

DISS. ETH NO. 28640

TOWARDS NEXT GENERATION
TIME-DEPENDENT EARTHQUAKE
FORECASTING

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

Leila MIZRAHI
MSc UZH in Mathematics, University of Zurich
born on 20.03.1992

accepted on the recommendation of

Prof. Dr. Stefan WIEMER
Dr. Shyam NANDAN
P.D. Dr. Sebastian HAINZL

November 2022

ACKNOWLEDGMENTS

First and foremost, I thank Stefan Wiemer, my supervisor, for giving me the opportunity of doing this PhD. For trusting that I can become a seismologist despite my non-seismological background, and for supporting my various earthquake-related endeavors on the way there. Thank you for your feedback that always allows me to view things from a much wider perspective. At the same time, I highly value the freedom you give me to pursue the topics I am most passionate about, even if not all of them are according to the initial plan.

I thank my second supervisor, Shyam Nandan, for transferring a huge amount of knowledge to me, be it about how to make my Python code faster (although using Matlab would always be your preferred way), about where to find good coffee around the NO building, or of course about anything related to ETAS. As soon as you are presented with a tricky question, you are enthusiastically discussing solutions with me, and often times putting everything into a philosophical context. I deeply appreciate these discussions, your sharing of ideas, and the time you make for me. Without you this thesis would be a very different one today.

Thank you to Yehuda Ben-Zion for hosting me at USC during times of Covid, for allowing me to travel and make connections during a time when travelling and making connections was not easy. Thank you for encouraging me to understand why things aren't working in order to make them work, and for the opportunity to get to know Californian seismicity first hand.

Thank you Bill Savran, for sharing your immense knowledge and experience with the topic of earthquake forecasting with me. Our discussions on Chapter 5 but also other, more or less closely related topics, always feel symbiotic and meaningful. I also thank you for being my friend, for going to the beach with me even on “bad weather” [*sic*] days, and for introducing me to the concept of dive bars.

I thank Sebastian Hainzl for taking the role of the external examiner at my defense, and I thank Andreas Fichtner for chairing it.

An incredible thank you and a big hug goes to Pilar Sánchez-Pastor, for being my office mate but also my mentor in anything related to academic etiquette. Thank you for reminding me to add my co-authors' names to my first ever presentation, and for helping me formulate rude emails without sounding rude. But more importantly, for being my friend, my companion with whom I shared countless hours in the office, in the bar, or in the middle of nowhere in Hengill getting lost in the fog, discussing life, love, and science.

Thank you Sydney Luca-Lion, for sharing our WG, and therefore our office, during that time when we were not allowed to leave it. You made this time tremendously more enjoyable and less isolating than it could have been. Thank you for the many coffee breaks on the roof terrace, for making the most delicious cookies, and sharing them, for our Wednesday lunch adventures when we would leave the house and realize other people still exist (!). I could certainly not have wished for a better person to stay at home with.

Thank you Sofia Brinck, for living the L.A. experience with me, on campus, at the beach, in traffic, and at home. For being the foreigner with me on Halloween, Thanksgiving, or when decorating the Christmas tree. We definitely immersed well. Also, you may not be

aware of it, but the key idea to Chapter 5 was born during our (thankfully joint) quarantine, when I was trying to put my knotty thoughts into simple words to let you know what I am doing all the time up there.

A huge thank you to my family: My parents Barbara Fuchs and David Mizrahi, my brother Kilian Mizrahi, my grandparents, Marie Fuchs-Fluri, Andrea Fuchs, and Clara Mizrahi, for raising me to be a curious person who finds joy in solving problems, and who is convinced that she can find a solution, which, in the worst case, is to re-formulate the problem. For supporting and encouraging me to pursue whatever dream I can think of, throughout my life. For trusting in me and for listening, for sharing sad moments and happy ones, for reminding me of where I come from and how I arrived at where I am today.

Thank you to my friends, peers, colleagues, teachers, role models, mentors: Anna Ehrler, Antonio Rinaldi, Ben Melechin, Christian Grimm, Conni Mertens, Danielle Bazzi, Dominik Inauen, Erich Weihs, Federica Lanza, Florian Fuchs, Florian Haslinger, Gwen Gubler, Johannes Abegglen, John Platt, Jonas Ruh, Jonathan Müller, Judith Egli, Katinka Tuinstra, Kilian Ashley, Laurentiu Danciu, Lauro Böni, Leandra Eberle, Luigi Passarelli, Lukas Nibourel, Martina Wundling, Max Klingler, Men-Andrin Meier, Michael Strupler, Michelle Meier, Nicola Gruber, Nicolas Schmid, Oleg Zlydenko, Pascal Graf, Paulina Janusz, Patrick Zöchbauer, Phil Kong, Philipp Kästli, Philippe Roth, Priyanka Soni, Rahel Blank, Roman Flury, Roman Küpper, Romano Meier, Sabine Brühwiler, Sämi Nussbaum, Dr. Thomas Wolters (the only real doctor here), Urs Vögeli, Vanille Ritz, Yara Rossi, Yifan Yin. I could write another thesis describing in detail how grateful I am to each of you. I hope you'll understand that I prefer to thank you collectively. For discussing science, wondering, roaming, complaining and celebrating with me. For occasionally being math nerds with me, and for helping out a mathematician who wants to understand the Earth. For your highly valued advice, and for valuing my advice when I give it. For the countless times you make me laugh. For making the past three years what they have been, at work and off work.

Thank you also to The Pohl family: Chris Carlson, Demetrius, Yani, and Jordan Pohl, for accommodating me in L.A., for teaching me how not to be killed riding a bike in this city, for explaining American football, Los Angeles, its people and its history to me. For sharing your captivating stories about getting lost in Chile, or Tarantino's favorite kind of Champagne. For letting me be a part of your lives during the months I was there.

Thank you Judith Lion for giving me a hideaway in your beautiful home in Santa Barbara. I will never forget the feeling of relaxation and relief I felt the first time we sat on your terrace, with a glass of Prosecco and elderberry syrup, and the scent of the pepper tree in the air. Thank you for naturally making me part of the family on Christmas, and generally throughout my time in California. Thank you for taking me on hikes, for discussing Swiss and American culture, and for becoming my friend and later on my Gotti by choice!

And lastly, thank you Paul Hammer, for everything. For our many shared adventures and for unconditionally supporting me on my very personal adventure of this PhD. For your patience, your advice, your planning, your ideas, your love. For your feedback that is often more honest than hoped for, but, admittedly, always useful. For being who you are, and for sharing your life and your view of the world with me. I love you. And soon I'll be able to tell you about things that are not related to ETAS.

CONTENTS

Acknowledgments	i
Abstract	v
Zusammenfassung	vii
1 Introduction	1
1.1 Basic statistical seismology	2
1.2 Probabilistic Seismic Hazard Analysis	3
1.3 The Epidemic-Type Aftershock Sequence model	4
1.4 Earthquake forecasting	5
1.5 Thesis outline	7
2 The Effect of Declustering on the Size Distribution of Mainshocks	13
Abstract	14
2.1 Introduction	14
2.2 Data	17
2.3 Method	18
2.4 Results and discussion	20
2.5 Conclusion	24
2.6 Data and resources	25
Acknowledgments	26
3 Embracing Data Incompleteness for Better Earthquake Forecasting	31
Abstract	32
Plain language summary	32
3.1 Introduction	32
3.2 Data	35
3.3 Model	35
3.4 Synthetic tests	42
3.5 Application to California	48
3.6 Pseudo-prospective forecasting experiments	54
3.7 Conclusion	59
Acknowledgments	60
4 Developing and Testing ETAS-Based Earthquake Forecasting Models for Switzerland	67
Abstract	68
4.1 Introduction	68
4.2 Selection of catalog region, time window, magnitude range	69
4.3 From catalog to forecast	72
4.4 Testing of model variants	75
4.5 Results and discussion	79

4.6	Conclusions	85
4.7	Data and resources	86
	Acknowledgments	86
5	Question-Driven Ensembles of Flexible ETAS Models	91
	Abstract	92
5.1	Introduction	92
5.2	Flexible ETAS models	94
5.3	Forecasting experiments	99
5.4	Results and discussion	101
5.5	Conclusions	106
5.6	Data and resources	106
	Acknowledgments	106
6	Conclusions and Outlook	111
	Appendices	119
A	Supporting Information for	
	Chapter 2	121
A.1	b -value estimation	121
A.2	Completeness magnitude	122
A.3	Detailed description of declustering methods	123
A.4	ETAS model	125
A.5	Sensitivity of b -value to m_c	130
A.6	Supplemental figures	130
A.7	Supplemental tables	132
B	Supporting Information for	
	Chapter 3	133
B.1	Catalog simulation	133
B.2	Forecast evaluation	134
C	Supporting Information for	
	Chapter 4	137
C.1	Supplemental figures	137
D	Supporting Information for	
	Chapter 5	145
D.1	Polygons	145
D.2	Inverted parameters	147
E	GitHub repositories	149
E.1	lmizrahi/etas	149
E.2	lmizrahi/petai	150
	Curriculum Vitae	

ABSTRACT

The ancient problem of earthquake prediction is simultaneously fascinating and disheartening. Disheartening, because after millennia of observing and documenting earthquakes, we are still caught by surprise every time tremendous amounts of energy happen to be released from the Earth's crust, destroying buildings, entire cities, and lives in the process. Fascinating, because the seemingly chaotic occurrence of earthquakes does exhibit astonishingly patterned behavior.

Instead of earthquake prediction, which is commonly viewed to be the specification of location, time, and magnitude of the next devastating event, the prevalent practice today is that of earthquake forecasting. That is, for a pre-specified space-time-magnitude domain, the earthquake probability is calculated, ideally using all knowledge regarding occurrence patterns of seismic events.

Two main branches of earthquake forecasting can be distinguished: Probabilistic Seismic Hazard Analysis (PSHA) combines time-independent earthquake probabilities with expected ground motions to form the basis of long-term protective measures in the form of building codes. Time-dependent operational earthquake forecasting (OEF) on the other hand aims to capture temporal fluctuations of earthquake probabilities based on which short-term protective measures mostly affecting people's behavior for a narrow period of time can be taken.

While PSHA is widely used around the world, OEF is by and large still in its infancy. Perhaps in its adolescence; models for time-dependent earthquake forecasting exist since decades and have been thoroughly tested ever since they were first described. All the more surprising is the lack of operational implementations of such models, with few, relatively recent exceptions.

The most prominent models for time-dependent earthquake forecasting are Epidemic-Type Aftershock Sequence (ETAS) models. They model seismicity as the sum of background earthquakes and their cascades of aftershocks. Aftershock triggering is described using few empirically derived laws regarding the number of triggered events as well as their occurrence in space and time around their parent event. Earthquake magnitudes follow, like many other quantities in nature, a power law.

In this thesis, the advancement of the field of earthquake forecasting is approached from multiple perspectives, with ETAS models taking a key role throughout.

First, we address the topic of declustering, that is, identifying and removing earthquake clusters from catalogs, in the context of PSHA. We systematically assess the effect of declustering on the size distribution of the remaining mainshocks. By declustering synthetically generated earthquake catalogs, and afterwards calculating their b -values, we find that declustering introduces a bias to the earthquake size distribution of a catalog. This highlights the problematic practice of using declustered catalogs for PSHA.

A possible approach to PSHA which does not rely on declustering is to quantitatively model earthquake clusters instead of removing them. The ETAS model provides a simple method to characterize the clustering behavior of a region through a set of parameters to be

calibrated using a complete earthquake catalog of that region. Because the completeness of earthquake catalogs can vary in short-term and long-term time horizons for distinct reasons, we propose two extended parameter calibration techniques suited to deal with each of these cases. Both methods are shown to accurately invert the data-generating parameters from synthetic catalogs, and one of the methods is further tested in pseudo-prospective forecasting experiments in California. The results thereof reveal that accounting for short-term aftershock incompleteness when forecasting yields significantly improved forecasts. This improvement vanishes as the difference in magnitude between the newly included small events and the target forecasted events increases, which is possibly explained through a tendency of earthquakes to trigger similarly sized aftershocks.

After discussing and refining existing methods for earthquake forecasting, we commit to advancing the operationalization of time-dependent earthquake forecasting in Switzerland. We develop six variants of ETAS models using the Swiss earthquake catalog, additional information extracted from the time-independent SUIhaz2015 model, or generic ETAS parameters calibrated on Californian data. One of the models uses the extended parameter calibration technique mentioned earlier. In seven-day pseudo-prospective forecasting experiments and 30- and 50-year retrospective consistency tests, we find that the purely ETAS-based models are best suited for the first Swiss OEF system. The poor performance of the model using Californian ETAS parameters highlights the importance of locally applicable parameter inversion methods.

Finally, we propose a novel approach to combine existing forecasting models into new ones. A question-driven ensemble (QDE) model is constructed by combining parameters of ingredient models, where the rules to combine parameters are driven by questions about the number of expected events, the spatio-temporal distribution of forecasted background earthquakes, and the spatio-temporal distribution of forecasted aftershocks. A QDE model can be viewed as a model which answers each question with a different ingredient model. We first describe flexible ETAS (fETAS) models which use nonparametric formulations of background seismicity and aftershock productivity during parameter inversion. The QDE approach is then tested by comparing combinations of fETAS models in pseudo-prospective forecasting experiments in Southern California and Italy. Our results show significant superiority of certain QDE models compared to their ingredient models, and we find striking similarities between the results in the two regions. This emphasizes the usefulness of the approach not only for the development of future earthquake forecasting models, but also for understanding strengths and weaknesses of existing ones.

ZUSAMMENFASSUNG

Das uralte Problem der Erdbebenvorhersage ist faszinierend und entmutigend zugleich. Entmutigend, weil wir nach jahrtausendelangem Beobachten und Dokumentieren von Erdbeben immer noch davon überrascht werden, wenn plötzlich enorme Mengen an Energie von der Erdkruste freigesetzt werden und dabei Gebäude, ganze Städte und Leben zerstört werden. Faszinierend, weil das scheinbar chaotische Auftreten von Erdbeben doch erstaunliche Regelmässigkeiten aufweist.

Anstelle der Erdbebenvorhersage, die üblicherweise als die Spezifikation von Ort, Zeitpunkt und Magnitude des nächsten verheerenden Ereignisses definiert wird, ist die heute gängige Praxis diejenige der Erdbebenprognose. Für einen vorgängig spezifizierten Raum-Zeit-Magnituden-Bereich wird also die Wahrscheinlichkeit eines Erdbebens berechnet, bestenfalls unter Berücksichtigung des gesamten verfügbaren Wissens über die Auftretensmuster solcher seismischen Ereignisse.

Zwei Hauptzweige der Erdbebenprognose können unterschieden werden: Die probabilistische seismische Gefährdungsanalyse (Probabilistic Seismic Hazard Analysis, PSHA) kombiniert zeitunabhängige Erdbebenwahrscheinlichkeiten mit zu erwartender Bodenbewegung und bildet so die Basis für Langzeitschutzmassnahmen in der Form von Baunormen. Die zeitabhängige operative Erdbebenprognose (operational earthquake forecasting, OEF) hingegen versucht zeitliche Veränderungen der Erdbebenwahrscheinlichkeiten zu erfassen, aufgrund derer kurzfristige Schutzmassnahmen ergriffen werden können, die vor allem das Verhalten von Personen für beschränkte Zeiträume beeinflussen.

Während PSHA weltweit breit eingesetzt wird, steckt OEF noch weitestgehend in Kinderschuhen. Vielleicht auch im Jugendalter; Modelle für zeitabhängige Erdbebenprognosen gibt es seit Jahrzehnten, und diese wurden seit ihrer ersten Beschreibung immer wieder gründlich getestet. Umso überraschender ist das Fehlen von operationellen Implementierungen solcher Modelle, von wenigen, relativ jungen Ausnahmen abgesehen.

Die bedeutendsten Modelle für zeitabhängige Erdbebenprognosen sind ETAS (Epidemic-Type Aftershock Sequence) Modelle. Sie modellieren Seismizität als die Summe von Hintergrunderdbeben und deren Kaskaden von Nachbeben. Das Auslösen von Nachbeben wird durch wenige, empirisch erlangte Gesetze bezüglich der Anzahl Nachbeben, sowie derer Auftreten in Raum und Zeit um ihr Vorgängerbeben beschrieben. Erdbebenmagnituden folgen, wie viele Grössen in der Natur, einem Potenzgesetz.

In dieser Dissertation wird die Weiterentwicklung des Gebietes der Erdbebenprognose aus verschiedenen Perspektiven angegangen, wobei ETAS-Modelle eine durchgehend zentrale Rolle spielen.

Als erstes behandeln wir die Thematik des Entclusterns, also des Erkennens und Entfernens von Erdbebenclustern, im Kontext von PSHA. Wir untersuchen den Effekt des Entlusterns auf die Stärkenverteilung der übrigbleibenden Hauptbeben systematisch. Indem wir synthetisch generierte Erdbebenkataloge entclustern und daraufhin deren b -Werte berechnen, stellen wir fest, dass das Entclustern einen Verzerrungseffekt auf die Magnitudenverteilung eines Erdbebenkatalogs hat. Dies unterstreicht die Problematik der gängigen Praxis des Verwendens von entclusterten Katalogen für PSHA.

Ein möglicher Ansatz für PSHA, der nicht auf das Entclustern angewiesen ist, ist das quantitative Modellieren von Erdbebenclustern anstelle derer Entfernung. Das ETAS-Modell bietet eine einfache Methode um das Clusteringverhalten einer Region zu charakterisieren, indem einige Parameter anhand eines vollständigen Erbebenkatalogs dieser Region kalibriert werden. Da die Vollständigkeit eines Erdbebenkatalogs in kurz- und langfristigen Zeiträumen variieren kann, schlagen wir zwei erweiterte Kalibrierungsverfahren vor, die die jeweiligen Fälle behandeln. Es wird aufgezeigt, dass beide Methoden die datenerzeugenden Parameter von synthetischen Katalogen korrekt invertieren können, und eine der Methoden wird zusätzlich in pseudoprospektiven Prognoseexperimenten in Kalifornien getestet. Die Ergebnisse dieser Experimente zeigen, dass das Miteinkalkulieren von kurzfristiger Nachbebenunvollständigkeit (Short-Term Aftershock Incompleteness, STAI) signifikant verbesserte Prognosen liefert. Diese Verbesserung verschwindet mit steigender Magnitudendifferenz zwischen den neu miteinbezogenen Kleinbeben und den prognostizierten Zielbeben, was möglicherweise durch eine Tendenz von Erdbeben, ähnlich grosse Nachbeben auszulösen, erklärbar ist.

Nach der Diskussion und Verfeinerung existierender Erdbebenprognosemethoden, verpflichten wir uns dem Vorantreiben der Operationalisierung von zeitabhängigen Erdbebenprognosen in der Schweiz. Wir entwickeln sechs Varianten von ETAS-Modellen unter Verwendung des schweizerischen Erdbebenkatalogs, zusätzlicher Informationen des zeitunabhängigen SUThaz2015 Modells, oder generischer ETAS-Parameter, die mit kalifornischen Daten kalibriert wurden. Eines der Modelle verwendet das vorhin erwähnte erweiterte Kalibrierungsverfahren. In siebentägigen pseudoprospektiven Prognoseexperimenten sowie 30- und 50-jährigen retrospektiven Konsistenztests stellen wir fest, dass die rein ETAS-basierten Modelle am besten geeignet sind für das erste schweizerische OEF-System. Das schwache Abschneiden des Modells, welches kalifornische Parameter verwendet, hebt die Wichtigkeit lokal anwendbarer Parameterinversionsmethoden hervor.

Zuletzt schlagen wir einen neuartigen Ansatz vor anhand dessen neue Prognosemodelle aus bestehenden kombiniert werden können. Ein fragestellungsgetriebenes Ensemblemodell (question-driven ensemble model, QDE model) wird durch das Kombinieren der Parameter seiner Ursprungsmodelle kreiert, wobei die Regeln bezüglich der Kombination von Parametern auf Fragen zur Anzahl erwarteter Beben, zur räumlich-zeitlichen Verteilung von Hintergrundbeben, und zur räumlich-zeitlichen Verteilung von Nachbeben basieren. Ein QDE-Modell kann als ein Modell angesehen werden, welches jede Frage mit einem anderen Ursprungsmodell beantwortet. Wir beschreiben zunächst flexible ETAS (fETAS) Modelle, welche in der Parameterinversion nichtparametrische Formulierungen der Hintergrundseismizität und Nachbebenproduktivität verwenden. Der QDE-Ansatz wird anschliessend getestet, indem verschiedene Kombinationen von fETAS-Modellen in pseudoprospektiven Prognoseexperimenten in Südkalifornien und Italien verglichen werden. Unsere Ergebnisse zeigen eine signifikante Überlegenheit gewisser QDE-Modelle im Vergleich mit deren Ursprungsmodellen, und wir stellen bemerkenswerte Ähnlichkeiten zwischen den Resultaten in den zwei Regionen fest. Dies betont den Nutzen dieses Ansatzes, nicht nur für das Entwickeln künftiger Erdbebenprognosemodelle, sondern auch für ein besseres Verständnis von Stärken und Schwächen bestehender Modelle.

CHAPTER 1

INTRODUCTION

“Is nature chaotic, or is nature patterned?”

— Douglas R. Hofstadter,
Gödel, Escher, Bach: an Eternal Golden Braid

Humans are curious. We want to understand the world around us. When we lay beneath the clear night sky, we do not just enjoy the view. We wonder what those shiny dots are and how they were formed. We wonder why the Sun rises every morning. We wonder why objects fall to the ground when we stop holding on to them. We wonder what the ground on which they fall is made of. We wonder how life came to existence, how pea plants inherit traits from their parents. We wonder whether we have free will, or whether our thoughts are mere consequences of natural laws acting on the current state of the world. We wonder how many prime numbers there are, what the price of the Bitcoin is tomorrow, and whether there is an alternative to capitalism. And we wonder how earthquakes occur.

One way to try to understand the world is to find patterns. There are patterns in our behavior, there are patterns in our thoughts, there are patterns in nature. We use patterns to make predictions about the future. Though it has been wondered whether observing the past allows us to make conclusions about the future (Hume, 1739). Or, to come back to the question at the very beginning of this chapter, whether all observed patterns are purely a result of chaos and chance. So as not to be halted by philosophical questions already, we shall for now concern ourselves with finding patterns. With their propensity to destroy and their intriguing property of occurring in seemingly chaotic yet beautifully patterned ways, earthquakes are a fascinating topic to study.

The focus of this thesis will lie on finding patterns related to earthquakes.

Already in 1756, Immanuel Kant studied patterns in the occurrence of ground shaking, correctly noting, for instance, that the earthquake frequency of a region is *not* determined by the degree of Christianness of its population (Kant, 1756). He noted that it is, however, related to the region’s proximity to mountain ranges. He noted that sometimes, the movements of the ground are accompanied by movements of the ocean, and that earthquakes are somehow related to fire-spitting mountains. While many of his observations were correct, we know today that his theory of underground caves containing flammable gases, whose explosion causes the earth to shake, was not. It wasn’t until 1912 though that Alfred Wegener formulated the hypothesis of moving crust to explain the formation of continents (Wegener, 1912). Due to the lack of a physical explanation for such “continental drift”, his theory was widely disputed for another few decades. Upon the discovery of magnetic patterns on the ocean bottom which supported the idea of seafloor spreading (Dietz, 1961; Morley and Larochelle, 1964; Vine and Matthews, 1963), the modern concept of plate tectonics

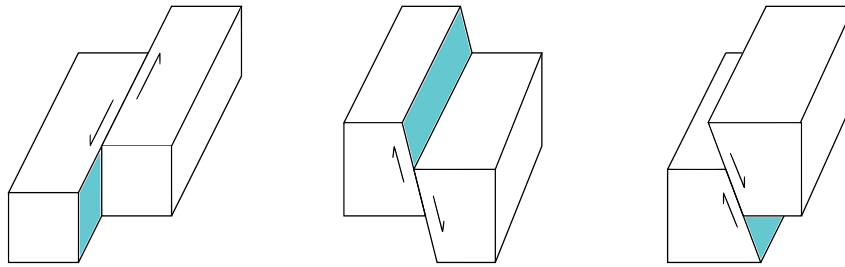


Figure 1.1: Schematic illustration of blocks moving relative to each other, representing possible earthquake mechanisms.

as the theory of rigid blocks of crust moving relatively to each other over the mantle, was independently formulated by Morgan (1968) and McKenzie and Parker (1967).

Nowadays, it is undisputed that the main driver of seismic activity is plate tectonics, and we know much more about how earthquakes nucleate than Kant did. We know that earthquakes occur when two blocks of rock suddenly move relatively to each other (see Figure 1.1). We know that when an earthquake occurs, P- and S-waves travel through the Earth's interior at different speeds. We build instruments to record those waves at the surface, and we use those recordings to find out where and when an earthquake occurred, what its direction of movement was, and what size it had. We can even infer knowledge about the structure of the Earth's interior from seismic recordings, or the interior of Mars for that matter. (Aki and Richards, 2002; Stein and Wysession, 2009; Shearer, 2019; Stähler et al., 2021)

1.1 Basic statistical seismology

As much as we do know about earthquakes, we are still far from being able to predict them, even though, as noted earlier, predicting the future is one of the key activities us humans would like to do after finding patterns in nature. In the case of earthquakes in particular, a successful prediction would be especially desirable as it could save the lives and homes of thousands of people in a single earthquake. In fact, however, we are so far from predicting them, that sometimes we unwillingly induce them (Häring et al., 2008; Foulger et al., 2018).

To address the goal of earthquake prediction, what we can do is to specifically look for patterns in the occurrence of earthquakes. Kant (1756) already noted that in some areas they occur more frequently than in others. Omori (1894) later more quantitatively described patterns in the temporal occurrence of earthquakes. Once a relatively big earthquake occurs, so-called *aftershocks* tend to follow. Most of them occur very soon after their parent event, and with time, their frequency decreases. The Omori-Utsu law (Omori, 1894; Utsu, 1961) expresses the rate $r(t)$ of aftershocks at a time delay t after the parent event as

$$r(t) = \frac{k}{(t + c)^p}, \quad (1.1)$$

with parameters k , c and p .

Gutenberg and Richter (1944) discovered an additional important pattern in the frequency of earthquakes in relation to their size. The Gutenberg-Richter (GR) law, possibly the most well-known law of statistical seismology, states that

$$N(m) = 10^{a-b \cdot m}, \quad (1.2)$$

the number $N(m)$ of earthquakes of magnitude $M \geq m$ decreases exponentially with m . The values a and b are commonly referred to as the a - and b -value, where the a -value determines the number of earthquakes of magnitude $M \geq 0$ in the region and time horizon intended to be described with Equation (1.2), and the b -value establishes the exponential dependency of $N(m)$ on m .

Given enough data, a - and b -value of any region can be estimated, and, as Kant would likely have suspected, the a -value can vary strongly between different regions. The b -value on the other hand tends to take values close to 1 (Y. Kagan, 1999), which means that for an increase of m by one magnitude unit, the number $N(m)$ of earthquakes larger than or equal to m decreases by a factor of roughly 10. Of course, b is not always exactly equal to 1, and dependencies of b on time, region, tectonic setting and various physical quantities have been proposed in the past (Henderson et al., 1992; Main et al., 1992; Frohlich and Davis, 1993; Wiemer and Wyss, 1997; Wyss et al., 1997; Schorlemmer et al., 2005; Petruccioli et al., 2019).

1.2 Probabilistic Seismic Hazard Analysis

Earthquake prediction is usually viewed as the specification of the time, location and magnitude of the next large event. Using the tools introduced thus far, we are closer to being able to produce what is generally called an earthquake forecast. Given certain assumptions, we can calculate the probability of an earthquake of a certain magnitude in a certain space-time domain.

Cornell (1968) described the basic approach to Probabilistic Seismic Hazard Analysis (PSHA) that is still applied today. In this approach, it is assumed that the long-term rate of earthquakes is constant, that their magnitudes follow the GR law, and that their occurrence times are representable by a stationary Poisson process. Expected ground motions are calculated based on the resulting earthquake probabilities to understand what forces our buildings should be able to withstand.

The assumption of Poissonianity is however not compatible with Omori's pattern regarding the temporal clustering behavior of earthquakes. Besides temporal clustering, aftershocks tend to also occur close in space to their parent event. We can describe the rate $r(x)$ of aftershocks at a distance x from the parent event as

$$r(x) = \frac{k}{(x+d)^q}, \quad (1.3)$$

with parameters k, d and q , where k can differ from the one in Equation (1.1). The parameter k in both Equations (1.1) and (1.3) reflects the dependency of the aftershock rate on the overall number of aftershocks of a parent event, its so-called *productivity*. For this productivity, there is yet another law describing its dependency on the magnitude of the parent event,

$$K(m) = k \cdot e^{\alpha \cdot m}, \quad (1.4)$$

with $K(m)$ denoting the productivity of an event of magnitude m , and k and α parameters. Again, k can differ from k in previous equations.

Equations (1.1-1.4) will be revisited later in this thesis, and slightly more complicated versions of them may be used. For now, they describe the clustering of earthquakes sufficiently well for us to recognize that earthquake occurrence times are not Poissonian. Additionally, regions with recent large sequences may be overrepresented in an earthquake catalog. For these reasons, the PSHA approach by Cornell (1968) is commonly done using earthquake catalogs that have been declustered (Pace et al., 2006; Wiemer et al., 2009; Beauval et al.,

2013; Field et al., 2014; Woessner et al., 2015; Meletti et al., 2017; Akinici et al., 2018; Petersen et al., 2018; Sesetyan et al., 2018; Waseem et al., 2019; Drouet et al., 2020). In this process, clusters of earthquakes are identified, and only the largest event of each cluster is kept in the remaining catalog.

What may sound simple is in reality not simple at all. To partition earthquakes into discrete clusters is a nontrivial task to which no objectively correct solution exists, which can lead to difficulties. In Chapter 2, we will focus on the effect of declustering on the size distribution of mainshocks, and the implications of these effects for PSHA.

1.3 The Epidemic-Type Aftershock Sequence model

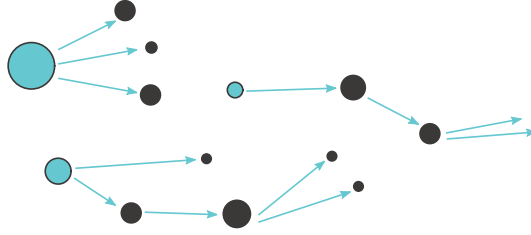


Figure 1.2: Schematic illustration of earthquake triggering in the ETAS model.

To avoid the problems posed by declustering, and to not neglect the seismic hazard that comes from aftershocks (Marzocchi and Taroni, 2014; Iervolino et al., 2018), one alternative is to understand clusters and quantitatively model their behavior instead of removing them. The Epidemic-Type Aftershock Sequence (ETAS) model, first introduced by Ogata (1988), aims to do exactly that. Its foundation are the empirical laws stated in Equations (1.1-1.4). In ETAS, each earthquake is considered to be either a background earthquake, or an aftershock of a previous one. Aftershock magnitudes are independent of the size of their parent event, and both follow the same GR distribution. All aftershocks can trigger aftershocks of their own, which can lead to cascades of aftershocks as illustrated in Figure 1.2. Background earthquakes are modeled, in the style of PSHA, to occur uniformly in time, while the distribution of aftershocks in number, space and time is given by the productivity, temporal, and spatial laws discussed earlier. We will use the formulation of Nandan et al. (2017), where the total rate of events $\lambda(x, y, t|\mathcal{H}_t)$ at location (x, y) and time t , provided the history \mathcal{H}_t , is given by

$$\lambda(x, y, t|\mathcal{H}_t) = \mu + \sum_{i:t_i < t} g(x - x_i, y - y_i, t - t_i), \quad (1.5)$$

where

$$g(\Delta x, \Delta y, \Delta t) = \frac{k_0 \cdot e^{a \cdot (m - m_{ref})} \cdot e^{-\Delta t / \tau}}{(\Delta x^2 + \Delta y^2 + d \cdot e^{\gamma \cdot (m - m_{ref})})^{1+\rho} \cdot (\Delta t + c)^{1+\omega}}. \quad (1.6)$$

Here, $g(\Delta x, \Delta y, \Delta t)$ is the overall rate of aftershocks of a parent earthquake with magnitude m , at a distance $(\Delta x, \Delta y)$, and a time delay Δt . Marked in blue is the part describing the productivity of an event of magnitude m . Note the term $m - m_{ref}$ which signals that we are interested in events above a certain reference magnitude m_{ref} . Marked in turquoise is the part describing the spatial distribution of aftershocks. As an additional feature not captured in Equation (1.3), this part also depends on the size m of the parent event. Marked in yellow is the part describing the temporal distribution of aftershocks, with an exponential

taper that was not present in Equation (1.1). It has the aim to avoid heavy tails which imply large portions of aftershocks at very long time delays. The parameter μ represents the rate of background earthquakes.

Using Equation (1.5), the expected seismicity rate at any time t in a region can be described. This requires knowledge of the parameters $\mu, k_0, a, \tau, c, \omega, d, \gamma$ and ρ , and the complete history \mathcal{H}_t of events that occurred prior to time t . The knowledge of ETAS parameters and the history \mathcal{H}_t are both hard to acquire. The latter depends on seismic instruments and our interpretation of their recordings, which realistically will never be perfect. The former are usually inferred from exactly these imperfect datasets.

More specifically, to invert the parameters of an ETAS model, and to calculate the expected rate of events, we need a catalog that is complete. An instrumental catalog can impossibly be complete in the sense that it can't contain earthquakes that occurred before the seismic network was operating. But, as the aftershock rates decrease with time, and especially thanks to the exponential taper $e^{-\Delta t/\tau}$, the contribution of events that lie very far in the past is negligible in the sum in Equation (1.5). A more important reason why a catalog could be incomplete is that even with a fully operational seismic network, small events are not always being detected.

The magnitude of completeness, m_c , of a catalog is usually defined as the magnitude above which all earthquakes are registered by the network. As it is impossible to know what was not observed, the estimation of m_c is another problem to which no objectively correct solution exists. The value of m_c is usually jointly estimated with the b -value (Rydelek and Sacks, 1989; Wiemer and Wyss, 1997), based on the assumption that a catalog is complete above the magnitude at which it starts exhibiting the GR-law behavior.

The term $m - m_{ref}$ in Equation (1.6) may have already given it away: In most formulations of the ETAS model, it is assumed that only earthquakes above a reference magnitude m_{ref} can trigger and be triggered, and the value of m_{ref} is generally set to be equal to m_c . As a consequence of this assumption, earthquakes that were triggered by events of smaller magnitudes will be interpreted as background earthquakes.

While the estimation of m_c is already relatively complex, more complexity is added to the problem due to the variation of m_c with time. In the long term, m_c can fluctuate with changes of the seismic network. In the short term, a main driver of m_c fluctuation is Short-Term Aftershock Incompleteness (STAI), which, broadly speaking, results when seismic waves of large earthquakes cover those of smaller ones (Y. Y. Kagan, 2004).

The completeness of a catalog is not only a requirement to calculate the expected seismicity rate of a region according to Equation (1.5), but also when estimating the model parameters. Thus, assuming a constant value for m_c for the catalog used to train an ETAS model forces a modeler to either temporally over- or underestimate its true value. In Chapter 3, we thoroughly address the topic of ETAS parameter inversion in the context of temporally varying catalog completeness. We propose extended parameter calibration methods to deal with short- and long-term fluctuations of m_c , and test them using synthetic experiments and in forecasting experiments with Californian data.

1.4 Earthquake forecasting

Equation (1.5) (together with a complete catalog and a set of ETAS parameters), in principle already provides an earthquake forecast. Note however that to calculate the seismicity rate at time t , all events prior to time t need to be known. Hence, $\lambda(x, y, t|\mathcal{H}_t)$ is only valid until time t^* , the time when the next event (after time t) occurs.

If one is interested in the expected rate of earthquakes to occur in the week, month, or in 50 years after time t , the cascades of aftershocks of all events which fall into the forecasting period are not captured by Equation (1.5). To account for these events in a forecast, we can simulate many (hundreds of thousands or even millions of) possible continuations of the current catalog. This includes simulating the aftershocks of events in the training catalog plus their cascades of aftershocks, and background events that occur during the forecasting period plus their cascades of aftershocks. With this approach, forecasts for periods of arbitrary length can be generated.

So, we have found a solution to the problem mentioned at the very end of Section 1.2. Instead of removing earthquake clusters, we model them, and to avoid a spatial bias in our long-term forecast, we simulate a large number of possible catalogs, each of which may contain large sequences in different places.

The core strength of ETAS lies in its ability to model aftershocks, which makes it particularly suited to model the shorter-term behavior of seismicity. ETAS models have been tested in various forecasting experiments (Woessner et al., 2011; Ogata et al., 2013; Strader et al., 2017; Cattania et al., 2018; Taroni et al., 2018; Nandan et al., 2019; Mancini et al., 2019; Mancini et al., 2020; Savran et al., 2020) and are used or considered for operational earthquake forecasting (OEF) (Marzocchi et al., 2014; Rhoades et al., 2016; Field et al., 2017; van der Elst et al., 2022). They represent the undisputed state-of-the-art of modern earthquake forecasting.

In Chapter 4, we describe the development of a first time-dependent earthquake forecasting model for Switzerland. Using 7-day forecasting experiments and 30 to 50-year consistency tests, we compare several ETAS variants, including one developed in Chapter 3, to the current time-independent state-of-the-art.

Certainly, ETAS models are a major simplification of reality. I have already mentioned a few model assumptions such as the existence of a minimum triggering magnitude or the time-independence of m_c in general formulations of ETAS. Other model assumptions such as the spatially isotropic distribution of aftershocks, or a spatially and temporally uniform background rate, are known to cause biases in the estimated parameters (Hainzl et al., 2008; Hainzl et al., 2013; Seif et al., 2017; Nandan et al., 2021).

The data used to calibrate ETAS models is by no means perfect; earthquake catalogs suffer from various kinds of uncertainties and inconsistencies. Researchers continuously enhance the methods to locate earthquakes, calculate their magnitudes and generally improve earthquake catalog quality (e.g. Bagagli et al., 2019; Ross et al., 2019; Mousavi and Beroza, 2020).

But even with a perfectly estimated set of parameters, ETAS models would still provide the total expected seismicity rate at a given time and location of interest, and rely on the GR law to describe the distribution of magnitudes. That is, besides the few patterns incorporated in ETAS, namely those given in Equations (1.1-1.4), we still model the nature of earthquake occurrence to be chaotic. And it is not clear whether earthquakes will ever be more predictable than that.

So then, being humans, we are inclined to search for further patterns. We can search for different kinds of patterns motivated by how we plan to make use of them. The discovery of a certain type of patterns will satisfy our desire to better *understand* the world around us. The discovery of other types of patterns will allow us to *forecast* future earthquakes more accurately. Some patterns might be capable of both. In Chapter 5, we look for patterns of the latter kind by testing model combinations which can reveal strengths and weaknesses of ingredient models in an interpretable way.

1.5 Thesis outline

The remainder of this thesis is structured as follows.

In Chapter 2, we systematically assess the effect of declustering on the b -value of mainshocks. For this, we generate synthetic catalogs with known b -value using ETAS, and decluster them using commonly used methods with a wide range of hyperparameters. We then calculate the b -values of the declustered catalogs and compare them to the data-generating b -values.

In Chapter 3, we describe two methods for the estimation of ETAS parameters when m_c varies with time. The first method accepts $m_c(t)$ as an input and is especially suited to account for long-term changes in m_c . The second method generalizes the concept of m_c to address STAI. It considers a rate- and magnitude-dependent detection probability function, and jointly estimates ETAS parameters and a high-frequency detection probability at the time of each event. Both methods are then tested on synthetic data. Finally, the usefulness for forecasting of the second method is tested in a pseudo-prospective forecasting experiment in California.

In Chapter 4, we describe the development of ETAS-based time dependent earthquake forecasting models for Switzerland. Besides a basic ETAS model, we describe several variants of it: Some use information from the time-independent SUIhaz2015 model (Edwards et al., 2016), one uses generic model parameters calibrated for California, and one uses the extended calibration technique of Chapter 3 to estimate ETAS parameters with long-term changes of m_c . All model variants are finally tested in pseudo-prospective one-week forecasting experiments and 30- and 50-year retrospective consistency tests.

In Chapter 5, we present the novel approach of question-driven ensemble (QDE) modeling. Different variants of ETAS models are combined into new models by combining their parameters according to certain rules. These rules are driven by questions about the number of earthquakes expected by a given model, the spatio-temporal distribution of background events, and the spatio-temporal distribution of aftershocks. QDE models can be interpreted as models which answer each question with a different ingredient model. A suite of ensemble models is created through the combination of flexible ETAS models, and their performance is then compared in pseudo-prospective forecasting experiments for Southern California and Italy. Besides an optimized forecasting performance, this approach simultaneously provides insight into which model variants tend to answer which questions well, which hopefully enhances our understanding of seismogenesis.

I summarize the main conclusions of this thesis in Chapter 6 and give an outlook on promising future developments in the field of earthquake forecasting.

References

- Aki, K., & Richards, P. G. (2002). *Quantitative seismology*. 2
- Akinci, A., Moschetti, M. P., & Taroni, M. (2018). Ensemble smoothed seismicity models for the new italian probabilistic seismic hazard map. *Seismological Research Letters*, 89(4), 1277–1287. 4
- Bagagli, M., Molinari, I., Diehl, T., Kissling, E., et al. (2019). Towards a new semi-automated consistent multiphase picking algorithm for local and regional seismic networks. *Geophysical Research Abstracts*, 21. 6

- Beauval, C., Yepes, H., Palacios, P., Segovia, M., Alvarado, A., Font, Y., Aguilar, J., Troncoso, L., & Vaca, S. (2013). An earthquake catalog for seismic hazard assessment in ecuador. *Bulletin of the Seismological Society of America*, *103*(2A), 773–786. [3](#)
- Cattania, C., Werner, M. J., Marzocchi, W., Hainzl, S., Rhoades, D., Gerstenberger, M., Liukis, M., Savran, W., Christophersen, A., Helmstetter, A., et al. (2018). The forecasting skill of physics-based seismicity models during the 2010–2012 canterbury, new zealand, earthquake sequence. *Seismological Research Letters*, *89*(4), 1238–1250. [6](#)
- Cornell, C. A. (1968). Engineering seismic risk analysis. *Bulletin of the seismological society of America*, *58*(5), 1583–1606. [3](#)
- Dietz, R. S. (1961). Continent and ocean basin evolution by spreading of the sea floor. *Nature*, *190*(4779), 854–857. [1](#)
- Drouet, S., Ameri, G., Le Dortz, K., Secanell, R., & Senfaute, G. (2020). A probabilistic seismic hazard map for the metropolitan france. *Bulletin of Earthquake Engineering*, *18*(5), 1865–1898. [4](#)
- Edwards, B., Cauzzi, C., Danciu, L., & Fäh, D. (2016). Region-specific assessment, adjustment, and weighting of ground-motion prediction models: Application to the 2015 swiss seismic-hazard maps. *Bulletin of the Seismological Society of America*, *106*(4), 1840–1857. [7](#)
- Field, E. H., Arrowsmith, R. J., Biasi, G. P., Bird, P., Dawson, T. E., Felzer, K. R., Jackson, D. D., Johnson, K. M., Jordan, T. H., Madden, C., et al. (2014). Uniform california earthquake rupture forecast, version 3 (ucerf3)—the time-independent model. *Bulletin of the Seismological Society of America*, *104*(3), 1122–1180. [4](#)
- Field, E. H., Jordan, T. H., Page, M. T., Milner, K. R., Shaw, B. E., Dawson, T. E., Biasi, G. P., Parsons, T., Hardebeck, J. L., Michael, A. J., et al. (2017). A synoptic view of the third uniform california earthquake rupture forecast (ucerf3). *Seismological Research Letters*, *88*(5), 1259–1267. [6](#)
- Foulger, G. R., Wilson, M. P., Gluyas, J. G., Julian, B. R., & Davies, R. J. (2018). Global review of human-induced earthquakes. *Earth-Science Reviews*, *178*, 438–514. [2](#)
- Frohlich, C., & Davis, S. D. (1993). Teleseismic b values; or, much ado about 1.0. *Journal of Geophysical Research: Solid Earth*, *98*(B1), 631–644. [3](#)
- Gutenberg, B., & Richter, C. F. (1944). Frequency of earthquakes in california. *Bulletin of the Seismological Society of America*, *34*(4), 185–188. [2](#)
- Hainzl, S., Christophersen, A., & 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. [6](#)
- Hainzl, S., Zakharova, O., & 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. [6](#)
- Häring, M. O., Schanz, U., Ladner, F., & Dyer, B. C. (2008). Characterisation of the basel 1 enhanced geothermal system. *Geothermics*, *37*(5), 469–495. [2](#)
- Henderson, J., Main, I., Meredith, P., & Sammonds, P. (1992). The evolution of seismicity at parkfield: Observation, experiment and a fracture-mechanical interpretation. *Journal of Structural Geology*, *14*(8-9), 905–913. [3](#)
- Hume, D. (1739). A treatise of human nature (book 1). *JM Dent & Sons, London*. [1](#)
- Iervolino, I., Chioccarelli, E., & Giorgio, M. (2018). Aftershocks’ effect on structural design actions in italy. *Bulletin of the Seismological Society of America*, *108*(4), 2209–2220. [4](#)
- Kagan, Y. Y. (2004). Short-term properties of earthquake catalogs and models of earthquake source. *Bulletin of the Seismological Society of America*, *94*(4), 1207–1228. [5](#)
- Kagan, Y. (1999). Universality of the seismic moment-frequency relation. *Seismicity patterns, their statistical significance and physical meaning* (pp. 537–573). Springer. [3](#)

- Kant, I. (1756). Geschichte und naturbeschreibung der merkwürdigsten vorfälle des erdbebens welches an dem ende des 1755ten jahres einen grossen theil der erde erschüttert hat. *Band 1 Vorkritische Schriften I. 1747–1756*, 417–504. 1, 2
- Main, I. G., Meredith, P. G., & Sammonds, P. R. (1992). Temporal variations in seismic event rate and b-values from stress corrosion constitutive laws. *Tectonophysics*, 211(1-4), 233–246. 3
- Mancini, S., Segou, M., Werner, M., & Cattania, C. (2019). Improving physics-based aftershock forecasts during the 2016–2017 central italy earthquake cascade. *Journal of Geophysical Research: Solid Earth*, 124(8), 8626–8643. 6
- Mancini, S., Segou, M., Werner, M. J., & Parsons, T. (2020). The predictive skills of elastic coulomb rate-and-state aftershock forecasts during the 2019 ridgecrest, california, earthquake sequence. *Bulletin of the Seismological Society of America*, 110(4), 1736–1751. 6
- Marzocchi, W., & Taroni, M. (2014). Some thoughts on declustering in probabilistic seismic-hazard analysis. *Bulletin of the Seismological Society of America*, 104(4), 1838–1845. 4
- Marzocchi, W., Lombardi, A. M., & Casarotti, E. (2014). The establishment of an operational earthquake forecasting system in italy. *Seismological Research Letters*, 85(5), 961–969. 6
- McKenzie, D. P., & Parker, R. L. (1967). The north pacific: An example of tectonics on a sphere. *Nature*, 216(5122), 1276–1280. 2
- Meletti, C., Marzocchi, W., Albarello, D., D’Amico, V., Luzi, L., Martinelli, F., Pace, B., Pignone, M., Rovida, A. N., Visini, F., et al. (2017). The 2016 italian seismic hazard model. *16th World Conference on Earthquake Engineering*. 4
- Morgan, W. J. (1968). Rises, trenches, great faults, and crustal blocks. *Journal of Geophysical Research*, 73(6), 1959–1982. 2
- Morley, L., & Larochele, A. (1964). Paleomagnetism as a means of dating geological events. *Geochronology in Canada*, 8, 39–51. 1
- Mousavi, S. M., & Beroza, G. C. (2020). A machine-learning approach for earthquake magnitude estimation. *Geophysical Research Letters*, 47(1), e2019GL085976. 6
- Nandan, S., Ouillon, G., Sornette, D., & Wiemer, S. (2019). Forecasting the rates of future aftershocks of all generations is essential to develop better earthquake forecast models. *Journal of Geophysical Research: Solid Earth*, 124(8), 8404–8425. 6
- Nandan, S., Ouillon, G., Wiemer, S., & Sornette, D. (2017). Objective estimation of spatially variable parameters of epidemic type aftershock sequence model: Application to california. *Journal of Geophysical Research: Solid Earth*, 122(7), 5118–5143. 4
- Nandan, S., Ram, S. K., Ouillon, G., & Sornette, D. (2021). Is seismicity operating at a critical point? *Physical Review Letters*, 126(12), 128501. 6
- Ogata, Y., Katsura, K., Falcone, G., Nanjo, K., & Zhuang, J. (2013). Comprehensive and topical evaluations of earthquake forecasts in terms of number, time, space, and magnitude. *Bulletin of the Seismological Society of America*, 103(3), 1692–1708. <https://doi.org/10.1785/0120120063>. 6
- Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical association*, 83(401), 9–27. 4
- Omori, F. (1894). On after-shocks. *Seismological journal of Japan*, (19), 71–80. 2
- Pace, B., Peruzza, L., Lavecchia, G., & Boncio, P. (2006). Layered seismogenic source model and probabilistic seismic-hazard analyses in central italy. *Bulletin of the Seismological Society of America*, 96(1), 107–132. 3
- Petersen, M. D., Mueller, C. S., Moschetti, M. P., Hoover, S. M., Rukstales, K. S., McNamara, D. E., Williams, R. A., Shumway, A. M., Powers, P. M., Earle, P. S., et al. (2018). 2018 one-year seismic hazard forecast for the central and eastern united states from induced and natural earthquakes. *Seismological Research Letters*, 89(3), 1049–1061. 4

- Petrucelli, A., Gasperini, P., Tormann, T., Schorlemmer, D., Rinaldi, A. P., Vannucci, G., & Wiemer, S. (2019). Simultaneous dependence of the earthquake-size distribution on faulting style and depth. *Geophysical Research Letters*, *46*(20), 11044–11053. [3](#)
- Rhoades, D., Liukis, M., Christophersen, A., & Gerstenberger, M. (2016). Retrospective tests of hybrid operational earthquake forecasting models for canterbury. *Geophysical Journal International*, *204*(1), 440–456. [6](#)
- Ross, Z. E., Trugman, D. T., Hauksson, E., & Shearer, P. M. (2019). Searching for hidden earthquakes in southern california. *Science*, *364*(6442), 767–771. [6](#)
- Rydelek, P. A., & Sacks, I. S. (1989). Testing the completeness of earthquake catalogues and the hypothesis of self-similarity. *Nature*, *337*(6204), 251–253. [5](#)
- Savran, W. H., Werner, M. J., Marzocchi, W., Rhoades, D. A., Jackson, D. D., Milner, K., Field, E., & Michael, A. (2020). Pseudoprospective evaluation of ucerf3-etas forecasts during the 2019 ridgecrest sequence. *Bulletin of the Seismological Society of America*, *110*(4), 1799–1817. [6](#)
- Schorlemmer, D., Wiemer, S., & Wyss, M. (2005). Variations in earthquake-size distribution across different stress regimes. *Nature*, *437*(7058), 539–542. [3](#)
- Seif, S., Mignan, A., Zechar, J. D., Werner, M. J., & Wiemer, S. (2017). Estimating etas: The effects of truncation, missing data, and model assumptions. *Journal of Geophysical Research: Solid Earth*, *122*(1), 449–469. [6](#)
- Sesetyan, K., Demircioglu, M. B., Duman, T. Y., Can, T., Tekin, S., Azak, T. E., & Fercan, Ö. Z. (2018). A probabilistic seismic hazard assessment for the turkish territory—part i: The area source model. *Bulletin of Earthquake Engineering*, *16*(8), 3367–3397. [4](#)
- Shearer, P. M. (2019). *Introduction to seismology*. Cambridge university press. [2](#)
- Stähler, S. C., Khan, A., Banerdt, W. B., Lognonné, P., Giardini, D., Ceylan, S., Drilleau, M., Duran, A. C., Garcia, R. F., Huang, Q., et al. (2021). Seismic detection of the martian core. *Science*, *373*(6553), 443–448. [2](#)
- Stein, S., & Wysession, M. (2009). *An introduction to seismology, earthquakes, and earth structure*. John Wiley & Sons. [2](#)
- Strader, A., Schneider, M., & Schorlemmer, D. (2017). Prospective and retrospective evaluation of five-year earthquake forecast models for california. *Geophysical Journal International*, *211*(1), 239–251. [6](#)
- Taroni, M., Marzocchi, W., Schorlemmer, D., Werner, M. J., Wiemer, S., Zechar, J. D., Heiniger, L., & Euchner, F. (2018). Prospective csep evaluation of 1-day, 3-month, and 5-yr earthquake forecasts for italy. *Seismological Research Letters*, *89*(4), 1251–1261. [6](#)
- Utsu, T. (1961). A statistical study on the occurrence of aftershocks. *Geophys. Mag.*, *30*, 521–605. [2](#)
- van der Elst, N. J., Hardebeck, J. L., Michael, A. J., McBride, S. K., & Vanacore, E. (2022). Prospective and retrospective evaluation of the us geological survey public aftershock forecast for the 2019–2021 southwest puerto rico earthquake and aftershocks. *Seismological Society of America*, *93*(2A), 620–640. [6](#)
- Vine, F., & Matthews, D. (1963). Magnetic anomalies over oceanic ridges. *A century*. [1](#)
- Waseem, M., Lateef, A., Ahmad, I., Khan, S., & Ahmed, W. (2019). Seismic hazard assessment of afghanistan. *Journal of Seismology*, *23*(2), 217–242. [4](#)
- Wegener, A. (1912). Die entstehung der kontinente. *Geologische Rundschau*, *3*(4), 276–292. [1](#)
- Wiemer, S., Giardini, D., Fäh, D., Deichmann, N., & Sellami, S. (2009). Probabilistic seismic hazard assessment of switzerland: Best estimates and uncertainties. *Journal of Seismology*, *13*(4), 449. [3](#)
- Wiemer, S., & Wyss, M. (1997). Mapping the frequency-magnitude distribution in asperities: An improved technique to calculate recurrence times? *Journal of Geophysical Research: Solid Earth*, *102*(B7), 15115–15128. [3](#), [5](#)

- Woessner, J., Hainzl, S., Marzocchi, W., Werner, M., Lombardi, A., Catalli, F., Enescu, B., Cocco, M., Gerstenberger, M., & Wiemer, S. (2011). A retrospective comparative forecast test on the 1992 landers sequence. *Journal of Geophysical Research: Solid Earth*, *116*(B5). [6](#)
- Woessner, J., Laurentiu, D., Giardini, D., Crowley, H., Cotton, F., Grünthal, G., Valensise, G., Arvidsson, R., Basili, R., Demircioglu, M. B., et al. (2015). The 2013 european seismic hazard model: Key components and results. *Bulletin of Earthquake Engineering*, *13*(12), 3553–3596. [4](#)
- Wyss, M., Shimazaki, K., & Wiemer, S. (1997). Mapping active magma chambers by b values beneath the off-ito volcano, japan. *Journal of Geophysical Research: Solid Earth*, *102*(B9), 20413–20422. [3](#)

CHAPTER 2

THE EFFECT OF DECLUSTERING ON THE SIZE DISTRIBUTION OF MAINSHOCKS

Published as:

Mizrahi, L., Nandan, S., & Wiemer, S. (2021). The Effect of Declustering on the Size Distribution of Mainshocks. *Seismological Research Letters*, 92 (4): 2333–2342.

<https://doi.org/10.1785/0220200231>

Abstract

Declustering aims to divide earthquake catalogs into independent events (mainshocks), and dependent (clustered) events, and is an integral component of many seismicity studies, including seismic hazard assessment. We assess the effect of declustering on the frequency-magnitude distribution of mainshocks. In particular, we examine the dependence of the b -value of declustered catalogs on the choice of declustering approach and algorithm-specific parameters. Using the catalog of earthquakes in California since 1980, we show that the b -value decreases by up to 30% due to declustering with respect to the undclustered catalog. The extent of the reduction is highly dependent on the declustering method and parameters applied. We then reproduce a similar effect by declustering synthetic earthquake catalogs with known b -value, which have been generated using an Epidemic-Type Aftershock Sequence (ETAS) model. Our analysis suggests that the observed decrease in b -value must, at least partially, arise from the application of the declustering algorithm on the catalog, rather than from differences in the nature of mainshocks versus fore- or aftershocks. We conclude that declustering should be considered as a potential source of bias in seismicity and hazard studies.

2.1 Introduction

Models for probabilistic seismic hazard analysis (PSHA), e.g. (Petersen et al., 2018; Pace et al., 2006; Wiemer, Giardini, et al., 2009; Gerstenberger et al., 2020), are commonly based on the approach described by Cornell (1968), which assumes earthquake occurrence times to be representable by a stationary Poisson process. The long-term seismicity rate in a region is considered to be constant in time, reflecting a constant deformation rate and hence constant energy input at any given location, driven by plate tectonics. In reality, earthquakes trigger aftershocks, which in turn trigger their aftershocks, and so on, leading to intense clustering of earthquakes in space and time (Ogata, 1998; Jackson and Kagan, 1999; Helmstetter and Sornette, 2003). Earthquakes can also occur in swarms (Hainzl and Fischer, 2002; Hainzl, 2004), lasting days to months, sometimes comprising thousands of earthquakes in one location, which are followed by long periods of quiescence. Consequently, the recorded earthquake catalogs, especially modern instrumental ones that are complete down to small magnitudes, always show conspicuous deviations from Poissonianity. Average seismicity rates in regions with recent large sequences are therefore not representative of the long-term seismic hazard, indicating a potentially substantial location-dependent bias of seismicity rates.

Aims and challenges of declustering

So-called declustering algorithms aim to divide earthquake catalogs into clusters of dependent events and retain only the independent event of each such cluster. Although Luen and Stark (2012) find that Poissonianity depends on “the declustering method, the catalog, the magnitude range, and the statistical test”, it is generally assumed that a properly declustered earthquake catalog satisfies the condition of being Poissonian (Gardner and Knopoff, 1974, van Stiphout et al., 2012). Because of the requirement of Poissonianity for the current approach to PSHA, rate estimation for hazard assessment is often done on the basis of declustered catalogs (Field et al., 2014; Petersen et al., 2018; Pace et al., 2006; Wiemer, Giardini, et al., 2009, Drouet et al., 2020; Meletti et al., 2017; Akinci et al., 2018; Sesetyan et al., 2018; Beauval et al., 2013; Waseem et al., 2019; Woessner et al., 2015). In this sense, PSHA approaches estimate mainshock rates rather than total seismicity rates.

