Aalborg Universitet

Statistics for point processes on linear networks and on the space cross sphere

Christensen, Heidi Søgaard

Publication date:
2019

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

Link to publication from Aalborg University

Citation for published version (APA):
Christensen, H. S. (2019). *Statistics for point processes on linear networks and on the space cross sphere*. Aalborg Universitetsforlag. Ph.d.-serien for Det Ingeniør- og Naturvidenskabelige Fakultet, Aalborg Universitet

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

? Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
? You may not further distribute the material or use it for any profit-making activity or commercial gain
? You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: November 25, 2020

STATISTICS FOR POINT PROCESSES ON LINEAR NETWORKS AND ON THE SPACE CROSS SPHERE

BY
HEIDI SØGAARD CHRISTENSEN

DISSERTATION SUBMITTED 2019

AALBORG UNIVERSITY
DENMARK

Statistics for point processes on linear networks and on the space cross sphere

PhD Dissertation

Heidi Søgaard Christensen

Dissertation submitted August 30, 2019

Dissertation submitted: August 30, 2019

PhD supervisor: Prof. Jesper Møller
Aalborg University

Assistant PhD supervisor: Assoc. Prof. Jakob Gulddahl Rasmussen
Aalborg University

PhD committee: Associate Professor Poul Svante Eriksen (chairman)
Aalborg University

Doc. Tomás Mrkvicka
University of South Bohemia

Associate Professor Ute Hahn
Aarhus University

PhD Series: Faculty of Engineering and Science, Aalborg University

Department: Department of Mathematical Sciences

ISSN (online): 2446-1636
ISBN (online): 978-87-7210-488-1

Published by:
Aalborg University Press
Langagervej 2
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

© Copyright: Heidi Søgaard Christensen

Printed in Denmark by Rosendahls, 2019

Abstract

This thesis concerns point processes which are used for modelling a certain type of data, where each data point describes the location of an object or the time of an event. Such data are called point patterns and arise within a variety of applications as for example forestry, astronomy, neuroscience, criminology, or seismology, where the data points can describe the location of trees in a forest, the time and position of fireballs observed from Earth, the location of neurons in the brain, the site of crimes within a city, or the time of an earthquake. Thus point pattern data can be observed on a wide range of spaces such as one-, two-, and three-dimensional Euclidean spaces, spheres or linear networks. Depending on the nature of this space - and the way distance is measured - different tools and models exist for analysing the point pattern. Most of the existing literature on point processes concerns Euclidean spaces, but some advances has also been made for point processes on linear networks or on spheres. In some cases, theory from the Euclidean space set-up can easily be extended or modified to other types of spaces, but in other cases such extensions are not straightforward or even possible. The thesis concerns modelling and inference for point patterns on linear networks, the Euclidean space, or the product space between the Euclidean space and there sphere.

The thesis consists of two parts, where Part I is a brief introduction to relevant concepts and methods aiming at giving the reader the necessary background for understanding the scientific contribution of Papers A–D in Part II. Specifically, Part I reviews existing theory for point processes on the Euclidean space, linear networks, or the sphere and also introduces two specific types of point patterns of particular interest in the thesis: point patterns on the product space between the three-dimensional Euclidean space and the two-dimensional sphere describing the location of pyramidal cells in a part of a human brain, and point patterns on linear networks describing the location of spines on a dendrite tree from a mouse neuron.

The second-order moment properties of a point process describe how a pair of points from the point process interact. It is common to investigate such an interaction for an observed point pattern by considering the empirical

estimate of the K-function, which is a second-order summary function. The K-function has already been defined for point processes on Euclidean spaces, spheres, and linear networks, and in Paper A a K-function for point processes on the product space between a Euclidean space and a sphere is presented. This new K-function is used in Paper A for a preliminary analysis of the pyramidal cell data. The analysis is continued in Paper B, where modelling the location of the pyramidal cells is in focus. Specifically, a hierarchical model that exhibit cylindrical clusters and allow for complicated interaction structures is proposed.

In the literature, only a handful of models have been discussed for point processes on linear networks and of these it seems that only the Poisson process has been applied for modelling data. In Paper C, an extension of the conditional intensity function for point processes on the time line is used for constructing point processes on directed acyclic linear networks, and it is discussed how such a model can be fitted. Further, in the search for a suitable model for the spine data, a Cox process on linear networks is presented (and fitted) in Paper D. Further in Paper D, new empirical second-order summary functions are introduced for linear networks for the purpose of model checking.

Resumé

Denne afhandling omhandler punktprocesser, som bruges til at modellere en bestemt type af data, hvor hvert datapunkt beskriver placeringen af et objekt eller tidspunktet for en hændelse. Sådanne data kaldes punktmønstre og opstår inden for en bred vifte af fagområder som for eksempel skovbrug, astronomi, neurovidenskab, kriminologi, eller seismologi, hvor datapunkterne kan beskrive placering af træer i en skov, tidspunkt og position af meteorer, placering af neuroner i hjernen, forbrydelser i en by, eller tidspunkt for jordskælv. Punktmønstre bliver således observeres på forskellige typer af rum, som en-, to-, og tre-dimensionelle Euklidiske rum, sfærer og lineære netværk. Afhængig af hvilken type rum der betragtes - og alt efter hvordan afstande måles - findes der forskellige værktøjer og modeller til at analysere og beskrive punktmønstre. Størstedelen af den eksisterende litteratur om punktprocesser vedrører Euklidiske rum, men nogle fremskridt er dog også blevet gjort for punktprocesser på lineære netværk eller sfærer. I nogle tilfælde kan teori udviklet for punktprocesser på Euklidiske rum let udvides eller modificeres til andre typer af rum, men i andre tilfælde er sådan en udvidelse ikke ligetil eller slet ikke mulig. Denne afhandling vedrører modellering samt inferens for punktmønstre både på lineære netværk, Euklidiske rum, og produktrummet mellem et Euklidisk rum og en sfære.

Afhandlingen starter ud med en kort introduktion til relevante koncepter og metoder med det formål at give læseren mulighed for at forstå det videnskabelige bidrag af artikel A–D set i forhold til eksisterende litteratur. I introduktionen gennemgås eksisterende teori for punktprocesser på Euklidiske rum, sfærer eller lineære netværk og der præsenteres to typer af punktmønstre af særlig interesse for denne afhandling: punktmønstre på produktrummet mellem det tre-dimensionelle Euklidiske rum og den to-dimensionelle sfære, som beskriver placering og orientering af pyramideceller i et udsnit af en menneskehjerne, samt punktmønstre på lineære netværk, som beskriver placeringen af spines på et dendrittræ fra en museneuron.

Andenordens momentegenskaberne for en punktproces beskriver hvordan par af punkter fra punktprocessen interagerer. For at afdække og klassificere sådanne interaktioner i et observeret punktmønster betragtes ofte et

empirisk estimat af K-funktionen. Denne funktion er blevet præsenteret for punktprocesser på Euklidiske rum, sfærer, og lineære netværk, og i artikel A introduceres en K-funktion for punktprocesser på produktrummet mellem et Euklidisk rum og en sfære. Denne nye K-funktion bruges i artikel A til en indledende analyse af datasættet vedrørende pyramidecellers placering samt orientering. Analysen fortsættes i artikel B, hvor der fokuseres på at modellere placeringen af pyramidecellerne. Specifikt foreslås en hierarkisk model som udviser cylindriske klyngedannelser og som tillader komplicerede interaktionsstrukturer.

I litteraturen beskrives kun en lille håndful af modeller for punktprocesser på lineære netværk, og af disse er det tilsyneladende kun Poisson-processen som hidtil er blevet brugt til at modellere data. I artikel C bruges en udvidelse af den betingede intensitetsfunktion for punktprocesser på tidslinjen til at konstruere punktprocesser på acykliske lineære netværk med retninger, og det diskuteres hvorledes sådanne modeller tilpasses data. Desuden, i søgen efter en passende model til at beskrive spine-datasættet, bliver der i artikel D præsenteret og tilpasset en Cox-proces for lineære netværk. I artikel D introduceres der ydermere nye empiriske opsummerende funktioner for punktmønstre på lineære netværk med henblik på modeltjek.

Contents

Abstract iii

Resumé v

Preface xi

I Background 1

Introduction 3
- 1 Statistics for point processes on different spaces 3
  - 1.1 Intensity function 4
  - 1.2 Pair correlation function 5
  - 1.3 K-functions 5
  - 1.4 Other second-order summary functions 6
- 2 Point process models 7
  - 2.1 Poisson processes 7
  - 2.2 Cox processes 8
  - 2.3 Models on linear networks 8
  - 2.4 Models on the time line specified by a conditional intensity function 9
- 3 Examples of point pattern data sets 11
  - 3.1 The pyramidal cell data 11
  - 3.2 The spine data 12
- References 12

II Papers 17

A Structured space-sphere point processes and K-functions 19
- 1 Introduction 21
- 2 Preliminaries 24

2.1 Setting 24
2.2 Examples 26
3 The space-sphere K-function 28
3.1 Definition 28
4 Separability 30
4.1 First order separability 30
4.2 Second order separability 31
4.3 Assuming both SOIRS and first and second order separability 32
5 Estimation of K-functions 33
6 Data examples 34
6.1 Fireball locations over time 34
6.2 Location and orientation of pyramidal neurons 34
6.3 Simulation study 40
7 Additional comments 42
References 46

B Modelling columnarity of pyramidal cells in the human cerebral cortex **51**
1 Introduction and conclusions 53
1.1 Data 53
1.2 Background and purpose 54
1.3 Hierarchical point process models 54
1.4 Model fitting 59
1.5 Outline 59
2 Preliminaries 59
2.1 Moments 60
2.2 Functional summaries 60
2.3 Minimum contrast estimation 62
3 Complete spatial randomness 62
4 The degenerate Poisson line cluster point process 65
5 A generalisation of the degenerate PLCPP 67
5.1 A determinantal point process model for the centre points 68
5.2 A Markov random field model for the z-coordinates . . 69
References 75

C Point processes on directed linear networks **79**
1 Introduction 81
2 Point processes on directed linear networks 82
2.1 Directed linear networks 82
2.2 Directed paths and partial orders 83
2.3 Point patterns and point processes on linear networks . 85

2.4 Conditional intensity functions on directed linear networks . . . 86
2.5 Marks . . . 87
3 Likelihood function . . . 87
4 Simulation . . . 88
5 Residual analysis . . . 90
6 Models . . . 92
6.1 Poisson process . . . 93
6.2 Hawkes process . . . 93
6.3 Non-linear Hawkes process . . . 94
6.4 Self-correcting process . . . 95
6.5 Marked models . . . 95
7 Data Analysis . . . 97
7.1 Simulated data . . . 97
7.2 Dendrite data . . . 101
8 Extensions and future research . . . 104
References . . . 106

D Modelling spine locations on dendrite trees using inhomogeneous Cox point processes 109
1 Introduction . . . 111
2 Dendritic spine data . . . 114
3 Point processes on linear networks . . . 115
3.1 Summary functions for first and second-order moment properties . . . 115
3.2 New empirical summary functions . . . 116
4 Modelling spine locations . . . 117
4.1 Fitting a Poisson process model . . . 117
4.2 Fitting a Cox process model . . . 120
5 Discussion . . . 125
A Expressions for K . . . 126
B Simulation study concerning estimation procedure . . . 127
B.1 Second order composite likelihood . . . 127
B.2 Simulation study . . . 128
C Analysis of spine locations . . . 136
References . . . 144

Preface

This thesis contains scientific work conducted by me (and collaborators) during my PhD study at Department of Mathematics, Aalborg University, which was supported by Centre for Stochastic Geometry and Advanced Bioimaging (funded by the Villum foundation). The work concerns the modelling and inference of point processes on the Euclidean space, linear networks or on the product space between the Euclidean space and the sphere. The thesis consists of two parts: Part I is a short introduction to relevant theory, state-of-the-art as well as the thesis' scientific contribution, and Part II constitutes the main part of the thesis containing the four research papers:

> J. Møller, H. S. Christensen, F. Cuevas-Pacheco, and A. D. Christoffersen. Structured space-sphere point processes and K-functions, To appear in *Methodology and Computing in Applied Probability*.
>
> A. D. Christoffersen, J. Møller, and H. S. Christensen. Modelling columnarity of pyramidal cells in the human cerebral cortex. Submitted to *Journal of the Royal Statistical Society Series C (Applied Statistics)*.
>
> J. G. Rasmussen and H. S. Christensen. Point processes on directed linear networks. Submitted to *Methodology and Computing in Applied Probability*.
>
> H. S. Christensen and J. Møller. Modelling spine locations on dendrite trees using inhomogeneous Cox point processes. Submitted to *Biometrics*.

The papers are included in a version similar to the journal/submitted version (with only minor changes such as page numbering and layout), and therefore there may be inconsistency in the notation used across papers and some material is presented multiple times. On the other hand, each paper is fully self-contained and may be read independently of the others.

I wish to thank my supervisor Jesper Møller who have guided me through my PhD study and whose dedication has been a source of inspiration. I would like to thank my co-supervisor Jakob Gulddahl Rasmussen for the collaboration and for inviting me to contribute to his scientific idea. I also wish to thank Ethan Anderes for hosting me during my stay in Davis.

A special thanks to my boyfriend and partner in crime, Andreas, for lending me an endless supply of love, support, and stress-relieving hugs. I would also like to thank my fellow PhD students for our social gatherings and in particular my office mate Kristian with whom I have had many good conversations. You all helped me getting through my three years as a PhD student.

Heidi S. Christensen
Aalborg University, August 30, 2019

Part I

Background

Introduction

A *point pattern* is a collection of coordinates or points within a chosen region; this region can have many forms, e.g. a subset of the time line, a linear network, or a three-dimensional box, and the points can describe the location of any chosen event or object observed within that region. Thus, point pattern data arise in an endless number of applications for example as the occurrence of earthquakes over time, the location of spines on a dendrite tree, or the location of neurons in a section of the brain. In spatial statistics we analyse such point patterns to get a grasp of the underlying random mechanism that produces the events/objects; this mechanism is referred to as a *point process*. A point process may produce point patterns, where the points tend to fall in clusters, repel each other, or fall completely independent of each other with no sort of interaction between them. Detecting and modelling such behaviours is a central part of a point pattern analysis.

The following introduction serves as a background for understanding and putting into perspective the contribution of this thesis. The introduction is kept brief and non-technical, while more details are given in each of the subsequent papers. Notation used in the introduction and in Papers A–D may to some extent be different, but each paper is fully self-contained.

1 Statistics for point processes on different spaces

Briefly speaking, a *point process X on S* is a random countable subset of the space S, and the realisations of X are called *point patterns*; for a more technical introduction to point processes see e.g. Daley and Vere-Jones (2003). Most commonly, S is a subset of the d-dimensional Euclidean space, $\mathbb{R}^d$, where distance between two points is measured by the Euclidean distance. However, other spaces and distance metrics may be of interest such as the k-dimensional sphere $\mathbb{S}^k = \{u \in \mathbb{R}^{k+1} : \|u\| = 1\}$ along with the great circle distance or a linear network L along with the shortest path distance; a *linear network* is a finite union of line segments only intersecting at their end points. Depending on the character of S and the associated distance metric, different tools and models are available for analysing point patterns on S. Most existing point process theory has been developed for Euclidean spaces, while only few advances have been made for point processes on linear networks and the sphere. In this thesis point processes on two non-Euclidean spaces S are considered: a linear network and the product space $\mathbb{R}^d \times \mathbb{S}^k$.

In Papers A and B, we consider datasets concerning the location and orientation of pyramidal cells in a section of a human brain, where the cell locations are given by three-dimensional coordinates and the orientations by unit vectors. The data can thus be viewed as a point pattern on (a subset of) $\mathbb{R}^3 \times \mathbb{S}^2$; more details are given in Section 3.1. In Papers C and D, point patterns on linear networks are considered. There each of the linear networks is an approximation of an apical dendrite of a mouse neuron, and the points describe the locations of spines along the network; more details are given in Section 3.2.

1.1 Intensity function

A central part of analysing an observed point pattern is to fit a point process model and next to check whether the fitted model adequately describe the point pattern. For this, so-called *summary functions* and their empirical estimates are vital tools. A summary function summarises certain characteristics of a point process; one such function is the *intensity function*, whose empirical estimate can be used to detect the presence and form of inhomogeneity in an observed point pattern.

If the intensity function $\rho : S \to [0, \infty)$ of a point process X on S exists, it describes the first-order moment properties of X. That is,

$$\mathrm{E}\, n(X \cap A) = \int_A \rho(u) \,\mathrm{d}_S u < \infty,$$

where $n(X \cap A)$ is the number of points from X falling in $A \subseteq S$, and d_S denotes integration with respect to a suitable measure on S, e.g. the Lebesgue measure if $S \subseteq \mathbb{R}^d$, one-dimensional arc length if S is a linear network, and surface measure if $S \subseteq \mathbb{S}^k$. Intuitively, $\rho(u) \,\mathrm{d}_S u$ can be interpreted as the probability of observing a point from X in an infinitesimal region of size $\mathrm{d}_S u$ that contains u. If the intensity function $\rho(\cdot) = \rho$ is constant, X is said to be *(first-order) homogeneous* with intensity ρ, and then ρ is the expected number of points per unit area/volume/length/surface area; otherwise the point process is said to be *inhomogeneous*.

1.2 Pair correlation function

To describe the second-order moment properties of X we can consider the *pair correlation function* $g : S \times S \to [0, \infty)$, given that it exists. That is,

$$\mathrm{E}\{n(X \cap A)n(X \cap B)\} = \int_A \int_B g(u,v)\rho(u)\rho(v)\,\mathrm{d}_S u\,\mathrm{d}_S v < \infty$$

for disjoint $A, B \subset S$. We can interpret $g(u,v)\rho(u)\rho(v)\,\mathrm{d}_S u\,\mathrm{d}_S v$ as the probability of simultaneously observing a point from X in each of two infinitesimal regions of size $\mathrm{d}_S u$ and $\mathrm{d}_S v$ containing u and v, respectively. Alternatively, we can heuristically understand $g(u,v)$ as the probability of simultaneously observing points from X in these infinitesimal regions relative to the probability of such points occurring independently of each other. If the pair correlation function $g(u,v) = g(d_S(u,v))$ only depends on the distance between u and v, we say that it is isotropic; here d_S is a distance metric related to S, and specifically (unless otherwise stated) we let in the following d_S be the Euclidean distance on $\mathbb{R}^d$ when $S \subseteq \mathbb{R}^d$, the great circle distance on $\mathbb{S}^k$ when $S \subseteq \mathbb{S}^k$, and the shortest path distance when S is a linear network. If g is isotropic, then $g(r) > 1$ indicates that a pair of points with distance r is more likely to occur in X than for a point process with independent points, and we say that X posses clustering/aggregation at scale r. Similarly, $g(r) < 1$ indicates regularity/repulsion at scale r.

1.3 K-functions

Another second-order summary function is the *K-function*, which was presented for point processes on $\mathbb{R}^d$ first by Ripley (1976) in the homogeneous case and later extended by Baddeley et al. (2000) to the inhomogeneous case. Lawrence et al. (2016) and Møller and Rubak (2016) present an analogue of the K-function for point processes on $\mathbb{S}^k$. These two K-functions, in the following referred to as the spatial and spherical K-function and denoted by $K_{\mathbb{R}^d}$ and $K_{\mathbb{S}^k}$, respectively, can be defined by

$$K_S(r) = \frac{1}{\nu_S(B)} \mathrm{E} \sum_{u \in X \cap B} \sum_{v \in X \setminus \{u\}} \frac{\mathbb{I}(d_S(u,v) \le r)}{\rho(u)\rho(v)}, \qquad S = \mathbb{R}^d, \mathbb{S}^k, \tag{1}$$

under the assumption that the right-hand side does not depend on the choice of $B \subseteq S$ for $0 < \nu_S(B) < \infty$, where $\nu_S(\cdot)$ is the Lebesgue measure on $\mathbb{R}^d$ for $S = \mathbb{R}^d$ and the surface measure on $\mathbb{S}^k$ for $S = \mathbb{S}^k$. Further, $\mathbb{I}(\cdot)$ denotes the indicator function. If the distribution of X is invariant under translations when $S = \mathbb{R}^d$ (in which case X is said to be *stationary*) or invariant under rotations when $S = \mathbb{S}^k$, then X is homogeneous and $\rho K_S(r)$ can be interpreted as the expected number of further points from X occurring within distance r of a typical point of X (when measuring distance with respect to d_S).

For homogeneous point processes on linear networks, Okabe and Yamada (2001) suggest a K-function directly adapted from the spatial K-function. However, their network K-function depends on the geometry of the network and thus cannot be compared for point processes on different networks. Later, a modified version of the network K-function, K_L, was introduced by Ang et al. (2012), which removes the dependency on the network geometry and extends the definition to inhomogeneous point process (fulfilling certain assumptions). For point processes on $\mathbb{R}^d \times \mathbb{S}^k$, there exists no analogue of the K-function. In Paper A, we extend the notion of a K-function to point processes on $S \subseteq \mathbb{R}^d \times \mathbb{S}^k$ and show that if the point process possesses a certain separable structure, then this *space-sphere K-function* is proportional to the product of the spatial and spherical K-functions.

For the spatial, spherical and network K-functions, the assumptions required to define them are fulfilled whenever the pair correlation function is isotropic with respect to $d_{\mathbb{R}^d}$, $d_{\mathbb{S}^k}$, and d_L, respectively. In that case, the pair correlation function and the K-function are related by

$$K_{\mathbb{R}^d}(r) = \frac{2\pi^{d/2}}{\Gamma(d/2)} \int_0^r t^{d-1} g(t)\,\mathrm{d}t, \tag{2}$$

cf. Baddeley et al. (2000),

$$K_{\mathbb{S}^2}(r) = 2\pi \int_0^r g(t)\sin(t)\,\mathrm{d}t, \tag{3}$$

cf. Lawrence et al. (2016), and

$$K_L(r) = \int_0^r g(t)\,\mathrm{d}t, \tag{4}$$

cf. Ang et al. (2012). Thus, the K-function and the pair correlation function are closely related.

1.4 Other second-order summary functions

Other commonly used second-order summary functions for point processes on $\mathbb{R}^d$, include the F-, G-, and J-functions defined by van Lieshout and Baddeley (1996) in the stationary case and by van Lieshout (2011) in the non-stationary case. Shortly, for a stationary point process $X \subset \mathbb{R}^d$, $F(r)$ is the probability of observing a point from X within distance $r > 0$ of a fixed point in $\mathbb{R}^d$, $G(r)$ can be interpreted as the probability of observing a point from $X \backslash \{x\}$ with distance r of a fixed point $x \in X$, and $J(r) = [1 - G(r)]/[1 - F(r)]$ for $F(r) < 1$.

If the isotropic pair correlation function, the K-function, or another relevant summary function can be expressed on closed form for the point process model of interest, it and its empirical estimates can be used for fitting the model to an observed point pattern. One option is to perform *minimum contrast estimation* (see for example Guan, 2009; Diggle, 2014), where the estimates of the model parameters are chosen such that the discrepancy between the curve for the theoretical summary function and the empirical summary function for the data is as small as possible. After a model has been fitted, the next natural step is to perform a model check, that is, to investigate how well the fitted model actually describes the data. One possibility is to use a *global rank envelope* procedure Myllymäki et al. (2017), where the shape of the empirical summary function for the data is compared to that of point patterns simulated from the fitted/hypothesised model; this procedure provides both a confidence region for the empirical summary function under the hypothesised model as well as a p-interval encompassing the most liberal and conservative p-values of the associated test. However, if a summary function first has been used for model fitting it should not be used for model checking too. Moreover, if the pair correlation function or the K-function has been used for model fitting, neither should be used for model checking due to their close relation, see (2)–(4). For point processes on $\mathbb{R}^d$, we can then use the F-, G-, and J-functions, but for point processes on $\mathbb{S}^k$ or on a linear network there are no further second-order summary functions available complicating a rigorous model check.

In Paper D, we suggest new empirical summary functions useful for model checking when S is a linear network. These are adaptations of the F-, G-, and J-functions for point processes on $\mathbb{R}^d$ to linear networks.

2 Point process models

2.1 Poisson processes

One of the simplest but yet most important point process models is the *Poisson process*, which is easily defined both on $\mathbb{R}^d$, $\mathbb{S}^k$, $\mathbb{R}^d \times \mathbb{S}^k$, and on a linear network. This process is fully specified by its intensity function and possesses no sort of interaction between points. More precisely, X is a Poisson process on S if $n(X \cap A)$ follows a Poisson distribution for any bounded $A \subseteq S$, and if further the points in $X \cap A$ are independent and identically distributed with density proportional to ρ when conditioning on $n(X \cap A)$. For a Poisson process, the pair correlation function $g \equiv 1$, reflecting that the model does not posses repulsion or aggregation at any scale. Note that other processes may have the same pair correlation function (Baddeley and Silverman, 1984). The Poisson process is often to simple to describe real point pattern data, but can then serve as a reference mark when investigating whether a point pattern exhibits repulsion or aggregation. Additionally, the Poisson process is often used as a building stone for more interesting models such as the *Cox processes*.

2.2 Cox processes

The Cox processes were first introduced by Cox (1955) and constitute a flexible class of point processes often used for modelling clustered point patterns. The Cox process is an extension of the Poisson process, where the intensity function now is stochastic and determined by a non-negative *random field* Λ; that is, $\Lambda = \{\Lambda(u) : u \in S\}$ is a collection of non-negative random variables indexed by the elements in S. More precisely, X is said to be a *Cox process driven by* Λ, if X conditioned on Λ is a Poisson process with intensity function Λ. Then X has intensity function $\rho(u) = \mathrm{E}\Lambda(u)$ and pair correlation function $g(u,v) = \mathrm{E}[\Lambda(u)\Lambda(v)]/[\rho(u)\rho(v)]$. If $Z = \log(\Lambda)$ is a Gaussian random field, meaning that any finite collection of elements from Z follows a (multivariate) Gaussian distribution, X is called a *log Gaussian Cox process* (LGCP; see Møller et al., 1998). The distribution of Z (and thus X) is fully specified by its mean and covariance function. Often the covariance function is assumed to depend only on the distance between the points, in which case we say that the covariance function is isotropic; this property entails an isotropic pair correlation function. Depending on the choice of distance metric and the space we are considering, there are different valid isotropic covariance functions to choose from; see for example Gneiting (2013) when $S = \mathbb{S}^k$, Berg and Porcu (2017) when $S = \mathbb{R} \times \mathbb{S}^k$, and Anderes et al. (2017) when S is a linear networks.

2.3 Models on linear networks

Besides the Poisson and Cox processes, there exists a large number of point process models for describing both clustered and regular point patterns on the Euclidean space. However, for linear networks only few point process models have been considered in the literature: Baddeley et al. (2014) use the Poisson process to analyse spines on a dendrite tree, and Baddeley et al. (2017) discuss some point process constructions extended from the Euclidean space to linear networks; these constructions yield point processes in $\mathbb{R}^d$ with an isotropic pair correlation function (when measuring distance with respect to $d_{\mathbb{R}^d}$), but mimicking these constructions for a linear network L do in general not lead to point processes on L with an isotropic pair correlation function (when measuring distance with respect to d_L). Furthermore, their constructions lead to point processes on L which are either clustered or with the same first- and second-order moments as the Poisson process; none of the constructions lead to regular point processes. Finally, considering both the shortest path distance and the so-called resistance metric (a metric adapted from electrical network theory to linear networks), Anderes et al. (2017) present isotropic covariance functions on linear networks and briefly mention that these can be used to construct LGCPs (which in turn will have an isotropic pair correlation function).

In Paper D, we adapt a clustered Cox point process on $\mathbb{R}^d$ proposed by Lavancier and Møller (2016) to the case of a linear network. The model, which is not a LGCP, has an isotropic pair correlation function and is obtained by randomly removing points from a Poisson process with probabilities determined by a spatially correlated random field. Further, in Paper C we extend both clustered and repulsive models for point processes on $\mathbb{R}$ specified by a conditional intensity function to directed acyclic linear networks; more details are given in Section 2.4.

2.4 Models on the time line specified by a conditional intensity function

On the time line, $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$ (or alternatively the whole real line, $\mathbb{R}$), a point process can be constructed by specifying the so-called *conditional intensity function*. This construction exploits that $\mathbb{R}_+$ has an evolutionary character, meaning that we can talk about points from $\mathbb{R}_+$ coming before or after one another.

Let X be a point process on $S \subseteq \mathbb{R}_+$. Then the events/points of X can be strictly ordered, say $X = (t_1, t_2, \dots)$ such that $t_i < t_{i+1}$ for $i = 1, 2, \dots$, and we let X_{t_-} denote the events or *history* up to time t, that is, $X_{t_-} = \{t_i \in X : t_i < t\}$. Heuristically, the conditional intensity function (also called the hazard function) λ^* can be defined as

$$\lambda^*(t)dt = \mathrm{E}[n(X \cap [t, t+dt])|X_{t_-}],$$

where the $*$ indicates that $\lambda^*(t)$ depends on the history up to time t; for a more formal definition of the conditional intensity function see Daley and Vere-Jones (2003). Thus, $\lambda^*(t)dt$ can be interpreted as the expected number of points falling in an infinitesimal interval containing t and of length dt when conditioning on events occurring before time t. If $\lambda^*(t) = \lambda(t)$ does not depend on the history, the events occur independently of each other and the resulting process is a Poisson process with intensity function λ.

For a conditional intensity function $\lambda^*(\cdot) = \lambda^*(\cdot;\theta)$, depending on a parameter vector θ, and a point pattern x observed on a bounded interval $W \subseteq S$, the likelihood function is given by

$$\mathcal{L}(\theta; x) = \prod_{t \in x} \lambda^*(t;\theta) \exp\Big(- \int_W \lambda^*(s;\theta)ds \Big). \tag{5}$$

Thus, parameter estimation for a model specified by a conditional intensity function is simply a matter of maximising the likelihood. However, this can rarely be done analytically – one exception being the homogeneous Poisson process – and therefore numerical methods are usually needed for obtaining the maximum likelihood estimates.

One common model for clustered point processes is given by the conditional intensity function

$$\lambda^*(t) = \mu + \alpha \sum_{t_i \in X_{t_-}} \exp(t_i - t),$$

where $\mu, \alpha > 0$ are model parameters. This conditional intensity function increases with α whenever a new point occurs, whereafter it decreases exponentially towards the baseline intensity μ. This model is a special case of the *Hawkes* (or *self-exciting*) *process* (Hawkes, 1971a,b).

The *self-correcting process* introduced by Isham and Westcott (1979) is an example of a regular point process and is defined by the conditional intensity function

$$\lambda^*(t) = \exp\Big(\mu t - \sum_{t_i \in X_{t_-}} \alpha\Big)$$

for model parameters $\mu, \alpha > 0$. Note that the logarithm of the conditional intensity function increases linearly with time and with rate μ, while it decreases with α each time a point occurs.

Simulations from a fitted model are useful for model checking as discussed in Section 1, where an empirical summary function can be used for a global rank envelope procedure; the *inverse method* (see e.g. Daley and Vere-Jones, 2003) and *Ogata's modified thinning algorithm* (Ogata, 1981) are two algorithms for simulating point processes specified by a conditional intensity function. Another option for model checking, is to exploit that if $X = (t_1, t_2, \dots)$ follows a point process model specified by λ^*, then the transformation $\tilde{t}_i = \int_0^{t_i} \lambda^*(s)$ yields a point process $\tilde{X} = \{\tilde{t}_1, \tilde{t}_2, \dots\}$ that follows a unite-rate Poisson process (see e.g. Daley and Vere-Jones, 2003, Proposition 7.4.IV), entailing that the interevent times $\tilde{t}_{i+1} - \tilde{t}_i$, $i = 1, 2, \dots$, are independent and exponentially distributed with mean 1. Thus, for an observed point pattern $x = (t_1, \dots, t_n)$, we can perform model checking by investigating whether the interevent times of the transformed point pattern $\tilde{x} = (\tilde{t}_1, \dots, \tilde{t}_n)$ (calculated using the fitted conditional intensity function) are independent and exponentially distributed with mean 1. This sort of model checking is referred to as *residual analysis*.

Introducing directions to an acyclic linear network give the network an evolutionary character similar to the time line, and this enables us in Paper C to extend the definition of a conditional intensity function to apply for a directed acyclic linear network. Specifically, we define models inspired by the Hawkes and self-correcting processes, find an expression of the likelihood similar to (5), and adapt the inverse method, Ogata's modified thinning algorithm, and the residual analysis procedure to directed acyclic linear networks.

3 Examples of point pattern data sets

3.1 The pyramidal cell data

The pyramidal cell data briefly mentioned in Section 1 come from a small section of the primary motor cortex of a brain from a healthy (but deceased) human and contain information on the cells lying within this section. Specifically, two things are recorded for each pyramidal cell: a set of three-dimensional coordinates describing the *location* of its nucleolus and a unit vector pointing from its nucleolus toward its apical dendrite, giving a sense of its *orientation*. Thus, the data set can be viewed as a point pattern on $S \subset \mathbb{R}^3 \times \mathbb{S}^2$ with the locations in a subset of $\mathbb{R}^3$ and the orientations in $\mathbb{S}^2$.

The pyramidal cell data set are not only interesting as an example of a point pattern on $\mathbb{R}^3 \times \mathbb{S}^2$ but also for modelling, since finding a suitable model can give a better understanding of the structuring of neurons (of which 75% to 80% are pyramidal cells). A hypothesis, called the *minicolumn hypothesis*, suggests that neurons are organised in columns perpendicular to the pial surface of the brain. Biological studies have been performed to investigate the minicolumn hypothesis (see e.g. Lorente de Nó, 1938; Mountcastle, 1978; Buxhoeveden and Casanova, 2002) and it is believed that neurological diseases such as Alzheimers and schizophrenia can be linked with deviations from such a columnar structure. Only few statistical studies have been made to investigate the minicolumn hypothesis, see Skoglund et al. (2004), Cruz et al. (2005), Cruz et al. (2008), and Rafati et al. (2016). Of these, Rafati et al. (2016) is the only study considering neurons from a human brain (Skoglund et al., 2004; Cruz et al., 2005, 2008, consider either rat or monkey brains), and Rafati et al. (2016) further use tools from point process theory for the analysis. Specifically, using the *cylindrical K-function*, a summary function introduced by Møller et al. (2016) for detecting columnar structures, Rafati et al. (2016) find evidence of columnar structures. However, they do not find a suitable point process model for characterising the neuronal structuring in more details. Finding an appropriate model for the structuring of neurons in healthy humans can not only serve as evidence of or against the minicolumn hypothesis but also as a benchmark for detecting and classifying abnormal changes of the neuronal organisation in humans with neurological diseases.

In Paper A, the space-sphere K-function (also proposed in Paper A as discussed in Section1.3) is used to investigate whether the location and orientation of the pyramidal cells are independent. If this is the case it eases the modelling as the locations and orientations may be modelled separately. Further, we find in Paper A an inhomogeneous Poisson model that seem to fit the orientations adequately. Next, in Paper B we consider the locations only and try to model these with a number of point process models with columnar structures in $\mathbb{R}^3$. The main idea behind these models is to first find a suitable model for the point pattern projected onto a plane in $\mathbb{R}^2$ perpendicular to the pial surface, and second to build a model for the third dimension conditional on the planar point process.

3.2 The spine data

Spines are small protrusions on the dendrite trees of a neuron that among other help transmitting electrical signals to the cell body. If the dendrite tree is approximated by a linear network L, we can view the locations of spines as a point pattern on L. Previous studies analysing the distribution of spine locations as a point pattern include Jammalamadaka et al. (2013) and Baddeley et al. (2014), where dendrite trees extracted from rat neurons grown in vitro are considered. Specifically, Jammalamadaka et al. (2013) investigate the distribution of spine locations for a large number of neurons, and they conclude (by considering the network K-function) that a homogeneous Poisson model adequately describes the spine locations. In Baddeley et al. (2014) the focus is mostly on presenting new tools for *multitype point processes* on linear networks; that is, a point process where the points are classified into a number of types. For illustrating the use of these new tools, Baddeley et al. (2014) consider one of the datasets from Jammalamadaka et al. (2013), where each spine is classified according to its shape giving a multitype point pattern. Regardless of the spine type and considering both homogeneous and inhomogeneous Poisson processes, Baddeley et al. (2014) find some evidence that the spines are spatially clustered but they do not search for a suitable model. In conclusion, the only model that has been fitted to the spine data (and seemingly to any point pattern on a linear network) is a Poisson process.

In Paper D, we consider six point pattern datasets each describing the location of spines on a linear network that represents a dendrite tree. These dendrite trees are apical dendrites from mouse neurons grown in vivo. Jammalamadaka et al. (2013) mention that their results for in vitro grown neurons do most likely not hold for the in vivo setting. In Paper D we therefore set out to find an adequate point process model for the in vivo spine data. In that effort, we suggest a Cox process as briefly mentioned in Section 2.3.

The spine data is also shortly considered in Paper C to illustrate the concepts presented therein. Specifically, we introduce directions to the linear network representing the dendrite tree (yielding a directed acyclic linear network as the linear network forms a tree) and fit a model specified by an analogue to the conditional intensity function (as discussed in Section 2.4).

References

Anderes, E., Møller, J., and Rasmussen, J. G. (2017). Isotropic covariance functions on graphs and their edges. Available at arXiv:1710.01295.

Ang, Q. W., Baddeley, A. J., and Nair, G. (2012). Geometrically corrected second order analysis of events on a linear network, with applications to ecology and criminology. *Scandinavian Journal of Statistics*, 39:591–617.

Baddeley, A. J., Jammalamadaka, A., and Nair, G. (2014). Multitype point process analysis of spines on the dendrite network of a neuron. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 63:673–694.

Baddeley, A. J., Møller, J., and Waagepetersen, R. P. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54:329–350.

Baddeley, A. J., Nair, G., Rakshit, S., and McSwiggan, G. (2017). "Stationary" point processes are uncommon on linear networks. *Stat*, 6:68–78.

Baddeley, A. J. and Silverman, B. (1984). A cautionary example on the use of second-order methods for analyzing point patterns. *Biometrics*, 40:1089–1094.

Berg, C. and Porcu, E. (2017). From Schoenberg coefficients to Schoenberg functions. *Constructive Approximation*, 45:217–241.

Buxhoeveden, D. P. and Casanova, M. F. (2002). The minicolumn hypothesis in neuroscience. *Brain*, 125:935–951.

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

Cruz, L., Buldyrev, S. V., Peng, S., Roe, D. L., Urbanc, B., Stanley, H. E., and Rosene, D. L. (2005). A statistically based density map method for identification and quantification of regional differences in microcolumnarity in the monkey brain. *Journal of Neuroscience Methods*, 141:321–332.

Cruz, L., Urbanc, B., Inglis, A., Rosene, D. L., and Stanley, H. E. (2008). Generating a model of the three-dimensional spatial distribution of neurons using density maps. *NeuroImage*, 40:1105–1115.

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

Diggle, P. J. (2014). *Statistical Analysis of Spatial and Spatio-temporal Point Patterns*. Chapman & Hall/CRC Press, Boca Raton, Florida.

Gneiting, T. (2013). Stricly and non-strictly positive definite functions on spheres. *Bernoulli*, 19:1327–1349.

Guan, Y. (2009). A minimum contrast estimation procedure for estimating the second-order parameters of inhomogeneous spatial point processes. *Statistics and Its Interface*, 2:91–99.

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

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

Isham, V. and Westcott, M. (1979). A self-correcting point process. *Stochastic Processes and their Applications*, 8:335–347.

Jammalamadaka, A., Banerjee, S., Manjunath, B. S., and Kosik, K. S. (2013). Statistical analysis of dendritic spine distributions in rat hippocampal cultures. *BMC Bioinformatics*, 14:287.

Lavancier, F. and Møller, J. (2016). Modelling aggregation on the large scale and regularity on the small scale in spatial point pattern datasets. *Scandinavian Journal of Statistics*, 43:587–609.

Lawrence, T., Baddeley, A. J., Milne, R. K., and Nair, G. (2016). Point pattern analysis on a region of a sphere. *Stat*, 5:144–157.

van Lieshout, M. N. M. (2011). A J-function for inhomogeneous point processes. *Statistica Neerlandica*, 65:183–201.

van Lieshout, M. N. M. and Baddeley, A. J. (1996). A nonparametric measure of spatial interaction in point patterns. *Statistica Neerlandica*, 50:344–361.

Lorente de Nó, R. (1938). The cerebral cortex: Architecture, intracortical connections, motor projections. In Fulton, J. F., editor, *Physiology of the Nervous System*, pages 274–301. Oxford University Press, Oxford.

Møller, J. and Rubak, E. (2016). Functional summary statistics for point processes on the sphere with an application to determinantal point processes. *Spatial Statistics*, 18:4–23.

Møller, J., Safavimanesh, F., and Rasmussen, J. G. (2016). The cylindrical K-function and Poisson line cluster point processes. *Biometrika*, 104:937–954.

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

Mountcastle, V. B. (1978). *The Mindful Brain: Cortical Organization and the Group-selective Theory of Higher Brain Function*. MIT Press, Cambridge.

Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H., and Hahn, U. (2017). Global envelope tests for spatial processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79:381–404.

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

Okabe, A. and Yamada, I. (2001). The K-function method on a network and its computational implementation. *Geographical analysis*, 33:271–290.

Rafati, A. H., Safavimanesh, F., Dorph-Petersen, K., Rasmussen, J. G., Møller, J., and Nyengaard, J. R. (2016). Detection and spatial characterization of minicolumnarity in the human cerebral cortex. *Journal of Microscopy*, 261:115–126.

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability*, 13:255–266.

Skoglund, T. S., Pascher, R., and Berthold, C. H. (2004). Aspects of the organization of neurons and dendritic bundles in primary somatosensory cortex of the rat. *Neuroscience Research*, 50:189–198.

Part II

Papers

Paper A

Structured space-sphere point processes and *K*-functions

Jesper Møller, Heidi S. Christensen, Francisco Cuevas-Pacheco,
and Andreas D. Christoffersen

The paper is to appear in
Methodology and Computing in Applied Probability.

© Springer Science + Business Media, LLC, part of Springer Nature 2019
The layout has been revised.

Abstract

This paper concerns space-sphere point processes, that is, point processes on the product space of $\mathbb{R}^d$ *(the d-dimensional Euclidean space) and* $\mathbb{S}^k$ *(the k-dimensional sphere). We consider specific classes of models for space-sphere point processes, which are adaptations of existing models for either spherical or spatial point processes. For model checking or fitting, we present the space-sphere K-function which is a natural extension of the inhomogeneous K-function for point processes on* $\mathbb{R}^d$ *to the case of space-sphere point processes. Under the assumption that the intensity and pair correlation function both have a certain separable structure, the space-sphere K-function is shown to be proportional to the product of the inhomogeneous spatial and spherical K-functions. For the presented space-sphere point process models, we discuss cases where such a separable structure can be obtained. The usefulness of the space-sphere K-function is illustrated for real and simulated datasets with varying dimensions d and k.*

1 Introduction

Occasionally point processes arise on more complicated spaces than the usual space $\mathbb{R}^d$, the d-dimensional Euclidean space, as for spatio-temporal point processes, spherical point processes or point processes on networks (see e.g. Dvořák and Prokešová, 2016; Lawrence et al., 2016; Møller and Rubak, 2016; Baddeley et al., 2017, and the references therein for details on such point processes). In this paper we consider *space-sphere point processes* that live on the product space $S = \mathbb{R}^d \times \mathbb{S}^k$, where $\mathbb{S}^k = \{u \in \mathbb{R}^{k+1} : \|u\|_{k+1} = 1\}$ is the k-dimensional unit sphere, $\|\cdot\|_k$ denotes the usual distance in $\mathbb{R}^k$, and $d, k \in \{1, 2, \ldots\}$. For each point $(y, u) \in S$ belonging to a given space-sphere point process, we call y its spatial component and u its spherical component. Assuming local finiteness of a space-sphere point process, the spatial components constitute a locally finite point process in $\mathbb{R}^d$, but the spherical components do not necessarily form a finite point process on $\mathbb{S}^k$. However, in practice the spatial components are only considered within a bounded window $W \subset \mathbb{R}^d$, and the associated spherical components do constitute a finite point process.

