

Statistical Modelling of Induced Earthquakes

Zak Varty, M.Sci, M.Res



Submitted for the degree of Doctor of
Philosophy at Lancaster University.

July 2021

Abstract

Earthquakes induced by human activities present a unique set of challenges to the statistical modeller. Relative to tectonic earthquakes, the recorded number of induced earthquakes can be very small, while interventions to better record and prevent these earthquakes make the use of stationary models either statistically inefficient or inappropriate. On the other hand, the human activity causing seismicity is often well documented and can be a valuable resource that is not available in the tectonic setting.

This thesis focuses on how to model anthropogenic earthquakes while making best use of the limited available data. This research provides three main contributions to statistical seismology, each motivated by the induced earthquakes in the Groningen gas field.

Firstly, we consider the link between earthquake locations and gas extraction, using a state-of-the-art, physically-motivated model as our baseline. We investigate model simplifications to ensure parsimony of the baseline model and explore model extensions that assess the statistical evidence for additional physical characteristics that are not currently represented.

Secondly, we consider how to include developments to the earthquake detection network when modelling earthquake magnitudes. We develop a method for selecting a

time-varying threshold above which the earthquake catalogue may be considered complete. This allows small magnitude events, unused by existing analyses, to contribute to our understanding of the largest events.

Finally, we turn our focus to aftershock activity and the Epidemic Type Aftershock Sequence (ETAS) model. The use of this model is widespread, but the conventional formulation represents a narrow model class with strong parameter dependence and assumes independent and identically distributed magnitudes. We introduce a reparameterisation and two extensions of the conventional ETAS model, along with efficient inference procedures, which alleviate these issues.

Acknowledgements

I would firstly like to thank my supervisors Jon Tawn, Pete Atkinson and Stijn Bierman for their support, guidance and encouragement over the course of my PhD. I have learned a great deal from each of you and I am very grateful for all the opportunities that have come my way as a result of working with you.

I would also like to thank the wider group of people that I have had the pleasure to work alongside both within STOR-i and at Shell. Research is much more enjoyable when surrounded by such exceptional people.

I give special thanks to the friends who saw me through an isolating final year with coffee chats, pomodoros, book clubs and film nights. Thank you to Edwin, Rob, Georgia, Euan, Hankui, Sean, Srshti and Anja. Thank you also to Jamie-Leigh, Aaron, Emma, Tom, Christian, Jon and Julia for the Friday night social calls.

A huge thank you must go to my family and in particular to my parents, Tracey and Alan; without your dedication, love and support I would not be where I am today. This PhD would have been a lot easier, and shorter, if the answer was always six.

Finally, and especially, I would like to give thanks to my partner Paul. Thank you for supporting me when I thought this was impossible and for always being my favourite human.

Declaration

I declare that the work in this thesis has been done by myself and has not been submitted elsewhere for the award of any other degree.

A version of Chapter 5 has been submitted for publication as: Varty, Z., Tawn J.A., Atkinson P.M. and Bierman S. (2021). Inference for extreme earthquake magnitudes accounting for a time-varying measurement process.

- Zak Dennis Varty

"Only fools, liars and charlatans predict earthquakes."

- Charles Francis Richter.

Contents

Abstract	I
Acknowledgements	III
Declaration	IV
Contents	XII
List of figures	XVIII
List of tables	XIX
1 Introduction	1
1.1 Motivation	1
1.1.1 A brief history of the Groningen gas field	4
1.1.2 Geology of the Groningen gas field	6
1.1.3 From gas extraction to induced earthquakes	8
1.1.4 Shell seismic risk analysis	10
1.2 Thesis Outline	12
2 Data	14
2.1 Outline	14

2.2	Earthquake catalogue	15
2.3	Static covariates	16
2.3.1	Fault structure	16
2.3.2	Reservoir thickness and topographic gradient	17
2.4	Extraction covariates	19
2.5	Exploratory analysis	21
2.5.1	Outline	21
2.5.2	Exploratory spatial analysis	21
2.5.3	Exploratory temporal analysis	25
2.5.4	Exploratory magnitude analysis	28
2.5.5	Review	34
3	Literature review	36
3.1	Point process models	36
3.1.1	Overview	36
3.1.2	Poisson point processes	37
3.1.3	Generalisations of the Poisson process	42
3.1.4	Measures of clustering	45
3.2	Extreme value methods	46
3.2.1	Overview	46
3.2.2	Block maxima approach	47
3.2.3	Peaks over threshold	49
3.2.4	Point process representation	52
3.2.5	Inference for extreme value models	53
3.3	Earthquake modelling	55
3.3.1	Overview	55
3.3.2	Epidemic Type Aftershock Sequence models	55

3.3.3	Physics-based modelling	58
3.4	Recent work on Groningen seismicity	62
4	Covariate-based models for induced earthquake locations	64
4.1	Introduction	64
4.1.1	Induced earthquakes	64
4.1.2	Motivation and aims	65
4.1.3	Outline	67
4.2	Background	68
4.2.1	Point process models for earthquakes	68
4.2.2	Baseline intensity model	70
4.3	Alternative models	72
4.3.1	Approach outline	72
4.3.2	Model simplifications	73
4.3.3	Model extensions	74
4.4	Results	80
4.4.1	Outline	80
4.4.2	Covariate smoothing	80
4.4.3	Model reductions	82
4.4.4	Model extensions	83
4.5	Conclusions and further work	85
5	Inference for extreme earthquake magnitudes accounting for a time-varying censoring process	88
5.1	Introduction	88
5.1.1	Aims and motivation	88
5.1.2	Earthquake data	89
5.1.3	Magnitude of completion	90

5.1.4	Extreme value methods	91
5.1.5	Shortcomings of current methods	92
5.1.6	Contributions and outline	93
5.2	Motivating data and model formulation	94
5.2.1	Data description	94
5.2.2	Data model and inference	95
5.3	Motivating the inclusion of small magnitudes	97
5.3.1	Simulation study overview	97
5.3.2	Simulation study results	99
5.4	Threshold selection	101
5.4.1	Overview	101
5.4.2	Graphical assessment	102
5.4.3	Metric-based assessment	103
5.4.4	Minimisation procedure	107
5.5	Threshold selection on simulated catalogues	108
5.5.1	Simulation study overview	108
5.5.2	Constant threshold, hard censoring	109
5.5.3	Constant threshold, phased censoring	111
5.5.4	Non-constant threshold selection	112
5.6	Application to Groningen earthquakes	114
5.6.1	Validating data model for Groningen catalogue	114
5.6.2	Parametric threshold forms	116
5.6.3	Threshold selection	117
5.7	Discussion / Conclusion	122
6	Improving and extending the ETAS formulation	125
6.1	Introduction	125

6.1.1	The ETAS model	125
6.1.2	Magnitude modelling	127
6.1.3	Bayesian ETAS modelling	130
6.1.4	Contributions and chapter outline	132
6.2	Estimation of ETAS parameters	133
6.2.1	Direct estimation: ETAS as a point process	133
6.2.2	Latent estimation: ETAS as a branching process	135
6.2.3	Benefits of the latent estimation approach	137
6.3	Extreme value reparameterisation of empirical laws	139
6.3.1	Empirical laws	139
6.3.2	Proposed parameterisation	140
6.3.3	Comparing parameterisations	143
6.3.4	Demonstration on simulated catalogue	146
6.4	Extensions of the magnitude model	157
6.4.1	Dual magnitude extension	157
6.4.2	Correlated magnitude extension	159
6.5	Application of extended ETAS models to simulated catalogues	164
6.5.1	Dual magnitude extension	164
6.5.2	Correlated magnitude extension	175
6.6	Conclusions and further work	183
7	Conclusions and further work	186
A	Supplementary materials to Chapter 4	194
A.1	Integrated intensity functions	194
A.2	Maps of annual expected earthquake counts	196
B	Supplementary materials to Chapter 5	201

B.1	Assessing i.i.d. assumption for Groningen earthquakes	201
B.1.1	Connection to main text	201
B.1.2	Exploratory analysis	201
B.2	Bootstrap datasets and parameter estimates	203
B.2.1	Relation to main text	203
B.2.2	Generating bootstrapped data-sets	203
B.2.3	Generating bootstrap maximum likelihood estimates	208
B.3	Sampling standardised threshold exceedances	208
B.3.1	Connection to main text	208
B.3.2	Sampling unrounded threshold exceedances	209
B.3.3	Transformation onto common margins	210
B.4	Supplementary figures	212
C	Supplementary materials to Chapter 6	215
C.1	B conditional posterior, dual magnitude model	215
C.2	ψ conditional posterior, correlated magnitudes	218
C.3	B conditional posterior, correlated magnitudes	220
D	Supplementary materials to Chapter 7	224
D.1	Outline for combined covariate and aftershock model	224
D.1.1	Link to main text	224
D.1.2	Motivation	225
D.1.3	Suggested model	226
D.1.4	Likelihood function	230
D.1.5	Condition for sub-criticality	232
D.1.6	Parametric forms for the temporal change in stress	233
D.1.7	Parametric forms for the spatial change in stress	235
D.1.8	Parametric forms for the effect of magnitude on additional stress	238

D.1.9 Summary	239
Bibliography	242

List of Figures

1.1.1 Schematic cross-section through the Groningen field indicating the main stratigraphic intervals. Source: TNO (2017)	7
1.1.2 Lithostratigraphic subdivision of the Rotliegend in the Groningen area. Source: NAM (2016a)	7
1.1.3 Schematic diagram of seismic risk assessment procedure after modelling of pressure depletion and reservoir compaction.	11
2.3.1 Maps of human settlements, gas production clusters, fault lines and earthquakes in the Groningen region.	18
2.3.2 Topographic gradient and initial thickness of Groningen reservoir. . .	18
2.5.1 Contour plots of the spatial distribution of earthquakes in the Groningen gas field	22
2.5.2 Estimated Ripley K-function for Groningen earthquakes	23
2.5.3 Boundary of high density regions HDR1 and HDR2 and temporal event density in each region.	25
2.5.4 Histogram displaying the number of earthquakes each year since 1995.	26
2.5.5 Estimated earthquake intensity through time.	27
2.5.6 Quantile-quantile plot of inter-event times.	27
2.5.7 Kernel density estimated magnitude distributions using all events and only those exceeding $1.5M_L$	29

2.5.8 Magnitudes of Groningen earthquakes through time.	31
2.5.9 Estimated magnitude distributions for events exceeding $1.5M_L$	33
2.5.10 Quantile-quantile plot of magnitude exceedances of $1.5M_L$	34
3.2.1 Data used in block maxima and peaks over threshold analysis	50
4.3.1 Groningen earthquakes and line used to separate lower and upper modes of earthquake activity.	76
4.3.2 Maps of expected event counts under the fitted baseline model for years 2005-2007.	77
5.2.1 Event times and magnitudes of Groningen earthquakes.	95
5.3.1 Catalogue structure and conditional return level plots for conservative, extended and stepped approaches to inference.	100
5.4.1 Examples of modified PP and QQ plots for valid and invalid flat thresholds.	104
5.5.1 PP- and QQ- distance metrics to select a constant threshold for a single simulated catalogue.	110
5.5.2 Sampling distribution of threshold selection methods for quantile-based metrics and catalogues simulated with a constant threshold and hard censoring.	111
5.5.3 Example simulated catalogues with hard censoring and phased censoring applied to a stepped threshold.	113
5.5.4 Marginal sampling distributions of errors in the selected values of $v^{(1)}$, $v^{(2)}$ and τ^* for simulated catalogues with change-point type thresholds and either hard or phased censoring.	113
5.6.1 Modified PP (left) and QQ (right) plots for Groningen magnitudes exceeding $1.45M_L$ under the GPD model.	116

5.6.2 Threshold selection and conditional return level plots for Groningen earthquakes and a constant threshold.	118
5.6.3 Selected sigmoid thresholds fro Groningen earthquakes using Bayesian optimisation from 5 random initial parameter sets.	119
5.6.4 Bootstrap GPD parameter estimates and conditional return levels based on Groningen earthquake magnitudes exceeding the selected conservative, flat and sigmoid thresholds	122
6.2.1 Graphical representation of a toy example from the ETAS model. . .	136
6.3.1 Simulated earthquake catalogue of 863 events used to compare θ_{conv} and θ_{prop} parameterisations of the ETAS model.	147
6.3.2 Contour plots of pair-wise marginal posterior distributions of ETAS parameters θ using the θ_{conv} and θ_{prop} parameterisations.	148
6.3.3 Estimated posterior distributions using samples from the θ_{prop} and θ_{conv} parameterisations, shown on each parameter space.	149
6.3.4 Posterior distributions of physically meaningful quantities θ_{phys} using estimated using samples from the θ_{prop} and θ_{conv} parameterisations. .	151
6.3.5 Trace plots and posterior pmfs for branching vector elements B_{51} , B_{81} and B_{199}	154
6.3.6 Proportion of sampled B_i values equal to the true value in the θ_{conv} and θ_{prop} chains.	155
6.3.7 Proportion of updates at which each branching vector element was unchanged for the θ_{conv} chain and θ_{prop} chain.	156
6.5.1 Marginal posterior distributions of ETAS parameters θ_{prop} given a dual magnitude earthquake catalogue, estimated using a dual magnitude model or single magnitude model.	165

6.5.2 Marginal and joint posterior distributions of magnitude parameters based on a simulated dual magnitude earthquake catalogue.	167
6.5.3 Return level plots for background magnitudes, triggered magnitudes, and all magnitudes.	167
6.5.4 Posterior proportions of triggered events and correctly allocated elements of B , based the dual magnitude and single magnitude models.	169
6.5.5 Marginal posterior distributions of ETAS parameters θ_{prop} given an earthquake catalogue with i.i.d. magnitudes, estimated using a dual magnitude model or single magnitude model.	171
6.5.6 Marginal and joint posterior distributions of magnitude parameters shown by event type under a dual or single magnitude model fitted to a catalogue with i.i.d. magnitudes.	172
6.5.7 Joint posterior distributions of the differences in magnitude parameters.	173
6.5.8 Return level plots for background magnitudes, triggered magnitudes, and all magnitudes fitted using a catalogue with i.i.d. magnitudes. . .	173
6.5.9 Posterior proportions of triggered events and correctly allocated elements of B , based the dual magnitude and single magnitude models and a simulated catalogue with i.i.d. magnitudes.	174
6.5.10 Marginal posterior distributions of ETAS parameters θ_{prop} given an earthquake catalogue with dual magnitudes, estimated using a dual magnitude model and correlated magnitude model.	177
6.5.11 Marginal posterior distributions of magnitude parameters ψ given a earthquake catalogue with dual magnitudes, estimated using a dual magnitude model and correlated magnitude model.	177

6.5.12 Posterior proportions of triggered events and correctly allocated elements of B , based the dual magnitude and correlated magnitude models for a simulated catalogue with dual magnitudes.	178
6.5.13 Marginal posterior distributions of ETAS parameters θ_{prop} given an earthquake catalogue with correlated magnitudes, estimated using a dual magnitude model and correlated magnitude model.	180
6.5.14 Marginal posterior distributions of magnitude parameters ψ given an earthquake catalogue with correlated magnitudes, estimated using a dual magnitude model and correlated magnitude model.	180
6.5.15 Posterior proportions of triggered events and correctly allocated elements of B , based the dual magnitude and correlated magnitude models for a simulated catalogue with correlated magnitudes.	181
A.2.1 Observed and expected event counts in each year (1995-2006) under the fitted baseline model B0.	197
A.2.2 Observed and expected event counts in each year (2007-2016) under the fitted baseline model B0.	198
A.2.3 Observed and expected event counts in each year (1995-2006) under the fitted model extension E2.	199
A.2.4 Observed and expected event counts in each year (2007-2016) under the fitted model extension E2.	200
B.1.1 Frequency plots of the number of earthquakes exceeding $1.45M_L$ that separate earthquakes exceeding $1.65M_L$, $1.75M_L$, and $1.85M_L$ in the Groningen earthquake catalogue.	202
B.2.1 Threshold exceedances as point process.	204
B.4.1 Sampling distribution of threshold selection methods for PP-based metrics for simulated catalogues with constant threshold and hard censoring.	213

B.4.2 Sampling distribution of Maximum likelihood estimates of GPD parameters obtained using a conservative, stepped and extended approach, along with mean squared error decomposition under each approach.	213
B.4.3 Bootstrap maximum likelihood estimates for GPD parameters and conditional return levels for Groningen magnitudes above 1.45ML assuming GPD and exponential models.	214
B.4.4 Sampling distribution of threshold selection methods for QQ-based and PP-based metrics for catalogues simulated with a constant threshold and phased censoring.	214
D.1.1 Schematic of ETAS model inputs, component models, and outputs.	227
D.1.2 Schematic of combined model inputs, component models, and outputs.	227
D.1.3 Proposed forms for additional stress and stress rate as functions of time.	234
D.1.4 Proposed forms for additional stress as a function of distance x	237
D.1.5 Proposed forms for the relationship between magnitude and total additional stress.	240

List of Tables

3.3.1 Reservoir models considered by Bourne and Oates (2017a)	60
4.3.1 Intensity functions for sub-models, the baseline model and model extensions.	75
4.4.1 Summaries of fitted sub-models, baseline model and model extensions.	81
6.3.1 Element-wise relative root mean squared errors ($\times 10^3$) of θ_{conv} , θ_{prop} and θ_{phys} for MCMC chains on the θ_{prop} and θ_{prop} parameter spaces.	150
6.3.2 Element-wise effective sample sizes of θ_{conv} , θ_{prop} and θ_{phys} for MCMC chains of 10,000 sampled values on the θ_{prop} and θ_{prop} parameter spaces.	152
A.1.1 Integrated intensity functions for sub-models, the baseline model and model extensions.	195

Chapter 1

Introduction

1.1 Motivation

Statistical seismology is the study of earthquakes as a stochastic phenomenon. This approach to earthquake modelling is both useful and necessary because the physical processes that cause earthquakes are highly complex. These processes cannot be observed or modelled at the necessary scale and precision for seismicity to be effectively modelled deterministically. Taking a more descriptive, statistical approach permits modelling using the noisy, incomplete knowledge that is available.

Much of the literature in statistical seismicity was developed for earthquakes caused by the motion of the Earth's tectonic plates. The catalogues of earthquakes used to construct these models consist of many, large magnitude earthquakes that are observed on an expansive spatio-temporal region. We are interested in modelling induced earthquakes, which are caused by human activity. While these are related to tectonic earthquakes, they present their own unique set of modelling issues and opportunities.

Catalogues of induced earthquakes are typically much smaller than those of tectonic earthquakes. They are smaller in terms of the number of earthquakes, the magnitude of those earthquakes and the region on which they are observed. Smaller data sets make estimating model parameters and comparison of models more difficult, because there is less information on which to base these decisions. Smaller magnitude events are also more difficult to discern from background vibrations, making them harder to measure and locate. A smaller observation window increases the importance of boundary effects and measurement errors. When modelling induced earthquakes we are not in a data-rich setting, and so making the most of the available data and expert knowledge is of paramount importance.

Models for tectonic earthquakes often assume that the system being modelled has settled to a steady state. Changes to this system caused by human behaviour are rapid on a geological time-scale. Steady state models for induced earthquakes are therefore rarely appropriate and models for non-stationary behaviour are required. This presents a modelling challenge in a low-data setting, but presents an opportunity from the application perspective; with the correct intervention there is potential to prevent earthquakes by changing the human activity that is causing them, rather than only defending against them.

This ability to alter the earthquake generating process can lead to greater non-stationarity in the data. Following one or more large earthquakes the generating process is likely to be changed in an attempt to prevent further such events, based on the available data. These changes might be single or multiple, synchronous or asynchronous. Their effect on seismicity might be smooth or sudden, instantaneous or lagged, or only have a cumulative effect. Making good, data-driven decisions on how to intervene requires high quality detection systems. Improvements to these systems over time are therefore likely to increase the ability to detect small earthquakes. This

will result in further non-stationarity in the catalogue through an improvement in both the quality and quantity of earthquake observations within the record of seismic activity. Separating the developments in the earthquake generating and detection processes is a challenge that is not faced when modelling tectonic earthquakes.

One substantial advantage when modelling induced earthquakes is that the human activity which causes the earthquakes is often reliably recorded. When combined with geological expertise, these records can be used to measure or estimate covariates relevant to the earthquake generating process. These covariates are rarely available in the tectonic setting and can help to complement the small amount of data in the earthquake catalogue itself.

There are two main purposes for modelling induced seismicity. The first is to develop understanding of the process that is generating earthquakes. The second is to draw on this understanding to forecast earthquakes under different scenarios. Each of these, in their own way, allows informed decision making on which actions or interventions should be taken in order to keep seismic hazard at an acceptable level.

Modelling induced earthquakes presents many challenges. Some of these are unique to a particular application and require solutions tailored to that application. Other challenges are common to all earthquake catalogues and could be adapted to similar data structures in other applications. This thesis presents solutions to several problems motivated by earthquakes caused by gas extraction in the Netherlands. The remainder of this chapter gives the context for these problems and then describes the thesis structure.

1.1.1 A brief history of the Groningen gas field

At its time of discovery in 1959, the Groningen gas field was the largest in the world. Today it remains the largest natural gas field discovered in Europe and among the top ten worldwide (TNO, 2017). It was estimated that the field initially contained close to 3000 billion cubic metres (bcm) of gas. Extraction of this gas commenced in 1963 and by the beginning of 2015, 2115bcm of gas had been extracted from the field - around 75% of the initial volume (van Thienen-Visser and Breunese, 2015). This extraction is operated by the Dutch Petroleum Society (Nederlandse Aardolie Maatschappij, NAM), a joint venture between Royal Dutch Shell and Exxon Mobil.

The Groningen field is located in the north-east of the Netherlands. It is situated in the geological region named the Groningen high. This region is tectonically inactive and does not experience earthquakes caused by the motion of the Earth's tectonic plates. Following extraction of gas from the field, concerns were raised about the possibility of induced earthquakes; earthquakes caused by human activity. Since 1986 the region has been monitored for the presence of low magnitude seismic events that might occur as a result of the gas extraction. The network of geophones used for this monitoring is owned and maintained by the Royal Netherlands Meteorological Institute (KNMI). This network extends across the Netherlands but repeated investment in, and improvement of, the network were made in the Groningen region in order to better detect and understand induced earthquakes around the Groningen gas field. These improvements mean that this region of the Netherlands now has the highest resolution geophone network globally.

The first recorded earthquake in the the Groningen region occurred in 1991. To date, the largest recorded event occurred on the 16th of August 2012, with a local magnitude of 3.6M_L. While this is not a large magnitude in relation to tectonic seismic events, these induced events occur at shallow depths which, combined with the soft surface

soil in the area, lead to housing damage (van Thienen-Visser and Breunese, 2015). The gas from the Groningen field was used extensively both within the Netherlands and as an export to Germany, Belgium and Northern France. Within these regions adaptations were made to the majority of domestic and industrial gas appliances to account for the relatively low calorific value of gas from the Groningen field (NAM, 2017; TNO, 2017). This meant that switching to gas from another source was a prohibitively expensive option and extraction from the site was expected to continue until around 2080 subject to safety and the recommendations of the state supervision of mines (TNO, 2017).

Further earthquakes in the years that followed lead to cap on total annual extraction from the Groningen field. In addition, changes were made to the method of gas extraction in an attempt to extract more evenly across the field and through the year in an attempt to mitigate the negative effects of gas extraction. Magnitude 3.4 events occurring in January 2018 and May 2019 contributed to the decision to move away from Groningen gas and transition to using more costly imported gas. The initial deadline for this transition was then accelerated so that from 2022 gas will only be extracted from the field to supply extreme demand caused by the most severe seasonal weather.

The mechanism linking gas extraction to induced seismicity is highly complex and not well understood. The central aim of this thesis is to contribute to the understanding of this relationship and the models that are used to describe it. The original purpose of this was to allow informed decision making when selecting between a set of future gas extraction scenarios. This understanding is useful even when production is terminated, because in the years that follow gas will continue to redistribute within the reservoir and potentially cause further seismic events.

1.1.2 Geology of the Groningen gas field

The Groningen gas field is located on the Groningen High, a region that is fault-closed and lies within a tectonically stable block. The gas reservoir is located at a vertical depth that varies from 2.6-3.2km below surface level, in the Rotliegend layer as shown in Figure 1.1.1. The fault closures around the field prevent movement of gas out of the region, while a perfect top seal for the reservoir is provided by a layer of Zechstein salt above (NAM, 2016a). The Rotliegend reservoir is composed of porous sandstone and claystone. The pore space in the structure of this rock layer is filled with gas emanated from the carboniferous layer below. The thickness of the porous Rotliegend layer varies across the field, from its thinnest of 140m in the south-south-east to thickest of 300m in the north-north-west of the field. As well as varying in thickness, the composition of the layer also changes from predominantly sandstone to predominantly claystone along this gradient, as shown in Figure 1.1.2. The porosity of the sandstone rock is high while in the claystone it is low. The composition and thickness of the reservoir partly determine the potential for compaction of the porous layer on the removal of gas from its pore space. This is pertinent because pore pressure depletion and the resulting compaction of the reservoir are thought to be drivers of the seismicity observed in the Groningen region (NAM, 2016a).

Other reservoirs have been observed with greater levels of compaction than seen in Groningen without any earthquakes being induced (Davison et al., 2010). Earthquakes occur because the Groningen reservoir contains a great many pre-existing faults. These faults are cracks in the reservoir rock structure that pre-date gas extraction and are thought to be necessary for induced seismicity to occur. In excess of 1700 faults have been interpreted in the Groningen field as a part of the addendum to the 2013 Wulingsplan Groningen (van Elk et al., 2013). Of the faults in the Groningen reservoir, 707 are included within the static and dynamic reservoir models

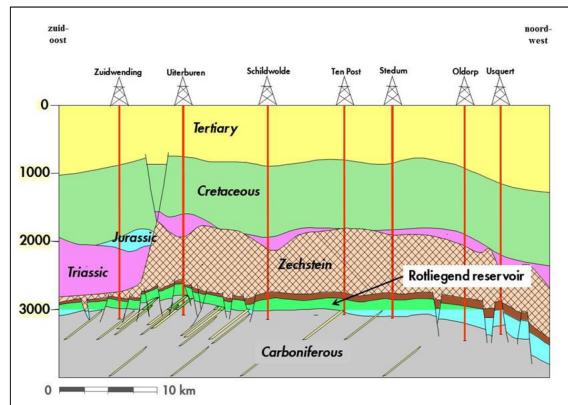


Figure 1.1.1: Schematic cross-section through the Groningen field indicating the main stratigraphic intervals. Source: TNO (2017)

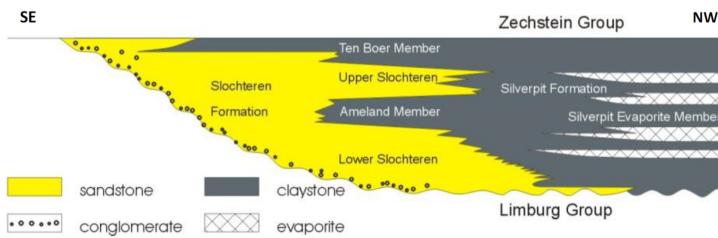


Figure 1.1.2: Lithostratigraphic subdivision of the Rotliegend in the Groningen area.
Source: NAM (2016a)

of Shell. The selection of these faults was justified by the discarded faults adding geometric detail but not contributing significantly to seismic risk at the scale being considered. Even with this up-scaling, it is currently not computationally feasible to include all 707 of these faults in geomechanical models of the reservoir. Simplifications of this fault structure are used in the geomechanical models, with ongoing work to lessen these simplifications. It should be noted that at the resolution of the available data it is not possible to assign earthquakes to individual faults. These limitations to a deterministic, physical model for induced earthquake activity further motivates a statistical approach to this problem.

1.1.3 From gas extraction to induced earthquakes

Gas is extracted from the Groningen reservoir via 258 wells, which are located across the field as 22 production clusters. The gas at each of these wells comes from a common connected source. As gas is removed at wells, the remaining gas will move from areas of high to low concentration and equalise the pressure gradient caused by extraction. However, due to hydraulic resistance across the reservoir this pressure equalisation is far from instantaneous (van Thienen-Visser and Breunese, 2015).

The pore pressure in the reservoir is the pressure of the gas that is located in the pore space of its rocky structure. Localised and field-wide reduction of pore pressure are problematic. This is because the layers of rock and soil above the reservoir are held in place by a combination of the normal force provided by the reservoir structure and the pore pressure of the gas within that structure. As gas is removed from the reservoir the pore pressure reduces because the number of molecules per unit volume is decreased. The resultant force causes compaction of the reservoir structure until the forces are again balanced. As an illustration of this concept, consider letting air out of a balloon. The pressure inside the balloon reduces while the air pressure on the

outside of the balloon remains constant. This net force causes the balloon to contract until the forces are once again in balance.

This analogy also introduces the idea of elastic deformation; the balloon is able to stretch and compress, subject and proportional to the forces exerted upon it. Such elastic deformations are often assumed when constructing physics-based models for induced seismicity. An associated concept is the bulk modulus of a material. This gives the relative change in volume of the material per unit of compressive force exerted upon it. In the Groningen reservoir, it has been argued that pressure depletion brings about compaction in two ways; not only does it cause the net force on the reservoir but it is also thought to reduce the bulk modulus of the composite material. In other words, the reservoir compacting is thought to make it more susceptible to further compaction (Bourne and Oates, 2017a).

This compaction alone is not necessarily problematic. Much of the 900 km² area of the Groningen field has historically experienced subsidence due to the compaction of the overburden, comprised of the layers above the reservoir. This means that building methods and interventions are already used and available in the area to cope with compaction and the resulting subsidence. It is the pre-existing faults which present a danger.

When compaction occurs in the reservoir, as opposed to the overburden, additional shear stresses are placed across the pre-existing faults within this layer. Initially these are accommodated by static friction and the walls of the fault remain in place. When these stresses become large enough to overcome the static friction, the fault will slip, releasing energy and causing an induced earthquake (van Thienen-Visser and Breunese, 2015). While this general process is widely agreed upon, the specifics of how this happens are not well understood. Additionally, there is not agreement upon which reservoir properties and features of the extraction scheme that are most important

drivers of induced earthquake activity. For example, Bourne et al. (2014) and Bourne and Oates (2017a) argue in favour of seismicity being dependent on compaction while van Thienen-Visser et al. (2016) provides a counter-argument, favouring seismicity being dependent on pressure depletion. Despite this lack of understanding and agreement, predictions of subsidence and seismic hazard must be produced. These must be provided to regulators for a proposed production plan and must be updated and plans be amended as extraction proceeds (van Thienen-Visser and Breunese, 2015).

1.1.4 Shell seismic risk analysis

Extraction of gas from the Groningen field has in the past been an essential component of the Dutch economy and continues to be an important reserve for times of crisis. It is widely agreed that this process has led to induced earthquakes in the Groningen region, but the specifics of this process are not well understood.

While operating and eventually closing the Groningen gas field, Shell and NAM are invested in minimising the risk associated with the production of natural gas. The production plans of NAM include predictions of the subsidence and seismicity caused by the proposed gas extraction and the associated the hazards and risks. The modelling and assessment of risks are covered in greater technical detail in the Technical Addendum to the Winningsplan. This large document is split into sections covering methods for: planning gas production, forecasting and reducing seismic risk NAM (2016a); subsidence modelling NAM (2016b); hazard assessment NAM (2016c); risk assessment NAM (2016d) and damage assessment NAM (2016e).

A schematic diagram of the risk assessment procedure is shown in Figure 1.1.3. The process begins by forecasting pressure depletion and compaction for three levels of production, at each level considering two production distributions across the reservoir. This gives six possible production scenarios for risk comparison. These models

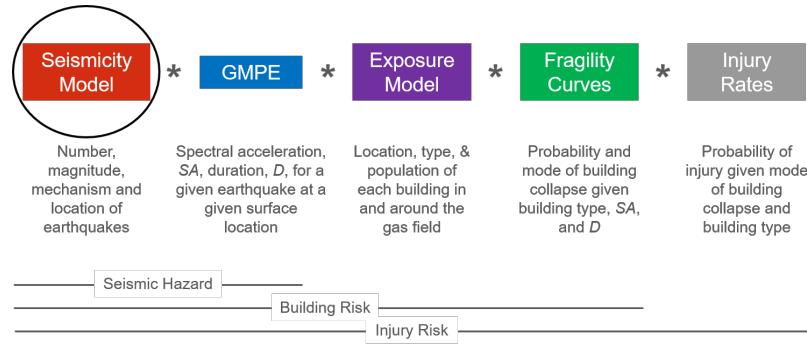


Figure 1.1.3: Schematic diagram of seismic risk assessment procedure after modelling of pressure depletion and reservoir compaction.

for pressure depletion and compaction of the reservoir then feed into the probabilistic forecasting, under each production scenario, of seismic events within the reservoir. These forecasts are then translated into the effects seen at surface level through the use of ground motion prediction equations (GMPEs). These ground motions, along with the density of population and development across the region, are used to calculate the risk and hazard profiles of each production scenario. The assessment of production scenarios can therefore be seen as a modular exercise, with each component interchangeable with another method that performs the same function and provides as output the required inputs of future modelling stages.

This framework is adaptable to changes in internal structure but also can affect external change. As a result of the first Winningsplan published in 2013, action was taken in 2014 to reduce the risk associated with induced seismic activity. Production was then reduced and redistributed within the field, with dramatic reductions in the areas with greatest compaction (van Thienen-Visser and Breunese, 2015). This action followed from the theory, later published in (Bourne et al., 2014), that the amount of seismic energy that can be released in an earthquake is an exponential function of the cumulative compaction of the reservoir.

This thesis aims to provide improvements for the first component of this workflow: modelling the number, magnitude, mechanism and location of induced earthquakes.

1.2 Thesis Outline

This thesis has two primary aims. The first of these is to improve the understanding of the process by which earthquakes are induced in the Groningen gas field through examination, comparison and deconstruction of existing models. The second aim is to improve these models in terms of their ability to represent reality and in terms of their statistical properties.

Chapter 2 introduces and describes in detail the data that are available for modelling earthquakes in the Groningen gas field. Features of both the data and the collection process are discussed.

Chapter 3 introduces the relevant theory on stochastic point processes and extreme value theory that provide the building blocks of the statistical models used in this thesis. In particular Poisson, self-exciting and marked point processes are described along with the peaks-over-threshold and point process approaches to univariate extreme value modelling. Also introduced are current approaches to modelling induced and tectonic earthquakes, with a focus on statistical approaches.

In Chapter 4 earthquake locations are considered in isolation from their magnitudes. A physical hybrid model of earthquake locations is deconstructed to identify important features that it captures in the earthquake generating process, and features it fails to capture that present potential areas for improvement.

In Chapter 5 the development of the earthquake detection network in Groningen is considered. A statistical approach is developed for estimating the time-varying

magnitude threshold above which all earthquakes are detected. The benefits of using these additional small magnitude events when modelling large magnitude event is demonstrated using both simulated and Groningen data.

In Chapter 6 takes the popular epidemic type aftershock sequence (ETAS) model for aftershocks and proposes a reparameterisation and extension of the model. The reparameterisation simplifies inference by rendering the parameters near-orthogonal while increasing flexibility in the model. The extension provides a simple way to move the ETAS model beyond the assumption of independent, identically distributed magnitudes and allows interaction between earthquakes.

Chapter 7 concludes this thesis with a summary of the contributions that have been made, a discussion of the shortcomings of the presented models, and opportunities for future work.

Chapter 2

Data

2.1 Outline

This chapter will introduce the data available for modelling earthquake activity in the Groningen region. The earthquake catalogue used in this thesis is publicly available from the Royal Dutch Meteorological Institute, (KNMI, 2020). The data on gas extraction and subsidence were provided by Shell and are available as part of published technical reports, e.g. Bourne and Oates (2017b). Supplementary information concerning the development of the earthquake detection and the gas production networks were also provided by Shell. These datasets will be outlined in Sections 2.2 - 2.4, where descriptions are given on both the information recorded and how this was collected or constructed. Following this, Section 2.5 provides an exposition of the key features of these data sets.

2.2 Earthquake catalogue

The earthquake catalogue is freely available from the Royal Dutch Meteorological Institute KNMI (2020). It details the location, time and magnitude of all recorded earthquakes across the European Netherlands since the 5th of December 1991.

Earthquake locations are given by Northing-Easting pairs in the RD coordinate system. Earthquake locations are stated to be accurate to within 500m. This has previously been interpreted as locations having isotropic standard errors of 250m (Bourne and Oates, 2015). The RD coordinate system is a planar projection of the European Netherlands and the areas of land and sea around its borders. The errors in location incurred because of this projection are up to 0.25m, which are inconsequential as compared to the measurement error. In addition to planar locations, earthquake depths are available but these measurements have low resolution compared to the depth of the reservoir. Earthquakes within the Groningen field are therefore all assigned a nominal depth of 3km and so earthquake locations are represented as points on a plane.

Each recorded earthquake has an associated magnitude. This is a measure of the energy released by the seismic event and is reported in units of local magnitude (M_L) to one decimal place. Local magnitude is a logarithmic scale used to measure earthquake severity. It is important to note that an earthquake can only be included in the catalogue if it is of sufficiently large magnitude for its location to be determined. This requires that the earthquake to be detected by three or more geophones. The *magnitude of completion* for a given region and period is the smallest magnitude seismic event that can be detected at any location or time within that interval. The magnitude of completion therefore varies in space and time according to the number, location and quality of geophones within the interval.

The magnitude of completion for the Groningen gas field is a property that must be estimated. It is widely accepted that since 1995 a conservative estimate of the magnitude of completion in the Groningen region is $1.5M_L$. Since then significant investment in the geophone network has lead to an increased ability to detect small magnitude events, reducing the magnitude of completion. The time at which this was reduced and the value that it was reduced to are disputed. Statistical analyses of the Groningen catalogue are therefore typically performed only using events of magnitude $1.5M_L$ or greater to avoid the bias induced by incorrectly assuming that the catalogue is complete.

Figure 2.3.1 shows the locations of town- and city-centres and the locations of all observed earthquakes in the period 1991-2015. The magnitude of each observed earthquake is indicated by the radius of the circle representing it. The highest density regions of earthquakes and population are aligned, which makes induced earthquakes in the region a particular concern.

2.3 Static covariates

2.3.1 Fault structure

The fault structure of the reservoir was determined using tomography (also known as seismic imaging) and can be separated into closing faults and internal faults. Gas cannot move across closing faults, which represent the perimeter of the gas field and determine its spatial extent. A polygonal approximation of the perimeter of the gas field is shown in Figure 2.3.1. Internal faults are cracks in the rock structure of the reservoir and are represented by sets of connected line segments. The 707 internal faults that are included in the static and dynamic reservoir models of the 2013 Winningsplan Groningen van Elk et al. (2013) are shown in Figure 2.3.1. It

can be seen that these major faults are not evenly distributed throughout the field, being more heavily concentrated in the south, with a band running NW-SE across the field.

This set of faults was chosen by expert opinion to be a representative and influential collection of faults. However, the selection is not exhaustive and this is for two reasons. Firstly, the fractal nature of faulting means that there will always be faults smaller than the resolution of the chosen imaging technique. It is thought that the failure of several of these undetectable faults could cause an earthquake that is not attributable to any of the larger, detected faults. Secondly, the computational intensity of the reservoir models limits the scale at which faults can be included while maintaining reasonable computation times. The threshold for fault inclusion is therefore somewhat arbitrary but should ideally be as low as possible within the constraints of the available measurement and modelling techniques.

2.3.2 Reservoir thickness and topographic gradient

The initial thickness of the reservoir, measured in metres, is given on a 500×500 metre grid across the reservoir. The topographic gradient of the reservoir is the gradient of the upper surface of the reservoir, and is provided on the same 500×500 metre grid. Topographic gradient is a dimensionless scalar value giving the absolute value of the steepest gradient at each point in the reservoir. The top surface of the reservoir is measured by tomography, from which the spatial gradients used to calculate the topographic gradient are obtained by finite differencing methods.

The topographic gradient and initial reservoir thickness are static reservoir properties, they do not change over time, and are plotted in Figure 2.3.2.

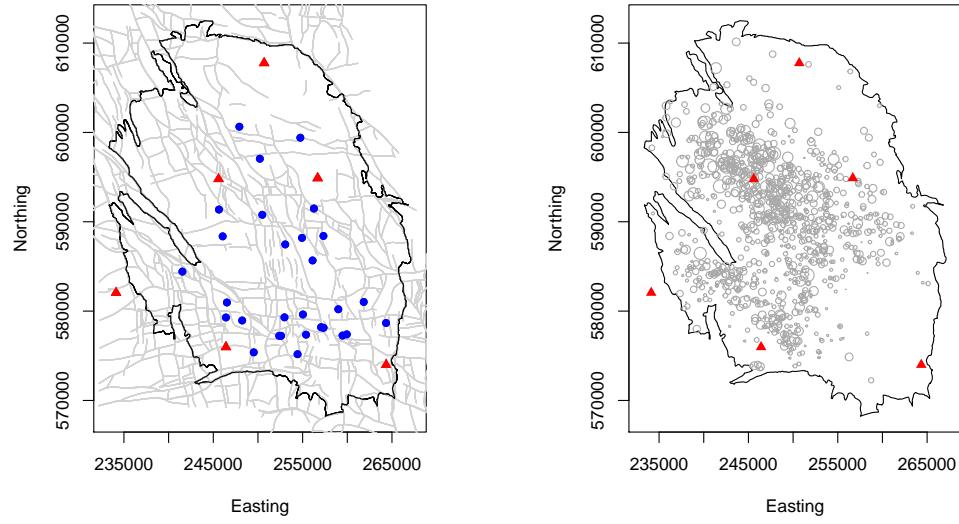


Figure 2.3.1: Locations of towns and cities around the Groningen field, shown by red triangles. [Left] Locations of active production clusters shown as blue circles. Major mapped faults shown in grey. [Right] Locations of all Earthquakes since 1995 shown as grey circles with areas proportional to magnitude.

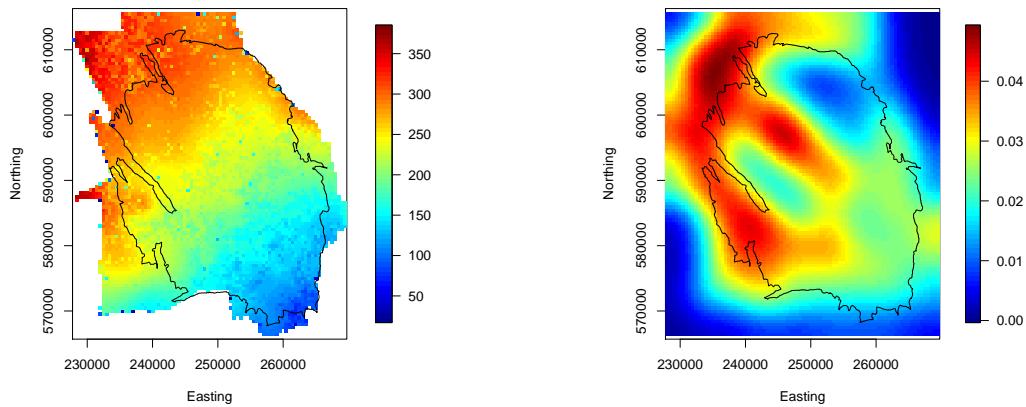


Figure 2.3.2: [Left] Thickness of Rotliegend reservoir in metres. [Right] Topological gradient, the gradient of the upper surface of the Rotliegend reservoir.

2.4 Extraction covariates

Information on how gas extraction impacts the reservoir is also available in the form of gridded covariate data. The spatial grid on which these covariates are available is the same as for the static covariates. Temporally, these values are reported at each location, interpolated based on model output to the start of each year or to the start of each week.

Extraction covariates are outputs of physics-based reservoir models which interpolate and infer values on a regular grid, based on irregularly spaced measurements of: pressures at well-heads, surface-level displacements, and seismic images of the reservoir's top surface. The scale of errors on these covariates is unknown, but is likely to be large, and is frequently disregarded. Bourne and Oates (2017a) acknowledges the spatial uncertainty in both earthquake location and covariate value by selecting the level of spatial smoothing in the covariate to optimise predictive performance. What follows is a brief description of four possible covariates and their methods of estimation.

Pressure depletion (Bar): Pressure in the reservoir pore space is measured at well-heads at infrequent, irregular intervals. The MoRes finite element reservoir model interpolates these values to return pressure estimates at 3 dimensional points within the reservoir (NAM, 2016b). These estimates are based on strong modelling assumptions about reservoir properties and are aggregated vertically and then laterally interpolated to give estimates on a regular lattice. Smoothing is applied to these lattices before use as a covariate.

Compaction (Metres): Compressibility of the reservoir, which is assumed to be temporally constant but spatially varying, is estimated on a 2500m grid using surface

displacement measurements and estimated pressure depletion. Estimates of compaction are achieved on a 500 metre grid by multiplying the pressure depletion grids by the compressibility grid and then smoothing. For further details see Bierman et al. (2015).

Strain Thickness (Metres): The vertical integral of strain over the thickness of the reservoir at a given point. When there is only vertical strain this reduces to compaction, otherwise it is the product of compaction and the topographic gradient. Topographic gradient is the absolute value of the slope of the reservoir's upper surface, as measured using seismic imaging. To avoid over-fitting based on this image a smoothing is usually applied. For further details see Bourne and Oates (2015).

Incremental Coulomb stress (Bar): The additional stress placed along a fault by compaction across the fault. For steeply dipping faults, the confining stress will be small as compared to shear stress. In this case incremental coulomb stress will be equal to a material constant multiplied by vertical strain (compaction per unit original length) multiplied by topographic gradient. Again smoothing is used to avoid overfitting to the seismic image.

The pressure depletion and the resulting derived values of compaction, strain thickness and incremental Coulomb stress were initially only available aggregated to an annual scale. Later a weekly version of pressure depletion was made available, which presented issues that were not present in the aggregated data. These issues stem from the historically variable extraction of gas to supply a strongly seasonal demand. This presents as large amounts of gas being extracted during the winter months and relatively little during the summer. During the period of high extraction pockets of low pressure are created which gas then moves into, equalising the pressure gradient during the summer months. This leads to localised but extensive pore pressure

increases in the weekly data. How these areas of ‘pressure-up’ should be related to earthquake activity is not well understood.

2.5 Exploratory analysis

2.5.1 Outline

The aim of this section is to provide an exploratory analysis of the Groningen earthquake catalogue in order to improve familiarity with the available data and identify potential challenges developing models for this earthquake catalogue. This exploratory analysis has been decomposed to consider spatial, temporal and magnitude features separately.

As mentioned in Section 2.2 the magnitude of completion is not constant through the observation period. This complicates the interpretation of many aspects of the exploratory analysis, because not all events below the magnitude of completion are recorded and therefore a naive use of all data will result in biased conclusions. To demonstrate this, several of the following analyses are presented for all events since 1995, irrespective of their magnitudes, and also for events since 1995 which exceed a conservative threshold magnitude of $1.5M_L$. This choice of threshold follows Bourne and Oates (2017a) and Dost et al. (2012), where the magnitude of completion is taken to be $1.5M_L$ in the Groningen field for this period.

2.5.2 Exploratory spatial analysis

Following from Figure 2.3.1, which showed the location of events of all magnitudes, we begin our exploratory analysis with a simple investigation of the spatial density of events throughout the Groningen region. Figure 2.5.1 shows Gaussian kernel density estimates of the spatial distribution of seismic events for all events since 1995 and those

exceeding magnitude $1.5M_L$. Most earthquakes occur in the north-west of the gas field with some evidence of another intensity mode in the south-west. The earthquakes in the south tend to be smaller and this second mode is less distinct when considering events above magnitude $1.5M_L$.

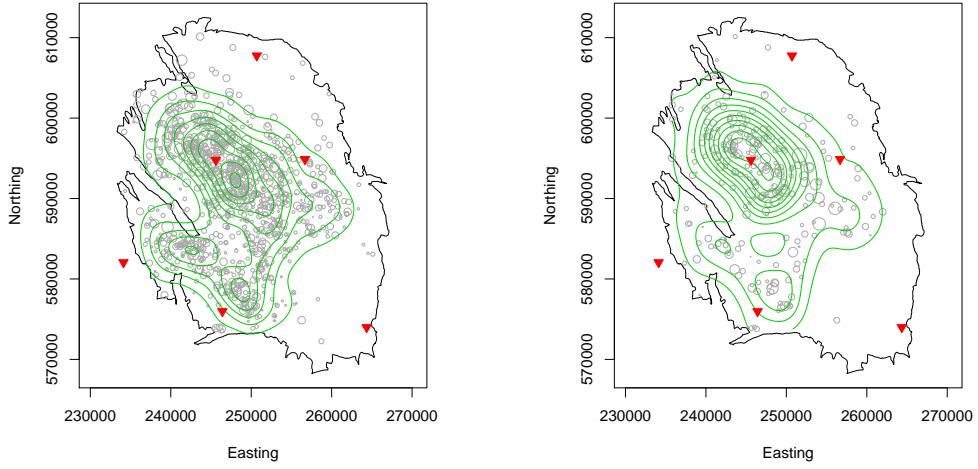


Figure 2.5.1: Contour plots of the spatial earthquake density estimated using: [left] all events since 1995, [right] all events with magnitude of at least $1.5M_L$ since 1995.

Although these plots show the spatial variability in the observed earthquake density through the field, they do not indicate whether these peaks in earthquake activity would be surprising if events were located at random across the gas field. As a measure of how unusual such a point pattern would be under complete spatial randomness, we look at Ripley's K -function (described fully in Section 3.1.4) as a measure of spatial clustering. This function describes the expected number of further events within a distance $r > 0$ of an arbitrary event in the process. Figures 2.5.2 shows the K -function estimates using the observed point patterns of all events and those above the $1.5M_L$ threshold. For values of r that are above 200 m, the estimated K -function of both point patterns exceeds the simulation envelope calculated under the assumption of complete spatial randomness. This indicates that there are a greater number of

events within a short distance of each other than would be expected under this very simple model, confirming the initial appearance of clustering. Failure to accept complete spatial randomness is unsurprising, but gives justification for investigating other models which might better describe the clustering of events. This might be through the use of either spatially inhomogeneous Poisson processes, self-exciting processes or a combination of the two.

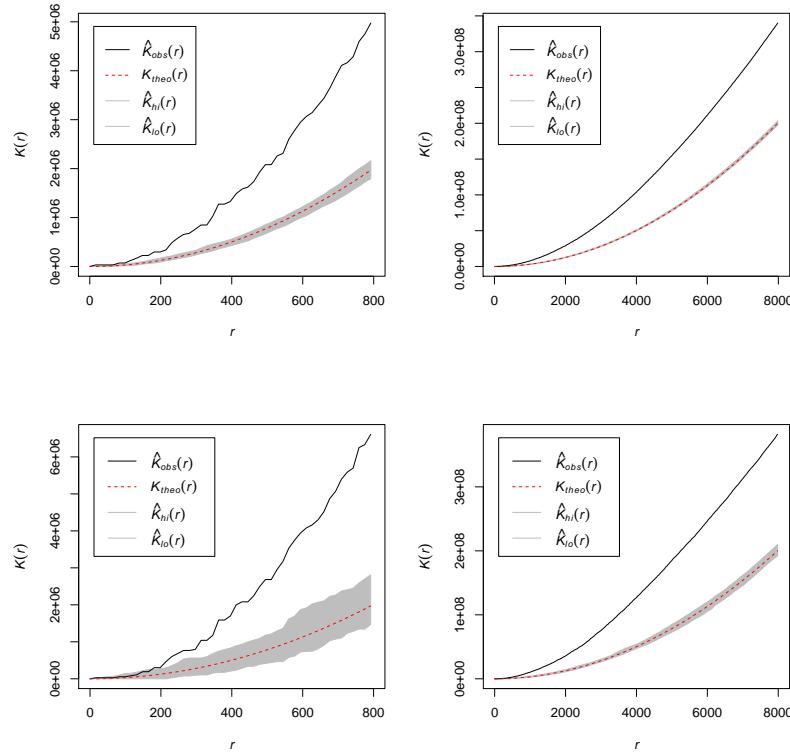


Figure 2.5.2: Estimated Ripley K -function for observed data as compared to simulation envelope for 99 realisations of complete spatial randomness: [upper] all events since 1995 and [lower] events with magnitude of at least $1.5M_L$ since 1995. Left and right plots show function estimates over short (800m) and moderate (8km) distances relative to the size of gas field.

To introduce a temporal component to the exploratory analysis, a first question of

interest is whether or not the intensity of events is changing in the same way through time at all locations. In order to test this, we consider a simple division of the field separating the two high density regions seen previously in Figure 2.5.1. This is shown in Figure 2.5.3, with the spatial modes separated by the line $y = 73500 - 0.6x$. The density of events in each of these regions through time is also shown in Figure 2.5.3.

A Monte Carlo test was carried out to test the null hypothesis that the earthquake intensity in each of these regions is changing in the same way through time. Each earthquake was labelled as belonging to one of the two high density regions. Under the null hypothesis, that each region is developing in the same way through time, estimates of the temporal intensity should remain valid under permutation of these labels.

The integrated squared error between temporal density estimates $f(t)$ and $g(t)$ on the time interval $[t_{\min}, t_{\max}]$ is

$$\text{ISE}(f(x), g(x)) = \int_{t_{\min}}^{t_{\max}} [f(t) - g(t)]^2 dt. \quad (2.5.1)$$

The integrated squared error was used as a measure of discrepancy between the temporal earthquake densities in each of the high density regions, which were each scaled to have unit integral over the observation period. The integrated squared error of the observed intensity estimates, shown in Figure 2.5.3, was compared to those obtained from 10,000 random permutations of the region labels. The probability of an integrated squared error at least as large as that observed was 0.005, strongly suggesting that the intensity of events changes differently though time depending on location. Although the significance of this difference will depend on the chosen time-frame and measure of discrepancy, this provides motivation for further work investigating the way in which spatial location interacts with changes in intensity through time.

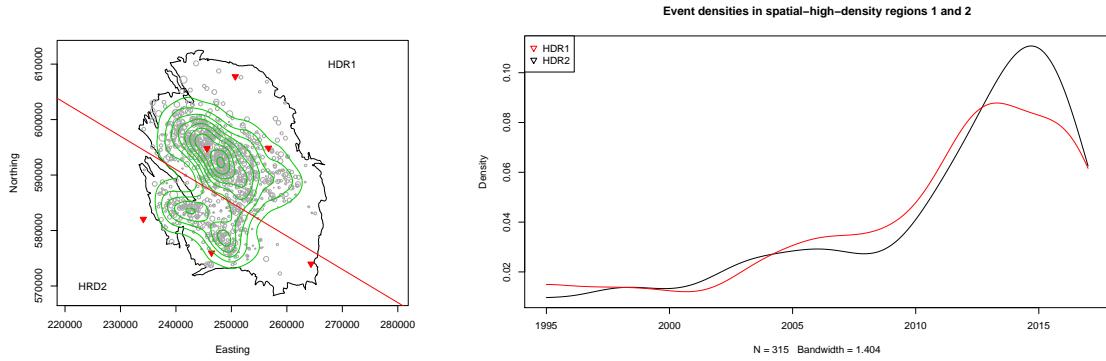


Figure 2.5.3: [Left] Splitting of field into two high density regions HDR1 and HDR2. [Right] Temporal event density estimates for HDR1 and HDR2 through time.

2.5.3 Exploratory temporal analysis

In the same way that the exploratory spatial analysis began by considering the distribution of earthquakes over the region, we begin the temporal analysis by considering their distribution through time. Figure 2.5.4 shows a histogram of the number of events per year, which gives a simple first visualisation. While the rate of events above the magnitude of completion of $1.5M_L$ is potentially constant over time, the overall rate of events appears to be increasing through time. Two points are of note here. Firstly that the catalogue up until March 2017 is used in this exploratory analysis, which explains the low event counts for 2017. Secondly, annual aggregation of event counts is motivated by the seasonal nature of gas extraction but the division into *calendar* years is an arbitrary choice. From this plot we can see that the total event count increases more dramatically over time than that of the largest events; that is likely through a combination of smaller events being detected more consistently as the sensor network developed and because small events occur with higher relative frequency.

Rather than aggregating events by year, in Figure 2.5.5 a smoothed intensity estimate for each dataset is considered instead. These are obtained using Gaussian kernel den-

sity estimation and show a similar structure whether including or excluding small magnitude events. Added to each plot are 95% simulation intervals for the estimated intensity if the event rate were truly constant over the entire period. The observed intensity estimates are both systematically below the simulation interval at the start of the period and above at the end of the period. This indicates that the intensity of earthquakes is increasing through time at both large and small magnitudes, though this change is more apparent when considering all events than only those above magnitude $1.5M_L$.

An artefact of this density estimation method is that the intensity estimates will be lower at the beginning and end of the estimation interval, because the events that might have occurred just before or after the interval do not contribute to the intensity estimate. This edge effect could somewhat be overcome by using a density estimate that uses boundary reflection methods (Silverman, 1986), but since our focus here is on the values relative to a homogeneous process this is not strictly necessary.

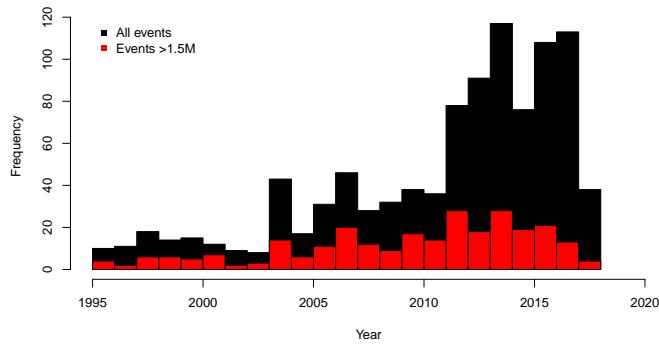


Figure 2.5.4: Histogram displaying the number of earthquakes each year since 1995.

In addition to the trend in intensity through time, the inter-arrival time between events is of interest as an indicator of the degree of temporal clustering or regularity. In such comparisons, the homogeneous Poisson process is typically used as a comparator and has exponentially distributed inter-arrival times. To compare these

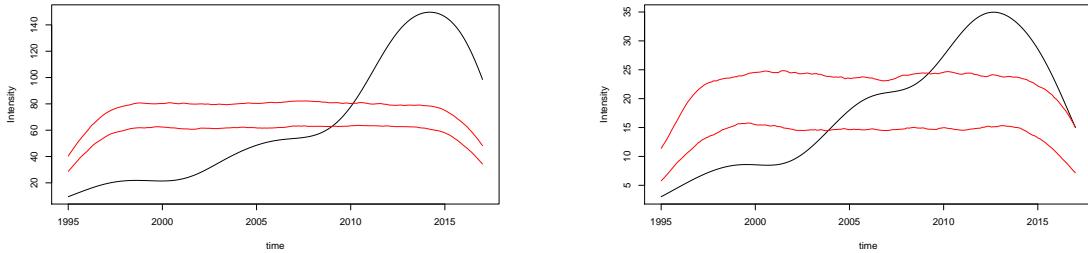


Figure 2.5.5: Estimated intensity of events through time. Red lines indicate the 95% simulation intervals under the assumption of constant intensity. [Left] All events since 1995. [Right] Events since 1995 of at least magnitude $1.5M_L$.

distributions, the observed inter-arrival times were first standardised to have mean one. The quantiles of these standardised intervals were then compared to those of the standard exponential distribution. This is shown in Figure 2.5.6. We consider only events above magnitude $1.5M_L$ in this plot, as the increasing capability to detect small earthquakes leads to identification of clustering behaviour that is actually an artefact of the detection method.

