UNIVERSITAT
JAUME • **I**

Doctoral Programme in Science
Doctoral School of the University Jaume I

Statistical tests for comparisons of spatial and spatio-temporal point patterns

This dissertation is submitted by **Jonatan A. González** for the
degree of Doctor of Philosophy at University Jaume I

Jonatan A. González

Prof. Dr. Jorge Mateu
Prof. Dr. Ute Hahn

Castellón de la Plana, January 2018

Acknowledgements

Institutional Acknowledgements

Personal Acknowledgements

Declaration

The director and the main author hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Castellón, January 2018.

Jorge Mateu
Director

Jonatan González
Author

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Point processes constitute a powerful tool for the analysis of sets of randomly located objects in a plane, in the 3D-space, or in general in an abstract space. These objects can be defined rigorously as a locally finite random counting measure on a locally compact completely separable Hausdorff space (Daley and Vere-Jones, 2003, 2007). The applications of this theory are widely diverse and include a considerable variety of situations that come from reality. These applications include astronomy (Neyman and Scott, 1958), biology (Diggle, 2013), ecology (Thorsten Wiegand, 2014), economic sciences (Bowsher, 2007), epidemiology (Diggle, Guan, Hart, Paize and Stanton, 2010), materials engineering (Ohser and Mücklich, 2000), medicine (Andersen and Hahn, 2016; Diggle et al., 1991; Hahn, 2012), seismology (Tranbarger and Schoenberg, 2010), amongst many others (see e.g. Illian et al., 2008 for a nice treatment of theory and practice). There is also an extensive bibliography covering a broad spectrum, both theoretical (Chiu et al., 2013; Daley and Vere-Jones, 2003, 2007; Møller and Waagepetersen, 2004) and practical. The practical aspect also covers from the resolution of problems that come from the reality and increasingly complex and demanding datasets (Diggle, 2013), to the implementation of several methodologies in user-friendly, free and affordable software (Baddeley et al., 2015; Gabriel et al., 2013).

One of the assumptions that is usually made in working with point processes is that there is a single pattern observed in reality, i.e. a single set of points located in an observation window. The distribution of this pattern is mathematically complex, this has led to propose in the literature some *summary-statistics* that can show or describe the nature of the distribution. Normally, these statistics are functions or real numbers that are based on distances between the points within the pattern or in counts. For instance, the first-order intensity describes the dens-

ity of the points of a point pattern. The well-known second-order characteristics (Chiu et al., 2013; Illian et al., 2008), are extensively used by scientists since these characteristics describe the attraction or repulsion that may exist between the points of the point pattern. In general these functions or descriptors can be non-parametrically estimated and some of the most famous are Ripley's *K*-function (Ripley, 1977) and its standardised version, the *L*-function (Besag and Diggle, 1977); the *J*-function (van Lieshout and Baddeley, 1996), the pair-correlation function, etc.

Some datasets of point patterns contain more than one observed pattern, and in this case, answering the questions that usually are associated with data is not easy and only few authors have worked on the subject. Baddeley et al. (1987, 1993) presented an approach using ratio-regressions in three-dimensional replicated point patterns by using three-dimensional *K*-functions. Diggle et al. (1991) presented an approach on analysis of variance in clinical neuroanatomy through the comparison of estimated *K*-functions and non-parametric inference through a Monte Carlo bootstrap test. This method was later improved by Diggle et al. (2000) and by Hahn (2012) who developed a studentized permutation test for the comparison of several groups of spatial point patterns. Some authors (Bagchi and Illian, 2015; Landau and Everall, 2008; Landau et al., 2004; Myllymäki et al., 2014) have included, for example, categorical and continuous predictors in linear models of mixed effects or discrete marks for the explanation of the variation in the spatial structure of replicated patterns.

The fundamental aim of this thesis is to introduce a new set of tests to compare estimated *K*-functions (or in general any suitable functional descriptor). Two parallel lines are available for the development of the new tests.

In the first place, the studentized permutation test proposed by Hahn (2012) is extended to the spatio-temporal case. Since the study of spatio-temporal point processes has not been as widely covered in the literature as its spatial counterpart, a complete review is made. The review is as a reference paper of the available techniques and approaches regarding the spatio-temporal context. Some of these techniques are applied to interesting datasets (González et al., 2016). The new spatio-temporal permutation test shows an accurate performance in terms of empirical level and power. This test is motivated by a complex dataset: the locations of tornadoes in the U.S. in a period of 36 years. This is why in addition to the test itself, some additional tools have been developed as a non-

separable estimator of the first-order spatio-temporal intensity, which allows a much more realistic analysis of the phenomenon through the new test.

Secondly, an ANOVA two-way design is considered, where the observations are spatial point patterns and where in addition, replicates are available for each of the combinations of the treatments. This methodology is motivated by a materials engineering experiment, where the locations of bubbles in a flotation cell have been measured in a fixed time and the levels of two factors involved in the experiment have been varied. A whole scheme of analysis of variance is developed for factorial experiments, that is, we develop efficient statistics to test, in a non-parametric way, the influence of the factors and the possible interaction effect. For this, several possibilities are proposed to make inference through random permutations and demonstrate that they are useful in practice.

The thesis is organized as follows. First of all, we present some introductory information (Chapters 1 and 2). The particularity of the subsequent chapters is that they are self-contained, since they correspond to articles already published (in the case of Chapter 3), or they are in the process of being published (Chapters 4 and 5). Therefore, they can be addressed independently.

In Chapter 2, we provide a brief summary of the theory of spatial point processes. The most basic concepts are presented at an abstract level, some important typical moment-measures are also introduced. We define theoretically some of the classical models of point processes. The spatial summary statistics are presented and finally, we show some non-parametric estimators of these summary statistics.

In Chapter 3 we start with a brief description of the datasets that are analysed. Subsequently, we present an introduction to spatio-temporal point processes. Then, we introduce some characteristics (descriptors) of first- and second-order for spatio-temporal point processes. A review of the empirical and mechanistic models is made and, finally, we conclude with a general discussion and some future research ideas.

In Chapter 4, we first consider a description of the tornado data in the U.S. We proceed to study of the first-order spatio-temporal intensity function and to analyse the hypothesis of separability in two yearly seasons of tornadoes, indeed cold- and warm- seasons. We consider a spatial partition of the map regions in order to simplify the analysis by considering the elements of the partition as possible replicates of an underlying point process. We check the performance of the spatio-temporal permutation test and we apply it to the dataset. Finally,

some questions and problems are discussed. This chapter motivates interesting lines of research. For instance, developing suitable tests that do not require the exchangeability hypothesis (Hahn, 2012). Additionally, the selection of bandwidth in the spatio-temporal context and the possible inclusion of additional dimensions in the observed point patterns as covariates or marks.

Chapter 5 begins with an introduction to the engineering problem that generates the bubble data. Afterwards, we make a complete description of the database. A small section is presented where the mathematical tools used are described. We give a motivation of the factorial analysis of two factors starting from the simplest case to arrive at the balanced two-way ANOVA design. The necessary statistics for the analysis are proposed, and a whole simulation outline is presented to verify that the statistics work properly. Finally, we apply the procedures and statistics to the bubbles dataset. This chapter also motivates some lines of research such as the extension to more complex designs as nested designs and designs with fixed and mixed effects through the use of Fisher-type statistics. The use of more complex functional descriptors (including marks, for example) in the ANOVA design also constitutes an open field of research. Finally, the definition of the factorial analysis statistics but using complex functional distances between patterns (Mateu et al., 2015) or between functions, could be a challenging field of research.

Chapters 3,4 and 5 are independent articles that are meant to be self-contained. Therefore, they have their own notation, introduction and conclusion sections.

Chapter 2

Basic framework of point processes

2.1 Point process methodology

In order to provide a unified and proper context to develop our methodology, we consider in this chapter a set of definitions and important results that provide subsequently a nice grammatical and mathematical basis. We take Stoyan and Stoyan (1994), Møller and Waagepetersen (2004), Illian et al. (2008), Chiu et al. (2013) and Baddeley et al. (2015) texts as fundamental references, treating deeply and rigorously all the concepts that we need in the rest of the work.

We begin recalling the definition and basic concepts of point processes in general metric spaces. We give the theoretical description of first- and second-order characteristics such as the intensity function, the pair correlation function, K-function and J-function, which is not first- or second-order but it looks at all orders. We also provide the definition of certain types of models for spatial point patterns. In addition, we present an extension of this theory to the case of multivariate point processes (Lotwick and Silverman, 1981). We show some statistical estimators of these first- and second-order characteristics in the spatial context by using non-parametric methods.

2.1.1 Point processes on metric spaces

Let S be a metric space with metric $d\left(\cdot,\cdot\right)$ and $\mathscr{B}$ be the Borel σ-algebra (generated by open sets) in S. Let $\mathscr{B}_0 \subseteq \mathscr{B}$ be the system of all bounded Borel sets. We define

the space of locally finite subsets of S as

$$N_{\mathrm{lf}} = \{x \subseteq S : n(x_B) < +\infty, \ \forall B \in \mathscr{B}_0\},$$

where $x_B = x \cap B$ and $n(y)$ denotes the cardinality of the set y. Elements of N_{lf} are called *locally finite point configurations*. We equip N_{lf} with the σ-algebra

$$\mathscr{N}_{\mathrm{lf}} = \sigma\{\{x \in N_{\mathrm{lf}} : n(x_B) = m\}, m \in \mathbb{N}_0, B \in \mathscr{B}_0\},$$

where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. A point process is defined as a random locally finite point configuration $\{\xi_i\}_{i=1}^n = \{\mathbf{u}_i\}_{i=1}^n \subseteq S$. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be an abstract probability space, a point process X is a measurable mapping

$$X : (\Omega, \mathscr{F}, \mathbb{P}) \longrightarrow (N_{\mathrm{lf}}, \mathscr{N}_{\mathrm{lf}}).$$

The distribution of simple point process is a measure P_X defined on $(N_{\mathrm{lf}}, \mathscr{N}_{\mathrm{lf}})$ by the relation

$$P_X(F) = \mathbb{P}(X \in F) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in F\}), \qquad F \in \mathscr{N}_{\mathrm{lf}}.$$

We say that the point process is finite if $n(X) < +\infty$ almost surely. For a point process X, we will denote the number of points in the set B by $N(B) = n(X_B) = n(X \cap B)$ and refer to the function N as a counting function. Thus X is a point process if and only if $N(B)$ is a random variable for any $B \in \mathscr{B}_0$. By *void probabilities* we understand probabilities $\mathbb{P}(N(B) = 0), B \in \mathscr{B}_0$. Whereby, a point process is uniquely determined by its void probabilities.

A point process X is stationary if its distribution is translation invariant, i.e. $X + \mathbf{u}$ has the same distribution as X for any $\mathbf{u} \in S$. A point process X is isotropic if its distribution is invariant under rotations around the origin, i.e. $\mathbf{r}X = \{\mathbf{r}\mathbf{u} : \mathbf{u} \in X\}$ has the same distribution as X.

The *intensity measure* μ on S is given by

$$\mu(B) = \mathbb{E}[N(B)], \qquad B \in \mathscr{B}.$$

If the intensity measure μ can be written as

$$\mu(B) = \int_B \lambda(\mathbf{u}) \, d\mathbf{u} \qquad B \in \mathscr{B},$$

where λ is a non-negative function, then λ is called the *intensity function*. If λ is constant, then X is said to be *homogeneous* or *first-order stationary* with intensity λ; otherwise X is said to be inhomogeneous. Heuristically, $\lambda(\mathbf{u})\,\mathrm{d}\mathbf{u}$ is the probability for the occurrence of a point in an infinitesimally small ball with centre $\mathbf{u}$ and volume $|\mathrm{d}\mathbf{u}| = \mathrm{d}\mathbf{u}$. For a homogeneous point process, λ is the mean number of points per unit volume and the constant is called *intensity* or *rate*.

If X is a stationary point process with locally finite intensity measure μ, then μ is proportional to the Lebesgue measure. Since the intensity measure of a stationary point process is proportional to the Lebesgue measure, the intensity function is constant and equal to this proportionality constant. It means that every stationary point process is homogeneous.

2.1.2 Moment measures

First- and higher-order moments of the counts $N(B)$ with $B \in \mathscr{B}$, can be expressed by the following measures. For a point process X on S and each $m \in \mathbb{N}$, define the m-th order moment measure $\mu^{(m)}$ on S^m by

$$\mu^{(m)}(D) = \mathbb{E}\left[\sum_{\xi_1,\dots,\xi_m \in X} \mathbf{1}\left[(\xi_1,\dots,\xi_m) \in D\right]\right], \qquad D \subseteq S^m,$$

and the m-th order factorial moment measure $\alpha^{(m)}$ on S^m by

$$\alpha^{(m)}(D) = \mathbb{E}\left[\sum_{\xi_1,\dots,\xi_m \in X}^{\neq} \mathbf{1}\left[(\xi_1,\dots,\xi_m) \in D\right]\right], \qquad D \subseteq S^m,$$

where the $\sum^{\neq}$ over the summation sign means that the m points $\xi_1,\dots,\xi_m$ are pairwise distinct and $\mathbf{1}[\cdot]$ denotes the indicator function.

In particular, $\mu = \mu^1 = \alpha^1$ is called the *intensity measure*. The m-th order moment measure $\mu^{(m)}$ determines the m-th order moments of the count variables $N(B)$, $B \subseteq S$, since

$$\mu^{(m)}(B_1 \times \cdots \times B_m) = \mathbb{E}\left[\prod_{i=1}^{m} N(B_i)\right], \qquad B_i \in \mathscr{B}.$$

For any $m \in \mathbb{N}$, there is a one-to-one correspondence between $(\mu^{(1)},\dots,\mu^{(m)})$ and $(\alpha^{(1)},\dots,\alpha^{(m)})$. It is often more convenient to work with the reduced moment

measures. The above definition immediately extends to

$$\mathbb{E}\left[\sum_{\xi_1,\dots,\xi_m \in X}^{\neq} h(\xi_1,\dots,\xi_m)\right] = \int_S \cdots \int_S h(\xi_1,\dots,\xi_m)\,\mathrm{d}\alpha^{(m)}(\xi_1,\dots,\xi_m),$$

for non-negative functions h.

The second-order reduced moment measure

Let $\mathscr{B}^d$ be the Borel σ-algebra (generated by open sets) in $\mathbb{R}^d$. If the second-order factorial moment measure $\alpha^{(2)}$ can be written as

$$\alpha^{(2)}(C) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbf{1}\left[(\xi,\eta) \in C\right]\lambda^{(2)}(\xi,\eta)\,\mathrm{d}\xi\mathrm{d}\eta, \qquad C \subseteq \mathbb{R}^d \times \mathbb{R}^d,$$

where $\lambda^{(2)}$ is a non-negative function, then $\lambda^{(2)}$ is called the *second-order product density function*. If both λ and $\lambda^{(2)}$ exist, the pair correlation function is defined by

$$g(\xi,\eta) = \frac{\lambda^{(2)}(\xi,\eta)}{\lambda(\xi)\lambda(\eta)},$$

where we take $a/0 = 0$ for $a \geq 0$. Suppose that X has intensity function λ and that the measure

$$\mathscr{K}(B) = \frac{1}{|A|}\mathbb{E}\left[\sum_{\xi,\eta \in X}^{\neq} \frac{\mathbf{1}\left[\xi \in A, \eta - \xi \in B\right]}{\lambda(\xi)\lambda(\eta)}\right], \qquad B \subseteq \mathbb{R}^d$$

does not depend on the choice of A, wher $A \subseteq \mathbb{R}^d$ and $0 < |A| < +\infty$, furthermore, we take $a/0 = 0$ for $a \geq 0$. Then X is said to be *second-order intensity reweighted stationary* and $\mathscr{K}$ is called the second-order reduced moment measure.

If the pair correlation function exists and is invariant under translations, then we have second-order intensity reweighted stationarity and

$$\mathscr{K}(B) = \int_B g(\xi)\mathrm{d}\xi, \qquad B \subseteq \mathbb{R}^d.$$

Campbell measures and Palm distributions

Campbell measures and Palm distributions are very important tools as they play a fundamental role in the definition and calculation of characteristics of point processes or their estimators. We present the main results related to these

important concepts. Assume that μ is σ-finite, i.e. $\mu(B_i) < +\infty$ for a countable partition B_i of S (this is e.g. satisfied if μ is locally finite). The Campbell measure is defined by

$$C(D) = \mathbb{E}\left[\sum_{\xi \in X} \mathbf{1}\left[(\xi, X) \in D\right]\right], \qquad D \in \mathcal{B} \times \mathcal{N}_{\mathrm{lf}}.$$

For a point process X on S, define the reduced Campbell measure $C^!$ on $S \times N_{\mathrm{lf}}$ by

$$C^!(D) = \mathbb{E}\left[\sum_{\xi \in X} \mathbf{1}\left[(\xi, X \setminus \{\xi\}) \in D\right]\right], \qquad D \in \mathcal{B} \times \mathcal{N}_{\mathrm{lf}}.$$

We have

$$\mathbb{E}\left[\sum_{\xi \in X} h(\xi, X \setminus \{\xi\})\right] = \int h(\xi, x)\, \mathrm{d}C^!(\xi, x),$$

for non-negative functions h. Note that $C^!$ determines (μ, α^2), since clearly $\mu(\cdot) = C^!(\cdot \times N_{\mathrm{lf}})$ and

$$\alpha^{(2)}(B_1 \times B_2) = \int \mathbf{1}\left[\xi \in B_1\right] n(x_{B_2})\, \mathrm{d}C^!(\xi, x).$$

For each $F \in \mathcal{N}_{\mathrm{lf}}$, $C^!(\cdot \times F) \leq \mu(\cdot)$, so $C^!(\cdot \times F)$ is absolutely continuous $\left(C^!(\cdot \times F) \ll \mu\right)$ with respect to μ. Then by the Radon-Nikodym theorem, there exists a μ-almost surely unique integrable function $P^!_{\xi}$ such that

$$C^!(B \times F) = \int_B P^!_{\xi}(F)\, \mathrm{d}\mu(\xi),$$

where $P^!_{\xi}(\cdot)$ is a probability measure for each $\xi \in S$ (see e.g. Daley and Vere-Jones (2007)). The probability measure $P^!_{\xi}(\cdot)$ on $\mathcal{N}_{\mathrm{lf}}$ is called a *reduced Palm distribution at point* ξ. Based on the above, we obtain the *Campbell-Mecke formula* as

$$\mathbb{E}\left[\sum_{\xi \in X} h(\xi, X \setminus \{\xi\})\right] = \iint h(\xi, x)\, \mathrm{d}P^!_{\xi}(x)\, \mathrm{d}\mu(\xi),$$

for non-negative functions h. If $X \sim Poisson(S, \lambda)$, then for functions $h : S \times N_{\mathrm{lf}} \to [0, \infty)$,

$$\mathbb{E}\left[\sum_{\xi \in X} h(\xi, X \setminus \{\xi\})\right] = \int_S \mathbb{E}\left[h(\xi, x)\, \lambda(\xi)\right] \mathrm{d}\xi.$$

Assume that X is a stationary point process on $\mathbb{R}^d$ with intensity $0 < \lambda < +\infty$. For $\xi \in \mathbb{R}^d$, $x \in N_{\mathrm{lf}}$, y $F \subseteq N_{\mathrm{lf}}$, let $x + \xi = \{\eta + \xi : \eta \in x\}$ denote the translation of

the point configuration x by ξ, and $F + \xi = \{x + \xi : x \in F\}$ the translation of F by ξ. In the stationary case

$$P_0^!(F) = \mathbb{E}\left[\sum_{\xi \in X_B} \frac{\mathbf{1}\left[X\setminus\{\xi\} \in F + \xi\right]}{\lambda\,|B|}\right], \qquad F \subseteq N_{\mathrm{lf}},$$

for an arbitrary set $B \subseteq S$ with $0 < |B| < +\infty$, and $P_\xi^!(F) = P_0^!(F - \xi)$. Moreover

$$\mathbb{E}\left[\sum_{\xi \in X} h(\xi, X\setminus\{\xi\})\right] = \lambda \iint h(\xi, x + \xi)\,\mathrm{d}P_0^!(x)\,\mathrm{d}\xi,$$

for non-negative functions h.

Consider the problem of estimating $P_0^!(F)$ for some $F \in \mathscr{B}$. Since X is stationary we may consider $X - \xi$, $\xi \in X$, for a bounded B, as representing observations of X conditional on $0 \in X$. Thus a natural estimator of $P_0^!(F)$ is the empirical average

$$\frac{1}{N(B)}\sum_{\xi \in X_B} \mathbf{1}\left[X\setminus\{\xi\} \in F + \xi\right].$$

If B is large, we may expect that $\lambda \approx N(B)/|B|$, and so we obtain the estimator

$$\frac{1}{\lambda\,|B|}\sum_{\xi \in X_B} \mathbf{1}\left[X\setminus\{\xi\} \in F + \xi\right].$$

Interpretation of $\mathscr{K}$ as Palm expectation

The second-order reduced moment measure $\mathscr{K}$ has an interpretation as a Palm expectation, since

$$\mathscr{K}(B) = \int_{\mathbb{R}^d} \sum_{\eta \in X_B} \frac{\mathbf{1}\left[\eta - \xi \in B\right]}{\lambda(\eta)}\,\mathrm{d}P_\xi^!(x),$$

for almost all $\xi \in \mathbb{R}^d$. In the stationary case, it follows that

$$\mathscr{K}(B) = \frac{\mathbb{E}_0^!(N(B))}{\lambda}, \qquad B \subseteq \mathbb{R}^d,$$

where $E_0^!$ denotes expectation with respect to $P_0^!$.

2.1.3 Models

Poisson point process

Poisson point processes play a fundamental role. They serve as a tractable model class for *no interaction* or *complete spatial randomness* in spatial point patterns. They also serve as reference processes when summary statistics are studied. A point process with *n i.i.d* points with density f is called a binomial point process with *n* points in *B*. We write $X \sim Binomial\,(B, n, f)$.

A point process X on S is a Poisson point process with intensity function λ if the following properties are satisfied

   i. For any $B \in \mathscr{B}$ with $\mu\,(B) < +\infty$, $N\,(B)$ follows a Poisson distribution with mean $\mu\,(B)$.

   ii. For any $n \in \mathbb{N}$ and $B \in \mathscr{B}$ with $0 < \mu\,(B) < +\infty$, conditional on $N(B) = n$, X_B follows a Binomial distribution with $f\,(\xi) = \lambda\,(\xi)\,/\mu\,(B)$ and number of points n.

We write $X \sim Poisson(S, \lambda)$. If λ is constant, the process $Poisson(S, \lambda)$ is called a homogeneous Poisson process on S with rate or intensity λ; else it is said to be an inhomogeneous Poisson process on S. A homogeneous Poisson point process is a stationary and isotropic process. A Poisson point process exists and it is uniquely determined by its intensity measure.

If $X_i \sim Poisson(S, \lambda_i)$, for $i \in \mathbb{N}$ mutually independent and

$$\lambda = \sum_{i=1}^{+\infty} \lambda_i,$$

is locally integrable, then with probably one,

$$X = \bigcup_{i=1}^{+\infty} X_i,$$

is a disjoint union, and $X \sim Poisson(S, \lambda)$.

The Neyman-Scott process

We consider here Neyman-Scott processes, proposed initially by Neyman and Scott (1958). Let C be a stationary Poisson process on $\mathbb{R}^d$ with intensity $\kappa > 0$.

Conditional on C, let $X_c, c \in C$, be independent Poisson processes on $\mathbb{R}^d$ where X_c has intensity function

$$\lambda_c(\xi) = \alpha \kappa(\xi - c),$$

where $\alpha > 0$ is a parameter and κ is a kernel (i.e. for all $c \in \mathbb{R}^d$, $\kappa(\xi - c)$ is a density function). Then

$$X = \bigcup_{c \in C} X_c,$$

is a special case of a *Neyman-Scott process* with cluster centres C and clusters $X_c, c \in C$ (in the general definition of a Neyman-Scott process, $n(X_c)$ given C is not restricted to be a Poisson variate, see e.g. Chiu et al. (2013).

Matérn cluster process is a special case of a Neyman-Scott process where the density function $\kappa(\xi)$, is the uniform density on the ball with centre zero and radius r. Let ω_d and σ_d be the volume and the surface area of the d-dimensional unit ball, respectively. Then

$$\omega_d = \frac{\pi^{d/2}}{\Gamma(1 + d/2)}, \quad \sigma_d = \frac{2\pi^{d/2}}{\Gamma(d/2)},$$

where Γ is the classical Gamma function. Under the definition of the Neyman-Scott process, the Matérn cluster process density is given by (Matérn, 1986)

$$\kappa(\xi) = \frac{\mathbf{1}[\|\xi\| \leq r]}{\omega_d r^d},$$

the uniform density on the ball $b(\mathbf{0}, r)$.

The hard-core process

The construction of a hard-core process is based on the sequential approach. Let $r > 0$ and $B \in \mathscr{B}_0$ be given. A simple sequential inhibition (SSI) process in the set B is constructed in the following way:

  i. choose $\xi_1 \in B$ uniformly at random,

 ii. if $k - 1$ points are chosen, choose ξ_k uniformly in $B \setminus \bigcup_{i=1}^{k-1} b(\xi_i, r)$,

iii. the construction ends in n steps, if $B \subseteq \bigcup_{i=1}^{n} b(\xi_i, r)$.

Since $B \setminus \bigcup_{i=1}^{k-1} b(\xi_i, r)$ can have complicated geometrical shape, in practice the process is simulated by the rejection method. A point ξ_i is generated uniformly in

the window B and if it lies closer than r from an existing point, then it is rejected and a new point is generated.

2.2 Summary statistics

2.2.1 Second-order summary statistics

The K- and L-functions for a second-order reweighted stationary point process are defined by

$$K(r) = \mathcal{K}(b(\mathbf{0},r)), \qquad L(r) = \left(\frac{K(r)}{\omega_d}\right)^{1/d}, \quad \text{for } r > 0.$$

This definition, which extends the definition of Ripley's K-function (Ripley (1976, 1977)) for the stationary case to the case of second-order intensity re-weighted stationarity, is due to Baddeley et al. (2000). In the stationary case, $\lambda K(r)$ is the expected number of further points within distance r from the origin given that X has a point at the origin. For a stationary and isotropic process

$$K(r) = \sigma_d \int_0^r u^{d-1} g(u) \mathrm{d}u \quad \text{or} \quad g(r) = \frac{1}{\sigma_d r^d} \frac{\mathrm{d}}{\mathrm{d}r} K(r).$$

For a stationary Poisson point process

$$K(r) = \omega_d r^d = \frac{\pi^{d/2} r^d}{\Gamma\left(1 + \frac{d}{2}\right)} \qquad \text{and} \qquad L(r) = r.$$

So, in the particular case of $d = 2$,

$$K(r) = \pi r^2 \qquad \text{and} \qquad g(r) = 1.$$

The K- and L-functions are in one-to-one correspondence, and in applications the L-function is often used instead of the K-function. One reason is that L is the identity for a Poisson process. In general, at least for small values of r, $L(r) - r > 0$ indicates aggregation or clustering at distances less than r, and $L(r) - r < 0$ regularity at distances less than r. This may be due to certain latent processes or attraction or repulsion between the points. Moreover, for a homogeneous Poisson process, the transformation $K \to L$ is variance stabilising when K is estimated by non-parametric methods (Besag, 1977).

2.2.2   Summary statistics based on inter-point distances

Assume that X is stationary. The *empty space function F* is the distribution function of the distance from the origin (or another fixed point $\mathbb{R}^d$) to the nearest point in X, i.e.

$$F(r) = \mathbb{P}(X \cap b(\mathbf{0}, r) \neq \emptyset), \qquad r > 0.$$

The *nearest-neighbour function G* is

$$G(r) = \frac{1}{\lambda |A|} \mathbb{E} \left[\sum_{\xi \in X \cap A} \mathbf{1}[(X \setminus \{\xi\}) \cap b(\xi, r) \neq \emptyset] \right], \qquad r > 0,$$

for an arbitrary set $A \subset \mathbb{R}^d$ with $0 < |A| < +\infty$. For the nearest-neighbour function G, it follows that

$$G(r) = \mathbb{P}_0^!(N(b(\mathbf{0}, r)) > 0), \qquad r > 0.$$

The *J-function* is defined by

$$J(r) = \frac{1 - G(r)}{1 - F(r)}, \quad \text{for} \quad F(r) < 1.$$

The J-function was suggested by van Lieshout and Baddeley (1996). For a stationary Poisson process on $\mathbb{R}^d$ with intensity $\lambda < \infty$,

$$F(r) = G(r) = 1 - \exp(-\lambda \omega_d r^d) \quad \text{and} \quad J(r) = 1 \quad r > 0,$$

where the equality for G comes from the *Slivnyak-Mecke formula*.

In general, at least for small values of $r > 0$, $F(r) < G(r)$ (or $J(r) < 1$) indicates *clustering*, and $F(r) > G(r)$ (or $J(r) > 1$) indicates *regularity*, but if $J(r) = 1$ does not imply that X is a stationary Poisson process (see Bedford and van den Berg (1997)). For Neyman-Scott point processes, the J-function can be expressed as

$$J(r) = \int k(\xi) \exp \left(-\alpha \int_{\|\eta\| \leq r} k(\xi + \eta) \mathrm{d}\eta \right) \mathrm{d}\xi,$$

thus $J(r)$ is non-increasing for $r > 0$ with range $(\exp(-\alpha), 1)$. So $F(r) < G(r)$ for $r > 0$.

2.3 Extension to multivariate processes

Let Y be a point process on $U \subseteq \mathbb{R}^d$. Given some space M, if a random *mark* $m_\xi \in M$ is attached to each point $\xi \in Y$, then

$$X = \left\{ (\xi, m_\xi) : \xi \in Y \right\},$$

is called a marked point process with points in U and mark space M. One simple example is a *multitype point process*, where $M = \{1, \ldots, n\}$ and the marks specify k different types of points. This is equivalent to a *k-dimensional multivariate point process*, that is a tuple $(X_1, \ldots, X_k)$ of point processes $X_1, \ldots, X_k$ corresponding to the k different types of points.

2.3.1 Marked Poisson processes

Consider a marked point process $X = \left\{ (\xi, m_\xi) : \xi \in Y \right\}$ with points in U and mark space M. Suppose that Y is $Poisson(U, \phi)$, where ϕ is a locally integrable intensity function, and conditional on Y, the marks $\left\{ m_\xi : \xi \in Y \right\}$ are mutually independent. Then X is a marked Poisson process. If the marks are identically distributed with a common distribution Q, then Q is called the mark distribution.

Let X be a marked Poisson process with $M \in \mathscr{B}^p$ and where, conditional on Y, each mark m_ξ has a discrete or continuous density λ_ξ which does not depend on $Y \backslash \{\xi\}$. Let $\lambda(\xi, m) = \phi(\xi) \lambda_\xi(m)$. Then

i. $X \sim Poisson(U \times M, \lambda)$.

ii. If the density on M defined by $\kappa(m) = \int \lambda(\xi, m) \, \mathrm{d}\xi$ is locally integrable, then $\left\{ m_\xi : \xi \in Y \right\} \sim Poisson(M, \kappa)$.

2.3.2 Multivariate Poisson processes

By a multivariate Poisson process it is usually meant that each X_i is a stationary Poisson process on $\mathbb{R}^d$ with intensity $0 < \lambda_i < +\infty$ for $i = 1, \ldots, k$, see e.g. Diggle (2013). We have the equivalence between the following two properties:

1. $\mathbb{P}\left(m_\xi = i | Y = y \right) = p_\xi(i)$ depends only on ξ for realisations y of Y and $\xi \in y$.

2. $(X_1, \ldots, X_k)$ is a multivariate Poisson process with independent components $X_i \sim Poisson(U, \lambda_i)$ where $\lambda_i(\xi) = \phi(\xi) p_\xi(i)$, $i = 1, \ldots, k$.

A common hypothesis for marked point processes $\{m_\xi : \xi \in Y\}$ is that of random labelling which means that conditional on Y, the marks m_ξ are mutually independent and the distribution of m_ξ does not depend on Y.

2.3.3 Cross-moment measure

Consider a multivariate point process $X = \{X_i\}_{i=1}^k$, we assume that each X_i is a point process in $\mathbb{R}^d$ with intensity function λ^i and count function N_i. Let $i, j \in \{1, \ldots, k\}$ be different, and set $a/0 = 0$, for $a \geq 0$.

i. We define the cross-moment measure for points of types i and j by

$$\alpha^{ij}(C) = \mathbb{E}\left[\sum_{\xi \in X_i, \eta \in X_j} \mathbf{1}\left[(\xi, \eta) \in C\right]\right], \qquad C \in \mathscr{B}^d \times \mathscr{B}^d.$$

ii. If α^{ij} can be written as

$$\alpha^{ij}(C) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{1}\left[(\xi, \eta) \in C\right] \lambda^{ij(2)}(\xi, \eta)\,\mathrm{d}\xi\,\mathrm{d}\eta, \qquad C \in \mathscr{B}^d \times \mathscr{B}^d,$$

where $\lambda^{ij(2)}$ is a non-negative function, then $\lambda^{ij(2)}$ is called the cross second-order product density.

The cross pair correlation function for points of types i and j is defined by

$$g^{ij}(\xi, \eta) = \frac{\lambda^{ij(2)}(\xi, \eta)}{\lambda^i(\xi)\lambda^j(\eta)}.$$

Suppose that the measure

$$\mathscr{K}^{ij}(B) = \frac{1}{|A|} \sum_{\xi \in X_i, \eta \in X_j} \frac{\mathbf{1}\left[\xi \in A, \eta - \xi \in B\right]}{\lambda^i(\xi)\lambda^j(\eta)}, \qquad B \in \mathscr{B}^d$$

does not depend on the choice of $A \in \mathscr{B}^d$ with $0 < |A| < +\infty$. Then (X_i, X_j) is said to be *cross second-order intensity reweighted stationary*.

Consider a multivariate point process $X = \{X_i\}_{i=1}^k$, we assume that each X_i is a point process in $\mathbb{R}^d$ with intensity function λ^i and count function N_i. The cross K- and L-functions are defined by

$$K^{ij}(r) = \mathscr{K}^{ij}(b(\mathbf{0}, r)), \qquad L^{ij}(r) = \left(\frac{K^{ij}(r)}{\omega_d}\right)^{1/d} \qquad r > 0.$$

If X_i and X_j are independent, then (X_i, X_j) is cross second-order reweighted stationary,

$$K^{ij}(B) = K^{ji}(B) = |B|, \qquad \text{and} \qquad L^{ij}(r) = r.$$

Assume that (X_i, X_j) is stationary with intensities $0 < \lambda^i < +\infty$ and $0 < \lambda^j < +\infty$. The nearest-neighbour function G^{ij} is the distribution function for the distance from a typical type i point to its nearest type j point, i.e.

$$G^{ij}(r) = \frac{1}{\lambda^i |A|} \mathbb{E}\left[\sum_{\xi \in X_i \cap A} \mathbf{1}[X_j \cap b(\xi, r) \neq \emptyset] \right], \qquad r > 0,$$

for an arbitrary set $A \subset \mathbb{R}^d$ with $0 < |A| < +\infty$. Moreover, define

$$J^{ij}(r) = \frac{1 - G^{ij}(r)}{1 - F^j(r)}, \quad \text{for} \quad F^i(r) < 1.$$

The cross statistics G^{ij} and J^{ij} are not symmetric in i and j.

2.4   Non-parametric estimation

The non-parametric methods do not assume a specific parametric model for the first- and second-order behaviour. The estimation approach is typically based on simple counts and kernel methods. We consider methods for analysing spatial point pattern data not linked to special families of parametric models.

Let X be a spatial point process on $\mathbb{R}^d$ with intensity function λ; if X is stationary, λ is assumed to be a constant with $0 < \lambda < +\infty$. Whenever needed we assume that the measure $\mathcal{K}$ exists. We confine to the case where a single point pattern $X_W = x$ is observed in a bounded window $W \in \mathcal{B}^d$ with $|W| > 0$, and discuss non-parametric estimation of λ, K, g, L and J functions in the isotropic case. Higher-order summary statistics can be introduced as well, but the corresponding non-parametric estimators may be less stable if the number of points observed is not sufficiently large; see Peebles and Groth (1975), Stoyan and Stoyan (1994), Møller et al. (1998).

2.4.1   Estimation of intensity functions

In the homogeneous case, a natural unbiased estimate of the intensity function
is

$$\hat{\lambda} = \frac{N(X_W)}{|W|}.$$

This estimator is unbiased and, if N is ergodic, then it is consistent, i.e. as
W increases it converges to the true value λ. This holds independent of the
specific distribution of N, whereas the variability of the estimator $\hat{\lambda}$ is of course
distribution-dependent (Illian et al., 2008). This is in fact the maximum likelihood
estimate if X is a homogeneous Poisson process. In the inhomogeneous case, a
non-parametric kernel estimate of the intensity function, following Diggle (1985),
is

$$\hat{\lambda}_\varepsilon(\xi) = \sum_{\eta \in X_W} \frac{\kappa_\varepsilon(\xi - \eta)}{C_{W,\varepsilon}(\eta)}, \qquad \xi \in W.$$

Here κ_ε is a kernel with bandwidth $\varepsilon > 0$, i.e.

$$\kappa_\varepsilon(\xi) = \frac{1}{\varepsilon^d} \kappa\left(\frac{\xi}{\varepsilon}\right),$$

where κ is a given density function, and

$$C_{W,\varepsilon}(\eta) = \int_W \kappa_\varepsilon(\xi - \eta) \, d\xi,$$

is an *edge-correction factor*.

The estimate of the intensity is usually sensitive to the choice of ε. When $d = 2$,
it is common to use a product kernel given by $\kappa(\xi) = e(\xi_1) e(\xi_2)$ for $\xi = (\xi_1, \xi_2) \in \mathbb{R}^d$,
where

$$e(u) = \frac{3}{4}(1 - |u|) \mathbf{1}[|u| \le 1], \qquad u \in \mathbb{R},$$

is the *Epanechnikov kernel* (Epanechnikov, 1969). $\int_W \hat{\lambda}_\varepsilon(\xi) \, d\xi$ is an unbiased
estimate of $\mu(W)$. A simple stationarity test is closely related to the intensity,
and can be built by calculating an estimator of the intensity $\lambda(\cdot)$ and verify that
a plot of the obtained estimate shows only local irregularities but not a general
trend.

2.4.2 Estimation of K-, L- and g-functions

For non-parametric estimation of $\mathcal{K}$, we consider the following steps. Let $|\Phi|_{\alpha}$ denote the α-dimensional Hausdorff measure of $\Phi \in \mathscr{B}^d$. See Stoyan and Stoyan (1994). Let $\xi, \eta \in W$, we define the *Ripley's isotropic edge-correction factor*, (see Ripley (1976), Illian et al. (2008)) as

$$w^d(\xi, \eta) = \frac{|\partial b(\xi, \|\xi - \eta\|) \cap W|_{d-1}}{|\partial b(\xi, \|\xi - \eta\|)|_{d-1}}.$$

Here $|\cdot|_0$ and $|\cdot|_1$ are the zero- and one-dimensional Hausdorff measures respectively in $\mathbb{R}$ and $\mathbb{R}^2$. The zero-dimensional Hausdorff measure in $\mathbb{R}$ is the number of points of the set. The one-dimensional Hausdorff measure of a simple curve in $\mathbb{R}^2$ is equal to the length of the curve.

Suppose that X is second-order intensity reweighted stationary, then

$$\sum_{\xi, \eta \in X_W}^{\neq} \frac{\mathbf{1}[\eta - \xi \in B]}{\lambda(\xi)\lambda(\eta)w^d(\xi, \eta)},$$

is an unbiased estimator for $\mathcal{K}(B)$. In practice λ is not known, so $\lambda(\xi)\lambda(\eta)$ must be replaced by the estimator $\widehat{\lambda(\xi)\lambda(\eta)}$. The combined estimator

$$\hat{\mathcal{K}}(B) = \sum_{\xi, \eta \in X_W}^{\neq} \frac{\mathbf{1}[\eta - \xi \in B]}{\widehat{\lambda(\xi)\lambda(\eta)}w^d(\xi, \eta)},$$

is then biased (see Illian et al. (2008)). In fact unbiasedness is usually unobtainable for many estimators in spatial statistics, but instead they are often ratio-unbiased, i.e. of the form $\hat{\theta} = Y/Z$ where $\theta = \mathbb{E}[Y]/\mathbb{E}[Z]$. For example, in the homogeneous case, if $\widehat{\lambda(\xi)\lambda(\eta)} = \widehat{\lambda^2}$ is unbiased, then it is ratio-unbiased. Illian et al. (2008) discuss various possibilities for the homogeneous case: one possibility is to transform the estimate to obtain $(N(X_W))^2 / |W|^2$ as an estimate of λ^2; an alternative is

$$\widehat{\lambda^2} = \frac{N(X_W)(N(X_W) - 1)}{|W|^2},$$

which is unbiased for a Poisson process. For the inhomogeneous case, Baddeley et al. (2000) propose to use $\widehat{\lambda(\xi)\lambda(\eta)} = \bar{\lambda}_\varepsilon(\xi)\bar{\lambda}_\varepsilon(\eta)$ where

$$\bar{\lambda}_\varepsilon(\xi) = \sum_{\eta \in X_W \setminus \{\xi\}} \frac{\kappa_\varepsilon(\xi - \eta)}{C_{W,\varepsilon}(\eta)}, \qquad \xi \in W,$$

is a slight modification. Also, Baddeley et al. (2000) show that for an inhomogeneous Poisson process, $\bar{\lambda}_{\varepsilon}(\xi)$ is less biased than $\hat{\lambda}_{\varepsilon}(\xi)$ when ξ is a data point.

Consider $\hat{K}(r) = \hat{\mathscr{K}}(b(\mathbf{0}, r))$:

1. From its definition, it is clear that $1/w^d(\xi, \eta) \geq 1$. Also, the larger weights tend to be associated with pairs of events separated by large distances. Typically, $\mathbb{V}\text{ar}\{\hat{K}(r)\}$ tends to increase with r.

2. The dimension of W clearly limits the range of values of r which can be considered. In practice, the increasing variance of $\hat{K}(r)$ is a more serious limitation. As a rough guide, for data on a rectangle W, it is usually not worth trying to estimate $K(r)$ at values of r bigger than one-half the length of the shorter side of W.

The estimate of $L(r)$ obtained from transforming that of $K(r)$ is in general biased.

An alternative edge-corrected estimator for the pair correlation function (Fiksel (1988), Stoyan and Stoyan (1994), Baddeley et al. (2000)) is given by

$$\hat{g}(r) = \frac{1}{\sigma_d r^d |W|} \sum_{\xi, \eta \in X_W}^{\neq} \frac{\kappa_b(\|\xi - \eta\| - r)}{\widehat{\lambda(\xi)\lambda(\eta)} w^d(\xi, \eta)}.$$

Here $\kappa_b(u) = \kappa(u/b)/b$, $u \in \mathbb{R}$ and bandwidth $b > 0$, (see Illian et al. (2008)).

2.4.3 Estimation of F-, G- and J-functions

Reduced-sample estimators of F and G are derived using minus sampling. Let

$$d(\xi, B) = \inf\{\|\xi - \eta\| \mid \eta \in B\},$$

be the shortest distance from a point $\xi \in \mathbb{R}^d$ to a set $B \subset \mathbb{R}^d$. Let $I \subset \mathbb{R}^d$ denote a finite regular grid of points (chosen independently of X), and let $\#I_r$ denote the cardinality of the set $I_r = I \cap W_{\ominus} r$, where $W_{\ominus} r = \{\xi \in W : b(\xi, r) \subseteq W\}$ for $r > 0$.

The following estimator is unbiased for F

$$\hat{F}(r) = \sum_{\xi \in I_r} \frac{\mathbf{1}[d(\xi, X_W) \leq r]}{\#I_r},$$

for $\#I_r > 0$, and the next one is a ratio-unbiased estimator for G

$$\hat{G}(r) = \sum_{\xi \in X_W \cap W_\ominus} \frac{\mathbf{1}[d(\xi, X_W) \le r]}{\hat{\lambda}\,|W_\ominus|},$$

for $|W_\ominus| > 0$, so that $\hat{J}(r)$ is given by

$$\hat{J}(r) = \frac{1 - \hat{G}(r)}{1 - \hat{F}(r)}, \quad \text{for} \quad \hat{F}(r) < 1.$$

2.4.4 Multivariate estimators

If (X_i, X_j) is cross second-order reweighted stationary, then for $B \in \mathscr{B}^d$

$$\sum_{\substack{\xi \in X_i \cap W \\ \eta \in X_j \cap W}}^{\ne} \frac{\mathbf{1}[\eta - \xi \in B]}{\lambda_i(\xi)\,\lambda_j(\eta)\,w^d(\xi, \eta)},$$

is an unbiased estimator of $\mathscr{K}^{ij}(B)$. We can substitute $\lambda_i(\xi)\,\lambda_j(\eta)$ with a non-parametric estimator $\widehat{\lambda_i(\xi)\,\lambda_j}(\eta)$.

If $X = (X_1, \ldots, X_k)$ is a multivariate point process, then the $k(k+1)/2$ functions $K^{ij}(r)$ with $1 \le i \le j < k$, completely describe the *second-order* properties of the process. A necessary, but not sufficient, condition for the process of points of type i to be independent of the process of points of type j is that $K^{ij}(r) = \omega_d r^d$ for all r. Not surprisingly, description and estimation of the second-order structure of a multitype process requires consideration of only two of the types at a time. Therefore it will be sufficient when discussing the problem of estimation to consider a two-type process consisting of n_1 points $\{\varkappa_1, \ldots, \varkappa_{n_1}\}$ and n_2 points $\{\varsigma_1, \ldots, \varsigma_{n_2}\}$. Suppose that we observe such a process over a plane region W.

Lotwick and Silverman (1982), and Diggle (2013), give a combined estimator for the bivariate K- function:

$$\hat{K}^{ij}(r) = \frac{|W|}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \frac{\mathbf{1}[\|\varkappa_i - \varkappa_j\| \le r]}{w^{d*}(\varkappa_i, \varsigma_j)},$$

where

$$w^{d*}(\varkappa_i, \varsigma_j) = \frac{n_1 w^d(\varsigma_i, \varkappa_j) + n_2 w^d(\varkappa_i, \varsigma_j)}{n_1 + n_2}.$$

If g^{ij} is isotropic, a kernel estimator for the cross pair correlation is given by

$$\hat{g}^{ij}(r) = \frac{1}{\sigma_d r^d |W|} \sum_{\xi, \eta \in X_W}^{\neq} \frac{\kappa_b(\|\xi - \eta\| - r)}{\lambda^i(\widehat{\xi}) \widehat{\lambda^j}(\eta) w^d(\xi, \eta)}.$$

Similarly, for (X_i, X_j) the reduced-sample estimator of $G^{ij}(r)$ is

$$\hat{G}^{ij}(r) = \sum_{\xi \in X_i \cap W_{\ominus}r} \frac{\mathbf{1}[d(\xi, X_j \cap W) \leq r]}{\hat{\lambda}^i |W_{\ominus}|},$$

By substitution of $\widehat{F}^j(r)$ and $\widehat{G}^{ij}(r)$ functions the estimator for $\hat{J}^{ij}(r)$ is obtained.

2.5 Envelopes procedure

Consider a simple hypothesis H_0. Confidence intervals and other distributional characteristics associated with some non-parametric estimate $\hat{R}$ of a second-order summary can be obtained by a *bootstrap* using simulation under H_0. For a given distance $r \in B$, let $T_0(r) = T(X, r)$ denote any statistic obtained from the point process X observed within the window W. Let $\Theta = \{T_i(r)\}_{i=1}^n$ be obtained from *i.i.d.* simulations $X_1, \ldots, X_n$ under H_0. From the empirical distribution of Θ we can estimate any quantile for the distribution of $T_0(r)$ under H_0, and we can do this with any desired precision if n is large enough. Notice that although $T_1, \ldots, T_n$ are *i.i.d.*, the random vectors $(T_1(r), \ldots, T_n(r))$ considered for different values of r are dependent. So some caution should be taken when we compare the results for different values of r

If the computation of $T_i(r)$, $i = 1, \ldots, n$, is time consuming, the following envelopes may be used where n is small. Let

$$T_{\min}(r) = \min \Theta, \qquad \text{and} \qquad T_{\max}(r) = \max \Theta,$$

under H_0,

$$\mathbb{P}(T_0(r) < T_{\min}(r)) = \mathbb{P}(T_0(r) > T_{\max}(r)) \leq \frac{1}{n+1},$$

with equality if $T_0(r), T_1(r), \ldots, T_n(r)$ are almost surely different. The bounds $T_{\min}(r)$ and $T_{\max}(r)$ are called the $100/(n+1)\%$-lower and the $100n/(n+1)\%$-upper envelopes.

2.6 Replicated spatial point patterns

In the literature (Baddeley et al., 1993; Diggle et al., 1991, 2000; Hahn, 2012), second-order characteristics have been used to find a methods to discriminate between groups of replicated patterns. Usually these methods have been focused on Ripley's K function by defining a series of statistics analogous to those used in classic ANOVA analysis and by taking into account that $\mathbb{Var}(\hat{K}(r))$ is roughly proportional to r^2/n^2 if r is small.

2.6.1 Diggle et al.'s Monte Carlo test

Consider an experiment with g groups of point patterns, each one containing m_i replicates. Let $w_{ij} = n_{ij}/n_i$ ($n_i = \sum_{j=1}^{m_i} n_{ij}$), and $n = \sum_{i=1}^{g} n_i$. Following Diggle et al. (1991, 2000), we can define the estimated group-specific mean function as

$$\bar{K}_i(r) = \sum_{j=1}^{m_i} w_{ij} \hat{K}_{ij}(r), \qquad i = 1,...,g, \tag{2.1}$$

and an overall average mean function as

$$\bar{K}(r) = \frac{1}{n} \sum_{i=1}^{g} n_i \bar{K}_i(r), \tag{2.2}$$

and then consider the statistic

$$D_g = \sum_{i=1}^{g} \int_0^{r_0} n_i \frac{1}{r^2} \left[\bar{K}_i(r) - \bar{K}(r) \right]^2 \mathrm{d}r. \tag{2.3}$$

The statistic D_g is an intuitively sensible measure of the difference between K-functions along the groups and is loosely analogous to a residual sum of squares in a conventional one-way ANOVA.

The interest focuses on testing the null hypothesis that the expectations of the empirical K-functions do not differ between groups. The analytical form of the distribution of D_g is not known. A pure randomisation test is available in which one permutes the $\hat{K}_{ij}(r)$ across groups and recomputes D_g to obtain its exact conditional distribution.

Define

$$\hat{R}_{ij}(r) = \sqrt{n_{ij}} \left[\hat{K}_{ij}(r) - \bar{K}_i(r) \right]. \tag{2.4}$$

$R_{ij}(r)$ should be approximately exchangeable quantities because the sampling variance of each $\hat{K}_{ij}(r)$ is proportional to n_{ij}^{-1}. They call the $\hat{R}_{ij}(r)$ residual K-functions. Obtain an empirical approximation to the distribution of D_g by recomputing its values from a large number of bootstrap samples

$$\hat{K}_{ij}^*(r) = \bar{K}(r) + \frac{\hat{R}_{ij}^*(r)}{\sqrt{n_{ij}}}, \tag{2.5}$$

where the $\hat{R}_{ij}^*(r)$ are obtained by drawing at random and without replacement, keeping group sizes fixed, from the empirical distribution of the $\hat{R}_{ij}(r)$.

2.6.2 Studentized permutation test

Hahn (2012) stated that non-uniformity of p-values under the null hypothesis as observed with Diggle's et al.'s Monte Carlo test is apparently a general problem of bootstrap tests based on small samples. As an alternative, Hahn (2012) proposed two *pure permutation tests*. Such tests have uniformly distributed rejection rates by construction even when sample sizes are small, as long as the samples are exchangeable. In order to achieve robustness of the test towards heteroscedasticity, they suggest to use a statistic related to the Behrens-Fisher-Welch t-statistic, or alternatively the corresponding F-statistic. This is generalised to the functional data case by considering the L_2-norm of the t-statistic, i.e. the integral over the squared studentized differences between the group means

$$T = \sum_{1 \leq i < j \leq g} \int_0^{r_0} \frac{\left(\bar{K}_i(r) - \bar{K}_j(r)\right)^2}{\frac{1}{m_i}s_i^2(r) + \frac{1}{m_j}s_j^2(r)}\,\mathrm{d}r, \tag{2.6}$$

where

$$\bar{K}_i(r) = \frac{1}{m_i}\sum_{j=1}^{m_i}\hat{K}_{ij}(r) \qquad \text{and} \qquad s_i^2(r) = \frac{1}{m_i-1}\sum_{j=1}^{m_i}\left(\hat{K}_{ij}(r) - \bar{K}_i(r)\right)^2.$$

Tests using the statistic T are still sensitive to pronounced heteroscedasticity. In these cases, its better to use the following statistic instead

$$\overline{T} = \sum_{1 \leq i < j \leq g} \int_0^{r_0} \frac{\left(\bar{K}_i(r) - \bar{K}_j(r)\right)^2}{\frac{1}{m_i}\overline{s_i^2}(r) + \frac{1}{m_j}\overline{s_j^2}(r)}\,\mathrm{d}r,$$

where

$$\overline{s_i^2}(r) = \frac{r^2}{r_0} \int_0^{r_0} \frac{s_i^2(t)}{t^2} \mathrm{d}t.$$

The variance estimator used in $\overline{T}$ is more stable than the individual denominators in the statistic T. $\overline{T}$ should therefore be preferred to T in the cases of heteroscedasticity.

