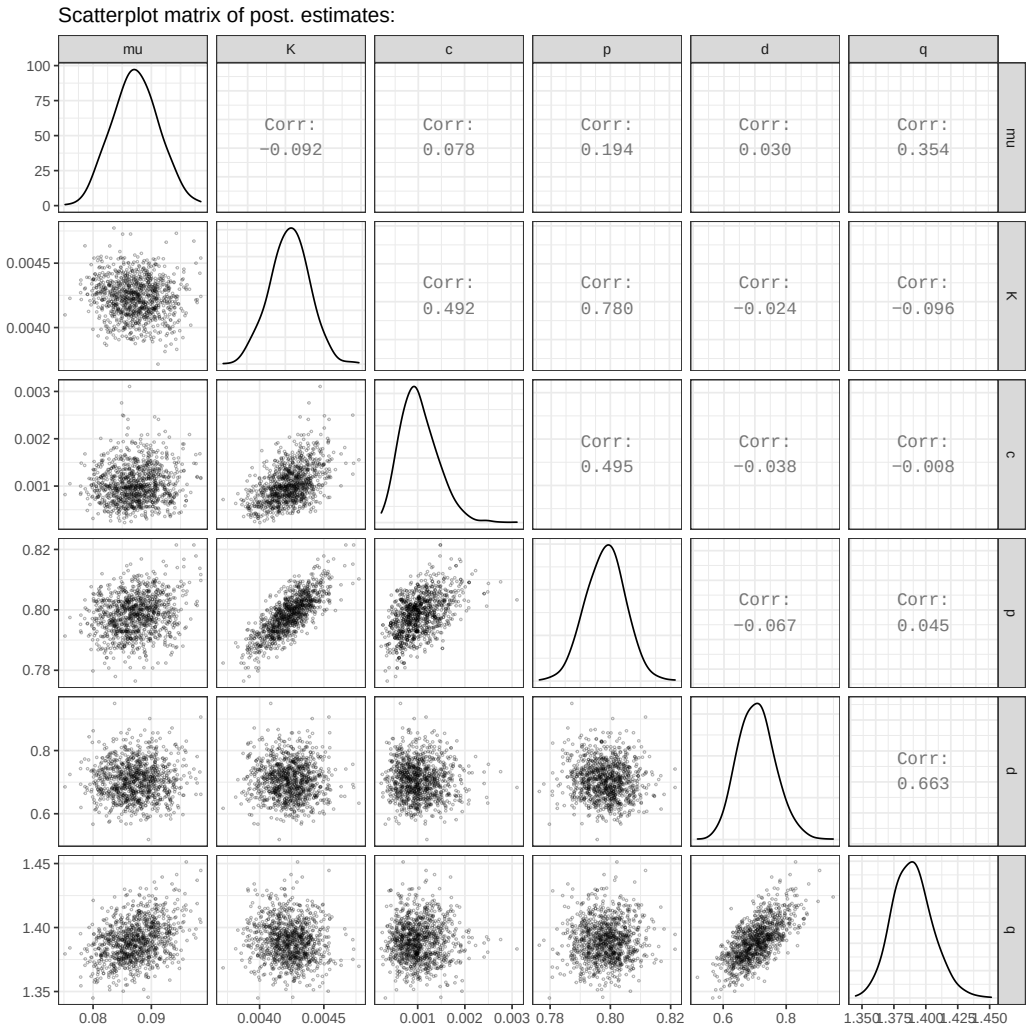

Figure 3.13: Scatterplot matrices for Bayesian inference for ETAS parameters for crustal catalog, PNW North.

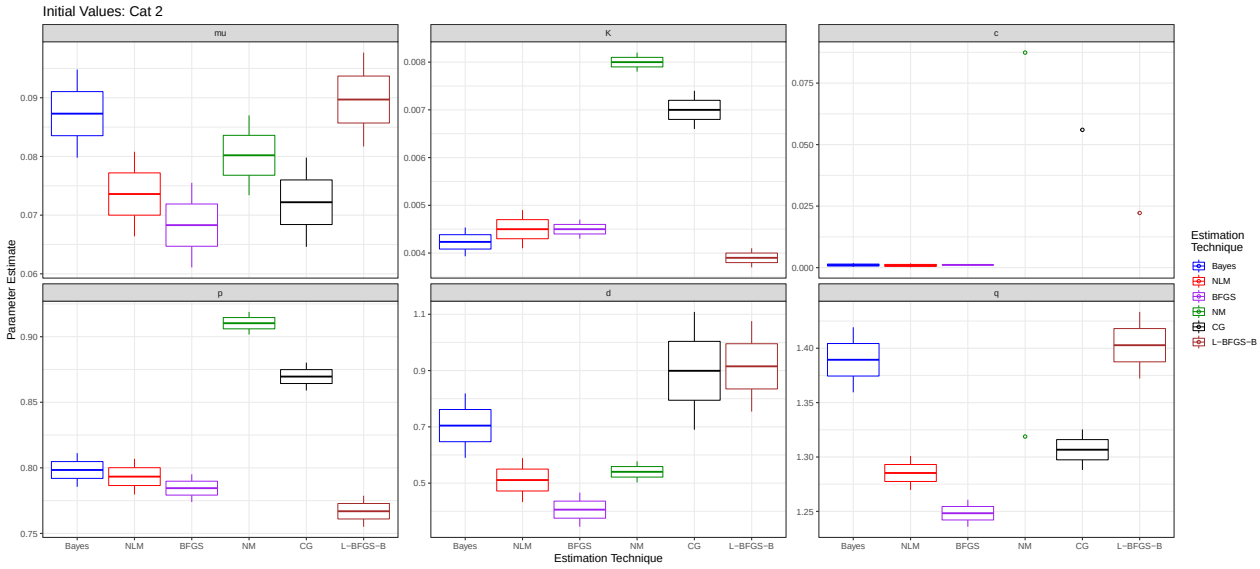

Figure 3.14: Parameter estimation for crustal catalog, PNW North (baseline). Boxplots show the 67% (inner box) and 95% credible intervals (whiskers) for Bayesian estimates. They show the $\pm$ 1 standard error (inner box) and 2 standard error (whiskers) confidence intervals for MLEs, using different optimizers, using Catalog 2 initial values (see Table 3.8). Empty circles represent an estimate where the Hessian-based standard error was not available, signifying a lack of convergence in estimating that parameter.

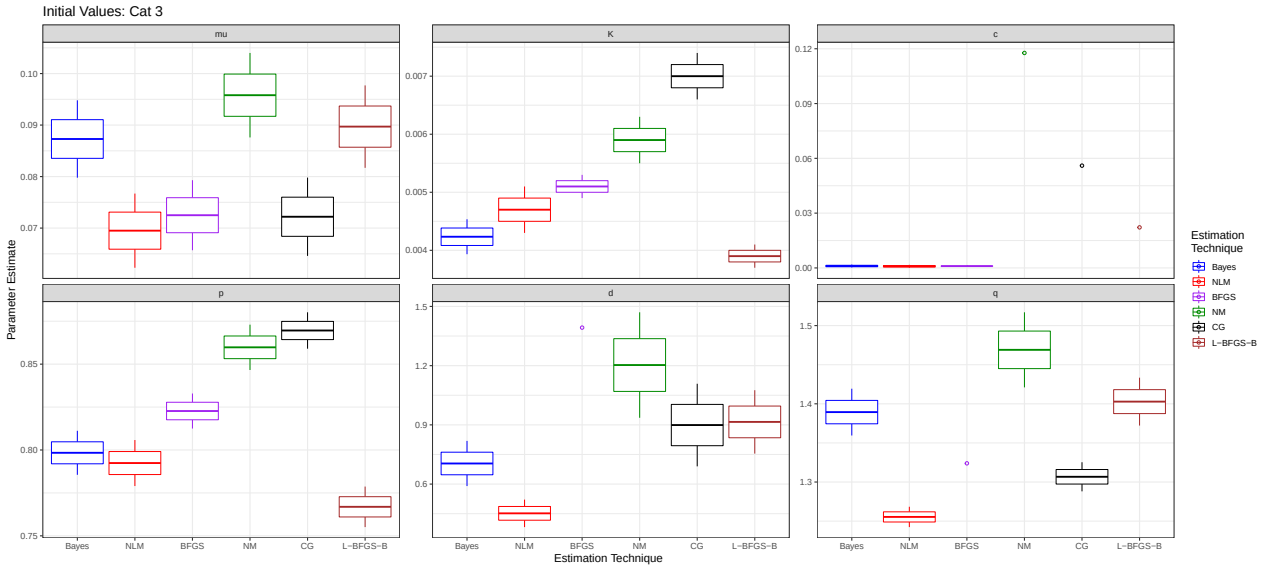

Figure 3.15: Parameter estimation for crustal catalog, PNW North (baseline), using Catalog 3 initial values (see Table 3.8).

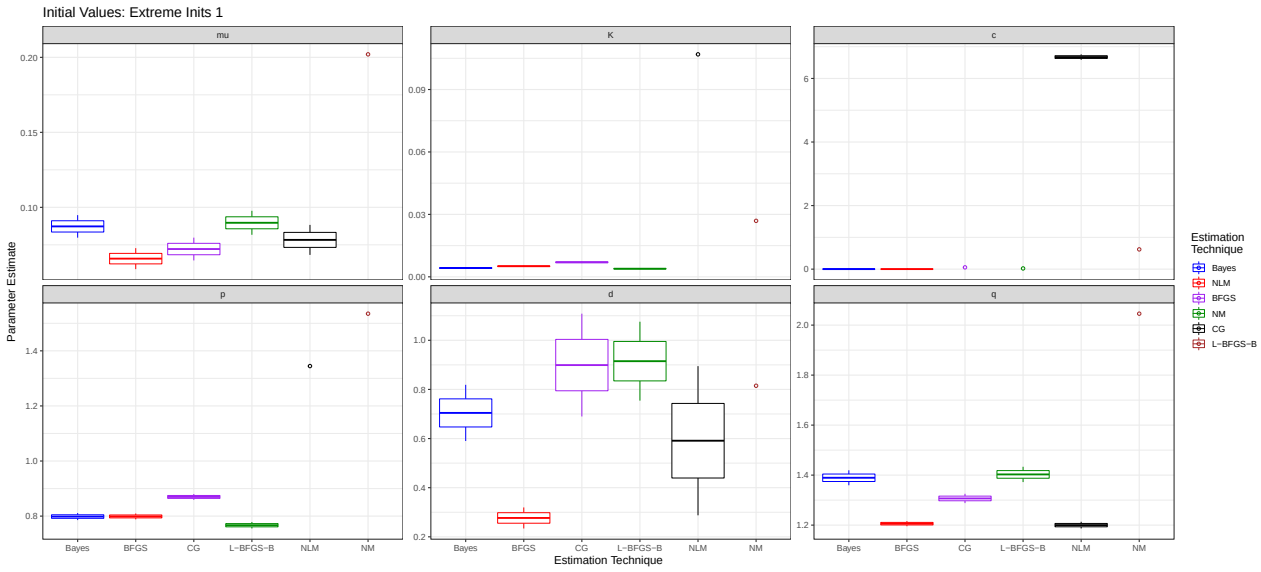

Figure 3.16: Parameter estimation for crustal catalog, PNW North (baseline), using extreme initial values (see Table 3.8).

	μ	K	α	c	p	d	q
Catalog 2	0.1	0.02	$\hat{b}_{MLE} \cdot \ln(10)$	0.05	1.08	0.1	1.5
Catalog 3	0.1	0.006	$\hat{b}_{MLE} \cdot \ln(10)$	0.05	1.08	1	2.0
Extreme inits 1	10	0.00002	$\hat{b}_{MLE} \cdot \ln(10)$	0.005	1.5	0.1	1.5
Posterior medians	0.086	0.0041	$\hat{b}_{MLE} \cdot \ln(10)$	0.0008	0.79	0.7	1.39

Table 3.8: Different initial value sets. Posterior medians are means of the Bayesian posterior distributions, with fixed α.

3.18), with all other parameters fixed to their MLEs, as calculated by the NLM optimizer. Figure 3.19 show the same plots for the (negative) branching log-likelihood, with other parameters fixed to posterior medians. These indicate that the MLE solutions and Bayesian solutions are both in high-likelihood (low-negative-likelihood) zones for their respective likelihood functions.

We evaluate the goodness of fit of each estimate set by considering its transformed times against the unit-rate Poisson model, plotted in Figure 3.20. While both approaches underpredict at the beginning of the period (transformed times above the diagonal line), the Bayesian estimates end up better fitting the catalog as time goes on, with a difference between the estimated and actual total earthquake counts of only 3.9 earthquakes (95% credible interval of (-91.3)-143.1). The MLE solutions have similar trajectories for their transformed times, slightly overpredicting the total number of earthquakes compared to the Bayesian solution (transformed times below the diagonal line). The difference between the estimated and actual total earthquake is 100.3 (95% confidence interval: 1.3-142.0), when averaging over the five optimizers and three initial values given in Figures 3.15 to 3.17. MLEs under different optimizers lead to similar transformed times, despite their difference in estimates (see, e.g., Figure 3.14). This suggests that the different parameter estimates may compensate for each

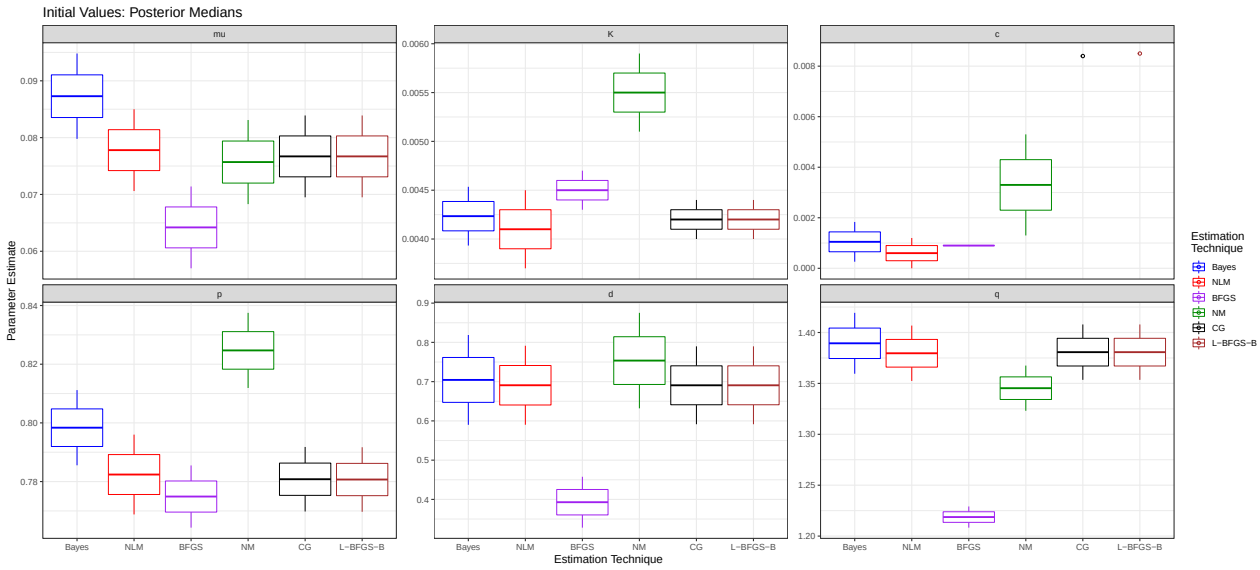

Figure 3.17: Parameter estimation for crustal catalog, PNW North (baseline), using posterior medians as initial values (see Table 3.8).

other (i.e., lower $\hat{\mu}$ combining with a higher $\hat{K}$), but come out to a similar overall fit to the catalog.

Comparing the ETAS and NHPP models We evaluated the need for the ETAS model by comparing its goodness of fit to the catalog against the simpler NHPP model. Specifically, we computed K-function scores to assess whether our ETAS model, which explicitly models aftershock dynamics and magnitude relationships but omits spatial heterogeneity of the background events (constant μ_{ST}) fits the spatial clustering of the catalog better than a model that solely represents the spatial heterogeneity of all events (ignoring earthquake magnitudes and the presence of aftershocks).

We used the testing/training model evaluation framework, splitting the crustal PNW North catalog at year 2002 to have equally sized testing/training catalogs. We simulated

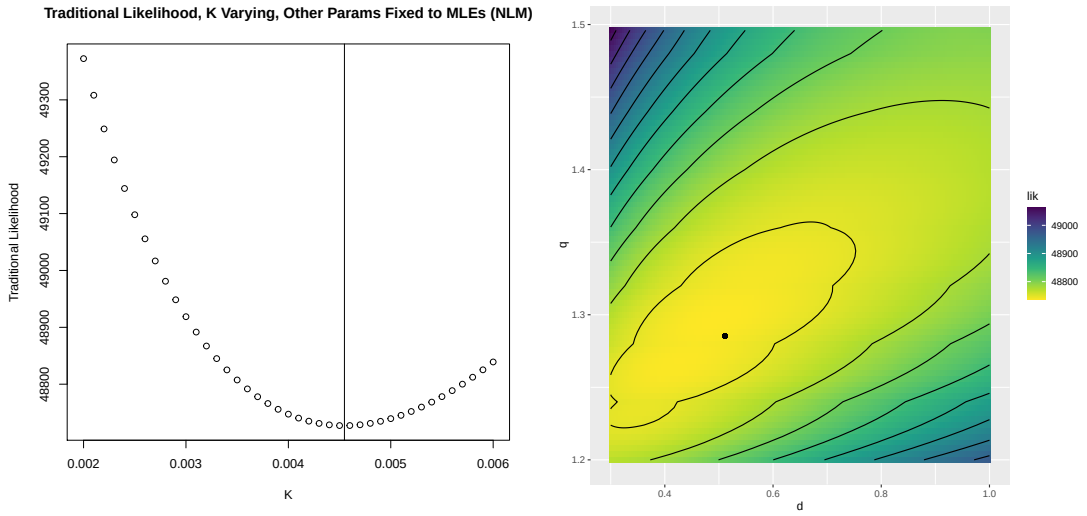

Figure 3.18: Negative likelihood profiles for the traditional likelihood of Ogata (1998) for varying K (top) and d/q (bottom), with all other parameter values fixed to the MLEs, as found by the NLM optimizer. Contour lines in the d/q plot are for every 30 units of likelihood. The MLEs for K and d/q are marked on the plots.

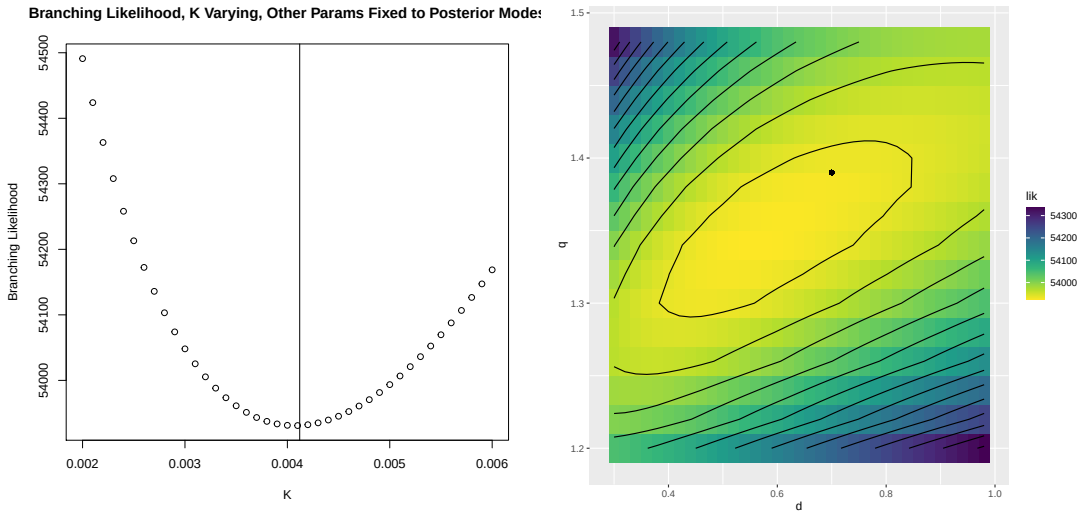

Figure 3.19: Negative likelihood profiles for the branching likelihood for varying K (top) and d/q (bottom), with all other parameter values fixed to Bayesian posterior medians. Contour lines in the d/q plot are for every 30 units of likelihood. The posterior medians for K and d/q are marked on the plots.

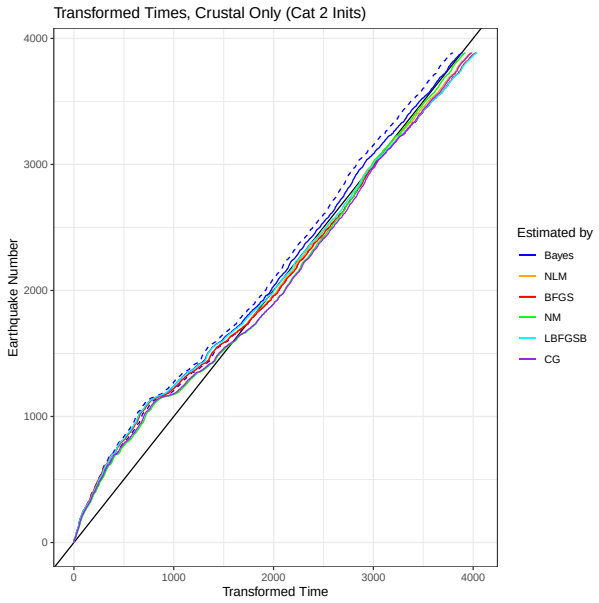

Figure 3.20: Transformed times for Bayesian (blue) and MLE fits to the baseline catalog (PNW North, crustal-only). Blue dotted lines show transformed times using parameter values from the 2.5% and 97.5% percentiles of each parameter's posterior distributions (the 95% credible interval for the transformed time).

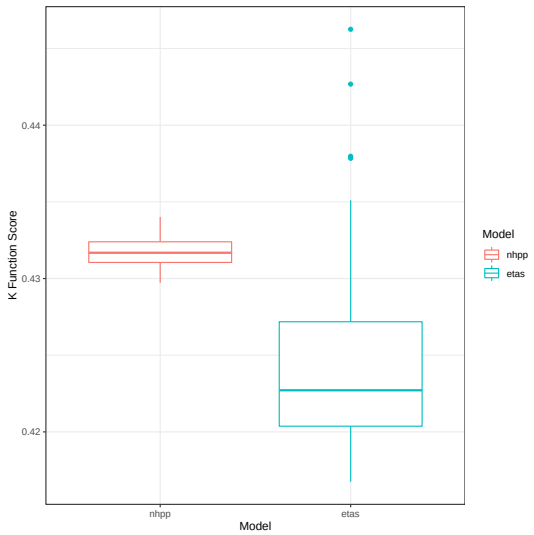

Figure 3.21: K-function scores between the ETAS model, estimated with the Bayesian method, and a NHPP model. Lower scores indicate a better fit of the spatial clustering in the testing catalog.

ETAS catalogs for the PNW North region using Bayesian posterior medians (as fit to the training catalog), using the simulation procedure in Section 3.5, but truncating the Gutenberg-Richter magnitude distribution at $M = 5.8$, which is 0.2 units above the maximum observed magnitude for this catalog (Seif et al., 2017). We also simulated catalogs from a NHPP model fit by kernel density estimation to the training catalog. Figure 3.21 shows boxplots of the K-function scores over simulations from the two models. The mean score is slightly lower for the ETAS model, though the spread of scores is larger. This indicates that simulations from the fitted ETAS model do an overall slightly better job at capturing the spatial clustering of the testing catalog than does NHPP, but with more variability.

3.8.2 Modelling the PNW South

We also fit ETAS models using Bayesian and MLE inference for the much sparser PNW South catalog. We considered three different magnitudes of completeness as described in Section 3.6.1. Bayesian estimates for the $M_c = 2$ catalog (crustal-only, n=289) are summarized in the posterior distributions in Figure 3.22, with trace plots and scatterplot matrices in Figures 3.23 and 3.24. Parameter estimates are notably different between the northern and southern regions, particularly for parameters K, p and q, though there is much greater uncertainty about all triggering parameter estimates in the PNW South.

Bayesian results for crustal-only catalogs with $M_c = 2.3$ (arguably complete, n= 164) and $M_c = 2.5$ (likely complete, n= 104) are given in Figures B.6 through B.11 in Appendix, Section B.10. Some parameter estimates (e.g., p and q) are similar between the $M_c = 2$ and $M_c = 2.3$ catalogs, just with wider posteriors indicating greater uncertainty around the estimate; other parameters, like μ have noticeably different estimates when raising the completeness magnitude. Posteriors for all triggering parameters are much wider for the $M_c = 2.5$ catalog, nearly spanning the range of the (uninformative) priors and the MCMC sampler is not as well-mixed (see Figure B.10). This indicates that there is insufficient data to meaningfully update the posterior from the prior.

In Figure 3.25, we compare the Bayesian estimates to MLEs (for multiple optimizers) under all three magnitude cutoffs. Though parameter uncertainty is larger for higher M_c for Bayesian estimates, this is not evident for all MLEs. While some optimizers (e.g., NLM and L-BFGS-M), have somewhat larger confidence intervals for increasing M_c, this difference is much smaller than for Bayesian inference; other optimizers (e.g., BFGS and Nelder-Mead) have inconsistent trends in standard errors as M_c rises.

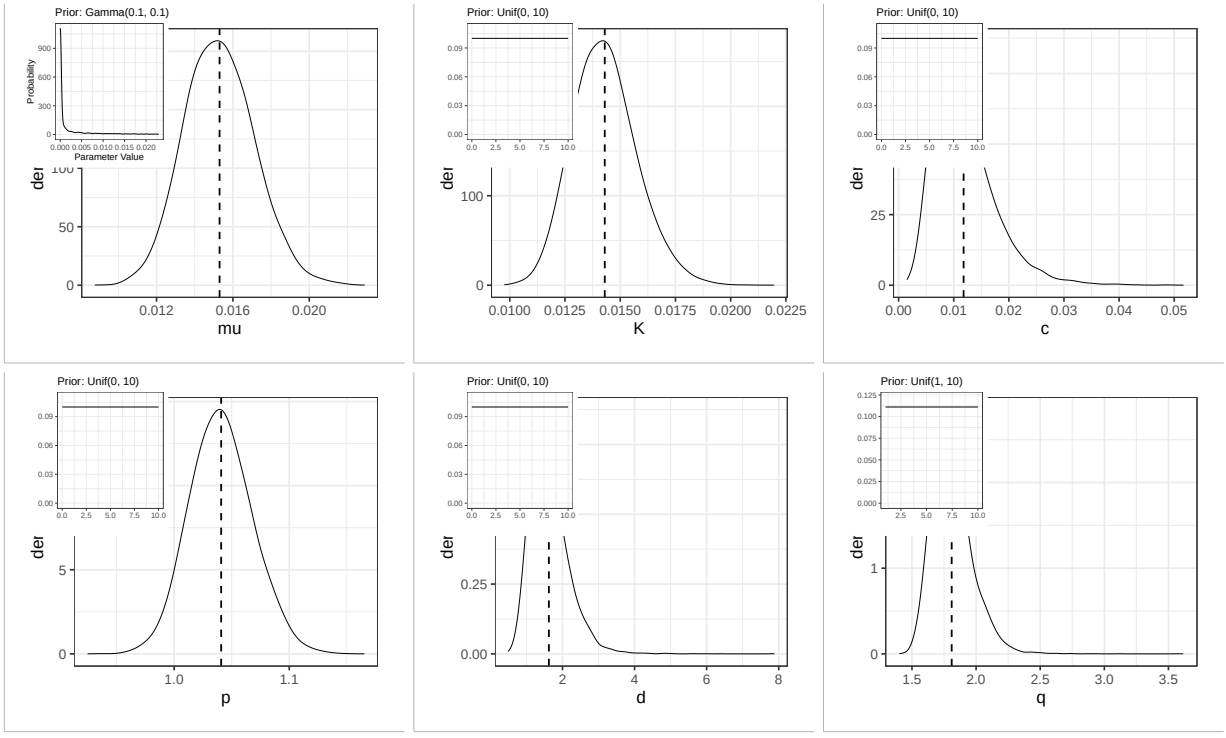

Figure 3.22: Sampled posteriors for ETAS parameters for PNW South crustal catalog (M2+), without maybe/confirmed swarms, with α fixed to $\hat{b}_{MLE} \ln(10)$. Posterior medians marked with black lines.

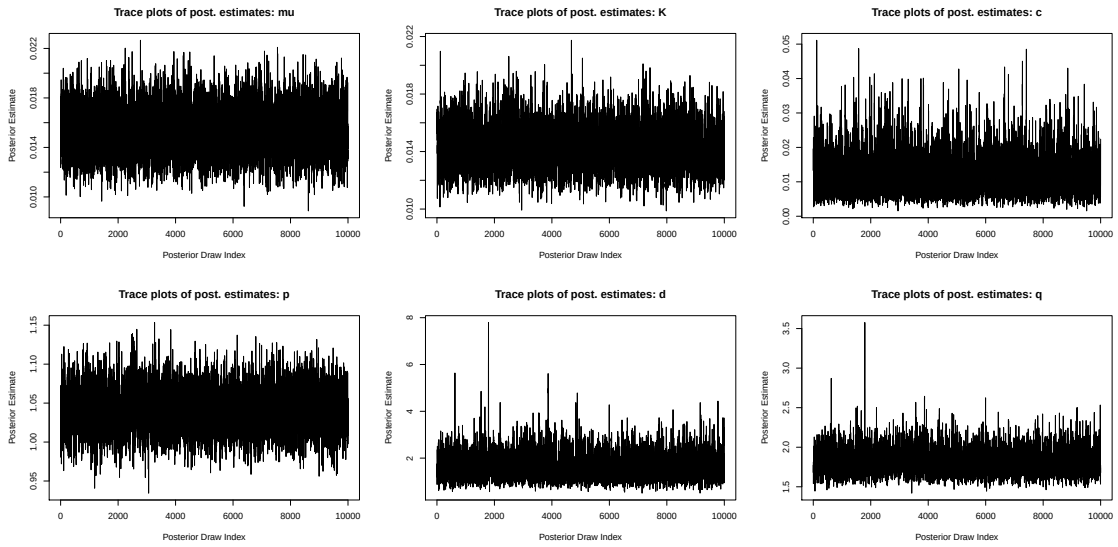

Figure 3.23: Trace plots for Bayesian inference for ETAS parameters for crustal catalog, PNW South (M2+).

Model evaluation

We evaluated the Bayesian and MLE fits to the PNW South catalog using transformed times (see Figure 3.26). The fit of the ETAS model is worse than for the PNW North, especially for $M_c > 2$, regardless of the estimation approach. For the $M_c = 2$ catalog, Bayesian inference slightly underpredicts for much of the catalog period, though ends at a total number of earthquakes similar to the observed number. For the $M_c = 2.3$ catalog, Bayesian inference overpredicts for the entirety of the period, with the problem even worse by the end of the period. MLE fits struggle even more than Bayesian inference, usually having worse overprediction for all three catalogs.

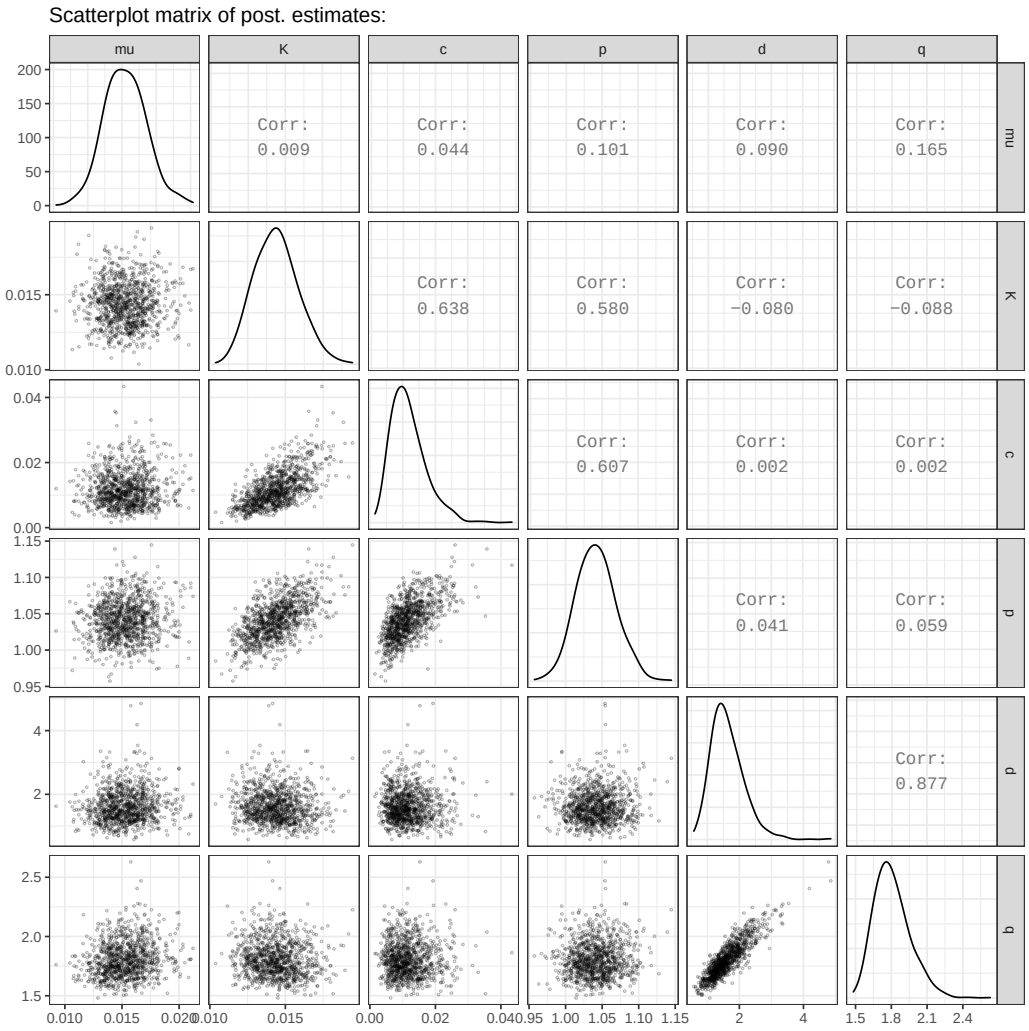

Figure 3.24: Scatterplot matrices for Bayesian inference for ETAS parameters for crustal catalog, PNW South (M2+).

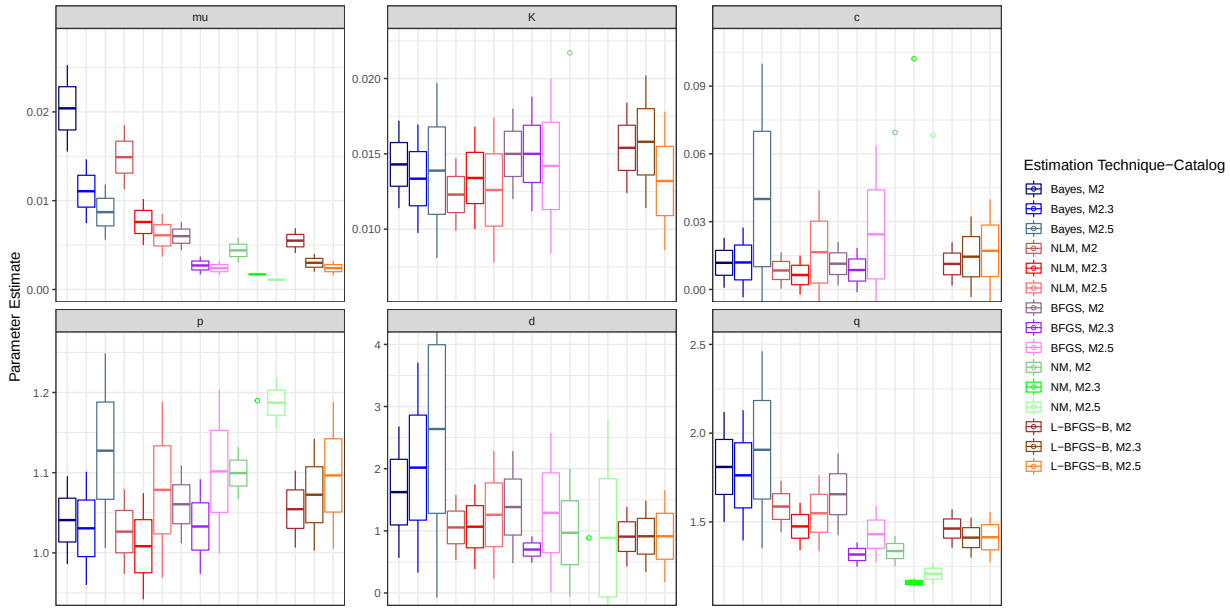

Figure 3.25: Parameter estimation for crustal catalog, PNW South. 95% credible intervals for Bayesian estimates and confidence intervals for MLEs, using different optimizers, using Catalog 2 initial values (see Table 3.8). We give results for each catalog grouped by estimation technique (i.e., the three blue lines are Bayesian estimates for $M_c = 2+$, $M_c = 2.3+$, and $M_c = 2.5+$ and the pattern is repeated across optimizers. Results for the M2.5+ catalog are often outside the bounds of the plot due to their much larger uncertainties.

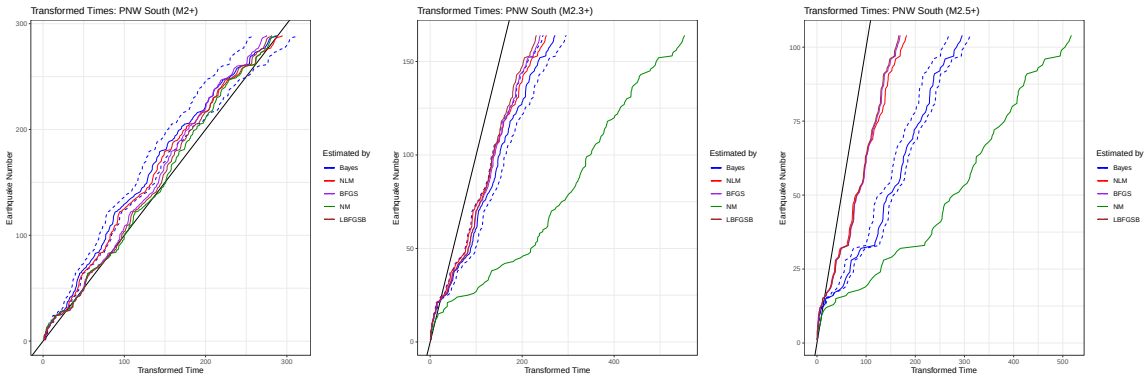

Figure 3.26: Transformed times for Bayesian (blue) and MLE fits to the baseline catalog (PNW South, M2+, M2.3+ and M2.5+). Blue dotted lines show transformed times using parameter values from the 2.5% and 97.5% percentiles of posterior distributions.

3.8.3 Bayesian inference with informed priors

We considered Bayesian models using both sets of informed priors, described in Section 3.3. The PNW-specific priors are much broader (see Figure 3.3) than the ones based on the subduction zone ETAS literature (see Figure 3.4); the PNW priors also cover the values of parameter posterior distributions found under uninformative priors, while the subduction zone priors do not.

Figure 3.27 shows posteriors under the three sets of priors. While the parameter inference is nearly identical between uninformative priors and those with PNW-specific priors, results differ when using subduction zone priors. Parameter estimates are higher under the subduction zone priors; the posteriors for parameters p and q are at the lowest extreme of their priors (see Figure 3.4) and with very short credible intervals, indicating that the posterior density is concentrated at the lowest possible values of these priors for parameters p and q. The fact that the posteriors lie at the the boundary of the priors suggests that priors drawn from previous models of subduction zones may not be compatible with the PNW catalog.

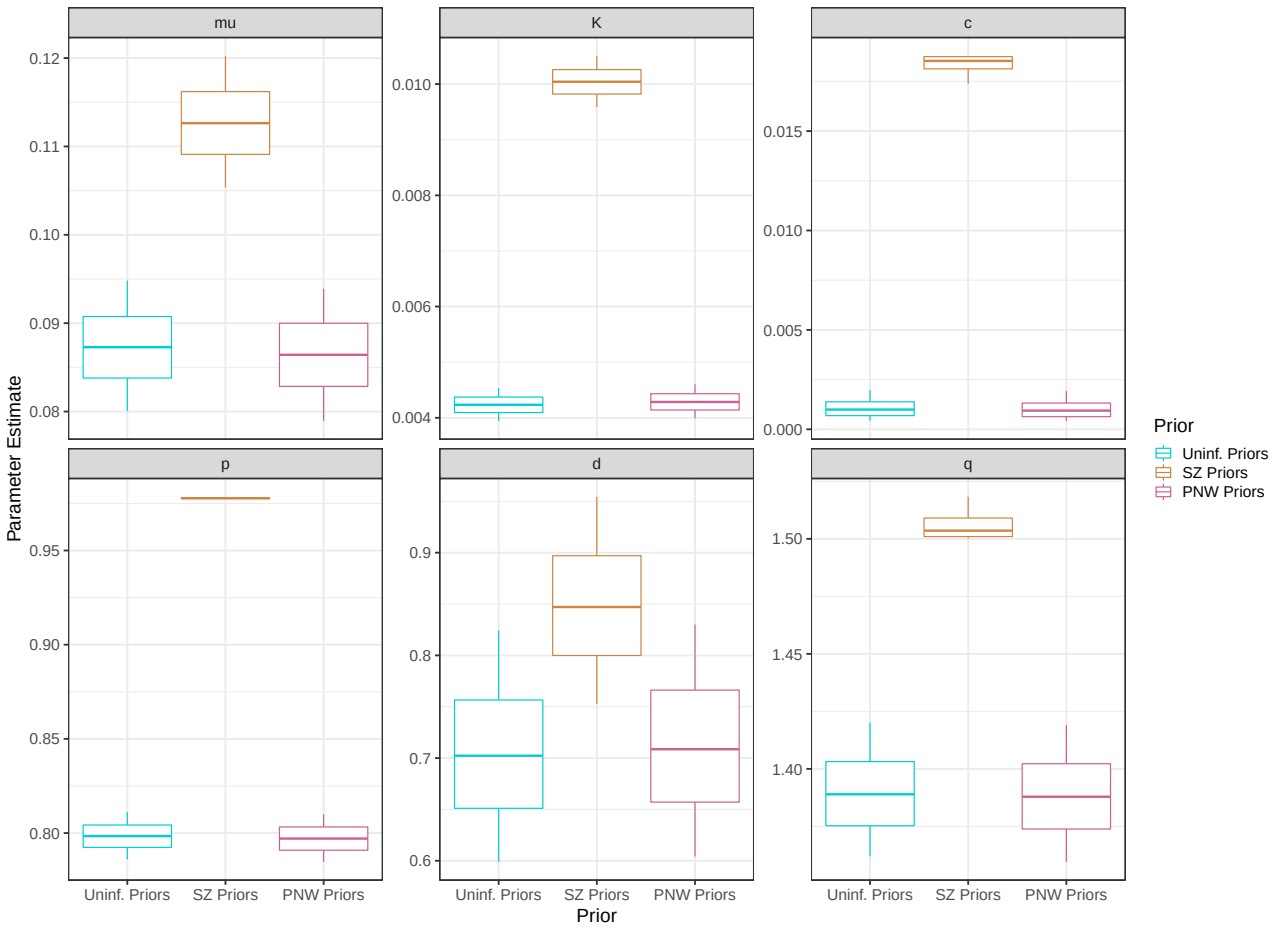

Figure 3.27: Bayesian credible intervals for ETAS parameters for the baseline catalog, PNW North, using different sets of priors. Prior distributions in Figures 3.3 and 3.4.

3.8.4 Quantifying uncertainty from catalog choices

We ran a series of analyses to assess how considering additional sources of uncertainty in the catalog would affect parameter estimates. We considered how results for the PNW North changed when adding events in:

- earthquakes in the deep tectonic regime
- potential or confirmed swarms (with a value of 1 or 2 on the scale described in Section 3.6, respectively)
- the margin zone (the gray events in Figure 3.7, top); that is, modelling the auxiliary zone (target + margin)

Tectonic Regimes Figure 3.28 shows Bayesian inference for the catalog that combines deep and crustal seismicity compared to those for the baseline catalog (crustal only). Most parameters are somewhat different between the two sets; $\hat{\mu}$ is higher for the combined catalog and it has a lower $\hat{K}$ and higher $\hat{p}$ and $\hat{q}$.