One example is the data shown in Figure A.1 that consists of the location and orientation of a number of pyramidal neurons found in a small area of a healthy human's primary motor cortex. More precisely, the locations are three-dimensional coordinates each describing the placement of a pyramidal neuron's nucleolus, and the orientations are unit vectors pointing from a neuron's nucleolus toward its apical dendrite. These data can be considered as a realisation of a space-sphere point process with dimensions $d = 3$ and $k = 2$, where the spatial components describe the nucleolus locations and the spherical components are the orientations. How neurons (of which around 75% to 80 % are pyramidal neurons) are arranged have been widely discussed in the literature. Specifically, it is hypothesised that neurons are arranged in columns perpendicular to the pial surface of the brain. This hypothesis, referred to as the minicolumn hypothesis, have been studied for more than half a century (see e.g. Lorente de Nó, 1938; Mountcastle, 1978; Buxhoeveden and Casanova, 2002), and it is believed that deviation from such a columnar structure is linked with neurological diseases such as Alzheimers and schizophrenia.

Another example is the time and geographic location of fireballs, which are bright meteors reaching a visual magnitude of -3 or brighter. They are continually recorded by U.S. Government sensors and made available at `http://neo.jpl.nasa.gov/fireballs/`. We can consider fireball events as a space-sphere point process with dimensions $d = 1$ and $k = 2$, where the time and locations are the spatial and spherical components, respectively. Figure A.2 shows the location of fireballs on the globe (identified with the unit sphere) observed over a time period of about 606 weeks.

The paper is organised as follows. In Section 2, we define concepts related to space-sphere point processes and give some natural examples of such processes. In Section 3, we define the space-sphere K-function, a functional summary statistic which is analogue to the space-time K-function when $d = 2$ and $\mathbb{S}^k$ is replaced by the time axis (Diggle et al., 1995; Gabriel and Diggle, 2009; Møller and Ghorbani, 2012). The space-sphere K-function is defined in terms of the pair correlation function which is assumed to have a certain stationary form. In the case where both the intensity and pair correlation function have a specific separable structure discussed in Section 4, the space-sphere K-function is shown to be proportional to the product of the spatial K-function (Baddeley et al., 2000) and the spherical K-function (Lawrence et al., 2016; Møller and Rubak, 2016). Further, an unbiased estimate is given in Section 5. In Section 6, the usefulness of the space-sphere K-function is illustrated for the fireball and neuron data as well as for simulated data, and it is e.g. seen how the K-function may be used to test for independence between the spatial and spherical components.

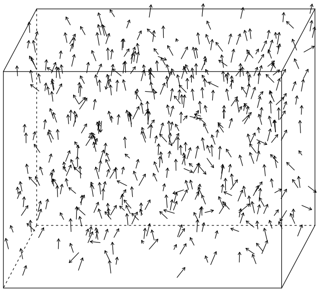


Figure A.1: Location and orientation of pyramidal neurons in a small section of a human brain. For details, see Section 6.2.

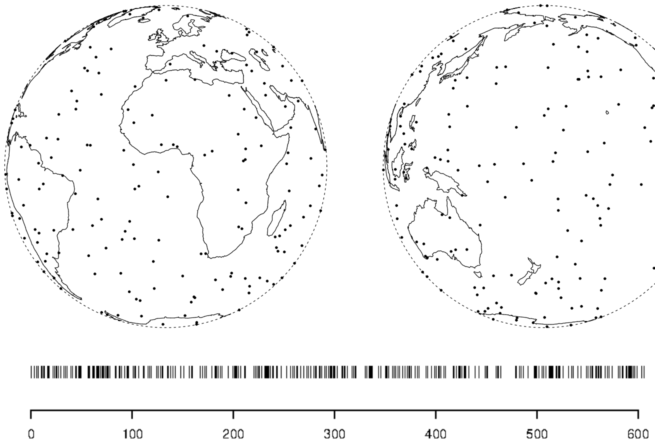


Figure A.2: Top: orthographic projection of the fireball locations. Bottom: time of fireball events measured in weeks. For details, see Section 6.1.

2 Preliminaries

2.1 Setting

Throughout this paper we consider the following setting.

Equip $\mathbb{R}^d$ with the Lebesgue measure $|A| = \int_A \mathrm{d}y$ and $\mathbb{S}^k$ with Lebesgue/surface measure $\nu(B)$, where $A \subseteq \mathbb{R}^d$ and $B \subseteq \mathbb{S}^k$ are Borel sets. Thus, the product space $S = \mathbb{R}^d \times \mathbb{S}^k$ is equipped with Lebesgue measure μ given by $\mu(A \times B) = |A|\nu(B)$.

Let X be a simple locally finite point process on S, that is, we can view X as a random subset of S such that the restriction $X_B = X \cap B$ of X to any bounded set $B \subset S$ is finite. We call X a *space-sphere point process*, and assume that it has intensity function ρ with respect to μ and pair correlation function g with respect to the product measure $\mu \otimes \mu$. That is, for any Borel function $h : S \mapsto [0, \infty)$,

$$\mathrm{E}\Big\{ \sum_{(y_i,u_i)\in X} h(y_i, u_i) \Big\} = \int h(y,u)\rho(y,u)\,\mathrm{d}\mu(y,u), \tag{A.1}$$

provided this integral is finite. We say that X is *(first order) homogeneous* if ρ is a constant function. Furthermore, for any Borel function $k : S \times S \mapsto [0, \infty)$,

$$\begin{aligned} &\mathrm{E}\Big\{ \sum_{(y_i,u_i),(y_j,u_j)\in X}^{\neq} k(y_i,u_i,y_j,u_j) \Big\} \qquad \text{(A.2)} \\ &= \iint k(y_1,u_1,y_2,u_2)\rho(y_1,u_1)\rho(y_2,u_2)g(y_1,u_1,y_2,u_2)\,\mathrm{d}\mu(y_1,u_1)\,\mathrm{d}\mu(y_2,u_2), \end{aligned}$$

provided this double integral is finite. Here, we set $g(y_1,u_1,y_2,u_2) = 0$ if $\rho(y_1,u_1)\rho(y_2,u_2) = 0$, and $\sum^{\neq}_{(y_i,u_i),(y_j,u_j)\in X}$ means that we sum over pairs of distinct points $(y_i,u_i),(y_j,u_j) \in X$.

The functions ρ and g are unique except for null sets with respect to μ and $\mu \otimes \mu$, respectively. For ease of presentation, we ignore null sets in the following. Note that $g(y_1,u_1,y_2,u_2) = g(y_2,u_2,y_1,u_1)$ is symmetric on $S \times S$. We say that X is *stationary in space* if its distribution is invariant under translations of its spatial components; this implies that $\rho(y,u)$ depends only on u, and $g(y_1,u_1,y_2,u_2)$ depends only on (y_1,y_2) through the difference $y_1 - y_2$. If the distribution of X is invariant under rotations (about the origin in $\mathbb{R}^d$) of its spatial components, we say that X is *isotropic in space*. Stationarity and isotropy in space imply that $g(y_1,u_1,y_2,u_2)$ depends only on (y_1,y_2) through the distance $\|y_1 - y_2\|_d$. We say that X is *isotropic on the sphere* if its distribution is invariant under rotations (on $\mathbb{S}^k$) of its spherical components; this implies that $g(y_1,u_1,y_2,u_2)$ depends only on (u_1,u_2) through the geodesic (great circle/shortest path) distance $d(u_1, u_2)$ on $\mathbb{S}^k$. If X is stationary in space and isotropic on the sphere, then ρ is constant and

$$g(y_1, u_1, y_2, u_2) = g_0\{y_1 - y_2, d(u_1, u_2)\}, \qquad y_1, y_2 \in \mathbb{R}^d,\ u_1, u_2 \in \mathbb{S}^k, \quad \text{(A.3)}$$

depends only on (y_1, y_2) through $y_1 - y_2$ and on (u_1, u_2) through $d(u_1, u_2)$ (this property is studied further in Section 3). If it is furthermore assumed that X is isotropic in space, then

$$g(y_1, u_1, y_2, u_2) = g_*\{\|y_1 - y_2\|_d, d(u_1, u_2)\}, \qquad y_1, y_2 \in \mathbb{R}^d,\ u_1, u_2 \in \mathbb{S}^k,$$

depends only on (y_1, y_2) through $\|y_1 - y_2\|_d$ and on (u_1, u_2) through $d(u_1, u_2)$.

The spatial components of X constitute a usual *spatial point process* $Y = \{y : (y, u) \in X\}$, which is locally finite, whereas the spherical components constitute a *point process on the sphere* $U = \{u : (y, u) \in X\}$ that may be infinite on the compact set $\mathbb{S}^k$. Let $W \subset \mathbb{R}^d$ be a bounded Borel set, which we may think of as a window where the spatial components $Y_W = Y \cap W$ are observed. As X is locally finite, the spherical components associated with Y_W constitute a finite point process $U_W = \{u : (y, u) \in X, y \in W\}$ on $\mathbb{S}^k$. Let $N = N(W)$ denote the cardinality of Y_W. To avoid trivial and undesirable cases, we assume that $|W| > 0$ and that the following inequalities hold:

$$0 < \mathrm{E}(N) < \infty \qquad \text{(A.4)}$$

and

$$0 < \mathrm{E}\{N(N-1)\} < \infty, \qquad \text{(A.5)}$$

where, by (A.1)–(A.2),

$$\mathrm{E}(N) = \int_{W \times \mathbb{S}^k} \rho(y, u)\, \mathrm{d}\mu(y, u)$$

and

$$\mathrm{E}\{N(N-1)\} = \int_{W \times \mathbb{S}^k} \int_{W \times \mathbb{S}^k} \rho(y_1, u_1)\rho(y_2, u_2) g(y_1, u_1, y_2, u_2)\, \mathrm{d}\mu(y_1, u_1)\, \mathrm{d}\mu(y_2, u_2).$$

Note that Y has intensity function ρ_1 and pair correlation function g_1 given by

$$\rho_1(y) = \int \rho(y, u)\, \mathrm{d}\nu(u), \qquad y \in \mathbb{R}^d, \qquad \text{(A.6)}$$

and

$$\rho_1(y_1)\rho_1(y_2)g_1(y_1, y_2) = \iint \rho(y_1, u_1)\rho(y_2, u_2) g(y_1, u_1, y_2, u_2)\, \mathrm{d}\nu(u_1)\, \mathrm{d}\nu(u_2) \qquad \text{(A.7)}$$

for $y_1, y_2 \in \mathbb{R}^d$, where we set $g_1(y_1, y_2) = 0$ if $\rho_1(y_1)\rho_1(y_2) = 0$. This follows from (A.1)–(A.2) and definitions of the intensity and pair correlation function for spatial point processes (see e.g. Møller and Waagepetersen, 2004) Clearly, if X is stationary in space, then Y is stationary, ρ_1 is constant, and $g_1(y_1, y_2)$ is stationary, that is, it depends only on $y_1 - y_2$. If in addition X is isotropic in space, then $g_1(y_1, y_2)$ is isotropic, that is, it depends only on $\|y_1 - y_2\|_d$. On the other hand if Y is stationary (or isotropic) and the spherical components are independent of Y, then X is stationary (or isotropic) in space.

Similarly, using definitions of the intensity and pair correlation function for point processes on the sphere (Lawrence et al., 2016; Møller and Rubak, 2016), U_W has intensity function ρ_2 (with respect to ν) and pair correlation function g_2 (with respect to $\nu \otimes \nu$) given by

$$\rho_2(u) = \int_W \rho(y, u)\,\mathrm{d}y, \qquad u \in \mathbb{S}^k, \tag{A.8}$$

and

$$\begin{aligned}\rho_2(u_1)\rho_2(u_2)g_2(u_1, u_2)& \\ = \int_W \int_W \rho(y_1, u_1)\rho(y_2, u_2)g(y_1, u_1, y_2, u_2)\,\mathrm{d}y_1\,\mathrm{d}y_2&\end{aligned} \tag{A.9}$$

for $u_1, u_2 \in \mathbb{S}^k$, where we set $g_2(u_1, u_2) = 0$ if $\rho_2(u_1)\rho_2(u_2) = 0$. Note that we suppress in the notation that ρ_2 and g_2 depend on W. Obviously, if X is isotropic on the sphere, then ρ_2 is constant and $g_2(u_1, u_2)$ is isotropic as it depends only on $d(u_1, u_2)$.

2.2 Examples

The following examples introduce the point process models considered in this paper.

Example 1 (Poisson and Cox processes). First, suppose X is a *Poisson process* with a locally integrable intensity function ρ. This means that, the count $N(A) = \#X_A$ is Poisson distributed with mean $\int_A \rho(y, u)\,\mathrm{d}\mu(y, u)$ for any bounded Borel set $A \subset S$ and, conditional on $N(A)$, the points in X_A are independent and identically distributed (IID) with a density proportional to ρ restricted to A. Note that $g = 1$. Further, X is stationary in space and isotropic on the sphere if and only if ρ is constant, in which case we call X a *homogeneous Poisson process* with intensity ρ. Furthermore, Y and U_W are Poisson processes, so $g_1 = 1$ and $g_2 = 1$.

Second, let $\Lambda = \{\Lambda(y, u) : (y, u) \in S\}$ be a non-negative random field so that with probability one $\int_A \Lambda(y, u)\,\mathrm{d}\mu(y, u)$ is finite for any bounded Borel set $A \subset S$. If X conditioned on Λ is a Poisson process with intensity function

Λ, then X is said to be a *Cox process driven by* Λ (Cox, 1955). Clearly, the intensity and pair correlation functions of X are

$$\rho(y,u) = \mathrm{E}\{\Lambda(y,u)\}, \qquad y \in \mathbb{R}^d,\ u \in \mathbb{S}^k, \tag{A.10}$$

and

$$\rho(y_1,u_1)\rho(y_2,u_2)g(y_1,u_1,y_2,u_2) = \mathrm{E}\{\Lambda(y_1,u_1)\Lambda(y_2,u_2)\},$$

for $y_1,y_2 \in \mathbb{R}^d$ and $u_1,u_2 \in \mathbb{S}^k$. To separate the intensity function ρ from random effects, it is convenient to work with a so-called residual random field $R = \{R(y,u) : (y,u) \in S\}$ fulfilling $\Lambda(y,u) = \rho(y,u)R(y,u)$, so $\mathrm{E}\{R(y,u)\} = 1$ (see e.g. Møller and Waagepetersen, 2007; Diggle, 2014). Then

$$g(y_1,u_1,y_2,u_2) = \mathrm{E}\{R(y_1,u_1)R(y_2,u_2)\}, \qquad y_1,y_2 \in \mathbb{R}^d,\ u_1,u_2 \in \mathbb{S}^k, \tag{A.11}$$

whenever $\rho(y_1,u_1)\rho(y_2,u_2) > 0$.

Note that projected point processes Y and U_W are Cox processes driven by the random fields $\{\int_{\mathbb{S}^k} \Lambda(y,u)\,\mathrm{d}\nu(u) : y \in \mathbb{R}^d\}$ and $\{\int_W \Lambda(y,u)\,\mathrm{d}y : u \in \mathbb{S}^k\}$, respectively. Their intensity and pair correlation functions are specified by (A.6)–(A.9).

Example 2 (Log Gaussian Cox processes). A Cox process X is called a *log Gaussian Cox process* (*LGCP*; Møller et al., 1998) if the residual random field is of the form $R = \exp(Z)$, where Z is a Gaussian random field (GRF) with mean function $\mu(y,u) = -c(y,u,y,u)/2$, where c is the covariance function of Z. Note that X has pair correlation function

$$g(y_1,u_1,y_2,u_2) = \exp\left[c\{(y_1,u_1),(y_2,u_2)\}\right] \tag{A.12}$$

for $y_1,y_2 \in \mathbb{R}^d$ and $u_1,u_2 \in \mathbb{S}^k$.

Example 3 (Marked point processes). It is sometimes useful to view X as a marked point process (see e.g. Daley and Vere-Jones, 2003; Illian et al., 2008), where the spatial components are treated as the ground process and the spherical components as marks. Often it is of interest to test the hypothesis H_0 that the marks are IID and independent of the ground process Y. Under H_0, with each mark following a density p with respect to ν, the intensity is

$$\rho(y,u) = \rho_1(y)p(u), \qquad y \in \mathbb{R}^d,\, u \in \mathbb{S}^k,$$

and the pair correlation function

$$g(y_1,u_1,y_2,u_2) = g_1(y_1,y_2), \qquad y_1,y_2 \in \mathbb{R}^d,\, u_1,u_2 \in \mathbb{S}^k,$$

does not depend on (u_1,u_2).

In some situations, it may be more natural to look at it conversely, that is, treating U_W as the ground process and Y_W as marks. Then similar results for ρ and g may be established by interchanging the roles of points and marks.

Example 4 (Independently marked determinantal point processes). Considering a space-sphere point process X as a marked point process that fulfils the hypothesis H_0 given in Example 3, we may let the ground process Y be distributed according to any point process model of our choice regardless of the marks U. For instance, in case of repulsion between the points in Y, a *determinantal point process (DPP)* may be of interest because of its attractive properties (see Lavancier et al., 2015, and the references therein). Briefly, a DPP is defined by a so-called kernel $C : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$, which we assume is a complex covariance function, that is, C is positive semi-definite and Hermitian. Furthermore, let $\rho_1^{(n)}$ denote the nth order joint intensity function of Y, that is, $\rho_1^{(1)} = \rho_1$ is the intensity and $\rho_1^{(2)}(y_1, y_2) = \rho_1(y_1)\rho_1(y_2)g_1(y_1, y_2)$ for $y_1, y_2 \in \mathbb{R}^d$, while we refer to Lavancier et al. (2015) for the general definition of $\rho_1^{(n)}$ which is an extension of (A.6)–(A.7). If for all $n = 1, 2, \ldots,$

$$\rho_1^{(n)}(y_1, \ldots, y_n) = \det\{C(y_i, y_j)\}_{i,j=1,\ldots,n}, \qquad y_1, \ldots, y_n \in \mathbb{R}^d,$$

where $\det\{C(y_i, y_j)\}_{i,j=1,\ldots,n}$ is the determinant of the $n \times n$ matrix with (i, j)-entry $C(y_i, y_j)$, we call Y a DPP with kernel C and refer to X as an *independently marked DPP*. It follows that Y has intensity function $\rho(y) = C(y, y)$ and pair correlation function

$$g_1(y_1, y_2) = 1 - |R(y_1, y_2)|^2, \qquad y_1, y_2 \in \mathbb{R}^d,$$

whenever $\rho(y_1)\rho(y_2) > 0$, where $R(y_1, y_2) = C(y_1, y_2)/\sqrt{C(y_1, y_1)C(y_2, y_2)}$ is the correlation function corresponding to C and $|z|$ denotes the modulus of $z \in \mathbb{C}$.

Alternatively, we may look at a DPP on the sphere (Møller et al., 2018), that is, modelling U_W as a DPP while considering Y_W as the marks and impose the conditions of IID marks independent of U_W.

3 The space-sphere K-function

3.1 Definition

When (A.3) holds we say that the space-sphere point process X is *second order intensity-reweighted stationary* (SOIRS) and define the *space-sphere K-function* by

$$K(r, s) = \int_{\|y\|_d \le r,\, d(u,e) \le s} g_0\{y, d(u, e)\}\, \mathrm{d}\mu(y, u), \qquad r \ge 0,\ 0 \le s \le \pi, \tag{A.13}$$

where $e \in \mathbb{S}^k$ is an arbitrary reference direction. This definition does not depend on the choice of e, as the integrand only depends on $u \in \mathbb{S}^k$ through its geodesic distance to e and $\nu(\cdot)$ is a rotation invariant measure. For example, we may let $e = (0, \ldots, 0, 1) \in \mathbb{S}^k$ be the "North Pole".

Let $\sigma_k = \nu(\mathbb{S}^k) = 2\pi^{(k+1)/2}/\Gamma\{(k+1)/2\}$ denote the surface measure of $\mathbb{S}^k$. For any Borel set $B \subset \mathbb{R}^d$ with $0 < |B| < \infty$, we easily obtain from (A.2) and (A.13) that

$$
\begin{aligned}
&K(r,s) \\
&= \frac{1}{|B|\sigma_k} \iint_{y_1 \in B,\, \|y_1 - y_2\|_d \le r,\, d(u_1,u_2) \le s} g_0\{y_1 - y_2, d(u_1,u_2)\} \,\mathrm{d}\mu(y_1,u_1)\,\mathrm{d}\mu(y_2,u_2) \\
&= \frac{1}{|B|\sigma_k} \mathrm{E}\Big[\sum_{(y_i,u_i),(y_j,u_j) \in X}^{\neq} \frac{\mathbb{I}\{y_i \in B,\, \|y_i - y_j\|_d \le r,\, d(u_i,u_j) \le s\}}{\rho(y_i,u_i)\rho(y_j,u_j)} \Big] \qquad \text{(A.14)}
\end{aligned}
$$

for $r \ge 0$, $0 \le s \le \pi$, where $\mathbb{I}(\cdot)$ denotes the indicator function. The relation given by (A.14) along with the requirement that the expression in (A.14) does not depend on the choice of B could alternatively have been used as a more general definition of the space-sphere K-function. Such a definition is in agreement with the one used in Baddeley et al. (2000) for SOIRS of a spatial point process. It is straightforward to show that (A.14) does not depend on B when X is stationary in space.

For $r, s > 0$ and $(y_1,u_1), (y_2,u_2) \in S$, we say that (y_1,u_1) and (y_2,u_2) are (r,s)-close neighbours if $\|y_1 - y_2\|_d \le r$ and $d(u_1,u_2) \le s$. If X is stationary in space and isotropic on the sphere, then (A.14) shows that $\rho K(r,s)$ can be interpreted as the expected number of further (r,s)-close neighbours in X of a typical point in X. More formally, this interpretation relates to the reduced Palm distribution (Daley and Vere-Jones, 2003).

Some literature treating marked point processes discuss the so-called *mark-weighted K-function* (see e.g. Illian et al., 2008; Koubek et al., 2016), which to some extent resembles the space-sphere K-function in a marked point process context; both are cumulative second order summary functions that consider points as well as marks. However, the mark-weighted K-function has an emphasis on the marked point process setup (and considers e.g. ρ_1 rather than ρ), whereas the space-sphere K-function is constructed in such a way that it is an analogue to the planar/spherical K-function for space-sphere point processes.

Example 1 continued (Poisson and Cox processes). A Poisson process is clearly SOIRS and $K(r,s)$ is simply the product of the volume of a d-dimensional ball with radius r and the surface area of a spherical cap given by $\{u \in \mathbb{S}^k : d(u,e) \le s\}$ for an arbitrary $e \in \mathbb{S}^k$ (see Li, 2011, for formulas of this area). Thus, for $r \ge 0$, the space-sphere K-function is

$$
K_{Pois}(r,s) = \begin{cases} \frac{r^d \pi^{(d+k+1)/2}}{\Gamma(1+d/2)\Gamma\{(k+1)/2\}} I_{\sin^2(s)}\left(\frac{k}{2}, \frac{1}{2}\right), & 0 \le s \le \frac{\pi}{2}, \\ \frac{r^d \pi^{(d+k+1)/2}}{\Gamma(1+d/2)\Gamma\{(k+1)/2\}} \{2 - I_{\sin^2(\pi - s)}(\frac{k}{2}, \frac{1}{2})\}, & \frac{\pi}{2} < s \le \pi, \end{cases}
$$

where $I_x(a,b)$ is the regularized incomplete beta function. In particular, if $k=2$,

$$\left.\begin{array}{ll} I_{\sin^2(s)}(\frac{k}{2},\frac{1}{2}), & 0 \leq s \leq \frac{\pi}{2} \\ 2 - I_{\sin^2(\pi-s)}(\frac{k}{2},\frac{1}{2}), & \frac{\pi}{2} < s \leq \pi \end{array}\right\} = 1-\cos(s).$$

If the residual random field R in (A.11) is invariant under translations in $\mathbb{R}^d$ and under rotations on $\mathbb{S}^k$, then the associated Cox process is SOIRS. The evaluation of g (and thus K) depends on the particular model of R as exemplified in Example 2 below and in Section 7.

Example 2 continued (LGCPs). Suppose that the distribution of R is invariant under translations in $\mathbb{R}^d$ and under rotations on $\mathbb{S}^k$, and recall that R is required to have unit mean. Then the underlying GRF Z has a covariance function of the form

$$c(y_1,u_1,y_2,u_2) = c_0\{y_1 - y_2, d(u_1,u_2)\}, \qquad y_1,y_2 \in \mathbb{R}^d,\ u_1,u_2 \in \mathbb{S}^k,$$

and $\mathrm{E}Z(y,u) = -\sigma^2/2$ for all $y \in \mathbb{R}^d$ and $u \in \mathbb{S}^k$, where $\sigma^2 = c_0(0,0)$ is the variance. It then follows from (A.12) that X is SOIRS with

$$g_0(y,s) = \exp\{c_0(y,s)\}, \qquad y \in \mathbb{R}^d,\ 0 \leq s \leq \pi. \tag{A.15}$$

4 Separability

4.1 First order separability

We call the space-sphere point process X *first order separable* if there exist non-negative Borel functions f_1 and f_2 such that

$$\rho(y,u) = f_1(y)f_2(u), \qquad y \in \mathbb{R}^d,\ u \in \mathbb{S}^k.$$

By (A.4), (A.6), and (A.8) this is equivalent to

$$\rho(y,u) = \rho_1(y)\rho_2(u)/\mathrm{E}(N), \qquad y \in \mathbb{R}^d,\ u \in \mathbb{S}^k, \tag{A.16}$$

recalling that ρ_2 and N depend on W, but $\rho_2/\mathrm{E}N$ does not depend on the choice of W. Then, in a marked point process setup where the spherical components are treated as marks, $\rho_2(\cdot)/\mathrm{E}(N)$ is the density of the mark distribution. First order separability was seen in Example 3 to be fulfilled under the assumption of IID marks independent of the ground process. Moreover, any homogeneous space-sphere point process is clearly first order separable. In practice, first order separability is a working hypothesis which may be hard to check.

4.2 Second order separability

If there exist Borel functions k_1 and k_2 such that

$$g(y_1, u_1, y_2, u_2) = k_1(y_1, y_2) k_2(u_1, u_2), \qquad y_1, y_2 \in \mathbb{R}^d,\ u_1, u_2 \in \mathbb{S}^k,$$

we call X *second order separable*. Assuming first order separability, it follows by (A.5), (A.7), (A.9), and (A.16) that second order separability is equivalent to

$$g(y_1, u_1, y_2, u_2) = \beta g_1(y_1, y_2) g_2(u_1, u_2), \qquad y_1, y_2 \in \mathbb{R}^d,\ u_1, u_2 \in \mathbb{S}^k, \quad \text{(A.17)}$$

where

$$\beta = \mathrm{E}(N)^2 / \mathrm{E}\{N(N-1)\}$$

and noting that β and g_2 depend on W, but βg_2 does not depend on the choice of W. The value of β may be of interest: for a Poisson Process, $\beta = 1$; for a Cox process, $\mathrm{var}(N) \geq \mathrm{E}(N)$ (see e.g. Møller and Waagepetersen, 2004), so $\beta \leq 1$; for an independently marked DPP, $\beta \geq 1$ (Lavancier et al., 2015).

Example 1 continued (Poisson and Cox processes). Clearly, when X is a Poisson process, it is second order separable. Assume instead that X is a Cox process and the residual random field is separable, that is, $R(y, u) = R_1(y) R_2(u)$, where $R_1 = \{R_1(y) : y \in \mathbb{R}^d\}$ and $R_2 = \{R_2(u) : u \in \mathbb{S}^k\}$ are independent random fields. Then, by (A.11), X is second order separable and

$$g(y_1, u_1, y_2, u_2) = \mathrm{E}\{R_1(y_1) R_1(y_2)\} \mathrm{E}\{R_2(u_1) R_2(u_2)\}$$

for $y_1, y_2 \in \mathbb{R}^d$ and $u_1, u_2 \in \mathbb{S}^k$.

Example 2 continued (LGCPs). If X is a LGCP driven by $\Lambda(y, u) = \rho(y, u) \cdot \exp\{Z(y, u)\}$, second order separability is implied if $Z_1 = \log R_1$ and $Z_2 = \log R_2$ are independent GRFs so that $Z(y, u) = Z_1(y) + Z_2(u)$. Then, by the imposed invariance properties of the distribution of the residual random field, Z_1 must be stationary with a stationary covariance function $c_1(y_1, y_2) = c_{01}(y_1 - y_2)$ and mean $-c_{01}(0)/2$, and Z_2 must be isotropic with an isotropic covariance function $c_2(u_1, u_2) = c_{02}\{d(u_1, u_2)\}$ and mean $-c_{02}(0)/2$. Consequently, in (A.15), $c_0(y, s) = c_{01}(y) + c_{02}(s)$ for $y \in \mathbb{R}^d$ and $0 \leq s \leq \pi$.

Example 3 continued (Marked point processes). Consider the space-sphere point process X as a marked point process with marks in $\mathbb{S}^k$. As previously seen, first and second order separability is fulfilled under the assumption of IID marks independent of the ground process, but we may in fact work with weaker conditions to ensure the separability properties as follows. Assume that each mark is independent of the ground process Y and the marks are identically distributed following a density function p with respect to ν. Then the first order separability condition (A.16) is satisfied with $\rho_2(u) = \mathrm{E}(N)p(u)$ for $u \in \mathbb{S}^k$. In addition, assuming the conditional distribution of the marks given Y is such that any pair of marks is independent of Y and follows the same joint density $q(\cdot,\cdot)$ with respect to $\nu \otimes \nu$, it is easily seen that the second order separability condition (A.17) is satisfied with

$$g_2(u_1, u_2) = \frac{q(u_1, u_2)}{\beta p(u_1) p(u_2)}, \qquad u_1, u_2 \in \mathbb{S}^k,$$

whenever $\rho_2(u_1)\rho_2(u_2) > 0$. If we also have pairwise independence between the marks, that is, $q(u_1, u_2) = p(u_1)p(u_2)$, then the pair correlation function $g(y_1, u_1, y_2, u_2) = g_1(y_1, y_2)$ does not depend on (u_1, u_2) and $g_2(u_1, u_2) = 1/\beta$ is constant. Note that this implies $g_2 \leq 1$ for an independently marked DPP, reflecting that even when the marks are drawn independently of Y the behaviour of the points implicitly affects the marks as the number of points is equal to the number of marks.

Again, the roles of points and marks may be switched resulting in statements analogue to those above.

4.3 Assuming both SOIRS and first and second order separability

Suppose that X is both SOIRS and first and second order separable. Then the space-sphere K-function can be factorized as follows. Note that Y and U_W are SOIRS since there by (A.3), (A.7), (A.9), and (A.16) exist Borel functions g_{01} and g_{02} such that

$$\begin{aligned} g_1(y_1, y_2) &= g_{01}(y_1 - y_2) \\ &= \iint \frac{\rho_2(u_1)}{\mathrm{E}(N)} \frac{\rho_2(u_2)}{\mathrm{E}(N)} g_0\{y_1 - y_2, d(u_1, u_2)\} \,\mathrm{d}\nu(u_1) \,\mathrm{d}\nu(u_2) \end{aligned} \tag{A.18}$$

for $y_1, y_2 \in \mathbb{R}^d$ with $\rho_1(y_1)\rho_1(y_2) > 0$, and

$$\begin{aligned} g_2(u_1, u_2) &= g_{02}\{d(u_1, u_2)\} \\ &= \int_W \int_W \frac{\rho_1(y_1)}{\mathrm{E}(N)} \frac{\rho_1(y_2)}{\mathrm{E}(N)} g_0\{y_1 - y_2, d(u_1, u_2)\} \,\mathrm{d}y_1 \,\mathrm{d}y_2 \end{aligned} \tag{A.19}$$

for $u_1, u_2 \in \mathbb{S}^k$ with $\rho_2(u_1)\rho_2(u_2) > 0$. Hence, the inhomogeneous K-function for the spatial components in Y (introduced in Baddeley et al., 2000) is

$$K_1(r) = \int_{\|y\|_d \leq r} g_{01}(y) \,\mathrm{d}y, \qquad r \geq 0,$$

and the inhomogeneous K-function for the spherical components in U_W (introduced in Lawrence et al., 2016; Møller and Rubak, 2016) is

$$K_2(s) = \int_{d(u,e) \leq s} g_{02}\{d(u, e)\} \,\mathrm{d}\nu(u), \qquad 0 \leq s \leq \pi,$$

where $e \in \mathbb{S}^k$ is arbitrary. Combining (A.13) and (A.17)–(A.19), we obtain

$$K(r,s) = \beta K_1(r) K_2(s), \qquad r \geq 0,\, 0 \leq s \leq \pi.$$

Note that, if X is a first order separable Poisson process, then $D(r,s) = K(r,s) - K_1(r)K_2(s)$ is 0, and an estimate of D may also be used as a functional summary statistic when testing a Poisson hypothesis.

5 Estimation of K-functions

In this section, we assume for specificity that the observation window is $W \times \mathbb{S}^k$, where $W \subset \mathbb{R}^d$ is a bounded Borel set, and a realisation $X_{W\times\mathbb{S}^k} = x_{W\times\mathbb{S}^k}$ is observed; in Section 7, we discuss other cases of observation windows. We let $Y_W = y_W$ and $U_W = u_W$ be the corresponding sets of observed spatial and spherical components.

First, assume that ρ_1 and ρ_2 are known. Following Baddeley et al. (2000), we estimate K_1 by

$$\hat{K}_1(r) = \sum_{y_i, y_j \in y_W}^{\neq} \frac{\mathbb{I}(\|y_i - y_j\|_d \leq r)}{w_1(y_i, y_j)\rho_1(y_i)\rho_1(y_j)}, \qquad r \geq 0, \tag{A.20}$$

where w_1 is an edge correction factor on $\mathbb{R}^d$. If we let $w_1(y_i, y_j) = |W \cap W_{y_i - y_j}|$ be the translation correction factor (Ohser, 1983), where $W_y = \{y + z : z \in W\}$ denotes the translation of W by $y \in \mathbb{R}^d$, then $\hat{K}_1$ is an unbiased estimate of K_1 (see e.g. Lemma 4.2 in Møller and Waagepetersen, 2004). For $d = 1$, we may instead use the temporal edge correction factor with $w_1(y_i, y_j) = |W|$ if $[y_i - y_j, y_i + y_j] \subseteq W$ and $w_1(y_i, y_j) = |W|/2$ otherwise (Diggle et al., 1995; Møller and Ghorbani, 2012). Moreover, for estimation of K_2, we use the unbiased estimate

$$\hat{K}_2(s) = \frac{1}{\sigma_k} \sum_{u_i, u_j \in u_W}^{\neq} \frac{\mathbb{I}\{d(u_i, u_j) \leq s\}}{\rho_2(u_i)\rho_2(u_j)}, \qquad 0 \leq s \leq \pi, \tag{A.21}$$

cf. Lawrence et al. (2016) and Møller and Rubak (2016). A natural extension of the above estimates gives the following estimate of K:

$$\hat{K}(r,s) = \frac{1}{\sigma_k} \sum_{(y_i,u_i),(y_j,u_j) \in x_{W\times\mathbb{S}^k}}^{\neq} \frac{\mathbb{I}\{\|y_i - y_j\|_d \leq r,\, d(u_i, u_j) \leq s\}}{w_1(y_i, y_j)\rho(y_i, u_i)\rho(y_j, u_j)} \tag{A.22}$$

for $r \geq 0$, $0 \leq s \leq \pi$. This is straightforwardly seen to be an unbiased estimate when w_1 is the translation correction factor.

Second, in practice we need to replace ρ_1 in (A.20), ρ_2 in (A.21), and ρ in (A.22) by estimates, as exemplified in Section 6. This may introduce a bias.

6 Data examples

6.1 Fireball locations over time

Figure A.2 shows the time and location of $n = 344$ fireballs observed over a time period from `2005-01-01 03:44:09` to `2016-08-12 23:59:59` corresponding to a time frame W of about 606 weeks. The data can be recovered at `http://neo.jpl.nasa.gov/fireballs/` using these time stamps. Figure A.2 reveals no inhomogeneity of neither fireball locations or event times. Therefore we assumed first order homogeneity, and used the following unbiased estimates for the intensities:

$$\hat{\rho}_1 = n/|W| = 0.57, \;\; \hat{\rho}_2 = n/(4\pi) = 27.37, \;\; \hat{\rho} = n/(4\pi|W|) = 0.05.$$

Then $\hat{K}_1$, $\hat{K}_2$, and $\hat{K}$ (with w_1 in (A.20) and (A.22) equal to the temporal edge correction factor) were used as test functions in three different global rank envelope tests for testing whether fireball event times, locations, and locations over time each could be described by a homogeneous Poisson model with estimated intensity $\hat{\rho}_1$, $\hat{\rho}_2$, and $\hat{\rho}$, respectively. Appendix A provides a brief account on global rank envelope tests; see also Myllymäki et al. (2017). Under each of the three fitted Poisson processes and using 2499 simulations (as recommended in Myllymäki et al., 2017), we obtained p-intervals of $(0.028, 0.040)$ for the event times, $(0.908, 0.908)$ for the locations, and $(0.445, 0.516)$ for the locations over time. The associated 95% global rank envelopes for $\hat{K}_1$ and $\hat{K}_2$ are shown in Figure A.3, and the difference between $\hat{K}$ and the upper and lower 95% global rank envelope is shown in Figure A.4. Since $\hat{K}_2$ and $\hat{K}$ stay inside the 95% global rank envelopes for the considered distances on $\mathbb{S}^k$ and $\mathbb{R} \times \mathbb{S}^k$, there is no evidence against a homogeneous Poisson model for neither locations or locations over time. On the other hand, with a conservative p-value of 4%, the global rank envelope test based on $\hat{K}_1$ indicates that a homogeneous Poisson model for the event times is not appropriate. However, the observed test function $\hat{K}_1(r)$ falls only outside the envelope in Figure A.3 for large values of r. Thus, choosing a slightly smaller interval of r-values would lead to a different conclusion.

As an alternative to the space-sphere K-function, we considered the summary function $D(r,s)$ which in case of a Poisson process is 0. Estimating D by $\hat{D}(r,s) = \hat{K}(r,s) - \hat{K}_1(r)\hat{K}_2(s)$, we performed a global rank envelope test with D as test function. The resulting test gave a p-interval of $(0.537, 0.564)$ which is similar to the one obtained using $\hat{K}$ as test function.

6.2 Location and orientation of pyramidal neurons

We now return to the space-sphere point pattern concerning location and orientation of pyramidal neurons described in Section 1, which is a data set

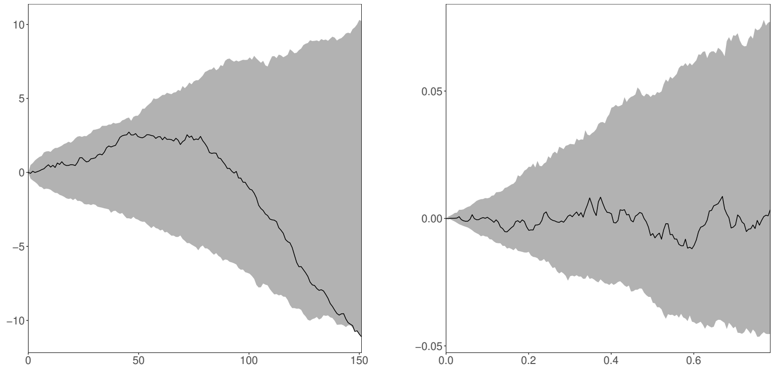


Figure A.3: Left: $\hat{K}_1(r) - 2r$ for the fireball event times (solid curve) along with a 95% global rank envelope (grey area) under a homogeneous Poisson model on the time interval for the observed events. Right: $\hat{K}_2(s) - 2\pi\{1 - \cos(s)\}$ for the fireball locations (solid curve) along with a 95% global rank envelope (grey area) under a homogeneous Poisson model on $\mathbb{S}^2$.

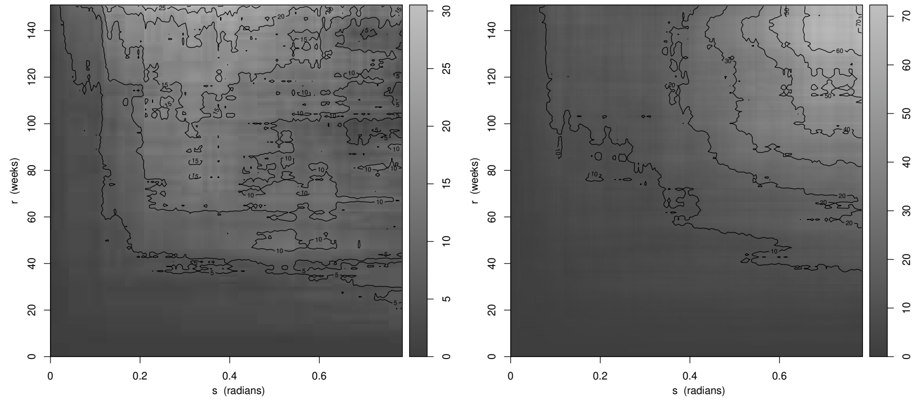


Figure A.4: Difference between $\hat{K}(r,s)$ for the observed fireball locations over time and the lower ($\hat{K}_{\text{low}}$) or upper ($\hat{K}_{\text{upp}}$) 95% global rank envelope under a homogeneous Poisson model on $W \times \mathbb{S}^2$. Left: $\hat{K}(r,s) - \hat{K}_{\text{low}}(r,s)$. Right: $\hat{K}_{\text{upp}}(r,s) - \hat{K}(r,s)$.

collected by Ali H. Rafati, a biomedical and clinical scientist. The point pattern is observed on $W \times \mathbb{S}^2$, where $W \subset \mathbb{R}^3$ is the rectangular box shown in Figure A.1 with side lengths 492.7 µm, 132.0 µm, and 407.7 µm. Due to the way data was collected, 46 neurons in the original dataset had an orientation/unit vector lying exactly in the x-z plane, meaning that the orientations of these 46 neurons are located only on a great circle of $\mathbb{S}^2$. To keep the analysis simple we disregarded these neurons, resulting in a dataset consisting of $n = 504$ neurons. Below we initially discuss an appropriate parametric forms of the intensity for the locations and orientations and how to estimate the intensity parameters. Then we investigate whether the orientations and locations can be described by a Poisson model with the proposed intensity, where we first consider the data as two separate point patterns (a spatial point pattern describing the locations and a spherical point pattern describing the orientations) and next as a space-sphere point pattern.

