Aspects of temporal and spatio-temporal processes

Rasmussen, Jakob Gulddahl

Publication date:
2007

Document Version
Publisher's PDF, also known as Version of record

Aspects of Temporal and Spatio-temporal Processes

Jakob Gulddahl Rasmussen

PhD Thesis

August 2006

Department of Mathematical Sciences
Aalborg University
Denmark

Preface

This thesis is the result of my PhD-study at the Department of Mathematical Sciences, Aalborg University, Denmark. The topic of the thesis is temporal and spatio-temporal processes.

Chapter 1 contains short introductions to point processes and lattice processes, focusing on concepts that are used in later chapters.

Chapter 2–5 contains the following papers:

2. Møller, J. & Rasmussen, J. G. (2005). Perfect Simulation of Hawkes Processes. *Advances in Applied Probability.* Vol. 37, no. 3, pp. 629-646.

3. Møller, J. & Rasmussen, J. G. (2006). Approximate Simulation of Hawkes Processes. *Methodology and Computing in Applied Probability.* Vol. 8, no. 1, pp. 53-64.

4. Zhu, J., Møller, J., Rasmussen, J. G., Aukema, B. H. & Raffa, K. F. (2006). Spatial-temporal modeling of forest gaps generated by colonization from below- and above-ground bark beetle species. *Research Report R-2006-4*, Department of Mathematical Sciences, Aalborg University.

5. Rasmussen, J. G., Møller, J., Aukema, B. H., Raffa, K. F. & Zhu, J. (2006). Bayesian inference for multivariate point processes observed at sparsely distributed times. *Research Report R-2006-24*, Department of Mathematical Sciences, Aalborg University.

The published papers and research reports have been included only changing layout, numbering, and a few minor details from the version that appears in the above list. As a consequence the notation is not consistent between the chapters, and some material is repeated in different chapters. On the other hand, the chapters may be read independently of each other.

Chapter 6 contains a section on work in progress, and since this thesis includes much programming to do the heavy computations involved, it also contains a section describing the use of some of these programs, so that others may be able to use or be inspired by these.

I wish to thank my supervisor Jesper Møller, who has proposed many of the problems and solutions addressed in this thesis and has been a great source of inspiration. I also wish to thank Jun Zhu for hosting me in my stay at the University of Wisconsin-Madison, for many useful discussions, and for collaborating with Jesper and me on several projects. I also would like to thank Brian Aukema and Kenneth Raffa for their participations in these projects. Furthermore, I would like to thank Øivind Skare for letting me use his software and for providing help with this. Finally, I wish to thank the rest of my colleagues with whom I have had many small but useful discussions on theoretical and practical issues.

Aalborg, August 2006 Jakob G. Rasmussen

Minor corrections have been made to this revised version.

Aalborg, February 2007 Jakob G. Rasmussen

Summary in English

This thesis treats temporal and spatio-temporal processes, particularly point processes and lattice processes. Such processes are useful models for many types of data of events occurring in time and space. However, inference using these processes is often difficult, even when one employs simulation. Various aspects of simulation algorithms and simulation based inference are treated in this thesis from theoretical and practical points of view.

The thesis begins with describing well-known theory of point processes and lattice processes. The focus is (marked) temporal point processes and, to a lesser extent, lattice processes, and the concepts considered consist primarily of theory that is used in later chapters. Furthermore, various simulation algorithms for point processes and lattice processes are considered. The introduction is concluded by considering various kinds of missing data, since the main issues in this thesis are strongly related to this subject.

One of the two primary issues addressed in this thesis is edge effects. Roughly speaking, edge effects means that a process is observed on a time interval (or a region in space), but what happens outside this interval may influence the process within the interval causing errors in calculations or simulation algorithms.

Specifically, simulation algorithms for a particular class of point processes called Hawkes processes are developed and examined. A Hawkes process is an important (marked) temporal point process with interesting theory and several practical applications. By viewing the Hawkes process as a Poisson cluster process with a certain

branching and conditional independence structure, a straightforward simulation algorithm for Hawkes processes is easily constructed, but this algorithm suffers from edge effects. One solution to this problem, a perfect simulation algorithm, is constructed, where "perfect" here refers to the fact that the algorithm does not suffer from edge effects. Furthermore, various propositions and theorems are proved to make this algorithm practically useful. Illuminating examples are presented showing the usefulness as well as the limitations of the algorithm, and empirical results are obtained.

The straightforward simulation algorithm is also examined in more detail, and various measures for the error committed when using this algorithm are discussed. Theoretical results are obtained for calculating these measures, and the examples used to illustrate the perfect simulation algorithm are revisited. Empirical results regarding the measures are obtained, and we show empirically that this algorithm is much faster than the perfect simulation algorithm.

The other primary issue addressed in this thesis is continuous time processes observed only at a discrete set of times. Many data sets have events that may occur at any time point on the continuous time line; that is, the natural model for such a data set would be a (marked) temporal point process. However, some of these data sets require the physical presence of an observer and are thus only observed at a rather sparse set of discrete times. In this thesis two modelling approaches for such data sets are considered. The first approach is a discrete time process, where the situation at the observation times is modelled, and the fact that the events actually happens between the observation times is ignored. The second approach is a continuous time process, where what happens between the observation times is regarded as missing data. Both approaches are explored from an applied point of view, using a data set involving a plantation of red pine trees and two types of beetles attacking these trees.

In the discrete time case, a complicated Bayesian autoregressive type of spatio-temporal model is constructed, quantifying the rela-

tions among mortality of the red pine trees and attacks of the two types of beetles including correlation across space and over time. For inference, Markov chain Monte Carlo algorithms are used, and practical results are obtained from these. Furthermore, simulation based model checking is made.

In the continuous time case, the focus is changed from the application to the methodology. The continuous and discrete time processes are compared, and the strengths and limitations of the two approaches are discussed. A subset of the data set used in the discrete case is modelled using a multivariate temporal point process, mainly with the purpose of illustrating the use of such a process in the setup with sparsely distributed observation times. We here show empirically that the continuous time processes can indeed be a useful and efficient alternative to the discrete time processes.

All of the problems treated in this thesis have required programing for calculations, and in the last part of the thesis the use of some of these programs is explained. Furthermore, three works in progress are described. These are a spatial point process model for the location of barrows in a part of Jutland, a description of the biological implications of the discrete and continuous time models for insect spread, and finally some ideas on applications for a spatial version of the Hawkes process.

Summary in Danish

Denne afhandling omhandler tids- og rumtidsprocesser, specielt punktprocesser og lattice processer. Disse processer er anvendelige som modeller for mange typer data, som består af begivenheder i rum og tid. Imidlertidigt er inferens for disse processer ofte problematisk, selv når der anvendes simulation. Forskellige aspekter af simulationsalgoritmer og simulationsbaseret inferens bliver undersøgt i denne afhandling fra teoretiske og praktiske indgangsvinkler.

Afhandlinger starter med en gennemgang af velkendt teori for punktprocesser og lattice processer. Fokus lagt på (mærkede) punktprocesser i tid og i et mindre omfang lattice processer, og de koncepter, der bliver betragtet, består primært af teori, der er brugt i senere kapitler. Desuden bliver forskellige simulationsalgoritmer for punktprocesser og lattice processer gennemgået. Introduktionen bliver afsluttet med at se på forskellige former for manglende data, da de primære problemstillinger i denne afhandling er grundet i dette emne.

Et af de to primære problemer, der bliver undersøgt i denne afhandling, er kanteffekter. Groft sagt betyder kanteffekter, at en proces bliver observeret på et tidsinterval (eller en område i rummet), men det, der sker udenfor tidsintervallet, kan influere processen indenfor tidsintervallet og skabe fejl i beregninger og simulationsalgoritmer.

Mere specifikt bliver simulationsalgoritmer for en speciel klasse af punktprocesser kaldet Hawkes processer udviklet og undersøgt. En Hawkes proces er en vigtig (mærket) punktproces på tidslinjen med

interessant teori og adskillige praktiske anvendelser. Ved at betragte Hawkes processen som en Poisson klyngeproces med visse forgrenings- og betingede uafhængigheds-strukturer, kan en simpel simulationsalgoritme for Hawkes processer let konstrueres, men denne algoritme er påvirket af kanteffekter. En løsning på dette problem, en perfekt simulationsalgoritme, bliver konstrueret, hvor "perfekt" her refererer til det faktum, at algoritmen ikke bliver påvirket af kanteffekter. Derudover bliver forskellige sætninger bevist med henblik på at gøre denne algoritme praktisk anvendelig. Illuminerende eksempler, som viser anvendeligheden samt begrænsningerne for algoritmen, bliver præsenteret, og empiriske resultater bliver opnået.

Den simple simulationsalgoritme bliver også undersøgt i flere detaljer, og forskellige mål for fejlen begået ved at bruge denne algoritme bliver diskuteret. Teoretiske resultater til at beregne disse mål bliver lavet, og eksemplerne brugt til at illustrere den perfekte simulationsalgoritme bliver betragtet igen. Empiriske resultater omkring målene bliver lavet, og vi viser empirisk, at denne algoritme er meget hurtigere end den perfekte simulationsalgoritme.

Den anden primære problemstilling i denne afhandling er stokastiske processer i kontinuert tid observeret kun til diskrete tider. Mange datasæt har begivenheder, der kan ske på et hvilket som helst tidspunkt på den kontinuerte tidslinje; det vil sige, den naturlige model for et sådan datasæt er en (mærket) punktproces på tidslinjen. Dog kræver nogle datasæt den fysiske tilstedeværelse af en observatør, og bliver derfor kun observeret på et ret begrænset antal diskrete tidspunkter. I denne afhandling bliver to metoder til modellering af sådanne datasæt betragtet. Den første metode er processer i diskret tid, hvor situationen til observationstiderne bliver modelleret, og det faktum, at begivenheder rent faktisk finder sted mellem observationstiderne, bliver ignoreret. Den anden metode er processer i kontinuert tid, hvor det, der sker mellem observationstiderne, bliver betragtet som manglende data. Begge metoder bliver undersøgt fra et anvendt synspunkt, hvor der bruges et datasæt, der involverer en plantage af fyrtræer og to typer biller, som angriber disse træer.

I tilfældet med diskret tid bliver der konstrueret en kompliceret Bayesiansk autoregressiv type af rumtidsmodel, som kvantificerer relationerne mellem trædødelighed og billeangreb og inkluderer korrelation over rum og tid. Til inferens bruges Markov kæde Monte Carlo algoritmer, og der opnås praktiske resultater ved hjælp af disse. Desuden bliver der lavet simulationsbaseret model tests.

I tilfældet med kontinuert tid bliver fokus ændret fra anvendelse til metode. Tilfældene med kontinuert og diskret tid bliver sammenlignet, og styrkerne og begrænsningerne af de to indgangsvinkler bliver diskuteret. En delmængde af datasættet brugt i forbindelse med det diskrete tilfælde bliver modelleret ved hjælp af en multivariat punktproces i tid, primært med det formål at illustrere brugen af denne type proces i tilfældet med et lille antal observationstider. Vi viser her empirisk, at processer i kontinuert tid virkeligt kan være et brugbart og effektivt alternativ til processer i diskret tid.

Alle problemstillingerne i denne afhandling har krævet programmering til beregningerne, og i den sidste del af afhandlingen beskrives nogle af disse programmer. Desuden bliver der beskrevet tre projekter under udarbejdelse. Disse er en rumlig punktproces til modellering af placeringen af gravhøje i en del af Jylland, en beskrivelse af de biologiske konsekvenser af modellerne for insektspredning i kontinuert og diskret tid, og sidst nogle ideer til anvendelser af en rumlig version af Hawkes processen.

Table of Contents

Chapter 1

Introduction

1.1 Temporal and spatio-temporal processes

Many data sets consists of events or objects scattered randomly throughout time and space. In order to obtain an understanding of such data sets, it is important to construct realistic models of them. In this thesis I work with various aspects of two general classes of stochastic processes for modelling such data sets: point processes and lattice processes.

A point process is a stochastic process whose realisations are patterns of points in some arbitrary set. In practice, this set is often a subset of the time line, in which case I use the term temporal point process, or a subset of the physical space, in which case I use the term spatial point point process. In a temporal point process the points represent the times of events (and are thus often referred to as events). Many types of data can be modelled by such processes, for example times of earthquakes or other disasters, arrivals of customers in a queuing system, or failures in a computer network. In a spatial point process the points represent the location of events or objects. In applications the points usually fall within some subset of

$\mathbb{R}^2$ or $\mathbb{R}^3$. Examples of data in $\mathbb{R}^2$ include the position of trees in a forest, and the position of plants infected with a certain disease, and in $\mathbb{R}^3$ an example is the position of stars. Note that the term spatial point process is often used for a point process defined on a general space (including temporal point processes), but to keep the distinction between point processes in time and in space clear, I will use it to refer to processes modelling locations in the physical space. Since temporal point processes are the primary focus of this thesis, I will often work with the temporal cases and just briefly discuss the spatial cases. In Section 1.2, I introduce some theory of point processes; for more comprehensive introductions to the subject, see Møller and Waagepetersen (2004) or Daley and Vere-Jones (2003).

In applications we often have other information connected to an event or object which is also of interest or which has impact on other events or objects. For example, when modelling earthquakes, the magnitude of the earthquake and the location of the epicentre may be included to make a more realistic and useful model, or when modelling the locations of trees, we may wish to include some measure of the size of a tree. A point process including such additional information is called a marked point process. If a process includes information on both time and space, we call it a spatio-temporal (or spatial-temporal) point process. In Section 1.2.3, some theory of marked point processes is described; again more comprehensive introductions can be found in Møller and Waagepetersen (2004) or Daley and Vere-Jones (2003).

Unlike point processes, which are (usually) used for modelling events or objects on a continuous space, lattice processes are used for modelling events or objects on a discrete space. Example of data modelled by a lattice process include the number of beetles attacking each tree in a plantations, or the number of cases of a disease in different regions of a country. The spatio-temporal lattice processes which we will encounter in this thesis are extensions of a spatial lattice process, so I will focus on the spatial case in this chapter. Lattice processes are treated in Section 1.3; for a more thorough treatment,

see Besag (1974).

1.2 Point processes

1.2.1 Definition

One way of defining a *point process* is by using counting measures. If we think of a *temporal point process* as a random countable set of times (or events) $X = \{t_i\}$, where each t_i belongs to some subset of the real numbers, say $S \subseteq \mathbb{R}$, we can define such a point process by the number of events located in various subsets of S. More precisely, let $\mathcal{B}$ denote the set of Borel sets in S, and let $N(B)$ be the number of events falling in any $B \in \mathcal{B}$. Furthermore, we restrict the attention to locally finite point processes, i.e. if $\mathcal{B}_0$ denotes the bounded Borel sets in S, then $N(B) < \infty$ for any $B \in \mathcal{B}_0$. Technically speaking, N is an stochastic counting measure on $\mathcal{B}$.

There are in fact many ways of defining a point process, for example at page 41 in Daley and Vere-Jones (2003) four equivalent ways of defining a temporal point process are shown, including the above definition where the point process is defined by a stochastic counting measure. One advantage of using counting measures is that this definition immediately extends to *spatial point processes*: simply let S be a subset of $\mathbb{R}^d$ instead of $\mathbb{R}$ (and in this case the points are called x_i instead of t_i and are not referred to as events).

1.2.2 The temporal dimension

Usually we think of time as having an *evolutionary character*: what happens now may depend on what happened in the past, but not on what is going to happen in the future. Many important classes of temporal point processes also has such an evolutionary character, where the natural order of time is respected. I will use the term *evolutionary (temporal) point process* for these processes. The central

point of this section is that such processes can be described using the so-called *conditional intensity function*.

To understand what the conditional intensity function is, we first have to define the *history* of the process. The history $\mathcal{H}_{t-}$ is the σ-algebra of events that has occurred strictly before time t ($t-$ denotes the "time just before time t" and should not be confused with t_- used in Section 1.5 and Chapters 2 and 3). In practice we can usually just think of this as the set of times of all events that have occurred before t. Using the history, the $\lambda(t)$ can be defined somewhat heuristically:

$$\lambda(t) = \mathbb{E}[N(\mathrm{d}t)|\mathcal{H}_{t-}]/\mathrm{d}t,$$

i.e. the risk of an event occurring at time t conditional on what has occurred before t. Note that the dependence of the conditional intensity on the history is suppressed in the notation. For a more strict definition of the conditional intensity function, see Daley and Vere-Jones (2003).

The conditional intensity function turns out to have many uses. Firstly, it is a convenient way of defining an evolutionary temporal point process, since it describes what is happening locally at time t and often is fairly easy to interpret. For example, the model used in Chapter 5 is defined by specifying a conditional intensity function that fits various aspects of the data set in that chapter. Secondly, the conditional intensity function can be used for simulation-based model checking and prediction, since some simulation algorithms are based on the conditional intensity function (see Section 1.4.2). Thirdly, the likelihood function can be expressed on closed form using the conditional intensity function; if the point process is defined on the interval $S = [0, t_+)$ for some fixed $t_+ > 0$, then the likelihood function is given by

$$L = \left(\prod_{i=1}^{N([0,t_+))} \lambda(t_i) \right) \exp\left(- \int_0^{t_+} \lambda(s)\mathrm{d}s \right).$$

Finally, there are other uses of the conditional intensity function, for example a goodness-of-fit test known as residual analysis for point

processes (Ogata, 1988), or the distribution of the length of the time intervals between subsequent events (see e.g. Daley and Vere-Jones (2003)). It should be obvious from all of this that the conditional intensity function is a powerful tool for dealing with evolutionary temporal point processes.

Not all point processes have a temporal (or similar) dimension with an evolutionary character. Purely spatial point processes usually have no natural order on the space. As a result of this, there is no such thing as a history or a conditional intensity; actually there is a similar concept called the Papangelou conditional intensity, see Papangelou (1974), but, although it is quite useful, it does not have the same general usage as the conditional intensity described here. Thus we do not have an easy and general way of obtaining likelihood functions, simulation algorithms, etc., but we can usually get around the problems caused by this. For example, for many classes of spatial point processes the likelihood is known only up to an *unknown normalising constant*. There are many ways of approximating or avoiding unknown normalising constants when the likelihood function is used in practice, for example importance sampling (see e.g. Møller and Waagepetersen (2004)), bridge or path sampling (Gelman and Meng, 1998), and auxiliary variables (Møller *et al.*, 2006), but they are usually rather computationally heavy.

1.2.3 Marked point processes

Sometimes we wish to include other information into a temporal point process (or spatial point process; see e.g. Møller and Waagepetersen (2004)) about an event than just its time, since this other information may have impact on the times of later events or it may be of separate interest. This extra information about an event is handled by introducing *marks*: to each event t_i is attached a mark $z_i \in M$ where M is a probability space called a mark space equipped with a probability measure Q. Furthermore, the mark space is also equipped with a reference measure ℓ_M chosen such that Q has a density with

respect to ℓ_M. Typically the reference measure is the Lebesgue measure if $M = \mathbb{R}^d$, or counting measure if M is a discrete space. For marked processes, $N(B)$ is redefined to be the number of marked events $(t_i, z_i) \in B$ where B is a measurable set in $\mathbb{R} \times M$, and the *ground process*, i.e. the marginal process consisting of the unmarked events, is denoted N_g.

Much of the theory of temporal point processes generalises to the marked case, including the theory of evolutionary temporal point processes described in Section 1.2.2. For evolutionary marked temporal point processes, the conditional intensity generalises to

$$\lambda(t, z) = \mathbb{E}[N(\mathrm{d}t \times \ell_M(\mathrm{d}z))|\mathcal{H}_{t-}]/(\mathrm{d}t\ \ell_M(\mathrm{d}z)),$$

where the history now contains information on the marks as well as times of past events; for a strict definition of the conditional intensity function for a marked temporal point process, see Daley and Vere-Jones (2003). The *ground intensity*, i.e. the conditional intensity for the ground process, is denoted λ_g for marked processes. It is often convenient to factorise the conditional intensity into the ground intensity and the conditional density with respect to ℓ_M for the mark given the time t and the history $\mathcal{H}_{t-}$,

$$\lambda(t, z) = \lambda_g(t) f(z|t).$$

Note that just as in the case with the conditional intensities, the dependence on the past in the conditional mark density has been suppressed in the notation. If the marked point process is defined on $[0, t_+) \times M$ for some fixed $t_+ > 0$, then the likelihood function is given by

$$L = \left(\prod_{i=1}^{N_g([0,t_+))} \lambda(t_i, z_i) \right) \exp\left(- \int_0^{t_+} \lambda_g(s) \mathrm{d}s \right).$$

Most of the other uses of the conditional intensity function for the unmarked case also generalises to the marked case, for example one

of the simulation algorithms in Section 1.4.2 is described for marked processes. See Daley and Vere-Jones (2003) for more details on the conditional intensity function for the marked case.

If the distribution of an arbitrary mark z_i is independent of the past, i.e. (t_j, z_j) for all $t_j < t_i$, then the marks are said to be *unpredictable*. Unpredictable marks may depend on the future times (but not marks), or, reformulated in the spirit of the evolutionary character of time, the future times may depend on the unpredictable marks. Unpredictable marks are particularly easy to handle, since we can treat them as a sequence of independent random variables.

An important special case of marked temporal point processes is *spatio-temporal point processes*, where the marks denote spatial locations. A spatial location typically means a point in a rectangle or a lattice in $\mathbb{R}^2$ or $\mathbb{R}^3$. Another important special case is *multivariate temporal point processes*, where we have multiple dependent temporal point processes. Mathematically a multivariate temporal point process is simply a marked temporal point process where the mark space has a finite number of states.

1.2.4 Examples of point processes

There are many types of point processes, and I will only consider a few examples below. For comprehensive introductions to different classes of point processes, see e.g. Møller and Waagepetersen (2004), Daley and Vere-Jones (2003), or van Lieshout (2000).

1.2.4.1 Poisson processes

The most basic example of temporal (or spatial) point processes is the *Poisson process*. It is best thought of as a model for events occurring independently of each other. To define it, let λ be a non-negative measurable function on S. A point process is a Poisson process with *intensity function* λ if for any disjoint $B_1, \ldots, B_n \in \mathcal{B}_0$, $N(B_1), \ldots, N(B_n)$ are independent and $N(B_i)$ is Poisson distributed

with mean $\int_{B_i} \lambda(t)\mathrm{d}t$ for $i = 1, \ldots, n$. The intensity function turns out to be the conditional intensity function and hence I use the notation λ for both functions. If λ is constant, the process is called a *homogeneous* Poisson process, otherwise it is *inhomogeneous*. A homogeneous temporal Poisson process with intensity λ has a particularly simple characterisation: the time intervals between subsequent events are independent exponentially distributed random variables with inverse mean λ. This characterisation is very useful for constructing a simulation algorithm for a homogeneous temporal Poisson process (see Section 1.4.1).

Poisson processes have the advantage of being simple - many interesting quantities can be analytically derived. However, Poisson processes are not very useful for modelling real data, since all events are assumed to happen independently. They are, however, a very useful starting point for defining other processes; for example, the Poisson process is used to define the class of Hawkes processes in Chapters 2 and 3.

Poisson processes are also used as a reference mark for defining the concepts of a *clustered* or *regular* point process. If the points of a point pattern tend to fall in clusters more than a Poisson process we say that the point pattern is clustered. Conversely, if the points instead tend to be more evenly distributed than a Poisson process, we say that it is regular. Figure 1.1 shows a realisation of a temporal Poisson process. Furthermore, two clustered point processes are shown; these processes are introduced in Sections 1.2.4.2 and 1.2.4.3. From the figure it is clear that the events of the two clustered processes tend more to fall in clusters than the events of the Poisson process.

Poisson processes are a huge subject, and e.g. Kingman (1993) is dedicated to the study of these. Furthermore, many books on point processes have a chapter on Poisson processes, see e.g. Møller and Waagepetersen (2004) or Daley and Vere-Jones (2003).

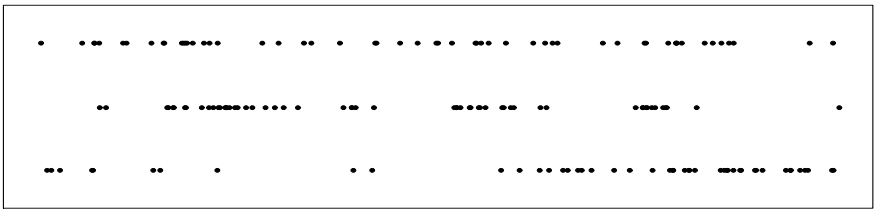


Figure 1.1: Three temporal point processes: a Poisson process, a shot noise Cox process and a Hawkes process (top to bottom).

1.2.4.2 Cox processes

A rich class of temporal (or spatial) point processes that is constructed from Poisson processes is *Cox processes* (or doubly stochastic Poisson processes), which were originally defined in Cox (1955). It is a natural generalization of the Poisson process obtained by letting the intensity function λ be stochastic; we say that the Cox process is driven (or directed) by the intensity function λ. A very simple example of a Cox process is obtained by letting the intensity of a homogeneous Poisson process be a nonnegative random variable.

More interesting examples include *shot noise Cox processes* (also called trigger processes). Following Daley and Vere-Jones (2003), a temporal shot noise Cox process is driven by

$$\lambda(t) = \sum_{s_i<t} \alpha_i g(t - s_i),$$

where g is a density function on $(0, \infty)$, α_i is a non-negative random variable and $\{s_i\}$ is a homogeneous Poisson process on $\mathbb{R}$; g is sometimes called a kernel. Figure 1.1 shows a realisation of a shot noise Cox process with constant α_i and the density of the exponential distribution as the kernel g. For more on temporal shot noise Cox processes, see Daley and Vere-Jones (2003), and for a spatial (and indeed more general) version, see Møller (2003).

Another interesting example is *log Gaussian Cox processes*. The log Gaussian Cox processes are defined by their intensity, which is the exponential of a Gaussian field. The theory of log Gaussian processes can be found in Møller *et al.* (1998).

1.2.4.3 Cluster processes

A *cluster process* is a point process intended for modelling clustered point patterns by constructing clusters individually. The definition of a cluster process is divided into two parts. Firstly a point process known as a centre process is generated. Then each point (or centre) in the centre process generates a new point process called a component process or a cluster. The cluster process consists of the superposition of the clusters, where the centres may or may not be included. The centres are also sometimes called ancestors, parents or immigrants, the points in the clusters are sometimes called children or offspring, and if the centres follow a Poisson process, the cluster process is called a *Poisson cluster process*.

An example of a Poisson cluster process is the *Thomas process* (Thomas, 1949). Here the clusters are i.i.d. relative to their centres and each cluster is itself a Poisson process where the intensity function is proportional to the normal density function with the cluster centre as mean and some fixed variance. Another example is the shot noise Cox process described in Section 1.2.4.2.

An example of a Poisson cluster process where the clusters are not simply Poisson processes is the *Hawkes process* (or self-exciting process). There are two equivalent ways of defining the Hawkes process. The first way is by specifying its conditional intensity function; for example, in the unmarked case, this is given by

$$\lambda(t) = \mu(t) + \sum_{t_i<t} \gamma(t - t_i),$$

where μ and γ are non-negative, measurable functions and $\gamma(t) = 0$ for $t \leq 0$. The conditional intensity function (or, more precisely,

the ground intensity) for the marked case is given by formula (2.1) in Chapter 2. The other way is to use a construction of clusters with a certain branching structure of Poisson processes; this is done in Definition 2.1 in Chapter 2. Figure 1.1 shows a realisation of an unmarked Hawkes process, where γ is an exponentially decaying function on $[0, \infty)$, and μ is constant. Since the Hawkes process is the focus of Chapters 2 and 3, I will postpone all further details until these chapters. For more theory on Hawkes processes, see Hawkes (1971a,b, 1972); Hawkes and Oakes (1974); Brémaud and Massoulié (2001, 2002); Torrisi (2002), and for applications of Hawkes processes, see e.g. Vere-Jones and Ozaki (1982); Chornoboy *et al.* (1988); Ogata (1988, 1998). A purely spatial version of this process also exists, see Møller and Torrisi (2005).

1.3 Lattice processes

1.3.1 Definition

A *lattice process* (or lattice model) is used to model a system of random variables associated to a set of sites, where sites for example could represent a fixed set of physical locations or times. Basically, a lattice process is just a random vector X, where each entry in the vector x_i correspond to a site i. Thus we could define a lattice process by its joint distribution. However, as Besag (1974) points out, it is usually more convenient to specify a lattice process by the conditional distribution P_i of the random variable x_i at a site i given the set of all other sites. In Chapter 4 we refer to this as the *local characteristic* since it describes the local behaviour of the process. Note that P_i has much the same role as the conditional intensity defined for temporal processes in Section 1.2.2. Much like the conditional intensity, it provides an easy way of defining a model on a complex system by defining what is happening locally. However, it does not have the evolutionary character which makes the conditional intensity so useful.

In practice, the conditional distribution P_i is usually assumed to depend only on sites that are located close to site i. For each site i let N_i be the *neighbourhood* of i, i.e. the set of sites that are close to i. Of course what "close" means depend on the application at hand. For example, on a regular square lattice it could mean the four nearest sites; this is called the first order neighbourhood. Figure 1.2 illustrates the zeroth- to fifth-order neighbourhoods on a five times five regular square lattice. Here the sites are indicated by numbers which tell the order of the neighbourhood that the site belongs to in relation to the site in the middle; for convenience the middle site is said to be a zeroth order neighbour of itself. Finally, note that a neighbourhood relation should be symmetric: $i \in N_j$ if and only if $j \in N_i$ for all pairs of sites (i, j).

5	4	3	4	5
4	2	1	2	4
3	1	0	1	3
4	2	1	2	4
5	4	3	4	5

Figure 1.2: The zeroth- to fifth-order neighbours of the middle site in a five times five regular square lattice.

1.3.2 Auto-models

An important class of lattice processes is the class of *auto-models*. Auto-models can be specified by letting the conditional distribution

of a site i given its neighbours $P_i(x_i|N_i)$ have the form

$$P_i(x_i|N_i) \propto \exp\left(x_i G_i(x_i) + \sum_{j \in N_i} \beta_{i,j} x_i x_j\right),$$

where G_i is a real-valued function and $\beta_{i,j} \in \mathbb{R}$ fulfills that $\beta_{i,j} = \beta_{j,i}$ for each pair of sites (i, j).

One example of an auto-model is the *autologistic model*. For each site i, let x_i be Bernoulli distributed and $G_i(x_i) = \alpha_i$ be constant. The autologistic model can then be specified through its conditional distribution at site i,

$$P_i(x_i|N_i) = \frac{\exp(\alpha_i x_i + \sum_{j \in N_i} \beta_{i,j} x_i x_j)}{1 + \exp(\alpha_i + \sum_{j \in N_i} \beta_{i,j} x_j)}.$$

Locally $\beta_{i,j}$ is easily interpreted: a positive $\beta_{i,j}$ means that x_i will tend to have the same value as x_j, whereas a negative $\beta_{i,j}$ means it will tend to have a different value than x_j. Globally an autologistic model with numerically large, positive values for $\beta_{i,j}$ has large clusters of zeros and ones. In the case of a first order neighbourhood on a square lattice, an autologistic model with numerically large, negative values for $\beta_{i,j}$ tends to have chess board patterns. These two cases of patterns of zeros and ones are analogous to clustered and regular point patterns. The size of the parameter space is usually reduced for this model to be practically useful, for example by setting $\beta_{i,j} = \beta$ for all neighbour pairs (i, j) and $\alpha_i = \alpha$ for all i. Figure 1.3 shows two realisations of autologistic processes with a first order neighbourhood on a 20×20 lattice; on the left-hand side β is positive, and on the right-hand side β is negative.

One problem associated with the autologistic model is that the likelihood function is known only up to an unknown normalizing constant; by the Hammersley-Clifford theorem (see e.g. Besag (1974)), the likelihood function can be shown to be

$$L = \frac{1}{c} \exp\left(\sum_i \alpha_i x_i + \sum_i \sum_{j \in N_i, j > i} \beta_{i,j} x_i x_j\right),$$

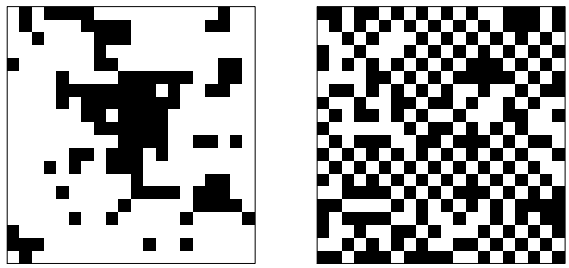


Figure 1.3: Left: a realisation of the autologistic model with positive $\beta_{i,j} = \beta$. Right: as left, but with negative $\beta_{i,j} = \beta$.

where c denotes the normalizing constant. It is only feasible to calculate the normalizing constant through brute force in cases with small lattices. Otherwise techniques such as those mentioned at the end of Section 1.2.2 are needed.

There are of course many other auto-models than the autologistic model, for example the auto-Poisson model, where the conditional distribution of x_i is Poisson. For details on this and other auto-models, see Besag (1974).