The model estimated on the combined catalog has worse fit to this data, as seen in the transformed times in Figure 3.29. There is a sharp increase around the transformed time of 1600, which corresponds to when the M6.8 Nisqually earthquake occurred (in the deep regime). The model expects a large number of aftershocks that did not actually occur, leading to worse model fit.

Swarms Figure 3.30 shows Bayesian inference for a catalog that includes swarms compared to the baseline catalog (no confirmed swarms), and those where confirmed or possible swarms are removed. There are no differences in parameter estimates when we exclude just confirmed or both confirmed and possible swarms. There is a slight increase in $\hat{K}$ and $\hat{p}$ when we include swarms in the estimation, but all other parameters have 95% credible intervals that overlap.

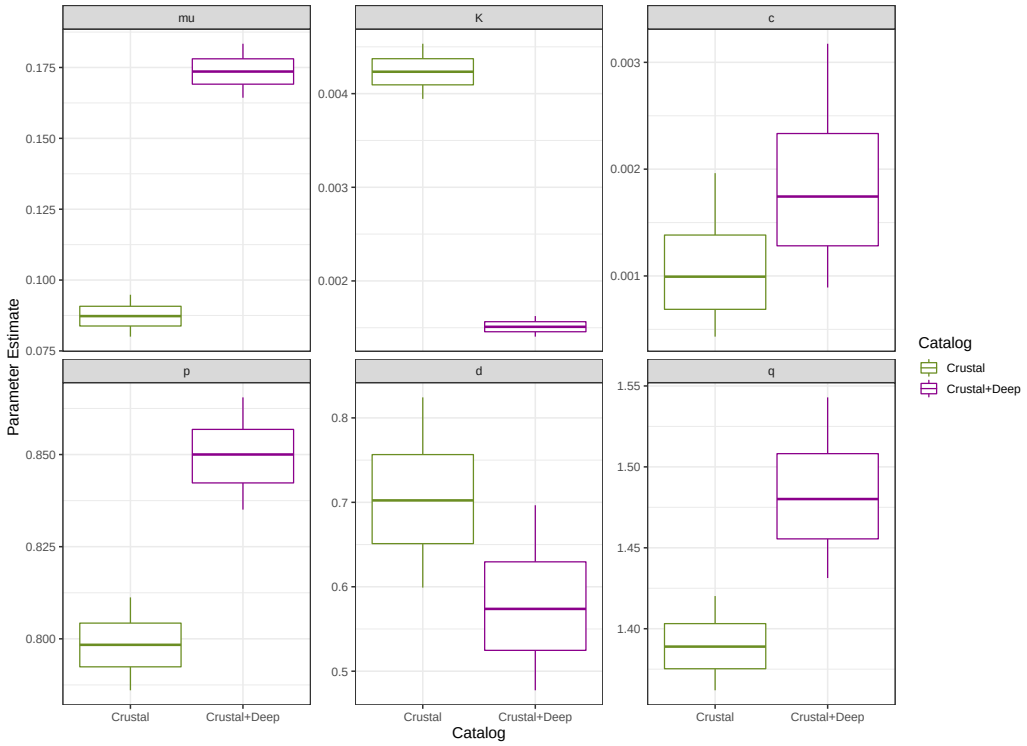

Figure 3.28: Bayesian credible intervals for ETAS parameters for the baseline catalog, PNW North, using a crustal seismicity-only catalog or one that combines crustal and deep seismicity.

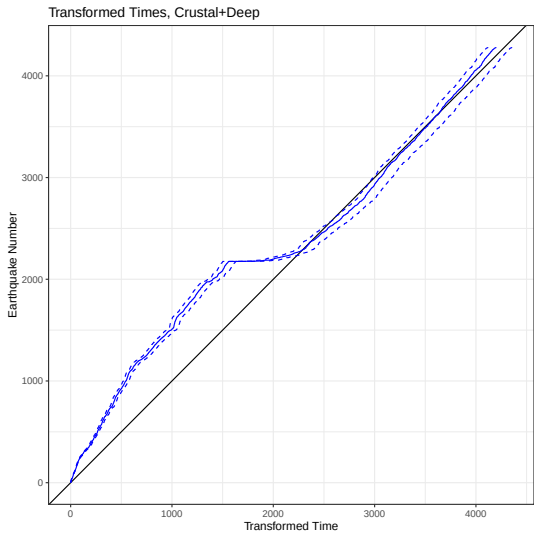

Figure 3.29: Transformed times for the Bayesian fit to the combined crustal and deep seismicity catalog.

Spatial Zones Figure 3.31 shows Bayesian inference for the target baseline catalog and the auxiliary catalog (for crustal events in PNW North). Estimates are somewhat different for parameters μ, K and q. When we model the auxiliary zone, $\hat{\mu}$ and $\hat{K}$ are substantially lower, and $\hat{q}$ is somewhat lower as well. The other parameter estimates are largely unchanged.

3.8.5 Quantifying uncertainty from measurement error

Bayesian inference also enables quantifying how catalog measurement error affects ETAS estimates. We fit the ETAS model to 100 jittered catalogs as described in Section 3.6.2 and show posteriors for nine randomly selected jittered catalogs in Figure 3.32. The 95% credible interval for the grand posterior encompassing the reported and all jittered catalogs is plotted in black. The uncertainty for all parameter estimates increased when considering measurement error (black line). The estimates for d and q are raised for jittered catalogs and the estimate for K is slightly reduced.

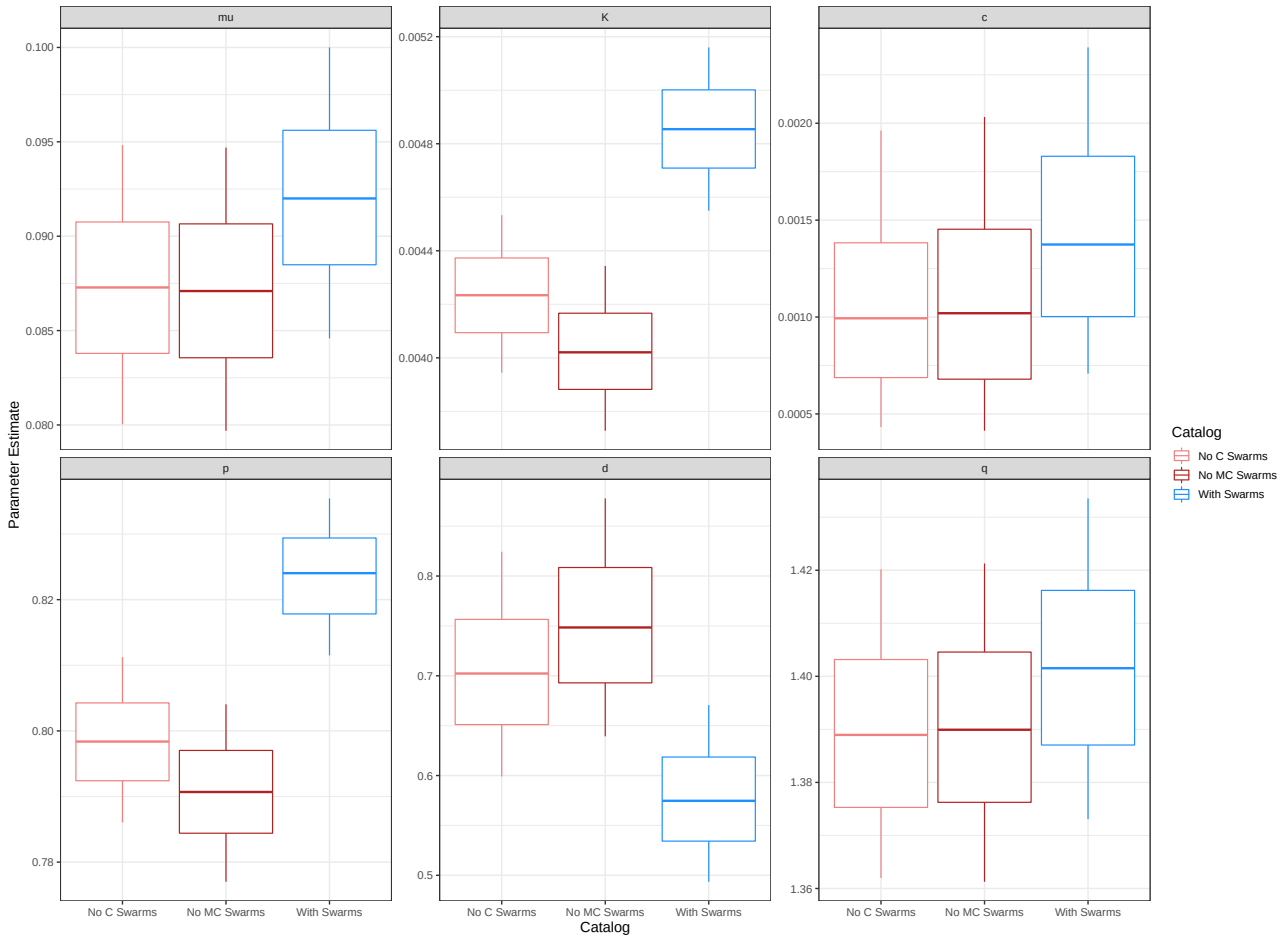

Figure 3.30: Bayesian credible intervals for ETAS parameters for the baseline catalog, PNW North, using catalogs that either exclude confirmed (C) swarms, maybe and confirmed (MC) swarms, or include all swarms.

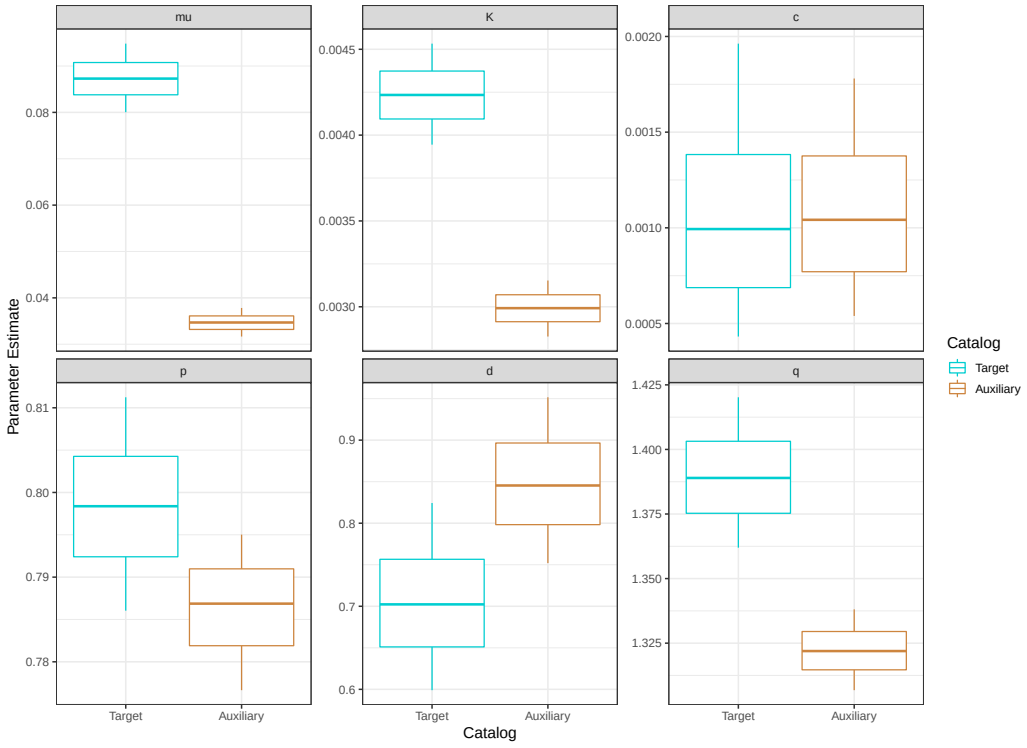

Figure 3.31: Bayesian credible intervals for ETAS parameters for the target (baseline) catalog and the auxiliary catalog.

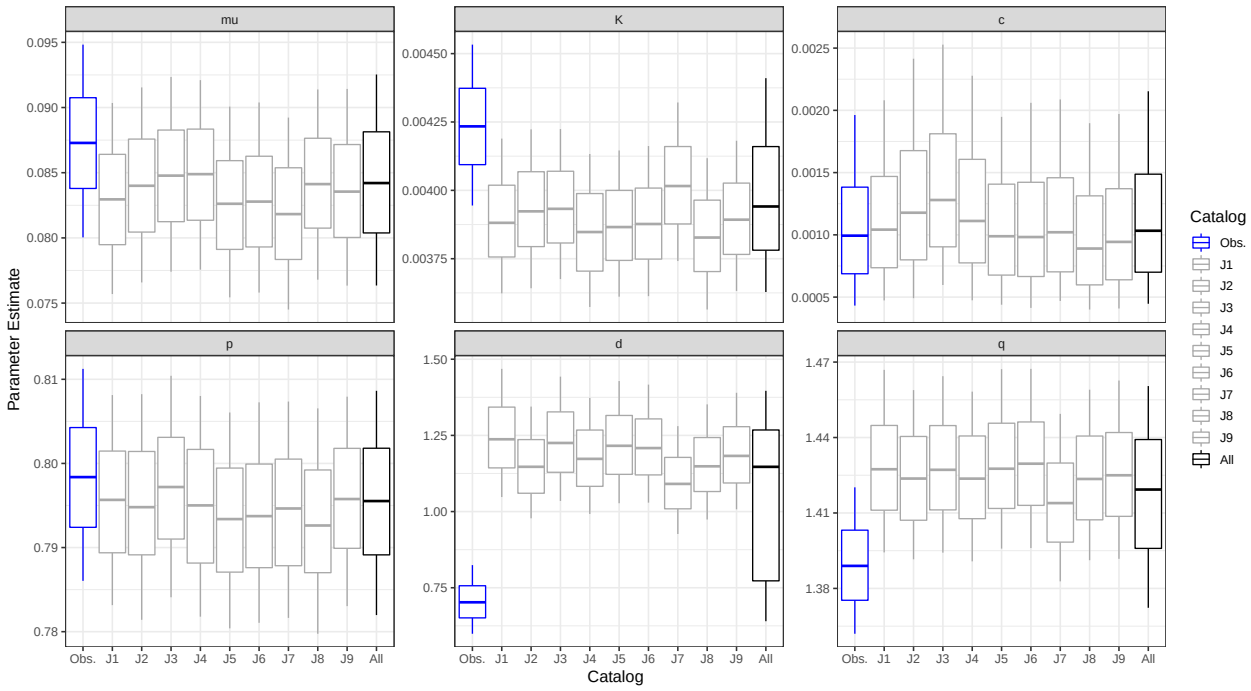

Figure 3.32: Bayesian credible intervals for ETAS parameters for the reported (baseline) catalog, as well as nine jittered versions of this catalog.

3.9 Discussions

3.9.1 Bayesian inference for improved ETAS modelling

We present a Bayesian inference method for parameter estimation and uncertainty quantification for the ETAS model. This approach is attractive because it does not have the problems long reported for ETAS parameter MLEs, namely that they are unstable and depend on numerical optimization procedures and their initial values. Furthermore, we can quantify uncertainty in a straightforward way using parameter posteriors, without appealing to the tenuous asympototic arguments necessary for Hessian-based standard errors. Our approach has two components: first, we derive a simpler likelihood function for the ETAS model, conditional on the catalog branching structure. This branching likelihood function does not have the level of parameter correlation as the traditional likelihood, though it is also flat around the likelihood maxima for some parameters (Figure 3.19).

Second, we adopt the Bayesian paradigm for parameter estimation. Rather than numerically maximizing the likelihood function to estimate parameters, we simply evaluate it when drawing from the posterior distribution for each parameter. This avoids the numerical instability of either ETAS likelihood and thus removes the effect of initial values on ETAS estimation and uncertainty quantification (Figures 3.15, 3.16). The iterative nature of our Bayesian approach is appropriate for estimation conditional on a branching structure; that is, the standard (gradient descent) approaches to optimize likelihoods would not work with the branching likelihood, as results would be entirely dependent on the given branching structure, which would need to be based on initial values for ETAS parameters. Any frequentist estimation using the branching likelihood would need to use an Expectation-Maximization style approach, like in Veen and Schoenberg (2008).

Though this work is based on the approach presented in Ross (2021), we tailored it to the specifics of the PNW. Unlike previous studies, we developed a new solution for the spatial

integral in the likelihood that is applicable to any region, like the PNW, that has ample seismicity at its borders. This improves on the previous best-practice solution (Schoenberg, 2013), which may have caused bias in the PNW, which has substantial seismicity at its borders. We also elicited priors from experts that reflected two competing sets of beliefs about what is known about the PNW. The work of Ross (2021) and earlier temporal-only Bayesian ETAS models (Ross, 2017; Kolev and Ross, 2019) use uninformative priors, even when modeling real catalogs, for which prior beliefs are likely to exist.

We apply this approach to model a new catalog for the PNW. We split the catalog into northern and southern regions, due to distinct differences in network coverage and earthquake detection between them. After evidencing the need for the ETAS model structure by model comparison to a simpler NHPP model, we find that our Bayesian approach for ETAS fits the PNW North slightly better than MLEs from different optimizers (Figure 3.20). The fit to the much sparser PNW South is somewhat worse for all estimation techniques, though the Bayesian solution has a good fit for the lowest M_c.

The Bayesian approach has the additional benefit of being able to quantify how different model and catalog choices affect ETAS estimates and their uncertainty. We can specify prior distributions that represent particular beliefs about earthquake and aftershock dynamics in the studied region. We can also quantify how such catalog issues as swarms or different tectonic regimes affects parameter estimates and their uncertainty. We have also demonstrated one way to use our approach to propagate catalog measurement error into ETAS estimation.

3.9.2 PNW aftershock parameters in context

The Bayesian approach provides robust estimates of ETAS parameters for the PNW, that are insensitive to initial values. We can compare them to other reported estimates for the PNW or other subduction zones. Gomberg and Bodin (2021) built a Reasenberg-Jones model for individual sequences in the PNW, which uses the productivity law as parametrized in

Equation 3.1. Our K estimates for the crustal regime are consistent with the range of a estimates given in Gomberg and Bodin (2021) for this regime. For example, taking our 95% credible interval for K of 0.0039-0.0044 (with α fixed to $\hat{b}_{MLE}\log(10) = 0.935 \cdot \log(10)$), our expected total number of aftershocks for an M6 mainshock in the crustal regime is between 21.3-24.2, which is consistent with the a estimates of -2.5 to -3 reported for these mainshocks in Gomberg and Bodin (2021).

Authors have previously argued that the PNW is somewhat of an outlier from other subduction zones, particularly due to its lower overall seismicity level and aftershock productivity (Gomberg and Bodin, 2021). Our results offer some support to this argument. When specifiying a prior based on previously reported ETAS estimates from other subduction zones, posterior estimates for p and q were concentrated at the extremely low end of the prior. This suggests that using aftershock parameters based solely on other subduction zones to model the PNW (as is currently being done in the USGS operational aftershock forecasting system (Page et al., 2016)) could lead to systematically incorrect results. With that said, the small set of reported estimates that were available may have artificially limited the true ranges of subduction zone parameters. Furthermore, we find some similarities between our results for the PNW North and subduction zone results from previous global studies; for example, Page et al. (2016) reports a $\hat{p}$=0.81 in a Reasenberg-Jones model for all subduction zone (onshore) regions, which is similar to our Bayesian posterior median for $\hat{p}$ (Figure 3.14).

3.9.3 Effects of catalog choices and issues on ETAS parameter estimation

Using Bayesian inference allows us to quantify the effects of specific catalog choices on ETAS parameter estimation. Including deep earthquakes in the catalog had a meaningful effect on estimates for several parameters, especially μ, K, p and q (Figure 3.28). The background seismicity rate nearly doubled when combining both regimes and smaller aftershock sequences

were expected, that should decay more quickly in time and space. This is consistent with other reported findings on the smaller productivity of deep earthquakes (Gomberg and Bodin, 2021), most of which belong to the background set in the estimated branching structure. It should be noted that splitting the catalog into the two distinct regimes ignores the fact that crustal aftershocks have been associated with deep mainshocks in other subduction zones (Bilek and Lay, 2018).

Adding swarms to the catalog also made an important change in estimates of K and p parameters, but left the other parameters nearly identical (Figure 3.30). Somewhat larger aftershock sequences are expected when swarms are included, which are expected to decay slightly slower. This may be caused by several large swarms in the catalog (three of these had at least 25 events).

Adding the seismicity in a margin zone around the target region also affected estimates of the μ, K and q parameters, though leaving the other parameters nearly identical (Figure 3.31), while also reducing the uncertainty for these parameters. When considering this larger area, the background rate was more than halved, and somewhat shorter sequences are expected which decay slightly more quickly in space. This may be due to this margin zone generally having sparser seismicity (see Figure 3.7), outside of several isolated pockets (e.g., the Mendocino Triple Junction).

We propagated the effects of catalog measurement error by jittering catalogs based on reported location and magnitude errors and combining posterior distributions modelled over numerous jittered catalogs. Estimates for d and q were slightly higher for the jittered catalogs, so sequences are expected to have a larger spatial offset before decaying in space slightly more quickly. This may reflect two distinct outcomes of the jittering process: the larger spatial separation of likely aftershocks from their mainshocks (larger d) and broken links between likely aftershocks and mainshocks caused by jittering (a possible explanation for the larger q). Posteriors for all parameters were also wider than when modelling the reported catalog,

reflecting the greater uncertainty in parameter estimation.

3.9.4 Limitations and future work

This study has a number of limitations that should trigger future work. It is possible that another specification of the ETAS model may better describe the seismicity of the PNW. Other ETAS models allow μ_{ST} to be spatially varying; that is, estimating $\mu_{ST}(x, y)$ across the spatial zone using, e.g., bivariate kernel density estimation. The level of smoothing in such a kernel density estimation should be carefully chosen as it can bias estimates for triggering parameters (Harte, 2013), especially those more correlated with μ_{ST}. In addition, the aftershock area has been described as "highly correlated with the mainshock's magnitude" (Ogata and Zhuang, 2006); therefore, the characteristic distance for a mainshock before aftershocks decay in space (represented by parameter d) can be specified to vary with magnitude. The parameter d is often replaced by $d\exp(\gamma \cdot M)$ in other ETAS models (e.g., (Nicolis et al., 2015) and (Zhuang, 2011)), which can be done as a future step. Furthermore, we have imposed the assumption of self-similarity by fixing α to $\hat{b}_{MLE}\log(10)$. MLEs for α are known to be biased downwards (Seif et al., 2017; van der Elst, 2017) to "favor branching structures where many aftershocks of large events are assigned to smaller, nearer events" (Schoenberg, 2013). This problem is especially present when assuming isotropic aftershock zones, as we have done (Hainzl, 2016), which may be another source of bias, as aftershocks are expected to spread anisotropically around the fault rupture (Hainzl et al., 2008). It remains to be seen if this problem persists, as well as the others mentioned, for Bayesian fits to this branching likelihood. Future model development for the PNW should explore models where μ_{ST} is spatially-varying, d is magnitude-dependent, aftershock locations spread anisotropically around the mainshock and α is estimated along with the other parameters.

We only performed basic model evaluation on our fitted Bayesian ETAS models and more work should be done here. We can assess the model's fit of the spatial distribution

of the observations using a similar proper score to the one described in Section 3.4.1, but based on the fitted conditional intensity (Heinrich-Mertsching et al., 2021). Since aftershock forecasting is a key application for these models, we may also generate aftershock forecasts retrospectively (for large aftershock sequences classified in Chapter 2, Section 2.4.1) or for future mainshocks in a prospective manner. We may use likelihood-based tests (Schorlemmer et al., 2018) or spatial residual scores (Clements et al., 2011) to evaluate different dimensions of our model's forecasting performance (i.e., the forecasted number of aftershocks and spatial and magnitude distributions) against observations.

We split the PNW at 45°N into northern and southern regions, based on the completeness study in Chapter 2, Section 2.8.1. This can be refined by adopting the approach of Nicolis et al. (2015), which estimated ETAS parameters separately for overlapping latitude bands over Chile. It could be useful to investigate in a similar fashion at which point of latitude the patterns (parameter values) of the PNW North change, or degrade into higher uncertainty.

Our investigation of measurement error also uncovered that it does indeed affect parameter estimation, both for spatial parameters (d, q) and non-spatial parameters. This is consistent with other literature on the importance of magnitude and epicentral errors on seismicity model estimates (Wang et al., 2010). Extensions of the ETAS model are needed that explicitly account for measurement error. Spatial point process models exist that incorporate noise (Lund and Rudemo, 2000) and Bayesian estimation has previously been developed for these (Chakraborty and Gelfand, 2010). These studies lay out general methodology for spatial point process models and future work should extend these for spatiotemporal ETAS models.

Chapter 4

EFFECTIVE UNCERTAINTY VISUALIZATION FOR AFTERSHOCK FORECAST MAPS

And I'm still picking up the rubble from my last quake,
And I was sitting on the shadow coming from my last break.
Had to take a second look, then a third, then etcetera.
He said, don't let your eyes get ahead of ya.

"Bet," Mereba (2018)

4.1 Introduction

[1] In the previous chapter, we developed statistical machinery to do Bayesian inference on the spatiotemporal ETAS model for the Pacific Northwest (PNW), using it to quantify how catalog issues affect parameter estimates and their uncertainty. Such a model can be used to issue probabilistic forecasts for the number, locations and times of aftershocks following a large earthquake. Aftershock forecasts are associated with large uncertainties, particularly early in the sequence (McBride et al., 2020), which also vary over space. In this chapter, we design several uncertainty visualizations for aftershock forecasts and test them with human users in a controlled experiment using several cognitive tasks.

[1]Note: This chapter is in revision for publication in *Natural Hazards and Earth System Sciences* as Schneider, M., McDowell, M., Guttorp, P., Steel, E. A., Fleischhut, N. (2022). Effective uncertainty visualization for aftershock forecast maps. *Natural Hazards and Earth System Sciences*.

4.1.1 Aftershock forecasts and their uncertainty

Forecasts are generally produced from statistical seismicity models via simulation. Once a large earthquake (mainshock) happens, the model is fit to the catalog containing this mainshock. Synthetic catalogs for a given future period can be generated through a statistical simulation procedure, like the one described in Chapter 3, Section 3.5, using the fitted parameter values and their uncertainties (Milner et al., 2020). For example, using the ETAS model fit with Bayesian inference, we can simulate a given period of future activity (say, one week or one year) with parameters randomly drawn from their posterior densities (e.g., Chapter 3, Figure 3.11). If we repeat this procedure many times (say 10,000), we get a distribution of aftershock forecasts for the given time period (possible futures that are consistent with the model built on past data). We can then summarize the number, time and space dimensions of these simulated aftershock sequences; for example, understanding the aftershock time duration is often of interest (Michael, 2018), as is capturing the total number of expected aftershocks (McBride et al., 2020). We can then issue forecasts that have both the most probable guess (the central tendency of this distribution) as well as some measure of its uncertainty (the spread of this distribution) in the full set of simulated sequences. Importantly, the spatial rate of earthquakes (the number of earthquakes per unit area) during an aftershock sequence is known to follow a highly skewed distribution (Saichev and Sornette, 2007), where many more earthquakes may occur if aftershocks trigger their own sequences of additional aftershocks.

Such forecasts are highly sought after by diverse user groups, such as emergency managers and the media (Gomberg and Jakobitz, 2013), in order to inform decisions about disaster declarations and crisis information delivery (Becker et al., 2020; McBride et al., 2020). Recently, several national scientific agencies, including New Zealand's GNS Science (Becker et al., 2020) and the United States Geological Survey (Michael et al., 2020), also began releasing aftershock forecasts to the public. In these public communications, the forecast

distribution has been represented in tables of either (1) the expected number of earthquakes above some magnitude for fixed time periods (e.g., one day, one week) after the mainshock or (2) the probability of an event above some magnitude occurring within a time period.

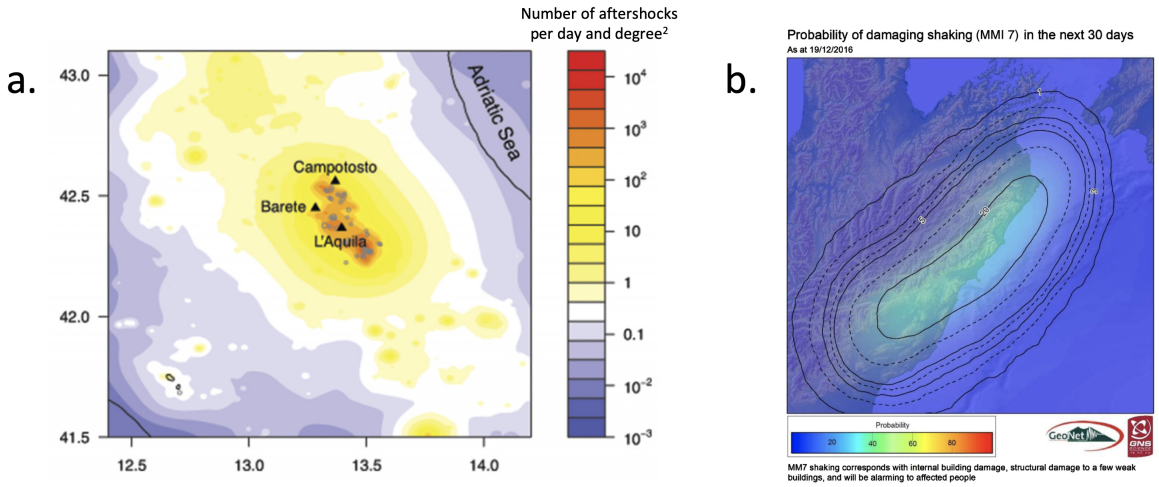

Figure 4.1: Two examples of aftershock forecasts. (a) A map showing the expected daily number of aftershocks above magnitude 2.0 several weeks after the L'Aquila, Italy earthquake of 2006 (reprinted from Murru et al. (2015)). (b) A map showing the forecasted probability of a damaging aftershock (defined as having an Modified Mercalli Intensity score above 7) in the month following the Kaikōura, New Zealand earthquake of 2016, released to the public by New Zealand's GNS Science (reprinted from Becker et al. (2020)).

While state-of-the-art models can forecast aftershocks with some degree of accuracy (Schorlemmer et al., 2018), they also have substantial uncertainty. Aftershock models like ETAS are built on catalogs in a given seismic region; however, the considerable variation between aftershock sequences in any region contributes to a large spread in parameter estimates (e.g., see Chapter 3, Figures 3.14). This contributes to a large spread in the forecasted number of aftershocks following any mainshock. Forecast uncertainty can be communicated

by directly giving this spread for the forecasted number of aftershocks. It can also be communicated implicitly by giving the probability that an aftershock above some magnitude may occur, where a very high or low probability can imply lower uncertainty and a middle probability (e.g., 50%) can imply higher uncertainty. Uncertainty has already been communicated for tabular forecasts by providing these probabilities and ranges of the expected number of aftershocks across the entire forecast region (e.g., Becker et al. (2020); Michael et al. (2020)).

Aftershock activity varies over space (Zhuang, 2011) and forecast maps are commonly requested by users (Michael et al., 2020). Figure 4.1 shows example maps following the L'Aquila earthquake and the magnitude 7.8 earthquake in 2016 in Kaikōura, New Zealand. Aftershock forecast models will be better calibrated in areas with frequent aftershock sequences than where aftershock activity is sparse, meaning that forecast uncertainty also varies over space. But the uncertainty of the forecast distribution is usually not communicated in aftershock forecast maps; for example, the map in Figure 4.1a shows the expected number but not its spread and the probability map in Figure 4.1b does not make uncertainty explicit.

As evident in the response to the L'Aquila earthquake, omitting uncertainty from forecasts can affect people's perceptions and responses to the associated risks. When uncertainty is not displayed, users tend to form their own understanding of where uncertainty is higher or lower, which may not coincide with its actual patterns (Ash et al., 2014; Mulder et al., 2017). Aftershock forecasts that do not communicate uncertainty may therefore result in users misunderstanding the forecast (Fleischhut et al., 2020; Spiegelhalter et al., 2011); for example, in previous studies in the weather domain, users incorrectly expected wind and snow to be lower than forecasted when these had high forecasts, but not when the forecast was low (Joslyn and Savelli, 2010). Misinterpreting a forecast could become particularly problematic when, due to the skewed distributions of aftershock rates, high uncertainty could mean a

greater chance that more aftershocks will occur than forecasted.

While studies across multiple domains have found that displaying uncertainty when communicating forecasts can improve responses related to judgment and decision-making (Kinkeldey et al., 2017; Joslyn and LeClerc, 2012; Nadav-Greenberg and Joslyn, 2009), there is a paucity of literature on uncertainty visualization for spatial aftershock forecasts, or other seismic communications (Pang, 2008). Thus, despite recommendations to incorporate uncertainty into communications from earthquake models (Bostrom et al., 2008), there are currently no guidelines on how to do so or on what visualization techniques could help users to understand and incorporate uncertainty to inform their judgments (Doyle et al., 2019). The purpose of the present study is to develop and evaluate different approaches to visualizing uncertainty in the distribution of spatial aftershock forecasts. These approaches may also be useful for other natural hazards whose forecasts follow a similar skewed distribution.

4.1.2 Visualizing uncertainty for natural hazards

Aftershock forecasts maps can show how many aftershocks are expected for some time period at different locations throughout a region (e.g., Figure 4.1a). Uncertainty visualizations have already been designed for similar maps for other geohazards and evaluated using task-based experiments (Kinkeldey et al., 2014). These previous studies can serve as a natural starting point for visualizing uncertainty for aftershock forecast maps. We review the literature that evaluates uncertainty visualizations in geoscientific domains.

One approach to representing uncertainty in geospatial forecasts is by plotting the center of the forecast distribution (e.g., the median or mean) and uncertainty as the spread of the distribution (e.g., the standard deviation or width of a credible interval). The uncertainty is either visualized (1) in adjacent maps, where the center and uncertainty are displayed in separate maps or (2) within the same map, using, for example, color, patterns, opacity or symbols to visualize the uncertainty (Pang, 2008). When the forecast and its uncertainty are

represented within the same map, designs using color lightness or transparency (i.e., fading the color to a background color like white or gray) have been found to be effective. For instance, Retchless and Brewer (2016) evaluated nine designs for forecast maps of long-term temperature change. They tested an adjacent design against designs that varied color properties (hue, lightness, and saturation) or used textured patterns to display the uncertainty in a single map together with the forecast. Participants had to rank several zones on the map(s), first by their forecasted temperature change and then by their uncertainty. Visualizations where the transparency of the color increased with uncertainty led to more accurate uncertainty rankings than other color-based designs, although the adjacent approach was most accurate.

When forecast uncertainty is represented using the probability of exceeding a threshold value (such as Figure 4.1b), studies have also found visualizations using transparency to work well. For instance, Ash et al. (2014) found that a forecast map using red shades of decreasing lightness to depict the forecasted probabilities of a tornado's location were associated with greater willingness to take protective action than forecast maps using rainbow hues or a deterministic map showing only the boundary of the zone of elevated tornado probability (which omitted uncertainty altogether). Similarly, Cheong et al. (2016) found that maps using color hue or lightness to represent the probability, in this case of a bushfire, led to better decisions about whether to evacuate from marked locations on a map (based on realizations of bushfires from the model), compared to a boundary design that omitted uncertainty. A review of dozens of geospatial uncertainty visualization evaluations has similarly concluded that color transparency or lightness can be effective in communicating uncertainty (Kinkeldey et al., 2017).

A third approach to communicate the uncertainty in a forecast distribution is by visualizing the bounds of an interval describing the distribution (e.g., a 95% confidence interval). Although such interval-based maps have recently been proposed, for instance, to communi-

cate public snowfall forecasts by the US National Oceanic and Atmospheric Administration (Waldstreicher and Radell, 2020), there are few studies evaluating the effectiveness of this approach. Nadav-Greenberg et al. (2008) tested uncertainty visualizations that paired a map of median wind speed forecasts with either an adjacent map of the forecast's margin of error, a "worst case" map showing the 90th percentile of the forecast distribution, or boxplots of the forecast distribution at multiple locations. Participants tended to predict higher wind speeds at a given location using the worst case map relative to the other visualizations. We are not aware of any other studies that have investigated the effects of interval-based uncertainty visualizations for forecast maps.

In the present study, we evaluate three approaches to visualizing uncertainty in aftershock forecast maps. We focus on representing the distribution's spread rather than probability of exceedance because there has been less research on this way of communicating forecast uncertainty (Spiegelhalter et al., 2011) and it is of general interest to other natural hazard forecasts as well. Specifically, we compare a novel interval-based approach where uncertainty can be inferred from the bounds of a 95% confidence interval, a commonly-used approach that uses color transparency to display the uncertainty within the forecast map and the classic adjacent display. These uncertainty visualizations are compared against a forecast depicted without uncertainty, which is most common in practice and thus the natural baseline (see Figure 4.1).

4.1.3 Evaluating the effectiveness of uncertainty visualizations

An effective uncertainty visualization should facilitate not only the reading of the forecast and uncertainty off of a map but also help users to apply this information appropriately. While designing map-reading tasks is more straightforward (e.g., asking users to read or locate particular areas on a map), previous judgment or decision-making tasks have not been designed to systematically evaluate how users interpret uncertainty given different designs

(Hullman et al., 2018; Kinkeldey et al., 2017). Further, without defining tasks in line with specific communication goals, it is not possible to identify what constitutes an effective uncertainty visualization.

First, tasks should be designed such that the effect of uncertainty visualizations can be disentangled from other features of the maps or the task. For instance, Viard et al. (2011) asked participants to rank the risk of over-pressure based on an adjacent or a pattern-based uncertainty visualization of estimated pressure of oil reservoirs. Several locations were compared that varied both in mean pressure and uncertainty. While a difference in rankings was found between the visualization conditions, it was impossible to conclude whether either visualization led to more reasonable responses, as the selected locations did not vary systematically and no normative ranking was specified. Other studies have implemented similar ranking tasks using locations that do not uncover how different designs affect how users understand forecast uncertainty (e.g., Deitrick and Edsall (2006); Scholz and Lu (2014)). Additional issues include tasks that ask for forecast and uncertainty information to be used separately (Retchless and Brewer, 2016), or results that are aggregated across many trials without accounting for the locations' forecast/uncertainty levels (e.g., Cheong et al. (2016)).

Second, evaluation tasks should be designed to link to specific communication goals relevant across user groups, in order to inform designs that could serve a wider audience. In Padilla et al. (2017), participants had to decide whether to move an oil rig in the face of an oncoming hurricane, receiving the hurricane track forecast either with a cone of uncertainty or an ensemble of possible hurricane track curves. By systematically moving the oil rig's location across trials, the authors found that decisions were influenced by whether the oil rig was inside the cone of uncertainty or directly on top of an ensemble hurricane track. But the generalizability of these results outside of this highly specific decision task is debatable. Other studies similarly evaluate uncertainty visualizations with well-designed decision tasks inspired by specialist use cases but it is unclear what their results indicate for designing pub-

lic forecast maps (Seipel and Lim, 2017; Correll et al., 2018). In contrast, other experimental literature on text-based uncertainty communication (e.g., Burgeno and Joslyn (2020); Joslyn and LeClerc (2012); Van Der Bles et al. (2019)) works with tasks that are relevant across user groups and designed to reveal how the communication was understood. Visualization scholars have also urged more generalizable tasks in evaluation experiments for uncertainty visualizations (Crisan and Elliott, 2018; Meyer and Dykes, 2019).

Building off this literature, we seek to improve how uncertainty visualizations are evaluated, using an experimental task framed around ubiquitous goals that an uncertainty visualization should ideally achieve. We frame our task around targeted needs users have from aftershock forecast maps.

4.1.4 Communication needs for aftershock forecasts

To understand the informational needs of a common user group of aftershock forecasts, we interviewed five emergency management officials from Washington State, USA, four of whom were directors of either state or county-level emergency management offices (the fifth participant was a volunteer at a county-level emergency management office). The interviews focused on how scientific information is used when responding to natural disasters, and how visual features of previous disaster scientific communications affected how the emergency managers used them for decision-making needs. We also asked about how uncertainty in scientific communications affects their use for disaster response. We began by posing general questions and then narrowed in on previous experiences with non-earthquake hazards, as most participants had not yet worked in earthquake response.

Interviews were recorded and then summarized across participants, finding similarities and differences across the individual responses. After synthesizing the interviews, we isolated two commonly mentioned questions that an aftershock forecast should facilitate answering and posited that these communication needs would also be relevant to a general user audience:

1. Where is it likely that aftershocks will or will not take place ("sure bets"; i.e., areas with high/low forecasted aftershock rates and low uncertainty)?

2. Where is a bad surprise possible due to the high uncertainty of the forecast ("surprises"; i.e., areas with high uncertainty, that can yield an aftershock rate higher than forecasted)?

Using these communication needs as a guide, we designed a comparative judgment task to evaluate how different visualization approaches can support these communication needs. Within this task (see Section 2.3), participants judge two locations with systematically varying forecasted aftershock rates and uncertainty levels.

Given locations with *equally low uncertainty* but *different forecasted rates*, users should be able to correctly identify that more aftershocks are expected at the location with the higher forecasted rate. We refer to this as a sure bet trial (addressing communication goal 1). Given locations with *equal forecasted rates* but *different uncertainty*, it is more likely that more aftershocks than forecasted will occur at the location with higher uncertainty, compared to the location with lower uncertainty. As this could lead to a bad surprise, we refer to this as a surprise trial (addressing communication goal 2).

For surprise trials, the forecast level determines whether a response exists that can be considered "correct". When both locations have *low forecasted rates*, users should identify the location with low uncertainty (a sure bet to have few aftershocks) as having less potential for aftershocks compared to the location with high uncertainty. This is because high uncertainty means that in the long run, much higher aftershock rates are possible than forecasted due to the skewed distribution of aftershock rates. When both locations have *higher forecasted aftershock rates*, it is not possible to define a correct answer for where users should expect fewer or more aftershocks. In this case, comparing locations with lower uncertainty (where high rates are more certain) to locations with higher uncertainty (where even higher rates are possible) may be subjective to each user, for instance, based on their risk preferences.

It is thus even more important to understand how different uncertainty visualizations affect user judgments within this situation and which visualizations lead users to recognize that forecasts with high uncertainty could result in more aftershocks than forecasts with low uncertainty.

Using the comparative judgment task, we methodically investigate how effectively different uncertainty visualizations fulfill both communication needs. Unlike previous tasks in the uncertainty visualization literature, our task allows to infer which uncertainty visualizations produce responses about uncertainty that are consistent with the forecast distribution and for communication needs relevant across user groups.

4.1.5 Research aims and contributions

In the present experiment, we test three uncertainty visualizations for aftershock forecast maps to evaluate which visualization approach best serves the above communication goals. Specifically, we seek to answer the following research questions:

1. How well can people read off the forecasted aftershock rate and its level of uncertainty from the different visualizations?

2. How do the visualizations affect people's judgments about where to expect more aftershocks?

    (a) How accurate are people's expectations when comparing locations with varying forecasted rates but low uncertainty, and how does accuracy differ by visualization?

    (b) Where do people expect more aftershocks when comparing locations with the same forecasted rates but different uncertainty, and how does this differ by visualization?

We evaluate the aftershock forecast visualizations in an online experiment using a broad sample of participants from the United States. Our experiment allows to infer how both classical and novel uncertainty visualizations affect potential users' perceptions of aftershock forecasts. Moreover, our work adds to the literature on visualizing uncertainty for natural hazards by using a systematic judgment task based on user needs which may be applicable across hazards.