While Poissonianity of the declustered catalog is necessary for a declustering method to serve its purpose, this condition does not ensure a unique solution to the declustering prob-

lem. To avoid inadvertently rewarding the excessive removal of events from the catalog, an additional criterion is required. However, as the actual triggering processes are not currently known and nature does not provide us with labels such as ‘mainshock’, ‘aftershock’, ‘foreshock’, or ‘swarm member’, we lack an objective criterion for the performance evaluation of declustering methods. Several algorithms have been proposed and used in the past (Gardner and Knopoff, 1974; Gruenthal, 1985; Uhrhammer, 1986; Reasenber, 1985; Zaliapin et al., 2008; Zhuang et al., 2002; Marsan and Lengline, 2008; see van Stiphout et al. (2012) for an overview).

Effects of declustering on PSHA

In a study on the effect of declustering on hazard results for the city of Istanbul, Eroglu Azak et al. (2018) found that peak ground acceleration values vary by up to 20% depending on the declustering method. Marzocchi and Taroni (2014) discuss the need for declustering for PSHA, concluding that it is only necessary to avoid a bias in the spatial distribution of earthquake occurrences. Furthermore, considering that aftershocks can also cause considerable damage, they find that the neglecting of aftershock effects due to declustering may lead to significant underestimation of seismicity rates and hence of seismic hazard. In this regard, Iervolino et al. (2018) and Iervolino (2019) have proposed a generalization of the hazard integral to re-introduce aftershock hazard in PSHA. Moreover, van Stiphout et al. (2011) found that the choice of declustering method has a major effect on seismicity rate-change estimations. On the other hand, sensitivity studies to different declustering approaches in Switzerland have shown that the impact of declustering on the hazard is often negligible (Wiemer, Garcia-Fernandez, et al., 2009). The need for, potential biases introduced by, and alternatives to declustering have also been discussed in the context of seismicity forecasting (Nandan, Ouillon, Sornette, and Wiemer, 2019a; Schorlemmer et al., 2007). In particular, the issue is raised that a mainshock forecast can only be tested against a mainshock ‘truth’ which is inherently dependent on the somewhat arbitrary choice of declustering method, yielding full seismicity forecasts the only objectively testable type of forecast.

Effects of declustering on the b -value

A major role in the calculation of seismicity rates is played by the b -value of the empirical Gutenberg-Richter (GR) law (Gutenberg and Richter, 1944), which describes the frequency distribution of earthquake magnitudes. Typically, b -values of earthquake catalogs lie close to 1 (Y. Kagan, 1999; Kamer and Hiemer, 2015), but have been found to vary with time, region, depth, and stress regime. Several studies have also reported higher b -values during swarms or in volcanic areas (Main et al., 1992; Henderson et al., 1992; Frohlich and Davis, 1993; Wiemer and Wyss, 1997; Schorlemmer et al., 2005; Petruccelli et al., 2019; Wyss et al., 1997). Y. Kagan (1999), Kamer and Hiemer (2015), and Marzocchi et al. (2020) discussed a variety of potential technical causes of b -value variations, such as magnitude binning, network coverage, catalog incompleteness, or the finiteness of data. Moreover, imposing a GR law on declustered catalogs, as is commonly done in seismic hazard analysis, often results in a significantly lower b -value compared to full catalogs (Y. Y. Kagan, 2010; Christophersen et al., 2011; Field et al., 2014; Petersen et al., 2018). Some argue that this behavior is a property naturally inherent to mainshocks (Knopoff, 2000). On a similar note, Gulia et al. (2018) suggested that the b -value of typical aftershock sequences is on average 20% higher than the mainshock b -value, and that this increase in b -value is a long-lasting effect for several years.

However, it is debatable whether the b -value of a declustered catalog is at all meaningful. Most declustering methods define a mainshock as the largest event of an independent cluster. If one assumes a GR law-type Pareto distribution of magnitudes on the full catalog, one should not at the same time assume a GR law-type Pareto distribution of mainshock

magnitudes. The distribution of the maximum of a set of independent and identically distributed random variables, i.e., the distribution of mainshock magnitudes, can be derived from fundamental principles of probability theory (Kolmogoroff, 1934). Lombardi (2003) gave a mathematical description of how mainshock magnitude distribution follows from the full-catalog GR law. She showed that the difference in b -value between mainshocks and all events becomes minimal when a corrected log-likelihood function is used in the maximum likelihood estimation of the mainshock b -value. The mainshock magnitude distribution she proposed depends on the empirical distribution of cluster sizes emerging from the declustering process. Given this result and assuming that different declustering algorithms will lead to different cluster size distributions, it is expected that different declustering methods will also lead to different mainshock magnitude distributions. Hence, b -values of mainshocks, when estimated in the usual way, are expected to be biased as an artifact of declustering.

Similarly, Zhuang and Ogata (2006) found that the magnitude distribution of mainshocks defined via the Epidemic-Type Aftershock Sequence (ETAS) model (see their article for the definition or Section A.4 in the Supporting Information for this chapter for details on the ETAS model) departs from the GR law and that the full catalog b -value is valid for mainshocks in the asymptotic case where $m \rightarrow \infty$. For lower magnitude mainshocks, one could argue that a GR law with lower a - and b -values than those of the full catalog presents an acceptable approximation of the true, non-Pareto distribution of mainshock magnitudes. However, when the logarithms of the numbers $N(m)$ and $N_{main}(m)$ of earthquakes and mainshocks of magnitude $M > m$ are both described by linear terms of the form

$$\log_{10}(N) = a - b \cdot m, \quad (2.1)$$

$$\log_{10}(N_{main}) = a_{main} - b_{main} \cdot m, \quad (2.2)$$

where $b_{main} \neq b$, the two lines intersect at a point

$$m_x = \frac{a - a_{main}}{b - b_{main}}. \quad (2.3)$$

If $b_{main} < b$, this means that the expected number of mainshocks of magnitude $M > m_x$ is larger than the expected number of total earthquakes of magnitude $M > m_x$, even though the observed number of mainshocks can never be larger than the observed number of earthquakes.

Paper outline

Considering the importance of the b -value for seismicity studies and seismic hazard estimates, we here systematically assess the influence of declustering on mainshock size distribution. To do so, we first verify that imposing a GR law on mainshocks yields a b -value that does indeed depend on the choice of the declustering method applied. Then, we show that a similar effect is observed for synthetic catalogs with known b -value, whose magnitude distribution by design does not distinguish mainshocks and other events. Furthermore, we illustrate the consequences of approximating mainshock magnitude distribution with a Pareto distribution and calculate the tipping point magnitude m_x , above which the bias introduced by declustering cannot be interpreted as mainshock-specific behavior.

The rest of the paper is organized as follows. In Section 2.2, we describe the earthquake catalog used for this study and discuss the coupled estimation of completeness magnitude and b -value. In Section 2.3, we describe the declustering methods and corresponding parameter choices. There, we also describe the ETAS model, which is used for the simulation of

synthetic catalogs and furthermore serves as the basis for two of the declustering methods. We then present and discuss our main results in Section 2.4 and state our conclusion in Section 2.5. The Supporting Information for this chapter (Appendix A) contains a more detailed description of all methods and algorithms used. Moreover, it contains analyses of the sensitivity of full catalog b -value and mainshock b -value on the completeness magnitude m_c .

2.2 Data

In this study, we use the ANSS Comprehensive Earthquake Catalog (ComCat) provided by the U.S. Geological Survey (see Section 2.6) with ‘preferred’ magnitudes as defined in ComCat, in the collection area around the state of California as in the RELM testing center (Schorlemmer and Gerstenberger, 2007). The choice of the study region is motivated mainly by the high seismicity in the area and by completeness at low magnitudes of the catalog for several decades (Hutton et al., 2006), both ensuring that a large and representative amount of data can be used in our study. We consider events of magnitude $M \geq 0.0$, with magnitudes rounded into bins of size $\Delta M = 0.2$. Figure A.1 in the Supporting Information for this chapter shows that the b -value is insensitive to bin size for reasonable choices of m_c . It also shows that $b(m_c)$ is more stable for $\Delta M = 0.2$ compared to $\Delta M = 0.1$. The time frame used is January 1, 1970 until September 30, 2019, of which only the events on or after January 1, 1980 are used for the estimation of b -values. We subsequently call this set of events the *incomplete primary catalog*. The earlier events make up the *incomplete auxiliary catalog*. As earthquake clusters may occur close to the start of the primary time period, ignoring auxiliary events could lead to unwanted deficiencies in cluster detection and mainshock identification (Wang et al., 2010; Schoenberg et al., 2010; Nandan, Ouillon, and Sornette, 2019). Our choice of time periods aims to achieve balance between long enough primary and auxiliary periods, and low completeness magnitude in the primary catalog thanks to seismic network configuration (see e.g. Hutton et al., 2006).

b -value estimation and completeness magnitude

Estimating the b -value of a catalog requires knowledge of its completeness magnitude m_c , the magnitude threshold above which all events are assumed to be detected. Assuming too low values for m_c can cause severe underestimation of the b -value (see Figure A.1 in the Supporting Information for this chapter). On the other hand, assuming overly conservative values for m_c leads one to discard a large portion of the data, making b -value estimates imprecise. In reality, m_c is not known and has to be estimated itself. Several methods to do so have been proposed; see Mignan and Woessner (2012) for an overview. Commonly, m_c is estimated by defining it as the magnitude threshold above which earthquakes follow the GR law. In this sense, the estimation of b -value and m_c becomes a coupled problem; one cannot be estimated without knowledge of the other. In Sections A.1 and A.2 and Figure A.2 in the Supporting Information for this chapter, we adapt the method proposed by Clauset et al. (2009) to jointly estimate m_c and b -value, ultimately arriving at a value of 3.6 for m_c . A sensitivity analysis (see Figure A.3 in the Supporting Information for this chapter) shows that the results presented in the following sections are insensitive to reasonable choices of m_c .

Setting the value of m_c to 3.6 implies that we subsequently use the subset of events with magnitude $M \geq 3.6$ of the previously described (binned) catalog. This filter is applied to both the incomplete primary and the incomplete auxiliary catalog, yielding the (*complete*) *primary* and (*complete*) *auxiliary catalog*. Figure 2.1(a) and (c) show the spatial and temporal distribution of events in the catalog with magnitude $M \geq 3.6$, where auxiliary events

Table 2.1: ETAS parameters used for catalog simulation, obtained by expectation maximization.

Parameter	Value
$\log_{10}(k_0)$	-2.49
a	1.69
$\log_{10}(c)$	-2.95
ω	-0.03
$\log_{10}(\tau)$	3.99
$\log_{10}(d)$	-0.35
γ	1.22
ρ	0.51
$\log_{10}(\mu)$	-7.17

are highlighted in yellow. For our primary catalog we obtain a b -value of 1.01, as illustrated in Figure 2.1(b).

2.3 Method

To better understand the influence of declustering on the b -value, we first apply five often used declustering techniques with different parameter and window choices to the same real catalog. We then apply the same declustering methods, with standard parameters, to a set of 2000 synthetic catalogs. The synthetic catalogs are generated using a basic ETAS model (see Ogata, 1998; Veen and Schoenberg, 2008; Nandan et al., 2017; Nandan, Ouillon, Sornette, and Wiemer, 2019b), which is described in Section A.4 in the Supporting Information for this chapter. Table 2.1 shows the parameters used in the simulation of synthetic catalogs. They were obtained by applying expectation maximization (Veen and Schoenberg, 2008; Nandan et al., 2017) to the primary and auxiliary California catalog, to support the comparability of real and synthetic catalogs. For a detailed description of the ETAS model, as well as the algorithms used for inversion and simulation, see Section A.4 in the Supporting Information for this chapter. We use the synthetic catalogs to test whether declustering introduces any systematic bias to the mainshock size distribution. As in the case of synthetic catalogs, the distribution from which magnitudes are drawn is known and is assumed to be the same for mainshocks and aftershocks, any changes in b -value observed after declustering must have their origin in the application of declustering algorithms.

To further understand the consequences of approximating mainshock magnitude distribution with a lower- b -value GR law, we compare the ratio

$$r(m) = \frac{N_{main}(m)}{N(m)} \quad (2.4)$$

of mainshocks among earthquakes of magnitude $M_{i,m}$ between observation and approximation, for the different declustering methods with standard parameter settings applied to the Californian catalog. We calculate m_x (see Equation (2.3)), above which $r(m) > 1$ implies that the introduced bias can impossibly be supported by observations.

We examine the declustering methods proposed by Reasenberg (1985), Zaliapin et al. (2008), and window methods as proposed by Gardner and Knopoff (1974), Gruenthal (1985), and Uhrhammer (1986). We also consider two versions of declustering based on the ETAS model (Zhuang et al., 2002). For the detailed description of all declustering algorithms and parameter ranges applied, see Sections A.3 and A.4 and Tables A.1 to A.3 in the Supporting

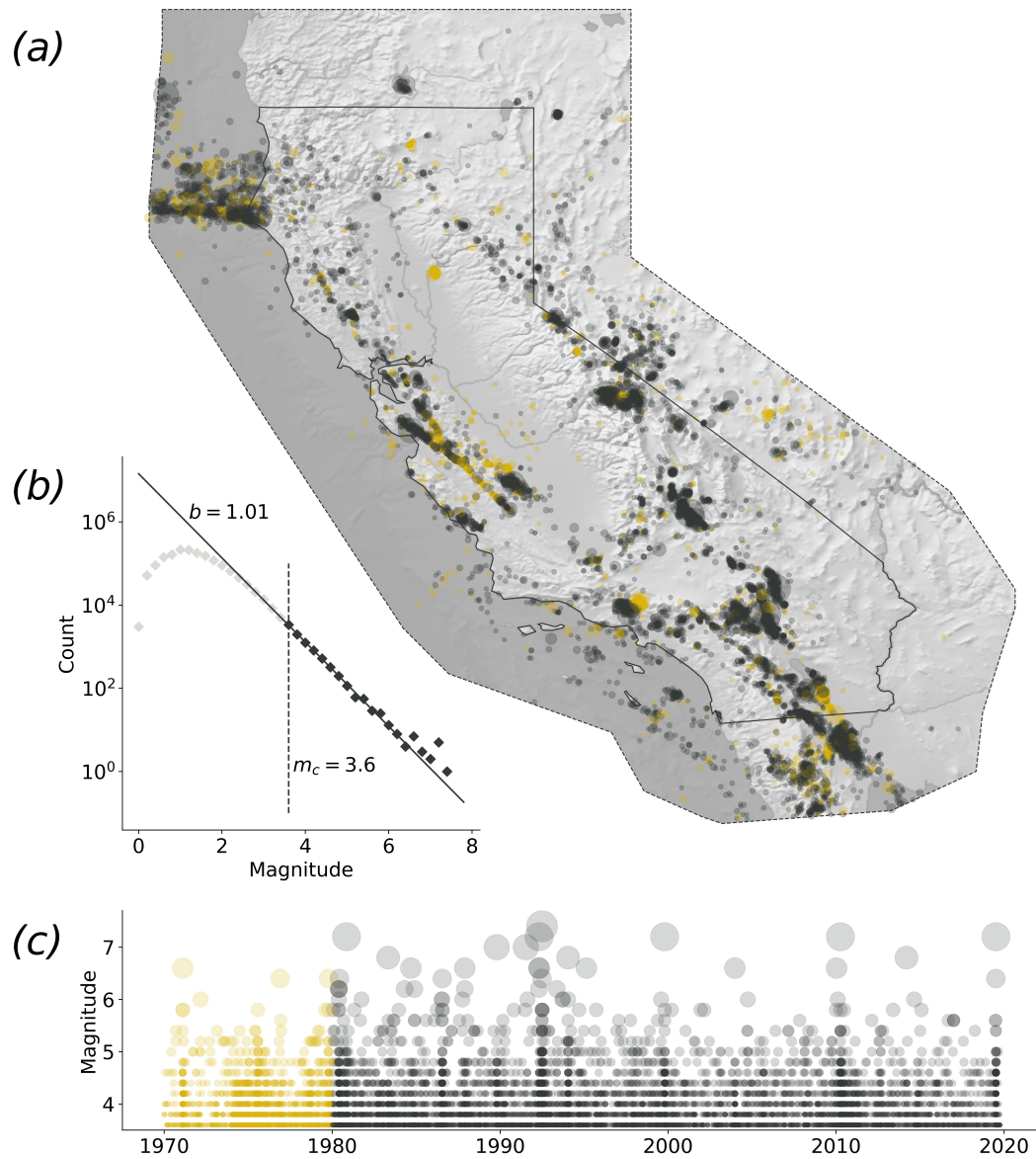


Figure 2.1: Earthquake catalog used in this analysis. (a) Seismicity map. Dots represent earthquakes in the catalog with $M \geq 3.6$, where dot size indicates magnitude. Events between 1970 and 1980, which serve as auxiliary data, are marked in yellow. Solid black line marks the California state boundary, dotted line marks the boundary of the considered region. (b) Absolute frequency distribution of magnitudes above and below m_c (black versus grey diamonds). Solid black line shows the GR law fitted to the catalog of events with $M \geq 3.6$. (c) Temporal distribution of the events shown in (a), with identical size and color coding.

Information for this chapter; we give a short description of each method below. Note that the non-parametric stochastic declustering algorithm proposed by Marsan and Lengline (2008) is not used here. This is because of its similarity to the already considered parametric stochastic declustering alternative provided by the ETAS model. The main difference to ETAS declustering is that the triggering rate, described as $g(t, x, y, m)$ in Section A.4 in the Supporting Information for this chapter, is there obtained empirically, without presuming the laws (A.13-A.15). In their analysis of southern California seismicity, they observe that their empirically derived triggering rate follows laws similar to those described in (A.13-A.15), which, likewise, were originally discovered empirically.

Short descriptions of the declustering methods applied in this article

1. Reasenber (1985) introduced an algorithm that has been used in numerous studies and recent PSHA, e.g. in Ecuador (Beauval et al., 2013) or Afghanistan (Waseem et al., 2019). It defines earthquake interaction zones in space and time. Here, we apply the spatial interaction relationships proposed by Reasenber (1985), and Wells and Coppersmith (1994), and the parameter ranges for temporal interaction zones recommended by Schorlemmer and Gerstenberger (2007).
2. Window methods, as first described by Gardner and Knopoff (1974) use space-time windows around large events to identify their fore- and aftershocks. Different formulations of such window boundaries have been suggested and are applied in this study (Gardner and Knopoff, 1974; Gruenthal, 1985; Uhrhammer, 1986; see van Stiphout et al. (2012)). We use the original formulation by Gardner and Knopoff (1974) as the standard window, which is used in the Uniform California Earthquake Rupture Forecast (UCERF3, Field et al., 2014). Generally, window methods are widely used in modern regional and national seismic hazard models, see Drouet et al. (2020) for France, Meletti et al. (2017) for Italy, Sesetyan et al. (2018) for Turkey, Woessner et al. (2015) for Europe (ESHM13).
3. The alternative approach of Zaliapin et al. (2008) applies a Gaussian mixture model on space-time nearest-neighbor distances between events to distinguish independent from dependent events.
4. The ETAS model is used here in two ways. Firstly, it is used to simulate synthetic earthquake catalogs upon which declustering methods are applied to study their effects. Secondly, the ETAS model induces an alternative, parametric approach to declustering, which was introduced by Zhuang et al. (2002). We consider two versions of declustering based on the ETAS model, which differ in their definition of mainshocks and are described in detail in Section A.4 in the Supporting Information for this chapter. ‘ETAS-Main’ defines the largest event of a cluster to be the mainshock, while ‘ETAS-Background’ defines events to be mainshocks if they are not triggered. The definition used in ETAS-Background is in the spirit of the ETAS model, where background earthquakes of any size can trigger cascades of aftershocks. ETAS-Main, on the other hand, imposes the mainshock definition used in the other methods, in the interest of comparability. We subsequently call those methods which define mainshock as the largest events ‘mainshock methods’. Note that because of its different definition of mainshocks, ETAS-Background is unsuited to be applied in the standard PSHA approach, which is designed to work with mainshock methods.

2.4 Results and discussion

The Disparity Between Declustering Methods The cumulative number of mainshocks for different declustering methods with standard parameter and window choices, compared to

the full California catalog, is shown in Figure 2.2(a). The diversity among the resulting declustered catalogs is remarkable. Mainshock rates vary by a factor of 6.1 between the most and least ‘aggressive’ algorithm. Moreover, while the removal of temporal clusters is the primary goal of the declustering process, some are still clearly visible after declustering with Reasenberg’s method, and still recognizable, though less pronounced, after applying Zaliapin’s method. Gardner-Knopoff and ETAS (Main and Background) appear to be more successful at achieving temporal Poissonianity.

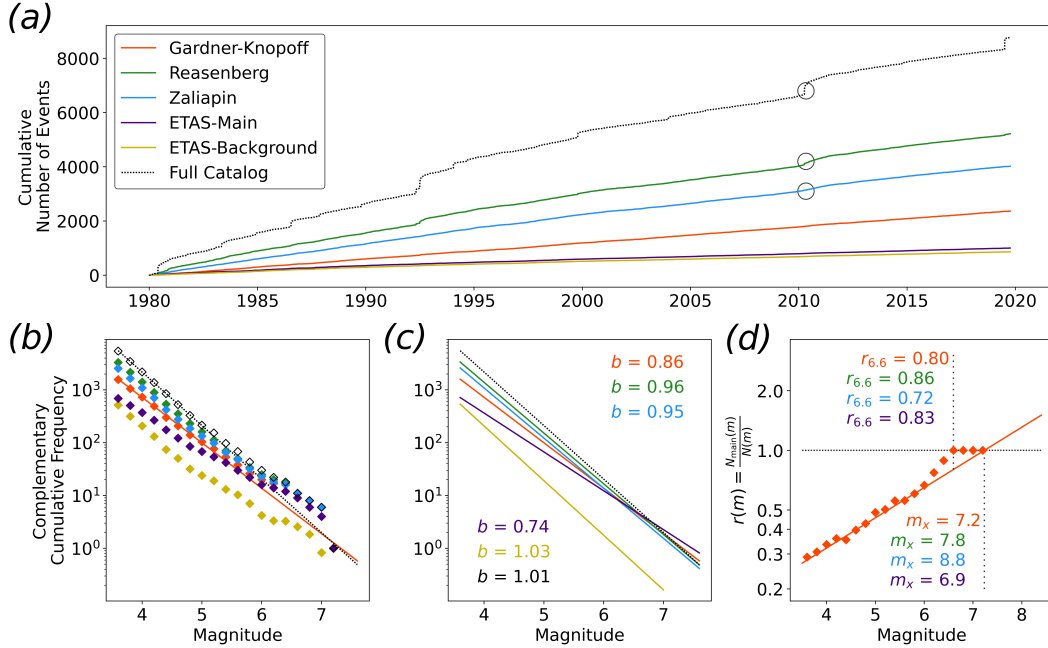


Figure 2.2: Properties of the California catalog of mainshocks larger than or equal to $M_{3.6}$, depending on declustering method. Standard parameter settings (and standard window) of each method are used for declustering. (a) Cumulative number of mainshocks. Dotted black line represents the full (non-declustered) catalog. The rapid increase in seismicity highlighted in circles corresponds to the 2010 (M_w 7.2) El Mayor-Cucapah earthquake in Baja California, Mexico. (b) Empirical Complementary Cumulative Frequency Function (CCFF, diamonds) of mainshock magnitudes. Empty black diamonds represent the full California catalog. The two lines are the fitted CCFF for Gardner-Knopoff declustered and full catalog. (c) Fitted CCFF for declustered catalogs compared to full catalog GR law fit. Fitted b -values are given. (d) Observed (diamonds) and approximated (line) evolution of $r(m) = \frac{N_{main}(m)}{N(m)}$ for the Gardner-Knopoff declustered catalog. Black dotted line marks $r(m) \equiv 1$. m_x and $r_{6.6}$ are given, also for Reasenberg, Zaliapin and ETAS-Main declustered catalogs. Note the different x-axis for (d) compared to (b) and (c).

The observed and fitted complementary cumulative frequency functions (CCFFs) of mainshock magnitudes are shown in Figure 2.2(b) and (c). Observed absolute frequencies of large events ($M \geq 6.4$) are somewhat similar for all declustering methods, with the exception of ETAS-Background. Relative frequencies of large events versus small events vary strongly between methods, which manifests itself in slope differences between the CCFFs. Note that for the mainshock methods, the aggressiveness of the method coincides with the extent of slope decrease. This effect can be explained as a consequence of the methods’ mainshock definition. Since small events are less likely to be identified as mainshocks, they are more likely to be removed from the catalog, increasing relative frequencies of large events.

ETAS-Background, in contrast, does not seem to preferentially remove events from specific magnitude ranges.

Figure 2.2(d) illustrates the consequences of estimating seismic hazard based on a mainshock GR law with lower b -value. For the California catalog, observed and approximated evolutions of $r(m) = \frac{N_{main}(m)}{N(m)}$ are shown for mainshocks obtained by Gardner-Knopoff declustering. The magnitude m_x (see Equation (2.3)) above which rate overestimation cannot be denied, is given also for Reasenber, Zaliapin and ETAS-Main declustered catalogs. m_x varies between 6.9 and 8.8, where higher values of m_x are predominantly observed for declustering methods which do not succeed at achieving Poissonianity in time. Furthermore, most declustering methods show considerable deviations of observed $r(m)$ from its approximation already at lower magnitudes. For instance, the approximation of $r_{6.6} = r(6.6)$ lies between 0.72 and 0.86, depending on mainshock definition, even though all definitions except ETAS-Main classified all $M > 6.6$ events to be mainshocks. Note that ETAS-Background is excluded from this part of the analysis due to its inapplicability in the standard PSHA approach.

b -value of declustered catalogs

Observations on real data

The relative frequency increase of large events translates into a lower b -value when a GR law is imposed on the frequency-magnitude distribution of mainshocks. In Figure 2.3(a), b -value is plotted against a -value of the declustered California catalog, comparing the effects of varying declustering methods and parameters. We find that b -values of declustered catalogs vary strongly with declustering algorithms. Values between 0.73 and 1.00 are attained without any significant gap. A general trend is recognizable among the mainshock methods: removal of more events correlates with lower b -values, indicating a penchant of these methods to relatively remove more smaller events than larger ones. The b -value obtained with ETAS-Background does not significantly differ from the full catalog b -value. These observations are in line with the explanation given above, which describes the b -value decrease as a consequence of the mainshock definition, and are expected knowing the results by Lombardi (2003), Zhuang and Ogata (2006), Kagan (2010) and van Stiphout et al. (2011). A sensitivity analysis of the b -value to the completeness magnitude m_c (see Figure A.3 in the Supporting Information for this chapter) shows that the b -value decrease after declustering is an effect that is observed regardless of the reasonable choice of m_c , with the extent of the decrease being characteristic of each method.

Observations on simulated data

Synthetic catalogs, where all magnitudes are drawn from one single distribution, show lower b -values after declustering. In Figure 2.3(b), the distribution of mainshock b -values of 2000 ETAS-simulated catalogs is shown for different declustering methods with standard parameter settings, aligned according to the median observed a -value of the respective method. The mainshock b -values of the same methods applied to the California catalog are indicated with stars; error bars mark the estimated standard error. If no declustering, or ETAS-Background declustering, is applied, the estimated b -value of synthetic catalogs is consistent with the b -value used in their simulation. At the same time, b -values of synthetic catalogs declustered with mainshock methods are always lower than the b -value used in their simulation. Comparing the extent of the effect across different declustering methods, synthetic and real data have the same qualitative behavior. Similarly, the a -value decrease is observed to be method-characteristic.

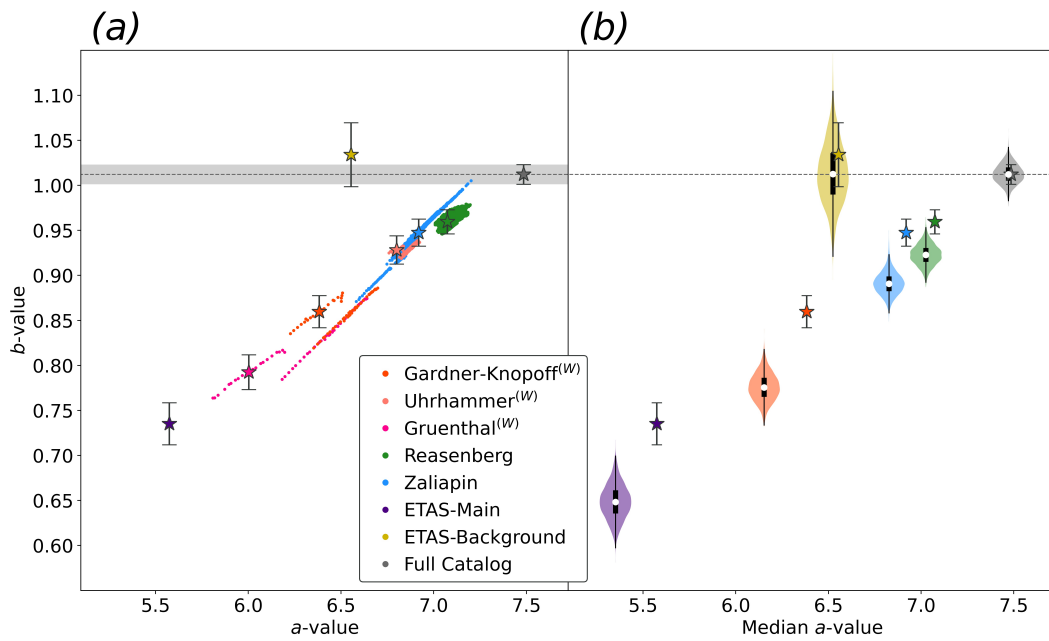


Figure 2.3: (a) a -value versus b -value of the declustered California catalog depending on declustering method and parameters. Each dot represents one variation of parameter settings, stars with error bars represent standard parameter settings. Marked with (W) are window methods. The dotted grey line and grey area indicate the b -value of the non-declustered catalog and its uncertainty. (b) Distribution of mainshock b -values of 2000 simulated catalogs, depending on declustering method (with standard parameter settings and standard (Gardner-Knopoff) window), plotted against median a -value per method. Stars with error bars represent the a - and b -value of the regional earthquake catalog from (a) for the respective methods. White dot, black box and black line represent median, interquartile range and adjacent values. The dotted line displays the b -value used for catalog generation, which corresponds to the full-catalog b -value observed in the Californian primary catalog.

The effect of declustering on the b -value is more pronounced in synthetic data, for all methods. A possible explanation for this is that all declustering methods assume isotropic spatial distribution of aftershocks, which is known to be wrong in reality, but valid for synthetic catalogs. Hence, cluster detection is facilitated in synthetic catalogs, resulting in more effective removal of small events compared to the real catalog.

2.4.1 Remarks regarding ETAS declustering

1. Despite ETAS being the generative process of synthetic catalogs, a large difference in b -value is observed after ETAS-Main declustering. This is not a flaw in the generative process or the parameter inversion. On the contrary, this behavior is expected. The low b -value of ETAS-Main-declustered catalogs is due to the imposed definition of ‘mainshock’ as the largest event of a cluster, not to be confused with ETAS’ notion of background events. This concept of mainshocks is not relevant in the generation of catalogs. Imposing such a definition leads to selective removal of small events rather than to the removal of aftershocks in the true ETAS sense. With ETAS, aftershocks are temporally restricted to occur after their triggering events but may have larger magnitudes.
2. Declustering with ETAS-Background allows a comparison between the b -value of background events according to the ETAS definition and the full catalog. No significant difference is observed, both in the case of synthetic and real catalogs. The difference only arises when the rule of maximum magnitude is applied.
3. For synthetic data, the underlying branching structure is known by design of the experiment. In contrast, cluster detection for real data requires the lengthy process of inversion and probabilistic cluster assignment. Thus, compared to synthetic data, cluster detection is intrinsically less correct for real data, and declustering is inclined to be less effective. It is reasonable to assume that this circumstance explains the particularly pronounced difference in b -value in the case of ETAS-Main declustering.

ETAS simulations do not distinguish the magnitude distribution of mainshocks versus aftershocks. A difference in pre- and post-declustering b -value of a declustered catalog that was generated using ETAS can, therefore, only have its origin in the systematic selection of large events as mainshocks. The purely declustering-induced and strongly method-dependent decrease in b -value suggests that other potential causes, such as a different nature of mainshocks compared to fore- or aftershocks, have negligible effects on the mainshock b -value. While the possibility cannot be precluded that a part of the effect is due to the change in stress state before and after major events (e.g. Gulia et al., 2018), the notably arbitrary effect of declustering should not be ignored. The mere observation of a lower b -value of mainshocks is no proof of its meaningfulness; the observation of artifactual effects of declustering on the mainshock b -value, however, is a reason to doubt its meaningfulness.

2.5 Conclusion

We demonstrate that a decrease in overall b -value of the California catalog after declustering is observed for a variety of declustering methods and parameter settings. Furthermore, the extent of the decrease is highly dependent on the algorithm applied. A general trend is observed, suggesting that more ‘aggressive’ algorithms tend to be accompanied by a more pronounced b -value decrease, ETAS-Background being the only exception to this rule. With a medial resulting a -value among the methods considered, it leaves the b -value unchanged. Finally, we find that all the above-described effects can be reproduced in synthetic data, which was generated using a constant b -value for all events.

Our results indicate that declustering substantially affects the earthquake size distribution. Imposing a GR law on declustered catalogs, therefore, leads to b -values which are biased to a somewhat arbitrary and not immediately apparent extent. This bias leads to an overestimation of seismic hazard above a certain magnitude m_x . Thus, we can conclude that the current state of practice of equating seismic hazard with mainshock rates which follow a GR law can be accused of three deficiencies. One is the non-verifiability of any mainshock definition. Secondly, fore- or aftershocks can be large and devastating. Neglecting aftershock effects may give rise to a substantial underestimation of seismic hazard (Marzocchi and Taroni, 2014). And lastly, the earthquake size distribution resulting from the procedure causes hazard overestimation for events above a certain size.

One may argue that increasing the relative frequency of large events and decreasing the absolute frequency of all events have antagonistic effects on the absolute frequency of large events, justifying any choice of declustering method. Indeed, most hazard studies ignore previous findings and continue to calculate hazard in the usual way. However, we believe that two wrongs do not make a right. To be precise, two wrongs make a right only for one particular magnitude, m_x . Our analysis suggests that above m_x , classical hazard studies overestimate the seismic hazard, whereas below m_x , they underestimate it. While the accusation of underestimation could partially be rejected by insisting that declustering reveals the true mainshocks and that aftershock effects are deliberately excluded from the scope of hazard assessment, we have shown that the overestimation cannot be similarly attributed to a true b -value that is revealed by declustering, but that this resulting b -value is biased to a non-negligible extent.

It is crucial to be aware of this issue when estimating seismic hazard. While analysis of earthquake dependency is inevitable to eliminate spatial bias for the calculation of seismicity rates (Marzocchi and Taroni, 2014), basing calculations solely on declustered catalogs is not an appropriate approach. One alternative possibility is to use ETAS models to assess seismicity rates (see e.g. Field et al. (2015)). In a pseudo-prospective forecasting experiment on the Californian catalog conducted by Nandan, Ouillon, Sornette, and Wiemer (2019b), ETAS models generally outperform all competing smoothed seismicity models and models based on strain rates. Using hundreds of thousands of simulations of possible scenarios as the basis for a forecast, they intrinsically account for the spatiotemporal clustering of earthquakes. This approach incorporates the non-Poissonian nature of reality while reducing the spatial bias encountered in undeclustered catalogs. At the same time, ETAS relies only on the GR law of the full catalog and therefore avoids making assumptions on the frequency-magnitude distribution of somewhat arbitrarily selected large events.

Other ways to address this matter may exist. What is essential is to recognize the problematic aspects of doing hazard assessment based on declustered catalogs and to find a way to address the issues presented here.

2.6 Data and resources

The ANSS Comprehensive Earthquake Catalog (ComCat) provided by the U.S. Geological Survey was searched using <https://earthquake.usgs.gov/data/comcat/> (last accessed on November 30, 2019).

The electronic supplement to this article contains a more detailed description of all methods and algorithms used. Moreover, it contains analyses of the sensitivity of full catalog b -value and mainshock b -value on the completeness magnitude m_c . Finally, observed and

approximated ratio of mainshocks among earthquakes of magnitude $M_{i,m}$ is shown using different declustering methods for mainshock definition.

Acknowledgments

The authors wish to thank Celso Reyes for providing python implementations of the declustering algorithms for Reasenber, Zaliapin, and window method declustering, as well as Laurentiu Danciu, Arnaud Mignan, as well as two anonymous reviewers for their helpful feedback on an earlier version of this manuscript. This work has received funding from the ETH research grant for project No. 2018-FE-213, “Enabling dynamic earthquake risk assessment (DynaRisk)” and from the European Union’s Horizon 2020 research and innovation programme under grant agreement No.821115, Real-Time Earthquake Risk Reduction for a Resilient Europe (RISE).

References

- Akinci, A., Moschetti, M. P., & Taroni, M. (2018). Ensemble smoothed seismicity models for the new italian probabilistic seismic hazard map. *Seismological Research Letters*, 89(4), 1277–1287. [14](#)
- Beauval, C., Yepes, H., Palacios, P., Segovia, M., Alvarado, A., Font, Y., Aguilar, J., Troncoso, L., & Vaca, S. (2013). An earthquake catalog for seismic hazard assessment in ecuador. *Bulletin of the Seismological Society of America*, 103(2A), 773–786. [14](#), [20](#)
- Christophersen, A., Gerstenberger, M., Rhoades, D., & Stirling, M. (2011). Quantifying the effect of declustering on probabilistic seismic hazard. *Proc. of the Ninth Pacific Conf. on Earthquake Engineering: Building an Earthquake-Resilient Society*. [15](#)
- Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, 51(4), 661–703. [17](#)
- Cornell, C. A. (1968). Engineering seismic risk analysis. *Bulletin of the seismological society of America*, 58(5), 1583–1606. [14](#)
- Drouet, S., Ameri, G., Le Dortz, K., Secanell, R., & Senfaute, G. (2020). A probabilistic seismic hazard map for the metropolitan france. *Bulletin of Earthquake Engineering*, 18(5), 1865–1898. [14](#), [20](#)
- Eroglu Azak, T., Kalafat, D., Şeşetyan, K., & Demircioğlu, M. (2018). Effects of seismic declustering on seismic hazard assessment: A sensitivity study using the turkish earthquake catalogue. *Bulletin of Earthquake Engineering*, 16(8), 3339–3366. [15](#)
- Field, E. H., Arrowsmith, R. J., Biasi, G. P., Bird, P., Dawson, T. E., Felzer, K. R., Jackson, D. D., Johnson, K. M., Jordan, T. H., Madden, C., et al. (2014). Uniform california earthquake rupture forecast, version 3 (ucerf3)—the time-independent model. *Bulletin of the Seismological Society of America*, 104(3), 1122–1180. [14](#), [15](#), [20](#)
- Field, E. H., Biasi, G. P., Bird, P., Dawson, T. E., Felzer, K. R., Jackson, D. D., Johnson, K. M., Jordan, T. H., Madden, C., Michael, A. J., et al. (2015). Long-term time-dependent probabilities for the third uniform california earthquake rupture forecast (ucerf3). *Bulletin of the Seismological Society of America*, 105(2A), 511–543. [25](#)
- Frohlich, C., & Davis, S. D. (1993). Teleseismic b values; or, much ado about 1.0. *Journal of Geophysical Research: Solid Earth*, 98(B1), 631–644. [15](#)
- Gardner, J., & 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. [14](#), [15](#), [18](#), [20](#)
- Gerstenberger, M. C., Marzocchi, W., Allen, T., Pagani, M., Adams, J., Danciu, L., Field, E. H., Fujiwara, H., Luco, N., Ma, K.-F., et al. (2020). Probabilistic seismic hazard analysis at regional and national scales: State of the art and future challenges. *Reviews of Geophysics*, 58(2), e2019RG000653. [14](#)

- Gruenthal, G. (1985). State of knowledge about the earthquake activity in the vogtland/nw-bohemia region prior to the earthquake swarm in 1985/86. *Monitoring and analysis of the earthquake swarm*, 86. 15, 18, 20
- Gulia, L., Rinaldi, A. P., Tormann, T., Vannucci, G., Enescu, B., & Wiemer, S. (2018). The effect of a mainshock on the size distribution of the aftershocks. *Geophysical Research Letters*, 45(24), 13–277. 15
- Gutenberg, B., & Richter, C. F. (1944). Frequency of earthquakes in california. *Bulletin of the Seismological Society of America*, 34(4), 185–188. 15
- Hainzl, S. (2004). Seismicity patterns of earthquake swarms due to fluid intrusion and stress triggering. *Geophysical Journal International*, 159(3), 1090–1096. 14
- Hainzl, S., & Fischer, T. (2002). Indications for a successively triggered rupture growth underlying the 2000 earthquake swarm in vogtland/nw bohemia. *Journal of Geophysical Research: Solid Earth*, 107(B12), ESE–5. 14
- Helmstetter, A., & Sornette, D. (2003). Importance of direct and indirect triggered seismicity in the etas model of seismicity. *Geophysical Research Letters*, 30(11). 14
- Henderson, J., Main, I., Meredith, P., & Sammonds, P. (1992). The evolution of seismicity at parkfield: Observation, experiment and a fracture-mechanical interpretation. *Journal of Structural Geology*, 14(8-9), 905–913. 15
- Hutton, K., Hauksson, E., Clinton, J., Franck, J., Guarino, A., Scheckel, N., Given, D., & Yong, A. (2006). Southern california seismic network update. *Seismological Research Letters*, 77(3), 389–395. 17
- Iervolino, I. (2019). Generalized earthquake counting processes for sequence-based hazard. *Bulletin of the Seismological Society of America*, 109(4), 1435–1450. 15
- Iervolino, I., Chioccarelli, E., & Giorgio, M. (2018). Aftershocks' effect on structural design actions in italy. *Bulletin of the Seismological Society of America*, 108(4), 2209–2220. 15
- Jackson, D. D., & Kagan, Y. Y. (1999). Testable earthquake forecasts for 1999. *Seismological Research Letters*, 70(4), 393–403. 14
- Kagan, Y. Y. (2010). Earthquake size distribution: Power-law with exponent $\beta \equiv 12$? *Tectonophysics*, 490(1-2), 103–114. 15
- Kagan, Y. (1999). Universality of the seismic moment-frequency relation. *Seismicity patterns, their statistical significance and physical meaning* (pp. 537–573). Springer. 15
- Kamer, Y., & Hiemer, S. (2015). Data-driven spatial b value estimation with applications to california seismicity: To b or not to b. *Journal of Geophysical Research: Solid Earth*, 120(7), 5191–5214. 15
- Knopoff, L. (2000). The magnitude distribution of declustered earthquakes in southern california. *Proceedings of the National Academy of Sciences*, 97(22), 11880–11884. 15
- Kolmogoroff, A. (1934). Grundbegriffe der wahrscheinlichkeitsrechnung. *Bull. Am. Math. Soc.*, 522–523. 16
- Lombardi, A. M. (2003). The maximum likelihood estimator of b-value for mainshocks. *Bulletin of the Seismological Society of America*, 93(5), 2082–2088. 16
- Luen, B., & Stark, P. B. (2012). Poisson tests of declustered catalogues. *Geophysical journal international*, 189(1), 691–700. 14
- Main, I. G., Meredith, P. G., & Sammonds, P. R. (1992). Temporal variations in seismic event rate and b-values from stress corrosion constitutive laws. *Tectonophysics*, 211(1-4), 233–246. 15
- Marsan, D., & Lengline, O. (2008). Extending earthquakes' reach through cascading. *Science*, 319(5866), 1076–1079. 15, 20
- Marzocchi, W., & Taroni, M. (2014). Some thoughts on declustering in probabilistic seismic-hazard analysis. *Bulletin of the Seismological Society of America*, 104(4), 1838–1845. 15, 25

- Marzocchi, W., Spassiani, I., Stallone, A., & Taroni, M. (2020). How to be fooled searching for significant variations of the b-value. *Geophysical Journal International*, 220(3), 1845–1856. [15](#)
- Meletti, C., Marzocchi, W., Albarello, D., D’Amico, V., Luzi, L., Martinelli, F., Pace, B., Pignone, M., Rovida, A. N., Visini, F., et al. (2017). The 2016 italian seismic hazard model. *16th World Conference on Earthquake Engineering*. [14](#), [20](#)
- Mignan, A., & Woessner, J. (2012). Estimating the magnitude of completeness for earthquake catalogs. *Community Online Resource for Statistical Seismicity Analysis*, 1–45. [17](#)
- Nandan, S., Ouillon, G., & Sornette, D. (2019). Magnitude of earthquakes controls the size distribution of their triggered events. *Journal of Geophysical Research: Solid Earth*, 124(3), 2762–2780. [17](#)
- Nandan, S., Ouillon, G., Sornette, D., & Wiemer, S. (2019a). Forecasting the full distribution of earthquake numbers is fair, robust, and better. *Seismological Research Letters*, 90(4), 1650–1659. [15](#)
- Nandan, S., Ouillon, G., Sornette, D., & Wiemer, S. (2019b). Forecasting the rates of future aftershocks of all generations is essential to develop better earthquake forecast models. *Journal of Geophysical Research: Solid Earth*, 124(8), 8404–8425. [18](#), [25](#)
- Nandan, S., Ouillon, G., Wiemer, S., & Sornette, D. (2017). Objective estimation of spatially variable parameters of epidemic type aftershock sequence model: Application to california. *Journal of Geophysical Research: Solid Earth*, 122(7), 5118–5143. [18](#)
- Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2), 379–402. [14](#), [18](#)
- Pace, B., Peruzza, L., Lavecchia, G., & Boncio, P. (2006). Layered seismogenic source model and probabilistic seismic-hazard analyses in central italy. *Bulletin of the Seismological Society of America*, 96(1), 107–132. [14](#)
- Petersen, M. D., Mueller, C. S., Moschetti, M. P., Hoover, S. M., Rukstales, K. S., McNamara, D. E., Williams, R. A., Shumway, A. M., Powers, P. M., Earle, P. S., et al. (2018). 2018 one-year seismic hazard forecast for the central and eastern united states from induced and natural earthquakes. *Seismological Research Letters*, 89(3), 1049–1061. [14](#), [15](#)
- Petrucelli, A., Gasperini, P., Tormann, T., Schorlemmer, D., Rinaldi, A. P., Vannucci, G., & Wiemer, S. (2019). Simultaneous dependence of the earthquake-size distribution on faulting style and depth. *Geophysical Research Letters*, 46(20), 11044–11053. [15](#)
- Reasenber, P. (1985). Second-order moment of central california seismicity, 1969–1982. *Journal of Geophysical Research: Solid Earth*, 90(B7), 5479–5495. [15](#), [18](#), [20](#)
- Schoenberg, F. P., Chu, A., & Veen, A. (2010). On the relationship between lower magnitude thresholds and bias in epidemic-type aftershock sequence parameter estimates. *Journal of Geophysical Research: Solid Earth*, 115(B4). [17](#)
- Schorlemmer, D., Gerstenberger, M., Wiemer, S., Jackson, D., & Rhoades, D. (2007). Earthquake likelihood model testing. *Seismological Research Letters*, 78(1), 17–29. [15](#)
- Schorlemmer, D., & Gerstenberger, M. (2007). Relm testing center. *Seismological Research Letters*, 78(1), 30–36. [17](#), [20](#)
- Schorlemmer, D., Wiemer, S., & Wyss, M. (2005). Variations in earthquake-size distribution across different stress regimes. *Nature*, 437(7058), 539–542. [15](#)
- Sesetyan, K., Demircioglu, M. B., Duman, T. Y., Can, T., Tekin, S., Azak, T. E., & Ferkan, Ö. Z. (2018). A probabilistic seismic hazard assessment for the turkish territory—part i: The area source model. *Bulletin of Earthquake Engineering*, 16(8), 3367–3397. [14](#), [20](#)
- Uhrhammer, R. (1986). Characteristics of northern and central california seismicity. *Earthquake Notes*, 57(1), 21. [15](#), [18](#), [20](#)

- van Stiphout, T., Schorlemmer, D., & Wiemer, S. (2011). The effect of uncertainties on estimates of background seismicity rate. *Bulletin of the Seismological Society of America*, *101*(2), 482–494. [15](#)
- van Stiphout, T., Zhuang, J., & Marsan, D. (2012). Seismicity declustering. *Community Online Resource for Statistical Seismicity Analysis*, *10*(1). [14](#), [15](#), [20](#)
- Veen, A., & 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. [18](#)
- Wang, Q., Jackson, D. D., & Zhuang, J. (2010). Missing links in earthquake clustering models. *Geophysical Research Letters*, *37*(21). [17](#)
- Waseem, M., Lateef, A., Ahmad, I., Khan, S., & Ahmed, W. (2019). Seismic hazard assessment of afghanistan. *Journal of Seismology*, *23*(2), 217–242. [14](#), [20](#)
- Wells, D. L., & Coppersmith, K. J. (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of the seismological Society of America*, *84*(4), 974–1002. [20](#)
- Wiemer, S., Garcia-Fernandez, M., & Burg, J.-P. (2009). Development of a seismic source model for probabilistic seismic hazard assessment of nuclear power plant sites in switzerland: The view from pegasos expert group 4 (eg1d). *Swiss Journal of Geosciences*, *102*(1), 189–209. [15](#)
- Wiemer, S., Giardini, D., Fäh, D., Deichmann, N., & Sellami, S. (2009). Probabilistic seismic hazard assessment of switzerland: Best estimates and uncertainties. *Journal of Seismology*, *13*(4), 449. [14](#)
- Wiemer, S., & Wyss, M. (1997). Mapping the frequency-magnitude distribution in asperities: An improved technique to calculate recurrence times? *Journal of Geophysical Research: Solid Earth*, *102*(B7), 15115–15128. [15](#)
- Woessner, J., Laurentiu, D., Giardini, D., Crowley, H., Cotton, F., Grünthal, G., Valensise, G., Arvidsson, R., Basili, R., Demircioglu, M. B., et al. (2015). The 2013 european seismic hazard model: Key components and results. *Bulletin of Earthquake Engineering*, *13*(12), 3553–3596. [14](#), [20](#)
- Wyss, M., Shimazaki, K., & Wiemer, S. (1997). Mapping active magma chambers by b values beneath the off-ito volcano, japan. *Journal of Geophysical Research: Solid Earth*, *102*(B9), 20413–20422. [15](#)
- Zaliapin, I., Gabrielov, A., Keilis-Borok, V., & Wong, H. (2008). Clustering analysis of seismicity and aftershock identification. *Physical review letters*, *101*(1), 018501. [15](#), [18](#), [20](#)
- Zhuang, J., & Ogata, Y. (2006). Properties of the probability distribution associated with the largest event in an earthquake cluster and their implications to foreshocks. *Physical Review E*, *73*(4), 046134. [16](#)
- Zhuang, J., Ogata, Y., & Vere-Jones, D. (2002). Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, *97*(458), 369–380. [15](#), [18](#), [20](#)

CHAPTER 3

EMBRACING DATA INCOMPLETENESS FOR BETTER EARTHQUAKE FORECASTING

Published as:

Mizrahi, L., Nandan, S., & Wiemer, S. (2021). Embracing Data Incompleteness for Better Earthquake Forecasting. *Journal of Geophysical Research: Solid Earth*, 126, e2021JB022379.
<https://doi.org/10.1029/2021JB022379>

Abstract

We propose two methods to calibrate the parameters of the epidemic-type aftershock sequence (ETAS) model based on expectation maximization (EM) while accounting for temporal variation of catalog completeness. The first method allows for model calibration on long-term earthquake catalogs with temporal variation of the completeness magnitude, m_c . This calibration technique is beneficial for long-term probabilistic seismic hazard assessment (PSHA), which is often based on a mixture of instrumental and historical catalogs. The second method generalizes the concept of m_c , considering rate- and magnitude-dependent detection probability, and allows for self-consistent estimation of ETAS parameters and high-frequency detection incompleteness. With this approach, we aim to address the potential biases in parameter calibration due to short-term aftershock incompleteness, embracing incompleteness instead of avoiding it. Using synthetic tests, we show that both methods can accurately invert the parameters of simulated catalogs. We then use them to estimate ETAS parameters for California using the earthquake catalog since 1932. To explore how model calibration, inclusion of small events, and accounting for short-term incompleteness affect earthquakes' predictability, we systematically compare variants of ETAS models based on the second approach in pseudo-prospective forecasting experiments for California. Our proposed model significantly outperforms the ETAS null model, with decreasing information gain for increasing target magnitude threshold. We find that the ability to include small earthquakes for simulation of future scenarios is the primary driver of the improvement and that accounting for incompleteness is necessary. Our results have significant implications for our understanding of earthquake interaction mechanisms and the future of seismicity forecasting.

Plain language summary

Our capability to detect earthquakes varies with time, on one hand because more and better instruments are being deployed over time, leading to long-term changes of detection capability. On the other hand, earthquakes are more difficult to be detected when seismic activity is high, which manifests in short-term changes of detection capability. Incomplete detection can lead to biases in epidemic-type aftershock sequence (ETAS) models used for earthquake forecasting. We propose two methods that allow us to calibrate these models while accounting for long-term (first method) and short-term (second method) changes in detection capability, which allows us to use a larger and more representative fraction of the available data. We test both methods on synthetic data and then apply them to the Californian earthquake catalog. Using the second method, we test how small earthquakes can help us improve ETAS forecasts. We find that the ability to include small earthquakes in simulations yields superior ETAS forecasts, and that it is necessary to correct for short-term incompleteness to achieve this superiority. The positive effect on forecasting is strongest when forecasting relatively small events, and decreases when forecasting larger events. These results have important implications for our understanding of earthquake interactions and for the future of earthquake forecasting.

3.1 Introduction

One of the key challenges in the field of statistical seismology, or seismology in general, is the development of accurate earthquake forecasting models, with epidemic-type aftershock sequence (ETAS) models (see Ogata, 1998; Veen and Schoenberg, 2008; Nandan et al., 2017) currently being widely used for this purpose. ETAS models account for the spatio-temporal clustering of earthquakes, and they have been shown in retrospective, pseudo-prospective, and prospective forecasting experiments to be among the best-performing earthquake fore-

casting models available today (Nandan, Ouillon, Sornette, and Wiemer, 2019b; Taroni et al., 2018, Woessner et al., 2011; Cattania et al., 2018; Mancini et al., 2019; Mancini et al., 2020). Furthermore, they are used for operational earthquake forecasts by the USGS (Field et al., 2017), in Italy (Marzocchi et al., 2014), and New Zealand (D. Rhoades et al., 2016).

A fundamental requirement for reliable parameter estimation of the ETAS model is the completeness of the training catalog above the magnitude of completeness, m_c . As we can not know with certainty what we did not observe, m_c itself needs to be estimated, and numerous approaches to this problem have been proposed (Wiemer and Wyss, 2000; Cao and Gao, 2002; Woessner and Wiemer, 2005; Amorese, 2007; Rydelek and Sacks, 1989; see Mignan and Woessner (2012) for an overview). m_c is known to vary in space and time, due to gradual improvement of the seismic network, software upgrades, and so on. Several methods have been proposed to estimate its spatial and temporal variation (Wiemer and Wyss, 2000; Mignan et al., 2011; Woessner and Wiemer, 2005; Amato and Mele, 2008; Nanjo et al., 2010; Hutton et al., 2010; Mignan and Chouliaras, 2014; Schorlemmer and Woessner, 2008; Hainzl, 2016b). In particular, Hainzl (2016b) addresses an additional important cause of variation in time of m_c , short-term aftershock incompleteness (STAI). Because earthquakes strongly cluster in time, seismic networks can only capture a subset of events during periods of high activity (Kagan, 2004).

As mentioned earlier, a reliable estimation of ETAS parameter depends on a reliable estimate of m_c . Although the biasing effects on ETAS parameter estimates caused by data incompleteness are known and discussed (Hainzl, 2016b; Seif et al., 2017; Zhuang et al., 2017), nearly all applications of the ETAS model assume for simplicity a global magnitude of completeness for the entirety of the training period. This assumption is problematic in several ways.

First, in order to be complete for the entire training period, the modeller is often forced to use very conservative estimates of m_c , as a result completely ignoring abundant and high-quality data from more complete periods. Furthermore, m_c is often assumed to be equal to the minimum magnitude of earthquakes that can trigger aftershocks, m_0 , and this conservative assumption can introduce a bias to ETAS parameter estimates. This idea that earthquakes below m_c are relevant for our understanding of earthquakes' clustering behavior was thoroughly discussed by Sornette and Werner (2005a), who pointed out the important distinction between m_0 and m_c , also providing constraints for m_0 . Although small earthquakes trigger fewer aftershocks than large ones do, Marsan (2005), as well as Helmstetter et al. (2005), found that small earthquakes, being more numerous, are as important as large ones for earthquake triggering. Thus, it is natural to assume that a larger difference between m_c and m_0 will lead to a larger bias in the estimated parameters.

Alternatively, one may estimate ETAS parameters from catalogs with restricted space-time volume which can have low overall m_c values. Parameters estimated in this way can however be dominated by one or two sequences and may not represent long-term behavior, thus making the use of ETAS models non-ideal for long-term probabilistic seismic hazard assessment (PSHA). Instead, the modellers rely on smoothed seismicity approaches based on declustered catalogs (see e.g. Gerstenberger et al., 2020; Petersen et al., 2018; Wiemer et al., 2009), which is a problematic approach due to the biasing effects of declustering on the size distribution of mainshocks, and thus on the estimated seismic hazard (Mizrahi et al., 2021). In this regard, Marzocchi and Taroni (2014) discussed the need for spatial declustering so as not to distort future seismic hazard, and Llenos and Michael (2020) proposed an approach to calculate regionally optimized background earthquake rates from ETAS to be used for the U.S. Geological Survey National Seismic Hazard Model (NSHM), stressing the need for methods to address catalog heterogeneities such as time-dependent incompleteness.