Figure A.1 reveals no inhomogeneity for the neuron locations, whereas it is evident that the orientations are inhomogeneous pointing mostly toward the pial surface of the brain (the plane perpendicular to the z-axis). Thus, we estimated the intensity of the locations by $\hat{\rho}_1 = n/|W| = 1.9 \times 10^{-5}$, and for the orientations we let the intensity be $\rho_2(u) = nf(u)$, where f is a density on $\mathbb{S}^2$ which we model as follows. Figure A.5 indicates that the orientations arise from a mixture of two distributions; one distribution with points falling close to the North Pole and another with points falling in a narrow girdle. Therefore, we let $f(u) = pf_K(u) + (1-p)f_W(u)$ be the mixture density of a Kent and a Watson distribution on $\mathbb{S}^2$ (see e.g. Fisher et al., 1987) for a detailed description of these spherical distributions). In brief, the Kent density, f_K, depends on five parameters (three directional, one concentration, and one ovalness parameter), and its contours are oval with centre and form specified by the directional parameters. Depending on the values of the ovalness and concentration parameter, the Kent distribution is either uni- or bimodal. Here, to account for the large number of points centred around the North Pole, we consider the unimodal Kent distribution. Furthermore, the Watson density, f_W, depends on two parameters; a directional parameter determining the centres of the density's circular contours, and a concentration parameter controlling where and how fast the density peaks. Depending on the sign of the concentration parameter, the density either decreases or increases as the geodesic distance to the centres of the contours increases, giving rise to either a bimodal or girdle shaped distribution. Since the Watson distribution shall describe the orientations on the girdle, the concentration parameter must be negative.

The eight parameters of the proposed intensity function ρ_2 were estimated as follows. The orientations occurring on the southern hemisphere are presumed to come from the Watson distribution, while the orientations on the northern hemisphere come from both distributions. Therefore, and because the Watson density on the northern hemisphere is a reflection of the southern hemisphere, we simple estimated the mixture probability by $\hat{p} = 1 - 2n_s/n = 0.94$, where n_s is the observed number of points on the southern hemisphere. The directional parameters were chosen based on expectations expressed by the scientist behind the data collections, which were supported by visual inspection of the data; e.g. the directional parameter for the Kent distribution that determines the centre of the contours was chosen as the North Pole corresponding to the direction perpendicular to the pial surface and consistent with Figure A.5. Finally, the concentration and ovalness parameters were estimated by numerical maximization of the profile likelihood, giving the estimated density

$$\hat{f}(u) = 0.94\, C_K \exp\{14.89 u_3 + 2.69(u_1^2 - u_2^2)\} + 0.06\, C_W \exp(-7.88 u_2^2),$$

where $u = (u_1, u_2, u_3) \in \mathbb{S}^2$ and C_K, C_W are normalising constants (see Fisher et al., 1987, for details). Figure A.5 suggests that the fitted density (and associated marginal densities found by numerical integration of $\hat{f}$) adequately describe the distribution of the observed orientations. Therefore, we now turn to investigate whether the locations and orientations can be described by Poisson models with the estimated intensities.

First, we considered the locations and orientations separately and used $\hat{K}_1$ and $\hat{K}_2$, respectively, as test functions for the global rank envelope procedure. Using 2499 simulations from a homogeneous Poisson process on W, we obtained a global rank envelope test with p-interval $(0.000, 0.020)$. Thus, we reject that the locations follow a homogeneous Poisson model. The associated 95% global rank envelopes in Figure A.6 show that the rejection is due to the observed $\hat{K}_1(r)$ falling below the envelope for r-values between 10 µm to 25 µm. This suggests that the observed locations exhibit some degree of regularity that needs to be modelled. For the orientations, a global rank envelope test based on 2499 simulations under an inhomogeneous Poisson model on $\mathbb{S}^2$ with intensity $\hat{\rho}_2(u) = n\hat{f}(u)$ gave a p-interval of $(0.475, 0.481)$ and thus no evidence against the proposed model. Figure A.6 shows the associated 95% global rank envelope.

Second, we considered the data as a space-sphere point pattern and used $\hat{K}$ and $\hat{D}$ to test for a Poisson model with a separable intensity estimated by $\hat{\rho}(y, u) = \hat{\rho}_1 \hat{\rho}_2(u)/n = n\hat{f}(u)/|W|$, cf. (A.16). As the test functions $\hat{K}(r, s)$ and $\hat{D}(r, s)$ depend on the two-dimensional argument (r, s) and they are nonsmooth with large jumps due to few orientations occurring in areas with low intensity, we increased the number of simulations to 49999 in order to improve the quality of the global rank envelope test (49999 is a much higher number than recommended in Myllymäki et al., 2017, for test functions depending on one argument only). This resulted in the p-intervals $(0.547, 0.549)$ for $\hat{K}$ and $(0.000, 0.003)$ for $\hat{D}$; plots of the difference between the associated

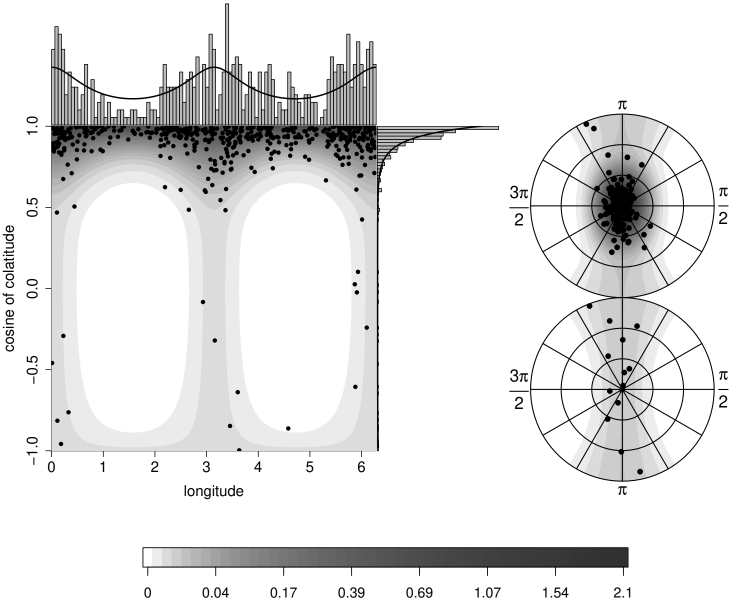


Figure A.5: Plots of the observed neuron orientations (dots) and the fitted mixture density $\hat{f}$ (grey scale). Left: plot of orientations represented as cosine-colatitude and longitude along with their marginal histograms and marginal fitted densities (solid curves) found by numerically integrating $\hat{f}$. Right: stereographic projection of the northern (top) and southern (bottom) hemisphere.

envelopes and the observed test function are shown in Figure A.7. In conclusion, the test based on $\hat{K}$ reveals no evidence against the proposed space-sphere Poisson model even though the corresponding Poisson model for the locations was rejected by the test based on $\hat{K}_1$. However, the test based on $\hat{D}$ provides a great deal of evidence against the model. This conclusion is probably due to the fact that for this data set $\hat{K}_1(r)\hat{K}_2(s) \gg \hat{K}(r,s)$, meaning that the test based on $\hat{D}$ is highly controlled by the values of $\hat{K}_1$ and $\hat{K}_2$, which results in a rejection for r-values from 10 µm to 20 µm, in line with the test based on $\hat{K}_1$.

It is unsatisfactory that $\hat{K}$ does not detect any deviation from Poisson when $\hat{K}_1$ clearly does, but we expect that the large jumps in $\hat{K}(r,s)$, caused by (r,s)-close neighbours with low intensity, may explain why no evidence against the model is detected. The few orientations that were modelled using a Watson distribution mostly fall in places with very low intensity. Therefore, we independently thinned the space-sphere point pattern with retention probability $\hat{p}\hat{f}_K(u)/\hat{f}(u)$ for $u \in \mathbb{S}^2$. Thereby (with high probability) we removed neurons with orientations that were most likely generated by the Watson distribution. This lead to removal of 26 neurons. For the thinned data, the global rank envelope test based on $\hat{K}$ for testing the hypothesis of an inhomogeneous Poisson process with intensity proportional to a Kent density gave a p-interval of $(0.052, 0.058)$. Still, the model was not rejected at a 5% significance level, but we at least got closer to a rejection; and so we continued the analysis with the thinned data. The analysis here indicates that, at least in some cases, the power of the global rank envelope test based on $\hat{K}$ may be small. This is investigated further in Section 6.3.

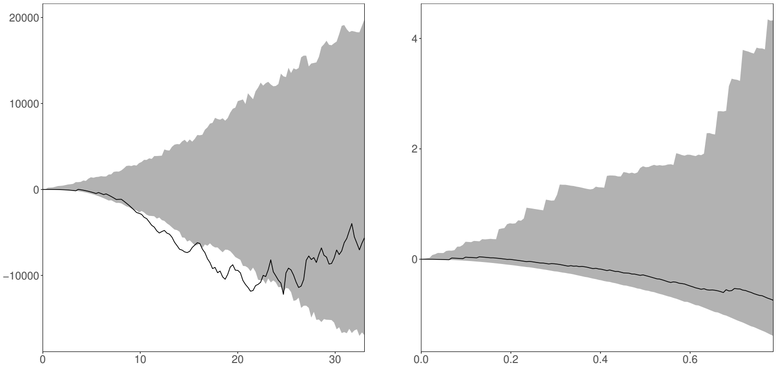


Figure A.6: Left: $\hat{K}_1(r) - 4\pi r^3/3$ for the neuron locations (solid curve) along with a 95% global rank envelope (grey area) under a homogeneous Poisson model on $W \subseteq \mathbb{R}^3$. Right: $\hat{K}_2(s) - 2\pi\{1 - \cos(s)\}$ for the neuron orientations (solid curve) along with a 95% global rank envelope (grey area) under the fitted inhomogeneous Poisson model on $\mathbb{S}^2$.

As we have seen, a homogeneous Poisson model is not adequate for the locations, and thus a Poisson model with intensity $\hat{\rho}$ as described above is not suitable for describing the space-sphere point pattern. To investigate whether orientations and locations can be modelled separately, that is, whether the locations and orientations are independent, we kept the locations fixed, and independent of the locations we simulated IID orientations from the fitted Kent distribution. The resulting global rank envelope test based on 49999 of such simulations gave a p-interval of $(0.9255, 0.9258)$ for $\hat{K}$ and $(0.1265, 0.1266)$ for $\hat{D}$, showing no evidence against the hypothesis of independence between locations and orientations. Alternatively, if one does not have a suitable model to simulate the spherical (or spatial) components from, the independence test may be performed by randomly permuting the components. Formally, this tests only the hypothesis of exchangeability; a property that is fulfilled under independence. Performing such a permutation test for our data where the locations were fixed and the orientations permuted 49999 times resulted in a p-interval of $(0.5431, 0.5445)$ using either $\hat{K}$ or $\hat{D}$ (as $\hat{K}$ and $\hat{D}$ only differ by a constant under permutation of the orientations and thus lead to equivalent tests).

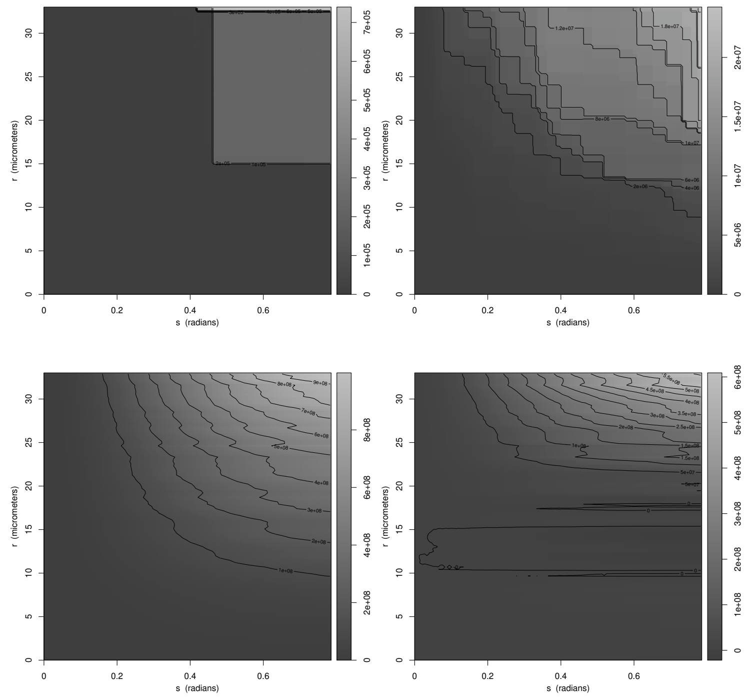


Figure A.7: For $\hat{T} = \hat{K}$ (first row) and $\hat{T} = \hat{D}$ (second row): difference between $\hat{T}(r,s)$ for the observed neuron locations and orientations and the lower ($\hat{T}_{\text{low}}$) or upper ($\hat{T}_{\text{upp}}$) 95% global rank envelope under the fitted inhomogeneous Poisson model on $W \times \mathbb{S}^2$. Left: $\hat{T}(r,s) - \hat{T}_{\text{low}}(r,s)$. Right: $\hat{T}_{\text{upp}}(r,s) - \hat{T}(r,s)$.

6.3 Simulation study

In the data analyses in Sections 6.1–6.2, the tests based on $\hat{K}$ failed to reject the proposed Poisson models in cases where the corresponding spatial model was rejected when using $\hat{K}_1$. To investigate whether the space-sphere K-function is a valuable addition to the existing functional summary statistics on the space and sphere, we performed a simulation study comparing the power of global rank envelope tests based on either $\hat{K}$, $\hat{D}$, or a combination of $\hat{K}_1$ and $\hat{K}_2$. (The combined test function is simply a concatenation of $\hat{K}_1$ and $\hat{K}_2$. Mrkvička et al., 2017, recommended using such a combination rather than $\hat{K}_1$ or $\hat{K}_2$ as a test function.)

Specifically, we consider a homogeneous LGCP X driven by a random field $\Lambda(y,u) = \rho \exp\{Z(y,u)\}$, where $\rho > 0$ and

$$Z(y,u) = \alpha + \sigma_1 Z_1(y) + \sigma_2 Z_2(u) + \delta Z_3(y,u), \qquad y \in \mathbb{R},\ u \in \mathbb{S}^2,$$

for parameters $\sigma_1, \sigma_2 > 0$, $\delta \geq 0$, and $\alpha = -(\sigma_1^2 + \sigma_2^2 + \delta^2)/2$. Further,

Z_1, Z_2, and Z_3 are independent GRFs with mean 0 and covariance functions $c_1(y_1, y_2) = \exp(-\|y_1 - y_2\|_d/\phi_1)$, $c_2(u_1, u_2) = \exp(-d(u_1, u_2)/\phi_2)$, and $c_3(y_1, u_1, y_2, u_2) = c_1(y_1, y_2)c_2(u_1, u_2)$, respectively, with parameters $\phi_1, \phi_2 > 0$. Note that the resulting LGCP is homogeneous (and thus first order separable) and SOIRS for any value of $\delta \geq 0$. In addition, by (A.12), the process is second order separable if and only if $\delta = 0$, in which case X has pair correlation function

$$\begin{aligned} &g_\theta(y_1, u_1, y_2, u_2) \\ &\quad = \exp\{\sigma_1^2 c_1(y_1, y_2) + \sigma_2^2 c_2(u_1, u_2)\}, \qquad y_1, y_2 \in \mathbb{R},\ u_1, u_2 \in \mathbb{S}^2, \end{aligned} \tag{A.23}$$

where $\theta = (\sigma_1, \phi_1, \sigma_2, \phi_2)$.

For each value of $\delta = 0, 0.5, 1, 1.5, 2$, we simulated 100 realisations of a LGCP on $[0, 1] \times \mathbb{S}^2$ with $\rho = 1000$, $\sigma_1 = \sigma_2 = 0.5$, $\phi_1 = 0.05$, and $\phi_2 = 0.132$. Then for each of these simulations, we fitted the LGCP model with $\delta = 0$ using a second order composite likelihood approach proposed by Guan (2006) to estimate θ. In the present time-sphere setting, for a finite point pattern $x \subset [0, 1] \times \mathbb{S}^2$, the log second order composite likelihood is given by

$$\begin{aligned} \mathrm{CL}(\theta; x) = &\sum_{(y_i,u_i),(y_j,u_j)\in x}^{\neq} w(y_i, u_i, y_j, u_j) \log\{\rho_\theta^{(2)}(y_i, u_i, y_j, u_j)\} \\ &- n_{r,s} \log\bigg\{ \int_{[0,1]\times\mathbb{S}^2} \int_{[0,1]\times\mathbb{S}^2} w(y_1, u_1, y_2, u_2) \\ &\qquad \cdot \rho_\theta^{(2)}(y_1, u_1, y_2, u_2)\, \mathrm{d}\mu(y_1, u_1)\, \mathrm{d}\mu(y_2, u_2) \bigg\}. \end{aligned} \tag{A.24}$$

Here, for user specified distances r and s, $w(y_1, u_1, y_2, u_2) = \mathbb{I}\{\|y_1 - y_2\|_d < r,\ d(u_1, u_2) < s\}$, $n_{r,s}$ is the number of (r, s)-close neighbours, and $\rho_\theta^{(2)}$ is the second order joint intensity function, which for the homogeneous LGCP presented above is $\rho_\theta^{(2)}(y_1, u_1, y_2, u_2) = \rho^2 g_\theta(y_1, u_1, y_2, u_2)$. Then (A.24) is easily seen not to depend on ρ, and by (A.23) the composite likelihood can be written as

$$\mathrm{CL}(\theta; x) = l_1(\sigma_1, \phi_1; x) + l_2(\sigma_2, \phi_2; x) \tag{A.25}$$

for functions l_1 and l_2. Thus, maximising the composite likelihood with respect to θ can be split into two maximisation problems; that is, maximising l_1 with respect to (σ_1, ϕ_1) and l_2 with respect to (σ_2, ϕ_2). Finally, we tested the null hypothesis $\delta = 0$ using the global rank envelope test with 4999 simulations from the fitted model using either $\hat{K}$, $\hat{D}$, or a combination of $\hat{K}_1$ and $\hat{K}_2$ as test functions.

Table A.1 gives an overview of the conclusions reached by these tests. Note that the power of the tests based on either of the three test functions in

	Test function	$\delta = 0$	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$	$\delta = 2$
Liberal	$\hat{K}$	4%	7%	42%	75%	98%
	$\hat{D}$	2%	45%	92%	97%	100%
	$\hat{K}_1, \hat{K}_2$	10%	11%	29%	28%	42%
Conservative	$\hat{K}$	2%	5%	32%	72%	90%
	$\hat{D}$	0%	26%	77%	82%	86%
	$\hat{K}_1, \hat{K}_2$	10%	11%	29%	28%	40%

Table A.1: Power of tests for different values of δ when using the global rank envelope test with either $\hat{K}$, $\hat{D}$, or $\hat{K}_1$ combined with $\hat{K}_2$. The decision was made using a significance level of 5% for both the liberal and conservative tests.

general increases with δ, both for the liberal and conservative test (for details see Appendix A). Thus, with increasing degree of non-separability the tests more often detect deviation from the separable model. However, tests based on $\hat{K}$ and particularly $\hat{D}$ seem preferable in this setup as they have a higher power than tests based on $\hat{K}_1$ combined with $\hat{K}_2$.

Obviously, the conservative p-value always lead to fewer rejections than the liberal, giving a lower power. However, if the global rank envelope procedure is based on a higher number of simulations, then the conservative and liberal test will more often lead to the same conclusion.

7 Additional comments

Section 2.2 introduced examples of space-sphere point processes for which the second order separability property described in Section 4 seems natural. However, for other classes of point processes a different structure of the pair correlation function may be more interesting. For example, suppose X is a Cox process driven by

$$\Lambda(y,u) = \sum_{(y',u',\gamma')\in\Phi} \gamma' k(y',u',y,u), \qquad y \in \mathbb{R}^d,\, u \in \mathbb{S}^k, \tag{A.26}$$

where Φ is a Poisson process on $S \times (0,\infty)$ with intensity function ζ, and $k(y',u',\cdot,\cdot)$ is a density with respect to μ. Then X is called a *shot noise Cox process (SNCP)* with kernel k (Møller, 2003). The process has intensity function

$$\rho(y,u) = \iint \gamma' \zeta(y',u',\gamma') k(y',u',y,u)\, \mathrm{d}\mu(y',u')\, \mathrm{d}\gamma', \qquad y \in \mathbb{R}^d,\, u \in \mathbb{S}^k,$$

and pair correlation function

$$\begin{aligned} &g(y_1,u_1,y_2,u_2) \\ &\quad = 1 + \frac{\iint \gamma'^2 \zeta(y',u',\gamma') k(y',u',y_1,u_1) k(y',u',y_2,u_2) \,\mathrm{d}\mu(y',u') \,\mathrm{d}\gamma'}{\rho(y_1,u_1)\rho(y_2,u_2)} \end{aligned} \tag{A.27}$$

for any $y_1, y_2 \in \mathbb{R}^d$ and any $u_1, u_2 \in \mathbb{S}^k$ with $\rho(y_1,u_1)\rho(y_2,u_2) > 0$. In the trivial case where the kernel $k(y',u',y,u)$ in (A.26) does not depend on u (or y), the SNCP is both first and second order separable, with intensity and pair correlation functions that do not depend on the spherical (or spatial) components, and the process thus fulfils second order separability in the sense of (A.17). However, the specific structure of the pair correlation function for a SNCP in (A.27) makes it more natural to look for a product structure in $g - 1$ rather than g. That is, we may say that X is second order separable if there exist Borel functions h_1 and h_2 such that

$$g(y_1,u_1,y_2,u_2) - 1 = h_1(y_1,y_2) h_2(u_1,u_2) \tag{A.28}$$

for $y_1, y_2 \in \mathbb{R}^d$, $u_1, u_2 \in \mathbb{S}^k$. This property is naturally fulfilled whenever we consider a Poisson process or any marked point process with marks that are IID and independent of the ground process as described in Example 3. Now, think of Φ in (A.26) as a marked point process with ground process $\{(y,u) : (y,u,\gamma) \in \Phi\}$ and marks $\{\gamma : (y,u,\gamma) \in \Phi\}$, and assume that the ground process and the marks are independent processes, the ground process is a homogeneous Poisson process on S with intensity $\alpha > 0$, and the marks are IID with mean m_1 and second moment m_2. If in addition $k(y',u',y,u) = k_0\{y - y', d(u,u')\}$, then Λ and thus X is stationary in space and isotropic on the sphere. Further, X is homogeneous with intensity $\rho = m_1\alpha$ and pair correlation function

$$\begin{aligned} &g(y_1,u_1,y_2,u_2) \\ &\quad = 1 + \frac{m_2}{\alpha m_1^2} \int k_0\{y_1 - y', d(u_1,u')\} k_0\{y_2 - y', d(u_2,u')\} \,\mathrm{d}\mu(y',u') \end{aligned} \tag{A.29}$$

for $y_1, y_2 \in \mathbb{R}^d$ and $u_1, u_2 \in \mathbb{S}^k$. Clearly, (A.29) depends only on (y_1, y_2) through $y_1 - y_2$, and on (u_1, u_2) through $d(u_1, u_2)$, although there is no simple expression for these dependencies in general. Furthermore, separability in the form of (A.28) is fulfilled if the kernel k in (A.29) factorizes such that

$$k_0\{y - y', d(u,u')\} = k_{01}(y - y') k_{02}\{d(u,u')\}, \qquad y, y' \in \mathbb{R}^d,\ u, u' \in \mathbb{S}^k,$$

for Borel functions k_{01} and k_{02}. Then by (A.6)–(A.9), the pair correlations functions for Y and U_W are

$$g_1(y_1,y_2) = 1 + c_1 \frac{m_2}{\alpha m_1^2} \int k_{01}(y_1 - y') k_{01}(y_2 - y') \,\mathrm{d}y', \qquad y_1, y_2 \in \mathbb{R}^d,$$

and

$$g_2(u_1, u_2) = 1 + c_2 \frac{m_2}{\alpha m_1^2} \int k_{02}\{d(u_1, u')\} k_{02}\{d(u_2, u')\} \, \mathrm{d}\nu(u'), \qquad u_1, u_2 \in \mathbb{S}^k,$$

where

$$c_1 = \frac{1}{\sigma_k^2} \iiint k_{02}\{d(u_1, u')\} k_{02}\{d(u_1, u')\} \, \mathrm{d}\nu(u_1) \, \mathrm{d}\nu(u_2) \, \mathrm{d}\nu(u')$$

and

$$c_2 = \frac{1}{|W|^2} \int \int_W \int_W k_{01}(y_1 - y') k_{01}(y_1 - y') \, \mathrm{d}y_1 \, \mathrm{d}y_2 \, \mathrm{d}y'.$$

That is, g_1 depends only on the spherical components through the constant c_1, and similarly g_2 depends only on the spatial components through c_2. Prokešová and Dvořák (2014) discuss how an analogue property can be exploited to estimate the parameters of space-time SNCPs using minimum contrast estimation for the projected processes. A similar procedure will be applicable for space-sphere point processes, but we have not investigated this further.

Section 5 considered the situation where the spatial components of X are observed on a subset of $\mathbb{R}^d$ and the spherical components are observable on the entire sphere. In more general applications, the spherical components may only be observable on a subset of $\mathbb{S}^k$ leading to edge effects on the sphere too. To account for this, edge correction methods for the sphere should be used when estimating K_2 (see Lawrence et al., 2016) and K. If X is observable on a product space $W_1 \times W_2$, where $W_1 \subset \mathbb{R}^d$ and $W_2 \subset \mathbb{S}^k$, then an edge corrected estimate for K may be obtained by combining edge corrected estimates for K_1 and K_2 analogous to (A.22). Concerning the specific choice of edge correction method, Baddeley et al. (2016) mentioned for planar point processes that, "So long as some kind of edge correction is performed ..., the particular choice of edge correction technique is usually not critical". We expect that the situation is similar for our setting.

For one-dimensional test functions, Myllymäki et al. (2017) recommend using 2499 simulations to perform a global rank envelope test, and Mrkvička et al. (2017) discuss the appropriate number of simulations when using a multivariate test function (as the empirical space-sphere K-function). In Section 6.2, we used 49999 simulations for the global rank envelope test based on $\hat{K}$, since $\hat{K}$ had steep jumps. To avoid this large number of simulations, a refinement of the global rank envelope test discussed in Mrkvička et al. (2018) can be applied.

In Example 1 we noticed that if a space-sphere point process is a Poisson process, then the spatial and spherical components are Poisson processes as well. Nevertheless, using $\hat{K}$ for testing a space-sphere Poisson model may lead to a different conclusion than using $\hat{K}_1$ and $\hat{K}_2$ for testing whether the corresponding Poisson models for the spatial and spherical components, respectively, are appropriate. Indeed, in the case of Figures A.3–A.4, the test based on $\hat{K}_1$ showed some evidence against a homogeneous Poisson model for the fireball event times, while no evidence against a homogeneous Poisson model for the locations over time was seen with the test based on $\hat{K}$. This observation together with the results in Section 6.2 was our motivation for making the simulation study in Section 6.3, where we investigated the power of global rank envelope tests based on either $\hat{K}$, $\hat{D}$, or a combination of $\hat{K}_1$ and $\hat{K}_2$, and we concluded that tests based on $\hat{K}$ and in particular $\hat{D}$ seem preferable.

In Section 6.3, we utilised homogeneity and second order separability to speed up optimisation of the second order composite likelihood proposed in Guan (2006). If the process is inhomogeneous but first (and second) order separable, we still get a separation of the composite likelihood similar to (A.25), where l_1 and l_2 now may depend also on intensity parameters. As an alternative, the second order composite likelihood discussed in Waagepetersen (2007) can be used. However, in that case, first and second order separability do not yield a separable likelihood as in (A.25), and for our simulation study it resulted in unstable estimates (and thus it was discarded in favour of the one proposed by Guan, 2006). Furthermore, one may investigate whether the adaptive procedure discussed in Lavancier et al. (2018) will provide stable estimates in the space-sphere setting. In short, Lavancier et al. (2018) consider the score function related to (A.24) and introduce a modified weight function w depending on g.

In this paper, we considered point processes living on $\mathbb{R}^d \times \mathbb{S}^k$. Naturally, we may extend the results/methods to more general metric spaces $\mathbb{R}^d \times \mathbb{M}$, where $\mathbb{M}$ is a compact set (e.g. a torus). However, we need to require some invariance property for the metric space $\mathbb{M}$ and its metric under a group action, such that we can define an equivalence of the SOIRS property needed to define K.

Acknowledgements The authors are grateful to Jiří Dvořák for helpful comments and to Ali H. Rafati for collecting the pyramidal cell data. This work was supported by The Danish Council for Independent Research | Natural Sciences, grant DFF – 7014-00074 "Statistics for point processes in space and beyond", and by the "Centre for Stochastic Geometry and Advanced Bioimaging", funded by grant 8721 from the Villum Foundation.

Appendix A

In Sections 6.1–6.3, we used the global rank envelope test presented in Myllymäki et al. (2017) to test for various point process models. In this appendix, we briefly explain the idea and use of such a test. A global rank envelope test compares a chosen test function for the observed data with the distribution of the test function under the null model; as this distribution is typically unknown it is approximated using a Monte Carlo approach. The comparison is based on a rank that only gives a weak ordering of the test functions. Thus, instead of a single p-value, the global rank envelope test provides an interval of p-values, where the end points specify the most liberal and conservative p-values of the test. A narrow p-interval is desirable as the test is inconclusive if the p-interval contains the chosen significance level. The width of the p-interval depends on the number of simulations, smoothness of the test functions and dimensionality. Myllymäki et al. (2017) recommended to use 2499 simulations for one-dimensional test functions and a significance level of 5%.

An advantage of the global rank envelope procedure is that it provides a graphical interpretation of the test in form of critical bounds (called a global rank envelope) for the test function. For example, if the observed test function is not completely inside the 95% global rank envelope, this corresponds to a rejection of the null hypothesis at a significance level of 5%. Furthermore, locations where the observed test function falls outside the global rank envelope reveal possible reasons for rejecting the null model.

In their supplementary material, Myllymäki et al. (2017) discussed two approaches for calculating test functions that rely on an estimate of the intensity. One approach is to reuse the intensity estimate for the observed point pattern in calculation of all the test functions, another is to reestimate the intensity for each simulation and then use this estimate when calculating the associated test function. For the L-function, which is a transformation of K_1, Myllymäki et al. (2017) concluded that the reestimation approach give the more powerful test. In this paper, we have therefore based all our global rank envelope tests on that approach.

References

Baddeley, A., Rubak, E., and Turner, R. (2016). *Spatial Point Patterns: Methodology and Applications with R*. Chapman & Hall/CRC, New York.

Baddeley, A. J., Møller, J., and Waagepetersen, R. P. (2000). Non- and semiparametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54:329–350.

Baddeley, A. J., Nair, G., Rakshit, S., and McSwiggan, G. (2017). "Stationary" point processes are uncommon on linear networks. *Stat*, 6:68–78.

Buxhoeveden, D. P. and Casanova, M. F. (2002). The minicolumn hypothesis in neuroscience. *Brain*, 125:935–951.

Cox, D. R. (1955). Some statistical models related with series of events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 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-Verlag, New York, second edition.

Diggle, P. J. (2014). *Statistical Analysis of Spatial and Spatio-temporal Point Patterns*. Chapman & Hall/CRC Press, Boca Raton, Florida.

Diggle, P. J., Chetwynd, A. G., Häggkvist, R., and Morris, S. E. (1995). Second-order analysis of space-time clustering. *Statistical Methods in Medical Research*, 4:124–136.

Dvořák, J. and Prokešová, M. (2016). Parameter estimation for inhomogeneous space-time shot-noise Cox point processes. *Scandinavian Journal of Statistics*, 43:939–961.

Fisher, N. I., Lewis, T., and Embleton, B. J. J. (1987). *Statistical Analysis of Spherical Data*. Cambridge University Press, New York.

Gabriel, E. and Diggle, P. J. (2009). Second-order analysis of inhomogeneous spatio-temporal point process data. *Statistica Neerlandica*, 63:43–51.

Guan, Y. (2006). A composite likelihood approach in fitting spatial point process models. *Journal of the American Statistical Association*, 101:1502–1512.

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*. Statistics in Practice. Wiley, New York.

Koubek, A., Pawlas, Z., Brereton, T., Kriesche, B., and Schmidt, V. (2016). Testing the random field model hypothesis for random marked closed sets. *Spatial Statistics*, 16:118–136.

Lavancier, F., Møller, J., and Rubak, E. (2015). Determinantal point process models and statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77:853–877.

Lavancier, F., Poinas, A., and Waagepetersen, R. P. (2018). Adaptive estimating function inference for non-stationary determinantal point processes. Available on arXiv:1806.06231.

Lawrence, T., Baddeley, A. J., Milne, R. K., and Nair, G. (2016). Point pattern analysis on a region of a sphere. *Stat*, 5:144–157.

Li, S. (2011). Concise formulas for the area and volume of a hyperspherical cap. *Asian Journal of Mathematics & Statistics*, 4:66–70.

Lorente de Nó, R. (1938). The cerebral cortex: Architecture, intracortical connections, motor projections. In Fulton, J. F., editor, *Physiology of the Nervous System*, pages 274–301. Oxford University Press, Oxford.

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

Møller, J. and Ghorbani, M. (2012). Aspects of second-order analysis of structured inhomogeneous spatio-temporal point processes. *Statistica Neerlandica*, 66:472–491.

Møller, J., Nielsen, M., Porcu, E., and Rubak, E. (2018). Determinantal point process models on the sphere. *Bernoulli*, 24:1171–1201.

Møller, J. and Rubak, E. (2016). Functional summary statistics for point processes on the sphere with an application to determinantal point processes. *Spatial Statistics*, 18:4–23.

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

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

Møller, J. and Waagepetersen, R. P. (2007). Modern statistics for spatial point processes. *Scandinavian Journal of Statistics*, 34:643–684.

Mountcastle, V. B. (1978). *The Mindful Brain: Cortical Organization and the Group-selective Theory of Higher Brain Function*. MIT Press, Cambridge.

Mrkvička, T., Myllymäki, M., and Hahn, U. (2017). Multiple Monte Carlo testing, with applications in spatial point processes. *Statistics and Computing*, 27:1239–1255.

Mrkvička, T., Hahn, U., and Myllymäki, M. (2018). A one-way ANOVA test for functional data with graphical interpretation. Available on arXiv:1612.03608.

Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H., and Hahn, U. (2017). Global envelope tests for spatial processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79:381–404.

Ohser, J. (1983). On estimators for the reduced second moment measure of point processes. *Mathematische Operationsforschung und Statistik, series Statistics*, 14:63–71.

Prokešová, M. and Dvořák, J. (2014). Statistics for inhomogeneous space-time shot-noise Cox processes. *Spatial Statistics*, 10:76–86.

Waagepetersen, R. P. (2007). An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics*, 63:252–258.

Paper B

Modelling columnarity of pyramidal cells in the human cerebral cortex

Andreas D. Christoffersen, Jesper Møller, and Heidi S. Christensen

The paper has been submitted to
Journal of Royal Statistical Society: Series C (Applied Statistics).

The layout has been revised.

Abstract

For modelling the location of pyramidal cells in the human cerebral cortex we suggest a hierarchical point process in $\mathbb{R}^3$*. The model consists first of a generalised shot noise Cox process in the xy-plane, providing cylindrical clusters, and next of a Markov random field model on the z-axis, providing either repulsion, aggregation, or both within specified areas of interaction. Several cases of these hierarchical point processes are fitted to two pyramidal cell datasets, and of these a model allowing for both repulsion and attraction between the points seem adequate.*

1 Introduction and conclusions

The structuring of neurons in the human brain is a subject of great interest since abnormal structures may be linked to certain neurological diseases (see Casanova, 2007; Esiri and Chance, 2006; Casanova et al., 2006; Buxhoeveden and Casanova, 2002). A specific structure that has been extensively studied in the biological literature is the so called ’minicolumn’ structure of the cells in the cerebral cortex (see Buxhoeveden and Casanova, 2002; Rafati et al., 2016, and references therein). Rafati et al. (2016) characterised these minicolumns as ‘linear aggregates of neurons organised vertically in units that traverse the cortical layer II–VI, and have in humans a diameter of 35–60 µm and consist typically of 80–100 neurons’.

1.1 Data

In this paper we analyse the structuring of pyramidal cells, which make up approximately 75 % to 80 % of all neurons (Buxhoeveden and Casanova, 2002) and are pyramid shaped cells, where the apical dendrite extends from the top of the pyramid. Specifically, the paper is concerned with two datasets consisting of the locations and orientations of pyramidal cells in a section of the third, respectively, fifth layer of Brodmann’s fourth area of the human cerebral cortex. Here, each location is a three-dimensional coordinate representing the centre of a pyramidal cell’s nucleolus and each orientation is a unit vector representing the apical dendrite’s position relative to the corresponding nucleolus.

Figure B.1 shows the two point pattern datasets of 634 and 548 nucleolus locations which will be referred to as `L3` and `L5`, respectively (for plot of the orientations for `L3`, see Møller et al., 2019). Note that the observation window W for the cell locations is a rectangular region with side lengths 492.70 µm, 132.03 µm, and 407.70 µm for `L3` and 488.40 µm, 138.33 µm, and 495.40 µm for `L5`. Notice also that the nucleolus locations are recorded such that the z-axis is perpendicular to the pial surface of the brain. In accordance to the minicolumn hypothesis, this implies that the minicolumns extend parallel to the z-axis.

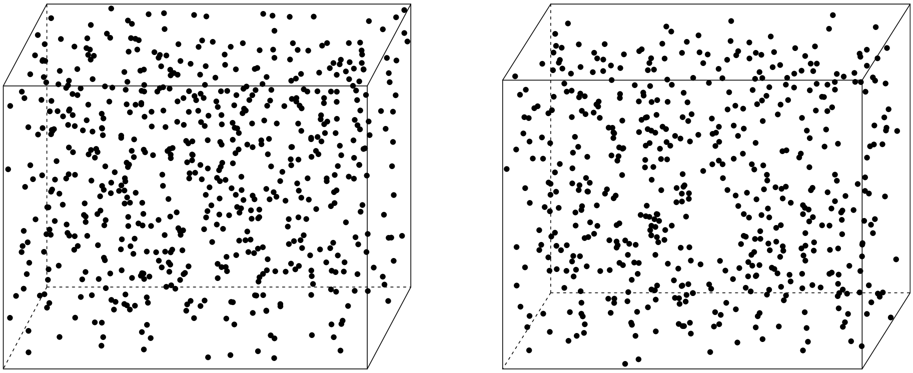


Figure B.1: Visualisations of the nucleolus locations for datasets `L3` (left) and `L5` (right).

1.2 Background and purpose

Møller et al. (2019) found independence between locations and orientations for `L3` meaning that the two components may be modelled separately; the same conclusion has afterwards been drawn for `L5`. As they also found a suitable inhomogeneous Poisson process model for the orientations, and since it is hard by eye to see much structure in the point patterns shown in Figure B.1, the focus of this paper is on modelling the nucleolus locations. In particular, we aim at modelling the nucleolus locations for `L3` respective `L5` by a spatial point process with a columnar structure and discuss to what extent this relates to the minicolumn hypothesis. Note that for the two datasets we use the same notation X for the spatial point process, and we view X as a random finite subset of W.

To the best of our knowledge the so-called Poisson line cluster point process (see Møller et al., 2016) is the only existing point process model for modelling columnar structures. This model was considered by Rafati et al. (2016) in connection to the pyramidal nucleolus locations, but was not fitted to data. For each point pattern considered in the present paper, we notice later that a more advanced model than the Poisson line cluster point process is needed; below we describe such a model for X.

1.3 Hierarchical point process models

Briefly, we consider a hierarchical model for X (further details are given in Sections 3–5), noting that the observation window is a product space, $W = W_{xy} \times W_z$, where W_{xy} is a rectangular region in the xy-plane and W_z is an interval on the z-axis. First, we model the point process X_{xy} given by the projection of X onto the xy-plane; second, conditioned on X_{xy}, we model the vector X_z consisting of the z-coordinates of the points in X. Note that the dimension of X_z agrees with the number of points in X_{xy} and is denoted by n.

The model for X_{xy}

For X_{xy} we consider the restriction of a cluster point process to W_{xy} defined briefly as follows (further details are given in Sections 4–5). Let $\Phi \subset \mathbb{R}^2$ be a stationary point process with intensity $\kappa > 0$, and associate to each point $(\xi, \eta) \in \Phi$ a point process $X_{(\xi,\eta)} \subset \mathbb{R}^3$ that is concentrated around the line in $\mathbb{R}^3$ which is perpendicular to the xy-plane, with intersection point $(\xi, \eta, 0)$. We refer to $X_{(\xi,\eta)}$ as the cylindrical cluster associated to (ξ, η). Let $P_{xy}(X_{(\xi,\eta)} \cap W)$ denote the projection onto the xy-plane of the observed part of the cylindrical cluster. For short we refer to the non-empty $P_{xy}(X_{(\xi,\eta)} \cap W)$ as the projected cluster with centre point (ξ, η). Then we let

$$X_{xy} = \bigcup_{(\xi,\eta)\in\Phi} P_{xy}(X_{(\xi,\eta)} \cap W).$$

Further, conditional on Φ, we assume that the projected clusters are independent and each non-empty $P_{xy}(X_{(\xi,\eta)} \cap W)$ is distributed as the intersection of W_{xy} with a finite planar Poisson process translated by the centre point (ξ, η); this Poisson process has intensity function $a\alpha f$, where a is the length of the interval W_z, $\alpha > 0$ is a parameter, and f is the probability density function of a bivariate zero-mean isotropic normal distribution with standard deviation $\sigma > 0$. Thus, ignoring boundary effects, αa is the expected size (or number of points) of a projected cluster and σ controls the spread of points in a projected cluster. Specifically, we let first Φ be a planar stationary Poisson process and second a stationary determinantal point process (Lavancier et al., 2015), since we observe in the first case a very low expected number of points in a projected cluster and because in the second case we want a repulsive model in order to obtain less overlap between the projected clusters.

The special case with Φ a planar stationary Poisson process and X_z a homogeneous binomial point process (that is, the n points in X_z are independent and uniformly distributed on W_z) which is independent of X_{xy} corresponds to a degenerate case of a Poisson line cluster point process as considered in Møller et al. (2016). This becomes clear in Section 4.