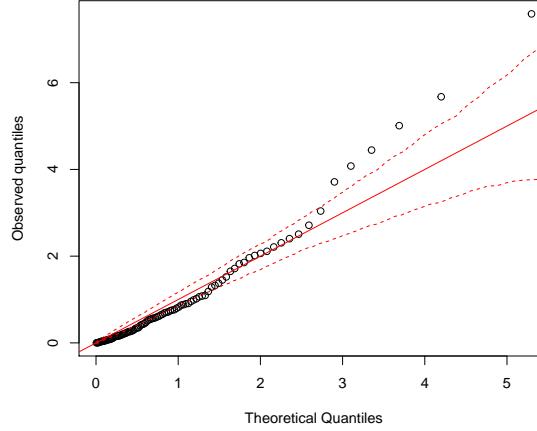


Figure 2.5.6: Quantile-quantile plot of standardised observed interval lengths against standard exponential distribution. Solid line shows $y = x$, dashed lines give 95% simulation envelope.

From Figure 2.5.6 we can see that the observed quantiles deviate significantly from an exponential distribution. The largest quantiles are larger than expected and the smallest are smaller than expected, which is consistent with temporal clustering of events. The idea of temporal clustering has the intuitive appeal that in the context of earthquakes it may be interpreted physically as aftershocks, earthquakes that are triggered by another recent earthquake. However this is not the only possible explanation, the increasing intensity over time seen in Figure 2.5.4 would also induce clustering, though of a different type.

If the mainshocks could be well modelled as a homogeneous Poisson process, so that the overall intensity was constant, then Figure 2.5.6 could be used to estimate the proportion of aftershocks within the data. Assuming that delays between a main-and aftershocks are small compared to the time between mainshocks, long inter-event intervals would consist mainly of the intervals between main-shocks. Since these follow an exponential distribution for a homogeneous Poisson process, the upper quantiles of the inter-event time distribution would therefore form a straight line on the right of the QQ-plot. The quantile at which this line begins would give an estimate of the proportion of events which are aftershocks (Ferro and Segers, 2003).

In Figure 2.5.6, there appears to be three rather than two straight line sections. There are several possible models which may yield such an interval distribution, for example a combination of a clustered point process and a background intensity that changes gradually over time. This suggests that inhomogeneous and self-exciting process models are both worth further investigating in the context of Groningen earthquakes.

2.5.4 Exploratory magnitude analysis

In addition to exploring the spatial and temporal structure of event occurrences, we can also look at the structure of the marks associated with these events. In the

assessment of seismic risk, being able to model accurately the markings of events as well as their locations is of great importance. This is because the mark associated with each event denotes the magnitude of the earthquake, indicating the amount of energy released and therefore to some extent the potential of the event to cause damage at surface level.

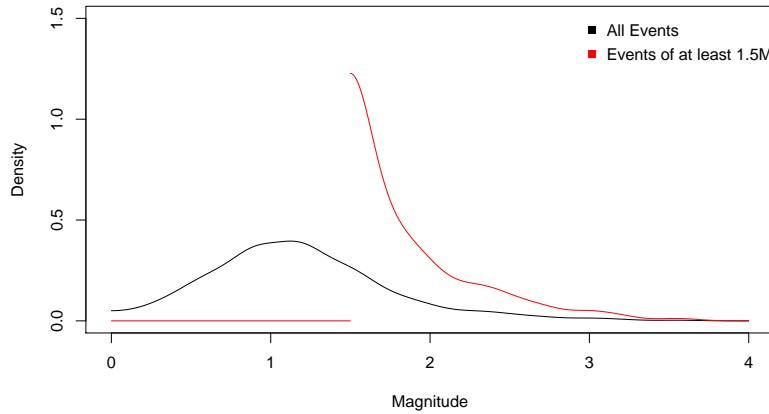


Figure 2.5.7: Gaussian kernel density estimates of magnitude distributions using all events (black) and only those of at least magnitude $1.5M_L$ (red).

Figure 2.5.7 shows kernel density estimates of the magnitude distributions of all recorded events since 1995, and the subset of those which reached at least magnitude $1.5M_L$. The Gutenberg-Richter law is a widely used descriptive model for magnitudes of seismic events. It states that in any given earthquake catalogue, the number of events N of magnitude of M or greater is given by

$$N = 10^{a-bM}, \quad (2.5.2)$$

for some constants a and b specific to the region and period. The term 10^a is a normalising constant equal to the total number of events in the catalogue N_{total} . Reworking equation (2.5.2) to be in terms of probabilities and natural logarithms, the

magnitude survivor function $\bar{F}_M(m)$ is

$$\bar{F}_M(m) = \frac{N}{N_{total}} = 10^{-a}N = (10^{-b})^m = e^{-\beta m}, \quad (2.5.3)$$

where $\beta = b \log_e 10$. This is the survivor function of an exponential distribution and so the Gutenberg-Richter law can be restated as magnitudes following an exponential distribution.

In many earthquake catalogues, the distribution of magnitudes deviates from an exponential at both the small and large magnitudes. At the lower end this is usually because the network of geophones lacks the sensitivity to detect small events and so under-counts as compared to the Gutenberg-Richter model. At the higher end, the deviation occurs because the maximum potential energy that can be stored within a region before an earthquake occurs has some finite limit. This effectively means that there is a maximum magnitude of earthquake which can be observed, whereas there is no upper limit on the tail of the exponential distribution (Vere-Jones, 2010).

In Figure 2.5.7, we can see that for events of at least magnitude $1.5M_L$, the monotonic decreasing form of the Gutenberg-Richter model seems as though it may be appropriate. When using the entire catalogue, however, the density estimate becomes uni-modal. This is likely due to events with small magnitudes typically being missed by the detection network. A change in this behaviour over time can be seen in Figure 2.5.8; small magnitude events are detected more often in the later part of the observation period, following improvements to the sensor network.

Figure 2.5.8 also demonstrates that the censoring of small magnitude events is not of a simple ‘cut-off’ form; some events below the magnitude of completion are detected. These low magnitude events are increasingly detected at later times, which may be a result of continued/changing gas production or else due to the increasing sensitivity of the geophone network. It is thought that all earthquakes of this magnitude $1.5M_L$

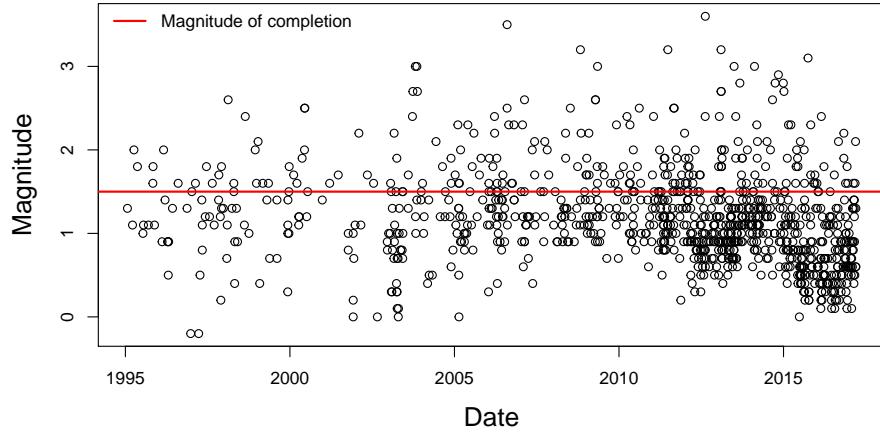


Figure 2.5.8: Magnitudes of earthquakes through time. Red line indicates the magnitude of completion used by Bourne and Oates (2017a) for this period, $1.5M_L$.

or greater have been recorded in the Groningen region since 1995. The non-standard censoring of this non-stationary process means that seismicity models for the Groningen region are typically fitted using only events of at least magnitude $1.5M_L$. This avoids biasing in the resulting parameter estimates by using incomplete data but at the cost of only using 27% of the available data, resulting in parameter estimates with greater uncertainties.

Figure 2.5.7 provides high-level evidence in support of the Gutenberg-Richter law being a suitable model for the magnitude distribution in the Groningen region above the nominal magnitude of completion. We therefore investigate the appropriateness and the goodness-of-fit of this exponential model to events large enough to be completely recorded.

As a first step, we consider whether the distribution of large magnitudes remains constant through time. Since the intensity of events is increasing over time, when making this assessment we do not split the Groningen catalogue into equal time intervals but rather into 6 consecutive intervals that each containing the same number

of events. This helps to ensure that our ability to identify the form of the magnitude distribution is similar in each interval.

Figure 2.5.9 shows, for each of the 6 time intervals, Gaussian kernel density estimates for the conditional distribution of magnitudes exceeding $1.5M_L$ using two techniques. In the first approach, only events exceeding $1.5M_L$ are used to construct the density estimate, using the boundary reflection technique of Silverman (1986) to account for edge effects. This approach avoids the inclusion of data below the stated magnitude of completion but also assumes that the density is symmetric about the boundary. This is inconsistent with the monotonically decreasing Gutenberg-Richter model and may be causing the flattening of density estimates near the boundary value of $1.5M_L$. The second approach includes events below $1.5M_L$ when estimating the conditional distribution above $1.5M_L$. If many events are missing below the boundary, this risks introducing a bias that reduces density estimates near the boundary.

Comparing the density estimates in Figure 2.5.9, including events below the magnitude threshold seems to cause less distortion than the boundary reflection method; the density estimates for each interval using the second method retain an exponential-like trend close to the threshold magnitude. This suggests that $1.5M_L$ is a conservative estimate for the magnitude of completion for the period since 1995, or that only a small proportion of events just below this level are censored. Using either method, the estimated magnitude distributions are in good agreement with one another across all time intervals, indicating that above $1.5M_L$ magnitudes follow the same distribution over time. We therefore have evidence that while the proportion of all recorded events that are above the threshold is decreasing through time, the distribution of events conditional on being above this threshold remains constant.

The Gutenberg-Richter law asserts not only that the exceedances follow the same distribution through time, but that this is an exponential distribution. To assess this

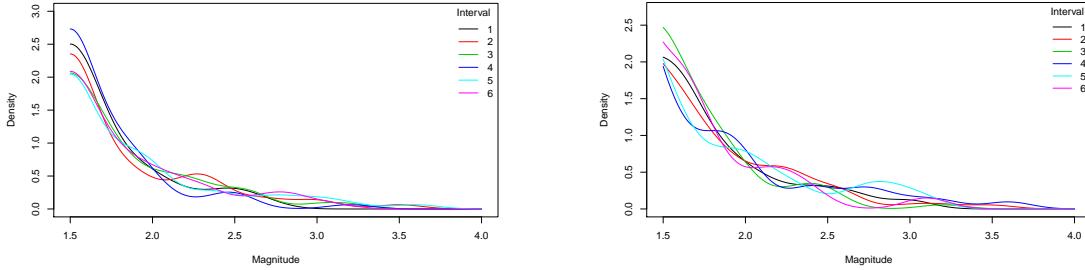


Figure 2.5.9: Gaussian kernel density estimates of the magnitude distribution in each of six consecutive intervals, conditional on events exceeding magnitude $1.5 M_L$. [Left] Density estimate excluding events below magnitude $1.5 M_L$ with boundary correction. [Right] Density estimate including events below magnitude $1.5 M_L$.

assumption we use a quantile-quantile plot for magnitudes exceeding $1.5 M_L$, shown in Figure 2.5.10. Issues arise here due to many sample quantiles taking the same value because the magnitude data is reported to only one decimal place. Therefore, both a standard (red) and a corrected (blue) tolerance interval are shown in Figure 2.5.10. The standard tolerance interval shows the typical range for quantiles of unrounded exponential data. The corrected tolerance interval show similar ranges but for exponential data that have been rounded to one decimal place. Including rounding explains some, but not all, of the additional probability mass around the threshold magnitude as compared to an exponential distribution. This is indicated by the greater coverage of the data by the rounding-corrected interval as compared to the standard exponential interval.

Taken collectively, these exploratory plots indicate that the Gutenberg-Richter model provides a reasonable starting point for modelling Groningen earthquake magnitudes but that further investigations in which rounding of the observations is properly accounted for are also warranted.

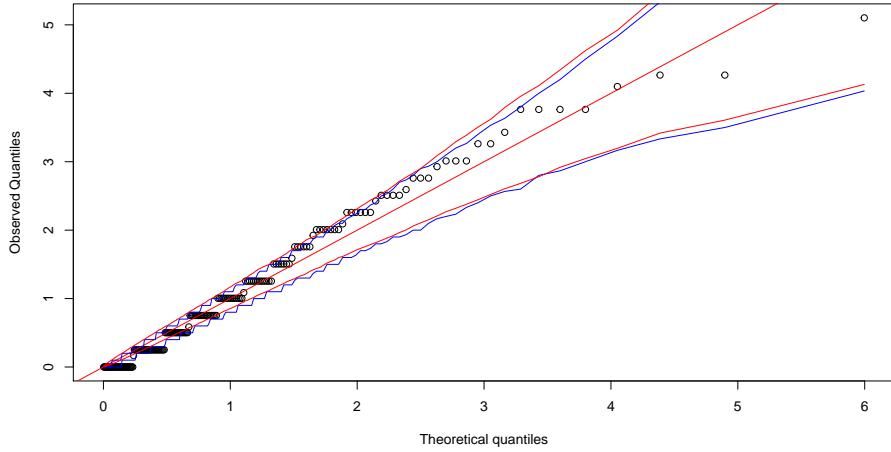


Figure 2.5.10: Quantile-quantile plot of standardised threshold exceedances against standard exponential distribution. 95 % tolerance intervals for standard exponential and rounded standard exponential quantiles are shown in red and blue respectively.

2.5.5 Review

While by no means exhaustive, this exploratory analysis has revealed many features of the data and answered several of the most pressing question regarding its modelling.

After exploratory analysis we have found that the spatial intensity of events is inhomogeneous. This was confirmed by considering the probability, under complete spatial randomness, of the originally apparent clustering of large magnitude events in the north-west of the field and the secondary cluster of smaller events in the south-west, by way of the Ripley K -function. The intensity of recorded events was also found to be increasing through time, with the additional possibility of temporal clustering which requires further investigation. Furthermore, the two high density regions of recorded earthquake activity appear to have intensity functions that are changing differently through time.

An exploratory analysis of the earthquake magnitudes revealed issues relating to the censoring of small magnitude events throughout the observation period. This censoring is incomplete and varies through time, caused by the increasing sensitivity of the geophone network. The apparent change in modal magnitude is perhaps also an artefact of improved detection of small magnitude events because the conditional distribution of magnitudes exceeding $1.5M_L$ appears to be stationary. Despite this stationarity, the Gutenberg-Richter law has limitations when modelling these large magnitude events, assigning lower than probability to magnitudes close to the threshold value.

Chapter 3

Literature review

3.1 Point process models

3.1.1 Overview

Point processes are a special case of stochastic processes. As such, they can be described by a collection of scalar- or vector-valued random variables $\mathcal{Y} = \{Y_1, \dots, Y_N\}$. Both the elements of \mathcal{Y} and the number these elements $N = |\mathcal{Y}|$ are random. Each element of \mathcal{Y} represents a point in some mathematical space, typically this observation window W will be either all or a subset of \mathbb{R}^d for $d \geq 1$. One realisation of such process is a collection of point locations in the observation window $\mathbf{y} = \{y_1, \dots, y_n\} \in W^n$.

Point processes are useful models for localised events which occur across space, time or both. Point process models have been used in forestry, epidemiology and neuroscience, and form the basis for a popular class of earthquake models. This section aims to give a brief introduction to important examples of point processes, building from simple to specialised models. Further details on point process theory are given in Cox and Isham

(1980), while Diggle (1983) covers statistical analysis of point process data.

3.1.2 Poisson point processes

Homogeneous Poisson processes A homogeneous Poisson process (HPP) is the simplest type of point process; in a HPP both the event locations and their count, $N(W) = |\{Y_i \in \mathcal{Y} : Y_i \in W\}|$, are both random. In a HPP events are located independently of one another and uniformly at random across W . The intensity of the point process λ determines the expected number of events per unit volume in W .

Consider events occurring only in time so that $W = \mathbb{R}$. In this case, it is usual to relabel the elements of \mathcal{Y} so that event indices impose a temporal ordering and $Y_1 \leq \dots \leq Y_N$. The history of the point process at time t is the set of events that occur up to time t and is denoted by $\mathcal{H}_t = \{Y_i : Y_i \leq t\}$. The intensity of the Poisson process, $\lambda > 0$, gives the expected number of events in the process per unit time. The HPP with intensity λ on \mathbb{R} can be defined by the conditions that for all $t \in W$, as $\delta \rightarrow 0_+$:

$$\mathbb{P}(N(t, t + \delta) = 1 | \mathcal{H}_t) = \lambda\delta + o(\delta); \quad (3.1.1)$$

$$\mathbb{P}(N(t, t + \delta) > 1 | \mathcal{H}_t) = o(\delta); \quad (3.1.2)$$

$$\mathbb{P}(N(t, t + \delta) = 0 | \mathcal{H}_t) = 1 - \lambda\delta + o(\delta). \quad (3.1.3)$$

It follows from conditions (3.1.1) - (3.1.3) that:

- The waiting time until the next event from an arbitrary time t , $T_t \sim \text{Exp}(\lambda)$;
- The intensity and waiting time distribution do not depend on the history of the process;
- The counting measure $N(A)$ on $A \subset \mathbb{R}$ is distributed $\text{Pois}(\lambda|A|)$, where $|A|$ is

the Lebesgue measure of A ;

- For disjoint sets $A, B \subset W$, $N(A)$ and $N(B)$ are independent.

These properties allow us to specify a HPP by giving one or more of the following properties: the intensity for every point in W , the distribution of inter-event times, or the joint distribution of the counting measure N on all subsets of W .

Inhomogeneous Poisson processes The HPP can be generalised by allowing the intensity of events to vary as a function of time or covariates. This results in an inhomogeneous Poisson process (IHPP) with intensity function $\lambda(t) : W \rightarrow \mathbb{R}_0^+$. This intensity function is defined as the instantaneous rate of events at t , i.e.,

$$\lambda(t) = \lim_{\delta \rightarrow 0_+} \mathbb{E} \left[\frac{N(t, t + \delta)}{\delta} \right]. \quad (3.1.4)$$

The survivor function of the waiting time T_t retains an exponential form but with a rate parameter that now depends on both the starting point t and the intensity function. Define the integrated intensity function on a set A as

$$\Lambda(A) = \int_A \lambda(a) da$$

and as $\Lambda(a, b)$ on open intervals of the form (a, b) . Then the survivor function of T_t has the form

$$\bar{F}_{T_t}(\tau) = \mathbb{P}(T_t \geq \tau) = 1 - \exp\{-\Lambda(t, t + \tau)\}.$$

It follows that an IHPP, also satisfies the memory-less property and has independent event counts on disjoint subsets. However, for an IHPP the distribution of counting measure $N(A)$ depends on the set A that is being considered, where $N(A) \sim \text{Pois}(\Lambda(A))$.

Time rescaling theorem Through a transformation of the time axis it is possible to transform between \mathcal{Y}_1 , an IHPP with known intensity $\lambda(t) : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, and \mathcal{Y}_2 , a

HPP on \mathbb{R}^+ with unit rate. The transformation from \mathcal{Y}_1 to \mathcal{Y}_2 is given by

$$\mathcal{Y}_2 = \tau_\lambda(\mathcal{Y}_1) := \{\Lambda(0, Y_i) : Y_i \in \mathcal{Y}_1\}. \quad (3.1.5)$$

While the reverse the transformation from a HPP with unit rate \mathcal{Y}_2 to \mathcal{Y}_1 , an IHPP with known intensity $\lambda(t) : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, is

$$\mathcal{Y}_1 = \tau_\lambda^{-1}(\mathcal{Y}_2) := \{\Lambda^{-1}(0, Y_i) : Y_i \in \mathcal{Y}_2\}, \quad (3.1.6)$$

where $\Lambda^{-1}(0, y)$ is the value of t which solves $\Lambda(0, t) = y$.

The transformation (3.1.5) is provided by the time rescaling theorem (Brown et al., 2002) and can be used to assess the fit of an IHPP model to a point pattern; the observed point pattern is transformed using the fitted intensity function and the properties of the transformed pattern can be compared to those of a homogeneous Poisson process. For example, the inter-event time distribution can be compared to an $\text{Exp}(1)$ distribution. The reverse transformation (3.1.6) gives can be used to simulate a one dimensional IHPP by transforming a HPP.

Random thinning and superposition In a random thinning of a point process, each event is either retained in the process or removed with some stated probability. A randomly thinned Poisson process remains a Poisson process. If the original process has intensity $\lambda_1(t)$ and each point is retained with probability $p(t)$, then the thinned process has intensity $\lambda_2(t) = \lambda_1(t)p(t)$. This result gives a second method of generating IHPPs with intensity function $\lambda_2(t)$: first simulate a HPP with rate $\lambda_1(t) = m$ where $m \geq \lambda_2(t)$ for all t , then retain each event in this process with probability $p(t) = \lambda_2(t)/m$.

A related operation on point processes is superposition, where events from two or more point processes are combined into a single point process. The superposition of independent Poisson processes remains a Poisson processes and the resulting intensity

is given by the sum of the component intensities. A limit theorem exists for point process superposition, which states that for a set of k suitably well-behaved, non-Poisson processes the process formed by their superposition, after sufficient scaling, converges to a Poisson process as $k \rightarrow \infty$ (Cox and Isham, 1980). This theorem helps to explain the wide applicability of Poisson processes to natural phenomena, which can often be thought of as the superposition of many sub-processes.

Spatial and spatio-temporal Poisson processes The definitions and properties of Poisson processes can be extended to spatial and spatio-temporal observation windows. For models in two or more dimensions, the interval specification of the point process becomes less useful because each inter-event distance corresponds to a set of possible locations and points which are at a similar distance from a particular location are not necessarily close to one another. The intensity specification remains valid for point processes in more than one dimension, with slight alterations to the definition.

For a spatial point process on $W \subset \mathbb{R}^2$ the intensity function at $x \in \mathbb{R}^2$ is

$$\lambda(x) = \lim_{\epsilon \rightarrow 0_+} \mathbb{E} \left[\frac{N(b_2(x, \epsilon))}{|b_2(x, \epsilon)|} \right],$$

where where $b_d(x, r) = \{s \in \mathbb{R}^d : |s - x| < r\}$ is the d -dimensional ball centred at x of radius r .

The spatial analogue of the waiting time is the radial contact distance, R_x . This is the the ℓ_2 -distance from planar location x to the nearest event. The radial contact distance R_x has distribution function:

$$\mathbb{P}(R_x \leq r) = 1 - \exp \{-\Lambda(b_2(x, r))\}. \quad (3.1.7)$$

When the Poisson process is homogeneous with rate λ , the distribution (3.1.7) no longer depends on location x and simplifies to $R \sim \text{Exp}(\lambda\pi r^2)$.

For a spatio-temporal process on $W \subset \mathbb{R}^2 \times \mathbb{R}$ the intensity function at $(x, t) \in \mathbb{R}^2 \times \mathbb{R}$ is defined using a space-time cylinder surrounding that location:

$$\lambda(x, t) = \lim_{\epsilon, \delta \rightarrow 0_+} \mathbb{E} \left[\frac{N(b_2(x, \epsilon) \times (t, t + \delta))}{|b_2(x, \epsilon) \times (t, t + \delta)|} \right].$$

For spatio-temporal processes, the waiting time distribution is defined by aggregating events over spatial dimensions at each time point and the radial contact distance by aggregating over time. Note that this does not preserve *orderliness* of the process because multiple events can occur at the same location in space or time following aggregation of the point process.

Inference for Poisson processes Let $\mathbf{y} = \{y_1, \dots, y_n\}$ be a point pattern on observation window W , from which a parametric Poisson process intensity function $\lambda(w; \theta) : W \rightarrow \mathbb{R}_0^+$ is to be inferred, where the parameters θ are in some specified parameter space. The event count and event locations are independent for a Poisson process. Therefore, the likelihood function for θ is

$$\begin{aligned} L(\theta; \mathbf{y}) &= \mathbb{P}(N(W) = n | \theta) \prod_{i=1}^n \mathbb{P}(Y_i = y_i | \theta), \\ &= \frac{\Lambda(W; \theta)^n \exp\{-\Lambda(W; \theta)\}}{n!} \prod_{i=1}^n \frac{\lambda(y_i; \theta)}{\Lambda(W; \theta)}, \\ &= (n!)^{-1} \exp\{-\Lambda(W; \theta)\} \prod_{i=1}^n \lambda(y_i; \theta); \end{aligned} \tag{3.1.8}$$

where $\Lambda(W; \theta) = \int_W \lambda(w; \theta) dw$. The corresponding log-likelihood is

$$\ell(\theta; \mathbf{y}) = \sum_{i=1}^n \{\log \lambda(y_i; \theta)\} - \log(n!) - \Lambda(W; \theta). \tag{3.1.9}$$

The parameters θ of the intensity model can be estimated by maximum-likelihood or Bayesian inference. For maximum likelihood estimation, closed form estimators can only be obtained for very simple intensity functions and so numerical optimisation routines are required. Similarly, in a Bayesian analysis conjugate prior distributions

are not available except in the simplest of cases and Markov chain Monte Carlo (MCMC) methods must be used to estimate the posterior distribution of θ .

One difficulty that can arise when fitting such models is the computational cost when evaluating the integral $\Lambda(W; \theta)$. If this integral does not have a closed form then numerical integration is required at each step of the optimisation or Markov chain. This can be very costly depending on complexity of the intensity model and the shape of the observation window. Another difficulty is that point pattern data sets are often small, making the intensity parameters difficult to estimate precisely. A Bayesian approach to modelling can ease this problem if domain specific knowledge can be used to help constrain parameter values through the choice of prior distribution.

3.1.3 Generalisations of the Poisson process

3.1.3.1 Renewal Processes

Renewal processes generalise the interval specification of homogeneous Poisson processes in one dimension. For a one-dimensional HPP, the intervals between events are independent and identically distributed exponential random variables (Cox and Isham, 1980). Renewal processes generalise this to allow independent intervals with some other distribution function G .

The dispersion of G controls the degree of regularity or clustering in the process. If G is less dispersed than the exponential density then the renewal process will be more regular than a Poisson process. If the dispersion is larger, then the interval lengths in the renewal process are more irregular, leading to clustering of events in the process. Renewal processes encompass Poisson processes but allow greater flexibility in clustering behaviour. However, intervals lengths remain independent and the model does not easily extend to higher dimensional spaces.

3.1.3.2 Linear self-exciting processes

The intensity function definition of a point process (3.1.4), can be generalised to allow dependence on the history of the process, so that

$$\lambda(t|\mathcal{H}_t) = \lim_{\delta \rightarrow 0+} \frac{1}{\delta} \mathbb{E}[N(t, t + \delta)|\mathcal{H}_t].$$

Poisson process intensities are independent of their history, and renewal process intensities at time t depend only on the instant before t . Self-exciting processes extend this dependence to allow some or all previous events to influence the intensity at time t . A non-negative background intensity $\mu(t)$ is supplemented by contributions to the integrated intensity by each previous event. The amount by which the integrated intensity is increased by each previous event is denoted by $r_0 \geq 0$, which gives the expected number of further events triggered by each event. The allocation of this additional intensity over time is determined by the kernel function, $w(\tau)$, which is defined on $\tau \geq 0$ and integrates to 1 over this support. The amount and allocation of additional intensity determines the level of clustering within the process.

A linear self-exciting process has an intensity function of the form

$$\lambda(t; \mathcal{H}_t) = \mu(t) + \sum_{Y_i \leq t} r_0 w(t - Y_i). \quad (3.1.10)$$

This intensity function can be viewed as the superposition of $N + 1$ IHPPs. Each component represents one cause of events; the background intensity and each of the N events themselves. The process can therefore be seen as a composite arrival and branching process, where each event generates a Poisson number of offspring. The expected number of offspring per event, r_0 , must be below 1 for the process to be sub-critical and have stable long-term properties.

The intensity specification of self-exciting processes allows the extension of the model to spatial and spatio-temporal supports. Self-exciting processes can be made more

flexible by allowing $\mu(t)$, r_0 or $w(\tau)$ to depend on covariates. In the modelling of earthquakes, self-exciting models provide a way of jointly modelling earthquakes and aftershock activity.

3.1.3.3 Doubly stochastic point processes

The final extension of a Poisson process treats the intensity as a random function. These point processes are called doubly stochastic point processes or Cox processes.

Let $\{\Lambda(t)\}$ be a real-valued non-negative stochastic process of preassigned structure, with history at time t given by $\mathcal{H}_t^\Lambda = \{\Lambda(\tau) : \tau < t\}$. Then the complete intensity function of the doubly stochastic Poisson process is:

$$\lambda(t; \mathcal{H}_t, \mathcal{H}_t^\Lambda) = \lim_{\delta \rightarrow 0+} \frac{1}{\delta} \mathbb{E} [N(t, t + \delta) \mid \mathcal{H}_t, \mathcal{H}_t^\Lambda] = \lambda(t).$$

Unlike the realisation of the events from the point process, the realisation of the intensity function is not typically observed. Cox processes therefore belong to the class of latent models and, depending on the specification of $\{\Lambda(t)\}$, are capable of representing a variety of complex intensity functions. A common form is to suppose that the log intensity is a Gaussian process (Baddeley, 2008). The flexibility of these models means that very large datasets, repeated observations or strong prior knowledge are required to constrain the model.

3.1.3.4 Multi-type and marked processes

All previous examples of point processes have considered events to consist only of a location in space, time or both. Covariate information is often available on the events themselves, as well as on the observation window. To distinguish between these, event covariates are described as *marks*. A marked point process has one or more marks

assigning to each event. These marks could be external covariates or they could be used to extend point process models. For example, integer valued marks can be used to allow multiple events to occur at the same location, where the mark denotes event multiplicity. Multi-type point processes can be modelled by using categorical marks denote the type of each event. This allows, for example, within-type clustering but between-type independence.

3.1.4 Measures of clustering

To determine if models that allow clustering are appropriate for an observed point pattern requires a method to measure the observed degree of clustering. This measurement can then be compared to the expected behaviour of a Poisson process to give evidence of relative clustering or regularity in the pattern.

The second order intensity function describes the level of clustering of events in a point process. For a point process on $W \subset \mathbb{R}^2$ and arbitrary locations $x_1, x_2 \in W$ the second order intensity function is given by

$$\lambda_2(x_1, x_2) = \lim_{\delta, \epsilon \rightarrow 0_+} \left\{ \frac{\mathbb{E}[N(b_2(x_1, \delta))N(b_2(x_2, \epsilon))]}{|b_2(x_1, \delta)||b_2(x_2, \epsilon)|} \right\}.$$

The second order intensity function lacks the easy physical interpretation of the first order intensity function. For this reason the reduced second moment function, or Ripley K -function, was introduced. For a stationary, isotropic and orderly process with intensity λ , the function λ_2 depends on x_1 and x_2 only through the distance between them, $r = |x_1 - x_2|$. The Ripley K -function is then

$$K(s) = 2\pi\lambda^{-2} \int_0^s \lambda_2(r)rdr. \quad (3.1.11)$$

The K -function has a physical interpretation as the expected number of further events within distance s of an arbitrary event, per unit intensity. At small distances this

function will therefore have relatively high values for clustered patterns, and low for regular ones. To make this comparison a homogeneous Poisson process is used as a benchmark, with $K(s) = \lambda\pi s^2$.

The physical interpretation of the K -function suggests a means of its estimation from a point pattern. The K -function may be estimated by first estimating the intensity and then finding the sample mean number of further events within distance s of an arbitrary event. When considering a finite observation window this estimate should be corrected for the bias caused by not observing events outside that window. Let d_{ij} be the distance between events i and j . Also let w_{ij} be the proportion of the circumference of the circle centred at event i of radius d_{ij} that is contained in the window. An edge corrected estimate for the Ripley K -function is then:

$$\hat{K}(s) = \frac{1}{\hat{\lambda}n} \sum_{i=1}^n \sum_{j \neq i} w_{ij} \mathbb{I}\{d_{ij} \leq s\},$$

where $\hat{\lambda}$ is an estimate of the intensity λ and n is the number of events within the window. Baddeley et al. (2000) and Marcon and Puech (2009) later considered generalisations of the K -function to inhomogeneous processes. The second order intensity function and K -function may be extended to spatio-temporal point process. Definitions and techniques for estimating these extensions are detailed in Dorai-Raj (2001).

3.2 Extreme value methods

3.2.1 Overview

The usual aim of statistical modelling is to closely represent the centre of a probability distribution or the typical values of a stochastic process. Standard statistical methods were developed with this aim in mind, but in many applications it is not typical values

that are of interest. Rather, it is the particularly high or low *extreme* values that are of interest. A model is therefore required for the tails rather than the body of the probability distribution. It is not appropriate to use standard modelling approaches in these settings because they are driven primarily by the large number of non-extreme observations.

Extreme value methods provide statistically rigorous models for the tails a probability distribution or random process. These models are fitted exclusively to, or with strong emphasis on, data from the tail of the distribution. They are therefore not compromised by the abundance of central values in the observed data. Extreme value models also provide a principled way of extrapolating beyond the observed levels of the process. This is justified by deriving asymptotic models for extremal behaviour and then using these models as approximations for the behaviour at high but finite levels of a process.

The ability to extrapolate is particularly important when models are being used to assess hazard and risk, which are strongly influenced by the extremal properties of both the damage and protective mechanisms. As such, extreme value methods have been used widely in areas where risk estimation is important, including: finance, hydrology, and process control (Coles, 2001). To date, the use of these methods within seismology has been limited.

This section provides an overview of the asymptotic motivation for commonly used univariate extreme value models and their estimation.

3.2.2 Block maxima approach

Let X_1, \dots, X_n be a sequence of independent, identically distributed (i.i.d.) random variables with unknown distribution function F . Define the maximum of this sequence

to be $M_n = \max(X_1, \dots, X_n)$. The distribution function of M_n is then given by

$$\begin{aligned}\Pr(M_n \leq x) &= \Pr(X_1 \leq x, \dots, X_n \leq x) \\ &= \Pr(X_1 \leq x) \dots \Pr(X_n \leq x) \\ &= \{F(x)\}^n.\end{aligned}$$

The distribution function of M_n could be estimated by constructing an estimator \hat{F} for F , but this approach is highly sensitive to changes in the estimated distribution function \hat{F} . An alternative approach is to consider the distribution function of M_n as the length of the sequence X_1, \dots, X_n grows. Unfortunately, the distribution of M_n converges to a point mass on the upper end point of F :

$$M_n \rightarrow x^F \text{ as } n \rightarrow \infty \text{ where } x^F = \sup \{x : F(x) < 1\}.$$

This issue can be overcome by obtaining a sequence of linear transformations on M_n that result in a non-degenerate limit distribution. Define M_n^* to be

$$M_n^* = \frac{M_n - b_n}{a_n},$$

for sequences of constants $a_n > 0$ and $b_n \in \mathbb{R}$, which stabilise the location and scale of M_n^* to avoid the degeneracy of its distribution as n increases. The Extremal Types Theorem (Fisher and Tippett, 1928) states that if these sequences of normalising constants exist, then as $n \rightarrow \infty$:

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) \rightarrow G(x), \quad (3.2.1)$$

where G is distribution function of a Fréchet, Gumbel or negative Weibull random variable. These distributional forms are united in a single parameterisation by the Unified Extremal Types Theorem. The resulting generalised extreme value (GEV) family of distribution functions has the form

$$G(x) = \exp \left\{ - \left[1 + \xi \frac{x - \mu}{\sigma} \right]_+^{-1/\xi} \right\}, \quad (3.2.2)$$

where $x_+ = \max(x, 0)$, $\sigma \in \mathbb{R}^+$ and $\mu, \xi \in \mathbb{R}$. The parameters μ, σ and ξ have respective interpretations as location, scale and shape parameters. Positive values of ξ correspond to a Fréchet distribution and a heavy upper tail. When $\xi = 0$ the GEV is equivalent to a Gumbel distribution and has an exponential upper tail. Negative values of ξ correspond to a negative Weibull distribution, which is light tailed and has a finite upper end point.

A linear transformation of a GEV random variable remains within the GEV family but has different parameter values; this means that if the distribution of M_n^* can well approximated by a GEV distribution then so can the distribution of M_n . This result motivates the use of the GEV distribution as an asymptotic model for finite sample maxima, analogous to the central limit theorem motivating a Gaussian model for finite sample means.

To estimate the parameters of the GEV distribution, the observed sequence X_1, \dots, X_n is separated into m blocks of equal length as in Figure 3.2.1. The sample maxima in each of these blocks may be treated as an approximate sample from the GEV distribution of interest and used to estimate its parameters in a frequentist or Bayesian framework.

3.2.3 Peaks over threshold

An alternative approach to modelling extremal behaviour uses all observations that exceed some suitably high threshold u . This approach can make use of more of the available data when multiple extreme events occur within a single block, as shown in Figure 3.2.1, and can also be used when data are not regularly sampled.

To derive an asymptotic model for this type of data, let X_1, \dots, X_n be a sequence of

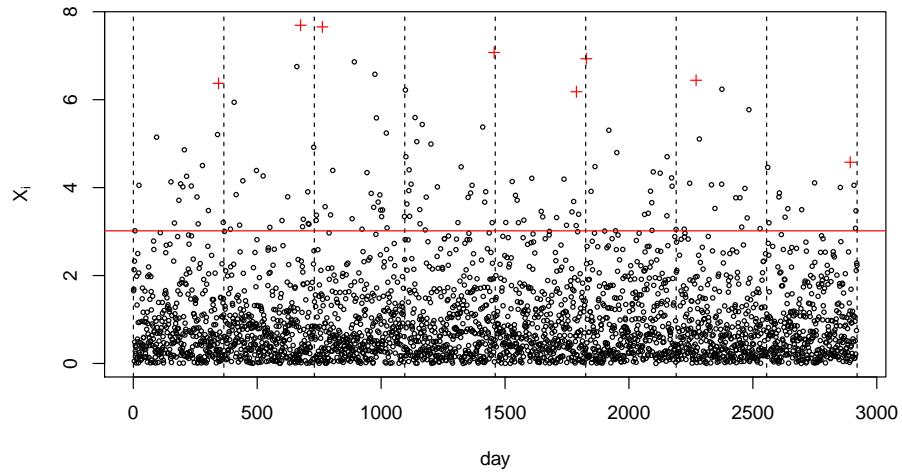


Figure 3.2.1: Simulated daily data. Red crosses show the values in an annual maxima extreme value analysis. Red line indicates threshold value $u = 3.0$, all exceedances of which are used in a peaks over threshold or point process approach to extreme value analysis.

i.i.d. random variables with common distribution function F . Let

$$N_n(x) = \sum_{i=1}^n \mathbb{I}\{X_i > a_n x + b_n\},$$

where the normalising constants a_n and b_n satisfy the conditions of limit (3.2.1) and $\mathbb{I}\{A\}$ is an indicator of event A . This random variable counts the exceedances of $a_n x + b_n$ among the sequence X_1, \dots, X_n . Since the X_i are i.i.d. random variables,

$$N_n(x) \sim \text{Binom}(n, 1 - F(a_n x + b_n)). \quad (3.2.3)$$

Taking logs of the limiting result (3.2.2), we know that

$$n \log F(a_n x + b_n) \rightarrow \log G(x).$$

Applying a first-order Taylor expansion to the left hand side and then negating gives that, for all values of x ,

$$n [1 - F(a_n x + b_n)] \rightarrow -\log G(x) = \left[1 + \xi \frac{x - \mu}{\sigma} \right]_+^{-1/\xi}. \quad (3.2.4)$$

It follows from (3.2.3) and (3.2.4) that as $n \rightarrow \infty$ the exceedance count

$$N_n(x) \rightarrow N(x), \quad \text{where} \quad N(x) \sim \text{Pois} \left(\left[1 + \xi \frac{x - \mu}{\sigma} \right]_+^{-1/\xi} \right).$$

Additionally, for values of $x > u$:

$$\mathbb{P}(X_i > a_n x + b_n | X_i > a_n u + b_n) \rightarrow \frac{\log G(x)}{\log G(u)} = 1 - H_u(x),$$

where

$$H_u(x) = \begin{cases} 1 - \left[1 + \xi \frac{x-u}{\sigma_u} \right]_+^{-1/\xi} & \text{if } \xi \neq 0, \\ 1 - \exp \left(-\frac{x-u}{\sigma_u} \right) & \text{if } \xi = 0. \end{cases} \quad (3.2.5)$$

The distribution function (3.2.5) defines a generalised Pareto random variable with parameters for the threshold $u \in \mathbb{R}$, scale $\sigma \in \mathbb{R}$, and shape $\xi \in \mathbb{R}$ of the distribution.

The shape parameter is equal to that of the corresponding GEV distribution for block maxima and does not depend on the choice of threshold. The scale parameter is dependent on the choice of threshold and is linked to that of the corresponding GEV distribution by $\sigma_u = \sigma + \xi(u - \mu)$. Additionally, the threshold stability property of the GPD (Davison and Smith, 1990) states that if $X - u | X > u \sim \text{GPD}_u(\sigma_u, \xi)$ then for a higher threshold v , $X - v | X > v \sim \text{GPD}_v(\sigma_u + \xi(v - u), \xi)$.

The GPD family has three sub-classes depending on the value of ξ .

- If $\xi < 0$ then the GPD is heavy-tailed and $X - u | X > u$ is Pareto distributed with scale parameter σ/ξ and shape parameter $1/\xi$.
- If $\xi = 0$ then the GPD has an exponential tail and $X - u | X > u$ is exponentially distributed with expectation $1/\sigma$.
- If $\xi > 0$ then the GPD has a light tail and the distribution has a finite upper end point $x^G = u - \sigma_u/\xi$.

The limit (3.2.5) motivates the use of the GPD as a model for the exceedances of a high threshold u . This model does not depend on the generating distribution F ,

and so is widely applicable. After choosing a threshold value, exceedances of this level may be treated as approximate samples from a GPD and used to estimate its parameters.

3.2.4 Point process representation

A further generalisation of the extreme value model is to consider exceedances of a high threshold u as a point process and to examine the properties of this process as the number of observations becomes large and the threshold approaches x^F .

Let X_1, \dots, X_n be a sequence of independent random variables with common distribution function F . Assume that the conditions for the limit distribution (3.2.1) hold and define the sequence of point processes P_1, P_2, \dots on $[0, 1] \times \mathbb{R}$ where

$$P_n = \left\{ \left(\frac{i}{n+1}, \frac{X_i - b_n}{a_n} : i = 1, \dots, n \right) \right\}. \quad (3.2.6)$$

This sequence converges to a non-degenerate, inhomogeneous Poisson process P as $n \rightarrow \infty$. Large values of X_i are retained in P , while small values are normalised to a common value b_l . The limit process P has intensity function

$$\lambda(t, x) = \frac{1}{\sigma} \left[1 + \xi \frac{x - \mu}{\sigma} \right]_+^{-1-1/\xi}, \quad (3.2.7)$$

on $(t, x) \in [0, 1] \times [b_l, \infty)$. The link to the peaks over threshold model can be seen by considering regions of the form $A = (t_1, t_2) \times [u, \infty)$. The integrated intensity function on such regions is

$$\Lambda(A) = \int_u^\infty \int_{t_1}^{t_2} \lambda(t, x) \, dt \, dx = (t_2 - t_1) \left[1 + \xi \left(\frac{u - \mu}{\sigma} \right) \right]_+^{1/\xi}.$$

The Poisson distribution for the total exceedance counts (3.2.3) can be shown by considering $(t_1, t_2) = (0, 1)$. This limit process also demonstrates the independence and identical distribution of exceedance sizes, which correspond to the generalised Pareto survivor function (3.2.5).

Parameters of the extreme value model can be estimated by assuming that the Poisson process limit holds exactly above some high threshold u . Standard estimation methods for parametric point process intensities may then be used to fit the extreme value model to an observed point pattern.

Extreme value models are often reported in terms of the distribution for annual maxima. This can be easily incorporated into the point process model by introducing a scaling factor m into the intensity function (3.2.7);

$$\lambda(t, x) = \frac{m}{\sigma_m} \left[1 + \xi \frac{x - \mu_m}{\sigma_m} \right]_+^{-1-1/\xi}. \quad (3.2.8)$$

The resulting parameter estimates are equivalent to the GEV parameters for annual maxima if m is chosen to be the duration (in years) of the data. If m is chosen to be the number of threshold exceedances then the parameter estimates are equivalent to those of the corresponding GPD. For any value of m the resulting parameter estimates are do not dependent on the choice of threshold u . The point process formulation is therefore particularly useful when the threshold, size or rate of exceedances change over time.

3.2.5 Inference for extreme value models

Each of the extreme value models can be fitted in a likelihood or Bayesian framework, with a frequentist approach being most common. Closed forms are not available for the maximum likelihood estimators in any of these models and so to find estimates numerical optimisation routines are required. When the shape parameter $\xi > -1/2$ the maximum likelihood estimators are asymptotically Gaussian in the number of block maxima or threshold exceedances (Smith, 1985). This property holds in the majority of physical applications of extreme value models and allows standard confidence interval construction using the delta method or the profile deviance function.

The case where $\xi \leq -1/2$ corresponds to an upper tail that is very short and the estimators converge at a greater rate.

Bayesian modelling requires a prior distribution to be specified on the model parameters. Conjugate priors do not exist for these models and so Markov chain Monte Carlo (MCMC) methods are required for inference. A Bayesian approach can therefore be computationally costly, but does allow expert knowledge of the process to be included through the choice of prior. This can be highly beneficial to inference in the low data, extreme value setting. Bayesian methods also provide natural estimation of parameter uncertainty, avoiding the theoretical complications of maximum likelihood.

A detailed description of the frequentist inference procedure for each of these models is given in Coles (2001), and the Bayesian analogues in Coles and Tawn (1996) and Sharkey and Tawn (2017).

In each of these models the threshold value, or equivalently block length, must be chosen. This choice presents a trade-off between bias and variance. A low threshold risks the asymptotic model being valid for only a portion of the data used, biasing parameter estimates. Conversely, a high threshold reduces the amount of available data and provides lower precision estimates of the parameters. There are no exact methods for choosing the threshold value. Diagnostics used to guide this choice are usually based on demonstrating deviation from the limit distribution and are reviewed in Scarrott and MacDonald (2012).

3.3 Earthquake modelling

3.3.1 Overview

The are a great many models for seismicity in either space or time, with far fewer spatio-temporal models available. A reason for this is the large computational cost of fitting and implementing spatio-temporal models and the relatively small amount of information available to fit these. Earthquake models may be descriptive or conceptual. Descriptive models aim only to capture the important properties of the earthquake catalogue, while conceptual models aim to describe the process generating the earthquakes. The mechanics of geological systems are not fully understood and so all models must fall somewhere between these two extremes. Appropriately combining a descriptive model with physical insights is therefore a key component of statistical seismicity modelling.

Point process models of seismicity can range from purely descriptive to very conceptual depending on the intensity specification. Epidemic type aftershock sequence (ETAS) point process models are toward the descriptive end of this scale and are widely used to model tectonic earthquakes (Zhuang et al., 2012). These models are special cases of the self-exciting processes introduced in Section 3.1.3.2 and will form the focus of Section 3.3.2. Following this, Section 3.3.3 considers physically motivated statistical models specific to the Groningen field, which aim to explain as well as describe earthquake occurrences.

3.3.2 Epidemic Type Aftershock Sequence models

Labelling earthquakes as foreshocks, mainshocks and aftershocks is not a straightforward task, even retrospectively, and remains open area of research (van Stiphout et al., 2012; Benali et al., 2020). ETAS models draw inspiration from epidemiology

and avoid the need for this classification. Instead, all earthquakes are treated equally. Each earthquake has the potential to trigger further earthquakes and the propensity to do so is determined by its magnitude.

The initial ETAS model of Ogata (1988) is a temporal marked point process model for earthquake occurrences times and their magnitudes. In this model, each event represents a time-magnitude pair $Y_i = (T_i, M_i)$ for $i = 1, \dots, N$ and the history of the process includes the marks as well as times of previous events; $\mathcal{H}_t = \{Y_i = (T_i, M_i) : T_i \leq t\}$. The point pattern and associated marks are modelled independently, with marks assumed to be i.i.d. with probability density function $f(m)$. This magnitude distribution is often taken to be the Gutenberg-Richter model,

$$f(m) = \begin{cases} \beta \exp\{-\beta(m - m_0)\} & \text{for } \beta > 0 \text{ and } m \geq m_0, \\ 0 & \text{otherwise,} \end{cases} \quad (3.3.1)$$

where m_0 is the minimum event magnitude. The Gutenberg-Richter model is equivalent to an $\text{Exp}(\beta)$ density translated by m_0 , where earthquakes with magnitudes less than m_0 are considered too small to cause a hazard or to induce further earthquakes. When this magnitude model is used, $Y_i \in W \times (m_0, \infty)$ where $W \subseteq \mathbb{R}$.

The intensity function of this ETAS model is given in equation (3.3.2). This intensity is a superposition of a background process with constant rate, μ , and intensity contributions from the previous events depending on their times t_i and magnitudes m_i ;

$$\lambda(t; \mathcal{H}_t) = \left[\mu + \sum_{i: t_i < t} \kappa(m_i) h(t - t_i) \right], \quad (3.3.2)$$

where κ and h are functions defined as follows.

The productivity function $\kappa(m)$ is a function giving the expected number of events that are triggered directly by an earthquake of magnitude m . The expected number of triggered events is dependent on the triggering event magnitude through a relation

of the form:

$$\kappa(m) = \begin{cases} A \exp\{\alpha(m - m_0)\} & \text{for } m \geq m_0 \text{ and } A \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

The allocation function $h(t)$ is a probability distribution describing the time delay between triggering and triggered events. The time delays until these aftershocks are usually described by the Omori-Utsu law. When modelling the temporal intensity of aftershocks t time units after a main-shock, $\lambda(t)$, for the large Nobi earthquake of 1891, Omori (1894) found that a relation of the form:

$$\lambda(t) = \frac{K}{(t + c)} \mathbb{I}\{t > 0\} \quad \text{for } c, K \geq 0, \quad (3.3.3)$$

provided a good fit to the observations. This was later generalised by Utsu (1957), who suggested that the decay through time could vary across catalogues and proposed a relation of the form

$$\lambda(t) = \frac{K}{(t + c)^p} \mathbb{I}\{t > 0\} \quad \text{for } c, K \geq 0 \text{ and } p \geq 0. \quad (3.3.4)$$

This relationship is known as the Omori-Utsu or modified Omori law. It was shown by Utsu et al. (1995) to describe many aftershock sequences and that the temporal decay in intensity was independent of the magnitude of the initial earthquake.

The temporal ETAS model was extended to space and time in Ogata (1998). The intensity function of this process is given in equation (3.3.5), where $x \in \mathbb{R}^2$. This model allows for a spatially varying background intensity $\mu(x)$ and the spatial distribution of triggered events depend on the magnitude and location of the triggering earthquake through the spatial kernel $g(x, m)$. Again, the magnitudes are assumed to be independent of time, location and cause of event. This formulation gives the model

$$\lambda(x, t, m | \mathcal{H}_t) = f(m) \left[\mu(x) + \sum_{i: t_i < t} \kappa(m_i) h(t - t_i) g(x - x_i, m_i) \right]. \quad (3.3.5)$$

The ETAS model is now well studied, both in theory and in practice. This model and its extensions are commonly used to describe tectonic seismicity. A review of further extensions to the ETAS model is given in Zhuang et al. (2011). These extensions include models for $\mu(x)$ using splines, Gaussian processes, and adaptive piecewise-constant functions (Kolev and Ross, 2020; Molkenthin et al., 2020; Ogata, 2011). Models with spatially varying kernels and magnitude parameters have also been proposed. Such model extensions are only feasible for large earthquake catalogues. The ETAS model is not the only descriptive model of seismicity available. Zhuang et al. (2011) reviews alternative models, with references to more detailed descriptions.

3.3.3 Physics-based modelling

3.3.3.1 Elastic thin-sheet models

Physics-based seismicity models must be developed in the context of the geological structures and the mechanism driving seismicity in a particular study region. In this review, we therefore focus on the elastic thin-sheet models for the Groningen reservoir developed in Bourne and Oates (2017a). These models build upon previous work in Bourne et al. (2014) by incorporating additional assumptions about reservoir properties to better link the gas extraction and earthquake processes. In Bourne and Oates (2017a), the reservoir is modelled as a porous, elastic, thin sheet which deforms due to the observed pore pressure depletion. This deformation causes additional stress on a heterogeneous network of faults, which each have some initial stress state. When the combined initial and added stress exceeds a critical value, the faults will slip and release the stored potential energy as an earthquake. Several inhomogeneous Poisson process models are constructed to describe resulting spatio-temporal intensity of induced earthquakes based on variations in the assumptions about reservoir properties.

The simplest form of the model considers a uniform pressure change within an homogeneous, isotropic, linear-elastic reservoir of infinite extent, which contains a network of preexisting faults with i.i.d. initial stresses. Since the reservoir is assumed to be linear-elastic, the additional stresses acting on faults are proportional to pressure depletion. Under this model a fault will slip and cause an earthquake after a pressure depletion of Δp if its initial stress C that was within $m\Delta p$ of its critical stress c_{crit} , where m is a material constant. In this way, induced earthquakes correspond to those faults with the highest initial stresses. The fraction of faults which fail at a particular level of reservoir depletion therefore depends on the upper tail of the initial stress distribution. This initial stress distribution is unknown, but extreme value theory provides an asymptotically motivated form for its tail.

If C is the random initial stress on a particular fault and u is a high quantile of the initial stress distribution, then $C - u | C > u \sim \text{GPD}_u(\sigma, \xi)$ provides an asymptotically motivated model for extreme initial stresses. Since a fault fails if it has initial stress within $m\Delta p$ of c_{crit} , the probability of failure is

$$\begin{aligned} P_f &= \mathbb{P}(C + m\Delta p \geq c_{\text{crit}}) \\ &= \mathbb{P}(C \geq c_{\text{crit}} - m\Delta p | C > u) \mathbb{P}(C > u) \\ &= \begin{cases} (1 + \frac{\xi}{\sigma}(c_{\text{crit}} - m\Delta p - u))_+^{-1/\xi} \mathbb{P}(C > u) & \text{if } \xi \neq 0, \\ \exp\left\{-\frac{1}{\sigma}(c_{\text{crit}} - m\Delta p - u)\right\} \mathbb{P}(C > u) & \text{if } \xi = 0. \end{cases} \end{aligned}$$

Fault failure probabilities are derived in a similar way for cases where the properties of the reservoir are less restricted; these allow for heterogeneous reservoir thickness and for there to be preexisting vertical offsets across faults. Using these failure probabilities, a Poisson point process for extreme threshold failures may be specified. The intensity function $\lambda(t)$ at time t for this process can be expressed in terms of the fault

failure probability:

$$\lambda(t) = \rho h \frac{\partial P_f}{\partial t} = \rho h \frac{\partial P_f}{\partial (\Delta p)} \frac{\partial (\Delta p)}{\partial t},$$

where ρ is the volume density of faults and h is reservoir thickness.

This type of intensity function was constructed for a range of reservoir models, the properties of which are detailed in Table 3.3.1. In these models the bulk modulus of the reservoir may be homogeneous or may vary laterally, the deformation of the reservoir may be elastic or plastic (elastic deformation models are based on pressure depletion while plastic deformation models are based on vertical compaction), and the failure probability may respond to depletion or compaction in a linear, exponential or generalised Pareto relationship.

Heterogeneity				
Model	Geometric	Elastic	Covariate	P_f response
Homogeneous	None	None	None	Constant
PT	None	None	Pressure Depletion	Linear
EPT	None	None	Pressure Depletion	Exponential
CT	None	✓	Compaction	Linear
ECT	None	✓	Compaction	Exponential
EST	✓	✓	Strain	Exponential
GPST	✓	✓	Strain	Generalised Pareto

Table 3.3.1: Reservoir models considered by Bourne and Oates (2017a)

Each of these models were fitted to the earthquake catalogue for the Groningen reservoir. To avoid the issues associated with incomplete earthquake detection, the models were fitted using events of magnitude $1.5M_L$ or greater in the period between 1995 and 2017. The models were fitted in a Bayesian framework using independent uni-

form prior distributions for the parameters. Exponential prior distributions were also considered, and achieved the same model rankings as the uniform prior.

For evaluation of the models, the data were split into training and test periods. The division between these was taken as the 1st of January 2012, to match with the five year predictions required for the Groningen production plans. The models were then evaluated by likelihood- and simulation-based testing. For likelihood-based testing, the posterior predictive distribution of the likelihood of the test data was calculated for each model in Table 3.3.1. These distributions were then used for model comparison, where models with larger modal values and small variability about this value are preferred. Simulation testing was also performed, to compare the properties of catalogues simulated under each fitted model to in the test and training portions of the catalogue. These properties included the temporal intensity of events, and the spatial distribution of events.

The likelihood-based testing revealed the following order of model performances using the same model codes as in Table 3.3.1, where $A < B$ and $A \leq B$ show strong and weak preferences for model B over model A :

$$\text{PT} < \text{homogeneous} < \text{CT} < \text{EPT} < \text{ECT} < \text{GPST} \leq \text{EST}.$$

Simulation testing revealed that the exponential and generalised Pareto trend models were better able to describe the temporal developments in earthquake intensity. It also revealed that strain-based models were better able describe the spatial distribution of seismic events, which gives some reasoning for the observed ordering of events. The exponential strain model (EST), was selected to be the preferred model because of its parsimony and slightly better performance in the likelihood testing. From the relative performance of these models, Bourne and Oates (2017a) conclude that there is strong evidence for including the inhomogeneous reservoir properties when

modelling seismicity. Further, it was suggested that additions to the model might include stress-transfer within the reservoir, the possibility of self-excitation of the point process and covariate dependent event magnitudes. Later model developments (Bourne and Oates, 2017c; Bourne et al., 2018; Bourne and Oates, 2020) investigate these additional features using the exponential trend intensity model, or one of similar derivation, for mainshocks.

