Christian Grimm

How Frequent Are Earthquake Doublets Worldwide? Advancing the Epidemic Type Aftershock Sequence Model to Improve Forecasts of Multi-Mainshock Sequences

Dissertation an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig-Maximilians-Universität München

Eingereicht am 20.05.2022

How Frequent Are Earthquake Doublets Worldwide?
Advancing the Epidemic Type Aftershock Sequence Model to Improve Forecasts of Multi-Mainshock Sequences

Dissertation
an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig-Maximilians-Universität München

eingereicht von

Christian Grimm

20.05.2022

Erster Berichterstatter: Prof. Dr. Helmut Küchenhoff
Zweiter Berichterstatter: Prof. Dr. Martin Käser
Dritter Berichterstatter: apl. Prof. Dr. Gert Zöller

Tag der mündlichen Prüfung: 26.10.2022

Acknowledgments

I would like to express my heartfelt thanks to the many people who have supported and accompanied me over the past three years.

*Lieber **Martin**, ich möchte Dir sehr herzlich für deine tolle, ausdauernde und motivierende Unterstützung danken! Du hast mich seit 2015 als Betreuer im Praktikum, in der Bachelor- und Masterarbeit sowie jetzt in der PhD fast durch mein ganzes Studium begleitet. Ich habe in dieser Zeit viel von Dir gelernt, und denke an viele nette Momente und Anekdoten zurück. Ich hätte mir keinen besseren Betreuer wünschen können!*

*Auch Dir, lieber **Helmut**, ein spezielles Dankeschön, dass Du dieses spannende Forschungsthema im Schnittfeld von Statistik und Geophysik, zwischen Akademie und Industrie, vom ersten Moment an unterstützt hast! Vielen Dank, dass ich Teil des StaBLab-Teams sein durfte - ich denke besonders an die gemeinsamen Team-Sitzungen vor Ort und die geselligen Wanderausflüge zurück.*

*Ein dickes Dankeschön geht nach Potsdam zu Dir, lieber **Sebastian**. Danke für Deine vielen Stunden aufopferungsvolle Feinschliff-Diskussionen für die ETAS-Modellentwicklung! Danke für Deine außergewöhnlich detaillierten Paper-Korrekturen. Ich habe viel von Dir gelernt. Du warst, und das meine ich von ganzem Herzen, einer meiner wichtigsten Kontakte während der PhD!*

*Caro **Marco**, grazie per il tuo importante supporto, soprattutto all'inizio del mio dottorato. In particolare vorrei ringraziarvi per il tempo fantastico e davvero istruttivo a Pavia - sono molto felice di poter finalmente vivere questa esperienza.*

*Ein herzlicher Dank an das StaBLab-Team sowie alle Doktoranden-Kollegen am Institut für die schöne Zeit mit Euch. Herausheben möchte ich hier **Andi** und **Alex** - ihr habt mir auf den letzten Metern der PhD mit vielen Diskussionen wirklich sehr weitergeholfen.*

*A big thank to you, **Kendra** and **Teresa**, for having supported me so well on my last paper!*

*Many thanks to **Munich Re colleagues** for the wonderful working atmosphere over the past three years, for the nice lunches and coffee breaks. I miss you!*

*Grazie mille to the **GEM team** in Pavia for the great and instructive time!*

*¡Gracias a ti, mi **Amor** por animarme cuando mi código no funcionó y celebrar conmigo cuando funcionó! No puedo expresar con palabras lo importante que eres para mí. ¡Te amo mucho!*

*Danke **Mama** und **Papa** und **Björn**, dass ihr immer da seid und mich unterstützt!*

*Danke an Euch liebe **Freunde**, für die vielen wunderschönen Momente in den letzten Jahren. Ihr seid wichtig in meinem Leben!*

Summary

As energy is released in the event of a strong earthquake, tectonic stress redistributes in the surroundings of the initial rupture and typically results in further aftershocks. Studies have shown that these aftershocks can substantially increase damage in buildings and infrastructure due to prior destabilization of the structure through the mainshock. This has implications to a wide range of disciplines such as seismic engineering, emergency response management and insurance loss models. The main interest is typically in the strongest aftershock of a sequence.

Thus, the focus of this dissertation is on the statistical analysis and modeling of so-called earthquake doublets, which are generally defined as a pair of two similarly strong earthquakes, occurring in temporal and spatial proximity to each other. The goal is to develop statistical models that predict the long-term occurrence probabilities of doublets, forecast the spatio-temporal evolution of triggered sequences and explain the main drivers and geophysical conditions that favor the occurrence of such multi-mainshock sequences.

To date, the literature only provides a starting point for an understanding of earthquake doublets, based on non-consistent definitions and regional studies. In particular, no parametric distribution of the magnitude difference ΔM between the mainshock and its strongest aftershocks has been proposed so far. This dissertation aims to close this gap by suggesting two comprehensive, parametric approaches for modeling and predicting earthquake doublet probabilities worldwide.

In the first approach, I develop two advanced versions of the so-called *Epidemic Type Aftershock Sequence (ETAS)* model, which describes the spatio-temporal evolution and self-exciting nature of earthquake sequences as a special case of a Hawkes process. The *ETAS-Anisotropic* model generalizes the spatial aftershock distribution, conventionally assumed to be isotropic, to more adequately reflect the observed anisotropic shape of aftershock clouds. The *ETAS-Incomplete* model *additionally* accounts for typically short-term incomplete aftershock records, and therefore solves three major ETAS model biases at once.

In the second approach, I propose the innovation of adapting *survival models*, originally developed for medical applications, in order to estimate the fully parametric distribution of the magnitude differences ΔM. The observations of ΔM are partially *right-censored*, as the exact value is unknown if the largest aftershock was not recorded due to the magnitude completeness threshold of the underlying catalog. A simulation model demonstrates the leverage effect of the two main drivers, aftershock productivity and frequency-magnitude distribution, on ΔM. The variation of aftershock counts is analyzed by a generalized additive model (GAM).

Results indicate that approximately 20% of the global $M \geq 6$ mainshocks trigger doublets. The preferred ETAS-Incomplete model substantially improves both doublet frequency predictions and spatio-temporal sequence forecasts, while the conventional ETAS model provides poor estimates due to its assumptional biases. The survival model approach suggests that ΔM is best modeled by a Gompertz distribution. Earthquakes at larger depths tend to trigger less aftershocks and have larger ΔM observations. Additionally, the GAM results suggest that triggered events may produce two to three times more aftershocks than background events, which would substantially increase cluster sizes and doublet occurrence probabilities.

Future research should further investigate the latter observation and analyze the impact of the frequency-magnitude distribution of triggered events on ΔM. In particular, potential magnitude correlations between the mainshock and aftershocks may have a substantial impact on doublet occurrence chances.

Zusammenfassung

Beim Auftreten eines starken Erdbebens wird Energie freigesetzt, die sich in Form von tektonischem Stress in der Umgebung des Erdbeben-Bruchs verteilt und typischerweise zu weiteren Nachbeben führt. Studien haben gezeigt, dass diese Nachbeben die Schäden an Gebäuden und Infrastruktur erheblich verstärken können. Dies hat Auswirkungen auf eine Vielzahl seismologiebezogener Disziplinen wie Seismic Engineering, Emergency Response Management und Insurance Loss Modeling. Das Hauptinteresse gilt typischerweise dem stärksten Nachbeben einer Sequenz.
Der Fokus dieser Dissertation liegt daher auf der statistischen Analyse und Modellierung sogenannter Erdbeben-Dubletten, die allgemein definiert werden als ein Paar von zwei ähnlich starken Erdbeben, die in zeitlicher und räumlicher Nähe zueinander auftreten. Ziel ist es, statistische Modelle zu entwickeln, die die langfristigen Auftrittswahrscheinlichkeiten von Dubletten sowie die räumlich-zeitliche Entwicklung getriggerter Sequenzen vorhersagen und die geophysikalischen Bedingungen erklären, die das Auftreten solcher Multi-Beben-Sequenzen begünstigen.
Bisher bietet die Literatur nur einen Ausgangspunkt für das Verständnis von Erdbeben-Dubletten, basierend auf uneinheitlichen Definitionen und regionalen Studien. Insbesondere wurde bisher keine parametrische Verteilung der Magnitudendifferenz ΔM zwischen dem Hauptbeben und seinen stärksten Nachbeben vorgeschlagen. Diese Dissertation zielt darauf ab, diese Lücke zu schließen, indem sie zwei umfassende, parametrische Ansätze zur Modellierung und Vorhersage der Wahrscheinlichkeit von Erdbeben-Dubletten weltweit vorschlägt.
Im ersten Ansatz entwickle ich zwei erweiterte Versionen des sogenannten *Epidemic Type Aftershock Sequence (ETAS)*-Modells, das die räumlich-zeitliche Entwicklung von Erdbebensequenzen über mehrere Trigger-Generationen als Spezialfall eines Hawkes-Prozess beschreibt. Das *ETAS-Anisotropic*-Modell verallgemeinert die im konventionellen Modell isotrope räumliche Nachbebenverteilung durch eine anisotrope Form, die beobachtete Erdbeben-Cluster realistischer widerspiegelt. Das *ETAS-Incomplete*-Modell berücksichtigt *zusätzlich* die typischerweise unvollständigen Nachbebenaufzeichnungen kurz nach einem Hauptbeben und behebt damit die drei wichtigsten Verzerrungen im konventionellen ETAS-Modell mit einem Ansatz.
Im zweiten Ansatz schlage ich die Innovation vor, die parametrische Verteilung der Magnitudendifferenzen ΔM mittels *Survival-Modellen* zu schätzen. Die Beobachtungen von ΔM sind teilweise *rechtszensiert*, da der genaue Wert unbekannt ist, wenn das größte Nachbeben aufgrund der unteren Magnituden-Schwelle des zugrunde liegenden Katalogs nicht aufgezeichnet wurde. Ein Simulationsmodell demonstriert die Hebelwirkung der Nachbebenproduktivität sowie der Magnituden-Verteilung mit Bezug auf ΔM. Die Nachbebenproduktivität wird zusätzlich durch ein generalisiertes additives Modell (GAM) analysiert.
Die Ergebnisse zeigen, dass ungefähr 20% der globalen Hauptbeben Dubletten auslösen. Verglichen mit dem konventionellen Modell liefert das ETAS-Incomplete-Modell deutlich verbesserte Frequenzvorhersagen sowie räumlich-zeitliche Vorhersagen von Erdbeben-Sequenzen. Der Survival-Modell-Ansatz legt nahe, dass ΔM am besten durch eine Gompertz-Verteilung modelliert wird. Tiefe Erdbeben neigen dazu, weniger Nachbeben zu triggern und haben tendenziell größere ΔM. Darüber hinaus deuten die GAM-Ergebnisse darauf hin, dass getriggerte Beben zwei- bis dreimal mehr Nachbeben erzeugen als sogenannte Background-Beben, was die Clustergröße und somit die Wahrscheinlichkeit des Auftretens von Dubletten erheblich erhöhen würde.
Zukünftige Forschungsarbeiten sollten die letztgenannte Beobachtung weiter untersuchen und den Einfluss der Magnituden-Verteilung auf ΔM analysieren. Insbesondere potenzielle Korrelationen zwischen der Nachbeben- und Hauptbeben-Magnitude können einen erheblichen Einfluss auf die Wahrscheinlichkeit des Auftretens von Dubletten haben.

Contents

Part I.

Introduction

1. Motivation and Scope

1.1. Basics of Earthquake Clustering

Earthquake Sequences As energy is released in the event of a strong earthquake, tectonic stress redistributes in the surroundings of the initial rupture and typically results in further earthquakes, so-called *aftershocks* (Utsu et al., 1995).
The cascade of aftershocks is commonly referred to as an *earthquake sequence*, and the strongest event of the sequence is called the *mainshock*. Typically, events that occurred shortly before the mainshock, so-called *foreshocks*, are included in the sequence since they are believed to be physically related to the upcoming major earthquake (e.g. Helmstetter and Sornette, 2003). The dependent occurrence behavior of earthquakes is hereafter referred to as *earthquake clustering*.

Spatio-Temporal Clustering Strong earthquakes are usually observed to cause a pronounced spatio-temporal pattern of aftershocks. According to the Omori-Utsu Law (Utsu et al., 1995), the temporal aftershock rate is subject to a power law decrease with time after the triggering event, i.e., the aftershock sequence is dominated by events occurring shortly after the mainshock.
The observed spatial patterns of aftershock sequences stem from their tendency to occur on or close to the mainshock rupture plane (Marsan and Lengliné, 2008). The larger the length-to-width ratio of this plane gets, the more elongate the typical aftershock region becomes. In addition, a higher dip angle reduces the width of the 3D-to-2D projection of the rupture plain to the earth's surface and therefore results in a scatter of two-dimensional aftershock epicenters that can be increasingly well approximated by a line segment. For instance, the prevailing continental tectonic regime in southern California with typically steep, strike-slip faulting favors such elongated aftershock patterns in this region (Marsan and Ross, 2021).

1.2. Relevance for Insurance Loss Models

Earthquake sequences have the potential to increase the risk of both human fatalities and economic loss induced by a major earthquake. This applies in particular to the strongest aftershocks, which will therefore be the focus of this dissertation. Studies have shown that strong aftershocks can substantially increase damage in buildings and infrastructure due to prior destabilization of the structure through the mainshock. Similarly, foreshocks can set the stage for more severe mainshock damage (Abdelnaby, 2012; Kagermanov and Gee, 2019; Papadopoulos et al., 2020).
Therefore, a better understanding and predictability of the spatio-temporal pattern and the largest expected aftershock (or foreshock) of an earthquake sequence is relevant to a wide range of disciplines and applications related to seismic hazard, including seismic engineering, emergency response management and insurance loss models.

Here I give a short, compact overview of insurance loss models for earthquake risks, which provided the main motivation and inspiration for this dissertation. The summary is based on Mitchell-Wallace et al. (2017) (especially Section 3.7). Earthquake risk models are typically based on the following three model

- The **hazard model** analyses the seismic hazard, including statistics of the frequency, severity and spatial distribution of earthquakes in the modelled region. The result of the hazard analysis is a so-called *event table*, that lists a large amount of synthetic earthquakes and provides an annual occurrence rate estimate for each of them. Additionally, a so-called *ground motion footprint* is computed for each earthquake, specifying the estimated severity of ground shaking as a function of the distance to the presumed earthquake location, the earthquake's magnitude and the local soil condition.

- The **vulnerability model** translates the footprints, evaluated at the various locations of insured objects, into a damage estimate. The event loss, which is the aggregate of the financial loss to all insured buildings due to a single event, is then added to the event table, hereafter referred to as *event loss table*.

- Finally, the **financial model** applies (re-)insurance contract terms such as the *event excess of loss (EventXL)*, that applies a retention and a limit to the aggregated loss of an earthquake event. Formally, an "event" is typically defined as the mainshock, including all aftershocks occurring within typically 72 hours, which consequently count toward the mainshock loss. Conversely, later aftershocks are counted as new loss events to which the EventXL term is reapplied. In practice, however, it usually takes several weeks to few months to quantify the losses of a major mainshock, which makes it almost impossible to distinguish between the damage caused by later aftershocks and the mainhock.

Status Quo For reasons of simplicity and computational efficiency, earthquake risk models typically rely on the probabilistic seismic hazard analysis (PSHA) approach (Cornell, 1968), which builds upon *declustered* earthquake catalogs and therefore only accounts for the mainshock hazard. Mainshock occurrences are often modeled by an independent Poisson process (e.g. McGuire, 2008) or a Brownian passage time model (e.g. Zöller, 2018) for earthquake recurrence probabilities. In doing so, traditional PSHA neglects the additional hazard due to aftershock sequences and underestimates the risk of multiple strong earthquakes hitting the same region within short time.

Including Earthquake Clustering Accounting for earthquake clustering in insurance loss models may affect all of the three abovementioned model components.

- When accounting for earthquake clustering in the hazard model, earthquake occurrences could no longer be modeled independently of each other. A possible approach, which is demonstrated in the case study in Subsection 2.6, is to model mainshock events according to PSHA, and then to add simulated aftershock sequences for each of the mainshocks.

- In the vulnerability model, building damage induced by a simulated aftershock would need to be modeled depending on the pre-existing damage induced by the former mainshock. A possible approach are so-called *damage state-dependent fragility curves* (Papadopoulos et al., 2021). In Subsection 2.6, I will apply a much simplified assumption. A relevant factor for damage correlations is the distance of both events to the affected location, highlighting the importance of an adequate model for the spatial aftershock distribution.

- For the financial model, the relevant criterion is the inter-event time between the mainshock and the aftershock, highlighting the importance of an adequate model for the temporal aftershock distribution.

1.3. Definition of Earthquake Doublets

Section 1.2 highlighted the relevance of a suitable model for both the expected number and severity of aftershocks as well as their spatio-temporal distribution in an on-going earthquake sequence. In nearly all risk-related contexts, the main interest is in the strongest aftershock (or foreshock) of the sequence. Therefore, the focus of this dissertation is on the statistical analysis and modeling of so-called *earthquake doublets*, which are generally defined as (e.g. Felzer et al., 2004; Kagan and Jackson, 1999; Gibowicz and Lasocki, 2005; Lay and Kanamori, 1980)

- a pair of two similarly strong earthquakes,
- occurring in temporal and
- spatial proximity to each other.

Unfortunately, there is no uniform specification of the terms ”similarly strong” and ”temporal and spatial proximity” in the literature. Kagan and Jackson (1999) defined doublets as pairs of earthquakes with magnitude $M_w \geq 7.5$, that are no more than one rupture size apart and whose inter-event time is less than their recurrence time derived from plate motion, which is typically in the range of decades to centuries. In contrast, Felzer et al. (2004) specified *multiplets* (generalization of doublets) as a mainshock together with two or more aftershocks within 0.4 magnitude units, occurring on a much shorter time scale of two days, and within a spatial box centered in the mainshock's epicenter. The distance of the mainshock's epicenter to the sides of the box is 2.5 times the estimated fault length. Gibowicz and Lasocki (2005) interpreted doublets as a pair of earthquakes with no more than 0.25 magnitude units difference, applying magnitude-dependent stepwise spatial and temporal constraints of 40-90 kilometers and 200-450 days.

Definition in this Dissertation In the first contribution to this dissertation (Grimm et al., 2021), I defined earthquake doublets as a pair of events with a magnitude difference of no more than 0.4 units, occurring within 365 days and within a circular radius of 2.5 times the estimated rupture length of the earlier event. The temporal constraint of one year is derived from the typically modeled time span in a (re-)insurance loss model.
Hereafter, I use the term *multi-mainshock sequence* synonymously for earthquake doublets.

1.4. Research Questions

The goal of this dissertation is to contribute to a better understanding of the occurrence frequencies and driving forces of worldwide earthquake doublets. The main research questions addressed in this dissertation are:

1. How frequent are earthquake doublets worldwide?
2. In which tectonic regions, and under which geophysical conditions, are doublets most likely to occur?

3. Which are the main drivers of earthquake doublet occurrences?

4. Can we develop an adequate model to predict long-term occurrence probabilities of doublets?

5. Can we develop an adequate model to forecast the spatio-temporal evolution of a particular earthquake sequence?

1.5. Literature Review

To date, the literature only provides a starting point for an understanding of earthquake doublets. Kagan and Jackson (1999) found that approximately 22% of the $M > 7.5$ earthquakes worldwide occurred accompanied by another $M > 7.5$ event within a distance of one rupture length. Felzer et al. (2004) demonstrated statistical evidence that foreshocks, aftershocks, and doublets occur due to the same physical triggering mechanism and that the number of times that doublets occur increases linearly with the number of aftershocks observed. They inferred that certain regions in the world, such as the Solomon Islands, show an increased doublet rate due to higher aftershock rates and earthquake density, rather than unique seismic fault structures that support the occurrence of doublets. Gibowicz and Lasocki (2005) analyzed the occurrence frequency of earthquake doublets in the Fiji-Tonga-Kermadec region and found that 36% of all shallow, 14% of all intermediate, and 27% of all deep events were associated with a doublet. Note that these results are not directly comparable due to the varying doublet definitions (see Section 1.3).

A more general statement on the expected magnitude difference between a mainshock and its strongest aftershock is made by the well-established *Bath's law* (Bath, 1965). It states that the average difference is roughly 1.2, *independently* of the size of the mainshock. The main challenge in calculating this value is the bias introduced by missing data, if no aftershock was observed above the cut-off magnitude M_c of the catalog and therefore ΔM cannot be computed. Ignoring missing values leads to a systematic error, because the data points removed are those with particularly large magnitude differences ΔM. Several authors found that the statistics is robust, if we restrict the sample to mainshocks at least two magnitude units above M_c, but then the sample size is strongly reduced (e.g. Tahir et al., 2012; Zakharova et al., 2013). Another workaround was suggested by Zakharova et al. (2013), who modeled the seismic moment ratio between aftershocks and the mainshock, rather than ΔM, approximating the ratio by zero if no aftershocks were recorded.

Note that Bath's law only makes a statement about the average value of the ΔM, but not about their distribution (and its parameters) or any important quantiles in the lower tail of the distribution. Similarly, the available literature on earthquake doublets only delivers regional estimates of doublet frequencies based on different thresholds for ΔM, but does not provide a parametric and globally applicable concept to predict doublet probabilities. This dissertation aims to close this gap and to create a comprehensive approach to modeling and predicting earthquake doublet probabilities.

Further literature on earthquake cluster models and the analysis of the variation of aftershock productivity is highlighted in the discussion sections 2.7 and 3.4, respectively.

1.6. Approaches and Outline

This dissertation comprises three contributions made through papers published in or submitted to peer-reviewed journals. The papers analyze clustered seismicity for different tectonic settings in Japan, New Zealand and Southern California. The aforementioned research questions (see Section 1.4) and goals are tackled by two distinct approaches, which are outlined here in short.

Approach 1 - Epidemic Type Aftershock Sequence Model
The first and second contribution (Grimm et al., 2021, 2022a) aim to further develop the so-called *Epidemic Type Aftershock Sequence (ETAS)* model, which describes the spatio-temporal evolution and self-exciting nature of earthquake sequences as a special case of a Hawkes process. Innovations include two enhanced versions of the conventional model by applying more realistic, anisotropic and locally restricted spatial aftershock distributions and accounting for typically observed short-term incomplete aftershock records that may otherwise lead to substantial bias in the estimation results. The model is used to predict long-term occurrence probabilities of doublets and to forecast the spatio-temporal evolution of a local earthquake sequence.

Approach 2 - Statistical Regression Models
The third contribution (Grimm et al., 2022b) presents an innovative approach for modelling the magnitude difference ΔM between a mainshock and the second strongest event in the sequence, by adapting methods for time-to-event data, which often suffers from incomplete observation (censoring). The model is applied to a global set of earthquake clusters to identify in which tectonic regions, and under which geophysical conditions, ΔM may tend to be decreased. Additionally, this contribution investigates the main drivers of doublet occurrences and applies a *generalized additive model* to analyze the variation in aftershock productivity.

Structure of this Dissertation
This dissertation is structured as follows. Part I gives an overview of the research and contributions to this dissertation. Within this Part, there are four chapters. The current chapter 1 gives an introduction and motivation to the research topic. Chapter 2 expands on the first approach, providing a thorough introduction to the conventional ETAS model and defining the advanced model versions suggested in the first and second contribution (Grimm et al., 2021, 2022a). Then, chapter 3 summarizes the third contribution (Grimm et al., 2022b) and gives an overview of the statistical regression models for the second approach. Finally, chapter 4 provides a short, overarching conclusion and gives answers to the research questions posed in Section 1.4.
Part II lists the published papers underlying the first approach (Grimm et al., 2021, 2022a). Part III provides the manuscript of the third contribution (Grimm et al., 2022b), which is currently under revision in a peer-reviewed journal. Finally, the Part IV (Appendix) provides a non-published ETAS formulary with derivations of formulas and derivatives needed to implement a gradient based optimization method for various ETAS model versions.
In the spirit of open science and to ensure full reproducibility, the `Matlab` (Matlab, 2019) source code for ETAS model estimations and simulations, as performed in the contributions Grimm et al. (2021, 2022a), is made publically available in a Github repository.

2. Approach 1: Epidemic Type Aftershock Sequence (ETAS) Model

This chapter summarizes the first and second contribution, in which I developed advanced versions of the ETAS model in order to improve long-term predictions of earthquake doublet frequencies (Grimm et al., 2021) and enhanced spatio-temporal forecasts of the local 2019 Ridgecrest aftershock sequence (Grimm et al., 2022a). The numerical code was implemented using the software `Matlab` (Matlab, 2019).
Section 2.1 lays the methodological foundation, first introducing Hawkes processes, then giving a thorough formulation of the conventional ETAS model as a special case of the latter, and finally explaining the three major biases of this conventional model. In Section 2.2, I rigorously explain the advanced *ETAS-Anisotropic* and *ETAS-Incomplete* model versions, that aim to solve the major model biases of the conventional model. Sections 2.3 and 2.4 summarize selected results and conclusions of the first (Grimm et al., 2021) and second contribution (Grimm et al., 2022a). Next, Section 2.5 briefly describes the computational efficiency of the code, and Section 2.6 shows an illustrative case study that demonstrates the impact of earthquake clustering on the annual loss curve of an insurance. Finally, Section 2.7 discusses alternative approaches.

2.1. Statistical Model

2.1.1. Hawkes Processes

Model formulation The ETAS model is a special case of a Hawkes process that models the spatio-temporal occurrence rate of earthquakes as a self-exciting Poisson point process with a four-dimensional, space-time-magnitude mark (Jalilian, 2019; Daley and Vere-Jones, 2003). Hawkes processes go back to their pioneer Alan Hawkes (1971) and were originally developed for modeling epidemics. The process describes the event rate at time t as a function of the occurrence history of events before that time, $\mathcal{H}_t = \{t_i \in \mathbb{R}_+ : t_i < t\}$. The intensity function is specified as

$$\lambda(t|\mathcal{H}_t) = \mu(t) + \sum_{t_i:t_i<t} \lambda_{trig}(t - t_i), \qquad t > 0, \tag{2.1}$$

where $\mu(t)$ is the baseline event rate (independent of previous occurrences) and $\lambda_{trig}(t - t_i)$ is the additional event rate at time t, triggered by the previous event i at time t_i. The infinitesimal probability of an event occurring in the time window $[t, t + dt)$ is $\lambda(t|\mathcal{H}_t)\ dt$. Given the current example of the Corona pandemic, the self-exciting nature of the point process modeling contagion times can be interpreted as a chain of infection, where an infected individual spreads the virus and therefore (temporarily) increases the risk of triggering new cases.

Applications Both the baseline and triggered event rate functions of the Hawkes process can assume various shapes tailored to the problem at hand. For that reason, Hawkes processes are applicable in any other field where an event occurrence temporarily increases the chance of a subsequent event, such as in mathematical finance (e.g. trading orders Hawkes, 2018), military and terrorism (Tench et al., 2016) or earthquake triggering (Ogata, 1988, 1998; Jalilian, 2019). In the latter, as briefly outlined in Section 1.1, the occurrence of an earthquake redistributes stress in the surroundings of the fault, typically triggering numerous further events close in time and space.

2.1.2. Conventional ETAS Model

The ETAS model was first introduced by Ogata (1988, 1998) and expands the temporal Hawkes intensity function (2.1) by a two-dimensional space component (x, y) and a magnitude-size component m (Daley and Vere-Jones, 2003). In the following, I introduce the *conventional* ETAS model as described and implemented in the *R* package *ETAS* by Jalilian (2019), but use notation similar to the one in my second contribution (Grimm et al., 2022a). In the following, let M_c be the cut-off magnitude of the analyzed earthquake catalog.

Model Formulation In the conventional ETAS model approach, the occurrence rate of an earthquake with *magnitude* m, occurring at *time* t and at *location* (x, y) is modeled by an inhomogeneous Poisson process with a time-space-magnitude dependent intensity function

$$\lambda(t, x, y, m|\mathcal{H}_t) = f_0(m)\, R_0(t, x, y|\mathcal{H}_t), \tag{2.2}$$

where $\mathcal{H}_t = \{(t_i, x_i, y_i, m_i) : t_i < t\}$ is the history of events prior to time t,

$$f_0(m) = \beta\, e^{-\beta(m-M_c)}, \qquad \beta > 0,\ m \geq M_c, \tag{2.3}$$

is the probability density function (pdf) of the frequency-magnitude distribution (FMD) with b-value equal to $\beta/ln(10)$ above lower cut-off magnitude M_c (Gutenberg and Richter, 1944), and

$$R_0(t, x, y|\mathcal{H}_t) = \mu\, u(x, y) + \sum_{i:t_i<t} R_0^{trigg}(t, x, y, i) \tag{2.4}$$

is the occurrence rate of events with magnitude $m \geq M_c$, at time t and at location (x, y). This event rate is modeled by a superposition of the time-invariant seismic background rate $\mu\, u(x, y)$, where $\mu > 0$ is the total rate of background events and $u(x, y)$ is a non-parametrical spatial pdf, and the sum of the trigger rate contributions $R_0^{trigg}(t, x, y, i)$ of all events i that occurred prior to time t.

Trigger Rate Contributions The trigger rate contribution of event i is computed as the product of the aftershock productivity of event i and the pdfs of the temporal and spatial aftershock distributions evaluated at time t and location (x, y),

$$R_0^{trigg}(t, x, y, i) = k_{A,\alpha}(m_i)\, g_{c,p}(t - t_i)\, h_{D,\gamma,q}(r_i, m_i, l_i). \tag{2.5}$$

The *aftershock productivity* function

$$\kappa_{A,\alpha}(m_i) = A\, \exp(\alpha(m_i - M_c)), \qquad m_i \geq M_c;\ \ A,\ \alpha > 0 \tag{2.6}$$

describes the expected number of *direct* aftershocks triggered by event i with magnitude m_i. Such an exponential growth of the productivity is in good agreement with observations (see e.g. the summary provided by Hainzl and Marsan (2008)).

The *temporal trigger function* $g_{c,p}(t - t_i)$ is based on the well-known empirical Omori-Utsu law

$$\tilde{g}_{c,p}(t - t_i) = (t - t_i + c)^{-p}, \qquad t \geq t_i; \quad c, \ p > 0, \tag{2.7}$$

describing the power-law decay of aftershock rates with increasing time t after the occurrence time t_i of the triggering event i (Utsu et al., 1995). To ensure that $g_{c,p}(t - t_i)$ is a pdf, the Omori-Utsu law $\tilde{g}_{c,p}(t - t_i)$ is normalized by the integral of aftershock rates in the time window $t - t_i \in [0, T]$, where T is the assumed maximum duration of aftershock triggering (in days; e.g. $T = 365$). The c-value defines the delay of the onset of the power-law decay (typically a few minutes to hours). It is likely related to short-time incompleteness of earthquake catalogs after mainshocks (Hainzl, 2016b). The p-value is in the range 0.8–1.2 in most cases (Utsu et al., 1995).

The *spatial kernel* $h_{D,\gamma,q}(r_i, m_i, l_i)$ models the 2D-distribution of aftershocks locations. In conventional ETAS model approaches, the triggering event is assumed to be a point source, distributing its offsprings isotropically around its epicenter. A classical definition of an isotropic kernel (see also Ogata, 1998) is

$$h^{iso}_{D,\gamma,q}(r_i(x,y), m_i) := \frac{q-1}{D\,\exp(\gamma(m_i - M_c))} \left(1 + \frac{\pi\, r_i(x,y)^2}{D\,\exp(\gamma(m_i - M_c))}\right)^{-q} \tag{2.8}$$

where $r_i(x, y)$ denotes the point-to-point distance between a potential aftershock location (x, y) and the coordinates (x_i, y_i) of the triggering event i, and m_i is the magnitude of the event i. The kernel is constrained by the parameters D and γ that control the magnitude-dependent width of the kernel, and parameter q that describes the exponential decay of the function with growing spatial distance.

Numerical Optimization Following the implementation in the R package *ETAS*, the parameter vector $\theta = \{\mu, A, \alpha, c, p, D, \gamma, q\}$ is optimized by maximizing the log-likelihood function

$$LL = \sum_j ln\left(R_0(t_j, x_j, y_j)\right) - \int_{\mathbb{T}} \iint_{\mathbb{S}} R_0(t, x, y|\mathcal{H}_t)\, dx\, dy\, dt. \tag{2.9}$$

where $\mathbb{T}$ and $\mathbb{S}$ denote the target time and space window over which the model is fitted. The parameters are optimized by the iterative, gradient-based Davidson-Fletcher-Powell algorithm.

2.1.3. Three Major Model Biases

Within the limits of being a non-physical model, the ETAS approach delivers overall convincing representations of earthquake clustering, and has substantially contributed to analyze and compare seismic clustering behavior, e.g. comparing different tectonic regions of the world (Chu et al., 2011). Nevertheless, the Standard ETAS model as described in Subsection 2.1.2 is built upon three simplifying assumptions that can lead to significant biases in the model estimation results.

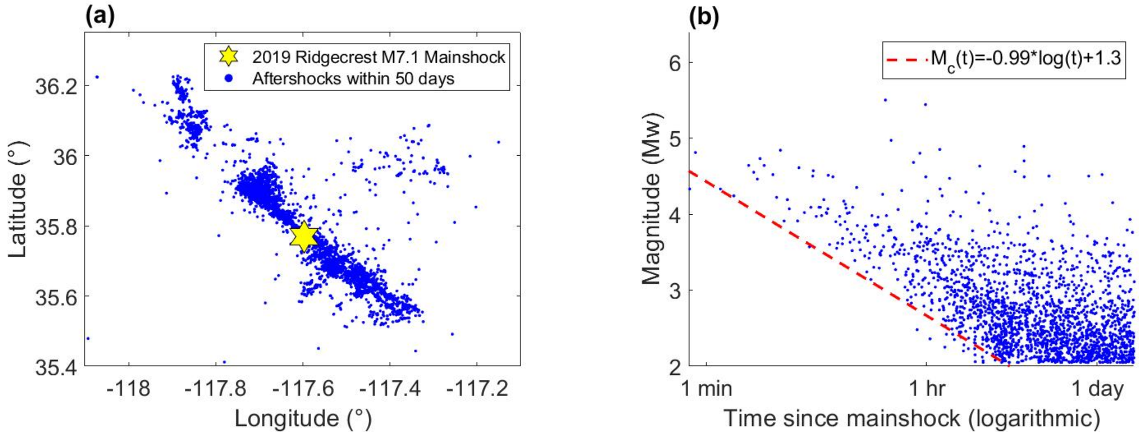


Figure 2.1.: Illustration of ETAS model biases for the example of the 2019 Ridgecrest Mw7.1 aftershock sequence. (a) Aftershock locations, showing a pronounced anisotropic shape of the spatial distribution; (b) Aftershock magnitudes vs. logarithmic event times, demonstrating short-term incomplete aftershock records for several hours.

Bias 1: Isotropic spatial distribution The spatial kernel defined in Equation (2.8) reflects the common assumption in ETAS models that aftershock locations are distributed isotropically around the triggering event, which is assumed to be a point source. This assumption is named as a shortcoming in many publications because it contradicts the observation that aftershocks tend to occur close to the typically elongate rupture plane of the triggering event (Ogata, 1998, 2011; Ogata and Zhuang, 2006; Hainzl et al., 2008, 2013; Seif et al., 2017; Zhang et al., 2018). The assumption of isotropy is reasonably valid for weak earthquakes with small rupture extensions, but becomes problematic for larger magnitudes, e.g. see the spatial pattern of the 2019 Ridgecrest sequence in Fig. 2.1(a). It has been shown that inadequate spatial models can lead to an underestimation of the productivity parameter α (Equation 2.6) because the numerous small events take over the role of mimicking the "true" anisotropic distribution (Hainzl et al., 2008, 2013).

Bias 2: Infinite spatial kernel In my first contribution (Grimm et al., 2021) I showed that, under the usual definition of an infinite range spatial kernel, aftershock triggering is disproportionately associated with the more numerous small events, that can more flexibly mimic anisotropic event alignments than the few strong mainshocks. This can lead to unrealistically far trigger impact of small magnitudes and thus to a substantial underestimation of the aftershock productivity of strong earthquakes (small α), resulting in a smoothing of temporal event occurrences.

Bias 3: Short-term aftershock incompleteness Strong earthquakes typically cause incomplete aftershock records immediately after their occurrence, mainly due to an overlap of coda waves (Hainzl, 2016b; de Arcangelis et al., 2018). Fig. 2.1(b) demonstrates this phenomenon for the aftershock sequence of the 2019 Ridgecrest M7.1 mainshock. Apparently, records of smaller sized aftershocks are missing in the first minutes to hours, somewhat foiling the power law decay of event rates expected from the Omori-Utsu law (Equation 2.7). The short-term incomplete event records can therefore hide to a large extent both the "true" Omori Law decay and the "true" aftershock productivity of the trigger event (Equation 2.6) and lead to an overestimation of Omori parameter c and an underestimation of productivity parameter α (Hainzl, 2021, 2016a; Page et al., 2016; Seif et al., 2017).

Further Uncertainties As every statistical model, the ETAS model relies on a sufficient sample size and quality of input data. The required input to an ETAS model is a so-called *earthquake catalog*, which is a list of earthquakes recorded in a specified region and time window and should minimally contain the occurrence time (accurate down to minutes or seconds), event location (in geographical coordinates) and magnitude.
While event times can typically be determined very precisely, location uncertainties are in the range of a few kilometers, depending on the quality and density of the network of seismic stations in the modeled area. Magnitudes may be specified in different scales and should be homogeneized. An in-depth discussion of ETAS model uncertainties, including a discussion of the assumption of a *time-invariant seismic background*, is given in Harte (2013, 2016); Seif et al. (2017).

2.2. Model Advancements

In the contributions Grimm et al. (2021, 2022a), I developed two advanced versions of the ETAS model to solve the estimation biases introduced above. Subsection 2.2.1 defines the ETAS-Anisotropic model with a generalized anisotropic, locally restricted spatial kernel. Then, subsection 2.2.2 introduced the formulation of the ETAS-Incomplete model that *additionally* accounts for short-term incomplete aftershock records.

2.2.1. ETAS-Anisotropic Model

Anisotropic Spatial Kernel In the contribution Grimm et al. (2021), I introduced an anisotropic generalization of the isotropic spatial kernel (Equation 2.8),

$$h_{D,\gamma,q}(r_i(x,y), m_i, l_i) := \frac{q-1}{D\,\exp(\gamma(m_i - M_c))} \left(1 + \frac{2\,l_i\,r_i(x,y) \;+\; \pi\,r_i(x,y)^2}{D\,\exp(\gamma(m_i - M_c))}\right)^{-q}. \tag{2.10}$$

In this spatial model, the distance term $r_i(x, y)$ denotes the point-to-line distance between the potential aftershock location (x, y) and the estimated rupture line segment of triggering event i with length l_i. Fig. 2.2 illustrates the different shapes of the two spatial kernels.

Note that

$$h_{D,\gamma,q}(r_i(x,y), m_i, 0) = h^{iso}_{D,\gamma,q}(r_i(x,y), m_i),$$

i.e. the anisotropic kernel collapses to the isotropic model if the triggering location is assumed to be a point source with rupture extension $l_i = 0$. Therefore, the generalized spatial kernel (2.10) allows for mixed approaches, modeling larger earthquakes anisotropically, but weaker events isotropically. The rupture length scaling relationship for anisotropic events is taken from the estimate of subsurface ruptures, provided in Wells and Coppersmith (1994).

Locally Restricted Spatial Kernel Both the conventional isotropic and the generalized anisotropic spatial kernel are defined in terms of a pdf over infinite space. Realistically, however, small earthquakes should exert only a locally restricted trigger influence.
A manual analysis of the spatial aftershock patterns of the six largest Californian mainshocks since 1987 (see Hainzl, 2021) shows that the cloud of potential aftershocks typically lies within one estimated rupture length from the epicenter. Acknowledging the particularly steep faulting systems prevailing in California, globally I suggest a radius of 2.5 rupture lengths, consistent with

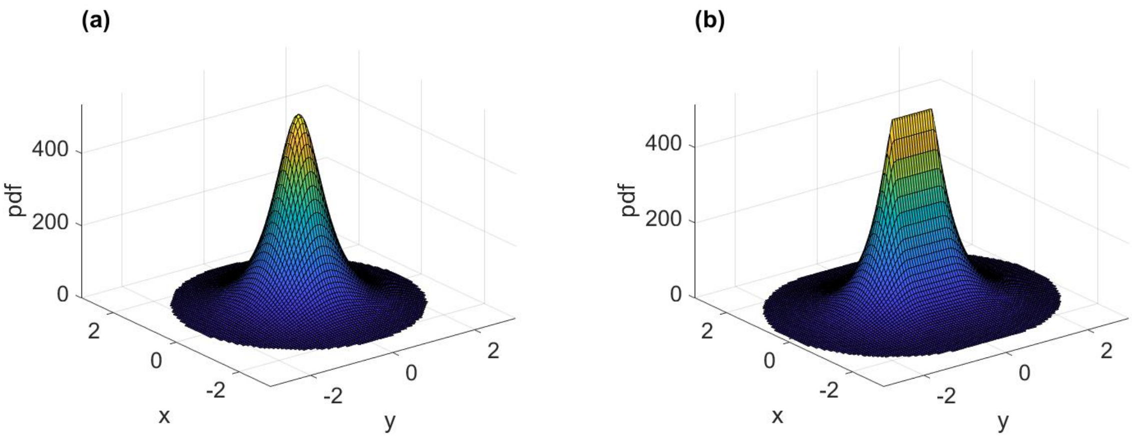


Figure 2.2.: Locally restricted (a) isotropic (Equation 2.8) and (b) anisotropic (Equation 2.10) spatial kernel.

Felzer et al. (2004). Given an arbitrary local restriction $R_i > 0$, the spatial kernel for event i is then only defined in the restricted area

$$\mathbb{S}_i(R_i) := \{(x,y) \in \mathbb{R}^2 | r_i(x,y) \leq R_i\}$$

and set to 0 outside of it. Note that the restricted area $\mathbb{S}_i(R_i)$ can assume isotropic and anisotropic shapes, depending on the point-to-point or point-to-line definition of the distance term $r_i(x,y)$. In order to retain the property of a pdf, we need to rescale the kernel within the restricted area by its analytical integral (for a derivation, see the ETAS formulary in the Appendix)

$$H_{D,\gamma,q}(R_i, m_i, l_i) := \iint_{\mathbb{S}_i(R_i)} h_{D,\gamma,q}(r_i(x,y), m_i, l_i)\, dx\, dy = 1 - \left(1 + \frac{2\, l_i\, R_i \,+\, \pi\, R_i^2}{D\, \exp(\gamma(m_i - M_c))}\right)^{1-q}.$$

The integral term holds true for both isotropic ($l_i = 0$) and anisotropic triggers ($l_i > 0$). We obtain the generalized, restricted and anisotropic spatial kernel

$$h^{restr}_{D,\gamma,q}(r_i(x,y), m_i, l_i) = \begin{cases} \frac{h^{restr}_{D,\gamma,q}(r_i(x,y), m_i, l_i)}{H_{D,\gamma,q}(R_i, m_i, l_i)}, & \text{if}\, r_i(x,y) \leq R_i, \\ 0, & \text{if}\, r_i(x,y) > R_i. \end{cases} \tag{2.11}$$

2.2.2. ETAS-Incomplete Model

Rate-Dependent Incompleteness The concept of rate-dependent earthquake record incompleteness assumes that the ”true” (i.e. physical) relationships underlying $f_0(m)$ and $R_0(t,x,y)$ are not accurately identifiable in available earthquake catalogs because especially small magnitudes are detected with lower probability in periods of high seismic activity. In these periods, the detection ability is limited typically due to overlapping seismic waves (Hainzl, 2016b, 2021).
Fitting the ”true” relationships to incomplete data records may therefore lead to significantly biased parameter estimates (Hainzl, 2016b,a; Page et al., 2016; Seif et al., 2017; Hainzl, 2021).

Model Formulation The working assumption of the ETAS-Incomplete model is that an earthquake with magnitude m, occurring at time t, can only be detected by the operating seismic network if no event of equal or larger magnitude occurred within the blind time $[t - T_b, t]$, where T_b is typically in the range of some seconds to few minutes (Hainzl, 2021). Similar assumptions have formerly been formulated by Lippiello et al. (2016), de Arcangelis et al. (2018) and Hainzl (2016b).

Let $N_0(t)$ be the expected number of events occurring within the entire spatial window $\mathbb{S}$ during blind time $[t - T_b, t]$,

$$N_0(t) = \int_{t-T_b}^{t} \iint_{\mathbb{S}} R_0(t,x,y) dx\, dy\, dt \approx T_b \iint_{\mathbb{S}} R_0(t,x,y)\, dx\, dy,$$

where the approximation holds under the assumption that event rates are approximately constant during the blind time (Hainzl, 2021). According to the "true" FMD (Equation 2.3), each of the $N_0(t)$ events has a probability of $e^{-\beta\,(m-M_c)}$ to exceed magnitude m. Then, the detection probability $p_d(m,t)$ of an earthquake at time t with magnitude m is the probability that no equal or larger event occurred during blind time T_b, i.e.

$$p_d(m,t) = e^{-N_0(t)\, e^{-\beta\,(m-M_c)}}.$$

Following the derivations in Hainzl (2016a, 2021), we obtain the "apparent", incompleteness-biased FMD

$$f(m,t) := f_0(m)\, N_0(t)\, \frac{p_d(m,t)}{1 - e^{-N_0(t)}}$$

and the "apparent" event rate

$$R(t,x,y) := \frac{R_0(t,x,y)}{N_0(t)} \left(1 - e^{-N_0(t)}\right).$$

The term "apparent" signalizes that the functions f and R do not represent the "true", but the observable relationships that are possibly distorted by short-term aftershock incompleteness. In periods of high seismic activity, the "apparent" FMD exhibits a larger relative frequency of strong events (because they are more likely to be detected) and an event rate lowered by detection capacity. We obtain the ETAS-Incomplete intensity function

$$\lambda(t,x,y,m) = f(m,t)\, R(t,x,y) = f_0(m)\, R_0(t,x,y)\, p_d(m,t)$$

The log-likelihood function of the ETAS-Incomplete model has the form

$$LL = \sum_j ln\,(\, f_0(m_j)\, R_0(t_j,x_j,y_j)\, p_d(m_j,t)\,) - \frac{T_2 - T_1}{T_b} - \frac{1}{T_b} \int_{T_1}^{T_2} e^{-T_b \iint_{\mathbb{S}} R_0(t,x,y)\, dx\, dy} dt$$

where the second term is an approximation of the integral of λ over the entire magnitude-space-time model space $[M_c, \infty) \times [T_1, T_2] \times \mathbb{S}$.

The two underlying, simplifying assumptions in the ETAS-Incomplete model are that (1) the blind time T_b is magnitude-independent, which Hainzl (2021) justifies by typically shorter source durations than travel times of coda waves, and (2) that the seismic network is equally occupied for blind time T_b by any event in the entire investigated spatial window. The second assumption is reasonable for a small spatial window, e.g. when analyzing an isolated sequence. When fitting the ETAS-Incomplete model over a larger region, it needs to be checked that relevant clusters do not evolve at the same time but at distinct locations as they would be assumed to simultaneously occupy the entire seismic network. A reasonable approach to prevent undesired biases is to choose a larger cut-off magnitude.

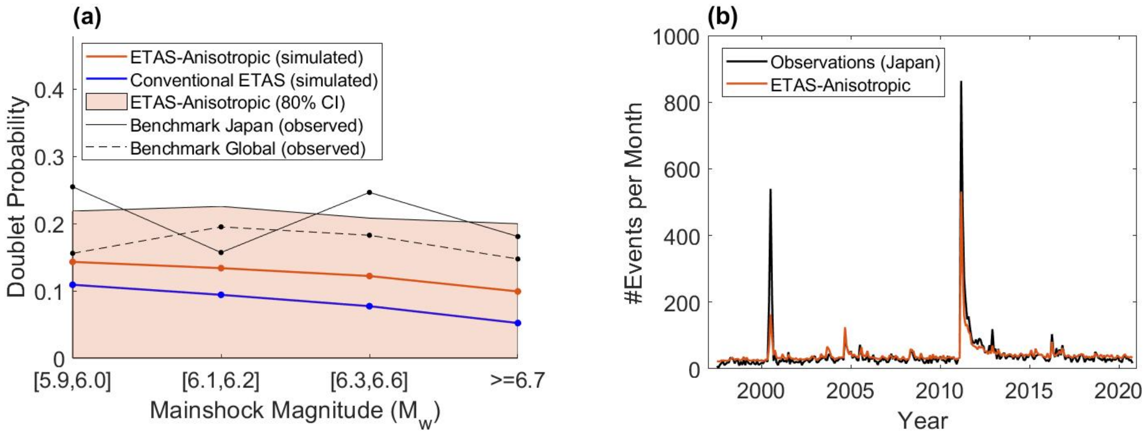


Figure 2.3.: (a) Comparison of doublet probability predictions predicted by the ETAS-Anisotropic model (red line) and the conventional ETAS model (blue line). The shaded range represents the 80% confidence interval for the anisotropic model. Observed doublet frequencies in a local and global catalog are shown by black solid and dashed lines, respectively. Mainshock magnitudes are grouped in intervals due to relatively small sample sizes in the benchmark catalogs. (b) Comparison of monthly event occurrences between the observed Japan catalog (black line) and the event rates predicted by the ETAS-Anisotropic model, given the observed event history at any time (red line).

2.3. Contribution 1: Improving Doublet Frequency Predictions

In the first contribution (Grimm et al., 2021), I developed the ETAS-Anisotropic model accounting for anisotropic and locally restricted aftershock distributions (see Subsection 2.2.1). The model was fitted to local earthquake catalogs for Japan and Southern California. Then, we used the estimated parameters to forward simulate 10,000 synthetic catalogs and statistically analyzed how well the models reproduce the spatio-temporal clustering of the original catalogs, respectively.
In particular, we compared the simulated probabilities that a mainshock triggers an earthquake doublet, with observations from the corresponding local and a global benchmark catalog. We used the doublet definition provided in Section 1.3.

Fig. 2.3(a) shows the predicted probabilities that a mainshock of a given magnitude interval triggers an earthquake doublet. The ETAS-Anisotropic model (red line) estimates that roughly 14% of the $M \geq 6$ mainshocks trigger a doublet. Despite an improvement compared to the conventional ETAS model (10%, blue line), it still underestimates the relative frequency of about 20% observed in the local and global benchmark catalogs.
The reason for this may be found in underestimated cluster sizes of strong mainshocks. As Fig. 2.3(b) shows, the ETAS-Anisotropic model underestimates the number of events in peak periods, but overestimates event rates in quiet periods. In other words, the model smooths out the temporal distribution of events.

Results indicate that the novel spatial kernel promotes more realistic estimates of cluster sizes by reducing the bias of inadequate spatial fits. Nevertheless, this contribution highlights the need to account for short-term aftershock incompleteness as the remaining major bias of ETAS estimations.

2.4. Contribution 2: Improving Forecasts of the 2019 Ridgecrest Sequence

In the second contribution (Grimm et al., 2022a), we developed the ETAS-Incomplete model (see Subsection 2.2.2) and combined it with the anisotropic spatial kernel of the ETAS-Anisotropic model (Subsection 2.2.1). The resulting model was fitted to a long-term earthquake catalog for Southern California. Then, we used the estimated parameters to perform 10,000 forecasts of the aftershock sequences of both the 2019 Ridgecrest M6.4 foreshock (July 4) and the M7.1 mainshock (July 6, see spatio-temporal pattern in Fig. 2.1).

Figure 2.4(a) depicts the empirical cumulative distribution functions of the number of aftershocks of the M6.4 foreshock event, for simulations based on the conventional ETAS (black line), the ETAS-Anisotropic (red line) and the preferred ETAS-Incomplete model (blue line), compared to the observed value (vertical gray line: 633 events). The conventional model provides entirely inappropriate estimates, and the ETAS-Anisotropic model underestimates the size of the sequence in more than 96% of the simulation runs. Only the ETAS-Incomplete model provides a very good prediction. This indicates that the ETAS-Incomplete model may solve the underestimation of cluster sizes by the ETAS model, observed in the first contribution (Grimm et al., 2021).

Fig. 2.4(b) shows the kernel density estimators for the simulated largest aftershock magnitude of the M6.4 event. It reveals that none of the models is capable to forecast the following M7.1 mainshock, that occurred 34 hours after the M6.4 foreshock. The reason is that, while the simulations were based on a magnitude distribution (see Equation 2.3) with $b = 1.01$, as estimated from the long-term California catalog, the M6.4 aftershock sequence was characterized by a particularly low b-value of 0.79, favoring the occurrence of stronger aftershocks.

Fig. 2.4(c) and (d) show the predicted spatial aftershock distributions of the M6.4 event, averaged over the 10,000 simulation runs of (c) the conventional, isotropic ETAS and (d) the anisotropic ETAS-Incomplete model. We overlaid the observed event locations (black scatter points) to the logarithmic heat map of predicted probabilities on a 1 km $\times$ 1 km grid. The M6.4 foreshock is a special case of anisotropic triggering in the sense that it simultaneously ruptured two almost orthogonal faults, leading to a double pattern of separate linearly elongate aftershock clouds (Marsan and Ross, 2021). Therefore, we further generalized the anisotropic spatial kernel (Equation 2.10) to allow for an equally weighted superposition of two rupture segments. The figures demonstrate that an anisotropic spatial kernel substantially improves the spatial forecast of the M6.4 sequence. Fig. 2.4(e) and (f) confirm this conclusion for the aftershocks of the M7.1 mainshock.

In summary, the ETAS-Incomplete provides a better understanding of the spatio-temporal evolution of earthquake sequences solves three major biases of the conventional ETAS model at once. Particularly, it leads to better forecasts of the size of a sequence and the spatial distribution of aftershocks. These improvements may be of major interest for short-term risk assessment during an on-going aftershock sequence, particularly for assessing the risk of a second, damaging earthquake. The anisotropic spatial forecast of aftershock locations enables desaster response managers to take actions in areas at risk where particularly high aftershock activity is expected.

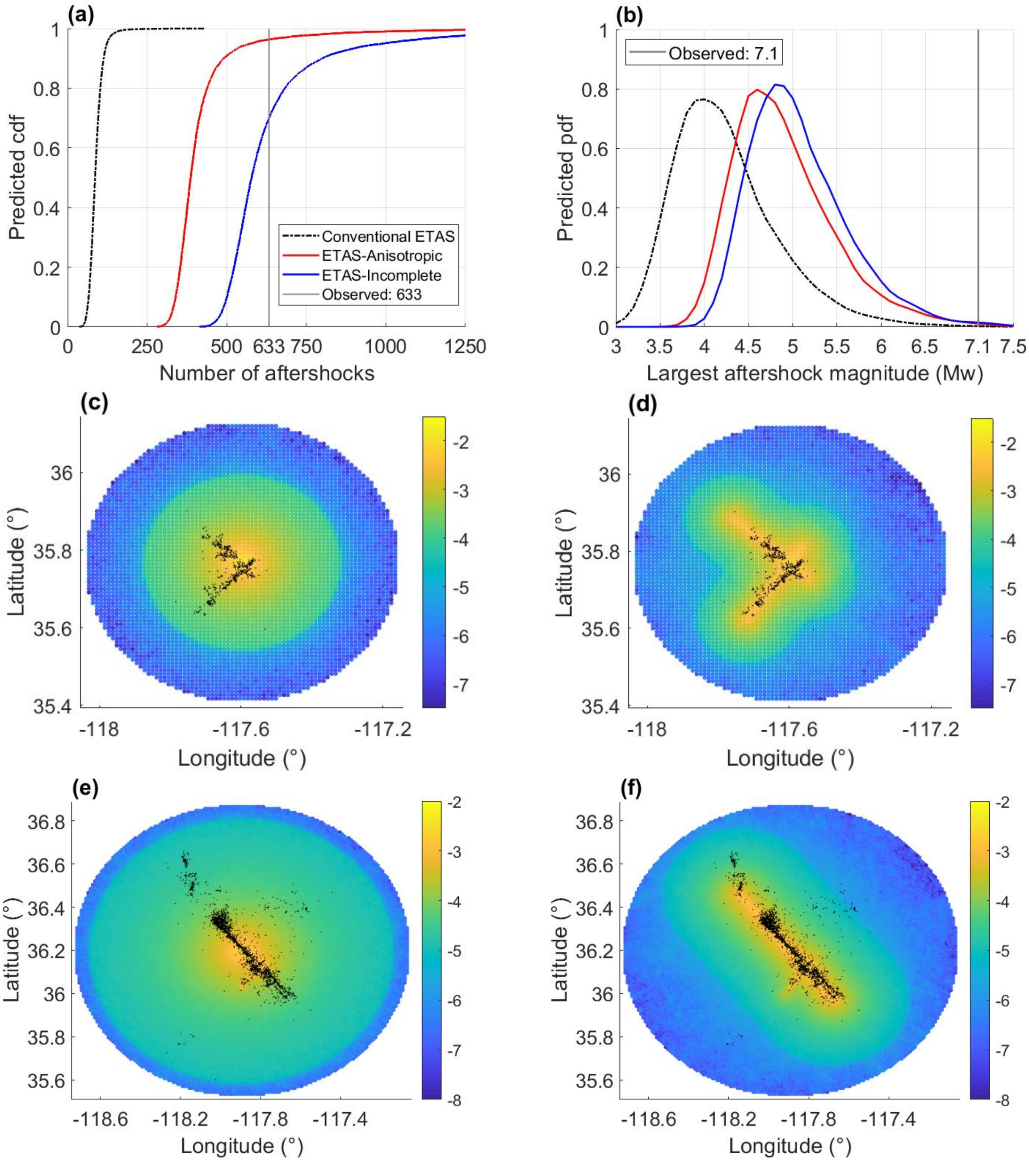


Figure 2.4.: (a) Predicted cumulative distribution functions of the number of aftershocks of the Ridgecrest M6.4 earthquake. (b) Kernel density estimators of the largest aftershock magnitude triggered by the Ridgecrest M6.4 earthquake. (c-d) Predicted spatial event distributions of the Ridgecrest M6.4 aftershock sequence, simulated by (c) the conventional ETAS model and (d) the ETAS-Incomplete model with anisotropic kernel. The color bar indicates the predicted, logarithmic probability that an event occurs at the respective grid point. (e-f) Same as (c-d), but for simulations of the Ridgecrest M7.1 event.