Chapter 3

On spatio-temporal point process statistics: a review

Abstract

Spatio-temporal point process data have been analysed quite a bit in specialised fields, with the aim of better understanding the inherent mechanisms that govern the temporal evolution of events placed in a planar region. In particular, in the last decade there has been an acceleration of methodological developments, accompanied by a broad collection of applications as spatio-temporally indexed data have become more widely available in many scientific fields. We present a self-contained review describing statistical models and methods that can be used to analyse patterns of points in space and time when the questions of scientific

interest concern both their spatial and their temporal behaviour. We revisit moment characteristics that define summary statistics, as well as conditional intensities which uniquely characterise certain spatio-temporal point processes. We make use of these concepts to describe models and associated methods of inference for spatio-temporal point process data. Three new motivating real-data examples are described and analysed throughout the paper to illustrate the most relevant techniques, discussing the pros and cons of the different considered approaches.

Keywords   Edge-correction; Empirical models; Intensity function; Mechanistic models; Second-order properties; Separability.

3.1   Introduction

The term *spatio-temporal/spatial-temporal point pattern* reveals two main pieces of information (regarding space and time) about the data considered. We are considering a collection of data which may be treated as 1) the realisation of a random collection of points, which 2) somehow evolves in space and time. The notion of something evolving in space and time is rather vague in the sense that it is not directly revealed how that evolution actually occurs. Consulting the literature, it becomes clear that the term spatio-temporal point pattern or, the term *spatio-temporal point process* if referring to the data generating mechanism, has been used for an array of different entities. As indicated in e.g. Cox and Isham, 1980; Illian et al., 2008, considering a non-discrete spatial study region W and a time frame/temporal study period T, we may roughly put them into the following main categories in which we refer to the locations as *events*:

1. The data $\{(\mathbf{u}_i, v_i)\}_{i=1}^n \subseteq W \times T$ are treated as being generated as a snapshot in space-time. More specifically, the data are treated as a collection of instantaneous events, each occurring at a given spatial location $\mathbf{u}_i$, with a given associated time point/event time v_i. In other words, at a given time point we observe at most one event. A point is thus not considered to remain in W after its occurrence. Typical applications include earthquakes and disease outbreaks. Essentially, this category may be viewed as a spatial point process with a further (temporal) dimension (see e.g. Daley and Vere-Jones, 2003; Diggle et al., 1995; Greenspan, 2013; Schoenberg et al., 2006; Vere-Jones, 2009). The questions posed are in many ways the same as those

posed in the study of spatial point patterns, as well as those in the study of events occurring according to a temporal point process; two main factors in the analysis are related to intensity variations and (spatio-temporal) interaction.

2. During T we observe a total of n points $\{\mathbf{u}_i\}_{i=1}^n \subseteq W$. The ith point arrives at some given time $v_i \in T$, obtains a spatial location $\mathbf{u}_i \in W$ and stays there for some given period of time $l_i > 0$, after which it is removed from W. Hence, we may express the data generating process as $Z(t) \subseteq \{\mathbf{u}_i\}_{i=1}^n \subseteq W, t \in T$. Here there are, in essence, two things of interest. Firstly, one is interested in understanding the mechanisms behind the arrivals of new points as well as those behind the removals of points (births and deaths). This basically boils down to modelling the total number of points present in W, $|Z(t)| \in \{0, 1, \ldots\}$, at any time point $t \in T$; this is done by means of some stochastic process. Secondly, there is an interest in the probabilistic laws governing the spatial aspects of additions and removals of points. The latter part has more of a spatial point process nature. Typically such a set-up is governed by a *spatial jump process*, where a subclass is given by the family of *spatial birth-death processes*; these models have been used extensively e.g. to simulate spatial point processes (Berthelsen and Møller, 2002; Daley and Vere-Jones, 2007; van Lieshout, 2000; Møller and Sørensen, 1994; Møller and Waagepetersen, 2004).

   Such an approach has been employed for e.g. the modelling of forest stands, where v_i refers to the birth time of the ith tree and l_i refers to its life time.

3. Objects move (more or less) continuously through the spatial domain W and form paths $Y_i(t) \in W$, $t \in T$, $i = 1, \ldots, m$. By sampling these movements at discrete times $T_1, \cdots, T_k \in T$ we obtain as end result the collection $\{\mathbf{u}_i\}_{i=1}^n = \bigcup_{i=1}^m \bigcup_{j=1}^k Y_i(T_j) \subseteq W$. Typical applications can be found in e.g. movement ecology (Preisler et al., 2004).

   Here one is mainly interested in understanding the underlying movement processes, which often are modelled as spatial stochastic processes (Cressie and Wikle, 2011). This analysis helps us clarify e.g. migratory patterns of animals. Statistically, the analysis is basically focussed on fitting spatial stochastic processes to discretely sampled paths.

   Note that this setting may be fitted into both of the above categories: in the latter case we let each l_i be so large that we do not remove any of the points

during T and in the former case we include the associated sampling times as v_i, $i = 1, \ldots, n$.

As already indicated, there are naturally intersections between the three categories, i.e. there are situations where the modelling approach of one category fits the data of another. However, as we have indicated, the questions of interest and the methods employed are essentially different in the three categories. Throughout we will focus on category one.

It is the intention of this review to give a thorough overview of the different tools and methods available to analyse and model spatio-temporal point patterns $\{(\mathbf{u}_i, v_i)\}_{i=1}^n \subseteq W \times T$. The main reason for producing this summary at this moment is to emphasise important contributions delivered in the last years, such as Cronie and van Lieshout (2015), as well as providing a catalogue of sort, which may be employed for the statistical analysis of the many interesting datasets appearing nowadays; the "new" datasets presented in this paper highlight this. After such an explosion of methodological and practical developments in the field of spatio-temporally indexed data, we aim at presenting a self-contained review describing statistical models and methods that can be used to analyse spatio-temporal point patterns when the questions of scientific interest concern both their spatial and temporal behaviour.

Note that many purely spatial datasets often in fact have an evolutionary nature, being the result of temporally evolving processes; spatial point patterns can appear quite differently over disjoint time windows and explaining these differences may be an essential aspect of the analysis. In essence, this aspect could be resolved by aggregating the patterns over time (Banerjee et al., 2014). Hence, adding a temporal dimension to the study of spatial point patterns can reveal many interesting and important features, which help us understand the inherent data generating mechanisms. Many datasets can be analysed as purely spatial data only if and when in so doing we can address interesting scientific questions. Despite this importance of spatio-temporal analysis, studies of spatio-temporal models have lagged a bit behind those of simple temporal models, as well as those of purely spatial models. Undoubtedly, the reasons have been largely practical, notably the difficulty lying in combining good spatio-temporal datasets and the heavy computations needed to analyse them. The scope of our discussions throughout the paper is to treat distributions in two-plus-one-dimensional space and time, i.e. W being two-dimensional (planar) and T being one-dimensional. Paraphrasing Diggle (2013), it should be emphasised that in this context two plus

one does not equal three, since the time dimension is fundamentally different from either of the two spatial dimensions. This equally happens in the marked and 3D cases, i.e. the cases 2D space + time and 2D space + marks are similar in the sense that the third dimension differs from the spatial dimension. However, the 3D case is different since all three directions are spatial and, indeed, in this case two plus one would equal three. This observation is key, even in the mathematical formulation of different related methods.

As already indicated, in the last decade there has been an acceleration of methodological development, accompanied by a diverse array of application as spatio-temporal indexed data have become more widely available in many scientific fields. Book-length treatments are now beginning to appear, including the edited collection by Finkenstadt et al. (2007), several chapters of Gelfand et al. (2010), Cressie and Wikle (2011) and, most recently, Diggle (2013) and Banerjee et al. (2014). It is evident that spatio-temporal point process techniques have covered a wide spectrum of scientific research. Environmental problems such as wildfires, earthquakes, lightning-caused fires, tornadoes and radioactive particles, among others, which represent important aspects of e.g. ecology, economy and social damage, have been well studied by many authors (see, for instance, Altieri et al., 2015; Clements et al., 2011; Karpman et al., 2013; Møller and Díaz-Avalos, 2010; Pereira et al., 2013 and the many references therein). A wealth of epidemiological problems have also been handled by employing spatio-temporal point process methodologies. Good and complete accounts of methods and models can be found in e.g. Diggle (2013) and Banerjee et al. (2014), where an impressive number of problems and data coming from the study of the spread of infections in animals and humans, and public health are treated.

Turning to some of the specifics discussed in the body of the text, we first note that in certain situations, where we treat the spatial locations $\mathbf{u}_i$ as marks of a purely temporal point process $\{v_i\}_{i=1}^n \subseteq T$, it is possible to define a given point process through its so-called conditional intensity function, which, heuristically, is the function governing the expected number of future events at some fixed time point, given all the previously observed events. Note that additional characteristics of an event, such as size, magnitude, spatial extent, or even duration, can be added as marks (additional dimensions). Daley and Vere-Jones (2003, Chapter 7) gives a rather extensive overview of the history of conditional intensities, as well as a good probabilistic coverage of the theory.

Following the terminology of Diggle (2013), we will refer to conditional intensity-based models as *mechanistic models*. This set-up may be considered the somewhat classical approach to analysing spatio-temporal point patterns (see e.g. Cox and Isham, 1980; Daley and Vere-Jones, 2003; Diggle, 2013; Karr, 1991; Snyder and Miller, 1991; Vere-Jones, 2009 and the references therein for details). Note that since the distribution of such a process is completely governed by its conditional intensity function, most of the statistical analysis and inference here reduce to analysis of conditional intensity functions. Hence, this approach presents one possible route to statistical analysis, and in particular likelihood analyses, of spatio-temporal point processes. Ogata (1998) wrote a summary paper on parametric and maximum likelihood techniques. Furthermore, regarding non-parametric estimation of the conditional intensity function, Choi and Hall (1999) considered a kernel estimation approach. The employment of mechanistic models has been considered extensively in the context of e.g. earthquake data (Choi and Hall, 1999). To exemplify the treatment of other applications, e.g., Rathbun and Cressie (1994) discuss spatio-temporal point processes in the context of tree growth and Tamayo-Uria et al. (2014) analyse the spatio-temporal distribution of rat sightings, which are directly related to rat infestation.

The arguably most prominent class of mechanistic models are the Epidemic-Type Aftershock Sequence (ETAS) models (see e.g. Daley and Vere-Jones, 2003; Ogata, 1988, 1998; Ogata and Zhuang, 2006), which have become somewhat the main tool for the analysis of earthquakes (Adelfio and Chiodi, 2015; Adelfio and Ogata, 2010; Marsan and Lengliné, 2008, 2010; Mohler et al., 2011); they belong to the more general family of Hawkes processes (see e.g. Daley and Vere-Jones, 2003; Marsan and Lengliné, 2008, 2010). In particular, hazard maps, declustering, diagnostic methods and other methods have been developed within the framework of ETAS models (Adelfio and Chiodi, 2010; Musmeci and Vere-Jones, 1992; Peng et al., 2005; van Lieshout and Stein, 2012; Zhuang et al., 2002). Also, regarding mechanistic models in general, spatio-temporal separability is a key issue (see Assunção and Maia, 2007; Chang and Schoenberg, 2011; Díaz-Avalos et al., 2014; Ogata, 1988; Schoenberg, 2003, 2004). Here separability refers to the conditional intensity function being expressed as the product of a purely spatial and a purely temporal component.

For reasons that will be noted, it is not always possible and/or convenient to treat spatio-temporal point processes by means of conditional intensities. Instead, it is often more suitable to treat them in the way they are defined, as a random

collections of (dependent) points, observed in some region, where one of the dimensions represents time. This leads us to the other approach for analysing spatio-temporal point patterns. This approach is very similar to that of purely spatial point processes, as opposed to the conditional intensity approach, which is more of an extension of the purely temporal point process approach, exploiting the natural ordering of time.

We next indicate some important aspects that will be treated in this paper. Often the analysis starts by estimating and/or modelling the intensity function, which governs the univariate properties of a spatio-temporal point process. Diggle (2013) provides an account on fully-, semi-, and non-parametric approaches to intensity estimation for spatio-temporal point patterns. Turning to the quantification of higher moments, i.e. the space-time interactions between the points, different summary statistics have been proposed. Under the assumption of stationarity, Diggle et al. (1995) considered the problem of detecting and describing spatio-temporal interactions in point process data. They extended existing second-order methods for purely spatial point process data to the spatio-temporal setting. This extension allows one to estimate spatio-temporal interaction and express it as a function of spatial and temporal lags. Furthermore, Gabriel and Diggle (2009) extend the inhomogeneous K-function of Baddeley et al. (2000) for inhomogeneous spatial point process data to the spatio-temporal setting, under the assumption of so-called second-order intensity-reweighted moment stationarity. Extending those ideas, Møller and Ghorbani (2012) study further the inhomogeneous K-function as well as the pair correlation function, under the assumption of so-called first- and second-order separability (the first- and second-order moments may be expressed as space-time products) and clarify the different consequences of such separability assumptions. Ghorbani (2013) proposes a weak stationarity test for spatio-temporal point processes and Gabriel (2014) presents an intensive simulation study to show the efficiency of the second-order estimators on different scenarios for various spatio-temporal edge-corrections. In order to account for possible anisotropy in the spatial domain, Comas et al. (2015) extend the spatial point pair orientation distribution (see Illian et al., 2008, Section 6.10) to the spatio-temporal context. The summary statistics just mentioned are all based on pairwise interaction and thus detect at most second-order interactions. For higher-order interactions, higher-order summary statistics, which take the whole point process distribution into account, are required. The J-function is a key example of these statistics. In the inhomogeneous spatio-temporal setting, it

has been defined by Cronie and van Lieshout (2015), under the assumption of so-called intensity-reweighted moment stationarity, and it reduces to a homogeneous version under the assumption of stationarity; c.f. van Lieshout and Baddeley (1996). A non-parametric estimator is also derived in the same paper. The above mentioned statistical methods will be presented in a stepwise fashion in the text. Note that the above mentioned assumptions of second-order intensity-reweighted stationarity and intensity-reweighted moment stationarity have not yet been widely tested or evaluated statistically in the available literature, with the exception of Hahn and Vedel Jensen (2016). Hence, these should be viewed as pragmatic assumptions, which are convenient starting points, and the related statistical analysis may be treated as an exciting open field of research.

In practice, one often deals with inhomogeneity, which manifests itself through e.g. parts of the spatial study region having no points and/or time periods where no events are occurring, or e.g. trends in the form of an increasing number of points in some direction or an increasing/decreasing number of points over time. Hence, depending on what interactions one is interested in studying, it is wise to proceed by assuming second-order intensity-reweighted stationarity or intensity-reweighted moment stationarity. In addition to the mechanistic models presented above, in this review we further devote a substantial part of the paper to presenting different parametric spatio-temporal models, such as Poisson processes, Cox processes, hard-core processes and inhibition processes.

Regarding the mathematical treatment of the concepts above, we recall characteristics such as intensity functions, product densities, conditional intensities, Papangelou conditional intensities and reduced Palm distributions, since these are all crucial for the development of the statistical methods presented. We then proceed to reviewing models and methods for spatio-temporal point process data. To illustrate the relevant techniques presented, we apply most of them to the three new real-data examples previously mentioned.

It should finally be mentioned that we have purposely chosen not to consider the case where we include marks in the analysis (although they are mentioned here and there). The reason is twofold; partly, the paper would become too long and, partly, for the non-conditional intensity-based set-up the study of marked spatio-temporal point patterns is rather limited (there is, however, ongoing work on the topic).

The plan of the paper is the following. We start in Section 3.2 with a description of our datasets and in Section 3.3 we present a set-up of spatio-temporal point

processes. Section 3.4 considers some characteristics of spatio-temporal point processes, discussing summary statistics, product densities, k-point correlation functions, Palm distributions and Papangelou conditional intensities for spatio-temporal point processes. Spatio-temporal empirical models are summarised in Section 3.5, and mechanistic models in Section 3.6. The paper ends with an overall discussion and depicts some existing and ongoing research lines.

3.2 Datasets

Below we describe the three data sets that are analysed throughout the paper. It should be noticed that there are cases where we only know the time up to a year and, therefore, several points appear at the same time.

3.2.1 Human outbreaks of Ebola

This dataset, obtained from Mylne et al. (2014), collates existing knowledge on the geographic spread of past Ebola outbreaks in a standardised format. They outlined simple procedures for data abstraction and each outbreak is summarised with a map and a brief text description. These data are useful for conducting spatio-temporal analyses of Ebola outbreak spread. They include every outbreak preceding the atypical 2013 Guinea epidemic which has spread further and faster than any previous epidemic. We have a temporal period from 1976 to 2012 with 96 records corresponding to the centre of each outbreak, where the time is reported as the year of the first reported case in the occurrence. Three outliers were removed in order to make easier the interpretation of results. The data locations are depicted in Figure 3.1.

3.2.2 Euphausia glacialis

Euphausia glacialis is a type of Antarctic krill of the family *Euphausiidae* and a member of the species *Euphausia superba* (Board, 2015), which is the dominant herbivore of the Southern Ocean. It is a small, swimming crustacean that lives in large schools, called swarms, which reach densities of $10000 - 30000$ individual animals per cubic metre (see Hamner et al., 1983).

The Euphausia glacialis dataset is extracted from *biodiversity.aq* (Van de Putte et al., 2015), which is the Antarctic biodiversity information system, which
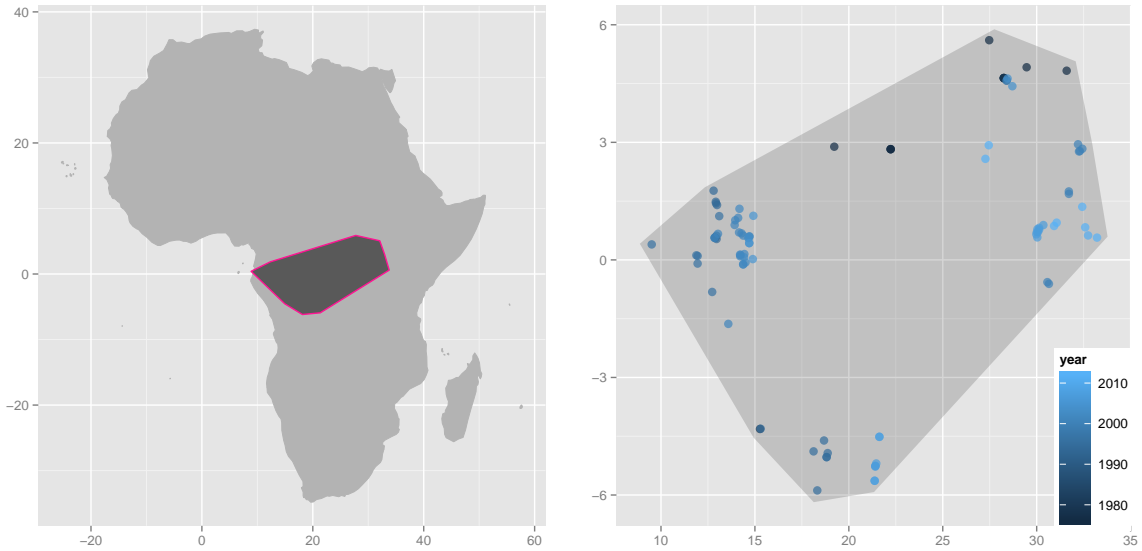
Figure 3.1 Locations of 96 past Ebola outbreaks occurred in a large region of Africa during the years $1976 - 2012$, the region of interest is formed by the smallest polygon containing all locations (left panel), here time is treated as a quantitative mark; dark dots correspond to the oldest events and light dots correspond to the most recent outbreaks (right panel).

gives access to a distributed network of contributing datasets, according to the principles of the *Global Biodiversity Information Facility* (GBIF). It is an international open data infrastructure, funded by governments and available at http://www.gbif.org/.

We have the locations, given in geographic latitude and longitude coordinates, of 57006 swarms reported between 1980 and 2008, year by year. Each location has its associated year as a third coordinate. Due to the inherent computing problems when working with such a large number of locations, we apply a *completely random thinning* with probability $p = 0.13$ of retention over the whole set of spatio-temporal points (see Baddeley et al., 2015), reducing it to 7263 sample points, see Figure 3.2.

3.2.3   Tornadoes in South Carolina

This dataset is provided by the Storm Prediction Center (SPC, http://www.spc.noaa.gov) of the National Oceanic and Atmospheric Administration (NOAA). It contains information of all starting locations of tornado records with the year
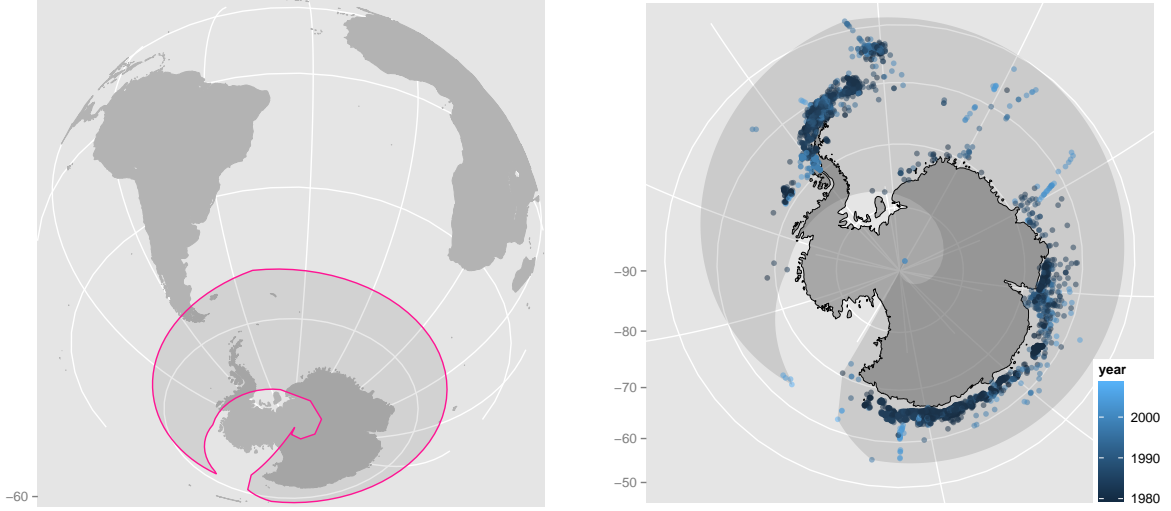
Figure 3.2 Left panel: sampling area of swarms of Euphausia glacialis in the Antarctic marine environment. Right panel: orthogonal projection of locations corresponding to a sample of 7263 swarms enclosed by the smallest polygon containing all the events taken between $1980 - 2008$; the darker points correspond to older records.

of its respective occurrence along fifty nine years $(1953 - 2012)$ occurring in the region of South Carolina. The number of records is 890 and the pattern associated is displayed in Figure 3.3.

3.3 Fundamentals of spatio-temporal point processes

Throughout the paper we assume every subset of $\mathbb{R}^2$ and $\mathbb{R}$ to be a Borel set and every function is assumed to be measurable. We also assume that $W \subseteq \mathbb{R}^2$ and that $T \subseteq \mathbb{R}$ is an interval and consider as spatio-temporal point pattern data a collection of points $\{\xi_i\}_{i=1}^n = \{(\mathbf{u}_i, v_i)\}_{i=1}^n \subseteq W \times T$. Formally, a spatio-temporal point process X is a random countable subset of $\mathbb{R}^2 \times \mathbb{R}$, for which $|X \cap (A \times B)| < \infty$ for bounded $A \times B \subseteq \mathbb{R}^2 \times \mathbb{R}$ (Hereinafter we use $|\cdot|$ to denote both cardinality of a set and absolute value of a real number). Although we here assume that T is an interval in $\mathbb{R}$, it is still possible to construct everything considered in this paper for, say, $T \subseteq \mathbb{Z}$. However, we could then instead treat all event times t_i as marks
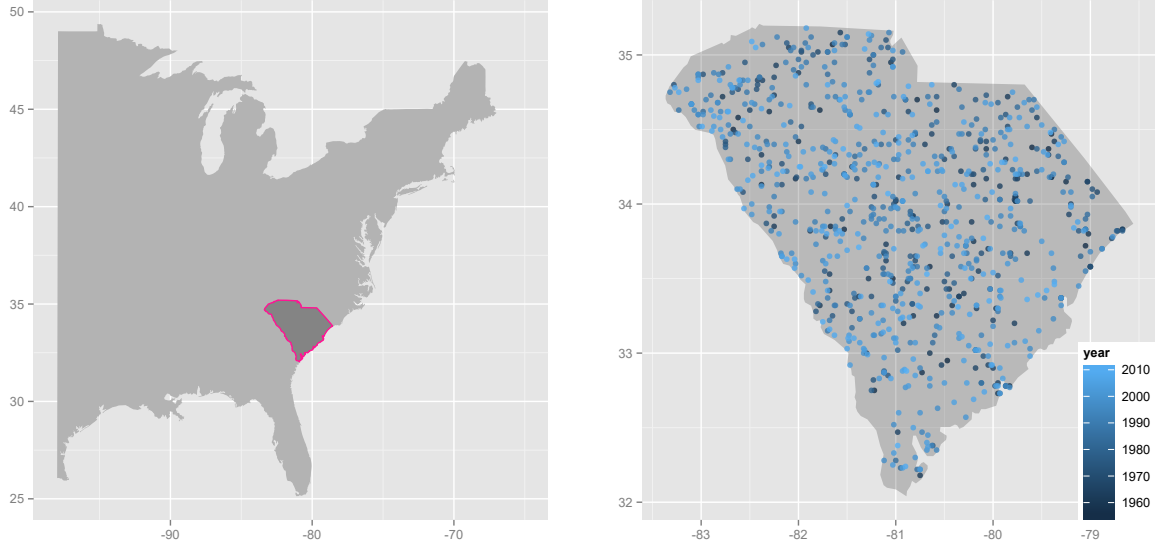
Figure 3.3 Left panel: South Carolina state within U.S. Right panel: starting locations of tornado occurrences reports over South Carolina between $1953-2012$; the darker points correspond to older records.

of a purely spatial point pattern/process with locations $\mathbf{u}_i$ (Daley and Vere-Jones, 2007; Vere-Jones, 2009). Recall that in our datasets we do not have the exact times but only the times with one year accuracy.

From a practical point of view, we treat a spatio-temporal point pattern observed in $W \times T$ either directly as the realisation of a spatio-temporal point process in $W \times T$ or as the realisation of the restriction of a spatio-temporal point process in $\mathbb{R}^2 \times \mathbb{R}$ to $W \times T$. Note that, depending on the modelling assumptions, one is better suited than the other.

The *cylindrical neighbourhood* $B[(\mathbf{u}, v), r, t]$, centred at $(\mathbf{u}, v) \in W \times T$ with spatial radius $r > 0$ and temporal radius $t > 0$, is defined as

$$B[(\mathbf{u}, v), r, t] = B[\mathbf{u}, r] \times [v - t, v + t] = \{(\mathbf{a}, b) \in W \times T : \|\mathbf{u} - \mathbf{a}\| \le r, |v - b| \le t\}, \quad (3.1)$$

where $B[\mathbf{u}, r] = \{\mathbf{a} \in W : \|\mathbf{u} - \mathbf{a}\| \le r\}$ is the Euclidean ball, centred at $\mathbf{u} \in W$, with radius r. We set $B_{rt} = B[(\mathbf{0}, 0), r, t]$. Hereby, given our way of measuring spatio-temporal distances (Cronie and van Lieshout, 2015; Møller and Ghorbani, 2012), the closed ball $B[(\mathbf{u}, v), r]$ of radius $r > 0$, which is centred at $(\mathbf{u}, v) \in W \times T$, is precisely the cylindrical neighbourhood $B[(\mathbf{u}, v), r, r]$.

Henceforth $N(A \times B)$ denotes the number of points of a set $(A \times B) \cap X$, where $A \subseteq W$ and $B \subseteq T$. As usual (Daley and Vere-Jones, 2003), when $N(W \times T) < \infty$ with probability one, which holds e.g. if X is defined on a bounded set, we call X a *finite* spatio-temporal point process.

Next, we turn to the distribution that X induces on the space of point configurations. In particular, we consider different forms of stationarity and isotropy for a spatio-temporal point process. In the case of isotropy, we note that some care has to be taken since $W \times T$ is non-Euclidean.

Definition 1 *Let X be a spatio-temporal point process on $W \times T \subseteq \mathbb{R}^2 \times \mathbb{R}$. X is called* (spatio-temporally) stationary *if the shifted counterpart process $(\mathbf{u}, v) + X$ has the same distribution as the original process X for any $(\mathbf{u}, v) \in W \times T$. We say that X is* (spatially) isotropic *if, for any rotation $\mathbf{r}$ around the origin, the rotated point process $\mathbf{r}X = \{(\mathbf{r}\mathbf{u}, v) : (\mathbf{u}, v) \in X\}$ has the same distribution as X.*

Note that one can define explicit spatial stationarity or temporal stationarity by assuming, respectively, that the definition of stationarity holds only for translations $(\mathbf{u}, 0) + X$, $\mathbf{u} \in \mathbb{R}^2$, or $(\mathbf{0}, v) + X$, $v \in \mathbb{R}$.

When X is a finite spatio-temporal point process or taken as the restriction of a spatio-temporal point process to $W \times T$, it may be natural, at times, to project X onto W and T, and thus deal with the space and time components of X separately. Following Møller and Ghorbani (2012), let

$$X_{\text{space}} = \{\mathbf{u} : (\mathbf{u}, v) \in X, v \in T\}, \quad X_{\text{time}} = \{v : (\mathbf{u}, v) \in X, \mathbf{u} \in W\}.$$

Note that these projections are not well defined unless we have a finite total number of points. The rotated version of X_{space} will be denoted by $\mathbf{r}X_{\text{space}}$.

3.4 Characteristics of spatio-temporal point processes

Having established the set-up of spatio-temporal point processes, we next turn to some point process characteristics that we need in order to define, for instance, the different summary statistics considered in this paper. Below we define product densities, k-order correlation functions, Palm distributions and Papangelou conditional intensities for spatio-temporal point processes.

3.4.1   Product densities

Arguably, the main tools in the statistical analysis of point processes are the *product densities* $\lambda^{(k)}$, $k \geq 1$. $\lambda^{(k)}$ may be defined through the so-called *Campbell theorem* (see Daley and Vere-Jones, 2007, p. 268), which states that, given a spatio-temporal point process X, for any non-negative function h on $(\mathbb{R}^2 \times \mathbb{R})^k$,

$$\mathbb{E}\left[\sum_{\xi_1,\dots,\xi_k \in X}^{\neq} h(\xi_1,\dots,\xi_k) \right] = \int_{\mathbb{R}^2 \times \mathbb{R}} \cdots \int_{\mathbb{R}^2 \times \mathbb{R}} h(\xi_1,\dots,\xi_k) \lambda^{(k)}(\xi_1,\dots,\xi_k) \prod_{i=1}^{k} \mathrm{d}\xi_i,$$

where the left hand side is infinite if and only if the right hand side is. This constitutes an essential result in spatio-temporal point process theory. Here $\sum^{\neq}$ indicates that the summation is taken over distinct k-tuples of spatio-temporal events i.e. points of X.

Since we assume that the point process is simple and that the product density $\lambda^{(k)}$ exists and is finite, then

$$\mathbb{P}(N(\mathrm{d}\xi_1) = 1,\dots,N(\mathrm{d}\xi_k) = 1) = \mathbb{P}(X \cap \mathrm{d}\xi_1 \neq \emptyset,\dots,X \cap \mathrm{d}\xi_k \neq \emptyset)$$

$$= \lambda^{(k)}(\xi_1,\dots,\xi_k) \prod_{i=1}^{k} \mathrm{d}\xi_i,$$

for infinitesimal spatio-temporal regions $\mathrm{d}\xi_1,\dots,\mathrm{d}\xi_k \subseteq W \times T$ with $\mathrm{d}\xi_i = \mathrm{d}\mathbf{u}_i \times \mathrm{d}v_i$ and size $|\mathrm{d}\xi_i| = \mathrm{d}\mathbf{u}_i \mathrm{d}v_i$, $i = 1,\dots,k$. Note that $\mathrm{d}\xi_i$ denotes both the infinitesimal spatio-temporal regions and its Lebesgue measure (Daley and Vere-Jones, 2003). Hence, provided $\lambda^{(k)}$ exists, it governs the infinitesimal k-dimensional joint distributions of the points of X in $W \times T$.

3.4.2   Intensity functions

We next turn to the intensity measure and intensity function, which govern the univariate distributions of the points of X in $W \times T$. The first step of a statistical analysis is usually to estimate and model the intensity function.

Definition 2 *Considering the* intensity measure $\mu(A \times B) = \mathbb{E}[N(A \times B)]$, $A \times B \subseteq W \times T$, *when* $\lambda = \lambda^{(1)}$ *exists, we have that*

$$\mu(A \times B) = \int_A \int_B \lambda(\mathbf{u}, v) \mathrm{d}\mathbf{u} \mathrm{d}v,$$

and we refer to $\lambda(\mathbf{u}, v)$ *as the* intensity function *of* X.

When X is stationary, also referred to as X being *homogeneous* (or completely stationary, see Illian et al., 2008), then $\lambda(\mathbf{u}, v) \equiv \lambda > 0$. This constant is referred to as the *intensity* of X.

There is an alternative heuristic definition of the spatio-temporal intensity (see Diggle, 2013; Diggle et al., 1995), given by

$$\lambda(\mathbf{u}, v) = \lim_{|\mathrm{d}\mathbf{u} \times \mathrm{d}v| \to 0} \frac{\mathbb{E}\left[N(\mathrm{d}\mathbf{u} \times \mathrm{d}v)\right]}{\mathrm{d}\mathbf{u}\mathrm{d}v}.$$

First-order spatio-temporal separability

If the first-order intensity function of a spatio-temporal temporal point process can be factorised (almost everywhere) as

$$\lambda(\mathbf{u}, v) = \lambda_1(\mathbf{u})\lambda_2(v), \tag{3.2}$$

whereby

$$\mu(A \times B) = \int_A \lambda_1(\mathbf{u})\mathrm{d}\mathbf{u} \int_B \lambda_2(v)\mathrm{d}v,$$

where $\lambda_1(\cdot)$ and $\lambda_2(\cdot)$ are non-negative functions, then the process is referred to as *first-order spatio-temporal separability*. Note that these functions are not unique.

If this identity is taken as assumption, it implies that effects that are non-separable could be interpreted as second-order effects, rather than first-order effects (see Gabriel, 2014; Gabriel and Diggle, 2009; Møller and Ghorbani, 2012). We assume this along the paper (unless specific mentioning of non-separability).

A stationary spatio-temporal point process X is automatically first-order separable since its intensity $\lambda = \lambda_1\lambda_2 \geq 0$ is constant. If X is *space-stationary* (recall Definition 1), implying that $\lambda(\mathbf{u}, v)$ depends only on v, it is also first-order separable with λ_1 being a non-negative constant. Similarly, when X is *time-stationary*, implying that $\lambda(\mathbf{u}, v)$ depends only on $\mathbf{u}$, first-order separability holds with λ_2 being a non-negative constant. When we have obtained X_{space} and X_{time}, we may also define the marginal spatial and temporal intensity functions λ_{space} and λ_{time}, respectively, as

$$\lambda_{\mathrm{space}}(\mathbf{u}) = \lambda_1(\mathbf{u}) \int_T \lambda_2(v)\mathrm{d}v \quad \text{and} \quad \lambda_{\mathrm{time}}(v) = \lambda_2(v) \int_W \lambda_1(\mathbf{u})\mathrm{d}\mathbf{u},$$

whereby $\lambda(\mathbf{u}, v) \propto \lambda_{\text{space}}(\mathbf{u})\lambda_{\text{time}}(v)$, with λ, λ_{space}, λ_{time} all being constant when X is homogeneous.

3.4.3   Estimation of first-order intensity functions

When estimating the first-order intensity function we are challenged with the task of finding an estimate $\hat{\lambda} : W \times T \to \mathbb{R}$, taking into account that usually we have one single realisation. Suppose we have obtained unbiased estimators given by $\hat{\lambda}_{\text{space}}(\cdot)$ and $\hat{\lambda}_{\text{time}}(\cdot)$. If we assume separability, the estimator of the spatio-temporal first-order intensity function is given by

$$\hat{\lambda}(\mathbf{u}, v) = \frac{1}{n}\left(\hat{\lambda}_{\text{space}}(\mathbf{u})\,\hat{\lambda}_{\text{time}}(v)\right). \tag{3.3}$$

This also constitutes an unbiased estimator of the expected number of points. For non-parametric estimation of the spatial intensity function, it is common to follow Diggle (1985), Berman and Diggle (1989) and Choi and Hall (1999) in using a kernel estimate,

$$\hat{\lambda}_{\text{space}}(\mathbf{u}) = \sum_{i=1}^{n} \frac{\kappa_{\varepsilon}(\mathbf{u} - \mathbf{u}_i)}{c_{W\varepsilon}(\mathbf{u}_i)}, \qquad \mathbf{u} \in W. \tag{3.4}$$

Here

$$\kappa_{\varepsilon}(\mathbf{u}) = \frac{1}{\varepsilon^2}\kappa\left(\frac{\mathbf{u}}{\varepsilon}\right),$$

where $\kappa(\cdot)$ is a bivariate kernel and $\varepsilon > 0$ is the *bandwidth*, a smoothing parameter, and

$$c_{W\varepsilon}(\mathbf{u}_i) = \int_W \kappa_{\varepsilon}(\mathbf{u} - \mathbf{u}_i)\,\mathrm{d}\mathbf{u}$$

is an edge-correction factor included in the estimation to guarantee that

$$\int_W \hat{\lambda}_{\text{space}}(\mathbf{u})\,\mathrm{d}\mathbf{u} = n$$

(Diggle, 1985; Ghorbani, 2013; van Lieshout, 2011). Similarly, we may also estimate $\lambda_{\text{time}}(v)$ non-parametrically by means of kernel estimators (Gabriel et al., 2013). Note that the specification of the bandwidth of the kernel is debatable. It should be emphasised that large values of spatial or temporal bandwidths produce very smooth estimates, whereas very small values produce noisy and unrealistic estimates. This is the most problematic aspect of the estimation (see Baddeley et al., 2000; Illian et al., 2008). Although these non-parametric

estimators may only lead to approximately unbiased estimates, we will still
employ Eq.(3.3) for the estimation of $\lambda(\mathbf{u}, v)$.

In the literature one finds that it is common to model both the spatial and the
temporal intensity components by using either a kernel estimators or parametric
methods, which in the latter case gives rise to a semi-parametric estimation
approach for $\lambda(\mathbf{u}, v)$. For example, Gabriel and Diggle (2009) estimate the tem-
poral intensity using a log-linear regression model to analyse the pattern of
cases of human *Campylobacter jejuni* infections reported in Preston, Lancashire,
UK over three years. Incidence of such infections is sporadic, with a seasonal
variation which rises in spring and peaks in summer. This suggests an annual
and four-monthly periodicity, so they fitted a harmonic linear model with one-year
period

$$\log \lambda_{\text{time}}(v) = \delta_{d(v)} + \sum_{k=1}^{3} \left(\alpha_k \cos(k\omega v) + \beta_k \sin(k\omega v) \right) + \varepsilon v,$$

where $\omega = 2\pi/365$, ε denotes the trend and $d(v)$ identifies the day of the week for
day $v = 1, \ldots, 1096$. See also Diggle (2013) and Tamayo-Uria et al. (2014).

Estimation of first-order intensities for the considered examples

Following what has been proposed in the literature, we restrict the analysis in this
section to the case of first-order separability. We show the non-parametric estim-
ation of the spatial and temporal intensities of the Ebola outbreaks described in
Section 3.2.1 by means of kernel smoothing (Diggle, 2013 and references therein).
A two-dimensional kernel intensity estimator, with bandwidth $\varepsilon = (2.64, 0.45)$
degrees (remember we are dealing with data given in long-lat coordinates here
and later on), with an axis-aligned bivariate Gaussian kernel, evaluated on a
square grid, is used for the spatial intensity. Also, a one-dimensional kernel
intensity estimator, with bandwidth $\delta = 3.13$ years, with a univariate Gaussian
kernel is used for the temporal intensity. To choose the bandwidth of the Gaus-
sian kernel intensity estimator we use a *rule of thumb*, which defaults to 0.9
times the minimum of the standard deviation and the interquartile range divided
by 1.34 times the sample size to the negative one-fifth power (Silverman, 1986,
p. 48). Figure 3.4 illustrates the estimates of the spatial (left) and temporal
(right) densities, understood as the intensities normalised by the total number
of cases. We can see in Figure 3.4 (left), three isolated areas, each one with a
high intensity. For example, one of these areas corresponds to the region around
Gabon and the Democratic Republic of Congo; here the probability of infection is

higher. Looking at Figure 3.4 (right) we note high rates of infection around the year 2000 indicating a clear degree of temporal inhomogeneity.
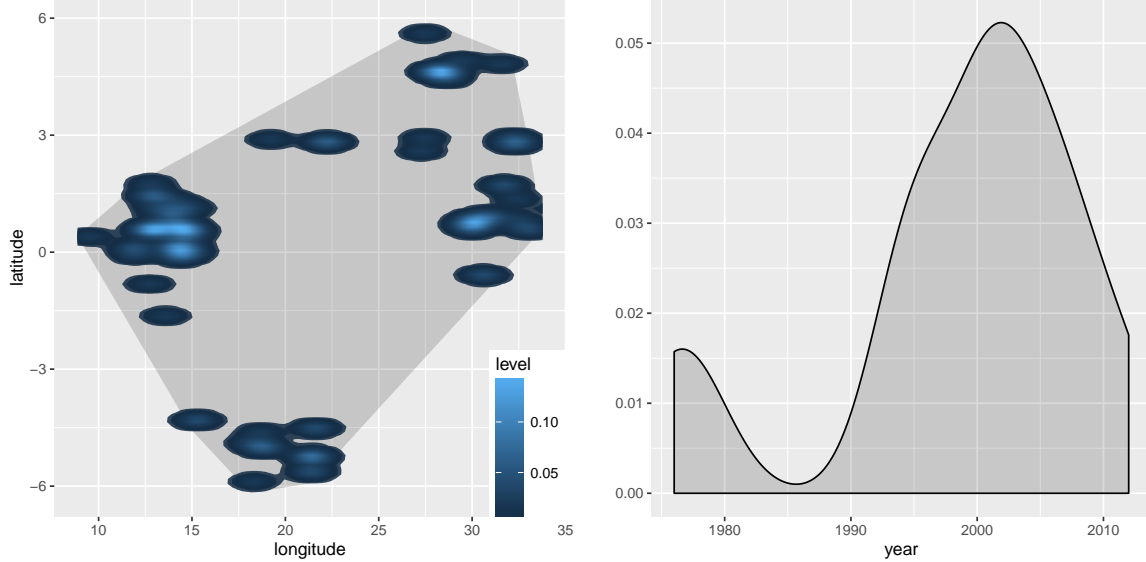
Figure 3.4 Spatial (left) and temporal (right) kernel-based estimates of the normalised intensity functions for the Ebola outbreaks data.

For the estimation of the intensity functions of the Euphausia glacialis dataset described in Section 3.2.2, we again use a two-dimensional kernel with an axis-aligned bivariate Gaussian kernel for the spatial intensity and a bandwidth of $\varepsilon = (13.18, 0.43)$ degrees. For the temporal intensity we use a one-dimensional Gaussian kernel for the temporal intensity, with a bandwidth of $\delta = 0.8$ years. Both intensities (normalised by the total number of cases) are displayed in Figure 3.5. In Figure 3.5 (left), a fairly high intensity in the region of Antarctica is observed near the American continental southern part, i.e., on the Antarctic Peninsula, specifically in Shetland and over the Weddell Sea. Notice an increase in the number of swarms in the region closest to Oceania. This may be due to the geography of the ocean floor, in terms of there being better conditions close to the surrounding continents. In Figure 3.5 (right), there is a rapid decrease in the last fifteen years that may be due to many factors, including environmental ones and human intervention.

The corresponding estimated intensity functions for the Tornado dataset described in Section 3.2.3, in which we set a spatial bandwidth of $\varepsilon = (0.24, 0.16)$ degrees and a temporal bandwidth of $\delta = 3.70$ years, are depicted in Figure 3.6. In this case we note that there is small variation throughout the entire spatial
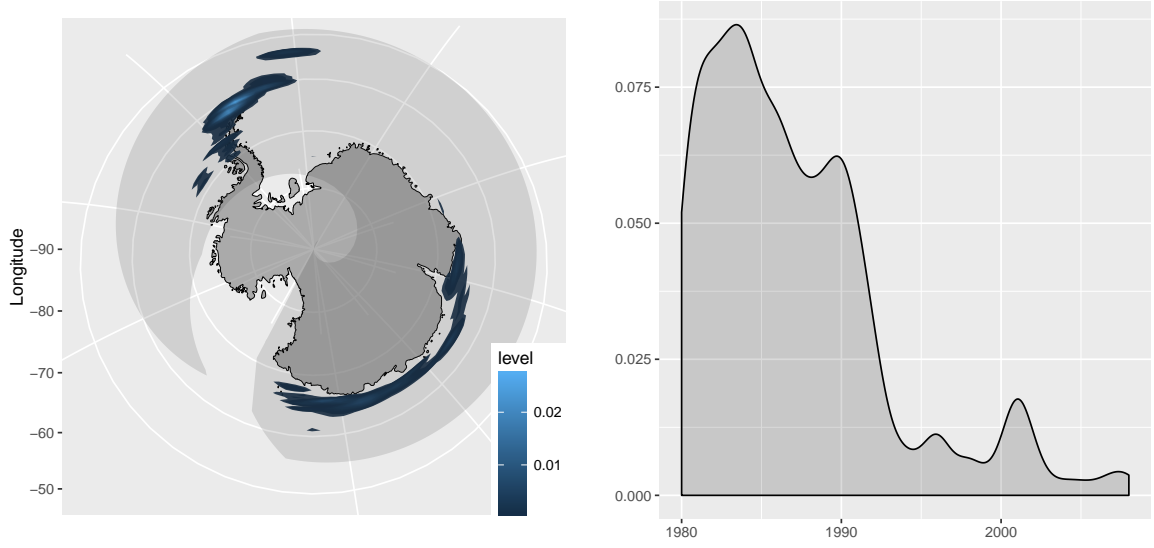
Figure 3.5 Spatial (left) and temporal (right) kernel-based estimation of the normalised intensity functions for the Euphausia glacialis data.

region, meaning that the geography of the state of South Carolina is probably not very complicated and tornadoes can occur almost anywhere with an almost constant probability. This is not so in the temporal domain, where we note rapid growth in the last twenty-five years, likely due to various climatic factors (including global warming). Another possibility is that the measurements have improved during this period, making possible the reporting of more cases.

3.4.4   Conditional intensity functions

Assume next explicitly that X is a spatio-temporal point process on $W \times T \subseteq \mathbb{R}^2 \times [0, \infty)$, such that X_{time} is well defined. X may be treated as a temporal point process with corresponding marks X_{space} and we may define the cumulative process $X_{\text{time}}(t) := |X_{\text{time}} \cap [0,t]|$, $t \in T$. This is the classical approach for spatio-temporal point processes, see for example Cox and Isham (1980).

The *conditional intensity function* $\lambda^*(\mathbf{u}, v | \mathscr{H}_v)$ of a spatio-temporal point process is the expected rate that points occur around the spatio-temporal location $(\mathbf{u}, v)$, conditionally on the *history* $\mathscr{H}_v$, $v \in T$, consisting of the set of locations and times of all events of the process that occur prior to time v. In other words, $\mathscr{H}_v$ is the family of σ-algebras generated by the events occurring at times up to, but not including v. Following e.g. Daley and Vere-Jones (2003), Diggle (2013) and
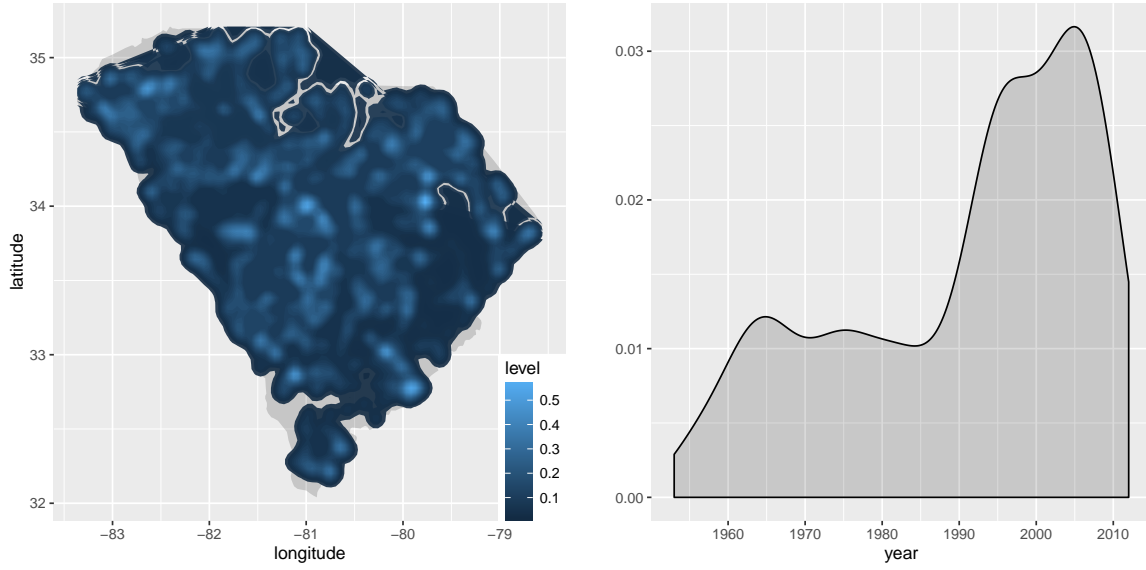
Figure 3.6 Spatial (left) and temporal (right) kernel-based estimation of normalised intensity functions for the tornado in South-Carolina data.

Møller et al. (2016), we have that

$$\lambda^*(\mathbf{u},v|\mathscr{H}_v)\,\mathrm{d}\mathbf{u}\mathrm{d}v = \mathbb{E}\left[N(\mathrm{d}\mathbf{u}\times\mathrm{d}v)|\mathscr{H}_v\right], \qquad (\mathbf{u},v)\in\mathrm{d}\mathbf{u}\times\mathrm{d}v\subseteq W\times T.$$

We are assuming that the underlying process X is *orderly* in the sense of Diggle, Guan, Hart, Paize and Stanton (2010); Diggle, Kaimi and Abellana (2010), which means that the probability of observing more than one point in a time interval is decreasing with the order of the size of the the interval. Denoting by $l+V$ and $\mathbf{U}$ the time and location, respectively, of the first event that occurs after time l, it follows that

$$\mathbb{P}(V > v) = \exp\left\{-\int_l^{l+v}\int_W \lambda^*(\mathbf{u},t|\mathscr{H}_l)\,\mathrm{d}\mathbf{u}\mathrm{d}t\right\},$$

the conditional probability density of $\mathbf{U}$ given $V = v$ is proportional to the conditional intensity, $\lambda^*(\mathbf{u},l+v|\mathscr{H}_{l+v})$, $\mathbf{u}\in W$.

3.4.5   Papangelou conditional intensities

Analogously to the purely spatial context, the Papangelou conditional intensity $\lambda^\dagger(\mathbf{u},v|X)$, $(\mathbf{u},v)\in W\times T$, of a spatio-temporal point process may be defined through the reduced Campbell-Mecke formula (see Cronie and van Lieshout,

2015). Heuristically,

$$\lambda^\dagger(\mathbf{u},v|X)\mathrm{d}\mathbf{u}\mathrm{d}v = \mathbb{P}\left[N(\mathrm{d}\mathbf{u}\times\mathrm{d}v)=1\,\middle|\,(X\setminus\mathrm{d}\mathbf{u}\times\mathrm{d}v)\right],$$

i.e. $\lambda^\dagger(\mathbf{u},v|X)\mathrm{d}\mathbf{u}\mathrm{d}v$ gives the conditional probability of finding a point of X in the infinitesimal spatio-temporal region $\mathrm{d}\mathbf{u}\times\mathrm{d}v$, with size $\mathrm{d}\mathbf{u}\mathrm{d}v$, given the process outside $\mathrm{d}\mathbf{u}\times\mathrm{d}v$.