4.2 Experimental and statistical methods

4.2.1 Participants

Participants were recruited via Amazon Mechanical Turk (MTurk) to complete an online study about map-reading and judgments about future aftershocks using different forecast visualizations. We restricted recruitment to the western U.S. states of California, Oregon and Washington, as these states are seismically active and participants would likely have some earthquake awareness. While the MTurk population we sample does not match these states' populations, comparisons between MTurk-based and probability-based samples of the US population often yield similar results (Zack et al., 2019), and MTurk has been used to recruit participants in previous uncertainty visualization evaluations (Retchless and Brewer, 2016; Correll et al., 2018). MTurk workers were eligible to participate if they had an approval rating of $\geq$99% (a best practice for screening participants in the research community (Paolacci and Chandler, 2014)) and were using a computer screen $\geq$ 13 inches in diagonal (to ensure that our maps could be seen sufficiently well). We also required participants to answer four multiple choice control questions about the study, based on initial instructions. Participants were given two attempts to answer the questions.

Of the 1392 participants who consented to participate in the experiment, 941 passed the control questions and 893 completed the full experiment. Seven participants self-reported to have an age less than 18 and were excluded from analysis, and two participants were

excluded because they took too many attempts at control questions, leaving a final sample of 884 participants (46.7% female, median age of 32 years (range: 18-77 years)), split evenly by condition. The size per condition was pre-specified by a power analysis, using effect sizes found in a pilot study.

The experiment was incentivized and participants were informed of this before consenting to the experiment. To calculate incentive bonuses, we randomly drew two map-reading trials and three comparative judgment trials per participant (see Section 2.3). We gave a bonus USD 0.10 for each correct map-reading response and for each judgment response that matched the outcome of a hypothetical week of aftershock activity simulated from the presented forecast map. Including the baseline payout (USD 1.80), participants could earn over U.S. minimum wage for just two correct trials (e.g., the map-reading trials), matching community standards (Paolacci et al., 2010).

We pre-registered the experiment and analysis plan in the Open Science Framework. The study was approved by the Max Planck Institute for Human Development ethics committee.

4.2.2 Materials

Creating forecast maps from earthquake model output

Each uncertainty visualization (UV) showed a weekly forecast for the number of aftershocks following a major earthquake together with its estimated model uncertainty. These hypothetical forecasts were created from the output of an ETAS model (see Chapter 3, section 3.1.1) built for a catalog for the entire Pacific Northwest of earthquakes of magnitude ≥ 2.5. This was an earlier version of the model (Schneider and Guttorp, 2018), where parameters were fit by maximum likelihood estimation (MLE) and the background rate μ_{ST} was allowed to vary spatially (Ogata, 2011). We computed the estimated total seismicity rate $\hat{\lambda}$ using MLEs for all parameters and with $\hat{\mu}_{ST}$ fit with kernel density estimation. The values of $\hat{\lambda}$ thus also varied over space and could be mapped.

We cropped an area of roughly 2000 km^2 from this model output to represent a forecast for the number of aftershocks above magnitude 2.5. This output differs from real probabilistic aftershock forecasts, which are typically computed from many simulations from a forecast model (as described in Section 4.1.1); however, this model output still maintained characteristic spatial features of aftershock forecasts. In particular, it showed a rapid and isotropic spatial decay from higher to lower forecasted rates, fading into a low (but non-zero) background rate (see e.g., Marzocchi et al. (2014); Zhuang (2011) and Figure 4.1). We scaled the model output up to achieve aftershock rates similar to the first week after recent major sequences (e.g., 2018 Anchorage, AK and 2019 Ridgecrest, CA, which had 500-1000 aftershocks above magnitude 2.5 in similarly sized zones around the mainshock).

We cropped two distinct $\sim$2000 km^2 areas of the map to represent forecasts with different spatial patterns (hereafter referred to as "forecast regions"). Maps were labelled with "Longitude" and "Latitude" along the axes, without tick marks. We randomized experimental trials between these forecast regions to allow for more trials due to expected variability in task response within participants. As the focus of the study was on how to visualize a forecast and its uncertainty, geographical features (e.g., topography, cities, roads) were omitted in forecasts maps to avoid these potentially confounding task responses (Nadav-Greenberg et al., 2008).

Visualization of forecasted rate

To visualize *forecasted rates*, the spatial rate distribution derived from the model was mapped onto a grid of 20×20 cells (each grid cell thus represented an area of approximately 2.8 km × 1.8 km). We then classified the numeric rates into categories, as is commonly done for natural hazards forecasts and recommended by the visualization literature (Correll et al., 2018; Thompson et al., 2015). We classified rates into five categories, to keep complexity manageable while maintaining sufficient information to display realistic trends. A sixth

(uppermost) class was subsequently added that was solely used in the Bounds UV (see below). The category cutoffs came from quantiles of the scaled rate distribution. These cutoff values thus revealed the skewed and concentrated nature of aftershock rates. We refer to this map as the *most likely forecast* map, as it is based on parameter MLEs, and labelled it "Forecasted Number of Aftershocks". We interpreted each location's most likely forecasted aftershock rate as the center (i.e., the mean) of the forecast distribution of aftershock rates for that location (we hereafter abbreviate "most likely forecast of aftershock rate" simply with "rate"). Rates were visualized using a 6-color palette with colors increasing uniformly from yellow to red and ending with brown (see Section C.1 in the Appendix for additional detail on color selection).

Figure 4.2 (left panel only) shows the Rate Only condition, which omits forecast uncertainty. This is the current manner of displaying aftershock forecasts to public audiences, when maps are used (see Figure 4.1).

Visualization of forecast uncertainty

To visualize *forecast uncertainty*, we created uncertainty maps that followed realistic patterns and would also have design features needed for experimental tasks. Uncertainty maps had the same dimensions as the rate maps; that is, our uncertainty maps are not based on ETAS model outputs. Uncertainty values were generated for each grid cell using random number generators and interpreted as the standard deviation of the forecast distribution. These maps adhered to the smoothly varying patterns expected of forecast uncertainty, with higher uncertainties sometimes, but not always, corresponding to areas of higher rate. These uncertainty values were then binned into three categories ("Low", "Medium", "High") to allow for sufficient information to display the above-described patterns. In this article, we present maps and describe experimental materials using an "uncertainty" framing for consistency with terminology used in the uncertainty visualization literature. However, in

the experiment, the uncertainty information was presented under a "certainty" framing (e.g., "the forecast is more certain for some locations than for others"), based on feedback from participants during extensive piloting of study materials.

Three UVs were constructed using the rate and uncertainty maps.

Adjacent UV. Uncertainty was presented in a separate map adjacent to the rate map (Figure 4.2). To display uncertainty, we used the Hue-Saturation-Lightness color model to select three perceptually uniform shades of gray, with constant hue and uniformly increasing lightness.

Transparency UV. The Transparency UV (Figure 4.3) was developed from the rate map, by changing the alpha levels of the rate colors based on their uncertainty level. Alpha is a graphical parameter that alters color lightness by fading to white at minimal alpha. Many previous UV evaluations for natural hazards have found this aspect of color to be an effective visual metaphor for uncertainty (Cheong et al., 2016; Retchless and Brewer, 2016). We chose the three alpha levels visually to maximize discriminability of the transparency levels across all colors. We included these in the legend by adding columns of colors with corresponding alpha for each uncertainty level.

Bounds UV. To create the Bounds UV (Figure 4.4), we developed two maps: one for the lower bound (most likely forecast minus 2 $\times$ standard deviation) and one for the upper bound (most likely forecast plus 2 $\times$ standard deviation) of a 95% confidence interval for each location's forecasted aftershock rate. While this assumes the forecast distribution is symmetric rather than skewed, it was the most reasonable approximation available to us without a complete forecast distribution to draw percentiles over. The lower and upper bound maps were labelled "Optimistic Forecast: Lowest Number of Aftershocks Reasonably Likely" and "Pessimistic Forecast: Highest Number of Aftershocks Reasonably Likely", respectively.

In this visualization, uncertainty is not conveyed explicitly but can be inferred from the color differences between the lower and upper bounds maps (i.e., a large color difference at a given location represents greater uncertainty about its forecast). Participants were informed of this interpretation of large color differences in a short tutorial preceding the study. A location with high uncertainty corresponded to a color difference of at least 4-5 colors between the lower and upper bound maps, regardless of its rate level in the most likely forecast. That is, for areas of high uncertainty, the pessimistic forecasted rate was much higher than the optimistic forecasted rate. The exception was in areas where both the most likely forecast and uncertainty is high (see caption to Figure 4.4).

The lower and upper bound maps used the same categorical color scheme as the rate map; however, they are calculated with continuous rather than categorical uncertainty values, meaning that they show more continuous variation than the uncertainty maps provided in the Adjacent and Transparency UVs, where uncertainty is provided with three categories. We considered maps using categorical uncertainty values (i.e., the average uncertainty within each uncertainty category); however, this created abruptly decaying zones of high or low forecasted rates. The ETAS model and the seismological laws it is based on ((Ogata, 1998), see Chapter 3, Section 3.1.1) would yield forecasts that decay spatially in a smooth way. Thus, lower and upper bounds maps using continuous uncertainties can be argued to be more externally valid for such a forecast, compared to those using categorical uncertainties.

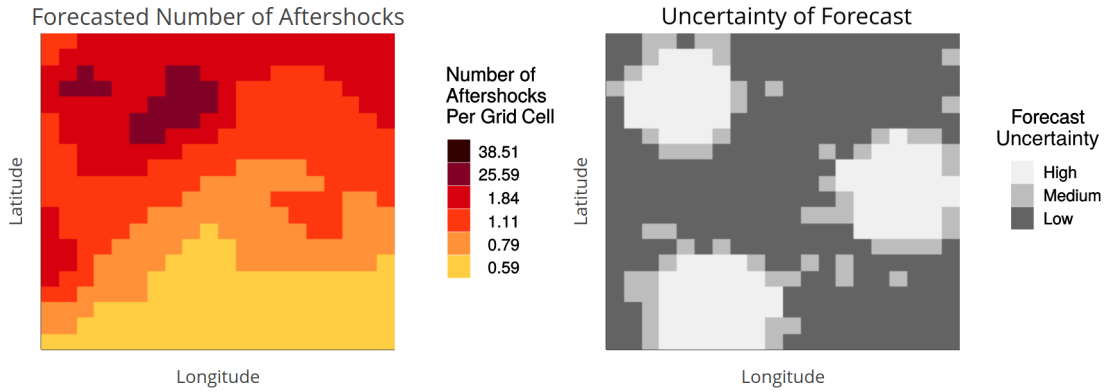

Figure 4.2: Adjacent UV: Most likely forecasted aftershock rate map next to a map of its model uncertainty. Figures 4.2-4.4 show UVs for one of the two forecast regions used in the experiment. Since we adopted a "certainty" framing in the experiment, the color palette for the right map (showing certainty rather than uncertainty) proceeded from light to dark for areas of lower to higher certainty, matching the color direction of the rate map. (See experiment screenshots in Figures C.1-C.3.)

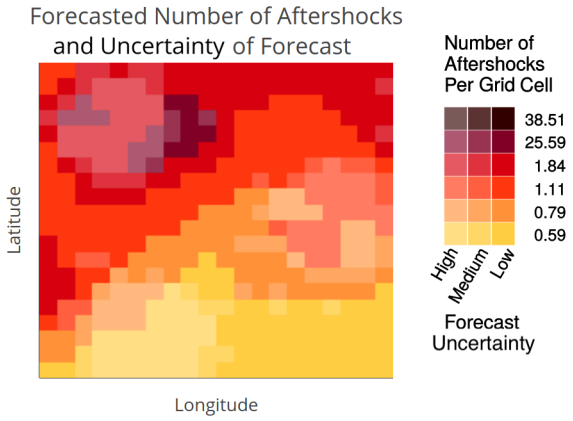

Figure 4.3: Transparency UV: Most likely forecast map in color and uncertainty levels shown by transparency (alpha level) of rate color.

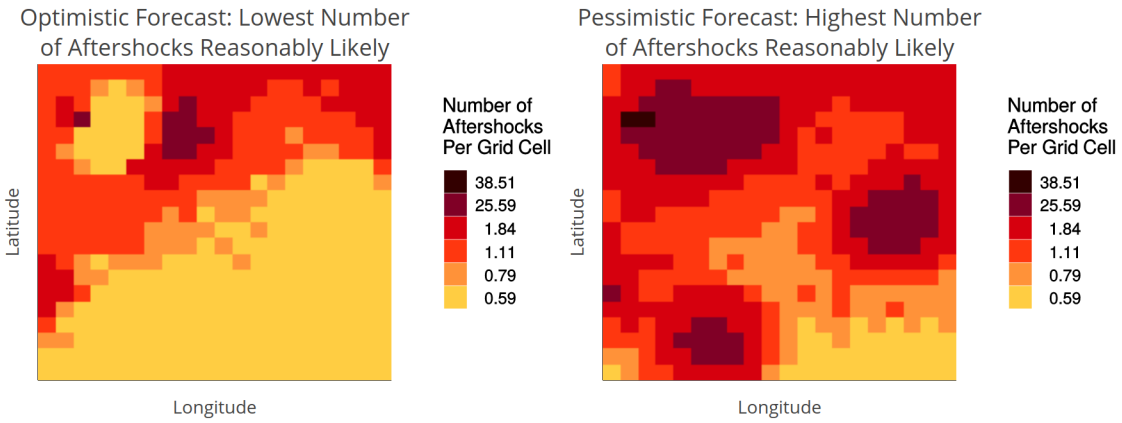

Figure 4.4: Bounds UV: Lower and upper bound maps of a 95% confidence interval around the most likely forecast. The forecast uncertainty at each location is shown through its difference in colors between the maps. Locations with high uncertainty have a 4-5 color difference, regardless of their most likely forecasted rate (e.g., lower left and middle right zones). We designed another scenario for when both the most likely forecasted rate and forecast uncertainty are high: the lower and upper bounds showed dark red and brown, respectively (e.g., several grid cells in the upper left zone). This scenario represents a case where due to the high uncertainty, the forecast's lower bound is still very high but its upper bound is extremely high.

4.2.3 Design and procedure

Following consent, participants were randomly assigned to one of the four visualization conditions in a between-participants design (Rate Only: n=217; Adjacent: n=221; Transparency: n=221; and Bounds: n=225). Participants were presented with some basic introductory information about aftershocks, how they are forecast and why forecasts are uncertain. In particular, participants were informed that, in areas where forecast uncertainty is high, many more aftershocks than forecasted could actually occur. This aligns with the skewed nature of the distribution of aftershock rates for spatial grid cells. To ensure they understood this introduction, participants had to correctly answer four multiple choice questions about the information provided. One of these questions related to this interpretation of high uncertainty.

Participants then received a visual tutorial explaining how to read the visualization that they were randomly assigned to. We explained the maps as displaying a forecast of how aftershocks will be distributed across a region in the next week following a major earthquake. We used arrows and highlighting to sequentially introduce the elements of the visualization (rate levels, uncertainty levels, legend), in a standardized way across all conditions.

Following this introduction, participants completed three tasks (see Table 4.1). Two map-reading tasks evaluated how well participants could retrieve and integrate the rate and uncertainty levels from the map(s). A comparative judgment task then tested how participants utilized the depicted forecast, together with its uncertainty, to make judgments about future aftershock occurrence.

Within each participant, we randomized forecast region across the two map-reading tasks to counter-balance any potential effects of forecast region on reading the visualizations. For the comparative judgment task, forecast region was randomized across trials within-participants. To isolate the effect of the rate level on judgment, we repeated trials for low, medium and high rate levels (across both forecast regions), which we set at 0.59, 1.11

and 25.59 aftershocks/grid cell in the most likely forecast map, respectively. To restrict the number of trials, we only focused on low and high uncertainty levels. The rate and uncertainty levels were also randomized across trials within-participants.

Task	Instruction	Rate/Uncertainty Levels for Trials
1. Read off (variables)	For a particular location on the map, return its rate or uncertainty level.	**Rate**: low (0.59 aftershocks/grid cell), medium (1.11), and high (25.59) rates, with low uncertainty only (3 trials) **Uncertainty**: low, medium, and high rates, with low and high uncertainty (6 trials)
2. Read between (variables)	Mark a location on the map with a particular rate and uncertainty level.	Low, medium, and high rates, with low and high uncertainty (6 trials)
3. Comparative judgment	Given the aftershock rate and uncertainty level, in which marked location will there be more aftershocks in the next week?	Depending on judgment type (see Table 4.2), locations varied forecast or uncertainty systematically (22 trials)

Table 4.1: Tasks to evaluate effects of UVs on map-reading (Tasks 1 and 2, for Research Question 1) and judgment (Task 3, for Research Question 2) with aftershock forecasts.

Read Off task. The first map-reading task required participants to provide the rate level or uncertainty level of a marked location (grid cell). For the Bounds UV, we asked participants to provide the rate level for the marked location on the upper bound map. Each participant

provided the rate level for three trials (low, medium, high rate for low-uncertainty locations) and uncertainty levels for six trials (full factorial, see Table 4.1). Participants in the Rate Only condition were not asked to provide the uncertainty level, as this was not depicted in their map. To increase difficulty, we used locations that bordered other rate or uncertainty levels.

Participants answered using multiple choice response options and responses were scored against the correct answer. Each participant's accuracy (calculated separately for the three rate and six uncertainty level responses) was averaged within visualization condition.

Read Between task. The second map-reading task required participants to integrate over both rate and uncertainty variables. In this task, participants were asked to click on a location that matched specific rate and uncertainty combinations. Participants in the Bounds condition were asked for a specific lower bound[2]/uncertainty combination, and participants in the Rate Only condition were asked only for a specific rate level. We asked for 6 locations (full factorial, see Table 4.1) and switched the forecast region from the one randomly assigned for the Read Off task.

We scored participants' responses on whether the clicked location's rate and uncertainty levels matched what was requested. We again averaged participant accuracy within visualization condition, separately for rate and uncertainty levels.

Comparative Judgment task. In the final experimental task, two marked locations were shown and participants were asked to select in which one they would expect more aftershocks to occur in the following week. Specifically, we asked: "Where will there be more aftershocks next week: Location 1 or Location 2? Please make a prediction." We further asked participants to rate their confidence in each judgment, using a six point scale

[2]Since there were no high and medium lower bound locations that also had high uncertainty, we asked for those two trials on the upper bound map.

with equidistantly spaced verbal labels: “completely guessing”; “mostly guessing”; “more guessing than sure”; “more sure than guessing”; “mostly sure”; “completely sure”(coded as 1-6, respectively).

We varied the rate and uncertainty levels of locations across trials to evaluate three distinct types of judgment. *Baseline trials* assessed how map features determined judgments. In these trials, participants chose between two low-uncertainty locations with identical rate levels (see Figure C.1 in the Appendix for an experiment screenshot of an example baseline trial). While we do not analyze these to answer our research questions, they provide a control against which to understand the next trials.

Surprise trials tested how a change in uncertainty led to a change in judgment. For these trials, we moved one of the baseline trial locations from low to high uncertainty, but kept the same rate (Figure C.2). Thus, each baseline trial yielded two distinct repetitions for each surprise trial (see Table 4.2), where one of the baseline locations remained constant and the other moved to a high-uncertainty zone. Lastly, *sure bet trials* were used to see how selection changes when rate level changes, but uncertainty levels were held constant. Participants compared two locations with low uncertainty, where the rate is higher for one and lower for the other (Figure C.3). Thus sure bet and surprise judgments correspond directly to communication needs 1 and 2, respectively (see Table 4.2).

Each trial type had multiple trials, where we varied the locations’ rate levels to assess how this affects the perception of its uncertainty (e.g., surprise trials 2 and 5 in Table 4.2). This also allowed multiple trials within a rate level (e.g., surprise trials 2 and 3 in Table 4.2, with low rates). We also repeated each trial in Table 4.2 across both forecast regions to manage expected within-participant variability in task response. There were thus (3×2) baseline trial repetitions, ($3 \times 2 \times 2$) surprise trial repetitions and (2×2) sure bet trial repetitions. These 22 total trial repetitions were randomized within-participant and the location labels (1 and 2) were also assigned randomly (though recoded in our analysis as described in Table

Trial	Type	Location 1		Location 2	
		Rate	Uncertainty	Rate	Uncertainty
1	Baseline	Low*	Low*	Low$^+$	Low$^+$
2	Surprise	Low	High	Low$^+$	Low$^+$
3	Surprise	Low*	Low*	Low	High
4	Baseline	Medium*	Low*	Medium$^+$	Low$^+$
5	Surprise	Medium	High	Medium$^+$	Low$^+$
6	Surprise	Medium*	Low*	Medium	High
7	Baseline	High*	Low*	High$^+$	Low$^+$
8	Surprise	High	High	High$^+$	Low$^+$
9	Surprise	High*	Low*	High	High
10	Sure Bet	High	Low	Medium	Low
11	Sure Bet	Medium	Low	Low	Low

Table 4.2: Trials in judgment task by type. Two trial types correspond to UV communication goals. Participants had to judge which of two locations will have more aftershocks in the next week, given the forecast map(s). For each rate level, locations marked with a star/plus sign were identical between the baseline and surprise trials, meaning that only the other location was moved to create a surprise trial.

4.2).

For baseline and surprise judgments, we attempted to select location pairs with balanced distances to, and symmetry around, the map center and zones of high uncertainty or high rate. For their trial repetitions, we sought to keep distance between location pairs constant while covering different parts of the map. For sure bet judgments, we again selected locations bordering other rate/uncertainty levels to increase difficulty.

Response time and covariates Since it may be faster to read and use forecast and uncertainty information from particular designs, previous UV evaluations have measured participant response times (Kinkeldey et al., 2017). To explore UV effects in response time, we recorded response times for each trial across all tasks, or the elapsed time between when a participant sees the trial and when they select their response. We compare the response times of participants in the UV conditions relative to those in Rate Only condition.

Following the judgment task, participants were asked several demographic questions (age, gender, education level, state of residence). We also asked participants how many earthquakes they previously experienced, as past experience has been shown to affect earthquake forecast perception (Becker et al., 2019).

4.2.4 Methods of statistical analysis

We specified a set of confirmatory statistical analyses aimed at answering our research questions, prior to conducting the experiment. We also specified a set of exploratory statistical analyses to complement the confirmatory analyses. These were aimed at investigating explanations for potential UV effects to generate hypotheses for future experiments. Both sets of analyses were pre-registered prior to conducting the experiment.

Confirmatory analysis

We performed confirmatory analysis using confidence intervals to answer our research questions about how UVs affected map-reading accuracy and comparative judgment. In the Read Off task, we aggregated average accuracy in responses about locations' rate levels within each condition. We calculated the 95% *simultaneous* confidence interval for the mean difference between each UV condition and Rate Only.

We use simultaneous inference to examine the relationship between UVs and experimental outcomes because we are simultaneously testing three hypotheses (comparing the three UV

groups to Rate Only). In order to maintain a 95% confidence level (or an α=5% chance of a Type I error) across the family of comparisons, we must thus adjust the familywise confidence level used. We used the Bonferonni correction and divided α by the number of comparisons. There is no single definition for a family of hypotheses and it has been described as "an individual experiment of a single researcher" (Miller, 1981). Since we randomized locations and forecast regions across the trials of each task, we assumed independence between the map-reading tasks and the three sets of trials in the judgment task. Thus, we took the number of comparisons to be three (the number of UV groups being compared to the Rate Only group).

We then compared confidence intervals to zero to infer differences between groups. We conducted the same confidence interval analysis for responses about locations' uncertainty levels, but comparing across the three UV conditions only. We repeated this confirmatory analysis on rate and uncertainty accuracies in the Read Between task. Across all analyses, we use a 5% significance level to determine statistically significant differences between conditions.

We evaluated comparative judgments by calculating the percentage of trials where the participant selects the location with the higher rate or uncertainty (see Table 4.2), which we then averaged across participants and conditions. Differences between these percentages for UV and Rate Only conditions were computed separately for sure bet and surprise judgments. We again inferred differences between groups using 95% simulataneous confidence intervals. For surprise judgments, we computed differences between UVs and Rate Only specifically for each rate level. This tests whether UV effects on judgments of high uncertainty locations are consistent when the forecasted rate is low, medium or high.

Exploratory analysis

As an exploratory analysis, we explored whether individual differences affected responses to forecast uncertainty.

We expected participant heterogeneity in the comparative judgment task, irrespective of visualization condition or the rate level. We did not experimentally control for the myriad sources of this heterogeneity, nor did we have any *a priori* expectations of which sources will affect judgment, so we omitted participant-level effects from our main confirmatory analysis. We did control several features of each judgment trial (e.g., the forecast region and location characteristics), but again, did not have any *a priori* expectations of how these may affect judgment.

We built multilevel models of participant judgments to explore how they were affected by these two sets of variables and by overall heterogeneity by participant. We then performed a model selection analysis to identify drivers of task response when accounting for individual differences. We investigated which participant- or trial-level variables led to best model fit (across all participants and trials).

We first built a baseline multilevel regression model on Y_{ij} (participant i's judgment in trial j), with a fixed effect for condition, rate level, and their interaction as well as a random participant-level intercept. We used a treatment contrast for the visualization condition, with reference category being Rate Only. The baseline multilevel model (built using the glmer R package, version 1.1-27) is:

$$
\begin{aligned}
logit(P(Y_{ij} = 0)) = \beta_0 + \zeta_{0i} + \\
&\beta_b I(UV_i = Bounds) + \beta_t I(UV_i = Transparency) + \beta_a I(UV_i = Adjacent) + \\
&\beta_L I(Rate_j = Low) + \beta_H I(Rate_j = High) + \\
&\beta_{bL} I(UV_i = Bounds) * I(Rate_j = Low) + \\
&\beta_{bH} I(UV_i = Bounds) * I(Rate_j = High) + \\
&\beta_{tL} I(UV_i = Transparency) * I(Rate_j = Low) + \\
&\beta_{tH} I(UV_i = Transparency) * I(Rate_j = High) + \\
&\beta_{aL} I(UV_i = Adjacent) * I(Rate_j = Low) + \\
&\beta_{aH} I(UV_i = Adjacent) * I(Rate_j = High), \\
\zeta_{0i} \sim N(0, \sigma_\zeta).
\end{aligned}
$$

We considered the estimated fixed effect coefficients and compared them with the confirmatory analysis. We then built models with additional fixed effects (trial-level variables, participant-level covariates) and compared them to the baseline, using model performance metrics. We only considered multilevel models with random intercepts only (as a random "slope" on UV condition would not be identifiable with the data). Thus, we simply added fixed effects stepwise to the baseline model and assessed if this produced a better model fit, using multiple metrics.

A common model performance metric is the modified Bayesian Information Criterion (mBIC), which balances model goodness of fit with parsimony (Müller et al., 2013). For a multilevel model where f is the number of fixed effects and r is the number of random effects estimated, this is defined as

$$mBIC = -2LL + log(n_t) \cdot (f + r),$$

where LL is the model log-likelihood and n_t is the number of trials for which the model is built.

Since this is a classification (binary response) model, we also used common metrics for classification. For all of these scores, we calculated the predicted probability of selecting the location of higher uncertainty, p_{ij}, for a given participant and trial, based on the given model. We binarized these at p_{ij}=0.5, obtaining binary predicted judgments b_{ij}. These were then compared to the observed selections o_{ij} with several goodness-of-fit metrics. The Brier score (Brier et al., 1950) is a standard classification metric for a set of predicted probabilities and observations. $Brier(p_{ij}, o_{ij})$ is defined as

$$Brier(p_{ij}, o_{ij}) = \frac{1}{n_t n_p} \sum_{i=1}^{n_p} \sum_{j=1}^{n_t} (p_{ij} - o_{ij})^2,$$

where n_p is the number of participants for which the model is built.

The correct classification rate (CCR) for a set of binary predictions and observations is another standard classification metric, defined as

$$CCR(b_{ij}, o_{ij}) = \frac{1}{n_t n_p} \sum_{i=1}^{n_p} \sum_{j=1}^{n_t} I(b_{ij} == o_{ij}).$$

Finally, we consider the area under the Receiver Operating Characteristic curve. Consider that we can binarize the predicted probability at any value p^*, not necessarily 0.5. Let $TPR(p^*)$ (true positive rate) be the proportion of binarized judgments at p^* which correctly match a judgment of 1 (so $o_{ij} = 1$ and $b_{ij} = 1$). Conversely, let the $FPR(p^*)$ (false positive rate) be the proportion of binarized judgments at p^* which misclassify a judgment of 0 (so $o_{ij} = 0$ and $b_{ij} = 1$). The Receiver Operating Characteristic curve is the curve of $FPR(p^*)$ vs. $TPR(p^*)$ for all possible values of p^* and the area under this curve (AUC) also measures

model performance. The AUC for the best possible classifier should be as close to 1 as possible, as this would maximize the TPR and minimize the FPR.

This model selection analysis solely focused on identifying other variables beyond those in the baseline model that explain the experimental data based on in-sample goodness of fit. Significant effects in the best-fitting model could be considered as key determinants of task response, and built into future experimental designs. We were specifically interested in how participant covariates and characteristics of the locations used for judgment trials (distance from map center and from high rate or uncertainty zones) influenced judgment pattern.

Finally, ordinal confidence ratings were analyzed in exploratory fashion across sure bet and surprise judgments and other trial subsets, comparing each UV group to the Rate Only baseline. We also compared how patterns in response times differ by condition and between judgment and map-reading tasks.

4.3 Results

4.3.1 Map-reading tasks

Read Off task Accuracy in reading rate levels was high across all conditions in both map-reading tasks; however, accuracy differed across conditions in reading the uncertainty level (see Table 4.3 and Figure 4.5).

In the Read Off task, participants had a high accuracy in reading off the rate in all conditions, giving correct responses in between 88.8%-92.8% of trials. Accuracy for participants in the UV conditions was thus statistically indistinguishable from Rate Only, with differences of less than 2.0 percentage points, on average (see Table 4.3).

Accuracy in reading off the uncertainty level was lower in the Transparency and Bounds conditions, relative to the Adjacent condition. Participants gave correct uncertainty responses in 82.7%-94.9% of trials, across all conditions. Compared to the Adjacent condition, which was most accurate, participants were on average 6.8 percentage points less accurate

	Read Off				Read Between			
	Rate		Uncertainty		Rate		Uncertainty	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Rate Only	90.8%	(17.8)	-		97.7%	(8.0)	-	
Adjacent	92.6%	(19.0)	94.9%	(15.3)	94.7%	(12.3)	95.7%	(13.2)
Transparency	92.8%	(17.3)	88.1%	(17.9)	93.8%	(13.1)	87.4%	(14.6)
Bounds	88.8%	(19.7)	82.7%	(29.1)	94.0%	(13.0)	70.0%	(19.0)

Table 4.3: Percentage of trials answered correctly by condition for both map-reading tasks, separately for rate and uncertainty levels. Participants in the Rate Only condition were not asked to read uncertainty in either task.

(95% confidence interval (CI) = 3.4-10.2) in the Transparency condition, and 12.2 percentage points less accurate (95% CI = 7.5-16.9) in the Bounds condition. Across all conditions, the majority of inaccurate responses came from misreading the location's uncertainty level as medium, rather than high or low. This was the case for 57.6% to 80.9% of inaccurate responses within the three UV conditions.

Read Between task Results were similar for the Read Between task. Participants identified the correct rate across all conditions. Participants in the Rate Only condition were most accurate for the rate level, and accuracy in UV conditions was no more than 3.9 percentage points lower, on average.

Just as in the Read Off task, fewer participants identified the correct uncertainty levels in the Transparency and Bounds conditions than in the Adjacent condition. There was a wider range in accuracy across conditions for the Read Between task, with participants giving correct uncertainty responses in 69.8%-95.6% of trials. Compared to the Adjacent

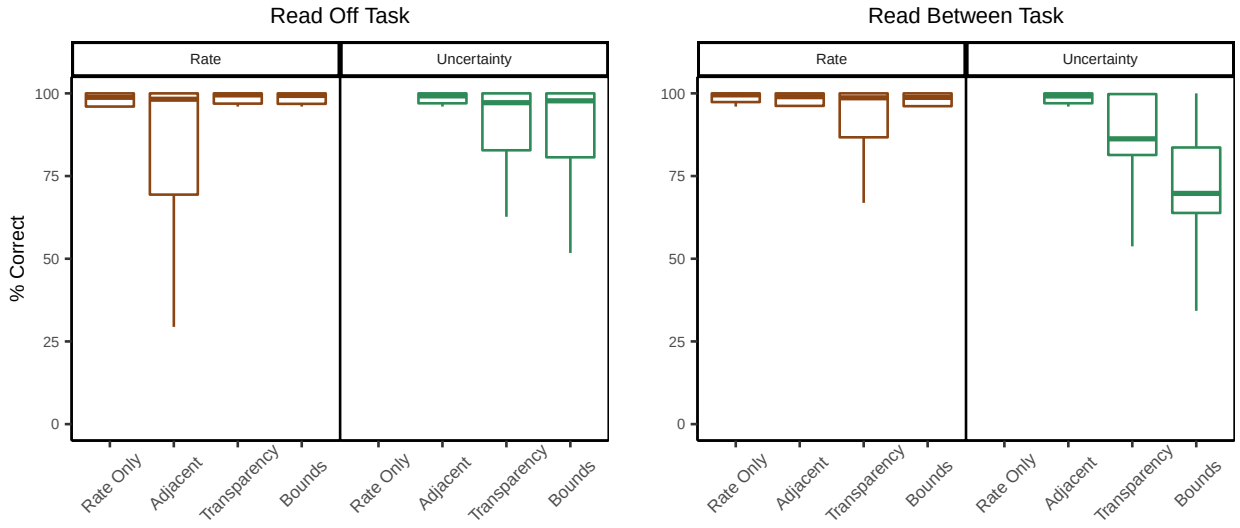

Figure 4.5: Accuracy in the Read Off (left) and Read Between (right) tasks. Boxplots show the percentage of trials answered correctly by participants, grouped by condition. We show results for reading rate and uncertainty separately for both tasks, slightly jittered to reveal trends. Participants in the Rate Only condition were not asked to read uncertainty in either task.

condition, participants were 8.3 percentage points less accurate (95% CI = 5.5-11.1) in the Transparency condition, and 25.7 percentage points less accurate in the Bounds condition (95% CI = 22.4-29.0). The majority of inaccurate responses were again caused by providing a location with medium uncertainty, across all conditions.

4.3.2 Comparative judgment task

In the comparative judgment task, participants compared two locations that differed systematically in either rate or uncertainty levels and judged where they would expect more aftershocks. For locations with *different rates* but *low uncertainty* (sure bet trials), almost

all participants in all conditions correctly selected the higher-rate locations (participants selected the higher-rate location in 92.2%-97.8% of trials). Yet for locations with *differing uncertainty levels* but *identical rates* (surprise trials), participants in the Bounds condition were more likely than participants in the other two UV conditions to select the higher-uncertainty location (Table 4.4 and Figure 4.6).

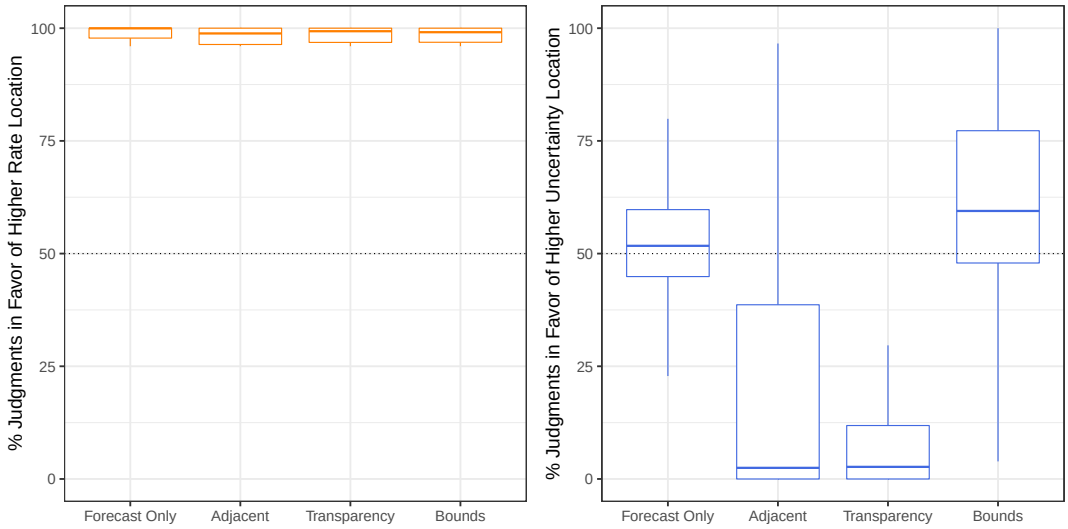

Figure 4.6: Proportion of trials where the location of higher rate (Sure Bet trials, orange) or higher uncertainty (Surprise trials, blue) was selected in the comparative judgment task. Boxplots are over mean proportion across participants by condition, slightly jittered to reveal trends.

In surprise trials, participants compared two locations with different uncertainty but identical rate levels. In the Rate Only condition, participants selected the location of higher uncertainty in 52.0% of the trials, on average (see Figure 4.6). Note that uncertainty was not visualized in this condition, so participants saw two locations of equal rate level but may nevertheless infer differences between them based on other information (Morss et al., 2010). Participants in the Bounds condition selected the higher-uncertainty location 6.9

	Sure Bets		Surprises	
	Mean	(SD)	Mean	(SD)
Rate Only	97.8%	(10.9)	52.0%	(11.7)
Adjacent	92.2%	(18.1)	22.8%	(36.3)
Transparency	93.4%	(17.4)	12.8%	(24.6)
Bounds	95.0%	(14.0)	58.9%	(22.8)

Table 4.4: Comparative judgment task: Percentage of trials where participants judged the higher-rate location (sure bet trials) or higher-uncertainty location (surprise trials) to have more aftershocks, per condition.

percentage points more often (95% CI = 3.2-10.5). In contrast, participants in the Adjacent and Transparency condition selected this location 29.3 (95% CI = 23.8-34.7) and 39.2 (95% CI = 35.3-43.1) percentage points less than Rate Only, respectively.

Moreover, 68.0% of the participants in the Adjacent and 75.7% in the Transparency condition selected the lower-uncertainty location in at least 11 of 12 surprise trials. In contrast, participants' judgments in the Bounds and Rate Only conditions were more variable, with at least 89.0% of these participants selecting both locations multiple times across the 12 trials.

Looking at the surprise trials separately for each of the three rate levels (low, medium, high), the difference between Bounds and the other UV conditions was found for each rate level, but was considerably stronger for trials with low rates and varying uncertainty (see Figure 4.7). Participants in the Rate Only condition selected the higher-uncertainty location in 85.9% of the low rate trials and those in the Bounds condition selected it 1.7 percentage points more often, on average (95% CI = (-3.3)-7.2). In contrast, participants in the Adjacent and Transparency conditions selected the higher-uncertainty location 59.2 percentage points (95% CI = 52.6-65.9) and 72.1 percentage points (95% CI = 66.7-77.5) less often, respectively.

The surprise trials with high rates and varying uncertainty contained cases where the higher-uncertainty location had a possible extreme outcome (high lower bound and extremely high upper bound), which was only visible in the Bounds condition (see caption of Figure 4.4). Participants using Bounds tended to select this higher-uncertainty location for extreme cases (at least 79.0% of participants selected it in these trials) but not for the other high rate trials (7.0% or fewer selected it in these trials), resulting in their average percentage for all high rate trials to be near 50% (Figure 4.7, right).

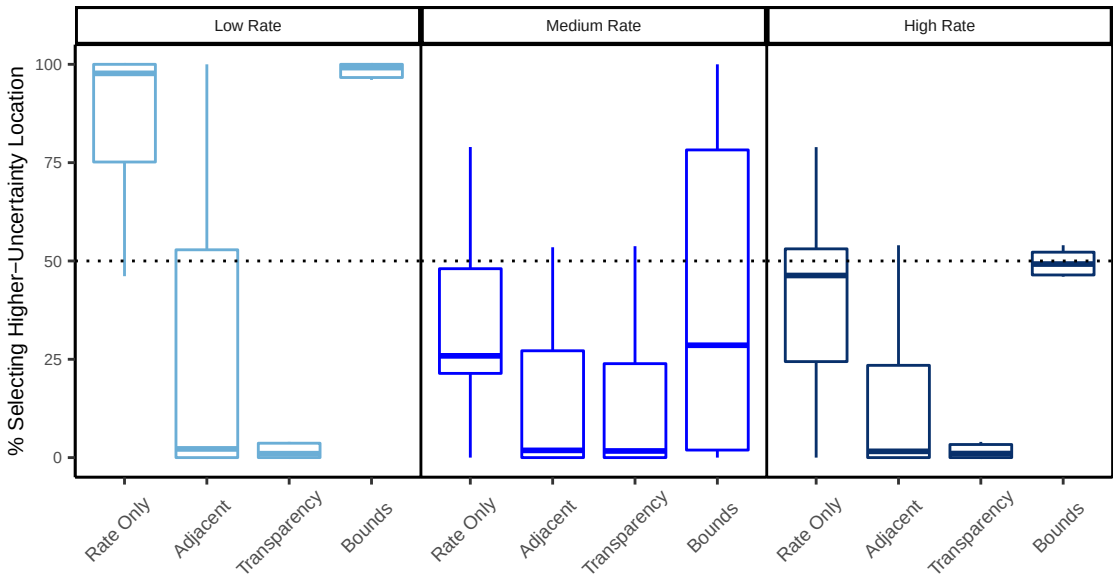

Figure 4.7: Proportion of trials where the higher-uncertainty location was selected between two locations of equal rate level (surprise trials), grouped by condition and rate level.

4.3.3 Multilevel modeling of judgments

Accounting for variability within participants The effect of visualization on judgments also held after we accounted for variation in judgments on an individual level. We

built a multilevel logistic model for judgments in situations with varying uncertainty and identical rates (surprise trials), with fixed effects for visualization condition, rate level and their interaction, and a random intercept per participant. Because the confirmatory analysis found no evidence of an effect by condition for sure bet trials, we report exploratory analysis on models for judgments in surprise trials only.[3]

While using the Bounds UV slightly raised the probability of selecting the higher-uncertainty location compared to Rate Only, the opposite was true for Transparency and Adjacent. These conditions had significantly negative estimated coefficients (see Table C.1 in the Appendix), suggesting, as in Figure 4.6, that participants in these conditions were much less likely to select the higher-uncertainty location. This location was more likely to be selected for trials with low rate over medium rate (highly positive estimated coefficient) and high rate, but this effect was dampened by the interaction terms of low rate with Adjacent and Transparency visualizations (highly negative estimated coefficients). These results match those from the confirmatory analyses in the previous section.

Effect of trial characteristics on judgments Next, we explored whether the spatial patterns of the forecast region and characteristics of the trial locations, such as their distances to important map features, also influenced forecast judgment. These trial-level features did have a measurable effect on participants' judgments, better explaining the variation in judgments than visualization and rate level alone.

To assess the importance of these variables, we included the forecast region of each trial, the locations' rate level and their distances to the map center and zones of high uncertainty or rate as fixed effects into the model. We performed a stepwise model comparison between the baseline model and models with these additional fixed effects (see Section C.2 in the Appendix for details on the distance measures and Table C.2 for model selection results).

[3]We also built corresponding models using both judgment trials with an interaction term for the trial type, and results were qualitatively similar to those for surprise trials only.

We again report results for surprise trials only, as results for models including both trial types were qualitatively similar. The best-fitting model across all four metrics included rate level and visualization, as well as having a location which was closer to the map center or to a high-rate zone (Table 4.5).

Fixed Effect	Estimated Coefficient	95% CI	
Intercept	-2.56	[-2.95,	-2.16]
Visualization			
Adjacent	-1.65	[-2.18,	-1.11]
Transparency	-2.01	[-2.54,	-1.48]
Bounds	0.68	[0.22,	1.13]
Rate Level			
Low	4.77	[4.39,	5.15]
High	1.12	[0.87,	1.38]
*Visualization*Rate Level*			
Adjacent*Low	-2.51	[-2.95,	-2.06]
Adjacent*High	-0.92	[-1.36,	-0.49]
Transparency*Low	-3.37	[-3.82,	-2.91]
Transparency*High	-1.61	[-2.07,	-1.16]
Bounds*Low	0.03	[-0.39,	0.45]
Bounds*High	-0.61	[-0.94,	-0.29]
Which Location Closer to Center			
Higher-Uncertainty Location Closer	-0.05	[-0.26,	0.17]
Lower-Uncertainty Location Closer	1.89	[1.68,	2.10]
Which Location Closer to High Rate Zone			
Higher-Uncertainty Location Closer	-0.17	[-0.43,	0.09]
Lower-Uncertainty Location Closer	1.01	[0.71,	1.31]
Random effects			
Intercept σ^2	4.37		

Table 4.5: Most likely estimates and 95% confidence intervals using Wald standard errors for fixed effects in best-fitting multilevel model. The intercept is the logistic of the probability of selecting the higher-uncertainty location for the Rate Only condition and medium-rate trials, with both locations being equidistant from map features (reference level for visualization, rate level and the map features variables). Each fixed effect gives the change in probability of selecting this location from Rate Only to UV conditions, from medium rate to other rate levels, or from equidistant locations to one location being closer to the map feature (all else being held equal).