2.5. Computational Efficiency and Code

I made the corresponding ETAS estimation and simulation source code publically available in the Github repository `https://github.com/ChrGrimm/ETASanisotropic`. The code package is implemented in the software `Matlab` and enables the estimation of ETAS model parameters with very flexible, user-defined input settings, e.g.

- with (or without) ETAS-Incomplete functionality,
- with (or without) anisotropic spatial kernel
- anisotropy can be assumed for earthquakes above a user-defined magnitude threshold
- with a user-defined local restriction of the spatial kernel
- with a user-defined temporal restriction of triggering

The proposed local restriction of the spatial kernel avoids the computation of trigger rate contributions between distant events, thereby breaking the quadratic growth of runtime with increasing catalog size. In consequence, the code is computationally very efficient and can compute an ETAS model (without ETAS-Incomplete functionality) for a catalog of more than 10,000 events within only few minutes. The ETAS-Incomplete model is numerically more challenging, since the temporal integral cannot be calculated analytically and needs to be approximated. Still, an ETAS-Incomplete model estimation of a similarly sized catalog can be performed in a convenient time like an hour.
The Appendix of this dissertation (see Chapter 8) provides an ETAS model formulary, including a comprehensive summary of formulas and derivations of all partial derivatives needed to implement a gradient-based optimization process for various ETAS models.

2.6. Case Study: Earthquake Clustering in Insurance Loss Model

In this case study, I demonstrate the impact of earthquake clustering on the annual exceedance probability (AEP) curve which is one of the central results of an insurance loss model (see Section 1.2). Forward simulations of earthquake sequences are based on a fit of the ETAS-Incomplete model to a local earthquake catalog for Japan.

Algorithm to compute the AEP Curve:

1. Simulate 12,500 annual periods of independent mainshocks as a Poisson process
2. Simulate an aftershock sequence for each mainshock using the ETAS-Incomplete model
3. Determine event losses for mainshocks and aftershocks
4. For each period, compute the annual insurance loss as the sum of the individual event losses
5. Sort the periods by descending annual losses and determine the corresponding return periods: The i-th largest annual loss is exceeded i times in 12,500 simulated annual periods. The corresponding exceedance probability is $i/12,500$, and the expected return period $12,500/i$.
6. Plot the annual insurance loss as a function of the return period.

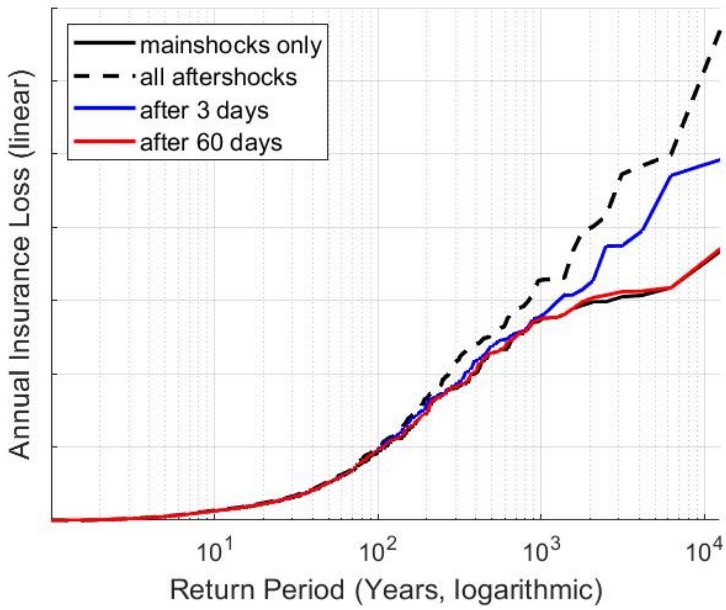


Figure 2.5.: Annual exceedance probability curves estimated from the case study. The various lines represent different assumptions on the contribution of aftershock losses (see text). The y axis is linear and shows annual losses. Labels are hidden for data protection reasons.

Determine Aftershock Losses Event losses of the simulated mainshocks are directly taken from the event loss table (ELT). The loss values provided by the ELT are based on historical loss data, seismic engineering models and expert opinion (Mitchell-Wallace et al., 2017).

At one extreme, we could similarly add event losses for *all* simulated aftershocks. However, this approach would be based on the assumption that the loss values in the ELT provide clean loss estimates for single events. In practice, damage from multiple events occurring within a short period of time in the same region can hardly be distinguished from one another and is therefore typically fully counted towards the mainshock event. Therefore, it is likely that the loss statistics in the ELT is biased by implicit clustering of losses.

To account for that, we only added losses of aftershocks that occurred more than three days after the mainshock or outside of its "damage region", which is defined as the area where the footprint of the mainshock predicts a peak ground acceleration larger than $1\frac{m}{s^2}$. The time criterion is inspired by the typical 72 hours clause in the definition of an earthquake "event" in the financial (re-)insurance terms (see Section 1.2). A stricter time criterion is 60 days, which represents the expected time that loss adjusters may need to quantify the damage of the mainshock.

Fig. 2.5 shows the resulting AEP curves. For data protection reasons, the curves are based on a randomized ELT and the (linear) y axis is hidden. The black solid line represents the annual insurance loss expected from mainshocks only. If adding *all* aftershock losses (dashed black line), the annual losses are 7-10% larger at return periods smaller than 200. In the less frequent scenarios, the additional loss through aftershocks would increase substantially. If we add aftershocks only after a three days waiting time, the effect is dampened to less than 5% for return periods smaller than 1000. If applying the stricter criterion of 60 days, the increase of annual losses is limited to about 2-3% across all return periods.

2.7. Alternative Approaches

Alternatives for modeling earthquake clustering include purely statistical or geophysical as well as mixed models. Statistical approaches include Markovian arrival processes (Bountzis et al., 2021; Bountzis and Tsaklidis, 2021; Bountzis et al., 2022) and hidden Markov models (Wu, 2010; Yip et al., 2018). Geophysical models typically consider stress changes induced by seismicity. For instance, Pope and Mooney (2020) investigated Coulomb stress changes for the 2019 Ridgecrest sequence, which was under statistical consideration in my second contribution (Grimm et al., 2022a). The Third Uniform California Earthquake Rupture Forecast (UCERF3) ETAS model, developed by the United States Geological Survey (USGS), combines the statistical ETAS model with regional, geophysical knowledge such as fault locations and interactions (Field et al., 2017). Long-term recurrence intervals of strong earthquakes, excluding short-term triggering, are often modeled by Brownian passage time distributions (e.g. Zöller, 2018).

Different ETAS model estimation approaches are possible. Instead of using the conventional maximum likelihood based optimization, parameters can be estimated by the expectation maximization algorithm of Veen and Schoenberg (2008), which is said to be more robust with respect to poor choices of initial parameters. However, in the context of my contributions, various sensitivity tests confirmed stable parameter estimation results for different starting values, using the maximum likelihood approach. Molkenthin et al. (2022) and Schneider (2021) proposed Bayesian inference procedures, which are less dependent on initial values, allow to account for prior knowledge of the spatio-temporal clustering behavior and provide stable estimates of parameter uncertainties. The Bayesian approach seems particularly attractive when fitting the model to a local sequence in a region which has been analyzed by a long-term clustering model before.

A few alternative approaches to model anisotropic spatial aftershock triggering have been proposed by other authors. Zhang et al. (2018) pursued an approach that assumed constant trigger rate in the entire rupture plane, with power-law decay outside of it. Different versions of elliptic Gaussian distributions were introduced and discussed by Ogata (1998, 2011) and Ogata and Zhuang (2006). The latter approaches successfully modeled spatial aftershock patterns, however, they require a new set of parameters and are therefore not flexibly combinable with the conventional, isotropic functionality. In contrast, the generalized spatial kernel suggested in the contributions to this dissertation allows for simultaneous anisotropic modeling of some events (e.g. above a certain magnitude threshold) and isotropic modeling of the rest. In order to address the abovementioned particularity of the M6.4 Ridgecrest foreshock, rupturing two almost orthogonal faults, the kernel can additionally reflect a weighted superposition of two distinct rupture line segments.

In recent years, there has been growing research interest in how to account for short-term incomplete datasets. For instance, Zhuang et al. (2017) developed a replenishment algorithm to fill up likely incomplete time intervals by simulated events, in order to obtain artificially complete pseudo-records. A rather simple workaround approach is to remove likely incomplete time periods from the fitted time interval using empirical completeness functions, such as performed in Hainzl et al. (2013). Other authors, particularly mentionable Omi et al. (2013, 2014), Lippiello et al. (2016), de Arcangelis et al. (2018), Mizrahi et al. (2021) and Hainzl (2021), tried to incorporate aftershock incompleteness directly into the ETAS model fit. For simplicity, Hainzl (2021) neglected the space dimension in his model. As my second contribution Grimm et al. (2022a) combines the ETAS-Incomplete time model of Hainzl (2021) with an adequate, anisotropic spatial kernel, it can be seen as the space-including extension of the latter.

3. Approach 2: Statistical Regression Models

This chapter summarizes the third contribution (Grimm et al., 2022b), which consists of two studies. First, I presented the innovative approach of adapting survival regression models to analyze the magnitude difference between a mainshock and the second strongest event of an earthquake sequence. Additionally, I fitted a generalized additive model (GAM) to investigate the aftershock productivity in a subduction region in New Zealand. All statistical analyses were performed with the open source software `R` (R Core Team, 2021).
Section 3.1 provides a compact introduction to the statistical models. In Section 3.2, I summarize the data compilation, model formulation and selected results of the survival model for magnitude differences. Section 3.3 gives an overview of the GAM for the aftershock productivity. Finally, Section 3.4 discusses alternative approaches in the literature. In this chapter, I use the terms *sequence* and *cluster* synonymously.

3.1. Statistical Models

3.1.1. Parametric Survival Models

This compact overview of the main types of survival model approaches is based on Klein and Moeschberger (2003). For a more complete overview, I refer the reader to the aforementioned book.
Survival models are a class of regression models that account for data with a censored (or truncated) response variable. As the term "survival" suggests, these models were originally developed in applications where the response represents the non-negative lifetime of a patient in medical studies or the lifetime of a device in engineering contexts (so-called *failure time analysis*). The above applications have in common that the exact value of the response is unknown, if the event has not occurred until the end of the study period. This is called *right-censoring.*

Non-Parametric Models Non-parametric models make no assumption about the exact shape of the distribution of the response. They include descriptive statistics such as the Kaplan Meier, life table or Nelson-Aalen estimators, for instance to analyze the median and other quantiles of lifetime based on a single categorical factor of interest. However, they do not provide effect size estimates and are limited to univariate models.

Semi-Parametric Models The most commonly used semi-parametric approach is the *Cox Proportional Hazard* model. Let $S(t) = P(T > t)$ denote the *survival function* of the lifetime T, and $f(t)$ be the probability density function (pdf), then the Cox model describes the event hazard function $h(t) = f(t)/S(t)$ by

$$h(t|\mathbf{X}) = h_0(t)\ \exp(\mathbf{X}\beta) = h_0(t)\, exp(\beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p)$$

where $\mathbf{X}$ is the covariate matrix and $p > 0$ is the number of covariates. The baseline hazard $h_0(t)$ is modeled non-parametrically, i.e., no particular assumption is made for the statistical distribution of survival times. The proportional hazards assumption means that covariate effects scale the entire hazard function up or down by the same factor time-independently. Testing approaches for this central assumption comprise Kaplan-Meier survival curve estimates or an analysis of Schoenfeld residuals against time.

Fully-Parametric Models In fully parametric models, the baseline hazard in the hazard function is specified, which requires the additional assumption on a specific statistical distribution of the lifetime T. Let the pdf of this distribution be given by

$$f(t|\mu(\mathbf{X}), \alpha(\mathbf{X})), \qquad t \geq 0,$$

with scale parameter μ and shape parameters $\alpha = (\alpha_1, ..., \alpha_R)$. Both scale and ancilliary parameters can be flexibly modeled depending on the scalar covariate matrix $\mathbf{X}$ by

$$g_0(\mu(\mathbf{X})) = \beta_0 + \sum_{k=1}^{K} f_{k0}(\mathbf{X}_k), \tag{3.1}$$

$$g_r(\alpha_r(\mathbf{X})) = \beta_r + \sum_{k=1}^{K} f_{kr}(\mathbf{X}_k), \tag{3.2}$$

where $\beta_0, \beta_1, ..., \beta_R$ are the intercept coefficients, $g_0(\cdot), ..., g_R(\cdot)$ are link functions and $f_{k0}(\cdot), ..., f_{kR}(\cdot)$ denote functions that specify the effect structure of the k-th covariate $\mathbf{X}_k$ (Jackson, 2016).

The fully parametric approach provides a more informative model, for instance enabling the use for predicting hazard rates as well as mean and median survival times. If the parametric distribution is correctly specified, parametric models have more power than semi-parametric approaches and lead to smaller standard errors. However, they are sensitive to miss-specifications of the distribution.

3.1.2. Generalized Additive Models

GAMs are an extension of generalized linear model approaches, allowing for flexible nonlinear covariate effects (Hastie and Tibshirani, 1990). The following short overview is based on the comprehensive works of Wood (2017) and Fahrmeir et al. (2013), section 5.2.
Let the conditional scalar response $Y_i|\mathbf{X}_i$ be distributed according to an exponential family distribution $F(\mu_i, \nu)$, where $\mu_i = \mathbb{E}(Y_i|\mathbf{X}_i)$ is the conditional mean and ν denotes further parameters of the distribution, given the scalar covariate matrix $\mathbf{X}$ and observation index i.
Then, given the link function $g(\cdot)$, the GAM estimates an intercept β_0 and potentially smooth functions $f_k(\cdot)$ for the k-th covariate $\mathbf{X}_k$ according to

$$g(\mu_i) = \beta_0 + \sum_{k=1}^{K} f_k(\mathbf{X}_{ki}), \qquad i = 1, ..., N \tag{3.3}$$

where N is the sample size. The unspecified smooth functions are typically modeled by penalized splines based on a basic spline basis (P-Splines), penalizing the second-order differences between coefficients of adjacent basis functions. (Fahrmeir et al., 2013, section 8.1).

3.2. Contribution 3a: Survival Models for Magnitude Differences

In the first study of the third contribution (Grimm et al., 2022b), I presented the approach of using a survival regression model to estimate the parametric distribution of the magnitude differences ΔM between a mainshock and its strongest aftershock (or foreshock).

3.2.1. Data Compilation

The underlying dataset for this study was extracted from a global earthquake catalog, with events between 1973 and 2021, cut-off magnitude $M_c = 5.0$ and depths $D \leq 70$ km, that occurred close to a tectonic plate boundary. We declustered this catalog using a window method (see e.g. van Stiphout et al., 2012; Uhrhammer, 1986; Gardner and Knopoff, 1974) with a time window of 100 days and a magnitude-dependent spatial radius of $R(m) = 2.5\, L(m)$, where $L(m)$ is the expected earthquake rupture length according to Wells and Coppersmith (1994). In doing so, we obtained a set of 2,933 *independent* earthquake sequences (clusters) with mainshock magnitudes $M > 6.0$. Foreshocks and aftershocks are complete down to M_c.
For each cluster, we computed the magnitude difference ΔM between the mainshock and the second-strongest event and defined it as the response variable. Additionally, we enriched the cluster dataset by additional geophysical site information such as a classification of *plate boundary types* as well as an estimation of the *relative plate velocity*, the *sea floor age* (all from Bird, 2003) and the *heat flow* (Bird et al., 2008).

3.2.2. Why Using a Survival Model for Magnitude Differences?

The magnitude difference ΔM of a cluster is only known, if at least one foreshock or aftershock was observed and assigned to the mainshock. Indeed, 1,180 out of 2,933 clusters are *single-event* sequences, i.e., no associated event was found in the corresponding time-space window. Based on seismological reasoning we can assume that these mainshocks actually triggered aftershocks, but that those had magnitudes smaller than M_c and were not recorded. For these clusters, we have the partial information that $\Delta M_i > M_i - M_c$, where M_i is the magnitude of mainshock i, i.e., the observations are right-censored (Klein and Moeschberger, 2003, section 3.2).
Our data meets the necessary requirements of a survival model, as it has *non-negative* ($\Delta M \geq 0$) and *independent* responses (declustered catalog) and the censoring is *non-informative*, i.e., censored clusters are not suspected to deviate structurally in their ΔM-distribution from non-censored clusters. Classical statistical models would substantially underestimate ΔM due to the large proportion of censored observations.

3.2.3. Choice of Distribution

The magnitude of the strongest aftershock can be interpreted as the largest order statistics from the sample of magnitudes in the aftershock sequence. Consequently, the main drivers of the magnitude difference ΔM between the mainshock and the second strongest event are (1) the number of aftershocks (hereafter called *aftershock productivity*) and (2) the frequency-magnitude distribution (FMD) of the triggered events. A simple simulation experiment shall illustrate the effect of both drivers on the distribution of ΔM.

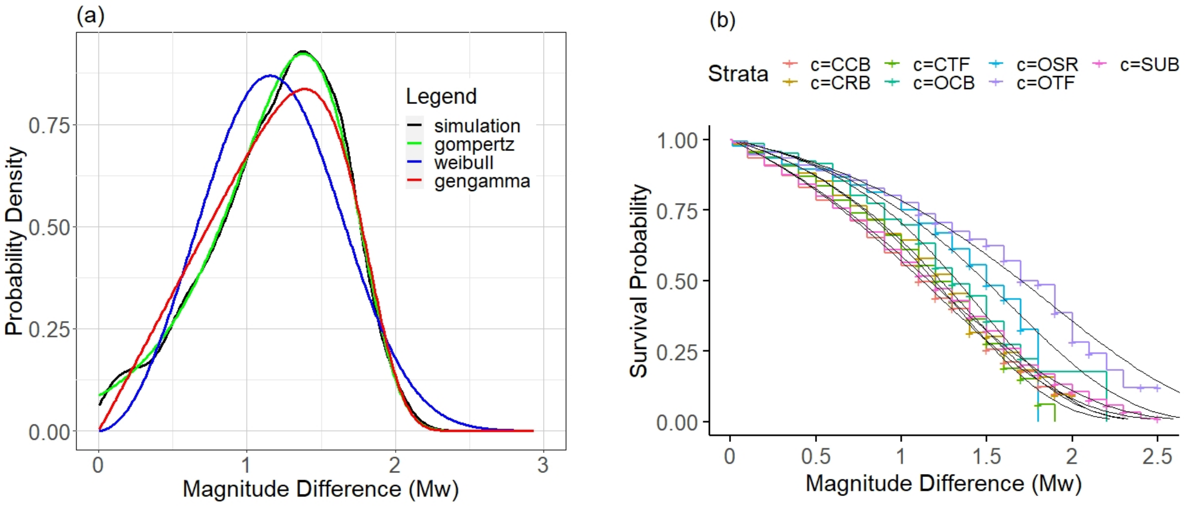


Figure 3.1.: (a) Fits of a Gompertz, Weibull and Generalized Gamma distribution to simulated magnitude differences ΔM, represented by the kernel density estimator (black curve). (b) Comparison of survival curves estimated from a Gompertz model and a non-parametric Kaplan-Meier estimator, stratified for plate boundary classes c.

Assume an initial mainshock of magnitude $M = 8$. For simplicity, let the number of aftershocks be Poisson distributed with a magnitude-dependent parameter $\lambda(M)$, being specified by the aftershock productivity function $\lambda(M) = A\, e^{\alpha\,(M-M_c)}$ (see Equation 2.6). Additionally, let the FMD of aftershocks be an exponential distribution with pdf $f(M) = \beta\, e^{-\beta(M-M_c)}$ (see Equation 2.3), where the Gutenberg-Richter b-value is $b = \beta/ln(10)$ (Gutenberg and Richter, 1944).
If we assume realistic parameters $A = 0.13$, $\alpha = 2.0$ and $b = 1.0$, we can simulate earthquake sequences including secondary triggering and derive a sampled distribution of the ΔM with a mean of 1.2, i.e. consistent with Bath's Law. Fig. 3.1(a) shows the fits of a Gompertz, Weibull and Generalized Gamma distribution to the kernel density estimator of the simulated magnitude differences. The Gompertz distribution clearly provides the best fit to the moderately negatively-skewed data. The distribution assumption is supported based on the actual dataset by fitting a univariable Gompertz survival model for the categorical variable *plate boundary class*, and comparing the predicted survival curves to those provided by the non-parametric Kaplan-Meier estimator (Klein and Moeschberger, 2003, ch. 4). Fig. 3.1(b) shows generally good agreement between the two approaches.

3.2.4. Model Formulation and Software

The Gompertz distribution is defined on $(0, \infty)$. Therefore, data points with $\Delta M = 0$ were substituted by the value 0.01. In the `R` package *flexsurv* (Jackson, 2016), the Gompertz distribution is parameterized by its probability density function

$$f(x|a, b) = be^{ax} \exp\left(-\frac{b}{a}(e^{ax} - 1)\right)$$

with shape parameter $a \in \mathbb{R}$ and scale parameter $b > 0$. We regressed the scale parameter b by all variables. Effects of plate boundary classes are linear, and the effects of the metric covariates

(mainshock magnitude, depth, relative plate velocity, heat flow and sea floor age) are modeled by penalized spline functions. The shape parameter a was additionally modeled only depending on the linear effects of the plate boundary class.
In this work, we fitted models using the function *flexsurvreg* from the *flexsurv* package, which estimates parameters by optimizing a parametric likelihood adapted for censored data (Jackson, 2016). To model flexible non-linear effects, we chose penalized B-splines using the function *pspline* from the `R` package *survival* (Therneau, 2016), consistenly specifying $df = 2$ degrees of freedom and $n = 2.5 \times df$ splines in the basis (Eilers and Marx, 1996; Hurvich et al., 1998).

3.2.5. Selected Results

The regression results show that larger ΔM are expected at higher depths and areas with large heat flow and young sea floor age, which are typically found in oceanic spreading ridges and transform faults. These observations may be an indication that aftershock productivity is a relevant driver of ΔM, as under these conditions lower aftershock productivity is expected due to reduced seismic coupling (Hainzl et al., 2019). The mainshock magnitude shows no structural effect, which confirms the hypothesis of the Bath's Law that the average magnitude difference is independent of the absolute size of the mainshock. After consideration of the heat flow and sea floor age effect, plate boundary classes show no further particularities.
In summary, the model suggests effects that explain *increased* magnitude differences, but it is not capable in adequately predicting the occurrence of particularly small ΔM such as earthquake doublets ($\Delta M \leq 0.4$). For future research, I therefore suggest to include more local, high resolution covariates or compile event-specific properties such as stress drop or the size of the rupture.

3.3. Contribution 3b: Modeling Aftershock Productivity

In the second study of the third contribution (Grimm et al., 2022b), I used a GAM to analyze the variation of aftershock productivity, estimated from a local catalog declustered by the ETAS-Incomplete model (see Section 2.2 and Grimm et al., 2022a).

3.3.1. Data Compilation

For this study, we chose a local earthquake catalog for the Hikurangi subduction zone in New Zealand, comprising 11,091 events between 1987 and the end of 2020, at depths down to 80 km and above cut-off magnitude $M_c = 3.5$.
For the regression of aftershock counts, we cannot use the window declustering method, as it does not distinguish *direct* from *secondary* aftershocks. Instead, we used the ETAS model based stochastic declustering approach, introduced by Zhuang et al. (2002). From Equation (2.4), they concluded that the probability, that the event j at time t_j and location (x_j, y_j) was an aftershock of the prior event i, is

$$P_{j,i} = \frac{R_0^{trig}(t_j, x_j, y_j, i)}{R_0(t_j, x_j, y_j | H_t)}$$

Similarly, the probability that event j is an independent seismic background event is

$$P_{j,backgr} = 1 - \frac{\sum_{i:t_i<t} R_0^{trigg}(t,x,y,i)}{R_0(t_j,x_j,y_j|H_t)} = \frac{\mu\, u(x,y)}{R_0(t_j,x_j,y_j|H_t)}.$$

We defined the response variable as the estimated *number of direct aftershocks* for each event i in the catalog. We did this by counting the number of subsequent events j for which i is the most probable trigger event, i.e. $P_{j,i} > P_{j,k} \;\; \forall\, k \neq i$, and that are more likely triggered by i than being a background event, i.e. $P_{j,i} > P_{j,backgr}$. In order to estimate the probabilistic trigger relations, we used the ETAS-Incomplete model from the second contribution (Grimm et al., 2022a), see also Section 2.2. By accounting for short-term incomplete aftershock records, this model avoids a substantial bias on the estimation of our response variable.

Covariate data includes the magnitude and depth of the triggering earthquake, a classification of events into tectonic region categories as well as an estimation of the slip types based on available local focal mechanism data. Additionally, if the triggering event i was itself already triggered by a previous event, we traced back the trigger sequence and identified the largest magnitude in the cluster, that occurred before event i. This covariate tests whether a triggered earthquake is itself more or less productive than an independent background event, and whether its aftershock productivity is influenced by the previous mainshock magnitude.

3.3.2. Model Formulation and Software

For modeling the aftershock count data, we tested a Poisson, Quasi-Poisson and Negative Binomial distribution as well as zero-inflated approaches to fit the data. According to Equation (3.3), we modeled the corresponding expected value with linear effects for the categorical variables tectonic region and slip type, as well as unspecified smooth effects for the magnitude and depth of the triggering event and the prior mainshock magnitude, if the triggering event was triggered itself.

To fit the model, we used the function *gam* from the `R` package *mgcv* (Wood, 2017), using a logarithmic link function and the restricted maximum likelihood estimator (REML) for the smoothing parameter estimation. We used function *s* from the *mgcv* package to setup the smooth terms based on P-splines, choosing $k = 5$ and $k = 8$ (for depth) as the dimensions of the basic spline basis.

3.3.3. Selected Results

The results confirm that aftershock counts can be better modelled by a Negative Binomial distribution (with dispersion parameter 2.3) rather than a Poisson distribution, as has already been suggested by Kagan (2017) and Shebalin et al. (2018). Alternative approaches such as a Quasi-Poisson or a zero-inflated model did not stand out substantially from the respective basic models. From a substantive point of view, there seems to be no causal reason for ”excess zeros” that would suggest zero-inflation.

Crustal events show an almost doubled aftershock productivity. A possible explanation for this effect is the rather dense network of faults in the modeled crustal region, that could be brought closer to failure by a change in stress conditions due to mainshock earthquakes. Slip types show no structural effects on the aftershock productivity.

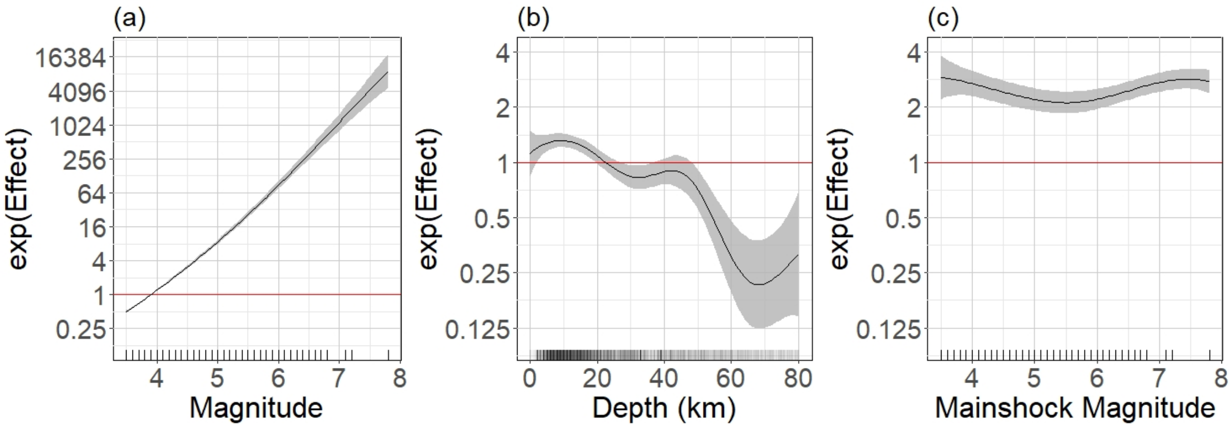


Figure 3.2.: Exponential, multiplicative effects of the metric covariates (a) *magnitude* (y axis is log2-transformed) and (b) *depth* of the triggering event as well as (c) *mainshock magnitude*, given that the triggering event was already part of a triggered sequence. Rug lines on the x axis visualize the marginal distributions of the corresponding covariate.

Fig. 3.2(a) shows that the number of expected aftershocks grows exponentially with the triggering magnitude. However, this effect is anyways enforced by the declustering approach that applies the exponential aftershock productivity function (2.6) in the ETAS parameter optimization procedure. The aftershock productivity decreases substantially for events deeper than approximately 45 km (see Fig. 3.2(b)), which supports the argumentation in the discussion of the ΔM regression results, that increasing magnitude differences may be related to reduced aftershock productivity at higher depths.

Finally, Fig. 3.2(c) shows that, independently of the mainshock size, triggered events appear to be generally two to three times more productive than a comparable background event. This finding has two possible explanations. On the one hand, it may be an indicator that the ETAS based declustering does not properly disentangle trigger chains in the catalog, and incorrectly overestimates secondary aftershock triggering at the cost of the mainshock. Such a rearrangement of trigger relations would have a strongly distorting effect on our model.

On the other hand, Zhuang et al. (2004) proposed that triggered events are more productive than background events, based on a similar study. It seems reasonable that during an on-going sequence the aftershock productivity could temporarily increase due to a higher level of energy prevalent in the tectonic system. A doubling of the productivity parameter A in the simulation model in Subsection 3.2.3, applied only to secondary triggering, would lead to a reduction of the expected magnitude difference ΔM from 1.2 to below 0.9 due to the increasing cluster sizes. This additional "boost" in triggering illustrates the relevance of the observed effect. The finding may also contribute to an explanation as to why the ETAS model tends to underestimate cluster sizes and doublet probabilities in forward simulations, as observed in my first contribution (Grimm et al., 2021). Further research is recommended to evaluate this finding.

3.4. Alternative Approaches

This work consists of two regression studies. The focus is on the innovate approach to estimate a fully parametric distribution of ΔM, using survival models that take into account right-censored data. To my knowledge, no similar approach has been made in literature so far. The chosen covariates represent rather large-scale regional effects. Attempts to consider small-scale variations of these covariates or to include further event specific data are out of the scope of this dissertation, but are recommended for future research.

Extensive research has been done on analyzing the variation of aftershock productivity. Kagan (2017) and Shebalin et al. (2018) showed that aftershock counts are best modeled by the Negative Binomial distribution due to their large variance. This assumption was confirmed by the results in this contribution. Page et al. (2016) found that aftershock productivity may regionally vary by a factor of almost 10, which would explain the variation in ΔM to large extent. To account for variation during on-going sequences, they suggested a Bayesian updating approach for sequence forecasts. Marsan and Helmstetter (2017) found that 40-80% of the aftershock variability may be related to variation in the mainshock stress drop. Dascher-Cousineau et al. (2020) investigated a large number of source and site effects on aftershock productivity using various machine learning algorithms and confirmed individual correlations of stress drop and rupture dimension with the number of aftershocks. Wetzler et al. (2016) suggested a larger productivity in subduction zones of the western circum-Pacific, compared to the eastern side. Based on an ETAS model approach, Zhuang et al. (2004) proposed that triggered events produce more aftershocks than comparable background events.

4. Overarching Conclusions

This dissertation contributes to a better understanding of earthquake doublets and the spatio-temporal evolution of earthquake sequences by two innovative approaches. In my contributions Grimm et al. (2021, 2022a), I developed advanced versions of the ETAS model. The ETAS-Anisotropic model (see Subsection 2.2.1) generalizes the spatial aftershock distribution, conventionally assumed to be isotropic, to more adequately reflect the observed elongate shape of aftershock clouds by applying an anisotropic, locally restricted spatial kernel. The ETAS-Incomplete model (Subsection 2.2.2) *additionally* accounts for typically short-term incomplete aftershock records, and therefore solves three of the major ETAS model biases at once.
In the third contribution, I proposed the innovative approach of using survival regression to model the right-censored observations of the magnitude differences ΔM between the mainshock and the second strongest event of the sequence. A generalized additive model was fitted to estimated aftershock counts, suggesting a substantially increased aftershock productivity of triggered earthquakes.

The contributions provide the following answers to the research questions stated in Section 1.4:

1. *How frequent are earthquake doublets worldwide?*
   The contribution Grimm et al. (2021) shows that roughly 20% of the global $M \geq 6$ mainshocks trigger earthquake doublets. In Japan, it tend to be slightly more. Despite being based on different definitions of a doublet, these values are consistent with Kagan and Jackson (1999) who found that approximately 22% of the global $M \geq 7.5$ earthquakes occurred in doublets.

2. *In which tectonic regions, and under which geophysical conditions, are doublets most likely to occur?*
   This question cannot be answered conclusively. The regression study in the contribution Grimm et al. (2022b) shows that deep earthquakes and events in oceanic spreading ridges or transform faults typically have larger ΔM. The contribution could not indicate variables that could explain particularly small ΔM.

3. *Which are the main drivers of earthquake doublet occurrences?*
   The contributions Grimm et al. (2021, 2022b) interpret the strongest aftershock as the largest order statistics of a sample of N aftershocks with magnitudes sampled from a common frequency-magnitude distribution (FMD). Thus, the main drivers of ΔM are the aftershock productivity and the b-value governing the Gutenberg-Richter type FMD. Contribution Grimm et al. (2022a) showed that an overestimation of b lead to a substantial underestimation of the largest magnitude triggered by the 2019 Ridgecrest M6.4 sequence. Contribution Grimm et al. (2022b) demonstrated that a doubled aftershock productivity decreases the expected ΔM from 1.2 to below 0.9.

4. *Can we develop an adequate model to predict long-term occurrence probabilities of doublets?* Yes, the advanced ETAS-Anisotropic and ETAS-Incomplete versions proposed in this dissertation can be used to estimate clustering in a catalog, to forward simulate synthetic clustered catalogs and to analyze and predict doublet occurrence probabilities. The conventional ETAS model provides poor estimates, mainly due to its assumptional biases that are solved by my suggested versions.

5. *Can we develop an adequate model to forecast the spatio-temporal evolution of a particular earthquake sequence?*
   Yes, the advanced ETAS-Anisotropic and ETAS-Incomplete versions can be used to forecast local sequences, as shown on the 2019 Ridgecrest sequence in the contribution Grimm et al. (2022a). Especially, the spatial forecasts show good agreement with observations, which may be of major interest for short-term risk assessment by desaster response managers.

Future Research This dissertation has focused on the impact of aftershock productivity on the occurrence of earthquake doublets. Future work should investigate the impact of the FMD by exploring potential correlations between the magnitudes of an aftershock and its mother event, as have been proposed by Gulia et al. (2018) and Nandan et al. (2019). Positively correlated magnitudes would increase the chance of a doublet occurrence after a strong mainshock.
Based on such an analysis, it would be interesting to identify whether small magnitude differences ΔM are typically characterized rather by above-average aftershock productivity or by magnitude size distributions favoring large aftershocks. To do so, one could compile a sufficiently large set of global earthquake sequences and analyze the correlation of their ΔM with estimates of the aftershock productivity and FMD.
Additionally, the finding in the contribution Grimm et al. (2022b), that triggered earthquakes have a substantially larger aftershock productivity than background events and therefore provide a ”boost” to the cluster evolution, should be verified based on alternative declustering methods. Forward simulations could then show the impact on ΔM expectations.
Finally, an extension of the ΔM survival regression model using small-scale covariate data could certainly contribute to shed more light on the question which geophysical conditions favor doublet occurrences.

Part II.

Approach 1: Epidemic Type Aftershock Sequence (ETAS) Model

5. Improving Earthquake Doublet Frequency Predictions by Modified Spatial Trigger Kernels in the Epidemic-Type Aftershock Sequence (ETAS) Model

Contributing article

Grimm, C., Käser, M., Hainzl, S., Pagani, M., Küchenhoff, H. (2021). Improving Earthquake Doublet Frequency Predictions by Modified Spatial Trigger Kernels in the Epidemic-Type Aftershock Sequence (ETAS) Model. *Bulletin of the Seismological Society of America*

Code repository

`https://github.com/ChrGrimm/ETASanisotropic`

Copyright information

Author contributions

Christian Grimm (CG) brought up the initial idea of using the ETAS model and locally restricting the spatial kernel. He designed the novel spatial kernel and derived the algorithm to numerically approximate its spatial integral. Additionally, CG programmed the estimation code, designed and implemented the simulation experiment, conducted the study and first interpreted the results. Furthermore, CG wrote the initial draft of the paper and prepared figures. CG, Martin Käser (MK) and Marco Pagani (MP) jointly contributed to the idea of using anisotropic spatial distributions, Helmut Küchenhoff contributed to the initial methodological discussions. Sebastian Hainzl (SH) conducted comparative ETAS calculations and extensively consulted on the detailed choice of model settings and study design. All authors thoroughly proofread the paper. In particular, SH largely improved the design of the figures. Multiple discussions between CH, SH and MK led to substantial improvements of the interpretations.

Improving Earthquake Doublet Frequency Predictions by Modified Spatial Trigger Kernels in the Epidemic-Type Aftershock Sequence (ETAS) Model

Christian Grimm*[1], Martin Käser[2,3], Sebastian Hainzl[4], Marco Pagani[5], and Helmut Küchenhoff[1]

ABSTRACT

Earthquake sequences add a substantial hazard beyond the solely declustered perspective of common probabilistic seismic hazard analysis. A particularly strong driver for both social and economic losses are so-called earthquake doublets (more generally multiplets), that is, sequences of two (or more) comparatively large events in spatial and temporal proximity. Without differentiating between foreshocks and aftershocks, we hypothesize three main influencing factors of doublet occurrence: (1) the number of direct and secondary aftershocks triggered by an earthquake; (2) the occurrence of independent clusters and seismic background events in the same time–space window; and (3) the magnitude size distribution of triggered events (in contrast to independent events). We tested synthetic catalogs simulated by a standard epidemic-type aftershock sequence (ETAS) model for both Japan and southern California. Our findings show that the common ETAS approach significantly underestimates doublet frequencies compared with observations in historical catalogs. In combination with that the simulated catalogs show a smoother spatiotemporal clustering compared with the observed counterparts. Focusing on the impact on direct aftershock productivity and total cluster sizes, we propose two modifications of the ETAS spatial kernel to improve doublet rate predictions: (a) a restriction of the spatial function to a maximum distance of 2.5 estimated rupture lengths and (b) an anisotropic function with contour lines constructed by a box with two semicircular ends around the estimated rupture segment. These modifications shift the triggering potential from weaker to stronger events and consequently improve doublet rate predictions for larger events, despite still underestimating historic doublet occurrence rates. Besides, the results for the restricted spatial functions fulfill better the empirical Båth's law for the maximum aftershock magnitude. The tested clustering properties of strong events are not sufficiently incorporated in typically used global catalog scale measures, such as log-likelihood values, which would favor the conventional, unrestricted models.

KEY POINTS

- Relative to its importance for hazard, strong event clustering is underrepresented in ETAS model fit measures.
- The ETAS model underrates doublet occurrences by assigning too low trigger productivity to strong events.
- Doublet occurrence rate predictions should be considered as a quality measure for seismic hazard models.

INTRODUCTION

Sequences of strong earthquakes within a relatively narrow time–space window can cause dramatic social and economic damage to our society. The financial losses produced by such multiplets are of particular interest to the risk assessment of governments and in the insurance industry. Recent examples of short-term clusters containing several strong, damaging earthquakes are

1. Department of Statistics, Ludwig-Maximilians-University Munich, Munich, Germany, https://orcid.org/0000-0002-2190-2981 (CG); 2. Department of Earth and Environmental Sciences, Geophysics, Ludwig-Maximilians-University Munich, Munich, Germany; 3. Munich Re, Section GeoRisks, Munich, Germany; 4. Section Physics of Earthquakes and Volcanoes, GFZ German Research Centre for Geoscience, Potsdam, Germany, https://orcid.org/0000-0002-2875-0933 (SH); 5. Global Earthquake Model Foundation, Pavia, Italy, https://orcid.org/0000-0001-8125-1925 (MP)

*Corresponding author: Christian.Grimm@stat.uni-muenchen.de

Cite this article as Grimm, C., M. Käser, S. Hainzl, M. Pagani, and H. Küchenhoff (2021). Improving Earthquake Doublet Frequency Predictions by Modified Spatial Trigger Kernels in the Epidemic-Type Aftershock Sequence (ETAS) Model, *Bull. Seismol. Soc. Am.* **XX**, 1–20, doi: 10.1785/0120210097

the Kumamoto (Japan, 2016) sequence with a magnitude M_{JMA} 7.3 mainshock (i.e., MA refers to Japanese Meteorological Agency.) preceded by M_{JMA} 6.4 and 6.5 foreshocks within 28 hr (Zhuang *et al.*, 2017), and the Ridgecrest (California, 2019) sequence with a mainshock M_w 7.1 preceded by an M_w 6.4 event about 34 hr earlier (Hauksson *et al.*, 2020).

Most typically, sequences of strong and destructive foreshocks, mainshocks, and aftershocks occur within several hours or few days and can therefore be assumed to be controlled by a physical triggering mechanism. However, it is well known that aftershock sequences can increase seismicity locally for years or even decades. In case that two strong events occur in spatial proximity but months apart, the second event may be an offspring of the ongoing sequence of the first or may have happened coincidentally due to independent background seismicity or as a part of an unrelated sequence.

However, from a risk management perspective, the question of physical causality and the particular interevent time seems rather irrelevant. In both cases, the repeated destruction may affect the same governmental budgets and (re)insurance contracts within a relatively short time, and thus presents a comparably severe risk. Reliable predictions of the likelihood of any strong event cluster, both triggered and coincidental, are therefore an important task for risk managers in governments and the insurance industry.

A suitable term for strong event clusters is given by so-called earthquake *doublets*, sometimes more generally referred to as *multiplets*. Although exact specifications are highly inconsistent in the literature, they are generally defined as pairs (doublets) or sets (multiplets) of similarly strong earthquakes in spatiotemporal proximity (Lay and Kanamori, 1980; Kagan and Jackson, 1999; Felzer *et al.*, 2004; Gibowicz and Lasocki, 2005).

Kagan and Jackson (1999) defined doublets as pairs of earthquakes with magnitude $M_w \geq 7.5$ that are no more than one rupture size apart and for which interevent time is less than their recurrence time derived from plate motion. They found that approximately 22% of worldwide events with $M_w \geq$ 7.5 occur in doublets, with a maximum interevent time of doublet pairs of almost 17 yr.

In contrast, Felzer *et al.* (2004) specified multiplets as a potential mainshock together with all aftershocks within 0.4 magnitude units, occurring during the following two days and within a spatial box centered in the mainshock's epicenter. The distance of the mainshock's epicenter to the sides of the box is set to 2.5 times the estimated fault length, which is justified by the hypothesis that aftershocks are generally expected to occur within two fault lengths, with an extra half a length accounting for location uncertainty. They demonstrated statistical evidence that foreshocks, aftershocks, and multiplets occur due to the same physical triggering mechanism, and that the number of times that multiplets occur increases linearly with the number of aftershocks observed. Felzer *et al.* (2004) infer that certain regions in the world, such as Solomon Islands, show an increased multiplet rate due to higher aftershock rates and earthquake density, rather than unique seismic fault structures that support the occurrence of multiplets.

Gibowicz and Lasocki (2005) defined doublets as a pair of trigger-related earthquakes with no more than 0.25 magnitude units difference, applying magnitude-dependent stepwise spatial and temporal constraints of 40–90 km and 200–450 days.

Although the concept of earthquake triggering is well known and the potential of additional damage due to ongoing seismic sequences has been shown in recent studies (Abdelnaby, 2012; Kagermanov and Gee, 2019; Papadopoulos *et al.*, 2020), seismic hazard is typically computed considering only independent (i.e., mainshock) earthquakes, for example, in probabilistic seismic hazard analysis (PSHA) approaches (Cornell, 1968; McGuire, 2008). PSHA traditionally not only neglects contributions to hazard from supposedly triggered sequences and therefore underestimates chances of doublet and multiplet occurrences, but it is also based on the highly subjective and influential selection of a declustering method (van Stiphout *et al.*, 2011; Marzocchi *et al.*, 2014; Zhang *et al.*, 2018).

A prominent and extensively studied method to analyze earthquake sequences is the epidemic-type aftershock sequence (ETAS) model (Ogata, 1988, 1998). ETAS accounts for earthquake clustering in terms of a branching process, and models the number of aftershocks as well as their spatial and temporal distribution depending on the magnitude of the trigger. The spatiotemporal event rate is formed by the sum of a triggered rate and a time-independent seismic background rate contribution (Zhuang *et al.*, 2002; Kagan *et al.*, 2010; Chu *et al.*, 2011; Jalilian, 2019). ETAS model estimations can be used for both short-term aftershock forecasts and the simulation of long-term synthetic catalogs.

The goodness of ETAS model fits is typically assessed by the log-likelihood function (LLF), Akaike's information criterion (AIC), or the degree of spatial clustering, expressed by Ripley's K-function (Veen, 2006; Chu *et al.*, 2011). Besides, visual tools such as spatial plots of the estimated conditional intensity and a comparison of the ETAS triggering function with observed aftershock rates in the historic catalog can be used (Chu *et al.*, 2011). All previously mentioned have in common that they assess the model fit on a global catalog scale, that is, they test whether the synthetic catalogs sufficiently well represent the observed spatiotemporal clustering behavior in the full magnitude range.

The log-likelihood and AIC measures are related to the joint probability of all earthquakes, and thus mainly determined by the fit to the more numerous smaller magnitude events. This might be problematic concerning earthquake risk, which is mainly related to large events. For example, Hainzl *et al.* (2008, 2013) showed that the common ETAS assumption of isotropic aftershock triggering leads to a biased magnitude scaling of the aftershock productivity in which the trigger potential of small magnitudes is overestimated to better adapt to realistic anisotropic aftershock distributions. Therefore, it is also desirable to assess synthetic ETAS catalogs on their capability to predict realistic occurrence rates of large-magnitude doublets and multiplets.

In this article, we present a new concept of assessing the quality of synthetic catalogs generated by ETAS with respect to doublet and multiplet rates. We introduce three novel and more realistic designs of the ETAS spatial kernel that improve predictions of the respective rates: (1) an anisotropic spatial distribution, (2) an isotropic but finite spatial distribution, and (3) a finite anisotropic spatial distribution. We then test our new model approaches for 24 and 39 yr lasting earthquake catalogs recorded in Japan and southern California, respectively.

In the following section, we derive a doublet and multiplet definition that is used in this article, and comprehensively discuss the main influencing factors for doublet and multiplet occurrences. Next, we briefly describe the utilized earthquake catalogs. We then describe the common ETAS model, define the tested variants of the spatial kernel, and introduce the quality measures applied in our analysis of model fits and simulation results. Finally, we present and discuss the results of all four studied ETAS model versions, and we interpret the findings related to the initial motivation in the Conclusion section.

EARTHQUAKE DOUBLETS

Definition

For the sake of simplicity, in this work, we waive the term multiplet and define an earthquake doublet more generally as a pair or set of events with a magnitude difference of less than 0.4, occurring within 1 yr (starting from the occurrence time of the earlier event) and within a circular radius of 2.5 times the estimated rupture length of the earlier event.

The temporal constraint of 1 yr is derived from the typical length of a (risk) budget period or reinsurance contract. We limit our investigation to strong events with magnitude $M_w \geq 5.9$, therefore allowing for doublet and multiplet partner events down to $M_w \geq 5.5$.

Doublets may either occur within a supposed triggered sequence (mainshock and aftershock) or among independent clusters. To avoid doublets built by two aftershocks, being both related to a stronger mainshock prior to them, we only count doublets in which the earlier event is not contained in the time–space domain of a previous, stronger event. This is consistent with our motivation drawn from a risk management perspective, because the damage caused by an aftershock–aftershock doublet is likely to be overshadowed by the mainshock.

Main influencing factors

Assuming equal physical triggering mechanisms of foreshocks and aftershocks (Felzer *et al.*, 2004), we propose the following three main factors for doublet occurrences:

1. the *aftershock productivity*, that is, the number of direct and secondary offsprings triggered by an earthquake,
2. the *number of independent events* in the same time–space window, that is, the occurrence of clustered and background events that are unrelated to the triggering of the event under consideration, and
3. the *magnitude size distribution* of triggered and independent background events.

It is evident that a higher amount of earthquakes within the time–space window of an investigated event increases the probability of a doublet occurrence. Therefore, an increased aftershock productivity and background activity (the first two factors mentioned before), increase the likelihood that a doublet partner is found. Clearly, the aftershock productivity has a much stronger effect than the time-homogeneous seismic background rate, because it directly increases the local and short-term cluster size. It is important to mention that triggered and background seismicity are interacting like competing contributors to event rates in ETAS, so an increase of aftershock productivity is generally going along with a decrease of background seismicity and vice versa. The overlapping of the considered event sequence by a second, unrelated cluster evolving in the same time–space window increases the doublet probability substantially if the second cluster is approximately equal or larger.

Regarding the magnitude size distribution (the third factor mentioned before), it is still controversially debated in the literature whether the magnitude size distribution of a triggered event depends on the magnitude of its trigger. Although Felzer *et al.* (2004) assume that the magnitude size distribution of a triggered event follows a constant Gutenberg–Richter relationship and therefore is independent of the trigger's magnitude, Nandan *et al.* (2019) find triggered magnitudes clustering around the triggering magnitude by a kinked magnitude size distribution, which would mean that the triggering event tends to reproduce similar magnitudes with increased probability. In contrast to that, based on a stacking-approach analysis of $M_w \geq 5.9$, therefore allowing for doublet and multiplet partner events down to $M_w \geq 5.5$, Gulia *et al.* (2018) argue that the b-value on average shows a temporal 20%–30% increase compared with the premainshock time, with more significant increases for stronger events nearby the mainshock epicenter location. We point out that the kinked magnitude size distribution by Nandan *et al.* (2019) would increase chances of doublet and multiplet occurrence, whereas the temporal b-value increase suggested by Gulia *et al.* (2018) significantly lowers their likelihood.

In this article, however, we assume a unique magnitude size distribution for all events according to the Gutenberg–Richter relationship (Gutenberg and Richter, 1944) as done so in the vast majority of ETAS studies. Instead, we are focusing our study on the impact of the aftershock productivity in ETAS on doublet and multiplet occurrence rates.

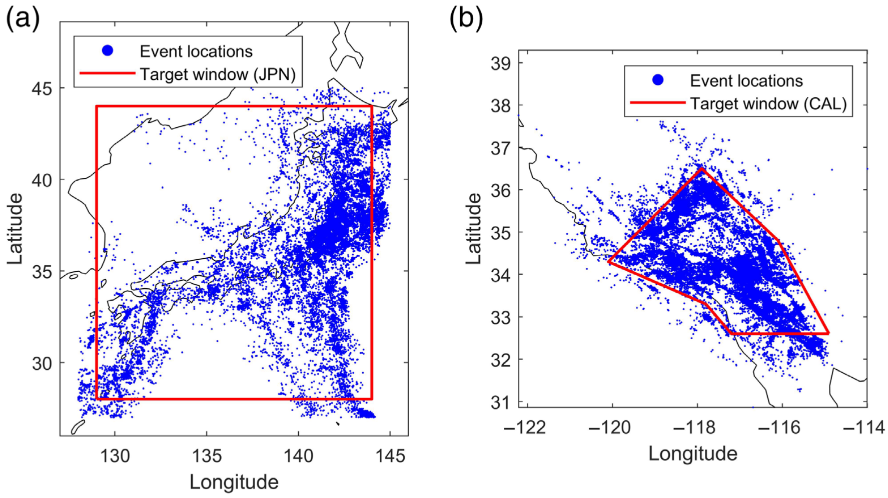


Figure 1. Event locations in the two utilized earthquake catalogs, including both target and complementary events. Red polygons represent the respective spatial target window. (a) Events in National Research Institute for Earth Science and Disaster Resilience (NIED) catalog for Japan, $M_w \geq 4.0$, target period from 1 July 1997 to 31 October 2020, complementary period from 1 January 1997 to 30 June 1997. (b) Events in Southern California Earthquake Data Center (SCEDC) catalog for southern California, $M_w \geq 2.8$, target period from 1 July 1981 until 31 December 2019, complementary period from 1 January 1981 to 30 June 1981. The color version of this figure is available only in the electronic edition.

SELECTION OF EARTHQUAKE CATALOGS

We perform our study in two regions with distinct tectonic environments and faulting types—Japan and southern California. The seismicity in Japan is complex, hosting reverse-faulting subduction zone events (particularly along the coast) with relatively flat dips and broader, more isotropically shaped spatial distributions of aftershock epicenters, as well as in-slab normal-faulting earthquakes and crustal events with varying depths and mechanisms. Southern California has mostly steep faults with strike-slip rupturing mechanisms in a continental tectonic regime, promoting narrower, elongate distributions of epicenters.

Regional catalogs

In the following, we describe the regional earthquake catalogs used for the estimations of the ETAS model. For each dataset, we define a time–space *target* window, which is constructed by a time span and a geographical polygon. This window comprises the so-called *target events* that are used to fit the model. The additional *complementary* window is built by the preceding six months and a 1° bounding box around the polygon in the geographic coordinate system. The so-called *complementary events* are not fitted by the ETAS model estimations but may contribute to the estimated trigger rate of events in the target domain.

We downloaded the Japan earthquake catalog from the National Research Institute for Earth Science and Disaster Resilience (NIED; see Data and Resources; Kubo *et al.*, 2002). The catalog provides both moment tensor magnitudes and JMA scale magnitudes. For our study, we chose the moment magnitude data, which are complete from $M_c = 4.0$ according to the fit of the model of Ogata and Katsura (1993). We define the time–space target window from 1 July 1997 to 31 October 2020, and for a longitude–latitude range from 129° to 144° E and from 28° to 44° N, respectively. Figure 1a shows the selected event locations with the corresponding boundaries of the spatial polygon.

The focal mechanism catalog for southern California was obtained from the Southern California Earthquake Data Center (See Data and Resources; Hauksson *et al.*, 2012; Yang *et al.*, 2012). Magnitudes are provided in moment magnitude scale. The completeness magnitude is estimated to be $M_c = 2.8$ using the Ogata and Katsura (1993) model. We defined the target window from 1 July 1981 to 31 December 2019 and by a hexagonal polygon (Hutton *et al.*, 2010), which is depicted in Figure 1b together with all event locations.

Both catalogs provide nodal-plane solutions for each event. Because the accuracy of focal mechanisms cannot be guaranteed, especially for smaller magnitude events, we used the given sets only as additional candidates in our algorithm to determine the strike angle needed for the anisotropic ETAS model version (See the ETAS Model section.).

Short-term incompleteness

Short-term incompleteness in earthquake catalogs can be defined as the deficiency of events above the general completeness level M_c for a limited time after a relatively large event. The phenomenon appears to mainly result from the overlap of seismic records that are dominated by the coda waves of the preceding strong event, and therefore let subsequent, weaker events remain undetected (de Arcangelis *et al.*, 2018).

Short-term incompleteness in the underlying earthquake catalogs has been identified as a major source of bias in the ETAS estimation process (Kagan, 2004; Hainzl, 2016a,b; Page *et al.*, 2016; Seif *et al.*, 2017). For $M \geq 6$ earthquakes in southern California, Helmstetter *et al.* (2006) estimated the duration of temporary catalog incompleteness (in days) above a given magnitude threshold M_c as

$$t = 10^{(m-4.5-M_c)/0.75}. \quad (1)$$

For instance, it means that a catalog with cutoff magnitude M_c is incomplete for about one day after an event with magnitude $m = M_c + 4.5$. The duration of incompleteness exceeds one minute for magnitudes $m \geq M_c + 2.2$.