3.4.6 Edge-correction

Edge-correction methods/factors have been widely studied in the spatial case (see e.g. Baddeley et al., 2000; Ripley, 1988; Stoyan and Stoyan, 1994 and Illian et al., 2008). There are some approaches dealing with three-dimensional data (Baddeley et al., 1993; Jafari-Mamaghani et al., 2010), with spatio-temporal data (Gabriel, 2014) and with marked and spatio-temporal data (Cronie and Särkkä, 2011). In particular, Gabriel (2014) extends three classical spatial edge-correction factors to the spatio-temporal context and compares the performance of the related estimators of several second-order characteristics for stationary/non-stationary and/or isotropic/anisotropic spatio-temporal point processes. It is common to consider correcting edge-effects separately (Diggle et al., 1995), thus, the edge-correction factor is the product of a spatial and a temporal edge-correction factor. The behaviour of the edge-correction should depend on how well the underlying assumptions hold. However, it is known that different edge-correction methods provide similar results when they are used in estimation procedures, in particular the isotropic edge-correction is one of the most widely used in practice (see e.g. Gabriel, 2014). In general, we denote an edge-correction factor by a weight w_{ij}, where i and j represent two different points of the pattern.

Isotropic correction

In this case, the weight is proportional to the product between the Ripley edge-correction factor (see Ripley, 1977) for the spatial region, and its one-dimensional analogue, giving

$$w_{ij} = |W\times T|w_{ij}^{(\mathbf{u})}w_{ij}^{(v)}.$$

Here $w_{ij}^{(\mathbf{u})}$ is the proportion of the circumference of a circle centred at the location $\mathbf{u}_i$ with radius $||\mathbf{u}_i-\mathbf{u}_j||$ that lies within W. The temporal edge-correction factor

$w_{ij}^{(v)} = 1$ if both ends of the interval of length $2|v_i - v_j|$ that is centred at v_i lie within T, and $w_{ij}^{(v)} = 1/2$ otherwise (Gabriel and Diggle, 2009).

Border method

Let $W_{\ominus r} = \{\mathbf{u} \in W : B[\mathbf{u}, r] \subseteq W\}$ and $T_{\ominus t} = \{v \in T : B[v, t] \subseteq T\}$, be eroded spatial and temporal regions, obtained by trimming off a margin of width $r \geq 0$ and $t \geq 0$ from the borders of W and T, respectively. Note that $W_{\ominus r} \times T_{\ominus t}$ may be visualised by taking the flat trimmed region $W_{\ominus r} \times \{0\}$ and stretching it in the t-dimension until its height reaches $|T| - 2t$. This method restricts attention to those events lying more than r units away from the boundary of W (see Diggle, 1979) and more than t units away from the boundary of T. For the border method we have that

$$w_{ij} = \frac{\sum_{j=1}^{n} \mathbf{1}\{(\mathbf{u}_j, v_j) \in W_{\ominus r} \times T_{\ominus t}\}/\lambda(\mathbf{u}_j, v_j)}{\mathbf{1}\{(\mathbf{u}_i, v_i) \in W_{\ominus r} \times T_{\ominus t}\}}, \quad r, t \geq 0.$$

Modified border method

As an extension of the method proposed by Baddeley and Turner (2000), Gabriel (2014) gives a spatio-temporal version of the border method by considering

$$w_{ij} = \frac{|W_{\ominus r}||T_{\ominus t}|}{\mathbf{1}\{(\mathbf{u}_i, v_i) \in W_{\ominus r} \times T_{\ominus t}\}}, \quad r, t \geq 0.$$

This edge-correction considers the eroded domain instead of the whole domain to avoid the edge-effect of the general term $|W \times T|$.

Translation correction

Ohser and Stoyan (1981) proposed a correction based on the proportion of translation of regions. Gabriel (2014) defined the weights for the spatio-temporal case, which are given by the proportion of translations of $((\mathbf{u}_i, v_i), (\mathbf{u}_j, v_j))$ which have both $(\mathbf{u}_i, v_i)$ and $(\mathbf{u}_j, v_j)$ inside $W \times T$. More specifically,

$$w_{ij} = |W \cap W_{\mathbf{u}_i - \mathbf{u}_j}||T \cap T_{v_i - v_j}|,$$

where $W_{\mathbf{u}_i - \mathbf{u}_j}$ and $T_{v_i - v_j}$ are, respectively, the translated spatial and temporal regions along the vectors $\mathbf{u}_i - \mathbf{u}_j$ and $v_i - v_j$.

3.4.7 The pair correlation function

Turning to measures of second-order spatio-temporal interaction, in particular in presence of inhomogeneity, the *pair correlation function* (Gabriel and Diggle, 2009; Illian et al., 2008; Møller and Ghorbani, 2012) is defined as

$$g(\xi_1, \xi_2) = \frac{\lambda^{(2)}(\xi_1, \xi_2)}{\lambda(\xi_1)\lambda(\xi_2)}, \qquad \xi_1, \xi_2 \in W \times T. \tag{3.5}$$

For a spatio-temporal Poisson process (i.e. a completely random process, see Sections 3.5.1 and 3.5.2 for details), the pair correlation function is identically 1. Hence, larger or smaller values than this benchmark indicate, informally, how much more or less likely it is that a pair of events will occur at the specified locations, than in a Poisson process with the same intensity function.

Similarly to the case of first-order separability (equation (3.2)), the pair correlation function is said to be *separable* (Møller and Ghorbani, 2012) if

$$g((\mathbf{u}, v), (\mathbf{s}, l)) = g_1(\mathbf{u}, \mathbf{s})g_2(v, l), \tag{3.6}$$

where g_1 and g_2 are non-negative functions. Before turning to the different summary statistics used to quantify interactions in inhomogeneous spatio-temporal point processes, we first consider so-called *second-order intensity-reweighted stationarity*, which has to be imposed when we consider some of the different inhomogeneous summary statistics. Following Baddeley et al. (2000), Gabriel and Diggle (2009) and Cronie and van Lieshout (2015), we have the following definition.

Definition 3 *A spatio-temporal point process X is second-order intensity-reweighted stationary (SOIRS) if*

$$g((\mathbf{u}, v), (\mathbf{s}, l)) = \bar{g}(\mathbf{u} - \mathbf{s}, v - l),$$

for any $(\mathbf{u}, v), (\mathbf{s}, l) \in W \times T$, *where* $\bar{g}$ *is some non-negative function. If the process is also isotropic, then* $\bar{g}(\mathbf{u} - \mathbf{s}, v - l) = g_0(r, t)$, *i.e.* $g(\cdot, \cdot)$ *depends only on the distances* $r = \|\mathbf{u} - \mathbf{s}\|$ *and* $t = |v - l|$, *where* g_0 *is some non-negative function.*

The pair correlation function can be extended to general orders $k \geq 2$ by defining

$$g^{(k)}(\xi_1, \ldots, \xi_k) = \frac{\lambda^{(k)}(\xi_1, \ldots, \xi_k)}{\prod_{i=1}^{k} \lambda(\xi_i)}, \qquad \xi_1, \ldots, \xi_k \in W \times T.$$

Estimation of the pair correlation function

Non-parametric estimation of pair correlation functions is usually based on kernel methods (see Diggle, 2013; Gabriel and Diggle, 2009; Stoyan and Stoyan, 1994).

Given $X = \{(\mathbf{u}_i, v_i)\}_{i=1}^n$, the spatio-temporal pair correlation function defined in (4.1) can be estimated by

$$\hat{g}(r,t) = \frac{1}{4\pi r} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \frac{\kappa_{1\varepsilon}(\|\mathbf{u}_i - \mathbf{u}_j\| - r)\kappa_{2\delta}(|v_i - v_j| - t)}{\hat{\lambda}(\mathbf{u}_i, v_i)\hat{\lambda}(\mathbf{u}_j, v_j)w_{ij}}, \quad r > \varepsilon, t > \delta, \qquad (3.7)$$

where $\kappa_{1\varepsilon}$ and $\kappa_{2\delta}$ are one-dimensional kernel functions with spatial and temporal bandwidths ε and δ, respectively. For details see Cressie and Collins (2001); Gabriel (2014); Gabriel et al. (2013) and Rodríguez-Cortés et al. (2014). As in Section 3.4.6, w_{ij} are edge-correction factors, which correct for the loss of information regarding the interaction occurring between points close to the border of $W \times T$ and those (unobserved) ones outside.

3.4.8   The spatio-temporal K-function

Continuing the quantification of interactions between pairs of events, we may also consider the inhomogeneous spatio-temporal K-function defined in Gabriel and Diggle (2009) and Møller and Ghorbani (2012). Let X be a SOIRS (according to Definition 3) spatio-temporal point process. Then define

$$K_{\text{inhom}}(r,t) = \int_{\mathbb{R}^2} \int_{\mathbb{R}} \mathbf{1}\{(\mathbf{u},v) \in B_{rt}\}\,\bar{g}(\mathbf{u},v)\mathrm{d}\mathbf{u}\mathrm{d}v, \qquad (3.8)$$

for $r > 0$ and $t > 0$. Note that in (3.8), for convenience we write $\mathbf{u} = \mathbf{u} - \mathbf{s}$ and $v = v - l$. This form was suggested by Møller and Ghorbani (2012). Note that, as usual, the K-function can be defined based on the Palm distribution, expectations and first-order intensities (Baddeley and Turner, 2000). There also exists a weighted version of K_{inhom} where the first-order intensity λ is replaced by the conditional intensity function λ^*, as described in Veen and Schoenberg (2006) and Adelfio and Schoenberg (2009); this version is only considered for planar point processes and has not been extended to the spatio-temporal context so far.

Assuming further that the process is isotropic, we obtain the original definition of the spatio-temporal inhomogeneous K-function, which was given in Gabriel

and Diggle (2009),

$$K_{\mathrm{inhom}}(r,t) = 2\pi \int_0^r \int_{-t}^t s\,g_0(s,l)\mathrm{d}s\mathrm{d}l.$$

where g_0 is taken as in Definition 3. Note that Gabriel and Diggle (2009) originally suggested a different form, which only takes the present and future events into account, but its value only differs from the present form by a factor of $1/2$. For a spatio-temporal Poisson process (see Sections 3.5.1 and 3.5.2 for details), $K(r,t) = \pi r^2 t$, so that $K(r,t) - \pi r^2 t$ can be used as a measure of spatio-temporal aggregation or regularity (aggregation and regularity refer to variations in the density of points that cannot be explained by inhomogeneity alone), using the Poisson process as a reference. Following Møller and Ghorbani (2012), when X_{space} and X_{time} are defined, under the assumption of separability we can write the spatial and temporal components of the K-function as

$$K_{\mathrm{space}}(r) = \int_{\|u\|\leq r} g_{\mathrm{space}}(\mathbf{u})\,\mathrm{d}\mathbf{u} \quad\text{and}\quad K_{\mathrm{time}}(t) = \int_{-t}^t g_{\mathrm{time}}(v)\,\mathrm{d}v, \quad r,t > 0,$$

where

$$
\begin{aligned}
g_{\mathrm{space}}(\mathbf{u},\mathbf{s}) &= g_{\mathrm{space}}(\mathbf{u}-\mathbf{s}) = \int_T \int_T p_2(v)p_2(l)g(\mathbf{u}-\mathbf{s},v-l)\mathrm{d}v\mathrm{d}l, \\
g_{\mathrm{time}}(v,l) &= g_{\mathrm{time}}(v-l) = \int_W \int_W p_1(\mathbf{u})p_1(\mathbf{s})g(\mathbf{u}-\mathbf{s},v-l)\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{s},
\end{aligned}
$$

with

$$p_1(\mathbf{u}) = \frac{\lambda_1(\mathbf{u})}{\int_W \lambda_1(\mathbf{u})\mathrm{d}\mathbf{u}} \quad\text{and}\quad p_2(v) = \frac{\lambda_2(v)}{\int_T \lambda_2(v)\mathrm{d}v}.$$

Let $\{\mathbb{P}^{!(\mathbf{u},v)}(X \in \cdot) : (\mathbf{u},v) \in \mathbb{R}^2 \times \mathbb{R}\}$ be the family of reduced Palm distributions of X (see Cronie and van Lieshout 2015; Daley and Vere-Jones 2003, 2007); $\mathbb{P}^{!(\mathbf{u},v)}(X \in \cdot)$ may be interpreted as the distribution of X on the (rather abstract) space of point configurations, conditional on there being a point $(\mathbf{u},v) \in X$, which we neglect. Let $\mathbb{E}^{!(\mathbf{u},v)}[\cdot]$ denote expectation under $\mathbb{P}^{!(\mathbf{u},v)}(\cdot)$.

In the stationary case $K(r,t) = K_{\mathrm{inhom}}(r,t)$, i.e. the spatio-temporal version of Ripley's K-function (Ripley, 1977), defined in Diggle et al. (1995). $\lambda K(r,t) = \mathbb{E}^{!(\mathbf{0},0)}[N(B_{rt})]$ is simply the expected number of further points within distance r and time lag t from the origin, given that X has a point at the origin.

Note that when X is a SOIRS, the pair correlation function is proportional to the derivative of $K(r,t)$ with respect to r and t (see Rodríguez-Cortés et al. (2014)),

i.e.

$$g(r,t) = \frac{1}{4\pi r} \frac{\partial^2 K(r,t)}{\partial r \partial t}, \qquad r > 0, \quad t > 0. \tag{3.9}$$

Estimation of $K_{\text{inhom}}(r,t)$

Letting $X = \{(\mathbf{u}_i, v_i)\}_{i=1}^n$, a general estimator of $K_{\text{inhom}}(r,t)$ is given by

$$\hat{K}(r,t) = \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \frac{\mathbf{1}\left[\|\mathbf{u}_i - \mathbf{u}_j\| \leq r\right] \mathbf{1}\left[|v_i - v_j| \leq t\right]}{\hat{\lambda}(\mathbf{u}_i, v_i)\,\hat{\lambda}(\mathbf{u}_j, v_j)\,w_{ij}}. \tag{3.10}$$

Following common practice in the literature of spatio-temporal point processes, if we assume that w_{ij} is Ripley's spatial edge-correction factor, we obtain an approximately unbiased non-parametric estimator of $K_{\text{inhom}}(r,t)$ (Gabriel, 2014). We also have an alternative estimator, which matches the original definition of Gabriel and Diggle (2009) and does not take into account the past of the process. It is given by

$$\hat{K}^*(r,t) = \frac{n}{n_t} \sum_{i=1}^{n_t} \sum_{j>i} \frac{\mathbf{1}\left[\|\mathbf{u}_i - \mathbf{u}_j\| \leq r\right] \mathbf{1}\left[v_j - v_i \leq t\right]}{\hat{\lambda}(\mathbf{u}_i, v_i)\,\hat{\lambda}(\mathbf{u}_j, v_j)\,w_{ij}},$$

where n_t is the number of events $v_i \leq b - t$ whenever $T = [a,b] \subseteq \mathbb{R}_+$.

3.4.9 Spatio-temporal nearest neighbour distance distribution and empty- space functions

Second-order summary statistics should be applied mainly when one knows/believes that higher-order interactions do not exist. Or simply if one wants to quantify second-order effects explicitly. Hence, to get a general quantification of higher-order interactions as well, one should not limit oneself to finite-order interaction summary statistics (Cronie and van Lieshout, 2015). When this is the case, just as in the SOIRS case we need to consider some form of reweighted stationarity. It turns out that the assumption of *intensity-reweighted moment stationarity (IRMS)* is what is needed to be imposed on X. Note that for the purpose of clarity of exposition, we choose to define things a bit differently than originally done by Cronie and van Lieshout (2015).

Definition 4 *Given a spatio-temporal point process X on $\mathbb{R}^2 \times \mathbb{R}$, if the intensity is bounded away from 0, i.e. $\underline{\lambda} = \inf_{(\mathbf{u},v)} \lambda(\mathbf{u},v) > 0$, and*

$$g^{(k)}(\xi_1,\ldots,\xi_k) = g^{(k)}(c+\xi_1,\ldots,c+\xi_k)$$

almost everywhere for any $c \in \mathbb{R}^2 \times \mathbb{R}$ and any $k \geq 2$, we say that X is intensity-reweighted moment stationary (IRMS).

Note that stationarity implies IRMS, which in turn implies SOIRS. Throughout this Section we assume that X is IRMS.

Under IRMS, Cronie and van Lieshout (2015) defined the *inhomogeneous spatio-temporal nearest neighbour distance distribution function* as

$$G_{\text{inhom}}(r,t) = 1 - \mathbb{E}^{!(\mathbf{a},b)}\left[\prod_{(\mathbf{u},v)\in X}\left(1 - \frac{\underline{\lambda}\,\mathbf{1}\{(\mathbf{u},v)\in B[(\mathbf{a},b),r,t]\}}{\lambda(\mathbf{u},v)}\right)\right],$$

and the *inhomogeneous spatio-temporal empty space function* as

$$F_{\text{inhom}}(r,t) = 1 - \mathbb{E}\left[\prod_{(\mathbf{u},v)\in X}\left(1 - \frac{\underline{\lambda}\,\mathbf{1}\{(\mathbf{u},v)\in B[(\mathbf{a},b),r,t]\}}{\lambda(\mathbf{u},v)}\right)\right],$$

for $(\mathbf{a},b) \in \mathbb{R}^2 \times \mathbb{R}$ and $r,t \geq 0$, under the convention that empty products take the value one. It turns out that under IRMS the above functions are (almost everywhere) constant with respect to $(\mathbf{a},b)$. Note that when X is stationary, $G_{\text{inhom}}(r,t)$ and $F_{\text{inhom}}(r,t)$ reduce to spatio-temporal versions $G(r,t)$ and $F(r,t)$ of the classical nearest neighbour distance distribution function and the empty-space function, respectively.

3.4.10 The spatio-temporal *J*-function

The *inhomogeneous spatio-temporal J-function* is given as the ratio between $1 - G_{\text{inhom}}(r,t)$ and $1 - F_{\text{inhom}}(r,t)$, i.e.

$$J_{\text{inhom}}(r,t) = \frac{1 - G_{\text{inhom}}(r,t)}{1 - F_{\text{inhom}}(r,t)}, \tag{3.11}$$

for all $r,t \geq 0$ such that $F_{\text{inhom}}(r,t) \neq 1$. By the reduction of $G_{\text{inhom}}(r,t)$ and $F_{\text{inhom}}(r,t)$ to $G(r,t)$ and $F(r,t)$ under stationarity, $J_{\text{inhom}}(r,t)$ reduces to a spatio-temporal extension $J(r,t)$ of the original *J*-function of van Lieshout and Baddeley (1996) under

stationarity. Hence, $J_{\text{inhom}}(r,t)$ is truly an extension of $J(r,t)$ to the inhomogeneous setting.

To see that $K_{\text{inhom}}(r,t)$ is closely related to $J_{\text{inhom}}(r,t)$, it may be shown that

$$J_{\text{inhom}}(r,t) - 1 = \underline{\lambda}(2\pi r^2 t - K_{\text{inhom}}(r,t)) + \beta, \tag{3.12}$$

where β represents the interaction terms of order > 2 (Cronie and van Lieshout, 2015). Hence, for $K_{\text{inhom}}(r,t)$ we neglect all interaction terms of order higher than two.

The intuition behind $J_{\text{inhom}}(r,t)$ and $J(r,t)$ is that we look at whether conditioning on having a point at, say, the origin increases/decreases the probability of finding further points within B_{rt}, $r,t > 0$. Note that for a Poisson process, $J_{\text{inhom}}(r,t) = 1$. Hence, if $J_{\text{inhom}}(r,t) < 1$ we conclude that there is clustering at the spatio-temporal lag pair (r,t). Conversely, $J_{\text{inhom}}(r,t) > 1$ indicates regularity.

Estimation of $J_{\text{inhom}}(r,t)$

Turning next to the estimation of the inhomogeneous spatio-temporal summary statistics above, assume that we observe X on $W \times T$. In order to simplify the expressions, we first define some important quantities. For $r,t \geq 0$, let $L \subseteq W \times T$ be a fine grid, $\Psi_X = X \cap (W_{\ominus r} \times T_{\ominus t})$, $\Psi_L = L \cap (W_{\ominus r} \times T_{\ominus t})$,

$$
\begin{aligned}
\Omega_-^{(\mathbf{s},l)} &= (X \setminus \{(\mathbf{s},l)\}) \cap B[(\mathbf{s},l),r,t], \\
\Omega_+^{(\mathbf{s},l)} &= X \cap B[(\mathbf{s},l),r,t].
\end{aligned}
$$

A minus sampling estimator of $1 - G_{\text{inhom}}(r,t)$ is defined as

$$\frac{1}{|\Psi_X|} \sum_{(\mathbf{s},l) \in \Psi_X} \prod_{(\mathbf{u},v) \in \Omega_-^{(\mathbf{s},l)}} \left(1 - \frac{\hat{\lambda}}{\hat{\lambda}(\mathbf{u},v)}\right), \tag{3.13}$$

and a minus sampling estimator of $1 - F_{\text{inhom}}(r,t)$ as

$$\frac{1}{|\Psi_L|} \sum_{(\mathbf{s},l) \in \Psi_L} \prod_{(\mathbf{u},v) \in \Omega_+^{(\mathbf{s},l)}} \left(1 - \frac{\hat{\lambda}}{\hat{\lambda}(\mathbf{u},v)}\right), \tag{3.14}$$

where $\hat{\underline{\lambda}} = \min_{(\mathbf{u},v)} \hat{\lambda}(\mathbf{u},v)$. The ratio of the estimators given in (3.13) and (3.14) provides an estimator of $J_{\text{inhom}}(r,t)$. The latter is unbiased and the former is ratio-unbiased (see Cronie and van Lieshout, 2015).

3.4.11 Estimation of summary statistics for the considered examples

We keep here our assumption of first-order separability, that is, Eq. (3.2) is assumed to be true, but it is not required to assume that the separability is satisfied in the second-order terms, i.e., Eq. (3.6) is not necessarily true. We use a combination of Gaussian kernels, where the bandwidths are selected by using the rule-of-thumb, which is specially designed to be used with Gaussian kernels (see Section 3.4.3). Note that here we are estimating second-order characteristics, whose estimators are mainly based on distances between points, so that the bandwidths for such descriptors are quite different in general from those chosen to estimate the first-order intensity function. In addition, we use other kernels than the Gaussian kernel along this section and the bandwidths here are selected based on a mean square error minimisation approach (Berman and Diggle, 1989).

Figure 3.7 illustrates the estimates of g (left) and K (right) for the Ebola outbreaks pattern described in Section 3.2.1. The K-function is illustrated by subtracting the theoretical surface under an inhomogeneous Poisson process with the same intensity function (see Sections 3.5.1 and 3.5.2), i.e., $\hat{K}(r,t) - 2\pi r^2 t$. For the pair correlation function estimation, we use an Epanechnikov kernel for space and a biweight kernel for time, with the selected bandwidths being $\varepsilon = 0.99$ degrees and $\delta = 1.57$ years respectively. The $\hat{g}(r,t)$ surface, which describes the spatio-temporal structure of the pattern, shows the typical form of a cluster process for small distances, since $\hat{g}(r,t) \geq 1$ for small r and t. We can see certain regular behaviour when the spatial and temporal distances become large, this is mainly because the few clusters observed in the spatial window (Figure 3.2.1) tend to be very distant from each other. This conclusion is also supported by the $\hat{K}$-function with $2\pi r^2 t$ subtracted, which deviates positively from the plane $K = 0$, reinforcing the fact that we have clustering at short distances, and a tendency towards regularity over long distances.

Analogously, the estimates of the g and K summary descriptors for the Euphausia glacialis dataset described in Section 3.2.2 are displayed in Figure 3.8. We use a spatial Epanechnikov kernel and a temporal biweight kernel for the pair
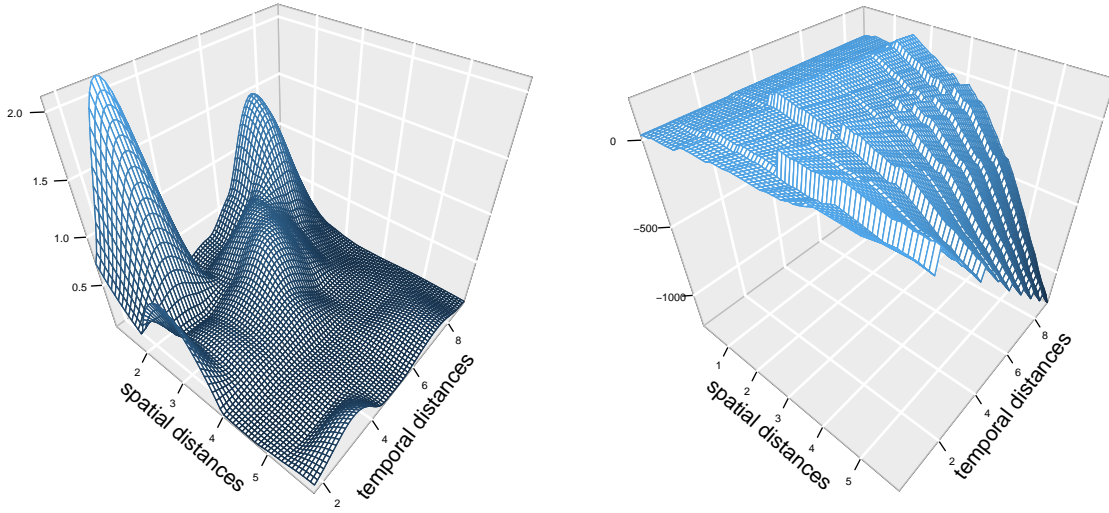
Figure 3.7 Spatio-temporal summary statistics $\hat{g}(r,t)$ (left) and $\hat{K}(r,t) - 2\pi r^2 t$ (right) for the Ebola outbreaks data.

correlation estimation, where the bandwidths are $\varepsilon = 0.01$ degrees and $\delta = 1.2$ years. Here the analysis is basically the same, with both graphs describing strong aggregation in the pattern for short and middle distances, whereas for large distances the interactions seem to be stable and more regular. Note that in this analysis we are assuming riskily that the underlying point process is isotropic, that is not necessarily true and we can suspect it by inspecting Figure 3.2. However, we perform this analysis as a first descriptive approach to understand the nature of the pattern interactions. In Section 3.4.12 we focus in the case in which the isotropy assumption is violated and this dataset is analysed more deeply.

The estimates of the descriptors g and K for the tornado dataset described in Section 3.2.3 are depicted in Figure 3.9. We use a spatial Epanechnikov kernel and a temporal biweight kernel with bandwidths $\varepsilon = 0.12$ degrees and $\delta = 1.55$ years for the pair correlation function. The $\hat{g}$-function seems to fluctuate over the plane $g = 1$ (the plane is shown with the contours in Figure 3.9 (left)). We might think that the proximity of the $\hat{g}$ function to the plane $g = 1$ could indicate complete spatio-temporal randomness (hereinafter CSTR, see Section 3.5.1 for details). To justify this conclusion, one could plot envelopes under CSTR and observe the behaviour of the surface with respect to the envelopes. However, in this case, even the $\hat{K}$-function is difficult to interpret because of the jumps. We can appreciate a small scale of growth on the selected mesh, leading to a reconfirmation of our
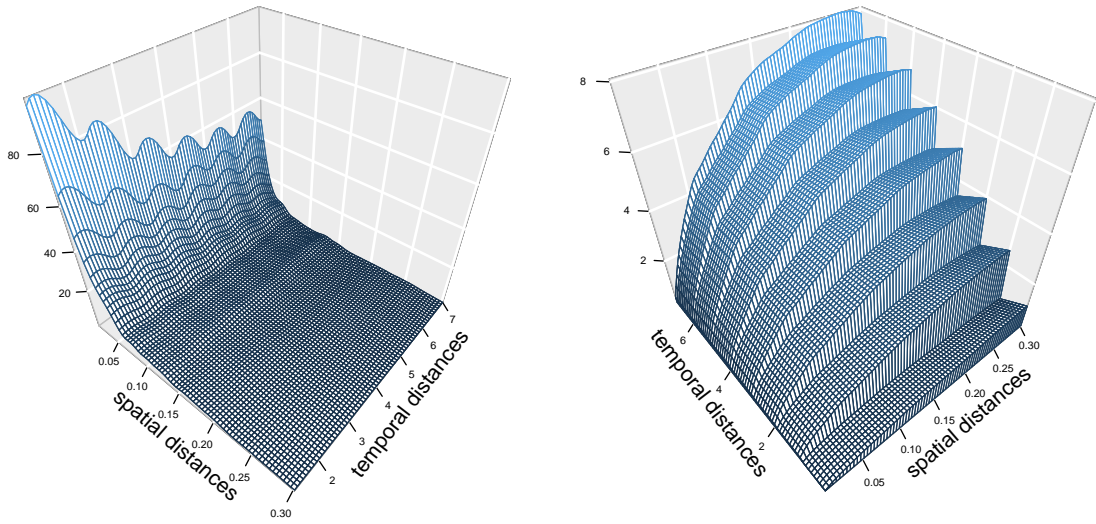
Figure 3.8 $\hat{g}(r,t)$ (left) and $\hat{K}(r,t) - 2\pi r^2 t$ (right) summary statistics for the Euphausia glacialis data.

suspicions of not having CSTR and reinforcing the notion that there is some degree of clustering in the spatio-temporal pattern. We conclude that there is enough information in favour of clustering (although not very strong in this case).
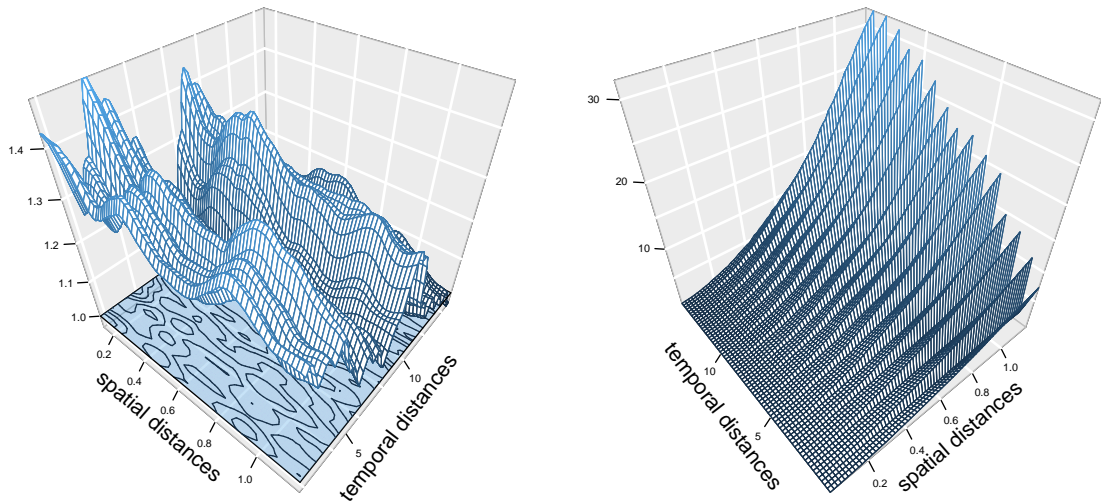
Figure 3.9 $\hat{g}(r,t)$ (left) and $\hat{K}(r,t) - 2\pi r^2 t$ (right) summary statistics for tornadoes in South-Carolina data. The plane $g = 1$ is shown with the contours

Figure 3.10 illustrates the estimated *J*-function for both the Ebola and tornadoes datasets. Since for both $J_{\text{inhom}}(r,t) < 1$ for almost all ranges r and t, we see that both point patterns exhibit clustering at certain distances, as expected and already seen in the estimated *K*-function. Note that the *J*-function for the tornadoes case takes values close to one, indicating a Poisson process behaviour (see Sections 3.5.1 and 3.5.2 for details). Note also that the jumps seen are an artefact of the time scale of the data being discrete and that high peaks in the right plot are an effect of the lags considered being too large; recall that we are employing a minus sampling estimator.
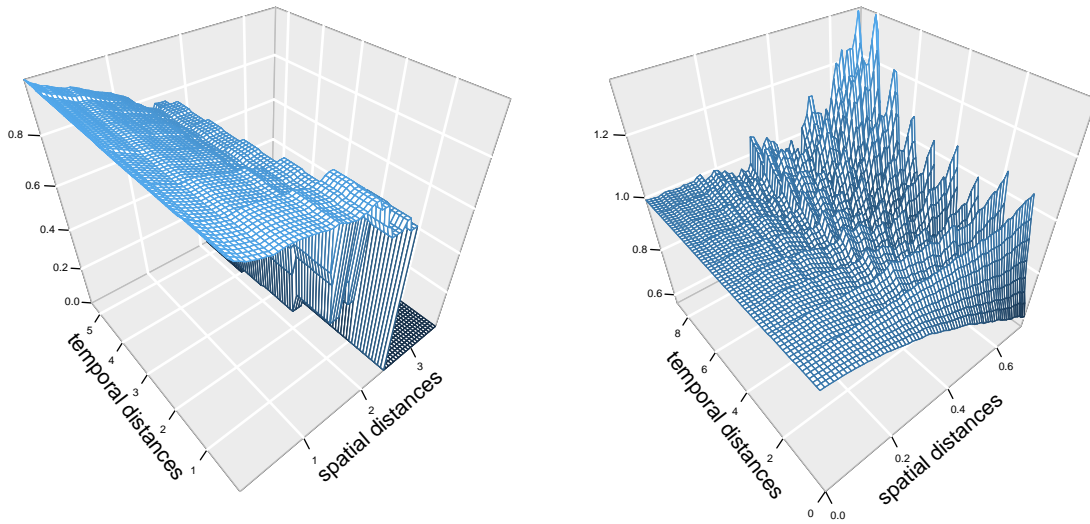
Figure 3.10 $\hat{J}(r,t)$ function estimated for the Ebola outbreaks (left) data and tornado data (right).

3.4.12   Directional second-order summary statistics

In the case of anisotropic planar point processes, Ohser and Stoyan (1981) defined a reduced second moment measure and provided an estimator for the orientation analysis when the intensity is known. Comas et al. (2015) consider a similar approach but assume a SOIRS and anisotropic spatio-temporal point process. The directional *K*-function (hereinafter K_φ-function) should be proportional to the mean number of points in a cylindrical sector with spatial distance r, angle ϕ, and time lag t, centred at an arbitrary point of the spatio-temporal point process X. Let $\theta(\mathbf{u},\mathbf{s})$ be the least angle between the *x*-axis and the line connecting the

points $\mathbf{u}$ and $\mathbf{s}$. The pair (r, ϕ), where $r > 0$, $0 \leq \phi \leq \pi$, $t > 0$ denotes the point
with polar coordinates r and ϕ. The definition of $K_\varphi(r, t, \phi)$ differs from that of
$K_{\text{inhom}}(r, t)$ in that we restrict the points of X to have inter-point angles at most
ϕ. In the stationary and isotropic case the relation between them is given by
$K(r, t) = K_\varphi(r, t, 2\pi) = 2K_\varphi(r, t, \pi)$.

For a SOIRS and anisotropic spatio-temporal point process, we present an
approximately non-parametric edge-corrected estimator of the K_φ-function, which
is a straightforward generalisation of the presented in Comas et al. (2015). The
expression includes a general first-order intensity function estimator (see Section
3.4.3) and considers an edge-correction according to the nature of the underlying
process. The estimator is given by

$$\hat{K}_\varphi(r, t, \phi) = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{\mathbf{1}\{\|\mathbf{u}_i - \mathbf{u}_j\| \leq r, |\theta(\mathbf{u}_i, \mathbf{u}_j)| \leq \phi, |v_i - v_j| \leq t\}}{\hat{\lambda}(\mathbf{u}_i, v_i) \hat{\lambda}(\mathbf{u}_j, v_j) w_{ij}}, \tag{3.15}$$

where w_{ij} is the translation edge-correction if the process is stationary, or the
border or modified border edge-corrections for other cases.

It is possible to detect the predominant directions in spatio-temporal point
patterns through the orientation analysis suggested by Ohser and Stoyan (1981).
They used the directional distribution of line segments connecting point pairs of
the point pattern. The corresponding distribution function is called point pair
orientation distribution and is equal to the probability that a randomly chosen
line segment forms an angle with the x-axis, that is smaller than $\phi \in [0, \pi]$. A
spatio-temporal counterpart version can be obtained via

$$\vartheta_{(r_1, t_1), (r_2, t_2)}(\phi) = \frac{\int_0^\phi \int_{t_1}^{t_2} \int_{r_1}^{r_2} \mathrm{d}K_\varphi(r, t, \psi)}{\int_0^\pi \int_{t_1}^{t_2} \int_{r_1}^{r_2} \mathrm{d}K_\varphi(r, t, \psi)}, \quad r_2 > r_1 \geq 0, t_2 > t_1 \geq 0. \tag{3.16}$$

For two suitable positive values t_1 and t_2, $\vartheta_{(0, t_1), (r, t_2)}(\phi)$ describes a *short-range*
spatial directionality in the point pattern, $\vartheta_{(r_1, t_1), (r_2, t_2)}(\phi)$ provides a *middle-range*
spatial orientation for $r_1 < r_2$, while $\vartheta_{(r_2, t_1), (\infty, t_2)}(\phi) = \lim_{r \to \infty} \vartheta_{(r_2, t_1), (r, t_2)}(\phi)$ describes
long-range spatial directionality.

Short-range orientation provides information about the clustering degree of
the point pattern for short distances in a particular direction at some fixed time.
This kind of information can be useful as a first exploratory analysis looking for
predominant directions in the spatio-temporal point pattern. Given the nature

of the *K*-function as a cumulative function, proportional to the mean number of points in a cylindrical sector, the middle-range orientation reveals whether such directionality in terms of interaction (aggregation degree) is maintained for middle distances or if it has changed in a certain range of distances. Finally, long-range orientation describes the asymptotic orientation (residual orientation) of the point pattern and together with short-range and middle-range orientations provide the whole anisotropic behaviour of the spatio-temporal point pattern.

Furthermore, combinations of spatial and temporal intervals are not worth to be described given their lack of practical relevance. In the case of isotropy, all these distributions coincide with the uniform distribution on $[0, \pi]$. As (3.16) is a cumulative measure for a given angle ϕ, it can be useful to consider a cylindrical sector instead to better highlight the possible directional components (as in Møller et al. 2015),

$$\vartheta^*_{(r_1,t_1),(r_2,t_2)}(\phi) = \vartheta_{(r_1,t_1),(r_2,t_2)}(\phi + \alpha) - \vartheta_{(r_1,t_1),(r_2,t_2)}(\phi - \alpha), \quad 0 \le \phi \le \pi, \qquad (3.17)$$

where $0 < \alpha < \phi$ is a fixed prescribed angle interval, which provides the direction in which anisotropic effects are tested. Using (3.15), we obtain estimators of (3.16) and (3.17), that is, $\hat{\vartheta}_{(r_1,t_1),(r_2,t_2)}$ and $\hat{\vartheta}^*_{(r_1,t_1),(r_2,t_2)}(\phi)$, respectively.

We consider the Euphausia glacialis dataset and calculate $\hat{\vartheta}_{(0,1),(0.1,5)}$ and $\hat{\vartheta}_{(0.1,1),(0.3,5)}$, and obtain the rose histograms (see e.g. Baddeley et al., 2015; Ohser and Stoyan, 1981) depicted in Figure 3.11. They show that the pattern is oriented, with a remarkably clear main direction around $38°$ (0.21π-rad) and $114°$ (0.63π-rad), both for short as well as for long distances, i.e. pairs of Euphausia glacialis swarms in the Antarctic marine environment tend to be either above or below each other. It is noteworthy that the polar coordinate system (very convenient to visualise) shown in Figure 3.5 (left), is not the same we used here for calculations. Here we worked with a polygon that stretches across the coastline of Antarctica in the long-lat system. Interestingly these two graphs highlight that there is a grouping in the direction of about $47°$ (0.26π-rad), a fact that already was suspected from the intensity in Section 3.4.3.
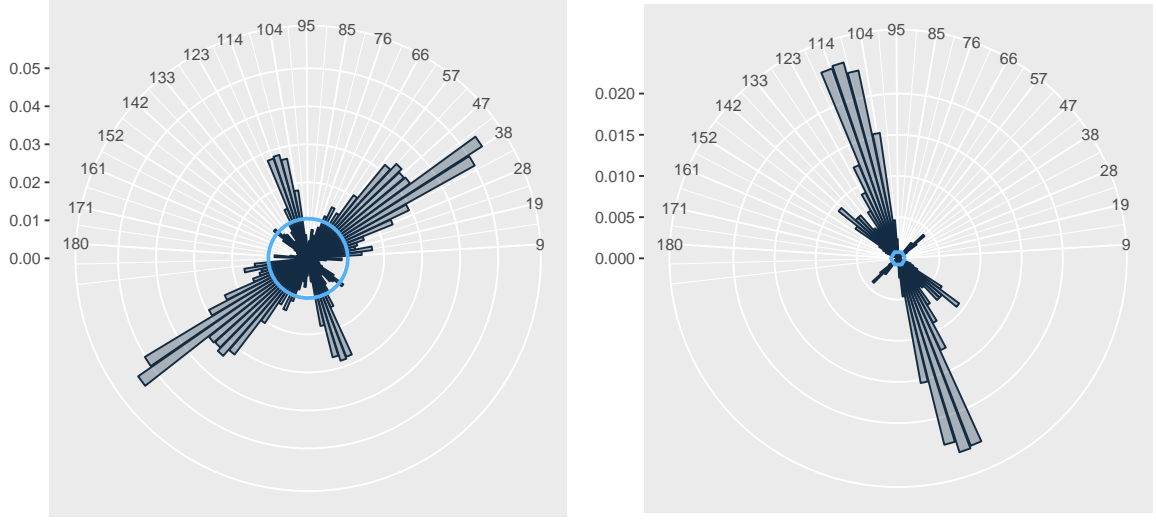
Figure 3.11 Anticlockwise rose histogram of the short and middle orientation distribution $\hat{\vartheta}_{(0,1),(0.1,5)}$ (left) and of the middle range orientation distribution $\hat{\vartheta}_{(0.1,1),(0.3,5)}$(right), for the sample of Euphausia glacialis displayed in the right-side of Figure 3.2, and described in Section 3.2.2. The blue circle corresponds to the median of each range.

3.5   Spatio-temporal empirical models

3.5.1   Spatio-temporal homogeneous Poisson processes

Poisson processes are considered benchmark models for spatio-temporal point pattern data. They are rarely realistic models for data but they do, however, provide a proxy for complete spatio-temporal randomness (CSTR). Explicitly, a spatio-temporal homogeneous Poisson process with intensity $\lambda > 0$ is defined as a spatio-temporal point process X satisfying:

i. Given any disjoint $A_1 \times B_1, \ldots, A_m \times B_m \subseteq W \times T$, the corresponding random variables $N(A_1 \times B_1), \ldots, N(A_m \times B_m)$ follow independent Poisson distributions with the respective means $\mu(A_i \times B_i) = \lambda |A_i \times B_i|$, $i = 1, \ldots, m$.

ii. Conditioned on $N(A \times B)$, the points falling in $A \times B$ form an independent random sample from the uniform distribution on $A \times B$.

It follows that all product densities exist and are given by

$$\lambda^{(k)}(\xi_1, \cdots, \xi_k) \equiv \lambda^k, \quad k \geq 1.$$

Figure 3.12 shows the locations of points in the unit cube (meaning that even the time axes is one unit long) and their respective projections, according to three homogeneous Poisson process realisations, with varying intensities.
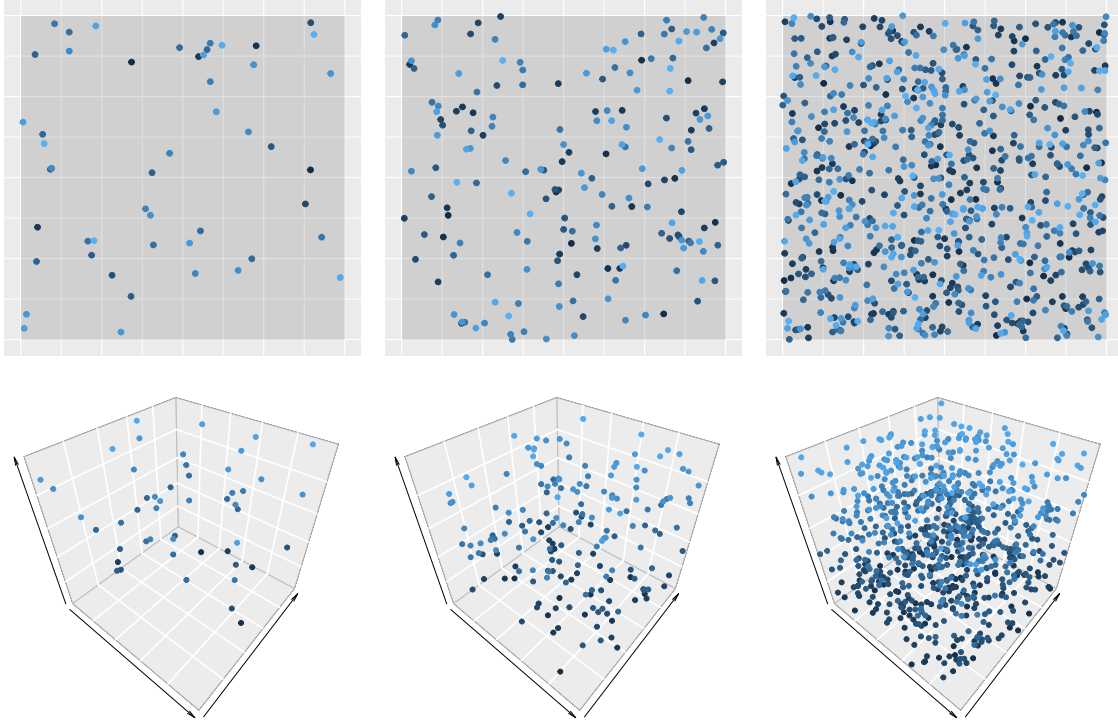
Figure 3.12 Simulated realisations of a spatio-temporal homogeneous Poisson point process with $\lambda = 50$, $\lambda = 150$ and $\lambda = 800$ in left, central and right panels respectively in $W \times T = [0,1]^2 \times [0,1]$, the darker points correspond to older events.

3.5.2   Spatio-temporal inhomogeneous Poisson processes

The spatio-temporal inhomogeneous Poisson process is the simplest non-stationary spatio-temporal point process. It is obtained by replacing the constant intensity of a homogeneous Poisson process by a spatially and/or temporally varying intensity function $\lambda(\mathbf{u},s)$, $(\mathbf{u},s) \in W \times T$. Inhomogeneous Poisson processes are defined by the following postulates:

  i. Given any disjoint $A_1 \times B_1, \ldots, A_m \times B_m \subseteq W \times T$, the corresponding random variables $N(A_1 \times B_1), \ldots, N(A_m \times B_m)$ follow independent Poisson distributions with the respective means

$$\int_{A_i} \int_{B_i} \lambda(\mathbf{u},v)\mathrm{d}\mathbf{u}\mathrm{d}v, \quad i = 1, \ldots, m.$$

ii. Given $N(W \times T) = n$, the n events in $W \times T$ form an independent random sample from the distribution on $W \times T$ which has density function

$$f(\mathbf{u}, v) = \frac{\lambda(\mathbf{u}, v)}{\int_W \int_T \lambda(\mathbf{u}, v) \mathrm{d}\mathbf{u}\mathrm{d}v}.$$

Clearly, we obtain the homogeneous Poisson process by setting $\lambda(\mathbf{u}, v) \equiv \lambda > 0$. Similarly to the homogeneous case, it follows that the product densities exist and are given by

$$\lambda^{(k)}((\mathbf{u}_1, v_1), \ldots, (\mathbf{u}_k, v_k)) = \prod_{i=1}^{k} \lambda(\mathbf{u}_i, v_i), \quad k \geq 1.$$

As an example, we consider a time-stationary, spatially inhomogeneous Poisson process with

$$\lambda(x, y, v) = a \sin\left(\pi \sqrt{\left(\tfrac{3\pi}{2}x - 2\right)^2 + \left(\tfrac{3\pi}{2}y - 2\right)^2} - 1\right) + 2, \tag{3.18}$$

where the Cartesian horizontal and vertical coordinates $(x, y) \in W$, the time $v \in T$ and a is a constant governing the average number of points lying in $[0, 1]^2$. Note that the intensity is separable and $\lambda_{\text{time}}(v) \equiv 1$. For $a = 1000$, in Figure 3.13 we find the intensity together with a realisation of such a process and its cumulative times. We are aware that this form of intensity function is hardly realistic in practice. However some intensities coming from rare phenomena, could well be fitted by this model through a good parameter choice.

Likelihood inference for inhomogeneous spatio-temporal point processes

An instance where the likelihood function is tractable is the inhomogeneous Poisson process with intensity function $\lambda(\mathbf{u}, v)$. Essentially, the distribution associated with a partial realisation of X on a bounded region $W \times T$ can be factorised as the product of a Poisson distribution with mean $\int_W \int_T \lambda(\mathbf{u}, v) \mathrm{d}\mathbf{u}\mathrm{d}v$ for the number of events n, and a set of mutually independent spatio-temporal locations $(\mathbf{u}_i, v_i)$ whose common distribution has density

$$\frac{\lambda(\mathbf{u}, v)}{\int_W \int_T \lambda(\mathbf{u}, v) \mathrm{d}\mathbf{u}\mathrm{d}v}.$$

Following Daley and Vere-Jones (2003) and Diggle (2013), the likelihood may be defined as the probability of obtaining a given number of points in the spatio-
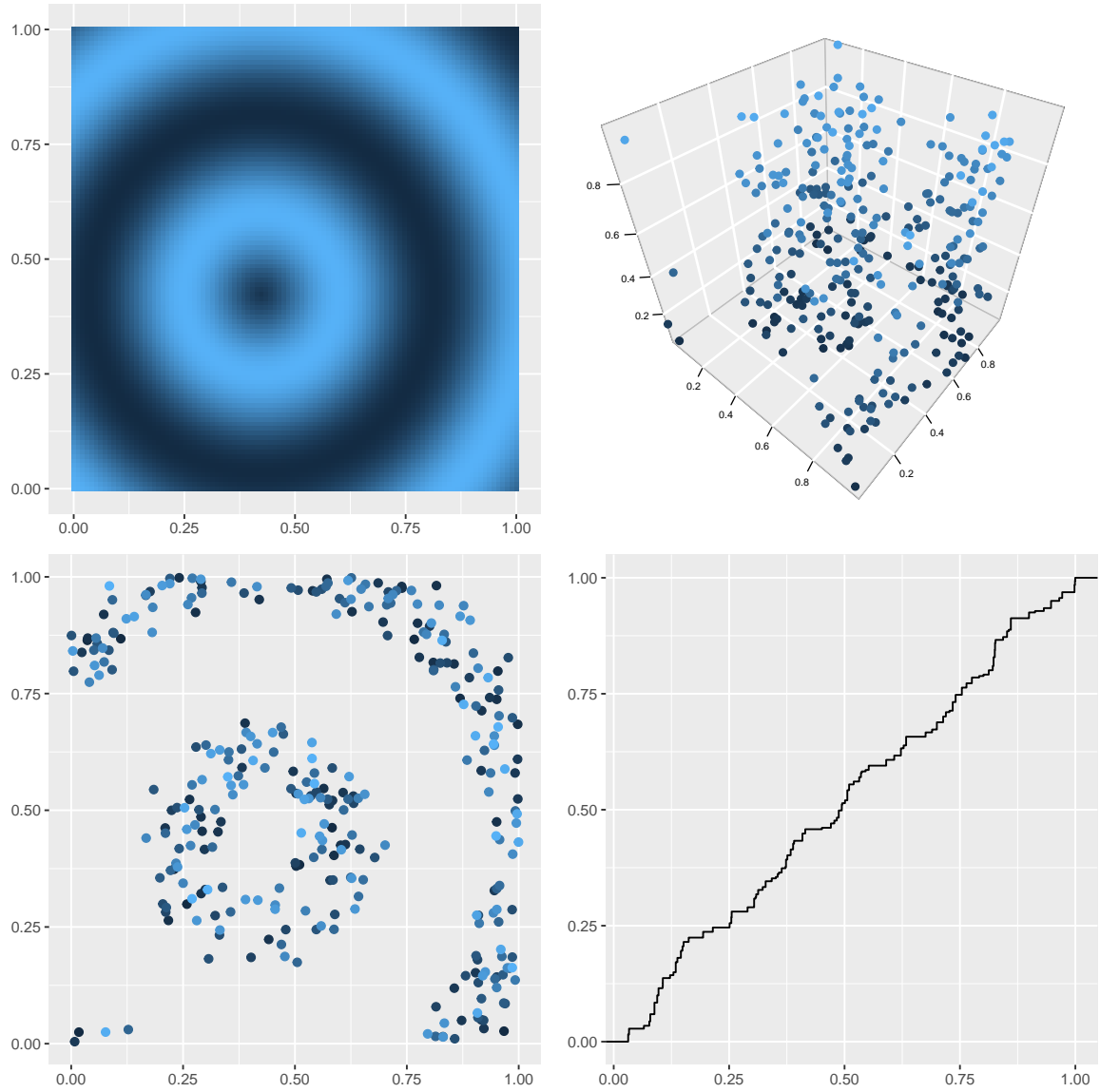
Figure 3.13 Simulated realisation of a spatio-temporal inhomogeneous Poisson point process in $W \times T = [0,1]^2 \times [0,1]$, with λ given by (3.18). On the upper left-side panel the spatial intensity function is displayed, the points of the realisation on the right-side panel represented in a three-dimensional plane with darker points being older. The projection in the plane is displayed in the down left-side panel. Cumulative distribution of times is displayed in the right-side panel

temporal observation window, times the joint conditional density for the locations of those points, given their number. Suppose that there are n observations on $W \times T$ at spatio-temporal points $\{(\mathbf{u}_i, v_i)\}_{i=1}^{n}$. Since the distribution of the number of points is Poisson, then the probability of obtaining single points in some differential volume Δ centred at $(\mathbf{u}_i, v_i)$ and no points on the remaining part of

$W \times T$ is given by

$$\exp\left\{-\int_W \int_T \lambda(\mathbf{u}, v)\mathrm{d}\mathbf{u}\mathrm{d}v\right\} \prod_{i=1}^n \lambda(\mathbf{u}_i, v_i)\Delta.$$

Hence, dividing by Δ^n, letting $\Delta \to 0$ and taking logs, we have that the log-likelihood for $\lambda(\cdot, \cdot)$ based on data is given by

$$L(\lambda) = \sum_{i=1}^n \log \lambda(\mathbf{u}_i, v_i) - \int_W \int_T \lambda(\mathbf{u}, v)\mathrm{d}\mathbf{u}\mathrm{d}v.$$

In practice, it is particularly useful if $\lambda(\mathbf{u}, v)$ can be specified through a regression model, e.g.

$$\log \lambda(\mathbf{u}, v) = \sum_{j=1}^p \beta_j z_j(\mathbf{u}, v), \tag{3.19}$$

where the $z_j(\mathbf{u}, v)$ are covariates that may vary in space and time (Diggle, 2013).