Additionally, with the assumption of a constant overall m_c , the crucial requirement of completeness of the training catalog is not fulfilled during large aftershock sequences, which can bias the estimated parameters. Several studies have highlighted the importance of considering short-term variation of m_c in the context of ETAS models. These include Hainzl (2016b) and Hainzl (2016a), who modeled STAI based on the short-term rate of earthquakes, bringing into relation true and apparent triggering laws; Stallone and Falcone (2020), who proposed a method to stochastically replenish catalogs suffering from STAI, to be used for better operational earthquake forecasting and hazard assessment, albeit without addressing the effectiveness of the method in this regard; Zhuang et al. (2017), who showed that estimating ETAS parameters using a replenished catalog is more stable with respect to cutoff magnitude; Omi et al. (2014), who proposed a method to estimate parameters of the ETAS model from incompletely observed aftershock sequences, by statistically modelling detection deficiency.

In this article, we thoroughly address the use of small earthquakes for seismic hazard forecasting. For this, we develop two complementary methods with which long-term (first method) and short-term (second method) temporal variations of m_c can be accounted for when calibrating ETAS models and when issuing ETAS-based forecasts. The first method extends the expectation maximization scheme for ETAS parameter inversion (Veen and Schoenberg, 2008) for application to training catalogs with time-varying completeness magnitude $m_c(t)$. This simultaneously allows the inclusion of historical data in the parameter inversion, as well as the inclusion of small magnitude events, which make up a large fraction of data and can enable the ability to more clearly illuminate faults. ETAS models can hence be trained on a more representative and informative set of data, which in some areas facilitates a more appropriate approach to PSHA. With the second method proposed in this article, we want to utilize the knowledge about clustering derived using the ETAS model to quantitatively estimate the level of completeness of a catalog at any given time, and then use this knowledge to minimize the incompleteness-induced bias in the ETAS model. We approach this issue by generalizing the notion of m_c , moving from a binary completeness space (complete versus incomplete) to a continuous-valued completeness space by means of a magnitude-dependent detection probability – embracing incompleteness instead of avoiding it, as has been proposed previously by Ogata and Katsura (2006) and Omi et al. (2014). While the first method described in this article allows $m_c(t)$ as an input to the ETAS parameter calibration, which makes it powerful in a long-term context, the second method addresses the additional challenge of estimating short-term variations of completeness. To understand their abilities and limitations, we subject both methods to rigorous synthetic tests. Then, we apply them to Californian earthquake data and interpret the results in light of the findings of the synthetic tests. Using the second approach, we systematically assess how the inclusion of small earthquakes, which may be incompletely detected, affects the performance of earthquake forecasts. We conduct pseudo-prospective 30-day forecasting experiments for California, designed to answer several questions: Does our new model outperform the current state of the art? If so, what is the role of the newly estimated ETAS parameters in this improvement? Similarly, what is the role of newly included small earthquakes in this improvement, and the role of the estimated high-frequency detection incompleteness? How do the models perform for different target magnitude thresholds?

The remainder of the paper is structured as follows. Section 3.2 describes the earthquake catalog that was used in this analysis. The modified ETAS parameter inversion methods are presented in Section 3.3.1 for time-varying m_c , and in Section 3.3.2 for time-varying probabilistic detection incompleteness. Sections 3.3.3 and 3.3.4 describe the formulation of probabilistic detection incompleteness and the algorithm for joint estimation of ETAS parameters and detection probability. Section 3.4 presents synthetic tests for both methods, and Section 3.5 presents applications of both methods to the Californian data. Section 3.6

describes pseudo-prospective forecasting experiments used to assess the impact of the newly acquired information on the forecastability of earthquakes in California. Finally, in Section 3.7, we present our conclusions.

3.2 Data

In this article, we use the ANSS Comprehensive Earthquake Catalog (ComCat) provided by the U.S. Geological Survey. We adopt the preferred magnitudes as defined in ComCat, and use as study region the collection area around the state of California as proposed in the RELM testing center (Schorlemmer and Gerstenberger, 2007). We consider events of magnitude $M \geq 0.0$, with magnitudes rounded into bins of size $\Delta M = 0.1$. For the major part of the study, the time frame used is January 1, 1970 until December 31, 2019. For the analysis of long-term variations in m_c , we extend the time frame to start on January 1, 1932, when instrumentation was introduced to the Californian seismic network (Felzer, 2007).

Whenever ETAS parameters are inverted, we use the first fifteen years of data to serve as auxiliary data. Thus, the start of the primary catalog is either January 1985, or January 1947. Earthquakes in the auxiliary catalog may act as triggering earthquakes in the ETAS model, but not as aftershocks.

To estimate a constant magnitude of completeness of the catalog, we use the method described by Mizrahi et al. (2021) with an acceptance threshold value of $p = 0.1$, which yields $m_c = 3.1$ for the time period between 1970 and 2019. This method is adapted from Clauset et al. (2009) and jointly estimates m_c and the b -value of the Gutenberg-Richter law (Gutenberg and Richter, 1944) describing earthquake size distribution. It compares the Kolmogorov-Smirnov (KS) distance between the observed cumulative distribution function (CDF) and the fitted GR law to KS distances obtained for magnitude samples simulated from said GR law. A value of m_c is accepted if at least a fraction of $p = 0.1$ of KS distances is larger than the observed one.

3.3 Model

3.3.1 ETAS parameter inversion for time-varying m_c

Consider an earthquake catalog

$$C = \{e_i = (m_i, t_i, x_i, y_i), i \in \{1, \dots, n\}\} \quad (3.1)$$

consisting of events e_i of magnitudes m_i which occur at times t_i and locations (x_i, y_i) . Furthermore, consider a time-varying magnitude of completeness $m_c(t)$ defined for all t_i . We say that the catalog is complete if $m_i \geq m_c(t_i) \forall i$.

The ETAS model describes earthquake rate as

$$l(t, x, y) = \mu + \sum_{i:t_i < t} g(m_i, t - t_i, x - x_i, y - y_i). \quad (3.2)$$

That is, the sum of background rate μ and the rate of all aftershocks of previous events e_i . The aftershock triggering rate $g(m, \Delta t, \Delta x, \Delta y)$ describes the rate of aftershocks triggered by an event of magnitude m , at a time delay of Δt and a spatial distance $(\Delta x, \Delta y)$ from the triggering event. We here use the definition

$$g(m, \Delta t, \Delta x, \Delta y) = \frac{k_0 \cdot e^{a(m-m_{ref})}}{\frac{(\Delta t+c)^{1+\omega}}{e^{-\frac{\Delta t}{\tau}}} \cdot ((\Delta x^2 + \Delta y^2) + d \cdot e^{\gamma(m-m_{ref})})^{1+\rho}}, \quad (3.3)$$

as in Nandan et al. (2017).

To calibrate the ETAS model, the nine parameters to be optimized are the background rate μ and the parameters $k_0, a, c, \omega, \tau, d, \gamma, \rho$ which parameterize the aftershock triggering rate $g(m, t, x, y)$ given in Equation (3.3). Implicitly, the model assumes that all earthquakes with magnitude larger than or equal to m_{ref} can trigger aftershocks. We build on the expectation maximization (EM) algorithm to estimate the ETAS parameters (Veen and Schoenberg, 2008). In this algorithm, the expected number of background events \hat{n} and the expected number of directly triggered aftershocks \hat{l}_i of each event e_i are estimated in the expectation step (E step), along with the probabilities p_{ij} that event e_j was triggered by event e_i , and the probability p_j^{ind} that event e_j is independent. Following the E step, the nine parameters are optimized to maximize the complete data log likelihood in the maximization step (M step). E and M step are repeated until convergence of the parameters. The usual formulation of the EM algorithm defines

$$p_{ij} = \frac{g_{ij}}{\mu + \sum_{k:t_k < t_j} g_{kj}}, \quad (3.4)$$

$$p_j^{ind} = \frac{\mu}{\mu + \sum_{k:t_k < t_j} g_{kj}}, \quad (3.5)$$

with $g_{kj} = g(m_k, t_j - t_k, x_j - x_k, y_j - y_k)$ being the aftershock triggering rate of e_k at location and time of event e_j . For a given target event e_j , Equations (3.4-3.5) define p_{ij} to be proportional to the aftershock occurrence rate g_{ij} , and p_j^{ind} to be proportional to the background rate μ . As an event must be either independent or triggered by a previous event, the normalization factor $\Lambda_j := \mu + \sum_{k:t_k < t_j} g_{kj}$ in the denominator of Equations (3.4-3.5) stipulates that $p_j^{ind} + \sum_{k:t_k < t_j} p_{kj} = 1$. This relies on the assumption that all potential triggering earthquakes of e_j were observed, that is, all events prior to t_j above the reference magnitude (minimum considered magnitude), m_{ref} were observed. To fulfill this requirement, most applications of the method define m_{ref} to be equal to the constant value of m_c .

For the case of time-varying $m_c(t)$, we define $m_{ref} := \min_i \{m_c(t_i)\}$, the minimum $m_c(t_i)$ for times t_i of events in the complete catalog. This implies that for the times when $m_c(t) > m_{ref}$ the requirement of complete recording of all potential triggers may be violated. Events whose magnitudes fall between m_{ref} and $m_c(t)$ are not part of the complete catalog and are considered to be unobserved (even though they may have been detected by the network). Hence, the normalization factor Λ_j (the denominator of Equations (3.4-3.5)) needs to be adapted to account for the possibility that e_j was triggered by an unobserved event.

Consider

$$\xi(t) = \frac{\int_{m_{ref}}^{m_c(t)} f_{GR}(m) \cdot G(m) dm}{\int_{m_c(t)}^{\infty} f_{GR}(m) \cdot G(m) dm}, \quad (3.6)$$

the ratio between the expected number of events triggered by an unobserved event and the expected number of events triggered by an observed event at time t . Here, $f_{GR} = \beta \cdot e^{-\beta \cdot (m-m_{ref})}$ is the probability density function of magnitudes according to the GR law, and $G(m) = \int_0^\infty \iint_R g(m, t, x, y) dx dy dt$ is the total number of expected aftershocks larger than m_{ref} of an event of magnitude m . Note that in the calculation of $G(m)$ we make

the simplifying assumption that the considered region R extends infinitely in all directions, allowing a facilitated, asymptotically unbiased estimation of ETAS parameters (Schoenberg, 2013). Analogously,

$$\zeta(t) = \frac{\int_{m_{ref}}^{m_c(t)} f_{GR}(m) dm}{\int_{m_c(t)}^{\infty} f_{GR}(m) dm} \quad (3.7)$$

is the ratio between the expected fraction of unobserved events and the expected fraction of observed events at time t . If $\beta > a - \rho\gamma$, both $\xi(t)$ and $\zeta(t)$ are well-defined and we have that

$$\xi(t) = e^{-(a-\beta-\rho\gamma)\cdot\Delta m(t)} - 1, \quad (3.8)$$

$$\zeta(t) = e^{\beta\cdot\Delta m(t)} - 1, \quad (3.9)$$

where $\Delta m(t) = m_c(t) - m_{ref}$. Consider the productivity exponent $\alpha := a - \rho \cdot \gamma$, which describes the exponential relationship between aftershock productivity and magnitude of an event. The condition that β is larger than the productivity exponent α is generally fulfilled in naturally observed catalogs (Helmstetter, 2003). If this were not the case, earthquake triggering would be dominated by large events and one would need to introduce a maximum possible magnitude for both denominators to be finite (see available equations in Sornette and Werner (2005b); Sornette and Werner (2005a)). The normalization factor Λ_j consists of the sum of background rate and aftershock rates of all events which happened prior to e_j . In the case of time-varying m_c , besides the possibilities of being a background event or being triggered by an observed event, the event e_j can also be triggered by an unobserved event. We thus generalize Λ_j by adding to the rate of aftershocks g_{kj} of each observed triggering event e_k the expected rate of aftershocks of unobserved triggering events at that time, $g_{kj} \cdot \xi(t_k)$. This yields $\Lambda_j = \mu + \sum_{k:t_k < t_j} g_{kj} \cdot (1 + \xi(t_k))$ and thus the generalized definition of p_{ij} and p_j^{ind} is given by

$$p_{ij} = \frac{g_{ij}}{\mu + \sum_{k:t_k < t_j} g_{kj} \cdot (1 + \xi(t_k))}, \quad (3.10)$$

$$p_j^{ind} = \frac{\mu}{\mu + \sum_{k:t_k < t_j} g_{kj} \cdot (1 + \xi(t_k))}. \quad (3.11)$$

Note that the probability p_{uj} that event e_j was triggered by an unobserved event is given such that $p_j^{ind} + p_{uj} + \sum_{k:t_k < t_j} p_{kj} = 1$. In the above equations, the special case of $m_c(t) \equiv m_{ref}$ is accounted for when $\xi(t) \equiv 0$. In this special case, \hat{n} and \hat{l}_i are obtained by summing independence probabilities ($\hat{n} = \sum_j p_j^{ind}$) and triggering probabilities ($\hat{l}_i = \sum_j p_{ij}$), respectively. In the generalized case however, \hat{n} and \hat{l}_i are the estimated number of background events and aftershocks of event e_i above m_{ref} , which includes unobserved events. Similarly to inflating the triggering power, we hence inflate the observed event numbers to account for unobserved events. Whenever an event is observed at time t_j , we expect that $\zeta(t_j)$ events occurred under similar circumstances (with same independence and triggering probabilities), but were not observed. This yields

$$\hat{n} = \sum_j p_j^{ind} \cdot (1 + \zeta(t_j)), \quad (3.12)$$

$$\hat{l}_i = \sum_j p_{ij} \cdot (1 + \zeta(t_j)). \quad (3.13)$$

With these adapted definitions of p_{ij} , p_j^{ind} , \hat{n} and \hat{l}_i (Equations (3.10 - 3.13)), ETAS parameters can be inverted using the procedure described by Veen and Schoenberg (2008).

3.3.2 ETAS parameter inversion for time-varying probabilistic detection

To overcome the binary view of completeness which forces us to disregard earthquakes which were detected but happen to fall between m_{ref} and $m_c(t)$, we can take the generalization of the EM algorithm for ETAS parameter inversion one step further by introducing a time and magnitude-dependent probability of detection,

$$f: \mathbb{R}_{\geq m_{ref}} \times \mathbb{R} \longrightarrow [0, 1] \\ (m, t) \mapsto p.$$

To be able to account for such a probabilistic concept of catalog completeness in the ETAS inversion algorithm, one needs to generalize $\xi(t)$ and $\zeta(t)$ (Equations (3.6) and (3.7)). In contrast to before, the magnitude of an event does not determine whether or not the event has been detected. We therefore adapt the bounds of integration in numerator and denominator such that all events above magnitude m_{ref} are considered. To obtain the expected number of earthquakes triggered by observed and unobserved events, the integrands are multiplied by the probability of the triggering events to be observed, $f(m, t)$, or unobserved, $(1 - f(m, t))$, respectively. The generalized formulations of $\xi(t)$ and $\zeta(t)$ then read

$$\xi(t) = \frac{\int_{m_{ref}}^{\infty} (1 - f(m, t)) \cdot f_{GR}(m) \cdot G(m) dm}{\int_{m_{ref}}^{\infty} f(m, t) \cdot f_{GR}(m) \cdot G(m) dm}, \quad (3.14)$$

and

$$\zeta(t) = \frac{\int_{m_{ref}}^{\infty} (1 - f(m, t)) \cdot f_{GR}(m) dm}{\int_{m_{ref}}^{\infty} f(m, t) \cdot f_{GR}(m) dm}. \quad (3.15)$$

For compatible choices of $f(m, t)$, $f_{GR}(m)$, $G(m)$, we find that $\xi(t)$ and $\zeta(t)$ are well-defined. Consider for instance the special case of binary detection, where $f(m, t)$ is defined via the Heaviside step function H as $f_{bin}(m, t) = H(m - m_c(t))$, which is equal to 1 if $m \geq m_c(t)$ and 0 otherwise. This is the case discussed in the previous section, for which we have well-definedness if $\beta > a - \rho\gamma$.

The reference magnitude m_{ref} is a model constant. Smaller values of m_{ref} allow the modeller to use a larger fraction of the observed catalog, which can be especially useful in regions with less seismic activity.

Note that both generalizations of the ETAS inversion algorithm (for time-varying completeness or for time-varying probabilistic detection) can without further modification be applied when m_c or detection probability vary with space. The formulation is based on the assumption that the behavior of observed events is locally representative (in space and/or time) of the behavior of unobserved events.

3.3.3 Rate-dependent probabilistic detection incompleteness

In this section we present our approach to define $f(m, t)$, where the temporal component is purely driven by the current rate of events. Note that this means we only capture changes in detection due to changes in short-term circumstances, and neglect long-term changes due to network updates. We make the following simplifying assumptions.

- Any earthquake will obstruct the entire seismic network from detecting smaller earthquakes for a duration of t_R (recovery time of the network).
- Magnitudes of events that are simultaneously blocking the network are distributed according to the time-invariant Gutenberg-Richter law which also describes the magnitude distribution of the full catalog (Gutenberg and Richter, 1944).

De Arcangelis et al. (2018) found that short-term aftershock incompleteness can be well explained in terms of overlapping seismic records, while instrumental coverage of an area plays a subsidiary role. Nevertheless, assuming t_R to be independent of the magnitude of the event, and independent of the spatial distance between the event and the locations of interest, is certainly a major simplification which could be refined in subsequent studies.

Conveniently, the ETAS model provides a simple way of calculating the current rate of events in the region R as

$$\lambda(t) = \iint_R l(t, x, y) dx dy = \iint_R \mu + \sum_{i:t_i < t} g(m_i, t - t_i, x - x_i, y - y_i) dx dy. \quad (3.16)$$

For the remainder of this paper, we will refer to the current rate of events in the region, $\lambda(t)$, simply as the current rate of events. The probability $f(m, t)$ of an earthquake to be detected is then given by the probability of it being the largest of all the earthquakes that are currently blocking the network. Consider

$$f(m, t) = \left(1 - e^{-\beta \cdot (m - m_{ref})}\right)^{t_R \cdot \lambda(t)}. \quad (3.17)$$

Here, $t_R \cdot \lambda(t)$ is an approximation of the expected number of events blocking the network at time t , and the term $1 - e^{-\beta \cdot (m - m_{ref})}$ is the probability of any given earthquake's magnitude falling between m_{ref} and m , where $\beta = b \cdot \ln 10$ is the exponent in the GR law with basis e . Thus, $f(m, t)$ is the probability that in the set of $t_R \cdot \lambda(t)$ events currently blocking the network, all of them have a magnitude of less than m , which is the condition for an event of magnitude m to be detected. Because the time-dependence of $f(m, t)$ is solely controlled by the time-dependence of λ , we here use the terms $f(m, t)$ and $f(m, \lambda)$ interchangeably.

Plugging this definition of $f(m, t)$ into Equations (3.14) and (3.15), and setting $\kappa := -\frac{a - \rho\gamma}{\beta}$, we obtain

$$\xi(t) = \frac{1}{(\kappa + 1) \cdot \text{B}(\kappa + 1, t_R \cdot \lambda(t) + 1)} - 1, \quad (3.18)$$

$$\zeta(t) = t_R \cdot \lambda(t), \quad (3.19)$$

so long as $\beta > a - \rho\gamma$, where B is the Beta function. A positive background rate $\mu > 0$ ensures $\lambda(t) > 0 \quad \forall t$. Expressions analogous to (3.18) and (3.19) hold when alternative exponents are chosen instead of $t_R \cdot \lambda$ in the definition of $f(m, \lambda)$ (Equation (3.17)).

The network recovery time t_R and the current event rate $\lambda(t)$ at the times t_i of all earthquakes e_i need to be estimated from the data.

3.3.4 Estimating probabilistic epidemic-type aftershock incompleteness (PETAI)

Estimation of (t_R, β) when λ_i are known

The function $f(m, t)$ brings with it two parameters, t_R and β , which need to be estimated in addition to the ETAS parameters. We here describe how t_R and β can be jointly estimated using a maximum likelihood approach for the case when current event rates $\lambda_i = \lambda(t_i)$ are known. In reality, the λ_i have to be estimated themselves. This is described in Section 3.3.4.

In the case when the true ETAS parameters, as well as the current event rates $\lambda(t_i)$ for all events e_i in the primary catalog $\{e_1, \dots, e_n\}$, are known, the GR-law exponent β and the network recovery time t_R can be estimated by optimizing the log-likelihood $\mathcal{L}\mathcal{L}$ of observing the catalog at hand.

$$\begin{aligned} \mathcal{L}\mathcal{L} &= \sum_{i=1}^n (\ln(\nu_i + 1) - \ln N) \\ &\quad + \sum_{i=1}^n \left(\nu_i \cdot \ln(1 - e^{-\beta \cdot (m_i - m_{ref})}) \right) \\ &\quad + \sum_{i=1}^n (\ln \beta - \beta \cdot (m_i - m_{ref})), \end{aligned} \quad (3.20)$$

where $N = \sum_{i=1}^n (\nu_i + 1)$, and $\nu_i = t_R \cdot \lambda(t_i)$ is an approximation of the expected number of events blocking the network at time t_i . The expression for $\mathcal{L}\mathcal{L}$ given above is valid in general for alternative exponents ν_i in the definition of detection probability (Equation (3.17)). $\mathcal{L}\mathcal{L}$ is derived from the likelihood \mathcal{L}_i of an event to have magnitude m_i and to be observed during a current event rate of $\lambda_i = \lambda(t_i)$, and the current event rate being λ_i ,

$$\mathcal{L}_i = f_{emp}(\lambda_i) \cdot f_{GR}(m_i) \cdot f_{det}(m_i, \lambda_i), \quad (3.21)$$

where $f_{GR}(m)$ is the probability density function of magnitudes given by the GR law, $f_{det}(m, \lambda)$ is the detection probability as defined in Equation (3.17), and

$$f_{emp}(\lambda) = \begin{cases} \frac{t_R \cdot \lambda + 1}{\sum_i (t_R \cdot \lambda_i + 1)}, & \text{if } \lambda \in \{\lambda_1, \dots, \lambda_n\} \\ 0, & \text{otherwise} \end{cases} \quad (3.22)$$

is the empirical density function of event rates. $f_{emp}(\lambda)$ is defined such that

$$\sum_{i=1}^n f_{emp}(\lambda_i) = 1 \quad (3.23)$$

and

$$f_{emp}(\lambda_i) \propto \frac{1}{\int_{m_{ref}}^{\infty} f_{GR}(m) \cdot f_{det}(m, \lambda_i) dm} = \lambda_i \cdot t_R + 1, \quad \forall i = 1, \dots, n. \quad (3.24)$$

Without the latter condition (Equation (3.24)), we would wrongly assume that the values $\lambda(t_i)$ were uniformly drawn from the true distribution of event rates. However, in our sample of λ_i , large values of λ are underrepresented, because during times t when $\lambda(t)$ is high, events are less likely to be detected, and those times and their corresponding rates are thus less likely to be part of our sample. Defining $f_{emp}(\lambda_i)$ to be inversely proportional to the fraction of events that are observed when the current rate is λ_i corrects for this under-representation. This yields

$$\mathcal{L}_i = \frac{\nu_i + 1}{\sum_j (\nu_j + 1)} \cdot \beta \cdot e^{-\beta \cdot (m_i - m_{ref})} \cdot \left(1 - e^{-\beta \cdot (m_i - m_{ref})}\right)^{\nu_i}, \quad (3.25)$$

which explains the term for $\mathcal{L}\mathcal{L}$ (Equation (3.20)). Figure B.1 shows the log likelihood of a synthetic test catalog for different values of t_R and β when λ_i are known. The resulting estimators match the data-generating parameters.

Estimation of λ_i when ETAS parameters and (t_R, β) are known

On one hand, the λ_i depend on the ETAS parameters (see Equation (3.16)). On the other hand, the sum of aftershocks of previous earthquakes in the definition of $\lambda(t)$ (Equation (3.16)) does not account for aftershocks of events that were not detected. As in the ETAS parameter inversion, to account for aftershocks of undetected events in the calculation of $\lambda(t)$, we inflate the triggering power of each event e_i by a factor of $1 + \xi(t_i)$ and define

$$\lambda(t) = \iint_R \mu \, dx \, dy + \sum_{i: t_i < t} (1 + \xi(t_i)) \cdot \iint_R g(m_i, t - t_i, x - x_i, y - y_i) \, dx \, dy. \quad (3.26)$$

Estimation of λ_i and (t_R, β) when ETAS parameters are known

$\xi(t)$ however requires knowledge of (t_R, β) (see Equation (3.18)). This implies that even when ETAS parameters are fixed, an additional, lower-level circular dependency dictates the relationship between $(\lambda_i)_{i=1, \dots, n}$ and (t_R, β) .

To fully estimate the high-frequency probabilistic detection incompleteness, given fixed ETAS parameters, we recursively re-estimate $(\lambda_i)_{i=1, \dots, n}$ (see Section 3.3.4) and (t_R, β) (see Section 3.3.4), until (t_R, β) meets a convergence criterion, starting with an informed or random initial guess for (t_R, β) .

3.3.5 PETAI inversion algorithm

The overarching joint inversion of ETAS parameters (\mathcal{E}) and high-frequency detection incompleteness ($\mathcal{I} = (\lambda_i, t_R, \beta)$) starts with estimating ETAS parameters in the usual way, i.e. using the algorithm described in Section 3.3.1, with a time-independent completeness magnitude $m_c (= m_{ref})$ above which all events are detected. It then recursively re-estimates \mathcal{I} (see Section 3.3.4) and \mathcal{E} (see Section 3.3.2) until convergence of the ETAS parameters.

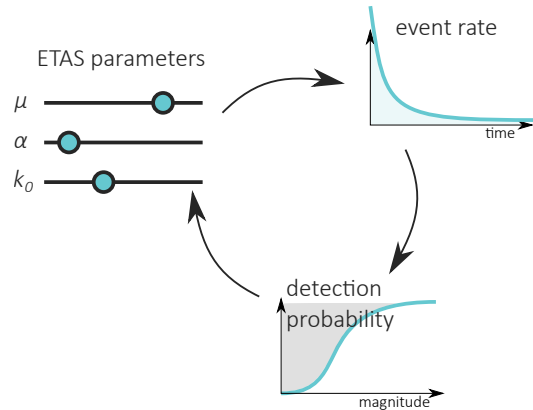


Figure 3.1: Simplified schematic illustration of PETAI inversion.

A simplified illustration of the inversion algorithm is shown in Figure 3.1. Starting with the initial ETAS parameters obtained assuming constant m_c , event rates can be calculated at each point in time. Given these event rates, the detection probability function is calibrated, which then provides insight into the temporal evolution of catalog (in-)completeness. ETAS parameters can then be re-estimated, now also using data below m_c , by accounting for the estimated incompleteness. With this new set of ETAS parameters, event rates can be re-calculated, upon which detection probability is re-calibrated, and so on, until all convergence criteria are satisfied. Figure 3.2 shows the detailed flow diagram of the PETAI inversion algorithm.

3.4 Synthetic tests

3.4.1 Synthetic Test for ETAS model with long-term variation of m_c (ST1)

To test the ETAS parameter inversion for time-varying m_c , we generate 400 complete synthetic catalogs using ETAS and then artificially impose a given $m_c(t)$ on the catalogs. Assuming $m_c(t)$ to be known, we use the method described in Section 3.3.1 to infer the parameters used in the simulation.

We estimate $m_c(t)$ based on the Californian catalog described in Section 3.2 with a time horizon from 1932 to 2019. Fixing the b -value we had estimated for the main catalog (1970 - 2019, $M \geq 3.1$, $b = 1.01 \pm 0.006$, see Mizrahi et al. (2021) for the method used), we estimate m_c for successive 10 year periods starting in 1932. The last period then comprises only 8 years of data. Estimation of m_c is analogous to the main catalog, using the method of Mizrahi et al. (2021) with an acceptance threshold of $p = 0.1$, but keeping $b = 1.01$ fixed.

This yields

$$m_c(t) = \begin{cases} 4.3 & \text{for } t \text{ between 1932 and 1941,} \\ 3.9 & \text{for } t \text{ between 1942 and 1951,} \\ 4.3 & \text{for } t \text{ between 1952 and 1961,} \\ 3.4 & \text{for } t \text{ between 1962 and 1971,} \\ 3.1 & \text{for } t \text{ between 1972 and 1981,} \\ 3.3 & \text{for } t \text{ between 1982 and 1991,} \\ 2.4 & \text{for } t \text{ between 1992 and 2001,} \\ 2.8 & \text{for } t \text{ between 2002 and 2011,} \\ 3.6 & \text{for } t \text{ between 2012 and 2019.} \end{cases} \quad (3.27)$$

The large increase in m_c for the years 2012 to 2019 is due to the Ridgecrest events in 2019. Although the period affected by aftershock incompleteness only makes up a small fraction of the 8 year period, our method with an acceptance threshold of $p = 0.1$ yields a conservative estimate of m_c . To avoid such an effect, one could use shorter than 10 year periods, or use different methods to estimate time-varying m_c .

Note that our method to invert ETAS parameters for time-varying m_c (Section 3.3.1) accepts $m_c(t)$ as an input and works independently of how this $m_c(t)$ was obtained. We here want to keep the focus on the parameter inversion and thus choose the described approach to estimate $m_c(t)$ due to its simplicity.

To mimic a realistic scenario, we simulate the synthetic catalogs using parameters obtained after applying ETAS parameter inversion for time-varying m_c on the California data, with two manual corrections.

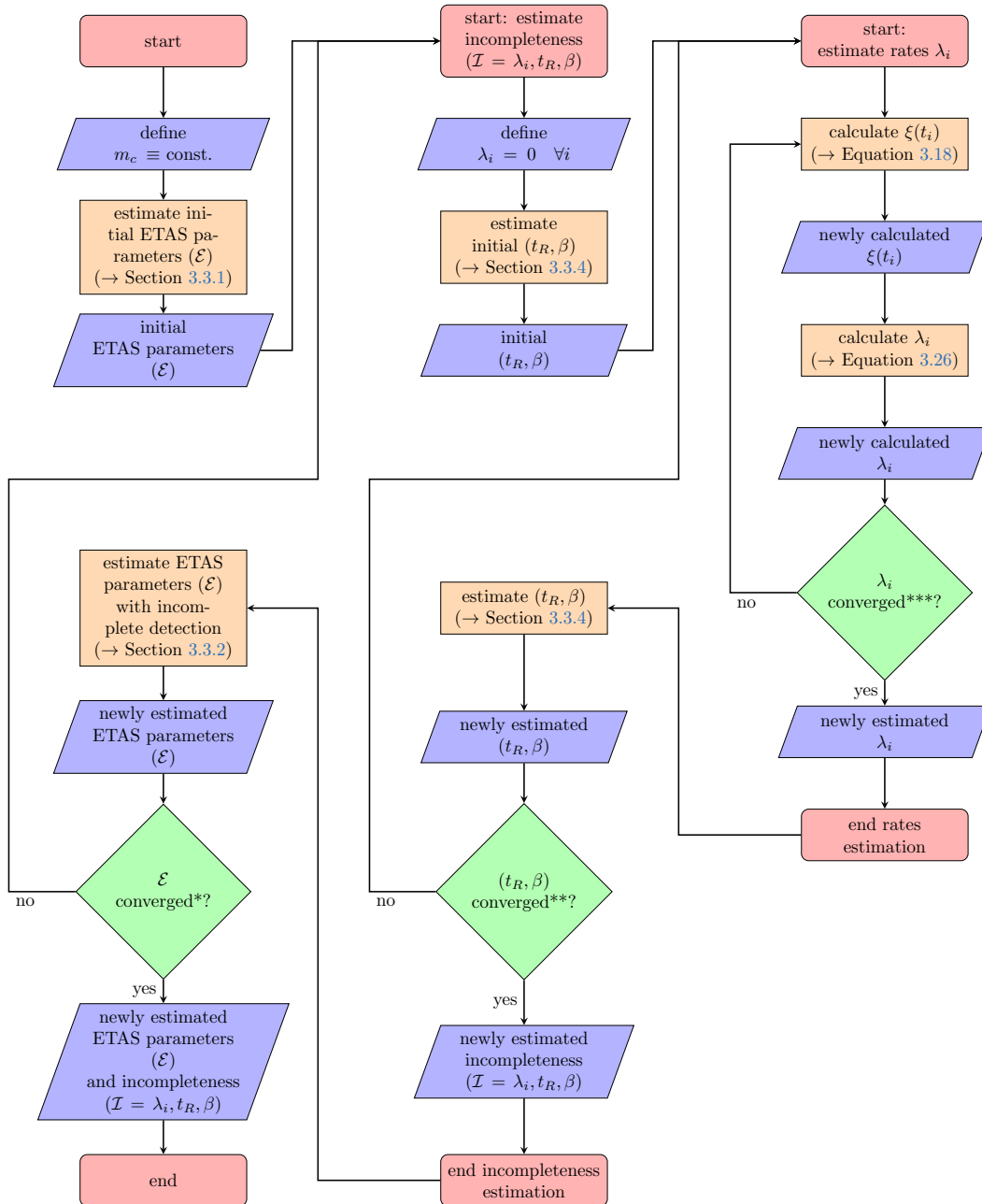


Figure 3.2: Flow diagram of PETAI inversion. Caption on next page.

Figure 3.2: (Previous page.) Flow diagram of PETAI inversion. Main algorithm starts at top left and ends at bottom left. The middle column describes the estimation of incompleteness ($\mathcal{I} = \lambda_i, t_R, \beta$) when ETAS parameters (\mathcal{E}) are given. Note that the estimation of $(\lambda_i)_{i=1, \dots, n}$ when ETAS parameters and (t_R, β) are fixed requires yet another loop to obtain self-consistency, as updating λ_i (step Λ) leads to changes in the inflation factor $1 + \xi(t_i)$, which forces one to update $(\lambda_i)_{i=1, \dots, n}$. This sub-sub-algorithm is visualized in the right column of the flow diagram. Process boxes are linked to corresponding methods and equations described in this article.

*, **, ***: Convergence is reached when the estimated values of the k^{th} iteration, \hat{a}_k , lie very close to the estimated values of the previous iteration, that is, if $\sum_{a \in A} |\hat{a}_k - \hat{a}_{k-1}| \leq \theta$. Here, A is the set of values that are tested for convergence, $*A = \mathcal{E}$, $**A = \{t_R, \beta\}$, $***A = \{\lambda_i, i = 1, \dots, n\}$. For convergence threshold θ we use $*\theta = 10^{-3}$, $**\theta = 10^{-12}$, $***\theta = 1$.

The first correction is done because it has been shown that certain assumptions in the ETAS model such as a spatially isotropic aftershock distribution or a temporally stationary background rate, as well as data incompleteness can lead to biased estimations of the productivity exponent (Hainzl et al., 2008; Hainzl et al., 2013; Seif et al., 2017). This bias can lead to a lack of clustering when catalogs are simulated. We thus use an artificially increased productivity exponent α' for our simulations as follows.

Consider the branching ratio η , defined as the expected number of direct aftershocks (larger than m_{ref}) of any earthquake larger than m_{ref} ,

$$\eta = \int_{m_{ref}}^{\infty} f_{GR}(m) \cdot G(m) dm. \quad (3.28)$$

It follows easily that

$$\eta = \frac{\beta \cdot k_0 \cdot \pi \cdot d^{-\rho} \cdot \tau^{-\omega} \cdot e^{c/\tau} \cdot \Gamma(-\omega, c/\tau)}{\rho \cdot (\beta - (a - \rho \gamma))}, \quad (3.29)$$

if $\beta > a - \rho \cdot \gamma$, where $\Gamma(s, x) = \int_x^{\infty} t^{s-1} e^{-t} dt$ is the upper incomplete gamma function.

We fix $\alpha' = 2.0$ (based on Helmstetter (2003); Z. Guo and Ogata (1997)) and from this derive new values for a and k_0 , keeping the branching ratio η constant. In particular, we define

$$a' := \alpha' + \rho \cdot \gamma, \quad (3.30)$$

$$k_0' := k_0 \cdot \frac{\beta - (a' - \rho \cdot \gamma)}{\beta - (a - \rho \cdot \gamma)}. \quad (3.31)$$

It can be easily shown that in this way, the branching ratio η remains the same as long as $\beta - (a - \rho \cdot \gamma) > 0$.

Secondly, we reduce the background rate μ . In this way, the size of the simulated catalogs is reduced such that inversion requires a reasonable amount of computational power, even for large regions and time horizons. The final parameters used for the simulation of the catalogs can be found in Figure 3.3 (black crosses).

400 catalogs of events of magnitude $M \geq 2.4 = m_{ref}$ are simulated as described in Section B.1 for the time period of January 1832 to December 2019 in a square of 40° lat \times 40° long. Because of missing long-term aftershocks in the beginning of the simulated catalogs, we allocate a burn period of 100 years in the beginning of the simulated period and are left with catalogs from 1932 to 2019. The starting year of our synthetic catalogs coincides with the introduction of instrumentation in California (Felzer, 2007). This allows

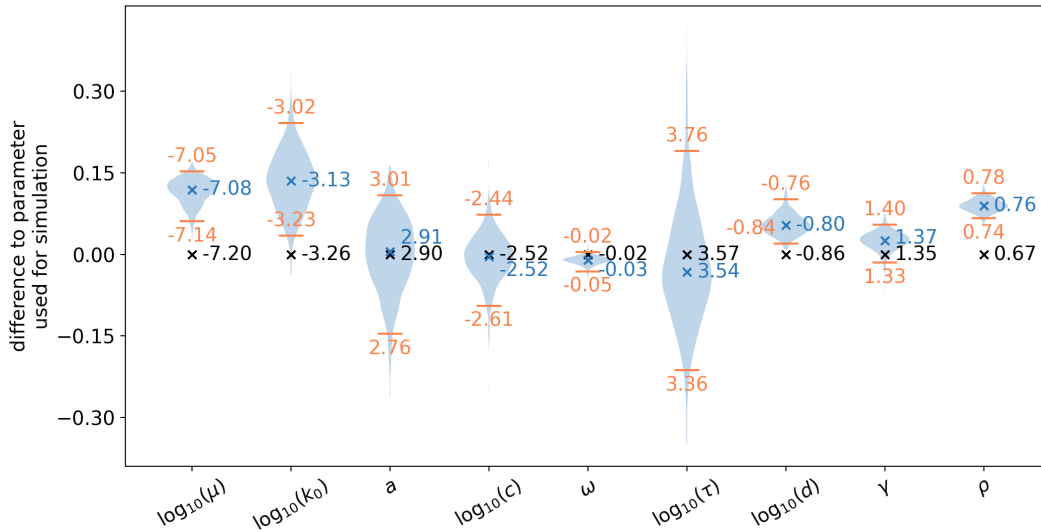


Figure 3.3: ETAS parameters used and inferred in synthetic test 1 (ST1). Black crosses indicate parameters used in simulation of 400 catalogs, blue crosses indicate median inverted parameters. Violins show the distribution of obtained parameters for 400 catalogs, with orange lines marking the 2.5% and 97.5% percentiles. Note that the y-axis gives the difference to parameters used for simulation, the actual values are written next to their marks.

us to impose the $m_c(t)$ history observed in California on the synthetic catalogs by discarding all events e_i for which $m_i < m_c(t_i)$.

We apply the ETAS inversion for time-varying m_c with the here-obtained $m_c(t)$ (see Equation (3.27)) to the synthetic catalogs.

3.4.2 Synthetic test for PETAI (ST2)

To test the PETAI inversion algorithm, 500 synthetic catalogs are created as follows. We use the parameters obtained after applying the PETAI inversion algorithm to the California data (1970 to 2019) with $m_{ref} = 2.5$. The value of m_{ref} is chosen to achieve a balance between the amount of data available for the inversion and the computational power required to process such an amount of data. For the reasons described in Section 3.4.1, we reduce the background rate μ and modify the parameters to obtain a corrected productivity exponent as described in Equations (3.30 - 3.31). The final parameters used for the simulation of the catalogs can be found in Figure 3.4(d) - (l).

Using these parameters, we simulate as described in Section B.1, 500 synthetic catalogs that resemble the Californian catalog, for the period between 1850 and 2020 in a square of $40^\circ \text{ lat} \times 40^\circ \text{ long}$. As in the previous case, because of missing long-term aftershocks in the beginning of the simulated catalogs, we discard the first 100 years of data and are left with catalogs from 1950 to 2020. For each of these catalogs and given the ETAS parameters used for simulation, we calculate the current event rate at the time of each event in the catalogs (Equation (3.16)). As the current event rate is to a large extent driven by aftershock rates of earlier events, we expect overestimation of detection probabilities, as well as overestimation of independence probabilities, during the beginning of the time period (Wang et al., 2010; Schoenberg et al., 2010; Nandan, Ouillon, and Sornette, 2019). For this reason, we allocate another 20 years of burn period, leaving us with catalogs starting in 1970.

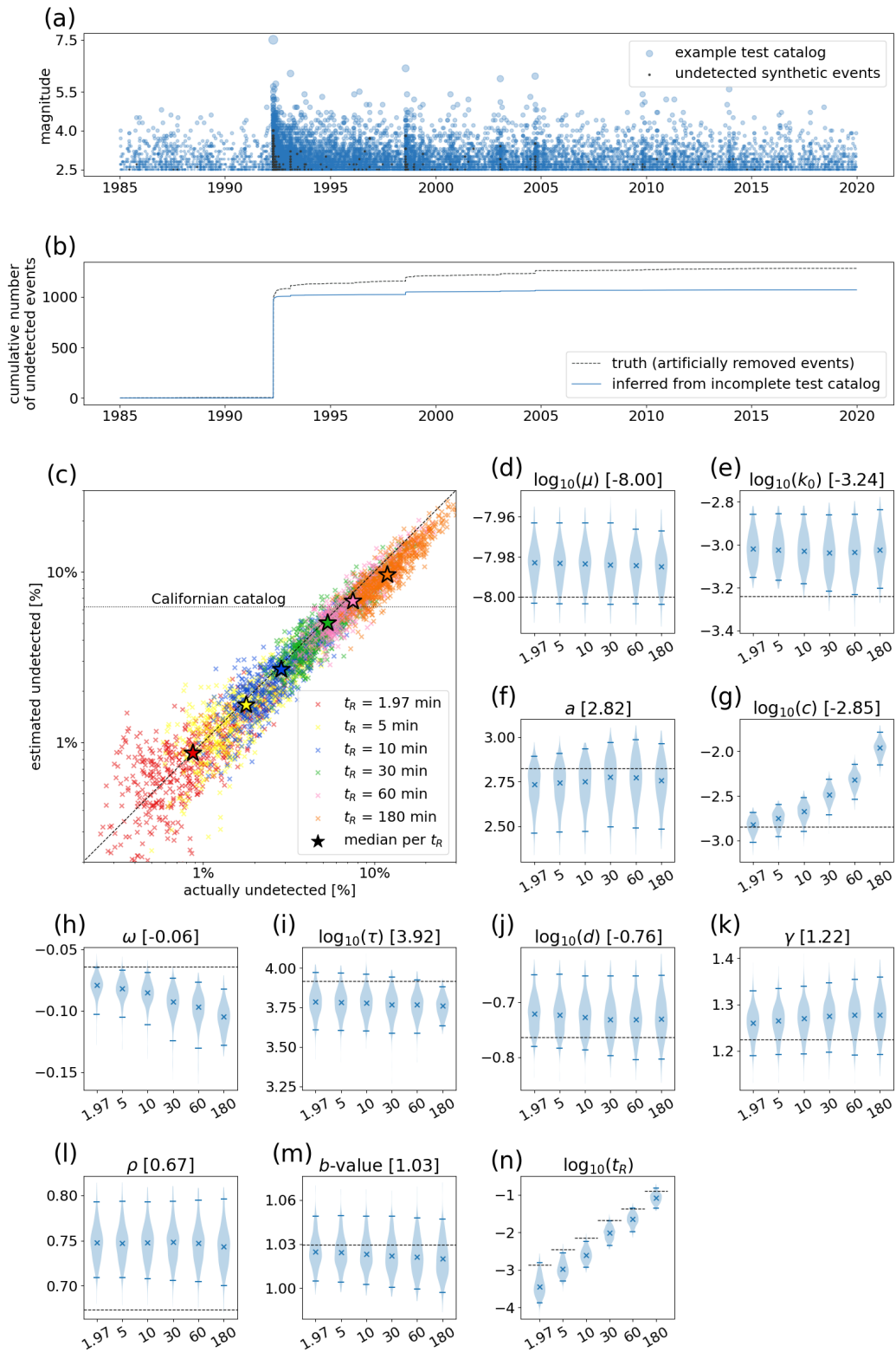


Figure 3.4: Results of synthetic test 2 (ST2). Caption on next page.

Figure 3.4: (Previous page.) Results of synthetic test 2 (ST2). (a) Earthquake magnitudes over time for one example test catalog (blue). Events marked in black were simulated, but declared as undetected. (b) Cumulative number of unobserved events over time for the catalog shown in (a). Black line marks the truth, blue line is inferred from the test catalog using PETAI. (c) Estimated fraction of undetected events versus actually removed fraction of events, for 3000 test catalogs. Different colors indicate different assumed detection efficiencies. Stars mark the median actual and estimated fraction of undetected events per t_R . Dashed line indicates where actual and estimated fraction coincide, dotted horizontal line indicates the estimated fraction for California. (d) - (n) ETAS and PETAI parameters inferred in synthetic test. Panel title indicates the parameter name and in square brackets the value used for simulation. Violins show the distribution of the parameter inferred from test catalogs per value of t_R used, with the value of t_R given on the x-axis in minutes. Crosses indicate median obtained value, blue lines indicate 95% confidence interval, dashed line indicates the value used for simulation.

Each of the 500 catalogs are then artificially made incomplete as follows. Using the detection probability function given by Equation (3.17), and the b -value of 1.03 estimated from the Californian catalog using PETAI inversion, we calculate for each event its probability of being detected. According to this probability we randomly decide for each event whether it has been detected or not. The subset of all events that were detected is then used as a test catalog. This is done assuming different values for t_R of 1.97 (as estimated from the Californian catalog), 5, 10, 30, 60, and 180 minutes, yielding six variations of the test catalog per originally simulated catalog, which makes a total of 3000 test catalogs. The value of t_R greatly influences the fraction of undetected events in the resulting catalog, and we chose to investigate different values of t_R to ensure there are sufficiently many test catalogs with a fraction of undetected events similar to the fraction of estimated undetected events inferred for California. This estimated number of undetected events is obtained by summing $\zeta(t_i)$, the expected number of unobserved events per observed event, which is estimated as a component of the PETAI inversion, over all occurrence times t_i of events in the primary catalog.

3.4.3 Results for ST1

Figure 3.3 shows the ETAS parameters used in the simulation of the synthetic catalogs, and the median, distribution, and 95% confidence intervals of the parameters inverted from the synthetic catalogs. The parameters estimated from the synthetic catalogs lie reasonably close to the data-generating parameters. In particular, a , c , ω , τ and γ are accurately inverted, while μ , k_0 , d and ρ tend to be overestimated. The reason for the overestimation of ρ stems from a computational simplification made during inversion. In order to avoid extremely large triggering probability matrices, we only consider pairs of source and target events with a spatial distance of less than 50 source lengths, where one source length is defined using the magnitude to length scaling relations defined in Wells and Coppersmith (1994). This upper limit for distances between event pairs translates to an exaggeration of the estimated values of ρ . We confirmed that as we gradually relax the cutoff criterion, the estimated value of ρ moves closer to the true values used for generating the synthetic catalog. The regularizer of the spatial kernel, d , is positively correlated with ρ , hence an overestimation of the latter translates to an overestimation of the former. The overestimation of μ can also be explained, considering that distant aftershocks have a higher tendency to appear independent due to the artificially imposed cutoff criterion.

3.4.4 Results for ST2

Inverted number of undetected events

Figure 3.4 (a) shows the series of events of one example synthetic test catalog over the primary time period in blue, with the undetected synthetic events marked in black. The number of undetected events is 1282, which makes up 6.25% of the original synthetic catalog. Figure 3.4 (b) shows cumulative number of undetected synthetic events over time in black, compared to the cumulative inferred number of undetected events in blue for the same example catalog. Overall, it is estimated as a result of applying the PETAI inversion that 1068.88 events were undetected in the example catalog. While this underestimates the true number of 1282 undetected events, the major part of events can be reconstructed, with accurate timing.

Figure 3.4 (c) shows inferred versus actual number of undetected events for 3000 test catalogs assuming different detection efficiencies. The estimated fraction of undetected events is distributed around the actual fraction of undetected events, and the median estimated fraction matches well the median actual fraction, with a slight tendency towards underestimation.

Accuracy of inverted parameters

Figure 3.4 (d) - (n) shows the ETAS parameters and (t_R, β) that were used in the simulation of the synthetic catalogs, and the parameters inverted from these synthetic catalogs. In general, the inverted parameters correspond well to the parameters used in the simulation, although some of the estimates are slightly biased. The parameters c and ω , both describing the temporal decay of aftershock rate, show a trend of increasing bias with increasing t_R , that is, with increasing incompleteness. For the other parameters, no clear dependency of the bias on t_R is recognizable. The estimate of c matches the true value almost perfectly for $t_R = 1.97$ minutes, but starts being overestimated for larger values of t_R above 30 minutes. On the other hand, ω shows an increasing tendency of being underestimated with increasing values of t_R . Earlier aftershocks have a larger tendency to be missing due to STAI, which leads to a seemingly slower decay of aftershock rate in time. As the PETAI algorithm has a tendency to underestimate STAI (Figure 3.4 (c)), and this tendency increases with increasing t_R , this translates into an increasing negative bias in the inferred values of ω .

Qualitatively, the tendencies to over- or underestimate the remaining parameters are identical with the tendencies observed in ST1 (Section 3.4.3). It is therefore plausible that these tendencies are consequences of a finite time horizon and finite spatial window used in the simulation of the synthetics, rather than being artifacts of the PETAI inversion algorithm.

Finally, we observe a tendency to underestimate t_R , which means that detection probabilities tend to be overestimated. This is in line with our previous observation that the fraction of undetected events tends to be slightly underestimated, suggesting the PETAI inversion to be slightly conservative.

3.5 Application to California

We calculate ETAS parameters, β and t_R (if applicable) using different inversion algorithms to Californian data. Additionally, we provide the resulting values for productivity exponent $\alpha = a - \rho\gamma$ and branching ratio η (see Equation (3.29)).

First, we apply usual inversion method as described in Section 3.3.1 with a constant completeness magnitude of $m_c \equiv 3.1$ to the main catalog (1970 to 2019). Then, we invert the

parameters by accounting for long-term time-variation of completeness (Equation (3.27)). In this case, the extended catalog from 1932 to 2019 can be used with a reference magnitude of $m_{ref} = 2.4$. Finally, we apply PETAI inversion to the main catalog (1970 to 2019) with a reference magnitude of $m_{ref} = 2.5$. Note that the estimation of β is independent of the ETAS parameter estimates for the first two applications, but not in the case of PETAI inversion (see Section 3.3.4).

To allow a better comparison between parameters inverted using different methods when m_{ref} varies, we translate the parameters to a reference magnitude of $m_{ref} = 3.1$ as follows. With the exception of μ , k_0 and d , all parameters are m_{ref} -agnostic, and the three exceptions can easily be adjusted. Denote by Δm the difference between new and original reference magnitude, $\Delta m = m'_{ref} - m_{ref}$. Then,

$$d' := d \cdot e^{\Delta m \cdot \gamma} \quad (3.32)$$

ensures that

$$d \cdot e^{\gamma \cdot (m - m_{ref})} = d' \cdot e^{\gamma \cdot (m - m'_{ref})}. \quad (3.33)$$

Stipulating that the branching ratio η (Equation (3.29)) remains unchanged, it follows that

$$k'_0 := k_0 \cdot e^{\Delta m \cdot \gamma \cdot \rho}. \quad (3.34)$$

The adaptation of the background rate μ follows trivially from the GR law,

$$\mu' = \mu \cdot e^{-\beta \cdot \Delta m}. \quad (3.35)$$

3.5.1 Interpretation of inverted parameters

Table 3.1: ETAS and PETAI parameters inferred for California. First column shows parameters when constant m_c of 3.1 is assumed. Second and third column show parameters when time-varying m_c is accounted for, and fourth and fifth column show parameters when PETAI inversion is applied. Note that the originally derived parameters are given in Columns 1, 2, and 4. Columns 3 and 5 show the parameters of Columns 2 and 4, transformed (as described in Equations (3.32 - 3.35)) to a reference magnitude of 3.1 to allow comparison with Column 1. Productivity exponent $\alpha = a - \rho\gamma$ and branching ratio η are not directly inverted but inferred from the inverted parameters.

parameter	$m_c \equiv \text{const.}$	$m_c(t)$		$f(m, t)$	
m_{ref}	3.1	2.4	3.1	2.5	3.1
$\log_{10}(\mu)$	-6.86	-5.97	-6.68	-6.35	-6.97
$\log_{10}(k_0)$	-2.53	-2.63	-2.36	-2.70	-2.49
a	1.74	1.86	1.86	1.92	1.92
$\log_{10}(c)$	-2.97	-2.52	-2.52	-2.85	-2.85
ω	-0.05	-0.02	-0.02	-0.06	-0.06
$\log_{10}(\tau)$	4.03	3.57	3.57	3.92	3.92
$\log_{10}(d)$	-0.51	-0.86	-0.45	-0.76	-0.45
γ	1.19	1.35	1.35	1.22	1.22
ρ	0.60	0.67	0.67	0.67	0.67
$\log_{10}(t_R)$	n/a	n/a	n/a	-2.86	-2.86
β	2.33	2.32	2.32	2.37	2.37
$a - \rho\gamma$	1.03	0.95	0.95	1.09	1.09
η	0.94	0.95	0.95	0.93	0.93

Table 3.1 shows the estimated values of ETAS parameters, β , and t_R (if applicable) obtained using different inversion algorithms to Californian data. Additionally, the resulting values for the productivity exponent $\alpha = a - \rho\gamma$ and branching ratio η (see Equation (3.29)) are provided. The first, second, and fourth column show the parameters obtained from applying the method with $m_c \equiv 3.1$, when using long-term-variations of m_c , and when using PETAI, respectively. Columns three and five contain the parameters of columns two and four after having been transformed to a reference magnitude of $m_{ref} = 3.1$.

Overall, the inverted parameters are roughly consistent among the three algorithms. Although there are slight differences between the estimated parameters, they can plausibly be attributed to different input datasets, which vary for the three algorithms in either time-span or magnitude range. In the following, we present some speculative explanations of the observed differences.

We find that the estimate of τ obtained from the ETAS model calibrated on the extended catalog (1932 to 2019) with the long-term time variation of m_c is smaller than in the other two cases, to an extent that the uncertainties obtained in the synthetic tests cannot explain this decrease. This decrease despite the use of a catalog spanning a longer duration compared to the other two cases, shows that τ may actually better reflect the long-term behavior of earthquake interaction, rather than being determined by the finite duration of the catalog. Note that if the temporal finiteness of the catalog was the dominant factor in the determination of τ , one would expect an increase of τ with increasing time spanned by the catalog. Furthermore, the less pronounced decrease of τ in case of the PETAI inversion speaks against the possibility that the decrease is caused by inclusion of lower magnitude earthquakes revealing previously unseen earthquake interactions.

A somewhat counter-intuitive observation is the increase of c for both new inversion techniques. For the case of long-term variation of m_c , in particular, c shows a significant increase considering the expected uncertainties. The parameter c has been interpreted to reflect aftershock incompleteness (Kagan, 2004; Lolli and Gasperini, 2006; Hainzl, 2016a) and would thus be expected to decrease when this effect is accounted for by the model (Seif et al., 2017). The observed higher value of c even after accounting for STAI thus requires a different interpretation of c . Narteau et al. (2009) found a dependency of c on faulting style, and brought the parameter in relation with differential stress and the intensity of stress re-distribution. Another possible interpretation provided by Lippiello et al. (2007) is based on the dynamical scaling hypothesis in which time differences relate to magnitude differences. Shcherbakov et al. (2004) proposed a generalized Omori law which incorporates three empirical scaling laws (Gutenberg and Richter, 1944, Båth, 1965, Utsu, 1961) with a dependence of c on the cutoff magnitude which can qualitatively explain our observations: The value inverted for c is highest in the case of $m_{ref} = 2.4$, and lowest for $m_{ref} = 3.1$. Overall, one should be careful to not over-interpret this estimate of c . After all, c is overestimated for large values of t_R in the PETAI synthetic test and hence an observed increase in c might be a consequence of complex interdependencies of all parameters involved.

While the branching ratio η does not substantially vary with the different inversion methods, we observe a slightly increased productivity exponent for the PETAI inversion. Although the increase lies within expected uncertainty, such an increase is expected given the results of Seif et al. (2017), with the extent of the observed increase being in line with their estimated extent of underestimation for the productivity exponent.

The background rate μ shows a significant increase when a longer time horizon is considered, and decreases significantly when STAI is accounted for. As μ is clearly overestimated in the synthetic test with long-term variation of m_c , and only slightly overestimated in the

case of PETAI, we may suspect that the increased value for μ in the first case is an artifact of the inversion method, while the decrease in background rate with PETAI could suggest that including smaller magnitude events in our model by accounting for incompleteness reveals previously hidden earthquake interactions, resulting in a lower μ .

The parameter γ , which describes the exponential relationship between earthquake magnitude and the distance to the event at which the aftershock rate starts to decrease faster, is significantly increased in the case of long-term variation of m_c . Slight overestimation is expected based on the synthetic tests, but not to this extent.

At the same time, ρ increases for both new inversion techniques. Again, overestimation of ρ is expected given the results of the synthetic tests and the observed values might thus be artifacts of the algorithms applied. As the problem of the finite spatial region applies in the same way to standard ETAS as well as the other two methods, this is unlikely to be the cause of the difference in parameter estimates.

The value of β shows an increase from 2.33 to 2.37, which translates to a b -value increase from 1.01 to 1.03, when STAI is accounted for in the PETAI inversion. This is expected due to the underestimated number of small events caused by STAI.

