University College London

Department of Statistical Science

PhD Statistical Science

Extensions of Self-Exciting Point processes with Applications in Seismology and Ecology

by Aleksandar Atanasov Kolev

supervised by
Dr. Gordon Ross
Prof. Richard Chandler

September 27, 2020

Declaration of Authorship

I, Aleksandar Kolev, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgements

This work would not have been possible without the support of a number of organisations and individuals. I would like to thank ESPRC for the provision of PhD funding to UK graduates. Without their financial support I would not have been able to support my studies. Further, I would like to thank Yonita Carter and Prof. Philip Treleaven for accepting me into their family of PhD candidates, for their mentorship and ability to address critical issues in a timely manner.

I am grateful to my alma mater, University College London, for the provision of infinite opportunities for personal and professional development; knowledge acquisition; working in a vibrant and diverse environment; meeting so many incredible people. I would like to thank Prof. Serge Guillas who not only welcomed me at UCL but also directed me towards my primary supervisor - Dr. Gordon Ross.

I would like to thank Gordon for more things that I can summarise with words. He is an endless source of knowledge, combined with out-of-the box work flow and an incredible ability to amuse. We faced and overcame a lot of challenges together. All this developed me as a researcher, educator and individual. I wish I had more time to work with you personally, do more projects, and produce even higher quality work. However, while on the topic of quality work I cannot continue the logical flow of facts without mentioning Prof. Richard Chandler, my second supervisor. Thank you for being so incredibly thorough. Your way of work thought me how to write properly, be extremely self-critical and research up to the highest standards.

My research would not have been possible without the support of my non-academic 'supervisors'- my family and friends. I am both grateful and guilty to my family and more specifically to my parents and grandparents. I highly appreciate their immense support and ability to handle me in the occasions where I am unbearable. Thank you for raising the man that I am, thank you for supporting me in my crazy endeavours

and for allowing me to pursue my dreams. I am sorry that I missed so many important moments in the last several years.

Some say that friends are just like family. This applies directly to a few people in my inner circle who are the go-to in almost every situation. Thank you for being such good listeners, distractors and supporters. Among them all, I would like to thank especially Viktoria Nikolova for the support. I would like to also thank Dr. Ilko Marinov who saved my life numerous times and provided me with the much needed care when the NHS was completely incapable of acting adequately. There are many other people including Simeon, Simon, Petya, Krasi, Kiril, Val, Hristo, Xiaochen, Theo, Anna, Marta and others who contributed a lot to keep me sane during this insane endeavour.

I few honourable mentions from academia. I would like to thank the two anonymous reviewers from the Statistics and Computing journal for the detailed comments on the paper associated with the third Chapter of this thesis. Their immense contribution helped me address very critical issues that resulted in substantial improvements to my work. I am grateful to Dr. Katerina Stavrianaki for the constructive feedback provided during my upgrade Viva. A big thank you to Prof. Janine Illian and Prof. Gianluca Baio for their time and efforts in reviewing and examining this thesis. I will remember my Viva for the rest of my life! It was a great pleasure. I am very grateful to Prof. Ruth King for the informal support in the very last stages of my thesis development and for the enriching conversations. Without her immense help I would have probably never finished this work. The intellectually stimulating discussion with Prof. David Borchers were immensely helpful for me for which I would like to thank him. I would also like to express my appreciation to Prof. John McColl and Dr. Vincent Macaulay who offered me the first glance at Statistical research as my undergrad thesis supervisors. Last but not least, I would like to thank who ever is reading this, as I do not expect that many will end up doing it.

I know I missed a lot of important people - apologies in advance.

Abstract

This thesis focuses on extending the ETAS model. ETAS is a special case of the Hawkes process - a self-exciting point process that provides the opportunity for a multilayered intensity structure that addresses the rate of events as a function of previous events' history. Triggering and clustering behaviours are naturally captured. The most simplistic version of the Hawkes process takes into account a single temporal sequence. Additional features such as marks, spatial information, other labels and multivariate scenarios can be considered. In this thesis we contribute primarily to three main aspects of a Hawkes process - temporal, spatial and multivariate analyses. Each of these challenges were addressed by incorporating new functionalities into the base process. Then we also solved the emerging estimation needs.

We began by exploring a renewal immigration concept where the main (immigrant) events follow a non-Poissonian distribution that provides an inhomogeneous temporal ground modelling. Then we explored a non-parametric spatial kernel estimation for the inference of the main events spatial aggregation. This Bayesian density estimation relies on a Dirichlet process application in a multivariate Normal distribution mixture modelling. Finally, we explored the application of self-exciting process in the context of spatially explicit capturing data. We introduced discrete space, continuous time, multivariate Hawkes process that is tailored towards limited number of observations from multiple objects that share common behaviour.

The introduced models and methods suggest superior performance compared to conventional techniques. They are directly applicable to fields where spatio-temporal clustering is observed. Some of the examples include crime, financial indicators change, earthquake modelling, people and animal movement.

Impact Statement

Nowadays data analysis is an inherent part of every endeavour. Unseen patterns naturally trigger phenomenon development. Every *event* could potentially influence the occurrence of other events or initiate a stochastic change in the pattern that was present prior to its occurrence. Such patterns are hard to track although extremely beneficial for inferential purposes due to the inherent flexibility of the model construction. In our work we focus on one of the most popular model constructions for addressing such patterns – the Hawkes process.

Hawkes process is a method that describes natural occurring events' behaviour. Unlike some more advanced models, it also possess inherent statistical properties. Hawkes process can address complex patterns that are commonly hidden for standard modelling techniques. The specific results presented in this thesis can be used directly for mitigation of seismic hazards applicable to structural engineering and (re-)insurance purposes. Further, our methods provide a novel estimation of short term seismic hazards which enhances the forecasting probability of an aftershock tremor. This way the risk of human life loss can be mitigated further.

Bayesian Statistics on its own is a topic that is becoming very relevant to contemporary research. However, due to its demanding nature Hawkes processes are usually estimated in a more conservative, frequentist manner. Providing a framework for robust and resilient methods for Bayesian analysis improves considerably the understanding of parameter uncertainty on multiple levels. Hence, this uncertainty can be propagated towards the quantity of interest for the specific project. Having a clear perspective on the possible flaws of a target estimation can considerably increase our confidence in decision making and catastrophe prevention. Obtaining unreasonable parameter uncertainties drives engineers to increase unreasonably the embedded functionality in every component which increases unnecessary production costs.

The Hawkes process applications in ecology context are heavily under explored. Our work on animal movement outlines one of the many direct opportunities for major improvement of the underlying methodology in standard ecology techniques. We believe that the Hawkes process can be applied to vegetation spread, mutation and extinction; land irrigation, forestoration and aridation; animal species extinction, migration and adaptation.

We believe that this thesis will influence applications of the Hawkes process in other fields, further improve the Bayesian analysis application to the self-exciting point process and by this increase the awareness and attractiveness of the class of Hawkes processes to the wider analytical community.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis explores random processes describing the collection of point occurrences along time, space and other dimensions of interest. The observations in these sequences could represent, for example, coal mining disasters in the UK in the 19th century, a radioactive emission above a certain level in Europe, a financial loss or gain, customer's arrival and service time. In this work we explore two very different application fields that actually have similar underlying patterns that we address using hybrid Hawkes processes.

The first area of application that we consider is related to earthquake modelling. We would like to estimate the number of earthquakes that would occur in the next day or month based on the recently detected earthquakes. This is achieved by developing models in which an earthquake occurrence increases the short term (in space, time or both) earthquake occurrence probability. Such patterns are present in the earthquake data where the observations are typically clustered in groups. On Figure 1.1 are displayed the time and density of all earthquakes with magnitude greater than 3 that were observed in Bulgaria and Romania within 5 year (top) and 20 year (bottom) periods. If we consider only the 5 year catalogue we could identify two large clusters of seismic activity. However, the bottom figure clearly outlines that those peaks are negligible compared to the activity in 2004-2009. Thus the number of earthquakes in each group is only relative to those within close proximity (in this case time) as the number of detections varies between periods.

The second application field that we explore is related to animal movement. Consider a reserve in which animals (e.g. tigers) are kept to be preserved. Detecting the passage of an animal in close proximity to a stationary camera, within a spatio-temporal interval,

provides sufficient information to estimate animal abundance in the whole reserve. The exact spotting data, however, follows a similar cluster-based pattern - an animal detected once is more likely to be detected again on the same camera or on those nearby within some short period of time. If an animal is not active within the area in which the cameras are placed we might only expect occasional observations, clustered within the temporal span of each passage. For example, on Figure 1.2 are shown the detections of two randomly picked animals from an animal spotting data in the Nagarahole reserve (see Section 5.6.3). The animal detection information shown on the top figure has elevated detection density during two long periods while the other one (bottom figure) has considerably more aggregated detections within several shorter periods. An animal detection increases the short term probability of it being spotted again although the scale of this effect might differ between animals.

The Hawkes process first introduced by Hawkes (1971) is a widely used statistical model that addresses the concept of clustered point processes. Different extensions of this model have been applied to analyse various data from a vast range of areas such as credit risk [Errais et al., 2010], criminology [Mohler et al., 2011], finance [Chavez-Demoulin et al., 2005, Embrechts et al., 2011, Bacry et al., 2015], genome analysis [Reynaud-Bouret et al., 2010], neuroscience [Chornoboy et al., 1988], social interaction modelling [Crane and Sornette, 2008], terrorist activity modelling [Porter et al., 2012] and many others. Hawkes process's key feature is the definition of a hidden structure that addresses events' influence. The basic Hawkes process develops this relationship solely based on the event arrival times. However, a number of extensions are present towards multivariate analysis, marked processes and spatial analysis.

The general model describing clustering behaviour assigns all events into two possible categories - uncaused and caused events. The caused events, are also know as cluster centres, are a realisation of unobserved parent (ground) process. Each of them generates an offspring process centred at each of them that allows for further excitation from a multiple generations [Daley and Vere-Jones, 2003]. All events generated from the offspring process are referred to as caused events as they are triggered by the occurrence of another event in the sequence. Probably the most simplistic representation of such a process is when both cluster centres and offspring generations are drawn from a Poisson process. This is an example of Bartlett-Lewis [Neyman and Scott, 1952] or Neyman Scott [Rodriguez-Iturbe et al., 1987a, Rodriguez-Iturbe et al., 1987b] process depending on the spatial paradigm that is taken into account [Ritschel et al., 2017, Islam et al.,

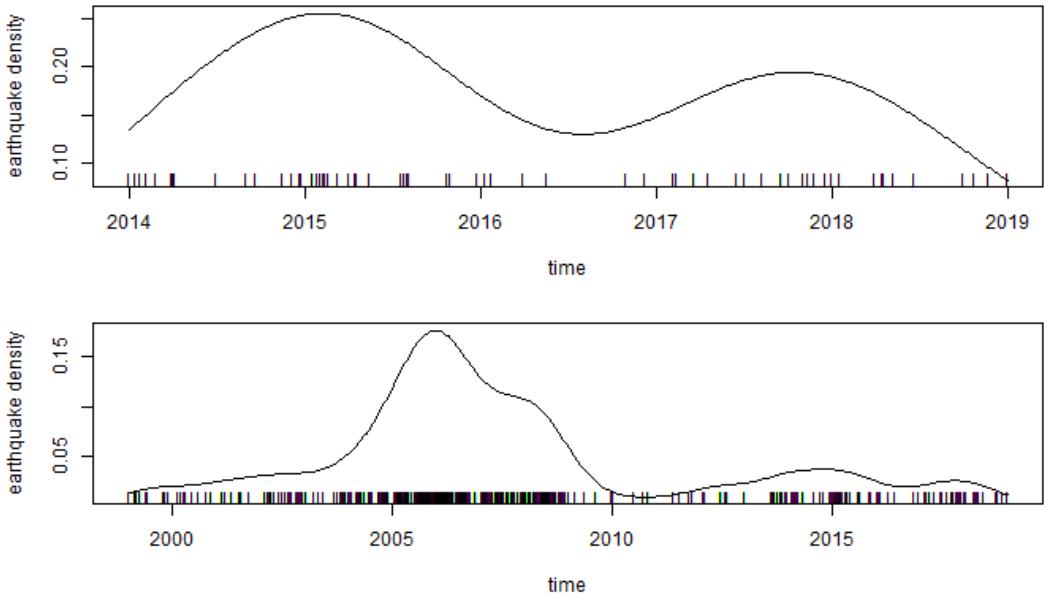

Figure 1.1: Earthquake temporal occurrence and density. Based on a catalogue that covers Bulgaria and Romania with magnitude larger than 3 within 2014-2019 (top) and 1999-2019 (bottom), where | indicates an earthquake occurrence.

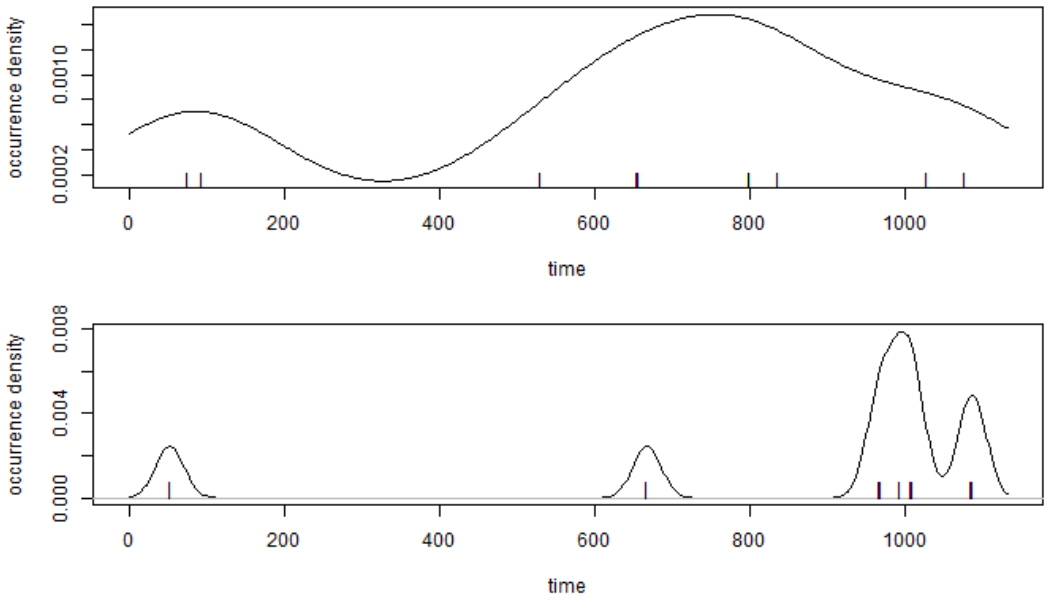

Figure 1.2: Animal spotting data for two different animals (top and bottom respectively) from the Nagarahole reserve. Here | indicates an animal spotting across the same spatio-temporal interval. There are 10 observations displayed on the top figure and 9 - on the bottom one.

1990]. The case in which the uncaused events follow a Poisson process while every event in the sequence can generate caused events that follow an uniform distribution is

referred to as Matérn cluster process. A general description of the most widely used cluster processes were reviewed by [González et al., 2016].

The prime focus of this thesis is to address novel extensions of the Hawkes process based on the Epidemic Type Aftershock Sequence (ETAS) model [Ogata, 1988]. This model is widely used in seismology [Rotondi and Varini, 2019, Ross, 2018a, Omi et al., 2014, Omi et al., 2015, Ebrahimian et al., 2013, Schoenberg, 2013, Vargas and Gneiting, 2012] to study earthquake arrival times. Most applications of the standard ETAS model rely on point estimates for the model parameters, which ignore the inherent uncertainty that arises as part of the model structure. Thus fitting ETAS to an earthquake catalogues can result in misleading forecasts that under or over estimate the process' multilayered intensity structure. In contrast, Bayesian statistics allows parameter uncertainty to be explicitly incorporated. These estimates can be used for detailed forecasts that characterise the process uncertainty in several stages such as uncaused and offspring events productivity, spatial aggregation and boundary condition based restrictions.

Despite ETAS' growing popularity in seismology, a Bayesian treatment of the ETAS model has been limited by the complex nature of the resulting posterior distribution, which makes it infeasible to apply to catalogues containing more than a few hundred earthquakes. To combat this, we develop a new framework for estimating the ETAS model in a fully Bayesian manner, which can be efficiently scaled up to large catalogues containing thousands of earthquakes. More details regarding Hawkes and ETAS processes are presented in Chapter 2.

The basic temporal ETAS model assumes that all uncaused events follow a Poisson process, with aftershocks triggered via a parametric kernel function. However, the Poissonian assumption contradicts with several aspects of seismological theory, which suggest that the arrival time of the main earthquakes that trigger all other earthquakes in the sequence instead follows alternative renewal distributions that address better the inherited seismic behaviour of strain accumulation and relaxation. In Chapter 3 the standard temporal ETAS process is extended to allow for non-Poissonian distributions by introducing a dependence based on the underlying process' behaviour. We introduce two fundamental model structures that can be further extended toward an ensemble of probability distributions. Then, we provide an illustrative application to two real earthquake catalogues that further hone the introduced model's benefits compared to the standard ETAS model. The introduced methods in Chapter 3 were published in *Statistics and Computing* journal [Kolev and Ross, 2019].

A further extension of the standard temporal ETAS model is to include spatial information and thus obtain a spatio-temporal point, rather than purely temporal, context. So far the literature has addressed its estimation primarily in a frequentist manner. The spatial ETAS model relies on the estimation of spatial density of the uncaused events in the catalogue. We address this problem in Chapter 4 using a mixture model based on a Dirichlet process with base measure the Normal-Inverse-Wishart distribution. Hence, we provide a Bayesian non-parametric estimation algorithm as part of a larger Bayesian framework for the parameter estimation of the model's parameters. Neither the number of the components nor their explicit centres are restricted in our framework which provides data-driven estimation with unregulated spatial clustering behaviour of the uncaused events. This approach direct extends the previously introduced spatial ETAS models [Ogata, 1998, Ogata and Zhuang, 2006, Schoenberg, 2013]. We also provide a kernel density estimate alternative of the model, as well as techniques for out of sample performance testing and forecasting quantification. A critical study on simulated data and real earthquake catalogues is reported.

In Chapter 5 we explore the possibility of an ETAS process application towards a discrete space, continuous time multivariate data. We address multiple subjects of interest that are observed at specific locations. Their spotting time is recorded. An example of such data structure is the spatially-explicit capture re-capture where animal movement is captured by traps (cameras) at specific locations over a period of time to quantify population dynamics. Each unique data object (animal) is considered to follow a separate Hawkes process, hence the inherent multivariate structure of the proposed model. However, a basic multivariate ETAS process is unlikely to be able to address such a pattern due to the limited number of observations for each animal. Hence, we propose a hybrid model that has some subject-specific components, while others are pulled across the entire population. We reported the performance of our method across two catalogues.

Chapter 2

Methods and Techniques

This Chapter reviews ETAS models, starting with some general theory of point processes and considering the specification and theoretical properties of the models, as well as issues related to inference and simulation.

2.1 Temporal Point process

Point processes are stochastic processes that can be used to represent patterns of points within a space of interest. Examples of such patterns include temporal occurrences such as arrival times of people, locations of plants within a specific area, and times and spatial locations of earthquakes. The first example is a temporal point process, the second is in space and the final one is in both space and time. This initial review focuses on the simplest of these cases: the temporal point process.

An accessible introduction on point processes is provided by [Cox and Isham, 1980], while a more advanced treatments can be found in [Daley and Vere-Jones, 2003, Daley and Vere-Jones, 2007].

2.1.1 Counting process

A point process is also referred to as a counting process as it counts the number of events that occur in an interval of interest. Consider a sequence of ordered data $0 < t_1 < ... < t_i < t_{i+1}$ for $i \in \mathbb{Z}$. The point process that models these data are associated with a random variable $N(a, b)$ that captures the number of observations within an interval (a, b) as follows:

$$N(0,t) = \sum_i \mathbb{1}_{t_i \in (0,t)}, \tag{2.1}$$

where $\mathbb{1}$ is an indicator function which takes value 1 if its corresponding statement is true and 0 otherwise. An alternative notation is used with respect to sets, where $N(A)$ records the number of observations within a set A as follows:

$$N(A) = \sum_i \mathbb{1}_{t_i \in A}.$$

We are going to use both of these interchangeably depending on the context with default values assumed to follow Equation 2.1 unless specified otherwise.

2.1.2 Intensity function

One of the most fundamental properties of a point process is its first-order intensity function, informally representing the rate of event occurrence. Formally, it can be represented with respect to the event detection probability in an infinitesimal interval:

$$\lambda^*(t) = \lim_{\epsilon \to 0+} \frac{E[N(t, t+\epsilon)]}{\epsilon}, \tag{2.2}$$

where $E[\cdot]$ denotes the expectation function.

Some point processes are memoryless and the occurrence of an event is unrelated to the occurrence of others. An example of such a process is provided in Section 2.1.4. In practice, however, there are many situations in which events are associated with each other. The occurrence of one or more events can increase directly the subsequent event rate. For example, in a population events might correspond to the birth times of new individuals who themselves can subsequently have offspring of their own. Alternatively, the event rate over time might fluctuate depending on some underlying process. This thesis focuses on associations of the first type. There are various ways of characterising such associations, but in the current context a particularly useful one is the conditional intensity function.

The conditional intensity function represents a modification of the intensity function (as of Equation 2.2) to account for the influence of the events that occur prior to the current time. It is defined with respect to the expected number of events in an infinitesimal interval, given the $\mathscr{H}_t$ - the point-process' history up to t:

$$\lambda(t|\mathscr{H}_t) = \lim_{\epsilon \to 0+} \frac{E[N(t, t+\epsilon)|\mathscr{H}_t]}{\epsilon}, \tag{2.3}$$

where $E[\cdot]$ denotes the expectation function.

An important class of point processes is specified explicitly via the conditional intensity function: these are Hawkes process. They address the inherent pattern of 'birth' of events. A special case of a Hawkes process is an ETAS model that is predominantly used in this thesis, which is illustrated in greater detail in Section 2.2. Hawkes process' conditional intensity $\lambda(t)$ depends on all data observations up to t, denoted with $\mathscr{H}_t$. In a slight abuse of notation, throughout this thesis the dependence on $\mathscr{H}_t$ will be suppressed and we write $\lambda(t) \equiv \lambda(t|\mathscr{H}_t)$ for the conditional intensity function. In this data-driven scheme, the occurrence of each event leads to an increase in the rate of occurrence of subsequent events, with the effect usually dying out gradually over time. The overall intensity function of a Hawkes process at any time point is a superposition of a baseline intensity together with the conditional intensities arising from all previously-occurring events.

The majority of the work in the thesis will focus on models that are specified in terms of the conditional intensity function. In general, it is necessary to impose restrictions on the form of this conditional intensity in order for the resulting process to be well-defined and to ensure that properties such as stationarity hold: these will be discussed in the context of the specific models introduced later.

Subsequently in this thesis, extensive use will be made of renewal processes in which the inter-arrival times between successful events are not identically distributed. A detailed introduction to renewal processes is provided by Cox (1962). Suppose the waiting intervals between successive events (W) follow a continuous distribution $f(\cdot)$ with a respective cumulative distribution function $F(\cdot)$, then Equation 2.3 with respect to a renewal process is equivalent to:

$$\begin{aligned}
\lambda(t) &= \lim_{\epsilon \to 0+} \frac{E[N(t, t+\epsilon)|\mathscr{H}_t]}{\epsilon} && (2.4)\\
&= \lim_{\epsilon \to 0+} \frac{P[N(t, t+\epsilon) > 0|\mathscr{H}_t]}{\epsilon} && (2.5)\\
&= \lim_{\epsilon \to 0+} \frac{P[t < W < t+\epsilon | W > t]}{\epsilon} && (2.6)\\
&= \lim_{\epsilon \to 0+} \frac{P[t < W < t+\epsilon]}{\epsilon} \frac{1}{P[W > t]} = \frac{f(t)}{1 - F(t)}. && (2.7)
\end{aligned}$$

2.1.3 Stationary process

An important class of point processes are those that are stationary in a sense that their stochastic structure is unchanged over time. The formal definition of a stationary point process given in Definition 3.2.I. by Daley and Vere-Jones (2003b) states that a point process is stationary if for every $s = 1, 2...$ and all subsets $A_1, ..., A_s$ on the real line, the joint distribution of

$$\{N(A_1 + t), ..., N(A_r + t)\}$$

does not depend on $t \in (-\infty, \infty)$. An immediate consequence is the distribution of the number of events in an interval depends on the length of the interval, not on its location.

2.1.4 Poisson process

Let us describe the relationship in Equation 2.4 with respect to a simple point process - the Poisson process. This process is defined based on the following four properties [Ross, 2014]:

(i) $N(0) = 0$.

(ii) $P(N(t, t + \delta) = 1) = \lambda\delta + o(\delta)$. We write $x = o(\delta)$ if $\lim_{\delta \to 0} x/\delta = 0$.

(iii) $N(A)$ and $N(B)$ are independent for disjoint sets A and B.

(iv) $P(N(t, t + \delta) > 1) = o(\delta)$.

The probability distributions of the number of events in any interval of length t are the same. The Poisson process is a memoryless process. This implies that the dependence in Equation 2.1 can be simplified to $P[N(a, a + t) = k] = P[N(0, t) = k] = P[N(t) = k]$ for $k \in \{\mathbb{N}, 0\}$. Let $p_n(t)$ be the probability of observing n events in a temporal interval with length t. The distribution of $p_n(t)$ is Poisson with mean λt:

$$p_n(t) = P[N(t) = n] = \frac{(\lambda t)^n e^{-\lambda t}}{n!}.$$

Let $T_1, T_2, \ldots$ be the times of successive events in a Poisson process of rate λ, starting at time zero. Respectively $W_n = T_n - T_{n-1}$ is the waiting time between event occurrence, with $T_0 = 0$. The cumulative distribution function of W_n is

$$F(t) = P[W_n \leq t] = P[T_1 \leq t] = 1 - p_0(t) = 1 - e^{-\lambda t},$$

with a corresponding density function of

$$f(t) = \frac{dF(t)}{dt} = \frac{d}{dt}\left[1 - e^{-\lambda t}\right] = \lambda e^{-\lambda t} \sim \text{Exp}(\lambda),$$

which represents the density function of an exponential distribution with rate λ. Hence, the waiting times between event occurrences of a Poisson process follow an Exponential distribution. Based on Equation 2.4 the intensity of a Poisson process is

$$\lambda(t) = \frac{f(t)}{1 - F(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda,$$

which recovers the intensity that the process originates from. An illustrative example of a simulated Poisson process is shown on Figure 2.1. These data consist of 1000 observations (events) from a Poisson process with intensity $\lambda = 1$. As expected, the events' density is relatively constant.

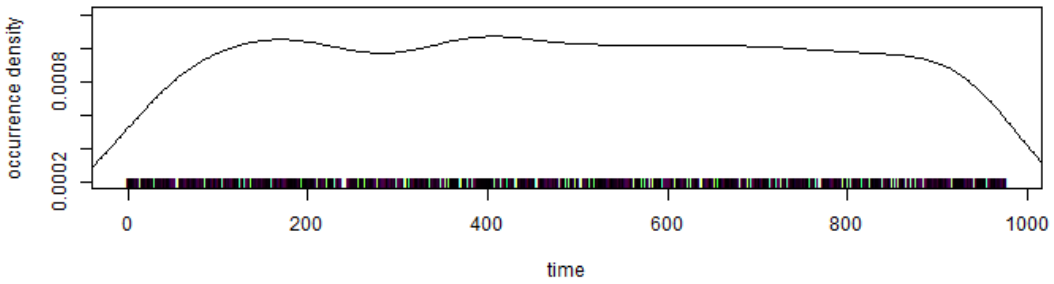

Figure 2.1: Example of a Poisson process based on 1000 sampled events with $\lambda = 1$, where | indicates an event.

A useful Poisson process property is that event arrival times in an interval with length t follow an uniform distribution given $N(t) = n$. A derivation of this result is as follows:

For $k \geq 1$, the joint density of the first k event times can be written as

$$f^{(k)}(t_1, \ldots, t_k) = f_1(t_1)\, f_2(t_2|T_1 = t_1) \ldots f_k(t_k|T_1 = t_1, \ldots, T_{k-1} = t_{k-1}) \ , \qquad (2.8)$$

based on the generalised multiplication law, where $0 < t_1 < \ldots < t_k$. Now $T_1 \sim \text{Exp}(\lambda)$ so that $f_1(t_1) = \lambda \exp(-\lambda x)$ for $t_1 > 0$. Also, for $i > 1$, $T_i - T_{i-1} \sim \text{Exp}(\lambda)$ so that

$$f_i(t_i|T_1 = t_1, \ldots, T_{i-1} = t_{i-1}) = f_i(t_i|T_{i-1} = t_{i-1}) = \lambda \exp\left[-\lambda\left(t_i - t_{i-1}\right)\right]$$

for $t_i > t_{i-1}$. It follows that the joint density as of Equation 2.8 is

$$\begin{aligned} f^{(k)}(t_1, \ldots, t_k) &= \lambda^k \exp[-\lambda t_1] \prod_{i=2}^{k} \exp[-\lambda(t_i - t_{i-1})] \\ &= \lambda^k \exp[-\lambda t_k] \ . \end{aligned} \tag{2.9}$$

Consider the joint density of the event times $T_1, \ldots, T_n$ conditional on $N(t) = n$, where $N(t)$ is again the number of events in the interval $(0, t)$. This conditional joint density has support on the region $0 < t_1 \ldots < t_n < t$, and can be defined as

$$\lim_{\delta t_1 \to 0, \ldots, \delta t_n \to 0} \frac{P[T_1 \in (t_1, t_1 + \delta t_1), \ldots, T_n \in (t_n, t_n + \delta t_n) \,|N(t) = n]}{\delta t_1 \ldots \delta t_n} \ . \tag{2.10}$$

Noting that $N(t) = n$ if and only if $T_n < t < T_{n+1}$ and that the first of these inequalities is automatically satisfied by the support of the conditional joint density, the numerator here is:

$$\begin{aligned} & \frac{P[T_1 \in (t_1, t_1 + \delta t_1), \ldots, T_n \in (t_n, t_n + \delta t_n), N(t) = n]}{P[N(t) = n]} \\ = \ & \frac{P[T_1 \in (t_1, t_1 + \delta t_1), \ldots, T_n \in (t_n, t_n + \delta t_n), T_{n+1} > t]}{P[N(t) = n]} \\ = \ & \frac{\int_{t_{n+1}=t}^{\infty} f^{(n+1)}(t_1, \ldots, t_k)\, dt_{n+1} \delta t_1 \ldots \delta t_n + o(\delta t_1 \ldots \delta t_n)}{P[N(t) = n]} \\ = \ & \frac{\int_{t_{n+1}=t}^{\infty} \lambda^{n+1} \exp[-\lambda t_{n+1}]\, dt_{n+1} \delta t_1 \ldots \delta t_n + o(\delta t_1 \ldots \delta t_n)}{(\lambda t)^n \exp[-\lambda t]/n!} \\ = \ & \frac{\lambda^n \exp[-\lambda t]\, \delta t_1 \ldots \delta t_n + o(\delta t_1 \ldots \delta t_n)}{(\lambda t)^n \exp[-\lambda t]/n!} \ , \end{aligned}$$

the penultimate step following from Equation 2.9 and the fact that $N(t) \sim \text{Poi}(\lambda t)$. Simplifying and substituting into Equation 2.10, the required conditional joint density is:

$$\lim_{\delta t_1 \to 0, \ldots, \delta t_k \to 0} \frac{n!/t^n \left[\delta t_1 \ldots \delta t_n + o(\delta t_1 \ldots \delta t_n)\right]}{\delta t_1 \ldots \delta t_n} = \frac{n!}{t^n} \quad (0 < t_1 < \ldots < t_n < t) \ .$$

This expression corresponds to the joint density of the order statistics from n independent $U(0, t)$ random variables.

The simplicity of a Poisson process is beneficial for obtaining computationally undemanding results which makes it a natural starting point for the development of more complex processes such as the ETAS model.

The Poisson process is an example of a renewal process in which the inter-event times are exponential. In Section 3.3 we extend this to consider alternative waiting-time distributions that allow us to construct more flexible models which improve the standard ETAS model.

2.1.5 Non-temporal extensions

So far we discussed point processes solely in terms of occurrence times of events. There are many ways in which this basic structure can be extended. For example, each event may be associated with a "mark" or label that carries additional information about the event and which may or may not influence the subsequent evolution of the process. In a financial context marks can correspond to financial loss or gain, while in seismology we usually refer to the earthquakes' magnitude. The introduced models in Chapters 3 and 4 explore a marked point process.

A further extension is to consider counting processes not only on temporal intervals, but also on a spatial region or a mixture on a space-time continuum. We do not use spatial-only point processes within this thesis. Chapter 3 explores a temporal point process, while Chapters 4 and 5 use spatio-temporal constructions.

A point process can be considered to count points in multiple sequences that share common features or influence each other. This corresponds to a multivariate point process. A special case of this construction is used in Chapter 5 based on shared parameters across dimensions with no dependence among them. This simplified version of a multivariate ETAS model is beneficial for data with limited information across some of the dimensions.

2.2 Epidemic Type Aftershock Sequence (ETAS)

The Epidemic Type Aftershock Sequence (ETAS) model is commonly used for studying and forecasting the occurrence times of earthquakes in a geographical region of interest [Ogata, 1988, Marzocchi and Lombardi, 2009, Omi et al., 2014, Omi et al., 2015]. It aims to represent the mechanism of seismic activity in which the occurrence of large earthquakes in particular is associated with an increased rate of subsequent events, or aftershocks. In this construction, event occurrences are assumed to follow a self-exciting marked point process governed by a conditional intensity function $\lambda(t|\mathscr{H}_t)$ which defines the probability of an event occurring at each infinitesimal time interval around point t

based on the catalogue $\mathscr{H}_t = \{(t_1, m_1), (t_2, m_2), \ldots : t_i < t\}$, where t_i and m_i respectively denote the time and magnitude of the i^{th} previous event. The ETAS model and a general Hawkes process share the same conditional intensity construction:

$$\lambda(t|\mathscr{H}_t) = \mu(t) + \sum_{t_i<t} \nu(t - t_i, m_i), \tag{2.11}$$

where $\mu(\cdot)$ and $\nu(\cdot)$ on their own are intensity functions. $\mu(t)$ is the intensity of the uncaused (immigrant) events and $\nu(t - t_i, m_i)$ represents an increase in intensity caused by event i i.e. the intensity of the caused (children) events. The background intensity $\mu(\cdot)$ can be allowed to vary in time and space, although it is usually taken to be constant i.e. $\mu(\cdot) \equiv \mu$. As already introduced in Section 2.1.4 a constant intensity function corresponds to a Poisson process, hence the standard ETAS process assumes that the inter-arrival times of all uncaused events follow an Exponential distribution. $\mu(\cdot)$ is typically referred to as ground intensity as this is the minimum intensity that the model has. This model formulation has the effect that each event occurrence increases the subsequent event rate: the processes are sometimes called 'self-exciting' processes for this reason. The base structure of an ETAS model relies on main temporal offspring intensity proportional to the Omori law [Guglielmi, 2017]:

$$\nu(t - t_i, m_i) \propto \frac{k}{(t - t_i + c)^p},$$

where c and p are parameters controlling the temporal decay rate, while k controls the average productivity (i.e. the expected number of children of each event). A detailed treatment of the development of the Omori law is provided by [Utsu et al., 1995].

A complete conditional intensity function depends on both arrival times and marks. The latter are assumed to be realised independently for each event and commonly follow a scaled Gutenberg-Richter law [Gutenberg and Richter, 1944, Fox et al., 2016]:

$$m_i - M_0 \sim \text{Exp}(\beta),$$

where M_0 is the minimum magnitude that is taken into account. Typically, small events are excluded from studies of earthquake catalogues. This may be due to concerns that some such events may be missing from the catalogue because they were not detected using the technology available at the time; alternatively, in some applications small events are of limited interest because they are not a major source of risk. It is common, there-

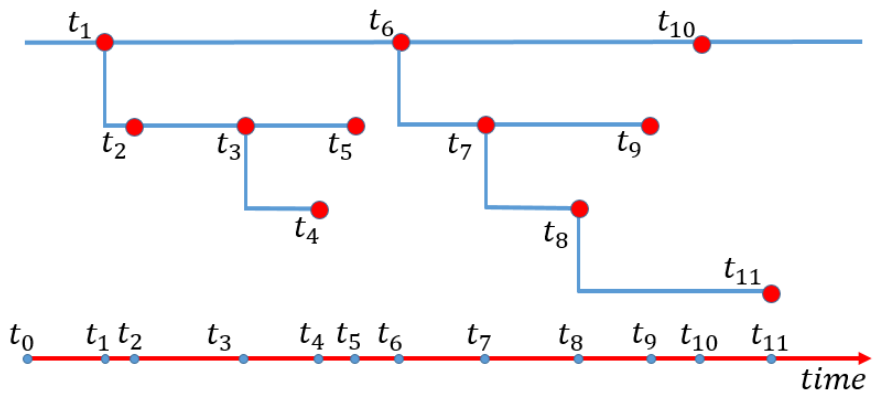

Figure 2.2: Example of a Branching structure

fore, to restrict attention to events with magnitudes above a user-specified threshold, denoted M_0 in this thesis [Gutenberg and Richter, 1944, Ogata, 1988].

2.2.1 Branching structure

In the ETAS model formulation, every event in a catalogue can generate offspring events that produce further aftershocks (also known as 'offspring' or 'children'), and so on. A visual example of a possible branching structure is shown on Figure 2.2. Here events t_1, t_6 and t_{10} are the uncased events (immigrants) which initiated the other events. They were generated from a homogeneous Poisson process with rate μ. Then, each of the events in the sequence produces offspring events according to an inhomogeneous Poisson process with rate $\nu(\cdot)$ that can further cause offsprings of their own and so on. As of Figure 2.2 events t_2, t_3 and t_5 are children of t_1, while t_4 is a child of t_3. Similarly t_7 and t_9 are off-springs of t_6, while t_{11} is caused by t_8, which is a child of t_7. There are no detected children events for t_{10} in the temporal interval that we currently observe, although offsprings from all events can occur outside of the observed period of interest. Further, we can define a 'dynasty' to consist of all events associated with an uncaused event. Then, the introduced branching structure consists of 3 dynasties - $\{t_1, t_2, t_3, t_4, t_5\}$, $\{t_6, t_7, t_8, t_9, t_{11}\}$ and $\{t_{10}\}$.

A procedure to sample a branching structure is derived with respect to all introduced models in this thesis as outlined in Sections 3.5.2, 4.6.1 and 5.5.1. Based on the branching structure we can develop inference regarding the number of caused events every event in the sequence is likely to generate. This way, the branching structure offers an opportunity to obtain a computationally tractable way of model estimation and uncertainty propagation.

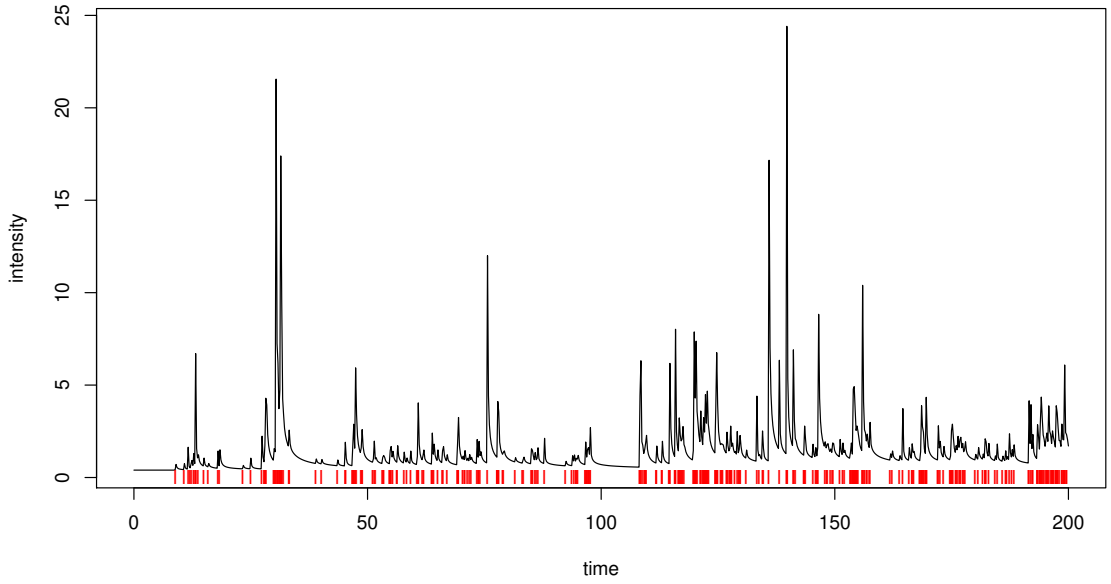

Figure 2.3: Example of an ETAS model conditional intensity function. The red lines indicate event occurrence times.

2.2.2 Intensity structure

Based on the conditional intensity function as of Equation 2.11, we can illustrate the process' self exciting nature as shown on Figure 2.3. The minimum intensity level is $\mu(t) = \mu$ which is fixed over time. Then, each event leads to an instantaneous increase in the subsequent intensity, gradually decaying over time and eventually becoming zero according to the form of the $\nu(\cdot)$ function. Each of these additive intensities will decay over time and eventually become zero. Hence, the process is self-excited when an event occurs and will gradually return to its base, ground intensity μ if no further events occur for a long enough period of time.

Obtaining a finite catalogue depends on the form of $\nu(\cdot)$. It is sufficient to obtain an offspring decay for which every new event would be associated with less than 1 additional event on average. This corresponds to the following expression:

$$\int_0^\infty \nu(z)dz < 1.$$

Under a stationarity assumption the ETAS model overall intensity is:

$$\tilde{\lambda} = \frac{\mu}{1 - \int_0^\infty \nu(z)dz}.$$

The corresponding expected number of events associated with each uncaused event N_μ is:

$$E[N_\mu] = \frac{1}{1 - \int_0^\infty \nu(z)dz}. \tag{2.12}$$

These results were established more rigorously in the literature [Cox and Isham, 1980, Hawkes and Oakes, 1974]. Hawkes and Oakes (1974) further derive the interval of the expected duration of a dynasty as:

$$\left(A\frac{\int_0^\infty \nu(z)dz}{\exp\left[-2\int_0^\infty \nu(z)dz\right]}, A\frac{\int_0^\infty \nu(z)dz}{1 - \int_0^\infty \nu(z)dz} \right), \tag{2.13}$$

where $A = \int_0^\infty z\nu(z)dz / \int_0^\infty \nu(z)dz$.

Let us consider a very simplistic unmarked ETAS example, where the intensity of the process is :

$$\lambda(t) = \mu + \sum_{t_i<t}(p-1)c^{p-1}\frac{K}{(t-t_i+c)^p} \qquad p > 1, c > 0, \tag{2.14}$$

where $\nu(t) = (p-1)c^{p-1}\frac{K}{(t-t_i+c)^p}$ for $p > 1, c > 0$ is the modified Omori law [Vere-Jones and Davies, 1966, Ross, 2018b]. This intensity corresponds to cumulative intensity of the offspring process of a single event of:

$$\int_0^\infty \nu(z)dz = K. \tag{2.15}$$

Further,

$$\int_0^\infty z\nu(z)dz = \frac{Kc}{p-2} \qquad p > 2. \tag{2.16}$$

Combining the results in Equations 2.12-2.16, we obtain the expected dynasty size for the discussed ETAS model to be:

$$E(N_\mu) = \frac{1}{1-K} \qquad K > 1,$$

with corresponding expected dynasty duration interval of:

$$\left(\frac{cK\exp(-2K)}{p-2}, \frac{Kc}{(p-2)(1-K)} \right) \qquad p > 2, K > 1, c > 0. \tag{2.17}$$

The discussed key properties highlight some restrictions on the parameter values ($K < 1$ for finite catalogue and $p > 2$ for finite expectation) that are needed to ensure

that the process behaviour is 'realistic' when considered as a model with respect to real data. However, p is typically unrestricted in the literature and commonly obtains values close to 1. Further, the finite expectation is not guaranteed in a seismic context because an event can potentially cause another one forever. The restriction of $K < 1$ is essential to obtain a 'finite' catalogue i.e. a catalogue in which the number of earthquakes above certain magnitude is countable in a closed interval. More specific details on this topic related to marked, spatio-temporal ETAS model are introduced in Section 4.7.

2.2.3 ETAS structural extensions

The basic ETAS model is applicable in situations where it is of interest to model a sequence of earthquake occurrence times and magnitudes. In other situations, however, it may be of interest to study the events' spatial locations as well: in this case it is necessary to extend the basic model to incorporate spatial locations as well as occurrence times. Furthermore, we might consider multiple sites of interests that share similar but not necessarily identical behaviour or multiple temporal point processes e.g. representing occurrences at a discrete set of locations, or events experienced by different individuals in a population. Such linked patterns can be modelled based on a multivariate point process with either directly shared parameter values or to introduce an interaction between parameter values across different dimensions. In this thesis are explored both of these extensions: spatio-temporal ETAS and multivariate ETAS.

Spatio-temporal ETAS

In its essence the spatio-temporal ETAS, also referred to as spatial ETAS, model consists of additional intensity functions that incorporate the location information in each of the model components. In a space-time setting the intensity is a function of the spatial co-ordinates x and y as well as the event arrival time t:

$$\lambda(t, x, y|\mathscr{H}_t) = \mu(t, x, y) + \sum_{t_i<t} \nu(t - t_i, x - x_i, y - y_i, m_i). \tag{2.18}$$

where the component functionalities of Equation 2.18 are the same as the ones from Equation 2.11, namely $\mu(\cdot)$ is the intensity of the uncaused (immigrant) events and $\nu(\cdot)$ represents an increase in intensity caused by event i. For this construction, the catalogue information to consider is $\mathscr{H}_t = \{(t_1, x_1, y_1, m_1), (t_2, x_1, y_1, m_2), \ldots : t_i < t\}$, where t_i, (x_i, y_i) and m_i respectively denote the time, location and magnitude of the i^{th}

earthquake. Then the cumulative intensity of the model will depend on space and time instead of just time. A detailed review of the sptio-temporal ETAS model is present in Section 4.2.