In our study, we assume that relation (1) is approximately valid for the region of Japan as well. Events occurred during periods of temporary incompleteness are not used for the maximum-likelihood ETAS fit, but still contribute to the ETAS event rates of future target events, which means, technically speaking, that they are downgraded from target to complementary events. To avoid excessive fragmentation of the target time window, we applied short-term incompleteness only to events with magnitudes $M_c \geq 6.2$ for Japan and $M_c \geq 5.0$ for southern California, which is 2.2 magnitude units above the respective catalog thresholds.

Global ISC-GEM catalog

For the comparison with more long-term regional and global doublet occurrence rates, we utilize the International Seismological Centre–Global Earthquake Model (ISC-GEM) Global Instrumental Earthquake Catalogue with events from 1 January 1904 (See Data and Resources; Storchak *et al.*, 2015; Di Giacomo *et al.*, 2018). Magnitudes are provided in moment magnitude scale. According to the catalog description and Di Giacomo *et al.* (2018), the ISC-GEM catalog is stepwise complete from $M_c = 7.5$ (before 1918), $M_c = 6.25$ (from 1918 to 1959), and $M_c = 5.5$ (since 1960). Significant continental earthquakes with magnitude 6.5 or larger are included before 1918.

ETAS MODEL

The initial ETAS model implemented in this study is based on the *R* package ETAS, as presented by Jalilian (2019) (See Data and Resources.). It estimates the model parameters using a maximum-likelihood approach and the stochastic declustering method introduced by Zhuang *et al.* (2002).

In ETAS, the occurrence rate of an earthquake at a given time t and location (x, y) corresponds to the sum of two overlaying components: (a) the coincidental, time-independent background seismicity rate and (b) the sum of dynamic trigger rate contributions from all events occurred before time t (i.e., the event history H_t). The combined occurrence rate is, therefore, modeled by a nonhomogeneous Poisson process with intensity function:

$$\lambda(t, x, y|H_t) = \mu h(x, y) + \sum_{i:t_i<t} \kappa_{A,\alpha}(m_i) g_{c,p}(t - t_i) f_{D,\gamma,q}(x, y, i), \quad (2)$$

in which μ is the total rate of $m \geq M_c$ background events in the whole region, and $h(x, y)$ denotes the spatial probability density function (pdf) of the background seismicity.

The term within the sum describes the trigger rate contribution of an event i, occurred at time $t_i < t$ and location (x_i, y_i) with magnitude m_i, to the rate of $m \geq M_c$ events at time t and location (x, y).

The *aftershock productivity* function

$$\kappa_{A,\alpha}(m_i) = A \exp(\alpha(m_i - M_c)) \qquad (m_i \geq M_c; A, \alpha > 0), \quad (3)$$

describes the average number of direct aftershocks (offsprings) triggered by an event i with magnitude m_i. Such an exponential growth of the productivity is in good agreement with observations (see, e.g., the summary provided by Hainzl and Marsan, 2008).

The *temporal trigger function*

$$g_{c,p}(t - t_i) = (t - t_i + c)^{-p} \qquad (t \geq t_i; c, p > 0), \quad (4)$$

is the well-known empirical Omori–Utsu law for the decay of aftershock rates with increasing time t after the occurrence time t_i of the triggering event i (Utsu *et al.*, 1995). The c-value defines the delay of the onset of the power-law decay and is typically much less than one day. It is likely related to short-time incompleteness of earthquake catalogs after mainshocks (Hainzl, 2016a). The p-value is in the 0.8–1.2 range in the most cases (Utsu *et al.*, 1995).

Finally, the spatial trigger function $f_{D,\gamma,q}(x, y, i)$ is conventionally designed as an isotropic pdf and models the decay of aftershock rates depending on the distance of (x, y) to the epicenter of the triggering event (x_i, y_i). The ETAS model with an isotropic spatial kernel is the hereinafter called isotropic reference model M_0.

However, the assumption of an isotropic distribution is considered to be a weak point in many publications throughout the literature (Ogata, 1998, 2011; Ogata and Zhuang, 2006; Hainzl *et al.*, 2008, 2013; Bach and Hainzl, 2012; Seif *et al.*, 2017; Zakharova *et al.*, 2017; Zhang *et al.*, 2018, 2020). To name a few, Zhang *et al.* (2018) emphasize that isotropy may be specifically unsuitable for subduction zone events above a magnitude of approximately M_w 7.5, because estimated rupture lengths and widths are diverging increasingly. They suggest a uniform spatial density in the rupture area with power-law decay outside. Moreover, because ETAS usually neglects the depth dimension, increasing dip angles can already lead to a clearly elongate and thus anisotropic projection shapes of the rupture plane for even smaller events. Another prominent design is the elliptic Gaussian distribution introduced by Ogata (1998) and further studied by Ogata and Zhuang (2006) and Ogata (2011).

In the previous references, anisotropic models are generally found to lead to more accurate ETAS model estimates. In particular, Hainzl *et al.* (2008, 2013) emphasize that the assumption of isotropy can lead to an underestimation of the aftershock productivity parameter α, resulting in underpredicted cluster sizes of stronger events. Given our particular interest in strong events, this gives the motivation to apply an anisotropic alternative in this study.

Besides, in preliminary analyses of a standard ETAS model, we observed that small events are typically assigned a much wider reach of spatial triggering relative to their estimated rupture size than large events. We hypothesize that this might similarly promote disproportionate triggering of smaller events, because it might be easier for the ETAS algorithm to model unique spatial cluster patterns by the overlapping spatial kernels of a large number of smaller events than by the rather inflexible spatial shapes of fewer, but stronger events.

Therefore, in this article, we propose two modifications of the conventional, isotropic reference model M_0: First, we apply an anisotropic spatial kernel constructed around the surface projection of the estimated rupture segment, which is assumed to be parallel to the strike and passing through the epicenter. Second, we introduce a magnitude-dependent spatial restriction threshold to the spatial kernel that prevents events from triggering outside of the specified surrounding area.

In the following, we introduce the finite and infinite isotropic and anisotropic kernels. Next, we present the algorithm to estimate the rupture length as well as the strike angle and epicenter position along the rupture line in the anisotropic model case. Then, we define the set of four models that were tested in this study. Ultimately, we account for a rescaling of the aftershock productivity.

Isotropic versus anisotropic spatial kernel

Consider a triggering event i with magnitude m_i and epicenter location (x, y). Furthermore, in the isotropic case, let $r_i(x, y)$ be the point-to-point distance of a point (x, y) to the epicenter location of event i. We define the standard isotropic spatial kernel following Jalilian (2019) by

$$f_{D,\gamma,q}(x,y,i) := \frac{q-1}{D\exp(\gamma(m_i-M_c))}\left(1+\frac{\pi r_i(x,y)^2}{D\exp(\gamma(m_i-M_c))}\right)^{-q}, \tag{5}$$

with spatial parameters $q > 1$ and $D, \gamma > 0$. The characteristic length of the power-law decay, $\sqrt{D\exp(\gamma(m_i - M_c))/\pi}$, scales with the trigger magnitude, which accounts for the observed exponential increase of the rupture dimensions with earthquake magnitude (Wells and Coppersmith, 1994).

For the anisotropic case, let l_i be the estimated rupture length of event i and $r_i(x, y)$ denote the nearest point-to-segment distance of a point (x, y) to the estimated rupture segment of event i. Then, we construct the anisotropic spatial kernel by

$$f_{D,\gamma,q}(x,y,i) := \frac{q-1}{D\exp(\gamma(m_i-M_c))}\left(1+\frac{2l(m_i)r_i(x,y)+\pi r_i(x,y)^2}{D\exp(\gamma(m_i-M_c))}\right)^{-q}, \tag{6}$$

with the same parameter constraints $q > 1$ and $D, \gamma > 0$. The anisotropic kernel (equation 6) is a generalisation of the isotropic kernel (equation 5) for rupture lengths $l(m_i) > 0$. In contrast to the isotropic function, the contour lines of the anisotropic kernel are not centered around the epicenter but constructed as a box with two semicircular ends around the estimated rupture line of the triggering event. Both kernels are pdfs over infinite space.

Spatial restriction

We can restrict the spatial extent of both the isotropic and anisotropic spatial kernel by setting $f_{D,\gamma,q}$ equal to 0, if the respective distance term exceeds a certain magnitude-dependent threshold $\tilde{r}(m_i)$, that is,

$$\tilde{f}_{D,\gamma,q}(x,y,i) = \begin{cases} \frac{f_{D,\gamma,q}(x,y,i)}{F_{D,\gamma,q}(m_i)} & \text{if } r_i(x,y) \le \tilde{r}(m_i) \\ 0 & \text{otherwise} \end{cases}, \tag{7}$$

in which $\tilde{f}_{D,\gamma,q}(x,y,i)$ is normalized by the integral of $f_{D,\gamma,q}(x,y,i)$ over the area up to the cutoff distance $\tilde{r}(m_i)$ to retain a pdf:

$$F_{D,\gamma,q}(m_i) = \begin{cases} 1-\left(1+\frac{\pi\tilde{r}(m_i)^2}{D\exp(\gamma(m_i-M_c))}\right)^{1-q} & \text{(isotropic model)} \\ 1-\left(1+\frac{2l(m_i)\tilde{r}(m_i)+\pi\tilde{r}(m_i)^2}{D\exp(\gamma(m_i-M_c))}\right)^{1-q} & \text{(anisotropic model)} \end{cases}.$$

In this study, we use a threshold that is proportional to the magnitude-dependent rupture length $l(m_i)$ of event i, that is, $\tilde{r}(m_i) = kl(m_i)$, to correlate the spatial trigger extent to the estimated rupture dimension. Figure 2 visualizes the shapes of isotropic and anisotropic spatial kernels, restricted to a distance of $\tilde{r}(m_i) = 2.5l(m_i)$, for the exemplary magnitudes M_c 5.0 and 7.5, using initial spatial parameter guesses $D = 2.0$, $\gamma = 2.1$, and $q = 1.5$.

Estimation of rupture length, strike, and position of rupture line

The anisotropic spatial kernel defined in equation (6) requires an estimation of the ruptured segment, in particular, its central location, the length, and the strike angle to locate the rupture line segment of an earthquake. To obtain magnitude-dependent estimates of the subsurface rupture lengths l of all events, we use the scaling relations:

$$\log_{10}(l(m)) = \begin{cases} -2.37 + 0.57m & \text{reverse faulting} \\ -2.57 + 0.62m & \text{strike-slip faulting} \end{cases}, \tag{8}$$

in which, for the sake of simplicity, we selected the reverse-faulting scaling relations for subduction environments, provided by Blaser *et al.* (2010), for all events in the Japan catalog and the strike-slip faulting equations for continental regimes, given by Wells and Coppersmith (1994), for all southern Californian events.

The strike angles are selected such that the corresponding rupture line fits well to the cloud of potential aftershocks.

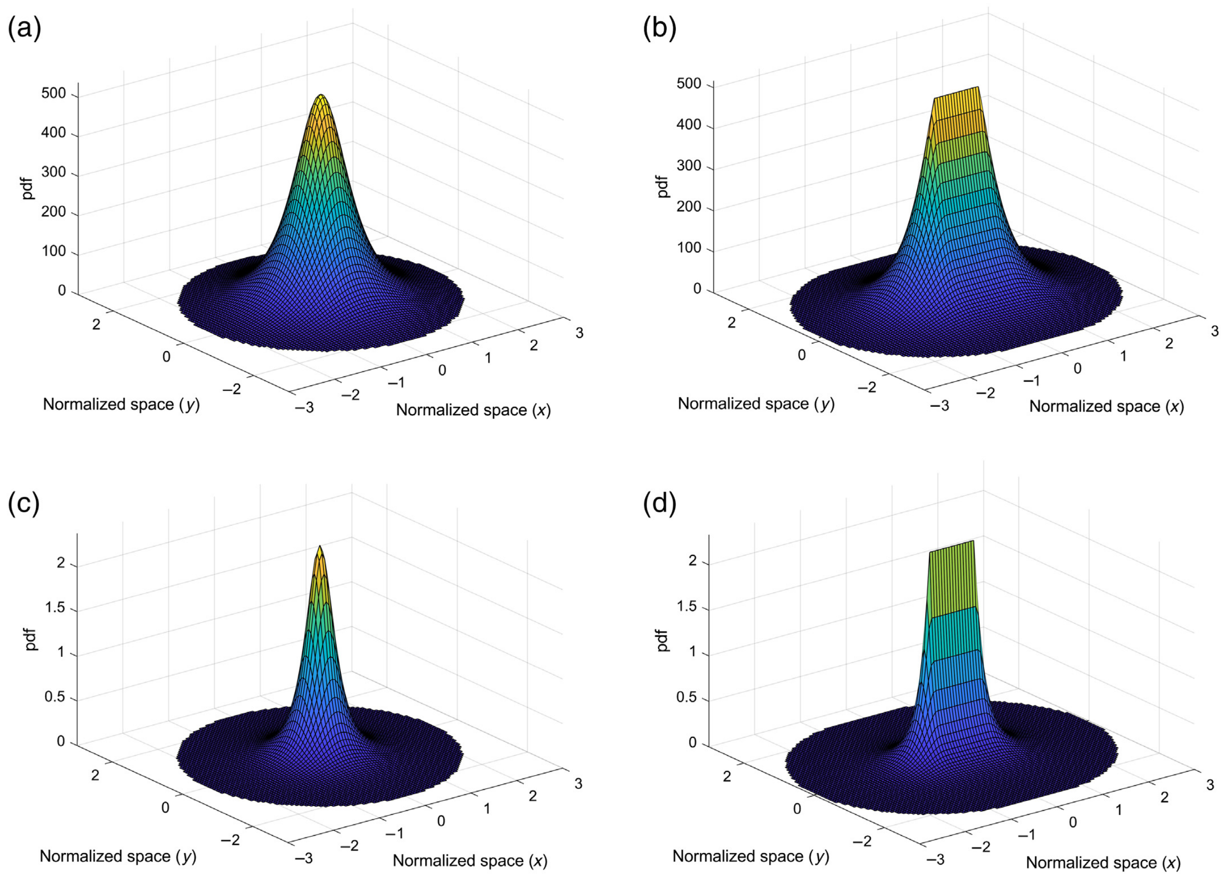


Figure 2. Visualization of the spatial kernels restricted to a distance of $\tilde{r}(m_i) = 2.5l(m_i)$: (a) isotropic kernel for magnitude M_c 5.0, (b) anisotropic kernel for magnitude M_c 5.0, (c) isotropic kernel for magnitude M_c 7.5, and (d) anisotropic kernel for magnitude M_c 7.5. The 3D probability density functions (pdfs) result from equation (7), using the initial parameter guesses $D = 2.0$, $\gamma = 2.1$, and $q = 1.5$. The color version of this figure is available only in the electronic edition.

Therefore, we test the given focal mechanism data in the earthquake catalogs and compute the summed trigger rates for the subsequent events like in equation (2), assuming initial parameter guesses:

$$(A_0, \alpha_0, c_0, p_0, D_0, \gamma_0, q_0) = (0.02, 1.6, 0.02, 1.0, 2.0, 2.1, 1.5).$$

In addition, we go through all strikes from 0° to 175° in 5° steps and compute the initial trigger rates accordingly. From all candidates, we choose the one that leads to the maximum sum of occurrence rate contributions to future events and, therefore, is in best agreement with presumed offsprings. Because we do not consider the rupture's dip, strikes above 180° coincide with the tested set of angles. Once we have optimized the strikes via the earlier approach, we additionally test five different positions of the rupture line relative to the corresponding epicenter location of the trigger event. Thus, we allow the epicenter to lie either right at the start, center, or end of the rupture line, or a quarter or three quarters along the rupture line. The aforementioned selection algorithm clearly represents a manipulation of the initial model conditions. In fact, the so-selected strike angles show only moderate agreement with the originally provided strikes. Compared to using only nodal-plane solutions given in the catalogs, we observed negligible effects for Japan and moderately increasing estimates of the aftershock productivity in southern California, where potential aftershocks were more likely to scatter along a clearly identifiable line. In any case, the impact of optimized strike selection was much smaller than the effect of the introduced spatial restrictions or the anisotropic shape of the spatial kernel itself.

Mismodeling of the spatial aftershock distribution leads to biased model estimates (Hainzl *et al.*, 2008). To minimize this problem, we refrained from directly using the strike values provided in the catalogs due to the large uncertainties in the source inversions. Instead, the optimized selection of strike angles assures that the event's rupture line passes through the cloud of its potential aftershocks, which we visually confirmed for individual sequences.

TABLE 1
Overview of the Specifications of the Four Epidemic-Type Aftershock Sequence (ETAS) Model Variants Tested in This Article

Model	Spatial Design	Restriction Factor	Strike Estimation	Epicenter Location
M_0	Isotropic	100	—	—
M_1	Anisotropic	100	Optimized	Optimized
M_2	Isotropic	2.5	—	—
M_3	Anisotropic	2.5	Optimized	Optimized

Choice of four model designs

In this article, we analyze four different variants of the ETAS model regarding their ability to predict realistic doublet and multiplet rates. Table 1 lists the model design specifications made for each approach.

The reference model M_0 represents the standard isotropic design in equation (5) with event-specific spatial restriction:

$$\tilde{r}_0(m_i) = 100 l(m_i).$$

The restriction $\tilde{r}_0(m_i)$ is only of technical nature and has negligible impact on results, while considerably improving code performance by avoiding the computation of extremely distant interevent triggering relations over the entire catalog size. Hereinafter, we will therefore refer to models with spatial extent $\tilde{r}_0(m_i)$ as unrestricted.

Using the same isotropic kernel (equation 5), in model M_1, we test the spatial restriction:

$$\tilde{r}_1(m_i) = \min\{2.5 l(m_i), 1\},$$

in which the lower limit of one kilometer guarantees a minimum spatial extent to the smallest events in the catalog. The goal of the restricted extent of the spatial kernel is to avoid wrong associations of distant events as aftershocks. It gives more triggering power to the stronger events (which may trigger in a larger area) and takes away triggering potential from the weaker events. The metric of 2.5 rupture lengths goes back to the assumption in Felzer *et al.* (2004) that aftershocks are expected to mainly occur within this distance, including a buffer of half a rupture length for location uncertainties.

Model M_2 builds upon the anisotropic spatial kernel (equation 6) with optimized strikes and relative rupture locations and is unrestricted ($\tilde{r}_0(m_i)$).

Finally, model M_3 tests the anisotropic spatial kernel with restriction $\tilde{r}_1(m_i)$.

For the sake of consistency, we applied the anisotropic spatial kernels to all events disregarding their magnitude in models M_1 and M_3. For small rupture lengths, however, the shape is similar to an isotropic kernel.

Subsequent rescaling of ETAS functions

The temporal trigger function (equation 4) is not a pdf, because its integral over infinite time typically amounts to a number larger than 1 (for $p > 1$) or infinity (for $p < 1$). Therefore, the excessive density in equation (4) downscales the estimates of parameter A in the productivity function (equation 3).

In favor of better interpretability of the model results, it is useful to cut off the temporal trigger function (equation 4) at the length of the entire catalog T (in days) and normalize it by the integral over the time range from 0 to T, that is,

$$G_{c,p}(T) = \frac{1}{1-p}((T+c)^{1-p} - c^{1-p}).$$

Accordingly, we rescale the absolute aftershock productivity parameter A by

$$\tilde{A} = A G_{c,p}(T). \tag{9}$$

QUALITY MEASURES

In this section, we introduce the quality measures used to assess and compare the goodness of the selected models. We start with a short description of the log-likelihood value and branching ratio, designed to assess the goodness of fit and the detected trigger portion on a global catalog scale. These properties are widely used in ETAS analysis but have the disadvantage that they do not provide any detailed information on how well the model represents the critical triggering behavior of particularly strong earthquakes, which is of interest in this study.

Therefore, we add tools to more specifically evaluate the models' capability of representing strong event clusters. First, the expected, magnitude-dependent cluster size is derived. Next, we outline the ETAS forward simulation procedure for both single sequences and synthetic catalogs based on the model estimates. Then, we suggest visual and semiquantitative measures (e.g., Båth's law, degree of temporal and spatial clustering) that help understand clustering properties in the simulated catalogs. Finally, we describe the evaluation of doublet probabilities from simulated catalogs and sequences.

LLF and integrated event rate

The set of ETAS parameters, $\theta = (\mu, A, \alpha, c, p, D, \gamma, q)$, is optimized by maximizing the LLF:

$$l(\theta|H_T) = \sum_{j=1}^{N} \ln(\lambda_\theta(t_j, x_j, y_j|H_{t_j})) - \Lambda_\theta(\mathbb{T}, \mathbb{S}|H_{\mathbb{T}}), \tag{10}$$

in which the first term sums up the logarithmic event rates (equation 2) at the exact times t_j and locations (x_j, y_j) of the N target events that occurred in the time–space window specified for each catalog. The second term

$$\Lambda_\theta(\mathbb{T}, \mathbb{S}|H_\mathbb{T}) = \int_\mathbb{T} \iint_\mathbb{S} \lambda_\theta(t, x, y|H_t)dxdydt, \quad (11)$$

represents the total event rate integrated over the (stepwise) target time window $\mathbb{T}$ and the target space window $\mathbb{S}$ based on the estimated background seismicity rate and the triggering-induced rate resulting from contributions of both target and complementary events in the original catalog. In other words, $\Lambda_\theta(\mathbb{T}, \mathbb{S}|H_\mathbb{T})$ represents the expected total number of events to occur within the modeled target time–space domain (Ogata, 1988, 1998; Jalilian, 2019).

In general, a larger LLF value $l(\theta|H_T)$ implies a better fit to the event occurrence in the original catalog. The LLF value is comparable only for identical data inputs, that is, model runs for Japan and southern California cannot be cross compared. Because all model approaches are based on the same number of free parameters, the information from the AIC criterion is redundant and therefore not shown.

Branching ratio

We modeled the magnitude size distribution by the pdf derived from the Gutenberg–Richter relationship, that is,

$$\rho(m) = \begin{cases} \beta \exp(-\beta(m - M_c)) & \text{if } m \geq M_c \\ 0 & \text{otherwise,} \end{cases} \quad (12)$$

thus assuming $M_{\max} = \infty$ as the maximum magnitude for each region. The maximum-likelihood estimator for parameter β is

$$\hat{\beta} = \frac{N}{\sum_{i=1}^{N}(m_i - M_c)},$$

in which N is the number of fitted events, and m_i denotes the respective event magnitudes (Jalilian, 2019). Applied to the magnitudes of all target events in our regional catalogs, we obtained $\hat{\beta}_{\mathrm{JPN}} = 2.36$ for Japan and $\hat{\beta}_{\mathrm{CAL}} = 2.73$ for southern California.

The branching ratio measures the mean direct aftershock productivity of an arbitrary event, averaged over the entire magnitude range. It is computed by the integral of the estimated aftershock productivity with parameters α and rescaled $\tilde{A}$ weighted by the pdf of the magnitude size distribution $\rho(m)$, that is (Seif *et al.*, 2017; Jalilian, 2019)

$$\nu_{\mathrm{branch}} = \int_{M_c}^{\infty} \tilde{A} e^{\alpha(m - M_c)} \rho(m) dm = \frac{\tilde{A}\beta}{\beta - \alpha}, \quad (13)$$

for $\alpha < \beta$.

Cluster size

Based on the estimates of the (direct) aftershock productivity function (equation 3) and the branching ratio (equation 13), we obtain the expected cluster size:

$$\widehat{N}_c(m) = \frac{\tilde{A} e^{\alpha(m - M_c)}}{1 - \nu_{\mathrm{branch}}}, \quad (14)$$

including secondary triggering by use of the geometric series (Helmstetter and Sornette, 2003).

ETAS forward simulation process

For every model and region, we used the fitted ETAS parameters to forward-simulate both single synthetic sequences and entire catalogs. We generated single trigger sequences to study the results without the impact of background seismicity and independent clusters. Each of these sequences is initiated by a mainshock of varying magnitude, starting from M_w 5.5 and incrementally increasing in tenths of a magnitude unit. For each region and model, a set of 5000 sequences was simulated for each mainshock magnitude.

In addition, we simulated 10,000 realizations of an entire synthetic catalog, including background seismicity and simultaneously evolving trigger sequences. As a time–space window for the simulations, we chose the identical constraints for which the ETAS models were fitted (See the Selection of Earthquake Catalogs section.), including the semiyear complementary time window as an initialization period of pre-existing seismicity. The background seismicity rate is distributed over the spatial window by a superposition of bivariate, isotropic Gaussian kernels, centered in the original event occurrences (Jalilian, 2019).

In both types of simulations, the number of offsprings is drawn from a Poisson distribution with an expected value equal to the magnitude-dependent aftershock productivity estimate. We used the inversion method to sample the spatial and temporal distance of an offspring to the trigger, and then sampled uniformly from the respective contour line of the spatial distribution. The magnitudes of both triggered and independent events were sampled from the Gutenberg–Richter distribution (equation 12), with β as estimated for the respective region.

Because the original Japan catalog contains the extreme Tohoku earthquake (11 March 2011; according to the catalog M_w 8.7) that is very unlikely to be sampled from the Gutenberg–Richter distribution, we manually added the Tohoku event to all synthetic catalogs for Japan.

Båth's law

An important property of an earthquake cluster is the magnitude difference between the mainshock and the strongest aftershock, as it can serve as an indicator of how much hazard is added by the ongoing triggering of a sequence.

Historical observations show that this magnitude difference is, on average, approximately 1.2 magnitude units independently of the absolute magnitude of the trigger event, which is referred to as Båth's Law (Vere-Jones, 1969; Helmstetter and Sornette, 2003; Shearer, 2012).

For observed and synthetic catalogs, we approximate the magnitude difference by applying the time–space constraints of our doublet definition to any event under consideration, and computing the magnitude difference between the considered event and the strongest of all events that occurred in the specified time–space domain. Clearly, this selection can include independent background events or events occurred in unrelated clusters. To constrain the Båth law statistics to mainshocks, we skip earthquakes that are supposed aftershocks (i.e., which are contained in the time–space range of a previous, stronger event) or foreshocks (i.e., which contain a stronger event in their own time–space domain).

For synthetic sequences, we apply the same filtering algorithm to each simulated sequence with its known initiating magnitude.

Coefficient of variation

We measure the degree of temporal clustering of event occurrences by decomposing the time domain into a monthly grid and computing the variation of the numbers of events falling into the time intervals. To account for varying overall catalog sizes, we use the coefficient of variation (CV), which is a measure of the relative dispersion of a random distribution sample X standardized by its mean. It is computed as $\mathrm{CV} = \frac{\sqrt{\mathrm{Var}(X)}}{\mathrm{Mean}(X)}$, in which $\mathrm{Var}(X)$ denotes the variance of the sample X.

Ripley's *K*

The degree of spatial clustering of the event locations can be expressed by Ripley's K-function (Ripley, 1976; Veen, 2006). The K-function computes the average number of additional event locations within a distance h of any given event, normalized by the overall number of events per space unit N/A, that is,

$$K(h) = \frac{A}{N^2} \sum_i \sum_{j \neq i} \mathbb{1}(r(i,j) \leq h), \tag{15}$$

in which $\mathbb{1}$ is the indicator function.

If the investigated catalog was produced by a homogeneous Poisson process with no spatial clustering inherent, $K(h)$ would be asymptotically normal with $K(h) \sim N(\pi h^2, \frac{2\pi h^2 A}{N^2})$ (Chu *et al.*, 2011). The more $K(h)$ exceeds πh^2, the more clustered the event locations are. Values of $K(h) < \pi h^2$ signify inhibition.

Doublets probability

The most important measure for our study's purpose is the probability that an event is part of an earthquake doublet, according to our definition. Similar to the Båth law evaluation, we searched all events within the specified time–space window spanned by the earthquake under consideration in the synthetic catalogs. We counted the earthquake as a doublet event if any of the potential partners fulfill the magnitude criterion.

Similarly, for synthetic sequences, we applied the previously mentioned algorithm to the known sequence initiating events.

RESULTS AND DISCUSSION

In the following, we discuss the results obtained from the four tested models. We start by comparing the ETAS estimation results on a global catalog and model scale by looking at the log-likelihood values, the branching ratios, the general shapes of the fitted spatial kernels, and the average cluster sizes depending on the trigger magnitude. Then, we move on to the analysis of the synthetic results from simulated sequences and catalogs. Herein, we first analyze the consistency of simulation results with Båth's law and observed magnitude differences in the original catalogs, respectively. We continue with an analysis of the degree of temporal and spatial clustering in simulated catalogs compared with the original event sets. Finally, we evaluate doublet frequencies in simulated catalogs and compare them with historical observations.

Model fit

Table 2 lists the results from models M_0, M_1, M_2, and M_3 for both regions—Japan and southern California, including the LLF values and the branching ratios.

Regarding the log-likelihood values $l(\theta|H_T)$, in both regions, we observe the order $M_1 > M_3 > M_0 > M_2$. According to the log-likelihood measure, we can conclude that the anisotropic shape of the spatial kernel leads to an improved performance, whereas the spatial restriction detracts the quality of the model fits.

One reason for the better performance of the anisotropic models can be found in the optimization process used to define the strike. In fact, the advantage of the anisotropic over the isotropic models was moderately reduced when we ran the models with the originally provided strike angles rather than the optimized ones. However, also in the case of original strikes, the anistropic models were superior with regard to the log-likelihood value.

On the other hand, more generally, the anisotropic shape of the spatial kernel leads to an improved adaptation to the aftershock clouds for the most events. For two exemplary magnitudes M_c 5.0 and 7.5, Figure 3a,b depicts the cumulative distribution functions of the spatial kernels against the normalized distance to the event location (for isotropic models) or rupture segment (for anisotropic models). In both the regions, we can see that the anisotropic models show a significantly narrower shape, which suggests that the estimated rupture segments fit the potential aftershock clouds better than the isotropic point sources and, therefore, tend to bring possible offsprings closer to the trigger source. Although in the

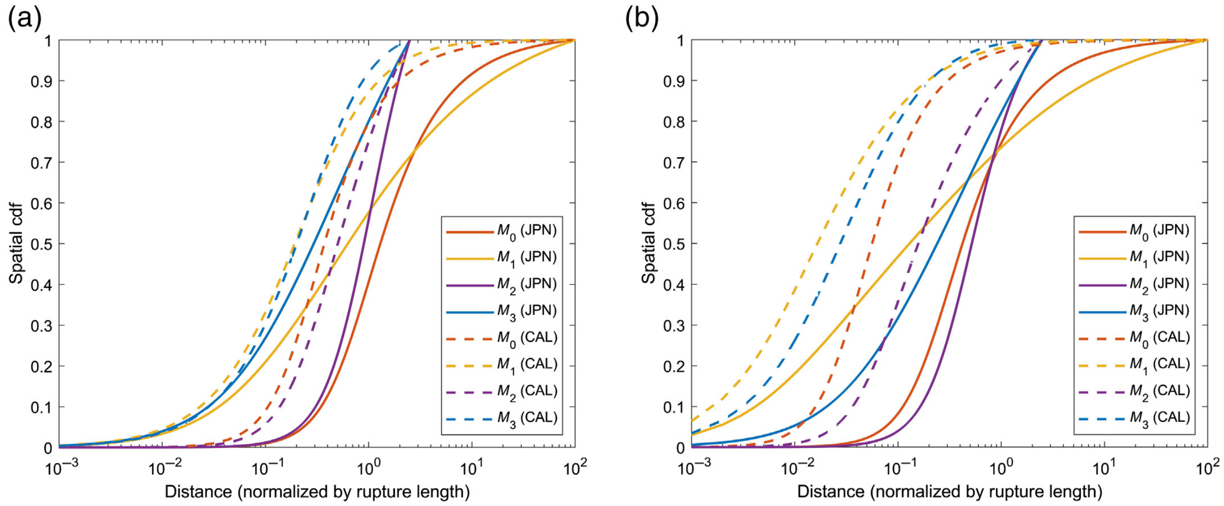


Figure 3. Cumulative distribution functions (cdf) of spatial kernels for trigger magnitudes (a) M_c 5.0 and (b) M_c 7.5. Solid lines show Japan (JPN) models. Dashed lines represent southern California (CAL) models. The x axis is defined as the distance to the point source location (for isotropic models M_0 and M_2) or rupture line (for anisotropic models M_1 and M_3), normalized by the rupture length estimate for the respective region. The color version of this figure is available only in the electronic edition.

Japan models, this narrowing effect is characterized by the dramatic decrease of parameter D, in the southern California results it is modeled by the increase of parameter q.

The characteristic length of the power-law decay, $\sqrt{D\exp(\gamma(m_i - M_c))/\pi}$, has unit km^2, so it really has the dimension of an area. Therefore, its exponential increase does not fully compensate for the faster exponential growth of the 1D rupture length estimates in equation (8). Consequently, especially the anisotropic spatial distributions tend to get narrower relative to the rupture length with increasing trigger magnitude, as observable in Figure 3b. We conclude from this that the anisotropic kernel gains relevance in the upper magnitude ranges. Moreover, we observe that southern Californian models generally fit narrower shapes than Japan's models. This agrees with the predominant faulting style. In California, strike-slip events on approximately vertical faults dominate, whereas shallow-dipping mechanisms are common in Japan, widening the epicentral aftershock distributions.

The generally inferior log-likelihood values of the restricted models can be explained by the additional constraint imposed to the model by the limitation of the extent of the spatial kernels. Any decline of flexibility inevitably leads to a lesser (or equal) overall model performance.

In this context, we observe that the parameter estimates of $\tilde{A}$, which represent the average number of aftershocks triggered by an event with threshold magnitude $m = M_c$, are substantially lower, in Japan even more than halved, when comparing a restricted model with the according unrestricted model. On

TABLE 2
Overview of Model Fit Results for Japan and Southern California

	Japan				Southern California			
Outcomes	M_0	M_1	M_2	M_3	M_0	M_1	M_2	M_3
$l(\theta\|H_T)$	−21,063	−18,626	−22,684	−19,814	28,444	30,266	27,144	30,003
ν_{branch}	0.52	0.52	0.45	0.45	0.60	0.57	0.54	0.53
$\mu(day^{-1})$	0.51	0.54	0.61	0.64	0.18	0.19	0.21	0.21
$\tilde{A}$	0.26	0.24	0.12	0.11	0.34	0.27	0.25	0.22
$\alpha(mag^{-1})$	1.21	1.28	1.78	1.84	1.18	1.41	1.48	1.59
c (days)	0.015	0.017	0.013	0.014	0.011	0.012	0.013	0.012
p	1.02	1.05	1.00	1.03	1.07	1.08	1.08	1.09
$D_{M=4.0}(km^2)$	2.274	0.194	2.466	0.117	0.441	0.584	0.403	0.849
$\gamma(mag^{-1})$	1.72	1.73	2.05	2.48	1.37	1.78	1.86	1.95
q	1.43	1.20	1.60	1.21	1.48	1.71	1.19	1.95

The parameter D has been scaled to $D_{M=4.0} = D\exp(\gamma(4.0 - M_c))$ to make results cross-comparable among regions (for Japan $D_{M=4.0} = D$ since $M_c = 4.0$).

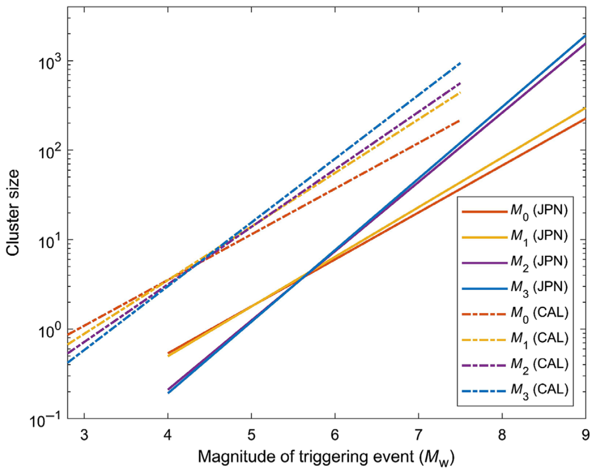


Figure 4. Expected cluster sizes according to equation (14). The x axis states the magnitude of the sequence-initiating mainshock event. The y axis is on logarithmic scale and denotes the average number of cluster members. Solid lines show Japan (JPN) models, starting from catalog cutoff magnitude $M_c = 4.0$. Dashed lines represent southern California (CAL) models, starting from catalog cutoff magnitude $M_c = 2.8$ and ending at the assumed maximum magnitude M_c 7.5. The color version of this figure is available only in the electronic edition.

the other hand, the restricted models lead to highly increased estimates for parameter α signifying an acceleration of the exponential increase of aftershock productivity with growing trigger magnitudes. Figure 4 displays the exponential relation of the expected cluster sizes according to equation (14), that is, including direct and secondary aftershocks to the initiating mainshock magnitude on a logarithmic scale. Although the restricted models start at a lower base, they cross the lines of unrestricted models at about magnitude $M_{cross} = 5.6$ for Japan models and $M_{cross} = 4.4$ for southern California models. In other words, on average, events with $m \geq M_{cross}$ are expected to trigger more aftershocks in restricted models than in unrestricted models.

Because the vast majority of events in the original catalogs have magnitudes $m < M_{cross}$, we may expect that the observed shift of aftershock productivity from small to large events leads to, in total, fewer identified trigger relations in restricted models. Indeed, the restricted models reveal smaller branching ratios and, contrarily, larger background rates. Therefore, we may conclude that the spatial restriction eliminates some trigger relations between more distant events with relatively small magnitudes that are consequently either associated with the background seismicity or with the another stronger trigger event with larger spatial extent. In particular, the latter case provides an explanation for the greater estimates of parameter α. Furthermore, under the realistic assumption that there were more trigger relations in reality than identified in the models, the absolute loss of identified trigger relations to background seismicity would explain the inferior log-likelihood values.

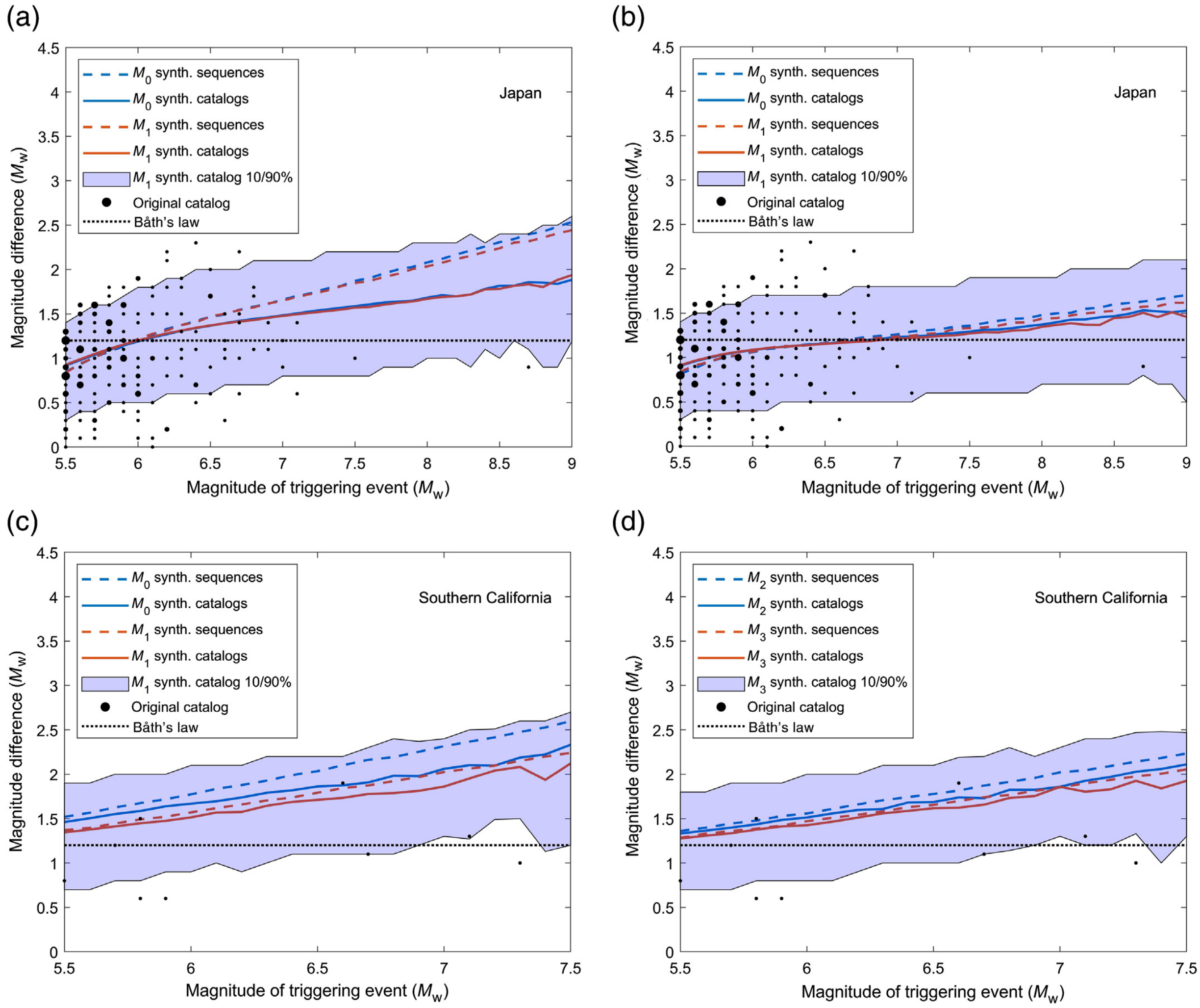
We further notice in Figure 4 and Table 2 that in southern California models the anisotropy of the spatial kernels has far more impact on expected cluster sizes than in Japan models. Cross-comparisons of cluster sizes between the two regions are only valid if the cluster sizes of southern California are downscaled by $\exp(-1.2\alpha)$, accounting for the difference of 1.2 of the magnitude thresholds. The clustering is on a generally comparable level, despite we note a more gradual growth due to smaller α estimates for restricted models in southern California.

Båth's law

Figure 5 depicts the mean magnitude differences between an earthquake and the strongest event following in the specified time–space domain in simulated catalogs in comparison with those in the respective original catalog. The corresponding algorithm is outlined in the Quality Measures section. Figure 5a presents the results for the unrestricted models M_0 and M_1 in Japan. Both models appear to estimate almost identical magnitude differences, with a significant slope for increasing reference magnitudes of the triggering event. On average, the simulated catalogs seem to continuously overestimate the magnitude difference for magnitudes $M_c > 6.8$ compared with the original dataset, with some data points located even outside of the 10%–90% confidence interval for model M_1. The divergence between the results of the simulated catalogs and sequences can be explained by the impact of independent events that are not contained in the pure sequences. The effect intensifies with increasing magnitudes due to the exponential growth of the spatial window size.

According to Figure 5b, Japan's restricted models show a better agreement with the original catalog. There are no data points for magnitudes $M_c \geq 6.8$ outside of the 10%–90% confidence interval for model M_3. The slope of the curves is smaller, which suggests better accordance with Båth's law hypothesis that the magnitude difference is independent of the trigger magnitude. The smaller divergence between catalog and simulation results emphasizes that the improvement is caused by the increase of the average cluster sizes for the investigated magnitude ranges, as shown in Figure 4. This increases the chance of strong aftershocks, and it reduces the relative impact of independent events in the considered time–space domain, at the same time.

The results for southern California, depicted in Figure 5c,d, show similar trends. Approximately half of the historic events have magnitude differences outside of the 10%–90% confidence interval in both the models. In general, southern California models estimate considerably larger and faster-growing magnitude differences than Japan models, reaching up to two magnitude units for the maximum magnitude M_w 7.5. This observation can be explained by the more moderate increase of cluster sizes due to smaller estimates of α. Comparing the two regions, we conclude that the restricted



models work better and lead to more pronounced improvements in Japan than in California.

Figure 5. Approximations of the average magnitude difference between a considered mainshock event and the strongest event following in the specified time–space window, for (a) unrestricted models M_0 and M_1 in Japan (JPN), (b) restricted models M_2 and M_3 in JPN, (c) unrestricted models M_0 and M_1 in southern California (CAL), and (d) restricted models M_2 and M_3 in CAL. Solid lines show catalog simulations, and dashed lines represent sequence simulations. The shaded range visualizes the 10%–90% confidence interval of the respective catalog simulation. Black dots represent observations in the underlying original catalogs and are sized according to the number of points stacked. The horizontal dotted line is consistent with the Båth's law prediction of a magnitude difference of 1.2 units independent of the absolute size of the trigger magnitude. The color version of this figure is available only in the electronic edition.

Spatial and temporal clustering

Figure 6 analyzes the degree of temporal and spatial clustering in synthetic catalogs compared with the respective original catalog. For these plots only, we generated the synthetic catalogs with magnitudes sampled from the empirical magnitude distribution observed in the respective original catalogs, instead of using the Gutenberg–Richter distribution (equation 12) with estimated parameter β.

The reason is that, using magnitudes sampled from equation (12), we observed a deficiency of extremely strong events in the synthetic catalogs compared with the original catalogs, which suggests that equation (12) tends to underestimate the tail of the empirical magnitude size distribution in the observational data. Consequently, the synthetic catalogs would lack some influential trigger events that would otherwise cause sporadic peaks in the spatiotemporal distribution of event occurrences.

Figure 6a depicts boxplots of the CV of event occurrence numbers in monthly time intervals of synthetic catalogs for

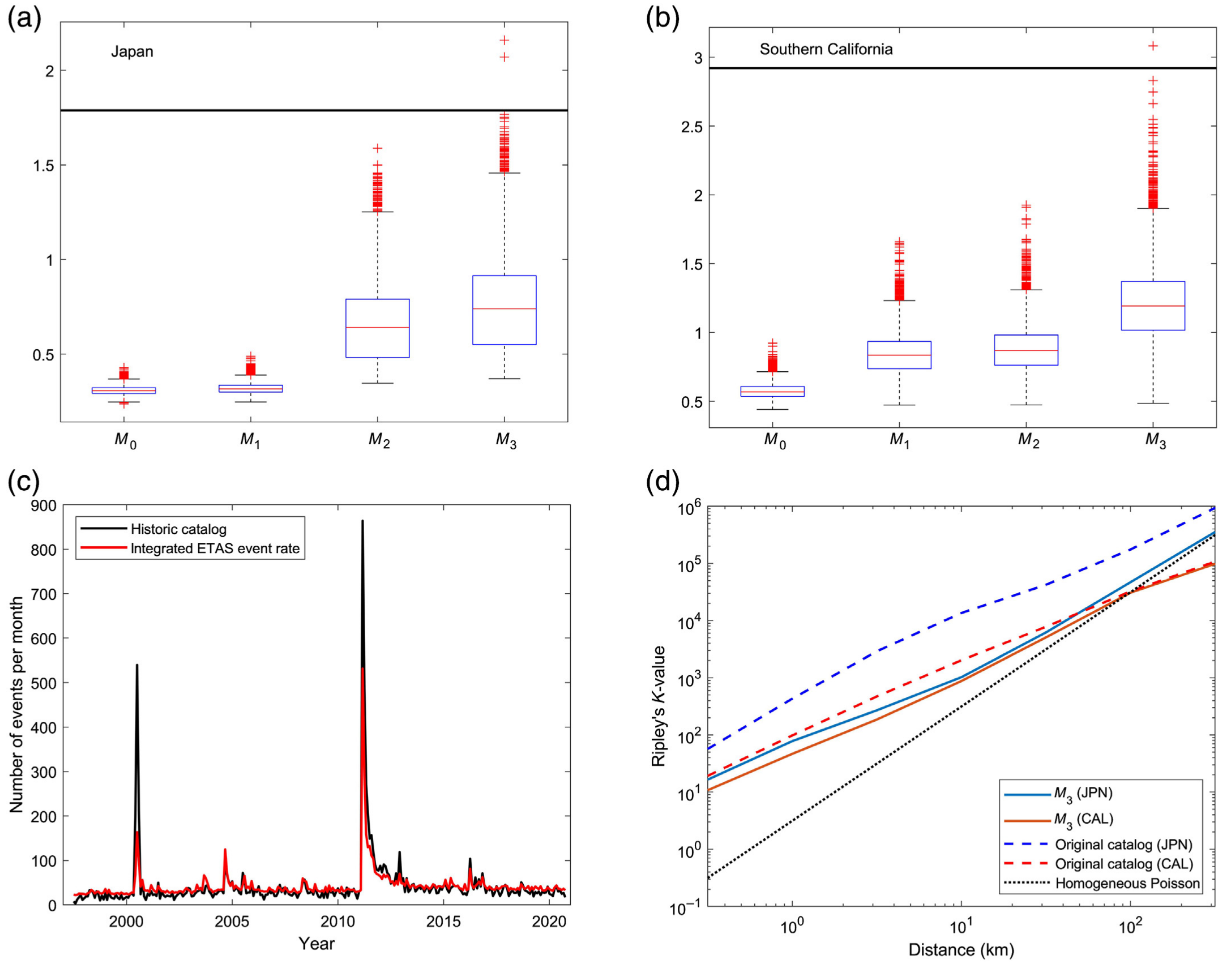


Figure 6. Boxplot representation of the coefficients of variation (CV) of monthly numbers of event occurrences in the simulated catalogs, based on the four estimated models, for (a) Japan and (b) southern California. The black horizontal line represents the CV of the respective original earthquake catalog. The red plus symbols represent outliers. (c) Comparison of monthly event occurrences between the original Japan catalog (black line) and the epidemic-type aftershock sequence (ETAS) rate for Japan's model M_3, integrated piecewise for the monthly integrals, based on trigger contributions of the original history of events (red line). (d) Analysis of the degree of spatial clustering by Riley's *K*-function. Solid lines represent results for synthetic catalogs, generated by model M_3 for Japan (JPN) and southern California (CAL). Dashed lines show results for the respective original earthquake catalogs. The dotted black line represents Riley's *K*-function values for a homogeneous Poisson process. Values above indicate clustering, values below signify inhibition. The color version of this figure is available only in the electronic edition.

Japan. We observe that, on average, the variance of monthly event occurrences in simulations is by factors smaller than in the original catalog, displayed by the horizontal black line. However, the restricted models tend to produce considerably more temporal variation, with some pronounced outliers, than the unrestricted models. The same observation is made for southern California in Figure 6b. Furthermore, the CVs seem to correlate with the expected cluster sizes of strong events, as shown in Figure 4. For instance, the anisotropy of the spatial kernel leads to a stronger increase of both productivity parameter α and the temporal clustering in southern California than in Japan.

Figure 6c demonstrates that the observed smoothing of temporal event occurrences is not a pure side effect of catalog simulations. Exemplary for model M_3 in Japan, we plotted the curve of monthly event occurrences in the original catalog against the expected number of event occurrences predicted by the ETAS event rate. More precisely, the latter is computed as the total ETAS event rate (see $\Lambda_\theta(\mathbb{T}, \mathbb{S}|H_\mathbb{T})$ in equation 11), stepwise integrated over the monthly intervals instead of the entire target time window $\mathbb{T}$, which provides us an estimate of the expected number of events occurring in the considered month. This monthly forecast is thus purely based on the fit of the model parameters and the original, nonsimulated history of events.

On the one hand, the integrated rate clearly underpredicts event occurrences in peak months, whereas, on the other hand, it overrates the seismicity in relatively calm months. This contrast is an immediate consequence of the log-likelihood-based model estimation algorithm, which requires that for the optimal set of parameters, the ETAS rate, integrated over the entire time–space target window, equalizes the exact number of target events. Thus, once underestimating the pronounced peaks in the most active months, the rate needs to compensate for this inaccuracy by predicting larger occurrence rates in rather inactive months, which ultimately leads to a clear smoothing of the temporal occurrence curve. We hypothesize that this compensation is caused by an overprediction of both the background seismicity and the triggering potential of small events and an underprediction of the triggering power of strong events.

Figure 6d sheds light on the degree of spatial clustering, measured by Ripley's K-function (equation 15). For the sake of clearer visualization, we only present the spatially most strongly clustered models M_3 in both the regions. Nevertheless, we observe that spatial clustering is underestimated compared with the respective original catalogs in both the regions. In general, event occurrences in southern Californian seem less intensely clustered in space than in Japan. The kink in the curves, which in the case of southern California even suggests inhibition, is a boundary effect due to the limited polygon areas.

Earthquake doublets

Finally, Figure 7 analyzes the occurrence rates of doublets in the simulated catalogs and sequences. Figure 7a compares the percentage that an event finds a doublet partner depending on its magnitude for the four models and both simulated sequences and catalogs in Japan. For the sake of clarity, the data are smoothed by aggregating magnitude intervals.

We observe that the restricted models show substantially larger doublet chances than the unrestricted models, which is consistent with our previous findings regarding the larger cluster sizes, the larger degree of temporal and spatial clustering, and the lower average magnitude differences to the strongest event in the time–space domain spanned by an event. Also, doublet percentages decrease with growing magnitudes, which accompanies the earlier observation of increasing the Båth law magnitude differences.

It is also worth mentioning that the proportion of events that find a doublet partner is considerably larger within a simulated catalog than in a synthetic sequence. This implies that independent seismic background events or unrelated clusters generate a nonnegligible fraction of doublets.

Figure 7b shows this aspect in more detail for models M_0 and M_3 in Japan. Conditional on realized doublet pairs, it shows the inverse proportions of doublets consisting of two events from the same cluster and doublets composed by two independent events. The corresponding triggering relationship is known in simulations. For small triggering magnitudes, in-cluster doublets make up a much larger proportion. The share declines with increasing trigger magnitude, however, much stronger for model M_0 than for model M_3. In the case of model M_0, independent doublets get even more likely than in-cluster doublets for triggering magnitudes larger than M_w 7.6.

These observations can be explained by the more rapid exponential growth of the area of the spatial window than the aftershock productivity and expected cluster sizes. According to the scaling relations (equation 8), the area of the spatial window covering the surrounding of two and a half rupture lengths is $\pi(2.510^{-2.37+0.57m})^2$. Consequently, the area grows by factor $(10^{0.57})^2 = 10^{1.14} \approx 13.8$, which is faster than the magnitude-dependent growth of aftershock productivity and expected cluster sizes, $\exp(\alpha)$, for all $\alpha < 2.62$. Following this line of argument, we can explain the growing impact of independent and unrelated events with increasing trigger magnitudes. The curves for model M_3 are more robust compared with M_0, because the larger aftershock productivity and expected cluster sizes resulting from greater estimates of parameter α better balance out the growth of the spatial window.

Figure 7c,d compares the doublet rate predictions for the Japan models M_0 and M_3 with analogously measured doublet percentages in historic catalogs. As benchmarks, we use the original NIED Japan catalog used for the ETAS model estimation, as well as a regional and a global extract from the ISC-GEM catalog. Respecting the stepwise completeness levels in the ISC-GEM catalog, we counted doublets for events with magnitudes $M_w \geq 5.9$ from the year 1960 and for events with magnitudes $M_w \geq 6.7$ starting in 1918. In particular, this allows for a reliable search of doublet partners with a maximum magnitude difference of M_w 0.4 units. Because of the relatively small sample sizes in historical data, we grouped the events in the four magnitude intervals [5.9,6.0], [6.1,6.2], [6.3,6.6], and $[6.7, \infty)$. Because of its limited time (24 yr), the regional NIED catalog provides only between 18 and 37 events in the respective magnitude intervals and, therefore, has limited statistical significance, especially in the higher magnitude ranges. Furthermore, we obtained 70–105 events in the regional extract for Japan of the ISC-GEM catalog from 1918 and 1362 to 2219 events in the entire ISC-GEM dataset. In the simulated catalogs, we isolated all events with magnitude M_w 8.7 from the last interval, because they would dominate the statistic due to the manual sampling of the Tohoku event.

Figure 7c demonstrates that model M_0 tends to underestimate the doublet occurrence probabilities observed in the three benchmark catalogs. The simulations accurately fit two out of four data points of the original NIED catalog, which, however, is an uncertain statistic due to its small sample size. The more stable curves of the long-term Japan and global benchmark catalogs are mostly located outside of the 10%–90% confidence interval.