The effects of long-term gas extraction or cessation of extraction are the same under each of these models. Reducing or stopping production would reduce the rate of earthquakes, but not stop them entirely, because pore pressure would continue to change across the reservoir as spatial pressure gradients equalise. The effect of long-term gas production is that fault failures will begin on faults that are in the body of the initial stress distribution, rather than the tail. The failure trend would then be expected to fall from exponential to linear (Bourne and Oates, 2017a) and so the model would over-predict seismicity. This is not necessarily negative, but could lead to the implementation of overly conservative production plans.

3.4 Recent work on Groningen seismicity

We conclude this chapter by giving a brief overview of the breadth of current research topics that focus on the Groningen gas field and the earthquakes that occur there. The induced earthquakes in the Groningen gas field have received much and varied attention by the statistical seismology community in recent years. The reasons for this are at least threefold; the detection of earthquakes in this region is second to none, providing a world-leading earthquake catalogue in terms of completeness; the high quality covariate information on gas extraction presents new opportunities that are not present when modelling tectonic earthquakes; and finally, the possibility for human intervention in the earthquake generating process means that work in this area

has a potentially huge impact.

There is continuing research to further improve the completeness of the earthquake catalogue in the Groningen region as well as the quality of the covariate information, which describes how the gas field is changing due to gas extraction. This has led to active research in signal processing to better detect earthquakes from the background vibrations of the Earth's surface (Paap et al., 2020; Waheed et al., 2020) and in remote sensing to better measure the compaction of the gas field (Hol et al., 2018; Hadi Mehranpour et al., 2020).

Physics-based statistical models continue to be a popular approach to modelling the times, locations and magnitudes of induced earthquakes (Dempsey and Suckale, 2017; Richter et al., 2020; Smith et al., 2020). The application of machine learning techniques is a novel approach to this same task, which is taken by Limbeck et al. (2021).

Finally, a major focus of research regarding the Groningen gas field is to characterise the earthquake magnitudes in the region. This is particularly important in assessing the risk posed to buildings overlying the gas field. Attention is often given to estimating the largest possible earthquake within the region or the largest earthquake expected during a given time interval, with a workshop having been dedicated to addressing these challenges (Zöller and Holschneider, 2016). A recent treatment of this problem is given by Beirlant et al. (2019) who compare a range of approaches. Estimating these largest magnitudes links closely with the extreme value techniques introduced earlier in this chapter; Shcherbakov et al. (2019) uses associated methods to address this same problem in the tectonic setting.

Chapter 4

Covariate-based models for induced earthquake locations

4.1 Introduction

4.1.1 Induced earthquakes

Tectonic vs Induced Seismicity Catalogues of earthquakes induced by human activity differ in several important ways from those caused by the motion of the Earth’s tectonic plates. Induced earthquake catalogues are typically composed of fewer, smaller earthquakes that occur closer to the Earth’s surface. Their proximity to the surface means that they pose a hazard to infrastructure despite their relatively small magnitudes because their effect is dispersed over a smaller spatial extent. Appropriate modelling of earthquake occurrences and magnitudes is foundational to the appropriate protection of infrastructure against seismic hazards. When modelling either type of seismicity it is common to assume that earthquake counts and locations may be modelled separately from their magnitudes (Zhuang et al., 2011). Here, we

focus on models for the locations and counts of induced earthquakes.

In the tectonic setting, the earthquake generating process is often modelled as having reached steady-state. Fluctuations from this are then attributed to aftershock activity: earthquakes that are triggered by slip rather than drift of tectonic plates. The combination of large catalogue sizes and the assumption of temporal stationarity facilitate the use of highly flexible semi-parametric occurrence models for tectonic earthquakes (Kolev and Ross, 2020; Molkenthin et al., 2020). When modelling induced earthquakes, the steady-state assumption is rarely appropriate because the human activity that causes earthquakes changes over time. Additionally, the small number of earthquakes available in catalogues of induced seismicity makes the use of such highly flexible models particularly challenging.

Despite the above challenges, specific opportunities exist that are accessible only in the context of induced earthquakes. Notably, the seismic process is driven by human actions, and so if these have been sufficiently monitored then there is the potential to include these actions as covariates within an earthquake model. The small size of the earthquake catalogue necessitates a structured modelling approach. This additional model structure can be beneficial as it can allow us to focus on interpretable model forms that can increase our understanding of the earthquake triggering process, relative to a purely descriptive modelling approach. Increased understanding is important here, because it can potentially support informed intervention into the human activities that are driving the earthquake activity.

4.1.2 Motivation and aims

This research focuses on earthquakes that are induced by extraction of natural gas from the Groningen gas field. This gas field lies approximately 3 km below a densely populated region in the north-east of the Netherlands that does not experience any

other form of seismic activity (NAM, 2016a; van Thienen-Visser et al., 2016). As a result, the structures overlying the gas field were not designed to withstand such events. Understanding the link between gas extraction and the locations of induced earthquakes is therefore important for at least two reasons. Firstly, it can help to inform which areas should be prioritised when retrofitting buildings to withstand induced earthquakes. Secondly, it allows investigation of whether the number or location of earthquakes could be influenced by following different gas extraction scenarios.

The ground and structures above the gas field are collectively known as the overburden. The overburden is supported by the reservoir, which is comprised of porous rock where the pore space is filled with natural gas. The overburden is supported by both the structural integrity of the porous rock and also by the pressure exerted by the gas within the pore space. Extracting gas reduces the pressure in the pore space and increases the load on preexisting faults within the reservoir structure. When the shear force on these faults becomes sufficiently large to overcome static friction the fault will slip, releasing the potential energy as an induced earthquake (van Thienen-Visser and Breunese, 2015; Bourne and Oates, 2017c).

Detailed information on gas extraction from the Groningen field is available along with other key reservoir properties (Bourne and Oates, 2017b). Such covariate information is rarely available in the tectonic setting and is central to the modelling approach for small catalogues of induced earthquakes. Rather than considering very flexible model forms, parsimonious parametric models can be constructed based on the physical process that is causing earthquakes. This model structure supports limited earthquake data within a physically motivated framework, which can be adapted to answer questions of interest about the earthquake generating process.

Bourne and Oates (2017a) developed such a physically-motivated point process model for earthquakes within the Groningen gas field. In this chapter, we take this as our

baseline model with two primary aims. Firstly, we assess the parsimony of the baseline model. We do so by considering in turn each component in the baseline point process intensity function and exploring its relative importance to model fit. Secondly, we incorporate a range of additional physical and mathematical features into the baseline model, for example a lagged effect of gas extraction or spatially varying model parameters. We investigate whether incorporating these additional features of the earthquake generating process into the baseline model leads to significant improvements in model fit. This modelling addresses questions of practical and scientific interest. Specifically, we assess whether there is sufficient evidence to suggest answer the following questions:

- Is there spatial variation in the parameters of the baseline model?
- Does the level of smoothing applied to the gas extraction activities limit the baseline model?
- Is there a temporal lag or spatial displacement between gas extraction and induced earthquake activities?
- Does the gas extraction rate influence the resulting induced earthquake count?

Full descriptions are given in Chapter 2 for the available covariates on gas extraction when answering these questions.

4.1.3 Outline

The rest of this chapter is organised as follows: Section 4.2 introduces covariate-based point process models and describes the baseline model; Section 4.3 gives descriptions of model simplifications and extensions that will be investigated; Section 4.4 gives the results and a discussion of fitting these models; Section 4.5 summarises our findings and proposes potentially fruitful areas of further research.

4.2 Background

4.2.1 Point process models for earthquakes

A point process is a stochastic process that provides a statistical model for the count and locations of localised events within a fixed region. As such, point processes are frequently used as a stochastic model for earthquake epicentres. Poisson processes are one of the simplest, most well studied and most widely used point process models. A Poisson process, defined on a region A , may be defined completely by its intensity function $\lambda(x) : A \rightarrow [0, \infty)$. The number of point events in the process and their locations on A are random. The number of events on A is denoted by $N(A)$ and follows a Poisson distribution where the expected event count is equal to the intensity function integrated over A :

$$N(A) \sim \text{Poisson}(\Lambda(A)) \quad \text{where} \quad \Lambda(A) = \int_{x \in A} \lambda(x) dx.$$

Each event in a Poisson process is located independently of all other events and of the event count. The locations of events within a particular realisation of the point process are distributed over A in proportion to the intensity function λ . Throughout, $\mathbf{X} = \{X_i : i = 1, \dots, N\}$ denotes the event locations in a stochastic point process, where $X_i \in A$ and $N \in \mathbb{R}_0^+$. The observed point pattern on A , which is a particular realisation of the point process, will be denoted by $\mathbf{x} = \{x_1, \dots, x_n\}$.

Point process models can be fitted to an observed point pattern using standard methods from either frequentist or Bayesian inference. In either case, it is usual to select a flexible parametric or semi-parametric model for the intensity function and to estimate the vector of parameter values θ for $\lambda(x; \theta)$ under the chosen framework. In a similar manner to generalised linear modelling, covariate effects may be incorporated into the intensity function; an equivalent of the link function may be used to ensure that the non-negativity condition on λ is not violated by the inclusion of covariates.

Letting \mathbf{z} denote one or more covariates that are measured on A , the intensity function may then be specified as $\lambda(x; \theta, \mathbf{z})$. The corresponding integrated intensity function on the entire region is given by

$$\Lambda(A; \theta, \mathbf{z}) = \int_{x \in A} \lambda(x; \theta, \mathbf{z}) dx.$$

The likelihood function for a Poisson process with intensity function $\lambda(x; \theta, \mathbf{z})$ and integrated intensity function $\Lambda(A; \theta, \mathbf{z})$ is then given by

$$\begin{aligned} L(\theta; \mathbf{x}, \mathbf{z}) &= \Pr(N(A) = n \cap \mathbf{X} = \mathbf{x} | \theta, \mathbf{z}) \\ &= \Pr(N = n | \theta, \mathbf{z}) \prod_{i=1}^n \Pr(X_i = x_i | \theta, \mathbf{z}) \\ &= \frac{\Lambda^n(A; \theta, \mathbf{z}) \exp\{-\Lambda(A; \theta, \mathbf{z})\}}{n!} \prod_{i=1}^n \frac{\lambda(x_i; \theta, \mathbf{z})}{\Lambda(A; \theta, \mathbf{z})}. \end{aligned}$$

The log-likelihood of the Poisson process is therefore given by

$$\ell(\theta; \mathbf{x}, \mathbf{z}) = -\log(n!) - \Lambda(A; \theta, \mathbf{z}) + \sum_{i=1}^n \log \lambda(x_i; \theta, \mathbf{z}).$$

For particularly simple choices of the intensity function it is possible to find the maximum likelihood estimator for θ analytically or to specify a conjugate prior distribution for θ . For most parametric forms for λ that are useful in practice, numerical optimisation routines or Markov chain Monte Carlo methods must be used to estimate the parameter values.

Specifying a Poisson process thorough its intensity function allows the same modelling framework to be used for point patterns observed on regions A of one, two or many dimensions. In the context of the Groningen earthquakes, the region A is the spatial extent of the gas field, $W \subset \mathbb{R}^2$, over the time interval $(0, t_{\max})$, so that $A = W \times (0, t_{\max})$. Earthquakes are then described as point events (x, t) in three dimensions, where x represents the planar location and t represents occurrence time.

4.2.2 Baseline intensity model

We use the physically motivated Poisson process model developed in Bourne and Oates (2017a) as our baseline model. This model was derived based on the physics of the gas extraction process and a thin-sheet approximation of the reservoir structure to describe vertical reservoir compaction due to gas extraction. Based on weak assumptions about the initial distribution of stresses on pre-existing faults within the reservoir, a covariate-driven intensity model was derived based on the additional stresses applied to these faults due to gas extraction.

The baseline model was constructed based on the earlier observation in Bourne et al. (2014) that the earthquake count (per unit reduction in reservoir volume) increases exponentially with cumulative reservoir compaction. The covariates \mathbf{z} in the resulting intensity function are $c(x, t)$, the cumulative compaction until time t at location x , and $\dot{c}(x, t) = \frac{d}{dt}c(x, t)$ the instantaneous compaction at time t and location x . Consider a small spatial extent around a point $x \in W$ in which reservoir compaction is approximately constant and denote this region by $B(x, \delta) = \{x' \in W : \|x - x'\|_2 < \delta\}$ for $\delta > 0$. By letting $|B(x, \delta)|$ be the spatial area of $B(x, \delta)$, the observed relationship between compaction and earthquake counts leads to an integrated intensity function of the form:

$$\Lambda(B(x, \delta) \times (0, t); \boldsymbol{\beta}, \mathbf{z}) = \beta_0 |B(x, \delta)| c(x, t) \exp\{\beta_1 c(x, t)\}, \quad (4.2.1)$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1)$ is a vector of model parameters to be estimated. To ensure non-negativity of the corresponding intensity function, it is required that $\beta_0 \geq 0$ and $\beta_1 \geq -\max_A c(x, t)$. It is useful to consider the integrated intensity function on regions of this form because covariate values are given annually on a spatial grid with fine spatial resolution. The integrated intensity over $W \times (0, t)$ can then be well approximated by a temporal interpolation of the cumulative covariate value in

each pixel to time t and then summing expressions of the form (4.2.1) over all pixels. The integrated intensity function (4.2.1) corresponds to an intensity function of the form:

$$\lambda(x, t; \boldsymbol{\beta}, \mathbf{z}) = \beta_0 \dot{c}(x, t) [1 + \beta_1 c(x, t)] \exp\{\beta_1 c(x, t)\}. \quad (4.2.2)$$

The later modelling approach Bourne and Oates (2017a) uses an intensity function of the same form but using strain thickness, the product of reservoir compaction and the topographic gradient of the reservoir surface, in place of compaction. In their approach, the strain thickness covariate is spatially smoothed using an isotropic Gaussian kernel where the bandwidth is chosen to optimise model performance.

Work done by Shell subsequent to the initial development of this material has resulted in further changes to the covariates used during model fitting (Bourne et al., 2018). There has been a shift to using the smoothed incremental Coulomb stress (ICS) in place of smoothed strain thickness. The ICS is the product of compaction, topographic gradient and a spatially variable reservoir property (a poroelastic modulus) that describes the proportion of reservoir compressibility attributable to each of the reservoir rock structure and remaining gas pressure. When constructing the incremental Coulomb stress covariate, three properties are selected to optimise model performance: the largest fault offset, as a proportion of reservoir thickness, on which induced earthquakes can occur (termed the maximum fault throw); the value of the poroelastic modulus; and the bandwidth of the spatial smoothing kernel. In the most recent implementation, uncertainty in these values is reflected within a Bayesian inference framework. This is a thorough but computationally expensive approach that is not suitable for the exploratory nature of the work presented in this chapter; namely to investigate a wide range of physical processes which might improve the model formulation.

In this research we address the above developments by select the smoothing length scale σ applied to gas extraction covariates as part of our inference procedures. To emphasise this, we denote (for example) the cumulative ICS at location $x \in W$ and time $t \in (0, t_{\max})$ by $s(x, t; \sigma)$. The model extensions that we consider here are already extensive in number and complexity, and so to limit the computational intensity of this exploratory work we use the incremental Coulomb stress for fixed values of the maximum fault throw and poroelastic modulus.

4.3 Alternative models

4.3.1 Approach outline

We have two aims when developing alternative forms for the intensity function λ : firstly to determine which mathematical components of the physically motivated intensity model (4.2.2) are most important to model fit, and secondly to investigate whether there are physical features that could be included to improve the intensity model. This section is therefore divided into an investigation of sub-models to address the former question and model extensions to address the latter. Here, we discuss model motivations and formulations while the results of fitting these models to the Groningen earthquake catalogue are presented in Section 4.4.

Table 4.3.1 gives the intensity function corresponding to each model, while Table A.1.1 in Appendix A.1 gives the corresponding integrated intensity function for an individual spatial pixel. The baseline model introduced in Section 4.2.2 is described by model B0 in each of these tables.

4.3.2 Model simplifications

A Poisson process model with intensity of the form (4.2.2) was proposed by (Bourne and Oates, 2017a) for earthquake locations in the Groningen gas field. Up to proportionality, this intensity function has three terms: the instantaneous covariate rate, a linear term in the cumulative covariate and an exponential term in the cumulative covariate. Since the model was derived based on physical considerations, we aim to assess the importance of each of these terms to the overall model fit. To do this we consider four sub-models.

The first sub-model that we consider, S1 in Table 4.3.1, has an intensity proportional to the smoothed topographic gradient of the upper surface of the gas field, $g(x)$. Since this covariate does not change in time, the model does not allow for any temporal variation in earthquake locations and counts. This is a deliberately over-simplified model that we do not expect to perform well. The purpose of including this model is to provide a comparison for the second sub-model, S2 in Table 4.3.1. This is again a very simple model, where earthquake intensity is proportional to the ICS. This is likely a more meaningful model as it allows the changes in gas extraction over time to be represented. Models S1 and S2 represent the most parsimonious inclusion of covariates within the point process intensity.

We also consider the two sub-models, S3 and S4, formed by respective elimination of the exponential and linear terms from the intensity function. This is motivated by the the linear term providing a first order approximation to the exponential term for small covariate values. If this approximation is good then this may allow a more parsimonious representation of the intensity function. Otherwise, these models provide a means to assess the relative contributions of these two terms to the overall model fit.

Evaluation and comparison of these models is somewhat complicated by the small number of observed events. Additionally, these sub-models are not formally nested within the baseline model B0. This means that likelihood ratio tests are not appropriate. For these reasons we do not split the earthquake catalogue into training and evaluation sets. Instead, we use the AIC as a metric to reward goodness-of-fit while penalising model complexity. Other metrics or information criteria such as the BIC might also have been considered. In this exploratory work it is preferable to identify potentially promising model forms to then be put forward for more thorough examination, and so the less conservative AIC metric is used for model comparison (Pawitan, 2001). In subsequent discussions comparing model performance, we deem a better fitting model to be one with a lower AIC value.

4.3.3 Model extensions

In addition to understanding the most important components of the baseline model, we want to investigate whether the inclusion of a range of additional physical features into the intensity function might improve model fit.

Spatial variation in model parameters. The first feature we investigate is whether there is spatial variability in the model coefficients; that is we wish to investigate whether the effect of gas extraction on induced seismicity is the same throughout the gas field. There are two spatial modes of earthquake activity within the gas field, as shown in Figure 4.3.1. A question of interest is therefore whether the seismicity in these two regions is better described by a unified model or two distinct models. As an exploratory approach to this problem we use a simple linear partitioning of the gas field into lower and upper regions, each containing one mode of earthquake activity. We allow distinct coefficients in the parametric intensity function between regions but maintain a shared smoothing scale across the entire gas field.

Model	$\lambda(x, t \boldsymbol{\beta}, \sigma)$
S1	$\beta_0 g(x; \sigma)$
S2	$\beta_0 \dot{s}(x, t; \sigma)$
S3	$\beta_0 \dot{s}(x, t; \sigma)[1 + \beta_1 s(x, t; \sigma)]$
S4	$\beta_0 \dot{s}(x, t; \sigma) \exp\{\beta_1 s(x, t; \sigma)\}$
B0	$\beta_0 \dot{s}(x, t; \sigma)[1 + \beta_1 s(x, t; \sigma)] \exp\{\beta_1 s(x, t; \sigma)\}$
E1	$(\beta_0 \mathbb{I}_{x \in W_L} + \beta_1 \mathbb{I}_{x \in W_U}) \dot{s}(x, t; \sigma)$
E2	$\beta_0 \dot{s}(x, t; \sigma)[1 + \beta_1 s(x, t; \sigma)] \exp\{\beta_1 s(x, t; \sigma)\} \mathbb{I}_{x \in W_L} +$ $\beta_2 \dot{s}(x, t; \sigma)[1 + \beta_3 s(x, t; \sigma)] \exp\{\beta_3 s(x, t; \sigma)\} \mathbb{I}_{x \in W_U}$
E3	$\beta_0 \dot{s}(x, t; \sigma_1)[1 + \beta_1 s(x, t; \sigma_2)] \exp\{\beta_1 s(x, t; \sigma_2)\}$
E4	$\beta_0 [\beta_2 \dot{s}(x, t; \sigma) + (1 - \beta_2) \dot{s}(x, t - 1; \sigma)] [1 + \beta_1 s(x, t; \sigma)] \exp\{\beta_1 s(x, t; \sigma)\}$
E5	$\beta_0 \dot{s}(x, t; \sigma) s(x, t; \sigma)^{\alpha-1} [\alpha + \beta_1 \gamma s(x, t; \sigma)^\gamma] \exp\{\beta_1 s(x, t; \sigma)^\gamma\}$
E6	$\beta_0 [\dot{s}(x, t; \sigma)(1 + \beta_1 s(x, t; \sigma)) + \beta_2 s(x, t; \sigma) \ddot{s}(x, t; \sigma)] \times$ $\exp\{\beta_1 s(x, t; \sigma) + \beta_2 \dot{s}(x, t; \sigma)\}$

Table 4.3.1: Intensity functions for sub-models (S1-S4), the baseline model (B0) and model extensions (E1 - E6). The topographic gradient is denoted by $g(x)$, while $s(x, t; \sigma)$ denotes the cumulative incremental Coulomb stress smoothed using an isotropic Gaussian kernel with standard deviation σ . The first and second temporal derivatives of cumulative ICS are given by $\dot{s}(x, t; \sigma)$ and $\ddot{s}(x, t; \sigma)$. Regions W_L and W_U for models E1 and E2 are defined in Section 4.3.3.

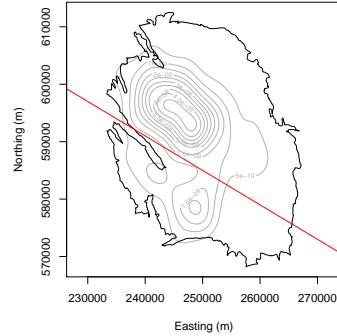


Figure 4.3.1: Field outline with superposed Gaussian kernel density estimate of spatial earthquake distribution. The red line shows the line $y = 735000 - 0.6x$, which is used to separate lower and upper modes of earthquake activity.

Formally, we partition the gas field into lower and upper regions, $W = W_L \cup W_U$, where $W_U = \{(x, y) = w \in W : y > ax + b\}$ and $W_L = W \setminus W_U$ for $a = -0.6$ and $b = 735000$. Alternative partitioning methods are of course possible, but we limit ourselves here to testing sensitivity to the choices of a and b which separate the spatial modes of earthquake activity. This restriction to the boundary forms can be motivated by the principle of parsimony; more complex boundary forms are no less arbitrary than a linear division unless they have a sound physical motivation, for example using sealing faults across which gas pressure gradients can not equalise. Since this type of fault information is not available, we proceed with the simplest form of boundary. We investigate the advantages of this approach for the baseline model and also for the intensity model that is proportional to ICS, which are respectively named models E1 and E2 in Table 4.3.1.

Rate smoothing. The second type of model extension we investigate is related to the smoothing of the rate and cumulative ICS covariates. It has been shown in the context of kernel density estimation that the optimal kernel bandwidths for estimation of a function and its derivative are not the same (Ramsay, 2006). This suggests that

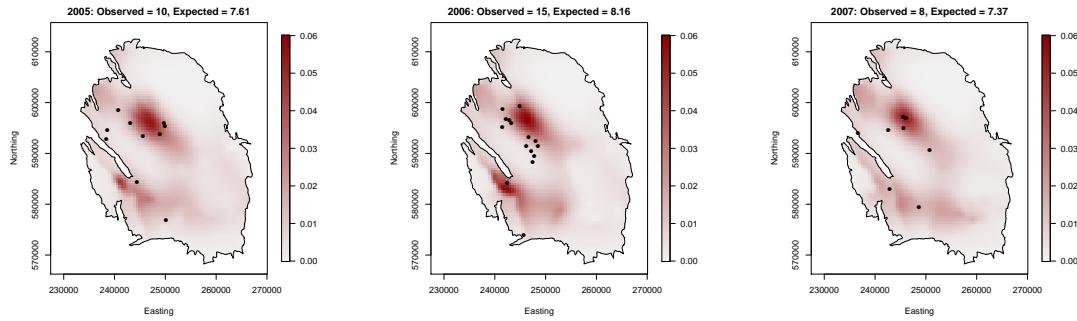


Figure 4.3.2: Maps of expected event counts under the fitted baseline model for years 2005-2007. Observed event locations are overlaid.

for an intensity function that includes both a covariate and its cumulative value, using a single smoothing level might be sub-optimal in both cases. We therefore investigate whether selecting separate smoothing parameters for ICS and cumulative ICS can provide a better representation of these covariates and improve model performance. This is represented by model E3 in Table 4.3.1.

Lag and displacement effects. The next proposed model extensions arise from observed discrepancies between the fitted baseline model and the observed earthquake locations. The number and location of events observed in each year are shown against the fitted baseline model in Figures A.2.1 and A.2.2 of Appendix A.2. The observed events appear to be displaced relative to peaks in the fitted intensity. This is highlighted in Figure 4.3.2 which focuses on the years 2005-2007. Spatially, events appear to surround peaks in intensity rather than occurring at the apex. The years 2005 and 2006 demonstrate this, particularly around the northern mode of earthquake activity. There also appears to be several observed events in regions with low expected counts that were higher in the previous time period. This potential lagged effect can be observed for events in the south-west of the field in 2007.

The above observations may correspond to two physical phenomena that may not have

been properly represented in the model. Firstly, there may be some delay between gas extraction causing an increase in ICS and this presenting as induced seismicity. This has important implications, particularly when trying to establish whether changes in gas extraction methods are effective in influencing the number or location of induced earthquakes. If there is a time lag then it will take longer for these changes to become apparent. Secondly, earthquakes may be triggered not at peak values of ICS but where the spatial gradient of this covariate is large. This might suggest that induced earthquake activity could be decreased by extracting gas more equally across wells to reduce spatial stress gradients.

To investigate the evidence for a lagged effect, we replace the current incremental Coulomb stress in the intensity function by a weighted combination of the current value and the value one year previously. The relative weight given to the current and previous covariate values is determined as a part of model fitting. This is model E4 in Table 4.3.1.

To investigate the clustering of events around (but not directly upon) the peaks of the fitted intensity, we fit a model in which incremental Coulomb stress values are raised to a fractional power. This has the effect of flattening peaks in the intensity function so that more intensity is allocated to areas surrounding the peak. Since it is not clear whether this should be applied to the ICS, cumulative ICS or both, we use a model form which allows for each of these possibilities. Model E6 of Table 4.3.1 considers a fractional power transformation applied to both the instantaneous and cumulative covariates, where the same or different exponents may be applied to each covariate.

Effect of extraction rate. Under the baseline model the rate at which gas is extracted does not influence the total number of earthquakes triggered; this is influenced

only by the total amount of gas extracted. This is known as a *film-rate effect*: it is the total amount of gas extracted and not the time frame over which extraction occurs that determines the total amount of induced seismic activity. In contrast, we consider a non-film-rate effect model, where the total number of induced events depends on the rate at which gas is extracted. If there is evidence for such a non-film-rate model, this clearly has important repercussions when deciding between different gas extraction plans; it determines whether scenarios with greater extraction rates should be viewed negatively. To incorporate this into the baseline model we include both the cumulative and instantaneous ICS within the exponential term of the integrated intensity function. This model is described as E9 in Table 4.3.1.

Summary. By fitting the proposed sub-models we establish which terms within the baseline intensity function are most important to model fit. This may allow a more parsimonious representation of the baseline model and will certainly increase understanding of this model.

The proposed model extensions aim to identify or exclude future areas of model development. This is done through the use of simple statistical tests to establish the merit of including additional physical features within the baseline intensity model. The physical features that we investigate include spatial variation in model parameters, a time-lag between gas extraction and induced seismicity, displacement of induced earthquakes from the area of greatest stress and the extraction rate impacting the total number of induced earthquakes.

4.4 Results

4.4.1 Outline

For each model simplification or extension the resulting AIC value relative to the baseline model B0 is given in Table 4.4.1 along with the maximum likelihood estimate and standard error of the associated smoothing parameter, σ . Interpretation and discussion of the results are organised as follows: Section 4.4.2 addresses the choice of smoothing parameter, σ ; Section 4.4.3 addresses model simplifications; and Section 4.4.4 addresses model extensions.

4.4.2 Covariate smoothing

Our first finding based on Table 4.4.1 is that the estimated smoothing parameter does not differ significantly across models. It should be noted that for all models the standard error of the estimated smoothing scale σ is large and the point estimate is close to the lower boundary of the parameter space, where $\sigma = 0$. Under none of the models does the smoothing of the covariate produce a significant improvement in model fit when tested at the 5% level. The standard errors of the estimated smoothing parameter are included in the table to illustrate this point, but should be interpreted with caution since the sample size on which they are based is small and the point estimate to which they pertain is close to the boundary of the parameter space. This finding suggests that it is not necessary to separately optimise the smoothing scale for each model and that, for the range of models considered, optimising the smoothing scale does not significantly improve model fit.

Model	ΔAIC	$\hat{\beta}$ [std error]	$\hat{\sigma}$ [std error]
S1	160.30	$0.33 [2.3 \times 10^{-2}]$	49 [750]
S2	119.87	$9.6 \times 10^{-2} [6.6 \times 10^{-3}]$	252 [213]
S3	76.98	$(7.0 \times 10^{-4}, 21.8) [2.0 \times 10^{-6}, 1.5]$	228 [157]
S4	-0.82	$(4.0 \times 10^{-3}, 0.45) [7.3 \times 10^{-4}, 2.0 \times 10^{-2}]$	407 [250]
B0	0.00	$(2.4 \times 10^{-3}, 0.36) [2.0 \times 10^{-4}, 1.2 \times 10^{-2}]$	484 [273]
E1	94.15	$(5.3 \times 10^{-2}, 0.12) [8 \times 10^{-3}, 9.5 \times 10^{-3}]$	251 [216]
E2	-27.83	$(2.8 \times 10^{-3}, 2.7 \times 10^{-1}, 2.6 \times 10^{-3}, 3.8 \times 10^{-1})$	511 [401]
—	—	$[6.5 \times 10^{-4}, 2.8 \times 10^{-2}, 2.7 \times 10^{-4}, 1.4 \times 10^{-2}]$	—
E3	9.93 (± 14)	$(3.1 \times 10^{-3}, 0.33)$	(487, 484)
—	—	$[3.7 \times 10^{-4}, 1.4 \times 10^{-2}]$	[674, 223]
E4	5.86 (± 12)	$(3.3 \times 10^{-3}, 0.33, 0.59)$	489 [210]
—	—	$[4.3 \times 10^{-4}, 1.5 \times 10^{-2}, \text{NA}]$	—
E5	3.88	$(2.4 \times 10^{-3}, 0.36, 0.95)$	494 [271]
—	—	$[2.0 \times 10^{-4}, 1.1 \times 10^{-2}, 1.3]$	—
E6	1.58	$(2.5 \times 10^{-3}, 0.36, 1.1 \times 10^{-5})$	126 [320]
—	—	$[1.3 \times 10^{-4}, 7.4 \times 10^{-3}, 2.0 \times 10^{-6}]$	—

Table 4.4.1: Summaries of fitted sub-models (S1-S4), baseline model (B0) and model extensions (E1 - E6). The second and third columns give the numerically maximised log-likelihood value and the change in AIC relative to the baseline model. The fourth and fifth columns give point estimates for the model parameters and covariate smoothing scales; approximate standard errors are given in square brackets.

4.4.3 Model reductions

When considering model reductions, the sub-models S1-S3 each resulted in a worse model fit relative to the baseline model. This is to be expected, particularly for models S1 and S2, which were deliberate oversimplifications. Within the sequence of sub-models S1-S3 we have increasing goodness of fit. The improvement of S2 over S1 demonstrates what might reasonably be expected: that including information on gas extraction over time leads to a better model of induced seismicity than using only the topographic gradient, which is a reservoir property that does not change over time. The improvement when moving from S2 to S3 then demonstrates that the number and location of induced earthquakes is not influenced only by gas extraction (through the ICS) but also depends on the cumulative amount of gas that has been extracted. Finally, the lower AIC value of model S4 over S3 shows that this dependence on the cumulative ICS is better described by an exponentially increasing trend than by a linear approximation to this trend.

It should be noted that model S4, which removes the linear term from the baseline intensity function, provides a slight improvement in model fit over the baseline model B0. Model S4 improves parsimony in terms of the expression for the intensity function, but leads to more complicated interpretation of the model parameters. This is because under S4 both β_0 and β_1 scale the integrated intensity function, as can be seen in Table A.1.1 of Appendix A.1. Since the improvement in model fit by S4 is modest, we choose to develop our model extensions on the baseline model B0 and prioritise the desire for ease of interpretation and physical derivation over the desire to use strictly the most parsimonious model.

4.4.4 Model extensions

The majority of the proposed model extensions did not provide a significant improvement to model fit. In particular, it should be noted that the models E3 and E4 had maximised log-likelihood values that were less than that of the baseline model B0, which they extend; this is counter-intuitive but is explained as follows. These models respectively extend the baseline model by including separate smoothing parameters or a lag between gas extraction and induced seismicity. The decrease in the maximum log-likelihood value for these models can be attributed to the integrated intensity function for these models not having a closed form. This necessitates the use of numerical integration to calculate the integrated intensity term of their log-likelihoods. From the available information, the cumulative ICS value is known for each spatial pixel on the first of January each year. Since the evolution of ICS within each year is not known, the mean of the initial and final values is used to construct an approximation of the integrated intensity. Bounds on the likelihood value can be obtained by using the initial or final value of ICS within each year, these bounds correspond to a step change in cumulative ICS at the end or beginning of each year. The AIC values for models E3 and E4 are less than or greater than that of the baseline model B0, depending on which approximation to the integral is used. There is therefore not enough information available using these covariate grids to definitively establish whether including separate smoothing parameters or a lag between gas extraction improve the model of induced seismicity. However, when using the numerical integration scheme under which the AIC is reduced, the size of the reduction is modest. This suggests that these features do not influence strongly the induced seismic activity.

Model E5 considered raising either the cumulative or instantaneous ICS value to a fractional power, in order to address the observation that earthquakes cluster around peaks in the fitted baseline intensity, rather than occurring at the peaks. This model

did not reduce the AIC relative to the baseline model. This is perhaps because such flattening of peaks in the intensity function is also achievable through increasing the length scale of the smoothing kernel. This could be further explored by investigating models which include the spatial gradient of the covariates, $\frac{d}{dx}s(x, t; \sigma)$ or $\frac{d}{dx}\dot{s}(x, t; \sigma)$, in some way.

Model E6 investigated whether there was significant evidence that the ICS rate influenced the total number of induced earthquakes. From the available data, there was insufficient evidence to suggest a non-film-rate effect. It should be noted that there were some difficulties when fitting of this model, due to the presence of second derivative of cumulative ICS within the intensity function. This covariate is difficult to obtain accurately due to the coarse temporal resolution of the cumulative ICS data, which might be a limiting factor in assessing this modelling approach. Additionally, the second derivative of ICS is negative in many locations and periods (roughly corresponding to areas or periods where the gas depletion slows down). This presented problems when ensuring non-negativity of the fitted intensity everywhere on A . In principle, this can be achieved through constraints on the coefficient values β , but this caused the numerical optimisation routines used to fit the model to fail. To work around this issue, we approximated the model by taking the pointwise maximum of zero and the intensity term that depends on $\ddot{s}(x, t; \sigma)$ (the term given in square brackets) at each time and location to ensure non-negativity. Then numerical integration was used to calculate the integrated intensity term of the log-likelihood. Alternative methods of ensuring non-negativity were also investigated, where $\ddot{s}(x, t; \sigma)$ was replaced by $|\ddot{s}(x, t; \sigma)|$ or $\max\{\ddot{s}(x, t; \sigma), 0\}$; neither approach led to a significant reduction in AIC.

The models E1 and E2 allow the model parameters to vary spatially across the gas field while maintaining the same model forms as S1 and B0. Both model extensions

led to reduced AIC values relative to their equivalent spatially constant models. In particular, the extension of the baseline model to have separate parameters in the north-east and south-west of the gas field, E2, was the only model extension to provide a marked improvement in model fit over the baseline model. This suggests that the two spatial modes of earthquake activity within the gas field are responding differently to gas extraction. The sensitivity of this finding was tested by considering 4 alternative boundary lines that separated the two modes of intensity. Maps of annual expected earthquake counts under this model are given in Figures A.2.3 and A.2.4 of Appendix A.2.

The improvement of this simple spatial model over the baseline suggests that a more in-depth investigation might be worthwhile, for example, variations of the baseline model with parameters that vary smoothly over space. This will require a careful balance between achieving flexibility in the model form and the ability to fit the model using the limited available data. Our linear partition clearly leans toward the latter consideration but serves as motivation for a model using either a physically motivated reservoir partitioning or smoothly varying parameters as further work.

4.5 Conclusions and further work

The first aim of this work was to establish which sub-components of the baseline intensity function were the most important drivers of model fit. Through the investigation of four sub-models, we found that the terms concerning the ICS and exponential trend in cumulative ICS are the main contributors to model fit. The linear term in cumulative ICS is dominated by the exponential term to the extent that a small improvement in model fit may be achieved by removing the linear term from the intensity function. However, removing this term complicates the interpretation of model parameters since this leads to both β_0 and β_1 scaling the integrated intensity function.

The second aim of this work was to investigate a range of extensions to the baseline intensity model. The extensions investigated were selected based on physical phenomena that might be missing from the model, or mathematical transformations motivated by potential shortcomings of the baseline model. From the data used, there is insufficient evidence to conclude that there is a lagged effect of ICS on induced seismicity or that this can be better described by using some fractional power of ICS. There is also insufficient evidence to conclude that model fit can be significantly improved by using separate smoothing levels, either within or across the models considered.

It does, however, appear that the relationship between induced seismicity and ICS is not constant across the gas field. This finding was tested using a simple linear partitioning of the gas field to separate modes of earthquake activity. The finding was not highly sensitive to the choice of boundary line, however, the use of a linear form for the boundary lacks a physical basis and was chosen for reasons of parsimony. A potentially fruitful area of further work might therefore be to investigate models that rely on a physically motivated partitioning of the gas field (for example using sealing faults across which pressure gradients can not easily equalise), or indeed a model where parameters vary smoothly over space. The challenge in each of these cases will be striking a balance between model complexity and the small number of available data.

We identified that induced events appear to cluster around modes of fitted intensity in the baseline model, but that model fit can not be significantly improved by using a power transformation of the ICS. In further work, this phenomenon could be further investigated by including the spatial gradient of ICS within the intensity or integrated intensity function - although how best to include this covariate is not obvious. Another route to extending the work here would be to use covariates on a finer temporal grid, for example using weekly or monthly values. This would have the benefit of allowing

variations in gas extraction within each year to be represented within the models and might allow first and second temporal derivatives of ICS to be better represented. However, using monthly values comes with an increased computational burden and leads to extensive issues related to ensuring non-negativity of the fitted intensity. For example, seasonal or reduced gas extraction can lead to areas of “pressure-up” as gas pressure equalises across the field resulting in widespread areas where the ICS rate is negative.

One model extension which is not considered in this work is the inclusion of aftershock activity, such as within an epidemic type aftershock sequence model (Ogata, 1988, 2011). While this was considered by Bourne et al. (2018), we were unable to replicate such a model extension without imposing strong constraints on aftershock parameter values. There are a range of issues associated with fitting this type of aftershock model, such as the likelihood function having locally flat regions, being costly to evaluate and difficulty in separating parameter effects (Veen and Schoenberg, 2008; Schoenberg, 2013; Ross, 2016). In this work we have restricted focus to models for independent events but in Chapter 6 of this thesis we go some way to addressing these issues.

Chapter 5

Inference for extreme earthquake magnitudes accounting for a time-varying censoring process

5.1 Introduction

5.1.1 Aims and motivation

The observational nature of environmental data can lead to challenges during statistical modelling and inference. In particular, improved measurement of an environmental process within a dataset should be acknowledged as part of any inference. Failing to do so leads to biased inference, while including data based only on the initial quality of measurements is overly conservative, leads to inefficient inference, and makes financial investment into the measurement process redundant. We consider how to include changing data quality in an extreme value analysis where low data quality is present as the partial censoring of rounded data. Here and throughout, censored data

are values that are missing-not-at-random (Little and Rubin, 2019). This chapter is motivated by the modelling of earthquake catalogues, but results in a method that is applicable more widely where these data features are present. This new threshold selection method should also be of value in more general extreme value analyses.

5.1.2 Earthquake data

Earthquakes are recorded if their locations and magnitudes can be inferred from ground vibrations at sensor locations; this requires an earthquake to be detected by multiple sensors. An earthquake is detected or missed depending on its magnitude and location relative to the sensor network. A low sensitivity network of sensors therefore leads to the partial or complete censoring of small magnitude seismic events. As the network is extended or upgraded over time the censoring of small events is reduced. It is usual in earthquake catalogues for magnitudes to be reported to one decimal place; this data feature is often overlooked during statistical analyses (Marzocchi et al., 2019). Using these rounded, incomplete observations we seek to understand the tail behaviour of the magnitude distribution.

Since 1991 the Groningen region of the Netherlands has experienced induced earthquakes. These seismic events are caused by gas extraction and have relatively small magnitudes compared to tectonic events. However, they also occur at much shallower depths than their tectonic equivalents. This means that for equal magnitudes they pose a greater hazard than their tectonic counterparts because their impact is spread over a smaller spatial extent. These small earthquakes are therefore both hazardous and difficult to detect. This has led to continued investment in the geophone network around the Groningen gas field to increase detection of small earthquakes and to better understand earthquake activity in the region. Estimating high quantiles of the magnitude distribution, and quantifying their uncertainty, is instrumental to appropriate

design and improvement of buildings to withstand these earthquakes.

5.1.3 Magnitude of completion

The magnitude of completion m_c is the lowest magnitude above which all earthquakes are certain to be recorded in a given area and time interval. The magnitude of completion therefore depends on the density and sensitivity of the sensor network as well as the local geology. When a sensor network changes substantially over time t , the magnitude of completion in that region can be considered as a function of time, denoted $m_c(t)$. The magnitude of completion is not a quantity that can be determined experimentally, it must be inferred from the set of observed event magnitudes.

Existing methods for statistical estimation of a constant m_c use parametric or non-parametric methods to detect deviations from the assumed monotonicity of the magnitude distribution (Mignan and Woessner, 2012). Parametric methods typically assume an exponential magnitude distribution, based on the empirical magnitude-frequency relationship of Gutenberg and Richter (1956). Heuristic techniques are used to detect deviations from this model based on maximum curvature, goodness-of-fit, or parameter stability.

Several methods exist to estimate a spatially varying magnitude of completion (Wiemer and Wyss, 2000; Mignan et al., 2011). In contrast, little attention has been given to estimating a changing magnitude of completion over time. Where it has been considered, focus has been on temporary increases in $m_c(t)$ due to residual vibrations following large earthquakes (Woessner and Wiemer, 2005; Utsu et al., 1995). Long-term changes in $m_c(t)$ have been considered by assuming a constant value within a pre-determined temporal partitioning (Hutton et al., 2010) or a locally constant value estimated using a rolling window (Mignan and Woessner, 2012).

5.1.4 Extreme value methods

To specify a model for earthquake magnitudes we adapt a model from extreme value theory. An asymptotic argument justifies the use of the generalised Pareto distribution (GPD) to model the excesses of a continuous random variable over a suitably chosen threshold, under weak assumptions on the distribution of that random variable (Pickands, 1975). The distribution function of a random variable Y that follows a GPD, given that it is above the threshold u , is

$$F(y; \sigma, \xi) = \begin{cases} 1 - [1 + \xi(y - u)/\sigma]_+^{-1/\xi} & \text{for } \xi \neq 0, y \geq u, \\ 1 - \exp[-(y - u)/\sigma] & \text{for } \xi = 0, y \geq u; \end{cases} \quad (5.1.1)$$

where the shape parameter $\xi \in \mathbb{R}$, scale parameter $\sigma > 0$ and $y_+ = \max(0, y)$. The distribution is exponential when $\xi = 0$, heavy-tailed when $\xi > 0$ and decays to a finite upper end point $y^+ = u - \sigma/\xi$ when $\xi < 0$ (Davison and Smith, 1990). The GPD generalises the Gutenberg-Richter model, in which magnitudes are independent and identically distributed (i.i.d.) exponential random variables, by allowing greater flexibility in the tail behaviour of the distribution.

Standard extreme value modelling deals with i.i.d. data, observed at regular intervals without rounding or censoring. The standard approach is to select a constant threshold u that is a fixed, high quantile of the empirical distribution. Heuristic methods are used to select an appropriate quantile, see Scarrott and MacDonald (2012) for a review. These methods can be based on the stability of parameter estimates, goodness-of-fit measures, or the mean threshold exceedance size (Coles, 2001). When interest lies in estimating a particular extreme value property, such as the shape parameter, an alternative strategy is to select the threshold that optimises inference for that property (Danielsson et al., 2001).

Using a constant threshold is inefficient when the data distribution changes over time. This type of change is likely to alter the quantile value above which a GPD is ap-

propriate and cause the GPD parameters to change over time. To avoid this issue, quantiles can be estimated locally as a function of time and a global decision can be made on which quantile to use as a time-varying threshold $u(t)$ (Eastoe and Tawn, 2009; Northrop and Jonathan, 2011).

5.1.5 Shortcomings of current methods

Estimating the magnitude of completion and selecting an extreme value threshold are closely linked problems. Both aim to select a value (possibly time-varying) above which a probability model is appropriate. Standard methods from either setting do not meet our modelling needs, for the reasons that follow.

Methods assuming an exponential magnitude distribution are problematic for two reasons. Firstly, an exponential tail model can lead to bias and false confidence in quantile estimates. Coles and Pericchi (2003) demonstrated in a hydrological context the benefits of using the encompassing generalised Pareto model to properly represent uncertainty in the tail shape. Secondly, the exponential distribution does not account for rounding of the data, resulting in biased parameter estimates (Marzocchi et al., 2019; Rohrbeck et al., 2018). Failing to acknowledge this rounding can therefore also cause bias in threshold selection.

Methods to select a static threshold are also unsuitable for our problem. To obtain precise estimates of the GPD parameters and high quantiles, as much data as possible should be used in the analysis. However, this must be balanced by the need to represent model uncertainty and avoid bias from incorrectly including small magnitude events. This bias has two sources: using either data values for which the extreme value model does not apply or values that are below the magnitude of completion at the time of their occurrence. The optimal choice of time-varying threshold is therefore $v(t) = \max\{m_c(t), u(t)\}$. Methods for selecting a static modelling threshold v are

inefficient when the true threshold varies with time, since the static threshold must satisfy $v \geq \max_t v(t)$ and so excludes viable data from the analysis.

Finally, current approaches to selecting or estimating time-varying thresholds are also unsuitable for our problem; methods for estimating $m_c(t)$ consider only a small portion of the data at once, while the selection of $u(t)$ by a local quantile approach is impeded by the temporal development of the censoring process.

5.1.6 Contributions and outline

In this chapter we develop an automated method to select a dynamic threshold for rounded GPD data. This is, to our knowledge, the first time that data rounding has been considered during threshold selection. Our proposed threshold selection method uses as much data as possible while guarding against the use of values where a tail model is not appropriate or observations are not complete. This threshold choice leads to more precise estimation of high magnitude quantiles, properly represents their uncertainty, and can also suggest how the magnitude of completion changes over time. The selection method is developed for earthquake data, but the core idea of the method can be applied to extreme value threshold selection more generally. We demonstrate, via simulation, the benefits of including additional, small magnitude events in an extreme value analysis to both parameter recovery and return level estimation. We go on to select dynamic thresholds for partially censored earthquake catalogues and investigate the impact of this threshold when estimating high quantiles of the magnitude distribution.

This chapter is structured as follows. Section 5.2 describes the Groningen earthquake catalogue that motivates the proposed methodology, the model for observed magnitudes, and the novel inference for the underlying parameters. Section 5.3 demonstrates the benefits of including small magnitude events into an extreme value analysis. Sec-

tion 5.4 introduces our proposed method of threshold selection. The method is applied to simulated earthquake catalogues in Section 5.5 and to the Groningen catalogue in Section 5.6. Concluding remarks are given in Section 5.7.

5.2 Motivating data and model formulation

5.2.1 Data description

We study the induced earthquakes in the Groningen region of the Netherlands from January 1st 1995 to December 31st 2019. Compared to tectonic earthquakes, these are close to the surface and can cause damage despite their relatively small magnitudes. This has led the Royal Dutch Meteorological Institute (KNMI) to invest heavily in the earthquake detection infrastructure in the Groningen region. Over time, more and better sensors have been added in the region to increase the detection and reporting of small earthquakes. The resulting earthquake catalogue is publicly available and magnitudes are reported in units of local magnitude (M_L) to one decimal place (KNMI, 2020).

Figure 5.2.1 shows Groningen earthquake magnitudes against both occurrence time and earthquake index, along with smoothed estimates of their mean using a generalised additive model with cubic-spline basis. Assuming that magnitudes are i.i.d. (which is supported by the exploratory analysis in Appendix B.1 for Groningen earthquakes exceeding $1.5M_L$) and that departures from this are due to the partial censoring of small magnitude events, the reduction in mean magnitude indicates that fewer small magnitude events were censored at later times. It is unclear whether this change in detection was sudden or gradual. The KNMI report that $m_c(t) \leq 1.5M_L$ for the entire period (Dost et al., 2012). Paleja and Bierman (2016) and Dost et al. (2017) used a fixed temporal partitioning and conclude that for the period 2014-09-

24 to 2016-09-27 the magnitude of completion was likely to be below $1.0M_L$. Since sensors have not been removed from the network, this suggests that the magnitude of completion should be less than or equal to this in the period following their analysis (i.e. to 2020 in Figure 5.2.1).

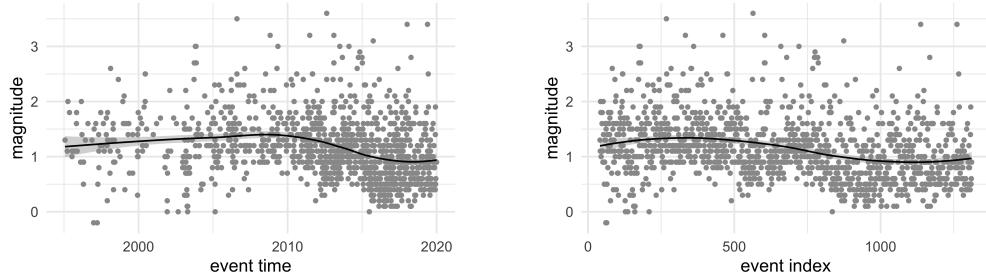


Figure 5.2.1: Full Groningen earthquake catalogue, with magnitudes reported in M_L and smoothed mean estimate; shown using natural- [left] and index-times [right].

5.2.2 Data model and inference

This section introduces our notation and data model for threshold selection and inference on extreme earthquake magnitudes. We define an earthquake catalogue to be the set of n recorded time-magnitude pairs $\{(t_i, x_i) : i = 1, \dots, n\}$ where the recorded magnitudes $\mathbf{x} = (x_1, \dots, x_n)$ are given rounded to the nearest 2δ ($\delta > 0$) and the event times $\mathbf{t} = (t_1, \dots, t_n)$ are each within the observation interval (t_{\min}, t_{\max}) . The unrounded magnitudes associated with each event are represented by the vector $\mathbf{y} = (y_1, \dots, y_n)$. An event (t_i, x_i) therefore corresponds to an earthquake of magnitude $y_i \in (x_i - \delta, x_i + \delta]$ that occurred at time t_i and that was not censored.

Recall from Figure 5.2.1 that earthquake intensity is not constant over the observation period. To separate exposition of our threshold selection method from estimation of this temporally-varying earthquake rate, we map each event time to its corresponding index. This transforms event times \mathbf{t} from an irregular sequence on the natural

timescale t to a regular sequence $\boldsymbol{\tau}$ on the index scale τ , where observed events occur at $\tau = 1, \dots, n$. A modelling threshold $v(\tau)$ is then specified for the transformed observation period $\tau \in (0, \tau_{\max})$ and the threshold values at each event time are given by the vector $\mathbf{v} = (v(1), \dots, v(n)) = (v_1, \dots, v_n)$. The threshold function $v(\tau)$ and threshold vector \mathbf{v} will be initially treated as known until threshold selection is discussed in Section 5.4.

The probability $\alpha(\tau, y)$ that an event is detected by the sensor network and included in the earthquake catalogue is an unknown function of its time and magnitude. It is expected that for the Groningen catalogue $\alpha(\tau, y)$ is a non-decreasing function in each of τ and y ; larger or later earthquakes are more likely to be detected. We make two assumptions on $\alpha(\tau, y)$: firstly that observation is complete above the modelling threshold, so that $\alpha(\tau, y) = 1$ for $y \geq v(\tau)$; secondly that censoring begins gradually so that for all τ , $\alpha(\tau, y) \approx 1$ when $y \in [v(\tau) - \delta, v(\tau)]$. This allows rounded magnitudes within δ of the modelling threshold to be included during inference without constructing a full model for the censoring process.

In constructing our model for magnitudes exceeding $v(\tau)$, we assume that the unrounded magnitudes \mathbf{y} may be modelled as i.i.d. GPD random variables (Y_1, \dots, Y_n) with parameters $\boldsymbol{\theta} = (\sigma_u, \xi)$ when they exceed a constant, lower threshold of $u < \min_\tau v(\tau) - \delta$. Formally, we assume $Y_i - u | Y_i > u \sim \text{GPD}(\sigma_u, \xi)$. Since events exceeding $v(\tau)$ are never censored, excess magnitudes of $v(\tau)$ may also be modelled using a GPD but with threshold dependent scale parameters $\sigma_{v_i} = \sigma_u + \xi(v_i - u)$, so that $Y_i - v_i | Y_i > v_i \sim \text{GPD}(\sigma_{v_i}, \xi)$.

When using this probability model to construct a likelihood function for the GPD parameters, rounded magnitudes x_i should contribute only if the latent value $y_i > v_i$. Events with $x_i > v_i + \delta$, should certainly contribute to the likelihood function and events with $x_i < v_i - \delta$ certainly should not. When $|x_i - v_i| < \delta$ it is uncertain

whether $y_i > v_i$ and whether event i should contribute to the likelihood. Each event is therefore weighted in the log-likelihood by $w_i = \Pr(Y_i > v_i | x_i, \boldsymbol{\theta})$, the probability it truly exceeds $v(\tau)$. This is equivalent to using the expected likelihood over all possible unrounded magnitude vectors. The resulting log-likelihood function for the parameters $\boldsymbol{\theta} = (\sigma_u, \xi)$ of F , the GPD (5.1.1) is:

$$\begin{aligned}\ell(\boldsymbol{\theta} | \mathbf{x}, \mathbf{v}) &= \sum_{i=1}^n w_i \log \Pr(X_i = x_i | Y_i > v_i, \boldsymbol{\theta}) \\ &= \sum_{i=1}^n w_i \log \Pr(\max(v_i, x_i - \delta) < Y_i < x_i + \delta | \boldsymbol{\theta}) \\ &= \sum_{i=1}^n w_i \log [F(x_i + \delta - v_i; \sigma_{v_i}, \xi) - F(\max(v_i, x_i - \delta) - v_i; \sigma_{v_i}, \xi)],\end{aligned}\tag{5.2.1}$$

where

$$\begin{aligned}w_i &= \frac{\Pr(\max(v_i, x_i - \delta) < Y_i < x_i + \delta | \boldsymbol{\theta})}{\Pr(x_i - \delta < Y_i < x_i + \delta | \boldsymbol{\theta})} \\ &= \frac{F(x_i + \delta - u; \sigma_u, \xi) - F(\max(v_i, x_i - \delta) - u; \sigma_u, \xi)}{F(x_i + \delta - u; \sigma_u, \xi) - F(x_i - \delta - u; \sigma_u, \xi)}.\end{aligned}\tag{5.2.2}$$

The maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ can be found using numerical optimisation of this function. Confidence intervals may be obtained based on asymptotic normality, but this approximation can be poor for the estimated shape parameter $\hat{\xi}$ and quantile values. To avoid this and to ensure that confidence bounds on $\hat{\sigma}_u$ are positive, we use a parametric bootstrap approach to describe parameter uncertainty, as described in Appendix B.2.

5.3 Motivating the inclusion of small magnitudes

5.3.1 Simulation study overview

Here we show that using a non-constant threshold to include additional, small magnitude earthquakes in an extreme value analysis can be beneficial to both parameter

and quantile estimation. We compare three approaches to inference on 1000 simulated earthquake catalogues that have a known, stepped threshold. Each catalogue is simulated by first generating 1000 latent magnitudes as independent GPD exceedances of $u = 1.05M_L$ with parameters $\boldsymbol{\theta} = (\sigma_u, \xi) = (0.4, 0.1)$. Each event $i = 1, \dots, 1000$ is censored if $\tau_i \leq 500$ and $y_i < 1.65M_L$. The retained magnitudes are then rounded to the nearest $2\delta = 0.1M_L$, resulting in a catalogue of the form shown in Figure 5.3.1 (left). The size of the retained catalogue depends on the simulated magnitudes, and so varies between catalogues.

A GPD model is fitted to each of the simulated catalogues by maximising the log-likelihood (5.2.1) under each of three approaches. The first, conservative approach to inference uses only exceedances of the flat modelling threshold $v(\tau) = 1.65M_L$ for $0 \leq \tau \leq 1000$. The second approach uses exceedances of the stepped threshold where $v(\tau) = 1.65M_L$ for $0 \leq \tau \leq 500$ and $v(\tau) = 1.05M_L$ for $500 < \tau \leq 1000$. The number of data points used by the stepped approach will be at least as large as by the conservative approach. A third approach, possible in simulation but not practice, is also considered. In this third approach, additional earthquakes are simulated above the conservative level to extend the simulated catalogue until the number of exceedances of $1.65M_L$ matches the number of events used by the stepped approach. A GPD model is then fitted to the extended set of earthquakes that exceed $1.65M_L$.

We compare the three approaches to inference in terms of parameter and quantile estimation. The conclusion of each comparison can differ because of the non-linear relationship between GPD parameters and quantiles, which are also sensitive to small changes in the estimated shape parameter ξ . Parameter estimates are compared using their bias and variance over the 1000 simulated catalogues. To be able to compare quantile estimates across modelling thresholds we consider the conditional quantiles above the conservative threshold level, using conditional return levels. The conditional

p -quantile above some magnitude $c > u$ is the magnitude $y_{p,c}$ that satisfies

$$\Pr(Y \leq y_{p,c} | Y > c) = p.$$

Letting $\zeta_c = \Pr(Y > c | Y > u) = 1 - F(c; \boldsymbol{\theta})$, where F is the distribution function (5.1.1), $y_{p,c}$ can be expressed as a function of $\boldsymbol{\theta}$:

$$y_{p,c}(\boldsymbol{\theta}) = \begin{cases} u + \frac{\sigma_u}{\xi} ((\zeta_c p)^{-\xi} - 1) & \text{for } \xi \neq 0, \\ u + \log(\zeta_c p) & \text{for } \xi = 0. \end{cases} \quad (5.3.1)$$

An alternative representation of conditional quantiles, more in-keeping with the extreme value approach, is the m -event conditional return level above c . This can be found by setting $p = 1 - 1/m$ in equation (5.3.1) and interpreted as the magnitude exceeded (on average) by one in every m events that exceed c . We compare point estimates and confidence intervals of conditional return levels under the three approaches to inference.