Multivariate ETAS model

The process above is a spatio-temporal marked model. An alternative way to extend the basic spatio-temporal ETAS is to introduce a multivariate setting. Chapter 5 introduces a multivariate construction of the unmarked, spatio-temporal ETAS model. Consider a collection of m separate temporal point processes that are linked in some way: each of the m processes will be referred to subsequently as a "dimension" of the combined process as of Equation 2.1. Then for each dimension j in $\{1, ..., m\}$

$$N^{(j)}(0,t) = \sum_i \mathbb{1}_{t_i^{(j)}<t}, \tag{2.19}$$

where $t^{(j)}$ corresponds to an event arrival time in the j^{th} dimension and $N^{(j)}(\cdot)$ indicates the counting process of the j^{th} dimension. Then the multivariate counting construction is an aggregation of the individual dimensions i.e. $\boldsymbol{N}(\cdot) = \{N^{(1)}(\cdot), ..., N^{(m)}(\cdot)\}$. Similarly, the history to be considered is an ordered aggregation of the individual dimensions' history $\mathscr{H}_t = \cup_{j=1}^m \mathscr{H}_t^{(j)}$, where the data in the j^{th} dimension is $\mathscr{H}_t^{(j)} = \{(t_1^{(j)}, x_1^{(j)}, y_1^{(j)}), (t_2^{(j)}, x_2^{(j)}, y_2^{(j)}), \ldots : t_i^{(j)} < t\}$, where $t_i^{(j)}$ and $(x_i^{(j)}, y_i^{(j)})$ respectively denote the time and location of the i^{th} observation across the j^{th} dimension. The intensity measure of the process $\boldsymbol{\lambda}(t,x,y)$ is also multivariate with m dimensions.

$$\boldsymbol{\lambda}(t,x,y) = (\lambda^{(1)}(t,x,y), ..., \lambda^{(m)}(t,x,y))'$$

where each of the individual dimensions' intensity functions are of the form

$$\lambda^{(i)}(t,x,y) = \lim_{\epsilon \to 0} \frac{E[N^{(i)}(t, t+\epsilon) | \mathscr{H}_t^{(i)}]}{\epsilon}.$$

2.3 Simulation and restrictions

In this Section we discuss several basic methods for sampling from a general Hawkes process. They can be applied in a wider ETAS context with minor modifications. Simulating from a specific process is essential for sanity checks of code performance, estimation methods' performance and predictions. In a Hawkes process setup, simulated data

can further exhibits information that is typically unknown such as the full branching structure of the process and the true underlying intensity function.

2.3.1 Simulation techniques

A number of techniques can be used to simulate a realisation from a general Hawkes process. The two most recognised methods for model simulation are *simulation by inversion* and *simulation by thinning* [Rasmussen, 2011,Rasmussen, 2018]. Let us define a temporal region of interest $[0, T]$ in which we would like to obtain a Hawkes process realisation.

Simulation by inversion

1. Set $t = 0$, $t_0 = 0$ and $n = 0$ (note that t_0 is a starting point rather than an actual event).

2. Repeat while $t < T$ i.e. current time is before the maximum time of the sequence:

   (a) Generate $s \sim \text{Exp}(1)$.

   (b) Calculate t, where $t = \Lambda^{-1}(s)$, where $t = \Lambda^{-1}(\cdot)$ is the inverse of the cumulative intensity function, $\lambda(\cdot)$ i.e. $\Lambda(t) = \int_0^t \lambda(s|\mathscr{H}_s)ds$

   (c) if $t < T - t_n$, set $t_{n+1} = t_n + t$ and $n = n + 1$.

3. The sequence $\{t_1, ..., t_n\}$ is generated by a Hawkes process and is in the interval from 0 to T.

However, the evaluation of $\Lambda^{-1}(t)$ is not trivial for the general Hawkes process. Thus, alternatives such as Ogata's modified thinning algorithm are preferred.

Simulation by thinning

1. Set $t = 0$ and $n = 0$.

2. Repeat until $t < T$ i.e. current observation is smaller than the maximum value we want to take into account.

   (a) Compute the value of $m(t)$ and $l(t)$ such that $m(t) \geq \sup_{s \in [t, t+l(t)]} \lambda(s|\mathscr{H}_s)$.

   (b) Generate i.i.d. variables such that $s \sim Exp[m(t)]$ and $U \sim U(0, 1)$.

(c) If $s > l(t)$, then $t = t + l(t)$, else if $t + s > T$ or $U > \lambda(t + s)/m(t)$, then $t = t + s$, else $n = n + 1$, $t_n = t + s$, $t = t + s$.

3. The sequence $\{t_1, ..., t_n\}$ is generated by a Hawkes process and is in the interval from 0 to T.

In terms of a typical Hawkes process, the $m(t)$ is simply the intensity function and $l(t) = \infty$.

Although the Ogata thinning algorithm provides a viable simulation option, it tends to be relatively slow compared to a simulation technique that explores the process inherent branching structure. Further, the thinning algorithm does not provide direct information of the processes' underlying branching structure, which is required for the inference methods developed later in the thesis. Throughout all code implementations in this thesis the Ogata thinning algorithm was used solely for comparison purposes with the method introduced below, which we used for the simulation of all stress tests that were carried out.

Simulation by clustering

An alternative method for obtaining a realisation from ETAS is to simulate events based on their underlying branching structure (as of Figure 2.2). We do that by initiating all immigrant events and then allowing each of them to excite further events which on their own can produce more events and so on. Such simulation provides intuition into the underlying model since the true branching structure is known [Dassios et al., 2013,Harte et al., 2010].

1. Sample the number of uncaused events, m, in the temporal detection region $[0, T]$ from a Poisson distribution with mean $\int_0^T \mu(t)dt$. Then assign uniformly the events' attributes within their detection range.

2. Create offspring generation G_0 that contains all uncaused events (generation 0).

3. Repeat while the i^{th} ($i \in \{0, 1, ...\}$) generation is non-empty (i.e. $|G_i| > 0$):

   (a) For all elements in $G_i = \{t_1, ..., t_{|G_i|}\}$:

      i. Sample the number of offsprings for each event in G_i from a Poisson distribution with mean $\int_0^{T-t_j} \nu(z)dz$.

ii. Sample temporal lags from $\nu(\cdot)$ on temporal interval $(0, T - t_j]$.

iii. Record this realisation in $G_i^{(j)}$.

(b) Create generation $G_{i+1} = \cup_{j=1}^{|G_i|} G_i^{(j)}$.

(c) Order chronologically G_{i+1}.

4. Create the full catalogue by merging all generations $G = \cup_i G_i$.

5. Order chronologically G and record the full branching structure that created this realisation.

More details related to the specific simulation techniques are present in Section 4.4 that directly address the large simulation study that we conduct with respect to the spatio-temporal ETAS process (Section 4.9). We also outline a simulation mechanism for the multivariate ETAS process in Section 5.3.

2.3.2 Simulation considerations

The properties of a general ETAS model that were discussed in Section 2.2.2 place restrictions on parameter values for which the overall intensity of the process is finite. Furthermore, there are parameter values for which the process does not have any appreciable clustering structure, for example when the expected dynasty (as of Equation 2.12) size is close to zero.

A set of parameter estimates, $\bar{\theta}$, can provide an optimum with respect to an objective function of interest with offspring productivity that does not guarantee a finite catalogue. Such parameter sets cannot be used for the development of a predictive study.

None of the three simulation algorithms introduced in this section allow for descendants of immigrants that occur prior to time 0. This provides leaner catalogues until the mean offspring causality is reached. In estimation context this phenomenon will cause difficulties with respect to the proportion of uncaused events in the first part of the catalogue. For studies that span a small temporal interval is beneficial to sample a larger set and discard a catalogue length larger than the expected dynasty duration.

2.4 Inference

In this subsection we will address the basic estimation techniques of a Hawkes process. Then we will discuss methods for comparison and diagnostics across a set of proposed models.

2.4.1 Likelihood

A natural way to demonstrate the form of a point process' log-likelihood is to consider dividing the observation interval into a large number of small intervals, each of length δt. Consider aggregating the data by counting the numbers of events in each of these small intervals. Then, if δt is small and conditional on the history up to $k\delta t$, the number of events in the interval $\big(k\delta t, (k+1)\delta t\big)$, for $k = \{0, 1, ...\}$, has approximately a Poisson distribution with mean $\lambda(t|\mathscr{H}_{k\delta t})\delta t$, where $\mathscr{H}_{k\delta t}$ denotes the processes history up to $k\delta t$. However, this Poisson distribution approximates the conditional distribution of the number of events given the numbers of events in all previous intervals, together with other elements of the history $\mathscr{H}_{k\delta t}$.

The joint probability mass function of the aggregated observational counts can be approximated by the product of the individual probabilities over all the intervals (by the generalised multiplication law). As δt tends to zero, the approximation gets more and more accurate; moreover, the number of events in each interval will eventually become 0 (with probability $\exp[-\lambda(t|\mathscr{H}_{k\delta t})\delta t]$ or 1 (with probability $\lambda \exp[-\lambda(t|\mathscr{H}_{k\delta t})\delta t]\delta t)$).

The logarithm of the joint distribution is thus approximated by:

$$\sum_{k:N(k\delta t,(k+1)\delta t)=0} -\lambda(t|\mathscr{H}_{k\delta t})\delta t + \sum_{k:N(k\delta t,(k+1)\delta t)>0} \log \lambda(t|\mathscr{H}_{k\delta t}) - \lambda(t|\mathscr{H}_{k\delta t})\delta_t + \log \delta t =$$

$$\sum_{k} -\lambda(t|\mathscr{H}_{k\delta t})\delta t + \sum_{k:N(k\delta t,(k+1)\delta t)>0} \log \lambda(t|\mathscr{H}_{k\delta t}) + \log \delta t.$$

The final term of the second sum ($\log \delta t$) does not depend on the model parameters, we can consider the remaining expression as defining an approximation to the log-likelihood function which becomes more accurate as δt goes to 0. In the limit, the first sum becomes an integral and the second becomes a sum over the actual event times.

The conditional Poisson marginal inter-event time density is the probability of observing an event at time t_i and not observing any other events between the last detected event t_{i-1} and t_i:

$$f(t_i|\mathscr{H}_{t_i}) = \lambda(t_i) \exp\left(-\int_{t_{i-1}}^{t_i} \lambda(u)du \right) \text{ for } i \in \{1, \ldots, n\},$$

where $t_0 = 0$. Combining that with the density of not detecting an event after the last detected event, t_n, until the end of the catalogue T we obtain:

$$f(T|\mathscr{H}_T) = \exp\left(- \int_{t_n}^{T} \lambda(u)du \right),$$

which leads to the following overall form of the likelihood:

$$\begin{aligned} \mathcal{L}(\theta; \mathscr{H}_T) &= f(T|\mathscr{H}_T) \prod_{i=1}^{n} f(t_i|\mathscr{H}_{t_i}) \\ &= \exp\left(- \int_{0}^{T} \lambda(u)du \right) \prod_{i=1}^{n} \lambda(t_i), \end{aligned} \tag{2.20}$$

with corresponding log-likelihood of:

$$\log\left(\mathcal{L}(\theta; \mathscr{H}_T)\right) = \ell(\theta; \mathscr{H}_T) = - \int_{0}^{T} \lambda(u)du + \sum_{i=1}^{n} \lambda(t_i). \tag{2.21}$$

A more formal definition of a point process' (log-)likelihood is provided by Daley and Vere-Jones [Daley and Vere-Jones, 2003] with a detailed treatment in the work of Rasmussen [Rasmussen, 2018].

The likelihood is a function of the parameters, θ, for the data of interest, in our case $\mathscr{H}_T$. This leads to the general likelihood notation of $\mathcal{L}(\theta; \mathscr{H}_T)$. According to the *likelihood principle* for a given sample of data, $\mathscr{H}_T$, the inference for the parameter set θ across any two probability models $P(\mathscr{H}_T|\theta)$ is the same if their likelihood functions are the same [Gelman et al., 2014]. The model function $P(\mathscr{H}_T|\theta)$, commonly referred to as a sampling (or data) distribution in a Bayesian context, equates to the (log-)likelihood function if considered as a function of θ for fixed data $\mathscr{H}_T$.

2.4.2 Parameter uncertainty

Parameter uncertainty can be propagated either in frequentist (classic) or Bayesian manner. The former one considers the information carried out by each parameter. Based on the law of large numbers, the parameters' distribution is approximated to Gaussian with mean the maximum likelihood estimate and standard deviation proportional to the information that it carries. The procedure for this is the following:

1. Evaluate the Hessian matrix.

Let us define a function $f : \mathbb{R}^n \to \mathbb{R}$ for which all second partial derivatives exist and are continuous over the support of the function. Then the Hessian matrix $\mathbf{H}$ of the function $f(\theta)$ is:

$$\mathbf{H}(\theta) = \begin{bmatrix} \frac{\partial^2 f(\theta)}{\partial x_1^2} & \frac{\partial^2 f(\theta)}{\partial x_1 \, \partial x_2} & \cdots & \frac{\partial^2 f(\theta)}{\partial x_1 \, \partial x_n} \\ \frac{\partial^2 f(\theta)}{\partial x_2 \, \partial x_1} & \frac{\partial^2 f(\theta)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\theta)}{\partial x_2 \, \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\theta)}{\partial x_n \, \partial x_1} & \frac{\partial^2 f(\theta)}{\partial x_n \, \partial x_2} & \cdots & \frac{\partial^2 f(\theta)}{\partial x_n^2} \end{bmatrix},$$

where $\theta = \{x_1, ..., x_n\}$.

2. Calculate the Fisher Information matrix. Let us consider the function $f(\theta)$ that was introduced in the previous step to be the log-likelihood function of a valid probability density function $g(X|\theta)$, where $\theta = \{x_1, ..., x_n\}$ is its parameter set. The Fisher information measures the amount of information that a random variable X carries about each of the parameters in θ and it has the following form

$$\mathbf{I}(\hat{\theta}) = -\mathbf{H}(\hat{\theta}),$$

where $\hat{\theta}$ indicates the maximum likelihood estimate of θ.

3. Approximate distribution of $\hat{\theta}$ is

$$\hat{\theta} \propto N\big(\hat{\theta}, \mathbf{I}(\hat{\theta})^{-1}\big),$$

with each of the parameters in $\hat{\theta}$ having estimated standard errors equal to the inverse of the square root of the corresponding diagonal element from the full information matrix.

There are situations in which the maximum likelihood provides reasonable results and others in which it does not. The classic approach is viable if, for example, we consider the point process as of Equation 2.14 with respect to model parameters for which the expected dynasty duration (Equation 2.17) is relatively short and all dynasties are non-overlapping. However, such patterns are not commonly observed. Without considering the finite dynasty expectation constraint, the performance of frequentist

maximum likelihood estimation for the general ETAS model based on directly maximising the likelihood function discovered that the resulting parameter estimates often differed substantially from their true values [Schoenberg, 2013]. The prime cause is the likelihood function's multi-modality which is primarily driven by the overlapping dynasties. Further, the components of the parameter vector can be moderately correlated. This issue primarily occurs for parameters c and p of the modified Omori law. The correlation that they experience makes parameter-wise optimisation problematic. Further, the starting values of a maximum likelihood optimisation can dictate the convergence results [Harte et al., 2010].

Although the overall uncertainty of the model parameters can be obtained in a classical manner, it cannot be fully implemented in the context of ETAS model usage for risk calculations. In order this to be achieved, we have to propagate the parameter uncertainty through the subsequent risk analysis which is very difficult. If risk could be represented as a simple function of the ETAS model parameters, we could in principle use standard results for transformation of variables to get approximate uncertainties on the risk: however, in general this is not the case. Rather, risk calculations require a combination of hazard (represented by the ETAS model) with exposure, vulnerability and loss models [Kron, 2002]. Such relationships are complex with the easiest way to get at the risk will often be to simulate multiple catalogues and to run each of these through the subsequent risk calculations [Shapira, 1983, Crowley et al., 2013]. In this case, the parameter uncertainties must be incorporated into the simulations somehow which is not trivial in a classical framework, essentially because we have to sample from the predictive distribution of the point process and such predictive distributions are hard to calculate in all but the simplest statistical models [Cox and Hinkley, 1974]. The Bayesian approach via Markov chain Monte-Carlo (MCMC) offers an alternative solution in which we can generate samples from the posterior predictive distribution of the process easily given a sample from the posterior of the parameters.

2.4.3 Bayesian paradigm

Bayesian statistics represents an alternative statistical framework for reasoning about uncertainty, which is becoming increasingly popular both in seismology and in ETAS contexts [Faenza et al., 2010, Holschneider et al., 2012, Shcherbakov, 2014, Ross, 2018a, Kolev and Ross, 2019, Rotondi and Varini, 2019]. In the Bayesian paradigm, we do not

work with only a single estimate of θ (with its corresponding symmetric uncertainties) but instead consider the whole posterior distribution $P(\theta|\mathscr{H}_t)$ which represents our uncertainty about θ based on both the observed data and any prior knowledge we have based on previously conducted studies. This uncertainty can then be incorporated into forecasts in a straightforward manner [Glickman and Van Dyk, 2007]. Despite its advantages, the Bayesian framework is difficult to apply since the posterior distribution in the ETAS model is highly complex. As such, many studies which attempt Bayesian earthquake forecasting have had to resort to using frequentist-style point estimates for θ, which mitigates the benefits of the Bayesian framework [Ebrahimian et al., 2013, Ogata, 2011]. An attempt at providing a fully Bayesian treatment of the ETAS model is the unpublished thesis [Vargas and Gneiting, 2012] which proposed using a computational simulation based on the framework from [Rasmussen, 2013] for parameter estimation. However, their approach is not scalable to catalogues containing more than a few hundred earthquakes, which limits its applicability.

The main aim of Bayesian analysis is to fully explore the parameters' distribution. A prior distribution $\pi(\theta)$ is set based on our prior knowledge of the parameters' distribution. Then, using the Bayes theorem, the posterior distribution of the parameter θ can be represented as follows:

$$P(\theta|\mathscr{H}_T) = \frac{P(\mathscr{H}_T|\theta)\pi(\theta)}{\int_\Theta P(\mathscr{H}_T|\theta)\pi(\theta)d\theta}, \tag{2.22}$$

where $P(\mathscr{H}_T|\theta)$ is the sampling (or data) distribution with respect to the observed data $\mathscr{H}_T$ up to time T given a set of model parameters θ. As outlined in Section 2.4.1, the sampling distribution equates to the likelihood function when regarded as a function of θ, for fixed data $\mathscr{H}_T$. The association of a model parameter set θ, given an observed data $\mathscr{H}_T$ is represented by the posterior distribution $P(\theta|\mathscr{H}_T)$.

The multi-dimensional integral in Equation 2.22 is extremely hard to handle in its general form with respect to even a very simple example of a self-exciting point process [Rasmussen, 2013, Veen and Schoenberg, 2008, Ross, 2018a]. For this reason, we can use the well-known Metropolis-Hastings (MH) algorithm for sampling the Markov Chain of interest [Chib and Greenberg, 1995,Hamra et al., 2013,Rotondi and Varini, 2007,Rotondi and Varini, 2019]. According to this method, we firstly initialise the parameter set with some reasonable values $\theta^{(0)}$. At step i we would like to propose a new sampled value of $\theta^{(i)}$ based on $\theta^{(i-1)}$. For example we might consider a random walk transformation such

as $\theta^{(i)} = \theta^{(i-1)} + \epsilon$ where $\epsilon \sim N(0, \sigma^2)$. The acceptance probability of the proposed value $\theta^{(i)}$ is $P(\theta^{(i)}|\mathscr{H}_T)/P(\theta^{(i-1)}|\mathscr{H}_T)$. If the value is rejected, we fail to obtain a new sample at this step and assign the $(i-1)^{st}$ sample to the i^{th} (i.e. $\theta^{(i)} = \theta^{(i-1)}$) and repeat the procedure for the next step until we obtain the required number of MCMC steps. After sufficient number of parameter updates we will obtain samples from $P(\theta|\mathscr{H}_T)$ that represent an equilibrium (stationary) distribution which addresses the true shape of the parameters' distribution.

The obtained parameter chains can be further subject to *burn-in* in which we erase the initial samples for each of them. This is done to consider only parameter samples that are obtained after convergence to the respective parameters' true posterior distributions. Further, every sample depends on the previous one which naturally induces autocorrelation which can be reduced by *thinning* the obtained chains. This is done by removing all obtained parameter samples except every n^{th} one. Both of these methods are commonly used for obtaining good samples from the respective posterior distributions.

A maximum *a posterior* probability (MAP) estimate is a point estimate of model parameters obtained as the mode of its posterior distribution after thinning and burn-in. MAP technique is commonly used for obtaining point estimates similar to those from a classical maximum likelihood procedure.

It may initially seem feasible to use direct Metropolis-Hastings to sample from the posterior distribution (Equation 2.22). A general MCMC algorithm, as the one introduced in this Section, will involve iterative parameter proposal with acceptance dependent on the change of the full likelihood (Equation 2.21). Proposing updates for all parameters simultaneously requires a single calculation of the likelihood. However, this is likely to cause a lack of parameter acceptance because each of the parameters can cause major shifts in the likelihood value. An alternative approach will be to propose and accept parameters independently, which is computationally slow because the full likelihood function has to be evaluated for every proposed parameter.

Further, the direct Bayesian approach suffers from serious convergence problems that arise even in the most simplistic parametric case of ETAS model. A general exception can be obtained for non-overlapping dynasty (Section 2.2.2). Such restrictions are typically too restrictive and do not address fully the underlying ETAS behaviour. The efficiency of MCMC algorithms for ETAS type models can be drastically improved by including the branching structure as a latent variable. For all these reasons, we instead

propose a reparametrisation of the model based on the process' underlying branching structure (see Section 2.2.1) that aims to break the parameter correlation and lead to a usable Metropolis-Hastings algorithm for posterior sampling. A direct comparison between the standard MCMC and its latent variable alternative indicates faster computation time and greater effective sample size [Ross, 2018a]. Conditional on the branching structure, the parameter updates are independent of each other. Then, the updates for each of them are dependent on separate likelihood functions that are computationally inexpensive. This greatly improves the convergence and computational time. More details on the ETAS model latent variable MCMC approach are present in Sections 3.5, 4.6 and 5.5.

2.4.4 Model comparison methods

In this section are discussed multiple methods for model comparison. Some of them are applicable to any model, while others are only feasible in a parametric context.

Log-likelihood

Models of interest might be compared directly based on the maximum (log-)likelihood value that they can achieve with respect to data of interest. Although a comparison between the ratio of two likelihoods can be interpreted directly with the likelihood ratio test [Gourieroux et al., 1982], this method neglects completely the concept of overfitting. A model with more model parameters is usually better simply because it can adjust better to the data. Thus, alternatives for model comparison based on penalisation of the number of included parameters are introduced.

Akaike Information Criterion (AIC)

Probably the most widely used method for model comparison that penalises for the number of used parameters is the Akaike Information Criterion (AIC) [Akaike, 1973]. It was introduced as a method for allocation of the best fit across an ensemble of proposed models by measuring the distance between the unknown true likelihood function and the fitted one. It asymptotically corrects the bias that occurs when using the in-sample likelihood to estimate the out-of-sample likelihood. For every set of model parameters θ the model's AIC value is the following:

$$\mathrm{AIC}(\theta) = -2\ell(\theta; \mathscr{H}_T) + 2d,$$

where d is the number of free model parameters i.e. $d = |\theta|$ and $\ell(\theta; \mathscr{H}_T)$ is the log-likelihood value. The best model across a set of proposed models is associated with the lowest value of AIC coefficient. However, AIC does not explicitly address the posterior distribution association with the data. This can be done based on Bayes factor, or its approximation the Bayesian Information Criteria (BIC). Further, a hierarchical model selection generalisation of the AIC can be addressed based on the Deviance Information Criteria (DIC).

Bayes factor

Another method for model comparison based on Bayesian parameter estimation is the Bayes factor approach. It examines which model is more feasible for the specific data based on the underlying posterior probability. For every two models $M = \{M_1, M_2\}$ we have to firstly find their posterior probabilities given the data $P(M|\mathscr{H}_T)$ which is the following:

$$P(M|\mathscr{H}_T) = \frac{P(\mathscr{H}_T|M)P(M)}{P(\mathscr{H}_T)} \propto P(\mathscr{H}_T|M)P(M),$$

where the term $P(\mathscr{H}_t|M)$ is model distribution under model M. Then, the Bayes factor for M_1 versus M_2 is the following:

$$BF = \frac{P(\mathscr{H}_T|M_1)}{P(\mathscr{H}_T|M_2)}.$$

The above expression is the likelihood-ratio test statistics, which represents the difference between the log-likelihoods of two models. The result can be interpreted in terms of the log value of BF, where if $0 < log(BF) < 1$ indicates a minor difference between the two models, $1 < log(BF) < 3$ - an evident positive difference, $3 < log(BF) < 5$ - a strong difference and $5 < log(BF)$ - a very strong difference [Kass and Raftery, 1995]. From a Bayesian perspective, the marginal likelihood is expressed in terms of Bayesian evidence for a model M:

$$P(\mathscr{H}_t|M) = \int_{\Theta_M} P(\mathscr{H}_t|\theta_M, M)P(\theta_M|M)d\theta_M, \tag{2.23}$$

where with θ_M is notated the parameter set of model M, with corresponding support Θ_M. The above expression provides a better model comparison as it fully explores the parameters' support compared with the standard likelihood ratio test, in which we compare the ratio between the likelihood of two models rather than the one between their posteriors. The integral in Equation 2.23 provides a challenging numerical problem.

Obtaining a large sample of the model parameters' set θ_M indicates that its posterior distribution is approximately Gaussian around the MAP estimates ($\hat{\theta}_M$) of a model M. Then:

$$P(\theta_M|\mathscr{H}_T, M) \approx (2\pi)^{\frac{-d}{2}}|A|^{\frac{1}{2}} \exp\left[-\frac{1}{2}(\theta_M - \hat{\theta}_M)'A(\theta_M - \hat{\theta}_M)\right],$$

where d is the number of model's parameters i.e. $d = |\theta_M|$ and A is a $d \times d$ Hessian matrix of $P(\theta_M|\mathscr{H}_T, M)$ evaluated at the MAP estimates i.e. $A_{ij} = -\frac{\partial^2}{\partial\theta_M^{(i)}\partial\theta_M^{(j)}}P(\theta_M|\mathscr{H}_T, M)|_{\hat{\theta}_M}$. Combining this with the fact that:

$$P(\mathscr{H}_T|M) = \frac{P(\theta_M, \mathscr{H}_T|M)}{P(\theta_M|\mathscr{H}_T, M)},$$

we can evaluate the $\log(P(\theta_M|\mathscr{H}_T, M))$ at $\hat{\theta}_M$ as follows:

$$\log(P(\mathscr{H}_T|M)) \approx \log(P(\hat{\theta}_M|M)) + \log(P(\mathscr{H}_T|\hat{\theta}_M, M)) + \frac{d}{2}log(2\pi) - \frac{1}{2}\log(|A|). \quad (2.24)$$

An approximation of the result in Equation 2.24 can be obtained for large sample ($n \to \infty$) since for a fixed matrix A_0 the Hessian matrix A grows as nA_0 [Ghahramani, 2005]. Thus, the term $\log(|A|)$ can be expressed as follows:

$$\log(|A|) \to \log(|nA_0|) = \log(n^d|A_0|) = d\log(n) + \log(|A_0|).$$

Applying this results in Equation 2.24 and further retain only parameters that grow in n, we obtain:

$$\log(P(\mathscr{H}_T|M)) \approx \log(P(\mathscr{H}_T|\hat{\theta}_M, M)) - \frac{d}{2}\log(n),$$

which returns the Bayesian Information Criterion (BIC). Hence, for a large sample size the Bayes factor converge to the Bayesian Information Criterion (BIC).

Bayesian Information Criterion (BIC)

The Bayesian Information Criterion (BIC) is a popular penalised likelihood technique which incorporates a penalty based on the number of parameters to reduce the risk of overfitting [Schwarz et al., 1978]. Given a model with parameter vector θ, the models BIC is defined as:

$$\text{BIC}(\theta) = -\ell(\theta; \mathscr{H}_T) + \frac{d}{2}\log(n),$$

where d is the number of free model parameters i.e. $d = |\theta|$, $\ell(\theta; \mathscr{H}_T)$ is the log-likelihood value evaluated at the Maximum Likelihood estimates (MLE) $\hat{\theta}$ and n is the number of observations. BIC provides stronger penalty than AIC for $n > 7$ with respect to models with more parameters. The best model with respect to this criterion is associated with the lowest value of BIC coefficient.

Deviance Information Criterion (DIC)

The DIC is a fully Bayesian alternative to the AIC. It replaces the maximum likelihood parameter estimates of θ with their posterior mean $\bar{\theta}$. The correction associated with the number of model parameters is replaced with a measure of parameter adequacy based on the goodness of sample of θ in terms of log-likelihood [Gelman et al., 2014]. DIC informally addressed the extend to which parameters are approximated well based on the obtained MCMC samples. For every set of model parameters θ the model's DIC value is:

$$\mathrm{DIC}(\theta) = -2\ell(\bar{\theta}; \mathscr{H}_T) + 2p_{DIC},$$

where $\ell(\theta; \mathscr{H}_T)$ is the log-likelihood function and p_{DIC} is the effective number of parameters, which evaluates the number of independent samples the MCMC draws are equivalent to. It is defined as:

$$p_{DIC} = 2\ell(\bar{\theta}; \mathscr{H}_T) - 2\,\mathbb{E}[\ell(\theta; \mathscr{H}_T)] \cong 2\ell(\bar{\theta}; \mathscr{H}_T) - 2\frac{1}{S}\sum_{s=1}^{S} \ell(\theta_s; \mathscr{H}_T),$$

where θ_s indicates the s^{th} parameters' sample in the considered MCMC chain. Alternatively, we can compute the effective sample size as the variance of the obtained log-likelihood values for all sampled parameters as follows:

$$p_{DICalt} = 2Var[\ell(\theta; \mathscr{H}_T)].$$

This method is not as numerically stable as the other one but it is easier to compute as it does not require the allocation of $\bar{\theta}$ as well as the calculation of the likelihood function for this set of parameters. It further guarantees to provide positive values. There are many different alternatives of DIC that address specific data and model prerequisites [Spiegelhalter et al., 2014].

2.4.5 Diagnostic measures for model checking

In the previous section we discussed methods that allocate the best model across a range of proposed models. However, in order a model to be considered as adequate we want to formally examine its appropriateness. We do that with respect to the introduced in this Section measures that we use for model checking.

Time re-scaling residuals

The time re-scaling concept aims to re-scale the observations from a point process based on its conditional intensity function, to produce residuals which follow a homogeneous Poisson process [Brown et al., 2002, Lallouache and Challet, 2016]. This way we can evaluate whether a model produces undesired patterns that indicate major flows with its performance. For a given temporal point process sequence $0 = t_0 \leq t_1 <, ..., < t_n \leq T$ with corresponding conditional intensity $\lambda(t|\mathscr{H}_t) > 0$ for $t \in (0, T]$, the residuals are defined for each $k \in \{1, ..., n\}$ as:

$$\Lambda(t_k) = \int_0^{t_k} \lambda(u|\theta, \mathscr{H}_u)du.$$

Assuming that the cumulative intensity is finite, i.e. $\lambda(\cdot) < \infty$, then all $\Lambda(\cdot)$s are a realisation from a Poisson process with rate 1. The inter arrival times $\delta_k = \Lambda(t_k) - \Lambda(t_{k-1})$ (assuming that $t_0 = 0$) are hence independent Exponentially distributed with mean and standard deviation of 1. As such, testing these residuals to check they follow an Exponential distribution is equivalent to testing whether the conditional intensity function describes the data well.

Proof. A sufficient proof to the above statement that time re-scaling residuals follow a homogeneous Poisson process is to show that δ_i, $i \in \{1, ..., n\}$ are independently and identically distributed, following an Exponential distribution with a unit rate. We set an additional time-residual that captures the time elapsed between the last event t_n and the stopping time T as $\delta_T = \int_{t_n}^{T} \lambda(u|\theta, \mathscr{H}_u)du$. The joint probability of all δ_ks is the probability of obtaining the specific time residuals, combined with having a time residual for the $n + 1^{st}$ event greater than $T - t_n$

$$f(\delta_1, \delta_2, ..., \delta_n, \delta_{n+1} > T - t_n) = f(\delta_1, ..., \delta_n)P(\delta_{n+1} > \delta_T). \tag{2.25}$$

However, the probability of having $\delta_{n+1} > \delta_T$ is equivalent to $t_{n+1} > T$ or in no events to be detected in the interval $(t_n, T]$ as previously outlined in Section 2.4.1. Then

$$P(\delta_{n+1} > \delta_T) = P(t_{n+1} > T) = \exp\left(- \int_{t_n}^{T} \lambda(u|\theta, \mathscr{H}_u) du \right) = \exp(-\delta_T). \tag{2.26}$$

We further perform a multivariate change of variables from $\delta_.$ to $t_.$

$$f(\delta_1, ..., \delta_n) = |J| f(t_1, ..., t_n, N(0, t_n) = n),$$

where $f(t_1, ..., t_n, N(0, t_n) = n)$ is the joint density of interest, J is a Jacobian matrix between t_i and δ_i for $i \in \{1, ..., n\}$. δ_k is a one-to-one function of t_k, hence J is a lower triangular matrix with determinant $|J| = |\prod_{i=1}^{n} J_{ii}|$ where $J_{ii} = \frac{\partial t_i}{\partial \delta_i} = \lambda(t_i|\theta, \mathscr{H}_{t_i})^{-1}$. Then

$$f(\delta_1, ..., \delta_n) = \prod_{i=1}^{n} \lambda(t_i|\theta, \mathscr{H}_{t_i})^{-1} \prod_{i=1}^{n} \lambda(t_i|\theta, \mathscr{H}_{t_i}) \exp\left(- \int_{t_{i-1}}^{t_i} \lambda(u|\theta, \mathscr{H}_u) du \right) = \prod_{i=1}^{n} \exp(-\delta_i).$$

Substituting this result with Equation 2.26 into Equation 2.25 we obtain:

$$\begin{aligned} f(\delta_1, ..., \delta_n, \delta_{n+1} > T - t_n) =& P(t_{n+1} > T) f(\delta_1, ..., \delta_n) \\ =& \exp(-\delta_T) \prod_{i=1}^{n} \exp(-\delta_i). \end{aligned} \tag{2.27}$$

The provided joint density of the time-residuals δ_i, $i \in \{1, ..., n\}$ is recognised as the density function of an Exponential distribution with unit rate which establishes the result. □

The time re-scaling test can be carried out using a goodness-of-fit that examines how close the obtained sequence of time residuals follows the unit Poisson process. This can be achieved using any of the following tests: Kolmogorov-Smornov test, Cramér-Von Mises (CVM), Anderson-Darling, Ljung-Box or Engle Russell Excess Dispersion.

An alternative informal method for examining the goodness of fit is to define residuals related to the expected number of events at specific time versus the actual number of events that occur [Andersen et al., 2012]. We aim to assign residuals value to every model, based on the average number of observations up to point t based on their intensity

functions as follows:

$$I(t|\theta) = N(0,t) - \int_0^t \lambda(s|\theta, \mathscr{H}_s)ds.$$

By definition, the intensity $\lambda(t) = \lim_{\delta t \to 0} \frac{1}{\delta t} P([N(0,t+\delta t) - N(0,t)] = 1)$ where $N(0,t)$ is random variable that captures the population at time t as introduced in Equation 2.1. Thus, $E[I(t|\theta)] = 0$. The obtained value of $I(t)$, which is also known as a raw residuals process, reaches zero for the best model and it can also be informally used for direct comparison between two models at specific times of interest, where the superior model is given by the smaller absolute value of the raw residual process ($|I(t)|$) [Baddeley et al., 2005].

Kolmogorov-Smirnov test

Kolmogorov-Smirnov (KS) test [Chakravarti and Laha, 1967] was introduced to check whether a set of observations follows a specific distribution of interest. For an ordered data $X = (x_1, ..., x_n)$ we are interested in examining whether the sample cumulative distribution function (CDF) $F(\cdot)$ is close to a specific CDF $F_0(\cdot)$. The corresponding test statistic is the following:

$$KS = \max_{1 \leq i \leq n} \left(F_0(x_i) - \frac{i-1}{n}, \frac{i}{n} - F_0(x_i) \right).$$

It is applicable to continuous distributions only, such as the exponential distribution which is of interest in our case. The KS test is more sensitive in the centre of the distribution compared to its tails.

Anderson-Darling test

The Anderson-Darling (AD) [Anderson and Darling, 1954] is an extension of the CVM with a varying weighting function $w(\cdot)$. The chosen functional form is

$$w(x) = \left[F_0(x)(1 - F_0(x)) \right]^{-1},$$

where $F_0(\cdot)$ is a specific CDF that we would like to compare with the sample one $F(\cdot)$ based on the expression in Equation 2.28. This way the AD test is penalising more with respect to the tails' fit.

Cramér-Von Mises test

The Cramér-Von Mises (CVM) test [Stephens, 1970] compares a set of observations to a hypothesised distribution function by computing the average distance between the empirical and hypothesised distributions. For ordered data $X = (x_1, ..., x_n)$ we are interested in examining whether the sample cumulative distribution function (CDF) $F(\cdot)$ is close to a specific CDF $F_0(\cdot)$. The CVM test statistics is:

$$CVM = n \int_{-\infty}^{\infty} \Big(F(x) - F_0(x)\Big)^2 w(x) dF_0(x), \tag{2.28}$$

where $w(x)$ is a weight function which is assumed to be equal to 1 in the standard CVM test. It can be considered that CVM provides a test which is in between the KS and the AD tests [Laio, 2004].

Ljung-Box test

The Ljung-Box (LB) test was introduced by [Box and Pierce, 1970], with a more detailed treatment in [Ljung and Box, 1978], to test whether time series residuals satisfy the white noise assumption. In practise it is commonly used to examining the adequacy of a null hypothesis of independence in a given time series [Mahdi and McLeod, 2019]. This is done by examining m autocorrelations of the residuals. The test statistic is:

$$Q = n(n+2) \sum_{k=1}^{m} \frac{\hat{\rho_k}^2}{n-k},$$

where n is the length of the data, $\hat{\rho}_k$ is the estimated autocorrelation at the lag of interest k with respect to the number of lags (m) that are taken into account. The choice of appropriate number of lags m is critical for obtaining adequate test results [Hyndman, 2014]. Given that the data are expected to be non-seasonal and always have more than 200 observations, we always use 10 lags i.e. $m = 10$.

Engle Russell Excess Dispersion test

Excessive dispersion of the exponentially distributed residuals can be examined using the Engle Russell Excess Dispersion (ER) test [Lallouache and Challet, 2016, Engle and Russell, 1998]. It takes into account only the sample variance $\hat{\sigma}^2$ and has the following

test statistics:

$$ER = \sqrt{n}\frac{\hat{\sigma}^2 - 1}{\sqrt{8}}.$$

Under the null hypothesis of lack of excess dispersion, $\hat{\sigma}$ is distributed approximately in large samples as a standard Normal random variable.

Chapter 3

Inference for ETAS Models With Non-Poissonian Mainshock Arrival Times

The Hawkes process is a widely used statistical model for point processes which produce clustered event times. A specific version known as the ETAS model is used in seismology to forecast earthquake arrival times under the assumption that main shocks are triggered by a ground level Poisson process, with aftershocks triggered via a parametric kernel function. However, this Poissonian assumption contradicts several aspects of seismological theory which suggest that the arrival time of mainshocks instead follows alternative renewal distributions such as the Gamma or Brownian Passage Time (BPT). We hence show how the standard ETAS/Hawkes process can be extended to allow for non-Poissonian distributions by introducing a dependence based on the underlying process' behaviour. Direct maximum likelihood estimation of the resulting models is not computationally feasible in the general case, so we also present a novel Bayesian MCMC algorithm for efficient estimation using a latent variable representation.

3.1 Background

The Epidemic Type Aftershock Sequence (ETAS) model is commonly used for studying and forecasting the occurrence of earthquakes in a geographical region of interest [Ogata, 1988]. Since the ETAS model assumes that the uncaused earthquakes follow a Poisson process with constant intensity μ_0 (Equations 2.11 and 3.1), this implies that they occur

completely at random, i.e. that an uncaused event is equally likely to occur at each point in time, and that the time between each pair of uncaused events (known as the 'inter-arrival times') follows a time-independent Exp(μ_0) distribution. However, this conflicts with findings elsewhere in the seismology literature, where there is substantial doubt over whether the occurrence times of mainshock earthquakes is really Poissonian [Tahernia et al., 2014, Ordaz and Arroyo, 2016, Marzocchi and Taroni, 2014]. Although ETAS uncaused events are not strictly equivalent to mainshocks [Rotondi and Varini, 2006, Rotondi and Varini, 2019] as defined elsewhere in the seismology literature (since there is no requirement that an ETAS main (uncaused) event should have larger magnitude than its offspring [Ogata, 1988, Ogata, 1998]), this still seems to cast some doubt on the Poissonian assumption.

The concept of Stress Release (SR) suggests that the mainshock arrival times instead follow a renewal process that has a time-dependent hazard function, with inter-event times following a distribution such as the Weibull, Gamma, or Brownian Passage Times (BPT). Stress release models (SRMs) were a representation of Reid's elastic rebound theory [Reid, 1910] and were fully described by [Isham and Westcott, 1979] as a self correcting point process which is updated after every event occurrence. They were introduced to seismology by [Vere-Jones, 1978] who developed them to address Reid's theory that earthquakes occur due to a release of energy which was previously accumulated strain energy along faults. SRMs were used in many locations to implement the elastic rebound theory due to their solid physical background. As outlined in [Varini and Rotondi, 2015] some of the examples of such implementations are present for the following countries: China [Yang et al., 2000, Liu et al., 1998, Xiaogu and Vere-Jones, 1994], Greece [Rotondi and Varini, 2006], Iran [Xiaogu and Vere-Jones, 1994], Italy [Rotondi and Varini, 2007, Varini and Rotondi, 2015, Rotondi and Varini, 2019], Japan [Imoto, 2001, Lu et al., 1999, Xiaogu and Vere-Jones, 1994], New Zealand [Yang et al., 2000] and Taiwan [Zhu and Shi, 2002].

SRMs are primarily applied to a sequence of earthquakes with large magnitudes, rather than to the full seismic sequences that are commonly used to fit ETAS models. In this project we will develop a new class of ETAS models which we call SR-ETAS (Stress Release ETAS) that improve on standard ETAS models by incorporating time-dependent, SRM based, inter-arrival distributions. We explore two different formulations of SR-ETAS, which differ based on how they handle the interevent time that is taken into account when calculating the main event intensity. The first formulation is simpler to

estimate. It addresses the Reid's elasticity rebound theory directly for all events in the catalogue. The second one is harder to estimate due to its dependence on the branching structure. It assumes that Reid's theory is applicable only for the main events, making direct maximum likelihood estimation impossible.

A model which is closely related to our SR-ETAS was proposed by [Wheatley et al., 2016], who considered a Hawkes process with a renewal immigration process, which they call Renewal Hawkes (RHawkes). The authors proposed an Expectation Maximisation (EM) algorithm for parameter estimation. However, as pointed out by Wheatley [Wheatley, 2017, Wheatley, 2016] their approach crucially exploited the Markovian properties used by the Exponential offspring density $g(\cdot)$ that they considered, which leads to instability when this is replaced by a heavy-tailed alternative such as the Omori law used in the ETAS model [Oakes, 1975, Filimonov and Sornette, 2015]. To mitigate this, they suggest that such heavy-tailed densities should be approximated by a sum of weighted exponential kernels [Hardiman et al., 2013]. Further, simulation studies found that their EM algorithm performs poorly even for the more simplistic Renewal Immigration Hawkes process in the case where the dynasties are heavily overlapping (Section 2.2.2), which is inevitable in the case of seismic sequences. To correct this, [Chen and Stindl, 2018] provided a direct maximum likelihood optimisation, as well as some conceptional corrections to the method proposed by [Wheatley et al., 2016]. However, both methods fail to address two fundamental issues. The first one is the potential multimodality of the ETAS model likelihood. As discussed in [Rasmussen, 2013, Veen and Schoenberg, 2008, Ross, 2018a], such numerical instabilities can be tackled using an MCMC sampler. The second, and probably more important problem, is the lack of discussion regarding the numerical stability of the uncaused events intensity (Equations 2.4 and 3.3) as a function of the proposed SR density. This ratio is used if the intensity cannot be factorised into a single equation i.e. it has to be evaluated as a ratio between the probability density and complementary cumulative distribution functions (CCDF). The problem occurs since the denominator of Equation 3.3 (the SR distribution CCDF) is approaching zero for large time lag.

Since the existing Expectation Maximisation (EM) and Direct Maximum Likelihood Estimation algorithms lead to either poor or limited estimation of the SR-ETAS model, we instead propose a novel Bayesian inference algorithm which uses latent variables to allow for computationally efficient inference using a Gibbs sampler, which is an extension of that proposed for the standard Hawkes process by [Ross, 2018a].

The remainder of this Chapter proceeds as follows. In Section 3.2 we review the standard ETAS model in more detail. The SR-ETAS models are fully introduced in Section 3.3, and we discuss different choices for the uncaused events' process in Section 3.4. The parameter estimation technique are present in Section 3.5. In Section 3.6 we apply SR-ETAS models and compare their performance to standard ETAS using real earthquake data from the New Madrid and the North California seismic sequences. We evaluate the models' performance based on the previously introduced in Sections 2.4.4 and 2.4.5 Goodness-of-Fit tests. We conclude with a short summary of our findings in Section 3.7.