3.5.2 Incompleteness insights through PETAI

In addition to a new set of estimated ETAS parameters, applying the PETAI inversion to the Californian catalog produces further interesting outputs. Similarly to the case of the synthetic catalog, Figure 3.5 (a) shows the estimated cumulative number of undetected events over time. As expected, the increase is predominantly step-wise, caused by short, incomplete periods during aftershock sequences, and long, complete periods in-between. While the total expected number of undetected events is at 5041.74, the extrapolated number obtained from a GR law fitted on $M \geq 3.1$ events is only 88.91. This estimate of the number of unobserved events differs from the PETAI estimate in that it assumes perfect detection above $M3.1$. Although the true number of undetected events can never be known, the synthetic test suggests that the PETAI result is reliable and even slightly conservative, and thus the GR law extrapolation would be a severe underestimation of the true number of undetected events.

The magnitude-dependent detection probability evolution is illustrated in Figure 3.5 (b). In around 84% of event times t_i , events of magnitude $M \geq 4$ are expected to be detected with a probability of 99.9% or more. Similarly, in 82% of event times t_i , $M \geq 3$ events are expected to be detected with a probability of 99% or more. Spikes of incompleteness during large sequences lead to detection probabilities of less than 50% for smaller events, in the most extreme case for events of magnitude $M \leq 3.47$.

As expected, periods of elevated incompleteness coincide with the periods of rapid increase in undetected events shown in (a). The last step in (a), which corresponds to the 2019 Ridgecrest sequence, is extraordinarily large compared to all previous steps. This is most likely explained by the fact that the sequence was better recorded than comparable sequences in previous years. When the detection capability of the seismic network improves, the recovery time t_R becomes shorter. Because we have assumed t_R to be stationary for simplicity, a larger number of recorded events will lead to a smaller estimated detection probability, which in turn leads to larger numbers of expected undetected events. In future versions of the model, to avoid such artifacts, it would be advisable to combine the possibility of including long-term changes in completeness (as in the model described in Section 3.3.1) with rate-dependent aftershock incompleteness by means of a non-stationary t_R .

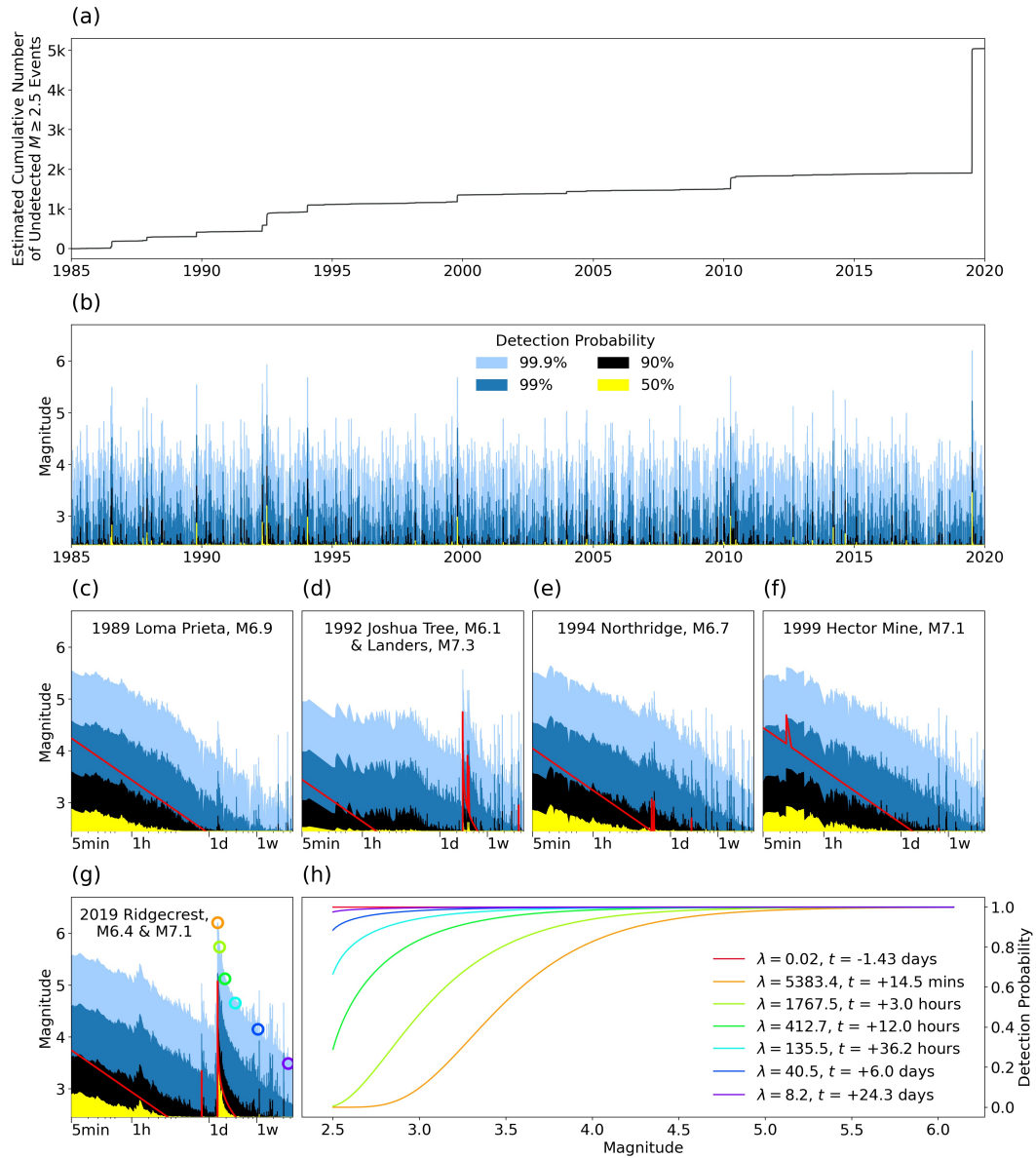


Figure 3.5: Aftershock Incompleteness in California. (a) Estimated cumulative number of undetected events over time. (b) Evolution of magnitude-dependent detection probability. Yellow indicates a detection probability of 50% or less. Black, dark blue, and light blue indicate detection probabilities of up to 90%, 99%, and 99.9%, respectively. White area represents detection probabilities higher than 99.9%. (c)-(g) Excerpts of (b) for selected large events. x -axes are logarithmic and show time since (first) mainshock, and range from 5 minutes to 30 days after that mainshock. Red lines indicate $m_c(t)$ as described by Helmstetter et al. (2006), including the effect of all $M \geq 5$ events. Colored circles in (g) represent selected times t_i and corresponding magnitude of 99.9% detection. (h) Detection probability function $f(m, \lambda = \lambda(t_i))$ snapshots for the times that are highlighted in (g), plus a time prior to both mainshocks (in red). Time deltas are given with respect to the M7.1 mainshock. $\lambda(t_i)$ are as estimated during PETAI inversion.

Figure 3.5 (c) - (g) shows excerpts of Figure 3.5 (b) for the 1989 M6.9 Loma Prieta, the 1992 M6.1 Joshua Tree and 7.3 Landers, the 1994 M6.7 Northridge, the 1999 M7.1 Hector Mine, and the 2019 M6.4 and 7.1 Ridgecrest events, in comparison to the $m_c(t)$ estimate given by the formulation of Helmstetter et al. (2006) which was provided for Southern California. While their definition is not probabilistic, we observe that their m_c 5 minutes after the mainshock lies between 90% and 99% detection according to PETAI. The shape of the recovery from incompleteness does not fully coincide for the two methods, with generally slower recovery in the case of PETAI for the shown excerpts. Helmstetter et al. (2006) use a simpler formulation, and do not provide arguments for their specific choice of parameterization of $m_c(t)$. On the other hand, the parametric description of the magnitude of 50% detection by Ogata and Katsura (2006), presented for the example of the 2003 Miyagi-Ken-Oki earthquake, takes a shape similar to the one obtained through PETAI, although it is not quantitatively comparable to the case of California.

The range of observed states of detection efficiency during the 2019 Ridgecrest sequence is visualized in Figure 3.5 (h). Prior to the large events, detection is almost perfect for all magnitudes. After the M6.4 event, detection is weakened and recovers with time, until the M7.1 mainshock, when it is again weakened. Around 15 minutes after the earthquake, events of magnitude below 3.0 still have almost no chance to be detected, with M3.5 events having roughly a 50% chance to be detected. After three hours, detection has already clearly improved, although M2.5 events are still almost surely not detected. After six days, the detection probability function almost corresponds to the perfect detection state, which was in place prior to the main events.

3.5.3 Comments on computational time

There are two aspects to consider when discussing the computational time of the parameter inversion techniques presented here. On one hand, the increased complexity of the algorithms plays an important role. In particular, the PETAI inversion comprises multiple loops of ETAS and incompleteness estimation. Although convergence was usually reached after 4 iterations, this still implies a minimum factor of 4 in terms of computation time which is only required for ETAS inversion, on top of which comes the time needed for the estimation of detection parameters and event rates. The second factor, which contributes even more to an increase of computation time, is the increased size of the catalog which is available to be used. For our application to Californian data, the number of events used in the PETAI inversion increases by a factor of 3.78 because the minimum considered magnitude is reduced from 3.1 to 2.5. This leads the number of pairs of potentially related events to increase from 7.3 million to 47.1 million. Note that these numbers are obtained after imposing the 50 source length cutoff criterion described in 3.4.3. While this increase in the number of potentially related event pairs causes a substantial increase in run time, educated initial guesses for ETAS parameter inversions can substantially reduce run time without affecting the results. Our Python 3.8 implementation of the PETAI inversion, run with a single core (Intel Xeon E5-2697v2) of the Euler high-performance computing (HPC) cluster at ETH Zurich, took 23 hours. Roughly 20% of this time was spent on the optimization of event rates and detection parameters, and 80% on the optimization of ETAS parameters.

In contrast to the PETAI inversion, the run time of the ETAS parameter inversion with time-varying m_c is barely affected by model complexity. During synthetic experiments, we found the run time to be comparable to the run time of the usual ETAS inversion when the number pairs of potentially related events was similar.

3.6 Pseudo-prospective forecasting experiments

To better understand if and how the PETAI model can improve earthquake forecasts, we conduct pseudo-prospective forecasting experiments. Note that as these experiments are computationally expensive, we conduct them for the PETAI method only. As most aftershocks occur soon after their triggering event, accounting for STAI in ETAS simulations seems promising for forecasting. The parameter inversion for long-term variations of m_c is mainly intended as a tool to obtain ETAS parameters in regions where data is sparse and a model inversion would not be possible otherwise.

3.6.1 Competing models

We compare five models.

1. The base ETAS model assumes perfect detection above a constant $m_c = 3.1$ and is used as the null model.
2. PETAI, the alternative model, has two modifications to the null model. First, it uses improved ETAS parameter estimates that were obtained in the PETAI inversion with a reference magnitude m_{ref} of 2.5. Second, magnitude $M \geq 2.5$ earthquakes are allowed to trigger and be triggered. For this, the events in the training catalog, which act as triggering earthquakes in the simulation, have their triggering capability inflated by $1 + \xi(t)$, as estimated in the PETAI inversion.

Two intermediate models are assessed to dissect the effect of the two modifications.

3. `par_only` uses ETAS parameter estimates obtained from PETAI, but only $M \geq 3.1$ events are allowed to trigger and be triggered, assuming perfect detection there (i.e. $\xi(t) \equiv 0$). In this case, the parameters obtained for the PETAI model have to be transformed to be compatible with a reference magnitude of $m_{ref} = 3.1$ as described in Equations (3.32 - 3.35).
4. Vice-versa, `trig_only` allows $M \geq 2.5$ events to trigger and be triggered, using the inverted $\xi(t)$ for inflated triggering, but does not use the improved ETAS parameter estimates. In this case, the parameters obtained for the null model have to be transformed to be compatible with a reference magnitude of $m_{ref} = 2.5$ as described in Equations (3.32 - 3.35).

Lastly, we assess an additional benchmark model to test whether deliberately underestimating m_c is an appropriate alternative to the rather complex PETAI model.

5. `low_mc` assumes perfect detection above a constant $m_c = 2.5$. This model uses neither the parameter estimates obtained from PETAI, nor the inverted $\xi(t)$ for inflated triggering, but it allows $M \geq 2.5$ events to trigger and be triggered and thus is based on the same data as the PETAI-based models.

3.6.2 Experiment setup

For a testing period length of 30 days, we define a family of training and testing periods such that the testing periods are consecutive and non-overlapping. Each training period ends with the starting date of its corresponding testing period. The starting date of the first testing period is January 1st, 2000. The end date of the last of the 244 testing periods is January 16th, 2020.

For each testing period, all competing models are trained based on the corresponding training data. Then, forecasts are issued with each model through simulation of 100,000 possible continuations of the training catalog. Because the testing data is ignored when the models are calibrated, these forecasts are pseudo-prospective. This is done by simulating Type I earthquakes (the cascade of aftershocks of earthquakes in the training catalog) and Type II earthquakes (simulated background earthquakes and their cascade of aftershocks) similarly to how it is described by Nandan, Ouillon, Sornette, and Wiemer (2019a). The algorithm for simulation is described in detail in Section B.1.

The performance of each model is evaluated by calculating the log-likelihood of the testing data given the forecast. See Section B.2 for details on the calculation of the log-likelihood using the full distribution approach as described by Nandan, Ouillon, Sornette, and Wiemer (2019a) for a fair evaluation of ETAS-based models. Two competing models can be compared by calculating the information gain (IG) of the alternative model M_{alt} over the null model M_0 , which is simply the difference in log-likelihood of observing the testing data. The mean information gain (MIG) is calculated as the mean over all testing periods. This evaluation metric is similar to other metrics that have been used for model comparison, such as the total information gain or information gain per earthquake (IGPE) used in the CSEP T-test (Harte and Vere-Jones, 2005; D. A. Rhoades et al., 2011; Zechar et al., 2013; Strader et al., 2017, see Savran et al., 2020 for recent complementary CSEP testing metrics) or the residual-based log-likelihood ratio score (Clements et al., 2011; Bray et al., 2014; Gordon et al., 2015; Gordon et al., 2021).

As an additional benchmark, we calculate the total IGPE of the ETAS null model versus a spatially and temporally homogeneous Poisson process (STHPP) model. Note that the STHPP model is not considered a participant of the forecasting experiment and superiority is always discussed relative to the ETAS null model.

For details on the STHPP model and on the conditions under which one model is considered superior over another, see Section B.2 and Nandan, Ouillon, Sornette, and Wiemer (2019a).

3.6.3 Time evolution of the parameters of the competing models

Figure 3.6 shows the parameter evolution with increasing training period obtained with standard ETAS ($m_c = 2.5$ and $m_c = 3.1$) and PETAI inversion. Two parameters, namely μ and τ , show a systematic decrease and increase, respectively, with growing time horizon of the training catalog. When compared to the uncertainties in the synthetic tests, the extents of the changes of μ are larger than the 95% confidence intervals, while the changes of τ lie within the expected uncertainties. A possible explanation for this observation is that an increased time horizon of the training catalog reveals more long-term earthquake interactions, leading to a higher value of τ , that is a later onset of the exponential taper in the temporal aftershock density, and simultaneously to a lower background rate μ , as more events can be interpreted as aftershocks of previous earthquakes.

Nearly all parameter estimates show a jump in 2010, caused by the 2010 El-Mayor Cupacah earthquake sequence, and a second jump in 2019, caused by the 2019 Ridgecrest sequence. There are several reasons why such jumps in parameter estimates could occur. In the case of the 2010 events, the main earthquake occurred outside of California and thus network coverage can play a role, as well as the absence of a large fraction of aftershocks due to the boundaries of the considered region. Furthermore, triggering parameters can differ between regions, sequences and can also depend on the magnitude of the mainshocks (Nandan, Ouillon, and Sornette, 2019; Nandan, Kamer, et al., 2021; Ouillon and Sornette, 2005;

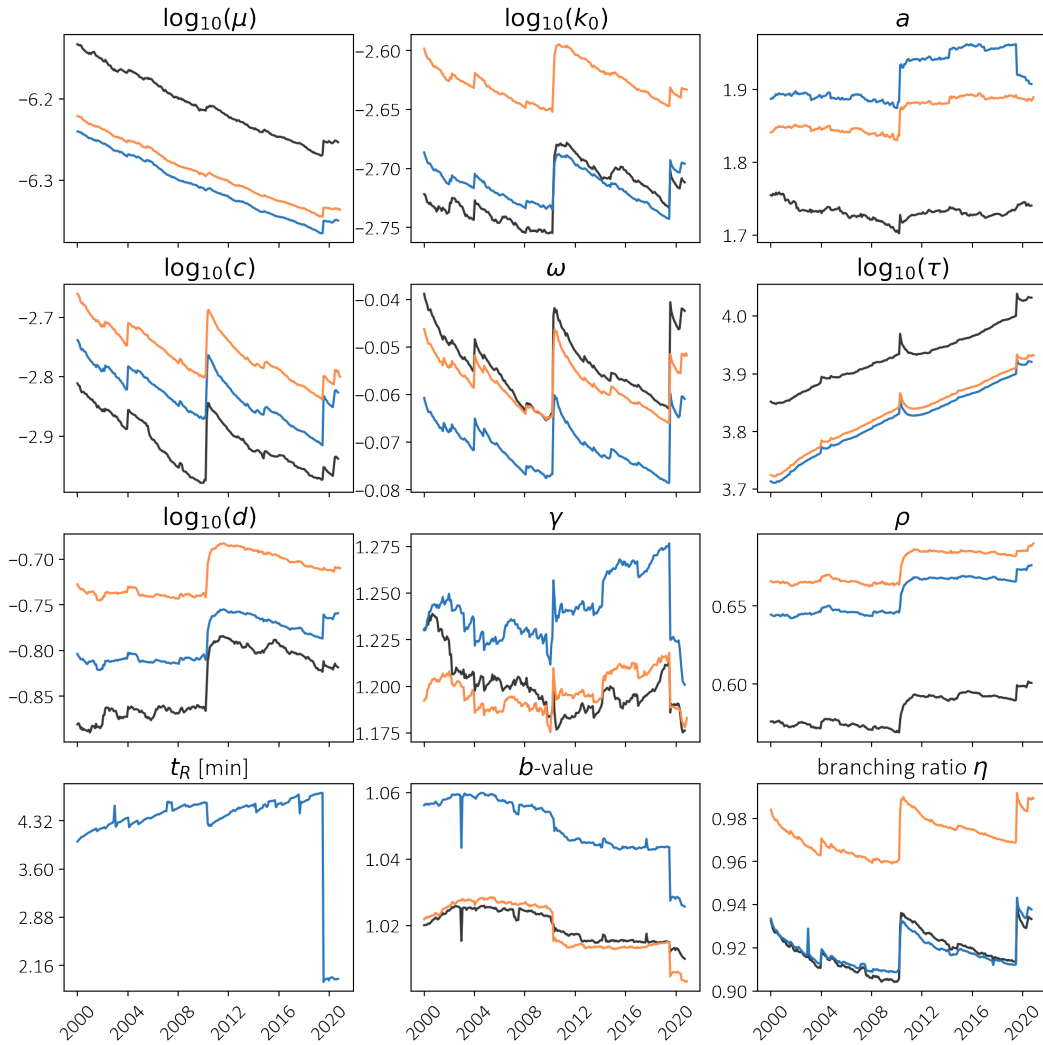


Figure 3.6: Evolution of ETAS and PETAI parameter estimates with increasing training catalog, when using standard inversion with $m_c = 3.1$ (black lines) or $m_c = 2.5$ (orange lines) and when using PETAI inversion (blue lines). The evolution for t_R is only given for PETAI inversion because it does not exist in standard ETAS. Parameters are with respect to $m_{ref} = 2.5$, transformed using Equations (3.32 - 3.35) if necessary.

Sornette and Ouillon, 2005; Nandan, Ouillon, and Sornette, 2021). These dependencies can increase the representation of the active region and particular sequences in the catalog and lead to sudden changes in the overall parameters.

3.6.4 Forecasting performance of the competing models

Figure 3.7 shows the results of the pseudo-prospective forecasting experiments. For a target magnitude threshold of $m_t = 3.1$, PETAI as well as trig_only significantly outperform the ETAS null model with p-values of virtually 0 and a mean information gain of 0.97 and 0.94, respectively. Note that this improvement is over a very strong null model, which has a total information gain of 49'246 (i.e. a MIG of 202.66 or an IGPE of 5.62) over the STHPP model. PETAI has a slightly positive but not statistically significant information gain compared to trig_only. On the other hand, par_only and low_mc do not significantly outperform the ETAS null model. This suggests that the main driver of the improvement of the forecast is the inclusion of small events between $M2.5$ and $M3.1$ in the simulations, rather than the newly obtained parameter estimates. It also indicates that accounting for incompleteness, which is possible due to the estimated $\xi(t_i)$ obtained in the PETAI inversion, is necessary for this improved forecast. The sole inclusion of events between $M2.5$ and $M3.1$ in the simulations assuming completeness above $M2.5$ is not sufficient to obtain significant improvements. For all considered values of m_t , PETAI and trig_only rank higher than low_mc in terms of MIG, which further supports the idea that accounting for STAI is relevant for improved ETAS-based earthquake forecasting.

The temporal evolution of the cumulative information gain of the two superior models shows a decrease during the 2010 El Mayor-Cucapah and the 2019 Ridgecrest sequences. Those sequences were most active in Southern California, where the seismic network is much denser than in the rest of the considered region (Hutton et al., 2010; Schorlemmer and Woessner, 2008). The assumption of spatially homogeneous detection incompleteness is thus inaccurate and may be the reason for over-inflation of the aftershock productivity during these sequences, explaining the decrease in information gain. One can therefore expect that accounting for spatial variation of STAI in subsequent models may lead to even better forecasts.

With increasing values of m_t to 3.5, 4.0, 4.5, and 5.0, the IGPE of the ETAS null model versus the STHPP model increases to 5.92, 5.92, 6.62, and 7.44, respectively. At the same time, the mean information gain values between the competing ETAS-based models generally decrease, and almost no model significantly outperforms any other competing model. Occasionally, par_only is outperformed by the ETAS null model or by trig_only. These observations suggest that taking into account information about smaller earthquakes mainly helps improving ETAS-based forecasts of smaller earthquakes. More precisely, simulating aftershocks of small earthquakes is the key ingredient for improved forecasting of similarly-sized events. Although within the framework of the standard ETAS model, small earthquakes can trigger large ones, and their relative abundance implies significant contribution to the overall triggering (Marsan, 2005; Helmstetter et al., 2005; Sornette and Werner, 2005a), we find that the beneficial effect vanishes when forecasting large events. Additional ways exist in which small earthquakes can contribute to improving forecasting models. Besides their potential to cumulatively contribute to aftershock triggering, the large number of earthquakes below m_c can help to highlight the underlying fault structure, which, when accounted for, can significantly improve forecasting performance (Gordon et al., 2021; Bach and Hainzl, 2012; Cattania et al., 2018; Y. Guo et al., 2015). In fact, small earthquakes have been shown to improve forecasts in the context of other models (Mancini et al., 2019; Mancini et al., 2020), and somewhat mixed overall results but a clear signal that small earthquakes do contribute to triggering through the redistribution of static stresses have been reported (Meier et al., 2014; Segou et al., 2013; Nandan et al., 2016).

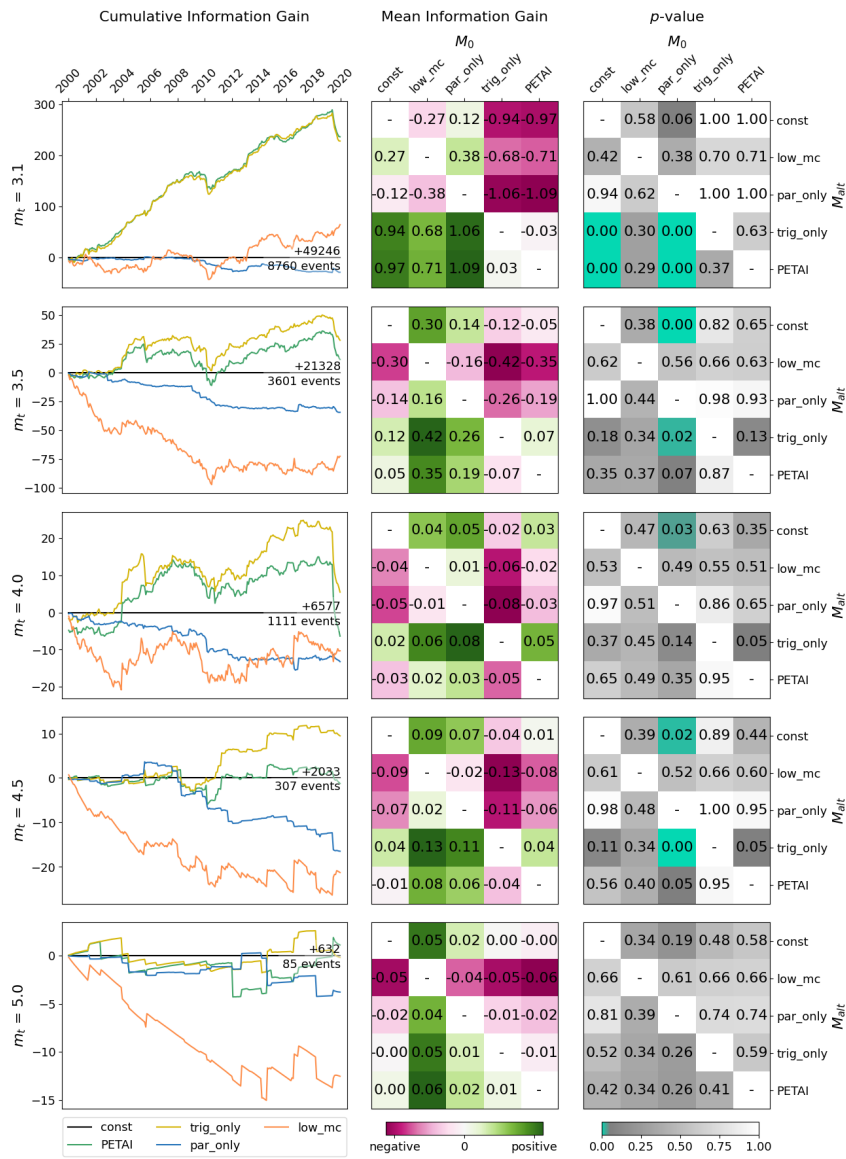


Figure 3.7: From left to right: Cumulative information gain for the alternative models versus the ETAS null model, mean information gain matrix, and corresponding p -value matrix comparing all competing models. Matrix entries represent the test of superiority of M_{alt} (y -axis) versus M_0 (x -axis). From top to bottom: target magnitude thresholds m_t of 3.1, 3.5, 4.0, 4.5, and 5.0. Indicated as text in the left panels is the cumulative information gain of the ETAS null model versus the STHPP model, and the number of events in all testing periods combined. Note the different y -axes for the left panels. Also note that the color scheme for the middle panels is different between threshold magnitudes m_t and normalized with respect to the maximum absolute mean information gain for that m_t . Color coding for the panels on the right is such that p -values of 0.05 and below are green, and transition from grey to white between 0.05 and 1.

Helmstetter et al. (2006) compared the probability gain of their time-dependent model versus their similar but time-independent model and found that probability gain decreases with an increasing target magnitude threshold. They speculated that this observation may be due to a smaller sample size when the target magnitude threshold increases. Helmstetter and Werner (2014) found the same decrease in the context of a different, non-parametric kernel space-time smoothing model. Although the study was based on a larger amount of data than Helmstetter et al. (2006), they likewise speculated that this decrease is due to a small sample size. In our case, the same effect is observed at considerably large sample sizes of 3601, 1111, 307, and 85 events for $m_t = 3.5, 4.0, 4.5,$ and 5.0 .

Another possible explanation for this effect is provided by the findings of multiple previous studies using both non-parametric (Nichols and Schoenberg, 2014; Spassiani and Sebastiani, 2016) and parametric (Nandan, Ouillon, and Sornette, 2019; Nandan, Ouillon, and Sornette, 2021) approaches, that earthquakes tend to preferentially trigger aftershocks of similar size. Their results can explain the improved forecast of small events when small events are used for simulation, as well as the vanishing of this improvement when the magnitude difference between newly included events and target events becomes large. This could furthermore serve as an alternative explanation of the results of Helmstetter et al. (2006) and Helmstetter and Werner (2014).

Note that the IGPE of the ETAS null model against the STHPP model increases with increasing m_t , while Helmstetter et al. (2006) observe a decrease in probability gain against a temporally homogeneous and spatially variable Poisson model. This suggests that spatial inhomogeneity becomes more important with increasing target magnitude, while temporal inhomogeneity based on small earthquakes becomes less important with increasing target magnitude. It is important to highlight that the 30-day testing periods of the present study, in contrast to 1-day periods as in Helmstetter et al. (2006), prevent ETAS models from being updated after large events. This likely understates the extent of superiority that could be achieved by models which include $M2.5$ to $M3.1$ events, and thus the role of small earthquakes, in daily forecasts.

3.7 Conclusion

We propose a modified algorithm for the inversion of ETAS parameters when m_c varies with time, and an algorithm for the joint inversion of ETAS parameters and probabilistic, epidemic-type aftershock incompleteness. We test both methods on synthetic catalogs, concluding that they can accurately invert the parameters used for simulation of the synthetics. The given formulations are rather general and can equally be applied to spatial or spatio-temporal variations of m_c , as well as to any suitable definition of a detection probability function.

Two potential use cases are the estimation of ETAS parameters based on the Californian catalog since 1932 with long-term fluctuations of m_c between 4.3 and 2.4, and the estimation of ETAS parameters and short-term aftershock incompleteness based on the incomplete Californian catalog of events above $M2.5$. The latter is further used to test the forecasting power of small earthquakes. Results of numerous pseudo-prospective forecasting experiments suggest that

- Information about small earthquakes significantly and substantially improves forecasts of similar-sized events.
- Main driver of this improvement is the simulation of aftershocks of small events.
- Accounting for incompleteness when simulating aftershocks of small events is necessary to achieve this improvement.

- Information about small earthquakes does not significantly affect the performance of large event forecasts.

A possible explanation for these results is provided by previous findings (Nichols and Schoenberg, 2014; Spassiani and Sebastiani, 2016; Nandan, Ouillon, and Sornette, 2019; Nandan, Ouillon, and Sornette, 2021), that earthquakes preferentially trigger aftershocks of similar size.

Our results have potentially significant implications for the future of earthquake forecasting. Thanks to the here-presented algorithms, ETAS models may be calibrated for regions with low seismicity where the usual inversion algorithms would fail due to missing data. To facilitate the embracing of data incompleteness in such cases, our inversion codes will be made openly available after publication of the article through github.com/lmizrahi/etas and github.com/lmizrahi/petai.

The newly gained insights from forecasting experiments guide us in the search of the next generation earthquake forecasting models. Besides other discussed topics such as anisotropy, temporally or spatially non-stationary background rate (Hainzl et al., 2008; Hainzl et al., 2013; Nandan et al., 2020), the importance of accounting for short-term incompleteness when simulating, as well as a magnitude-dependent distribution of aftershock magnitudes are emphasized.

Acknowledgments

The data used for this analysis is available through the website earthquake.usgs.gov/earthquakes/search/ (U.S. Geological Survey, Earthquake Hazards Program, 2017). The authors wish to thank Sebastian Hainzl, Andrew Michael and Andrea Llenos for insightful discussions and helpful feedback on earlier versions of this article. We also want to thank the editor Rachel Abercrombie, the anonymous associate editor, one anonymous reviewer, and Max Werner for their constructive feedback which greatly improved this article. This work has received funding from the Eidgenössische Technische Hochschule (ETH) research grant for project number 2018-FE-213, “Enabling dynamic earthquake risk assessment (DynaRisk)” and from the European Union’s Horizon 2020 research and innovation program under Grant Agreement Number 821115, real-time earthquake risk reduction for a resilient Europe (RISE).

References

- Amato, A., & Mele, F. (2008). Performance of the ingv national seismic network from 1997 to 2007. *Annals of Geophysics*, 33
- Amorese, D. (2007). Applying a change-point detection method on frequency-magnitude distributions. *Bulletin of the Seismological Society of America*, 97(5), 1742–1749. 33
- Bach, C., & Hainzl, S. (2012). Improving empirical aftershock modeling based on additional source information. *Journal of Geophysical Research: Solid Earth*, 117(B4). 57
- Båth, M. (1965). Lateral inhomogeneities of the upper mantle. *Tectonophysics*, 2(6), 483–514. 50
- Bray, A., Wong, K., Barr, C. D., & Schoenberg, F. P. (2014). Voronoi residual analysis of spatial point process models with applications to california earthquake forecasts. *The Annals of Applied Statistics*, 8(4), 2247–2267. 55
- Cao, A., & Gao, S. S. (2002). Temporal variation of seismic b-values beneath northeastern japan island arc. *Geophysical research letters*, 29(9), 48–1. 33

- Cattania, C., Werner, M. J., Marzocchi, W., Hainzl, S., Rhoades, D., Gerstenberger, M., Liukis, M., Savran, W., Christophersen, A., Helmstetter, A., et al. (2018). The forecasting skill of physics-based seismicity models during the 2010–2012 canterbury, new zealand, earthquake sequence. *Seismological Research Letters*, 89(4), 1238–1250. [33](#), [57](#)
- Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, 51(4), 661–703. [35](#)
- Clements, R. A., Schoenberg, F. P., & Schorlemmer, D. (2011). Residual analysis methods for space-time point processes with applications to earthquake forecast models in california. *The Annals of applied statistics*, 2549–2571. [55](#)
- De Arcangelis, L., Godano, C., & Lippiello, E. (2018). The overlap of aftershock coda waves and short-term postseismic forecasting. *Journal of Geophysical Research: Solid Earth*, 123(7), 5661–5674. [39](#)
- Felzer, K. R. (2007). Appendix i: Calculating california seismicity rates. *US Geol. Surv. Open-File Rept. 2007-1437I*. [35](#), [44](#)
- Field, E. H., Jordan, T. H., Page, M. T., Milner, K. R., Shaw, B. E., Dawson, T. E., Biasi, G. P., Parsons, T., Hardebeck, J. L., Michael, A. J., et al. (2017). A synoptic view of the third uniform california earthquake rupture forecast (ucerf3). *Seismological Research Letters*, 88(5), 1259–1267. [33](#)
- Gerstenberger, M. C., Marzocchi, W., Allen, T., Pagani, M., Adams, J., Danciu, L., Field, E. H., Fujiwara, H., Luco, N., Ma, K.-F., et al. (2020). Probabilistic seismic hazard analysis at regional and national scales: State of the art and future challenges. *Reviews of Geophysics*, 58(2), e2019RG000653. [33](#)
- Gordon, J. S., Clements, R. A., Schoenberg, F. P., & Schorlemmer, D. (2015). Voronoi residuals and other residual analyses applied to csep earthquake forecasts. *Spatial Statistics*, 14, 133–150. [55](#)
- Gordon, J. S., Fox, E. W., & Schoenberg, F. P. (2021). A nonparametric hawkes model for forecasting california seismicity. *Bulletin of the Seismological Society of America*, 111(4), 2216–2234. [55](#), [57](#)
- Guo, Y., Zhuang, J., & Zhou, S. (2015). An improved space-time etas model for inverting the rupture geometry from seismicity triggering. *Journal of Geophysical Research: Solid Earth*, 120(5), 3309–3323. [57](#)
- Guo, Z., & Ogata, Y. (1997). Statistical relations between the parameters of aftershocks in time, space, and magnitude. *Journal of Geophysical Research: Solid Earth*, 102(B2), 2857–2873. [44](#)
- Gutenberg, B., & Richter, C. F. (1944). Frequency of earthquakes in california. *Bulletin of the Seismological Society of America*, 34(4), 185–188. [35](#), [39](#), [50](#)
- 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. [34](#), [50](#)
- Hainzl, S. (2016b). Rate-dependent incompleteness of earthquake catalogs. *Seismological Research Letters*, 87(2A), 337–344. [33](#), [34](#)
- Hainzl, S., Christophersen, A., & 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. [44](#), [60](#)
- Hainzl, S., Zakharova, O., & 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. [44](#), [60](#)
- Harte, D., & Vere-Jones, D. (2005). The entropy score and its uses in earthquake forecasting. *Pure and Applied Geophysics*, 162(6), 1229–1253. [55](#)
- Helmstetter, A. (2003). Is earthquake triggering driven by small earthquakes? *Physical review letters*, 91(5), 058501. [37](#), [44](#)

- Helmstetter, A., Kagan, Y. Y., & Jackson, D. D. (2006). Comparison of short-term and time-independent earthquake forecast models for southern california. *Bulletin of the Seismological Society of America*, *96*(1), 90–106. [52](#), [53](#), [59](#)
- Helmstetter, A., Kagan, Y. Y., & Jackson, D. D. (2005). Importance of small earthquakes for stress transfers and earthquake triggering. *Journal of Geophysical Research: Solid Earth*, *110*(B5). [33](#), [57](#)
- Helmstetter, A., & Werner, M. J. (2014). Adaptive smoothing of seismicity in time, space, and magnitude for time-dependent earthquake forecasts for california. *Bulletin of the Seismological Society of America*, *104*(2), 809–822. [59](#)
- Hutton, K., Woessner, J., & Hauksson, E. (2010). Earthquake monitoring in southern california for seventy-seven years (1932–2008). *Bulletin of the Seismological Society of America*, *100*(2), 423–446. [33](#), [57](#)
- Kagan, Y. Y. (2004). Short-term properties of earthquake catalogs and models of earthquake source. *Bulletin of the Seismological Society of America*, *94*(4), 1207–1228. [33](#), [50](#)
- Lippiello, E., Bottiglieri, M., Godano, C., & de Arcangelis, L. (2007). Dynamical scaling and generalized omori law. *Geophysical Research Letters*, *34*(23). [50](#)
- Llenos, A. L., & Michael, A. J. (2020). Regionally optimized background earthquake rates from etas (robere) for probabilistic seismic hazard assessment. *Bulletin of the Seismological Society of America*, *110*(3), 1172–1190. [33](#)
- Lolli, B., & Gasperini, P. (2006). Comparing different models of aftershock rate decay: The role of catalog incompleteness in the first times after main shock. *Tectonophysics*, *423*(1-4), 43–59. [50](#)
- Mancini, S., Segou, M., Werner, M., & Cattania, C. (2019). Improving physics-based aftershock forecasts during the 2016–2017 central italy earthquake cascade. *Journal of Geophysical Research: Solid Earth*, *124*(8), 8626–8643. [33](#), [57](#)
- Mancini, S., Segou, M., Werner, M. J., & Parsons, T. (2020). The predictive skills of elastic coulomb rate-and-state aftershock forecasts during the 2019 ridgecrest, california, earthquake sequence. *Bulletin of the Seismological Society of America*, *110*(4), 1736–1751. [33](#), [57](#)
- Marsan, D. (2005). The role of small earthquakes in redistributing crustal elastic stress. *Geophysical Journal International*, *163*(1), 141–151. [33](#), [57](#)
- Marzocchi, W., & Taroni, M. (2014). Some thoughts on declustering in probabilistic seismic-hazard analysis. *Bulletin of the Seismological Society of America*, *104*(4), 1838–1845. [33](#)
- Marzocchi, W., Lombardi, A. M., & Casarotti, E. (2014). The establishment of an operational earthquake forecasting system in italy. *Seismological Research Letters*, *85*(5), 961–969. [33](#)
- Meier, M.-A., Werner, M., Woessner, J., & Wiemer, S. (2014). A search for evidence of secondary static stress triggering during the 1992 mw7. 3 landers, california, earthquake sequence. *Journal of Geophysical Research: Solid Earth*, *119*(4), 3354–3370. [57](#)
- Mignan, A., Werner, M., Wiemer, S., Chen, C.-C., & Wu, Y.-M. (2011). Bayesian estimation of the spatially varying completeness magnitude of earthquake catalogs. *Bulletin of the Seismological Society of America*, *101*(3), 1371–1385. [33](#)
- Mignan, A., & Chouliaras, G. (2014). Fifty years of seismic network performance in greece (1964–2013): Spatiotemporal evolution of the completeness magnitude. *Seismological Research Letters*, *85*(3), 657–667. [33](#)
- Mignan, A., & Woessner, J. (2012). Estimating the magnitude of completeness for earthquake catalogs. *Community Online Resource for Statistical Seismicity Analysis*, 1–45. [33](#)
- Mizrahi, L., Nandan, S., & Wiemer, S. (2021). The Effect of Declustering on the Size Distribution of Mainshocks. *Seismological Research Letters*. <https://doi.org/10.1785/0220200231>. [33](#), [35](#), [42](#)

- Nandan, S., Kamer, Y., Ouillon, G., Hiemer, S., & Sornette, D. (2021). Global models for short-term earthquake forecasting and predictive skill assessment. *The European Physical Journal Special Topics*, *230*(1), 425–449. [55](#)
- Nandan, S., Ouillon, G., & Sornette, D. (2019). Magnitude of earthquakes controls the size distribution of their triggered events. *Journal of Geophysical Research: Solid Earth*, *124*(3), 2762–2780. [45](#), [55](#), [59](#), [60](#)
- Nandan, S., Ouillon, G., & Sornette, D. (2021). Triggering of large earthquakes is driven by their twins. *arXiv preprint arXiv:2104.04592*. [57](#), [59](#), [60](#)
- Nandan, S., Ouillon, G., Sornette, D., & Wiemer, S. (2019a). Forecasting the full distribution of earthquake numbers is fair, robust, and better. *Seismological Research Letters*, *90*(4), 1650–1659. [55](#)
- Nandan, S., Ouillon, G., Sornette, D., & Wiemer, S. (2019b). Forecasting the rates of future aftershocks of all generations is essential to develop better earthquake forecast models. *Journal of Geophysical Research: Solid Earth*, *124*(8), 8404–8425. [33](#)
- Nandan, S., Ouillon, G., Wiemer, S., & Sornette, D. (2017). Objective estimation of spatially variable parameters of epidemic type aftershock sequence model: Application to california. *Journal of Geophysical Research: Solid Earth*, *122*(7), 5118–5143. [32](#), [36](#)
- Nandan, S., Ouillon, G., Woessner, J., Sornette, D., & Wiemer, S. (2016). Systematic assessment of the static stress triggering hypothesis using interearthquake time statistics. *Journal of Geophysical Research: Solid Earth*, *121*(3), 1890–1909. [57](#)
- Nandan, S., Ram, S. K., Ouillon, G., & Sornette, D. (2020). Is the earth crust operating at a critical point? *arXiv preprint arXiv:2012.06013*. [60](#)
- Nanjo, K., Ishibe, T., Tsuruoka, H., Schorlemmer, D., Ishigaki, Y., & Hirata, N. (2010). Analysis of the completeness magnitude and seismic network coverage of japan. *Bulletin of the Seismological Society of America*, *100*(6), 3261–3268. [33](#)
- Narteau, C., Byrdina, S., Shebalin, P., & Schorlemmer, D. (2009). Common dependence on stress for the two fundamental laws of statistical seismology. *Nature*, *462*(7273), 642–645. [50](#)
- Nichols, K., & Schoenberg, F. P. (2014). Assessing the dependency between the magnitudes of earthquakes and the magnitudes of their aftershocks. *Environmetrics*, *25*(3), 143–151. [59](#), [60](#)
- Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, *50*(2), 379–402. [32](#)
- Ogata, Y., & Katsura, K. (2006). Immediate and updated forecasting of aftershock hazard. *Geophysical research letters*, *33*(10). [34](#), [53](#)
- Omi, T., Ogata, Y., Hirata, Y., & Aihara, K. (2014). Estimating the etas model from an early aftershock sequence. *Geophysical Research Letters*, *41*(3), 850–857. [34](#)
- Ouillon, G., & Sornette, D. (2005). Magnitude-dependent omori law: Theory and empirical study. *Journal of Geophysical Research: Solid Earth*, *110*(B4). [55](#)
- Petersen, M. D., Mueller, C. S., Moschetti, M. P., Hoover, S. M., Rukstales, K. S., McNamara, D. E., Williams, R. A., Shumway, A. M., Powers, P. M., Earle, P. S., et al. (2018). 2018 one-year seismic hazard forecast for the central and eastern united states from induced and natural earthquakes. *Seismological Research Letters*, *89*(3), 1049–1061. [33](#)
- Rhoades, D., Liukis, M., Christophersen, A., & Gerstenberger, M. (2016). Retrospective tests of hybrid operational earthquake forecasting models for canterbury. *Geophysical Journal International*, *204*(1), 440–456. [33](#)
- Rhoades, D. A., Schorlemmer, D., Gerstenberger, M. C., Christophersen, A., Zechar, J. D., & Imoto, M. (2011). Efficient testing of earthquake forecasting models. *Acta Geophysica*, *59*(4), 728–747. [55](#)
- Rydelek, P. A., & Sacks, I. S. (1989). Testing the completeness of earthquake catalogues and the hypothesis of self-similarity. *Nature*, *337*(6204), 251–253. [33](#)

- Savran, W. H., Werner, M. J., Marzocchi, W., Rhoades, D. A., Jackson, D. D., Milner, K., Field, E., & Michael, A. (2020). Pseudoprospective evaluation of ucerf3-etas forecasts during the 2019 ridgecrest sequence. *Bulletin of the Seismological Society of America*, *110*(4), 1799–1817. [55](#)
- Schoenberg, F. P. (2013). Facilitated estimation of etas. *Bulletin of the Seismological Society of America*, *103*(1), 601–605. [37](#)
- Schoenberg, F. P., Chu, A., & Veen, A. (2010). On the relationship between lower magnitude thresholds and bias in epidemic-type aftershock sequence parameter estimates. *Journal of Geophysical Research: Solid Earth*, *115*(B4). [45](#)
- Schorlemmer, D., & Gerstenberger, M. (2007). Relm testing center. *Seismological Research Letters*, *78*(1), 30–36. [35](#)
- Schorlemmer, D., & Woessner, J. (2008). Probability of detecting an earthquake. *Bulletin of the Seismological Society of America*, *98*(5), 2103–2117. [33](#), [57](#)
- Segou, M., Parsons, T., & Ellsworth, W. (2013). Comparative evaluation of physics-based and statistical forecasts in northern california. *Journal of Geophysical Research: Solid Earth*, *118*(12), 6219–6240. [57](#)
- Seif, S., Mignan, A., Zechar, J. D., Werner, M. J., & Wiemer, S. (2017). Estimating etas: The effects of truncation, missing data, and model assumptions. *Journal of Geophysical Research: Solid Earth*, *122*(1), 449–469. [33](#), [44](#), [50](#)
- Shcherbakov, R., Turcotte, D. L., & Rundle, J. B. (2004). A generalized omori’s law for earthquake aftershock decay. *Geophysical research letters*, *31*(11). [50](#)
- Sornette, D., & Ouillon, G. (2005). Multifractal scaling of thermally activated rupture processes. *Physical review letters*, *94*(3), 038501. [57](#)
- Sornette, D., & Werner, M. J. (2005a). Constraints on the size of the smallest triggering earthquake from the epidemic-type aftershock sequence model, bath’s law, and observed aftershock sequences. *Journal of Geophysical Research: Solid Earth*, *110*(B8). [33](#), [37](#), [57](#)
- Sornette, D., & Werner, M. J. (2005b). Apparent clustering and apparent background earthquakes biased by undetected seismicity. *Journal of Geophysical Research: Solid Earth*, *110*(B9). [37](#)
- Spassiani, I., & Sebastiani, G. (2016). Exploring the relationship between the magnitudes of seismic events. *Journal of Geophysical Research: Solid Earth*, *121*(2), 903–916. [59](#), [60](#)
- Stallone, A., & Falcone, G. (2020). Missing earthquake data reconstruction in the space-time-magnitude domain. [34](#)
- Strader, A., Schneider, M., & Schorlemmer, D. (2017). Prospective and retrospective evaluation of five-year earthquake forecast models for california. *Geophysical Journal International*, *211*(1), 239–251. [55](#)
- Taroni, M., Marzocchi, W., Schorlemmer, D., Werner, M. J., Wiemer, S., Zechar, J. D., Heiniger, L., & Euchner, F. (2018). Prospective csep evaluation of 1-day, 3-month, and 5-yr earthquake forecasts for italy. *Seismological Research Letters*, *89*(4), 1251–1261. [33](#)
- U.S. Geological Survey, Earthquake Hazards Program. (2017). Advanced national seismic system (anss) comprehensive catalog of earthquake events and products: Various. <https://doi.org/10.5066/F7MS3QZH>. [60](#)
- Utsu, T. (1961). A statistical study on the occurrence of aftershocks. *Geophys. Mag.*, *30*, 521–605. [50](#)
- Veen, A., & 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. [32](#), [34](#), [36](#), [38](#)
- Wang, Q., Jackson, D. D., & Zhuang, J. (2010). Missing links in earthquake clustering models. *Geophysical Research Letters*, *37*(21). [45](#)

- Wells, D. L., & Coppersmith, K. J. (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of the seismological Society of America*, *84*(4), 974–1002. [47](#)
- Wiemer, S., Giardini, D., Fäh, D., Deichmann, N., & Sellami, S. (2009). Probabilistic seismic hazard assessment of switzerland: Best estimates and uncertainties. *Journal of Seismology*, *13*(4), 449. [33](#)
- Wiemer, S., & Wyss, M. (2000). Minimum magnitude of completeness in earthquake catalogs: Examples from alaska, the western united states, and japan. *Bulletin of the Seismological Society of America*, *90*(4), 859–869. [33](#)
- Woessner, J., Hainzl, S., Marzocchi, W., Werner, M., Lombardi, A., Catalli, F., Enescu, B., Cocco, M., Gerstenberger, M., & Wiemer, S. (2011). A retrospective comparative forecast test on the 1992 landers sequence. *Journal of Geophysical Research: Solid Earth*, *116*(B5). [33](#)
- Woessner, J., & Wiemer, S. (2005). Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty. *Bulletin of the Seismological Society of America*, *95*(2), 684–698. [33](#)
- Zechar, J. D., Schorlemmer, D., Werner, M. J., Gerstenberger, M. C., Rhoades, D. A., & Jordan, T. H. (2013). Regional earthquake likelihood models i: First-order results. *Bulletin of the Seismological Society of America*, *103*(2A), 787–798. [55](#)
- 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 and Space*, *69*(1), 36. [33](#), [34](#)

CHAPTER 4

DEVELOPING AND TESTING ETAS-BASED
EARTHQUAKE FORECASTING MODELS FOR
SWITZERLAND

To be submitted as:

Mizrahi, L., Nandan, S., & Wiemer, S. (2022). Developing and Testing ETAS-Based Earthquake Forecasting Models for Switzerland.

Abstract

We present the development and testing of multiple ETAS-based earthquake forecasting models for Switzerland. The simplest of these only require a sufficiently large catalog of earthquakes as an input and use Python code which is publicly available on GitHub. Besides the standard ETAS model, we consider five model variants. Three of these variants use information from the current state-of-the-art time-independent earthquake rate forecast of SUIhaz2015. The fourth model uses generic parameters fitted to Californian data, and the fifth variant is based on an extended calibration technique that allows for temporal variation of the completeness magnitude. We test all models using pseudo-prospective forecasting experiments and retrospective consistency tests. The results of 5 years of pseudo-prospective 7-day forecasting experiments suggest that all ETAS-based models outperform the time-independent SUIhaz2015 forecast. Based on this result and the results of the retrospective 30-year and 50-year consistency tests, we find that the purely ETAS-based models which do not include additional information and which use parameters calibrated on Swiss data are the best candidates for the first Swiss operational earthquake forecasting (OEF) system.

4.1 Introduction

Seismic risk is one of the largest threats posed to society by natural hazards in Switzerland and many other countries in Europe and worldwide. While earthquakes cannot be prevented nor predicted, an increased understanding of the probability of a devastating event can help societies prepare and mitigate the risk.

The most widely used approach to address seismic risk is through Probabilistic Seismic Hazard Analyses (PSHA, Cornell, 1968), in which long-term earthquake rate forecasts (also called earthquake source models) are the basis upon which seismic hazard maps are established and building codes are defined (see Woessner et al., 2015 for Europe; Drouet et al., 2020 for France; Meletti et al., 2017 for Italy; Sesetyan et al., 2018 and Demircioğlu et al., 2018 for Turkey; Mosca et al., 2022 for the UK; Wiemer et al., 2009 for Switzerland; Irsyam et al., 2020 for Indonesia; Stirling et al., 2021 for New Zealand; Petersen et al., 2018 for the US; Gupta et al., 2021 for India; etc.).

Such long-term studies, which are primarily time-independent, are a key ingredient for societies' resilience to seismic hazard. Recently, significant progress has been made towards the development of time-dependent operational earthquake forecasting (OEF) systems to supplement the existing time-independent hazard models, for example in Italy (Marzocchi et al., 2014), in the US (Field et al., 2017), New Zealand (Rhoades et al., 2016) and for the entire globe (Nandan, Kamer, et al., 2021; Kamer et al., 2021). The main benefit of OEF systems is that they enable reactions to temporally increased seismic hazard, such as evacuating buildings or shutting down power plants. Note however, that the development of appropriate strategies to react to elevated hazard is a highly nontrivial challenge that comes in addition to the challenge that is the development of the OEF system itself. In this article, we focus on the latter.

The most commonly used models used for time-dependent earthquake forecasting, and the basis of the four OEF systems or candidates mentioned above, are Epidemic-Type Aftershock Sequence (ETAS, Ogata, 1988) models. They account for the spatio-temporal clustering of earthquakes based on a few simple empirical principles. Having been around for several decades, ETAS models have been tested extensively, and remain the most successful earthquake forecasting models available today (Cattania et al., 2018; Mancini et al., 2019; Mancini et al., 2020; Nandan et al., 2019b; Taroni et al., 2018; Woessner et al., 2011), which makes them ideal candidates for OEF.

Most countries to date do not have OEF systems in place. A possible explanation for this is the seemingly large effort or simply the lack of resources and know-how required for their development. ETAS models, although having their focus on the short-term aspects

of seismic hazard, can also produce long-term forecasts which avoid the issue of b -value biases as artifacts of the common practice of catalog declustering in the standard PSHA approach (Mizrahi et al., 2021b). This is an additional incentive for seismic hazard analysts to strive to implement an ETAS model for their region of interest. In this paper, we describe the necessary steps to develop an ETAS-based OEF model, showcased on the example of Switzerland. This description will facilitate the implementation of ETAS-based forecasting models in other regions for which a reasonably sized earthquake catalog is available.

The current state-of-the-art earthquake source model in Switzerland is described in SUIhaz2015 (see Section 4.7 and Edwards et al., 2016). It combines the previous area source model (Giardini and Sismologique Suisse, 2004) and an updated version thereof, the relevant area sources of the 2013 SHARE model (Woessner et al., 2015), and a smoothed-seismicity model similar to the one presented by Hiemer et al. (2014).

We proceed as follows to develop an ETAS-based earthquake forecasting model for Switzerland to supplement the time-independent SUIhaz2015 model. We first discuss how the region, time window and magnitude range of the training catalog are selected in Section 4.2. Then we describe how the ETAS model is calibrated and how a forecast is issued using the ETAS model in Section 4.3. In addition to a general ETAS model, we describe multiple alternative versions allowing time-varying completeness, using generic parameters calibrated for California, or using additional information from the SUIhaz2015 model, such as the spatial distribution of background earthquakes or regionally varying magnitude frequency distribution. In Section 4.4 we describe pseudo-prospective and retrospective tests to verify the usefulness and suitability for the first Swiss OEF system. Our results are presented and discussed in Section 4.5, and we give our conclusions in Section 4.6.

4.2 Selection of catalog region, time window, magnitude range

To select the catalog on which we will calibrate an ETAS model for Switzerland, we start with the most recent earthquake catalog available (ECOS-09, Fäh et al., 2011). Several factors need to be considered when specifying the subset of the initial catalog used during model calibration.

The catalog should be complete above a threshold magnitude of completeness m_c (which may or may not be space- or time-dependent). This is a requirement of the ETAS model inversion. On top of this, the selection of region and time horizon determine the used catalog and their choice can largely affect the results of the parameter calibration. As Switzerland has relatively few large events when comparing to more seismically active regions such as California or Italy, we aim to choose our catalog such that the number of events is maximized.

The selection of region and time horizon directly influence the number of events in the catalog. They also influence the completeness magnitude of the catalog, which additionally indirectly affects the number of events in the remaining catalog.

4.2.1 Choice of region

Most straightforward is the choice of region. We choose the rectangle around the authoritative borders of Switzerland in which the smoothed seismicity model CH15 described in SUIhaz2015 is defined, as depicted in Figure 4.1(a). With this choice, our catalog spans the region of Switzerland and has a buffer zone around it so that earthquake interactions that cross the Swiss borders can still be accounted for in the calibration.

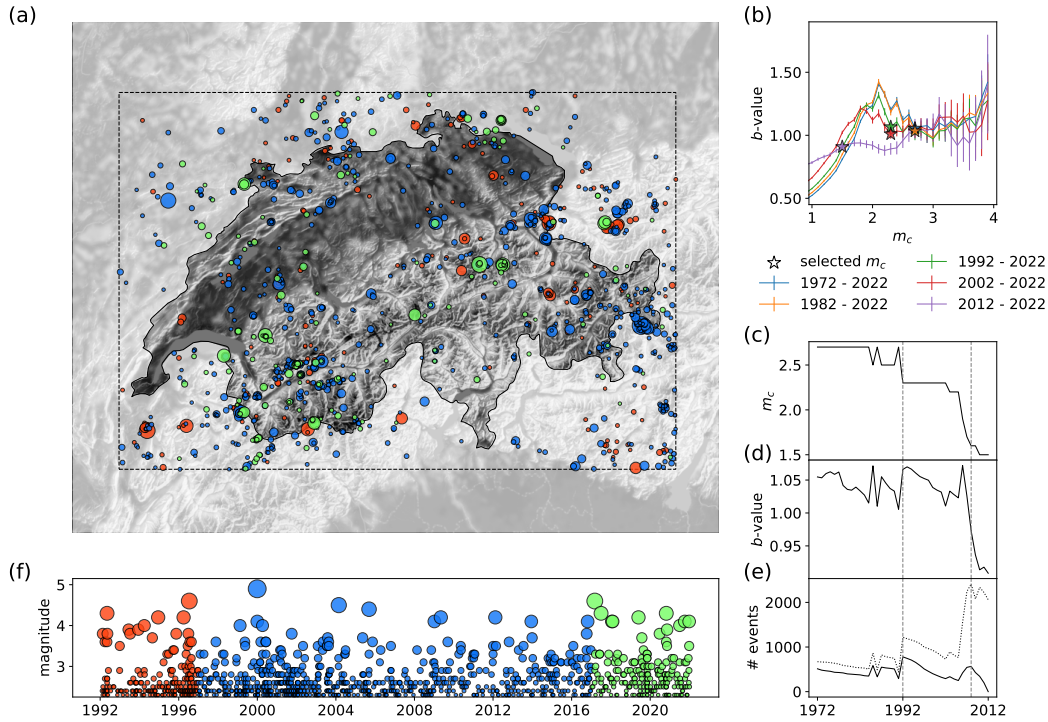


Figure 4.1: (a) Map of Switzerland. The dashed line indicates the border of the considered region; the solid line indicates the authoritative borders of Switzerland. Colored dots represent events in the Swiss catalog (ECOS-09, Fäh et al., 2011), where orange, blue, and green symbols represent events in the auxiliary, primary, and testing data set, respectively, and symbol size represents event magnitude. (b) b -value depending on m_c for different choices of time horizon (between 10 and 50 years) in Switzerland. Stars mark the estimated completeness magnitude for that time horizon. (c)-(e) Estimated completeness magnitude (c), corresponding b -value (d), and the number of events (e) contained in the complete catalog (dotted line) and complete primary catalog (solid line) for time horizons starting between 1972 and 2012, ending in 2022, with 1-year steps. The starting year is indicated on the x -axis. Dashed grey lines mark the time horizons with the highest number of events in the complete catalog (2008-2022) and the complete primary catalog (1992-2022). (f) Temporal distribution of events in the catalog, with identical size and color coding as in (a).