The model for X_z conditioned on X_{xy}

We consider several other cases than a homogeneous binomial point process for X_z which is independent of X_{xy}. In general, conditioned on $X_{xy} =$

$\{(x_i, y_i)\}_{i=1}^n$, we propose a Markov random field model, where the conditional probability density function of X_z is of the form

$$f((z_i)_{i=1}^n \mid (x_i, y_i)_{i=1}^n) \propto \gamma_1^{s_{B_1,\theta_1}((z_i)_{i=1}^n \mid (x_i,y_i)_{i=1}^n)} \gamma_2^{s_{B_2,\theta_2}((z_i)_{i=1}^n \mid (x_i,y_i)_{i=1}^n)} \\ \times \mathbb{I}(\|(x_i, y_i, z_i) - (x_j, y_j, z_j)\| > h \text{ for } 1 \leq i < j \leq n), \tag{B.1}$$

with notation defined as follows. We consider $\{(x_i, y_i, z_i)\}_{i=1}^n$ as a realisation of X, where (x_i, y_i) is the xy-point associated to z_i, the realisation of the i'th point in X_z (as a technical detail, unless X_z is a binomial point process, (B.1) is not invariant under permutations of $z_1, \ldots, z_n$ since we have associated (x_i, y_i) to z_i, so we cannot view (B.1) as the density of a point process where we are conditioning on the number of points). Note that the right hand side in (B.1) is an unnormalised density and e.g. $(z_i)_{i=1}^n$ is short hand notation for $(z_1, \ldots, z_n)$. We let $\mathbb{I}(\cdot)$ be the indicator function. Further, $\gamma_1 > 0$, $\gamma_2 > 0$, and $h \geq 0$ are unknown parameters; if $h > 0$, it is a hard core parameter ensuring a minimum distance h between all pair of points in X; for the pyramidal cell data it seems natural to include a hard core condition since cells cannot overlap; and when $\gamma_1 = \gamma_2 = 1$ and $h = 0$, the conditional model simply reduces to the homogeneous binomial point process. Furthermore, for $k = 1, 2$,

$$s_{B_k,\theta_k}((z_i)_{i=1}^n \mid (x_i, y_i)_{i=1}^n) = \sum_{1 \leq i < j \leq n} \mathbb{I}((x_i, y_i, z_i) \in B_k(x_j, y_j, z_j; \theta_k)),$$

where $B_k(x, y, z; \theta_k) \subset \mathbb{R}^3$ is an interaction region, with centre of mass (x, y, z) and a 'size and shape parameter' θ_k, that determines the interaction between points. It is additionally assumed that the hard core ball, given by the three-dimensional closed ball of radius h and centre (x, y, z) does not contain $B_1(x, y, z; \theta_1)$ or $B_2(x, y, z; \theta_2)$. Finally, it is assumed that the symmetry condition

$$(x_i, y_i, z_i) \in B_k(x_j, y_j, z_j; \theta_k) \quad \text{if and only if} \quad (x_j, y_j, z_j) \in B_k(x_i, y_i, z_i; \theta_k)$$

and the disjointness condition

$$B_1(x, y, z; \theta_1) \cap B_2(x, y, z; \theta_2) = \emptyset$$

are satisfied.

These conditions ensure that we can view X_z conditioned on X_{xy} as a Markov random field with second order interactions: for $1 \leq i < j \leq n$, two z-coordinates z_i and z_j interact (in Markov random field terminology, z_i and z_j are neighbours) if and only if $\|(x_i, y_i, z_i) - (x_j, y_j, z_j)\| \leq h$ (that is, the hard core condition is not satisfied, which happens with probability 0) or (x_i, y_i, z_i) lies within the region of interaction of z_j given by the union of $B_1(x_j, y_j, z_j; \theta_1)$ and $B_2(x_j, y_j, z_j; \theta_2)$ (here the symmetry condition is needed

	$\hat{\kappa}$	$\hat{\sigma}$	$\widehat{\alpha a}$
L3	0.0040	5.45	2.42
L5	0.0021	6.53	3.87

Table B.1: Minimum contrast estimates for the final model of X_{xy} (the DLCPP model in Section 5.1) for the datasets L3 and L5.

to ensure that we can interchange the roles of i and j). The interaction can either cause repulsion/inhibition or attraction/clumping of the points in X depending on whether $\gamma_k < 1$ or $\gamma_k > 1$ for $k = 1, 2$. Thus, apart from the hard core condition, the model allows for both repulsion and attraction but within different interaction regions B_1 and B_2.

The final hierarchical model and results

At the end of the paper (Section 5) we obtain a satisfactory fit of the following hierarchical model, with the following interpretation of the estimated parameters.

First, the model for X_{xy} is given as in Section 1.3 where the centre process Φ is a most repulsive determinantal point process (as detailed in Section 5.1). The parameter estimates are given in Table B.1, where the estimated expected cluster size $\widehat{\alpha a}$ is much smaller than expected for a minicolumn when restricting it to the observation window – provided the minicolumn hypothesis is true; cf. personal communication with Jens R. Nyengaard. So we neither claim that we have a fitted model for minicolumns nor that the minicolumn hypothesis is true. Instead we have fitted a model with cylindrical clusters: from Table B.1 we see, if $|W_{xy}|$ denotes the area of W_{xy}, the estimated number of projected clusters is $|W_{xy}|\hat{\kappa}$, which is approximately 260 for L3 and 142 for L5; the estimated expected size of a projected cluster is only 2.42 for L3 and 3.87 for L5.

Second, the model of X_z conditioned on X_{xy} has cylindrical interaction regions as illustrated in Figure B.2, and (B.1) is the pairwise interaction Markov random field density

$$f((z_i)_{i=1}^n \mid (x_i, y_i)_{i=1}^n) \propto \prod_{1 \le i < j \le n} \mathbb{I}(\|(x_i, y_i, z_i) - (x_j, y_j, z_j)\| > h) \\ \times \gamma_1^{\mathbb{I}(\|(x_i,y_i)-(x_j,y_j)\| \le r_1, |z_i - z_j| \le t_1)} \\ \times \gamma_2^{\mathbb{I}(\|(x_i,y_i)-(x_j,y_j)\| \le r_2, t_1 < |z_i - z_j| \le t_2)},$$

where $\gamma_1 > 0$ and $\gamma_2 > 0$ are interaction parameters and $0 < r_2 \le r_1$ and $0 < t_1 < t_2$ are parameters which determine the 'range of interaction' such

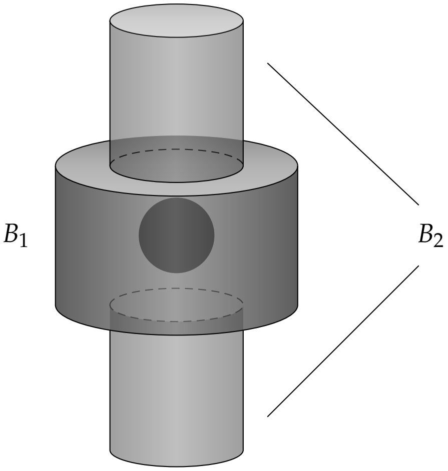


Figure B.2: Visualisation of the hard core region ball (in dark) and the cylindrical interaction regions B_1 (the cylinder) and B_2 (the union of the two elongated cylinders) used in our final model for `L3`.

	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{h}$	$\hat{r}_1$	$\hat{t}_1$	$\hat{r}_2$	$\hat{t}_2$
L3	0.41	1.78	6.25	20	11.5	11	35.5
L5	0.51	1.68	6.77	24.25	15.5	14.75	37.25

Table B.2: Pseudo likelihood estimates of the final model (model 5 from Table B.4 in Section 5.2) for the datasets `L3` and `L5`.

that $h < \sqrt{t_k^2 + r_k^2}$ for $k = 1, 2$. The restrictions on r_1, r_2, t_1, and t_2 are empirically motivated by use of functional summaries as detailed in Section 5.2. The final fitted model have parameter estimates as displayed in Table B.2 where most notably $\hat{\gamma}_1 < 1$ and $\hat{\gamma}_2 > 1$. In particular the final fitted model is in accordance to the empirical findings as noted later when the so-called cylindrical K-function of Figure B.3 is discussed: we have modelled repulsion within stunted cylinders (corresponding to B_1) and aggregation within elongated cylinders (corresponding to B_2), see again Figure B.2. Moreover, the estimated hard core $\hat{h}$ is greater than 6 µm, which is in accordance with 'distance between the nucleolus and the membrane of a pyramidal cell' (personal communication with Jens R. Nyengaard). Note that the hard core ball is much smaller than the interaction region B_1: $2\hat{h}$ (the diameter of the hard core ball) is about half as small as $2\hat{t}_1$ (the height of B_1). Finally, comparing Tables B.1–B.2, we note that the two 'clustering parameters' $2\hat{\sigma}$ and $\hat{r}_2$ are of the same order.

In conclusion, for each dataset we have fitted a rather complex hierarchical point process model describing columnar structures of the nucleolus locations. This model included repulsion between nucleolus locations given by a hard core condition on a small scale and a stunted cylindrical interaction region on a larger scale, as well as clustering between nucleolus locations given by an elongated cylindrical interaction region.

1.4 Model fitting

In Møller et al. (2016) parameter estimation for the degenerate PLCPP model was simply done by a moment based procedure which included minimisation of a certain contrast between a theoretical second order moment functional summary and its empirical estimate. In the present paper we use a similar minimum contrast procedure for estimating the parameters of models for X_{xy}. For the models of X_z conditioned on X_{xy} we find it convenient to use a maximum pseudo likelihood procedure as detailed in Section 5.2. Moreover, each fitted model is evaluated by considering informative global extreme rank length (GERL) envelope procedures (Mrkvička et al., 2018; Myllymäki et al., 2017) for various functional summaries.

1.5 Outline

The remainder of this paper explains how we arrive at the final model given in Section 1.3 after fitting several other models. In Section 2 we introduce some basic concepts and definitions needed for the models in the subsequent sections. In Section 3 we investigate how the nucleolus locations deviate from complete spatial randomness (that is, when X is a homogeneous Poisson process), and in Section 4 we also notice a deviation from a fitted degenerate PLCPP model. Finally, in Section 5 we introduce and fit various generalisations of the degenerate PLCPP model as briefly described in Sections 1.3–1.4.

2 Preliminaries

The point processes X, X_{xy}, and X_z introduced above are viewed as the restriction to the bounded sets W, W_{xy}, and W_z of a locally finite point process $Y \subset \mathbb{R}^d$ with $d = 3, 2, 1$, respectively. Briefly speaking, this means that Y is a random subset of $\mathbb{R}^d$ satisfying that $Y_B = Y \cap B$ is finite for any bounded set $B \subset \mathbb{R}^3$; for a more rigorous definition of point processes, see e.g. Daley and Vere-Jones (2003) or Møller and Waagepetersen (2004). Below we recall a few basic statistical tools needed in this paper, using the generic notation Y for a locally finite point process defined on $\mathbb{R}^d$ (apart from the cases above, we have in mind that Y could also be the centre process Φ from Section 1.3).

2.1 Moments

For each integer $n \geq 1$, to describe the n'th order moment properties of Y, we consider the so-called n'th order intensity function $\lambda^{(n)} : (\mathbb{R}^d)^n \to [0, \infty)$ given that it exists. This means that for any pairwise distinct and bounded Borel sets $B_1, \ldots, B_n \subset \mathbb{R}^d$,

$$\mathrm{E}\left[n(Y_{B_1}) \cdots n(Y_{B_n})\right] = \int_{B_1} \cdots \int_{B_n} \lambda^{(n)}(x_1, \ldots, x_n) \,\mathrm{d}x_1 \cdots \mathrm{d}x_n$$

is finite, where $n(Y_B)$ denotes the cardinality of Y_B.

The first order intensity function $\lambda^{(1)} = \lambda$ is of particular interest and is simply referred to as the intensity function. Heuristically, $\lambda(u)\,\mathrm{d}u$ can be interpreted as the probability of observing a point from Y in the infinitesimal ball of volume $\mathrm{d}u$ centred at u. If the intensity function $\lambda(\cdot) \equiv \lambda$ is constant, then $\lambda|B| = \mathrm{E}\left[n(Y_B)\right]$ for any bounded Borel set $B \subset \mathbb{R}^d$, where $|\cdot|$ is the Lebesgue measure. In this case Y is said to be homogeneous and otherwise inhomogeneous. Clearly, stationarity of Y (meaning that its distribution is invariant under translations in $\mathbb{R}^d$) implies homogeneity.

2.2 Functional summaries

In order to determine an appropriate model for an observed point pattern, we consider functional summaries, which reflect/summarise different properties of the point pattern and are useful for model fitting and control. The main examples are considered below.

To summarise the second order moment properties, it is custom to consider the pair correlation function (PCF), g, which is defined as the ratio of the second and first order intensity function, that is,

$$g(x_1, x_2) = \frac{\lambda^{(2)}(x_1, x_2)}{\lambda(x_1)\lambda(x_2)}, \qquad x_1, x_2 \in \mathbb{R}^d.$$

Heuristically, $g(x_1, x_2)$ can be interpreted as the probability of simultaneously observing a point from X in each of the two infinitesimal balls of volume $\mathrm{d}x_1$ and $\mathrm{d}x_2$ centred at respectively x_1 and x_2 relative to the probability of independently observing a point in the two infinitesimal balls. The PCF is said to be stationary when (with abuse of notation) $g(x_1, x_2) = g(x_1 - x_2)$ and isotropic when $g(x_1, x_2) = g(\|x_1 - x_2\|)$.

If the PCF is stationary, it is closely related to the so-called second order reduced moment measure, $\mathcal{K}$, given by

$$\mathcal{K}(B) = \int_B g(x)\,\mathrm{d}x,$$

where $B \subset \mathbb{R}^d$ is a Borel set (see Møller and Waagepetersen, 2004). If Y is stationary and B has centre of mass at the origin of $\mathbb{R}^d$, then $\lambda\mathcal{K}(B)$ can be interpreted as the expected number of further points falling within B given that Y has a point at the origin; and when considering scalings of B, we refer to B as a structuring element. The simplest example occurs when B is a ball centred at the origin and with radius $r > 0$; then $K(r) = \mathcal{K}(B)$ becomes the K-function introduced by Ripley (1976); and often we instead consider a transformation of the K-function, which is called the L-function and defined by $L(r) = (K(r)/\omega_d)^{1/d}$, where ω_d is the volume of the d-dimensional unit ball. In particular, if Y is a stationary Poisson process, then $L(r) = r$.

For detecting cylindrical structures, Møller et al. (2016) introduced the cylindrical K-function which corresponds to $\mathcal{K}(B)$ when B is a cylinder of height $2t$, base-radius r, and centre of mass at the origin. Note that Ripley's K-function depends only on one argument, r, while the cylindrical K-function depends both on r, t, and the direction of the cylinder. However, when $d = 3$ and since the minicolumns are expected to extend along the z-axis, we only consider cylinders extending in this direction, effectively reducing the number of arguments to two.

We will also consider the commonly used F-, G-, and J-functions when performing model control; see van Lieshout and Baddeley (1996) for definitions. Briefly, if Y is stationary, $F(r)$ is the probability that Y has a point within distance $r > 0$ from a fixed location in $\mathbb{R}^d$; $G(r)$ is the probability that Y has another point within distance $r > 0$ from an arbitrary fixed point in Y; and $J(r) = (1 - G(r))/(1 - F(r))$ when $F(r) < 1$.

The functional summaries will in the following be used both for model fitting as described in Section 2.3 and for model checking using GERL envelope procedures as mentioned in Section 1.4. In the GERL envelope procedure, the distribution of the empirical functional summary under the hypothesis of interest is estimated by simulations. The procedure is a refinement of the global rank envelope procedure (Myllymäki et al., 2017), where it is recommended to use 2499 simulations for a single one-dimensional functional summary and at least 9999 simulations for a single two-dimensional functional summary (Mrkvička et al., 2016). However, we consider a concatenation of the L-, G-, F-, and J-functions, as well as the cylindrical K-function in which case Mrkvička et al. (2017) recommend using more simulations. Particularly for a concatenation of k one-dimensional summary functions they recommend using $k \times 2499$ simulations. We do however have a different setup since we are concatenating both one- and two-dimensional summary functions. For the GERL envelope procedure, Mrkvička et al. (2018) suggest that a lower number of simulations may be enough. Therefore, we use 9999 simulations. Since we consider a concatenation of one- and two-dimensional functional summaries, we ensure that each of the functional summaries are weighted equally in the GERL envelope test by evaluating them at the same number of arguments (Mrkvička et al., 2017). Specifically we consider 64^2 r-values for each of the L-, G- F-, and J-functions and a square grid over 64 r-values and

64 t-values for the cylindrical K-function.

2.3 Minimum contrast estimation

For parametric point process models, minimum contrast estimation is a computationally simple fitting procedure introduced by Diggle and Gratton (1984) that is applicable when a closed form expression of a functional summary, T, exists. The idea is to minimise the distance from the theoretical function T to its empirical estimate $\hat{T}$ for the data. Specifically, if T depends on the parameter vector θ and is a function of 'distance' $r > 0$ (as for example in case of Ripley's K-function), the minimum contrast estimate of θ is given by

$$\hat{\theta} = \operatorname{argmin}_\theta \int_{r_{\min}}^{r_{\max}} \left| T(\theta, r)^q - \hat{T}(r)^q \right|^p \mathrm{d}r, \tag{B.2}$$

where $r_{\min} < r_{\max}$, q, and p are positive tuning parameters. General recommendations on q are given in Guan (2009) and Diggle (2014), when $T(r) = g(r)$ or $T(r) = K(r)$. Unless otherwise stated, we let $p = 2$, $q = 1/4$, $r_{\min} = 0$, and $r_{\max}$ be one fourth of the shortest side length of the relevant observation window (the rectangular window W_{xy} in our case).

When the PCF has a closed form expression, alternative estimation procedures can be used, including the second order composite likelihood (see Guan, 2006; Waagepetersen, 2007), adapted second order composite likelihood (see Lavancier et al., 2018), and Palm likelihood (see Ogata and Katsura, 1991; Prokešová et al., 2016; Baddeley et al., 2016).

3 Complete spatial randomness

The most natural place to begin our point pattern analysis is by testing whether a homogeneous Poisson process X with intensity $\lambda > 0$ (we then view Y as a stationary Poisson process with the same intensity), also called complete spatial randomness (CSR), adequately describe each nucleolus point pattern dataset. Recall that this means that $n(X)$ is Poisson distributed with parameter $\lambda|W|$ and conditional on $n(X)$ the points in X are independent and uniformly distributed within W. Even when CSR is not an appropriate model, deviations from the model can be useful for determining whether the points of a homogeneous point pattern tend to e.g. attract or repel each other.

The CSR model is fully specified by its intensity, which naturally is estimated by $n(X)/|W|$, which is equal to 2.37×10^{-5} for `L3` and 1.63×10^{-5} for `L5`. For this fitted model Figure B.3 summarises the results of the GERL envelope procedure based on the concatenation of the empirical L-, G-, F-, J-, and cylindrical K-functions as discussed in Section 2.2. Particularly, the left column depicts the part concerning the empirical functional summaries

$\hat{L}(r) - r$, $\hat{G}(r)$, $\hat{F}(r)$, and $\hat{J}(r)$ along with the corresponding 95% envelope. The right column depicts the empirical cylindrical *K*-function along with the areas at which it falls outside the 95% envelope. It is observed that the empirical functional summaries $\hat{L}$, $\hat{F}$, and $\hat{J}$ fall strictly outside the envelope for midrange values of r in a manner that indicates repulsion between points at this range. For small and large r-values the observed point patterns resemble the Poisson process. This behaviour could suggest a kind of clustering, where clusters of points from a Poisson process are somewhat separated. The separation of these clusters seems to be more pronounced for `L3` than for `L5`. Further, in the right column of Figure B.3, the empirical cylindrical *K*-function falls above the upper global rank envelopes for cylinders that have a height larger than approximately 35 µm for both datasets and a base radius of approximately 5 µm to 15 µm for `L3` and 5 µm to 20 µm for `L5`. Furthermore, the observed cylindrical *K*-functions falls below the lower 95% GERL envelope for cylinders with a height of approximately 10 µm to 30 µm and a base radius larger than 5 µm. Hence, for elongated cylinders extending in the z-direction, we tend to see more points in the data than we expect under CSR, while for stunted cylinders we tend to see fewer points. This seems to be in correspondence with columnar structures where the columns extend in the z-direction.

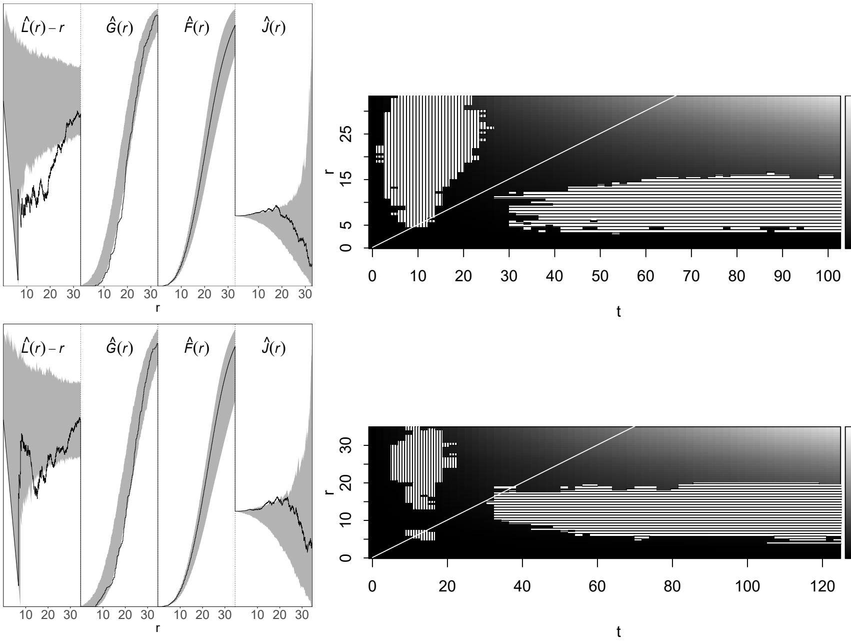


Figure B.3: Results of the GERL envelope procedure under CSR based on a concatenation of the empirical *L*-, *G*-, *F*-, *J*-, and cylindrical *K*-functions. Left: concatenation of the one-dimensional empirical functional summaries for the data (solid line) together with 95% envelopes (grey region); for ease of visualisation, the functions have been scaled. Right: empirical cylindrical *K*-function (grey scale) where shaded vertical/horizontal lines indicate that the function falls above/below the 95% envelope. The white line indicates the values for which the cylinder height is equal to the base diameter. Top: results for the dataset `L3`. Bottom: results for the dataset `L5`.

4 The degenerate Poisson line cluster point process

Møller et al. (2016) presented the so-called Poisson line cluster point process (PLCPP) which is useful for modelling columnar structures. Specifically, we consider a degenerate PLCPP $Y \subset \mathbb{R}^3$ constructed as follows.

1. Generate a stationary Poisson process $\Phi = \{(\xi_i, \eta_i)\}_{i=1}^{\infty} \subset \mathbb{R}^2$ with finite intensity $\kappa > 0$. Each point $(\xi_i, \eta_i) \in \Phi$ corresponds to an infinite line l_i in $\mathbb{R}^3$ which is perpendicular to the xy-plane, that is, $l_i = \{(\xi_i, \eta_i, z) \,|\, z \in \mathbb{R}\}$.
2. Conditional on Φ, generate independent stationary Poisson processes $L_1 \subset l_1, L_2 \subset l_2, \ldots$ with identical and finite intensity $\alpha > 0$.
3. Generate point processes $X_1, X_2, \ldots \subset \mathbb{R}^3$ by independently displacing the points of $L_1, L_2, \ldots$ by the zero-mean isotropic normal distribution with standard deviation $\sigma > 0$.
4. Finally, set $Y = \bigcup_{i=1}^{\infty} X_i$ and $X = Y_W$.

Some comments to the construction in items 1–4 are in order.

In the general definition of the PLCPP in Møller et al. (2016), the lines $l_1, l_2, \ldots$ are modelled as a stationary Poisson line process. That is, the lines are not required to be perpendicular to the xy-plane nor does the Poisson line process need to be degenerate (meaning that the lines are not required to be mutually parallel). Further, the dispersion density (used in item 3) can be arbitrary. However, the construction is still such that Y becomes stationary. Furthermore, it turns out that it does not matter whether we consider a three-dimensional normal distribution for displacements in in item 3 or a bivariate normal distribution with displacements of the xy-coordinates for the points of $L_1, L_2, \ldots$.

Returning to the degenerate PLCPP of items 1–4, we imagine that each X_i is a cylindrical cluster of points around the line l_i, where these cylindrical clusters are parallel to the z-axis. Furthermore, the interpretation of the parameters κ, α, and σ in terms of a Poisson cluster point process is similar to that in Section 1.3 except that we now also consider lines not intersecting W: if Y as defined by items 1–4 is restricted to a subset $S \subset \mathbb{R}^3$ bounded by two planes parallel to the xy-plane, for specificity $S = \{(x, y, z) \in \mathbb{R}^3 \,|\, z \in W_z\}$, this restricted point process can be seen as a (modified) Thomas process (see Thomas, 1949; Møller and Waagepetersen, 2004) on $\mathbb{R}^2$ along with independent z-coordinates following a uniform distribution on W_z.

To see this, first note that conditional on $\Phi = \{(\xi_i, \eta_i)\}_{i=1}^{\infty}$ and for all $i = 1, 2, \ldots$, X_i is a Poisson process in $\mathbb{R}^3$ with intensity function $\lambda_i((x, y, z)) =$

	$\hat{\kappa}$	$\hat{\sigma}$	$\widehat{\alpha a}$
L3	0.027	2.86	0.36
L5	0.0085	4.58	0.95

Table B.3: Minimum contrast estimates of the degenerate PLCPP.

$\alpha f(x - \xi_i, y - \eta_i)$, where f is the probability density function of the bivariate isotropic normal distribution given in item 3. In turn, this implies that Y conditioned on Φ is a Poisson process in $\mathbb{R}^3$ with intensity function $\sum_{i=1}^{\infty} \lambda_i((x, y, z))$. Further, since $\lambda_i(x, y, z) = \lambda_i(x, y)$ does not depend on z for all $i = 1, 2, \ldots$, the projection of Y_S onto the xy-plane, $P_{xy}(Y_S)$, conditioned on Φ is a Poisson process with intensity $a \sum_{i=1}^{\infty} \lambda_i(x, y)$, where a is the length of the interval W_z. Since Φ is a stationary Poisson process, $P_{xy}(Y_S)$ is a Thomas process with centre process intensity κ and expected cluster size αa (that is, the expected number of points in $X_i \cap S$). Finally, from items 2–4 it follows that the z-coordinates of X_z are independent and uniformly distributed on W_z, and they are independent of X_{xy}.

Consequently, simulating $X = Y_W$ is straightforwardly done by simulating a Thomas point process (on a larger set than W_{xy} in order to avoid boundary effects) along with independent uniform z-coordinates on W_z. For simulating the Thomas point process we apply standard software from the `R`-package `spatstat` (Baddeley et al., 2016). Similarly, fitting a degenerate PLCPP to a realisation of X is simply a matter of fitting a Thomas process to the point pattern consisting of the xy-coordinates of the points in that realisation. Since the K-function of the Thomas process has a closed form expression, the model can be fitted using minimum contrast estimation with $T(r) = K(r)$ in (B.2). Table B.3 summarises the parameter estimates, where most notably the expected cluster size $\widehat{\alpha a}$ is < 1 for both `L3` and `L5`. Understanding each cylindrical cluster within W as (a part of) a minicolumn, 'these parameter estimates result in very unnatural models for the datasets, since each minicolumn within W is expected to consist of less than one point' (personal communication with Jens R. Nyengaard).

Despite the fact that the fitted degenerate PLCPP models are somewhat unnatural and hardly can be interpreted as a model with (hypothesised) minicolumnar structures, GERL envelope procedure based on a concatenation of the F-, G-, and J-functions show that the Thomas process suitably fit the projected locations with a p-value of 0.76 for `L3` and 0.87 for `L5`. However, results from the concatenated GERL envelope procedure described in Section 2.2 indicated that the model did not suitably describe the three-dimensional nucleolus locations with a p-value of 10^{-4} for both `L3` and `L5`. Specifically, Figure B.4 shows the empirical cylindrical K-function and indicates where it deviates from the 95% envelope. Clearly, the model does

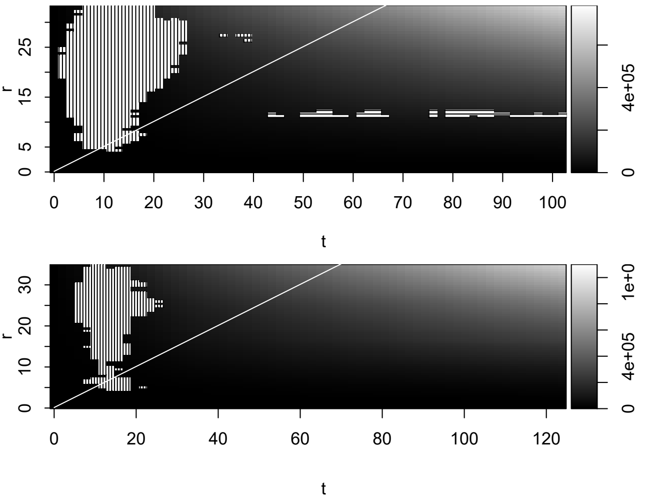


Figure B.4: Empirical estimates of the cylindrical K-function (grey scale) where shaded vertical/horizontal lines indicate that the function falls above/below the 95% GERL envelope under the fitted degenerate PLCPP and based on the concatenation described in Section 2.2. The white line indicates the values for which the cylinder height is equal to the base diameter. Top: results for the dataset `L3`. Bottom: results for the dataset `L5`.

account for some of the columnarity of the data as opposed to CSR, but the empirical cylindrical K-function for `L3` still falls above the 95% envelope. Furthermore, the empirical cylindrical K-function for both datasets falls below the 95% envelope similar to what was seen under CSR, indicating a lack of regularity, which in fact is supported by the one-dimensional functional summaries (not shown). This could suggest that the cylindrical clusters should be more distinct; motivating us to generalise the degenerate PLCPP model as in the following section.

5 A generalisation of the degenerate PLCPP

As some but not all features of the data were explained by the degenerate PLCPP fitted in Section 4, we propose in this section two generalisations as follows.

1. The centre process Φ is a planar stationary point process.
2. X_z conditioned on X_{xy} follows a Markov random field model.

The first modification is straightforward and for this specific application we chose a repulsive centre process to obtain more distinguishable cylindrical clusters; this is detailed in Section 5.1. Further, the assumption of stationarity of Φ is made in order to apply a similar minimum contrast estimation procedure as in Section 4, so implicitly we make the assumption that the PCF or the K-function is expressible on closed form. For the second modification we suggest a conditional model inspired by the multiscale point process and particularly the Strauss hard core point process (see e.g. Møller and Waagepetersen, 2004) which will allow for further repulsion or even aggregation between the points; this is detailed in Section 5.2.

5.1 A determinantal point process model for the centre points

Consider a point process $Y \subset \mathbb{R}^3$ specified by items 1–4 in Section 4 with the exception that the centre process Φ now is an arbitrary stationary planar point process. Then, recalling the notation from Section 4, $P_{xy}(Y_S)$ is a planar Cox process (see Møller and Waagepetersen, 2004) and even a planar generalised shot-noise Cox process (see Møller and Torrisi, 2005) driven by the random intensity function $\Lambda(x,y) = a \sum_{i=1}^{\infty} \lambda_i(x,y)$ for $(x,y) \in \mathbb{R}^2$. Moreover, $P_{xy}(Y_S)$ corresponds to the Thomas process, but with a different centre point process (unless of course Φ is a stationary Poisson process).

In this section we focus on the case where Φ is a stationary determinantal point process (DPP; see Lavancier et al., 2015), in which case we will refer to Y as the determinantal line cluster point process (DLCPP). A DPP is defined in terms of its n'th order intensity function for $n = 1, 2, \ldots$: let $C : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{C}$ be a function and $\lambda^{(n)}$ the n'th order intensity function of Φ, then Φ is called a DPP with kernel C if

$$\lambda^{(n)}(x_1, \ldots, x_n) = \det[C](x_1, \ldots, x_n) \qquad \text{for } n = 1, 2, \ldots,\ x_1, \ldots, x_n \in \mathbb{R}^2,$$

where $\det[C](x_1, \ldots, x_n)$ is the determinant of the $n \times n$ matrix with (i,j)'th entry $C(x_i, x_j)$. For further details on DPPs, we refer to Lavancier et al. (2015) and the references therein. When Φ is a DPP, we call $P_{xy}(Y_S)$ a determinantal Thomas point process (DTPP). The DTPP is discussed to some extent in Møller and Christoffersen (2018), where a closed form expression of its PCF is found. Thus, the DLCPP can be fitted by fitting a DTPP to the projected data using a minimum contrast procedure (see Section 2.3).

For our data we want to obtain a DLCPP with as much repulsion as possible between the centre lines of the cylindrical clusters. Therefore, we let Φ be the 'most repulsive DPP' (in the sense of Lavancier et al., 2015), which is the jinc-like DPP given by the kernel $C(x_1, x_2) = \sqrt{\rho/\pi} J_1 \left(2\sqrt{\pi\rho}\|x_1 - x_2\|\right) / \|x_1 - x_2\|$, where J_1 is the first order Bessel function of the first kind and $\|\cdot\|$ denotes the usual planar distance (for more information on this particular DPP, see Lavancier et al., 2018; Biscio and Lavancier, 2016).

Simulation of the DTPP is done by first simulating a DPP with intensity κ (on a larger region than W_{xy} in order to avoid boundary effects), for which we use the functionality of `spatstat`, then secondly generating for each cluster a

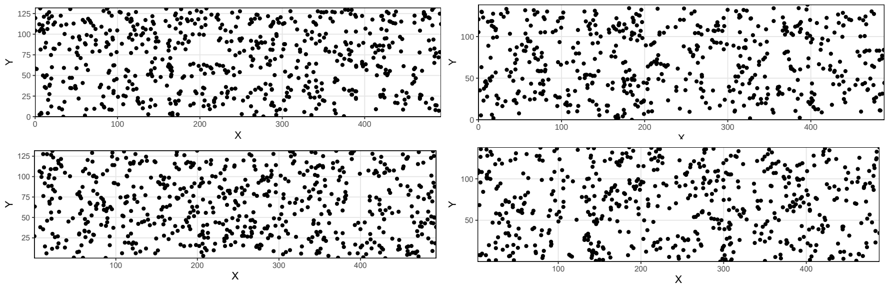


Figure B.5: Projection of observed nucleolus locations onto the xy-plane (left) and simulations from the fitted jinc-like DTPP (right) for the datasets `L3` (top) and `L5` (bottom).

Poisson distributed number of points with intensity αa, and finally displacing these points by a bivariate zero-mean isotropic normal distribution.

The parameter estimates of the jinc-like DTPP model were obtained by minimum contrast with $T(r) = g(r)$; see Table B.1 for the results and the accompanying discussion in Section 1.3. Despite the expectation under the minicolumn hypothesis of having much higher values of $\widehat{\alpha a}$ than in Table B.1 (see again Section 1.3), simulations of the fitted jinc-like DPP in the xy-plane seem in reasonable correspondence to the projected data; see Figure B.5. Furthermore, results from the GERL envelope procedure based on a concatenation of the F-, G-, and J-functions do not provide any evidence against the jinc-like DPP model for the projected points with p-values of 0.67 for `L3` and 0.83 for `L5`.

Since the jinc-like DTPP model fits the projected data well, we proceeded and added independent uniform z-coordinates on W_z to the simulations, thereby obtaining simulations of the jinc-like DLCPP. Figure B.6 summarises the result of the 95% GERL test based on the concatenation of functional summaries as described in Section 2.2. The left column depicts the part of the one-dimensional functional summaries along with 95% envelopes, while the right column shows the empirical cylindrical K-function along with shaded regions that indicate where it deviates from the corresponding envelope. These plots show that the models do not account for the regularity of the data. This leads us to our next generalisation in Section 5.2.

5.2 A Markov random field model for the z-coordinates

Motivated by the observations at the end of the previous section, in this section we propose to model the z-coordinates conditioned on the xy-coordinates by a pairwise interaction point process as given in (B.1). Thereby, our hierarchical model construction yields a more flexible model for X but we ignore edge effects in the sense that we have only specified a model for first $P_{xy}(Y_S)$

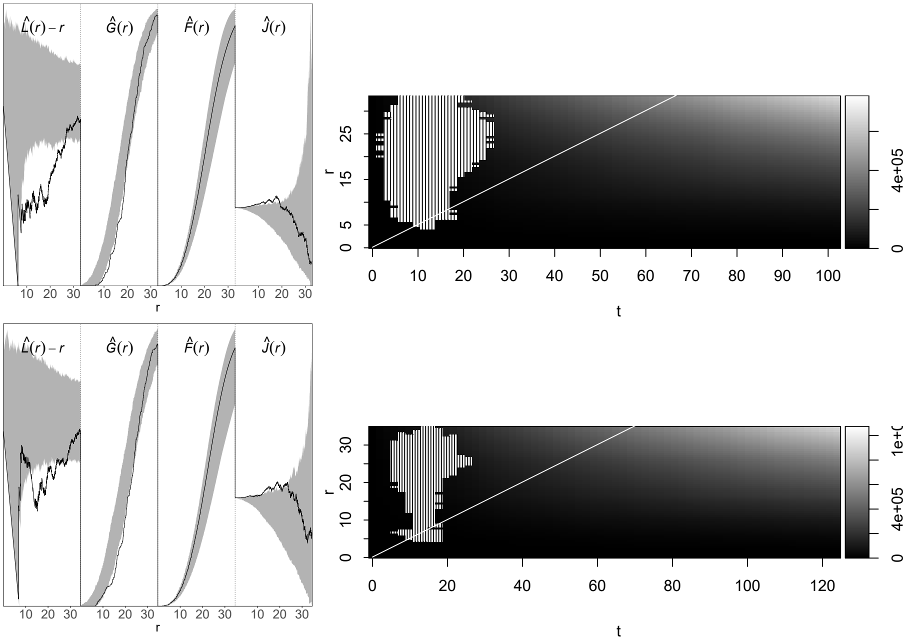


Figure B.6: Results of the GERL envelope procedure under the fitted DLCPP based on a concatenation of the empirical *L*-, *G*-, *F*-, *J*-, and cylindrical *K*-functions. Left: concatenation of the one-dimensional empirical functional summaries for the data (solid line) together with 95% envelopes (grey region); for ease of visualisation, the functions have been scaled. Right: empirical cylindrical *K*-function (grey scale) where shaded vertical/horizontal lines indicate that the function falls above/below the 95% envelope. The white line indicates the values for which the cylinder height is equal to the base diameter. Top: results for the dataset `L3`. Bottom: results for the dataset `L5`.

and second X_z conditioned on $X_{xy} = P_{xy}(Y_S) \cap W_{xy}$, thereby ignoring a possible influence of points in $Y \setminus W$ when (B.1) is used in the latter step (unless it specifies a binomial point process). This simplification is just made for mathematical convenience; indeed it would be interesting to construct a model taking edge effects into account so that Y becomes stationary, but we leave this challenging issue for future research. Below we first specify the ingredients of the conditional probability density function given in (B.1) for various models and discuss the overall conclusions, next describe how to find parameter estimates, and finally discuss how well the estimated models fit the data. Note that although we have not specified a stationary model for Y, it may still make sense to interpret plots of empirical cylindrical K-functions and $\hat{F}$, $\hat{G}$, $\hat{J}$, and $\hat{L}$-functions, since we have stationarity in the xy-plane and approximately stationarity in the z-direction (as the density (B.1) is invariant under 'translations of $(z_1, \dots, z_n)$ within W_z').

In our search for a suitable model for the nucleolus locations, we considered many special cases of (B.1). Table B.4 summarises five selected models, where $b((x,y,z);r)$ is the ball with centre (x,y,z) and radius r, and where $c((x,y,z);r,t)$ and $d((x,y,z);r,t)$ denote the cylinder and double cone, respectively, with centre of mass at (x,y,z), height $2t$, base radius r, and extending in the z-direction. First, we considered model 1 which is a hard core model if $h > 0$ and one of the simplest ways of modelling regularity; note that model 1 with $h = 0$ is the binomial point process with a uniform density as considered in Section 4. Though accounting for small distance repulsion, when fitted to the data, model 1 turned out not to account for the repulsion at larger scales. Second, we considered model 2 which is a conditional Strauss model with a hard core condition (see Møller and Waagepetersen, 2004, and the references therein). For this model the scale of repulsion for the z-coordinates seemed too great for points with similar xy-coordinates, and therefore we found it natural to replace the spherical interaction region with a cylinder, yielding model 3. However, model 3 did not correct the problem, and continuing with a single region of interaction we next suggested model 4 with a region given by a cylinder minus a double cone. Model 4 does to a smaller degree penalise the occurrence of points with similar xy-coordinates. However, this model was not suitable either. Models 1–4 were discarded by GERL tests with extremely small p-values. Finally, we considered model 5 which is a more flexible model that allows for both repulsion and aggregation within cylinder shaped interaction regions, cf. the discussion in Section 1.3. For simplicity all the models were also considered without a hard core condition, that is $h = 0$, but was in every case found inadequate.

The likelihood function corresponding to (B.1) involves a normalising constant which needs to be approximated by Markov chain Monte Carlo methods. We propose an easier alternative based on the pseudo likelihood function (Besag, 1975) defined as follows when the data is given by $\{(x_i, y_i, z_i)\}_{i=1}^{n} \subset$

Model	γ_1	γ_2	$B_1(\,\cdot\,;\theta_1)$	$B_2(\,\cdot\,;\theta_2)$	θ_1
1	1	1	$\emptyset$	$\emptyset$	–
2	>0	1	$b(\cdot;r)$	$\emptyset$	$r > h$
3	>0	1	$c(\cdot;r,t)$	$\emptyset$	$r,t > 0$
4	>0	1	$c(\cdot;r,t) \setminus d(\cdot;r,t)$	$\emptyset$	$r,t > 0$
5	>0	>0	$c(\cdot;r_1,t_1)$	$c(\cdot;r_2,t_2) \setminus c(\cdot;r_1,t_1)$	$r_1,t_1 > 0$

Table B.4: Specific choices of the parameters $\gamma_1,\gamma_2,\theta_1,\theta_2$ and the interaction regions $B_1(\cdot;\theta_1), B_2(\cdot;\theta_2)$ for five models given by the density (B.1). For each model, a hard core parameter $h \geq 0$ is included. Apart from the specified restrictions, it is required for models 2–5 that $B_1(\cdot;\theta_1) \not\subseteq b(\cdot;h)$ (for model 2 this means that $r > h$ as already indicated) and in addition for model 5 that $B_2(\cdot;\theta_2) \not\subseteq b(\cdot;h)$ where $\theta_2 = (r_2,t_2)$ with $r_1 \geq r_2 > 0$ and $t_2 > t_1$.