1.4 Simulation

Most point processes of any applicational value are quite complex stochastic processes, and often it is impossible to derive quantities of interest analytically. Instead one must turn to simulation algorithms for approximating such quantities. Simulation is an invaluable tool for example in parameter estimation, model checking, and prediction. Even something as simple as getting an idea of what a typical realisation of a particular type of point process looks like may require the use of a simulation algorithm.

There are many different simulation algorithms available for simulating a point process with a given set of parameters; many of these

algorithms are tailored to specific classes of point processes, while others are rather general but not always very efficient. I will in no way try to give a comprehensive introduction to the subject; instead I will focus on algorithms that are relevant to the later chapters.

1.4.1 Poisson and related processes

Since the Poisson process is one of the simplest examples of point processes, it is probably not much of a surprise that it is also one of the simplest processes to simulate. The homogeneous temporal Poisson process with intensity λ is easily simulated on an interval $[0, t_+)$ by using that the lengths of the time intervals between subsequent events are i.i.d. exponentially distributed random variables. First simulate i.i.d. exponential variables Y_i with inverse mean λ. Each event are then given by $t_i = \sum_{j=1}^{i} Y_j$. The algorithm terminates once the end of the time interval t_+ has been reached. Another way of simulating a homogeneous Poisson process with intensity λ is as follows: first simulate the number of events as a Poisson variable with mean λt_+ and then simulate the times of the events as i.i.d. uniform variables on $[0, t_+)$. This latter method generalizes directly to the spatial Poisson process, where we need to simulate uniform variables on the simulation region.

Inhomogeneous Poisson processes can be simulated in several ways depending on the shape of the intensity function. A quite general way is *independent thinning*. Independent thinning means that we take a point process $X_1 = \{t_i\}$ and obtain another process X_2 by independently keeping each event t_i from X_1 with a specified probability $p(t_i)$ (called the retention probability), or otherwise throw t_i away. If X_1 is a Poisson process with intensity function $\lambda_1(t)$, then X_2 is a Poisson process with intensity $\lambda_2(t) = p(t)\lambda_1(t)$ (see e.g. Møller and Waagepetersen (2004)). This means we can simulate an inhomogeneous Poisson process X_2 with a bounded intensity $\lambda_2(t)$ by using independent thinning on a simulation of a homogeneous Poisson process X_1 with intensity $\lambda_1 = \sup(\lambda_2(t))$. For specific functional forms

of the intensity function, it may be more convenient to use other algorithms. For example, if the intensity has the shape of an unnormalized exponential density function, then it is easiest to simulate it by first simulating the number of events as the appropriate Poisson variable, and then simulate the time of each event as i.i.d. exponential variables.

Since many processes are defined using the Poisson process as a starting point, it is often possible to simulate these directly by exploiting the definition. For example, conditionally on the intensity, a Cox process is a Poisson process, so if we can simulate the intensity, then we can simply simulate the Poisson process afterwards using one of the above techniques. Another example is the Hawkes process. If the Hawkes process is defined using Poisson processes as in Chapter 3, then Algorithm 3.1 is an straightforward procedure for simulating the process once one knows a simulation algorithm for the Poisson processes used in the definition.

1.4.2 Simulation using the conditional intensity

One of the reasons that the conditional intensity function is so useful is that it leads to two rather general simulation algorithms for evolutionary temporal point processes.

One of these algorithms is *Ogata's modified thinning algorithm* (Ogata, 1981), or rather, its generalization to marked processes (see e.g. Daley and Vere-Jones (2003), page 273). This algorithm is a generalization of the independent thinning algorithm used for simulating inhomogeneous Poisson processes as described in Section 1.4.1. The basic idea behind Ogata's modified thinning algorithm is that we start at the first time in the simulation interval (e.g. time 0 in the interval $[0, t_+)$) and go forward in time. We propose new events one after another, and check right away whether to keep an event, since this changes the conditional intensity function for later times. More specifically, the algorithm requires that there exists two functions $m(t) = m(t|\mathcal{H}_{t-})$ and $l(t) = l(t|\mathcal{H}_{t-})$ such that $\lambda_g(t+u) \leq m(t)$

for $0 \leq u < l(t)$. The function $l(t)$ is the maximum length of time we will go forward in time in one step of the algorithm, and $m(t)$ is the maximum value that the conditional intensity may attain in this interval. For simulating a process on the interval $[0, t+)$, the algorithm for the marked case is as follows:

Algorithm 1.1 (*Ogata's modified thinning algorithm.*)

1. Set t=0 and n=0.
2. Repeat until $t > t_+$:
   (a) Compute $m(t)$ and $l(t)$.
   (b) Generate independent random variables $T \sim \text{Exp}(m(t))$ and $U \sim \text{Unif}([0, 1])$.
   (c) If $T > l(t)$, set $t = t + l(t)$.
   (d) Else if $t + T > t_+$ or $U > \lambda_g(t + T)/m(t)$, set $t = t + T$.
   (e) Otherwise, set $n = n+1$, $t_n = t+T$, $t = t+T$ and simulate $z_n \sim Q$.
3. Output is $\{t_1, \ldots, t_n\}$.

Here $\text{Exp}(m(t))$ denotes the exponential distribution with inverse mean $m(t)$, and $\text{Unif}([0, 1])$ denotes the uniform distribution on the interval $[0, 1]$. This algorithm is very useful for simulating the point process used in Chapter 5, but it is much less efficient than Algorithm 3.1 in Chapter 3 for simulating Hawkes processes.

The other algorithm that uses the conditional intensity function for simulating evolutionary temporal point processes is *simulation by inversion*. The basic idea in this algorithm is that we simulate a number of unit exponential variables, and then we transform these into time intervals for the process we are simulating. Assuming that we want to simulate the process on $[0, t+)$, the algorithm is as follows:

Algorithm 1.2 (*Simulation by inversion.*)

1. Set $t = 0$, $t_0 = 0$ and $n = 0$ (note that t_0 is not an event).
2. Repeat until $t > t_+$:
   (a) Generate $T \sim \text{Exp}(1)$.
   (b) Calculate t, where $T = \int_{t_n}^{t} \lambda(s)\mathrm{d}s$.
   (c) If $t < t_+$, set $n = n + 1$ and $t_n = t$.
3. Output is $\{t_1, \ldots, t_n\}$.

The difficult part of this algorithm is of course calculating t in step (2b) since this requires finding the inverse of the integrated conditional intensity function.

1.4.3 Simulation using MCMC

It often happens that a stochastic process is too complex to simulate directly, but when all else fails, one can usually turn to *Markov chain Monte Carlo (MCMC)* techniques for simulating a point process. Spatial point processes and spatial lattice processes are two types of stochastic processes that frequently require the use of an MCMC algorithm for simulation. Furthermore, inference for most point processes or lattice processes is often not possible analytically. Instead MCMC based simulation is useful for obtaining, for example, an approximation of the posterior distributions of parameters in the case of Bayesian inference (for an introduction to Bayesian inference, see e.g. Gelman *et al.* (2004)). I will assume that the concepts of MCMC is well-known to the reader; for an introduction to Markov chains, see e.g. Meyn and Tweedie (1993) or Norris (1997), and for an introduction to MCMC in practice with a focus on posterior distributions, see Tierney (1994).

Since we never meet any point process that requires the use of an MCMC algorithm for simulation in this thesis, I will not go

into details with simulation algorithms for these; see e.g. Møller and Waagepetersen (2004) or van Lieshout (2000) for introductions to this subject. Instead I will consider the following algorithm for simulating a lattice process, since this is used in Chapter 4. We start at some configuration (say all $x_i = 0$) and use a *Gibbs sampler* where we update each x_i one at a time in some random or fixed order. Each x_i is updated by drawing a new value for x_i from it conditional distribution given the state of all the other sites P_i.

One of the problems with MCMC based approaches is that in principle the Markov chains have to run infinitely long before reaching the target distribution. One solution to this problem is to put the MCMC algorithm into the framework of *perfect simulation*. By perfect simulation is meant that the output of the algorithm follows the target distribution exactly. In Chapter 4, the Gibbs sampler described above is combined with a perfect simulation technique known as Propp-Wilson's coupling from the past (Propp and Wilson, 1996) to obtain perfect simulations. Other examples of perfect simulation include dominated coupling from the past (Kendall and Møller, 2000), Wilson's read-once algorithm (Wilson, 2000), and perfect simulation using clans of ancestors (Fernández *et al.*, 2002). For an overview of various perfect simulation algorithms in relation to point processes, see Møller and Waagepetersen (2004).

1.5 Missing data

Often it is the case that a useful model needs more information than the observed data provides. That is, we have a so-called *missing data* problem. In this thesis, missing data in various forms are central parts of the problems studied, and therefore I will introduce the two missing data problems which will be treated in Chapters 2–5.

1.5.1 Edge effects

Many point processes are defined on an infinite set, e.g. the time line $\mathbb{R}$, but in practice we only observe data on a finite observation region. However, the observed data may depend on the unobserved data outside the observation region. If we try to simulate new data on the observation region without taking this fact into account, then the simulation will not be exact. Furthermore, calculations of various functions, such as the likelihood function, may also contain errors. The errors resulting from only observing the process on a limited region without taking the region outside the observation region into account are known as *edge effects*. Many kinds of stochastic processes suffer from edge effects, but I will only consider point processes, and furthermore I will focus on edge effects in relation to simulation.

Evolutionary temporal point processes are usually simulated on a time interval, say $[0, t_+)$, but what happens during this time interval may depend on the past. For example, we may wish to simulate a temporal point process model for the times of earthquakes and their aftershocks aftershocks. If we simply simulate the times of earthquakes and aftershocks on $[0, t_+)$ and completely ignore that there may have been earthquakes before time 0, then the aftershocks of earthquakes occurring before time 0 will be missing. Results based on such a simulation will be biased and may lead to the wrong conclusions. Since simulation is an important tool for prediction and model checking for point processes, methods are needed for dealing with edge effects.

An easy, but only approximate, method for simulating a point process without edge effects is to simulate it on a larger region. In the case of an evolutionary temporal point process on $[0, t_+)$, this means that we would start our simulation at some time before time 0, say $t_- < 0$; the question is what value of t_- we should choose to obtain a simulation without significant edge effects. Algorithm 3.1 in Chapter 3 simulates a Hawkes process by starting at time t_-. In that chapter, various ways of quantifying edge effects are considered

to obtain an idea of what t_- to choose to make a simulation with insignificant edge effects (see also Brémaud and Massoulié (2002) and Brémaud *et al.* (2002)). In principle, most simulation algorithms, for example Ogata's modified thinning algorithm, can be used on a larger region, but in general it is difficult to determine how large the simulation region should be.

There are also algorithms which can simulate specific classes of point processes completely without edge effects. Brix and Kendall (2002) (see also Møller (2003)) introduces a clever way of simulating point processes that are simultaneously Poisson cluster processes and Cox processes without any edge effects. They simulate exactly those clusters which have one or more events inside the simulation region, whether or not the cluster centre is inside the simulation region. In Chapter 2 we make a non-trivial extension to this simulation method and obtain a simulation algorithm for the Hawkes process that does not suffer from edge effects.

There are many other methods for dealing with edge effects in various contexts. Møller and Waagepetersen (2004), Daley and Vere-Jones (2003), and some of the references therein provide various techniques of dealing with edge effects in simulations and calculations, e.g. minus sampling, edge correction factors, and periodic boundaries.

1.5.2 Continuous time processes observed at discrete times

Stochastic processes may be defined on a continuous subset of the time line $\mathbb{R}$, but sometimes it may only be possible to observe the process on a discrete set of observation times. For example, an event may be that a tree has become infested with insects, but in order to detect the time of this event, we will need to observe the tree continuously, which may not possible in practice. If the data is only observed at discrete times, it only provides partial information about the events, for example whether or not one or more events have occurred between two observation times. Obviously, such a missing

data setup of discrete observation times can apply to many types of continuous time stochastic processes, but in this thesis the attention will be restricted to multivariate temporal point processes.

There are several ways of dealing with this missing data problem for a multivariate temporal point process. One way is simply to ignore that the observed and unobserved data are continuous and instead model only the observed data using a discrete time process, i.e. a lattice process can be used instead of a multivariate temporal point process. This seems to be the standard approach for modelling such systems, and has for example been done in Besag and Tantrum (2003). However, this approach is not unproblematic. In Chapter 4, a complex system of trees and insects with annual observations is modelled using a spatio-temporal autoregressive type of model. However, in this particular model unknown normalising constants appear in the likelihood function, which complicates the computations considerably.

Another way is to specify a continuous time model and then simulate the missing data using MCMC. In general, this approach may seem to be conceptually and computationally more difficult than the discrete time approach, and this approach seems to be avoided in the literature. However, in some cases the inclusion of continuous time may pay off. In Chapter 5, a part of the discretely observed data from Chapter 4 is modelled using a multivariate temporal point process. There the inclusion of continuous time means that the conditional intensity function (see Section 1.2.2) can be used for specifying a model, which implies that an expression for the likelihood function is available without any unknown normalizing constants. In Chapter 5 the advantages and disadvantages of the discrete and continuous time approaches to the missing data are discussed.

References

Besag, J. and Tantrum, J. (2003). Likelihood analysis of binary data in space and time. In P. J. Green, N. L. Hjort, and S. Richardson, editors, *Highly Structured Stochastic Systems*, pages 289–295. Oxford University Press, Oxford.

Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society Series B*, **36**, 192–236.

Brémaud, P. and Massoulié, L. (2001). Hawkes branching point processes without ancestors. *Journal of Applied Probability*, **38**, 122–135.

Brémaud, P. and Massoulié, L. (2002). Power spectra of general shot noises and Hawkes point processes with a random excitation. *Advances in Applied Probability*, **34**, 205–222.

Brémaud, P., Nappo, G., and Torrisi, G. (2002). Rate of convergence to equilibrium of marked Hawkes processes. *Journal of Applied Probability*, **39**, 123–136.

Brix, A. and Kendall, W. (2002). Simulation of cluster point processes without edge effects. *Advances in Applied Probability*, **34**, 267–280.

Chornoboy, E. S., Schramm, L. P., and Karr, A. F. (1988). Maximum likelihood identification of neural point process systems. *Advances in Applied Probability*, **34**, 267–280.

Cox, D. R. (1955). Some statistical models related with series of events. *Journal of the Royal Statistical Society Series B*, **17**, 129–164.

Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods*. Springer, New York, 2nd edition.

Fernández, R., Ferrari, P. A., and Garcia, N. L. (2002). Perfect simulation for interacting point processes, loss networks and ising models. *Stochastic Processes and Their Applications*, **102**, 63–88.

Gelman, A. and Meng, X. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science*, **13**(2), 163–185.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis*. Chapman & Hall, Boca Raton, 2nd edition.

Hawkes, A. G. (1971a). Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society Series B*, **33**, 438–443.

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

Hawkes, A. G. (1972). Spectra of some mutually exciting point processes with associated variables. In P. A. W. Lewis, editor, *Stochastic Point Processes*, pages 261–271. Wiley, New York,.

Hawkes, A. G. and Oakes, D. (1974). A cluster representation of a self-exciting process. *Journal of Applied Probability*, **11**, 493–503.

Kendall, W. S. and Møller, J. (2000). Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Advances in Applied Probability*, **32**, 844–865.

Kingman, J. F. C. (1993). *Poisson processes*. Oxford University Press, Oxford.

Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer Verlag, London.

Møller, J. (2003). Shot noise Cox processes. *Advances in Applied Probability*, **35**, 614–640.

Møller, J. and Torrisi, G. L. (2005). Second order analysis for spatial hawkes processes. Research Report R-2005-20, Department of Mathematical Sciences, Aalborg University.

Møller, J. and Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall, Boca Raton, Florida.

Møller, J., Syversveen, A. R., and Waagepetersen, R. P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, **25**, 451–482.

Møller, J. M., Pettitt, A. N., Berthelsen, K. K., and Reeves, R. W. (2006). An efficient markov chain monte carlo method for distributions with intractable normalising constants. *Biometrika*, **93**. To appear.

Norris, J. R. (1997). *Markov Chains*. Cambridge University Press, New York.

Ogata, Y. (1981). On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, **IT-27**(1), 23–31.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, **83**(401), 9–27.

Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, **50**(2), 379–402.

Papangelou, F. (1974). The conditional intensity of general point processes and an application to line processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **28**, 207–226.

Propp, J. G. and Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, **9**, 223–252.

Thomas, M. (1949). A generalization of Poisson's binomial limit for use in ecology. *Biometrika*, **36**, 18–25.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, **22**(4), 1701–1762.

Torrisi, G. L. (2002). A class of interacting marked point processes: rate of convergence to equilibrium. *Journal of Applied Probability*, **39**, 137–160.

van Lieshout, M. (2000). *Markov Point Processes and Their Applications*. Imperial College Press, London.

Vere-Jones, D. and Ozaki, T. (1982). Some examples of statistical inference applied to earthquake data. *Annals of the Institute of Statistical Mathematics*, **34**, 189–207.

Wilson, D. B. (2000). How to couple from the past using a read-once source of randomness. *Random Structures and Algorithms*, **16**, 85–113.

Chapter 2

Perfect simulation of Hawkes processes

Jesper Møller & Jakob G. Rasmussen

Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg, Denmark.

Email addresses: jm@math.aau.dk & jgr@math.aau.dk

Abstract

Our objective is to construct a perfect simulation algorithm for unmarked and marked Hawkes processes. The usual straightforward simulation algorithm suffers from edge effects, whereas our perfect simulation algorithm does not. By viewing Hawkes processes as Poisson cluster processes and using their branching and conditional independence structures, useful approximations of the distribution function for the length of a cluster are derived. This is used to construct upper and lower processes for the perfect simulation algorithm. A tail-lightness condition turns out to be of importance for the applicability of the perfect simulation algorithm. Examples of applications and empirical results are presented.

Keywords: Approximate simulation; dominated coupling from the past; edge effects; exact simulation; marked Hawkes process; marked point process; perfect simulation; point process; Poisson cluster process; thinning; upper process; lower process

2000 Mathematical Subject Classification: Primary 60G55
Secondary 68U20

2.1 Introduction

Unmarked and marked Hawkes processes (Hawkes, 1971a,b, 1972; Hawkes and Oakes, 1974) play a fundamental role for point process theory and its applications, cf., for example, Daley and Vere-Jones (2003), and they have important applications in seismology (Hawkes and Adamopoulos, 1973; Ogata, 1988, 1998; Vere-Jones and Ozaki, 1982) and neurophysiology (Brémaud and Massoulié, 1996; Chornoboy *et al.*, 2002). There are many ways to define a marked Hawkes process, but for our purpose it is most convenient to define it as a marked Poisson cluster process $X = \{(t_i, Z_i)\}$ with events (or times) $t_i \in \mathbb{R}$ and marks Z_i defined on an arbitrary (mark) space M equipped with a probability distribution Q. The cluster centres of X are given by certain events called *immigrants*, while the other events are called *offspring*.

Definition 2.1 (*Hawkes process with unpredictable marks.*)

(a) The immigrants follow a Poisson process with a locally integrable intensity function $\mu(t)$, $t \in \mathbb{R}$.

(b) The marks associated to the immigrants are independent and identically distributed (i.i.d.) with distribution Q, and are independent of the immigrants.

(c) Each immigrant t_i generates a *cluster* C_i, which consists of marked events of generations of order $n = 0, 1, \ldots$ with the following *branching structure* (see Figure 2.1). We first have

(t_i, Z_i), which is said to be of generation 0. Given the $0, \ldots, n$ generations in C_i, each $(t_j, Z_j) \in C_i$ of generation n recursively generates a Poisson process Φ_j of offspring of generation $n+1$ with intensity function $\gamma_j(t) = \gamma(t - t_j, Z_j)$, $t > t_j$. Here γ is a non-negative measurable function defined on $(0, \infty)$. We refer to Φ_j as an *offspring process*, and to γ_j and γ as *fertility rates*. Furthermore, the mark Z_k associated to any offspring $t_k \in \Phi_j$ has distribution Q and Z_k is independent of t_k and all (t_l, Z_l) with $t_l < t_k$. As in Daley and Vere-Jones (2003), we refer to this as the case of *unpredictable marks*.

(d) Given the immigrants, the clusters are independent.

(e) Finally, X consists of the union of all clusters.

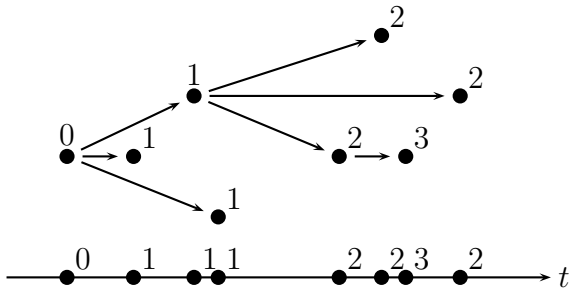


Figure 2.1: The branching structure of the various generations of events in a cluster (ignoring the marks)(*top*), and the events on the time axis (*bottom*).

The independence assumptions in (c) and (d) imply that we have i.i.d. marks. In the special case where $\gamma(t, z) = \gamma(t)$ does not depend on its second argument (or if just $\mathrm{P}(\gamma(t, Z) = \gamma(t)$ for Lebesgue almost all $t > 0$ $) = 1$ where Z denotes a generic mark), the events follow an unmarked Hawkes process. Apart from in that case, the events and the marks are dependent processes. Another way of defining the process is as follows (see e.g. Daley and Vere-Jones (2003)):

The marks are i.i.d. and the conditional intensity function $\lambda(t)$ at time $t \in \mathbb{R}$ for the events given the previous history $\{(t_k, Z_k) : t_k < t\}$ is given by

$$\lambda(t) = \mu(t) + \sum_{t_i<t} \gamma(t - t_i, Z_i). \tag{2.1}$$

Simulation procedures for Hawkes processes are needed for various reasons: Analytical results are rather limited due to the complex stochastic structure; statistical inference, especially model checking and prediction, require simulations; and displaying simulated realizations of specific model constructions provides a better understanding of the model. The general approach for simulating a (marked or unmarked) point process is to use a thinning algorithm such as the Shedler-Lewis thinning algorithm or Ogata's modified thinning algorithm (see e.g. Daley and Vere-Jones (2003)). However, Definition 2.1 immediately leads to the following approximate simulation algorithm, where $t_- \in [-\infty, 0]$ and $t_+ \in (0, \infty]$ are user-specified parameters, and the output consists of all marked points (t_i, Z_i) with $t_i \in [0, t_+)$.

Algorithm 2.1 The following steps (i)-(ii) generate an approximate simulation of those marked events $(t_i, Z_i) \in X$ with $0 \le t_i < t_+$.

(i) Simulate the immigrants on $[t_-, t_+)$.

(ii) For each such immigrant t_i, simulate Z_i and those $(t_j, Z_j) \in C_i$ with $t_i < t_j < t_+$.

In general Algorithm 2.1 suffers from edge effects, since clusters generated by immigrants before time t_- may contain offspring in $[0, t_+)$. Brémaud *et al.* (2002) studied the 'rate of installation', i.e. they considered a coupling of X, after time 0, with the output from Algorithm 2.1 when $t_+ = \infty$. Under a tail-lightness assumption (see the paragraph after Proposition 2.3, below) and other conditions, they established an exponentially decreasing bound for the probability $\mathrm{P}(t_-, \infty)$, say, that X, after time 0, coincides with the output of the algorithm. Algorithm 2.1 is also investigated in our own work

(Møller and Rasmussen, 2004) where various measures for edge effects, including refined results for $\mathrm{P}(t_-, \infty)$, were introduced.

Our objective in this paper is to construct a perfect (or exact) simulation algorithm. Perfect simulation has been a hot research topic since the seminal Propp-Wilson algorithm (Propp and Wilson, 1996) appeared, but the areas of application have so far been rather limited and many perfect simulation algorithms proposed in the literature are too slow for real applications. As demonstrated in Møller and Rasmussen (2004) our perfect simulation algorithm can be practical and efficient. Moreover, apart from the advantage of not suffering from edge effects, our perfect simulation algorithm can also be useful in quantifying the edge effects suffered by Algorithm 2.1 (see Møller and Rasmussen (2004)).

The perfect simulation algorithm is derived using similar principles as in Brix and Kendall (2002), but our algorithm is a non-trivial extension, since the Brix-Kendall algorithm requires the knowledge of the cumulative distribution function (CDF) F for the length of a cluster, and F is unknown even for the simplest examples of Hawkes processes. By establishing certain monotonicity and convergence results, we are able to approximate F to any required precision, and, more importantly, to construct a dominating process and upper and lower processes in a similar fashion as in the dominated-coupling-from-the-past algorithm of Kendall and Møller (2000). Under a tail-lightness condition, our perfect simulation algorithm turns out to be feasible in applications, while in the heavy-tailed case, we can at least say something about the approximate form of F, cf. Example 2.7.

The paper is organised as follows. Section 2.2 contains some preliminaries, including illuminating examples of Hawkes process models used throughout the paper to illustrate our results. In Section 2.3, we describe the perfect simulation algorithm, assuming that F is known, while the above-mentioned convergence and monotonicity results are established in Section 2.4. Section 2.5 completes the perfect simulation algorithm, using dominated coupling from the past. Finally, Section 2.6 contains a discussion of our algorithm and results, and

suggestions on how to extend these to more general settings.

2.2 Preliminaries and examples

2.2.1 The branching structure and self-similarity property of clusters

By Definition 2.1, we can view the marked Hawkes process $X = \{(t_i, Z_i)\}$ as a Poisson cluster process with cluster centres given by the immigrants, where the clusters, given the immigrants, are independent. In this section, we describe a self-similarity property resulting from the specific branching structure within a cluster.

For events $t_i < t_j$, we say that (t_j, Z_j) has *ancestor* t_i *of order* $n \geq 1$ if there is a sequence $s_1 \ldots, s_n$ of offspring such that $s_n = t_j$ and s_k, $k = 1, \ldots, n$, is one of the offspring of s_{k-1}, with $s_0 = t_i$. We then say that t_j is an *offspring of* n*th generation with respect to* t_i; for convenience, we say that t_i is of zeroth generation with respect to itself. Now we define the *total offspring process* C_i as all those (t_j, Z_j) such that t_j is an event of generation $n \in \mathbb{N}_0$ with respect to t_i (note that $(t_i, Z_i) \in C_i$). The *clusters* are defined as those C_i for which t_i is an immigrant (see Definition 2.1).

The total offspring processes have the same branching structure relative to their generating events. More precisely, since $\gamma_i(t) = \gamma(t - t_i, Z_i)$ for any event t_i, we see by Definition 2.1 that conditional on events $t_i < t_j$, the translated total offspring processes $C_i - t_i \equiv \{(t_l - t_i, Z_l) : (t_l, Z_l) \in C_i\}$ and $C_j - t_j \equiv \{(t_l - t_j, Z_l) : (t_l, Z_l) \in C_j\}$ are identically distributed.

In particular, conditional on the immigrants, the clusters relative to their cluster centres (the immigrants) are i.i.d. with distribution P, say. Furthermore, conditional on a cluster's nth generation events $\mathcal{G}_n$, say, in a cluster, the translated total offspring processes $C_j - t_j$ with $t_j \in \mathcal{G}_n$ are i.i.d. with distribution P. We refer to this last property as the i.i.d. self-similarity property of offspring processes or,

for short, the *self-similarity property*. Note that the assumption of unpredictable marks is essential for these properties to hold.

2.2.2 A basic assumption and some terminology and notation

Let F denote the CDF for the length L of a cluster, i.e. the time between the immigrant and the last event of the cluster. Consider the mean number of events in any offspring process $\Phi(t_i)$, $\bar{\nu} \equiv \mathrm{E}\nu$, where

$$\nu = \int_0^\infty \gamma(t, Z)\,\mathrm{d}t$$

is the *total fertility rate of an offspring process* and Z denotes a *generic mark* with distribution Q. Henceforth, we assume that

$$0 < \bar{\nu} < 1. \tag{2.2}$$

The condition $\bar{\nu} < 1$ appears commonly in the literature on Hawkes processes (see e.g. Brémaud *et al.* (2002), Daley and Vere-Jones (2003), and Hawkes and Oakes (1974)), and is essential to our convergence results in Section 2.4.2. It implies that

$$F(0) = \mathrm{E}\mathrm{e}^{-\nu} > 0 \tag{2.3}$$

where $F(0)$ is the probability that a cluster has no offspring. It is equivalent to assuming that $\mathrm{E}S < \infty$, where S denotes the number of events in a cluster: by induction on $n = 0, 1 \ldots$, because of the branching and conditional independence structure of a cluster, $\bar{\nu}^n$ is the mean number of generation n events in a cluster, meaning that

$$\mathrm{E}S = 1 + \bar{\nu} + \bar{\nu}^2 + \cdots = 1/(1 - \bar{\nu}) \tag{2.4}$$

if $\bar{\nu} < 1$, while $\mathrm{E}S = \infty$ otherwise.

The other condition, $\bar{\nu} > 0$, excludes the trivial case where there are almost surely no offspring. It is readily seen to be equivalent to

$$F < 1. \tag{2.5}$$

Furthermore,

$$h(t) = \mathrm{E}[\gamma(t, Z)/\nu], \qquad t > 0, \tag{2.6}$$

and

$$\bar{h}(t) = \mathrm{E}\gamma(t, Z)/\bar{\nu}, \qquad t > 0, \tag{2.7}$$

are well-defined densities (with respect to the Lebesgue measure). The density $\bar{h}$ will play a keyrole later in this paper; it can be interpreted as the *normalised intensity function for the first generation of offspring in a cluster started at time* 0. Note that h specifies the *density of the distance R from an arbitrary offspring to its nearest ancestor*. In the sequel, since the clusters, relative to their cluster centers, are i.i.d. (see Section 2.2.1), we assume without loss of generality that L, R and S are defined with respect to the same immigrant $t_0 = 0$, with mark $Z_0 = Z$.

Clearly, if $L > 0$ then $R > t$ implies $L > t$, meaning the distribution of L has a thicker tail than that of R. The probability function for S is given by

$$P(S = k) = P(S_{n+1} = k - 1 | S_n = k)/k, \qquad k \in \mathbb{N},$$

where S_n denotes the number of events of nth generation and $n \in \mathbb{N}$ is arbitrary (see Dwass (1969) or Theorem 2.11.2 in Jagers (1975)). Thus,

$$\mathrm{P}(S = k) = \mathrm{E}\left[\mathrm{e}^{-k\nu}(k\nu)^{k-1}/k!\right], \qquad k \in \mathbb{N}. \tag{2.8}$$

2.2.3 Examples

Throughout the paper, we illustrate the results with the following cases.

Example 2.1 (*Unmarked process*) An unmarked Hawkes process with exponentially decaying fertility rate is given by

$$\bar{\nu} = \nu = \alpha, \qquad \bar{h}(t) = h(t) = \beta \mathrm{e}^{-\beta t},$$

where $0 < \alpha < 1$ and $\beta > 0$ are parameters. Here $1/\beta$ is a scale parameter for both the distribution of R and the distribution of L.

The left-hand panel of Figure 2.2 shows perfect simulations of this process on $[0, 10]$ when $\mu(t) = 1$ is constant, $\alpha = 0.9$, and $\beta = 10, 5, 2, 1$. By (2.4), we expect to see about 10 clusters (in total) and 100 events. The clusters of course become more visible as β increases.

The left-hand panel of Figure 2.3 shows six simulations of clusters with $\alpha = 0.9$. Here, α is an inverse scaling parameter; β is irrelevant since, to obtain comparable results for this example and the following two examples, we have omitted showing the scale. All the clusters have been simulated conditional on $S > 1$ to avoid the frequent and rather uninteresting case containing only the immigrant. These few simulations indicate the general tendency of L vary widely. □

Example 2.2 (*Birth-death process*) Consider a marked Hawkes process with

$$\gamma(t, Z) = \alpha \mathbf{1}(t \le Z)/\mathrm{E}Z,$$

where α, $0 < \alpha < 1$, is a parameter, Z is a positive random variable, and $\mathbf{1}(\cdot)$ denotes the indicator function. Then X can be viewed as a birth-death process with birth at time t_i and survival time Z_i of the i'th individual. The birth rate is

$$\lambda(t) = \mu(t) + (\alpha/\mathrm{E}Z)\mathrm{card}\big(\{i : t_i < t \le t_i + Z_i\}\big), \qquad t \in \mathbb{R},$$

(cf. (2.1)). Moreover,

$$\nu = \alpha Z/\mathrm{E}Z, \quad \bar{\nu} = \alpha, \quad h(t) = \mathrm{E}(\mathbf{1}(t \le Z)/Z), \quad \bar{h}(t) = \mathrm{P}(Z \ge t)/\mathrm{E}Z.$$

Since ν is random, the distribution of S is more dispersed than in the unmarked case (cf. (2.8)).