4.2.2 Choice of time horizon and magnitude range

The time horizon and magnitude range selection can be tricky as the two are closely intertwined. For a first basic ETAS model calibration, we make the simplifying assumption of a spatially and temporally constant magnitude of completeness. A more general case will be discussed in Section 4.2.2. Furthermore, we use binned magnitudes with bin size $\Delta m = 0.1$.

Seismic networks usually improve over time. Thus, longer periods (i.e. more data) and increased completeness magnitude (i.e. less data) have antagonistic effects on the number of events in the final complete catalog. Figure 4.1(b) shows the evolution of the estimated b -value in Switzerland depending on assumed m_c , for different time horizons. Typically, the b -value is underestimated when m_c is underestimated and stabilizes above a certain m_c . For the time periods starting earlier than 2012, the increase in b -values is followed by a decrease before stabilization is reached, which may be due to mixing of different magnitude scales. The period of 2012-2022 does not show this behavior and consists mostly of events of the same magnitude type. The stars in Figure 4.1(b) indicate the overall m_c values estimated

for the respective time horizons using the method of Mizrahi et al. (2021b). The method selects m_c such that the observed sample of magnitudes is plausible to have emerged from a GR law with the b -value associated with this value of m_c .

To select the time horizon, we systematically assess the dependency of the size of the resulting complete catalog on the time horizon.

Figures 4.1(c)-(e) show how the estimated magnitude of completeness, the corresponding b -value, and the corresponding number of events in the complete catalog (dotted line) vary with the choice of time horizon. If one were to maximize this number, the clear best choice of time horizon would be to use data between 2008 and 2022 with a m_c value of 1.6, which results in a catalog containing 2413 events.

As will be described later in Section 4.3.1, the first few years of the catalog will be used as auxiliary data so that events at the beginning of the primary time period are not wrongly interpreted as background events due to the lack of potential parent events in the data set. Furthermore, as will be described in Section 4.4, a calibrated model should be tested using data that has not been used in the calibration, which shortens the primary time period available for model inversion from the other end. If five years of seismicity are allocated both as auxiliary data and as testing data, the time horizon of 2008-2022 yields a primary time period that only spans four years, from 2013 to 2017. The solid line in Figure 4.1(e) shows the number of events in the complete primary catalog that remains after removing the first and last five years of the complete catalog. For the overall time horizon of 2008-2022, the corresponding primary time period contains 566 events. The overall time horizon of 1992-2022 maximizes the number of events in the complete primary catalog (780 events) at a completeness magnitude of $m_c = 2.3$ at a b -value of 1.07. Besides the advantage of a larger data set, this choice of time horizon has the additional advantage that the primary time window spans 20 years (1997-2017). Parameters estimated based on this data are therefore more representative of the longer-term behavior of seismicity.

With this catalog, we proceed to calibrate our ETAS model. Figure 4.1(a) and (f) show the spatial and temporal distribution of events in our catalog, where dot size represents event magnitude, and orange, blue, and green dots represent events in the auxiliary, primary, and testing data set, respectively.

Spatially or temporally varying catalog completeness

An extra layer of complexity may arise when working with a larger region with a more heterogeneous seismic network, such as for example the European one. In this case, the completeness magnitude and time horizon of complete catalogs can vary substantially between sub-regions. For the case of Switzerland, we assume spatial homogeneity of the completeness magnitude.

Besides spatial variations, completeness also shows variations in time. The topic of estimating ETAS parameters while accounting for such temporal variations has recently been thoroughly addressed by Mizrahi et al. (2021a) and Hainzl (2022). In addition to a basic ETAS model based on the catalog described above, we here want to consider a simple case of temporally varying completeness magnitude for Switzerland. With the intention to include data from a longer time horizon, we proceed as follows. We leave the catalog described above from 1992 to 2022 unchanged with $m_c \equiv 2.3$ and extend it towards the past. Estimating the completeness of a catalog is highly nontrivial and becomes more difficult when fewer data points are available, because statistical tests lose their power. For this reason, we limit ourselves to using the instrumental catalog of Switzerland, which starts around 1972 (Fäh et al., 2011). We apply the same technique as before to estimate m_c for the catalog between 1972-1992, arriving at a value of 2.7 with a b -value of 1.15. The true completeness certainly varies during this period, and the assumption of a constant m_c is a simplification of a much more complex reality. The higher b -value compared to the later period is possibly explained by the different magnitude types being used in different periods. However, to develop a simple model, we decide not to introduce further complexity into our

m_c model. Combining the catalogs of 1972-1992 and 1992-2022 and discounting the first and last five years of data yields a primary catalog of 1186 events, spanning a time window of 40 years between 1977-2017.

4.3 From catalog to forecast

In this section we describe how an ETAS model is calibrated based on the given catalog, and how a forecast can be issued given a calibrated ETAS model.

4.3.1 ETAS parameter calibration

In ETAS, each earthquake is either a background event or an aftershock of a previous event, and each earthquake can trigger aftershocks. Background events are assumed to occur uniformly in time, and aftershocks follow a certain distribution in time and space around their parent event. The number of aftershocks of any parent event is dependent on the magnitude of the parent event. Earthquake magnitudes follow the Gutenberg-Richter law (Gutenberg and Richter, 1944), independently of whether the event is a background or a triggered event. To calibrate ETAS parameters, we here use the model described by Nandan et al. (2017) or Mizrahi et al. (2021b), which defines the rate g of aftershocks triggered by an event of magnitude m , at a time delay of Δt and a spatial distance $(\Delta x, \Delta y)$ from the parent event as

$$g(m, \Delta t, \Delta x, \Delta y) = \frac{k_0 \cdot e^{a(m-m_{ref})} \cdot e^{\Delta t/\tau}}{((\Delta x^2 + \Delta y^2) + d \cdot e^{\gamma(m-m_{ref})})^{1+\rho} \cdot (\Delta t + c)^{1+\omega}}. \quad (4.1)$$

Here, m_{ref} is the reference magnitude, which is an important quantity and deserves some discussion. In general ETAS formulations, it is assumed that earthquakes can trigger and be triggered if and only if their magnitude is greater than or equal to the reference magnitude. Moreover, the catalog used to calibrate the model must be complete above m_{ref} , as otherwise, earthquakes will be wrongly interpreted to be background events, even though they may have been triggered by an event that is missing in the catalog. For simplicity, we choose m_{ref} to be equal to the estimated magnitude of completeness $m_c = 2.3$ (see Sornette and Werner, 2005 for implications of this assumption). Note that we implicitly assume that earthquakes of magnitudes below 2.3 do not trigger any aftershocks. In this sense, earthquakes that are triggered by events smaller than $M = 2.3$ are considered background events.

The total rate of events at any time t and location (x, y) is then given by

$$\lambda(t, x, y) = \mu + \sum_{i:t_i < t} g(m_i, t - t_i, x - x_i, y - y_i). \quad (4.2)$$

That is, the seismicity rate at a given time and location is the sum of the background rate μ and the rate of all aftershocks of previous events e_i with magnitudes m_i , which occurred at times t_i and at locations (x_i, y_i) .

For the calibration of the ETAS parameters $\mu, k_0, a, \tau, c, \omega, d$, and γ , we use the Python implementation of the expectation maximization algorithm, which is available on the author's [GitHub repository](#) (algorithm: Veen and Schoenberg, 2008, implementation: Mizrahi and Schmid, 2022, see Appendix E).

In the case of spatially or temporally varying completeness as discussed in Section 4.2.2, one can select $m_{ref} := \min_i(m_c^i)$, the minimum value of m_c across all sub-regions or time-intervals, and use the ETAS parameter inversion of Mizrahi et al. (2021a) for spatially varying completeness when $m_c(x, y, t)$ is given as an input to the inversion algorithm (also available on [GitHub](#), Mizrahi and Schmid, 2022, see Appendix E). This inversion technique

Table 4.1: ETAS parameters, b -value, branching ratio η and $\log_{10}(\Lambda)$ (defined in Equation (4.3)) estimated from the Swiss catalog using ETAS and complETAS inversion, using data until 01 Jan, 2017 and until 01 Jan, 2022. The fifth column shows generic parameters estimated from Californian data, adapted to match the Swiss background rate and a reference magnitude of 2.3.

	ETAS		complETAS		caliETAS
	as of 2017	as of 2022	as of 2017	as of 2022	
$\log_{10}(\mu)$	-6.23	-6.21	-6.16	-6.16	-6.17
$\log_{10}(k_0)$	-2.56	-2.75	-2.71	-2.85	-2.84
a	0.94	1.13	1.29	1.37	1.69
$\log_{10}(c)$	-2.92	-2.85	-2.79	-2.76	-2.95
ω	-0.16	-0.13	-0.13	-0.1	-0.03
$\log_{10}(\tau)$	3.55	3.57	3.56	3.58	3.99
$\log_{10}(d)$	-0.28	-0.51	-0.37	-0.56	-1.04
γ	0.09	0.15	0.37	0.38	1.22
ρ	0.65	0.63	0.64	0.62	0.51
b -value	1.12	1.05	1.13	1.08	from ETAS
η	0.60	0.57	0.48	0.48	0.84
$\log_{10}(\Lambda)$	-5.83	-5.84	-5.88	-5.88	-5.37

assumes that the observed part of the catalog (above time-dependent completeness) is representative of the unobserved part at all times. The estimated triggering power of observed events and the estimated number of potential parent events are then inflated according to this assumption to account for unobserved data. In the following, we will refer to this version of the ETAS model as *complETAS*.

Table 4.1 shows the estimated ETAS parameters, as well as the b -value and branching ratio η for a primary time horizon starting in 1997 and ending in January 2017 (start of the first testing period, first column) and January 2022 (second column). Similarly, the third and fourth columns show the resulting parameters when time-varying completeness is accounted for and the start of the primary time horizon is in 1977. These are the parameters as they would have been estimated on the indicated as-of-date, not knowing the events after that day. For the evolution of parameters for increasing primary time horizons ending between 2017 and 2022, see Figure D.1. As a reference, the fifth column of Table 4.1 shows generic parameters estimated from Californian data by Mizrahi et al. (2021b), adjusted for a lower reference magnitude of $m_{ref} = 2.3$ using the transformation method described by Mizrahi et al. (2021a). The background seismicity rate is inferred from the SUIhaz2015 model.

4.3.2 Discussion of inverted parameters

Except for c , the values of all calibrated complETAS parameters (i.e. $\mu, k_0, a, \tau, \omega, d$, and γ) lie between regular ETAS and caliETAS parameters. A possible interpretation of this observation is that the longer term seismicity behaviour in Switzerland may be more similar to California than what has been observed in the more recent past. Alternatively, increasing the fraction of larger magnitude events in the training catalog used in complETAS could explain more similar parameters to the parameters obtained in California, which were estimated from a catalog with a higher completeness magnitude of $m_c = 3.6$.

Despite most complETAS parameters being closer to caliETAS parameters than ETAS parameters are to caliETAS parameters, the branching ratio η , which is calculated theoretically from the individual parameters as described in Mizrahi et al. (2021a), is lower for complETAS and much higher in California. This indicates that the shorter-term ETAS seismicity after all may be more similar to Californian seismicity than complETAS seismicity is to Californian seismicity. Complex interdependencies between parameters could

cause similarities of individual parameters discussed above, and we shall be reminded not to over-interpret them.

Comparing ETAS with complETAS, a salient difference lies between the values of the background rate μ . On average between 2017 and 2022, complETAS shows a background rate that is 15% higher than ETAS, where the difference is decreasing with increasing time horizon. The higher background rate is accompanied by a smaller value for k_0 (indicating fewer aftershocks) and a larger value for a (indicating a stronger magnitude dependency of the aftershock productivity). The overall branching ratio is lower for complETAS than ETAS, which can be viewed as a compensation for the higher background rate. Hypothetically, if every earthquake triggered exactly η aftershocks, the total rate of events could be calculated as

$$\Lambda = \mu \cdot \lim_{n \rightarrow \infty} \sum_{i=0}^n \eta^i = \mu \cdot \frac{1}{1 - \eta}, \quad (4.3)$$

provided that $\eta < 1$. In this scenario, $\log_{10}(\Lambda)$ values for ETAS range between -5.83 and -5.85, and between -5.88 and -5.89 for complETAS. Thus, despite the much higher background seismicity rate of complETAS, the hypothetical overall seismicity rate Λ is 10% lower for complETAS than ETAS. caliETAS in contrast has a much higher $\log_{10}(\Lambda)$ than ETAS or complETAS, which amounts to almost three times the expected seismicity than that of ETAS. This is due to the combination of a high branching ratio and a background rate obtained from SUIhaz2015 which is not designed to have aftershock sequences added to it.

4.3.3 Issuing a forecast through catalog simulation

A seismicity forecast is given as a probability density function (PDF) of the number of earthquakes to occur in a predefined space-time-magnitude range. To issue a forecast using ETAS, one needs to stochastically simulate possible continuations of the given catalog for the duration of the forecast time range. This is done by simulating cascades of aftershocks of events that occurred before the start of the forecasting time range and background events that occur during the forecasting time range plus their cascades of aftershocks. A detailed description of the catalog simulation algorithm used here can be found in Section B.1 in the Supporting Information for Chapter 3 (Mizrahi et al., 2021a), and a Python code to stochastically simulate catalog continuations can be found on [GitHub](#) (Mizrahi and Schmid, 2022, Appendix E).

Note that the locations of background events are simulated by sampling from the locations of the events in the training catalog. The likelihood of an event location being selected is proportional to the probability of this event to be a background event. The sampled locations are then slightly dislocated by a vector drawn from a Gaussian distribution such that the mean dislocation is 0.

A common number of simulations done is 100,000 (Savran et al., 2020; Nandan, Kamer, et al., 2021), although more simulations may be necessary in certain cases. If one is interested in a forecast with a high spatial resolution, the chance of an event to occur in a spatial cell of interest may be lower than 1 in 100,000. Thus, 100,000 simulations are not enough to estimate the probability of such an event. Similarly, if one is interested in forecasting large magnitude events, those may have low chances of occurrence, and more simulations are needed to estimate their probability of occurrence.

The PDF $p(k)$ of the number of events to occur in a space-time-magnitude range can be derived through the simulated catalogs as follows. We define $p(k)$ as the empirical distribution resulting from n_{sim} simulations. Consider $n(k)$, the number of simulations among n_{sim} for which k events are observed. To avoid zero probabilities for values of k that do not appear in the simulations, i.e. $n(k) = 0$, we define a large number k^* until which the probability should be nonzero. We define the PDF of the number of earthquakes k as

$$p(k) = \begin{cases} \frac{n(k)}{n_{sim}+1} & \text{if } n(k) > 0 \\ \frac{1}{m \cdot (n_{sim}+1)} & \text{if } n(k) = 0 \text{ and } k \leq k^* \\ 0 & \text{otherwise,} \end{cases} \quad (4.4)$$

where m is the number of values k between zero and k^* for which $n(k) = 0$.

4.4 Testing of model variants

In this section, we describe six ETAS variants to be considered for the first Swiss OEF system. We then describe pseudo-prospective 7-day forecasting experiments and retrospective consistency tests spanning several decades. In this way, we thoroughly test all candidate models before selecting the best suited one(s).

4.4.1 Competing models

Null Model: SUIhaz2015

Before we start describing the six ETAS model variants, we define a null model to which the six candidates will be compared in the pseudo-prospective tests. For this purpose we consider the time-independent earthquake rate forecast from the SUIhaz2015 hazard maps, defined on 0.05×0.05 degrees latitude and longitude. The PDF $p_{null}(k)$ of the number of earthquakes k can be derived analytically as a Poisson distribution with the mean being the expected number of events to occur in the space-time-magnitude range under consideration.

In the SUIhaz2015 model, earthquake rates are defined for events of magnitude 4.0 or above. We will be interested in forecasting the number of $M \geq 2.5$ events (see Section 4.4.2). Thus, we extrapolate from the $M \geq 4.0$ rates using the Gutenberg-Richter distribution, where the b -value is estimated separately for each zone of the seismogenic zonation model SEIS-15.

We calculate the zone-wise b -value estimator which accounts for binned magnitudes with bin size Δm (Tinti and Mulargia, 1987) as

$$\hat{b} = \frac{\log_{10}(\hat{p})}{\Delta m}, \quad (4.5)$$

where

$$\hat{p} = 1 + \frac{\Delta m}{\bar{m} - m_0}, \quad (4.6)$$

and \bar{m} is the mean of the binned magnitudes m_i , weighted by the rate λ_i associated to each magnitude bin,

$$\bar{m} = \frac{\sum_i m_i \cdot \lambda_i}{\sum_i \lambda_i}. \quad (4.7)$$

In our case, we have magnitude bins of size $\Delta m = 0.1$, starting at $m_0 = 4.0$. Figure C.2 shows the resulting b -values for the relevant SEIS-15 zones.

Not all cells within one seismogenic zone have the same earthquake rates. The extrapolation is therefore done for each cell individually, using the b -value of its zone.

ETAS

The first alternative model is the most basic ETAS model, which uses the training catalog shown in Figure 4.1 (a) and (f) of earthquakes of magnitude $M \geq 2.3$ since 1992, where the

primary time period starts in 1997. The end of the training catalog is the start date of the forecasting period. Thus, parameters may vary from one forecasting period to the next.

The simulation of background events is done as described in Section 4.3.3 using the inferred background probabilities of events which occurred in the primary time period of the training catalog.

For this model, and analogously for all ETAS-based models described below, we define $p_{ETAS}(k)$ as in Equation 4.4.

compleTAS

The compleTAS model is analogous to the ETAS model, with the difference that it uses a larger training catalog with time-varying completeness as described in Section 4.2.2 with the start of the primary time period in 1977. The parameters are estimated using the generalized inversion algorithm of Mizrahi et al. (2021a).

hazETAS-BG

As a modification of the standard ETAS model, we also consider hazETAS-BG, where the locations of background events are simulated based on the rates provided by SUIhaz2015, all else staying identical. In particular, the total rate of background events is still determined by the parameter μ given by the ETAS model. Only the relative spatial differences in SUIhaz2015 rates are relevant for the simulation of background event locations. Note that we use the rates extrapolated to $M \geq 2.5$ as described in Section 4.4.1.

The Swiss Seismic Hazard Maps 2015 (SUIhaz2015) are based on expert knowledge and data from a much longer time scale. Considering this ETAS variant will allow us to quantitatively assess potential improvements in forecasting due to the inclusion of this additional information.

hazETAS-mag

With the hazETAS-mag model, a second way to combine ETAS with SUIhaz2015 is tested. It uses the magnitude distribution provided by SUIhaz2015 when simulating earthquake magnitudes, all else staying identical. In particular, background event locations are simulated as in the ETAS model.

The magnitude simulation for hazETAS-mag is done as follows. After simulating the location of an event, its magnitude is drawn from the distribution valid in the SEIS-15 zone into which the event falls.

The magnitude frequency distribution of a zone is obtained by taking the sum over all cells within that zone of the SUIhaz2015 rates per magnitude bin. This distribution is then extrapolated to magnitudes below 4.0 using the zone-wise b -values calculated in Section 4.4.1.

Note that magnitudes for which rates are zero according to the SUIhaz2015 model (i.e. magnitudes above 7.3) will not be drawn during ETAS simulation. This is in contrast to the previously described ETAS models which use a pure GR law, in which very large magnitudes have a low chance of being simulated but are technically possible.

hazETAS-full

The last model of the hazETAS group is hazETAS-full, which uses both aspects of SUIhaz2015 of the previous two models. It simulates background event locations according to SUIhaz2015, and it simulates magnitudes according to SUIhaz2015.

caliETAS

The final ETAS-based model considered in our experiment is caliETAS. This model is included to challenge and/or confirm the necessity of calibrating ETAS parameters using regional data.

Instead of using ETAS parameters which are obtained from the most recent Swiss catalog, it uses generic ETAS parameters that were estimated for California by Mizrahi et al. (2021b), adjusted to the reference magnitude of $m_{ref} = 2.3$ using the transformation method described by Mizrahi et al. (2021a).

As California is much more seismically active than Switzerland, and for this model, we do not use information that results from the ETAS parameter inversion applied to the Swiss catalog, we use the background seismicity rate provided by SUIhaz2015. For this, we calculate the background rate μ in the entire region based on the SUIhaz2015 rates extrapolated down to $M \geq 2.3$ as described in Section 4.4.1. The simulation of background event location is then done in the same manner as for the hazETAS-BG model.

The parameters used for simulation can be found in Table 4.1.

4.4.2 Pseudo-prospective tests

The first test is to compare ETAS variants to the SUIhaz2015 null model in pseudo-prospective one-week forecasting experiments. For this, we issue a forecast using data until a certain time t , and then compare the actual catalog between time t and $t + \Delta t$ with the forecast. This is done for all considered models and for several consecutive testing periods to obtain a representative sample of the quality of the forecast. This concept is illustrated Figure 4.2 (top).

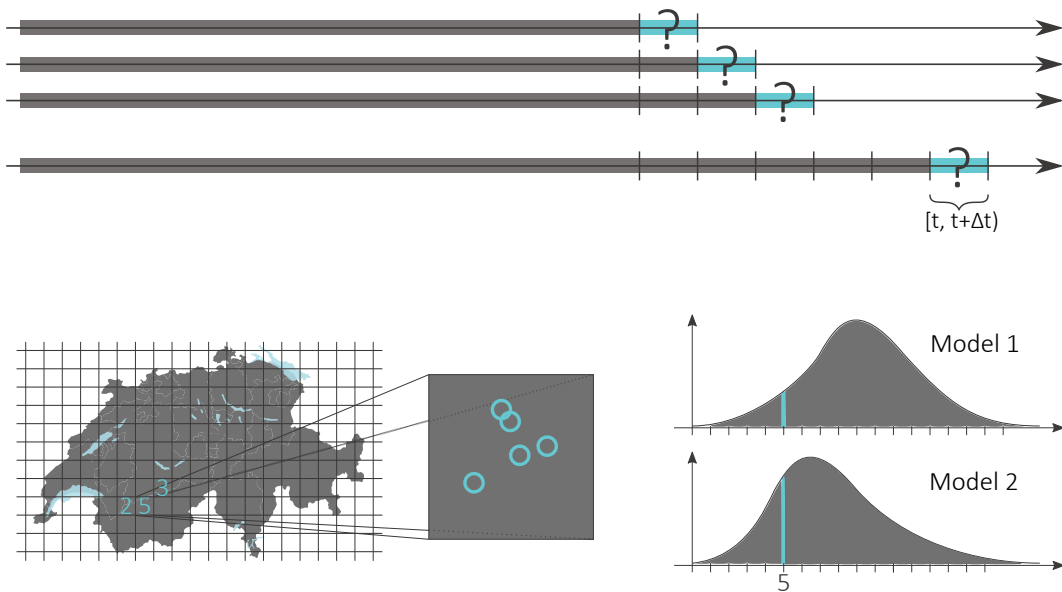


Figure 4.2: Schematic illustration of the concept of pseudo-prospective forecasting experiments (top) and model comparison (bottom).

The beginning of our first testing period is January 1st, 2017, and we consider consecutive non-overlapping testing periods of length $\Delta t = 7$ days, with the start of the last testing period on December 26th, 2021.

The metric used to evaluate a forecast is the information gain over a null model as described by Nandan et al. (2019a). Its concept is schematically illustrated in Figure 4.2 (bottom). The goal is to correctly forecast the number of earthquakes above a magnitude

threshold m_{thr} which occur in the time interval $[t, t + \Delta t)$, in each grid cell of a pre-specified spatial grid covering the region of interest. In each grid cell, the probability density function $p(k)$ of the number of earthquakes k is calculated for both the alternative and the null model. The information gain of the alternative versus the null model is defined as

$$IG = \sum_i \ln(p_{alt}^i(k_{true}^i)) - \ln(p_{null}^i(k_{true}^i)), \quad (4.8)$$

where summation is over all grid cells.

We are interested in correctly forecasting the number of earthquakes of magnitude $m_{thr} = 2.5$ or above, which is the threshold magnitude for which media alerts are released in Switzerland. Because of the limited amount of data in the five year testing period, a higher threshold would make the test less powerful. We will address the adequacy of the tested models at a higher magnitude threshold using retrospective 30- and 50-year tests described in Section 4.4.3.

As a spatial grid, we choose a grid of 0.05×0.05 degrees latitude and longitude, which is the resolution of the SUIhaz2015 hazard maps. We set the number of simulations to be performed for each testing period to be $n_{sim} = 100,000$. We consider the grid cells in which our newly calibrated ETAS model and the SUIhaz2015 model are defined. Finally, in each grid cell, we choose $k^* = 500$ to be the maximum value for which $p(k)$ (see Equation (4.4)) shall return a nonzero value.

In summary, we consider 216 testing periods and 4075 grid cells, in which 164 earthquakes with $M \geq 2.5$ occurred. We call the area consisting of these 4075 grid cells the testing region.

The alternative model is considered superior to the null model if the mean information gain averaged over all testing periods is positive, where significance is calculated through a paired one-sided t-test with a significance threshold of 0.1.

4.4.3 Retrospective tests

In addition to the pseudo-prospective tests, we perform retrospective tests at a larger magnitude threshold of $m_{thr} = 4.5$. In particular, we use the catalog-based forecast evaluation tests provided by the pyCSEP toolbox (Savran et al., 2022) and described by Savran et al. (2020). They allow us to test whether an observed catalog is an outlier when compared with the simulated catalogs which constitute a forecast. If this is not the case, the observation shall be considered consistent with the forecast. Different tests address the consistency of the magnitude distribution, the spatial distribution, the number distribution, and the overall pseudolikelihood of observed events. We use the same spatial grid cells as in the pseudo-prospective test, and magnitude bins of size 0.1.

The different model variants are tested as follows. Starting at the end of the pseudo-prospective testing period (end of 2021), we issue a $50 \cdot 365$ days $\simeq 50$ years long forecast by generating 100,000 simulated catalogs, for each of the ETAS variants. We then compare these catalogs to the observed catalog which ends in 1992 and goes back $50 \cdot 365$ days. The observed catalog thus starts in 1942. According to the method described by Mizrahi et al. (2021b), this catalog is complete at the magnitude threshold of 4.5. Besides the 50 year test, we analogously perform a 30 year test with the observed catalog starting in 1962.

Although these tests are retrospective, they can be considered partially out-of-sample, as the observed catalog used to evaluate the forecasts was not used in the calibration of the ETAS parameters. Note however that the data was used to construct the SUIhaz2015 model, and that part of the data was used in the calibration of completAS parameters.

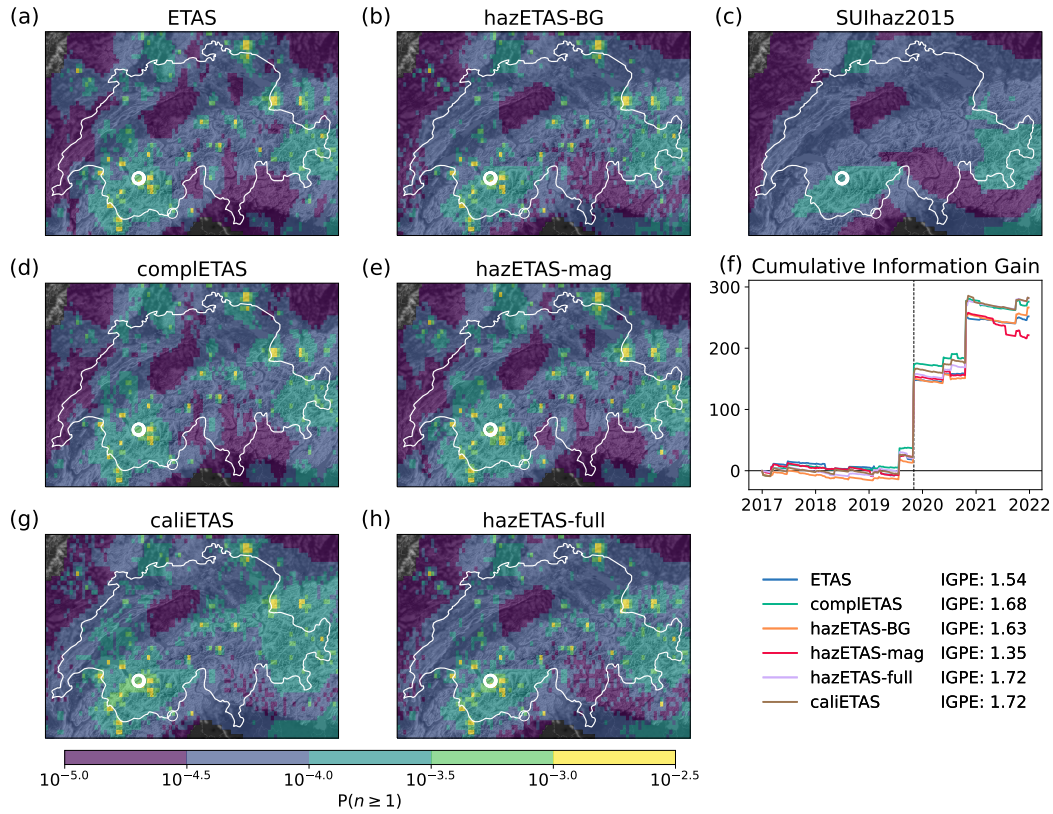


Figure 4.3: (a-e, g, h) Examples of 7-day seismicity forecasts produced by the considered models for the week starting on 03 Nov, 2019. The color of each pixel of $0.05^\circ \text{lat} \times 0.05^\circ \text{lon}$ ($\approx 5 \text{ km} \times 5 \text{ km}$) corresponds to the probability of one or more $M \geq 2.5$ earthquakes to occur in the pixel during 7 days. White circles represent 15 $M \geq 2.5$ earthquakes which actually occurred during the period for which the forecast was issued. (f) Cumulative information gain (CIG) over time of the ETAS-based models versus the time-independent SUIhaz2015 model represented by the black horizontal line. Vertical line marks the date of the forecasts shown in (a-e, g, h). Information gain per earthquake (IGPE) over SUIhaz2015 is given in the legend.

4.5 Results and discussion

4.5.1 Results of the pseudo-prospective tests

Figure 4.3 shows example forecasts produced by the considered models in the testing region (colored region) for the week starting on 03 Nov 2019, in which a large number of 15 $M \geq 2.5$ events occurred, with the locations of those 15 events indicated by white circles on the maps. As most events have occurred very close to each other, it is difficult to distinguish the individual circles.

The time-dependent forecasts show similar features of increased probability compared to the time-independent forecast. They must be caused by seismic activity prior to the forecasting period. Meanwhile, certain differences can be visually identified in-between the time-dependent forecasts. For instance, the different spatial distributions of background seismicity of the models can be recognized (see also Figure C.3 for the forecasted background seismicity only). caliETAS shows large areas of increased earthquake probabilities compared to other forecasts. Such features will however be analyzed more quantitatively and based

Table 4.2: Summary of forecasting results for the considered models. The first four columns show the probability of one or more earthquake of magnitude $M \geq 2.5$ to occur in the next seven days, starting on 03 Nov 2019 for the first and third columns, and starting on 10 Nov for the second and fourth columns. The first two columns show the probability for the entire testing region, the third and fourth columns for the single pixel of $0.05^\circ\text{lat} \times 0.05^\circ\text{lon}$ ($\simeq 5\text{km} \times 5\text{km}$) centered at $46.325^\circ\text{lat}, 7.375^\circ\text{long}$. The fifth, sixth, and seventh columns show the total information gain (IG) of ETAS-based models vs SUIhaz2015 for the entire testing period, the information gain per earthquake (IGPE), and the mean information gain (MIG) averaged over the testing periods. The last column indicates the p-value of a paired one-sided t-test, testing the null hypothesis of whether the alternative model’s MIG is less than or equal to zero.

	$P(n \geq 1)$ in CH		$P(n \geq 1)$ in pixel		total IG	IGPE	MIG	p-value
	Nov 03	Nov 10	Nov 03	Nov 10				
SUIhaz2015	24%	24%	0.03%	0.03%	0	0	0	-
ETAS	29%	55%	0.06%	28%	251.98	1.54	0.97	0.063
complETAS	34%	56%	0.05%	25%	276.06	1.68	1.06	0.053
hazETAS-BG	35%	61%	0.07%	29%	266.53	1.63	1.02	0.063
hazETAS-mag	38%	63%	0.08%	31%	221.24	1.35	0.85	0.097
hazETAS-full	37%	64%	0.07%	31%	281.28	1.72	1.08	0.055
caliETAS	63%	74%	0.17%	42%	282.02	1.72	1.08	0.055

on more than one week of data using the consistency tests.

Figure 4.3(f) shows the cumulative information gained over time by the ETAS-based models compared to the time-independent SUIhaz2015 model. The mean information gain per earthquake (IGPE) of the ETAS-based models against the null model is given in the legend. All ETAS-based models have a positive IGPE compared to SUIhaz2015. Among them, IGPE values range between 1.35 and 1.72.

Table 4.2 provides a summary of the forecasting results of the considered models. The first four columns show the 7-day forecasted probabilities of one or more magnitude $M \geq 2.5$ events for two different forecasting periods and two different regions. The first two columns give probabilities for the entire testing region, while the third and fourth columns give probabilities for the single pixel centered at $46.325^\circ\text{lat}, 7.375^\circ\text{long}$. In this pixel, 13 out of 15 earthquakes occurred during the first shown testing week starting on 03 Nov 2019. It was the week with the highest number of events among the tested forecasting periods of this experiment. Two out of two earthquakes in the following week starting on 10 Nov also occurred in the same pixel.

All ETAS-based models forecast a higher probability of occurrence than the null model for both weeks and regions, and all of them show an increase from the first to the second week. The highest increase is observed with the ETAS model. For the entire testing region, it gives a 91% increase of the occurrence probability in the second week compared to the first week, and in the active pixel the probability is more than five times as high as for the first week.

All ETAS-based models outperform the null model with a high IGPE as well as mean information gain (MIG) averaged over the testing periods (see Table 4.2, columns six and seven). The significance levels indicated by the p-values in the last column of Table 4.2 show that all alternative models achieve a significance level of < 0.1 .

Figure 4.4 shows the pairwise comparison of all competing models. The clearest signal is the superiority of all time-dependent models over the time-independent null model. This main result of our 7-day forecasting experiment was expected and is likely to be explained by the fact that the main strength of ETAS lies in capturing short-term earthquake triggering behavior accurately.

In-between the ETAS-based models, the only significant performance difference with a

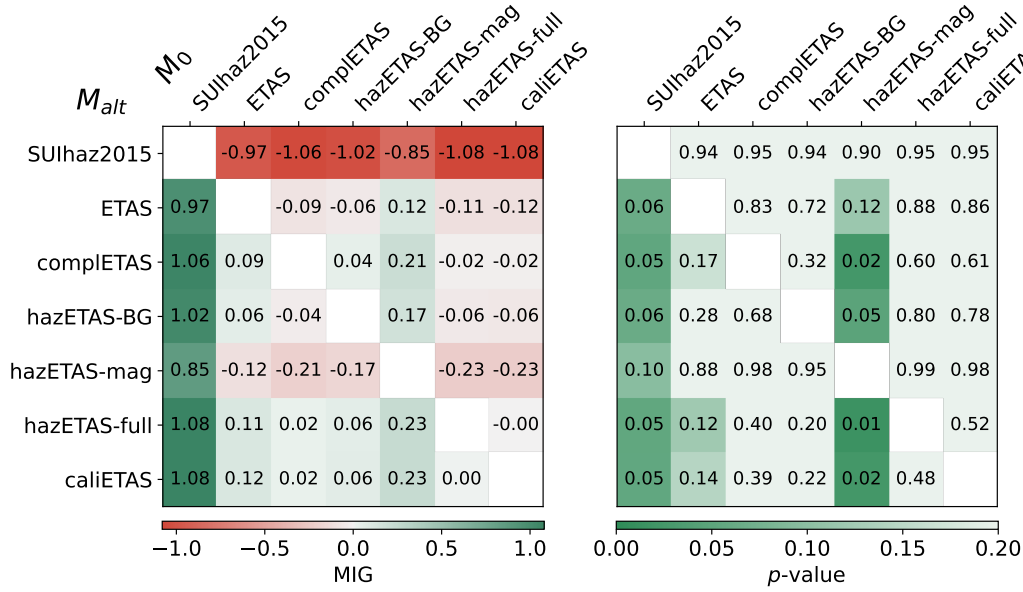


Figure 4.4: Pairwise comparison of all competing models. Left: Mean information gain (MIG) of the alternative model (M_{alt} , rows) versus the null model (M_0 , columns). Right: Same as left, but showing p -value of the paired one-sided t-test.

p -value of less than 0.1 is observed when comparing to hazETAS-mag. It is significantly outperformed by four out of five ETAS-based models. We conclude that hazETAS-mag is not the ideal candidate for an OEF model in Switzerland.

4.5.2 Results of the consistency tests

Figure 4.5 shows the results of the pyCSEP consistency tests for a 30 year and a 50 year testing period. In these tests, a test statistic is calculated for each simulated catalog and for the observed catalog. The quantile score of the observed statistic among the simulated test statistics is then used to decide whether a model passes the test. If the quantile score, which is given in the cells in Figure 4.5, lies between 0.05 and 0.95, we consider the test to be passed, indicated by a green cell color.

For the 30 year test, ETAS, completETAS and hazETAS-BG pass all tests. The remaining models hazETAS-mag, hazETAS-full and caliETAS all fail the number test. Their low quantile scores indicate that the failure is caused by a severe overestimation of the expected event number with these three models.

hazETAS-mag and hazETAS-full fail the magnitude test. They use b -values obtained from the SUIhaz2015 model which are much lower than what is estimated from a pure GR law used by the other models. b -values are known to be biased when they are calculated based on declustered catalogs (Mizrahi et al., 2021b). That hazETAS-mag and hazETAS-full fail the magnitude test is therefore not unforeseen. Lower b -values lead to a higher expected number of large events. In the ETAS-based models, this leads to a higher number of aftershock sequences and thus to an overall higher number of events, which explains why the two models overestimate the event number.

In the case of caliETAS, the overestimation of the event rate most likely stems from its high branching ratio of 0.84 (see Table 4.1), combined with a background rate that was inferred from SUIhaz2015 and which is not meant to be complemented with aftershock sequences.

Figure 4.6 shows that the model pairs ETAS and hazETAS-BG, and hazETAS-mag

	30 years				50 years			
	N	S	M	P	N	S	M	P
ETAS	0.63	0.41	0.43	0.43	1.00	0.91	0.33	0.02
compleTAS	0.83	0.48	0.30	0.23	1.00	0.94	0.21	0.00
hazETAS-BG	0.63	0.38	0.43	0.43	1.00	1.00	0.33	0.03
hazETAS-mag	0.00	0.50	0.99	1.00	0.00	0.86	0.99	1.00
hazETAS-full	0.00	0.56	0.99	1.00	0.01	1.00	0.99	1.00
caliETAS	0.02	0.48	0.94	0.99	0.13	1.00	0.95	0.97

Figure 4.5: Results of the pyCSEP consistency tests for the considered models. For a 30 year (left) and 50 year (right) testing period, the quantile score of the observed catalog statistic among the catalog statistics of simulated catalogs is given for the number (N), spatial (S), magnitude (M), and pseudolikelihood (P) test. Cell color indicates whether a model passed (green) or failed (red) the test.

and hazETAS-full exhibit the same behavior in the magnitude test. This is expected as they only differ from each other in the spatial distribution of background seismicity, which does not affect the magnitude distribution. Interestingly, caliETAS, which uses the same magnitude distribution as ETAS and hazETAS-BG, behaves differently in the test. This is an artifact of the test and can be explained by the much larger number of events forecasted by caliETAS compared to ETAS or hazETAS-BG, which forecast the same number of events. In the magnitude test, the test statistic quantifies how different the magnitude distribution of an individual catalog is from the overall magnitude distribution of the 100,000 simulated catalogs. Less deviation from the underlying GR law is expected when simulating many events instead of few. At the same time, the test statistic of the observed catalog remains the same, as it is comparing the observed magnitude distribution to the same underlying distribution.

The analogous figures to Figure 4.6 for the remaining tests can be found in the Supporting Information for this chapter (Figures C.4 - C.10).

For the 50 year test, all models except for caliETAS fail the number test. The quantile scores in Figure 4.5 show that ETAS, compleTAS and hazETAS-BG fail the test for opposite reasons than hazETAS-mag and hazETAS-full. The former three underestimate the number of events, while the latter two overestimate it. While caliETAS passes the test, the observed number of events is also considered to be at the lower end of expected event numbers. The 50 year observed catalog which spans the years 1942-1992 contains the Sierre 1946 M5.8 event. The entire ECOS-09 catalog contains only 8 events of this size or larger in our region of consideration, where the first one is a M6 event in the year 250. Although we cannot claim the catalog to be complete at this threshold since the year 250, the small number observed of events of this size in Switzerland indicates that the presence of the Sierre 1946 event in our 50 year testing catalog is relatively exceptional. In this sense, it is more plausible that the observed number of events is among the highest expected rather than being at the lower end of expectations.

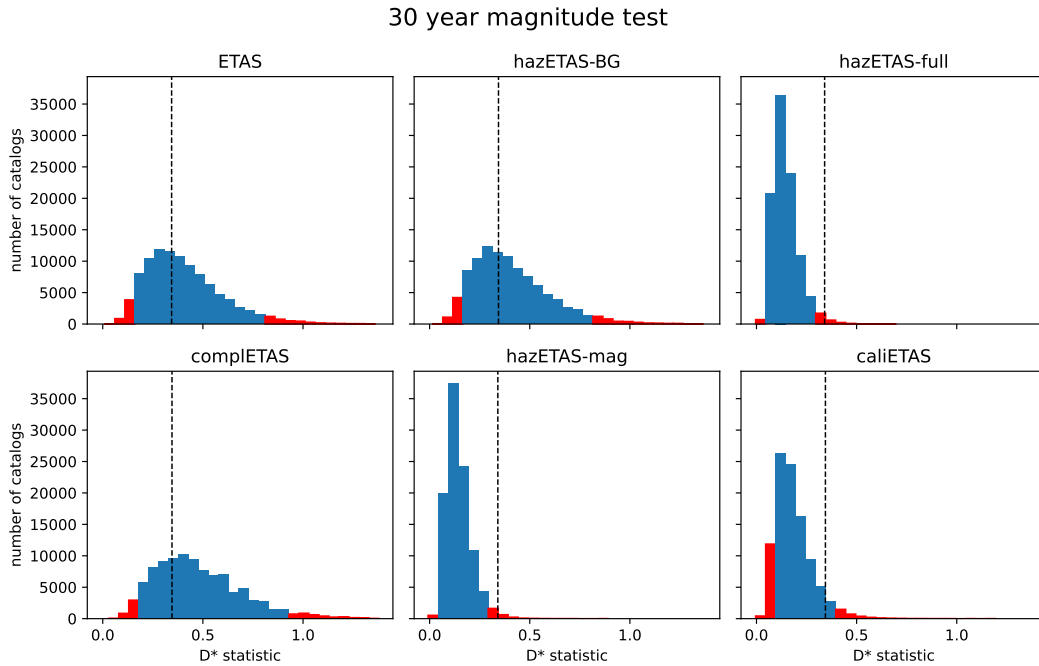


Figure 4.6: Results of the 30 year magnitude test. Histograms show distribution of the test statistic in 100,000 simulated catalogs, vertical dashed line indicates the observed statistic. Red areas of the histograms mark values below the 5th or above the 95th percentile for which the test is failed.

Based on these results, we decide to reject hazETAS-mag, hazETAS-full, and caliETAS. Among the remaining models, ETAS and completETAS are the two models with the highest number of six out of eight passed tests. They fail the number test in the 50 year case because they consider the observed catalog an exceptionally eventful catalog, which may actually be the truth. As a consequence of this, they also fail the pseudolikelihood test. While ETAS and completETAS pass the spatial test in the 50 year case, hazETAS-BG fails it with a quantile score of 1.00. This means that the observed catalog is more localized than the simulated catalogs. As the only difference between ETAS and hazETAS-BG is the spatial distribution of background seismicity, this suggests that the additional feature of using SUIhaz2015 background seismicity locations does not improve the spatial distribution of the model. Compared to the hazETAS-BG model, which ranks third according to its number of five passed consistency tests, ETAS and completETAS have the additional benefit of simplicity; they are both pure ETAS models and do not require the additional information of SUIhaz2015. For this reason, we propose these two models as suitable candidate OEF models for Switzerland.

4.5.3 Scenarios

Table 4.3 shows earthquake probabilities in Switzerland following hypothetical events, calculated with the two recommended models ETAS and completETAS. Four different scenarios are investigated: How would earthquake probabilities evolve after an event of magnitude $M^* = 4.5, 5.0, 5.5,$ or 6.0 occurred at the location of the 1946 Sierre earthquake? The probabilities are given for the candidate models ETAS and completETAS for the first day, week, and month after the event occurred. From left to right, Table 4.3 shows the probability that an event of $M \geq M^* - 1.0$, $M \geq M^*$, or an even larger event of magnitude $M \geq M^* + 1.0$ occurs anywhere in the testing region.

Table 4.3: Earthquake probabilities following hypothetical events at the location of the 1946 Sierre earthquake.

M^*	days	$M \geq M^* - 1.0$			$M \geq M^*$			$M \geq M^* + 1.0$		
		1	7	30	1	7	30	1	7	30
4.5	ETAS	6.78%	15.12%	31.46%	0.69%	1.56%	3.54%	0.06%	0.15%	0.34%
	completETAS	6.58%	14.04%	28.45%	0.55%	1.30%	2.95%	0.05%	0.11%	0.23%
5.0	ETAS	3.11%	6.39%	12.93%	0.27%	0.60%	1.31%	0.03%	0.06%	0.14%
	completETAS	3.13%	6.06%	11.52%	0.29%	0.55%	1.05%	0.02%	0.04%	0.08%
5.5	ETAS	1.47%	2.83%	5.30%	0.12%	0.23%	0.47%	0.01%	0.02%	0.05%
	completETAS	1.48%	2.72%	4.74%	0.13%	0.25%	0.43%	0.01%	0.03%	0.03%
6.0	ETAS	0.71%	1.31%	2.22%	0.07%	0.14%	0.23%	0.00%	0.01%	0.02%
	completETAS	0.74%	1.26%	2.04%	0.05%	0.08%	0.14%	0.00%	0.00%	0.01%

Generally, ETAS gives slightly higher probabilities than completETAS. As the magnitude M^* of the scenario earthquake increases, the chance of a subsequent event of similar size decreases. This is because $\beta = \ln 10 \cdot b$ is larger than the productivity exponent $a - \rho \cdot \gamma$ which describes the growth of the aftershock number with the magnitude of the triggering event (see Mizrahi et al., 2021a for details). To maintain similar probabilities of $M \geq M^*$ events as M^* increases, the number of aftershocks generated by an M^* event should increase by the same factor as the fraction of $M \geq M^*$ decreases according to the GR law. This would be the case for $a - \rho \cdot \gamma = \beta$.

4.5.4 Learnings for future model development

The result that the simplest models perform best in our tests raises the question of how to proceed with the development of future forecasting models. Which route to take to go beyond simple ETAS models?

All ETAS-based models perform similarly well in the pseudo-prospective tests, while no significant differences between the time-dependent models are observed. Only through a process of elimination using consistency tests the most suitable models can be identified. At the moment we are lacking criteria for selecting the ‘best’ model among the data-consistent ones. One possible reason for the similar performance of ETAS and completETAS in the pseudo-prospective tests is the relatively limited size of the testing dataset. In a few more years, we might be able to identify a clearer signal as to which model is to be preferred.

We may as well spend those years developing and testing further models. The results of the consistency tests clearly show that incorporating external data sources into ETAS models does not guarantee success. The lack of success of such models in our case may be due to the way in which the external data was incorporated. If additional information could be included in the model already during the parameter inversion process, the resulting model parameters could be more consistent with the newly included data. This would possibly produce models which can pass the consistency tests and simultaneously perform better in forecasting experiments.

The parameters of ETAS and completETAS differ substantially. This manifests in a 10% lower overall expected seismicity for completETAS compared to ETAS as quantified by Λ (see Equation (4.3) and Table 4.1). And while the main benefit of ETAS-based models over the SUIhaz2015 is their time dependency, the precision of the parameters describing the time-dependent behavior, namely aftershock triggering, does not seem to play a sufficiently crucial role to significantly determine the extent of their success. Another way of improving the time-dependent aspect the ETAS model besides parameter optimization is to address one of its current shortcomings: the time dependency of background seismicity.

In an ETAS sense, background seismicity is the remaining seismicity that could not be explained by triggering. Future background seismicity is modelled as the average unex-

plained seismicity of the training period. However, background seismicity can have strong long-term fluctuations or be influenced in the short term by human interference in the form of induced seismicity (Király-Proag et al., 2016), by Earth tides (Cochran et al., 2004; Métivier et al., 2009), or earthquake swarms (Hainzl and Fischer, 2002; Hainzl, 2004), which lack the burst-like behavior of typical aftershock sequences (see Zaliapin and Ben-Zion, 2013) and may be mixed into ETAS’s background seismicity. Recent advances in earthquake detection through deep learning have greatly improved the level of detail with which seismicity and in particular earthquake swarms or tidal earthquake triggering can be studied (Ross et al., 2019; Ross et al., 2020; Ross and Cochran, 2021; Wang et al., 2022). A continued increase in our understanding of the geophysical processes that influence seismicity, together with the results of Nandan, Ram, et al. (2021) that more accurate background seismicity models bear great potential for improved ETAS forecasts, suggests that next-generation earthquake forecasting models will no longer view background seismicity as the bulk of seismicity which was not triggered by previous events, and will capture more (space- and) time-dependency than “just” aftershock forecasting.

4.6 Conclusions

In this article, we describe the development and testing of six ETAS-based OEF candidate models for Switzerland and give our recommendation for two models to be used in the first Swiss OEF system.

We propose a systematic approach to choose the region, time horizon and magnitude range of the catalog used to calibrate an ETAS model. The approach is designed to require a minimum of choices influenced by personal preferences or subconscious biases, so that it can be applied to other regions without major modification.

Given the selected training catalog, we describe how an ETAS model can be calibrated, and possible catalog continuations which form the basis of a forecast can be simulated.

In pseudo-prospective forecasting experiments with five years of consecutive 7-day testing periods, we find that all ETAS-based models outperform the time-independent null model with large IGPE values between 1.35 and 1.72.

Retrospective 30 year and 50 year consistency tests reveal that three of the six model variants drastically overestimate the number of earthquakes. One of them uses ETAS parameters calibrated on Californian data. This result highlights the importance of using regionally calibrated parameters when using the ETAS model. The other two models which overestimate earthquake rate are using b -values inferred from the SUIhaz2015 model. These b -values are lower than the ones used in the other model variants which are estimated directly from the undeclustered training catalog. This implies that using model-consistent b -values is crucial to avoid biased forecasts.

Among the remaining models, the two purely ETAS-based models (ETAS and compleTAS, which is calibrated using the method of Mizrahi et al. (2021a) that allows a time-dependent magnitude of completeness) pass the highest number of consistency tests. No significant difference can be detected between the performance of the two models neither in the pseudo-prospective nor in the retrospective tests. For this reason, we propose both of them as the ideal candidates for the first operational earthquake forecasting system in Switzerland. A slight preference may be given to the most basic ETAS model for its simplicity.

A script which calibrates the Swiss ETAS model and runs the simulations is publicly available on [GitHub](#) (Mizrahi and Schmid, 2022, Appendix E).

Despite, or especially because the hybrid models which use ETAS and SUIhaz2015 information, were not able to outperform the more simple ETAS models in this experiment, we

believe that there lies great potential in ETAS-based forecasting models in which more effort goes into modeling what currently falls under the broad umbrella of background seismicity.

4.7 Data and resources

A summary of the Swiss National Seismic Hazard Maps 2015 can be viewed online on http://www.seismo.ethz.ch/export/sites/sedsite/knowledge/.galleries/pdf_knowledge/SUIhaz2015_final-report_16072016_2.pdf_2063069299.pdf (last accessed March 2022).

Acknowledgments

This study has been funded by the Eidgenössische Technische Hochschule (ETH) research grant for project number 2018-FE-213, “Enabling dynamic earthquake risk assessment (DynaRisk)” and the European Union’s Horizon 2020 research and innovation program under Grant Agreement Number 821115, real-time earthquake risk reduction for a resilient Europe (RISE). The authors thank Bill Savran for his support with the interpretation of the results of the pyCSEP consistency tests.