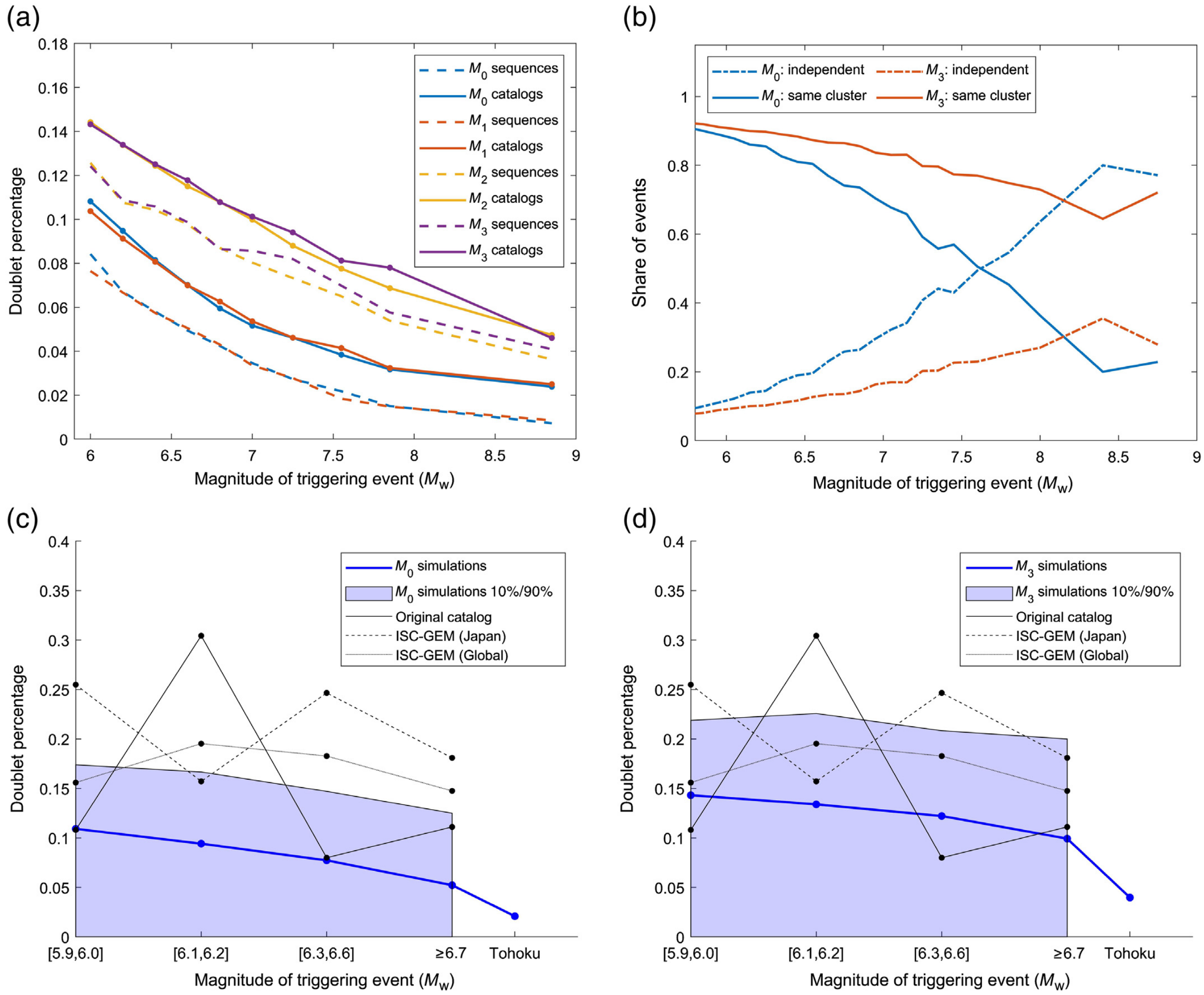


Figure 7. (a) Percentages of doublet occurrences, depending on the considered event magnitude, for the four model variants in Japan. Solid lines represent simulated catalogs. Dashed lines show simulated sequences. Magnitudes are aggregated in 0.2 magnitude unit steps from M_w 5.9 to 7.1, then in 0.3 unit steps up to M_w 8.0, followed by one interval for all magnitudes above. (b) Proportions of doublet pairs generated by (1) independent seismic background events or unrelated clusters (dashed-dotted lines), or (2) events of the same cluster (solid lines). Results are presented for models M_0 and M_3 in Japan. (c,d) Comparison of the doublet occurrence frequencies in synthetic catalogs (blue lines) to historic catalogs (black lines), for (c) model M_0 and (d) model M_3, both Japan. Shaded ranges represent 10%–90% confidence interval (CI) of the synthetic catalogs. Events are aggregated in the magnitude intervals labeled on the *x* axis. Tohoku events are extracted in simulated catalogs. The color version of this figure is available only in the electronic edition.

Model M_3, shown in Figure 7d, moves considerably closer to the long-term benchmark catalog curves and appears to provide a rather adequate prediction of doublet probabilities compared with the original NIED event set. The 10%–90% confidence interval covers all data points of the global catalog and two out of four samples from the long-term Japan dataset.

Both models show a downtrend of doublet rates with increasing magnitudes, which reveals itself particularly in the small fraction of doublets initiated by the sampling of the Tohoku event. In contrast, the probability of doublet occurrences seems magnitude-independent, at least in the lower three magnitude ranges, for the ISC-GEM catalog extracts, which reminds us of the self-similarity of earthquake clustering observed according to Båth's law. The comparison, however, is unavoidably biased due to the subjective specification of our time–space domain in the doublet definition and because of the fact that we do not prohibit doublets produced by independent events not belonging to the same triggered sequence.

The historical observations for southern California do not provide a sufficient database for benchmarking. We only observed seven events in the overall magnitude range from

$M_w \geq 5.9$ in the original catalog, with two of them being a doublet (both in magnitude range [6.1,6.2]). In the regional extract of the ISC-GEM catalog since 1918, we found an overall number of 15 events, with one of them being doublets (in the third magnitude range). In the latter, this would signify a chance of 6.7% that an event finds a doublet partner, which is less than half of the global percentages shown in Figure 7c,d. However, the Californian models predict a chance of only 3% (model M_3) or even 1.6% (model M_0) for doublet occurrences.

Sensitivity of results

The results described previously, especially the estimated doublet probabilities, are clearly dependent on the rather subjective definition of the temporal and spatial constraints of 365 days and 2.5 rupture lengths as well as the magnitude window of M_w 0.4 units. In accordance with intuition, sensitivity tests have shown that a decrease of one of the three criteria led to lower doublet probabilities in both the simulated and historical data, and vice versa. However, the relative behavior of the four models under consideration, among each other and in comparison with historical catalogs, and therefore the central conclusions, remain the same.

SUMMARY AND CONCLUSION

We compared seismicity generated with four variants of the ETAS model to earthquake catalogs for Japan and southern California. More precisely, we tested isotropic and anisotropic as well as unrestricted and restricted spatial kernels. The central objective of this study was to find out which of the four models the best describes the clustering of particularly strong events and leads to the most realistic predictions of the occurrence probabilities of earthquake doublets. Rather subjectively, we defined a doublet as a pair of an earthquake with any other event occurring during the next 365 days and within a distance of 2.5 rupture lengths to the considered event, with a magnitude difference of no more than 0.4 units. By assuming an identical magnitude size distribution for triggered and independent events, we analyze the impact of aftershock productivity and cluster sizes on cluster properties and doublet occurrences.

The results indicate that the conventional, unrestricted isotropic model poorly represents clusters triggered by particularly large-magnitude earthquakes. We found that this model estimates too large-magnitude differences between a strong earthquake and the largest event in the specified time–space window that it tends to highly underestimate the degree of temporal and spatial clustering by smoothing out the occurrence times and locations, and that it tends to underestimate the chances of doublet occurrence. This stands in contrast to the global catalog scale measures such as the log-likelihood value, which do not incorporate these weaknesses and would attest a relatively high quality to the conventional model.

The anisotropic spatial kernel improves the overall fit of the model but cannot noticeably alleviate the weaknesses of the unrestricted model variants. Perhaps, it shows its strengths primarily in combination with Uniform California Earthquake Rupture Forecast–ETAS models in which crustal fault structures, subduction zones, and multisegment ruptures are incorporated on a detailed level (Field *et al.*, 2017).

By shifting triggering potential from smaller to larger events and therefore increasing cluster sizes of strong trigger events, the restriction of the spatial kernel to 2.5 rupture lengths promotes more realistic estimations of the magnitude difference to the strongest following event and of the doublet probability, compared with historical observations. The temporal and spatial variability of event occurrences rises, additionally indicating more pronounced clustering. However, the improvements in the representation of strong earthquake clusters are at the expense of a decline of the log-likelihood value, because trigger relations in the smallest magnitude ranges get lost.

Again, the anisotropic model variant improves the overall fit of the model but has negligible impact on the temporal and spatial clustering and the doublet's occurrence.

We conclude that global catalog scale measures such as the log-likelihood value or the AIC criterion are not an adequate tool for evaluating ETAS model fits if the representation of strong event clusters is of particular interest. It is in the nature of these measures that they show better performance when more trigger relations are detected. Consequently, a model that is given more freedom, such as the unrestricted variants, will always outperform the more conditioned variants, such as the restricted variants in our study. However, this may lead to trigger relations between events that are, from a standpoint of reason, improbable. In other words, the conventional model does a good job in identifying triggered events, but it does a relatively poor job in assigning the aftershocks to their most realistic triggers, which goes to the benefit of the smaller events.

Certainly, this deficiency can be partly explained by the well-known and extensively studied biases in the use of the ETAS model, such as earthquake location uncertainty, the catalog cutoff magnitude, and short-term incompleteness. In our study, we have accounted for the latter by applying blind periods after strong events according to Helmstetter *et al.* (2006).

The spatial restriction tested in our models, however, demonstrates that we can improve aftershock to trigger assignments and, therefore, strengthen the aftershock productivity of strong events by giving the ETAS model more guidance in terms of conditions. Given the assumption of an identical magnitude size distribution for triggered and independent events, aftershock productivity becomes the dominant driver for cluster properties. The larger the size of a cluster, the smaller the magnitude difference to the strongest following event and the larger the chance of a doublet to occur. At the same time, a larger cluster size decreases the relative relevance of independent seismicity in the considered time–space window around an earthquake.

Even the restricted models reveal a persistent underestimation of the cluster properties of large earthquakes. We hypothesize that, in reality, the exponential growth of the aftershock productivity with increasing trigger magnitudes should be even larger. This would also increase the underrepresented clustering of events both in time and space.

Future work should emphasize the importance of a correct representation of strong event clusters by the ETAS model. Using only goodness-of-fit measures operating on a global catalog scale provides an inherent risk that a poor representation of extreme clusters remains undetected.

This work has analyzed the impact of aftershock productivity and cluster sizes on the occurrence of earthquake doublets. It has, however, neglected the influence of potentially varying magnitude size distributions that may lead to a correlation of triggering and triggered magnitudes (Gulia *et al.*, 2018; Nandan *et al.*, 2019) and may, therefore, result in modified doublet occurrence probabilities. Positively correlated magnitudes could, therefore, contribute to closing the gap between simulated and observed doublet frequencies. Another, however more profound, research topic is the further evaluation of the impact of faulting types, event characteristics (e.g., dip, rake, depth, and so forth), and local geophysical parameters (e.g., strain rates, heat flow, tectonic plate velocities, and so forth) on the aftershock productivity and ultimately strong event clustering. This could also close the current gap in the most seismic hazard models and lead to a better risk assessment by considering modeled damage based on more realistic, synthetic catalogs, including increased earthquake clustering and doublet occurrences.

DATA AND RESOURCES

The National Research Institute for Earth Science and Disaster Resilience (NIED) earthquake mechanism catalog for Japan (Kubo *et al.*, 2002) was downloaded from www.fnet.bosai.go.jp/event/search.php?LANG=en. The Southern California Earthquake Data Center (SCEDC) focal mechanism (Hauksson *et al.*, 2012) was searched using scedc.caltech.edu/data/alt-2011-yang-hauksson-shearer.html. Global earthquake data were obtained from the International Seismological Centre–Global Earthquake Model (ISC-GEM) Global Instrumental Earthquake Catalogue (Di Giacomo *et al.*, 2018) at www.isc.ac.uk/iscgem/download.php. The epidemic-type aftershock sequence (ETAS) model code used for this research was initially based on the CRAN R package repository ETAS (Jalilian, 2019), available at https://CRAN.R-project.org/package=ETAS. The package is based on the original Fortran implementation etas8p, available at http://bemlar.ism.ac.jp/zhuang/software.html. All websites were last accessed in January 2021.

DECLARATION OF COMPETING INTERESTS

The authors acknowledge that there are no conflicts of interest recorded.

ACKNOWLEDGMENTS

The authors thank the anonymous reviewers for their helpful and constructive feedback that has significantly improved the work. Financial support for this work was provided by Munich Re through a scholarship granted to the first author and by the Department of Statistics at Ludwig Maximilian University of Munich.

REFERENCES

Abdelnaby, A. E. (2012). Multiple earthquake effects on degrading reinforced concrete structures, *Ph.D. Thesis*, University of Illinois at Urbana-Champaign.

Bach, C., and S. Hainzl (2012). Improving empirical aftershock modeling based on additional source information, *J. Geophys. Res.* **117,** no. B04312, doi: 10.1029/2011JB008901.

Blaser, L., F. Krüger, M. Ohrnberger, and F. Scherbaum (2010). Scaling relations of earthquake source parameter estimates with special focus on subduction environment, *Bull. Seismol. Soc. Am.* **100,** no. 6, 2914–2926, doi: 10.1785/0120100111.

Chu, A., F. P. Schoenberg, P. Bird, D. D. Jackson, and Y. Y. Kagan (2011). Comparison of ETAS parameter estimates across different global tectonic zones, *Bull. Seismol. Soc. Am.* **101,** no. 5, 2323–2339, doi: 10.1785/0120100115.

Cornell, C. A. (1968). Engineering seismic risk analysis, *Bull. Seismol. Soc. Am.* **58,** no. 5, 1583–1606.

de Arcangelis, L., C. Godano, and E. Lippiello (2018). The overlap of aftershock coda waves and short-term postseismic forecasting, *J. Geophys. Res.* **123,** 5661–5674, doi: 10.1029/2018JB015518.

Di Giacomo, D., E. R. Engdahl, and D. A. Storchak (2018). The ISC-GEM earthquake catalogue (1904-2014): Status after the extension project, *Earth Syst. Sci. Data* **10,** 1877–1899, doi: 10.5194/essd-10-1877-2018.

Felzer, K. R., R. E. Abercrombie, and G. Ekström (2004). A common origin for aftershocks, foreshocks, and multiplets, *Bull. Seismol. Soc. Am.* **94,** no. 1, 88–98, doi: 10.1785/0120030069.

Field, E. H., K. R. Milner, J. L. Hardebeck, M. T. Page, N. van der Elst, T. H. Jordan, A. J. Michael, B. E. Shaw, and M. J. Werner (2017). A spatiotemporal clustering model for the third Uniform California Earthquake Rupture Forecast (UCERF3-ETAS): Toward an operational earthquake forecast, *Bull. Seismol. Soc. Am.* **107,** no. 3, 1049–1081, doi: 10.1785/0120160173.

Gibowicz, S. J., and S. Lasocki (2005). Earthquake doublets and multiplets in the Fiji-Tonga-Kermadec region, *Acta Geophys. Pol.* **53,** no. 3, 239–274.

Gulia, L., A. P. Rinaldi, T. Tormann, G. Vannucci, B. Enescu, and S. Wiemer (2018). The effect of a mainshock on the size distribution of the aftershocks, *Geophys. Res. Lett.* **45,** no. 24, 13,277–13,287, doi: 10.1029/2018GL080619.

Gutenberg, B., and C. F. Richter (1944). Frequency of earthquakes in California, *Bull. Seismol. Soc. Am.* **34,** 185–188, doi: 10.1038/156371a0.

Hainzl, S. (2016a). Rate-dependent incompleteness of earthquake catalogs, *Seismol. Res. Lett.* **87,** no. 2A, 337–344, doi: 10.1785/0220150211.

Hainzl, S. (2016b). Apparent triggering function of aftershocks resulting from rate-dependent incompleteness of earthquake catalogs, *J. Geophys. Res.* **121,** no. 9, 6499–6509, doi: 10.1002/2016JB013319.

Hainzl, S., and D. Marsan (2008). Dependence of the Omori-Utsu law parameters on main shock magnitude: Observations and modeling, *J. Geophys. Res.* **113,** no. B10309, doi: 10.1029/2007JB005492.

Hainzl, S., A. Christophersen, and B. Enescu (2008). Impact of earthquake rupture extensions on parameter estimations of point-process models, *Bull. Seismol. Soc. Am.* **98,** no. 4, 2066–2072, doi: 10.1785/0120070256.

Hainzl, S., O. Zakharova, and D. Marsan (2013). Impact of aseismic transients on the estimation of aftershock productivity parameters, *Bull. Seismol. Soc. Am.* **103,** no. 3, 1723–1732, doi: 10.1785/0120120247.

Hauksson, E., W. Yang, and P. M. Shearer (2012). Waveform relocated earthquake catalog for Southern California (1981 to June 2011), *Bull. Seismol. Soc. Am.* **102,** no. 5, 2239–2244, doi: 10.1785/0120120010.

Hauksson, E., C. Yoon, E. Yu, J. R. Andrews, M. Alvarez, R. Bhadha, and V. Thomas (2020). Caltech/USGS Southern California Seismic Network (SCSN) and Southern California Earthquake Data Center (SCEDC): Data availability for the 2019 Ridgecrest sequence, *Seismol. Res. Lett.* **91,** no. 4, 1961–1970, doi: 10.1785/0220190290.

Helmstetter, A., and D. Sornette (2003). Båth's law derived from the Gutenberg-Richter law and from aftershock properties, *Geophys. Res. Lett.* **30,** no. 20, 2069, doi: 10.1029/2003GL018186.

Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2006). Comparison of short-term and time-dependent earthquake forecast models for Southern California, *Bull. Seismol. Soc. Am.* **96,** no. 1, 90–106, doi: 10.1785/0120050067.

Hutton, K., J. Woessner, and E. Hauksson (2010). Earthquake monitoring in Southern California for seventy-seven years (1932-2008), *Bull. Seismol. Soc. Am.* **100,** no. 2, 423–446, doi: 10.1785/0120090130.

Jalilian, A. (2019). ETAS: An R package for fitting the space-time ETAS model to earthquake data, *J. Stat. Softw.* **88,** no. 1, 1–39, doi: 10.18637/jss.v088.c01.

Kagan, Y. Y. (2004). Short-term properties of earthquake catalogs and models of earthquake source, *Bull. Seismol. Soc. Am.* **94,** no. 4, 1207–1228, doi: 10.1785/012003098.

Kagan, Y. Y., and D. D. Jackson (1999). Worldwide doublets of large shallow earthquakes, *Bull. Seismol. Soc. Am.* **89,** no. 5, 1147–1155.

Kagan, Y. Y., P. Bird, and D. D. Jackson (2010). Earthquake patterns in diverse tectonic zones of the globe, *Pure Appl. Geophys.* **167,** no. 6, 721–741, doi: 10.1007/s00024-010-0075-3.

Kagermanov, A., and R. Gee (2019). Cyclic pushover method for seismic assessment under multiple earthquakes, *Earthq. Spectra* **35,** no. 4, 1541–1558, doi: 10.1193/010518EQS001M.

Kubo, A., E. Fukuyama, H. Kawai, and K. Nonomura (2002). NIED seismic moment tensor catalogue for regional earthquakes around Japan: Quality test and application, *Tectonophysics* **356,** 23–48, doi: 10.1016/S0040-1951(02)00375-X.

Lay, T., and H. Kanamori (1980). Earthquake doublets in the Solomon Islands, *Phys. Earth Planet. In.* **21,** 283–304.

Marzocchi, W., A. M. Lombardi, and E. Casarotti (2014). The establishment of an operational earthquake forecasting system in Italy, *Seismol. Res. Lett.* **85,** no. 5, 961–969, doi: 10.1785/0220130219.

McGuire, R. K. (2008). Probabilistic seismic hazard analysis: Early history, *Earthq. Eng. Struct. Dynam.* **37,** 329–338, doi: 10.1002/eqe.765.

Nandan, S., G. Ouillon, and D. Sornette (2019). Magnitude of earthquakes controls the size distribution of their triggered events, *J. Geophys. Res.* **124,** no. 3, 2762–2780, doi: 10.1029/2018JB017118.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes, *J. Am. Stat. Assoc.* **83,** no. 401, 9–27.

Ogata, Y. (1998). Space-time point-process models for earthquake occurrences, *Ann. Inst. Stat. Math.* **50,** no. 2, 379–402.

Ogata, Y. (2011). Significant improvements of the space-time ETAS model for forecasting of accurate baseline seismicity, *Earth Planets Space* **63,** no. 3, 217–229, doi: 10.5047/eps.2010.09.001.

Ogata, Y., and K. Katsura (1993). Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogues, *Geophys. J. Int.* **113,** no. 3, 727–738.

Ogata, Y., and J. Zhuang (2006). Space-time ETAS models and an improved extension, *Tectonophysics* **413,** nos. 1/2, 13–23, doi: 10.1016/j.tecto.2005.10.016.

Page, M. T., N. van Der Elst, J. Hardebeck, K. Felzer, and A. J. Michael (2016). Three ingredients for improved global aftershock forecasts: Tectonic region, time-dependent catalog incompleteness, and intersequence variability, *Bull. Seismol. Soc. Am.* **106,** no. 5, 2290–2301, doi: 10.1785/0120160073.

Papadopoulos, A. N., P. Bazzurro, and W. Marzocchi (2020). Exploring probabilistic seismic risk assessment accounting for seismicity clustering and damage accumulation: Part I. Hazard analysis, *Earthq. Spectra* **37,** doi: 10.1177/8755293020957338.

Ripley, B. D. (1976). The second-order analysis of stationary point processes, *J. Appl. Probab.* **13,** no. 2, 255–266.

Seif, S., A. Mignan, J. D. Zechar, M. J. Werner, and S. Wiemer (2017). Estimating ETAS: The effects of truncation, missing data, and model assumptions, *J. Geophys. Res.* **122,** no. 1, 449–469, doi: 10.1002/2016JB012809.

Shearer, P. M. (2012). Self-similar earthquake triggering, Bath's law, and foreshock/aftershock magnitudes: Simulations, theory, and results for southern California, *J. Geophys. Res.* **117,** no. B06310, doi: 10.1029/2011JB008957.

Storchak, D. A., D. Di Giacomo, E. R. Engdahl, J. Harris, I. Bondár, W. H. Lee, P. Bormann, and A. Villaseñor (2015). The ISC-GEM global instrumental earthquake catalogue (1900-2009): Introduction, *Phys. Earth Planet. In.* **239,** 48–63, doi: 10.1016/j.pepi.2014.06.009.

Utsu, T., Y. Ogata, and R. S. Matsu'ura (1995). The centenary of the Omori formula for a decay law of aftershock activity, *J. Phys. Earth* **43,** 1–33.

van Stiphout, T., D. Schorlemmer, and S. Wiemer (2011). The effect of uncertainties on estimates of background seismicity rate, *Bull. Seismol. Soc. Am.* **101,** no. 2, 482–494, doi: 10.1785/0120090143.

Veen, A. (2006). Some methods of assessing and estimating point processes models for earthquake occurrences, *Ph.D. Thesis*, University of California, Los Angeles.

Vere-Jones, D. (1969). A note on the statistical interpretation of Bath's law, *Bull. Seismol. Soc. Am.* **59,** no. 4, 1535–1541.

Wells, D. L., and K. J. Coppersmith (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacements, *Bull. Seismol. Soc. Am.* **84,** no. 4, 974–1002.

Yang, W., E. Hauksson, and P. M. Shearer (2012). Computing a large refined catalog of focal mechanisms for Southern California (1981-2010): Temporal stability of the style of faulting, *Bull. Seismol. Soc. Am.* **102,** no. 3, 1179–1194, doi: 10.1785/0120110311.

Zakharova, O., S. Hainzl, D. Lange, and B. Enescu (2017). Spatial variations of aftershock parameters and their relation to geodetic slip models for the 2010 M_w 8.8 Maule and the 2011 M_w 9.0 Tohoku-oki earthquakes, *Pure Appl. Geophys.* **174,** 77–102, doi: 10.1007/s00024-016-1408-7.

Zhang, L., M. J. Werner, and K. Goda (2018). Spatiotemporal seismic hazard and risk assessment of aftershocks of M 9 megathrust earthquakes, *Bull. Seismol. Soc. Am.* **108,** no. 6, 3313–3335, doi: 10.1785/0120180126.

Zhang, L., M. J. Werner, and K. Goda (2020). Variability of ETAS parameters in global subduction zones and applications to mainshock–aftershock hazard assessment, *Bull. Seismol. Soc. Am.* **110,** 191–212, doi: 10.1785/0120190121.

Zhuang, J., Y. Ogata, and D. Vere-Jones (2002). Stochastic declustering of space-time earthquake occurrences, *J. Am. Stat. Assoc.* **97,** no. 458, 369–380, doi: 10.1198/016214502760046925.

Zhuang, J., Y. Ogata, and T. Wang (2017). Data completeness of the Kumamoto earthquake sequence in the JMA catalog and its influence on the estimation of the ETAS parameters, *Earth Planets Space* **69,** no. 36, doi: 10.1186/s40623-017-0614-6.

Manuscript received 31 March 2021
Published online 7 September 2021

6. Solving three major biases of the ETAS model to improve forecasts of the 2019 Ridgecrest sequence

Contributing article

Grimm, C., Hainzl, S., Käser, M., Küchenhoff, H. (2022a). Solving three major biases of the ETAS model to improve forecasts of the 2019 Ridgecrest sequence. *Stochastic Environmental Research and Risk Assessment*

Code repository

`https://github.com/ChrGrimm/ETASanisotropic`

Copyright information

Author contributions

Christian Grimm (CG), Martin Käser (MK) and Sebastian Hainzl (SH) brought up the idea of applying the ETAS-Anisotropic model, developed in the first contribution (Grimm et al., 2021), to the 2019 Ridgecrest sequence and testing its (spatial) forecasting ability. Helmut Küchenhoff contributed to the initial methodological discussions. CG accounted for the bias of short-term aftershock incompleteness and derived the joint concept of combining ETAS-Anisotropic and ETAS-Incomplete model. CG also implemented the additional numerical code based on the methodological description in the manuscript of SH (Hainzl, 2021). CG designed and conducted the simulation experiments, interpreted results, wrote the initial draft of the paper and prepared figures. SH extensively consulted on the detailed choice of model settings and study design. All authors thoroughly proofread the paper. Multiple discussions between CG, SH and MK led to substantial improvements of figures and interpretations.

Stochastic Environmental Research and Risk Assessment
https://doi.org/10.1007/s00477-022-02221-2

ORIGINAL PAPER

Solving three major biases of the ETAS model to improve forecasts of the 2019 Ridgecrest sequence

Christian Grimm[1] · Sebastian Hainzl[2] · Martin Käser[3,4] · Helmut Küchenhoff[1]

Accepted: 21 March 2022

Abstract

Strong earthquakes cause aftershock sequences that are clustered in time according to a power decay law, and in space along their extended rupture, shaping a typically elongate pattern of aftershock locations. A widely used approach to model earthquake clustering, the *Epidemic Type Aftershock Sequence* (ETAS) model, shows three major biases. First, the conventional ETAS approach assumes isotropic spatial triggering, which stands in conflict with observations and geophysical arguments for strong earthquakes. Second, the spatial kernel has unlimited extent, allowing smaller events to exert disproportionate trigger potential over an unrealistically large area. Third, the ETAS model assumes complete event records and neglects inevitable short-term aftershock incompleteness as a consequence of overlapping coda waves. These three aspects can substantially bias the parameter estimation and lead to underestimated cluster sizes. In this article, we combine the approach of Grimm et al. (Bulletin of the Seismological Society of America, 2021), who introduced a generalized anisotropic and locally restricted spatial kernel, with the *ETAS-Incomplete* (ETASI) time model of Hainzl (Bulletin of the Seismological Society of America, 2021), to define an ETASI space-time model with flexible spatial kernel that solves the abovementioned shortcomings. We apply different model versions to a triad of forecasting experiments of the 2019 Ridgecrest sequence, and evaluate the prediction quality with respect to cluster size, largest aftershock magnitude and spatial distribution. The new model provides the potential of more realistic simulations of on-going aftershock activity, e.g. allowing better predictions of the probability and location of a strong, damaging aftershock, which might be beneficial for short term risk assessment and disaster response.

Keywords ETAS · Short-term incompleteness · Anisotropic spatial kernel · Ridgecrest

1 Introduction

Strong earthquakes are usually observed to cause a pronounced spatio-temporal pattern of aftershocks. More precisely, according to the Omori-Utsu Law (Utsu et al. 1995), the temporal aftershock rate is subject to a power law decrease with time $t - t_{main}$ after the main triggering event, that is,

$$g(t - t_{main}) = (t - t_{main} + c)^{-p} \tag{1}$$

✉ Christian Grimm
christian.grimm@stat.uni-muenchen.de

Sebastian Hainzl
hainzl@gfz-potsdam.de

Martin Käser
martin.kaeser@geophysik.uni-muenchen.de

Helmut Küchenhoff
Kuechenhoff@stat.uni-muenchen.de

[1] Department of Statistics, Ludwig-Maximilians-University Munich, Ludwigstraße 33, 80539 Munich, Germany

[2] GFZ German Research Centre for Geoscience, Physics of Earthquakes and Volcanoes, Helmholtzstraße 6/7, 14467 Potsdam, Germany

[3] Department of Earth and Environmental Sciences, Geophysics, Ludwig-Maximilians-University Munich, Theresienstraße 41, 80333 Munich, Germany

[4] Present Address: Munich Re, Section GeoRisks, Königinstr. 107, 80802 Munich, Germany

Published online: 16 April 2022 Springer with the delay parameter $c > 0$ (usually a few minutes to hours) and the exponent p (usually in the range between $0.8 - 1.2$). It means that the temporal pattern of aftershocks is dominated by events occurring within short time after the mainshock. Figure 1a demonstrates this temporal behavior for the Ridgecrest sequence in California, which produced an M6.4 foreshock on July 4, 2019, followed by an M7.1 mainshock within 34 hours on July 6, 2019.

The observed spatial patterns of aftershock sequences stem from their tendency to occur on or close to the mainshock rupture plane (Marsan and Lengliné 2008). The larger the length-to-width ratio of this plane gets, the more elongate the typical aftershock region becomes. In addition, a higher dip angle reduces the width of the 3D-to-2D projection of the rupture plain to the earth's surface and therefore results in a scatter of two-dimensional aftershock epicenters that can be increasingly well approximated by a line segment.

The prevailing continental tectonic regime in southern California with typically steep, strike-slip faulting favors such elongated aftershock patterns in this region. With the exception of the M6.7 1994 Northridge earthquake, all of the most prominent mainshock-aftershock sequences of the last 40 years (M6.6 1987 Superstition Hill, M7.3 1992 Landers, M7.1 1999 Hector Mine, M7.2 2010 Baja California, M7.1 2019 Ridgecrest) demonstrate distinct linearly elongate scattering of aftershock locations (Hainzl 2021).

In this context, the Ridgecrest sequence is a special case as the M6.4 foreshock simultaneously ruptured two almost orthogonal faults, leading to a double pattern of separate linearly elongate aftershock clouds (Marsan and Ross 2021). Fig. 1b shows that the triggering M6.4 event (yellow pentagram) is located close to the intersection of the two ruptured faults. In contrast, the M7.1 mainshock (yellow hexagram) ruptured only one fault which appears to be the extension of one of the faults activated by the foreshock.

Analyzing and forecasting clustered seismicity is an established discipline in seismological research. Its goal is to understand the evolution of large aftershock sequences and to predict their size, largest aftershock magnitude, spatial distribution etc. A prominent approach to model clustered seismicity is the so-called *Epidemic Type Aftershock Sequence (ETAS)* model, which describes earthquake records as a superposition of independent background seismicity and triggered earthquake sequences (Ogata 1988, 1998). The earthquake triggering component is designed in terms of a branching process and characterized by the triad of (1) trigger-magnitude dependent aftershock productivity, (2) a temporal distribution of aftershock times typically derived from the Omori Law (see Eq. 1), and (3) an usually isotropic spatial distribution of aftershock locations (e.g. Zhuang et al. 2002; Jalilian 2019). Particularly, the aftershock productivity (i.e. expected number of offsprings) for a trigger event with magnitude m is

$$k_{A,\alpha}(m) = A \exp(\alpha (m - M_c)), \tag{2}$$

where parameters $A > 0$ and $\alpha > 0$ control the exponential growth of the trigger potential and M_c is the cut-off magnitude of the analyzed earthquake catalog.

Despite generally producing successful and insightful estimation and forecast results, ETAS models are known to be limited by a number of potential biases. In this article, we present an approach that combines solutions for three main short-comings of the conventional ETAS model, (1) the isotropic spatial aftershock distribution, (2) the infinite extent of the spatial kernel and (3) the short-term incompleteness of earthquake records after strong triggering events.

1.1 Bias 1: isotropic spatial distribution

The common assumption in ETAS models is that spatial aftershock locations are distributed isotropically around the triggering event. It is named as a shortcoming in many publications because it stands in conflict with the above-mentioned observation that aftershocks tend to occur close to the (elongate) rupture plane of the triggering event (Ogata 1998, 2011; Ogata and Zhuang 2006; Hainzl et al. 2008, 2013; Seif et al. 2017; Zhang et al. 2018). The assumption of isotropy is reasonably valid for weak earthquakes with small rupture extensions, but becomes problematic for larger magnitudes, e.g. see the spatial pattern of the Ridgecrest sequence in Fig. 1b. It has been shown that inadequate spatial models can lead to an underestimation of the productivity parameter α (Eq. 2) because the numerous small events take over the role of mimicking the "true" anisotropic distribution (Hainzl et al. 2008, 2013; Grimm et al. 2021).

1.2 Bias 2: infinite spatial extent

Under the common definition of an inifinite spatial kernel, aftershock triggering is disproportionately associated with the more numerous small events, that can more flexibly mimic anisotropic event alignments than the few strong mainshocks. This can lead to unrealistically far trigger impact of small magnitudes and to a substantial underestimation of the direct aftershock productivity of strong events, resulting in a smoothing of temporal event distributions (Grimm et al. 2021).

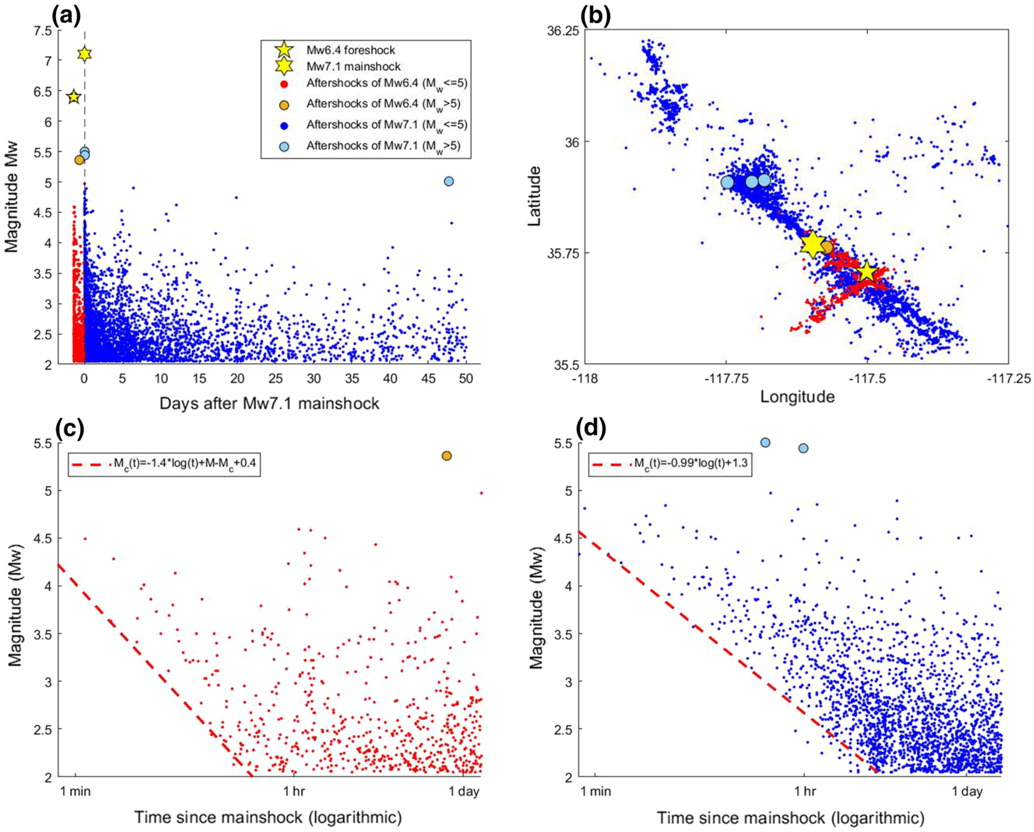


Fig. 1 **a** Magnitudes versus event times of Ridgecrest Mw6.4 (red dots) and Mw7.1 (blue dots) aftershock sequences. Event times are denoted in days before/after Mw7.1 mainshock, the dashed black line represents the time origin (M7.1 event time). Light blue and light red dots mark aftershocks with magnitudes larger than 5. Yellow pentagram symbolizes the Mw6.4 foreshock, and yellow hexagram marks the Mw7.1 mainshock. **b** Aftershock locations of the Ridgecrest Mw6.4 and Mw7.1 sequences. Legend as in **a**. **c** Magnitudes versus logarithmic event times of Ridgecrest Mw6.4 sequence. The dashed red line represents a manually fitted estimate of the empirical completeness function $M_c(t)$. **d** Magnitudes versus logarithmic event times of Ridgecrest Mw7.1 sequence. The dashed red line represents a manually fitted estimate of the empirical completeness function $M_c(t)$

1.3 Bias 3: short-term aftershock incompleteness (STAI)

Strong earthquakes typically cause incomplete aftershock records immediately after their occurrence, mainly due to an overlap of coda waves (Hainzl 2016a; de Arcangelis et al. 2018). Figure 1c and (d) confirms this phenomenon for the aftershock sequences of the M6.4 and M7.1 Ridgecrest events, respectively. Apparently, records of smaller sized aftershocks are missing in the first minutes to hours, somewhat foiling the power law decay of event rates expected from the Omori-Utsu law (Eq. 1). The short-term incomplete event records can therefore hide to a large extent both the ”true” Omori Law decay (Eq. 1) and the ”true” aftershock productivity of the trigger event (Eq. 2) and lead to an overestimation of Omori parameter c and an underestimation of productivity parameter α (Hainzl 2021, 2016b; Page et al. 2016; Seif et al. 2017).

Data-driven uncertainties of event locations and cut-off magnitude as well as the assumption of a *time-invariant seismic background* may lead to further inaccuracies in the parameter estimation (Harte 2013, 2016; Seif et al. 2017). However, they can be neglected in our study because they are either expected to be small in southern California datasets (e.g. location and magnitude uncertainty) or do not apply in an isolated sequence analysis (background miss-specification).

1.4 Scope of this article

In this article, we combine an ETAS approach accounting for short-term incomplete event records with the application of a generalized, anisotropic spatial model that restricts the spatial kernel to the local surrounding of the trigger source. We demonstrate the functionality and superiority of our approaches over the conventional, isotropic ETAS model by means of forecasting experiments for the Ridgecrest sequence.

We utilize the generalized anisotropic and locally restricted spatial kernel suggested by Grimm et al. (2021), which assumes uniform trigger density along an estimated rupture line segment, with power-law decay to the sides and at the end points of the rupture. Zhang et al. (2018) pursued an even more detailed approach, which assumed constant trigger rate in the entire rupture plane, with power-law decay outside of it. Different versions of elliptic Gaussian distributions were introduced and discussed by Ogata (1998, 2011) and Ogata and Zhuang (2006). The latter approaches successfully modeled spatial aftershock patterns, however, they require a new set of parameters and are therefore not flexibly combinable with the conventional, isotropic functionality. In contrast, the kernel of Grimm et al. (2021) represents a generalization of the isotropic function and therefore allows simultaneous anisotropic modeling of some events (e.g. above a certain magnitude threshold) and isotropic modeling of the rest. In order to address the abovementioned particularity of the M6.4 Ridgecrest foreshock, rupturing two almost orthogonal faults, we further generalize the approach by allowing a spatial kernel composed by a weighted superposition of two distinct rupture line segments.

Additionally, we accounts for STAI by applying an ETAS model version that incorporates rate-dependent incompleteness of event records. Recognizing alternative approaches that will be briefly described in the *Methods* section, we choose for the *ETAS-Incomplete (ETASI)* model as recently suggested by Hainzl (2021). For simplicity and to sharpen its focus on the incompleteness detection, Hainzl (2021) neglected the space dimension in his model. As this article combines the ETASI time model of Hainzl (2021) with an adequate, anisotropic spatial kernel it can be seen as the space-including extension of the latter. The focus of this study, however, is on the benefit of modeling the spatial aftershock distribution by a generalized anisotropic spatial kernel, rather than the benefit of the ETASI model.

This article is structured as follows. In the *Methods* section, we introduce the conventional ETAS model and its ETASI extension and define the anisotropic, locally restricted spatial kernel. This section includes a description of the estimation procedures for strikes and rupture positions and the spatial integral over anisotropic kernels. Next, the *Application* section explains the three forecasting experiments, introducing the data and time-space windows for the parameter estimation and forward simulations. Finally, we interpret and discuss our forecasting results and draw our conclusions. Source codes for model estimation and simulation are freely available in a Github repository (see Data and resources).

2 Methods

The ETAS model, first introduced by Ogata (1988, 1998), is a branching-tree type model which describes clustered earthquake occurrences by consecutive triggering evolving over multiple parent-child generations (i.e. allowing secondary aftershocks). The triggered seismicity is overlaying a time-invariant background process.

In this section, we will first introduce the conventional, isotropic ETAS model approach. Next, we will extend the model to obtain a time-space version of the ETASI model suggested by Hainzl (2021), which involves STAI into parameter estimation. Mostly, notations are consistent with Hainzl (2021). We will then define the anisotropic generalization of the spatial kernel, which is compatible with both the ETAS and ETASI model, and introduce the local restriction of the kernel. Finally, we explain the fitting algorithm for the strike angle and rupture position of anisotropic trigger events and the methods for spatial integral estimation.

2.1 ETAS-model

In the conventional ETAS model approach, the occurrence rate of an earthquake with *magnitude m*, occurring at *time t* and at *location* (x, y) is modeled by an inhomogeneous Poisson process with a time-space-magnitude dependent intensity function

$$\lambda(t,x,y,m) = f_0(m)\,R_0(t,x,y)$$

where

$$f_0(m) = \beta\,e^{-\beta(m-M_c)} \tag{3}$$

is the "true" probability density function (pdf) of the frequency-magnitude distribution (FMD) with Gutenberg-Richter parameter $b = \beta/ln(10)$ (Gutenberg and Richter 1944), and

$$R_0(t,x,y) = \mu\,u(x,y) + \sum_{i:t_i<t} k_{A,\alpha}(m_i)\,g_{c,p}(t-t_i)\,h_{D,\gamma,q}(r_i(x,y),m_i,l_i) \tag{4}$$

is the "true" occurrence rate of events with magnitude $m \geq M_c$, at time t and at location (x, y). The "true" event rate is modeled by a superposition of the time-invariant *seismic background rate* $\mu\,u(x,y)$ with parameter $\mu > 0$ and a sum of the trigger rate contributions of all events i that occurred prior to current time t. $k_{A,\alpha}(m_i)$ and $g_{c,p}(t-t_i)$ denote the aftershock productivity and Omori-Utsu Law decay functions as defined in Eqs. (1) and (2), respectively. $h_{D,\gamma,q}(r_i(x,y),m_i,l_i)$ models distribution of aftershock locations triggered by event i, with parameters D,γ and

q. The precise inputs and shape of the spatial kernel are discussed later.

The term "true" means that the (physical) relationships are expected to be observed with perfect earthquake records. The main assumption of the conventional ETAS model is that STAI does not significantly distort the "true" magnitude distribution and the "true" event rates.

2.2 ETASI model

2.2.1 Rate-dependent iIncompleteness

The concept of rate-dependent earthquake record incompleteness assumes that the "true" relationships underlying $f_0(m)$ and $R_0(t,x,y)$ are not accurately identifiable in available earthquake catalogs because especially events with small magnitudes are detected with lower probability in periods of high seismic activity. In these periods, the detection ability is limited typically due to overlapping seismic waves (Hainzl 2016a, 2021).

Fitting the "true" relationships to incomplete data records may therefore lead to significantly biased parameter estimates (Hainzl 2016a, b; Page et al. 2016; Seif et al. 2017; Hainzl 2021).

In recent years, there has been growing research interest in how to account for short-term incomplete datasets. For instance, Zhuang et al. (2017) developed a replenishment algorithm to fill up likely incomplete time intervals by simulated events, in order to obtain artificially complete pseudo-records. Other authors, particularly mentionable Omi et al. (2013, 2014), Lippiello et al. (2016), de Arcangelis et al. (2018), Mizrahi et al. (2021) and Hainzl (2021), tried to incorporate STAI directly into the ETAS model fit. A rather simple workaround approach is to remove likely incomplete time periods from the fitted time interval using empirical completeness functions, such as performed in Hainzl et al. (2013) and Grimm et al. (2021). A comprehensive discussion and comparison of various ETASI models is not in the scope of this article. The choice for the ETASI model proposed by Hainzl (2021) was made for rather practical reasons, mainly because of its compatibility with existing code.

2.2.2 Model formulation

The working assumption of the ETASI model described here is that an earthquake with magnitude m and occurring at time t can only be detected by the operating seismic network if no event of equal or larger magnitude occurred within the blind time $[t-T_b,t]$, where T_b is typically in the range of some seconds to few minutes (Hainzl 2021). Similar assumptions have formerly been formulated by Lippiello et al. (2016), de Arcangelis et al. (2018) and Hainzl (2016a).

Let $N_0(t)$ be the expected number of events occurring within the entire spatial window S during blind time $[t-T_b,t]$,

$$N_0(t) = \int_{t-T_b}^{t} \iint_S R_0(t,x,y)dx\,dy\,dt \approx T_b \iint_S R_0(t,x,y)\,dx\,dy,$$

where the approximation holds under the assumption that event rates are approximately constant during the blind time (Hainzl 2021). According to the "true" FMD (Eq. 3), each of the $N_0(t)$ events has a probability of $e^{-\beta(m-M_c)}$ to exceed magnitude m. We define the detection probability $p_d(m,t)$ of an earthquake at time t with magnitude m as the probability that no equal or larger event occurred during blind time T_b, i.e.

$$p_d(m,t) = e^{-N_0(t)\,e^{-\beta(m-M_c)}}.$$

Following the derivations in Hainzl (2016b, 2021), we obtain the "apparent", incompleteness-biased FMD

$$f(m,t) := f_0(m)\,N_0(t)\,\frac{p_d(m,t)}{1-e^{-N_0(t)}}$$

and the "apparent" event rate

$$R(t,x,y) := \frac{R_0(t,x,y)}{N_0(t)}\left(1-e^{-N_0(t)}\right).$$

The term "apparent" signalizes that the functions f and R do not represent the "true", but the observable relationships that are possibly distorted by rate-dependent record incompleteness. In periods of high seismic activity, the "apparent" FMD exhibits a larger relative frequency of strong events (because they are more likely to be detected) and an event rate lowered by detection capacity. We obtain the ETASI intensity function

$$\lambda(t,x,y,m) = f(m,t)\,R(t,x,y) = f_0(m)\,R_0(t,x,y)\,p_d(m,t)$$

The two underlying, simplifying assumptions in the ETASI model are that (1) the blind time T_b is magnitude-independent, which Hainzl (2021) justifies by typically shorter source durations than travel times of coda waves, and (2) that the seismic network is equally occupied for blind time T_b by any event in the entire investigated spatial window. The second assumption is reasonable for a small spatial window, e.g. when analyzing an isolated sequence. When fitting the ETASI model over a larger region, it needs to be checked that relevant clusters do not evolve at the same time but at distinct locations as they would be assumed to simultaneously occupy the entire seismic network. A reasonable approach to prevent undesired biases is to choose a larger cut-off magnitude.

2.2.3 Log-likelihood optimization

The parameter vector $\theta = \{\mu, A, \alpha, c, p, D, \gamma, q, \beta, T_b\}$ of the ETASI model is estimated by maximizing its log-likelihood function $LL = LL_1 - LL_2$ with

$$LL_1 = \sum_{events\ j} ln\big(f_0(m_j)\, R_0(t_j, x_j, y_j)\, p_d(m_j, t)\big),$$

$$LL_2 = \int_{M_c}^{\infty}\int_{T_1}^{T_2}\iint_S \lambda(t,x,y,m)\,dx\,dy\,dt\,dm \approx \frac{T_2 - T_1}{T_b} - \frac{1}{T_b}\int_{T_1}^{T_2} e^{-T_b \iint_S R_0(t,x,y)\,dx\,dy}\,dt \tag{5}$$

where the sum in LL_1 goes over all *target events* in the magnitude-time-space window $[M_c, \infty) \times [T_1, T_2] \times S$ and LL_2 integrates over this model space. In our study we optimized the parameter vector θ using the gradient-based Davidson-Fletcher-Powell algorithm (Ogata 1998; Zhuang et al. 2002; Jalilian 2019).

2.3 Generalized anisotropic spatial kernel

2.3.1 Conventional isotropic kernel

The *spatial kernel* $h_{D,\gamma,q}(r_i, m_i, l_i)$ in Eq. (4) models the 2D-distribution of aftershocks locations. In conventional ETAS model approaches, the triggering event is assumed to be a point source, distributing its offsprings isotropically around its epicenter. A classical definition of an isotropic kernel (see Ogata 1998; Grimm et al. 2021; Jalilian 2019) is

$$h^{iso}_{D,\gamma,q}(r_i(x,y), m_i) := \frac{q-1}{D\,\exp(\gamma(m_i - M_c))}\left(1 + \frac{\pi\, r_i(x,y)^2}{D\,\exp(\gamma(m_i - M_c))}\right)^{-q}$$

where $r_i(x,y)$ denotes the point-to-point distance between a potential aftershock location (x, y) and the coordinates (x_i, y_i) of the triggering event i, and m_i is the magnitude of the event i. The kernel is constrained by the parameters D and γ that control the magnitude-dependent width of the kernel, and parameter q that describes the exponential decay of the function with growing spatial distance.

2.3.2 Anisotropic generalization

Here we use the anisotropic generalization of the spatial kernel that was first introduced by Grimm et al. (2021),

$$h_{D,\gamma,q}(r_i(x,y), m_i, l_i) := \frac{q-1}{D\,\exp(\gamma(m_i - M_c))}\left(1 + \frac{2\,l_i\, r_i(x,y) + \pi\, r_i(x,y)^2}{D\,\exp(\gamma(m_i - M_c))}\right)^{-q}.$$

In this spatial model, the distance term $r_i(x,y)$ denotes the point-to-line distance between the potential aftershock location (x, y) and the estimated rupture segment of triggering event i with length l_i. That is, the kernel assigns constant density along the rupture line segment, with a power-law decay to the sides. Note that

$$h_{D,\gamma,q}(r_i(x,y), m_i, 0) = h^{iso}_{D,\gamma,q}(r_i(x,y), m_i),$$

i.e. the anisotropic kernel is a generalization and collapses to the isotropic model if the triggering location is assumed to be a point source with rupture extension $l_i = 0$. Therefore, the generalized spatial model can be used for mixing approaches of both kernels, e.g. applying anisotropy to events i with magnitudes $m_i \geq M_{aniso}$:

$$l_i = \begin{cases} 0, & \text{for } m_i < M_{aniso}, \quad \text{(isotropic trigger)} \\ 10^{-2.57+0.62 m_i}, & \text{for } m_i \geq M_{aniso}, \quad \text{(anisotropic trigger)} \end{cases} \tag{6}$$

The scaling relationship for anisotropic events is taken from the estimate of subsurface rupture lengths for strike-slip faulting events, provided in Wells and Coppersmith (1994). Alternative relationships can be applied, too, but are not expected to impact results.

2.3.3 Local spatial restriction

Both the conventional isotropic and the generalized anisotropic kernels are defined in terms of a probability density function (pdf) over infinite space. Realistically, however, small earthquakes should exert only a locally restricted trigger influence. Grimm et al. (2021) showed that an infinite spatial extent may lead to an underestimation of the aftershock productivity parameter α because it overestimates the triggering power of smaller events at the cost of the larger events. A manual analysis of the spatial aftershock patterns of the six Californian mainshocks named in the introduction shows that the cloud of potential aftershocks typically lies within one estimated rupture length (by Wells and Coppersmith 1994) from the epicenter. In case of an anisotropic shape of the kernel, the area of half a rupture length around the extended rupture segment seems sufficient. According to this observation, we choose a spatial restriction R_i for event i according to

$$R_i := \begin{cases} 10^{-2.57+0.62 m_i}, & \text{for } m_i < M_{aniso}, \quad \text{(isotropic trigger)} \\ 0.5 \cdot 10^{-2.57+0.62 m_i}, & \text{for } m_i \geq M_{aniso}, \quad \text{(anisotropic trigger)} \end{cases} \tag{7}$$

where again we use the strike-slip faulting subsurface rupture length scaling from Wells and Coppersmith (1994).

In other words, the spatial kernel for event i is only defined in the restricted area

$$S_i(R_i) := \{(x,y) \in \mathbb{R}^2 | r_i(x,y) \leq R_i\}$$

and set to 0 outside of it. Note that the restricted area $S_i(R_i)$ can assume isotropic and anisotropic shapes, depending on the point-to-point or point-to-line definition of the distance term $r_i(x,y)$. In order to retain the property of a pdf, we need to rescale the kernel within the restricted area by its analytical integral

$$\begin{aligned} H_{D,\gamma,q}(R_i,m_i,l_i) :&= \iint_{S_i(R_i)} h_{D,\gamma,q}(r_i(x,y),m_i,l_i)\,dx\,dy \\ &= 1 - \left(1 + \frac{2\,l_i\,R_i\, +\, \pi R_i^2}{D\,\exp(\gamma(m_i - M_c))}\right)^{1-q}. \end{aligned} \tag{8}$$

The integral term holds true for both isotropic ($l_i = 0$) and anisotropic triggers ($l_i > 0$). We obtain the generalized, restricted and anisotropic spatial kernel

$$h^{restr}_{D,\gamma,q}(r_i(x,y),m_i,l_i) = \begin{cases} \dfrac{h^{restr}_{D,\gamma,q}(r_i(x,y),m_i,l_i)}{H_{D,\gamma,q}(R_i,m_i,l_i)}, & \text{if } r_i(x,y) \leq R_i, \\ 0, & \text{if } r_i(x,y) > R_i. \end{cases} \tag{9}$$

2.4 Estimation of strike and epicenter location

The restricted, anisotropic spatial kernel in Eq. (9) requires a strike angle to define the orientation of the extended rupture for anisotropic triggers with $l_i > 0$. In retrospect, the strike angle could be taken from one of the numerous publications about the Ridgecrest sequence or from focal mechanism datasets such as the Global Moment Tensor Catalog, the ISC-GEM Global Instrumental Earthquake Catalog or from local datasets of the Southern California Earthquake Data Center (SCEDC). In order to perform a realistic forecasting test case, however, we should build upon instantaneous strike estimates such as from local agencies (e.g. the United States Geological Survey), which are typically available within several minutes to hours.

Here, we used the strike estimation algorithm proposed by Grimm et al. (2021), that optimally fits the rupture segment through the cloud of early aftershock locations by maximizing the contributed spatial rate under initial spatial parameter guesses. To be more precise, we ran through possible strikes in 1° steps (i.e. $\{1°, ..., 180°\}$ where we can neglect all strikes larger than 180° because we do not account for dip direction in our model). We also moved the rupture along each strike angle in order to test different positions of the rupture segment relative to the fix epicenter. Here, we go through possible relative positions in 0.01-steps (i.e. $\{0, 0.01, 0.02, ..., 0.99, 1\}$), where 0 and 1 means that one of the ends of the rupture segment coincides with the epicenter, and 0.5 denotes the situation where the rupture embeds the epicenter directly in its center. Among all combinations, we searched the orientation and rupture position that maximizes the forward trigger contribution of the anisotropic event i to subsequent events j within a fixed duration $\Delta t = 1$ *hour*, i.e. with $t_i < t_j < t_i + \Delta t$. The forward trigger contribution of event i is computed as

$$\sum_{t:t_i<t_j<t_i+\Delta t} h^{rest}_{D,\gamma,q}(r_i(x_j,y_j),m_i,l_i). \tag{10}$$

In order to avoid that the rupture orientation and position is dominated by single events that occurred very close to the segment candidate, we applied a minimum distance of 0.2 kilometers.

Here, we use the initial spatial parameters $D = 0.0025$, $\gamma = 1.78$ and $q = 1.71$ derived from the results of an isotropic ETAS model for a long-term California dataset, locally restricted to $R = 2.5$ rupture lengths, by Grimm et al. (2021). Tests have shown that modified initial parameters changed the level of the sum of forward rate contributions, but led to similar strike and epicenter location estimates. We also tested multiple durations Δt up to 30 hours and found that the estimation procedure provided very similar estimates for strike and rupture position. It shows that the rupture orientation and position can be appropriately identified soon after the trigger event occurred.