The special case where $\mu(t) = \mu$ is constant and Z is exponentially distributed with mean $1/\beta$ is considered at page 136 in Brémaud *et al.* (2002). In this case, X is a time-homogeneous Markov birth-death process with birth rate $\mu + \alpha\beta n$ and death rate βn, where

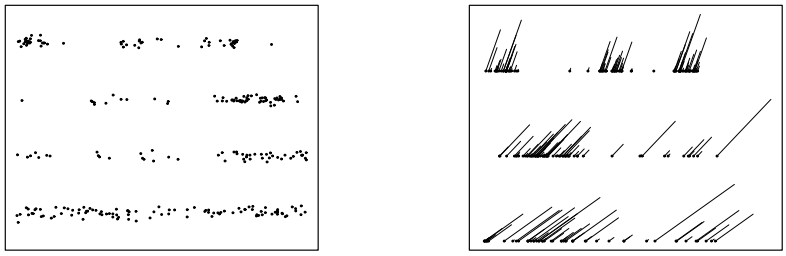


Figure 2.2: On the left, we display four perfect simulations on $[0, 10]$ of the unmarked Hawkes process (Example 2.1) with parameters $\alpha = 0.9$, $\mu = 1$, and $\beta = 10, 5, 2, 1$ (top to bottom). Random jitter has been added in the vertical direction to help distinguishing events located close together. On the right, we display three perfect simulations on $[0, 10]$ of the birth-death Hawkes process (Example 2.2) with parameters $\alpha = 0.9$, $\mu = 1$, and $\beta = 5, 2, 1$ (top to bottom), where the projections of the lines onto the horizontal axis show the size of the marks.

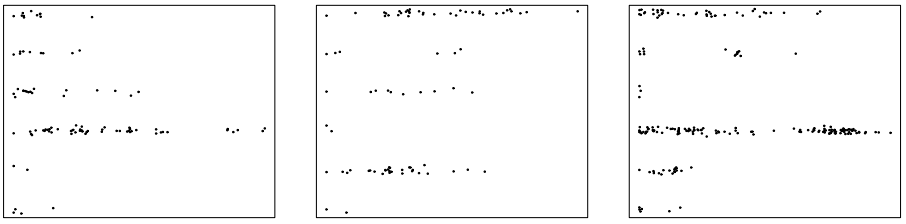


Figure 2.3: On the left, we display six simulations of clusters started at 0 and conditioned on $S > 1$ in the unmarked case with $\alpha = 0.9$. In the centre, we display the same simulations, in the birth-death case, and, on the right, in the heavy-tailed case. Different scalings are used in the three cases.

n is the number of living individuals. Furthermore, $\bar{h}(t) = \beta \mathrm{e}^{-\beta t}$ and $h(t) = \beta E_1(\beta t)$, where $E_1(s) = \int_s^\infty \mathrm{e}^{-t}/t \, \mathrm{d}t$ is the exponential integral function. As in Example 2.1, $1/\beta$ is a scale parameter for the distribution of L. As discussed in Example 2.8, below, the stationary distribution (i.e. the distribution of X at any fixed time) is known up to a constant of proportionality, and it is possible to simulate from this by rejection sampling.

The right-hand panel of Figure 2.2 shows three perfect simulations in the Markov case on $[0, 10]$ with $\mu = 1$, $\alpha = 0.9$, and $\beta = 5, 2, 1$, where the marks are indicated by line segments of different lengths. The centre panel of Figure 2.3 shows six simulations of clusters (with marks excluded) with $\alpha = 0.9$, simulated conditional on $S > 1$. These simulations indicate that L is slightly more dispersed than in Example 2.1, since the marks introduce additional variation in the cluster lengths. In fact, the coefficient of variation estimated from 10000 perfect simulations is 1.92 for Example 2.1 and 2.85 for the present case. □

Example 2.3 (*Heavy-tailed distribution for L*) Suppose that $\gamma(t, Z) = \alpha Z \mathrm{e}^{-tZ}$, where $\alpha \in (0, 1)$ is a parameter and Z is exponentially distributed with mean $1/\beta$. Then $\bar{\nu} = \nu = \alpha$ is constant, meaning that the distribution of S is the same as in the unmarked case (cf. (2.8)). Furthermore,

$$h(t) = \bar{h}(t) = \beta/(t + \beta)^2$$

specifies a Pareto density. This is a heavy-tailed distribution, as it has infinite Laplace transform ($\mathcal{L}(\theta) = \mathrm{E}\mathrm{e}^{\theta R} = \infty$ for all $\theta > 0$). Moreover, it has infinite moments ($\mathrm{E}\big(R^p\big) = \infty$ for all $p \geq 1$). Consequently, L also has a heavy-tailed distribution with infinite moments and infinite Laplace transform. Note that β is a scale parameter for the distribution of L.

The right-hand panel of Figure 2.3 shows six simulations of clusters with $\alpha = 0.9$ and $\beta = 1$. These indicate that L is much more dispersed than in Examples 2.1 and 2.2 (in fact, the dispersion is infinite in the present case). □

2.3 Perfect Simulation

Assuming for the moment that F (the CDF for the length of a cluster) is known, the following algorithm for perfect simulation of the marked Hawkes process is similar to the algorithm for simulation of Poisson cluster processes without edge effects given in Brix and Kendall (2002) (see also Møller (2003) and Møller and Waagepetersen (2004)).

Algorithm 2.2 Let I_1 be the point process of immigrants on $[0, t_+)$, and let I_2 be the point process of immigrants $t_i < 0$ such that $\{(t_j, Z_j) \in C_i : t_j \in [0, \infty)\} \neq \emptyset$.

1. Simulate I_1 as a Poisson process with intensity function $\lambda_1(t) = \mu(t)$ on $[0, t_+)$.

2. For each $t_i \in I_1$, simulate Z_i and those $(t_j, Z_j) \in C_i$ with $t_i < t_j < t_+$.

3. Simulate I_2 as a Poisson process with intensity function $\lambda_2(t) = (1 - F(-t))\mu(t)$ on $(-\infty, 0)$.

4. For each $t_i \in I_2$, simulate Z_i and $\{(t_j, Z_j) \in C_i : t_j \in [0, t_+)\}$ conditional on the event that $\{(t_j, Z_j) \in C_i : t_j \in [0, \infty)\} \neq \emptyset$.

5. The output is all marked points from (1), (2), and (4).

Remark 2.1 In steps (1) and (2) of Algorithm 2.2, we use Algorithm 2.1 (with $t_- = 0$). In step (4), it is not obvious how to construct an elegant approach ensuring that at least one point will fall after 0. Instead, we use a simple rejection sampler: we repeatedly simulate Z_i from Q and the successive generations of offspring t_j to t_i (together with their marks Z_j) until there is at least one event of C_i after time 0.

The key point is how to simulate I_2 in step (3), since this requires the knowledge of F, which is unknown in closed form (Remark 2.3, below). In Section 2.4, we address this problem and, in Section 2.5, we construct an algorithm for simulating I_2.

In practice we must require that I_2 is (almost surely) finite or, equivalently, that

$$\int_{-\infty}^{0} (1 - F(-t))\mu(t)\,\mathrm{d}t < \infty. \tag{2.9}$$

In the case that $\mu(t)$ is bounded, (2.9) is satisfied if $\sup_{t \geq 0} \mu(t)\,\mathrm{E}L < \infty$. A condition for finiteness of $\mathrm{E}L$ is established in Lemma 2.1 and Remark 2.2, below. □

Proposition 2.1 *The output of Algorithm 2.2 follows the distribution of the marked Hawkes process.*

Proof. The immigrant process minus $I_1 \cup I_2$ generates clusters with no events in $[0, t_+)$. Since I_1 consists of the immigrants on

$[0, t_+)$, it follows directly that I_1 is a Poisson process with intensity $\lambda_1(t) = \mu(t)$ on $[0, t_+)$. Since I_2 is those immigrants on $(-\infty, 0)$ with offspring after 0, I_2 can be viewed as an independent thinning of the immigrant process with retention probability $p(t) = 1 - F(-t)$, and, thus, I_2 is a Poisson process with intensity $\lambda_2(t) = (1 - F(-t))\mu(t)$. Since I_1 and I_2 are independent, it follows from Section 2.2.1 that $\{C_i : t_i \in I_1\}$ and $\{C_i : t_i \in I_2\}$ are independent. Viewing the marked Hawkes process as a Poisson cluster process, it follows from Remark 2.1 that the clusters are generated in the right way in steps (2) and (4) of Algorithm 2.2 when we only want to sample those marked points (t_j, Z_j) with $t_j \in [0, t_+)$. Thus, Algorithm 2.2 produces realizations from the distribution of the marked Hawkes process. □

Using the notation of Section 2.2.2, the following lemma generalises and sharpens a result of Hawkes and Oakes (1974) about the mean length of a cluster.

Lemma 2.1 *We have that*

$$\frac{1}{\mathrm{E}\mathrm{e}^{-\nu}} \mathrm{E}\left[(1 - \mathrm{e}^{-\nu})\mathrm{E}[R|Z]\right] \leq \mathrm{E}L \leq \frac{\bar{\nu}}{1 - \bar{\nu}} \mathrm{E}\bar{R}. \tag{2.10}$$

Proof. Consider a cluster starting with an immigrant at time $t_0 = 0$, with mark $Z_0 = Z$ (cf. Section 2.2.1). For $t_j \in \mathcal{G}_1$, let R_j denote the distance from t_j to 0, and L_j the length of the total offspring process C_j started by t_j. Then $L = \max\{R_j + L_j : t_j \in \mathcal{G}_1\}$, so, if we condition on Z, and let $R_{j,z}$ be distributed as R_j, conditional on the event $Z = z$, then

$$\mathrm{E}L = \mathrm{E}\mathrm{E}[L|Z] = \mathrm{E}\left[\sum_{i=1}^{\infty} \frac{\mathrm{e}^{-\nu}\nu^i}{i!} \mathrm{E}\left[\max\{R_{j,Z} + L_j : j = 1, \ldots, i\}\right]\right]. \tag{2.11}$$

To obtain the upper inequality, observe that

$$\mathrm{E}L \leq \mathrm{E}\left[\sum_{i=1}^{\infty} \frac{\mathrm{e}^{-\nu}\nu^i}{i!} \mathrm{E}\left[\sum_{j=1}^{i} (R_{j,Z} + L_j)\right]\right] = \mathrm{E}[\nu \mathrm{E}[R|Z]] + \bar{\nu}\mathrm{E}L,$$

where we have used that the L_j are identically distributed and has the same distribution as L because of the self-similarity property (Section 2.2.1), and the fact that the R_j are identically distributed when conditioned on Z. Hence,

$$\mathrm{E}L \leq \frac{1}{1-\bar{\nu}}\mathrm{E}[\nu\mathrm{E}[R|Z]] = \frac{1}{1-\bar{\nu}}\mathrm{E}\left[\int_0^\infty s\gamma(s,Z)\mathrm{d}s\right] = \frac{\bar{\nu}}{1-\bar{\nu}}\mathrm{E}\bar{R},$$

which verifies the upper inequality. Finally, by (2.11),

$$\begin{aligned}\mathrm{E}L &\geq \mathrm{E}\left[\sum_{i=1}^{\infty}\frac{\mathrm{e}^{-\nu}\nu^i}{i!}(\mathrm{E}[R|Z]+\mathrm{E}L)\right]\\ &= \mathrm{E}\big[(1-\mathrm{e}^{-\nu})\mathrm{E}[R|Z]\big] + \mathrm{E}[1-\mathrm{e}^{-\nu}]\mathrm{E}L,\end{aligned}$$

which reduces to the lower inequality. □

Remark 2.2 If either ν or γ/ν is independent of Z (in other words, either the number or the locations of offspring in an offspring process are independent of the mark associated to the generic event), then it is easily proven that $\bar{h} = h$ and, thus, (2.10) reduces to

$$\left(\frac{1}{\mathrm{E}\mathrm{e}^{-\nu}} - 1\right)\mathrm{E}R \leq \mathrm{E}L \leq \frac{\bar{\nu}}{1-\bar{\nu}}\mathrm{E}R.$$

Consequently, $\mathrm{E}L < \infty$ if and only if $\mathrm{E}R < \infty$. This immediately shows that $\mathrm{E}L < \infty$ in Example 2.1 and $\mathrm{E}L = \infty$ in Example 2.3. In Example 2.2, when Z is exponentially distributed with mean $1/\beta$, (2.10) becomes

$$\frac{\alpha(\alpha+2)}{2(\alpha+1)\beta} \leq \mathrm{E}L \leq \frac{\alpha}{\beta(1-\alpha)},$$

so in this case $\mathrm{E}L < \infty$. Not surprisingly, apart from for small values of $\alpha \in (0,1)$, the bounds are rather poor and of little use except in establishing finiteness of $\mathrm{E}L$. □

2.4 The distribution of the length of a cluster

In this section we derive various distributional results concerning the length L of a cluster. The results are needed in Section 2.5 to complete step (3) in Algorithm 2.2; however, many of the results are also of independent interest.

2.4.1 An integral equation for F

Below, in Proposition 2.2, an integral equation for F is derived, and we discuss how to approximate F by numerical methods, using a certain recursion. Proposition 2.2 is a generalisation of Theorem 5 of Hawkes and Oakes (1974), which is proved using void probabilities obtained from a general result for the probability-generating functional for an unmarked Hawkes process. However, as was pointed out in Daley and Vere-Jones (2003), the probability generating functional for the marked Hawkes process is difficult to obtain. We give a direct proof based on void probabilities.

For $n \in \mathbb{N}_0$, let 1_n denote the CDF for the length of a cluster when all events of generation $n+1, n+2, \ldots$ are removed (it becomes clear in Section 2.4.2 why we use the notation 1_n). Clearly, 1_n is decreasing in n, $1_n \to F$ pointwise as $n \to \infty$, and

$$1_0(t) = 1, \qquad t \geq 0. \tag{2.12}$$

Let $\mathcal{C}$ denote the class of Borel functions $f : [0, \infty) \mapsto [0, 1]$. For $f \in \mathcal{C}$, define $\varphi(f) \in \mathcal{C}$ by

$$\varphi(f)(t) = \mathrm{E}\left[\exp\left(-\nu + \int_0^t f(t-s)\gamma(s, Z)\,\mathrm{d}s\right)\right], \qquad t \geq 0. \tag{2.13}$$

Proposition 2.2 *We have*

$$1_n = \varphi(1_{n-1}), \qquad n \in \mathbb{N}, \tag{2.14}$$

and

$$F = \varphi(F). \tag{2.15}$$

Proof. As in the proof of Lemma 2.1, we can consider a cluster started at time $t_0 = 0$, with associated marks $Z_0 = Z$. For a fixed $t \geq 0$ and $n \in \mathbb{N}$, split $\Phi(0)$ into three point processes Φ_1, Φ_2, Φ_3: Φ_1 consists of those first generation offspring $t_i \in \Phi(0) \cap [0,t)$ that do not generate events of generation $n-1$ or lower with respect to t_i on $[t,\infty)$; $\Phi_2 = (\Phi(0) \cap [0,t)) \setminus \Phi_1$ consists of the remaining first generation offspring on $[0,t)$; and $\Phi_3 = \Phi(0) \cap [t,\infty)$ consists of the first generation offspring on $[t,\infty)$. Conditional on Z, Φ_1, Φ_2, and Φ_3 are independent Poisson processes with intensity functions $\lambda_1(s) = \gamma(s,Z)1_{n-1}(t-s)$ on $[0,t)$, $\lambda_2(s) = \gamma(s,Z)(1-1_{n-1}(t-s))$ on $[0,t)$, and $\lambda_3(s) = \gamma(s,Z)$ on $[t,\infty)$, respectively. This follows by an independent thinning argument since, conditional on $\mathcal{G}_n$ (the n-th generation of offspring in C_0), the processes $C_j - t_j$ with $t_j \in \mathcal{G}_n$ are i.i.d. and distributed as C_0 (this is the self-similarity property from Section 2.2.1). Consequently,

$$\begin{aligned} 1_n(t) &= \mathrm{E}[\mathrm{P}(\Phi_2 = \emptyset | Z)\mathrm{P}(\Phi_3 = \emptyset | Z)] \\ &= \mathrm{E}\exp\left(-\int_0^t \lambda_2(s,Z)\,\mathrm{d}s - \int_t^\infty \lambda_3(s,Z)\,\mathrm{d}s\right) \end{aligned}$$

which reduces to (2.14). Taking the limit as $n \to \infty$ on both sides of (2.14), we obtain (2.15) by monotone convergence, since $1_n(t) \leq 1_{n-1}(t)$ for all $t \geq 0$ and $n \in \mathbb{N}$. □

Remark 2.3 As illustrated in the following example, we have been unsuccessful in using (2.15) to obtain a closed form expression for F even for simple choices of γ. Fortunately, the recursion (2.14) provides a useful numerical approximation to F. As the integral in (2.13) with $f = 1_{n-1}$ quickly becomes difficult to evaluate analytically as n increases, we compute the integral numerically, using a quadrature rule. □

Example 2.4 (*Unmarked process*) Consider Example 2.1 with $\beta = 1$. Then (2.15) is equivalent to

$$\int_0^t F(s)\mathrm{e}^s \,\mathrm{d}s = \frac{\mathrm{e}^t}{\alpha} \ln(\mathrm{e}^\alpha F(t))$$

which is not analytically solvable. □

2.4.2 Monotonicity properties and convergence results

As established in Theorem 2.1 below, many approximations of F other than 1_n exist, and their rates of convergence may be geometric with respect to different norms. Notice that certain monotonicity properties are fulfilled by φ, where, for functions $f : [0,\infty) \mapsto [0,1]$, we recursively define $\varphi^{[0]}(f) = f$ and $\varphi^{[n]}(f) = \varphi(\varphi^{[n-1]}(f))$, $n = 1, 2, \ldots$ and set $f_n = \varphi^{[n]}(f), n = 0, 1, \ldots$. Note that $F_n = F$ for all $n \in \mathbb{N}_0$. As $1_n = \varphi^{[n]}(1)$ is decreasing towards the CDF F, cases in which G is a CDF and G_n increases to F are of particular interest.

Lemma 2.2 *For any $f, g \in \mathcal{C}$, we have*

$$f \le g \quad \Rightarrow \quad f_n \le g_n, \qquad n \in \mathbb{N}, \tag{2.16}$$

$$f \le \varphi(f) \quad \Rightarrow \quad f_n \text{ is nondecreasing in } n, \tag{2.17}$$

$$f \ge \varphi(f) \quad \Rightarrow \quad f_n \text{ is nonincreasing in } n. \tag{2.18}$$

Proof. Equation (2.16) follows immediately from (2.13) when $n = 1$, and then by induction in the remaining cases. Equations (2.17) and (2.18) follow from (2.16). □

Theorem 2.1 *With respect to the supremum norm $\|f\|_\infty = \sup_{t\ge 0} |f(t)|$, φ is a contraction on $\mathcal{C}$, that is, for all $f, g \in \mathcal{C}$ and $n \in \mathbb{N}$, we have that $f_n, g_n \in \mathcal{C}$ and*

$$\|\varphi(f) - \varphi(g)\|_\infty \le \bar{\nu}\|f - g\|_\infty. \tag{2.19}$$

Furthermore, F is the unique fixpoint, i.e.

$$\|F - f_n\|_\infty \to 0 \qquad \text{as } n \to \infty, \tag{2.20}$$

and

$$\|F - f_n\|_\infty \le \frac{\bar\nu^n}{1-\bar\nu}\|\varphi(f) - f\|_\infty, \tag{2.21}$$

where $\|\varphi(f) - f\|_\infty \le 1$. Furthermore, if $f \le \varphi(f)$ or $f \ge \varphi(f)$, then f_n converges to F from below or, respectively, above.

Proof. Let $f, g \in \mathcal{C}$. Recall that, by the mean value theorem (e.g. Theorem 5.11 in Apostol (1974)), for any real numbers x and y, we have $\mathrm{e}^x - \mathrm{e}^y = (x-y)\mathrm{e}^{z(x,y)}$, where $z(x,y)$ is a real number between x and y. Thus by (2.13),

$$\|\varphi(f)-\varphi(g)\|_\infty = \sup_{t\ge 0}\left|\mathrm{E}\left[\mathrm{e}^{-\nu}\mathrm{e}^{c(t,f,g)}\int_0^t (f(t-s)-g(t-s))\gamma(s,Z)\mathrm{d}s\right]\right|$$

where $c(t,f,g)$ is random variable between $\int_0^t f(t-s)\gamma(s,Z)\,\mathrm{d}s$ and $\int_0^t g(t-s)\gamma(s,Z)\,\mathrm{d}s$. Since $f, g \le 1$, we obtain $\mathrm{e}^{c(t,f,g)} \le \mathrm{e}^\nu$, cf. (2.2). Consequently,

$$\begin{aligned}
\|\varphi(f) - \varphi(g)\|_\infty &\le \sup_{t\ge 0}\left|\mathrm{E}\left[\int_0^t (f(t-s)-g(t-s))\gamma(s,Z)\,\mathrm{d}s\right]\right| \\
&\le \mathrm{E}\left[\int_0^\infty \|f-g\|_\infty\,\gamma(s,Z)\,\mathrm{d}s\right] \\
&= \bar\nu\,\|f-g\|_\infty.
\end{aligned}$$

Thereby, (2.19) is verified. Since $\mathcal{C}$ is complete (see e.g. Theorem 3.11 in Rudin (1987)), it follows, from the fixpoint theorem for contractions (see e.g. Theorem 4.48 in Apostol (1974)), that the contraction has a unique fixpoint: by (2.15), this is F.

Since $f \in \mathcal{C}$ implies that $\varphi(f) \in \mathcal{C}$, we find that $f_n \in \mathcal{C}$ by induction. Hence, using (2.15), (2.19) and induction, we have

$$\|f_n - F\|_\infty = \|\varphi(f_{n-1}) - \varphi(F)\|_\infty \le \bar\nu\|f_{n-1} - F\|_\infty \le \bar\nu^n\|f - F\|_\infty, \tag{2.22}$$

for $n \in \mathbb{N}$. Since $\bar{\nu} < 1$, we recover (2.20).

Similarly to (2.22), we have

$$\|f_n - f_{n-1}\|_\infty \leq \bar{\nu}^{n-1} \|f_1 - f\|_\infty, \qquad n \in \mathbb{N}. \tag{2.23}$$

Furthermore, by (2.20), we have

$$\|F - f\|_\infty = \lim_{m \to \infty} \|f_m - f\|_\infty.$$

So, by the triangle inequality and (2.23), we have

$$\begin{aligned} \|F - f\|_\infty &\leq \lim_{m \to \infty} \big(\|f_1 - f\|_\infty + \|f_2 - f_1\|_\infty + \cdots + \|f_m - f_{m-1}\|_\infty \big) \\ &\leq \lim_{m \to \infty} \|f_1 - f\|_\infty \big(1 + \bar{\nu} + \cdots + \bar{\nu}^{m-1} \big) \\ &= \frac{\|f_1 - f\|_\infty}{1 - \bar{\nu}} \end{aligned}$$

(see (2.2)). Combining this with (2.22), we obtain (2.21). Finally, if $f \leq \varphi(f)$ or $f \geq \varphi(f)$ then by (2.17) or, respectively, (2.18)) and (2.20), f_n converges from below or, respectively, above. □

Similar results to those in Theorem 2.1, but for the L^1-norm, were established in Møller and Rasmussen (2004). The following remark and proposition show how to find upper and lower bounds on F in many cases.

Remark 2.4 Consider a function $f \in \mathcal{C}$. The conditions $f \leq \varphi(f)$ and $f \geq \varphi(f)$ are satisfied for the extreme cases $f = 0$ and $f = 1$, respectively. The upper bound $f = 1$ is useful in the following sections, but the lower bound $f = 0$ is too small a function for our purposes; if we require that $\mathrm{E}L < \infty$ (cf. Remark 2.1) then $f = 0$ cannot be used (in fact we use only $f = 0$ when producing the right-hand plot in Figure 2.4, below). To obtain a more useful lower bound, observe that $f \leq \varphi(f)$ implies $f \leq F < 1$ (cf. (2.5) and Theorem 2.1). If $f < 1$ then a sufficient condition for $f \leq \varphi(f)$ is

$$\frac{1}{\bar{\nu}} \geq \frac{\int_0^t (1 - f(t-s)) \bar{h}(s) \, \mathrm{d}s + \int_t^\infty \bar{h}(s) \, \mathrm{d}s}{1 - f(t)}, \qquad t \geq 0. \tag{2.24}$$

This follows readily from (2.7) and (2.13), using the fact that $e^x \geq 1 + x$.

The function f in (2.24) is closest to F when f is a CDF G and we have equality in (2.24). Equivalently, G satisfies the renewal equation

$$G(t) = 1 - \bar{\nu} + \bar{\nu} \int_0^t G(t-s)\bar{h}(s)\,\mathrm{d}s, \qquad t \geq 0,$$

which has the unique solution

$$G(t) = 1 - \bar{\nu} + \sum_{n=1}^{\infty} (1-\bar{\nu})\bar{\nu}^n \int_0^t \bar{h}^{*n}(s)\,\mathrm{d}s, \qquad t \geq 0, \tag{2.25}$$

where $*n$ denotes n-times convolution (see Theorem IV2.4 in Asmussen (1987)). In other words, G is the CDF of $\bar{R}_1 + \cdots + \bar{R}_K$ (setting $\bar{R}_1 + \cdots + \bar{R}_K = 0$ if $K = 0$), where $K, \bar{R}_1, \bar{R}_2, \ldots$ are independent random variables, each $\bar{R}_i$ has density $\bar{h}$, and K has a geometric density $(1-\bar{\nu})\bar{\nu}^n$. Interestingly, this geometric density is equal to $\mathrm{E}S_n/\mathrm{E}S$ (see (2.4)).

The next proposition shows that, in many situations, $G \leq \varphi(G)$ when G is an exponential CDF with a sufficiently large mean. In such cases, F has no heavier tails than such an exponential distribution. □

Denote by

$$\mathcal{L}(\theta) = \int_0^{\infty} \mathrm{e}^{\theta t}\bar{h}(t)\,\mathrm{d}t, \qquad \theta \in \mathbb{R},$$

the Laplace transform of $\bar{h}$.

Proposition 2.3 *If $G(t) = 1 - \mathrm{e}^{-\theta t}$ for $t \geq 0$, where $\theta > 0$ and $\mathcal{L}(\theta) \leq 1/\bar{\nu}$, then $G \leq \varphi(G)$.*

Proof. Upon inserting $f = G$ into the right-hand side of (2.24), we obtain

$$\int_0^t \mathrm{e}^{\theta s}\bar{h}(s)\,\mathrm{d}s + \mathrm{e}^{\theta t} \int_t^{\infty} \bar{h}(s)\,\mathrm{d}s.$$

Since this is an increasing function of $t > 0$, (2.24) is satisfied if and only if $\mathcal{L}(\theta) \leq 1/\bar{\nu}$. □

Note that Proposition 2.3 always applies for sufficiently small $\theta > 0$, except in the case where $\bar{h}$ is heavy-tailed in the sense that $\mathcal{L}(\theta) = \infty$ for all $\theta > 0$. The condition $\mathcal{L}(\theta) \leq 1/\bar{\nu}$ is equivalent to the tail-lightness condition (2.1) in Brémaud *et al.* (2002).

2.4.3 Examples

In Examples 2.5 and 2.6 below, we let

$$G(t) = 1 - \mathrm{e}^{-\theta t}, \qquad t \geq 0, \tag{2.26}$$

be the exponential CDF with parameter $\theta > 0$.

Example 2.5 (*Unmarked process*) For the case in Example 2.1, $\mathcal{L}(\theta) = \beta/(\beta - \theta)$ if $\theta < \beta$, and $\mathcal{L}(\theta) = \infty$ otherwise. Interestingly, for 'the best choice' $\theta = \mathcal{L}^{-1}(1/\bar{\nu}) = \beta(1-\alpha)$, (2.26) becomes the CDF for R times ES, which is easily seen to be the same as the CDF in (2.25).

The left-hand panel of Figure 2.4 shows 1_n and G_n when $\theta = \beta(1-\alpha)$ and $(\alpha, \beta) = (0.9, 1)$. The convergence of 1_n and G_n (with respect to $\|\cdot\|_\infty$) and the approximate form of F are clearly visible. Since G is a CDF and $G_{n+1} \geq G_n$, we find that G_n is also a CDF. The centre panel of Figure 2.4 shows the density $F'(t)/(1 - F(0))$ $(t > 0)$ approximated by

$$\frac{1}{2}\left(\frac{1_n'(t)}{1 - 1_n(0)} + \frac{G_n'(t)}{1 - G_n(0)}\right)$$

when $n = 50$ (in which case $1_n'(t)/(1 - 1_n(0))$ and $G_n'(t)/(1 - G_n(0))$ are effectively equal). As shown in the plot, the density is close to the exponential density with the same mean, but the tail is slightly thicker. □

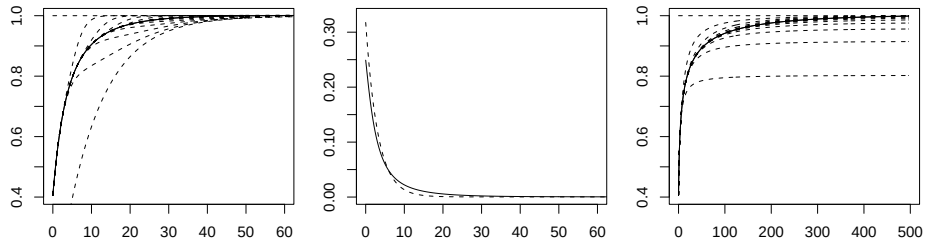


Figure 2.4: On the left, we display plots of 1_n and G_n for $n = 0, 5, \ldots, 50$ in the unmarked case with $\alpha = 0.9$ and $\beta = 1$ (see Example 2.5); 1_{50} and G_{50} are drawn solid to illustrate the approximate form of F, whereas the other curves are dashed. In the centre, we display plots of the density $\frac{1}{2}[1_n'/(1 - 1_n(0)) + G_n'/(1 - G_n(0))]$ when $n = 50$ (solid) and the exponential density with the same mean (dashed). On the right, we display the same plots as on the left, for Example 2.7 with $\alpha = 0.9$ and $\beta = 1$ using 1_n and 0_n as approximations of F.

Example 2.6 (*Birth-death process*) For the case in Example 2.2,

$$\mathcal{L}(\theta) = \mathrm{E}\int_0^Z \frac{\mathrm{e}^{\theta s}}{\mathrm{E}Z}\,\mathrm{d}s = \frac{\mathcal{L}_Z(\theta) - 1}{\theta \mathrm{E}Z}$$

where $\mathcal{L}_Z(\theta) = \mathrm{E}\mathrm{e}^{\theta Z}$ is the Laplace transform of Z. In the special case where Z is exponentially distributed with mean $1/\beta$, $\mathcal{L}(\theta) = \mathcal{L}_Z(\theta) = \beta/(\beta - \theta)$ is of the same form as in Example 2.5. Plots of 1_n, G_n, and

$$\frac{1}{2}\left(\frac{1_n'}{1 - 1_n(0)} + \frac{G_n'}{1 - G_n(0)}\right)$$

for $n = 0, 5, \ldots, 50$ and $(\alpha, \beta) = (0.9, 1)$ are similar to those in the left-hand and centre panels of Figure 2.4 and are therefore omitted. □

Example 2.7 (*Heavy-tailed distribution for L*) For the case in Example 2.3, Proposition 2.3 does not apply, as $\mathcal{L}(\theta) = \infty$ for all $\theta > 0$. The CDF in (2.25) is not known on closed form, since the convolutions are not tractable (in fact, this is the case when $\bar{h}$ specifies any known heavy-tailed distribution, including the Pareto, Weibull, log-normal or loggamma distribution). Nonetheless, it is still possible to get an idea of what F looks like: the right-hand panel of Figure 2.4 shows 1_n and 0_n for $n = 0, 5, \ldots, 50$ in the case $(\alpha, \beta) = (0.9, 1)$. As in Examples 2.5 and 2.6, the convergence of 1_n and G_n (where, now, $G = 0$) and the approximate form of F are clearly visible. However, as indicated by the plots and verified in Møller and Rasmussen (2004), $\lim_{t\to 0} G_n(t) < 1$ when $G = 0$, meaning that G_n is not a CDF. □

2.5 Simulation of I_2

To complete the perfect simulation algorithm (Algorithm 2.2), we need a useful way of simulating I_2. Our procedure is based on a

dominating process and the use of coupled upper and lower processes, in the spirit of the dominated-coupling-from-the-past algorithm (Kendall and Møller, 2000).

Suppose that $f \in \mathcal{C}$ is in closed form, with $f \leq \varphi(f)$, and that (2.9) is satisfied when we replace F by f (situations in which these requirements are fulfilled are considered in Sections 2.3, 2.4.2 and 2.4.3). Particularly, if μ is constant and f is a CDF, (2.9) implies that f has a finite mean. Now, for $n \in \mathbb{N}_0$, let U_n and L_n denote Poisson processes on $(-\infty, 0)$ with intensity functions

$$\lambda_n^u(t) = (1 - f_n(-t))\mu(t) \qquad \text{and} \qquad \lambda_n^l(t) = (1 - 1_n(-t))\mu(t),$$

respectively. By Theorem 2.1, λ_n^u is nonincreasing and λ_n^l is nondecreasing in n, and they both converge to λ_2 (geometrically fast with respect to the supremum norm). Consequently, we can use independent thinning to obtain the following sandwiching/funneling property (cf. Kendall and Møller (2000)):

$$\emptyset = L_0 \subseteq L_1 \subseteq L_2 \subseteq \cdots \subseteq I_2 \subseteq \cdots \subseteq U_2 \subseteq U_1 \subseteq U_0. \tag{2.27}$$

The details are given by the following algorithm.