W. For $i = 1,\dots,n$, the i'th full conditional density associated to (B.1) is

$$\begin{aligned} &f(z_i \mid (z_1,\dots,z_{i-1},z_{i+1},\dots,z_n),(x_j,y_j)_{j=1}^n) \\ &\qquad = \mathbb{I}(\|(x_i,y_i,z_i) - (x_j,y_j,z_j)\| > h \text{ for } j \neq i)\gamma_1^{s_{1,i}}\gamma_2^{s_{2,i}}/c_i \end{aligned} \tag{B.3}$$

where we define

$$s_{k,i} = \sum_{j:\, j\neq i} \mathbb{I}\left((x_j,y_j,z_j) \in B_k((x_i,y_i,z_i);\theta_k)\right), \qquad k = 1,2,$$

and where the normalising constant is given by

$$\begin{aligned} c_i = \sum_{k=0}^{n-1}\sum_{l=0}^{n-1} \gamma_1^k\gamma_2^l \int_{W_z} &\mathbb{I}(\|(x_i,y_i,z) - (x_j,y_j,z_j)\| > h \text{ for } j \neq i) \\ &\times \mathbb{I}\Big(\sum_{j:\, j\neq i} \mathbb{I}\left((x_j,y_j,z_j) \in B_1((x_i,y_i,z);\theta_1)\right) = k \Big) \\ &\times \mathbb{I}\Big(\sum_{j:\, j\neq i} \mathbb{I}\left((x_j,y_j,z_j) \in B_2((x_i,y_i,z);\theta_2)\right) = l \Big)\,\mathrm{d}z. \end{aligned}$$

To estimate the model parameters we maximise the log pseudo likelihood given by

$$\begin{aligned} &LP(\gamma_1,\gamma_2,h,\theta_1,\theta_2) \\ &\qquad = \sum_{i=1}^n \log f(z_i \mid \{z_1,\dots,z_{i-1},z_{i+1},\dots,z_n\},(x_j,y_j)_{j=1}^n). \end{aligned} \tag{B.4}$$

Clearly, by (B.3) the maximum pseudo likelihood estimate (MPLE) $\hat{h}$ of h is the minimum distance between any distinct pair of points (x_i,y_i,z_i) and

(x_j, y_j, z_j) in the data. This in fact also corresponds to the maximum likelihood estimate. For $h = \hat{h}$ and for fixed θ_1 and θ_2, we easily obtain the following. For each of models 2–4, the MPLE of γ_1 exists if and only if $s_{1,i} \neq 0$ for some i, and then the log pseudo likelihood function is strictly concave with respect to $\log \gamma_1$. For model 5, the MPLE of (γ_1, γ_2) exists if and only if $s_{1,i} \neq 0$ for some i and $s_{2,j} \neq 0$ for some j, and then the log pseudo likelihood function is strictly concave with respect to $(\log \gamma_1, \log \gamma_2)$. Therefore, the (profile) log pseudo likelihood can be maximised by a combination of a grid search over θ_1 and θ_2 and numerical optimisation with respect to γ_1 and γ_2. Table B.2 shows the maximum pseudo likelihood estimates of model 5 for the two datasets, where for the numerical optimisation we used `optim` (a general-purpose optimisation function from the `R`-package `stats`).

Each of the five models in Table B.4 were fitted to `L3` and `L5` by finding the maximum pseudo likelihood estimate, and model checking was performed using GERL envelope procedures based on the concatenation of functional summaries as discussed in Section 2.2. For the fitted models, model 5 was the most appropriate with p-values of 0.34 for `L3` and 0.03 for `L5` when using the GERL envelope procedure; the 95% GERL envelope is visualised in Figure B.7. Thus no evidence is seen against the fitted models summarised in Table B.2 for `L3` while only slight evidence is present for `L5`. We note that for both datasets the fitted models are such that B_1 is a stunted cylinder and models repulsion since $\hat{\gamma}_1 < 1$, while $c(\cdot, r_2, t_2)$ is elongated and B_2 models aggregation, since $\hat{\gamma}_2 > 1$. Hence, when standing in some point $(x_1, y_1, z_1) \in X$ it is less likely to observe a z-coordinate if the corresponding xy-coordinates are similar to (x_1, y_1). Specifically, if (x_1, y_1) and (x_2, y_2) lies within distance 20 µm for `L3` and 24.25 µm for `L5`, it is less likely to observe a z-coordinate z_2 (associated to (x_2, y_2)) with $|z_1 - z_2|$ less than 11.5 µm for `L3` and 15.5 µm for `L5`. Analogously, given that (x_1, y_1) and (x_2, y_2) lies within distance 11 µm for `L3` and 14.75 µm `L5`, it is more likely to observe z_2 if $|z_1 - z_2|$ is in the interval from 11.5 µm to 35.5 µm for `L3` or from 15.5 µm to 37.25 µm for `L5`.

Finally, note that simulations from each of models 1–5 can straightforwardly be obtained using a Metropolis-Hastings algorithm for a fixed number of points and given a realisation of the xy-coordinates. Specifically, we used (Algorithm 7.1 in Møller and Waagepetersen, 2004) but with a systematic updating scheme cycling over the point indexes 1 to n, using a uniform proposal for a new point in W_z and a Hastings ratio calculated from the full conditional (B.3). We successively updated each point 100 times under the systematic updating scheme, corresponding to 63400 and 54800 point updates for `L3` and `L5`, respectively.

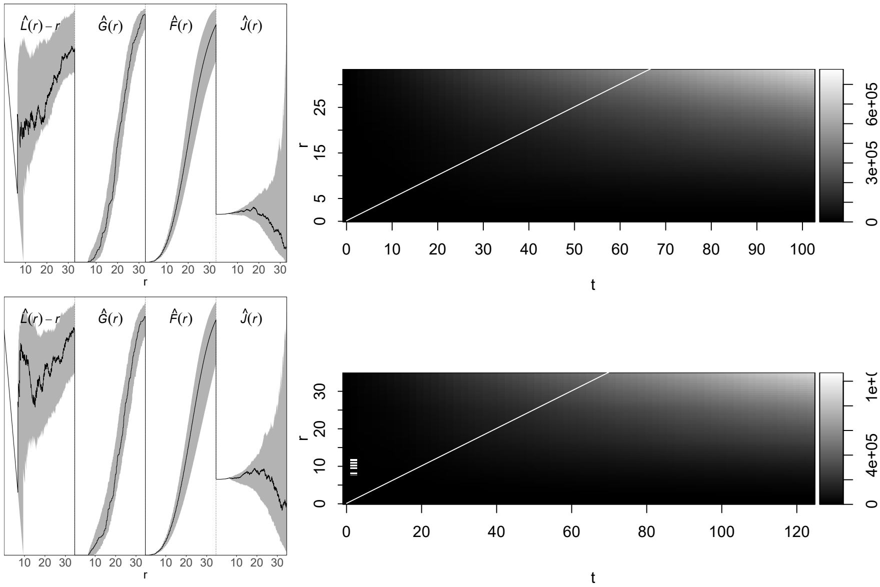


Figure B.7: Results of the GERL envelope procedure under the fitted model 5 based on a concatenation of the empirical *L*-, *G*-, *F*-, *J*-, and cylindrical *K*-functions. Left: concatenation of the one-dimensional empirical functional summaries for the data (solid line) together with 95% envelopes (grey region); for ease of visualisation, the functions have been scaled. Right: empirical cylindrical *K*-function (grey scale) where shaded vertical/horizontal lines indicate that the function falls above/below the 95% envelope. The white line indicates the values for which the cylinder height is equal to the base diameter. Top: results for the dataset `L3`. Bottom: results for the dataset `L5`.

Acknowledgements This work was supported by The Danish Council for Independent Research | Natural Sciences, grant DFF – 7014-00074 'Statistics for point processes in space and beyond', and by the 'Centre for Stochastic Geometry and Advanced Bioimaging', funded by grant 8721 from the Villum Foundation. We are thankful to Ali H. Rafati for collecting the data analysed in this paper and to Jens R. Nyengaard and Ninna Vihrs for helpful comments.

References

Baddeley, A., Rubak, E., and Turner, R. (2016). *Spatial Point Patterns: Methodology and Applications with R*. Chapman & Hall/CRC, New York.

Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician*, 24:179–195.

Biscio, C. A. N. and Lavancier, F. (2016). Quantifying repulsiveness of determinantal point processes. *Bernoulli*, 22:2001–2028.

Buxhoeveden, D. P. and Casanova, M. F. (2002). The minicolumn hypothesis in neuroscience. *Brain*, 125:935–951.

Casanova, M. F. (2007). Schizophrenia seen as a deficit in the modulation of cortical minicolumns by monoaminergic systems. *International Review of Psychiatry*, 19:361–372.

Casanova, M. F., van Kooten, I. A. J., Switala, A. E., van Engeland, H., Heinsen, H., Steinbusch, H. W. M., Hof, P. R., Trippe, J., Stone, J., and Schmitz, C. (2006). Minicolumnar abnormalities in autism. *Acta Neuropathologica (Berl)*, 112:287–303.

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

Diggle, P. J. (2014). *Statistical Analysis of Spatial and Spatio-temporal Point Patterns*. Chapman & Hall/CRC Press, Boca Raton, Florida.

Diggle, P. J. and Gratton, R. J. (1984). Monte Carlo methods of inference for implicit statistical models (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 46:193–227.

Esiri, M. M. and Chance, S. A. (2006). Vulnerability to Alzheimer's pathology in neocortex: the roles of plasticity and columnar organization. *Journal of Alzheimer's Disease*, 9:79–89.

Guan, Y. (2006). A composite likelihood approach in fitting spatial point process models. *Journal of the American Statistical Association*, 101:1502–1512.

Guan, Y. (2009). A minimum contrast estimation procedure for estimating the second-order parameters of inhomogeneous spatial point processes. *Statistics and Its Interface*, 2:91–99.

Lavancier, F., Møller, J., and Rubak, E. (2015). Determinantal point process models and statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77:853–877.

Lavancier, F., Poinas, A., and Waagepetersen, R. P. (2018). Adaptive estimating function inference for non-stationary determinantal point processes. Available on arXiv:1806.06231.

van Lieshout, M. N. M. and Baddeley, A. J. (1996). A nonparametric measure of spatial interaction in point patterns. *Statistica Neerlandica*, 50:344–361.

Møller, J., Christensen, H. S., Cuevas-Pacheco, F., and Christoffersen, A. D. (2019). Structured space-sphere point processes and K-functions. *Methodology and Computing in Applied Probability*. Available at https://doi.org/10.1007/s11009-019-09712-w.

Møller, J. and Christoffersen, A. D. (2018). Pair correlation functions and limiting distributions of iterated cluster point processes. *Journal of Applied Probability*, 55:789–809.

Møller, J., Safavimanesh, F., and Rasmussen, J. G. (2016). The cylindrical K-function and Poisson line cluster point processes. *Biometrika*, 104:937–954.

Møller, J. and Torrisi, G. L. (2005). Generalised shot noise Cox processes. *Advances in Applied Probability*, 37:48–74.

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

Mrkvička, T., Myllymäki, M., and Hahn, U. (2017). Multiple Monte Carlo testing, with applications in spatial point processes. *Statistics and Computing*, 27:1239–1255.

Mrkvička, T., Soubeyrand, S., Myllymäki, M., Grabarnik, P., and Hahn, U. (2016). Monte Carlo testing in spatial statistics, with applications to spatial residuals. *Spatial Statistics*, 18:40–53.

Mrkvička, T., Hahn, U., and Myllymäki, M. (2018). A one-way ANOVA test for functional data with graphical interpretation. Available on arXiv:1612.03608.

Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H., and Hahn, U. (2017). Global envelope tests for spatial processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79:381–404.

Ogata, Y. and Katsura, K. (1991). Maximum likelihood estimates of the fractal dimension for random spatial patterns. *Biometrika*, 78:463–474.

Prokešová, M., Dvořák, J., and Jensen, E. B. V. (2016). Two-step estimation procedures for inhomogeneous shot-noise Cox processes. *Annals of the Institute of Statistical Mathematics*, 69:1–30.

Rafati, A. H., Safavimanesh, F., Dorph-Petersen, K., Rasmussen, J. G., Møller, J., and Nyengaard, J. R. (2016). Detection and spatial characterization of minicolumnarity in the human cerebral cortex. *Journal of Microscopy*, 261:115–126.

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability*, 13:255–266.

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

Waagepetersen, R. P. (2007). An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics*, 63:252–258.

Paper C

Point processes on directed linear networks

Jakob G. Rasmussen and Heidi S. Christensen

The paper has been submitted to
Methodology and Computing in Applied Probability.

The layout has been revised.

Abstract

In this paper we consider point processes specified on directed linear networks, i.e. linear networks with associated directions. We adapt the so-called conditional intensity function used for specifying point processes on the time line to the setting of directed linear networks. For models specified by such a conditional intensity function, we derive an explicit expression for the likelihood function, specify two simulation algorithms (the inverse method and Ogata's modified thinning algorithm), and consider methods for model checking through the use of residuals. We also extend the results and methods to the case of a marked point process on a directed linear network. Furthermore, we consider specific classes of point process models on directed linear networks (Poisson processes, Hawkes processes, non-linear Hawkes processes, self-correcting processes, and marked Hawkes processes), all adapted from well-known models in the temporal setting. Finally, we apply the results and methods to analyse simulated and neurological data.

1 Introduction

Point processes on linear networks are important for modelling events or objects on a real network, such as a road or river network. In recent years there have been a fair amount of papers on functional summary statistics and models for point processes specified on linear networks (see Okabe and Yamada, 2001; Ang et al., 2012; Baddeley et al., 2014; McSwiggan et al., 2016; Baddeley et al., 2017; Rakshit et al., 2017). Specifically, Okabe and Yamada (2001) present a network analogue of Ripley's K-function of which Ang et al. (2012) suggest a correction that compensates for the geometry of the network, making it possible to compare K-functions for different networks directly. For these K-functions it is required that the point process is second-order pseudostationary, meaning that the intensity is constant and the pair correlation function depends only on the geodesic distance. However, Baddeley et al. (2017) discuss the difficulties of finding such point processes, and Rakshit et al. (2017) discuss using alternative distance metrics and present analogues of the K-function and pair correlation function wrt. these metrics. Further, Baddeley et al. (2014) present methods for analysing multitype point processes on networks, and McSwiggan et al. (2016) address problems with existing kernel estimates of the intensity point processes and further develop a new kernel estimate eluding these problems.

In the present paper we consider directed linear networks, i.e. networks consisting of line segments with an associated direction. Such directions appear naturally in some applications, while directions cannot be used, or at the very least are rather artificial, in other applications. For example, river networks have a natural direction following the flow of water, while the bidirectionality of (most) roads means that directed networks are not particularly useful as models for road networks. Ver Hoef et al. (2006), Garreta et al. (2010), and Ver Hoef and Peterson (2012) consider Gaussian processes and covariance functions on so-called stream networks, which are special cases of directed linear networks. In the present paper, however, we focus on point processes specified by a modified version of the conditional intensity function often used for point processes on the time line (see e.g. Daley and Vere-Jones, 2003, Chapter 7, for an introduction to these). On the time line, the conditional intensity is based on conditioning on the past, and for a directed linear network the directions enable us to modify the notion of past and thereby to extend the definition of a conditional intensity.

There are many types of data suitable for modelling by a point process specified by a conditional intensity function on a directed linear network. One example is the locations of spines along a dendritic tree, where we can introduce directions going from the root of the tree towards the leaves of the network. Spines play a role in e.g. memory storage, and changes in the spine distribution and shape have been linked to neurological diseases (Irwin et al., 2001). Only a few studies (Jammalamadaka et al., 2013; Baddeley et al., 2014) model the distribution of spines using point processes on (undirected) linear networks. Further, these studies only consider the Poisson process and the multitype Poisson process when spines are classified into types depending on their shape.

The outline of the paper is as follows: In Section 2 we first define directed linear networks and a number of related concepts, and next what we mean by a conditional intensity function on a directed linear network. In Section 3 we derive the likelihood function for a point process specified by such a conditional intensity function, and in Section 4 we consider two simulation algorithms. In Section 5 we discuss a method for model checking based on residuals. In Section 6, using the conditional intensity function, we define a number of models for point processes on directed linear networks all inspired from similar models on the time line. In Section 7 we use the presented models and methods to analyse simulated datasets and a real dataset consisting of spines along a dendritic tree. Finally, we round off the paper by considering possible extensions and future research directions in Section 8.

2 Point processes on directed linear networks

2.1 Directed linear networks

Let $L_i \subseteq \mathbb{R}^d$, $i = 1, \dots, N$, denote an open line segment of finite length $|L_i|$, where $\mathbb{R}^d$ denotes the d-dimensional Euclidean space for $d \geq 2$. A direction is associated to each line segment L_i, where we denote the endpoints of L_i

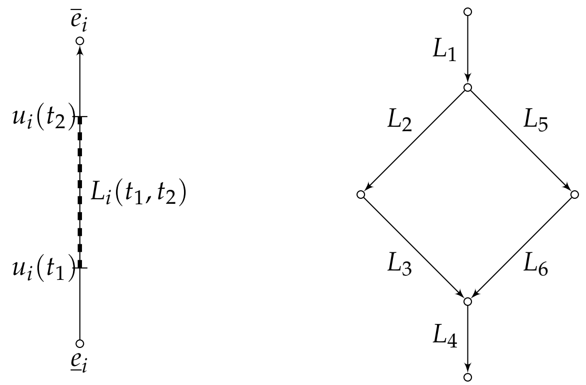


Figure C.1: Left: a directed line segment L_i with endpoints $\underline{e}_i$ and $\bar{e}_i$ and a partial line segment $L_i(t_1,t_2)$ starting in $u_i(t_1)$ and ending in $u_i(t_2)$. Right: a DALN consisting of the line segments $L_1,\dots,L_6$

by $\underline{e}_i, \bar{e}_i \in \mathbb{R}^d$ such that the direction goes from $\underline{e}_i$ to $\bar{e}_i$. Furthermore, we assume that the line segments do not overlap (but they are of course allowed to join at their endpoints in order to form a network). Any point that is the endpoint of at least one line segment is called a vertex. The line segment L_i is an outgoing line segment from the vertex at $\underline{e}_i$ and ingoing at the vertex at $\bar{e}_i$. The ith line segment can conveniently be represented by the parametrisation

$$u_i(t) = \underline{e}_i + t\frac{\bar{e}_i - \underline{e}_i}{\|\bar{e}_i - \underline{e}_i\|},$$

where $t \in (0, |L_i|)$ and $\|\cdot\|$ denotes Euclidean distance. Occasionally we will consider only parts of a line segment, where $L_i(t_1,t_2)$ denotes the set $\{u_i(t) : t \in (t_1,t_2)\}$ for $0 \le t_1 \le t_2 \le |L_i|$. See Figure C.1 for illustrations of the above concepts.

The union of the line segments is denoted by $L^\cup = \bigcup_{i=1}^N L_i$, while the set of line segments is denoted by $L = \{L_i : i = 1,\dots,N\}$. The terminology directed linear network may refer to either L or $L^\cup$ depending on the context.

We have used open line segments to build the directed linear network in order to avoid endpoints being part of multiple line segments. Obviously, when we later define point processes on networks, this means that there cannot be any points exactly at the vertices, but since we will anyway consider only point processes with a diffuse measure, such points would occur with probability zero.

2.2 Directed paths and partial orders

We define a directed path of line segments going from $L_i \in L$ to $L_j \in L$, where $i \neq j$, in the following way: Let $I \subseteq \{1,\dots,N\}$ be indices for a subset of L

with cardinality $|I|$ and $i, j \in I$. Then $(L_{\psi(1)}, \dots, L_{\psi(|I|)})$ is called a directed path from L_i to L_j if $\psi : \{1, \dots, |I|\} \to I$ is a bijection such that $\psi(1) = i$, $\psi(|I|) = j$, and $\underline{e}_{\psi(k)} = \overline{e}_{\psi(k-1)}$ for $k = 2, \dots, |I|$. In other words, you can get from any point in L_i to any point in L_j following the directions of the line segments in the directed path. If at least one such directed path exists, we write $L_i \to L_j$. For the DALN in Figure C.1, one possible directed path from L_1 to L_4 is (L_1, L_5, L_6, L_4), where $I = \{1, 4, 5, 6\}$ and ψ is specified by $\psi(1) = 1$, $\psi(2) = 5$, $\psi(3) = 6$, and $\psi(4) = 4$.

We extend the notion of a directed path to any pair of points in the network: Let $u = u_i(t_1) \in L_i$ and $v = u_j(t_2) \in L_j$ for $L_i, L_j \in L$ (not necessarily distinct), and let

$$u \to v \text{ if } \begin{cases} L_i \to L_j & \text{for } i \neq j \\ t_1 < t_2 & \text{for } i = j \end{cases}.$$

A path from u to v is denoted by $p_{u \to v}$ and consists of

$$(L_{\psi(1)}(t_1, |L_{\psi(1)}|), L_{\psi(2)}, \dots, L_{\psi(|I|-1)}, L_{\psi(|I|)}(0, t_2)),$$

where the first and last line segments have been restricted to points strictly between u and v. The length of a directed path $p_{u \to v}$ is the sum of the lengths of all line segments on the directed path (with the end line segments restricted as above) and will be denoted $|p_{u \to v}|$. Further, we let $P_{u \to v}$ denote the set of all directed paths from u to v. If $u \to v$, the length of the shortest directed path from u to v is denoted by $d_L^{\to}(u, v)$, and if $u \not\to v$ then we let $d_L^{\to}(u, v) = \infty$ (except if $u = v$ in which case $d_L^{\to}(u, v) = 0$). Note that $d_L^{\to}(\cdot, \cdot)$ is a metric except that it is not symmetric (i.e. it is a quasi-metric).

We restrict attention to a particular class of directed linear networks satisfying that if $u \to v$ for any $u, v \in L^{\cup}$, then $v \not\to u$; that is, there are no directed loops in the network. Such a network is called a directed acyclic linear network (DALN); two examples of DALNs are shown in Figure C.2. Most results in this paper depend on this assumption. For a DALN, the relation $\to$ is a strict partial order (i.e. it is non-reflexive, anti-symmetric and transitive) either on L (when considering line segments) or on $L^{\cup}$ (when considering points). Furthermore, we often consider special cases of DALNs for which we use standard terminology from graph theory: A directed linear network without loops (disregarding directions) is called a tree, and a tree where all directions go either away from or towards a single vertex (called the root) is called a rooted tree (or an out-tree when the directions are away from the root and in-tree when they are towards the root). The right panel of Figure C.2 shows an example of an out-tree, where the root is the vertex at the bottom.

The relation $\to$ induces a (not necessarily unique) order on all line segments in a DALN. Let $\omega : \{1, \dots, N\} \to \{1, \dots, N\}$ be a bijection such that $L_{\omega(i)} \to L_{\omega(j)}$ whenever $i < j$, i.e. we get a new ordering of the line segments $L_{\omega(1)}, \dots, L_{\omega(N)}$ that follows the partial order $\to$ whenever it applies to a

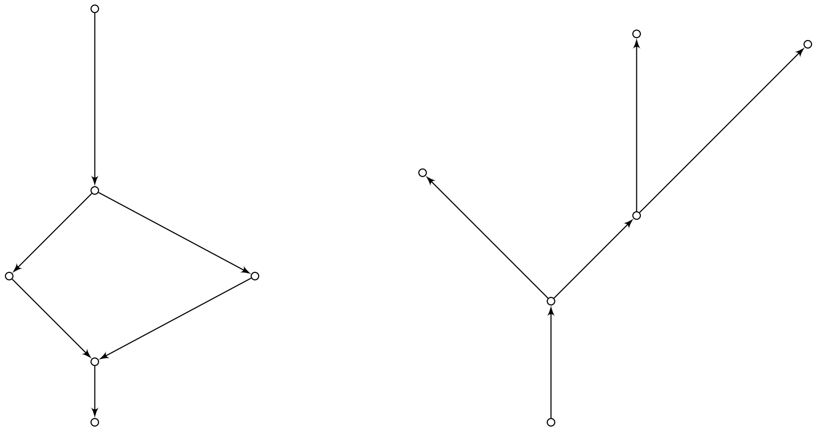


Figure C.2: Two examples of DALNs. These DALNs are also used in the simulation study in Section 7.1

pair of line segments. Denote the set of all such bijections by Ω. A bijection $\omega \in \Omega$ can be obtained by picking out an arbitrary line segment, then go against the directions until we reach a line segment L_i where no $L_j \to L_i$ for $j \in \{1, \dots, N\}$ (the existence of such a line segment follows since L is finite and $\to$ is a strict partial order). We then let $\omega(1) = i$, such that $L_{\omega(1)} = L_i$ is the first line segment in the new ordering of the line segments, and continue iteratively by considering the network $L \backslash \{L_i\}$. For the DALN in Figure C.1, one choice of $\omega \in \Omega$ is $\omega(1) = 1$, $\omega(2) = 2$, $\omega(3) = 3$, $\omega(4) = 5$, $\omega(5) = 6$, and $\omega(6) = 4$.

While most results in this paper need a choice of $\omega \in \Omega$, they do not depend on the actual choice of ω. Throughout the paper, it will be assumed that some choice of ω has been made whenever needed.

2.3 Point patterns and point processes on linear networks

A point pattern on a linear network is a finite set $\boldsymbol{x} \subset L^{\cup}$, and a point process on a linear network is a stochastic process $\boldsymbol{X}$ whose realisations are point patterns on the network. If we assume that the network is a DALN, we can use the order induced by $\to$ to specify a point process by a conditional intensity function as described in Section 2.4.

To describe how points in a point pattern on a directed linear network are located relative to each other, we adopt further terms from graph theory. Let $\boldsymbol{x}$ be a point pattern on a directed linear network and $x \in \boldsymbol{x}$. Then the ancestors of x are the set of points

$$\mathrm{an}(x) = \{y \in \boldsymbol{x} : y \to x\},$$

and the parents of x, $\text{pa}(x) \subseteq \text{an}(x)$, are the ancestors from which there exists at least one directed path to x containing no other points of $\boldsymbol{x}$. The descendants and children of x, denoted by $\text{de}(x)$ and $\text{ch}(x)$, are defined similarly, just reversing the direction.

2.4 Conditional intensity functions on directed linear networks

Let $L_{\to i} = \{L_j \in L : L_j \to L_i\}$ denote the set of line segments with directed paths to $L_i \in L$, and similarly let $L^{\cup}_{\to i} = \bigcup_{L_j \in L: L_j \to L_i} L_j$ denote the union of these. We now define a point process on L by defining a point process on each L_i conditionally on the points in $L_{\to i}$. First, recall that the conditional intensity function $\tilde{\lambda}^*$ of a temporal point process $\tilde{\boldsymbol{X}}$ is given by

$$\tilde{\lambda}^*(t) = \frac{\mathrm{E}[N([t, t+dt])|\tilde{\boldsymbol{X}}_t]}{dt},$$

i.e. the mean number of points $N([t, t+dt])$ falling in an infinitesimally small time interval $[t, t+dt]$ starting at time t conditional on the point process before time t denoted by $\tilde{\boldsymbol{X}}_t$ (see e.g. Daley and Vere-Jones, 2003, for more details on the conditional intensity function for temporal point processes). Next, to adapt this concept to directed linear networks, we now let $\tilde{\lambda}^*$ denote the conditional intensity function of a temporal point process restricted to an interval $(0, |L_i|)$, and allow it to depend on a point pattern on $L_{\to i}$. The resulting point pattern on $(0, |L_i|)$ is then mapped to L_i by $u_i(t)$. We call this a point process with conditional intensity function $\lambda^*(u)$ for $u \in L_i$, where $\lambda^*(u) = \tilde{\lambda}^*(t)$ for $u = u_i(t)$.

To obtain a point process on $L^{\cup}$, we define a point process with conditional intensity function λ^* as above recursively on $L_{\omega(1)}, \ldots, L_{\omega(N)}$, $\omega \in \Omega$. Note that following this order ensures that whenever we define a point process on a line segment L_i, we have already defined it on $L_{\to i}$, on which we condition. Further, the conditional intensity does not depend on the specific choice of permutation ω.

As discussed in Section 2.1, our definition of a directed linear network does not include the vertices as a part of the network. However, in practice we may have datasets containing points located exactly on the vertices, e.g. if the location of the points have been used as vertices when approximating the true network with line segments (this is indeed the case for the dendrite data considered in Section 7.2). Each of these points need to be allocated to a unique line segment such that the conditional intensity is correctly specified. How to do this depends on the nature of the network L. If L is an out-tree where the root is of degree 1, we naturally allocate a point falling at the root to the starting point of the line segment starting in the root. Any other points falling at a vertex will be allocated the endpoint of the ingoing line segment of that vertex. Thus, the line segment going from the root has been extended to include both endpoints, while any other line segment L_i include their second endpoint $\bar{e}_i$. Similar modifications can be made to other networks.

2.5 Marks

Often additional information, referred to as marks, are associated with each point in a point pattern $\boldsymbol{x} = \{x_1, \ldots, x_n\}$. Assume that the marks belong to a space $\mathbb{M}$, which we call the mark space, and that each point $x_i \in \boldsymbol{x}$ in the observed point pattern has an associated mark $m_i \in \mathbb{M}$. That is, a marked point pattern on a directed linear network is a finite set $\boldsymbol{y} = \{(x_1, m_1), \ldots, (x_n, m_n)\} \subset L^{\cup} \times \mathbb{M}$.

To define a marked point process, we let a mark associated with the point $u \in L$ follow a distribution, that may depend both on the location u and the marked point process on $L^{\cup}_{\to u} = L(t, |L_i|) \cup \bigcup_{L_j \to L_i} L_j$ for $u = u_i(t)$. The conditional intensity can then be generalised to the marked case by

$$\lambda^*(u, m) = \lambda^*(u) f^*(m|u),$$

where $\lambda^*(u)$ is the intensity defined in Section 2.4, except that the star now means that it may depend on marks of points on $L^{\cup}_{\to u}$ in addition to the points themselves, and $f^*(\cdot|u)$ is the conditional density function of the mark given the points and marks on $L^{\cup}_{\to u}$. Note that in the marked case, by a slight abuse of notation, we let λ^* denote the conditional intensity function both depending on the point u and mark m as well as the conditional intensity function depending only on the point u.

3 Likelihood function

We can obtain a closed form expression for the likelihood function for a point process on a DALN specified by a conditional intensity function. Firstly, consider the measure λ_1, where $\lambda_1(A)$ is the total length of a measurable subset $A \subseteq L^{\cup}$. Furthermore, we use the notation $\boldsymbol{x}_{(i)} = \boldsymbol{x} \cap L_i$ and $\boldsymbol{x}_{(\to i)} = \boldsymbol{x} \cap L_{\to i}$. Finally, assume that the conditional intensity function depends on some parameter vector, say $\boldsymbol{\theta}$.

Proposition 3.1 . Consider an unmarked point process $\boldsymbol{X}$ on a DALN L specified by a conditional intensity function λ^* depending on a parameter vector $\boldsymbol{\theta}$, and let $\boldsymbol{x}$ be an observed point pattern dataset. Then the likelihood function is given by

$$\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{x}) = \left(\prod_{x \in \boldsymbol{x}} \lambda^*(x; \boldsymbol{\theta}) \right) \exp\left(- \int_L \lambda^*(u; \boldsymbol{\theta}) \, \mathrm{d}\lambda_1(u) \right).$$

Similarly, if $\boldsymbol{Y}$ is a marked point process with conditional intensity function λ^* depending on $\boldsymbol{\theta}$, and $\boldsymbol{y}$ is an observed marked point pattern, then the likelihood function is given by

$$\mathcal{L}(\boldsymbol{\theta};\boldsymbol{y}) = \left(\prod_{(x,m)\in\boldsymbol{y}} \lambda^*(x,m;\boldsymbol{\theta})\right) \exp\left(-\int_L \lambda^*(u;\boldsymbol{\theta})\,\mathrm{d}\lambda_1(u)\right).$$

Proof. Consider first the unmarked case. Letting $\omega \in \Omega$, we split the likelihood into a product of density functions for the point pattern $\boldsymbol{x}_{(\omega(i))}$ on each line segment conditional on the points patterns earlier in ω given by $\boldsymbol{x}_{(\omega(1))}, \dots, \boldsymbol{x}_{(\omega(i-1))}$. That is,

$$\mathcal{L}(\boldsymbol{\theta};\boldsymbol{x}) = \prod_{i=1}^{N} f(\boldsymbol{x}_{(\omega(i))}|\boldsymbol{x}_{(\omega(1))}, \dots, \boldsymbol{x}_{(\omega(i-1))};\boldsymbol{\theta}).$$

Since $\boldsymbol{x}_{(\to i)} \subseteq \cup_{j=1}^{i-1}\boldsymbol{x}_{(\omega(j))}$ and $\boldsymbol{x}_{(\omega(i))}|\boldsymbol{x}_{(\omega(1))}, \dots, \boldsymbol{x}_{(\omega(i-1))}$ by construction depends only on $\boldsymbol{x}_{(\to i)}$, we get that

$$\mathcal{L}(\boldsymbol{\theta};\boldsymbol{x}) = \prod_{i=1}^{N} f(\boldsymbol{x}_{(\omega(i))}|\boldsymbol{x}_{(\to\omega(i))};\boldsymbol{\theta}) = \prod_{i=1}^{N} f(\boldsymbol{x}_{(i)}|\boldsymbol{x}_{(\to i)};\boldsymbol{\theta}). \tag{C.1}$$

By definition, a point process on L_i specified conditionally on $\boldsymbol{x}_{(\to i)}$ by $\lambda^*(u;\boldsymbol{\theta})$ is equivalent to a temporal point process specified by $\tilde{\lambda}^*(t;\boldsymbol{\theta})$ on $(0,|L_i|)$, where $u = u_i(t)$. Thus, by Daley and Vere-Jones (2003, Proposition 7.2.III), we get that

$$f(\boldsymbol{x}_{(i)}|\boldsymbol{x}_{(\to i)}) = \left(\prod_{x\in\boldsymbol{x}_{(i)}} \lambda^*(x;\boldsymbol{\theta})\right) \exp\left(-\int_{L_i} \lambda^*(u;\boldsymbol{\theta})\,\mathrm{d}\lambda_1(u)\right),$$

which together with (C.1) completes the proof for the unmarked case. The result for the marked case is proven in a similar manner, using Proposition 7.3.III instead of 7.2.III in Daley and Vere-Jones (2003). □

4 Simulation

There are two general methods for simulating temporal point processes specified by a conditional intensity function: the inverse method and Ogata's modified thinning algorithm. Both algorithms can be modified to work for a point process on a DALN by simulating the point process on one line segment at a time following the order given by $\omega \in \Omega$. So we focus on specifying how to simulate the point process on a line segment L_i conditional on the points already simulated on $L_{(\to i)}$.

First, we consider the inverse method in the unmarked case. Let $u = u_i(t)$ for $t \in (0, |L_i|)$, and let

$$\Lambda^*(u) = \int_{L_i(0,t)} \lambda^*(v) \, \mathrm{d}\lambda_1(v).$$

In the inverse method, independent and identically distributed (IID) unit-rate exponential random variables are simulated and transformed into the appropriate points on L_i by the inverse of Λ^*. More precisely, the algorithm is as follows:

```
1. Let $j = 0$.
2. Repeat:
    (a) Generate $Y_j \sim \mathrm{Exp}(1)$.
    (b) Find $t$ such that $\Lambda^*(u_i(t)) = \sum_{k=0}^{j} Y_k$.
    (c) If $t < |L_i|$, let $j = j + 1$ and $x_j = u_i(t)$. Else end repeat loop.
3. Output $(x_1, \ldots, x_j)$.
```

Next, for Ogata's modified thinning algorithm in the unmarked case, we use that $\lambda^*(u) = \tilde{\lambda}^*(t)$ when $u = u_i(t)$ and assume that for any $t \in (0, |L_i|)$, there exist functions $L^*(t) > 0$ and $M^*(t) \geq \tilde{\lambda}^*(s)$ for any $s \in [t, t + L^*(t)]$ (here * means that these functions may depend on the already simulated point patterns $\boldsymbol{x}_{(\to i)}$ and the part of $\boldsymbol{x}_i$ in $L_i(0, t)$). The algorithm is as follows:

```
1. Let $t = 0$ and $j = 0$.
2. Repeat:
    (a) Calculate $M^*(t)$ and $L^*(t)$.
    (b) Generate (independently) $T \sim \mathrm{Exp}(M^*(t))$ and $U \sim \mathrm{Unif}([0, 1])$.
    (c) If $t + T > |L_i|$, end repeat loop.
    (d) Else if $T > L^*(t)$, let $t = t + L^*(t)$.
    (e) Else if $U > \tilde{\lambda}^*(t + T)/M^*(t)$, let $t = t + T$.
    (f) Else, let $j = j + 1$, $t = t + T$, and $x_j = u_i(t)$.
3. Output $(x_1, \ldots, x_j)$.
```

Both the inverse method and Ogata's modified simulation algorithm can be extended to the marked case by the following two modifications: First, insert an extra step such that each time a point has been simulated (and either kept for Ogata's modified thinning algorithm or moved for the inverse method), its mark should be simulated using the mark density f^*. Second, note that any function with a star may depend on both points and marks, not just points as in the unmarked case.

The following proposition verifies that both of these algorithms produce a point process on a DALN with the correct distribution in both the unmarked and the marked case.

Proposition 4.1 . Let $\omega \in \Omega$ and L be a DALN, and produce point patterns on $L_{\omega(1)}, \ldots, L_{\omega(N)}$ recursively using either the inverse method or Ogata's modified simulation algorithm. Then the resulting simulation is a point process on L with conditional intensity function λ^*.

Proof. Consider first the inverse method in the unmarked case used on a single line segment L_i. By Daley and Vere-Jones (2003, Theorem 7.4.I) the algorithm produces a simulation on $(0, |L_i|)$ with conditional intensity $\tilde{\lambda}^*$ when we consider $\Lambda^*(u_i(t))$ as a function of t. By the definition of $\lambda^*(u) = \tilde{\lambda}^*(t)$ for $u = u_i(t)$, we then get a correct simulation on L_i.

Consider next Ogata's modified thinning algorithm in the unmarked case used on L_i. By Ogata (1981) and since $\lambda^*(u) = \tilde{\lambda}^*(t)$ for $u = u_i(t)$, a correct simulation is produced on L_i.

Finally, since we follow the order given by ω, $\boldsymbol{x}_{(\to i)}$ has always been simulated, when $\boldsymbol{x}_{(i)}$ has to be simulated, and thus by the above argument each $\boldsymbol{x}_{(i)}$ is simulated correctly, leading to a correct simulation of $\boldsymbol{x}$ in the unmarked case.

Turning to the marked case, we note that the above arguments still hold to prove that the points follow the correct distribution (where Proposition 7.4.IV in Daley and Vere-Jones, 2003, needs to be used for the inverse method, and the text accompanying Algorithm 7.5.V for Ogata's modified thinning algorithm). The proof is completed for the marked case by noting that the marks are always drawn from the correct distribution. ☐

5 Residual analysis

One way of checking the fit of a model specified by a conditional intensity function is to calculate residuals and check their distribution. Consider first the unmarked case, and assume that we have observed a point pattern $(x_{(i),1}, \ldots, x_{(i),n_i})$ on L_i for every $L_i \in L$, and that we have obtained a fitted model with conditional intensity function $\hat{\lambda}^*$ for this dataset. Let

$$\hat{\Lambda}^*(u) = \int_{L_i(0,t)} \hat{\lambda}^*(v) \, \mathrm{d}\lambda_1(v),$$

where $u = u_i(t)$ and $t \in (0, |L_i|)$. We then calculate the residuals for the points on L_i given by $(\hat{\Lambda}^*(x_{(i),1}), \ldots, \hat{\Lambda}^*(x_{(i),n_i}))$, which we (with a slight abuse of notation) denote by $\hat{\Lambda}^*(\boldsymbol{x}_{(i)})$. If the model is correct, then, by Daley and Vere-Jones (2003, Proposition 7.4.IV), $\hat{\Lambda}^*(\boldsymbol{x}_{(i)})$ is a unit-rate Poisson process on the interval $(0, \hat{\Lambda}^*(\bar{e}_i))$, and the residual processes $\hat{\Lambda}^*(\boldsymbol{x}_{(i)})$ are independent for $i = 1, \dots, N$.

In order to check the fit of the model given by $\hat{\lambda}^*$, we need to check whether the residuals form a Poisson process. The question is now how we best check this. One possibility is to check each process $\hat{\Lambda}^*(\boldsymbol{x}_{(i)})$ separately, but there may be few points in each $\hat{\Lambda}^*(\boldsymbol{x}_{(i)})$ and further, we lose information on any discrepancies around the junctions. For example, a particularly large gap around a junction may indicate a discrepancy between the model and the data, but since this gap will then be divided over several $\hat{\Lambda}^*(\boldsymbol{x}_{(i)})$ it may be hard to discover. A better approach might be to construct a network with the same connecting junctions as the original network, and place the residuals on this network. The main complication is that the lengths of the line segments have changed from $|L_i|$ to $\hat{\Lambda}^*(|L_i|)$, so we cannot use the original network, and indeed the changed lengths may imply that there exists no network in $\mathbb{R}^d$ with the correct line segments lengths and the correct connecting junctions. However, we can consider this network in a more abstract sense and apply any method for checking that a point pattern on a network comes from a unit-rate Poisson process, provided that the method does not rely on a correct Euclidean geometry of the network.

One method for testing whether the residuals follow a unit-rate Poisson process model is to perform a global rank envelope test (Myllymäki et al., 2017) with the empirical geometrically corrected K-function or pair correlation function (Ang et al., 2012) as test function. Note that this approach effectively ignores the directions present in the network (see Section 8 for further comments on including directions in the K-function).

Another method is based on interevent distances. To define these for a directed linear network, first recall that for point processes on the time line specified by a conditional intensity function, residual analysis often includes an investigation of the interevent times, i.e. the times between consecutive points of the residual process. If the proposed model is correct, the residuals constitute a unit-rate Poisson process which means that the interevent times are IID exponential variables with mean 1. In practice this can e.g. be checked visually by considering Q-Q-plots or histograms. For a point pattern $\boldsymbol{x}$ on a DALN L, we can define a similar concept, the interevent distances, as the set

$$\{d_L^{\rightarrow}(x_i, x_j) : x_i, x_j \in \boldsymbol{x},\, x_i \in \mathrm{pa}(x_j)\},$$

that is, the distance(s) to a point from its parent(s). If L is an out-tree there is at most one parent for each point in $\boldsymbol{x}$. For a unit-rate Poisson process on L, the interevent distances that corresponds to the distance between two consecutive points on the *same* line segment are independent exponentially distributed variables with mean 1. However, interevent distances going across the same junction are not independent, since a part of the network is shared by the intervals corresponding to the interevent distance. One possible solution is to exclude all such interevent distances when comparing interevent distances to the exponential distribution, but then information around the junctions is lost. Another solution is to consider all interevent distances and thus ignoring the dependency (which in practice may occur only for a small portion of the interevent distances depending on the number of junctions).

Generalizing this to the marked case in a sensible way is tricky. If we focus on the multivariate case, i.e. when the mark space is finite, some progress can be made. Assume that we have estimated the conditional intensity function by $\hat{\lambda}(\cdot,\cdot)$, and let

$$\hat{\Lambda}^*_m(u) = \int_{L_i(0,t)} \hat{\lambda}^*(v,m)\,\mathrm{d}\lambda_1(v)$$

for a fixed mark $m \in \mathbb{M}$. Then all points with mark m on a line segment L_i is transformed using $\hat{\Lambda}^*_m$ to an interval $(0, \hat{\Lambda}^*_m(|L_i|))$. Note in particular that the intervals have different lengths, so the residuals for points with different marks end up in differently sized networks. By Daley and Vere-Jones (2003, Proposition 7.4.VI.(a)), the residual processes thus obtained for different marks should behave like independent unit-rate Poisson processes provided the model is fitting well. We can apply the above techniques to each process separately to check whether these are unit-rate Poisson processes. Ideally we should also check whether each of these processes are independent of each other, but it seems to be hard to make a general test for this, since the processes are located on different networks.