5.3.2 Simulation study results

Figure B.4.1 in Appendix B.4 shows the sampling distribution of parameter estimates and an error decomposition under each approach to inference. The stepped threshold is best for parameter estimation, with the smallest bias and variance of the three approaches. The mean squared error of the stepped estimator is 9.6 times smaller than that of the conservative estimator, mainly due to its increased precision. For comparison, artificially extending the earthquake catalogue gives a reduction factor of only 4.2. In this example, each small magnitude event added by lowering the threshold is more than twice as valuable to parameter estimation than an additional observation above the conservative level.

Figure 5.3.1 (right) shows the conditional return levels for magnitudes above $c = 1.65M_L$ under each approach. Point estimates are qualitatively similar in each case,

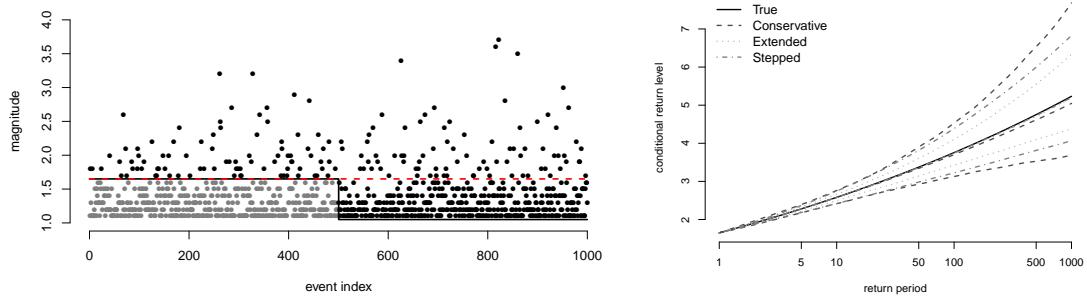


Figure 5.3.1: [Left] Simulated catalogue structure: events are censored (grey dots) if in the first 500 and below $1.65M_L$. For this catalogue, the conservative threshold (dashed red line) includes 181 events, while the stepped threshold (solid black line) includes 582 events. [Right] Magnitude conditional return level estimates in M_L against return period in number of earthquakes exceeding $1.65M_L$. Point estimates and 95% confidence intervals are given under conservative, extended and stepped approaches to inference, along with the true values.

but confidence intervals are narrowed by using the stepped rather than constant threshold. Confidence intervals are further narrowed by artificially extending the observation period. This is because of the additional large values in the extended data, which have a strong influence over the estimated return levels (Davison and Smith, 1990).

These results show clearly the benefits for parameter and quantile inference that can be achieved by using a dynamic modelling threshold to include additional small magnitude events in an extreme value analysis. Using a conservative constant threshold leads to wasteful inference and the squandering of these potential gains.

5.4 Threshold selection

5.4.1 Overview

In practice, the true modelling threshold $v(\tau)$ is always unknown. To choose between potential thresholds, we must define what it means for one threshold to be preferred over another. A generalised likelihood ratio test is not appropriate for this comparison because it compares nested models on the same data, rather than comparing the same model on nested data (Wadsworth and Tawn, 2012; Wadsworth, 2016).

To select a model that is robust to sampling variability, $v(\tau)$ should include as much data as possible in the model and therefore be chosen to be as low as possible. However, selecting $v(\tau) < \max(u(\tau), m_c(\tau))$ for any $0 < \tau < \tau_{\max}$ will cause bias in the fitted model, making it incapable of obtaining an asymptotically consistent estimator of the true parameter values. The best choice of $v(\tau)$ is therefore the threshold that includes the most data while maintaining a good agreement between observed threshold exceedances and the fitted GPD.

For i.i.d. continuous valued data, the distributional agreement with a probability model can be assessed graphically by using a PP- or QQ-plot and adding tolerance intervals to show expected behaviour under that model. Alternatively, the distributional fit can be summarised using a metric, such as the Anderson-Darling or Cramer-von Mises distances (Laio, 2004). Both graphical- and metric-based approaches can be adapted for data \mathbf{y} that are independent and continuous valued, but which do not have a shared distribution. This is achieved by using the probability integral transform and the fitted distribution to transform the data to have a shared marginal distribution before using methods for i.i.d. data to produce plots or metric values (Heffernan and Tawn, 2001).

We further adapt these methods to handle both rounded data and parameter un-

certainty, before showing how they can be used to inform selection of a modelling threshold. In doing so, we transform to standard Exponential margins because this distribution is central within the GPD family and follows the precedent of Heffernan and Tawn (2001). Alternative marginal distributions could be used; we additionally consider PP-plots, which correspond to the special case of uniform margins.

5.4.2 Graphical assessment

The observed magnitudes \mathbf{x} that exceed $v(\tau)$ do not have a shared marginal distribution when $v(\tau)$ is non-constant and they are not continuous-valued due to their rounding. This presents challenges when trying to create a PP- or QQ-plot for exceedances of the modelling threshold $v(\tau)$. Firstly, constructing these plots using rounded values can lead to many probabilities or quantiles of equal value and the plots being difficult to interpret. The second challenge relates to observed, rounded values close to the modelling threshold, $\{x_i : |x_i - v_i| < \delta, i = 1, \dots, n\}$; it is not known which, or how many, of these events satisfy $y_i \geq v_i$ and so should be included when constructing the plot.

To overcome these challenges we use simulation to construct Monte Carlo confidence intervals for the sample quantiles (or probabilities) of the unrounded threshold exceedances transformed onto shared exponential margins. The process is described in Appendix B.3 and leads to a modified plot with two sets of intervals; tolerance intervals show the expected variability of sample quantiles (or probabilities) under the fitted model while confidence intervals show the uncertainty about the observed sample quantile values. Confidence and tolerance intervals that do not overlap suggest that the distribution of the rounded exceedances is not coherent with the fitted GPD model.

Examples of such PP- and QQ-plots are shown in Figure 5.4.1. These use the simu-

lated catalogue shown in Figure 5.3.1 (left) and constant modelling thresholds of $v(\tau) = 1.85M_L$ and $1.15M_L$. For this catalogue, exceedances of a flat threshold should be consistent with a GPD model only if that threshold is of $1.65M_L$ or greater. For the higher threshold $v(\tau) = 1.85M_L$, the confidence intervals on sample probabilities and quantiles overlap with the tolerance intervals, indicating that exceedances of this threshold are consistent with the fitted GPD model. For the lower threshold $v(\tau) = 1.15M_L$ this is not the case, with the large sample quantiles bigger than expected under the fitted model. Notice the shape of the tolerance intervals in Figure 5.4.1; the largest deviations from the line $y = x$ are expected at central probabilities in the PP-plots and at the largest quantiles of the QQ-plots. This is important in Section 5.4.3 where we propose metrics to summarise these plots.

5.4.3 Metric-based assessment

Using a metric rather than a graphic to assess the distributional coherence of modelled and observed threshold exceedances facilitates the comparison of many thresholds. We therefore aim to summarise the PP- and QQ-plots using metrics that reward accurate estimation of the magnitude distribution function. An unbiased estimate results in a plot that covers the line $y = x$, while a precise estimate results in plots that are stable between sampled values for the mle $\hat{\theta}$ and unrounded data \mathbf{y} . Our approach to creating a metric that summarises these plots is novel in its design, which rewards large sample sizes through their effect to increase the precision of the distribution estimate.

We propose four metrics to summarise deviation from the line $y = x$ in PP- and QQ-plots using the mean absolute distance and mean squared distance in what follows. The calculation of these metrics is described below for a single sampled vector of threshold exceedances on exponential margins $\tilde{\mathbf{z}}$. Let d_0 be the realised metric value

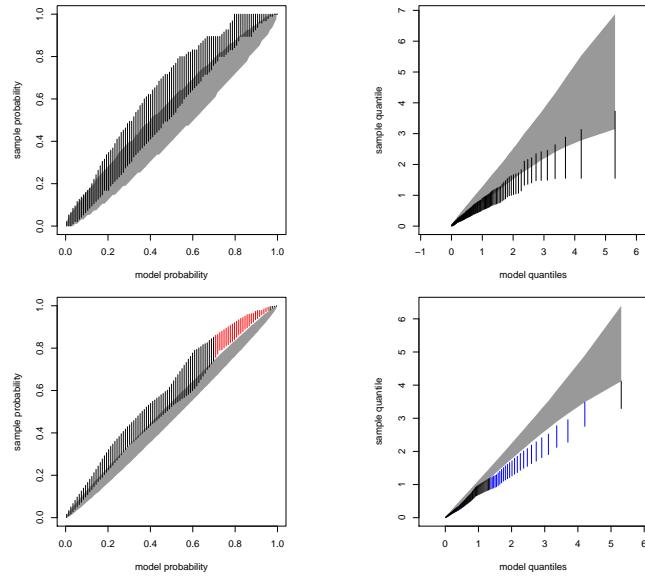


Figure 5.4.1: PP-plots [left] and QQ-plots [right] for threshold exceedance sizes shown on $\text{Exp}(1)$ margins for constant modelling thresholds $v(\tau) = 1.85\text{M}_L$ [top] and $v(\tau) = 1.15\text{M}_L$ [bottom]. 95% tolerance intervals are shown as grey regions, while 95% confidence intervals on each probability or quantile are shown as vertical lines. These are coloured red (blue) where the confidence interval is entirely above (below) the tolerance interval.

for an arbitrary dataset using one of the four methods, and $d = \mathbb{E}_{\mathbf{Y}, \hat{\boldsymbol{\theta}}|\mathbf{x}, \mathbf{v}}(d_0)$ be the expected value of d_0 over the joint distribution of $\mathbf{Y}, \hat{\boldsymbol{\theta}}|\mathbf{x}, \mathbf{v}$, thus accounting for the rounding and parameter uncertainties that are represented by the confidence intervals of Figure 5.4.1. We select a modelling threshold by minimising d and investigate which choice of d_0 is best. Here the expected values of the metrics are calculated by a Monte Carlo approximation.

Smaller values of each metric are preferable, with large values caused by the quantiles of the fitted model being either highly uncertain or incoherent with the observed data. Minimising these metrics provides a new approach to threshold selection, which rewards thresholds that give low sampling variability and small bias in the resulting estimator. The remainder of this section covers the calculation of these metrics, while Section 5.5 explores their relative performance on simulated data.

In the following, $\tilde{\mathbf{z}}^{(i)}$ is the i^{th} parametric-bootstrapped vector of threshold exceedances transformed onto exponential margins for independent, replicated samples $i = 1, \dots, k$. An algorithm to obtain these is given in Appendix B.3. Also let $H_{(i)}(y) : \mathbb{R}^+ \rightarrow [0, 1]$ and $Q_{(i)}(p) : [0, 1] \rightarrow \mathbb{R}^+$, respectively, be the empirical distribution function and the sample quantile function of $\tilde{\mathbf{z}}^{(i)}$ for $i = 1, \dots, k$. The sample quantile functions are defined as linear interpolations of the points $\left\{ \left(\frac{j-1}{\tilde{n}^{(i)}-1}, \tilde{z}_{(j)}^{(i)} \right) : j = 1, \dots, \tilde{n}^{(i)} \right\}$, where $\tilde{n}^{(i)}$ is the length of $\tilde{\mathbf{z}}^{(i)}$ and $\tilde{z}_{(j)}^{(i)}$ is the j^{th} order statistic of $\tilde{\mathbf{z}}^{(i)}$.

The quantile based distance metrics $d_{(i)}(q, 1)$ and $d_{(i)}(q, 2)$ summarise the expected deviation in the QQ-plot of $\tilde{\mathbf{z}}^{(i)}$ from the line $y = x$ at a set of $m \in \mathbb{N}^+$ equally spaced evaluation probabilities $\{p_j = j/(m+1) : j = 1, \dots, m\}$. The two metrics respectively give the mean absolute distance and mean squared distance between model and sample quantiles over the set of evaluation probabilities. They are given by

$$d_{(i)}(q, 1) = \frac{1}{m} \sum_{j=1}^m | -\log(1 - p_j) - Q_{(i)}(p_j) |$$

and

$$d_{(i)}(q, 2) = \frac{1}{m} \sum_{j=1}^m (-\log(1 - p_j) - Q_{(i)}(p_j))^2.$$

In a PP-plot the variance of deviations from the line $y = x$ is greatest when $p_j = 0.5$ and shrinks to 0 as p_j approaches 0 or 1. In the PP-based metrics we therefore weight the sum of the deviations to account for large discrepancies being less surprising for central probabilities. The metrics $d_{(i)}(p, 1)$ and $d_{(i)}(p, 2)$ are therefore calculated using, respectively, the weighted-absolute and weighted-squared errors:

$$d_{(i)}(p, 1) = \frac{1}{m} \sum_{j=1}^m \left[\left(\frac{p_j(1 - p_j)}{\sqrt{n^{(i)}}} \right)^{-1/2} |p_j - H_{(i)}(-\log(1 - p_j))| \right]$$

and

$$d_{(i)}(p, 2) = \frac{1}{m} \sum_{j=1}^m \left[\left(\frac{p_j(1 - p_j)}{\sqrt{n^{(i)}}} \right)^{-1/2} (p_j - H_{(i)}(-\log(1 - p_j)))^2 \right].$$

These deviations are again measured at m equally spaced evaluation probabilities denoted by p_1, \dots, p_m . In the quantile-based metrics the weighting is handled implicitly by choosing equally spaced evaluation probabilities, which gives dense evaluation where discrepancies from $y = x$ are expected to be small and sparse evaluation where they are expected to be large. In this way, the weights reflect the width of the tolerance intervals in Figure 5.4.1.

Uncertainties in the estimated GPD parameters, the size of the exceedance set and the values of the unrounded exceedances should all be accounted for when using a metric to select a modelling threshold. This can be achieved by calculating the distance metrics for each of k realisations of the vector $\tilde{\boldsymbol{z}}$, where each uses one of k bootstrap parameter estimates of $\hat{\boldsymbol{\theta}}$. The expected metric values over these realisations are denoted by $d(a, b)$, where $a \in \{p, q\}$ and $b \in \{1, 2\}$. The expected distance metric $d(q, 1)$ is defined as:

$$d(q, 1) = \frac{1}{k} \sum_{i=1}^k d_{(i)}(q, 1),$$

with the other expected distance metrics defined similarly.

5.4.4 Minimisation procedure

To select the most appropriate threshold, the threshold parameters which minimise the selected metric d must be found. Standard, gradient-based optimisation procedures are not well suited to this task because the censoring mechanism can cause multiple local minima and the Monte Carlo evaluation leads to local roughness over parameter values. When using a simple parametric form for the threshold, such as a constant or stepped threshold (where the change location is known), a simple grid search can be used to overcome these issues and find the threshold parameters that minimise the metrics. For more complex threshold forms, with a higher dimensional parameter space to optimise over, a grid search becomes prohibitively expensive.

To find the threshold parameter set for more complicated thresholds we explore the threshold parameter space in a more principled manner. To do this we use Bayesian optimisation (Snoek et al., 2012) as implemented in the `ParBayesianOptimization` R package (Wilson, 2020). The optimisation procedure begins by evaluating d at a small initial collection of randomly chosen parameter vectors within a bounded search space. Based on the resulting metric values, future evaluation points are selected sequentially as the parameter vector with the greatest expected reduction in d as compared to the current best value. This search method balances evaluations between parts of the parameter space where the metric is known to have low values and parts where it is most uncertain.

Bayesian optimisation is a heuristic search method but has been shown in other applications to find good parameter combinations using a relatively small number of function evaluations (Shahriari et al., 2015). To establish its suitability in our setting we compared Bayesian optimisation to a grid search for two sub-problems; catalogues with a flat threshold and catalogues with a stepped threshold with known change location. In both cases Bayesian optimisation performed favourably compared to grid

search, selecting thresholds close to the true value at a lower computational cost. We do not claim that Bayesian optimisation is the best method for optimising the proposed metrics over threshold parameters, only that it appears to be an efficient method of finding good thresholds.

5.5 Threshold selection on simulated catalogues

5.5.1 Simulation study overview

We consider the performance of the proposed threshold selection metrics on a collection of simulated data sets with either constant or stepped threshold forms. This simulation study illustrates the effectiveness of our method and establishes which of the distance metrics proposed in Section 5.4 is best.

We attempt to select the most appropriate threshold from a set of candidate thresholds. Two censoring types (hard and phased) are considered for magnitudes that are below the modelling threshold. For hard censoring, all simulated continuous magnitudes below the modelling threshold are undetected. In phased censoring the detection probability of each event, $\alpha(y_i, v_i) = \exp(-\lambda[v_i - y_i]_+)$, decreases as the simulated continuous magnitude falls further below the threshold, as controlled by the parameter $\lambda > 0$. The particular choices of exponential decay and the value of λ are arbitrary but were chosen to reflect, in a broad sense, the censoring observed in the Groningen earthquake catalogue. Note that either of these censoring types can result in some rounded magnitudes that are below the threshold even though their simulated continuous values are above the threshold.

5.5.2 Constant threshold, hard censoring

We first use the four proposed metrics to select a constant threshold for 1500 simulated i.i.d. GPD exceedances of the constant threshold $v(\tau) = 0.32M_L$, hard-censored below $v(\tau)$ and rounded to the nearest $0.1M_L$. We first consider the metrics for a single dataset. Expected metric values are calculated at the 41 equally spaced, constant candidate thresholds shown in Figure 5.5.1. The candidate threshold selected by minimising $d(q, 1)$ is the closest threshold on the grid to the true value. This metric also appears to provide the most clearly defined minimum, indicating that it penalises both thresholds that are too low and too high. All four metrics show clear increases in metric value for candidate thresholds that are too low, but not when the candidate threshold is too high. The probability-based metrics do not increase greatly when the candidate threshold is too high, and so fail to adequately reward the inclusion of valid events with smaller magnitudes. This is presumably because they do not sufficiently penalise the increased uncertainty in the estimated parameters when using a higher threshold.

When selecting a constant threshold, the standard approach is to exploit the well-established property that the GPD shape parameter is invariant to threshold choice (Coles, 2001). Point estimates and 95% confidence intervals for ξ were obtained using exceedances of each candidate threshold, accounting for the rounding of observations. The confidence intervals for ξ overlap for all candidate thresholds above $0.275M_L$, and so by the parameter stability method any greater threshold is also valid. The thresholds chosen by our proposed method are therefore consistent with the parameter stability method, but are preferable in that the selected thresholds are not below the true level. Our proposed selection method is also more general; it allows comparison of many non-constant thresholds without the need for subjective and time-consuming interpretation of parameter stability plots.

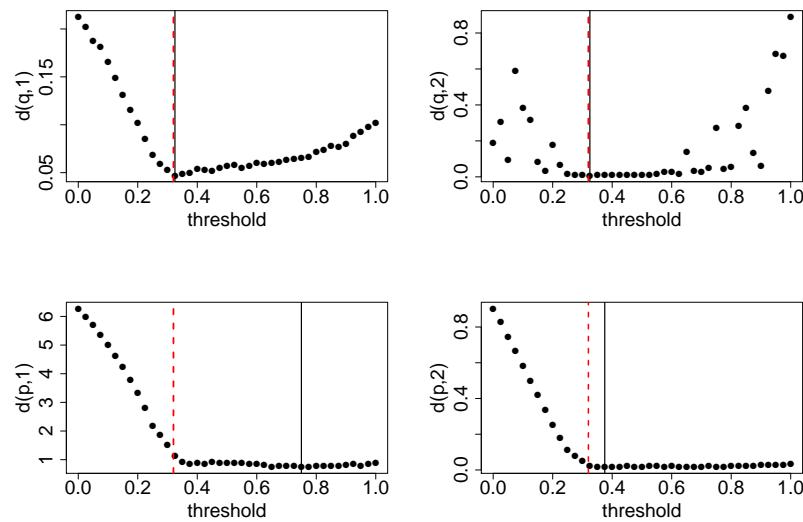


Figure 5.5.1: Flat threshold selection on a simulated catalogue. Top row: expected mean absolute [left] and expected mean squared [right] QQ-distances against threshold value. Bottom row: expected PP-distance metrics based on absolute [left] and squared [right] errors against threshold. Selected and true thresholds are indicated by solid black and dashed red lines.

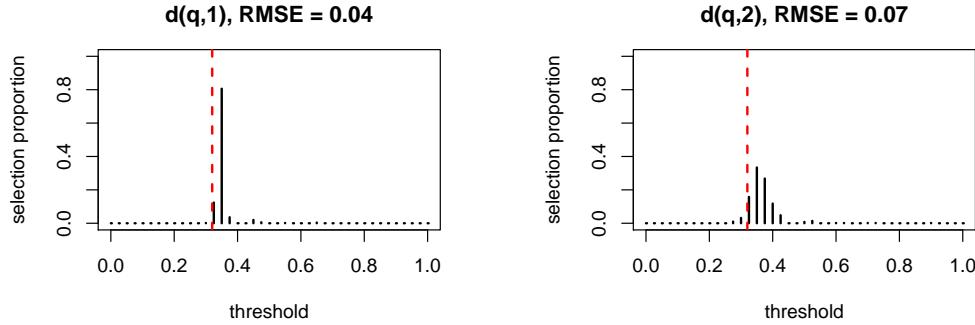


Figure 5.5.2: Sampling distribution of threshold selection methods for quantile-based metrics over 500 simulated catalogues with constant threshold and hard censoring. The true threshold is shown by a dashed red line and the root mean squared error (RMSE) for each method is given in plot titles.

Figure 5.5.2 presents the sampling distribution and RMSE of the thresholds selected from the candidate set by each of the QQ-based metrics over 500 replicated datasets, simulated as previously described. The thresholds chosen by the PP-based metrics are shown in Figure B.4.2 of Appendix B.4 and are frequently much higher than the true value, resulting in higher RMSE values of 0.34 for $d(p, 1)$ and 0.12 for $d(p, 2)$. The metric $d(q, 1)$ has the lowest RMSE and so appears to be the best of the proposed metrics in this case. All metrics have a tendency to overestimate the threshold value; this is likely to be attributable to the hard censoring process. We therefore also consider the performance of each metric using catalogues with phased censoring.

5.5.3 Constant threshold, phased censoring

To assess the performance of each metric on simulated catalogues with phased censoring, we consider the thresholds selected by each metric for each of 500 simulated catalogues. For each catalogue, 2400 i.i.d. GPD exceedances of $0M_L$ were simulated. Each exceedance was retained with probability $\alpha(y_i, v_i)$, as defined in Section 5.5.1 with $v(\tau) = 0.32M_L$ and $\lambda = 7$. This combination of simulated catalogue size and

censoring parameter gave an average catalogue size of 1500 recorded values, similar to those in Section 5.5.2.

The resulting RMSEs in threshold selection over these 500 catalogues were: 0.06 for $d(q, 1)$, 0.08 for $d(q, 2)$, 0.35 for $d(p, 1)$, and 0.12 for $d(p, 2)$. For all metrics the RMSE is slightly increased compared to hard censoring, as threshold selection is made more difficult by the retention of some events that are truly below the threshold. As with hard censoring, the metrics $d(p, 1)$ and $d(p, 2)$ were prone to selecting conservative threshold values and $d(q, 1)$ resulted in the lowest RMSE. Unlike for hard censoring, the sampling distributions of selected thresholds now cover the true threshold values, this is shown in Figure B.4.4 of Appendix B.4. Similar selection properties for each metric were seen when considering more complex threshold forms and so further exposition is limited to the metric $d(q, 1)$, and we subsequently refer to $d = d(q, 1)$.

5.5.4 Non-constant threshold selection

Here catalogues are simulated by generating 4000 i.i.d GPD exceedances of $0M_L$ and censoring (either hard or phased) based on a threshold with $v(\tau) = 0.83M_L$ for $0 < \tau \leq 2000$ and $v(\tau) = 0.42M_L$ for $2000 < \tau \leq 4000$, see Figure 5.5.3 where $\lambda = 7$.

We considered threshold selection behaviour over 500 earthquake catalogues simulated using the above change-point threshold for each of hard and phased censoring. Note that the number of retained events and the threshold change location τ^* within these will vary between simulations because they both depend on the simulated event magnitudes and on how many of these are retained. However, in each case the true value of τ^* is known.

For each simulated catalogue we selected a threshold of the form $v(\tau) = v^{(1)}$ for

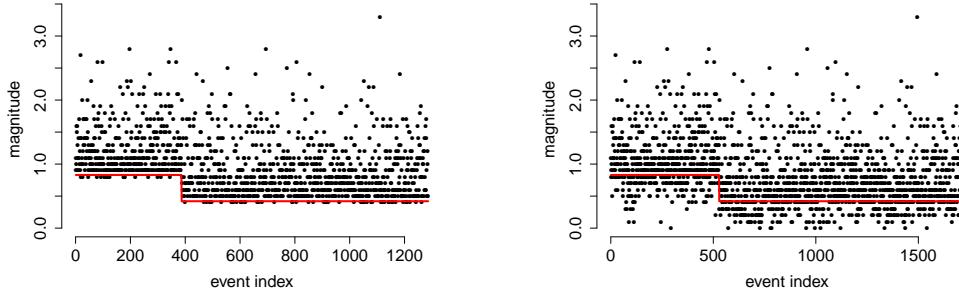


Figure 5.5.3: Example simulated catalogues with hard censoring [left] and phased censoring [right] for stepped thresholds of $(v^{(1)}, v^{(2)}) = (0.83, 0.42)$, shown as a red line, and phasing parameter $\lambda = 7$.

$0 < \tau \leq \tau^*$ and $v(\tau) = v^{(2)}$ for $\tau^* < \tau < \tau_{\max}$, where the threshold parameters $(v^{(1)}, v^{(2)}, \tau^*)$ are unknown. Threshold parameters were selected using the Bayesian optimisation method of Section 5.4.4 to minimise the metric d . The sampling distribution of the errors in the selected threshold parameters are shown in Figure 5.5.4, where it can be seen that our threshold selection method regularly recovers the non-constant modelling threshold to within $\delta/2$ of its true value.

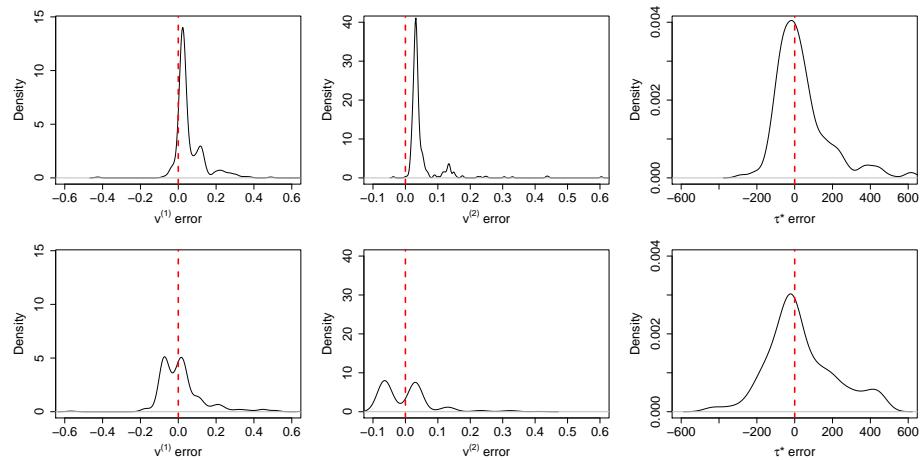


Figure 5.5.4: Marginal sampling distributions of errors in the selected values of $v^{(1)}$ (left), $v^{(2)}$ (center) and τ^* (right) for 500 simulated catalogues with change-point type thresholds and hard (top row) or phased (bottom row) censoring.

Specific findings vary by censoring type. For hard censoring, as would be expected, the threshold levels $v^{(1)}$ and $v^{(2)}$ are rarely selected to be below the true values. The error distribution of τ^* has, in both cases, a mode close to 0 but with large variance. As expected, the sampling variability of the error in each parameter is larger for phased censoring than for hard censoring, though it is reassuring to see that the distributions of selected threshold parameters are now centered on the true values. This demonstrates that the tendency to select threshold values too high for catalogues with hard censoring is a consequence of the censoring mechanism, not a bias in the selection method.

5.6 Application to Groningen earthquakes

5.6.1 Validating data model for Groningen catalogue

We compare GPD and exponential models for Groningen earthquake magnitudes. Rohrbeck et al. (2018) and Marzocchi et al. (2019) demonstrated the importance of acknowledging rounding of observations, and so this is accounted for within the inference for both models. We focus on earthquakes exceeding the constant conservative threshold of $1.45M_L$, subsequently referred to as v_C . This is the magnitude of completion stated by the KNMI (Dost et al., 2012), adjusted to account for rounding.

Both the GPD and exponential models assume that magnitudes are i.i.d.; this is supported by our exploratory analysis of the Groningen catalogue above $1.5M_L$ in Appendix B.1. The two models may be compared by considering the sampling distribution of the estimated shape parameter under a GPD model, because the exponential model is a special case of the GPD where $\xi = 0$. Fitting a GPD to the 311 exceedances of v_C leads to point estimates of $(\hat{\sigma}_{1.45}, \hat{\xi}) = (0.448, -0.018)$ with respective 95% bootstrap confidence intervals of $(0.399, 0.501)$ and $(-0.147, 0.086)$. Since the confidence

interval for ξ covers 0, the exponential model cannot be discounted at the 5% significance level using only exceedances of v_C . A second method of comparison is to fit both an exponential and GPD model to exceedances of v_C and, appealing to the asymptotic distribution of the MLE, perform a likelihood ratio test. This produces a likelihood ratio of 1.04 and associated p -value of 0.214, leading us to draw the same conclusion in both comparisons: that there is insufficient evidence to conclude that the Groningen magnitudes deviate from the Gutenberg-Richter law when using only exceedances of v_C .

However, if an exponential magnitude model is assumed then the uncertainty about ξ is ignored. This has the effect of dramatically, but artificially, narrowing the confidence intervals on the estimated magnitude quantiles, as shown in Figure B.4.3 of Appendix B.4. The potential repercussions of ignoring this uncertainty are described in detail in Coles and Pericchi (2003). A GPD model should therefore be used for the underlying magnitudes, to properly represent this uncertainty when selecting a modelling threshold for the Groningen gas field.

If the rounding of observations had been ignored in the fitting of the GPD model, the point estimates of the GPD parameters would be $(\hat{\sigma}_{1.45}, \hat{\xi}) = (0.453, -0.027)$ with respective standard errors of $(0.039, 0.066)$. The parameter estimates are not significantly different to those using the correct likelihood because the small number of threshold exceedances means that parameter uncertainty obscures the bias induced by neglecting to account for rounding.

Finally, in Figure 5.6.1 we check that the fitted GPD model is consistent with the empirical distribution of exceedances of $1.45M_L$ through the use of the modified QQ and PP plots introduced in Section 5.4.2. Since the tolerance intervals and confidence intervals overlap for both the sample quantiles and sample probabilities, we conclude that a GPD model is appropriate for Groningen earthquake rounded magnitudes

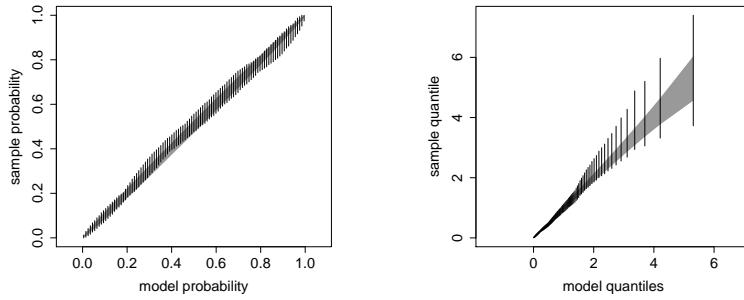


Figure 5.6.1: Modified PP (left) and QQ (right) plots for Groningen magnitudes exceeding $1.45M_L$ under the GPD model. Grey regions show 95% tolerance intervals while vertical lines show 95% confidence intervals on sample probabilities / quantiles. All confidence intervals overlap with the associated tolerance intervals.

exceeding $1.45M_L$.

5.6.2 Parametric threshold forms

Now we select thresholds of two parametric forms for the Groningen catalogue and explore the results of the subsequent inference. The first is a constant threshold, $v(\tau) = v$, where the level v is to be chosen. This will allow us to assess the level of conservatism in the conventional modelling threshold where $v = 1.45M_L$. The second form is a sigmoid-type threshold $v(\tau) = v_R + (v_L - v_R)\Phi((\mu - \tau)/\varsigma)$, with parameters $(v_L, v_R, \mu, \varsigma) \in \mathbb{R}^3 \times \mathbb{R}^+$ and where Φ is the standard Gaussian distribution function. This extends the idea of the change-point threshold to accommodate smooth change in the threshold value centred on μ . The threshold parameters may be interpreted as follows. The left and right asymptotic levels of the threshold are given by v_L and v_R , μ is the index-time at which the threshold takes the value $(v_L + v_R)/2$, and ς controls how rapidly the threshold changes about μ , with $\varsigma \rightarrow 0$ corresponding to a step change. In the context of the Groningen catalogue we expect that $v_R < v_L$.

5.6.3 Threshold selection

5.6.3.1 Constant threshold

A grid search was used to find the flat threshold that minimises the metric d , as shown in Figure 5.6.2. There are two local minima at $v = 0.85M_L$ and $v = 1.07M_L$, the latter being the global minimum. For thresholds greater than $1.07M_L$, including the conservative threshold of $1.45M_L$, the metric values are increasing as not all viable data are utilised. For thresholds less than $0.85M_L$ the metric also increases as the validity of the tail model breaks down. The small peak between these minima is likely attributable to the reduction of the m_c over time. In Figure 5.2.1 we saw that fewer small magnitude events are censored at later times. The minimum at $1.07M_L$ uses less data to achieve good distributional agreement for the entire period, while the minimum at $0.85M_L$ compromises on the distributional agreement at early times to retain a larger proportion of the data. As the threshold is lowered between magnitudes $0.95M_L$ and $0.85M_L$, enough additional data are added to more than compensate for the reduced goodness-of-fit in the early part of the observation period and so the metric value reduces. Since the global minimum corresponds to the more conservative threshold, we select $1.07M_L$ as our constant modelling threshold.

5.6.3.2 Sigmoid threshold

Bayesian optimisation was used to find the sigmoid threshold parameters $(v_L, v_R, \mu, \varsigma)$ that minimise the metric d , where the search space was constrained to the region $[0.4, 1.7]^2 \times [200, 1100] \times [1, 500]$. For an initial set of 20 randomly selected threshold parameter combinations, d was evaluated. A further fixed budget of 100 metric evaluations was allocated and the thresholds with the smallest metric value retained for further inspection. To assess the sensitivity of the selected threshold to the set of initial evaluation points, this was repeated for five initial parameter combination

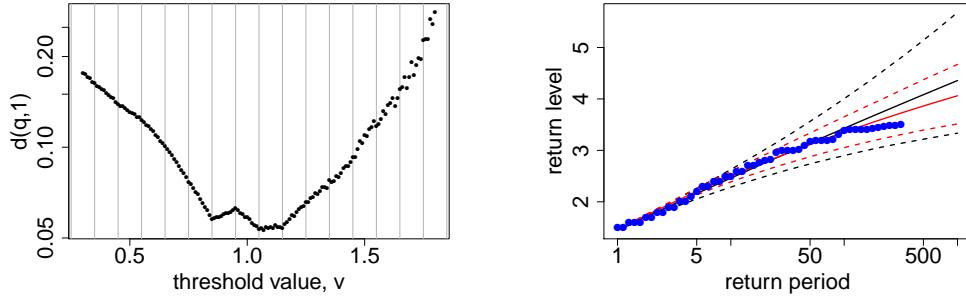


Figure 5.6.2: [Left] Grid search to minimise $d(q, 1)$ over threshold values v . Metric values are shown on log-scale and vertical lines mark the edges of magnitude rounding intervals. [Right] Point estimates (solid lines) and 95% confidence intervals (dashed lines) for the conditional return levels for exceedances of $1.45M_L$, using the conservative (black) and selected thresholds (red). Sample conditional return levels are shown in blue.

sets.

The thresholds with the lowest values of d based on each initialisation are shown in Figure 5.6.3 (left). The selected threshold values at the ends of the observation interval appear to be stable across initialisations, but the transition between these levels is not. Further investigation supports the stability of the end levels; the blue and turquoise thresholds have significantly greater metric values than the other thresholds, suggesting that these initialisations had too few evaluations to explore beyond a local minimum. These conclusions are consistent with the simulation study of Section 5.5.4, illustrating that threshold levels are more easily estimated than the change between those levels.

A second Bayesian optimisation was performed, fixing the end levels of the sigmoid threshold to the those shared by the best performing thresholds in the previous optimisation, namely $(v_L, v_R) = (1.15, 0.76)$. This reduces the dimension of the parameter space and simplifies the optimisation task. Using the same procedure as for the un-

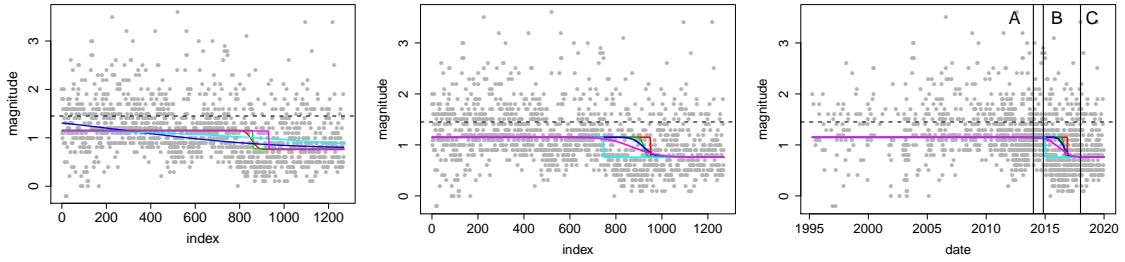


Figure 5.6.3: Selected sigmoid thresholds using Bayesian optimisation from 5 random initial parameter sets. [left] Optimising over all thresholds parameters. [centre, right] Optimising over (μ, σ) and fixing $(v_L, v_R) = (1.15, 0.76)$ on index- (centre) and natural- (right) timescales. Colours are comparable only between centre and right plots. Dashed horizontal lines show the conservative threshold value. Important dates relating to the development of the Groningen seismic detection network are shown as vertical lines: (A) development begins, (B) first additional sensors activated, (C) upgrade complete.

constrained optimisation, the resulting selected thresholds from each initialisation are shown in Figure 5.6.3 (centre). Upon repeated Monte Carlo evaluation of the metric value for each of these thresholds, there is insufficient evidence to select one over the others. When transformed onto the natural time scale, as shown in Figure 5.6.3 (right), the selected thresholds are all consistent with the known dates at which sensor installation occurred. This shows that from the earthquake catalogue alone our method is able to detect the starting and ending threshold levels and the period in which it changed. However, we cannot identify precisely the way in which the threshold changed during the installation period. This is not a major setback, since between the most and least conservative of the chosen thresholds (turquoise and red in the centre and right panels of Figure 5.6.3) the expected number of observations above the threshold differs by only 50 earthquakes. We fitted the GPD model using each of these five threshold functions, reaching similar conclusions, and so present

further results for only the turquoise threshold.

5.6.3.3 Threshold comparison

We compare the conservative, selected constant, and selected sigmoid thresholds, which are referred to as \hat{v}_C , \hat{v} and \hat{v}_S respectively. Comparisons are made on: the expected metric value, the number of events used to fit the GPD model, the estimated GPD parameter values, and the estimated return levels.

Metric evaluations are subject to Monte Carlo noise and so the metric value was evaluated 100 times for each threshold. The mean metric value and 95% Monte Carlo noise intervals were calculated to be: 0.091 (0.088, 0.096) for \hat{v}_C , 0.054 (0.053, 0.055) for \hat{v} , and 0.041 (0.039, 0.043) for \hat{v}_S . This suggests that the model fit using \hat{v}_S fits the observed data best, with \hat{v} being preferred over \hat{v}_C . These improvements in model fit may be attributable to the increased data usage of the selected thresholds. The threshold \hat{v}_C is at the edge of a rounding interval and so utilises 311 threshold exceedances in the resulting model. For thresholds \hat{v} and \hat{v}_S , the rounding of magnitudes means that the exact number of exceedances is unknown. The expected number of exceedances under the fitted magnitude models are 629 and 702 for \hat{v} and \hat{v}_S , respectively. By using either of the selected thresholds, we have more than doubled the size of usable catalogue as compared to the conservative threshold.

Figure 5.6.4 (left) shows the estimated parameter values under the fitted GPD model using each threshold. The uncertainty in both parameters is reduced when using \hat{v} rather than \hat{v}_C , and further reduced when using \hat{v}_S . To give a sense of scale in this reduction, we can calculate the number of additional exceedances of \hat{v}_C to which they are equivalent, under the assumption that the standard error of parameter estimates scales with exceedance count n as $n^{-1/2}$. In doing this, the additional 318 and 391 small magnitude earthquakes included by, respectively, using \hat{v} or \hat{v}_S are equivalent

to 363 or 509 additional events above \hat{v}_C . Therefore, point-for-point, the small magnitude earthquakes are at least as valuable as additional data above v_C for parameter estimation.

When modelling exceedances of \hat{v} or \hat{v}_S the respective point estimates and 95% confidence intervals for the shape parameter are -0.084 (-0.168, -0.017) and -0.069 (-0.144, -0.008). Using exceedances of \hat{v} or \hat{v}_S leads to only 0.5% or 1.5% of the sampling distribution for $\hat{\xi}$ being above 0. This provides empirical evidence that the Groningen magnitude distribution has a finite upper endpoint, unlike the conventional Gutenberg Richter magnitude model. This dramatic conclusion could not be reached using the smaller dataset exceeding \hat{v}_C , where 33% of the sampling distribution for $\hat{\xi}$ lay above 0.

Similar conclusions can be reached by using likelihood ratio tests to compare Exponential and GPD models for exceedances each of v_C , \hat{v} and \hat{v}_S ; the respective p -values are 0.78, 0.046, and 0.064. By using more of the available data, we have increased our ability to discern between an exponential model and the observed magnitude distribution. The conclusions that can be drawn from this test are in agreement with, but are less strong than, those of the previous comparison: a Gutenberg Richter magnitude model is likely inferior to a GPD. The discrepancy in conclusion strength between the two comparisons is likely due to the asymptotic assumptions of the likelihood ratio test not being met by our finite sample size.

The estimated conditional return levels above $1.45M_L$ are shown using each threshold in Figure 5.6.4 (right). The estimated return levels are similar when using \hat{v} and \hat{v}_S , but confidence intervals for large return periods are narrower when using \hat{v}_S . In either case, the return levels have both smaller point estimates and uncertainties by using our threshold selection method than when using the conservative threshold. This is an important finding when deciding what measures to take when designing or retrofitting

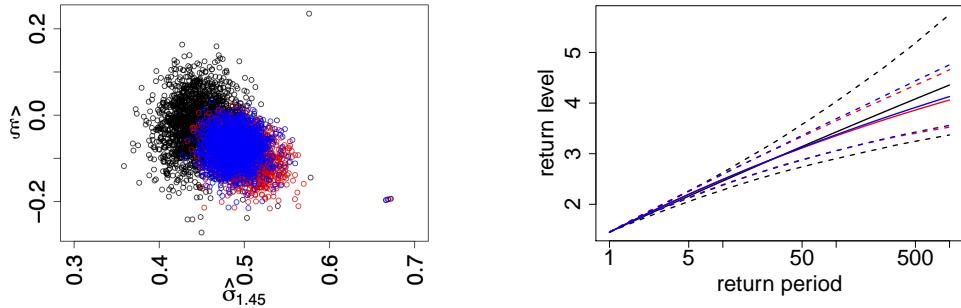


Figure 5.6.4: Bootstrap GPD parameter estimates based on exceedances of the conservative (black), flat (red) and sigmoid (blue) thresholds [left]. Estimated return levels in M_L and 95% confidence intervals for magnitudes exceeding $1.45M_L$ [right].

earthquake defences for buildings.

5.7 Discussion / Conclusion

This chapter introduced a principled method to select a time-varying modelling threshold for an extreme value analysis. The effectiveness and value of using this method to include additional, less extreme events in the analysis were demonstrated through simulation studies. Although the method was developed in the context of earthquake catalogues, and to accommodate the additional challenges to inference that these pose, the core of our method is applicable to extreme value threshold selection more generally and we anticipate it having a much broader impact.

Using the new threshold selection method, we have been able to identify the period in which the Groningen sensor network was being improved by using the earthquake catalogue alone. Our threshold selection method more than doubled the usable size of the Groningen earthquake catalogue compared to using the conservative threshold given by the KNMI, whilst also improving model fit. This has several important implications beyond the direct improvement to statistical inference.

The use of these additional small magnitude earthquakes leads to greater precision in the estimates of high magnitude quantiles, which is potentially a huge benefit by reducing the cost of designing, constructing or retrofitting earthquake defences. Following threshold selection, a Bayesian modelling approach would allow quantile uncertainty to be included naturally when designing defences against natural hazards (Coles and Tawn, 1996; Fawcett and Green, 2018; Jonathan et al., 2021) and estimates with greater precision can reduce the cost required to provide protection with equivalent confidence. The gain we have made in the efficiency of statistical inference can be translated to a tangible economic benefit of using the additional data recorded by improving the censor network. The more efficient use of the available data has allowed us to conclude, for the first time based on empirical evidence alone, that Groningen earthquake magnitudes are likely to have a light-tailed distribution. Using the conservative threshold level this conclusion could only have been achieved by waiting many years to observe additional large magnitude earthquakes. Finally, using a less conservative modelling threshold provides a return on the substantial investment into the earthquake detection network around the Groningen gas field. When a non-constant threshold is selected, the added value of the network improvements is exploited fully and the subsequent modelling threshold can also offer insights into the reduction of m_c over time.

A limitation of the work is that the computational effort required to select a modelling threshold is relatively high. We do not view this as a large drawback since threshold selection must be performed only once through the modelling process. An area for further development would be to investigate alternative, exact methods to optimise the expected selection metric over the threshold parameters. One possible extension to our approach would be to adapt the data model to account for magnitude measurement error causing events to be recorded within incorrect rounding intervals. Another,

more ambitious, extension might consider a selection of spatio-temporal threshold function to describe spatial variability as well as the temporal evolution of event detection. Finally, an extensive comparison of our proposed and standard extreme value threshold selection methods would be a valuable piece of further work, given its critical importance in extreme value methods.

Chapter 6

Improving and extending the ETAS formulation

6.1 Introduction

6.1.1 The ETAS model

The epidemic-type aftershock sequence (ETAS) model describes marked, clustered point processes. In the simplest form of the ETAS model, each event is a time-mark pair (t, m) where typically $t \in (0, t_{\max})$ and $m \in \mathbb{R}^+$. A single realisation of the point process is then a collection of $n \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$ such pairs $\{(t_i, m_i) : i = 1, \dots, n\}$, where the number of events n is a random variate. The ETAS model was developed for earthquake catalogues by Ogata (1988), but has since found applications in finance, the natural- and social-sciences (Reinhart, 2018). In the context of seismicity, the ETAS model can be used descriptively or predictively. Descriptive modelling can be used to address questions of scientific interest about the earthquake generating process, while predictive modelling can inform earthquake defence policies that help

to safeguard against future hazards.

A point process model can be defined by its intensity function, which determines the expected number of events in the process and how those events are located across time or space. In the ETAS model this intensity function is increased locally after each event, resulting in a clustered point process. The size of the additional intensity contribution from each event is determined by the value of its mark, where larger mark values are expected to yield a larger number of additional events. In the context of seismicity modelling, each event represents an earthquake and the associated mark represents its magnitude. In this way, each event in an ETAS point process may be considered as either a background or triggered event; background events are attributable to the initial intensity component and triggered events are attributable to an intensity component caused by a previous event. Within the standard ETAS model, event magnitudes are independent of one another and of event type. As with background events, triggered events provide an additional intensity contribution and so can also trigger further events. The ETAS model may therefore be interpreted as either the superposition of point processes or as a branching process.

In the simplest form of the ETAS model, background events come from a Poisson process with constant intensity μ . This intensity is then augmented by each of the events, indexed by i , according to their magnitude and occurrence time. The magnitude productivity function $\kappa : \mathbb{R} \rightarrow \mathbb{R}_0^+$ determines the expected number of events triggered at a given magnitude. The probability density function $h : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ characterises the delays between triggering and triggered events. The resulting conditional intensity function depends on the history of the process at time τ , $\mathcal{H}_\tau = \{(t_i, m_i) : t_i < \tau\}$ and is parameterised by the vector $\theta = (\mu, \theta_\kappa, \theta_h)$. Specifically, the conditional intensity λ

at time τ is defined as:

$$\lambda(\tau|\mathcal{H}_\tau; \theta) = \mu + \sum_{i:t_i < \tau} \kappa(m_i; \theta_\kappa) h(\tau - t_i; \theta_h). \quad (6.1.1)$$

In seismological applications of the ETAS model the most common parametric forms for κ and h are the empirical earthquake ‘laws’ which will be introduced in Section 6.3.1. Broadly, these forms for κ and h impose that the increment in intensity following each event should increase with the magnitude of the event and should diminish with time since that event. The first contribution of this chapter is to show that these empirical earthquake laws may be represented by the single, encompassing model provided by the generalised Pareto distribution (GPD). We show that using this encompassing model better represents epistemic uncertainty (by broadening the class of models which can be represented) and increases the statistical efficiency of parameter inference.

6.1.2 Magnitude modelling

In the standard ETAS formulation, event marks are modelled as an independent component of the point process. That is, marks are modelled as independent and identically distributed (i.i.d.) random variables with a common probability density function $f(m; \psi)$.

Within the seismology literature, previous studies have investigated the validity of earthquake magnitudes being identically distributed, i.e., that there is a single magnitude distribution for all earthquakes. These studies rely on ‘declustering’ the earthquake catalogue into mainshock and aftershock events and then testing for differences in the distributions of the two samples (Stallone and Marzocchi, 2019). This clustering is frequently done using a window- or distance-based approach following the events of largest magnitude, in which case mainshocks and aftershocks are not coherent with

the complex triggering structure of the ETAS model (Varini et al., 2020). An alternative approach to declustering is to use point estimates of the ETAS parameters to perform a model-based stochastic declustering (Zhuang et al., 2002). This approach yields, for each event, a probability distribution that gives the probability that the event is attributable to each of the previous earthquakes or is a background event. This stochastic declustering respects the branching structure of ETAS point processes and gives some measure of the uncertainty on the estimated cluster allocation. However, this approach importantly fails to capture the joint uncertainty across all events or to include the additional uncertainty arising from the ETAS parameters themselves being estimated.

There have also been previous investigations into the assumed independence between earthquake magnitudes (Gulia et al., 2018; Stallone and Marzocchi, 2019; Cai et al., 2021). These rely on the previously mentioned declustering methods and so suffer from the same limitations. Additionally, attention is often restricted to dependence between the magnitudes of the main-shock and largest aftershock within each identified cluster. This focus may be motivated by mathematical convenience or else by an empirical relation known as Båth's law (Båth, 1965), which relates these quantities. However, this approach fails to use all available data and the motivation based on Båth's law is unsound; Båth's law can be shown to arise as an artefact of the window-based declustering method when magnitudes are truly i.i.d. (Lombardi, 2002; Helmstetter and Sornette, 2003).

A different dependence structure was investigated by Chavez-Demoulin et al. (2005), who incorporated auto-regressive dependence into a financial application of the ETAS model. In this setting, events correspond to financial losses exceeding a given threshold and marks give the size of this exceedance. The event times are modelled using an ETAS point process, modified so that each mark has a generalised Pareto conditional

distribution (as given in equation (6.3.6)) when the value of the previous mark is known. The scale parameter of this conditional distribution is linked to the previous mark so that $M_i|M_{i-1} = m_{i-1} \sim \text{GPD}(\exp\{a + bm_{i-1}\}, \xi)$, where a, b and ξ are parameters to be estimated. This is a relaxation of the assumption that marks are i.i.d. under the standard ETAS model that leads to marginal mark distributions that are not within the GPD family.

A third dependence structure, which addresses the branching structure of the ETAS process, was investigated by Žugec (2019). Their approach assumes that the marginal distribution of all magnitudes can be described by a single exponential distribution. The dependence between the magnitudes of triggering and triggered events is incorporated through the use of a Farlie - Gumbel - Morgenstern copula, constrained to only allow positive dependence. This dependence structure has the benefit of preserving the marginal magnitude distributions and respecting the branching structure of the ETAS process. However, the choice of copula used to model dependence does not allow magnitudes to be dependent at extreme levels, allowing only ‘near’ extremal independence (Ledford and Tawn, 1997). Additionally, Žugec (2019) presents only theoretical results for their model and does not provide an inference method with which to fit their model to an observed earthquake catalogue.

In Section 6.4 we introduce an alternative relaxation of the i.i.d. mark assumption. This is designed to ensure that: dependence is based on the branching structure of the process, rather than the temporal ordering of events; the marginal mark distributions are in the GPD family; and strong dependence at extreme magnitudes may be represented. We begin Section 6.4 by introducing an extension of the ETAS model to allow separate magnitude distributions for background and triggered events. This is based on the estimated branching structure of the ETAS process and accounts for all sources of uncertainty within this model. We then consider a further extension by

allowing dependence between the magnitudes of triggered events and the magnitude of the event that triggered them. We present inference methods for each model. These methods allow for simple hypothesis tests on the presence of each of these magnitude features, while properly reflecting uncertainties within the ETAS paradigm.

6.1.3 Bayesian ETAS modelling

When fitting the ETAS model to an observed catalogue of earthquakes, one must estimate the parameter vectors θ and ψ that describe the point process intensity $\lambda(t; \theta)$ and magnitude distribution $f(m; \psi)$. Direct maximisation of the likelihood function is the most common approach to parameter estimation for the ETAS model. This approach is described in Section 6.2.1. There are three main issues with the direct approach to inference for the standard ETAS model. Firstly, parameter uncertainty is difficult to propagate into earthquake forecasts. This means that in many cases only the point estimates of parameters are retained (Ogata, 1988; Veen and Schoenberg, 2008). This is a particular issue when the model is used to predictively simulate an ensemble set of futures to aid decision-making; using only point estimates will lead to an overly narrow set of possible outcomes being represented. Secondly, the likelihood function requires numerical maximisation. Direct numerical maximisation is unreliable for the ETAS likelihood because the parameters are not orthogonal, the likelihood function can have multiple modes and local regions may be almost flat (Veen and Schoenberg, 2008). These features also mean that measures of parameter uncertainty based on asymptotic standard errors can be unsuitable or unreasonable. Finally, each evaluation of the likelihood for the ETAS model is computationally expensive, and this becomes the increasingly prohibitive factor as the number of events in the observed catalogue grows. This final issue can be compensated for through use of the expectation-maximisation algorithm introduced by Veen and Schoenberg

(2008). An additional, fourth issue arises when the i.i.d. magnitude assumption is relaxed: the likelihood becomes intractable making evaluation, and therefore direct maximisation, of the likelihood function impossible.

The aforementioned challenges presented by the ETAS likelihood motivate a Bayesian approach to modelling. This avoids reliance on asymptotic results for parameter uncertainty and makes propagating these uncertainties into predictions straightforward. It also offers the possibility to incorporate expert knowledge into the fitting procedure, which could be particularly helpful when modelling point processes with relatively few observed events. Fitting the ETAS model in a Bayesian framework requires Markov Chain Monte Carlo (MCMC) sampling methods, which makes it more computationally challenging than a frequentist approach. However, this cost can be greatly reduced through the use of the Gibbs sampling scheme introduced by Ross (2016).

Ross (2016) interprets the ETAS model as a branching process and introduces a latent vector to describe the branching structure. This vector defines the graph representation of triggering and triggered events and is described fully in Section 6.2. Introducing this branching vector facilitates inference for the ETAS model by allowing alternate estimation of the branching structure and ETAS parameters. This provides a method of probabilistically declustering earthquake sequences that fits naturally into the Bayesian paradigm. The approach was shown to reduce dependence between groups or ‘blocks’ of the ETAS parameters and was shown experimentally to reduce the cost of evaluating the likelihood function. A remaining issue, which we address through our proposed reparameterisation in Section 6.3, is that parameters remain highly dependent on one another within these blocks.

6.1.4 Contributions and chapter outline

We first show that the earthquake laws used in Ross (2016) and many other applications of the ETAS model are restricted forms of the generalised Pareto distribution. We reformulate the ETAS intensity in terms of this model, which reduces within-block parameter dependence and leads to more efficient inference. The reformulation properly accounts for both model-uncertainty and parameter-uncertainty, leading to conclusions and predictions that are more robust.

We then propose two extensions of the ETAS model that build explicitly on its representation as a branching process. Our first extension allows distinct magnitude distributions for background and triggered earthquakes. The second extension allows magnitudes of triggered earthquakes to be dependent on the magnitude of the triggering event. These extensions allow simple, structured tests for the additional magnitude properties that could not be included in the ETAS model when taking a direct approach to fitting the model.

The temporal ETAS model (Ogata, 1988) has many extensions, including those which allow a variable background rate and those which model events over space and time (Ogata, 2011; Kolev and Ross, 2020). The ideas presented in this paper extend readily to these settings and so for simplicity of exposition we restrict our attention to the temporal ETAS model.

The rest of this chapter is set out as follows. In Section 6.2 we outline in detail both the direct and latent variable inference procedures for the ETAS model. We identify the source of the efficiency gain that was demonstrated numerically by Ross (2016) and calculate the order of this efficiency gain. In Section 6.3 we propose a more general parameterisation of the empirical laws that are commonly used for the productivity function, delay and magnitude distributions. The proposed extension broadens the

class of models that can be represented within the ETAS framework and aims to make inference more efficient by reducing the dependence within groups of parameters. In Section 6.4, we introduce two extensions to the standard ETAS magnitude model and describe estimation procedures for these extensions. The first extension utilises the latent branching structure in an ETAS point process to allow distinct magnitude distributions for background and triggered events. The second extension moves beyond the assumption of independence to allow triggered magnitudes to depend on the magnitude of their triggering event. Section 6.5 demonstrates the benefits of these approaches using simulated earthquake catalogues. Finally, Section 6.6 gives concluding remarks and suggests areas for further work.