Algorithm 2.3 (*Simulation of I_2.*)

1. Generate a realization $\{(t_1, Z_1), \ldots, (t_k, Z_k)\}$ of U_0, where $t_1 < \ldots < t_k$.

2. If $U_0 = \emptyset$, then return $I_2 = \emptyset$ and stop; otherwise generate independent uniform numbers $W_1, \ldots, W_k$ on $[0, 1]$ (independently of U_0), and set $n = 1$.

3. For $j = 1, \ldots, k$, assign (t_j, Z_j) to L_n or U_n if $W_j \lambda_0^u(t_j) \leq \lambda_n^l(t_j)$ or, respectively, $W_j \lambda_0^u(t_j) \leq \lambda_n^u(t_j)$.

4. If $U_n = L_n$, then return $I_2 = L_n$ and stop; otherwise increase n by 1 and repeat steps (3)–(4).

Proposition 2.4 *Algorithm 2.3 works correctly and terminates almost surely within finite time.*

Proof. To see this, imagine that, regardless of whether $U_0 = \emptyset$ in step (2) or $U_n = L_n$ in step (4), we continue to generate (U_1, L_1), (U_2, L_2), etc. Furthermore, add an extra step: for $j = 1, \ldots, k$, assign (t_j, Z_j) to I_2 if and only if

$$W_j \lambda_n^u(t_j) \leq \lambda_2(t_j).$$

Then clearly, because of the convergence properties of λ_n^u and λ_n^l (see the discussion above), (2.27) is satisfied and, conditional on $t_1, \ldots, t_k$,

$$\begin{aligned} & \mathrm{P}(L_n \neq U_n \text{ for all } n \in \mathbb{N}_0) \\ & \quad \leq \sum_{j=1}^{k} \lim_{n\to\infty} \mathrm{P}(W_j \lambda_0^u(t_j) \leq \lambda_n^u(t_j),\ W_j \lambda_0^u(t_j) > \lambda_n^l(t_j)) \\ & \quad = \sum_{j=1}^{k} \mathrm{P}(\lambda_2(t_j) < W_j \lambda_0^u(t_j) \leq \lambda_2(t_j)) = 0. \end{aligned}$$

Thus, almost surely Algorithm 2.3 terminates within finite time and the output equals I_2. □

Remark 2.5 We compute 1_n and f_n numerically, using a quadrature rule (see Remark 2.3). After step (1) in Algorithm 2.3, we let the last quadrature point be given by $-t_1$ (since we do not need to calculate $1_n(t)$ and $f_n(t)$ for $t > -t_1$). Since we have to calculate 1_n and f_n recursively for all $n = 0, 1, 2, \ldots$ until Algorithm 2.3 terminates, there is no advantage in using a doubling scheme for n, as in the Propp-Wilson algorithm (Propp and Wilson, 1996). □

Example 2.8 (*Birth-death process.*) We have checked our computer code for Algorithms 2.2 and 2.3 by comparing with results produced by another perfect simulation algorithm. Consider the case in Example 2.2 when $\mu(t) = \mu$ is constant and Z is exponentially distributed with mean $1/\beta$. If N denotes the number of events alive at time 0,

we have the following detailed balance condition for its equilibrium density π_n:

$$\pi_n(\mu + \alpha\beta n) = \pi_{n+1}\beta(n+1), \qquad n \in \mathbb{N}_0.$$

This density is well defined, since $\lim_{n\to\infty} \pi_{n+1}/\pi_n = \alpha < 1$. Now, choose $m \in \mathbb{N}_0$ and $\epsilon \geq 0$ such that $a = \alpha + \epsilon < 1$ and $\pi_{n+1}/\pi_n \leq a$ whenever $n \geq m$. If $\mu \leq \alpha\beta$, we can take $\epsilon = m = 0$; otherwise, we can use $m \geq (\mu - \alpha\beta)/(\beta\epsilon)$ for some $\epsilon > 0$. Define an unnormalized density $\pi'_n,\ n \in \mathbb{N}_0$, by $\pi'_n = \pi_n/\pi_0$ if $n \leq m$, and $\pi'_n = a^{n-m}\pi_m/\pi_0$ otherwise. We can easily sample from π'_n by inversion (see Ripley (1987)), since we can calculate

$$\sum_0^\infty \pi'_n = \sum_0^m \frac{\pi_n}{\pi_0} + \frac{a}{1-a}\,\frac{\pi_m}{\pi_0}.$$

Then, since $\pi'_n \geq \pi_n/\pi_0$, we can sample N from π_n by rejection sampling (see Ripley (1987)). Furthermore, conditional on $N = n$, we generate n independent marks $Z'_1, \ldots, Z'_n$ that are exponentially distributed with mean $1/\beta$ (here, we exploit the memoryless property of the exponential distribution). Finally, we simulate the marked Hawkes process with events in $(0, t_+]$, using the conditional intensity

$$\lambda'(t) = \mu + \alpha\beta \left(\sum_{i=1}^n \mathbf{1}[t < Z'_i] + \sum_{0<t_i<t} \mathbf{1}[t < t_i + Z_i] \right).$$

We have implemented this algorithm for comparison with our algorithm. Not surprisingly, it is a lot faster than our perfect simulation algorithm (roughly 1200 times as fast in the case $\alpha = 0.9$, $\beta = \mu = 1$, and $t_+ = 10$), since it exploits the fact that we know the stationary distribution in this special case. □

2.6 Extensions and open problems

Except for the heavy-tailed case, our perfect simulation algorithm is feasible in the examples we have considered. However, simulation

of the heavy-tailed cases is an unsolved problem. In these cases, we can only say something about the approximate form of F (see Example 2.7).

For applications such as in seismology (Ogata, 1998), extensions of our results and algorithms to the heavy-tailed cases are important. The epidemic type aftershock sequences (ETAS) model (Ogata (1988)) used for modeling times and magnitudes of earthquakes, is a heavy-tailed marked Hawkes process. Its spatio-temporal extension, which also includes the locations of the earthquakes (see Ogata (1998)), furthermore involves the problem of predictable marks (the location of an aftershock depends on the location of the earthquake that causes it). This problem is easily solved, though, since the times and magnitudes are independent of the locations and can be simulated without worrying about these. This, of course, still leaves the unsolved problem of the heavy tails.

Extensions to nonlinear Hawkes processes (Brémaud and Massoulié, 1996; Daley and Vere-Jones, 2003) would also be interesting. However, things again become complicated, since a nonlinear Hawkes process is not even a Poisson cluster process.

Simulations of Hawkes processes with predictable marks can, in some cases, be obtained by using a thinning algorithm, if it is possible to dominate the Hawkes process with predictable marks by a Hawkes process with unpredictable marks. We illustrate the procedure with a simple birth-death example.

Example 2.9 (*Birth-death process)* Consider two birth-death Hawkes processes as defined in Example 2.2. Let Ψ_1 have unpredictable marks, with $Z_i^1 \sim \text{Exp}(\beta)$, and let Ψ_2 have predictable marks, with $Z_i^2 \sim \text{Exp}(\beta + 1/Z_{An}^2)$, where Z_{An}^2 is the mark of the first order ancestor of t_i. Both models have $\gamma(t, Z) = \alpha\beta\mathbf{1}[t < Z]$, with the same α and β, and they also have the same $\mu(t)$. The model Ψ_2 has the intuitive appeal that long-living individuals have long-living offspring. Note that the intensity of Ψ_1 dominates the intensity of Ψ_2 if the marks are simulated such that $Z_i^1 > Z_i^2$.

To simulate Ψ_2, we first simulate Ψ_1 using Algorithm 2.3, with the modifications that we associate both marks Z_i^1 and Z_i^2 to the event t_i, and we keep all events from the algorithm whether they fall in or before $[0, t_+)$. Each marked event (t_j, Z_j^1) is then included in Ψ_2 with retention probability

$$\frac{\mu(t) + \alpha\beta \sum_{t_i<t_j, t_i\in\Psi_2} \mathbf{1}[t_j - t_i < Z_i^2]}{\mu(t) + \alpha\beta \sum_{t_i<t_j} \mathbf{1}[t_j - t_i < Z_i^1]}$$

and the final output is all marked events from Ψ_2 falling in $[0, t_+)$. It is easily proven that these retention probabilities result in the correct process Ψ_2. □

Another process that would be interesting to obtain by thinning is the Hawkes process without immigrants considered in Brémaud and Massoulié (2001); this process has $\mu(t) = 0$ for all t. However, for this to be non-trivial (i.e. not almost surely empty), it is necessary that $\bar{\nu} = 1$, which means that any dominating Hawkes process has $\bar{\nu} \geq 1$ and, thus, cannot be simulated by Algorithm 2.3.

Many of our results and algorithms can be modified if we slightly extend the definition in Section 2.1 of a marked Hawkes process, as follows. For any event t_i with associated mark Z_i, let n_i denote the number of (first generation) offspring generated by (t_i, Z_i), and suppose that n_i, conditional on Z_i, is not necessarily Poisson distributed, but n_i is still conditionally independent of t_i and the previous history. A particular simple case occurs when n_i is either 1 or 0, and

$$\bar{p} = \mathrm{E}[\mathrm{P}(n_i = 1|Z_i)]$$

is assumed to be strictly between 0 and 1 (here $\bar{p}$ plays a similar role as $\bar{\nu}$ introduced in Section 2.4). Then we redefine φ by

$$\varphi(f)(t) = 1 - \bar{p} + \bar{p} \int_0^t f(t-s)\bar{h}(s)\,\mathrm{d}s$$

where, now,

$$\bar{h}(s) = \mathrm{E}(p(Z)h(s, Z))/\bar{p}.$$

Since φ is now linear, the situation is much simpler. For example, F is given by G in (2.25) (with $\bar{\nu}$ replaced by $\bar{p}$).

Another extension of practical relevance is to consider a non-Poisson immigrant process, e.g. a Markov or Cox process. The results in Section 2.4 do not depend on the choice of immigrant process, and the straightforward simulation algorithm (Algorithm 2.1) applies provided it is feasible to simulate the immigrants on $[t_-, t_+)$. However, the perfect simulation algorithm relies much on the assumption that the immigrant process is Poisson.

Finally, we notice that it would be interesting to extend our ideas to spatial Hawkes processes (see Møller and Torrisi (2004a) and Møller and Torrisi (2004b)).

Acknowledgements

We are grateful to Søren Asmussen, Svend Berntsen, Horia Cornean, Martin Jacobsen, Arne Jensen, and Giovanni Luca Torrisi for helpful discussions. The research of Jesper Møller was supported by the Danish Natural Science Research Council and the Network in Mathematical Physics and Stochastics (MaPhySto), funded by grants from the Danish National Research Foundation.

References

Apostol, T. M. (1974). *Mathematical Analysis*. Addison-Wesley, Reading.

Asmussen, S. (1987). *Applied Probability and Queues*. Wiley, Chichester.

Brémaud, P. and Massoulié, L. (1996). Stability of nonlinear Hawkes processes. *Ann. Prob.*, **24**, 1563–1588.

Brémaud, P. and Massoulié, L. (2001). Hawkes branching point processes without ancestors. *J. Appl. Prob.*, **38**, 122–135.

Brémaud, P., Nappo, G., and Torrisi, G. (2002). Rate of convergence to equilibrium of marked Hawkes processes. *J. Appl. Prob.*, **39**, 123–136.

Brix, A. and Kendall, W. (2002). Simulation of cluster point processes without edge effects. *Adv. Appl. Prob.*, **34**, 267–280.

Chornoboy, E. S., Schramm, L. P., and Karr, A. F. (2002). Maximum likelihood identification of neural point process systems. *Adv. Appl. Prob.*, **34**, 267–280.

Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods*. Springer, New York, 2nd edition.

Dwass, M. (1969). The total progeny in a branching process and a related random walk. *J. Appl. Prob.*, **6**, 682–686.

Hawkes, A. G. (1971a). Point spectra of some mutually exciting point processes. *J. Roy. Statist. Soc. Ser. B*, **33**, 438–443.

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

Hawkes, A. G. (1972). Spectra of some mutually exciting point processes with associated variables. In P. A. W. Lewis, editor, *Stochastic Point Processes*, pages 261–271. Wiley, New York,.

Hawkes, A. G. and Adamopoulos, L. (1973). Cluster models for earthquakes – regional comparisons. *Bull. Int. Statist. Inst.*, **45**, 454–461.

Hawkes, A. G. and Oakes, D. (1974). A cluster representation of a self-exciting process. *J. Appl. Prob.*, **11**, 493–503.

Jagers, P. (1975). *Branching Processes with Biological Applications*. John Wiley & Sons, London.

Kendall, W. S. and Møller, J. (2000). Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Adv. Appl. Prob.*, **32**, 844–865.

Møller, J. (2003). Shot noise Cox processes. *Adv. Appl. Prob.*, **35**, 614–640.

Møller, J. and Rasmussen, J. G. (2004). Approximate simulation of Hawkes processes. Research report R-2004-28, Department of Mathematical Sciences, Aalborg University. Available at http://www.math.aau.dk/ jm.

Møller, J. and Torrisi, G. L. (2004a). Generalised shot noise Cox processes. Research report R-2004-07, Department of Mathematical Sciences, Aalborg University. Available at http://www.math.aau.dk/ jm.

Møller, J. and Torrisi, G. L. (2004b). Perfect and approximate simulation of spatial Hawkes processes. In preparation.

Møller, J. and Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall, Boca Raton, Florida.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Assoc.*, **83**(401), 9–27.

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

Propp, J. G. and Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, **9**, 223–252.

Ripley, B. D. (1987). *Stochastic Simulation*. Wiley, New York.

Rudin, W. (1987). *Real and Complex Analysis*. McGraw-Hill, New York.

Vere-Jones, D. and Ozaki, T. (1982). Some examples of statistical inference applied to earthquake data. *Ann. Inst. Statist. Math.*, **34**, 189–207.

Chapter 3

Approximate simulation of Hawkes processes

Jesper Møller & Jakob G. Rasmussen

Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg, Denmark.

Email addresses: jm@math.aau.dk & jgr@math.aau.dk

Abstract

Hawkes processes are important in point process theory and its applications, and simulation of such processes are often needed for various statistical purposes. This article concerns a simulation algorithm for unmarked and marked Hawkes processes, exploiting that the process can be constructed as a Poisson cluster process. The algorithm suffers from edge effects but is much faster than the perfect simulation algorithm introduced in our previous work Møller and Rasmussen (2004). We derive various useful measures for the error committed when using the algorithm, and we discuss various empirical results for the algorithm compared with perfect simulations. Extensions of

the algorithm and the results to more general types of marked point processes are also discussed.

Keywords: Edge effects; Hawkes process; marked point process; Poisson cluster process; simulation.

2000 Mathematical Subject Classification: Primary 60G55
Secondary 68U20

3.1 Introduction

This paper concerns a useful simulation algorithm for unmarked and marked Hawkes processes (Hawkes, 1971a,b, 1972; Hawkes and Oakes, 1974; Daley and Vere-Jones, 2003). Such processes are important in point process theory and its applications, cf., for example, p. 183 in Daley and Vere-Jones (2003). Particularly, marked Hawkes processes have applications in seismology (Hawkes and Adamopoulos, 1973; Vere-Jones and Ozaki, 1982; Ogata, 1988, 1998) and neurophysiology (Chornoboy *et al.*, 2002). The algorithm in this paper suffers from edge effects but is of more practical importance than the perfect simulation algorithm introduced in Møller and Rasmussen (2004).

There are many ways to define a marked Hawkes process, but for our purpose it is most convenient to define it as a marked Poisson cluster process $X = \{(t_i, Z_i)\}$ with events (or times) $t_i \in \mathbb{R}$ and marks Z_i defined on an arbitrary (mark) space M equipped with a probability distribution Q. The cluster centres of X correspond to certain events called *immigrants* and the rest of the events are called *offspring*.

Definition 3.1 (*Hawkes process with unpredictable marks*)

(a) The immigrants follow a Poisson process with a locally integrable intensity function $\mu(t)$, $t \in \mathbb{R}$.

(b) The marks associated to the immigrants are i.i.d. with distribution Q and independent of the immigrants.

(c) Each immigrant t_i generates a *cluster* C_i, which consists of marked events of generations of order $n = 0, 1, \ldots$ with the following *branching structure*: First we have (t_i, Z_i), which is said to be of generation zero. Recursively, given the $0, \ldots, n$ generations in C_i, each $(t_j, Z_j) \in C_i$ of generation n generates a Poisson process Φ_j of offspring of generation $n+1$ with intensity function $\gamma_j(t) = \gamma(t - t_j, Z_j)$, $t > t_j$. Here γ is a non-negative measurable function defined on $(0, \infty)$. We refer to Φ_j as an *offspring process*, and to γ_j and γ as *fertility rates*. Furthermore, the associated mark Z_k to any offspring $t_k \in \Phi_j$ has distribution Q and Z_k is independent of t_k and all (t_l, Z_l) with $t_l < t_k$. As in Daley and Vere-Jones (2003) we refer to this as the case of *unpredictable marks*.

(d) The clusters given the immigrants are independent.

(e) Finally, X consists of the union of all clusters.

Simulation procedures for Hawkes processes are needed for various reasons: Analytical results are rather limited due to the complex stochastic structure; statistical inference, especially model checking and prediction require simulations; displaying simulated realisations of specific model constructions provide a better understanding of the model. The general approach for simulating a (marked or unmarked) point process is to use a thinning algorithm such as Shedler-Lewis thinning algorithm or Ogata's modified thinning algorithm, see e.g. Daley and Vere-Jones (2003). However, Definition 3.1 immediately leads to the following simpler and more efficient simulation algorithm, where $t_- \in [-\infty, 0]$ and $t_+ \in (0, \infty]$ are user-specified parameters, and the output is all marked points (t_i, Z_i) with $t_i \in [0, t_+)$.

Algorithm 3.1 The following steps (i)-(ii) generate a simulation of those marked events $(t_i, Z_i) \in X$ with $0 \leq t_i < t_+$.

(i) Simulate the immigrants on $[t_-, t_+)$.

(ii) For each such immigrant t_i, simulate Z_i and those $(t_j, Z_j) \in C_i$ with $t_i < t_j < t_+$.

Usually in applications steps (i) and (ii) are easy because (a)–(c) in Definition 3.1 are straightforward. As discussed in Section 3.4.4, Algorithm 3.1 and many of our results apply or easily extend to the case where the immigrant process is non-Poisson.

Ideally we should take $t_- = -\infty$, but in practice we need to determine t_- such that $\int_{t_-}^{0} \mu(t)\,\mathrm{d}t < \infty$. When $\int_{-\infty}^{t_-} \mu(t)\,\mathrm{d}t > 0$, Algorithm 3.1 suffers from edge effects, since clusters generated by immigrants before time t_- may contain offspring in $[0, t_+)$. The objective in this paper is to quantify these edge effects and to compare Algorithm 3.1 with the perfect simulation algorithm in Møller and Rasmussen (2004).

The paper is organised as follows. Section 3.2 contains some preliminaries. Section 3.3 contains some convergence results needed in this paper. In Section 3.4 various quantitative results for edge effects are introduced, and among other things we relate our results to those in Brémaud *et al.* (2002) (which concerns approximate simulation of a stationary marked Hawkes process with unpredictable marks). Section 3.5 presents various examples of applications and empirical results for Algorithm 3.1, Ogata's modified thinning algorithm and the perfect simulation algorithm in Møller and Rasmussen (2004).

3.2 Preliminaries

Let F denote the c.d.f. (cumulative distribution function) for L, the length of a cluster, i.e. the time between the immigrant and the last event of the cluster. Consider the mean number of events in any offspring process Φ_i, $\bar{\nu} \equiv \mathrm{E}\nu$, where

$$\nu = \int_0^\infty \gamma(t, Z)\,\mathrm{d}t$$

is the total fertility rate of an offspring process and Z denotes a generic mark with distribution Q. We assume that

$$0 < \bar{\nu} < 1, \tag{3.1}$$

which among other places is needed in Proposition 3.1. This assumption is discussed in detail in Brémaud and Massoulié (2001) and Møller and Rasmussen (2004). Finally, let

$$\bar{h}(t) = \mathrm{E}\gamma(t, Z)/\bar{\nu}, \qquad t > 0, \tag{3.2}$$

which can be interpreted as the normalised intensity function for the first generation of offspring in a cluster started at time 0.

3.3 Approximations of F

It turns out that F is unknown even for very simple cases of Hawkes processes, cf. Møller and Rasmussen (2004).

We first recall some convergence results from Møller and Rasmussen (2004) and next establish a new useful result (Proposition 3.1) which provide useful approximations of F.

For $n \in \mathbb{N}_0$, let 1_n denote the c.d.f. for the length of a cluster when all events of generation $n+1, n+2, \ldots$ are removed. Clearly, 1_n is decreasing in n, $1_n \to F$ pointwise as $n \to \infty$, and

$$1_0(t) = 1, \qquad t \geq 0. \tag{3.3}$$

Let $\mathcal{C}$ denote the class of Borel functions $f : [0, \infty) \mapsto [0, 1]$. For $f \in \mathcal{C}$, define $\varphi(f) \in \mathcal{C}$ by

$$\varphi(f)(t) = \mathrm{E}\left[\exp\left(-\nu + \int_0^t f(t-s)\gamma(s, Z)\,\mathrm{d}s\right)\right], \qquad t \geq 0. \tag{3.4}$$

Then, as verified in Møller and Rasmussen (2004) the assumption of unpredictable marks implies that

$$1_n = \varphi(1_{n-1}), \qquad n \in \mathbb{N}, \tag{3.5}$$

and

$$F = \varphi(F). \tag{3.6}$$

The recursion (3.5) provides a useful numerical approximation to F. As the integral in (3.4) with $f = 1_{n-1}$ quickly becomes difficult to evaluate analytically as n increases, we compute the integral numerically, using a quadrature rule.

Convergence with respect to the supremum norm of 1_n and certain other functions towards F is established in Møller and Rasmussen (2004). In this paper establishing convergence with respect to the L^1-norm becomes relevant. We let $\mathcal{C}_1$ denote the class of functions $f \in \mathcal{C}$ with $\|F - f\|_1 < \infty$, where $\|g\|_1 = \int_0^\infty |g(t)| \, \mathrm{d}t$ is the L^1-norm.

Proposition 3.1 *With respect to the L^1-norm, φ is a contraction on $\mathcal{C}_1$, that is, for all $f, g \in \mathcal{C}_1$ and $n \in \mathbb{N}$, we have that $f_n, g_n \in \mathcal{C}_1$ and*

$$\|\varphi(f) - \varphi(g)\|_1 \leq \bar{\nu} \|f - g\|_1. \tag{3.7}$$

Furthermore, F is the unique fixpoint,

$$\|F - f_n\|_1 \to 0 \qquad \text{as } n \to \infty, \tag{3.8}$$

and if either $f \leq \varphi(f)$ or $f \geq \varphi(f)$, then f_n increases respectively decreases towards F with a geometric rate:

$$\|F - f_n\|_1 \leq \frac{\bar{\nu}^n}{1 - \bar{\nu}} \|\varphi(f) - f\|_1. \tag{3.9}$$

Proof. Let $f, g \in \mathcal{C}_1$. Recall that by the mean value theorem (e.g. Theorem 5.11 in Apostol (1974)), for any real numbers x and y, $\mathrm{e}^x - \mathrm{e}^y = (x - y)\mathrm{e}^{z(x,y)}$, where $z(x, y)$ is a real number between x and y. Thus by (3.4),

$$\begin{aligned} &\|\varphi(f) - \varphi(g)\|_1 \\ &\quad = \int_0^\infty \left| \mathrm{E}\left[\mathrm{e}^{-\nu} \mathrm{e}^{c(t,f,g)} \int_0^t (f(t-s) - g(t-s)) \gamma(s, Z) \mathrm{d}s \right] \right| \mathrm{d}t \end{aligned} \tag{3.10}$$

where $c(t, f, g)$ is a random variable between $\int_0^t f(t-s)\gamma(s, Z)\,\mathrm{d}s$ and $\int_0^t g(t-s)\gamma(s, Z)\,\mathrm{d}s$. Since $f, g \leq 1$, we obtain $\mathrm{e}^{c(t,f,g)} \leq \mathrm{e}^{\nu}$, cf. (3.1). Consequently,

$$\|\varphi(f) - \varphi(g)\|_1 \leq \int_0^\infty \left| \mathrm{E}\left[\int_0^t (f(t-s) - g(t-s))\gamma(s, Z)\,\mathrm{d}s \right] \right| \mathrm{d}t \tag{3.11}$$

$$\leq \mathrm{E}\left[\int_0^\infty \int_0^\infty |f(u) - g(u)|\,\mathrm{d}u\,\gamma(s, Z)\,\mathrm{d}s \right] = \bar{\nu}\,\|f - g\|_1 \tag{3.12}$$

where in the latter inequality we have used first the triangle inequality, next Fubini's theorem, and finally a simple transformation. Thereby (3.7) is verified. The remaining part is verified along similar lines as in the proof of Theorem 1 in Møller and Rasmussen (2004) (with the minor observations that F is the unique fixpoint because of (3.8), and that we use monotone convergence when establishing (3.9)). □

Remark 3.1 The following observation motivates why we restrict attention to the class $\mathcal{C}_1$ in Proposition 3.1, at least when considering functions $f \in \mathcal{C}$ such that $f \leq F$: For such functions f convergence fails as

$$\|F - f\|_1 = \infty \;\Rightarrow\; \|F - f_n\|_1 = \infty, \qquad n \in \mathbb{N}. \tag{3.13}$$

To verify this, consider two non-negative Borel functions $f \leq g$ defined on $[0, \infty)$. Then as in (3.10)–(3.12), but now observing that $c(t, f, g)$ is between 0 and ν,

$$\begin{aligned} \|\varphi(f) - \varphi(g)\|_1 &\geq \mathrm{E}\left[\int_0^\infty \int_0^\infty (g(u) - f(u))\mathrm{e}^{-\nu}\gamma(s, Z)\,\mathrm{d}s\,\mathrm{d}u \right] \\ &= \|f - g\|_1 \mathrm{E}[\nu \mathrm{e}^{-\nu}]. \end{aligned}$$

By (3.1), $\mathrm{E}[\nu \mathrm{e}^{-\nu}] > 0$, and so letting $g = F$, we obtain (3.13) when $n = 1$, whereby (3.13) follows by induction.

As noted the sequence $f_n = 1_n$ decreases towards F pointwise. In order to obtain L^1-convergence by Proposition 3.1 we need $1_0 \in \mathcal{C}_1$, that is, $\mathrm{E}L = \|1 - F\|_1$ is finite. A sufficient and necessary condition for this is given in Lemma 1 in Møller and Rasmussen (2004).

To construct a sequence f_n which increases towards F in the L^1-norm, it suffices to find $f \in \mathcal{C}_1$ such that $f \leq \varphi(f)$. Methods for finding a c.d.f. G with $G \leq \varphi(G)$ are discussed in Møller and Rasmussen (2004) (see in particular Proposition 3 in Møller and Rasmussen (2004)), in which case $G \leq F$ (see Theorem 1 in Møller and Rasmussen (2004)). Note that if $G \leq F$ is a c.d.f. and $\|1-F\|_1 < \infty$, then G needs to have a finite mean, since $\|1-G\|_1 = \|F-G\|_1 + \|1-F\|_1$. □

3.4 Edge effects

Let $N(t_-, t_+)$ denote the number of missing events when using Algorithm 3.1. In this section we consider the mean number of missing offspring, $\mathrm{E}(t_-, t_+) \equiv \mathrm{E}N(t_-, t_+)$, and the probability of having any missing offspring, $\mathrm{P}(t_-, t_+) \equiv \mathrm{P}(N(t_-, t_+) > 0)$. Furthermore, we relate these to the total variation distance between simulations and the target distribution.

3.4.1 The mean number of missing offspring

Consider a cluster $C_0 = \{(s_i, Z_i)\}$ started at time $t_0 = 0$. This has conditional intensity function

$$\lambda_0(t) = \gamma(t, Z_0) + \sum_{0<s_i<t} \gamma(t - s_i, Z_i), \qquad t \geq 0, \tag{3.14}$$

and unpredictable marks with distribution Q. For $t > 0$, let $\lambda(t) = \mathrm{E}\lambda_0(t)$ be the intensity function of the offspring in C_0, and $\bar{\gamma}(t) = \mathrm{E}\gamma(t, Z) = \bar{\nu}\bar{h}(t)$ be the intensity function of the first generation of offspring in C_0. The following proposition expresses $\mathrm{E}(t_-, t_+)$ and $\lambda(t)$ in terms of μ and $\bar{\gamma}$.

Proposition 3.2 *We have that*

$$\lambda(t) = \sum_{n=1}^{\infty} \bar{\gamma}^{*n}(t) = \sum_{n=1}^{\infty} \bar{\nu}^n \bar{h}^{*n}(t), \qquad t \geq 0, \tag{3.15}$$

where $*n$ *denotes convolution* n *times, and*

$$\mathrm{E}(t_-, t_+) = \int_{-\infty}^{t_-} \left(\int_{-t}^{t_+ - t} \lambda(s)\,\mathrm{d}s \right) \mu(t)\,\mathrm{d}t. \tag{3.16}$$

Proof. We claim that $\rho_n = \bar{\gamma}^{*n}$ is the intensity function of $\mathcal{G}_n$, the n-th generation of offspring in the cluster C_0: This is clearly true for $n = 1$, and so by induction

$$\begin{aligned} \rho_{n+1}(t) &= \mathrm{E} \sum_{s_i \in \mathcal{G}_n} \gamma(t - s_i, Z_i) = \mathrm{E} \sum_{s_i \in \mathcal{G}_n} \mathrm{E}[\gamma(t - s_i, Z_i) | s_i] \\ &= \mathrm{E} \sum_{s_i \in \mathcal{G}_n} \bar{\gamma}(t - s_i) = \int_0^t \rho_n(s) \bar{\gamma}(t - s)\,\mathrm{d}s = \bar{\gamma}^{*(n+1)} \end{aligned}$$

where we have used Campbell's theorem in the second last equality and the induction hypothesis in the last equality. Thereby (3.15) follows. Finally, if I denotes the Poisson process of immigrants,

$$\begin{aligned} \mathrm{E}(t_-, t_+) &= \mathrm{E} \sum_{t_i \in I} \sum_{s \in C_i} \mathbf{1}[t_i < t_-,\, 0 \leq s < t_+] \\ &= \mathrm{E} \sum_{t_i \in I:\, t_i < t_-} \mathrm{E} \left[\sum_{s \in C_i} \mathbf{1}[0 \leq s < t_+] \,\middle|\, t_i \right] \\ &= \mathrm{E} \sum_{t_i \in I:\, t_i < t_-} \int_{-t_i}^{t_+ - t_i} \lambda(u)\,\mathrm{d}u \end{aligned}$$

which reduces to (3.16) by Campbell's theorem. □

The convolution $\bar{\gamma}^{*n}$ is only computable for a few kinds of models (see Section 3.5).

Remark 3.2 It follows immediately that

$$\rho = \mu + \mu * \lambda \tag{3.17}$$

is the intensity function of all events. When quantifying edge effects it is natural to consider $\mathrm{E}(t_-, t_+)/\mathrm{E}(t_+)$, where

$$\mathrm{E}(t_+) = \int_0^{t_+} \rho(t)\,\mathrm{d}t$$

is the expected number of events on $[0, t_+]$. □

3.4.2 The probability of having any missing offspring

Obviously, $\mathrm{P}(t_-, t_+)$ is an increasing function of $t_+ \in (0, \infty]$. Proposition 3.3 gives an expression and upper and lower bounds for $\mathrm{P}(t_-, \infty)$.

Proposition 3.3 *We have that*

$$\mathrm{P}(t_-, \infty) = 1 - \exp\left(-\int_{-\infty}^{t_-} (1 - F(-t))\mu(t)\,\mathrm{d}t\right). \tag{3.18}$$

Further, for any $f \in \mathcal{C}_1$ such that $f \leq \varphi(f)$, we have an upper bound,

$$\mathrm{P}(t_-, \infty) \leq 1 - \exp\left(-\int_{-\infty}^{t_-} (1 - f_n(-t))\mu(t)\,\mathrm{d}t\right), \tag{3.19}$$

which is a decreasing function of n, and a lower bound

$$\mathrm{P}(t_-, \infty) \geq 1 - \exp\left(-\int_{-\infty}^{t_-} (1 - 1_n(-t))\mu(t)\,\mathrm{d}t\right), \tag{3.20}$$

which is a increasing function of n.