3.5.3   Spatio-temporal Neyman-Scott processes

We define a spatio-temporal Poisson cluster process as the following direct generalisation of its spatial counterpart (Gabriel and Diggle, 2009):

  i. Parents form a Poisson process with intensity $\lambda_{\mathrm{p}}(\mathbf{u}, v)$.

 ii. The number of offspring per parent is a random variable N_c with mean m_c, realised independently for each parent.

iii. The locations and times of the offspring relative to their parents are independently and identically distributed according to a trivariate probability density function $\Phi : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$.

 iv. The final process is composed of the superposition of the offspring only.

   Analogously to the spatial case, the process formed by the parents is taken as an auxiliary construction, and the parents are an unobservable part of the resulting pattern. The shape of a cluster depends on the probability distribution of the offspring so, for example, in a generalisation of the *Matérn cluster process* to the spatio-temporal domain, the offspring are independently and uniformly distributed on disks with a fixed radius around each parent for a fixed time. If the process is time-stationary, the shape of the whole cluster would be cylindrical. In

a possible generalisation of the *modified Thomas process*, the offspring locations follow a normal distribution around each parent for each time, so if the process is time-stationary, then the shape of the cluster would also be cylindrical but with a tendency to accumulate along the temporal axis of the cylinder.

Some examples of patterns which follow this kind of spatio-temporal structure are illustrated in Figure 3.14, where we set an intensity for parents of $\lambda_p(x, y, v) = 6v|\cos(x + y)|$, where $x, y \in [0, \pi]$ and $v \in [0, 1]$. We use different distributions for the offspring: normal, exponential and uniform, respectively. Note that in this case we have no time-stationary process.
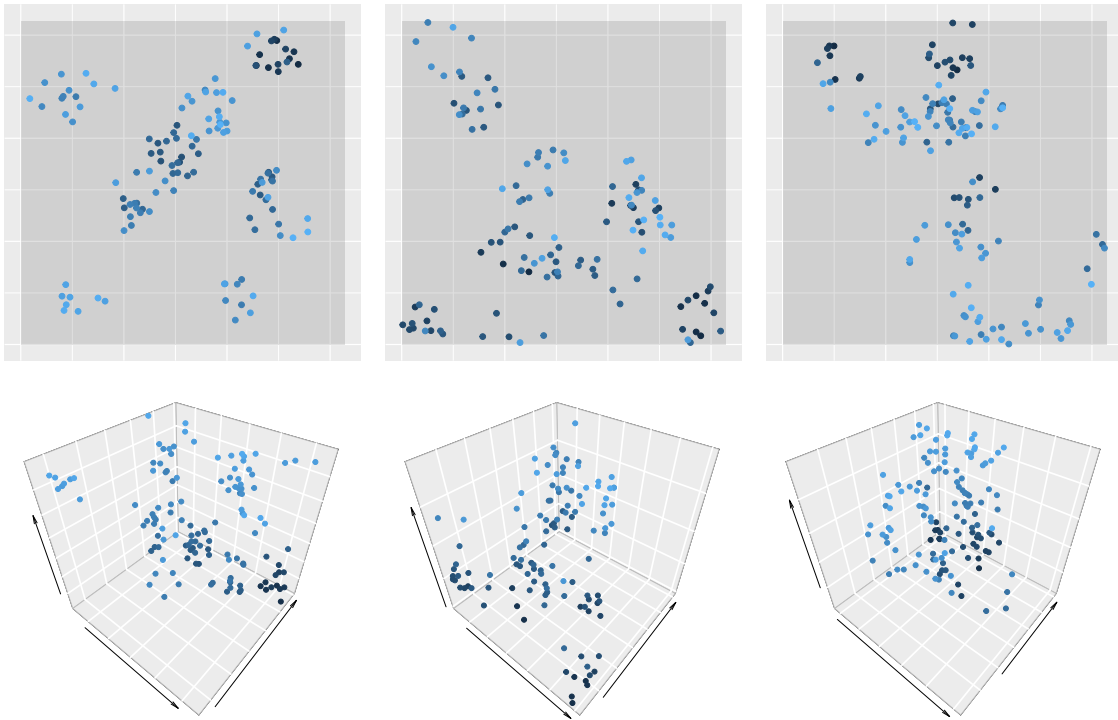
Figure 3.14 Three examples of spatio-temporal Neyman-Scott cluster patterns over the rectangular region $[0, \pi]^2$ and along the unit temporal interval $[0, 1]$, darker points correspond to older occurrences, and $m_c = 8$. The distributions of offspring are: normal (left patterns), exponential (middle patterns), and uniform (right patterns).

3.5.4 Spatio-temporal geometric anisotropic Poisson cluster processes

This type of process is defined as a spatio-temporal stationary Poisson cluster process (Neyman-Scott process) with pair correlation function given by

$$g(\mathbf{u}, v) = g_0\left(\sqrt{\mathbf{u}\Sigma^{-1}\mathbf{u}'}, v\right),$$

where $\mathbf{u}' \in \mathbb{R}^2$ is the transpose of $\mathbf{u} \in \mathbb{R}^2$ and $g_0 : \mathbb{R} \times \mathbb{R} \to [0, \infty]$ satisfies the integrability condition $\int_0^r \int_0^t s g_0(s, l) \mathrm{d}s \mathrm{d}l < \infty$, for $r, t \in (0, \infty)$. The matrix Σ is 2×2 symmetric positive definite and has the form $\Sigma = \omega^2 U_\theta \mathrm{diag}(1, \zeta^2) U_\theta'$, with ζ being the anisotropy factor. The ellipse $E = \{\mathbf{u} : \mathbf{u}\Sigma^{-1}\mathbf{u}'\}$ has semi-major axis ω corresponding to the angle θ and semi-minor axis $\omega\zeta$ corresponding to the angle $\theta + \pi/2$ and

$$U_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

The processes here have a shape of an ellipse at each fixed time. If, further, the process is time-stationary, we have spatio-temporal elliptical cylinder shapes. For details see Gabriel (2014) and Møller and Toftaker (2014).

Figure 3.15 shows a set of three realisations of an anisotropic Poisson cluster process with an average number of parents of 14, and an average of offspring of $m_c = 17$, and using a normal distribution for the locations of offspring with a standard deviation of $\sigma = 0.13$. Here we fix $\omega = 5$ and a rotation angle of $\theta = \pi/8$, and we vary the anisotropy factor $\zeta \in \{0.05, 0.30, 0.80\}$.

3.5.5 Spatio-temporal inhibition processes

Spatio-temporal inhibition processes were presented by Gabriel et al. (2013). They either prevent (strict inhibition) or make unlikely the occurrence of pairs of close events, resulting in patterns that are more regular in space and/or in time than a Poisson process of the same intensity. In a spatial simple sequential inhibition process, also called *RSA* as an abbreviation for "random sequential adsorption" (a term used in physics and chemistry Chiu et al., 2013), which gives strict inhibition, let $\delta_{\mathbf{u}}$ denote the minimum permissible distance between events and λ_{space} the spatial intensity of the process. The proportion of the plane covered
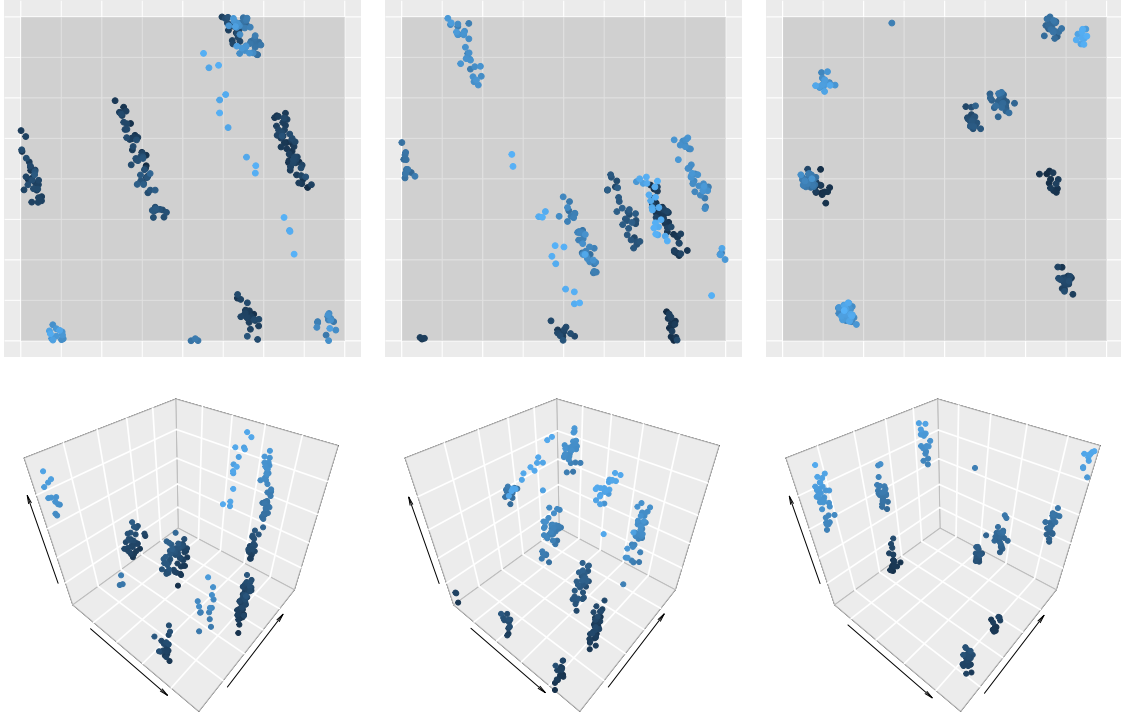
Figure 3.15 Realisations of geometric anisotropic Poisson cluster processes in $W \times T = [0,1]^2 \times [0,1]$. The three cases correspond to $\zeta = 0.05, 0.30, 0.80$ and left, central and right patterns, respectively. Dark dots correspond to older events in time.

by non-overlapping discs of radius $\delta_{\mathbf{u}}/2$ is

$$\lambda^p = \frac{\lambda_{\text{space}} \pi \delta_{\mathbf{u}}^2}{4},$$

which Gabriel et al. (2013) call the *packing density*. The maximum achievable packing density is obtained for a pattern of points in a regular triangular lattice at spacing $\delta_{\mathbf{u}}$, for which $\lambda^p = \sqrt{3}/2$. Depending on how the points are generated, even this value of $\delta_{\mathbf{u}}$ may not be feasible. Simple sequential inhibition processes in space and time are defined by the following algorithm. Consider a sequence of m events $(\mathbf{u}_i, v_i) \in W \times T$. Then,

   i. $\mathbf{u}_1$ and v_1 are uniformly distributed in W and T, respectively.

ii. At the k-th step of the algorithm, $k = 2, \ldots, m$, $\mathbf{u}_k$ is uniformly distributed on $W \cap \Delta_{\text{space}}$, where

$$\Delta_{\text{space}} = \left\{ \mathbf{u} : \left\| \mathbf{u} - \mathbf{u}_j \right\| \geq \delta_{\mathbf{u}}, \quad j = 1, \ldots, k - 1 \right\},$$

and v_k is uniformly distributed on $T \cap \Delta_{\text{time}}$, where

$$\Delta_{\text{time}} = \left\{ v : |v - v_j| \geq \delta_v, \quad j = 1, \ldots, k - 1 \right\}.$$

To obtain a larger class of inhibition processes, it is possible to extend condition (ii) of the above algorithm definition by introducing functions $p_{\mathbf{u}}(s)$ and $p_v(l)$ that together determine the probability that a potential point at location $\mathbf{u}$ and time v will be accepted as a point of the process. Figure 3.16 shows examples of three different single realisations of simple sequential inhibition point processes.
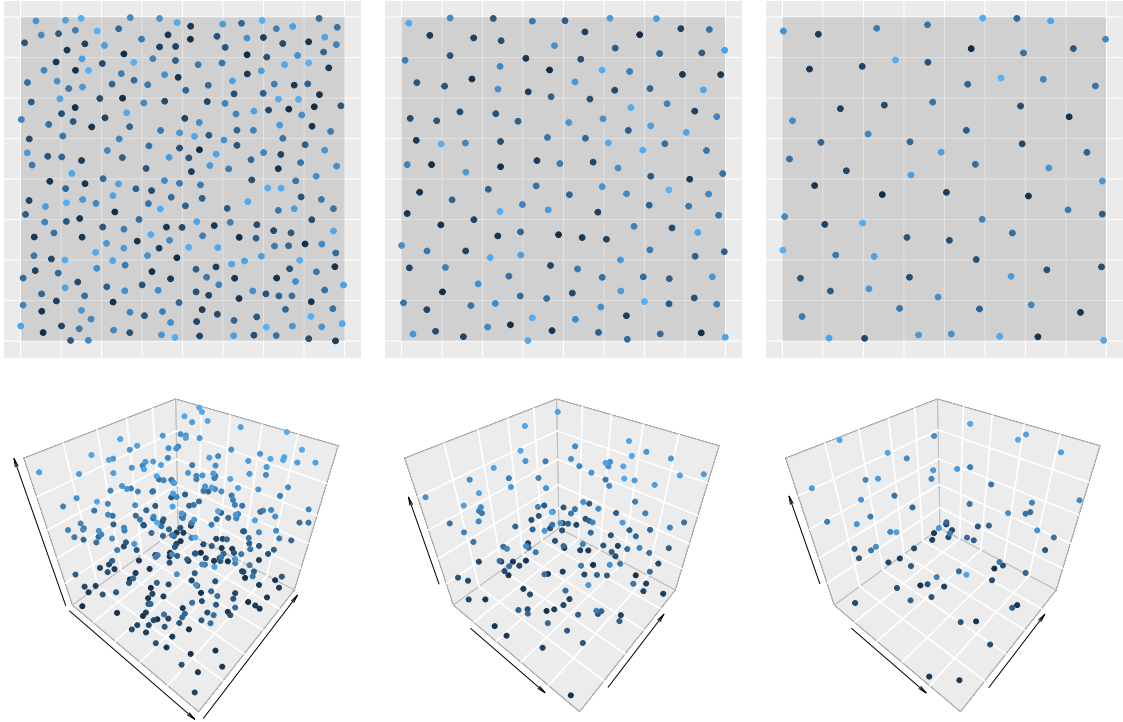
Figure 3.16 Simulated realisations of simple sequential spatio-temporal inhibition point processes with $(\delta_{\mathbf{u}}, \delta_v, n) = (0.04, 0.001, 300)$ (left), $(\delta_{\mathbf{u}}, \delta_v, n) = (0.07, 0.005, 150)$ (center) and $(\delta_{\mathbf{u}}, \delta_v, n) = (0.10, 0.001, 75)$ (right) in $W \times T = [0,1]^2 \times [0,1]$. The darker points correspond to older locations.

3.5.6 Spatio-temporal Strauss processes

Cronie and van Lieshout (2015) define the spatio-temporal hard-core process as a stationary spatio-temporal point process, specified by the Papangelou conditional intensity

$$\lambda^{\dagger}(\mathbf{u}, v | X) = \beta\, \mathbf{1}\{X \cap B[(\mathbf{u}, v), R_W, R_T] = \emptyset\}$$
$$= \beta \prod_{(\mathbf{s}, l) \in X} \mathbf{1}\{(\mathbf{s}, l) \notin B[(\mathbf{u}, v), R_W, R_T]\},$$

where $(\mathbf{u}, v) \in W \times T$, $\beta > 0$ is a model parameter, and $R_W > 0$ and $R_T > 0$ are, respectively, the spatial and the temporal hard core distances. Hence, we have inhibition since

$$\mathbb{P}^{!(\mathbf{0}, 0)}(N(B[(\mathbf{0}, 0), R_W, R_T]) > 0) = 0.$$

The authors further introduce inhomogeneity into the spatio-temporal hard-core process by applying independent thinning. They show that it is IRMS and that both for the thinned and the original hard-core process the corresponding *J*-functions are increasing and larger than 1, given certain set-ups of $R_W > 0$ and $R_T > 0$. This was finally verified numerically and it was shown that the hard-core distances R_W, R_T were well estimated through the inhomogeneous *J*-function.

Since the spatial hard-core process is a particular case of a spatial Strauss process (van Lieshout, 2000), it is easily realised that we may define spatio-temporal Strauss process by

$$\lambda^{\dagger}(\mathbf{u}, v | X) = \beta\, \gamma^{N(B[(\mathbf{u}, v), R_W, R_T])}, \quad \beta > 0, \gamma \in [0, 1].$$

Note that this replacing of Euclidean balls by spatio-temporal cylindrical neighbourhoods clearly provides a recipe for extending certain spatial Gibbs/Markov processes to the spatio-temporal context.

3.5.7 Spatio-temporal Cox processes

Cox processes are natural models for point patterns that are thought to be determined by environmental variability (see Cox, 1955; Diggle, 2013 and Diggle et al., 2013 for a nice exposition of definitions and properties of Cox processes).

A Cox process is a "doubly stochastic" process formed as an inhomogeneous Poisson process with an intensity function coming from some stochastic mechanism. A spatio-temporal Cox process can be defined by the following two postulates:

   i. $\{\Lambda(\mathbf{u},v):(\mathbf{u},v)\in\mathbb{R}^2\times\mathbb{R}\}$ is a non-negative-valued stochastic process.

  ii. Conditionally on $\{\Lambda(\mathbf{u},v)=\lambda(\mathbf{u},v):(\mathbf{u},v)\in\mathbb{R}^2\times\mathbb{R}\}$, the events form an inhomogeneous spatio-temporal Poisson process with intensity function $\lambda(\mathbf{u},v)$.

The moment properties of a Cox process are inherited from those of the process $\Lambda(\mathbf{u},v)$, and thus first- and second-order properties are obtained from those of the inhomogeneous Poisson process by taking expectations with respect to $\{\Lambda(\mathbf{u},v)\}$. Assuming that the covariance structure $\gamma(r,t)=\mathrm{Cov}\{\Lambda(\mathbf{u}_1,v_1),\Lambda(\mathbf{u}_2,v_2)\}$, for $r=||\mathbf{u}_1-\mathbf{u}_2||$ and $t=|v_1-v_2|$, is stationary, a convenient reparametrisation is

$$\Lambda(\mathbf{u},v)=\lambda(\mathbf{u},v)S(\mathbf{u},v), \tag{3.20}$$

where $S(\mathbf{u},v)$ is a stationary process with expectation 1 and covariance function $\gamma(r,t)=\sigma^2 s(r,t)$, where σ^2 is the variance of $S(\mathbf{u},v)$ and $s(\cdot,\cdot)$ is a spatio-temporal correlation function. It follows that $\lambda(\mathbf{u},v)$ is the first-order intensity of the point process, and the stationarity of $S(\mathbf{u},v)$ implies that the point process is intensity-reweighted stationary (Diggle, 2013).

    Given the parametrisation (3.20), we can say that $\Lambda(\mathbf{u},v)$ is *first-order separable* if (3.2) holds, and *second-order separable* if $\gamma(r,t)=\sigma^2 s_1(r)s_2(t)$, where $s_1(r)$ and $s_2(t)$ can be chosen as any pair of valid correlation functions in $\mathbb{R}^2$ and $\mathbb{R}$, respectively. Following (Diggle, 2013; Møller and Díaz-Avalos, 2010), the assumption of first-order separability can be supported from a practical point of view. However, the assumption of second-order separability is more difficult to deal with, but undeniably convenient.

    The *K*-function of an intensity-reweighted stationary Cox process, parametrised according to (3.20), is given by

$$K(r,t)=\pi r^2 t+2\pi\lambda^{-2}\sigma^2\int_0^t\int_0^r xs(x,y)\mathrm{d}x\mathrm{d}y.$$

3.5.8   Spatio-temporal log-Gaussian Cox processes

Møller et al. (1998) introduced the class of log-Gaussian Cox processes. The construction has an elegant simplicity. One of its attractive features is that the tractability of the multivariate Normal distribution carries over, to some extent, to the associated Cox process.

A spatio-temporal log-Gaussian Cox process is a spatio-temporal Poisson process, conditional on the realisation of a stochastic intensity function $\log \Lambda(\mathbf{u}, v)$ (see Diggle et al., 2013). In the intensity-reweighted stationary case with the parametrisation (3.20), for a log-Gaussian Cox process it is possible to write $S(\mathbf{u}, v) = \exp\{Y(\mathbf{u}, v)\}$, where $Y(\mathbf{u}, v)$ is a Gaussian process with expectation $-\tau^2/2$, variance τ^2 and correlation function $\breve{g}(r, t)$. It follows that $S(\mathbf{u}, v)$ has variance $\sigma^2 = \exp\{\tau^2\} - 1$.

Diggle et al. (2013) state that, any valid family of spatio-temporal correlation functions can be used to define a valid class of spatio-temporal log-Gaussian Cox processes. The study of such families is reviewed in Gneiting and Guttorp (2010), where they make a distinction between physical and empirical formulation.

Diggle et al. (2005) used the spatio-temporal Ornstein-Uhlenbeck process approach proposed by Brix and Diggle (2001) to model the underlying spatio-temporal stochastic process component $Y(\mathbf{u}, v)$. However, this model only accommodates separable covariance functions. As an example of a physically motivated construction, Brown et al. (2000) propose models based on a dispersion process. Other parametric families of non-separable models are studied in Cressie and Huang (1999); Gneiting (2002); Ma (2003, 2008) and Rodrigues and Diggle (2010). Alternatively, Rodrigues and Diggle (2012) use a class of low-rank, convolution-based models proposed by Rodrigues and Diggle (2010) to alleviate the computational burden involved in applying likelihood-based methods to full-rank models.

As an illustration, we generated realisations of two log-Gaussian Cox processes, both separable and non-separable. For the separable case, the spatial and temporal covariances are assumed to belong to the class of *exponential* covariance functions. For the non-separable models we have chosen the so-called *Gneiting covariance function* (see details and options in Gabriel et al., 2013). In Figure 3.17 we display the realisations together with the spatial random intensities of two log-Gaussian Cox processes.

Spatio-temporal log-Gaussian Cox processes are quite generally easy-going for statistical inference. The literature provides three different parameter estimation methods: moment-based estimation, maximum likelihood estimation, and Bayesian estimation.

Moment-based estimation. In the stationary case, moment-based estimation consists of minimising a measure of the discrepancy between the empirical and theoretical second-moment properties. One class of such measures is a weighted
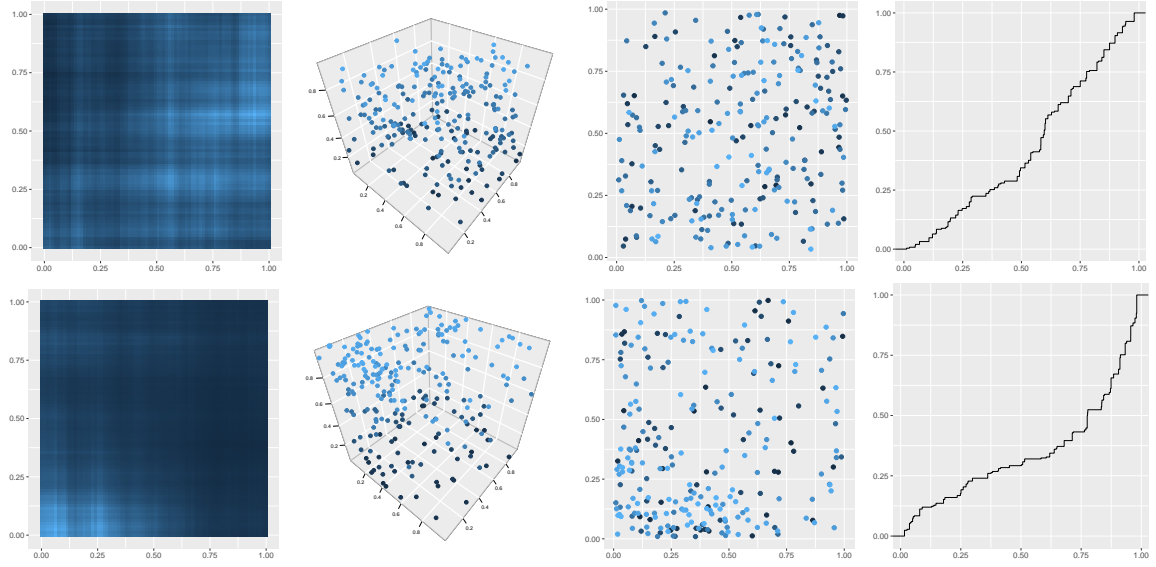
Figure 3.17 Spatial intensities (left), realisations (two central panels), and cumulative distributions of times (right) of log-Gaussian Cox processes. Non-separable (top) and the separable (bottom) cases.

least squares criterion (Diggle et al., 2013). In the intensity-reweighted case, this criterion can still be used after separately estimating a regression model for a spatially varying $\lambda(\mathbf{u}, v)$ under the working assumption that the data are a partial realisation of an inhomogeneous Poisson process. In any case, this method of estimation has an obvious *ad hoc* quality and depends on appropriate choices of several tuning parameters.

Maximum likelihood estimation. The general form of the Cox process likelihood associated with a point pattern $X = \{(\mathbf{u}_i, v_i) \in W \times T : i = 1, \cdots, n\}$ is given by (Diggle, 2013; Diggle et al., 2013)

$$\ell(\boldsymbol{\theta}; X) = \mathbb{E}_{\Lambda|\theta}(\ell^*(\Lambda; X)), \tag{3.21}$$

where

$$\ell^*(\Lambda; X) = \prod_{i=1}^{n} \Lambda(\mathbf{u}_i, v_i) \left(\int_W \int_T \Lambda(\mathbf{u}, v) \mathrm{d}\mathbf{u}\mathrm{d}v \right)^{-n}$$

is the likelihood for an inhomogeneous Poisson process with intensity $\Lambda(\mathbf{u}, v)$. The evaluation of (3.21) involves integration over the infinite-dimensional distribution of Λ. Diggle et al. (2013) describe an implementation in which the continuous region of interest is approximated by a finely spaced regular lattice, hence replacing Λ by a finite set of values covering the region. Even so, the high

dimensionality of the implied integration presents obstacles to analytic progress. One solution is to use Monte-Carlo methods. A crude Monte-Carlo method is proved to be inefficient in practice (Diggle et al., 2013), and a better approach is to use the method of Geyer (1999) (as cited in Diggle et al., 2013).

Bayesian estimation. One way to implement Bayesian estimation would be directly to combine Monte-Carlo evaluation of the likelihood with a prior for θ (Diggle et al., 2013). However, it turns out to be more efficient to incorporate Bayesian estimation and prediction into a single MCMC algorithm. Rodrigues and Diggle (2012) adopt a Bayesian approach to parameter estimation and spatial prediction using low-rank, convolution-based models. Diggle et al. (2013) provide a nice account of the pros and cons of implementing Bayesian inference, MCMC or INLA (integrated nested Laplace approximations, see Rue et al., 2009). See also Taylor and Diggle (2014) for a comparison of the performance of MCMC and INLA for a spatial log-Gaussian Cox process.

Prediction. Diggle et al. (2013) depict very nicely a comparison between plug-in and Bayesian prediction. Suppose that data Z are to be used to predict a target $\mathcal{T}$ under an assumed model with parameters θ. Then, plug-in prediction consists of a series of probability statements within the conditional distribution $[\mathcal{T}|Z;\hat{\theta}]$, where $\hat{\theta}$ is a point estimate of θ, whereas Bayesian prediction replaces $[\mathcal{T}|Z;\hat{\theta}]$ by

$$[\mathcal{T}|Z] = \int_{\Theta} [\mathcal{T}|Z;\theta][\theta|Z]\mathrm{d}\theta, \tag{3.22}$$

where Θ is the domain of the parameter vector θ. Thus Bayesian prediction is a weighted average of plug-in predictions, with different values of θ weighted according to the Bayesian posterior for θ. The Bayesian solution (3.22) incorporates parameter uncertainty in a way that is both natural and elegant.

3.5.9 Spatio-temporal stationary Poisson cluster and shot-noise Cox processes

These cluster processes are built as follows. The spatial distribution of the off-spring is zero-mean bivariate, diagonal with identity diagonal entries, normally distributed with standard deviation σ. The temporal distribution is exponential with rate α. The expected number of offspring per parent follows a Poisson distribution with mean m_c. This process has an interpretation as a spatio-temporal shot-noise Cox process (see Møller and Díaz-Avalos, 2010, and Gabriel, 2014). We consider Cox processes in Section 3.5.7. Here, the residual process $S(\cdot)$ in

equation (3.20) is given by

$$S(\zeta) = \frac{1}{\lambda} \sum_{\xi \in X} \varphi(\zeta - \xi).$$

Here X is a stationary Poisson process in $\mathbb{R}^2 \times \mathbb{R}$ with intensity λ and φ is a density function given by

$$\varphi(\zeta) = \varphi(\mathbf{u}, v) = \phi_{\sigma^2}(\|\mathbf{u}\|)\mathscr{E}_\alpha(v),$$

where $\phi_{\sigma^2}(\|\mathbf{u}\|)$ is the density of a zero-mean bivariate isotropic normal distribution with variance σ^2 and $\mathscr{E}_\alpha$ is the density of a exponential distribution with rate α. For such a process, the main summary statistics are

$$g(\mathbf{u}, v) = 1 + \frac{\alpha}{8\pi\sigma^2\lambda} \exp\left\{ -\frac{\|\mathbf{u}\|^2}{4\sigma^2} - \alpha|v| \right\},$$

and

$$K(r,t) = 2\pi r^2 t + \frac{1}{2\lambda}(\exp\{\alpha t\} - \exp\{-\alpha t\})\left(1 - \exp\left\{ -\frac{r^2}{4\sigma} \right\} \right).$$

From the above, we see that $K(r,t) \geq 2\pi r^2 t$, thus we have clustering by construction.

Spatio-temporal Cox processes show enough flexibility to adapt to a variety of practical situations driven by data. For instance, Prokešová and Dvořák (2013) introduce a flexible inhomogeneous spatio-temporal shot-noise Cox process model where the inhomogeneity is estimated by means of a Poisson score estimating equation. They use minimum contrast estimation based on second-order properties to obtain estimates of the clustering parameters. These authors suggested a non-separable model and use the spatial and temporal projections of the process for parameter estimation.

Now, let $\mathscr{L}$ be a *Lévy basis*, which is defined as an independently scattered, infinitely divisible random measure (Hellmund et al., 2008). It can be shown that Lévy basis include Poisson, Gaussian and mixed Poisson random measures, among others. When we consider a spatio-temporal Cox process in which the random field $\{\Lambda(\xi), \xi \in W \times T\}$ has a driving field of the form

$$\Lambda(\xi) = \int_{W \times T} \kappa(\xi, \eta)\mathscr{L}(\mathrm{d}\eta),$$

where κ is a kernel function (weight) and $\mathscr{L}$ is a non-negative Lévy basis (see Beneš et al., 2015; Hellmund et al., 2008), we have a *spatio-temporal Lévy-driven Cox process*. Under some regularity conditions, Λ has an equivalent shot-noise representation (Møller, 2003) with additional random noise.

3.6   Spatio-temporal mechanistic models

3.6.1   Poisson processes

For a Poisson process, by the independence of the events, we have that

$$\lambda^*(\mathbf{u},v|\mathscr{H}_v)\,\mathrm{d}\mathbf{u}\mathrm{d}v = \mathbb{E}\left[N(\mathrm{d}\mathbf{u}\times\mathrm{d}v)|\mathscr{H}_v\right] = \mathbb{E}\left[N(\mathrm{d}\mathbf{u}\times\mathrm{d}v)\right] = \lambda\left(\mathbf{u},v\right)\mathrm{d}\mathbf{u}\mathrm{d}v,$$

i.e, the conditional intensity function and the first-order intensity function are the same.

   We recall from Section 3.4.2 that the intensity function can be estimated either by smoothing the observations or by fitting some parametric model, and here the situation is similar. In the conditional intensity setting, the non-parametric estimation procedure of the first-order intensity (see Eq (3.3)) is directly followed (see Choi and Hall, 1999).

3.6.2   Self-exciting processes

Stationary spatio-temporal point processes are sometimes described by the covariance between the number of points in some spatio-temporal regions $A \times B$ and $A \times B + (\mathbf{u},v)$ ($A \times B$ shifted by $(\mathbf{u},v)$). A spatio-temporal point process X is called *self-exciting (underdispersed)* if

$$\mathrm{Cov}[N(A \times B), N(A \times B + (\mathbf{u},v))] > 0$$

for small values of $(\mathbf{u},v)$; on the other hand, X is *self-correcting (overdispersed)* if such covariance is negative. Thus the occurrence of points in a self-exciting point process causes other points to be more likely to occur in space-time, whereas in a self-correcting process, the points have an inhibitory effect (see Schoenberg et al., 2010). These models are commonly used in seismology.

Hawkes processes

Hawkes processes (Marsan and Lengliné, 2008, 2010; Schoenberg et al., 2010), are of interest in general point process modelling. There are some alternative ways to define a Hawkes process, here we follow Møller and Rasmussen (2006). Let $X = \{(\mathbf{u}_i, v_i)\}$ be a Poisson cluster process with events $(\mathbf{u}_i, v_i) \in \mathbb{R}^2 \times \mathbb{R}$. The cluster centres of X are given by certain events known as *immigrants*, while the other events are known as *offspring*. A spatio-temporal Hawkes process X satisfies:

   i. The immigrants follow a Poisson process with intensity function $\psi(\mathbf{u}, v)$.

  ii. Each immigrant $(\mathbf{u}_i, v_i)$ generates a cluster C_i, which consists of events of generations of order $m = 0, 1, ...$ with the following branching structure. We first have $(\mathbf{u}_i, v_i)$, which is said to be of generation 0. Given the $0, ..., m$ generations in C_i, each $(\mathbf{u}_i, v_i) \in C_i$ of generation m recursively generates a Poisson process X_j of offspring of generation $m+1$ with intensity function $\kappa_i(\mathbf{u}, v) = \kappa(\mathbf{u} - \mathbf{u}_i, v - v_i)$. Here, κ is a non-negative function defined on $(0, \infty)$.

 iii. Given the immigrants, the clusters are independent.

 iv. X consists of the union of all clusters.

Marsan and Lengliné (2008, 2010) used such processes to investigate how aftershocks are spatially distributed relative to the mainshock. They analysed a regional earthquake dataset, using non-parametric estimations of probabilities of finding aftershocks relative to mainshocks.

Epidemic-Type Aftershock Sequence (ETAS) processes

The Epidemic-Type Aftershock Sequence (ETAS) model is considered the main tool for the spatio-temporal analysis of earthquakes. These models are Hawkes processes and were introduced by Ogata (1988) to describe the times and magnitudes of earthquakes and they were extended to the spatio-temporal setting by Ogata (1998); Ogata and Zhuang (2006), and have since then been widely used to describe earthquakes (see, for instance, Adelfio and Chiodi, 2015; Adelfio and Ogata, 2010; Marsan and Lengliné, 2008, 2010; Mohler et al., 2011 and references therein).

A wealth of analysis techniques have been developed, taking as a starting point an ETAS model; for example Hazard maps, declustering, diagnostic methods,

among others (Adelfio and Chiodi, 2010; Musmeci and Vere-Jones, 1992; Peng et al., 2005; van Lieshout and Stein, 2012; Zhuang et al., 2002).

The magnitude of an earthquake is treated as a mark M_i, associated with its event $(\mathbf{u}_i, v_i)$, and the conditional intensity becomes $\lambda^*(\mathbf{u}, v, M | \mathscr{H}_v) = j(M)\lambda^*(\mathbf{u}, v | \mathscr{H}_v)$, where

$$\lambda^*(\mathbf{u}, v | \mathscr{H}_v) = \psi(\mathbf{u}) + \sum_{\{i: v_i < v\}} \kappa(\mathbf{u} - \mathbf{u}_i, v - v_i | M_i).$$

Note that we here have deviated from the previously indicated omission of marks. Hence, ETAS models work by dividing earthquakes into two categories: *background events*, which occur independently by means of a stationary Poisson process $\psi(\mathbf{u})$, with magnitudes distributed independently of $\psi(\cdot)$ according to the density $j(M)$, and *aftershock events* which represent the risk of aftershocks where the increased risk spreads in space and time following the kernel $\kappa(\cdot, \cdot)$.

Zhuang et al. (2002) dealt with estimation of the spatial intensity function of the background earthquake occurrences from an earthquake catalogue that includes numerous clustered events in space and time. An ETAS model is used for describing how each event generates offspring events. They combined a parametric maximum likelihood estimate for the clustering structures using the space-time ETAS model with a non-parametric estimate of the background seismicity.

3.6.3 Likelihood inference

We assume that the process is orderly (see Diggle, Guan, Hart, Paize and Stanton, 2010; Diggle, Kaimi and Abellana, 2010 and the references therein). It follows that for data $\{(\mathbf{u}_i, v_i)\}_{i=1}^n \subseteq W \times T$, the log-likelihood is given by

$$L(\lambda^*) = \sum_{i=1}^n \log \lambda^*(\mathbf{u}_i, v_i | \mathscr{H}_{v_i}) - \int_T \int_W \lambda^*(\mathbf{u}, v | \mathscr{H}_v) \, \mathrm{d}\mathbf{u} \mathrm{d}v. \qquad (3.23)$$

Hence, likelihood-based inference is quite straightforward for any model for which the conditional intensity is specified. It is only required that $\lambda^*(\mathbf{u}, v | \mathscr{H}_v)$ is non-negative and integrable over W, for any possible history at any time $l \in T$. For a counterexample, see Diggle (2013).

To make inference two further assumptions must be hold (for an example see Peng et al., 2005). First, if a model is not directly specified through its conditional intensity, we must have an explicit expression for $\lambda^*(\mathbf{u}, v | \mathscr{H}_v)$, and this may be difficult, even impossible. Secondly, the integrand of the integral term on the

right-hand side of Eq. (3.23) is often a complicated function with many local modes, so the accurate evaluation of the integral becomes very complicated.

3.6.4   Partial likelihood

As a direct adaptation of an efficient method originally proposed by Cox (1975), for proportional hazard modelling of survival data, which is much more convenient for computing (Diggle, 2013; Tamayo-Uria et al., 2014), we have a variant of the log-likelihood given in Eq. (3.23) called *partial likelihood*.

Conditioning on the times v_i and considering the resulting log-likelihood for the observed time-ordering of events $1, ..., n$, the individual contribution to the partial likelihood is (Diggle, Kaimi and Abellana, 2010)

$$p_i = \frac{\lambda^*\left(\mathbf{u}_i, v_i | \mathscr{H}_{v_i}\right)}{\int_W \lambda^*\left(\mathbf{x}, v_i | \mathscr{H}_{v_i}\right) d\mathbf{x}}, \tag{3.24}$$

and the partial log-likelihood is given by

$$L_p(\lambda^*) = \sum_{i=1}^{n} \log p_i.$$

Usually, the integral in Eq. (3.24) is not tractable from the analytical point of view, but it can be approximated by using some numerical integration technique. When we have a spatio-temporal point process in which W is a finite set of locations $\mathbf{u}_j$ with $j = 1, ..., N$ for some $N \geq n$, $L_p(\lambda^*)$ becomes (see Diggle, 2006; Møller and Sørensen, 1994)

$$L_p(\lambda^*) = \sum_{i=1}^{n} \log \frac{\lambda^*\left(\mathbf{u}_i, v_i | \mathscr{H}_{v_i}\right)}{\sum_{j \in \mathscr{R}_i} \lambda^*\left(\mathbf{u}_j, v_i | \mathscr{H}_{v_i}\right)},$$

where $\mathscr{R}_i$ denotes the risk set at time v_i, and typically $\mathscr{R}_i = \{i, i+1, \ldots, N\}$.

3.6.5   Separability of conditional intensities

Separability in the context of mechanistic models assumes, similarly to the first-order intensity case (Eq. (3.2)), a multiplicative form for the conditional intensity function

$$\lambda^*(\mathbf{u}, v | \mathscr{H}_t) = \lambda_1^*(\mathbf{u} | \mathscr{H}_t) \lambda_2^*(v | \mathscr{H}_t), \quad (\mathbf{u}, v) \in W \times T,$$

where $\lambda_1^*(\cdot | \mathscr{H}_t)$ and $\lambda_2^*(\cdot | \mathscr{H}_t)$ are two non-negative given functions. This hypothesis is especially convenient since each component of a separable process may be

modelled and estimated individually, and this greatly facilitates model building, fitting, and assessment. Díaz-Avalos et al. (2014) considered non-parametric kernel-based estimators; their approach calculates thinning probabilities under the conditions of separability and non-separability and compares them through divergence measures.

The separability assumption is, in fact, quite restrictive. However, few works have addressed a rigorous analysis of separability. Some authors, e.g. Ogata (1988) and Schoenberg (2003), used parametric methods to analyse departures from separability in the ETAS models, for earthquake occurrences. Assunção and Maia (2007); Chang and Schoenberg (2011); Schoenberg (2004) and Díaz-Avalos et al. (2014) developed non-parametric separability tests for the conditional intensity of spatio-temporal point processes. The two latter authors developed Monte-Carlo separability tests based on the comparison between the separable and non-separable kernel estimators of the conditional intensity function, which should match if the point process is separable. These tests have been developed for spatio-temporal marked point processes and for point processes that depend on covariates, but we consider the particular case of separability between the spatial and temporal components of unmarked spatio-temporal point processes.

Assuming that $|\{i : v_i < t\}| = n$, consider a three-dimensional kernel estimator of the non-separable spatio-temporal conditional intensity function $\hat{\lambda}^*_{\text{NS}}(\mathbf{u}, v | \mathscr{H}_t)$, and its separable counterpart estimator $\hat{\lambda}^*_{\text{S}}(\mathbf{u}, v | \mathscr{H}_t)$ consisting of the product of two kernels (two-dimensional and one-dimensional, respectively). Schoenberg (2004) proposes separability tests based on the standardised maximum and minimum absolute distances, a Cramér-Von-Mises type statistic, and a log-likelihood separability test. Díaz-Avalos et al. (2014) take advantage of the fact that the ratio between the intensity and its integral over the whole spatio-temporal region is a density. They consider

$$\hat{p}_i^{\text{NS}} = \frac{\hat{\lambda}^*_{\text{NS}}(\mathbf{u}_i, v_i | \mathscr{H}_t)}{\sum_{j=1}^n \hat{\lambda}^*_{\text{NS}}(\mathbf{u}_j, v_j | \mathscr{H}_t)} \quad \text{and} \quad \hat{p}_i^{\text{S}} = \frac{\hat{\lambda}^*_{\text{S}}(\mathbf{u}_i, v_i | \mathscr{H}_t)}{\sum_{j=1}^n \hat{\lambda}^*_{\text{S}}(\mathbf{u}_j, v_j | \mathscr{H}_t)}.$$

Under the null hypothesis of separability we have that $\hat{p}^{\text{NS}} = \hat{p}^{\text{S}}$, and the Kullback-Leibler and Hellinger divergences (see Deza and Deza, 2009) are given by

$$KL = \sum_{i=1}^n \log \left\{ \frac{\hat{p}_i^{\text{NS}}}{\hat{p}_i^{\text{S}}} \right\} \hat{p}_i^{\text{NS}} \quad \text{and} \quad H = \sqrt{2 \sum_{i=1}^n \left(\sqrt{\hat{p}_i^{\text{NS}}} - \sqrt{\hat{p}_i^{\text{S}}} \right)^2},$$

respectively. They found that *KL* and *H* are competitive with the former tests. In particular, for both Poisson and clustered point processes, they found that when testing for separability between the spatial and temporal components, the probability of type II error for the Monte-Carlo tests, *KL* and *H* decreased faster than for any other test statistic.

3.7   Graphical means of assessing goodness-of-fit

Some recent model evaluation tools in the assessment of spatio-temporal point process models include, in addition to the spatio-temporal summary statistics, residual point process methods such as thinning, superposition and rescaling, comparative quadrat methods such as Pearson residuals and deviance residuals, and weighted second-order statistics for assessing particular features of a model such as its background rate or the degree of spatial clustering. We present here some examples.

Clements et al. (2011) reviewed modern model evaluation techniques for spatio-temporal point processes and demonstrated their use and practicality on earthquake forecasting models.

Schorlemmer et al. (2007) proposed originally an *L*-test, which works by simulating some fixed number *m* of realisations from the forecast model. The log-likelihood *L* is computed for the observed earthquake catalogue (L_{obs}) and each simulation (L_j, for $j = 1, 2, ..., m$). The quantile score, Υ_L, is defined as the fraction of simulated likelihoods that are less than the observed catalogue likelihood

$$\Upsilon_L = \frac{1}{m} \sum_{j=1}^{m} \mathbf{1}\{L_j < L_{\text{obs}}\}.$$

If Υ_L is close to zero, then the model is considered to be inconsistent with the data, and can be rejected. Otherwise, the model is not rejected and further tests are necessary. In the *N*-test, the quantile score examined is instead the fraction of simulations that contain fewer points than the actual observed number of points in the catalogue, N_{obs}. The quantile score Υ_N is defined in a similar way, and the model is rejected if Υ_N is close to 0 or 1.

Methods for residual analysis of spatial point processes have been introduced by Baddeley et al. (2005). Such methods extend readily to the spatio-temporal case. Consider a model $\hat{\lambda}^\dagger(\mathbf{u}, v | X)$ for the Papangelou intensity at any location $\mathbf{u}$ and time *v*. *Raw residuals* may be defined as the difference between the number

of observed points and the number of expected points in any set $A \times B \subseteq W \times T$, that is,

$$R(A \times B) = N(A \times B) - \int_A \int_B \hat{\lambda}^\dagger(\mathbf{u}, v | X) \, d\mathbf{u} dv.$$

Zhuang (2006) extended the definition to the spatio-temporal case using the conventional conditional intensity function. With the purpose of having raw residuals with mean 0, it is possible to rescale them so that Pearson's residuals are defined as

$$R_P(A \times B) = \sum_{(\mathbf{u}_i, v_i) \in X \cap (A \times B)} \frac{1}{\sqrt{\hat{\lambda}^\dagger(\mathbf{u}_i, v_i | X)}} - \int_A \int_B \sqrt{\hat{\lambda}^\dagger(\mathbf{u}, v | X)} d\mathbf{u} dv,$$

for all $\hat{\lambda}^\dagger(\mathbf{u}, v | X) > 0$. These residuals constitute an analogue for spatio-temporal point patterns of the comprehensive strategy for model criticism in the linear model, which uses tools such as residual plots and influence diagnostics to identify unusual or influential observations, to assess model assumptions one by one, and to recognise forms of departure from the model (Baddeley et al., 2005). In analogy with deviances defined for generalised linear models, a good option for comparing models is using deviance residuals (Clements et al., 2011). The spatio-temporal window is divided into evenly spaced bins, and the differences between the log-likelihoods within each bin for the two competing models are analysed. Given two models for the conditional intensity, $\hat{\lambda}_1^*$ and $\hat{\lambda}_2^*$, the deviance residual in each bin $\Pi_i = (A \times B)_i$, of $\hat{\lambda}_1^*$ against $\hat{\lambda}_2^*$ is given by

$$\begin{aligned} R_D(\Pi_i) &= \sum_{i:(\mathbf{u}_i, v_i) \in \Pi_i} \log \hat{\lambda}_1^*(\mathbf{u}_i, v_i | \mathcal{H}_v) - \int_{\Pi_i} \hat{\lambda}_1^*(\mathbf{u}, v | \mathcal{H}_v) \, d(\mathbf{u}, v) \\ &\quad - \sum_{i:(\mathbf{u}_i, v_i) \in \Pi_i} \log \hat{\lambda}_2^*(\mathbf{u}_i, v_i | \mathcal{H}_v) + \int_{\Pi_i} \hat{\lambda}_2^*(\mathbf{u}, v | \mathcal{H}_v) \, d(\mathbf{u}, v). \end{aligned}$$

Positive residuals imply that the model $\hat{\lambda}_1^*$ fits better in the given bin and negative residuals imply that $\hat{\lambda}_2^*$ provides a better fit. By taking the sum of the deviance residuals, one gets a log-likelihood ratio score, which gives an overall impression of the improvement in fit from the best fitted model.

As Clements et al. (2011) establish, the distribution of raw and Pearson residuals tend to be skewed when the spatio-temporal bins are small. When bins are larger, a drawback of bins-based residuals is that considerable information is lost in aggregating over the bins. Instead, it can be possible to look at how the data and model agree, without relying on such aggregation. One way to perform such

an assessment is to transform the points of the process, by rescaling, thinning, superposition or super-thinning, to form a new point process that should be a homogeneous Poisson process if and only if the model used to govern this transformation is correct. The residual points can be assessed for inhomogeneity as a means of evaluating the goodness-of-fit of the underlying model.

3.8   Conclusion

This paper presents a review of known developments for spatio-temporal point process statistics. We have covered aspects of summary statistics, assumptions often considered in this context, statistical models and inference. However we have not entered into details of marked spatio-temporal point process as these still need quite a bit of additional development. We have resisted the temptation of considering novel ideas and concepts, and we have focused on writing a paper which aims at providing the state-of-the-art in the analysis of spatio-temporal point patterns. Another point that we have not covered in the paper is providing a deep modelling framework for the considered datasets. This would have enlarged the present paper, and it is itself the core of another papers.

It is easy to see that all concepts and contexts referred to in this review are closely interconnected. Together, they play an important role in application and current development of spatio-temporal point processes. Of course, they do not represent the only fields where statistical methods in spatio-temporal point processes have been applied with success.

Of the spatial data types, as recently recognised by Banerjee et al. (2014), spatial and spatio-temporal point patterns are the least developed in terms of the use of Bayesian methodology and its application. In this recent and updated book, the authors specifically outline the hierarchical approach through fully Bayesian modelling, which has received much less attention. The emphasis on hierarchical modelling through the Bayesian paradigm in the context of spatio-temporal point processes opens new avenues for future research.

Current open topics of interest, some of them suggested in Banerjee et al. (2014), are the following: (a) Handling measurement error in spatio-temporal point patterns, although this is a quite unstudied area even in the purely spatial context; one of the few papers in which this measurement error is discussed is due to Lund and Rudemo (2000). Consider the setting where the observed locations are measured with error perhaps in both space and time, and we seek to

assess the effect on the intensity function or on second-order summary statistics; (b) Presence-only spatio-temporal data. Analysis of presence-only data (see examples in Chakraborty et al., 2011) has become popular in recent years, and model-based strategies are needed; (c) Preferential sampling. The choice of the sampling locations in a spatio-temporal point pattern rises the question of how to adapt this concept (originally born in the geostatistical field) into the context of point patterns. Once again, preferential sampling has not been discussed so much in the spatial case, even though some studies exist (see e.g. Diggle, Menezes and Su, 2010; Ho and Stoyan, 2008).

Acknowledgements

Chapter 4

Analysis of tornado reports through replicated spatio-temporal point patterns

Abstract

Understanding the spatio-temporal distribution of tornado events is increasingly imperative not only because of the natural phenomenon itself and its tremendous complexity, but also because we can potentially reduce the risks that they entail. In particular, the U.S. regions are particularly susceptible to the occurrence of tornadoes and they are the focus and the motivation of our statistical analysis.