3.2 Standard ETAS model

The standard ETAS model was introduced by Ogata [Ogata, 1988], and assumes earthquakes follow a marked point process with conditional intensity function:

$$\lambda(t|\mathscr{H}_t) = \mu_0 + \sum_{t_i<t} g(t-t_i)\kappa(m_i), \tag{3.1}$$

where t_i and m_i denote the occurrence time and magnitude of earthquake i. All magnitudes are assumed to independently follow the Gutenberg-Richter law, which corresponds to a shifted Exp(β) distribution with lower bound M_0. The μ_0 parameter specifies the intensity of the homogenous point process governing the uncaused events, while $g(\cdot)$ is a kernel function specifying how the effect of each earthquake on the intensity decays over time. It is usually taken to be the Omori law [Guglielmi, 2017]:

$$g(z) = \frac{k}{(z+c)^p},$$

where c and p are parameters controlling the decay rate, while k controls the average productivity.

The magnitude kernel $\kappa(m_i)$ determines how the magnitude of each earthquake affects the intensity and is usually defined as:

$$\kappa(m_i) = e^{\alpha(m_i - M_0)},$$

where α provides similar functionality to those of k, and M_0 is the catalogue's magnitude of completeness i.e. the minimum magnitude above which is considered that no events

are missing due to physical limitations in the earthquake detection system. The unknown parameter set of the standard ETAS model is hence: $\theta = \{\mu_0, \alpha, c, p, k\}$. All parameters have positive support, while p should be greater than 1.

Note that the form of the conditional intensity function in Equation 3.1 is equivalent to a branching process, as discussed in Section 2.2. Suppose that at some time point t there have been n_t previous earthquakes. Then, the process intensity at t can be viewed as a linear superposition of the uncaused process with intensity μ_0 and the n_t processes associated with each previous event, each contributing an intensity of $g(t-t_i)$. It can hence be seen this formulation is equivalent to assuming that the uncaused events follow a homogenous Poisson process with intensity μ_0, and hence have Exponentially distributed inter-event times.

The standard ETAS model can be generalised to include a space component which is discussed in Sections 2.2.3 and 4.2. For simplicity and ease of both simulation and computation, in this Chapter we only consider the original temporal ETAS model rather than its spatio-temporal extension, although our model could be extended to the spatial version without difficulty assuming the provided spatial kernels are independent from those introduced in this Chapter.

3.3 SR-ETAS models

A largely discussed concept in the seismology literature is the "crustal strain budget" that could be addressed by a Stress Release (SR) model that provides a possible description of the seismic elasticity as introduced by Reid in his elasticity rebound theory [Reid, 1910]. In it, earthquake inter-arrival times are described as a ratio of tectonic strain accumulation and strain release, without any statistical association of factors such as duration, time, space, and size of the seismicity. We use a Stress Release distribution for modelling immigration mainshock events rather than the typically used Exponential distribution implied by the homogenous Poisson process assumption of the standard ETAS model. SR-ETAS models provide an alternative for modelling the structure of the uncaused events arrival process. Specifically, we will assume that the intensity is time-varying and hence specify a time-dependent $\mu(t)$, leading to the following specification of the conditional intensity:

$$\lambda(t|\mathscr{H}_t) = \mu(t) + \Phi(t|\mathscr{H}_t)) = \mu(t) + \sum_{t_i<t} g(t-t_i)\kappa(m_i). \tag{3.2}$$

Although time-varying specifications of $\mu(t)$ have been considered before in the literature [Imoto, 2001, Johnson et al., 2005, Varini and Rotondi, 2015], they typically try to capture structural changes in the long-term earthquake rate, for example modelling $\mu(t)$ as a step function. Instead, following stress-release concepts, we assume that the probability of a mainshock earthquake occurring at time t depends on the time at which the last mainshock was detected. To make this clearer, we introduce the following notation. For each earthquake i, let B_i denote the index of its parent earthquake in the branching structure, with $B_i = 0$ if it has no parent (i.e. if earthquake i is uncaused). We hence have the branching vector $B = (B_1, \ldots, B_n)$ [Ross, 2018a]. For example, in Figure 2.2, $B = (0, 1, 1, 3, 1, 0, 6, 7, 6, 0, 8)$.

Using this notation, at each time t we write the occurrence time of the last previous uncaused event prior to t_i is $t_{I_{[i]}}$ where $I_{[i]} = \max_j\{j|t_j < t_i \text{ and } B_j = 0\}$. Similarly, the amount of time which has elapsed since the previous uncaused event – known as the waiting time – is given by:

$$w_{t_i} = t_i - t_{I_{[i]}}.$$

Based on the usual point process theory, as previously introduced in Section 2.1.2, $\mu(t)$ can then be defined as the hazard function:

$$\mu(t) = \mu(t|w_t) = \frac{f_w(w_t)}{1 - F_w(w_t)}, \tag{3.3}$$

where $F_w(w_t)$ is the waiting time distribution and $f_w(w_t)$ is its corresponding density. The above expression can be simplified for some distribution choices although for more complex ones such as the BPT it has to be numerically evaluated since no explicit form is present. As the CDF goes closer to 1, the expression becomes unstable due to numerical underflow caused by the numerator being effectively 0, which cannot be avoided by transforming into the space of logarithms. Since the proposed EM and MLE algorithms depend on estimating this quantity for every time lag, they will not work for general seismological-based Stress Release distributions [Wheatley et al., 2016, Chen and Stindl, 2018]. However we will show that our Bayesian updates do not need a full exploration of all possible inheritance structures, thus the waiting times that are taken into account are much smaller. As such, there is no numerical instability for any reasonable parametrisation of the uncaused events' distribution and we can evaluate numerically the above function as part of our MCMC sampler.

Under the definition provided by Equation 3.3, the probability of an uncaused event occurrence depends on the time which has elapsed since the previous uncaused event, in a manner which is consistent with SR theory since it can be interpreted with respect to Reid rebound theory where the ground state level is reached only for uncaused events and all other events are causing smaller impact on the strain accumulation/reduction. Since the branching structure is used to determine the time of the last uncaused event, we will refer to this model as the B-SR-ETAS model (Branched SR-ETAS).

When working with real earthquake catalogues we do not know which events in the sequence are mainshocks since we do not have access to the true branching structure. Indeed, the branching structure is usually estimated as a byproduct of the standard algorithms used to estimate the ETAS model [Ross, 2018a, Rasmussen, 2013, Veen and Schoenberg, 2008]. However we cannot use this idea directly since we are caught in a vicious circle: our parameter estimation requires access to the branching structure to define the mainshock earthquakes, but we cannot get the branching structure without first estimating the model parameters! One approach is to marginalise the branching structures out of the joint distribution by summing over all 2^{n-1} unique branching structures, for a catalogue with length n. However this is computationally intractable for even a moderate value of n. As such, we will instead introduce a Monte Carlo approach for performing this inference in a computationally tractable way.

Since defining waiting times based on the previous uncaused event hence leads to computationally difficult parameter inference, we could instead define the waiting time w_t based on the occurrence time of the last earthquake prior to t, regardless of whether it was an uncaused or an offspring. At time t, the time of the last event is given by t_E where $E = \max\{i|t_i < t\}$. The waiting time in this case is hence:

$$w_t = t - t_E,$$

with $\mu(t)$ defined according to Equation 3.3 as before. Under an SR interpretation, this implies that the strain accumulated with respect to the uncaused events causation can be assumed to reduce to a ground level - the minimum strain that can be observed in the system after every earthquake in the sequence, which corresponds to Reid's elasticity rebound theory in which an event occurs when a specific intensity threshold is reached [Reid, 1910, Matthews et al., 2002]. We denote this model by F-SR-ETAS (Full SR-ETAS).

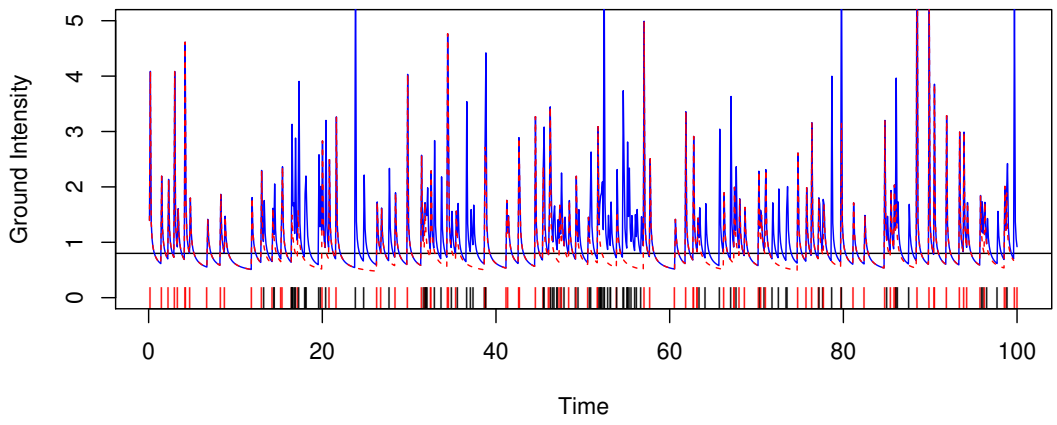

Figure 3.1: The ground (uncaused) intensity with respect to simulated data for which the uncaused events are illustrated with | and the caused ones with |, for each of the three models: $-$ μ_0 standard ETAS; - - $\mu(t - t_{I_t})$ B-SR-ETAS and $-$ $\mu(t - t_E)$ F-SR-ETAS.

The difference between ETAS, B-SR-ETAS and F-SR-ETAS models can be outlined clearly with respect to the shape of equivalent uncaused events intensity functions for each of them μ_0, $\mu(t - t_{I_t})$ and $\mu(t - t_E)$ respectively. On Figure 3.1 are shown the three intensity curves. Although the area under each of the curves is the same, the very spiky peaks for SR-ETAS based models are very evident. Further, the difference between F-SR-ETAS and B-SR-ETAS can be clearly outlined when a group of caused events is present. Such can be seen for $t \in (17, 19) \cup (35, 40) \cup (52, 58)$. Combining the SR-ETAS uncaused events spiky intensity structure with the inherently spiky ETAS intensity structure (see Figure 2.3) creates a very flexible model that can address data with varying behaviour.

The previously introduced in Section 3.2 concept of parameter set θ can be adapted for both SR-ETAS models as $\theta = \{\theta_{SR}, \alpha, c, p, k\}$, where θ_{SR} is taking the parameters of the waiting time distribution $F_w(\cdot)$. For ease of notation from here onward in this Chapter we will notate the ground intensity at time t with $\mu(t)$ instead of a $\mu(t|w_t)$ where w_t is the waiting time taken into interest with respect to event with time t.

3.4 Waiting Time Distributions

Regardless of which of the two approaches (B-SR-ETAS or F-SR-ETAS) we take when defining the waiting times w_t, we must specify a probability model F_w which governs

their distribution. In standard ETAS, the Poisson assumption results in a memoryless Exponential distribution. In contrast, the SR approach implies other forms of distributions with non-constant hazard rate. There is some controversy in the seismological literature over the appropriate waiting time distribution for modelling the time between mainshocks. As such, we will consider two different distribution which have been found to have strong empirical support: the Brownian Passage Time, and the Gamma.

3.4.1 Brownian Passage Times (BPT) immigration

The Brownian Passage Times concept was introduced to describe the inter-arrival times of earthquake events by [Ellsworth et al., 1999] and [Matthews et al., 2002]. It is a probabilistic physically based approach for addressing event recurrence based on long-term, load-state process assuming behaviour similar to those of the Brownian relaxation oscillator (BRO). In it, earthquakes are assumed to be an energy release of a tectonic system that accumulates strain. This approximates the event inter-arrival time probability density function as:

$$f(w_t;\lambda,\nu) = \left[\frac{\lambda}{2\pi\nu^2 w_t^3}\right]^{\frac{1}{2}} e^{-\frac{(w_t-\lambda)^2}{2\lambda\nu^2 w_t}}.$$

The cumulative distribution function of the BPT has a closed form which is the following:

$$\begin{aligned} F(w_t) = P(T \leq w_t) &= \int_0^{w_t} f(u)du \\ &= \Phi[u_1(w_t)] + e^{\frac{2}{\tilde{\alpha}^2}}\Phi[-u_2(w_t)], \end{aligned}$$

where

$$\Phi(w_t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{w_t} e^{-\frac{u^2}{2}}du, \tag{3.4}$$

for

$$u_1(w_t) = \nu^{-1}[w_t^{1/2}\lambda^{-1/2} - w_t^{-1/2}\lambda^{1/2}],$$

$$u_2(w_t) = \nu^{-1}[w_t^{1/2}\lambda^{-1/2} + w_t^{-1/2}\lambda^{1/2}].$$

The main attributes of BPT compared to other SR distributional alternatives are the following:

1. The mean waiting time, λ, of the events of interest provides a threshold until which the probability of event occurrence is continuously increasing. After reaching the mean waiting time, the conditional probability of occurrence is time independent

and relies only on the aperiodicity parameter, ν, which is associated with the scaling of the Brownian motion on which the process relies on.

2. Earthquake occurrence corresponds to immediate stress release to ground base level. Thus, the probability of immediate events recurrence is zero.

From here after the BPT based SR-ETAS models will be referred to as F-B-ETAS for the Full Stress Release Brownian Passage Time Epidemic After Shock Sequence model and B-B-ETAS for the branched one.

3.4.2 Gamma process immigration

The Gamma distribution is an exponential family distribution with two parameters, namely shape parameter $a > 0$ and scale parameter $s > 0$. It has been found by [Kagan and Knopoff, 1984, Chen et al., 2013, Wang et al., 2012] to provide a good model for main shock inter-arrival times. The Gamma distribution probability density function is:

$$f(w_t) = \frac{1}{s^a \Gamma(a)} w_t^{a-1} e^{-w_t/s},$$

where $\Gamma(a) = \int_0^\infty u^{a-1} e^{-u} du$ is the Gamma function. The corresponding cumulative distribution function is:

$$F(w_t) = \frac{1}{s^a \Gamma(a)} \int_0^{w_t} u^{a-1} e^{-u/s} du.$$

From here after the Gamma based SR-ETAS models will be referred to as F-G-ETAS for the Full SR Gamma ETAS model and B-G-ETAS for the branched one.

3.5 Estimation

We now consider parameter estimation for the SR-ETAS models. This includes estimating the ETAS model parameters $\theta_\Phi = (\alpha, c, p, k)$, as well as θ_{SR}, the parameters of the waiting time distribution F_w. Let $\theta = (\theta_{SR}, \theta_\Phi)$ denote the full set of unknown parameters. We perform Bayesian inference for the model parameters by developing a latent variable MCMC scheme that allows sampling from the full posterior.

3.5.1 Likelihood Function

The likelihood function of an ETAS process within period $[0, T]$ is the probability of the process's associated detection history, $\mathscr{H}_T$, combined with the probability of not detecting any other events within the period of interest (see Section 2.4.1 for more details):

$$\mathcal{L}(\theta; \mathscr{H}_T) = \prod_{i=1}^{n} \lambda(t_i|\mathscr{H}_T) e^{-\int_0^T \lambda(z|\mathscr{H}_T)dz}, \tag{3.5}$$

with corresponding log-likelihood:

$$\ell(\theta; \mathscr{H}_T, Z) \quad = \quad \sum_{i=1}^{n} \log(\lambda(t_i|\mathscr{H}_T)) \; - \; \int_0^{t_n} \mu(s)ds \; - \; \sum_{i=1}^{n-1} \kappa(m_i) \int_0^{t_n - t_i} g(s)ds. \tag{3.6}$$

Plugging in the specific choices for the offspringing functions $\kappa(\cdot)$ and $g(\cdot)$, as introduced in Section 3.2, in the ETAS model gives:

$$\begin{aligned} \ell(\theta; \mathscr{H}_T, Z) = \sum_{i=1}^{n} log \left[\mu(t) + \sum_{j=1}^{i-1} \frac{k e^{\alpha(m_j - M_0)}}{(t_i - t_j + c)^p} \right] \\ - \int_0^{t_n} \mu(s)ds - \sum_{i=1}^{n} k e^{\alpha(m_i - M_0)} \left(1 - \frac{c^{p-1}}{(t_n - t_i + c)^{p-1}} \right), \end{aligned} \tag{3.7}$$

where θ is the set of all parameters in the model and $Z_{1:n} \in \{0, 1\}^n$ is a vector with length n indicating whether each event is main/uncaused (1) or not (0). As of the branching structure introduced in Figure 2.2, the causality information is $Z = \{1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0\}$.

Note that $\mu(\cdot)$ depends on the branching vector Z for B-SR-ETAS since the intensity of the background process depends on the time at which the last uncaused event occurred. However since the true branching structure is not known in practice, it must be marginalised out by summing over all 2^{n-1} possible values. Therefore, the log-likelihood function of the B-SR-ETAS model is:

$$\ell(\theta; \mathscr{H}_T) = \sum_{j=1}^{2^{n-1}} \ell(\theta; \mathscr{H}_T, Z = z_j) P(Z = z_j|\theta).$$

In practice, this summation is likely to be intractable for even moderate n. As such, we will instead use a latent variable formulation where the unknown branching vector Z is treated as a parameter to be learned. In order to evaluate this quantity we can either

use a single "best" quantity or to provide a Monte Carlo approximation of it based on sampling multiple branching structures based on the true/optimised parameters θ.

While the proposed by [Wheatley et al., 2016] log-likelihood function is conceptually the same as the one shown above, [Chen and Stindl, 2018] Section 3, Remark 1, claims that the log-likelihood form is wrong with respect to the examined by them RHawkes process. In its essence, Rhawkes process is an unmarked branched-SR-ETAS model. The full algorithm that is proposed for the calculation of the (log-)likelihood of RHawkes by [Chen and Stindl, 2018] is the following.

Direct log-likelihood calculation for RHawkes

The overall scope of the defined by [Wheatley et al., 2016, Chen and Stindl, 2018] RHawkes process is very similar to B-SR-ETAS and relies on the following intensity function:

$$\lambda^{Rh}(t) = \mu(t - t_{I_{[t]}}) + \sum_{t_i<t} \eta r(t - t_i),$$

where $\int_0^\infty r(u)du = 1$, $I_{[i]} = \max_j\{j|t_j < t_i \text{ and } B_j = 0\}$ and $B = \{B_1, ..., B_n\}$ is a branching realisation as described in Section 3.3.

The proposed form for the likelihood is the following:

$$\mathcal{L}(\theta; \mathscr{H}_T) = \begin{cases} \exp(-U(T)), & n = 0 \\ \mu(t_1 - t_0)\exp(-U(t_1) \times -U(T - t_1) - \eta R(T - t_1)), & n = 1 \\ \mu(t_1 - t_0)\exp(-U(t_1) \times \prod_{i=2}^{n} \sum_{j=1}^{i-1} p_{ij} d_{ij} \times \sum_{j=1}^{n} s_{n+1} p_{n+1,j}, & n \geq 2 \end{cases} \tag{3.8}$$

where

$$d_{ij} = (\mu(t_i - t_j) + \sum_{t<t_i} \eta r(t_i - t)) \times \exp(-U(t_i - t_j) + U(t_{i-1} - t_j) - \eta \sum_{t<t_i} R(t_i - t) + \eta \sum_{t<t_{i-1}} R(t_{i-1} - t)) \tag{3.9}$$

and

$$s_{n+1,j} = \exp\Big\{ - \big[U(T - t_j) - U(t_n - t_j)\big] - \eta\Big[\sum_{t<T} R(T - t) - \sum_{t<t_n} R(t_n - t)\Big]\Big\}. \tag{3.10}$$

The $p_{ij}, i = \{2, ..., n+1\}, j = \{1, ..., i-1\}$ are obtained by initiating $p_{21} = 1$ and updated recursively as follows

$$p_{ij} = \begin{cases} \frac{\sum_{t<t_{i-1}} \eta r(t_{i-1}-t))}{\mu(t_{i-1}-t_j)+\sum_{t<t_{i-1}} \eta r(t_{i-1}-t))} \times \frac{d_{i-1,j} p_{i-1,j}}{\sum_{j=1}^{i-2} p_{i-1,j} d_{i-1,j}}, & j = 1, ..., i-2 \\ 1 - \sum_{k=1}^{j-2} p_{ik} & j = i-1 \end{cases} \tag{3.11}$$

for $i = 3, ..., n+1$;

$$U(t) = \int_0^t \mu(s)ds \quad \text{and} \quad R(t) = \int_0^t r(s)ds.$$

The evaluation of d_{ij} is only feasible based on Stress Release/ Renewal density that has explicit intensity function. Otherwise, the quantity in Equation 3.3 will be numerically unstable as previously discussed. Thus, the method is not directly applicable to a general family of renewal process distributions (e.g. B-B-ETAS).

Limitations of the Direct log-likelihood approach

The direct log-likelihood calculation introduced above requires the calculation of probabilities associated with all possible inheritance structures. The ground intensity as of Equation 3.3 has to be evaluated for all possible temporal lags when in the calculation of Equations 6-8, Section 3 of [Chen and Stindl, 2018]. As discussed before, such expression cannot be evaluated for immigrant distributions that do not have explicit intensity function (Equation 3.3). However, from a Bayesian prospective the branching structure is a feature that we learn. Rather than being an unknown quantity, it is a data characteristics that we evaluate based on our inheritance believes. Thus, the provided log-likelihood function in Equation 3.6 is feasible for the scope of a Bayesian algorithm.

3.5.2 Bayesian analysis

We will use Markov chain Monte-Carlo techniques to obtain samples from the posterior distribution of the ETAS parameters θ. As shown by [Ross, 2018b], the efficiency of MCMC algorithms for ETAS type models can be drastically improved by including the branching structure as a latent variable. What is more, the full conditional MCMC MH techniques are only applicable for the ETAS and F-SR-ETAS, while the B-SR-ETAS is only feasible to be implemented based on the latent variable approach since the branching structure is needed to obtain the last immigrant event time used in the

evaluation of $\mu(\cdot)$. To introduce this latent variable approach, we first make some small reparameterizations of the introduced in Section 3.2 (SR-)ETAS intensity function:

$$\begin{aligned}\Phi(t|\mathscr{H}_t)) &= \sum_{t_i<t} g(t-t_i)\kappa(m_i) \\ &= \sum_{t_i<t} \frac{k}{(t-t_i+c)^p} e^{\alpha(m_i-M_0)} \\ &= \sum_{t_i<t} \frac{(p-1)c^{p-1}}{(t-t_i+c)^p} K e^{\alpha(m_i-M_0)} \\ &= \sum_{t_i<t} h(t-t_i)\iota(m_i),\end{aligned}$$

where $\iota(m_i) = Ke^{\alpha(m_i-M_0)}$, $K = \frac{k}{(p-1)c^{p-1}}$, and $h(z) = (p-1)c^{p-1}\frac{1}{(z+c)^p}$ is a reparameterisation of $g(\cdot)$ that now integrates to 1 [Vere-Jones and Davies, 1966]. The log-likelihood function as of Equation 3.7 is then:

$$\ell(\theta; \mathscr{H}_T, Z) = \sum_{i=1}^{n} log\left[\mu(t) + \sum_{j=1}^{i-1} \frac{K(p-1)c^{p-1}e^{\alpha(m_j-M_0)}}{(t_i-t_j+c)^p}\right] - \int_0^{t_n} \mu(s)ds - \sum_{i=1}^{n} Ke^{\alpha(m_i-M_0)}\left(1 - \frac{c^{p-1}}{(t_n-t_i+c)^{p-1}}\right). \quad (3.12)$$

Performing MCMC directly is difficult due to the correlation between some of the model parameters. We employ the introduced in Section 2.4.3 latent variable MCMC sampler. For its application, we have to sample branching structures from their full posterior distribution.

Branching procedure

Let B denote the branching structure vector where $B_i = j$ indicates that the i^{th} event in the sequence is a descendant of the j^{th} event ($j < i$). Immigrant events are notated as uncaused i.e. caused by an event with index 0. If we refer again to the branching structure, that was introduced on Figure 2.2, we can visually assign corresponding values for our branching inheritance measure B_i as follows $B = \{0, 1, 1, 3, 1, 0, 6, 7, 6, 0, 8\}$. The immigrant events are coming from a in-homogeneous Poisson process with intensity function $\mu(\cdot)$ while the offspring events of the j^{th} event are generated from in-homogeneous Poisson process with intensity $h(t_i - t_j)\iota(m_j)$. Assuming that each event in the sequence is generated by a single process, we can assign probabilities distribution to each event

with respect to its branching pedigree and therefore sample a branching structure from its conditional posterior as follows:

1. Initiate the branching by setting $B_1 = 0$ as we assume that always the first term is immigrant.

2. Sample each B_i in turn from $P(B_i|\mathscr{H}_T, \theta, B_{1:(i-1)})$.

3. Return the sequence of generated B_is.

The form of $P(B_i|\mathscr{H}_T, \theta, B_{1:(i-1)})$ in the general SR-ETAS model is substantially more complex than for the standard ETAS, since using a general renewal process for the mainshocks introduces substantial dependence in the process. It was previously shown in [Ross, 2018a] that for a standard ETAS model, the conditional posterior for each B_i is independent of all other B_j for $i \neq j$ and can be written as:

- $P(B_i = 0|\mathscr{H}_T, \theta, B_{1:(i-1)}) = \frac{\mu(t_i - t_{I_{[i]}})}{\mu(t_i - t_{I_{[i]}}) + \Phi(t_i|\mathscr{H}_{t_i}))}$
- $P(B_i = j|\mathscr{H}_T, \theta, B_{1:(i-1)}) = \frac{h(t_i - t_j)\iota(m_j)}{\mu(t_i - t_{I_{[i]}}) + \Phi(t_i|\mathscr{H}_{t_i}))}$ for j in 1 to $i-1$

where $I_{[i]}$ takes the last immigrant event index before the i^{th} event to obtain branching for B-SR-ETAS, and $I_{[i]} = i - 1$ for F-SR-ETAS and ETAS models.

However for our more general SR-ETAS models with renewal process immigration, this independence no longer holds. In order to illustrate the problem, we have to re-derive the above expression first. To evaluate the branching probabilities we have to address the conditional probability of the branching (B) given the data Y, model parameters $\theta = \{\theta_{SR}, \theta_\Phi\}$ and potentially the previous obtained branching structure(s). The general likelihood function as introduced in Equation 3.5 can be re-arranged to include the information carried by the uncaused and offspring processes as follows:

$$\mathcal{L}(\theta; \mathscr{H}_T, B) = e^{-\int_0^T \mu(z|\mathscr{H}_T, B)dz} \prod_{i:B_i=0} \mu(t_i|\mathscr{H}_T, B) \times \prod_{j=1}^{n} \left[e^{-\int_0^{T-t_j} \Phi(z|\mathscr{H}_T)dz} \prod_{i:B_i=j}^{n} \Phi(t_i - t_j|\mathscr{H}_T, B) \right]. \quad (3.13)$$

For ETAS and Full-ETAS models, the branching structure (B) is not needed for the evaluation of Equation 3.3, thus the term $\int_0^T \mu(z|\mathscr{H}_T)dz$ is independent of B. The conditional distribution of the branching structure under a flat prior assumption is:

$$P(B|\mathscr{H}_T, \theta_{SR}, \theta_\Phi) \propto \prod_{i:B_i=0} \mu(t_i) \prod_{j=1}^{n} \prod_{i:B_i=j}^{n} \Phi(t_i - t_j),$$

which lead us, as required, to the following branching updates:

- $P(B_i = 0|\mathscr{H}_T, \theta, B_{1:(i-1)}) = \frac{\mu(t_i - t_{I_{[i]}})}{\mu(t_i - t_{I_{[i]}}) + \Phi(t_i|\mathscr{H}_{t_i}))}$
- $P(B_i = j|\mathscr{H}_T, \theta, B_{1:(i-1)}) = \frac{h(t_i - t_j)\iota(m_j)}{\mu(t_i - t_{I_{[i]}}) + \Phi(t_i|\mathscr{H}_{t_i}))}$ for j in 1 to $i-1$

where $I_{[i]} = \max_j\{j|t_j < t_i \text{ and } B_j = 0\}$.

In the case of Branched-ETAS or Rhawkes, as defined by [Wheatley et al., 2016,Chen and Stindl, 2018], $\int_0^T \mu(z|\mathscr{H}_T, B)$ requires a branching realisation B for the calculation of the ground intensity as of Equation 3.3. Let us define $z_0, z_1, ..., z_m + 1$ to be, $z_0 = 0$, $z_1, ..., z_m$ - all m immigrant events' times that were estimated to occur in the catalogue and $z_{m+1} = T$. Thus, we can re-write the full data likelihood components that depend on $\mu(\cdot)$ as follows:

$$\begin{aligned} e^{-\int_0^T \mu(z|\mathscr{H}_T,B)dz} \prod_{i:B_i=0} \mu(t_i|\mathscr{H}_T, B) = \\ e^{-\int_0^{z_1} \mu(z)dz - \int_{z_1}^{z_2} \mu(z)dz - ... - \int_{z_m}^{z_{m+1}} \mu(z)dz} \prod_{i=1}^{m} \mu(z_i - z_{i-1}) \propto \\ e^{-\int_{z_{i-1}}^{z_i} \mu(z)dz - \int_{z_i}^{z_{i+1}} \mu(z)dz} \mu(z_i - z_{i-1})\mu(z_{i+1} - z_i). \quad (3.14) \end{aligned}$$

Based on the above approximation, the full conditional posterior distribution of the branching structure is:

$$\begin{aligned} P(B_i|Y, \theta, B) \propto \left[e^{-\int_{t_{I_{[i]}}}^{t_i} \mu(z)dz - \int_{t_i}^{t_{I^*_{[i]}}} \mu(z)dz} \mu(t_i - t_{I_{[i]}})\mu(t_{I^*_{[i]}} - t_i)\right]^{\mathbb{1}_{B_i=0}} \times \\ \left[e^{-\int_{t_{I_{[i]}}}^{t_{I^*_{[i]}}} \mu(z)dz} \mu(t_{I^*_{[i]}} - t_{I_{[i]}})\right]^{\mathbb{1}_{B_i\neq 0}} \times \prod_{j=1}^{n} \Phi(t_i - t_j)^{\mathbb{1}_{B_i=j}}, \quad (3.15) \end{aligned}$$

where $I^*_{[i]} = \min_j\{j|t_j > t_i \text{ and } B_j = 0\}$ and $\mathbb{1}_x$ is 1 if x is True and 0 otherwise.

Thus, the required branching probability updates are:

- $P(B_i = 0|\mathscr{H}_T, \theta, B) \propto \mu(t_i - t_{I_{[i]}}) \times e^{-\int_{t_{I_{[i]}}}^{t_i} \mu(z)dz - \int_{t_i}^{t_{I^*_{[i]}}} \mu(t|B_i=0)dt} \mu(t_{I^*_{[i]}} - t_i)$
- $P(B_i = j|\mathscr{H}_T, \theta, B) \propto \Phi(t_i|\mathscr{H}_{t_i})) \times e^{-\int_{t_{I_{[i]}}}^{t_{I^*_{[i]}}} \mu(t|B_i\neq 0)dt} \mu(t_{I^*_{[i]}} - t_{I_{[i]}})$ for j in 1 to $i-1$

The expressions above can be further simplified to those stated in Equations 3.16 and 3.17 since $e^{-\int_{t_{I_{[i]}}}^{t_i} \mu(z)dz}$ has the same value regardless of the state of the branching assignment of t_i.

- $P(B_i = 0|\mathscr{H}_T, \theta, B) \propto \mu(t_i - t_{I_{[i]}}) \times e^{-\int_{t_i}^{t_{I^*_{[i]}}} \mu(t|B_i=0)dt} \mu(t_{I^*_{[i]}} - t_i)$ (3.16)
- $P(B_i = j|\mathscr{H}_T, \theta, B) \propto \Phi(t_i|\mathscr{H}_{t_i})) \times e^{-\int_{t_i}^{t_{I^*_{[i]}}} \mu(t|B_i\neq 0)dt} \mu(t_{I^*_{[i]}} - t_{I_{[i]}})$ (3.17)

  for j in 1 to $i-1$

where at each time t_i we write the occurrence time of the first uncaused event after t_i as $t_{I^*_{[i]}}$ where $I^*_{[i]} = \min_j\{j|t_j > t_i \text{ and } B_j = 0\}$.

Log-likelihood latent variable transformations

The (log-)likelihood function of the process can be rewritten conditional on the branching structure. From Equation 3.1 the process intensity at time t is a sum of the contribution of $\mu(t)$ from the background process, and a contribution of $h(t - t_i)$ for each of the previous event t_i. Let us define S_0 to be the set of all uncaused events (conditional on the branching structure), and S_i to be the set of all events triggered by each event t_i. We write $|S_i|$ to denote the number of events in each set. For a given branching structure B, the likelihood function can then be rewritten as:

$$\mathcal{L}(\theta; \mathscr{H}_T, B) = e^{-\sum_{t_i \in S_0} \int_0^{t_i - t_{i-1}} \mu(u)du} \prod_{s \in S_0} \mu(s) \times \prod_{j=1}^{n} \left(e^{-\iota(m_j) \int_0^{t_n - t_{i-1}} h(u)du} \iota(m_j)^{|S_j|} \prod_{t_i \in S_j} h(t_i - t_j) \right). \quad (3.18)$$

In this notation B is a full branching structure realisation, and the integrals are summed over all immigrant events except the first one since there is no waiting time for the first event. The permutation over $\mu(s)$ is a permutation of the spot values of $\mu(\cdot)$ at the triggering times of all immigrant events in the catalogue and θ_{SR} represents the parameter set of the chosen SR distribution. Note that $\mu(t)$ in this case is actually $\mu(t|w_t) = \frac{f_w(w_t)}{1-F_w(w_t)}$, where w_t is the waiting time from the last immigration for the B-SR-ETAS model and the waiting time between every event for the F-SR-ETAS. The $f_w(\cdot)$ and $F_w(\cdot)$ are the corresponding PDF and CDF of the candidate immigration distribution (SR). Additional approximation can be obtained based on the so-called "infinite time assumption". It is considered to hold true for large catalogue end time

($t_n \to \infty$) and it states that the integral over the Modified Omori law ($h(\cdot)$) for the range of values in the catalogue converges to 1:

$$\lim_{t_n \to \infty} \int_0^{t_n - t_{i-1}} h(u)du = 1. \tag{3.19}$$

Based on the branching structure, the likelihood function (and hence the posterior, given independent priors) in Equation 3.18 is separable into three functions that can update separately the following parameters' sets θ_{SR}, $\{K, \alpha\}$ and $\{c, p\}$. They have the following posterior probabilities that will be used for the Metropolis-Hastings accept-reject ratios:

$$\log(P(\theta_{SR}|\mathscr{H}_T, \theta, B)) \propto \log(\pi(\theta_{SR})) + \sum_{t_i \in S_0} \Big(\log(\mu(t_i)) - \int_0^{t_i - t_{i-1}} \mu(u)du \Big) \tag{3.20}$$

$$\log(P(K, \alpha|\mathscr{H}_T, \theta, B)) \propto \\ \log(\pi(K, \alpha)) - \sum_{j=1}^{n} \left(\iota(m_j) \Big(1 - \frac{c^{p-1}}{(t_n - t_j + c)^{p-1}} \Big) - |S_j| \log(\iota(m_j)) \right) \tag{3.21}$$

$$\log(P(c, p|\mathscr{H}_T, \theta, B)) \propto \\ \log(\pi(c, p)) - \sum_{j=1}^{n} \left(\iota(m_j) \Big(1 - \frac{c^{p-1}}{(t_n - t_j + c)^{p-1}} \Big) - \sum_{t_i \in S_j} \log(h(t_i - t_j)) \right) \tag{3.22}$$

Note that based on the infinite time assumption, as defined in Equation 3.19, the term $\frac{c^{p-1}}{(t_n - t_j + c)^{p-1}}$ is effectively zero and as such the above expressions can be simplified so that every posterior probability to be independent from the other parameters' behaviour. Based on the infinite time assumption are achieved three independently updated chains that have interaction only when a new full branching structure is sampled. The conducted analysis was carried out without the infinite time assumption for all datasets because it appeared that there are large discrepancies for the North California catalogue that we use in Section 3.6.2 to illustrate and compare the introduced models' behaviour.

Choice of Prior and Proposal distributions

The (SR-ETAS) parameter estimates in this Chapter were obtained by running a latent variable MCMC for each of the 5 proposed models - ETAS, B-B-ETAS, F-B-ETAS, B-G-ETAS and F-G-ETAS. We use non-informative priors for all (SR-) ETAS parameters. For the standard ETAS model there exists a conjugate Gamma prior for the fixed ground intensity μ [Ross, 2018a]. However, we decided to use the same prior distribution for all models to obtain results that only differentiate based on the underlying models' structure. We used a flat Uniform prior for θ_{SR}, α, $\log(c)$, $\log(p)$ and $\log(K)$ with bounds $\alpha \in [0, 10], c \in [0, 10], p \in [1, 30], K \in [0, \infty]$, although more informative priors could be used if desired. These bounds present unrestricted estimation in terms of maximised likelihood and posterior values. Further, for a infinite time catalogue, the overall productivity of the offspring decay, i.e. the mean number of off springs by every event is approximated by K and α thus we might want to restrict their values to provide average offspring productivity smaller than 1. However, this is beyond the scope of this Chapter because developing such framework might further induce dependencies between model parameters and could affect negatively MCMC mixing. The support of all parameters is greatly influenced by the potential multimodality and were taken to be identical to those used in the bayesianETAS R package [Ross, 2018a].

We use as a proposal distribution a Normal with standard deviation of 0.1 for all parameters that require Metropolis-Hastings updates. The New Madrid catalogue parameters' sequences are with overall length of 15000 after burn-in of 5, 100 and 100 for the θ_{SR}, $\{K, \alpha\}$ and $\{c, p\}$ respectively. The branching structure was sampled from its conditional posterior at every iteration. The North California catalogue is much larger, thus we updated the branching structure less frequently at every 20 iterations of the Gibbs sampler, overall 12000 parameter sets were obtained after burn in of 4, 100, 20 for the θ_{SR}, $\{K, \alpha\}$ and $\{c, p\}$ respectively.

3.6 Applications

In this section we discuss and compare the model fit across ETAS-based models on two seismic catalogue of interest. The first one is the New Madrid catalogue which is much smaller than a conventional earthquake catalogue but of great interest for underwriting community. This causes a lot of difficulties in estimation context due to the lack of

consistent presence of large earthquakes. The second dataset, the North California, is more dense and should behave similarly to a typical single fault catalogue.

3.6.1 New Madrid seismic sequence

We first compare the performance of the ETAS and SR-ETAS models on the catalogue of New Madrid earthquakes obtained from The University of Memphis website `http://www.memphis.edu/ceri/seismic/catalog.php`. This catalogue starts on 29/06/1974 and ends on 23/02/2017. Only earthquakes of magnitude greater than 3 are considered since smaller ones are typically considered harmless. The resulting catalogue contains 308 events. We fit the ETAS model and BPT and Gamma based SR-ETAS models to this catalogue.

Figure 3.2 shows how the sequence of log-likelihoods for each model evolves over each iteration of the Gibbs sampler (after convergence). It is clearly observable that there is a difference between the overall fitting capabilities among the 5 models. What is more, the overall mixing for branched SR models is greater and relatively more symmetric. Figure 3.4. plots the posterior distribution of the model parameters for the B-B-ETAS, which is the most difficult model to estimate due to the need to estimate the unknown branching structure. The posterior distributions for the parameters in the other models are similar. As expected, the obtained parameters' distributions are smooth, symmetric and not very different from a bell-shaped based form.

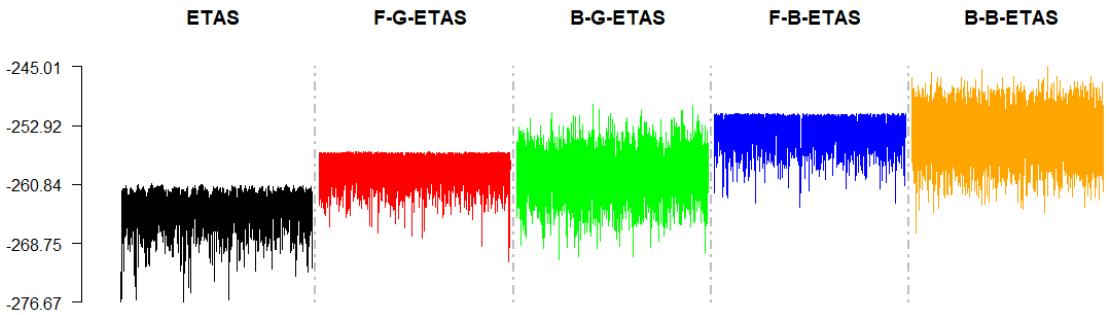

Figure 3.2: Log-likelihood of the MCMC sequences based on the used full branching structures for the New Madrid catalogue with respect to ETAS/F-G-ETAS/B-G-ETAS/F-B-ETAS/B-B-ETAS

The Goodness-of-fit and model comparison results are shown in Table 3.1. Among all ETAS–based models it appears that SR-ETAS models are superior to the standard ETAS model according to both BIC and DIC. BPT based models are better compared to their corresponding Gamma alternatives and B-SR-ETAS models are superior to the

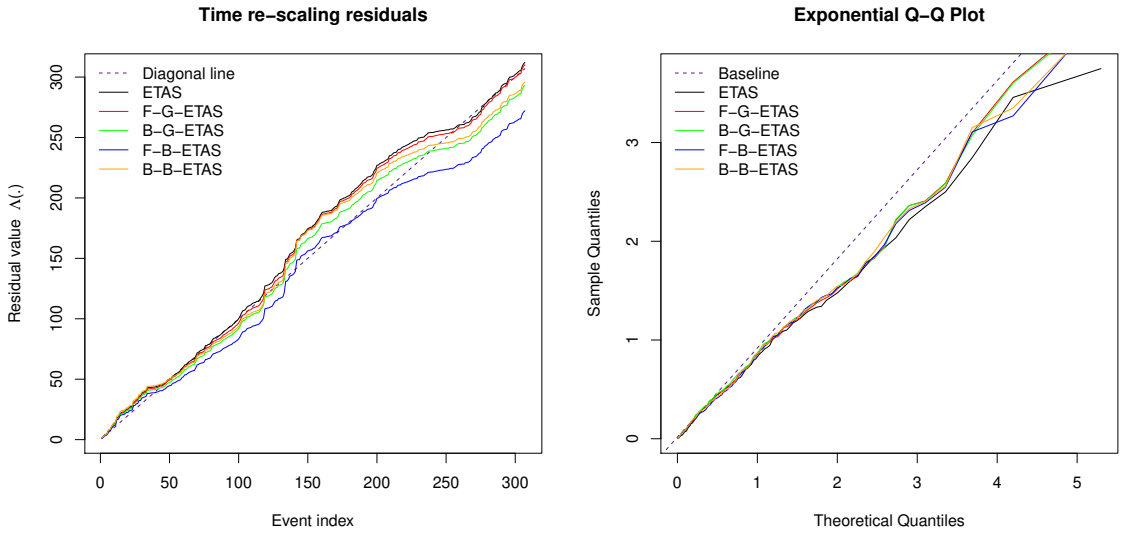

Figure 3.3: Time re-scaling diagnostic plots for the New Madrid catalogue.

F-SR-ETAS. The best model within all examined models is evidently the B-B-ETAS. However, the time-residual analysis at every 10% of the data suggest that some of the SR-ETAS might not be superior to standard ETAS model. The standard ETAS provides best fit for $T = \{20\%, 30\%, 90\%\}$ of the overall temporal catalogue length while F-G-ETAS is superior for $T = \{100\%\}$, B-G-ETAS for $T = \{10\%, 40\%, 50\%, 60\%, 70\%, 80\%\}$. B-G-ETAS and B-B-ETAS are never better than all the rest. Hence, we expect that informally B-G-ETAS provides the best fit with respect to the raw time re-scaling.

However, Figure 3.3 presents the informal diagnostic plots of the time residuals. On the left are the raw time residuals for all 5 models versus a diagonal line. Ideally these should overlap. The overall pattern is very similar for all models. They all experience a bias towards the middle of the catalogue which might indicate a potential minor nonstationarity in the data [Kumazawa and Ogata, 2013]. The right part of Figure 3.3 shows a Q-Q plot for the residuals of all 5 models versus Exp(1) distribution. All 5 models behave similarly, with minor spread from the expected results for large quantiles.

The formal time re-scaling diagnostic tests conclude that the KS, CVM, AD and ER tests were passed by all 5 models at the 5% significance level. The LB test is passed only by F-B-ETAS at 5% significance level, while all other models pass it at 1% significance level. Thus, there might be minor dependence in the residuals which we believe is negligible.

	ETAS	F-G-ETAS	B-G-ETAS	F-B-ETAS	B-B-ETAS
Log-likelihood	-260.75	-256.44	−250.13*	-251.28	−245.01*
Parameter count	5	6	6	6	6
BIC	275.05	273.23	−267.32*	268.47	−262.20*
DIC	519.82	515.47	484.28*	507.60	483.80*
$I(t_{1:\langle\frac{n}{10}\rangle}\|\theta)$	7.20	6.42	4.76*	6.30	8.90*
$I(t_{1:\langle\frac{2n}{10}\rangle}\|\theta)$	-1.02	-2.41	−5.14*	-1.93	−1.25*
$I(t_{1:\langle\frac{3n}{10}\rangle}\|\theta)$	-0.88	-3.79	−7.09*	-2.64	−3.75*
$I(t_{1:\langle\frac{4n}{10}\rangle}\|\theta)$	6.42	3.22	−0.86*	3.76	1.32*
$I(t_{1:\langle\frac{5n}{10}\rangle}\|\theta)$	23.64	22.59	16.98*	19.36	19.56*
$I(t_{1:\langle\frac{6n}{10}\rangle}\|\theta)$	20.82	18.95	12.19*	16.27	14.78*
$I(t_{1:\langle\frac{7n}{10}\rangle}\|\theta)$	24.03	21.65	14.42*	19.37	15.69*
$I(t_{1:\langle\frac{8n}{10}\rangle}\|\theta)$	13.14	9.90	1.27*	7.90	2.31*
$I(t_{1:\langle\frac{9n}{10}\rangle}\|\theta)$	-3.89	-6.23	−15.24*	-8.31	−16.36*
$I(t_{1:n}\|\theta)$	1.20	-1.03	−10.31*	-2.64	−15.04*

Table 3.1: Goodness-of-fit Summary - New Madrid; ETAS, BPT and Gamma based SR-ETAS. Lower values of the BIC/DIC indicate superior fit. $\langle x\rangle$ corresponds to the nearest integer larger than x.