Proof. Let I_{t_-} be the point process of immigrants $t_i < t_-$ with $\{(t_j, Z_j) \in C_i : t_j \geq 0\} \neq \emptyset$. Then I_{t_-} is a Poisson process with intensity function $\lambda_{t_-}(t) = (1 - F(-t))\mu(t)$ on $(-\infty, t_-)$, since we can view I_{t_-} as an independent thinning of the immigrant process on $(-\infty, t_-)$, with retention probabilities $p(t) = 1 - F(-t)$, $t < t_-$. Hence, since $\mathrm{P}(t_-, \infty)$ equals the probability that $I_{t_-} \neq \emptyset$, we obtain (3.18). Thereby (3.19) and (3.20) follows from (3.18) and Proposition 3.1. □

Remark 3.3 Proposition 3.1 ensures that the upper bound in (3.19) and the lower bound in (3.20) converge monotoneously to $\mathrm{P}(t_-, \infty)$ provided e.g. that μ is bounded and $\mathrm{E}L < \infty$, cf. Remark 3.1. □

3.4.3 The total variation distance between simulations and the target distribution

Recently, Brémaud *et al.* (2002) derived related results to Propositions 3.2 and 3.3 when $\mu(t)$ is constant and $t_+ = \infty$. Proposition 3.4 below generalises certain inequalities in Brémaud *et al.* (2002) (page 133, line 3 from below, and equation (3.13)) to the situation in the present paper where $\mu(t)$ is not necessarily constant and t_+ may be finite. Moreover, our proof is much simpler.

We let $\tilde{X}$ be another marked Hawkes process obtained from X by removing all clusters C_i with immigrants $t_i < t_-$. Furthermore, we let Y and $\tilde{Y}$ denote the restriction of X and $\tilde{X}$ to the marked events on $[0, t_+)$, and denote their distributions by $\pi(t_-, t_+)$ and $\tilde{\pi}(t_-, t_+)$. Thus the output of Algorithm 3.1 follows $\tilde{\pi}(t_-, t_+)$, which approximates the target distribution $\pi(t_-, t_+)$.

Proposition 3.4 *Let* $\|\cdot\|_{\mathrm{TV}}$ *denote the total variation distance, then*

$$\|\pi(t_-, t_+) - \tilde{\pi}(t_-, t_+)\|_{\mathrm{TV}} \leq \mathrm{P}(t_-, t_+) \leq \mathrm{E}(t_-, t_+). \tag{3.21}$$

Proof. By the construction of $\tilde{Y}$, we have that $\tilde{Y} \subseteq Y$. The first inequality then follows immediately from the coupling inequality (see

e.g. Lindvall (1992)), while the second inequality is trivially satisfied. □

Remark 3.4 In contrast to the first upper bound in (3.21) the second upper bound does not depend on knowing F or any approximation of F, cf. Propositions 3.2 and 3.3. □

3.4.4 Extensions and open problems

It would be of practical importance to extend our results to the case of predictable marks. Proposition 3.4 is still true if the conditional intensity function for X is larger than or equal to the conditional intensity function for $\tilde{X}$; this follows by a thinning argument, cf. Daley and Vere-Jones (2003). However, this observation seems of little use, since the assumption of unpredictable marks is essential in the proofs of (3.15) in Proposition 3.2 and (3.19)–(3.20) in Proposition 3.3. Moreover, though (3.18) in Proposition 3.3 remains true, it is expected to be of limited use, since F is expected to be of a more complicated form in the case of predictable marks.

The following observations may also be of practical relevance.

Algorithm 3.1 applies for a non-Poisson immigrant process, e.g. a Markov or Cox process provided it is feasible to simulate the immigrants on $[t_-, t_+)$. Furthermore, Proposition 3.2 remains true for any immigrant process with intensity function μ. Finally, Proposition 3.3 partly relies on the immigrants being a Poisson process: for instance, if now μ is a random intensity function and the immigrant process is a Cox process driven by μ, then (3.18)–(3.20) should be modified by taking the mean of the expressions on the right hand sides.

3.5 Examples and comparison with perfect simulation

Illustrative examples of specific unmarked and marked Hawkes processes (with plots showing perfect simulations) are given in Møller and Rasmussen (2004). In this section we consider the same examples of models and demonstrate the use and limitations of our results in Section 3.4. We also demonstrate the practical differences between Algorithm 3.1 and the perfect simulation algorithm in Møller and Rasmussen (2004).

3.5.1 An unmarked Hawkes process model

The events and marks of X are independent if and only if $\gamma(t, z) = \gamma(t)$ does not depend on the mark z (for almost all z) in which case the events form an unmarked Hawkes process. In this section we consider an unmarked Hawkes process with exponentially decaying fertility rate given by $\gamma(t) = \alpha\beta \mathrm{e}^{-\beta t}$, where $0 < \alpha < 1$ and $\beta > 0$ are parameters.

Note that $1/\beta$ is a scale parameter for the distribution of L, $\bar{\nu} = \nu = \alpha$, and $\bar{h} = \beta \mathrm{e}^{-\beta t}$. Hence $\bar{h}^{*n}$ is the density for a gamma distribution with shape parameter n and inverse scale parameter β. Using (3.15), we obtain $\lambda(t) = \alpha\beta \mathrm{e}^{(\alpha-1)\beta t}$. Inserting this into (3.16), assuming that $t_- > -\infty$ and $\mu(t) = \delta \mathrm{e}^{\kappa t}$ where $\delta > 0$ and $\kappa > (\alpha - 1)\beta$ are parameters, we obtain that

$$\mathrm{E}(t_-, t_+) = \frac{\alpha\delta}{(1-\alpha)((1-\alpha)\beta + \kappa)}(1 - \mathrm{e}^{(\alpha-1)\beta t_+})\mathrm{e}^{((1-\alpha)\beta+\kappa)t_-}.$$

Here the restriction on κ is equivalent to that ρ is finite, in which case $\rho(t) = \delta \mathrm{e}^{\kappa t}(\kappa + \beta)/(\kappa + (1-\alpha)\beta)$, cf. (3.17).

Figure 3.1 shows $\mathrm{E}(t_-, t_+)/\mathrm{E}(t_+)$ as a function of $-t_- \geq 0$ in the case $\alpha = 0.9$, $\delta = \beta = 1$, $t_+ = 10$, and for different values of κ. As expected numerically smaller values of t_- are needed as

κ increases. For $\kappa \geq 0$, effectively perfect simulation are produced when $t_- = -50$.

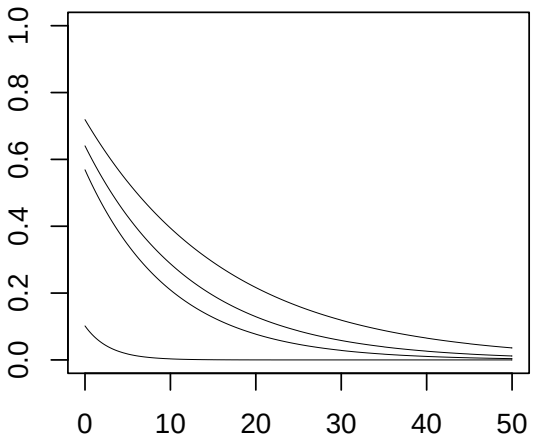


Figure 3.1: Plot of $\mathrm{E}(t_-, t_+)/\mathrm{E}(t_+)$ versus $-t_-$ for the unmarked case with parameters $\alpha = 0.9$, $\delta = \beta = 1$, $t_+ = 10$, and $\kappa = -0.04, -0.02, 0, 0.25$ (top to bottom).

Let $f(t) = 1 - \mathrm{e}^{-\theta t}$ be the c.d.f. for an exponential distribution with parameter $\theta = \beta(1 - \alpha)$. As verified in Møller and Rasmussen (2004), $f \leq \varphi(f)$, and so the bounds of $\mathrm{P}(t_-, \infty)$ in Proposition 3.3 hold. Figure 3.2 shows these bounds when $\alpha = 0.9$, $\beta = \delta = 1$ and $\kappa = 0$ (i.e. $\mu = 1$), and $n = 0, 7, \ldots, 70$. The convergence of the bounds to $\mathrm{P}(t_-, t_+)$ is clearly visible, and for $n = 70$ both bounds are practically equal. Also the plot reveals that for the present choice of parameters, the probability for having one or more missing events is effectively 0 for $t_- = -50$.

We can determine $N(t_-, t_+)$, or at least its distribution, from the perfect simulation algorithm in Møller and Rasmussen (2004). Figure 3.3 shows one minus the corresponding empirical distribution

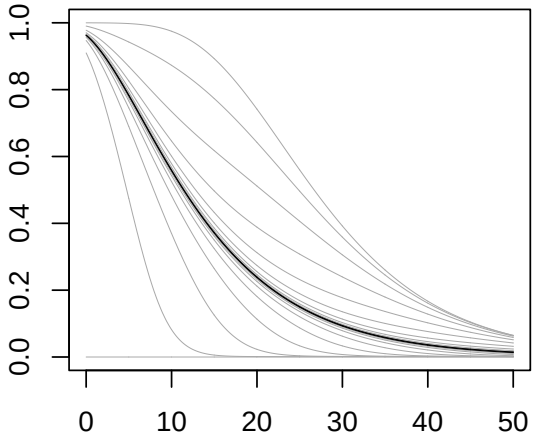


Figure 3.2: Upper and lower bounds (3.19) and (3.20) of $P(t_-, t_+)$ versus $-t_-$ in the unmarked case with $\alpha = 0.9$, $\mu = \beta = 1$, $t_+ = \infty$, and $n = 0, 7, \ldots, 70$. The bounds using $n = 70$ are shown in black to illustrate the approximate form of $\mathrm{P}(t_-, t_+)$, whereas the rest are shown in gray.

function based on 10000 perfect simulations when $\alpha = 0.9$, $\beta = \delta = 1$, $\kappa = 0$, $t_+ = 10$, and $t_- = 0, -10$, or -50. In each of the three cases, since $\mathrm{E}(t_+) = 100$, the number of missing events in the case $t_- = 0$ is substantially reduced, but still too large, when $t_- = -10$, while edge effects are practically non-existent for $t_- = -50$.

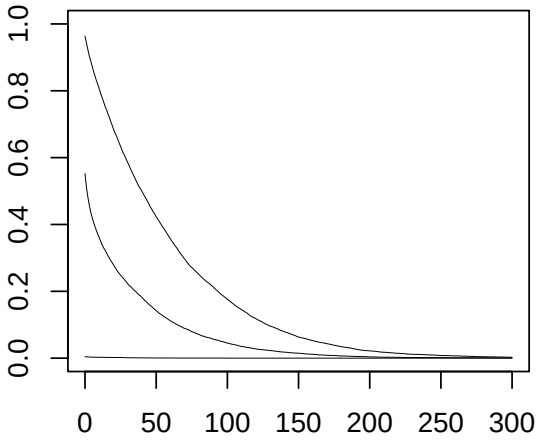


Figure 3.3: One minus the empirical distribution function for $N(t_-, t_+)$ in the unmarked case with $\alpha = 0.9$, $\beta = \delta = 1$, $\kappa = 0$, $t_+ = 10$, and $t_- = 0, -10, -50$ (top to bottom).

Comparing Figures 3.1–3.3 when for example $\alpha = 0.9$, $\beta = \delta = 1$, $\kappa = 0$, $t_+ = 10$, and $t_- = -50$, Algorithm 3.1 and the perfect simulation algorithm from Møller and Rasmussen (2004) are effectively producing identical results. Algorithm 3.1 uses roughly one-thousandth of a second for each simulation in our implementation, while the perfect simulation algorithm uses one-tenth of a second. For comparison we have implemented Ogata's modified thinning algorithm (Ogata, 1981), which is a general simulation algorithm for unmarked processes. Ogata's algorithm and our Algorithm 3.1 produce simu-

lations with the same distribution; however, Ogata's algorithm uses approximately one-twentieth of a second to make an approximate simulation.

3.5.2 A marked Hawkes process model with birth and death transitions

Consider a marked Hawkes process with

$$\gamma(t, z) = \alpha \mathbf{1}[t \leq z]/\mathrm{E}Z,$$

where $0 < \alpha < 1$ is a parameter, Z is a positive random variable with distribution Q, and $\mathbf{1}[\cdot]$ denotes the indicator function. Then X can be viewed as a birth and death process, with birth at time t_i and survival time Z_i of the i'th individual (more general birth and death processes (even non-Markovian) are considered in Brémaud and Massoulié (2002) and Torrisi (2002)).

The special case where $\mu(t) = \mu$ is constant and Z is exponentially distributed with mean $1/\beta$ is considered at page 136 in Brémaud *et al.* (2002). Since $\bar{h}(t) = \beta \mathrm{e}^{-\beta t}$ is the same function as in Section 3.5.1, $\mathrm{E}(t_-, t_+)$ is also the same as in Section 3.5.1. Further, a plot of $P(t_-, t_+)$ (omitted here) is similar to Figure 3.2 (when using the same parameters). Also a plot of the empirical distribution function of $N(t_-, t_+)$ (omitted here) is similar to Figure 3.3.

When for example $\alpha = 0.9$, $\beta = \mu = 1$, $t_+ = 10$, and $t_- = -50$, Algorithm 3.1 uses roughly one-five hundredth of a second for each simulation, and the perfect simulation algorithm uses just under three seconds. As in the unmarked case both algorithms are feasible, but the difference is much more clear in the present case.

3.5.3 A heavy-tailed distribution for L

We conclude by observing that heavy-tailed cases of the distribution of L are problematic. For instance, suppose that

$$\gamma(t, z) = \alpha z \mathrm{e}^{-tz}, \tag{3.22}$$

where $\alpha \in (0,1)$ is a parameter, and let Q be the exponential distribution with mean $1/\beta$. As argued in Møller and Rasmussen (2004), $\bar{h}(t) = \beta/(t+\beta)^2$ is a Pareto density and L has a heavy-tailed distribution with infinite moments and infinite Laplace transform. As $\mathrm{E}L = \infty$, Proposition 3.1 and hence Proposition 3.3 seem of rather limited use, cf. Remark 3.1. Proposition 3.2 in Brémaud *et al.* (2002) establish and upper bound for $P(t_-, \infty)$ for such heavy-tailed cases under certain mild conditions, however, in the case of 3.22 their result does not apply. Proposition 3.2 is also not applicable, since λ is not known on closed form, cf. Example 7 in Møller and Rasmussen (2004). It is a challenging open problem to handle such heavy-tailed cases.

Acknowledgements

The research of Jesper Møller was supported by the Danish Natural Science Research Council and the Network in Mathematical Physics and Stochastics (MaPhySto), funded by grants from the Danish National Research Foundation.

References

Apostol, T. M. (1974). *Mathematical Analysis*. Addison-Wesley, Reading.

Brémaud, P. and Massoulié, L. (2001). Hawkes branching point processes without ancestors. *J. Appl. Prob.*, **38**, 122–135.

Brémaud, P. and Massoulié, L. (2002). Power spectra of general shot noises and Hawkes point processes with a random excitation. *Adv. Appl. Prob.*, **34**, 205–222.

Brémaud, P., Nappo, G., and Torrisi, G. (2002). Rate of convergence

to equilibrium of marked Hawkes processes. *J. Appl. Prob.*, **39**, 123–136.

Chornoboy, E. S., Schramm, L. P., and Karr, A. F. (2002). Maximum likelihood identification of neural point process systems. *Adv. Appl. Prob.*, **34**, 267–280.

Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods*. Springer, New York, 2nd edition.

Hawkes, A. G. (1971a). Point spectra of some mutually exciting point processes. *J. Roy. Statist. Soc. Ser. B*, **33**, 438–443.

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

Hawkes, A. G. (1972). Spectra of some mutually exciting point processes with associated variables. In P. A. W. Lewis, editor, *Stochastic Point Processes*, pages 261–271. Wiley, New York,.

Hawkes, A. G. and Adamopoulos, L. (1973). Cluster models for earthquakes – regional comparisons. *Bull. Int. Statist. Inst.*, **45**, 454–461.

Hawkes, A. G. and Oakes, D. (1974). A cluster representation of a self-exciting process. *J. Appl. Prob.*, **11**, 493–503.

Lindvall, T. (1992). *Lectures on the Coupling Method*. Wiley, New York.

Møller, J. and Rasmussen, J. G. (2004). Perfect simulation of Hawkes processes. Research report R-2004-18, Department of Mathematical Sciences, Aalborg University. Available at http://www.math.aau.dk.

Ogata, Y. (1981). On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, **IT-27**(1), 23–31.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Assoc.*, **83**(401), 9–27.

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

Torrisi, G. L. (2002). A class of interacting marked point processes: rate of convergence to equilibrium. *J. Appl. Prob.*, **39**, 137–160.

Vere-Jones, D. and Ozaki, T. (1982). Some examples of statistical inference applied to earthquake data. *Ann. Inst. Statist. Math.*, **34**, 189–207.

Chapter 4

Spatial-temporal modeling of forest gaps generated by colonization from below- and above-ground bark beetle species

Jun Zhu
Department of Statistics, University of Wisconsin-Madison.
Email addresses: jzhu@stat.wisc.edu

Jesper Møller & Jakob G. Rasmussen
Department of Mathematical Sciences, Aalborg University.
Email addresses: jm@math.aau.dk & jgr@math.aau.dk

Brian H. Aukema & Kenneth F. Raffa
Department of Entomology, University of Wisconsin-Madison.
Email addresses: aukema@entomology.wisc.edu
& raffa@entomology.wisc.edu

Abstract

Studies of forest declines are important, because they both reduce timber production and affect successional trajectories of landscapes and ecosystems. Of particular interest is the decline of red pines which is characterized by expanding areas of dead and chlorotic trees in plantations throughout the Great Lakes Region. Here we examine the impact of two bark beetle groups, namely red turpentine beetles and pine engraver bark beetles, on tree mortality and the subsequent gap formation over time in a plantation in Wisconsin. We construct spatial-temporal statistical models that quantify the relations among red turpentine beetle colonization, pine engraver bark beetle colonization, and mortality of red pine trees, while accounting for correlation across space and over time. For statistical inference, we adopt a Bayesian hierarchical model and devise Markov chain Monte Carlo algorithms for obtaining the posterior distributions of model parameters as well as posterior predictive distributions. Our data analysis results suggest that red turpentine beetle colonization is associated with higher likelihood of pine engraver bark beetle colonization and pine engraver bark beetle colonization is associated with higher likelihood of red pine tree mortality, whereas there is no direct association between red turpentine beetle colonization and red pine tree mortality. There is strong evidence that red turpentine beetle colonization does not kill a red pine tree directly, but rather predisposes the tree to subsequent colonization by pine engraver bark beetles. The evidence is also strong that pine engraver bark beetles are the ultimate mortality agents of red pine trees.

Keywords: Autologistic model, Bayesian inference, forest entomology, Markov chain Monte Carlo, perfect simulation, spatial-temporal processes.

4.1 Introduction

Studies of forest declines are of great interest to the timber industry, natural resource managers, and ecologists alike, because these declines both reduce timber production and affect successional trajectories of landscapes and ecosystems. Decline syndromes occur in forests throughout the world, and occur at a variety of scales (Auclair, 2005). Declines due to soil acidification and atmospheric pollution may affect large areas (Battles and Fahey, 2000; Drohan *et al.*, 2002; Purdon *et al.*, 2004) while declines due to insect and/or disease complexes may exhibit smaller mosaics of mortality such as gap formation, which is our focus here (Klepzig *et al.*, 1991; Erbilgin and Raffa, 2003). In some systems, areas of large-scale mortality to insects and pathogens may originate from such small-scale mosaics. Characterizing spatial patterns and gaining inference about the processes that may create such patterns may assist in policy and management decisions when dealing with declines. Indeed, linking pattern and process is a key goal in ecology.

In particular we examine tree mortality and the subsequent gap formation over time in red pine forests. Decline of red pines is characterized by expanding areas, termed "pockets" of dead, slow growing, and chlorotic trees in plantations throughout the Great Lakes Region (Klepzig *et al.*, 1991). Abiotic factors such as soil characteristics and drought stress can predispose trees to biotic mortality agents such as insects and root pathogens (Klepzig *et al.*, 1991; Erbilgin and Raffa, 2002). Here we focus on the impact of two bark beetle groups, called "turpentine beetles" and "*Ips* spp." (for details of the species, see Section 4.2), on the decline of red pines in a plantation in Wisconsin.

Past studies on red pine decline have yielded valuable insights on individual components of this system by examining multiple levels of scale, from detailed studies on individual trees (Klepzig *et al.*, 1995; Raffa and Smalley, 1995; Klepzig *et al.*, 1996) to regional studies comparing multiple sites (Klepzig *et al.*, 1991; Erbilgin and Raffa, 2002, 2003). Despite these advances, elucidation of exact mechanisms

of pocket development and expansion remain elusive since a single site has never been observed over more than three years. In the present study, we examine a seven-year data set of annual surveys of all trees in a plantation. Each year, each of the 2,715 trees was examined for presence/absence of *Ips* spp., tree condition (alive/dead), and the number of pitch tubes, each of which signifies colonization by a turpentine beetle. We attempt to answer several important ecological questions. Of most interest is how the mortality rate of a tree is associated with the turpentine beetle and *Ips* spp. colonization. For example, how different are the mortality rates between a tree that has been colonized by both groups and a tree that has been colonized by just one group of bark beetles? Related to these questions are the degree of association between turpentine beetle and *Ips* spp. For example, what is the likelihood of a tree that has been colonized by turpentine beetles will be colonized by *Ips* spp. subsequently? Moreover it is also of interest to quantify the spatial and temporal relations among turpentine beetle colonization, *Ips* spp. colonization, and tree mortality.

The study of red pine declines poses statistical challenges, in that processes giving rise to patterns of mortality may act at different levels of temporal and spatial scales. Here we construct spatial-temporal models that quantify the relations among the activities of turpentine beetle, the activities of *Ips* spp., and the conditions of red pine trees. Furthermore, we introduce spatial and temporal terms into the model that account for correlations across space and over time. For statistical inference, we adopt a Bayesian hierarchical model and Markov chain Monte Carlo (MCMC) algorithms that enable us to obtain the posterior distributions of the model parameters and posterior predictive distributions. The model for *Ips* spp. also involves an unknown normalizing constant. Thus when we use a Metropolis-Hastings algorithm, we approximate a ratio of normalizing constants by path sampling (Gelman and Meng, 1998) combined with the Propp-Wilson algorithm for perfect simulation (Propp and Wilson, 1996; Møller, 1999).

The remainder of the paper is organized as follows. In Section 4.2, we describe some more biological background and the data. In Section 4.3, we specify a set of spatial-temporal models for the data. The Bayesian model and simulation algorithms are specified in Section 4.4. We describe the results of the data analysis and address the ecological questions in Section 4.5. Section 4.6 contains concluding remarks.

4.2 Bark beetle and red pine data

4.2.1 Background

Bark beetle species are characterized by their ability to mine and reproduce below the surface of the bark of trees. The red turpentine beetle (*Dendroctonus valens* (LeConte)), which we in short call "turpentine beetle", is one of the most widely distributed bark beetles in North America. Colonization by turpentine beetle adults concentrate in the lower stems of pine trees. The larvae breed largely below the soil line in the root collar and primary lateral root regions. An external indicator of colonization by the turpentine beetle is a pitch tube on the outer surface of the bark just above the soil line or pitch pellets on the ground. Peak flight and colonization in Wisconsin occur in late April and May. Turpentine beetles colonize primarily trees that are weakened by drought and fire, for example, but may also colonize apparently healthy trees. These beetles vector the staining fungi *Leptographium terebrantis* and *L. procerum* (Klepzig *et al.*, 1991). It is hypothesized that a colonization of a healthy tree by turpentine beetles does not kill the tree but may predispose it to subsequent colonization by other bark beetles such as engraver beetles.

Engraver beetles (predominantly *Ips pini* (Say), although additionally some *Ips grandicollis* (Eichhoff) in our study site (Klepzig *et al.*, 1991)), which we in short call "*Ips* spp.", may have up to three generations from spring to early fall (Erbilgin *et al.*, 2002; Erbilgin and Raffa, 2002; Aukema *et al.*, 2005). Successful colonization

by the *Ips* spp. is indicated by fine sawdust shavings pushed to the outer surface of the bark and galleries inside the tree. *Ips* spp. beetles produce aggregation pheromones as they enter host trees, thus attacking trees *en masse* over very short periods. These mass attacks typically result in complete utilization of the resource within a single generation, making it unlikely that subsequent *Ips* spp. or turpentine beetles will enter. *Ips* spp. also vector the fungal associate *Ophiostoma ips*, whose colonization may impede the upward flow of water to the tree crowns. Lack of water leads the needles to wither and die while the color characteristically fades from green to red to brown. *Ips* spp. brood adults leave the tree through emergence holes on the surface of the outer bark, the most apparent external indicator that a tree has been colonized by *Ips* spp. The tree is most likely to die within a few weeks after an attack.

4.2.2 Description of data

The study area is a red pine plantation near Spring Green, Wisconsin. In 1986, each of the 2,715 trees in the plantation was surveyed and its condition (alive/dead) was recorded. From 1987 to 1992, subsequent surveys were conducted not only of the tree condition, but also about the colonization of turpentine beetles and *Ips* spp. For turpentine beetles, the number of pitch tubes on the outer surface of a bark was recorded, while for *Ips* spp., an indicator variable of *Ips* spp. colonization (yes/no) was recorded. The survey took place in autumn of each year, after beetles had become dormant.

Selected image plots in Figure 4.1 illustrate the nature of the data. For 1987, colonization of turpentine beetles (zero or positive number of pitch tubes), colonization of *Ips* spp. (yes/no), and condition of trees (alive/dead) are shown (Figure 4.1(a)–(c)). For 1992, similar plots are shown, except that colonization of *Ips* spp. here includes colonization from 1987 to 1992 (Figure 4.1(d)–(f)). There is clear indication of spatial dependence among tree conditions, *Ips* spp. colonization, and turpentine beetle colonization. A gap of dead trees

was evident in the southeastern part of the plantation in the beginning and expanded over the years. Furthermore there was a strong association between the spatial pattern of *Ips* spp. colonization and that of tree mortality, but the link was not as obvious between turpentine beetle colonization and tree mortality.

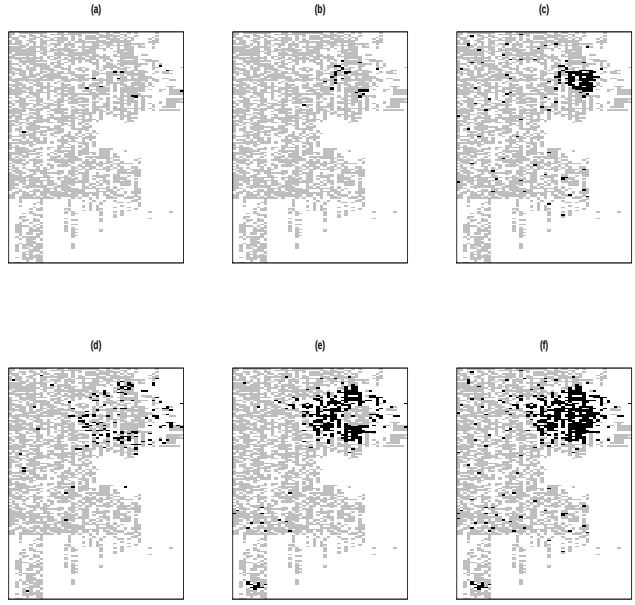
There were 126 dead trees in 1986. From 1987 to 1992, a total of 344 trees died, 339 trees were colonized by *Ips* spp., and 152 trees were colonized by turpentine beetles. Among the 344 dead trees, a majority of 330 were colonized by *Ips* spp. and 73 were colonized by turpentine beetles. Only 9 out of 339 trees that were colonized by *Ips* spp. survived by 1992, whereas 79 out of 152 of the trees that were colonized by turpentine beetles survived. *Ips* spp. colonization seemed to associate more with those trees with a larger number of pitch tubes of turpentine beetles, although the evidence was subtle due to the small number of trees that had a large number of pitch tubes.

4.3 Observation model

4.3.1 Notation

Let $t = -1, 0, \ldots, 5$ index the time of survey from 1986 to 1992 and let $i = 1, \ldots, 2715$ index the sites of 2,715 red pine trees in the plantation that were surveyed. For the purpose of modeling, we consider time points $t = \ldots, -1, 0, 1, \ldots$ and define $x_{t,i}$, $y_{t,i}$, $z_{t,i}$, and $u_{t,i}$ as follows. Since the survey was conducted in autumn, after insect and tree dormancy for any given year, the data reflect insect activities and tree conditions throughout that year. At time t and site i, let $x_{t,i}$ denote the turpentine beetle colonization variable such that $x_{t,i}$ is the cumulative number of turpentine beetle pitch tubes on the bark. Further, let $y_{t,i}$ denote the *Ips* spp. colonization variable such that $y_{t,i} = 0, 1, 2$ correspond respectively to no *Ips* spp. colonization, colonization by *Ips* spp. in year t, and colonization by *Ips* spp. in a previous year. Let $u_{t,i}$ denote an indicator variable of whether *Ips*

Figure 4.1: (a) and (d) turpentine beetle colonization, (b) and (e) *Ips* spp. colonization, and (c) and (f) tree condition by 1987 (top row) and 1992 (bottom row). The site of a tree is colored black if the tree was colonized by turpentine beetles ((a) and (d)), colonized by *Ips* spp. ((b) and (e)), or dead ((c) and (f)); all other sites are colored grey.



spp. colonization took place during year t at site i, i.e. $u_{t,i} = 1$ if $y_{t,i} = 1$ and $u_{t,i} = 0$ otherwise. In consistency with the data, we assume that *Ips* spp. colonization could only occur once at a given site and after colonization of a tree, *Ips* spp. leaves the tree before the end of the flight season of the same year (before the annual survey). Thus $u_{t,i} = 1$ for at most one year t. Finally, let $z_{t,i}$ denote the tree condition variable such that $z_{t,i} = 0$ if the tree was alive and $z_{t,i} = 1$ if the tree was dead at time t and site i.

We let $x_t = (x_{t,1}, \ldots, x_{t,2715})$, $y_t = (y_{t,1}, \ldots, y_{t,2715})$, and $z_t = (z_{t,1}, \ldots, z_{t,2715})$ denote the vectors of respectively the turpentine beetle colonization variables, the *Ips* spp. colonization variables, and the tree condition variables at time t and all the sites. Further, let $u_t = (u_{t,1}, \ldots, u_{t,2715})$ and $w_t = (x_t, y_t, z_t)$. Since turpentine beetle colonization typically precedes *Ips* spp. colonization, which in turn precedes the death of a tree, we order the variables x_t, y_t, z_t such that x_t precedes y_t and y_t precedes z_t. Thus, the data under study are ordered as $(z_{-1}, x_0, y_0, z_0, \ldots, x_5, y_5, z_5)$, while the unobserved data in the past are ordered as $(\ldots, x_{-2}, y_{-2}, z_{-2}, x_{-1}, y_{-1})$.

4.3.2 Temporal model

In Sections 4.3.3–4.3.5, we shall construct a set of spatial-temporal models to capture the relations among the variables $x_{t,i}$, $y_{t,i}$, and $z_{t,i}$, while accounting for spatial and temporal dependence. Before specifying these details, it is useful to give a brief description of the temporal process w_t and how the likelihood factorizes.

In equations (4.1)–(4.3) below we naturally consider a sequential model construction such that for each time t, we specify the conditional distribution of x_t first, y_t second, z_t third given the relevant past. The detailed model descriptions (4.5), (4.8), and (4.12) in Sections 4.3.3–4.3.5 imply the following conditional independence structure for the temporal process. Let $[a|b]$ denote the conditional distribution of a random component a given another random component b. For the turpentine beetle colonization variables at time

t,

$$[x_t|(w_s)_{s=t-1,t-2,\ldots}] \sim [x_t|x_{t-1}, z_{t-1}] \tag{4.1}$$

depends on a parameter θ as specified in Section 4.3.3; for the *Ips* spp. beetle colonization variables at time t,

$$[y_t|x_t, (w_s)_{s=t-1,t-2,\ldots}] \sim [y_t|x_t, y_{t-1}, z_{t-1}] \tag{4.2}$$

depends on a parameter ψ (Section 4.3.4); for the tree condition variables at time t,

$$[z_t|x_t, y_t, (w_s)_{s=t-1,t-2,\ldots}] \sim [z_t|x_t, y_t, z_{t-1}] \tag{4.3}$$

depends on a parameter φ (Section 4.3.5). For the corresponding likelihood terms, we write $L^{(1)}(\theta) = L^{(1)}(\theta; x_t|x_{t-1}, z_{t-1})$, $L^{(2)}(\psi) = L^{(2)}(\psi; y_t|x_t, y_{t-1}, z_{t-1})$, and $L^{(3)}(\varphi) = L^{(3)}(\varphi; z_t|x_t, y_t, z_{t-1})$.