The fact that points with different marks end up in different networks for the multivariate case hints at the difficulty in getting anything useful out of residual analysis for the general marked case. Proposition 7.4.VI.(b) in Daley and Vere-Jones (2003) can be used in this case, but if we for example have a continuous mark distribution, typically no marks are equal, so each residual point will end up in intervals of different lengths, and it is in no way obvious how to combine this into something useful for model checking.

6 Models

New models for point processes on a DALN specified by a conditional intensity function essentially boils down to giving a mathematical expression for the conditional intensity function. There is a rich selection of standard models for temporal point processes that can be expressed using the conditional intensity function. The main problem in adapting them to the case of a DALN is dealing with the fact that at junctions the network may join several line segments and/or split into several line segments. We consider a few examples of models here.

6.1 Poisson process

If the conditional intensity function λ^* is a deterministic non-negative valued measurable function on L, say λ, that does not depend on points further up the network, then we get a Poisson process on L with intensity function λ. For this particular model, the point process does not depend on the directions and is equivalent to a Poisson process specified on an undirected linear network (see e.g. Ang et al., 2012).

For a homogeneous Poisson process on L with constant intensity λ, the maximum likelihood estimate of λ is simply $n/|L|$, where n is the observed number of points (this follows from Proposition 3.1).

6.2 Hawkes process

Another common temporal point process is the Hawkes process or self-exciting process (Hawkes, 1971a,b, 1972; Hawkes and Oakes, 1974). We can extend it to a DALN using the conditional intensity function

$$\lambda^*(u) = \mu + \alpha \sum_{x_i \in \boldsymbol{x}\colon x_i \to u} \gamma(d_L^{\rightarrow}(x_i, u)), \tag{C.2}$$

where $\mu, \alpha > 0$ are parameters and γ is a density function on $(0, \infty)$ that may depend on additional parameters. In the temporal case the model has the interpretation that immigrants appear according to a Poisson process with intensity μ, then each immigrant, say t_i, produces a Poisson process of offspring with intensity $\alpha\gamma(\cdot - t_i)$, and each offspring produces another Poisson process of offspring, and so on (Møller and Rasmussen, 2005, 2006). In particular α can be interpreted as the mean number of offsprings produced by each point. However, for a DALN containing a diverging junction, that is, a vertex with in-degree $= 1$ and out-degree > 1, the offspring process is copied to each outgoing direction, thus giving many more offsprings in mean. In the case that there are multiple paths from x_i to u only the shortest path count, meaning that clusters die out if they encounter themselves further down the network.

If we want a version of the Hawkes process where clusters are split equally when a diverging junction is met, and superposed when a converging junction, i.e. a vertex with indegree > 1 and outdegree $= 1$, is met, we can let

$$\lambda^*(u) = \mu + \alpha \sum_{x_i \in \boldsymbol{x}\colon x_i \to u} \; \sum_{p \in P_{x_i \to u}} g_p(x_i, u)\gamma(|p|), \tag{C.3}$$

where $g_p(x_i, u) = 1/n_p(x_i, u)$, and $n_p(x_i, u)$ is the product of the number of outgoing line segments met on each junction on the path p. For this model, α

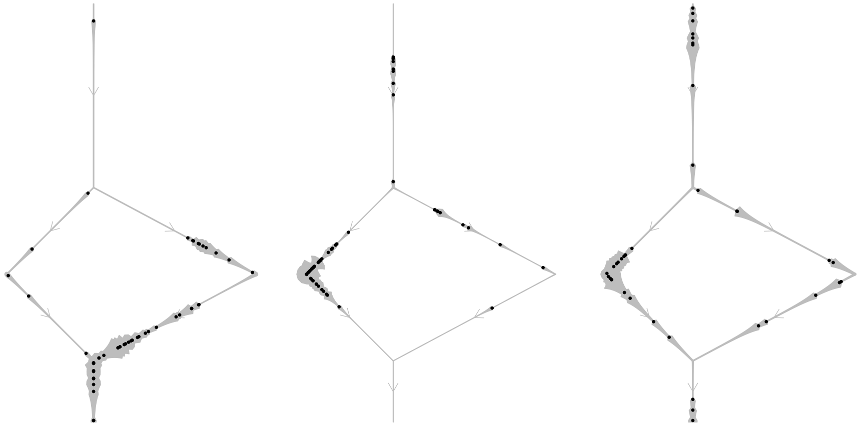


Figure C.3: Simulations of Hawkes processes (on the DALN left of Figure C.2) with parameters specified as follows. Left: $\mu = 1$, $\alpha = 0.8$ and $\gamma(t) = 5\exp(-5t)$. Middle: $\mu = 1$, $\alpha = 0.8$, and $\gamma(t) = 10\exp(-10t)$. Right: $\mu = 1$, $\alpha = 0.9$, and $\gamma(t) = 5\exp(-5t)$. Width of grey regions are proportional to the conditional intensity

is the mean number of offspring resulting from each point (or more precisely, the mean number of offspring is less than or equal to α since the network is finite, and the offspring processes thus get truncated). Using other functions g_j may give other interpretations that are useful for various datasets.

For $u = u_i(t)$, the integrated conditional intensity for (C.2) is given by

$$\Lambda^*(u) = \mu t + \alpha \left[\sum_{x_j \in \boldsymbol{x}\colon x_j \to u} \Gamma(d_L^{\rightarrow}(x_j, u)) - \sum_{x_j \in \boldsymbol{x}\colon x_j \in L^{\cup}_{\to i}} \Gamma(d_L^{\rightarrow}(x_j, \underline{e}_i)) \right],$$

where Γ is the distribution function associated with γ.

6.3 Non-linear Hawkes process

A non-linear Hawkes process (Brémaud and Massoulié, 1994, 1996) is obtained by inserting the conditional intensity function of the Hawkes process into a function $g : \mathbb{R} \to [0, \infty)$ such as the exponential function, that is,

$$\lambda^*(u) = \exp\left[\mu + \alpha \sum_{x_i \in \boldsymbol{x}\colon x_i \to u} \gamma(d_L^{\rightarrow}(x_i, u)) \right]. \tag{C.4}$$

The non-linear Hawkes process does not have a clustering and branching structure as the Hawkes process, so the modification done in (C.3) do not lead to any nice interpretations. On the other hand, if $\alpha > 0$ the point process given by the conditional intensity function (C.4) is clustered, if $\alpha < 0$ it is regular, and if $\alpha = 0$ it is a homogeneous Poisson process, so the model is rather flexible.

6.4 Self-correcting process

To model regular point patterns on DALNs, we further introduce a modification of the self-correcting process from the temporal setting (Isham and Westcott, 1979). The conditional intensity of a self-correcting process increases exponentially as the distance to the starting point increases, while it decreases whenever a point occurs. To adapt such a process to a DALN, we need to specify a meaningful starting point from which we measure distance. Therefore, we require that the DALN L has a vertex v_0 such that $d_L^{\rightarrow}(v_0, u) < \infty$ for all $u \in L$. Note, for an out-tree, v_0 is simply the root of the tree, and for consistency we use the terminology root for v_0 even if the network is not an out-tree.

Then we specify the self-correcting process by

$$\lambda^*(u) = \exp\left\{\mu d_L^{\rightarrow}(v_0, u) - \alpha|\boldsymbol{x} \cap \mathrm{sp}(v_0, u)|\right\}, \tag{C.5}$$

where $|\boldsymbol{x} \cap \mathrm{sp}(v_0, u)|$ is the number of points from $\boldsymbol{x}$ on the shortest directed path $\mathrm{sp}(v_0, u) = \arg\min_{p \in P_{v_0 \to u}} |p|$ from v_0 to u, and $\mu, \alpha > 0$ are parameters controlling the overall intensity and the degree of repulsion. With this definition, only the points lying on the shortest directed path between v_0 and u affect $\lambda^*(u)$. For networks with paths of same length joining u_1 and u_2, the conditional intensity specified by (C.5) is somewhat ambiguous as $\mathrm{sp}(u_1, u_2)$ is not necessarily unique. An alternative, that may be more natural for some applications, is to count events on all paths from v_0 to u (and not only on the shortest directed path). However, if $\boldsymbol{X}$ is specified by (C.5), the restriction of $\boldsymbol{X}$ to $p_{v_0 \to u}$ for any $u \in L$ is a temporal self-correcting process on the interval $(0, |p_{v_0 \to u}|)$.

Another possible alteration of (C.5), is to substitute the exponential function with some positive function g.

To obtain an expression for Λ^*, let $\{x_1^i, \ldots, x_{n_i}^i\} = \boldsymbol{x} \cap L_i(0, t)$ denote the n_i events falling on the partial line segment $L_i(0, t)$, while $x_0^i = u_i(0)$ and $x_{n_i+1}^i = u_i(t)$ denote the endpoints of $L_i(0, t)$. Then, for $u = u_i(t)$,

$$\Lambda^*(u) = \frac{c(\boldsymbol{x}, i)}{\mu} \sum_{j=0}^{n_i} \exp(-\alpha j) \left\{\exp\left[\mu d_L^{\rightarrow}(x_0^i, x_{j+1}^i)\right] - \exp\left[\mu d_L^{\rightarrow}(x_0^i, x_j^i)\right]\right\},$$

where $c(\boldsymbol{x}, i) = \exp\left\{\mu d_L^{\rightarrow}(v_0, x_0^i) - \alpha|\boldsymbol{x} \cap \mathrm{sp}(v_0, x_0^i)|\right\}$.

6.5 Marked models

Any of the models in Sections 6.1–6.4 can be extended to the marked set up. The simplest case is to use so-called independent marks (see e.g. Daley

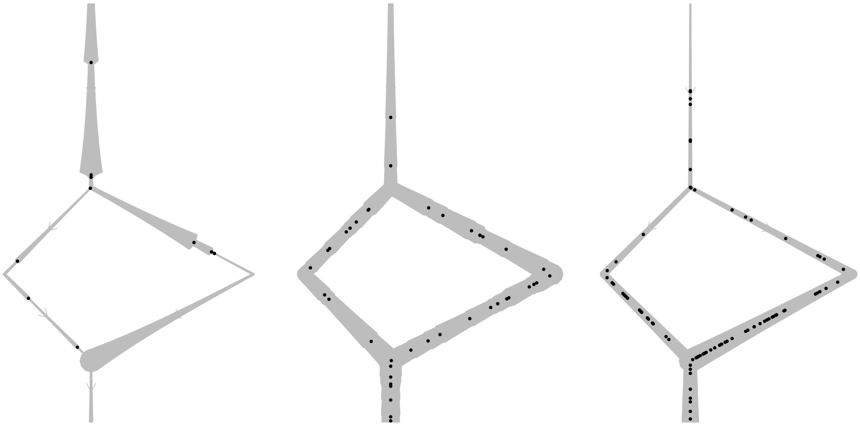


Figure C.4: Simulations of self-correcting processes (on the DALN left of Figure C.2) with parameters specified as follows. Left: $\mu = 0.8$ and $\alpha = 1$. Middle: $\mu = 0.4$ and $\alpha = 0.1$. Right: $\mu = 0.3$ and $\alpha = 0$

and Vere-Jones, 2003), where the marks are independent of each other and independent of the points, with the sole exception that a mark is allowed to depend on the location of the point to which it is associated. More interesting cases can be made by letting the conditional intensity at u depend on the marks associated to the points in $L_{\to u}$ (this is known as unpredictable marks if the other independence assumptions mentioned above still hold) and/or letting the mark associated to a point at u depend on points on $L_{\to u}$ and/or their associated marks.

For an example of a marked point process, consider the Hawkes process given by (C.2) and assume that we are trying to model a dataset that has K different types of points, denoted $1, \dots, K$. Then the marked Hawkes process can be defined using the conditional intensity function

$$\lambda^*(u, m) = \mu_m + \sum_{x_i \in \boldsymbol{x} \colon x_i \to u} \alpha_{m_i, m} \gamma_{m_i, m}(d_L^{\to}(x_i, u)), \tag{C.6}$$

where for $m, m' \in \{1, \dots, K\}$ the parameters in the model are given by $\mu_m, \alpha_{m,m'} > 0$, and $\gamma_{m,m'}$ are density functions on $(0, \infty)$. This generalization has a high number of parameters, and for practical use assumptions that some of these parameters are equal would typically be made.

Similarly, the other models presented in this paper can be extended to multitype cases or more general marked cases, and the primary difficulty is producing practically relevant models with nice interpretations and a reasonably low number of parameters. Obviously, what this is depends on the data at hand.

7 Data Analysis

7.1 Simulated data

To investigate properties of the maximum likelihood estimates for parameters in the Hawkes and self-correcting model, we performed a simulation study using the two DALNs shown in Figure C.2. As results for the two networks are very similar, we only present results for the DALN to the left in Figure C.2. In order to investigate increasing-domain asymptotic properties, we increase the size of the network seven times by 50% each time and denote the resulting networks by sizes $s = 1, \dots, 7$. For each s we simulate 1000 Hawkes processes $\mu = 1$, $\alpha = 0.8$, and $\gamma(t; \kappa) = \kappa \exp(-\kappa t)$, where $\kappa = 5$, and 1000 self-correcting processes with $\mu = 0.4$ and $\alpha = 0.1$ using the inverse method. For each simulation, the parameters have been estimated 1) jointly, by numerically maximising the log-likelihood simultaneously for all parameters, and 2) marginally, by fixing all but one parameter at the true value and then numerically maximising the log-likelihood with respect to the remaining parameter.

Figure C.5 shows box plots of the joint estimates for the simulated Hawkes processes; these suggest that the maximum likelihood estimator of (α, μ, κ) is consistent. Estimating the parameters marginally give similar results (not shown here) but with a slightly lower empirical variance.

For the simulated self-correcting processes, the joint estimates shown in Figure C.6–C.7 are clearly positively correlated, and both μ and α are grossly overestimated. This behaviour may be explained by the way μ and α influence the conditional intensity in (C.5). Specifically, μ controls how much the conditional intensity increases as the distance to the root grows, while α determines how much the conditional intensity decreases when a new point is met. As more points will occur when the distance to the root grows, an increase in μ may to some extent be balanced out by an increase in α. The ridge seen in Figure C.8, displaying contours of the log-likelihood for one of the simulations, confirms that it may be hard to identify the true values of α and μ as the estimates will be chosen somewhere along that ridge. The marginal estimates, shown in first panel of Figure C.6, are less extreme and on average closer to the true value. Specifically, fixing α, the marginal estimates of μ seem unbiased, while fixing μ give positively biased estimates of α but with a smaller bias as the network grows.

This short simulation study, indicates that the behaviour of the maximum likelihood estimates are quite model dependent, and thus it may be hard to say anything about the distribution of these in general. This is discussed further in Section 8.

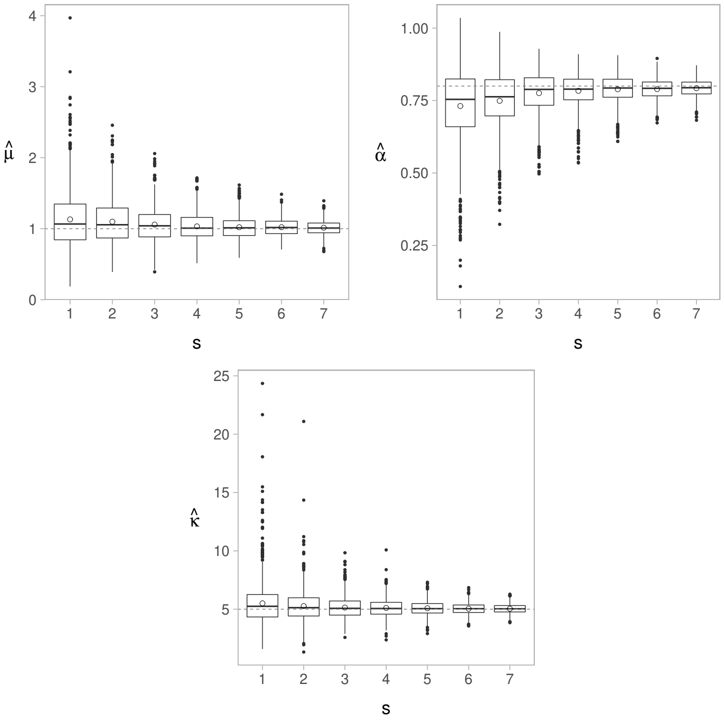


Figure C.5: Results from simulation study: Box plot of joint parameter estimates, $\hat{\mu}$ (top left), $\hat{\alpha}$ (top right), and $\hat{\kappa}$ (bottom), for the simulated Hawkes processes for each network of size s. Here $\circ$ is the empirical mean of the estimates

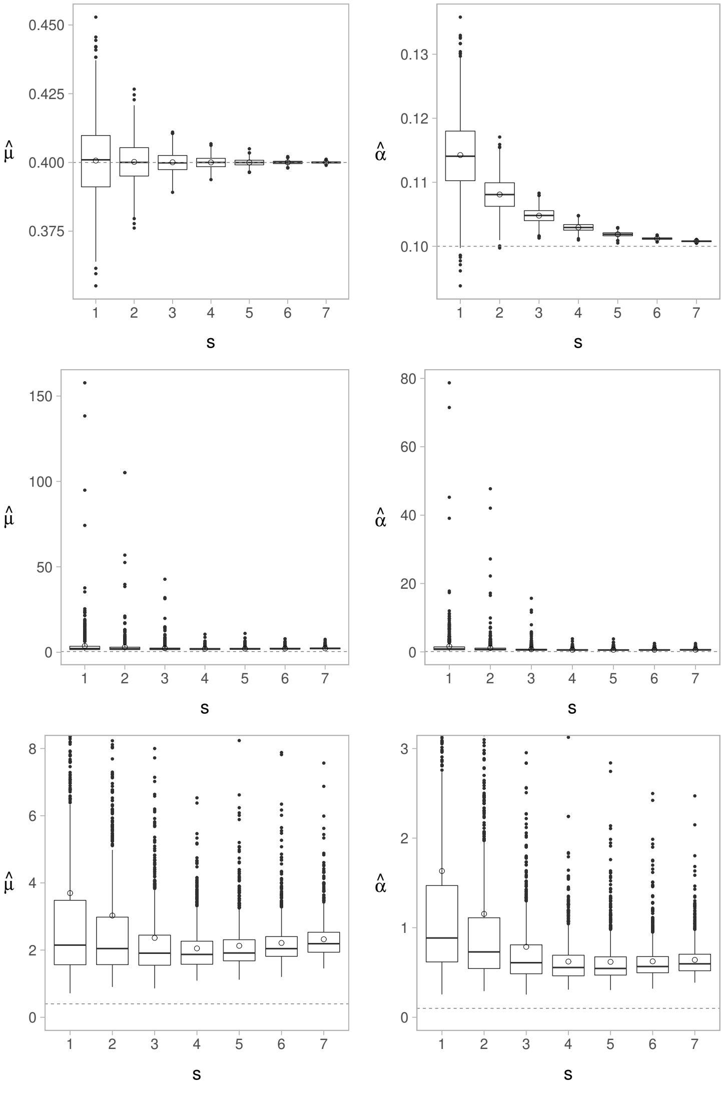


Figure C.6: Box plots of parameter estimates, $\hat{\mu}$ (left) and $\hat{\alpha}$ (right), for the self-correcting process on each network of size s. First panel: marginal parameter estimates. Second panel: joint estimates. Third panel: zoom of second panel. Here ○ denotes the mean estimate

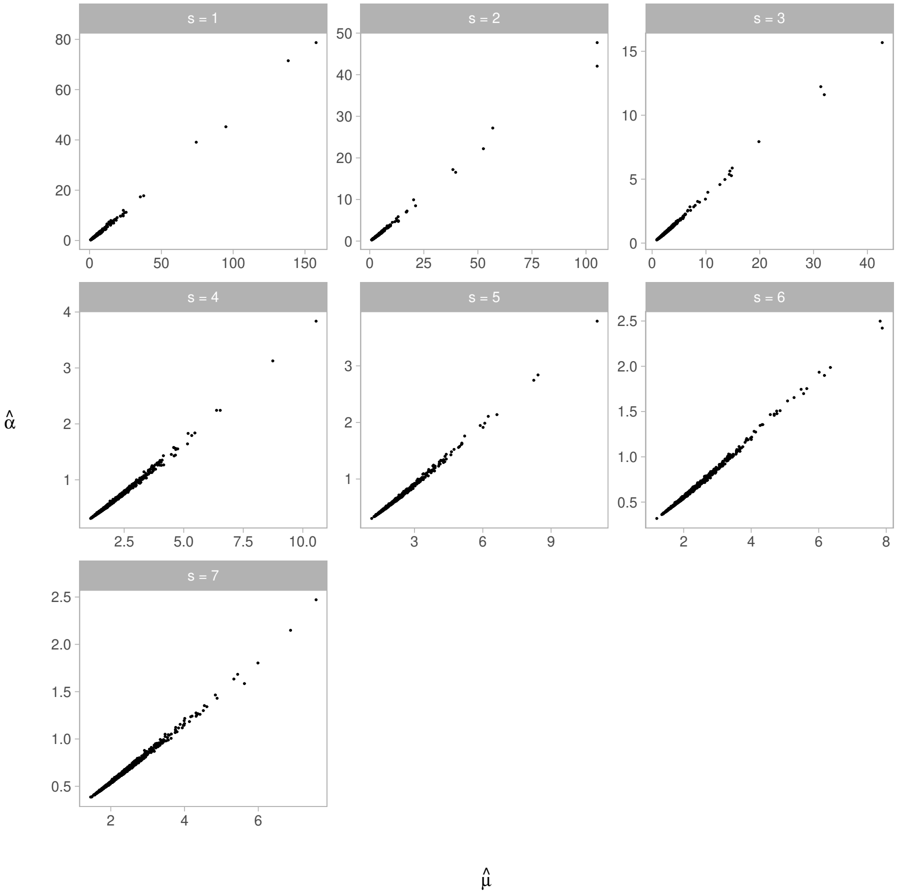


Figure C.7: Plots of joint estimates from the self-correcting simulations. For each size of network s considered in the simulation study, a plot of $\hat{\alpha}$ against $\hat{\mu}$

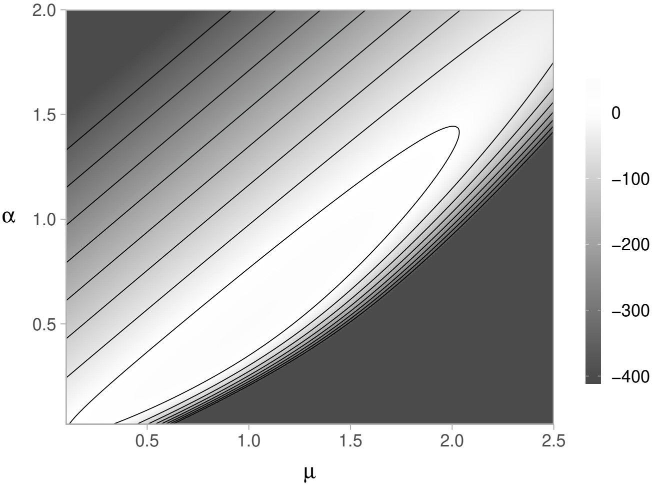


Figure C.8: Contours of the log-likelihood for a simulated self-corrected process with $\mu = 0.4$ and $\alpha = 0.1$

7.2 Dendrite data

In this section we consider a point pattern describing spine locations on an apical dendrite tree from a mouse neuron. The dendrite tree was first approximated by a linear network in $\mathbb{R}^3$ (see Figure C.9). Next, a simplified version of the network with fewer vertices was obtained by joining edges meeting at a vertex of degree two. Then the network was embedded in $\mathbb{R}^2$ (see Figure C.9) in order to directly use functionalities from the `R`-package `spatstat` (Baddeley et al., 2016). The embedding preserves distances, entailing that distance-based analyses on the original network in $\mathbb{R}^3$ and the embedded network in $\mathbb{R}^2$ are equivalent. For example, the geometrically corrected network K-function (Ang et al., 2012) is invariant under this kind of embedding. Letting the dendrite's attachment point to the cell body be the root vertex of the network tree, we can naturally consider the network as a DALN by introducing directions going away from the root (hence the network thus obtained is an out-tree).

First, we tested whether the spine locations can be described by a homogeneous Poisson model with estimated intensity $n/|L|$, where $n = 341$ is the number of spines and $|L| = 876$ μm is the total network length. The empirical geometrically corrected K-function (we return to the issue that this K-function ignores directions in Section 8), $\hat{K}$, may be used as a test function in a global rank envelope test (Myllymäki et al., 2017), where a Monte Carlo approach is applied for approximating the distribution of the test function under the null model. The global rank envelope procedure both give critical bounds for the test function as well as an interval going from the most liberal to the most conservative p-value of the associated test. The p-interval associated with the global rank envelope test for the homogeneous Poisson model is $(0, 0.0096)$, indicating that the model is not appropriate. Distances r, for which $\hat{K}(r)$ falls outside the critical bounds (also called a global rank envelope) shown in Figure C.10, reveal possible reasons for rejecting the model; in this case $\hat{K}(r)$ falls above the envelope for r-values up to approximately $50\,\mu m$, indicating clustering at this scale.

To model the clustering, we next consider the dendrite tree as a directed network and fit a Hawkes model, where we let γ in (C.2) be the density of an exponential distribution with parameter κ. The three parameters, μ, α and κ, are estimated by numerically optimizing the log-likelihood. The resulting estimates are $\hat{\mu} = 0.11$, $\hat{\alpha} = 0.84$, and $\hat{\kappa} = 0.073$. According to Section 5, we can check whether the model adequately describe our data by looking at the residuals. Again, we use the global rank envelope procedure with $\hat{K}$ as test function, but now for testing whether the residuals follow a unit-rate Poisson model on the transformed network. The resulting 95%-global rank envelope, shown in Figure C.10, has an associated p-interval of $(0, 0.0068)$. However, the only discordance detected between the residuals and the unit-rate Poisson model with the global rank envelope is for r-values less than $1\,\mu m$. This may indicate that there is a small-scale repulsion between the spines, which is not accounted for in the Hawkes model.

In Figure C.11, a Q-Q-plot of all interevent distances in the residual process is shown along with labels indicating whether the interevent distance is across a junction or not. Regardless of whether we include these crossing interevent distances or not, the distribution of the interevent distances seems to deviate only slightly from the exponential distribution with mean 1.

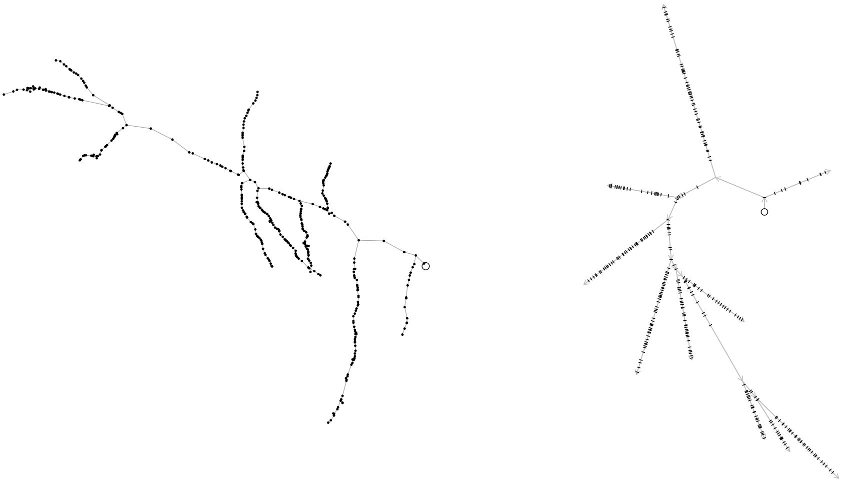


Figure C.9: Left: projection of the approximated dendrite tree onto $\mathbb{R}^2$. Right: a distance-preserving embedding of the network into $\mathbb{R}^2$. Here ○ identify the root of the dendrite tree, while ● (left) and **|** (right) are spine locations

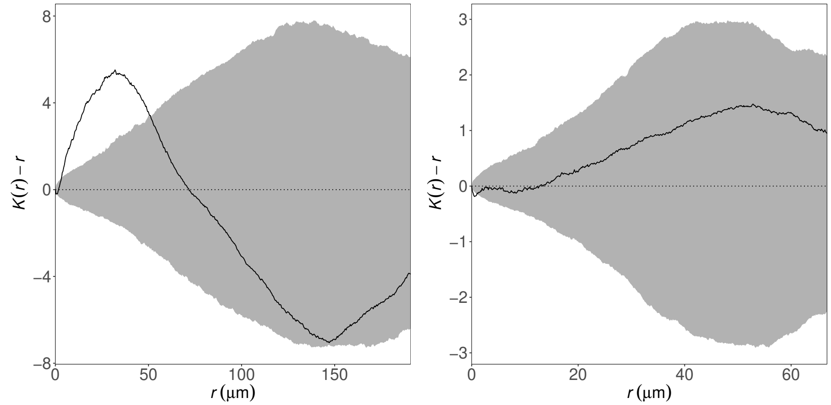


Figure C.10: $\hat{K}(r) - r$ (solid curve) for the observed spine locations (left) and the residuals from the fitted Hawkes model (right). The grey regions are 95%-global rank envelopes based on 4999 simulations from a homogeneous Poisson model with intensity $n/|L|$ (left) and from a unit-rate Poisson (right), respectively

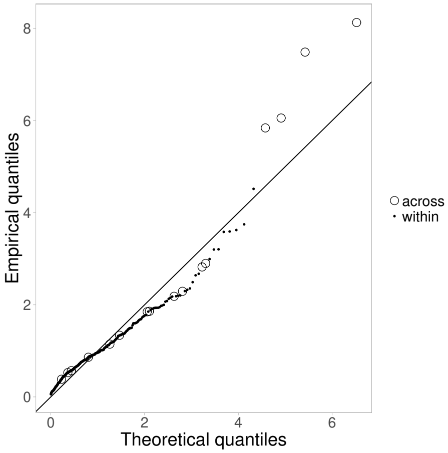


Figure C.11: Q-Q-plot of quantiles for the interevent distances in the residual process for the spine locations vs. theoretical quantiles from the exponential distribution with mean 1. The labels *across* and *within* refer to interevent distances for points on different line segments or on the same line segment, respectively

8 Extensions and future research

In this final section we consider extensions and modifications of the models and methods as well as future research directions.

Recall that the so-called geometrically corrected K-function is introduced in Ang et al. (2012) for analysing point processes on linear networks, and among other things, this can be used for analysing a point pattern dataset for example to determine whether the points are more clustered or more regular than a homogeneous Poisson process. This information is relevant for choosing an appropriate model for modelling a dataset. For example, the Hawkes process in Section 6.2 is more clustered than a homogeneous Poisson process, the self-correcting process in Section 6.4 is more regular, and the non-linear Hawkes process in Section 6.3 can be both. We used this in Section 7.2 for checking whether the residual process behaved like a Poisson process as is expected if an adequately well-fitting model has been used for modelling the data. However, for a directed linear network the dependence structure is completely changed, and as a consequence the appropriate concepts of clustering and regularity are also changed. While the (undirected) geometrically corrected K-function certainly gives a good idea of the amount of clustering and regularity, the development of a directed geometrically corrected K-function is useful for quantifying such concepts in a more appropriate manner. We leave this as an object of future research.

Essentially a linear network consists of a superposition of line segments in $\mathbb{R}^d$, but in Anderes et al. (2017) they have been generalized to graphs with Euclidean edges, which extends the linear network in various ways to include curve segments, crossing (but unconnected) segments, etc. Such a generalization can rather easily be made to directed linear networks to obtain a directed version of graphs with Euclidean edges, and all results in the present paper immediately extends to this case (we have only avoided making this extension to avoid a more cumbersome notation in this paper).

Furthermore, Anderes et al. (2017) also consider two different metrics on graphs with Euclidean edges: the shortest path metric, i.e. the length of the shortest path along the edges of the graph, and the resistance metric, i.e. the metric corresponding to the resistance in an electrical network (see also Rakshit et al., 2017, for use of various metrics on linear networks). We note that the quasi-metric $d_L^{\rightarrow}$ is the natural directed counterpart of the shortest path metric, and it would be the natural choice for modelling many kinds of point pattern data on a directed linear network. However, any other quasi-metric on L can be used as a basis for building models on directed linear networks and may be relevant for practical applications where $d_L^{\rightarrow}$ is not appropriate.

The results in Section 7.1 suggest that at least in some cases the maximum likelihood estimator has nice asymptotic properties. Specifically, the maximum likelihood estimates for the Hawkes processes seem unbiased and consistent, while the estimates in the self-correcting process are strongly biased and correlated. A proper development of asymptotic theory in the line of Ogata (1978) and Rathbun (1996) is important to establish properties of the maximum likelihood estimator for point processes specified by a conditional intensity function on a directed linear network.

Section 7.2 presents a very short analysis of a dendrite dataset using the conditional intensity function to build a point proces model on a directed linear network. The main purpose of this is to illustrate that the models and methods can be applied to real data. As a future research direction we plan to make a much more thorough analysis of the presented dendrite dataset and other similar datasets, where we will also model the spine types as marks, and derive practical results from the models with biological relevance.

Acknowledgements The authors would like to thank Abdel-Rahman Al-Absi who collected the dendrite data, and Adrian Baddeley, Gopalan Nair and Valerie Isham for preliminary discussions on ideas for the theory in the paper. This work was supported by The Danish Council for Independent Research - Natural Sciences, grant DFF - 7014-00074 "Statistics for point processes in space and beyond", by the "Centre for Stochastic Geometry and Advanced Bioimaging", funded by grant 8721 from the Villum Foundation, and by the Australian Research Council grant DP130102322, "Statistical methodology for events on a network, with application to road safety".

References

Anderes, E., Møller, J., and Rasmussen, J. G. (2017). Isotropic covariance functions on graphs and their edges. Available at arXiv:1710.01295.

Ang, Q. W., Baddeley, A. J., and Nair, G. (2012). Geometrically corrected second order analysis of events on a linear network, with applications to ecology and criminology. *Scandinavian Journal of Statistics*, 39:591–617.

Baddeley, A., Rubak, E., and Turner, R. (2016). *Spatial Point Patterns: Methodology and Applications with R*. Chapman & Hall/CRC, New York.

Baddeley, A. J., Jammalamadaka, A., and Nair, G. (2014). Multitype point process analysis of spines on the dendrite network of a neuron. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 63:673–694.

Baddeley, A. J., Nair, G., Rakshit, S., and McSwiggan, G. (2017). ”Stationary” point processes are uncommon on linear networks. *Stat*, 6:68–78.

Brémaud, P. and Massoulié, L. (1994). Imbedded construction of stationary point processes and sequences with random memory. *Queueing Systems*, 17:213–234.

Brémaud, P. and Massoulié, L. (1996). Stability of nonlinear Hawkes processes. *The Annals of Probability*, 24:1563–1588.

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

Garreta, V., Monestiez, P., and Hoef, J. M. V. (2010). Spatial modelling and prediction on river networks: up model, down model or hybrid? *Environmetrics*, 221:439–4562.

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

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

Hawkes, A. G. (1972). Spectra of some mutually exciting point processes with associated variables. In Lewis, P. A. W., 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.

Irwin, S. A., Patel, B., Idupulapati, M., Harris, J. B., Crisostomo, R. A., Larsen, B. P., Kooy, F., Willems, P. J., Cras, P., Kozlowski, P. B., Swain, R. A., Weiler, I. J., and Greenough, W. T. (2001). Abnormal dendritic spine characteristics in the temporal and visual cortices of patients with fragile-X syndrome: a quantitative examination. *American Journal of Medical Genetics*, 98:161–167.

Isham, V. and Westcott, M. (1979). A self-correcting point process. *Stochastic Processes and their Applications*, 8:335–347.

Jammalamadaka, A., Banerjee, S., Manjunath, B. S., and Kosik, K. S. (2013). Statistical analysis of dendritic spine distributions in rat hippocampal cultures. *BMC Bioinformatics*, 14:287.

McSwiggan, G., Baddeley, A. J., and Nair, G. (2016). Kernel density estimation on a linear network. *Scandinavian Journal of Statistics*, 44:324–345.

Møller, J. and Rasmussen, J. G. (2005). Perfect simulation of Hawkes processes. *Advances in Applied Probability*, 37:629–646.

Møller, J. and Rasmussen, J. G. (2006). Approximate simulation of Hawkes processes. *Methodology and Computing in Applied Probability*, 8:53–64.

Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H., and Hahn, U. (2017). Global envelope tests for spatial processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79:381–404.

Ogata, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*, 30:243–261.

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

Okabe, A. and Yamada, I. (2001). The *K*-function method on a network and its computational implementation. *Geographical analysis*, 33:271–290.

Rakshit, S., Nair, G., and Baddeley, A. J. (2017). Second-order analysis of point patterns on a network using any distance metric. *Spatial Statistics*, 22:129–154.

Rathbun, S. L. (1996). Asymptotic properties of the maximum likelihood estimator for spatio-temporal point processes. *Journal of Statistical Planning and Inference*, 51:55–74.

Ver Hoef, J. M. and Peterson, E. E. (2012). A moving average approach for spatial statistical models of stream networks. *Journal of the American Statistical Association*, 105:6–18.

Ver Hoef, J. M., Peterson, E. E., and Theobald, D. (2006). Spatial statistical models that use flow and stream distance. *Environmental and Ecological Statistics*, 13:449–464.

Paper D

Modelling spine locations on dendrite trees using inhomogeneous Cox point processes

Heidi S. Christensen and Jesper Møller

The paper has been submitted to
Biometrics.

The layout has been revised.

Abstract

Dendritic spines, which are small protrusions on the dendrites of a neuron, are of interest in neuroscience as they are related to cognitive processes such as learning and memory. We analyse the distribution of spine locations on six different dendrite trees from mouse neurons using point process theory for linear networks. Besides some possible small-scale repulsion, we find that one of the spine point pattern data sets may be described by an inhomogeneous Poisson process model, while the other point pattern data sets exhibit clustering between spines at a larger scale. To model this we propose an inhomogeneous Cox process model constructed by thinning a Poisson process on a linear network with retention probabilities determined by a spatially correlated random field. For model checking we consider network analogues of the empirical F-, G-, and J-functions originally introduced for inhomogeneous point processes on a Euclidean space. The fitted Cox process models seem to catch the clustering of spine locations between spines, but also posses a large variance in the number of points for some of the data sets causing large confidence regions for the empirical F- and G-functions.

1 Introduction

Point patterns on linear networks arise in a broad range of fields, where the network for example represents roads, a river network, or a dendrite tree. This paper focuses on the latter type of data: the left panel in Figure D.1 shows six linear networks each representing a dendrite tree from a mouse neuron grown in vivo. On the dendrites small protrusions called spines are found that among other things help transmitting electrical signals to the soma. In neuroscience, the behaviour of spines are of interest as changes can be linked to changes in cognitive processes. The spine locations can be viewed as a point pattern on the dendrite tree and thus analysed using point process theory for linear networks.

Over the last two decades, methods for analysing point patterns on linear networks have been developed. Particularly, a network analogue of Ripley's K-function was first presented in Okabe and Yamada (2001) and later modified and extended to the inhomogeneous case in Ang et al. (2012). When defining the K-function, Ang et al. (2012) required that the underlying point process model fulfils an invariance property called second-order pseudo-stationarity (an analogue to second-order intensity-reweighted stationarity as introduced in Baddeley et al., 2000). This property is fulfilled whenever the pair correlation function is isotropic, i.e. when it only depends on the shortest path distance. Baddeley et al. (2017) showed that certain constructions, e.g. special types of Cox point processes that lead to point processes in the Euclidean space with an isotropic pair correlation function rarely result in second-order pseudo-stationary point processes when adapted to linear networks. Even without the requirement of pseudo-stationarity, there are only a limited number of point process models available for linear networks. For point processes on directed acyclic linear networks, Rasmussen and Christensen (2019) presented both regular and clustered models defined by a generalisation of the conditional intensity function for temporal point processes. Anderes et al. (2017) supplied a list of valid isotropic covariance functions for connected linear networks that can be used to construct Cox point processes, particularly log Gaussian Cox processes (LGCPs; see also Møller et al., 1998).

Only few studies use point process theory to analyse the behaviour of spines: treating the dendrite tree as a directed tree, Rasmussen and Christensen (2019) analysed one of the six spine point pattern data sets (hereafter, the 'spine data') from Figure D.1. The distribution of spines (and their shape) have further been investigated using point process theory in Jammalamadaka et al. (2013) (testing a homogeneous Poisson process model) and Baddeley et al. (2014) (using multitype Poisson process models to account for the shape classification) for in vitro grown neurons. Based on the network *K*-function, Jammalamadaka et al. (2013) concluded that a homogeneous Poisson process model seems adequate to describe the spine locations. However, Jammalamadaka et al. (2013) also stated that their results for the in vitro setting are unlikely to hold in an in vivo setting.

Instead of Poisson process models, this paper suggests a new class of Cox process models on a linear network. Such a model applies for an undirected graph and is not a LGCP, but its construction still exploits a Gaussian random field so that the covariance functions from Anderes et al. (2017) become useful. Moreover, seemingly for the first time in connection to point process model fitting on linear networks, we demonstrate the use of minimum contrast and composite likelihood estimation procedures. Finally, we introduce new empirical summary functions and demonstrate their usefulness for model checking.

The paper is organised as follows. The spine data is described in more detail in Section 2 along with the general notion of a linear network. In Section 3 we discuss existing as well as our new summary functions for point processes on linear networks; these are used for analysing the spine locations in Section 4. We initially suggest to model the spine locations by an inhomogeneous Poisson process model in Section 4.1, but due to clustering between spines we propose in Section 4.2 an inhomogeneous Cox process model. Lastly, we discuss in Section 5 possible extensions and future research directions.