* - The value is approximate

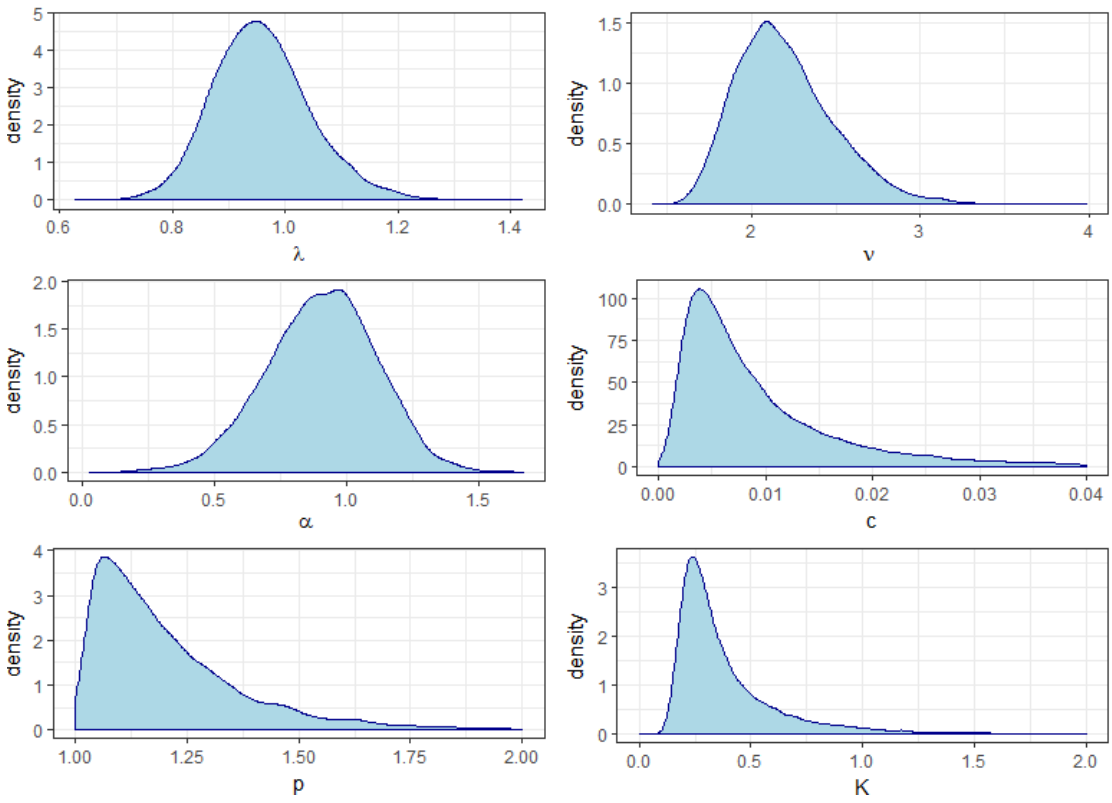

Figure 3.4: B-B-ETAS MCMC parameters' density for the New Madrid catalogue.

3.6.2 North California seismic sequence

The previous analysis was repeated using the North California seismic sequence. The catalogue of earthquake events can be obtained from `http://www.ncedc.org/ncedc/`

`catalog-search.html`. We took into account all detected events from 01/01/1987 until 31/12/2015, with magnitude of completeness of 3.5. This created a catalogue consisting of 3442 events.

	ETAS	F-G-ETAS	B-G-ETAS	F-B-ETAS	B-B-ETAS
Log-likelihood	-109.32	-103.68	-90.61^*	-110.58	-64.83^*
Parameter count	5	6	6	6	6
BIC	129.68	128.11	115.04^*	135.01	89.26^*
DICalt	1108.51	799.57	725.33^*	553.48	338.07^*
$I(t_{1:\langle\frac{n}{10}\rangle}\|\theta)$	-87.61	-70.80	-70.89^*	-75.71	-56.14^*
$I(t_{1:\langle\frac{2n}{10}\rangle}\|\theta)$	-140.34	-84.17	-84.36^*	-87.85	-64.46^*
$I(t_{1:\langle\frac{3n}{10}\rangle}\|\theta)$	-232.66	-116.09	-119.27^*	-112.73	-97.89^*
$I(t_{1:\langle\frac{4n}{10}\rangle}\|\theta)$	-344.43	-189.92	-195.67^*	-177.58	-170.48^*
$I(t_{1:\langle\frac{5n}{10}\rangle}\|\theta)$	-346.34	-169.76	-176.48^*	-151.79	-145.02^*
$I(t_{1:\langle\frac{6n}{10}\rangle}\|\theta)$	-371.49	-176.82	-184.65^*	-156.14	-145.84^*
$I(t_{1:\langle\frac{7n}{10}\rangle}\|\theta)$	-373.28	-166.22	-172.70^*	-152.49	-136.76^*
$I(t_{1:\langle\frac{8n}{10}\rangle}\|\theta)$	-299.36	-57.28	-61.52^*	-53.69	-30.97^*
$I(t_{1:\langle\frac{9n}{10}\rangle}\|\theta)$	-297.49	-14.87	-20.23^*	-11.35	15.01^*
$I(t_{1:n}\|\theta)$	-301.14	8.73	2.71^*	12.90	42.58^*

Table 3.2: Goodness-of-fit Summary - North California; ETAS, BPT and Gamma based SR-ETAS. Lower values of the BIC/DIC indicate superior fit. $\langle x\rangle$ corresponds to the nearest integer larger than x.

* - The value is approximate

The full sequences of the log-likelihood calculated using the Gibbs sampler are shown on Figure 3.5. As before, all the SR-ETAS models appear to give substantial improvements over the basic ETAS model. Again we decided to report the posterior density only for B-B-ETAS which are shown on Figure 3.7. The heavy tails that appeared for the New Madrid catalogue are not present. Overall the shapes of all 6 parameters appear to be roughly symmetric. The Goodness-of-fit results of the un-simplified (finite time) runs are shown on Table 3.2.

According to the BIC, the F-B-ETAS is the worst model while all other SR models are slightly better than the standard ETAS. Due to the larger number of observations in this catalogue, we decided to apply the DICalt that depends on the previously defined in Section 2.4.4 p_{DICalt}. According to it, the branched SR models are providing a considerable performance improvement compared to the full models while the standard ETAS performs the worst. For this catalogue BPT based models are no longer superior to their corresponding Gamma alternatives. Interestingly, F-B-ETAS is currently the worst model. This is probably attributed to the fact that Gamma-SR-ETAS models

are guaranteed to be at least as good as the ETAS model since they can reduce to it, since the Exponential interarrival time distribution used in the standard ETAS is nested inside the Gamma distribution. It is clear that B-B-ETAS has a great advantage among all other models.

Figure 3.6 presents the time residuals informal diagnostic plots. On the left are shown the residuals and on the right is shown the Q-Q plot for all 5 models. The residuals plot indicates that all SR-ETAS models behave very similarly, while standard ETAS indicates a flow from the desired pattern. Similarly to the results for the New Madrid catalogue, all models experience a bias towards the middle of the catalogue which might indicate a potential minor nonstationarity in the data [Kumazawa and Ogata, 2013]. The Q-Q plot again shows that SR-ETAS models provide a better realisation to Exp(1) distribution with minor overestimation for large quantiles, while standard ETAS underestimates it for a large proportion of the support.

The time-residual analysis at every 10% of the data suggest that the B-B-ETAS is dominating across all other models. It is superior to the rest for $T = \{10\%, 20\%, 30\%, 40\%, 50\%, 60\%, 70\%, 80\%, \}$, while F-B-ETAS and B-G-ETAS models are the best only at $T = 90\%$ and $T = 100\%$ respectively. The ETAS and F-G-ETAS are never superior across the range.

However, the formal time re-scaling diagnostic tests are not very conclusive. The KS test was failed by the standard ETAS model, passed at 1% significance level by F-B-ETAS and excelled at 5% significance level by the remaining 3 models. The CVM and AD tests were passed by all models except standard ETAS at the 5% significance level. All five models failed the LB test. We believe this was caused by the large sample size since overall there are no indications for dependence in the residuals. The ER test was passed by all non-Gamma based models at the 5% significance level indicating that Gamma distribution might induce an excessive dispersion in the residuals. All things considered, we conclude that ETAS model is not providing adequate fit to the North California catalogue, Gamma-SR-ETAS models are superior to it but some researchers might disregard them due to the present excessive dispersion in the residuals. The B-SR-ETAS is providing the most stable results, with Branched BPT ETAS being the superior model that should be chosen for this dataset.

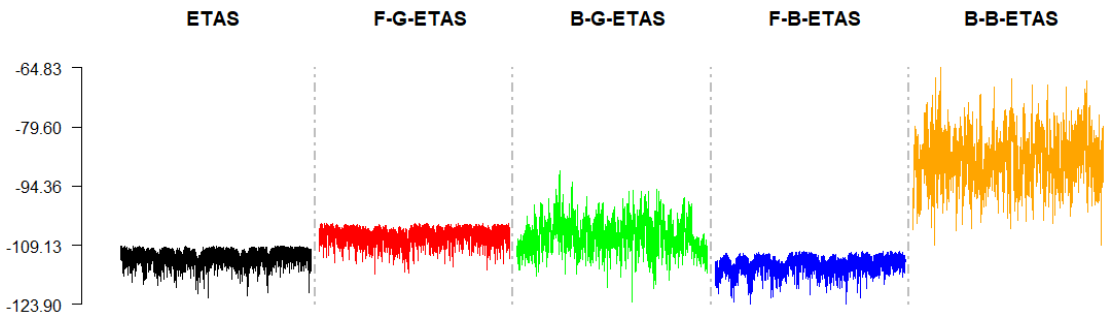

Figure 3.5: Log-likelihood of the MCMC sequences based on the used full branching structures for the North California catalogue with respect to ETAS/F-G-ETAS/B-G-ETAS/F-B-ETAS/B-B-ETAS

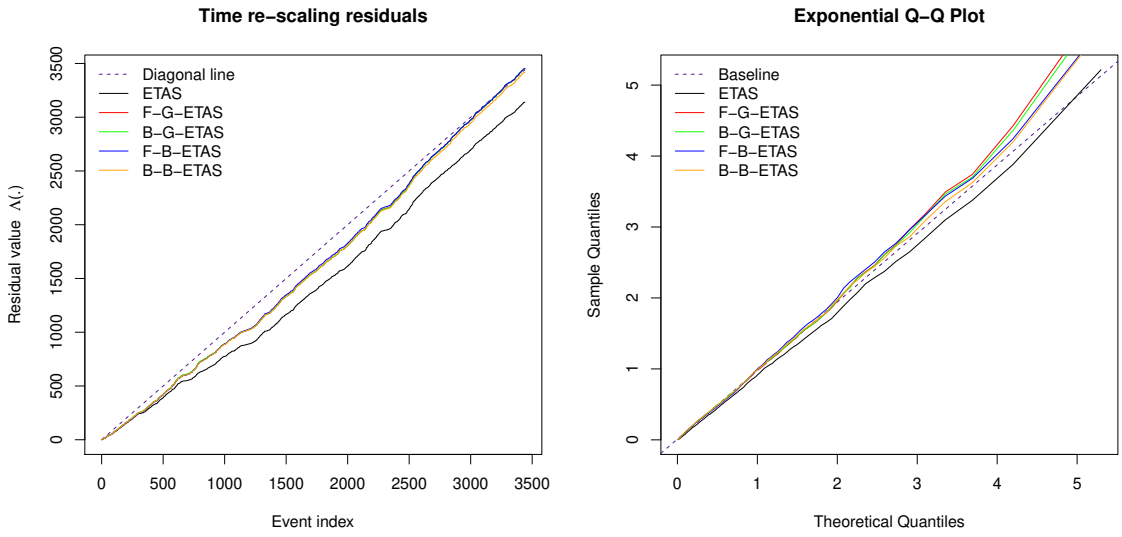

Figure 3.6: Time re-scaling diagnostic plots for the North California catalogue.

3.7 Conclusion

The ETAS model is one of the most widely used tools for modelling seismic activity in terms of both capturing specific features of interest and forecasting future events. Its estimation can be considered challenging due to identifiability issues. In this Chapter, we introduced the concept of temporally variable ground intensity based on Stress Release modelling. In it we specified two families of SR-ETAS model that depend on either the occurrence time of the previous event in the sequence (Full-SR-ETAS), or the elapsed time from the last uncaused (main) event (Branched-SR-ETAS). Our experimental results suggest that these models capture observed features of real earthquake catalogues that the standard ETAS model does not.

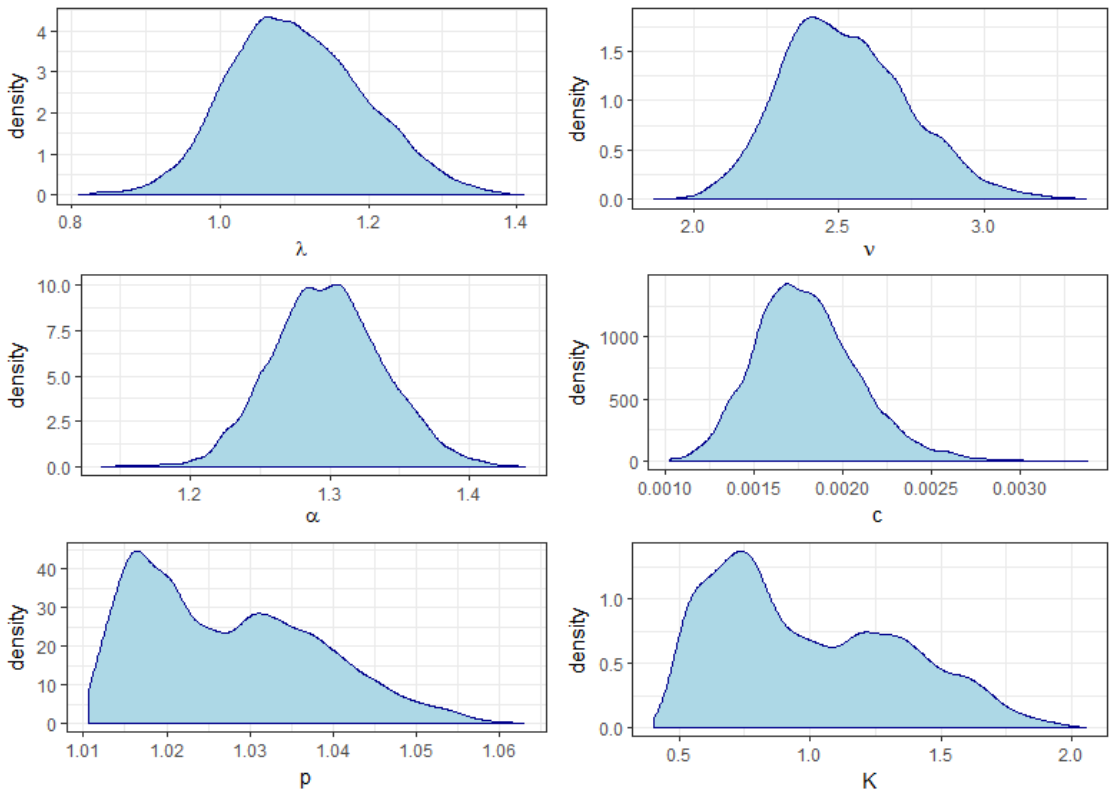

Figure 3.7: B-B-ETAS MCMC parameters' density for the North California catalogue.

Currently we examined a single fault, non-spatial occurrence that is typically used by general seismologist for the analysis of a seismic fault activity. All concepts are directly applicable to the spatial extension of ETAS that is introduced in Chapter 4. There are many alternatives of the spatial component(s) of the standard ETAS that provide a great differentiation amongst them which makes direct comparison of the introduced family of model impractical. Overall, the non-spatial alternative introduced in this Chapter will provide excellent results as long as there are no strong non-linear or non-uniform patterns in the spatial distribution of the earthquakes along the fault of interest.

All methods are introduced for a general distribution, as such the SR-ETAS family can grow very quickly to accommodate the modelling needs of any sort of data. Direct application to stock daily changes, insurance claims, fraud and terrorist threats is feasible.

Chapter 4

Semi-parametric Bayesian Forecasting of Spatial Earthquake Occurrences

In this Chapter we extend the temporal ETAS model for the incorporation of a spatial component. The spatial ETAS model relies on the estimation of spatial density of the main (uncaused) events in the catalogue. Typically the spatial background density is estimated based on a number of restrictive assumptions that either ignore the difference between caused and uncaused events or classifies them based on a holistic approach. Our work proposes a method for modelling the true uncaused events' spatial density as a non-parametric kernel which is based either on a Kernel density estimator or on a Dirichlet process with base measure the Normal Inverse Wishart distribution. This way we provide a flexible spatial measure that can be interpreted from a Bayesian perspective without inducing holistic bias.

4.1 Background

Considerable efforts are put into modelling efficiently natural hazards due to the increased occurrence of hazardous events that cause immense human and material losses. In this Chapter we will focus our study on earthquakes which are one of the most analysed natural catastrophe phenomena. Modelling earthquakes as a spatio-temporal stochastic point process is the key component of the quantification of seismic risk hazard [Brillinger, 1993, Ogata, 2011].

In Chapter 3 we applied and extended the latent variable MCMC algorithm that incorporates the information carried out by the processes' underlying branching structure to the context of SR-ETAS. In this Chapter we extend the latent variable MCMC algorithm for an application to spatial ETAS model. We will further propose new alternatives to the spatio-temporal ETAS model that incorporate non-parametric components.

The Dirichlet Process (DP) provides a framework for obtaining a discrete sample from a continuous distribution. It is discretising a distribution into fragments with variable bandwidth which on its own can be used as a non-parametric mixture model that we use as a component in our Spatial ETAS formulation. This way is obtained a very flexible, data driven, non-parametric model that provide estimation more roboust to overfitting compared to standard non-parametric methods.

We begin with a brief introduction of the spatio-temporal ETAS point process (Section 4.2). Then in Section 4.3 we review the introduction of three different uncaused events, non-parametric special density measures - Uniform, Kernel density estimation and Dirichlet process. After explicitly specifying the models that we will use, we address the so important concept of catalogue simulation and out-of-sample periods generation that are commonly used for prediction (Section 4.4). A justification for the usage of the latent variable MCMC within the context of Space-time ETAS is provided in Section 4.5, followed by a very detailed treatment of the exact method that we propose (Section 4.6). Further, we provide implementation consideration of the Latent variable method in Section 4.7. Then we present the relevant model diagnostic choices in Section 4.8. We study the application of Spatial ETAS Latent variable MCMC methods on simulated data (Section 4.9), as well as real earthquake data in Section 4.10, followed by Section 4.11 - a conclusion summarising all our findings.

4.2 Spatial ETAS model

In this Chapter is considered the improved extension of the space-time ETAS model as defined in [Ogata and Zhuang, 2006]. In this ETAS alternative, the probability of an earthquake occurring at time t depends on the previous seismicity $\mathscr{H}_t$, which is a collection of occurrences times $t_i < t$, marks m_i and locations (x_i, y_i). The earthquake's depth is usually ignored and instead only events within certain depth range are taken into account. Thus for a sequence of space-time observations in a region of interest Σ

the required information up to time t is the following:

$$\mathscr{H}_t = \{(t_i, m_i, x_i, y_i); t_i < t\}.$$

In addition to that, a 2×2 matrix $\mathscr{S}_i$ can be defined that provides the opportunity for anisotropic clusters by scaling the distance measure between observations. Although this is a parameter that has to be estimated from the raw data, it is a feature that is considered fixed prior to the estimation of all other parameters [Ogata and Zhuang, 2006]. As such it can be considered a data feature. The majority of the literature either ignores this term or challenges its effectiveness [Ogata and Zhuang, 2006, Ogata, 1998, Schoenberg, 2013, Lippiello et al., 2014, Fox et al., 2016].

One of the main issues related to the estimation of the space-time ETAS model is the opportunity to trigger an event within the area of interest, which was caused by an earthquake not present in Σ. Furthermore, the reverse process can occur when an event within the area of interest triggers earthquakes which are not part of Σ. In this Chapter, we address these problems in both simulation and estimation contexts.

The space component of the ETAS model should allow recovery of the basic ETAS model (e.g. Equation 2.11) after integrating over the spatial region of interest i.e.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \lambda(t, m, x, y)dxdy = \lambda(t, m).$$

4.2.1 Specific functional form

Let us consider the following general form of the spatio-temporal ETAS intensity function, $\lambda(t, m, x, y)$:

$$\lambda(t, m, x, y) = \mu(t, x, y) + \sum_{t_i<t} r(t - t_i) \times s(m_i - M_0, x - x_i, y - y_i, \mathscr{S}_i),$$

where $r(\cdot)$ is the Omori law:

$$r(z) = \frac{K}{(z + c)^p},$$

for which c and p are parameters controlling the decay rate, while K regulates the average productivity (i.e. the expected number of children of each event). Different forms of this function have been used for more than 100 years for various seismological studies across the world [Utsu et al., 1995, Guglielmi, 2017].

The ground intensity is given by the term $\mu(t, x, y)$ from which are triggered all immigrant events. It can either depend on time, space or both. Let us define $\mu(t, x, y) = \mu(t)\phi(x, y)$. There are many ways to shape the immigrant events spatial dependence, $\phi(x, y)$. The specific ones that are used in this Chapter are discussed in Section 4.3. All of them integrate to 1 over infinite spatial region. The temporal dependence of $\mu(t)$ is usually ignored [Ogata, 1988, Ogata, 1998, Ogata, 2004, Ogata and Zhuang, 2006, Ogata, 2011, Schoenberg, 2013, Holschneider et al., 2012, Fox et al., 2016], thus for simplicity we take $\mu(t, x, y) = \mu_0\phi(x, y)$. Additional information on the representation of $\mu(t)$ as a time-varying function is discussed in Chapter 3 and by [Chen and Stindl, 2018, Wheatley et al., 2016].

The aftershock events are triggered as an inhomogeneous Poisson process, with respective spatial intensity rate $s(m_i - M_0, x - x_i, y - y_i, \mathscr{S}_i)$ and temporal rate controlled by the Omori law.

The most widely considered general functional form of the spatial intensity [Ogata and Zhuang, 2006, Ogata, 1998, Schoenberg, 2013, Lippiello et al., 2014] is the following:

$$s(m_i - M_0, x - x_i, y - y_i, \mathscr{S}_i) = \frac{e^{\alpha(m_i - M_0)}}{\left[(x - x_i, y - y_i)\mathscr{S}_i(x - x_i, y - y_i)^t + d\right]^q}, \tag{4.1}$$

where α provides a similar functionality to those of K, and M_0 is the magnitude of completeness of the catalogue, which is determined empirically and corresponds to the minimum magnitude above which all earthquakes are successfully detected [Gutenberg and Richter, 1944, Wiemer and Wyss, 2000]. The magnitudes are then assumed to follow the usual Gutenberg-Richter law $m_i - M_0 \sim \text{Exp}(\beta)$ [Gutenberg and Richter, 1944, Fox et al., 2016]. The $\mathscr{S}_i$ is a positive definite matrix indicating the offspring inheritance. In Ogata's work [Ogata and Zhuang, 2006, Ogata, 2011] is suggested that the $\mathscr{S}_j$ matrix can be shared across all events that are in the same dynasty (see Section 2.2.1). Estimating this matrix requires the true underlying branching process without considering any model parameters. This implies that the branching structure should be obtained based on a holistic method that does not directly address the Hawkes process paradigm of self-excitation. Ogata further suggests to use the earthquakes with high magnitudes as immigrant shocks and then calculate $\mathscr{S}_j$ for all events in the dynasty of the j^{th} event, rather than for all events in the catalogue. Thus, the estimate of $\mathscr{S}_j$ is

typically approximated as:

$$\mathscr{S}_j = \begin{bmatrix} {\sigma_x}^2 & \rho\sigma_y\sigma_x \\ \rho\sigma_y\sigma_x & {\sigma_y}^2 \end{bmatrix},$$

where ρ is the correlation between all events' x and y coordinates in the dynasty of the j^{th} event, with their corresponding standard deviations $\sigma_{\cdot}$ in x and y.

Evidently, large earthquakes provide greater aftershock intensity but they are not necessarily immigrant events in a Hawkes process context. Assigning events as immigrant solely based on their features restricts the model branching structure, initiates bias and contradicts with direct parameters' interpretation. Further, the branching conditioning cannot be fully applied since there will already be a partial, holistic based branching structure which does not depend on the naturally occurring clustering process of a Hawkes process [Hawkes and Oakes, 1974]. This phenomenon directly contradicts the ETAS process' fundamental concepts of clustering and branching structures, hence it makes latent variable Bayesian analysis [Ross, 2018a] impossible.

For all these reasons, we will focus primarily on the following simplified alternative of the space-time ETAS model:

$$\lambda(t, m, x, y) = \mu_0\phi(x, y) + \sum_{t_i<t} \frac{K}{(t - t_i + c)^p} e^{\alpha(m_i - M_0)} \Big\{(x - x_i)^2 + (y - y_i)^2 + d\Big\}^{-q}, \quad (4.2)$$

with a parameter vector $\theta = (\mu_0, \alpha, c, p, K, d, q)$ and a background kernel $\phi(x, y)$.

We prefer to work with this version of spatial ETAS, as it incorporates a sound interpretation of the ETAS model's underlying assumptions related to event causality without negatively influencing the model's flexibility. The spatial ETAS model is reduced to the standard ETAS model if the x, y locations are set to 0, as well as $q = 2$ and $d = 1/\pi$ with respect to an infinite spatial domain (commonly referred to as infinite space assumption).

4.2.2 (Log-)Likelihood

The likelihood function of a space-time point process with intensity $\lambda(\cdot)$ for data $\mathscr{H}_t = \{(t_i, m_i, x_i, y_i); t_i < t\}$ is given by the following expression [Daley and Vere-Jones, 2003]:

$$\mathcal{L}(\theta; \mathscr{H}_T) = \prod_{i=1}^{n} \lambda(t_i, m_i, x_i, y_i | \mathscr{H}_T, \theta) e^{-\int\int\int \lambda(z, m_i, x_i, y_i | \mathscr{H}_T, \theta) dz dx dy}, \quad (4.3)$$

with respective log-likelihood form of

$$\ell(\theta; \mathscr{H}_T) = \sum_{i=1}^{n} \log(\lambda(t_i, m_i, x_i, y_i | \mathscr{H}_T, \theta)) \\ - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{0}^{T} \lambda(z, m, x, y | \mathscr{H}_T, \theta) dz dx dy. \quad (4.4)$$

The evaluation of the triple integral term is slow and numerically unstable. Thus a number of approximations are provided in the literature [Harte, 2012, Ogata, 1998, Schoenberg, 2013, Lippiello et al., 2014]. This calculation will be considerably neater if the spatial and temporal kernels are valid density functions. This can be achieved if the parameter K is represented as a product of normalisation parameters $K = \bar{K} K_t K_r$ where:

$$K_t \int_{0}^{\infty} (z + c)^{-p} dz = 1$$

for

$$K_t = (p - 1) c^{p-1}$$

and

$$K_r \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x^2 + y^2 + d)^{-q} dx dy = 1$$

for

$$K_r = \frac{q - 1}{\pi d^{1-q}}.$$

The above expressions assume infinite temporal and spatial domains (infinite time and space assumptions). Thus we can re-assign a new parametrisation to the temporal kernel $r(\cdot)$, commonly referred to as modified Omorri law [Vere-Jones and Davies, 1966]

$$g(z) = \frac{K_t}{(z + c)^p}.$$

Furthermore, the marked spatial kernel $s(\cdot)$ is split into a separate spatial kernel

$$h(x, y) = \frac{K_r}{(x^2 + y^2 + d)^q}$$

and an event marks kernel

$$\iota(m) = \bar{K} e^{\alpha(m - M_0)}.$$

This way we arrange the kernels' functionality to $r(\cdot)s(\cdot) = g(\cdot)h(\cdot)\iota(\cdot)$ to address better the underlying clustering functionality of the spatial ETAS model. The newly defined parameters $\bar{K}, K_r, K_t \in \theta$ since they represent a simple transformation of the standard parameter set. Hence, from here onward $\theta = \{\mu_0, \alpha, c, p, \bar{K}, d, q\}$ which is also sufficient information for the evaluation of K_r, K_t and K.

Based on the assumed shapes of $g(\cdot)$ and $h(\cdot)$ we establish that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} g(z) \times h(x,y) \times \iota(m) dz dx dy \approx \bar{K} e^{\alpha(m-M_0)} \tag{4.5}$$

under the infinite time and space assumptions. The approximation relies on the assumption that the space-time domain of interest is sufficiently large, which is not always true. In Chapter 3 we observed that infinite time assumption provides a poor performance regarding the North California seismic sequence, thus we prefer not to use it whenever possible. However, the infinite space assumption reduces considerably the computational cost. We did not experience any computational issues with it and we rely on its validity from here onward.

The log-likelihood of the Spatial ETAS model based on the finite time and infinite space assumptions can be approximated by the following expression:

$$\ell(\theta; \mathscr{H}_T) = \sum_{i=1}^{n} \log(\lambda(t_i, m_i, x_i, y_i | \mathscr{H}_T, \theta)) \\ - \mu_0 T + \bar{K} \sum_{i=1}^{n} e^{\alpha(m_i - M_0)} \left(1 - \frac{c^{p-1}}{(T - t_i + c)^{p-1}}\right), \tag{4.6}$$

with a parameter vector $\theta = (\mu_0, \alpha, c, p, \bar{K}, d, q)$ along with the specification of the background kernel $\phi(x, y)$.

4.3 Non-parametric Estimation of Background Intensity

In this Chapter we will consider a variety of methods for estimating the spatial background rate $\mu(x, y)$ in the ETAS model. Using the parametrisation above, we can write:

$$\mu(x, y) = \mu_0 \phi(x, y),$$

where μ_0 is a scaling constant and $\phi(x, y)$ is a probability density that integrates to 1. We will consider three different Bayesian models for the background rate. In all three,

the constant μ_0 will be estimated separately, with the models varying in terms of how they treat $\phi(x, y)$:

1. $\phi(x, y) \propto 1$, in which case the background intensity is constant over space. This is highly unrealistic due to the known fact that seismic activity is highly spatially depend, and we use this model only as a baseline. We refer to this model as Uniform ETAS.

2. A non-parametric model where $\phi(x, y)$ is learned using Kernel Density Estimation (KDE). Different versions of this method are fairly common in the seismology literature, albeit in a non-Bayesian context [Zhuang et al., 2002,Marsan and Lengline, 2008,Sornette and Utkin, 2009,Marsan and Lengliné, 2010,Fox et al., 2016]. However as we will show, the usual procedure suffers from a serious limitation where all the earthquakes in the catalogue are used in the estimation. This approach is technically incorrect, since $\phi(x, y)$ is specifically a model for the background events, rather than the triggered events. As such, KDE will result in a highly biased estimate of $\phi(x, y)$ since it treats background and immigrant events indistinguishably.

3. A non-parametric model where $\phi(x, y)$ is learned in a fully Bayesian manner, based on a Dirichlet Process, in a way which distinguishes between background and triggered events. This is substantially more complex than the KDE approach since it require declustering the earthquakes into background and immigrant events, with only the background events used to estimate $\phi(x, y)$. This corrects the bias in the KDE approach.

The first model using the uniform density is self-explanatory. We will now discuss the other two in more detail.

4.3.1 KDE ETAS

The second method described above uses kernel density estimation to learn $\phi(x, y)$, which is a classical method for non-parametric estimation of an unknown density function. Suppose we observe n observations (t_i, m_i, x_i, y_i) from the point process. Let $\mathbf{z}_i = (x_i, y_i)$ be the spatial coordinates of the i^{th} earthquake. Then a KDE estimate of $\phi(\cdot)$ can be given by:

$$\hat{\phi}(\mathbf{z}) = \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{z} - \mathbf{z}_i)$$

with

$$K_{\mathbf{H}}(\mathbf{x}) = |\mathbf{H}|^{-1/2} K(\mathbf{H}^{-1/2}\mathbf{x}),$$

where $\mathbf{H}$ is $d \times d$, symmetric and positive-definite which is also referred to as bandwidth matrix and $K(\cdot)$ is a symmetric kernel function. Without loss of generality we can choose it to be:

$$K(\mathbf{x}) = (2\pi)^{-1} \exp(-\frac{1}{2}\mathbf{x}'\mathbf{x}).$$

We refer to this model as KDE ETAS. Although Kernel Density Estimation is a powerful and flexible non-parametric method and it has two main drawbacks [Marsan and Lengline, 2008, Sornette and Utkin, 2009]. First, it can be difficult to choose $\mathbf{H}$ in a manner which produces an accepted level of smoothing across the whole spatial domain rather than under/over fitting in particular regions. Second, the resulting KDE estimate of $\phi(\cdot)$ is based on smoothing over all n earthquakes in the historical catalogue. However this is not the correct behaviour in the context of an ETAS model, since the $\phi(\cdot)$ function only specifies the occurrence of background, rather than triggered events. As such, we would expect the KDE estimate to be biased, and assign too much probability mass to spatial regions where large magnitude earthquakes have occurred, since their large number of triggered aftershocks will be incorporated into the estimate of $\phi(\cdot)$.

4.3.2 DP ETAS

The Dirichlet process (DP) was introduced by [Ferguson, 1973, Antoniak, 1974] as a probability distribution over probability distributions, and is commonly used as a prior in Bayesian non-parametric modelling. If a probability distribution G has a DP prior then we write:

$$G \sim DP(\chi, G_0),$$

where G_0 is the base distribution which defines the expected value of the DP and $\chi > 0$ is a measure of the variance. The DP is a conjugate prior in the following sense: suppose that $\varphi_1, \ldots, \varphi_n \sim G$ where $G \sim DP(\chi, G_0)$. Then the posterior distribution of G is:

$$G|\varphi_1, \cdots, \varphi_n \sim DP\left(\chi + n, \frac{\chi G_0 + \sum_{i=1}^{n} \delta_{\varphi_i}}{\chi + n}\right),$$

where $\delta_\cdot$ is a Dirac delta function [Dirac, 1947].

A constructive definition of the DP was given by [Sethuraman, 1994], who showed that samples from a DP can be written in stick breaking form:

$$G = \sum_{i=1}^{\infty} \pi_i \delta_{\psi_i}, \quad \psi_k \sim G_0,$$

where $\{\beta_i\}_{i=1}^{\infty} \sim Beta(1, \chi)$ $\pi_k = \beta_k \prod_{i=1}^{k-1}(1-\beta_i)$, and δ_{ψ_k} is the Dirac delta function [Dirac, 1947]. This provides a practical method for drawing a sample from a DP, by approximating the stick breaking as a finite sum:

$$G = \sum_{i=1}^{N} \pi_i \delta_{\psi_i}, \quad \psi_k \sim G_0.$$

Combining this with the conjugacy result above, we can hence sample G from its posterior distribution given some observed data as:

$$G|\varphi_1, \cdots, \varphi_n = \sum_{i=1}^{N} \pi_i \delta_{\psi_i}, \quad \psi_k \sim \frac{\chi G_0 + \sum_{i=1}^{n} \delta_{\varphi_i}}{\chi + n},$$

where $\{\beta_i\}_{i=1}^{N} \sim Beta(1, \chi + n)$ and $\pi_k = \beta_k \prod_{i=1}^{k-1}(1-\beta_i)$.

An alternative representation of the DP is based on the Chinese restaurant process [Neal, 2000], which shows that the marginal prior distribution of the samples $\varphi_1, \ldots, \varphi_n$ (with G integrated out) can be written as:

$$\varphi_i|\varphi_1, ..., \varphi_{i-1} \sim \frac{1}{i-1+\chi} \sum_{j=1}^{i-1} \delta_{\varphi_i} + \frac{\chi}{i-1+\chi} G_0,$$

where $\varphi_i \sim G_0$.

The Dirichlet Process as a Spatial ETAS Prior In this Chapter, we propose to use the Dirichlet Process (DP) as a non-parametric prior for the background ETAS intensity $\phi(x, y)$. From the above results, we can see that samples from a DP follow a discrete distribution. In order to adapt the DP to continuous data, it is common to instead use it as a prior distribution for a mixture model. This leads to the following model:

$$\phi(x, y) = \int k(x, y|\varphi) dG(\varphi)$$

$$G \sim DP(\chi, G_0),$$

where $k(\cdot)$ is a mixture kernel. This formulation corresponds to an infinite dimensional mixture model where the DP is used as a prior on the mixing distribution parameter.

Since $\phi(x, y)$ is a two-dimensional spatial distribution, we will model it as a mixture of bivariate Gaussians, where $\varphi = (\boldsymbol{\mu}, S)$ is the mean vector and precision matrix. For conjugacy, we choose G_0 to be the Normal Inverse-Wishart distribution. This leads to the following model:

$$x_i, y_i | c_i \sim N(\boldsymbol{\mu_i}, (S_i)^{-1})$$

$$\boldsymbol{\mu_i}, S_i \sim G$$

$$G \sim DP\Big(\chi, NW(\boldsymbol{\xi}, \rho, \beta, \beta V)\Big),$$

where $\boldsymbol{\xi}, \rho, \beta$ and V are the parameters of the Normal Inverse-Wishart distribution where the mean $\boldsymbol{\mu_k}$ follow a imp

From here after, this ETAS alternative will be referred to as DP ETAS model.

4.4 Catalogue Simulation

In this section we discuss how to approach the problem of simulating a catalogue based on a specific parameter set θ and a spatial immigration distribution $\phi(x, y)$. We also address the concept of extending a given catalogue, assuming that θ and $\phi(x, y)$ are known. This methodology is essential to obtain predictions based on any given dataset. Later in Section 4.9 we develop a study based on simulated data. Based on the obtained MCMC chains we can then generate catalogues based on each of them to propagate key hazard metrics of interest by extending the catalogues that were used for estimation.

4.4.1 Simulation

We base our simulation method on the spatial ETAS process' underlying inheritance structure (illustrated on Figure 2.2) as previously outlined in Section 2.3 . Firstly, all immigrant events are initiated. Then every event in the sequence is allowed to generate events from multiple generations based on its offspring intensity $g(\cdot) \times h(\cdot) \times \iota(\cdot)$. The obtained catalogue not only represents the fundamental ETAS clustering process but also recovers the usually unknown true branching structure.

1. Input

   (a) The parameter set θ.

(b) Immigrant events' density $\phi(x, y)$.

(c) Spatial detection region for immigrant events Σ.

(d) Temporal detection interval $\tau = [0, T]$.

2. Simulate values from the target distributions of interest. Using rejection sampling, we create sufficiently large samples from the following distributions:

   (a) $K_t(z + c)^{-p}$ - temporal lags with respect to any target point in time. A reasonable approach is to use as a proposal distribution Exponential with expected value that matches the expected causality time for an event in the catalogue. However, this is only applicable for parameters that guarantee finite expectation.

   (b) $K_r(x^2+y^2+d)^{-q}$ - spatial lags with respect to a given point in space. Rather than restricting all offspring events to be only in the spatial region Σ, we allow them to generate events outside of the area as long as the obtained time-lags come from the required target distribution. The proposal distribution was set to Pareto with parameters $(2, 2)$ to address the heavy tailed behaviour of the spatial kernel.

   In the simulation process we propose independently values for the spatial lags in both x and y directions. For every sampled pair (x_i, y_i) we can sample uniformly r points of the circumference of a circle with centre $(0, 0)$ and radius $\sqrt{x_i^2 + y_i^2}$. Thus, for every successfully sampled realisation from our target distribution we can generate additional r samples from it. In order for this method to work, the overall number of true samples generated from the initial rejection sampling should be a lot larger than r. If the opposite occurs i.e. r is relatively large while the overall number of samples is just a few times larger than r, then the density will be concentrated over the circumference of multiple circles with common centre $(0, 0)$. We set $r = 4$.

   (c) The productivity of every event (Section 2.3.2) associated with its magnitude m_i follows $\iota(m_i) = \bar{K}e^{\alpha(m_i - M_0)}$, where the marks are greater than a certain threshold M_0. However, the marks in an ETAS context are assumed to be a realisation from the Gutenberg-Richter law [Gutenberg and Richter, 1944,Fox et al., 2016]. This implies that for a productivity parameter β we obtain the required marks $(m_.)$ as a realisation from $m_i - M_0 \sim \text{Exp}(\beta)$.

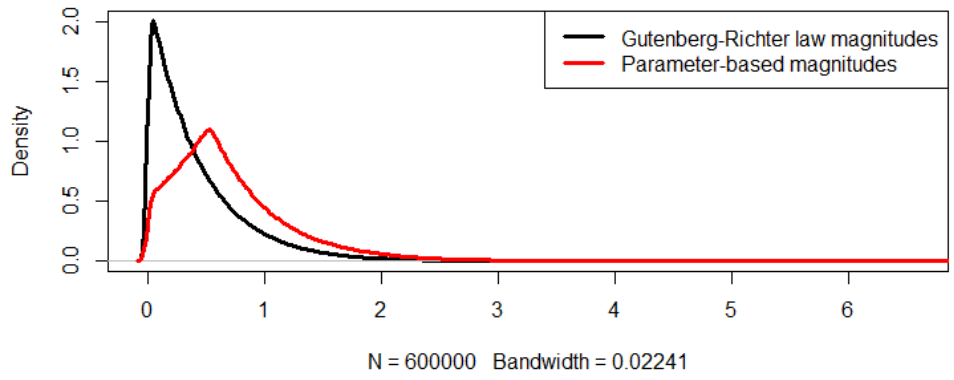

Figure 4.1: Comparison between magnitudes density obtained from simulation from Gutenberg-Richter law and true model parameters

In a simulation context, [Fox et al., 2016] generates the raw magnitudes $(m_i - M_0)$ from Gutenberg-Richter law with $b-$value of 1 [Gutenberg and Richter, 1944], which is equivalent to Exponential distribution $P(m-M_0|\beta) = \beta e^{-\beta(m-M_0)}$ for $\beta = ln(10)$. If instead the magnitudes were simulated from $\iota(\cdot)$ with parameter set $\{\alpha = 1.407, \bar{K} = 0.322, M_0 = 0\}$ using rejection sampling, the overall productivity is increased from 0.8 to 1.5. The difference in the raw magnitudes' density for 6×10^6 samples from each of the two methods are shown on Figure 4.1. Evidently, the two densities are different. To be more precise, not using the Gutenberg-Richter law makes this specific parameter estimation useless due to a infinite expected catalogue length.

3. Creating a set of uncaused events

   (a) Sample immigrant event time lags with a cumulative sum not exceeding the maximum of the temporal detection interval τ. In this Chapter , we focus on the case in which all immigrant events are a realisation of a Poisson process with constant rate μ_0. Hence, the inter-arrival times of all immigrant events are a realisation from exponential distribution with mean μ_0. The maximum time in the realisation should not exceed T, the maximum of the detection interval τ.

   (b) Allocate at random a magnitude for each of the simulated temporal realisations based on the already sampled magnitudes.

(c) Sample the required number of target locations from $\phi(x, y)$ that lie in the target area of interest Σ.

This sample is the so-called 0^{th} 'inheritance' level i.e. all events in it are uncaused. Each of them can initiate events that can further create descendants on their own. We describe this phenomenon with respect to the specific inheritance level which corresponds to the number of predecessors the event has. For example, on Figure 2.2 events t_1, t_6 and t_{10} are the 0^{th} inheritance level; events t_2, t_3, t_5, t_7 and t_9 are the 1^{st} inheritance level; events t_4 and t_8 are the 2^{nd} inheritance level; while t_{11} is the only event in inheritance level 3.

4. Generating inheritance structure of multiple generations.

The following iterative procedure is repeated until there exists a non-empty population at the i^{th} level for $i \in \{0; \mathbb{N}\}$.

(a) While N_i, the length of the i^{th} sub-population, is positive

(b) Evaluate the offspring intensity for each event $\{t_j, m_j, x_j, y_j\}$ for $j = 1, .., N_i$:

$$\bar{\lambda_j} = \int_0^{\infty} \int_0^{\infty} \int_0^{T-t_j} g(z) \times h(x, y) \times \iota(m) dz dx dy.$$

(c) Generate the number of offsprings for each of the events in the i^{th} sub-population as a realisation from $\varrho_j \sim \text{Poi}(\bar{\lambda_j})$ for $j = 1, .., N_i$.

(d) Create the sub-population of level $i + 1^{st}$ by assigning to each of the events of sub-population i, number of offsprings equal to ϱ_j. Hence, we have to sample $\sum_{j=1}^{N_i} \varrho_j$ magnitudes and lags in space and time. Then we assign them to the corresponding parent events to obtain the exact events from generation $i + 1^{st}$.

(e) The $i+1^{st}$ sub-population is only restricted in terms of the temporal detection range τ - we record events with arrival time not greater than T and allow the offsprings' spatial locations to be outside of the immigrant detection range Σ.

4.4.2 Extending a catalogue

We note that since our models are fully specified, we can use them to create forecasts about future earthquakes, e.g. the probability of an earthquake with magnitude greater

than M_0 occurring during some future time period $\mathscr{H}_{new}$ (e.g. the next year). This can be done by simulation, based on the samples drawn from the posterior using MCMC to produce simulated point process trajectories over the time period of interest, and then extracting the quantities to be forecasted as summary statistics. This is essentially a Monte Carlo approximation to the forecasting distribution:

$$P(\mathscr{H}_{new}|\mathscr{H}_t) = \int P(\mathscr{H}_{new}|\theta)P(\theta|\mathscr{H}_t)d\theta \approx \frac{1}{M}\sum_{i=1}^{M} P(\mathscr{H}_{new}|\theta^{(i)})$$

$$\text{for} \quad \theta^{(i)} \sim P(\theta|\mathscr{H}_t).$$

While on real catalogues we split the data in train and test sets, on simulated data we can sample multiple point process trajectories based on the true parameter set. All obtained catalogue extensions can be used to compare out-of-sample model performance. We can obtain such sample based on a branching-based sampling mechanism.