The UV effects on judgment did not change even when considering the other important variables in this optimal model. The estimated coefficients for visualizations were similar to the baseline model (compare Table C.1 to Table 4.5). Similarly, the coefficients for rate level and its interaction with UV group were equivalent to the baseline model, reflecting the findings in Figure 4.7. The best-fitting model also found that in trials where the lower-uncertainty location was closer to the center or to a high-rate zone, participants tended to select the higher-uncertainty location (highly positive estimated coefficients), though these variables only minorly improved model performance (see Table C.2 in the Appendix).

Participants also seemed to use features of the map to make consistent judgments between two locations of the same rate and uncertainty levels (baseline trials). Figure 4.8 shows that even though these locations have the same characteristics, participants tended to select one or the other for some trials, sometimes differently across conditions. Selection in these trials had no consistent pattern based on forecast region or locations' distances to map features.

Individual differences between participants' judgments To explore whether the participant-level variables we measured could account for differences in surprise judgments, we considered additional fixed effects for the number of earthquakes participants previously experienced, education, age, gender, and state of residence, performing the same model comparison as described above. None of these individual differences improved the model fit across multiple metrics (see model comparison in Table C.3 in the Appendix). Thus, differences in visualization and rate level explain participants' judgments better than any of these participant-level variables.

We also found no systematic differences on these variables between participants who consistently selected either location in 11 out of 12 trials and those with more variation in judgments.

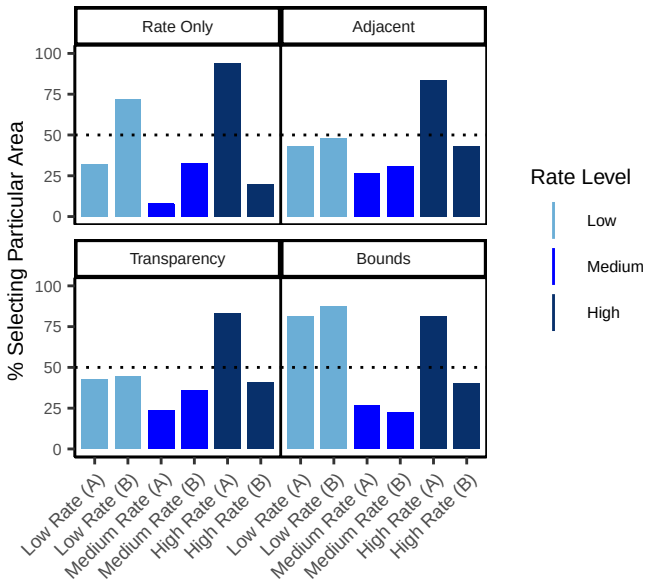

Figure 4.8: Percentage of participants selecting a particular location in the baseline trials, where both locations have the same rate and same low uncertainty. This is Location 1 in the baseline trials described in Table 4.2. Trials were repeated for forecast regions A and B.

4.3.4 Confidence and response time

Confidence ratings We asked participants to rate their confidence after each judgment and we investigated these by visualization condition and between sure bets and surprise judgments.

We calculated median confidence ratings for each participant within each trial type. Confidence did not differ by condition for the sure bet judgments, with identical medians and equivalent distributions. For the surprise judgments, participants in the Rate Only condition (who did not see the difference in uncertainty between the two locations) were generally less confident than participants in the three UV conditions. The median confidence rating was 2 points lower for Rate Only compared to all UVs.

We also investigated whether surprise judgments were made more confidently in favor of higher- or lower-uncertainty locations, as suggested in Hullman et al. (2018). Again, we calculate median confidence ratings per participant for those trial subsets. Participants using Rate Only and Bounds had similar confidence regardless of their judgment; however, confidence appeared to differ by judgment for participants using Adjacent and Transparency (see Figure 4.9). When they selected the higher-uncertainty location, participants in these two conditions had median confidence ratings that were at least 1 point lower than when they selected the lower-uncertainty location. The spread of their confidence ratings was also lower (shorter boxplots) in those trials where they selected the lower-uncertainty location.

Response times We also investigated trial response times by calculating medians within participant across trials, separately for each of the three tasks. There was no meaningful difference across the UV conditions in the Comparative Judgment and Read Off tasks, with differences in median response times to Rate Only of less than 1.5 sec, across all trial types (figure omitted). In the Read Between task, response times were slightly shorter for participants in the Rate Only condition (who were only asked to identify a particular rate level)

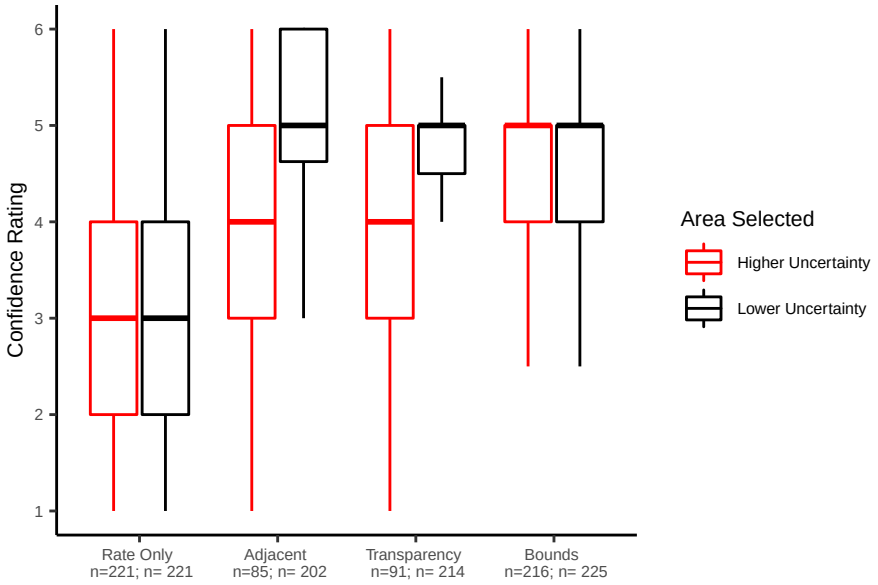

Figure 4.9: Boxplots for surprise judgment confidence ratings (median confidence by participant) by condition, binned between which location was selected. Sample sizes are given for each condition, the number of participants in that condition that selected the given location at least once.

than participants in the UV conditions (who were asked to identify a particular rate and uncertainty level).

4.4 Discussion

Uncertainty visualization is critical for forecast communication, as studies have shown that communicating uncertainty improves user decision-making and prevents users from building their own representation of where uncertainty is high or low (Joslyn and LeClerc, 2012; Ash et al., 2014). We compared three visualizations of uncertainty for aftershock forecast maps against a visualization of the forecast without uncertainty. The experimental tasks were designed to systematically evaluate how well users could read off the forecasts and how effectively the visualizations served two user-generated communication needs. In particular, aftershock forecast maps should communicate where the certainty is high that aftershocks will or will not occur, as well as where outcomes worse than forecasted are possible, due to high uncertainty.

The results show that all three uncertainty visualizations led to correct judgments about where to expect more aftershocks when both locations had low uncertainty and a difference in forecasted aftershock rate. However, the uncertainty visualizations resulted in significantly different judgments when the locations had forecasts where the uncertainty varied. Although users of the visualization showing the bounds of a forecast interval (Bounds UV) could read off the uncertainty less accurately than the other visualizations, it was the only one where users demonstrated an understanding that forecasts with high uncertainty could have outcomes worse than forecasted.

4.4.1 Effects of visualization

While there was no difference in how well participants could read the rate information from the different visualizations, accuracy in reading the uncertainty information differed depend-

ing on the visualization participants saw. Consistent with previous research, the adjacent design was associated with greater accuracy compared to the transparency-based design (Retchless and Brewer, 2016) and the interval-based design (Nadav-Greenberg et al., 2008). These differences held across both map-reading tasks. The majority of inaccurate responses in these tasks, across all visualizations, were in misreading the uncertainty level as being medium rather than high or low. For the Transparency UV, participants may have had difficulty distinguishing between the non-transparent version of one color (low uncertainty) and the more transparent version of the next darker color (one transparency level lighter, i.e., medium uncertainty). For the Bounds UV, since uncertainty was not depicted explicitly in the legend, participants may have had trouble classifying a location into the correct uncertainty category. Consistent with past research, aftershock forecast users were most accurate reading uncertainty information off adjacent displays compared to those that used transparency or represented uncertainty implicitly through a forecast interval.

In the judgment task, there were differences by visualization in how participants judged higher-uncertainty locations compared to lower-uncertainty locations. Most users of the adjacent and transparency visualizations expected the lower-uncertainty location to have more aftershocks than the higher-uncertainty location. This judgment pattern was so consistent that the majority of these participants selected the lower-uncertainty location in at least 11 of the 12 trials. In contrast, users of the Bounds UV were more likely to expect that the higher-uncertainty location would have more aftershocks, relative to the Adjacent and Transparency UVs.

Users of the Bounds visualization might have had this different judgment pattern, compared to the other visualizations, as a result of Bounds explicitly displaying the extremes of the forecast distribution rather than the uncertainty; the uncertainty must be inferred by comparing the lower and upper bound maps. The word uncertainty was also not used in the legend in the Bounds visualization, and we instructed participants that locations with

higher uncertainty can only be identified by their large color differences between the two maps. In contrast, the other uncertainty visualizations explicitly depicted the uncertainty and required users to infer where high uncertainty could yield extreme outcomes.

There are three potential explanations for why explicitly depicting the extremes rather than the uncertainty could account for our results. First, consistent with the results of Nadav-Greenberg et al. (2008), the upper bound map may have served as an anchor point that biased users' perceptions towards higher values. Users of the Bounds visualization may have expected more aftershocks in higher-uncertainty locations because they focused on the worst case scenario map (the upper bound) and did not even pay attention to differences in color to infer the uncertainty.

Second, the explicit depiction of uncertainty in the adjacent and transparency designs may have led users to associate high uncertainty with low forecast quality. That is, despite the study instructions explaining that locations with high uncertainty could have worse outcomes due to the skewed distribution of aftershock rates (which participants had to answer correctly before continuing to the study), participants may have nevertheless interpreted lower-uncertainty locations as being more reliable. Furthermore, uncertainty may have been interpreted differently based on whether the visualization presented it with words or numbers. Consistent with previous research (e.g., Retchless and Brewer (2016)), we used an ordinal uncertainty scale with verbal labels ("Low", "Medium", "High") in the adjacent and transparency designs. In contrast, the interval-based design used numeric legends in both lower and upper bounds maps. Previous experiments in non-spatial uncertainty communication have found that verbal labels can lower perceptions of forecast reliability compared to numeric ranges (Van Der Bles et al., 2019), which might explain the difference in judgments between the uncertainty visualizations.

Third, differences in judgment patterns between the visualizations may have resulted from differences in how the uncertainty visualizations used color to signify uncertainty.

Higher-uncertainty areas are always marked by darker colors in the upper bound map than lower-uncertainty areas of the same rate level. In contrast, higher uncertainty is marked by lighter colors in both the adjacent and transparency visualizations. These differences in color lightness may have affected how participants interpreted higher uncertainty in the visualizations. Previous studies have reported a dark-is-more bias in how people interpret color scales (Silverman et al., 2016; Schloss et al., 2018). Users of the adjacent and transparency visualizations may have thus perceived the lighter-colored zones of high uncertainty as having lower potential of aftershocks.

4.4.2 Effects of rate level

While users of the Adjacent and Transparency visualizations had the same patterns of selecting the higher-uncertainty location regardless of its most likely aftershock rate (rate level), judgment differed by rate level for the Bounds visualization. When the compared locations had *low rate levels* (most likely aftershock rate = 0.59 aftershocks/grid cell; yellow color), the majority of participants using the Bounds visualization correctly expected more aftershocks at the higher-uncertainty location, which had the much higher upper bound. That is, in these comparisons, the lower bounds of both locations were of equal color (both yellow) but their upper bounds showed a near-maximal color difference (from yellow to red), indicating the higher-uncertainty location's potential for much higher rates.

When the compared locations had *medium rate levels* (most likely aftershock rate = 1.11 aftershocks/grid cell; middle orange color), users of the Bounds visualization varied between selecting the lower-uncertainty (showing a medium rate in both lower and upper bound maps) and higher-uncertainty (showing low rate in lower bound, high rate in upper bound map) locations. When the compared locations had *high rate levels* (most likely aftershock rate = 25.59 aftershocks/grid cell; second-darkest red color), these participants only selected the higher-uncertainty location when the upper bound had an extremely high rate (38.51

aftershocks/grid cell; brown color, see Figure 4.4). In these situations, the lower-uncertainty location always showed a high rate (red color) across both lower and upper bound maps, indicating that a high number of aftershocks is very likely to occur. When the upper bound for the higher-uncertainty location showed an extremely high rate (brown color), participants were likely to select it. However, when the upper bound of the higher-uncertainty location only showed a high rate (red color) and the lower bound showed a low rate (yellow color), users of the Bounds visualization almost never selected this higher-uncertainty location.

These results indicate that the Bounds visualization led to judgments that recognized the relationship between high uncertainty and the potential for outcomes that could be worse than forecasted. For judgments where the higher-uncertainty location was much higher in the upper bound than the lower-uncertainty location, the majority of users of the Bounds UV selected the higher-uncertainty location. Users of the other uncertainty visualizations, who did not see the extremes explicitly, made the opposite judgment. This difference between uncertainty visualizations held even when accounting for within-participant variability and other potential participant-level determinants of judgments. Thus, the estimated visualization effects were robust across the sampled population, with respect to the studied covariates.

If highlighting the potential for (even) higher aftershock rates in cases of high forecast uncertainty is critical for a decision at hand, then our results support displaying forecast uncertainty with maps showing forecast intervals. Where locations have the same low uncertainty, the higher-rate location may be interpreted to have more aftershock potential and where locations have the same rate level, the higher-uncertainty location may be interpreted to have more aftershock potential. In contrast, adjacent and transparency-based displays appear to lead to an opposite response to high uncertainty.

4.4.3 Effects of map features

Previous studies have found that risk perception about a location can be impacted by its distance to risky areas (Ash et al., 2014; Mulder et al., 2017). We also found some evidence indicating that map features of the trial locations influenced judgments using aftershock forecasts. The model selection analysis found that when a lower-uncertainty location was closer to the map's center or to a high rate zone, the higher-uncertainty location was slightly more likely to be selected, irregardless of visualization (see Table 4.5). Furthermore, participants did not make the same judgments when comparing locations of equal rate and uncertainty, with clear differences by visualization in some trials (see Figure 4.8). It is possible that other map-related features that we did not measure could account for these differences in judgments. Future research could explore systematically how judgments are affected by map features or location characteristics.

4.4.4 Judgment confidence and response time

We found only minor differences between visualizations in user confidence in judgments. Not surprisingly, confidence ratings were higher when comparing low-uncertainty locations with different rates, than when comparing locations with different uncertainties. In general, participants using the forecast depicted without uncertainty made judgments between locations with different uncertainities with lower confidence than those using UVs. This suggests that omitting uncertainty lowers confidence for judgments between two locations whose uncertainties differ but rates do not. We observed higher and less variable confidence ratings for judgments in favor of the lower-uncertainty location, compared to those for the higher-uncertainty location, but only for users of the adjacent and transparency visualizations. These designs appear to encourage a confident interpretation of uncertainty, leading lower-uncertainty locations to be confidently associated with more aftershocks. Response time distributions were equivalent between visualizations, meaning that no evidence was

found to suggest visualizing uncertainty increases the time needed for map-reading or making comparative judgments. This result mirrors findings summarized in a review of previous UV evaluations (Kinkeldey et al., 2017).

4.4.5 Limitations and future research

Our study sought to examine the effects of UV designs on specific communication needs for aftershock forecast maps. To do so systematically, we had to fix one variable, either the forecasted aftershock rate or its uncertainty, in the comparative judgment task. In real-world decisions, locations must often be compared where both the rate and uncertainty vary, for example, comparing a location of medium rate/low uncertainty against one of low rate/high uncertainty. Future studies should explore these comparisons by systematically testing location pairs with meaningful differences, as recommended in prior reviews (Hullman et al., 2018; Kinkeldey et al., 2017).

Geographical features, such as roads and landmarks, were omitted from our maps in order to avoid potential confounding effects on judgments, as in Nadav-Greenberg et al. (2008). However, omitting these features lowers the ecological validity of the study as geographical features are generally included in public forecast communications. Future experiments could add standard map layers to visualizations and evaluate their effects on interpretations of the forecast made using an uncertainty visualization. We found some evidence that map features influence forecast perception, especially when locations with the same forecasted rate and uncertainty were compared. Controlled experiments should target the effects of map features on task response, both with and without additional map layers.

Finally, our evaluation used a single judgment task and only considered three uncertainty visualizations. More tasks would elucidate various other effects of candidate uncertainty visualizations, especially tasks that move beyond judgment and towards decision-making. For example, if aiding resource allocation is a communication goal, participants could be

asked to allocate search and rescue teams across the map, based on the forecasted rate and its uncertainty. As aftershock and other natural hazard forecasts are increasingly released via public portals with interactive capabilities (Marzocchi et al., 2014), tasks with interactivity can assess how this moderates an uncertainty visualization design's effects. A greater variety of uncertainty visualizations designs can also be evaluated, particularly those that represent uncertainty using patterns or other overlays easier to separate from the forecast's color than color transparency. Expressing uncertainty in a way that does not use color lightness could investigate whether the dark-is-more bias or something else affects the interpretation of high uncertainty.

4.5 Conclusions and practical implications

Our experiment found that the three approaches for visualizing uncertainty in an aftershock forecast map differed substantially in their effects on map-reading and judgment responses. These visualizations included one of the most effective designs from the visualization literature (color transparency) and a newer approach with recent operational use (forecast intervals). Our work suggests several practical implications for the design of aftershock forecast maps for different users. If the accurate reading of uncertainty is the most important aim (e.g., for technical users who may need the aftershock forecast distribution as a direct input into their own models), our results support communicating this explicitly and with a separate and adjacent map. Map-reading accuracy was high across all visualizations, but uncertainty was read most accurately for adjacent designs, compared to transparency- and interval-based designs.

If instead the aim is to communicate that higher-uncertainty locations could lead to more rather than fewer aftershocks, and some inaccuracy in the (implicit) communication of uncertainty is acceptable, then our results support representing uncertainty using forecast intervals. While all visualizations were able to convey that higher-rate locations are sure

bets to have more aftershocks when uncertainty is low, only the intervals-based design communicated that high uncertainty locations were potential surprises to have more aftershocks than were forecasted. As this interpretation of high uncertainty is consistent with the skewed distribution of aftershock rates, the intervals-based visualization may improve judgments by non-technical users. For example, emergency managers rely on aftershock forecasts to decide whether to issue disaster declarations during an aftershock sequence or not. The intervals-based design may help these decisions to be more consistent with the forecast. In a crisis management situation, providing a pessimistic map may be useful for rapid risk assessment, a hypothesis which should be tested in future research.

Chapter 5

CONCLUSIONS

I'm off the deep end, watch as I dive in,
I'll never meet the ground.
Crash through the surface, where they can't hurt us,
We're far from the shallow now.

"Shallow," Lady Gaga (2018)

This dissertation makes several substantial contributions to the statistical seismology of the Pacific Northwest (PNW), a region with large seismic risk. It also opens a new line of research in how to visualize uncertainty in spatial earthquake models, contributing to the larger literature on uncertainty visualization for natural hazards. In this final section, we discuss the research conducted in the previous chapters and shine light on promising future directions for inquiry.

5.1 Contributions and future directions for statistical seismology of the PNW

The originating goal of the work in Chapters 2 and 3 was to build a statistical seismicity model for the PNW. Along the way to accomplishing this goal, we had to address challenges in the catalog, which required merging multiple data sources, addressing spatial heterogeneities, handling swarms, and separating the multiple tectonic regimes that the PNW hosts. We also had to address challenges in the model; the top-performing statistical seismicity model (ETAS) has numerical issues with its likelihood function, spurring us to develop a Bayesian inference approach. This approach was well-suited to quantify how the challenges present in

the PNW catalog affect its parameter estimates.

In Chapter 2, we created a new catalog for the PNW, which required combining three different data sources. To accomplish this, we created reproducible systems to merge different catalogs and find duplicates and other discrepancies. We then performed a completeness study, finding two complete subcatalogs within the PNW and estimating Gutenberg-Richter b-values for them. To our knowledge, this is both the first-ever harmonized international catalog for the PNW that contains information on measurement error and is the first rigorous study of its completeness and the b-value (previous work like Gomberg and Bodin (2021) fix b to 1, its canonical value). Far more work can be done with these complete catalogs, not least of which is understanding how seismicity rate estimates with the more sophisticated ETAS model affect seismic hazard assessment for the PNW.

Chapter 2 also contained several methods for identifying clusters in the PNW catalog, and for subsequent labelling as aftershock sequences or swarms. This was initially done to limit how the presence of swarms can bias ETAS estimates. But it catalyzed methodological development of a new semi-parametric approach to cluster detection, which was relatively successful at detecting true clusters in the PNW. This method can be further tested on other real catalogs (e.g., southern California, which has a much higher rate of both overall and clustered seismicity) and synthetic catalogs. The result of our collaboration with the PNSN is a catalog of aftershock sequences and swarms within the PNW, which can be explored and modelled to more directly understand overarching patterns, e.g., recurrence rates, that an ETAS model is not built to do.

In Chapter 3, we used our complete catalogs to build the first ETAS models for the PNW. To accomplish this, we had to address the well-documented issues with the ETAS likelihood (namely that it is flat near parameter optima and that its parameter correlation leads to a multimodal likelihood) which make numerical optimization very challenging and unstable. Furthermore, maximum likelihood estimates only have good statistical properties (consis-

tency and asymptotic normality) asymptotically, which holds for few catalogs, and likely not smaller ones like the PNW. Building on recent methodological work (Ross, 2021), we customized a Bayesian method of inference for the ETAS parameters for the PNW, developing a novel solution for the spatial integral in the likelihood function and eliciting informative priors for the region. The process of prior elicitation also involved the collection of all reported ETAS estimates from the literature on subduction zones. We implemented a set of C++/R functions for this Bayesian method, which was validated on synthetic catalogs. Future work should concentrate on model development, especially in making the background rate parameter μ_{ST} spatially-varying and the spatial-offset parameter d magnitude-dependent.

We applied the Bayesian model to the PNW and found results to be quite insensitive to initial values, in contrast to the MLE solutions. We then used the Bayesian model as a tool to quantify the effects of the previously-mentioned catalog issues on parameter estimates. We found that parameters were quite different between the PNW North and South (crustal regime only), with the south having unsurprisingly lower background rates and aftershock parameters that suggest that sequences there should be larger and decay in time and space more quickly. We also found that, in the PNW North, parameters were sensitive to the previously-mentioned catalog issues. When earthquakes in the deep and crustal regimes were modelled together, the background rate increases and sequences are expected to be smaller and decay more quickly in time and space, which supports the hypothesis that deeper earthquakes are less aftershock-productive. Furthermore, when swarms are included in the catalog, aftershock sequences were expected to be slightly larger and decay more quickly in time than when they are removed; better understanding how patterns of swarms differ from aftershock sequences is key in the PNW, where swarms are ubiquitous. Finally, we also showed using a number of jittered catalogs that parameter estimates were affected by catalog measurement error. This joins other statistical work that underscores the importance of this catalog issue on parameter estimates. A new version of the ETAS model that explicitly

accounts for measurement error should also be the focus of future work.

5.2 Contributions and future directions for uncertainty visualization for aftershock forecasts

Aftershock forecasts use statistical seismicity models like the one developed in Chapter 3, and have been increasingly represented using forecast maps; however, despite the many sources of uncertainty we detail in Chapter 3, forecast uncertainty is rarely visualized on these maps. In an international collaboration, we designed the first experiment to find effective uncertainty visualizations for aftershock forecast maps, presented in Chapter 4. This required us first to design three candidate visualizations, which included a basic approach (adjacent), an approach validated by visualization research (transparency) and a novel approach with recent operational use (lower and upper bounds).

We then designed an experiment to compare these visualizations against showing the forecast map without uncertainty, which is the current practice. We developed tasks for this experiment in a novel way, by first interviewing emergency managers (a key user group); synthesizing these interviews into a set of communication goals that were relevant to other users; and then structuring the tasks to elicit responses connected to each of these goals. We designed a controlled and pre-registered online experiment to robustly identify effects of the different uncertainty visualizations on the task outcomes. The experiment found the visualization using lower and upper bounds could communicate both goals, while the other two visualizations could not. These results have practical implications for aftershock forecasts, but are also relevant across other natural hazards.

Future work in this research topic can investigate how visualization can better communicate other aspects of aftershock forecasts. For example, we might be interested in focusing a forecast product on extreme cases (high forecast with a very high uncertainty), which may require novel color scales or utilize a dashboard with multiple plots. These approaches should be optimized with empirical user research. Furthermore, as public-facing forecasts

become more sophisticated, we should explore how animation or interactivity affects a given visualization's effectiveness. Better visualization of aftershock forecasts can, as Jones (2020) advocates, "provide actionable information to help vulnerable societies prevent further damage", which should be a service of scientific research to society.

BIBLIOGRAPHY

Aki, K. (1965). Maximum likelihood estimate of b in the formula log N= a-bM and its confidence limits. *Bull. Earthq. Res. Inst., Tokyo Univ.*, 43:237–239.

Ammon, C. J., Velasco, A. A., Lay, T., and Wallace, T. C. (2020). *Foundations of Modern Global Seismology*. Academic Press.

Ash, K. D., Schumann III, R. L., and Bowser, G. C. (2014). Tornado warning trade-offs: Evaluating choices for visually communicating risk. *Weather, climate, and society*, 6(1):104–118.

Baddeley, A. and Turner, R. (2005). Spatstat: an R package for analyzing spatial point patterns. *Journal of statistical software*, 12(1):1–42.

Becker, J. S., Potter, S., McBride, S., Wein, A., Doyle, E., and Paton, D. (2019). When the earth doesn't stop shaking: How experiences over time influenced information needs, communication, and interpretation of aftershock information during the Canterbury Earthquake Sequence, New Zealand. *International journal of disaster risk reduction*, 34:397–411.

Becker, J. S., Potter, S. H., McBride, S. K., H. Doyle, E. E., Gerstenberger, M. C., and Christophersen, A. (2020). Forecasting for a fractured land: A case study of the communication and use of aftershock forecasts from the 2016 Mw 7.8 Kaikōura earthquake in Aotearoa New Zealand. *Seismological Society of America*, 91(6):3343–3357.

Bilek, S. L. and Lay, T. (2018). Subduction zone megathrust earthquakes. *Geosphere*, 14(4):1468–1500.

Bilham, R. (2009). The seismic future of cities. *Bulletin of earthquake engineering*, 7(4):839–887.

Bostock, M. G., Christensen, N. I., and Peacock, S. M. (2019). Seismicity in Cascadia. *Lithos*, 332:55–66.

Bostrom, A., Anselin, L., and Farris, J. (2008). Visualizing seismic risk and uncertainty: A review of related research. *Annals of the New York Academy of Sciences*, 1128(1):29–40.

Brewer, C. A. and Harrower, M. (2009). Colorbrewer 2.0: Color advice for cartography. *The Pennsylvania State University. http://colorbrewer2. org/. Accessed*, 6(02):2010.

Brier, G. W. et al. (1950). Verification of forecasts expressed in terms of probability. *Monthly weather review*, 78(1):1–3.

Brillon, C. (2020). Private communication.

Burgeno, J. N. and Joslyn, S. L. (2020). The impact of weather forecast inconsistency on user trust. *Weather, climate, and society*, 12(4):679–694.

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing*, 16(5):1190–1208.

Chakraborty, A. and Gelfand, A. E. (2010). Analyzing spatial point patterns subject to measurement error. *Bayesian Analysis*, 5(1):97–122.

Chang, S. E. and Falit-Baiamonte, A. (2002). Disaster vulnerability of businesses in the 2001 Nisqually earthquake. *Global Environmental Change Part B: Environmental Hazards*, 4(2):59–71.

Cheong, L., Bleisch, S., Kealy, A., Tolhurst, K., Wilkening, T., and Duckham, M. (2016). Evaluating the impact of visualization of wildfire hazard upon decision-making under uncertainty. *International Journal of Geographical Information Science*, 30(7):1377–1404.

Chiodi, M. and Adelfio, G. (2021). Package 'etasflp'.

Chu, A., Schoenberg, F. P., Bird, P., Jackson, D. D., and Kagan, Y. Y. (2011). Comparison of ETAS parameter estimates across different global tectonic zones. *Bulletin of the Seismological Society of America*, 101(5):2323–2339.

Clements, R. A., Schoenberg, F. P., and Schorlemmer, D. (2011). Residual analysis methods for space-time point processes with applications to earthquake forecast models in California. *The Annals of applied statistics*, pages 2549–2571.

Console, R., Murru, M., and Falcone, G. (2010). Retrospective forecasting of M $\geq$ 4.0 earthquakes in New Zealand. *Pure and Applied Geophysics*, 167(6):693–707.

Correll, M., Moritz, D., and Heer, J. (2018). Value-suppressing uncertainty palettes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 642. ACM.

Crisan, A. and Elliott, M. (2018). How to evaluate an evaluation study? Comparing and contrasting practices in vis with those of other disciplines: Position paper. In *2018 IEEE Evaluation and Beyond-Methodological Approaches for Visualization (BELIV)*, pages 28–36. IEEE.

Cucala, L. (2008). Intensity estimation for spatial point processes observed with noise. *Scandinavian journal of statistics*, 35(2):322–334.

Deitrick, S. and Edsall, R. (2006). The influence of uncertainty visualization on decision making: An empirical evaluation. In *Progress in spatial data handling*, pages 719–738. Springer.

Doore, G. et al. (1993). Guidelines for using color to depict meteorological information. *Bull. Am. Meteor. Soc*, 74:1709–1713.

Doyle, E. E., Johnston, D. M., Smith, R., and Paton, D. (2019). Communicating model uncertainty for natural hazards: a qualitative systematic thematic review. *International journal of disaster risk reduction*, 33:449–476.

Earthquakes Canada, G. (2020). Earthquake search (on-line bulletin). data retrieved from http://earthquakescanada.nrcan.gc.ca/stndon/NEDB-BNDS/bulletin-en.php,.

Ebrahimian, H. and Jalayer, F. (2017). Robust seismicity forecasting based on Bayesian parameter estimation for epidemiological spatio-temporal aftershock clustering models. *Scientific reports*, 7(1):1–15.

Farrell, J., Husen, S., and Smith, R. B. (2009). Earthquake swarm and b-value characterization of the Yellowstone volcano-tectonic system. *Journal of Volcanology and Geothermal Research*, 188(1-3):260–276.

Fausett, L. V. (2008). *Applied numerical analysis using MATLAB*. Pearson.

Felzer, K. R. (2007). Appendix I: calculating California seismicity rates. *US Geol. Surv. Open-File Rept. 2007-1437I*.

Felzer, K. R., Abercrombie, R. E., and Ekstrom, G. (2003). Secondary aftershocks and their importance for aftershock forecasting. *Bulletin of the Seismological Society of America*, 93(4):1433–1448.

Fleischhut, N., Herzog, S. M., and Hertwig, R. (2020). Weather literacy in times of climate change. *Weather, Climate, and Society*, 12(3):435–452.

Fletcher, R. (1987). *Practical methods of optimization. 1987*. John and Sons, Chichester.

Fletcher, R. and Reeves, C. M. (1964). Function minimization by conjugate gradients. *The computer journal*, 7(2):149–154.

Frankel, A. (1995). Mapping seismic hazard in the central and eastern United States. *Seismological Research Letters*, 66(4):8–21.

Gardner, J. and Knopoff, L. (1974). Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian? *Bulletin of the seismological society of America*, 64(5):1363–1367.

Garthwaite, P. H., Kadane, J. B., and O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, 100(470):680–701.

Goltz, J. D. (2015). A further note on operational earthquake forecasting: An emergency management perspective.

Gomberg, J. and Bodin, P. (2021). The productivity of Cascadia aftershock sequences. *Bulletin of the Seismological Society of America*, 111(3):1494–1507.

Gomberg, J. and Jakobitz, A. (2013). A collaborative user-producer assessment of earthquake-response products. Technical report, US Geological Survey.

Gomberg, J., Sherrod, B., Trautman, M., Burns, E., and Snyder, D. (2012). Contemporary seismicity in and around the Yakima Fold-and-Thrust belt in eastern Washington. *Bulletin of the Seismological Society of America*, 102(1):309–320.

Guo, Z. and Ogata, Y. (1997). Statistical relations between the parameters of aftershocks in time, space, and magnitude. *Journal of Geophysical Research: Solid Earth*, 102(B2):2857–2873.

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

Guttorp, P. (1987). On least-squares estimation of b values. *Bulletin of the Seismological Society of America*, 77(6):2115–2124.

Guttorp, P. (1995). *Stochastic modeling of scientific data*. CRC Press.

Guttorp, P. and Hopkins, D. (1986). On estimating varying b values. *Bulletin of the Seismological Society of America*, 76(3):889–895.

Guttorp, P. and Thorarinsdottir, T. L. (2012). Bayesian inference for non-Markovian point processes. In *Advances and Challenges in Space-time Modelling of Natural Events*, pages 79–102. Springer.

Hainzl, S. (2016). Apparent triggering function of aftershocks resulting from rate-dependent incompleteness of earthquake catalogs. *Journal of Geophysical Research: Solid Earth*, 121(9):6499–6509.

Hainzl, S., Christophersen, A., and Enescu, B. (2008). Impact of earthquake rupture extensions on parameter estimations of point-process models. *Bulletin of the Seismological Society of America*, 98(4):2066–2072.

Hardebeck, J. L., Llenos, A. L., Michael, A. J., Page, M. T., and Van Der Elst, N. (2019). Updated California aftershock parameters. *Seismological Research Letters*, 90(1):262–270.

Harte, D. (2013). Bias in fitting the ETAS model: a case study based on New Zealand seismicity. *Geophysical Journal International*, 192(1):390–412.

Harte, D. (2018). Effect of sample size on parameter estimates and earthquake forecasts. *Geophysical Journal International*, 214(2):759–772.

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90.

Hayes, G. P., Moore, G. L., Portner, D. E., Hearne, M., Flamme, H., Furtney, M., and Smoczyk, G. M. (2018). Slab2, a comprehensive subduction zone geometry model. *Science*, 362(6410):58–61.

Heinrich-Mertsching, C., Thorarinsdottir, T. L., Guttorp, P., and Schneider, M. (2021). Validation of point process predictions with proper scoring rules. *arXiv preprint arXiv:2110.11803*.

Holtkamp, S. and Brudzinski, M. (2011). Earthquake swarms in circum-Pacific subduction zones. *Earth and Planetary Science Letters*, 305(1-2):215–225.

Hullman, J., Qiao, X., Correll, M., Kale, A., and Kay, M. (2018). In pursuit of error: A survey of uncertainty visualization evaluation. *IEEE transactions on visualization and computer graphics*, 25(1):903–913.

Husen, S. and Hardebeck, J. (2010). Earthquake location accuracy. *CORSSA*.

Jacobs, K. M., Smith, E. G., Savage, M. K., and Zhuang, J. (2013). Cumulative rate analysis (CURATE): A clustering algorithm for swarm dominated catalogs. *Journal of Geophysical Research: Solid Earth*, 118(2):553–569.

Jaiswal, K., Bausch, D., Rozelle, J., Holub, J., and McGowan, S. (2017). Hazus® estimated annualized earthquake losses for the United States. Technical report, Federal Emergency Management Agency.

Jalilian, A. (2019). Etas: an R package for fitting the space-time ETAS model to earthquake data. *Journal of Statistical Software*, 88(1):1–39.

Johnson, S. R., Tomlinson, G. A., Hawker, G. A., Granton, J. T., and Feldman, B. M. (2010). Methods to elicit beliefs for Bayesian priors: a systematic review. *Journal of clinical epidemiology*, 63(4):355–369.

Jones, L. M. (2020). Empowering the public with earthquake science. *Nature Reviews Earth & Environment*, 1(1):2–3.

Joslyn, S. and Savelli, S. (2010). Communicating forecast uncertainty: Public perception of weather forecast uncertainty. *Meteorological Applications*, 17(2):180–195.

Joslyn, S. L. and LeClerc, J. E. (2012). Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error. *Journal of experimental psychology: applied*, 18(1):126.

Kasahara, A., Yagi, Y., and Enescu, B. (2016). etas_solve: A robust program to estimate the ETAS model parameters. *Seismological Research Letters*, 87(5):1143–1149.

Kim, A. Y. and Wakefield, J. (2010). R data and methods for spatial epidemiology: The spatialepi package. *Dept of Statistics, University of Washington*.

Kinkeldey, C., MacEachren, A. M., Riveiro, M., and Schiewe, J. (2017). Evaluating the effect of visually represented geodata uncertainty on decision-making: systematic review, lessons learned, and recommendations. *Cartography and Geographic Information Science*, 44(1):1–21.

Kinkeldey, C., MacEachren, A. M., and Schiewe, J. (2014). How to assess visual communication of uncertainty? A systematic review of geospatial uncertainty visualisation user studies. *The Cartographic Journal*, 51(4):372–386.

Kolev, A. A. and Ross, G. J. (2019). Inference for ETAS models with non-Poissonian mainshock arrival times. *Statistics and Computing*, 29(5):915–931.

Krauskopf, J., Zaidi, Q., and Mandlert, M. B. (1986). Mechanisms of simultaneous color induction. *JOSA A*, 3(10):1752–1757.

Lee, W. H. K., Bennett, R., and Meagher, K. (1972). *A method of estimating magnitude of local earthquakes from signal duration*. Citeseer.

Lepage, G. P. (1980). Vegas-an adaptive multi-dimensional integration program. Technical report.

Lippiello, E., Giacco, F., Arcangelis, L. d., Marzocchi, W., and Godano, C. (2014). Parameter estimation in the ETAS model: Approximations and novel methods. *Bulletin of the Seismological Society of America*, 104(2):985–994.

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

Llenos, A. L. and Michael, A. J. (2013). Modeling earthquake rate changes in Oklahoma and Arkansas: Possible signatures of induced seismicity. *Bulletin of the Seismological Society of America*, 103(5):2850–2861.

Llenos, A. L. and Michael, A. J. (2019). Ensembles of ETAS models provide optimal operational earthquake forecasting during swarms: Insights from the 2015 San Ramon, California swarm. *Bulletin of the Seismological Society of America*, 109(6):2145–2158.

Lombardi, A. M. (2015). Estimation of the parameters of ETAS models by simulated annealing. *Scientific reports*, 5(1):1–11.

Lombardi, A. M. (2017). The epistemic and aleatory uncertainties of the ETAS-type models: an application to the Central Italy seismicity. *Scientific reports*, 7(1):1–9.

Ludwin, R. (2004). PNSN, Pacific Northwest Seismograph Network. Technical report.

Lund, J. and Rudemo, M. (2000). Models for point processes observed with noise. *Biometrika*, 87(2):235–249.

Malone, S. (2020). Private communication.

Malone, S., Oppenheimer, D., Gee, L., and Neuhauser, D. (1996). The Council of the National Seismic System and a composite earthquake catalog for the United States. *IRIS Newsletter, XV*, 1:6–9.

Malone, S. D. (2019). An ominous (?) quiet in the Pacific Northwest.

Marin, J.-M. and Robert, C. (2007). *Bayesian core: a practical approach to computational Bayesian statistics*. Springer Science & Business Media.

Marzocchi, W., Lombardi, A. M., and Casarotti, E. (2014). The establishment of an operational earthquake forecasting system in Italy. *Seismological Research Letters*, 85(5):961–969.

McBride, S. K., Llenos, A. L., Page, M. T., and Van Der Elst, N. (2020). # earthquakeadvisory: Exploring discourse between government officials, news media, and social media during the 2016 bombay beach swarm. *Seismological Research Letters*, 91(1):438–451.

McFee, S. (1997). hp adaptive numerical integration techniques for finite element analysis in magnetics. *IEEE Transactions on Magnetics*, 33(5):4116–4118.

Meyer, M. and Dykes, J. (2019). Criteria for rigor in visualization design study. *IEEE transactions on visualization and computer graphics*, 26(1):87–97.

Michael, A. (2021). Private communication.

Michael, A. J. (2014). How complete is the ISC-GEM global earthquake catalog? *Bulletin of the Seismological Society of America*, 104(4):1829–1837.

Michael, A. J. (2018). *On the potential duration of the aftershock sequence of the 2018 Anchorage earthquake*. US Department of the Interior, US Geological Survey.

Michael, A. J., McBride, S. K., Hardebeck, J. L., Barall, M., Martinez, E., Page, M. T., van der Elst, N., Field, E. H., Milner, K. R., and Wein, A. M. (2020). Statistical seismology and communication of the USGS operational aftershock forecasts for the 30 November 2018 Mw 7.1 Anchorage, Alaska, earthquake. *Seismological Research Letters*, 91(1):153–173.

Miller, R. (1981). *Simultaneous statistical inference*. Springer.

Milner, K. R., Field, E. H., Savran, W. H., Page, M. T., and Jordan, T. H. (2020). Operational earthquake forecasting during the 2019 Ridgecrest, California, earthquake sequence with the UCERF3-ETAS model. *Seismological Research Letters*, 91(3):1567–1578.

Morris, D. E., Oakley, J. E., and Crowe, J. A. (2014). A web-based tool for eliciting probability distributions from experts. *Environmental Modelling & Software*, 52:1–4.

Morss, R. E., Lazo, J. K., and Demuth, J. L. (2010). Examining the use of weather forecasts in decision scenarios: Results from a US survey with implications for uncertainty communication. *Meteorological Applications*, 17(2):149–162.

Mueller, C. S. (2019). Earthquake catalogs for the USGS national seismic hazard maps. *Seismological Research Letters*, 90(1):251–261.

Mulder, K. J., Lickiss, M., Harvey, N., Black, A., Charlton-Perez, A., Dacre, H., and McCloy, R. (2017). Visualizing volcanic ash forecasts: scientist and stakeholder decisions using different graphical representations and conflicting forecasts. *Weather, Climate, and Society*, 9(3):333–348.

Müller, S., Scealy, J. L., Welsh, A. H., et al. (2013). Model selection in linear mixed models. *Statistical Science*, 28(2):135–167.

Murru, M., Zhuang, J., Console, R., and Falcone, G. (2015). Short-term earthquake forecast-

ing experiment before and during the L'Aquila (central Italy) seismic sequence of April 2009. *Annals of geophysics*, 57(6).

Nadav-Greenberg, L. and Joslyn, S. L. (2009). Uncertainty forecasts improve decision making among nonexperts. *Journal of Cognitive Engineering and Decision Making*, 3(3):209–227.

Nadav-Greenberg, L., Joslyn, S. L., and Taing, M. U. (2008). The effect of uncertainty visualizations on decision making in weather forecasting. *Journal of Cognitive Engineering and Decision Making*, 2(1):24–47.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The computer journal*, 7(4):308–313.

Nicolis, O., Chiodi, M., and Adelfio, G. (2015). Windowed ETAS models with application to the Chilean seismic catalogs. *Spatial Statistics*, 14:151–165.

NOAA Office of Coast SurveyNOAA Office of Coast Survey (2020). Maritime limits and boundaries. data retrieved from NOAA Office of Coast Survey, https://nauticalcharts.noaa.gov/data/us-maritime-limits-and-boundaries.html#general-information.

Oakley, J. (2010). Eliciting univariate probability distributions. *Rethinking risk measurement and reporting*, 1:155–177.

Ogata, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*, 30(2):243–261.