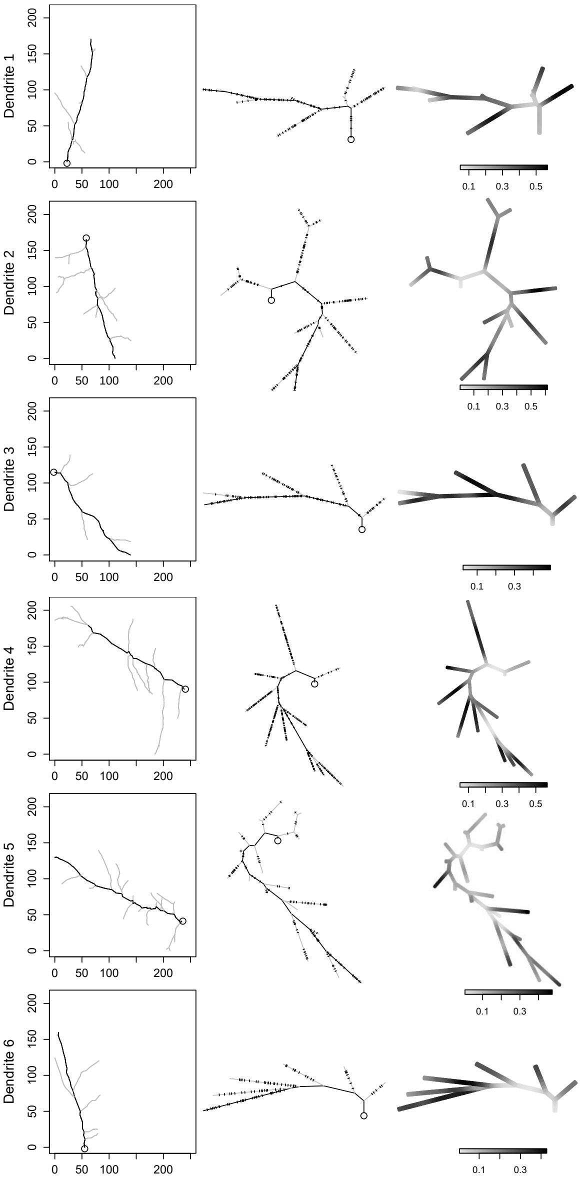


Figure D.1: Spine data sets for dendrite 1 to 6 (from top to bottom), where each main branch is coloured black and the side branches grey and the ○ marks the vertex closest to the dendrite's attachment to soma. Left: projection of the original three-dimensional network onto a plane. Middle: spine locations on the simplified networks embedded in $\mathbb{R}^2$ so that distances are preserved. Right: non-parametric kernel intensity estimates.

2 Dendritic spine data

Figure D.1 shows six examples of a linear network. Specifically, a linear network is a union $L = \bigcup_{i=1}^{N} L_i$ of a finite number N of line segments $L_i \subset \mathbb{R}^d$, $d \geq 2$, with finite length and intersecting only at the end points. A linear network may also be viewed as a graph consisting of a set of vertices and a set of weighted edges, where the edges coincide with the line segments $L_1, \ldots, L_N$, the vertices correspond to the end points of these line segments, and the weight of an edge is the length of the corresponding line segment.

Throughout this paper the distance between two points $u, v \in L$ is measured by the shortest path distance and is denoted by $d_L(u, v)$. For the spine data, the linear network L is a tree, meaning that there is only one path between any pair of points in L. Naturally, in other applications more complicated networks than a tree occur in which case we may need to take more care when choosing the distance metric d_L (see Section 5 for details).

The linear networks visualised in the left column of Figure D.1 are approximations of the underlying apical dendrite trees which were extracted from six different mouse neurons grown in vivo. The vertices of each of the linear networks are described by three-dimensional coordinates which represent a spine location or another point chosen to obtain the approximation. To simplify each of the networks, for any pair of edges meeting at a vertex with degree two, we replace the two edges and the vertex with one single edge, and to preserve distances within the network, we let the weight of the new edge be the sum of the two old edges. Note that this transformation of the original linear network into a simpler one, relies on the fact that the linear network forms a tree. Further, to utilise the functionalities of the `R`-package `spatstat` (Baddeley et al., 2016), we embed the simplified network in $\mathbb{R}^2$ in a way that also preserves distances. The simplified and embedded versions of the networks are shown in the middle column of Figure D.1 along with the spine locations. As the models and tools we use in Section 4 to analyse the spine data do not directly depend on the three-dimensional coordinates but on distances, we can without loss of information consider the spine locations as a point pattern on the simplified and embedded network.

The spine data origin from six neurons, with two neurons from each of three different mice. The numbering of the dendrites is as follows: dendrite 1 and 2 come from mouse no. 1; dendrite 4 and 5 from mouse no. 2; and dendrite 3 and 6 from mouse no. 3.

For each dendrite tree, we talk about two subsets: the main branch and the side branches. *Main branch* refers to the tree's stem, while *side branches* constitute the rest of the tree. Figure D.1 shows which parts of the original trees and the simplified embedded trees belong to the main branch and which to the side branches. In the following, $L = L_m \cup L_s$ denotes the whole dendrite tree, where L_m is the main branch, and L_s is the union of the side branches. Further, n_m and n_s denote the number of spines on L_m and L_s, respectively. Lastly, we let $|B|$ denote the size of $B \subseteq L$ or more precisely the total length of the (partial) line segments constituting $B \subseteq L$; note that $|L| = |L_m| + |L_s|$. Table D.1 summarises the number of spines and sizes for each dendrite tree.

Dendrite	n_m	n_s	$\lvert L_m\rvert$	$\lvert L_s\rvert$	$\hat{\rho}_m$	$\hat{\rho}_s$
1	51	72	212 µm	202 µm	0.240	0.356
2	36	145	204 µm	430 µm	0.176	0.337
3	69	63	211 µm	202 µm	0.328	0.312
4	33	308	225 µm	652 µm	0.134	0.477
5	34	83	286 µm	450 µm	0.119	0.184
6	30	62	178 µm	250 µm	0.168	0.248

Table D.1: Number of spines, length, and intensity estimates for the main and side branches separately.

3 Point processes on linear networks

The six spine point pattern data sets are modelled as realisations of six point processes defined on the six dendrite trees. In general, by a point process X on a linear network L we mean a random finite subset of L; we use this generic notation throughout this paper. In this section we consider summary functions useful for analysing point processes on linear networks, including the introduction of new empirical summary functions.

3.1 Summary functions for first and second-order moment properties

We assume that X has intensity ρ, that is, for $B \subseteq L$,

$$\mathrm{E}\, n(X \cap B) = \int_B \rho(u)\, \mathrm{d}_L u < \infty, \tag{D.1}$$

where $n(X \cap B)$ is the number of points from X falling in B and d_L denotes integration with respect to one-dimensional arc-length along L. Intuitively, $\rho(u)\, \mathrm{d}_L u$ is the probability of X having a point in an infinitesimal small subset of L that contains u and has size $\mathrm{d}_L u$. If the intensity $\rho(\cdot) \equiv \rho$ is constant, we say that X is homogeneous; otherwise X is said to be inhomogeneous. In case of homogeneity, ρ is the expected number of points per unit length.

We also assume that X has pair correlation function g, that is, for disjoint $A, B \subset L$,

$$\mathrm{E}\left\{n(X \cap A)n(X \cap B)\right\} = \int_A \int_B g(u,v)\rho(u)\rho(v)\, \mathrm{d}_L u\, \mathrm{d}_L v < \infty.$$

We can interpret $g(u,v)\rho(u)\rho(v)\, \mathrm{d}_L u\, \mathrm{d}_L v$ as the joint probability that two infinitesimal small regions around u and v of size $\mathrm{d}_L u$ and $\mathrm{d}_L u$, respectively, each contains a point from X.

If the pair correlation function only depends on the shortest path distance, we say that it is isotropic and write $g(u,v) = g_0\{d_L(u,v)\}$. When X has an isotropic pair correlation function, the (geometrically corrected network) K-function introduced by Ang et al. (2012) can be expressed as

$$K(r) = \int_0^r g_0(s)\, \mathrm{d}s, \qquad r \geq 0. \tag{D.2}$$

Alternatively, $K(r)$ may be written as an expectation with respect to a Palm distribution, see Ang et al. (2012, Theorem 3). If the K-function or the pair correlation function is expressible on closed form, we can use a minimum contrast or composite likelihood procedure to estimate the model parameters; this is described further in Section 4.2.

The simplest point process model is a Poisson process, which is characterised by that $n(X)$ follows a Poisson distribution with mean given by (D.1) with $B = L$ and further that the points of X conditioned on $n(X)$ are independent and identically distributed, with density proportional to ρ. For a Poisson process, $g \equiv 1$ and $K(r) = r$.

3.2 New empirical summary functions

For estimating the pair correlation function and the K-function we follow Ang et al. (2012). These empirical summary functions can be used in minimum contrast or composite likelihood estimation procedures as well as for model checking. Obviously, if the K-function or pair correlation function have been used to fit the model, neither should be used to check the adequacy of the model. Due to the shortage of summary functions for point processes on linear networks, we may let a simple visual comparison of the observed point pattern and simulations from the fitted model serve as a model check. It is needless to say that a more rigorous model checking would be preferred.

Therefore, we now introduce three purely empirical summary functions. These are obtained by modifying the empirical F-, G-, and J-functions for inhomogeneous point patterns on a Euclidean space (introduced by van Lieshout, 2011) to linear networks. The modification simply consists of replacing the Euclidean space with the linear network, introducing the shortest path distance instead of the Euclidean distance, and adapting the notion of an eroded set to linear networks. The functions are then defined as follows. Assume that the intensity ρ is known or has been estimated by $\hat{\rho}$ and that $\bar{\rho} = \inf_{u \in L} \tilde{\rho}(u) > 0$, where either $\tilde{\rho} = \rho$ or $\tilde{\rho} = \hat{\rho}$. For $r \geq 0$, let $L_{\ominus r}$ consist of the points in L with distance greater than r to any vertex of L with degree one. Furthermore, let $H \subset L$ be a finite 'lattice'. For an observed point pattern $X = x$, the empirical summary functions $\hat{F}$, $\hat{G}$, and $\hat{J}$ are then defined for $r \geq 0$ by

$$\hat{F}(r) = 1 - \frac{\sum_{v \in H \cap L_{\ominus r}} \prod_{u \in x:\, d_L(u,v) \leq r} \left\{1 - \frac{\bar{\rho}}{\tilde{\rho}(u)}\right\}}{\#(H \cap L_{\ominus r})}, \tag{D.3}$$

$$\hat{G}(r) = 1 - \frac{\sum_{v \in x \cap L_{\ominus r}} \prod_{u \in x \setminus \{v\}:\, d_L(u,v) \leq r} \left\{1 - \frac{\bar{\rho}}{\tilde{\rho}(u)}\right\}}{\#(x \cap L_{\ominus r})}, \tag{D.4}$$

$$\hat{J}(r) = \frac{1 - \hat{G}(r)}{1 - \hat{F}(r)}, \tag{D.5}$$

where we restrict attention to r-values small enough to ensure that $\#(H \cap L_{\ominus r}) > 0$ for $\hat{F}(r)$ and that $\hat{F}(r) < 1$ for $\hat{J}(r)$.

In Ang et al. (2012), the K-function and its empirical estimate include a factor that corrects for the network geometry, such that its shape can be compared for point patterns on different networks. As it was not obvious to us how to extend such a correction to $\hat{F}$, $\hat{G}$, and $\hat{J}$ in a meaningful way, our definitions in (D.3)–(D.5) do not correct for the network geometry. Further, we do not have any theoretical counterpart to $\hat{F}$, $\hat{G}$, and $\hat{J}$ and therefore their shapes alone can in general not be used to conclude anything about e.g. the presence of regularity or clustering. However, $\hat{F}$, $\hat{G}$, and $\hat{J}$ are still useful tools for providing a so-called global rank envelope; this is a confidence region for a given test function obtained from simulations under a fitted model (for details, see Myllymäki et al., 2017). In a global rank envelope procedure, the shape of the test function for the data is compared to that of the simulations and Myllymäki et al. (2017) discussed how this provides a test and an interval with lower and upper bounds given by a liberal and a conservative p-value, respectively.

4 Modelling spine locations

In this section each of the six data sets is analysed with the aim of finding a model that adequately describe the spine locations.

4.1 Fitting a Poisson process model

The simplest model we can propose is a Poisson process. To investigate the behaviour of the spine intensity, we calculated the non-parametric intensity estimate suggested by McSwiggan et al. (2016); the resulting estimates are seen in the right panel of Figure D.1. The spine intensity tend to be higher on the side branches than on the main branch, except perhaps for dendrite 3. Therefore, we allowed the intensity of the Poisson process to be different on the main and side branches, that is, recalling the notation in Section 2,

$$\rho(u) = \rho_m^{\mathbb{I}(u \in L_m)} \rho_s^{\mathbb{I}(u \in L_s)}, \qquad u \in L, \tag{D.6}$$

for non-negative parameters ρ_m and ρ_s; here $\mathbb{I}(\cdot)$ denotes the indicator function. The maximum likelihood estimates of the intensity parameters are easily found and given by

$$\hat{\rho}_m = \frac{n_m}{|L_m|}, \qquad \hat{\rho}_s = \frac{n_s}{|L_s|}, \tag{D.7}$$

cf. the notation in Section 2. These estimates are shown in Table D.1.

To test whether the proposed inhomogeneous Poisson process model adequately describes the spine locations, we performed global rank envelope tests using K as test function, cf. Section 3.2. Results from these tests are shown in Figure D.2. For all dendrites the conservative p-value is below 3%, suggesting that the fitted Poisson process models do not describe the spine locations adequately. Specifically, the empirical K-functions for dendrite 2, 4, 5, and 6 fall above the global rank envelopes for certain r-values, indicating that the spines tend to cluster at these distances. Further, for all six spine data sets the empirical K-function falls below the global rank envelope for small r-values, indicating a small-scale repulsion between spines. For dendrite 1 and 3, this small-scale repulsion is the only deviation from the Poisson process model revealed by the global rank envelope test. Disregarding the small distances ($r < 1\,\mu\text{m}$) for the global rank envelope test with K as test function, does not change the p-intervals significantly for dendrite 2, 4, 5, and 6. For dendrite 1 and 3 on the other hand, the p-intervals change from $(0.024, 0.040)$ to $(0.048, 0.060)$ and from $(0, 0.019)$ to $(0.160, 0.168)$, respectively, giving (most clearly for dendrite 3) no evidence against the proposed Poisson process model. Global rank envelopes with a concatenation of $\hat{F}$, $\hat{G}$, and $\hat{J}$ as test function, and where distances less than $1\,\mu\text{m}$ were disregarded, are shown in Figure D.9 in Appendix C; these do not provide any evidence against the Poisson process model for dendrite 1 and 3 either.

As the physical scale of the spine data is quite small (the dendrites range in size from $412\,\mu\text{m}$ to $876\,\mu\text{m}$, cf. Table D.1) and as there is uncertainty in the exact choice of the point representing a spine's location, we must expect some degree of imprecision and therefore we may not want to put too much value into the observed small-scale repulsion. In the following we will not take the small-scale repulsion into account but rather focus on modelling the large scale clustering.

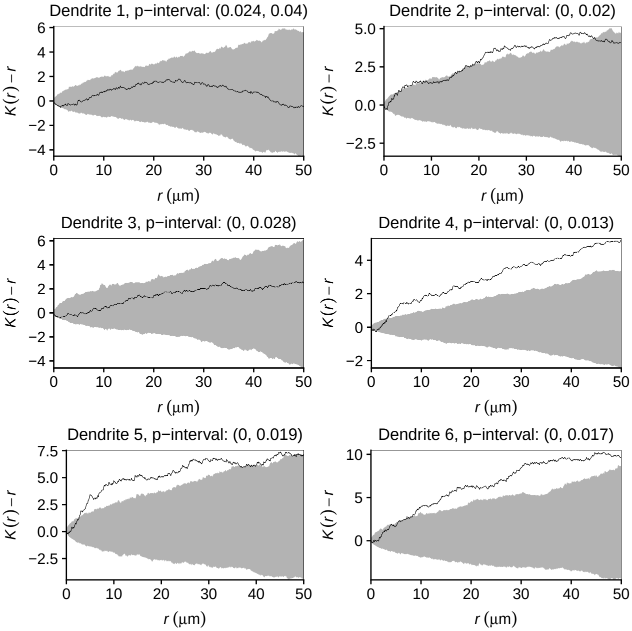


Figure D.2: For each spine data set: the empirical *K*-function for the data minus the *K*-function for a Poisson process (solid line) along with 95% global rank envelopes (grey region) based on 2499 simulations from the fitted inhomogeneous Poisson process model; *p*-intervals for each of the associated global rank envelope tests are also displayed.

4.2 Fitting a Cox process model

In addition to inhomogeneity in the location of spines, Figure D.1 also shows a tendency to large bare areas where no spines occur. To model this behaviour we considered a point process model introduced by Lavancier and Møller (2016) for the Euclidean space, which is easily adapted to a linear network L. The point process is constructed as a thinning of a Poisson process Y on L with intensity function ρ_Y, where the retention probabilities are determined by a random field $\Pi = \{\Pi(u) : u \in L\}$ that may be spatially correlated. That is, the point process is given by $X = \{u \in Y : \Pi(u) \geq R(u)\}$, where $R = \{R(u) : u \in L\}$ consist of independent uniform random variables on $[0, 1]$, and where Π and R are independent. Thus X is a Cox point process driven by the random field $\Lambda = \{\rho_Y(u)\Pi(u) : u \in L\}$. We let

$$\Pi(u) = \exp\left\{-\frac{\sigma^2}{2}\sum_{j=1}^{k} Z_j^2(u)\right\}, \qquad u \in L, \tag{D.8}$$

where $k \in \{1, 2, \dots\}$ and $\sigma^2 > 0$ are parameters, and $Z_1, \dots, Z_k$ are IID zero-mean unit-variance Gaussian random fields with correlation function c. If $c(u,v) = c_0\{d_L(u,v)\}$ depends only on the shortest path distance, we say that c is isotropic; see Anderes et al. (2017) for a list of isotropic correlation functions for linear networks. For the spine data, we considered the exponential correlation function, that is,

$$c(u,v) = \exp\{-\beta d_L(u,v)\}, \qquad u, v \in L, \tag{D.9}$$

which is a valid correlation function for any parameter value $\beta > 0$ and any tree network but not necessarily for other kinds of linear networks (see Section 5 for a comment on this).

We have that $\mathrm{E}\,\Pi(u) = (1 + \sigma^2)^{-k/2}$ and

$$\mathrm{E}\,\{\Pi(u)\Pi(v)\} = \{(1 + \sigma^2)^2 - (\sigma^2)^2 c(u,v)^2\}^{-k/2},$$

implying that X has intensity

$$\rho(u) = (1 + \sigma^2)^{-k/2}\rho_Y(u), \qquad u \in L, \tag{D.10}$$

and pair correlation function

$$g(u,v) = \left\{\frac{(1 + \sigma^2)^2}{(1 + \sigma^2)^2 - (\sigma^2)^2 c(u,v)^2}\right\}^{k/2}, \qquad u, v \in L,\ u \neq v. \tag{D.11}$$

If c is isotropic, then g is isotropic and the K-function can be expressed by (D.2). Closed form expressions of the K-function are given in Appendix A for c equal to the exponential correlation function and $k = 1, \dots, 5$.

Note that for each $u \in L$, $\sigma^2 \sum_{j=1}^{k} Z_j^2(u)$ in (D.8) follows a $\sigma^2\chi^2$-distribution with k degrees of freedom. That is, the skewness decreases as k increases, while the range is stretched/compressed depending on the value of σ^2. The larger σ^2 is, the more Y is thinned to obtain X and also the more variation in the thinning probabilities. The pair correlation function in (D.11) is an increasing function of both σ^2 and k. When c is given by (D.9), β controls the correlation of the retention probabilities: the smaller β, the longer range of correlation in Π and thus larger coherent bare/populated areas in X. Finally, the pair correlation function decreases towards 1, when β increases.

For the spine data, we still want to model the intensity function by (D.6), which by (D.10) requires Y to have a similar intensity structure, that is,

$$\rho_Y(u) = \rho_{Y,m}^{\mathbb{I}(u \in L_m)} \rho_{Y,s}^{\mathbb{I}(u \in L_s)}, \qquad u \in L, \tag{D.12}$$

for non-negative parameters $\rho_{Y,m}$ and $\rho_{Y,s}$.

Simulation

To perform model checking with global rank envelopes or to carry out simulation studies, we need to be able to simulate point patterns from the model of interest. Fortunately, it is straightforward to simulate a point pattern on L from the proposed Cox process model by the following three steps:

(a) Simulate a discretised version of the random field Π by first simulating the independent Gaussian random fields Z_j, $j = 1, \ldots, k$, at chosen grid locations along the network and then transforming the random fields according to (D.8) to obtain the retention probabilities.

(b) Simulate a point pattern y from a Poisson process on L with intensity ρ_Y.

(c) Thin y using the retention probabilities simulated in (a).

Estimation procedure

In the following we describe a procedure for estimating the model parameters of the proposed Cox process model where. For specificity we consider the case where c is given by (D.9) and where ρ_Y is given by (D.12).

To begin we assume that k is known, whereas the remaining parameters are estimated through a two-step procedure (Waagepetersen, 2007; Waagepetersen and Guan, 2009). In short, we first estimate (ρ_m, ρ_s) and then plug in these estimates in a second-order procedure where (σ^2, β) is estimated. Lastly, an estimate of $(\rho_{Y,m}, \rho_{Y,s})$ can be found by using (D.10).

First, to estimate (ρ_m, ρ_s) we use the first order composite likelihood (Waagepetersen, 2007) which simply corresponds to a Poisson likelihood yielding the estimates in (D.7).

Second, as we know explicit formulas for the pair correlation function and the K-function, we can estimate (σ^2, β) using a minimum contrast procedure (Guan, 2009; Diggle, 2014) or a second-order composite likelihood approach (Waagepetersen, 2007; Lavancier et al., 2018). The latter is not considered here but described in Appendix B, where results from a simulation study comparing the two approaches also can be found. The simulation study suggests that the minimum contrast procedure provides far more reliable estimates than the second-order composite likelihood.

For a chosen summary function $T_{(\sigma^2,\beta)}$ which depends on (σ^2, β), the minimum contrast estimate of (σ^2, β) is given by

$$(\hat{\sigma}^2, \hat{\beta}) = \underset{(\sigma^2,\beta)}{\arg\min} \int_{r_l}^{r_u} \left\{ \hat{T}(r)^p - T_{(\sigma^2,\beta)}(r)^p \right\}^2 \, \mathrm{d}r, \tag{D.13}$$

where $0 \leq r_l < r_u$ and $p > 0$ are user-specified tuning parameters, and $\hat{T}$ is an empirical estimate of the summary function. In our case, $T_{(\sigma^2,\beta)}$ is given by K or g_0, and $\hat{T}$ is given as in Ang et al. (2012). A frequently seen choice is $r_l = 0$, while general recommendations of r_u and p can be found in Guan (2009) and Diggle (2014) for point patterns on the Euclidean space.

The minimum contrast procedure can easily be extended to include estimation of k too, but a simulation study indicated that it may be difficult to estimate σ^2 and k simultaneously as an increase in k seemingly can be balanced out by an increase in σ^2. In practice we may therefore simply make a choice of k; for simplicity we chose $k = 1$ in the following. Note that for both $T = K$ and $T = g_0$ the estimate $\hat{T}$ in (D.13) depends on the intensity (Ang et al., 2012); here we simply plug-in the estimated intensity obtained in the first step of the estimation procedure.

A drawback of using $T = g_0$ is the need of choosing a bandwidth for the non-parametric kernel estimate $\hat{g}_0$ presented in Ang et al. (2012). However, the simulation study in Appendix B suggests that using $T = g_0$ with the default bandwidth and kernel from the `spatstat`-package generally performs better than $T = K$ when fitting the proposed Cox process model. This is consistent with results from a simulation study in Lavancier and Møller (2016) for point processes on a Euclidean space.

In the simulation study found in Appendix B, we also investigated how different choices of r_l, r_u, and p affect the estimates of σ^2 and β given by (D.13). We observed that the choice of r_u often is a matter of trade-off between bias and variance: a large value of r_u may entail a large bias, while a small r_u often leads to a greater variance of the estimates. Furthermore, the best choice of r_u seems to be quite depending on what the true underlying model parameters are. For example, a larger range of correlation in the retention probabilities, that is, a smaller value of β, requires a larger r_u. Naturally we should also take the size of the network into consideration when choosing

Dendrite	$\hat{\rho}_{Y,m}$	$\hat{\rho}_{Y,s}$	$\hat{\sigma}^2$	$\hat{\beta}$
1	0.312	0.463	0.686	0.037
2	0.275	0.525	1.427	0.020
3	0.328	0.312	5.170×10^{-8}	30.178
4	0.197	0.701	1.159	0.010
5	0.266	0.413	4.023	0.030
6	0.322	0.474	2.662	0.013

Table D.2: Estimates of $\rho_{Y,m}$, $\rho_{Y,s}$, σ^2, and β for each spine data set.

r_u. In the simulation study we found that $r_l = 0$ gives the best estimates, and that $p = 1$ or $p = 1/2$ behave equally well for $T = g_0$, while $p = 1/4$ is preferred over $p = 1/2$ for $T = K$. For parameter values yielding models close to the Poisson process model, that is, when σ^2 is close to zero or β is large, the estimation procedures were not very successful regardless of the tuning parameters. This does not come as a surprise as many combinations of σ^2- and β-values yield similar Poisson processes. Lastly, the estimation procedure seems quite stable with respect to the choice of start parameter values for the optimisation algorithm (`optim` in `R`) used to minimize (D.13).

Model fit and model check

The Cox process model was fitted to each of the spine data sets using the two-step procedure with $k = 1$ fixed, cf. Section 4.2. For the minimum contrast procedure we let $T = g_0$, $p = 1$, and $r_l = 0$ in accordance with the simulation results discussed in Section 4.2. Further, we initially let $r_u = 15$ and obtained a set of initial parameter estimates for each data set. Then we performed a small simulation study based on 500 simulations from the initially fitted models to investigate which of $r_u = 15, 30, 50$ results in the best estimates (in terms of bias and variance) for these specific models. For dendrite 1, 3, and 6, the initially fitted models are close to the Poisson process case, and as a consequence the model parameters are hard to estimate regardless of the choice of r_u. However, for dendrite 2, 4, and 5, it seems that $r_u = 50$ is the best choice. Using $r_u = 50$ for all six data sets, we obtained the parameter estimates in Table D.2. The fitted model for dendrite 3 is practically a Poisson process model (in consistency with the conclusions made in Section 4.1) while the remaining fitted models are not. In Figure D.10 in Appendix C, one simulation from each of the fitted random fields Π is shown to illustrate the behaviour of the retention probabilities. For example, $\hat{\sigma}^2$ is considerably larger for dendrite 5 than dendrite 1, resulting in more fluctuating retention probabilities.

As discussed in Section 3.2, for the statistical analyses of point patterns on

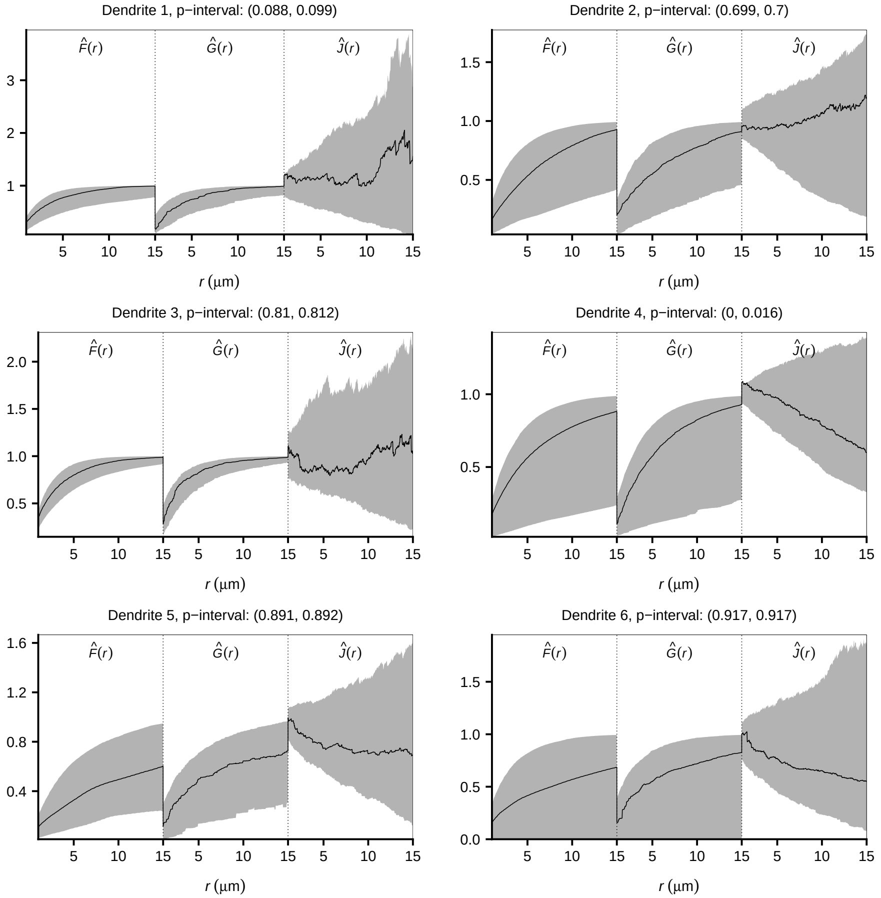


Figure D.3: For each spine data set: the concatenation of $\hat{F}$, $\hat{G}$, and $\hat{J}$ for the spine locations (black solid line) along with 95% global rank envelopes (grey region) based on 2499 simulations from the fitted Cox process model; p-intervals for each of the associated global rank envelope tests are also displayed.

linear networks there is only a limited number of options for model checking, especially when the K-function or the pair correlation function have already been used to estimate the model parameters. One simple option is to look at simulations from the fitted model as in Figures D.11–D.16 in Appendix C. Comparing these simulations visually to the observed point patterns, it seems that the simulations mimic the behaviour of the data quite well. For a more rigorous model checking, we performed global rank envelope tests with a concatenation of $\hat{F}$, $\hat{G}$, and $\hat{J}$ as test function, where distances less than 1 µm were disregarded as discussed in Section 4.1; results are shown in Figure D.3. Except from dendrite 4, where the data curve for $\hat{J}$ falls below the global rank envelope for some of the smaller r-values, the tests do not reveal any evidence against the fitted models. However, for some of the dendrites (especially dendrite 6) the global rank envelopes for the part concerning $\hat{F}$ and $\hat{G}$ are very wide due to a large variance in the number of points.

5 Discussion

As discussed in Section 4.1, the spines seem to posses a small-scale repulsion, which was not modelled by the Cox process proposed in Section 4.2. To accommodate the repulsion, the inhomogeneous Poisson process Y used to build the Cox process could be replaced by an inhomogeneous and repulsive point process. The simplest case may be to use a dependent thinning as in a Matérn hard core process of type I (Matérn, 1960; Matérn, 1986): let $\tilde{Y}$ be an inhomogeneous Poisson process (with constant intensity on the main branch respectively the side branches), and let

$$Y = \{u \in \tilde{Y} : d_L(u,v) > h \text{ for all } v \in \tilde{Y} \setminus \{u\}\},$$

where $h > 0$ is a hard core parameter; that is, a point in $\tilde{Y}$ is included in Y if and only if no other point in $\tilde{Y}$ is within distance h. However, it is doubtful whether an expression for the K-function or the pair correlation function can be found for such a Cox process model, posing new challenges with respect to parameter estimation.

To avoid using the rather ad hoc created summary functions $\hat{F}$, $\hat{G}$, and $\hat{J}$, it is needed to develop new summary functions for (inhomogeneous) point processes on linear networks for which we both have a theoretical interpretation and an estimate that do not depend on the geometry of the network. We leave this challenging issue for future research.

Rakshit et al. (2017) discussed the importance of how distance is measured when analysing point patterns on a linear network and they generalised the K-function to allow the use of any distance metric. In fact, following Anderes et al. (2017) all methods as well as Poisson and Cox process models in this paper immediately apply for more general linear networks, called graphs with

Euclidean edges, when the correlation function c is isotropic with respect to the shortest path distance as well as another metric called the resistance metric. For the dendrite networks or any other tree network, the resistance metric is equivalent to the shortest path distance. Anderes et al. (2017) showed that correlation functions that are isotropic with respect to the shortest path distance only are guaranteed to be valid for a small class of linear networks, whereas they are valid for any linear network when considering the resistance metric instead. Thus, depending on the network, it may be preferable to consider the resistance metric over the shortest path distance when specifying a correlation function. Anderes et al. (2017) provided a list of valid isotropic covariance functions for graphs with Euclidean edges.

Acknowledgements We would like to thank Abdel-Rahman Al-Absi who collected the spine data. We have been supported by The Danish Council for Independent Research | Natural Sciences, grant DFF – 7014-00074 "Statistics for point processes in space and beyond", and by the "Centre for Stochastic Geometry and Advanced Bioimaging", funded by grant 8721 from the Villum Foundation.

A Expressions for K

For the Cox point process presented in Section 4.2 with c equal to the exponential correlation function in D.9, the K-function is

$$K(r) = \int_0^r \left\{ 1 - \frac{\exp(-2\beta t)}{(1+1/\sigma^2)^2} \right\}^{-k/2} \mathrm{d}t. \tag{D.14}$$

Let $\alpha = \left(1 + 1/\sigma^2\right)^{-2}$, then $K(r)$ is given by

$$\frac{1}{\beta} \left\{ \log\left(\sqrt{\mathrm{e}^{2\beta r} - \alpha} + \mathrm{e}^{\beta r} \right) - \log\left(\sqrt{1-\alpha} + 1 \right) \right\}$$

if $k = 1$,

$$\frac{1}{2\beta} \log\left(\frac{\mathrm{e}^{2\beta r} - \alpha}{1 - \alpha} \right)$$

if $k = 2$,

$$\frac{1}{\beta} \left\{ \frac{\sqrt{\mathrm{e}^{2\beta r} - \alpha}\log\big(\sqrt{\mathrm{e}^{2\beta r} - \alpha} + \mathrm{e}^{\beta r}\big) - \mathrm{e}^{\beta r}}{\mathrm{e}^{\beta r}\sqrt{1 - \alpha \mathrm{e}^{-2\beta r}}} + \frac{1 - \sqrt{1-\alpha}\log\big(\sqrt{1-\alpha} + 1\big)}{\sqrt{1-\alpha}} \right\}$$

if $k = 3$,

$$\frac{1}{2\beta} \left\{ \frac{\alpha}{\alpha - \mathrm{e}^{2\beta r}} - \frac{\alpha}{\alpha - 1} + \log\Big(\frac{\mathrm{e}^{2\beta r} - \alpha}{1 - \alpha} \Big) \right\}$$

if $k = 4$, and

$$\frac{1}{\beta}\left\{\frac{\alpha + \mathrm{e}^{-\beta r}(\mathrm{e}^{2\beta r}-\alpha)^{3/2}\log(\sqrt{\mathrm{e}^{2\beta r}-\alpha}+\mathrm{e}^{\beta r}) - \frac{4}{3}\mathrm{e}^{2\beta r}}{\sqrt{1-\alpha \mathrm{e}^{-2\beta r}}(\mathrm{e}^{2\beta r}-\alpha)} \right.$$
$$\left. - \log(\sqrt{1-\alpha}+1) + \frac{\frac{4}{3}-\alpha}{(1-\alpha)^{3/2}}\right\}$$

if $k = 5$.

B Simulation study concerning estimation procedure

B.1 Second order composite likelihood

In Section 4.2 we fitted the parameters of the Cox process models using a two step procedure involving a minimum contrast procedure for estimating σ^2 and β. Another option is to consider a second order composite likelihood approach (adapted from Waagepetersen, 2007, to a point pattern on a linear network). That is, for an observed point pattern $x \subset L$, the maximum composite likelihood estimate is obtained by maximizing the log composite likelihood

$$CL(\sigma^2, \beta)$$
$$= \sum_{u,v \in x}^{\neq} w(u,v)\log\{\hat{\rho}(u)\hat{\rho}(v)g(u,v)\} - \int_L \int_L w(u,v)\hat{\rho}(u)\hat{\rho}(v)g(u,v)\mathrm{d}_L u\, \mathrm{d}_L v,$$

where w is a weight function and $\neq$ over the summation sign means that $u \neq v$. We can for example let

$$w(u,v) = \mathbb{I}(d_L(u,v) \leq r_0) \tag{D.15}$$

for some user specified value r_0. To estimate (σ^2, β), we either directly maximise B.1 or alternatively solve the associated estimating equation obtained by setting the score equal to 0. The score function is in our set-up given by

$$\nabla CL(\sigma^2,\beta) = \tag{D.16}$$
$$\sum_{u,v \in x}^{\neq} w(u,v)\frac{\nabla g(u,v)}{g(u,v)} - \int_L \int_L w(u,v)\hat{\rho}(u)\hat{\rho}(v)\nabla g(u,v)\, \mathrm{d}_L u\, \mathrm{d}_L v.$$

To improve the composite likelihood estimation procedure, Lavancier et al. (2018) suggested an adaptive version of D.16 where the weight function w

depends on the model parameters. We may for example let w be the indicator function given by

$$w(u,v) = \mathbb{I}\left(\frac{|g(u,v) - 1|}{M(u,v)} > \epsilon \right), \tag{D.17}$$

where $M(u,v) = \max_{s \in \{u,v\}} |g(s,s) - 1|$ and $\epsilon \in (0,1)$ is a small user-specified number, e.g. $\epsilon = 0.01$ or $\epsilon = 0.05$. Note that for an isotropic correlation function $g(u,v) = g_0\{d_L(u,v)\}$, we have $M(u,v) = |g_0(0) - 1|$. Another weight function suggested in Lavancier et al. (2018) is

$$w(u,v) = \begin{cases} \exp\left[1/\{h(u,v)^2 - 1\}\right] & \text{for } -1 \leq h(u,v) \leq 1, \\ 0 & \text{else,} \end{cases} \tag{D.18}$$

where $h(u,v) = \epsilon M(u,v)/\{g(u,v) - 1\}$.

For approximating the double integral in D.16 (or in the adapted version), note that this is of the form $\int_L \int_L f(u,v)\, \mathrm{d}_L u\, \mathrm{d}_L v$. We split up the integration area into the line segments constituting L, that is,

$$\int_L \int_L f(u,v)\, \mathrm{d}_L u\, \mathrm{d}_L v = \sum_{i,j} \int_{L_i} \int_{L_j} f(u,v)\, \mathrm{d}_L u\, \mathrm{d}_L v.$$

Note that $\hat{\rho}(\cdot)$ for the spine data is constant on any line segment L_i (as the line segment is either fully contained in L_m or L_s). Further, if L is a tree, w is given by D.15, D.17, or D.18, and g is isotropic, then $f(u,v) = f_0\{d_L(u,v)\}$ depends only on distance; this will ease the approximation of the integral:

$$\int_{L_i} \int_{L_j} f(u,v)\, \mathrm{d}_L u\, \mathrm{d}_L v = \begin{cases} \int_0^{|L_i|} \int_0^{|L_j|} f_0(d_{i,j} + x + y)\, \mathrm{d}x\, \mathrm{d}y & \text{if } i \neq j, \\ \int_0^{|L_i|} \int_0^{|L_j|} f_0(|x - y|)\, \mathrm{d}x\, \mathrm{d}y & \text{if } i = j, \end{cases}$$

where $d_{i,j} = \min_{u \in L_i, v \in L_j} d_L(u,v)$. Each of these integrals can then be approximated by Monte Carlo integration using uniform variables on $[0, |L_i|]$ and $[0, |L_j|]$.

B.2 Simulation study

In the following we describe and summarise results from a simulation study investigating how well σ^2 and β are estimated using either the minimum contrast procedure with $T = K$ or $T = g_0$ (as described in Section 4.2) or the adaptive composite likelihood procedure using D.17 or D.18. We considered the network for dendrite 4 and simulated from the Cox process described in Section 4.2 with different parameter values given in Table D.3 and when $k = 1$ was fixed.

Run no.	σ^2	β	$\rho_{Y,m}$	$\rho_{Y,s}$	p for MCE-g (-K)	r_l	r_u	(σ^2_*, β_*)
1	5	0.1	0.8	1.2	1 (0.25)	0	30	(0.5, 0.5)
2	5	0.1	0.8	1.2	0.5 (0.5)	0	30	(0.5, 0.5)
3	5	0.1	0.8	1.2	1 (0.25)	0	50	(0.5, 0.5)
4	5	0.1	0.8	1.2	1 (0.25)	0	20	(0.5, 0.5)
5	5	0.1	0.8	1.2	1 (0.25)	0	30	(3, 0.2)
6	5	0.1	0.8	1.2	1 (0.25)	0	30	(0.2, 3)
7	5	0.1	0.3	0.7	1 (0.25)	0	30	(0.5, 0.5)
8	5	0.1	1	1	1 (0.25)	0	30	(0.5, 0.5)
9	5	0.5	0.8	1.2	1 (0.25)	0	30	(0.5, 0.5)
10	5	1	0.8	1.2	1 (0.25)	0	30	(0.5, 0.5)
11	1	0.1	0.8	1.2	1 (0.25)	0	30	(0.5, 0.5)
12	1	0.5	0.8	1.2	1 (0.25)	0	30	(0.5, 0.5)
13	5	0.1	0.8	1.2	1 (0.25)	$2 * bw$	30	(0.5, 0.5)
14	5	0.1	0.8	1.2	1 (0.25)	$0.5 * bw$	30	(0.5, 0.5)
15	5	0.1	0.8	1.2	1 (0.25)	2	30	(0.5, 0.5)
16	5	0.1	0.8	1.2	1 (0.25)	0.5	30	(0.5, 0.5)

Table D.3: Overview of runs made in the simulation study for the minimum contrast procedures. Here (σ^2_*, β_*) denote the start parameters for the optimisation algorithm and bw is the automatically selected bandwidth used to calculate $\hat{g}$ in `spatstat`.

For the minimum contrast procedure we investigated different values of the tuning parameters r_u, r_l, and p as well as different start parameters for the optimisation algorithm, see Table D.3. Here run no. 1 is the reference run from which one (or two) model or tuning parameters are changed at the time. For run no. 1, we chose $\sigma^2 = 5$ and $\beta = 0.1$ resulting in a model rather far away from the case of a Poisson process. Note that decreasing β will increase the range of correlation in the thinning probability, whilst increasing σ^2 will increase the probability of thinning. Thus a small β and a large σ^2 yield a model very different from the Poisson process. For each choice of model parameters we simulated 500 point patterns and estimated (σ^2, β) using minimum contrast and for a few selected runs we also estimated (σ^2, β) using the adaptive composite likelihood method.

For the adaptive composite likelihood method the integral in D.16 was approximated using 10^6 simulations. Estimates of (σ^2, β) were found by minimising the length of the score over a 100×100 grid centred around the true values of σ^2 and β. The finer and broader grid, the better, but as a 100×100 grid was already quite time consuming we settled with that.

Results using the minimum contrast procedure

In the following we let MCE-K and MCE-g refer to minimum contrast estimation with $T = K$ and $T = g_0$, respectively.

Histograms of the obtained estimates are shown in Figures D.4–D.7. In general it seems that MCE-g performs better than MCE-K. For parameter values that result in models close to the Poisson model, neither of the estimation procedures estimate (σ^2, β) successfully. For simulations more distinguishable from Poisson (as for example run no. 1), the MCE-g gives more satisfactory estimates.

Neither MCE-g or MCE-K seem to be sensitive to the choice of start parameters for the reference run. Further, for the reference run it seem that $r_u = 20$ was the best choice, but in general this depend on the model we are trying to fit and naturally on the size of the network. The best choice of r_l seem to be $r_l = 0$, while $p = 0.25$ seems preferable over $p = 0.5$ for MCE-K, and $p = 0.5$ and $p = 1$ perform equally well for the MCE-g. The choice of r_u seem to be important with respect to bias and variance: a too high r_u may lead to a large bias, while a too smale r_u may lead to a large variance in the estimates. It is therefore recommendable to perform a small simulation study for the specific network and proposed model at hand, such that the best choice of r_u can be made.