Creating an extension can be done by making minor amendments to the simulation algorithm that was introduced in Section 4.4.1. The temporal detection range has to change from $\tau = [0, T]$ to $\bar{\tau} = (T, T + M]$, where M is the temporal extension length. The only major difference with respect to the simulation is in step 3. It is associated with the fact that the events that were already observed (τ) can excite event occurrence in the current detection interval ($\bar{\tau}$). We obtain a realisation for these events by executing once steps 4 (b)-(e) where the offspring detection range is only in the interval $\bar{\tau}$ for all events in τ. We execute steps 4 (b)-(e) only once because further triggered events will be generated later. We merge the obtained one level offspringing events with the sampled uncaused events for the extension interval $\bar{\tau}$ and assign them to generation 0 within the extension interval. Then step 4 from the simulation algorithm is performed.

4.5 Posterior Simulation

As previously discussed, although the direct Metropolis-Hastings applications for obtaining draws from the parameters' posterior distribution is tempting at first, it is extremely ineffective to be applied in real-world examples. Recall from Equation 4.6, that

the log-likelihood of the ETAS model is:

$$\ell(\theta;\mathscr{H}_T) \approx -\mu_0 T + \bar{K}\sum_{i=1}^{n} e^{\alpha(m_i - M_0)}\left(1 - \frac{c^{p-1}}{(T-t_i+c)^{p-1}}\right) + \sum_{j=1}^{n} log\Big(\mu(x_j,y_j) + \sum_{t_i<t_j} \iota(m_i - M_0)\times r(t_j - t_i)\times s(x_j - x_i, y_j - y_i)\Big). \tag{4.7}$$

The first problem is that evaluating the likelihood function requires a double summation and this evaluation must take place each time a new parameter value is proposed. This makes direct MCMC computationally very demanding, and cannot feasibly be run on a catalogue containing more than a few hundred earthquakes. The second problem is that [Schoenberg, 2013] studied the performance of frequentist maximum likelihood estimation for the ETAS model based on directly maximising the above likelihood function when $\mu(x,y) = \mu_0$ was a constant value found that the resulting parameter estimates often differed substantially from their true values. This is because the likelihood function is very complex and the components of the parameter vector are highly correlated. Since MCMC methods can also suffer from serious convergence issues when the parameters are correlated, it is reasonable to believe that this direct MCMC procedure will suffer from the same problem. Since this problem is already present in the simple parametric case with constant μ_0, it will be even worse in the more complex non-parametric setting.

We instead propose a reparametrisation of the model based on latent variables that aims to break the parameter correlation and lead to an efficient Metropolis-Hastings algorithm for posterior sampling. This is an extension of the method proposed by [Ross, 2018a] for the temporal Hawkes process.

4.6 Latent Variable Formulation

We now develop an alternative sampling posterior scheme based on introducing latent variables. These have the effect of breaking the dependence between the parameters in the likelihood function. We will show that conditional on the latent variables, the parameter sets $\{\mu_0, \phi(x,y)\}$, $\{\bar{K}, \alpha, p, c\}$, and $\{d, q\}$ are all conditionally independent of each other, which improves considerably the convergence of MCMC sampling.

As discussed in Section 2.2 the ETAS model can be reinterpreted as a branching process in the following sense. Suppose that the i^{th} earthquake occurs at time t_i, so that $i-1$ earthquakes have occurred previously. Equation 4.2 can be interpreted as

showing that the ETAS intensity function at time t_i is a sum of i different Poisson processes. The first is a homogenous Poisson process with intensity $\mu(x, y)$, while the other $i-1$ each correspond to one of the previous earthquakes. Specifically, for each $1 \leq j \leq i-1$, the earthquake at time t_j triggers an inhomogeneous Poisson process with intensity:

$$\lambda_p(t, x, y) = Ke^{\alpha(m_j - M_0)}(t - t_j + c)^{-p}\Big\{(x - x_j)^2 + (y - y_j)^2 + d\Big\}^{-q}. \tag{4.8}$$

Based on standard results about the superposition of Poisson processes [Daley and Vere-Jones, 2003] we can interpret event t_i as having been generated by a single one of these i processes. We hence introduce the latent branching variables $B = \{B_1, \ldots, B_n\}$ where $B_i \in \{0, 1, \ldots, i-1\}$ indexes the process which generated t_i:

$$B_i \sim \begin{cases} 0 & \text{if } t_i \text{ was produced by the background process} \\ j & \text{if } t_i \text{ was triggered by the previous earthquake at time } t_j \end{cases}$$

Conditional on knowing B, we can partition the earthquakes into $n+1$ sets $S_0, \ldots, S_n$ where:

$$S_j = \{t_i; B_i = j\}, \quad 0 \leq j < n,$$

so that S_0 is the set of immigrant events which were not triggered by previous earthquakes, and S_j is the set of direct aftershocks triggered by the earthquake at time t_j. It is clear that these sets are mutually exclusive and that their union contains all the earthquakes in the catalogue. Additionally, we can see that the earthquakes in set S_0 are generated by an inhomogenous Poisson process with intensity $\mu(x, y)$, while the events in each set S_j for $j > 0$ are generated by a single inhomogenous Poisson process with intensity given by Equation 4.8. The ETAS likelihood function from Equation 4.6 can hence be rewritten (conditional on knowing the latent branching variables) as:

$$\begin{aligned} \mathcal{L}(\theta; \mathscr{H}_T, B) =& e^{-\mu_0 T} \prod_{t_i \in S_0} \mu(x_i, y_i) \\ & \prod_{j=1}^{n} \left(e^{-\bar{K}e^{\alpha(m_j - M_0)}\left(1 - \frac{c^{p-1}}{(t_n - t_i + c)^{p-1}}\right)} \{\bar{K}e^{\alpha(m_j - M_0)}\}^{|S_j|} \right) \\ & \prod_{j=1}^{n} \prod_{t_i \in S_j} \left(\frac{K_t K_r}{(t_i - t_j + c)^p} ((x_i - x_j)^2 + (y_i - y_j)^2 + d)^{-q} \right), \end{aligned} \tag{4.9}$$

where as before $\mu(x,y) = \mu_0\phi(x,y)$ where μ_0 is a constant and $\phi(x,y)$ has a KDE or DP mixture prior. From this factorisation, we can see that μ_0 and $\phi(x,y)$ are independent of the other model parameters in the likelihood, and hence will be independent in the posterior assuming prior independence. This latent variable formulation hence breaks the dependency which would have made the KDE or DP part of the model difficult to learn, and further weakens the dependence between $\{c,p\}$, $\{\alpha,\bar{K}\}$ and $\{d,q\}$ which will allow for more efficient MCMC sampling.

We hence propose a Gibbs sampler which samples the parameters in the following conditionally independent blocks:

- $P(B|\mu_0,\phi(x,y),\bar{K},\alpha,c,p,d,q,\mathscr{H}_t)$
- $P(\phi(x,y)|B,\mathscr{H}_t)$
- $P(\mu_0|B,\phi(x,y),\mathscr{H}_t)$
- $P(\bar{K},\alpha|B,c,p,\mathscr{H}_t)$
- $P(c,p|B,\bar{K},\alpha,\mathscr{H}_t)$
- $P(d,q|B,\mathscr{H}_t)$

Given a set of model parameters $\theta^{(k)}$ at iteration k of the Gibbs sampler, we now explain how to sample the next value $\theta^{(k+1)}$ from the above full conditional distributions.

4.6.1 Sampling B

As shown by [Zhuang et al., 2002], in the context of stochastic declustering, each individual branching variable $B_i^{(k+1)}$ can be sampled exactly from its conditional posterior. Note that each B_i can take values only in the discrete set $\{0,1,\ldots,j-1\}$, i.e. each earthquake can only be triggered by either a previous earthquake, or the background process. Assuming a uniform prior on each B_i, the probability of it being caused by any of the i processes is simply the proportion of the overall intensity that can be attributed to that process, i.e.:

$$P(B_i^{(k+1)} = j|\mathscr{H}_t,\theta^{(k)}) = \begin{cases} \dfrac{\mu_0\phi(x_i,y_i)}{\lambda(t_i,m_i,x_i,y_i|\mathscr{H}_{t_i},\theta^{(k)})} & \text{for } j=0 \\ \dfrac{g(t_i-t_j)h(x_i-x_j,y_i-y_j)\iota(m_i)}{\lambda(t_i,m_i,x_i,y_i|\mathscr{H}_{t_i},\theta^{(k)})} & \text{for } j\neq 0 \end{cases} \tag{4.10}$$

Each B_i can hence be drawn independently with weights given by Equation 4.10. A proof of this is shown in Section 3.5.2.

4.6.2 Update $\phi(x, y)$

This step is required only for DP ETAS since $\phi(\cdot)$ is constant in the KDE version of the model. Recall from Section 4.3.2 that

$$\phi(x, y) = \int N(x, y|\varphi)d\varphi$$

$$\varphi \sim G$$

$$G \sim DP(\chi, G_0),$$

where $\phi(x, y)$ is the generating function for the $|S_0|$ immigrant earthquakes that are assigned to the background process based on the current branching structure.

In order to simulate a value of $\phi(\cdot)$ from its conditional posterior, we first simulate values of $\varphi_i, i \in \{1, 2, \ldots, |S_0|\}$ from their posterior distributions given the earthquakes which are assigned to the background process, using the usual Chinese Restaurant process sampler. Given these values, we then have from Section 4.3.2 that:

$$G|\varphi_1, \ldots, \varphi_{|S_0|} \sim DP\left(\chi + n, \frac{\chi G_0 + \sum_{i=1}^{n} \delta_{\varphi_i}}{\chi + n}\right).$$

We can hence sample a value of G from its posterior using truncated stick breaking, i.e:

$$G = \sum_{i=1}^{N} \pi_i \delta_{\psi_i}.$$

This hence fully defines a realisation of $\phi(x, y)$ from its posterior. Note that the reason why we need to simulate a realisation of $\phi(x, y)$ (and hence G) rather than working only with the φ_i samples is that we need to have a realisation of $\phi(x, y)$ to evaluate the branching posterior in Equation 4.10.

As part of this step, we can also perform an update of the hyperparameters of G_0 and of χ, if we wish to work with the full hierarchical version of the DP. This can be done by assigning them a sensible prior distribution, such as $\chi \sim \text{Gamma}(1, 1)$. The updates in this case are standard in the DP literature and are described in detail in [Görür and Rasmussen, 2010].

4.6.3 Update the value of μ_0

Using Equation 4.9 we can observe that μ_0 only depends on the number of events in the background process S_0, hence:

$$P(\mu_0|\mathscr{H}_t,\theta,B) \propto \pi(\mu_0)e^{-\mu_0 T}\prod_{t_i\in S_0}\mu(x_i,y_i) = \\ \pi(\mu_0)e^{-\mu_0 T}\mu_0^{|S_0|}\prod_{t_i\in S_0}\phi(x_i,y_i) \propto \pi(\mu_0)e^{-\mu_0 T}\mu_0^{|S_0|}.$$

This is equivalent to estimating the intensity μ_0 of a homogeneous Poisson process on $[0,T]$, with event times S_0. In this case, the Gamma distribution is the conjugate prior: $\pi_{\mu_0} = Ga(\alpha_{\mu_0},\beta_{\mu_0})$. The posterior distribution is then $p(\mu_0|\mathscr{H}_t,\theta,B) = Ga(\alpha_{\mu_0}+|S_0|,\beta_{\mu_0}+T)$ which can be sampled from directly [Ross, 2018a].

Update the values of $\bar{K}$ and α

Similar to the process for sampling μ_0, we can sample new values of $\bar{K}$ and α from $p(\bar{K},\alpha|\mathscr{H}_t,\theta,B)$. Based on Equation 4.9, we conclude that:

$$P(\alpha,\bar{K}|\mathscr{H}_t,\theta,B) \propto \pi(\bar{K},\alpha) \\ \prod_{j=1}^{n}\left(e^{-\bar{K}e^{\alpha(m_j-M_0)}\left(1-\frac{c^{p-1}}{(t_n-t_i+c)^{p-1}}\right)}\{\bar{K}e^{\alpha(m_j-M_0)}\}^{|S_j|}\right).$$

Although there is no conjugate prior in this case, it is straightforward to use (e.g.) random walk MCMC to draw a sample from this posterior as described in the beginning of this Section.

4.6.4 Update the values of c and p

Again, based on Equation 4.9, we can see that the posterior distribution of c and p is given by:

$$P(c,p|\mathscr{H}_t,\theta,B) \propto \pi(c,p) \\ \prod_{j=1}^{n}\left(e^{-\bar{K}e^{\alpha(m_j-M_0)}\left(1-\frac{c^{p-1}}{(t_n-t_i+c)^{p-1}}\right)}\prod_{t_i\in S_j}\frac{K_t}{(t_i-t_j+c)^p}\right).$$

The parameter sampling can be done using (e.g.) standard random walk MCMC sampler.

4.6.5 Update the values of d and q

As a last step of our MCMC sampler we update the offspring kernel space parameters d and q. The expression below is a simplified approximation that depends on an infinite space approximation which was discussed in Section 4.2.2

$$P(d,q|\mathscr{H}_t,\theta,B) \propto \pi(d,q)\prod_{j=1}^{n}\Big(\prod_{t_i\in S_j} K_r\Big((x_i-x_j)^2+(y_i-y_j)^2+d\Big)^{-q}\Big).$$

4.7 Prior Choice, and Implementation Details

The most common problem with the estimation and simulation of an ETAS model is that certain parameter values can result in infinitely many earthquakes being generated from the process with non-zero probability. As discussed in 2.2.2, with a more detailed treatment in [Helmstetter and Sornette, 2002], that to guarantee a finite catalogue, the average number of aftershocks produced by each earthquake in the catalogue must be less than 1. The intuition for this result is that if the average number of aftershocks is greater than 1, then the cluster process representation of the process may never converge.

The offspring intensity is only dependent on the magnitudes and two model parameters $\{\alpha,\bar{K}\}$ based on the infinite space and time assumption. Then, we can evaluate the productivity (Section 2.3.2) of each event as follows:

$$\int_{M_0}^{M_{max}} \bar{K}e^{\alpha(m-M_o)}P(m-M_0)dm =$$

$$\int_{M_0}^{M_{max}} \beta\bar{K}e^{(\alpha-\beta)(m-M_o)}dm = \begin{cases} \Big[\frac{\beta\bar{K}e^{(\alpha-\beta)(m-M_o)}}{\alpha-\beta}\Big]_{m=M_0}^{M_{max}} & \alpha\neq\beta \\ \Big[\alpha\bar{K}(m-M_0)\Big]_{m=M_0}^{M_{max}} & \alpha=\beta \end{cases} \tag{4.11}$$

The simulation of marks is assumed to be independent from the ETAS parametrisation, hence the case in which $\alpha=\beta$ is applicable only in an estimation context, where β is the parameter in the Gutenberg-Richter distribution $P(m-M_0|\beta)$ (Section 4.4.1). Choosing the magnitudes' upper bound to go to infinity corresponds to the following reduced form of the productivity dependence:

$$\frac{\bar{K}\beta}{\beta-\alpha}.$$

Then a finite catalogue will be guaranteed, based on the infinite time-space assumption, if the following two statements hold true

$$\beta < \alpha$$

and

$$\frac{\bar{K}\beta}{\beta - \alpha} < 1.$$

However, restriction the constrains based on infinite time, space and mark can be overly restrictive depending on the specific parameter scenario with the prime influence associated with the later one. For known magnitudes, we can use a deterministic approach for the offspring productivity evaluation based on the mean value of the offspring productivity:

$$\frac{\bar{K}}{N_m} \sum_{i=1}^{N_m} e^{\alpha(m_i - M_0)}(1 - c^{p-1}(c+T)^{1-p}), \tag{4.12}$$

where N_m is the number of magnitudes that we take into account. For simulation purposes we can evaluate the above expression directly from the generated values in step 2 of the simulation algorithm introduced in Section 4.4.1. In an estimation context the catalogue's magnitudes shall be used. We base finite catalogue restrictions based on Equation 4.12

As such, we will choose our prior distributions so that positive mass is only assigned to regions where the above parameter relations are satisfied. We choose to use relatively uninformative Uniform priors: $\alpha \in (0, 10)$, $c \in (0, 10)$, $p \in (1, 30)$, $\bar{K} \in (0, 30)$, $d \in (0, \infty)$ and $q \in (1, \infty)$, with the regions not satisfying the above relations assigned zero mass. Note that the priors for c and p are slightly informative which is required since these parameters are only weakly identifiable [Holschneider et al., 2012]

Finally in the above discussion of the MCMC sampler, we mentioned that random walk Metropolis-Hastings was used to update some blocks of parameters. For these, we used a Normal proposal distribution with standard deviation of 0.1 and mean the most recently obtained parameter value.

4.8 Model comparison

We have proposed three different versions of the Bayesian ETAS model, which respectively use the Uniform distribution, KDE and a DP mixture model to represent $\phi(x, y)$.

Across all discussed methods for model performance and evaluation in Section 2.4.4 we can only effectively apply in- and out-of-sample log-likelihood and DIC. In this section we address these techniques in the context of Spatio-temporal ETAS model.

4.8.1 Deviance Information Criterion (DIC)

Recall that for a given a set of model parameters θ the model's DIC value is:

$$\text{DIC}(\theta) = -2\ell(\bar{\theta}; \mathscr{H}_t) + 2p_{DIC},$$

where $\ell(\theta; \mathscr{H}_T)$ is the log-likelihood function and p_{DIC} is the effective sample size, which evaluates the number of independent samples the MCMC draws are equivalent to. It is defined as:

$$p_{DIC} = 2\ell(\bar{\theta}; \mathscr{H}_t) - 2\,\mathbb{E}(\ell(\theta; \mathscr{H}_T) \cong 2\ell(\bar{\theta}; \mathscr{H}_t) - 2\frac{1}{S}\sum_{s=1}^{S}\ell(\theta_s; \mathscr{H}_t),$$

where θ_s indicates the s^{th} parameters' sample in the considered MCMC chain. Alternatively, we can compute the effective sample size as the variance of the obtained log-likelihood values for all sampled parameters as follows:

$$p_{DICalt} = 2Var[\ell(\theta; \mathscr{H}_T)].$$

This method is not as numerically stable as the other one but it is easier to compute because it does not require the allocation of $\bar{\theta}$, which is a computationally very demanding task with respect to the $\phi(x, y)$ of DP ETAS model, hence we will use this alternative (p_{DICalt}) of the DIC metric through this Chapter.

4.8.2 Out-of-sample log-likelihood

A common way to evaluate the performance of earthquake models is to consider the out-of-sample predictive distributions, i.e. how well we can predict the occurrence time and locations of earthquakes in the time window $[T, U]$ given that we have fitted a model to the time window $[0, T]$. Several versions of this approach have been used in the literature, with a summary given in [Bray and Schoenberg, 2013]

Since all of our models are fully Bayesian with completely specified probability distributions, we will compare based on the out-of-sample posterior predictive likelihood

(i.e. the likelihood of the future observations averaged over the samples which have been drawn from the posterior).

4.9 Simulation Study

In this section we will use synthetic (simulated) data to evaluate and compare the performance of the three Bayesian ETAS models for $\phi(x, y)$ using 1) the Uniform distribution, 2) KDE, and 3) a Dirichlet process mixture. The next section will similarly compare them using real earthquake catalogues.

4.9.1 Initial Comparison

We simulate three catalogues, each with a different choice for the density $\phi(x, y)$. The first density that we consider follows a standard bivariate normal distribution i.e.

$$\phi_1(x, y) \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right). \tag{4.13}$$

The second one is a mixture of two Normal distributions. The first of them has a mean of $(-1, -1)$ and the second one $(1, 1)$. They share a common covariance matrix that comprises zero covariance and 0.4 standard deviation in each dimension i.e.

$$\phi_2(x, y) \sim N\left(\begin{bmatrix} -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0.4 & 0 \\ 0 & 0.4 \end{bmatrix} \right) + N\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0.4 & 0 \\ 0 & 0.4 \end{bmatrix} \right). \tag{4.14}$$

The third density aims to simulate a seismic fault - all events are uniformly distributed on a line with known boundary conditions. This requires the specification of a fixed spatial region Σ. We sample uniformly a realisation of the x range of Σ. Then we transform it into a point on a line defined by an intercept a and a slope b and further scale it by an error component $\epsilon \sim N(0, \sigma_\epsilon^2)$ i.e.

$$\phi_3(x, y) = \phi_x(x)\phi_y(y), \tag{4.15}$$

for $\phi_x(x) \sim Unif(\Sigma_x)$, where Σ_x is the range in x dimension and $\phi_y(y) \sim a + bx + \epsilon$. We chose $a = 1$, $b = 2$, $\sigma_\epsilon = 0.5$ and $\Sigma_x = (-2, 2)$.

For the remainder of the ETAS parameters, we choose parameters based on the Tohoku District, Japan catalogue from $1926 - 1995$ over 36^o, 42^o N and 141^o, 145^o,

E, which were estimated using maximum likelihood by [Ogata, 1998]. These are: $(\bar{K}, \alpha, p, c, d, q) = (0.322, 1.407, 1.121, 0.0353, 0.0159, 1.531)$ with ground intensity constant $\mu_0 = 0.854 \times 10^{-4}$ and margin of completion $M_0 = 5$.

The same parametrisation is used in [Fox et al., 2016] subject to the following amendments - $M_0 = 0$, $\Sigma = [0, 4] \times [0, 6]$, and $\mu_0\phi(x_i, y_i) = 0.001 + 0.004 \times \mathbb{1}_{(x;y)}\{([0, 2]; [3, 6]) \cup ([2, 4]; [0, 3])\}$ where $\mathbb{1}_{(\cdot)}\{\cdot\}$ is an indicator function. These simulations spread over temporal interval $[0, 25000]$.

For our simulation, we choose to set $\mu_0 = 0.325$ to provide denser catalogues within a shorter period of time. The overall event rate has increased by 0.35×10^4 which allows us to run simulations for a shorter period of time compared to the previously introduced examples. However, this is not going to affect negatively the performance of the remaining parameters.

All simulated catalogues in this section have magnitudes following the Gutenberg-Richter law [Gutenberg and Richter, 1944] with b-value of 1 i.e. $m_i - M_0 \sim \text{Exp}\big[\beta = ln(10)\big]$. Hence, all marks m are greater than the specific margin of completeness used for the simulation, M_0. Within the simulation study, we set temporal window in $t \in (0, 300)$ with extension interval $\tau \in [300, 350)$ and magnitude of completeness of $M_0 = 2$.

4.9.2 Model Fitting and Results

For each of the three datasets, we used MCMC to draw $12,000$ samples from the posterior (after thinning). The branching structure was sampled from its conditional posterior only at every 50 iterations of the latent variable MCMC algorithm, since this is an $O(n^2)$ operation and slower than the other updates. For the DP ETAS model, we also resample the immigrant events density function $\phi(x, y)$ when a new branching structure is sampled.

For the KDE ETAS model, the estimate of the immigrant spatial density $\phi(x, y)$ is based on the whole catalogue of observations (and is hence biased), and so is estimated prior to running the MCMC and set to a fixed value, as in [Zhuang et al., 2002, Marsan and Lengline, 2008, Sornette and Utkin, 2009, Marsan and Lengliné, 2010, Fox et al., 2016].

In the simulation example we developed an out-of sample comparison with respect to 30 out-of-sample periods for each dataset. In order to evaluate the estimate performance we used every 50^{th} parameter set across the $10,000$ sets that were obtained as part of the MCMC procedure. This way were obtained 200 estimates of the out-of-

sample log-likelihood for each of the 20 catalogue out-of-sample periods. Hence, we can examine the out-of-sample log-likelihood with respect to either each extension for the mean performance across selected MCMC samples or the opposite - evaluate the mean performance across all out-of-sample periods for each of the MCMC samples. An overall measure is either the mean out-of-sample log-likelihood or the maximum one across all MCMC sets and all catalogue out-of-sample periods.

$\phi.$	DIC_U	DIC_K	DIC_D	$\bar{l}^o_U$	$\bar{l}^o_K$	$\bar{l}^o_D$
ϕ_1	5985.64	2495.93	4347.91	-1150.57	-1079.11	-1084.31
ϕ_2	5984.60	2372.27	4583.25	-913.29	-862.19	-832.30
ϕ_3	5918.93	1780.47	3670.92	-830.16	-777.72	-778.89

Table 4.1: Comparison between the performance of Unif (U), KDE (K) and DP (D) ETAS models across three uncaused events' spatial distributions ($\phi.$) with respect to the Tohoku District [Ogata, 1998] MLE estimated based simulated catalogues.

The obtained results for $\phi_1(\cdot)$ and $\phi_3(\cdot)$ show that KDE ETAS model outperformed DP ETAS model, and that both outperformed the Fixed ETAS model (Table 4.1). However, the results obtained based on $\phi_2(\cdot)$ show that DP is better than KDE with respect to all diagnostic tests. On Figure 4.2 are shown the spatial distribution of the obtained data based on $\phi_2(\cdot)$, as well as the sequence of log-likelihoods for each model evolving over each iteration of the Gibbs sampler (after convergence). It is clearly observable that there is a difference between the overall fitting capabilities between the 3 spatial ETAS models. The out-of-sample diagnostics are shown on Figure A.1.

Since these results are mixed and show that both DP and KDE are capable of outperforming each other depending on the model parameters, we will now develop a larger simulation study to gain insight into the factors which determine when each is most suitable.

4.9.3 Large Scale Simulation Study

In order to examine further the behaviour of the spatial ETAS models, we created a number of simulated data sets by varying the model parameters. We set $\mu_0 = 0.325$, $c = 0.0353$ and $p = 1.121$ to be constant. Then we consider: $\alpha \in \{1.0, 1.3, 1.6, 1.9\}$, $\bar{K} \in \{0.1, 0.3, 0.5\}$, $d \in \{0.01, 0.255, 0.5\}$ and $d \in \{1.10, 1.55, 2\}$. We exclude the combinations of parameters which result in an expected productivity greater than 1 since these can potentially generate infinite catalogues as discussed in Section 4.7. This resulted in 63 different parameter sets.

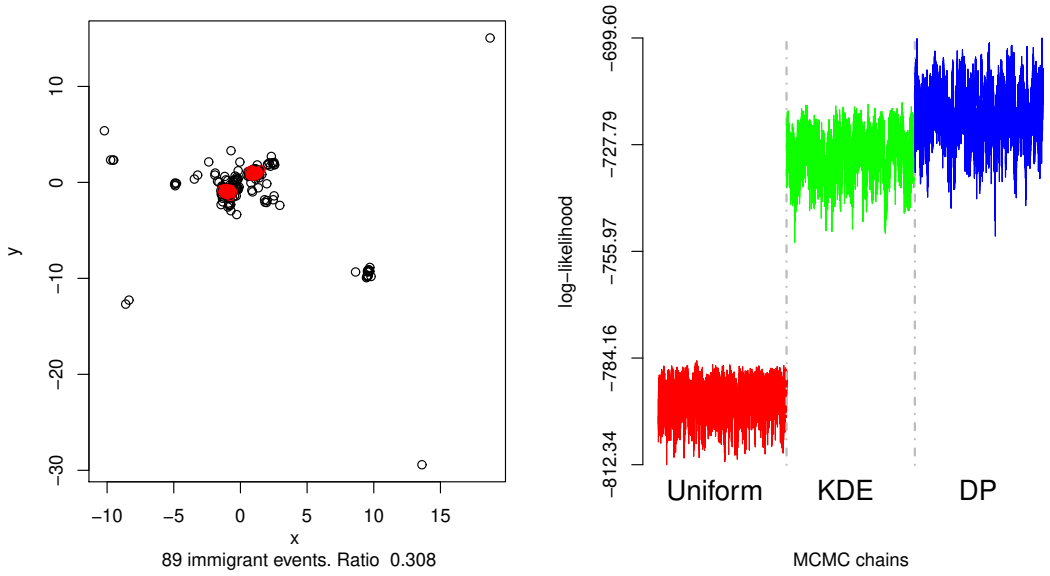

Figure 4.2: Descriptive plots of $\phi_2(\cdot)$ with parameter set $(\mu_0, \alpha, \bar{K}, c, p, d, q) = (0.325, 1.407, 0.0353, 1.121, 0.322, 0.0159, 1.531)$. Left: Data spatial distribution. In Red are all immigrant event while all others are displayed in Black. There are 89 immigrant events and 200 offsprings which corresponds to a ratio of 0.308. Right: The obtained log-likelihood with respect to the three different version of ETAS for 10,000 MCMC simulations.

On Tables A.1, A.2 and A.3 are shown the obtained results with respect to the 63 simulations for each of the three uncaused events' spatial densities as of Equations 4.13, 4.14 and 4.15 respectively. The first column consists of the specific values for parameters α, $\bar{K}$, d and q. The next one is the total number of events n obtained by this simulation. p_μ is the proportion of uncaused events, followed by the overall *Area* that the catalogue spans. All goodness-of-fit diagnostics are provided with respect to the Uniform, KDE and DP ETAS models as introduced in Section 4.3, which correspond to subscripts U, K and D respectively. $\hat{l}_\cdot$ provides the highest log-likelihood value across the MCMC. The DIC value for the corresponding model is shown under $DIC_\cdot$ as introduced in Section 4.8.1. The log-likelihood summary across all extensions (out-of-sample) and all sampled parameters is shown as $\hat{l}^o_\cdot$ (the maximum value) and $\bar{l}^o_\cdot$ (the mean value).

Direct comparison between all results shown on Tables A.1, A.2 and A.3 is a challenge since every goodness-of-fit metric can be considered on its own as sufficient for the allocation of the best model. However, the Fixed model is never superior across the other two. We examined the performance of KDE vs DP spatial ETAS models with respect to whether they excel in a single metric or across all provided metrics. The

summary of this comparison is provided on Table 4.2. This table presents the number of datasets that allocate either KDE or DP as the best model based on either maximum log-likelihood $\hat{l}$ or DIC or out-of-sample maximum likelihood $\hat{l}^o$ or out-of-sample mean log-likelihood $\bar{l}^o$ or with respect to all previous metrics (referred to as *best*). We further provided the aggregated counts across all 189 simulations. It can be seen that both the DP and KDE versions of the model can outperform each other for different values of the model parameters. This shows that they are both likely to have value for estimating real-world catalogues.

It is interesting to understand the factors which makes DP superior to KDE for particular data-sets, since this will give us a general rule for deciding which one is most appropriate to use. We propose the following hypothesis, which seems intuitively reasonable: since KDE forms its estimate of $\phi(x, y)$ by using all the earthquakes in the catalogue rather than only the immigrant events, we would expect it to perform well either when most earthquakes are immigrants, or when the true distribution $\phi(x, y)$ of mainshocks is not too dissimilar to the overall distribution of earthquakes in the catalogue.

We would expect the DP approach to perform better when K is large (since this results in a higher proportion of aftershocks relative to immigrants), and also when the parameters d and q are large since this results in the distribution of triggered events being spread out over a wider areas, which increases the spatial discrepancy between the mainshock distribution and the overall catalogue distribution.

subset	model	$max(\hat{l}_\cdot)$	$min(DIC)$	$max(\hat{l}^o_\cdot)$	$max(\bar{l}^o_\cdot)$	best
$\phi_1(\cdot)$	KDE	49	54	49	52	47
	DP	14	9	14	11	8
$\phi_2(\cdot)$	KDE	26	33	25	27	23
	DP	37	30	38	36	29
$\phi_3(\cdot)$	KDE	32	35	26	27	24
	DP	31	28	37	36	28
All	KDE	107	122	100	106	94
	DP	82	67	89	83	65

Table 4.2: Number of datasets that allocate either KDE or DP as the best model based on either maximum log-likelihood $\hat{l}$ or DIC or out-of-sample maximum likelihood $\hat{l}^o$ or out-of-sample mean log-likelihood $\bar{l}^o$ or with respect to all previous metrics (*best*).

To test this hypothesis, we will try to create a single measure which represents the discrepancy between $\phi(x, y)$ and the overall catalogue distribution. This relationship is

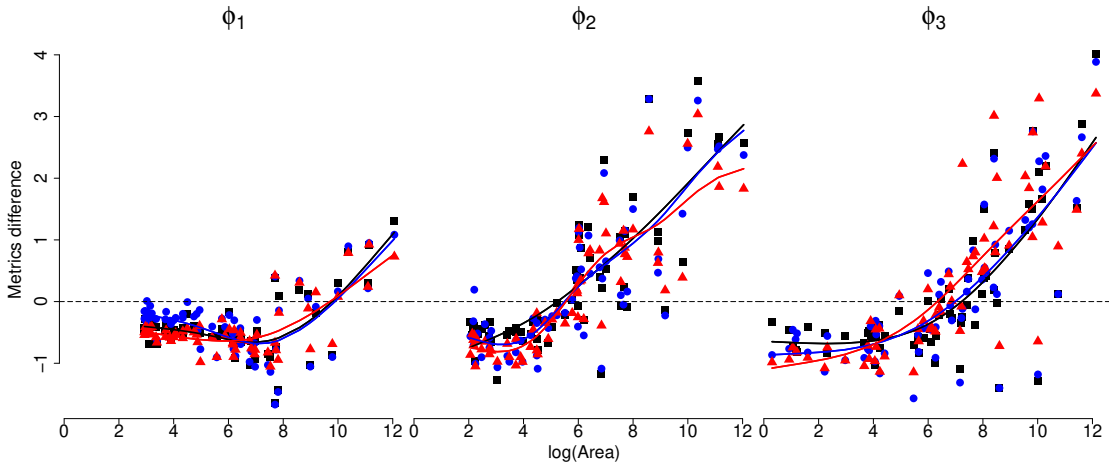

Figure 4.3: Standardised differences of performance metrics of DP ETAS related to KDE ETAS with respect to the logarithmic transformation of every catalogue overall *Area* across the three uncaused events' spatial densities $\phi_{\cdot}(\cdot)$ (Section 4.9.1). ■ stands for the difference between in-sample log-likelihood values for DP ETAS minus KDE ETAS; ● stands for the difference between DIC values for KDE ETAS minus DP ETAS; ▲ stands for the difference between out-of-sample log-likelihood values for DP ETAS minus KDE ETAS. For ease of display all values are re-scaled to follow a zero mean, unit variance Normal distribution. The three solid lines on each sub-plot represent the fitted lines of the pattern with respect to the three discussed difference (in their respective colours). The horizontal dashed line indicate the threshold for which DP ETAS will be considered to outperform KDE ETAS.

primarily influenced by the overall area that the catalogue spans. Since all immigrant events are restricted to lie within the same area (Σ), the overall area of the catalogue is driven primarily by the values of d and q and $\bar{K}$, which affects how the triggered events spread out relative to the immigrants. As such, we compute the resulting areas for each of the 63 catalogues which were simulated by varying the parameter values. In Figure 4.3, we plot how the area of the catalogue relates to the degree to which DP performed KDE. Specifically, we plot the difference in the DIC and out-of-sample log-likelihood values between DP and KDE, as a function of catalogue area. It can be seen that there is a clear relationship between the performance measures and the overall area of a catalogue - a larger area is associated with a better performance of DP ETAS. The correlation between these measures and the proportion of immigrant events (p_{μ_0}) is -0.27 and -0.26 for the DIC and ML differences respectively. This largely confirms our previous hypothesis; in general, the DP model outperforms KDE when there are a large number of aftershocks that are spread out over a wide area, while KDE performs best when the number of aftershocks is smaller.

4.10 Real earthquake sequences

In this Section we explore the performance of the spatio-temporal ETAS model across real earthquake catalogues. As expected, the Uniform ETAS model performed poorly compared to the other two models. For clarity we are not going to discuss further this model and we will focus on the results from the other two. The summary of all diagnostic measures across all analysed datasets are shown on Table 4.3.

4.10.1 Italian catalogue

Italy has suffered recently from a rather misfortunate phenomenon. On the 24^{th} of August 2016 was registered an earthquake of 6M, followed by more than 2500 aftershocks that caused the death of nearly 300 people [Luzi et al., 2017]. Non-self exciting point processes cannot address a short term prediction of such a sort since its intensity cannot increase rapidly enough. Further, a self-exciting point process without a spatial component cannot address such patterns as well due to the large seismic activity across multiple rupture areas of the Apennines peninsula. For all these reasons, we decided to apply the novel models introduced in the Chapter to the Italian seismic sequence.

We obtained the required data from the Italian National Institute of Geophysics and Volcanology (Instituto Nazionale Di Geofisica e Vulcanalogia `http://www.ingv.it`). The data were gathered from 01/04/1999 to 01/04/2019 that spans Italy within 35^o, 49^o N and 5^o, 20^o E with minimum magnitude of $M_0 = 3$. Then we split the data into a train set from 01/04/1999 to 01/04/2014 (4669 events) and a test set from 01/04/2014 to 01/04/2019 (2171 events).

The goodness-of-fit results are shown on Table 4.3. From a standard Bayesian perspective DP ETAS outperforms KDE ETAS because the only truly Bayesian measure across all model diagnostics is DIC in which DP ETAS has shown superior results. However, KDE ETAS outperforms DP ETAS in all other metrics both in- and out-of-sample. Therefore we conclude that KDE ETAS provides a better fit for the Italian catalogue. The main reason for this result is based on the way the catalogue was accumulated - it consists of events that were primarily detected on shore. Given the density of the catalogue, spatial distribution of the events and possible omission of caused events further away from the triggering origin, it is evident that KDE will outperform DP in this scenario.

4.10.2 Friuli, Italy

This is an area in Italy that is primarily known for the 6.5 M earthquake that occurred on 06/05/1976, followed by multiple aftershocks with considerably large magnitude. In the past 40 years, there were no detected earthquakes exceeding 5 M. This catalogue is very challenging because the available data prior to the 1976 Friuli Earthquake are limited.

We based our study on the earthquake occurrence from 01/01/1975 to 01/01/2019 that covers the area within $46^o36'$, 46^o N and $12^o18'$, $13^o30'$ E with minimum magnitude of $M_0 = 3$. For inferential purposes we split the data into a train set $01/01/1975 - 01/01/2004$ with 310 events and a test set $01/01/2004 - 01/01/2019$ that comprises of 20 events. The data were obtained from the United States Geological Survey (USGS) catalogue (`http://earthquake.usgs.gov/`).

As before, the goodness-of-fit results are shown on Table 4.3. KDE ETAS outperforms DP ETAS in terms of DIC value but suffers in the out-of-sample study. This indicates that KDE is overfitting the uncaused events spatial density. However, the train test is rather small and for that reason it is beneficial to examine the performance on catalogues with denser test sets.

4.10.3 Vrancea, Romania

Vrancea is an area in Romania that has a strategic importance on South-Eastern Europe. On 4/3/1977 was detected the 7.2 M Vrancea earthquake that caused large destruction and human loss in both Bulgaria and Romania. Examining this region is of critical interest because there were no large earthquakes in the area occurring for a prolonged period of time. There are a number of very ambitious projects which were developed or started development during the 20^{th} century in both Bulgaria and Romania. Some of them were proven to be poorly executed, causing major disasters in the recent years. Probably the most debatable projects are related to the nuclear power plants in these countries. At the moment, there are two nuclear power stations operating, one in each country - Kozloduy in Bulgaria and Cernavoda in Romania. An additional power plant is under development in Bulgaria near the town of Belene. A possible malfunction caused in any of them can be devastating not only for Bulgaria and Romania, but also for the entire Black Sea basin. There are a number of sites that can provide a smaller, yet considerably negative impact on the two countries' development if the appropriate

earthquake hazard measure is not taken into account. Some of them include water dams and reservoirs, artificial rivers and channel barriers, and gas lines.

This study analyses the earthquake information 01/01/1974−01/01/2019 that covers 46^o, $45^o18'$ N and 27^o, 26^o E with a minimum magnitude of $M_0 = 2.5$. The data were split into a train set 01/01/1975 − 01/01/2014 with 529 events and a test set 01/01/2014 − 01/01/2019 that comprises 46 events. The data were obtained from the United States Geological Survey (USGS) catalogue (`http://earthquake.usgs.gov/`).

The goodness-of-fit information is illustrated on Table 4.3. Similarly to the previous section, KDE ETAS outperforms DP ETAS in sample and provides a poor out-of-sample performance. This indicates that KDE is overfitting the uncaused events spatial density. We conclude that DP ETAS is more useful for this specific dataset.

4.10.4 Zakynthos and Kefalonia, Greece

Zakynthos and Kefalonia are subject to prolonged seismic activity. The area of interest spans $38^o33.54'$, $47^o14.34'$ N and $21^o36.96'$, $19^o39.96'$ E. The most important event in the region is the 6.8 M Ionian earthquake that occurred on 12/08/1953. Including data from this period is very challenging due to the advancements of detection methodology since then. For this reason we focused our study on a more recent time frame (1/1/1069 − 1/1/2019) and further increased the detection threshold ($M_0 = 4.5$) to ensure consistency throughout the catalogue. The data were split into a train set 01/01/1969 − 01/01/2018 with 343 events and a test set 01/01/2018 − 01/01/2019 that comprises 109 events. The data were obtained from the United States Geological Survey (USGS) catalogue (`http://earthquake.usgs.gov/`). KDE again outperformed DP in sample due to overfitting the data since DP ETAS performs better out-of-sample (see. Table 4.3).

4.10.5 Kyushu, Japan

The last catalogue that we analyse in this Chapter is based on the seismicity in the area around the Kyushu island in Japan. This seismic sequence is of prime importance due to the escalation of seismic activity in early 2019. We worked on the temporal interval 1/1/1069−1/1/2019 within $36^o36.9'$, $29^o59.58'$ N and $134^o9.9'$, $127^o44.94'$ E region, with detection threshold $M_0 = 4.5$. The obtained from the United States Geological Survey (USGS) catalogue (`http://earthquake.usgs.gov/`) consists of 761 events that we split

into a train set 01/01/1969 − 01/01/2016 (594 events) and a test set 01/01/2016 − 01/01/2019 (167 events). Goodness-of-fit results are illustrated on Table 4.3. They clearly indicate that KDE again outperformed DP in sample while DP ETAS clearly fits the the date better given its supremacy in the out-of-sample characteristics.

Data	DIC_K	DIC_D	$\bar{l}^o_K$	$\bar{l}^o_D$
Italy	- 6318.75	-7403.23	11580.85	11552.51
Friuli	946.66	995.58	-36.36	-29.11
Vrancea	2710.23	2731.23	-43.88	-37.86
Zakynthos	846.51	903.64	-36.66	-31.00
Kyushu	1811.71	2278.01	-29.95	-13.72

Table 4.3: KDE and DP based spatial ETAS model comparison across real catalogues. Lower values of the DIC and larger (less negative) values of the out-of-sample likelihood indicate superior performance. The large value of the likelihood for the catalogue that represents whole of Italy is due to the very large number of events compared to the other catalogues.

4.11 Conclusions

In this Chapter, we explored the most commonly used version of the spatio-temporal ETAS model. We further extended its uncaused events' density distribution modelling using a Dirichlet process non-parametric model and further compared it to an Uniform distribution and Kernel density estimation. The introduced posterior sampling scheme can easily be deployed on realistic seismic catalogues due to its scalability.

Our experimental results suggest that the KDE provides better results, while DP excels in the case of large-scale offspring kernel causality. The Uniform ETAS usually exhibits poor performance across all analysed scenarios which further strengthens the usefulness of the introduced DP and KDE based spatial ETAS models. KDE ETAS is typically better than DP ETAS models based on standard in-sample goodness-of-fit tests. A general exception of this pattern is where a large proportion of caused events are present in a catalogue. Further, the caused events should express a large-lag spatial offspringing behaviour. DP ETAS usually excels in the out-of-sample tests. Both KDE and DP based ETAS models provide an exceptional performance and they both should be considered in the applications of the spatio-temporal ETAS model on real catalogues.

Chapter 5

Spatially Explicit Capture Recapture as a Self-Exciting Point Process

In this Chapter we extended the most recent developments of the spatial-ETAS model to address the concept of animal movement as a point process. Spatially explicit capture recapture (SECR) methods are widely used to estimate animal density from trap surveys models. Nowadays traps are usually substituted with cameras that can provide detailed temporal occurrence information.

Here we extend continuous time SECR models to address dependence between detections of every animal across all cameras, and estimate model parameters that are shared between all animals which allows information to be pooled across individuals. This is achieved by a hybrid Self-exciting point process. We allow the intensity of the point process for an animal at each camera to depend on the times and locations of previous detections as well as on the distances from the animal's activity centre to each camera trap. The data for each animal can be sparse which we overcome by hierarchical pooling of offspring model parameters across all animals in the region, since all individuals from a certain species share similar behavioural patterns.

All introduced methods are applied to two datasets. The first examines leopards in Royal Manas National Park, Bhutan. The second one consists of tigers in Nagarahole reserve, India. Our study suggests that the introduced self-excitation component adds value to the model specification compared to simpler, non-self-excitation point processes.

5.1 Background

Camera traps record the exact time of each detection, but this information is typically discarded and the data reduced to numbers of detections within a time interval of a day, a week, or longer, for SECR analysis. There are a few exceptions. [Borchers et al., 2014] developed a maximum likelihood SECR estimator that uses exact detection times, and used it to estimate jaguar abundance. [Dorazio and Karanth, 2017] developed a Bayesian SECR estimator that does the same, although in their analysis of a camera trap survey of tigers in the Nagarahole Reserve in India, they reduced time to a factor with two levels: "day" and "night". Both these methods assume that the hazard of detecting an animal by any camera at any time does not depend on where or when the animal was previously detected. This is not a realistic assumption. Animals take time to move from one camera to another and will tend to be detected at traps closer to the trap of their last detection before being detected at those farther away. [Borchers et al., 2014] showed that under this unrealistic assumption, data on exact times of detection does not improve density estimation, although it does contain information on how animal activity varies over time. In this Chapter we develop a method that overcomes this assumption, by using ETAS to model the spatio-temporal dependence in detections that arises as animals move through the study area.