In the *Application* section we present the strike and rupture position estimation for the example of the M6.4 and M7.1 Ridgecrest events.

2.5 Estimation of spatial integral

The computation of the log-likelihood function in Eq. (5) requires the estimation of the spatial integral of R_0 and therefore $h^{restr}_{D,\gamma,q}$.

In the isotropic case, the integral can be estimated semi-analytically by the radial partitioning method suggested by Ogata (1998) and applied in the *R* package *ETAS* (Jalilian 2019). It uses the property, that the isotropic spatial kernel can be integrated analytically over circular areas $S_i(R)$, according to Eq. (8). As Fig. 2a illustrates, the polygon defining the spatial window S is partitioned into a fine grid, with two neighboring boundary grid points having approximately equal distances $\tilde{R}$ to the point source coordinate of event i. The integral of $h^{restr}_{D,\gamma,q}$ over each of these thin segments of a circle can then be approximated by the analytical full integral, weighted by the ratio of the circle segment $\phi/360°$, where ϕ is the angle enclosed by the circle segment (Jalilian 2019; Ogata 1998).

Similarly, an anisotropic spatial kernel can be integrated analytically over an anisotropic region $S_i(\tilde{R})$ with maximum distance $\tilde{R}$ to the extended rupture. Due to the non-circular shape of region $S_i(R)$ for anisotropic triggers, radial partitioning can be only performed at both ends of the rupture segment. As Fig. 2b illustrates, in a similar way we use "bin partitioning" in the space orthogonal to the rupture. Unfortunately, in the anisotropic case, the weights $\phi/360°$ of the circle segments at both ends of the rupture only relate to the part of the integral that is estimated by radial partitioning. Similarly, the weight of a bin of size Δl is $\frac{\Delta l}{2 l_i}$ relative to only the orthogonal space on both sides of the rupture segment. In each iteration of the parameter estimation, the shares of the radial and the orthogonal integral parts change and need to be determined numerically before each iteration. This comes at the computational cost of approximately doubled run-time, given that only the minority of strong earthquakes with magnitude $M \geq M_{aniso}$ are modelled anisotropically.

3 Application

We carry out three forecasting experiments to check the quality of the previously defined models in predicting the observed Ridgecrest M6.4 and M7.1 sequences. Each forecasting experiment consists of three main steps, represented as blue boxes in Fig. 3:

- **Parameter Estimation:** Estimate model parameters for a specified event sub-set of southern Californian earthquake data
- **Forward Simulation:** Use the fitted model parameters to conduct 10,000 forward simulations of the Ridgecrest M6.4 or M7.1 sequence, respectively.
- **Evaluation:** Analyze simulated sequences and compare to the observation.

In the following, we first introduce the basic earthquake event set for California underlying this study, and define the time-space windows used to obtain the event sub-sets applied for parameter estimation. Next, we describe the three forecasting experiments, rigorously defining the magnitude-time-space windows applied for parameter estimation and forward simulations. Each forecasting experiment is repeated for five versions of the models introduced in the *Methods* section, which are defined in detail. Finally, we specify the forward simulation process and attributes and measures to assess the quality of the forecasts. Here, we also give an example of the estimation of strike angles and rupture positions for the Ridgecrest M6.4 and M7.1 events.

3.1 Data

As our basic event set, we downloaded the earthquake catalog from the Southern California Earthquake Data Center (SCEDC, Hauksson et al. 2012). The entire dataset comprises 699,175 event occurrences between January 1, 1981, and December 31, 2019. Because magnitudes seem to be clustered at values with one decimal, we round all

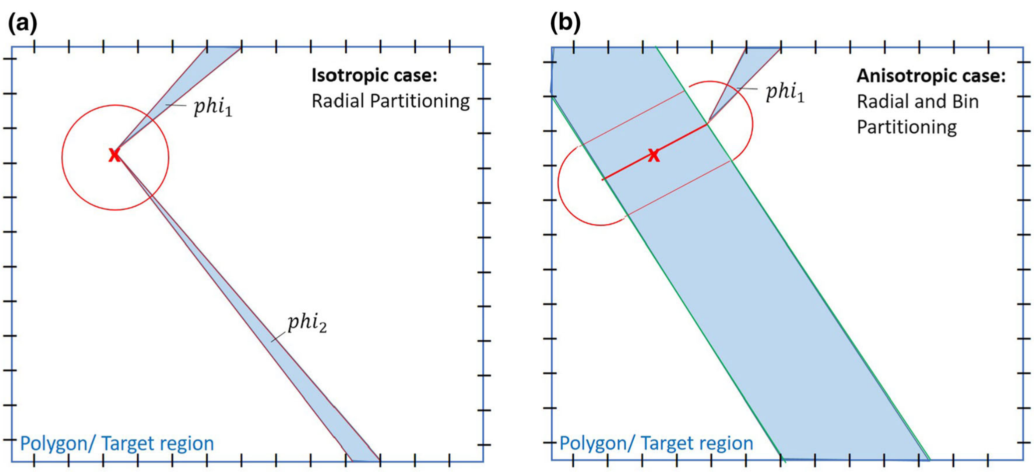


Fig. 2 Visualization of the spatial integral estimation needed for computing the log-likelihood function (Eq. 5) for **a** isotropic triggers and **b** anisotropic triggers. The box represents the boundary of the spatial target region (polygon), gridded into small intervals. Red crosses symbolize the epicenters of the triggering events. In **a**, the red circle around the event represents the contour lines of an isotropic spatial kernel and the shaded segments illustrate the radial partitioning method. In (b), the red box and semi-circles symbolize the contour lines of the anisotropic spatial kernel, and the shaded segments illustrate the radial and bin partitioning method

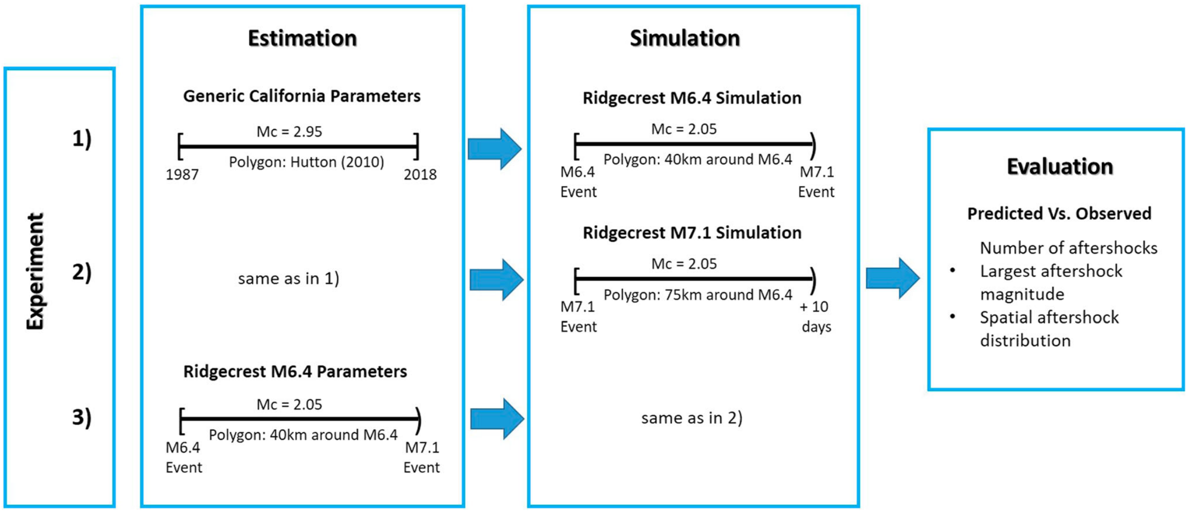


Fig. 3 Summary of the forecasting experiments (from left to right): The five model versions, listed in Table 1, are fitted to a long-term California event sub-set (*Experiments 1* and *2*) and to the local M6.4 Ridgecrest sequence (*Experiment 3*). The estimated parameters are applied to forward simulations of the Ridgecrest M6.4 sequence (*Experiment 1*) and the Ridgecrest M7.1 sequence (*Experiments 2* and *3*). The predicted sequences are compared to the observed ones with respect to three attributes, further described in the *Attributes and Measures* section

magnitudes to one decimal and use the cut-off magnitude $M_c = 2.05$ (Hutton et al. 2010; Hainzl 2021). We remove events at depths larger than 40 km to avoid completeness issues.

3.2 Forecasting experiments

Here, we describe in detail the design of the forecasting experiments, summarized in Fig. 3.

3.2.1 Experiment 1

We estimate generic, long-term California model parameters within the hexagonal polygon of southern California defined in Hutton et al. (2010). In order to mitigate computational costs, we restrict the time window to the period between January 1, 1987, and December 31, 2018, including the five prominent earthquake sequences before Ridgecrest as named in the *Introduction* section, and choose the larger cut-off magnitude $M_c = 2.95$. The cut-off magnitude is a trade-off between too large and too small event record sizes that ensures reasonable run-time and statistical robustness of parameter estimates. Additionally, it avoids potentially biased estimates of the blind time parameter T_b in large spatial areas due to simultaneous clustering. The magnitude-time-space window contains 7,215 fitted *target* events. We account for external triggering impact by including *complementary* events that occurred after January 1, 1986, and in the surrounding of 0.5° geographical degrees of the fitted area.

The estimated models are then applied to forecast the Ridgecrest M6.4 foreshock sequence above cut-off

Table 1 Overview of the model variants tested in this paper

Name	Model version	M_{aniso}	R_i (isotropic triggers)	R_i (anisotropic triggers)
ETAS conventional	ETAS	-	∞	-
ETAS iso-r	ETAS	-	$1\,RL_i$	-
ETAS aniso-r	ETAS	6.0	$1\,RL_i$	$0.5\,RL_i$
ETASI iso-r	ETASI	-	$1\,RL_i$	-
ETASI aniso-r	ETASI	6.0	$1\,RL_i$	$0.5\,RL_i$

Non applicable cases are filled with "-". Spatial restrictions R_i of event i are denoted in terms of the estimate rupture length (RL_i)

magnitude $M_c = 2.05$, within a circular polygon with radius 40 km centered in the M6.4 event location. The simulated time window starts in the moment of the M6.4 event (July 4, 2019) and ends at the M7.1 mainshock event time (July 6, 2019), thus it has a duration of approximately 34 hours. We initialize triggering seismicity by the event history from June 1, 2019.

3.2.2 Experiment 2

In the second experiment, we use the same set of generic, long-term California parameters, but apply it in a forecast of the Ridgecrest M7.1 mainshock sequence above cut-off magnitude $M_c = 2.95$, starting at the M7.1 event time for a duration of ten days. The spatial simulation window is defined by a disk with radius of 75 km, centered in the M7.1 event location. Again, we initialize triggering seismicity by the event history from June 1, 2019, here until the M7.1 event time.

3.2.3 Experiment 3

In the third experiment, we simulate Ridgecrest M7.1 sequences with the same settings as for *Experiment 2*, but based on parameter estimates fitted over the immediately preceding M6.4 foreshock sequence. For the parameter estimation, we use the same magnitude-time-space target window as for the M6.4 sequence simulations in *Experiment 1*. We account for external triggering by including complementary events that occurred after June 1, 2019, and within a disk with increased radius of 50 km.

3.3 Fitted models

Each forecasting experiment is carried out for five different versions of the model introduced in the *Methods* section, summarized in Table 1. The "ETAS conventional" model serves as our benchmark and uses the most standard set-up of an ETAS model (e.g. Ogata 1998; Zhuang et al. 2002; Jalilian 2019). It applies an isotropic spatial kernel with infinite spatial extent to all triggers. The "ETAS iso-r" model applies an isotropic kernel, but restricts the spatial extent to one rupture length for all events, according to Eq. (7). In the "ETAS aniso-r" model, all events with magnitudes $m_i \geq M_{aniso} = 6.0$ are modeled as anisotropic trigger source with a spatial restriction to half a rupture length (Eqs. 6 and 7). The other events are modeled as isotropic triggers, restricted to one rupture length. The "ETASI iso-r" and "ETASI aniso-r" models combine the spatial kernel settings of the latter models with the ETASI approach accounting for STAI.

3.4 Simulation process

For each forecasting experiment and model version, we carry out 10,000 realizations of synthetic sequences to obtain statistically stable results. At the beginning of each simulation, we distribute the Poisson-sampled number of background events, scaled by the size of the spatial area, uniformly over space and time. The assumption of an uniform spatio-temporal background event distribution appears plausible within the short and small space-time simulation windows.

Next, we sample the numbers of offsprings for the initiating event history and the background events. The number of offsprings, depending on trigger magnitude m, is drawn from a Poisson distribution with expected value

$$N(m) = k(m)\,\frac{1}{1-p}\left((T+c)^{1-p} - c^{1-p}\right). \tag{11}$$

where $k(m)$ is the aftershock productivity function in Eq. (2) and the latter term is the integral from $t = 0$ to a maximum trigger duration $t = T$ (in days) over the Omori-Utsu function in Eq. (1). We need to rescale the aftershock productivity to obtain the expected number of offsprings within T days, because the Omori-Utsu law is not normalized (no pdf) and, therefore, typically does not integrate to one. Thus, it interacts with the scaling parameter A of the productivity function.

Each triggered event is then assigned an event time and location according to inversion sampling from the (rescaled) Omori-Utsu law and the spatial kernel. The magnitude is sampled by the inversion method from the estimated FMD in Eq. (3), applying a maximum magnitude of 7.5 for California. The aftershock sampling routine is repeated for every newly triggered event in the simulated time-space window until no more events are sampled.

In order to make fair comparisons of simulated sequences with the observed ones, we need to consider the implications of STAI in the forecasts. The ETASI models account for incomplete records in the parameter estimation and therefore forecast the "true", i.e. complete, aftershock sequence. According to its definition of event detectability, we would need to delete all events that occurred within the blind time T_b of an earlier event with larger magnitude.

For the sake of transparency and consistency with the observations, we used an alternative approach and manually fitted empirical magnitude completeness functions

$$M_c(t) = \begin{cases} -1.4\,log_{10}(t) + M - M_c - 4.75, & \text{(Ridgecrest M6.4)}, \\ -0.99\,log_{10}(t) + M - M_c - 3.75, & \text{(Ridgecrest M7.1)}. \end{cases} \tag{12}$$

to the logarithmic event time-magnitude scatter data of the observed Ridgecrest M6.4 and Ridgecrest M7.1 sequences in Fig. 1c and d.

In the forecasts generated by the ETASI iso-r and aniso-r models, we delete all events that fall in the supposedly undetected range below the line of the simulated sequence. In contrast, ETAS models estimate STAI-biased aftershock productivities and therefore lead to predictions of the incomplete, rather than the "true" size of the sequence. Therefore, in forecasts generated by these models we do not delete events.

3.4.1 Attributes and measures

For each model version and experiment, we want to assess the quality of the forecasts with respect to three attributes, in comparison with the observed sequence evaluated over the same magnitude-time-space window.

We compute the predicted cumulative distribution function (cdf) of the number of aftershocks and the predicted pdf of the largest aftershock magnitude out of the 10,000 forecasted sequences. As a quantitative measure of the fit, we determine the exceedance probability that the predicted distribution would forecast a larger or the observed value. Extreme exceedance probabilities, either close to 0 or 1, indicate an inadequate prediction of the attribute.

To test the spatial distribution of aftershock locations, we define a 1km $\times$ 1km spatial grid over the spatial simulation window of the forecasting experiment and count the number of aftershocks in each simulation run, that occurred closest to the respective grid points. We determine the spatial distribution D_{ij} of the i-th simulation run by dividing the number of events occurred at each grid point j, N_{ij}, by the number of events in the i-th simulation run, N_i, i.e.

$$D_{ij} = N_{ij}/N_i.$$

By repeating the same procedure for each simulation run, we obtain 10,000 simulated spatial distributions D_{ij} for each model version. Finally, we average the individual simulated distributions to obtain the predicted probability P_j that an event occurs at grid point j.

The more events of the observed sequence have occurred at grid points with high predicted probabilities P_j, the better is the forecast. Therefore, we quantify the goodness of the spatial fit with the likelihood $L_{space} = \prod_{grid\ points\ j} P_j^{N_j^{obs}}$ where N_j^{obs} is the number of observed events at grid point j with corresponding log-likelihood

$$LL_{space} = \sum_{grid\ points\ j} N_j^{obs}\, ln(P_j).$$

We compute the information gain of the models' spatial predictions relative to the ETAS conventional model by

$$IG = \frac{LL_{space} - LL_{space}^0}{N_{obs}}$$

where LL_{space}^0 is the benchmark result for the ETAS conventional model (Hainzl 2021; Rhoades et al. 2014).

3.4.2 Strike and rupture position estimates

For anisotropic models, both the parameter estimation and the forward simulations of the Ridgecrest M6.4 and M7.1 sequences require estimates of the strike angle and rupture position of all events with magnitude $M > 6.0$.

Figure 4a demonstrates the methodology, described in the *Methods* section, for the Ridgecrest M6.4 foreshock. The forward trigger rate contribution (y axis) from Eq. (10) is plotted against the strike sample (x axis) and the sample of relative rupture positions (red lines). The curves clearly show a bi-modal shape, with peaks at strikes 34° and 132° and relative rupture positions 0.76 and 0.77. Fig. 4c visualizes the optimized rupture orientation and position as a fit through the cloud of potential aftershocks within the first hour (red) or within 30 hours (yellow). It confirms the earlier mentioned particularity of two almost orthogonally ruptured faults. The strike 34° rupture segment does not perfectly fit the aftershock alignment, as segment must go through the fixed M6.4 epicenter location which seems to be slightly off the ruptured fault. Apparently, later aftershocks have a very similar spatial distribution as the events occurred within the first hour. For larger Δt, the M6.4 strike estimates would vary by only 1° or 2°, respectively.

Figure 4b shows the analogous analysis for the M7.1 Ridgecrest mainshock. Here, the maximizing properties are strike 142° and a relatively central rupture position 0.55. The M7.1 event ruptured a single fault, resulting in an unimodal shape of the forward trigger contribution curves.

4 Results and discussion

In this section, we analyze and discuss the results of the three forecasting experiments, summarized in Fig. 3. We use the attributes and measures presented in the *Application* section to evaluate the quality of the forecasts, compared to the observed sequences. The model parameter estimation results of both the generic California and the Ridgecrest M6.4 sequence parameter fits are listed in Table 2 and will help us to understand and explain features in the forecasts. After a rigorous discussion of the forecasting results, we will mention some sensitivity tests that we applied to check the robustness of our findings.

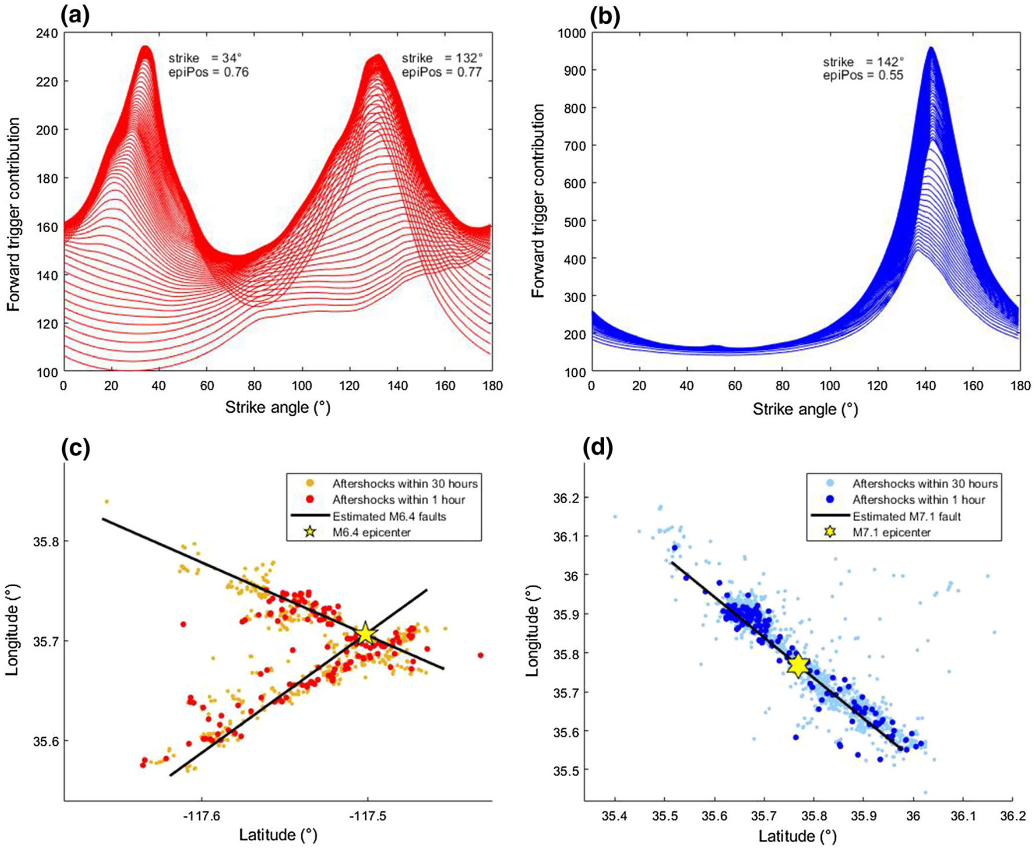


Fig. 4 Strike and relative rupture position optimization using initial ETAS parameter guesses $D = 0.0025, \gamma = 1.78, q = 1.71$. **a**, **b**: Sum of forward trigger rate contributions to events within one hour against strike sample (x axis) and relative rupture position sample (curves) for **a** the M6.4 Ridgecrest foreshock and **b** the M7.1 Ridgecrest mainshock. Text boxes show strike and relative rupture position estimates at the curve maxima. **c**, **d**: Fitted rupture segments through cloud of aftershocks after **c** the M6.4 Ridgecrest foreshock and **d** the M7.1 Ridgecrest mainshock. Darker red and blue points represent aftershocks within the first hour after the respective trigger event, yellow and lighter blue points represent aftershocks within the first 30 hours. Yellow pentagram symbolizes Mw6.4 foreshock, and yellow hexagram marks Mw7.1 mainshock. Thick black lines represent estimated rupture locations according to the strikes and relative rupture positions estimated in **a** and **b**.

4.1 Forecasting experiment 1

In the first forecasting experiment, we simulated the Ridgecrest M6.4 sequence, starting at the time of the M6.4 earthquake occurrence, based on generic parameters, fitted on a long-term and spacious Californian event set. The simulation period covers the 34 hours until (but non-including) the occurrence of the M7.1 mainshock.

4.1.1 Predicted aftershock productivity

Figure 5a shows the predicted cdfs of the number of aftershocks for each model, compared to the observed M6.4 sequence, which produced 633 events in the same time-space window. Evidently, the ETAS conventional model (with isotropic, unlimited spatial kernel) provides a very inappropriate estimate, as it does not reach the observed number in any of the 10,000 simulations. According to the ETAS iso-r and ETAS aniso-r models, the observed number of events would be a possible, but rather unlikely scenario, with approximately 2.4 and 3.7% probability to exceed the observed value. The ETASI models tend to only moderately (ETASI iso-r) or slightly (ETASI aniso-r) underestimate the observed number.

To explain this observation, we consider that the size of this relatively short sequence is predominantly influenced by the amount of direct aftershocks of the initial M6.4 trigger event. According to the model parameter estimates in Table 2 and Eq. (11), the M6.4 trigger event would only produce approximately 17 direct aftershocks in the ETAS conventional model, compared to 46 (ETAS iso-r), 49 (ETAS aniso-r), 66 (ETASI iso-r) and 74 (ETASI aniso-r)

Table 2 Overview of model results for generic (long-term) California and Ridgecrest M6.4 parameter estimation

Parameter		Generic California Estimates					Ridgecrest M6.4 Estimates				
		ETAS			ETASI		ETAS			ETASI	
		conv	iso-r	aniso-r	iso-r	aniso-r	conv	iso-r	aniso-r	iso-r	aniso-r
μ	$\frac{1}{days}$	0.16	0.21	0.21	0.21	0.21	0.11	0.30	0.29	0.18	0.30
A		0.027	0.012	0.011	0.010	0.009	0.052	0.024	0.022	0.022	0.019
α	$\frac{1}{mags}$	1.30	1.87	1.92	1.98	2.05	1.13	1.71	1.75	1.76	1.83
c	$\frac{1}{days}$	0.004	0.010	0.010	0.005	0.005	0.008	0.015	0.014	0.010	0.007
p		1.06	1.08	1.08	1.09	1.09	1.16	1.09	1.06	1.07	1.04
D	Km^2	0.085	0.037	0.110	0.037	0.107	0.135	0.085	0.469	0.080	0.399
γ	$\frac{1}{mag}$	1.60	1.86	2.09	1.88	2.10	1.15	1.43	1.55	1.44	1.57
q		1.51	1.03	2.14	1.07	2.20	1.93	1.73	8.98	1.72	8.79
T_b	sec			112.8	114.0				18.1	21.1	
b		0.98	0.98	0.98	1.01	1.01	0.72	0.72	0.72	0.77	0.79
LL		20,806	17,478	18,209	16,321	17,107	6524	6322	6433	6013	6131
ν_{branch}		0.73	0.60	0.59	0.61	0.60	1.38	1.76	1.89	1.54	1.52

in the other models. The larger the parameter α, the more direct aftershocks are expected for the M6.4 event.

As argued in the *Methods* section, the local restriction of the spatial kernels prevents a disproportionate triggering power of small events and in return increases the direct aftershock productivity of the stronger events, characterized by a considerable increase of the parameter α in the non-conventional models (Grimm et al. 2021). Besides, the application of the ETASI model accounts for missing aftershock records after strong trigger events and corrects for the biased, underestimated aftershock productivity, leading to an additional increase of α (Hainzl 2021). Finally, we note that the majority of the $M > 6$ mainshocks included in the estimation time window from 1987 until 2018, are characterized by anisotropic aftershock patterns. Consequently, more events are associated as direct aftershocks of the strong mainshocks when we estimate the parameters with the ETAS aniso-r or the ETASI aniso-r model.

4.1.2 Predicted largest aftershock magnitude

Figure 5b shows the predicted pdfs of the largest aftershock magnitude in the synthetic sequences, assuming that the Gutenberg-Richter distribution holds over the entire magnitude range up to the largest values. For each of the five models, a kernel density function was computed for the 10,000 largest magnitude samples. In all models, the observed M7.1 event would have been an extremely rare case, with exceedance probabilities $P(largest\ magnitude \geq 7.1) \leq 0.43\%$. Even the second largest, observed aftershock magnitude ($M = 5.4$) was not reached in approximately 75% of the simulations of the best model (ETASI aniso-r).

To interpret this result, think of the largest aftershock magnitude as the largest order statistic of the sample of simulated events in a simulation run. Then, the expected value of the sample maximum (i.e. the largest aftershock) increases if (1) the sequence size becomes larger or (2) if the magnitudes in the sample are distributed in a way that they favor high values.

The underestimations of the observed sequence size, shown in Fig. 5a and discussed earlier, cannot sufficiently explain the miss-match of the predicted largest aftershock magnitudes. Even the observed sample size (633 events) would produce a M7.1 event with a probability of less than 1%, given the generic California estimates for the FMD with $b = 0.98$ (ETAS models) or $b = 1.01$ (ETASI models; see Table 2). If $b = 1$, then each magnitude increment by 1 leads to a 10 times smaller probability of occurrence. Therefore, one M7.1 event is only obtained, on average, for a sequence with 100,000 aftershocks.

According to the results in Table 2, all models estimated significantly smaller values $b < 0.8$ for the observed Ridgecrest M6.4 sequence, which favors the occurrence of strong events. Note that the b estimates of the three ETAS models are biased, because they are fitted for the ”true” FMD using the evidently short-term incomplete M6.4 sequence event record (see Fig. 1c). The ETASI models account for the missing smaller magnitudes and therefore lead to corrected, larger b values.

If we would simulate the Ridgecrest M6.4 sequence using its own estimation results (note that this is not a valid forecasting experiment, but used for illustration purposes),

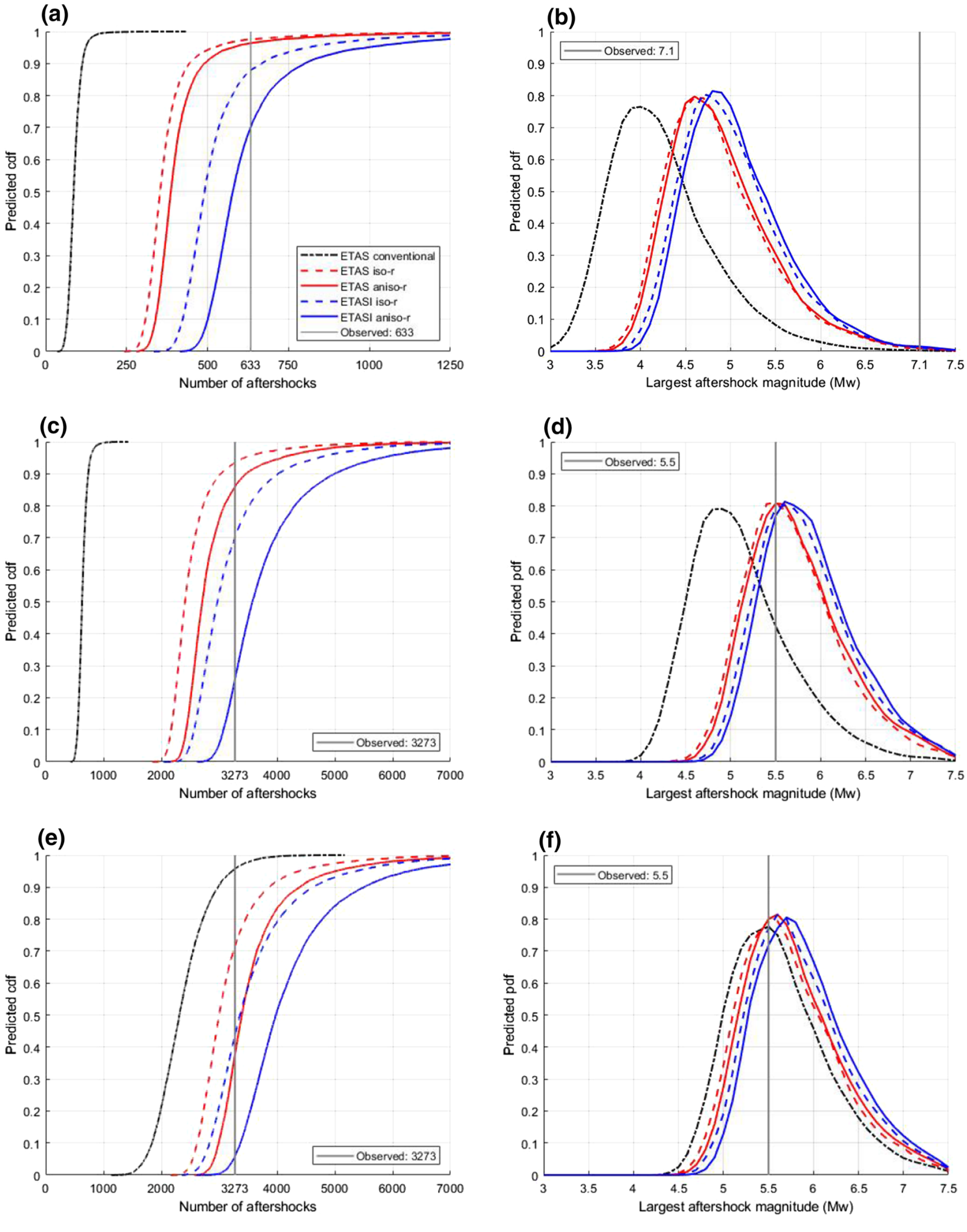


Fig. 5 Predicted cdfs of the number of aftershocks (**a**, **c**, **e**) and predicted pdfs of the largest aftershock magnitude (**b**, **d**, **f**) for *Experiment 1* (**a**, **b**), *Experiment 2* (**c**, **d**) and *Experiment 3* (**e**, **f**). Each predicted distribution is based on 10,000 simulated forecasts of the Ridgecrest M6.4 sequence (**a**, **b**) and the Ridgecrest M7.1 sequence (**c**–**f**), using the models indicated in the legend in the top left figure. Vertical gray lines show the value of the observed sequence

we would obtain an $M \geq 7.1$ event with 10.0% (ETAS conventional), 25.9% (ETAS iso-r), 53.7% (ETAS aniso-r), 15.6% (ETASI iso-r) and 25.3% (ETASI aniso-r) chance.

4.1.3 Criticality

The branching ratios ν_{branch}, i.e. the average number of direct aftershocks per event, clearly exceed 1 in each model (see Table 2). According to these models, an earthquake would trigger more than one direct aftershock on average, which would cause the sequence to be unstable, with exponentially increasing activity. The large branching ratios are mainly driven by the small *b* values, which substantially increase the occurrence probability of the more productive, strong earthquakes.

The instability of the M6.4 sequence could be interpreted as an indication of an imminent, strong mainshock. On the other hand, it is unlikely that the instability is based on a model error, e.g. due to a substantial misfit of the b-value due to few magnitude outliers. First, the FMD is estimated accounting for all earthquakes at equal weights, regardless of their magnitude. Therefore, the *b* value

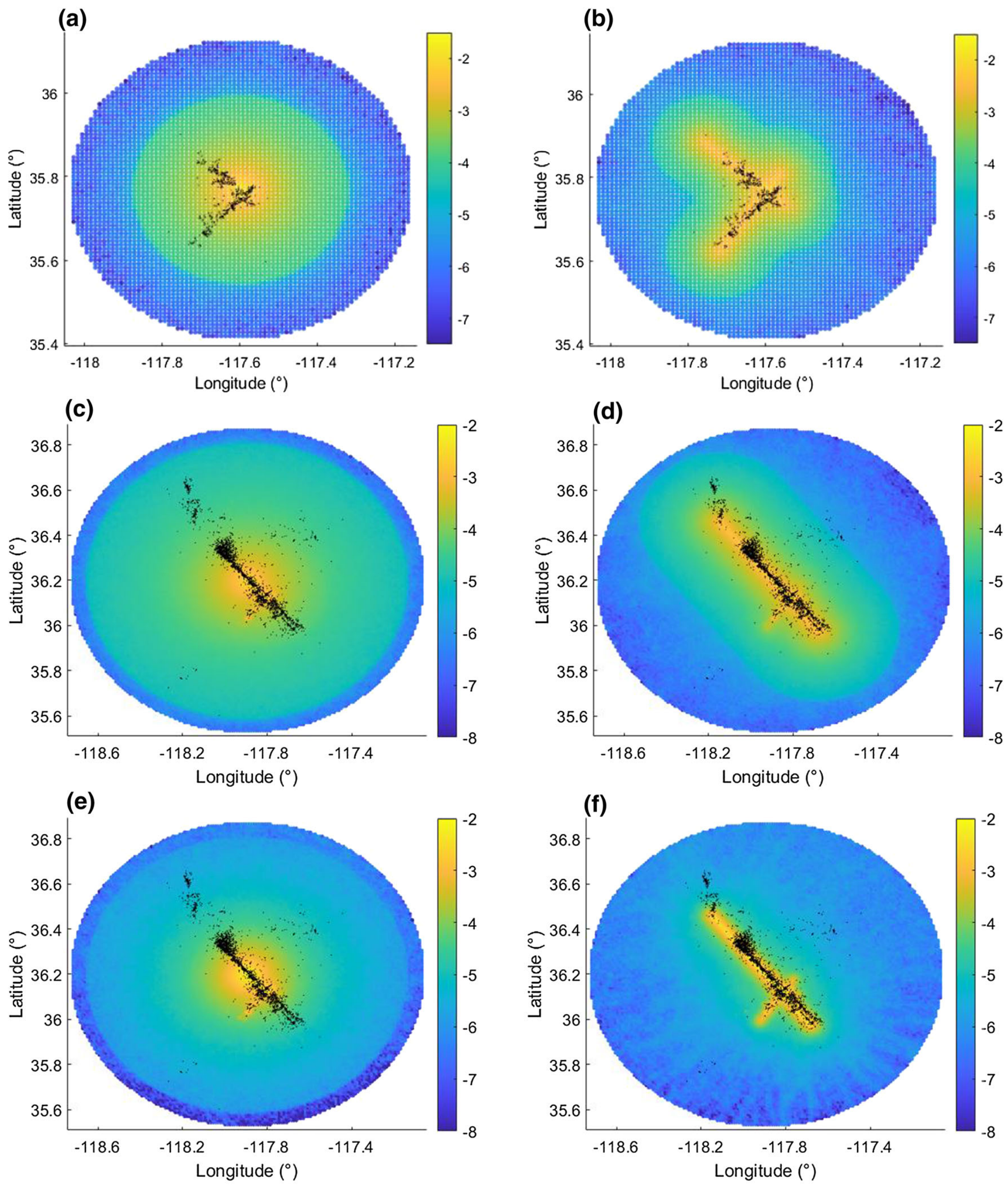


Fig. 6 Predicted spatial event distributions for *Experiment 1* (**a**, **b**), *Experiment 2* (**c**, **d**) and *Experiment 3* (**e**, **f**). Each predicted distribution is averaged over 10,000 simulated forecasts of the Ridgecrest M6.4 sequence (**a**, **b**) and the Ridgecrest M7.1 sequence (**c**–**f**), based on the ETASI iso-r model (**a**, **c**, **e**) and the ETASI aniso-r model (**b**, **d**, **f**) . The color bar indicates the predicted, logarithmic probability that an event occurs at the respective grid point

estimate is primarily controlled by the more numerous, small magnitudes. Secondly, the M7.1 event magnitude was *not* included in the b value estimation.

In summary, the generic California parameters are fitted to a long-term catalog mainly consisting of stable earthquake sequences and seismically quiet periods. Therefore, it cannot adequately predict the magnitude distribution of the M6.4 foreshock sequence, which is characterized by instability due to a particularly flat FMD.

4.1.4 Spatial distribution

Figure 6a and b show the predicted spatial event distributions, averaged over the 10,000 simulation runs and evaluated on the 1 km × 1 km grid described in the *Application* section, for the ETASI iso-r model (in (a)) and the ETASI aniso-r model (in (b)). We overlay the observed event locations to the logarithmic heat map of grid cell probabilities. At first glance, the anisotropic spatial forecast in (b) fits the observed, and clearly non-isotropical event distribution much better than the isotropic counterpart in (a).

In the isotropic model, all direct aftershocks are distributed point-symmetrically around the M6.4 trigger event. Subsequent secondary triggering then takes place around the new initiators. In the anisotropic model, the direct aftershocks are distributed around the fitted rupture segments of the two orthogonal faults (see Fig. 4). Subsequent trigger generations then spread isotropically (if $M < M_{aniso}$) or anisotropically (if $M \geq M_{aniso}$) around their new initiators. In both plots, we can see a pronounced boundary from green to blue color, indicating the spatial restriction to one rupture length (isotropic model) and half a rupture length (anisotropic model) around the trigger source, according to Eq. (7). Spatial grid cells outside of this boundary can only be activated by secondary triggering or background seismicity.

To quantify the quality of the spatial forecasts, we computed the information gains relative to the ETAS conventional model as described in the *Application* section. Figure 7c shows the results for *Experiment 1* in the left group of bars. Both anisotropic models lead to a pronounced improvement, which confirms the visual impression in Fig. 6a and b. The ETAS and ETASI iso-r models, which differ from the conventional approach in terms of the local spatial restriction, show a small information gain. As we can see in Fig. 6a, none of the observed events occurred outside of the spatial restriction. Therefore, the restriction leads to a slightly more pronounced accumulation of simulated event locations closer to the M6.4 trigger, which coincides better with the observation.

4.2 Forecasting experiment 2

In the second forecasting experiment, we simulated the Ridgecrest M7.1 sequence for a duration of 10 days based on the same generic California parameters as used for *Experiment 1*.

4.2.1 Predicted aftershock productivity

Figure 5c compares the number of aftershocks, predicted by the five models, to the observed number of 3,273 events. The forecasts show a very similar setup of curves as in *Experiment 1* (see Fig. 5a). The ETAS conventional model clearly underestimates the observed number of events. The ETAS iso-r and aniso-r models reach the observation in

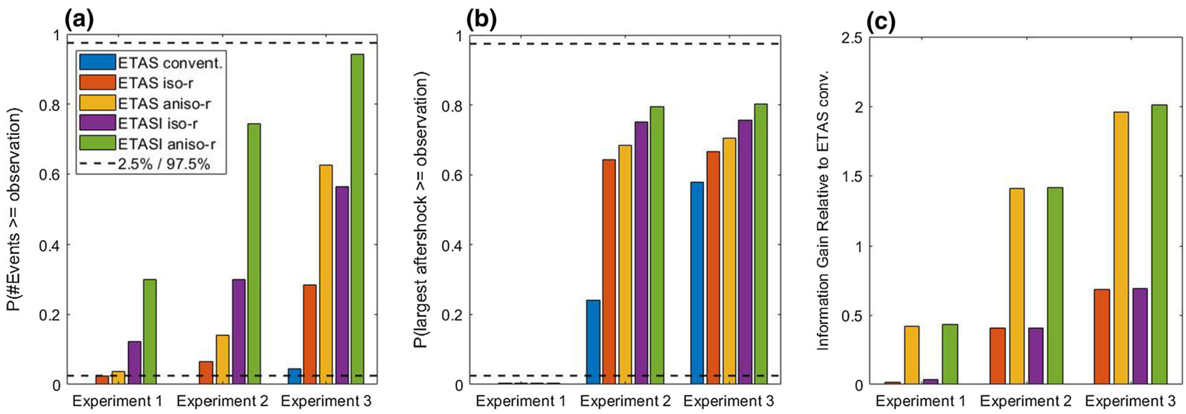


Fig. 7 Summary plots of forecasting results. Predicted probabilities per model that **a** the number of aftershocks exceeds the observation (633 for Ridgecrest M6.4; 3,273 for Ridgecrest M7.1) and **b** the largest aftershock magnitude exceeds the observation (7.1 for Ridgecrest M6.4; 5.5 for Ridgecrest M7.1). Dashed horizontal lines represent 2.5% and 97.5% quantiles. **c** Spatial information gains relative to the ETAS conventional model prediction for the same experiment. Legend in **a** holds for all plots

about 6.5 and 14.1% of the simulation runs. Again, the ETASI models provide the best approximations.

According to Eq. (11), the M7.1 trigger event would on average trigger only roughly 53 direct aftershocks in the ETAS conventional model, compared to 219 in the ETAS iso-r, 242 in the ETAS aniso-r, 328 in the ETASI iso-r and 387 in the ETASI aniso-r model. As explained in detail for *Experiment 1*, the reason is found in the parameter estimate for α.

4.2.2 Predicted largest aftershock magnitude

Figure 5d shows the predicted pdfs for the largest aftershock magnitude of the Ridgecrest M7.1 sequence. In contrast to *Experiment 1*, all but the conventional model provide very good forecasts, indicating that the generic, long-term California estimates of the FMD with $b \approx 1$ coincide well with the FMD of the Ridgecrest M7.1 sequence and the instability of the sequence ended with the occurrence of the M7.1 mainshock. The moderate underestimation by the ETAS conventional model can be explained by the underestimated sequence size, which substantially reduces the sample size of event magnitudes.

4.2.3 Spatial distribution

Figure 6c and d show the predicted spatial distributions of aftershock locations, again for the ETASI iso-r and aniso-r model. The visual impression, that the anisotropic model provides a substantially better forecast, is confirmed by the bar plot in Fig. 7c. The information gain by the anisotropic models is more pronounced for the Ridgecrest M7.1 sequence, because it has a longer rupture extension ($\sim 68km$ by Wells and Coppersmith 1994) than the M6.4 event and it did not rupture two orthogonal faults, which can be approximated more easily by an isotropic kernel.

4.3 Forecasting experiment 3

In the third forecasting experiment, we simulated the 10-days Ridgecrest M7.1 sequence based on the parameters fitted to the local Ridgecrest M6.4 foreshock sequence. Since the instability of the sequence would lead to exploding forecasts, we assumed the long-term estimated FMD with $b = 1$.

4.3.1 Predicted aftershock productivity

Figure 5e shows that the number of aftershocks is predicted much more similarly by the five models than in *Experiments 1* and *2*. It suggests that the particular features of the model versions play a smaller role in the estimation over a closed, local sequence than in the generic fit over a long-term catalog with several sequences and seismically quiet periods in between. The ETAS conventional model reaches the observation in 4.4% of the simulation runs, the ETASI aniso-r even overestimates the size of the sequence in 94.1% of the simulations. The other models show very good predictions.

4.3.2 Predicted largest aftershock magnitude

According to Fig. 5f, our manual choice of $b = 1$ led to very realistic predictions of the largest aftershock magnitude. Together with the results for the number of aftershocks, it shows that the Ridgecrest left the unstable state after the M7.1 event by returning to the generic FMD, while retaining a similar structure of aftershock productivity.

4.3.3 Spatial distribution

Finally, Fig. 6e and f suggests that, compared to *Experiment 2*, the spatial kernels fitted over the Ridgecrest M6.4 sequence are much narrower than those coming from the generic, long-term model fit. This is confirmed by the larger estimates of q and the smaller estimates of γ in Table 2. Figure 7c shows that the narrower spatial distribution leads to a more pronounced information gain by the local restriction and the anisotropy, relative to the ETAS conventional model.

Note that, to some extent, the predicted spatial distributions show traces of late or secondary aftershocks triggered along the orthogonal M6.4 Ridgecrest fault, in contrast to very few observed events in that area. This might be an indication of an underestimated Omori parameter p or an overestimated c, favoring pronounced triggering over a longer time period.

4.4 Summary of forecast quality

Figure 7 shows a summary of the quality measures for the three experiments, with respect to the predicted number of aftershocks in Fig. 7a, largest aftershock magnitude in Fig. 7b and spatial aftershock distribution in Fig. 7c. The conventional model scores worst in each category. It confirms the results in Grimm et al. (2021), who argued that the isotropic and unlimited spatial kernel assumes an implausibly far trigger reach and leads to underestimated cluster sizes.

According to Fig. 7a, the ETASI models seem to predominantly estimate more realistic aftershock productivties than the ETAS models when fitted over the long-term Californian catalog (see *Experiments 1* and *2*). When fitted over the specific Ridgcrest M6.4 sequence, the bias of an underestimated aftershock productivity seems to be balanced out by not cutting out undetected events. Anisotropic models always lead to larger predicted sequence sizes, in the case of *Experiment 3* even to a substantial overestimation.

The predictions of the largest aftershock magnitude, shown in Fig. 7b, are reasonable for all but *Experiment 1*. Apparently, the short-term incompleteness bias in the ETAS models is of much less consequence for the FMD than for the aftershock productivity.

According to Fig. 7c, as expected, the anisotropic models predict more realistic spatial event distributions. The spatial restriction leads to a much smaller improvement.

4.5 Sensitivity of results

As a sensitivity study, we forecasted the Ridgecrest M7.1 sequence over a duration of 50 days. In a longer time window, direct aftershock productivity has less dominance, and is being displaced more and more by secondary triggering. The underestimation of direct aftershock productivity (e.g. in the ETAS conventional model) typically goes along with more pronounced secondary triggering, characterized by larger estimates of the productivity constant A, see Table 2. Therefore, we observed that the ETAS conventional model caught up the missing events over time. On the other hand, this indicates a temporal distribution of aftershocks which is not in agreement with the observations. Other sensitivity tests, such as the model estimation with varying cut-off magnitudes M_c or different time windows for the generic California estimates showed generally stable results.

5 Conclusion

In this article, we combined an ETAS approach with generalized anisotropic and locally restricted spatial kernels (Grimm et al. 2021) with the ETASI time model of Hainzl (2021). The new features have the objective to solve the three major biases of the conventional ETAS model, which are the isotropic and spatially unlimited kernel as well as the neglection of short-term incompleteness in the fitted event records.

We estimated five different versions of the new ETASI time-space model to a generic, long-term Californian event set and to the specific Ridgecrest M6.4 foreshock sequence. Then, we applied the fitted model parameters to generate synthetic forecasts of the Ridgecrest M6.4 and the M7.1 sequences, which we analyzed regarding the predicted size of the sequence, largest aftershock magnitude and spatial aftershock distribution.

The results indicate that the ETAS conventional model leads to a substantial underestimation of the number of aftershocks due to its disproportionately small estimates of the direct aftershock productivity for the M6.4 and M7.1 trigger events. The locally restricted ETAS models without ETASI-extension provide more realistic, but still underestimated predictions, as they are affected by the bias of short-term incomplete event sequences in the fitted event set. The combination of ETASI model with locally restricted spatial kernel seems to solve the bias and provides the most robust predictions in the forecasting experiments. The anisotropy of the spatial kernel has a positive impact on the model estimation, however, shows its strength more clearly in the prediction of the spatial event distribution of aftershocks.

More as a by-product, we find that the Ridgecrest M6.4 foreshock sequence showed instable behavior, favoring strong aftershocks by a small Gutenberg-Richter parameter $b < 0.8$. The instability of the foreshock sequence can be interpreted as an indication of an imminent strong mainshock. In consequence, models fitted on the long-term, stable Californian event records cannot adequately predict the magnitude distribution of this sequence.

The new model provides a better understanding of spatio-temporal earthquake clustering and solves three major biases of the conventional ETAS model at once. Particularly, it leads to better estimates of the aftershock productivity and to improved forecasts of the size of a sequence and the spatial distribution of aftershocks. These improvements may be of major interest for short-term risk assessment during an on-going aftershock sequence, particularly for the risk of a second, damaging earthquake following the initial trigger event. The anisotropic spatial forecast of aftershock locations enables desaster response managers to take actions in areas at risk where particularly high aftershock activity is expected.

Future work should test the forecast quality for other earthquake sequences. It would be interesting to address the question whether the ETASI time-space model can be used to reliably detect the instability of a live sequence, which could have positive impacts on emergency management during on-going sequences. An evaluation of the goodness of fit for the temporal event distribution should be included into such analyses.

6 Data and resources

The earthquake event set for California has been downloaded from the Southern California Earthquake Data Center (https://scedc.caltech.edu/data/alt-2011-dd-hauksson-yang-shearer.html, last accessed on October 25, 2021). Results and figures were produced using Matlab software. The source code for model estimation and simulation is made freely available by the first author in the Github repository https://github.com/ChrGrimm/ETASanisotropic.

Acknowledgements We thank the anonymous reviewers for their helpful and constructive feedback that has significantly improved the work. Particularly, we thank Marco Pagani for valuable discussions during the on-going research. Financial support for this work was provided by Munich Re through a scholarship granted to the first author, and by the Department of Statistics at Ludwig-Maximilians-University Munich. S.H. was supported by the Deutsche Forschungsgemeinschaft (DFG) Collaborative Research Centre 1294 (Data Assimilation – The seamless integration of data and models, project B04) and the European Unions Horizon 2020 research and innovation program under Grant Agreement Number 821115, real-time earthquake risk reduction for a resilient Europe (RISE).

Authors contribution CG (first author) derived and programmed the model, conducted simulation experiments, prepared figures and wrote the paper. SH provided theory of ETASI time model, consulted in programming and conducted comparative calculations. CG, SH and MK designed the simulation experiments and interpreted the results. HK consulted on statistical questions and was involved in the concept of the study.

Funding Open Access funding enabled and organized by Projekt DEAL. Financial support for this work was provided by Munich Re through a scholarship granted to the first author, and by the Department of Statistics at Ludwig-Maximilians-University Munich (e.g. travel costs, conference and journal fees).

Declarations

Conflict of interest The authors acknowledge that there are no relevant financial or non-financial interests to disclose.

Human or animal rights This article does not contain any studies involving human participants or animals performed by any of the authors.

References

de Arcangelis L, Godano C, Lippiello E (2018) The overlap of aftershock coda waves and short-term postseismic forecasting. J Geophys Res Solid Earth 123:5661–5674. https://doi.org/10.1029/2018JB015518 (**ISSN: 21699356**)

Grimm C, Käser M, Hainzl S, Pagani M, Küchenhoff H (2021) Improving earthquake doublet frequency predictions by modified spatial trigger kernels in the epidemic-type aftershock sequence (ETAS) model. Bull Seismol Soc Am. https://doi.org/10.1785/0120210097 (**ISSN: 0037-1106**)

Gutenberg B, Richter CF (1944) Frequency of earthquakes in California. Bull Seismol Soc Am 34:185–188. https://doi.org/10.1038/156371a0 (**ISSN: 00280836**)

Hainzl S (2016) Rate-dependent incompleteness of earthquake catalogs. Seismol Res Lett 87(2A):337–344. https://doi.org/10.1785/0220150211 (**ISSN: 19382057**)

Hainzl S (2016) Apparent triggering function of aftershocks resulting from rate-dependent incompleteness of earthquake catalogs. J Geophys Res Solid Earth 121(9):6499–6509. https://doi.org/10.1002/2016JB013319 (**ISSN: 21699356**)

Hainzl S (2021) ETAS-approach accounting for short-term incompleteness of earthquake catalogs. Bull Seismol Soc Am. https://doi.org/10.1785/0120210146 (**ISSN: 0037-1106**)

Hainzl S, Christophersen A, Enescu B (2008) Impact of earthquake rupture extensions on parameter estimations of point-process models. Bull Seismol Soc Am 98(4):2066–2072. https://doi.org/10.1785/0120070256

Hainzl S, Zakharova O, Marsan D (2013) Impact of aseismic transients on the estimation of aftershock productivity parameters. Bull Seismol Soc Am 103(3):1723–1732. https://doi.org/10.1785/0120120247 (**ISSN: 00371106**)

Hainzl S, Zakharova O, Marsan D (2013) Impact of aseismic transients on the estimation of aftershock productivity parameters. Bull Seismol Soc Am 103(3):1723–1732. https://doi.org/10.1785/0120120247 (**ISSN: 00371106**)

Harte DS (2016) Model parameter estimation bias induced by earthquake magnitude cut-off. Geophys J Int 204(2):1266–1287. https://doi.org/10.1093/gji/ggv524 (**ISSN: 1365246X**)

Hauksson E, Yang W, Shearer PM (2012) Waveform relocated earthquake catalog for Southern California (1981 to June 2011). Bull Seismol Soc Am 102(5):2239–2244. https://doi.org/10.1785/0120120010 (**ISSN: 00371106**)

Hutton K, Woessner J, Hauksson E (2010) Earthquake monitoring in Southern California for seventy-seven years (1932–2008). Bull Seismol Soc Am 100(2):423–446. https://doi.org/10.1785/0120090130 (**ISSN: 00371106**)

Jalilian A (2019) ETAS: An R package for fitting the space-time ETAS model to earthquake data. *J Stat Softw*. https://doi.org/10.18637/jss.v088.c01. ISSN: 15487660

Lippiello E, Cirillo A, Godano G, Papadimitriou E, Karakostas V (2016) Real-time forecast of aftershocks from a single seismic station signal. Geophys Res Lett 43(12):6252–6258. https://doi.org/10.1002/2016GL069748 (**ISSN: 19448007**)

Marsan D, Ross ZE (2021) Inverse migration of seismicity quiescence during the 2019 Ridgecrest sequence. J Geophys Res Solid Earth 126(3):1–13. https://doi.org/10.1029/2020JB020329 (**ISSN: 21699356**)

Marsan D, Lengliné O (2008) Extending earthquakes' reach through cascading. *Science*, 319(5866):1076–1079. https://doi.org/10.1126/science.1148783. ISSN: 00368075

Mizrahi L, Nandan S, Wiemer S (2021) The effect of declustering on the size distribution of mainshocks. Seismol Res Lett. https://doi.org/10.1785/0220200231 (**ISSN: 23318422**)

Ogata Y (1988) Statistical models for earthquake occurrences and residual analysis for point processes. J Am Stat Assoc 83(401):9–27 (**ISSN: 0162-1459**)

Ogata Y (1998) Space-time point-process models for earthquake occurrences. Ann Inst Stat Math 50(2):379–402 (**ISSN: 00203157**)

Ogata Y (2011) Significant improvements of the space-time ETAS model for forecasting of accurate baseline seismicity. Earth

Planets Space 63(3):217–229. https://doi.org/10.5047/eps.2010.09.001 (**ISSN: 18805981**)
Ogata Y, Zhuang J (2006) Space-time ETAS models and an improved extension. Tectonophysics 413(1–2):13–23. https://doi.org/10.1016/j.tecto.2005.10.016 (**ISSN: 00401951**)
Omi T, Ogata Y, Hirata Y, Aihara K (2013) Forecasting large aftershocks within one day after the main shock. Sci Rep 3:1–7. https://doi.org/10.1038/srep02218 (**ISSN: 20452322**)
Omi T, Ogata Y, Hirata Y, Aihara K (2014) Estimating the ETAS model from an early aftershock sequence. Geophys Res Lett 41:850–857. https://doi.org/10.1002/2013GL058958 (**ISSN: 19448007**)
Page MT, van Der Elst N, Hardebeck J, Felzer K, Michael AJ (2016) Three ingredients for improved global aftershock forecasts: tectonic region, time-dependent catalog incompleteness, and intersequence variability. Bull Seismol Soc Am 106(5):2290–2301. https://doi.org/10.1785/0120160073 (**ISSN: 19433573**)
Rhoades DA, Gerstenberger MC, Christophersen A, Zechar JD, Schorlemmer D, Werner MJ, Jordan TH (2014) Regional earthquake likelihood models II: information gains of multiplicative hybrids. Bull Seismol Soc Am 104(6):3072–3083. https://doi.org/10.1785/0120140035 (**ISSN: 19433573**)
Seif S, Mignan A, Zechar JD, Werner MJ, Wiemer S (2017) Estimating ETAS: the effects of truncation, missing data, and model assumptions. J Geophys Res Solid Earth 122(1):449–469. https://doi.org/10.1002/2016JB012809 (**ISSN: 21699356**)
Utsu T, Ogata Y, Matsu'ura RS (1995) The centenary of the Omori formula for a decay law of aftershock activity. J Phys Earth 43:1–33
Wells DL, Coppersmith KJ (1994) New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacements. Bull Seismol Soc Am 84(4):974–1002 (**ISSN 0037-1106**)
Zhang L, Werner MJ, Goda K (2018) Spatiotemporal seismic hazard and risk assessment of aftershocks of M 9 megathrust earthquakes. Bull Seismol Soc Am 108(6):3313–3335. https://doi.org/10.1785/0120180126
Zhuang J, Ogata Y, Vere-Jones D (2002) Stochastic declustering of space-time earthquake occurrences. J Am Stat Assoc 97(458):369–380. https://doi.org/10.1198/016214502760046925 (**ISSN: 01621459**)
Zhuang J, Ogata Y, Wang T (2017) Data completeness of the Kumamoto earthquake sequence in the JMA catalog and its influence on the estimation of the ETAS parameters. Earth Planets Space 69:36. https://doi.org/10.1186/s40623-017-0614-6 (**ISSN: 18805981**)

Part III.