For statistical inference we condition on

$$e = (z_{-1}, x_0, y_0)$$

since by (4.1)–(4.3), the remaining data

$$d = (z_0, x_1, y_1, z_1, \ldots, x_5, y_5, z_5)$$

are conditionally independent of the unobserved ($\ldots$, x_{-2}, y_{-2}, z_{-2}, x_{-1}, y_{-1}). We let $L(\theta, \psi, \varphi) = L(\theta, \psi, \varphi; d|e)$ denote the likelihood based on the conditional distribution of d given e. By (4.1)–(4.3), this factorizes into

$$L(\theta, \psi, \varphi; d|e) = L^{(1)}(\theta)L^{(2)}(\psi)L^{(3)}(\varphi)$$

given by the likelihood terms for each type of data

$$L^{(1)}(\theta) = \prod_{t=1}^{5} L_t^{(1)}(\theta), L^{(2)}(\varphi) = \prod_{t=1}^{5} L_t^{(2)}(\varphi), L^{(3)}(\psi) = \prod_{t=0}^{5} L_t^{(3)}(\psi), \tag{4.4}$$

where $L_t^{(1)}(\theta) = L_t^{(1)}(\theta; x_t|x_{t-1}, z_{t-1})$, $L_t^{(2)}(\psi) = L_t^{(2)}(\psi; y_t|x_t, y_{t-1}, z_{t-1})$, and $L_t^{(3)}(\varphi) = L_t^{(3)}(\varphi; z_t|x_t, y_t, z_{t-1})$ are specified at the end of Sections 4.3.3–4.3.5.

In Sections 4.3.3–4.3.5, our strategy is for each time, site, and type of data $x_{t,i}$, $y_{t,i}$, or $z_{t,i}$ to specify a "local characteristic" which only depends on "local information". For example, by the local characteristic of $y_{t,i}$, we mean the conditional distribution of $y_{t,i}$ given the other $y_{t,j},\ j \neq i$ and the previous history $x_t, (w_s)_{s=t-1,t-2,\ldots}$. We express the local information with respect to the grid of tree locations and consider for site i its first-, second-, ... order neighbors which are the (up to) four nearest, four second nearest, ... sites to i.

4.3.3 Likelihood based on turpentine beetle colonization

The cumulative number of turpentine beetles at time t and site i is assumed to depend on local information such that

$$[x_{t,i}|(x_{t,j})_{j\neq i}, (w_s)_{s=t-1,t-2,\ldots}] \sim \left[x_{t,i}|x_{t-1,N_i^x}, z_{t-1,i}\right] \tag{4.5}$$

where x_{t-1,N_i^x} is the vector of variables x_j with $j \in N_i^x$. Here N_i^x consists of i and its neighbors up to the fifth order, and we assume that the conditional distribution of turpentine beetle colonization at time t depends only on turpentine beetle colonization at time $t-1$ and at sites in N_i^x, since this neighborhood is fairly large but is still interpretable biologically (see Section 4.5 for further details). Since turpentine beetles colonize red pines during only one brief period per year, and a tree can be colonized by multiple turpentine beetles, we assume conditional independence among nearby sites within the same year. On the other hand, turpentine beetles that colonize a tree in one spring tend to colonize nearby trees in the next spring. Thus we build into the model a possible relation between turpentine beetle colonization at time t and at time $t-1$.

The local characteristic $[x_{t,i}|x_{t-1,N_i^x}, z_{t-1,i}]$ is specified as follows. If the tree at site i was dead at time $t-1$ (i.e. $z_{t-1,i} = 1$), the local characteristic is deterministic with $x_{t,i} = x_{t-1,i}$, since turpentine beetles will not colonize a dead tree. Turpentine beetles could theoretically colonize a tree that dies from competitive thinning, i.e., a

process in which the growth of neighboring trees blocks out necessary sunlight. However, such events were rare in the stand, as the insects would likely colonize the weakened tree in advance of tree death. Moreover, the diameter and subcortical tissues of trees that have been crowded to death are frequently too thin to serve as a suitable breeding substrate for this insect. Turpentine beetles could also colonize a healthy tree that was killed suddenly, such as by a lightning strike or during a wind storm. However, we did not find any visual evidence of lightning (e.g., shredded bark, burn marks, or shattered limbs) or windthrow (other than trees that had already been killed) in any of our annual surveys. Hence, focusing on the colonization of live trees, if the tree at site i was alive at time $t-1$ (i.e. $z_{t-1,i}=0$), we assume that

$$\left[x_{t,i}-x_{t-1,i}|x_{t-1,N_i^x}, z_{t-1,i}=0\right] \sim \text{Poisson}(\lambda_{t,i})$$

where

$$\log(\lambda_{t,i}) = \theta_0 + \theta_1 \sum_{j\in N_i^x} x_{t-1,j}. \tag{4.6}$$

Thus, given the past, the $x_{t,i}-x_{t-1,i}$ with $z_{t-1,i}=0$ form a sample from a Poisson regression, so

$$\begin{aligned} L_t^{(1)}(\theta) &\propto \prod_{i:z_{t-1,i}=0} \lambda_{t,i}^{x_{t,i}-x_{t-1,i}} \exp(-\lambda_{t,i}) \\ &= \exp\left(\sum_{i:z_{t-1,i}=0}\left[(x_{t,i}-x_{t-1,i})\left(\theta_0+\theta_1\sum_{j\in N_i^x} x_{t-1,j}\right)\right.\right. \\ &\quad \left.\left. -\exp\left(\theta_0+\theta_1\sum_{j\in N_i^x} x_{t-1,j}\right)\right]\right). \end{aligned} \tag{4.7}$$

4.3.4 Likelihood based on *Ips* spp. colonization

The conditional dependence structure for whether colonization by *Ips* beetles has occurred is assumed to be

$$\begin{aligned} & [y_{t,i}|x_t, (y_{t,j})_{j\neq i}, (w_s)_{s=t-1,t-2,\ldots}] \\ \sim \; & [y_{t,i}|x_{t,i}, u_{t,N_i^y}, u_{t-1,N_i^y}, y_{t-1,i}, z_{t-1,i}]. \end{aligned} \tag{4.8}$$

Thus we assume that the conditional distribution of *Ips* spp. colonization at time t depends on turpentine beetle colonization at time t, *Ips* spp. colonization at sites $j \in N_i^y$ at both time $t-1$ and t, where N_i^y consists of the first and second order neighbors to i (note that N_i^y does not include i). Since it is hypothesized that turpentine beetles predispose red pines to colonization by *Ips* spp., we include in the model a possible relation to the number of turpentine beetle pitch tubes on the tree. Since *Ips* spp. attack different red pines 1–3 times per year and can overwinter near the trees they have colonized, we assume relations among neighboring sites for both time t and $t-1$ and that a first- and second-order neighborhood would suffice to capture spatial dependence in this study.

The local characteristic $[y_{t,i}|x_{t,i}, u_{t,N_i^y}, u_{t-1,N_i^y}, y_{t-1,i}, z_{t-1,i}]$ is specified as follows. If the tree at site i was dead at time $t-1$ (i.e. $z_{t-1,i} = 1$) or was colonized by *Ips* spp. at previous times (i.e., $y_{t-1,i} = 1$ or 2), the local characteristic is deterministic with $y_{t,i} = 0$ or 2, since *Ips* spp. will not colonize a dead tree. *Ips* spp. could theoretically colonize a tree that dies from competitive thinning, i.e., overshadowing and crowding by more dominant neighbors, although in practice the insects would likely find and colonize a weakened tree in advance of tree death, and would colonize only if the subcortical tissue was sufficiently thick. Such trees also contribute little to the ecological dynamics of the system, as they are commonly colonized by competing species of insects against which *Ips* spp. fare poorly. We also disregard the possibility that *Ips* spp. colonize lightning strikes or recent windthrow of live trees, due to the absence of such events observed during our annual surveys. Hence, focusing on colonization

of live trees, if the tree at site i was alive at time $t-1$ (i.e. $z_{t-1,i}=0$) and was not colonized previously ($y_{t,i}=0$), the local characteristic is assumed to be a logistic regression,

$$\left[y_{t,i}|x_{t,i},u_{t,N_i^y},u_{t-1,N_i^y},y_{t-1,i}=0,z_{t-1,i}=0\right]\sim \text{Bernoulli}(p_{t,i}) \quad (4.9)$$

where

$$\text{logit}(p_{t,i})=\psi_0+\psi_1 x_{t,i}+\psi_2\sum_{j\in N_i^y}u_{t-1,j}+\psi_3\sum_{j\in N_i^y}u_{t,j}. \quad (4.10)$$

Since $u_{t,i}=y_{t,i}$ in (4.9), by the Hammersley-Clifford theorem, $L_t^{(2)}(\psi)$ is equal to

$$\exp\left(\sum_{i:y_{t-1,i}=z_{t-1,i}=0}\left[\psi_0+\psi_1 x_{t,i}+\psi_2\sum_{j\in N_i^y}u_{t-1,j}\right]u_{t,i} + \psi_3\sum_{i<j:\,j\in N_i^y}u_{t,i}u_{t,j}\right)\Big/\, c(x_t,y_{t-1},z_{t-1},\psi) \quad (4.11)$$

where $c(x_t,y_{t-1},z_{t-1},\psi)$ is a normalizing constant (note that $j\in N_i^y \Leftrightarrow i\in N_j^y$). In other words, given the past, the $u_{t,i}$ with $y_{t,i}=z_{t,i}=0$ form an autologistic model (Besag, 1974).

4.3.5 Likelihood based on tree condition

The conditional dependence structure for tree condition is assumed to be

$$[z_{t,i}|x_t,y_t,(z_{t,j})_{j\neq i},(w_s)_{s=t-1,t-2,\ldots}]\sim[z_{t,i}|x_{t,i},u_{t,i},z_{t-1,N_i^z},z_{t-1,i}] \quad (4.12)$$

where the neighborhood N_i^z consists of the neighbors up to the fifth order. If the tree at site i was dead at time $t-1$ (i.e. $z_{t-1,i}=1$), the local characteristic is deterministic with $z_{t,i}=1$, because a dead

tree remains dead. But if the tree at site i was alive at time $t-1$ (i.e. $z_{t-1,i}=0$), the local characteristic is assumed to be a logistic regression,

$$\left[z_{t,i}|x_{t,i},u_{t,i},z_{t-1,N_i^z},z_{t-1,i}=0\right]\sim\text{Bernoulli}(q_{t,i})$$

where

$$\text{logit}(q_{t,i})=\varphi_0+\varphi_1x_{t,i}+\varphi_2u_{t,i}+\varphi_3\sum_{j\in N_i^z}z_{t-1,j}. \tag{4.13}$$

That is, mortality rate of a tree depends on both turpentine beetle colonization and *Ips* spp. colonization. The additional term involving the tree condition at time $t-1$ is a way of accounting for any potential spatial dependence. Again we consider a fairly large neighborhood that consists of neighbors up to the fifth order. Conditional on the past, the $z_{t,i}$ with $z_{t-1,i}=0$ form a sample from a logistic regressions, so

$$\begin{aligned}L_t^{(3)}(\varphi)&=\prod_{i:z_{t-1,i}=0}\frac{\exp(z_{t,i}\text{logit}(q_{t,i}))}{1+\exp(\text{logit}(q_{t,i}))}\\&=\prod_{i:z_{t-1,i}=0}\frac{\exp(z_{t,i}(\varphi_0+\varphi_1x_{t,i}+\varphi_2u_{t,i}+\varphi_3\sum_{j\in N_i^z}z_{t-1,j}))}{1+\exp(\varphi_0+\varphi_1x_{t,i}+\varphi_2u_{t,i}+\varphi_3\sum_{j\in N_i^z}z_{t-1,j})}.\end{aligned} \tag{4.14}$$

4.4 Bayesian model and posterior simulations

We assume independent improper uniform priors

$$p(\theta)\propto 1,\quad \theta\in\mathbb{R}^2;\quad p(\psi)\propto 1,\quad \psi\in\mathbb{R}^4;\quad p(\varphi)\propto 1,\quad \varphi\in\mathbb{R}^4.$$

Thus θ,ψ,φ are *a posteriori* independent with densities

$$\begin{aligned}&\pi(\theta)\propto L^{(1)}(\theta),\quad \theta\in\mathbb{R}^2;\quad \pi(\psi)\propto L^{(2)}(\psi),\quad \psi\in\mathbb{R}^4;\\&\pi(\varphi)\propto L^{(3)}(\varphi),\quad \varphi\in\mathbb{R}^4.\end{aligned} \tag{4.15}$$

For a discussion of posterior properity, see Appendix A. For the remaining discussion of MCMC simulations, we assume the reader is familiar with MCMC methods (e.g. Robert and Casella (2004)).

For turpentine beetles, we will simulate from the marginal posterior distribution of θ using a Metropolis within Gibbs algorithm, where we alternate between updating θ_0 and θ_1. Since the full conditional for $\lambda_0 = \exp(\theta_0)$ is a Gamma distribution with shape parameter $\sum_{t,i}(x_{t,i} - x_{t-1,i})$ and inverse scale parameter $\sum_{t,i} \exp(\theta_1 \sum_{j \in N_i^x} x_{t-1,j})$, where in both cases the sum $\sum_{t,i}$ is over those t, i with $z_{t-1,i} = y_{t-1,i} = 0$, we use a Gibbs update for this component. The full conditional for the other parameter θ_1 is not a standard distribution, so here we use a Metropolis random walk algorithm with a normal proposal distribution, cf. Robert and Casella (2004).

For *Ips* spp., suppose we use a Metropolis-Hastings algorithm to simulate from the marginal posterior distribution of ψ. Let $L^{(2)}_{\text{unnorm}}(\psi; u)$ denote $L^{(2)}$ in (4.4) but without the unknown normalizing constant

$$c(\psi) = \prod_{t=1}^{5} c(x_t, y_{t-1}, z_{t-1}, \psi)$$

from (4.11); here u denotes the vector of all observed $u_{t,i}$ values. If ψ is the current and ψ' is the proposed parameter values in the Metropolis-Hastings algorithm, then the Hastings ratio depends on the ratio $c(\psi')/c(\psi)$ of unknown normalizing constants. This can be approximated by path sampling (e.g. Gelman and Meng (1998)),

$$\log \frac{c(\psi')}{c(\psi)} \approx \frac{1}{n} \sum_{k=1}^{n} \left[\frac{\mathrm{d}}{\mathrm{d}s} \log L^{(2)}_{\text{unnorm}}(\psi(s_k); \upsilon_k) \right]. \tag{4.16}$$

Here we let $s_1, \ldots, s_n$ be independent and uniformly distributed on $[0, 1]$, and $\psi(s) = s\psi' + (1 - s)\psi, 0 \leq s \leq 1$ is a line segment. Further, each υ_k is a perfect simulation of $u = (u_1, \ldots, u_5)$ where u_t given the past follows the autologistic model (4.11) with parameter $\psi(s_k)$ (Propp and Wilson, 1996; Møller, 1999). Furthermore, given $s_1, \ldots, s_n$, the perfect simulations $\upsilon_1, \ldots, \upsilon_n$ are independent.

We use a Metropolis random walk algorithm with independent normal proposal distributions for $\psi_0, \psi_1, \psi_2, \psi_3$, where we propose to change all four parameters at the same time, since the main part of the running time of the algorithm is by far used in generating the perfect simulations, and this is the same amount of work whether we are changing one or all four parameters.

In the case of φ, we use a Metropolis within Gibbs algorithm, where we alternate between simulating from the marginal posterior distribution of $\varphi_0, \varphi_1, \varphi_2, \varphi_3$, respectively. Neither of these parameters have standard distributions, so for each parameter we use a Metropolis random walk update with a normally distributed proposal.

When running the Metropolis random walk algorithm for either $\theta_1, \psi, \varphi_0, \varphi_1, \varphi_2$, or φ_3, the standard deviation of the normal proposal distribution is chosen to reach an average acceptance probability about 0.3 (Roberts *et al.*, 1997).

4.5 Statistical inference and discussion of the ecological questions

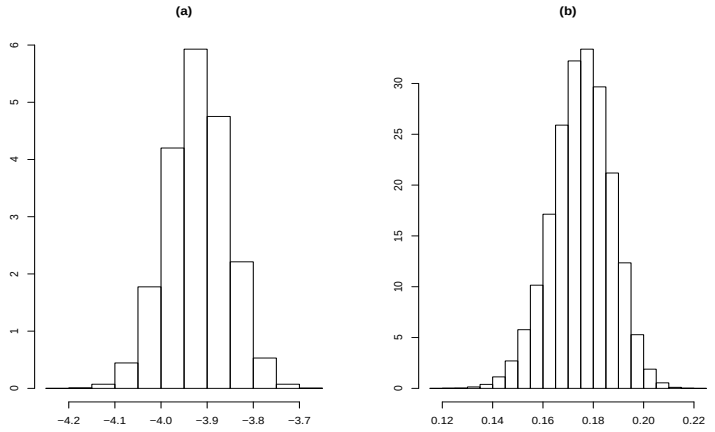
4.5.1 Posterior distributions of the model parameters

For inference of the parameters θ in the turpentine beetle colonization model, Figure 4.2 gives the posterior distributions based on an MCMC run length of 100,000 with a burn-in length of 1,000. The results suggest that there is a positive relation between the new turpentine beetle colonization and the number of turpentine beetle tubes in the previous year, at not only the same site, but also the sites that are up to the fifth-order neighbors. That is, the more turpentine beetles there were in the previous year on a tree and its neighboring trees, the more new colonization can be expected to occur on this tree in the current year. Here the extent of local temporal dependence

is captured by a 1-year lag and that of local spatial dependence by about 5.14 m, which is the distance between a fifth-order neighbor and the site of a tree. We have also fitted a model that has one term for the zero-, first-, second-order neighbors and another term for the third- to fifth-order neighbors. The results there (not shown) suggest that the regression coefficients for the two types of neighborhoods are similar and thus we combine all the neighbors up to the fifth order. This phenomenon is consistent with a hypothesis in which turpentine beetles colonize trees that are being slowly weakened by the spread of a root fungus, such as *L. terebrantis* or *L. procerum*. These fungi are introduced to trees by the beetles and spread via root grafts at a rate of 5m per year, according to our best estimates based on root excavations and fungal isolations (Klepzig *et al.*, 1991; Erbilgin and Raffa, 2002). This hypothesis is consistent with the work of Erbilgin and Raffa (2003), who found that the probability of tree death falls below 50% at a distance of 5 m from the outer edge of the pocket margin.

For inference of the parameters ψ in the *Ips* spp. colonization model, Figure 4.3 gives the posterior distributions based on an MCMC run length of 60,000 with a burn-in length of 1,000. For the approximation (4.16), we use only $n = 10$ perfect simulations, which seems to give a reasonable approximation of the normalizing constant ratio. The results suggest that there is a positive relation between the *Ips* spp. colonization in the current year and the number of turpentine beetle tubes in the same year at the same site, *Ips* spp. colonization in the previous year at the neighboring sites (excluding the same site), and *Ips* spp. colonization in the current year at the neighboring sites (excluding the same site), up to the second-order neighbors. In other words, the more turpentine beetles there are on a tree, the more likely that the tree will be colonized by *Ips* spp. Thus there is strong evidence that turpentine beetles pre-dispose trees to colonization by *Ips* spp. Moreover, there is clear spatial and temporal dependence in the *Ips* spp. colonization. The more trees in the neighborhood that were colonized by *Ips* spp. in the previous year, the more likely that

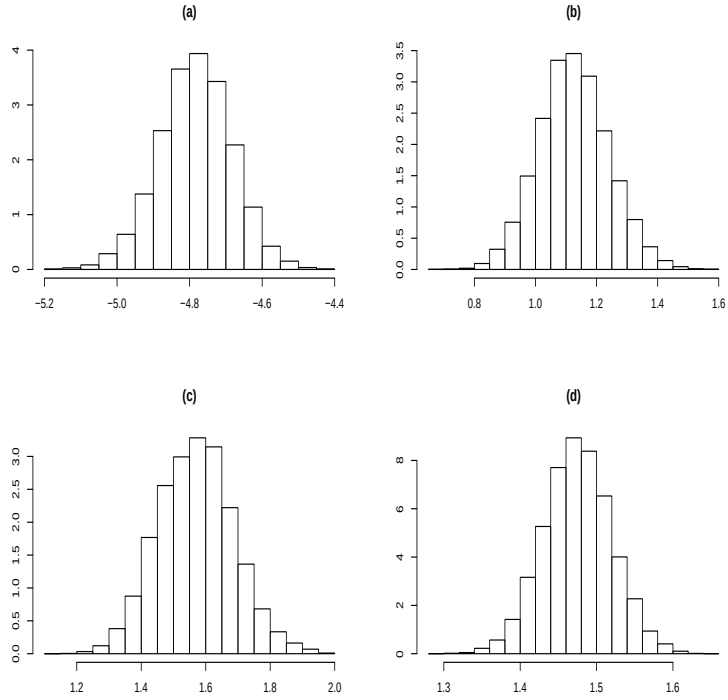
Figure 4.2: Posterior distribution of (a) θ_0; (b) θ_1 in the turpentine beetle colonization model.



the tree is colonized by *Ips* spp. in the current year. Similarly the more trees in the neighborhood that are colonized by *Ips* spp. in the current year, the more likely that the tree is colonized by *Ips* spp. in the current year. Here the extent of local temporal dependence is captured by a 1-year lag and that of local spatial dependence by about 2.07 m, which is the distance between a second-order neighbor and the site of the tree. Concentration of *Ips* spp. attacks among nearby trees may occur for three reasons, none of which are mutually exclusive. First, insect brood emerging from a previously colonized tree may preferentially colonize nearby trees. This may occur, for example, if brood adults from late fall overwinter in the duff around the base of their brood tree, and then emerges to colonize nearby trees in the spring. Although little is known about relations between brood tree and overwintering locations, inclement weather and predators exert substantial mortality on bark beetles engaging in host seeking behaviors (Berryman, 1979). Second, localized attacks may

occur when high numbers of bark beetles are attracted by aggregation pheromones of a successful attack and begin to attack nearby trees, a phenomena known as "switching" (Geiszler *et al.*, 1980). Third, turpentine beetles, and/or fungal root pathogens, may weaken trees in local neighborhoods and make them more susceptible to attacks and colonizations by *Ips* spp. (Owen, 1985).

Figure 4.3: Posterior distribution of (a) ψ_0; (b) ψ_1; (c) ψ_2; (d) ψ_3 in the *Ips* spp. colonization model.



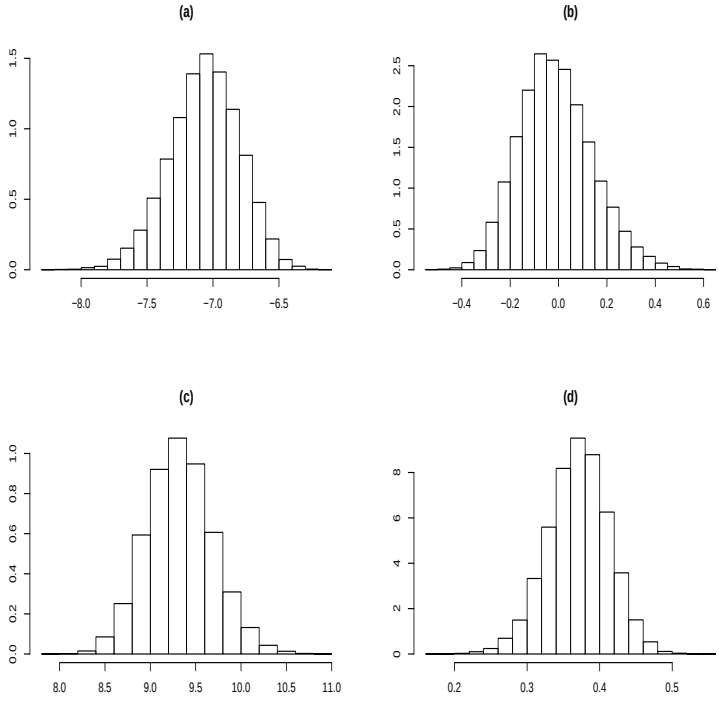
For inference of the parameters φ in the tree condition model, Figure 4.4 gives the posterior distributions based on an MCMC run

length of 100,000 with a burn-in length of 1,000. The results suggest that there is no evidence of a direct relation between a tree's condition and the number of its turpentine beetle tubes, but there is a strong positive relation between *Ips* spp. colonization and subsequent tree death. That is, the number of turpentine beetles does not directly influence the mortality of tree, but there is a very large increase in the probability that a tree dies after colonized by *Ips* spp. in the same year. This is not surprising, as trees may survive colonization of the root collar by turpentine beetles for more than one year. However, *Ips* spp. utilize aggregation pheromones to attract high numbers of conspecifics that quickly colonize all available subcortical tissue. The water-conducting tissues are mined by the developing larvae, and the tree dies soon thereafter. Furthermore it appears necessary to account for the spatial-temporal dependence among the tree conditions.

4.5.2 Empirical and predictive rates of mortality and *Ips* spp. colonization

In this and the next subsection we check important aspects of the model that correspond to the ecological questions of interest, particularly the relations among turpentine beetle colonization, *Ips* spp. colonization, and tree conditions (see Section 4.1). The model checking is based on posterior predictive distributions obtained by a Monte Carlo sample $(x^{(s)}, u^{(s)}, z^{(s)})$, $s = 1, \ldots, S$ where the Monte Carlo sample size is chosen to be $S = 100$. More precisely, since inference is performed conditional on e, given a posterior simulation $(\theta^{(s)}, \psi^{(s)}, \varphi^{(s)})$, we simulate "new data" $(x^{(s)}, u^{(s)}, z^{(s)})$ from the conditional distribution of d given e as specified in Section 4.3. This is done using the sequential model construction in Section 4.3, where simulation of x_t and z_t given their relevant past is straightforward (see Sections 4.3.3 and 4.3.5), while we use perfect simulation for y_t given the relevant past (see Section 4.3.4). Note that $x_0^{(s)} = x_0$, $y_0^{(s)} = y_0$, and $z_{-1}^{(s)} = z_{-1}$. The samples $(\theta^{(s)}, \psi^{(s)}, \varphi^{(s)})$, $s = 1, \ldots, S$ are chosen such that they

Figure 4.4: Posterior distribution of (a) φ_0; (b) φ_1; (c) φ_2; (d) φ_3 in the tree condition model.



are effectively independent posterior simulations. Moreover, we let $(x^{(0)}, y^{(0)}, z^{(0)})$ denote the data.

In this section, we consider the posterior predictive distribution of various statistics related to mortality rates of trees and rates of *Ips* spp. colonization. First, define

$$\begin{aligned}
\mathcal{I}_{0,0} &= \{i : z_{-1,i} = 0, x_{0,i} = 0, u_{0,i} = 0\},\\
\mathcal{I}_{0,1} &= \{i : z_{-1,i} = 0, x_{0,i} = 0, u_{0,i} = 1\},\\
\mathcal{I}_{1,0} &= \{i : z_{-1,i} = 0, x_{0,i} > 0, u_{0,i} = 0\},\\
\mathcal{I}_{1,1} &= \{i : z_{-1,i} = 0, x_{0,i} > 0, u_{0,i} = 1\},
\end{aligned}$$

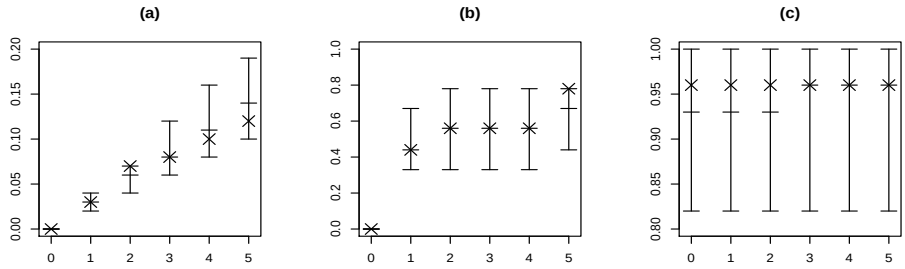
and

$$p_{k,l}^{(s)}(t) = \frac{1}{|\mathcal{I}_{k,l}|} \sum_{i \in \mathcal{I}_{k,l}} \mathbf{1}[z_{t,i}^{(s)} = 1], \quad s = 0, \ldots, S, \quad t = 0, \ldots, 5, \quad k, l = 0, 1$$

where $|A|$ denotes the cardinality of a finite set A. Then $p_{0,0}^{(0)}(t)$ is the observed tree mortality rate of trees, which were alive at time -1 and had no bark beetle colonization by time 0; $p_{0,1}^{(0)}(t)$ is the observed mortality rate of trees that were colonized by *Ips* spp.; $p_{1,0}^{(0)}(t)$ is the observed mortality rate of trees that were colonized by turpentine beetles; and $p_{1,1}^{(0)}(t)$ is the observed mortality rate of trees that were colonized by both turpentine beetles and *Ips* spp. Figure 4.5 shows for each value of $(k, l) = (0, 0), (0, 1), (1, 0)$ and $t = 0, \ldots, 5$ the observed mortality rate $p_{k,l}^{(0)}(t)$ and the $2.5^{th}, 50^{th}, 97.5^{th}$ percentiles of the posterior predictive distribution obtained from $p_{k,l}^{(s)}(t)$, $s = 1, \ldots, S$. Further, for the case $(k, l) = (1, 1)$ (not shown in Figure 4.5), the $2.5^{th}, 50^{th}, 97.5^{th}$ percentiles for the mortality rates are for all times $t = 0, \ldots, 5$ given by 0.50, 1.00, and 1.00, respectively, and the corresponding observed values are all 1.00. For all values of (k, l), the observed rates lie in the centers of the corresponding predictive distributions. Thus overall there is no evidence against our model. Compared to $p_{0,0}^{(0)}(t)$, which may be interpreted as a kind of observed

baseline mortality rate, $p_{0,1}^{(0)}(t)$ increased greatly and the large increase occurred within the same year of *Ips* spp. colonization; $p_{1,0}^{(0)}(t)$ increased at time 1 and the increase leveled off at time 2; and $p_{1,1}^{(0)}(t)$ is nearly 100% within the same year of the colonization. The predictive distributions show a similar behavior. The fact that deaths of trees occur in both the first and the second year after turpentine beetle colonization gives further evidence that turpentine beetles predispose a tree to death rather than killing a tree directly. The result here also supports the theory that *Ips* spp., unlike turpentine beetles, are the ultimate mortality agents of red pines.

Figure 4.5: Central 95% prediction intervals and medians (indicated by bars) for the tree mortality rates over time $t = 0, \ldots, 5$ among those trees that were alive at $t = -1$ and, (a) were not colonized ($x_{i,0} = u_{i,0} = 0$), (b) were colonized by turpentine beetles ($x_{i,0} > 0, u_{i,0} = 0$), and (c) were colonized by *Ips* spp. ($x_{i,0} = 0, u_{i,0} = 1$) at $t = 0$. The corresponding observed tree mortality rates are indicated by crosses. Note the different scales on the y-axes.



Next, let

$$\mathcal{I}_k = \{i : z_{-1,i} = 0, x_{0,i} = k\}, \quad k = 0, 1,$$

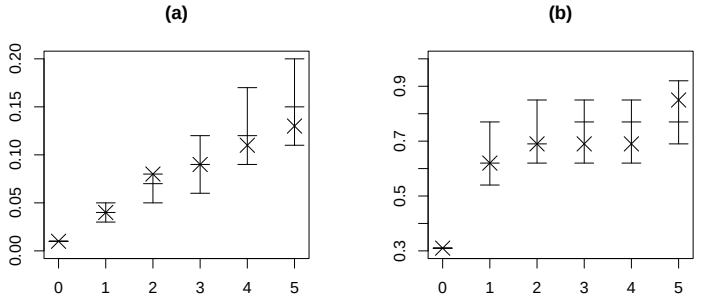
denote the collection of sites where a tree was alive at time -1 and was ($k = 1$) or was not ($k = 0$) colonized by turpentine beetles by

time 0, and let

$$p_k^{(s)}(t) = \frac{1}{|\mathcal{I}_k|} \sum_{i \in \mathcal{I}_k} \mathbf{1}[u_{t,i}^{(s)} = 1], \quad s = 0, \ldots, S, \quad t = 0, \ldots, 5, \quad k = 0, 1.$$

Then $p_k^{(0)}(t)$ is the observed rate of *Ips* spp. colonization of a tree from $\mathcal{I}_k$ by time $t = 0, \ldots, 5$. Figure 4.6 is similar to Figure 4.5 but concerns $p_k^{(s)}(t)$ for $k = 0, 1$ and $t = 0, \ldots, 5$ Again there is no evidence against our model. Compared to $p_0^{(0)}(t)$, the rates of *Ips* spp. colonization $p_1^{(0)}(t)$ are much higher and leveled off after 2–3 years, which support the theory that turpentine beetles pre-dispose the trees to subsequent colonization and thus kill by *Ips* spp.

Figure 4.6: Central 95% prediction intervals and medians (indicated by bars) for the rate of *Ips* spp. colonization over time $t = 0, \ldots, 5$ among those trees that were alive at $t = -1$ and (a) were not colonized by turpentine beetles ($x_{i,0} = 0$) or (b) were colonized by turpentine beetles ($x_{i,0} > 0$). The corresponding observed rate of *Ips* spp. colonization are indicated by crosses. Note the different scales on the y-axes.