SECR models for camera trap surveys model animal activity centres as arising from a spatial point process [Johnson et al., 2010, Buckland et al., 2016, Borchers and Marques, 2017]. Continuous time SECR models like those of [Borchers et al., 2014, Dorazio and Karanth, 2017] model the detection process at each camera as independent temporal point processes that are time-invariant but depend on the distance of each camera from an animal's activity centre. The much more common discrete-time SECR models [Yip, 1991, Barbour et al., 2013] aggregate over time within subjectively chosen time intervals and model the counts of detections at each camera in each interval as independent realisations of counting processes with expected values that depend on the distance of each camera from an animal's activity centre.

We use the self-exciting nature of ETAS model to address the elevated probability of detecting an animal in the vicinity of a camera that has just detected it, for some period after the initial detection. However the data for each animal can be sparse (few detections of each animal). Camera trap surveys would not normally generate enough data to estimate separate processes for each animal. We deal with this by pooling the

ETAS model offspring parameters across all animals in the survey region, assuming that the movement of all animals in the region is governed by the same process. This kind of assumption is common with SECR models, which typically assume that all animals share the same parameters either with respect to a counting process (discrete-time SECR) or temporal point process (continuous-time SECR), conditional on their activity centre locations. There are SECR models that allow animal-level random effects in the detection process but we do not consider such models here.

The remainder of this Chapter proceeds as follows. In Section 5.2 we describe the SECR ETAS model with its properties. Then we address the simulation mechanism of this process in Section 5.3 followed by introduction of a non-self-excitation model alternative (Section 5.4). In Section 5.5, we introduce the methods for model estimation based on the Bayesian paradigm. In Section 5.6, we apply all introduced concepts to two datasets to show the strengths and weaknesses of the SECR ETAS model. Finally, in Section 5.7 we summarise all our findings.

5.2 The structure of SECR ETAS model

Most, if not all, SECR models have a hierarchical structure in which some elements are shared between animals while others are individual-specific. The data consist of arrival times for all animals, camera indexes and locations at which they were spotted as well as an actual animal identifier. We denote by ζ the total number of cameras in the detection region, with $c_i = \{x_{c_i}, y_{c_i}\}$ the co-ordinates of the i^{th} camera location so that the set of all camera locations is $\{c_1, \ldots, c_\zeta\} = \boldsymbol{\zeta}$. The data associated with the animals that were observed consists of detection times $\{t.\}$, locations $\{x., y.\}$ of the camera on which they were captured and the exact animal index $\{a.\}$ that illustrates from which animal this realisation is considered to be obtained. Hence, the overall data available for the model calibration is $\mathscr{H}_t = \{(t_1, x_1, y_1, a_1), (t_2, x_2, y_2, a_2), \ldots : t_i < t\} \cup \boldsymbol{\zeta}$, where t_i is the occurrence time at location $\{x_i, y_i\}$ of an animal with index a_i. All of the observed animals are, of course, a realisation across the camera space $\boldsymbol{\zeta}$. This restricts our study only to locations covered by cameras since it would not be possible to observe animals at any locations not covered by them. Although cameras are typically positioned in a manner aiming to detect the majority of the animals in the area of interest, it is likely that some animals will be undetected due to various reasons [Borchers and Marques, 2017, Goldberg et al., 2015b, Dorazio and Karanth, 2017].

5.2.1 Single animal representation

The data from each animal are considered to be a realisation from a spatio-temporal ETAS model. The basic construction of the causality function follows closely the widely used characteristics of ETAS models commonly applied in seismology, while the uncaused events dependence is tailored to the needs of SECR. This way animal occurrences are initiated based on their home-range centres while further actions are primarily driven by species dependent characteristics that are modelled by the offspringing behaviour.

SECR ETAS shared components with standard ETAS model

The general structure of the point process that models every animal behaviour has the following intensity function:

$$\lambda(t, x, y|\mathscr{H}_t) = \mu(x, y) + \sum_{t_i<t} g(t - t_i) \times s(x - x_i, y - y_i). \tag{5.1}$$

We set the function $g(\cdot)$ to be the modified Omori law:

$$g(\Delta t) = \frac{K(p-1)}{(\Delta t + c)^p c^{(1-p)}}, \tag{5.2}$$

where $\Delta t > 0$ is the corresponding time-lag between two events. The modified Omori law (Equation 5.2) is commonly used for modelling aftershocks of earthquakes [Omori, 1894], however it could be applied in any context. A further study into different functional forms of $g(\cdot)$ could improve considerably the SECR ETAS model performance. However, this is beyond the scope of our work.

The spatial offsprings distribution is modelled as being proportional to zero mean multivariate normal distribution:

$$s(x_j - x_i, y_j - y_i) \propto N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} d & 0 \\ 0 & d \end{bmatrix} \right) = s_t(x_j - x_i, y_j - y_i|x_j, y_j), \tag{5.3}$$

for $i < j$, where x_j and y_j correspond to the location of the event from which we evaluate the spatial offspring causality function $s(\cdot)$. However, the causality behaviour is shared only across camera observation centres which implies a discrete support of the spatial lags between the location of the parent ($\{x_j, y_j\}$ in Equation 5.3) and all candidate transfer locations (e.g. $\{x_{c_i}, y_{c_i}\}$ for camera c_i) represented by the set of all camera

locations $\boldsymbol{\zeta}$. Then the spatial kernel for the observation with index j with respect to previous detection with index i is:

$$s(x_j - x_i, y_j - y_i) = \frac{s_t(x_j - x_i, y_j - y_i | x_j, y_j)}{s_c(x_j, y_j)},$$

where $s_c(x, y) = \sum_{i=1}^{\zeta} s_t(x - x_{c_i}, y - y_{c_i} | x, y)$. Thus $s(\cdot)$ is a valid discrete probability function.

The evaluation of both $g(\cdot)$ and $s(\cdot)$ is only feasible for sequences that consist of more than one event (detection). However, some of the animals that are observed are detected only once. This provides no data for which to estimate their offspring parameters. To overcome this problem we assume that all animals have similar offspring dependence. Hence, the offspring parameters (K, c, p, d) will be common for all animals that are present in the study region.

Uncaused events spatial dependence

The immigrant events' conditional intensity $\mu(\cdot)$ comprises a constant, temporal arrival rate μ_0 and a spatial function $\phi(x, y)$ based on the following relationship:

$$\mu(x_i, y_i | \boldsymbol{h}) = \mu_0 \times \phi(x_i, y_i). \tag{5.4}$$

The immigrant spatial distribution $\phi(x, y)$ for every unique animal can be interpreted as a discrete uniform distribution across all placed cameras $\boldsymbol{\zeta}$. This will suggest that a single animal can appear on any camera regardless of its home-range. However, the uniformity assumption is rather unrealistic as it suggest that an animal's home range centre is irrelevant for animal movement [Borchers and Marques, 2017, Dorazio and Karanth, 2017]. On the contrary, assigning a multivariate normal density provides a rather strict alternative in which the animal is unlikely to be seen on any camera outside of its home-range which is also not ideal. Since an animal is predominantly in its home-range and occasionally visiting a new territory, it is reasonable to consider a mixture model between these two components. This way we will assign large density to the area closest to every animal's home-range but also guarantee a fair chance of visiting a new territory. The rate between the two components is controlled by scaling parameter $\gamma \in [0, 1]$. Then, the uncaused events' probability mass function conditioned

on the animal of interest home-range centre $\boldsymbol{h}$ is proportional to:

$$\phi(x_i, y_i|\boldsymbol{h}) = \Big[(1-\gamma)\frac{1}{\zeta} + \gamma f(\{x_i, y_i\}|\boldsymbol{h}, \Sigma)\Big], \tag{5.5}$$

where ζ is the number of the cameras that were placed in the experiment regardless of whether an animal was detected on them or not. $f(\cdot)$ is the density of a bivariate Normal distribution, Σ is the common full covariance matrix for all animals and $\boldsymbol{h}$ is the current animal's home-range centre that is fixed across all observations from this animal. If the overall number of observations is too small we can restrict further the shape of the common covariance matrix Σ by setting all non-diagonal elements to zero.

We placed a discrete point mass on every camera that depends on a continuous random variable. This suggests that the cumulative intensity of the uncaused events for every animal is aggregation of the intensity across every placed camera:

$$\phi_c(\boldsymbol{h}) = \sum_{i=1}^{\zeta} \phi(x_i, y_i|\boldsymbol{h}).$$

Likelihood contribution from a single animal

The likelihood of the process for a single animal observation sequence follows the general point process paradigms (see Section 2.4.1). The probability of observing the data $\mathscr{H}_T^{(j)} = \{(t_1, x_1, y_1, j), (t_2, x_2, y_2, j), \ldots : t_i < T\} \cup \boldsymbol{\zeta}$ associated with a single animal j across all placed cameras $\boldsymbol{\zeta}$ in a temporal interval $[0, T]$, given the model parameters is:

$$\mathcal{L}(\theta_j; \mathscr{H}_T^{(j)}) = \prod_{i=1}^{n_j} \lambda(t_i, x_i, y_i|\mathscr{H}_T^{(j)}, \theta_j) e^{-\int \sum_{i=1}^{\zeta} \lambda(z, x_i, y_i|\mathscr{H}_T^{(j)}, \theta_j) dz} \tag{5.6}$$

with respective log-likelihood function $\ell(\theta_j; \mathscr{H}_T^{(j)}) = \log(\mathcal{L}(\theta_j; \mathscr{H}_T^{(j)}))$ of

$$\ell(\theta_j; \mathscr{H}_T^{(j)}) = \sum_{i=1}^{n_j} \log(\lambda(t_i, x_i, y_i|\mathscr{H}_T^{(j)}, \theta_j)) - \mu_0 T \phi_c(\boldsymbol{h}) \\ - K \sum_{t_i < t} \left(1 - \frac{c^{p-1}}{(T - t_i + c)^{p-1}}\right), \tag{5.7}$$

where n_j is the total number of observations of the j^{th} animal and θ_j is the full parameter set of the model with respect to the j^{th} animal i.e. $\theta_j = \{\mu_0, \gamma, \boldsymbol{h_j}, \Sigma, c, p, K, d\}$.

Animal detection probability

Evaluating the probability to detect (or undetected) an animal within the study area is essential for inferential purposes as it could be used for estimation of the full population. Animal detection depends on the duration of the temporal interval in which the data are collected, the number and density of the placed cameras in the area and, most importantly, animal's home-range centre position with respect to all traps.

Any conducted experiment cannot guarantee to observe all animals within a detection area of interest. Therefore, there exists a non-zero probability of not detecting an animal. Let us define a dichotomous random variable $S \in \{0, 1\}$ where state 0 indicates that an animal is undetected and state 1 if it is detected. The probability of an animal not being spotted is equivalent to the probability of not obtaining any events from the point process defined in Section 5.2.1. Since the animal is not spotted, the intensity as of Equation 5.1 reduces to $\mu(x, y)$ because the self-exciting term of the intensity function disappears due to the lack of observations to initiate the excitation component.

The probability of not detecting an animal is $e^{-\Lambda(\boldsymbol{h})}$ where $\Lambda(\boldsymbol{h})$ is the cumulative intensity of the process with home-range centre $\boldsymbol{h}$ over the spatial region with area A and temporal interval $[0, T]$:

$$\Lambda(\boldsymbol{h}) = \int_0^T \sum_{i=1}^{\zeta} \mu(x_i, y_i|\boldsymbol{h})dt = T\mu_0 \sum_{i=1}^{\zeta} \phi(x_i, y_i|\boldsymbol{h}) = \mu_0 T\phi_c(\boldsymbol{h}).$$

The above expression holds true since $\phi_c(\boldsymbol{h}) = \sum_{i=1}^{\zeta} \phi(x_i, y_i|\boldsymbol{h})$ and μ_0 is a constant over time. Then, the probability of an animal to be detected given its home-range centre $\boldsymbol{h}$ is equivalent to:

$$P(S = 1|\boldsymbol{h}) = 1 - \exp\Big(- T \sum_{i=1}^{\zeta} \mu(x_i, y_i|\boldsymbol{h})\Big) = 1 - \exp(-\mu_0 T\phi_c(\boldsymbol{h})).$$

The equation above represents the probability of an any outcome from which is subtracted the probability of obtaining no events from a Poisson distribution with rate $\Lambda(\boldsymbol{h})$.

Although every animal has a home-range centre $\boldsymbol{h}$ associated with it, we can gather information for the home-ranges only for the animals that we observe. In order to quantify the unconditional probability of an animal to be spotted we have to integrate out the distribution of the home-range centres, $P(\boldsymbol{h})$. Then the unconditional probability

to observe an animal is:

$$P(S=1) = \int_A P(\boldsymbol{h})P(S=1|\boldsymbol{h})d\boldsymbol{h}.$$

In this study we assign $P(\boldsymbol{h})$ to be a uniform distribution over the area of interest A i.e. $P(h) = 1/A$. Then the probability of detecting an animal is:

$$P(S=1) = 1 - \frac{1}{A}\int_A \exp(-\mu_0 T\phi_c(\boldsymbol{h}))d\boldsymbol{h}. \tag{5.8}$$

The uniformity of home-range centres is a simplifying assumption that eases further derivations. Without loss of generality we could assign another distribution given sufficient ecological reasoning.

5.2.2 Multiple animals

Let us consider the likelihood function of all observed animals. Based on the assumption that all animal occurrences are independent, the likelihood function is the product of the likelihood functions for every observed animal as introduced in Equation 5.6. For a point process that consists of m unique animals, where n_j is the total number of observations of the j^{th} animal, the likelihood function is:

$$\begin{aligned}\mathcal{L}_o(\theta; \mathscr{H}_T) &= \prod_{j=1}^{m} \mathcal{L}(\theta_j; \mathscr{H}_T^{(j)}) \\ &= \prod_{j=1}^{m}\prod_{i=1}^{n_j} \lambda_j(t_i, x_i, y_i|\mathscr{H}_T^{(j)}, \theta_j) e^{-\int_0^T \sum_{s=1}^{\zeta} \lambda_j(z,x_i,y_i|\mathscr{H}_T^{(j)},\theta_j)dz},\end{aligned} \tag{5.9}$$

where $\mathscr{H}_T$ is the full data across all animals and all placed traps, θ is the full parameter set of the model for all animals i.e. $\theta = \cup_{j=1}^{m}\theta_j = \{\mu_0, \gamma, h, \Sigma, c, p, k, d\}$ and h is the collection of the home-range centre for all animals $h = \{\boldsymbol{h_1}, ..., \boldsymbol{h_m}\}$. This corresponds to a log-likelihood form of:

$$\ell_o(\theta; \mathscr{H}_T) = \sum_{j=1}^{m}\left(\sum_{i=1}^{n_j} \log(\lambda_j(t_i, x_i, y_i|\mathscr{H}_t, \theta_j) - \int_0^T \sum_{s=1}^{\zeta} \lambda_j(z, x_s, y_s|\mathscr{H}_t, \theta_j)dz\right). \tag{5.10}$$

The inner sum goes through the indexes of each of the observed animal sequences. Hence $\sum_{i=1}^{m} n_i = n$ where n is the total number of observations in the full dataset.

However, the number of detected animals m is a subset of the full animal population M in the area of interest.

Unobserved animals

The log-likelihood as of Equation 5.10 takes into account only the information that we observed. The fact that an animal is not captured on camera is information which should contribute to the likelihood of the data given the model parameters. Then, the log-likelihood of the undetected animals is the probability of detecting m out of M animals and to further not detected $M - m$ of them with unit detection probability following Equation 5.8. This returns the following undetected events likelihood function:

$$\mathcal{L}_u(\theta; \mathscr{H}_T) = \binom{M}{m} \left(\frac{1}{A} \int_A \exp(-\mu_0 T \phi_c(\boldsymbol{h})) d\boldsymbol{h} \right)^{(M-m)}, \tag{5.11}$$

with corresponding log-likelihood representation of

$$\ell_u(\theta; \mathscr{H}_T) = (M-m) log \left(\frac{1}{A} \int_A \exp(-\mu_0 T \phi_c(\boldsymbol{h})) d\boldsymbol{h} \right) + log \binom{M}{m}.$$

A typical capture recapture study is primarily interested in evaluating the overall population size M. The main link between this quantity and the observed data are the estimation of the proportion of unseen animals [Goldberg et al., 2015b].

Overall likelihood of the process

The full log-likelihood that combines both the observed and unobserved information is the summation of the two separate log-likelihoods $\ell(\theta; \mathscr{H}_T) = \ell_o(\theta; \mathscr{H}_T) + \ell_u(\theta; \mathscr{H}_T)$ as follows:

$$\begin{aligned} \ell(\theta; \mathscr{H}_T) = & \sum_{j=1}^{m} \left(\sum_{i=1}^{n_j} \log(\lambda_j(t_i, x_i, y_i | \mathscr{H}_T^{(j)}, \theta_j)) - \int \sum_{s=1}^{\zeta} \lambda_j(z, x_s, y_s | \mathscr{H}_T^{(j)}, \theta_j) dz \right) \\ & + (M-m) log \left(\frac{1}{A} \int_A \exp(-\mu_0 T \phi_c(\boldsymbol{h})) d\boldsymbol{h} \right) + log \binom{M}{m}, \end{aligned} \tag{5.12}$$

where M is the parameter with greatest importance for a SECR study. Its point estimate can be obtained as a proportion of observed animals and detection probability as defined in Equation 5.8. In Section 5.5 we introduce a novel MCMC algorithm that will provide

a more robust estimation technique which further propagates the inherent parameter uncertainty.

5.3 Simulation

In this Section we provide a brief description of the methods that could be employed for creating a sample from SECR ETAS.

We begin by specifying the overall number of animals M in the area of interest A with respect to all placed cameras $\boldsymbol{\zeta}$. Events' locations are observable only at specific camera instances. All animal sequences should be in a temporal interval $\tau \in [0, T]$. Every animal is a realisation from its own ETAS model.

We firstly sample a home range centre $\boldsymbol{h}$ from $P(\boldsymbol{h})$. We continue with sampling uncaused events and their corresponding offsprings from multiple generations. If we fail to initiate the sequence, then the animal is considered undetected.

5.3.1 Uncaused events

Based on the already sampled animal home-range centre $\boldsymbol{h}$ we can evaluate the cumulative intensity across space and time as:

$$\int \sum_{i=1}^{\zeta} \mu_0 \phi(x_i, y_i \boldsymbol{h}) dt = \mu_0 T \phi_c(\boldsymbol{h}).$$

The number of immigrant events follows a Poisson distribution with a rate $\mu_0 T \phi_c(\boldsymbol{h})$. The corresponding arrival times are uniformly sampled in the temporal region $(0, T)$. Given the discrete spatial nature of the process, we can sample the camera labels directly from their probability distribution. For every camera location we can evaluate the probability of occurrence based on $\phi(\cdot)$ as of Equation 5.5 and then sample the required number of camera indexes.

5.3.2 Offspring events

Each event in the sequence can generate offspring events from multiple generations. The productivity of the j^{th} event of the sequence is spatio-temporally invariant since the modified Omori law integrates to K for infinite time and the spatial density adds

up to 1 by construction:

$$\sum_{i=1}^{\zeta} s(x_i - x_j, y_i - x_j) = 1$$

and

$$\int_0^{\infty} g(z)dz = K.$$

Again as described in Section 2.2.2, we restrict the average productivity to be smaller than 1 event i.e. $K < 1$. We then sample the proposed lags in space and time. However, the support of the spatial measure $s(\cdot)$ as of Equation 5.3 is discrete. Further, we allow for multiple instances from one camera which can cause difficulties in the evaluation scheme due to occurrence of multiple zero lag spatial realisations.

After obtaining the offspring set of the uncaused population of events, we repeat the procedure for every subsequent generation of observations while the offspring set is non-empty.

5.4 Spatial Poisson Mixture process

The effectiveness of the proposed self-exciting construction can be evaluated by comparing it to a nested model without self-excitation. This interesting simplification of the SECR ETAS model arises by considering all events as uncaused. Then this alternative model intensity coincides with the uncaused events' intensity as of Equation 5.4:

$$\lambda_j^{mp}(t_i, x_i, y_i) = \mu_0 \times \phi(x_i, y_i), \tag{5.13}$$

where $\phi(\cdot)$ is the same function as the one introduced for the full model as of Equation 5.5. This Poisson mixture model can be obtained from the SECR ETAS model by removing the offspring kernel, which is achieved by setting either $K = 0$ or $p = 1$ as of Equations 5.1 and 5.2. This model resembles more closely the standard SECR model specified in the literature [Dorazio and Karanth, 2017, Borchers and Marques, 2017].

The detection probability again remains unchanged from the one introduced in Equation 5.8 since the components related to it are the same. The log-likelihood function

with respect to all observed animals across all placed cameras of this process is:

$$\ell^{mp}(\theta;\mathscr{H}_T) = \sum_{j=1}^{m}\left(\sum_{i=1}^{n_j}\log(\lambda_j^{mp}(t_i,x_i,y_i|\mathscr{H}_T^{(j)},\theta_j)) - \int\sum_{s=1}^{\zeta}\lambda_j^{mp}(z,x_s,y_s|\mathscr{H}_T^{(j)},\theta_j)dz\right) + (M-m)log\left(\frac{1}{A}\int_A \exp(-\mu_0 T\phi_c(\boldsymbol{h}))d\boldsymbol{h}\right) + log\binom{M}{m}. \tag{5.14}$$

5.5 Bayesian methods for SECR

In this Section we introduce the specific latent variable methods employed for the estimation of SECR-ETAS model. The main difference between this algorithm and all previously introduced is its multi-dimensional structure. We merge individual branchings to facilitate the MCMC sampler needs. All introduced concepts rely on the strong assumption of independence of animals' behaviour.

5.5.1 Sampling a branching structure

The first step of the latent variable MCMC sampler is to sample a branching structure based on the previous iteration parameter set θ and space density $\phi(\cdot)$. Similarly to methods for obtaining a branching structure in the previous two Chapters (see Sections 3.5.2 and 4.6.1), a method for sampling a realisation of the underlying branching structure of the SECR ETAS can be recovered with respect to a given parameter set θ.

Let us consider a new version of the vector $B^{(k)} = \{B_1^{(k)},...,B_{n_k}^{(k)}\}$ where $B_i^{(k)} \in \{0,1,...,n_{k-1}\}$ which now collects information regarding the parenthood of each event for the k^{th} animal. If $B_i^{(k)} = 0$, the i^{th} event is uncaused (it is immigrant), otherwise $B_i^{(k)} = j$ and we say that i^{th} event is caused by j^{th} event (or that j^{th} event is a parent of the i^{th}). The conditional posterior for each $B_i^{(k)}$ is independent of all other $B_\cdot^{(\cdot)}$and can be written as:

$$P(B_i^{(k)} = j|\mathscr{H}_T^{(k)},\theta) = \begin{cases} \dfrac{\mu(t_i,x_i,y_i)}{\lambda(t_i,x_i,y_i|\mathscr{H}_{t_i}^{(k)},\theta)} & \text{for } j=0 \\ \dfrac{g(t_i-t_j)s(x_i-x_j,y_i-y_j)}{\lambda(t_i,x_i,y_i|\mathscr{H}_{t_i}^{(k)},\theta)} & \text{for } j\neq 0 \end{cases}$$

This way we can obtain a branching structure for a single animal. However, since the full dataset $\mathscr{H}_T$ is a collection of events for which we have to update the parameters of interest, we have to develop a branching method for the full dataset. Since we already

assumed that all animals have similar behaviour and cannot influence the occurrence from one another, we can directly merge all of the individual branching structures into one $B = \{B^{(1)}, ..., B^{(m)}\}$. The pooled animal branching structure B asserts that only realisations from the same animal can excite one another. For a given dataset $\mathscr{H}_T$ and parameter set θ, we can sample a branching structure B based on which to evaluate the number of children of every event of the full catalogue across every animal. Thus, let us define the following variable that collects all children of an event associated with the j^{th} animal observation in the full catalogue $\mathscr{H}_T$:

$$S_j = \{t_i | B_i = j\}.$$

The above expression is very useful as it collects information that can be used for estimating the SECR ETAS likelihood as of Equation 5.12, based on a specific branching structure. In order to improve notation, let us define the operation $|S_j|$ that returns the number of events in S_j. This construction allows the recovery of the causality with respect to every unique animal. For example, the set of uncaused events for the k^{th} animal is:

$$S_0^{(k)} = \{t_i | B_i^{(k)} = 0\},$$

where $S_0 = \{S_0^{(1)}, ..., S_0^{(m)}\}$. Based on the branching representation of the process we can modify the likelihood of a SECR ETAS model as of Equation 5.12 and further incorporate the point process representation of the undetected animals likelihood as of Equation 5.11 with respect to a given branching structure B. Using the likelihood as a sampling distribution we obtain the following full model posterior function:

$$\begin{aligned} P(\mathscr{H}_T|\theta, B) \propto & \pi(\theta) \binom{M}{m} e^{-T(M-m)\mu_0} \\ & \mu_0^{|S_0|} \prod_{j=1}^{m} P(\boldsymbol{h_j}) e^{-\mu_0 T \phi_c(\boldsymbol{h_j})} \prod_{t_i \in S_0} \phi(t_i, x_i, y_i) \\ & \prod_{j=1}^{n} \left(e^{-K\left(1 - \frac{c^{p-1}}{(t_n - t_i + c)^{p-1}}\right)} K^{|S_j|} \right) \\ & \prod_{j=1}^{n} \prod_{t_i \in S_j} s(x_i, y_i). \end{aligned} \tag{5.15}$$

5.5.2 Parameter updates

In this Section we illustrate the exact mechanism employed for updating all model parameters based on the already sampled full branching structure. We begin by sampling a branching for all observations for each animal. Then we update each of the model parameters in groups that is governed by their exact usage based on the full branching structure B. We iterate this procedure until we obtain the required number of parameter samples. More details related to the latent variable formulation are present in Sections 2.4.3, 3.5 and 4.6. Further implementation details, with respect to the Applications that we discuss in this Chapter, are present in Section 5.6.1.

Update the value of M

Among all model parameters only the overall number of animals M is independent from the full branching structure B. Its posterior distribution, under uniform prior, is

$$\begin{aligned} P(M|\mathscr{H}_T, \theta, B) &\propto \pi(M)\binom{M}{m}\left(\frac{1}{A}\int_A \exp(-\mu_0 T\phi_c(\boldsymbol{h}))d\boldsymbol{h}\right)^{(M-m)} \\ &= \binom{M}{m}\Big(1 - P(S=1)\Big)^{(M-m)}. \end{aligned} \tag{5.16}$$

This expression suggests a Gibbs update for the parameter M is equivalent to sampling the number of failures required to obtain m successes with success probability of $P(S=1)$. This is a Negative Binomial distribution, where

$$P(M-m|\mathscr{H}_T, \theta, B) \propto NB\Big(m, P(S=1)\Big).$$

However, large standard deviation of the samples in M can cause a disruption of the MCMC sampler on real catalogues. For example, large values for M directly reduce the values of μ_0 which can very quickly reach floating-point imprecision. For this reason, we employ a more conservative Metropolis-Hastings sampler based on Equation 5.16.

Update the value of μ_0

We aim to update the immigrant events' intensity parameter μ_0 based on the obtained branching structure. Using Equation 5.15 we can observe that μ_0 only enters the pos-

terior in conjunction with the background process S_0, hence:

$$P(\mu_0|\mathscr{H}_T,\theta,B) \propto \pi(\mu_0)\mu_0^{|S_0|}e^{-\mu_0 T\left[(M-m)\mu_0+\sum_{j=1}^{m}\phi_c(\boldsymbol{h_j})\right]}.$$

This conditional distribution is the same that would be obtained based on intensity function μ of a homogenous Poisson process on $[0,T]$, with event times S_0. In this case, the Gamma distribution is the conjugate prior: $\pi_{\mu_0} = Ga(\alpha_{\mu_0},\beta_{\mu_0})$. The full conditional distribution is then $P(\mu_0|\mathscr{H}_T,\theta,B) = Ga(\alpha_{\mu_0}+|S_0|,\beta_{\mu_0}+T\left[(M-m)\mu_0+\sum_{j=1}^{m}\phi_c(\boldsymbol{h_j})\right])$ which can be sampled from directly based on a Gibbs sampler.

Update the values of γ

There is no conjugate prior and posterior combination for the update of γ. Hence, we will use Metropolis-Hastings updates. Of course, these updates also depend on the current values and all means $\boldsymbol{h}.$ and of the covariance matrix Σ. The full conditional distribution is:

$$P(\gamma|\mathscr{H}_T,\theta,B) \propto \pi(\gamma)e^{-\mu_0 T\sum_{j=1}^{m}\phi_c(\boldsymbol{h_j})}\prod_{j=1}^{m}\prod_{t_i\in S_0^{(j)}}\phi(x_i,y_i|\boldsymbol{h_j},\boldsymbol{\theta}).$$

Sampling γ is done based on a Metropolis-Hastings MCMC sampler.

Update the value of $\boldsymbol{\Sigma}$.

The covariance matrix of data with zero mean has a conjugate update. However, in our model Σ is part of a discretised mixture model described by Equation 5.5. Thus, the full conditional of this covariance matrix is:

$$P(\Sigma|\mathscr{H}_T,\theta,B) \propto \pi(\Sigma)e^{-\mu_0 T\sum_{j=1}^{m}\phi_c(\boldsymbol{h_j})}\prod_{j=1}^{m}\prod_{t_i\in S_0^{(j)}}\phi(x_i,y_i|\boldsymbol{h_j},\boldsymbol{\theta}),$$

which directly resembles the full conditional for γ. However, the proposal for Σ can be obtained based on the approximately conjugate update with respect to $\gamma = 1$ since the uncaused spatial density $\phi(x,y)$ is reduced to bivariate normal distribution with mean $\boldsymbol{h}$ and covariance Σ. The corresponding prior is the Inverse-Wishart distribution ($IW(\nu,\Psi)$) with full conditional distribution being the following Inverse-Wishart

distribution:

$$IW\Big(n+\nu, \Psi + \sum_{i \in S_0}((x_i, y_i) - \boldsymbol{\mu})(\ (x_i, y_i) - \boldsymbol{\mu})^T\Big). \tag{5.17}$$

The data used for this estimation is obtained by subtracting the current estimate of the home-range centre for each animal $\boldsymbol{h}$ to obtain data with zero expectation. This suggests that the mean parameter $\boldsymbol{\mu}$ in the full conditional distribution above, will be also zero. The value of Ψ is set to be the covariance matrix of the zero mean transformation of the data that comprises the spatial realisation of all animals.

Update the value of $\boldsymbol{h}.$ for all animals.

Similar to the previous case, there is no conjugate prior distribution for the home-range centres $\{\boldsymbol{h_j} : j \in 1, ..., m\}$ updates. The full conditional distribution is:

$$P(\boldsymbol{h}.|\mathscr{H}_T, \theta, B) \propto \pi(\boldsymbol{h}.)e^{-\mu_0 T \phi_c(\boldsymbol{h}.)} \prod_{t_i \in S_0^{(\cdot)}} \phi(x_i, y_i|\boldsymbol{h}., \boldsymbol{\theta}).$$

However, similarly to the previous case, there is an approximate conjugate update for $\gamma = 1$. We use this approximation as a proposal distribution as part of a Metropolis-Hastings sampler. For a known covariance matrix Σ, we set the prior distribution of the j^{th} animal with a home-range centre $\boldsymbol{h_j}$ to

$$\boldsymbol{h_j} \sim N(\boldsymbol{\mu_j}, \boldsymbol{\Sigma}_0) \text{ where } \boldsymbol{\mu_j} = \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} \text{ and } \boldsymbol{\Sigma}_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \epsilon_j,$$

where $\bar{x}$ and $\bar{y}$ correspond to the mean value of the spatial occurrences for this animal in x and y direction and ϵ_j is the step of the updates. Then the approximate full conditional distribution that we will use as a proposal in our Metropolis-Hastings updates is:

$$N\Big(\left(\Sigma_0^{-1} + n_j\Sigma^{-1}\right)^{-1}\left(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_j + n_j\Sigma^{-1}\bar{\mathbf{x}}\right), \left(\Sigma_0^{-1} + n_j\Sigma^{-1}\right)^{-1}\left(\Sigma_0^{-1} + n_j\Sigma^{-1}\right)^{-1}\Big), \tag{5.18}$$

where n_j is the number of observations in the catalogue associated with the j^{th} animal. Σ_0 is set to be the covariance matrix of the zero mean transformation of the data that comprises the spatial realisation of all animals.

Update the values of d.

The offspring kernel space parameter d is shared across both x and y dimension. Then the data from each dimension is following a univariate normal distribution with zero mean and variance d. This allows us to gather the lags in space into one vector. Let us define $\Delta x = \{x_j - x_i\} : i \in \{1, ..., n\}, j \in I(S_i)$, where $I(S.)$ returns the indexes of all elements in $S.$; the same patterns with respect to the y dimension is notated analogically as Δy. Then $\Delta Y = \Delta x \cup \Delta y$ is a column vector collecting all spatial lags. We set the prior distribution to be Inverse-Gamma(α_d, β_d) which has a posterior Inverse-Gamma distribution. The lags in space have by definition zero mean. Hence, the full conditional of d is the following:

$$P(d|\mathscr{H}_T, \theta, B) \propto \text{InvGamma}\bigg(\alpha_d + \frac{|\Delta Y|}{2},\, \beta_d + \frac{(\Delta Y)'(\Delta Y)}{2}\bigg),$$

where $|\Delta Y|$ returns the number of elements in ΔY.

Proof. If the prior is set to $d \sim \text{InvGamma}(\alpha_d, \beta_d)$, then $P(d|\alpha_d, \beta_d) = \frac{\beta_d^{\alpha_d}}{\Gamma(\alpha)} d^{-\alpha-1} \exp(\frac{-\beta_d}{d})$. The likelihood of $\Delta Y \sim N(0, d)$ is:

$$P(\Delta Y|d) = (d2\pi)^{-\frac{|\Delta Y|}{2}} \exp\Big(-\frac{1}{2d}(\Delta Y)'(\Delta Y)\Big).$$

Then the full conditional distribution is:

$$\begin{aligned} P(d|\mathscr{H}_T, \theta, B) \propto & \frac{\beta_d^{\alpha_d}}{\Gamma(\alpha)} d^{-\alpha-1} \exp\Big(\frac{-\beta_d}{d}\Big)(d2\pi)^{-\frac{|\Delta Y|}{2}} \exp\Big(-\frac{1}{2d}(\Delta Y)'(\Delta Y)\Big) \\ \propto & d^{-(\alpha+\frac{|\Delta Y|}{2})-1} \exp\Big(-\beta_d - \frac{1}{2d}(\Delta Y)'(\Delta Y)\Big) \\ \propto & \text{InvGamma}\bigg(\alpha_d + \frac{|\Delta Y|}{2},\, \beta_d + \frac{(\Delta Y)'(\Delta Y)}{2}\bigg), \end{aligned}$$

as required. □

Update the values of c, p and K

For newly proposed values of c, p and K we evaluate their suitability with respect to the full conditional distribution $P(c, p, K|\mathscr{H}_T, \theta, B)$. Again, based on Equation 5.15, we

can see that this full conditional distribution is given by:

$$P(c,p,K|\mathscr{H}_T,\theta,B) \propto \pi(c,p,K)\prod_{j=1}^{n}\left(e^{-K\left(1-\frac{c^{p-1}}{(t_n-t_i+c)^{p-1}}\right)}\prod_{t_i\in S_j}\frac{K}{(t_i-t_j+c)^p}\right).$$

Similarly, the parameter updates can be done using the basic random walk MCMC sampler (more advanced sampling techniques for univariate distributions could also be used to speed up computation). Additional restrictions could be implemented to guarantee that the obtained parameters are representing a finite catalogue by restricting $K < 1$ (see Section 2.2.2).

5.6 Applications

Fitting an univariate self-exciting model typically requires to have at least a few hundred events - there should be a sufficient number of uncaused events for the identification of μ_0 and each of them should on average produce some events that would further descend more events from multiple generations. In the context of the SECR ETAS model, the total number of observations across all animals should be also large enough to provide sufficient data for the estimation of the total animal population. However, all of these requirements are rather unrealistic given the available data granularity. We used two different datasets to illustrate the behaviour of SECR ETAS and compare it with the alternative model (Section 5.4). Neither of the two datasets provides occurrence patterns similar to the required one. However, SECR ETAS still delivers the best results according to the relevant goodness-of-fit measures that were previously introduced in Section 2.4.4. We strongly believe that placing more cameras in the detection region, combined with more detailed spatio-temporal data recording will greatly improve the model supremacy towards more conventional techniques.

5.6.1 MCMC tuning

The prior distribution choice can influence greatly the obtained MCMC samples. We combine our prior knowledge of the ETAS model behaviour with the context of the data to develop realistic MCMC framework [Ross, 2018a, Kolev and Ross, 2019, Dorazio and Karanth, 2017].

The MCMC sequences are with overall length of 10000 after 2000 samples burn-in and thinning of 20 units for each of the parameter sets $\{M\}$, $\{\gamma\}$, $\{\Sigma\}$, $\{\boldsymbol{h}.\}$ and $\{c,p,K\}$,

while μ_0 and d rely on conjugate updates. The branching structure was sampled from its conditional posterior at every 40 iterations of the MCMC algorithm. The MCMC algorithms for the Poisson mixture model that was introduced in Section 5.4 uses the same techniques as the one of the SECR ETAS model. However, in it all animals are considered uncaused and the missing parameters are fixed to the previously discussed values that remove the needed components from the SECR model.

We used a Uniform prior within reasonable bounds for $M \in [m, \infty)$, $c \in (0, 10)$, $p \in (1, 30)$, $K \in (0, 1)$ and $\gamma \in [0, 1]$, although more informative priors could be used if desired. We used as a proposal distribution for the Metropolis-Hastings updates a Normal distribution with standard deviation of 0.1 for $\{c, p, K\}$ and a Normal distribution with reduced standar deviation to 0.02 for γ. The latter was required due to the large influence of this parameter across the model fit. In its essence it controls the model fit between using zero parameters and $3 + 2 \times m$ parameters, where m is the number of detected animals. As such, a very small deviation can cause an enormous change in the value of the posterior distribution. The proposal for M, Σ and $\boldsymbol{h}_\cdot$ rely on the previously outline approximate results.

Tuning the prior distribution for μ_0 is essential to obtain reliable estimates of the underlying branching structure. The introduced relationship in Section 5.5.2 can be heavily influenced by either a single animal with large number of observations or multiple animals which appeared only for a short period of time across the temporal detection interval. For this reason, we could restrict the prior distribution in a relatively narrow neighbourhood. We chose to set $\alpha_{\mu_0} = 0.1$ and $\beta_{\mu_0} = 2$ across all models and all datasets that we address in this Section. The prior parameters for d are set to $\alpha_d = 2$ and $\beta_d = 0.5$.

5.6.2 Leopard data

We begin our study with the well-known Common Leopard (*Panthera pardus*) Dryad data [Goldberg et al., 2015a, Goldberg et al., 2015b]. This specie is present from sub-Saharan Africa to the Russian Far East. It also populates the islands of Sri Lanka and Java [Bailey, 1993, Uphyrkina et al., 2001]. The common Leopard is exposed to many treads similarly to other large carnivores. Many leopards were killed in Bhutan by human-wildlife conflicts [Wang and Macdonald, 2006, Sangay and Vernes, 2008]. However, the knowledge of leopard populations is limited despite the numerous encounters

with humans [Wang and Macdonald, 2009]. For all these reasons, the evaluation of the overall leopard population in Bhutan is needed.

The data were gathered during winter 2010–2011 (17-November-2010 to 15-February-2011) in Royal Manas National Park, Bhutan ($26^o47'31.27$"N, $90^o57'37.61$"E) and it contains 82 instances across 22 animals. 10 of the animals were captured only once, 2 of them - twice, 3 of them - three times, 2 of them four times. The remaining five animals have 5, 7, 8, 15 and 16 observations. Exact time at which the animal had passed through the camera region is unknown since only the date of the capture is recorded. This limits considerably the study since multiple detections of an animal in a single day could be counted as a single occurrence [Foster and Harmsen, 2012].

The original study of these data [Goldberg et al., 2015b] made conclusions with respect to the overall population within the whole park despite the limited number of cameras and small study region. The park overall area is $1057km^2$ which increases to $1551km^2$ by allowing a $10km$ buffer. The study region area is $162km^2$ and it was estimated that a density of 10.0 animals/$100km^2$ (95% credibility interval: 6.25–15.93) despite the detection of 13.58 animals/$100km^2$ in the study region during the observation period.

The data temporal interval has an overall length of 91 days. The large number of animals that were captured only once combined with large discrepancy between the number of captures of the remaining animals and inaccurate temporal information affect negatively the performance of the SECR ETAS model. In our study we transformed the camera locations to be distributed with zero mean and unit variance in each dimension. The overall spatial distribution of the camera locations after the transformation is illustrated on Figure 5.1. The study placed 29 camera, 25 from which captured animals. The MCMC parameter density for the number of uncaused events, population size and detection probability is shown on Figure 5.2. All other parameter densities are shown on Figure 5.3. We obtained 55 immigrant events on average. This is a very small number given that we observe data across 22 animals as the first observation from every animals is considered an uncaused event. The SECR ETAS model estimated that we most likely detect 34.5% of all animals, while the Poisson mixture model proposed a detection probability of 55.9%. The population size MAP estimate was estimate to be 47 animals, with 95% credible interval of (35, 77), and 37 animals, with 95% credible interval of (29, 53), with respect to the SECR ETAS and Poisson mixture respectively.

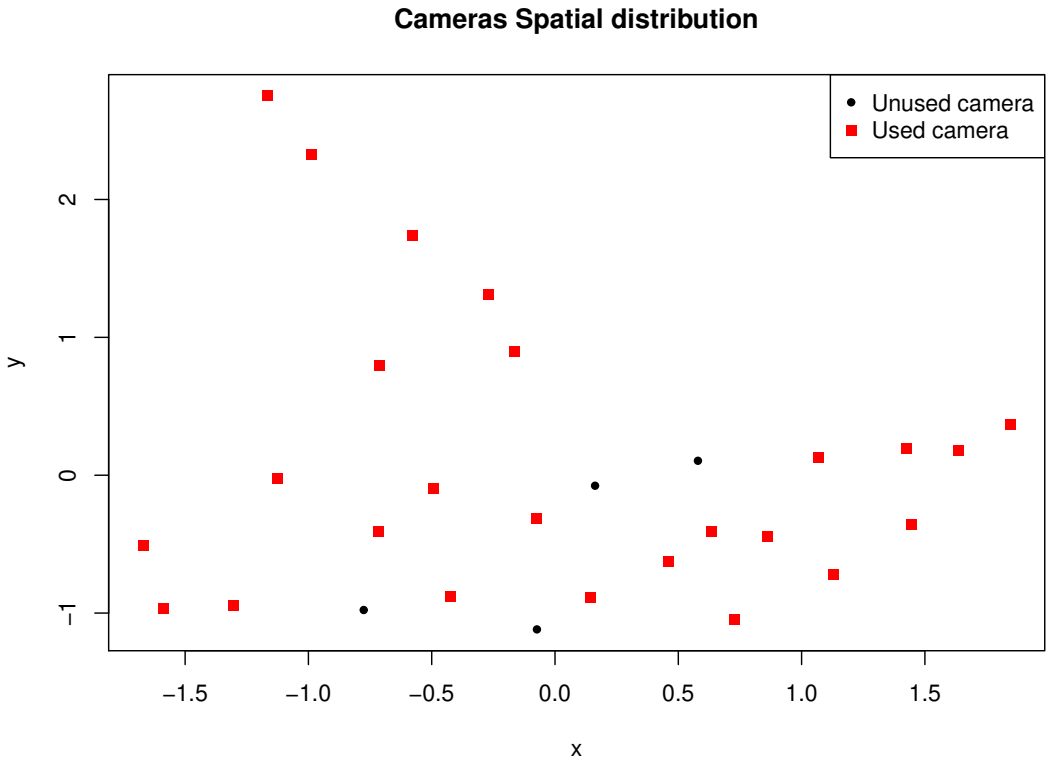

Figure 5.1: Leopard dataset camera distribution. Unused cameras are those that did not detect an animal during the study.

The differences between the two models is even more evident with respect to the formal model diagnostics that are presented on Table 5.2. SECR ETAS provides superior fit according to all them.

	SECR ETAS	Poisson Mixture
Log-likelihood	-448.08	-489.66
Number of parameters	54	50
AIC	1004.17	1079.33
BIC	567.06	599.83
DIC	951.81	993.12

Table 5.1: Goodness-of-fit Summary for the leopard Dataset. Lower values of the AIC, BIC and DIC indicate superior fit.

5.6.3 Tiger data

These data were based on the tiger (*Panthera tigris*) population within The Nagarahole tiger Reserve of Karnataka, India ($12^o01'53.7$"N, $76^o07'08.6$"E). The total area of the reserve is $860km^2$ which increases to $1130km^2$ under a $10km$ wide buffer zone inclusion.