Tornado reports can be seen and treated as spatio-temporal point patterns, and we aim at developing and applying some methods for the analysis of replicated spatio-temporal patterns in order to identify significant structural differences between cold- and warm- seasons along the sampling years. We extend some existing techniques for the spatial case based on the classical spatio-temporal K-function. In particular, we develop two non-parametric tests, one based on a

bootstrap procedure to approximate the null distribution of the test statistic, and the other based on a permutation test. We present intensive simulation studies that demonstrate the validity and power of our tests.

Keywords K-function, non-parametric test, permutation test, separability, spatio-temporal point process, tornadoes.

4.1 Introduction

Formally speaking, according to the Glossary of Meteorology (American Meteorological Society, 2013), a tornado is defined as "a rotating column of air, in contact with the surface, pendant from a cumuliform cloud, and often visible as a funnel cloud and/or circulating debris/dust at the ground". These terrible natural phenomena occur more frequently in certain regions of the world such as the United States, where there is an approximate annual average of 1200 reports and they kill an average of 60 people per year (data provided by National Severe Storms Forecast Center).

It is far from easy to mathematically study the occurrence of tornadoes, and one of the biggest problems is the considerable uncertainty due to problems in spotting and reporting occurrences (Doswell III and Burgess, 1988). The past officially recorded tornado climatologies are thought to be incomplete or with high error rates because, for instance, events can be missed or misclassified and some non-damaging tornadoes in remote areas could still be unreported (Grazulis, 1993). Therefore, studies of tornado climatology apply on reports of tornadoes (noisy data) with all the underlying errors that may have, rather than on the true occurrence of tornadoes (Karpman et al., 2013). The underlying error also widely motivates the need for statistical inference mechanisms to obtain valid conclusions (Wikle and Anderson, 2003).

There is a universally accepted scale known as *The Fujita (or Fujita-Pearson) F-scale* that allows the classification of tornadoes, and it is based largely on the damage inflicted by tornadoes to buildings and vegetation; this scale has five categories listed in increasing order of damage ($F0 - F5$).

Many authors have focused on large-scale studies of tornadoes, especially on the most deadly tornadoes ($F2 - F5$) in order to estimate and predict, for instance, daily climatological probability of tornadoes and yearly counts (Bruening et al., 2002; Concannon et al., 2000; Grazulis, 1993). Some other authors

have concentrated in particular regions (with similar weather conditions) and in specific seasons (e.g. season *El Niño* and *La Niña*), to study tornado counts in such spatio-temporal regions (Monfredo, 1999). There are also studies about the correlation between reports of tornadoes and, for example, the Pacific sea surface temperatures; Marzban and Schaefer (2001) found such a correlation to be negative and weak. However, it has also been found that certain exogenous climatic processes, such as the activity index of El Niño, are significantly associated with reports of tornadoes and this association depends on the regional variability (Wikle and Anderson, 2003).

If the tornado reports are based on spatio-temporal locations in form of coordinates in space and the time of occurrence, we might think about spatio-temporal point patterns as a way to represent and further analyse such tornado reports. However, little has been said about tornadoes in U.S. from the perspective of spatio-temporal point processes due to the high number of reports, the intrinsic complexity of the phenomenon and the large magnitude of the study area. An interesting first attempt is provided by Karpman et al. (2013), who proposed, focussing only on $F3, F4$ and $F5$ tornadoes, a non-parametric and separable model for the first-order intensity measure, that takes into account the topographical variability in the region (measured through an elevation factor), as well as spatial and seasonal effects.

The analysis and characterisation of the intensity of tornadoes and their spatio-temporal structure remains a challenging question for climatologists. Karpman et al. (2013) identified some strategies to estimate the spatio-temporal intensity using point process techniques. Here we go further in the statistical analysis of tornadoes by exploiting the spatio-temporal first- and second-order characteristics.

Motivated by the seasonal patterns of tornadoes, we present several statistical approaches for a better understand the behaviour of such patterns in space and time. We first aim at detecting seasonal effects in density. This has to do with first-order characteristics of the spatio-temporal patterns. Then we try to analyse differences in the spatio-temporal structure for two large seasonal point patterns selected as cold and warm seasons. This would allow us to a better understanding of the large-scale occurrence of tornadoes regardless of the huge difference in intensity over these two seasons, and without taking into account unobservable and some very complicated covariates as those that can describe geographic conditions of country.

Roughly speaking, we propose to test the null hypothesis that two (or more) observed spatio-temporal point patterns are realisations of point processes that have the same second-order descriptors. Following the ideas of Hahn (2012) and Hahn and Vedel Jensen (2016), we subdivide the observation windows into disjoint regions with certain conditions of regularity so that we will obtain several subpatterns that under the null hypothesis should have the same K-function, for instance. Estimates of second-order functions on subpatterns are approximately independent, and behave as if they resulted from replicates of the same underlying process. Therefore our motivating problem is approached by proposing new testing procedures for replicated spatio-temporal point patterns by using non-parametric methods.

In Section 4.2 we describe data from reports of tornadoes in the U.S. In Section 4.3, we consider several approaches for the statistical analysis without additional covariates. We first deal with the first-order intensity function of both the cold- and warm-seasons. We consider a partition of each season in order to carry out an analysis of spatio-temporal patterns to assess possible structural differences in the two yearly cold- and warm-seasons. Finally in Section 4.4 we address some open questions and issues.

4.2   Tornado reports in the U.S.

Tornado data were obtained directly from the performance branch of the Storm Prediction Center (SPC, http://www.spc.noaa.gov) of the National Oceanic and Atmospheric Administration (NOAA), which is part of the National Weather Service (NWS) and the National Centers for Environmental Prediction (NCEP). The data-set contains information on all tornado records in United States and their associated warnings and characteristics. We have extracted all reports collected between 1980 and 2016 because tornado reports prior to 1980 exhibit a high degree of inaccuracy due to record limitations. We have a sample of $n = 39864$ continental tornadoes (we removed all these tornadoes that took place at sea) along 36 years.

We focus on the starting location of each tornado report, which represents the most reliable entry (see e.g. Wikle and Anderson, 2003), in each of the regions identified by the corresponding 50 states. The locations are given by longitude and latitude of the touchdown of the tornadoes with a precision of one-hundredth of a degree. To make spatial analysis a bit more realistic, we might apply a map

projection considering that the U.S. is a fairly large proportion of the Earth's total area (about 1.79%)[1]. Additionally, we would have to consider, for instance, whether the best projection is one that preserves the area (Mollweide), shape (Lambert Conformal Conic), or the distances (Azimuthal Equidistant), see Fig 4.1. Since it is difficult to agree on what is the most important feature in the
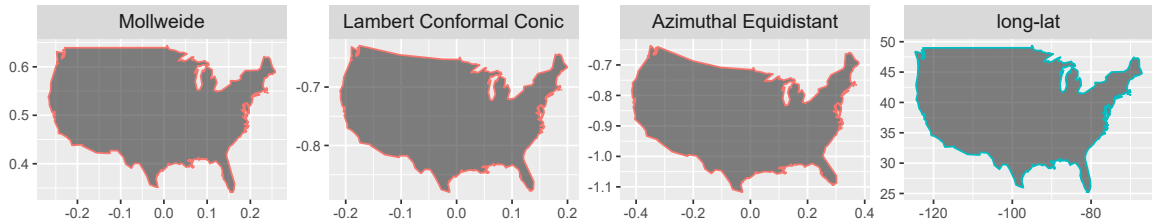
Figure 4.1 Left: Three different projections of the U.S. map with different properties, preserving area, shape and distances respectively. Right: A global coordinate system long-lat.

context of point processes, we prefer to use longitude and latitude as x and y coordinates of the point **u**. The whole data-set is displayed in Figure 4.2.
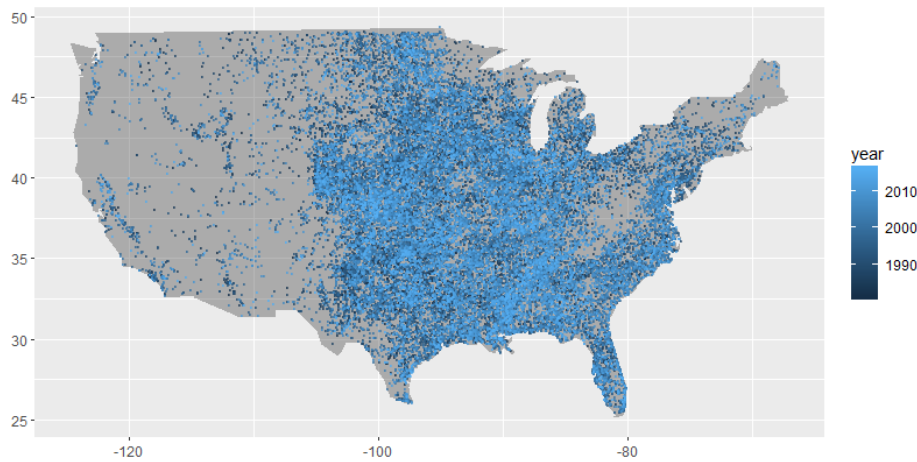
Figure 4.2 Observed point patterns of tornado occurrences in U.S. from 1980 to 2016. Darker points represent older records according to a yearly scale of 36 years.

---

[1]The measure of the total area of the U.S. should include Alaska, Hawaii and outlying territories as Guam and Puerto Rico. However, we avoid these territories because of the availability of tornadoes data. The total world area is 5.10×10^8 km^2 and the U.S. area is 9.15×10^6 km^2 (Central Intelligence Agency, 2017).

4.3 Spatio-temporal distribution

Let $W \subset \mathbb{R}^2$ and $T \subset \mathbb{R}_+ = [0, +\infty)$ be compact sets of positive Lebesgue measures $|W|$ and $|T|$, that are considered as spatial and temporal windows, respectively. Let **X** be a spatio-temporal point process, which stands as a random countable subset of $\mathbb{R}^2 \times \mathbb{R}_+$. The number of points in $\mathbf{X} \cap (W \times T)$ is finite and denoted by $N(W \times T)$. A realisation of $\mathbf{X} \cap (W \times T)$ has the form $X = \{(\mathbf{u}_1, v_1), \ldots, (\mathbf{u}_n, v_n)\} \subset W \times T$.

Let $\lambda(\mathbf{u}, v)$ be the first-order intensity function of the process **X**, so that $\lambda(\mathbf{u}, v)\mathrm{d}\mathbf{u}\mathrm{d}v$ is the mean number of points in an infinitesimal region around $(\mathbf{u}, v)$ with volume $\mathrm{d}\mathbf{u}\mathrm{d}v$. A spatio-temporal point process is called *first-order stationary* if its intensity function $\lambda(\mathbf{u}, v)$ is constant. Clearly, the tornado process does not possess this property. The intensity may have a parametric form, or it can be modelled through non-parametric components for spatial and seasonal effects as in Karpman et al. (2013). The interaction between points can be described, for instance, by the pair correlation function

$$g((\mathbf{u}, v), (\mathbf{s}, l)) = \frac{\lambda^{(2)}((\mathbf{u}, v), (\mathbf{s}, l))}{\lambda(\mathbf{u}, v)\lambda(\mathbf{s}, l)}, \qquad (\mathbf{u}, v), (\mathbf{s}, l) \in W \times T. \qquad (4.1)$$

where $\lambda^{(2)}(\cdot, \cdot)$ is the second-order product density function and

$$\lambda^{(2)}((\mathbf{u}, v), (\mathbf{s}, l))\mathrm{d}(\mathbf{u}, v)\mathrm{d}(\mathbf{s}, l)$$

for $(\mathbf{u}, v) \neq (\mathbf{s}, l)$, may be interpreted as the probability that there is a point from **X** in each of the infinitesimal regions around $(\mathbf{u}, v)$ and $(\mathbf{s}, l)$ of spatio-temporal volumes $\mathrm{d}(\mathbf{u}, v)$ and $\mathrm{d}(\mathbf{s}, l)$, respectively (see e.g González et al., 2016). The process X is called *second-order intensity-reweighted stationary (SOIRS)* if

$$g((\mathbf{u}, v), (\mathbf{s}, l)) = \bar{g}(\mathbf{u} - \mathbf{s}, v - l),$$

for any $(\mathbf{u}, v), (\mathbf{s}, l) \in W \times T$, where $\bar{g}$ is some non-negative function. For further details on spatio-temporal point processes, see e.g González et al. (2016) and references therein. The SOIRS property was defined for spatial point processes in the seminal paper by Baddeley et al. (2000). This assumption has since been widely adopted in the literature on inhomogeneous point processes. A formal test was given in Hahn and Vedel Jensen (2016).

The *K*-function of a spatio-temporal SOIRS process is given by

$$K(r,t) = \int_{c(r,t)} g(\mathbf{u},v) \mathrm{d}(\mathbf{u},v), \quad r > 0, t > 0, \tag{4.2}$$

where $c(r,t)$ represents a spatio-temporal cylinder of radius r and height $2t$, see Gabriel and Diggle (2009); Ghorbani (2013). In the stationary case, i.e. when λ is a constant over the spatio-temporal window, we have that

$$K(r,t) = \lambda^{-1} \mathbb{E}_0^! \left(N[c(r,t)] \right). \tag{4.3}$$

The conditional expectation $\mathbb{E}_0^!(\cdot)$ can be interpreted as the expected number of further events within distance r and time t of an arbitrary event (taken as the origin). Note that the *K*-function measures the interaction structure of the SOIRS spatio-temporal pattern independently of its first-order intensity function.

For a spatio-temporal homogeneous Poisson process, whose spatial and temporal components are independent homogeneous Poisson processes on $\mathbb{R}^2$ and $\mathbb{R}_+$, respectively, $g = 1$ and $K(r,t) = 2\pi r^2 t$, this represents the volume of a cylinder with base radius r and height $2t$. The homogeneous spatio-temporal Poisson process provides a benchmark of complete spatio-temporal randomness (CSTR) (Gabriel and Diggle, 2009). Clustering, which represents attraction between points within a spatial distance r and a temporal distance $2t$, is reflected by $K(r,t) > 2\pi r^2 t$, and values below $2\pi r^2 t$ indicate regularity or repulsion between points.

In order to estimate $K(r,t)$ from a given spatio-temporal point pattern in a bounded region $W \times T \subset \mathbb{R}^2 \times \mathbb{R}_+$, and to compensate for the unobservable points out of such region, usual estimators for $K(r,t)$ include a spatio-temporal edge-correction factor $e(\xi,\eta)$. A useful edge-correction is the isotropic edge-correction method (Gabriel, 2014; Gabriel and Diggle, 2009). The weight is proportional to the product between the spatial Ripley's isotropic edge-correction Ripley (1977) and its analogue for the temporal component. Let $(\mathbf{u},v)$ and $(\mathbf{s},l)$ be points of $W \times T$, so we have

$$e((\mathbf{u},v),(\mathbf{s},l)) = \frac{e^{\mathrm{S}}(\mathbf{u},\mathbf{s}) e^{\mathrm{T}}(v,l)}{|W \times T|},$$

where $e^{\mathrm{S}}(\mathbf{u},\mathbf{s})$ is the reciprocal of the proportion of the perimeter of the circle centred at location $\mathbf{u}$, with radius $||\mathbf{u}-\mathbf{s}||$ which lies within W, and $e^{\mathrm{T}}(v,l) = 1$ if both ends of the interval of length $2|v-l|$ centred at v lie within T, and 2 otherwise (see e.g. Diggle et al., 1995; Gabriel and Diggle, 2009). The estimator of $K(r,t)$ is

then given by

$$\hat{K}(r,t) = \sum_{(\mathbf{u},v)\neq(\mathbf{s},l)} \frac{\mathbf{1}\left[\|\mathbf{u}-\mathbf{s}\| \leq r\right]\mathbf{1}\left[|v-l| \leq t\right]e((\mathbf{u},v),(\mathbf{s},l))}{\hat{\lambda}(\mathbf{u},v)\hat{\lambda}(\mathbf{s},l)}, \tag{4.4}$$

where $\sum_{\neq}$ means that the sum is over all pairs $(\mathbf{u},v) \neq (\mathbf{s},l)$ of the data points. The variance of the estimator $\hat{K}(r,t)$ depends on the point process model, on the arguments r and t, and on the size and shape of the sampling window. For a homogeneous spatio-temporal point process conditional on the number of points $N(W \times T) = n$, the variance can be roughly approximated by

$$\mathbb{V}\mathrm{ar}_n\left(\hat{K}(r,t)\,|N(W \times T) = n\right) \approx \frac{4\pi r^2 t|W||T|}{n^2}\left[1 + 0.305\frac{|\partial W|r}{|W|} + 0.0414n\frac{|\partial W|r^3 t}{|W|^2|T|}\right]$$

where n is the number of points in the spatio-temporal window, and $|\partial W|$ is the perimeter length of W. See a sketch of the proof in the Appendix.

In the following sections we deal with first- and second-order properties of spatio-temporal point processes. Our methodological approaches for comparing groups of patterns come in Sections 4.3.3 and 4.3.3.

4.3.1 Sample

Tornado-Land

Tornadoes occur only rarely in the Western states, and are concentrated on few very restricted areas there. The data appear too sparse for a second-order analysis there. So we restrict the space-time analysis to a region that makes the pattern being closer to a second-order reweighted stationary pattern than the whole U.S. does. We therefore decided to restrict the space-time analysis to a region with relatively high tornado intensity. We thus divide the U.S. map into sensible tornado zones using an *ad hoc* method: define a cut-off at the median of the tornado density, and find the longitudes for which the density exceeds the median. As this is a contiguous interval, all steps have the same length, see Figure 4.3 (left). Therefore, we define the *tornado-land* as the region determined by the longitude range that is preferred by tornadoes, see Figure 4.3 (right).
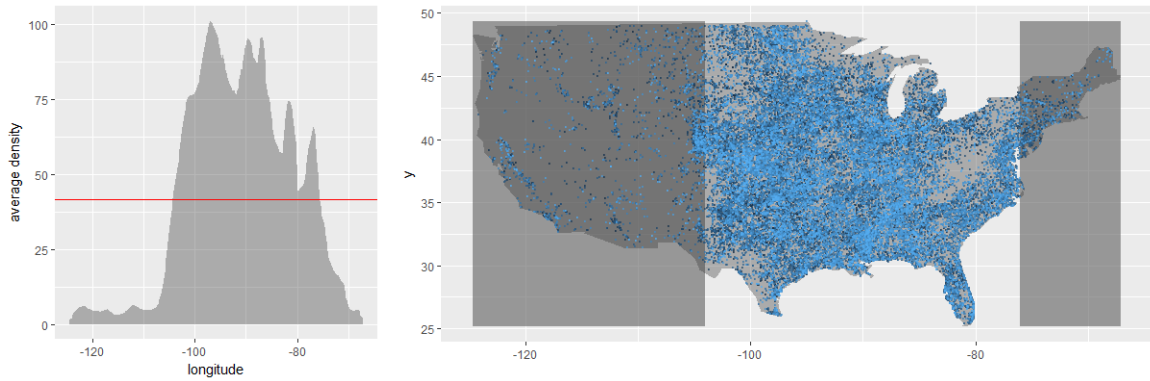
Figure 4.3 Left: Average density of tornadoes in their longitude coordinate. Right: Tornado-land.

Seasons

Our most relevant objective in this paper is to verify if structural seasonal differences exist in the patterns of tornadoes throughout the sampling time. To do this, we split our data-set (Tornado-land) in two exclusionary sets: a *cold-season* defined as the set of all those tornadoes that have occurred during the autumn and winter periods, and a *warm-season* as the set of tornadoes occurred during spring and summer periods.

We define the start of the warm-season as March 1st, and the start of the cold season as September 1st. These are the meteorological definitions of spring + summer, and autumn + winter on most countries of the northern hemisphere. We make bins starting 1979/01/01 (midst in the cold season), and then 1979/03/01, 1979/09/01, etc. This way dates falling into a bin with an odd number are the in cold-season, and dates in even numbers are in the warm-season. There is only one problem, the daylight saving. To avoid this, we let the cold season start one hour later and starts at midnight (+1). When calculating time differences, we can now decide on the time granularity, starting with seconds. The largest possible time unit is the week, because months and years vary in length. The final seasons along all the years are shown in Figure 4.4.

4.3.2 First-order separability

Computations simplify a lot when the spatio-temporal intensity can be assumed to be separable. We therefore check this assumption first. We say that X has a
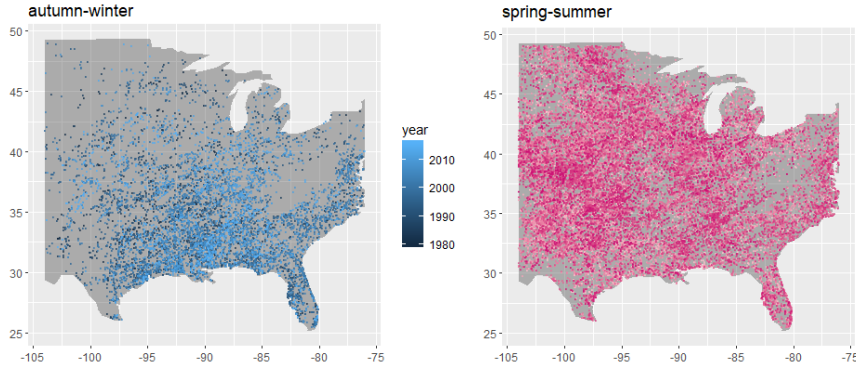
Figure 4.4 Left: The spatio-temporal pattern of tornadoes corresponding to cold-seasons in tornado-land. Right: The spatio-temporal pattern of tornadoes corresponding to warm-season.

separable first-order intensity function if

$$\lambda(\mathbf{u}, v) = \lambda_1(\mathbf{u})\lambda_2(v) \qquad \forall(\mathbf{u}, v) \in W \times T, \tag{4.5}$$

where λ_1 and λ_2 are non-negative functions; in fact, for spatio-temporal stationary and isotropic process, $\lambda(\mathbf{u}, v)$ assumes a constant value λ.

Some authors have suggested separable intensity models for the case of tornadoes. For example, Karpman et al. (2013) have proposed a separable model that has included not only spatio-temporal information but also the elevation variance in a neighbourhood of the location in a semi-parametric model given by

$$\lambda(\mathbf{u}, v) = \lambda \cdot \mathbb{M}^{\nu}(\mathbf{u}) \cdot \mathbb{S}^{\varphi}(v) \cdot \mathbb{D}^{\gamma}(\mathbf{u}),$$

where $\mathbb{M}(\mathbf{u})$ is the spatial first-order intensity function usually estimated by kernel-smoothing, $\mathbb{S}(v)$ is the seasonality component or temporal intensity function and $\mathbb{D}(\mathbf{u})$ is the component that takes into account the elevation variability, and exponents $\nu, \varphi,$ and γ are parameters to be estimated. So far, only models with separable intensity functions have been considered for tornadoes in the literature.

To assume that equation (4.5) holds, we have to study in some depth the tornado intensity functions. The separability hypothesis is particularly convenient since each component of a separable process can be modelled and estimated individually (see Schoenberg, 2004); this represents an important issue given the implications of the model. Despite the importance of this assumption, the separability of a process is rarely rigorously tested. The only available tests apply

to conditional intensity functions, as Díaz-Avalos et al. (2013) suggest. However, nothing has been said about first-order intensity functions. Under a separable first-order intensity model, for two different spatial regions W_1 and W_2, the ratio

$$\frac{\int_{W_1} \lambda(\mathbf{u}, v) \mathrm{d}\mathbf{u}}{\int_{W_2} \lambda(\mathbf{u}, v) \mathrm{d}\mathbf{u}},$$

does not change with time. This assumption is quite strong, in particular in our context as a separable model for tornadoes would imply that the ratio of the risk of a tornado occurrence in a certain area to that of a tornado in another area does not change over time. Thus, although the conditions that affect the risk of a tornado in the regions W_1 and W_2 can be different, their variation along time is similar. This would be the case if the risk of tornado occurrence is proportional to other climate factors related with the geography of the region. Moreover, a separable model is needed for processes in which there is a local temporal change that is more driven by the particular regional conditions such as precipitation and temperature.

We follow this approach to check for first-order separability; consider the projection of all points in space, that is, regardless of time, we divide each of the seasons into 4 groups corresponding to the decades over the years of sampling, and we estimate the spatial first-order intensity function in each of the decades to observe roughly how constant is the variation over time on a large scale. We consider an uniformly corrected kernel estimator for the spatial intensity (see e.g Baddeley et al., 2015; Diggle, 1985), given by

$$\hat{\lambda}_{\text{space}}(\mathbf{u}) = \frac{1}{e^2(\mathbf{u})} \sum_{i=1}^{n} \kappa_{\varepsilon}^2(\mathbf{u} - \mathbf{u}_i), \tag{4.6}$$

for any spatial location $\mathbf{u}$ inside the window W, where

$$e^2(\mathbf{u}) = \int_W \kappa_{\varepsilon}^2(\mathbf{u} - \mathbf{s}) \mathrm{d}\mathbf{s},$$

and where κ_{ε}^2 is the isotropic two-dimensional Gaussian kernel with bandwidth $\varepsilon = (\varepsilon_x, \varepsilon_y)$ calculated by a simple rule of thumb that depends on the short size of the window. The estimated intensities can be seen in Figure 4.5. The data look long-term non-separable: the early years had a focus on Florida, which disappeared in the later years for instance. This is consistent with the evidence of temporal trends in tornado counts, with spatial variation in the magnitude
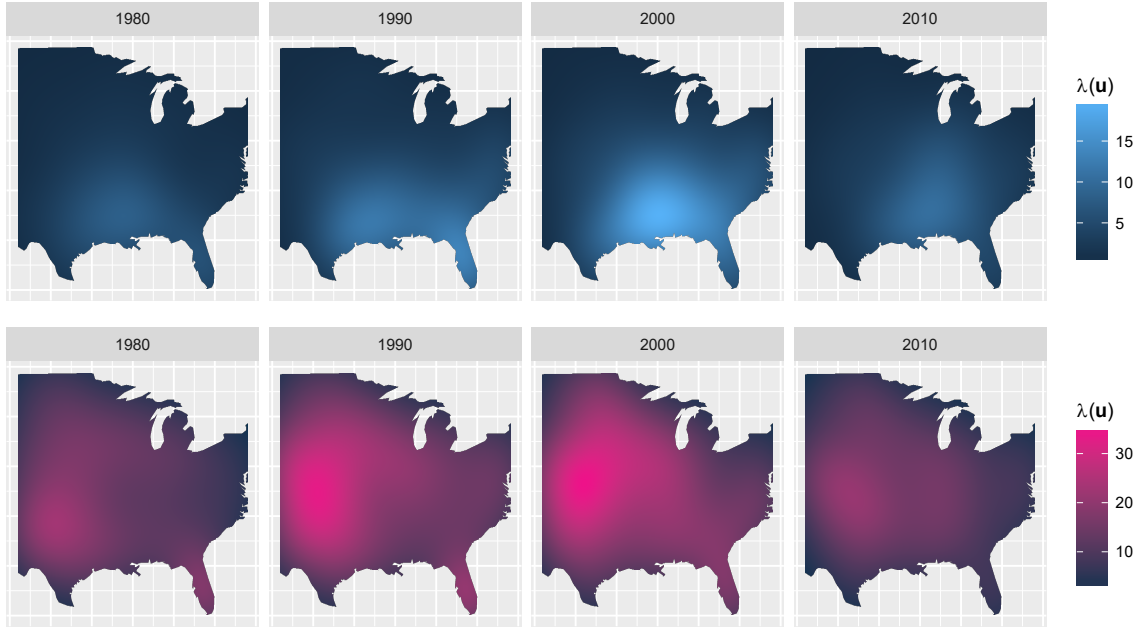
Figure 4.5 Estimation of the spatial intensity for decades (1980, 1990, 2000, 2010) of each tornado pattern given by the cold-(up) and warm-(down) seasons in Tornado-land.

and sign of the trend, see e.g. Wikle and Anderson (2003) and the references therein.

We then use a non-separable estimator which apply a Gaussian kernel for all coordinates. Let δ be the bandwidth for the time, let us define the following non-separable estimator for the spatio-temporal intensity function

$$\hat{\lambda}^{\mathrm{NS}}(\mathbf{u}, v) = \sum_{i=1}^{n} \frac{\kappa_{\varepsilon}^{2}(\mathbf{u} - \mathbf{u}_i)\kappa_{\delta}^{1}(v - v_i)}{e^{2}(\mathbf{u}_i)e^{1}(v_i)}, \tag{4.7}$$

where κ_{ε}^{1} denotes a one-dimensional Gaussian kernel density with bandwidth δ and $e^{1}(v)$ is the analogous to Diggle's edge correction for time.

Note that when the intensity is estimated non-parametrically, everything depends on selection of the bandwidth and this makes the second-order interpretation of inhomogeneous point processes an extremely delicate issue. For the tornado data, a large bandwidth will result in a clustered interpretation, and a small bandwidth leads everything to be interpreted as CSR. Some studies of simulation (not shown here) for the case of a non-separable intensity and some tests with different bandwidths led us to opt, in the case of the temporal dimension, for the use of an optimal bandwidth proposed by Sheather and Jones

(1991). We have then two bandwidths, one for the cold-season and another one for the warm, indeed, $\delta = 63$ days for cold-season and $\delta = 51$ days for warm-season.

4.3.3 Tiles

Spatio-temporal point patterns in Figure 4.4 seem to be inhomogeneous and possibly highly clustered, and trying to analyse the complete patterns makes modelling and inference a difficult task. These intrinsic issues motivate to propose some statistical procedures that relieve the computational burden of the tornadoes data-set and at the same time that is able to develop further statistical testing procedures dealing with replicates of spatio-temporal point patterns.

Therefore, as a variant of the method presented by Hahn (2012) and Hahn and Vedel Jensen (2016), our proposal is as follows. Consider partitioning the Tornado-land map into sub-regions based on deterministic rectangular spatial tiles that have (almost exactly) the same number of points. One can find the boundaries as quantiles, taking one coordinate after the other. These regions are season-specific spatial tiles. Figure 4.6 shows two of such partitions of the spatial region into tiles by splitting the latitudinal coordinate first.

(a)   (b)   (c)   (d)
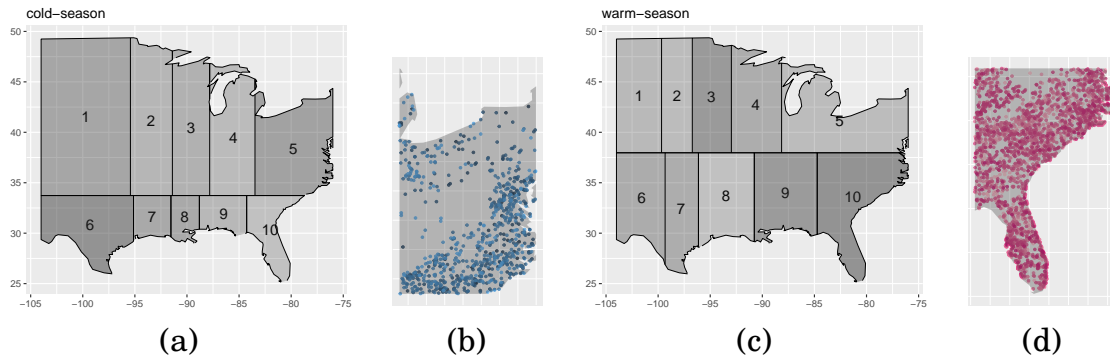
Figure 4.6 Deterministic tiles for warm and cold seasons by splitting the latitudinal coordinate first (a) and (c). Two enlargements from tessellations of the cold- and warm-seasons in (b) and (d), containing observed point patterns of tornado occurrences in the Tornado-land in arbitrary chosen tiles.

Once established this particular way of viewing and splitting the data into two groups of approximately replicated spatio-temporal point patterns, we proceed to show our methodological approaches to give answer to the main scientific question that is of concern in this paper: are there significant differences between the underlying process of the warm- and cold-seasons regardless of their intensity?

Spatio-temporal Diggle's Monte Carlo test

We develop a test in the spatio-temporal case, as a direct extension of a spatial version proposed by Diggle et al. (1991, 2000), who suggested a bootstrap procedure in order to test the differences between independent replicates of empirical spatial K-functions. We assume an original sample consisting of g groups of sizes $m_1, \ldots m_g$, let $n_{ij}(j = 1, \ldots, m_i)$ be a random variable representing the count of tornado occurrences for the jth replicate in the ith group (in our case we have two groups: cold- and warm-seasons respectively), n_i is the mean over all replicates in the group i and $n = \sum_{i=1}^{g} n_i$. We first compute a non-separable version of $\hat{\lambda}_{ij}(\mathbf{u}, v)$, and subsequently a descriptor $\hat{K}_{ij}(r,t)$ (as in Eq. (4.4)) for each pattern, and define the estimated group-specific and overall mean functions as usual in the context of one-way ANOVA by

$$\bar{K}_i(r,t) = \sum_{j=1}^{m_i} w_{ij}\hat{K}_{ij}(r,t) \qquad \text{and} \qquad \bar{K}(r,t) = \frac{1}{n}\sum_{i=1}^{g} n_i\bar{K}_i(r,t). \qquad (4.8)$$

The test statistic, referring to conventional between-treatment sum-of-squares, can be written as

$$BTSS_{\text{st}} = \sum_{i=1}^{g} \int_0^{r_0} \int_0^{t_0} \frac{n_i}{r^2 t}\left[\bar{K}_i(r,t) - \bar{K}(r,t)\right]^2 \mathrm{d}t\,\mathrm{d}r, \qquad (4.9)$$

which is a natural extension of that proposed by Diggle et al. (1991, 2000) to measure differences between groups.

The variance of $\hat{K}_{ij}(r,t)$ increases with r and t (at least in the CSTR case), so we use a weighting factor $(1/r^2 t)$, which down-weights the variance of the spatio-temporal K-function estimator for large r and t. The statistic $BTSS_{\text{st}}$ is a sensible measure of the extent to which the group-specific mean K-functions differ, and is analogous to a residual sum-of-squares in a conventional one-way ANOVA.

A bootstrap procedure   The interest focuses on testing the null hypothesis that the K-functions do not differ between groups, i.e.

$$\begin{aligned}
\mathscr{H}_0 \quad &: \quad \mathbb{E}\left(\bar{K}_1(r,t)\right) = \mathbb{E}\left(\bar{K}_2(r,t)\right) = \cdots = \mathbb{E}\left(\bar{K}_g(r,t)\right) \;\; \text{for all } r \text{ and } t \\
\mathscr{H}_1 \quad &: \quad \mathbb{E}\left(\bar{K}_i(r,t)\right) \neq \mathbb{E}\left(\bar{K}_j(r,t)\right) \;\; \text{for some } r, \text{ some } t \text{ and for some } i \text{ and } j.
\end{aligned}$$

The analytical form of the probability function of $BTSS_{st}$ is intractable, but we perform a pure randomisation test (Diggle et al., 1991) to permute the $\hat{K}_{ij}(r,t)$ across groups and recompute $BTSS_{st}$ in order to obtain its exact conditional distribution. We generate bootstrap samples as follows: in a first step, residual spatio-temporal functions are defined as

$$\hat{R}_{ij}(r,t) = n_{ij}^{1/2} \left(\hat{K}_{ij}(r,t) - \bar{K}_i(r,t) \right). \tag{4.10}$$

Under the null or the alternative hypotheses, the $\hat{R}_{ij}(r,t)$ are approximately exchangeable quantities since the sampling variance of each $K_{ij}(r,t)$ is proportional to n_{ij}^{-1}, although according to Hahn (2012), a more appropriate weight could be used, namely n^{-1}. Note that $\hat{K}_{ij}(r,t) = \bar{K}_i(r,t) + n_{ij}^{-1/2}\hat{R}_{ij}(r,t)$. Then, we obtain a random sample, without replacement, of functional residuals and define

$$\hat{K}_{ij}^{\text{boot}}(r,t) = \bar{K}(r,t) + n_{ij}^{-1/2}\hat{R}_{ij}^{\text{boot}}(r,t). \tag{4.11}$$

To determine the Monte Carlo bootstrap p-value, the observed value of $BTSS_{st}$ is ranked among the corresponding bootstrap values.

Empirical level   We are interested in determining the empirical level of the spatio-temporal version of Diggle's test, so we consider several simple situations that could be useful when analysing the tornado data. To validate the spatio-temporal bootstrap procedure, we apply it to 1000 simulations under the null hypothesis. We consider two groups, each group associated with a number of replicates of $m_1 = m_2 = 10$ under the same type of process, namely spatio-temporal Poisson, cluster and inhibition and three different settings as proposed in Hahn (2012).

A spatio-temporal Poisson cluster process *(Neyman-Scott)* is defined as follows. The parents constitute a Poisson process with intensity $\lambda_{\text{p}}(\mathbf{u},v)$. For each parent the number of *offspring* is a random variable with mean μ_{c}, realised independently. The positions and times of the offspring relative to their parents are independently and identically distributed according to a trivariate probability density function $\gamma(\mathbf{u},v)$. Finally, the process is composed by the superposition of the offspring points (Gabriel et al., 2013).

Spatio-temporal inhibition processes were also presented by Gabriel et al. (2013). In this case the points are prohibited from being closer than some distance apart, resulting in patterns that are more regular in space and time than a Pois-

son process of the same intensity. Let $\delta_{\mathbf{u}}$ and δ_v denote the minimum permissible spatial and temporal distances between events respectively, for a sequence of q events $(\mathbf{u}_i, t_i) \in W \times T$, where $(\mathbf{u}_1, t_1)$ is uniformly distributed in $W \times T$. Now each $(\mathbf{u}_k, v_k)$ is uniformly distributed on

$$(W \times T) \cap \left\{ (\mathbf{u}, v) : ||\mathbf{u} - \mathbf{u}_j|| \geq \delta_{\mathbf{u}}; |v - v_j| \geq \delta_v, j = 1, \ldots, k - 1. \right\}$$

Firstly, we preserve the homogeneity structure of the window, and the intensity of the processes. Secondly, we set different intensities but the same window, and finally we consider same intensity but with different windows.

For the first case (same intensity), we keep a constant intensity of $\lambda = 100$ for the two sets of patterns, and simulations in the *unit cube* ($W \times T = [0,1]^2 \times [0,1]$) and we set $r_0 = t_0 = 0.125$. For the spatio-temporal cluster process, we consider a constant intensity of the parent process ($\lambda_p = 20$) and an average number of offspring per parent of $\mu_c = 5$ with offspring following a normal distribution with $(\sigma_{\mathbf{u}}, \sigma_v) = (0.05, 0.05)$. For the inhibition process, we consider $\delta_s = 0.05$ as the maximum spatial distance of inhibition and $\delta_t = 0.001$ as the maximum temporal distance of inhibition and simulated 100 points per pattern.

For the second case (different intensities), we set for the Poisson case, one intensity of $\lambda = 100$ and another one of $\lambda = 150$. For the cluster process, we consider for the first group $\lambda_p = 20$ and $\mu_c = 5$ and for second group $\lambda_p = 20$ and $\mu_c = 7.5$. For the inhibition process, we set for the first group $n = 80$ and $n = 120$ for the second one.

In the third case, we set the spatio-temporal window to $W \times T = [0, 1.5]^2 \times [0, 1.5]$ for the first group, and for the second group we consider the unit cube, we keep a constant value for the intensity of $\lambda = 100$ in all cases. The results for the simulation experiments are summarised in Table 4.1.

Table 4.1 shows that in all cases the bootstrap procedure gives a somewhat conservative test and confirms the generally mildly good performance using the new statistic $BTSS_{\mathrm{st}}$.

We now show the influence of the upper limits on the integrals of the statistic (4.9) in connection with the empirical values. For this purpose, we vary spatial and temporal limits assigning three values for each, and setting the simulation scenario as in the first homogeneous case (see first row in Table 4.1). The results can be seen in Figure 4.7, where a similar performance is observed in all three cases, slightly better when both the spatial and the temporal limits tend to 0.2.

Nominal significance level α		Rejection Rates		
		0.01	0.05	0.10
Same intensity	Poisson process	0.020	0.067	0.121
	Cluster process	0.014	0.043	0.074
	Inhibition process	0.022	0.068	0.113
Different intensities	Poisson process	0.030	0.072	0.112
	Cluster process	0.027	0.061	0.104
	Inhibition process	0.041	0.092	0.144
Different spatio-temporal windows	Poisson process	0.015	0.041	0.074
	Cluster process	0.059	0.116	0.164
	Inhibition process	0.018	0.049	0.075

Table 4.1 Rejection rates from 1000 replicated simulations of the bootstrap procedure under the null hypothesis of no differences among the samples of *K*-functions of two groups of spatio-temporal point patterns.
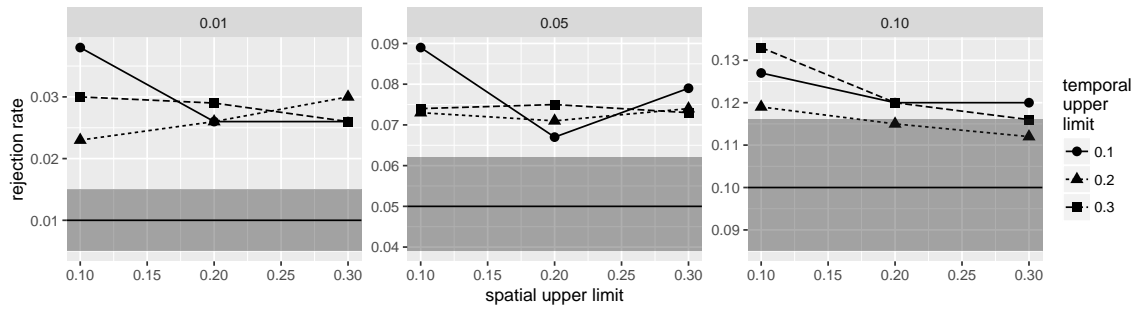
Figure 4.7 Empirical rejection rates of the Diggle's test under the null hypothesis as a function of the upper limits r_0 and t_0, for nominal significance levels $\alpha = 0.01, \alpha = 0.05$, and $\alpha = 0.10$. Results based on 1000 replications for Poisson models, on the unit cube with an intensity of $\lambda = 100$. Gray central area is an 90% interval around the true significance level under uniformity of *p*-values.

Power In order to assess the power of the bootstrap procedure, we simulated sets of 1000 replicates under the alternative hypotheses using five combinations of spatio-temporal processes with intensity $\lambda = 100$ on the unit cube and with 10 replicates in each group against CSTR.

We set a Poisson process as a standard group and for second group, three cluster processes with an average number of offspring per parent of $\mu_c = 1, 4$ and $\mu_c = 4$, respectively, and with a uniform distribution for the offspring with spatial and temporal dispersions of $(\sigma_u, \sigma_v) = (0.20, 0.20), (0.20, 0.20)$ and $(\sigma_u, \sigma_v) = (0.15, 0.15)$, respectively. We also consider a second case using two inhibition processes with $(\delta_u, \delta_v) = (0.02, 0.002)$ and $(\delta_u, \delta_v) = (0.05, 0.005)$ respectively to

		Significance level		
Underlying process	Parameters	0.01	0.05	0.10
Cluster process	$\mu_c = 1, \sigma_u = 0.20, \sigma_v = 0.20$	0.235	0.384	0.480
	$\mu_c = 4, \sigma_u = 0.20, \sigma_v = 0.20$	0.972	0.983	0.989
	$\mu_c = 4, \sigma_u = 0.15, \sigma_v = 0.15$	0.997	0.999	0.999
Inhibition process	$\delta_u = 0.02, \delta_v = 0.002$	0.301	0.494	0.612
	$\delta_u = 0.05, \delta_v = 0.005$	0.992	0.997	0.998

Table 4.2 Empirical power when the underlying spatio-temporal point processes are CSTR in the first group, and cluster and inhibition processes respectively for the second group. The upper limits are $r_0 = t_0 = 0.125$.

generate replicates for the second group. We fix $r_0 = t_0 = 0.125$ as upper bounds in the integrals in $BTSS_{\text{st}}$ given in (4.9). The empirical power estimates are tabulated in Table 4.2.

The results convey the general trends in power analysis in which, as we would expect, we have high values when specific spatio-temporal processes are clearly different in form. Of course, and as shown in the spatial case, the power decreases when we consider spatio-temporal processes that visually appear to be similar or have subtle differences.

Spatio-temporal studentized permutation test

We have been seen that the bootstrap procedure is somewhat conservative. As an alternative, we provide a spatio-temporal version of the studentized permutation test proposed by Hahn (2012), which exact by construction if the observations (K-functions in this case) are exchangeable. Consider the estimates of $K_{ij}(r,t)$ using an unbiased estimator as in (4.4). Let

$$\bar{K}_i(r,t) = \frac{1}{m_i} \sum_{j=1}^{m_i} \hat{K}_{ij}(r,t)$$

and

$$s_i^2(r,t) = \frac{1}{m_i - 1} \sum_{j=1}^{m_i} \left(\hat{K}_{ij}(r,t) - \bar{K}_i(r,t) \right)^2$$

denote the empirical mean and variance of the K-function estimates on a given group i. We define a statistic associated to the t-statistic as

$$T_{\text{st}} = \sum_{1 \leq i < j \leq g} \int_0^{r_0} \int_0^{t_0} \frac{\left(\bar{K}_i(r,t) - \bar{K}_j(r,t)\right)^2}{m_i^{-1} s_i^2(r,t) + m_j^{-1} s_j^2(r,t)} \, dt \, dr \qquad (4.12)$$

The use of the statistic T_{st} may lead to tests sensitive to heteroscedasticity. In these cases, we prefer using the modified version

$$\overline{T}_{\text{st}} = \sum_{1 \leq i < j \leq g} \int_0^{r_0} \int_0^{t_0} \frac{\left(\bar{K}_i(r,t) - \bar{K}_j(r,t)\right)^2}{m_i^{-1} \overline{s_i^2}(r,t) + m_j^{-1} \overline{s_j^2}(r,t)} \, dt \, dr, \qquad (4.13)$$

where

$$\overline{s_i^2}(r,t) = \frac{r^2 t}{r_0 t_0} \int_0^{r_0} \int_0^{t_0} \frac{s_i^2(u,v)}{u^2 v} \, dv \, du.$$

When working with a heteroscedastic sample, the use of $\overline{T}_{\text{st}}$ rather than T_{st} guarantees a better performance of the test (Hahn, 2012).

Empirical level To verify the performance of the spatio-temporal permutation test, we follow the same scenarios as the previous case with the spatio-temporal Diggle's test. Thus we consider two groups of patterns, each with ten replicates ($m_1 = m_2 = 10$), and the combinations of models, intensity functions and observation windows as the same as in the previous test. We compare the two groups using the statistic $\overline{T}_{\text{st}}$ given in (4.13). Note that we are not considering the statistic T_{st} because our tiles exhibit heteroscedasticity. The results for the simulation experiment are summarised in Table 4.3.

Table 4.3 shows that in all cases the permutation test has a good performance analogously to the spatial case (Hahn and Vedel Jensen, 2016), as long as the assumption of exchangeability is met. The empirical rejection rates usually exceed the nominal rejection rates, but it shows a clear improvement with respect to Diggle's test. Again, we consider the configurations of homogeneity in order to observe the influence of the upper limits on the integrals of the statistic (4.13). For this purpose, we vary spatial and temporal limits assigning three values for each case, see Figure 4.8. A similar performance is observed in all three cases for each type of process, although the performance is slightly better when both the spatial and the temporal limits tend to 0.1 and 0.2.

		Rejection Rates		
Nominal significance level α		0.01	0.05	0.10
Homoscedastic	Poisson process	0.010	0.055	0.100
	Cluster process	0.008	0.050	0.098
	Inhibition process	0.011	0.046	0.101
Different intensities	Poisson process	0.010	0.046	0.098
	Cluster process	0.011	0.053	0.099
	Inhibition process	0.012	0.068	0.115
Different spatio-temporal windows	Poisson process	0.019	0.071	0.130
	Cluster process	0.018	0.069	0.118
	Inhibition process	0.020	0.063	0.112

Table 4.3 Empirical rejection rates under null hypothesis of the spatio-temporal permutation test with different underlying processes, based on 1000 replications. The test is applied to two groups of ten realisations of each spatio-temporal point process model.
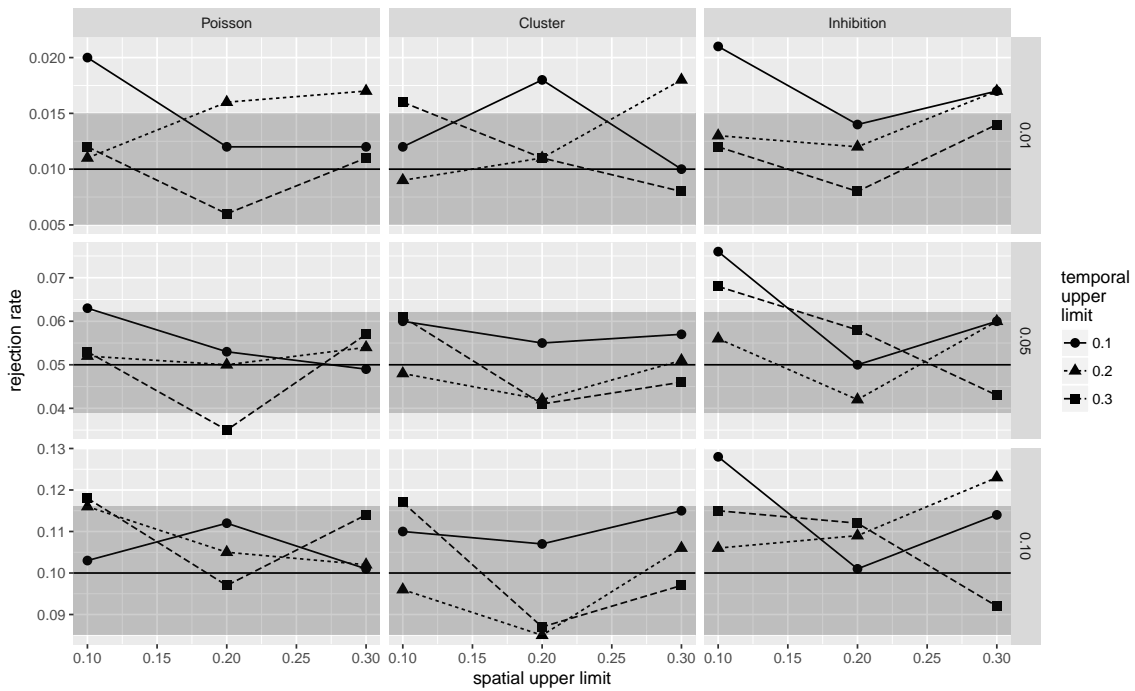
Figure 4.8 Empirical rejection rates of the permutation test under the null hypothesis as a function of the upper limits r_0 and t_0, for nominal significance levels $\alpha = 0.01$, $\alpha = 0.05$, and $\alpha = 0.10$. Results based on 1000 replications, for Poisson, cluster and inhibition models. Gray central area is a 90% interval around the true significance level under uniformity of p-values.

Underlying processes	Parameters	Significance level		
		0.01	0.05	0.10
Cluster process	$\mu_c = 1, \sigma_{\mathbf{u}} = 0.20, \sigma_v = 0.20$	0.229	0.395	0.537
	$\mu_c = 4, \sigma_{\mathbf{u}} = 0.20, \sigma_v = 0.20$	0.968	0.987	0.990
	$\mu_c = 4, \sigma_{\mathbf{u}} = 0.15, \sigma_v = 0.15$	0.995	0.998	0.998
Inhibition process	$\delta_{\mathbf{u}} = 0.02, \delta_v = 0.002$	0.224	0.476	0.631
	$\delta_{\mathbf{u}} = 0.05, \delta_v = 0.005$	0.971	0.990	0.993

Table 4.4 Empirical power of the spatio-temporal permutation test when the underlying processes are cluster and inhibition, respectively, for the second group against CSTR. The upper limits are $r_0 = t_0 = 0.125$.

Power   The power of the test is calculated considering, as an alternative hypothesis, various combinations of realisations of inhibition and cluster spatio-temporal point processes against CSTR patterns, and $r_0 = t_0 = 0.125$ as upper limits in the integrals involved in the statistic $\overline{T}_{\text{st}}$ (4.13). Once again, we simulate sets of 1000 replicates of alternative hypothesis and five combinations of spatio-temporal processes with intensity $\lambda = 100$ on the unit cube and with 10 replicates in each group against CSTR for clustered and inhibition models. We keep the same settings as in the study of the power in the case of Diggle's test. We set for the second group, three cluster processes with an average number of offspring per parent of $\mu_c = 1, 4, 4$ for each case with a uniform distribution for the offspring with spatial and temporal dispersions of $(\sigma_{\mathbf{u}}, \sigma_v) = (0.20, 0.20), (0.20, 0.20), (0.15, 0.15)$. In a second case we use two inhibition processes with $(\delta_{\mathbf{u}}, \delta_v) = (0.02, 0.002), (0.05, 0.005)$. The results are shown in Table 4.4.