6.2 Estimation of ETAS parameters

6.2.1 Direct estimation: ETAS as a point process

Use of the ETAS model requires estimation of the parameter vectors θ and ψ of the intensity function $\lambda(\tau; \theta)$ and mark distribution $f(m; \psi)$. This section describes how these parameter vectors can be estimated for the standard ETAS formulation (with i.i.d. magnitudes) using a direct point process representation. This is accomplished using a single, observed catalogue of n events $y = (y_1, \dots, y_n) = (\{t_1, m_1\}, \dots, \{t_n, m_n\})$, which is a realisation of the marked point process $Y = (Y_1, \dots, Y_N)$ where each element of Y is a time-magnitude pair and the number of events N is random. In order to be included in the catalogue, event magnitudes must exceed a minimum value of m_0 and must occur before the end of observation at time t_{\max} . The support for each element of Y is therefore the time-magnitude window $[0, t_{\max}] \times [m_0, \infty)$.

To find the joint posterior distribution of θ and ψ , we first require the likelihood function $\pi_{Y|\theta,\psi}(y)$. The corresponding log-likelihood has the same form as that of an

inhomogeneous Poisson process with independent marks and the history-dependent intensity function from expression (6.1.1), namely:

$$\log \pi_{Y|\theta,\psi}(y) = \sum_{i=1}^n \log f(m_i; \psi) - \int_0^{t_{\max}} \lambda(\tau | \mathcal{H}\tau; \theta) d\tau + \sum_{i=1}^n \log \lambda(t_i | \mathcal{H}t_i; \theta). \quad (6.2.1)$$

The log-likelihood (6.2.1) is separable in θ and ψ and so these vectors may be estimated separately. We focus in this section on estimation of θ , which is the more challenging aspect when magnitudes are independent and identically distributed. Expression (6.2.2) gives the log-likelihood of θ in expanded form, where $H(\tau; \theta_h)$ is the distribution function associated with the aftershock delay density $h(\tau; \theta_h)$.

$$\begin{aligned} \log \pi_{Y|\theta}(y) &= -\mu t_{\max} - \sum_{i=1}^n \kappa(m_i | \theta_\kappa) H(t_{\max} - t_i | \theta_h) \\ &\quad + \sum_{i=1}^n \log \left[\mu + \sum_{j: t_j < t_i} \kappa(m_j | \theta_\kappa) h(t_i - t_j | \theta_h) \right]. \end{aligned} \quad (6.2.2)$$

There is no conjugate form for this model and so Monte Carlo methods are required to obtain approximate samples from the posterior distribution. The standard and most direct approach is to use a random walk Metropolis algorithm to do so; Ross (2016) describes how this can be implemented for the ETAS model. As with likelihood-based inference, this approach suffers from the local flatness of the likelihood and strong, complex parameter dependence pointed out by Veen and Schoenberg (2008). The direct approach to inference requires a large number of evaluations of the log-likelihood (6.2.2). Each evaluation is an $O(n^2)$ operation due to double summation in the final term, which makes the sampling algorithm very slow for even moderately sized catalogues. These issues can be alleviated by instead considering the ETAS model as a branching process.

6.2.2 Latent estimation: ETAS as a branching process

Self-exciting point processes such as the ETAS model may be interpreted as branching processes (Kirchner, 2017). These processes can be represented as a graph, where events are nodes and directed edges connect an event to those that it triggers. This branching structure produces a collection of trees, which are unobserved but can be modelled by a latent n -vector, $B = (B_1, \dots, B_n)$. Element i of this vector, $B_i \in \{0, 1, \dots, i-1\}$, is equal to zero if event i is a background event (that is not triggered by a previous event) and otherwise gives the index the triggering event. The branching vector therefore identifies the set of any ‘child’ events triggered by each ‘parent’ event $i \in 1, \dots, n$, which we index by the sets $\mathcal{C}_i = \{j \in \{i+1, \dots, n\} : b_j = i\}$.

Figure 6.2.1 shows the graph representation and associated branching vector B of a toy example from the ETAS model, where Y_1, \dots, Y_7 are a temporally ordered set of events. The ‘root’ of each tree is formed by a background event. In this example there are three trees initiated by the background events Y_1 , Y_2 and Y_5 . The example shows that an event may directly trigger zero, one, or multiple further events, as demonstrated by the child sets $\mathcal{C}_2 = \emptyset$, $\mathcal{C}_1 = \{3\}$ and $\mathcal{C}_3 = \{4, 6\}$. Note that triggered events may in turn trigger further triggered events, as demonstrated by Y_3 . Finally, notice that triggering and triggered events are not necessarily consecutive or even contiguous in time. This is because events can trigger multiple further events and distinct trees can overlap in time. These effects can be seen in the toy example where Y_6 is separated in time from its triggering event Y_3 by both the background event Y_5 and also by Y_4 , another child event of Y_3 .

An alternative approach to estimation of the ETAS parameters utilises this branching process interpretation and in particular the latent branching vector, B . This method was proposed by Veen and Schoenberg (2008) and brought to the Bayesian setting by Ross (2016). In this approach, the conditional intensity function (6.1.1)

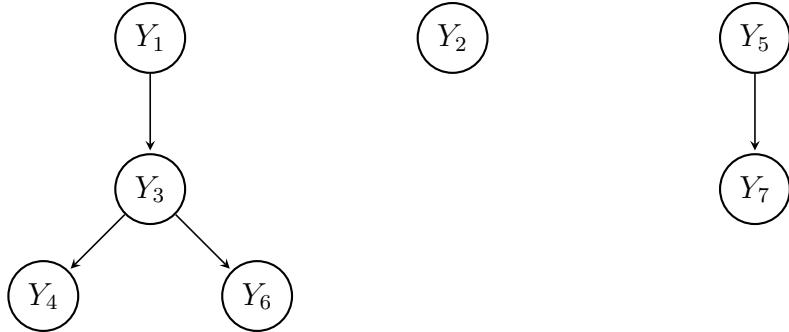


Figure 6.2.1: Graphical representation of a toy example from the ETAS model with branching vector $B = (0, 0, 1, 3, 0, 3, 5)$. Note that an event may trigger multiple further events, that event sequences may overlap temporally, and that both background and triggered events may trigger further events. Each event $Y_i = (T_i, M_i)$ represents a time-magnitude pair where $T_1 < T_2 < \dots < T_7$.

can be regarded as the superposition of $n + 1$ Poisson process intensities. These Poisson processes are indexed $0, \dots, n$, with intensity functions $\lambda_0, \dots, \lambda_n$ such that $\lambda(\tau|\mathcal{H}\tau; \theta) = \sum_{i=0}^n \lambda_i(\tau|y_i; \theta)$. Each Poisson process represents one source of seismicity: the background events, those triggered by the first earthquake, by the second earthquake and so on. The intensities of these Poisson processes, λ_0 to λ_n , are defined as:

$$\lambda_i(\tau|\mathcal{H}\tau; \theta) = \begin{cases} \mu & \text{if } i = 0 \text{ and } \tau \geq 0, \\ \kappa(m_i; \theta_\kappa)h(\tau - t_i|t_i; \theta_h) & \text{if } i \in \{1, \dots, n\} \text{ and } \tau > t_i, \\ 0 & \text{otherwise.} \end{cases} \quad (6.2.3)$$

Under this interpretation, the branching vector B specifies to which Poisson component of the ETAS process each event belongs. Given B , the full ETAS point process is simply the superposition of independent Poisson processes and the conditional log-likelihood given B is the sum of Poisson process log-likelihoods. By defining $y_0 = (t_0, m_0) = (0, 0)$ and letting $y_{b_i} = (t_{b_i}, m_{b_i})$ denote the parent of event i , the

conditional ETAS log-likelihood given B is:

$$\begin{aligned} & \log \pi_{Y|\theta,B}(y) \\ &= \sum_{i=1}^n \left\{ \log \lambda_{b_i}(t_i; \theta) - \int_0^{t_{\max}} \lambda_i(\tau) d\tau \right\} \\ &= \sum_{i=1}^n \{\log \kappa(m_{b_i}|\theta_\kappa) h(t_i - t_{b_i}|\theta_h)\} - \mu t_{\max} - \sum_{i=1}^n \{\kappa(m_i|\theta_\kappa) H(t_{\max} - t_i|\theta_h)\}. \quad (6.2.4) \end{aligned}$$

The conditional log-likelihood (6.2.4) no longer contains a double summation and so is much less costly to evaluate than the full likelihood (6.2.1). This reduced computational cost motivates a conditional approach to estimation of $\pi_{\theta|Y}$. This can be done by using a Metropolis-within-Gibbs sampler to alternately draw samples from the conditional posteriors $\pi_{B|Y,\theta}(b)$ and $\pi_{\theta|Y,B}(\theta)$ at the present value of the other parameter set. Implementing this sampler then requires a tractable form for $\pi_{B|Y,\theta}(b)$. This distribution has a particularly simple form if the priors on each B_i are pair-wise independent and discrete uniform on their respective supports. This choice of prior can be motivated by the inter-event times and magnitudes being unknown *a priori*.

In that case:

$$\pi_{B|Y,\theta}(b) = \prod_{i=1}^n \pi_{B_i|Y,\theta}(b_i) \quad \text{where } \pi_{B_i|Y,\theta}(b_i) = \frac{\lambda_{b_i}(t_i; \mathcal{H}t_i, \theta)}{\sum_{j=0}^{i-1} \lambda_j(t_i; \mathcal{H}t_i, \theta)}. \quad (6.2.5)$$

6.2.3 Benefits of the latent estimation approach

The computational benefits of the latent estimation approach were demonstrated empirically by Ross (2016). Here, we make some further comments to clarify the source of this improvement.

To sample values from $\pi_{\theta|Y}(\theta)$, the direct estimation procedure requires one $O(n^2)$ evaluation of the full log-likelihood at each step in the Markov chain. The latent estimation procedure requires, at each step in the Markov chain, one evaluation of the conditional likelihood $\pi_{\theta|Y,B}(\theta)$ and one sample from $\pi_{B|Y,\theta}(b)$. From expression

(6.2.4), it is evident that evaluating $\pi_{\theta|Y,B}(\theta)$ is an $O(n)$ operation. To draw the full vector B from this conditional posterior, we must sample each of its n elements. Sampling the element B_i draws a sample from a discrete distribution with i possible outcomes; this is an $O(i)$ operation (Walker, 1977) that must be done for each of the n earthquakes. Sampling the full vector B is therefore an $O(\log(n!))$ operation. Therefore, for large n , sampling the branching vector from $\pi_{B|Y,\theta}(b)$ is the most costly operation when obtaining each sample from the posterior distribution of θ . Sampling from $\pi_{\theta,B|Y}(\theta, b)$ by the latent variable method is therefore also an $O(\log(n!))$ operation, which has been reduced from $O(n^2)$ by the direct method.

This conditional estimation approach has two further benefits. The first is that sampling B alongside θ provides a stochastic declustering of the catalogue as part of the model fitting process. This declustering allows the branching structure to be estimated while fully accounting for parameter uncertainty. Being able to estimate the branching structure then permits generalisations of the ETAS model that are based on its representation as a branching process. The second benefit of the approach is that parameter vector θ can be decomposed into the near-orthogonal blocks $\theta = (\mu, \theta_\kappa, \theta_h)$ that may be updated separately (Ross, 2016). These blocks are exactly orthogonal if all aftershock activity is within the observation period, meaning that $H(t_{\max} - \tau|\theta_h) = 1$ for all $\tau \in [0, t_{\max}]$ (Schoenberg, 2013).

6.3 Extreme value reparameterisation of empirical laws

6.3.1 Empirical laws

The functions κ , h and f are usually chosen to coincide with empirical ‘laws’ from the seismology literature (Hainzl and Christoffersen, 2017). The productivity function $\kappa(m_i; K, a)$, is frequently chosen to be an exponentially increasing function of magnitude above some fixed threshold magnitude u :

$$\kappa(m_i; K, a) = K \exp\{a(m_i - u)\} \mathbb{I}_{m_i > u}, \quad (6.3.1)$$

where $K, a \geq 0$ and \mathbb{I}_A is an indicator function of the event A . The threshold u is typically taken to be the magnitude of completion, above which all earthquakes are recorded in the catalogue. The modified-Omori law is typically used for the aftershock delay distribution $h(\tau; c, p)$. This is a heavy tailed, power-law distribution that decays more slowly than an exponential distribution:

$$h(\tau; c, p) = (p - 1)c^{p-1} \left(1 + \frac{\tau}{c}\right)^{-p} \mathbb{I}_{\tau > 0}, \quad (6.3.2)$$

where $c > 0$ and $p > 1$. The (truncated) Gutenberg-Richter law is used for the magnitude distribution $f(m; \beta)$. This is a shifted and truncated exponential distribution with rate parameter $\beta > 0$. The support of the distribution is truncated to be in the range (m_{\min}, m_{\max}) where $m_{\max} \leq \infty$, resulting in the density:

$$f(m; \beta) = \frac{\beta \exp\{-\beta(m - m_{\min})\}}{1 - \exp\{-\beta(m_{\max} - m_{\min})\}} \mathbb{I}_{m_{\min} \leq m \leq m_{\max}}. \quad (6.3.3)$$

There are two main issues that arise from using these empirical laws in the point process intensity model (6.1.1).

When using the modified Omori law (6.3.2), the delay distribution is constrained to be heavy tailed. This means that each event influences the intensity at all future times

and that this influence decays very slowly. The modified-Omori law is restrictive in that it does not allow the influence to decay rapidly or to have a finite extent. The second issue with using this empirical law is that its heavy-tailed nature can lead to a large proportion of the intensity from each event being placed outside of the observation window. This increases dependence between the ETAS parameter blocks $\{\mu\}$, $\{K, a\}$ and $\{c, p\}$ and makes inference more challenging.

There are also issues with the choice of the Gutenberg-Richter magnitude model. Firstly, if $m_{\max} = \infty$ then the fitted magnitude distribution will always be unbounded, which is not physically reasonable. Secondly, if $m_{\max} < \infty$ then there exists some greatest possible magnitude but the density is discontinuous there, because the distribution is constrained to have an exponential form. This form of constraint has been shown in other environmental applications to result in underestimation of the severity of the largest events (Coles et al., 2003). When a maximum magnitude is used, it is estimated through a separate extrapolation and the point estimate is used, ignoring crucial uncertainty about the events with greatest potential to cause damage (Kijko and Singh, 2011; Beirlant et al., 2019).

6.3.2 Proposed parameterisation

We propose alternative parametric forms for the functions κ , h , and f , which maintain or extend the class of functions which can be represented by the model. The proposed forms remove the restrictions placed on the shape and upper endpoint of both the delay and magnitude distributions, while dramatically reducing dependence within parameter blocks.

Productivity function. The productivity function κ links the magnitude of an event to the expected number of additional events that it triggers. In the latent

approach to inference, the parameters of κ are estimated conditional on knowing the branching vector B . As discussed in Section 6.2, when B is known the ETAS process can be decomposed into $n + 1$ independent, inhomogeneous Poisson processes. The observed event counts of these processes, $\{N_i = |\mathcal{C}_i| : i = 0, \dots, n\}$, are independent observations from a Poisson generalised linear model, where the linear predictor is a function of the parent magnitude.

With this interpretation, the productivity function in expression (6.3.1) can be rewritten as the generalised linear model where the parameters K and a are to be estimated: $N_i \sim \text{Poisson}(\kappa(m_i))$, where

$$\kappa(m; K, a) = e^{\log(K) + a(m - m_0)} = e^{(\log(K) - am_0) + am}. \quad (6.3.4)$$

This view of the productivity function highlights that the effect of magnitude on productivity is described relative to the threshold magnitude m_0 . The intercept term in model (6.3.4) involves both K and a , which can result in strong dependence between these parameters and poor mixing when MCMC is used to sample from their joint posterior. To alleviate this issue, we can centre the magnitude effect at the mean observed magnitude \bar{m} :

$$\kappa(m_i; K, a) = e^{\log(C) + a(m_i - \bar{m})}, \quad (6.3.5)$$

where $C = K \exp\{a(\bar{m} - m_0)\} > 0$ and $a > 0$.

In model (6.3.5) the interpretation of a is unchanged from the previous model but the parameter C now represents the expected aftershock productivity of a mean magnitude event. The parameters C and a have distinct interpretations in relation to the observed counts. This reparameterisation reduces dependence between the parameters in the same way as centring a linear model and should lead to more efficient MCMC sampling behaviour. One slight disadvantage is that the interpretation of C is now

catalogue dependent. However, sampled parameter values may be back-transformed onto the original parameter space to avoid this issue.

Delay distribution and Magnitude distribution. For both the delay distribution and the magnitude distribution, we propose the use of the generalised Pareto distribution. This two parameter distribution is commonly used in extreme value theory as an asymptotically motivated model for the distribution of exceedances of a threshold u (Coles, 2001). Under mild regularity conditions, the density function of a random variable $(X - u|X \geq u) \sim \text{GPD}_u(\sigma, \xi)$ is given by:

$$f_{X-u|X>u}(x - u; \sigma, \xi) = \begin{cases} \frac{1}{\sigma} \left(1 + \frac{\xi(x-u)}{\sigma}\right)_+^{-\frac{1}{\xi}-1} & \xi \neq 0, x \geq u; \\ \frac{1}{\sigma} \exp\left(-\frac{x-u}{\sigma}\right) & \xi = 0, x \geq u; \\ 0 & \text{otherwise;} \end{cases} \quad (6.3.6)$$

where $x_+ = \max(x, 0)$. The shape parameter $\xi \in \mathbb{R}$ controls the tail decay behaviour while the scale parameter $\sigma > 0$ describes the typical size of excesses. When $\xi > 0$, the distribution is a heavy tailed power-law distribution, when $\xi = 0$ the distribution is exponential, and when $\xi < 0$ the distribution has a finite upper endpoint at $x_{\max} = u - \frac{\sigma}{\xi}$. Expression (6.3.6) gives the standard parameterisation of the generalised Pareto distribution, in which the parameters σ and ξ are correlated. Chavez-Demoulin and Davison (2005) introduced a parameterisation of the generalised Pareto distribution using the alternative scale parameter $\nu = \frac{\sigma}{1+\xi}$:

$$f_{X-u|X>u}(x - u; \nu, \xi) = \begin{cases} \frac{1+\xi}{\nu} \left(1 + \frac{\xi(1+\xi)(x-u)}{\nu}\right)_+^{-\frac{1}{\xi}-1} & \xi \neq 0, x \geq u; \\ \frac{1}{\nu} \exp\left(-\frac{x-u}{\nu}\right) & \xi = 0, x \geq u; \\ 0 & \text{otherwise.} \end{cases} \quad (6.3.7)$$

Under this parameterisation, the parameters ν and ξ are orthogonal when $\xi > -1/2$ and if $\xi < 0$, the upper endpoint of the distribution is $x_{\max} = u - \nu[\xi(1 + \xi)]^{-1}$.

The modified-Omori and Gutenberg-Richter laws of models (6.3.2) and (6.3.3) can

each be seen as restricted forms of the generalised Pareto distribution with thresholds $u = 0$ and parameters (ν_t, ξ_t) and (ν_m, ξ_m) , respectively. The modified-Omori law of aftershock delay times is a power-law distribution and so is equivalent to the distribution $\text{GPD}_0(\nu_t = \frac{c}{p-1}(1 + \frac{1}{p-1}), \xi_t = \frac{1}{p-1})$, constrained so that $\xi_t > 0$. Similarly, when $m_{\max} = \infty$, the Gutenberg-Richter law is a special case of the $\text{GPD}_0(\nu_m = \frac{1}{\beta}, \xi_m)$ distribution constrained so that $\xi_m = 0$. In the case where $\xi_m < 0$ the GPD provides a similar model to the truncated Gutenberg-Richter law, but with the added benefit of the density function being continuous at m_{\max} .

The previously described empirical law models are all nested within the generalised Pareto model. This means that using a generalised Pareto distribution in place of each empirical law will allow a broader class of models to be represented and better represent epistemic uncertainty. Additionally, the orthogonal representation of the generalised Pareto distribution (6.3.7) is likely to improve inference because of more efficient MCMC sampling. The empirical laws being nested within the GPD model allows this to be demonstrated by imposing parameter restraints on the the more flexible model and comparing the sampling efficiency under each parameterisation.

6.3.3 Comparing parameterisations

In this section, we compare the properties of MCMC chains for estimating the ETAS model parameters under two parameterisations. As described in the previous section, the conventional parameterisation of the ETAS model $\theta_{\text{conv}} = (\mu, K, a, c, p)$ is based on empirical ‘laws’, uses a power-law distribution for the aftershock delay distribution and describes the effect of magnitude on aftershock productivity relative to a threshold magnitude. The novel proposed parameterisation $\theta_{\text{prop}} = (\mu, C, a, \nu_t, \xi_t)$ centres the effect of magnitude on productivity at the mean magnitude and uses the encompassing generalised Pareto distribution for aftershock delay distribution. To compare the

models fairly, the parameters of the GPD are constrained so that $\xi_t > 0$. This ensures that the two parameterisations represent the same class of models, namely those with heavy tailed waiting-time distributions. By imposing this restriction there is a one-to-one mapping g between the two parameterisations so that

$$\theta_{\text{prop}} = g(\theta_{\text{conv}}) = \left(\mu, K \exp\{a(\bar{m} - m_0)\}, a, \frac{c}{p-1} \left(1 + \frac{1}{p-1} \right), \frac{1}{p-1} \right)$$

and

$$\theta_{\text{conv}} = g^{-1}(\theta_{\text{prop}}) = \left(\mu, C \exp\{a(m_0 - \bar{m})\}, a, \frac{\nu_t}{\xi_t(1 + \xi_t)}, 1 + \frac{1}{\xi_t} \right).$$

We base our comparison of the two parameterisations on two properties of the MCMC chain: recovery of the true parameter values from a simulated catalogue and efficient exploration of the parameter space. Effective parameter recovery is indicated by the posterior mode being close to the true parameter values and the posterior being concentrated around that true value. Efficient exploration of the parameter space by an MCMC chain is indicated by a moderate acceptance probability during Metropolis steps and sampled parameter values having auto-correlation functions that decay quickly as the lag increases.

Parameter recovery can be evaluated visually by overlaying true parameter values on contour plots of the univariate or bivariate marginal posteriors of each parameter. Good parameter recovery is then indicated by the combination of tightly spaced contours and the posterior probability density being high at the true value. This can be summarised numerically by using the mean squared error (MSE) across all sampled parameter values from their joint posterior distribution.

The exploration of the parameter space or ‘mixing’ of an MCMC chain can be measured using the effective sample size of each parameter chain. The sampled parameter values obtained using a Metropolis-Hastings MCMC scheme are usually positively

auto-correlated; a chain of s sampled values therefore conveys less information about the posterior distribution than would s independent samples from that posterior. The effective sample size gives the number of independent samples to which the auto-correlated MCMC chain is equivalent and a greater effective sample size is therefore preferable (O'Hagan and Forster, 2004).

Let $\tilde{\theta}_i$ denote the s -vector of sampled values for the i^{th} parameter of the ETAS model under a particular parameterisation and represent the lag- j sample auto-correlation of $\tilde{\theta}_i$ by $\text{acf}(\tilde{\theta}_i, j)$. The effective sample size of the vector $\tilde{\theta}_i$ is then defined to be

$$\text{ESS}(\tilde{\theta}_i) = \frac{s}{1 + 2 \sum_{j=1}^k \text{acf}(\tilde{\theta}_i, j)},$$

where in this expression k is the lag beyond which the sum of all auto-correlations is negligible. The MSE and ESS are useful tools for comparing two MCMC schemes on the same parameter space but are not particularly useful when comparing MCMC schemes that are implemented on separate parameter spaces. In particular, it is not meaningful to compare the MSE for posterior samples of θ_{conv} and θ_{prop} because these distances are calculated on different parameter spaces. To avoid this issue, both chains can be transformed onto a single parameter space and the ESS can then be compared. When proposing a new parameterisation of an existing model, the comparison is often made on the parameter space of the original model by applying g^{-1} to posterior samples of θ_{prop} . However, when the mapping onto the original parameter space is non-linear, this transformation can alter the auto-correlation of the transformed chain and the resulting ESS of transformed chain.

To address the above issues we begin by comparing the chains on both the conventional and proposed parameter spaces. We additionally compare the chains using a third, more meaningful representation. This representation is a collection of physical properties of the earthquake process that would be measurable were the branching

structure known. We do this because when fitting the ETAS models it is usually physical properties such as these that are of interest, rather than the model parameters themselves. We consider the recovery and mixing of the collection of physical properties described in the following paragraph, each of which can be calculated from a vector of sampled parameter values under either parameterisation.

The first property we use is already present in each parameterisation: the seeding rate μ . The second property we consider is h , the distribution of aftershock delay times. We investigate h through its 10th, 50th and 90th percentiles, which we denote by h_{10} , h_{50} and h_{90} . The third property we consider is the magnitude-productivity relationship, κ . We consider this relationship through the expected number of aftershocks triggered by events with magnitudes equal to the 10th, 50th and 90th sample percentiles. We denote these expected aftershock counts by κ_{10} , κ_{50} and κ_{90} . This gives a vector of physically meaningful properties that we would like to be able to recover from an observed catalogue: $\theta_{\text{phys}} = (\mu, h_{10}, h_{50}, h_{90}, \kappa_{10}, \kappa_{50}, \kappa_{90})$.

6.3.4 Demonstration on simulated catalogue

6.3.4.1 Description of simulated catalogue

To compare parameterisations, we consider a simulated earthquake catalogue on the time window $[0, 2000]$ and magnitude range $[3, \infty)$. This catalogue has the conventional ETAS parameters $\theta_{\text{conv}} = (\mu, K, a, c, p) = (0.2, 0.2, 1.5, 0.5, 2)$ and magnitude parameters $\psi = (\nu_m, \xi_m) = (0.42, 0)$. The catalogue has $n = 863$ events and is displayed in Figure 6.3.1. The mean magnitude of the simulated catalogue is $\bar{m} = 3.43$ and so the simulation parameters in the proposed parameterisation are $\theta_{\text{prop}} = (\mu, C, a, \nu_t, \xi_t) = (0.2, 0.38, 1.5, 1, 1)$. The model parameters were estimated from this catalogue using a Metropolis-within-Gibbs sampling scheme as described in Section 6.3 with independent flat priors and each of the θ_{conv} and θ_{prop} parameterisa-

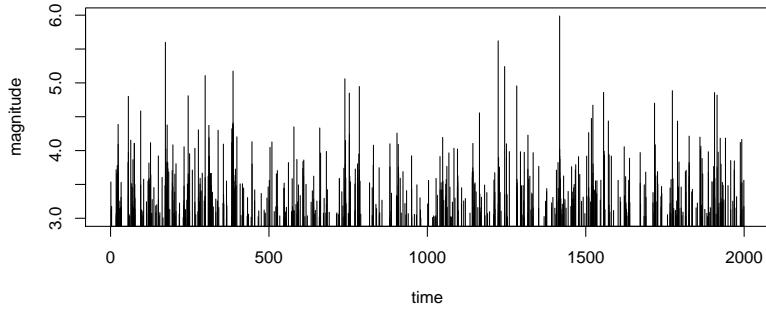


Figure 6.3.1: Simulated earthquake catalogue of 863 events used to compare θ_{conv} and θ_{prop} parameterisations of the ETAS model.

tions. In this section, we compare the resulting chains of sampled ETAS parameters and branching vectors from each parameterisation of the model.

6.3.4.2 Estimation of ETAS parameters

Marginal distributions. Figure 6.3.2 summarises, for each parameterisation θ_{conv} and θ_{prop} , the sampled values from the approximate joint posterior of the ETAS parameters. The plots on the diagonal show histograms of the univariate marginal posteriors, while the off-diagonal plots give contour plots and correlations for the pairwise marginal posterior distributions, where orange dots indicate the true parameter values. These plots reveal several important points.

Firstly, from the contour plots of Figure 6.3.2, we see that by using the θ_{prop} parameterisation the posteriors are better conditioned: the contours of pair-wise marginal distributions are elliptical around the modes and the posterior distributions are less concentrated at the edges of the parameter space. This is particularly apparent for the parameters of the delay distribution h . Secondly, the correlation within the parameter blocks controlling κ and h are both reduced by the reparameterisation. The correlations between the parameter blocks $\{\mu, \theta_\kappa, \theta_h\}$ have been increased, but recall that

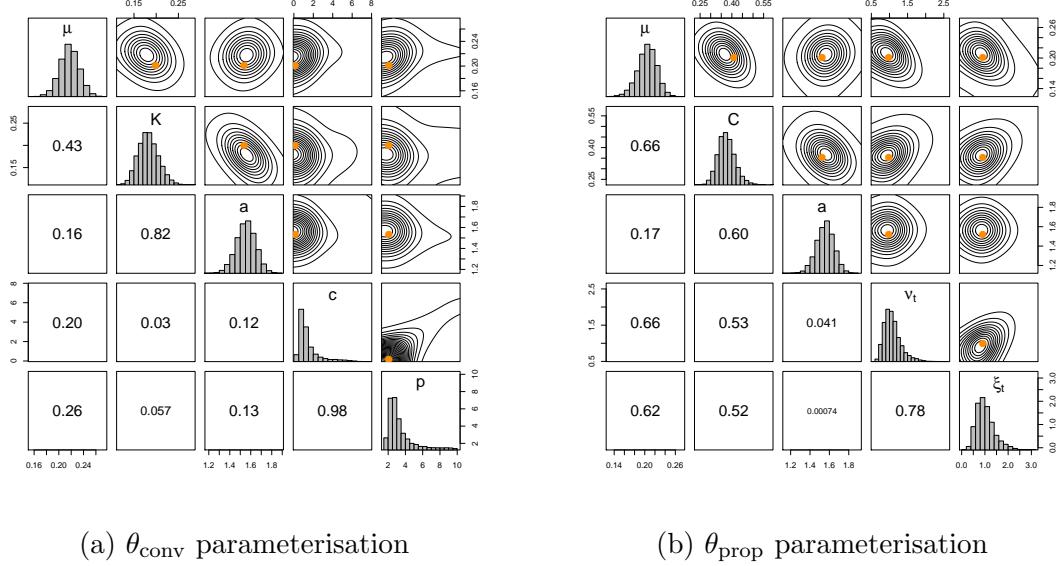


Figure 6.3.2: Contour plots of pair-wise marginal posterior distributions of ETAS parameters θ using the θ_{conv} and θ_{prop} parameterisations. True parameter values are indicated by orange dots.

these blocks of parameters are conditionally independent given the branching vector. Since these parameter blocks are updated conditional on both the other blocks and the branching vector, this should not impede effective inference. Finally, in the θ_{conv} parameterisation the posterior modes of the background event rate μ and the productivity intercept parameter K are misaligned with the true values; respectively these parameters are over- and under-estimated relative to their true values. This issue appears to be reduced for μ and C when the θ_{prop} parameterisation is used.

Within-block parameter recovery. Figure 6.3.3 focuses on the posterior distributions of θ_κ and θ_h under each parameterisation. As discussed in Section 6.3.3, each posterior is shown on both its original parameter space and transformed onto the alternative parameter space.

Under both parameterisations the intercept, a , and gradient terms, K or C , of the

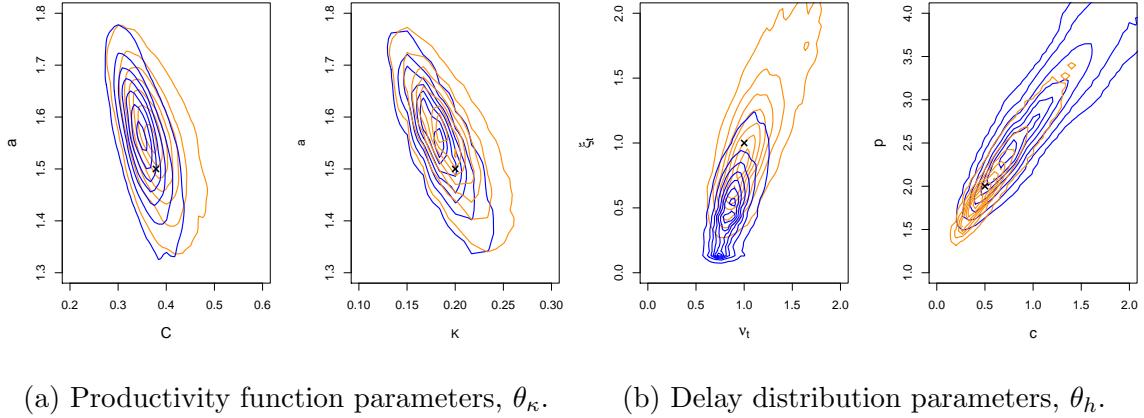


Figure 6.3.3: Estimated posterior distributions using samples from the θ_{prop} (orange) and θ_{conv} (blue) parameterisations. Both parameter spaces are shown and true values are given by black crosses.

magnitude-productivity relationship κ are overestimated. This suggests that this is a feature specific to this simulated catalogue. On both parameter spaces, the contours of θ_κ are wider and have a less steeply angled major axis for the θ_{prop} chain than for the θ_{conv} chain. These properties reflect, respectively, the better exploration of the parameter space of a and the reduced dependence between parameters when using the θ_{prop} chain. This results in the the posterior density being greater at the true value for the θ_{prop} chain than the θ_{conv} chain in both parameter spaces.

The parameters of the aftershock delay distribution h are poorly recovered by the θ_{conv} chain but successfully by the θ_{prop} chain, which has a posterior mode close to the true values in each parameter space. Each chain appears to explore its native parameter space more effectively than the transformed space, but the θ_{conv} chain does not well-explore heavy-tailed delay distributions (which correspond to large values of ξ_t or small values of p) as effectively as the θ_{prop} chain does in either space. This could be related to the over-estimation of the background rate μ by the θ_{conv} chain: long-term aftershock activity is being attributed to the background process because

RRMSE	μ	a	K	C	c	p	ν_t	ξ_t	h_{10}	h_{50}	h_{90}	κ_{10}	κ_{50}	κ_{90}
θ_{conv}	117	71	143	112	2954	1066	181	515	202	130	442	9	11	18
θ_{prop}	99	71	136	110	835	300	253	435	166	245	202	8	10	21

Table 6.3.1: Element-wise relative root mean squared errors ($\times 10^3$) of θ_{conv} , θ_{prop} and θ_{phys} for MCMC chains on the θ_{prop} and θ_{phys} parameter spaces.

heavy-tailed delay distributions are not being properly explored.

To compare parameter recovery across the joint distribution of all parameters, we calculate the relative root mean squared errors (RRMSE) of each chain on each parameter space. If $\{\tilde{\theta}^1, \dots, \tilde{\theta}^m\}$ is a set of m sampled s -dimensional parameter vectors from the posterior of θ , given data simulated with true parameter values θ^* , then the RRMSE is given by:

$$\text{RRMSE}(\{\tilde{\theta}^1, \dots, \tilde{\theta}^m\}, \theta^*) = \sum_{i=1}^s \sqrt{\frac{1}{m} \sum_{j=1}^m \left(\frac{\tilde{\theta}_i^j - \theta_i^*}{\theta_i^*} \right)^2}.$$

This metric gives a combined measure describing how close to, and concentrated about, the true parameter values are the sampled parameter vectors. Smaller values of the metric therefore represent more accurate and precise parameter recovery. On the θ_{conv} parameter space, the θ_{conv} chain had a RRMSE of 4.350 and the θ_{prop} chain had a RRMSE of 1.442. On the θ_{prop} parameter space, the θ_{conv} chain had a RRMSE of 0.995 and the θ_{prop} chain had a RRMSE of 0.969. In each parameter space the RRMSE is smaller for the θ_{prop} chain and so we conclude that overall, this chain is better able to recover the simulation parameters. The RRMSE values are decomposed into the contributions from each parameter in Table 6.3.1, from which it appears that most benefit comes from the improved estimation of the delay distribution h .

In addition to recovering the underlying model parameters, we would also like to be able to recover the physical quantities θ_{phys} , which would be measurable if the

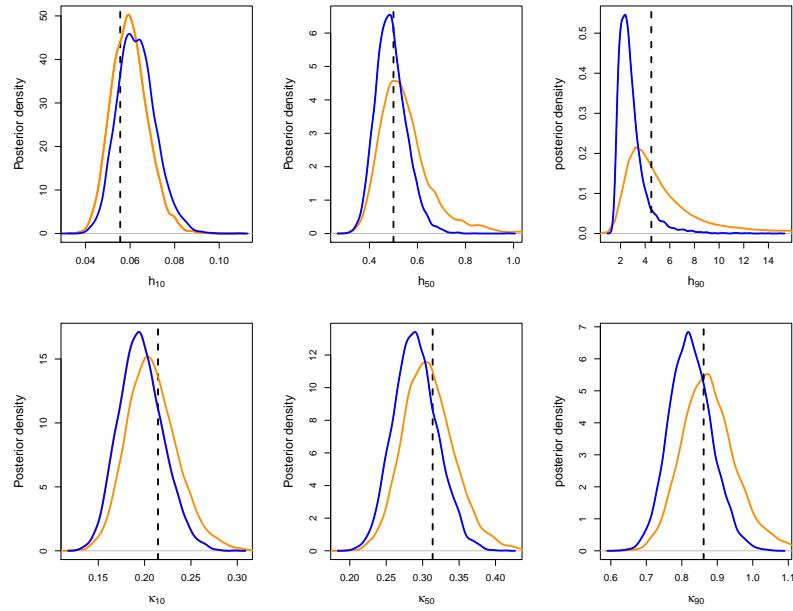


Figure 6.3.4: Posterior distributions of physically meaningful quantities θ_{phys} using estimated using samples from the θ_{prop} (orange) and θ_{conv} (blue) parameterisations. Upper plots show the 10th, 50th and 90th percentiles of the aftershock delay distribution. Lower plots show expected aftershock count at the 10th, 50th and 90th percentiles of the empirical magnitude distribution. True quantiles shown as dashed lines.

ESS	μ	a	K	C	c	p	ν_t	ξ_t	h_{10}	h_{50}	h_{90}	κ_{10}	κ_{50}	κ_{90}
θ_{conv}	1173	2451	1820	1573	516	478	551	413	1994	1103	419	1800	1635	1198
θ_{prop}	462	2212	827	593	813	622	343	372	2376	394	471	798	643	449

Table 6.3.2: Element-wise effective sample sizes of θ_{conv} , θ_{prop} and θ_{phys} for MCMC chains of 10,000 sampled values on the θ_{prop} and θ_{phys} parameter spaces.

true branching vector was known. Figure 6.3.4 shows the posterior distributions for quantiles of κ and h , based on the θ_{conv} and θ_{phys} chains, with the true values overlaid. The posterior modes for the θ_{prop} chain are closer to the true θ_{phys} values and, for all but the median of h , the posterior density is greater at the true value for the θ_{prop} chain. For all but h_{10} , the θ_{prop} posteriors have heavier right tails indicating that they better explore less optimistic regions of the parameter space, where triggered events form a greater proportion of the catalogue and can occur much later than their parent event. These effects are combined in the RRMSE values for each quantile, which are given in Table 6.3.1. According to the RRMSE values, the θ_{prop} chain better recovers all of the physical properties θ_{phys} except for h_{50} and κ_{90} .

Effective sample sizes. The effective sample size of each parameter in each parameter space is given in Table 6.3.2. For all parameters and all quantities, except h_{10} and h_{90} , the effective sample size of the θ_{conv} chain exceeds that of the θ_{prop} chain. This is surprising given that, based on the marginal plots considered previously, it appeared that the θ_{prop} chain was better exploring most parameter and quantile spaces. This might be explained by the ESS being a measure of *local* dependence within the chains; the θ_{conv} chain has weaker local dependence but this does not ensure the chain moves around the entire parameter space efficiently. The low ESS relative to chain lengths highlights the difficulty in efficiently sampling from the joint posterior of the ETAS parameters.

6.3.4.3 Branching vector recovery and mixing.

In addition to facilitating the estimation of the ETAS parameters θ , the conditional approach to inference also allows the branching structure B to be estimated as part of the inference procedure. The recovery of the true branching vector and proper representation of uncertainty in the estimated vector was not considered by Ross (2016). In this section, we introduce plots to diagnose branching vector recovery graphically and to assess effective mixing of the sampled chains of branching vector. We then use these to compare the performance of the θ_{conv} and θ_{prop} chains.

We begin by considering the marginal posterior of a single branching vector element B_i for a range of values of i . For a given event index i , this is a probability mass function (pmf) giving the posterior probability that event i was triggered by each previous earthquake, $\Pr(B_i = j)$ for $j \in \{1, \dots, i-1\}$, or comes from the background process (i.e. $\Pr(B_i = 0)$). The majority of probability mass will usually be allocated to recent previous events or the background process because the delay function h is monotonically decreasing over time. For large values of i there are many values of j (which have low probability mass) separating the most likely sources of event i , which can make reading the pmf difficult. To rectify this issue we therefore display $\Pr(B_i = 0)$ at i instead of at 0, so that the background process is the rightmost bar in the pmf. Figure 6.3.5 shows these posterior distributions along with trace-plots for a selection of events using each chain. The branching elements B_{51}, B_{81} and B_{199} are shown, which have true values of $b_{51} = 45$, $b_{81} = 0$ and $b_{199} = 0$.

From Figure 6.3.5 we can see that some events (such as event 199) are classified as a background event with high probability, while others (such as event 51) are more likely to be triggered events. Additionally, events which are labelled as triggered with a high probability might be attributable to a few recent events with high probability, as with event 81, or this might be split more evenly between many previous events, as

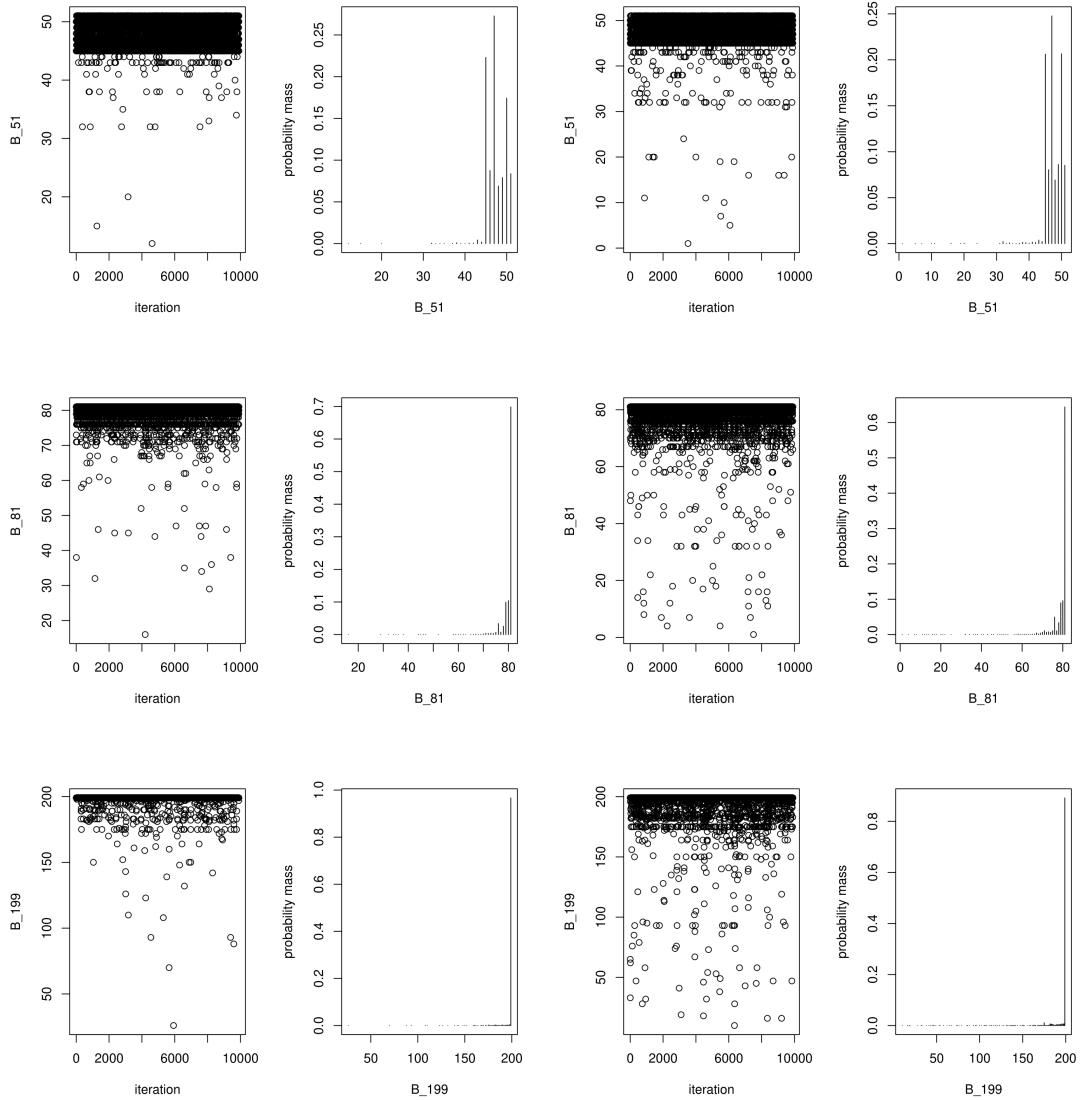


Figure 6.3.5: Trace plots and posterior pmfs for branching vector elements B_{51} , B_{81} and B_{199} . The θ_{conv} chain is shown on the left and the θ_{prop} chain on the right. The branching elements have true values of $b_{51} = 45$, $b_{81} = 0$ and $b_{199} = 0$.

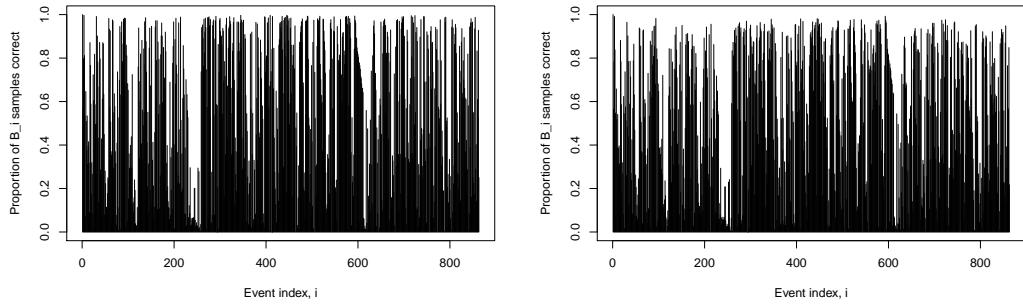


Figure 6.3.6: Proportion of sampled B_i values equal to the true value in the θ_{conv} (left) and θ_{prop} (right) chains.

with event 51. The interaction of the event magnitude and delay time can also be seen in these plots: considering event 51, we see that the most likely parent is not always the previous event and the triggering probability does not decrease monotonically with lag. Comparing the plots between parameterisations, it can be seen that for all three events, the θ_{prop} chain samples a greater proportion of b_i values for which $|i - b_i|$ is large. This again indicates that this chain better explores heavy-tailed distributions for h .

Since the catalogue is simulated, we know the true branching vector and can use this to assess the estimation of each element under each model parameterisation. To do this, we can consider the proportion of samples for which each branching element was correctly identified. This is shown in Figure 6.3.6. From this we can see that the source of each event is not equally easy to identify; some events are allocated correctly with high probability while others with very low probability. By considering the event-wise difference in the probability of correct allocation we can compare the two parameterisations; the correct parent is chosen on average 2% more often by the θ_{conv} chain, suggesting that this better recovers the true branching structure.

We can also assess the mixing of the branching vector chains by looking at the propor-

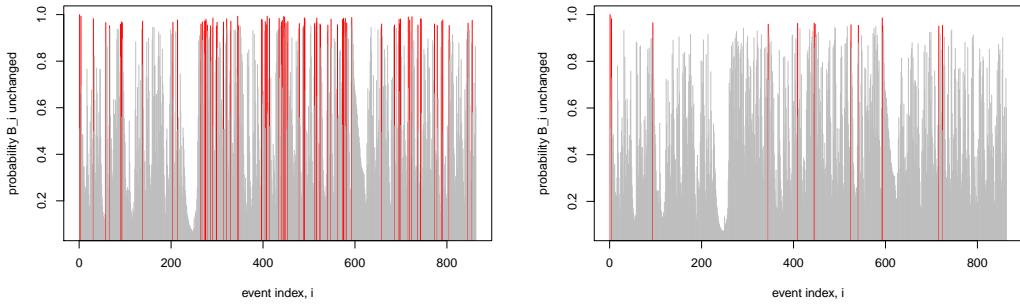


Figure 6.3.7: Proportion of updates at which each branching vector element was unchanged for the θ_{conv} chain (left) and θ_{prop} chain (right). Branching vector elements that change on less than 5% of updates are highlighted in red.

tion of updates of each branching element at which its value was unchanged. A high proportion of unchanged values indicates that either the chain is not mixing well or that the event is clearly attributable to a single source. Such events are highlighted in red in Figure 6.3.7, which shows the proportion of updates for which each branching element is unchanged for each of the θ_{conv} and θ_{prop} chains. By comparing these we can identify some common structures that are features of the data, such as the events around index 250 that cannot be attributed to a single source with high probability. This is contrasted by the number of highlighted elements, which is much lower for the θ_{prop} chain. This indicates that many of the elements of B were not mixing well using the θ_{conv} parameterisation, rather than it being the case that these events have a clear source.

Summary. Combining the previous findings we can see that the θ_{conv} chain recovers the true branching structure well. However, the chain does not mix as well as the θ_{prop} chain and does not reflect the full uncertainty about the branching structure. In this way it suffers, to a lesser extent, from the limitations of a deterministic declustering that were discussed in Section 6.1. This is likely linked to the chain not exploring

heavy-tailed aftershock delay distributions well. If heavy-tailed delay distributions are not being explored, then a greater proportion of probability mass will be allocated to branching elements corresponding to recent events or the background process. This in turn leads to overconfidence about which previous event was the parent event, overestimation of μ , and the consequent underestimation of K because of the negative correlation between these two parameters. This means that conclusions based on the θ_{conv} parameterisation risk under representing the possibility of long-term dependence in earthquake occurrences rates and overestimate the baseline level of seismicity.

Note that the conclusions here are based on a catalogue for which the delay distribution was truly a power-law distribution ($\xi_t > 0$). This class of distribution can be represented by both parameterisations and the less general θ_{conv} parameterisation can *only* represent these heavy-tailed distributions. The proposed parameterisation was constrained to this class of models to facilitate comparison. More generally the proposed parameterisation is also able to represent delay distributions that are exponentially decreasing over time or that have a finite upper end point (where $\xi_t \leq 0$). We have seen that the restricted θ_{prop} parameterisation better explores the space of heavy-tailed models. Further to this, the unconstrained θ_{prop} parameterisation is able to account for models that cannot be represented in the conventional framework. The unconstrained model would therefore represent model uncertainty more comprehensively in resulting the conclusions.

6.4 Extensions of the magnitude model

6.4.1 Dual magnitude extension

In the standard formulation of the ETAS model, introduced in Section 6.1, the marks associated with each event are assumed to be i.i.d. with a common probability density

function $f(m; \psi)$. Our first proposed extension to the magnitude model is to allow marks to be drawn independently from one of two possible distributions. Events from the background process have mark distribution $f_0(m; \psi_0)$; these are the events that are attributable to the intensity component $\lambda_0(t)$ and whose corresponding branching vector element is zero. All remaining events are triggered; they are attributable to one of the previous events and have a separate mark distribution that we denote by $f_1(m; \psi_1)$. The combined vector of magnitude parameters is now given by $\psi = (\psi_0, \psi_1)$ and must be estimated from the observed, marked point pattern together with the ETAS parameters and branching vector. Estimation of these parameters is simple when the branching vector B is known, since the distribution to which each magnitude belongs is also known.

In the i.i.d. mark formulation of the ETAS model, the mark distribution may be estimated independently of the intensity parameters θ and branching vector B . To fit the dual magnitude model, the magnitude parameters must be estimated jointly with the intensity parameters and the branching vector as ψ_0 and ψ_1 are not identifiable without knowledge of B . We therefore extend the Metropolis-within-Gibbs sampling scheme for the standard formulation to include the parameter blocks ψ_0 and ψ_1 ; each block of parameters $\{\mu, \theta_\kappa, \theta_h, B_1, \dots, B_n, \psi_0, \psi_1\}$ is updated in each iteration conditional on the values of all other blocks. Assuming independent, flat priors on the magnitude parameter blocks, the conditional posterior of the magnitude parameters is

$$\pi(\psi|Y, \theta, B) = \prod_{i:b_i=0} f_0(m_i|\psi_0) \prod_{i:b_i>0} f_1(m_i|\psi_1). \quad (6.4.1)$$

Under the dual magnitude model, the conditional log-likelihood (and therefore the conditional posterior) of the ETAS intensity parameters θ in expression (6.2.4) remains unchanged. With dual magnitude distributions, the elements of the branching vector B remain independent of one another, but the conditional posteriors are changed from

those in expression (6.2.5). This is because the observed magnitude m_i now carries information on whether $B_i = 0$ or $B_i > 0$. This modification leads to the magnitude-based weights in the conditional posterior of each branching vector element (6.4.2). These weights impart the additional information carried by the magnitudes and are obtained in Appendix C.1.

$$\Pr(B_i = b_i | Y, \theta, \psi) = \frac{i \cdot f_{\min(1, b_i)}(m_i | \psi)}{f_0(m_i | \psi_0) + (i - 1)f_1(m_i | \psi_1)} \frac{\lambda_{b_i}(t_i | \mathcal{H}t_i, t_i, \theta)}{\sum_{j=0}^{i-1} \lambda_j(t_i | \mathcal{H}t_i, t_i, \theta)}. \quad (6.4.2)$$

This dual magnitude model permits simple testing of the hypothesis that background and triggered events have different mark distributions through the use of Bayes factors. When f_0 and f_1 are nested, this becomes particularly simple because the standard ETAS formulation is then nested within the dual magnitude model. The evaluation can then be made directly using the values sampled from the posterior distribution of ψ .

6.4.2 Correlated magnitude extension

Our second proposed extension to the magnitude model allows for triggered magnitudes to be dependent on the magnitude of the event that triggers them. This dependence is modelled using a bivariate copula. This extension may be applied separately or together with the dual magnitude extension. In this section we outline the more general, combined model extension. In the following, we respectively denote the marginal distribution and probability density functions of background event magnitudes by $F_0(m; \psi_0)$ and $f_0(m; \psi_0)$, respectively. The corresponding functions for triggered events are denoted by $F_1(m; \psi_1)$ and $f_1(m; \psi_1)$. Correlated magnitudes may be introduced with a single distribution by fixing $f_0 = f_1$. We begin by describing the dependence model in detail before explaining how this alters the inference procedure from that of the dual magnitude extension. We introduce here an extension to the ETAS model in which magnitudes are not independent and where the magnitudes of

background and triggered events have separate distributions. A first-order Markov dependence structure is imposed on the graphical representation of the ETAS process. Under this dependence model, the magnitudes of events in the same tree are dependent on one another and the magnitudes of events in separate trees are independent. Additionally, events that are in the same tree but that are not adjacent to one another are conditionally independent given the magnitudes of their adjacent events: their ‘families’. Formally, the family of event $i = 1, \dots, n$ is indexed by the set $\mathcal{F}_i = \mathcal{C}_i = \{j \in \{i+1, \dots, n\} : b_j = i\}$ when event i is a background event and when event i is a triggered event then the family also includes the parent event of event i so that $\mathcal{F}_i = \mathcal{C}_i \cup \{b_i\}$.

To give a concrete example of this Markov dependence structure, we return to the toy example in Figure 6.2.1. In this example, the magnitude of event Y_2 is independent of the magnitudes of both of Y_5 and Y_7 because they belong to a different tree. The magnitudes of events Y_1, Y_3, Y_4 and Y_6 are all dependent on one another because they are in the same tree. However, given the magnitude of Y_3 is $M_3 = m_3$ the remaining magnitudes are pair-wise conditionally independent. This Markov dependence structure means that the joint density of all magnitudes may be written in terms of only marginal densities and bivariate joint densities. For the toy example we have that

$$\begin{aligned}
& f(m_1, \dots, m_7 | B, \psi) \\
&= f(m_1, m_3, m_4, m_6 | B, \psi) f(m_2 | B, \psi) f(m_5, m_7 | B, \psi) \\
&= \prod_{i \in \{1, 2, 5\}} \{f_0(m_i | \psi_0)\} f(m_3 | m_1, \psi) f(m_4 | m_3, \psi) f(m_6 | m_3, \psi) f(m_7 | m_5, \psi) \\
&= \prod_{i \in \{1, 2, 5\}} \{f_0(m_i | \psi_0)\} \frac{f(m_1, m_3 | \psi)}{f_0(m_1 | \psi_0)} \frac{f(m_3, m_4 | \psi)}{f_1(m_3 | \psi_1)} \frac{f(m_3, m_6 | \psi)}{f_1(m_3 | \psi_1)} \frac{f(m_5, m_7 | \psi)}{f_0(m_5 | \psi_0)}, \quad (6.4.3)
\end{aligned}$$

where the first two equalities follow from the independence between trees and conditional independence within trees. The final equality expresses the conditional densities

as the ratio of joint and marginal densities. Similarly, the conditional density of a single magnitude may also be written using only marginal densities and bivariate joint densities, as follows for M_3 in the toy example:

$$f(m_3|m_1, m_2, m_4, \dots, m_7, B, \psi) = f_1(m_3|\psi_1) \prod_{j \in \mathcal{F}_3 = \{1, 4, 6\}} \frac{f(m_j, m_3, \psi)}{f_1(m_j|\psi_1)}. \quad (6.4.4)$$

This Markov dependence structure can be achieved while preserving the marginal magnitude distributions of background and triggered events by using a copula to construct the joint distribution $f(\cdot, \cdot|\psi)$ of triggering and triggered magnitudes from their marginal distributions. A d -dimensional copula $C : [0, 1]^d \rightarrow [0, 1]$ is a multivariate distribution function with uniform margins; it is used to ‘couple’ or link $d \geq 2$ marginal distributions to give a joint distribution (Joe, 2014). The use of a copula allows the joint density of triggered and triggering magnitudes to be represented as the product of the marginal densities and the copula density. When $d = 2$ and the copula function C is parameterised by ρ , the copula density is given by $c(u, v; \rho) = \frac{\partial^2 C(u, v; \rho)}{\partial u \partial v}$. The joint density of two magnitudes M_i and M_j may therefore be expressed as

$$f(m_i, m_j|\psi, \rho) = c(F(m_i|\psi), F(m_j|\psi)|\rho) f(m_i|\psi) f(m_j|\psi),$$

where $f(m_i|\psi)$ is either $f_0(m_i|\psi_0)$ or $f_1(m_i|\psi_1)$, depending on whether event i is a background or triggered event, and $F(m_i|\psi)$ is used equivalently. The conditional distribution of M_i given M_j may be written as the product of the copula density and a single marginal density:

$$f(m_i|m_j, \psi, \rho) = c(F(m_i|\psi), F(m_j|\psi)|\rho) f(m_i|\psi).$$

The joint and conditional distributions (6.4.3) and (6.4.4) may therefore be expressed as the product of marginal densities and bivariate copula densities.