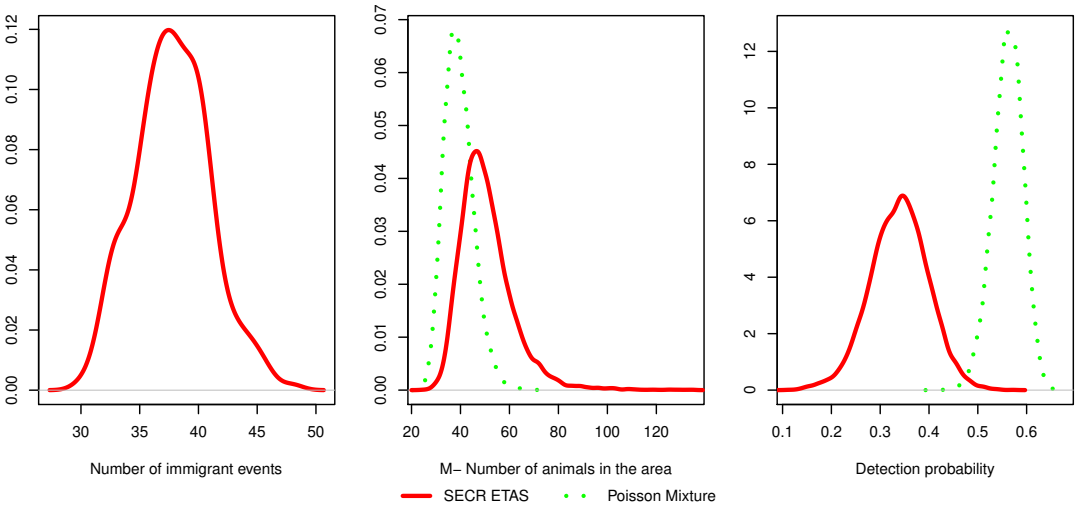

Figure 5.2: MCMC parameters density for the number of uncaused (immigrant) events, population size (M) and detection probability for the leopard dataset with respect to the introduced models

The main application for these data so far was the illustration of the continuous time SECR models [Dorazio and Karanth, 2017]. In their work was estimated expected tiger density of 11.3 animals/$100km^2$ with 95% credible interval of $(9.1, 13.9)$. The survey period was 45 days in winter 2014-2015 (26-November-14 to 13-January-15) and recorded continuously animal movement. However, the data that we used approximates time to the nearest minute. Individual animals were identified based on their unique stripe patterns [Karanth, 1995]. The data comprises 355 observations across 127 cameras with respect to 86 unique animals. However, the experiment has placed 162 cameras in the area of interest. This clearly indicates that neither the cameras are covering the entire area nor the tigers are uniformly exploring the target area of interest. The camera locations were re-scaled to have zero mean and unit standard deviation in each dimension. The obtained realisation is shown on Figure 5.4 where with black circles are indicated the 127 cameras on which the animals were not captured while with red squares are illustrated those that captured animals.

Similar to the previous dataset, the number of animals that were detected only once comprises the large proportion of the data - 32 out of 86 animals. The number of animals observed twice is 13; three times - 11; four times - 7; five and seven times - 4; eight, nine, ten, twelve, thirteen and twenty times - 2; fifteen and twenty-three times - 1. This catalogue should provide a better differentiation between the SECR ETAS model and its alternatives due to the opportunity for development of branching structures for each animal.

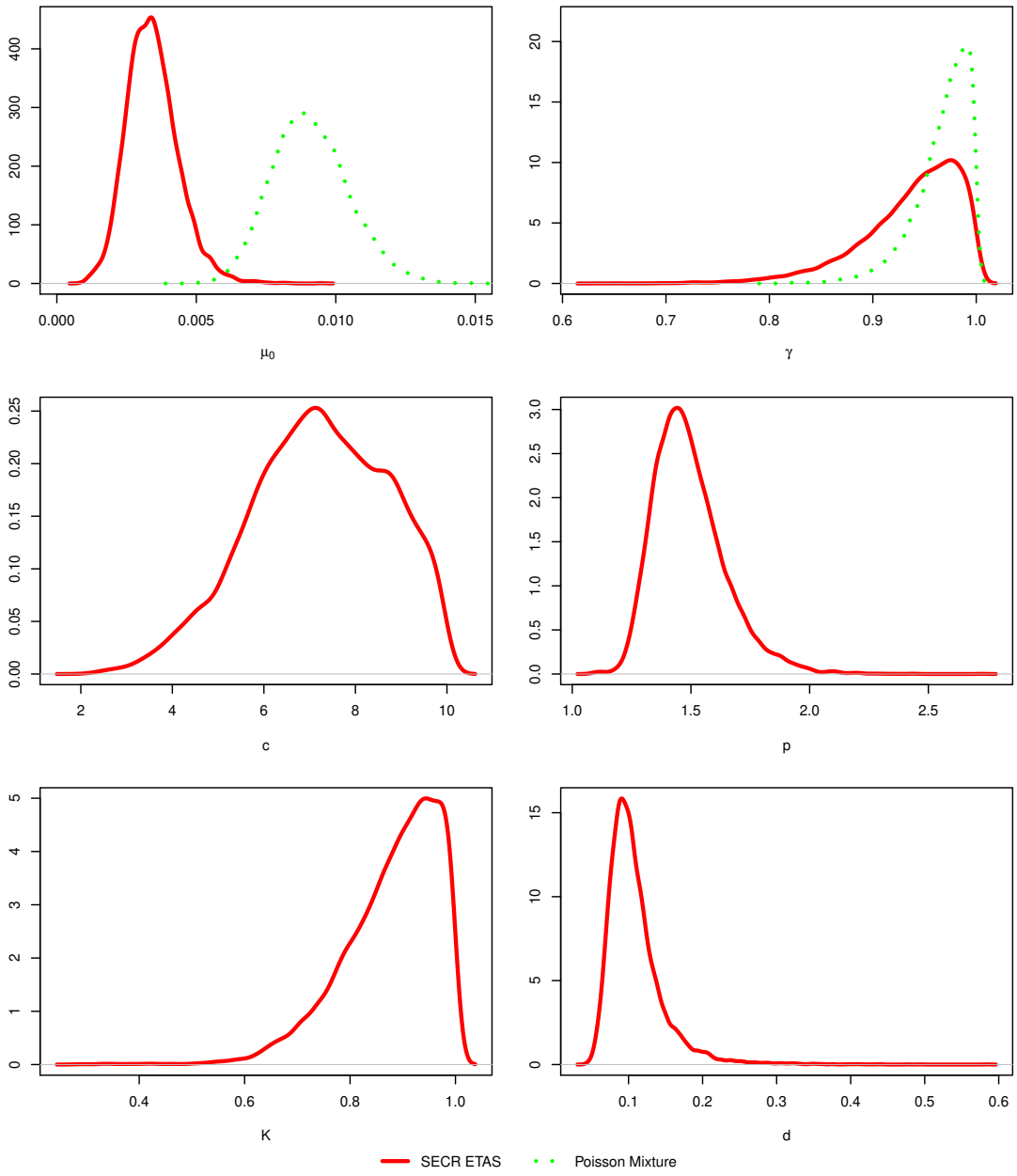

Figure 5.3: MCMC parameters density for the leopard dataset with respect to the introduced models. Parameters μ_0 and γ are shared across the two discussed models while c, p, K and d. The difference between μ_0 is primarily influenced by the value of K.

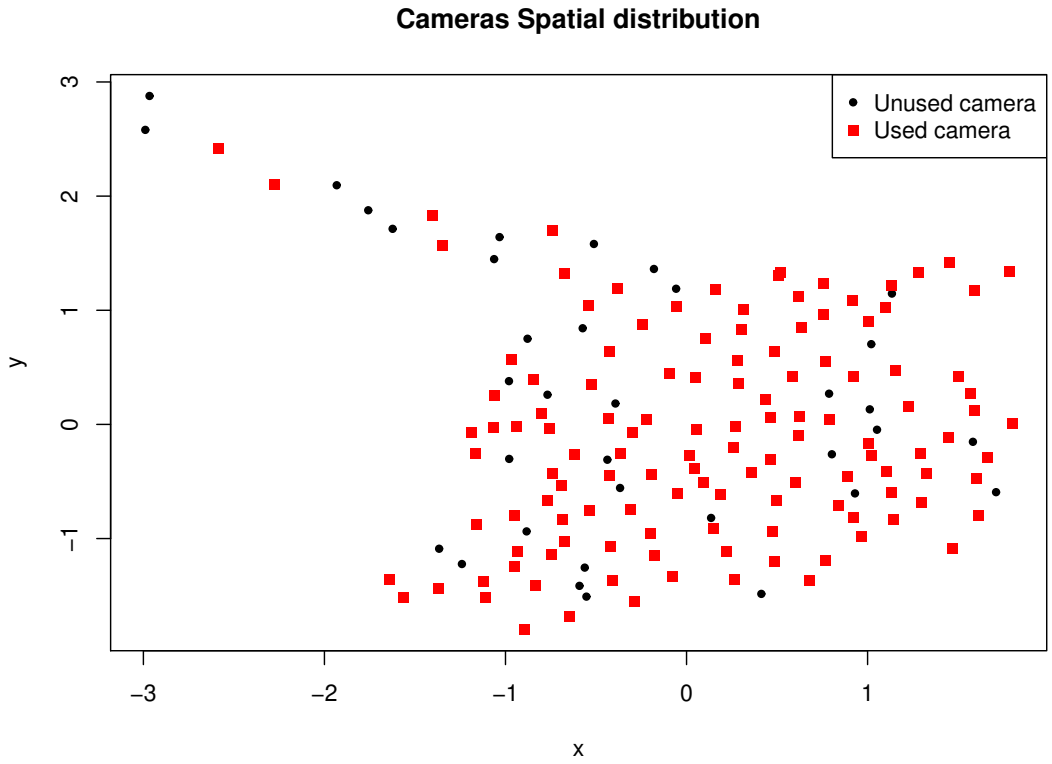

Figure 5.4: Tiger dataset camera distribution. Unused cameras are those that did not detect an animal during the study.

We converted the arrival times to hours which led us to obtain realisations with upper bound of 1130.67. The Poisson mixture and SECR ETAS were fit to data based on 10000 MCMC updates starting with the ML estimates. The MCMC parameter density for the number of uncaused events, population size and detection probability is shown on Figure 5.5. All other parameter densities are shown on Figure 5.6. The behaviour of the latent variable MCMC method can be analysed with respect to the number of uncaused events proposed by the branching simulation method through the MCMC updates. The mean number of uncaused events is 168. The SECR ETAS model suggests that we most likely detect 24.5% of all animals while the Poisson mixture model approximates it to 43.4%. The total population MAP estimated is approximated to 342 animals, with 95% credible interval of (260, 471), and 199 animals, with 95% credible interval of (168,235), with respect to the two models.

In order to comprehend the differentiation between the two models we have to investigate the formal model fit diagnostics that are presented on Table 5.2. SECR ETAS provides superior fit according to all diagnostic tests despite the inclusion of only 4 addi-

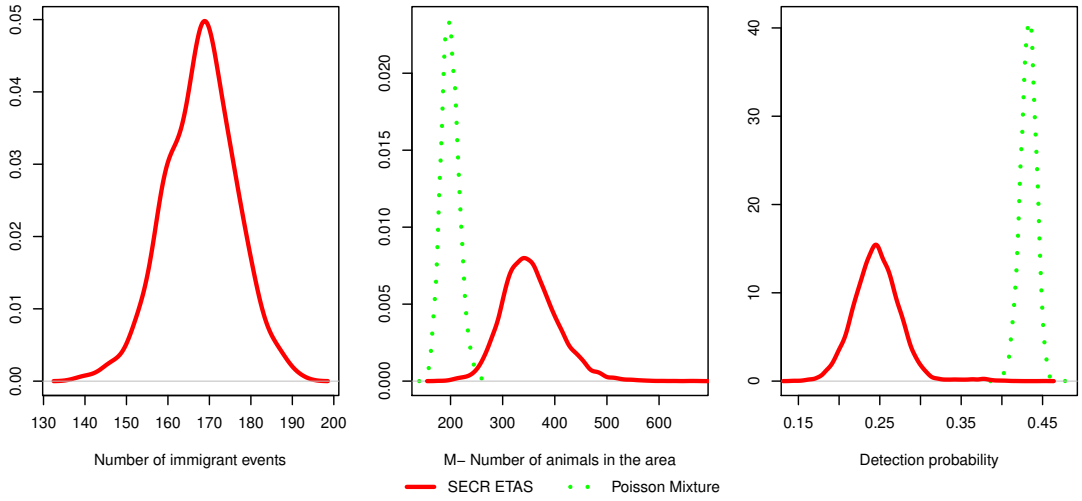

Figure 5.5: MCMC parameters density for the number of uncaused (immigrant) events, population size (M) and detection probability for the tiger dataset with respect to the introduced models

tional parameters. For the Poisson mixture we further evaluate γ, the three parameters of the common, full covariance matrix Σ and the two parameters of the home-range centre $\boldsymbol{h}.$ for every unique animal. SECR ETAS further introduces 4 parameters - c, p, K and d. Although the self-excitation component contributes with only a smaller number of additional parameters to estimate, the data fit has considerably improved not only the likelihood function but also the other metrics.

	SECR ETAS	Poisson Mixture
Log-likelihood	-2952.25	-3003.93
Number of parameters	182	178
AIC	6268.49	6367.69
BIC	3486.61	3528.42
DIC	5960.55	6063.46

Table 5.2: Goodness-of-fit Summary for the tiger dataset. Lower values of the AIC, BIC and DIC indicate superior fit.

5.7 Conclusion

In this Chapter, we introduced a novel multivariate ETAS model with discrete spatial support that addresses animal movement. We further specified an alternative model that simplify it to more conventional inhomogeneous Poisson processes. All goodness-of-fit metrics indicate that SECR ETAS addresses both analysed datasets better than

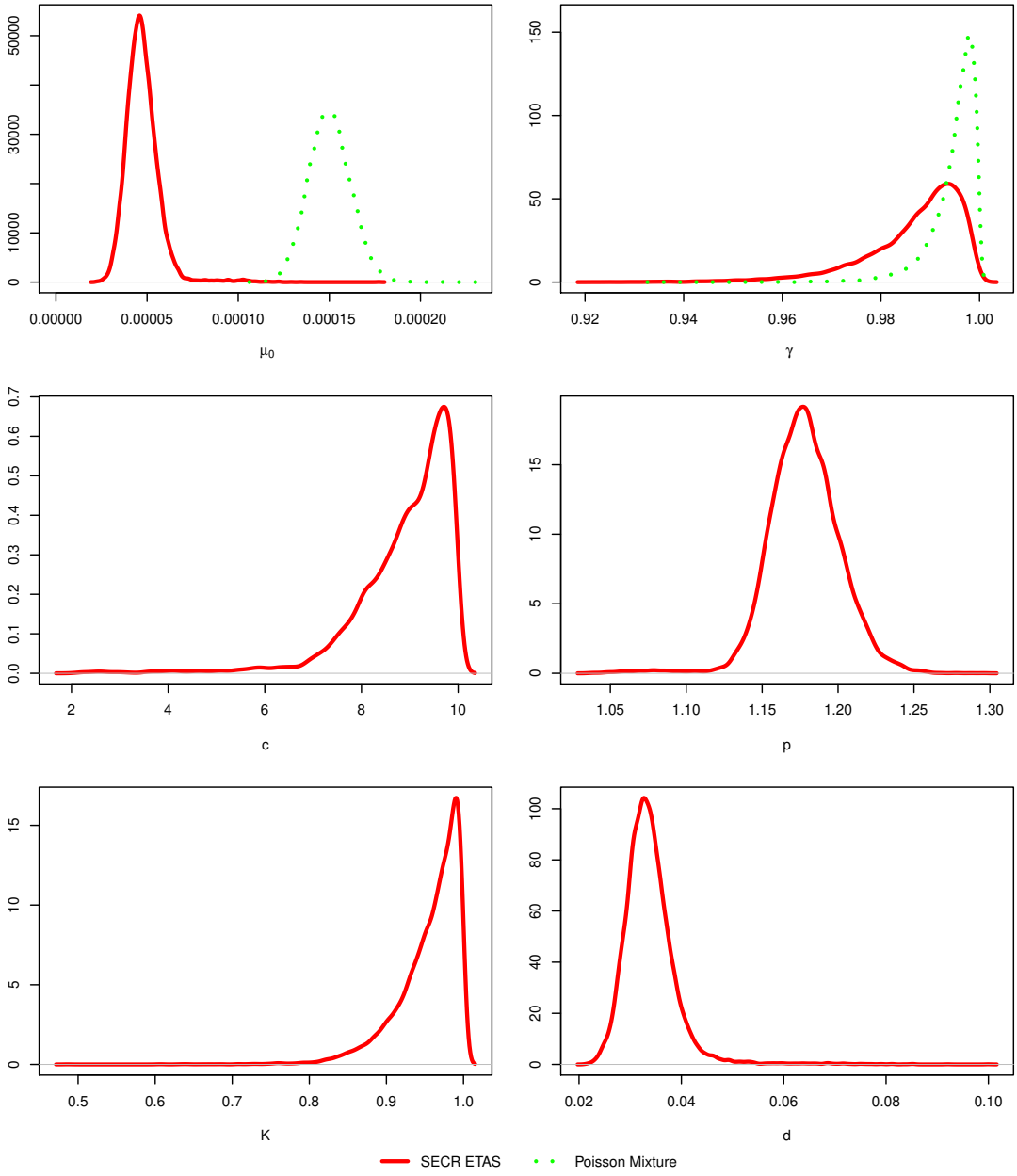

Figure 5.6: MCMC parameters density for the tiger dataset with respect to the introduced models. Parameters μ_0 and γ are shared across the two discussed models while c, p, K and d. The difference between μ_0 is primarily influenced by the value of K.

the alternative model. This further suggests that self-excitation phenomenon adds value in modelling animal movement.

Currently, we examined a uniform prior distribution for the home-range centres. A more involved approach will increase the insight with respect to the proportion of animals that were detected. The introduced model is directly applicable in continuous space set up in which we can track continuously animal movement across the region of interest. This way, we could evaluate the specific trip characteristics such as frequency and length of trips; habitat shape and spread; animal to animal influence. Such pattern can be obtained if we are solely interested in whether an animal will appear in specific areas that are fully covered by cameras. This will further change the actual spatial observation of the animals to their exact location instead of being approximated to the camera location. Additional alternatives of the offspring kernels can be analysed to fit the data better.

Chapter 6

Conclusion

This thesis studied in depth a specific class of point processes that provides a tractable methodology for linking inter event influence. Our work began with critical analysis of the literature associated with point processes; temporal/spatial analysis; estimation and optimisation methods. Then we allocated specific research questions. Our prime focus was on the ETAS model - a special form of the general Hawkes process that is predominantly used in seismology. Its inherent estimation difficulties placed an important emphasis in our work on point processes mechanisms, estimation algorithms and inferential results interpretation. However, all these issues left researching the ETAS model in greater depth rather undesired by the general research community which provides vast opportunities for fruitful research.

The novelty in our work consists of new structural forms of the general models, innovative estimation algorithms and some new areas for their application. The work in Chapter 3 was the most natural extension of the standard ETAS model, given the earthquake arrival time Poisson assumption challenge in literature in the last 20 years. Interestingly, the work on this project began prior to the discovery of the simplified alternative of B-SR-ETAS, namely RHawkes [Wheatley et al., 2016]. However, their work and the one of [Chen and Stindl, 2018] directly illustrated the crucial need of a latent variable-based Bayesian inference. The F-SR-ETAS model outlined an additional class of models that is still not illustrated in any other publicly available research article. In that Chapter we used two simplistic distributions to illustrate the benefits of SR-ETAS models over the standard ETAS. An interesting extension would be to substitute the functional form of the density component $f_w(w_t)$ as of Equation 3.3 that fits the

waiting time distribution, with a non-parametric kernel. This can be achieved (e.g.) by incorporating an univariate Dirichlet process over $f_w(w_t)$.

After completing the first project we switched our focus to the most challenging component of this thesis - proposing an interesting (and useful) spatial ETAS model that can be scaled by the latent variable approach. Prior to the development of Chapter 4 the Branching-based MCMC algorithm was solely defined for the temporal ETAS model [Ross, 2018a] and it was unclear what would be its direct application with respect to spatial kernels. The natural choice that we had was to define either Dirichlet or Gaussian process based density as part of our model. We decided to work with the first one to enhance the overall computation time of our algorithms. Gaussian process extension is an interesting comparison, however its application will not be feasible in a large earthquake sequence. We discovered that DP is not as good as we expected. In many cases KDE provided a superior performance. However, we also showed that in an out-of-sample study on real catalogues DP is the better choice. A natural extension of this project is to develop a library similar to the bayesianETAS R package [Ross, 2018a]. Further, this project can be combined with the previously defined SR-ETAS models for the development of even more powerful family of models for quantification of marked, spatio-temporal patterns.

The final project in this thesis outlines a novel application of the Hawkes process in an ecological context (Chapter 5). The rational behind is naively intuitive - if an animal is present in a specific time and space, then it is more likely to be spotted in the area nearby unless it decides to go 'home'. In this setup 'home' illustrates the uncaused events in the sequence while all other observations are caused by the previously detected animal location. We were challenged to incorporate the full Spatially explicit capture recapture framework within a multivariate ETAS models. We showed that SECR-ETAS model is indeed addressing better the paradigm of animal movement, compared with its non self-excited alternative. An open question is whether such a pattern is present across species. Further, a study with more granular data should outline the benefits of SECR-ETAS.

Bibliography

[Akaike, 1973] Akaike, H. (1973). Maximum likelihood identification of gaussian autoregressive moving average models. *Biometrika*, 60(2):255–265.

[Andersen et al., 2012] Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (2012). *Statistical models based on counting processes*. Springer Science & Business Media.

[Anderson and Darling, 1954] Anderson, T. W. and Darling, D. A. (1954). A test of goodness of fit. *Journal of the American statistical association*, 49(268):765–769.

[Antoniak, 1974] Antoniak, C. E. (1974). Mixtures of dirichlet processes with applications to Bayesian nonparametric problems. *The annals of statistics*, 2(6):1152–1174.

[Bacry et al., 2015] Bacry, E., Mastromatteo, I., and Muzy, J.-F. (2015). Hawkes processes in finance. *Market Microstructure and Liquidity*, 1(01):1550005.

[Baddeley et al., 2005] Baddeley, A., Turner, R., Møller, J., and Hazelton, M. (2005). Residual analysis for spatial point processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(5):617–666.

[Bailey, 1993] Bailey, T. N. (1993). *The African leopard: ecology and behavior of a solitary felid*. Columbia University Press.

[Barbour et al., 2013] Barbour, A. B., Ponciano, J. M., and Lorenzen, K. (2013). Apparent survival estimation from continuous mark–recapture/resighting data. *Methods in Ecology and Evolution*, 4(9):846–853.

[Borchers et al., 2014] Borchers, D., Distiller, G., Foster, R., Harmsen, B., and Milazzo, L. (2014). Continuous-time spatially explicit capture–recapture models, with an application to a jaguar camera-trap survey. *Methods in Ecology and Evolution*, 5(7):656–665.

[Borchers and Marques, 2017] Borchers, D. L. and Marques, T. A. (2017). From distance sampling to spatial capture–recapture. *AStA Advances in Statistical Analysis*, 101(4):475–494.

[Box and Pierce, 1970] Box, G. E. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, 65(332):1509–1526.

[Bray and Schoenberg, 2013] Bray, A. and Schoenberg, F. P. (2013). Assessment of point process models for earthquake forecasting. *Statistical science*, 28(4):510–520.

[Brillinger, 1993] Brillinger, D. R. (1993). Earthquake risk and insurance. *Environmetrics*, 4(1):1–21.

[Brown et al., 2002] 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.

[Buckland et al., 2016] Buckland, S. T., Oedekoven, C. S., and Borchers, D. L. (2016). Model-based distance sampling. *Journal of Agricultural, Biological, and Environmental Statistics*, 21(1):58–75.

[Chakravarti and Laha, 1967] Chakravarti, I. M. and Laha, R. G. (1967). Handbook of methods of applied statistics. In *Handbook of methods of applied statistics*. John Wiley & Sons.

[Chavez-Demoulin et al., 2005] 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.

[Chen et al., 2013] Chen, C.-H., Wang, J.-P., Wu, Y.-M., Chan, C.-H., and Chang, C.-H. (2013). A study of earthquake inter-occurrence times distribution models in taiwan. *Natural hazards*, 69(3):1335–1350.

[Chen and Stindl, 2018] Chen, F. and Stindl, T. (2018). Direct likelihood evaluation for the renewal Hawkes process. *Journal of Computational and Graphical Statistics*, 27(1):119–131.

[Chib and Greenberg, 1995] Chib, S. and Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The american statistician*, 49(4):327–335.

[Chornoboy et al., 1988] Chornoboy, E., Schramm, L., and Karr, A. (1988). Maximum likelihood identification of neural point process systems. *Biological cybernetics*, 59(4-5):265–275.

[Cox, 1962] Cox, D. R. (1962). *Renewal theory*. Methuen.

[Cox and Hinkley, 1974] Cox, D. R. and Hinkley, D. V. (1974). *Theoretical statistics*. Chapman and Hall/CRC.

[Cox and Isham, 1980] Cox, D. R. and Isham, V. (1980). *Point processes*, volume 12. CRC Press.

[Crane and Sornette, 2008] Crane, R. and Sornette, D. (2008). Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences*, 105(41):15649–15653.

[Crowley et al., 2013] Crowley, H., Pinho, R., Pagani, M., and Keller, N. (2013). Assessing global earthquake risks: the global earthquake model (gem) initiative. In *Handbook of seismic risk analysis and management of civil infrastructure systems*, pages 815–838. Elsevier.

[Daley and Vere-Jones, 2003] Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods, Second Edition*. New York). Springer-Verlag, New York,.

[Daley and Vere-Jones, 2007] Daley, D. J. and Vere-Jones, D. (2007). *An Introduction to the Theory of Point Processes: Volume II: General theory and structure*. Springer Science & Business Media.

[Dassios et al., 2013] Dassios, A., Zhao, H., et al. (2013). Exact simulation of Hawkes process with exponentially decaying intensity. *Electronic Communications in Probability*, 18.

[Dirac, 1947] Dirac, P. (1947). The principles of quantum mechanics (clarendon, oxford).

[Dorazio and Karanth, 2017] Dorazio, R. M. and Karanth, K. U. (2017). A hierarchical model for estimating the spatial distribution and abundance of animals detected by continuous-time recorders. *PloS one*, 12(5):e0176966.

[Ebrahimian et al., 2013] Ebrahimian, H., Jalayer, F., Asprone, D., Lombardi, A. M., Marzocchi, W., Prota, A., and Manfredi, G. (2013). Adaptive daily forecasting of seismic aftershock hazard. *Bulletin of the Seismological Society of America*, 104(1):145–161.

[Ellsworth et al., 1999] Ellsworth, W. L., Matthews, M. V., Nadeau, R. M., Nishenko, S. P., Reasenberg, P. A., and Simpson, R. W. (1999). A physically-based earthquake recurrence model for estimation of long-term earthquake probabilities. *US Geol. Surv. Open-File Rept. 99*, 522:23.

[Embrechts et al., 2011] Embrechts, P., Liniger, T., and Lin, L. (2011). Multivariate Hawkes processes: an application to financial data. *Journal of Applied Probability*, 48(A):367–378.

[Engle and Russell, 1998] Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, 66(5):1127–1162.

[Errais et al., 2010] Errais, E., Giesecke, K., and Goldberg, L. R. (2010). Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics*, 1(1):642–665.

[Faenza et al., 2010] Faenza, L., Meletti, C., and Sandri, L. (2010). Bayesian inference on earthquake size distribution: a case study in italy. *Bulletin of the Seismological Society of America*, 100(1):349–363.

[Ferguson, 1973] Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The annals of statistics*, 1(2):209–230.

[Filimonov and Sornette, 2015] Filimonov, V. and Sornette, D. (2015). Apparent criticality and calibration issues in the Hawkes self-excited point process model: application to high-frequency financial data. *Quantitative Finance*, 15(8):1293–1314.

[Foster and Harmsen, 2012] Foster, R. J. and Harmsen, B. J. (2012). A critique of density estimation from camera-trap data. *The Journal of wildlife management*, 76(2):224–236.

[Fox et al., 2016] Fox, E. W., Schoenberg, F. P., Gordon, J. S., et al. (2016). Spatially inhomogeneous background rate estimators and uncertainty quantification for nonparametric Hawkes point process models of earthquake occurrences. *The Annals of Applied Statistics*, 10(3):1725–1756.

[Gelman et al., 2014] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian data analysis*, volume 2. CRC press Boca Raton, FL.

[Ghahramani, 2005] Ghahramani, Z. (2005). Lecture notes in Unsupervised Learning.

[Glickman and Van Dyk, 2007] Glickman, M. E. and Van Dyk, D. A. (2007). Basic Bayesian methods. In *Topics in Biostatistics*, volume 404, pages 319–338. Springer.

[Goldberg et al., 2015a] Goldberg, J., Tempa, T., Norbu, N., Hebblewhite, M., Mills, L., Wangchuk, T., and Lukacs, P. (2015a). Data from: Examining temporal sample scale and model choice with spatial capture-recapture models in the common leopard panthera pardus.

[Goldberg et al., 2015b] Goldberg, J. F., Tempa, T., Norbu, N., Hebblewhite, M., Mills, L. S., Wangchuk, T. R., and Lukacs, P. (2015b). Examining temporal sample scale and model choice with spatial capture-recapture models in the common leopard panthera pardus. *PloS one*, 10(11):e0140757.

[González et al., 2016] González, J. A., Rodríguez-Cortés, F. J., Cronie, O., and Mateu, J. (2016). Spatio-temporal point process statistics: a review. *Spatial Statistics*, 18:505–544.

[Görür and Rasmussen, 2010] Görür, D. and Rasmussen, C. E. (2010). Dirichlet process gaussian mixture models: Choice of the base distribution. *Journal of Computer Science and Technology*, 25(4):653–664.

[Gourieroux et al., 1982] Gourieroux, C., Holly, A., and Monfort, A. (1982). Likelihood ratio test, wald test, and kuhn-tucker test in linear models with inequality constraints on the regression parameters. *Econometrica: journal of the Econometric Society*, pages 63–80.

[Guglielmi, 2017] Guglielmi, A. V. (2017). Omori's law: a note on the history of geophysics. *Physics-Uspekhi*, 60(3):319.

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

[Hamra et al., 2013] Hamra, G., MacLehose, R., and Richardson, D. (2013). Markov chain monte carlo: an introduction for epidemiologists. *International journal of epidemiology*, 42(2):627–634.

[Hardiman et al., 2013] Hardiman, S. J., Bercot, N., and Bouchaud, J.-P. (2013). Critical reflexivity in financial markets: a Hawkes process analysis. *The European Physical Journal B*, 86(10):442.

[Harte, 2012] Harte, D. (2012). Bias in fitting the ETAS model: A case study based on new zealand seismicity. *Geophysical Journal International*, 192(1):390–412.

[Harte et al., 2010] Harte, D. et al. (2010). Ptprocess: An r package for modelling marked point processes indexed by time. *Journal of Statistical Software*, 35(8):1–32.

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

[Hawkes and Oakes, 1974] Hawkes, A. G. and Oakes, D. (1974). A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11(3):493–503.

[Helmstetter and Sornette, 2002] Helmstetter, A. and Sornette, D. (2002). Subcritical and supercritical regimes in epidemic models of earthquake aftershocks. *Journal of Geophysical Research: Solid Earth*, 107(B10):ESE–10.

[Holschneider et al., 2012] Holschneider, M., Narteau, C., Shebalin, P., Peng, Z., and Schorlemmer, D. (2012). Bayesian analysis of the modified omori law. *Journal of Geophysical Research: Solid Earth*, 117(B6).

[Hyndman, 2014] Hyndman, R. J. (2014). Rob J Hyndman thoughts on the ljung-box test. `https://robjhyndman.com/hyndsight/ljung-box-test/`. Accessed: 2017-08-04.

[Imoto, 2001] Imoto, M. (2001). Application of the stress release model to the nankai earthquake sequence, southwest japan. *Tectonophysics*, 338(3):287–295.

[Isham and Westcott, 1979] Isham, V. and Westcott, M. (1979). A self-correcting point process. *Stochastic Processes and Their Applications*, 8(3):335–347.

[Islam et al., 1990] Islam, S., Entekhabi, D., Bras, R., and Rodriguez-Iturbe, I. (1990). Parameter estimation and sensitivity analysis for the modified bartlett-lewis rectangular pulses model of rainfall. *Journal of Geophysical Research: Atmospheres*, 95(D3):2093–2100.

[Johnson et al., 2010] Johnson, D. S., Laake, J. L., and Ver Hoef, J. M. (2010). A model-based approach for making ecological inference from distance sampling data. *Biometrics*, 66(1):310–318.

[Johnson et al., 2005] Johnson, N. L., Kemp, A. W., and Kotz, S. (2005). *Univariate discrete distributions*, volume 444. John Wiley & Sons.

[Kagan and Knopoff, 1984] Kagan, Y. and Knopoff, L. (1984). A stochastic model of earthquake occurrence. In *Proceedings of the Eighth International Conference on Earthquake Engineering*, volume 1, pages 295–302.

[Karanth, 1995] Karanth, K. U. (1995). Estimating tiger panthera tigris populations from camera-trap data using capture—recapture models. *Biological conservation*, 71(3):333–338.

[Kass and Raftery, 1995] Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association*, 90(430):773–795.

[Kolev and Ross, 2019] Kolev, A. A. and Ross, G. J. (2019). Inference for ETAS models with non-poissonian mainshock arrival times. *Statistics and Computing*, 29(5):915–931.

[Kron, 2002] Kron, W. (2002). Keynote lecture: Flood risk= hazard× exposure× vulnerability. *Flood defence*, pages 82–97.

[Kumazawa and Ogata, 2013] Kumazawa, T. and Ogata, Y. (2013). Quantitative description of induced seismic activity before and after the 2011 tohoku-oki earthquake by nonstationary ETAS models. *Journal of Geophysical Research: Solid Earth*, 118(12):6165–6182.

[Laio, 2004] 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).

[Lallouache and Challet, 2016] Lallouache, M. and Challet, D. (2016). The limits of statistical significance of Hawkes processes fitted to financial data. *Quantitative Finance*, 16(1):1–11.

[Lippiello et al., 2014] Lippiello, E., Giacco, F., De Arcangelis, L., Marzocchi, W., and Godano, C. (2014). Parameter estimation in the ETAS model: Approximations and novel methods. *Bulletin of the Seismological Society of America*, 104(2):985–994.

[Liu et al., 1998] Liu, J., Vere-Jones, D., Ma, L., Shi, Y.-L., and Zhuang, J.-C. (1998). The principle of coupled stress release model and its application. *Acta Seismologica Sinica*, 11(3):273–281.

[Ljung and Box, 1978] Ljung, G. M. and Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303.

[Lu et al., 1999] Lu, C., Harte, D., and Bebbington, M. (1999). A linked stress release model for historical japanese earthquakes: coupling among major seismic regions. *Earth, planets and space*, 51(9):907–916.

[Luzi et al., 2017] Luzi, L., Pacor, F., Puglia, R., Lanzano, G., Felicetta, C., D'Amico, M., Michelini, A., Faenza, L., Lauciani, V., Iervolino, I., et al. (2017). The central italy seismic sequence between august and december 2016: Analysis of strong-motion observations. *Seismological Research Letters*, 88(5):1219–1231.

[Mahdi and McLeod, 2019] Mahdi, E. and McLeod, A. I. (2019). Portmanteau test statistics. Accessed: 2019-09-18.

[Marsan and Lengline, 2008] Marsan, D. and Lengline, O. (2008). Extending earthquakes' reach through cascading. *Science*, 319(5866):1076–1079.

[Marsan and Lengliné, 2010] Marsan, D. and Lengliné, O. (2010). A new estimation of the decay of aftershock density with distance to the mainshock. *Journal of Geophysical Research: Solid Earth*, 115(B9):B09302.

[Marzocchi and Lombardi, 2009] Marzocchi, W. and Lombardi, A. M. (2009). Real-time forecasting following a damaging earthquake. *Geophysical Research Letters*, 36(21).

[Marzocchi and Taroni, 2014] Marzocchi, W. and Taroni, M. (2014). Some thoughts on declustering in probabilistic seismic-hazard analysis. *Bulletin of the Seismological Society of America*, 104(4):1838–1845.

[Matthews et al., 2002] Matthews, M. V., Ellsworth, W. L., and Reasenberg, P. A. (2002). A brownian model for recurrent earthquakes. *Bulletin of the Seismological Society of America*, 92(6):2233–2250.

[Mohler et al., 2011] Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., and Tita, G. E. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493):100–108.

[Neal, 2000] Neal, R. M. (2000). Markov chain sampling methods for dirichlet process mixture models. *Journal of computational and graphical statistics*, 9(2):249–265.

[Neyman and Scott, 1952] Neyman, J. and Scott, E. (1952). A theory of the spatial distribution of galaxies. *The Astrophysical Journal*, 116:144.

[Oakes, 1975] Oakes, D. (1975). The markovian self-exciting process. *Journal of Applied Probability*, 12(1):69–77.

[Ogata, 1988] 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, 1998] Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2):379–402.

[Ogata, 2004] Ogata, Y. (2004). Space-time model for regional seismicity and detection of crustal stress changes. *Journal of Geophysical Research: Solid Earth*, 109(B3).

[Ogata, 2011] Ogata, Y. (2011). Significant improvements of the space-time ETAS model for forecasting of accurate baseline seismicity. *Earth, planets and space*, 63(3):217–229.

[Ogata and Zhuang, 2006] Ogata, Y. and Zhuang, J. (2006). Space–time ETAS models and an improved extension. *Tectonophysics*, 413(1):13–23.

[Omi et al., 2014] Omi, T., Ogata, Y., Hirata, Y., and Aihara, K. (2014). Estimating the ETAS model from an early aftershock sequence. *Geophysical Research Letters*, 41(3):850–857.

[Omi et al., 2015] Omi, T., Ogata, Y., Hirata, Y., and Aihara, K. (2015). Intermediate-term forecasting of aftershocks from an early aftershock sequence: Bayesian and ensemble forecasting approaches. *Journal of Geophysical Research: Solid Earth*, 120(4):2561–2578.

[Omori, 1894] Omori, F. (1894). *On the after-shocks of earthquakes*, volume 7. The University.

[Ordaz and Arroyo, 2016] Ordaz, M. and Arroyo, D. (2016). On uncertainties in probabilistic seismic hazard analysis. *Earthquake Spectra*, 32(3):1405–1418.

[Porter et al., 2012] Porter, M. D., White, G., et al. (2012). Self-exciting hurdle models for terrorist activity. *The Annals of Applied Statistics*, 6(1):106–124.

[Rasmussen, 2011] Rasmussen, J. G. (2011). Temporal point processes the conditional intensity function.

[Rasmussen, 2013] Rasmussen, J. G. (2013). Bayesian inference for Hawkes processes. *Methodology and Computing in Applied Probability*, 15(3):623–642.

[Rasmussen, 2018] Rasmussen, J. G. (2018). Temporal point processes the conditional intensity function.

[Reid, 1910] Reid, H. F. (1910). *The mechanics of the earthquake*, volume 2. Carnegie institution of Washington.

[Reynaud-Bouret et al., 2010] Reynaud-Bouret, P., Schbath, S., et al. (2010). Adaptive estimation for Hawkes processes; application to genome analysis. *The Annals of Statistics*, 38(5):2781–2822.

[Ritschel et al., 2017] Ritschel, C., Ulbrich, U., Névir, P., and Rust, H. W. (2017). Precipitation extremes on multiple timescales–bartlett–lewis rectangular pulse model and intensity–duration–frequency curves. *Hydrology and Earth System Sciences*, 21(12):6501–6517.

[Rodriguez-Iturbe et al., 1987a] Rodriguez-Iturbe, I., Cox, D. R., and Isham, V. (1987a). Some models for rainfall based on stochastic point processes. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 410(1839):269–288.

[Rodriguez-Iturbe et al., 1987b] Rodriguez-Iturbe, I., De Power, B. F., and Valdes, J. B. (1987b). Rectangular pulses point process models for rainfall: analysis of empirical data. *Journal of Geophysical Research: Atmospheres*, 92(D8):9645–9656.

[Ross, 2018a] Ross, G. (2018a). Bayesian estimation of the ETAS model for earthquake occurrences. *Preprint*.

[Ross, 2018b] Ross, G. (2018b). Nonparametric Bayesian inference for the Hawkes process with seasonal event data. *Preprint*.

[Ross, 2014] Ross, S. M. (2014). *Introduction to probability models*. Academic press.

[Rotondi and Varini, 2006] Rotondi, R. and Varini, E. (2006). Bayesian analysis of marked stress release models for time-dependent hazard assessment in the western gulf of corinth. *Tectonophysics*, 423(1):107–113.

[Rotondi and Varini, 2007] Rotondi, R. and Varini, E. (2007). Bayesian inference of stress release models applied to some italian seismogenic zones. *Geophysical Journal International*, 169(1):301–314.

[Rotondi and Varini, 2019] Rotondi, R. and Varini, E. (2019). Failure models driven by a self-correcting point process in earthquake occurrence modeling. *Stochastic Environmental Research and Risk Assessment*, 3(3):709–724.

[Sangay and Vernes, 2008] Sangay, T. and Vernes, K. (2008). Human–wildlife conflict in the kingdom of bhutan: patterns of livestock predation by large mammalian carnivores. *Biological Conservation*, 141(5):1272–1282.

[Schoenberg, 2013] Schoenberg, F. P. (2013). Facilitated estimation of ETAS. *Bulletin of the Seismological Society of America*, 103(1):601–605.

[Schwarz et al., 1978] Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464.

[Sethuraman, 1994] Sethuraman, J. (1994). A constructive definition of dirichlet priors. *Statistica sinica*, 4(2):639–650.

[Shapira, 1983] Shapira, A. (1983). Potential earthquake risk estimations by application of a simulation process. *Tectonophysics*, 95(1-2):75–89.

[Shcherbakov, 2014] Shcherbakov, R. (2014). Bayesian confidence intervals for the magnitude of the largest aftershock. *Geophysical Research Letters*, 41(18):6380–6388.

[Sornette and Utkin, 2009] Sornette, D. and Utkin, S. (2009). Limits of declustering methods for disentangling exogenous from endogenous events in time series with foreshocks, main shocks, and aftershocks. *Physical Review E*, 79(6):061110.

[Spiegelhalter et al., 2014] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(3):485–493.

[Stephens, 1970] Stephens, M. A. (1970). Use of the kolmogorov–smirnov, cramer–von mises and related statistics without extensive tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 32(1):115–122.

[Tahernia et al., 2014] Tahernia, N., Khodabin, M., and Mirzaei, N. (2014). Non-poisson probabilistic seismic hazard assessment. *Arabian Journal of Geosciences*, 7(8):3259–3269.

[Uphyrkina et al., 2001] Uphyrkina, O., Johnson, W. E., Quigley, H., Miquelle, D., Marker, L., Bush, M., and O'Brien, S. J. (2001). Phylogenetics, genome diversity and origin of modern leopard, panthera pardus. *Molecular ecology*, 10(11):2617–2633.

[Utsu et al., 1995] Utsu, T., Ogata, Y., et al. (1995). The centenary of the omori formula for a decay law of aftershock activity. *Journal of Physics of the Earth*, 43(1):1–33.

[Vargas and Gneiting, 2012] Vargas, N. A. H. and Gneiting, T. (2012). Bayesian point process modelling of earthquake occurrences. Technical report, Ruprecht-Karls University Heidelberg.

[Varini and Rotondi, 2015] Varini, E. and Rotondi, R. (2015). Probability distribution of the waiting time in the stress release model: the gompertz distribution. *Environmental and Ecological Statistics*, 22(3):493–511.

[Veen and Schoenberg, 2008] 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, 1978] Vere-Jones, D. (1978). Earthquake prediction-a statistician's view. *Journal of Physics of the Earth*, 26(2):129–146.

[Vere-Jones and Davies, 1966] Vere-Jones, D. and Davies, R. (1966). A statistical survey of earthquakes in the main seismic region of new zealand: Part 2—time series analyses. *New Zealand journal of geology and geophysics*, 9(3):251–284.

[Wang et al., 2012] Wang, J.-H., Chen, K.-C., Lee, S.-J., Huang, W.-G., Wu, Y.-H., and Leu, P.-L. (2012). The frequency distribution of inter-event times of $m \geq 3$ earthquakes in the taipei metropolitan area: 1973-2010. *Terrestrial, Atmospheric & Oceanic Sciences*, 23(3):269–281.

[Wang and Macdonald, 2006] Wang, S. W. and Macdonald, D. (2006). Livestock predation by carnivores in jigme singye wangchuck national park, bhutan. *Biological Conservation*, 129(4):558–565.

[Wang and Macdonald, 2009] Wang, S. W. and Macdonald, D. W. (2009). The use of camera traps for estimating tiger and leopard populations in the high altitude mountains of bhutan. *Biological Conservation*, 142(3):606–613.

[Wheatley, 2016] Wheatley, S. (2016). *Extending the Hawkes process, a general outlier test, & case studies in extreme risk*. PhD thesis, ETH Zurich.

[Wheatley, 2017] Wheatley, S. (2017). Personal communication.

[Wheatley et al., 2016] Wheatley, S., Filimonov, V., and Sornette, D. (2016). The Hawkes process with renewal immigration & its estimation with an em algorithm. *Computational Statistics & Data Analysis*, 94:120–135.

[Wiemer and Wyss, 2000] 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.

[Xiaogu and Vere-Jones, 1994] Xiaogu, Z. and Vere-Jones, D. (1994). Further applications of the stochastic stress release model to historical earthquake data. *Tectonophysics*, 229(1-2):101–121.

[Yang et al., 2000] Yang, W.-z., Vere-Jones, D., Ma, L., and Liu, J. (2000). A method for locating the critical region of a future earthquake using the critical earthquake concept. *Earthquake*, 20(4):28–38.

[Yip, 1991] Yip, P. (1991). A method of inference for a capture-recapture experiment in discrete time with variable capture probabilities. *Stochastic Models*, 7(3):343–362.

[Zhu and Shi, 2002] Zhu, S.-b. and Shi, Y.-l. (2002). Improved stress release model: Application to the study of earthquake prediction in taiwan area. *Acta Seismologica Sinica*, 15(2):171–178.

[Zhuang et al., 2002] 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.