Approach 2: Statistical Regression Models

7. New Statistical Perspectives on Bath's Law and Aftershock Productivity

Contributing article

Grimm, C., Rupprecht, T., Johnson, K., Hainzl, S., Küchenhoff, H., Käser, M., Pagani, M. (2022b). New Statistical Perspectives on Bath's Law and Aftershock Productivity. *Manuscript submitted for publication.*

Author contributions

Survival model: Christian Grimm (CG) brought up the main idea to model magnitude differences ΔM globally. Helmut Küchenhoff suggested to use survival methods and consulted on the statistical analysis. CG designed the regression model, chose and prepared covariate data, declustered the global catalog and implemented the analysis code in R. Additionally, he ran and compared different model approaches, chose the final model, prepared figures and interpreted its results.

GAM: CG brought up the idea to model the variation of aftershock productivity as one of the main drivers of ΔM. CG declustered the local catalog using the ETAS-Incomplete method from the second contribution (Grimm et al., 2022a), and prepared most of the covariate data. Kendra Johnson (KJ) classified events in tectonic regions. Teresa Rupprecht (TR) designed and compared different model approaches, chose the final model, prepared figures and implemented the analysis code in R. CG and TR interpreted the results.

Both models: CG wrote the initial draft. KJ and Sebastian Hainzl (SH) substantially improved the geophysical interpretation of the results. Martin Käser and Marco Pagani consulted on seismic datasets and the study concept. All authors proofread the paper and suggested improvements.

New Statistical Perspectives on Bath's Law and Aftershock Productivity

Christian Grimm[a], Teresa Rupprecht[a], Kendra Johnson[b], Sebastian Hainzl[c], Helmut Küchenhoff[a], Martin Käser[d,e], Marco Pagani[b]

[a]Ludwig-Maximilians-University Munich, Department of Statistics, Ludwigstraße 33, 80539 Munich, Germany, christian.grimm@stat.uni-muenchen.de, https://orcid.org/0000-0002-2190-2981 (CG); Teresa.Rupprecht@campus.lmu.de (TR); Kuechenhoff@stat.uni-muenchen.de (HK)

[b]Global Earthquake Model Foundation, Via Ferrata 1, 27100 Pavia, Italy, kendra.johnson@globalquakemodel.org (KJ); marco.pagani@globalquakemodel.org (MP)

[c]GFZ German Research Centre for Geoscience, Physics of Earthquakes and Volcanoes, Helmholtzstraße 6/7, 14467 Potsdam, Germany; hainzl@gfz-potsdam.de, https://orcid.org/0000-0002-2875-0933 (SH)

[d]Ludwig-Maximilians-University Munich, Department of Earth and Environmental Sciences, Geophysics, Theresienstraße 41, 80333 Munich, Germany; martin.kaeser@geophysik.uni-muenchen.de (MK)

[e]also at Munich Re, Section GeoRisks, Königinstr. 107, 80802 Munich, Germany

Abstract. The well-established Bath's law states that the average magnitude difference between a mainshock and its strongest aftershock is roughly 1.2, independently of the size of the mainshock. The main challenge in calculating this value is the bias introduced by missing data points when the strongest aftershock is below the observed cut-off magnitude. Ignoring missing values leads to a systematic error, because the data points removed are those with particularly large magnitude differences ΔM. The error is minimized, if we restrict the statistics to mainshocks at least two magnitude units above the cut-off, but then the sample size is strongly reduced. This work provides an innovative approach for modelling ΔM by adapting methods for time-to-event data, which often suffers from incomplete observation (censoring). In doing so, we adequately account for unobserved values and estimate a fully parametric distribution of the magnitude differences ΔM for $M > 6$ mainshocks. Results show that magnitude differences are best modeled by the Gompertz distribution, and that larger ΔM are expected at increasing depths and higher heat flows. A simulation experiment suggests that ΔM is mainly driven by the number and the magnitude distribution of aftershocks. Therefore, in a second study, we modelled the variation of aftershock productivity in a stochastically declustered local catalog for New Zealand, using a generalized additive model approach. Results confirm that aftershock counts can be better modelled by a Negative Binomial than a Poisson distribution. Interestingly, there is indication that triggered earthquakes trigger themselves two to three times more aftershocks than comparable background events. This effect can either be an indicator for incorrect trigger pair assignments as a result of the declustering approach, or it may represent an actual effect due to a higher prevalent energy level in the tectonic system during on-going earthquake sequences. If confirmed, this effect must be carefully considered in forward simulations of earthquake sequences, as otherwise there is a risk of substantially underestimating cluster sizes and consequently overestimating ΔM.

Keywords: Bath's law, survival models, aftershock productivity, generalized additive models.

Main author contact information: Christian.Grimm@stat.uni-muenchen.de

1 Introduction

As energy is released in the event of a strong earthquake, tectonic stress redistributes in the surroundings of the initial rupture and typically results in further earthquakes, so-called *aftershocks* (Utsu et al., 1995). The cascade of aftershocks is commonly referred to as an *earthquake sequence*, and the strongest event of the sequence is called the *mainshock*. Typically, events that occurred shortly before the mainshock, so-called *foreshocks*, are included in the sequence since they are believed to be physically related to the upcoming major earthquake (e.g. Helmstetter and Sornette, 2003).
Extensive research has been carried out to analyze and model the spatio-temporal properties of earthquake sequences, e.g. through the Epidemic Type Aftershock Sequence (ETAS) model (Ogata, 1988, 1998; Zhuang et al., 2002). Studies found a well-established power-law decay of aftershock rates as a function of the time after the mainshock (Omori, 1895), while the spatial cluster is typically elongated rather than isotropic around the mainshock's rupture plane (e.g. Grimm et al., 2022, 2021; Hainzl et al., 2008; Ogata, 2011; Ogata and Zhuang, 2006; Zhang et al., 2018).
Aftershocks are a relevant risk driver since even moderate events can substantially increase damage in buildings and infrastructure destabilized by a prior mainshock. Similarly, foreshocks can set the stage for more severe mainshock damage (Abdelnaby, 2012; Kagermanov and Gee, 2019; Papadopoulos et al., 2020). Therefore, one of the central questions for emergency and risk management purposes is: *What is the second strongest magnitude to be expected in a sequence?*
To date, the literature only provides a starting point for answers to this question. The well-established Bath's law states that the average magnitude difference ΔM between a mainshock and its strongest aftershock is roughly 1.2, *independently* of the size of the mainshock (Bath, 1965). The main challenge in calculating this value is the bias introduced by missing data, if no aftershock was observed above the cut-off magnitude M_c of the catalog and therefore ΔM cannot be computed. We cannot simply ignore missing values, as these are the ones with particularly large magnitude differences ΔM. Therefore, leaving them out would lead to a systematic bias. Several authors found that the statistics is robust, if we restrict the sample to mainshocks at least two magnitude units above M_c, but then the sample size is strongly reduced (e.g. Tahir et al., 2012; Zakharova et al., 2013). Another work around was suggested by Zakharova et al. (2013), who modeled the seismic moment ratio between aftershocks and the mainshock, rather than ΔM, approximating the ratio by zero if no aftershocks were recorded.
In any case, Bath's law only makes a statement about the average value of the ΔM, but not about their distribution (and its parameters) or any important quantiles in the lower tail of the distribution.

Another term that appears occasionally in the literature is that of an *earthquake doublet*. Doublets are generally defined as a pair of two similarly strong earthquakes, occurring temporarly and spatially close to each other (e.g. Felzer et al., 2004; Grimm et al., 2021; Kagan and Jackson, 1999). Kagan and Jackson (1999) found that approximately 22% of the $M > 7.5$ earthquakes worldwide occurred accompanied by another $M > 7.5$ event within a distance of one rupture length and with an inter-event time of considerably less than their recurrence time estimated from plate motion. Grimm et al. (2021) showed that roughly 17% of the global $M \geqslant 6$ mainshocks and more than 20% of the mainshocks in Japan were part of an earthquake doublet, defining them as a pair of earthquakes with no more than 0.4 magnitude units difference, occurring within 365 days and a radius of 2.5 rupture lengths.

1.1 Survival Model Regression of ΔM

In the first part of this work, we propose an innovative approach that models the full, parametric distribution of ΔM by adapting so-called *survival models*, originally developed for medical applications. Survival models are a class of regression models that account for data with a censored (or truncated) response variable (see e.g. Klein and Moeschberger (2003) for comprehensive overviews). As the term "survival" suggests, these models were originally developed in applications where the response represents the non-negative lifetime of a patient in medical studies or the lifetime of a device in engineering contexts (so-called *reliability* or *failure time analysis*). The above applications have in common that the exact value of the response is unknown, if the event has not occurred until the end of the study period.
Replacing lifetimes by magnitude differences, we can therefore use survival models to account for the missing ΔM values where we only have the partial information that $\Delta M > M - M_c$, given mainshock magnitude M.
To do so, we decluster a global catalog using a window method, and compute the (partially right-censored) ΔM between the mainshock and the second strongest event of each cluster. Note that the latter may be a foreshock or an aftershock, as both are relevant in a risk management context. Then, we enrich the cluster set by a plate boundary classification, relative plate velocities, sea floor age and heat flow data, to investigate the regression effects of these large-scale geophysical conditions on the distribution of ΔM.

1.2 Drivers of ΔM

The magnitude difference ΔM is controlled by the two drivers (Grimm et al., 2021)

1. number of aftershocks (hereafter called *aftershock productivity*) and
2. frequency-magnitude distribution (FMD) of triggered events.

A simple simulation shall demonstrate the effect of both factors. Assume an initial mainshock of magnitude $M = 8$. Let the expected aftershock productivity of an earthquake with magnitude M be

$$k(M) = A \exp(\alpha \, (M - M_c)), \qquad M \geqslant M_c, \tag{1}$$

and the FMD be the exponential distribution with probability density function (pdf)

$$f(M) = \beta \, e^{-\beta(M - M_c)}, \qquad \beta > 0, \; M \geqslant M_c, \tag{2}$$

where β is related to the Gutenberg-Richter b-value by $b = \beta/ln(10)$ (Gutenberg and Richter, 1944). If we assume the realistic parameters $A = 0.13$, $\alpha = 2.0$ and $b = 1.0$ and, for simplicity, a Poisson distributed number of aftershocks with trigger-magnitude dependent parameter $\lambda = k(M)$, we can simulate a distribution of ΔM with a mean of 1.2, consistent with Bath's Law.
A doubled aftershock productivity (i.e. $A = 0.26$) would lead to a pronounced drop of the average magnitude difference down to 0.66. If the increase of A does only apply to secondary triggering, but the number of direct aftershocks of the $M8$ mainshock remains unchanged, the mean of ΔM still decreases to below 0.9. A similar reduction of the average ΔM is achieved, if instead we modify the FMD, sampling magnitudes according to $b = 0.85$.

1.3 Regression of Aftershock Productivity

The simplified simulation experiment above illustrates the leverage of both aftershock productivity and FMD on the magnitude differences ΔM. This gave motivation for a more in-depth analysis of the variation in aftershock productivity in the second part of this study. To do so, we decluster a local earthquake catalog for New Zealand and fit a generalized additive model (GAM) to the estimated number of direct aftershocks per event, comparing a Poisson with a Negative Binomial distribution. Here, we use the stochastic declustering method introduced by Zhuang et al. (2002), based on the ETAS-Incomplete model proposed by Grimm et al. (2022) that accounts for anisotropic ruptures as well as short-term aftershock incompleteness to reduce estimation biases. We enrich the local catalog by classifying the events to the main tectonic contexts in the surrounding of the Hikurangi fault, and by concluding the slip type from additional focal mechanism data.
Extensive research has been done on the variation of aftershock productivity. Kagan (2017) and Shebalin et al. (2018) showed that aftershock counts are best modeled by the Negative Binomial distribution due to their large variance. Page et al. (2016) found that aftershock productivity may regionally vary by a factor of almost 10, which would explain the variation in ΔM to large extent. Marsan and Helmstetter (2017) found that 40-80% of the aftershock variability may be related to variation in the mainshock stress drop. Dascher-Cousineau et al. (2020) investigated a large number of source and site effects on aftershock productivity and showed individual correlations of stress drop and rupture dimension with the number of aftershocks. Wetzler et al. (2016) suggest a larger productivity in subduction zones of the western circum-Pazific, compared to the eastern side. Zhuang et al. (2004) proposed that triggered events produce more aftershocks than comparable background events, which would provide an additional boost in clustering and decrease expected ΔM's.
Potential correlations of the FMD of triggered events with the magnitude of the direct ancestor or the cluster mainshock were found by Zhuang et al. (2004), Gulia et al. (2018) and Nandan et al. (2019), and may considerably increase the chance of small ΔM. However, they are out of the scope of this work.

1.4 Scope and Outline

This work consists of two regression studies, the analysis of magnitude differences (hereafter called ΔM*-regression*), and the analysis of the aftershock productivity (referred to as *productivity regression*). The focus is on the innovate approach to estimate a fully parametric distribution of ΔM, using survival models that take into account right-censored data rather than avoid it. To our

knowledge, no similar approach has been pursued in the literature so far. Especially in the ΔM-regression, covariates represent rather large-scale regional effects. Attempts to consider small-scale variations of these covariates or to include further event specific data are out of the scope of this paper.
Sections 2 and 3 introduce the utilized datasets, the declustering approaches and the compilation of the covariate datasets for both regression studies, respectively. Next, Section 4 rigorously explains the survival model and GAM methodological approaches. Then, the results of the regression studies are shown and discussed in Sections 5 (ΔM-regression) and 6 (productivity regression). Finally, conclusions are drawn from a joint interpretation and related future research topics are recommended.

2 Data for ΔM-Regression

This section summarizes the compilation of the regression dataset for the analysis of global magnitude differences between the mainshock and the second strongest event in the cluster. First, we justify the choice of the underlying global earthquake catalog. Then, we outline the window method to decluster the catalog, followed by the definition of the response variable. Finally, the enrichment of further geophysical variables as regression covariates is explained.

2.1 Global Earthquake Catalog

The choice of an appropriate global earthquake catalog for the regression of magnitude differences raises two requirements which, however, are not fully met by any currently available catalog and therefore requires a trade-off. On the one hand, the catalog should ideally have homogeneous magnitude scales, and be reliably complete in any part of the world, including far off-shore regions and aftershocks occurring shortly after the mainshock. On the other hand, it should be complete to the smallest possible magnitude level to assure a sufficient observable magnitude range of at least one unit below the smallest mainshock magnitude of interest, $M > 6.0$.
Despite not providing homogenized magnitude scales, we chose the U.S. Geological Survey National Earthquake Information Center (USGS-NEIC) catalog. We extracted all events from 1973 until 2021 with depths smaller than 70 km that occurred at a maximum of 300 km distance to a tectonic plate boundary according to the digital model by Bird (2003). The completeness magnitude of this dataset is $M_c = 5.0$ according to Kagan and Jackson (2010) and Tahir et al. (2012), which allows us to apply the regression model to clusters with mainshock magnitude larger than 6.0.
To test the influence of inhomogeneous magnitude scales, we performed sensitivity analyzes using the International Seismological Centre – Global Earthquake Model (ISC-GEM) instrumental catalogue, which is a relocated global event set with homogenized magnitude scales (Bondár et al., 2015; Di Giacomo et al., 2015a,b, 2018; Storchak et al., 2015). Due to its higher level of magnitude completeness, $M_c = 5.6$ according to Di Giacomo et al. (2015b) and $M_c = 6.0$ according to Michael (2014) since 1964, we have to limit our statistical analysis to mainshocks with $M \geqslant 6.6$.

2.2 Declustering of Global Catalog

In order to obtain a set of *independent* clusters, including the information about the magnitude difference ΔM between the mainshock and the largest aftershock (or foreshock), we declustered the global earthquake catalog using a rather simple window method (see e.g. Gardner and Knopoff,

1974; Uhrhammer, 1986; van Stiphout et al., 2012). To do so, we first sorted the catalog in descending magnitude order. Then, we consecutively searched aftershocks occurring within a time window of $T = 100$ days and a spatial radius of $R(m) = 2.5\,L(m)$, where $L(m) = 10^{-2.44+0.59m}$ is the expected rupture length of the mainshock, depending on its magnitude m, according to Wells and Coppersmith (1994). Similar to Reasenberg (1985), we linked clusters if an event *B* is found to trigger the potential aftershock *A*, but *A* is the mainshock of an already identified cluster and therefore, due to prior re-ordering of the catalog, has the larger magnitude, $m_A \geqslant m_B$. In this case, event *B* is called a foreshock of *A*.
We conducted sensitivity studies that showed that the regression results are insensitive to varying definitions such as $T = 365$ days and $R(m)$ varying between $1.0\,L(m)$ and $2.5\,L(m)$.

2.3 Response Variable

For each cluster, the magnitude difference ΔM is computed between the mainshock (i.e., the strongest event of the cluster) and the second-strongest event, be it a foreshock or aftershock. In total, we obtain 2,933 clusters with mainshock magnitudes $M > 6.0$.
Note that 1,180 of these are *single-event* clusters, i.e., no associated foreshock or aftershock was found in the corresponding time-space window. Based on seismological reasoning we can assume that these mainshocks actually triggered aftershocks, but that these were simply too weak to be recorded in the dataset, given its cut-off magnitude M_c. Therefore, if for a mainshock i with magnitude $M_i \geqslant M_c$ no second event is listed, we have the partial information that the magnitude difference is $\Delta M_i > M_i - M_c$. The single clusters are the reason why we need advanced regression models that can deal with censored data.

2.4 Covariates

We enriched the declustered catalog by additional geophysical site information interpolated to the mainshock locations by a nearest neighbor approach.
Using the digital plate boundary model of Bird (2003), we categorized each event into one of seven *plate boundary classes* continental convergence boundary (CCB), continental transform fault (CTF), continental rift boundary (CRB), oceanic spreading ridge (OSR), oceanic transform fault (OTF), oceanic convergent boundary (OCB) and subduction zone (SUB). Fig. 1 shows the mainshock locations of the declustered catalog, color-coded by the corresponding plate boundary class assigned to them. Table 1 lists the number of clusters with censored and observed ΔM value per boundary class, respectively. Note that almost half of the clusters are assigned to a subduction zone, and that oceanic spreading ridges and transform faults host more censored than observed data points.

From the same digital model, we assigned estimates of the *relative plate velocity* and *sea floor age* from the next boundary segment point to the mainshock locations. Likewise, using a nearest neighbor approach, we interpolated values from the scattered *heat flow* dataset of Bird et al. (2008), provided to us by the author. Fig. 2 illustrates the marginal distributions of the interpolated covariate data at the mainshock locations, grouped by the assigned plate boundary class. Subduction zones show the largest relative plate velocities (values range between 0.4 and 262 mm/a, see Fig. 2(a)), while oceanic spreading ridges and transform faults provide the youngest sea floor ages (between 0 and 262 Ma, see Fig. 2(b)) and the largest heat flows (between 0.025 and 0.3 Wm^{-2}, see Fig. 2(b)).

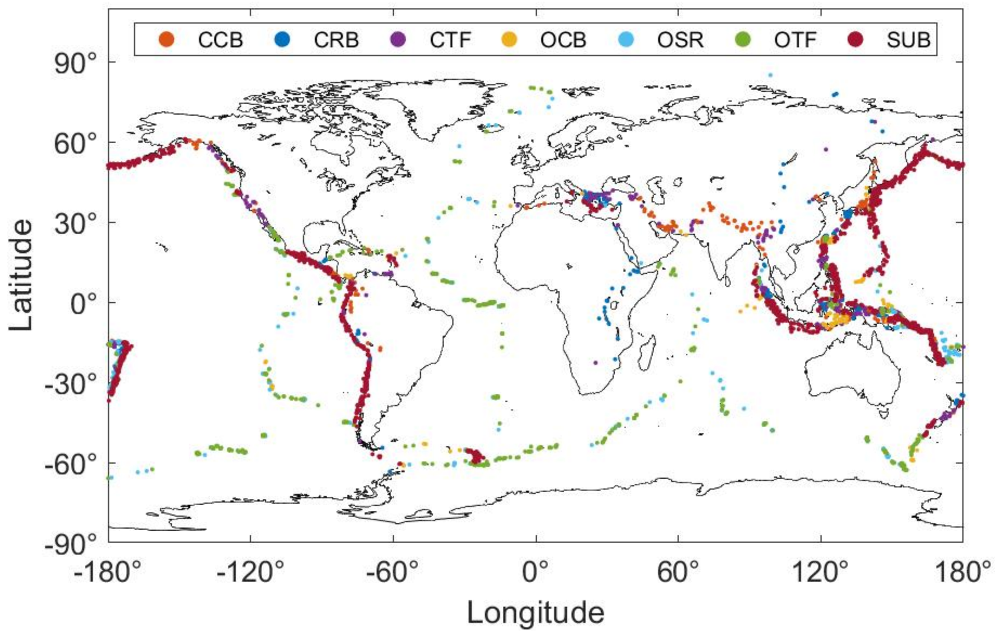


Fig 1: Locations of 2,933 global $M > 6$ mainshocks between 1973 and 2020 after declustering the USGS-NEIC catalog. Mainshocks are colour-coded according to their assignment to the plate boundary classes *continental convergence boundary (CCB), continental transform fault (CTF), continental rift boundary (CRB), oceanic spreading ridge (OSR), oceanic transform fault (OTF), oceanic convergent boundary (OCB) and subduction zone (SUB)*, introduced in the digital plate model of Bird (2003).

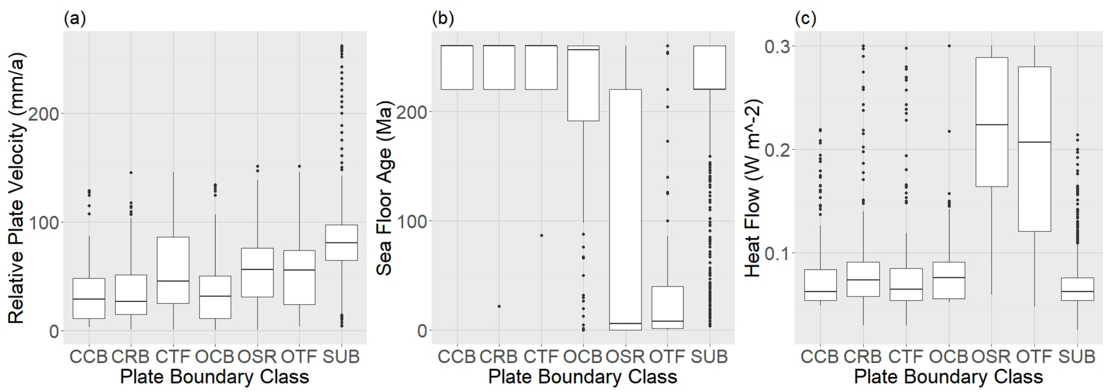


Fig 2: Boxplots of (a) relative plate velocity, (b) sea floor age, and (c) heat flow values assigned to cluster mainshocks by a nearest approach from original scatter data (Bird, 2003; Bird et al., 2008), grouped by the plate boundary class. Acronyms of boundary classes are spelled out in the caption of Fig. 1.

Table 1: Number of censored and observed ΔM data points, grouped by plate boundary class.

Plate Boundary	Number of Clusters	
Class	#censored	#observed
CCB	71	148
CTF	82	157
CRB	69	114
OSR	108	73
OTF	322	194
OCB	59	83
SUB	474	989

3 Data for Productivity-Regression

This section summarizes the dataset compilation for the regression of aftershock productivity. First, we briefly introduce the chosen local event set for New Zealand. Then, we rigorously describe the stochastic declustering method which is applied in order to estimate the number of aftershocks as the response variable. Finally, we describe the enrichment of the local event set by further geophysical properties.

3.1 Local New Zealand Catalog

We limited this study to the Hikurangi subduction zone in New Zealand. A local event set was provided by GNS Science as an input to the ongoing 2022 revision of the New Zealand National Seismic Hazard Model. Using the algorithm of Stepp (1972), we computed that the catalog is complete down to $M_c = 3.5$ from 1982; however, to be conservative, we assumed $M_c = 3.5$ from 1987, concurrent with an improvement to the seismic network. Fig. 3 shows the chosen extract of 11,091 events surrounding the Hikurangi fault, between 1987 and end of 2020, at depths down to 80 km.

3.2 Declustering of Local Catalog

For the regression of aftershock counts, we cannot use the window declustering method, as it does not distinguish *direct* from *secondary* aftershocks. Instead, we used the stochastic declustering approach based on the Epidemic Type Aftershock Sequence (ETAS) model, as introduced by Zhuang et al. (2002).
The ETAS model describes the spatio-temporal clustering behavior of the entire catalog, and models a dynamic event rate at time t and location (x, y), given the prior event history H_t, through two overlapping processes,

$$R(t, x, y|H_t) = u(x, y) + \sum_{i:t_i<t} R^{trig}(t, x, y, i), \tag{3}$$

where $u(x, y)$ denotes the time-invariant, aftershock-independent *seismic background* rate, and $\sum_{i:t_i<t} R^{trig}(t, x, y, i)$ is the sum of the *trigger rate* contributions by all events i that occurred prior

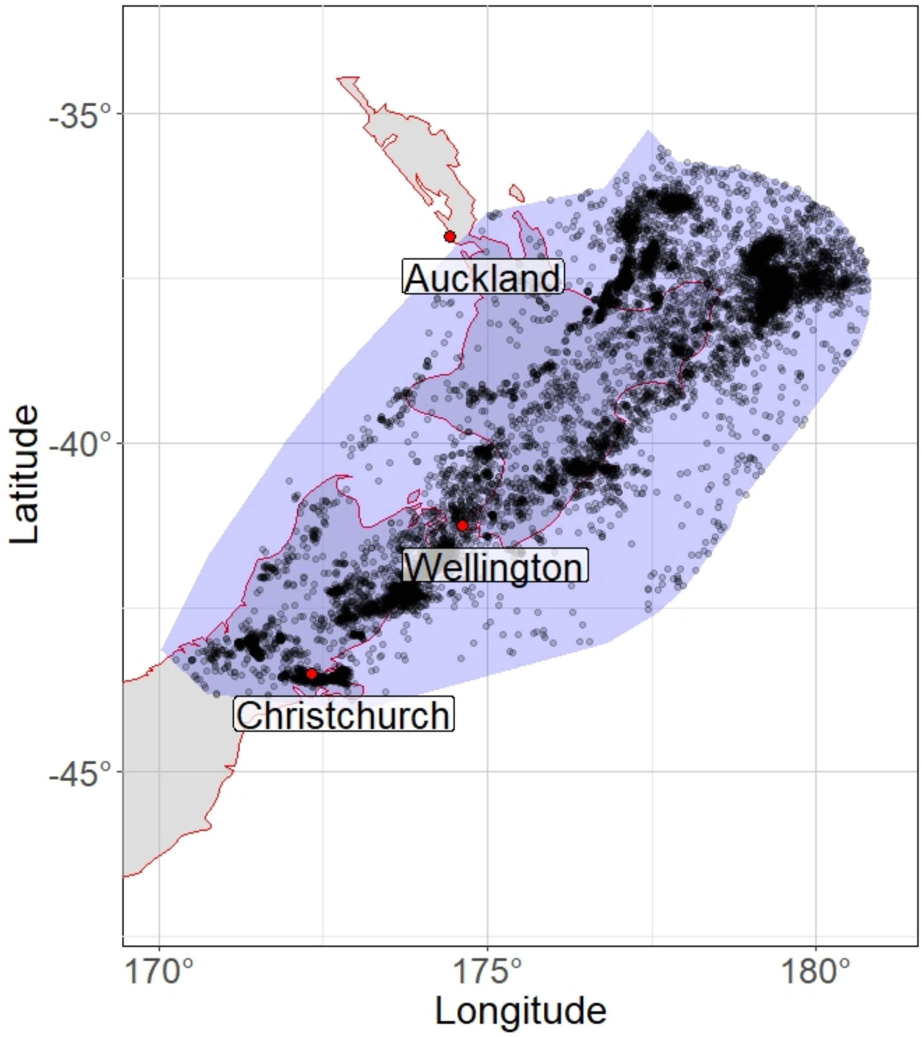


Fig 3: Spatial extract (blue polygon) of the Hikurangi subduction region in New Zealand, chosen for the aftershock productivity regression model. Black scatter points represent event locations of earthquakes with magnitude $M \geqslant 3.5$, depths $\leqslant 80$ km, that occurred between 1987 and 2020. The local event set was provided by GNS Science as an input to the ongoing 2022 revision of the New Zealand National Seismic Hazard Model.

to time t. For more details about the ETAS model, see e.g. Jalilian (2019); Ogata (1988, 1998); Zhuang et al. (2002).
From Equation (3), Zhuang et al. (2002) concluded that the probability, that the event j at time t_j and location (x_j, y_j) was an aftershock of the prior event i, is

$$P_{j,i} = \frac{R^{trig}(t_j, x_j, y_j, i)}{R(t_j, x_j, y_j|H_t)}.$$

Similarly, the probability that event j is a seismic background event and therefore independent of any prior trigger, is

$$P_{j,backgr} = \frac{u(x, y)}{R(t_j, x_j, y_j|H_t)}.$$

Thus, unlike the window method, the ETAS model provides probabilistic trigger associations between event pairs, and does not require the arbitrary choice of a fixed space-time window to search aftershocks. Instead, it optimizes built-in parametric functions that model the expected number of aftershocks (for trigger magnitude m_i),

$$K(m_i) = A\, e^{\alpha\, (m_i - M_c)}, \qquad A > 0, \quad \alpha > 0, \tag{4}$$

the temporal decay of aftershock rates (e.g. Omori-Utsu power law, see Omori, 1895), and a typically isotropic spatial distribution of aftershocks.
In this work, we used the *ETAS-Incomplete* model version of Grimm et al. (2022), who introduced a novel, anisotropic and locally restricted spatial kernel and accounted for incomplete records of aftershock sequences. The estimation source code is available in a public github repository (see Data and Resources). The new features solve the estimation biases due to the misfit of mostly elongated aftershock clouds by an isotropic kernel, and an underestimation of the trigger potential of strong mainshocks as a consequence of missing aftershock data. Both biases were shown to heavily affect our response variable, the aftershock productivity (Grimm et al., 2022, 2021; Hainzl, 2021; Hainzl et al., 2013; Page et al., 2016; Seif et al., 2017).

3.3 Response Variable

As the response variable of the regression model, we defined the estimated *number of direct aftershocks* for each event i in the catalog. We did this by counting the number of subsequent events j, for which i is the most probable trigger event, i.e. $P_{j,i} > P_{j,k} \;\; \forall\, k \neq i$, and that are more likely triggered by i than being a background event, i.e. $P_{j,i} > P_{j,backgr}$.
Note that the response is inevitably affected by the short-term incomplete records of aftershock sequences, as we can only count aftershocks that are recorded in the dataset. Nevertheless, the ETAS-Incomplete model approach avoids a biased parameter estimation, that would lead to manipulated declustering probabilities.

3.4 Covariates

The local catalog provides us the magnitude and depth for each event as immediate covariates for the regression model. Additionally, if the triggering event i was itself already triggered by a previous event, we traced back the trigger sequence and identified the largest magnitude in the

cluster, that occurred before event i. This covariate tests whether an event, that is a member of a cluster, is more or less productive than an independent background event, and whether its aftershock productivity is influenced by the previous mainshock.
A focal mechanism data set comprising 1,581 events of the chosen catalog extract was provided by GNS Science. We used the *nodal-plane.py* function from the public *GEMScienceTools/oq-mbtk* repository, based on the algorithm in Álvarez-Gómez (2019), to translate the given focal mechanisms into the slip type categories *normal (N)*, *strike-slip (SS)* and *reverse (R)* and mixed categories.
We classified each earthquake in the catalogue to the main tectonic regions. These regions are the *shallow crust*, *subduction interface*, *subduction intraslab deep* (within the subducting plate, but deeper than the zone of contact between the subducting and overriding plate), and *subduction intraslab shallow* (within the subducting plate, but in the shallow part of the plate beneath the interface, e.g. Reyners et al., 2010). To do so, we used the methodology described by Pagani et al. (2020), which classifies each earthquake based on its hypocentral position relative to surfaces (with buffers) that demarcate the boundaries of these regions. The surface used to represent the Hikurangi subduction interface and the top of the subducting plate was derived from a 3D model provided by GNS Science. Earthquakes shallower than 40 km and within 5 km of this surface were classified as interface; those shallower than 40 km and more than 5 km below the surface were classified as shallow slab; and those deeper than 40 km and within 5 km above or 60 km below this surface were classified as deep slab (the large below-slab buffer helps to capture earthquakes with large depth errors). Earthquakes shallower than the Moho (depths defined by LITHO1.0, Pasyanos et al., 2014) with a 10 km buffer were classified as crustal; crustal earthquakes were then sub-classified as occurring within or outside of the surface projection of the subduction zone. If an earthquake was classified into more than one tectonic region, then the following hierarchy was applied: interface is more likely than shallow slab, which is more likely than deep slab, which is more likely than shallow crustal. All other earthquakes were labelled as "unclassified" and not used in further analyses.

4 Regression Methods

In this section, we summarize the statistical models used in the two regression studies. Subsection 4.1 introduces survival regression models that can account for the censored ΔM response data due to unobserved aftershocks. Subsection 4.2 explains the use of a generalized additive model (GAM) for modeling aftershock counts in the local New Zealand catalog. All statistical analyses were performed with the open source software `R` (R Core Team, 2021).

4.1 Survival Models

4.1.1 Why Using a Survival Model for Earthquakes?

The magnitude difference ΔM between the mainshock and the second-largest earthquake of a sequence is only known, if at least one foreshock or aftershock was observed and assigned to the mainshock. Indeed, roughly 40% of the global clusters consist of a stand-alone mainshock. For these clusters, we can conclude that the second strongest event must be smaller than the cut-off magnitude M_c, i.e., that $\Delta M_i > M_i - M_c$, where M_i is the magnitude of mainshock i. In statistics, data points which are capped by such an upper observable threshold are called *right-censored*

(Klein and Moeschberger, 2003, section 3.2). Classical statistical models would substantially underestimate ΔM due to the relevant proportion of censored observations.
Replacing lifetimes by magnitude differences, our data meets the necessary requirements of a survival model,

- *non-negative responses* ($\Delta M \geqslant 0$)
- *independent responses* (mainshocks result from declustered catalog)
- *non-informative censoring* (i.e., conditional on covariates, censored clusters are not suspected to deviate structurally in their ΔM-distribution from non-censored clusters).

4.1.2 Model Formulation and Software

In order to estimate both covariate effects and the entire distribution of magnitude differences ΔM, we need a fully parametric survival model approach. As will be shown in the results section, the best model fits were achieved assuming a *Gompertz* distribution for the magnitude differences, rather than other candidates such as Weibull or Generalized Gamma. The Gompertz distribution is defined on $(0, \infty)$. Therefore data points with $\Delta M = 0$ were substituted by the value 0.01. In the `R` package *flexsurv* (Jackson, 2016), the Gompertz distribution is parameterized by its probability density function

$$f(x|a, b) = be^{ax} \exp\left(-\frac{b}{a}(e^{ax} - 1)\right)$$

with shape parameter $a \in \mathbb{R}$ and scale parameter $b > 0$. Besides the categorical plate boundary class, we modeled the effects of the mainshock magnitude $\mathbf{x}_{mag}$ and depth $\mathbf{x}_{depth}$, as well as the locally interpolated relative plate velocity $\mathbf{x}_{veloc}$, heat flow $\mathbf{x}_{heat}$ and sea floor age $\mathbf{x}_{age}$. In the resulting full Gompertz survival model, we regressed the scale parameter b through all covariates for observation i by

$$\begin{aligned} log(b(\mathbf{x_i})) =& \beta_0 + \beta_1 \, \mathbf{x}_{class=CCB,i} \; + ... + \beta_6 \, \mathbf{x}_{class=OTF,i} \; + \\ & f_{mag}(\mathbf{x}_{mag,i}) + f_{depth}(\mathbf{x}_{depth,i}) + f_{veloc}(\mathbf{x}_{veloc,i}) + f_{heat}(\mathbf{x}_{heat,i}) + f_{age}(\mathbf{x}_{age,i}), \end{aligned}$$

where $\beta_0, \beta_1, ..., \beta_6$ are the coefficients related to categorical variables, where boundary class "SUB" is the reference category, represented by the intercept β_0, and the f terms denote coeffiicients related to categorical variables. Similarly, we modeled the shape parameter a depending on the linear effects of the plate boundary class, i.e.

$$log(a(\mathbf{x})) = \alpha_0 + \alpha_1 \, \mathbf{x}_{class=CCB} \; + ... + \alpha_6 \, \mathbf{x}_{class=OTF}.$$

In this work, we fitted models using the function *flexsurvreg* from the *flexsurv* package, which estimates parameters by optimizing a parametric likelihood adapted for censored data (Jackson, 2016). To allow for flexible non-linear effects, all metric variables are modeled by the penalized spline function *pspline* from the `R` package *survival* (Therneau, 2016), consistenly using $df = 2$ degrees of freedom and $n = 2.5 \times df$ basis functions (Eilers and Marx, 1996; Hurvich et al., 1998).

4.2 Generalized Additive Count Models

In the second part of this study, we model the number of aftershocks N_i of each event i. The starting point for modeling count data response variables are so-called generalized linear models. Given covariate values $\mathbf{x}_{i1}, ..., \mathbf{x}_{ik}$, the expected aftershock productivity of event i is modeled by the log-linear relationship

$$\mathbb{E}[N_i] = \exp(\eta_i) \tag{5}$$

where $\eta_i = \beta_0 + \beta_1 \mathbf{x}_{i1} + ... + \beta_k \mathbf{x}_{ik}$ is the linear predictor and $\beta_0, \beta_1, ..., \beta_k$ are the estimated coefficients. Note that the covariates have an *exponentially multiplicative effect* on the expected number of aftershocks (see e.g. Fahrmeir et al., 2013, section 5.2).

In this work, we used a GAM approach and replaced the linear effects of all metric covariates by potentially smooth functions that can more flexibly represent varying effects in different value ranges of the covariates (e.g., see Fahrmeir et al., 2013, section 9.1). The full model is then specified by the predictor

$$\begin{aligned}\eta_i =& \beta_0 + \sum_{k=1}^{5} \beta_k \, I(\mathbf{x}_{TR,i} = k) + \sum_{h=6}^{12} \beta_h \, I(\mathbf{x}_{SL,i} = h) + ... \\ & f_{mag}(\mathbf{x}_{mag,i}) + f_{depth}(\mathbf{x}_{depth,i}) + ... \\ & f_{mainshMag}(\mathbf{x}_{mainshMag,i}) \, I(\mathbf{x}_{isBackground,i} = false),\end{aligned} \tag{6}$$

where the β's are the estimated coefficients for the linear categorical effects of the tectonic region (TR, $k = 1, ..., 5$) and slip type (SL, $h = 6, ..., 12$), and β_0 is the intercept representing the reference categories *TR=crustal outside* and *SL=unknown*. The functions f_{mag}, f_{depth} and $f_{mainshMag}$ represent the smooth effects of the magnitude and depth of the triggering event as well as of the prior mainshock magnitude, and $I(...)$ is the indicator function that is 1 if the inside condition is fulfilled, and 0 otherwise.

To fit the model, we used the function *gam* from the `R` package *mgcv* (Wood, 2017), using a logarithmic link function and the restricted maximum likelihood estimator (REML) as the smoothing parameter estimation method. Penalized splines based on a basic spline basis (P-Splines) were used to model the unspecified smooth functions (e.g., see Fahrmeir et al., 2013, section 8.1). We used the standard smooth term function of *mgcv*, choosing $k = 5$ and $k = 8$ (for depth) as the dimensions of the basis.

5 Results of the ΔM-Regression

In this section, we show and discuss the results of a parametric survival model fitted to the global declustered earthquake catalog in order to describe the magnitude difference ΔM between the mainshock and the second strongest event in the cluster. First, we justify and validate the distribution assumption for the response variable. Then we show and interpret the effects of the modeled covariates. Finally, we assess the explanatory power of the model using a response residual plot.

5.1 Choice of Distribution Family

Following the simple simulation model outlined in the section *Introduction*, with parameters $b = 1$, $A = 0.13$ and $\alpha = 2$, we fitted a Weibull, a Gompertz and a Generalized Gamma distribution to the simulated magnitude differences ΔM. Fig. 4(a) shows the fits of the three distributions to the

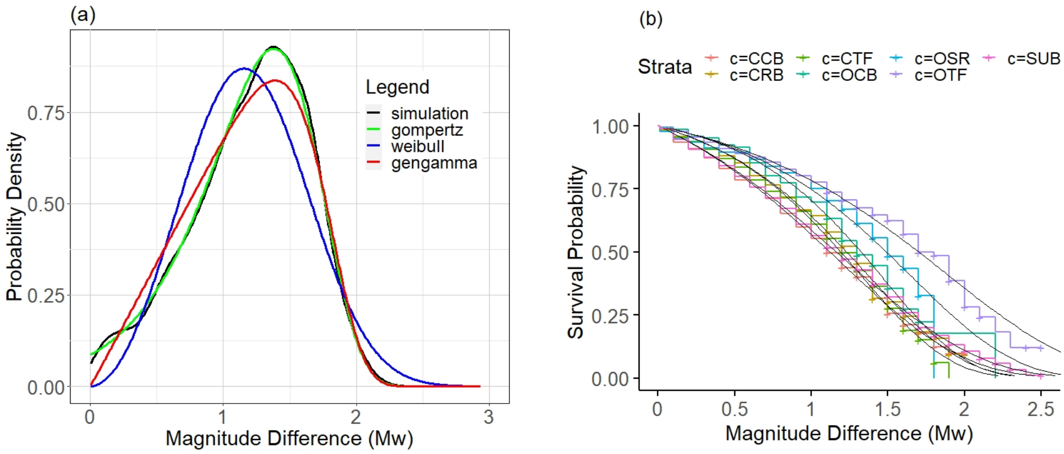


Fig 4: (a) Fits of a Gompertz, Weibull and Generalized Gamma distribution to simulated magnitude differences ΔM, represented by the kernel density estimator (black curve). (b) Comparison of survival curves estimated from a Gompertz model and a non-parametric Kaplan-Meier estimator, stratified for plate boundary classes (c).

kernel density estimator of the sampled data. The Gompertz distribution clearly provides the best fit to the moderately negatively-skewed data.

In order to confirm this assumption based on the actual dataset, we fitted a Gompertz survival model with only the scale parameter depending on the categorical plate boundary class, and compared the predicted survival curves to those provided by the non-parametric Kaplan-Meier estimator, which does not require a specific distribution assumption (Klein and Moeschberger, 2003, ch. 4). In Fig. 4(b), the step functions colored according to the seven boundary classes, refer to the Kaplan-Meier estimates. The Gompertz survival model survival curves are plotted on top by black lines, showing generally good agreement.

5.2 Covariate Effects

Fig. 5 shows the covariate effects for the full parametric Gompertz survival model. The categorical effects in Fig. 5(a) represent predictions of the response ΔM given the various boundary classes, if the other covariates are held fixed at their median values (magnitude=6.4, depth=23 km, velocity=66.5 mm/a, sea floor age=220 Ma, heat flow $\approx 0.07\ Wm^{-2}$). The effects of the metric covariates in Fig. 5(b-f) are similarly predicted for a fine grid of values of the considered variable, holding the other covariates fixed and assuming a subducting environment (i.e., boundary class *"SUB"*). Gray shades represent the 95% confidence interval.

5.2.1 Effect of Boundary Class

Fig. 5(a) reveals no structural effects of specific boundary classes. If we were fitting the same model, but leaving out sea floor age and heat flow, the boundary classes *OSR* and *OTF* would show a substantial and *OCB* a moderate increase in magnitude differences. In other words, mainshocks at oceanic, especially transform and divergent type boundaries, produce weaker second strongest events than those in continental zones, which fits with the generally limited magnitude sizes in

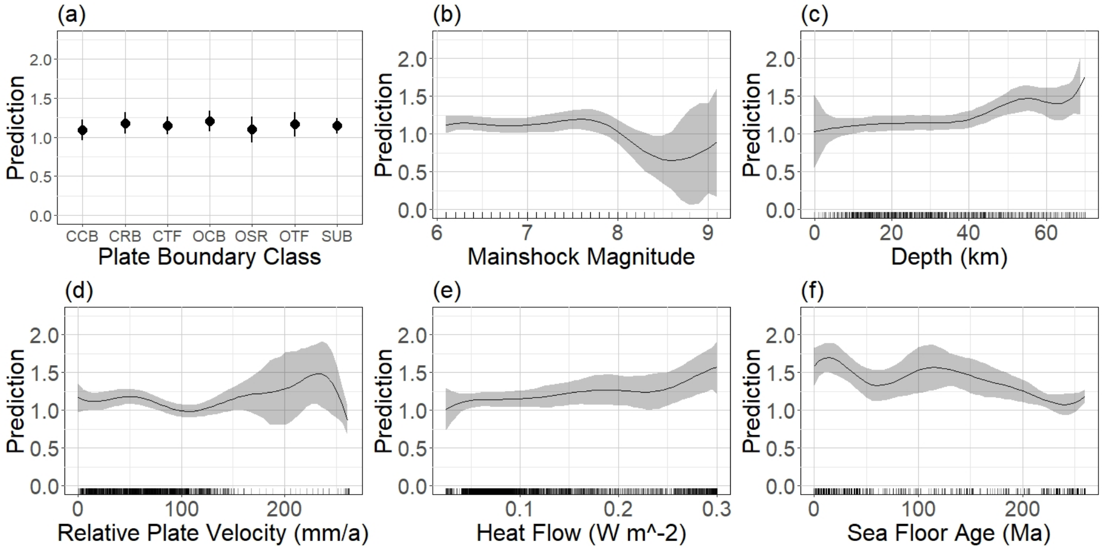


Fig 5: Covariate effects of the ΔM-Regression, by (a) plate boundary type (categorical), (b) mainshock magnitude, (c) mainshock depth, (d) relative plate velocity, (e) heat flow and (f) sea floor age on the magnitude difference between a mainshock and the second largest event of the cluster. For linear effects (a), 95% confidence intervals are represented by bars. For smooth effects (b-f), 95% confidence intervals are depicted by gray shades. The effects are computed as predictions of the response variable, fixing the other variables at their median values. Rug lines on the x axis visualize the marginal distributions of the corresponding metric covariate.

these two boundary classes (Bird et al., 2002; Boettcher and Jordan, 2004). However, this effect seems to be sufficiently represented by the added metric covariates.

5.2.2 Effect of Mainshock Magnitude

For values smaller than $M = 7.8$, the mainshock magnitude effect depicted in Fig. 5(b) confirms the well-established Bath's law hypothesis that the average magnitude difference ΔM is roughly 1.2, *independently* of the mainshock magnitude. For larger magnitudes, there seems to be a tendency toward smaller ΔM.
However, this effect is very uncertain for two reasons. First, the sample size of $M > 7.8$ events (41 data points) is very small compared to the lower magnitude ranges, leading to large standard errors. Second, the mainshock magnitude controls the radius of the spatial window in the declustering approach. Thus, larger mainshocks span an exponentially increasing area, in which potential aftershocks are searched. To test, whether the observed effect of strong mainshocks may be an artifact of a too generous choice of the spatial window radius, we repeated the study for an event set declustered with radius $R(m) = K(m)\,L(m)$, where the factor $K(m)$ gradually decreases from 2.5 to 1.0 for magnitudes between 6.0 and 9.0. This sensitivity study confirmed the shape of the effect curve, indicating that the second strongest event usually occurred relatively close to the mainshock.

5.2.3 *Effect of Mainshock Depth*

Fig. 5(c) shows that the effect of the mainshock depth is almost constant for depths smaller than 40 km. Between 40km und 50km, ΔM increases from roughly 1.2 to a new level of approximately 1.5. This effect is consistent with the observation of Hainzl et al. (2019), who showed that aftershock productivity decreases at higher depths due to reduced seismic coupling, i.e. the energy discharges increasingly through seismic creep rather than through aftershocks. Given a constant magnitude size distribution, this would immediately translate into higher magnitude differences ΔM. A connection with missing data at greater depths is unlikely, as we are only interested in the largest aftershock rather than the entire sequence.

5.2.4 *Effect of Relative Plate Velocity*

Plate velocities play an important role for the duration of stress re-accumulation at a fault after the occurrence of a large earthquake. However, recurrence intervals of so-called *characteristic earthquakes* are typically in the range of multiple decades or even centuries. For the short-term recurrence of strong aftershocks, Fig. 5(d) reveals no clear effect of the relative plate velocity. As an alternative covariate representing the velocity of deformation in the tectonic system, we tested global strain rate data (Kreemer et al., 2014), which similarly showed no structural effect.

5.3 *Effect of Heat Flow*

According to Fig. 5(e), regions with heat flow larger than $0.23W/m^{-2}$ show a substantial increase of magnitude differences. Warmer rock is known to be more viscous, which discharges stress through seismic creep rather than abrupt fractures, leading to the same aftershock productivity argument as for higher depths. As Fig. 2 shows, high heat flow values are typically prevalent in oceanic ridges and transform faults, which explains why the model predicts larger ΔM for the plate boundary classes *OSR* and *OTF* if heat flow is left out as a covariate.

5.3.1 *Effect of Sea Floor Age*

Fig. 5(f) shows that magnitude differences are substantially larger in young compared to old oceanic crusts. A potential causal reason for the effect of the plate age cannot be ruled out, but is unknown to the authors. Note that young sea floor typically comes with large heat flows. Therefore, the effects of the two variables are consistent. As new oceanic crust is formed at oceanic ridges, the effect also coincides with the increased magnitude differences in the nested model without sea floor age.

If we fit the full model to the subset of subduction zone mainshocks only, both heat flow and sea floor age show no clear signal. Thus, it is likely that their effect is mainly driven by their tails at oceanic ridges.

5.4 *Response Residuals*

Fig. 6 shows the response residuals (i.e., observed minus predicted values) plotted against predicted magnitude differences. Note that, as observations are censored, residuals are censored as well. Therefore, we show only residuals for *non-censored observations* here. This explains the superiority of negative residuals. The blue line represents the linear trend of the residuals.

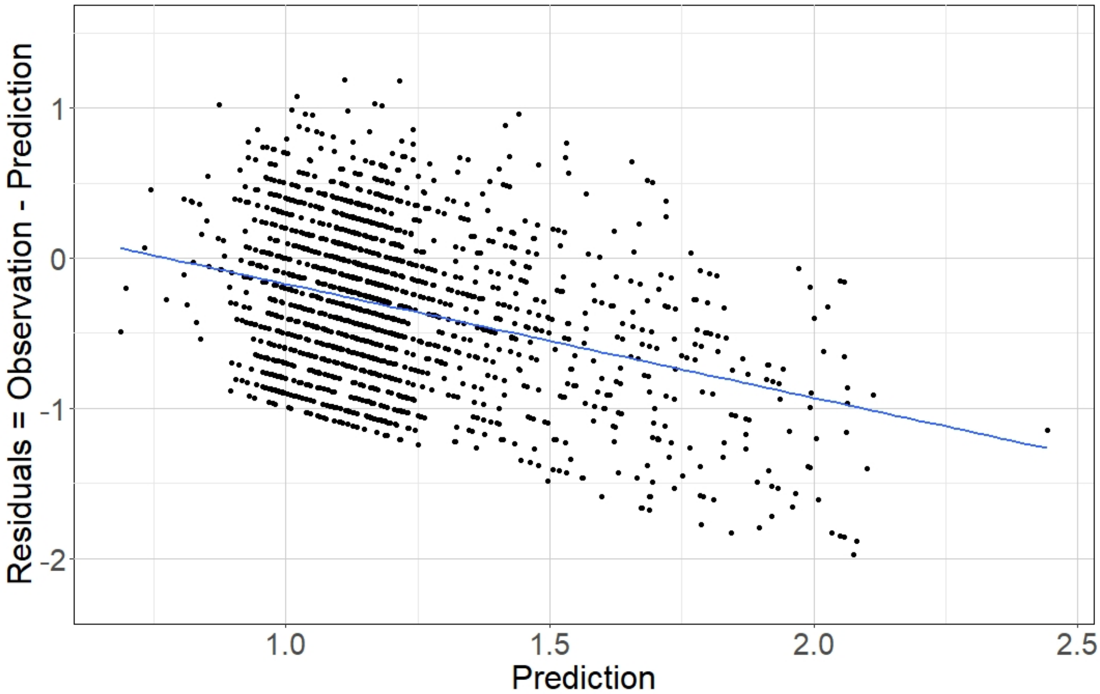


Fig 6: Response Residuals of the ΔM-regression for non-censored observations only, plotted against predicted values. The blue line represents the linear trend of the residuals. The row arrangement of the points is due to the rounding of the observed data to one decimal place. For instance, the bottom row represents observations where $\Delta M = 0$.

The large variation of the residuals suggests a weak predictive power of the model. Residuals of more than one magnitude unit are not rare, and can even reach up to almost two units. Small observations are typically substantially overestimated, and vice versa. The root mean square error for predictions by the full model, 0.62, is only minimally better than by a Gompertz intercept model, 0.63. However, these values only account for predictions of non-censored observations. The majority of substantial covariate effects identified above explain *increases* of the expected magnitude difference, which means that related observations (e.g. events with larger depth or heat flow, or at younger sea floors) are considerably more likely to be censored and therefore left out of the residuals statistics.

A similar argument holds for the negative linear trend of the residuals. For instance, the lowest mainshock magnitude, $M = 6.1$, can only have observed magnitude differences up to $\Delta M_{maxObs} = M - M_c = 1.1$. Therefore, if the model predicts $\Delta M > 1.1$, only negative residuals will occur in the statistic. As we move to larger predicted values on the x-axis, the selection bias affects even larger mainshock magnitudes.

The censoring of observations and residuals hinder a rigorous diagnosis of the model. Despite the covariates showing some relevant signals, it is evident that the model misses additional high-resolution geophysical variables for local site effects or event-specific properties that can help explain a larger proportion of the variance in the data.

5.5 Sensitivity Studies

As partly mentioned above, we tested the influence of varying time windows (e.g. $T = 365$ days) and spatial windows (e.g. $R(m) = L(m)$ or $R(m) = K(m)\,L(m)$ with gradually decreasing $K(m)$ as described above) in the declustering approach on the regression results. The covariate effects are very insensitive, indicating that the second strongest event typically occurs close to and shortly after (or before) the mainshock. In other words, the contamination of the response variable through background seismicity is negligible.

6 Results of Productivity-Regression

In this section, we present the results of the GAM regression of aftershock count data in New Zealand. First, we justify and validate our choice of the Negative Binomial distribution instead of the commonly used Poisson distribution for aftershock counts. Then we show and interpret the effects of the modeled covariates. Finally, we illustrate the impact of the results in a simulation experiment.

6.1 Choice of Distribution Family

Fig. 7(a) and (b) show the quantile-quantile (Q-Q) plots of the deviance residuals for the Poisson distribution (a) and the Negative Binomial distribution (b). The latter fits the data better, as it adjusts the variance independently of the mean and therefore allows for larger variation than the Poisson distribution, resulting in a more adequate representation of the upper and lower tails of the distribution. The dispersion parameter of the Negative Binomial fit is 2.3.