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

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical association*, 83(401):9–27.

Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2):379–402.

Ogata, Y. (2011). Significant improvements of the space-time ETAS model for forecasting of accurate baseline seismicity. *Earth, planets and space*, 63(3):217–229.

Ogata, Y. and Zhuang, J. (2006). Space–time ETAS models and an improved extension. *Tectonophysics*, 413(1-2):13–23.

Okutani, T. and Ide, S. (2011). Statistic analysis of swarm activities around the Boso Peninsula, Japan: Slow slip events beneath Tokyo Bay? *Earth, planets and space*, 63(5):419–426.

Omi, T., Ogata, Y., Hirata, Y., and Aihara, K. (2015). Intermediate-term forecasting of aftershocks from an early aftershock sequence: Bayesian and ensemble forecasting approaches. *Journal of Geophysical Research: Solid Earth*, 120(4):2561–2578.

Omori, F. (1894). On the after-shocks of earthquakes. *J. Coll. Sci., Imp. Univ., Japan*, 7:111–200.

Padilla, L. M., Ruginski, I. T., and Creem-Regehr, S. H. (2017). Effects of ensemble and summary displays on interpretations of geospatial uncertainty data. *Cognitive research: principles and implications*, 2(1):1–16.

Page, M. T., Van Der Elst, N., Hardebeck, J., Felzer, K., and Michael, A. J. (2016). Three ingredients for improved global aftershock forecasts: Tectonic region, time-dependent catalog incompleteness, and intersequence variability. *Bulletin of the Seismological Society of America*, 106(5):2290–2301.

Pang, A. (2008). Visualizing uncertainty in natural hazards. In *Risk Assessment, Modeling and Decision Support*, pages 261–294. Springer.

Paolacci, G. and Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current directions in psychological science*, 23(3):184–188.

Paolacci, G., Chandler, J., and Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision making*, 5(5):411–419.

Petersen, M., Mueller, C., Haller, K., Moschetti, M., Harmsen, S., Field, E., Rukstales, K., Zeng, Y., Perkins, D., Powers, P., et al. (2012). 2014 update of the United States national seismic hazard maps.

Reasenberg, P. (1985). Second-order moment of central California seismicity, 1969–1982. *Journal of Geophysical Research: Solid Earth*, 90(B7):5479–5495.

Reasenberg, P. A. and Jones, L. M. (1989). Earthquake hazard after a mainshock in California. *Science*, 243(4895):1173–1176.

Reich, B. J. and Ghosh, S. K. (2019). *Bayesian statistical methods*. CRC Press.

Reinhart, A. (2018). A review of self-exciting spatio-temporal point processes and their applications. *Statistical Science*, 33(3):299–318.

Retchless, D. P. and Brewer, C. A. (2016). Guidance for representing uncertainty on global temperature change maps. *International Journal of Climatology*, 36(3):1143–1159.

Richter, C. F. (1935). An instrumental earthquake magnitude scale. *Bulletin of the seismological society of America*, 25(1):1–32.

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of applied probability*, 13(2):255–266.

Robert, C. P. and Casella, G. (1999). The Metropolis—Hastings algorithm. In *Monte Carlo Statistical Methods*, pages 231–283. Springer.

Roland, E. and McGuire, J. J. (2009). Earthquake swarms on transform faults. *Geophysical Journal International*, 178(3):1677–1690.

Ross, G. J. (2017). *bayesianETAS: Bayesian Estimation of the ETAS Model for Earthquake Occurrences*. R package version 1.0.3.

Ross, G. J. (2021). Bayesian estimation of the ETAS model for earthquake occurrences. *Bulletin of the Seismological Society of America*, 111(3):1473–1480.

Saichev, A. and Sornette, D. (2007). Power law distributions of seismic rates. *Tectonophysics*, 431(1-4):7–13.

Schloss, K. B., Gramazio, C. C., Silverman, A. T., Parker, M. L., and Wang, A. S. (2018). Mapping color to meaning in colormap data visualizations. *IEEE transactions on visualization and computer graphics*, 25(1):810–819.

Schnabel, R. B., Koonatz, J. E., and Weiss, B. E. (1985). A modular system of algorithms for unconstrained minimization. *ACM Transactions on Mathematical Software (TOMS)*, 11(4):419–440.

Schneider, M. and Guttorp, P. (2018). Modelling PNW seismicity with HIST-ETAS: Towards improved aftershock forecasting. *Workshop on Probability and Statistical Modeling, and Applications (Smoegen, Sweden)*.

Schoenberg, F. P. (2013). Facilitated estimation of ETAS. *Bulletin of the Seismological Society of America*, 103(1):601–605.

Schoenberg, F. P. (2016). A note on the consistent estimation of spatial-temporal point process parameters. *Statistica Sinica*, pages 861–879.

Scholz, R. W. and Lu, Y. (2014). Uncertainty in geographic data on bivariate maps: an examination of visualization preference and decision making. *ISPRS International Journal of Geo-Information*, 3(4):1180–1197.

Schorlemmer, D., Werner, M. J., Marzocchi, W., Jordan, T. H., Ogata, Y., Jackson, D. D., Mak, S., Rhoades, D. A., Gerstenberger, M. C., Hirata, N., et al. (2018). The Collaboratory for the Study of Earthquake Predictability: achievements and priorities. *Seismological Research Letters*, 89(4):1305–1313.

Seif, S., Mignan, A., Zechar, J. D., Werner, M. J., and Wiemer, S. (2017). Estimating ETAS: The effects of truncation, missing data, and model assumptions. *Journal of Geophysical Research: Solid Earth*, 122(1):449–469.

Seipel, S. and Lim, N. J. (2017). Color map design for visualization in flood risk assessment. *International Journal of Geographical Information Science*, 31(11):2286–2309.

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(B6).

Sherman-Morris, K., Antonelli, K. B., and Williams, C. C. (2015). Measuring the effectiveness of the graphical communication of hurricane storm surge threat. *Weather, climate, and society*, 7(1):69–82.

Shi, Y. and Bolt, B. A. (1982). The standard error of the magnitude-frequency b value. *Bulletin of the Seismological Society of America*, 72(5):1677–1687.

Silverman, A., Gramazio, C., and Schloss, K. (2016). The dark is more (dark+) bias in colormap data visualizations with legends. *Journal of Vision*, 16(12):628–628.

Skoumal, R. J., Brudzinski, M. R., and Currie, B. S. (2016). An efficient repeating signal detector to investigate earthquake swarms. *Journal of Geophysical Research: Solid Earth*, 121(8):5880–5897.

Spiegelhalter, D., Pearson, M., and Short, I. (2011). Visualizing uncertainty about the future. *science*, 333(6048):1393–1400.

Thompson, M. A., Lindsay, J. M., and Gaillard, J.-C. (2015). The influence of probabilistic volcanic hazard map properties on hazard communication. *Journal of Applied Volcanology*, 4(1):1–24.

Utsu, T. (1961). A statistical study on the occurrence of aftershocks. *Geophys. Mag.*, 30:521–605.

Utsu, T. (1972). Aftershocks and earthquake statistics (3): Analyses of the distribution of earthquakes in magnitude, time and space with special consideration to clustering characteristics of earthquake occurrence. *Journal of the Faculty of Science, Hokkaido University. Series 7, Geophysics*, 3(5):379–441.

Utsu, T., Ogata, Y., et al. (1995). The centenary of the Omori formula for a decay law of aftershock activity. *Journal of Physics of the Earth*, 43(1):1–33.

Utsu, T. and Seki, A. (1955). Relation between the area of aftershock region and the energy of the main shock. *Zisin (Journal of the Seismological Society of Japan)*.

Van Der Bles, A. M., Van Der Linden, S., Freeman, A. L., Mitchell, J., Galvao, A. B., Zaval, L., and Spiegelhalter, D. J. (2019). Communicating uncertainty about facts, numbers and science. *Royal Society open science*, 6(5):181870.

van der Elst, N. J. (2017). Accounting for orphaned aftershocks in the earthquake background rate. *Geophysical Journal International*, 211(2):1108–1118.

van der Elst, N. J. (2021). B-positive: A robust estimator of aftershock magnitude distribution in transiently incomplete catalogs. *Journal of Geophysical Research: Solid Earth*, 126(2):e2020JB021027.

Veen, A. and Schoenberg, F. P. (2008). Estimation of space–time branching process models in seismology using an EM–type algorithm. *Journal of the American Statistical Association*, 103(482):614–624.

Viard, T., Caumon, G., and Levy, B. (2011). Adjacent versus coincident representations of geospatial uncertainty: Which promote better decisions? *Computers & Geosciences*, 37(4):511–520.

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

Waldstreicher, J. S. and Radell, D. B. (2020). New visualization techniques, verification tools and results from the NWS Probabilistic Snowfall Experiment. In *100th American Meteorological Society Annual Meeting*. AMS.

Wang, Q., Schoenberg, F. P., and Jackson, D. D. (2010). Standard errors of parameter estimates in the ETAS model. *Bulletin of the Seismological Society of America*, 100(5A):1989–2001.

Wei, M., McGuire, J. J., and Richardson, E. (2012). A slow slip event in the south central Alaska Subduction Zone and related seismicity anomaly. *Geophysical Research Letters*, 39(15).

Wells, D. L. and Coppersmith, K. J. (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of the seismological Society of America*, 84(4):974–1002.

Werner, M. J., Helmstetter, A., Jackson, D. D., and Kagan, Y. Y. (2011). High-resolution long-term and short-term earthquake forecasts for California. *Bulletin of the Seismological Society of America*, 101(4):1630–1648.

Wiemer, S. and Wyss, M. (2000). Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the western United States, and Japan. *Bulletin of the Seismological Society of America*, 90(4):859–869.

Wirth, E. A. and Frankel, A. D. (2019). Impact of down-dip rupture limit and high-stress drop subevents on coseismic land-level change during Cascadia megathrust earthquakes. *Bulletin of the Seismological Society of America*, 109(6):2187–2197.

Yazdi, P., Gaspar-Escribano, J. M., Santoyo, M. A., and Staller, A. (2019). Analysis of the 2014 Mw 7.3 Papanoa (Mexico) earthquake: Implications for seismic hazard assessment. *Seismological Research Letters*, 90(5):1801–1811.

Zack, E. S., Kennedy, J., and Long, J. S. (2019). Can nonprobability samples be used for social science research? a cautionary tale. In *Survey Research Methods*, volume 13, pages 215–227.

Zaliapin, I. and Ben-Zion, Y. (2013a). Earthquake clusters in southern California I: Identification and stability. *Journal of Geophysical Research: Solid Earth*, 118(6):2847–2864.

Zaliapin, I. and Ben-Zion, Y. (2013b). Earthquake clusters in southern California II: Classification and relation to physical properties of the crust. *Journal of Geophysical Research: Solid Earth*, 118(6):2865–2877.

Zaliapin, I. and Ben-Zion, Y. (2021). Perspectives on clustering and declustering of earthquakes. *Seismological Research Letters*.

Zhang, L., Werner, M. J., and Goda, K. (2020). Variability of ETAS parameters in global subduction zones and applications to mainshock–aftershock hazard assessment. *Bulletin of the Seismological Society of America*, 110(1):191–212.

Zhuang, J. (2006). Second-order residual analysis of spatiotemporal point processes and applications in model evaluation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(4):635–653.

Zhuang, J. (2011). Next-day earthquake forecasts for the Japan region generated by the ETAS model. *Earth, planets and space*, 63(3):5.

Zhuang, J. (2015). Weighted likelihood estimators for point processes. *Spatial Statistics*, 14:166–178.

Zhuang, J., Ogata, Y., and Vere-Jones, D. (2002). Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, 97(458):369–380.

Zhuang, J. and Touati, S. (2015). Stochastic simulation of earthquake catalogs. *Community Online Resource for Statistical Seismicity Analysis*, 29.

Appendix A

APPENDIX TO CHAPTER 2

A.1 Magnitude-frequency for PNSN in different decades

We examined magnitude frequency distributions using plots of logged magnitude tail frequencies, with simultaneous confidence bounds (see Figure A.1). There were no major differecnes distinguishable between the last decade (when the M_d magnitude scale was adopted), and the first three decades that used the M_l magnitude scale. We omit the first 9 years of the catalog period for space reasons.

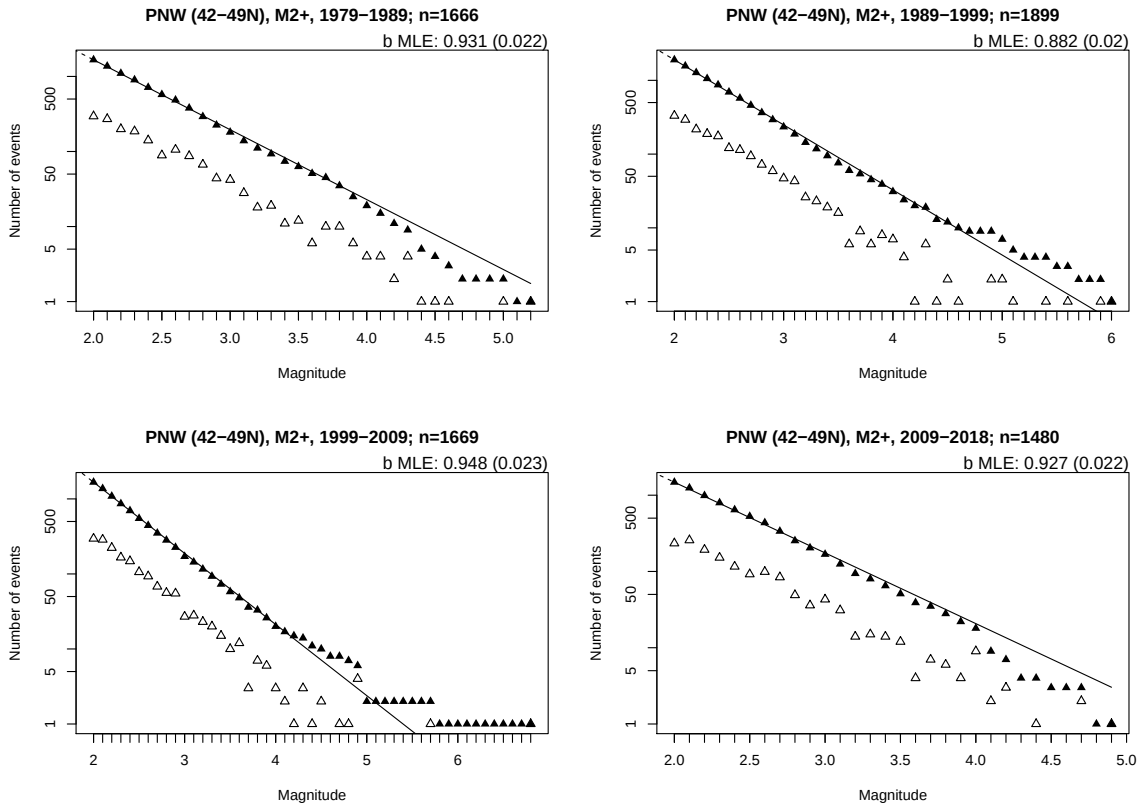

Figure A.1: Magnitude frequency distribution for the PNW, for consecutive 10-year periods. We plot the magnitude v. logged magnitude tail frequency and its simultaneous confidence bands (see explanation in Section 2.8.1).

A.2 PNSN station maps

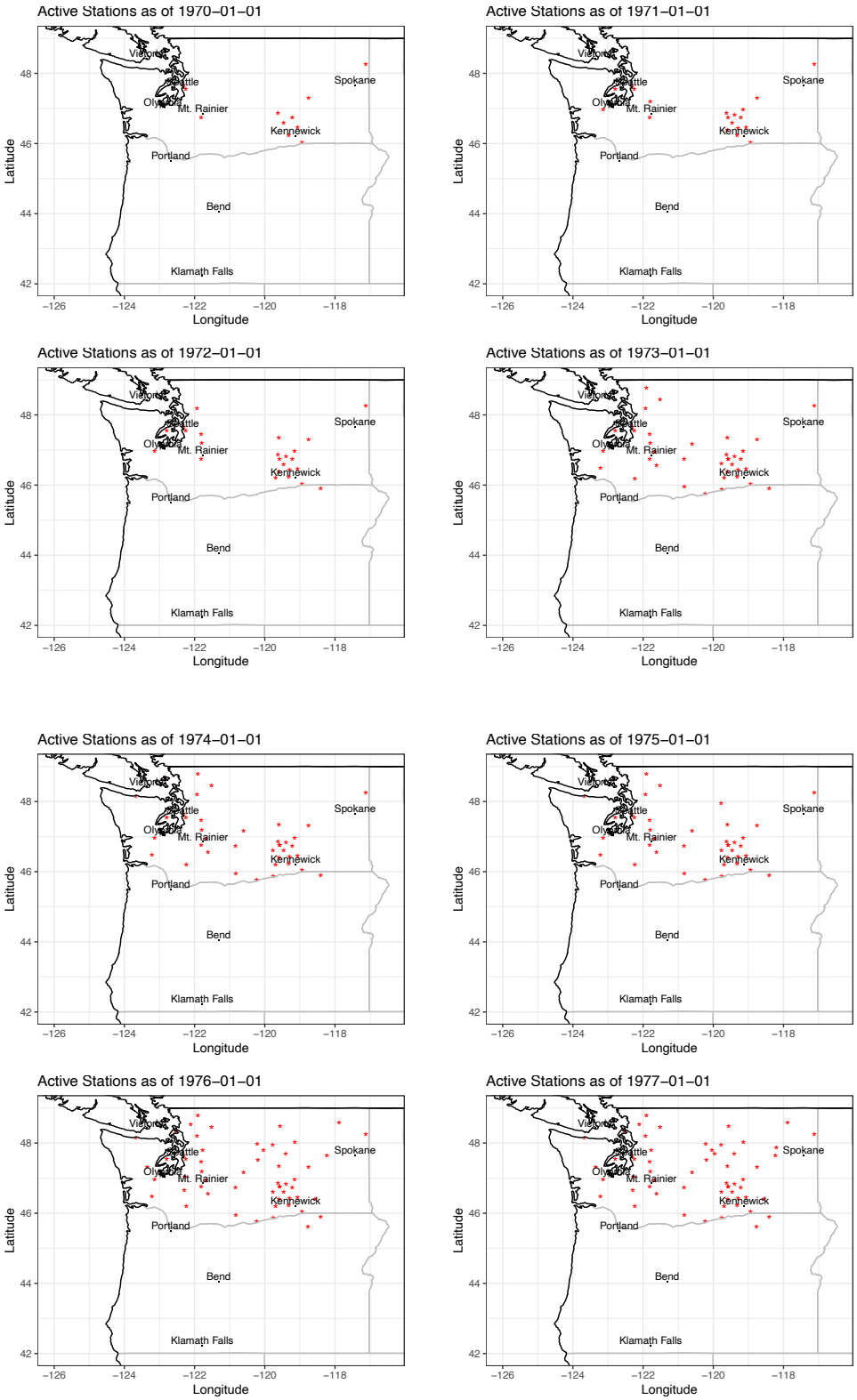

Figure A.2: Map of PNSN stations in network at start of each year.

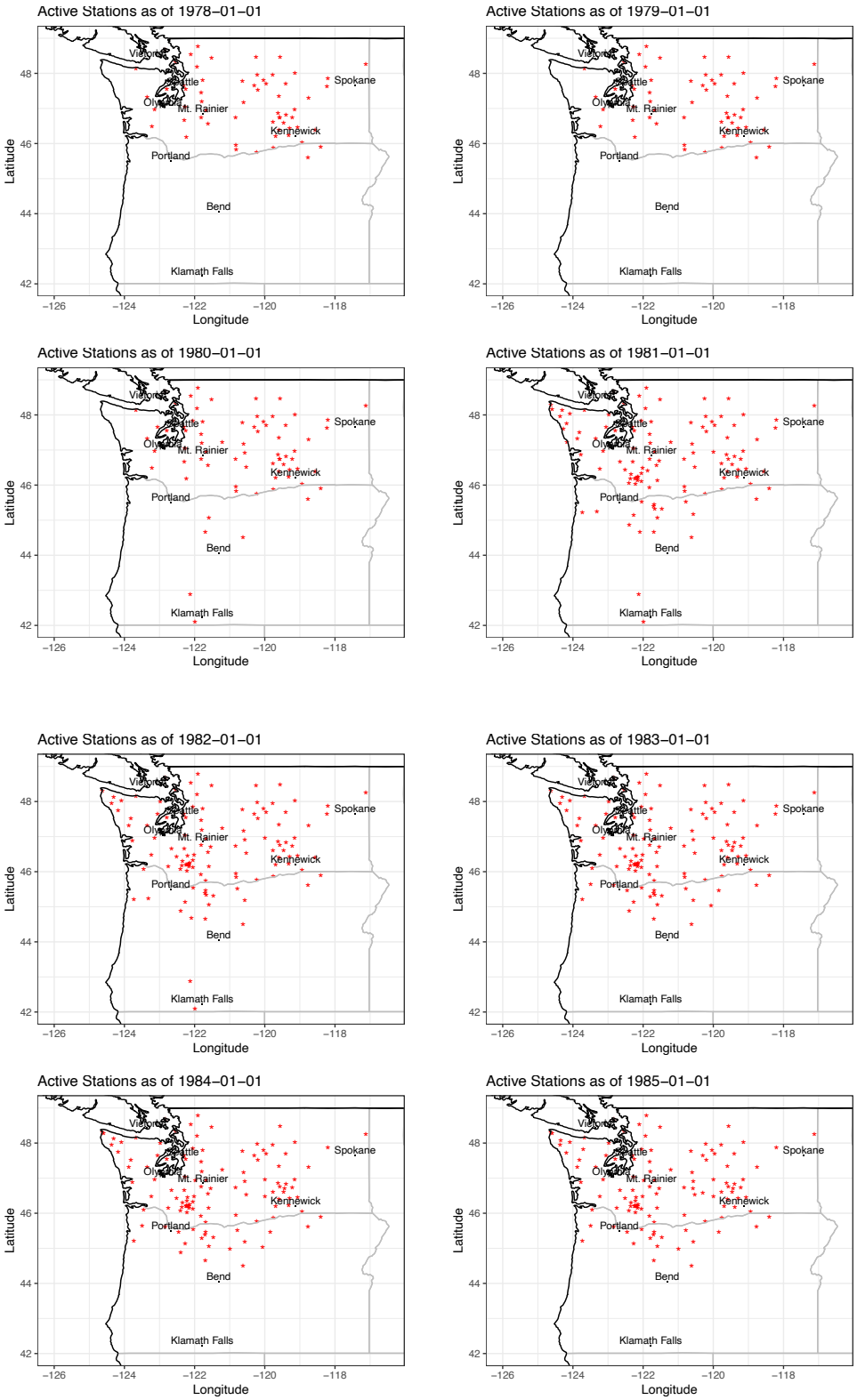

Figure A.3: Map of PNSN stations in network at start of each year (continued).

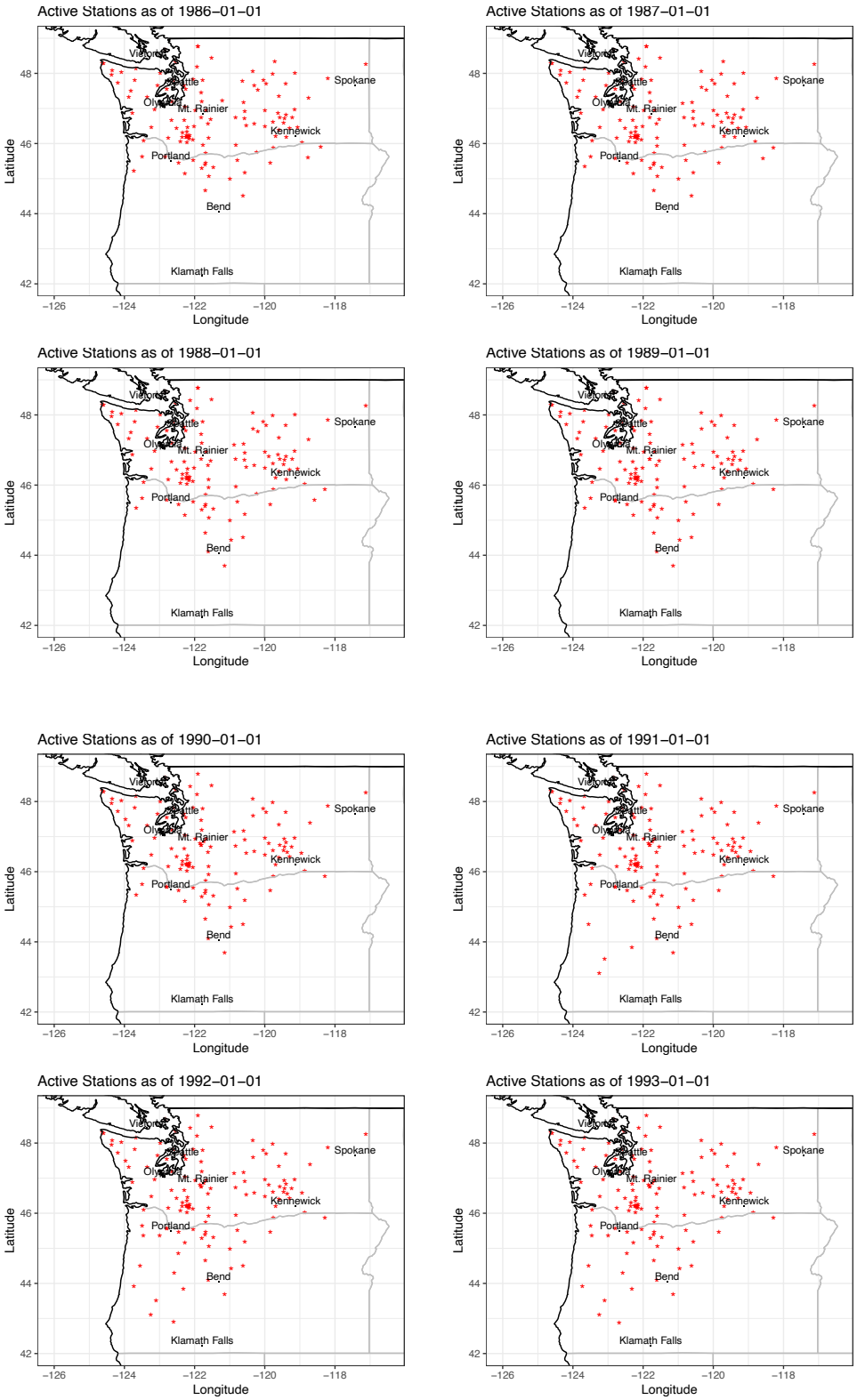

Figure A.4: Map of PNSN stations in network at start of each year (continued).

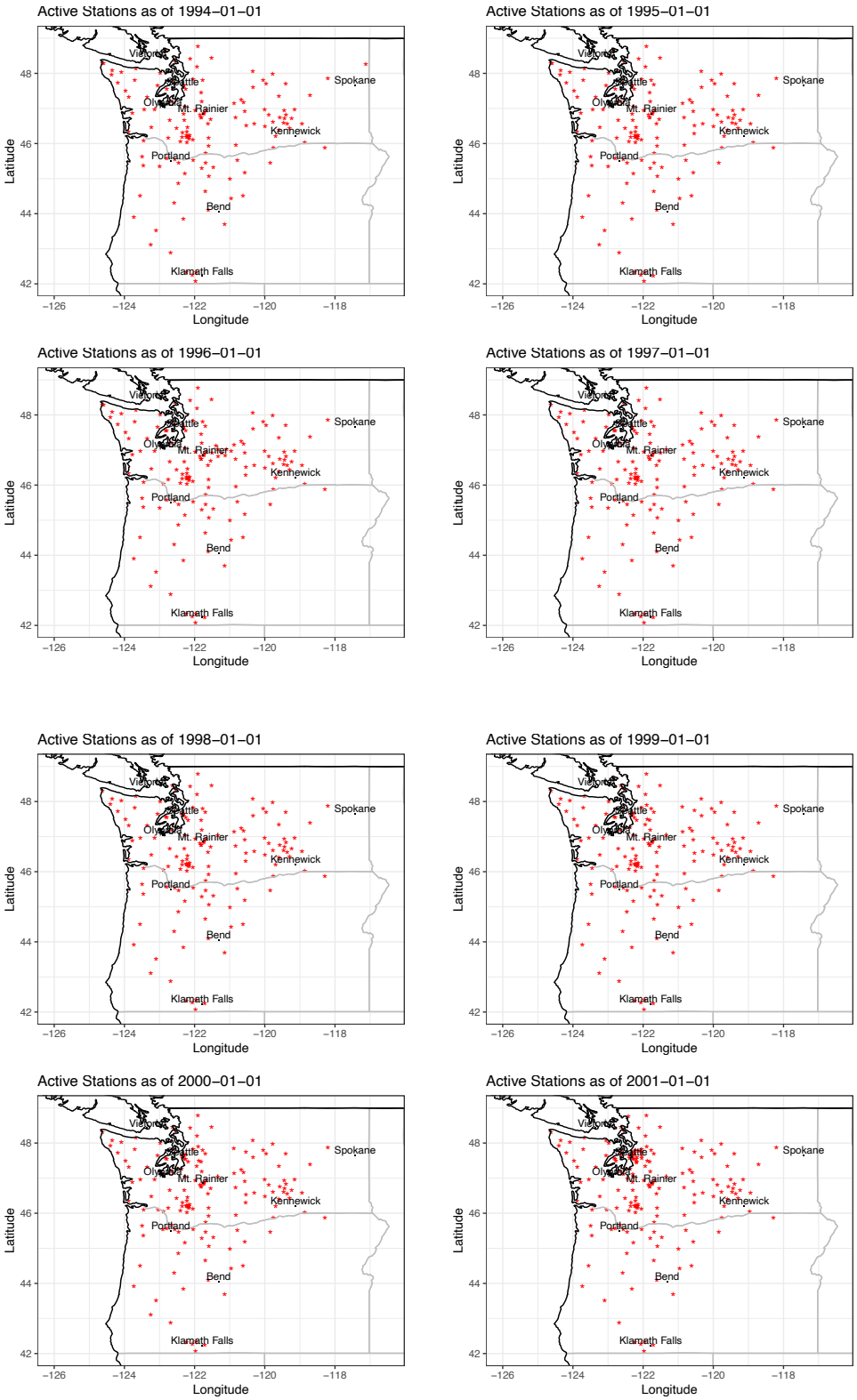

Figure A.5: Map of PNSN stations in network at start of each year (continued).

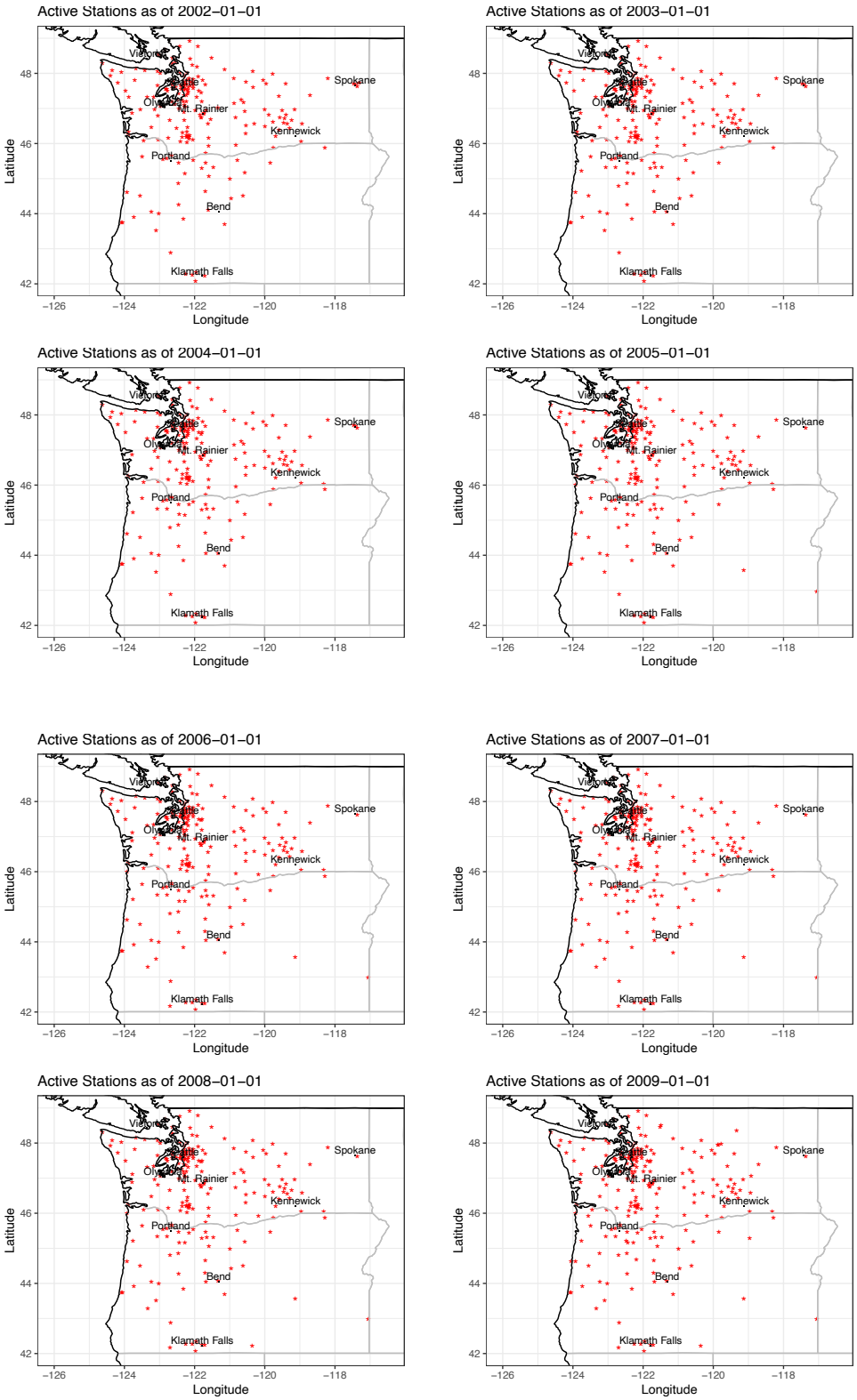

Figure A.6: Map of PNSN stations in network at start of each year (continued).

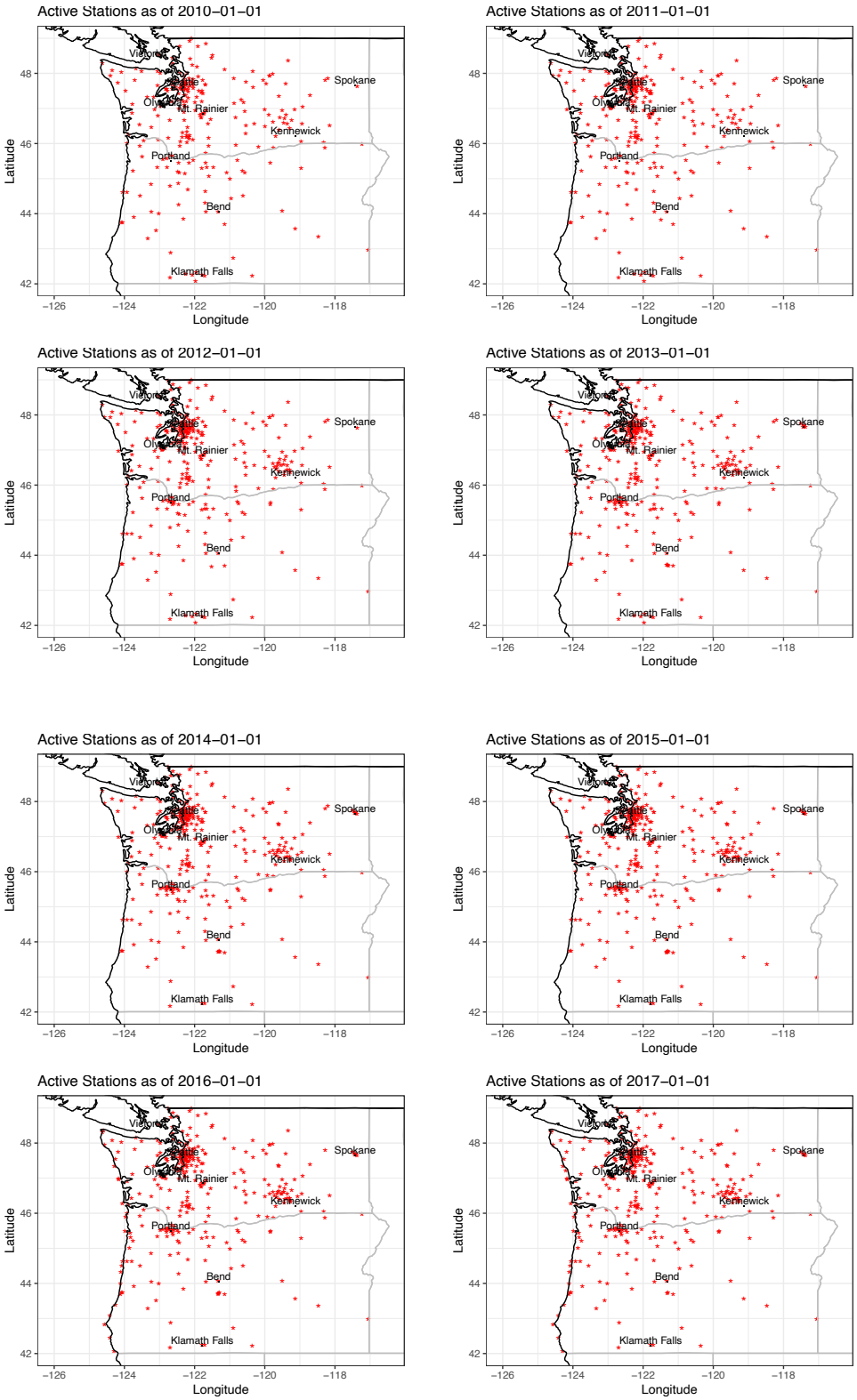

Figure A.7: Map of PNSN stations in network at start of each year (continued).

A.3 Additional materials for semi-parametric cluster detection

	10%	20%	25%	30%	35%	40%	50%
Area 1	0.14	0.61	0.90	1.20	1.57	1.96	2.94
Area 2	0.11	0.63	1.24	1.87	2.65	3.33	5.35
Area 3	0.01	0.02	0.03	0.04	0.07	0.11	0.33
Area 4	0.23	0.89	1.28	1.83	2.42	3.17	4.89
Area 5	0.19	1.15	1.85	4.51	8.22	10.73	26.81

Table A.1: Percentiles of $\Delta t_{i,i-1}$ (time differences, in decimal days), by area.

	10%	20%	25%	30%	35%	40%	50%
Area1	31.54	47.90	55.48	62.98	70.30	77.68	91.82
Area 2	29.05	52.94	62.53	70.58	79.01	88.43	110.76
Area 3	4.22	8.13	10.97	14.62	20.48	48.74	139.99
Area 4	49.36	80.02	92.81	103.26	112.72	122.05	145.19
Area 5	43.93	75.34	88.69	105.31	124.69	141.91	181.00

Table A.2: Percentiles of Δs_{ij} (spatial distances, in km), by area.

A.4 Additional plots for completeness study

A.5 Code appendix to Chapter 2

This appendix documents the code involved in processing the combined Pacific Northwest catalog. We followed the steps below in processing the component catalogs and merging them.

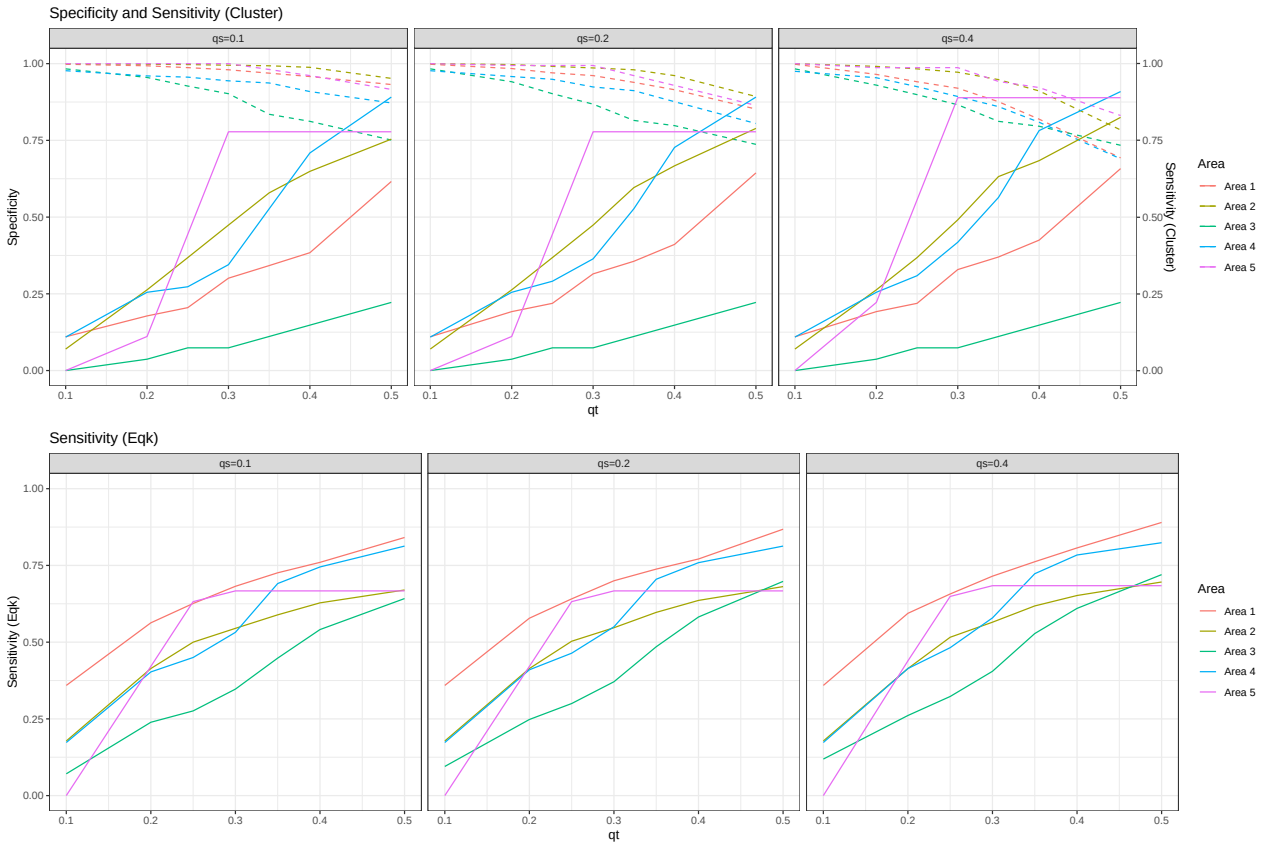

Figure A.8: Top: Cluster detection sensitivity (cluster level) in solid line and specificity in dotted line, compared to the true cluster, when $\tilde{t} = 2q_t, \tilde{s} = 2q_s$. Bottom: Cluster detection sensitivity at the earthquake level, when $\tilde{t} = 2q_t, \tilde{s} = 2q_s$..