Here, we use a bivariate Gaussian copula to link the marginal distributions of triggered and triggering magnitudes. The Gaussian copula defines the dependence between

triggering and triggered magnitudes when each is transformed to have a standard Gaussian marginal distribution. The copula has a single parameter $\rho \in [-1, 1]$ that defines the correlation between these transformed variables. The effect of the parameter ρ on the copula can be understood by considering three cases: if the magnitude of a triggering event is in a high quantile of its distribution and $\rho > 0$ then the triggered magnitude is likely to also be in a high quantile of its distribution; if $\rho < 0$ then the triggered magnitude is likely to be in a low quantile; finally if $\rho = 0$ then the two magnitudes are independent. This choice of copula has the practical benefit that associated conditional distributions can be found in closed form and with relative ease. This copula also has the benefit of being able to represent independence, positive or negative association, or extremal dependence between large magnitude events depending on the value of ρ (Ledford and Tawn, 1997).

Incorporating Markov dependence into the ETAS model using a Gaussian copula extends the vector of magnitude parameters, which is now $\psi = (\psi_0, \psi_1, \rho)$. By fixing $\rho = 0$, event magnitudes can be made independent and the correlated magnitude ETAS model reduces to the dual magnitude extension.

Conditional posterior distribution of the magnitude parameters. To fit the correlated magnitude model, we again extend the Metropolis-within-Gibbs sampling scheme to include the correlation parameter ρ . To do this, we require the conditional posteriors under the correlated magnitude model for each of: the magnitude parameters, $\pi_{\psi|Y,B,\theta}$; the branching vector, $\pi_{B|Y,\psi,\theta}$ and the ETAS parameters, $\pi_{\theta|Y,B,\psi}$. To define these conditional posteriors, we introduce the indicator variables $d_i = \mathbb{I}\{b_i > 0\}$ for $i = 1, \dots, n$, which give the index of the magnitude distribution associated with each event. We also define $g_i = \Phi^{-1}(F_{d_i}(m_i))$ to be the magnitude of event i transformed to have a standard Gaussian marginal distribution, which has distribution function Φ and density function ϕ .

Appendix C.2 derives the conditional posterior of the magnitude parameter vector when using independent flat priors on each parameter. The resulting conditional posterior of the magnitude parameter vector is

$$\pi_{\psi|Y,B,\theta} = \prod_{i:b_i=0} f_0(m_i|\psi_0) \prod_{i:b_i>0} \left\{ \phi\left(\frac{g_i - \rho g_{b_i}}{\sqrt{1-\rho^2}}\right) \frac{f_1(m_i;\psi_1)}{\phi(g_i)} \right\}. \quad (6.4.5)$$

The joint conditional posterior in (6.4.5) allows all magnitude parameters to be updated as a block using a single Metropolis step within the MCMC sampling scheme. This block can be further decomposed to sample from the conditional posteriors of ψ_0 , ψ_1 and ρ sequentially. Sampling from these smaller parameter blocks can improve mixing in the Metropolis component of the MCMC scheme and (when particularly simple magnitude distributions are used) may permit direct sampling from the conditional posteriors.

Conditional posterior distribution of the branching vector. When event magnitudes are i.i.d., the conditional posterior $\pi_{B|Y,\theta,\psi}$ can be factorised into terms corresponding to each event. The elements of the branching vector remained conditionally independent with dual magnitude distributions, given the magnitudes of each event. However, when correlation between magnitudes is introduced, elements of the branching vector are no longer conditionally independent given only Y, θ and ψ . However, we note that during the Metropolis-within-Gibbs sampling for the dual magnitude model, each element of B is sampled separately. This does not require the conditional distribution of the full branching vector, only the distribution of $B_i|Y, B_{-i}, \theta, \psi$; where $B_{-i} = \{B_j : j \neq i\}$. In a similar way as in the earlier case when magnitudes were independent, we can find the conditional posterior of B_i but now also conditioning on all other elements of B . This distribution has a closed form which is given in Appendix C.3.

Conditional posterior distribution of the ETAS parameters. The conditional distribution $\pi_{\theta|Y,B,\psi}$ is unchanged by the introduction of correlated magnitudes. We now have closed forms for all required distributions to implement a Metropolis-within-Gibbs sampler for the joint posterior $\pi_{\theta,B,\psi|Y}$. In this scheme each parameter block is sampled from its conditional posteriors (either directly or using a Metropolis step) given the current value of all other parameter blocks. The block structure for the correlated, dual magnitude model is $\{\mu, \theta_\kappa, \theta_h, \psi_0, \psi_1, \rho, B_1, B_2, \dots, B_n\}$.

6.5 Application of extended ETAS models to simulated catalogues

6.5.1 Dual magnitude extension

Section 6.4.1 introduced the dual magnitude extension of the ETAS model and described a Bayesian approach to inference for this model extension. In this section we use a simulated catalogue with dual magnitudes to demonstrate that the proposed inference method is able to recover the true parameter values used in the simulation. We also consider a second simulated catalogue, which has i.i.d. magnitudes, to investigate how estimation of the branching vector B and the magnitude distribution f are impacted by the false assumption of either a single or dual magnitude distribution.

6.5.1.1 Simulated catalogue with dual magnitude distribution

We consider a catalogue of earthquakes simulated on the time-magnitude interval $[0, 50000] \times (3, \infty)$ which has true ETAS parameters $\theta_{\text{prop}} = (\mu, C, a, \nu_t, \xi_t) = (0.02, 0.2, 0.1, 0.1, 0.1)$ and magnitude parameters $\psi = (\nu_{m0}, \xi_{m0}, \nu_{m1}, \xi_{m1}) = (0.6, 0, 0.1, 0)$. The catalogue contains $n = 1205$ earthquakes of which 239 are triggered events

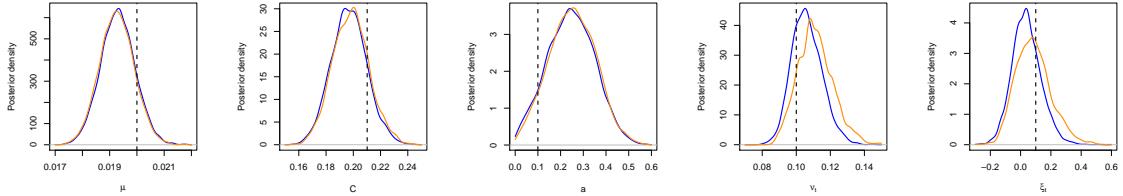


Figure 6.5.1: Marginal posterior distributions of ETAS parameters θ_{prop} given a dual magnitude earthquake catalogue, estimated using a dual magnitude model (blue) or single magnitude model (orange). True parameter values are shown as dashed lines - note that the true value of $\xi_{m0} = \xi_{m1} = 0$.

and where the mean magnitude is $\bar{m} = 1.99$. Both the dual magnitude and the single magnitude ETAS models were fitted to this catalogue using the θ_{prop} parameterisation, as outlined in Section 6.4; the former allows us to demonstrate recovery of the true parameters while the latter illustrates the bias that can occur from failure to account for dual magnitudes.

ETAS parameter recovery. Figure 6.5.1 shows the estimated marginal posterior distribution for each of the ETAS parameters θ_{prop} . For each of the ETAS parameters the true values are recovered well, falling within the 95% highest posterior density regions. Differences between true and estimated values of μ , C and a here are due to properties of the particular simulated catalogue and the marginal posteriors of these parameters are robust to mis-specification of the magnitude model. The parameters of the delay distribution, ν_t and ξ_t , are also well recovered from the simulated catalogue by both chains, but their marginal posterior distributions are more sensitive to mis-specification of the magnitude model.

Magnitude distribution recovery. Figure 6.5.2 shows the marginal and joint posterior distributions of the magnitude parameters estimated using a dual or single magnitude model along with the true values; for the dual magnitude model these

have been separated by event type. From these plots, we can see that the magnitude parameters for both the background and triggered events have been successfully recovered, as they are close to the posterior modes. We can also see that by falsely assuming a single magnitude distribution, the single scale parameter is estimated to be between those of the background and triggered distributions, while the single shape parameter is greater than that of either component distribution. Since the posteriors for background and triggered event magnitude parameters have very little overlap, we can conclude from these plots that the dual magnitude model is the more appropriate for this catalogue. While it is reassuring that we are able to recover the magnitude distribution of each event type, the aim of modelling is typically to estimate the combined distribution of all events. This combined distribution depends on each component magnitude distribution and also on the proportion of each event type.

In order to compare fitted and observed magnitude distributions we use return level plots for background, triggered and all magnitudes. The r -event magnitude return level m_r (with an associated return period of r events) is the quantile of the magnitude distribution that is exceeded in expectation once every r events. A return level plot shows, typically on a log-log scale, the return level against the return period. A return level plot allows assessment of model fit which emphasises the most extreme magnitudes by overlaying the fitted, empirical and (for simulated data) true return levels.

Figure 6.5.3 shows return level plots for the background event magnitudes, triggered event magnitudes and the combined magnitude distribution. Posterior mean and point-wise 95% credible intervals are shown for the fitted return levels along with the empirical values. Note that the empirical return levels for background and triggered events are obtained using the true branching vector, which is known only because

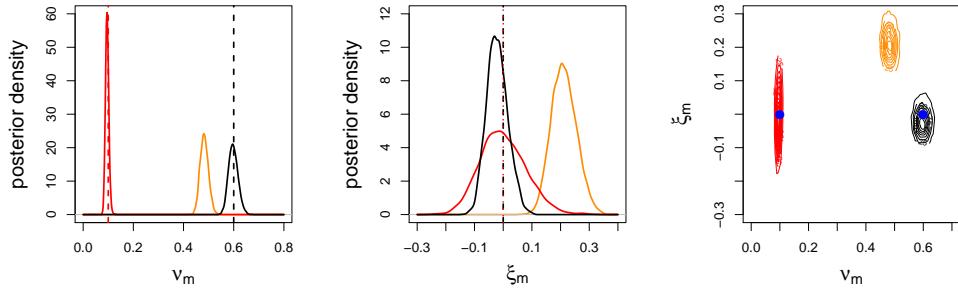


Figure 6.5.2: Marginal (left and centre) and joint (right) posterior distributions of magnitude parameters based on a simulated dual magnitude earthquake catalogue. Distributions are shown by event type under a dual or single magnitude model: distributions shown in black and red correspond to background and triggered event magnitude parameters in the dual magnitude model, (ν_{m0}, ξ_{m0}) and (ν_{m1}, ξ_{m1}) . Distributions shown in orange correspond to the mis-specified single magnitude distribution parameters (ν_m, ξ_m) . True parameter values are indicated by dashed lines / blue points.

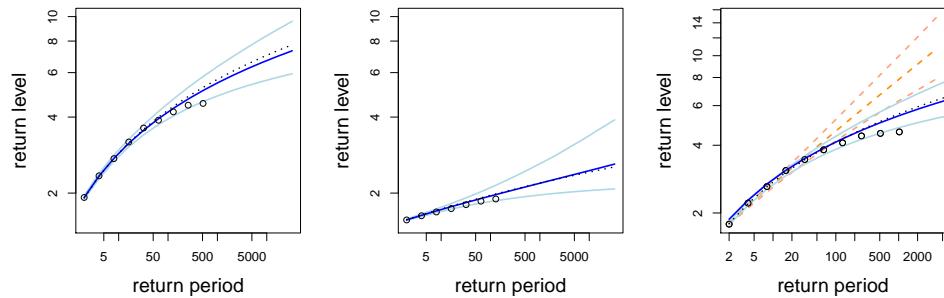


Figure 6.5.3: Return level plots for background magnitudes (left), triggered magnitudes (centre), and all magnitudes (right). Point-wise posterior means and 95% credible intervals for return levels are indicated by solid blue lines for the dual magnitude model and dashed orange lines for the single magnitude model. Empirical return level estimates are shown as points and true return levels as dotted lines.

this is a simulated catalogue. In recorded earthquake catalogues, only the combined empirical distribution would be known. In these plots we can see that for the dual magnitude posteriors, shown in blue, the empirical return levels are within the relevant credible regions and point estimates are close to the true values. This is not the case when a single magnitude distribution is falsely assumed. In that case, the fitted return levels of the combined magnitude distribution, shown in orange, are overestimated for events with long return periods; assuming a single magnitude distribution leads to inflation of the fitted return levels.

Inflated estimates for high return levels is not unique to this simulated catalogue, but are to be expected when fitting a single generalised Pareto distribution to data that truly comes from a mixture of generalised Pareto distributions. This effect can be linked to the fact that the generalised Pareto distribution is not sum stable; the weighted sum of two GPD random variables does not have a generalised Pareto distribution apart from in the trivial cases where all weight is allocated to one of the variables or where the parameters of the two distributions are equal. This means that, in general, the magnitude mixture distribution cannot be represented using the single magnitude model. This results in the parameter biases observed in Figure 6.5.2. The fitted scale parameter compromises between that of the background and triggered events, depending on their proportions and to compensate for this the shape parameter is overestimated. Since return levels are strongly driven by the shape parameter, particularly for long return periods, this leads us to the conclusion that falsely assuming a single mark distribution will inflate the estimated return levels. This reasoning suggests why this overestimation can be expected to occur across catalogues and gives further motivation for the use of our dual magnitude model extension, so as to avoid drawing overly conservative conclusions.

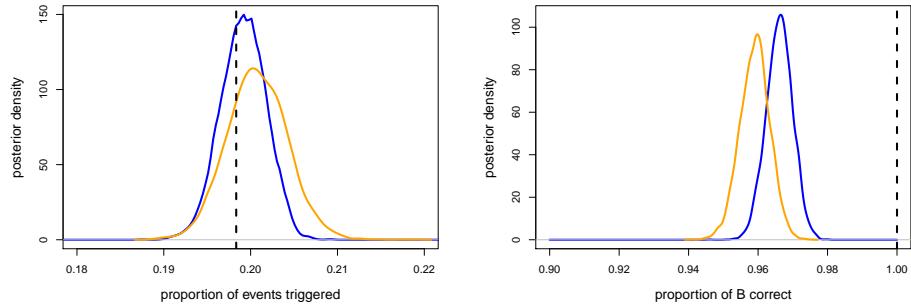


Figure 6.5.4: Posterior proportions of triggered events (left) and correctly allocated elements of B (right), based the dual magnitude (blue) and single magnitude (orange) models. Dashed lines represent perfect recovery.

Branching vector recovery. Figure 6.5.4 shows, for both a single and dual magnitude model, the posterior distributions for the proportion of triggered events and for the proportion of elements of B that are correctly recovered. This simulated catalogue has a low seeding rate and so has relatively few overlapping aftershock sequences. This means that a large proportion of events are correctly identified with high probability under both models - even when a single magnitude distribution is falsely assumed. Despite this, both properties of the branching vector are better recovered by the dual magnitude model, which has a posterior root mean squared error in the true proportion of triggered events of 0.30%, as compared to 0.42% for the single magnitude model. The dual magnitude model also better recovers the true proportion of triggered events: the posterior proportion of B correct has both a greater expectation and lower variance when using the dual magnitude model. We can also consider the recovery of the branching vector through the posterior probability that each branching element is allocated correctly. For events with different probabilities under each model, using the dual magnitude model leads to an average increase of 2.5% in the probability of correct allocation.

In all, we may therefore conclude that when there are truly two magnitude distri-

butions, we can effectively recover the ETAS, magnitude and branching parameters from a simulated catalogue using the proposed inference method. We have also shown that by acknowledging the two magnitude distribution within the inference procedure over-estimation of large return levels in the combined magnitude distribution can be avoided and estimates the branching structure can be improved.

6.5.1.2 Simulated catalogue with i.i.d. magnitudes

We also consider a simulated catalogue which has i.i.d. magnitudes. We aim show that when a dual magnitude model is used the parameters and distributions of magnitudes for background and triggered events are not significantly different from either one another or that obtained using the true, single magnitude model. We also investigate how falsely assuming a dual magnitude model impacts the estimation of the branching structure.

We use a catalogue that is simulated using the same parameters as in the previous simulated catalogue, but now with a single magnitude distribution with parameters $\psi = (\nu_m, \xi_m) = (0.4, 0)$. This simulated catalogue contains $n = 1206$ earthquakes of which 240 are triggered events and where the mean magnitude is $\bar{m} = 1.89$. Again, both the dual magnitude and single magnitude ETAS models were fitted to this catalogue using the θ_{prop} parameterisation. As can be seen in Figure 6.5.5, the ETAS intensity parameters were again successfully recovered under each model specification and so we focus attention on the magnitude model and the branching vector.

Magnitude distribution recovery. Figure 6.5.6 shows the marginal and joint posterior distributions of the magnitude parameters estimated under each model. There is substantial overlap between the posterior distributions of the magnitude parameters for background and triggered events, suggesting that the additional model complexity from allowing separate magnitude distributions is not necessary. Additionally, each

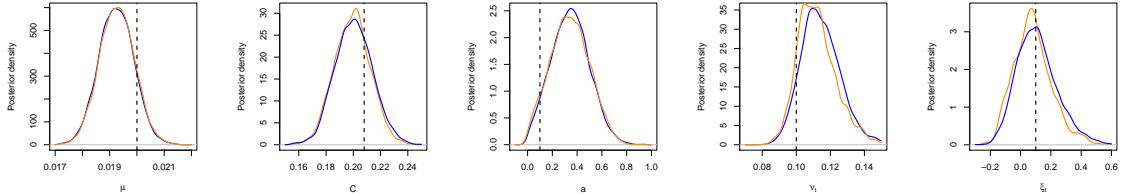


Figure 6.5.5: Marginal posterior distributions of ETAS parameters θ_{prop} given an earthquake catalogue with i.i.d. magnitudes, estimated using a dual magnitude model (blue) or single magnitude model (orange). True parameter values are shown as dashed lines.

of these posterior distributions has substantial overlap with that of the single magnitude model parameters, showing that we have still been able to recover the underlying magnitude distribution for this simulated catalogue, despite the magnitude model being over-parameterised and unnecessarily flexible. This comparison is formalised by Figure 6.5.7, which shows joint posteriors for the difference between parameters of the background magnitudes (ν_{m0}, ξ_{m0}) , triggered magnitudes (ν_{m0}, ξ_{m0}) and a single magnitude distribution (ν_m, ξ_m) . Since the origin lies within the 95% highest posterior density regions in each of these plots we have demonstrated that our ability to recover the magnitude parameters for each event type was not impaired by the model over-specification and that we can conclude that a single magnitude distribution is sufficient for this catalogue.

Figure 6.5.8 shows the estimated, empirical and true magnitude return level plots based on this single magnitude catalogue for background, triggered and all event magnitudes. Note that the credible intervals for background and triggered magnitude return levels would overlap if overlaid and that the point estimates and credible regions for the combined and single magnitude return levels are very similar. These features provide another route to the previous conclusion that a single magnitude distribution is sufficient and is not distorted by fitting the over-specified dual magnitude ETAS

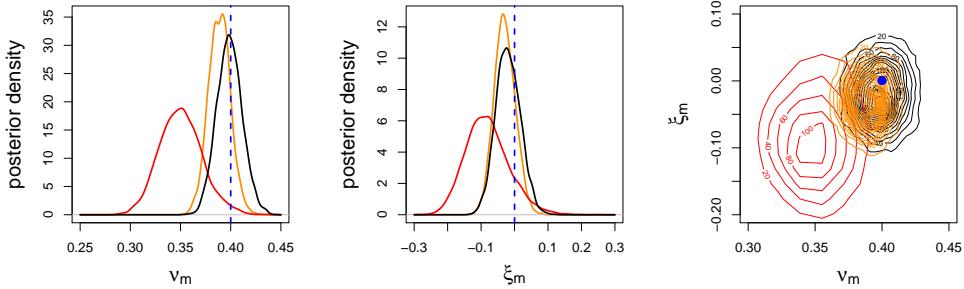


Figure 6.5.6: Marginal (left and centre) and joint (right) posterior distributions of magnitude parameters shown by event type under a dual or single magnitude model fitted to a catalogue with i.i.d. magnitudes. Distributions shown in black and red correspond to background and triggered event magnitude parameters in the over-specified dual magnitude model, (ν_{m0}, ξ_{m0}) and (ν_{m1}, ξ_{m1}) . Distributions shown in orange correspond to the single magnitude distribution parameters (ν_m, ξ_m) . True parameter values are indicated by dashed lines/ blue points.

model. This is in contrast to the previous conclusion from Figure 6.5.8, where falsely assuming a single magnitude distribution lead to overestimation of high return levels of the combined magnitude distribution.

Branching vector recovery. In Figure 6.5.9 we again consider the posterior proportion of events that are triggered and that are allocated to the correct parent process when using a single or dual magnitude model, this time for a catalogue with i.i.d. magnitudes. From this we can see that both models have very similar posterior distributions: fitting a model with two magnitude distributions does not detrimentally impact our ability to recover the branching structure. For the events which have a different probability of correct allocation under the two models, the expected increase in probability of correct allocation under the dual magnitude model is only 0.006%.

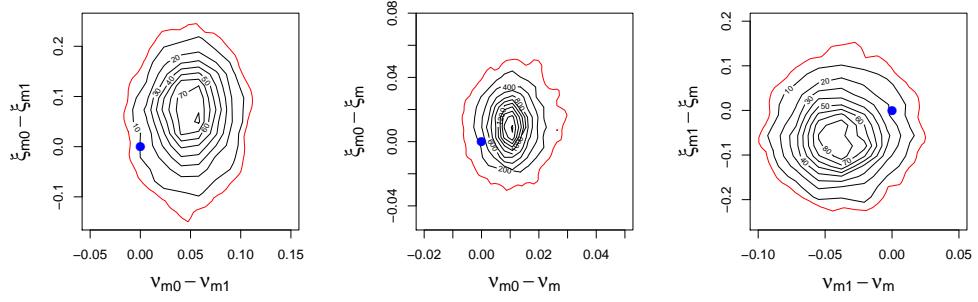


Figure 6.5.7: Joint posterior distributions of the difference in magnitude parameters for: (left) background and triggered events in the dual magnitude model, (centre) background events in the dual magnitude model and all events in the single magnitude model, (right) triggered events in the dual magnitude model and all events in the single magnitude model. 95% highest posterior density regions are shown in red and the origin is shown as a blue dots.

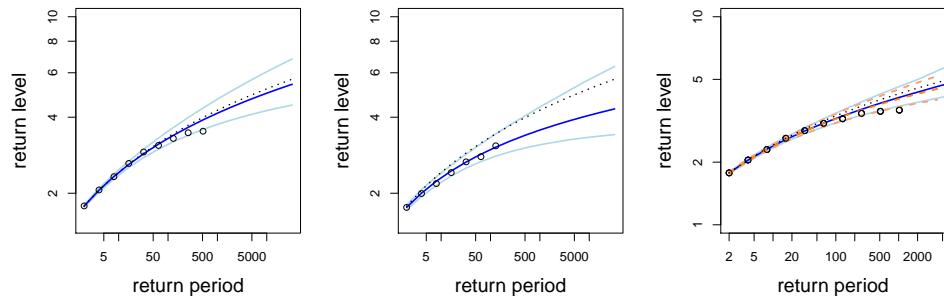


Figure 6.5.8: Return level plots for background magnitudes (left), triggered magnitudes (center), and all magnitudes (right) fitted using a catalogue with i.i.d. magnitudes. Point-wise posterior means and 95% credible intervals for return levels are indicated by lines: shown in solid blue for the dual magnitude model and dashed orange for the single magnitude model. Empirical return levels are shown as points and true return levels as dotted lines.

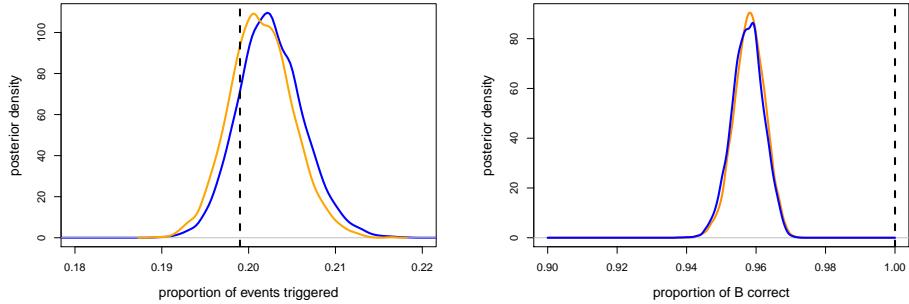


Figure 6.5.9: Posterior proportions of triggered events (left) and correctly allocated elements of B (right), based the dual magnitude (blue) and single magnitude (orange) models and a simulated catalogue with i.i.d. magnitudes. Dashed lines represent perfect recovery.

6.5.1.3 Summary

In this section we have demonstrated that our proposed inference procedure for the dual magnitude ETAS model is capable of recovering the ETAS, magnitude and branching parameters the proposed model extension. We have also shown that falsely assuming a single magnitude distribution inflated the estimated return levels for our simulated catalogue and have provided an outline argument for why this can be expected to occur more generally. We have shown that, conversely, the fitted magnitude distribution is robust to the false assumption of a dual magnitude model and also that simple testing procedures may be used to select between the dual and single magnitude ETAS models. In this case we used highest density posterior regions to inform our model selection but alternative approaches may have been used, for example Bayes factors or Bayesian model averaging (Gelman et al., 2013). Using these alternative approaches is less straightforward due to the conditional approach to inference; sampled joint posterior values are required but only conditional posterior values are calculated as part of the MCMC sampling scheme.

6.5.2 Correlated magnitude extension

Section 6.4.2 introduced an extension to the ETAS model in which the magnitudes of triggered events were allowed to depend on the triggering event magnitude through the introduction of a Gaussian copula. In this section, we demonstrate that the inference methods presented in Section 6.4.2 for the correlated ETAS model are able to recover the true values of the ETAS parameters, the magnitude parameters and the branching vector. We show that this is the case for a catalogue with either correlated or independent magnitudes. We also investigate how estimation of the branching vector B and the model parameters θ and ψ are impacted when magnitudes are assumed falsely to be independent.

In this section we use two simulated earthquake catalogues. The first simulated catalogue has independent magnitudes, so that $\rho = 0$, while the second has positive magnitude dependence with $\rho = 0.6$. We shall subsequently refer to these as the independent catalogue and the correlated catalogue. Both catalogues are simulated using the same ETAS parameters $\theta_{\text{prop}} = (\mu, C, a, \nu_t, \xi_t) = (0.02, 0.21, 0.1, 0.1, 0)$ and dual marginal magnitude distributions with parameters $\psi \setminus \{\rho\} = (\nu_{m0}, \xi_{m0}, \nu_{m1}, \xi_{m1}) = (0.6, 0, 0.1, 0)$. Both the independent catalogue and correlated catalogue are simulated on the time-magnitude interval $[0, 50000] \times (1.5, \infty)$.

The dual-magnitude and correlated-magnitude ETAS models are both fitted to each of simulated catalogues using the conditional Bayesian inference approach described in Section 6.4.2. The remainder of this section compares the resulting estimates of the model parameters and branching vector.

6.5.2.1 Simulated catalogue with dual, independent magnitudes

ETAS and marginal magnitude parameter recovery. Figure 6.5.10 shows the marginal posterior distributions of the ETAS parameters θ_{prop} given the indepen-

dent catalogue, obtained using the dual and correlated ETAS models. Figure 6.5.11 shows equivalent plots for the magnitude parameters ψ . From these plots we can see that both the dual magnitude model and the correlated magnitude model are able to recover the ETAS parameters and the parameters of the marginal magnitude distributions for background and triggered events. The posterior distributions are very similar under each model, which suggests that the greater flexibility afforded by the correlated ETAS model is not adversely impacting the estimation of the other model parameters.

The true correlation between triggered and triggering magnitudes in this catalogue is $\rho = 0$ and so the dual magnitude model recovers this parameter trivially. The posterior mode of ρ is close to zero under the correlated ETAS model, demonstrating that when the correlated model is applied to independent magnitudes it is capable of recovering that independence. Additionally, the posterior distribution of ρ using the correlated model is not entirely concentrated at 0, as it is under the dual model. This indicates that models with weak dependence (e.g. $\rho \approx 0.1$) are not implausible based on information conveyed by the independent catalogue. A risk-assessment using the dual magnitude model makes a hard assumption of independence and would therefore ignore all possibility of positive dependence, which can result in under-estimation of the largest earthquake magnitudes in the region and interval under assessment. A risk assessment based on the fitted correlated model would incorporate this possibility of mild positive dependence and should therefore be the preferred approach.

Branching vector recovery. Figure 6.5.12 shows plots summarising the recovery of the branching structure of the independent catalogue when using the dual and correlated ETAS models. The posterior distribution of the proportion of the branching vector that is correctly identified is similar under each model. Both models are able to recover the true proportion of events, with the dual magnitude model giving a lower

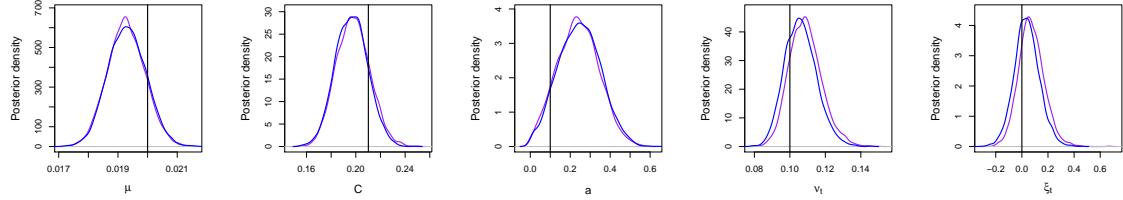


Figure 6.5.10: Marginal posterior distributions of ETAS parameters θ_{prop} given an earthquake catalogue with dual magnitudes, estimated using a dual magnitude model (blue) and correlated magnitude model (purple). True parameter values are shown by vertical lines.

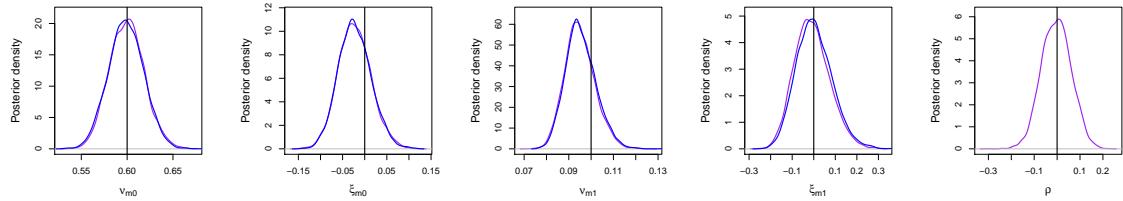


Figure 6.5.11: Marginal posterior distributions of magnitude parameters ψ given a earthquake catalogue with dual magnitudes, estimated using a dual magnitude model (blue) and correlated magnitude model (purple). True parameter values are shown by vertical lines.

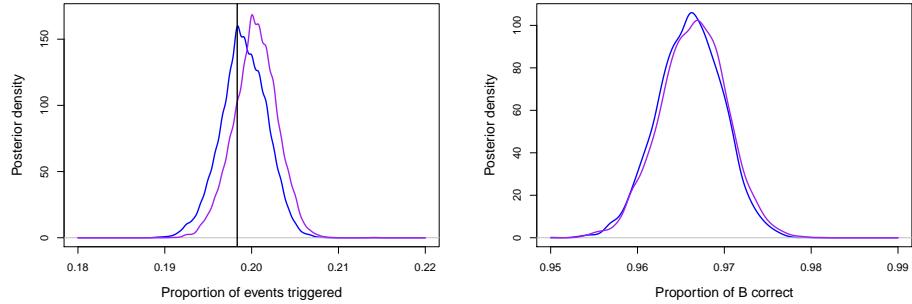


Figure 6.5.12: Posterior proportions of triggered events (left) and correctly allocated elements of B (right), based the dual magnitude (blue) and correlated magnitude (orange) models for a simulated catalogue with dual magnitudes. Vertical line represent true proportion of triggered events.

RMSE of 0.0030 compared to 0.0153 for the correlated model. This difference RMSE is due to slight over-estimation of the proportion of triggered events by the correlated ETAS model.

6.5.2.2 Simulated catalogue with dual, correlated magnitudes

ETAS and magnitude parameter recovery. Figure 6.5.13 shows the marginal posterior distributions of the ETAS parameters θ_{prop} given the correlated catalogue, under the dual and correlated ETAS models. Figure 6.5.14 shows equivalent plots for the magnitude parameters ψ . From these plots we can see that, as for the independent catalogue, both models are able to recover the ETAS parameters and the parameters of the marginal magnitude distributions for background and triggered events. There are some notable differences in the posteriors of ν_{m1} and ξ_{m1} between the two models, which we explore further using the ratio of root mean squared errors (RMSE) for each parameter under the correlated and dual models.

The elementwise ratios of RMSEs for the magnitude parameters are (1.03, 0.93, 1.03, 0.81, 0.07). This indicates that while the marginal scale parameters are slightly better

recovered by the dual model, this is more than offset by the marginal shape parameters, which are better recovered by the correlated magnitude model. The small value, 0.07, for the ratio of RMSEs for ρ should not be surprising; under the dual magnitude model ρ is effectively fixed at zero, making it impossible to recover the true correlation parameter $\rho = 0.6$. This should not discount the fact that the correlation parameter is accurately recovered by the correlated ETAS model with a RMSE of 0.044. As previously mentioned, falsely assuming independence of magnitudes can lead to overly optimistic assessments of the most extreme future magnitudes.

For the correlated catalogue, the RMSE of the marginal shape parameters is reduced when using the correlated (as opposed to the dual) ETAS model. This is because this model can fully utilise the additional information imparted by these correlations. To gain an intuition for this, consider (X_1, X_2) following a bivariate normal distribution with standard Gaussian margins and correlation ρ . When the dependence between X_1 and X_2 is ignored the standard deviation of X_1 is 1, but when the dependence is acknowledged this reduces to $\sqrt{1 - \rho^2}$. Acknowledging correlation reduces the standard deviation by a factor of $\sqrt{1 - \rho^2}$. Additionally, (Genest et al., 1995) showed that, when taking a step-wise approach to copula estimation by first estimating the marginal distributions, the estimated marginal distributions will be unbiased even when the copula is misspecified. Combining these two results, we might expect the RMSE of the marginal parameters, which combines the bias and standard deviation of their posterior distributions, to decrease by a factor of approximately $\sqrt{1 - 0.6^2} = 0.8$ when dependence in the correlated catalogue is acknowledged. Since there are two parameters in each marginal magnitude distribution, we cannot know *a priori* how this reduction will be attributed between those parameters. In this example, it appears that acknowledging dependence mainly benefits estimation of the shape parameters.

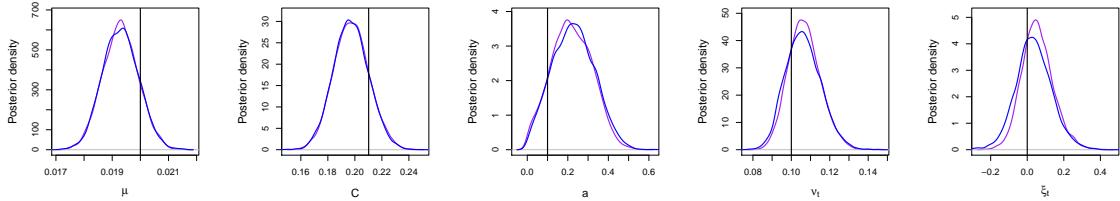


Figure 6.5.13: Marginal posterior distributions of ETAS parameters θ_{prop} given an earthquake catalogue with correlated magnitudes, estimated using a dual magnitude model (blue) and correlated magnitude model (purple). True parameter values are shown by vertical lines.

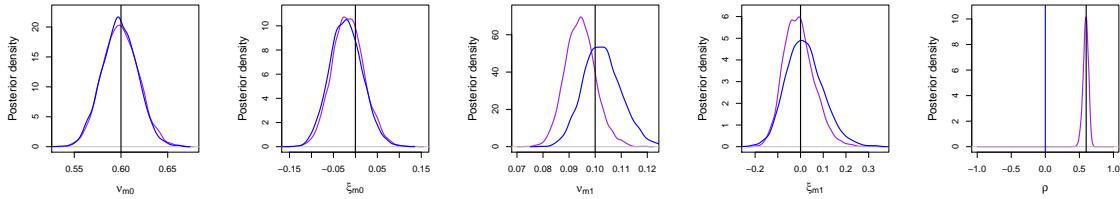


Figure 6.5.14: Marginal posterior distributions of magnitude parameters ψ given an earthquake catalogue with correlated magnitudes, estimated using a dual magnitude model (blue) and correlated magnitude model (purple). True parameter values are shown by black vertical lines.

The ratio of RMSEs for the 50th, 90th and 99th quantiles of the background and triggered magnitude distributions are (0.98, 1.03, 0.98) and (0.92, 0.93, 0.67) respectively. This indicates that acknowledging the correlation makes little difference to the estimation of low and moderate magnitude quantiles but greatly improves estimation of large aftershock magnitudes.

Branching vector recovery. In Figure 6.5.15 we consider again the posterior distributions for proportion of events that are triggered and the proportion that are allocated to the correct parent process; these distributions are examined when applying the dual or the correlated magnitude ETAS models to the correlated catalogue.

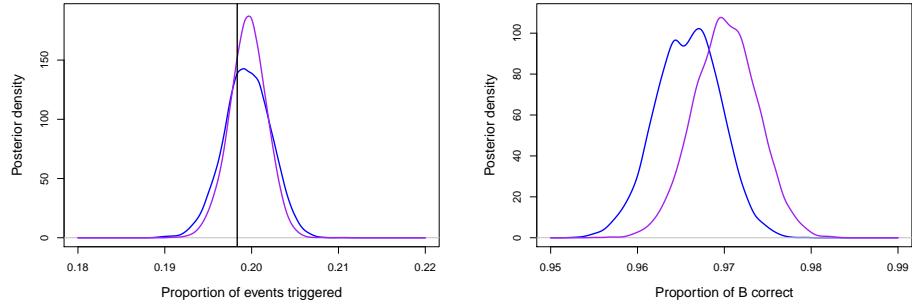


Figure 6.5.15: Posterior proportions of triggered events (left) and correctly allocated elements of B (right), based the dual magnitude (blue) and correlated magnitude (orange) models for a simulated catalogue with correlated magnitudes. Vertical line represent true proportion of triggered events.

Both models are able to recover the true proportion of triggered events, with the correlated model resulting in a smaller posterior variance and reduces the RMSE by a factor of 0.84 relative to the dual model, which falsely assumes that magnitudes are independent. Additionally, the expectation of the posterior proportion of the branching vector correct is increased when correlation is accounted for, while the variance of this posterior is slightly reduced. These effects combine to reduce the RMSE by a factor of 0.88 on the posterior proportion of B correctly allocated, as compared to the dual model. This demonstrates that the correlation between event magnitudes can be used to better estimate the branching structure of the point process, even when a large proportion of the branching structure can already be clearly identified when using the overly simplified dual model.

6.5.2.3 Summary

In this section we have demonstrated that, when magnitudes are truly independent, inference based on our correlated magnitude ETAS model is capable of identifying this. Testing for magnitude dependence is simple and equivalent to testing $\rho = 0$

against $\rho \neq 0$ under our correlated ETAS model. Additionally, we showed that fitting the correlated ETAS model to such a catalogue (rather than the true, dual magnitude model) does not prevent recovery of the ETAS parameters, marginal magnitude parameters or branching vector.

We have also shown that when magnitudes are truly dependent, our conditional Bayesian inference for the correlated ETAS model is capable of recovering the degree this dependence in addition to the ETAS, marginal magnitude and branching parameters. When magnitudes are truly dependent and independence is falsely assumed, then θ_{prop} , ψ_0 and ψ_1 may still be well estimated. However, making this false assumption is detrimental to the estimation of both ρ and B . This gives particular cause for concern. In simulated catalogues the true values of B and ρ are known but in recorded earthquake catalogues B and ρ are latent quantities, which must be inferred and cannot be measured directly. This makes validation of the independence assumption difficult and violations of the assumption more likely to go undetected.

Unlike the i.i.d. or dual magnitude ETAS models, the correlated ETAS model is capable of representing the possibility of weak magnitude dependence that cannot be excluded based the simulated dual-magnitude catalogue. Ignoring this potential dependence in a risk analysis can lead to underestimation of the true risk. (This is analogous to the underestimation of risk when assuming that a marginal magnitude distribution has a shape parameter exactly equal to zero when its credible interval contains both positive and negative values.) An assessment of risk based on the correlated magnitude model therefore more comprehensively reflects the epistemic uncertainty about whether earthquake magnitudes are truly independent.

6.6 Conclusions and further work

This chapter has focused on the Epidemic Type Aftershock Sequence model, which is used pervasively in the statistical modelling of both induced and tectonic earthquakes. In this final section, we summarise the contributions made by this chapter and outline potential areas for further work.

We first showed that the empirical laws that are conventionally used as components of the ETAS model are each nested within the generalised Pareto distribution. This led us to propose an orthogonal reparameterisation of the ETAS model in which each of these empirical laws is replaced by the encompassing GPD and the relationship between earthquake magnitudes and aftershock productivity is centred at the mean observed magnitude. This reparameterisation allows greater representation of model uncertainty by extending the representable class of ETAS models relative to the conventional parameterisation. Through simulation we showed that, even when constrained to the empirical-law model space, our reparameterised ETAS model leads to more efficient parameter inference, better exploration of heavy-tailed distributions and more accurate recovery of the latent branching structure.

The second contribution of this chapter was to introduce two extensions of the ETAS model, with dual and correlated magnitudes. These extensions were developed to ensure that their necessity over the standard ETAS model, with i.i.d. magnitudes, could be easily tested. Care was also taken to ensure that that a range of dependence strengths could be represented by our formulation of the correlated ETAS model (at both moderate and extreme values) and to ensure that the marginal magnitude distributions remained within the GPD family.

Efficient methods of conditional Bayesian inference were developed for both the dual- and correlated-magnitude ETAS models. We demonstrated that both the presence

and absence of these additional features can be identified through the use of simulated catalogues and considered, for the first time, the recovery of the latent branching structure of the ETAS model. Through these simulations we found that the additional model complexity does not impede inference when it is not required. Where non-identical margins or correlation are truly present, incorporating these into the inference lead to improved estimation of the branching structure and magnitude parameters.

There are many ways in which the new methodology and models presented in this chapter may be taken further. This chapter has focused on the temporal ETAS model with homogeneous background rate μ . This simplifies the exposition of our proposed methodology and models but this is likely to be overly simplified for direct use as a model for earthquake catalogues, as we saw in Chapter 2. A first extension might therefore consider a temporally varying background rate, estimated using a parametric or semi-parametric model. This would allow these models to be applied to induced earthquake catalogues, in which the seeding rate cannot reasonably be assumed constant because of changes in the human activity driving the seismicity. Care will be required to ensure that the background intensity model is sufficiently flexible to capture variations in the background event rate, but not so flexible as to obscure aftershock activity.

A second, and more challenging piece of further work would be to develop analogous extensions to the ETAS model in the spatio-temporal setting. This presents several challenges. Firstly, the dimensions of both the observation and parameter spaces are increased while the number of available data remains fixed. This could lead to less precise estimation of model parameters or require stronger prior distributions to be used to guide inference. Secondly, edge effects occur if additional intensity is allocated outside of the observation region and this is not accounted for during

inference. These edge effects have been shown to be highly influential on the estimated ETAS parameters (Schoenberg, 2013). These effects are easily handled in the one-dimensional temporal setting but calculating the amount of intensity allocated outside of a spatio-temporal region with an irregular spatial boundary would be a significant challenge.

A final piece of further work might consider how dual or correlated event magnitudes within the ETAS model influence the distribution of the largest magnitude in a given period, referred to as $M_{\max}(t_1, t_2)$ where $t_1 < t_2$ are the start and end of the period. This quantity is of particular interest in a seismological setting as it is often used to inform policy decisions relating to earthquake defences. Estimation of $M_{\max}(t_1, t_2)$ from an earthquake catalogue has been considered for i.i.d. magnitudes (Shcherbakov et al., 2019). Properties related to M_{\max} have also been considered for ETAS-type models used in financial applications; for example through the distribution of the sum or maximum of i.i.d. marks when the point process parameters are known (Bastrak et al., 2019; Žugec, 2019). What has not been established is how a mixture of magnitude distributions or dependence between earthquake magnitudes will impact these results. Theoretical or simulation based approaches to understanding this effect would be valuable further work.

Chapter 7

Conclusions and further work

In this final chapter, we summarise the contributions to the area of statistical seismicity that result from Chapters 4 - 6 of this thesis. We outline potential developments to each of these works individually and also comment on how these may be drawn together in future research.

In Chapter 4, we investigated simplifications and extensions to a state-of-the-art, physically motivated model for induced earthquake locations in the Groningen gas field, which we took as our baseline model. Model simplifications were motivated by a desire for parsimony in the statistical description of earthquake locations, while model extensions were proposed based on the addition of a range of physical characteristics that were not included in the baseline model.

From this work we found that, based on the available data, there was insufficient evidence to suggest that many of these model alterations provided a significant improvement to model fit. However, there was one notable exception to this finding: allowing the parameters of the baseline model to take different values in the upper and lower regions of the gas field resulted in a marked improvement to the baseline

model. In our model extension, spatial variability was permitted through a simple, linear partitioning of the gas field by dividing the field in such a way as to separate two modes of earthquake activity.

Although the improvement in model fit was not found to be overly sensitive to the choice of boundary line, the choice of a linear form for the boundary was arbitrary - as would have been any other division of the gas field that was not based on physical properties. Further work might investigate variations of the baseline model where parameters are allowed to vary smoothly over space, for example through the use of thin-plate splines. The challenge in successfully doing so will be balancing model flexibility against the small size of the Groningen earthquake catalogue.

In Chapter 5, we considered the development of the sensor network for detecting earthquakes in the Groningen gas field. In particular, we considered how investment in this network impacted the detection of small magnitude events and developed a method to select a time-varying threshold above which to model the magnitude distribution of earthquakes. This allowed small earthquakes, which would not have been detected at the start of the catalogue, to be included within the inference procedure thus reducing uncertainty in the estimated magnitude return levels as compared to a standard modelling approach.

When applying our threshold selection method to the Groningen earthquake catalogue we found that the constant threshold that is currently used as standard is overly conservative. Using the additional information provided by small magnitude earthquakes allowed us to conclude, for the first time based on empirical evidence alone, that the magnitude distribution of Groningen earthquakes has a finite upper end point.

This work and its conclusions have generated interest from stakeholders for the Groningen gas field. This interest lies both in the demonstrable return on investment this

gives for the developments to the sensor network and also in the implications about the magnitude return levels. The work presented in Chapter 5 prompted further work from Shell to investigate whether the measurement scale used to record earthquakes influences the selected threshold and resulting conclusions about the extreme value properties of the magnitude distribution. This work is inspired by Wadsworth et al. (2010), where Box-Cox transformations are used to represent the uncertainty in the measurement scale that leads to the most efficient extreme value analysis. Preliminary results from this further work suggest that the conclusions of Chapter 5, made on the local magnitude scale, are robust to moderate transformations of the chosen measurement scale.

There are many other possible avenues of additional further work leading from the methodology developed in Chapter 5. This work introduced a new method for selecting a time-varying threshold for a univariate extreme value analysis. This was motivated and explored in the case where observation is incomplete below the modelling threshold, but the method is applicable more generally. An important piece of further work would therefore be to perform an extensive simulation study to assess how the method compares to other approaches for selecting a time-varying threshold.

The method we have presented focuses the temporal evolution of earthquake detection. A second extension would be to develop a method to select a threshold that varies over space instead of or as well as over time, basing this on the same underlying principle of quantifying deviation from a fitted GPD model. This development is important since the spatial element in the development of the sensor network has not been accounted for in this first work.

In Chapter 6 we considered a reparameterisation and two extensions of the ETAS model for earthquakes and aftershock activity. We showed that inference for the

standard ETAS model could be improved while broadening the class of models that can be represented; this was achieved by centring the effect of earthquake magnitudes on aftershock productivity and by using generalised Pareto distributions in place of the empirical laws conventionally used to describe the aftershock-delay and magnitude distributions. We also introduced two extensions to the ETAS model: the dual magnitude extension, which allows background and triggered events to have distinct magnitude distributions, and the correlated magnitude extension which allows for dependence between triggered and triggering magnitudes. Each of these model extensions respects the complex branching structure of ETAS earthquake catalogues and efficient, conditional approaches to inference were developed for each extension.

A limitation in the current work is that we have restricted attention to the one dimensional case where background events come from a homogeneous Poisson process. The reparameterisation of the standard ETAS model could be extend readily to spatio-temporal point patterns with non-constant background rates, and could be of immediate practical benefit when fitting ETAS models to earthquake data from the Groningen field and beyond.

In order to be properly applied to catalogues of induced earthquakes, the model extensions must also be extended to accommodate a non-constant rate of background events. Translating the ETAS model extensions to this more general setting is conceptually simple, but will likely present practical difficulties in distinguishing between temporal (or indeed spatial) variability in the background event rate from clusters of aftershock activity. This same problem arises for the standard ETAS model when a flexible background intensity component is used and this identifiability issue is only exacerbated when catalogue sizes are small. Applying these model extensions to induced earthquake catalogues will likely be challenging without imposing strong assumptions on the form of the background event intensity to ensure sufficient smoothness in space

and/or time. Catalogues of tectonic earthquakes, where temporal stationarity may reasonably be assumed, might provide a suitable starting point for extensions of this type.

A final area of further work based on Chapter 6 would be to investigate the effects of a dual magnitude distribution or correlated earthquake magnitudes on the estimated values for the largest earthquake in a stated space-time interval, M_{\max} . This quantity is often of interest when planning earthquake defences and is dependent on all components of the ETAS model: the rate of background events, the recent event history, aftershock productivity, and of course the distribution and dependence structure of earthquake magnitudes. Because of these complex dependencies, M_{\max} is best estimated through the use of Monte Carlo simulations. It would be a worthwhile undertaking to conduct a thorough investigation on how estimates of M_{\max} are changed when, for example, dual or dependent magnitudes are truly present but are neglected in the modelling framework.

In addition to extending the work of each chapter of this thesis individually, there is also much further work that could be done to unite the methods and models developed in Chapters 4-6 of this thesis.

The work in Chapters 4 and 5 of this thesis could be combined to allow the use of later, small magnitude events when modelling earthquake locations. This additional information might materially alter the conclusions in Chapter 4 about which additional physical processes are detectable using the available data. The increased quantity of usable data might also facilitate the fitting of more complicated location models, such as the proposed extension with smoothly varying parameters. When combining location modelling with selection of a variable magnitude threshold, it will be important to proceed with great care because lowering the magnitude of completion will also increase the apparent rate at which events occur. A sensible starting point

for this work might be to first extend the models of Chapter 4 to consider the mark distribution as an additional dimension to the observation window and to allow the mark intensity to also vary according to a suitable covariate.

A limitation of the model extensions presented in Chapter 6 is the restriction of focus to the case where background events come from a homogeneous Poisson process with rate μ . This restriction was made to facilitate the presentation and implementation of these new model extensions, but in theory could be relaxed to allow parametric or semi-parametric modelling of the intensity of background events, such as those investigated in Chapter 4. This presents the additional challenge of selecting an appropriate level of smoothness in the intensity function of background events so that it may capture medium and long-term variation in the rate of background events without masking any short-term aftershock activity. A first attempt at this might assume a known background event rate, which could be estimated by spatially aggregating the fitted intensity of the baseline model from Chapter 4. Given this point estimate of baseline intensity, estimates of the ETAS, magnitude and branching parameters could be estimated from the observed catalogue. This might give a good starting point from which to relax the assumption that the background intensity is fixed and known.

Further work on modelling induced seismicity is not limited to modifications or extensions of the work presented in Chapters 4-6. The ETAS model is the dominant statistical model for aftershock activity, but when this model is used to augment a covariate-driven model of background events, such as those investigated in Chapter 4, it leads to inconsistent treatment of background and triggered events. The Poisson process describing background events is driven by the gas extraction process but the counts, locations and magnitudes of ETAS aftershocks are dependent on gas extraction only through the magnitude and locations of those background events. This is because

in the ETAS model earthquakes augment the intensity of the point process, rather than the covariate surface that is thought to be driving earthquake activity.

An alternative model formulation could be developed around a covariate-driven intensity model in which each earthquake may bring about further seismic events by locally altering the *covariate* values in the surrounding area and at subsequent times. This could reflect, for example, the additional stresses on the reservoir structure caused by that earthquake. The way in which the covariate surface is altered by seismic events should be chosen carefully, based on the form of the intensity model, to lead to a physically sensible local increase the covariate-driven intensity surface. Some initial ideas and discussions of this modelling approach are outlined Appendix D.1.

The main benefit of this type of combined model formulation for background and triggered events would be the increased coherence of the modelling framework for these two event types; each occurs when stresses across fault planes are increased beyond some critical value and this commonality should ideally be reflected in the point process model. However, such a combined model is not without its own challenges. The model would suffer from similar computational challenges to the direct fitting of the standard ETAS model, which could no longer be alleviated by a decomposition into simpler model components given a latent branching vector. There would also be the additional challenge of ensuring that changes to the covariates caused by earthquakes are coherent with subsequent, measured values of the covariates. These challenges could make implementation and application of such a combined aftershock model challenging.

At this point we hope that it is evident that the statistical modelling of earthquake activity is anything but straightforward. This modelling task is made yet more challenging in the setting of induced earthquakes, due to the small amount of available data and the variations in human activity that are driving earthquake occurrences.

This thesis has made several contributions to better our understanding and modelling of induced earthquakes, and toward overcoming some of the associated challenges.

As evidenced by the further work proposed in this chapter, many new questions and challenges about the modelling of induced earthquakes have been uncovered as the work in this thesis developed. In this way, statistical seismology remains an fecund, important and promising area for future research.

Appendix A

Supplementary materials to Chapter 4

A.1 Integrated intensity functions

Model	$\Lambda(B(x, \delta) \times (0, t) \boldsymbol{\beta}, \boldsymbol{\sigma})$
S1	$\beta_0 B(x, \delta) g(x; \sigma)$
S2	$\beta_0 B(x, \delta) s(x, t; \sigma)$
S3	$\beta_0 B(x, \delta) s(x, t; \sigma) [1 + \frac{\beta_1}{2} s(x, t, ; \sigma)]$
S4	$\beta_0 B(x, \delta) \beta_1^{-1} \exp\{\beta_1 s(x, t; \sigma)\}$
B0	$\beta_0 B(x, \delta) s(x, t; \sigma) \exp\{\beta_1 s(x, t; \sigma)\}$
E1	$(\beta_0 B(x, \delta) \cap W_L + \beta_1 B(x, \delta) \cap W_U) s(x, t; \sigma)$
E2	$\beta_0 B(x, \delta) \cap W_L s(x, t; \sigma) \exp\{\beta_1 s(x, t; \sigma)\} +$ $\beta_2 B(x, \delta) \cap W_U s(x, t; \sigma) \exp\{\beta_3 s(x, t; \sigma)\}$
E3	numerical integration required
E4	numerical integration required
E5	$\beta_0 B(x, \delta) s(x, t; \sigma)^\alpha \exp\{\beta_1 s(x, t; \sigma)^\gamma\}$
E6	$\beta_0 B(x, \delta) s(x, t; \sigma) \exp\{\beta_1 s(x, t; \sigma) + \beta_2 \dot{s}(x, t; \sigma)\}$

Table A.1.1: Integrated intensity functions for sub-models (S1-S4), the baseline model (B0) and model extensions (E1 - E6). The topographic gradient is denoted by $g(x)$, while $s(x, t; \sigma)$ denotes the cumulative incremental Coulomb stress smoothed using an isotropic Gaussian kernel with standard deviation σ . The first and second temporal derivatives of ICS are given by $\dot{s}(x, t; \sigma)$ and $\ddot{s}(x, t; \sigma)$. Regions W_L and W_U for models E1 and E2 are defined in Section 4.3.3.

A.2 Maps of annual expected earthquake counts

The following pages show maps of the annual expected earthquake counts on a 500×500 m grid over the Groningen gas field, under the fitted baseline model B0 and the fitted model extension E2.

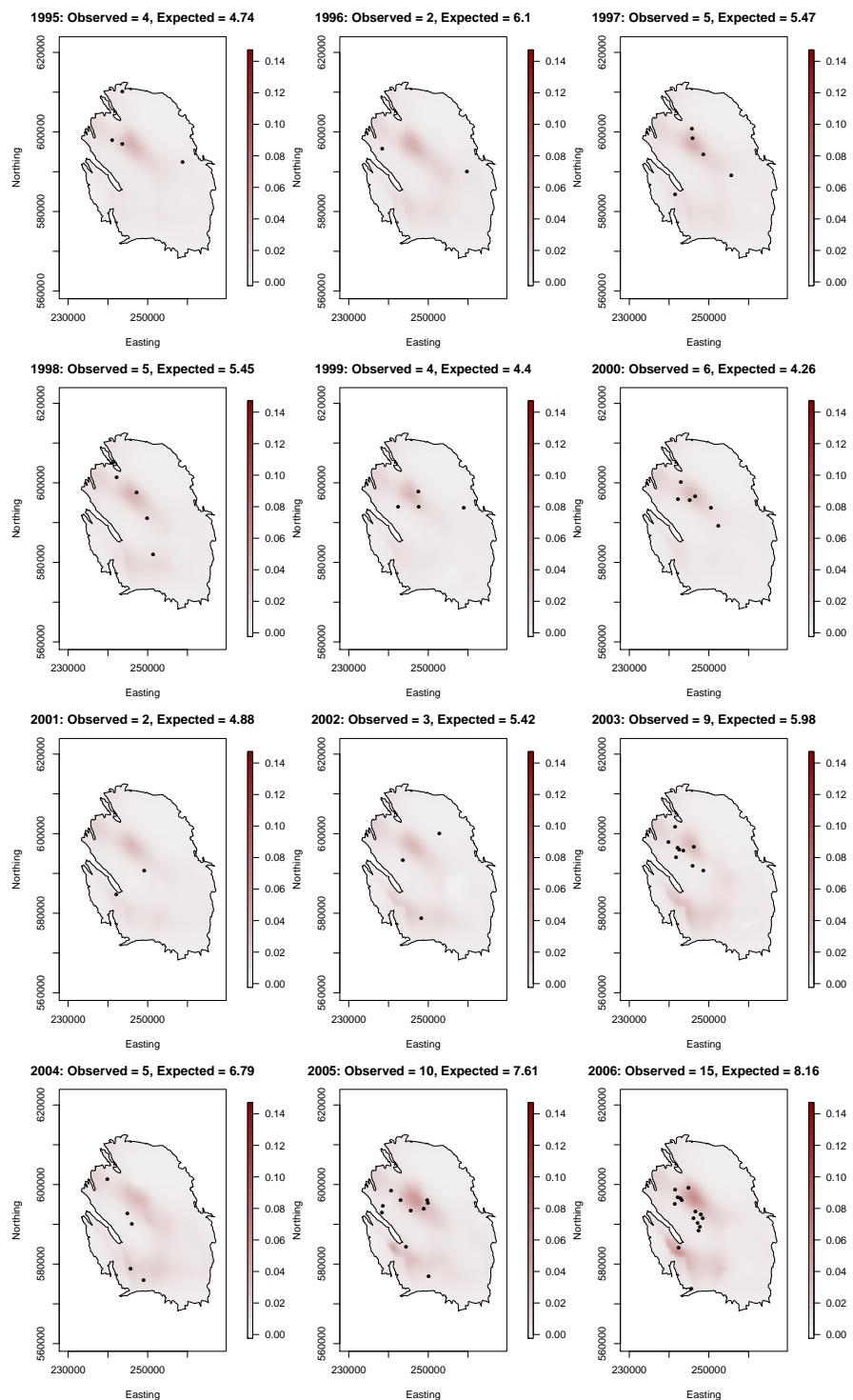


Figure A.2.1: Observed and expected event counts in each year (1995-2006) under the fitted baseline model B0.