In Fig. 7(c) and (d), the corresponding model residuals are plotted against the linear predictors η_i (see Equation 6) for the Poisson (c) and Negative Binomial distribution (d). The Negative Binomial fit shows substantially less spread in the residuals than the Poisson fit, confirming that aftershock count data is rather Negative Binomial distributed.

Note that, according to our simplified simulation model, a larger variance of aftershock counts directly translates into a larger variance of ΔM. In other words, the Negative Binomial distribution increases the likelihood of particularly small ΔM.

Alternative approaches such as a Quasi Poisson or a zero-inflated model were tested, but did not stand out substantially from the respective basic models. Additionally, from a substantive point of view, there seems to be no causal reason for ”excess zeros” that would suggest the use of zero-inflated approaches.

6.2 Covariate Effects

Fig. 8 and 9 present the exponential, multiplicative effects of the categorical and metrical covariates on the expected number of aftershocks according to the relationship in Equation (5). That is, if the exponential effect of a category is larger than 1, it has a positive impact on aftershock counts, and vice-versa. If the exponential effect is equal to 1, the model shows no effect of the respective category.

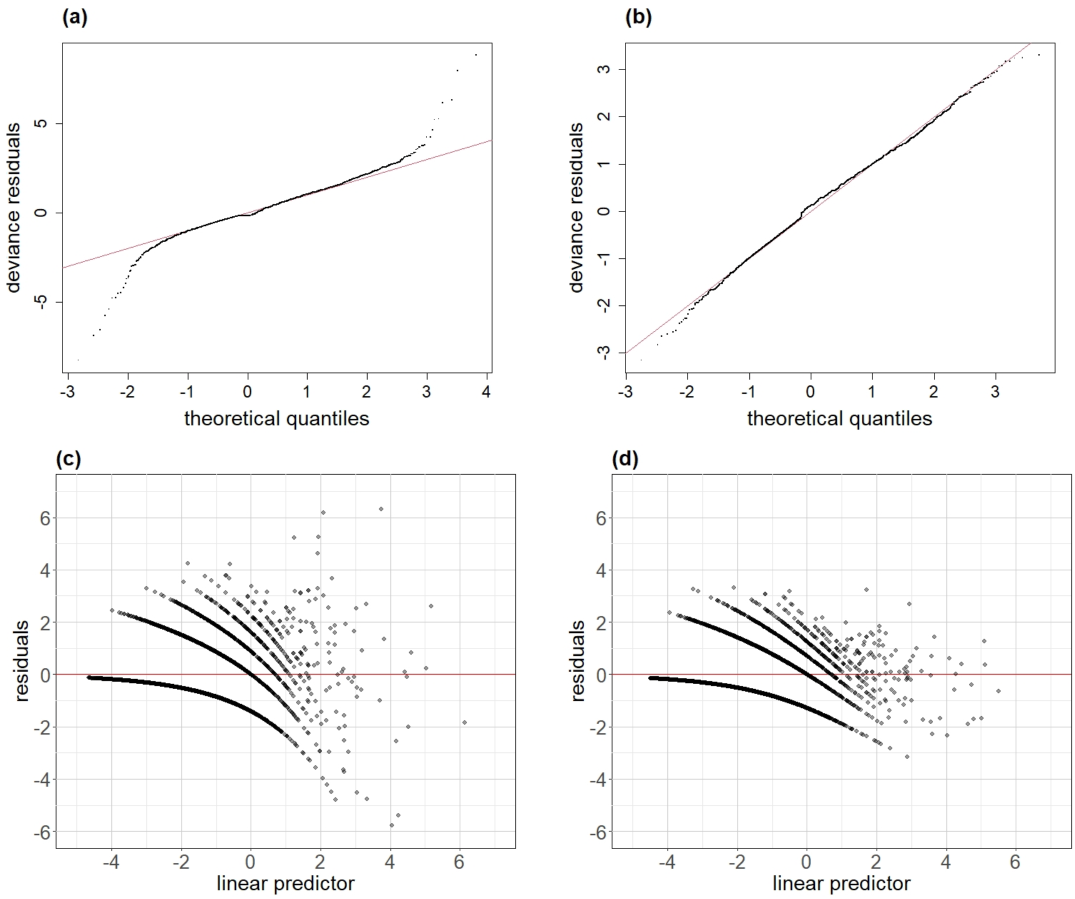


Fig 7: *Top row:* Quantile-Quantile plots of the deviance residuals for the (a) Poisson and the (b) Negative Binomial regression of the aftershock productivity. *Bottom row:* Corresponding model residuals plotted against the linear predictors η_i (see Equation 6) for the (c) Poisson and (d) Negative Binomial regression. The row arrangement of points is due to the count data structure of the response.

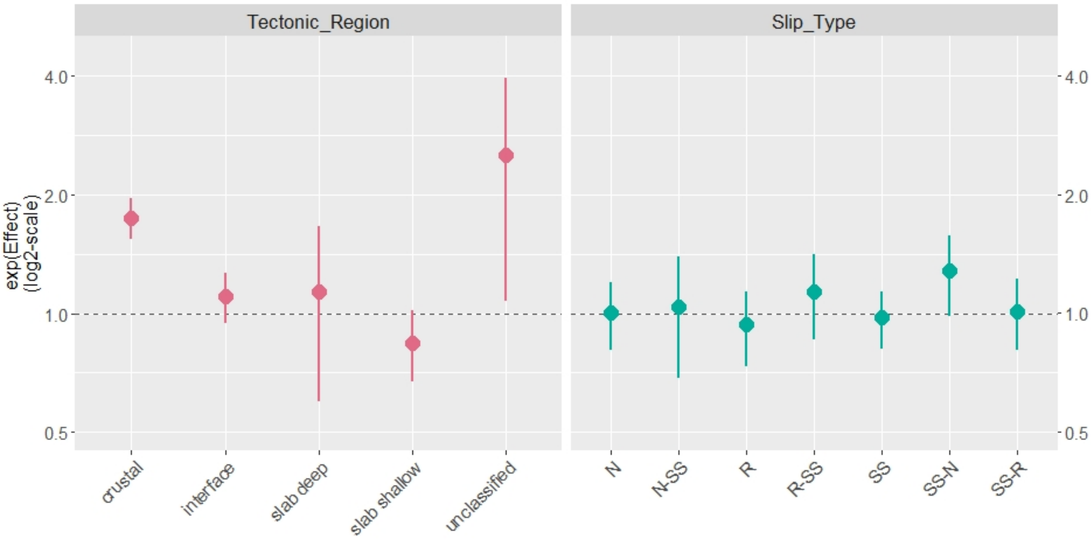


Fig 8: Exponential, multiplicative effects of the categorical covariates *tectonic region* and *slip type* relative to their reference categories ”crustal outside” and ”unkown”, respectively, according to Equation (5). Exponential effects larger than one signify a positive effect on aftershock productivity, and vice versa.

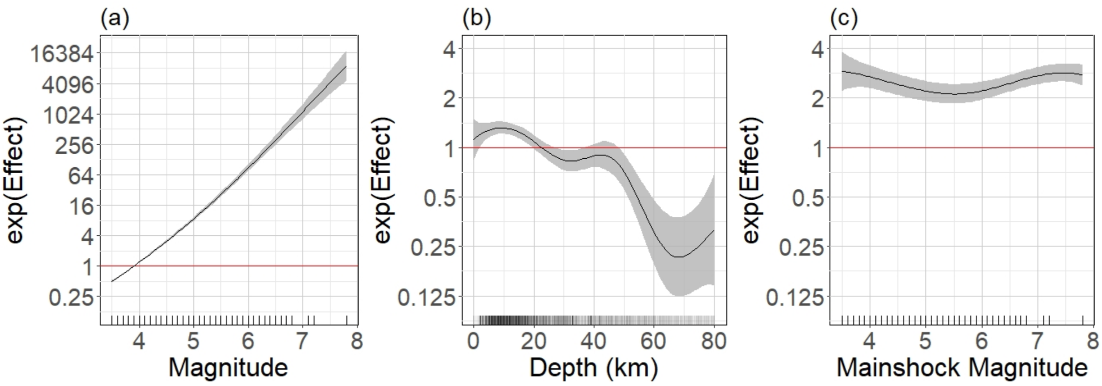


Fig 9: Exponential, multiplicative effects of the metric covariates *magnitude* and *depth* of the triggering event as well as *mainshock magnitude*, given that the triggering event was already part of a triggered sequence. Exponential effects larger than one signify a positive effect on aftershock productivity, and vice versa. Rug lines on the x axis visualize the marginal distributions of the corresponding covariate.

6.2.1 Effects of Categorical Variables

Fig. 8 shows the effects of the categorical covariates *Tectonic Region* and *Slip Type*.
The effects of the tectonic region are presented relative to their reference category *crustal outside*. Crustal events in the subduction zone show a substantially increased aftershock productivity. The expected number of aftershocks is approximately 1.8 times larger than for crustal events outside of the subduction zone. For interface and slab events, both shallow and deep, no clear signal is found, as their confidence intervals overlap with the reference line at $\exp(0) = 1$. Unclassified events appear to have a slightly positive effect, but the uncertainty is large. Sensitivity tests with different buffer sizes of the slab do not consistently confirm the effect for unclassified events. The positive effect of crustal events on the aftershock productivity, compared to interface, might be explained by reduced seismic coupling in subduction zones (Hainzl et al., 2019). On the interface, a substantial part of the deformation is often aseismic (Lay et al., 2012).
A second possibility is that, in the proximity of the study region, the crust contains a dense network of faults with a wide range of orientations, and therefore more structures that could be brought closer to failure by a change in stress conditions due to mainshock earthquakes. Slip type effects are depicted relatively to their reference category *unknown*. None of the focal mechanisms appears to be substantially more or less productive. Additionally, there seems to be no *selection effect* in the sense that events where the focal mechanism is known have a common effect on the aftershock productivity.

6.2.2 Effects of Magnitude and Depth

Fig. 9(a) shows an exponential effect of the triggering magnitude on aftershock productivity. However, this effect is enforced by the declustering approach, as the ETAS model fits the exponential aftershock productivity function (4) to optimize the aftershock trigger rates $R^{trig}(t, x, y, i)$. Therefore, the effect only has control character.
The effect of the depth, shown in Fig. 9(b), confirms the argumentation in the discussion of the ΔM regression results, that increasing magnitude differences may coincide with reduced aftershock productivity at higher depths. The physical reason may again be reduced seismic coupling along the subduction interface relative to the shallow crust (Hainzl et al., 2019; Lay et al., 2012).

6.2.3 Effect of Mainshock Magnitude

Fig. 9(c) shows no clear trend in the effect of varying mainshock magnitudes. However, independently of the size of the mainshock, triggered events in general appear to be two to three times more productive than a comparable background event.
This finding has two possible explanations. On the one hand, it may be an indicator that the ETAS model based declustering approach does not adequately disentangle spatio-temporal clusters in the catalog, and incorrectly assigns too much aftershock productivity to the smaller events, at the cost of the mainshock productivity. In other words, in seismically active periods, small events may be simply more likely to be assigned offsprings than in seismically quiet times, but in reality, many of these aftershocks should perhaps be direct rather than secondary ones of the mainshock. Such a rearrangement of trigger relationships would have a strongly distorting effect on our model, in which we consider the number of *direct* aftershocks. Note, however, that we used here an ETAS model approach that accounts for short-term aftershock incompleteness as well as locally restricted, anisotropic spatial kernels and therefore already improves some of the major biases of

common ETAS models (de Arcangelis et al., 2018; Grimm et al., 2022, 2021; Hainzl et al., 2008; Hainzl, 2021; Seif et al., 2017).
On the other hand, Zhuang et al. (2004) proposed that triggered events are more productive than background events, based on a similar study. It seems reasonable that during an on-going sequence the aftershock productivity could temporarily increase due to a higher level of energy prevalent in the tectonic system, compared to seismically quiet periods with occasional background activity. A doubling of the productivity parameter A in the simulation model (see Introduction), applied only to secondary triggering, led to a reduction of the expected magnitude difference ΔM from 1.2 to below 0.9 due to the increasing cluster sizes. This additional "boost" in triggering illustrates the relevance of the observed effect. The finding may also contribute to an explanation as to why the ETAS model tends to underestimate cluster sizes and doublet probabilities in forward simulations, as observed in my first contribution (Grimm et al., 2021). Further research is recommended to evaluate this finding.

7 Conclusions

We adapted a survival regression model approach from medical studies to estimate the parametric distribution of the magnitude difference ΔM between the mainshock and its strongest foreshock or aftershock. The highlight of this regression class is that it accounts for right-censored observations. In our case, these are mainshocks for which no aftershock or foreshock is recorded above cut-off magnitude M_c, and for which we therefore have only the partial information that $\Delta M > M - M_c$, where M is the mainshock magnitude.
We declustered a global earthquake catalog using a window method and computed ΔM for each of the independent clusters. Then, we enriched the cluster dataset with a plate boundary classification, relative plate velocities and sea floor ages obtained from the digital plate boundary model of Bird (2003) and with heat flow from Bird et al. (2008). From a simplified simulation model, assuming an exponential aftershock productivity law and the Gutenberg-Richter type magnitude size distribution, we concluded that the Gompertz distribution may be the better choice than Weibull or Generalized Gamma.
The regression results show that larger ΔM are expected at higher depths and in younger ocean crust. This may be an indication, that aftershock productivity is a relevant driver of ΔM, as in these conditions lower aftershock productivity is expected due to reduced seismic coupling (Hainzl et al., 2019).
In the second part of this study, we used the stochastic declustering method of Zhuang et al. (2002) to estimate the aftershock productivity per event in a local catalog for New Zealand. To do so, we used the anisotropic ETAS-Incomplete model (Grimm et al., 2022) to disentangle the trigger relations between events. We further enriched the event set by a categorization of events in tectonic regions and slip types and used a generalized additive regression approach to model the aftershock productivity.
The results clearly confirm that aftershock counts follow a Negative Binomial rather than a Poisson distribution (Kagan, 2017; Shebalin et al., 2018). Also, aftershock productivity decreases with increasing depth, supporting the reasoning regarding the depth effect on ΔM above. Furthermore, results indicate that triggered earthquakes trigger themselves two to three times more aftershocks than non-triggered ones. In other words, secondary aftershock triggering is substantially stronger than direct triggering by background events. This effect may either be an indicator for wrongly

disentangled sequences in the sense that secondary triggering is overestimated at the cost of the productivity of the mainshock, or it may represent an actual effect due to the temporarily higher energy level after the occurrence of a strong mainshock. A causal effect, if confirmed, would have an enormous impact on the expected cluster sizes (compare with Grimm et al., 2021) and could explain some of the rather small ΔM observations in the first study.
Future research should identify whether small magnitude differences ΔM are typically characterized rather by above-average aftershock productivity or by magnitude size distributions favoring large aftershocks. To do so, one could compile a sufficiently large set of earthquake sequences and analyze the correlation of their ΔM with estimates of the aftershock productivity (Equation 4) and frequency-magnitude distribution (Equation 2). Additionally, it should be verified whether triggered events indeed have a larger aftershock productivity, and how this effect impacts ΔM. Similarly, potential correlations of aftershock magnitudes with their ancestors should be evaluated. Finally, an extension of the ΔM-regression model using small-scale covariate data could certainly contribute to a better understanding of magnitude differences in different geophysical settings.

Data and Resources

The U.S. Geological Survey National Earthquake Information Center (USGS-NEIC) catalog has been downloaded from `https://earthquake.usgs.gov/earthquakes/search/` (last accessed on March 30, 2022). Global covariate data has been downloaded from `http://peterbird.name/publications/2003_pb2002/2003_pb2002.htm` (Bird, 2003, last accessed on March 30, 2022) or has been made available by Peter Bird after personal contact (heat flow data, Bird et al., 2008). The New Zealand event set and focal mechanism data was provided by GNS Science as an input to the ongoing 2022 revision of the New Zealand National Seismic Hazard Model. For the stochastic declustering, we used the ETAS-Incomplete model source code available in the Github repository `https://github.com/ChrGrimm/ETASanisotropic` (Grimm et al., 2022), implemented using the software *Matlab*. All statistical analyses were performed with the open source software *R*.

Acknowledgments

We sincerely thank Chris Rollins and Annemarie Christophersen of GNS Science for providing the local catalog for New Zealand, and for Chris' valuable guidance on questions regarding the data. A special thanks goes to Andreas Bender (LMU Munich) for numerous helpful discussions about the use of survival models in our study. The first author thanks all *GEM*-colleagues in Pavia for the fruitful collaboration.

Statements and Declarations

Financial support for this work was provided by Munich Re through a scholarship granted to the first author, and by the Department of Statistics at Ludwig-Maximilians-University Munich. S.H. was supported by the Deutsche Forschungsgemeinschaft (DFG) Collaborative Research Centre 1294 (Data Assimilation – The seamless integration of data and models, project B04) and the European Unions Horizon 2020 research and innovation program under Grant Agreement Number 821115, realtime earthquake risk reduction for a resilient Europe (RISE). The authors acknowledge that there are no relevant financial or non-financial interests to disclose. This article does not contain any studies involving human participants or animals performed by any of the authors.

Contribution of the authors

CG (first author) designed the study, compiled and declustered the datasets, conducted and evaluated the ΔM regression model, prepared figures and wrote the paper. TR conducted and evaluated the productivity regression model. SH, KJ and MK contributed significantly to the interpretation of the model results in the geophysical context. Additionally, KJ categorized New Zealand events into tectonic regions. MP und HK contributed to the project idea and advised on data and methodology.

References

Abdelnaby, A. E. (2012). Multiple earthquake effects on degrading reinforced concrete structures. PhD thesis, University of Illinois at Urbana-Champ.

Álvarez-Gómez, J. A. (2019). FCM - Earthquake focal mechanisms data management, cluster and classification. *SoftwareC*, **9**, 299–307.

Bath, M. (1965). Largest inhomogeneities of the upper mantle. *Tectonophysics*, **2**, 483–514.

Bird, P. (2003). An updated digital model of plate boundaries. *Geochemistry, Geophysics, Geosystems*, **4**(3). doi: 10.1029/2001GC000252.

Bird, P., Kagan, Y. Y., Jackson, D. D., Stein, S., and Freymueller, J. (2002). Plate Tectonics and Earthquake Potential of Spreading Ridges and Oceanic Transform Faults. In *AGU Monograph*, pages 203–218. doi: 10.1029/gd030p0203.

Bird, P., Liu, Z., and Rucker, W. K. (2008). Stresses that drive the plates from below: Definitions, computational path, model optimization, and error analysis. *Journal of Geophysical Research: Solid Earth*, **113**, B11406. doi: 10.1029/2007JB005460.

Boettcher, M. S. and Jordan, T. H. (2004). Earthquake scaling relations for mid-ocean ridge transform faults. *Journal of Geophysical Research: Solid Earth*, **109**(12), 1–21. doi: 10.1029/2004JB003110.

Bondár, I., Engdahl, E. R., Villaseñor, A., Harris, J., and Storchak, D. (2015). ISC-GEM: Global Instrumental Earthquake Catalogue (1900-2009), II. Location and seismicity patterns. *Physics of the Earth and Planetary Interiors*, **239**, 2–13. doi: 10.1016/j.pepi.2014.06.002.

Dascher-Cousineau, K., Brodsky, E. E., Lay, T., and Goebel, T. H. (2020). What controls variations in aftershock productivity? *Journal of Geophysical Research: Solid Earth*, **125**. doi: 10.1029/2019JB018111.

de Arcangelis, L., Godano, C., and Lippiello, E. (2018). The overlap of aftershock coda waves and short-term postseismic forecasting. *Journal of Geophysical Research: Solid Earth*, **123**, 5661–5674. doi: 10.1029/2018JB015518.

Di Giacomo, D., Bondár, I., Storchak, D. A., Engdahl, E. R., Bormann, P., and Harris, J. (2015a). ISC-GEM: global instrumental earthquake catalogue (1900-2009), III. Re-computed ms and mb, proxy mw, final magnitude composition and completeness assessment. *Physics of the Earth and*

Planetary Interiors, **239**, 33–47. doi: 10.1016/j.pepi.2014.06.005. URL http://dx.doi.org/10.1016/j.pepi.2014.06.005.

Di Giacomo, D., Harris, J., Villaseñor, A., Storchak, D. A., Engdahl, E. R., Lee, W. H., Verney, R., Safronova, N., Wylie, R., Baranauskaite, A., Wilson, J., and Simpson, H. (2015b). ISC-GEM: global instrumental earthquake catalogue (1900-2009), I. Data collection from early instrumental seismological bulletins. *Physics of the Earth and Planetary Interiors*, **239**, 14–24. doi: 10.1016/j.pepi.2014.06.003. URL http://dx.doi.org/10.1016/j.pepi.2014.06.003.

Di Giacomo, D., Robert Engdahl, E., and Storchak, D. A. (2018). The ISC-GEM earthquake catalogue (1904-2014): Status after the extension project. *Earth System Science Data*, **10**, 1877–1899. doi: 10.5194/essd-10-1877-2018.

Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, **11**, 89–121.

Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. (2013). *Regression - Models, Methods and Applications*. ISBN 9783642343322. doi: 10.1007/978-3-642-34333-9.

Felzer, K. R., Abercrombie, R. E., and Ekström, G. (2004). A common origin for aftershocks, foreshocks, and multiplets. *Bulletin of the Seismological Society of America*, **94**(1), 88–98. doi: 10.1785/0120030069.

Gardner, J. K. 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.

Grimm, C., Hainzl, S., Käser, M., and Küchenhoff, H. (2022). Solving three major biases of the ETAS model to improve forecasts of the 2019 Ridgecrest sequence. *Stochastic Environmental Research and Risk Assessment*, **2**. doi: 10.1007/s00477-022-02221-2.

Grimm, C., Käser, M., Hainzl, S., Pagani, M., and Küchenhoff, H. (2021). Improving Earthquake Doublet Frequency Predictions by Modified Spatial Trigger Kernels in the Epidemic-Type Aftershock Sequence (ETAS) Model. *Bulletin of the Seismological Society of America*, (Xx). doi: 10.1785/0120210097.

Gulia, L., Rinaldi, A. P., Tormann, T., Vannucci, G., Enescu, B., and Wiemer, S. (2018). The effect of a mainshock on the size distribution of the aftershocks. *Geophysical Research Letters*, **45** (24), 13,277–13,287. doi: 10.1029/2018GL080619.

Gutenberg, B. and Richter, C. F. (1944). Frequency of earthquakes in California. *Bulletin of the Seismological Society of America*, **34**, 185–188. doi: 10.1038/156371a0.

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. doi: 10.1785/0120070256.

Hainzl, S., Sippl, C., and Schurr, B. (2019). Linear Relationship Between Aftershock Productivity and Seismic Coupling in the Northern Chile Subduction Zone. *Journal of Geophysical Research: Solid Earth*, **124**(8), 8726–8738. doi: 10.1029/2019JB017764.

Hainzl, S. (2021). ETAS-Approach Accounting for Short-Term Incompleteness of Earthquake Catalogs. *Bulletin of the Seismological Society of America*. doi: 10.1785/0120210146.

Hainzl, S., Zakharova, O., and Marsan, D. (2013). Impact of aseismic transients on the estimation of aftershock productivity parameters. *Bulletin of the Seismological Society of America*, **103**(3), 1723–1732. doi: 10.1785/0120120247.

Helmstetter, A. and Sornette, D. (2003). Foreshocks explained by cascades of triggered seismicity. *Journal of Geophysical Research: Solid Earth*, **108**(B10). doi: 10.1029/2003jb002409.

Hurvich, C. M., Simonoff, J. S., and Tsal, C.-L. (1998). Smoothing parameter selection in non-parametric regression using an improved Akaike information criterion. *JRSSB*, **60**, 271–293.

Jackson, C. H. (2016). Flexsurv: A platform for parametric survival modeling in R. *Journal of Statistical Software*, **70**(8). doi: 10.18637/jss.v070.i08.

Jalilian, A. (2019). ETAS: An R package for fitting the space-time ETAS model to earthquake data. *Journal of Statistical Software*, **88**(1). doi: 10.18637/jss.v088.c01.

Kagan, Y. Y. and Jackson, D. D. (1999). Worldwide doublets of large shallow earthquakes. *Bulletin of the Seismological Society of America*, **89**(5), 1147–1155.

Kagan, Y. Y. and Jackson, D. D. (2010). Earthquake forecasting in diverse tectonic zones of the Globe. *Pure and Applied Geophysics*, **167**(6), 709–719. doi: 10.1007/s00024-010-0074-4.

Kagan, Y. Y. (2017). Earthquake number forecasts testing. *Geophysical Journal International*, **211**(1), 335–345. doi: 10.1093/gji/ggx300.

Kagermanov, A. and Gee, R. (2019). Cyclic pushover method for seismic assessment under multiple earthquakes. *Earthquake Spectra*, **35**(4), 1541–1558. doi: 10.1193/010518EQS001M.

Klein, J. and Moeschberger, M. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. ISBN 038795399X. doi: 10.1145/390011.808243.

Kreemer, C., Blewitt, G., and Klein, E. C. (2014). A geodetic plate motion and global strain rate model. *Geochemistry, Geophysics, Geosystems*, **15**. doi: 10.1002/2014GC005407.

Lay, T., Kanamori, H., Ammon, C. J., Koper, K. D., Hutko, A. R., Ye, L., Yue, H., and Rushing, T. M. (2012). Depth-varying rupture properties of subduction zone megathrust faults. *Journal of Geophysical Research: Solid Earth*, **117**(4), 1–21. doi: 10.1029/2011JB009133.

Marsan, D. and Helmstetter, A. (2017). How variable is the number of triggered aftershocks? *Journal of Geophysical Research: Solid Earth*, **122**(7), 5544–5560. doi: 10.1002/2016JB013807.

Michael, A. J. (2014). How complete is the ISC-GEM global earthquake catalog? *Bulletin of the Seismological Society of America*, **104**(4), 1829–1837. doi: 10.1785/0120130227.

Nandan, S., Ouillon, G., and Sornette, D. (2019). Magnitude of earthquakes controls the size distribution of their triggered events. *Journal of Geophysical Research: Solid Earth*, **124**(3), 2762–2780. doi: 10.1029/2018JB017118.

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. doi: 10.5047/eps.2010.09.001.

Ogata, Y. and Zhuang, J. (2006). Space-time ETAS models and an improved extension. *Tectonophysics*, **413**(1-2), 13–23. doi: 10.1016/j.tecto.2005.10.016.

Omori, F. (1895). On the aftershocks of earthquakes. *Journal of the College of Science, Imperial University of Tokyo*, **7**, 111–200.

Pagani, M., Johnson, K., and Garcia-Pelaez, J. (2020). Modelling subduction sources for probabilistic seismic hazard analysis. *Geological Society, London, Special Publications 501*.

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. doi: 10.1785/0120160073.

Papadopoulos, A. N., Bazzurro, P., and Marzocchi, W. (2020). Exploring probabilistic seismic risk assessment accounting for seismicity clustering and damage accumulation: Part I. Hazard analysis. *Earthquake Spectra*. doi: 10.1177/8755293020957338.

Pasyanos, M., Masters, T., Laske, G., and Ma, Z. (2014). LITHO1. 0: An updated crust and lithospheric model of the Earth. *Journal of Geophysical Research: Solid Earth*, **119**(3), 2153–2173.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `https://www.R-project.org/`.

Reasenberg, P. (1985). Second-order moment of central California seismicity, 1969–1982. *Journal of Geophysical Research: Solid Earth*, **90**, 5479–5495.

Reyners, M., Eberhart-Phillips, D., and Ristau, J. (2010). Final report on EQC Project BIE 08/549–Quantifying the hazard from New Zealand's shallow intraslab earthquakes. Technical report.

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. doi: 10.1002/2016JB012809.

Shebalin, P. N., Baranov, S. V., and Dzeboev, B. A. (2018). The law of the repeatability of the number of aftershocks. *Doklady Earth Sciences*, **481**(1), 963–966. doi: 10.1134/S1028334X18070280.

Stepp, J. C. (1972). Analysis of completeness of the earthquake sample in the Puget Sound area and its effect on statistical estimates of earthquake hazard. In *Proc. of the 1st Int. Conf. on Microzonazion*, volume 2, pages 897–910.

Storchak, D. A., Di Giacomo, D., Engdahl, E. R., Harris, J., Bondár, I., Lee, W. H., Bormann, P., and Villaseñor, A. (2015). The ISC-GEM global instrumental earthquake catalogue (1900-2009): Introduction. *Physics of the Earth and Planetary Interiors*, **239**, 48–63. doi: 10.1016/j.pepi.2014.06.009. URL `http://dx.doi.org/10.1016/j.pepi.2014.06.009`.

Tahir, M., Grasso, J. R., and Amorse, D. (2012). The largest aftershock: How strong, how far away, how delayed? *Geophysical Research Letters*, **39**(4), 1–5. doi: 10.1029/2011GL050604.

Therneau, T. (2016). survival: Survival Analysis, R package version 2.39-3.

Uhrhammer, R. A. (1986). Characteristics of northern and central California seismicity. *Earthquake Notes*, **57**(1), 21.

Utsu, T., Ogata, Y., and Matsu'ura, R. S. (1995). The centenary of the Omori formula for a decay law of aftershock activity. *J. Phys. Earth*, **43**, 1–33.

van Stiphout, T., Zhuang, J., and Marsan, D. (2012). Seismicity declustering. *Community Online Resource for Statistical Seismicity Analysis*. doi: 10.5078/corssa-52382934. URL `http://www.corssa.org/articles/themev/van{_}stiphout{_}et{_}al/vanstiphoutetal2012.pdf`.

Wells, D. L. and Coppersmith, K. J. (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacements. *Bulletin of the Seismological Society of America*, **84**(4), 974–1002.

Wetzler, N., Brodsky, E. E., and Lay, T. (2016). Regional and stress drop effects on aftershock productivity of large megathrust earthquakes. *Geophysical Research Letters*, **43**(23), 12,012–12,020. doi: 10.1002/2016GL071104.

Wood, S. (2017). Generalized Additive Models: An Introduction with R. *Boca Raton, FL: CRC Press.*

Zakharova, O., Hainzl, S., and Bach, C. (2013). Seismic moment ratio of aftershocks with respect to main shocks. *Journal of Geophysical Research: Solid Earth*, **118**(11), 5856–5864. doi: 10.1002/2013JB010191.

Zhang, L., Werner, M. J., and Goda, K. (2018). Spatiotemporal seismic hazard and risk assessment of aftershocks of M 9 megathrust earthquakes. *Bulletin of the Seismological Society of America*, **108**(6), 3313–3335. doi: 10.1785/0120180126.

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. doi: 10.1198/016214502760046925.

Zhuang, J., Ogata, Y., and Vere-Jones, D. (2004). Analyzing earthquake clustering features by using stochastic reconstruction. *Journal of Geophysical Research: Solid Earth*, **109**, B05301. doi: 10.1029/2003JB002879.

Contact to authors

Christian Grimm (main author), *Ludwig-Maximilians-University Munich, Department of Statistics, Ludwigstraße 33, 80539 Munich, Germany; Christian.Grimm@stat.uni-muenchen.de*,
Teresa Rupprecht, *Ludwig-Maximilians-University Munich, Department of Statistics, Ludwigstraße 33, 80539 Munich, Germany*,
Kendra Johnson, *Global Earthquake Model Foundation, Via Ferrata 1, 27100 Pavia, Italy*,
Sebastian Hainzl, *GFZ German Research Centre for Geoscience, Physics of Earthquakes and Volcanoes, Helmholtzstraße 6/7, 14467 Potsdam, Germany*,
Helmut Küchenhoff, *Ludwig-Maximilians-University Munich, Department of Statistics, Ludwigstraße 33, 80539 Munich, Germany*
Martin Käser, *Ludwig-Maximilians-University Munich, Department of Earth and Environmental Sciences, Geophysics, Theresienstraße 41, 80333 Munich, Germany*, also at *Munich Re, Section GeoRisks, Königinstr. 107, 80802 Munich, Germany*,
Marco Pagani, *Global Earthquake Model Foundation, Via Ferrata 1, 27100 Pavia, Italy*

Part IV.

Appendix

8. ETAS Formulary - Derivations for Gradient-Based Optimization

Non-Published Manuscript

This document gives a comprehensive summary of formulas and derivations of all partial derivatives needed to implement a gradient-based optimization process for the conventional ETAS model (e.g. Jalilian, 2019) or, in particular, for the ETAS-Anisotropic and ETAS-Incomplete model versions proposed in Grimm et al. (2021) and Grimm et al. (2022a), respectively.

It has not been published, but is uploaded as part of the instructions for the ETAS model source code in the Github repository `https://github.com/ChrGrimm/ETASanisotropic.`

Author contributions

All work has been done by Christian Grimm

ETAS Formulary - Derivations for Gradient-Based Optimization

Christian Grimm
(Christian.Grimm@stat.uni-muenchen.de)

May 16, 2022

Abstract

This document gives a comprehensive summary of formulas and derivations of all partial derivatives needed to implement a gradient-based optimization process for the conventional ETAS model (e.g. Jalilian, 2019) or, in particular, for the ETAS-Anisotropic and ETAS-Incomplete model versions proposed in Grimm et al. (2021) and Grimm et al. (2022), respectively.
Chapter 1 introduces the general notation of ETAS models. Then, Chapter 2 derives the log-likelihood functions for the corresponding model versions. Finally, Chapter 3 and 4 rigorously derive the derivatives of the log-likelihood function by the corresponding model parameters, as needed for the implementation of a gradient-based optimization method.
Note that I use the terms *conventional* and *standard* ETAS model synonymously.

Contents

Chapter 1

Definition and Notation of ETAS Models

In this chapter, two versions of the Epidemic Type Aftershock Sequence (ETAS) model are introduced:

- the *Conventional ETAS* model introduced by Ogata (1988, 1998) and implemented in the R package "ETAS" (Jalilian, 2019) and
- the *ETAS-Incomplete* model as described by (Grimm et al., 2022).

We define both models and derive the corresponding log-likelihood (LL) functions that are needed to optimize the model parameters by maximum likelihood estimation (MLE). Derivatives of the LL-functions, necessary for gradient-based optimization methods, are derived in later chapters.

Notations follow Jalilian (2019), Hainzl (2021) and Grimm et al. (2022).

1.1 Conventional ETAS Model

In the conventional ETAS approach, the occurrence rate of an earthquake

- with *magnitude* m,
- occurring at *time* t and
- at *location* (x, y)

is modeled by a non-homogeneous Poisson process with intensity function

$$\lambda(t, x, y, m) = f_0(m)\, R_0(t, x, y) \tag{1.1}$$

where

$$f_0(m) = \beta\, e^{-\beta(m-M_c)} \tag{1.2}$$

is the probability density function (pdf) of the frequency-magnitude distribution (FMD) following the Gutenberg-Richter law with parameter β (M_c denotes the pre-defined cut-off magnitude).

The overall event occurrence rate at time t and at location (x, y), denoted by $R_0(t, x, y)$ in Equ. (1.1), is modeled by a superposition of the (time-invariant) *seismic background rate*

$$\mu\, u(x, y)$$

and a sum of the trigger rate contributions $R_0^{trig}(t, x, y, i)$ of all events i that occurred prior to current time t, i.e. with $t_i < t$:

$$R_0(t, x, y) = \mu\, u(x, y) + \sum_{i:t_i<t} R_0^{trig}(t, x, y, i) \tag{1.3}$$

More precisely, the trigger rate contribution of a past event i to the current time t and location (x, y) is modeled by the product of the expected number of offsprings (*aftershock productivity*) of event i,

$$A\, e^{\alpha(m_i - M_c)},$$

and the *temporal* and *spatial trigger functions*

$$g_{c,p}(t - t_i) \qquad \textit{and} \qquad f_{D,\gamma,q}(r_i(x, y), m_i, l_i),$$

modeling the distribution of (relative) occurrence times/ locations of aftershocks triggered by event i. The precise inputs and shapes of the temporal and spatial trigger functions can assume different forms, and are discussed in detail in later chapters. The spatial seismic background rate $u(x, y)$ has no functional form, but rather is a set of values on a grid.
We obtain

$$R_0^{trig}(t, x, y, i) := A\, e^{\alpha(m_i - M_c)}\, g_{c,p}(t - t_i)\, f_{D,\gamma,q}(r_i(x, y), m_i, l_i).$$

to be plugged into Equ. 1.3. Thus, in total, the Standard ETAS model consists of the eight parameters

- β (frequency-magnitude distribution),
- μ (seismic background distribution),
- $A\,, \alpha$ (aftershock productivity),
- $c\,, p$ (temporal trigger function) and
- $D\,, \gamma\,, q$ (spatial trigger function).

1.2 ETAS-Incomplete Model

The *ETAS-Incomplete* model is based on the same "true" frequency-magnitude distribution $f_0(m)$ and spatio-temporal event rate $R_0(t, x, y)$ as introduced for the *Standard ETAS* model case.
However, the main idea of the incompleteness model is that these true (physical) relationships are not accurately identifiable in observed earthquake records due to missing events as a result of the bias of temporal record incompleteness. This incompleteness typically stems from overlapping coda-waves after large main-shock events (so-called *short-term aftershock incompleteness, STAI*), but can also be the result of intense seismic swarm activity.
Fitting the "true" relationships to incomplete data records may therefore lead to significantly biased parameter estimates.

Thus, the approach of the ETAS-Incomplete model is to adapt both the frequency-magnitude distribution f and event rate R to account for record incompleteness and fit these "apparent" relationships to the earthquake catalogs.

Following the *ETASI* model of Hainzl (2021), we define the "apparent" temporal frequency-magnitude distribution

$$\begin{aligned} f(m,t) &= f_0(m)\, N_0(t)\, \frac{e^{-N_0(t)\, e^{-\beta(m-M_c)}}}{1-e^{-N_0(t)}} \\ &= \beta\, e^{-\beta(m-M_c)}\, N_0(t)\, \frac{e^{-N_0(t)\, 10^{-b(m-M_c)}}}{1-e^{-N_0(t)}} \end{aligned}$$

and the "apparent" spatio-temporal event occurrence rate

$$R(t,x,y) = \frac{R_0(t,x,y)}{N_0(t)} \left(1 - e^{-N_0(t)}\right).$$

In both functions, $N_0(t)$ denotes the expected number of events occurring within the small blind time window $[t - T_b, t]$ in the entire target space, i.e.

$$N_0(t) = \int_{t-T_b}^{t} \iint_S R_0(t,x,y)\, dx\, dy\, dt \approx T_b \iint_S R_0(t,x,y)\, dx\, dy.$$

The approximation holds under the assumption that the event rate is approximately constant during the blind time.

In contrast to the Standard ETAS model, the ETAS rate is defined as the product of the apparent FMD and the apparent spatio-temporal event rate, rather than their "true" analogons. By canceling out terms, we obtain the ETAS rate

$$\lambda(t,x,y,m) = f(m,t)\, R(t,x,y) \approx \beta\, e^{-\beta(m-M_c)}\, R_0(t,x,y)\, e^{-N_0(t)\, e^{-\beta(m-M_c)}}$$

1.3 Temporal trigger function

1.3.1 Design and Notation

The *spatial trigger function* $g_{c,p}(t-t_i)$ models the distribution of the (relative) occurrence times $t-t_i$ of direct aftershocks triggered by event i. In both ETAS model versions, we use the *Omori law* to specify the temporal decay of aftershock intensity by defining

$$g_{c,p}(t-t_i) = (t-t_i+c)^{-p}, \qquad t-t_i > 0$$

with parameters c and p.

1.3.2 Temporal restriction T

We may want to restrict temporal triggering to a maximum number of days after the mainshock. In that case, we would set the Omori law equal to 0 for $t-t_i > T$. Howevr, this functionality is currently not implemented.

1.3.3 Temporal integral $\int_{T_1}^{T_2} g_{c,p}(t-t_i)\,dt$

It holds

$$\begin{aligned}\int_{T_1}^{T_2} g_{c,p}(t-t_i)\,dt &= \int_{T_1}^{T_2} (t-t_i+c)^{-p}\,dt \\ &= \frac{1}{1-p}\left[(t-t_i+c)^{1-p}\right]_{T_1}^{T_2} \\ &= \frac{1}{1-p}\left((T_2-t_i+c)^{1-p} - ((T_1-t_i)_{\geq 0}+c)^{1-p}\right).\end{aligned}$$

1.4 Spatial trigger function (Spatial Kernel)

1.4.1 Design and Notation

The *spatial kernel* $f_{D,\gamma,q}(r_i, m_i, l_i)$ models the distribution of the 2-dimensional locations (x,y) at which direct aftershocks of event i occur. We introduce here a novel design of the spatial kernel, that works in both the Standard and ETAS-Incomplete model and that may generally assume two shapes,

$$f_{D,\gamma,q}(r_i(x,y), m_i, l_i) := \begin{cases} \frac{q-1}{D\,\exp(\gamma(m_i-M_c))}\left(1+\frac{\pi\, r_i(x,y)^2}{D\,\exp(\gamma(m_i-M_c))}\right)^{-q} & (isotropic) \\ \frac{q-1}{D\,\exp(\gamma(m_i-M_c))}\left(1+\frac{2\,l_i\,r_i(x,y)+\pi\, r_i(x,y)^2}{D\,\exp(\gamma(m_i-M_c))}\right)^{-q} & (anisotropic). \end{cases} \tag{1.4}$$

Herein, the inputs of f are defined as follows:

- $r_i(x, y) := dist(x, y, i)$ is the point-to-point distance between a potential aftershock location (x, y) and the coordinates (x_i, y_i) of the triggering event i (isotropic case) or the shortest point-to-line distance of (x, y) to the estimated rupture segment of triggering event i (anisotropic case),
- m_i is the magnitude of the triggering event i and
- l_i is the rupture length of the triggering event i (only needed in the anisotropic case).

The kernel is constrained by the parameters D and γ that control the magnitude-dependent width of the kernel, and parameter q that describes the exponential decay of the function. The kernel is constructed in a way that they serve as a pdf, with

$$\int_0^\infty f(r, m_i, l_i)\, dr = \int_0^\infty \int_0^\infty f_{D,\gamma,q}(r_i(x, y), m_i, l_i)\, dx\, dy = 1.$$

Note that the isotropic kernel is a special case of the anisotropic one, if rupture length $l_i = 0$, i.e. if no extension of the rupture along a line segment is assumed. Denoting

$$E_\gamma(m_i) := \exp(\gamma(m_i - M_c)),$$

we can therefore work with the more general anisotropic function

$$f(r_i(x, y), m_i, l_i) = \frac{q-1}{D\, E_\gamma(m_i)} \left(1 + \frac{2\, l_i\, r_i(x, y) \,+\, \pi\, r_i(x, y)^2}{D\, E_\gamma(m_i)}\right)^{-q}.$$

1.4.2 Spatial restriction R

Following Grimm et al (2021), it may be useful to restrict the spatial kernel to a maximum distance $r_i(x, y) \leq R$. In order to retain a pdf that integrates to 1 over the restricted space, $f(r_i(x, y), m_i, l_i)$ needs to be normalized by the integral till distance R,

$$1 - \left(1 + \frac{2\, l_i\, R \,+\, \pi\, R^2}{D\, \exp(\gamma(m_i - M_c))}\right)^{1-q},$$

and we obtain the final kernel

$$f_{D,\gamma,q}(r_i(x, y), m_i, l_i) = \begin{cases} \frac{q-1}{D\, E_\gamma(m_i)} \; \frac{f^{inn}_{D,\gamma}(r_i(x,y), m_i, l_i)^{-q}}{1 - f^{inn}_{D,\gamma}(R, m_i, l_i)^{1-q}} & (r \leq R) \\ 0 & (r > R) \end{cases}$$

with the *inner function*

$$f^{inn}_{D,\gamma}(r_i(x, y), m_i, l_i) = 1 + \frac{2\, l_i\, r_i(x, y) \,+\, \pi\, r_i(x, y)^2}{D\, E_\gamma(m_i)}.$$

1.4.3 Spatial integral $\iint_{S_i(\tilde{r})} f_{D,\gamma,q}(r_i(x,y), m_i, l_i)\, dx\, dy$

Both ETAS model versions require the computation of the integral of the "true" event rate $R_0(t,x,y)$ over the target space region S,

$$\begin{aligned}\iint_S R_0(t,x,y)\, dx\, dy &= \mu \iint_S u(x,y)\, dx\, dy + \sum_{i:t_i<t} \iint_S R_0^{trig}(t,x,y,i)\, dx\, dy \\ &= \mu \iint_S u(x,y)\, dx\, dy + \sum_{i:t_i<t} A\, e^{\alpha(m_i - M_c)}\, g_{c,p}(t-t_i) \iint_S f_{D,\gamma,q}(r_i(x,y), m_i, l_i)\, dx\, dy\end{aligned}$$

The integral of the spatial background, $\iint_S u(x,y)\, dx\, dy$, is always computed numerically, since $u(x,y)$ has no functional expression.
The integral of the spatial kernel,

$$\iint_S f_{D,\gamma,q}(r_i(x,y), m_i, l_i)\, dx\, dy,$$

can be computed analytically for specific spatial areas $S_i(\tilde{r})$ covering all points $(x,y) \in \mathbb{R}^2$ up to a fixed distance $\tilde{r} \geq 0$ to the point coordinates (x_i, y_i) (isotropic case) or the rupture segment (anisotropic case) of a triggering event i:

$$S_i(\tilde{r}) := \{(x,y) \in \mathbb{R}^2 \,|\, r_i(x,y) \leq \tilde{r}\}.$$

Note that this area describes a circle in the case of an isotropic spatial kernel, and a box parallel to the rupture segment, closed by semi-circles on both sides, in the case of the anisotropic spatial kernel.

Due to identical values of $f_{D,\gamma,q}(r_i(x,y), m_i, l_i)$ along the (isotropic or anisotropic) contour lines with constant distance $r_i(x,y)$, one can convert the two-dimensional integral into a one-dimensional one integrating over the distance:

$$\iint_{S_i(\tilde{r})} f_{D,\gamma,q}(r_i(x,y), m_i, l_i)\, dx\, dy = \int_0^{\tilde{r}} (2\,\pi\, r + 2\, l_i)\, f_{D,\gamma,q}(r, m_i, l_i)\, dr$$

where the factor $2\,\pi\, r + 2\, l_i$ is the length of the isotropic ($l_i = 0$) or anisotropic ($l_i > 0$; two sides parallel to rupture segment + two semi-circular closings) contour lines for a given distance r to the trigger point source (isotropic case) or triggering rupture segment (anisotropic case).

Using the substitution rule for integrals with

$$\begin{aligned} u := u(r) := f_{inn}(r, m_i, l_i) &= 1 + \frac{2\, l_i\, r \,+\, \pi\, r^2}{D\, E_\gamma(m_i)} \\ \frac{d}{d\, r} u(r) &= \frac{2\,\pi\, r + 2\, l_i}{D\, E_\gamma(m_i)} \end{aligned}$$

in step (*) we obtain for the spatial integral (up to a distance smaller or equal to the spatial restriction, i.e. $\tilde{r} \leq R$)

$$
\begin{aligned}
&\int_0^{\tilde{r}} (2\,\pi\, r + 2\, l_i)\, f_{D,\gamma,q}(r, m_i, l_i)\, dr \\
&\quad = \int_0^{\tilde{r}} \frac{(q-1)\,(2\,\pi\, r + 2\, l_i)}{D\, E_\gamma(m_i)}\, \frac{f_{inn}(r_i, m_i, l_i)^{-q}}{1 - f_{inn}(R, m_i, l_i)^{1-q}}\, dr \\
&\quad \overset{(*)}{=} \int_0^{\tilde{r}} \frac{(q-1)\,(2\,\pi\, r + 2\, l_i)}{D\, E_\gamma(m_i)}\, \frac{u^{-q}}{1 - f_{inn}(R, m_i, l_i)^{1-q}}\, \frac{du}{(2\,\pi\, r + 2\, l_i)\,/(D\, E_\gamma(m_i))} \\
&\quad = \int_{u(0)}^{u(\tilde{r})} (q-1)\, \frac{u^{-q}}{1 - f_{inn}(R, m_i, l_i)^{1-q}}\, du \\
&\quad = \left[\frac{q-1}{1-q}\, \frac{u^{1-q}}{1 - f_{inn}(R, m_i, l_i)^{1-q}} \right]_{u(0)}^{u(\tilde{r})} \\
&\quad = \left[- \frac{f_{inn}(r, m_i, l_i)^{1-q}}{1 - f_{inn}(R, m_i, l_i)^{1-q}} \right]_0^{\tilde{r}} \\
&\quad = - \frac{f_{inn}(\tilde{r}, m_i, l_i)^{1-q} - 1}{1 - f_{inn}(R, m_i, l_i)^{1-q}} \\
&\quad = \frac{1 - f_{inn}(\tilde{r}, m_i, l_i)^{1-q}}{1 - f_{inn}(R, m_i, l_i)^{1-q}}.
\end{aligned}
$$

If we integrate up to the spatial restriction distance R, i.e. $\tilde{r} = 1$, follows

$$
\iint_{S_i(R)} f_{D,\gamma,q}(r_i(x,y), m_i, l_i)\, dx\, dy = \frac{1 - f_{inn}(R, m_i, l_i)^{1-q}}{1 - f_{inn}(R, m_i, l_i)^{1-q}} = 1.
$$

confirming that f is a pdf.

Please note:
For the rest of this manuscript, we will assume that all spatial integrals can be solved analytically over the specific region $S_i(\tilde{r})$ and will use the simplified notation S for any spatial constraints.

For more general spatial windows S, the area can be partitioned into many narrow pieces between any event and the target window boundary. The integral over each of these segments can be well approximated by the accordingly weighted analytical integral as derived above. The according methodology is called "radial partitioning" and is described in more detail by Jalilian (2019).

Chapter 2

Derivation of Log-Likelihood Functions

2.1 Conventional ETAS Model

The expected number of events occurring in the entire space-time-magnitude target window $[T_1, T_2] \times S \times [M_c, \infty)$ is

$$\begin{aligned} &\int_{M_c}^{\infty} \int_{T_1}^{T_2} \iint_S \lambda(t, x, y, m)\, dx\, dy\, dt\, dm \\ &= \int_{M_c}^{\infty} f_0(m) dm \int_{T_1}^{T_2} \iint_S R_0(t, x, y)\, dx\, dy\, dt \\ &= \int_{T_1}^{T_2} \iint_S R_0(t, x, y)\, dx\, dy\, dt \end{aligned} \tag{2.1}$$

since $\int_{M_c}^{\infty} f_0(m) dm = 1$ holds by construction of the PDF.

Thus, the probability of observing $N = n$ events in the remaining space-time target window, assuming a Poisson distribution with expected number

$$\mathbb{E}(N) = \lambda_{[T_1,T_2]\times S} := \int_{T_1}^{T_2} \iint_S R_0(t, x, y)\, dx\, dy\, dt$$

is

$$P(N = n) = e^{-\lambda_{[T_1,T_2]\times S}} \frac{\left(\lambda_{[T_1,T_2]\times S}\right)^n}{n!}.$$

Furthermore the probability density of a specific event i with (t_i, x_i, y_i, m_i) is

$$d_i = \frac{\lambda(t_i, x_i, y_i, m_i)}{\lambda_{[T_1,T_2]\times S}}.$$

Due to independence of observations, and $n!$ possible timely reorders of an observed event sample $\tilde{E} = \{(t_i, x_i, y_i, m_i) | i = 1, ..., n\}$, the likelihood of getting the very observed sample $\tilde{E}$ is

$$
\begin{aligned}
L(\theta) &= n! \ P(n) \prod_{i=1}^{n} d_i \\
&= n! \ e^{-\lambda_{[T_1,T_2]\times S}} \frac{\left(\lambda_{[T_1,T_2]\times S}\right)^n}{n!} \prod_{i=1}^{n} \frac{\lambda(t_i, x_i, y_i, m_i)}{\lambda_{[T_1,T_2]\times S}} \\
&= e^{-\lambda_{[T_1,T_2]\times S}} \prod_{i=1}^{n} \lambda(t_i, x_i, y_i, m_i) \\
&= e^{-\int_{T_1}^{T_2} \iint_S R_0(t,x,y)\, dx\, dy\, dt} \prod_{i=1}^{n} f_0(m_i)\, R_0(t_i, x_i, y_i).
\end{aligned}
$$

with $\theta = \{\beta, \mu, A, \alpha, c, p, D, \gamma, q\}$.

Given an earthquake catalog with N recorded events in the target time-space window, the log-likelihood function evaluates to

$$
\begin{aligned}
LL(\theta) = ln\,(L(\theta)) &= \sum_{j=1}^{N} ln\,(f_0(m_i)\, R_0(t_i, x_i, y_i)) - \int_{T_1}^{T_2} \iint_S R_0(t, x, y)\, dx\, dy\, dt \\
&= LL_0(\beta) + LL_1(\theta) - LL_2(\theta)
\end{aligned}
\tag{2.2}
$$

with

$$
\begin{aligned}
LL_0(\beta) &= \sum_{j=1}^{N} ln\,(f_0(m_j)) \\
LL_1(\theta) &= \sum_{j=1}^{N} ln\,(R_0(t_j, x_j, y_j)) \\
&= \sum_{j=1}^{N} ln \left(\mu\, u(x_j, y_j) + \sum_{i:t_i<t_j} A\, e^{\alpha(m_i - M_c)}\, g_{c,p}(t_j - t_i)\, f_{D,\gamma,q}(r_i(x_j, y_j), m_i, l_i) \right) \\
LL_2(\theta) &= \int_{T_1}^{T_2} \iint_S R_0(t, x, y)\, dx\, dy\, dt \\
&= \mu\, (T_2 - T_1) \iint_S u(x, y)\, dx\, dy \\
&+ \sum_{i:t_i<t} A\, e^{\alpha(m_i - M_c)} \left(\frac{(T_2 - t_i + c)^{1-p} - ((T_1 - t_i)_{\geq 0} + c)^{1-p}}{1-p} \right) \left(\frac{1 - f_{inn}(\tilde{r}, m_i, l_i)^{1-q}}{1 - f_{inn}(R, m_i, l_i)^{1-q}} \right)
\end{aligned}
\tag{2.3}
$$

Note that, while the outer sum in LL_0 and LL_1 only includes target events (i.e. events $j = 1, ..., N$ with $(t_j, x_j, y_j, m_j) \in [t_0, t_1] \times S \times [m_0, \infty)$), the sum within $R_0(t,x,y)$ may include trigger contributions from complementary events outside of the space-time-magnitude target window.

The optimal set of model parameters θ is found by maximizing $LL(\theta)$. For the use of gradient-based solvers, we need to know the derivatives of $LL(\theta)$ by the respective parameters. Derivations are shown in later chapters.