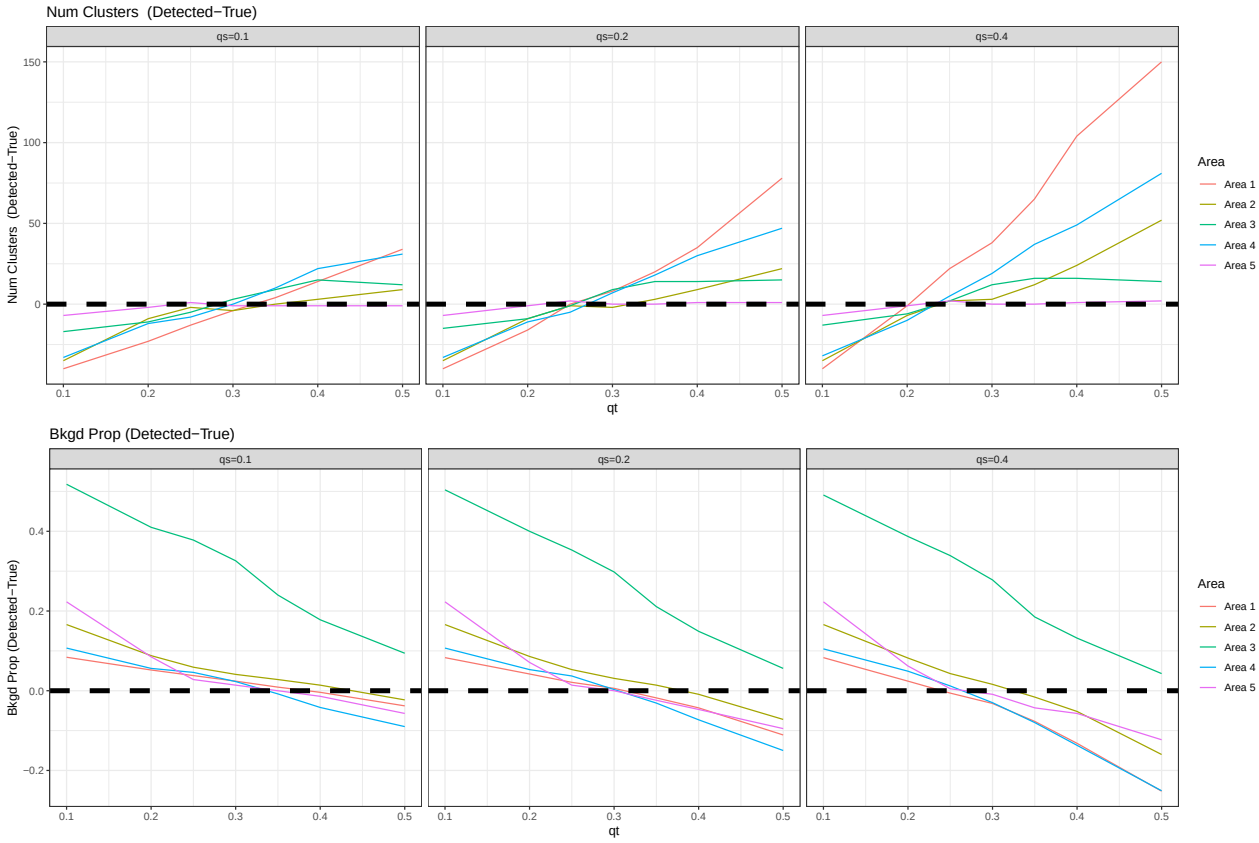

Figure A.9: Cluster detection performance metrics for number of clusters detected and proportion of background (non-clustered) events detected, when $\tilde{t} = 2q_t, \tilde{s} = 2q_s$..

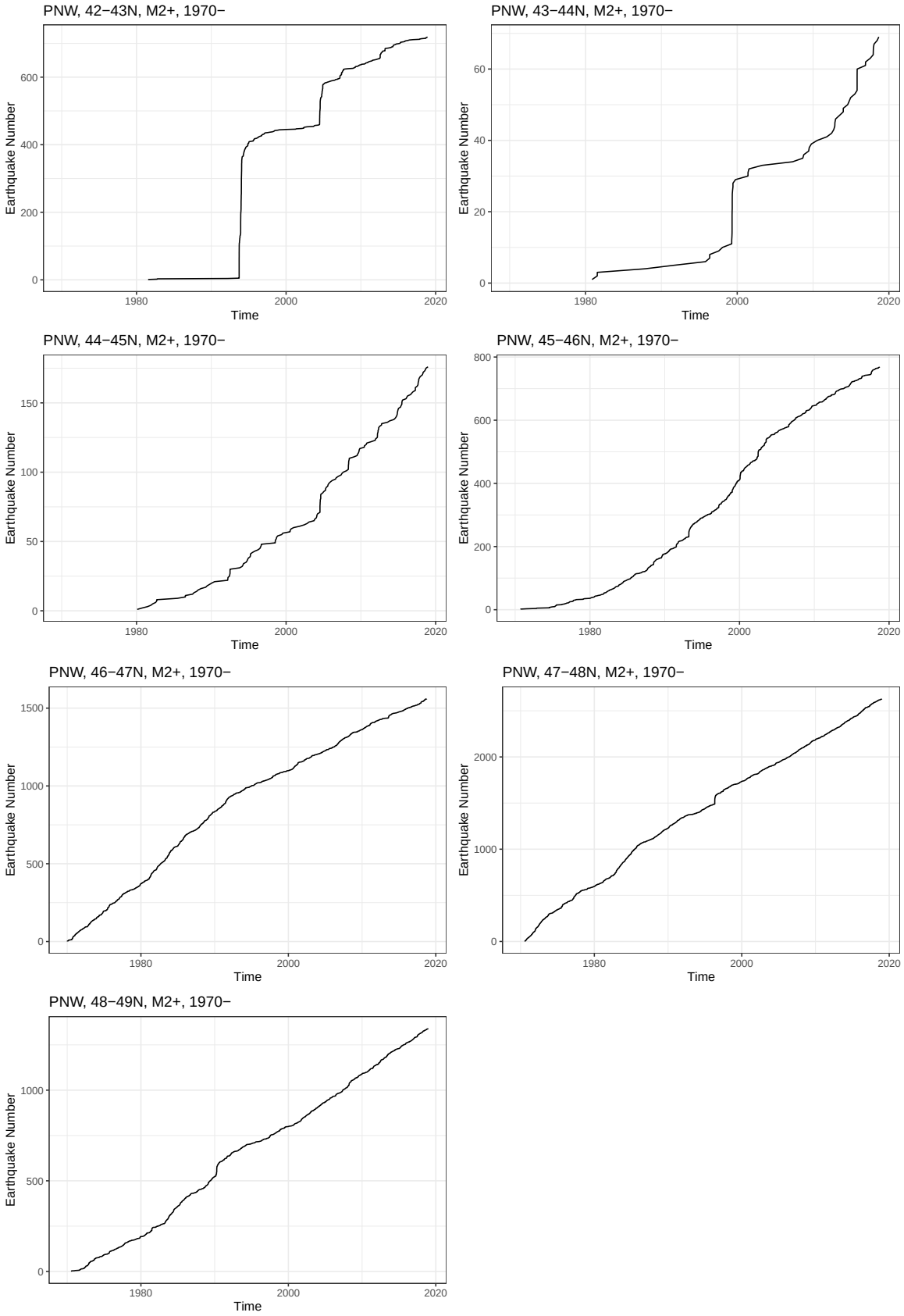

Figure A.10: Cumulative time plots for 1° bands.

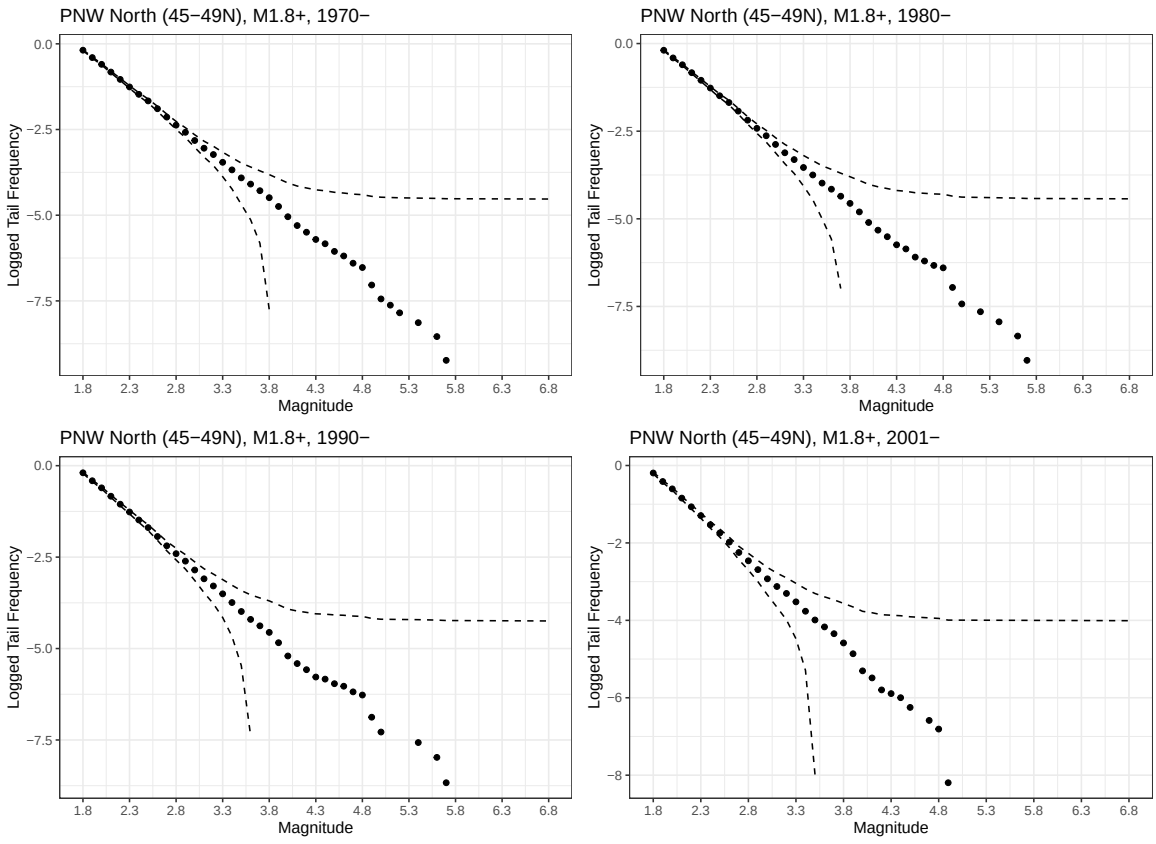

Figure A.11: Magnitude frequency distribution for the PNW North, for different start years. We plot the magnitude v. logged magnitude tail frequency and its simultaneous confidence bands.

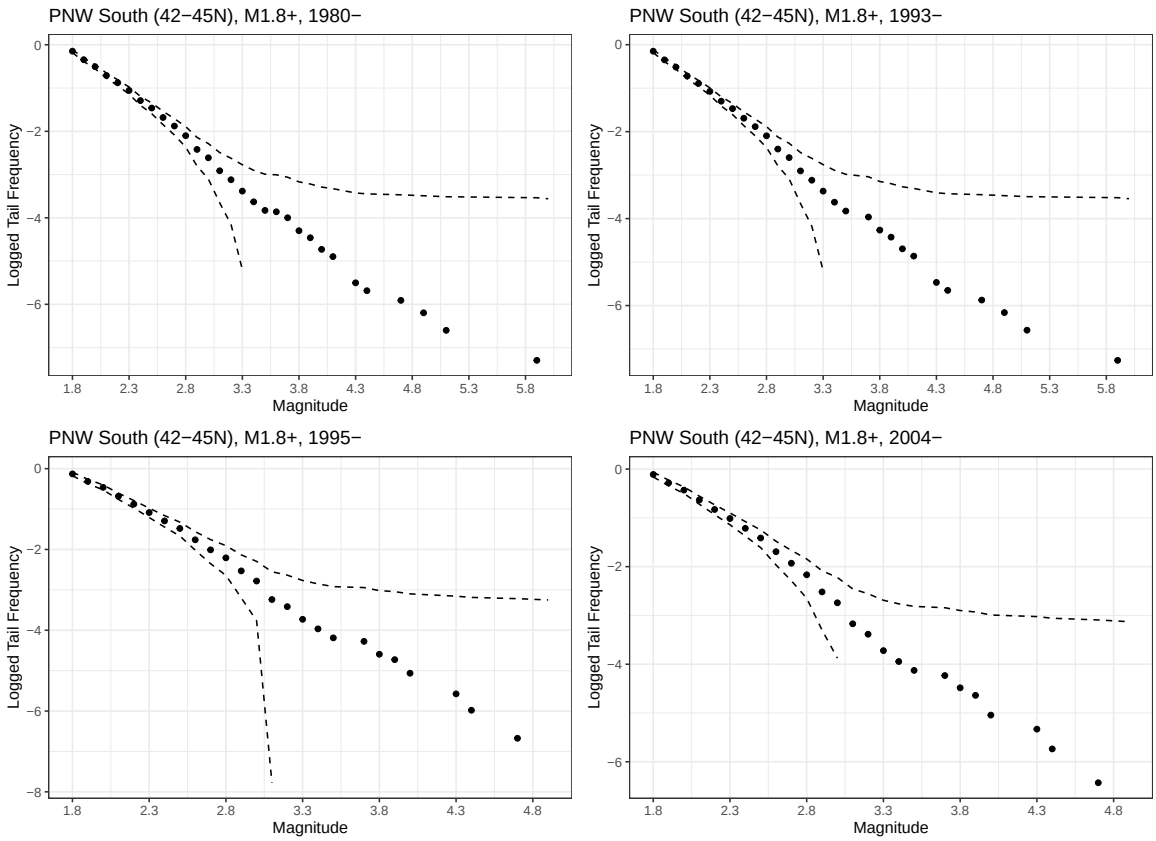

Figure A.12: Magnitude frequency distribution for the PNW South, for different start years. We plot the magnitude v. logged magnitude tail frequency and its simultaneous confidence bands.

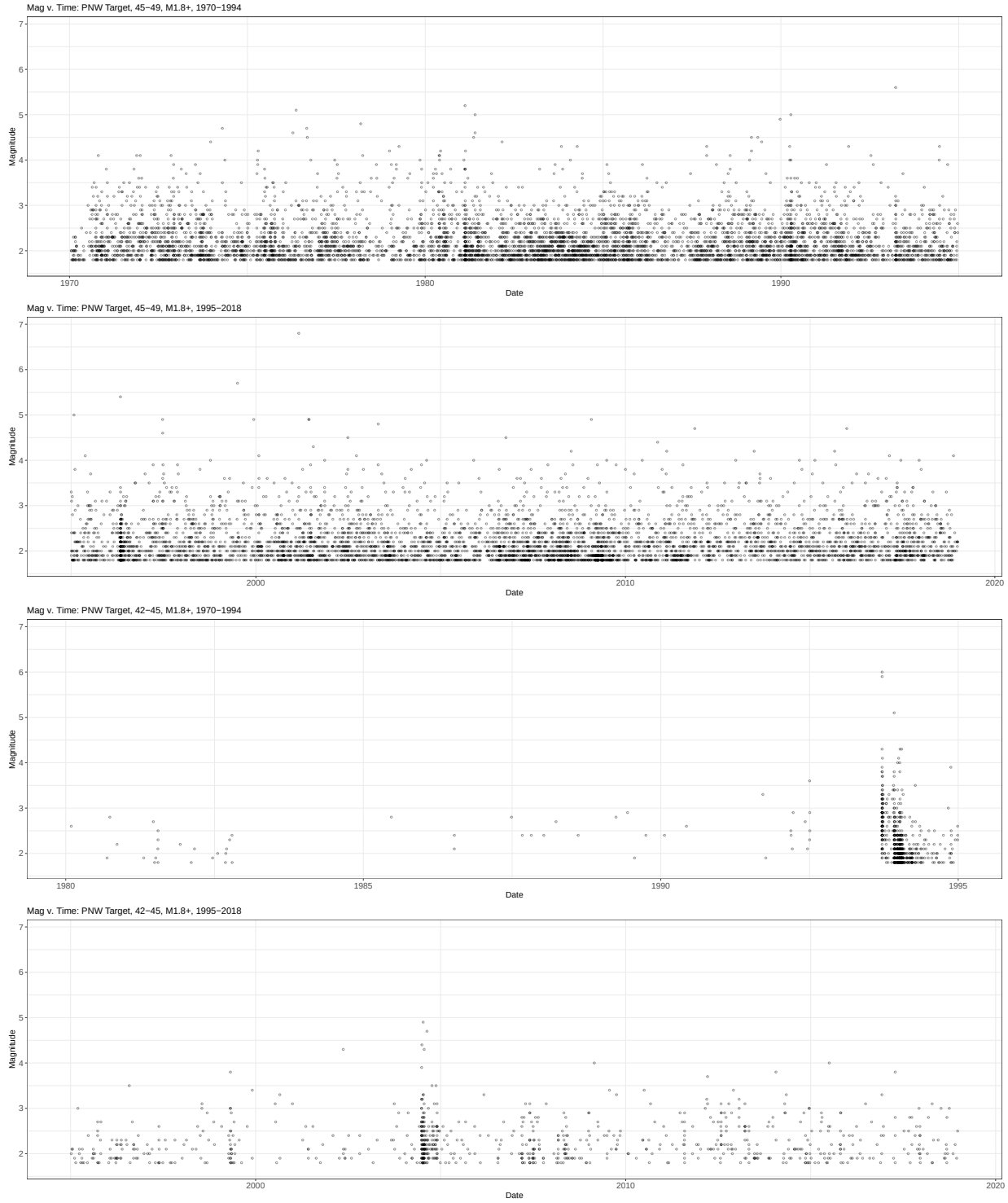

Figure A.13: Time v. magnitude for the entire catalog period for PNW North (top two panels) and PNW South (bottom two panels). No earthquakes were detected before 1980 in the PNW South which is why the plot only starts them.

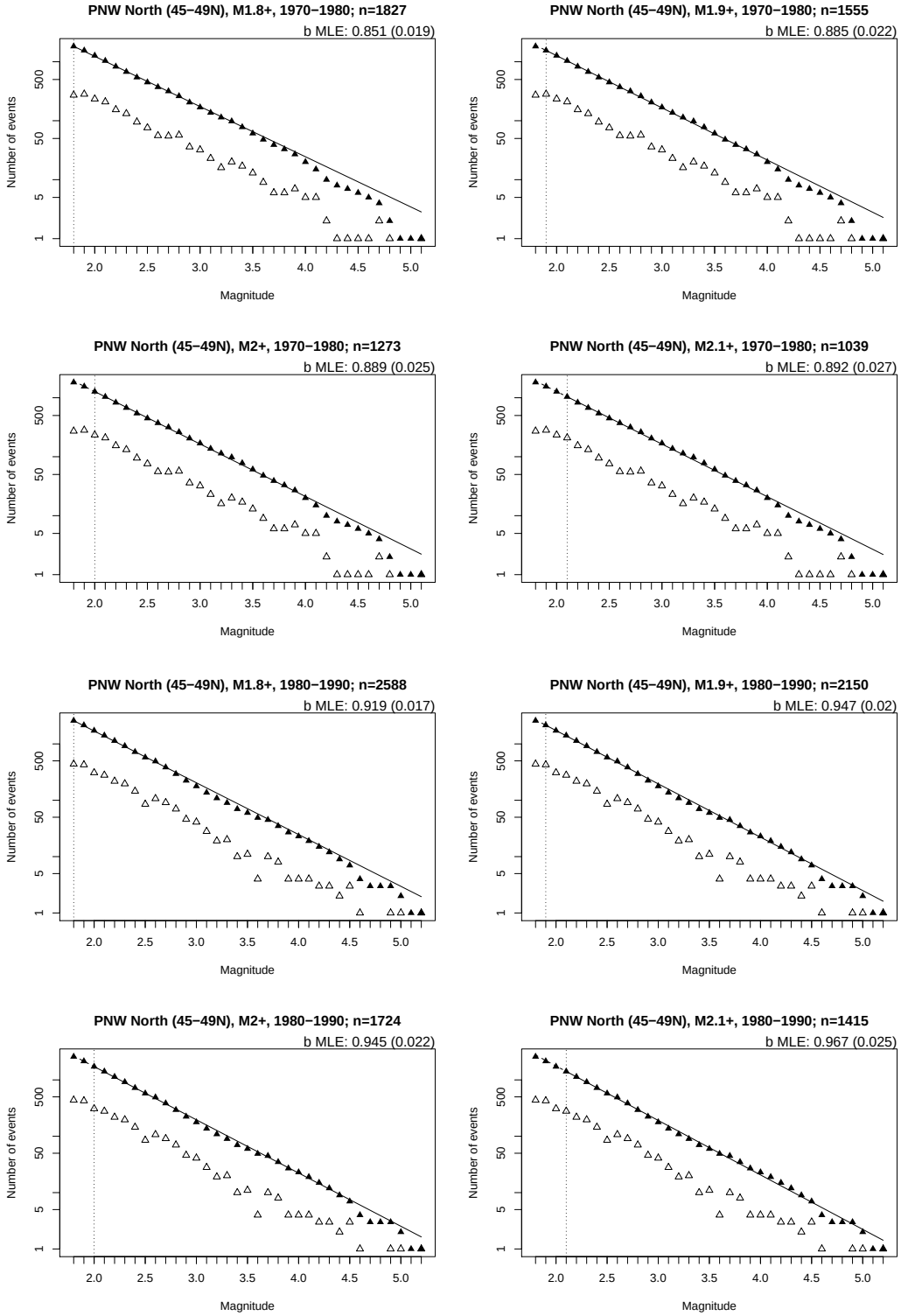

Figure A.14: Magnitude frequency distributions for the PNW North for different non-overlapping time windows and different magnitudes of completeness. The b value is estimated by MLE, with standard error of (Shi and Bolt, 1982) in parantheses.

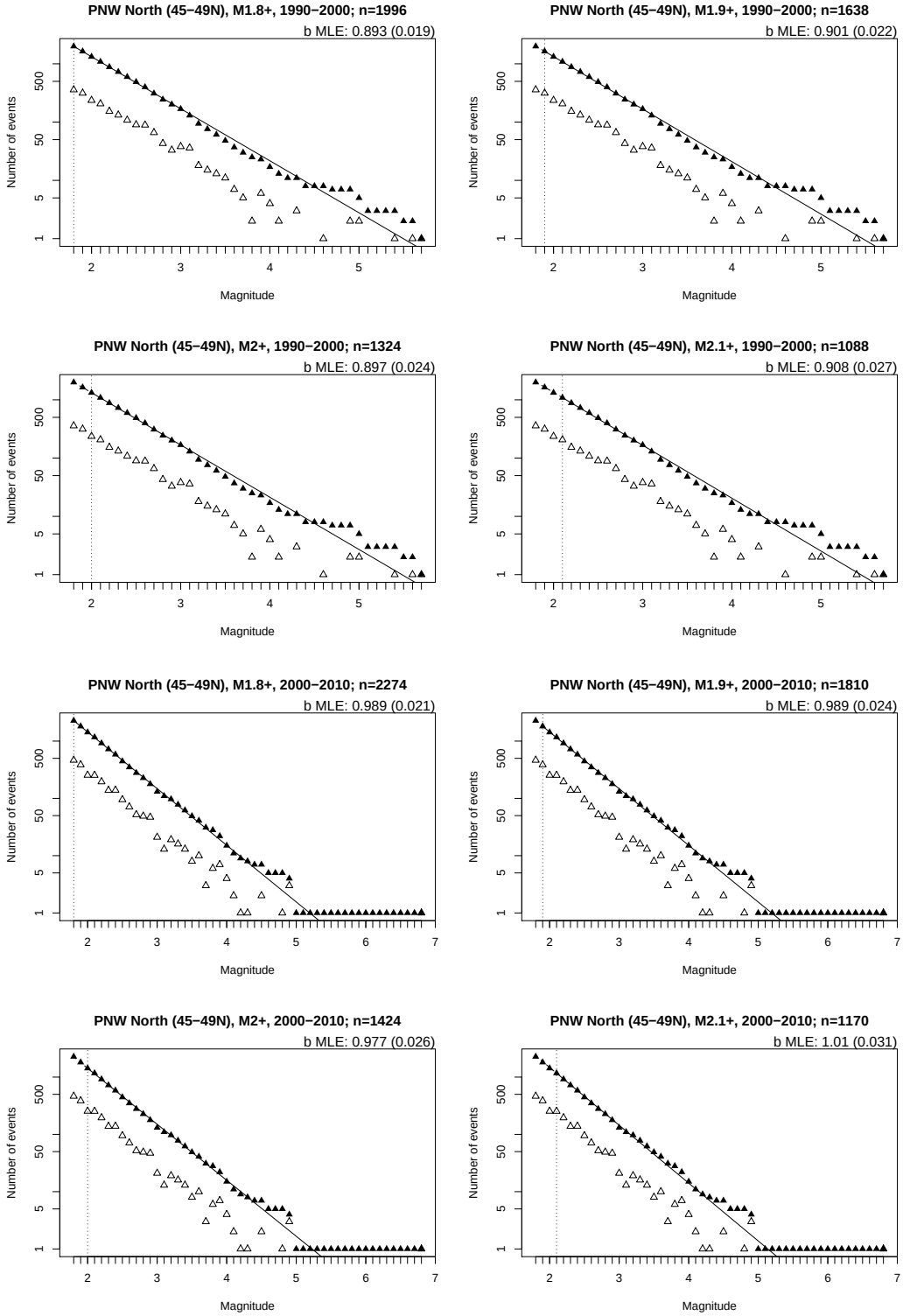

Figure A.15: Figure A.14 continued.

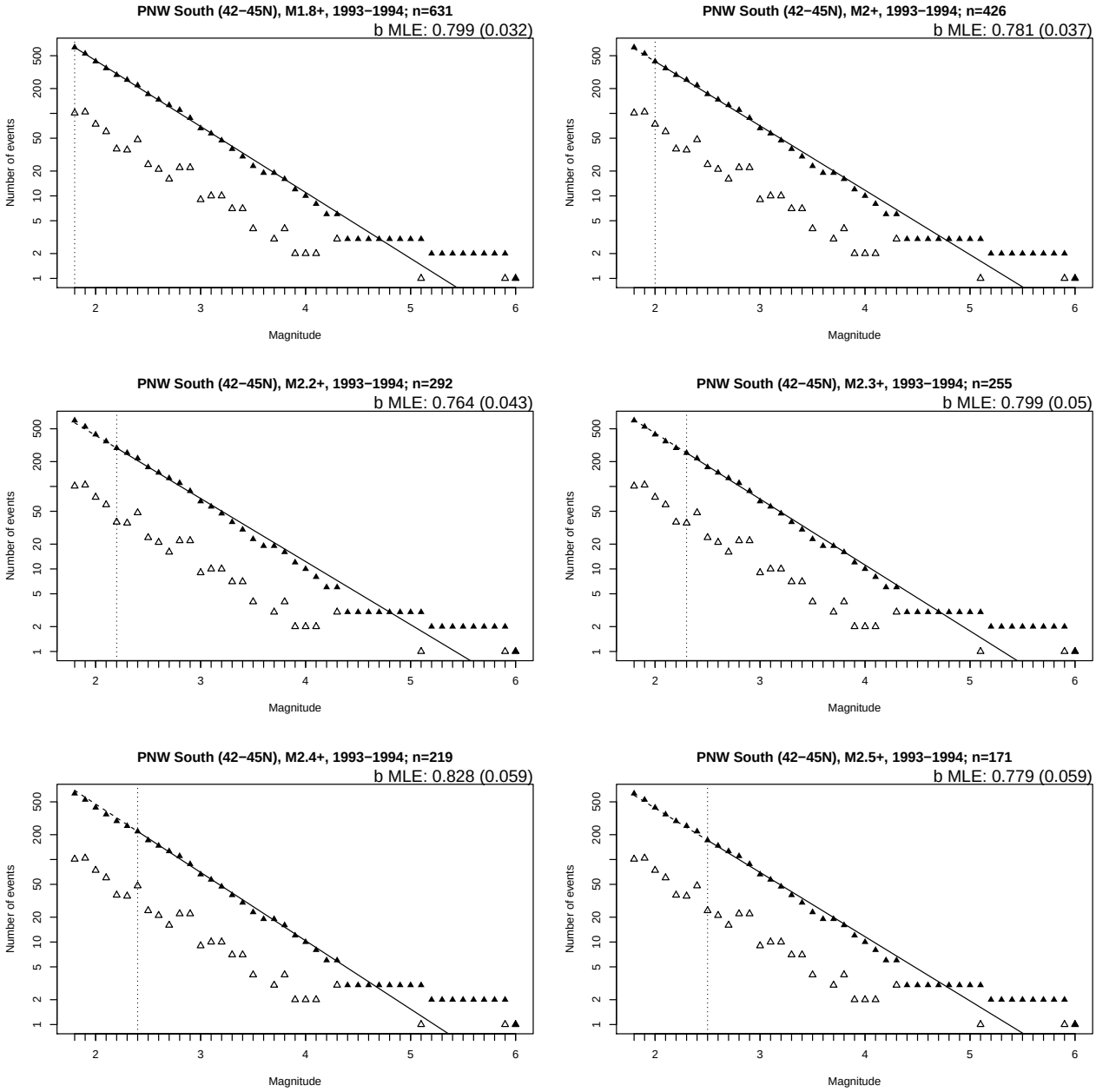

Figure A.16: Magnitude frequency distributions for the PNW South for different non-overlapping time windows and different magnitudes of completeness. The b value is estimated by MLE, with standard error of (Shi and Bolt, 1982) in parantheses.

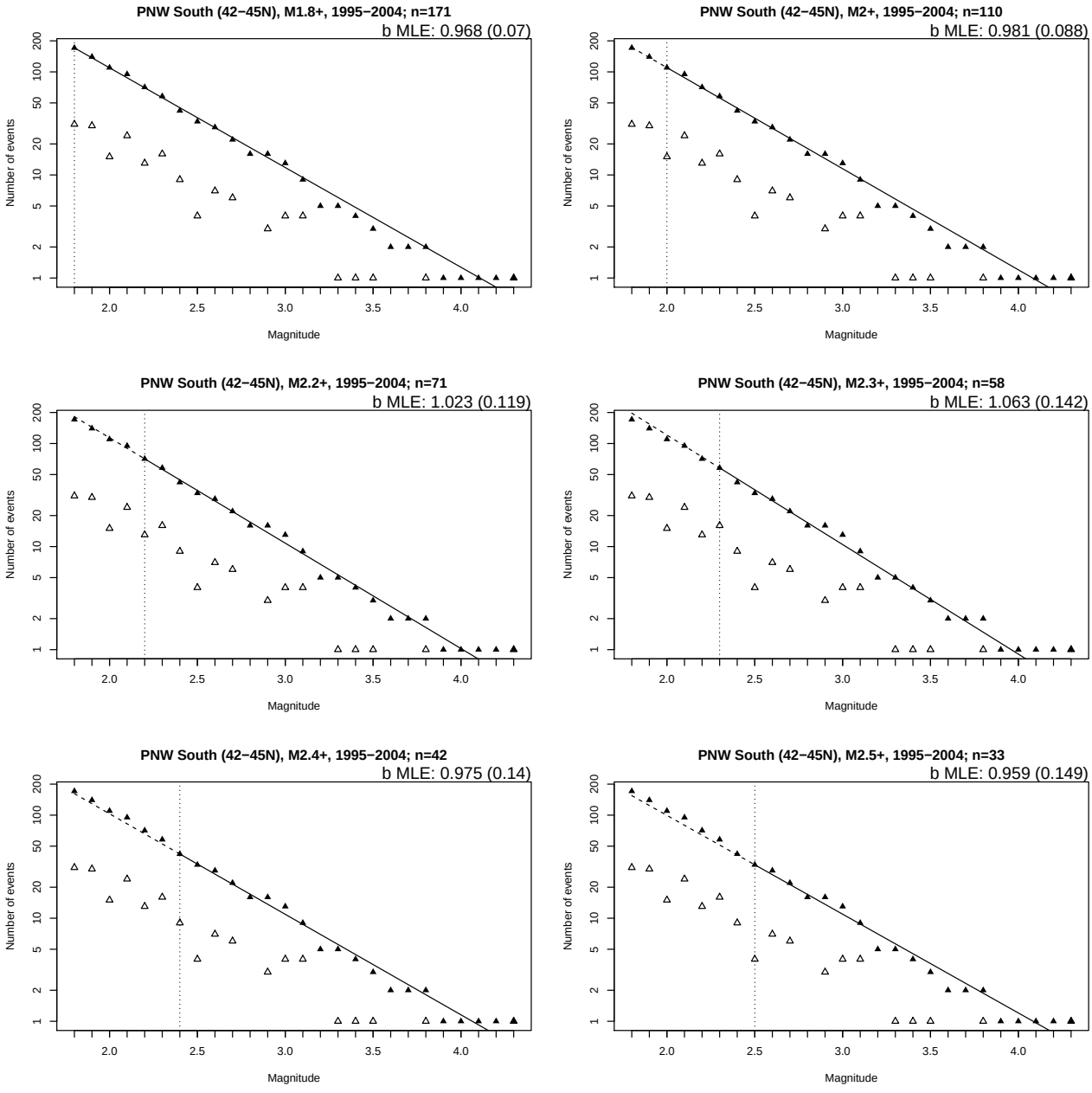

Figure A.17: Figure A.16 continued.

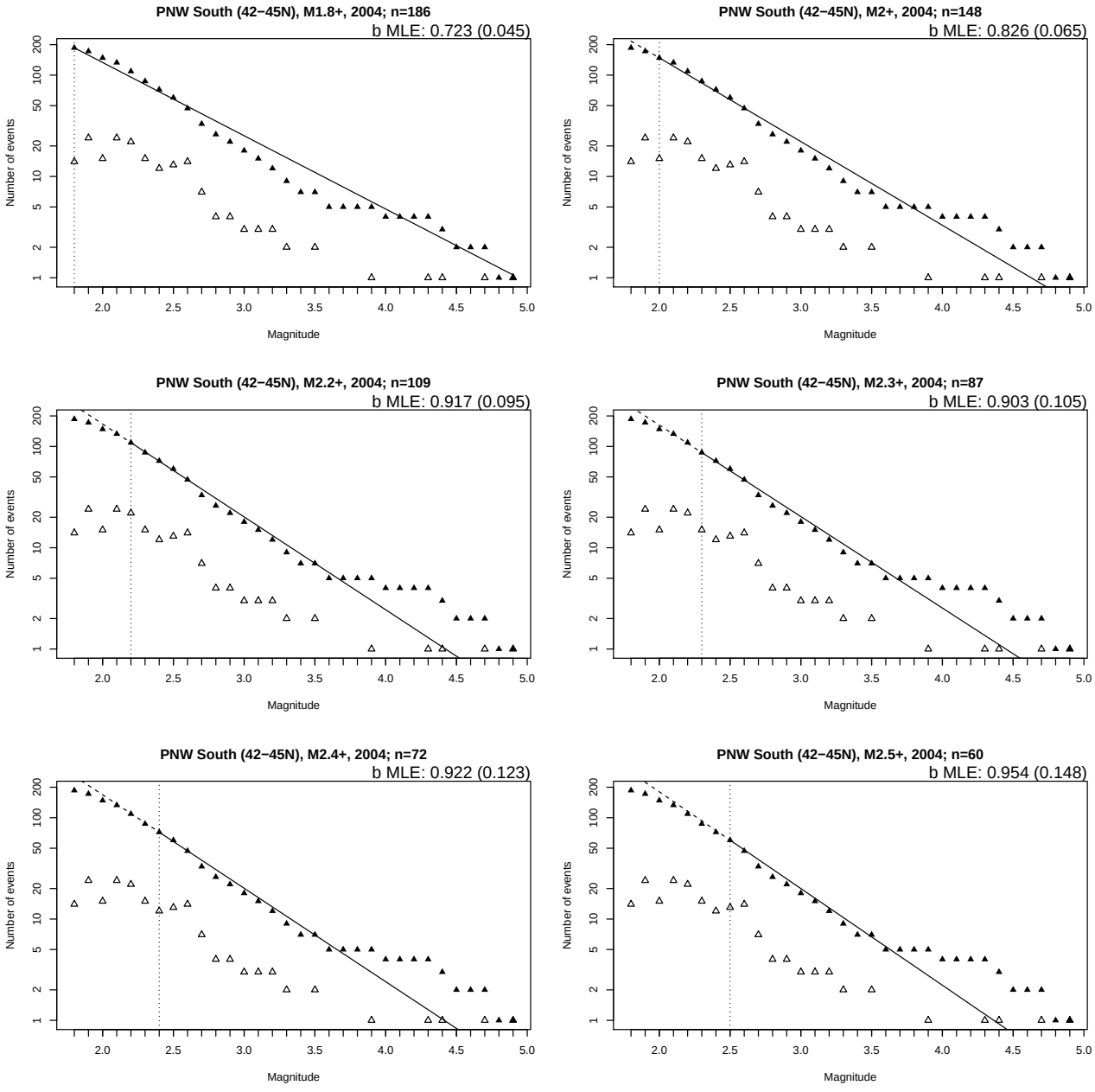

Figure A.18: Figure A.16 continued.

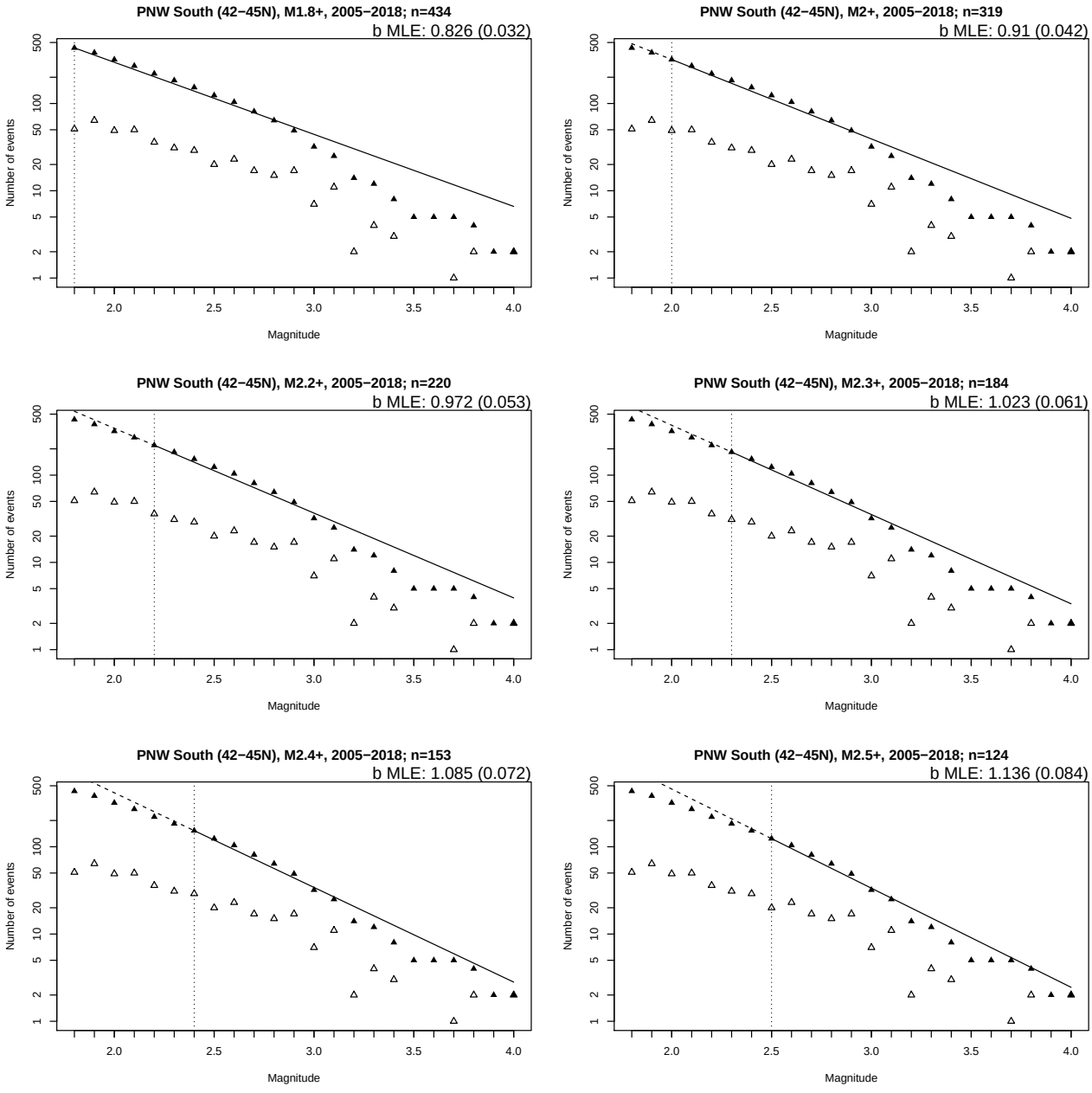

Figure A.19: Figure A.16 continued.

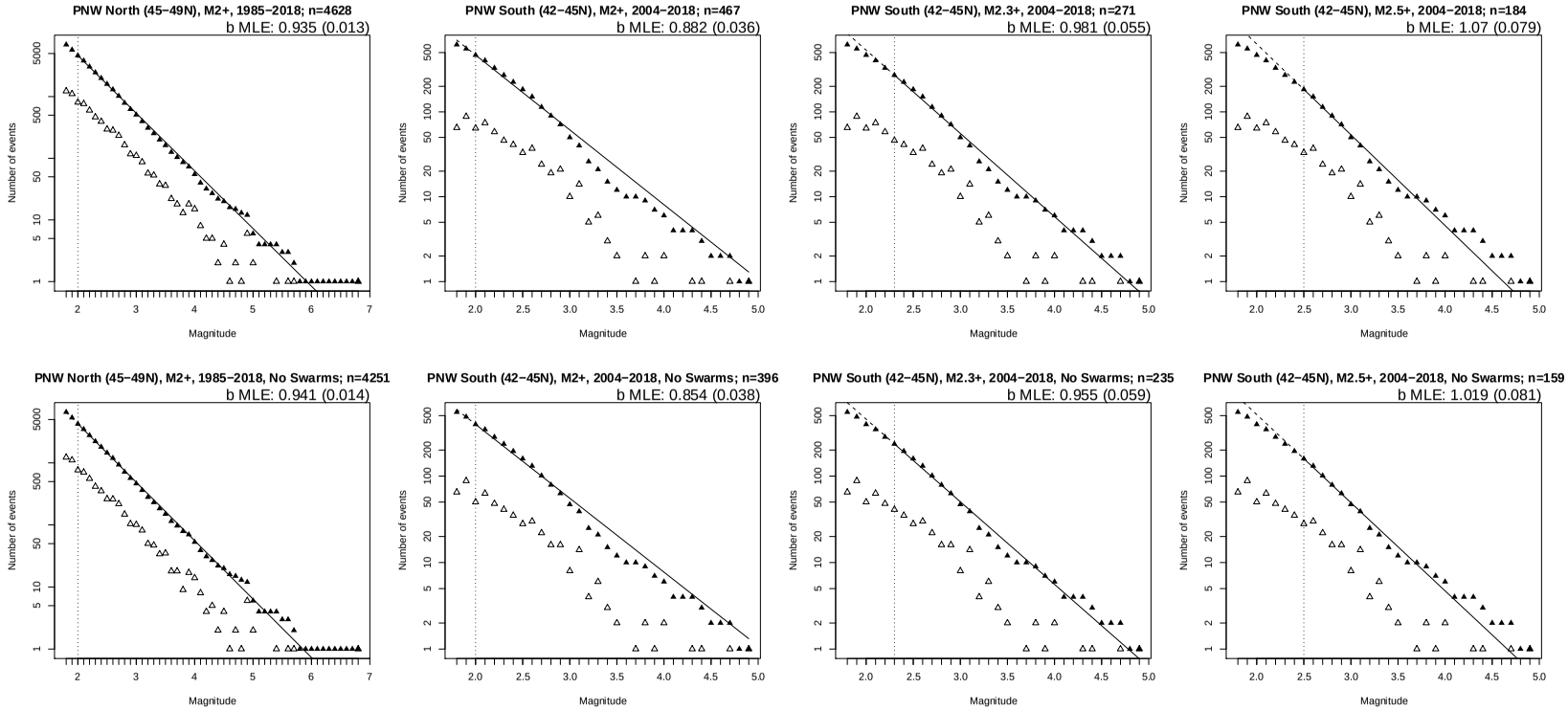

Figure A.20: Magnitude-frequency distributions for the PNW North complete catalog (first column of panels) and three PNW South catalogs, with $M_c = 2$ (second column), $M_c = 2.3$ (third column) and $M_c = 2.5$ (fourth column).

1. Harmonize each catalog's columns (variables) and rows (which earthquakes are included). This also involves removing the MSH events for every catalog.