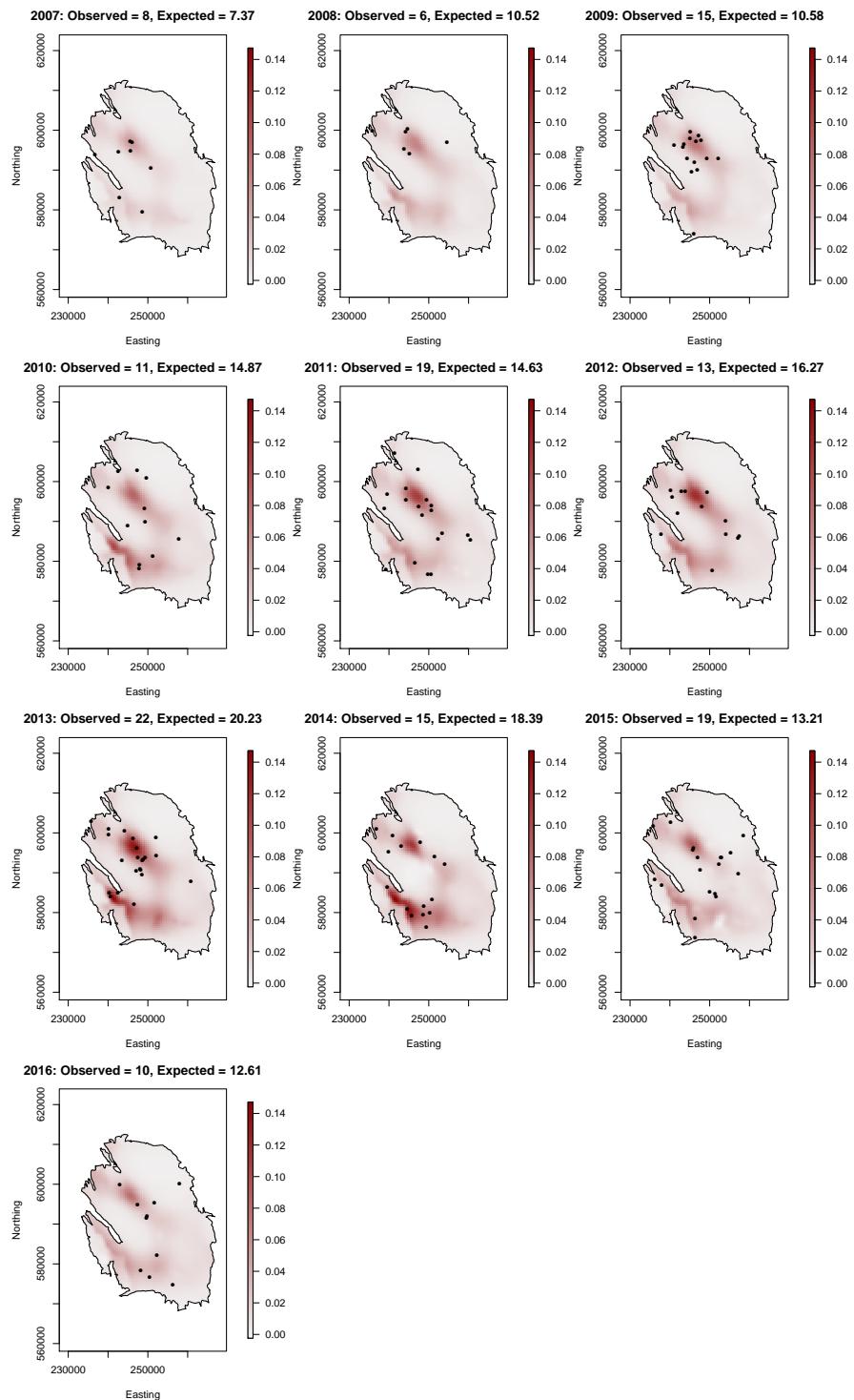


Figure A.2.2: Observed and expected event counts in each year (2007-2016) under the fitted baseline model B0.

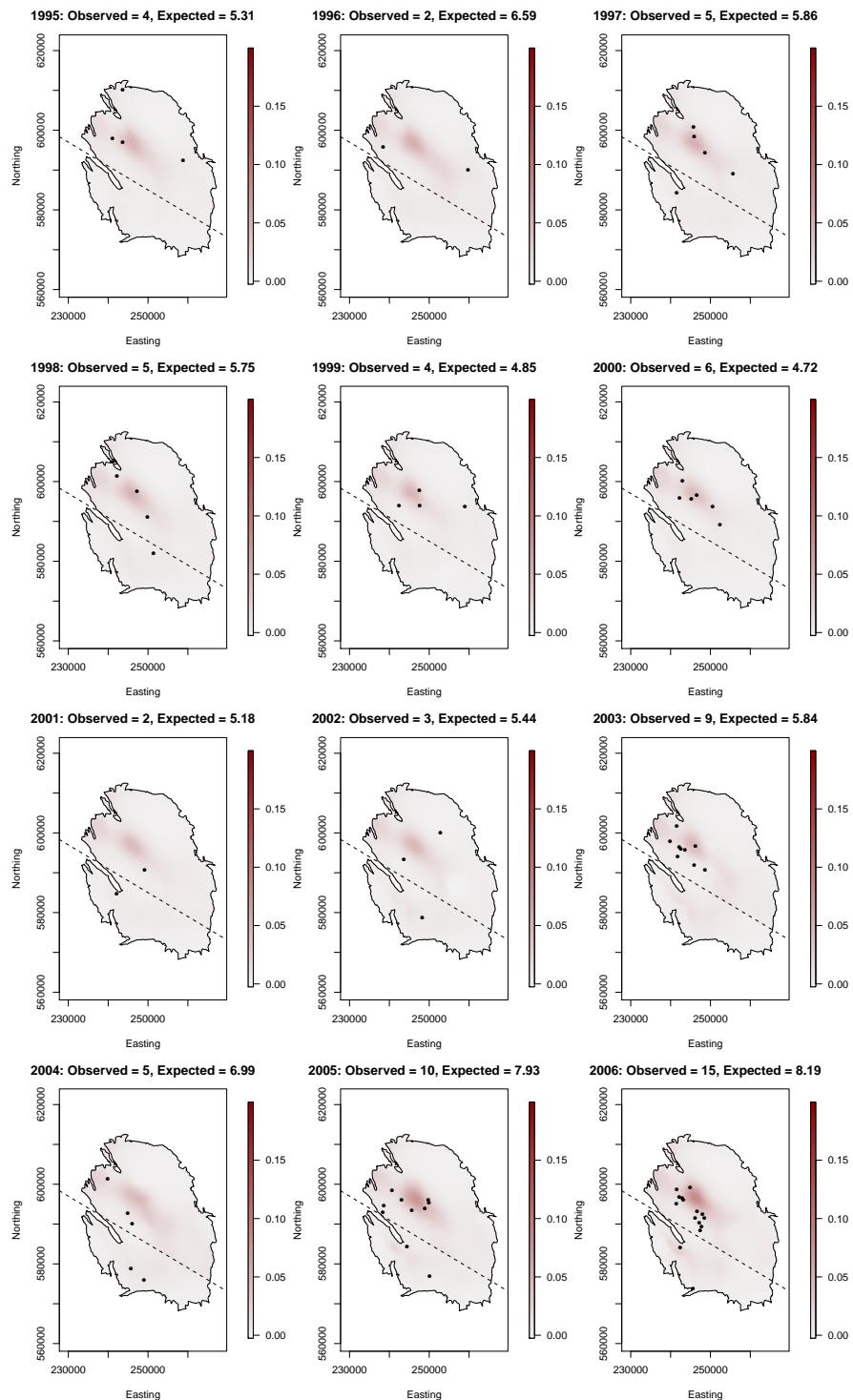


Figure A.2.3: Observed and expected event counts in each year (1995-2006) under the fitted model extension E2.

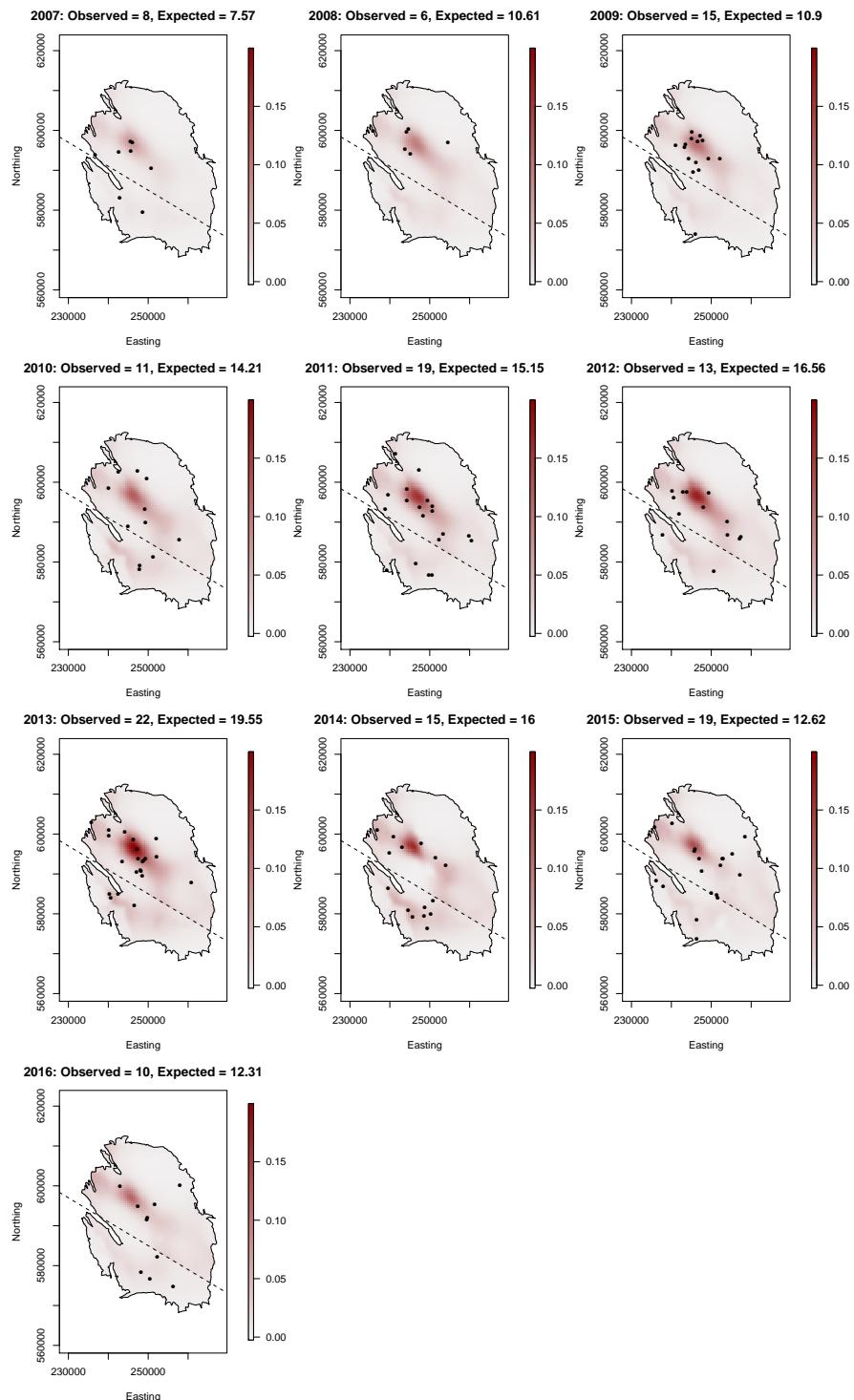


Figure A.2.4: Observed and expected event counts in each year (2007-2016) under the fitted model extension E2.

Appendix B

Supplementary materials to Chapter 5

B.1 Assessing i.i.d. assumption for Groningen earthquakes

B.1.1 Connection to main text

This appendix supports the claim made in Sections 5.2 and 5.6 of the main text that Groningen earthquakes exceeding $1.45M_L$ may be modelled as independent and identically distributed.

B.1.2 Exploratory analysis

Here we examine the validity of the assumption, common to both GPD and exponential models, that magnitudes are i.i.d. above $1.45M_L$. In the case of continuous-valued data that are completely observed, this assumption implies that inter-arrival times of threshold exceedances should approximately follow an exponential distribution. Due

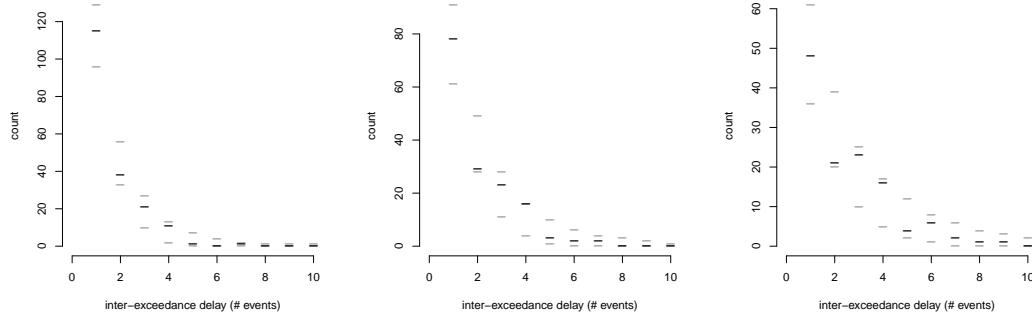


Figure B.1.1: Frequency plots of the number of earthquakes exceeding $1.45M_L$ that separate earthquakes exceeding $1.65M_L$ (left), $1.75M_L$ (centre), and $1.85M_L$ (right) for the Groningen earthquake catalogue. Observed frequencies (black lines) fall within the 95% confidence intervals under the fitted models (grey lines).

to incomplete observation below $1.45M_L$ and the transformed time scale, we instead consider the inter-arrival times of events exceeding magnitude $c \geq v_C$ measured in terms of the number of events with magnitudes between $1.45M_L$ and c . If events are i.i.d. then these inter-arrival times are geometrically distributed. It is important to investigate a range of values for c , since lower values lead to more (shorter) observed interval lengths but these are more concentrated about 0 making assessment of the geometric distribution more difficult. This trade-off can be seen by considering the edge-cases: if $c = 1.45M_L$ then each interval is of length 0, and if c is between the second and third largest observed magnitudes then there is a single observed interval.

The empirical distributions of interval lengths in the Groningen catalogue are shown in the panels of Figure B.1.1 for exceedances of $c = 1.65M_L$, $1.75M_L$, and $1.85M_L$. These are consistent with the 95% confidence interval for the fitted geometric distribution at each value of c . The same conclusion was found for $1.55M_L < c < 2.25M_L$. Events larger than this show mild evidence of clustering, but overall this suggests that it is reasonable to model magnitudes as i.i.d. above $c = 1.45M_L$.

B.2 Bootstrap datasets and parameter estimates

B.2.1 Relation to main text

This section supports the material presented in Section 5.2 of the main text. To represent the sampling variability in the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ associated with the log-likelihood (5.2.1) of the main text we take a parametric bootstrapping approach. This appendix describes the simulation of bootstrap catalogues and how they may be used to obtain bootstrap estimates of $\hat{\boldsymbol{\theta}}$. Bootstrap parameter estimates of this type are used throughout the main text.

B.2.2 Generating bootstrapped data-sets

B.2.2.1 Threshold exceedances and a point process

In bootstrap realisations of the earthquake catalogue, the number, timing and magnitudes of events exceeding $v(\tau)$ are all variable. In an alternate catalogue, events remain within the transformed observation interval $(0, \tau_{\max})$ but no longer form a regularly spaced sequence. Rather, the events in the region $A_v = \{(\tau, y) : 0 \leq \tau \leq \tau_{\max}, y \geq v(\tau)\}$, as shown in Figure B.2.1, are approximated by a Poisson process.

The Poisson process intensity on A_v is determined by the GPD parameters $\boldsymbol{\theta}$ and λ_u , where λ_u is the expected number of events exceeding $u \leq \min_{0 < \tau < \tau_{\max}} v(\tau)$ per unit τ (Coles, 2001). Since it is assumed that no censoring occurs on A_v and that the underlying magnitude distribution is identical over t , it follows that λ_u is constant over τ because time has been transformed so that earthquakes occur at a constant rate. The resulting intensity function on A_v is:

$$\lambda(\tau, y) = \frac{\lambda_u}{\sigma_u} \left[1 + \xi \frac{y - u}{\sigma_u} \right]_+^{1/\xi-1} \quad \text{for } (\tau, y) \in A_v. \quad (\text{B.2.1})$$

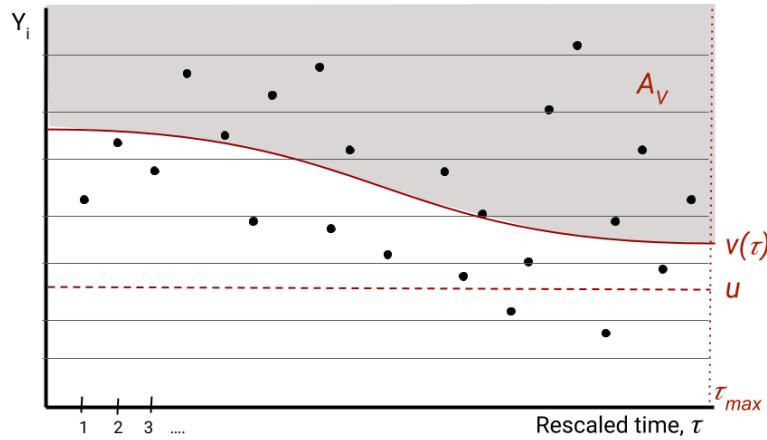


Figure B.2.1: Threshold exceedances as point process.

Bootstrap catalogues of earthquakes with magnitudes exceeding $v(\tau)$ can be obtained as realisations from the point process with intensity function (B.2.1) using the estimated parameters $\hat{\theta}$. We first describe how to generate such bootstrap catalogues and then how these can be used to obtain bootstrap estimates of $\hat{\theta}$.

B.2.2.2 Simulating the exceedance count

The first step in generating a bootstrap catalogue is to sample the number of events that occur on A_v . The number of events on A_v is Poisson distributed with expectation

$$\Lambda(A_v) = \int_{A_v} \lambda(\tau, y) d\tau dy, \quad (\text{B.2.2})$$

which must be estimated from the observed catalogue. This is complicated by the rounding of observed magnitudes \mathbf{x} to the nearest 2δ . Recall that for borderline events $\{x_i \in \mathbf{x} : |x_i - v_i| < \delta\}$ it is not known whether or not the corresponding unrounded magnitudes exceed $v(\tau)$ and place those events on A_v . Therefore n_v , the observed event count on A_v , is unknown and must itself be estimated.

Given the rounded magnitudes $\mathbf{x} = (x_1, \dots, x_n)$ and the estimated GPD parameters

$(\hat{\sigma}_u, \hat{\xi})$, events $i = 1, \dots, n$ each exceed their magnitude threshold and are on A_v independently with probability w_i as defined in expression (5.2.2) of the main text. Uncertainty in the observed event count due to magnitude rounding can be included in the generation of bootstrap sample sizes by simulating a value m_v for n_v , as the sum of n independent Bernoulli random variables with expectations w_1 to w_n .

This simulated value for the observed event count can be used as a point estimate $\hat{\Lambda}(A_v) = m_v$ for $\Lambda(A_v)$. Since $\hat{\Lambda}(A_v)$ is an estimate of based on a single observed count, it is important to also include uncertainty in the inferred value of $\Lambda(A_v)$ when generating bootstrap catalogue sizes. This can be done using a bootstrapped value for the value of the estimator $\hat{\Lambda}(A_v)$. Since inference is based on a single Poisson count, the bootstrap estimator $\tilde{\Lambda}(A_v)$ is obtained simply as a sampled value from the $\text{Poisson}(\hat{\Lambda}(A_v))$ distribution.

Finally, the event count \tilde{n} to be used for the bootstrapped catalogue can be sampled from a $\text{Poisson}(\tilde{\Lambda}(A_v))$ distribution. Doing so properly represents rounding, estimation and sampling uncertainties in the bootstrapped event counts.

B.2.2.3 Simulation of event times

Having simulated the exceedance count \tilde{n}_v , the times $\tilde{\tau} = (\tau_1, \dots, \tau_{\tilde{n}_v})$ of the events on A_v can sampled according to the marginal temporal intensity $\lambda(\tau)$. The marginal temporal intensity is found by integrating the joint intensity (B.2.1) over magnitudes. Noticing that $\lambda(\tau, y)$ is proportional to the GPD density, $\lambda(\tau)$ can be stated in terms of the GPD survivor function $\bar{F}(y; \sigma, \xi) = [1 + \xi y / \sigma]_+^{-1/\xi}$ to give:

$$\lambda(\tau) = \int_{v(\tau)}^{\infty} \lambda(\tau, y) dy = \lambda_u \bar{F}(v(\tau) - u; \sigma_u, \xi) \quad \text{for } 0 \leq \tau \leq \tau_{\max}. \quad (\text{B.2.3})$$

Sampling exceedance times from this intensity can be achieved through reverse application of the time-rescaling theorem (Brown et al., 2002). In general, this will require

numerical integration and can be computationally intensive. Depending on the form of $v(\tau)$ more efficient methods may be available. When $v(\tau)$ is a step function the step-wise integral of the temporal intensity has a simple form and event times can be simulated very efficiently; the \tilde{n}_v events are allocated independently to steps with probability proportional to the temporal intensity integrated over each step. Events are then located uniformly at random within their allocated step. When a more complex form is used for $v(\tau)$, alternative sampling approaches may reduce the computational cost of sampling event times. For example, exact samples may be obtained through rejection sampling and approximate samples by approximating $v(\tau)$ as a step function.

B.2.2.4 Conditional simulation of event magnitudes

The magnitude of each event in the bootstrap catalogue may then be simulated conditional on its occurrence time. Given a simulated occurrence time $\tilde{\tau} \in (0, \tau_{\max})$, the conditional magnitude intensity for magnitudes exceeding $v(\tilde{\tau})$ is simply the GPD density function with shape parameter ξ and time-dependent scale parameter $\sigma(\tilde{\tau}) = \sigma_u + \xi(v(\tilde{\tau}) - u)$:

$$\lambda(y|\tilde{\tau}) = \frac{\lambda(\tilde{\tau}, y)}{\lambda(\tilde{\tau})} = \frac{1}{\sigma(\tilde{\tau})} \left[1 + \xi \frac{y - v(\tilde{\tau})}{\sigma(\tilde{\tau})} \right]_+^{-1/\xi-1}, \quad \text{for } y \geq v(\tilde{\tau}). \quad (\text{B.2.4})$$

This allows easy simulation of magnitudes $\tilde{\mathbf{y}} = (\tilde{y}_1, \dots, \tilde{y}_{\tilde{n}_v})$ conditional on their occurrence times $\tilde{\tau}$, by generating random variates from the appropriate GPD. The simulation of a bootstrap earthquake catalogue is completed by rounding these to the nearest multiple of 2δ , to obtain $\tilde{\mathbf{x}} = (\tilde{x}_1, \dots, \tilde{x}_{\tilde{n}_v})$. Note that in a bootstrap catalogue $\tilde{y}_i > \tilde{v}_i$ for $i = 1, \dots, \tilde{n}_v$, but following rounding it is possible that some of the \tilde{x}_i are below their associated threshold value. The simulation of bootstrapped datasets is summarised in Algorithm 1, where $\hat{\sigma}(\tilde{\tau}) = \hat{\sigma}_u + \hat{\xi}(v(\tilde{\tau}) - u)$.

Algorithm 1: Simulation of GPD data with variable threshold and rounding.

Result: A bootstrapped dataset of rounded GPD observations $\tilde{\mathbf{x}}$, based on the threshold function $v(\tau)$, observations \mathbf{x} and parameter estimates $\hat{\boldsymbol{\theta}}$.

Sample m_v as the sum of independent Bernoulli(w_i) realisations where $i = 1, \dots, n$;

Sample the estimate of $\Lambda(A_v)$, $\hat{\Lambda}(A_v)$ from the Poisson(m_v) distribution;

Sample the bootstrap number of exceedances \tilde{n}_v from the Poisson($\hat{\Lambda}(A_v)$) distribution;

for $i = 1$ **to** $i = \tilde{n}_v$ **do**

sample u_i from a Uniform(0,1) distribution;

find the bootstrap occurrence time $\tilde{\tau}_i$ which satisfies

$$\int_0^{\tilde{\tau}_i} \lambda(\tau; \hat{\sigma}_u, \hat{\xi}, v(\tau)) d\tau = u_i \hat{\Lambda}(A_v);$$

sample the bootstrap magnitude exceedance \tilde{z}_i from the GPD($\hat{\sigma}(\tilde{\tau}_i), \hat{\xi}$) distribution;

calculate the bootstrap latent magnitude $\tilde{y}_i = v(\tilde{\tau}_i) + \tilde{z}_i$;

round \tilde{y}_i to the nearest 2δ to get the rounded bootstrap magnitude \tilde{x}_i (which may be less than $v(\tilde{\tau}_i)$).

end

B.2.3 Generating bootstrap maximum likelihood estimates

In the bootstrapped earthquake catalogues, some magnitudes $\tilde{\mathbf{x}}$ may be less than their respective threshold values. However, unlike in the original catalogue, each of the corresponding unrounded magnitudes $\tilde{\mathbf{y}}$ exceeds the respective modelling threshold. Therefore, an unweighted log-likelihood should be used when obtaining bootstrap maximum likelihood estimates. Letting $\tilde{v}_i = v(\tilde{\tau}_i)$ and $\sigma_{\tilde{v}_i} = \sigma_u + \xi(\tilde{v}_i - u)$, the unweighted log-likelihood function is

$$\ell(\boldsymbol{\theta}|\tilde{\mathbf{x}}, \tilde{\mathbf{v}}) = \sum_{i=1}^{\tilde{n}_v} \log [F(\tilde{x}_i + \delta - \tilde{v}_i; \sigma_{\tilde{v}_i}, \xi) - F(\max(\tilde{v}_i, \tilde{x}_i - \delta) - \tilde{v}_i; \sigma_{\tilde{v}_i}, \xi)].$$

The maximum likelihood estimates resulting from a collection of bootstrap catalogues can be used to represent the sampling uncertainty of the original maximum likelihood point estimate $\hat{\boldsymbol{\theta}}$. This is done in the main text when calculating confidence intervals on parameter values, conditional quantiles and return levels. The bootstrap parameter estimates are also used in the construction of adapted PP and QQ plots and when evaluating metric values.

B.3 Sampling standardised threshold exceedances

B.3.1 Connection to main text

This appendix describes how to sample a vector $\tilde{\mathbf{z}}$ of unrounded threshold exceedances transformed to have a common $\text{Exp}(1)$ marginal distribution. This is used in Sections 5.3 - 5.6 of the main text and requires: a single bootstrap estimate $\tilde{\boldsymbol{\theta}}$ of the estimated GPD parameters $\hat{\boldsymbol{\theta}}$ (obtained as described in Appendix B.2), the n -vector of rounded observations \mathbf{x} and the corresponding threshold vector \mathbf{v} . The process for sampling a single vector $\tilde{\mathbf{z}}$ is described and is formalised in Algorithm 2.

B.3.2 Sampling unrounded threshold exceedances

It is unknown which, if any, of the borderline values $\{x_i \in \mathbf{x} : |x_i - v_i| < \delta\}$ correspond to unrounded values $y_i \in \mathbf{y}$ that exceed the modelling threshold $v(\tau)$. The first step in sampling $\tilde{\mathbf{z}}$ is therefore to sample the set I that exceed the modelling threshold. This is done by simulating independent Bernoulli trials for each event $i = 1, \dots, n$ with success probabilities $w_i = \Pr(Y_i > v_i | x_i, \hat{\boldsymbol{\theta}})$ as defined in equation (5.2.2) of the main text. The vector $\tilde{\mathbf{z}}$ will therefore have a randomly sampled length $\tilde{m} = |I| \leq n$, where the distribution of \tilde{m} depends on \mathbf{x}, \mathbf{v} and $\hat{\boldsymbol{\theta}}$.

The unrounded magnitude values for events in I are then simulated from their conditional distribution given: their rounded values, the estimated GPD parameters and that they are threshold exceedances. Letting F be the GPD distribution function as in Equation (5.1.1) of the main text, the required conditional distribution function of $Y_i | x_i, \boldsymbol{\theta}, Y_i \geq v_i$ is given by:

$$G_{Y_i|x_i, \boldsymbol{\theta}, Y_i \geq v_i}(y) = \begin{cases} 0 & \text{for } y < b_i, \\ \frac{F(y-u; \boldsymbol{\theta}) - F(b_i-u; \boldsymbol{\theta})}{F(x_i+\delta-u; \boldsymbol{\theta}) - F(b_i-u; \boldsymbol{\theta})} & \text{for } b_i \leq y \leq x_i + \delta, \\ 1 & \text{for } y > x_i + \delta, \end{cases} \quad (\text{B.3.1})$$

where $i \in I$ and $b_i = \max(x_i - \delta, v_i)$ is the smallest value above the modelling threshold that results in the rounded observation x_i . The sampled values are combined to create $\tilde{\mathbf{y}}$, an \tilde{m} -vector of continuous-valued sampled threshold exceedances. Note that the length of this vector is a random variate when there are one or more borderline values. By repeated simulation using k bootstrap parameter estimates $\{\hat{\boldsymbol{\theta}}^{(1)}, \dots, \hat{\boldsymbol{\theta}}^{(k)}\}$, the resulting vectors of unrounded exceedances $\{\tilde{\mathbf{y}}^{(1)}, \dots, \tilde{\mathbf{y}}^{(k)}\}$ reflect uncertainty about both the GPD parameter values and the number of threshold exceedances.

B.3.3 Transformation onto common margins

When $v(\tau)$ is non-constant, the elements of $\tilde{\mathbf{y}}$ are random variates from the GPD family, but they do not share a common set of parameters. To resolve this issue, the probability integral transform can then be used to give each element of $\tilde{\mathbf{y}}$ an $\text{Exp}(1)$ marginal distribution using its fitted GPD parameters. This results in a vector $\tilde{\mathbf{z}}$ of standardised, sampled exceedances of the modelling threshold $v(\tau)$. The bootstrap simulation of one such vector is formalised in Algorithm 2.

Note that the transformation onto exponential margins is dependent on the estimated GPD parameters. The effect of parameter uncertainty on this transformation is represented across bootstrap catalogues by transforming each collection of bootstrapped threshold exceedances $\{\tilde{\mathbf{y}}^{(1)}, \dots, \tilde{\mathbf{y}}^{(k)}\}$ using their respective bootstrapped parameter estimates $\{\tilde{\boldsymbol{\theta}}^{(1)}, \dots, \tilde{\boldsymbol{\theta}}^{(k)}\}$. This yields a set of k sampled vectors of threshold exceedances on exponential margins, $\{\tilde{\mathbf{z}}^{(1)}, \dots, \tilde{\mathbf{z}}^{(k)}\}$. As with $\tilde{\mathbf{y}}$, the length of $\tilde{\mathbf{z}}$ is a \tilde{m} -vector and so is a random variate when there are one or more borderline events in \mathbf{x} .

These sampled vectors of standardised threshold excesses can be used to calculate expected metric values or to construct modified PP- and QQ-plots, as in Figures 5.4.1 and 5.6.1 of the main text. In those plots, the variability between the empirical quantiles (or probabilities) of each $\{\tilde{\mathbf{z}}^{(1)}, \dots, \tilde{\mathbf{z}}^{(k)}\}$ is shown by the confidence intervals on sample quantile values (or probabilities). The expected range of values for each sample quantile (or probability) is shown by the tolerance intervals. The uncertainty in the number of threshold exceedances is also incorporated when calculating the tolerance intervals; they are constructed using k sets of $\text{Exp}(1)$ random variates of lengths $\dim \tilde{\mathbf{z}}^{(1)}, \dots, \dim \tilde{\mathbf{z}}^{(k)}$.

Algorithm 2: Simulation of standardised threshold exceedance sets.

input : A bootstrap estimate $\tilde{\boldsymbol{\theta}} = (\tilde{\sigma}, \tilde{\xi})$ of the GPD parameters, the n -vector of rounded observed values \mathbf{x} , and their corresponding thresholds \mathbf{v} .

output: A vector $\tilde{\mathbf{z}}$ of length $\tilde{m} \leq n$ of sampled unrounded values, transformed to have an $\text{Exp}(1)$ distribution under the fitted model.

for $i = 1$ **to** n **do**

- | calculate $w_i = \Pr(Y_i > v_i | x_i, \tilde{\boldsymbol{\theta}})$, the probability that each rounded observation corresponds to an unrounded value on A_v , as in Equation (3) of the main text;

end

Generate n independent Uniform[0,1] random variates u_1, \dots, u_n ;

Sample the indexing set of events that are on A_v , $I = \{i \in (1, \dots, n) : u_i \leq w_i\}$ and let $\tilde{m} = |I|$;

Store the elements of I in the vector $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{\tilde{m}})$ so that $\tilde{y}_{\beta_i} > v_{\beta_i}$ for $i = 1, \dots, \tilde{m}$ and initialise $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{z}}$ as vectors of length \tilde{m} ;

for $j = 1$ **to** \tilde{m} **do**

- | Let $a = \beta_j$;
- | Sample the j^{th} unrounded exceedance \tilde{y}_j from its conditional distribution $G_{Y_a|x=x_a, \boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}, Y_a \geq v_a}(y)$ as in equation (B.3.1) ;
- | Let $\tilde{\boldsymbol{\theta}}_{v_a} = (\tilde{\sigma} + \tilde{\xi}(v_a - u), \tilde{\xi})$ be the bootstrapped GPD parameters for exceedances of v_a ;
- | Transform \tilde{y}_j onto $\text{Exp}(1)$ margins under this fitted model by letting F be the GPD distribution function (1) in the main text and setting

$$\tilde{z}_j = -\log \left[1 - F \left(\tilde{y}_j - v_a; \tilde{\boldsymbol{\theta}}_{v_a} \right) \right].$$

end

B.4 Supplementary figures

The following pages of this appendix contains supplementary figures relating to Chapter 5 of this thesis.

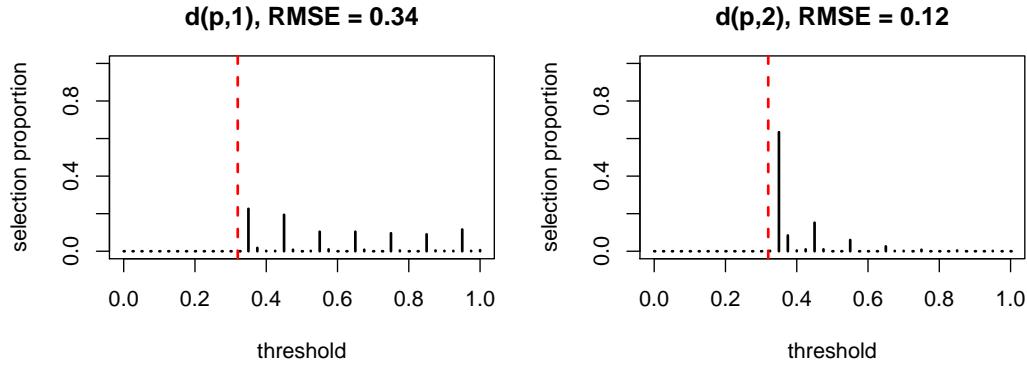


Figure B.4.1: Sampling distribution of threshold selection methods for PP-based metrics over 500 simulated catalogues with constant threshold and hard censoring. The true threshold is shown by a dashed red line and the root mean squared error (RMSE) for each method is given in plot titles.

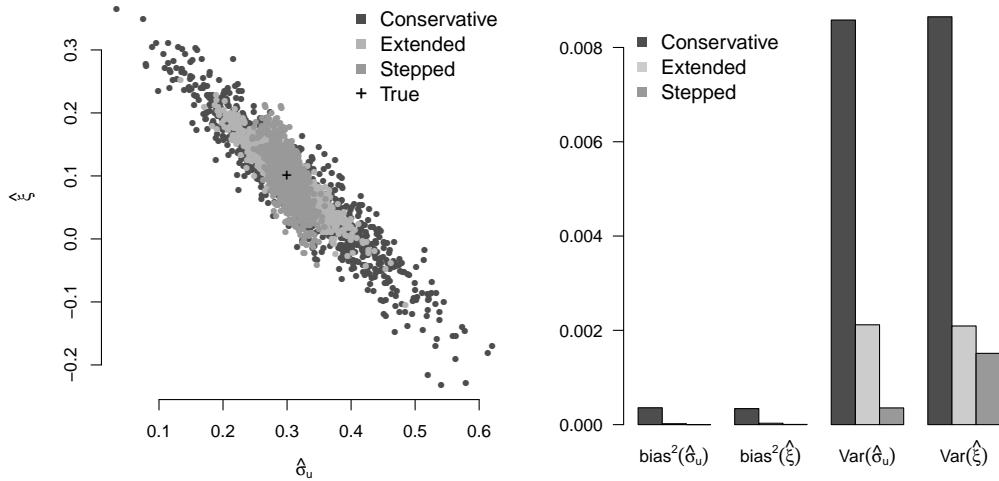


Figure B.4.2: [Left] Plot of maximum likelihood estimates of GPD scale and shape parameters for 1000 simulated catalogues obtained using a conservative, stepped and extended approach. [Right] Plot of mean squared error decomposition for maximum likelihood estimates by each approach. The error is decomposed into squared bias and variance terms for each of the GPD parameters, with variance terms having larger contributions.

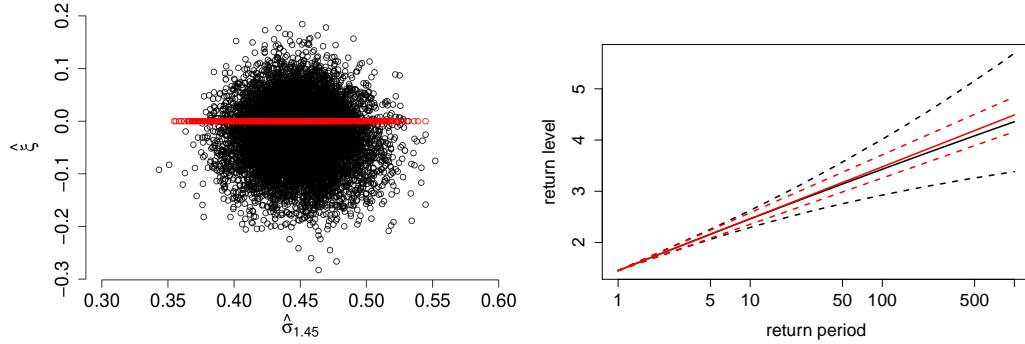


Figure B.4.3: [left] Bootstrap maximum likelihood estimates for GPD parameters for Groningen magnitudes above 1.45ML assuming GPD (black) and exponential (red) models. [right] Resulting conditional return level plot with return period measured in number of events exceeding 1.45 ML.

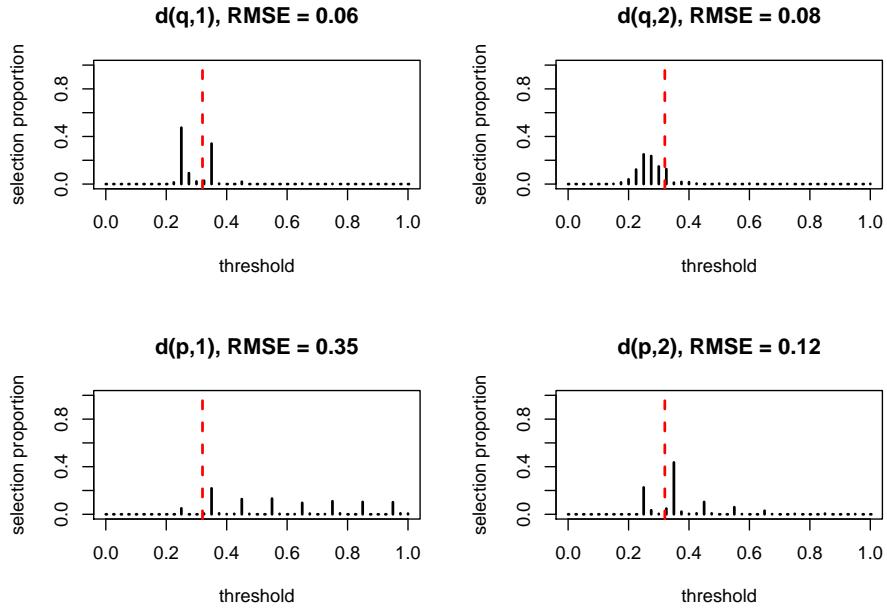


Figure B.4.4: Sampling distribution of threshold selection methods for QQ-based (top row) and PP-based (bottom row) metrics over 500 simulated catalogues with constant threshold and phased censoring. The true threshold is shown by a dashed red line and the root mean squared error (RMSE) for each method is given in plot titles.

Appendix C

Supplementary materials to Chapter 6

C.1 B conditional posterior, dual magnitude model

In the dual magnitude ETAS model there are two magnitude distributions: one for background events and another for triggered events. Each magnitude M_1, \dots, M_n remains independent of all other magnitudes, as in the standard ETAS model. However, introducing type-dependent magnitudes means that each element of the branching vector, B_i , is no longer independent of its corresponding magnitude, M_i for $i = 1, \dots, n$. This is because the observed magnitude of event i , m_i , carries information to discern whether $B_i = 0$ or $B_i > 0$ when the two magnitude distributions are not equal. Conversely, knowing that $b_i = 0$ or $b_i > 0$ is informative about the likely values of M_i .

Let $f_0(m; \psi_0)$ and $f_1(m; \psi_1)$ denote, respectively, the probability density functions for magnitudes of background events and triggered events, where $\psi = (\psi_0, \psi_1)$ is the vector of all magnitude parameters. Additionally, define $d_i = \min(1, b_i)$ to be the

index of the magnitude distribution to which event i belongs, where $i = 1, \dots, n$.

The conditional posterior for the full branching vector $\pi_{B|Y,\theta,\psi}(b)$ may still, as for the standard ETAS model, be written as the product of element-wise conditional posteriors:

$$\pi_{B|Y,\theta,\psi}(b) = \Pr(B = b|Y, \theta, \psi) = \prod_{i=1}^n \Pr(B_i = b_i|Y, \theta, \psi). \quad (\text{C.1.1})$$

In particular this means that the elements of the branching vector may be updated sequentially as part of the Gibbs sampling scheme used to fit the dual magnitude ETAS model. The element-wise conditional posteriors that constitute the product terms in equation (C.1.1) are less simple than in the case of i.i.d. magnitudes, but still have a closed form, which we derive presently.

Let $\mathcal{H}_i = \{Y_j \in Y : j = 1, \dots, i-1\}$, be the history of the point process up to event i for $i = 1, \dots, n$. The conditional posterior for a single element of the branching vector does not depend on events $\{Y_{i+1}, \dots, Y_n\}$ and so

$$\begin{aligned} \Pr(B_i = b_i|Y, \theta, \psi) &= \Pr(B_i = b_i|\mathcal{H}_i, Y_i, \theta, \psi) \\ &= \Pr(B_i = b_i|\mathcal{H}_i, T_i, M_i, \theta, \psi). \end{aligned}$$

Applying Bayes' rule and the law of total probability, we have that

$$\Pr(B_i = b_i|Y, \theta, \psi) = \frac{f(m_i|B_i = b_i, \mathcal{H}_i, T_i, \theta, \psi)\Pr(B_i = b_i|\mathcal{H}_i, T_i, \theta)}{\sum_{j=0}^{i-1} f_{\mathbb{I}\{j>0\}}(m_i|\psi_{\mathbb{I}\{j>0\}})\Pr(B_i = j|\mathcal{H}_i, T_i, \theta)}.$$

We may simplify the terms of both the numerator and denominator as follows. Consider the first term of the numerator, $f(m_i|B_i = b_i, \mathcal{H}_i, T_i, \theta, \psi)$. Since the branching element B_i is known, we know that M_i has density function $f_{d_i}(m; \psi_{d_i})$. Therefore conditioning of the remaining terms \mathcal{H}_i, T_i and θ contribute no further information about the distribution of M_i and these may be dropped. Now consider the second

term of the numerator, $\Pr(B_i = b_i | \mathcal{H}_i, T_i, \theta, \psi)$. Since the M_i is unknown the magnitude parameters ψ are not informative about B_i and may be dropped from the conditioning. Similar arguments apply to the denominator and so

$$\begin{aligned}\Pr(B_i = b_i | Y, \theta, \psi) &= \frac{f_{d_i}(m_i | \psi_{d_i}) \Pr(B_i = b_i | \mathcal{H}_i, T_i, \theta)}{\sum_{j=0}^{i-1} f_{\mathbb{I}\{j>0\}}(m_i | \psi_{\mathbb{I}\{j>0\}}) \Pr(B_i = j | \mathcal{H}_i, T_i, \theta)} \\ &= \frac{f_{d_i}(m_i | \psi_{d_i}) \lambda_{b_i}(t_i; \mathcal{H}_i, \theta) \left[\sum_{k=0}^{i-1} \lambda_k(t_i; \mathcal{H}_i, \theta) \right]^{-1}}{\sum_{j=0}^{i-1} f_{\mathbb{I}\{j>0\}}(m_i | \psi_{\mathbb{I}\{j>0\}}) \lambda_j(t_i; \mathcal{H}_i, \theta) \left[\sum_{k=0}^{i-1} \lambda_k(t_i; \mathcal{H}_i, \theta) \right]^{-1}}.\end{aligned}$$

The term in square brackets in the denominator may be brought out of the summation over j to cancel with the term in square brackets in the numerator to give

$$\Pr(B_i = b_i | Y, \theta, \psi) = \frac{f_{d_i}(m_i | \psi_{d_i}) \lambda_{b_i}(t_i; \mathcal{H}_i, \theta)}{\sum_{j=0}^{i-1} f_{\mathbb{I}\{j>0\}}(m_i | \psi_{\mathbb{I}\{j>0\}}) \lambda_j(t_i; \mathcal{H}_i, \theta)}.$$

Separating the normalising constant in the denominator into terms where $j = 0$ and where $j > 0$, since $\lambda_0(t_i, \mathcal{H}_i, \theta) = \mu$ for all $i = 1, \dots, n$, we find that

$$\Pr(B_i = b_i | Y, \theta, \psi) = \frac{f_{d_i}(m_i | \psi_{d_i}) \lambda_{b_i}(t_i; \mathcal{H}_i, \theta)}{\mu f_0(m_i; \psi_0) + f_1(m_i; \psi_1) \sum_{j=1}^{i-1} \lambda_j(t_i; \mathcal{H}_i, \theta)}, \quad (\text{C.1.2})$$

for $i = 1, \dots, n$, where the summation in the denominator is defined to be 0 when $i = 1$.

Expression (C.1.2) is similar to that found in (Ross, 2016) for i.i.d. magnitudes, but the probability mass function is now weighted to reflect the additional information that is conferred by the magnitude of event i about whether $B_i = 0$ or $B_i > 0$. Indeed, it is clear to see that when $f_0(m) = f_1(m)$, the conditional branching posterior (C.1.2) reduces to that of the i.i.d. case:

$$\frac{\lambda_{b_i}(t_i; \mathcal{H}_i, \theta)}{\sum_{j=0}^{i-1} \lambda_j(t_i; \mathcal{H}_i, \theta)}.$$

We can therefore draw the vector B directly from its conditional posterior as part of a Metropolis-within-Gibbs scheme, by drawing each element B_1, \dots, B_n according to its conditional posterior distribution, as given by expression (C.1.2). Note that for each event $i = 1, \dots, n$ the denominator terms in expression (C.1.2) are normalising constants for the probability mass function, taking the same value for each potential parent process $b_i = 0, \dots, i - 1$. Since sampling from this posterior distribution requires the probability mass function only up to a constant of proportionality, these denominator terms need not be calculated explicitly for each value of b_i .

C.2 ψ conditional posterior, correlated magnitudes

In the correlated magnitude ETAS model, triggering and triggered events may have distinct marginal magnitude distributions, which are coupled using a Gaussian copula. This extends the vector of magnitude parameters to $\psi = (\psi_0, \psi_1, \rho)$. The elements of ψ give parameters of the marginal distributions for background and triggered events as well as ρ , the correlation of triggered and triggering magnitudes when transformed onto standard Gaussian margins.

In the following, let $f_0(m; \psi_0)$ and $f_1(m; \psi_1)$ denote the marginal magnitude distributions of background and triggered events and $\phi(m)$ denote the standard Gaussian density. Let the corresponding cumulative distribution functions be denoted by $F_0(m|\psi_0)$, $F_1(m|\psi_1)$ and $\Phi(m)$. Finally, define $d_i = \min(1, b_i)$ to be the index of the magnitude distribution to which event i belongs, where $i = 1, \dots, n$.

Note that if the branching vector B is known, then the graphical representation of the point pattern as a collection of trees is also known. The likelihood of the magnitude

parameters given the magnitudes in a single tree may be calculated by multiplying the marginal density of the root magnitude (which corresponds to a background event) by the conditional densities of each triggered magnitude in the tree given the distribution and value of its parent, working from root to leaves. Since magnitudes in separate trees are independent, the likelihood of the magnitude parameters given all magnitudes is the product of such expressions over all trees. This combined likelihood may be written most clearly by collecting terms for background (root) and triggered (non-root) events. The conditional posterior of the magnitude parameters, assuming improper uniform priors, is then given by

$$\pi_{\psi|Y,B,\theta}(\psi) = \prod_{i:B_i=0} f_0(m_i; \psi_0) \prod_{i:B_i>0} f(m_i|m_{b_i}, b_i, b_{b_i}, \psi), \quad (\text{C.2.1})$$

where the terms of the second product are the conditional densities of triggered events, given the magnitude, index and marginal distribution of the corresponding parent events, in addition to the magnitude parameters. We can express this conditional density in terms of the marginal and copula densities.

To simplify notation, we will consider a parent-child magnitude pair pair of events (m_p, m_c) so that $1 \leq p < c \leq n$ and $b_c = p$. These magnitudes are transformed from their marginal distributions $f_{d_p}(m; \psi)$ and $f_{d_c}(m; \psi)$ onto standard Gaussian margins to give (g_p, g_c) . (Note that since event c is a child of event p we therefore know that $b_c = p$ and $d_c = 1$.) The transformation onto standard Gaussian margins can be achieved by application of the probability integral transform, where

$$g_p = \Phi^{-1} \circ F_{d_p}(m_p; \psi) \quad \text{and} \quad g_c = \Phi^{-1} \circ F_{d_c}(m_c; \psi). \quad (\text{C.2.2})$$

Then the conditional density that we require is

$$\begin{aligned} f(m_c|m_p, b_p, b_c = p, \psi) &= f(g_c|g_p, b_c = p, \psi) \left| \frac{dg_c}{dm_c} \right| \\ &= f(g_c|g_p, b_c = p, \psi) [\phi \circ \Phi^{-1} \circ F_{d_c}(m_c; \psi)]^{-1} f_{d_c}(m_c; \psi). \end{aligned}$$

By our choice of Gaussian copula linking the marginal distributions of triggering and triggered events it follows that $G_c|g_p, b_c = p, \psi \sim N(\rho g_p, 1 - \rho^2)$ and therefore

$$\begin{aligned} f(m_c|m_p, b_p, b_c = p, \psi) &= \phi\left(\frac{\Phi^{-1}(F_{d_c}(m_c; \psi)) - \rho\Phi^{-1}(F_{d_p}(m_p; \psi))}{\sqrt{1 - \rho^2}}\right) [\phi \circ \Phi^{-1} \circ F_{d_c}(m_c; \psi)]^{-1} f_1(m_c; \psi) \\ &= \phi\left(\frac{g_c - \rho g_p}{\sqrt{1 - \rho^2}}\right) \frac{f_{d_c}(m_c)}{\phi(g_c)}. \end{aligned}$$

We can therefore rewrite the conditional posterior of the magnitude parameters (C.2.1) in terms of the observed magnitudes transformed onto standard Gaussian margins. These transformed magnitudes are denoted by $\{g_i = \Phi^{-1}(F_{d_i}(m_i; \psi)) : i = 1, \dots, n\}$ and the conditional posterior of the magnitude parameter vector is:

$$\pi_{\psi|Y, B, \theta}(\psi) = \prod_{i:B_i=0} f_0(m_i; \psi_0) \prod_{i:B_i>0} \left\{ \phi\left(\frac{g_i - \rho g_{b_i}}{\sqrt{1 - \rho^2}}\right) \frac{f_1(m_i; \psi_1)}{\phi(g_i)} \right\}. \quad (\text{C.2.3})$$

Using this conditional posterior distribution, the magnitude parameters ψ may be updated within the Gibbs sampling scheme used to fit the correlated ETAS model.

C.3 B conditional posterior, correlated magnitudes

In the ETAS model with correlated magnitudes, the elements of the branching vector are no longer conditionally independent of one another given the ETAS and magnitude parameters θ and ψ . This makes the conditional posterior distribution for the full branching vector intractable. However, during the Metropolis-within-Gibbs sampling used to fit the model, each element of B is updated separately. Updating a single branching element B_i requires only the distribution of $B_i|Y, B_{-i}, \theta, \psi$ (up to proportionality) where $B_{-i} = \{B_j \in B : j \neq i\}$ is the branching vector B without its i^{th} element. Fortunately, the conditional posterior of each branching element given

the rest of the branching vector, θ and ψ may be obtained in a closed form, which is derived in this appendix.

As in Appendix C.2, we begin by simplifying the conditioning on subsequent events $\{Y_{i+1}, \dots, Y_n\}$. We may drop dependence on subsequent events which are not directly triggered by event i , but we must retain those ‘child’ events directly triggered by event i . These child events are denoted by the set $Y_{\mathcal{C}_i} = \{Y_j \in Y : j \in \mathcal{C}_i\}$ where $\mathcal{C}_i = \{j \in \{i+1, \dots, n\} : B_j = i\}$. This gives

$$\Pr(B_i = b_i | Y, \theta, \psi, B_{-i}) = \Pr(B_i = b_i | \mathcal{H}_i, Y_i, Y_{\mathcal{C}_i}, \theta, \psi, B_{-i}). \quad (\text{C.3.1})$$

Applying Bayes’ rule and the law of total probability to bring focus onto the joint conditional distribution of $\{M_i, M_{\mathcal{C}_i}\}$ we have that

$$\Pr(B_i = b_i | Y, \theta, \psi, B_{-i}) = \frac{f(m_i, m_{\mathcal{C}_i} | B_i = b_i, \mathcal{H}_i, T_i, T_{\mathcal{C}_i}, \theta, \psi, B_{-i}) \Pr(B_i = b_i | \mathcal{H}_i, T_i, T_{\mathcal{C}_i}, \theta, \psi, B_{-i})}{\sum_{j=0}^{i-1} f(m_i, m_{\mathcal{C}_i} | B_i = j, \mathcal{H}_i, T_i, T_{\mathcal{C}_i}, \theta, \psi, B_{-i}) \Pr(B_i = j | \mathcal{H}_i, T_i, T_{\mathcal{C}_i}, \theta, \psi, B_{-i})}. \quad (\text{C.3.2})$$

We will consider each of the numerator terms and the denominator of equation (C.3.2) individually.

In the following, let $\phi(g)$ and $\Phi(g)$ be the standard Gaussian probability density and cumulative distribution functions. Also let $f_0(m)$ and $f_1(m)$ be the marginal probability density functions of background and triggered magnitudes with corresponding distribution functions $F_0(m)$ and $F_1(m)$. Finally, let $g_{(i,j)} = \Phi^{-1}(F_{\mathbb{I}\{j>0\}}(m_i))$ be the magnitude of event i transformed onto standard Gaussian margins assuming the corresponding branching element $B_i = j$, where $0 \leq j < i \leq n$.