4.5.3 Checking further aspects of the model

To check whether the model captures the relation between turpentine beetle colonization and *Ips* spp. colonization and between colonization of *Ips* spp. and tree mortality, we consider

$$r_{x,u}^{(s)} = \sum_{t=1}^{5} \sum_{i} (x_{t,i}^{(s)} - x_{t-1,i}^{(s)}) u_{t,i}^{(s)}, \quad r_{u,z}^{(s)} = \sum_{t=1}^{5} \sum_{i} u_{t,i}^{(s)} z_{t,i}^{(s)}, \quad s = 0, \ldots, S.$$

Here $r_{x,u}^{(0)}$ summarizes the observed relation between new colonization of turpentine beetles and new colonization of *Ips* spp. in the same year and at the same site, and $r_{u,z}^{(0)}$ summarizes the observed occurrences of *Ips* spp. colonization that is involved in mortality of trees. Furthermore, for spatial dependence structure, we consider

$$\begin{aligned}
v_{x,x}^{(s)}(\delta) &= \sum_{i,j:d(i,j)\in N(\delta)} \mathbf{1}[x_{5,i}^{(s)} > 0, x_{5,j}^{(s)} > 0], \quad s = 0, \ldots, S, \quad \delta > 0, \\
v_{y,y}^{(s)}(\delta) &= \sum_{i,j:d(i,j)\in N(\delta)} \mathbf{1}[y_{5,i}^{(s)} > 0, y_{5,j}^{(s)} > 0], \quad s = 0, \ldots, S, \quad \delta > 0, \\
v_{z,z}^{(s)}(\delta) &= \sum_{i,j:d(i,j)\in N(\delta)} \mathbf{1}[z_{5,i}^{(s)} = 1, z_{5,j}^{(s)} = 1], \quad s = 0, \ldots, S, \quad \delta > 0,
\end{aligned}$$

where $d(i, j)$ denotes the Euclidean distance between sites i and j, and $N(\delta) = (\delta - 1, \delta]$ is a half-open interval. That is, $v_{x,x}^{(0)}(\delta)$ $(v_{y,y}^{(0)}(\delta), v_{z,z}^{(0)}(\delta))$ captures the observed spatial relation between turpentine beetle colonization (*Ips* spp. colonization, tree mortality) at two sites that are at least $\delta - 1$ and at most δ apart in distance. Here we focus on cumulative effect of all three variables for simplicity.

Finally, for temporal dependence structure, we consider

$$\begin{aligned} w_x^{(s)}(t) &= \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[x_{t,i}^{(s)} = 0], \\ w_y^{(s)}(t) &= \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[y_{t,i}^{(s)} = 0], \\ w_z^{(s)}(t) &= \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[z_{t,i}^{(s)} = 0], \end{aligned}$$

where $s = 0, \ldots, S$, $t = 0, \ldots, 5$ for $w_x(t)$ and $w_y(t)$, while $t = -1, \ldots, 5$ for $w_z(t)$, and $N = 2715$. That is, $w_x^{(0)}(t)$ ($w_y^{(0)}(t)$, $w_z^{(0)}(t)$) is the observed proportion of trees that are not colonized by turpentine beetles (that are not colonized by *Ips* spp., that are alive) by time t.

Figures 4.7 and 4.8 are similar to Figure 4.5 but concern the statistics above except $r_{x,u}^{(s)}$ and $r_{u,z}^{(s)}$, where the 2.5%, 50%, 97.5% percentiles are 14.0, 28.0, 149.0 for $r_{x,u}^{(s)}$, and 225.0, 314.5, 409.0 for $r_{u,z}^{(s)}$. Thus the observed values $r_{x,u}^{(0)} = 58$ and $r_{u,z}^{(0)} = 269$ fall well within the central 95% prediction intervals. Our model also adequately captures the spatial dependence for turpentine beetle colonization at all lag distances (see $v_{x,x}^{(s)}(\delta)$ in Figure 4.7). For *Ips* spp. colonization and tree condition (see $v_{y,y}^{(s)}(\delta)$ and $v_{z,z}^{(s)}(\delta)$ in Figure 4.7), the spatial dependence is well captured by the model when the lag distances are small. The observed values tend to be larger than what the model predicts, which may be a result of the large cluster of trees that were colonized by *Ips* spp. and/or were dead in the southeastern part of the plantation. Our model furthermore adequately captures the temporal dependence for *Ips* spp. colonization and tree condition at all time points (see $w_y^{(s)}(t)$ and $w_z^{(s)}(t)$ in Figure 4.8). But for turpentine beetle colonization (see $w_x^{(s)}(t)$ in Figure 4.8), the observed values tend to be slightly larger than what the model predicts.

Figure 4.7: Central 95% prediction intervals and medians (indicated by bars) for (a) $v^{(s)}_{x,x}(\delta)$, (b) $v^{(s)}_{y,y}(\delta)$, and (c) $v^{(s)}_{z,z}(\delta)$. The corresponding observed values are indicated by crosses. Note the different scales on the y-axes.

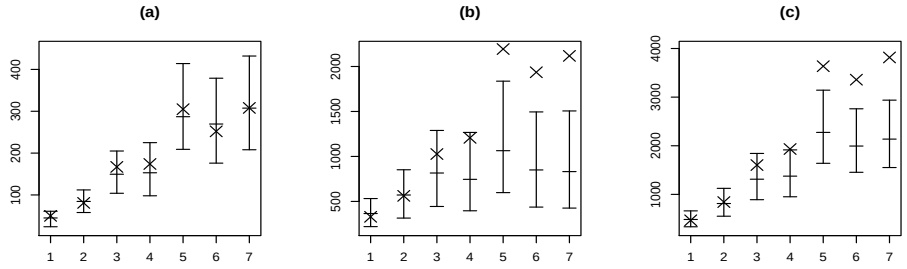
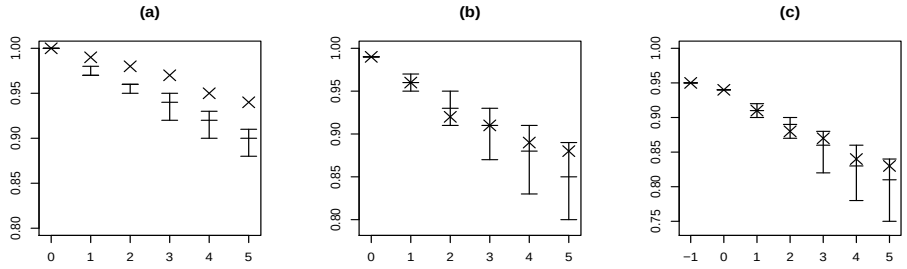


Figure 4.8: Central 95% prediction intervals and medians (indicated by bars) for (a) $w^{(s)}_{x}(t)$, (b) $w^{(s)}_{y}(t)$, and (c) $w^{(s)}_{z}(t)$. The corresponding observed values are indicated by crosses. Note the different scales on the y-axes.



4.6 Concluding remarks

In this article, we have examined the effect of two bark beetle groups on the mortality of red pine trees in a Wisconsin plantation. We have constructed spatial-temporal statistical models to quantify the relations among turpentine beetle colonization, *Ips* spp. colonization, and mortality of red pine trees, while accounting for correlation across space and over time. For statistical inference, we have adopted a Bayesian hierarchical model and devised MCMC algorithms for obtaining the posterior distributions of model parameters. Based on the results in Sections 4.5.2–4.5.3, our impression is that the spatial-temporal model in Section 4.3 has adequately captured the relations among the three variables, turpentine beetle colonization, *Ips* spp. colonization, and tree condition. Moreover, our model has often though not always captured adequately the spatial and temporal structure. The data analysis in Section 4.5 suggests that turpentine beetle colonization is associated with higher likelihood of *Ips* spp. colonization and *Ips* spp. colonization is associated with higher likelihood of red pine tree mortality, whereas there is no direct association between turpentine beetle colonization and red pine tree mortality. There is strong evidence that turpentine beetle colonization does not kill a red pine tree directly, but rather predisposes the tree to subsequent colonization by *Ips* spp. The evidence is also strong that *Ips* spp. are the ultimate mortality agents of red pine trees. The modeling approach here is of general utility to systems in which interactions among several species affect overall dynamics, but likewise generate spatial-temporal patterns that can complicate dissection of underlying processes. Such systems are quite likely common in forest ecosystems. Employment of this approach can help managers predict insect and pathogen dynamics as well as direct preventative and remedial measures against inciting rather than merely ultimate agents affecting forest health.

Appendix A

From a practical viewpoint, we would expect our MCMC runs to diverge if an improper posterior distribution had been specified. From a theoretical viewpoint, since the three likelihood functions in (4.15) are log concave, properity of the posteriors with uniform improper priors is equivalent to the existence of the maximum likelihood estimate (MLE) based on $L^{(1)}(\theta)$, $L^{(2)}(\psi)$, and $L^{(3)}(\varphi)$, respectively. This can be established as sketched below.

The likelihood functions $L^{(1)}(\theta)$, $L^{(2)}(\psi)$, $L^{(3)}(\varphi)$ in (4.7), (4.11), (4.14) are products of log concave functions $L_t^{(1)}(\theta)$, $L_t^{(2)}(\psi)$, $L_t^{(3)}(\varphi)$, respectively. Therefore, to verify the existence of the MLE based on $L^{(1)}(\theta)$, $L^{(2)}(\psi)$, $L^{(3)}(\varphi)$, it suffices for each $t = 1, \ldots, 5$ to verify the existence of the MLE based on $L_t^{(1)}(\theta)$, $L_t^{(2)}(\psi)$, $L_t^{(3)}(\varphi)$, respectively. This can easily be checked in the cases of the Poisson regression $L_t^{(1)}(\theta)$ based on the data x_t and the logistic regression $L^{(3)}(\varphi)$ based on the data z_t, either by theoretical results (Barndorff-Nielsen, 1978; Jacobsen, 1989) or using software for generalized linear models. Moreover, by (4.11), $L_t^{(2)}(\psi)$ is of regular exponential family form with canonical statistic

$$s_t^{(2)}(u_t) = \sum_{i: y_{t-1,i} = z_{t-1,i} = 0} \left(u_{t,i},\ x_{t,i} u_{t,i},\ \sum_{j \in N_i^y} u_{t-1,j} u_{t,i},\ \sum_{j:\, j \in N_i^y} u_{t,i} u_{t,j} \right).$$

Consequently, by a well-known result from exponential family theory (Barndorff-Nielsen, 1978), the MLE of ψ based on the data u_t exists if $s_t^{(2)}(u_t)$ belongs to the interior of the convex hull of its support. This condition seems less straightforward to check, so alternatively, MCMC methods for finding the MLE may be applied (Geyer and Thompson, 1992).

Acknowledgment

Funding has been provided for this research from NSF: DEB-0314215 and the Wisconsin Alumni Research Foundation. We thank the Wisconsin Department of Natural Resources for providing the research site.

References

Auclair, A. N. D. (2005). Patterns and general characteristics of severe forest dieback from 1950 to 1995 in the northeastern United States. *Canadian Journal of Forest Research*, **35**, 1342–1355.

Aukema, B. H., Clayton, M. K., and Raffa, K. F. (2005). Modeling flight activity and population dynamics of the pine engraver, *Ips pini*, in the Great Lakes region: Effects of weather and predators at short time scales. *Population Ecology*, **47**, 61–69.

Barndorff-Nielsen, O. E. (1978). *Information and Exponential Families in Statistical Theory*. Wiley, Chichester.

Battles, J. J. and Fahey, T. J. (2000). Gap dynamics following forest decline: a case study of red spruce forests. *Ecological Applications*, **10**, 760–774.

Berryman, A. A. (1979). Dynamics of bark beetle populations: Analysis of dispersal and redistribution. *Entomologique Suisse*, **52**, 227–234.

Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society Series B*, **36**, 192–236.

Drohan, P. J., Stout, S. L., and Petersen, G. W. (2002). A study of sugar maple (*Acer saccharum* Marsh) decline during 1979–1989

in northern Pennsylvania. *Forest Ecology and Management*, **170**, 1–17.

Erbilgin, N. and Raffa, K. F. (2002). Association of declining red pine stands with reduced populations of bark beetle predators, seasonal increases in root colonizing insects, and incidence of root pathogens. *Forest Ecology and Management*, **164**, 221–236.

Erbilgin, N. and Raffa, K. F. (2003). Spatial analysis of forest gaps resulting from bark beetle colonization of red pines experiencing belowground herbivory and infection. *Forest Ecology and Management*, **177**, 145–153.

Erbilgin, N., Nordheim, E. V., Aukema, B. H., and Raffa, K. F. (2002). Population dynamics of *Ips pini* and *Ips grandicollis* in red pine plantations in Wisconsin: within- and between-year associations with predators, competitors, and habitat quality. *Environmental Entomology*, **31**, 1043–1051.

Geiszler, D. R., Gallucci, V. F., and Gara, R. I. (1980). Modeling the dynamics of mountain pine beetle aggregation in a lodgepole pine stand. *Oecologia*, **46**, 244–253.

Gelman, A. and Meng, X. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science*, **13**, 163–185.

Geyer, C. J. and Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Society of Statistics Series B*, **54**, 657–699.

Jacobsen, M. (1989). Existence and unicity of MLE in discrete exponential family distributions. *Scandinavian Journal of Statistics*, **16**, 335–349.

Klepzig, K., Smalley, E. B., and Raffa, K. F. (1996). Combined chemical defenses against an insect-fungal complex. *Journal of Chemical Ecology*, **22**, 1367–1388.

Klepzig, K. D., Raffa, K. F., and Smalley, E. B. (1991). Association of insect-fungal complexes with red pine decline in Wisconsin. *Forest Science*, **41**, 1119–1139.

Klepzig, K. D., Kruger, E. L., Smalley, E. B., and Raffa, K. F. (1995). Effects of biotic and abiotic stress on induced accumulation of terpenes and phenolics in red pines inoculated with bark beetle-vectored fungus. *Journal of Chemical Ecology*, **21**, 601–626.

Møller, J. (1999). Perfect simulation of conditionally specified models. *Journal of the Royal Statistical Society Series B*, **61**, 251–264.

Owen, D. R. (1985). *The Role of Dendroctonus valens and Its Vectored Fungi in the Mortality of Ponderosa Pine.* Ph.D. dissertation, University of California-Berkeley, Berkeley, California.

Propp, J. G. and Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, **9**, 223–252.

Purdon, M., Cienciala, E., Metelka, V., Beranova, J., Hunova, I., and Cerny, M. (2004). Regional variation in forest health under long-term air pollution mitigated by lithological conditions. *Forest Ecology and Management*, **195**, 355–371.

Raffa, K. and Smalley, E. (1995). Interaction of pre-attack and induced monoterpene concentrations in host conifer defense against bark beetle-fungal complexes. *Oecologia*, **102**, 285–295.

Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer, New York, 2nd edition.

Roberts, G. O., Gelman, A., and Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Annals of Applied Probability*, **7**, 110–120.

Chapter 5

Bayesian inference for multivariate point processes observed at sparsely distributed times

Jakob G. Rasmussen & Jesper Møller
Department of Mathematical Sciences, Aalborg University.
Email addresses: jgr@math.aau.dk & jm@math.aau.dk

Brian H. Aukema
Canadian Forest Service and University of Northern British Columbia
Email addresses: baukema@nrcan.gc.ca

Kenneth F. Raffa
Department of Entomology, University of Wisconsin-Madison.
Email addresses: raffa@entomology.wisc.edu

Jun Zhu
Department of Statistics, University of Wisconsin-Madison.
Email addresses: jzhu@stat.wisc.edu

Abstract

We consider statistical and computational aspects of simulation-based Bayesian inference for a multivariate point process which is only observed at sparsely distributed times. For specificity we consider a particular data set which has earlier been analyzed by a discrete time model involving unknown normalizing constants. We discuss the advantages and disadvantages of using continuous time processes compared to discrete time processes in the setting of the present paper as well as other spatial-temporal situations.

Keywords: Bark beetle, conditional intensity, forest entomology, Markov chain Monte Carlo, missing data, prediction, spatial-temporal process.

5.1 Introduction

This paper concerns statistical inference of spatial-temporal processes that are observed on a spatial lattice but only at sparsely distributed time points. Zhu *et al.* (2006) analyzed such a data set using a spatial-temporal autoregressive type of model, which assumes that time is discrete and coincides with the observation times. Here we propose an alternative continuous time model, based on multivariate point processes (Daley and Vere-Jones, 2003), and develop Bayesian inference for estimating the model parameters and the times of events. The proposed methodology is illustrated by a subset of the data set featured in Zhu *et al.* (2006), but we do not attempt to address all the scientific questions in that paper. Instead the focus is on the methodology and our conclusion is that a multivariate point process model may have several advantages compared to a spatial-temporal autoregressive type of model.

The data set in Zhu *et al.* (2006) is from a study of a plantation of red pines located near Spring Green, Wisconsin, USA. Each tree was examined annually, and three types of data were recorded.

The tree's condition (alive or dead) was recorded from 1986 to 1992. The number of *Dendroctonus valens* (LeConte), a bark beetle hereafter called "turpentine beetle" that attacks the base of the tree, was recorded from 1987 to 1992. The presence or absence of *Ips* spp. (predominantly *Ips pini* (Say) and to a lesser extent *Ips grandicollis* (Eichhoff)), another bark beetle that mass attacks the main stem, was also recorded from 1987 to 1992. The turpentine beetle has one generation per year, with new attacks occurring from late April through June, and each beetle attacking only one tree (Furniss and Carolin, 1980). *Ips* spp. have two to three generations per year, depending on temperature. They become active in early May and dormant in September. Again, each beetle attacks only one tree. The primary objectives in Zhu *et al.* (2006) were quantifying the relations between the two types of beetles and the condition of the trees, as well as capturing the spatial and temporal structure of colonization by the two beetle types and the condition of the trees. Of special interest was the fact that a large gap of dead trees appeared.

In the present paper, we model *Ips* spp., using the turpentine beetles as a covariate. Furthermore, we only consider the area immediately around the gap, where Figure 5.1 shows the *Ips* spp. data. Focusing on a subset of the trees leads of course to loss of important biological information, but since we want to illustrate the methodological aspects, we do not aim at an overly complex model. We specify a multivariate point process model for *Ips* spp., using a Bayesian setting where we regard what happens between the observation times as missing data. For short we refer to the multivariate point process model as a continuous time model and the spatial-temporal autoregressive type of model in Zhu *et al.* (2006) as a discrete time model.

The paper is organized as follows. Section 5.2 introduces the needed notation. Section 5.3 specifies the multivariate point process model and prior assumptions, and Section 5.4 discusses simulation based Bayesian inference. Section 5.5 concludes with a comparison between continuous time and discrete time models for the specific models considered in Zhu *et al.* (2006) and the present paper as well

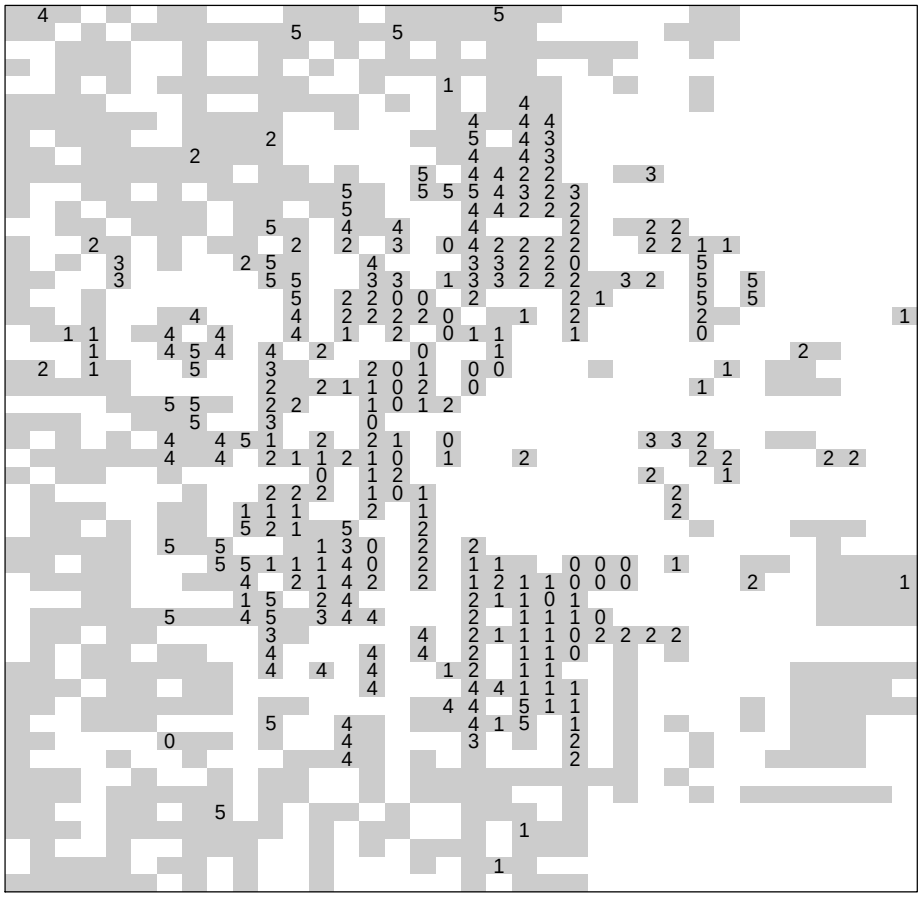


Figure 5.1: White squares are locations without trees or trees that were already dead in 1986, and gray squares are trees that were alive in 1986. The numbers indicate which year a tree has been attacked by *Ips* spp.; trees without numbers have not been attacked during the period of observation.

as for more general models.

5.2 Notation

The data were collected during autumn, when *Ips* spp. were dormant until the beginning of the next attack period the following spring. For convenience we can therefore assume that the data were observed at the times $k = 0, 1, \ldots, 5$ which correspond to the end of the years $1987, 1988, \ldots, 1992$. Further, we let $k = -1$ correspond to the end of year 1986, and we say that time t is in year k if $k - 1 < t \leq k$.

We let $i = 1, \ldots, 807$ index the sites, i.e. the locations with living trees at time -1. Let $y_{t,i} = 0$ if site i has not been previously attacked by *Ips* spp. at time t, $y_{t,i} = 1$ if it has been attacked earlier in the same year, and $y_{t,i} = 2$ if it has been attacked in a previous year. By an "event" at the tree i we mean a transition of the zero-one process $v_{t,i} = \mathbf{1}[y_{t,i} \geq 1]$, where $\mathbf{1}[\cdot]$ denotes the indicator function. We denote the time of this transition t_i. We do not consider a transition $1 \to 2$ for $y_{t,i}$ as an event, since it is certain that this transition happens at the end of the year at which the event took place. Note that there is a one-to-one correspondence between the process $v_{s,i}$ for $s < t$ (or $s \leq t$) and the process $y_{s,i}$ for $s < t$ (or $s \leq t$).

The process $v_s = (v_{s,1}, \ldots, v_{s,807})$ is a particular kind of multivariate point process (or counting process), where each $v_{t,i}$ is restricted to be either zero or one. Such a process is specified by the conditional intensity function (Daley and Vere-Jones, 2003): For each tree i, given the history of the process v_s for times $s < t$, let

$$\lambda_{t,i} = \mathbf{E}\left[\mathrm{d}v_{t,i} \middle| (v_s)_{s<t}\right] / \mathrm{d}t$$

denote the conditional intensity of the tree being attacked by *Ips* spp. In the next section we specify models for the conditional intensity, allowing $\lambda_{t,i}$ to depend on covariate information $(x_{s,i})_{s<t}$, where $x_{t,i}$ denotes the number of turpentine beetles at time t and site i. It will also depend on external information related to *Ips* spp. activity and

on "neighboring information". For a site i, consider its first-, second-,... order neighbors, which are the (up to) four nearest, four second nearest,... sites to i, and let N_i denote the set of first to fifth order neighbors of i. Finally, let $u_{t,i} = \mathbf{1}[y_{t,i} = 1]$ and let $n_{t,i} = \sum_{j \in N_i} u_{t,j}$ be the number of neighbors of i that are infested with *Ips* spp. by time t in the same year.

5.3 Model

As usual, $t-$ means the time just before time t. Assume that for t in year k,

$$\lambda_{t,i} = \mathbf{1}[v_{t-,i} = 0]\rho(t)\left\{\psi_0 + \psi_1 n_{t-,i}^{\alpha_1} + \psi_2 n_{(k-1)-,i}^{\alpha_2} + \psi_3 x_{k-,i}\right\} \quad (5.1)$$

where ρ is a non-negative function, $\psi_0, \psi_1, \psi_2, \psi_3$ are non-negative parameters, and $\alpha_1 = \alpha_2 = 2$ (this choice and alternative models are discussed at the end of this section). The term $\mathbf{1}[v_{t-,i} = 0]$ is included, since *Ips* spp. do not attack the same tree twice. The meaning of the other terms in (5.1) is given below.

The function ρ incorporates external information about *Ips* spp. activity due to seasonal variation. Figure 5.2 shows ρ and reflects the fact that *Ips* spp. has a window of activity about four-five months and normally peak around July. Specifically, for t in year k, $\rho(t) = \varphi((t - \mu_k)/\sigma_k)$ where φ is the standard normal density function and the parameters μ_k and σ_k are determined as follows. Aukema *et al.* (2005) modeled *Ips* spp. activity as the number of *Ips* spp. caught in traps every week during the flight period in 2001–2002, using a linear regression model with various explanatory variables, but only the temperature is available in our study. Therefore we refit the model with temperature as the only explanatory variable, and estimate μ_k and σ_k by the empirical mean and standard deviation obtained from predicting the number of *Ips* spp. that would have been caught during each week of the kth year. Since $\mu_k - k$ and σ_k do not depend

greatly on k, the five normal densities in Figure 5.2 look rather similar relative to the years.

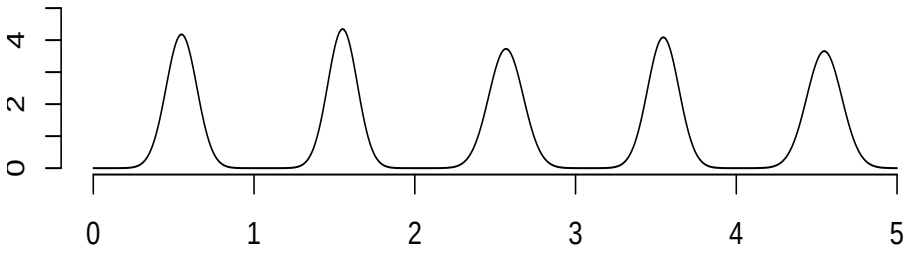


Figure 5.2: The function ρ.

The function ρ is the same for all sites, and is a rough description of *Ips* spp. activity depending only on the temperature. The term $\{\cdots\}$ in (5.1) adjusts for this. The individual terms in $\{\cdots\}$ play the following roles. The term ψ_0 is included since the other terms can be zero and we need then to scale the function ρ. The term $\psi_1 n_{t-,i}^{\alpha_1}$ appears, since *Ips* spp. may emerge from a tree attacked earlier in the year to attack nearby trees. The term $\psi_2 n_{(k-1)-,i}^{\alpha_2}$ appears, since *Ips* spp. overwinter in the ground close to a previously attacked tree and emerge to attack nearby trees in the following year. The term $\psi_3 x_{k-,i}$ reflects that *Ips* spp. tend to attack trees that previously have been attacked by turpentine beetles. Since we have only observed which trees the turpentine beetles have attacked at the end of the year, we should ideally also model the turpentine beetles as a continuous time process. However, to keep the model simple we refrain from doing that, and instead assume that the turpentine beetles contribute to the conditional intensity throughout the year. Since turpentine beetles usually attack earlier in the year than *Ips* spp. and $\rho(t)$ is close to being zero early in the year, this is probably not so unrealistic.

For each year k, the processes $(v_t)_{t<k}$ and $(v_t)_{t>k}$ are conditionally independent given v_k (however, v_t is not a continuous time Markov

process). We therefore condition on v_0 and consider the likelihood function based on the remaining data $d = (v_t)_{0<t\leq 5}$ (i.e. including the missing data on the time interval $[0, 5]$). Letting $\psi = (\psi_0, \psi_1, \psi_2, \psi_3)$, the likelihood is

$$L(\psi; d|v_0) = \prod_{k=1}^{5} \left[\prod_{i:\, k-1<t_i\leq k} \lambda_{t_i,i} \right] \exp\left(- \int_{k-1}^{k} \sum_i \lambda_{t,i} \, \mathrm{d}t \right). \quad (5.2)$$

Furthermore, we specify an improper uniform prior for ψ on $[0,\infty)^4$. Thus the conditional density of ψ given d (and v_0) is proportional to the likelihood (5.2). A rigorous proof that this conditional distribution of ψ is proper seems difficult, but from a practical point of view we would expect the MCMC runs described in Section 5.4.1 to diverge if the distribution was improper. This is not the case, and alternatively we could replace the range of ψ by a very large but bounded region.

Actually, before considering the model (5.1), we analyzed the model with

$$\lambda_{t,i} = \rho(t) \exp\left(\psi_0 + \psi_1 n_{t-,i} + \psi_2 n_{(k-1)-,i} + \psi_3 x_{k-,i}\right),$$

where $\psi_0, \psi_1, \psi_2, \psi_3$ are real parameters. This model is of a somewhat similar form as the model in Zhu *et al.* (2006), but a model check along similar lines as in Section 5.4.2 showed that the model did not fit the data well. Partly inspired by the form of the Hawkes process (Hawkes, 1971; Daley and Vere-Jones, 2003), we then turned to the model (5.1) but with $\alpha_1 = \alpha_2 = 1$. We observed a misfit and considered therefore alternative values $\alpha_1, \alpha_2 \in \{0, 1, 2\}$. We finally concluded that the model (5.1) with $\alpha_1 = \alpha_2 = 2$ fit best.

5.4 Inference

5.4.1 Posterior simulation and estimation

The posterior distribution is the conditional joint distribution of ψ and the missing data on $[0, 5]$ given the observations $y_0, y_1, \ldots, y_5$.

We use a Metropolis-within-Gibbs algorithm to simulate from the posterior distribution — we assume that the reader is familiar with Markov chain Monte Carlo (MCMC) methods; see e.g. Robert and Casella (2004).

Specifically, we propose to update each of the four parameters $\psi_0, \psi_1, \psi_2, \psi_3$ one at a time using Metropolis random walk steps with normal proposal distributions, where the proposal variances are chosen to reach an average acceptance ratio of approximately 0.25 (Roberts *et al.*, 1997). Moreover, within a given year k at a site i either one event $t_i \in (k-1, k]$ has happened, where we do not know the exact value of t_i, or nothing has happened. If an event has happened at site i, we need to simulate the unknown t_i from its conditional distribution given "everything else", i.e. from the density proportional to $\lambda_{t,i}$, with $t \in (k-1, k]$. We do this by visiting all the sites with events in some pre-determined order, updating t_i by an independent Metropolis sampler with proposal density $\rho(t)$, $t \in (k-1, k]$.

Figure 5.3 shows plots of the posterior distributions of ψ_0, ψ_1, ψ_2, ψ_3 based on an MCMC run length of 100,000 with a burn-in length of 1,000. Note that all of the parameters are clearly bounded away from zero. Obviously, since ψ_1 respective ψ_2 is significant, the probability that a tree is going to be attacked increases when more trees in the neighborhood have been attacked earlier in the year respective in the previous year. This leads to several potential biological mechanisms. One is that once a beetle enters a tree and emits aggregation pheromones (Wood, 1982), large numbers of beetles from previously attacked neighboring trees are available to respond and hence rapidly exhaust the tree's defenses (Raffa and Berryman, 1983). Second, factors that predispose trees to being susceptible to attack may be distributed in a highly clustered fashion, and the resident population of beetles again respond to the first entries thereby generating the observed pattern (Erbilgin and Raffa, 2003). Obviously these are not mutually exclusive. Finally, the significance of ψ_3 in (5.1) implies that trees are more susceptible to being attacked by *Ips* spp. if they have been attacked by turpentine beetles previously.

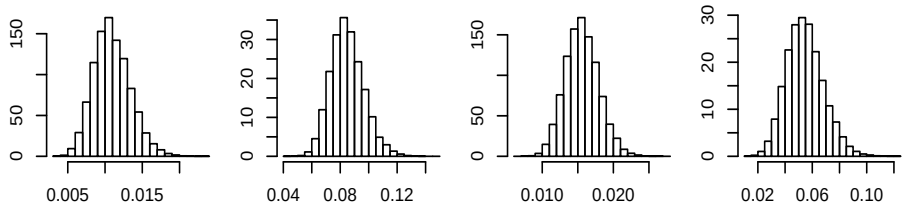


Figure 5.3: Posterior distributions of $\psi_0, \psi_1, \psi_2, \psi_3$.

Figure 5.4 shows estimated results for the missing event times t_i. The results are based on an MCMC run length of 100,000 with a burn-in length of 1,000 and sampling every 25th missing data set. The first row in the figure shows for each year 1–5 the estimated mean value of the t_i, using a gray scale where white means that no event happened at the site and darker values correspond to early events. During the five years the missing data are located further and further away from the initial gap of dead trees, and within each year the mean values tend to be larger at locations further away from this gap. The second row in Figure 5.4 shows the standard deviations of the event times, where dark values correspond to small standard deviations. For each year, the standard deviations are about twice as large for isolated attacks than for those occurring in clumps. Since the degree of clumping varies from year to year, the standard deviations in e.g. year 3 are higher than in the other years. The histograms in the last row in Figure 5.4 show for each year 1–5 the empirical distribution of all times of events during the year. The histograms are rather similar except for the difference in years, and they look much like the normal densities given by $\rho(t)$ except for a shift about half a month to the right.