In this case, we have the expected results in comparison with those obtained through the bootstrap procedure. Generally, but not always, a non-parametric bootstrap is less powerful than a permutation test (Good, 2005). As the parameters of the models tend to generate virtually indistinguishable realisations from a CSTR, the power of the test decreases. Obviously, and as was already studied in Hahn (2012); Ho and Chiu (2006), the power continuously depends on the upper limits of integrals (in both test studied), here we are using the recommendation of Ripley (1979) of setting $r_0 = 1.25/\sqrt{\lambda}$.

Spatio-temporal pattern analysis of tornado data

As described in earlier sections, we aim to analyse differences in the spatio-temporal structure for cold- and warm- seasons. Thus we split the tornado records into two major seasonal groups, then we subdivide each season into spatio-temporal tiles as shown in Figure 4.6.

We assume that the K-function estimates from different tiles are approximately independent. This assumption is motivated by the high number of points contained in each tile as well as the fact of asymptotic independence of the subpatterns defined by tiles on growing windows, cf. (Hahn, 2012).

A current use of our tests needs a proper selection of the upper limits r_0 and t_0 of the integrals involved in the statistics (4.9) and (4.13) in order to obtain reasonable p-values. Hahn and Vedel Jensen (2016) suggested to take $r_0 = 1.25/\sqrt{\sup_W \hat{\lambda}(\mathbf{u})}$, where $\sup_W \hat{\lambda}(\mathbf{u})$ denotes the supremum of the spatial intensity over its domain W. However, we have a non-separable spatio-temporal intensity function, so that for dealing with the spatio-temporal upper limits we define the marginal intensity functions as (Illian et al., 2008)

$$\lambda_W(\mathbf{u}) = \frac{1}{|T|} \int_T \lambda(\mathbf{u}, v) \mathrm{d}v \quad \text{and} \quad \lambda_T(v) = \frac{1}{|W|} \int_W \lambda(\mathbf{u}, v) \mathrm{d}\mathbf{u},$$

and we consider their natural estimators, which in this case come from our estimate of the intensity through Eq. (4.7). We have approximated the integrals through conventional Riemann sums. Thus we consider the following individual upper limits

$$r^* = \frac{1}{\sqrt{\sup_{\mathbf{u} \in W} \hat{\lambda}_W(\mathbf{u})}} \quad \text{and} \quad t^* = \frac{1}{\sqrt{\sup_{v \in T} \hat{\lambda}_T(v)}},$$

and set $r_0 = 1.25 r^*$ and $t_0 = 1.25 t^*$, although the constant 1.25 can be changed to obtain alternative p-values as indicated below.

For each pattern in each tile, we calculated the spatio-temporal inhomogeneous K-function based on the non-separable estimator of the first-order intensity function (see Eq. 4.7) with upper integration bounds $r_0 = 3.91$ degrees and $t_0 = 67.5$ days, see Figure 4.9. Thus, we perform both Diggle's and the permutation tests. We rank the observed value of the statistic $BTSS_{\mathrm{st}}$ among 90000 bootstrap resamples yielding a p-value of 0.0042, and rank the observed value of $\overline{T}_{\mathrm{st}}$ among all
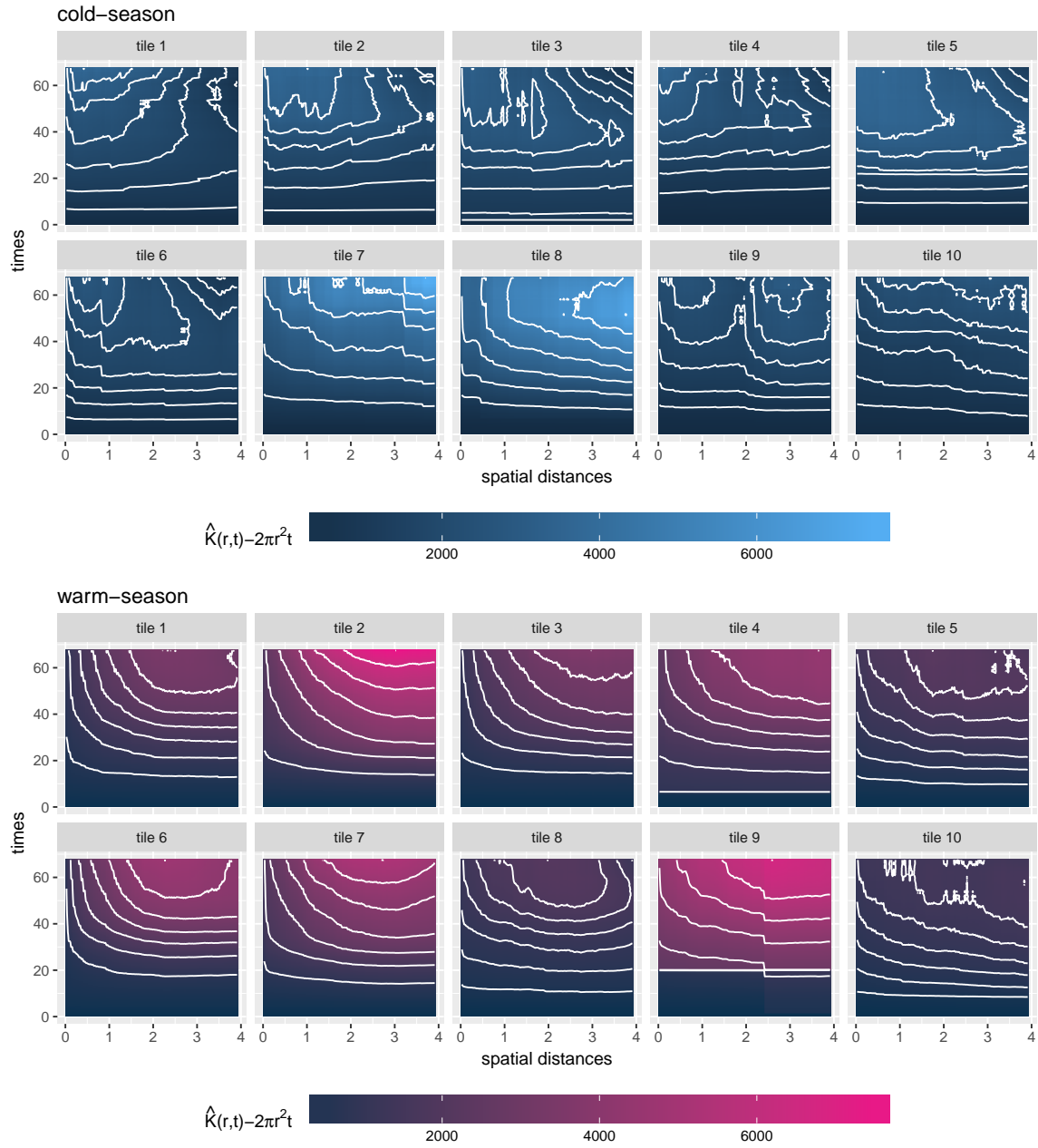
Figure 4.9 Estimates of the spatio-temporal K-function ($\hat{K}(r,t) - 2\pi r^2 t$) for Tornadoes in Tornado-land using Ripley's isotropic edge correction. Up (blue): cold-season K-functions. Down (magenta): warm-season.

possible outcomes, $\binom{20}{10}/2 = 92378$, yielding a p-value of 0.01514, so in both cases the tests give significant p-values

To further analyse the dependence of the results on the upper limits, we proceed to vary these limits and obtain the corresponding p-values. For the

spatial bounds, we range r_0 from $r_0 = 3.13$ degrees ($r_0 = r^*$), to $r_0 = 9.39$ degrees ($r_0 = 3r^*$), and the temporal bounds t_0, from $t_0 = 53.99$ days, ($t_0 = t^*$) to $t_0 = 161.99$ days ($t_0 = 3t^*$). The two tests based on the statistics $BTSS_{st}$ and $\overline{T}_{st}$, were carried out and the associated p-values are shown in Figure 4.10.
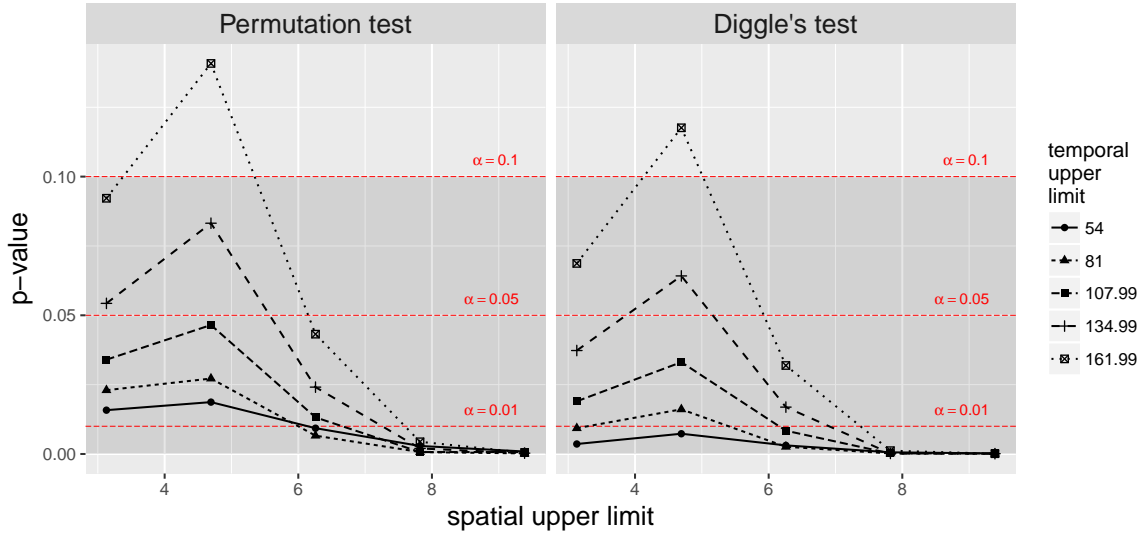
Figure 4.10 Empirical p-values of both the permutation and Diggle's tests as a function of the two upper integration bounds r_0 and t_0 of $BTSS_{st}$ and $\overline{T}_{st}$. Results based on 99999 bootstrapping resamples and 92378 (all) permutations.

Interestingly, both tests give somewhat significant p-values, and we may reject the null hypothesis regardless of the upper limits for the tests. So we can conclude that there are significant structural differences between the tornado cold-season and warm-seasons, meaning that the sources of spatio-temporal variation of interaction do change from one season to another, even though the number of tornadoes is much larger in the warm-season.

To take the data analysis a little further, it should be noticed that in the application of the tests, to avoid the loss of mass, we shift the data in the subsets together according to each season. However, for the estimation of the spatio-temporal inhomogeneous K-function, shifting times together creates an interpretation problem, because for instance, we consider any tornado happened in 30 August 2015 as one day apart from those happened on 1 March 2016. In order to avoid that strange neighbour tornadoes in time, one should estimate the K-function in temporal stripes corresponding to a season. The K-function estimates on the stripes are then pooled over all stripes from one season in each tile, where the weights are the number of tornadoes in that stripe in the respective season and

tile. We can perform the permutation test by using that pooled means. Only stripes with more than 30 points are included in the test, with upper integration bounds $r_0 = 3.13$ degrees and $t_0 = 54.00$ days, the permutation test yields a p-value of 8.66×10^{-5}. The results for using several upper integration bounds are depicted in Figure 4.11. The difference between K-functions for cold and warm seasons
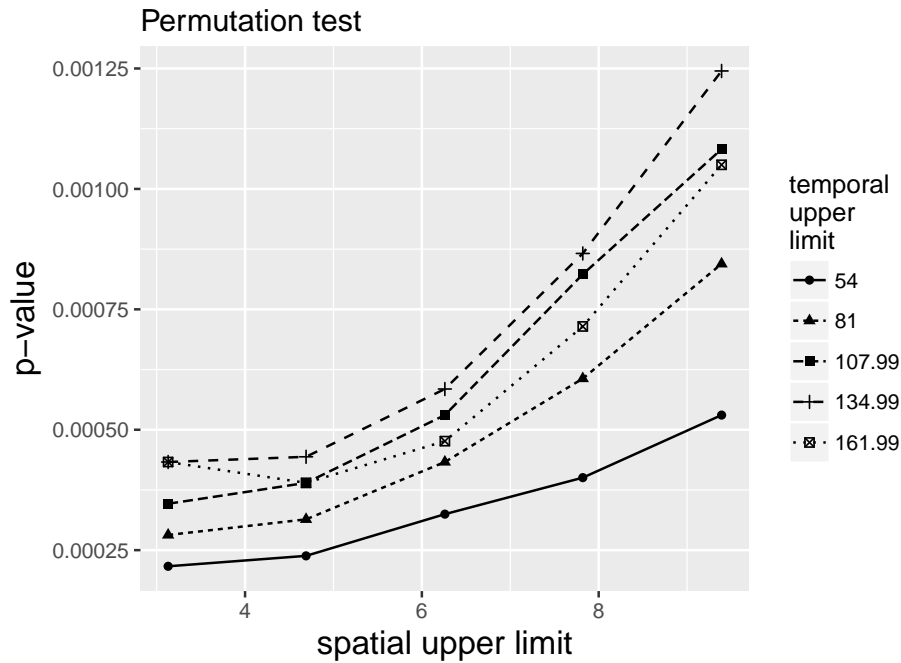
Figure 4.11 Empirical p-values of the permutation tests as a function of the two upper integration bounds of $\overline{T}_{\mathrm{st}}$. The permutation test is performed by using pooled K-functions over all temporal stripes from one season in each tile. Results based on 92378 (all) permutations.

was statistically significant, no matter how the spatial upper limit was chosen.

4.4   Discussion

The statistical framework we have adopted here corresponds to a set of extensions of spatial point process techniques to spatio-temporal point processes. We have divided the point pattern of occurrences of tornadoes in the U.S. in two seasons (cold and warm) to proceed to the comparison of the patterns in a structural (interaction) level through the partition of the spatial region into tiles.

In the first place, it is important to note that, strictly speaking, the Euclidean distance used for the calculations involved in the geometry of space is not strictly

fair. Although tornado-land is not particularly a huge land region, all our calculations are based on long-lat coordinates as we have highlighted in Section 4.2. For a more rigorous approach, we should be aware that Earth is a spaces with curvature, where straight lines should be replaced by geodesics. Thus we could use spherical trigonometry and replace the Euclidean distance by the *great circle distance*, which is the shortest distance between points $\mathbf{u}_1$ and $\mathbf{u}_2$ on the surface of the Earth measured along a path on the Earth's surface (Deza and Deza, 2009). It is the length of the great circle arc, passing through $\mathbf{u}_1$ and $\mathbf{u}_2$, in the spherical model of the planet. Another interesting question for further study would be to investigate how the analysis is affected by different projections of the data.

We have focused on reports of tornadoes rather than their actual locations. These reports could have some kind of noise that is difficult to control in practice. According to the Storm Prediction Center, tornado dataset has gone through many changes over the years. The source data ingested into the database are widely varied and leads to many questions about the precision and accuracy of the location data. Nevertheless, the reports have been vastly improved in matters of precision and automation since 1995 where reports are made through sophisticated technological tools. Handling of such noise would be a very interesting topic for future research since there is not much done in this area in point process context.

We have implemented a new estimator for non-separable first-order intensity function. The use of separable estimators for the first spatial-temporal order intensity is very common in the literature (Diggle et al., 1995; Gabriel, 2014; Gabriel and Diggle, 2009; Gabriel et al., 2013). However, neither the verification of this assumption nor its compliance in practice are trivial at all and if the data were separable, the analysis would be much easier. Separable intensities are a special case of the more general non-separable ones, so a non-separable estimator is always suitable in this context even though it could overfit the data somewhat if the true intensity function is separable.

Although a Monte Carlo test could be developed to verify that the assumption is fulfilled, it must be taken into account that the separability hypothesis is a composite hypothesis (i.e. when it is necessary to estimate parameters), this makes that the Monte Carlo inference has to be made through a nested two-stage test which performs well for composite null hypotheses as it is described in Baddeley et al., 2017. The test could be developed as follows. Let $\mathbf{X}$ be a

spatio-temporal point process with unknown intensity $\lambda(\mathbf{u}, v)$. The hypothesis

$$\mathscr{H}_0 : \lambda(\mathbf{u}, v) = \lambda_1(\mathbf{u})\lambda_2(v),$$

for two positive and measurable functions $\lambda_1(\cdot)$ and $\lambda_2(\cdot)$ and for all $(\mathbf{u}, v) \in W \times T$, is composed. A test statistic F can be given, for example, by an f-divergence (Kullback-Leibler divergence or Hellinger distance) between the estimated non-separable and separable intensities $\hat{\lambda}^{\mathrm{NS}}(\mathbf{u}, v)$ and $\hat{\lambda}^{\mathrm{S}}(\mathbf{u}, v) = \hat{\lambda}_1(\mathbf{u})\hat{\lambda}_2(v)$ respectively. For a Monte Carlo test, given the observed pattern X, we may estimate F simulate n point patterns $\{X_i\}_{i=1}^n$ with separable intensity. For each simulated pattern X_i we estimate the separable and non-separable intensities and simulate m patterns $\{X_{ij}\}_{j=1}^m$ with such separable intensities. We calculate F_{ij} for each pattern and obtain the Monte Carlo p-value of the test as

$$\hat{p} = \frac{1 + \sum_{i=1}^n \mathbf{1}\left[\frac{1 + \sum_{j=1}^m \mathbf{1}[F_{ij} \geq F_i]}{m+1} \leq F\right]}{n+1}.$$

The test of nominal size α rejects $\mathscr{H}_0$ if $p \leq \alpha$.

It is demonstrated that kernel selection is not as determinant as the bandwidth. Research in this field is still ongoing and there is no unique best recommendation. Although in the spatial context there are many available methods (Baddeley et al., 2015; Diggle, 2013; Illian et al., 2008), in the spatio-temporal context bandwidth is a subject not enough explored in the literature. In the paper we have used a rule of thumb implemented in `spatstat` package (Baddeley et al., 2015) for the spatial kernel. For the temporal part, we have used the optimum bandwidth of Sheather and Jones (1991). It should be noted that all the conclusions made throughout this paper depend on the estimated intensity function and therefore, on the choice of the bandwidths.

We have set the spatial region in such a way that we focus on a middle region with the largest number of tornadoes per area, see Fig.4.3. This selection of the region is motivated by the strong inhomogeneity shown in the complete map of the U.S. Considering the whole region for the analyses can lead to errors in the tests due to the huge variance of the descriptors.

Appendix

Theorem 5 *Let $E = W \times T \subset \mathbb{R}^2 \times \mathbb{R}$ be a spatio-temporal window and $\mathbf{X}$ a spatio-temporal process observed within E. Multiple points are not allowed, so each point of $\mathbb{R}^2 \times \mathbb{R}$ occurs at most once. Let $\mathbf{X}$ be a spatio-temporal binomial process, i.e., a stationary point process with a fixed number of points n, then*

$$\mathbb{V}ar_n(\hat{K}(r,t)\,|N(E) = n) \propto \frac{r^2 t}{n^2}. \tag{4.14}$$

Sketch of the proof. We mainly extend the proof provided by Ripley (1988) for the planar case by including time as an additional dimension and making use of his numerical approximations. Let $\xi_i = (\mathbf{u}_i, v_i), \xi_j = (\mathbf{u}_j, v_j) \in \mathbf{X}$ for $i, j = 1, ..., n$. Consider the estimator of $K(r,t)$ given by (4.4) for a constant spatio-temporal intensity $\lambda = n/|E|$ and isotropic spatio-temporal edge corrections $e^{\mathrm{S}}(\mathbf{u}_i, \mathbf{u}_j) e^{\mathrm{T}}(v_i, v_j) = e^{\mathrm{S}}_{ij} e^{\mathrm{T}}_{ij}$, then

$$\hat{K}(r,t) = \frac{|E|}{n(n-1)} \sum_{i \neq j} \mathbf{1}\left[||\mathbf{u}_i - \mathbf{u}_j|| \leq r\right] \mathbf{1}\left[|v_i - v_j| \leq t\right] e^{\mathrm{S}}_{ij} e^{\mathrm{T}}_{ij}. \tag{4.15}$$

Note that we can write (4.15) as

$$\hat{K}(r,t) = \frac{|E|}{n(n-1)} \sum_{i \neq j} \mathbf{1}\left[||\mathbf{u}_i - \mathbf{u}_j|| \leq r\right] \mathbf{1}\left[|v_i - v_j| \leq t\right] \tfrac{1}{2}\left(e^{\mathrm{S}}_{ij} e^{\mathrm{T}}_{ij} + e^{\mathrm{S}}_{ji} e^{\mathrm{T}}_{ji}\right).$$

Let

$$\psi(\xi_i, \xi_j) = \mathbf{1}\left[||\mathbf{u}_i - \mathbf{u}_j|| \leq r\right] \mathbf{1}\left[|v_i - v_j| \leq t\right] e^{\mathrm{S}}_{ij} e^{\mathrm{T}}_{ij} = \psi^{\mathrm{S}}(\mathbf{u}_i, \mathbf{u}_j)\, \psi^{\mathrm{T}}(v_i, v_j)$$

and

$$\phi(\xi_i, \xi_j) = \tfrac{1}{2}\left\{\psi(\xi_i, \xi_j) + \psi(\xi_j, \xi_i)\right\}.$$

Note that $\phi(\xi_i, \xi_j)$ is a symmetric function by construction thus (4.15) becomes

$$\hat{K}(r,t) = \frac{|E|}{n(n-1)} \sum_{i \neq j} \phi(\xi_i, \xi_j) = \frac{|E|}{n(n-1)} \Upsilon. \tag{4.16}$$

Then, using factorial moment measures $\alpha^{(k)}$ (see e.g. Daley and Vere-Jones, 2003),

$$\mathbb{E}(\Upsilon) = \int_{E^2} \phi(\xi, \eta)\, d\alpha^{(2)}(\xi, \eta) \tag{4.17}$$

and

$$
\begin{aligned}
\mathbb{E}\left(\Upsilon^2\right) &= \int_{E^4} \phi\left(\xi,\eta\right)\phi\left(\zeta,\vartheta\right)\mathrm{d}\alpha^{(4)}\left(\xi,\eta,\zeta,\vartheta\right) \\
&\quad +4\int_{E^3}\phi\left(\xi,\eta\right)\phi\left(\xi,\zeta\right)\mathrm{d}\alpha^{(3)}\left(\xi,\eta,\zeta\right) \\
&\quad +2\int_{E^2}\phi\left(\xi,\eta\right)^2\mathrm{d}\alpha^{(2)}\left(\xi,\eta\right).
\end{aligned}
$$

Let ν_d denote the Lebesgue measure on $\mathbb{R}^d$. Then for a spatio-temporal binomial process,

$$
\alpha^{(k)} = n(n-1)\cdots(n-k+1)|E|^{-k}\nu_{2(k+1)}
$$

and we define

$$
S = \int_{E^2}\phi\left(\xi,\eta\right)\mathrm{d}\xi\mathrm{d}\eta,
$$

$$
S_1 = \int_{E^3}\phi\left(\xi,\eta\right)\phi\left(\xi,\zeta\right)\mathrm{d}\alpha^{(3)}\left(\xi,\eta,\zeta\right) = \int_E\left(\int_E\phi\left(\xi,\eta\right)d\xi\right)^2\mathrm{d}\eta
$$

and

$$
S_2 = \int_{E^2}\phi\left(\xi,\eta\right)^2\mathrm{d}\xi\mathrm{d}\eta.
$$

For a binomial process,

$$
\mathbb{V}\mathrm{ar}_n\left(\hat{K}(r,t)\right) = \frac{2n-1}{n^3}\left(\frac{2(n-2)S_1}{|E|} - \frac{(2n-3)S^2}{|E|^2} + S_2\right). \tag{4.18}
$$

Since $\hat{K}(r,t)$ is an unbiased estimator for $K(r,t)$ (see Gabriel, 2014), then we have

$$
2\pi r^2 t = \mathbb{E}_n\left(\frac{|E|}{n(n-1)}\Upsilon\right) = \frac{|E|}{n(n-1)}\left(\frac{n(n-1)}{|E|^2}\right)S,
$$

where $\mathbb{E}_n$ is the conditional expectation given n events in the region E. So $S = 2\pi r^2 t|E|$.

For further cases we restrict our attention to the main terms of the integrals in order to simplify the proof. Define

$$
E^* = W^* \times T^* = \left\{\mathbf{u}\in W \,|\, b(\mathbf{u},r)\subset W\right\} \times \left\{v\in T \,|\, b(v,t)\subset T\right\},
$$

where b denotes, indistinctly, a ball in $\mathbb{R}^2$ or $\mathbb{R}$. Note that in W^*, $e^2(\mathbf{u}_i, \mathbf{u}_j) = 1$, and the same happens with $e^1(t_i, t_j)$ in T^*. Note also that

$$E = W \times T = (W^* \times T^*) \cup (W^* \times (T \setminus T^*)) \cup ((W \setminus W^*) \times T^*) \cup ((W \setminus W^*) \times (T \setminus T^*)).$$

Following Ripley (1988) we assume that the boundaries are straight. Then $v(W^*) \approx |W| - |\partial W| r$, where $|\partial W|$ stands for the perimeter length of W. For the time, and considering that we have only two points and that the correction only assigns 1 or 2, we assume that $v(T^*) \approx |T|$. Note that these conditions imply that we can avoid the further integrals in every region related with $T \setminus T^*$ in E.

For S_1 we have

$$
\begin{aligned}
S_1 &= \int_E \left(\int_E \phi(\xi, \eta) \, d\xi \right)^2 d\eta \\
&= \int_E \left(\int_E \phi(\xi, \eta) \, d\xi - 2\pi r^2 t \right)^2 d\eta \\
&\quad + 4|E|\pi r^2 t \int_{E^2} \phi(\xi, \eta) \, d\xi d\eta - |E| \left(2\pi r^2 t \right)^2 \\
&= 4\pi^2 r^4 t^2 |E| + \int_E \left(\int_E \phi(\xi, \eta) \, d\xi - 2\pi r^2 t \right)^2 d\eta,
\end{aligned}
$$

and $(\cdots)$ is non-zero only when $\mathbf{u} \notin W^*$ or $v \notin T^*$, so

$$
\begin{aligned}
S_1 &= 4\pi^2 r^4 t^2 |E| + \int_T \int_{W \setminus W^*} (\cdots)^2 d\mathbf{u} dv \\
&\approx 4\pi^2 r^4 t^2 |E| + 4\pi^2 r^5 t^2 |\partial W||T| \int_0^2 \left(\int_W \chi(h, \mathbf{s}) \, d\mathbf{s} - 1 \right)^2 dh \\
&\approx 4\pi^2 r^4 t^2 |E| \left(1 + 0.0066 \frac{|\partial W||T|r}{|E|} \right)
\end{aligned}
$$

where $\chi(h, \mathbf{s}) = \psi^{\mathrm{S}}(\mathbf{u}_h, \mathbf{s})$ for a point $\mathbf{u}_h$ distance h from the boundary. Also for S_2 we have

$$
\begin{aligned}
S_2 &= \int_{E^2} \phi(\xi, \eta)^2 \, d\xi d\eta = \tfrac{1}{2} \int_{E^2} \psi(\xi, \eta)^2 \, d\xi d\eta + \tfrac{1}{2} \int_{E^2} \psi(\xi, \eta) \psi(\eta, \xi) \, d\xi d\eta \\
&= \tfrac{1}{2} \int_{W^2} \psi^{\mathrm{S}}(\mathbf{u}, \mathbf{s}) \, d\mathbf{u} d\mathbf{s} \int_{T^2} \psi^{\mathrm{T}}(v, l) \, dv dl \\
&\quad + \tfrac{1}{2} \int_{W^2} \psi^{\mathrm{S}}(\mathbf{u}, \mathbf{s}) \psi^{\mathrm{S}}(\mathbf{s}, \mathbf{u}) \, d\mathbf{u} d\mathbf{s} \int_{T^2} \psi^{\mathrm{T}}(v, l) \psi^{\mathrm{T}}(l, v) \, dv dl.
\end{aligned}
$$

Since $\psi^{\mathrm{S}}(\mathbf{u},\mathbf{s}) = 1$ unless $\mathbf{s}$ is within distance $2r$ of ∂W, and taking into account the approximation of the temporal component, we have that

$$
\begin{aligned}
S_2 \;\approx\;& 2t|T|\left\{\pi r^2(|W| - 2|\partial W|r)\right.\\
&\left. +\pi r^3|\partial W|\left(1 + \int_0^1 \chi(h,\mathbf{s})\,\mathrm{d}\mathbf{s}\,\mathrm{d}h + \int_0^2 \chi(h,\mathbf{s})\chi(\mathbf{s},h)\,\mathrm{d}\mathbf{s}\,\mathrm{d}h\right)\right\}\\
\approx\;& 4\pi r^2 t|E|\left(1 + 0.305\frac{|\partial W||T|r}{|E|}\right).
\end{aligned}
$$

Finally, collecting the pieces for S, S_1, S_2 and from (4.18) we have

$$
\mathbb{V}\mathrm{ar}_n\left(\hat{K}(r,t)\,|\,N(E) = n\right) \approx \frac{4\pi r^2 t|E|}{n^2}\left[1 + 0.305\frac{|\partial W||T|r}{|E|} + 0.0414n\frac{|\partial W||T|r^3 t}{|E|^2}\right].\ \square
$$

Chapter 5

Factorial experiments for spatial point patterns in minerals Engineering

Abstract

Factorial experiments allow for evaluating combined effects of two or more experimental variables and their interactions. Here we are interested in factorial experiments when the observations are spatial point patterns rather than real-valued random variables. In particular, we develop a number of random permutation techniques to test main effects and interactions in a two-way ANOVA design under orthogonal experiments, where the observations correspond to observed functional second-order summary statistics such as Ripley's K-function. We conduct simulation experiments in order to demonstrate the statistical performance of the new statistics that we propose for the analysis of variance. Our method works well compared to existing methods.

Our work is motivated by the analysis of the spatial distribution of bubbles under three specific frother concentrations and three levels of volumetric air flow, in a flotation experiment. We analyse if there are significant main and interaction effects of these two factors, keeping all other physico-chemical and technological variables fixed when explaining the spatial patterns of bubbles.

Keywords   *Bootstrap hypothesis testing; Flotation bubble data; Frother concentration; K-function; Replicated point patterns; Volumetric air flow*

5.1   Introduction and data

The cooper production process consists of a large number of steps that allow to obtain cathodes from a mineral deposit. Once the ore has been extracted from a mine, it passes through successive stages of a comminution process. In the case of copper oxide ores, the subsequent processing steps correspond to hydrometallurgy and electrometallurgy. For sulphide ores, after the comminution, the processing requires steps of concentration by flotation and pyrometallurgy (Schlesinger et al., 2011, Chapter 3).

   The flotation process, as part of the production chain of copper sulphide ores, allows to separate mineral particles from non-metallic gangue particles. This is achieved by taking advantage of differences in hydrophobicity of the two species: the comminuted material is deposited in reactors (flotation cells) forming a pulp with water and some reagents. The particles of mineral which adhere to air bubbles and float towards a foam phase generated in the top of the cell, are recovered by injecting air bubbles (Schlesinger et al., 2011, Chapter 3).

   Contrary to sulphide minerals, the non-metallic gangue particles are not hydrophobic, so they are not collected by the air bubbles. It should be noticed that the flotation recovers gangue mineral found in mixed particles, i.e. particles with both mineral and gangue. From the foam phase, sulphides rich particles are separated from the non-metallic gangue, which is removed as part of the pulp in the bottom of the cells. The process is controlled by using several reagents: lime to regulate the pH, collectors favouring bonding stability bubble-particle, and foam which favours the formation of smaller air bubbles, which are the most efficient in the mineral particles collection.

   In order to quantify the hydrodynamic performance within flotation cells, a series of so-called *gas dispersion characteristics* are defined, namely, *gas holdup*

(volumetric fraction of gas in a gas-slurry mix, ε_g), *superficial gas velocity* (volumetric gas flowrate per cross-sectional area of a cell, J_g), *bubble size distribution* (*BSD*, characterised by a random bubble diameter d_b), and the *derived parameter bubble surface area flux* ($S_b = 6J_g/d_b$, where d_b is usually the Sauter mean diameter d_{32}) (Gómez and Finch, 2007; Nesset et al., 2006). S_b represents the flux of bubble surface area per cross-sectional area in the flotation machine, and it is related to the flotation rate constant (Gorain et al., 1997, 1999; Hernandez-Aguilar et al., 2004).

The bubble size distribution is a very complicated variable to be measured (Kracht et al., 2013). One of the sampling techniques for measuring it is the *sampling-followed-by-imaging* (Gómez and Finch, 2007 and Miskovic and Luttrell, 2012). Some further approaches have been applied in the literature to analyse *BSD* by using methods from stochastic geometry such as Boolean models (Emery et al., 2012; Kracht et al., 2013). These approaches lead to calculate, for instance, the diameters d_{32} and the *BSD* directly from the binary image by taking advantage of the assumption of complete spatial randomness, which is usually taken for granted.

High resolution images (170 pixels/mm) were recorded by using a Nikon D-5100 photo camera with a macro lens of 60mm, the camera was adjusted manually always with the aim of achieving the same dimensions in all the photographs of the experiment. Thus images were obtained with a rectangular box of exactly the same size for each pattern with dimensions 29.0mm × 19.2mm. Figure 5.1 shows two images of bubbles generated in a flotation machine.
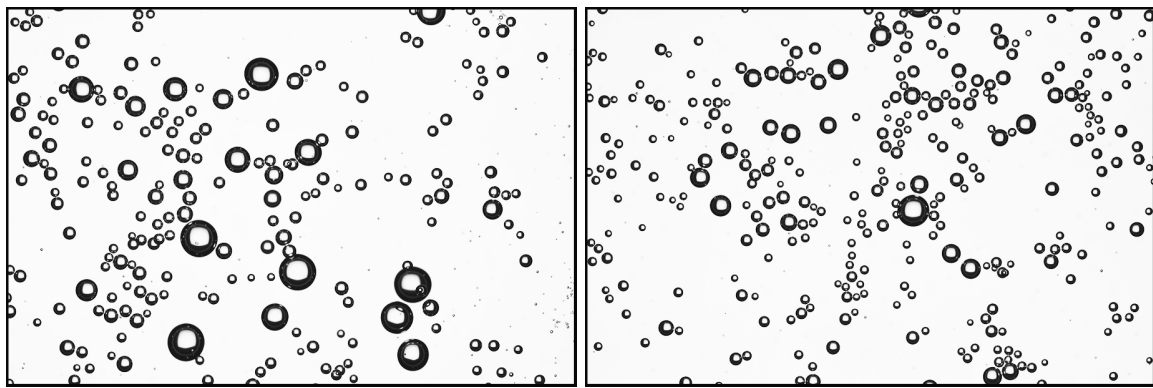
Figure 5.1 Two images of bubbles in a flotation machine.

When the bubble images have been successfully recorded through the procedure described above, they have been processed through classic image analysis

([Kracht et al., 2013](#)). The analysis of such images is performed to identify bubbles and some characteristics associated with them such as the location of their centroids, diameters, areas, etc. Our dataset consists of 54 images containing a total of 8385 floating bubbles. The images of bubbles can be regarded as spatial point patterns where locations are the centroids of the bubbles and where we can consider the further characteristics as quantitative marks. An example of a typical bubble point pattern is displayed in Figure 5.2. Flotation cell performance
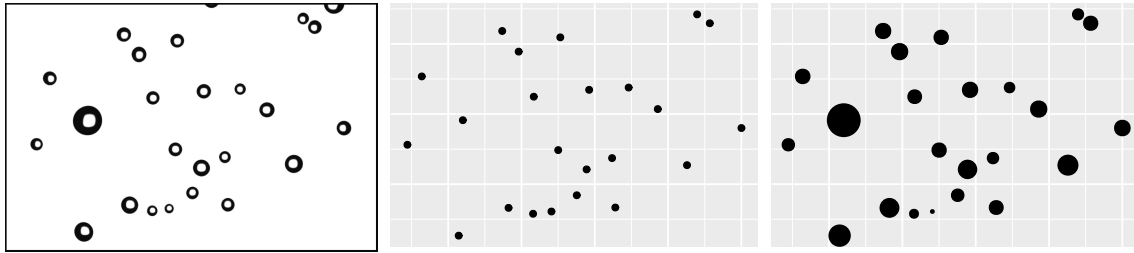
Figure 5.2 Two point patterns of bubbles in a flotation cell where points are centres of bubbles (left) and the bubbles size is attached to centres locations as marks (right).

is strongly dependent on a variety of important operating and design factors. For instance, a proper combination of gas rate and bubble size is required to provide a considerable gas holdup in the flotation pulp ([Miskovic and Luttrell, 2009](#)). The volumetric air flow ($L\min^{-1}$) and the specific frother concentration (ppm) are two factors that could influence the physical properties ([Gómez et al., 2016](#); [Laskowski, 2001](#)).

We are interested in these two factors that are potentially influential in the spatial distribution of the bubbles. A relevant question is if there is a significant interaction between volumetric air flow and frother concentration, keeping all other physico-chemical and technological variables fixed to explain the spatial pattern. Thus in the dataset we have three frother concentration levels $5\mathrm{ppm}, 10\mathrm{ppm}$ and $15\mathrm{ppm}$, as well as three volumetric air flowrate levels $5\mathrm{L}\min^{-1}, 8\mathrm{L}\min^{-1}$ and $10\mathrm{L}\min^{-1}$. Additionally, we have 6 replicates (point patterns) at each combination of levels of such factors. The treatment combinations of the experiment as well as the observed bubble point patterns are represented in Figure 5.3. In classical statistics, factorial experiments allow to evaluate the combined effects of two or more experimental variables. The information obtained in this type of experiments is much more complete than that obtained through a series of single-factor experiments, since factorial experiments allow the study of the interaction of the factors. The problem we are facing here is considering factorial experiments
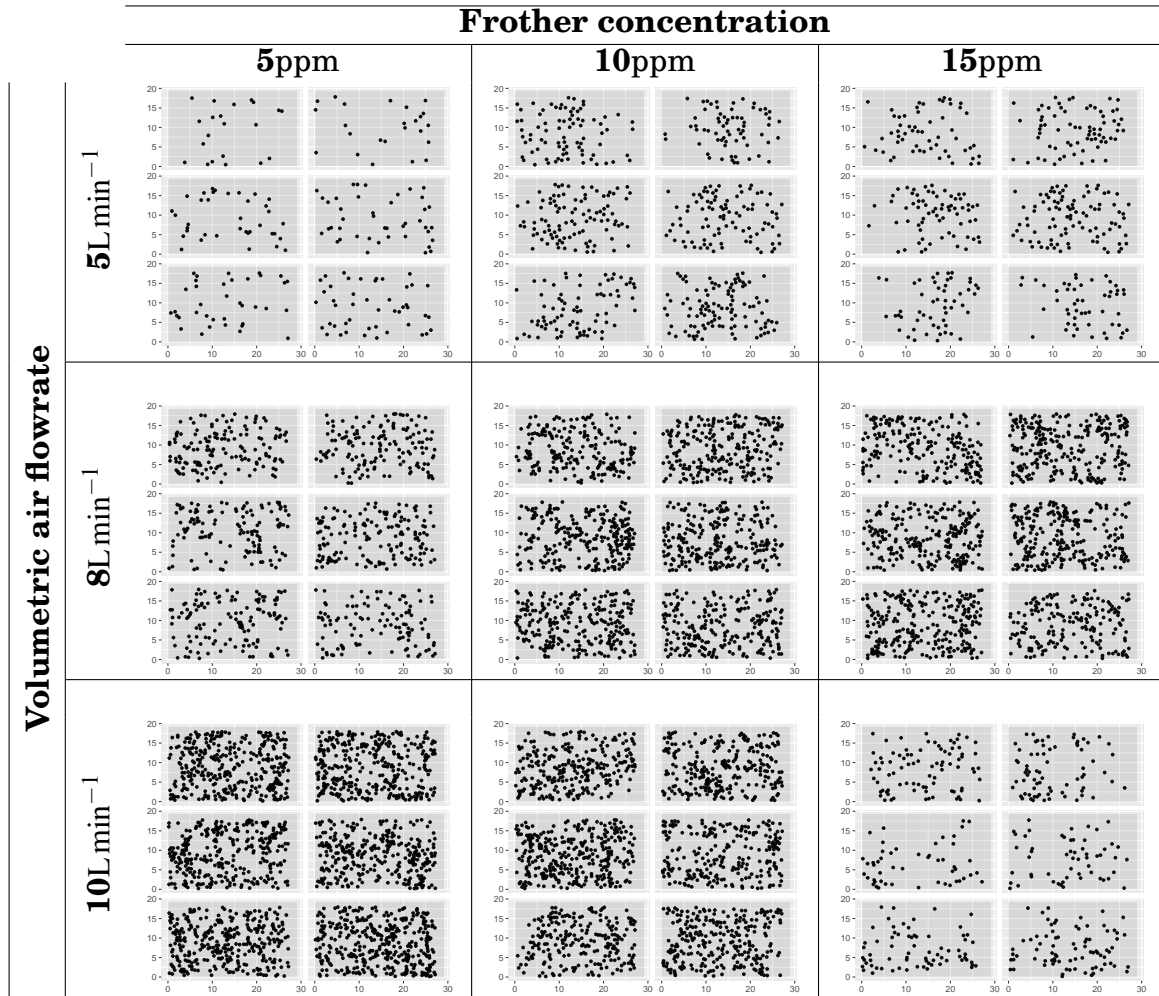
Figure 5.3 General arrangement for patterns of locations (centres) of floating bubbles from three concentration levels of certain type of frother and three volumetric air flow levels.

when the observations are spatial point patterns rather than numeric random variables.

The analysis of this type of experiments when the observations are not quantitative variables is in its infancy, specially in the field of point processes. A few authors have treated the observations in an experiment when they are point patterns (Baddeley et al., 1993; Diggle et al., 1991, 2000; Hahn, 2012; Hahn and Vedel Jensen, 2016), and their studies have concentrated mainly on the comparison of several groups of patterns (a single factor). Some other authors have included, for example, non-spatial variables or mixed effects in their models (Bagchi and Illian, 2015; Landau and Everall, 2008; Myllymäki et al., 2014). Finally, only a few works have been focused on factorial experiments, for instance

Ramón et al. (2016) have tried to extend these methods to the case of two factors but from a very practical approach.

The purpose of this paper is twofold. One is to develop a set of techniques that help us to answer the set of questions coming from the bubble data of the flotation experiment. The second one is to provide a methodology for the analysis of variance in the case of two factors in balanced experiments when observations are point patterns. Section 5.2 presents a necessary background. Section 5.3 develops the methodology for factorial experiments for spatial point patterns. In particular, we provide a new set of Fisher's test statistics as well as some new residuals scheme useful for inference. The bubbles dataset is analysed in Section 5.4. The paper ends with a discussion.

5.2 Background and Set-up

Throughout this paper, we assume that every subset A of $\mathbb{R}^2$ is measurable and any function f defined on $\mathbb{R}^2$ is integrable, see Daley and Vere-Jones (2003, 2007) for details. Let T be a finite interval, and $||f||$ be the L^2-norm of a measurable function $f(r), r \in T$, i.e.

$$||f|| = \sqrt{\int f^2}.$$

Provided that $||f|| < \infty$, f is said to be a squared integrable function and it belongs to the Hilbert space formed by all squared integrable functions over T. We use the notation $|| \cdot ||$ indistinctly to denote the Euclidean vector norm and the L^2-norm. Let $y(r), r \in T$ be a stochastic process with covariance function $\gamma(r,s), r,s \in T$. Consider the two-dimensional Euclidean space $\mathbb{R}^2$. A planar point process $W \subset \mathbb{R}^2$ is a random, finite or countable collection of points $\mathbf{X}$ with no accumulation points. A realisation X of a point process can be considered as a finite subset $\{\mathbf{u}_i\}_{i=1}^n \subset W$. Let $N(A)$ be the number of points of X in $A \subseteq W$. For real functions f defined on $\mathbb{R}^2$, we have $\sum f(\mathbf{u}_i) = \int f(\mathbf{u})N(d\mathbf{u})$.

5.2.1 Second-order summary statistics

We assume here that the point processes are stationary, so we use Ripley's K-function (Ripley, 1977) in its homogeneous version, given by

$$\mathscr{K}(r) = \frac{1}{\lambda}\mathbb{E}\left[\sum_{\mathbf{u}_i \in X} \mathbf{1}\{||\mathbf{u} - \mathbf{u}_i|| \le r\} \middle| \mathbf{u} \in \mathbf{X}\right], \qquad (5.1)$$

provided that this value does not depend on the choice of the location $\mathbf{u}$ (see e.g. Baddeley et al., 2015; Møller and Waagepetersen, 2004) and where $||\cdot||$ represents the Euclidean distance and λ is the first-order intensity of $\mathbf{X}$. In this context we set $a/0 = 0$ and $r \in T = [0, r_0]$. For homogeneous Poisson processes (Diggle, 2013), the K-function is $K(r) = \pi r^2$. Note that the the homogeneous Poisson processes are the archetype of complete spatial random processes (hereinafter CSR). A natural estimator for the K-function is given by

$$\hat{K}(r) = \frac{1}{\hat{\lambda}^2 |W|} \sum_{i=1}^{n} \sum_{j \neq i} \mathbf{1}\{||\mathbf{u}_i - \mathbf{u}_j|| \leq r\} e(\mathbf{u}_i, \mathbf{u}_j; r), \tag{5.2}$$

where $|\cdot|$ is, in this case, the set size, $e(\mathbf{u}, \mathbf{v}; r)$ is an edge-correction weight (see e.g. Baddeley et al., 2015; Ripley, 1988), and $\hat{\lambda}$ is an estimator of the first-order intensity function; λ^2 is usually estimated by

$$\hat{\lambda}^2 = \frac{n(n-1)}{|W|^2} \tag{5.3}$$

where n is the number of points of the pattern X.

5.2.2 Pooled estimators

If a summary-statistic function, usually given by a ratio $Y(r) = \mathbb{E}[U(r)]/\mathbb{E}[V(r)]$, is calculated for a set of n independent observations of a point process $\mathbf{X}$, then, according to Baddeley et al. (1987, 1993, 2015) (and references therein), the *pooled* summary-statistic across replicates, which is unbiased for $Y(r)$ and has minimum variance (given the denominators $V_i(r)$), is the weighted average of the individual ratios with weights proportional to $V_i(r)$, i.e.

$$\bar{Y}(r) = \frac{\sum_i \hat{U}_i(r)}{\sum_i \hat{V}_i(r)} = \frac{\sum_i \hat{V}_i(r) \hat{Y}_i(r)}{\sum_i \hat{V}_i(r)}.$$

If our summary-statistic is Ripley's K-function, the estimator given in Eq. (5.2) using the isotropic or the translation edge-correction factors, may be seen as the ratio of two estimators; the numerator is a sum over all pairs of data points and the denominator corresponds to the number of pairs of distinct points, i.e. $\hat{V}_i(r) = n_i(n_i - 1)$ where n_i is the number of points in the ith point pattern (see e.g Baddeley et al. (2015)). Thus the pooled K-function estimator for a sample $\{\hat{K}_i\}_{i=1}^{m}$

coming from m point patterns $\{X_i\}_{i=1}^m$, is given by

$$\bar{K}(r) = \frac{1}{\omega_{\cdot}} \sum_{i=1}^m \omega_i \hat{K}_i(r), \quad r \in T, \tag{5.4}$$

where

$$\omega_i = n_i(n_i - 1), \tag{5.5}$$

and $\omega_{\cdot} = \sum_{j=1}^m n_i(n_i - 1)$. For the sample covariance function $\hat{\gamma}(r,s)$ we have

$$\hat{\gamma}(r,s) = \widehat{\text{Cov}}\{\hat{K}(r), \hat{K}(s)\} = \frac{1}{m-1} \sum_{i=1}^m \left[\hat{K}_i(r) - \bar{K}(r)\right] \left[\hat{K}_i(s) - \bar{K}(s)\right], \quad r,s \in T. \tag{5.6}$$

5.3 Factorial experiment procedures for point patterns

In this section we extend the methodology of ANOVA procedures for general functional data provided by Zhang (2013), to the particular case of functional descriptors associated with replicated spatial point patterns.

5.3.1 One-sample tests

We start with a simple example meant to motivate our further analysis. Suppose that we have a functional sample of estimated K-functions $\{\hat{K}_i\}_{i=1}^m$ coming from m point patterns $\{X_i\}_{i=1}^m$, which in turn, are realisations of the same underlying point process $\mathbf{X}$ with unknown K-function $\mathscr{K}(r)$. In this case we want to test the following hypothesis

$$\mathscr{H}_0 : \mathscr{K}(r) = K_0(r), r \in T \quad \text{versus} \quad \mathscr{H}_1 : \mathscr{K}(r) \neq K_0(r), \text{ for some } r \in T, \tag{5.7}$$

where $K_0(r)$ is some known function by virtue of properties of the underlying process $\mathbf{X}$. For example, if we want to test the hypothesis of complete spatial randomness of the point process, we could extrapolate the technique used by Diggle (1986) (improved in Diggle, 2013, p.77), and exploit the fact of having replicates. Since in the Poisson case $K_0(r) = \pi r^2$, we can then define the following statistic to test (5.7)

$$\Delta(r) = \sqrt{m} \left[\bar{K}(r) - \pi r^2\right], \quad r \in T, \tag{5.8}$$

as a *pivotal function*, where $\bar{K}(r)$ is given in Eq. (5.4). A L^2-norm-based test uses the squared L^2-norm of $\Delta(r)$ as its test statistic

$$D_m = \left\| \frac{\Delta(r)}{r} \right\|^2 = m \int_T \frac{1}{r^2} \left[\bar{K}(r) - \pi r^2 \right]^2 dr, \qquad (5.9)$$

where the integration domain is some compact set $T \subset (0, r_0]$. Note that in this statistic, the term $1/r^2$ is thought to stabilise the variance of $\hat{K}_i(r)$, which is roughly proportional to r^2 (Ripley, 1988) in the case of CSR patterns. However, in the general case, it is analytically intractable so we can use the variance $\hat{\sigma}^2(r) = \hat{\gamma}(r, r)$ of the sample of K-functions as an adjustment factor and build a Fisher-test type statistic (in the sense of Hahn, 2012), as follows

$$F_m = \frac{||\Delta||^2}{||\sigma||^2}. \qquad (5.10)$$

5.3.2 One-way ANOVA

In the literature of point processes, some procedures have been developed to compare the expected values of second-order summary statistics of two or more samples or groups (see e.g Baddeley et al., 1993; Diggle et al., 1991, 2000; Hahn, 2012). To observe in a natural way these comparisons, we can consider that the different samples of point patterns are grouped by a categorical variable defined as a *factor* in classical linear models, whose different values are called *factor levels*. In this section we try to give some slight modifications to the statistics presented in the literature and establish a unified notation.

We assume that we have m independent samples (groups) $\{\hat{K}_{ij}\}, j = 1, \cdots, m_i, i = 1, \cdots, m$, with $\{\mathcal{K}_i\}_{i=1}^m$ the unknown group mean K-functions and $\{\gamma_i\}$ are the covariance functions. The main idea is to check this null hypothesis

$$\mathcal{H}_0 : \mathcal{K}_1(r) = \mathcal{K}_2(r) = \cdots = \mathcal{K}_m(r), \quad r \in T. \qquad (5.11)$$

The group weighted mean function of the m groups is given by

$$\bar{K}_{i\cdot}(r) = \frac{1}{\omega_{i\cdot}} \sum_{j=1}^{m_i} \omega_{ij} \hat{K}_{ij}(r), \quad i = 1, \ldots, m, \qquad (5.12)$$

where the weights ω_{ij} are defined as in Eq. (5.5). Note that we could make the convenient assumption of equal covariance functions between groups, and

estimate the covariance jointly. However, for practical purposes and due to the delicate structure of the variance function of Ripley's K-function (Ripley, 1988), we directly estimate covariances separately based only on each group (sample). Note also that under the independence of the patterns, the means of the groups remain independent. We may define the *pointwise between-subject variation* as

$$\text{SSH}_m(r) = \sum_{i=1}^{m} m_i \left[\bar{K}_{i\cdot}(r) - \bar{K}_{\cdot\cdot}(r) \right]^2, \tag{5.13}$$

where $\bar{K}_{\cdot\cdot}(r)$ denotes the overall mean function of the m groups, i.e.

$$\bar{K}_{\cdot\cdot}(r) = \frac{1}{\omega_{\cdot\cdot}} \sum_{i=1}^{m} \omega_{i\cdot} \bar{K}_{i\cdot}(r), \tag{5.14}$$

where

$$\omega_{\cdot\cdot} = \sum_{i=1}^{m} \sum_{j=1}^{m_i} n_{ij}(n_{ij} - 1). \tag{5.15}$$

In this case we can define a L^2-norm-based test through the test statistic given by

$$D_m = \int_T \frac{\text{SSH}_m(r)}{r^2} \mathrm{d}r = \sum_{i=1}^{m} m_i \int_T \frac{1}{r^2} \left[\bar{K}_{i\cdot}(r) - \bar{K}_{\cdot\cdot}(r) \right]^2 \mathrm{d}r.$$

Notice that the statistic D_m is Diggle's statistic (Diggle et al., 1991, 2000). Similarly, we can define a statistic that takes into account the variability of particular groups in the sample, i.e. where the patterns are observed prior to any treatment. Such statistic is the *pointwise within-subject variation* and it is given by

$$\text{SSE}_m(r) = \sum_{i=1}^{m} \left(1 - m_i/m \right) \hat{\sigma}_i(r).$$

We could include the last term into a Fisher-test type statistic in order to include the sampling variations of the variance functions $\hat{\sigma}_i, i = 1, \ldots, m$. In this case Fisher's statistic is

$$F_m = \frac{\int_T \text{SSH}_m(r) \mathrm{d}r}{\int_T \text{SSE}_m(r) \mathrm{d}r}. \tag{5.16}$$

5.3.3   Balanced two-way ANOVA

We have seen in the previous section a model for the analysis of variance with one factor whose categories are used to classify the actions of a spatial descriptor of a

point pattern (in our case the *K*-function) in groups whose functional means are compared. In this section, the model is extended to include two factors instead of only one.