Appendices

Appendix A

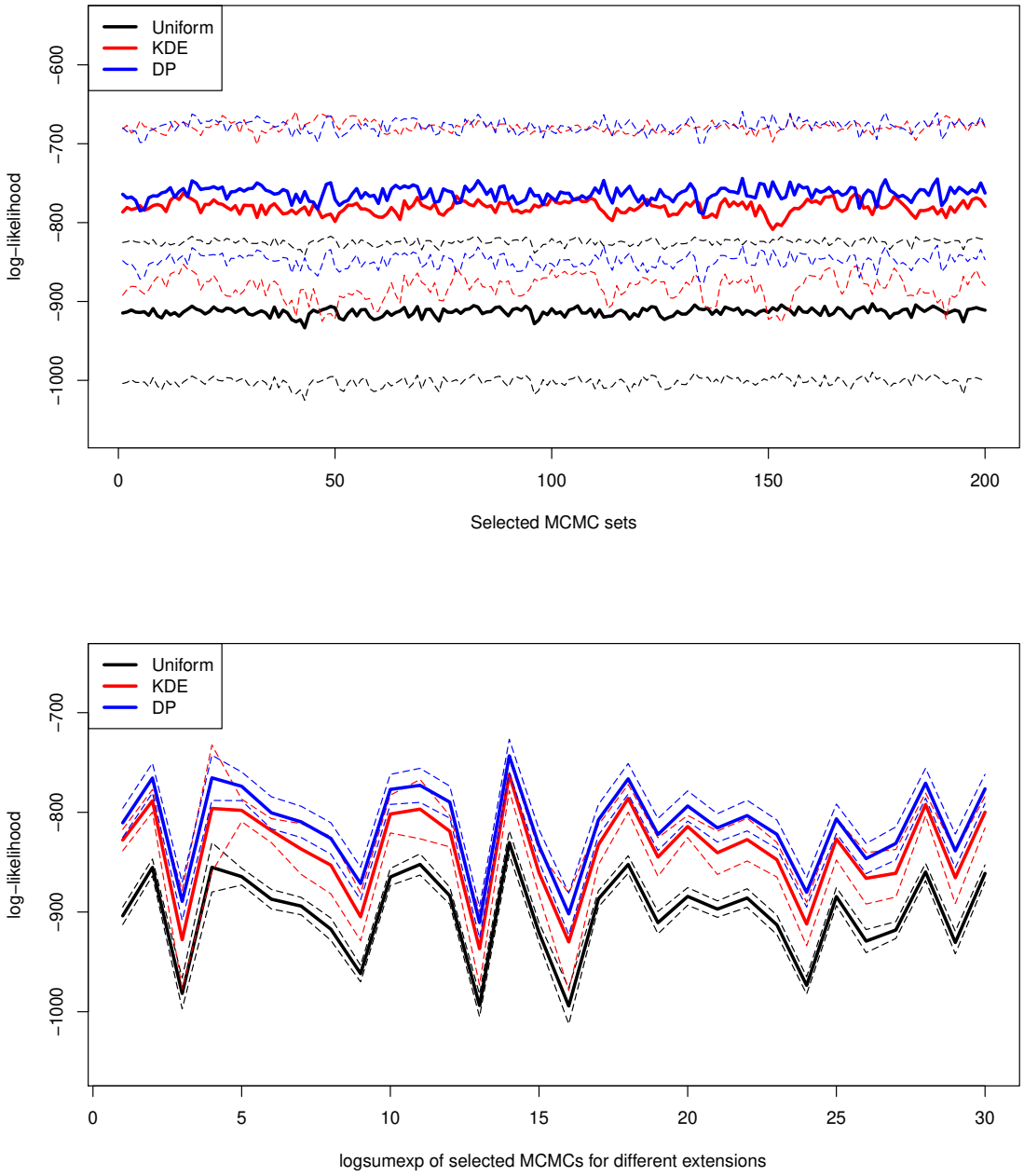

Figure A.1: Out-of-sample mean log-likelihood values for $\phi_2(x, y)$ for Uniform (Black), KDE (Red) and DP (Blue) based spatial ETAS model. The thick line indicates the mean value, while the dashed lines - the 95% confidence interval for the log-likelihood. Top: Log-likelihood averaged across all 30 out-of-sample periods for every 50^{th} MCMC sample. Bottom: Log-likelihood averaged across all 200 selected MCMC realisation across the 30 obtained out-of-sample periods.

(α ,$\bar{K}$, d, q)	n	p_μ	Area	$\hat{l}_U$	$\hat{l}_K$	$\hat{l}_D$	DIC_U	DIC_K	DIC_D	$\hat{l}^o_U$	$\hat{l}^o_K$	$\hat{l}^o_D$	$\bar{l}^o_U$	$\bar{l}^o_K$	$\bar{l}^o_D$
(1.00,0.10,0.01,1.10)	104	0.86	263.27	-572.97	-509.30	-517.02	1163.62	1031.08	1095.67	-636.06	-571.73	-590.32	-675.48	-613.25	-639.80
(1.30,0.10,0.01,1.10)	106	0.84	355.73	-581.33	-520.17	-526.21	1179.26	1050.82	1080.15	-640.88	-579.50	-585.64	-695.84	-633.15	-643.31
(1.60,0.10,0.01,1.10)	109	0.82	140.77	-583.55	-525.24	-531.52	1184.55	1068.07	1077.64	-625.71	-565.21	-571.07	-700.21	-651.52	-654.03
(1.90,0.10,0.01,1.10)	117	0.76	1703.85	-659.03	-580.60	-598.45	1331.66	1177.21	1250.73	-715.09	-640.10	-669.66	-786.19	-718.58	-741.63
(1.00,0.30,0.01,1.10)	193	0.46	9810.07	-1075.48	-1001.24	-996.49	2179.30	2035.65	2041.43	-1166.08	-1097.31	-1088.65	-1245.01	-1194.16	-1170.63
(1.30,0.30,0.01,1.10)	227	0.39	2210.55	-1232.05	-1113.47	-1157.87	2494.52	2251.24	2370.83	-1340.96	-1234.60	-1279.79	-1472.62	-1391.36	-1406.91
(1.00,0.50,0.01,1.10)	354	0.25	17919.23	-1941.87	-1837.92	-1855.92	3908.22	3701.29	3765.57	-2153.76	-2053.18	-2071.40	-2297.24	-2199.10	-2215.62
(1.00,0.10,0.26,1.10)	104	0.86	1048.75	-606.03	-533.97	-544.36	1228.24	1077.94	1126.11	-669.99	-594.79	-612.66	-715.19	-641.92	-666.40
(1.30,0.10,0.26,1.10)	106	0.84	475.12	-600.97	-529.78	-541.41	1218.21	1071.13	1116.59	-661.38	-588.90	-606.68	-720.78	-651.32	-668.20
(1.60,0.10,0.26,1.10)	109	0.82	114.39	-584.23	-526.71	-529.30	1187.35	1062.84	1079.93	-626.43	-566.84	-571.96	-703.88	-655.83	-663.32
(1.90,0.10,0.26,1.10)	117	0.76	521.32	-675.07	-593.38	-612.95	1365.40	1200.34	1263.42	-731.19	-652.97	-675.33	-818.53	-746.50	-767.40
(1.00,0.30,0.26,1.10)	193	0.46	68528.32	-1310.49	-1268.11	-1226.06	2644.71	2560.32	2491.93	-1407.28	-1371.03	-1323.84	-1513.34	-1486.43	-1425.94
(1.30,0.30,0.26,1.10)	227	0.39	21921.83	-1507.79	-1432.29	-1410.85	3045.11	2885.25	2873.17	-1620.43	-1547.58	-1533.10	-1787.37	-1716.97	-1696.94
(1.00,0.50,0.26,1.10)	354	0.25	65817.21	-2361.46	-2286.76	-2265.43	4744.79	4597.41	4581.74	-2668.99	-2600.14	-2576.59	-2839.82	-2774.58	-2746.81
(1.00,0.10,0.50,1.10)	104	0.86	632.75	-599.72	-532.54	-542.24	1216.58	1068.97	1113.40	-663.46	-594.46	-606.66	-709.09	-641.71	-654.96
(1.30,0.10,0.50,1.10)	106	0.84	927.61	-619.61	-550.28	-560.70	1255.38	1105.14	1162.84	-681.39	-609.29	-625.42	-742.56	-671.66	-691.55
(1.60,0.10,0.50,1.10)	109	0.82	1139.95	-625.77	-565.56	-570.86	1266.57	1141.07	1162.19	-668.25	-606.44	-613.08	-751.48	-694.97	-704.32
(1.90,0.10,0.50,1.10)	117	0.76	1066.74	-711.98	-631.76	-651.92	1439.38	1275.00	1350.54	-769.91	-693.14	-723.20	-861.54	-794.03	-814.99
(1.00,0.30,0.50,1.10)	193	0.46	172645.36	-1406.17	-1362.54	-1307.37	2835.36	2741.96	2663.93	-1503.28	-1458.80	-1400.19	-1600.25	-1556.37	-1505.17
(1.30,0.30,0.50,1.10)	227	0.39	5393.97	-1470.41	-1405.43	-1388.11	2965.71	2830.90	2806.43	-1600.57	-1551.26	-1521.59	-1764.18	-1710.81	-1679.85
(1.00,0.50,0.50,1.10)	354	0.25	32108.34	-2482.27	-2425.41	-2387.22	4987.90	4876.93	4812.40	-2790.20	-2735.60	-2692.28	-3004.20	-2956.75	-2902.71
(1.00,0.10,0.01,1.55)	104	0.86	23.81	-518.87	-476.10	-477.41	1053.93	970.82	988.81	-578.30	-532.28	-539.12	-612.15	-575.62	-583.85
(1.30,0.10,0.01,1.55)	106	0.84	20.54	-519.29	-489.31	-488.63	1055.62	993.89	992.99	-577.45	-548.98	-550.80	-625.46	-594.51	-598.34
(1.60,0.10,0.01,1.55)	109	0.82	144.83	-557.42	-482.87	-488.81	1131.79	988.47	1043.57	-599.59	-524.26	-545.34	-662.94	-590.52	-621.06
(1.90,0.10,0.01,1.55)	117	0.76	47.53	-572.90	-507.07	-507.23	1160.75	1030.88	1052.85	-620.96	-554.24	-553.48	-682.55	-617.33	-628.92
(1.00,0.30,0.01,1.55)	193	0.46	498.78	-807.80	-725.71	-737.97	1642.69	1481.07	1502.60	-890.38	-807.35	-812.46	-948.16	-870.27	-876.11
(1.30,0.30,0.01,1.55)	227	0.39	2207.51	-880.65	-822.59	-798.73	1790.04	1677.35	1647.90	-959.44	-900.85	-878.73	-1050.25	-1005.99	-969.55
(1.00,0.50,0.01,1.55)	354	0.25	873.50	-1077.82	-989.02	-1012.46	2178.89	2003.92	2073.78	-1212.06	-1122.64	-1152.97	-1293.14	-1222.15	-1236.42
(1.00,0.10,0.26,1.55)	104	0.86	416.37	-577.70	-525.44	-519.84	1174.64	1056.91	1072.41	-640.34	-585.12	-582.57	-680.25	-622.68	-627.76
(1.30,0.10,0.26,1.55)	106	0.84	41.41	-552.62	-512.51	-514.30	1123.32	1033.62	1045.26	-612.05	-571.33	-574.13	-662.92	-619.84	-625.81
(1.60,0.10,0.26,1.55)	109	0.82	485.26	-570.34	-509.26	-519.08	1157.88	1036.11	1090.97	-612.76	-550.61	-573.91	-683.78	-624.21	-650.72
(1.90,0.10,0.26,1.55)	117	0.76	126.23	-611.43	-559.29	-562.82	1240.50	1133.09	1147.21	-661.70	-611.71	-614.83	-732.81	-682.33	-691.66
(1.00,0.30,0.26,1.55)	193	0.46	7398.71	-1004.10	-959.51	-944.23	2030.66	1943.98	1940.29	-1089.68	-1048.22	-1030.47	-1155.22	-1115.14	-1104.13
(1.30,0.30,0.26,1.55)	227	0.39	628.84	-1030.52	-957.29	-968.56	2087.27	1932.43	1981.71	-1120.56	-1051.86	-1058.41	-1239.99	-1179.38	-1187.19
(1.00,0.50,0.26,1.55)	354	0.25	8026.94	-1680.80	-1596.70	-1620.18	3386.19	3223.25	3298.66	-1879.12	-1796.45	-1828.14	-1999.88	-1928.87	-1949.43
(1.00,0.10,0.50,1.55)	104	0.86	23.16	-527.27	-487.51	-490.24	1073.21	992.59	997.30	-589.85	-551.68	-553.71	-626.59	-591.88	-595.73
(1.30,0.10,0.50,1.55)	106	0.84	130.91	-573.46	-516.54	-520.15	1164.62	1039.69	1073.28	-639.13	-586.49	-596.04	-692.62	-637.19	-652.25
(1.60,0.10,0.50,1.55)	109	0.82	223.16	-587.40	-527.05	-535.57	1191.52	1064.17	1092.99	-629.22	-568.83	-576.14	-701.85	-642.69	-654.18
(1.90,0.10,0.50,1.55)	117	0.76	474.27	-614.85	-554.72	-564.35	1245.50	1122.76	1160.01	-668.52	-609.45	-621.77	-741.92	-682.70	-698.49
(1.00,0.30,0.50,1.55)	193	0.46	2285.25	-1064.54	-983.57	-997.45	2152.55	1979.26	2034.96	-1150.41	-1068.20	-1084.01	-1225.13	-1139.03	-1160.33
(1.30,0.30,0.50,1.55)	227	0.39	1868.59	-1171.00	-1084.97	-1104.10	2366.34	2183.23	2264.80	-1257.84	-1171.56	-1203.76	-1397.30	-1310.49	-1344.50
(1.00,0.50,0.50,1.55)	354	0.25	2469.55	-1697.75	-1591.85	-1628.96	3421.11	3212.24	3317.17	-1910.33	-1801.71	-1851.98	-2048.56	-1960.75	-1989.55
(1.00,0.10,0.01,2.00)	104	0.86	20.30	-500.70	-464.72	-465.25	1029.53	955.49	979.26	-561.21	-522.61	-528.20	-599.22	-564.31	-572.87
(1.30,0.10,0.01,2.00)	106	0.84	18.76	-499.77	-466.94	-470.53	1015.48	952.00	963.06	-560.05	-525.54	-530.22	-603.47	-571.06	-578.40
(1.60,0.10,0.01,2.00)	109	0.82	18.76	-497.89	-462.76	-467.74	1012.77	942.56	962.32	-538.99	-503.61	-508.12	-596.16	-562.78	-572.29
(1.90,0.10,0.01,2.00)	117	0.76	20.29	-525.71	-485.69	-488.76	1067.67	986.16	1004.68	-577.89	-532.55	-536.13	-625.13	-583.58	-590.21
(1.00,0.30,0.01,2.00)	193	0.46	22.18	-663.05	-632.04	-635.49	1360.77	1285.42	1299.25	-742.15	-712.08	-712.99	-791.14	-751.26	-758.33
(1.30,0.30,0.01,2.00)	227	0.39	29.62	-664.63	-599.28	-611.34	1352.51	1223.70	1268.46	-731.02	-670.99	-681.65	-797.34	-733.63	-749.02
(1.00,0.50,0.01,2.00)	354	0.25	22.57	-808.98	-763.30	-775.61	1638.26	1548.66	1585.26	-919.67	-878.09	-887.27	-960.60	-925.61	-934.72
(1.00,0.10,0.26,2.00)	104	0.86	28.72	-534.45	-491.15	-493.90	1085.92	998.96	1018.92	-594.66	-549.15	-555.28	-629.05	-587.32	-598.62
(1.30,0.10,0.26,2.00)	106	0.84	21.15	-523.60	-487.97	-490.17	1059.09	990.76	1003.21	-582.78	-546.66	-548.92	-629.98	-593.39	-598.56
(1.60,0.10,0.26,2.00)	109	0.82	24.62	-533.50	-490.17	-490.44	1080.25	995.48	1009.02	-574.42	-530.89	-531.27	-635.61	-600.66	-603.91
(1.90,0.10,0.26,2.00)	117	0.76	19.37	-545.31	-511.39	-509.90	1102.87	1036.92	1047.59	-596.21	-562.50	-561.63	-651.97	-626.02	-629.31
(1.00,0.30,0.26,2.00)	193	0.46	90.16	-875.97	-838.47	-834.01	1774.86	1692.87	1693.23	-957.33	-918.39	-915.32	-1016.51	-971.28	-975.87
(1.30,0.30,0.26,2.00)	227	0.39	49.46	-879.13	-831.68	-840.36	1782.78	1682.71	1708.55	-960.53	-915.01	-924.93	-1057.24	-1012.07	-1026.26
(1.00,0.50,0.26,2.00)	354	0.25	76.38	-1285.58	-1242.22	-1249.74	2589.98	2515.15	2529.19	-1435.86	-1391.47	-1400.71	-1519.15	-1489.28	-1496.03
(1.00,0.10,0.50,2.00)	104	0.86	40.33	-543.77	-494.15	-499.20	1104.13	1003.62	1024.09	-604.61	-553.79	-556.90	-641.32	-592.69	-603.75
(1.30,0.10,0.50,2.00)	106	0.84	54.57	-544.94	-496.22	-500.53	1108.94	1008.02	1029.59	-605.48	-554.84	-563.24	-655.85	-602.59	-613.79
(1.60,0.10,0.50,2.00)	109	0.82	76.12	-555.21	-502.20	-505.88	1128.19	1018.53	1037.12	-597.72	-542.24	-549.77	-664.69	-612.61	-624.57
(1.90,0.10,0.50,2.00)	117	0.76	58.67	-611.33	-560.13	-565.51	1238.31	1133.52	1152.19	-660.06	-612.26	-615.28	-729.87	-678.44	-686.59
(1.00,0.30,0.50,2.00)	193	0.46	2553.12	-933.78	-898.71	-884.84	1890.29	1816.18	1826.24	-1016.23	-982.36	-967.25	-1083.33	-1047.78	-1040.12
(1.30,0.30,0.50,2.00)	227	0.39	621.84	-1025.37	-966.49	-972.36	2076.19	1955.37	1986.07	-1103.29	-1044.10	-1051.13	-1224.20	-1164.83	-1176.49
(1.00,0.50,0.50,2.00)	354	0.25	317.30	-1477.52	-1424.55	-1428.54	2978.76	2871.34	2890.92	-1649.89	-1598.19	-1600.66	-1762.50	-1717.58	-1714.93

Table A.1: Obtained diagnostic results for uncaused events spatial density $\phi_1(\cdot)$ as of Equation 4.13

(α ,$\bar{K}$, d, q)	n	p_μ	Area	$\hat{l}_U$	$\hat{l}_K$	$\hat{l}_D$	DIC_U	DIC_K	DIC_D	$\hat{l}^o_U$	$\hat{l}^o_K$	$\hat{l}^o_D$	$\bar{l}^o_U$	$\bar{l}^o_K$	$\bar{l}^o_D$
(1.00,0.10,0.01,1.10)	104	0.86	323.35	-372.30	-349.73	-337.81	759.47	704.15	701.76	-406.38	-388.43	-373.77	-438.06	-420.15	-412.64
(1.30,0.10,0.01,1.10)	106	0.84	384.88	-392.40	-358.49	-331.04	803.02	721.29	691.61	-436.03	-399.81	-371.85	-473.03	-441.21	-416.42
(1.60,0.10,0.01,1.10)	109	0.82	134.51	-389.89	-328.43	-340.00	795.98	664.70	695.26	-414.89	-355.02	-364.99	-471.71	-417.86	-430.44
(1.90,0.10,0.01,1.10)	117	0.76	1905.36	-473.62	-430.87	-422.49	963.04	878.67	870.98	-515.93	-470.32	-465.81	-572.68	-557.54	-526.03
(1.00,0.30,0.01,1.10)	193	0.46	9611.96	-865.28	-806.97	-800.68	1756.60	1633.54	1649.51	-919.69	-861.40	-847.23	-992.66	-956.17	-931.11
(1.30,0.30,0.01,1.10)	227	0.39	2102.83	-1058.47	-984.72	-965.67	2146.98	1994.99	1998.32	-1154.55	-1090.13	-1065.51	-1261.29	-1245.12	-1173.90
(1.00,0.50,0.01,1.10)	354	0.25	18132.69	-1776.63	-1697.02	-1664.46	3576.41	3419.38	3316.94	-1967.40	-1890.09	-1860.28	-2101.42	-2031.49	-1996.59
(1.00,0.10,0.26,1.10)	104	0.86	1033.10	-424.96	-456.12	-367.81	867.23	917.27	767.50	-459.13	-504.91	-403.39	-502.13	-546.50	-453.05
(1.30,0.10,0.26,1.10)	106	0.84	385.60	-427.96	-384.05	-361.48	872.78	774.07	747.39	-472.55	-426.62	-404.64	-517.45	-478.65	-455.84
(1.60,0.10,0.26,1.10)	109	0.82	91.54	-406.82	-339.66	-355.28	832.05	683.11	760.99	-432.36	-364.49	-386.97	-493.00	-435.37	-459.86
(1.90,0.10,0.26,1.10)	117	0.76	414.39	-517.32	-506.57	-453.42	1057.15	1022.85	936.81	-558.11	-555.67	-493.42	-638.80	-646.60	-573.95
(1.00,0.30,0.26,1.10)	193	0.46	69110.36	-1175.01	-1136.22	-1035.13	2376.75	2298.74	2117.87	-1244.34	-1211.37	-1112.21	-1347.52	-1316.58	-1211.42
(1.30,0.30,0.26,1.10)	227	0.39	21874.10	-1392.09	-1335.59	-1232.36	2812.81	2706.47	2527.04	-1494.86	-1438.15	-1333.91	-1642.57	-1616.70	-1478.14
(1.00,0.50,0.26,1.10)	354	0.25	65852.66	-2291.77	-2214.13	-2117.12	4608.04	4456.05	4278.61	-2590.31	-2513.50	-2391.75	-2758.35	-2685.44	-2564.73
(1.00,0.10,0.50,1.10)	104	0.86	585.27	-414.10	-403.07	-351.38	845.66	810.20	733.11	-448.58	-445.94	-387.32	-489.94	-489.56	-435.78
(1.30,0.10,0.50,1.10)	106	0.84	878.68	-442.57	-397.54	-372.99	906.29	833.26	793.14	-486.40	-439.82	-412.66	-534.95	-526.45	-470.57
(1.60,0.10,0.50,1.10)	109	0.82	1108.75	-453.29	-411.57	-382.56	931.19	850.11	802.98	-480.66	-437.21	-408.24	-555.49	-556.76	-487.60
(1.90,0.10,0.50,1.10)	117	0.76	970.37	-558.04	-500.23	-481.89	1136.97	1017.46	990.48	-602.35	-549.31	-522.05	-685.52	-706.78	-610.35
(1.00,0.30,0.50,1.10)	193	0.46	169255.99	-1279.09	-1232.66	-1135.11	2582.50	2488.35	2317.65	-1345.85	-1300.01	-1188.93	-1440.28	-1395.09	-1291.31
(1.30,0.30,0.50,1.10)	227	0.39	5328.05	-1407.20	-1360.04	-1238.16	2842.64	2753.56	2517.71	-1525.82	-1488.79	-1353.13	-1678.12	-1648.91	-1500.76
(1.00,0.50,0.50,1.10)	354	0.25	31709.34	-2439.79	-2377.09	-2245.69	4904.16	4783.58	4549.50	-2740.33	-2679.89	-2524.99	-2953.12	-2897.10	-2735.70
(1.00,0.10,0.01,1.55)	104	0.86	15.76	-309.72	-263.27	-263.27	637.45	533.80	575.64	-343.06	-298.52	-304.08	-364.20	-317.88	-338.24
(1.30,0.10,0.01,1.55)	106	0.84	13.30	-321.65	-276.35	-278.00	660.25	563.83	589.93	-364.26	-324.44	-329.90	-389.65	-349.76	-364.47
(1.60,0.10,0.01,1.55)	109	0.82	119.73	-326.35	-285.16	-284.92	669.81	599.19	619.90	-350.15	-323.13	-314.79	-391.34	-363.80	-363.97
(1.90,0.10,0.01,1.55)	117	0.76	32.50	-357.06	-293.68	-298.49	728.51	609.18	678.92	-387.61	-326.77	-350.72	-427.19	-380.19	-406.83
(1.00,0.30,0.01,1.55)	193	0.46	492.59	-547.28	-492.72	-491.80	1115.63	1011.17	1050.29	-592.86	-540.22	-545.39	-637.84	-601.01	-598.19
(1.30,0.30,0.01,1.55)	227	0.39	2214.16	-629.77	-601.73	-552.07	1285.45	1230.75	1147.82	-689.41	-657.56	-613.16	-749.54	-732.28	-679.24
(1.00,0.50,0.01,1.55)	354	0.25	940.99	-809.33	-725.15	-753.51	1642.44	1476.14	1553.92	-897.42	-813.13	-844.65	-979.83	-926.58	-928.79
(1.00,0.10,0.26,1.55)	104	0.86	368.33	-402.17	-372.05	-343.70	825.74	747.89	720.56	-437.73	-412.39	-378.87	-471.73	-440.12	-413.58
(1.30,0.10,0.26,1.55)	106	0.84	30.21	-373.14	-310.48	-317.58	764.06	627.92	690.99	-416.59	-357.88	-365.99	-452.63	-395.01	-411.38
(1.60,0.10,0.26,1.55)	109	0.82	397.97	-399.72	-351.79	-343.12	815.36	718.49	747.39	-424.90	-378.50	-372.74	-479.63	-437.65	-433.43
(1.90,0.10,0.26,1.55)	117	0.76	155.01	-446.16	-381.89	-384.86	910.78	779.15	795.88	-484.65	-423.30	-423.42	-538.22	-485.30	-482.78
(1.00,0.30,0.26,1.55)	193	0.46	7400.09	-805.33	-775.66	-731.58	1636.05	1571.95	1538.11	-860.14	-832.86	-786.43	-922.55	-897.67	-851.39
(1.30,0.30,0.26,1.55)	227	0.39	620.57	-891.40	-809.11	-774.39	1813.76	1632.14	1599.32	-977.67	-898.54	-853.48	-1077.75	-1014.24	-958.15
(1.00,0.50,0.26,1.55)	354	0.25	7510.11	-1585.07	-1524.54	-1475.52	3191.18	3074.50	3024.71	-1768.63	-1707.89	-1663.30	-1880.68	-1831.44	-1777.08
(1.00,0.10,0.50,1.55)	104	0.86	15.71	-340.71	-286.57	-294.17	701.50	582.62	636.13	-374.40	-323.11	-339.82	-405.88	-356.76	-381.19
(1.30,0.10,0.50,1.55)	106	0.84	184.20	-397.51	-344.48	-334.37	811.05	693.04	708.71	-443.44	-393.24	-385.20	-484.67	-436.86	-428.95
(1.60,0.10,0.50,1.55)	109	0.82	248.33	-404.93	-351.27	-347.70	825.86	706.32	719.17	-429.32	-377.12	-371.71	-490.41	-437.11	-437.75
(1.90,0.10,0.50,1.55)	117	0.76	424.70	-452.25	-412.35	-388.31	925.49	841.02	777.88	-493.63	-458.18	-427.82	-556.25	-523.92	-491.94
(1.00,0.30,0.50,1.55)	193	0.46	2195.63	-941.13	-889.26	-841.41	1906.41	1795.34	1713.15	-999.07	-946.37	-895.09	-1071.42	-1034.43	-976.35
(1.30,0.30,0.50,1.55)	227	0.39	1827.01	-1058.93	-985.20	-938.81	2143.77	1984.69	1913.29	-1139.01	-1076.50	-1017.55	-1261.15	-1196.39	-1135.23
(1.00,0.50,0.50,1.55)	354	0.25	2411.41	-1604.27	-1499.69	-1491.58	3233.35	3031.90	3020.02	-1808.96	-1712.80	-1686.53	-1939.92	-1869.35	-1818.42
(1.00,0.10,0.01,2.00)	104	0.86	7.83	-286.78	-244.18	-244.18	590.97	511.69	551.81	-315.87	-283.72	-296.73	-338.30	-302.20	-327.09
(1.30,0.10,0.01,2.00)	106	0.84	8.74	-284.40	-248.76	-252.96	586.88	514.98	537.63	-323.76	-291.36	-293.02	-346.70	-314.42	-327.78
(1.60,0.10,0.01,2.00)	109	0.82	9.03	-290.66	-251.36	-253.47	598.33	533.76	519.73	-314.31	-288.81	-279.28	-346.71	-323.10	-332.86
(1.90,0.10,0.01,2.00)	117	0.76	8.41	-305.77	-262.45	-269.84	627.87	536.66	576.24	-331.21	-280.09	-290.13	-363.31	-334.72	-343.31
(1.00,0.30,0.01,2.00)	193	0.46	8.81	-394.46	-353.05	-358.76	813.12	722.55	773.22	-432.13	-389.28	-397.94	-465.50	-419.99	-437.48
(1.30,0.30,0.01,2.00)	227	0.39	9.41	-346.33	-293.04	-314.59	714.13	605.05	677.03	-372.65	-323.72	-343.40	-423.24	-372.79	-406.60
(1.00,0.50,0.01,2.00)	354	0.25	21.04	-463.16	-388.85	-420.71	943.75	800.65	876.87	-516.20	-439.68	-470.39	-551.79	-493.77	-524.01
(1.00,0.10,0.26,2.00)	104	0.86	41.25	-344.07	-289.01	-292.34	705.82	587.02	645.79	-377.18	-323.25	-346.39	-403.05	-346.88	-380.01
(1.30,0.10,0.26,2.00)	106	0.84	11.78	-332.31	-284.29	-288.29	683.77	575.24	639.74	-374.89	-326.68	-338.24	-405.65	-356.40	-377.04
(1.60,0.10,0.26,2.00)	109	0.82	10.17	-329.46	-284.22	-290.76	677.85	581.97	618.88	-354.15	-312.14	-317.70	-395.79	-356.33	-370.01
(1.90,0.10,0.26,2.00)	117	0.76	16.78	-383.73	-315.92	-327.00	783.62	648.76	682.60	-415.21	-348.83	-366.70	-461.25	-397.60	-419.30
(1.00,0.30,0.26,2.00)	193	0.46	78.00	-690.75	-622.31	-622.36	1401.33	1255.70	1274.89	-745.76	-676.83	-670.98	-792.17	-718.86	-724.29
(1.30,0.30,0.26,2.00)	227	0.39	41.74	-690.68	-600.58	-606.20	1409.94	1212.35	1256.92	-761.33	-674.74	-684.22	-834.23	-747.45	-759.95
(1.00,0.50,0.26,2.00)	354	0.25	88.54	-1129.41	-1057.98	-1057.03	2283.39	2140.09	2164.82	-1253.97	-1185.85	-1169.29	-1340.26	-1273.06	-1265.75
(1.00,0.10,0.50,2.00)	104	0.86	70.59	-366.90	-300.85	-307.68	749.83	610.81	649.62	-401.54	-337.60	-347.95	-430.90	-363.94	-382.75
(1.30,0.10,0.50,2.00)	106	0.84	49.12	-359.60	-299.58	-304.98	736.26	605.22	650.45	-403.22	-338.69	-353.49	-437.05	-370.75	-397.38
(1.60,0.10,0.50,2.00)	109	0.82	52.59	-374.12	-305.99	-316.25	765.74	623.97	693.76	-398.30	-334.46	-355.45	-449.44	-383.91	-414.26
(1.90,0.10,0.50,2.00)	117	0.76	88.80	-451.30	-374.83	-384.33	919.77	763.13	825.63	-483.91	-413.11	-430.59	-543.50	-471.15	-492.56
(1.00,0.30,0.50,2.00)	193	0.46	2974.76	-781.78	-759.19	-690.92	1587.73	1543.51	1435.67	-834.08	-814.86	-738.32	-896.09	-879.57	-807.38
(1.30,0.30,0.50,2.00)	227	0.39	447.65	-889.49	-813.81	-773.65	1807.17	1641.23	1603.52	-961.42	-885.61	-854.78	-1061.94	-984.08	-951.33
(1.00,0.50,0.50,2.00)	354	0.25	408.46	-1377.61	-1314.76	-1267.55	2779.47	2652.71	2571.67	-1523.34	-1466.52	-1410.95	-1640.68	-1589.51	-1525.43

Table A.2: Obtained diagnostic results for uncaused events spatial density $\phi_2(\cdot)$ as of Equation 4.14

(α ,$\bar{K}$, d, q)	n	p_μ	Area	$\hat{l}_U$	$\hat{l}_K$	$\hat{l}_D$	DIC_U	DIC_K	DIC_D	$\hat{l}^o_U$	$\hat{l}^o_K$	$\hat{l}^o_D$	$\bar{l}^o_U$	$\bar{l}^o_K$	$\bar{l}^o_D$
(1.00,0.10,0.01,1.10)	105	0.85	278.44	-345.83	-296.77	-304.93	712.02	597.33	635.93	-389.11	-347.41	-351.80	-430.26	-381.75	-386.72
(1.30,0.10,0.01,1.10)	116	0.77	452.49	-438.03	-385.09	-396.37	897.76	779.84	812.53	-481.16	-433.02	-439.76	-518.43	-467.11	-470.33
(1.60,0.10,0.01,1.10)	142	0.63	46536.83	-635.09	-599.43	-584.18	1293.88	1215.02	1206.04	-739.73	-718.99	-689.78	-786.17	-790.19	-731.13
(1.90,0.10,0.01,1.10)	194	0.46	1413.15	-685.69	-616.72	-614.28	1399.81	1263.63	1271.84	-786.66	-733.38	-713.27	-852.53	-905.14	-782.15
(1.00,0.30,0.01,1.10)	205	0.43	4409.35	-899.33	-839.26	-821.68	1828.71	1699.40	1682.76	-974.53	-933.95	-908.39	-1073.62	-1080.00	-1005.36
(1.30,0.30,0.01,1.10)	268	0.33	4969.82	-1126.76	-1050.93	-1040.47	2279.64	2135.51	2126.61	-1298.61	-1238.16	-1222.54	-1458.75	-1493.45	-1381.35
(1.00,0.50,0.01,1.10)	705	0.13	22237.82	-3154.71	-3013.03	-3045.61	6328.67	6046.73	6131.23	-3457.32	-3317.31	-3336.19	-3664.24	-3540.28	-3554.75
(1.00,0.10,0.26,1.10)	105	0.85	341.29	-375.90	-319.22	-327.97	773.93	642.39	673.35	-415.01	-363.15	-367.06	-469.28	-419.46	-418.18
(1.30,0.10,0.26,1.10)	116	0.77	2086.13	-513.70	-490.26	-446.13	1053.34	994.10	934.23	-566.75	-552.55	-505.09	-607.11	-588.24	-536.80
(1.60,0.10,0.26,1.10)	142	0.63	2258.53	-660.54	-594.17	-596.21	1342.88	1207.77	1216.53	-755.09	-696.34	-684.40	-805.33	-791.60	-737.15
(1.90,0.10,0.26,1.10)	194	0.46	18449.05	-1138.61	-1096.77	-992.75	2307.55	2227.20	2029.05	-1265.13	-1222.35	-1107.63	-1353.40	-1338.86	-1191.49
(1.00,0.30,0.26,1.10)	205	0.43	16076.94	-1230.53	-1168.66	-1104.27	2488.04	2351.59	2265.36	-1323.42	-1262.90	-1203.31	-1462.07	-1435.30	-1331.09
(1.30,0.30,0.26,1.10)	268	0.33	29346.94	-1570.80	-1508.76	-1423.72	3167.08	3049.43	2879.83	-1769.50	-1712.08	-1609.29	-1989.69	-1947.37	-1826.38
(1.00,0.50,0.26,1.10)	705	0.13	110391.08	-4731.19	-4639.72	-4531.70	9482.25	9298.44	9107.04	-5173.32	-5085.21	-4961.93	-5455.32	-5366.85	-5235.88
(1.00,0.10,0.50,1.10)	105	0.85	854.30	-412.33	-379.19	-361.32	850.78	765.42	730.03	-460.25	-445.74	-416.29	-515.32	-495.48	-457.92
(1.30,0.10,0.50,1.10)	116	0.77	7774.77	-532.14	-495.18	-455.47	1089.62	1015.46	932.81	-584.77	-543.40	-498.55	-625.70	-601.73	-542.02
(1.60,0.10,0.50,1.10)	142	0.63	91416.84	-746.87	-719.42	-657.02	1516.89	1457.35	1339.96	-843.82	-817.18	-741.65	-900.41	-885.98	-798.55
(1.90,0.10,0.50,1.10)	194	0.46	4469.48	-1120.77	-1070.78	-978.48	2268.83	2164.40	1997.93	-1262.79	-1219.76	-1112.63	-1348.55	-1357.52	-1197.29
(1.00,0.30,0.50,1.10)	205	0.43	13997.49	-1223.93	-1151.83	-1101.71	2472.21	2323.59	2228.42	-1325.10	-1265.22	-1204.83	-1467.81	-1445.86	-1332.42
(1.30,0.30,0.50,1.10)	268	0.33	23205.63	-1742.15	-1664.14	-1582.40	3511.21	3364.40	3200.93	-1962.89	-1896.02	-1788.89	-2213.15	-2202.37	-2028.71
(1.00,0.50,0.50,1.10)	705	0.13	185242.51	-4869.60	-4783.13	-4636.73	9757.00	9586.69	9307.63	-5360.87	-5281.71	-5107.73	-5662.67	-5586.33	-5408.89
(1.00,0.10,0.01,1.55)	105	0.85	643.21	-299.65	-277.67	-265.48	619.75	574.74	551.57	-328.00	-306.53	-297.69	-366.27	-343.90	-335.59
(1.30,0.10,0.01,1.55)	116	0.77	400.54	-316.77	-296.70	-285.57	650.71	626.03	592.66	-355.22	-350.71	-319.54	-382.40	-375.73	-349.82
(1.60,0.10,0.01,1.55)	142	0.63	60.60	-354.72	-299.84	-308.35	724.12	612.24	650.23	-404.19	-346.55	-357.88	-428.39	-375.47	-394.22
(1.90,0.10,0.01,1.55)	194	0.46	285.64	-484.59	-404.03	-420.87	998.89	851.07	909.41	-541.85	-471.79	-477.37	-579.48	-511.28	-525.66
(1.00,0.30,0.01,1.55)	205	0.43	237.08	-579.34	-487.55	-500.48	1184.11	989.07	1101.42	-632.50	-546.14	-595.83	-682.17	-598.11	-636.53
(1.30,0.30,0.01,1.55)	268	0.33	1292.55	-651.33	-563.34	-588.94	1323.45	1157.68	1251.68	-744.42	-651.38	-693.10	-800.09	-758.25	-764.65
(1.00,0.50,0.01,1.55)	705	0.13	5446.67	-1591.12	-1480.89	-1516.96	3202.87	2983.59	3083.93	-1733.29	-1623.98	-1664.62	-1832.34	-1745.82	-1763.89
(1.00,0.10,0.26,1.55)	105	0.85	44.55	-327.68	-276.26	-282.58	674.32	556.62	611.55	-363.47	-315.70	-333.44	-406.95	-359.48	-377.23
(1.30,0.10,0.26,1.55)	116	0.77	602.55	-411.78	-370.26	-366.51	842.38	763.21	754.84	-454.31	-415.46	-412.54	-488.00	-456.27	-444.54
(1.60,0.10,0.26,1.55)	142	0.63	1360.52	-521.93	-466.90	-467.57	1066.87	948.71	967.63	-604.98	-554.01	-548.66	-640.25	-596.22	-583.70
(1.90,0.10,0.26,1.55)	194	0.46	3303.29	-695.94	-642.53	-617.43	1421.23	1312.42	1279.15	-788.37	-732.13	-697.00	-849.13	-804.39	-762.89
(1.00,0.30,0.26,1.55)	205	0.43	2887.54	-821.50	-756.79	-741.43	1674.33	1536.42	1510.79	-886.37	-831.73	-810.46	-982.68	-932.50	-893.28
(1.30,0.30,0.26,1.55)	268	0.33	1656.34	-1077.88	-997.02	-987.85	2184.66	2025.44	2013.66	-1235.93	-1163.50	-1139.77	-1377.18	-1333.57	-1286.57
(1.00,0.50,0.26,1.55)	705	0.13	18188.41	-2779.57	-2716.82	-2655.04	5575.87	5454.56	5364.30	-3034.90	-2976.01	-2921.21	-3229.20	-3175.77	-3109.52
(1.00,0.10,0.50,1.55)	105	0.85	82.28	-351.10	-286.31	-293.90	722.47	578.65	638.67	-384.37	-320.74	-346.68	-430.67	-365.12	-391.26
(1.30,0.10,0.50,1.55)	116	0.77	433.98	-415.82	-361.88	-365.67	854.51	732.28	761.77	-462.77	-406.34	-413.22	-499.35	-440.49	-445.67
(1.60,0.10,0.50,1.55)	142	0.63	533.03	-568.06	-506.42	-513.99	1161.63	1025.24	1054.70	-654.09	-594.60	-588.54	-693.61	-634.75	-628.75
(1.90,0.10,0.50,1.55)	194	0.46	905.46	-848.57	-761.49	-757.48	1728.79	1547.80	1552.22	-960.43	-881.35	-853.92	-1025.86	-944.89	-916.97
(1.00,0.30,0.50,1.55)	205	0.43	956.21	-911.33	-822.10	-820.02	1849.68	1656.20	1677.92	-994.41	-907.05	-900.22	-1100.73	-1018.89	-1002.26
(1.30,0.30,0.50,1.55)	268	0.33	1751.44	-1178.34	-1094.70	-1070.83	2383.02	2202.52	2176.17	-1340.51	-1261.87	-1235.54	-1492.06	-1428.66	-1387.21
(1.00,0.50,0.50,1.55)	705	0.13	4682.72	-3292.11	-3207.78	-3170.36	6606.68	6436.25	6369.71	-3628.82	-3551.59	-3498.22	-3824.45	-3748.73	-3694.14
(1.00,0.10,0.01,2.00)	105	0.85	1.35	-226.77	-201.14	-201.14	452.99	412.52	474.42	-253.67	-229.26	-254.78	-287.17	-257.38	-288.04
(1.30,0.10,0.01,2.00)	116	0.77	4.97	-281.58	-237.78	-240.62	582.26	488.99	547.95	-319.52	-277.62	-297.07	-339.79	-293.09	-320.19
(1.60,0.10,0.01,2.00)	142	0.63	2.88	-304.46	-264.86	-277.95	627.17	542.89	575.45	-352.18	-308.93	-324.74	-370.42	-324.37	-343.51
(1.90,0.10,0.01,2.00)	194	0.46	3.35	-268.73	-224.13	-240.50	565.98	487.32	524.05	-222.73	-224.26	-218.21	-336.16	-294.21	-317.10
(1.00,0.30,0.01,2.00)	205	0.43	3.19	-317.85	-272.81	-288.34	658.60	563.50	607.15	-357.65	-315.81	-336.05	-391.55	-344.59	-367.56
(1.30,0.30,0.01,2.00)	268	0.33	10.16	-333.58	-287.57	-304.56	690.19	598.68	638.47	-214.93	-176.21	-171.32	-421.65	-383.73	-402.93
(1.00,0.50,0.01,2.00)	705	0.13	59.37	-664.06	-604.98	-608.40	1348.99	1228.08	1250.18	-696.14	-630.46	-638.75	-765.46	-707.00	-712.23
(1.00,0.10,0.26,2.00)	105	0.85	2.52	-275.40	-238.64	-243.31	573.69	481.25	535.97	-306.71	-266.83	-292.80	-339.38	-301.92	-330.50
(1.30,0.10,0.26,2.00)	116	0.77	46.82	-362.78	-302.77	-307.98	742.80	612.89	669.59	-405.45	-345.65	-366.18	-428.94	-362.79	-390.90
(1.60,0.10,0.26,2.00)	142	0.63	69.05	-419.27	-348.11	-362.41	854.47	708.34	791.69	-485.09	-410.50	-445.05	-508.17	-433.26	-471.23
(1.90,0.10,0.26,2.00)	194	0.46	67.85	-576.61	-504.39	-506.08	1182.75	1030.55	1058.70	-658.91	-584.45	-579.87	-694.08	-619.67	-624.22
(1.00,0.30,0.26,2.00)	205	0.43	57.60	-627.37	-537.53	-560.94	1277.78	1089.72	1159.09	-688.79	-601.76	-632.18	-754.02	-663.08	-692.85
(1.30,0.30,0.26,2.00)	268	0.33	49.46	-821.43	-756.06	-764.41	1665.82	1528.39	1556.97	-944.19	-874.84	-881.89	-1026.15	-972.16	-970.91
(1.00,0.50,0.26,2.00)	705	0.13	140.58	-1896.95	-1832.90	-1827.15	3812.95	3687.21	3680.50	-2091.11	-2033.02	-2020.56	-2199.49	-2147.27	-2126.03
(1.00,0.10,0.50,2.00)	105	0.85	9.22	-316.87	-255.14	-261.52	655.21	514.18	595.49	-352.40	-297.31	-331.04	-390.68	-329.75	-365.00
(1.30,0.10,0.50,2.00)	116	0.77	19.52	-359.56	-293.66	-301.64	738.02	592.38	660.40	-399.50	-329.41	-354.95	-429.85	-357.67	-386.89
(1.60,0.10,0.50,2.00)	142	0.63	39.12	-468.56	-387.53	-399.14	955.46	786.66	851.89	-535.58	-448.86	-480.33	-562.57	-476.50	-510.10
(1.90,0.10,0.50,2.00)	194	0.46	26199.05	-769.03	-751.16	-683.96	1567.88	1530.14	1399.41	-852.10	-832.43	-758.97	-905.19	-888.40	-810.83
(1.00,0.30,0.50,2.00)	205	0.43	3128.96	-788.35	-768.78	-707.41	1601.33	1557.94	1444.78	-862.22	-842.68	-782.04	-946.39	-927.36	-862.29
(1.30,0.30,0.50,2.00)	268	0.33	527.41	-945.41	-853.64	-875.99	1915.44	1722.81	1787.92	-1091.13	-1002.48	-1008.46	-1195.15	-1121.35	-1128.25
(1.00,0.50,0.50,2.00)	705	0.13	3156.33	-2353.46	-2284.98	-2261.27	4723.66	4587.43	4547.75	-2585.84	-2519.81	-2487.60	-2739.28	-2682.40	-2640.27

Table A.3: Obtained diagnostic results for uncaused events spatial density $\phi_3(\cdot)$ as of Equation 4.15