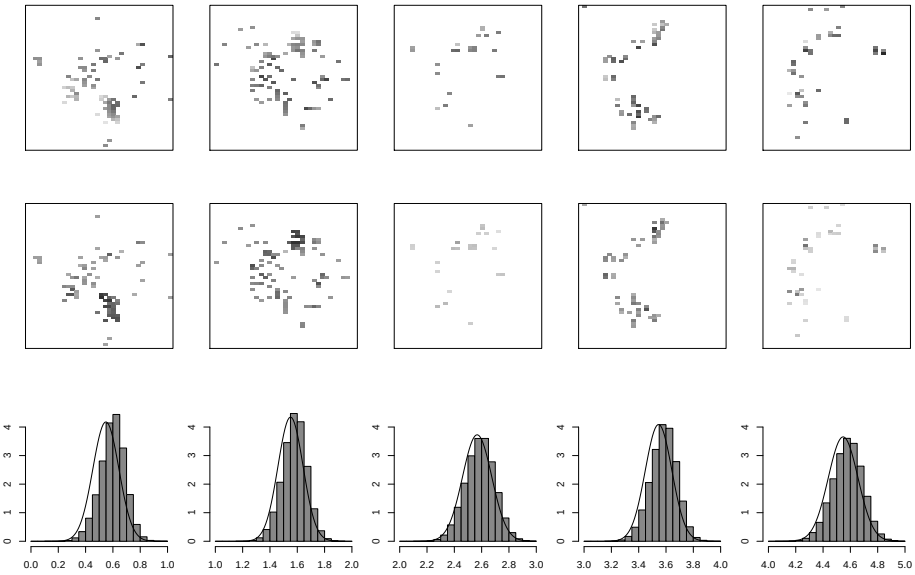


Figure 5.4: Columns: years $1, 2, 3, 4, 5$. First row: gray scale plots of estimated mean event times. Second row: gray scale plots of estimated standard deviations of event times. Last row: histograms of all event times and $\rho(t)$ within each year.

5.4.2 Model checking

Following the idea of posterior predictive model checking (Gelman *et al.*, 1996), we consider posterior predictions: for each $l = 1, 2, \ldots$, we simulate a realization from the posterior distribution and use this when simulating "new data" $u_{t,i}^{(l)}$ and $v_{t,i}^{(l)}$ from the observation model; see Appendix A. For convenience, let $u_{t,i}^{(0)} = u_{t,i}$ and $v_{t,i}^{(0)} = v_{t,i}$ denote the data.

To check how well the model fits the data, we first consider the number of attacks in each year,

$$w_1^{(l)}(t) = \sum_i u_{t,i}^{(l)},$$

where $t = 1, \ldots, 5$. The left plot in Figure 5.5 shows the $2.5, 50, 97.5$ percentiles estimated from $w_1^{(l)}(t)$ for $l = 0, \ldots, 199$. The values of $w_1^{(0)}(t)$ is also shown in the figure. In all years except year 3 where the number of *Ips* spp. attacks was particularly low, $w_1^{(0)}(t)$ is located in the central interval.

Second, let $d(i, j)$ denote Euclidean distance between site i and j. For $\delta \geq 1$,

$$w_2^{(l)}(\delta) = \sum_{i<j:\, d(i,j)\in(\delta-1,\delta]} v_{5,i}^{(l)} v_{5,j}^{(l)},$$

is the number of pairs of sites between $\delta - 1$ and δ apart and attacked at some time during the observation period. Thus $w_2^{(l)}(\delta)$ quantifies the degree of spatial clustering. The center plot in Figure 5.5 shows $w_2^{(l)}(\delta)$ for $\delta = 1, \ldots, 5$ in the same way as in the first plot. Again there are no discrepancies between the model and the data.

Third, in order to combine temporal and spatial information, consider the number of neighboring sites that are attacked in years time t apart,

$$w_3^{(l)}(t) = \sum_{i,j,k:\, j\in N_i, k=t,\ldots,5} u_{k,i}^{(l)} u_{k-t,j}^{(l)}$$

for $t = 1, \ldots, 4$, while

$$w_3^{(l)}(0) = \sum_{i,j,k:\, j \in N_i, k=0,\ldots,5} u_{k,i}^{(l)} u_{k,j}^{(l)} / 2$$

where we divide by 2 to avoid counting all the pairs twice. The right plot in Figure 5.5 shows $w_3^{(l)}(t)$ for $t = 0, \ldots, 4$ in the same way as the plot for $w_1^{(l)}(t)$, except that we have taken the logarithm of $w_3^{(l)}(t)$ to be able to see what happens at times $t = 3, 4$. For $t = 0, \ldots, 2$, $w_3^{(0)}(t)$ is located in the 95% central interval, but for $t = 3, 4$ the model underestimates the number of pairs.

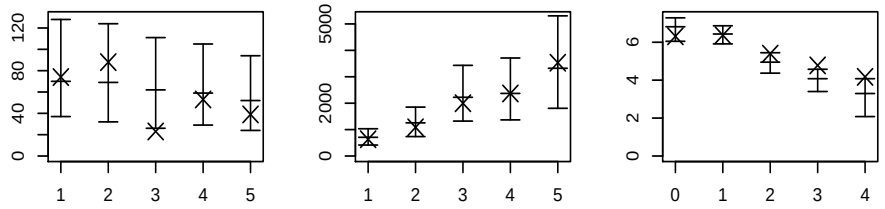


Figure 5.5: Plots of $w_1(t)$, $w_2(\delta)$, and $\log(w_3(t))$. The crosses indicate the data and the bars indicate the 95% central interval and the median.

Finally, the upper row in Figure 5.6 shows plots of the data for each year $t = 1, \ldots, 5$ where a site i at time t has been colored black if $v_{t,i}^{(0)} = 1$ and gray otherwise. Simulating 1000 posterior predictions, we let $\tilde{v}_{t,i} = 1$ if $v_{t,i}^{(l)} = 1$ in more than 50% of the simulations and $\tilde{v}_{t,i} = 0$ otherwise. The lower row in Figure 5.6 shows plots of $\tilde{v}_{t,i}$. Comparing the two rows in the figure, we see that both the data and the posterior predictive simulations show a clear formation of a large cluster of infested trees in the middle. Furthermore, the clusters seem to be roughly of the same size in the data and the simulations. On the other hand, there are some discrepancies between the shape

of the cluster in the data and in the simulations: the shape of the cluster is more circular in the data than in the simulations. However, in spite of this deviation, the general behavior still seems to have been captured adequately well by the model.

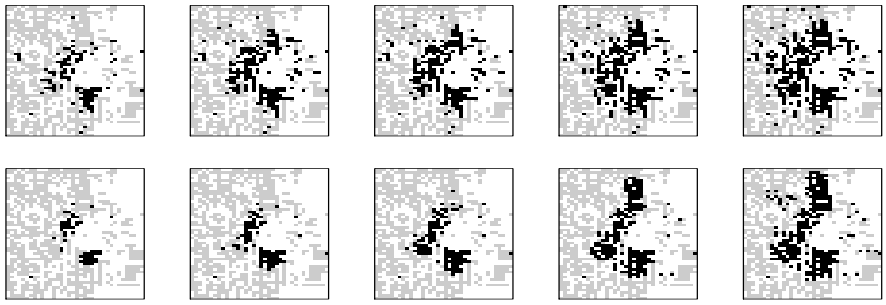


Figure 5.6: Upper row: A site is colored black if $v_{t,i} = 1$ and gray if $v_{t,i} = 0$ for times $t = 1, \ldots, 5$ (left to right). Lower row: same as upper row, but for $\tilde{v}_{t,i}$.

5.5 Comparison between continuous and discrete time models

We have illustrated how discrete time observations of a multivariate point process can be analyzed using a Bayesian missing data approach. We conclude with a discussion of the advantages and disadvantages of using continuous time processes compared to discrete time processes in the setting of the present paper as well as other spatial-temporal situations.

Computation: The most important advantage is the ease of computation. The likelihood function for a spatial-temporal process with discrete time often involves one or more unknown normalizing constants which need to be estimated using MCMC methods, see e.g.

the Markov random field models in Besag and Tantrum (2003) and Zhu *et al.* (2006). In contrast the likelihood function for a multivariate point process is completely specified by modeling the conditional intensity, cf. equation (5.2). In Zhu *et al.* (2006), we modeled the data set consisting of the three types of data given by turpentine beetles, *Ips* spp. and tree conditions. The likelihood function for this model factorized into three terms, one for each type of data, where we can compare the term corresponding to *Ips* spp. and the model used in the present paper, when we include the *Ips* spp. at all sites (rather than restricting the data to the subset of sites considered so far in the present paper). For this particular comparison, the computation time of the MCMC algorithm for posterior distributions for the discrete time model is roughly two hundred times longer than the corresponding computation time for the continuous time model. This improvement in speed means that we have been able to investigate several variations of the model (5.1) (see the end of Section 5.3), something we could not do within practical time limits in the case of the discrete time process for *Ips* spp. in Zhu *et al.* (2006). On the other hand, in Zhu *et al.* (2006) the likelihood terms corresponding turpentine beetle colonization and tree conditions were easy to specify, and calculations for these terms were very fast. Extending the continuous time model to include turpentine beetles and tree conditions may be much more involved, and a comparison between such a model and the full model in Zhu *et al.* (2006) may well turn out differently.

External information: For a discrete time process as compared to a continuous time process, it may be difficult to incorporate external (or, in a broad sense, covariate) information in the form of another stochastic process observed at a different time scale than the discrete time process. For example, the ρ term in (5.1) incorporates external information about *Ips* spp. activity due to seasonal variation; such information can only be incorporated into the model in Zhu *et al.* (2006) after some form of aggregation. On the other hand, we require the external information that was collected in the field across

a full season of *Ips* spp. activity modeled into ρ to obtain a realistic continuous time model for *Ips* spp. attacks. If we had no such information or only unreliable information available, the continuous time model approach would be problematic. It would also be a problem if the modeling form of ρ is misspecified. Specifying ρ is time consuming, and this partially offsets the advantage of the shorter computation time for the continuous time model.

Time scale comparability: While the parameters of two continuous time processes may be compared even if their time scales for observations (i.e. time lengths between observation times) are different, it may not be meaningful to compare the parameters of two discrete time processes with different time scales. The choice of observation times in Zhu *et al.* (2006) is biologically meaningful, since we have annual observation after cessation of insect activity; however, suppose that we had another data set with observations every second year. Further, imagine that we wish to use a model of the same form as in Zhu *et al.* (2006) for this other data set. Then, because of the different time scales, we cannot directly compare the parameters governing e.g. *Ips* spp. activity.

Consistency: Although observed only at discrete times, the types of data considered in the present paper come from an underlying continuous time process, and the existence of an underlying continuous time process is not ensured by specifying a discrete time process. Obviously, this is not an issue when we have specified a continuous time model, but our continuous time model only approximates the complexity of the system under study. Indeed, *Ips* spp. colonization does not occur exactly at one time point; it is a complicated process involving hundreds or thousands of beetles attacking over a short period of time. On the other hand, the discrete time model reflects the cumulative *Ips* spp. attacks throughout the past season, and in this sense it is a perfectly sensible model for the system under study.

Estimation of missing data: A continuous time process allows us to model what has happened between observation times, where in our case the event times of within-season *Ips* spp. attacks can be readily

estimated by the MCMC algorithm in Section 5.4.1. In contrast it is not possible to do this kind of estimation for missing data based on the model in Zhu *et al.* (2006), although it was not a part of the scientific objectives there. In the present paper, the estimates provide some trustful qualitative results regarding how the times of events depend on the distance to the gap of dead trees and the similar behavior over the years. However, quantitative results depend much on a careful modeling of external information, particularly the function ρ.

Predictions: It is straightforward to predict what may happen at any time after the final observation time, applying Ogata's modified thinning algorithm for the continuous time process in the present paper (see Appendix A). The discrete time model in Zhu *et al.* (2006) allows us only to predict what may happen annually.

In conclusion, spatial-temporal processes often suffer from being computationally intensive, and we have demonstrated that in the present case, using a continuous time process is indeed a viable alternative to a discrete time process.

Appendix A

For model checking in Section 5.4.2, we need to simulate new data from the observation model. For this we use Ogata's modified thinning algorithm extended to marked processes (Daley and Vere-Jones, 2003; Ogata, 1981).

Factorize the intensity, $\lambda_{t,i} = \lambda_t \times p(i|t)$, where $\lambda_t = \sum_i \lambda_{t,i}$ is the intensity for the temporal process of events disregarding the location, and $p(i|t) \propto \lambda_{t,i}$ is the probability function for the location given the time of the event. Note that the dependence on the past events (locations and times) of both functions are suppressed in the notation. Furthermore, define two functions $l(t)$ and $m(t)$ by $l(t) = t + 0.1$ for $k-1 < t \leq \mu_k$ and $l(t) = k$ for $\mu_k < t \leq k$, and $m(t) = \max_{s\in[t,l(t)]} \lambda_s$. Ogata's modified thinning algorithm is then started at time $t = 0$ and

the following steps are repeated until $t > 5$:

1. Compute $l(t)$ and $m(t)$.
2. Generate an exponentially distributed variable T with inverse mean $m(t)$ and a uniform variable U on $[0, 1]$ independently of each other.
3. If $T > l(t)$, set $t = t + l(t)$.
4. Else if $t + T > 5$ or $U > \lambda_{t+T}/m(t)$, set $t = t + T$.
5. Otherwise, let the next event be $t_i = t+T$, where i is generated using the probability function $p(i|t)$, and set $t = t + T$.

The output is the set of all t_i obtained in step 5.

Combining this algorithm with a sample of the posterior distributions in Section 5.4.1, we get posterior predictions. In practice we get a sample of the posterior distributions by taking values from the Markov chains at regular intervals chosen such that the sample is effectively independent.

Acknowledgments

This work was supported by the Danish Natural Science Research Council, USDA NRI (2003-3502-13528), NSF (DEB0314215), and the UW-Madison College of Agricultural and Life Sciences. We are grateful to the Wisconsin Department of Natural Resources for providing the study site.

References

Aukema, B. H., Clayton, M. K., and Raffa, K. F. (2005). Modeling flight activity and population dynamics of the pine engraver, *Ips pini*, in the Great Lakes region: Effects of weather and predators at short time scales. *Population Ecology*, **47**, 61–69.

Besag, J. and Tantrum, J. (2003). Likelihood analysis of binary data in space and time. In P. J. Green, N. L. Hjort, and S. Richardson, editors, *Highly Structured Stochastic Systems*, pages 289–295. Oxford University Press, Oxford.

Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods*. Springer, New York, 2nd edition.

Erbilgin, N. and Raffa, K. F. (2003). Spatial analysis of forest gaps resulting from bark beetle colonization of red pines experiencing belowground herbivory and infection. *For. Ecol. & Manag.*, **177**, 145–153.

Furniss, R. L. and Carolin, V. M. (1980). *Western Forest Insects*. USDA FS Misc. Publ. 1339. Washington DC. 654 pp.

Gelman, A., Meng, X. L., and Stern, H. S. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica*, **6**, 733–807.

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

Ogata, Y. (1981). On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, **IT-27**, 23–31.

Raffa, K. F. and Berryman, A. A. (1983). The role of host plant resistance in the colonization behavior and ecology of bark beetles. *Ecol. Monogr.*, **53**, 27–49.

Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer, New York, 2nd edition.

Roberts, G. O., Gelman, A., and Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Annals of Applied Probability*, **7**, 110–120.

Wood, D. L. (1982). The role of pheromones, kairomones, and allomones in the host selection and colonization behavoir of bark beetles. *Ann. Rev. of Entomol.*, **27**, 411–446.

Zhu, J., Rasmussen, J. G., Møller, J., Aukema, B. H., and Raffa, K. (2006). Spatial-temporal modeling of forest gaps generated by colonization from below- and above-ground beetle species. Research report R-2006-04, Department of Mathematical Sciences, Aalborg University. Available at `http://www.math.aau.dk/~jm`. (Submitted for publication).

Chapter 6

Software and work in progress

6.1 Software

Since all the papers included in this thesis have been computationally heavy and required much programming, I have added this section describing some (though far from all) of the many programs used. This allows other people to use or be inspired by the programs. Section 6.1.1 describes some general issues of the programs, and Sections 6.1.2–6.1.4 are dedicated to the specific programs used in each chapter. I assume that the reader is familiar with the corresponding theoretical chapters before reading the following sections.

6.1.1 Downloading, compiling and running programs

All the implementations have been done in `C++` on a Linux system, but should work equally well on Windows. The files needed can be downloaded from

`www.math.aau.dk/~jgr/downloads/`

where the name of the file needed is stated in each of the following sections. All files have been compressed using `zip`, and can be uncompressed using the command `unzip filename` (on a Linux system). The specific files retrieved from the compressed file are described in the following sections. To compile one of the programs on a Linux system, use the command

```
g++ -o outputfile inputfile -lm
```

where `inputfile` is the name of the `C++` source code file to be compiled, and `outputfile` is the user-specified name of the program file to be generated. The program is then run with the command `outputfile`. Any parameters should be set in `inputfile` prior to compiling and running the program; the parameters are described in the following sections. Note that only the files described as main programs in the following sections can be compiled and run; trying to compile other files will result in an error.

6.1.2 Software for Chapters 2 and 3

This section explains the use of some of the programs I have used for simulating Hawkes processes in Chapters 2 and 3. From Chapter 2, I have included the perfect simulation algorithms (Algorithms 2.2 and 2.3 combined) for the Hawkes process with exponentially decaying offspring intensity (Example 2.1) and the birth and death type Hawkes process with exponential life times (Example 2.2). From Chapter 3, I have included the approximate simulation algorithm (Algorithm 3.1) for the same examples.

All the files needed can be extracted from the file `SimHawkes.zip` as described in Section 6.1.1. There are two types of files included. The first type are the main programs for simulating the two processes with the two algorithms:

- `SimHawkesASExp.cpp`: Approximate simulation of the Hawkes process with exponential offspring intensity.

- `SimHawkesASBD.cpp`: Approximate simulation of the birth-death Hawkes process.

- `SimHawkesPSExp.cpp`: Perfect simulation of the Hawkes process with exponential offspring intensity.

- `SimHawkesPSBD.cpp`: Perfect simulation of the birth-death Hawkes process.

The other files contain the algorithms that the main programs use (these files cannot be compiled and run as separate programs). The files are:

- `SimHawkesTools.cpp`: Basic tools needed by the other programs.

- `SimHawkesExp.cpp`: Simulation algorithms for the Hawkes process with exponential offspring intensity.

- `SimHawkesBD.cpp`: Simulation algorithms for the birth-death Hawkes process with exponential life times.

Prior to compiling and running the programs, various parameters found in the main programs can be changed:

- `mu` > 0: The intensity of the immigrant process.

- $0 <$ `alpha` < 1: The mean number of offspring generated by any event.

- `beta` > 0: The parameter used in the exponential offspring intensity (for the Hawkes process with exponential offspring intensity) or the inverse mean of life times (for the birth-death Hawkes process).

- `tplus` > 0: Simulation interval is $[0, \texttt{tplus}]$.

- `tminus` ≤ 0: Starting value for approximate simulation algorithms, i.e. the approximate simulation algorithm simulates immigrants on the interval [`tminus`, `tplus`]
- `ngrid` $\in \mathbb{N}$: Number of grid points used in quadrature rule for approximating integrals in perfect simulation algorithms. Note that this should not be set too low.

See Chapters 2 and 3 for more details on the parameters.

The output of either one of the four main programs is a text-file called either `hawkesexp.dat` or `hawkesbd.dat` depending on whether it is a Hawkes process with exponential offspring intensity or a birth-death Hawkes process. The file `hawkesexp.dat` contains one column of simulated event times. The file `hawkesbd.dat` contains two columns of real numbers - the first one is the event times, and the second one is the marks (i.e. life times).

6.1.3 Software for Chapter 4

This section describes the implementation of one of the algorithms used in Chapter 4. This is the MCMC algorithm for approximating the posterior distribution of parameters $(\psi_0, \psi_1, \psi_2, \psi_3)$ (see Section 4.4 for details). The reason for only including this single program out of the many programs used is that it contains most of the interesting algorithms (e.g. perfect simulation and path sampling), and already it is rather comprehensive.

The files included can be extracted from `BeetlesTrees.zip` as described in Section 6.1.1. The following file contain the main program:

- `mcmc_psi_path.cpp`: The main program for the MCMC runs for approximating the posterior distributions of the parameters $(\psi_0, \psi_1, \psi_2, \psi_3)$ using path sampling for approximating the unknown normalizing constants.

The algorithms used by the main programs are included in the following files (these files cannot be compiled and run):

- `data.cpp`: Algorithms for loading data and creating neighbourhood structures.
- `sim_autolog_process.cpp`: Algorithms for making perfect simulations of autologistic processes.
- `random.c` and `random.h`: Algorithms for simulating random variables.

Finally all files with the extension `.dat` are data files. Note that since the data is not publicly available the files included are fake data, i.e. the data is fine for illustrative purposes, but will not give the same results as those obtained in Chapter 4.

- `turp87sim.dat`,...,`turp92sim.dat`: Matrices of turpentine beetle data for each year, where each entry indicates the state of the tree at a position in a regular grid; an entry of -1 means there is no tree at the position, 0 means that the tree at this position has not been attacked at this time, an entry larger than 0 means that this number of turpentine beetles is present in the tree at this position.
- `ips87sim.dat`,...,`ips92sim.dat`: Ips spp. colonization data represented as a matrix for each year; an entry of -1 means there is no tree at the position, 0 means that the tree has not been attacked at this time, 1 means it has been attacked.
- `trees86sim.dat`,...,`trees92sim.dat`: Matrices of tree condition data for each year; an entry of -1 means there is no tree at the position, 0 means that the tree at this position is alive, and 1 means that it is dead.

The program contains the following parameters that can be adjusted before compiling and running the programs:

- `NumSteps`: Number of steps used in the MCMC run. Note that the algorithm is rather slow, and setting `NumSteps` too high may result in a very long running time.

- `NumSim`: Number of perfect simulations used in the approximation of the normalizing constants when using path sampling.

- `sigma0`,...,`sigma3`: Standard deviations used in the normally distributed proposals for each of the four parameters.

The output of the program is a file named `mcmc_psi.dat` containing a matrix with one row for each step in the MCMC run. Each row has five columns: the first four entries are the values of $(\psi_0, \psi_1, \psi_2, \psi_3)$ and the last row contains the acceptance ratio for this particular MCMC step. This output is also written to the screen when the program is running.

6.1.4 Software for Chapter 5

This section describes some of the programs used in Chapter 5. Specifically, the software described is the implementations of the MCMC algorithms for approximating the posterior distribution of parameters (see Section 5.4.1) and Ogata's modified thinning algorithm for generating posterior predictions (see Appendix A in Chapter 5).

The files needed can be extracted from the file `BeetlesCont.zip` as described in Section 6.1.1. The following files contain the main programs:

- `mcmcclin.cpp`: MCMC algorithms for approximating posterior distributions of the parameters in the model.

- `csimclin.cpp`: Algorithms for simulating event times, i.e. the times of Ips spp. attacks.

The algorithms used by the main programs are included in the following files (these files cannot be compiled and run):

- `datac.cpp`: Algorithms for loading data and creating neighbourhood structures.

- `thinningclin.cpp`: Ogata's modified thinning algorithm used for simulating times of Ips spp. attacks.

- `random2.c` and `random2.h`: Algorithms for simulating random variables.

Finally all files with the extension `.dat` are data files. As in Section 6.1.3, the data is not the real data, and will not give the same results as those obtained in Chapter 5.

- `rho.dat`: Coefficients for $\rho(t)$.

- `ips87sim.dat`,...,`ips92sim.dat`: Ips spp. data represented as a matrix; an entry of -1 means there is no tree at the position, 0 means that the tree has not been attacked at this time, 1 means it has been attacked.

- `turp87sim.dat`,...,`turp92sim.dat`: Turpentine beetles data represented as a matrix; an entry of -1 means there is no tree at the position, 0 means that the tree has not been attacked at this time, an entry larger than 0 means that this number of turpentine beetles is present in the tree.

- `treessim.dat`: Tree data represented as a matrix; an entry of -1 means there is no tree at the position, 0 means that the tree is alive, 1 means it is dead. Only positions with living trees are used by the algorithm.

The program `mcmcclin.cpp` contains the following parameters that can be adjusted before compiling and running the programs:

- `NumSteps`: Number of steps used in the MCMC algorithm.

- `inc0`,...,`inc6`: Various models have been tried in Chapter 5, and the programs described in this section not only includes the four parameters $(\psi_0, \psi_1, \psi_2, \psi_3)$ from model (5.1) in Chapter 5, but also three other parameters. In the program the parameters

are numbered $0, \ldots, 6$, and the parameters numbered $(0, 2, 4, 5)$ correspond to $(\psi_0, \psi_1, \psi_2, \psi_3)$. Parameters $(1, 3)$ corresponds to (ψ_1, ψ_2) but without squaring, and parameter 6 corresponds to ψ_3 but with $x_{k-,i}$ squared. Setting `incx` equal to one, where `x` $= 0, \ldots, 6$, includes the corresponding term.

- `sd0`,. . . ,`sd6`: Standard deviations used for making normally distributed proposals in the MCMC algorithm for each of the seven parameters.

The output of the program `mcmcclin.cpp` is a file called `psi.dat` containing a matrix, which includes a row for each step in the MCMC run. Each row contains 14 entries of parameter values and acceptance ratios; the first two entries contain the parameter value and acceptance ratio for parameter 0, entries three and four contain the parameter value and acceptance ratio for parameter 1, and so on. Any parameter not included (i.e. by setting `incx` equal to zero, where `x` $= 0, \ldots, 6$) is set equal to zero in this file.

The program `csimclin.cpp` contains the following parameters:

- `type`: If `type` equals 0 the parameters used in the simulation is taken from the file `psi.dat` (i.e. it can be considered a random draw from the posterior distributions of the parameters), and if it equals 1 the parameters are user-specified.

- `step`: In the case of parameters being taken from `psi.dat`, this specifies which step in the Markov chain the parameters are taken from.

- `psi0`,. . . ,`psi6`: In the case of parameters being user-specified, these specifies the values of the seven parameters. Here the four parameters (`psi0`,`psi2`,`psi4`,`psi5`) correspond to (ψ_0,ψ_1,ψ_2,ψ_3) from Chapter 5 as described above.

The output of this program is a matrix of times of Ips spp. attacks saved in the file `simtimes.dat`. An entry of -1 specifies a position

where there is no tree, and entry of -0.5 means that the tree at this position has been attacked before time 0, and an entry of 1000 means that the tree at this position has not been attacked in this simulation. The values in the rest of the entries are the simulated times of Ips spp. attacks.

6.2 Work in progress

This section describes other projects that has been started but not finished during my PhD-study. These projects are at different stages of completion, and may well be the focus of future work.

6.2.1 Modelling the spatial distribution of barrows

One project concerns the spatial distribution of barrows (bronze age burial sites). Figure 6.1 shows the location of barrows on an observation region in a part of Jutland in Denmark. From the plot it is evident that, while some of the barrows are distributed fairly evenly across the observation region, many tend to form linear shapes. The classical theory states that the reason for these linear shapes is that barrows are located close to roads, cf. Müller (1904) (this and other theories are mentioned in Sahlquist (2001)). However, we are examining a model for the spatial distribution of barrows, where new barrows are placed close to older barrows. It turns out that this model results in linear shapes resembling those of the data.

More specifically, we have specified a point process model on a set $W \subset \mathbb{R}^2$ for the spatial point pattern of barrows, where the points are placed one at a time conditionally on previously placed points. Initially a new point x_i follows some distribution on W independently of everything else; denote this position y_i. We can think of y_i as the place were a individual dies and x_i as the place where this individual is buried. The first point $x_1 = y_1$ keeps this initial position. There-

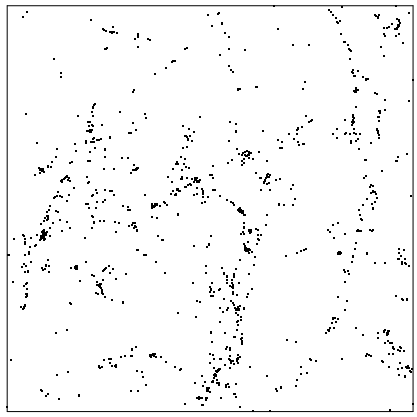


Figure 6.1: The locations of barrows.

after each point x_i has a small probability of staying at the initial placement, i.e. $x_i = y_i$; otherwise x_i will be moved close to the closest previously placed point, say $x_j \in \{x_1, \ldots, x_{i-1}\}$. We have tried various ways of moving a point x_i closer to a previous point x_j, and finally settled on letting x_i follow a truncated normal distribution on the line going through the initial placement y_i and the closest previously placed point x_j. This normal distribution has mean parameter x_j, its variance parameter is a parameter in the model, and it is truncated at the edges of the Voronoi cell of x_j in the Voronoi tessellation constructed from $x_1, \ldots, x_{i-1}$. Since not all points in the data form linear shapes, we superpose this process with another process where a point follows some distribution on W independently of everything else.

For inference we use a Bayesian approach, where we estimate the unknown parameters of the model using MCMC. This also includes the estimation of missing data; for examples, since we have no information on when the individual barrows were built, we do not know the order in which they should appear in the model. So far some

preliminary results have been obtained using simulation, including posterior distributions of the parameters of the model and model checking based on various summary statistics. These results have been obtained from programs I have made (except for the programs generating the Voronoi tessellations which have kindly been lent to us by Øivind Skare). They indicate that the model fits the data fairly well.

We have also considered including various covariates regarding the landscape to make a more realistic model. For example, the density of settlements is lower in very marshy terrains than on good soil, and it seems unlikely that people would cross a big river to bury a person on the other side. Therefore, extending the model to include information about population densities on different soil types and the difficulty of moving on certain terrain types would be useful to obtain a more realistic model. However, since the model without these extensions already involves many calculations (mainly because of the Voronoi tessellations), the computations are already somewhat slow. It seems difficult to incorporate the landscape in a way that is both realistic and computationally feasible. Fortunately, the data shown in Figure 6.1 seems to be quite homogeneous with respect to the landscape, and it should be possible to obtain a realistic model for this particular data set without considering these covariates.

6.2.2 Biological conclusions derived from beetle colonization models

The papers in Chapters 4 and 5 are treating the same data set from two different points of view. Chapter 4 is an applied statistical paper where the aim is a Bayesian spatio-temporal autoregressive type of model describing the locations of beetle attacks and tree deaths. In Chapter 5 the aim is an alternative statistical methodology, a multivariate point process model with missing data, for modeling a subset of this data. A third paper by the same authors is presently in preparation. This paper treats the biological implications of the

conclusions in the two previous papers (mainly the paper treated in Chapter 4). Furthermore, the paper includes an additional type of beetle known as root weevils (Coleoptera: Scolytidae; alt. Curculionidae: Scolytinae).

There are two primary biological issues treated in this paper. The first one is the impact of the different types of beetles on tree mortality. The question here is what the role of the various types of beetles are. For example, in Chapter 4 we discovered that *Ips* spp. are a cause of death of the trees, and, although turpentine beetles are not killing the trees directly, they predipose the trees to later attacks by *Ips* spp. The second issue is the spatio-temporal spread of the different kinds of beetles, and how they relate to gap formations. The full biological implications of the conclusions derived regarding both of these issues are important for formulating forest management strategies.

6.2.3 Spatial Hawkes processes and applications

The Hawkes process consider in Chapters 2 and 3 can easily be generalized to a spatial version on $\mathbb{R}^d$, see e.g. Møller and Torrisi (2005) for the definition. As Møller and Torrisi (2005) points out, this process is a natural process for modelling the spatial spread of epidemics or individuals from a reproducing population.

Recently the spatial Hawkes process has received some attention from a theoretical point of view. Møller and Torrisi (2005) obtain expressions for various summary statistics, while Brémaud *et al.* (2005) consider power spectra. However, there seems to be no practical work done using the spatial Hawkes processes so far. Jesper Møller and I have discussed various possible data sets. Particularly, we have considered the weed data set modelled in Brix and Møller (2001) and Brix and Chadoeuf (2002).

References

Brémaud, P., Massoulié, L., and Ridolfi, A. (2005). Power spectra of random spike fields and related processes. *Advances in Applied Probability*, **37**(4), 1116–1146.

Brix, A. and Chadoeuf, J. (2002). Spatio-temporal modeling of weeds and shot-noise G Cox processes. *Biometrical Journal*, **44**(1), 83–99.

Brix, A. and Møller, J. (2001). Space-time multitype log gaussian processes with a view to modelling weed data. *Scandinavian Journal of Statistics*, **28**, 471–488.

Møller, J. and Torrisi, G. L. (2005). Second order analysis for spatial hawkes processes. Research Report R-2005-20, Department of Mathematical Sciences, Aalborg University.

Müller, S. (1904). Vei og bygd i sten- og bronzealderen. *Aarbøger for Nordisk Oldkyndighed og Historie*, pages 1–64.

Sahlquist, L. (2001). Territorial behaviour and communication in a ritual landscape. *Geografiska Annaler*, **83 B**(2), 79–102.