References

- Cattania, C., Werner, M. J., Marzocchi, W., Hainzl, S., Rhoades, D., Gerstenberger, M., Liukis, M., Savran, W., Christophersen, A., Helmstetter, A., et al. (2018). The forecasting skill of physics-based seismicity models during the 2010–2012 canterbury, new zealand, earthquake sequence. *Seismological Research Letters*, 89(4), 1238–1250. 68
- Cochran, E. S., Vidale, J. E., & Tanaka, S. (2004). Earth tides can trigger shallow thrust fault earthquakes. *Science*, 306(5699), 1164–1166. 85
- Cornell, C. A. (1968). Engineering seismic risk analysis. *Bulletin of the seismological society of America*, 58(5), 1583–1606. 68
- Demircioğlu, M. B., Şeşetyan, K., Duman, T. Y., Can, T., Tekin, S., & Ergintav, S. (2018). A probabilistic seismic hazard assessment for the turkish territory: Part ii—fault source and background seismicity model. *Bulletin of Earthquake Engineering*, 16(8), 3399–3438. 68
- Drouet, S., Ameri, G., Le Dortz, K., Secanell, R., & Senfaute, G. (2020). A probabilistic seismic hazard map for the metropolitan france. *Bulletin of Earthquake Engineering*, 18(5), 1865–1898. 68
- Edwards, B., Cauzzi, C., Danciu, L., & Fäh, D. (2016). Region-specific assessment, adjustment, and weighting of ground-motion prediction models: Application to the 2015 swiss seismic-hazard maps. *Bulletin of the Seismological Society of America*, 106(4), 1840–1857. 69
- Fäh, D., Giardini, D., Kästli, P., Deichmann, N., Gisler, M., Schwarz-Zanetti, G., Alvarez-Rubio, S., Sellami, S., Edwards, B., Allmann, B., et al. (2011). Ecos-09 earthquake catalogue of switzerland release 2011 report and database. public catalogue, 17. 4. 2011. swiss seismological service eth zurich. *RISK*. 69–71
- Field, E. H., Jordan, T. H., Page, M. T., Milner, K. R., Shaw, B. E., Dawson, T. E., Biasi, G. P., Parsons, T., Hardebeck, J. L., Michael, A. J., et al. (2017). A synoptic view of the third uniform california earthquake rupture forecast (ucerf3). *Seismological Research Letters*, 88(5), 1259–1267. 68
- Giardini, D., & sismologique suisse, S. (2004). *Seismic hazard assessment of switzerland, 2004*. Swiss Seismological Service: ETH. 69

- Gupta, A., Gupta, I. D., & Gupta, V. K. (2021). Probabilistic seismic hazard mapping of national capital region of india using a modified gridded seismicity model. *Soil Dynamics and Earthquake Engineering*, 144, 106632. 68
- Gutenberg, B., & Richter, C. F. (1944). Frequency of earthquakes in california. *Bulletin of the Seismological Society of America*, 34(4), 185–188. 72
- Hainzl, S. (2004). Seismicity patterns of earthquake swarms due to fluid intrusion and stress triggering. *Geophysical Journal International*, 159(3), 1090–1096. 85
- Hainzl, S. (2022). Etas-approach accounting for short-term incompleteness of earthquake catalogs. *Bulletin of the Seismological Society of America*, 112(1), 494–507. 71
- Hainzl, S., & Fischer, T. (2002). Indications for a successively triggered rupture growth underlying the 2000 earthquake swarm in vogtland/nw bohemia. *Journal of Geophysical Research: Solid Earth*, 107(B12), ESE–5. 85
- Hiemer, S., Woessner, J., Basili, R., Danciu, L., Giardini, D., & Wiemer, S. (2014). A smoothed stochastic earthquake rate model considering seismicity and fault moment release for europe. *Geophysical Journal International*, 198(2), 1159–1172. 69
- Irsyam, M., Cummins, P. R., Asrurifak, M., Faizal, L., Natawidjaja, D. H., Widiyantoro, S., Meilano, I., Triyoso, W., Rudiyanto, A., Hidayati, S., et al. (2020). Development of the 2017 national seismic hazard maps of indonesia. *Earthquake Spectra*, 36(1-suppl), 112–136. 68
- Kamer, Y., Nandan, S., Ouillon, G., Hiemer, S., & Sornette, D. (2021). Democratizing earthquake predictability research: Introducing the richterx platform. *The European Physical Journal Special Topics*, 230(1), 451–471. 68
- Király-Proag, E., Zechar, J. D., Gischig, V., Wiemer, S., Karvounis, D., & Doetsch, J. (2016). Validating induced seismicity forecast models—induced seismicity test bench. *Journal of Geophysical Research: Solid Earth*, 121(8), 6009–6029. 85
- Mancini, S., Segou, M., Werner, M., & Cattania, C. (2019). Improving physics-based aftershock forecasts during the 2016–2017 central italy earthquake cascade. *Journal of Geophysical Research: Solid Earth*, 124(8), 8626–8643. 68
- Mancini, S., Segou, M., Werner, M. J., & Parsons, T. (2020). The predictive skills of elastic coulomb rate-and-state aftershock forecasts during the 2019 ridgecrest, california, earthquake sequence. *Bulletin of the Seismological Society of America*, 110(4), 1736–1751. 68
- Marzocchi, W., Lombardi, A. M., & Casarotti, E. (2014). The establishment of an operational earthquake forecasting system in italy. *Seismological Research Letters*, 85(5), 961–969. 68
- Meletti, C., Marzocchi, W., Albarello, D., D’Amico, V., Luzi, L., Martinelli, F., Pace, B., Pignone, M., Rovida, A. N., Visini, F., et al. (2017). The 2016 italian seismic hazard model. *16th World Conference on Earthquake Engineering*. 68
- Métivier, L., de Viron, O., Conrad, C. P., Renault, S., Diament, M., & Patau, G. (2009). Evidence of earthquake triggering by the solid earth tides. *Earth and Planetary Science Letters*, 278(3-4), 370–375. 85
- Mizrahi, L., Nandan, S., & Wiemer, S. (2021a). Embracing data incompleteness for better earthquake forecasting. *Journal of Geophysical Research: Solid Earth*, 126(12), e2021JB022379. <https://doi.org/10.1029/2021JB022379>. 71–74, 76, 77, 84, 85
- Mizrahi, L., Nandan, S., & Wiemer, S. (2021b). The Effect of Declustering on the Size Distribution of Mainshocks. *Seismological Research Letters*. <https://doi.org/10.1785/0220200231>. 69, 71–73, 77, 78, 81
- Mizrahi, L., & Schmid, N. (2022). *Lmizrahi/etas* (Version 3.1). Zenodo. <https://doi.org/10.5281/zenodo.6951562>. 72, 74, 85
- Mosca, I., Sargeant, S., Baptie, B., Musson, R., & Pharaoh, T. (2022). The 2020 national seismic hazard model for the united kingdom. *Bulletin of Earthquake Engineering*, 1–43. 68

- Nandan, S., Kamer, Y., Ouillon, G., Hiemer, S., & Sornette, D. (2021). Global models for short-term earthquake forecasting and predictive skill assessment. *The European Physical Journal Special Topics*, *230*(1), 425–449. [68](#), [74](#)
- Nandan, S., Ouillon, G., Sornette, D., & Wiemer, S. (2019a). Forecasting the full distribution of earthquake numbers is fair, robust, and better. *Seismological Research Letters*, *90*(4), 1650–1659. [77](#)
- Nandan, S., Ouillon, G., Sornette, D., & Wiemer, S. (2019b). Forecasting the rates of future aftershocks of all generations is essential to develop better earthquake forecast models. *Journal of Geophysical Research: Solid Earth*, *124*(8), 8404–8425. [68](#)
- Nandan, S., Ouillon, G., Wiemer, S., & Sornette, D. (2017). Objective estimation of spatially variable parameters of epidemic type aftershock sequence model: Application to california. *Journal of Geophysical Research: Solid Earth*, *122*(7), 5118–5143. [72](#)
- Nandan, S., Ram, S. K., Ouillon, G., & Sornette, D. (2021). Is seismicity operating at a critical point? *Physical Review Letters*, *126*(12), 128501. [85](#)
- Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, *83*(401), 9–27. [68](#)
- Petersen, M. D., Mueller, C. S., Moschetti, M. P., Hoover, S. M., Rukstales, K. S., McNamara, D. E., Williams, R. A., Shumway, A. M., Powers, P. M., Earle, P. S., et al. (2018). 2018 one-year seismic hazard forecast for the central and eastern united states from induced and natural earthquakes. *Seismological Research Letters*, *89*(3), 1049–1061. [68](#)
- Rhoades, D., Liukis, M., Christophersen, A., & Gerstenberger, M. (2016). Retrospective tests of hybrid operational earthquake forecasting models for canterbury. *Geophysical Journal International*, *204*(1), 440–456. [68](#)
- Ross, Z. E., & Cochran, E. S. (2021). Evidence for latent crustal fluid injection transients in southern california from long-duration earthquake swarms. *Geophysical Research Letters*, *48*(12), e2021GL092465. [85](#)
- Ross, Z. E., Cochran, E. S., Trugman, D. T., & Smith, J. D. (2020). 3d fault architecture controls the dynamism of earthquake swarms. *Science*, *368*(6497), 1357–1361. [85](#)
- Ross, Z. E., Trugman, D. T., Hauksson, E., & Shearer, P. M. (2019). Searching for hidden earthquakes in southern california. *Science*, *364*(6442), 767–771. [85](#)
- Savran, W. H., Werner, M. J., Marzocchi, W., Rhoades, D. A., Jackson, D. D., Milner, K., Field, E., & Michael, A. (2020). Pseudoprospective evaluation of ucerf3-etas forecasts during the 2019 ridgecrest sequence. *Bulletin of the Seismological Society of America*, *110*(4), 1799–1817. [74](#), [78](#)
- Savran, W. H., Werner, M. J., Schorlemmer, D., & Maechling, P. J. (2022). Pycsep: A python toolkit for earthquake forecast developers. *Journal of Open Source Software*, *7*(69). [78](#)
- Sesetyan, K., Demircioglu, M. B., Duman, T. Y., Can, T., Tekin, S., Azak, T. E., & Fercan, Ö. Z. (2018). A probabilistic seismic hazard assessment for the turkish territory—part i: The area source model. *Bulletin of Earthquake Engineering*, *16*(8), 3367–3397. [68](#)
- Sornette, D., & Werner, M. J. (2005). Constraints on the size of the smallest triggering earthquake from the epidemic-type aftershock sequence model, bath’s law, and observed aftershock sequences. *Journal of Geophysical Research: Solid Earth*, *110*(B8). [72](#)
- Stirling, M., Shaw, B., Fitzgerald, M., Ross, C., & Gerstenberger, M. (2021). The national seismic hazard model for new zealand: Contributions from the deep south. [68](#)
- Taroni, M., Marzocchi, W., Schorlemmer, D., Werner, M. J., Wiemer, S., Zechar, J. D., Heiniger, L., & Euchner, F. (2018). Prospective csep evaluation of 1-day, 3-month, and 5-yr earthquake forecasts for italy. *Seismological Research Letters*, *89*(4), 1251–1261. [68](#)
- Tinti, S., & Mulargia, F. (1987). Confidence intervals of b values for grouped magnitudes. *Bulletin of the Seismological Society of America*, *77*(6), 2125–2134. [75](#)

- Veen, A., & 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. [72](#)
- Wang, W., Shearer, P. M., Vidale, J. E., Xu, X., Trugman, D. T., & Fialko, Y. (2022). Tidal modulation of seismicity at the coso geothermal field. *Earth and Planetary Science Letters*, *579*, 117335. [85](#)
- Wiemer, S., Giardini, D., Fäh, D., Deichmann, N., & Sellami, S. (2009). Probabilistic seismic hazard assessment of switzerland: Best estimates and uncertainties. *Journal of Seismology*, *13*(4), 449. [68](#)
- Woessner, J., Hainzl, S., Marzocchi, W., Werner, M., Lombardi, A., Catalli, F., Enescu, B., Cocco, M., Gerstenberger, M., & Wiemer, S. (2011). A retrospective comparative forecast test on the 1992 landers sequence. *Journal of Geophysical Research: Solid Earth*, *116*(B5). [68](#)
- Woessner, J., Laurentiu, D., Giardini, D., Crowley, H., Cotton, F., Grünthal, G., Valensise, G., Arvidsson, R., Basili, R., Demircioglu, M. B., et al. (2015). The 2013 european seismic hazard model: Key components and results. *Bulletin of Earthquake Engineering*, *13*(12), 3553–3596. [68](#), [69](#)
- Zaliapin, I., & Ben-Zion, Y. (2013). Earthquake clusters in southern california ii: Classification and relation to physical properties of the crust. *Journal of Geophysical Research: Solid Earth*, *118*(6), 2865–2877. [85](#)

CHAPTER 5

QUESTION-DRIVEN ENSEMBLES OF FLEX-
IBLE ETAS MODELS

Under review at Seismological Research Letters as:

Mizrahi, L., Nandan, S., Savran, W., Wiemer, S. & Ben-Zion, Y.
(2022). Question-driven Ensembles of Flexible ETAS Models.
<https://arxiv.org/pdf/2207.06247.pdf>

Abstract

The development of new earthquake forecasting models is often motivated by one of the following complementary goals: to gain new insights into the governing physics and to produce improved forecasts quantified by objective metrics. Often, one comes at the cost of the other. Here, we propose a question-driven ensemble (QDE) modeling approach to address both goals. We first describe flexible ETAS models in which we relax the assumptions of parametrically defined aftershock productivity and background earthquake rates during model calibration. Instead, both productivity and background rates are calibrated with data such that their variability is optimally represented by the model. Then we consider 64 QDE models in pseudo-prospective forecasting experiments for Southern California and Italy. QDE models are constructed by combining model parameters of different ingredient models, where the rules for how to combine parameters are defined by questions about the future seismicity. The QDE models can be interpreted as models which address different questions with different ingredient models. We find that certain models best address the same issues in both regions, and that QDE models can substantially outperform the standard ETAS and all ingredient models. The best performing QDE model is obtained through the combination of models allowing flexible background seismicity and flexible aftershock productivity, respectively, where the former parameterizes the spatial distribution of background earthquakes and the partitioning of seismicity into background events and aftershocks, and the latter is used to parameterize the spatio-temporal occurrence of aftershocks.

5.1 Introduction

Earthquake forecasting is one of the defining problems of seismology. To provide useful solutions, forecasting models use a wide range of approaches: Coulomb rate-and-state (CRS) models (Cocco et al., 2010; Parsons et al., 2012; Mancini et al., 2019) calculate Coulomb stress changes and couple them with a lab-based constitutive friction law (Dieterich, 1994). On the other end of the spectrum are statistical models, with the Epidemic-Type Aftershock Sequence (ETAS) model being the best performing current statistical approach (Cattania et al., 2018; Taroni et al., 2018). First introduced by Ogata (1988), it models seismicity rate as the sum of background and aftershock events, where aftershocks are triggered according to regional empirical laws. In-between the purely physics-based and purely statistics-based approaches are models such as the short-term earthquake probability (STEP) model (M. C. Gerstenberger et al., 2005), the Inlabru model (Bayliss et al., 2020) and hybrid Coulomb/statistical models (Steady et al., 2014). The STEP model combines clustering principles with fault information in a statistical model to produce time-dependent forecasts. The Inlabru model more generally allows the inclusion of diverse data sets as covariates to issue time-independent seismicity forecasts. A hybrid Coulomb/statistical model redistributes seismicity forecasted by STEP according to Coulomb stress changes.

While physics-based models aim to describe the processes and mechanisms underlying seismogenesis, statistical models are generally more empirical and data-driven. Ultimately, “all models are wrong, but some are useful”, to cite the famous statistician George Box (1979). Usefulness can be viewed from different perspectives. Different forecasting models can be useful for gaining new scientific insight, for producing the most accurate forecasts, or for producing forecasts that are most suited for operational earthquake forecasting (OEF), given the trade-off between accuracy and computational cost. Cattania et al. (2018) found in a pseudo-prospective forecasting experiment for the 2010-2012 Canterbury, New Zealand earthquake sequence that hybrid Coulomb/statistical models have a similar forecasting skill as CRS models, at a lower computational effort. Mancini et al. (2019) and Mancini et al. (2020) conducted pseudo-prospective experiments for the 2016 central Italy and the 2019 Ridgecrest, California sequences, comparing CRS models of different complexity with ETAS forecasts. In both studies, the forecasting skill of CRS models increases with their

complexity, with the most complex CRS model performing similarly to ETAS. Hardebeck (2021) investigated possible reasons for the general underperformance of the physics-based models relative to statistical models and suggested that understanding and incorporating heterogeneities in background conditions into physical forecasting models may be key to improving their skill.

Having been tested thoroughly and systematically (Woessner et al., 2011; Ogata et al., 2013; Strader et al., 2017; Taroni et al., 2018; Nandan, Ouillon, Sornette, and Wiemer, 2019b; Savran et al., 2020), ETAS models meanwhile remain the state-of-the-art of earthquake forecasting and are being used or considered for OEF at various locations (Marzocchi et al., 2014; D. Rhoades et al., 2016; Field et al., 2017; Nandan, Kamer, et al., 2021; Kamer et al., 2021; van der Elst et al., 2022). Besides using the most basic formulation of ETAS, modelers also commonly refine the model. For instance, Bach and Hainzl (2012) enhanced ETAS with fault information, ShakeMaps, ground motion models, or Coulomb stress changes. Seif et al. (2017) assessed the biasing effects of data incompleteness and model assumptions on the estimated ETAS parameters. Several techniques have been proposed to address the effects of short-term aftershock incompleteness (Mizrahi et al., 2021b; Hainzl, 2022; Grimm et al., 2022) or the assumption of isotropic aftershock triggering (Grimm et al., 2022; Page and van der Elst, 2022). Other studies focus on deriving spatial variations of ETAS parameters or background seismicity (Nandan et al., 2017; Nandan, Ram, et al., 2021; Enescu et al., 2009), also relating parameter variations with physical quantities such as heat flow. Others have refined the standard ETAS model with a relationship between magnitudes of triggered and triggering earthquakes and a magnitude-dependent Omori kernel and found the resulting models to possess improved forecasting performance (Nandan, Kamer, et al., 2021; Nandan, Ouillon, and Sornette, 2019). A recent framework for modeling seismicity with an invariant Galton–Watson stochastic branching process provides a generalization of ETAS that is invariant with respect to various common deficiencies of earthquake catalogs (Kovchegov et al., 2022). However, this framework has not yet been used for forecasting seismicity.

A related forecasting topic which has recently received attention is ensemble modeling (D. A. Rhoades and Gerstenberger, 2009; Marzocchi et al., 2012; Taroni et al., 2014; Bird et al., 2015; Akinci et al., 2018; Llenos and Michael, 2019; Bayona et al., 2021). The idea, widely used for decades in the meteorological and climate forecasting community (Tracton and Kalnay, 1993; Leutbecher and Palmer, 2008; Eyring et al., 2016), is to combine different models in an overarching ensemble model to obtain more robust forecasts. Commonly, an ensemble is a linear or multiplicative combination of ingredient models (e.g. Bird et al., 2015), and the challenge is to optimize the weights given to each model. In a recent study, Bayona et al. (2021) found that the time-independent ensemble models WHEEL and GREAR1 (Bird et al., 2015) outperform the ingredient models of which they consist. Akinci et al. (2018) found that their time-independent ensemble model outperforms its ingredients and performs similarly to the best-performing time-independent model tested in the 2009 CSEP experiment (Zechar et al., 2010; Schorlemmer, Zechar, et al., 2010) for Italy. In the context of time-dependent models, Taroni et al. (2014) and M. Gerstenberger et al. (2014) used ensemble approaches, and Llenos and Michael (2019) found that ensembles of ETAS models perform best for the 2015 San Ramon, California Swarm. Shebalin et al. (2014) proposed an iterative method to combine forecasting models and found the resulting models to have advantageous properties compared to the ingredient models or traditional linear combinations thereof. The emerging consensus across the mentioned studies is that ensemble modeling is a promising path to use for earthquake forecasting; this is also demonstrated by the fact that they are currently implemented in Italy’s OEF system (Marzocchi et al., 2014). Yet, a breakthrough of ensemble models as established in the meteorological forecasting community is still pending.

For practical operational forecasting, especially in regions that are less studied due to a lack of data or resources, a balance must be achieved between model accuracy and simplicity. With this in mind, we relax some of the assumptions behind ETAS. We allow aftershock productivity and background seismicity to be described non-parametrically, providing event-specific productivity and background rates. This aims to better capture the real behavior of seismicity without making any choices on resolution, parametric form, etc. Using pseudo-prospective forecasting experiments in Southern California and Italy, we evaluate whether these flexible ETAS (fETAS) models provide superior forecasts.

We also propose a novel approach for question-driven ensemble (QDE) modeling, fundamentally different from traditional ensemble modeling approaches. In the QDE approach, models are combined in the parameter space as opposed to the solution space. Several ETAS-like models are fit to the observed data, yielding an individual set of parameters for each model. A QDE model is then created by defining a new set of parameters based on a combination of the ingredient model parameters. The rules to combine parameters are defined by dividing the forecasting problem into several sub-problems. Each sub-problem addresses a question regarding the number of forecasted events or the spatio-temporal distribution of either background earthquakes or aftershocks. A QDE model can be viewed as a model which addresses different questions with different ingredient models. This approach allows the combination of ETAS variants but can be extended to combining more general types of seismicity models.

By including such QDE models in the forecasting experiments, we assess their forecasting capability in comparison with their ingredient models, standard ETAS and fETAS. At the same time, the QDE approach helps to understand which ingredient models are best suited to solve different forecasting sub-problems, thus, making it useful from the perspective of gaining new scientific insight.

The remainder of this paper is structured as follows. We describe fETAS models and the QDE approach in Section 5.2. The setup for the forecasting experiments, the data analyzed and the metrics used to evaluate forecasting performance are described in Section 5.3. We present and discuss our results in Section 5.4 and finally provide our conclusions in Section 5.5.

5.2 Flexible ETAS models

The following sub-sections describe flexible ETAS models and explain the question-driven ensemble modeling. We begin by explaining the algorithm used to estimate the parameters of the ETAS model. Then, we describe how to relax some parametric assumptions of the ETAS model. Finally, we introduce a framework for question-driven ensemble modeling of flexible ETAS models.

5.2.1 Expectation Maximization algorithm

Consider an earthquake catalog

$$C = \{e_i = (m_i, t_i, x_i, y_i), i \in \{1, \dots, n\}\} \quad (5.1)$$

consisting of events e_i of magnitudes m_i which occur at times t_i and locations (x_i, y_i) .

The ETAS model describes earthquake rate as

$$\lambda(t, x, y | \mathcal{H}_t) = \mu + \sum_{i: t_i < t} g(m_i, t - t_i, x - x_i, y - y_i). \quad (5.2)$$

That is, the sum of background rate μ and the rate of all aftershocks of previous events e_i . The aftershock triggering rate $g(m, \Delta t, \Delta x, \Delta y)$ describes the rate of aftershocks triggered

by an event of magnitude m , at a time delay of Δt and a spatial distance $(\Delta x, \Delta y)$ from the triggering event. We use here the definition

$$g(m, \Delta t, \Delta x, \Delta y) = \frac{k_0 \cdot e^{a(m-m_{ref})} \cdot e^{-\Delta t/\tau}}{((\Delta x^2 + \Delta y^2) + d \cdot e^{\gamma(m-m_{ref})})^{1+\rho} \cdot (\Delta t + c)^{1+\omega}}, \quad (5.3)$$

as in Nandan, Kamer, et al. (2021) and Mizrahi et al. (2021a).

To calibrate the ETAS model, the nine parameters to be optimized are the background rate μ and $k_0, a, c, \omega, \tau, d, \gamma, \rho$, which parameterize the aftershock triggering rate $g(m, t, x, y)$ given in Equation (5.3). Implicitly, the model assumes that only earthquakes with magnitudes larger than or equal to m_{ref} can trigger aftershocks. Most applications of the method define m_{ref} as equal to the constant value of m_c .

We build on the expectation maximization (EM) algorithm to estimate the ETAS parameters (Veen and Schoenberg, 2008). In this algorithm, the expected number of background events \hat{n} and the expected number of directly triggered aftershocks \hat{l}_i of each event e_i are estimated in the expectation step (E step), along with the probabilities p_{ij} that event e_j was triggered by event e_i , and the probability p_j^{ind} that event e_j is independent. Following the E step, the nine parameters are optimized to maximize the complete data log-likelihood in the maximization step (M step). E and M steps are repeated until convergence of the parameters. The usual formulation of the EM algorithm defines

$$\hat{n} = \sum_j p_j^{ind}, \quad (5.4)$$

$$\hat{l}_i = \sum_j p_{ij}, \quad (5.5)$$

and

$$p_{ij} = \frac{g_{ij}}{\mu + \sum_{k:t_k < t_j} g_{kj}}, \quad (5.6)$$

$$p_j^{ind} = \frac{\mu}{\mu + \sum_{k:t_k < t_j} g_{kj}}, \quad (5.7)$$

with $g_{kj} = g(m_k, t_j - t_k, x_j - x_k, y_j - y_k)$ being the aftershock triggering rate of e_k at location and time of event e_j . For a given target event e_j , Equations (5.6-5.7) define p_{ij} to be proportional to the aftershock occurrence rate g_{ij} , and p_j^{ind} to be proportional to the background rate μ . As an event must be either independent or triggered by a previous event, the normalization factor $\Lambda_j := \mu + \sum_{k:t_k < t_j} g_{kj}$ in the denominator of Equations (5.6-5.7) stipulates that $p_j^{ind} + \sum_{k:t_k < t_j} p_{kj} = 1$.

5.2.2 Introducing flexibility

In the above formulation of the ETAS model, the the rate of background earthquakes is described by the parameter μ , which does not vary with space nor time. During the maximization step of the EM algorithm, μ can be estimated independently from the other parameters as

$$\mu = \frac{\hat{n}}{A_R \cdot T}, \quad (5.8)$$

where A_R and T denote the area of the study region and the length of the considered time window, respectively. In some approaches, the region of interest is divided into several

sub-regions which can have their own values for μ (Veen and Schoenberg, 2008). An iterative algorithm to estimate spatial variations of background rate based on maximum likelihood estimation used a Gaussian kernel smoothing (Zhuang, 2012) to the catalog event locations, weighted by their estimated independence probability, to obtain an estimate of $\mu(x, y)$. Here, we present a similar approach using expectation maximization, which has been shown to be more stable with respect to the initial conditions compared to maximum likelihood approaches (Veen and Schoenberg, 2008). Our approach is similar yet not identical to the one described by Nandan, Ram, et al. (2021) which uses a regularized inverse power law for smoothing the locations. We define the background rate at a location (x, y) as

$$\mu(x, y) = \frac{1}{T} \cdot \sum_j p_j^{ind} \cdot k(\Delta x_j, \Delta y_j), \quad (5.9)$$

where $k(\Delta x_j, \Delta y_j)$ is the Gaussian kernel with bandwidth σ applied to the distance $(\Delta x_j, \Delta y_j)$ of event e_j to the location (x, y) ,

$$k(\Delta x, \Delta y) = \frac{1}{2\pi\sigma^2} \cdot \exp\left(-\frac{1}{2} \cdot \frac{\Delta x^2 + \Delta y^2}{\sigma^2}\right). \quad (5.10)$$

The bandwidth σ determines the smoothness of the background event density. In principle, σ could be calibrated itself, but we choose to fix it to $5km$ for simplicity. Our next modification to the standard ETAS model is to allow flexibility of the aftershock probability. The number of directly triggered aftershocks \hat{l}_j is estimated during the expectation step of the EM algorithm as described in Equation (5.5). We can thus replace the term $k_0 \cdot e^{a(m-m_{ref})}$ in Equation (5.3) with κ_j , where κ_j is stipulated to be proportional to \hat{l}_j . Instead of parameterizing aftershock productivity to be exponentially increasing with the magnitude of the triggering event, we allow each event to have its own productivity. This yields

$$g_{j\theta, \kappa_j}(m, \Delta t, \Delta x, \Delta y) = \frac{\kappa_j \cdot e^{\Delta t/\tau}}{\left((\Delta x^2 + \Delta y^2) + d \cdot e^{\gamma(m-m_{ref})}\right)^{1+\rho} \cdot (\Delta t + c)^{1+\omega}} \quad (5.11)$$

for given parameters $\theta = (c, \omega, \tau, d, \gamma, \rho)$ and κ_j . The EM algorithm is adapted as follows:

1. Define initial estimates of κ_j as $\kappa_j = e^{a(m_j - m_{ref})}$ with a random guess for a .
2. Define initial estimates of independence probability $p_j^{ind} \equiv 0.1$. The inversion result is not sensitive to this choice.
3. Define random initial guesses for the parameters $\theta = (c, \omega, \tau, d, \gamma, \rho)$.
4. Expectation Step: Calculate $\hat{n}, \hat{l}_j, p_{ij}, p_j^{ind}$ using the current estimates of κ_j, θ , and p_j^{ind} . p_{ij}, p_j^{ind} are calculated using Equations (5.6-5.7), but using the flexible definitions of g_{ij} and $\mu(x, y)$ of Equations (5.9) and (5.11).
5. Maximization Step: Optimize the parameters θ to minimize the complete data log likelihood (see Mizrahi et al. (2021a) for details), given the current estimates of $\hat{n}, \hat{l}_j, p_{ij}, p_j^{ind}$.
6. Update κ_j^{new} to be $\kappa_j^{old} \cdot \frac{\hat{l}_j}{G_{j\theta, \kappa_j^{old}}}$, where $G_{j\theta, \kappa_j^{old}}$ is the expected total number of aftershocks of e_j , given θ and κ_j^{old} . This ensures that $\hat{l}_j = G_{j\theta, \kappa_j^{new}}$. We calculate $G_{j\theta, \kappa_j}$ as

$$G_{j\theta, \kappa_j} = \iint_R \int_0^{t_{end}-t_j} g_{j\theta, \kappa_j}(m_j, t, x, y) dt dx dy, \quad (5.12)$$

where t_{end} is the end time of the considered time window, and we assume the spatial region R to extend infinitely in space, allowing a facilitated, asymptotically unbiased estimation of ETAS parameters (Schoenberg, 2013).

7. Repeat from 4 until convergence of θ , i.e. until $\sum_{a_i \in \theta} |a_i^{new} - a_i^{old}| < 10^{-3}$.

After the inversion, we calibrate an overall productivity law for the fETAS models with free productivity to avoid over-fitting with event-wise productivity. From the individually estimated productivities κ_j of magnitude m_j events, we calibrate a law of the form

$$\kappa(m) = k_0 \cdot e^{a(m-m_{ref})} \quad (5.13)$$

by minimizing the sum of absolute residuals between the observed $\bar{\kappa}(m) = \frac{1}{n(m)} \sum_{i:m_i=m} \kappa_i$ and the theoretical $\kappa(m) = k_0 \cdot e^{a(m-m_{ref})}$, where $n(m)$ is the number of events with magnitude m .

Then, productivity is treated the same way as in the case of standard ETAS. In this way, the variability of productivity is only accounted for during the parameter inversion process and may lead to more accurate estimators of the productivity as well as the remaining ETAS parameters.

5.2.3 Question-driven ensemble (QDE) modeling

We propose a novel approach for question-driven ensemble (QDE) modelling, where a forecast is created by combining model parameters of different ingredient models. The rules for how parameters can be combined are defined by questions which divide the forecasting problem into several sub-problems: *How many background events are expected? Where are they expected? When are they expected? How many aftershocks are expected? Where are they expected? When are they expected?*

By answering each of these questions with different ingredient models, we create a suite of ensembles. The remainder of this section establishes rules to combine parameters based on the questions.

Consider a collection of ETAS or fETAS ingredient models, $(M_i)_{i=0, \dots, n_M}$. As they are sufficiently defined through their parameters, we can write

$$M_i = (\mu_i, \kappa_i, c_i, \omega_i, \tau_i, d_i, \gamma_i, \rho_i). \quad (5.14)$$

In case M_i is a fETAS model, $\mu_i = \mu_i(x, y)$ can vary with space. For simplicity, we denote with κ_i the function which assigns to each event its appropriate value to replace the term κ_j in Equation (5.11). In our case, this means that we define $\kappa_i(m) = k_{0_i} \cdot e^{a_i(m-m_{ref})}$, where k_{0_i} and a_i are either obtained during parameter inversion directly, or afterwards in case M_i is a fETAS model with free productivity. We chose the notation of κ_i instead of (k_{0_i}, a_i) to emphasize this possible distinction. We can then generally describe the aftershock triggering kernel g as

$$g_i(m, \Delta t, \Delta x, \Delta y) = \frac{\kappa_i \cdot e^{\Delta t/\tau_i}}{((\Delta x^2 + \Delta y^2) + d_i \cdot e^{\gamma_i(m-m_{ref})})^{1+\rho_i} \cdot (\Delta t + c_i)^{1+\omega_i}}. \quad (5.15)$$

Let us now revisit the questions above.

1. *How many background events are expected?*

More precisely, what we want to ask here is how many background events do we expect in total in the region R and forecasting horizon $[T_0, T_1]$ we are issuing a forecast for. The answer to this question, given out of the perspective of model M_i , is

$$N_{B_i} = \iint_R \int_{T_0}^{T_1} \mu_i(x, y) dt dx dy. \quad (5.16)$$

2. *Where and when are they expected?*

We address for now these two questions jointly. The spatio-temporal density of background events is given by

$$f_{B_i}(x, y, t) = \frac{\mu_i(x, y)}{\iint_R \int_{T_0}^{T_1} \mu_i(x, y) dt dx dy} = \frac{\mu_i(x, y)}{N_{B_i}}, \quad (5.17)$$

which is effectively time-independent due to our choice of a time-independent $\mu(x, y)$.

3. *How many aftershocks are expected?*

Again, what we want to ask here is how many aftershocks do we expect in total in the region R and forecasting horizon $[T_0, T_1]$ we are issuing a forecast for. For an individual event e_j , we expect it to have n_A aftershocks, where

$$n_{A_i}(e_j) = \iint_R \int_{T_0}^{T_1} g_i(m_j, t - t_j, x - x_j, y - y_j) dt dx dy. \quad (5.18)$$

The total number of aftershocks N_{A_i} is then given as the sum of aftershocks of all events

$$N_{A_i} = \sum_{j:t_j < T_1} n_{A_i}(e_j). \quad (5.19)$$

4. *Where and when are they expected?*

We again answer these two questions jointly. If we define

$$G_i(x, y, t) := \sum_{j:t_j < T_1} g_i(m_j, t - t_j, x - x_j, y - y_j) \quad (5.20)$$

as the total rate of aftershocks at time t and location (x, y) , consisting of the sum of aftershock rates of all events that occurred prior to the end T_1 of the forecasting horizon, the spatio-temporal density of aftershocks is given by

$$f_{A_i}(x, y, t) = \frac{G_i(x, y, t)}{\iint_R \int_{T_0}^{T_1} G_i(x, y, t) dt dx dy} = \frac{G_i(x, y, t)}{N_{A_i}}. \quad (5.21)$$

We now construct a question-driven ensemble (QDE) model E^{klm} as follows. The number questions (1) and (3) are answered with model M_k , the background density question (2) is answered with model M_l , and the aftershock density question (4) is answered with model M_m . Note that questions (1) and (3) are addressed with the same model. This is a choice made to avoid unrealistic event numbers. If one model interprets the majority of events as background, and another model interprets the majority of events to be aftershocks, answering the two questions with two different models would lead to exceptionally high or low total event numbers, which is not intended by the two ingredient models.

In the notation above, which identifies a model with its parameters, this would give us

$$E^{klm} = (\mu_l \cdot \frac{N_{B_k}}{N_{B_l}}, \kappa_m \cdot \frac{N_{A_k}}{N_{A_m}}, c_m, \omega_m, \tau_m, d_m, \gamma_m, \rho_m). \quad (5.22)$$

5.3 Forecasting experiments

To test whether fETAS models and QDE models which consist of ETAS and fETAS models provide better forecasts, we conduct pseudo-prospective forecasting experiments for Southern California and Italy.

5.3.1 Competing models

In these experiments, we consider the following four competing ingredient models.

- M_0 : standard ETAS
- M_1 : fETAS with free productivity and standard background
- M_2 : fETAS with standard productivity and free background
- M_3 : fETAS with free productivity and free background

Out of these, $4^3 = 64$ QDE models can be constructed.

Note that M_2 is conceptually close to the models described by Zhuang (2012) and Nandan, Ram, et al. (2021).

5.3.2 Evaluation metric

We use interevent time horizons: Whenever an event occurs, a forecast is issued, which is valid until the occurrence of the next event. A pseudo-prospective model evaluation then aims to capture how well a forecast issued using data until event e_{j-1} can describe the occurrence of the next event e_j .

An ETAS forecast always consists of the forecasted background seismicity rate plus the forecasted aftershock seismicity rate. With this flexible definition of forecasting horizon, our ETAS forecast can be calculated and evaluated analytically.

Consider $\lambda_i(t, x, y | \mathcal{H}_{t_{j-1}})$, the event rate under model M_i as of time t_{j-1} of the $(j-1)^{th}$ earthquake. This formulation of λ_i is valid for times $t \in (t_{j-1}, t_j]$ between the occurrence of event e_{j-1} and event e_j , and hence this is the forecasting horizon we consider.

For the traditional experiment settings where one is interested in the seismicity forecast of the next days, months, or years, such an analytical description of the forecasted seismicity is not possible. As soon as an event occurs during the forecasting period, its aftershocks are not part of the background seismicity, nor of the aftershock seismicity that was calculated at the start of the forecasting period. For this reason, ETAS forecasts for fixed forecasting horizons are usually produced through the simulation of a large number of possible continuations of the catalog.

In our case of flexible forecasting horizons, the log likelihood of observing e_j under model M_i is analytically defined (see Ogata et al., 2013; Daley, Vere-Jones, et al., 2003) as

$$\ln \mathcal{L}_i(e_j) = \ln \lambda_i(t_j, x_j, t_j | \mathcal{H}_{t_{j-1}}) - \iint_R \int_{t_{j-1}}^{t_j} \lambda_i(t, x, t | \mathcal{H}_{t_{j-1}}) dt dx dy. \quad (5.23)$$

We then define the information gain $IG_j^{i_1, i_2}$ of model i_1 over model i_2 during the j^{th} forecasting period $(t_{j-1}, t_j]$ as

$$IG_j^{i_1, i_2} = \ln \frac{\mathcal{L}_{i_1}(e_j)}{\mathcal{L}_{i_2}(e_j)} = \ln \mathcal{L}_{i_1}(e_j) - \ln \mathcal{L}_{i_2}(e_j). \quad (5.24)$$

The information gain per event (IGPE) over forecasting periods j_1, \dots, j_K is defined as

$$\frac{1}{K} \sum_{k=1, \dots, K} IG_{j_k}^{i_1, i_2}, \quad (5.25)$$

the average of IGs over those testing periods.

Compared to evaluation techniques based on the simulation of large numbers of possible catalog continuations such as in Nandan, Ouillon, Sornette, and Wiemer (2019a) and Mizrahi et al. (2021a), which are encouraged by CSEP (see Savran et al., 2022), this approach allows us to compare models much faster, accelerating the development and testing process. To apply these models operationally, where forecasts are required for a fixed time horizon, simulations would still be required. This evaluation approach allows us to save time when developing and selecting the model to be used operationally, and is especially useful for evaluating a large suite of QDE models.

5.3.3 Data

For Southern California, we consider the ANSS comprehensive earthquake catalog (Com-Cat), in the polygon given by the vertices in Table D.1. We consider earthquakes of magnitude $M \geq 2.0$ from January 1, 2010 until January 1, 2022. The first two years serve as auxiliary period in the ETAS and fETAS parameter inversion, and thus the start of the primary catalog is January 1, 2012. This means that the events between January 2010 and January 2012 can act as triggering events during the inversion, but not as triggered events. Using the method described by Mizrahi et al. (2021b), we find that the overall catalog is complete at this threshold, although there are likely periods during which the catalog is incomplete due to short-term aftershock incompleteness (STAI). Although Mizrahi et al. (2021a) have proposed a method to account for STAI in the ETAS model, we do not address this issue here.

For Italy, we consider the Italian Seismological Instrumental and Parametric Data-Base catalog (ISIDe, Group, 2007), in the area defined for the first CSEP experiment (Schorlemmer, Christophersen, et al., 2010, vertices given in Table D.2). We consider earthquakes of magnitude $M \geq 2.5$ from April 16, 2005 until July 1, 2021. This is the time horizon available to modelers in the upcoming prospective CSEP forecasting experiment in Italy, and the estimated magnitude of completeness provided in the experiment description. The start of the primary catalog is January 1, 2010.

5.3.4 Experiment setting

For Southern California, we consider 5 years of testing, with the start of the first forecasting period at the occurrence of event e_0 , the first event at or after January 1, 2017. In Italy, we consider 3 years of testing, starting at the occurrence of the first event at or after July 1, 2018. The idea of the pseudo-prospective experiments is to only use data that would have been available at the time the forecast is issued to calibrate the models. One could thus re-calibrate the model at the start of each forecasting period, whenever one more event becomes part of the catalog. To limit the number of computationally expensive parameter inversions for these experiments, we re-estimate the model parameters every 7 days in Southern California, and every day in Italy, and use the latest available set of parameters at the start time of each forecasting interval. Note that this does not mean that events between the calibration time and forecasting start are ignored. Their aftershocks are still considered in the calculated aftershock rate. We chose a shorter parameter updating

interval for Italy to mimic the conditions of the CSEP experiment, and a longer one for Southern California to limit computational cost.

We then calculate $IG_j^{i_1, i_2}$ for all j , and for all pairs of models M_{i_1}, M_{i_2} . If the IGPE over all forecasting periods of one model to another is positive, we consider the model to produce superior forecasts.

As one could argue that generating a large number of models and then selecting the best performing ones somewhat invalidates the pseudo-prospective nature of our experiments, we consider the following additional model. At the start of the j^{th} forecasting period, the total information gain of all QDE models during the last n forecasting periods, i.e. periods $j - (n + 1)$ to $j - 1$, is compared. The model with the highest IG is selected to produce the forecast for the j^{th} forecasting period. We call this model QDE- S_n .

This type of model, if capable of producing a powerful forecast, would be well suited to be used in an OEF context.

5.4 Results and discussion

The parameters that were obtained using the fETAS inversion algorithm are described in Section D.2 in the Appendix. Here, we present the results of the forecasting experiments.

5.4.1 Experiment results

Figure 5.1 compares the information gain per earthquake (IGPE) over the standard ETAS null model ($M_0 = E^{000}$) of all 64 QDE models in Italy and Southern California. The IGPE varies between -0.64 and 0.45 in Italy, and between -0.13 and 0.12 in Southern California. The best and worst performing QDE models are E^{221} and E^{112} , respectively, for both regions. The best performing model E^{221} uses the free background model M_2 to answer the number and background density questions, and the free productivity model M_1 to answer the aftershock density question. Vice versa, the worst performing model E^{112} uses M_1 to answer the number and background density questions, and model M_2 to answer the aftershock density question. Generally, the models which perform well or poorly in Italy are also performing similarly in Southern California.

The symbol shape, fill color, and edge color in the scatter plot of Figure 5.1 represent the ingredient model used to answer the background density (BG), number (N), and aftershock density (AS) questions, respectively. Models which perform well tend to answer the BG question with the free background ingredient model, and the AS question with the free productivity model. Conversely, models which address the BG question with the free productivity model, and those which address the AS question with the free background model, tend to perform poorly.

This is highlighted in the box plots of Figure 5.1. There, for each question, the distribution of IGPE of the 64 QDE models is given per possible answer. While for the number questions, no clear trend can be inferred, it is evident that the free background model serves well at answering the BG question and the free productivity model serves well at answering the AS question. These trends are qualitatively very similar in Southern California and Italy.

These results emphasize the added value generated by the fETAS approach, although most fETAS models individually do not outperform standard ETAS. Apparently, a model which gives full flexibility to the background rate during parameter inversion is more informative than others when addressing the background density question. And a model which is flexible at identifying aftershocks is more informative than others when answering the aftershock density question. These observations are made for both considered regions.

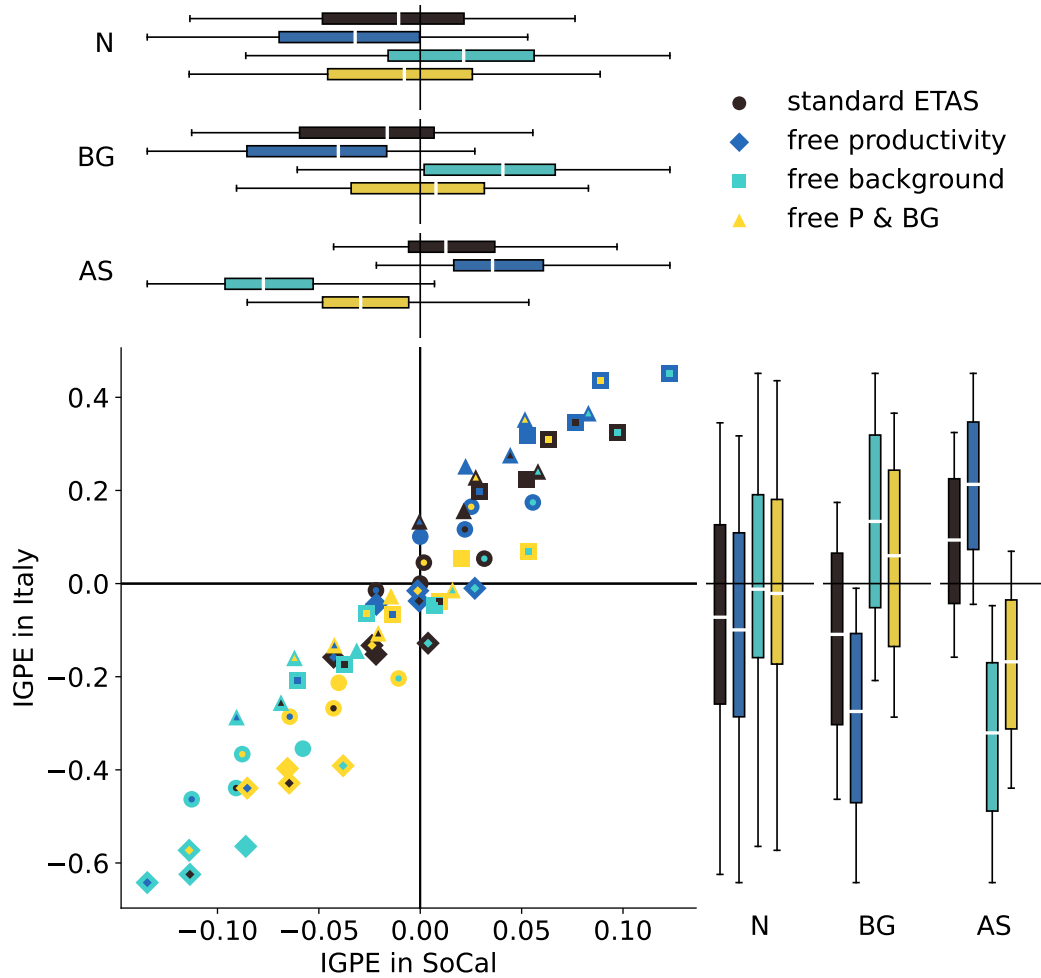


Figure 5.1: Scatter plot of IGPE over standard ETAS of the 64 QDE models in Italy and Southern California. Symbol shape, fill color and edge color describe the composition of the QDE. Shape, fill color, and edge color represent the ingredient model used to answer the background density (BG), number (N), and aftershock density (AS) questions, respectively. Box plots on top (for Southern California) and to the right (for Italy) of the scatter plot: For N, BG, and AS questions, the four boxes represent the IGPE of four groups of QDE models. Each group contains the 16 QDE models which use a specific ingredient model (indicated by box color) to answer the indicated question.

While conceptually it makes sense that a model which can more flexibly capture one particular aspect of seismicity is particularly successful at answering questions about this very aspect of seismicity, this is simultaneously a somewhat counter-intuitive result. If flETAS with free background is more successful than other models at identifying background events, one would expect it, due to the self-consistent nature of parameter inversion, to also be more successful at identifying aftershocks, and thus at describing their occurrence times and locations.

A possible interpretation of the observation that E^{221} , E^{220} , and even E^{223} can so clearly outperform E^{222} , is the following. Compared to the null model M_0 , model $M_2 = E^{222}$ allows the background seismicity to be free and therefore interprets a higher fraction of events in the training catalog to be background earthquakes, which manifests in a much higher background rate. M_2 can thus explain the spatial distribution of background events well,

as well as the partitioning of seismicity into background events and aftershocks. Possibly, M_2 overestimates the background portion of the training catalog due to “too much freedom”. The level of overestimation may be small enough so that M_2 still captures the fraction and locations of background earthquakes better than the other ingredient models do. Overestimation of the background seismicity comes with underestimation of the fraction of aftershocks in the training catalog. While this underestimation may have a minor biasing effect on the number of background earthquakes and aftershocks, the spatio-temporal distribution of aftershocks can be affected in a more harmful way. Aftershocks which occur in the tails of the spatial or temporal distributions have higher chances to be falsely identified as background events compared to aftershocks which are close to their parent event. This leads to a distorted characterization of the aftershock triggering behavior of model M_2 , which can be fixed by using the triggering parameters from models M_0 or M_1 , as indicated by the good performance of models E^{221} and E^{220} .

Another noteworthy observation is that model M_3 , which in principle has all the flexibility necessary to encompass the parameterization of model E^{221} , is clearly outperformed by E^{221} . We interpret this to be a consequence of the fact that the information which is optimized during model calibration and the information used for forecasting are not the same. This does not indicate a flaw in the method presented, but rather illustrates a complexity of the forecasting problem to which the QDE approach offers an apparently useful solution.

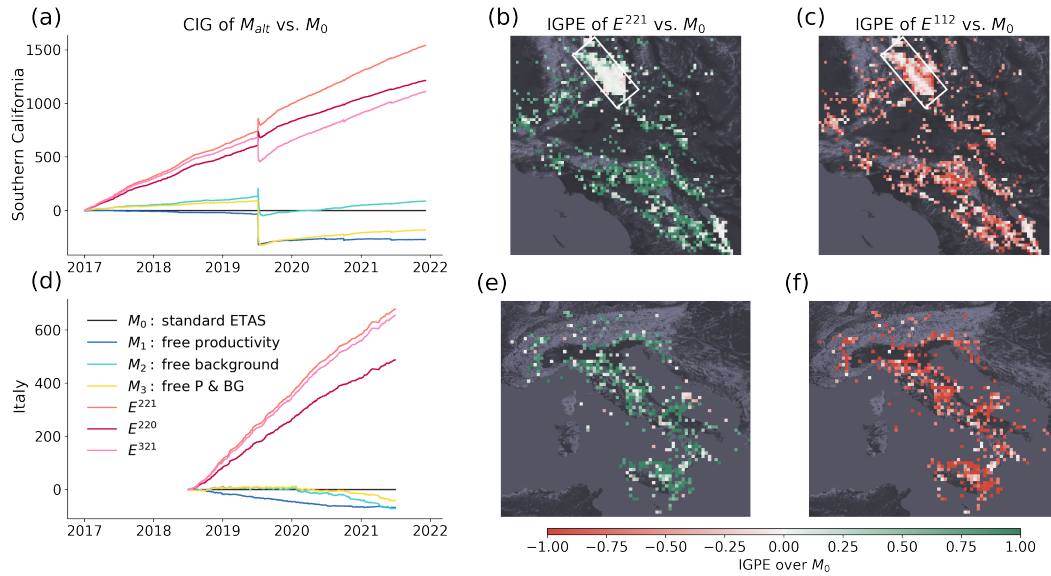


Figure 5.2: Top panels show results for Southern California, bottom panels for Italy. (a) and (d): Cumulative Information Gain (CIG) over time of the ingredient models and the three QDE models best performing in Southern California, compared to the standard ETAS model indicated by the black horizontal line. (b-c) and (e-f): Information gain per earthquake (IGPE) per spatial grid cell of the best performing QDE model (E^{221} , (b) and (e)) and the worst performing QDE model (E^{112} , (c) and (f)), compared to standard ETAS ($M_0 = E^{000}$). Grid cell resolution is 0.05×0.05 degrees in SoCal, and 0.2×0.2 degrees in Italy, chosen for best visibility. The white rectangle in (b-c) highlights the region of the Ridgecrest sequence in 2019.

Figure 5.2(a) shows the cumulative information gain (CIG) over the standard ETAS model over time of the three flETAS ingredient models, and the three best performing QDE models. The CIG of model i_1 over model i_2 at time t is given as the sum of IGs of all forecasting periods ending prior to time t ,

$$\sum_{j:t_j < t} IG_j^{i_1, i_2}. \quad (5.26)$$

In Southern California, the fETAS ingredient models have a negative information gain following the Ridgecrest events in July 2019, meaning that during this time, the standard ETAS model (M_0) is better performing. The free background model M_2 outperforms M_0 immediately after the onset of the sequence, and suffers from information loss later during the sequence. The other two ingredient models do not exhibit the initial information gain. Among the fETAS models, only M_2 can compensate for the information loss during the course of the 5 years of testing and ends up with a positive overall information gain.

Among the QDE models presented, models E^{221} and E^{220} show an initial information gain after the onset of the Ridgecrest sequence, followed by a period of information loss. In contrast to the ingredient models, the information loss during the sequence is smaller than the gain at the beginning of the sequence, such that these models show positive information gain during the Ridgecrest sequence. The three QDE models in Figure 5.2(a) also show a rapidly accumulating information gain throughout the testing period, arriving at an overall IGPE of 0.12, 0.10 and 0.09.

From Figure 5.2(b), it is clear that the IGPE is relatively close to zero in the Ridgecrest area, and the positive IG during the sequence must come from a few specific locations. In the rest of Southern California, higher IGPE values are achieved, with a median grid-cell-wise IGPE of 0.66 for model E^{221} shown in 5.2(b). Conversely, the median grid-cell-wise IGPE for the worst performing model E^{112} shown in 5.2(c) is -0.54. Generally, it performs poorly where E^{221} performs well.

In Italy, all fETAS models have negative total information gain over M_0 . Nevertheless, two of the top three QDE models which perform best in Southern California are also among the top three in Italy, with overall IGPE values of 0.45 and 0.44 for E^{221} and E^{321} . The second best model of SoCal, E^{220} , ranks sixth in Italy with an IGPE of 0.32. Similar to what can be observed in Southern California, the regions in Italy in which the best performing model E^{221} performs well coincide with the areas in which model E^{112} shown in Figure 5.2(f) performs poorly. The median grid-cell-wise IGPE of the two models are 0.76 and -0.82, respectively. Although these grid-cell-wise IGPE values cannot directly be compared between Italy and Southern California due to the different size of the grid cells, the results suggest a qualitatively more similar model performance between the two regions than what is shown by the overall IGPE shown in Figure 5.1. The lower IGPE in SoCal is likely caused by a relatively small IG during the Ridgecrest sequence when a large fraction of events occurred.

5.4.2 Pseudo-prospective model selection

Figure 5.3 illustrates the composition and performance of QDE- S_n models. The number n of past forecasting periods considered when selecting the forecasting model for the next period is in $\{1 = 2^0, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024 = 2^{10}\}$ for SoCal, and $n \in \{1 = 2^0, \dots, 512 = 2^9\}$ for Italy. We do not consider $n = 1024$ for Italy, as this would reduce the number of testing periods in which QDE- S_n is defined by more than half compared to the QDE models. The top, middle, and bottom parts of Figure 5.3(a) and (b) show the ingredient model used by QDE- S_n to answer the N, BG, and AS questions over time. Within each part, n increases from top to bottom. As expected, the composition of QDE- S_n is more stable as n increases, and is almost always defined via E^{221} for large n , in both regions.

In Southern California, a change in composition can be observed after the onset of the Ridgecrest sequence in July 2019. Specifically, the number questions are best answered by standard ETAS, free productivity fETAS, and free productivity and background fETAS, in this order, before moving back to answering with free background fETAS. The aftershock

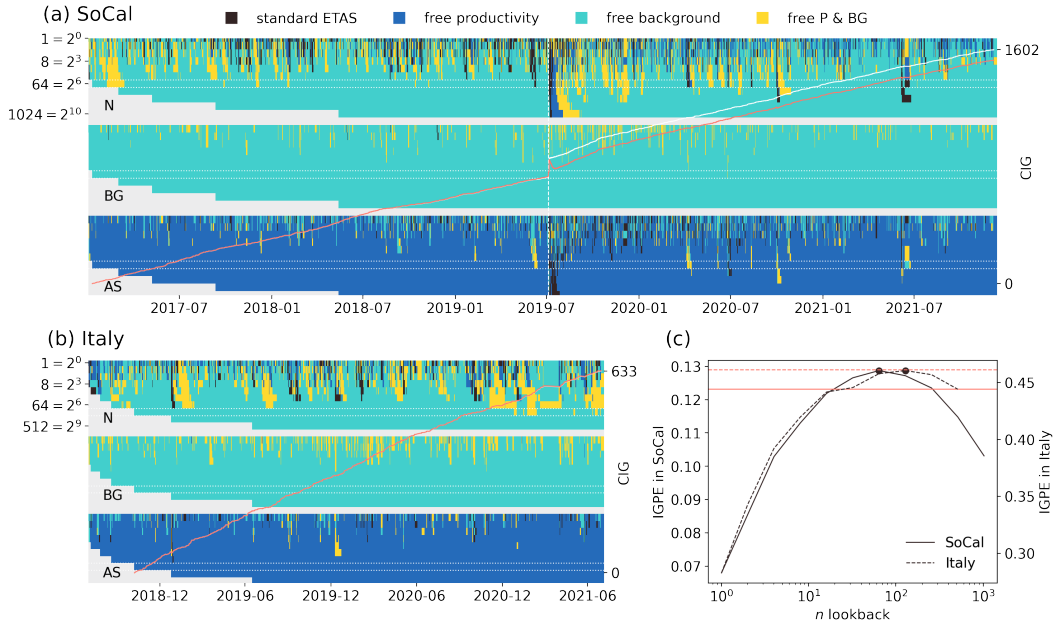


Figure 5.3: Composition and performance of QDE- S_n models. (a) and (b) for Southern California and Italy: Composition of QDE- S_n , where n takes values of powers of 2. Top, middle, and bottom part represent the ingredient model used to answer the number (N), background density (BG), and aftershock density (AS) questions. Within each part, n increases from top to bottom. Dotted white lines highlight the best performing QDE- S_n . Solid white and orange line show the cumulative information gain (CIG) of the best QDE- S_n and best QDE (E^{221}), respectively, for the period in which both are defined. White line is barely visible for Italy because it coincides with the orange line. Vertical dashed line indicates the occurrence time of the M6.4 Ridgecrest event on July 04, 2019. (c): IGPE of different QDE- S_n (black lines), for different values of n . Horizontal orange lines indicate IGPE of E^{221} for the period in which the best QDE- S_n is defined. Solid lines represent Southern California, dashed lines represent Italy.

question intermittently best answered by standard ETAS during the sequence. It is interesting to note here that the performance of E^{221} and QDE- S_{64} are almost identical throughout the 5 years of testing, with the difference that QDE- S_{64} does not show the information loss after the initial information gain after the onset of the sequence. This results in an overall IGPE of 0.13 and 0.12 for QDE- S_{64} and E^{221} , during the period in which both are defined, as is shown in Figure 5.3(c). Thus, the QDE- S_n model, which was originally designed to avoid a biased selection of the winning model after knowing the experiment outcome, is capable of outperforming the winning QDE model for good choices of n , and clearly outperforms all ingredient fETAS models for any tested choice of n .

In Italy, the best performing QDE- S_n model is QDE- S_{128} . It is almost always using E^{221} to issue a forecast for the next period, and thus unsurprisingly achieves the same IGPE. As in SoCal, all tested choices of n yield a model which clearly outperforms all ingredient fETAS models. The most simple QDE- S_n model, QDE- S_1 , which always selects the best QDE model of the previous forecasting period to issue the next forecast, already achieves a very high IGPE of 0.28.

5.5 Conclusions

We describe an adapted ETAS expectation maximization (EM) algorithm which allows a non-parametric inversion of aftershock productivity and/or background rate. Further, we introduce a novel approach of question-driven ensemble (QDE) modeling, which combines ingredient models by using them to answer different forecasting sub-problems. In pseudo-prospective forecasting experiments for Southern California and Italy, we compare the forecasting skill of three flexible ETAS (fETAS) models and a total of 60 nontrivial QDEs of fETAS and ETAS models, to that of the standard ETAS null model.

We find that the best models tend to use fETAS with free background to model the number of events and locations of background earthquakes, and fETAS with free productivity to model the time and location of aftershocks. The best model is the same in both regions and achieves an information gain per earthquake (IGPE) over standard ETAS of 0.12 in Southern California, and 0.45 in Italy.

To address the possible concern of a biased selection of the winning model after knowing the experiment outcome, we also test the forecasting skill of a model which pseudo-prospectively selects the currently best performing QDE model to issue the forecast for the next testing period. Depending on the criteria to identify the best QDE model, we find that the forecasting skill can be greater than that of the overall best QDE model. This approach thus provides a promising candidate for an operational earthquake forecast.

During the 2019 Ridgecrest sequence in Southern California, different ingredient models are best suited to model the number of events during different stages of the sequence. The idea of operationally selecting different QDE models (i.e. selecting different ETAS model parameters) based on their recent performance is in this case related to the idea of Page et al. (2016). They considered sequence-specific parameters to be sampled from an underlying distribution and described a Bayesian approach to update this distribution as aftershock data becomes available.

Our results can also be viewed as a first step toward developing a potentially fruitful branch of earthquake forecasting research. Several key questions remain open and are to be addressed in future studies: *Why do QDE models outperform ingredient models which were inverted in a self-consistent way? What drives the success of different QDE models during different phases of the Ridgecrest sequence? How does QDE performance increase when further ingredient models are considered? And what does all of this teach us about the dynamics of seismicity?*

5.6 Data and resources

The Advanced National Seismic System (ANSS) Comprehensive Earthquake Catalog (ComCat) provided by the U.S. Geological Survey (USGS) was searched using <https://earthquake.usgs.gov/data/comcat/> (last accessed January 2022). The Italian Seismological Instrumental and Parametric Data-Base (ISIDe) was used as provided by the organizers of the upcoming CSEP experiment in Italy, and can be accessed via <http://terremoti.ingv.it/en/search>.