2.2 ETAS-Incomplete Model

In the ETAS-Incomplete model, it is not as easily seen that we can waive the magnitude part in the integral of the time-space-magnitude intensity function $\lambda(t,x,y,m)$ over the respective target ranges $[T_1,T_2]\times S\times[M_c,\infty)$ such as

$$\int_{M_c}^{\infty}\int_{T_1}^{T_2}\iint_S \lambda(t,x,y,m)\,dx\,dy\,dt\,dm = \int_{T_1}^{T_2}\iint_S R(t,x,y)\,dx\,dy\,dt.$$

Indeed, it holds

$$\begin{aligned}
&\int_{M_c}^{\infty}\int_{T_1}^{T_2}\iint_S \lambda(t,x,y,m)\,dx\,dy\,dt\,dm\\
&\quad= \int_{M_c}^{\infty}\int_{T_1}^{T_2}\iint_S \beta\, e^{-\beta(m-M_c)}\,R_0(t,x,y)\,e^{-N_0(t)\,e^{-\beta(m-M_c)}}\,dx\,dy\,dt\,dm\\
&\quad= \int_{T_1}^{T_2}\left(\int_{M_c}^{\infty}\beta\, e^{-\beta(m-M_c)}\,e^{-N_0(t)\,e^{-\beta(m-M_c)}}\,dm\right)\left(\iint_S R_0(t,x,y)\,dx\,dy\right)\,dt
\end{aligned}$$

with

$$\int_{M_c}^{\infty}\beta\, e^{-\beta(m-M_c)}\,e^{-N_0(t)\,e^{-\beta(m-M_c)}}\,dm = \left[\frac{e^{-N_0(t)\,e^{-\beta(m-M_c)}}}{N_0(t)}\right]_{M_c}^{\infty} = \frac{1-e^{-N_0(t)}}{N_0(t)}$$

and consequently, using the aproximation

$$N_0(t) = \int_{t-T_b}^{t}\iint_S R_0(\tilde{t},x,y)\,dx\,dy\,d\tilde{t} \approx T_b\iint_S R_0(t,x,y)dxdy$$

the integral evaluates to

$$\begin{aligned}
\int_{M_c}^{\infty}\int_{T_1}^{T_2}\iint_S \lambda(t,x,y,m)\,dx\,dy\,dt\,dm &= \int_{T_1}^{T_2}\iint_S R(t,x,y)\,dx\,dy\,dt\\
&= \int_{T_1}^{T_2}\frac{1-e^{-N_0(t)}}{N_0(t)}\iint_S R_0(t,x,y)\,dx\,dy\,dt\\
&= \int_{T_1}^{T_2}\frac{1-e^{-T_b\iint_S R_0(t,x,y)\,dx\,dy}}{T_b\iint_S R_0(t,x,y)\,dx\,dy}\iint_S R_0(t,x,y)\,dx\,dy\,dt\\
&= \int_{T_1}^{T_2}\frac{1-e^{-T_b\iint_S R_0(t,x,y)\,dx\,dy}}{T_b}dt
\end{aligned}$$

Analogously to the Standard ETAS model case, we therefore obtain the likelihood function

$$L(\theta) = e^{-\int_{M_c}^{\infty} \int_{T_1}^{T_2} \iint_S \lambda(t,x,y,m)\, dx\, dy\, dt\, dm} \prod_{i=1}^{n} \lambda(t_i, x_i, y_i, m_i)$$

and the log-likelihood function

$$LL(\theta) = LL_1(\theta) - LL_2(\theta) \tag{2.4}$$

with

$$\begin{aligned}
LL_1(\theta) &= \sum_{j=1}^{N} ln\left(f(m_j, t_j)\, R(t_j, x_j, y_j)\right) \\
&= \sum_{j=1}^{N} ln\left(\beta\, e^{-\beta(m_j - M_c)}\, R_0(t_j, x_j, y_j)\, e^{-N_0(t_j)\, e^{-\beta(m_j - M_c)}}\right) \\
&= N\, ln(\beta) + \sum_{j=1}^{N} \left(ln(R_0(t_j, x_j, y_j)) - \beta(m_j - M_c) - N_0(t_j)\, e^{-\beta(m_j - M_c)}\right) \\
LL_2(\theta) &= \int_{T_1}^{T_2} \frac{1 - e^{-T_b \iint_S R_0(t,x,y)\, dx\, dy}}{T_b} dt \\
&= \frac{T_2 - T_1}{T_b} - \frac{1}{T_b} \int_{T_1}^{T_2} e^{-T_b \iint_S R_0(t,x,y)\, dx\, dy} dt
\end{aligned} \tag{2.5}$$

and

$$\begin{aligned}
\iint_S R_0(t,x,y)\, dx\, dy = \mu \iint_S u(x,y)\, dx\, dy \\
+ \sum_{i: t_i < t} A\, e^{\alpha(m_i - M_c)}\, g_{c,p}(t - t_i) \iint_S f_{D,\gamma,q}(r_i(x,y), m_i, l_i)\, dx\, dy
\end{aligned}$$

Note that the parameter vector $\theta = \{\beta, T_b, \mu, A, \alpha, c, p, D, \gamma, q\}$ comprises the additional parameter T_b. The sum in LL_1 is again computed over all target events (i.e. events j with $(t_j, x_j, y_j, m_j) \in [t_0, t_1] \times S \times [m_0, \infty)$), whereas the integral in LL_2 may include trigger contributions from complementary events outside of the space-time-magnitude target window.

The optimal set of model parameters θ is found by maximizing $LL(\theta)$. Derivatives of $LL(\theta)$ by the respective parameters, as needed for the use of gradient-based solvers, are derived in later chapters.

Chapter 3

Derivatives of LL-Functions (Conventional ETAS)

According to equations Equ. 2.2, 2.3, the LL-function of the *Standard-ETAS* model is decomposed into the three summands

$$LL(\theta) = LL_0(\beta) + LL_1(\theta) - LL_2(\theta)$$

where

$$LL_0(\beta) = \sum_{j=1}^{N} ln\,(f_0(m_j))$$

only depends on the Gutenberg-Richter parameter β and

$$LL_1(\theta) = \sum_{j=1}^{N} ln\,(R_0(t_j, x_j, y_j))$$

$$LL_2(\theta) = \int_{T_1}^{T_2} \iint_S R_0(t, x, y)\, dx\, dy\, dt$$

depend on the remaining eight parameters $\theta = \{\mu, A, \alpha, c, p, D, \gamma, q\}$.

3.1 Analytical Estimator for β

The Gutenberg-Richter parameter β appears only separated from the other parameters in $LL_0(\beta)$ and can therefore be optimized analytically by maximizing

$$LL_0(\beta) = \sum_{j=1}^{N} ln\,(f_0(m_j)) = N\, ln(\beta) - \beta \sum_{i=1}^{N} (m_i - M_c).$$

Solving the derivative

$$\frac{d}{d\beta} LL_0(\beta) = \frac{N}{\beta} - \sum_{i=1}^{N} (m_i - M_c) = 0$$

leads to the estimator

$$\beta = \frac{N}{\sum_{i=1}^{N}(m_i - M_c)}.$$

3.2 Derivatives of $LL_1(\theta)$

Via chain rule, any derivative of $LL_1(\theta)$ has the form

$$\frac{d\,LL_1(\theta)}{d\,\theta_k} = \sum_{j=1}^{N} \frac{\frac{d}{d\,\theta_k}\,R_0(t_j,x_j,y_j)}{R_0(t_j,x_j,y_j)}$$

with $\theta_k \in \theta \quad (k = 1, ..., 8)$ and

$$\frac{d}{d\,\theta_k}\,R_0(t_j,x_j,y_j) = \frac{d}{d\,\theta_k}\,\mu\, u(x_j,y_j) + \sum_{i:t_i<t_j} \frac{d}{d\,\theta_k}\,R_0^{trig}(t_j,x_j,y_j,i)$$

$$R_0^{trig}(t_j,x_j,y_j,i) = A\,e^{\alpha(m_i-M_c)}\,g_{c,p}(t_j-t_i)\,f_{D,\gamma,q}(r_i(x_j,y_j),m_i,l_i)$$

3.2.1 By μ

It holds

$$\frac{d}{d\,\mu}\,R_0(t_j,x_j,y_j) = u(x_j,y_j).$$

3.2.2 By A

It holds

$$\begin{aligned}\frac{d}{d\,A}\,R_0(t_j,x_j,y_j) &= \sum_{i:t_i<t_j} \frac{d}{d\,A}\,R_0^{trig}(t_j,x_j,y_j,i)\\ &= \sum_{i:t_i<t_j} \frac{R_0^{trig}(t_j,x_j,y_j,i)}{A}\end{aligned}$$

3.2.3 By α

It holds

$$\begin{aligned}\frac{d}{d\,\alpha}\,R_0(t_j,x_j,y_j) &= \sum_{i:t_i<t_j} \frac{d}{d\,\alpha}\,R_0^{trig}(t_j,x_j,y_j,i)\\ &= \sum_{i:t_i<t_j} (m_i-M_c)\,R_0^{trig}(t_j,x_j,y_j,i)\end{aligned}$$

3.2.4 By c

It holds

$$\begin{aligned}\frac{d}{d\,c}\,g(t_j-t_i) &= \frac{d}{d\,c}\,(c+t_j-t_i)^{-p}\\ &= (-p)\,(c+t_j-t_i)^{-p-1}\\ &= \frac{-p}{c+t_j-t_i}\,g(t_j-t_i)\end{aligned}$$

and therefore

$$\frac{d}{d\,c}\,R_0(t_j,x_j,y_j) = \sum_{i:t_i<t_j} \frac{d}{d\,c}\,R_0^{trig}(t_j,x_j,y_j,i)$$
$$= \sum_{i:t_i<t_j} \frac{-p}{c+t_j-t_i}\,R_0^{trig}(t_j,x_j,y_j,i)$$

3.2.5 By p

It holds

$$\frac{d}{d\,p}\,g(t_j-t_i) = \frac{d}{d\,p}\,(c+t_j-t_i)^{-p}$$
$$= \frac{d}{d\,p}\,e^{ln\left((c+t_j-t_i)^{-p}\right)}$$
$$= \frac{d}{d\,p}\,e^{(-p)\,ln(c+t_j-t_i)}$$
$$= -ln(c+t_j-t_i)\,g(t_j-t_i)$$

and therefore

$$\frac{d}{d\,p}\,R_0(t_j,x_j,y_j) = \sum_{i:t_i<t_j} \frac{d}{d\,p}\,R_0^{trig}(t_j,x_j,y_j,i)$$
$$= \sum_{i:t_i<t_j} -ln(c+t_j-t_i)\,R_0^{trig}(t_j,x_j,y_j,i)$$

3.2.6 Notations for Spatial Kernel

For better notation in long formula derivations, we decompose the spatial kernel

$$f_{D,\gamma,q}(r,m,l) = \frac{q-1}{D\,E_\gamma(m_i)}\,\frac{f_{D,\gamma}^{inn}(r,m,l)^{-q}}{1-f_{D,\gamma}^{inn}(R,m,l)^{1-q}}$$

into the numerator term

$$t_{numer} := \frac{q-1}{D\,E_\gamma(m)}\,f_{D,\gamma}^{inn}(r,m,l)^{-q}$$

and the denominator term

$$t_{denom} := 1-f_{D,\gamma}^{inn}(R,m,l)^{1-q},$$

i.e.

$$f_{D,\gamma,q}(r,m,l) = \frac{t_{numer}}{t_{denom}}.$$

The derivative of $f_{D,\gamma,q}(r,m,l)$ by any of the three spatial parameters D,γ,q is computed via the quotient rule

$$(f_{D,\gamma,q}(r,m,l))' = \frac{t'_{numer}\,t_{denom} - t_{numer}\,t'_{denom}}{t^2_{denom}}.$$

Also, in any spatial derivative we assume a distance smaller or equal to the spatial extent, i.e. $r_i(x, y) > R$ (otherwise the spatial kernel is 0, as its derivatives).

3.2.7 By D

Having the inner derivative

$$\frac{d}{dD} f^{inn}_{D,\gamma}(r, m, l) = -\frac{1}{D} \frac{\pi r^2}{D E_\gamma(m)} = -\frac{1}{D} \left(f^{inn}_{D,\gamma}(r, m, l) - 1\right)$$

we obtain by the use of product and chain rule in step (*)

$$\begin{aligned}
\frac{d}{dD} t_{numer} &= \frac{q-1}{E_\gamma(m)} \frac{d}{dD} \left(\frac{1}{D} f^{inn}_{D,\gamma}(r, m)^{-q} \right) \\
&\overset{(*)}{=} \frac{q-1}{E_\gamma(m)} \Bigg[-\frac{1}{D^2} f^{inn}_{D,\gamma}(r, m)^{-q} + \ldots \\
&\qquad \frac{1}{D} (-q) f^{inn}_{D,\gamma}(r, m)^{-q-1} \left(-\frac{1}{D}\right) \left(f^{inn}_{D,\gamma}(r, m) - 1\right) \Bigg] \\
&= \left[\frac{q-1}{D E_\gamma(m)} f^{inn}_{D,\gamma}(r, m)^{-q} \right] \frac{1}{D} \left(-1 + q\, f^{inn}_{D,\gamma}(r, m)^{-1} \left(f^{inn}_{D,\gamma}(r, m) - 1\right) \right) \\
&= \left(\frac{q-1}{D\, E_\gamma(m_i)} f^{inn}_{D,\gamma}(r_i, m_i, l_i)^{-q} \right) \frac{1}{D} \left(q \left(1 - f^{inn}_{D,\gamma}(r, m)^{-1}\right) - 1 \right)
\end{aligned}$$

and by the use of chain rule for the normalization term

$$\begin{aligned}
\frac{d}{dD} t_{denom} &= -(1-q)\, f^{inn}_{D,\gamma}(R, m)^{-q} \left(-\frac{1}{D}\right) \left(f^{inn}_{D,\gamma}(R, m) - 1\right) \\
&= \frac{1-q}{D} f^{inn}_{D,\gamma}(R, m)^{-q} \left(f^{inn}_{D,\gamma}(R, m) - 1\right).
\end{aligned}$$

By quotient rule, we can now conclude

$$\begin{aligned}
&\frac{d}{dD} f_{D,\gamma,q}(x, y, i) \\
&= \Bigg[t_{numer} \frac{1}{D} \left(q \left(1 - f^{inn}_{D,\gamma}(r, m)^{-1}\right) - 1 \right) t_{denom} \ldots \\
&\qquad - t_{numer} \frac{1-q}{D} f^{inn}_{D,\gamma}(R, m)^{-q} \left(f^{inn}_{D,\gamma}(R, m) - 1\right) \Bigg] \Big/ t^2_{denom} \\
&= \frac{t_{numer}}{t_{denom}} \frac{1}{D} \left[\left(q \left(1 - f^{inn}_{D,\gamma}(r, m)^{-1}\right) - 1 \right) - \frac{(1-q) \left(f^{inn}_{D,\gamma}(R, m) - 1\right)}{t_{denom}\, f^{inn}_{D,\gamma}(R, m)^q} \right] \\
&= f_{D,\gamma,q}(x, y, i) \frac{1}{D} \left[\left(q \left(1 - f^{inn}_{D,\gamma}(r, m)^{-1}\right) - 1 \right) - \frac{(1-q) \left(f^{inn}_{D,\gamma}(R, m) - 1\right)}{t_{denom}\, f^{inn}_{D,\gamma}(R, m)^q} \right]
\end{aligned}$$

Therefore, it holds

$$\frac{d}{d\,D} R_0(t_j, x_j, y_j) = \sum_{i: t_i < t_j} \frac{d}{d\,D} R^{trig}_0(t_j, x_j, y_j, i) = \sum_{i: t_i < t_j} \frac{1}{D} h_D\, R^{trig}_0(t_j, x_j, y_j, i)$$

with

$$h_D := q \left(1 - f^{inn}_{D,\gamma}(r,m)^{-1}\right) - 1 - \frac{(1-q)\ \left(f^{inn}_{D,\gamma}(R,m) - 1\right)}{t_{denom}\ f^{inn}_{D,\gamma}(R,m)^q}.$$

3.2.8 By γ

Having the inner derivative

$$\frac{d}{d\gamma} f^{inn}_{D,\gamma}(r,m) = (-(m-m_0))\ \frac{\pi r^2}{DE_\gamma(m)} = (-(m-m_0))\ \left(f^{inn}_{D,\gamma}(r,m) - 1\right) \tag{3.1}$$

we obtain by the use of product and chain rule in step (*)

$$\begin{aligned}
\frac{d}{d\gamma} t_{numer} &= \frac{q-1}{D}\ \frac{d}{d\gamma}\left(\frac{1}{E_\gamma(m)}\ f^{inn}_{D,\gamma}(r,m)^{-q}\right)\\
&\overset{(*)}{=} \frac{q-1}{D}\ \left[\frac{-(m-m_0)}{E_\gamma(m)}\ f^{inn}_{D,\gamma}(r,m)^{-q}\ \dots\right.\\
&\quad \left. + \frac{1}{E_\gamma(m)}\ (-q)\ f^{inn}_{D,\gamma}(r,m)^{-q-1}\,(-(m-m_0))\ \left(f^{inn}_{D,\gamma}(r,m) - 1\right)\right]\\
&= \left[\frac{q-1}{DE_\gamma(m)}\ f^{inn}_{D,\gamma}(r,m)^{-q}\right](m-m_0)\left(-1 + q\ f^{inn}_{D,\gamma}(r,m)^{-1}\ \left(f^{inn}_{D,\gamma}(r,m) - 1\right)\right)\\
&= t_{numer}\ (m-m_0)\ \left(q\ \left(1 - f^{inn}_{D,\gamma}(r,m)^{-1}\right) - 1\right)
\end{aligned} \tag{3.2}$$

and by the use of chain rule

$$\begin{aligned}
\frac{d}{d\gamma} t_{denom} &\overset{(*)}{=} -(1-q)\ f^{inn}_{D,\gamma}(R,m)^{-q}\,(-(m-m_0))\ \left(f^{inn}_{D,\gamma}(r,m) - 1\right)\\
&= (1-q)\ (m-m_0)\ f^{inn}_{D,\gamma}(R,m)^{-q}\left(f^{inn}_{D,\gamma}(R,m) - 1\right).
\end{aligned} \tag{3.3}$$

Comparing (3.2) and (**??**) with (**??**) and (**??**), we see that

$$\begin{aligned}
\frac{d}{d\gamma} t_{numer} &= D\ (m-m_0)\ \frac{d}{dD} t_{numer},\\
\frac{d}{d\gamma} t_{denom} &= D\ (m-m_0)\ \frac{d}{dD} t_{denom}.
\end{aligned}$$

Therefore it follows that

$$\frac{d}{d\gamma} f_{D,\gamma,q}(x,y,i) = D\ (m-m_0)\ \frac{d}{dD} f_{D,\gamma,q}(x,y,i)$$

and

$$\frac{d}{d\,\gamma}\,R_0(t_j,x_j,y_j) = \sum_{i:t_i<t_j} \frac{d}{d\,D}\,R_0^{trig}(t_j,x_j,y_j,i) = \sum_{i:t_i<t_j} (m_i - m_0)\,h_D\,R_0^{trig}(t_j,x_j,y_j,i)$$

with h_D as defined in the derivation of the derivative by D.

3.2.9 By q

We obtain by the use of product rule in step (*)

$$\begin{aligned}
&\frac{d}{dq} t_{numer} = \frac{1}{DE_\gamma(m)} \;\; \frac{d}{dq} \Big((q-1) \; f^{inn}_{D,\gamma}(r,m)^{-q} \Big) \\
&\overset{(*)}{=} \frac{1}{DE_\gamma(m)} \Big[f^{inn}_{D,\gamma}(r,m)^{-q} + (q-1) \;\; \big(-ln \left(f^{inn}_{D,\gamma}(r,m)\right)\big) \;\; f^{inn}_{D,\gamma}(r,m)^{-q} \Big] \\
&= \left[\frac{q-1}{DE_\gamma(m)} \; f^{inn}_{D,\gamma}(r,m)^{-q} \right] \left(\frac{1}{q-1} - ln \left(f^{inn}_{D,\gamma}(r,m)\right) \right) \\
&= t_{numer} \;\; \left(\frac{1}{q-1} - ln \left(f^{inn}_{D,\gamma}(r,m)\right) \right)
\end{aligned}$$

and by the use of chain rule

$$\begin{aligned}
\frac{d}{dq} t_{denom} &= - \big(-ln \left(f^{inn}_{D,\gamma}(R,m)\right)\big) \;\; f^{inn}_{D,\gamma}(R,m)^{1-q} \\
&= ln \left(f^{inn}_{D,\gamma}(R,m)\right) \;\; f^{inn}_{D,\gamma}(R,m)^{1-q}
\end{aligned}$$

By quotient rule, we can now conclude

$$\begin{aligned}
&\frac{d}{dq} f_{D,\gamma,q}(x,y,i) = \frac{d}{dq} \frac{t_{numer}}{t_{denom}} \\
&= \Bigg[t_{numer} \;\; \left(\frac{1}{q-1} - ln \left(f^{inn}_{D,\gamma}(r,m)\right) \right) \; t_{denom} \; \cdots \\
&- t_{numer} \; ln \left(f^{inn}_{D,\gamma}(R,m)\right) \;\; f^{inn}_{D,\gamma}(R,m)^{1-q} \Bigg] \;\; / \; t^2_{denom} \\
&= f_{D,\gamma,q}(x,y,i) \; \left[\frac{1}{q-1} - ln \left(f^{inn}_{D,\gamma}(r,m)\right) - \frac{ln \left(f^{inn}_{D,\gamma}(R,m)\right)}{t_{denom}} \;\; f^{inn}_{D,\gamma}(R,m)^{1-q} \right]
\end{aligned}$$

Therefore, it holds

$$\frac{d}{d\,q} R_0(t_j,x_j,y_j) = \sum_{i:t_i<t_j} \frac{d}{d\,q} R_0^{trig}(t_j,x_j,y_j,i) = \sum_{i:t_i<t_j} h_q \, R_0^{trig}(t_j,x_j,y_j,i)$$

with

$$h_q := \frac{1}{q-1} - ln \left(f^{inn}_{D,\gamma}(r,m)\right) - \frac{ln \left(f^{inn}_{D,\gamma}(R,m)\right)}{t_{denom}} \;\; f^{inn}_{D,\gamma}(R,m)^{1-q} \, .$$

3.3 Derivatives of $LL_2(\theta)$

The second summand of the LL-function is

$$\begin{aligned} LL_2(\theta) &= \int_{T_1}^{T_2} \iint_S R_0(t,x,y)\,dx\,dy\,dt \\ &= (T_2 - T_1)\,\mu \iint_S u(x,y)\,dx\,dy + \sum_{i=1}^{N} A\,e^{\alpha(m_i - M_c)}\,G_{c,p}(T_1,T_2,i)\,F_{D,\gamma,q}(S,i) \end{aligned}$$

with

$$\begin{aligned} G_{c,p}(T_1,T_2,i) &:= \int_{T_1}^{T_2} g(t-t_i)\,dt = \frac{1}{1-p}\left((T_2 - t_i + c)^{1-p} - ((T_1 - t_i)_{\geq 0} + c)^{1-p}\right), \\ F_{D,\gamma,q}(S,i) &:= \iint_S f_{D,\gamma,q}(r_i(x,y), m_i, l_i)\,dx\,dy = \frac{1 - \left(1 + \frac{2\,l_i\,\tilde{r} + \pi\,\tilde{r}^2}{D\,E_\gamma(m_i)}\right)^{1-q}}{1 - \left(1 + \frac{2\,l_i\,R + \pi\,R^2}{D\,E_\gamma(m_i)}\right)^{1-q}} \end{aligned}$$

We obtain the following derivatives.

3.3.1 By μ

It holds

$$\frac{d}{d\mu} LL_2(\theta) = (T_2 - T_1) \iint_S u(x,y)\,dx\,dy.$$

3.3.2 By A

It holds

$$\frac{d}{d\,A} LL_2(\theta) = \sum_{i=1}^{N} e^{\alpha(m_i - M_c)}\,G_{c,p}(T_1,T_2,i)\,F_{D,\gamma,q}(S,i)$$

3.3.3 By α

It holds

$$\frac{d}{d\,A} LL_2(\theta) = \sum_{i=1}^{N} A\,(m_i - M_c)\,e^{\alpha(m_i - M_c)}\,G_{c,p}(T_1,T_2,i)\,F_{D,\gamma,q}(S,i)$$

3.3.4 By c

It holds

$$\frac{d}{d\,c} LL_2(\theta) = \sum_{i=1}^{N} A\,e^{\alpha(m_i - M_c)} \left(\frac{d}{d\,c} G_{c,p}(T_1,T_2,i)\right) F_{D,\gamma,q}(S,i)$$

with

$$\begin{aligned}\frac{d}{d\,c}G_{c,p}(T_1,T_2,i) &= \frac{d}{d\,c}\,\frac{1}{1-p}\,\left((T_2-t_i+c)^{1-p}-((T_1-t_i)_{\geq 0}+c)^{1-p}\right)\\ &= (T_2-t_i+c)^{-p}-((T_1-t_i)_{\geq 0}+c)^{-p}.\end{aligned}$$

3.3.5 By p

It holds

$$\frac{d}{d\,p}\,x^{1-p}=\frac{d}{d\,p}\,e^{(1-p)\,ln(x)}=-ln(x)\,x^{1-p},$$

and via quotient rule

$$\begin{aligned}\frac{d}{d\,p}\,\frac{x^{1-p}}{1-p} &= \frac{-ln(x)\,x^{1-p}\,(1-p)-x^{1-p}\,(-1)}{(1-p)^2}\\ &= x^{1-p}\,\frac{1-ln(x)\,(1-p)}{(1-p)^2}.\end{aligned}$$

Therefore we obtain

$$\frac{d}{d\,p}LL_2(\theta)=\sum_{i=1}^{N}A\,e^{\alpha(m_i-M_c)}\,\left(\frac{d}{d\,p}G_{c,p}(T_1,T_2,i)\right)\,F_{D,\gamma,q}(S,i)$$

with

$$\begin{aligned}\frac{d}{d\,p}G_{c,p}(T_1,T_2,i) &= \frac{d}{d\,p}\,\frac{1}{1-p}\,\left((T_2-t_i+c)^{1-p}-((T_1-t_i)_{\geq 0}+c)^{1-p}\right)\\ &= \frac{d}{d\,p}\,\frac{(T_2-t_i+c)^{1-p}}{1-p}-\frac{((T_1-t_i)_{\geq 0}+c)^{1-p}}{1-p}\\ &= (T_2-t_i+c)^{1-p}\,\frac{1-ln(T_2-t_i+c)\,(1-p)}{(1-p)^2}\quad\ldots\\ &\quad-((T_1-t_i)_{\geq 0}+c)^{1-p}\,\frac{1-ln((T_1-t_i)_{\geq 0}+c)\,(1-p)}{(1-p)^2}.\end{aligned}$$

3.3.6 Notations for Spatial Integral

Again, for better notation in long formula derivations, we decompose the spatial integral

$$F_{D,\gamma,q}(S,i)=\iint_S f_{D,\gamma,q}(r_i(x,y),m_i,l_i)\,dx\,dy=\frac{1-\left(1+\frac{2\,l_i\,\tilde{r}+\pi\,\tilde{r}^2}{D\,E_\gamma(m_i)}\right)^{1-q}}{1-\left(1+\frac{2\,l_i\,R+\pi\,R^2}{D\,E_\gamma(m_i)}\right)^{1-q}}$$

into the numerator term

$$t_{numInteg}:=1-\left(f^{inn}_{D,\gamma}(r,m,l)\right)^{1-q}$$

and the denominator term

$$t_{denom} := 1 - \left(f^{inn}_{D,\gamma}(R,m,l)\right)^{1-q},$$

i.e.

$$F_{D,\gamma,q}(S,i) = \frac{t_{numInteg}}{t_{denom}}.$$

The derivative of $f_{D,\gamma,q}(r,m,l)$ by any of the three spatial parameters D,γ,q is computed via the quotient rule

$$\left(F_{D,\gamma,q}(S,i)\right)' = \frac{t'_{numInteg}\, t_{denom} - t_{numInteg}\, t'_{denom}}{t^2_{denom}}.$$

Also, in any spatial derivative we assume a distance smaller or equal to the spatial extent, i.e. $r_i(x,y) > R$ (otherwise the spatial integral has reached 1 and the derivative is 0).

3.3.7 By D

From subsection 3.2.7 we obtain

$$\frac{d}{dD} t_{numInteg} = \frac{1-q}{D}\; f^{inn}_{D,\gamma}(r,m)^{-q}\left(f^{inn}_{D,\gamma}(r,m) - 1\right)$$
$$\frac{d}{dD} t_{scale} = \frac{1-q}{D}\; f^{inn}_{D,\gamma}(R,m)^{-q}\left(f^{inn}_{D,\gamma}(R,m) - 1\right)$$

and consequently, via quotient rule,

$$\begin{aligned}
&\frac{d}{dD} F_{D,\gamma,q}(S,i) \\
&= \left[\frac{1-q}{D}\; f^{inn}_{D,\gamma}(r,m)^{-q}\left(f^{inn}_{D,\gamma}(r,m) - 1\right)\; t_{scale}\; \ldots \right. \\
&\quad \left. - t_{numInteg}\; \frac{1-q}{D}\; f^{inn}_{D,\gamma}(R,m)^{-q}\left(f^{inn}_{D,\gamma}(R,m) - 1\right)\right] / t^2_{scale}.
\end{aligned}$$

3.3.8 By γ

From subsection 3.2.8 we obtain

$$\frac{d}{d\gamma} t_{numInteg} = D\;(m - m_0)\; \frac{d}{dD} t_{numInteg},$$
$$\frac{d}{d\gamma} t_{scale} = D\;(m - m_0)\; \frac{d}{dD} t_{scale}.$$

Therefore it follows that

$$\frac{d}{d\gamma} F_{D,\gamma,q}(S,i) = D\;(m - m_0)\; \frac{d}{dD} F_{D,\gamma,q}(S,i).$$

3.3.9 By q

From subsection 3.2.9 we obtain

$$\frac{d}{dq} t_{integ} = ln \left(f^{inn}_{D,\gamma} (r, m) \right) \ f^{inn}_{D,\gamma} (r, m)^{1-q} \tag{3.4}$$

$$\frac{d}{dq} t_{scale} = ln \left(f^{inn}_{D,\gamma} (R, m) \right) \ f^{inn}_{D,\gamma} (R, m)^{1-q} \tag{3.5}$$

and consequently, via quotient rule,

$$\frac{d}{dq} F_{D,\gamma,q}(r, x, y) = \Bigg[ln \left(f^{inn}_{D,\gamma} (r, m) \right) \ f^{inn}_{D,\gamma} (r, m)^{1-q} \ t_{scale} \ ... \\ - t_{integ} \ ln \left(f^{inn}_{D,\gamma} (R, m) \right) \ f^{inn}_{D,\gamma} (R, m)^{1-q} \Bigg] / t^2_{scale}. \tag{3.6}$$

Chapter 4

Derivatives of LL-Functions (ETAS-Incomplete)

4.1 Derivatives of $N(t)$

In both LL-summands $LL_1(\theta)$ and $LL_2(\theta)$ occurs the term $N(t)$ representing the overall event rate at time t in the entire target space window,

$$\begin{aligned} N(t) &\approx T_b \iint_S R_0(t,x,y)\,dx\,dy \\ &= T_b \Big(\mu \iint_S u(x,y)\,dx\,dy + \sum_{i:t_i<t} A\,e^{\alpha(m_i-M_c)}\,g_{c,p}(t-t_i)\,F_{D,\gamma,q}(S,i)\Big). \end{aligned}$$

with

$$F_{D,\gamma,q}(S,i) = \iint_S f_{D,\gamma,q}(r_i(x,y),m_i,l_i)\,dx\,dy = \frac{1-\Big(1+\frac{2\,l_i\,\tilde{r}+\pi\,\tilde{r}^2}{D\,E_\gamma(m_i)}\Big)^{1-q}}{1-\Big(1+\frac{2\,l_i\,R+\pi\,R^2}{D\,E_\gamma(m_i)}\Big)^{1-q}}.$$

4.1.1 By μ

It holds

$$\frac{d}{d\,\mu}\,N(t) = T_b \iint_S u(x,y)\,dx\,dy.$$

4.1.2 By A

It holds

$$\frac{d}{d\,A}\,N(t) = T_b \sum_{i:t_i<t} e^{\alpha(m_i-M_c)}\,g_{c,p}(t-t_i)\,F_{D,\gamma,q}(S,i).$$

4.1.3 By α

It holds

$$\frac{d}{d\,\alpha}\,N(t) = T_b \sum_{i:t_i<t} A\,(m_i - M_c)\,e^{\alpha(m_i-M_c)}\,g_{c,p}(t-t_i)\,F_{D,\gamma,q}(S,i).$$

4.1.4 By c

By section 3.2, it holds

$$\begin{aligned}\frac{d}{d\,c}\,N(t) &= T_b \sum_{i:t_i<t} A\,e^{\alpha(m_i-M_c)}\,\left(\frac{d}{d\,c}\,g_{c,p}(t-t_i)\right)\,F_{D,\gamma,q}(S,i)\\ &= T_b \sum_{i:t_i<t} A\,e^{\alpha(m_i-M_c)}\,\frac{-p}{c+t-t_i}\,g_{c,p}(t-t_i)\,F_{D,\gamma,q}(S,i)\end{aligned}$$

4.1.5 By p

By section 3.2, it holds

$$\begin{aligned}\frac{d}{d\,p}\,N(t) &= T_b \sum_{i:t_i<t} A\,e^{\alpha(m_i-M_c)}\,\left(\frac{d}{d\,p}\,g_{c,p}(t-t_i)\right)\,F_{D,\gamma,q}(S,i)\\ &= T_b \sum_{i:t_i<t} A\,e^{\alpha(m_i-M_c)}\,\left(-ln(c+t-t_i)\right)\,g_{c,p}(t-t_i)\,F_{D,\gamma,q}(S,i)\end{aligned}$$

4.1.6 By D

It holds

$$\frac{d}{d\,D}\,N(t) = T_b \sum_{i:t_i<t} A\,e^{\alpha(m_i-M_c)}\,g_{c,p}(t-t_i)\,\left(\frac{d}{d\,D}\,F_{D,\gamma,q}(S,i)\right)$$

with $\frac{d}{d\,D}\,F_{D,\gamma,q}(S,i)$ as derived in section 3.3.

4.1.7 By γ

It holds

$$\frac{d}{d\,\gamma}\,N(t) = T_b \sum_{i:t_i<t} A\,e^{\alpha(m_i-M_c)}\,g_{c,p}(t-t_i)\,\left(\frac{d}{d\,\gamma}\,F_{D,\gamma,q}(S,i)\right)$$

with $\frac{d}{d\,\gamma}\,F_{D,\gamma,q}(S,i)$ as derived in section 3.3.

4.1.8 By q

It holds

$$\frac{d}{d\,q}\,N(t) = T_b \sum_{i:t_i<t} A\,e^{\alpha(m_i-M_c)}\,g_{c,p}(t-t_i)\,\left(\frac{d}{d\,q}\,F_{D,\gamma,q}(S,i)\right)$$

with $\frac{d}{d\,q}\,F_{D,\gamma,q}(S,i)$ as derived in section 3.3.

4.1.9 By T_b

It holds

$$\frac{d}{d\,T_b}\,N(t) = \frac{N(t)}{T_b} = \mu\,\iint_S u(x,y)\,dx\,dy + \sum_{i:t_i<t} A\,e^{\alpha(m_i-M_c)}\,g_{c,p}(t-t_i)\,F_{D,\gamma,q}(S,i).$$

4.1.10 By β

The parameter β does not occur in $N(t)$, i.e. $\frac{d}{d\,\beta}\,N(t) = 0$.

4.2 Derivatives of $LL_1(\theta)$

For the *ETAS-Incomplete* model, derivatives of the first LL-summand $LL_1(\theta)$ are most easily computed starting from (see Equ. 2.5)

$$LL_1(\theta) = N\, ln(\beta) + \sum_{j=1}^{N} ln(R_0(t_j, x_j, y_j)) - \beta(m_j - M_c) - N(t_j)\, e^{-\beta(m_j - M_c)}$$

In this case, the Gutenberg-Richter parameter β is not isolated from the other nine parameters and therefore the entire parameter set $\theta = \{\beta, T_b, \mu, A, \alpha, c, p, D, \gamma, q\}$ needs to be optimized numerically. The following gradients serve for gradient-based optimization methods.

4.2.1 By $\mu, A, \alpha, c, p, D, \gamma, q$

For any parameter other than β and T_b, we obtain the derivative by chain rule as

$$\frac{d}{d\,\theta_k} LL_1(\theta) = \sum_{j=1}^{N} \frac{\frac{d}{d\,\theta_k} R_0(t_j, x_j, y_j)}{R_0(t_j, x_j, y_j)} - \left(\frac{d}{d\,\theta_k} N(t_j)\right) e^{-\beta(m_j - M_c)} \tag{4.1}$$

with the inner derivatives $\frac{d}{d\,\theta_k} R_0(t_j, x_j, y_j)$ and $\frac{d}{d\,\theta_k} N(t_j)$ from section section 3.2 and 4.1, respectively, and therefore:

$$\frac{d}{d\,\mu} LL_1(\theta) = \sum_{j=1}^{N} \frac{\frac{d}{d\,\mu} R_0(t_j, x_j, y_j)}{R_0(t_j, x_j, y_j)} - \left(\frac{d}{d\,\mu} N(t_j)\right) e^{-\beta(m_j - M_c)}$$

$$\frac{d}{d\,A} LL_1(\theta) = \sum_{j=1}^{N} \frac{\frac{d}{d\,A} R_0(t_j, x_j, y_j)}{R_0(t_j, x_j, y_j)} - \left(\frac{d}{d\,A} N(t_j)\right) e^{-\beta(m_j - M_c)}$$

$$\frac{d}{d\,\alpha} LL_1(\theta) = \sum_{j=1}^{N} \frac{\frac{d}{d\,\alpha} R_0(t_j, x_j, y_j)}{R_0(t_j, x_j, y_j)} - \left(\frac{d}{d\,\alpha} N(t_j)\right) e^{-\beta(m_j - M_c)}$$

$$\frac{d}{d\,c} LL_1(\theta) = \sum_{j=1}^{N} \frac{\frac{d}{d\,c} R_0(t_j, x_j, y_j)}{R_0(t_j, x_j, y_j)} - \left(\frac{d}{d\,c} N(t_j)\right) e^{-\beta(m_j - M_c)}$$

$$\frac{d}{d\,p} LL_1(\theta) = \sum_{j=1}^{N} \frac{\frac{d}{d\,p} R_0(t_j, x_j, y_j)}{R_0(t_j, x_j, y_j)} - \left(\frac{d}{d\,p} N(t_j)\right) e^{-\beta(m_j - M_c)}$$

$$\frac{d}{d\,D} LL_1(\theta) = \sum_{j=1}^{N} \frac{\frac{d}{d\,D} R_0(t_j, x_j, y_j)}{R_0(t_j, x_j, y_j)} - \left(\frac{d}{d\,D} N(t_j)\right) e^{-\beta(m_j - M_c)}$$

$$\frac{d}{d\,\gamma} LL_1(\theta) = \sum_{j=1}^{N} \frac{\frac{d}{d\,\gamma} R_0(t_j, x_j, y_j)}{R_0(t_j, x_j, y_j)} - \left(\frac{d}{d\,\gamma} N(t_j)\right) e^{-\beta(m_j - M_c)}$$

$$\frac{d}{d\,q} LL_1(\theta) = \sum_{j=1}^{N} \frac{\frac{d}{d\,q} R_0(t_j, x_j, y_j)}{R_0(t_j, x_j, y_j)} - \left(\frac{d}{d\,q} N(t_j)\right) e^{-\beta(m_j - M_c)}$$

4.2.2 By T_b (Blind Time)

Having

$$\frac{d}{d\,T_b}\,N(t) = \frac{N(t)}{T_b}$$

from section 4.1 it holds

$$\frac{d}{dT_b}\,LL_1(\theta) = \sum_{j=1}^{N} -\left(\frac{d}{d\,T_b} N(t_j)\right)\, e^{-\beta(m_j - M_c)}$$

4.2.3 By β (Gutenberg-Richter)

It holds

$$\frac{d}{d\beta}\,LL_1(\theta) = \frac{N}{\beta} + \sum_{j=1}^{N} (m_i - M_c)\,\left(N(t_j)\, e^{-\beta(m_i - M_c)} - 1\right).$$

4.3 Derivatives of $LL_2(\theta)$

The second summand of the LL-function is

$$LL_2(\theta) = \frac{T_2 - T_1}{T_b} - \frac{1}{T_b} \int_{T_1}^{T_2} e^{-N_0(t)}\, dt$$

4.3.1 By $\mu, A, \alpha, c, p, D, \gamma, q$

For any parameter other than T_b, we obtain the derivative by chain rule as

$$\frac{d}{d\,\theta_k} LL_2(\theta) = \frac{1}{T_b} \int_{T_1}^{T_2} \left(\frac{d}{d\,\theta_k} N_0(t) \right) e^{-N_0(t)}\, dt. \tag{4.2}$$

with the inner derivatives $\frac{d}{d\,\theta_k} N_0(t)$ given in section 4.1:

$$\begin{aligned}
\frac{d}{d\,\mu} LL_2(\theta) &= \frac{1}{T_b} \int_{T_1}^{T_2} \left(\frac{d}{d\,\mu} N_0(t) \right) e^{-N_0(t)}\, dt \\
\frac{d}{d\,A} LL_2(\theta) &= \frac{1}{T_b} \int_{T_1}^{T_2} \left(\frac{d}{d\,A} N_0(t) \right) e^{-N_0(t)}\, dt \\
\frac{d}{d\,\alpha} LL_2(\theta) &= \frac{1}{T_b} \int_{T_1}^{T_2} \left(\frac{d}{d\,\alpha} N_0(t) \right) e^{-N_0(t)}\, dt \\
\frac{d}{d\,c} LL_2(\theta) &= \frac{1}{T_b} \int_{T_1}^{T_2} \left(\frac{d}{d\,c} N_0(t) \right) e^{-N_0(t)}\, dt \\
\frac{d}{d\,p} LL_2(\theta) &= \frac{1}{T_b} \int_{T_1}^{T_2} \left(\frac{d}{d\,p} N_0(t) \right) e^{-N_0(t)}\, dt \\
\frac{d}{d\,D} LL_2(\theta) &= \frac{1}{T_b} \int_{T_1}^{T_2} \left(\frac{d}{d\,D} N_0(t) \right) e^{-N_0(t)}\, dt \\
\frac{d}{d\,\gamma} LL_2(\theta) &= \frac{1}{T_b} \int_{T_1}^{T_2} \left(\frac{d}{d\,\gamma} N_0(t) \right) e^{-N_0(t)}\, dt \\
\frac{d}{d\,q} LL_2(\theta) &= \frac{1}{T_b} \int_{T_1}^{T_2} \left(\frac{d}{d\,q} N_0(t) \right) e^{-N_0(t)}\, dt
\end{aligned}$$

4.3.2 By T_b (Blind Time)

It holds

$$\frac{d}{d\,T_b} \frac{T_2 - T_1}{T_b} = -\frac{T_2 - T_1}{T_b^2}$$

and, via quotient rule,

$$\frac{d}{d\,T_b} \frac{e^{-N_0(t)}}{T_b} = \frac{\left(-\frac{d}{d\,T_b} N_0(t) \right) e^{-N_0(t)}\, T_b - e^{-N_0(t)}}{T_b^2} = -\frac{e^{-N_0(t)}}{T_b^2} \left(T_b \left(\frac{d}{d\,T_b} N_0(t) \right) + 1 \right).$$

We obtain

$$\frac{d}{d\,T_b} LL_2(\theta) = -\frac{T_2 - T_1}{T_b^2} + \frac{1}{T_b^2} \int_{T_1}^{T_2} \left(T_b \left(\frac{d}{d\,T_b}\, N_0(t) \right) + 1 \right) e^{-N_0(t)}\, dt.$$

4.3.3 By β (Gutenberg-Richter)

The parameter β does not occur in $LL_2(\theta)$, therefore

$$\frac{d}{d\,\beta} LL_2(\theta) = 0.$$

Bibliography

C. Grimm, S. Hainzl, M. Käser, and H. Küchenhoff. Solving three major biases of the ETAS model to improve forecasts of the 2019 Ridgecrest sequence. *Stochastic Environmental Research and Risk Assessment*, 2, 2022. doi: 10.1007/s00477-022-02221-2.

Christian Grimm, Martin Käser, Sebastian Hainzl, Marco Pagani, and Helmut Küchenhoff. Improving Earthquake Doublet Frequency Predictions by Modified Spatial Trigger Kernels in the Epidemic-Type Aftershock Sequence (ETAS) Model. *Bulletin of the Seismological Society of America*, (Xx), 2021. ISSN 0037-1106. doi: 10.1785/0120210097.

Sebastian Hainzl. ETAS-Approach Accounting for Short-Term Incompleteness of Earthquake Catalogs. *Bulletin of the Seismological Society of America*, 2021. ISSN 0037-1106. doi: 10.1785/0120210146.

Abdollah Jalilian. ETAS: An R package for fitting the space-time ETAS model to earthquake data. *Journal of Statistical Software*, 88(1), 2019. ISSN 15487660. doi: 10.18637/jss.v088.c01.

Yosihiko Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9–27, mar 1988. ISSN 0162-1459.

Yosihiko Ogata. Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2):379–402, 1998. ISSN 00203157.

Contributing Publications

Grimm, C., Käser, M., Hainzl, S., Pagani, M., Küchenhoff, H. (2021). Improving Earthquake Doublet Frequency Predictions by Modified Spatial Trigger Kernels in the Epidemic-Type Aftershock Sequence (ETAS) Model. *Bulletin of the Seismological Society of America*

Grimm, C., Hainzl, S., Käser, M., Küchenhoff, H. (2022a). Solving three major biases of the ETAS model to improve forecasts of the 2019 Ridgecrest sequence. *Stochastic Environmental Research and Risk Assessment*

Grimm, C., Rupprecht, T., Johnson, K., Hainzl, S., Küchenhoff, H., Käser, M., Pagani, M. (2022b). New Statistical Perspectives on Bath's Law and Aftershock Productivity. *Manuscript submitted for publication.*

References

Abdelnaby, A. E. (2012). *Multiple earthquake effects on degrading reinforced concrete structures*. Ph.D. thesis, University of Illinois at Urbana-Champ.

de Arcangelis, L., Godano, C., and Lippiello, E. (2018). The overlap of aftershock coda waves and short-term postseismic forecasting. *Journal of Geophysical Research: Solid Earth*, 123: 5661–5674.

Bath, M. (1965). Largest inhomogeneities of the upper mantle. *Tectonophysics*, 2: 483–514.

Bird, P. (2003). An updated digital model of plate boundaries. *Geochemistry, Geophysics, Geosystems*, 4(3).

Bird, P., Liu, Z., and Rucker, W. K. (2008). Stresses that drive the plates from below: Definitions, computational path, model optimization, and error analysis. *Journal of Geophysical Research: Solid Earth*, 113: B11406.

Bountzis, P., Kostoglou, A., Papadimitriou, E., and Karakostas, V. (2021). dentification of spatiotemporal seismicity clusters in central Ionian Islands (Greece). *Physics of the Earth and Planetary Interiors*, 312: 106675.

Bountzis, P., Papadimitriou, E., and Tsaklidis, G. (2022). Identification and Temporal Characteristics of Earthquake Clusters in Selected Areas in Greece. *Applied Sciences*, 12(4): 1908.

Bountzis, P. and Tsaklidis, G. (2021). Markovian Arrival Processes for Earthquake Clustering Analysis. *Statistical Methods and Modeling of Seismogenesis*, pages 241–270.

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.

Cornell, C. A. (1968). Engineering seismic risk analysis. *Bulletin of the Seismological Society of America*, 58(5): 1583–1606.

Daley, D. and Vere-Jones, D. (2003). *An introduction to the theory of point processes: Volume I: Elementary theory and methods.*

Dascher-Cousineau, K., Brodsky, E. E., Lay, T., and Goebel, T. H. (2020). What controls variations in aftershock productivity? *Journal of Geophysical Research: Solid Earth*, 125.

Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11: 89–121.

Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. (2013). *Regression - Models, Methods and Applications.*

Felzer, K. R., Abercrombie, R. E., and Ekström, G. (2004). A common origin for aftershocks, foreshocks, and multiplets. *Bulletin of the Seismological Society of America*, 94(1): 88–98.

Field, E. H., Milner, K. R., Hardebeck, J. L., Page, M. T., van der Elst, N., Jordan, T. H., Michael, A. J., Shaw, B. E., and Werner, M. J. (2017). A spatiotemporal clustering model for the third uniform California earthquake rupture forecast (UCERF3-ETAS): Toward an operational earthquake forecast. *Bulletin of the Seismological Society of America*, 107(3): 1049–1081.

Gardner, J. K. 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.

Gibowicz, S. J. and Lasocki, S. (2005). Earthquake Doublets and Multiplets in the Fiji-Tonga-Kermadec Region. *Acta Geophysica Polonica*, 53(3): 239–274.

Grimm, C., Hainzl, S., Käser, M., and Küchenhoff, H. (2022a). Solving three major biases of the ETAS model to improve forecasts of the 2019 Ridgecrest sequence. *Stochastic Environmental Research and Risk Assessment*, 2.

Grimm, C., Käser, M., Hainzl, S., Pagani, M., and Küchenhoff, H. (2021). Improving Earthquake Doublet Frequency Predictions by Modified Spatial Trigger Kernels in the Epidemic-Type Aftershock Sequence (ETAS) Model. *Bulletin of the Seismological Society of America*, (Xx).

Grimm, C., Rupprecht, T., Johnson, K., Hainzl, S., Küchenhoff, H., Käser, M., and Pagani, M. (2022b). New Statistical Perspectives on Bath's Law and Aftershock Productivity. *Manuscript submitted for publication.*

Gulia, L., Rinaldi, A. P., Tormann, T., Vannucci, G., Enescu, B., and Wiemer, S. (2018). The effect of a mainshock on the size distribution of the aftershocks. *Geophysical Research Letters*, 45(24): 13,277–13,287.

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

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.

Hainzl, S. and Marsan, D. (2008). Dependence of the Omori-Utsu law parameters on main shock magnitude: Observations and modeling. *Journal of Geophysical Research: Solid Earth*, 113: B10309.

Hainzl, S., Sippl, C., and Schurr, B. (2019). Linear Relationship Between Aftershock Productivity and Seismic Coupling in the Northern Chile Subduction Zone. *Journal of Geophysical Research: Solid Earth*, 124(8): 8726–8738.

Hainzl, S. (2016a). 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. (2016b). Rate-dependent incompleteness of earthquake catalogs. *Seismological Research Letters*, 87(2A): 337–344.

Hainzl, S. (2021). ETAS-Approach Accounting for Short-Term Incompleteness of Earthquake Catalogs. *Bulletin of the Seismological Society of America.*

Hainzl, S., Zakharova, O., and Marsan, D. (2013). Impact of aseismic transients on the estimation of aftershock productivity parameters. *Bulletin of the Seismological Society of America*, 103(3): 1723–1732.

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

Harte, D. S. (2016). Model parameter estimation bias induced by earthquake magnitude cut-off. *Geophysical Journal International*, 204(2): 1266–1287.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*, volume 43. CRC press.

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

Hawkes, A. G. (2018). Hawkes processes and their applications to finance: a review. *Quantitative Finance*, 18(2): 193–198.

Helmstetter, A. and Sornette, D. (2003). Foreshocks explained by cascades of triggered seismicity. *Journal of Geophysical Research: Solid Earth*, 108(B10).

Hurvich, C. M., Simonoff, J. S., and Tsal, C.-L. (1998). Smoothing parameter selection in non-parametric regression using an improved Akaike information criterion. *JRSSB*, 60: 271–293.

Jackson, C. H. (2016). Flexsurv: A platform for parametric survival modeling in R. *Journal of Statistical Software*, 70(8).

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

Kagan, Y. Y. and Jackson, D. D. (1999). Worldwide doublets of large shallow earthquakes. *Bulletin of the Seismological Society of America*, 89(5): 1147–1155.

Kagan, Y. Y. (2017). Earthquake number forecasts testing. *Geophysical Journal International*, 211(1): 335–345.

Kagermanov, A. and Gee, R. (2019). Cyclic pushover method for seismic assessment under multiple earthquakes. *Earthquake Spectra*, 35(4): 1541–1558.

Klein, J. and Moeschberger, M. (2003). *Survival Analysis: Techniques for Censored and Truncated Data.*

Lay, T. and Kanamori, H. (1980). Earthquake doublets in the Solomon Islands. *Physics ofthe Earth and Planetary Interiors*, 21: 283–304.

Lippiello, E., Cirillo, A., Godano, G., Papadimitriou, E., and Karakostas, V. (2016). Real-time forecast of aftershocks from a single seismic station signal. *Geophysical Research Letters*, 43(12): 6252–6258.

Marsan, D. and Helmstetter, A. (2017). How variable is the number of triggered aftershocks? *Journal of Geophysical Research: Solid Earth*, 122(7): 5544–5560.

Marsan, D. and Ross, Z. E. (2021). Inverse Migration of Seismicity Quiescence During the 2019 Ridgecrest Sequence. *Journal of Geophysical Research: Solid Earth*, 126(3): 1–13.

Marsan, D. and Lengliné, O. (2008). Extending earthquakes' reach through cascading. *Science*, 319(5866): 1076–1079.

Matlab. (2019). *MATLAB version 9.6.0.1174912 (R2019a) Update 5*. The Mathworks, Inc., Natick, Massachusetts.

McGuire, R. K. (2008). Probabilistic seismic hazard analysis: Early history. *Earthquake Engineering and Structural Dynamics*, 37: 329–338.

Mitchell-Wallace, K., Jones, M., Hillier, J., and Foote, M. (2017). Natural catastrophe risk management and modelling: A practitioner's guide. *John Wiley & Sons*.

Mizrahi, L., Nandan, S., and Wiemer, S. (2021). Embracing Data Incompleteness for Better Earthquake Forecasting. *Journal of Geophysical Research: Solid Earth*, 126(12): 1–26.

Molkenthin, C., Donner, C., Reich, S., Zöller, G., Hainzl, S., Holschneider, M., and Opper, M. (2022). GP-ETAS: semiparametric Bayesian inference for the spatio-temporal epidemic type aftershock sequence model. *Statistics and Computing*, 32(2): 1–25.

Nandan, S., Ouillon, G., and Sornette, D. (2019). Magnitude of earthquakes controls the size distribution of their triggered events. *Journal of Geophysical Research: Solid Earth*, 124(3): 2762–2780.

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.

Omi, T., Ogata, Y., Hirata, Y., and Aihara, K. (2013). Forecasting large aftershocks within one day after the main shock. *Scientific Reports*, 3: 1–7.

Omi, T., Ogata, Y., Hirata, Y., and Aihara, K. (2014). Estimating the ETAS model from an early aftershock sequence. *Geophysical Research Letters*, (41): 850–857.

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.

Papadopoulos, A. N., Bazzurro, P., and Marzocchi, W. (2020). Exploring probabilistic seismic risk assessment accounting for seismicity clustering and damage accumulation: Part I. Hazard analysis. *Earthquake Spectra*.

Papadopoulos, A. N., Eeri, M., and Eeri, M. (2021). Exploring probabilistic seismic risk assessment accounting for seismicity clustering and damage accumulation : Part II . Risk analysis.

Pope, N. and Mooney, W. D. (2020). Coulomb stress models for the 2019 Ridgecrest, California earthquake sequence. *Tectonophysics*, 791: 228555.

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Schneider, M. (2021). *Improving Uncertainty Quantification and Visualization for Spatiotemporal Earthquake Rate Models for the Pacific Northwest.* Dissertation, University of Washington.

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.

Shebalin, P. N., Baranov, S. V., and Dzeboev, B. A. (2018). The law of the repeatability of the number of aftershocks. *Doklady Earth Sciences*, 481(1): 963–966.

van Stiphout, T., Zhuang, J., and Marsan, D. (2012). Seismicity declustering. *Community Online Resource for Statistical Seismicity Analysis.*

Tahir, M., Grasso, J. R., and Amorse, D. (2012). The largest aftershock: How strong, how far away, how delayed? *Geophysical Research Letters*, 39(4): 1–5.

Tench, S., Fry, H., and Gill, P. (2016). Spatio-temporal patterns of IED usage by the Provisional Irish Republican Army. *European Journal of Applied Mathematics*, 27(3): 377–402.

Therneau, T. (2016). survival: Survival Analysis, R package version 2.39-3.

Uhrhammer, R. A. (1986). Characteristics of northern and central California seismicity. *Earthquake Notes*, 57(1): 21.

Utsu, T., Ogata, Y., and Matsu'ura, R. S. (1995). The centenary of the Omori formula for a decay law of aftershock activity. *J. Phys. Earth*, 43: 1–33.

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.

Wells, D. L. and Coppersmith, K. J. (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacements. *Bulletin of the Seismological Society of America*, 84(4): 974–1002.

Wetzler, N., Brodsky, E. E., and Lay, T. (2016). Regional and stress drop effects on aftershock productivity of large megathrust earthquakes. *Geophysical Research Letters*, 43(23): 12,012–12,020.

Wood, S. (2017). Generalized Additive Models: An Introduction with R. *Boca Raton, FL: CRC Press.*

Wu, Z. (2010). A hidden Markov model for earthquake declustering. *Journal of Geophysical Research: Solid Earth*, 115(3).

Yip, C., Ng, W., and Yau, C. (2018). A hidden Markov model for earthquake prediction. *Stoch Environ Res Risk Assess*, 32: 1415–1434.

Zakharova, O., Hainzl, S., and Bach, C. (2013). Seismic moment ratio of aftershocks with respect to main shocks. *Journal of Geophysical Research: Solid Earth*, 118(11): 5856–5864.

Zhang, L., Werner, M. J., and Goda, K. (2018). Spatiotemporal seismic hazard and risk assessment of aftershocks of M 9 megathrust earthquakes. *Bulletin of the Seismological Society of America*, 108(6): 3313–3335.

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., Ogata, Y., and Vere-Jones, D. (2004). Analyzing earthquake clustering features by using stochastic reconstruction. *Journal of Geophysical Research: Solid Earth*, 109: B05301.

Zhuang, J., Ogata, Y., and Wang, T. (2017). Data completeness of the Kumamoto earthquake sequence in the JMA catalog and its influence on the estimation of the ETAS parameters. *Earth, Planets and Space*, 69: 36.

Zöller, G. (2018). A Statistical Model for Earthquake Recurrence Based on the Assimilation of Paleoseismicity, Historic Seismicity, and Instrumental Seismicity. *Journal of Geophysical Research: Solid Earth*, 123(6): 4906–4921.

Eidesstattliche Versicherung (Affidavit)

(Siehe Promotionsordnung vom 12. Juli 2011, § 8 Abs. 2 Pkt. 5)

Hiermit erkläre ich an Eides statt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

München, den 20.05.2022

Christian Grimm