The first numerator term in equation (C.3.2) may be written as

$$\begin{aligned}
& f(m_i, m_{\mathcal{C}_i} | B_i = b_i, \mathcal{H}_i, T_i, T_{\mathcal{C}_i}, \theta, \psi, B_{-i}) \\
&= f(m_i, m_{\mathcal{C}_i} | B_i = b_i, M_{b_i}, \psi, B_{b_i}, B_{\mathcal{C}_i}) \\
&= f(m_i | B_i = b_i, M_{b_i}, B_{b_i}, \psi) \prod_{c \in \mathcal{C}_i} f(m_c | M_i, B_i = b_i, B_c = i, \psi) \\
&= \left[\phi \left(\frac{g_{(i,b_i)} - \rho g_{(b_i,b_{b_i})}}{\sqrt{1 - \rho^2}} \right) \frac{f_1(m_i)}{\phi(g_{(i,b_i)})} \right]^{\mathbb{I}(b_i > 0)} [f_0(m_i)]^{\mathbb{I}(b_i = 0)} \times \\
&\quad \prod_{c \in \mathcal{C}_i} \left\{ \phi \left(\frac{g_{(c,i)} - \rho g_{(i,b_i)}}{\sqrt{1 - \rho^2}} \right) \frac{f_1(m_c)}{\phi(g_{(c,i)})} \right\}. \tag{C.3.3}
\end{aligned}$$

The second numerator term in equation (C.3.2) may be written as

$$\Pr(B_i = b_i | \mathcal{H}_i, T_i, T_{\mathcal{C}_i}, \theta, \psi, B_{-i}) = \Pr(B_i = b_i | \mathcal{H}_i, T_i, \theta) = \frac{\lambda_{b_i}(t_i; \mathcal{H}_i, \theta)}{\sum_{k=0}^{i-1} \lambda_k(t_i; \mathcal{H}_i, \theta)}. \tag{C.3.4}$$

By similar arguments used to obtain (26) and (27), the denominator of equation (C.3.2) is equal to:

$$\begin{aligned}
& \sum_{j=0}^{i-1} \left\{ \left[\phi \left(\frac{g_{(i,j)} - \rho g_{(j,b_j)}}{\sqrt{1 - \rho^2}} \right) \frac{f_1(m_i)}{\phi(g_{(i,j)})} \right]^{\mathbb{I}(j > 0)} [f_0(m_i)]^{\mathbb{I}(j = 0)} \times \right. \\
&\quad \left. \prod_{c \in \mathcal{C}_i} \left\{ \phi \left(\frac{g_{(c,i)} - \rho g_{(i,j)}}{\sqrt{1 - \rho^2}} \right) \frac{f_1(m_c)}{\phi(g_{(c,i)})} \right\} \frac{\lambda_j(t_i; \mathcal{H}_i, \theta)}{\sum_{k=0}^{i-1} \lambda_k(t_i; \mathcal{H}_i, \theta)} \right\}. \tag{C.3.5}
\end{aligned}$$

The conditional posterior $\pi_{B_i | Y, B_{-i}, \theta, \psi}(b_i)$ may be obtained by combining expressions (C.3.3), (C.3.4) and (C.3.5). When this is done, the summations indexed by k in (C.3.4) and (C.3.5) cancel to give, up to proportionality:

$$\Pr(B_i = b_i | Y, \theta, \psi, B_{-i}) \propto \left[\phi \left(\frac{g_{(i,b_i)} - \rho g_{(b_i, b_{b_i})}}{\sqrt{1 - \rho^2}} \right) \frac{f_1(m_i)}{\phi(g_{(i,b_i)})} \right]^{\mathbb{I}(b_i > 0)} [f_0(m_i)]^{\mathbb{I}(b_i = 0)} \times \\ \prod_{c \in \mathcal{C}_i} \left\{ \phi \left(\frac{g_{(c,i)} - \rho g_{(i,b_i)}}{\sqrt{1 - \rho^2}} \right) \frac{f_1(m_c)}{\phi(g_{(c,i)})} \right\} \lambda_{b_i}(t_i; \mathcal{H}_i, \theta). \quad (\text{C.3.6})$$

Using this conditional posterior distribution, the elements of the branching vector B_1, \dots, B_n may be sequentially updated within the Gibbs sampling scheme used to fit the correlated ETAS model. As was noted in Appendix C.1, this requires $\pi_{B_i|Y,\theta,\psi,B_{-i}}(b_i)$ only up to a constant of proportionality, and so the normalising denominator term (C.3.6) (which for a given event $i = 1, \dots, n$ is equal for all $b_i = 0, \dots, i-1$) need not be evaluated.

Note that when $\rho = 0$ the conditional posterior of each branching element C.3.6 simplifies to

$$\Pr(B_i = b_i | Y, \theta, \psi, B_{-i}) \propto f_{\mathbb{I}(b_i > 0)}(m_i; \psi_{\mathbb{I}(b_i > 0)}) \prod_{c \in \mathcal{C}_i} \{f_1(m_c)\} \lambda_{b_i}(t_i; \mathcal{H}_i, \theta) \\ \propto f_{\mathbb{I}(b_i > 0)}(m_i; \psi_{\mathbb{I}(b_i > 0)}) \lambda_{b_i}(t_i; \mathcal{H}_i, \theta),$$

which is the corresponding expression for the dual magnitude model. As noted previously, this further simplifies to the standard ETAS model when the same distribution is used for magnitudes of background and triggered events so that $f_0(m; \psi_0) = f_1(m; \psi_1)$ for all m .

Appendix D

Supplementary materials to Chapter 7

D.1 Outline for combined covariate and aftershock model

D.1.1 Link to main text

In Chapter 7, one suggestion given for further work was to develop a combined model for induced earthquakes and aftershocks inspired by the ETAS model. In this suggested model, earthquakes do not induce further events by altering the point process intensity function directly, but instead by altering the covariates (such as incremental Coulomb stress) which feed into a parametric intensity model. In this appendix we give some initial thoughts to motivate such a model and consider how a model of this type might be constructed.

D.1.2 Motivation

To describe clustering of observed events in the Groningen earthquake catalogue, the epidemic type aftershock sequence (ETAS) model can be used. The intensity function for this point process has two components: a background intensity that changes in space and time based on gas extraction and additional intensity that is added around each past event. The background intensity is currently modelled as a function of the stress state of the reservoir. The additional intensity is allocated isotropically about the past events, where the amount of added intensity is an increasing function of event magnitude and a decreasing function of both distance and time since the event.

An issue with this model is that the occurrence of background and triggered events are treated differently. The background event intensity is driven by the stress state of the reservoir, but the additional aftershock intensity does not depend on the stress state of the reservoir at all. Here we describe a model in which all seismic activity is driven by the stress state of the reservoir. We propose to achieve this by having earthquakes increase the stress state around previous events, and therefore indirectly increase the point process intensity function, rather than increasing the intensity function around past events directly.

In this appendix we explore a few simple forms for how the additional stress might be allocated over the space and time around each event. We compare these suggestions based on their implications for the intensity function. We then state the generic likelihood function for the proposed model, without restricting to one choice of additional strain allocation, and discuss its potential issues. When giving an outline of this combined aftershock model, we consider a background intensity function where earthquake count increases exponentially with cumulative stress c , which motivates

an intensity model of the form:

$$\lambda_1(\mathbf{x}, t) = \dot{c}(\mathbf{x}, t)\beta_0(1 + \beta_1 c(\mathbf{x}, t)) \exp(\beta_1 c(\mathbf{x}, t)), \quad (\text{D.1.1})$$

where $c(\mathbf{x}, t)$ is the cumulative stress at location \mathbf{x} and time t caused by gas extraction, $\dot{c}(\mathbf{x}, t)$ is the stress rate caused by gas extraction and (β_0, β_1) are model parameters to be estimated. Where it clarifies exposition, the explicit dependence of the covariates on location and time will be dropped, so that equation (D.1.1) is equivalent to:

$$\lambda_1 = \dot{c}\beta_0(1 + \beta_1 c) \exp(\beta_1 c). \quad (\text{D.1.2})$$

D.1.3 Suggested model

D.1.3.1 Description

We suggest that rather than previous earthquakes directly adding to the intensity function, that they should instead add to the covariate that is driving seismic activity in the region of the event. This will contribute to the intensity through the same model as the stress caused by gas extraction, providing a model that is more cohesive and coherent with the physical process that is thought to be driving earthquake activity.

In the proposed model, the intensity function from equation (D.1.1) would be replaced by:

$$\lambda_2 = \dot{s}(\mathbf{x}, t, \mathcal{H}_t)\beta_0(1 + \beta_1 s(\mathbf{x}, t, \mathcal{H}_t)) \exp(\beta_1 s(\mathbf{x}, t, \mathcal{H}_t)) \quad (\text{D.1.3})$$

where $s(\mathbf{x}, t, \mathcal{H}_t)$ is the total stress, which is a function of location (\mathbf{x}), time (t) and now also depends on previous earthquakes (\mathcal{H}_t). The total stress $s(\mathbf{x}, t, \mathcal{H}_t)$ is composed of the stress caused by gas extraction $c(\mathbf{x}, t)$ and the stress caused by previous earthquakes $\sigma(\mathbf{x}, t, \mathcal{H}_t)$:

$$s(\mathbf{x}, t, \mathcal{H}_t) = c(\mathbf{x}, t) + \sigma(\mathbf{x}, t, \mathcal{H}_t). \quad (\text{D.1.4})$$

The total stress rate $\dot{s}(\mathbf{x}, t, \mathcal{H}_t)$ is the sum of the stress rates from extraction and previous earthquakes:

$$\dot{s}(\mathbf{x}, t, \mathcal{H}_t) = \dot{c}(\mathbf{x}, t) + \dot{\sigma}(\mathbf{x}, t, \mathcal{H}_t). \quad (\text{D.1.5})$$

Figures D.1.1 and D.1.2 graphically present the differences in dependence structure between the current ETAS model and the proposed combined model. In the ETAS model, induced events and aftershocks form separate model components, while under the combined model these are modelled together.

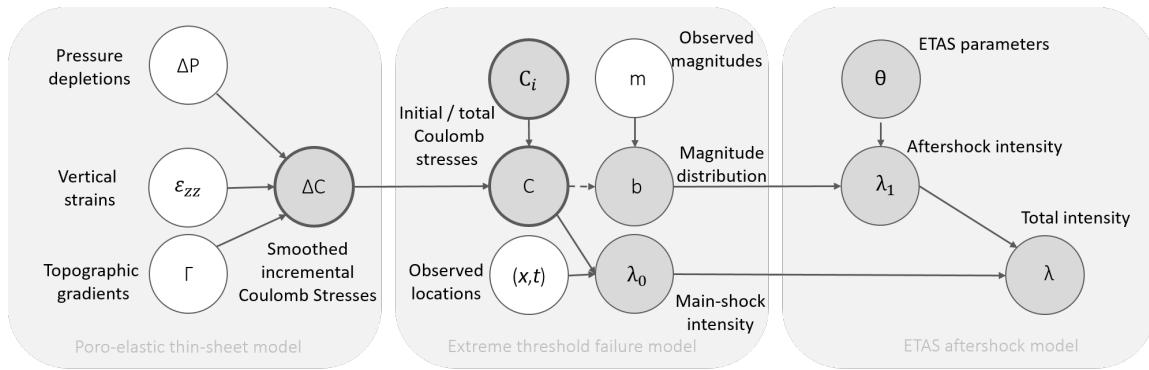


Figure D.1.1: Schematic of ETAS model inputs, component models, and outputs.

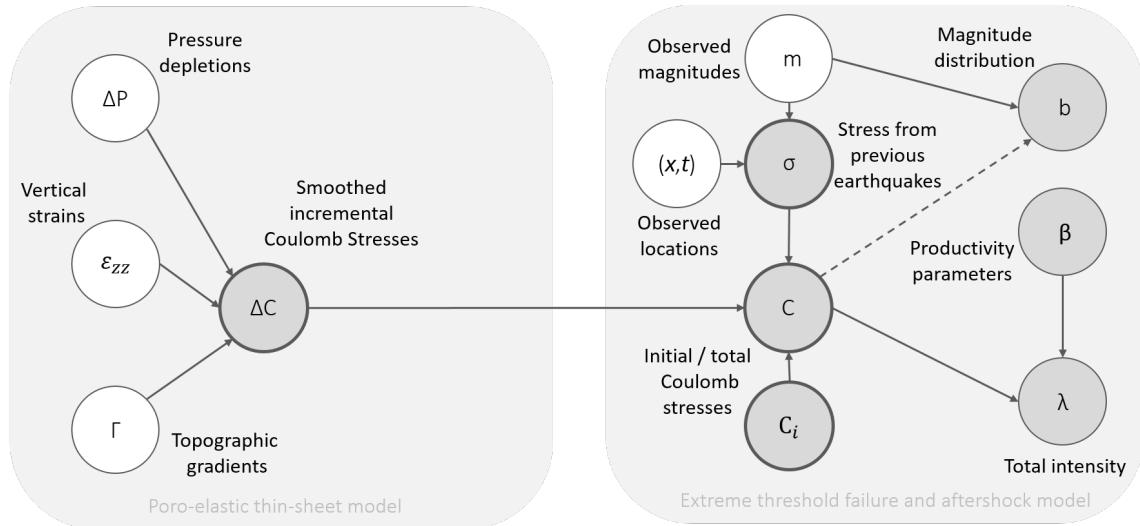


Figure D.1.2: Schematic of combined model inputs, component models, and outputs.

D.1.3.2 Modelling assumptions

In Sections D.1.6 and D.1.7 we will propose functional forms for the total additional stress at location \mathbf{x} and time t caused by all previous earthquakes $\sigma(\mathbf{x}, t, \mathcal{H}_t)$. We then discuss the implications of these forms on the point process intensity function. In doing so, we will make the following assumptions:

1. The additional stress from all previous earthquakes is small relative to the cumulative stress caused by gas extraction, $\sigma(\mathbf{x}, t, \mathcal{H}_t) \ll c(\mathbf{x}, t)$. This means that the cumulative total stress is approximated well by the cumulative stress from gas extraction:

$$s(\mathbf{x}, t, \mathcal{H}_t) \approx c(\mathbf{x}, t). \quad (\text{D.1.6})$$

2. The total additional stress is the sum of contributions from all of the previous earthquakes. The size of the contribution from the i^{th} event should depend the proximity of its space-time location (\mathbf{x}_i, t_i) , as well as that event's magnitude m_i . The form of this dependence is common to all earthquakes so that

$$\sigma(\mathbf{x}, t, \mathcal{H}_t) = \sum_{i:t_i < t} f(\mathbf{x} - \mathbf{x}_i, t - t_i, m_i) \quad (\text{D.1.7})$$

for some function $f : \mathbb{R}^2 \times \mathbb{R}_+ \times [M_c, M_{\max}] \rightarrow \mathbb{R}_+$, where the M_c is the magnitude of completion of the catalogue and M_{\max} is the upper end point of the magnitude distribution.

3. The effects of distance, time and magnitude on additional stress are independent, so that

$$f(\mathbf{x}, t, m) = \sigma_x(\mathbf{x})\sigma_t(t)\sigma_m(m). \quad (\text{D.1.8})$$

where $\sigma_x : \mathbb{R}^2 \rightarrow \mathbb{R}_+$, $\sigma_t : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ and $\sigma_m : [M_c, M_{\max}] \rightarrow \mathbb{R}_+$.

D.1.3.3 Intensity function

Under the assumptions D.1.6 to D.1.8, we may rewrite the intensity function for the proposed model in equation (D.1.3) as:

$$\begin{aligned}
 \lambda_2 &= (\dot{c} + \dot{\sigma})\beta_0(1 + \beta_1(c + \sigma)) \exp(\beta_1(c + \sigma)) \\
 &\approx (\dot{c} + \dot{\sigma})\beta_0(1 + \beta_1c) \exp(\beta_1c) \\
 &= \dot{c}\beta_0(1 + \beta_1c) \exp(\beta_1c) + \dot{\sigma}\beta_0(1 + \beta_1c) \exp(\beta_1c) \\
 &= \dot{c}\beta_0(1 + \beta_1c) \exp(\beta_1c) + \\
 &\quad \beta_0(1 + \beta_1c) \exp(\beta_1c) \sum_{i:t_i < t} \sigma_x(\mathbf{x} - \mathbf{x}_i) \sigma_m(m - m_i) \dot{\sigma}_t(t - t_i). \tag{D.1.9}
 \end{aligned}$$

The similarities differences between the proposed and ETAS models are apparent from inspection of this intensity function. As in the ETAS model, we have a baseline intensity driven by gas extraction (the first term in equation (D.1.9)) and additional contributions to the intensity function from previous events (the second term in equation (D.1.9)). However, in the combined aftershock model, the allocation of additional intensity through space is controlled by the amount of added stress at that distance and the allocation of additional intensity through time is controlled by the added stress rate.

A second difference from the ETAS model is that, under the combined aftershock model, additional intensity contribution from an earthquake is dependent on the cumulative stress in the area surrounding (and time period following) that event. This means that the expected number of aftershocks produced by an earthquake of known magnitude will be greater if that event occurs in a region or at a time where the cumulative stress is higher. This follows from the logic that in areas of higher stress a greater number of faults are close to slipping and so a given increase in stress (whether caused by gas extraction or other earthquakes) should result in a greater number of induced earthquakes in these areas, as compared to areas with lower stresses.

A model that includes this property is highly desirable, because it leads to a coherent description of the physical process which triggers both background events and aftershocks.

An additional benefit of the combined model form is that it leads to low aftershock productivity in regions or time periods with low cumulative stresses. This may help to mitigate against window- and edge-effects. This is because additional aftershock intensity will be better contained within the spatial extent of the gas field if the cumulative stresses in the area surrounding the field are comparatively low.

D.1.4 Likelihood function

In order to fit the proposed model, whether by maximum likelihood or Bayesian inference, the likelihood of the parameters of the model will have to be repeatedly calculated. Since the proposed model is still a Hawkes process, the overall form of the likelihood function is similar to that of the ETAS model. Let the parameters determining σ_x , σ_t and σ_m be represented by the vector ψ , and let $\theta = (\beta_0, \beta_1, \psi)$ be the vector of all parameters in the proposed model. For a catalogue of earthquakes $Y = \{(\mathbf{x}_i, t_i, m_i) : i = 1, \dots, n\}$ occurring in the spatial region A and time period $[0, T]$, which we call the observation window $W = A \times [0, T]$, and given the extraction related stresses on this window \mathbf{c} , then the likelihood function for the combined aftershock model is:

$$\ell(\theta | \mathbf{x}, \mathbf{t}, \mathbf{m}, \mathbf{c}) = \sum_{i=1}^n \log (\lambda_2(\mathbf{x}_i, t_i | \mathcal{H}_{t_i}, \theta, \mathbf{c})) - \int_A \int_0^T \lambda_2(\boldsymbol{\xi}, \tau | \mathcal{H}_\tau, \theta, \mathbf{c}) d\tau d\boldsymbol{\xi}. \quad (\text{D.1.10})$$

The double integral in this expression can be decomposed into two double integrals, respectively describing the expected event count from gas extraction and aftershocks.

Specifically,

$$\begin{aligned}
\Lambda(W) &= \int_A \int_0^T \lambda_2(\boldsymbol{\xi}, \tau | \mathcal{H}_\tau, \boldsymbol{\theta}, \mathbf{c}) d\tau d\boldsymbol{\xi} \\
&= \int_A \int_0^T \dot{c}(\boldsymbol{\xi}, \tau) \beta_0(1 + \beta_1 c(\boldsymbol{\xi}, \tau)) \exp(\beta_1 c(\boldsymbol{\xi}, \tau)) d\tau d\boldsymbol{\xi} + \\
&\quad \int_A \int_0^T \beta_0(1 + \beta_1 c(\boldsymbol{\xi}, \tau)) \exp(\beta_1 c(\boldsymbol{\xi}, \tau)) \sum_{i: t_i < \tau} \sigma_x(\boldsymbol{\xi} - \mathbf{x}_i) \dot{\sigma}_t(\tau - t_i) \sigma_m(m_i) d\tau d\boldsymbol{\xi} \\
&= \int_A \beta_0 [c(\boldsymbol{\xi}, T) \exp(\beta_1 c(\boldsymbol{\xi}, T)) - c(\boldsymbol{\xi}, 0) \exp(\beta_1 c(\boldsymbol{\xi}, 0))] d\boldsymbol{\xi} + \\
&\quad \int_A \int_0^T \beta_0(1 + \beta_1 c(\boldsymbol{\xi}, \tau)) \exp(\beta_1 c(\boldsymbol{\xi}, \tau)) \sum_{i: t_i < \tau} \sigma_x(\boldsymbol{\xi} - \mathbf{x}_i) \dot{\sigma}_t(\tau - t_i) \sigma_m(m_i) d\tau d\boldsymbol{\xi}
\end{aligned} \tag{D.1.11}$$

The first of the double integrals has (by design) a closed form in time, but still requires numerical integration over space. The second double integral must be evaluated numerically over both time and space. If all parameters are updated together then this integration will have to be performed at each evaluation of the log-likelihood, which will be computationally very demanding.

If parameters are updated sequentially as part of an MCMC scheme then by storing the value of integrals that are unchanged by the update we can reduce the computational burden of evaluating the log-likelihood. To clarify this point we can rewrite equation (D.1.11) as

$$\begin{aligned}
\Lambda(W) &= \beta_0 \int_A [c(\boldsymbol{\xi}, T) \exp(\beta_1 c(\boldsymbol{\xi}, T)) - c(\boldsymbol{\xi}, 0) \exp(\beta_1 c(\boldsymbol{\xi}, 0))] d\boldsymbol{\xi} + \\
&\quad \beta_0 \sum_{i=1}^n \left\{ \sigma_m(m_i) \int_A \sigma_x(\boldsymbol{\xi} - \mathbf{x}_i) \int_{t_i}^T (1 + \beta_1 c(\boldsymbol{\xi}, \tau)) \exp(\beta_1 c(\boldsymbol{\xi}, \tau)) \dot{\sigma}_t(\tau) d\tau d\boldsymbol{\xi} \right\}.
\end{aligned} \tag{D.1.12}$$

In this form we can see that in order to update β_0 or the parameters of σ_m , no additional numerical integration is required. When updating β_1 or the parameters of σ_x , then numerical integration is only required in two dimensions. It is only when

updating the parameters of σ_t that numerical integration over space and time will be required. Exploiting this structure might lead to significant reductions to the computational burden of evaluating the combined model.

D.1.5 Condition for sub-criticality

As with the ETAS model, the proposed model can be viewed as a branching process. In order for this process to be sub-critical, the expected number of aftershocks per triggering event must be strictly less than one. Let $g(m)$ denote a probability density function that represents the distribution of earthquake magnitudes. We will assume here that this distribution is independent of the stress state at the location of the event. This simplifying assumption can later be relaxed if the combined model improves upon the ETAS model with the same assumptions.

Conditional on the location and time of the triggering event, the expected number of aftershocks from a triggering event is:

$$\begin{aligned} \mathbb{E}[N_{\text{after}}|\boldsymbol{x}, t] &= \\ &= \int_0^\infty \int_A \int_0^T \beta_0(1 + \beta_1 c(\boldsymbol{\xi}, \tau)) \exp(\beta_1 c(\boldsymbol{\xi}, \tau)) \sigma_x(\boldsymbol{\xi} - \boldsymbol{x}) \dot{\sigma}_t(\tau - t) \sigma_m(m) g(m) d\tau d\boldsymbol{\xi} dm \\ &= \mathbb{E}[\sigma_m(M)] \int_A \int_0^T \beta_0(1 + \beta_1 c(\boldsymbol{\xi}, \tau)) \exp(\beta_1 c(\boldsymbol{\xi}, \tau)) \sigma_x(\boldsymbol{\xi} - \boldsymbol{x}) \dot{\sigma}_t(\tau - t) d\tau d\boldsymbol{\xi}. \end{aligned}$$

To get the unconditional expectation, we would need to integrate the product of this conditional expectation and the probability of an event occurring at each point over the observation window:

$$\mathbb{E}[N_{\text{after}}] = \int_A \int_0^T \mathbb{E}[N_{\text{after}}|\boldsymbol{\xi}, \tau] \frac{\lambda_2(\boldsymbol{\xi}, \tau)}{\Lambda(W)} d\tau d\boldsymbol{\xi}. \quad (\text{D.1.13})$$

To ensure sub-criticality of the branching process, it is required that $\mathbb{E}[N_{\text{after}}] < 1$. This condition that is much more difficult to verify than that for the ETAS model, which can be found by a similar derivation and is stated in Žugec (2019).

D.1.6 Parametric forms for the temporal change in stress

Figure D.1.3 shows six possibilities for how additional stress might be allocated over time following an earthquake. Each black curve shows a suggested additional stress allocation, and the additional stress rate associated with each of these is shown in red. It is the red curves that show the additional stress rate which will impact the intensity function, and which should be the focus here.

The first option, in Figure D.1.3a, is an immediate step change in stress. This is likely not reasonable because it implies a delta function peak in the stress rate, adding an infinite peak to the intensity at the instant of the triggering event but leaving it otherwise unchanged. The second option, in Figure D.1.3b is a more gradual transition to the higher stress state. The stress is initially loaded slowly before reaching some maximum loading rate, after which the loading rate decreases as the final stress state is approached. This leads to the additional intensity peaking at some time after the event and then diminishing again so that the intensity is not effected at long lead times. Figure D.1.3c shows a similar change in the stress state, an increase to plateau, but with a stress rate that is initially high and then reduces as the stress reaches its maximum. This option provides the closest analogue to the ETAS model.

Options four through to six, in Figures D.1.3d - D.1.3f, all suppose that at very long times after an event, the stress state is unchanged by the occurrence of the event. This causes these models to have a common deficiency in that they all require the stress state to decrease, which requires the stress rate to become negative. This has the effect of reducing the intensity function, possibly after some delay. This does not match with the idea that earthquakes should cause more seismic activity in the future, not less.

It seems that the additional stress function shown in Figure D.1.3c is the one that

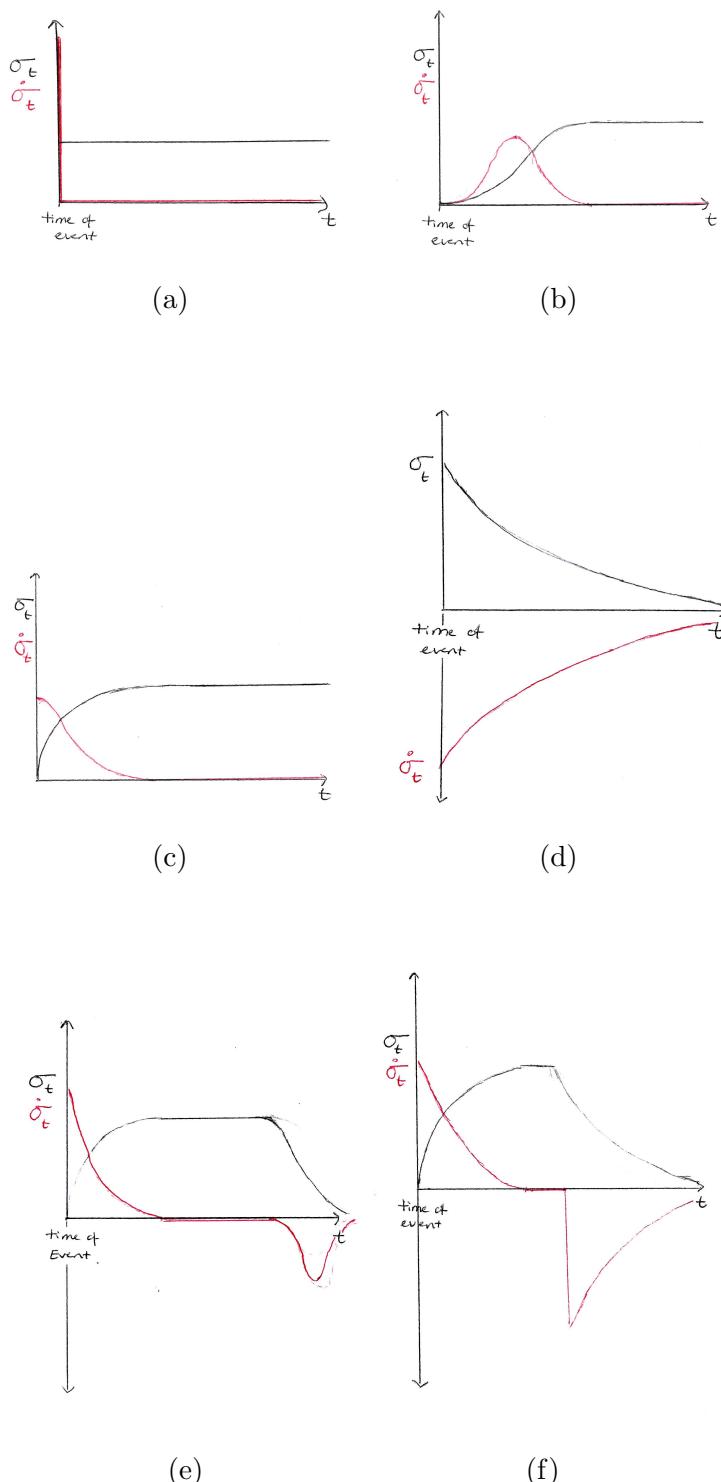


Figure D.1.3: Proposed forms for additional stress (black) and stress rate (red) as functions of time.

is most similar to current practice, but that Figure D.1.3b may also be a feasible option. One possible parameterisation of this function in figure D.1.3c would take the stress rate to be the normalised upper-half of a Gaussian distribution, and the stress function be its integral. If $\phi(\cdot)$ is the density function of standard Gaussian random variable and $\Phi(\cdot)$ is its cumulative distribution function, then the additional stress and stress rate functions are controlled by one parameter ς and are stated below.

$$\dot{\sigma}_t(t) = 2\phi\left(\frac{t}{\varsigma}\right) \quad \text{for } t \geq 0 \text{ and } \varsigma > 0. \quad (\text{D.1.14})$$

$$\sigma_t(t) = 2\varsigma \left(\Phi\left(\frac{t}{\varsigma}\right) - 0.5 \right) \quad \text{for } t \geq 0 \text{ and } \varsigma > 0. \quad (\text{D.1.15})$$

Following feedback from Shell on these model forms, it was suggested that if additional stress rate varied according to an inverse power law in time, then this would be consistent with the rate and state friction model. Motivated by this and the parallels that this form gives to the current modelling approach, we might process by a more flexible version of the inverse power law model. For example could describe the additional stress rate using a generalised Pareto density function and the additional stress by its cumulative density function.

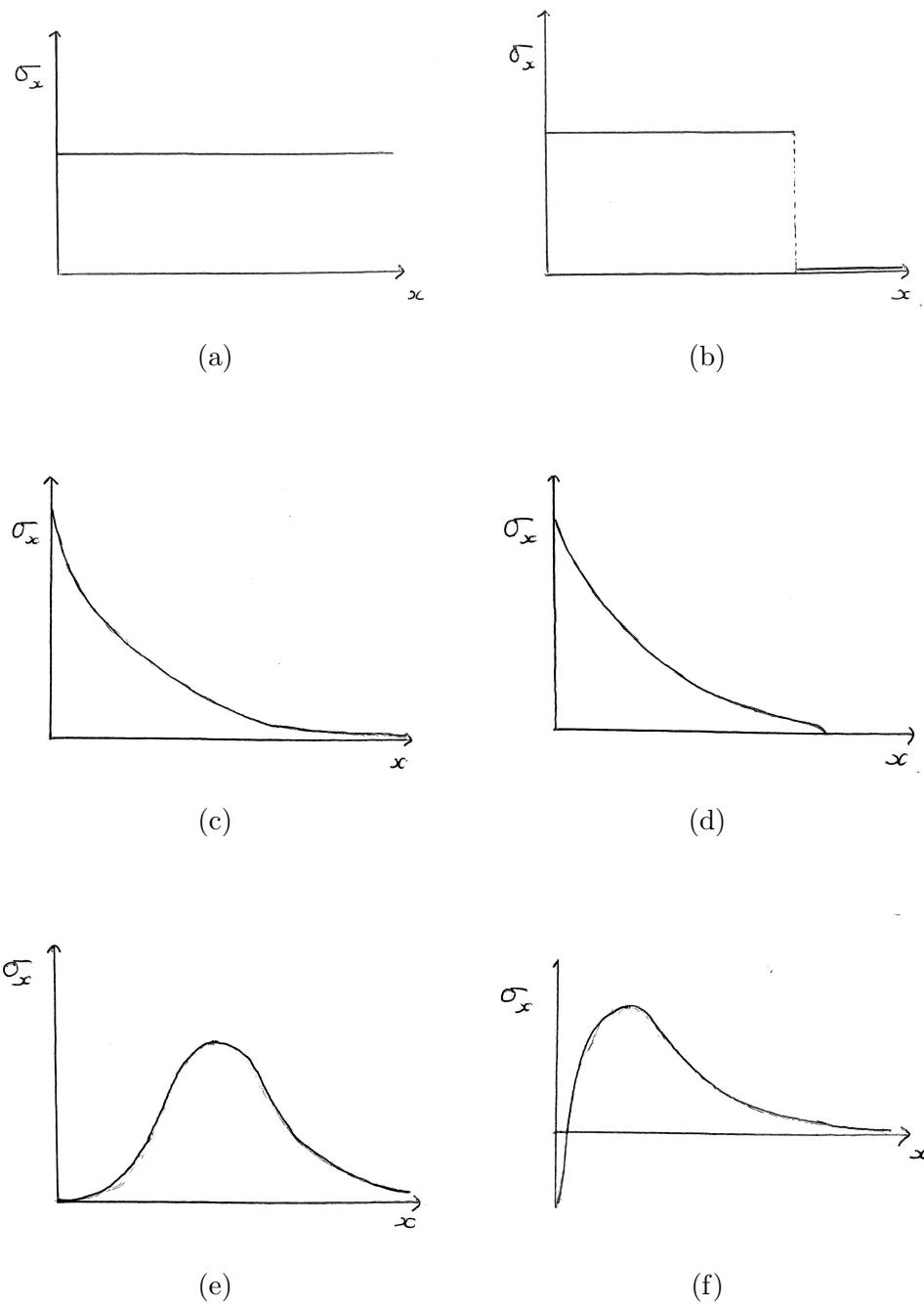
D.1.7 Parametric forms for the spatial change in stress

In the previous section, we considered several functional forms for how additional stress might be allocated through time. In this section we will consider how additional stress might be allocated over space. As a fist step we will consider the additional stress to be isotropic, so that it is a function of distance from a point source, the location of a previous event. Figure D.1.4 shows several potential functional forms for additional stress as a function of distance from the triggering earthquake. Over space, unlike for time, it is the amount of additional stress that impacts the intensity

function not the stress rate.

In the first function, shown in Figure D.1.4a, the stress is increased equally at all distances. In the second function, shown in Figure D.1.4b, the stress is increased equally but only within a fixed radius. The first of these does not seem sensible, because we would expect that the additional stress should be negligible at very large distances. The second seems more reasonable but we would also expect that additional stress to be a smooth, decreasing function of distance. The second function does not satisfy either of these, as it is piece-wise constant and contains a discontinuity. Figures D.1.4c and D.1.4d present functional forms for the additional stress which lead to intensity responses that are similar to the ETAS model. In these functions, additional stress is a continuous function of distance and is strictly decreasing when the additional stress is non-zero. At very large distances these functions respectively allow the additional stress to either be very small or zero. Both of these functional forms would be possible if the additional stress were allocated over space proportional to a generalised Pareto density. The functions presented in Figures D.1.4e and D.1.4f have additional stress peak some distance away from the previous event and then deplete as distance increases beyond this value. At very short distances these models respectively leave the stress unchanged or reduce stress. A justification for the decentralised additional stress might be that earthquakes releases some local stress which is either expressed in the earthquake or passed on to neighbouring faults.

Of these options, the generalised Pareto form seems like the least controversial choice as it is the most similar in structure to the currently used ETAS model. Following discussions with Shell, it was advised that a heavy-tailed generalised-Pareto allocation of the additional stress over space is consistent with the idea of previous events causing elastic dislocations within the reservoir. The generalised Pareto form seems to provide a sensible model for initial development of a combined aftershock model.

Figure D.1.4: Proposed forms for additional stress as a function of distance x .

D.1.8 Parametric forms for the effect of magnitude on additional stress

In the previous sections, we considered how additional stress is added through space and time. Another question of interest is how the magnitude of an earthquake impacts the total amount of additional stress that it causes. This is the purpose of the function σ_m . Since both σ_x and σ_t are density functions, these both integrate to one over their respective supports and the total additional stress from an earthquake is completely controlled by σ_m . Figure D.1.5 shows six suggestions for the form of this function.

If small magnitude earthquakes are not thought to add to the local stress state then there may be a threshold involved in the relationship; this is included in models shown in red in Figure D.1.5. This type of threshold is assumed in the ETAS model, where the threshold is chosen to be equal to the magnitude of completion but it could equally be chosen as a greater magnitude. Indeed, this will be the case in the latter portion of the Groningen catalogue since the magnitude of completion has been reduced. Above this chosen threshold level, increasing magnitude could change the stress state in a number of ways. Magnitude might have no effect as in Figure D.1.5a, a linear effect as in Figure D.1.5b, or exponential effect on the total additional stress as in Figure D.1.5c.

The exponential increase seems most sensible, since magnitudes are measured on a logarithmic scale. Since magnitude is a measure of the energy released by an earthquake and stress has units $\text{Nm}^{-2} = \text{Jm}^{-3}$ it seems reasonable that a ten-fold increase in the energy released would lead to a ten-fold increase in additional stress. As with the other model components, using a generalised Pareto form here rather than an exponential would allow for more flexibility, in particular it can account for the pos-

sibility of a sub-exponential relationship, which may arise if not all energy that is released is propagated as additional stress.

A similar energy based argument could be used in support support an inverse power-law style of decay of additional stress with distance. If this set amount of stress is dissipated radially, then it would seem reasonable that it decrease in an inverse-square or inverse-cube relation to distance. Again, this can be generalised by using an inverse power-law, which would allow the power to be selected as part of model fitting, or using a generalised Pareto form which allows more localised allocation.

Following discussions with Shell, we were advised that an exponentially increasing form for σ_m is equivalent to the constant stress drop model for slipping faults. For this reason and the similarities to the ETAS model it would be prudent to initial develop a model using an exponential form. At a later point the need for a more flexible model could be investigated.

D.1.9 Summary

This appendix has introduced some initial ideas about the form and implementation of a ETAS-style model which provides a coherent treatment for the source of both induced earthquakes and their aftershocks. While the combined modelling approach for all earthquakes is conceptually appealing, the resulting model is likely to presents many of the same challenges as direct estimation of the ETAS model: highly correlated parameters, issues with numerical optimisation routines, and a likelihood function that is computationally expensive to evaluate. Additionally, this combined aftershock model requires extensive use of numerical integration, which further increases computational burden, and the model is not amenable to the conditional inference approaches which facilitate estimation of the ETAS model.

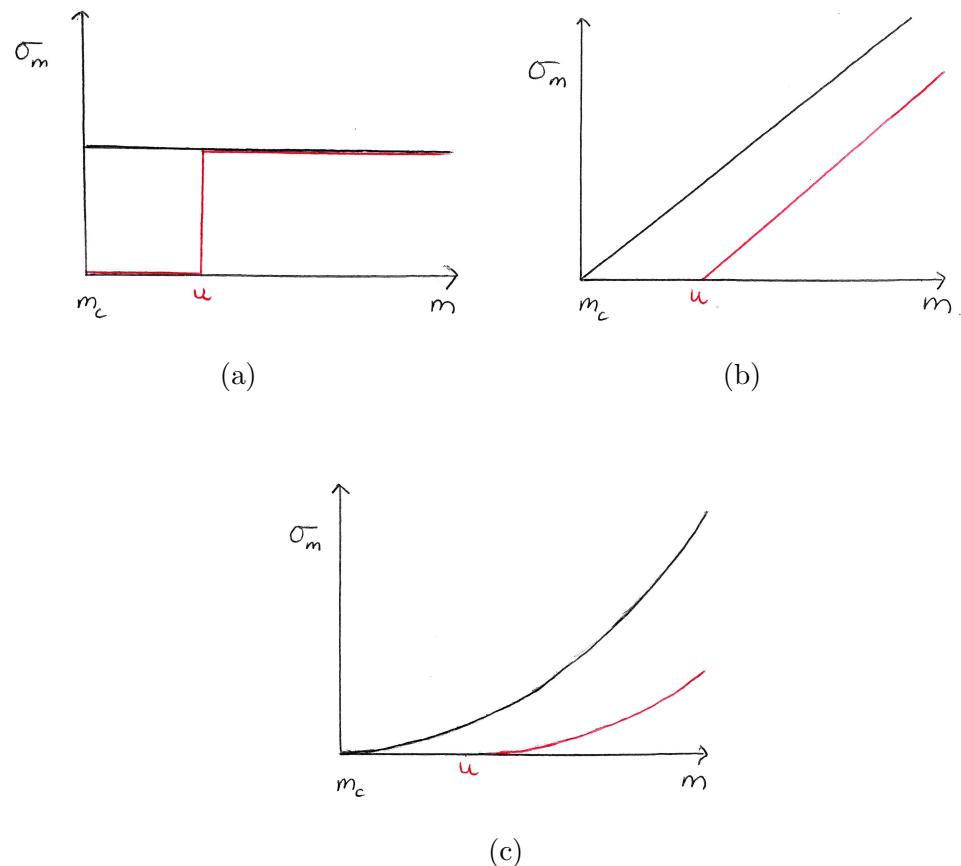


Figure D.1.5: Proposed forms for the relationship between magnitude and total additional stress. Additional stress added by events above the magnitude of completion M_c (black) or some higher magnitude u (red).

It is for the reasons listed above that we have chosen not to develop this model further, instead to focus on extending and improving inference for the ETAS model in Chapter 6 of this thesis. We present these initial ideas here in the hope that they might be informative to anyone aiming to develop a similar model in the future.

Bibliography

- Aki, K. (1965). Maximum likelihood estimate of b in the formula $\log n = a - bm$ and its confidence limits. (<http://hdl.handle.net/2261/12198>).
- Baddeley, A. (2008). Analysing spatial point patterns in R. Technical report, CSIRO, 2010. Version 4. (<https://research.csiro.au/software/r-workshop-notes>).
- Baddeley, A., Bárány, I., and Schneider, R. (2007). *Spatial Point Processes and their Applications*. Springer.
- Baddeley, A., Møller, J., and Waagepetersen, R. (2000). Non-and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54(3):329–350.
- Basrak, B., Wintenberger, O., and Žugec, P. (2019). On the total claim amount for marked poisson cluster models. *Advances in Applied Probability*, 51(2):541–569.
- Båth, M. (1965). Lateral inhomogeneities of the upper mantle. *Tectonophysics*, 2(6):483–514.
- Beirlant, J., Kijko, A., Reynkens, T., and Einmahl, J. H. (2019). Estimating the maximum possible earthquake magnitude using extreme value methodology: the Groningen case. *Natural Hazards*, 98(3):1091–1113.

- Benali, A., Peresan, A., Varini, E., and Talbi, A. (2020). Modelling background seismicity components identified by nearest neighbour and stochastic declustering approaches: the case of northeastern italy. *Stochastic Environmental Research and Risk Assessment*, 34:775–791.
- Bierman, S., Kraaijeveld, F., and Bourne, S. (2015). Regularised direct inversion to compaction in the Groningen reservoir using measurements from optical leveling campaigns. *NAM report SR*, 15:11194.
- Bourne, S. and Oates, S. (2015). An activity rate model of seismicity induced by reservoir compaction and fault reactivation in the Groningen gas field. NAM report.
- Bourne, S. and Oates, S. (2017a). Extreme threshold failures within a heterogeneous elastic thin sheet and the spatial-temporal development of induced seismicity within the Groningen gas field. *Journal of Geophysical Research: Solid Earth*, 122(12):10–299.
- Bourne, S. and Oates, S. (2017b). Induced seismicity within the Groningen gas field [data set]. *Zenodo*. <https://doi.org/10.5281/zenodo.103522>.
- Bourne, S., Oates, S., and Van Elk, J. (2018). The exponential rise of induced seismicity with increasing stress levels in the Groningen gas field and its implications for controlling seismic risk. *Geophysical Journal International*, 213(3):1693–1700.
- Bourne, S., Oates, S., van Elk, J., and Doornhof, D. (2014). A seismological model for earthquakes induced by fluid extraction from a subsurface reservoir. *Journal of Geophysical Research: Solid Earth*, 119(12):8991–9015.
- Bourne, S. J. and Oates, S. (2020). Stress-dependent magnitudes of induced earthquakes in the Groningen gas field. *Journal of Geophysical Research: Solid Earth*, 125(11):e2020JB020013.

- Bourne, S. J. and Oates, S. J. (2017c). Development of statistical geomechanical models for forecasting seismicity induced by gas production from the Groningen field. *Netherlands Journal of Geosciences*, 96(5):s175–s182.
- Brown, E. N., Barbieri, R., Ventura, V., Kass, R. E., and Frank, L. M. (2002). The time-rescaling theorem and its application to neural spike train data analysis. *Neural Computation*, 14(2):325–346.
- Cai, J.-J., Wan, P., and Ozel, G. (2021). Parametric and non-parametric estimation of extreme earthquake event: the joint tail inference for mainshocks and aftershocks. *Extremes*, 24(2):199–214.
- Carter, D. and Challenor, P. (1981). Estimating return values of environmental parameters. *Quarterly Journal of the Royal Meteorological Society*, 107(451):259–266.
- Chavez-Demoulin, V. and Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(1):207–222.
- Chavez-Demoulin, V., Davison, A. C., and McNeil, A. J. (2005). Estimating value-at-risk: a point process approach. *Quantitative Finance*, 5(2):227–234.
- Coles, S., Pericchi, L. R., and Sisson, S. (2003). A fully probabilistic approach to extreme rainfall modeling. *Journal of Hydrology*, 273(1-4):35–50.
- Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, New York.
- Coles, S. G. and Pericchi, L. (2003). Anticipating catastrophes through extreme value modelling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(4):405–416.

- Coles, S. G. and Tawn, J. A. (1996). A Bayesian analysis of extreme rainfall data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 45(4):463–478.
- Cox, D. R. and Isham, V. (1980). *Point Processes*, volume 12. CRC Press.
- Danielsson, J., de Haan, L., Peng, L., and de Vries, C. G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis*, 76(2):226–248.
- Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(3):393–425.
- Davison, M., Ita, J., Hofmann, R., Seli, P., and Ong, P. (2010). Ensuring production in the Malaysian F6 field using field monitoring and geomechanical forecasting. *Proceedings of the International Oil and Gas Conference and Exhibition in China*.
- Dempsey, D. and Suckale, J. (2017). Physics-based forecasting of induced seismicity at Groningen gas field, the Netherlands. *Geophysical Research Letters*, 44(15):7773–7782.
- Diggle, P. (1983). *Statistical Analysis of Spatial Point Patterns*. Mathematics in Biology. Academic Press.
- Dorai-Raj, S. (2001). First-and second-order properties of spatiotemporal point processes in the space-time and frequency domains. *Doctoral Dissertation, Virginia Polytechnic Institute and State University*.
- Dost, B., Goutbeek, F., van Eck, T., and Kraaijpoel, D. (2012). Monitoring induced seismicity in the North of the Netherlands: Status Report 2010. *KNMI scientific report*.

- Dost, B., Ruigrok, E., and Spetzler, J. (2017). Development of seismicity and probabilistic hazard assessment for the Groningen gas field. *Netherlands Journal of Geosciences*, 96(5):235–245.
- Eastoe, E. F. and Tawn, J. A. (2009). Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58(1):25–45.
- Fawcett, L. and Green, A. C. (2018). Bayesian posterior predictive return levels for environmental extremes. *Stochastic Environmental Research and Risk Assessment*, 32(8):2233–2252.
- Ferro, C. A. and Segers, J. (2003). Inference for clusters of extreme values. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):545–556.
- Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24:180–190.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. CRC press.
- Genest, C., Ghoudi, K., and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3):543–552.
- Gulia, L., Rinaldi, A. P., Tormann, T., Vannucci, G., Enescu, B., and Wiemer, S. (2018). The effect of a mainshock on the size distribution of the aftershocks. *Geophysical Research Letters*, 45(24):13–277.
- Gutenberg, B. and Richter, C. F. (1956). Earthquake magnitude, intensity, energy,

- and acceleration: (second paper). *Bulletin of the Seismological Society of America*, 46(2):105–145.
- Hadi Mehranpour, M., Hangx, S. J., and Spiers, C. J. (2020). Discrete element method modelling of Groningen reservoir compaction using a new contact model describing elastic and inelastic grain-scale interactions. In *EGU General Assembly Conference Abstracts*, page 13941.
- Hainzl, S. and Christoffersen, A. (2017). Testing alternative temporal aftershock decay functions in an etas framework. *Geophysical Journal International*, 210(2):585–593.
- Heffernan, J. E. and Tawn, J. A. (2001). Extreme value analysis of a large designed experiment: a case study in bulk carrier safety. *Extremes*, 4(4):359–378.
- Helmstetter, A. and Sornette, D. (2003). Båth’s law derived from the Gutenberg-Richter law and from aftershock properties. *Geophysical Research Letters*, 30(20).
- Hol, S., Van der Linden, A., Bierman, S., Marcelis, F., and Makurat, A. (2018). Proxies for quantifying depletion-induced reservoir compaction in the Groningen field. In *52nd US Rock Mechanics/Geomechanics Symposium*. OnePetro.
- Hutton, K., Woessner, J., and 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.
- Joe, H. (2014). *Dependence Modeling with Copulas*. CRC press.
- Jonathan, P., Randell, D., Wadsworth, J., and Tawn, J. (2021). Uncertainties in return values from extreme value analysis of peaks over threshold using the generalised Pareto distribution. *Ocean Engineering*, 220,107725.

- Kijko, A. and Singh, M. (2011). Statistical tools for maximum possible earthquake magnitude estimation. *Acta Geophysica*, 59(4):674–700.
- Kirchner, M. (2017). *Perspectives on Hawkes Processes*. PhD thesis, ETH Zürich.
<https://doi.org/10.3929/ethz-b-000161487>.
- KNMI (2020). Aardbevings catalogus. (<https://www.knmi.nl/kennis-en-datacentrum/dataset/aardbevingscatalogus>).
- Kolev, A. A. and Ross, G. J. (2020). Semiparametric Bayesian forecasting of spatial earthquake occurrences. *arXiv preprint arXiv:2002.01706*.
- Kostrov, V. (1974). Seismic moment and energy of earthquakes, and seismic flow of rock. *Physics of the Solid Earth*, 1:13–21.
- Laio, F. (2004). Cramer–von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters. *Water Resources Research*, 40(9).
- Ledford, A. W. and Tawn, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(2):475–499.
- Limbeck, J., Bisdom, K., Lanz, F., Park, T., Barbaro, E., Bourne, S., Kiraly, F., Bierman, S., Harris, C., Nevenzeel, K., et al. (2021). Using machine learning for model benchmarking and forecasting of depletion-induced seismicity in the groningen gas field. *Computational Geosciences*, 25(1):529–551.
- Little, R. J. and Rubin, D. B. (2019). *Statistical Analysis with Missing Data*, volume 793. John Wiley & Sons.
- Lombardi, A. M. (2002). Probabilistic interpretation of Båth’s law. *Annals of Geophysics*, 45(3-4):455–472.

- Marcon, E. and Puech, F. (2009). Generalizing Ripley's k function to inhomogeneous populations. HAL (<https://halshs.archives-ouvertes.fr/halshs-00372631>).
- Marzocchi, W., Spassiani, I., Stallone, A., and Taroni, M. (2019). How to be fooled searching for significant variations of the b-value. *Geophysical Journal International*, 220(3):1845–1856.
- McGarr, A. (1976). Dependence of magnitude statistics on strain rate. *Bulletin of the Seismological Society of America*, 66(1):33–44.
- Mignan, A., Werner, M., Wiemer, S., Chen, C., and Wu, Y. (2011). Bayesian estimation of the spatially varying completeness magnitude of earthquake catalogs. *Bulletin of the Seismological Society of America*, 101(3):1371–1385.
- Mignan, A. and Woessner, J. (2012). Estimating the magnitude of completeness for earthquake catalogs. *Community Online Resource for Statistical Seismicity Analysis*.
- Molkenthin, C., Donner, C., Reich, S., Zöller, G., Hainzl, S., Holschneider, M., and Opper, M. (2020). GP-ETAS: semiparametric Bayesian inference for the spatio-temporal Epidemic Type Aftershock Sequence model. *arXiv preprint arXiv:2005.12857*.
- NAM (2016a). Technical addendum to the Winningsplan Groningen 2016, Document 1: Summary and Production. Technical report, NAM.
- NAM (2016b). Technical addendum to the Winningsplan Groningen 2016, Document 2: Subsidence. Technical report, NAM.
- NAM (2016c). Technical addendum to the Winningsplan Groningen 2016, Document 3: Hazard. Technical report, NAM.

NAM (2016d). Technical addendum to the Winningsplan Groningen 2016, Document 4: Risk. Technical report, NAM.

NAM (2016e). Technical addendum to the Winningsplan Groningen 2016, Document 5: Damage and Appendices. Technical report, NAM.

NAM (2017). English information site. <https://www.nam.nl/english-information.html>. (26-07-2017).

Northrop, P. J. and Jonathan, P. (2011). Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights. *Environmetrics*, 22(7):799–809.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9–27.

Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2):379–402.

Ogata, Y. (2011). Significant improvements of the space-time ETAS model for forecasting of accurate baseline seismicity. *Earth, Planets and Space*, 63(3):6.

O'Hagan, A. and Forster, J. J. (2004). *Kendall's Advanced Theory of Statistics, volume 2B: Bayesian Inference*. Arnold.

Omori, F. (1894). On the aftershocks of earthquakes. *Journal of the College of Science, Imperial University of Tokyo*, 7:111–120.

Paap, B., Carpentier, S., van Maanen, P.-P., and Meekes, S. (2020). A neural network approach for improved seismic event detection in the Groningen gas field, the Netherlands. *arXiv preprint arXiv:2001.07027*.

- Paleja, R. and Bierman, S. (2016). Measuring changes in earthquake occurrence rates in Groningen. *Shell Global Solutions International BV (Amsterdam)*.
- Pawitan, Y. (2001). *In All Likelihood: Statistical Modelling and Inference using Likelihood*. Oxford University Press.
- Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, 3(1):119–131.
- Ramsay, J. (2006). *Functional Data Analysis*. Springer, 2nd ed. edition.
- Reinhart, A. (2018). A review of self-exciting spatio-temporal point processes and their applications. *Statistical Science*, 33(3):299–318.
- Richter, G., Hainzl, S., Dahm, T., and Zöller, G. (2020). Stress-based, statistical modeling of the induced seismicity at the Groningen gas field, the Netherlands. *Environmental Earth Sciences*, 79:1–15.
- Rohrbeck, C., Eastoe, E. F., Frigessi, A., and Tawn, J. A. (2018). Extreme value modelling of water-related insurance claims. *Annals of Applied Statistics*, 12(1):246–282.
- Ross, G. (2016). Bayesian estimation of the ETAS model for earthquake occurrences. *Technical Report*.
- Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT–Statistical Journal*, 10(1):33–60.
- Schoenberg, F. P. (2013). Facilitated estimation of ETAS. *Bulletin of the Seismological Society of America*, 103(1):601–605.

- Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., and De Freitas, N. (2015). Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175.
- Sharkey, P. and Tawn, J. A. (2017). A Poisson process reparameterisation for Bayesian inference for extremes. *Extremes*, 20(2):239–263.
- Shcherbakov, R., Zhuang, J., Zöller, G., and Ogata, Y. (2019). Forecasting the magnitude of the largest expected earthquake. *Nature communications*, 10(1):1–11.
- Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*, volume 26. CRC press.
- Smith, J. D., White, R. S., Avouac, J.-P., and Bourne, S. (2020). Probabilistic earthquake locations of induced seismicity in the Groningen region, the Netherlands. *Geophysical Journal International*, 222(1):507–516.
- Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90.
- Smith, R. L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science*, 4(4):367–377.
- Snoek, J., Larochelle, H., and Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, pages 2951–2959.
- Stallone, A. and Marzocchi, W. (2019). Empirical evaluation of the magnitude-independence assumption. *Geophysical Journal International*, 216(2):820–839.
- TNO (2017). Geological survey of the Netherlands, Groningen gasfield. (<http://nlog.nl/en/groningen-gasfield>).

- Utsu, T. (1957). Magnitude of earthquakes and occurrence of their aftershocks. *Zisin, Ser. 2*, 10:35–45.
- Utsu, T., Ogata, Y., and Matsu'ura, R. S. (1995). The centenary of the Omori formula for a decay law of aftershock activity. *Journal of Physics of the Earth*, 43(1):1–33.
- van Elk, J., Doornhof, D., Bourne, S., Oates, S., Bommer, J., Visser, C., van Eijs, R., and van den Bogert, P. (2013). Technical addendum to the Winningsplan Groningen 2013. Technical report, NAM.
- van Stiphout, T., Zhuang, J., and Marsan, D. (2012). Seismicity declustering. *Community Online Resource for Statistical Seismicity Analysis*.
- van Thienen-Visser, K. and Breunese, J. (2015). Induced seismicity of the Groningen gas field: History and recent developments. *The Leading Edge*, 34(6):664–671.
- van Thienen-Visser, K., Sijacic, D., van Wees, J., Kraaijpoel, D., and Roholl, J. (2016). TNO report R10425. *SodM*.
- Varini, E., Peresan, A., and Zhuang, J. (2020). Topological comparison between the stochastic and the nearest-neighbor earthquake declustering methods through network analysis. *Journal of Geophysical Research: Solid Earth*, 125(8):e2020JB019718.
- Veen, A. and Schoenberg, F. P. (2008). Estimation of space–time branching process models in seismology using an EM–type algorithm. *Journal of the American Statistical Association*, 103(482):614–624.
- Vere-Jones, D. (2010). How to educate yourself as a statistical seismologist. *Community Online Resource for Statistical Seismicity Analysis*.
- Wadsworth, J., Tawn, J., and Jonathan, P. (2010). Accounting for choice of measurement scale in extreme value modeling. *Annals of Applied Statistics*, 4:1558–1578.

- Wadsworth, J. L. (2016). Exploiting structure of maximum likelihood estimators for extreme value threshold selection. *Technometrics*, 58(1):116–126.
- Wadsworth, J. L. and Tawn, J. A. (2012). Likelihood-based procedures for threshold diagnostics and uncertainty in extreme value modelling. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3):543–567.
- Waheed, U. b., Shaheen, A., Fehler, M., and Fulcher, B. (2020). Winning with simple learning models: Detecting earthquakes in Groningen, the Netherlands. *arXiv preprint arXiv:2007.03924*.
- Walker, A. J. (1977). An efficient method for generating discrete random variables with general distributions. *ACM Transactions on Mathematical Software (TOMS)*, 3(3):253–256.
- Wiemer, S. and Wyss, M. (2000). Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the western United States, and Japan. *Bulletin of the Seismological Society of America*, 90(4):859–869.
- Wilson, S. (2020). *ParBayesianOptimization: Parallel Bayesian Optimization of Hyperparameters*. R package version 1.2.1.
- Woessner, J. and 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.
- Zhuang, J., Harte, D. S., Werner, M. J., Hainzl, S., and Zhou, S. (2012). Basic models of seismicity: Temporal models. *Community Online Resource for Statistical Seismicity Analysis*.
- Zhuang, J., Ogata, Y., and Vere-Jones, D. (2002). Stochastic declustering of space-

- time earthquake occurrences. *Journal of the American Statistical Association*, 97(458):369–380.
- Zhuang, J., Werner, M. J., Hainzl, S., Harte, D., and Zhou, S. (2011). Basic models of seismicity: Spatiotemporal models. *Community Online Resource for Statistical Seismicity Analysis*.
- Zöller, G. and Holschneider, M. (2016). The maximum possible and the maximum expected earthquake magnitude for production-induced earthquakes at the gas field in Groningen, the Netherlands. *Bulletin of the Seismological Society of America*, 106(6):2917–2921.
- Žugec, P. (2019). *Marked Poisson Cluster Processes and Applications*. PhD thesis, University of Zagreb. Faculty of Science. Department of Mathematics.