Acknowledgments

This study has been funded by the Eidgenössische Technische Hochschule (ETH) research grant for project number 2018-FE-213, “Enabling dynamic earthquake risk assessment (DynaRisk)”, the European Union’s Horizon 2020 research and innovation program under Grant Agreement Number 821115, real-time earthquake risk reduction for a resilient Europe

(RISE), the National Science Foundation (grant EAR-2122168), and the Southern California Earthquake Center (based on NSF Cooperative Agreement EAR-1600087 and USGS Cooperative Agreement G17AC00047). The paper benefited from constructive comments by the Associate Editor and two anonymous referees.

References

- Akinci, A., Moschetti, M. P., & Taroni, M. (2018). Ensemble smoothed seismicity models for the new Italian probabilistic seismic hazard map. *Seismological Research Letters*, 89(4), 1277–1287. [93](#)
- Bach, C., & Hainzl, S. (2012). Improving empirical aftershock modeling based on additional source information. *Journal of Geophysical Research: Solid Earth*, 117(B4). [93](#)
- Bayliss, K., Naylor, M., Illian, J., & Main, I. G. (2020). Data-driven optimization of seismicity models using diverse data sets: Generation, evaluation, and ranking using inlabru. *Journal of Geophysical Research: Solid Earth*, 125(11), e2020JB020226. [92](#)
- Bayona, J., Savran, W., Strader, A., Hainzl, S., Cotton, F., & Schorlemmer, D. (2021). Two global ensemble seismicity models obtained from the combination of interseismic strain measurements and earthquake-catalogue information. *Geophysical Journal International*, 224(3), 1945–1955. [93](#)
- Bird, P., Jackson, D. D., Kagan, Y. Y., Kreemer, C., & Stein, R. (2015). Gear1: A global earthquake activity rate model constructed from geodetic strain rates and smoothed seismicity. *Bulletin of the Seismological Society of America*, 105(5), 2538–2554. [93](#)
- Box, G. E. (1979). Robustness in the strategy of scientific model building. *Robustness in statistics* (pp. 201–236). Elsevier. [92](#)
- Cattania, C., Werner, M. J., Marzocchi, W., Hainzl, S., Rhoades, D., Gerstenberger, M., Liukis, M., Savran, W., Christophersen, A., Helmstetter, A., et al. (2018). The forecasting skill of physics-based seismicity models during the 2010–2012 Canterbury, New Zealand, earthquake sequence. *Seismological Research Letters*, 89(4), 1238–1250. [92](#)
- Cocco, M., Hainzl, S., Catalli, F., Enescu, B., Lombardi, A., & Woessner, J. (2010). Sensitivity study of forecasted aftershock seismicity based on Coulomb stress calculation and rate- and state-dependent frictional response. *Journal of Geophysical Research: Solid Earth*, 115(B5). [92](#)
- Daley, D. J., Vere-Jones, D. et al. (2003). *An introduction to the theory of point processes: Volume i: Elementary theory and methods*. Springer. [99](#)
- Dieterich, J. (1994). A constitutive law for rate of earthquake production and its application to earthquake clustering. *Journal of Geophysical Research: Solid Earth*, 99(B2), 2601–2618. [92](#)
- Enescu, B., Hainzl, S., & Ben-Zion, Y. (2009). Correlations of seismicity patterns in southern California with surface heat flow data. *Bulletin of the Seismological Society of America*, 99(6), 3114–3123. [93](#)
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, 9(5), 1937–1958. [93](#)
- Field, E. H., Jordan, T. H., Page, M. T., Milner, K. R., Shaw, B. E., Dawson, T. E., Biasi, G. P., Parsons, T., Hardebeck, J. L., Michael, A. J., et al. (2017). A synoptic view of the third uniform California earthquake rupture forecast (ucerf3). *Seismological Research Letters*, 88(5), 1259–1267. [93](#)
- Gerstenberger, M., McVerry, G., Rhoades, D., & Stirling, M. (2014). Seismic hazard modeling for the recovery of Christchurch. *Earthquake Spectra*, 30(1), 17–29. [93](#)
- Gerstenberger, M. C., Wiemer, S., Jones, L. M., & Reasenberg, P. A. (2005). Real-time forecasts of tomorrow's earthquakes in California. *Nature*, 435(7040), 328–331. [92](#)

- Grimm, C., Hainzl, S., Käser, M., & 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*, 1–20. [93](#)
- Group, I. W. (2007). Italian seismological instrumental and parametric database (iside). [100](#)
- Hainzl, S. (2022). Etas-approach accounting for short-term incompleteness of earthquake catalogs. *Bulletin of the Seismological Society of America*, *112*(1), 494–507. [93](#)
- Hardebeck, J. L. (2021). Spatial clustering of aftershocks impacts the performance of physics-based earthquake forecasting models. *Journal of Geophysical Research: Solid Earth*, *126*(2), e2020JB020824. [93](#)
- Kamer, Y., Nandan, S., Ouillon, G., Hiemer, S., & Sornette, D. (2021). Democratizing earthquake predictability research: Introducing the richterx platform. *The European Physical Journal Special Topics*, *230*(1), 451–471. [93](#)
- Kovchegov, Y., Zaliapin, I., & Ben-Zion, Y. (2022). Invariant galton-watson branching process for earthquake occurrence. *Geophysical Journal International*, *231*, 567–583. <https://doi.org/10.1093/gji/ggac204>. [93](#)
- Leutbecher, M., & Palmer, T. N. (2008). Ensemble forecasting. *Journal of computational physics*, *227*(7), 3515–3539. [93](#)
- Llenos, A. L., & Michael, A. J. (2019). Ensembles of etas models provide optimal operational earthquake forecasting during swarms: Insights from the 2015 san ramon, california swarmensembles of etas models provide optimal operational earthquake forecasting during swarms. *Bulletin of the Seismological Society of America*, *109*(6), 2145–2158. [93](#)
- Mancini, S., Segou, M., Werner, M., & Cattania, C. (2019). Improving physics-based aftershock forecasts during the 2016–2017 central italy earthquake cascade. *Journal of Geophysical Research: Solid Earth*, *124*(8), 8626–8643. [92](#)
- Mancini, S., Segou, M., Werner, M. J., & Parsons, T. (2020). The predictive skills of elastic coulomb rate-and-state aftershock forecasts during the 2019 ridgecrest, california, earthquake sequence. *Bulletin of the Seismological Society of America*, *110*(4), 1736–1751. [92](#)
- Marzocchi, W., Lombardi, A. M., & Casarotti, E. (2014). The establishment of an operational earthquake forecasting system in italy. *Seismological Research Letters*, *85*(5), 961–969. [93](#)
- Marzocchi, W., Zechar, J. D., & Jordan, T. H. (2012). Bayesian forecast evaluation and ensemble earthquake forecasting. *Bulletin of the Seismological Society of America*, *102*(6), 2574–2584. [93](#)
- Mizrahi, L., Nandan, S., & Wiemer, S. (2021a). Embracing data incompleteness for better earthquake forecasting. *Journal of Geophysical Research: Solid Earth*, *126*(12), e2021JB022379. <https://doi.org/https://doi.org/10.1029/2021JB022379>. [95](#), [96](#), [100](#)
- Mizrahi, L., Nandan, S., & Wiemer, S. (2021b). The Effect of Declustering on the Size Distribution of Mainshocks. *Seismological Research Letters*. <https://doi.org/10.1785/0220200231>. [93](#), [100](#)
- Nandan, S., Kamer, Y., Ouillon, G., Hiemer, S., & Sornette, D. (2021). Global models for short-term earthquake forecasting and predictive skill assessment. *The European Physical Journal Special Topics*, *230*(1), 425–449. [93](#), [95](#)
- Nandan, S., Ouillon, G., & Sornette, D. (2019). Magnitude of earthquakes controls the size distribution of their triggered events. *Journal of Geophysical Research: Solid Earth*, *124*(3), 2762–2780. [93](#)
- Nandan, S., Ouillon, G., Sornette, D., & Wiemer, S. (2019a). Forecasting the full distribution of earthquake numbers is fair, robust, and better. *Seismological Research Letters*, *90*(4), 1650–1659. [100](#)

- Nandan, S., Ouillon, G., Sornette, D., & Wiemer, S. (2019b). Forecasting the rates of future aftershocks of all generations is essential to develop better earthquake forecast models. *Journal of Geophysical Research: Solid Earth*, *124*(8), 8404–8425. [93](#)
- Nandan, S., Ouillon, G., Wiemer, S., & Sornette, D. (2017). Objective estimation of spatially variable parameters of epidemic type aftershock sequence model: Application to california. *Journal of Geophysical Research: Solid Earth*, *122*(7), 5118–5143. [93](#)
- Nandan, S., Ram, S. K., Ouillon, G., & Sornette, D. (2021). Is seismicity operating at a critical point? *Physical Review Letters*, *126*(12), 128501. [93](#), [96](#), [99](#)
- Ogata, Y., Katsura, K., Falcone, G., Nanjo, K., & Zhuang, J. (2013). Comprehensive and topical evaluations of earthquake forecasts in terms of number, time, space, and magnitude. *Bulletin of the Seismological Society of America*, *103*(3), 1692–1708. <https://doi.org/10.1785/0120120063>. [93](#), [99](#)
- Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, *83*(401), 9–27. [92](#)
- Page, M. T., Van Der Elst, N., Hardebeck, J., Felzer, K., & 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. [106](#)
- Page, M. T., & van der Elst, N. J. (2022). Aftershocks preferentially occur in previously active areas. *The Seismic Record*, *2*(2), 100–106. [93](#)
- Parsons, T., Ogata, Y., Zhuang, J., & Geist, E. L. (2012). Evaluation of static stress change forecasting with prospective and blind tests. *Geophysical Journal International*, *188*(3), 1425–1440. [92](#)
- Rhoades, D., Liukis, M., Christophersen, A., & Gerstenberger, M. (2016). Retrospective tests of hybrid operational earthquake forecasting models for canterbury. *Geophysical Journal International*, *204*(1), 440–456. [93](#)
- Rhoades, D. A., & Gerstenberger, M. C. (2009). Mixture models for improved short-term earthquake forecasting. *Bulletin of the Seismological Society of America*, *99*(2A), 636–646. [93](#)
- Savran, W. H., Werner, M. J., Marzocchi, W., Rhoades, D. A., Jackson, D. D., Milner, K., Field, E., & Michael, A. (2020). Pseudoprospective evaluation of ucerf3-etas forecasts during the 2019 ridgecrest sequence. *Bulletin of the Seismological Society of America*, *110*(4), 1799–1817. [93](#)
- Savran, W. H., Werner, M. J., Schorlemmer, D., & Maechling, P. J. (2022). Pycsep: A python toolkit for earthquake forecast developers. *Journal of Open Source Software*, *7*(69). [100](#)
- Schoenberg, F. P. (2013). Facilitated estimation of etas. *Bulletin of the Seismological Society of America*, *103*(1), 601–605. [97](#)
- Schorlemmer, D., Christophersen, A., Rovida, A., Mele, F., Stucchi, M., & Marzocchi, W. (2010). Setting up an earthquake forecast experiment in italy. *Annals of Geophysics*. [100](#)
- Schorlemmer, D., Zechar, J. D., Werner, M. J., Field, E. H., Jackson, D. D., Jordan, T. H., Group, R. W., et al. (2010). First results of the regional earthquake likelihood models experiment. *Seismogenesis and earthquake forecasting: The frank evison volume ii* (pp. 5–22). Springer. [93](#)
- Seif, S., Mignan, A., Zechar, J. D., Werner, M. J., & Wiemer, S. (2017). Estimating etas: The effects of truncation, missing data, and model assumptions. *Journal of Geophysical Research: Solid Earth*, *122*(1), 449–469. [93](#)
- Shebalin, P. N., Narteau, C., Zechar, J. D., & Holschneider, M. (2014). Combining earthquake forecasts using differential probability gains. *Earth, Planets and Space*, *66*(1), 1–14. [93](#)

- Steacy, S., Gerstenberger, M., Williams, C., Rhoades, D., & Christophersen, A. (2014). A new hybrid coulomb/statistical model for forecasting aftershock rates. *Geophysical Journal International*, *196*(2), 918–923. [92](#)
- Strader, A., Schneider, M., & Schorlemmer, D. (2017). Prospective and retrospective evaluation of five-year earthquake forecast models for california. *Geophysical Journal International*, *211*(1), 239–251. [93](#)
- Taroni, M., Zechar, J., & Marzocchi, W. (2014). Assessing annual global m 6+ seismicity forecasts. *Geophysical Journal International*, *196*(1), 422–431. [93](#)
- Taroni, M., Marzocchi, W., Schorlemmer, D., Werner, M. J., Wiemer, S., Zechar, J. D., Heiniger, L., & Euchner, F. (2018). Prospective csep evaluation of 1-day, 3-month, and 5-yr earthquake forecasts for italy. *Seismological Research Letters*, *89*(4), 1251–1261. [92](#), [93](#)
- Tracton, M. S., & Kalnay, E. (1993). Operational ensemble prediction at the national meteorological center: Practical aspects. *Weather and Forecasting*, *8*(3), 379–398. [93](#)
- van der Elst, N. J., Hardebeck, J. L., Michael, A. J., McBride, S. K., & Vanacore, E. (2022). Prospective and retrospective evaluation of the us geological survey public aftershock forecast for the 2019–2021 southwest puerto rico earthquake and aftershocks. *Seismological Society of America*, *93*(2A), 620–640. [93](#)
- Veen, A., & 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. [95](#), [96](#)
- Woessner, J., Hainzl, S., Marzocchi, W., Werner, M., Lombardi, A., Catalli, F., Enescu, B., Cocco, M., Gerstenberger, M., & Wiemer, S. (2011). A retrospective comparative forecast test on the 1992 landers sequence. *Journal of Geophysical Research: Solid Earth*, *116*(B5). [93](#)
- Zechar, J. D., Schorlemmer, D., Liukis, M., Yu, J., Euchner, F., Maechling, P. J., & Jordan, T. H. (2010). The collaboratory for the study of earthquake predictability perspective on computational earthquake science. *Concurrency and Computation: Practice and Experience*, *22*(12), 1836–1847. [93](#)
- Zhuang, J. (2012). Long-term earthquake forecasts based on the epidemic-type aftershock sequence (etas) model for short-term clustering. *Research in Geophysics*, *2*(1), e8–e8. [96](#), [99](#)

CHAPTER 6

CONCLUSIONS AND OUTLOOK

Human beings are frightened yet fascinated by earthquakes, and likely have been so ever since the first person experienced the immense forces which cause the usually immobile ground to shake. Pliny the Younger (~ 100 AD) in his letters to Tacitus described the tragic events of the eruption of Mount Vesuvius in the year 79 AD, famously known for causing the burial of the ancient city of Pompeii beneath several meters of volcanic ash. Almost 2000 years ago, Pliny and his mother left their house to protect themselves from collapsing buildings as they were woken up by strong earthquakes which accompanied the eruption. Today, they would still be advised to do exactly that.

One might wonder how thousands of years of human curiosity have hardly been able to produce any progress regarding the protective measures taken in case of earthquakes. One might wonder why the progress made towards accurate earthquake forecasting is so limited, and whether it is worth the effort to continue studying this topic.

This is not to say that the study of earthquakes in general is only justified through its potential to contribute to the advancement of earthquake prediction. We can use information from recorded earthquakes to better understand the structures beneath us (e.g. Lanza et al., 2022), which in itself is valuable for its ability to satisfy our curious minds. Seismic imaging techniques are used to identify attractive sites to be used for geothermal energy production (Sánchez-Pastor et al., 2021). In light of the urgent need to move to renewable energy sources in the hope to limit the effects of climate change, such developments are of undeniable importance for our society. With the increasing amounts of available high-quality data, individual faults and earthquake sequences can nowadays be studied in staggering detail (Ross et al., 2019), possibly revealing previously unknown physical mechanisms which lead up to or follow major events. In comparison, the development of forecasting models which inform us whether the probability of a magnitude 5.0 earthquake in Switzerland in the next 7 days is 1.31% or 0.06% (see Table 4.3) may seem unimportant.

However, the past has shown that experts' assessment of the seismic hazard during an ongoing sequence does influence people's behavior, which can subsequently cause or prevent casualties (Cartledge, 2016). It is for this reason that I believe that every quantitatively demonstrated improvement of a forecasting model, as minor as it may be, is valuable in the grand scheme of things.

In this thesis, I contributed to the advancement of the field of earthquake forecasting from several different perspectives. Chapter 2 focuses on analyzing existing methods for time-independent forecasting and identifies weaknesses therein. Chapter 3 addresses data incompleteness issues in short- and long-term time horizons and proposes improved methods for time-dependent earthquake forecasting. Chapter 4 takes earthquake forecasting in Switzerland to the next level through the development, testing, and finally the recommendation of the ideal candidate model for a first Swiss OEF system. Chapter 5 describes a novel approach to combine existing models into new ones, facilitating the search for better

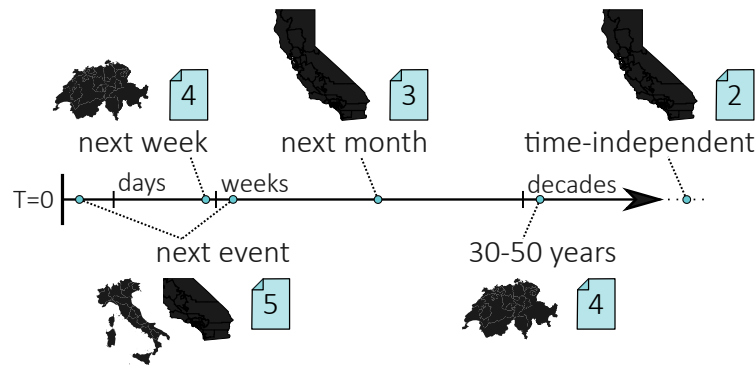


Figure 6.1: Schematic overview of forecasting horizons and regions covered in this thesis. Icons with numbers indicate the associated chapter.

earthquake forecasting models in the future.

The different modes of investigation in Chapters 2 - 5 are on one hand a testimony of my personal development from an outside observer to an active contributor to earthquake forecasting tools. They are on the other hand a reminder of the variety of directions in which this field can be advanced.

In this final chapter, I will revisit and summarize the results and conclusions of all preceding chapters, and finally give an outlook on the future of earthquake forecasting based on the findings in this thesis.

Biased b -values of declustered catalogs

In Chapter 2, it is found that declustering introduces a systematic bias to the size distribution of mainshocks. Depending on the declustering method and hyperparameters used, the b -value can be reduced by up to 30%. The findings of this chapter show that this reduction is, at least partially, an artifact of declustering, rather than an underlying property of mainshocks which is revealed by declustering.

It is common practice in PSHA to estimate seismic hazard based on declustered catalogs. The results of Chapter 2 suggest this to be a problematic approach for several reasons.

- There is a lack of an objective criterion based on which the declustering method is selected, but the selection has a significant influence on the resulting hazard.
- Neglecting aftershocks leads to significant underestimation of the seismic hazard, as brought up by Marzocchi and Taroni (2014).
- Imposing a Gutenberg-Richter distribution on the magnitudes of a declustered catalog leads to overestimation of the seismic hazard above a certain magnitude.

Conclusion 1. A re-thinking of the way in which PSHA is done needs to take place. While it is inevitable and crucial to analyze earthquake-earthquake interactions to understand the nature of long-term and short-term seismicity in a region, ignoring aftershocks that were identified with an arbitrarily selected method is not recommended.

ETAS models: a versatile approach to earthquake forecasting

One of the final considerations of Chapter 2 is that using ETAS models in the time-independent context may be the solution to the demonstrated issues of the current approach to PSHA. ETAS models are well suited to distinguish background and triggered seismicity, and to model long-term seismicity which includes the highly hazardous aftershock sequences.

However, there are more advantages to ETAS models than their independence of the process of declustering for time-independent hazard studies. At the core of ETAS models lies their ability to model the time-dependent behavior of seismicity at any short- or medium-term time scale of interest. Throughout this thesis, they play a fundamental role in modeling seismicity, and Figure 6.1 illustrates the diversity of forecasting horizons and regions that are addressed using ETAS models in the different chapters of the thesis.

While Chapter 2 discusses time-independent earthquake forecasting, in Chapters 3 - 5, ETAS models are used with time-dependent forecasting horizons ranging from inter-event time-scale to long-term horizons of 30-50 years. This highlights the universal applicability of these models for a broad spectrum of forecasting horizons.

Chapter 4 in particular showcases the applicability of ETAS models for both OEF and PSHA on the example of Switzerland. Several variants of a basic ETAS model are tested in pseudo-prospective one-week forecasting experiments and in 30- and 50-year retrospective consistency tests, whereupon the purely ETAS-based models are recommended as ideal candidates for the first Swiss OEF system.

Conclusion 2. The same simple ETAS models are shown to be useful in one-week and 50-year forecasting horizons. This emphasizes their applicability at both ends of the spectrum of forecasting horizons, bringing the realms of PSHA and OEF closer together through a consistent approach.

The importance of using locally calibrated ETAS parameters

In this thesis, models and ideas are tested for (Southern) California, Switzerland, and Italy. The versatility of ETAS models extends not only to the temporal, but also the spatial domain; methods developed and tested using data of one region are often also useful elsewhere. In Chapter 5, we demonstrate this by applying the same models to Southern California and Italy and finding striking similarities between the results in the two regions. In Chapter 4, a method developed in Chapter 3 for California is applied in Switzerland without modification, and the resulting forecasting model ends up being recommended for the first Swiss OEF system.

On the other hand, the ETAS variant which uses generic parameters inverted for California performs poorly in Switzerland.

Conclusion 3. While ETAS models are generally not bound to a particular region, using locally calibrated parameters is essential when issuing forecasts. This highlights the value of parameter inversion techniques which can be applied in the case of limited available data.

The role of small events in triggering large ones

One of the methods developed in Chapter 3 allows the estimation of ETAS parameters when the completeness of the catalog varies with space and/or time. This can be particularly useful in regions where data is relatively sparse and the amount of data usable for inversion greatly increases when such m_c -variations are accounted for. This approach was, as a result of the study in Chapter 4, suggested to be one of the candidates for the first time-dependent earthquake forecast of Switzerland.

The second approach introduced in Chapter 3 jointly estimates ETAS parameters and high-frequency detection incompleteness at each point in time. Both methods are shown to be able to accurately invert the data-generating parameters in synthetic tests. The second approach is then further used to test the forecasting power of small earthquakes in a pseudo-prospective forecasting experiment for California. The results thereof suggest that:

- Information about small earthquakes clearly improves forecasts of similarly-sized events.

- This improvement is mainly achieved through the simulation of aftershocks of small events.
- To observe the improvement, it is necessary to account for incompleteness when simulating.
- Information about small earthquakes does not significantly affect the performance of forecasts targeting larger magnitude events.

Previous results of Nichols and Schoenberg (2014), Spassiani and Sebastiani (2016), Nandan et al. (2019) and Nandan et al. (2021) offer a possible interpretation of these findings; a tendency of earthquakes to trigger similarly sized aftershocks.

The results of this chapter provide useful new insights that can point in the right direction, or one of the potentially many right directions, on the way to better earthquake forecasting. If earthquakes tend to trigger similarly sized aftershocks, forecasts are expected to improve when this property is captured by the models. On the other hand, it implies that ETAS-like modeling of small events' aftershocks will not make large events substantially more predictable, and other patterns in the occurrence of earthquakes need to be found.

Conclusion 4. Further research is required to investigate if and how the dependency of aftershock magnitudes on the magnitude of their parent event can be beneficially incorporated into existing models.

QDE modeling as a tool for the future

In Chapter 5, a novel approach to construct question-driven ensembles of ETAS models is described, along with flexible ETAS models which, during parameter inversion, use a nonparametric definition of background seismicity and aftershock productivity. While these fETAS models individually do not show significantly improved forecasting performance compared to the standard ETAS variant, the study shows that QDEs of fETAS models can clearly outperform their ingredient models.

A QDE model can be viewed as a model which answers questions about the number of expected earthquakes, the spatio-temporal distribution of background events, and the spatio-temporal distribution of aftershocks with different ingredient models. Testing all possible QDE combinations of the ingredients provides information about which ingredient models are suited to answer which questions well. This can help to shed light on the strengths and limitations of the ingredient models. The result that the same QDEs are best performing in Italy and Southern California implies the ability of this approach to reveal previously unknown but universally valid properties of seismicity.

Conclusion 5. Using the best available ETAS model as a benchmark, the QDE approach can help modelers to find seismicity patterns that go beyond what is currently incorporated in ETAS.

Outlook

While this thesis is a contribution to the advancement of earthquake forecasting, we are as a community still far from deterministically predicting the exact time and location of the next devastating earthquake. It is unknown today whether that will ever be possible and how predictable earthquakes will ever be. But what is clear is that probabilistic earthquake forecasting capabilities can still be improved. And based on them, relatively low-cost risk mitigation measures can be taken.

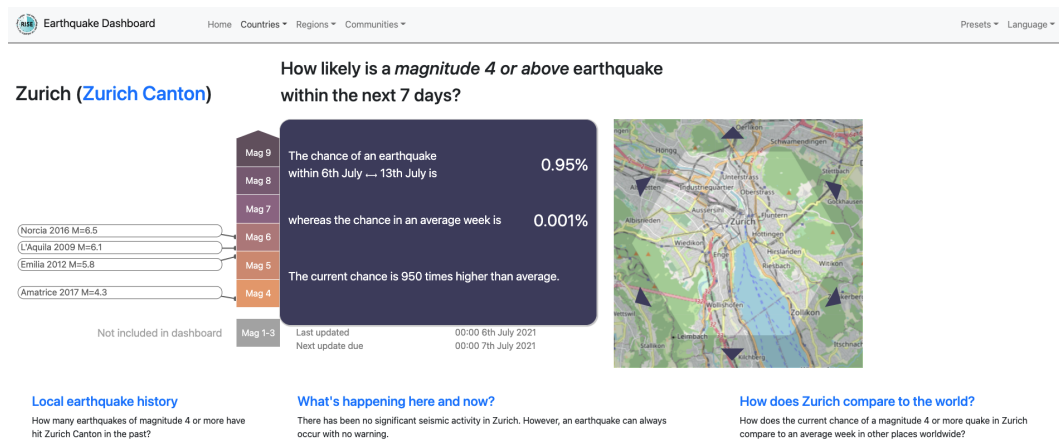


Figure 6.2: Screenshot of the preliminary Earthquake Dashboard by the Winton Centre for Risk & Evidence Communication, University of Cambridge, UK (last accessed on 27 July, 2022).

operationalize The most straightforward next steps to take based on the results of this thesis are to operationalize the forecasting model for Switzerland which is proposed in Chapter 4. Due to the general nature in which the models in Chapter 4 are described and developed, an analogous model can be developed and tested for Europe with few minor adaptations. Its operationalization and the establishment of rule-based response measures will positively influence the resilience of an even wider and partially more vulnerable society.

communicate Additionally, a cost-benefit analysis of possible measures taken as a response to a temporally elevated earthquake probability, as it is envisioned by Böse et al. (2022) in the context of earthquake early warning, enables a direct impact of our research on society. One possible measure is the communication of time-dependent earthquake probabilities to the public. For this measure to achieve the highest benefit, it is crucial to study how information is best communicated. Dallo et al. (2020) investigated how earthquake information is best integrated into multi-hazard platforms. Besides behavioral recommendations for the public, such platforms could in the future contain information about the time-dependent earthquake probability, provided that it is presented adequately. Figure 6.2 shows a screenshot of the preliminary Earthquake Dashboard which is being developed at the Winton Centre for Risk & Evidence Communication, University of Cambridge, UK. Currently fed with dummy data, it shows earthquake probabilities for the next seven days in the city of Zurich, compared to the probability in the same area during an average week. The user can also view the local earthquake history and compare local probabilities with the rest of the world.

refine The models presented in Chapter 4 are relatively basic versions of the ETAS model. Chapters 3 and 5 provide recipes for improved time-dependent earthquake forecasting models. The natural next steps after making a first earthquake forecast operational in Switzerland and Europe would thus be to address short-term aftershock incompleteness by calibrating and testing the PETAI model of Chapter 3, which is expected to produce superior forecasts. Furthermore, fETAS models and question-driven ensemble (QDE) models proposed in Chapter 5 shall be considered to be applied operationally.

advance Besides putting the models and findings of this thesis into action in Switzerland, Europe, and possibly elsewhere, the field of earthquake forecasting needs to develop further.

The QDE modeling approach provides a powerful tool to use in this regard. The more ingredient models that are being tested, the higher are the chances of obtaining better forecasts, while simultaneously the strengths and weaknesses of the individual models can be better understood.

To be able to combine models using the QDE approach, or any other approach, a crucial requirement is the development of new ingredient models. The more heterogeneous the pool of forecasting models becomes, and the more these models are tested and compared, the more will be learned from those tests and comparisons. The results of this thesis can offer guidance in the search of the best possible earthquake forecasting model.

Two key goals should be aimed for:

- *Improve the forecasted earthquake magnitude.* As societies' main interest lies in correctly forecasting large earthquakes, constraining the magnitudes of forecasted events beyond a simple Gutenberg-Richter law is strongly recommended. The results of Chapter 3 are in line with results of previous studies which found that earthquakes tend to preferentially trigger similarly sized aftershocks (Nichols and Schoenberg, 2014; Spasiani and Sebastiani, 2016; Nandan et al., 2019; Nandan et al., 2022). Shcherbakov et al. (2019) proposed a Bayesian method to forecast the magnitude of the largest expected earthquake. Gulia and Wiemer (2019) used temporal b -value variations to determine whether a subsequent larger event is to be expected. In a next step, such ideas should be tested (pseudo-)prospectively and their value for forecasting should be evaluated compared to current state-of-the-art models. Additionally, other ways to address the question of magnitude forecasting can and should be examined. In this way, robust and reliable methodologies which improve current forecasting models can be established.
- *Improve the modeled background seismicity.* At this point, the main focus of ETAS models lies on accurately modeling aftershock triggering behavior, while background seismicity can roughly be described as the part of seismicity which can't be explained otherwise. In reality, all earthquakes are triggered by underlying physical processes. Ideally, forecasting models should minimize the 'unexplained' part of seismicity. Physical covariates such as strain rates or fault proximity have been found to be useful in time-independent forecasting (Bayliss et al., 2020). A self-consistent ETAS parameter inversion which includes such additional data for modeling background seismicity is thus a natural next step to take to increase the explained portion of seismicity. Temporal trends in the background rate, for example caused by earthquake swarms, should also be examined and incorporated into current models. Ideas for such models exist and have been proposed (Llenos and Michael, 2019). A key step is to systematically test and evaluate them to establish the next generation of forecasting models.

In summary, the recommendation is to focus on the aspects that are not yet explicitly modeled by ETAS, but are of great importance for society: To forecast earthquake magnitudes more precisely, and to model the first event of potentially intense aftershock cascades more precisely.

share This thesis does not solve the ancient problem of earthquake prediction. Instead, it contributes a few bricks to a wall that protects our society from earthquake hazard. Maybe earthquakes will never be substantially more predictable than they are today. Maybe we are just one brick away from a breakthrough.

Whichever is the case: A forecasting model can only be improved or used as a benchmark by those who can access it. I therefore believe that it is our duty as earthquake forecasting modelers to make our codes freely available to everyone (see Appendix E) and foster interdisciplinary, inter-institutional and inter-national collaboration.

References

- Bayliss, K., Naylor, M., Illian, J., & Main, I. G. (2020). Data-driven optimization of seismicity models using diverse data sets: Generation, evaluation, and ranking using inlabru. *Journal of Geophysical Research: Solid Earth*, *125*(11), e2020JB020226. [116](#)
- Böse, M., Papadopoulos, A. N., Danciu, L., Clinton, J. F., & Wiemer, S. (2022). Loss-based performance assessment and seismic network optimization for earthquake early warning. *Bulletin of the Seismological Society of America*, *112*(3), 1662–1677. [115](#)
- Cartledge, E. (2016). Seven-year legal saga ends as italian official is cleared of manslaughter in earthquake trial. *Science*, *3*. [111](#)
- Dallo, I., Stauffacher, M., & Marti, M. (2020). What defines the success of maps and additional information on a multi-hazard platform? *International Journal of Disaster Risk Reduction*, *49*, 101761. [115](#)
- Gulia, L., & Wiemer, S. (2019). Real-time discrimination of earthquake foreshocks and aftershocks. *Nature*, *574*(7777), 193–199. [116](#)
- Lanza, F., Diehl, T., Deichmann, N., Kraft, T., Nussbaum, C., Schefer, S., & Wiemer, S. (2022). The saint-ursanne earthquakes of 2000 revisited: Evidence for active shallow thrust-faulting in the jura fold-and-thrust belt. *Swiss Journal of Geosciences*, *115*(1), 1–24. [111](#)
- Llenos, A. L., & Michael, A. J. (2019). Ensembles of etas models provide optimal operational earthquake forecasting during swarms: Insights from the 2015 san ramon, california swarmensembles of etas models provide optimal operational earthquake forecasting during swarms. *Bulletin of the Seismological Society of America*, *109*(6), 2145–2158. [116](#)
- Marzocchi, W., & Taroni, M. (2014). Some thoughts on declustering in probabilistic seismic-hazard analysis. *Bulletin of the Seismological Society of America*, *104*(4), 1838–1845. [112](#)
- Nandan, S., Ouillon, G., & Sornette, D. (2019). Magnitude of earthquakes controls the size distribution of their triggered events. *Journal of Geophysical Research: Solid Earth*, *124*(3), 2762–2780. [114](#), [116](#)
- Nandan, S., Ouillon, G., & Sornette, D. (2022). Are large earthquakes preferentially triggered by other large events? [116](#)
- Nandan, S., Ram, S. K., Ouillon, G., & Sornette, D. (2021). Is seismicity operating at a critical point? *Physical Review Letters*, *126*(12), 128501. [114](#)
- Nichols, K., & Schoenberg, F. P. (2014). Assessing the dependency between the magnitudes of earthquakes and the magnitudes of their aftershocks. *Environmetrics*, *25*(3), 143–151. [114](#), [116](#)
- Pliny the Younger, G. P. C. S. (~ 100 AD). *Epistularum liber sextus*. [111](#)
- Ross, Z. E., Idini, B., Jia, Z., Stephenson, O. L., Zhong, M., Wang, X., Zhan, Z., Simons, M., Fielding, E. J., Yum, S.-H., et al. (2019). Hierarchical interlocked orthogonal faulting in the 2019 ridgecrest earthquake sequence. *Science*, *366*(6463), 346–351. [111](#)
- Sánchez-Pastor, P., Obermann, A., Reinsch, T., Ágústsdóttir, T., Gunnarsson, G., Tómasdóttir, S., Hjörleifsdóttir, V., Hersir, G., Ágústsson, K., & Wiemer, S. (2021). Imaging high-temperature geothermal reservoirs with ambient seismic noise tomography, a case study of the hengill geothermal field, sw iceland. *Geothermics*, *96*, 102207. [111](#)
- Shcherbakov, R., Zhuang, J., Zöller, G., & Ogata, Y. (2019). Forecasting the magnitude of the largest expected earthquake. *Nature communications*, *10*(1), 1–11. [116](#)
- Spassiani, I., & Sebastiani, G. (2016). Exploring the relationship between the magnitudes of seismic events. *Journal of Geophysical Research: Solid Earth*, *121*(2), 903–916. [114](#), [116](#)

APPENDICES

APPENDIX A

SUPPORTING INFORMATION FOR CHAPTER 2

Description of the Supporting Information

This document contains a detailed description of the methodology, including the formula used for b -value estimation and the algorithm for joint estimation of b -value and completeness magnitude, the description of all declustering methods and parameter ranges applied, as well as a description of the ETAS model, parameter inversion, and catalog simulation algorithm applied. Finally, we analyze the sensitivity on the completeness magnitude m_c of the full catalog and mainshock b -value, and sensitivity on magnitude binning of the full catalog b -value.

A.1 b -value estimation

According to the Gutenberg-Richter law (Gutenberg and Richter, 1944), frequency-magnitude distribution can be described as

$$N(m) = N_0 \cdot e^{-\beta \cdot m} = 10^a \cdot 10^{-b \cdot m}, \quad (\text{A.1})$$

where $N(m)$ is the number of earthquakes of magnitude $M \geq m$, N_0 is the total number of earthquakes above magnitude $M = 0$. a and b will subsequently be called ‘ a -value’ and ‘ b -value’, respectively. Note that $\beta = \ln(10) \cdot b$.

β is estimated using the formula proposed in Tinti and Mulargia (1987) for binned magnitude values (in contrast to continuous values). It reads

$$\hat{\beta} = \frac{1}{\Delta M} \cdot \ln(p), \quad (\text{A.2})$$

where

$$p = 1 + \frac{\Delta M}{\hat{\mu} - m_c}. \quad (\text{A.3})$$

m_c is the completeness magnitude, $\hat{\mu}$ is the mean observed magnitude and ΔM is the bin size in which earthquake magnitudes are given, in our case 0.2.

A.2 Completeness magnitude

Estimating the b -value of a catalog requires knowledge of its completeness magnitude m_c , the magnitude threshold above which all events are assumed to be detected. Assuming too low values for m_c can cause severe underestimation of the b -value (see Figure A.2). On the other hand, assuming overly conservative values for m_c leads one to discard a large portion of the data, making b -value estimates imprecise. In reality, m_c is not known and has to be estimated itself. Commonly, m_c is estimated by defining it as the magnitude threshold above which earthquakes follow the GR law (Mignan and Woessner, 2012). In this sense, the estimation of b -value and m_c becomes a coupled problem; one cannot be estimated without knowledge of the other. In the following section, we adapt the method proposed by Clauset et al. (2009) to jointly estimate m_c and b -value.

For the estimation of completeness magnitude, we use all events above magnitude 0.0 in the primary time period, in the same collection area and using the same binning as described in Section 2.2 of the paper. For simplicity, m_c is chosen to be constant in time and space, with the knowledge that this assumption does not generally hold. Short-term aftershock incompleteness after large events and modifications to the seismic network or analysis procedures will increase (and decrease) the completeness magnitude (Woessner and Wiemer, 2005; Schorlemmer et al., 2010). We test a range of possible completeness magnitudes m_c^i between 2.0 and 6.0. For each choice of m_c^i , the b -value b^i is calculated for the events above m_c^i , using the formula proposed by Tinti and Mulargia, 1987, (see Section A.1). This yields a discretized GR law and its corresponding cumulative distribution function (CDF). Simultaneously, the observed cumulative distribution function of events above m_c^i is computed. Observed and theoretical CDF are then compared to each other using the Kolmogorov-Smirnov (KS) distance. We then estimate p -values p^i : the probability of observing a KS distance of $D_{m_c^i}$ or higher, under the assumption that the observed magnitudes were drawn from a discretized GR law with a b -value of b^i . This is done by generating 10,000 random samples of the size of the original sample. For each such sample, we compute the KS distance of observed CDF to the theoretical CDF they were drawn from. The p -value p^i is then calculated as the fraction of KS distances that are larger than $D_{m_c^i}$. Finally, we accept the completeness magnitude m_c^i if $p^i \geq 0.05$, and set the according b -value estimator to b^i . Note that this selection of m_c^i based on p^i is unlike Clauset's method, where m_c^i is chosen such that the KS distance $D_{m_c^i}$ is minimized.

Since the KS distance highly depends on sample size and can hence fluctuates strongly with m_c^i , this approach is more robust than simply minimizing the KS distance. After applying the described algorithm to our incomplete primary catalog, we choose $m_c = 3.6$ with $b \equiv 1.01$ and $p \equiv 0.09$. Figure A.2 illustrates the test for five cases of m_c^i . In Figure A.2(a), theoretical and observed CDFs are depicted as dotted black and red lines, respectively. Light grey and dark grey areas (barely visible for small m_c^i) mark the 5th to 95th, and 25th to 75th percentile of 10,000 randomly sampled CDFs. In Figure A.2(c), the histograms of KS distances of 10,000 randomly sampled CDFs are shown for the same five m_c^i . Red vertical lines mark $D_{m_c^i}$ values. In Figure A.2(b), the evolution of the p -value is shown for all m_c^i participating in the test, while Figure A.2(d) shows the corresponding b -values b^i . Note that $m_c^i = 3.0$ is the lowest value to pass the test. However, because completeness is strictly required for b -value calculation, we choose the more conservative value of 3.6, as it is the smallest m_c^i whose p -value is not exceeded by p -values of later m_c^i .

A.3 Detailed description of declustering methods

For Reasenber and Zaliapin’s declustering methods as well as window methods, we use the python codes developed for the 2020 update of the European Seismic Hazard Model (ESHM20, Danciu et al., 2021) available as an open-source contribution to the seismic hazard modeler’s toolkit of OpenQuake (Weatherill et al., 2014; Pagani et al., 2014). The implementation of the ETAS declustering algorithm is described separately.

Cluster detection is applied to the full time period including 10 years of auxiliary data, but only those mainshocks that fall into the primary time period are considered for further analysis such as b -value estimation. This ensures that aftershocks of events happening before the start of the primary time period are not mistakenly identified as mainshocks. With the exception of ETAS-Background, all methods define a mainshock to be the largest event of a cluster detected by the method. The set of all mainshocks, including events that form a cluster of size 1, make up the declustered catalog.

Reasenber

Reasenber (1985) introduced an algorithm that has been used in numerous studies, e.g. in Ecuador (Beauval et al., 2013) or Afghanistan (Waseem et al., 2019). It defines earthquake interaction zones in space and time. Clusters are built by linking dependent events as follows. If an event lies in the interaction zone of another, the two events are linked. Linked events build a cluster. If an event is linked to an event belonging to a cluster, it is added to the cluster. If two events belonging to different clusters are linked, the two clusters are merged to form one cluster.

Spatial interaction zones are defined around the current as well as the largest event of an ongoing sequence. Their boundaries depend on magnitude, and on the interaction formula that is being used. The two interaction formulae applied here are

$$f_{R_{1985}}(m) = 0.011 \cdot 10^{0.4m}, \quad (\text{A.4})$$

$$f_{W\&C_{1994}}(m) = 0.01 \cdot 10^{0.5m}, \quad (\text{A.5})$$

where $f_{R_{1985}}$ and $f_{W\&C_{1994}}$ are relationships proposed by Reasenber (1985), and an updated version using the scaling relationships by Wells and Coppersmith, 1994, respectively.

Temporal interaction zones are defined based on a probabilistic Omori law approach. The look-ahead time, within which events are being linked, is calculated such that the probability of detecting the next event in the cluster is equal to the parameter p , assuming a lower cut-off magnitude, xm_{eff} , an increase of the lower cut-off magnitude during clusters to $xm_{eff} + x_k \cdot M$, where M is the magnitude of the largest event of the current sequence. r_{fact} represents the number of crack radii around each earthquake within which new events of the cluster are considered (van Stiphout et al., 2012). τ_{min} and τ_{max} are the minimum and maximum allowed look-ahead time.

Table A.1 shows the parameter ranges applied in this study, with standard parameters in bold characters. The ranges and standard settings are adopted from Schorlemmer and Gerstenberger (2007). xm_{eff} is set to be the completeness magnitude m_c .

Window methods

Window methods, as first described by Gardner and Knopoff (1974), are widely used in regional and national seismic hazard models, see Drouet et al. (2020) for France, Akinci et al. (2018) for Italy, Sesetyan et al. (2018) for Turkey, Field et al. (2014) for California (UCERF3), Woessner et al. (2015) for Europe (ESHM13). They define space time windows depending on mainshock magnitude and denote events within the window of a large event as fore- or aftershocks of said mainshock. In contrast to Reasenbergs method, higher level aftershocks are not considered. Furthermore, large events become mainshocks by definition.

Spatial windows are circular around epicenters and vary with magnitude. Temporal windows for foreshocks are per default identical to their corresponding aftershock time window but can be proportionally shortened or extended by multiplication with a factor $p_{\tau_{foreshock}}$. Furthermore, there is the possibility to cap time window length at a maximum number of days τ_{max} . A variety of formulae for the definition of space time window boundaries has been proposed. In our analysis, we use the following three variants.

1. Gardner-Knopoff Window (Gardner and Knopoff, 1974)

$$\Delta_s(m) = 10^{0.1238m+0.983} \quad [km] \quad (A.6)$$

$$\Delta_t(m) = \begin{cases} 10^{0.5409m-0.547}, & \text{if } m < 6.5 \\ 10^{0.032m+0.983}, & \text{if } m \geq 6.5 \end{cases} \quad [days] \quad (A.7)$$

2. Gruenthal window (Gruenthal, 1985; see van Stiphout et al., 2012)

$$\Delta_s(m) = e^{1.77+\sqrt{0.037+1.02m}} \quad [km] \quad (A.8)$$

$$\Delta_t(m) = \begin{cases} \left| e^{-3.95+\sqrt{0.62+17.32m}} \right|, & \text{if } m < 6.5 \\ 10^{2.8+0.024m}, & \text{if } m \geq 6.5 \end{cases} \quad [days] \quad (A.9)$$

3. Uhrhammer window (Uhrhammer, 1986)

$$\Delta_s(m) = e^{-1.024+0.804m} \quad [km] \quad (A.10)$$

$$\Delta_t(m) = e^{-2.87+1.235m} \quad [days] \quad (A.11)$$

Table A.2 shows the parameter ranges applied in this study, with standard window and parameters in bold characters.

Zaliapin

An alternative approach to declustering was proposed in Zaliapin et al. (2008). Space time distances between pairs of events are calculated as

$$\eta_{ij} = \begin{cases} (t_{ij})^\theta \cdot (r_{ij})^d \cdot 10^{-bm_i}, & \text{if } t_{ij} > 0 \\ \infty, & \text{if } t_{ij} \leq 0 \end{cases} \quad (A.12)$$

where $t_{ij} = t_i - t_j$ is the time difference between event i and event j , r_{ij} is the spatial distance between their epicenters, and m_i is the magnitude of the earlier event. d is the fractal dimension of epicenters, θ is an exponent used to weight temporal distance relative to spatial distance, and b is the b -value of the Gutenberg Richter law.

For each event, its nearest neighbor with respect to this distance measure can be identified. Using a Gaussian mixture model, nearest-neighbor space time distances between events are then classified into two categories. Smaller distances are interpreted as distances between dependent events, and larger distances represent distances between independent events. The Gaussian mixture model yields two pairs of mean and standard deviation of distances; one for dependent events, and one for independent events. Nearest neighbor distances are classified as dependent event distances, if their likelihood under the dependent normal distribution is larger than under the independent normal distribution. Earthquake clusters are then defined as tree-like structures of dependent nearest-neighbors.

Table A.3 shows the parameter ranges applied in this study, with standard parameters in bold characters. These ranges were chosen based on recommendations given in the paper introducing the algorithm.

The other two declustering methods are quasi a side product of the ETAS inversion and are therefore described separately.

A.4 ETAS model

A basic epidemic-type aftershock sequence (ETAS, Ogata, 1998) model is used here in two ways. Firstly, it is used to simulate synthetic earthquake catalogs upon which declustering methods are applied to study their effects. Secondly, ETAS provides an alternative, parametric approach to declustering, which was introduced by Zhuang et al. (2002).

Remark: Note here the important distinction between independent events in the ETAS sense, and mainshocks in the declustering sense.

In ETAS, one distinguishes triggered events and events that are not triggered, so-called independent events. Strictly speaking, all earthquakes are triggered by some underlying physical processes. What we mean by “not triggered” in this context is that the event is unlikely to be triggered by another earthquake that was observed in the catalog. This can be because the triggering earthquake was too weak to be detected, or that other physical processes that are not captured in the model were its cause.

Independent events and their aftershocks are modelled as a marked self-exciting point-process as follows. While independent events are assumed to be uniformly distributed in time and space, the aftershock triggering process is modelled to follow three fundamental principles derived from empirical laws.

1. The Utsu aftershock productivity law (Utsu, 1970) describes the number of aftershocks of an event given its magnitude. It describes the aftershock productivity p_{AS} of an event to be exponentially increasing with the magnitude m of the triggering event.

$$p_{AS}(m) = K \cdot e^{a(m-m_c)}. \quad (\text{A.13})$$

2. Aftershock occurrence time is modelled by an exponentially tapered Omori kernel, i.e. aftershock occurrence rate after waiting time t is proportional to

$$q_{AS}(t) = \frac{e^{-t/\tau}}{(t+c)^{1+\omega}}. \quad (\text{A.14})$$

The non-tapered Omori law only allows for exponents larger than 1 in the above denominator, which seems to be an unreasonably strict mathematical condition. This can be bypassed by using the tapered Omori law.

3. Aftershock location is assumed to be distributed isotropically around the epicenter of the triggering event and aftershock occurrence rate r_{AS} decreases with distance as follows.

$$r_{AS}(x, y) = \frac{1}{((x^2 + y^2) + d \cdot e^{\gamma(m-m_c)})^{1+\rho}}, \quad (\text{A.15})$$

where x and y are spatial distances in x - and y -direction between two events. The assumption of isotropy is a simplification which is known to be wrong in reality, as aftershocks tend to occur along fault systems rather than in circles around their triggering events.

In summary, the aftershock triggering rate $g(m)$ is defined as

$$g(t, x, y, m) = \frac{k_0 \cdot e^{a(m-m_c)}}{\frac{(t+c)^{1+\omega}}{e^{-t/\tau}} \cdot ((x^2 + y^2) + d \cdot e^{\gamma(m-m_c)})^{1+\rho}}, \quad (\text{A.16})$$

where m is the magnitude of the triggering event, and $\sqrt{x^2 + y^2}$ and t are the spatial and temporal distance between triggering and triggered event, respectively. $k_0, a, c, \omega, \tau, d, \gamma$ and ρ are constants. Nandan et al. (2019b) show that models allowing these parameters to vary in space, as proposed in Nandan et al. (2017), significantly outperform the forecasting ability of models where they are spatially homogeneous. However, they also show that spatially homogeneous ETAS outperforms three declustering based smoothed seismicity models (SSMs), a simple SSM based on undeclustered data and a model based on strain rate data. For our purpose, spatially homogeneous ETAS is an adequate choice, as our main focus lies on the overall distribution of magnitudes, and not on spatial variations in seismicity. For analogous reasons, possible temporal variability of ETAS parameters is neglected in this analysis.

Furthermore, the constant μ represents the (spatially and temporally) uniform background intensity. The occurrence rate of earthquakes at a given time t and place (x, y) is then described as

$$\lambda(x, y, t) = \mu + \sum_{i:t_i < t} g(t - t_i, x - x_i, y - y_i, m_i), \quad (\text{A.17})$$

where $t_i, m_i, (x_i, y_i)$ are time, magnitude, and location of the i^{th} event. Essentially, the occurrence rate λ at a location (x, y) at time t is the sum of the independent earthquake rate and aftershock rates of all events preceding time t .

A.4.1 ETAS inversion

In order to obtain reasonable estimates for the above-mentioned ETAS parameters, one needs to solve an inversion problem: which set of parameters best describes the observed catalog? In this analysis, we use a so-called Expectation Maximization (EM) algorithm, as proposed by Veen and Schoenberg (2008). They find that, compared to commonly used maximum likelihood estimation, using EM for the inversion of ETAS parameters has substantial advantages in terms of convergence, bias, and robustness to the choice of starting values.

Starting with a set of randomly chosen initial parameters, we repeatedly undergo the expectation step E and the maximization step M, until a convergence criterion is met.

In the expectation step, probabilities p_{ij} of each event e_j to be triggered by each other event e_i are calculated, given the current parameter estimates. Earthquakes in the auxiliary catalog can only serve as triggering events, not as triggered events, while events of the primary catalog can assume both roles. For any given target event, triggering probabilities are proportional to aftershock occurrence rates g_{ij} . The probability of each event e_j to be an independent event (independence probability, p_j^{ind}) is proportional to the background rate. The normalization factor is chosen such that all triggering probabilities and the independence probability sum up to 1.

$$p_{ij} = \frac{g_{ij}}{\mu + \sum_{k:t_k < t_j} g_{kj}}, \quad (\text{A.18})$$

$$p_j^{ind} = \frac{\mu}{\mu + \sum_{k:t_k < t_j} g_{kj}}. \quad (\text{A.19})$$

This choice of normalization factor relies on the implicit assumption that all potential triggering events are captured in the catalog. For this reason, it is essential to ensure completeness of the primary catalog. Events in the auxiliary catalog serve as potential triggering events, in particular of earthquakes at the beginning of the primary time period. They may not take the role of triggered events, and hence independence probabilities can only be defined for primary events. Completeness of the auxiliary catalog is beneficial, but not strictly required.

The sum of independence probabilities of events in the primary catalog yields the expected number of independent events, n_{ind} . Similarly, summing up the triggering probabilities of each triggering event, one obtains its expected number of aftershocks.

In the maximization step, the parameters are optimized to maximize the log likelihood of the observed data, assuming the expected number of independent events and the expected number of aftershocks of each event resulting from the preceding expectation step.

We stop the algorithm as soon as the cumulative absolute difference between parameters of two consecutive maximization steps falls below a threshold of 10^{-3} . Table 2.1 shows the set of parameters obtained.

A.4.2 ETAS simulation

For the simulation of synthetic catalogs, we start by generating independent events. The number of independent events is drawn from a Poisson distribution. Its mean is calculated as the expected number of independent events for the region and time period in question. (Here, the parameter μ is used.) Time and location of each independent event are then drawn from uniform distributions. Especially for event location this is a massive generalization of reality. For the purpose of this analysis however, such a generalization will likely have a minor effect on the results. Clusters are detected depending on distances in-between events. If event locations are more evenly distributed, clusters might more easily be detected. The effect of declustering could therefore be more pronounced in synthetic catalogs compared to real catalogs. Since the purpose of using synthetic catalogs is to reproduce similar effects to those that are observed in real data, an amplification of the effects does not invalidate the argument.

Denoting independent events as events of generation 0, we recursively simulate events of subsequent generations until no more aftershocks are produced, or all aftershocks lie outside the relevant time window. For each event of generation i that lies in the relevant time window, we

- calculate its expected number of aftershocks as

$$n_{AS}(m, \Delta t_0, \Delta t_1) = k_0 \cdot e^{a \cdot (m - m_c)} \cdot \frac{\pi}{\rho} \cdot (d \cdot e^{\gamma \cdot (m - m_c)})^{-\rho} \cdot e^{c/\tau} \cdot \tau^{-\omega} \cdot \left(\Gamma\left(-\omega, \frac{\Delta t_0 + c}{\tau}\right) - \Gamma\left(-\omega, \frac{\Delta t_1 + c}{\tau}\right) \right), \quad (\text{A.20})$$

where $\Gamma(s, x)$ is the upper incomplete gamma function, and $\Delta t_0, \Delta t_1$ are the positive time difference between the event and the start and end of the primary time period, respectively.

- draw its actual number of aftershocks from a Poisson distribution with mean rate $n_{AS}(m, \Delta t_0, \Delta t_1)$.
- simulate aftershock occurrence times by simulating inter-event time differences following the Omori law. For this, we simulate uniformly distributed random numbers between 0 and 1 and apply the inverse of the Omori law CDF with parameters c, ω, τ .
- simulate aftershock locations isotropically around its epicenter by simulating inter-event distance and angle. Again, we simulate uniformly distributed random numbers between 0 and 1 and apply the inverse of the CDF induced by the spatial kernel with parameters d, γ, ρ , yielding the distance. The inter-event angle is sampled from a uniform distribution on the interval $[0, 2\pi)$.
- simulate aftershock magnitudes by simulating uniformly distributed random numbers and applying the inverse of the CDF induced by the GR-law with fixed b -value.

Because the time window is limited, more and more aftershocks will be simulated to lie outside the relevant period and hence the algorithm stops after having generated a finite number of earthquakes. Note that the time period is not required to be limited for the ETAS simulation to stop after a finite number of generations. One can calculate the so-called branching ratio

$$\eta = \int_{m_c}^{\infty} p(m) \cdot n_{AS}(m, 0, \infty) dm \quad (\text{A.21})$$

where $p(m) = \beta \cdot e^{-\beta \cdot (m - m_c)}$ is the probability that an earthquake of magnitude above m_c has magnitude m , and $\Gamma(-\omega, \infty) := 0$ in the calculation of $n_{AS}(m, 0, \infty)$. This branching ratio represents the average number of expected direct aftershocks of any earthquake. If $\eta < 1$, the sum of the geometric series of higher order aftershock numbers converges, and hence is finite. In other words, the ETAS process is in a subcritical regime. Therefore, the expected number of generations to be simulated before all events stop producing aftershocks, is also finite. Based on the ETAS parameters obtained in the inversion, we find that $\eta \equiv 0.89 < 1$, implying that simulations of aftershock chains will stop after finitely many generations.

Because the time delay between a triggering event and its aftershock can be in the order of months or even years, aftershock chains starting with independent events prior to the relevant time window are missing in such simulations. The characteristic waiting time τ is around 26.7 years in our case. Since aftershock chains can go on for many generations, we

choose the start of our simulation period generously early on January 1, 1850, 120 years before the start of the auxiliary time period used for further analysis.

A.4.3 ETAS declustering

It has been proposed by Zhuang et al. (2002) that ETAS can also be used for declustering. We consider two versions of ETAS-based declustering, which differ in their definition of mainshocks. As was mentioned above, one component of ETAS inversion is to calculate probabilities p_{ij} of each event e_j to be triggered by each other event e_i , and the probability p_i^{ind} of each event to be an independent event.

Regarding independent events to be mainshocks, one straightforward way to decluster is to weight each event in the catalog by its independence probability. The b -value can then be calculated using weighted magnitudes. We call this first version of ETAS declustering ‘ETAS-Background’.

However, declustering methods normally distinguish between mainshocks and fore- or aftershocks, rather than independent events and triggered events in the ETAS sense. Going from independent and triggered events to mainshocks, fore- and aftershocks requires the extra step of identifying earthquake clusters and then imposing the rule of maximum magnitude to identify mainshocks. The probabilities p_{ij} and p_i^{ind} can be utilized to identify clusters, yielding an additional ETAS-based method of declustering.