2. Do data processing on the harmonized catalogs:

    - Find duplicates inside each catalog
    - Find mismatched data between ComCat and PNSN
    - Find duplicates between ComCat and PNSN: matching by ID, by loose match, by manual match, then reaching out to PNSN seismologists
    - Handle events outside of authoritative network (target zone) between PNSN and NRCan
        - Classify them as duplicates (remove them for the record not in authoritative network) or additional events (that the authoritative network did not have)
    - Handle events outside of authoritative network (margin zone) between ComCat and NRCan
        - Classify them as duplicates (remove them for the record not in authoritative network) or additional events (that the authoritative network did not have)

3. The above processing results in each earthquake either being removed for a particular reason or staying in the catalog. Clean the catalog by adding each earthquake's reason for removing (NA means the earthquake stays).

4. For each of the PNSN and NRCan catalogs, search for clusters.

5. Merge the cleaned catalogs between PNSN and ComCat to get a single US catalog

6. Combine (rbind) the US catalog with the NRCan catalog to get the final international catalog

A.6 Code structure

Notes:

- In all these, the ... stands in for "[remainder of the task name after first letter]/[subfolder for those tasks in older directory system]/Src/lowM" and prior to the first letter is "/Users/maxs/Dropbox/PNW-Data-Processing/"
- Maritime borders are in input/USMaritimeLimits.../USMaritimeLimitsNBoundaries.shp
- All input data comes from input/lowM subdirectory
- All output data goes to output/lowM subdirectory

Important function files:

- Task C: src/lowM/functions-for-catalog-matching.R

Code File	Description	Input Data	Outputs
/A. ... /Process-PNSN-lowM-MS.R	Read in raw PNSN-downloaded data (two files), combine them, add in inUSA variable based on US maritime borders, check for exact duplicates, remove all MSH events.	PNSN-lowM-1-raw.csv, PNSN-lowM-2-raw.csv, Maritime borders	PNSN-full-processed-lowM.csv
/B. ... /Process-ComCat-lowM-MS.R	Read in raw ComCat-downloaded data (five files), combine them, add in inUSA variable, check for exact duplicates, remove all MSH events.	ComCat-lowM-1970-1984-raw.csv-ComCat-lowM-2015-2019-raw.csv, Maritime borders	ComCat-full-processed-lowM.csv
/F. ... /Process-NRCan-lowM-MS.R	Read in raw NRCan-downloaded data (two files), combine them, add in inUSA variable, check for exact duplicates, remove all MSH events.	ComCat-lowM-1970-1984-raw.csv-ComCat-lowM-2015-2019-raw.csv, Maritime borders	NRCan-full-processed-lowM.csv
/C. ... /find-dup-PNSN-lowM-MS.R	Find loosely matched duplicates in PNSN catalog	PNSN-full-processed-lowM.csv	PNSN-wiggle-duplicates.csv
/C. ... /find-dup-ComCat-lowM-MS.R	Find loosely matched duplicates in ComCat catalog	ComCat-full-processed-lowM.csv	ComCat-wiggle-duplicates-no-depth.csv, ComCat-wiggle-duplicates.csv
/C. ... /find-dup-NRCan-lowM-MS.R	Find loosely matched duplicates in NRCan catalog	NRCan-full-processed-lowM.csv	NRCan-wiggle-duplicates.csv

Code File	Description	Input Data	Outputs
/C. ... /PNSN-ComCat- // Matching-lowM-MS	Find all PNSN events that match by ID to a partner in ComCat. Characterize the level of matching by event variables for partnered matches. For unpartnered PNSN events, perform loose matching and then manual matching to try to find a match in ComCat. Output results of matches for unpartnered events for PNSN.	PNSN-full-processed-lowM.csv, ComCat-full-processed-lowM.csv	PNSN-Com-wiggle-duplicates.csv, PNSN-Com-manual-matches-lowM.csv
/C. ... /PNSN-NRCan- // Matching-lowM-MS	Search for out-of-network NRCan events in PNSN, then ComCat (first loosely, then manually). Output manual search results for resolution by PNSN. Search for out-of-network PNSN events in NRCan (first loosely, then manually) for US and Canadian events separately; output results for resolution by NRCan.	NRCan-full-processed-lowM.csv, PNSN-full-processed-lowM.csv, ComCat-full-processed-lowM.csv	PNSN-NRCan-Can-wiggle-duplicates.csv, PNSN-NRCan-US-wiggle-duplicates.csv, Can-NRCan-PNSN-Manual-Matches-lowM.csv, Can-NRCan-PNSN-No-Manual-Match-lowM.csv, PNSN-NRCan-Can-manual-matching-lowM.csv, US-NRCan-PNSN-Manual-Matches-lowM.csv, US-NRCan-PNSN-No-Manual-Match-lowM.csv, Canadian-PNSN-Target-Unmatchable.csv, US-NRCan-target-Com-matching-lowM.csv, US-NRCan-taget-manual-Com-matches-lowM.csv, US-NRCan-Target-In-ComCat-Not-PNSN.csv, US-NRCan-Target-Unmatchable.csv
/C. ... /NRCan-ComCat- // Matching-lowM-MS	Search for out-of-network NRCan events in ComCat, then PNSN (first loosely, then manually). Output manual search results for resolution by PNSN. Search for out-of-network ComCat events in NRCan (first loosely, then manually), for US and Canadian events separately; output results for resolution by NRCan.	NRCan-full-processed-lowM.csv, PNSN-full-processed-lowM.csv, ComCat-full-processed-lowM.csv	NRCan-ComCat-margin-wiggle-duplicates.csv, US-NRCan-ComCat-Manual-Matches-lowM.csv, US-NRCan-ComCat-No-Manual-Match-lowM.csv, Can-NRCan-ComCat-Manual-Matches-lowM.csv, Can-NRCan-ComCat-No-Manual-Match-lowM.csv, Canadian-ComCat-Margin-Unmatchable.csv, Can-NRCan-margin-manual-PNSN-matches-lowM.csv, US-NRCan-Margin-Unmatchable.csv, US-NRCan-Margin-In-PNSN-Not-ComCat.csv

Code File	Description	Input Data	Outputs
/A. ... /Clean-PNSN-lowM-MS.R	Flag PNSN events found in task C to be loose matches to a Canadian NRCan target event; to a Canadian NRCan event that is out of bounds; or that Amy told us is related to a meteor.	PNSN-full-processed-lowM.csv, PNSN-NRCan-Can-wiggle-duplicates.csv, Canadian-PNSN-Target-Unmatchable-20200706.cb-why.hf.csv, PNSN-Com-manual-matches-lowM-MS-AW-20210120.csv	PNSN-full-with-remove-lowM.csv
/B. ... /Clean-ComCat-lowM-MS.R	First, give all ComCat events that are loose or manual matches (and confirmed by Amy) to the PNSN (that is, not matches by ID), the ID of their PNSN partner for subsequent merging. Flag all duplicates (by ID and matching). Then clean the ComCat margin catalog based on task C results. Flag Canadian ComCat margin events that were matched to a Canadian NRCan margin event (loosely or manually); that Camille told us was a blast; or that Camille flagged as in NRCan and we could actually find in our NRCan catalog (only 3 such events); or that was a duplicate.	ComCat-full-processed-lowM.csv, PNSN-Com-wiggle-duplicates.csv, PNSN-Com-manual-matches-lowM-MS-AW-20210120.csv, NRCan-ComCat-margin-wiggle-duplicates.csv, Can-NRCan-ComCat-Manual-Matches-lowM.csv, Canadian-ComCat-Margin-Unmatchable-20200706.cb.csv, Canadian-ComCat-Margin-Unmatchable-20200706.cb.HF.csv, ComCat-duplicates-20200706-AW.csv	ComCat-full-with-remove-lowM.csv
/F. ... /Clean-NRCan-lowM-MS.R	Flag NRCan events found in task C to be duplicates; loose or manual matches to a US PNSN target event; were previously found to be matches to a US PNSN target event; loose or manual matches to a US ComCat margin event; implement Amy's feedback on unmatchable NRCan events in the target and margin zones, not flagging events that were set to remain in the final catalog; apply the additional analysis where Hank found 2 NRCan margin events in ComCat but there were four that were unfound (so they should remain in the catalog); apply Amy's feedback for those NRCan target events found in ComCat but not PNSN; apply two miscellaneous corrections. Finally, add a variable isTarget based on target spatiomagnitude restrictions.	NRCan-full-processed-lowM.csv, PNSN-NRCan-US-wiggle-duplicates.csv, US-NRCan-PNSN-Manual-Matches-lowM.csv, US-NRCan-PNSN-No-Manual-Match-lowM.csv, US-NRCan-target-Com-matching-lowM.csv, US-NRCan-taget-manual-Com-matches-lowM.csv, NRCan-ComCat-margin-wiggle-duplicates.csv, US-NRCan-ComCat-Manual-Matches-lowM.csv, US-NRCan-Target-Unmatchable-20200706-AW.csv, US-NRCan-Margin-Unmatchable-20200706-AW.csv, US-NRCan-ComCat-Amy-Matched-Final-6.HF.csv, US-NRCan-Target-In-ComCat-Not-PNSN-20200723-AW.csv	NRCan-full-with-remove-lowM.csv

Code File	Description	Input Data	Outputs
/D. ... /identify-swarms- // lowM-MS.R	Use the space/time/depth criteria to identify clusters in the PNSN catalog (target only). Remove obvious spatial outliers. (This is steps 1-5 of the process in Section 4.1.1)	PNSN-full-processed-lowM.csv	PNSN-swarm-ids.csv
/D. ... /identify- // archived-swarms.R	Use criteria given in Section 4.X to identify documented swarms in the PNSN and NRCan catalogs.	NRCan-Target-with-Remove.csv, PNSN-Full-with-Remove.csv,	Archived-Swarms-lowM.csv
/D. ... /classify- // cluster-types.R	Classify the identified clusters into four categories based on rules in Section 4.1.2. Read out summaries of clusters by categories.	PNSN-swarm-ids.csv, swarm-decisions-20191121- // RH-MS-20201110.csv	cluster-single-conc-10km-decisions-20201202.csv, pnw-catalog-20201202.csv, catalog-clusters-single-conc-10km-20201202.csv, catalog-clusters-single-conc-20210826.csv, same but for "multiple/multi-conc-20210130/0826", "diffuse-20210130/0826", "short-20210130/0826" pnsn-with-clusterIDs-20210306.csv
/D. ... /process-PNSN- // swarms-sequences.R	Read in Amy's feedback on clusters, by type. Remove outliers / split multiple clusters using spatial rules and relabel clusters as necessary. Merge this data with PNSN catalog by cluster ID.	pnsn-with-clusterIDs-20210306.csv, catalog-clusters-single-conc-20210826.csv, catalog-clusters-diffuse-20210826.csv, catalog-clusters-short-20210826.csv, catalog-clusters-multi-conc-20210826.csv, catalog-notclustered-20210826.csv, cluster-single-conc-10km-decisions-20201202-AW-updates.csv, cluster-diffuse-decisions-20210130-AW.csv, catalog-clusters-diffuse-20210130.csv, cluster-short-decisions-20210122-AW.csv, cluster-multiple-conc10km-decisions-20210130-AW.csv	pnsn-clusters-classified.csv

Code File	Description	Input Data	Outputs
/E. ... /merge-PNSN-ComCat- // lowM-MS.R	Read in cleaned PNSN and ComCat data. Check that the number of events without a remove value is as expected. Merge all PNSN and (target-only) ComCat data by ID. Harmonize the variable names with the ComCat catalog and bind in the margin ComCat catalog.	ComCat-full-with-remove-lowM.csv, ComCat-full-with-remove-lowM.csv	PNSN-ComCat-aux.csv
/E. ... /merge-US-NRCan- // lowM-MS.R	Read in cleaned NRCan catalog and merged PNSN-ComCat (US) catalog. Check that the number of NRCan events without a remove value is as expected. Harmonize NRCan variable names with the (US) catalog and bind all NRCan events with the US catalog.	PNSN-ComCat-aux.csv, NRCan-full-with-remove-lowM.csv	pnw-intl-aux.csv
/E. ... /merge-intl- // clusters.R	Read in cleaned intl catalog and (PNSN) catalog subsets with labelled clusters (swarms/sequences). Merge the intl catalog with the algorithmically-found and seismologist-classified clusters data. Then flag events in all documented swarms as being in confirmed swarms.	pnw-intl-aux.csv, pnsn-clusters-classified.csv, Archived-Swarms-lowM.csv	pnw-intl-aux-clusters.csv
/G. catalog-visualization/ // ... /visualize-pnw-cat-lowM.R	Vis	pnw-intl-aux-clusters.csv	

Below are the remove flags in the final data.
For PNSN:

- NRCanAlgo
- OutOfBounds, meaning it is a duplicate of an NRCan earthquake that is outside the spatiomagnitude auxiliary zone
- PNSNBlast

For ComCat:

- NRCanAlgo
- NRCanManual
- OutOfBoundsinCan, meaning it was in Canada but it had magnitude $<$ M1.8 and thus is out of bounds for the margin
- NRCanBlast

For NRCan:

- PNSNAlgo, ComAlgo
- PNSNManual, ComManual
- PrevProcessed, meaning it was previously flagged by Amy for removal in M2.5+ catalog processing (see file NRCan-Unmatched-20200416-AW.csv which got used to create US-NRCan-PNSN-No-Manual-Match-lowM.csv). This will get changed to PNSNManual.
- PNSNBlast, ComBlast
- OutOfBounds, meaning it is a duplicate of a PNSN/ComCat earthquake that is outside the spatiomagnitude auxiliary zone

Appendix B

APPENDIX TO CHAPTER 3

B.1 Additional material related to evaluating the spatial integral

B.1.1 Radially approximating the spatial integral

The spatial component of INT_{trig} (spatial integral), see Section 3.2.2 is particularly complicated and is not directly evaluable. Several authors (Ogata, 1998; Jalilian, 2019; Chiodi and Adelfio, 2021) advocate to handle the spatial integral by radially approximating it across the catalog zone.

For each mainshock j, partition the region $S^{(j)}$ into N subregions: $S_1^{(j)}, S_2^{(j)}, \ldots, S_N^{(j)}$. Do this by setting N knots along the boundary of $S^{(j)}$ and drawing a segment to connect the earthquake's location (the origin of $S^{(j)}$) to each knot. That is, we split $S^{(j)}$ into a set of N radial wedges (triangles) from the earthquake j in the center. See Figure B.1 for a visual example from Jalilian (2019). Each of these triangles is defined radially by its radial segment r_k and azimuth angle θ_k ($k = 1, \ldots, N$), and we order the N segments by θ, such that $0 < \theta_1 < \theta_2 < \ldots < \theta_N < 2\pi$.

Then, using Ogata (1998),

$$\int\int_{S^{(j)}} \frac{q-1}{\pi d^{1-q}} \frac{1}{(x^2+y^2+d)^q} dxdy \approx \sum_{k=1}^{N} G_k^{(i)}(x_i, y_i) \frac{\Delta_k}{2\pi},$$

where

$$G_k^{(i)} = \frac{\pi}{1-q}((r_k^2+d)^{1-q} - d^{1-q})$$

$$\Delta_k = \theta_{k+1} - \theta_k.$$

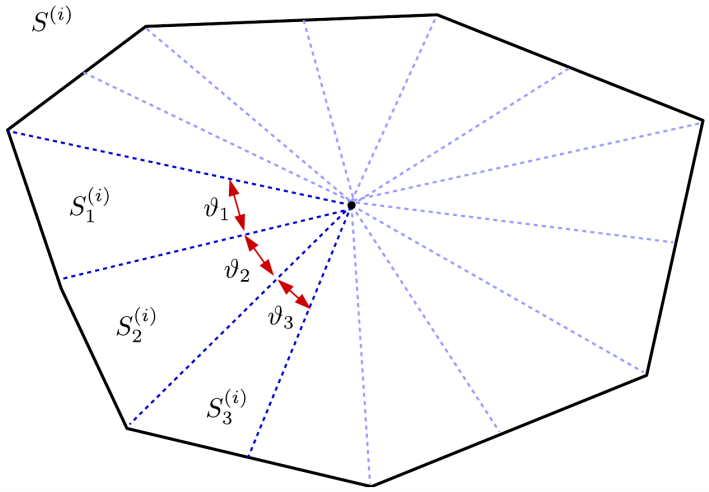

Figure B.1: Jalilian (2019), Figure 1. Note that here, θ refers to the triangle's central angle but in our (and Ogata (1998)'s) notation, it is the azimuth angle.

The first term in the sum evaluates the Utsu-Seki integral along the radial segment r_k and the second term scales it by a measure of the angular area of the wedge. This is the proportion of the circle (2π) corresponding to the difference between this central angle and the neighboring central angle. Ogata (1998) fails to specify the case of $k = N$, where $\Delta_N = \theta_{N+1} - \theta_N$ does not exist. We simply set $\Delta_N = \theta_N - \theta_{N-1}$.

The number of triangles N must be "sufficiently large for the accurate numerical approximation of the integral" (Ogata, 1998). It is unclear how one chooses N other than by trial and error (Schoenberg, 2013).

B.1.2 Direct evaluation of the spatial integral

The spatial integral $\int\int_{S^{(j)}} \frac{1}{(x^2+y^2+d)^q} dxdy$ may also be computed directly using numerical integral solvers in R. This would avoid some of the arbitrary choices that the radial approximation requires (e.g., setting N, the number of knots).

There are many approaches available for multidimensional integration. The R package cubature offers both deterministic and Monte Carlo (MC) direct integration. We examined two approaches:

- **hcubature**, which implements an h-adaptive cubature rule (McFee, 1997). A cubature rule is a quadrature rule (one-dimensional integral evaluation procedure) applied to a multidimensional space. The h-adaptive version recursively partitions the integration area into smaller subareas, applying the same cubature rule to each, until convergence is achieved within each subarea. Eventually, the integral over the entire area converges.

- **vegas**, which implements a Monte Carlo-based integration (Lepage, 1980). The integral is computed at random points from a probability dist'n related to the integrand, drawn using importance sampling. The integral points are summed to estimate the full function integral.

We evaluated the spatial integral for all earthquakes in synthetic catalogs 3A and 3B (see Section 3.5) under both hcubature and vegas; they yielded very similar values.

B.1.3 Tables and figures for d/q grids for evaluating the spatial integral

In order to save computation in evaluating the ETAS branching likelihood as parameter posteriors are sampled, we pre-compute G, the spatial integral, for each earthquake, under a grid of d and q values given in Table B.1. The surfaces of G values for two earthquakes in a synthetic catalog are given in Figure B.2.

B.2 Derivation of branching likelihood

We can directly derive the branching likelihood of Ross (2021) from Ogata (1998)'s likelihood function:

Catalog	true d	d sequence	true q	q sequence
A	1	from=0.3, to=1.75, by=0.05	2	from=1.25, to=2.75, by=0.05
B	1	from=0.3, to=1.75, by=0.05	1.5	from=1.05, to=2.25, by=0.05
C	0.1	from=0.025, to=0.175, by=0.005	2	from=1.25, to=2.75, by=0.05
D	0.1	from=0.025, to=0.175, by=0.005	1.5	from=1.05, to=2.25, by=0.05
PNW	-	from=0.3, to=2.0, by=0.08	-	from=1.2, to=1.6, by=0.04

Table B.1: d/q grids for pre-computed G evaluations.

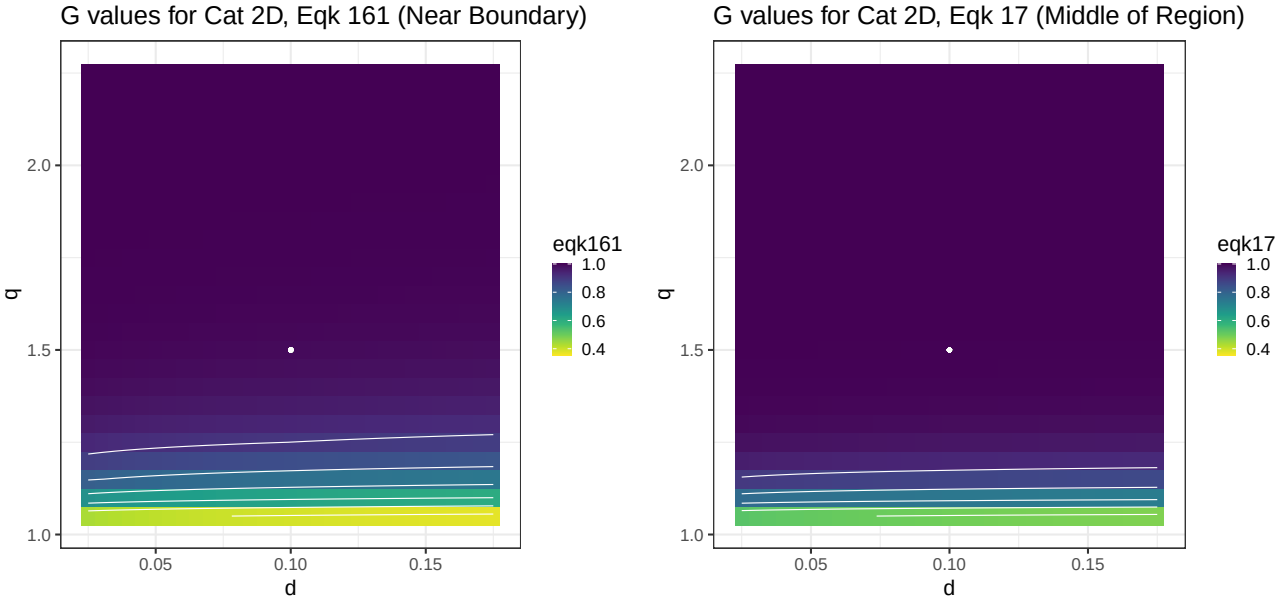

Figure B.2: G surfaces for varying d and q around the true values for Catalog 2D. The integral values for an earthquake near the border (left) are smaller than the values for an earthquake near the center of the region (right). The values increase for increasing q, but the rate of increase gets smaller for q values above the truth. White contours show increase in 0.1 steps of integral value.

$$
\begin{aligned}
L(Y|\theta) &= \prod_{i=1}^{n} \lambda(t_i, x_i, y_i, M_i|\mathcal{H}_t, \theta) \exp(-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} \lambda(t, x, y, M|\theta, \mathcal{H}_t) dt dx dy) \\
&= \prod_{i=1}^{n} \Big(\mu_{ST} + \sum_{j:t_j<t} k(M_j)h(t_j)g(x - x_j, y - y_j)\Big) \\
&\exp\Big(-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} \mu_{ST} + \sum_{j:t_j<t} k(M_j)h(t_j)g(x - x_j, y - y_j) dt dx dy\Big)
\end{aligned}
$$

Say we have $|A_0|$ background and $n - |A_0|$ triggered events in the catalog of size n. The background events are generated by a homogenuous Poisson process with intensity $\lambda(t|\mathcal{H}_t) = \mu_{ST}$. So,

$$
\begin{aligned}
L(Y|\theta, A_0) &= \prod_{i=1}^{|A_0|} \mu_{ST} \exp\Big(-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} \mu_{ST} dt dx dy\Big) \\
&= \prod_{i=1}^{|A_0|} \mu_{ST} \exp(-\mu_{ST} T|S|) \\
&= \exp(-\mu_{ST} T|S|) \prod_{i=1}^{|A_0|} \mu_{ST} \\
&= \exp(-\mu_{ST} T|S|) \mu_{ST}^{|A_0|},
\end{aligned}
$$

where $|S|$ is the area of the spatial region being modelled.

Now each of the events triggered by earthquake j (those in set A_j) are generated by a time-inhomogenuous Poisson process with $\lambda(t|\mathcal{H}_t) = k(M_j)h(t - t_j)g(x - x_j, y - y_j)$. The contribution to the likelihood function for each aftershock sequence is:

$$
\begin{aligned}
L(Y|\theta, A_j) &= \prod_{i=1}^{|A_j|} k(M_j) h(t_i - t_j) g(x_i - x_j, y_i - y_j) \\
&\qquad \exp\Big(- \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} k(M_j) h(t_j) g(x - x_j, y - y_j) dt dx dy \Big) \\
&= k(M_j)^{|A_j|} \prod_{i=1}^{|A_j|} h(t_i - t_j) g(x_i - x_j, y_i - y_j) \\
&\qquad \exp\Big(- k(M_j) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} h(t - t_j) g(x - x_j, y - y_j) dt dx dy \Big) \\
&= k(M_j)^{|A_j|} \prod_{i=1}^{|A_j|} h(t_i - t_j) g(x_i - x_j, y_i - y_j) \\
&\qquad \exp\Big(- k(M_j) \int_{0}^{\infty} h(t - t_j) dt \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x - x_j, y - y_j) dx dy \Big). \\
&= k(M_j)^{|A_j|} \prod_{i=1}^{|A_j|} h(t_i - t_j) g(x_i - x_j, y_i - y_j) \\
&\qquad \exp\Big(- k(M_j) H(T - t_j) G(x - x_j, y - y_j) \Big).
\end{aligned}
$$

As described in Section 3.2.2, the spatial integral $G(x - x_j, y - y_j)$ is evaluated using our analytical+numerical solution.

Now we combine the background and each aftershock contribution to the likelihood:

$$
\begin{aligned}
L(\theta|Z,B) &= L(Y|\theta,A_0)\cdot L(Y|\theta,A_1)\cdot L(Y|\theta,A_2)\cdot\ldots\cdot L(Y|\theta,A_n)\\
&= L(Y|\theta,A_0)\prod_{j=1}^{n} L(Y|\theta,A_j)\\
&= \exp(-\mu_{ST}T|S|)\mu_{ST}^{|A_0|}\prod_{j=1}^{n}\Big(\exp\Big((-k(M_j)H(T-t_j)G(x-x_j,y-y_j)\Big)k(M_j)^{|A_j|}\\
&\prod_{t_i\in A_j} h(t_i-t_j)g(x_i-x_j,y_i-y_j)\Big)\\
&=\Big[\mu_{ST}^{|A_0|}\prod_{j=1}^{n}k(M_j)^{|A_j|}\prod_{t_i\in A_j}h(t_i-t_j)g(x_i-x_j,y_i-y_j)\Big]\cdot\\
&\Big[\exp(-\mu_{ST}T|S|)\prod_{j=1}^{n}\exp\Big((-k(M_j)H(T-t_j)G(x-x_j,y-y_j)\Big)\Big]
\end{aligned}
$$

B.3 Derivations of log posterior density functions

Recall that posteriors for a given parameter pair are conditional on fixed values for the other ETAS parameters. In other words, for example, $p(K,\alpha) = p(K,\alpha|\mu,c,p,d,q)$. We remove this "conditional" part for notational ease.

$$
\begin{aligned}
\log(p(K,\alpha)) &\propto \log(\pi(K,\alpha)) + l(Y|K,\alpha,B)\\
&= 0 + (\sum_{j=1}^{n}|A_j|\log(k(M_j))) + INT_{trig}.\\
&= \Big(\sum_{j=1}^{n}|A_j|\log(K\exp(\alpha(M_j-M_0)))\Big) + INT_{trig}.
\end{aligned}
$$

$$\begin{aligned}\log(p(c,p)) &\propto \log(\pi(c,p)) + l(Y|c,p,B) \\ &= 0 + \sum_{j=1}^{n} \sum_{t_i \in A_j} \log(h(t_i - t_j)) + INT_{trig} \\ &= \Big(\sum_{j=1}^{n} \sum_{t_i \in A_j} -p(\log(t_i - t_j) + c) \Big) + INT_{trig}.\end{aligned}$$

$$\begin{aligned}\log(p(d,q)) &\propto \log(\pi(d,q)) + l(Y|d,q,B) \\ &= 0 + \sum_{j=1}^{n} \sum_{t_i \in A_j} \log(g(x_i - x_j, y_i - y_j)) + INT_{trig} \\ &= \Big(\sum_{j=1}^{n} \sum_{t_i \in A_j} \log(q-1) - \log(\pi) + (q-1)\log(d) + (-q(\log((x_i - x_j)^2 + \\ &(y_i - y_j)^2 + d)) \Big) + INT_{trig}.\end{aligned}$$

B.4 Equations used in prior elicitation

To derive t^*, the time at which the temporal rate of aftershocks equals the temporal background rate, we set these two rates equal. The aftershock rate is the expected number of aftershocks: $\int_0^T (t+c)^{-p} dt$, from the time of the mainshock ($t = 0$) to the catalog length ($t = T$). We find t^* by solving the expression for t.

$$\mu_{ST} S = \frac{(1-p)}{(T+c)^{1-p} - c^{1-p}} (t^* + c)^{-p} (K \exp(\alpha(M - M_0)))$$
$$\frac{S\mu_{ST}}{K \exp(\alpha(M - M_0))} \frac{(T+c)^{1-p} - c^{1-p}}{(1-p)} = (t^* + c)^{-p}$$
$$t^* = \Big(\frac{\mu}{K \exp(\alpha(M - M_0))} \frac{(T+c)^{1-p} - c^{1-p}}{(1-p)} \Big)^{-\frac{1}{p}} - c$$

We find $dist^*$, the distance at which the spatial rate of aftershock equals the spatial background rate, in an analogous way:

$$dist^* = \Big(\mu(\frac{\pi d^{1-q}}{q-1}) \frac{1}{K \exp(\alpha(M - M_0))} \Big)^{-\frac{1}{q}} - d$$

B.5 Reported ETAS parameters for subduction zones

See Table B.2.

B.6 Derivation of transformed times

Consider event times $t_i, i = 1, \ldots, n$. We define the *transformed time* τ_i as the integral of the fitted spatiotemporal ETAS conditional intensity λ from time 0 to t_i, over all space.

$$\tau_i = \int_0^{t_i} \int_S \hat{\mu}_{ST} + \sum_j^{t_j < t} \hat{k}(M_j) \hat{h}(t - t_j) \hat{g}(x - x_j, y - y_j) ds dt,$$

where $k(M_j) = K \exp(\alpha(M_j - M_0))$;

we use the **unnormalized** Omori temporal decay law: $h(t - t_j) = (t - t_j + c)^{-p}$;

and the **normalized** Utsu-Seki spatial decay law: $g(x - x_j, y - y_j) = \frac{q-1}{\pi d^{1-q}}((x - x_j)^2 - (y - y_j)^2 + d)^{-q}$;

and the hats mean that we use parameter estimates from an ETAS model fit to a catalog. The two parts of the formula for τ_i can be split up:

$$\begin{aligned}\tau_i &= \int_0^{t_i}\int_S \hat{\mu}_{ST} ds dt + \int_0^{t_i}\int_S \sum_j^{t_j<t} \hat{k}(M_j)\hat{h}(t-t_j)\hat{g}(x-x_j, y-y_j) ds dt \\ &= \tau_i^{bkgd} + \tau_i^{trig}.\end{aligned}$$

We first derive τ_i^{trig}:

$$\begin{aligned}\tau_i^{trig} &= \int_0^{t_i}\int_S \sum_j^{t_j<t} \hat{k}(M_j)\hat{h}(t-t_j)\hat{g}(x-x_j, y-y_j) ds dt \\ &= \sum_j^{t_j<t} \int_{t_j}^{t_i}\int_S \hat{k}(M_j)\hat{h}(t-t_j)\hat{g}(x-x_j, y-y_j) ds dt \\ &= \sum_j^{t_j<t} \hat{k}(M_j) \int_{t_j}^{t_i} \hat{h}(t-t_j) dt \int_S \hat{g}(x-x_j, y-y_j) ds.\end{aligned}$$

Let $\hat{G}_i = \sum_j^{t_j<t_i} \int_{ymin}^{ymax}\int_{xmin}^{xmax} \hat{g}(x-x_j, y-y_j) dx dy$ be the spatial integral. The spatial zone being modelled contains the aftershocks of all earthquakes far enough from a border; G_i is thus the same with the definite limits of integration or infinite limits, meaning we can set it to 1 since we use a normalized Utsu-Seki law (which is a probability density). While there are some events near a border for which this does not hold, we ignore this for now and set $\hat{G}_i = 1$ for all earthquakes i.

The analytical integral to the unnormalized Omori law is:

$$\begin{aligned}\int_{t_j}^{t_i} \hat{h}(t-t_j)dt &= \int_{t_j}^{t_i} (t-t_j+\hat{c})^{-\hat{p}}dt \\ &= \frac{1}{1-\hat{p}}(t-t_j+\hat{c})^{1-\hat{p}}|_{t=t_j}^{t=t_i} \\ &= \frac{1}{1-\hat{p}}((t_i-t_j+\hat{c})^{1-\hat{p}} - (\hat{c})^{1-\hat{p}}).\end{aligned}$$

So,

$$\tau_i^{trig} = \sum_j^{t_j<t} \hat{k}(M_j)\Big(\frac{1}{1-\hat{p}}((t_i-t_j+\hat{c})^{1-\hat{p}} - (\hat{c})^{1-\hat{p}}\Big). \tag{B.1}$$

For the background term, since we assume homogeneity in space for the background seismicity, $\mu_{ST} = \frac{\mu}{|S|}$, where $|S|$ is the spatial area of the modelled catalog ($|S| = (xmax - xmin) \cdot (ymax - ymin)$). So,

$$\begin{aligned}\tau_i^{bkgd} &= \int_0^{t_i}\int_S \frac{\hat{\mu}}{|S|}dsdt \\ &= \int_0^{t_i} \frac{\hat{\mu}}{|S|} \cdot |S|dt \\ &= \hat{\mu} \cdot t_i.\end{aligned}$$

The full τ_i is thus:

$$\tau_i = \hat{\mu}t_i + \sum_j^{t_j<t} \hat{k}(M_j)\Big(\frac{1}{1-\hat{p}}((t_i-t_j+\hat{c})^{1-\hat{p}} - (\hat{c})^{1-\hat{p}})\Big).$$

B.7 Derivation of inverse transforms

In simulation step 3a (see Section 3.5), we need to draw random triggering times as specified by the ETAS triggering equation, normalized over the time interval $[0, T]$. To obtain a probability density function (pdf) for this distribution, we divide the triggering equation by its integral over the time interval $[0, T]$, using the fact that $\int_0^T (t+c)^{-p} = ((T+c)^{1-p} - c^{1-p}) \cdot \frac{1}{1-p}$.

$$\begin{aligned} f(t) &= \frac{K \exp(\alpha \cdot (M_i - M_0)) \cdot (t+c)^{-p}}{\int_0^T K \exp(\alpha \cdot (M_i - M_0)) \cdot (t+c)^{-p} dt} \\ &= \frac{K \exp(\alpha \cdot (M_i - M_0)) \cdot (t+c)^{-p}}{K \exp(\alpha \cdot (M_i - M_0)) \cdot ((T+c)^{1-p} - c^{1-p}) \cdot \frac{1}{1-p}} \\ &= \frac{(t+c)^{-p}}{((T+c)^{1-p} - c^{1-p}) \cdot \frac{1}{1-p}}. \end{aligned}$$

From this, we can use the cumulative distribution function (cdf), $F(z) = \int_0^z f(t)dt$, to derive the inverse transform function, from which we can sample triggering times using a random uniform number.

$$\begin{aligned} F(z) &= \int_0^z (1-p)(\frac{(t+c)^{-p}}{((T+c)^{1-p} - c^{1-p})})dt \\ &= (1-p)(\frac{1}{((T+c)^{1-p} - c^{1-p})}) \int_0^z (t+c)^{-p} dt \\ &= (1-p)(\frac{1}{((T+c)^{1-p} - c^{1-p})})(\frac{1}{1-p}\Big((z+c)^{1-p} - c^{1-p}\Big)) \\ &= \frac{(z+c)^{1-p} - c^{1-p}}{(T+c)^{1-p} - c^{1-p}}. \end{aligned}$$

We draw $u \sim Unif(0,1)$ and solve $F(F^{-1}(u)) = u$, to find $F^{-1}(u)$, the inverse transform function with which we can draw a sample from the desired pdf $f(t)$.

$$\begin{aligned}
\frac{(F^{-1}(u)+c)^{1-p} - c^{1-p}}{(T+c)^{1-p} - c^{1-p}} &= u \\
(F^{-1}(u)+c)^{1-p} - c^{1-p} &= u((T+c)^{1-p} - c^{1-p}) \\
(F^{-1}(u)+c)^{1-p} &= u((T+c)^{1-p} - c^{1-p}) + c^{1-p} \\
F^{-1}(u)+c &= (u((T+c)^{1-p} - c^{1-p}) + c^{1-p})^{\frac{1}{1-p}} \\
F^{-1}(u) &= (u((T+c)^{1-p} - c^{1-p}) + c^{1-p})^{\frac{1}{1-p}} - c.
\end{aligned}$$

This is the uth percentile of the desired pdf.

We next derive the inverse transform for the Utsu-Seki spatial decay law (needed for simulation step 3b). First, we need to derive the normalization constant, Q_s, to get a spatial pdf:

$$\begin{aligned}
&Q_s \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x^2+y^2+d)^{-q}dxdy \\
&= Q_s \int_0^{2\pi}\int_0^{R}(r^2+d)^{-q}rdrd\theta \\
&= Q_s \int_0^{2\pi}\frac{1}{1-q}(\frac{(R+d)^{1-q}}{2} - \frac{d^{1-q}}{2})d\theta \\
&= \frac{Q_s}{1-q}\pi((R+d)^{1-q} - d^{1-q}).
\end{aligned}$$

If we set $Q_s = \frac{q-1}{\pi d^{1-q}}$, then we would get

$$\begin{aligned}
&\frac{q-1}{\pi d^{1-q}} \frac{\pi}{1-q}((R+d)^{1-q} - d^{1-q}) \\
&= \frac{-1}{\pi d^{1-q}}((R+d)^{1-q} - d^{1-q}) \\
&= -\frac{-d^{1-q}}{d^{1-q}} \text{ as } R \to \infty \\
&= 1.
\end{aligned}$$

So the spatial pdf is $g(x,y) = \frac{q-1}{\pi d^{1-q}}(x^2+y^2+d)^{-q}$ as this integrates to 1 over the entire domain. We use the cdf, $G(z)$ to derive the inverse transform function.

$$\begin{aligned}
G(z) &= \int_{-z}^{z}\int_{-z}^{z} \frac{q-1}{\pi d^{1-q}}(x^2+y^2+d)^{-q} dxdy \\
&= \frac{q-1}{\pi d^{1-q}} \int_0^{2\pi}\int_0^{z} (r^2+d)^{-q} r dr d\theta \\
&= \frac{q-1}{\pi d^{1-q}} \frac{1}{2(1-q)} \int_0^{2\pi} (z^2+d)^{1-q} - d^{1-q} d\theta \\
&= \frac{q-1}{\pi d^{1-q}} \frac{2\pi}{2(1-q)}((z^2+d)^{1-q} - d^{1-q}) \\
&= \frac{(z^2+d)^{1-q}}{-d^{1-q}} - \frac{d^{1-q}}{-d^{1-q}} \\
&= \frac{(z^2+d)^{1-q}}{-d^{1-q}} + 1.
\end{aligned}$$

We draw $u \sim Unif(0,1)$ and solve $G(G^{-1}(u)) = u$, to find $G^{-1}(u)$, the inverse transform function with which we can draw a sample from the desired pdf $g(x,y)$.

$$\frac{(G^{-1}(u)^2 + d)^{1-q}}{-d^{1-q}} + 1 = u$$
$$(G^{-1}(u)^2 + d)^{1-q} = (u-1)(-d^{1-q})$$
$$G^{-1}(u)^2 = [(u-1)(-d^{1-q})]^{1/1-q} - d$$
$$G^{-1}(u) = \sqrt{[(u-1)(-d^{1-q})]^{1/1-q} - d}.$$

This is the uth percentile of the desired spatial pdf.

B.8 Likelihood-based estimation for simulated catalogs

Evaluating the traditional likelihood function. We began by evaluating the traditional likelihood function that uses the analytical+numerical approach for the spatial integral. We focus on the negative log likelihood (NLL), as this is what is being optimized in our MLE routine. We evaluate it for a given simulated catalog, with K and α varying over a grid near their true values, and the other parameters fixed to their true values. The goal is to see whether the NLL reaches its minimum when K and α are close to their true values. Figure B.3 shows these for two synthetic catalogs.

Another way to check the likelihood values is to evaluate the sum and integral terms under specific conditions:

1. When $K = 0$, the triggered part of the sum term drops out and we have:

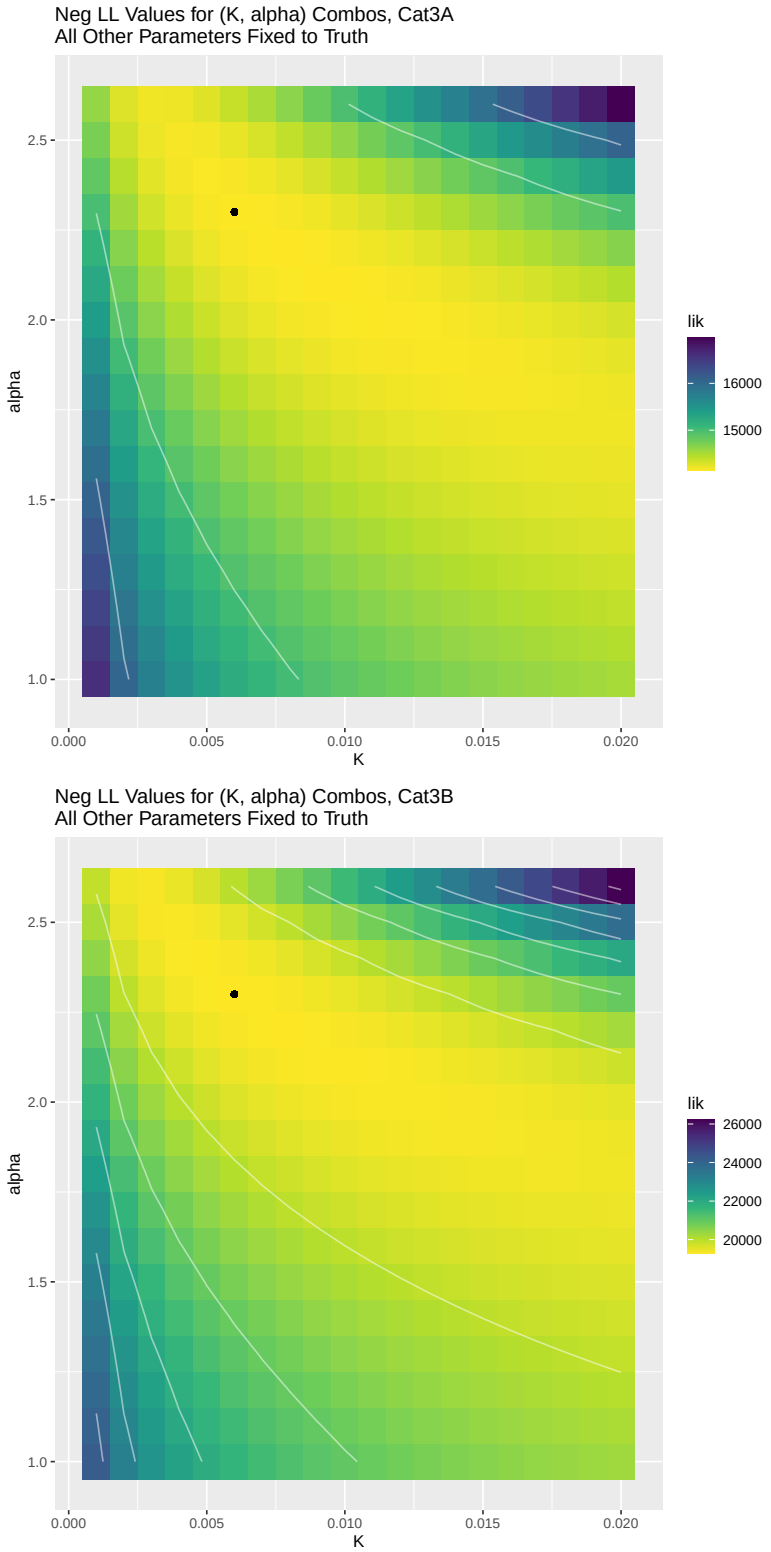

Figure B.3: (Negative log-)likelihood profiles with parameter K and α varying, all other parameters fixed to truth. Contours for every 500 unit change in likelihood. The point is the true values for K and α. Top plot is Catalog 3A, bottom is Catalog 3B.

$$
\begin{aligned}
SUM &= \sum_{i=1}^{n} \log(\lambda(t_i, x_i, y_i|_{t_i})) \\
&= \sum_{i=1}^{n} \log(\mu_{ST} + \sum_{j}^{t_j<t} k(M_j)h(t-t_j)g(x-x_j, y-y_j)) \\
&= \sum_{i=1}^{n} \log(\mu_{ST} + 0) \\
&= n \log(\mu_{ST}).
\end{aligned}
$$

2. When all parameters θ are at their true values $\theta = \theta^*$ (or optimal values when we do not know true values), the integral term should be similar to the number of total events in the catalog; that is $INT_{\theta=\theta^*} \approx n$ at the "correct" parameter values.