This inclusion is far from trivial, the problem of comparing groups of means of functional descriptors induced by two factors is not solved by simply extending the model of one factor. It is possible that both factors are fixed, both random, or one fixed and one random, and it could be the case that both factors act independently or that the combined action of two factors enhances or inhibits the action of each other in the response function. In the latter case we say that there is interaction, and we might consider an interaction model. Note that this interaction is not the interaction between points within a point process. When every category of one factor co-occurs in the design with every category of the other factor, then we talk about a crossed model or a factorial model. A first approach of this model was proposed by Wilson (1998) by using Ripley's *K*-function and it was later applied by Ramón et al. (2016). Throughout this work we focus on a factorial model, involving only two factors.

Suppose that factor *A* has *a* levels and factor *B* has *b* levels. Each realisation or replicate contains all *ab* factorial combinations. In general, there are m_{ij} replicates in each level combination, even though the analysis for unequal $\{m_{ij}\}$ becomes much more complicated than for equal $\{m_{ij}\}$. In this paper we assume the same number of observations in the cells, i.e, $m_{ij} \equiv c$ for all $i = 1, \ldots, a$ and $j = 1, \ldots, b,$. We have a functional descriptor sample

$$\{\hat{K}_{ijk}\}, i = 1, \ldots, a, j = 1, \ldots, b, k = 1, \ldots, c.$$

The functional observations can be described by the model

$$\hat{K}_{ijk}(r) = \mathscr{K}_{ij}(r) + \varepsilon_{ijk}(r), r \in T, \tag{5.17}$$

where $\varepsilon_{ijk}(r)$ is a random error with mean zero. Note that in this case the covariance functions are supposed to be different across the level combinations and the functional samples are assumed to be independent. Therefore, model (5.17) is a heteroscedastic two-way ANOVA model.

For a two-way ANOVA, the mean $\mathscr{K}_{ij}(r)$ can be expressed in the form

$$\mathscr{K}_{ij}(r) = \mathscr{K}_0(r) + \tau_i(r) + \beta_j(r) + (\tau\beta)_{ij}(r), \qquad i = 1, \ldots, a, j = 1 \ldots, b, r \in T, \tag{5.18}$$

where $\mathcal{K}_0(r)$ is the overall mean effect, $\tau_i(r)$ is the effect of the ith level of the *row factor A*, $\beta_j(r)$ is the effect of the jth level of *column factor B*, $(\tau\beta)_{ij}(r)$ is the effect of the interaction between $\tau_i(r)$ and $\beta_j(r)$. Both factors are assumed to be fixed, and the factor effects are defined as deviations from the overall mean. If nothing more is stated about the decomposition (5.18), the components of the decomposition are not uniquely defined. Some restrictions over the decomposition are required (see e.g Zhang, 2013). We set

$$\sum_i \tau_i(r) = \sum_j \beta_j(r) = \sum_i (\tau\beta)_{ij}(r) = \sum_j (\tau\beta)_{ij}(r) = 0.$$

So, hereafter, we limit our discussion to equi-replicated orthogonal ANOVA. We are then interested in testing equality of factor effects (hereafter *treatments*), i.e, that there are no treatment effects in both row treatments

$$\mathcal{H}_0^A : \tau_1(r) = \ldots = \tau_a(r) = 0, r \in T,$$
$$\mathcal{H}_1^A : \tau_i(r) \neq 0, \text{ for some } i, \text{ and for some } r \in T, \quad (5.19a)$$

and column treatments

$$\mathcal{H}_0^B : \beta_1(r) = \ldots = \beta_b(r) = 0, r \in T,$$
$$\mathcal{H}_1^B : \beta_j(r) \neq 0, \text{ for some } j, \text{ and for some } r \in T, \quad (5.19b)$$

as well as testing if the main-effects are simultaneously zero,

$$\mathcal{H}_0^{AB} : \tau_i(r) = 0 \text{ and } \beta_j(r) = 0 \quad \text{ for all } i, j, \text{ and for } r \in T,$$
$$\mathcal{H}_1^{AB} : \text{at least one } \tau_i(r) \text{ or } \beta_j(r) \neq 0, \text{ for some } r \in T. \quad (5.19c)$$

Finally, we could be interested in determining whether row and column factors interact, so testing

$$\mathcal{H}_0^I : (\tau\beta)_{ij}(r) = 0 \quad \text{ for all } i, j, \text{ and for } r \in T,$$
$$\mathcal{H}_1^I : \text{at least one } (\tau\beta)_{ij}(r) \neq 0, \text{ for some } r \in T. \quad (5.19d)$$

The estimators of cell means and variance functions are well defined whenever $c > 1$. We consider the same pooled estimators defined in Section 5.2.2 including factors and their levels, so that we have cell weighted mean and covariance

functions given by

$$\bar{K}_{ij\cdot}(r) = \frac{1}{\omega_{ij\cdot}} \sum_{k=1}^{c} \omega_{ijk} \hat{K}_{ijk}(r), \qquad i = 1, \ldots, a, j = 1, \ldots, b, \qquad (5.20)$$

and

$$\hat{\gamma}_{ij}(r,s) = \frac{1}{c-1} \sum_{i=1}^{c} \left[\hat{K}_{ijk}(r) - \bar{K}_{ij\cdot}(r) \right] \left[\hat{K}_{ijk}(s) - \bar{K}_{ij\cdot}(s) \right], \qquad (5.21)$$

where the weights ω_{ijk} are defined as in Eq. (5.5), and the number of points per pattern is denoted by n_{ijk}, where k is the individual within the ij cell (sample) and $i = 1, \ldots, a$ and $j = 1, \ldots, b$, and $\omega_{ij\cdot} = \sum_{k=1}^{c} n_{ijk}(n_{ijk} - 1)$. As in the classical ANOVA two-way analysis, we define $\bar{K}_{i\cdot\cdot}, \bar{K}_{\cdot j\cdot}$ and $\bar{K}_{\cdots}$ as the corresponding row, column, and grand weighted average K-functions. Thus,

$$\bar{K}_{i\cdot\cdot}(r) = \frac{1}{\omega_{i\cdot\cdot}} \sum_{j=1}^{b} \omega_{ij\cdot} \bar{K}_{ij\cdot}(r), \qquad i = 1, \ldots, a,$$

$$\bar{K}_{\cdot j\cdot}(r) = \frac{1}{\omega_{\cdot j\cdot}} \sum_{i=1}^{a} \omega_{ij\cdot} \bar{K}_{ij\cdot}(r), \qquad j = 1, \ldots, b, \qquad (5.22)$$

$$\bar{K}_{\cdots}(r) = \frac{1}{\omega_{\cdots}} \sum_{i=1}^{a} \sum_{j=1}^{b} \omega_{ij\cdot} \bar{K}_{ij\cdot}(r),$$

where

$$\omega_{i\cdot\cdot} = \sum_{j=1}^{b} \omega_{ij\cdot}, \quad \omega_{\cdot j\cdot} = \sum_{i=1}^{a} \omega_{ij\cdot} \quad \text{and} \quad \omega_{\cdots} = \sum_{i=1}^{a} \sum_{j=1}^{b} \omega_{ij\cdot}. \qquad (5.23)$$

From (5.22), the estimators of the general mean, main and interaction effects are then

$$\hat{\mathcal{K}}_0(r) = \bar{K}_{\cdots}(r),$$
$$\hat{\tau}_i(r) = \bar{K}_{i\cdot\cdot}(r) - \bar{K}_{\cdots}(r),$$
$$\hat{\beta}_j(r) = \bar{K}_{\cdot j\cdot}(r) - \bar{K}_{\cdots}(r), \qquad (5.24)$$
$$\widehat{(\tau\beta)}_{ij}(r) = \bar{K}_{ij\cdot}(r) - \bar{K}_{i\cdot\cdot}(r) - \bar{K}_{\cdot j\cdot}(r) + \bar{K}_{\cdots}(r).$$

Analogous to the classical functional data analysis, let us consider some fixed $r \in T$ and let $\text{SST}(r)$ be the pointwise total-sum-of-squares, $\text{SSA}(r), \text{SSB}(r)$ are the main-effect pointwise sum-of-squares, respectively, and let $\text{SSI}(r)$ be the interaction-effect pointwise sum-of-squares. Finally, let $\text{SSE}(r)$ denote the pointwise sum-of-squares due to errors. Following the classical balanced two-way

ANOVA, we can define the following estimators

$$
\text{SSA}(r) = bc \sum_{i=1}^{a} [\bar{K}_{i\cdot\cdot}(r) - \bar{K}_{\cdots}(r)]^2 = bc \sum_{i=1}^{a} \hat{\tau}_i^2(r),
$$

$$
\text{SSB}(r) = ac \sum_{j=1}^{b} [\bar{K}_{\cdot j\cdot}(r) - \bar{K}_{\cdots}(r)]^2 = ac \sum_{j=1}^{b} \hat{\beta}_j^2(r),
$$

$$
\text{SSI}(r) = \sum_{i=1}^{a} \sum_{j=1}^{b} [\bar{K}_{ij\cdot}(r) - \bar{K}_{i\cdot\cdot}(r) - \bar{K}_{\cdot j\cdot}(r) + \bar{K}_{\cdots}(r)]^2 = c \sum_{i=1}^{a} \sum_{j=1}^{b} \widehat{(\tau\beta)}_{ij}^{\,2}(r),
$$

$$
\text{SSE}(r) = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} \left[\hat{K}_{ijk}(r) - \bar{K}_{ij\cdot}(r)\right]^2 = (c-1) \sum_{i=1}^{a} \sum_{j=1}^{b} \hat{\sigma}_{ij}^2(r).
$$

(5.25)

The corresponding L^2-norm-based test statistics for our null hypotheses $\mathscr{H}_0^A, \mathscr{H}_0^B, \mathscr{H}_0^{AB}$ and $\mathscr{H}_0^I$ respectively, are

$$
D^A = \int_T \frac{\text{SSA}_n(r)}{r^2}\mathrm{d}r = bc \sum_{i=1}^{a} \int_T \frac{\hat{\tau}_i^2(r)}{r^2}\mathrm{d}r,
$$

$$
D^B = \int_T \frac{\text{SSB}_n(r)}{r^2}\mathrm{d}r = ac \sum_{j=1}^{b} \int_T \frac{\hat{\beta}_j^2(r)}{r^2}\mathrm{d}r,
$$

$$
D^{AB} = \int_T \frac{1}{r^2}[\text{SSA}(r) + \text{SSB}(r)]\mathrm{d}r = D^A + D^B \quad \text{and}
$$

$$
D^I = \int_T \frac{\text{SSI}_n(r)}{r^2}\mathrm{d}r = c \sum_{i=1}^{a} \sum_{j=1}^{b} \int_T \frac{\widehat{(\tau\beta)}_{ij}^{\,2}(r)}{r^2}\mathrm{d}r.
$$

(5.26)

Similarly, for the Fisher-test type statistics for $\mathscr{H}_0^A, \mathscr{H}_0^B, \mathscr{H}_0^{AB}$ and $\mathscr{H}_0^I$, respectively, we have

$$
F^A = \frac{\int_T \text{SSA}(r)\mathrm{d}r/(a-1)}{\int_T \text{SSE}(r)\mathrm{d}r/(ab(c-1))},
$$

$$
F^B = \frac{\int_T \text{SSB}(r)\mathrm{d}r/(b-1)}{\int_T \text{SSE}(r)\mathrm{d}r/(ab(c-1))},
$$

$$
F^{AB} = \frac{\int_T [\text{SSA}(r) + \text{SSB}(r)]\,\mathrm{d}r/(a+b-2)}{\int_T \text{SSE}(r)\mathrm{d}r/(ab(c-1))},
$$

$$
F^I = \frac{\int_T \text{SSI}(r)\mathrm{d}r/((a-1)(b-1))}{\int_T \text{SSE}(r)\mathrm{d}r/(ab(c-1))}.
$$

(5.27)

Remark 6 *Note that the entire previous theoretical approach can be reproduced in identical conditions by replacing Ripley's K-function with its respective weighted mark version. Undoubtedly, the variance in that case becomes increasingly difficult*

to be controlled since the distribution of the marks implies an additional source of variation. However, the inclusion of marks constitutes a much more realistic treatment whenever the marks are dependent from locations. See further comments in Section 5.5.

5.3.4   Random permutation tests

The null distribution of our test statistics is not analytically tractable in any of the cases. We assume that the observed K-functions come from the statistical model (5.17). In order to determine the null distribution of our test statistics, we use the so-called *the bootstrap based on residuals* method. This method was proposed by Efron (1979), and has been applied in the point process context by e.g. Diggle et al. (1991).

Many permutation strategies can be applied for tests for individual terms in classic ANOVA designs. Anderson and Braak (2003) provide a complete guide to build accurate and approximate permutation strategies for all terms in a classic two-way ANOVA. Following this motivation, we consider inference through the choice of several types of exchangeable units under the null hypothesis (residual in our case of approximate permutation tests).

We provide results of Monte Carlo simulations to show the empirical level and the power of our tests. We consider some standard cases, namely Poisson (Complete Spatial Random model), Cluster (Aggregation model) and Inhibition (Regular model) processes. The parameters are chosen according to our questions related to the bubbles dataset, i.e. an equi-replicated design with two factors with three levels each one, equal observation windows (unit square window $[0,1] \times [0,1]$) and small sample sizes per level of each factor $c = 6$. We are interested in those situations in which the overall hypotheses hold true: the underlying point processes have the same K-function, meaning they have the same approximated spatial distribution and such distribution is not affected by the levels of factors. Recall that we are investigating the second-order behaviour, i.e. the patterns can have different intensities. The integrals are numerically approximated by using the trapezoidal rule. Additionally, we set $T = (0, r_0] = (0, 0.15]$ according to the results obtained by Hahn (2012). We perform each value through 1000 random permutations of observed stationary K-functions estimated by using Ripley's edge-correction. We consider several scenarios to the null hypothesis in order to verify the performance of the test under homoscedastic and heteroscedastic cases.

Poisson model We first consider complete spatial random patterns (**Poisson patterns**) with some configurations of homoscedasticity and heteroscedasticity given by the intensity of the process in each cell of the design, such configurations are shown in Table 5.1.

Homoscedastic configurations		Heteroscedastic configurations	
SConfig.1 $\lambda = \begin{pmatrix} 50 & 50 & 50 \\ 50 & 50 & 50 \\ 50 & 50 & 50 \end{pmatrix}$		DConfig.1 $\lambda = \begin{pmatrix} 50 & 50 & 50 \\ 100 & 100 & 100 \\ 150 & 150 & 150 \end{pmatrix}$	
SConfig.2 $\lambda = \begin{pmatrix} 100 & 100 & 100 \\ 100 & 100 & 100 \\ 100 & 100 & 100 \end{pmatrix}$		DConfig.2 $\lambda = \begin{pmatrix} 50 & 100 & 150 \\ 50 & 100 & 150 \\ 50 & 100 & 150 \end{pmatrix}$	
SConfig.3 $\lambda = \begin{pmatrix} 150 & 150 & 150 \\ 150 & 150 & 150 \\ 150 & 150 & 150 \end{pmatrix}$		DConfig.3 $\lambda = \begin{pmatrix} 60 & 120 & 180 \\ 80 & 140 & 200 \\ 100 & 160 & 220 \end{pmatrix}$	

Table 5.1 Arrangement of independent configurations of cell intensities in a two-way ANOVA design in the Poisson case.

Hard-core and Cluster models We also consider point patterns coming from *Matérn hard-core* and *cluster* point processes (see e.g Chiu et al., 2013 and references therein) with parameters leading to different degrees of regularity or clustering. Both types of point processes come from a parent Poisson point process. In this case we use a Matérn hard-core obtained by dependent thinning. The points are first marked with independent, identically distributed random numbers and all points that have a higher mark than any neighbours (further points) within the hard-core distance h are removed.

On the other hand, the Matérn cluster point process consists of independent clusters of daughter points around each parent point. The numbers of daughter points per cluster are Poisson distributed with mean μ, and the points are independently uniformly scattered in the ball with centre in the parent point and radius r. The parent points are not included in the observed point pattern. Some realisations of these two processes and CSR are shown in Figure 5.4. The bootstrap tests are briefly described as follows. We take model (5.17) as a starting point. Let $\mathbb{P}_\varepsilon(r)$ be the unknown distribution of $\varepsilon_{ijk}(r)$, such distribution is assumed to be centred at zero. We estimate $\hat{\mathbb{P}}_\varepsilon(r)$ as the sample probability distribution of the residuals $\hat{\varepsilon}_{ijk}(r)$ defined as
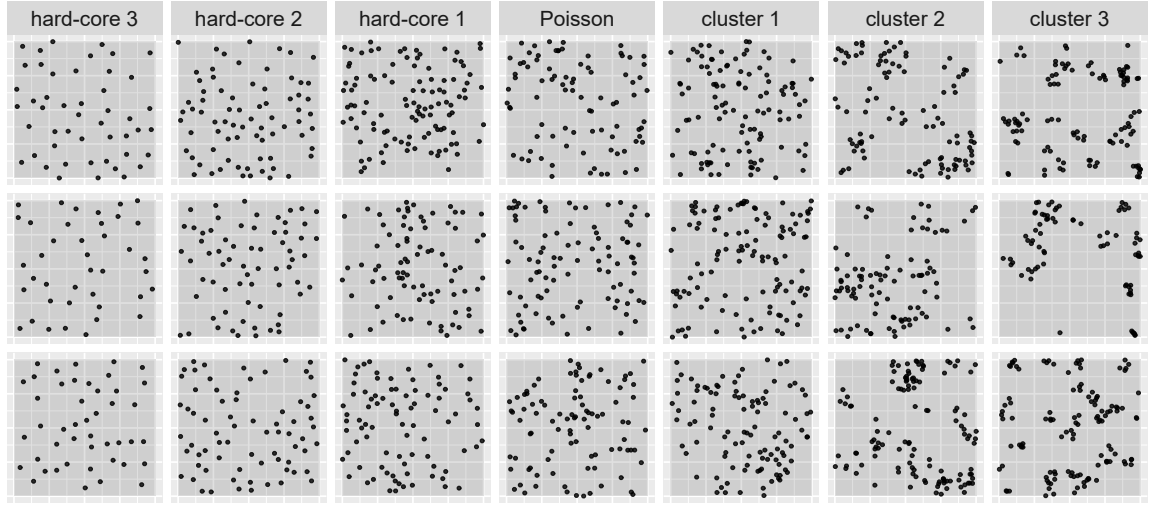
Figure 5.4 Sets of three independent realisations of Matérn hard-core, Poisson and Matérn cluster point processes, with intensity $\lambda = 100$ on a unit square. Model parameters: hard-core radius (1) $h = 0.02$, (2) $h = 0.05$ and (3) $h = 0.08$; mean number μ of points per cluster and cluster radius r (1) $\mu = 1, r = 0.1$, (2) $\mu = 4, r = 0.1$, (3) $\mu = 4, r = 0.05$.

$$\hat{\varepsilon}_{ijk}(r) = \sqrt{n_{ijk}(n_{ijk} - 1)} \left[\hat{K}_{ijk}(r) - \bar{K}_{ij\cdot}(r) \right],$$
$$i = 1, \dots, a, j = 1, \dots, b, k = 1, \dots c. \tag{5.28}$$

Note that we can not construct an exact test since the assumption of exchange-ability is not strictly satisfied, that is, the residuals are weakly correlated and the correction factor $\sqrt{n_{ijk}(n_{ijk} - 1)}$ is motivated by the asymptotic approximation of the variance of Ripley's K-function for Poisson patterns where the number of points has been observed (Ripley, 1988). We can calculate the statistics by using the residuals and repeat this process a large number of times (by drawing at random with replacement) so that we can obtain bootstrap samples that can be used to obtain an empirical approximation of the $(1 - \alpha)$-percentiles of our test statistics.

For the interaction effect we also consider a more sophisticated version of the residuals. Indeed, we use the model (5.18) to give a weighted version of the residuals (as long as the processes are stationary)

$$\hat{\varepsilon}_{ijk}^{\dagger}(r) = \sqrt{n_{ijk}(n_{ijk} - 1)} \left[\hat{K}_{ijk}(r) - \hat{K}_{i\cdot\cdot}(r) - \hat{K}_{\cdot j\cdot}(r) + \hat{K}_{\cdot\cdot\cdot}(r) \right],$$
$$i = 1, \dots, a, j = 1, \dots, b, k = 1, \dots c. \tag{5.29}$$

These residuals are meant to attempt to control for main effects and asymptotically approach the exact test because, although SSA and SSB are not kept constant, variability due to *A* and *B* are estimated and removed by subtracting weighted means.

We go further and propose some residuals to study the main effects. These model-based residuals have been motivated by those referred already by Diggle et al. (1991), although they have never been used in this context, not even in the one-way case. The residuals are given by

$$\hat{\varepsilon}_{ijk}^{\dagger A}(r) = \sqrt{n_{ijk}(n_{ijk}-1)} \left[\hat{K}_{ijk}(r) - \bar{K}_{i\cdot\cdot}(r) \right],$$
$$i = 1,\ldots,a, j = 1,\ldots,b, k = 1,\ldots c, \tag{5.30}$$

for the row effect, and

$$\hat{\varepsilon}_{ijk}^{\dagger B}(r) = \sqrt{n_{ijk}(n_{ijk}-1)} \left[\hat{K}_{ijk}(r) - \bar{K}_{\cdot j\cdot}(r) \right],$$
$$i = 1,\ldots,a, j = 1,\ldots,b, k = 1,\ldots c, \tag{5.31}$$

for the column effect.

Level accuracy under the null hypothesis

The available methodologies for the analysis of variance in the context of point processes are given by using *D*-type statistics, given in Eq. (5.26), whose null distribution is obtained by permuting residuals proportional to $\hat{\varepsilon}_{ijk}(r)$ for all effects (Wilson, 1998) or permuting residuals proportional to $\hat{\varepsilon}_{ijk}^{\dagger}(r)$ for interaction effects (Ramón et al., 2016). Therefore, we first study the empirical distribution of *D*-type statistics by permuting residuals $\hat{\varepsilon}_{ijk}(r)$ and by using the simplest scenario (Poisson) with configurations given in Table 5.1. The results are shown in Table 5.2. Unfortunately these test statistics show a poor performance in the homoscedastic case. As expected, we are also not doing better with the heteroscedasticity From now on, we concentrate on the study of our proposed *F*-statistics with several suitable residuals according to each effect. Indeed, Table 5.3 shows the performance in the homoscedastic and heteroscedastic Poisson cases, and Table 5.4 shows the performance in the specific cluster and hard-core models. Tables 5.3 and 5.4 show much better performance even in heteroscedastic cases. It is important to highlight the good performance of the test statistic F^I that measures the interaction effect. This statistic is probably the most important in an ANOVA

Sig. level		Rejection Rates		
	D	0.01	0.05	0.10
SConfig.1	D^A	0.024	0.098	0.164
	D^B	0.037	0.109	0.189
	D^{AB}	0.050	0.140	0.209
	D^I	0.047	0.129	0.205
SConfig.2	D^A	0.037	0.102	0.166
	D^B	0.038	0.109	0.191
	D^{AB}	0.049	0.147	0.223
	D^I	0.040	0.141	0.237
SConfig.3	D^A	0.035	0.109	0.170
	D^B	0.034	0.097	0.171
	D^{AB}	0.057	0.138	0.192
	D^I	0.055	0.132	0.099
DConfig.1	D^A	0.056	0.164	0.272
	D^B	0.019	0.072	0.129
	D^{AB}	0.061	0.175	0.278
	D^I	0.041	0.116	0.218
DConfig.2	D^A	0.020	0.067	0.116
	D^B	0.060	0.169	0.276
	D^{AB}	0.062	0.177	0.261
	D^I	0.044	0.129	0.199
DConfig.3	D^A	0.025	0.076	0.129
	D^B	0.047	0.133	0.219
	D^{AB}	0.056	0.139	0.226
	D^I	0.051	0.153	0.237

Table 5.2 Rejection rates from replicated simulations of the true null hypothesis of no differences amongst the samples of K-functions of two factors with 3 levels and 6 replicates (point patterns) in each cell by using naive statistics given in Eq. (5.26). The scenarios indicated in left columns come from configurations shown in Table 5.1. Exchangeable units for a test were chosen according to the residuals given in Eq. (5.28).

two-way design because if the interaction is significant, the interpretation of the individual effects of the factors becomes incomplete and misleading. So digging a bit more into the interaction effect, the performance of the test statistic F^I by using the residuals (5.29) is shown in Table 5.5. Once again, the test statistic F^I shows an absolutely satisfactory performance (even in the case of heteroscedasticity). This shows that for the interaction both types of residuals are efficient and robust against a moderate violation of the equality of variances.

Sig. level		Rejection Rates			
	F	0.01	0.05	0.10	
SConfig.1	F^A	0.012	0.056	0.093	*
	F^B	0.013	0.056	0.098	*
	F^{AB}	0.012	0.054	0.102	*
	F^I	0.012	0.053	0.091	*
Sconfig.2	F^A	0.018*	0.052	0.096	**
	F^B	0.013	0.053	0.103	*
	F^{AB}	0.013	0.062	0.108	*
	F^I	0.013	0.057	0.111	*
Sconfig.3	F^A	0.011	0.046	0.100	*
	F^B	0.010	0.039	0.089	*
	F^{AB}	0.014	0.044	0.087	*
	F^I	0.013	0.051	0.092	*
DConfig.1	F^A	0.026	0.080	0.151	
	F^B	0.010	0.037	0.073*	**
	F^{AB}	0.030	0.078	0.133	
	F^I	0.009	0.063	0.116	*
DConfig.2	F^A	0.004	0.031*	0.066*	**
	F^B	0.025	0.088	0.165	
	F^{AB}	0.023	0.078	0.128	
	F^I	0.012	0.056	0.104	*
DConfig.3	F^A	0.010	0.027*	0.068*	**
	F^B	0.031	0.089	0.171	
	F^{AB}	0.024	0.087	0.148	
	F^I	0.011	0.051	0.100	*

Table 5.3 Rejection rates from replicated simulations of the true null hypothesis of no differences amongst the samples of K-functions by using Fisher-test type statistics including the degrees of freedom within the statistic. Exchangeable units for tests were chosen according to the residuals given in Eq. (5.28). Values that lie outside the 95% confidence interval for type I error, which has a binomial distribution with parameters $n = 1000, p = 0.01, 0.05$ and 0.1 respectively, are indicated with no symbols in the case of violation of the three nominal levels, or with ** in the case of violation of at most two nominal levels. Finally, * indicates three good nominal levels.

By using the same simulation scheme, we study the model-based residuals (Eq. 5.30 and 5.31) that in theory are the most adequate to test the main effects in ANOVA two-way designs (Anderson and Braak, 2003). The performance is shown in Table 5.6. Interestingly, model-based residuals (Eq. 5.30 and 5.31) do

cluster					
Sig. level		Rejection Rates			
	F	0.01	0.05	0.10	
$\mu = 1, r = 0.1$	F^A	0.008	0.044	0.093	*
$\kappa = 100$	F^B	0.006	0.041	0.080	*
	F^{AB}	0.008	0.047	0.086	*
	F^I	0.009	0.044	0.098	*
$\mu = 4, r = 0.1$	F^A	0.014	0.056	0.101	*
$\kappa = 100$	F^B	0.016	0.063	0.104	*
	F^{AB}	0.012	0.052	0.111	*
	F^I	0.008	0.061	0.102	*
$\mu = 4, r = 0.05$	F^A	0.017*	0.054	0.097	**
$\kappa = 100$	F^B	0.009	0.049	0.097	*
	F^{AB}	0.011	0.055	0.098	*
	F^I	0.012	0.056	0.095	*
hard-core					
$h = 0.05$	F^A	0.009	0.041	0.091	*
	F^B	0.010	0.052	0.101	*
	F^{AB}	0.012	0.044	0.087	*
	F^I	0.015	0.047	0.098	*
$h = 0.02$	F^A	0.010	0.045	0.101	*
	F^B	0.011	0.044	0.098	*
	F^{AB}	0.009	0.055	0.101	*
	F^I	0.013	0.047	0.092	*

Table 5.4 Performance of the test for **cluster** (left) and **hard-core** (right) models whose parameters are given in the left columns. Exchangeable units for a test were chosen according to the residuals given in Eq. (5.28). Values are indicated as in Table 5.3.

not outperform the full-model residuals when using Fisher-test type statistics to test main-effects.

Power of the tests

Analysing the power of the ANOVA (even in the classical case) is far from being a simple task. The problem is that there are infinite possibilities to get away from the null hypotheses, and each one of those infinite differences implies a different power. The problem comes from trying to compare several groups according to all the levels of the factors and replicates.

Poisson					
Sig. level		Rejection Rates			
	F	0.01	0.05	0.10	
SConfig.1	F^I	0.008	0.042	0.094	*
SConfig.2	F^I	0.012	0.048	0.097	*
Sconfig.3	F^I	0.012	0.054	0.095	*
DConfig.1	F^I	0.011	0.052	0.103	*
DConfig.2	F^I	0.016	0.049	0.088	*
DConfig.3	F^I	0.010	0.049	0.088	*
cluster					
$\mu = 1, r = 0.1, \kappa = 100$	F^I	0.008	0.042	0.098	*
$\mu = 4, r = 0.1, \kappa = 25$	F^I	0.011	0.047	0.091	*
$\mu = 4, r = 0.05, \kappa = 25$	F^I	0.013	0.047	0.094	*
hard-core					
$h = 0.05$	F^I	0.006	0.048	0.097	*
$h = 0.02$	F^I	0.015	0.055	0.105	*

Table 5.5 Rejection rates from replicated simulations of the true null hypothesis of no interaction between factors A and B by using residuals given in Eq. (5.29). Three null models are considered: homogeneous **Poisson** (left), **cluster** (right-up) and **hard-core** (right-down) as underlying processes. Values are indicated as in Table 5.3.

In the classical literature, there is a suitable function, called *degree of falsity* (Feldt and Mahmoud, 1958) that can help us in detecting uniquely the power (for some fixed parameters). It is the measure of effect size (Anderson and Braak, 2003). This function measures the average of the quadratic differences between the means of the null hypothesis and those of the alternative hypothesis. Symbolically the function is defined for main and interaction effects as

$$f_A = \frac{1}{\sigma_\varepsilon} \sqrt{\frac{\sum_{i=1}^{a}(A_i - \bar{A})^2}{a}}, \quad \text{and}$$

$$f_{AB} = \frac{1}{\sigma_\varepsilon} \sqrt{\frac{\sum_{j=1}^{b}\sum_{i=1}^{a}(AB_{ij} - \overline{AB})^2}{ab}}.$$

(5.32)

In classical ANOVA it is defined $\theta_A = \sigma_\varepsilon f_A$ and analogously for the other cases. The index θ is used to measure the power. In order to extrapolate the function θ to our point process context, we point out the nature of the K-function, which makes it sensitive to the aggregation parameters of the underlying process and only to them. This is why for example, all Poisson processes have the same

Poisson					
Sig. level		Rejection Rates			
	F	0.01	0.05	0.10	
SConfig.1	F^A	0.008	0.046	0.087	*
	F^B	0.012	0.051	0.096	*
	F^{AB}	0.011	0.054	0.094	*
Sconfig.2	F^A	0.008	0.051	0.108	*
	F^B	0.014	0.046	0.095	*
	F^{AB}	0.007	0.047	0.102	*
Sconfig.3	F^A	0.009	0.051	0.098	*
	F^B	0.013	0.044	0.092	*
	F^{AB}	0.014	0.052	0.096	*
DConfig.1	F^A	0.032	0.095	0.163	
	F^B	0.005	0.027*	0.065*	**
	F^{AB}	0.031	0.087	0.143	
DConfig.2	F^A	0.008	0.034*	0.069*	**
	F^B	0.027	0.097	0.163	
	F^{AB}	0.029	0.091	0.151	
DConfig.3	F^A	0.008	0.043	0.078*	**
	F^B	0.018*	0.064	0.127*	**
	F^{AB}	0.013	0.061	0.115	*
cluster					
$\mu = 1, r = 0.1,$	F^A	0.013	0.058	0.104	*
$\kappa = 100$	F^B	0.010	0.047	0.082	*
	F^{AB}	0.014	0.054	0.102	*
$\mu = 4, r = 0.1$	F^A	0.008	0.040	0.075*	**
$\kappa = 100$	F^B	0.009	0.044	0.082	*
	F^{AB}	0.008	0.037	0.077	**
$\mu = 4, r = 0.05,$	F^A	0.013	0.048	0.089	*
$\kappa = 100$	F^B	0.009	0.044	0.088	*
	F^{AB}	0.007	0.057	0.094	*
hard-core					
$h = 0.05$	F^A	0.004	0.043	0.091	*
	F^B	0.009	0.049	0.096	*
	F^{AB}	0.010	0.041	0.095	*
$h = 0.02$	F^A	0.014	0.057	0.100	*
	F^B	0.017*	0.059	0.099	**
	F^{AB}	0.016	0.060	0.110	*

Table 5.6 Rejection rates from replicated simulations of the true null hypothesis of no main effects in the samples of K-functions. Three null models are considered: homogeneous **Poisson** (left), **cluster** (right-up) and **hard-core** (right-down) as underlying processes. Exchangeable units are the model-based residuals (Eq. 5.30 and 5.31). Values are indicated as in Table 5.3.

theoretical K-function (πr^2 in the planar case). So, we can move away from the null hypothesis by systematically increasing or decreasing these parameters.

In a Matérn-cluster process (as a particular case of the Neyman-Scott) the K-function depends on the intensity of the parents and the degree of aggregation of the offspring around the parents Diggle (2013). Therefore, we could easily establish structural differences between Matérn processes through the variation of these two parameters. On the other hand, in a Matérn-hard-core process, the K-function depends on both the intensity of the proposal-points and the inhibition distance Cressie (1993). So again we could fix structural differences through the modification of these sensitive parameters.

Our most general alternative hypothesis is that the levels of the factors as well as the interaction between them affect in a significant way these parameters. Therefore they also affect the K-function of the resulting patterns in equal measure. So we define such degree of falsity through the standard deviation of the parameters matrix in each scenario. Thus, for an array of Matérn cluster patterns with $\kappa = 25, \mu = 4$, and radius of the clusters in the patterns of the cells given by

$$\mathbf{r}_{AB} = \begin{pmatrix} 0.1 & 0.1 & 0.1 \\ 0.1 & 0.3 & 0.1 \\ 0.1 & 0.1 & 0.5 \end{pmatrix}, \mathbf{r}_A = \begin{pmatrix} 0.1 & 0.3 & 0.5 \\ 0.1 & 0.3 & 0.5 \\ 0.1 & 0.3 & 0.5 \end{pmatrix}, \text{ and } \mathbf{r}_B = \begin{pmatrix} 0.1 & 0.1 & 0.1 \\ 0.3 & 0.3 & 0.3 \\ 0.5 & 0.5 & 0.5 \end{pmatrix},$$

the degree of falsity with respect to $\mathbf{r}$ is given by $\theta_{AB} = \mathrm{sd}(\mathbf{r}_{AB}) = 0.1414, \theta_A = \theta_B = 0.2$. Note that any of the alternative hypotheses for rows or columns effects apply directly over the effects of addition of rows and columns, because in this case the effects of rows and columns are not simultaneously zero.

To study the power of the statistic for the interaction effects we fix a set of configurations as given in Table 5.7. The results of the simulations for the power of the interaction and main effects statistics are shown in Table 5.8.

Clearly, strongly aggregated patterns in the cells (patterns with larger K-functions) are discriminated from weakly aggregated patterns with an approximate probability of 1, and the same happens with strongly regular patterns. As might be expected, as the parameters of the models approach each other ($\theta \to 0$), and therefore generate virtually indistinguishable patterns, the power of the tests decreases.

Cluster Configurations		
$\theta_{AB} = 0.0014$	$\mathbf{r}_{AB} = \begin{pmatrix} 0 & & \\ & 0.002 & \\ & & 0.004 \end{pmatrix} + \phi_{\mathrm{c}}\mathbf{J}_3$	
$\theta_{AB} = 0.0707$	$\mathbf{r}_{AB} = \begin{pmatrix} 0 & & \\ & 0.1 & \\ & & 0.2 \end{pmatrix} + \phi_{\mathrm{c}}\mathbf{J}_3$	
$\theta_{AB} = 0.3536$	$\mathbf{r}_{AB} = \begin{pmatrix} 0 & & \\ & 0.5 & \\ & & 1.0 \end{pmatrix} + \phi_{\mathrm{c}}\mathbf{J}_3$	
$\theta_A = 0.0017$	$\mathbf{r}_A = \begin{pmatrix} 0 & 0.002 & 0.004 \\ 0 & 0.002 & 0.004 \\ 0 & 0.002 & 0.004 \end{pmatrix} + \phi_{\mathrm{c}}\mathbf{J}_3$	
$\theta_A = 0.0866$	$\mathbf{r}_A = \begin{pmatrix} 0 & 0.1 & 0.2 \\ 0 & 0.1 & 0.2 \\ 0 & 0.1 & 0.2 \end{pmatrix} + \phi_{\mathrm{c}}\mathbf{J}_3$	
$\theta_A = 0.4330$	$\mathbf{r}_A = \begin{pmatrix} 0 & 0.5 & 1.0 \\ 0 & 0.5 & 1.0 \\ 0 & 0.5 & 1.0 \end{pmatrix} + \phi_{\mathrm{c}}\mathbf{J}_3$	
Hard-Core Configurations		
$\theta_{AB} = 0.0014$	$\mathbf{h}_{AB} = \begin{pmatrix} 0 & & \\ & 0.002 & \\ & & 0.004 \end{pmatrix} + \phi_{\mathrm{i}}\mathbf{J}_3$	
$\theta_{AB} = 0.0567$	$\mathbf{h}_{AB} = \begin{pmatrix} 0 & & \\ & 0.08 & \\ & & 0.16 \end{pmatrix} + \phi_{\mathrm{i}}\mathbf{J}_3$	
$\theta_{AB} = 0.07071$	$\mathbf{h}_{AB} = \begin{pmatrix} 0 & & \\ & 0.1 & \\ & & 0.2 \end{pmatrix} + \phi_{\mathrm{i}}\mathbf{J}_3$	
$\theta_A = 0.0017$	$\mathbf{h}_A = \begin{pmatrix} 0 & 0.002 & 0.004 \\ 0 & 0.002 & 0.004 \\ 0 & 0.002 & 0.004 \end{pmatrix} + \phi_{\mathrm{c}}\mathbf{J}_3$	
$\theta_A = 0.0693$	$\mathbf{h}_A = \begin{pmatrix} 0 & 0.08 & 0.16 \\ 0 & 0.08 & 0.16 \\ 0 & 0.08 & 0.16 \end{pmatrix} + \phi_{\mathrm{c}}\mathbf{J}_3$	
$\theta_A = 0.0866$	$\mathbf{h}_A = \begin{pmatrix} 0 & 0.1 & 0.2 \\ 0 & 0.1 & 0.2 \\ 0 & 0.1 & 0.2 \end{pmatrix} + \phi_{\mathrm{c}}\mathbf{J}_3$	

Table 5.7 Configurations of systematic departures from null hypotheses of no interaction between factors in the ANOVA two-way design. The parameters of the cluster processes are $\mu \in (1,4,4); \kappa \in (100,25,25)$. $\mathbf{J}_3$ denotes the 3×3 matrix of ones and $\phi_{\mathrm{c}} \in (0.1, 0.05)$ and $\phi_{\mathrm{i}} \in (0.02, 0.05, 0.08)$.

5.4 Data analysis

5.4.1 Poisson log-linear model for the expected cell counts

One of the first questions that we are interested in, is whether there are significant differences in the intensities of the bubble patterns as the levels of the volumetric

Cluster Configurations

Sig. level			Rejection Rates		
Degree of falsity	Residual	(μ,κ,ϕ_c)	0.01	0.05	0.10
$\theta_{AB}=0.0014$	$\hat{\varepsilon}_{ijk}(r)$	(1,100,0.1)	0.011	0.054	0.108
		(4,25,0.1)	0.014	0.043	0.091
		(4,25,0.05)	0.007	0.044	0.094
	$\hat{\varepsilon}^{\dagger}_{ijk}(r)$	(1,100,0.1)	0.014	0.050	0.092
		(4,25,0.1)	0.011	0.044	0.095
		(4,25,0.05)	0.009	0.034	0.082
$\theta_{AB}=0.0707$	$\hat{\varepsilon}_{ijk}(r)$	(1,100,0.1)	0.056	0.217	0.365
		(4,25,0.1)	0.361	0.710	0.856
		(4,25,0.05)	0.624	0.903	0.966
	$\hat{\varepsilon}^{\dagger}_{ijk}(r)$	(1,100,0.1)	0.058	0.203	0.364
		(4,25,0.1)	0.326	0.700	0.856
		(4,25,0.05)	0.633	0.911	0.964
$\theta_{AB}=0.3536$	$\hat{\varepsilon}_{ijk}(r)$	(1,100,0.1)	0.140	0.409	0.587
		(4,25,0.1)	0.779	0.952	0.988
		(4,25,0.05)	0.941	0.989	0.998
	$\hat{\varepsilon}^{\dagger}_{ijk}(r)$	(1,100,0.1)	0.099	0.393	0.588
		(4,25,0.1)	0.754	0.961	0.987
		(4,25,0.05)	0.924	0.988	0.995
$\theta_A=0.0017$	$\hat{\varepsilon}_{ijk}(r)$	(1,100,0.1)	0.017	0.052	0.089
		(4,25,0.1)	0.014	0.047	0.098
		(4,25,0.05)	0.007	0.047	0.101
	$\hat{\varepsilon}^{\dagger A}_{ijk}(r)$	(1,100,0.1)	0.009	0.056	0.112
		(4,25,0.1)	0.009	0.063	0.110
		(4,25,0.05)	0.011	0.050	0.090
$\theta_A=0.0866$	$\hat{\varepsilon}_{ijk}(r)$	(1,100,0.1)	0.866	0.957	0.981
		(4,25,0.1)	1.000	1.000	1.000
		(4,25,0.05)	1.000	1.000	1.000
	$\hat{\varepsilon}^{\dagger A}_{ijk}(r)$	(1,100,0.1)	0.952	0.989	0.997
		(4,25,0.1)	1.000	1.000	1.000
		(4,25,0.05)	1.000	1.000	1.000
$\theta_A=0.4330$	$\hat{\varepsilon}_{ijk}(r)$	(1,100,0.1)	0.998	1.000	1.000
		(4,25,0.1)	1.000	1.000	1.000
		(4,25,0.05)	1.000	1.000	1.000
	$\hat{\varepsilon}^{\dagger A}_{ijk}(r)$	(1,100,0.1)	1.000	1.000	1.000
		(4,25,0.1)	1.000	1.000	1.000
		(4,25,0.05)	1.000	1.000	1.000

Hard-Core Configurations

Sig. level			Rejection Rates		
Degree of falsity	Residual	ϕ_1	0.01	0.05	0.10
$\theta_{AB}=0.0014$	$\hat{\varepsilon}_{ijk}(r)$	0.02	0.015	0.055	0.113
		0.05	0.019	0.065	0.121
		0.08	0.014	0.063	0.117
	$\hat{\varepsilon}^{\dagger}_{ijk}(r)$	0.02	0.015	0.055	0.099
		0.05	0.012	0.067	0.123
		0.08	0.013	0.069	0.123
$\theta_{AB}=0.0567$	$\hat{\varepsilon}_{ijk}(r)$	0.01	1.000	1.000	1.000
		0.05	0.986	0.986	0.986
		0.08	0.994	0.994	0.994
	$\hat{\varepsilon}^{\dagger}_{ijk}(r)$	0.01	0.999	0.999	0.999
		0.05	0.990	0.993	0.999
		0.08	0.994	0.994	0.996
$\theta_{AB}=0.07071$	$\hat{\varepsilon}_{ijk}(r)$	0.01	1.000	1.000	1.000
		0.05	1.000	1.000	0.988
		0.08	0.994	0.994	0.994
	$\hat{\varepsilon}^{\dagger}_{ijk}(r)$	0.01	0.999	0.999	0.999
		0.05	0.990	0.990	0.990
		0.08	0.931	0.994	0.994
$\theta_A=0.0017$	$\hat{\varepsilon}_{ijk}(r)$	0.02	0.001	0.009	0.150
		0.05	0.104	0.294	0.424
		0.08	0.111	0.303	0.423
	$\hat{\varepsilon}^{\dagger A}_{ijk}(r)$	0.02	0.020	0.073	0.139
		0.05	0.109	0.296	0.427
		0.08	0.111	0.270	0.398
$\theta_A=0.0693$	$\hat{\varepsilon}_{ijk}(r)$	0.02	1.000	1.000	1.000
		0.05	1.000	1.000	1.000
		0.08	0.999	0.999	0.999
	$\hat{\varepsilon}^{\dagger A}_{ijk}(r)$	0.02	1.000	1.000	1.000
		0.05	1.000	1.000	1.000
		0.08	0.987	0.987	0.987
$\theta_A=0.0866$	$\hat{\varepsilon}_{ijk}(r)$	0.02	1.000	1.000	1.000
		0.05	1.000	1.000	1.000
		0.08	1.000	1.000	1.000
	$\hat{\varepsilon}^{\dagger A}_{ijk}(r)$	0.02	1.000	1.000	1.000
		0.05	1.000	1.000	1.000
		0.08	1.000	1.000	1.000

Table 5.8 Empirical power for interaction and main effect statistics when the underlying point processes are cluster and hard-core with parameters given in Table 5.7.

air flow and the specific frother concentration vary. Since the spatial windows are fixed and constant throughout the experiment, it is sufficient to compare the count differences between the groups. Figure 5.5 shows violin-plots of the number of points in each pattern, broken down by the three levels of frother concentration levels, and faceted by the three levels of volumetric air flow levels. The violin plot allows us to suspect that the number of points, as well as their probability distributions, vary between the cells along the levels of the two factors. We use a quasi-Poisson regression model in order to deal with possible
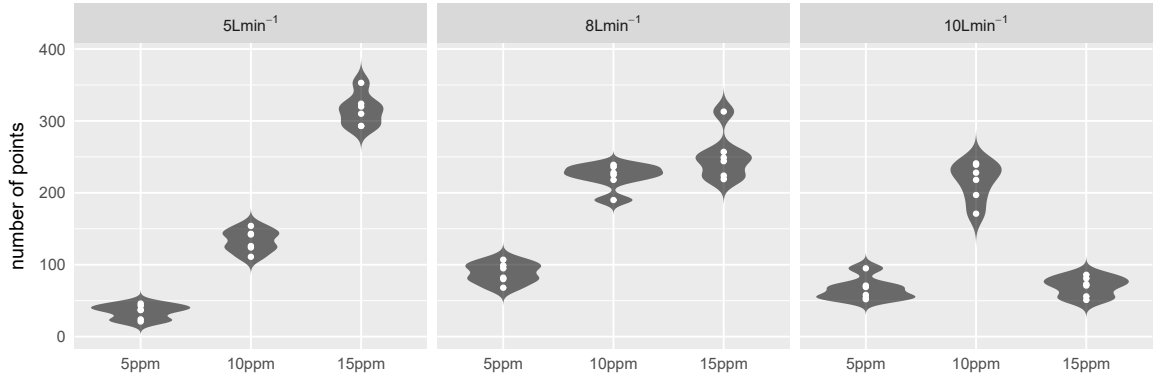
Figure 5.5 Violin-plots of numbers of bubbles per level of the two different factors. Each panel corresponds to a level of volumetric air flow and the horizontal axis corresponds to frother concentration.

overdispersion Diggle et al. (1991); McCullagh and Nelder (1989) with a quasi-likelihood approach. Let us denote as a random variable the count n_{ijk}, and we can assume that $\mathbb{E}(n_{ijk}) = \mu_{ij}$ and $\mathbb{V}\mathrm{ar}(n_{ijk}) = \phi\mu_{ij}$, where $\phi > 0$ is an overdispersion parameter. We fit the following model

$$\mu_{ij} = \exp\left\{\beta_0 + \beta_1 i + \beta_2 j + \beta_3 ij\right\}$$

where β_1 and β_2 are the parameters for the main effects, and β_3 is the parameter for the interaction. The estimated overdispersion parameter is $\hat{\phi} = 2.737$ which is quite high. Such overdispersion can be explained by non-measured covariates, random effects or by non-Poisson variation within bubble patterns (clustering or inhibition into the point patterns for instance, see Diggle et al., 1991). We test the overall effect of the two factors by comparing the deviance of the full model with the deviance of the model excluding each one (Table 5.9). The Chi-square test indicates that factors, taken together, are statistically significant predictors of the number of points despite the overdispersion.

Model	Deviance	Res. df	df	Res. Deviance	$\mathbb{P}(> X_{\mathrm{df}}^2)$
Null		53		3255.9	
Frother concentration	1738	51	2	1517.6	< 0.01
Volumetric air flow	299	49	2	1218.6	< 0.01
Interaction	1095	45	4	123.40	< 0.01

Table 5.9 Analysis of deviance table for the quasi-likelihood model.

Overall, the counts differ between frother concentration levels and volumetric air flow and their interaction is significant, i.e, there are combined effects of these two factors on the observed number of bubbles. This interaction is extremely important since when an interaction effect is present, the impact of one factor depends on the levels of the another factor.

5.4.2   Spatial distribution of bubble patterns

Once we have checked the influence of the factors on the densities of the patterns, the next question to answer is whether the factors or their interaction affect the second-order structure of the patterns. Indeed, the 54 point patterns shown in Figure 5.3 represent the positions of the bubbles in the flotation experiment described in Section 5.1. The estimated K-functions are shown in Figure 5.6. To measure the variability of the pooled estimates $\bar{K}_{ij\cdot}$, we used the delta-method approximation to the variance of a ratio described in Baddeley et al. (2015). We thus calculated a standard error and made approximated 95% confidence intervals (grey shadings in Figure 5.6) for the K-function. K-functions of the bubble experiment array indicate small-scale regularity with an inhibitory effect up to 0.5mm on average in most of the cells, except in cells $(5\mathrm{L\,min}^{-1}, 8\mathrm{ppm})$ and $(5\mathrm{L\,min}^{-1}, 10\mathrm{ppm})$, where the inhibitory effect is almost 1.0mm. For larger distances there is a tendency to clustering in most patterns, except perhaps in those of the first cell $(5\mathrm{L\,min}^{-1}, 5\mathrm{ppm})$ where it is seen that the K-functions oscillate around CSR line (horizontal zero). Furthermore, one would suspect that the 5ppm case behaves differently than the others. The first row indicates a (close to) Poisson pattern and the two other rows indicate a larger radius for the inhibition.