We split our primary catalog into independent and triggered events as follows. Those events with the highest independence probabilities are defined to form the set E_{ind} of independent events. The number of independent events is chosen to be $[n_{ind}]$, the closest integer to n_{ind} .

$$E_{ind} = \{e_i \mid p_i^{ind} \geq p_{thresh}\}, \quad (\text{A.22})$$

where p_{thresh} is the maximum threshold such that E_{ind} contains $[n_{ind}]$ events. Each independent event forms its own cluster. All remaining events are then chronologically added to one of the existing clusters. More precisely, each triggered event is added to the cluster with the highest responsibility for having triggered it. The responsibility r_{C_i} for cluster C_i for having caused event e_j is given as the cumulative triggering probability of C_i ,

$$r_{C_i} = \sum_{k:e_k \in C_i} p_{kj}. \quad (\text{A.23})$$

Events of the auxiliary catalog are allowed to have responsibility for a primary event. As there is no information on triggering probabilities p_{ij} between auxiliary events, they are not seen as part of any cluster, which is why each auxiliary event is interpreted as a separate cluster. In this way, primary events may be allocated to a cluster originating in the auxiliary time period. These events are excluded from further analysis, since they are assumed to be incomplete.

The ETAS method for declustering does not depend on any input parameters. Unlike with other methods, the optimal parameters for declustering can be naturally obtained by calibrating the model on the data. ‘ETAS-Main’, in agreement with the other three methods, defines the largest event of each cluster to be the mainshock of the cluster.

Note that the non-parametric stochastic declustering algorithm proposed by Marsan and Lengline (2008) is not used here. This is because of its similarity to the already considered

parametric stochastic declustering alternative provided by the ETAS model. The main difference to ETAS declustering is that the triggering rate, here denoted by $g(t, x, y, m)$, is there obtained empirically, without presuming the laws (A.13-A.15). In their analysis of southern California seismicity, they observe that their empirically derived triggering rate follows laws similar to those described in (A.13-A.15), which, likewise, were originally discovered empirically.

A.5 Sensitivity of b -value to m_c

A sensitivity analysis of the b -value to the completeness magnitude m_c , represented in Figure A.3, shows that the b -value decrease after declustering is an effect that is observed regardless of the choice of m_c . For the low value of $m_c = 3.2$, Reasenber and Zaliapin declustered catalogs' b -values do not significantly differ from the full catalog b -value. However, completeness at this value is not certain, especially not after large events. As we increase the completeness threshold and therewith the certainty of dealing with a complete catalog, b -values quickly start to differ substantially from the full-catalog b -value. Note that the extent of the decrease is characteristic of each method.

A.6 Supplemental figures

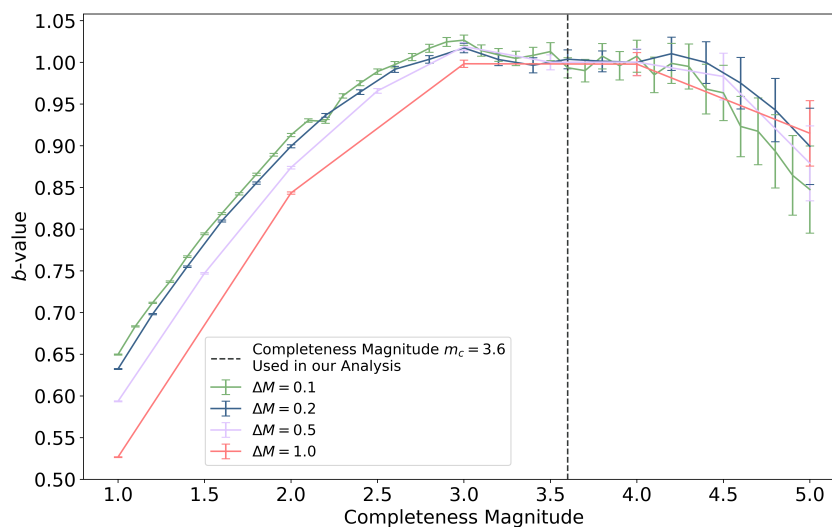


Figure A.1: Sensitivity of b -value on completeness magnitude and magnitude bin size ΔM .

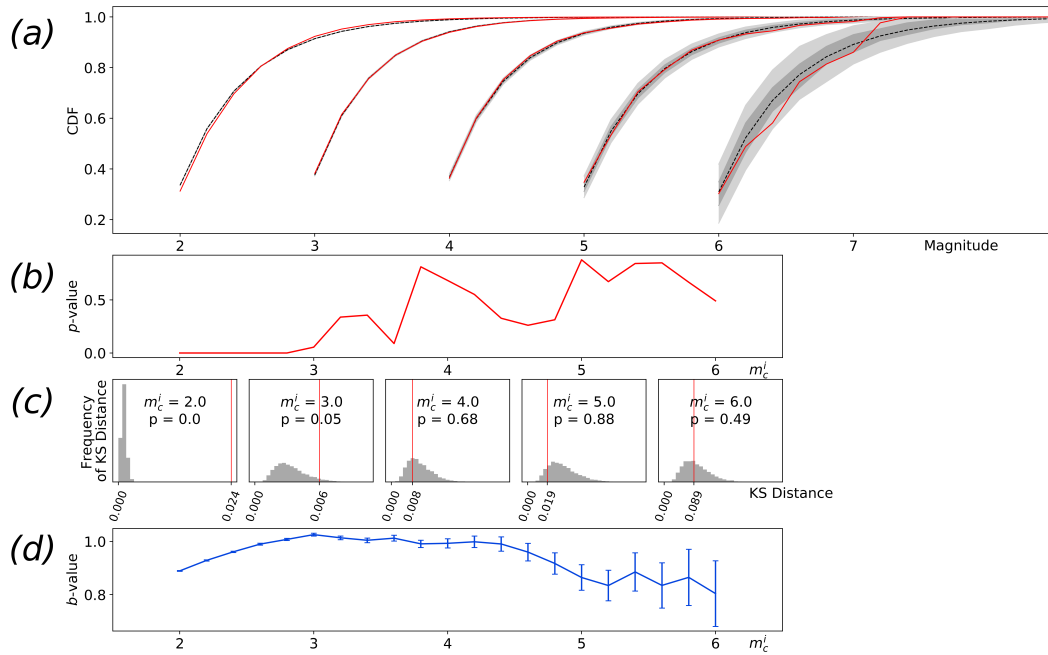


Figure A.2: Testing m_c^i between 2.0 and 6.0. (a): For integer values of m_c^i , theoretical and observed CDFs are represented as dotted black and red lines, respectively. Light grey and dark grey areas mark 5th to 95th, and 25th to 75th percentile of 10,000 randomly sampled CDFs. Grey areas are barely visible for small m_c^i , because they are in agreement with the theoretical CDF. (b): Evolution of p-value with m_c^i . (c): Derivation of p-value for integer values of m_c^i . Histogram shows the frequency distribution of KS distances of 10,000 randomly sampled CDFs to theoretical CDF. Red vertical line indicates KS distance of observed to theoretical CDF. (d) Evolution of b -value with m_c^i , with uncertainties.

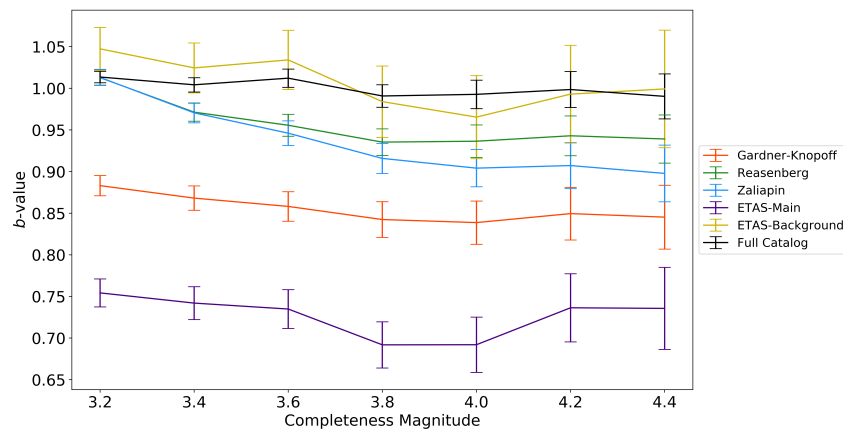


Figure A.3: Sensitivity of mainshock b -value on completeness magnitude for different declustering methods.

A.7 Supplemental tables

Table A.1: Parameter ranges applied in the Reasenbergl declustering algorithm. Standard parameter settings are given in bold characters.

Parameter Name	Value Range
Interaction Formula	Reasenbergl ₁₉₈₅ , WellsCoppersmith ₁₉₉₄
τ_{min}	0.5, 1.0 , 1.5, 2.0, 2.5
τ_{max}	3.0, 6.0, 8.0, 10.0 , 11.0, 13.0, 15.0
x_k	0.0, 0.167, 0.333, 0.5 , 0.667, 0.833, 1.0
p	0.75, 0.8, 0.85, 0.9, 0.95
r_{fact}	5.0, 10.0 , 15.0, 20.0

Table A.2: Parameter ranges and window types applied in the window declustering methods. Standard window and parameter settings are given in bold characters.

Parameter Name	Value Range
Window Method	GardnerKnopff , Gruenthal, Uhrhammer
τ_{max}	None , 15.0, 30.0
$p_{\tau_{foreshock}}$	0.0, 0.1, 0.2, ..., 1.0 , 1.1, 1.2, ..., 1.9, 2.0

Table A.3: Parameter ranges applied in the Zaliapin declustering algorithm. Standard parameter settings are given in bold characters.

Parameter Name	Value Range
d	1.0, 1.1, 1.2, 1.3, 1.4 , ..., 2.4, 2.5
b	0.0, 0.8, 0.85, 0.9, 0.95, 1.0 , 1.05, ..., 1.45, 1.5
θ	0.8, 0.85, 0.9, 0.95, 1.0 , 1.05, 1.1, 1.15, 1.2

APPENDIX B

SUPPORTING INFORMATION FOR CHAPTER 3

B.1 Catalog simulation

The following algorithm is used to simulate the continuation of a training catalog.

Note that the synthetic catalogs referred to in Sections 3.4.1 and 3.4.2 in Chapter 3 are not continuations of a training catalog, hence generation 0 (defined in the following) consists only of background events. The locations of these background events are uniformly distributed in the study region. Also in the case of synthetic catalog simulation, the “testing period”, which is referred to below, is the period for which one wishes to simulate a catalog. Where different models are mentioned, the base model is used for synthetic catalog simulation.

1. Background events are simulated for the testing period.
 - Number of background events is drawn from a Poisson distribution with mean as given by the ETAS background rate.
 - Occurrence times are drawn from a uniform distribution within the testing period.
 - Locations are drawn from the locations of events in the training catalog, weighted by their probability of being background events. The locations are then randomly displaced by a distance drawn from a normal distribution with mean 0 and standard deviation of 0.1° .
 - Magnitudes are drawn from a GR law with exponent β as estimated in the PETAI inversion (for PETAI and trig_only). For the base model and par_only, we use the β estimate obtained when using the formula proposed by Tinti and Mulargia (1987) for binned magnitude values, using magnitudes $M \geq 3.1$ in the training catalog.
2. The training catalog together with the simulated background events make up generation 0. $i_{gen} := 0$.
3. Expected number of aftershocks is calculated for all events of generation 0. In the case of the PETAI and the trig_only model, the average number of aftershocks triggered by any event e_i in the training catalog is inflated by $1 + \xi(t_i)$.
4. Actual number of aftershocks of each event is randomly drawn from a Poisson distribution with mean as calculated in the previous step.
5. Aftershocks of the current generation i_{gen} are simulated.
 - Aftershock time distance to its parent event is randomly generated according to the estimated ETAS time kernel. If aftershock time falls out of the testing period, this aftershock is discarded.

- Aftershock spatial distance to its parent event is randomly generated according to the estimated isotropic ETAS spatial kernel. If aftershock location falls out of the considered polygon, this aftershock is discarded.
 - Aftershock magnitude is generated according to the GR law with exponent β (same as for the background events).
6. The newly generated aftershocks now make up the next generation $i_{gen} + 1$.
 7. We move on to the next generation. $i_{gen} := i_{gen} + 1$
 8. Expected number of aftershocks is calculated for all events of generation i_{gen} . Continue with step 4.

The algorithm terminates when no aftershocks fall into the testing period anymore, which is expected to happen in a finite amount of time if the branching ratio $\eta < 1$.

B.2 Forecast evaluation

The performance of each model is evaluated by calculating the log-likelihood of the testing data given the forecast. Specifically, we calculate the log-likelihood of N_i earthquakes to occur in each bin b_i of a spatial grid of 0.1° latitude \times 0.1° longitude. Here, N_i is the number of earthquakes that actually occurred during the testing period in spatial bin b_i .

The log-likelihood for b_i is calculated based on the smoothed estimate of the probability of N_i earthquakes to occur in b_i , where the probability estimate is based on the 100,000 simulations of the model in question. For smoothing we use Gaussian kernels with adaptive bandwidth as described by Nandan et al. (2019a), with a fixed value of $\Omega = 3.0$. To avoid arbitrary likelihood values due to extrapolation, we define a water-level likelihood for event counts larger than the maximum simulated event count in the respective bin. This waterlevel probability is defined as a uniform value of $100,001^{-1}/n_{extr}$, where n_{extr} is the number of event counts larger than the maximum observed and smaller than a generously high maximum possible event count. Symbolically, this suggests that all other possible event counts could have been simulated in the $100,001^{st}$ simulation. Inevitably, the probabilities for non-extrapolated event counts are proportionally reduced such that the probabilities of all possible event counts add up to 1.

The total log-likelihood of the testing data is then given by the sum of log-likelihoods over all bins b_i .

Two competing models can be compared by calculating the information gain (IG) of the alternative model M_{alt} over the null model M_0 , which is simply the difference in log-likelihood of observing the testing data. The mean information gain (MIG) is calculated as the mean over all testing periods. We accept the superiority of M_{alt} over M_0 when we reject the null hypothesis that M_{alt} does not outperform M_0 . To decide whether to reject the null hypothesis, we perform a one-sided t-test on the set of IGs for all testing periods. This assumes that the IG values follow a normal distribution, and we test whether or not the MIG of M_{alt} versus M_0 is significantly positive. We reject the null hypothesis when a p -value of less than 0.05 is observed.

The spatially and temporally homogeneous Poisson process (STHPP) model forecasts the same number of events in all spatial bins. This number of events forecasted, N_{fc} , is given by

$$N_{fc} = \frac{N_{train} \cdot T_{test}}{T_{train} \cdot N_{bins}}, \quad (\text{B.1})$$

where n_{train} is the number of events observed in the training period, T_{train} is the length of the training period in days, $T_{test} = 30$ is the testing period length, and N_{bins} is the total number of spatial bins. The log-likelihood for the STHPP model is calculated assuming a Poisson distribution of event numbers with mean N_{fc} in each spatial bin.

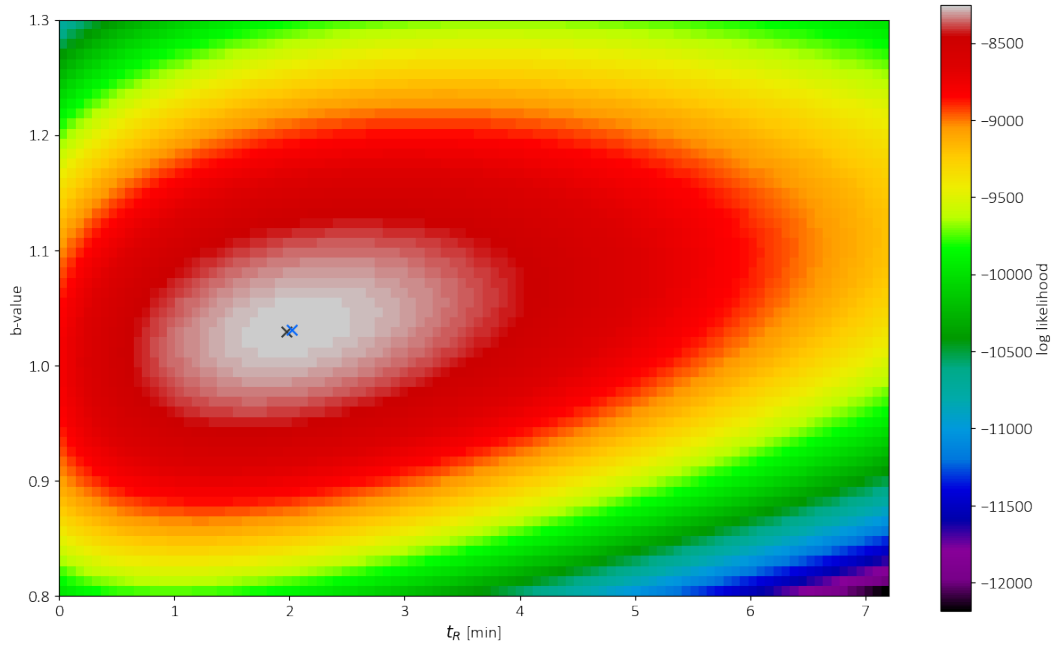


Figure B.1: Log likelihood of observing the test data for different values of t_R and b -value, when current rate is known. Black cross indicates true values used in simulation, blue cross indicates maximum likelihood estimators obtained.

APPENDIX C

SUPPORTING INFORMATION FOR
CHAPTER 4

C.1 Supplemental figures

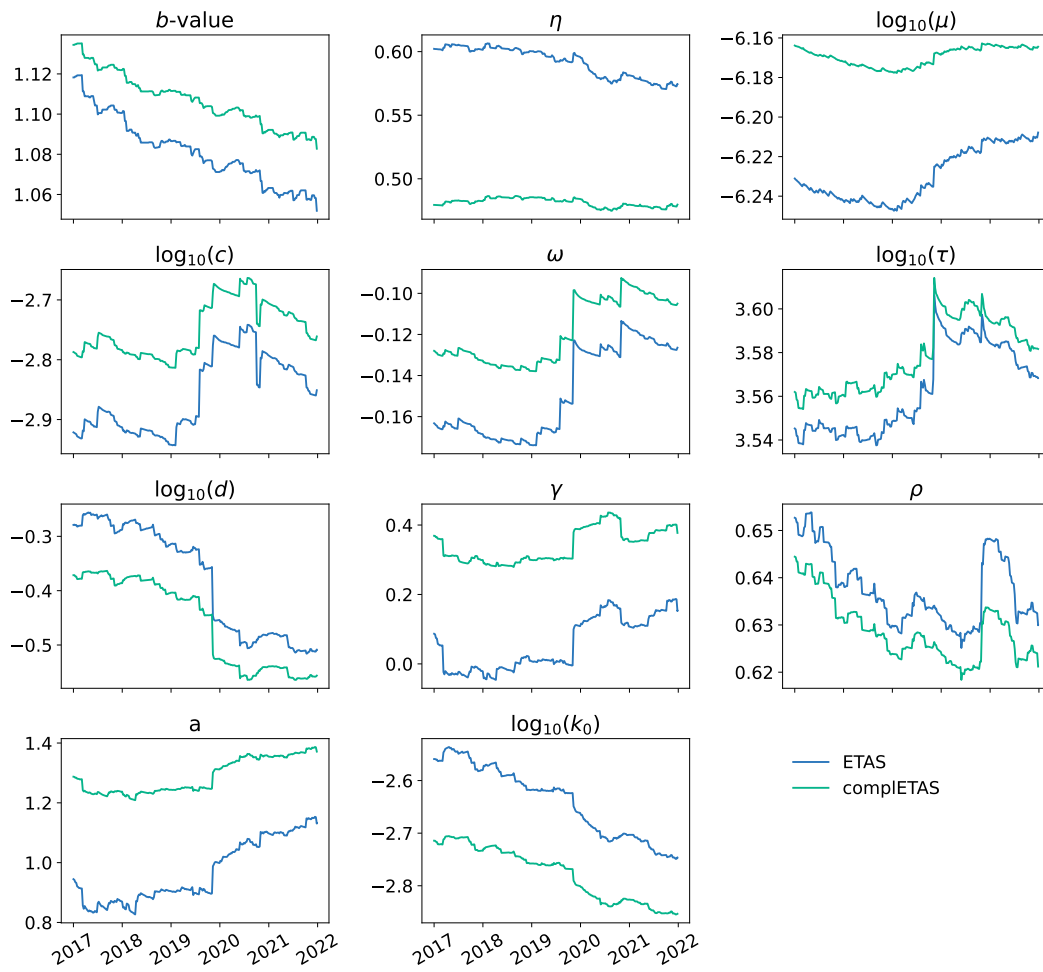


Figure C.1: Evolution of b -value, branching ratio η , and ETAS parameters when an increasing time horizon is used for calibration. For standard ETAS, start of the primary time horizon is January 1, 1997, for completETAS, it is January 1, 1977. End of time horizon is indicated on x -axis.

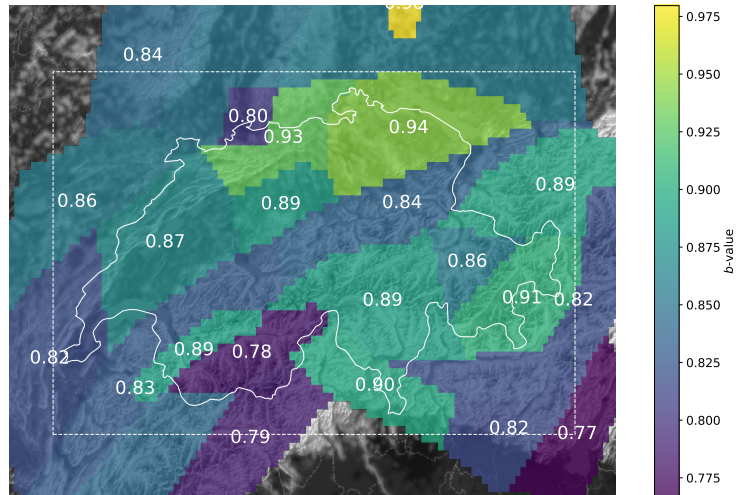


Figure C.2: b -values used to extrapolate the rates of $M \geq 2.5$ earthquakes in different SEIS-15 zones. In non-coloured regions, SUIhaz2015 is not defined. Dashed line marks the region in which the ETAS model is defined.

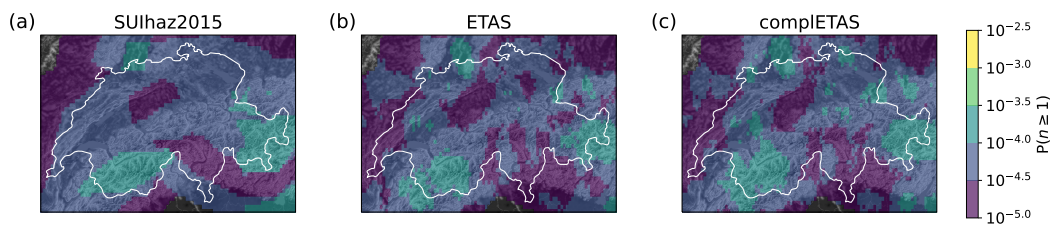


Figure C.3: 7-day background seismicity forecasts produced by the different models. The color of each pixel of $0.05^\circ\text{lat} \times 0.05^\circ\text{lon}$ ($\simeq 5\text{km} \times 5\text{km}$) corresponds to the probability of one or more $M \geq 2.5$ earthquakes to occur in the pixel during 7 days.

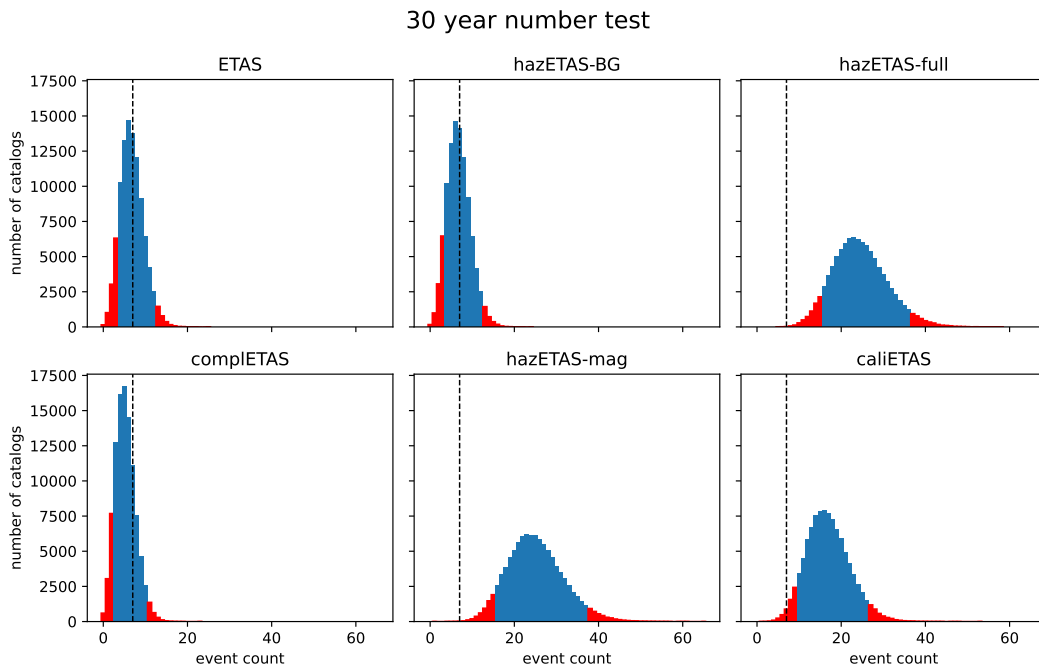


Figure C.4: Results of the 30 year number test. Histograms show the number of $M \geq 4.5$ events in 100,000 simulated catalogs, vertical dashed line indicates the observed number of 7 events. Red areas of the histograms mark event numbers below the 5th or above the 95th percentile for which the test is failed.

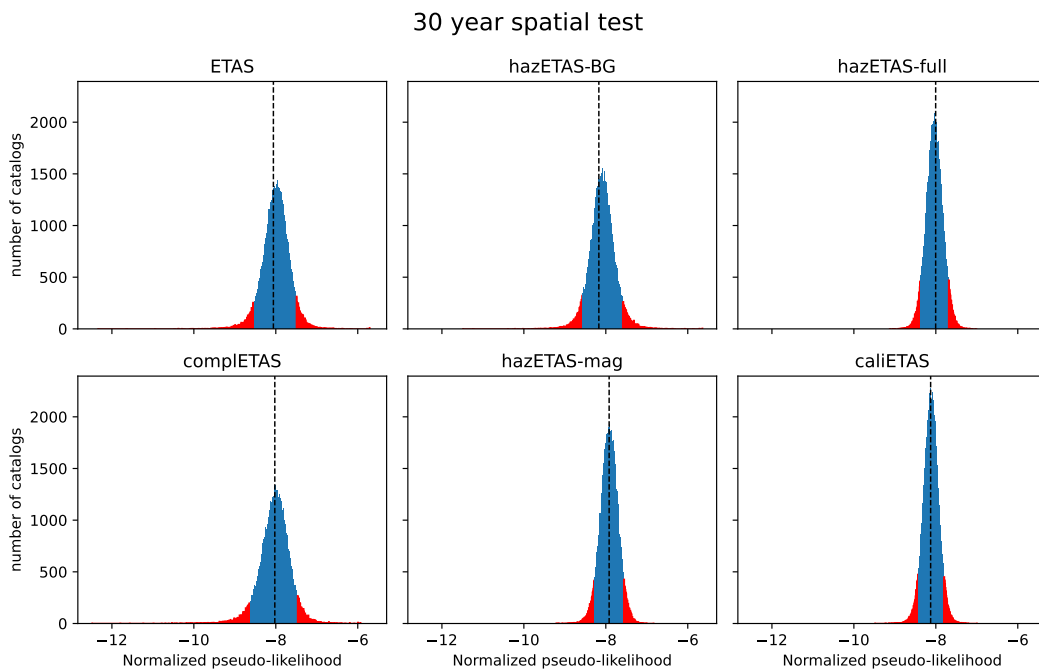


Figure C.5: Results of the 30 year spatial test. Histograms show distribution of the test statistic in 100,000 simulated catalogs, vertical dashed line indicates the observed statistic. Red areas of the histograms mark values below the 5th or above the 95th percentile for which the test is failed.

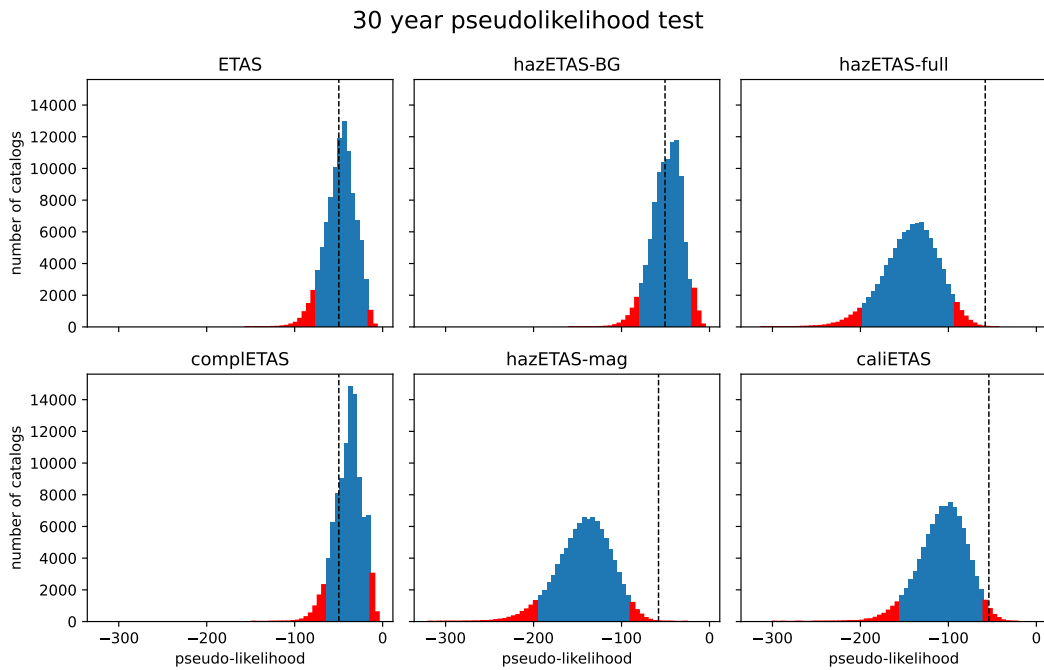


Figure C.6: Results of the 30 year pseudolikelihood test. Histograms show distribution of the test statistic in 100,000 simulated catalogs, vertical dashed line indicates the observed statistic. Red areas of the histograms mark values below the 5th or above the 95th percentile for which the test is failed.

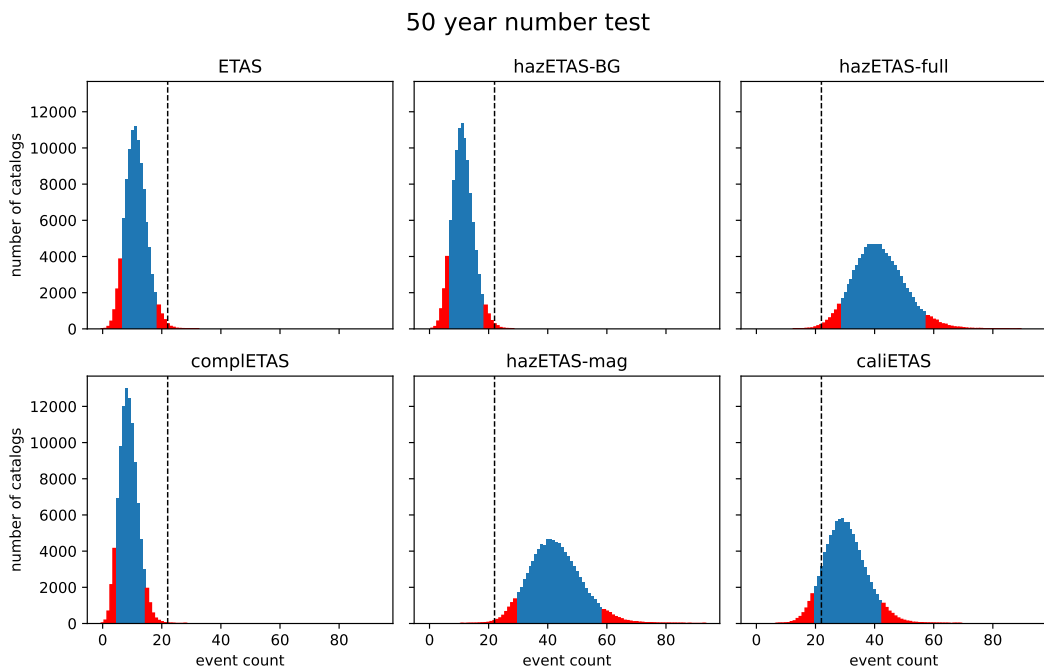


Figure C.7: Results of the 50 year number test. Histograms show the number of $M \geq 4.5$ events in 100,000 simulated catalogs, vertical dashed line indicates the observed number of 22 events. Red areas of the histograms mark event numbers below the 5th or above the 95th percentile for which the test is failed.

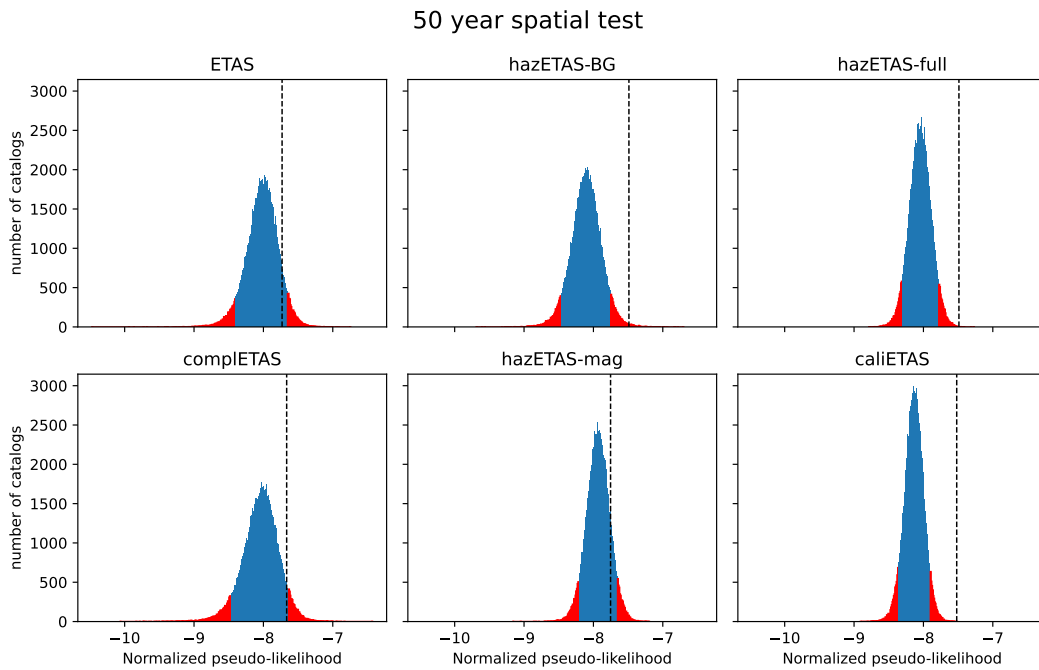


Figure C.8: Results of the 50 year spatial test. Histograms show distribution of the test statistic in 100,000 simulated catalogs, vertical dashed line indicates the observed statistic. Red areas of the histograms mark values below the 5th or above the 95th percentile for which the test is failed.

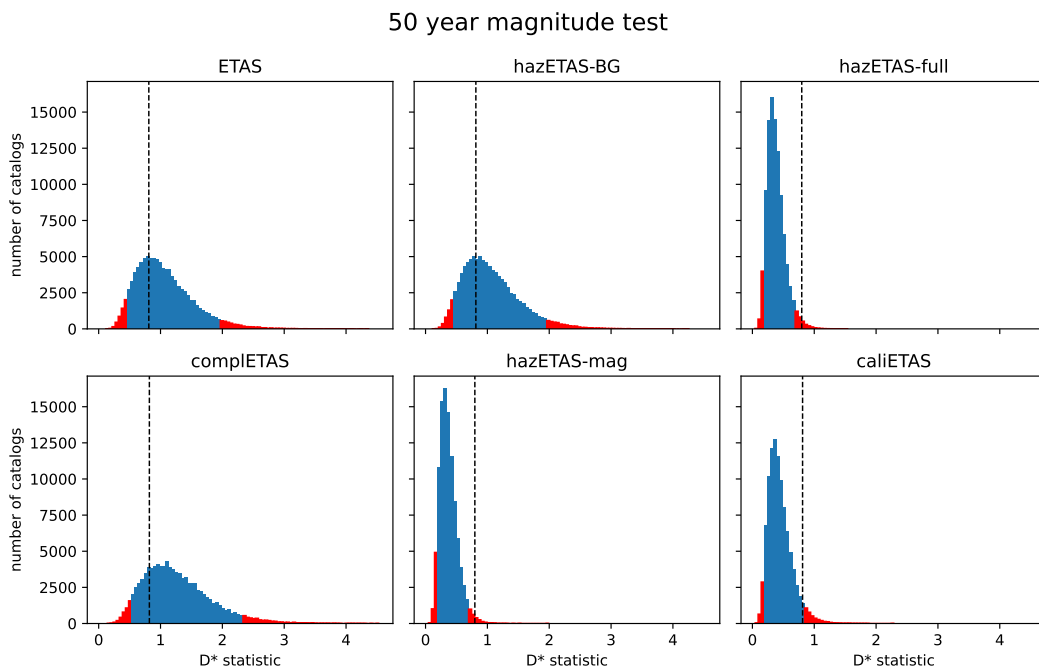


Figure C.9: Results of the 50 year magnitude test. Histograms show distribution of the test statistic in 100,000 simulated catalogs, vertical dashed line indicates the observed statistic. Red areas of the histograms mark values below the 5th or above the 95th percentile for which the test is failed.

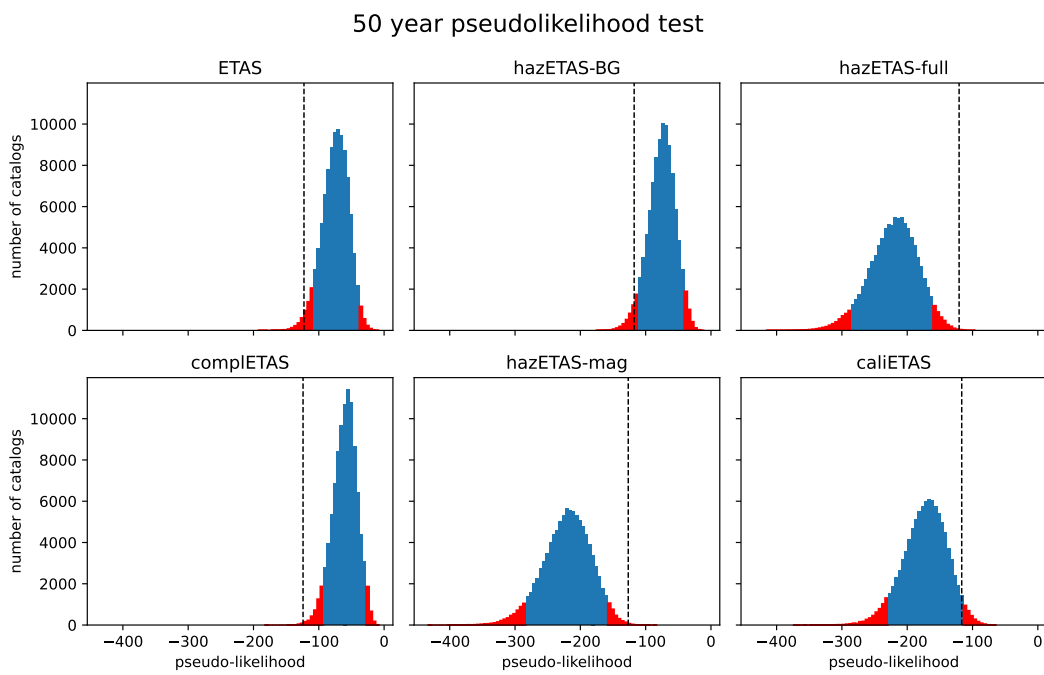


Figure C.10: Results of the 50 year pseudolikelihood test. Histograms show distribution of the test statistic in 100,000 simulated catalogs, vertical dashed line indicates the observed statistic. Red areas of the histograms mark values below the 5th or above the 95th percentile for which the test is failed.

APPENDIX D

SUPPORTING INFORMATION FOR CHAPTER 5

D.1 Polygons

The polygons used in this study are defined via the following lists of vertices.

Table D.1: Southern California polygon boundary vertices.

latitude	longitude
32.7219	-116.3004
33.7424	-117.6512
33.7958	-117.966
33.9322	-118.0775
34.0984	-118.2611
34.1755	-118.9365
34.6027	-118.8775
34.8281	-119.343
36.525	-119.1988
36.4835	-115.6381
34.128	-115.5463
32.7219	-115.2578
32.6922	-115.448
32.7753	-115.7234
32.8109	-115.8545

Table D.2: Italy polygon boundary vertices.

latitude	longitude
45.1	4.9
44.5	5.1
43.3	5.9
42.8	6.5
41.6	9.1
38.0	10.5
36.7	11.5
35.8	13.4
35.3	15.1
35.7	16.1
38.8	19.4
40.1	20.1
41.3	19.5
42.9	17.2
44.0	15.6
45.6	15.6
46.5	15.4
47.5	14.7
47.9	13.7
48.1	13.2
48.4	12.2
48.2	10.7
47.9	9.4
47.8	8.4
46.8	5.8
45.8	5.1
45.1	4.9

D.2 Inverted parameters

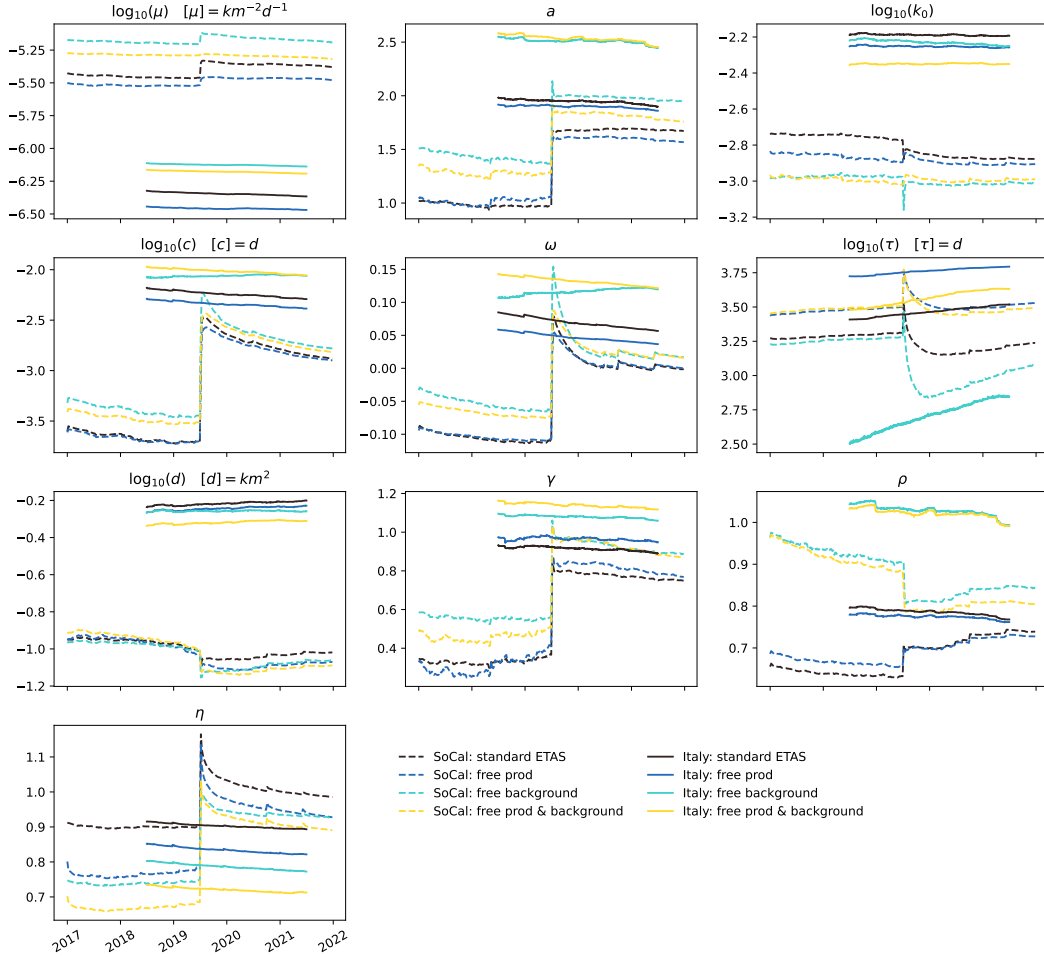


Figure D.1: Evolution of inverted parameters with increasing length of the training catalog, for the four ingredient models. The branching ratio η is not individually inverted, but is calculated from the other parameters. Dashed lines reflect Southern California parameters, solid lines reflect Italian parameters.

Figure D.1 shows the inverted parameters for the four ingredient models, with an increasing time horizon used for the calibration, for Southern California and Italy. Note that for the standard ETAS model and fETAS where only the background rate is free, the parameters a and k_0 are inverted directly during expectation maximization (EM), while for the fETAS models with free productivity, they are inferred afterwards based on the κ_j values that result from the EM inversion.

Most parameters show remarkable changes in time in Southern California, and generally, the parameters differ between Italy and Southern California. The differences between parameters obtained for different ingredient models show similar trends in both regions.

For instance, the background rate μ is highest for the model which only allows the background rate to be free, followed by the model where background and productivity are free,

and is lowest when only the productivity is free. This is expected, since allowing the background to be free will allow the model to classify more events to be background events, while allowing the productivity to be free will allow it to classify more events to be aftershocks.

The exponent of the productivity law, a , is larger in the fETAS models which allow the background to be free, indicating a stronger magnitude dependency of the number of aftershocks an earthquake is expected to generate. Those models also have larger γ and much larger ρ values, which translates to a stronger magnitude dependency of the spatial region in which aftershocks occur, and a stronger spatial decay of the aftershock rate.

Interestingly, the fETAS model in which only productivity is free shows smaller k_0 values than standard ETAS in both regions, accompanied by values of a that are similar to standard ETAS. Both these effects would suggest lower overall productivity. However, the value of τ is larger in this model, indicating a slower long-term tapering off of aftershock rate in time, and ω is smaller in Southern California (similar in Italy), further indicating a slower (similar) temporal decay of aftershock rate. Together with the observation that μ is smaller for this model, these results suggest that allowing productivity to free leads to an overall slower decay of aftershock rate, and thus a large fraction of aftershocks is expected to occur later in an ongoing sequence.

The branching ratio η , which captures the average expected number of aftershocks of any event, is highest for the standard ETAS model, followed by fETAS with free productivity, fETAS with free background, and fETAS with free productivity and background with the lowest branching ratio. Thus, the degree of flexibility of a model is qualitatively opposite to the degree of criticality of the system that is inferred with that model.

APPENDIX E

GITHUB REPOSITORIES

E.1 `lmizrahi/etas`

The GitHub repository [lmizrahi/etas](#) contains Python implementations of the most central algorithms used in this thesis. Figure E.1 shows a screenshot of the repository, which among other information displays its `README.md` file. The repository contains scripts which allow the user to run the following procedures.

- Jointly estimate a constant completeness magnitude m_c and corresponding b -value given a set of earthquake magnitudes, using the method described in Section A.2. Running `runnable_code/estimate_mc.py` will apply the method to the example magnitudes stored in `input_data/magnitudes.npy`.
- Invert ETAS parameters of an earthquake catalog, given certain input parameters such as the primary and auxiliary time period, completeness magnitude, regional polygon, etc., using the method described in Section A.4.1. Input parameters are specified in `config/invert_etas_config.json`. Running `runnable_code/invert_etas.py` will invert the parameters of the catalog stored in `input_data/example_catalog.csv`, and store the results in `output_data/`.
- Invert ETAS parameters of an earthquake catalog, given certain input parameters and a spatially and/or temporally varying magnitude of completeness, using the method described in Section 3.3.1. Input parameters are specified in `config/invert_etas_mc_var_config.json`. After uncommenting lines 31-32, running `runnable_code/invert_etas.py` will invert the parameters of the catalog stored in `input_data/example_catalog_mc_var.csv`, and store the results in `output_data/`. The varying completeness magnitude is given as an extra column named `mc_current` in the input catalog, where each row takes the value of m_c which is valid at the time and location of the event represented by that row.
- Simulate a synthetic catalog as described in Section A.4.2. The region, time horizon and minimum simulated magnitude, as well as ETAS parameters used for the simulation are specified in `config/simulate_catalog_config.json`. Running `runnable_code/simulate_catalog.py` will store a synthetic catalog in `output_data/`.
- Simulate a continuation of a catalog using the algorithm described in Section B.1, after its ETAS parameters have been inverted using one of the previously described scripts. The information for the duration of the catalog continuation, and the paths to the inversion output, are specified in `config/simulate_catalog_continuation_config.json`. Running `runnable_code/simulate_catalog_continuation.py` will store a simulated catalog in `output_data/`.
- Issue a pseudo-prospective forecast consisting of 100 simulated catalogs for Switzerland. Running `runnable_code/ch_forecast.py` will first invert the ETAS parameters

based on the catalog stored in `input_data/ch_catalog.csv`, and then simulate 100 continuations thereof which will be stored in `output_data/`. This is the basic ETAS model used in Section 4.4.1. Relevant input is specified in `config/ch_forecast_config.json`. Note that more than 100 simulations are recommended for a reliable forecast; this is just to limit the run time of the script when first trying it out.

E.2 [lmizrahi/petai](#)

The repository [lmizrahi/petai](#) contains scripts which allow the user run the PETAi inversion described in Section 3.3.4 and simulate a catalog continuation. Running `invert_petai.py` will run the inversion on a synthetic example catalog stored in `data/synthetic_catalog.csv`, and store the inverted ETAS and detection incompleteness in `data/`. Thereafter running `simulate_catalog_continuation.py` will store a catalog continuation as `data/my_catalog_continuation.csv`.

Imizrahi / etas (Public)

<> Code Issues Pull requests Actions Projects Security Insights

main 1 branch 5 tags Go to file Code

Imizrahi Merge pull request #18 from swiss-seismological-service/feature/appro... 2ac9a5c on 6 Sep 132 commits
config fix: default for ch forecast should be without free background or pro... 3 months ago
etas Merge pull request #18 from swiss-seismological-service/feature/ap... 2 months ago
input_data adding a script to run a forecast for switzerland, including necessar... 4 months ago
runnable_code doc: add comment on ho to store simulations instead of just poducing 3 months ago
.gitignore feat: create folder when storing data in case it doesn't exist yet 4 months ago
LICENSE Initial commit 2 years ago
README.md docs: updating resdme file and authors in setup.cfg 3 months ago
pyproject.toml build: add setuptools config files 4 months ago
requirements.txt docs: move dependency pins to requirements.txt 4 months ago
setup.cfg docs: updating resdme file and authors in setup.cfg 3 months ago

README.md

ETAS: Epidemic-Type Aftershock Sequence

DOI [10.5281/zenodo.6951562](https://doi.org/10.5281/zenodo.6951562)

This code was written for the following articles:

Leila Mizrahi, Shyam Nandan, Stefan Wiemer 2021;
The Effect of Declustering on the Size Distribution of Mainshocks.
Seismological Research Letters; doi: <https://doi.org/10.1785/0220200231>

The option for (space-time-)varying completeness magnitude in the parameter inversion is described in:

Leila Mizrahi, Shyam Nandan, Stefan Wiemer 2021;
Embracing Data Incompleteness for Better Earthquake Forecasting. (Section 3.1)
Journal of Geophysical Research: Solid Earth; doi: <https://doi.org/10.1029/2021JB022379>

To cite the code, please use its DOI, and if appropriate, please cite the article(s).
For more documentation on the code, see the (electronic supplement of the) articles.
For Probabilistic, Epidemic-Type Aftershock Incompleteness, see PETAL.
In case of questions or comments, contact me: leila.mizrahi@sed.ethz.ch.

Contents:

- runnable_code/ scripts to be run for parameter inversion or catalog simulation
 - ch_forecast.py estimates ETAS parameters and creates 100 simulations using the Swiss catalog
 - estimate_mc.py estimates constant completeness magnitude for a set of magnitudes
 - invert_etas.py calibrates ETAS parameters based on an input catalog (option for varying mc available)
 - simulate_catalog.py simulates a synthetic catalog
 - simulate_catalog_continuation.py simulates a continuation of a catalog, after the parameters have been inverted. if you run this many times, you get a forecast. **this only works if you run invert_etas.py beforehand.**
- config/ configuration files for running the scripts in runnable_code/
 - names should be self-explanatory.
- input_data/ input data to run example inversions and simulations

About

calibrate ETAS, simulate using ETAS, estimate completeness magnitude & magnitude frequency distribution

#seismology #earthquakes #statistical-models

Readme MIT license 50 stars 6 watching 15 forks

Releases 5

3.1: Merge pull request #10 fr... on 2 Aug (Latest)

+ 4 releases

Packages

No packages published

Contributors 2

Imizrahi Leila Mizrahi schmidni Nicolas Schmid

Languages

Python 100.0%

Figure E.1: Screenshot of the Imizrahi/etas repository on GitHub (last accessed on 15 September 2022).

LEILA MIZRAHI

PhD candidate, Swiss Seismological Service, ETH Zurich
Sonneggestrasse 5, 8092 Zurich, Switzerland
leila.mizrahi@sed.ethz.ch

Education

PhD in Statistical Seismology ETH Zurich (Switzerland), Swiss Seismological Service Visiting scholar at University of Southern California (Sep 2021-Jan 2022)	<i>Jul 2019-Nov 2022</i>
MSc in Mathematics University of Zurich (Switzerland) Master's Thesis: "Thoroughly Formalizing an Uncommon Construction of the Real Numbers"	<i>Feb 2014-Sep 2015</i>
BSc in Mathematics University of Zurich (Switzerland)	<i>Sep 2010-Feb 2014</i>

Professional Experience

Actuary Methods & Processes (Assistant Vice President) Swiss Re Ltd., Underwriting Strategy department, Zurich, Switzerland	<i>Mar 2019-Jun 2019</i>
Underwriting Strategy Graduate (graduates@swissre program) Swiss Re Ltd., Underwriting Strategy department, Zurich, Switzerland	<i>Sep 2017-Feb 2019</i>
Modelling Intern Swiss Re Ltd., Underwriting Strategy department, Zurich, Switzerland	<i>Feb 2016-Jan 2017</i>

Other Relevant Experience

Lecturer Department of Earth Science, ETH Zurich (Switzerland) Integrated Practical Course on Seismic Networks and Data	<i>May-Jun 2022</i>
Co-Convener SSA Annual Meeting Session: New Developments in Physics- and Statistics-based Earthquake Forecasting	<i>Apr 2022</i>
Reviewer Seismological Research Letters Earth and Planetary Science Letters Nature Communications Earth & Environment	<i>Jun 2021-present</i>
On-call Duty Seismologist Swiss Seismological Service, ETH Zurich (Switzerland) Communicate with Swiss authorities and with the public in case earthquakes in Switzerland or abroad.	<i>Jun 2020-present</i>
Teaching Assistant Department of Earth Science, ETH Zurich (Switzerland) Statistical Data Analysis with Matlab, Geophysical Field Course on Seismic Refraction	<i>Sep 2019-present</i>
Fieldwork Hengill geothermal area (Iceland), Swiss Seismological Service & Reykjavik Energy Assisted with the installation of a 500 node seismic array.	<i>Jun 2021</i>
Fieldwork Hengill geothermal area (Iceland), Swiss Seismological Service & ISOR Icelandic Geosurvey Assisted with the installation and dismantling of broadband seismic stations, including wind turbines and solar panels.	<i>Aug 2020</i>
Teaching Assistant Institute of Mathematics, University of Zurich (Switzerland) Analysis I&II, Number Theory, Logic and Set Theory	<i>Sep 2013-Sep 2015</i>

Publications

Articles

1. **Mizrahi, L.**, Nandan, S., Savran, W., Wiemer, S. and Ben-Zion, Y., 2022. Question-Driven Ensembles of Flexible ETAS Models. *arXiv preprint*.
arxiv.org/pdf/2207.06247.pdf
2. **Mizrahi, L.**, Nandan, S. and Wiemer, S., 2021. Embracing Data Incompleteness for Better Earthquake Forecasting. *Journal of Geophysical Research: Solid Earth*.
doi.org/10.1029/2021JB022379
3. **Mizrahi, L.**, Nandan, S. and Wiemer, S., 2021. The Effect of Declustering on the Size Distribution of Mainshocks. *Seismological Research Letters*.
doi.org/10.1785/0220200231

Selected Conference Abstracts

1. **Mizrahi, L.**, Nandan, S. Savran, W., Wiemer, S. and Ben-Zion, Y., 2022. Relaxing ETAS's Assumptions to Better Capture the Real Behavior of Seismicity. *Talk*, EGU General Assembly, May 23-27 2022, Vienna, Austria
2. **Mizrahi, L.**, Nandan, S. and Wiemer, S., 2021. The Role of HPC in the Search of Next-Generation Earthquake Forecasting Models. *Talk*, Platform for Advanced Scientific Computing (PASC) Conference, July 5-9 2021, Geneva (Switzerland)
3. **Mizrahi, L.**, Nandan, S. and Wiemer, S., 2021. Embracing Data Incompleteness for Better Earthquake Forecasting. *Talk*, SSA Annual Meeting, April 19-23 2021, online
Receiver of 2021 Student Presentation Award
4. **Mizrahi, L.**, Nandan, S. and Wiemer, S., 2020. The Effect of Declustering on the Size Distribution of Mainshocks. *Talk*, Swiss Geoscience Meeting, November 6-7 2020, online

Code Repositories

1. **Mizrahi, L.** and Schmid, N., 2022. Imizrahi/etas (3.1). *Zenodo*.
doi.org/10.5281/zenodo.6583992

Technical Skills

Sorted from most to least recently used

Python (pandas, numpy, matplotlib, scikit-learn, keras, PySpark, etc.), **Git** (GitLab, GitHub: [Imizrahi](#)), **distributed high performance computing** (using LSF workload management platform), **UNIX shell**, **LaTeX** (Overleaf), **Microsoft Excel**, **Matlab**, **JavaScript** (AngularJS), **SQL**, **MongoDB**

Languages

German (native), **English** (fluent), **French** (advanced), **Spanish** (basic), **Hebrew** (beginner)