We verify in Table B.4 that SUM and INT pass these sanity checks under the traditional likelihood function, for all synthetic catalogs.

Estimating K and α only. We then estimate K together with α, holding all other parameters fixed to their true values. This allows for a simpler algorithm to compute the likelihood function:

$$
l(\theta|_T) = \sum_{i=1}^{n} \log(\lambda(t_i, x_i, y_i|_{t_i})) - \int_0^T \int \int_S \lambda(t, x, y|_t) dx dy dt.
$$

1. $l(\theta|\mathcal{H}_t)$ consists of a sum term and integral term. Pre-compute the parts of each term that do not depend on K, as they do not need to be updated in each iteration of the likelihood optimization.

    - In the sum term, we precompute $h(t-t_j)$ and $g(x-x_j, y-y_j)$ with fixed (true) parameter values c, p, q, d

- In the integral term, we precompute $\int\int_{S^{(j)}} \frac{1}{(x^2+y^2+d)^q} dxdy$ for each mainshock j, with fixed (true) parameter values q, d, using our analytical+numerical approach

2. Maximize the log-likelihood function (minimize the negative log-lik. function) using a numerical optimizer in R.

    - Start with an initial value of K and compute the likelihood as a function of the catalog data (with the pre-computed fixed terms).

    - Update K by increasing/decreasing it by a (random) step size, based on the optimization procedure. We used the non-linear maximization (NLM) method in R, a Newton-type (gradient-based) optimization algorithm. It updates the parameter estimate using the Hessian and gradient of the likelihood function, evaluated at the current parameter estimate.

So the inputs to the $K - \mu$-only likelihood optimization are:

- initial value to start the algorithm

- step size tolerance, the "minimum allowable relative step length", or how large of a step the algorithm will take down the gradient of the likelihood function in every iteration. This can dictate whether the optimizer converges to a local minimum or a global minimum (or even moves from the initial values). Using a tolerance of 1e-6 generally led to good solutions for simulated catalogs.

Estimating μ, K, α, c, p only. See Table B.5.

	Article	ETAS Model Parametrization	Catalog Location	Catalog Time	Catalog M_c	Temporal Units	Spatial Units
1	(Nicolis et al., 2015)	Exponential productivity law with magnitude scaled by alpha-gamma. Magnitude-dependent spatial decay (magnitude scaled by gamma)	Chile	2007-2014	3	days	km^2
2	(Zhuang, 2015)	Exponential productivity law with magnitude scaled by alpha. Magnitude-dependent spatial decay (magnitude scaled by gamma)	Japan	1965-2009	4	days	deg^2
3	(Zhuang, 2011)	Exponential productivity law with magnitude scaled by alpha. Magnitude-dependent spatial decay (magnitude scaled by gamma)	Japan	1965-2003	4	days	deg^2
4	(Yazdi et al., 2019)	Exponential productivity law with magnitude scaled by alpha. Magnitude-dependent spatial decay (magnitude scaled by gamma)	Guerrero (Mexico)	1992-2018	3.8	days	deg^2?
5	(Harte, 2013)	Exponential productivity law with magnitude scaled by alpha. Magnitude-dependent aftersock spatial density (magnitude scaled by gamma), NOT decay (no q). Uses multiple spatial decayand background functions - I summarize results over all these (from Table 4, which uses both spatial and temporal boundary conditions, which is preferred)	New Zealand	1965-2011	4	days	deg^2?
6	(Console et al., 2010)	Linear productivity law with no magnitude-scaling (just productivity K). Magnitude-dependent spatial decay (magnitude scaled by gamma)	New Zealand	1960-2005	3.95	days	km^2
7	(Wei et al., 2012)	Temporal only, exponential productivity with magnitude scaled by alpha	Alaska	1990-2012	1.8	days	-

Table B.2: Collected literature on previous ETAS models built for other subduction zones.

	Article	μ_{low}	μ_{high}	K_{low}	K_{high}	α_{low}	α_{high}	c_{low}	c_{high}	p_{low}	p_{high}	d_{low}	d_{high}	q_{low}	qhigh	γ_{low}	γ_{high}
1	(Nicolis et al., 2015)	0.85	0.90	0.12	0.17	0.05	0.16	0.02	0.02	0.98	0.99	22.38	25.89	1.82	1.88	0.69	0.76
2	(Zhuang, 2015)	0.00	0.40	0.07	0.60	0.40	2.00	NA	NA	1.05	1.40	NA	NA	1.50	3.10	0.30	1.90
3	(Zhuang, 2011)	0.56	0.56	0.37	0.37	0.80	0.80	0.01	0.01	1.15	1.15	0.00	0.00	1.71	1.71	1.61	1.61
4	(Yazdi et al., 2019)	0.01	0.01	0.01	0.01	1.17	1.17	0.01	0.01	1.09	1.09	8.69	8.69	2.47	2.47	0.35	0.35
5	(Harte, 2013)	0.07	0.94	4.95		1.25	1.72	0.00	0.01	1.00	1.21	0.04	0.06	NA	NA	0.32	0.49
6	(Console et al., 2010)	NA	NA	0.00	0.00	NA	NA	0.01	0.01	1.77	1.77	3.30	3.30	1.77	1.77	0.50	0.50
7	(Wei et al., 2012)	0.89	0.89	0.21	0.21			0.02	0.02	1.00	1.00	NA	NA	NA	NA	NA	NA

Table B.3: Parameter values reported in ETAS literature. Most models using an ETAS specified with magnitude-dependent d value.

Catalog	$n\log(\mu_{ST})$	$SUM_{K=0}$	n	$INT_{\theta=\theta^*}$
Cat 2A	-23423.56	-25627.77	1590	1513.05
Cat 2B	-32542.55	-35604.87	2209	2089.9
Cat 2C	-23998.10	-26256.38	1629	1482.5
Cat 2D	-38037.51	-41616.92	2582	2395
Cat 3A	-20388.81	-22307.44	1384	1313.89
Cat 3B	-32409.96	-35459.81	2200	2052.71
Cat 3C	-22362.87	-24467.27	1518	1380.10
Cat 3D	-19917.40	-21791.67	1352	1187.18

Table B.4: Checks that the SUM and INT terms of the branching likelihood function match what they should produce under specific conditions.

	Value	μ	K	α	c	p	d	q
Cat 2A	True	0.1	0.02	1.7	0.05	1.08	1	2
Cat 2A	MLEs	0.108 (0.005)	0.018 (0.001)	1.719 (0.044)	0.04 (0.008)	1.05 (0.012)	Fixed: 1	Fixed: 2
Cat 2D	True	0.1	0.02	1.7	0.05	1.08	0.1	1.5
Cat 2D	MLEs	0.127 (0.005)	0.018 (0.001)	1.73 (0.025)	0.039 (0.006)	1.06 (0.009)	Fixed: 0.1	Fixed: 1.5
Cat 3A	True	0.1	0.006	ln(10)	0.05	1.08	1	2
Cat 3A	MLEs	0.102 (0.005)	0.008 (0.001)	2.162 (0.046)	0.043 (0.01)	1.061 (0.014)	Fixed: 1	Fixed: 2
Cat 3D	True	0.1	0.006	ln(10)	0.05	1.08	0.1	1.5
Cat 3D	MLEs	0.123 (0.005)	0.007 (0.001)	2.298 (0.046)	0.061 (0.014)	1.079 (0.016)	Fixed: 0.1	Fixed: 1.5

Table B.5: New MLEs for several synthetic catalogs - $\mu/K/\alpha/c/p$ estimation only.

Estimating d, q only. See Table B.6.

	Value	μ	K	α	c	p	d	q
Cat 2A	True	0.1	0.02	1.7	0.05	1.08	1	2
Cat 2A	MLEs	Fixed: 0.1	Fixed: 0.02	Fixed: 1.7	Fixed: 0.05	Fixed: 1.08	0.781 (0.100)	1.864 (0.064)
Cat 2D	True	0.1	0.02	1.7	0.05	1.08	0.1	1.5
Cat 2D	MLEs	Fixed: 0.1	Fixed: 0.02	Fixed: 1.7	Fixed: 0.05	Fixed: 1.08	0.0812 (0.008)	1.468 (0.022)
Cat 3A	True	0.1	0.006	ln(10)	0.05	1.08	1	2
Cat 3A	MLEs	Fixed: 0.1	Fixed: 0.006	Fixed: ln(10)	Fixed: 0.05	Fixed: 1.08	0.943 (0.134)	1.963 (0.081)
Cat 3D	True	0.1	0.006	ln(10)	0.05	1.08	0.1	1.5
Cat 3D	MLEs	Fixed: 0.1	Fixed: 0.006	Fixed: ln(10)	Fixed: 0.05	Fixed: 1.08	0.101 (0.0138)	1.522 (0.0358)

Table B.6: New MLEs for several synthetic catalogs - estimation of d and q only.

Estimating all 7 parameters together. See Table B.7.

	Value	μ	K	α	c	p	d	q
Cat 2A	True	0.1	0.02	1.7	0.05	1.08	1	2
Cat 2A	MLEs	0.108 (0.005)	0.019 (0.001)	1.716 (0.044)	0.04 (0.008)	1.049 (0.012)	0.78 (0.101)	1.861 (0.065)
Cat 2D	True	0.1	0.02	1.7	0.05	1.08	0.1	1.5
Cat 2D	MLEs	0.127	0.019	1.70	0.037	1.06	0.087	1.49
Cat 3A	True	0.1	0.006	ln(10)	0.05	1.08	1	2
Cat 3A	MLEs	0.102 (0.005)	0.008 (0.001)	2.161 (0.047)	0.043 (0.01)	1.061 (0.014)	0.912 (0.132)	1.953 (0.081)
Cat 3D	True	0.1	0.006	ln(10)	0.05	1.08	0.1	1.5
Cat 3D	MLEs	0.123 (0.005)	0.007 (0.001)	2.296 (0.046)	0.061 (0.014)	1.08 (0.016)	0.102 (0.014)	1.527 (0.037)

Table B.7: MLEs for several synthetic catalogs, estimating all 7 ETAS parameters. Cat 2D's estimate do not have estimated standard errors because the optimizer crashed after reaching the maximum number of iterations (100); the estimates at that point are given as MLEs.

B.9 Additional results for simulated catalogs

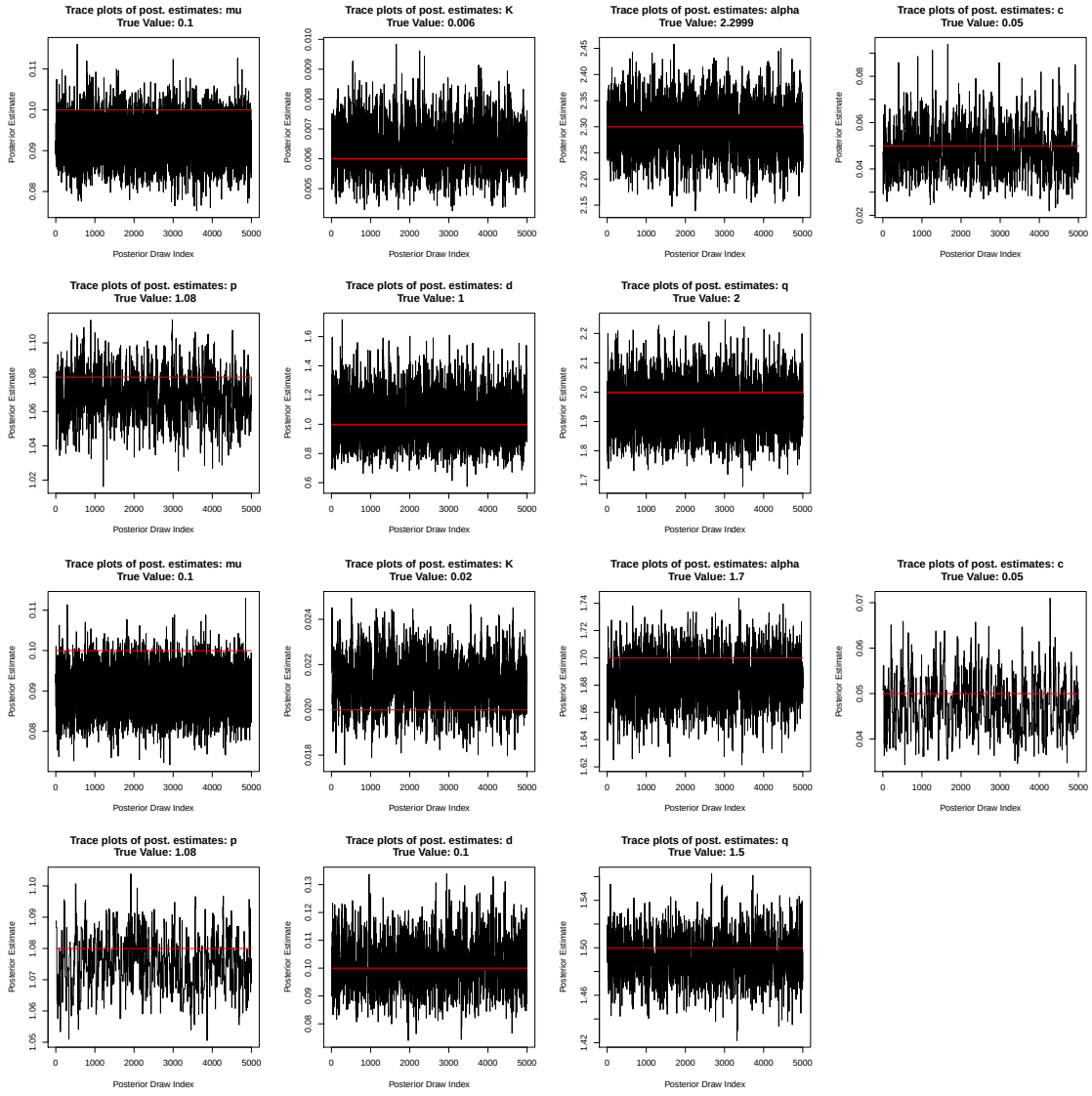

Figure B.4: Trace plots for posterior samples for Catalogs 3A (top) and 2D (bottom), with true parameter values in red. Results were similar for the other synthetic catalogs.

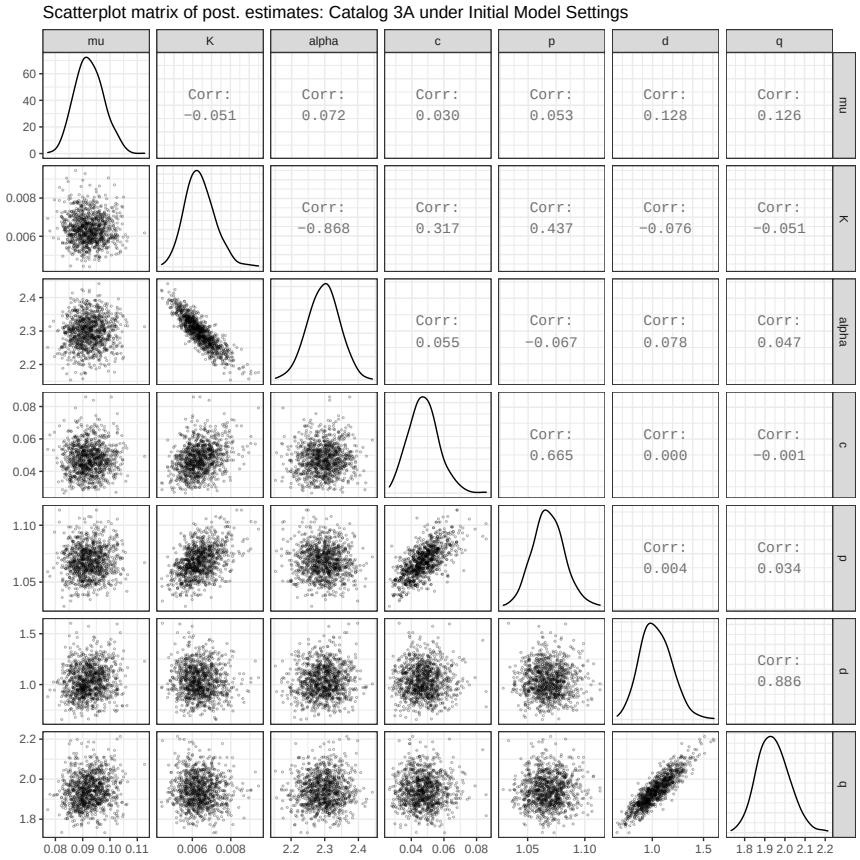

Figure B.5: Scatterplot matrices for posterior samples for pairs of parameters for Catalog 3A (scatterplot matrix for Catalog 2D omitted). Results were similar for the other synthetic catalogs.

	Value	μ	K	α	c	p	d	q	β	T	n	br. ratio	M_{max}
Cat 2A	True	0.1	0.02	1.7	0.05	1.08	1	2	2.3	5000	1725	0.677	6.14
Cat 2A	MLEs	0.108 (0.005)	0.019 (0.001)	1.716 (0.044)	0.04 (0.008)	1.049 (0.012)	0.78 (0.101)	1.861 (0.065)	β	T	n	br. ratio	M_{max}
Cat 2D	True	0.1	0.02	1.7	0.05	1.08	0.1	1.5	2.3	5000	1725	0.677	6.14
Cat 2D	MLEs	0.125 (0.005)	0.019 (0.001)	1.702 (0.019)	0.04 (0.005)	1.061 (0.009)	0.082 (0.007)	1.476 (0.018)	β	T	n	br. ratio	M_{max}
Cat 3A	True	0.1	0.006	2.2999	0.05	1.08	1	2	2.3	5000	1725	0.677	6.14
Cat 3A	MLEs	0.102 (0.005)	0.008 (0.001)	2.161 (0.047)	0.043 (0.01)	1.061 (0.014)	0.912 (0.132)	1.953 (0.081)	β	T	n	br. ratio	M_{max}
Cat 3D	True	0.1	0.006	2.2999	0.05	1.08	0.1	1.5	2.3	5000	1725	0.677	6.14
Cat 3D	MLEs	0.123 (0.005)	0.007 (0.001)	2.296 (0.046)	0.061 (0.014)	1.08 (0.016)	0.102 (0.014)	1.527 (0.037)	β	T	n	br. ratio	M_{max}

Table B.8: MLEs for spatiotemporal ETAS parameters in several synthetic catalogs, using the traditional log-likelihood $l(\theta|Z)$ and Hessian-based standard errors in parentheses. All runs have initial values close to true parameters, though results are insensitive to different initial values. Other catalogs not shown did equally well.

B.10 Additional results for PNW South subcatalogs

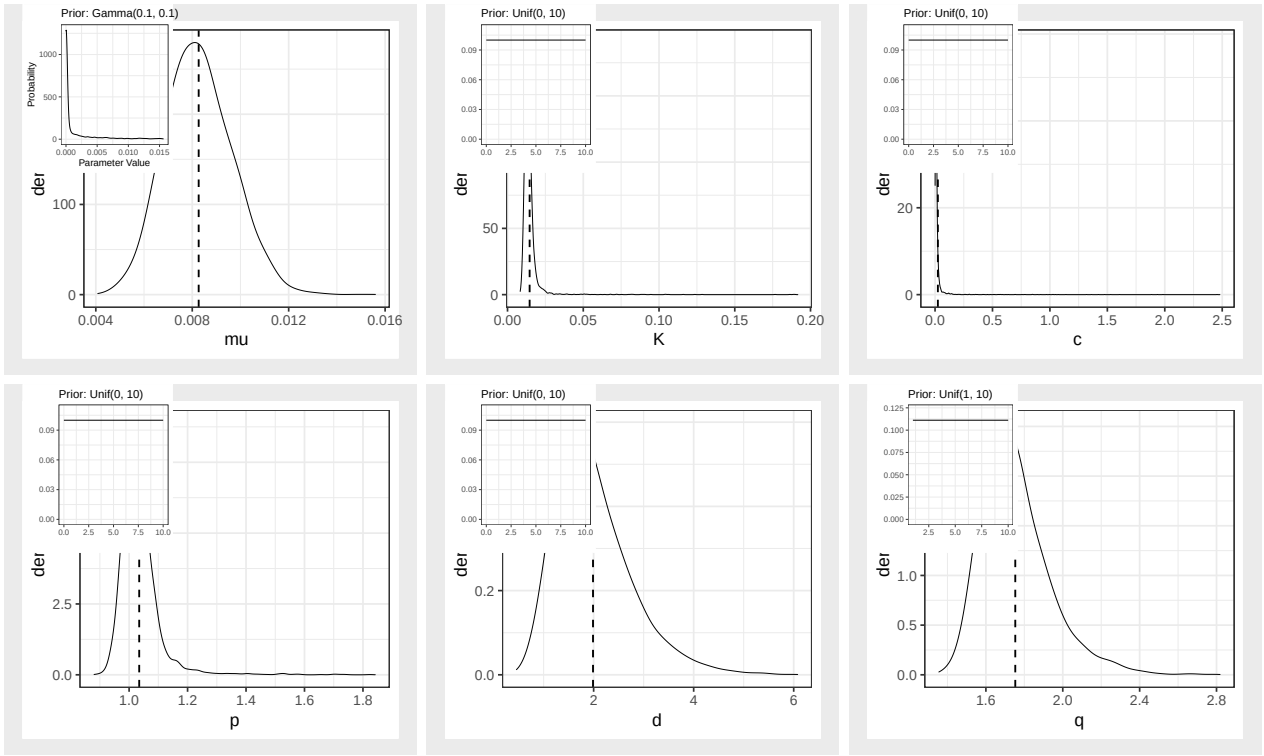

Figure B.6: Sampled posteriors for ETAS parameters for PNW South crustal catalog (M2.3+), without maybe/confirmed swarms, with α fixed to $\hat{b}_{MLE} \ln(10)$. Posterior medians marked with black lines.

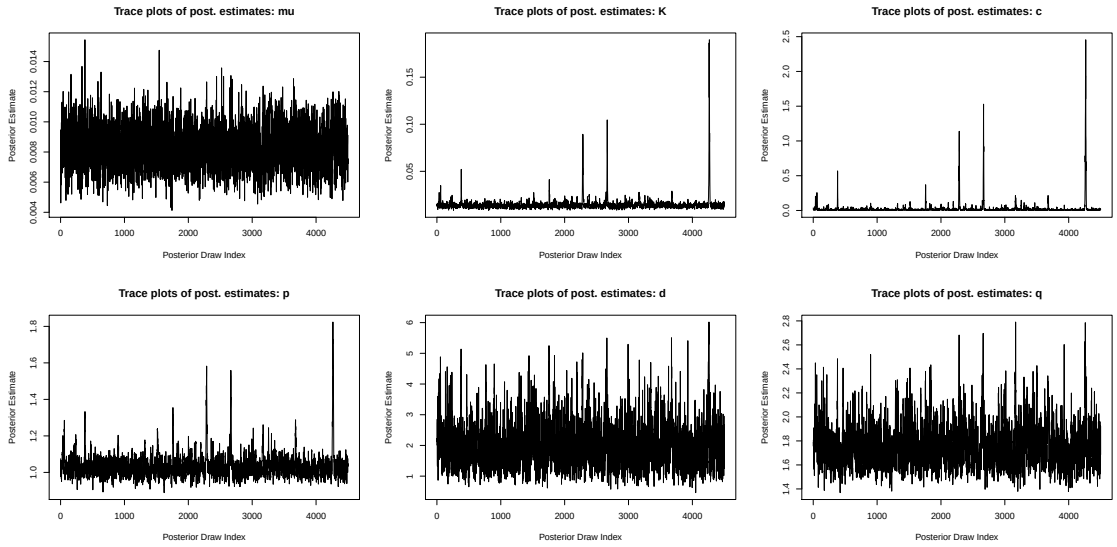

Figure B.7: Trace plots for Bayesian inference for ETAS parameters for crustal catalog, PNW South (M2.3+).

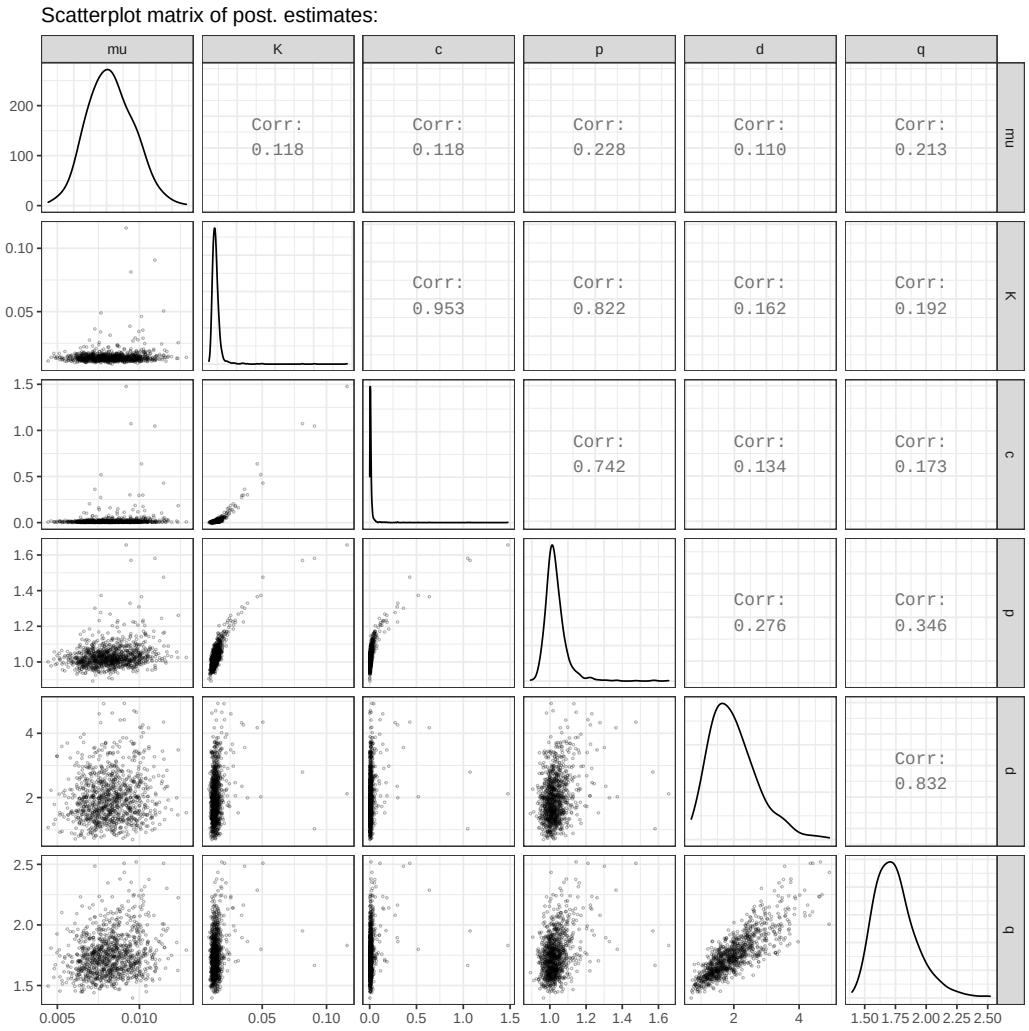

Figure B.8: Scatterplot matrices for Bayesian inference for ETAS parameters for crustal catalog, PNW South (M2.3+).

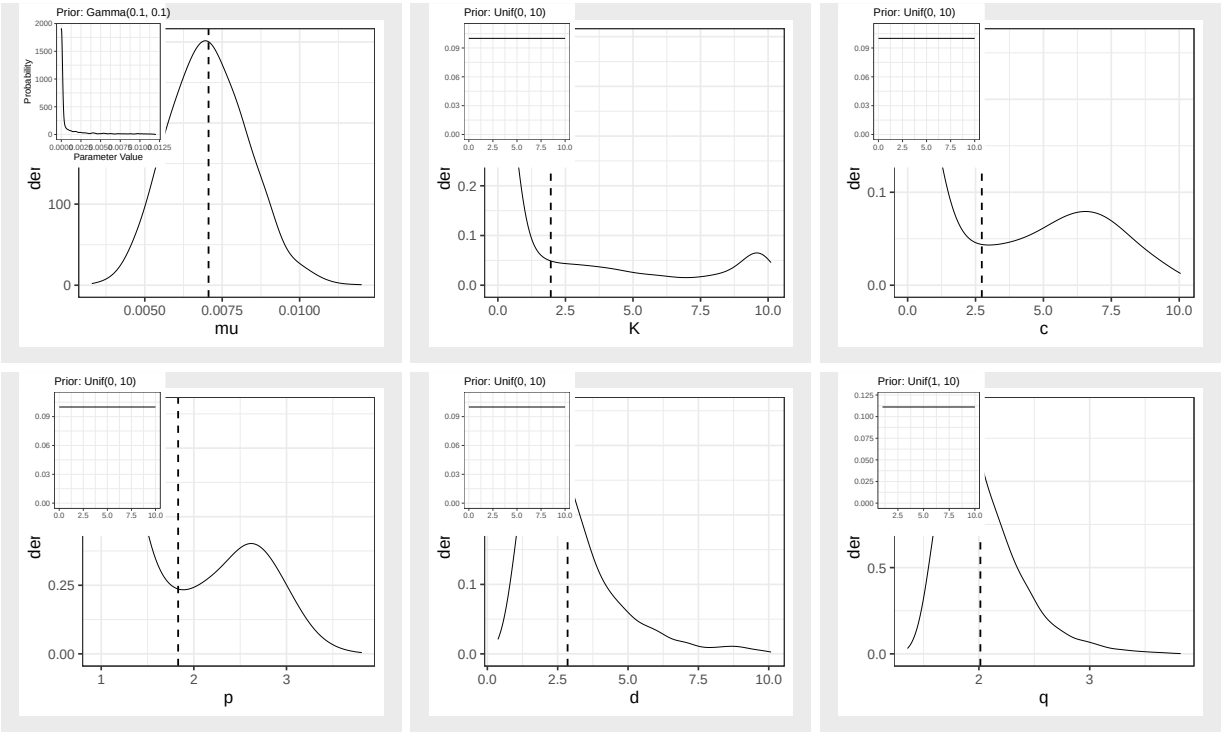

Figure B.9: Sampled posteriors for ETAS parameters for PNW South crustal catalog (M2.5+), without maybe/confirmed swarms, with α fixed to $\hat{b}_{MLE} \ln(10)$. Posterior medians marked with black lines.

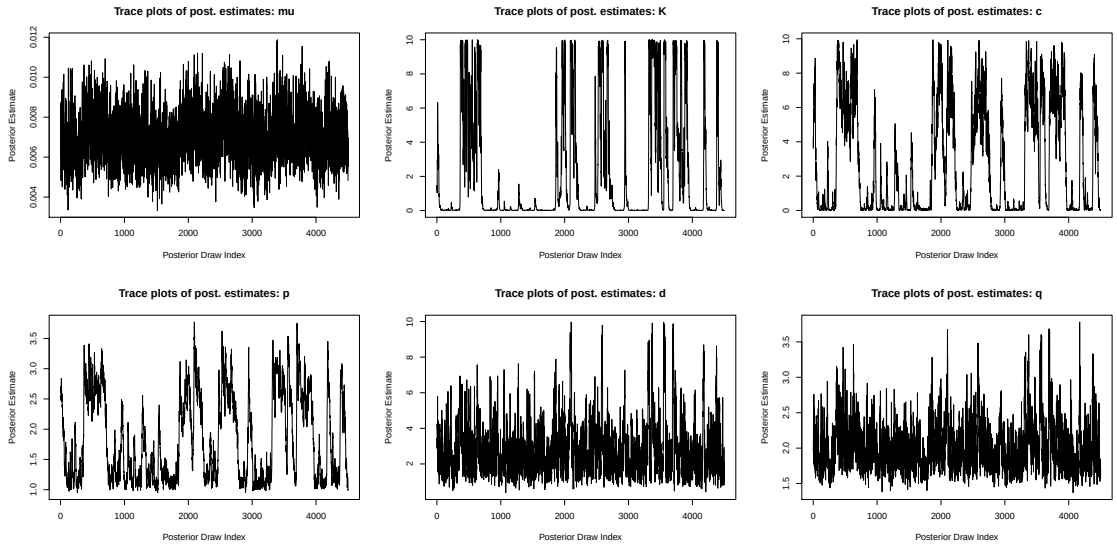

Figure B.10: Trace plots for Bayesian inference for ETAS parameters for crustal catalog, PNW South (M2.5+).

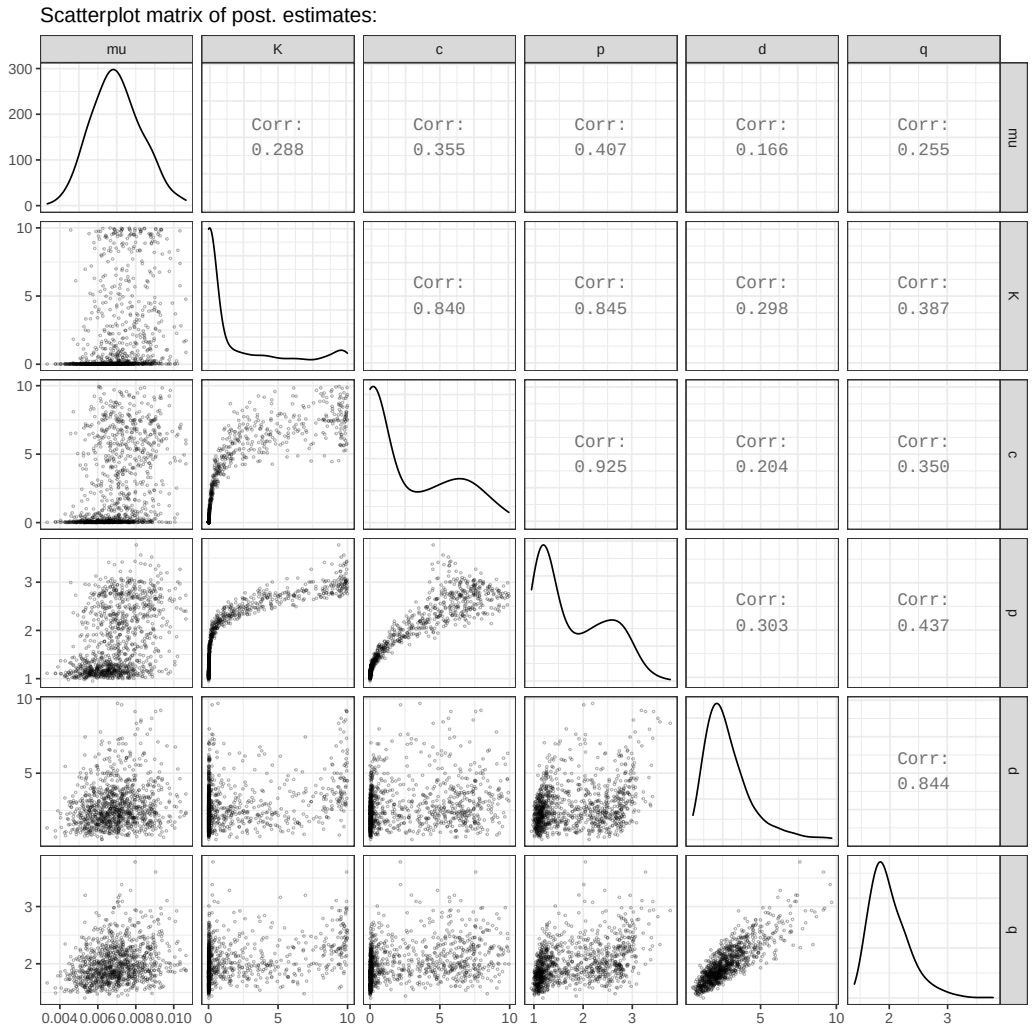

Figure B.11: Scatterplot matrices for Bayesian inference for ETAS parameters for crustal catalog, PNW South (M2.5+).

Appendix C

APPENDIX TO CHAPTER 4

C.1 Selection of colors for maps

The yellow-orange-red (Yl-Or-Rd) palette is often recommended for natural hazards (Doore et al., 1993) and has appropriate color connotations across Western cultures, especially for the lowest rate level (yellow, commonly connoting caution in the natural hazards context) and high rate level (red, commonly connoting danger) (Sherman-Morris et al., 2015; Thompson et al., 2015). We began with the Yl-Or-Rd RColorBrewer palette, known to be colorblind-friendly (Brewer and Harrower, 2009). We fixed the start and end colors as suggested by this palette: respectively, yellow (made darker to maintain distinguishability for different transparency levels) and a dark red. We then created a five-color palette that was perceptually uniform using the Hue-Saturation-Lightness color model. We uniformly decreased colors' lightness and increased hue, while keeping maximal saturation. The hue and lightness of all colors were then adjusted to be optimally discriminable in both the legends and maps. The sixth color (highest rate level) was made dark brown, a color off the Yl-Or-Rd color palette to further distinguish it from the previous colors, as it was not in any visualization aside from the intervals-based design. This was to avoid the false matching of the darkest color in the map with the darkest color on the legend, as occurred when piloting.

We used a white color with black shadow for the area marker and label in the Read Off and Comparative Judgment tasks to avoid simulatanous contrast affecting the perception of the color within the marker (Krauskopf et al., 1986).

C.2 Developing location distance measures

We investigated whether the characteristics of trial locations in the comparative judgment task influenced how participants judged between them. We focused on primary map features (map center, zones of high-rate and -uncertainty) and whether either location was substantially closer to them.

For each location, we calculated its Manhattan distance (number of grid cells) to the map center and the nearest zone of high rate or uncertainty. We then calculated the difference between these distances for the two locations in each trial. These distance differences are only meaningful when neither location itself has high rate or uncertainty and were only computed in those cases. Furthermore, we are interested in assessing whether locations were meaningfully closer to a given map feature, as we did not expect participants to be influenced by small differences in distances. Thus, we created a new categorical variable indicating which location was at least 3 grid cells (the median distance across all locations) closer to each map feature. If both locations were essentially equidistant from that map feature (difference of 2 grid cells or fewer) or if either location was itself in that zone, the trial's value of this variable was "neither", which was set as the reference category. These categorical variables entered our model selection procedure as potential fixed effects (see Text C.4).

C.3 Screenshots of conditions from experiment platform

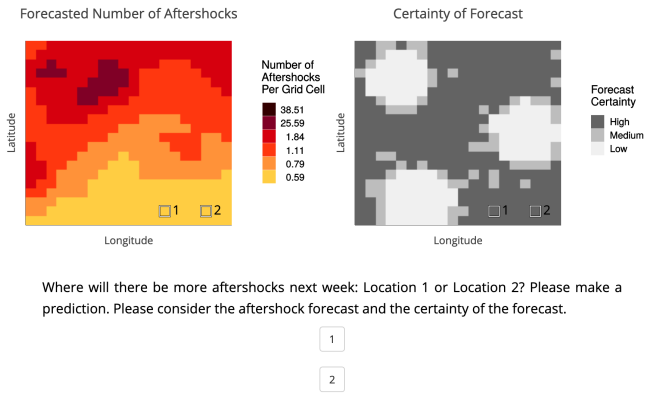

Figure C.1: Screenshot of example baseline trial for Comparative Judgment task with the Adjacent UV.

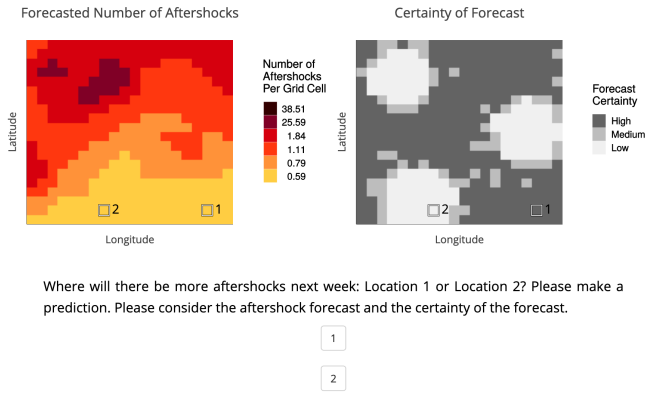

Figure C.2: Screenshot of example surprise trial for Comparative Judgment task with the Adjacent UV.

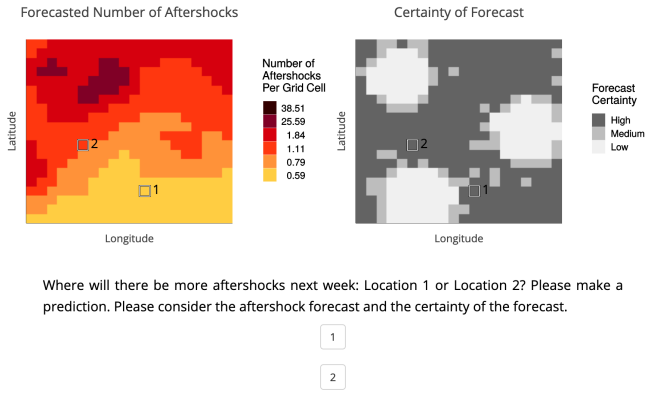

Figure C.3: Screenshot of example sure bet trial for Comparative Judgment task with the Adjacent UV.

C.4 Supplementary model and model selection results

Fixed Effect	Estimated Coefficient	95% CI
Intercept	-1.07	[-1.37, -0.76]
Visualization		
Adjacent	-1.54	[-2.04, -1.04]
Transparency	-1.90	[-2.40, -1.40]
Bounds	0.65	[0.22, 1.07]
Rate Level		
Low	3.15	[2.88, 3.41]
High	0.64	[0.43, 0.86]
*Visualization*Rate Level*		
Adjacent*Low	-2.49	[-2.92, -2.06]
Adjacent*High	-0.86	[-1.27, -0.44]
Transparency*Low	-3.31	[-3.75, -2.87]
Transparency*High	-1.43	[-1.87, -0.99]
Bounds*Low	0.01	[-0.40, 0.42]
Bounds*High	-0.61	[-0.92, -0.31]

Table C.1: Most likely estimates and 95% confidence intervals (using Wald standard errors) for fixed effects in baseline model. The intercept is the logistic of the probability of selecting the higher-uncertainty location for the Rate Only condition for medium-rate trials (reference levels for visualization and rate level). Each fixed effect give show the change in the logistic of the probability corresponding to a change in visualization, rate level or their interaction (all else being held equal).

Model	Fixed Effects	*CCR*	*AUC*	*Brier*	*mBIC*
Model 1	Visualization * Rate Level	0.851	0.931	0.102	17072.6
Model 2	Visualization * Rate Level + Location Closer to Center	0.858	0.941	0.094	16700.7
Model 3	Visualization * Rate Level + Location Closer to Center + Location Closer to High Rate Zone	0.871	0.942	0.093	16670.9

Table C.2: Model performance for baseline model and other models with trial-level fixed effects. CCR is the correct classification rate, AUC is the area under the ROC curve and $mBIC$ is the modified Bayesian information criterion (see Section 4.2.4 for definitions).

Model	Fixed Effects	CCR	AUC	$Brier$	$mBIC$
Model 0	Visualization * Rate Level	0.851	0.931	0.102	17072.6
Model 4	Visualization * Rate Level + Age	0.850	0.931	0.102	17078.6
Model 5	Visualization * Rate Level + log(numberEq + 0.01)	0.851	0.931	0.102	17080.3
Model 6	Visualization * Rate Level + Education	0.850	0.931	0.102	17112.6
Model 7	Visualization * Rate Level + Gender	0.851	0.931	0.102	17089.9
Model 8	Visualization * Rate Level + State	0.850	0.931	0.103	16948.4

Table C.3: Model performance for baseline model and other models with combos of participant-level covariates. Note: We attempted several transformations of the highly skewed numberEq variable, including sqrt() but it never performed better in a model. Model 1 from Table C.2, with an interaction term between Rate Level and UV outperforms these models with covariates. Results were similar when using Forecast Region in place of UV or Rate Level.