Results using the adaptive composite likelihood

For the CLE procedure we restricted ourselves to a small number of runs as the grid search was very time consuming. Specifically, we simulated 500 point patterns from the Cox process models with the parameters from run no. 1, 10, and 11 in Table D.3. For each choice of model parameters we estimated (σ^2, β) using both weight functions and $\epsilon = 0.05$. Further, we also ran the CLE procedure with $\epsilon = 0.01$ for the model parameters from run no. 1. Figure D.8 shows histograms of the resulting estimates. It is clear that for all three choices of model parameters the grid should be broader in order to find the parameter values that give the smallest length of D.16 and that the estimates are worse than the ones obtained by the minimum contrast procedures. Finally, there is no seemingly advantage of choosing one weight function over the other or of choosing $\epsilon = 0.01$ over $\epsilon = 0.05$.

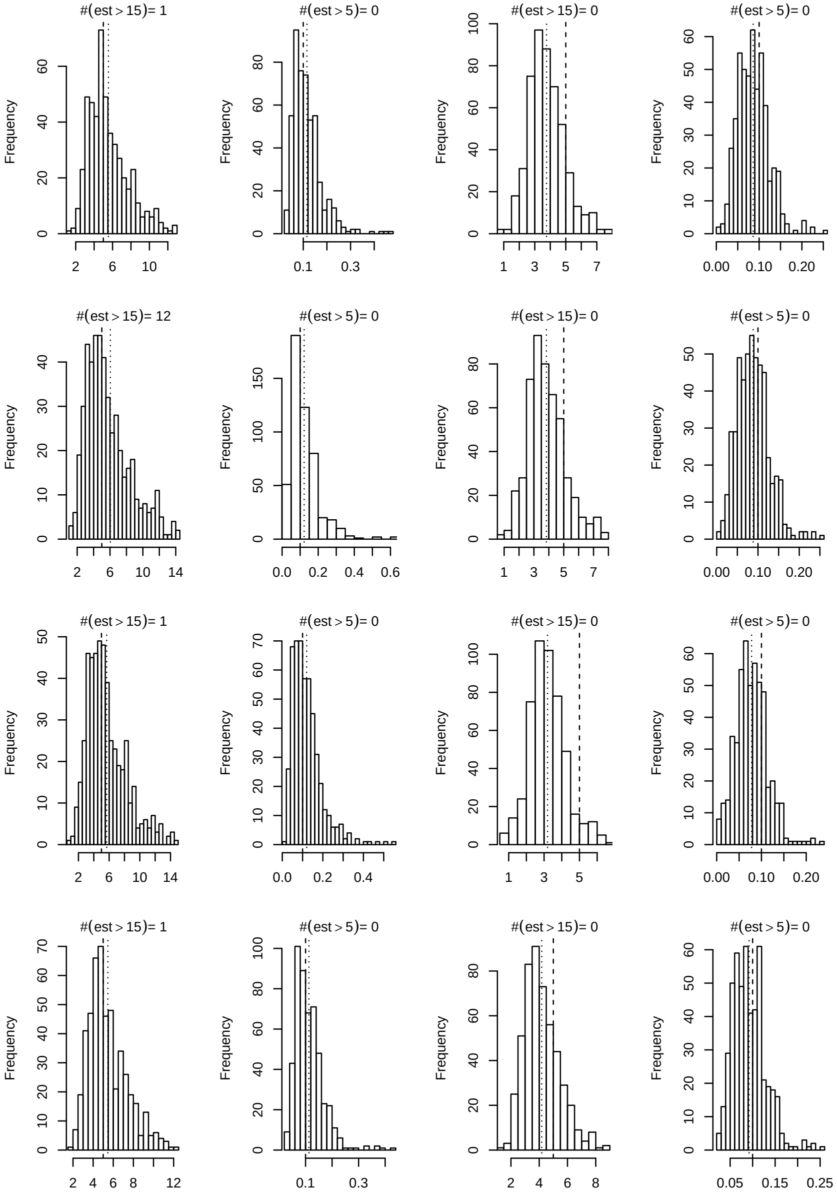


Figure D.4: Estimates of σ^2 and β using either MCE-g or MCE-K for 500 simulated point patterns of models with parameters no. 1–4 in Table D.3 (one set of parameters for each row, starting with no. 1 in the top). From left to right: estimates of σ^2 and β found by MCE-K (column 1 and 2), followed by estimates of σ^2 and β based on MCE-g (column 3 and 4). Dashed line is the true parameter value, and dotted line is the mean of the estimates. OBS: the histograms have been truncated such that estimates above 15 for σ^2 (column 1 and 3) and 5 for β (column 2 and 4) have been omitted in the frequency count; in each histogram it is stated how many values were discarded.

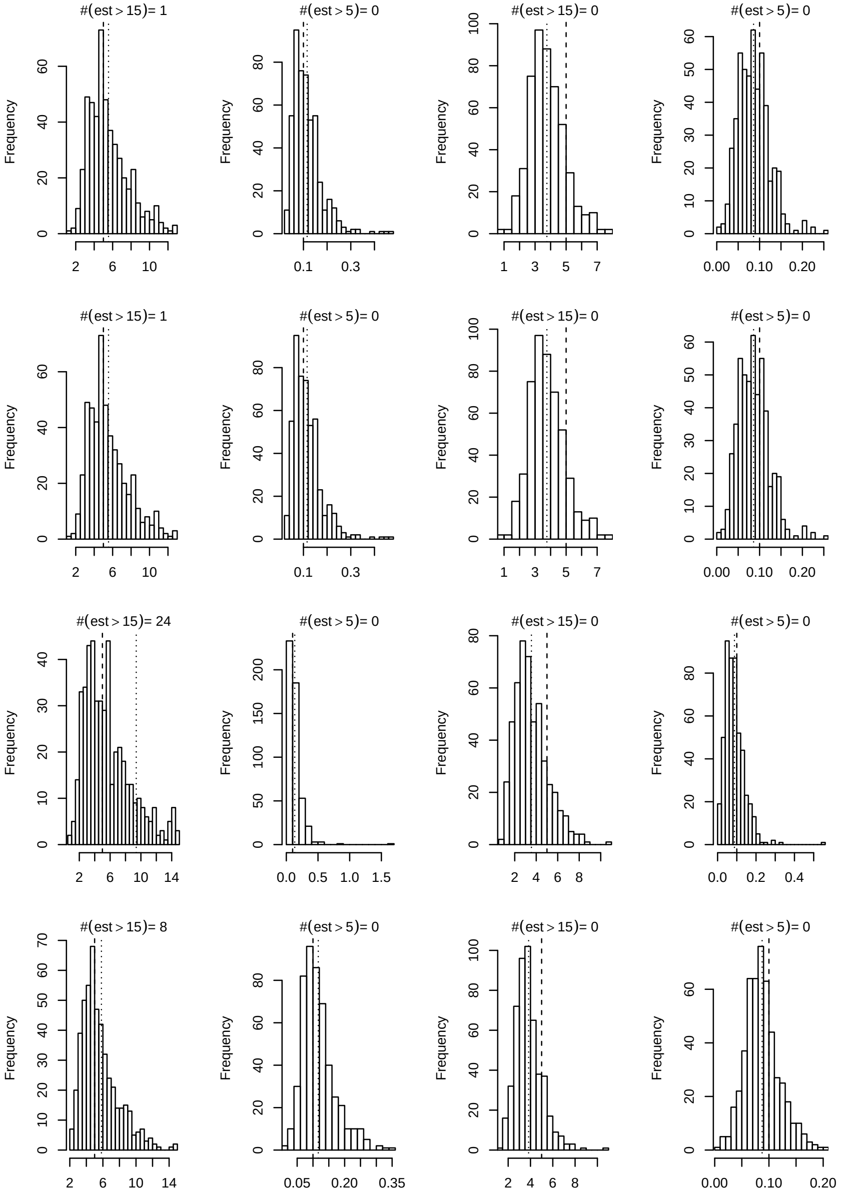


Figure D.5: Estimates of σ^2 and β using either MCE-g or MCE-K for 500 simulated point patterns of models with parameters no. 5–8 in Table D.3 (one set of parameters for each row, starting with no. 5 in the top). From left to right: estimates of σ^2 and β found by MCE-K (column 1 and 2), followed by estimates of σ^2 and β based on MCE-g (column 3 and 4). Dashed line is the true parameter value, and dotted line is the mean of the estimates. OBS: the histograms have been truncated such that estimates above 15 for σ^2 (column 1 and 3) and 5 for β (column 2 and 4) have been omitted in the frequency count; in each histogram it is stated how many values were discarded.

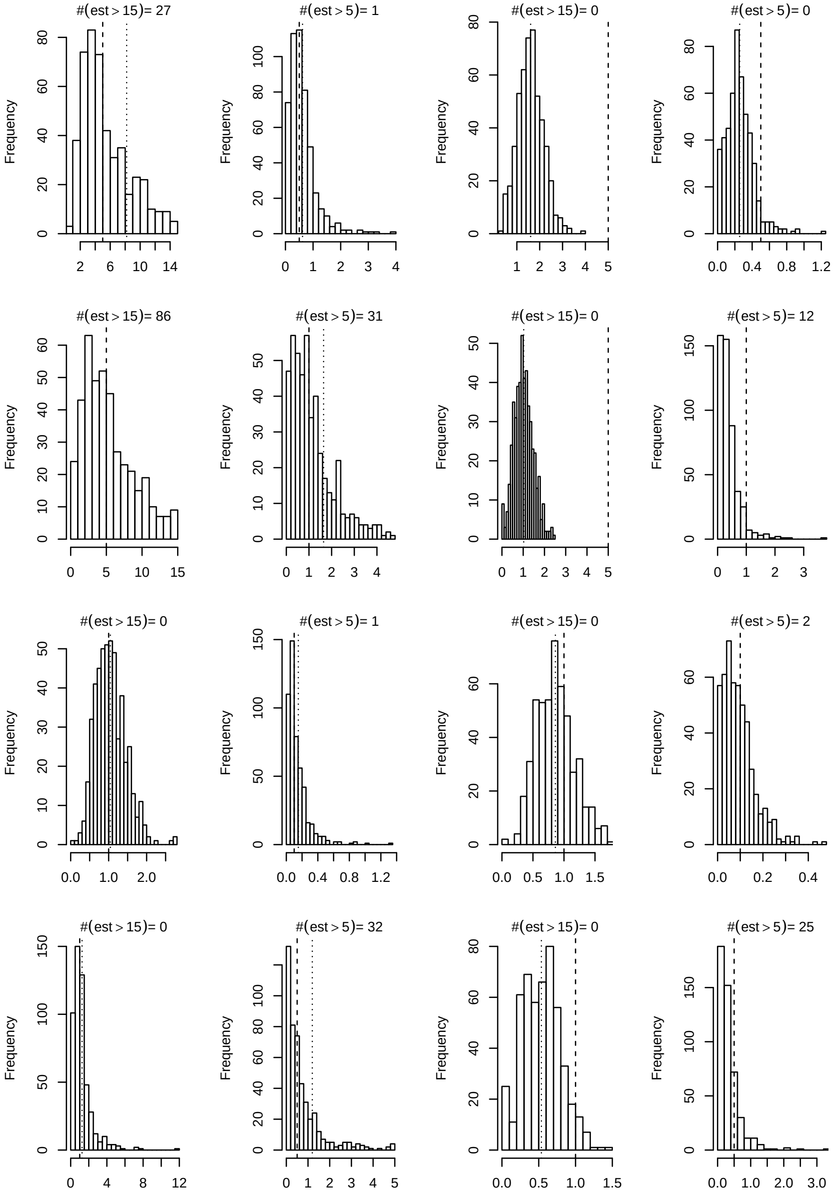


Figure D.6: Estimates of σ^2 and β using either MCE-g or MCE-K for 500 simulated point patterns of models with parameters no. 9–12 in Table D.3 (one set of parameters for each row, starting with no. 9 in the top). From left to right: estimates of σ^2 and β found by MCE-K (column 1 and 2), followed by estimates of σ^2 and β based on MCE-g (column 3 and 4). Dashed line is the true parameter value, and dotted line is the mean of the estimates. OBS: the histograms have been truncated such that estimates above 15 for σ^2 (column 1 and 3) and 5 for β (column 2 and 4) have been omitted in the frequency count; in each histogram it is stated how many values were discarded.

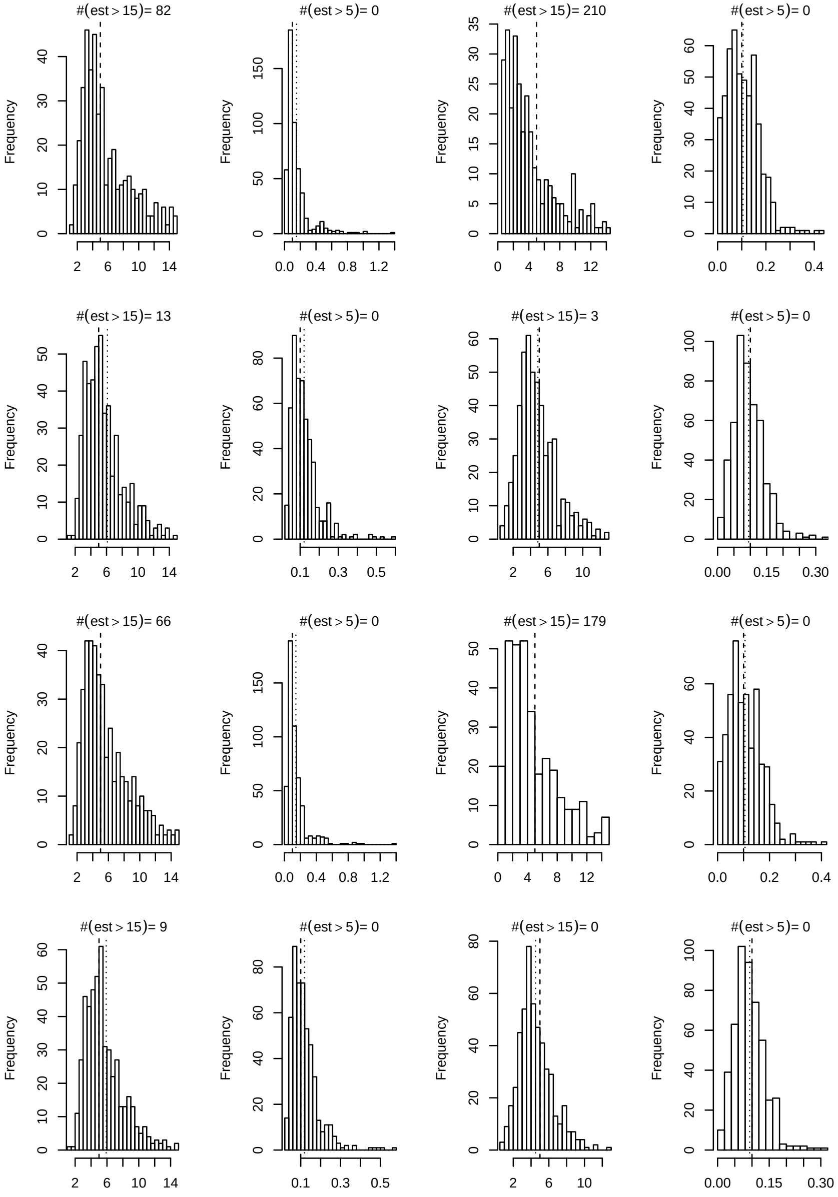


Figure D.7: Estimates of σ^2 and β using either MCE-g or MCE-K for 500 simulated point patterns of models with parameters no. 13–16 in Table D.3 (one set of parameters for each row, starting with no. 13 in the top). From left to right: estimates of σ^2 and β found by MCE-K (column 1 and 2), followed by estimates of σ^2 and β based on MCE-g (column 3 and 4). Dashed line is the true parameter value, and dotted line is the mean of the estimates. OBS: the histograms have been truncated such that estimates above 15 for σ^2 (column 1 and 3) and 5 for β (column 2 and 4) have been omitted in the frequency count ; in each histogram it is stated how many values were discarded.

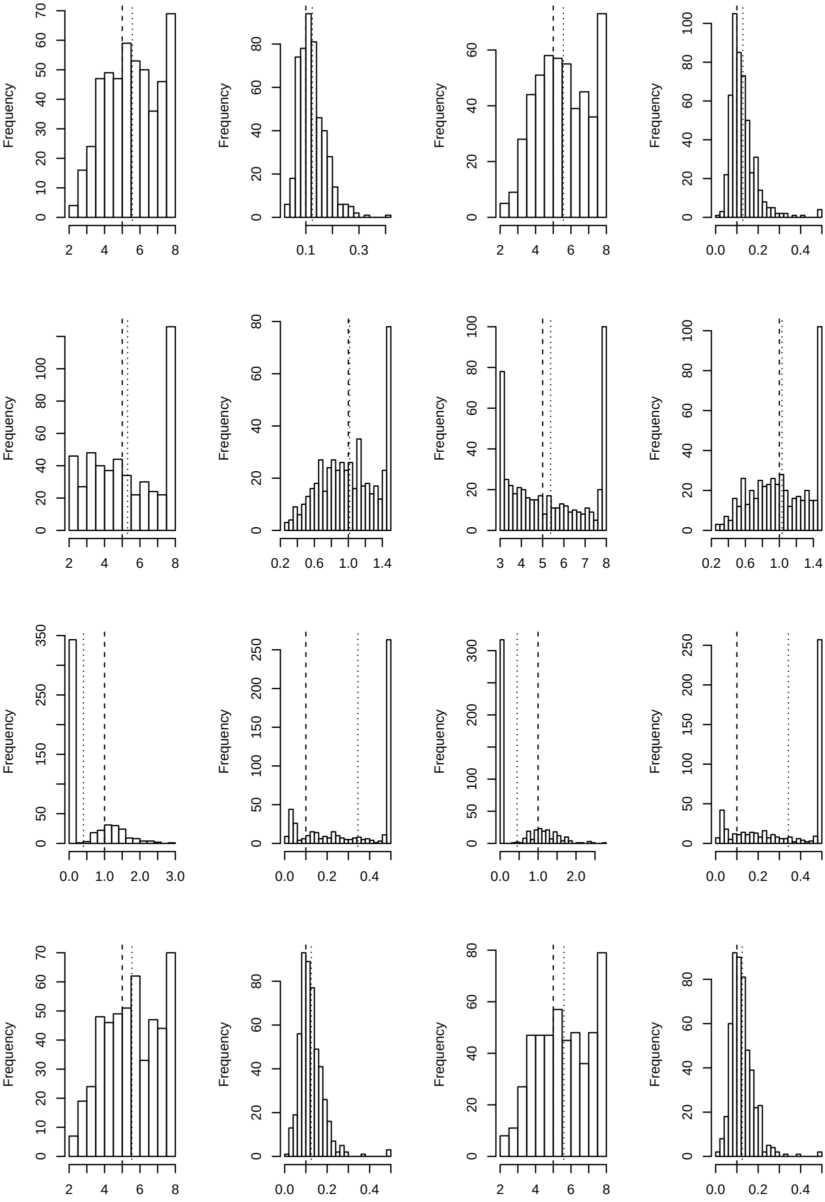


Figure D.8: Results from simulation study using adaptive composite likelihood: estimates of β and σ^2 for model parameters from run no. 1 (first row), 10 (second row), and 11 (third row) with $\epsilon = 0.05$ and run no. 1 with $\epsilon = 0.01$ (fourth row); see Table D.3. The two first columns display estimates of σ^2 and β (in that order) when using the indicator weight function for the CLE procedure, while estimates in the two right columns are found using CLE with the exponential weight function. Dashed line is the true parameter value, dotted line is the mean of the estimates.

C Analysis of spine locations

This Appendix contains figures related to the analysis of the six spine data sets. Figure D.9 shows 95% global rank envelopes under the fitted inhomogeneous Poisson model using a concatenation of $\hat{F}$, $\hat{G}$, and $\hat{J}$ as test function. Further, Figure D.10 show one simulation of the fitted random field Π for each network. Finally, each of Figures D.11–D.16 display the data along with five simulated point patterns from the fitted Cox process model.

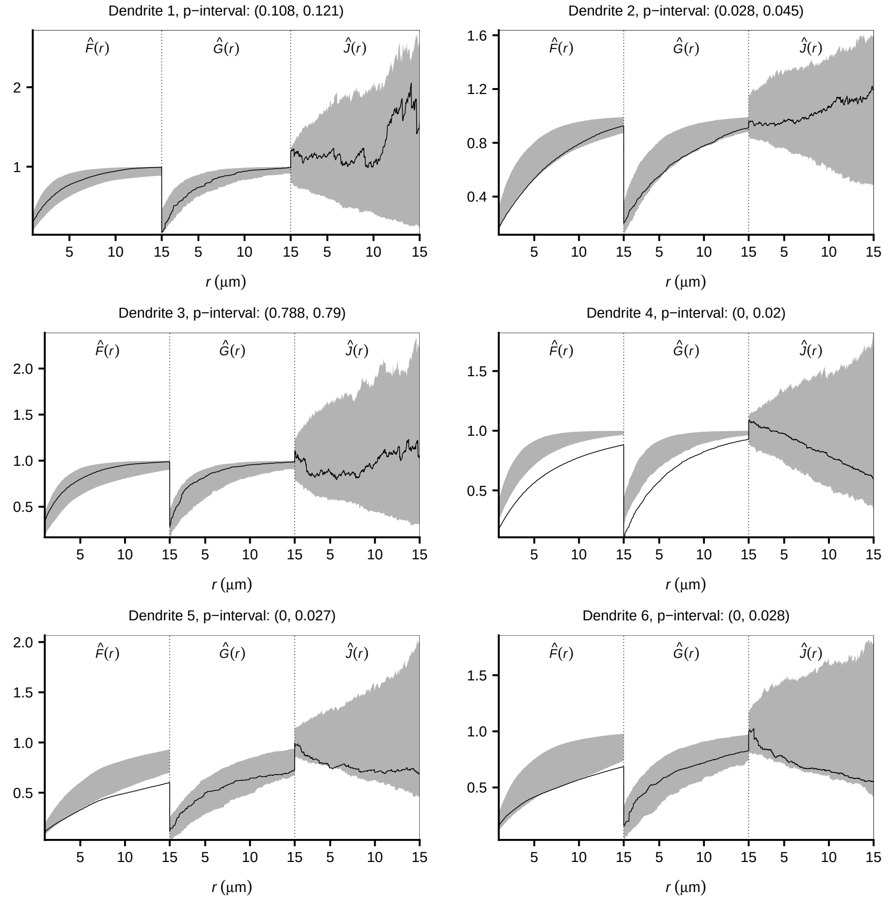


Figure D.9: For each spine data set: the concatenation of $\hat{F}$, $\hat{G}$, and $\hat{J}$ for the spine locations (black solid line) along with 95% global rank envelopes (grey region) based on 2499 simulations from the fitted inhomogeneous Poisson model; p-intervals for each of the associated global rank envelope tests are also displayed.

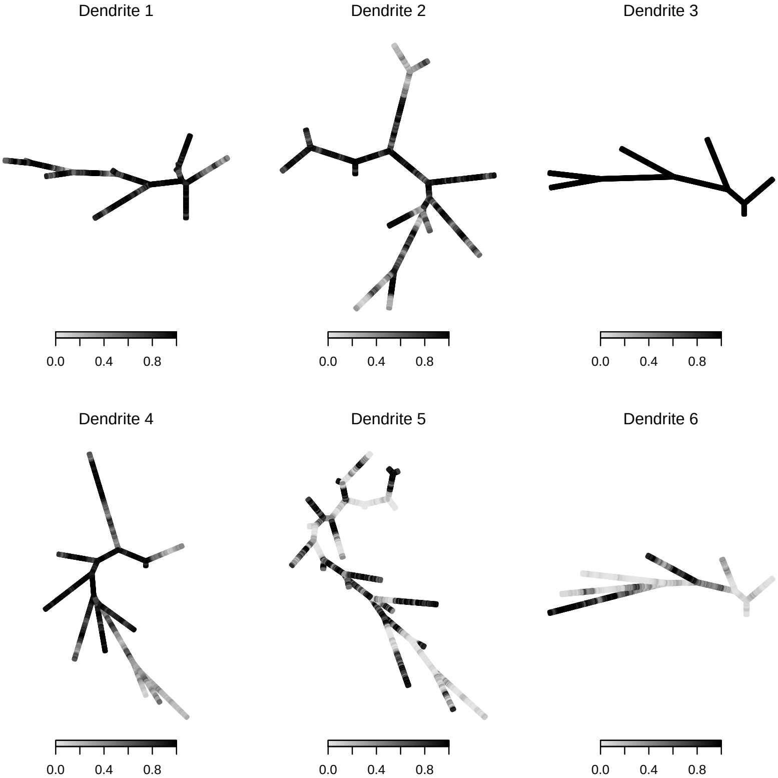


Figure D.10: For each dendrite tree, a simulated realisation of the random field Π determining the retention probabilities in the fitted Cox process models.

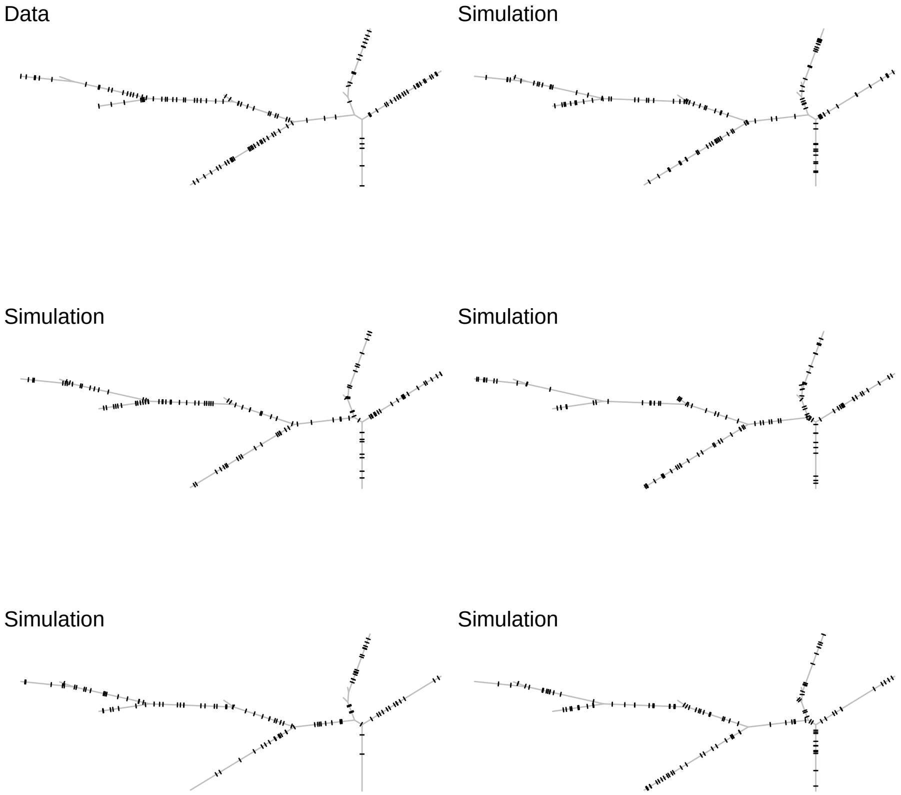


Figure D.11: Upper left corner: observed spine locations on dendrite 1. Remaining: simulated point patterns from the fitted Cox process model. The simulated retention probabilities used to obtain the point pattern in the upper right corner are shown in Figure D.10.

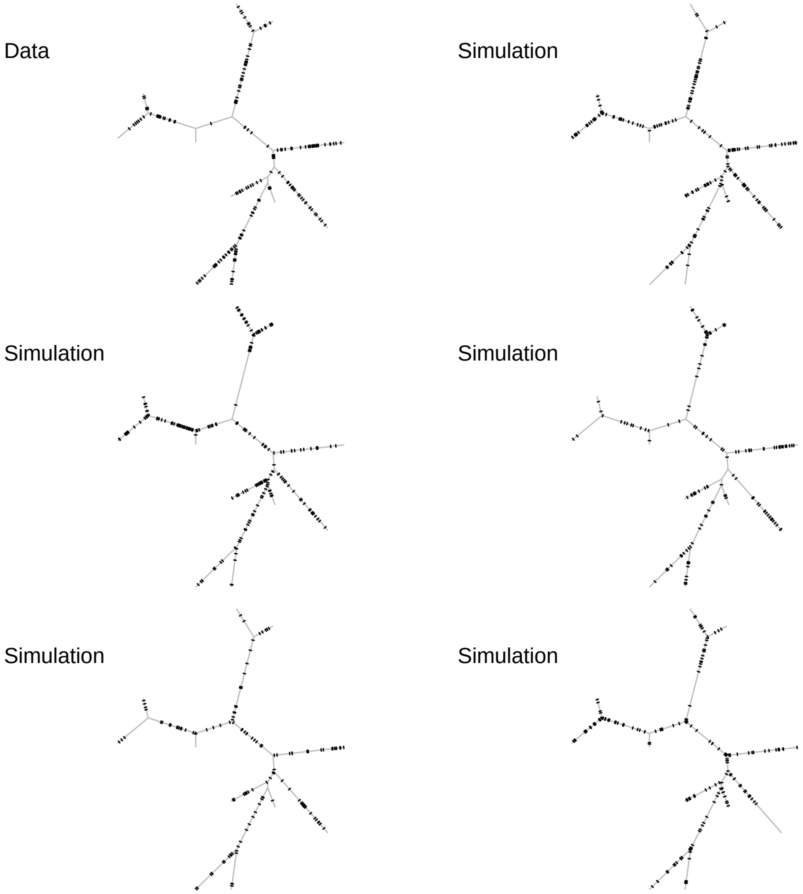
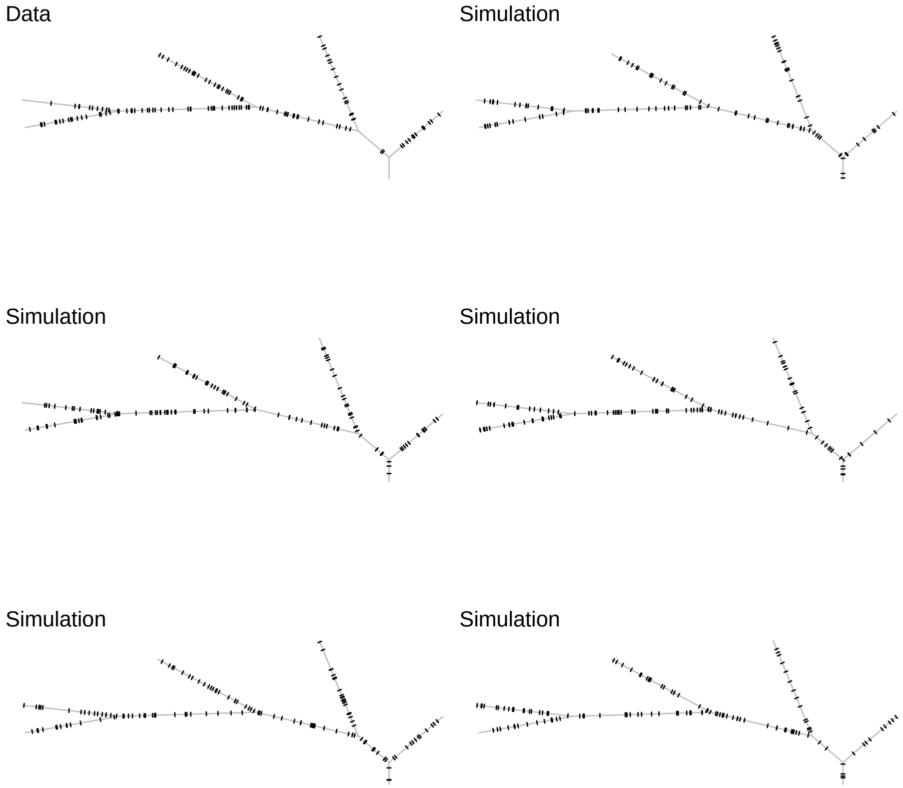


Figure D.12: Upper left corner: observed spine locations on dendrite 2. Remaining: simulated point patterns from fitted Cox process model. The simulated retention probabilities used to obtain the point pattern in the upper right corner are shown in Figure D.10.



Data

Simulation

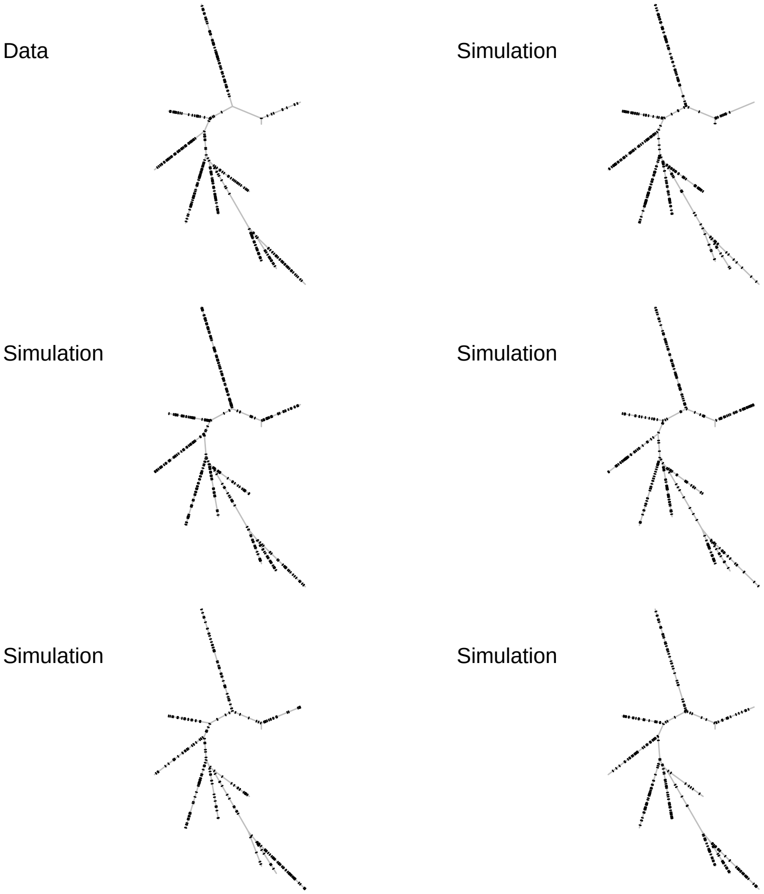
Simulation

Simulation

Simulation

Simulation

Figure D.13: Upper left corner: observed spine locations on dendrite 3. Remaining: simulated point patterns from fitted Cox process model. The simulated retention probabilities used to obtain the point pattern in the upper right corner are shown in Figure D.10.



Data

Simulation

Simulation

Simulation

Simulation

Simulation

Figure D.14: Upper left corner: observed spine locations on dendrite 4. Remaining: simulated point patterns from fitted Cox process model. The simulated retention probabilities used to obtain the point pattern in the upper right corner are shown in Figure D.10.

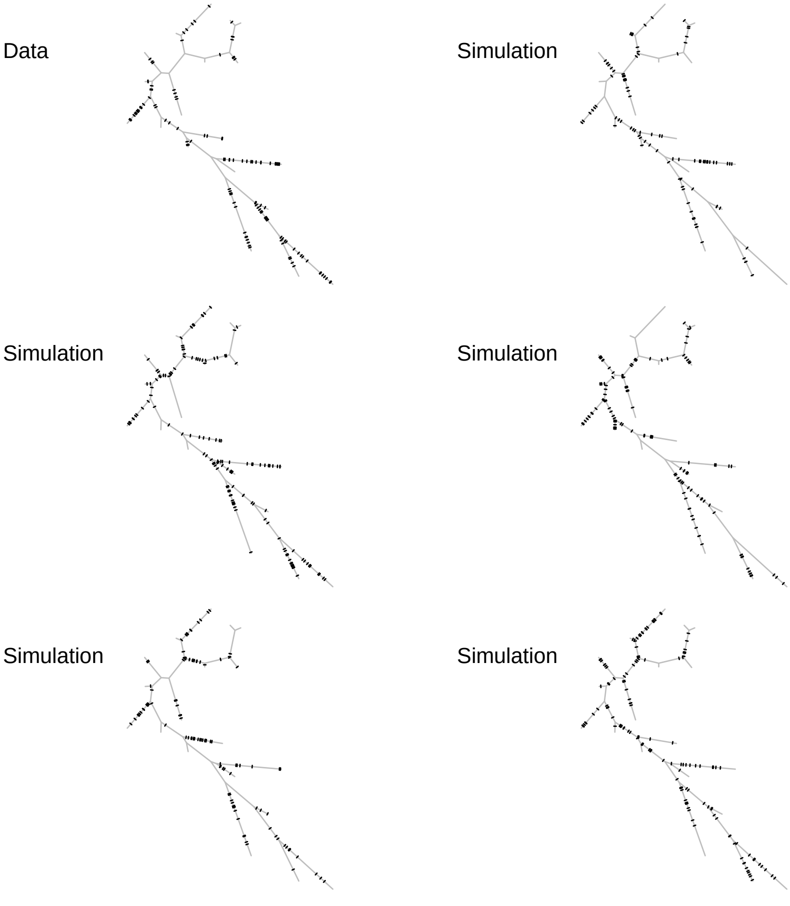


Figure D.15: Upper left corner: observed spine locations on dendrite 5. Remaining: simulated point patterns from fitted Cox process model. The simulated retention probabilities used to obtain the point pattern in the upper right corner are shown in Figure D.10.

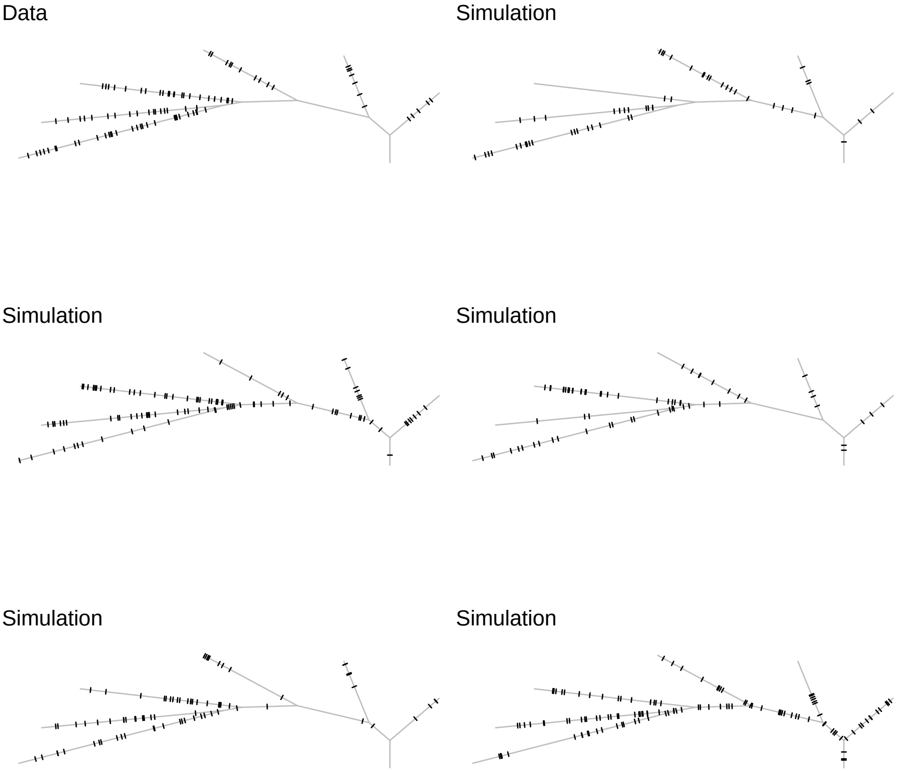

Data
Simulation
Simulation
Simulation
Simulation
Simulation

Figure D.16: Upper left corner: observed spine locations on dendrite 6. Remaining: simulated point patterns from fitted Cox process model. The simulated retention probabilities used to obtain the point pattern in the upper right corner are shown in Figure D.10.

References

Anderes, E., Møller, J., and Rasmussen, J. G. (2017). Isotropic covariance functions on graphs and their edges. Available at arXiv:1710.01295.

Ang, Q. W., Baddeley, A. J., and Nair, G. (2012). Geometrically corrected second order analysis of events on a linear network, with applications to ecology and criminology. *Scandinavian Journal of Statistics*, 39:591–617.

Baddeley, A., Rubak, E., and Turner, R. (2016). *Spatial Point Patterns: Methodology and Applications with R*. Chapman & Hall/CRC, New York.

Baddeley, A. J., Jammalamadaka, A., and Nair, G. (2014). Multitype point process analysis of spines on the dendrite network of a neuron. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 63:673–694.

Baddeley, A. J., Møller, J., and Waagepetersen, R. P. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54:329–350.

Baddeley, A. J., Nair, G., Rakshit, S., and McSwiggan, G. (2017). "Stationary" point processes are uncommon on linear networks. *Stat*, 6:68–78.

Diggle, P. J. (2014). *Statistical Analysis of Spatial and Spatio-temporal Point Patterns*. Chapman & Hall/CRC Press, Boca Raton, Florida.

Guan, Y. (2009). A minimum contrast estimation procedure for estimating the second-order parameters of inhomogeneous spatial point processes. *Statistics and Its Interface*, 2:91–99.

Jammalamadaka, A., Banerjee, S., Manjunath, B. S., and Kosik, K. S. (2013). Statistical analysis of dendritic spine distributions in rat hippocampal cultures. *BMC Bioinformatics*, 14:287.

Lavancier, F. and Møller, J. (2016). Modelling aggregation on the large scale and regularity on the small scale in spatial point pattern datasets. *Scandinavian Journal of Statistics*, 43:587–609.

Lavancier, F., Poinas, A., and Waagepetersen, R. P. (2018). Adaptive estimating function inference for non-stationary determinantal point processes. Available on arXiv:1806.06231.

van Lieshout, M. N. M. (2011). A J-function for inhomogeneous point processes. *Statistica Neerlandica*, 65:183–201.

Matérn, B. (1960). Spatial variation: Stochastic models and their application to some problems in forest surveys and other sampling investigations. *Meddelanden från Statens Skogforskningsinstitut*, 49:1–144.

Matérn, B. (1986). *Spatial Variation*. Lecture Notes in Statistics 36. Springer-Verlag, Berlin.

McSwiggan, G., Baddeley, A. J., and Nair, G. (2016). Kernel density estimation on a linear network. *Scandinavian Journal of Statistics*, 44:324–345.

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

Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H., and Hahn, U. (2017). Global envelope tests for spatial processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79:381–404.

Okabe, A. and Yamada, I. (2001). The K-function method on a network and its computational implementation. *Geographical analysis*, 33:271–290.

Rakshit, S., Nair, G., and Baddeley, A. J. (2017). Second-order analysis of point patterns on a network using any distance metric. *Spatial Statistics*, 22:129–154.

Rasmussen, J. G. and Christensen, H. S. (2019). Point processes on directed linear networks. Available at arXiv:1812.09071.

Waagepetersen, R. P. (2007). An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics*, 63:252–258.

Waagepetersen, R. P. and Guan, Y. (2009). Two-step estimation for inhomogeneous spatial point processes. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 71:685–702.

ISSN (online): 2446-1636
ISBN (online): 978-87-7210-488-1

AALBORG UNIVERSITY PRESS