Tests based on Fisher-type statistics were carried out using the integration interval $T = (0, r_0]$, whose upper limit is recommended to be taken as (Hahn, 2012)

$$r_0 = \frac{1.25}{\sqrt{\max \hat{\lambda}_{ij}}},$$
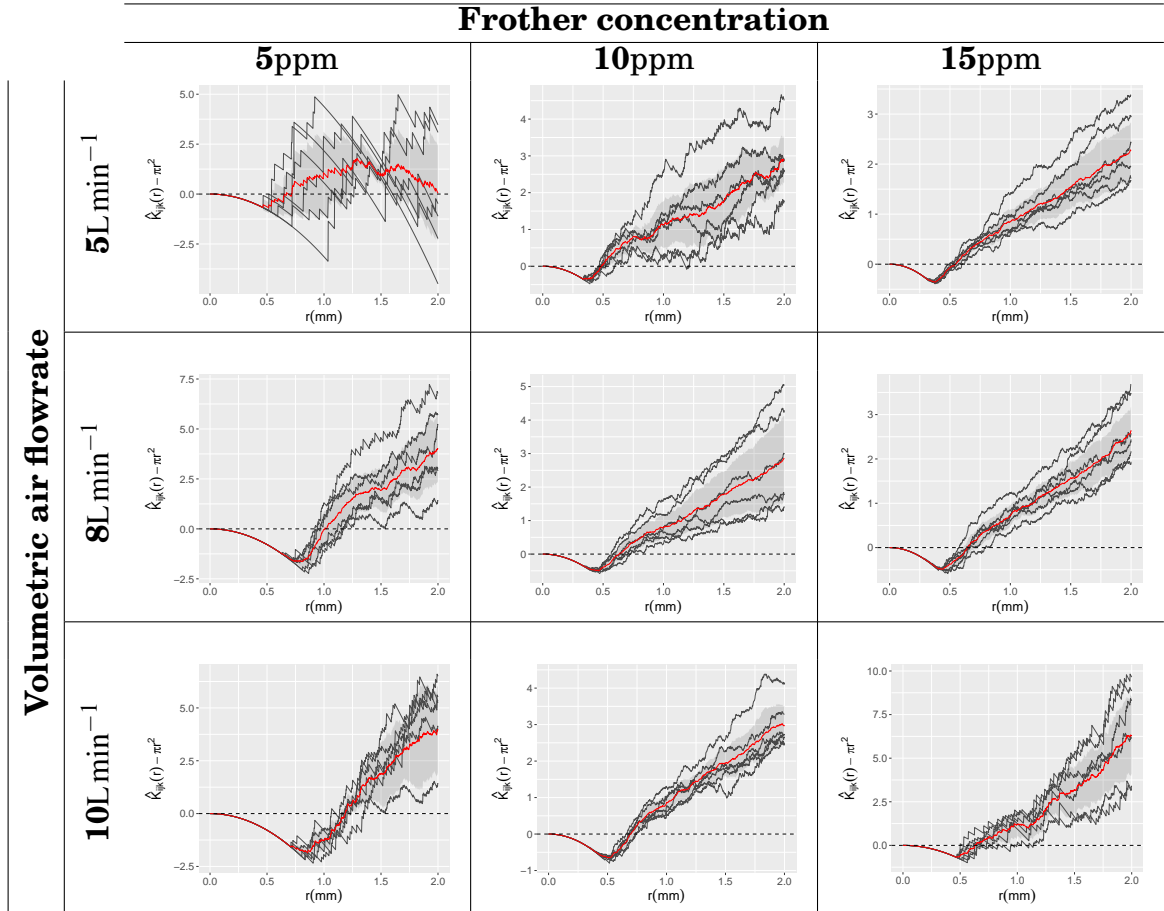
Figure 5.6 General arrangement of estimates of centred *K*-functions (black lines) and pooled *K*-functions (red lines) of point patterns for floating bubbles with three concentration levels and three gas flow levels. The grey shading in each panel corresponds to pointwise 95% confidence interval based on the observed within cell sample variance in the estimated *K*-function.

obtaining an upper integration bound $r_0 = 1.57$mm. We implemented the tests with 500000 random permutations. For the interaction effect, we have a significant *p*-value associated with the residuals given in Eq. (5.28), $\hat{p}_{\hat{\varepsilon}_{ijk}} = 0.0275$, analogously $\hat{p}_{\hat{\varepsilon}^{\dagger}_{ijk}} = 0.0153$ in the case of the residuals given in Eq. (5.29). For the main effects, we have *p*-values of zero in the case of the frother concentration factor for both residuals as well as *p*-values of zero in the additive effect. Finally, for the volumetric air flowrate factor, we have also significant *p*-values ($\hat{p}_{\hat{\varepsilon}_{ijk}} = 0.0110$ and $\hat{p}_{\hat{\varepsilon}^{\dagger A}_{ijk}} = 0.0059$).

Apparently the point patterns do not behave the same (in the sense of their structure) up to small distances, and this difference is explained by the levels of

the row and column factors. We conclude that the concentration and the volumetric air flow are significant terms in the model for the bubble point patterns.

We have seen that if the tests are made at the 0.05 level of significance, the critical p-values for the tests of the hypothesis that the action of the frother concentration is independent of the volumetric air flowrate category (ie, zero interaction) are significant. Hence bubble patterns do not support hypothesis of zero interaction. The K-functions indicate that the effects of the frother concentration differ for the three levels of the volumetric airflow rate. Contrary to classical analysis of variance, in this case it is quite difficult to indicate the nature of the interaction in a geometric fashion since the means of the cells are pooled K-functions, see Figure 5.7. For the geometric interpretation we define
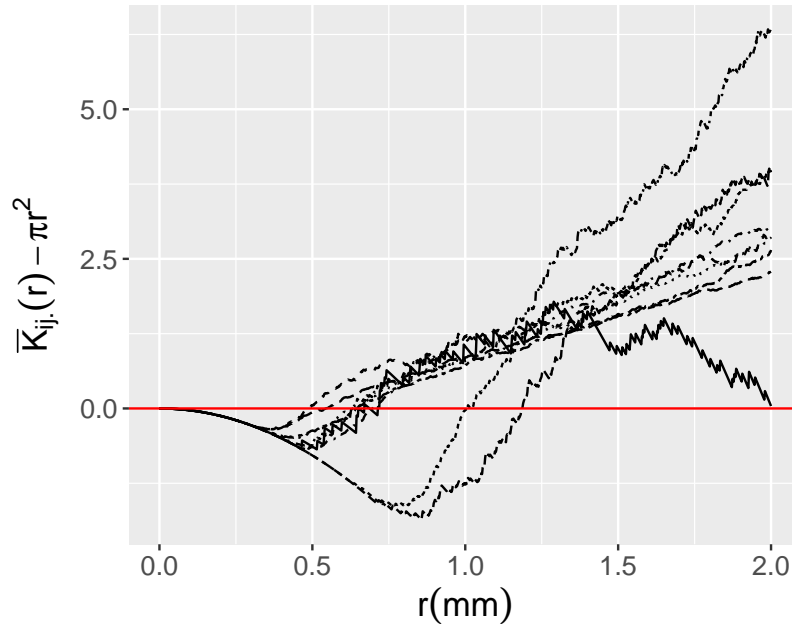
Figure 5.7 Pooled mean values of $\hat{K}_{ijk}(r)$ estimated on 9 cells of the flotation experiment. The red line represents the complete spatial randomness.

the following quantities

$$\Xi_{ij} = ||\bar{K}_{ij\cdot}(r) - \pi r^2||,$$

so we have a single number for each weighted cell mean, which may represent the overall behaviour of the pooled K-function. These quantities represent the quadratic differences between the estimates of the K-function and the theoretical K-function of a completely random pattern. This means that larger differences imply stronger interaction within the observed pattern. Values close to zero

indicate weak attraction or inhibition and, vice versa, high values indicate high degree of clustering or inhibition. We could relate this integral differences to the cell means of the classical experimental design. So a geometric representation of these Ξ_{ij} is given in Figure 5.8. This figure (in the left panel) represents the
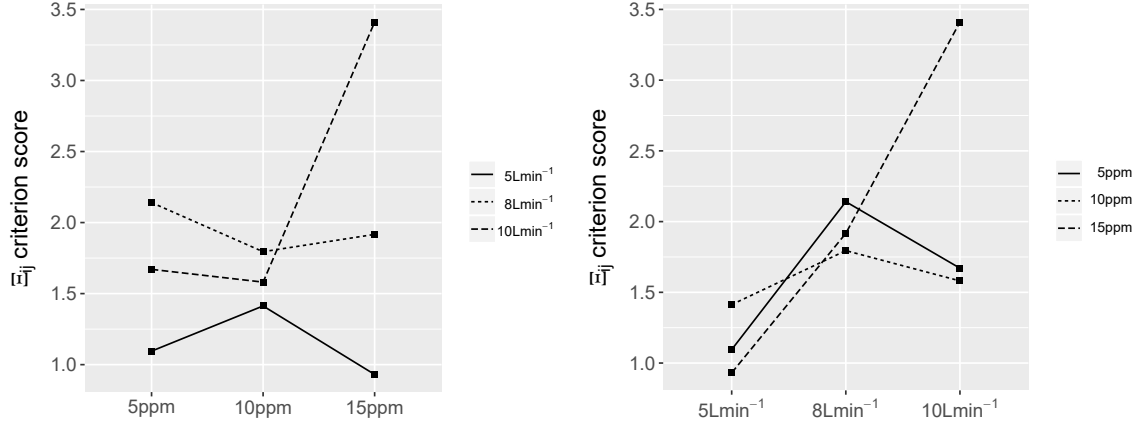
Figure 5.8 Profiles for simple effects for frother concentration (left) and for volumetric air flowrates (right), based on the measure Ξ_{ij}.

profiles corresponding to the simple effects of the frother concentration (column factor) for each of the volumetric air flowrates (row factor). An approximate graphical test for the presence of interaction might be equivalent to a test on the difference in the shapes of the profiles of these simple effects. A geometric representation (equivalent) of the profiles corresponding to the simple effects of the volumetric air flowrates for each of the frother concentrations is also displayed (right panel).

5.5   Discussion

We have implemented a collection of methods for the analysis of factorial experiments in the context of point processes where observations are functional second-order descriptors. Classical inference for spatial point processes is a quite delicate matter (see e.g Baddeley et al., 2017; Diggle et al., 1991; Hahn, 2012). We have implemented a new approach for the analysis of experiments with two fixed factors in the presence of replicates through approximate permutation tests. These tests have been demonstrated by simulation (the validity and power). Our tests require exchangeable units (functions), and this assumption can be guaranteed through the independence of point patterns in the sample. We have focused

on the case of balanced experiments with two factors and the same number of replicates within cells.

Our approach can be enhanced in many ways depending on the questions related to the experiment itself. Obviously a possible generalisation consists of considering unbalanced experiments, or more ambitiously, non-orthogonal designs. Our test statistics are defined through integral distances between the estimates of *K*-functions, their means and variances, implying that the power of the tests depends on the interval of integration $T = (0, r_0]$ (Choi and Hall, 1999; Hahn, 2012). So, further functional distances that are not as sensitive to the r_0 parameter could be used instead. Although we have focused mainly on Ripley's *K*-function, these methods could be used with other descriptors such as the *L*-function Besag (1977), the pair-correlation function $g(r)$, the nearest-neighbour distance distribution $F(r)$, the empty space function $G(r)$, or the *J*-function (Illian et al., 2008; van Lieshout and Baddeley, 1996).

Regarding the flotation experiment and the data of the bubbles, we applied a scaled Poisson model to the bubble counts in the flotation cells and we observed significant differences between the levels of the two factors as well as an extra-Poisson variation within the groups that suggests that a spatial structure may be achieved by the levels of the factors. We have discovered through the *K*-function that the patterns exhibit a small-scale regularity to move on to aggregation thereafter. This aspect represents an advance in the assumptions usually made in the literature of flotation experiments, since CSR patterns have been assumed (see e.g Emery et al., 2012; Kracht et al., 2013). We have shown how an analysis of variance for the effects of the frother concentration for each level of the volumetric air flowrate can explain the structural variations of the patterns. This variation is the product of the factors separately and, most importantly, of their interaction.

One of the important covariates that are associated with bubbles is their diameter. The diameter can be associated as a mark attached to the location of the centre of the bubble. Consider a space M in $\mathbb{R}$, called the *space of marks*, let $m(\mathbf{u}) \in M$ (hereinafter *mark*) be a random variable describing further information of $\mathbf{u} \in \mathbf{X}$. The set $\Phi = \{(\mathbf{u}, m(\mathbf{u})) | \mathbf{u} \in \mathbf{X}\}$, defined in $W \times M$, is a *planar marked point process* (Illian et al., 2008; Møller and Waagepetersen, 2004; Penttinen et al., 1992). It is natural to represent a marked point process as a collection of pairs $\{(\mathbf{u}_i, m_i)\}_{i=1}^{p} \subset W \times M$.

Let $t(\eta, \eta')$ be a non-negative function (called *test function*, see e.g. Stoyan and Stoyan, 1994) depending on the marks η, η'. Similarly to the unmarked case,

a second-order characteristic can be defined by the Ripley's *K*-function in the marked scenario. In this context we have the *Mark-weighted K-function* (see e.g Baddeley et al., 2015; Illian et al., 2008; Penttinen et al., 1992), defined as

$$\mathcal{K}_t(r) = \frac{1}{\lambda \mathbb{E}[t(\mathcal{N}, \mathcal{N}')]} \mathbb{E}\left[\sum_{\mathbf{u}_i \in X} t(\eta(\mathbf{u}), \eta(\mathbf{u}_i)) \mathbf{1}\{||\mathbf{u} - \mathbf{u}_i|| \leq r\} \, \middle| \, \mathbf{u} \in \mathbf{X} \right], \qquad (5.33)$$

where $\eta(\mathbf{u})$ and $\eta(\mathbf{u}_i)$ denote the mark values at the points $\mathbf{u}$ and $\mathbf{u}_i$, respectively. $\mathcal{N}$ and $\mathcal{N}'$ are independent random marks having the same distribution as the marks in the point process. Here the contribution from each pair of points is weighted by $t(\eta(\mathbf{u}), \eta(\mathbf{u}_i))$. A remarkable property is that under random labelling, $\mathcal{K}_t(r) = \mathcal{K}(r)$. Under stationarity, an unbiased estimator of $\mathcal{K}_t(r)$ is

$$\hat{K}_t(r) = \frac{1}{\hat{\lambda}^2 \hat{c}_t^*} \sum_{i=1}^{n} \sum_{j \neq i} t(\eta(\mathbf{u}_i), \eta(\mathbf{u}_j)) \mathbf{1}\{||\mathbf{u}_i - \mathbf{u}_j|| \leq r\} e(\mathbf{u}_i, \mathbf{u}_j; r), \qquad (5.34)$$

where

$$\hat{c}_t^* = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} t(\eta(\mathbf{u}_i), \eta(\mathbf{u}_j)),$$

and where $\hat{\lambda}^2$ is given by Eq. (5.3).

Before proposing an analysis through this marked methodology, it would be necessary to verify if the marks depend on the locations since, if not, the analysis proposed in this paper would be sufficient. To test this dependency Schlather et al. (2004) introduced two functional descriptors for homogeneous marked point processes describing the conditional expectation $E(r)$ and conditional variance $V(r)$ of a mark, given the existence of a further point of the process within a distance r. They presented tests based on such descriptors related with the null hypothesis that an independently marked point process has marks that are independent, identically distributed and independent of the locations. The descriptors are given, respectively, by

$$E(r) = \mathbb{E}\left[m(\mathbf{u}) \middle| \mathbf{u}, \mathbf{v} \in \mathbf{X}\right] \qquad (5.35)$$

$$V(r) = \mathbb{E}\left[(m(\mathbf{u}) - E(r))^2 \middle| \mathbf{u}, \mathbf{v} \in \mathbf{X}\right] \qquad (5.36)$$

where $\mathbf{u}, \mathbf{v}$ are arbitrary locations with $||\mathbf{u} - \mathbf{v}|| = r$. Under the null hypothesis $E(r)$ and $V(r)$ are constant. Thus, it can be tested by Monte Carlo tests based on measures of non-constancy of $\hat{V}(r) + \hat{E}^2(r)$ (a combined test), where the estimates

can be obtained through non-parametric smoothing methods as $E(r)$ is a especial case of the *mark-correlation function* with $t(m, m') = m$ and similarly for $V(r)$ (see e.g. Baddeley et al., 2015).

As an additional exercise, we perform a Monte Carlo test by using the test statistic

$$S^\dagger = \int_0^{r_0} \left[\hat{V}(r) + \hat{E}^2(r) - C \right]^2 dr, \qquad (5.37)$$

where the constant estimator C is calculated as the sum of the sample quadratic mean of the marks and their sample variance. The results are shown in Figure 5.9. It can be seen that the hypothesis that $E(r)$ and $V(r)$ are constant is not
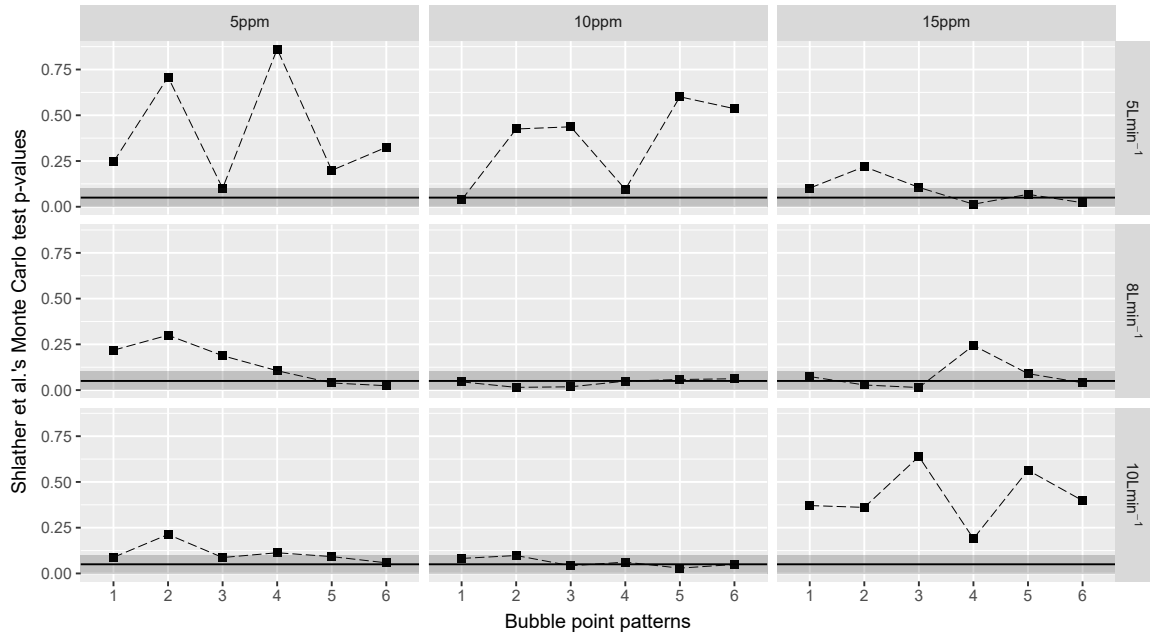
Figure 5.9 Array of p-values of the Schlather et al.'s test for each of the bubble point patterns in each cell of the experiment. Grey horizontal bands are critical regions ($\alpha \leq 0.1$) and the continuous horizontal lines corresponds to $\alpha = 0.05$.

always rejected. Apparently the dependence between marks and locations is also strongly influenced by the levels of the experiment factors. This inspires additional analysis approaches, for example by considering mark K-functions instead of classical K-functions to build Fisher-test type statistics for the factorial analysis. It is also possible to model the strength of the dependence as a functional response variable and implement Fisher-test type statistics in order to corroborate the effect of the interaction and the factors separately in the dependence structure of the array of the point patterns.

Acknowledgements

Chapter 6

Conclusions

In this thesis we have studied spatial and spatio-temporal point processes and how to analyse such point pattern data. Special attention has been paid to data with replicates. We have covered the analysis of spatio-temporal replicated point patterns and factorial procedures for spatial point patterns trying to answer statistical questions arisen from applications.

In Chapter 2, we have given a framework for spatial point processes and for the analysis of point patterns. Some basic definitions and concepts of point processes are recalled, both in stationary and inhomogeneous (second-order intensity-reweighted stationary) cases, and in the case of multivariate and general marked point processes. Definitions of first- and second-order characteristics are recalled and their non-parametric (edge-corrected) estimation discussed. In addition, some models for point patterns are given. Attention is carefully paid to define the different concepts for stationary and second-order intensity reweighted stationary (SOIRS) point processes, as they are both of interest along the thesis. Comments about replicated point patterns are made as they are essential in this thesis. This chapter is quite brief and intend to provide basic notions to make the thesis a more self-contained work. However, the extensive bibliography available in the field of spatial point processes, makes it very easy to go further in any of the exposed concepts.

Chapter 3 is a review paper, co-authored with Francisco J. Rodríguez-Cortés, Ottmar Cronie and Jorge Mateu. It has been published in *Spatial Statistics* in 2016. Spatio-temporal point processes are considered as coming from three categories and we have focused on the case where data can be thought as snapshots in space-time. We have presented the three data sets used for illustration through the paper and we have given a general definition for spatio-temporal

point processes as well as discussed on spatio-temporal separability. Some characteristics, such as intensity function and higher-order characteristics, and their edge-corrected estimation, are recalled. Depending on the summary statistic, stationarity, isotropy, second-order intensity-reweighted stationarity or intensity-reweighted moment stationarity has been assumed. Some attention has been paid also to anisotropic spatio-temporal point processes. Finally, some spatio-temporal empirical and mechanistic models, estimation of parameters, and testing goodness-of-fit of the models are discussed and some ongoing research is highlighted and discussed.

Chapters 4 and 5 are the principal focus of the thesis. The paper in Chapter 4 is co-authored with Ute Hahn and Jorge Mateu. We have considered tornado reports as a realisation of a SOIRS spatio-temporal point process. We have not assumed separability of the spatio-temporal first-order intensity function. We have defined a new non-parametric estimator of the spatio-temporal intensity function. In order to test if there are differences between the underlying processes of the warm- and cold-seasons, we have partitioned the study region (Central and Eastern U.S.) into spatial tiles with almost the same number of points per tile. Finally we have introduced new tests to compare the second-order structure (described by the spatio-temporal K-function) of such point patterns. A simulation study is carried out to study the performance of the new tests. A discussion about the assumptions and methods used is proposed.

Chapter 5 is composed by a paper co-authored with Bernardo M. Lagos-Álvarez and Jorge Mateu. We have extended the balanced two-way factorial analysis to the case of replicated spatial point patterns. The paper is motivated by a flotation experiment in minerals engineering affected by two experimental factors. We have assumed stationarity given the physical conditions of the experiment. The second-order structure of the point patterns is described by Ripley's K-function and it is investigated whether the structure is affected by the experimental factors. Different test statistics are derived and their distribution under the null hypothesis found by using bootstrap. A simulation study is performed in order to study the performance of the new tests. The methodology has been applied to the spatial distribution of bubbles in the flotation experiment. Some possible generalisations are discussed, in particular the fact of considering important properties of the bubbles along the experiment as marks attached to the locations and including them in the second-order descriptors.

Finally, we would like to emphasize that the fact of having replicates has not been widely exploited throughout the literature of point processes. However, in recent years, certain data sets have inspired researchers to become interested in developing methodologies in this context. There are still very interesting questions to be answered. For example, the violation of assumptions is not yet a subject sufficiently studied in the literature. The use of descriptors other than Ripley's K-function or its generalisations (marked, spatio-temporal etc.) has been proposed but never tested in practice. Support changes are not yet explored when there are replicated patterns. As authors, we believe that these methodologies are in their beginning and that they will increasingly be demanded and used by scientists from different branches of science.

Bibliography

Adelfio, G. and Chiodi, M. (2010). Diagnostics for nonparametric estimation in space-time seismic processes, *Journal of Environmental Statistics* **1**(2): 1–13.

Adelfio, G. and Chiodi, M. (2015). Alternated estimation in semi-parametric space-time branching-type point processes with application to seismic catalogs, *Stochastic Environmental Research and Risk Assessment* **29**(2): 443–450.

Adelfio, G. and Ogata, Y. (2010). Hybrid kernel estimates of space-time earthquake occurrence rates using the epidemic-type aftershock sequence model, *Annals of the Institute of Statistical Mathematics* **62**(1): 127–143.

Adelfio, G. and Schoenberg, F. P. (2009). Point process diagnostics based on weighted second-order statistics and their asymptotic properties, *Annals of the Institute of Statistical Mathematics* **61**(4): 929–948.

Altieri, L., Scott, E., Cocchi, D. and Illian, J. (2015). A changepoint analysis of spatio-temporal point processes, *Spatial Statistics* **14**: 197–207.

American Meteorological Society (2013). Glossary of meteorology.
URL: *http://glossary.ametsoc.org/wiki/Main_Page*

Andersen, I. T. and Hahn, U. (2016). Matérn thinned cox processes, *Spatial Statistics* **15**: 1–21.

Anderson, M. and Braak, C. T. (2003). Permutation tests for multi-factorial analysis of variance, *Journal of Statistical Computation and Simulation* **73**(2): 85–113.

Assunção, R. and Maia, A. (2007). A note on testing separability in spatial-temporal marked point processes, *Biometrics* **63**(1): 290–294.

Baddeley, A., Boyde, A., Reid, S. and Howard, C. (1987). Three-dimensional analysis of the spatial distribution of particles using the tandem-scanning reflected light microscope, *Acta Stereologica* **6**(supplement II): 87–100.

Baddeley, A., Hardegen, A., Lawrence, T., Milne, R. K., Nair, G. and Rakshit, S. (2017). On two-stage Monte Carlo tests of composite hypotheses, *Computational Statistics & Data Analysis* **114**: 75–87.

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

Baddeley, A., Moyeed, R., Howard, C. and Boyde, A. (1993). Analysis of a three-dimensional point pattern with replication, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **42**(4): 641–668.

Baddeley, A., Rubak, E. and Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with R*, Chapman & Hall Interdisciplinary Statistics Series, CRC Press, Boca Raton, Florida.

Baddeley, A. and Turner, R. (2000). Practical maximum pseudolikelihood for spatial point patterns, *Australian & New Zealand Journal of Statistics* **42**(3): 283–322.

Baddeley, A., Turner, R., Møller, J. and Hazelton, M. (2005). Residual analysis for spatial point processes, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(5): 617–666.

Bagchi, R. and Illian, J. B. (2015). A method for analysing replicated point patterns in ecology, *Methods in Ecology and Evolution* **6**(4): 482–490.

Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall Monographs on Statistics & Applied Probability, second edn, CRC Press, Boca Raton, Florida.

Bedford, T. and van den Berg, J. (1997). A remark on van Lieshout and Baddeley's *J*-function for point processes, *Advances in Applied Probability* **29**: 19–25.

Beneš, V., Prokešová, M., Helisová, K. S. and Zikmundová, M. (2015). *Space-Time Models in Stochastic Geometry*, Springer International Publishing, Cham.

Berman, M. and Diggle, P. J. (1989). Estimating weighted integrals of the second-order intensity of a spatial point process, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **51**(1): 81–92.

Berthelsen, K. and Møller, J. (2002). Spatial jump processes and perfect simulation, *in* K. R. Mecke and D. Stoyan (eds), *Morphology of condensed matter: Physics and geometry of spatially complex systems*, Spriger-Verlag, Berlin, pp. 391–417.

Besag, J. (1977). Contribution to the discussion of dr Ripley's paper, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **39**: 193–195.

Besag, J. and Diggle, P. J. (1977). Simple Monte Carlo tests for spatial pattern, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **26**(3): 327–333.

Board, W. E. (2015). World register of marine species (worms), =http://www.marinespecies.org.

Bowsher, C. G. (2007). Modelling security market events in continuous time: Intensity based, multivariate point process models, *Journal of Econometrics* **141**(2): 876–912.

Brix, A. and Diggle, P. J. (2001). Spatiotemporal prediction for log-Gaussian Cox processes, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63**(4): 823–841.

Brown, P., Karesen, K., Roberts, G. and Tonellato, S. (2000). Blur-generated non-separable space-time models, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62**(4): 847–860.

Bruening, S., Kay, M. and Brooks, H. (2002). A new perspective on the climatology of tornadoes in the united states, *Preprints, 16th Conference on Probability and Statistics, Orlando, Florida*, American Meteorological Society, pp. J96–J103.

Central Intelligence Agency (2017). *The World Factbook*, Washington, DC.
URL: *https://www.cia.gov/library/publications/the-world-factbook/index.html*

Chakraborty, A., Gelfand, A. E., Wilson, A. M., Latimer, A. M. and Silander, J. A. (2011). Point pattern modelling for degraded presence-only data over large regions, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **60**(5): 757–776.

Chang, C.-H. and Schoenberg, F. (2011). Testing separability in marked multidimensional point processes with covariates, *Annals of the Institute of Statistical Mathematics* **63**(6): 1103–1122.

Chiu, S. N., Stoyan, D., Kendall, W. S. and Mecke, J. (2013). *Stochastic Geometry and Its Applications*, Wiley Series in Probability and Statistics, third edn, John Wiley & Sons, Chichester.

Choi, E. and Hall, P. (1999). Nonparametric approach to analysis of space-time data on earthquake occurrences, *Journal of Computational and Graphical Statistics* **8**: 733–748.

Clements, R. A., Schoenberg, F. and Schorlemmer, D. (2011). Residual analysis methods for space-time point processes with applications to earthquake forecast models in California, *The Annals of Applied Statistics* **5**(4): 2549–2571.

Comas, C., Rodriguez-Cortes, F. J. and Mateu, J. (2015). Second-order analysis of anisotropic spatiotemporal point process data, *Statistica Neerlandica* **69**(1): 49–66.

Concannon, P. R., Brooks, H. E. and Doswell III, C. A. (2000). Climatological risk of strong and violent tornadoes in the united states, *Preprints, 2nd Symposium on Environmental Applications, Long Beach, California*, American Meteorological Society, pp. 212–219.

Cox, D. R. (1955). Some statistical methods connected with series of events, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **17**(2): 129–164.

Cox, D. R. (1975). Partial likelihood, *Biometrika* **62**(2): 269–276.

Cox, D. R. and Isham, V. (1980). *Point Processes*, Chapman & Hall. CRC Press, London.

Cressie, N. A. C. (1993). *Statistics for Spatial Data, revised ed*, Wiley, New York.

Cressie, N. and Collins, B. (2001). Analysis of spatial point patterns using bundles of product density lisa functions, *Journal of Agricultural, Biological, and Environmental Statistics* **6**(1): 118–135.

Cressie, N. and Huang, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions, *Journal of the American Statistical Association* **94**(448): 1330–1340.

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*, Wiley Series in Probability and Statistics, John Wiley & Sons, New Jersey.

Cronie, O. and Särkkä, A. (2011). Some edge correction methods for marked spatio-temporal point process models, *Computational Statistics & Data Analysis* **55**(7): 2209–2220.

Cronie, O. and van Lieshout, M. (2015). A *J*-function for inhomogeneous spatio-temporal point processes, *Scandinavian Journal of Statistics* **42**(2): 562–579.

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

Daley, D. and Vere-Jones, D. (2007). *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*, second edn, Springer-Verlag, New York.

Deza, M. M. and Deza, E. (2009). *Encyclopedia of Distances*, Springer-Verlag, Berlin.

Díaz-Avalos, C., Juan, P. and Mateu, J. (2013). Similarity measures of conditional intensity functions to test separability in multidimensional point processes, *Stochastic Environmental Research and Risk Assessment* **27**(5): 1193–1205.

Díaz-Avalos, C., Juan, P. and Mateu, J. (2014). Significance tests for covariate-dependent trends in inhomogeneous spatio-temporal point processes, *Stochastic Environmental Research and Risk Assessment* **28**(3): 593–609.

Diggle, P. J. (1979). On parameter estimation and goodness-of-fit testing for spatial point patterns, *Biometrics* **35**(1): 87–101.

Diggle, P. J. (1985). A kernel method for smoothing point process data, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **34**(2): 138–147.

Diggle, P. J. (1986). Displaced amacrine cells in the retina of a rabbit: analysis of a bivariate spatial point pattern, *Journal of Neuroscience Methods* **18**(1–2): 115–125.

Diggle, P. J. (2006). Spatio-temporal point processes, partial likelihood, foot and mouth disease, *Statistical Methods in Medical Research* **15**(4): 325–336.

Diggle, P. J. (2013). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*, Chapman & Hall Monographs on Statistics & Applied Probability, third edn, CRC Press, Boca Raton, Florida.

Diggle, P. J., Chetwynd, A., Häggkvist, R. and Morris, S. (1995). Second-order analysis of space-time clustering, *Statistical Methods in Medical Research* **4**(2): 124–136.

Diggle, P. J., Guan, Y., Hart, A., Paize, F. and Stanton, M. (2010). Estimating individual-level risk in spatial epidemiology using spatially aggregated information on the population at risk, *Journal of the American Statistical Association* **105**(492): 1394–1402.

Diggle, P. J., Kaimi, I. and Abellana, R. (2010). Partial-likelihood analysis of spatio-temporal point-process data, *Biometrics* **66**(2): 347–354.

Diggle, P. J., Lange, N. and Beneš, F. M. (1991). Analysis of variance for replicated spatial point patterns in clinical neuroanatomy, *Journal of the American Statistical Association* **86**(415): 618–625.

Diggle, P. J., Mateu, J. and Clough, H. (2000). A comparison between parametric and non-parametric approaches to the analysis of replicated spatial point patterns, *Advances in Applied Probability* **32**(2): 331–343.

Diggle, P. J., Menezes, R. and Su, T.-l. (2010). Geostatistical inference under preferential sampling, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **59**(2): 191–232.

Diggle, P. J., Moraga, P., Rowlingson, B. and Taylor, B. (2013). Spatial and spatio-temporal log-Gaussian Cox processes: Extending the geostatistical paradigm, *Statistical Science* **28**: 542–563.

Diggle, P. J., Rowlingson, B. and Su, T.-l. (2005). Point process methodology for on-line spatio-temporal disease surveillance, *Environmetrics* **16**(5): 423–434.

Doswell III, C. A. and Burgess, D. W. (1988). On some issues of united states tornado climatology, *Monthly Weather Review* **116**(2): 495–501.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife, *The Annals of Statistics* **7**(1): 1–26.

Emery, X., Kracht, W., Egaña, Á. and Garrido, F. (2012). Using two-point set statistics to estimate the diameter distribution in boolean models with circular grains, *Mathematical Geosciences* **44**(7): 805–822.

Epanechnikov, V. A. (1969). Non-parametric estimation of a multivariate probability density, *Theory of Probability & Its Applications* **14**(1): 153–158.

Feldt, L. S. and Mahmoud, M. W. (1958). Power function charts for specification of sample size in analysis of variance, *Psychometrika* **23**(3): 201–210.

Fiksel, T. (1988). Edge-corrected density estimators for point processes, *Statistics* **19**(1): 67–75.

Finkenstadt, B., Held, L. and Isham, V. (2007). *Statistical Methods for Spatio-Temporal Systems*, Chapman & Hall Monographs on Statistics & Applied Probability, CRC Press, Boca Raton, Florida.

Gabriel, E. (2014). Estimating second-order characteristics of inhomogeneous spatio-temporal point processes, *Methodology and Computing in Applied Probability* **16**: 411–431.

Gabriel, E. and Diggle, P. J. (2009). Second-order analysis of inhomogeneous spatio-temporal point process data, *Statistica Neerlandica* **63**(1): 43–51.

Gabriel, E., Rowlingson, B. and Diggle, P. J. (2013). stpp: An R package for plotting, simulating and analyzing spatio-temporal point patterns, *Journal of Statistical Software* **53(2)**(2): 1–29.

Gelfand, A. E., Diggle, P., Guttorp, P. and Fuentes, M. (eds) (2010). *Handbook of spatial statistics*, Chapman & Hall Handbooks of Modern Statistical Methods, CRC Press, Boca Raton, Florida.

Geyer, C. (1999). Likelihood inference for spatial point processes: Likelihood and computation, *in* O. E. Barndorff-Nielsen, W. S. Kendall and M. N. M. van Lieshout (eds), *Stochastic Geometry (Toulouse, 1996)*, Vol. 80 of *Chapman & Hall Monographs on Statistics & Applied Probability*, CRC Press, Boca Raton, Florida, pp. 79–140.

Ghorbani, M. (2013). Testing the weak stationarity of a spatio-temporal point process, *Stochastic Environmental Research and Risk Assessment* **27**(2): 517–524.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data, *Journal of the American Statistical Association* **97**(458): 590–600.

Gneiting, T. and Guttorp, P. (2010). Continuous parameter spatio-temporal processes, *in* A. E. Gelfand, P. J. Diggle, M. Fuentes and P. Guttorp (eds), *Handbook of Spatial Statistics*, Chapman & Hall Handbooks of Modern Statistical Methods, CRC Press, Boca Raton, Florida, pp. 427–436.

Gómez, C. and Finch, J. (2007). Gas dispersion measurements in flotation cells, *International Journal of Mineral Processing* **84**(1): 51–58.

Gómez, C., Mesías, J. and Álvarez, J. (2016). Bubble surface area flux and performance in laboratory flotation testing, *XXVIII International Mineral Processing Congress Proceedings*, Canadian Institute of Mining, Metallurgy and Petroleum.

González, J. A., Rodríguez-Cortés, F. J., Cronie, O. and Mateu, J. (2016). Spatio-temporal point process statistics: A review, *Spatial Statistics* **18**: 505–544.

Good, P. I. (2005). *Permutation, Parametric and Bootstrap Tests of Hypotheses*, Springer Series in Statistics, 3 edn, Springer-Verlag, New York.

Gorain, B., Franzidis, J. and Manlapig, E. (1997). Studies on impeller type, impeller speed and air flow rate in an industrial scale flotation cell. part 4: Effect of bubble surface area flux on flotation performance, *Minerals Engineering* **10**(4): 367–379.

Gorain, B., Franzidis, J.-P. and Manlapig, E. (1999). The empirical prediction of bubble surface area flux in mechanical flotation cells from cell design and operating data, *Minerals Engineering* **12**(3): 309–322.

Grazulis, T. P. (1993). *Significant Tornadoes, 1680-1991*, Environmental Films St. Johnsbury, VT.

Greenspan, B. M. (2013). *Survey of some recent advances in spatial-temporal point processes*, PhD thesis, University of California, Los Angeles.

Hahn, U. (2012). A studentized permutation test for the comparison of spatial point patterns, *Journal of American Statistical Association* **107**(498): 754–764.

Hahn, U. and Vedel Jensen, E. B. (2016). Hidden second-order stationary spatial point processes, *Scandinavian Journal of Statistics* **43**(2): 455–475.

Hamner, W., Hamner, P., Strand, S. and Gilmer, R. (1983). Behavior of antarctic krill, euphausia superba: Chemoreception, feeding, schooling, and molting, *Science* **220**(4595): 433–435.

Hellmund, G., Prokešová, M. and Jensen, E. B. V. (2008). Lévy-based Cox point processes, *Advances in Applied Probability* **40**(3): 603–629.

Hernandez-Aguilar, J., Coleman, R., Gomez, C. and Finch, J. (2004). A comparison between capillary and imaging techniques for sizing bubbles in flotation systems, *Minerals Engineering* **17**(1): 53–61.

Ho, L. P. and Chiu, S. N. (2006). Testing the complete spatial randomness by Diggle's test without an arbitrary upper limit, *Journal of Statistical Computation and Simulation* **76**(7): 585–591.

Ho, L. P. and Stoyan, D. (2008). Modelling marked point patterns by intensity-marked Cox processes, *Statistics & Probability Letters* **78**(10): 1194–1199.

Illian, J., Penttinen, P., Stoyan, H. and Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*, Statistics in Practice, Wiley.

Jafari-Mamaghani, M., Andersson, M. and Krieger, P. (2010). Spatial point pattern analysis of neurons using Ripley's K- function in $3D$, *Frontier in Neuroinformatics* **4**(9): 1–10.

Karpman, D., Ferreira, M. A. and Wikle, C. K. (2013). A point process model for tornado report climatology, *Stat* **2**(1): 1–8.

Karr, A. (1991). *Point Processes and Their Statistical Inference*, second edn, Marcel Dekker, New York.

Kracht, W., Emery, X. and Paredes, C. (2013). A stochastic approach for measuring bubble size distribution via image analysis, *International Journal of Mineral Processing* **121**: 6–11.

Landau, S. and Everall, I. P. (2008). Nonparametric bootstrap for *K*-functions arising from mixed-effects models with applications in neuropathology, *Statistica Sinica* **18**(4): 1375–1393.

Landau, S., Rabe-Hesketh, S. and Everall, I. P. (2004). Nonparametric one-way analysis of variance of replicated bivariate spatial point patterns, *Biometrical Journal* **46**(1): 19–34.

Laskowski, J. (2001). *Coal Flotation and Fine Coal Utilization*, Developments in Mineral Processing, Elsevier Science, Amsterdam.

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

Lotwick, H. W. and Silverman, B. W. (1981). Convergence of spatial birth-and-death processes, *Math. Proc. Cambridge Philos. Soc.* **90**: 155–165.

Lotwick, H. W. and Silverman, B. W. (1982). Methods for analysing spatial processes of several types of points, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **44**(3): 406–413.

Lund, J. and Rudemo, M. (2000). Models for point processes observed with noise, *Biometrika* **87**(2): 235–249.

Ma, C. (2003). Families of spatio-temporal stationary covariance models, *Journal of Statistical Planning and Inference* (116): 489–501.

Ma, C. (2008). Recent developments on the construction of spatio-temporal covariance models, *Stochastic Environmental Research and Risk Assessment* **22**(Supplement 1): S39–S47.

Marsan, D. and Lengliné, O. (2008). Extending earthquakes' reach through cascading, *Science* **319**(5866): 1076–1079.

Marsan, D. and Lengliné, O. (2010). A new estimation of the decay of aftershock density with distance to the mainshock, *Journal of Geophysical Research: Solid Earth* **115**(B9).

Marzban, C. and Schaefer, J. T. (2001). The correlation between U.S. tornadoes and pacific sea surface temperatures, *Monthly Weather Review* **129**(4): 884–895.

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

Mateu, J., Schoenberg, F. P., Diez, D. M., González, J. A. and Lu, W. (2015). On measures of dissimilarity between point patterns: Classification based on prototypes and multidimensional scaling, *Biometrical Journal* **57**(2): 340–358.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, Chapman & Hall Monographs on Statistics and Applied Probability, second edn, CRC Press, Boca Raton, Florida.

Miskovic, S. and Luttrell, G. (2009). Column and non-traditional flotation, *in* D. Malhotra, P. R. Taylor, E. Spiller and M. LeVier (eds), *Recent Advances in Mineral Processing Plant Design*, Society for Mining, Metallurgy, and Exploration. Littleton, Colorado.

Miskovic, S. and Luttrell, G. (2012). Comparison of two bubble sizing methods for performance evaluation of mechanical flotation cells, *in* C. A. Young and G. H. Luttrell (eds), *Separation Technologies for Minerals, Coal, and Earth Resources*, Society for Mining, Metallurgy, and Exploration. Englewood, Colorado.

Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P. and Tita, G. E. (2011). Self-exciting point process modeling of crime, *Journal of the American Statistical Association* **106**(493): 100–108.

Møller, J. (2003). Shot noise cox processes, *Advances in Applied Probability* **35**(3): 614–640.

Møller, J. and Díaz-Avalos, C. (2010). Structured spatio-temporal shot-noise Cox point process models, with a view to modelling forest fires, *Scandinavian Journal of Statistics* **37**(1): 2–25.

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., Ghorbani, M. and Rubak, E. (2016). Mechanistic spatio-temporal point process models for marked point processes, with a view to forest stand data, *Biometrics* **72**(3): 687–696.

Møller, J. and Rasmussen, J. G. (2006). Approximate simulation of hawkes processes, *Methodology and Computing in Applied Probability* **8**(1): 53–64.

Møller, J., Safavimanesh, F. and Rasmussen, J. (2015). The cylindrical K-function and poisson line cluster point processes, *ArXiv e-prints* .

Møller, J. and Sørensen, M. (1994). Statistical analysis of a spatial birth-and-death process model with a view to modelling linear dune fields, *Scandinavian Journal of Statistics* **21**: 1–19.

Møller, J., Syversveen, A. R. and Waagepetersen, R. P. (1998). Log Gaussian Cox processes, *Scandinavian Journal of Statistics* **25**(3): 451–482.

Møller, J. and Toftaker, H. (2014). Geometric anisotropic spatial point pattern analysis and Cox processes, *Scandinavian Journal of Statistics* **41**: 414–435.

Møller, J. and Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*, Chapman & Hall Monographs on Statistics & Applied Probability, CRC Press, Boca Raton, Florida.

Monfredo, W. (1999). Relationships between phases of the el niño-southern oscillation and character of the tornado season in the south-central united states, *Physical Geography* **20**(5): 413–421.

Musmeci, F. and Vere-Jones, D. (1992). A space-time clustering model for historical earthquakes, *Annals of the Institute of Statistical Mathematics* **44**(1): 1–11.

Myllymäki, M., Särkkä, A. and Vehtari, A. (2014). Hierarchical second-order analysis of replicated spatial point patterns with non-spatial covariates, *Spatial Statistics* **8**: 104–121.

Mylne, A., Brady, O. J., Huang, Z., Pigott, D. M., Golding, N., Kraemer, M. U. and Hay, S. I. (2014). A comprehensive database of the geographic spread of past human ebola outbreaks, *Scientific Data* .
URL: *http://www.nature.com/sdata*

Nesset, J. E., Hernandez-Aguilar, J. R., Acuna, C., Gomez, C. O. and Finch, J. A. (2006). Some gas dispersion characteristics of mechanical flotation machines, *Minerals Engineering* **19**(6): 807–815.

Neyman, J. and Scott, E. L. (1958). Statistical approach to problems of cosmology, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **20**(1): 1–43.

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

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

Ogata, Y. and Zhuang, J. (2006). Space-time ETAS models and an improved extension, *Tectonophysics* **413**(1): 13–23.

Ohser, J. and Mücklich, F. (2000). *Statistical Analysis of Microstructures in Materials Science*, Wiley, Chichester.

Ohser, J. and Stoyan, D. (1981). On the second-order and orientation analysis of planar stationary point processes, *Biometrical Journal* **23**: 523–533.

Peebles, P. J. E. and Groth, E. J. (1975). Statistical analysis of extragalactic objects. V: three-point correlation function for the galaxy distribution in the Zwicky catalog, *Astrophysical Journal* **196**: 1–11.

Peng, R. D., Schoenberg, F. P. and Woods, J. A. (2005). A space–time conditional intensity model for evaluating a wildfire Hazard index, *Journal of the American Statistical Association* **100**(469): 26–35.

Penttinen, A., Stoyan, D. and Henttonen, H. M. (1992). Marked point processes in forest statistics, *Forest Science* **38**(4): 806–824.

Pereira, P., Turkman, K. F., Turkman, M. A. A., Sá, A. and Pereira, J. M. (2013). Quantification of annual wildfire risk; a spatio-temporal point process approach., *Statistica* **73**(1): 55.

Preisler, H. K., Ager, A. A., Johnson, B. K. and Kie, J. G. (2004). Modeling animal movements using stochastic differential equations, *Environmetrics* **15**(7): 643–657.

Prokešová, M. and Dvořák, J. (2013). Statistics for inhomogeneous space-time shot-noise Cox processes, *Methodology and Computing in Applied Probability* **16**(2): 433–449.

Ramón, P., de la Cruz, M., Chacón-Labella, J. and Escudero, A. (2016). A new non-parametric method for analyzing replicated point patterns in ecology, *Ecography* **39**(11): 1109–1117.

Rathbun, S. L. and Cressie, N. (1994). A space-time survival point process for a longleaf pine forest in southern Georgia, *Journal of the American Statistical Association* **89**(428): 1164–1174.

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

Ripley, B. D. (1977). Modelling spatial patterns (with discussion), *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **39**(2): 172–212.

Ripley, B. D. (1979). Tests of "randomness" for spatial point patterns, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **41**(3): 368–374.

Ripley, B. D. (1988). *Statistical Inference for Spatial Processes*, Cambridge University Press. Cambridge.

Rodrigues, A. and Diggle, P. (2010). A class of convolution-based models for spatio-temporal processes with non-separable covariance structure, *Scandinavian Journal of Statistics* **37**: 553–567.

Rodrigues, A. and Diggle, P. J. (2012). Bayesian estimation and prediction for inhomogeneous spatiotemporal log-Gaussian Cox processes using low-rank models, with application to criminal surveillance, *Journal of the American Statistical Association* **107**(497): 93–101.

Rodríguez-Cortés, F. J., Ghorbani, M., Mateu, J. and Stoyan, D. (2014). On the expected value and variance for an estimator of the spatio-temporal product density function, *Department of Mathematical Sciences, Aalborg University* **Research Report Series**(0): 1 – 19.

Rue, H., Martino, S. and Chopin, N. (2009). Approximate bayesian inference for latent Gaussian models by using integrated nested Laplace approximations, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71**(2): 319–392.

Schlather, M., Ribeiro, P. J. and Diggle, P. J. (2004). Detecting dependence between marks and locations of marked point processes, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66**(1): 79–93.

Schlesinger, M. E., King, M. J., Sole, K. C. and Davenport, W. G. (2011). *Extractive Metallurgy of Copper*, fifth edn, Elsevier Science, Amsterdam.

Schoenberg, F. P. (2003). Multidimensional residual analysis of point process models for earthquake occurrences, *Journal of the American Statistical Association* **98**(464): 789–795.

Schoenberg, F. P. (2004). Testing separability in spatial-temporal marked point processes, *Biometrics* **60**(2): 471–481.

Schoenberg, F. P., Brillinger, D. R. and Guttorp, P. (2006). Point processes, spatial-temporal, *in* A. H. El-Shaarawi and W. W. Piegorsch (eds), *Encyclopedia of Environmetrics*, John Wiley & Sons, Chichester, chapter 3, pp. 1573–1577.

Schoenberg, F. P., Cochran, J. J., Cox, L. A., Keskinocak, P., Kharoufeh, J. P. and Smith, J. C. (2010). *Wiley encyclopedia of operations research and management science*, John Wiley & Sons, chapter Introduction to Point Processes, pp. 2438–2811.

Schorlemmer, D., Gerstenberger, M., Wiemer, S., Jackson, D. and Rhoades, D. (2007). Earthquake likelihood model testing, *Seismological Research Letters* **78**(1): 17–29.

Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **53**(3): 683–690.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, Chapman & Hall Monographs on Statistics & Applied Probability, CRC press, Boca Raton, Florida.

Snyder, D. L. and Miller, M. I. (1991). *Random Point Processes in Time and Space*, second edn, Springer-Verlag, New York.

Stoyan, D. and Stoyan, H. (1994). *Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics*, John Wiley & Sons, Chichester.

Tamayo-Uria, I., Mateu, J. and Diggle, P. J. (2014). Modelling of the spatio-temporal distribution of rat sightings in an urban environment, *Spatial Statistics* **9**: 192–206.

Taylor, B. and Diggle, P. (2014). INLA or MCMC? a tutorial and comparative evaluation for spatial prediction in log-Gaussian Cox processes, *Journal of Statistical Computation and Simulation* **84**(10): 2266–2284.

Thorsten Wiegand, K. A. M. (2014). *Handbook of Spatial Point-Pattern Analysis in Ecology*, Chapman & Hall Applied Environmental Statistics, CRC Press, Boca Raton, Florida.

Tranbarger, K. E. and Schoenberg, F. P. (2010). On the computation and application of prototype point patterns, *Open Applied Informatics Journal* **4**: 1–9.

Van de Putte, A., Youdjou, N. and Danis, B. (2015). The Antarctic biodiversity information facility.
URL: *http://www.biodiversity.aq*

van Lieshout, M. N. M. (2011). On estimation of the intensity function of a point process, *Methodology and Computing in Applied Probability* **14**(3): 567–578.

van Lieshout, M. N. M. and Baddeley, A. J. (1996). A nonparametric measure of spatial interaction in point patterns, *Statistica Neerlandica* **50**(3): 344–361.

van Lieshout, M. N. M. and Stein, A. (2012). Earthquake modelling at the country level using aggregated spatio-temporal point processes, *Mathematical Geosciences* **44**(3): 309–326.

Veen, A. and Schoenberg, F. P. (2006). *Assessing Spatial Point Process Models Using Weighted K-functions: Analysis of California Earthquakes*, Springer-Verlag, New York, pp. 293–306.

Vere-Jones, D. (2009). Some models and procedures for space-time point processes, *Environmental and Ecological Statistics* **16**: 173–195.

Wikle, C. K. and Anderson, C. J. (2003). Climatological analysis of tornado report counts using a hierarchical bayesian spatiotemporal model, *Journal of Geophysical Research: Atmospheres* **108**(D24): 9005–9019.

Wilson, H. E. (1998). *Statistical Analysis of Replicated Spatial Point Patterns. Ph. D Thesis.*, Lancaster University.

Zhang, J.-T. (2013). *Analysis of Variance for Functional Data*, Chapman & Hall Monographs on Statistics & Applied Probability, CRC Press, Boca Raton, Florida.

Zhuang, J. (2006). Second-order residual analysis of spatiotemporal point processes and applications in model evaluation, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**(4): 635–653.

Zhuang, J., Ogata, Y. and Vere-Jones, D. (2002). Stochastic declustering of space-time earthquake occurrences, *Journal of the American Statistical Association* **97**(458): 369–380.

Appendix A

Resumen

Los procesos puntuales constituyen una poderosa herramienta para el análisis de conjuntos de objetos aleatoriamente localizados en un plano, en el espacio, o en general en un espacio abstracto. Estos objetos pueden ser definidos rigurosamente como colecciones localmente finitas de espacios segundo-contables de Hausdorff (Daley and Vere-Jones, 2003, 2007). Las aplicaciones de esta teoría son diversas e incluyen una considerable variedad de situaciones que provienen de la realidad. Estas aplicaciones incluyen astronomía (Neyman and Scott, 1958), biología (Diggle, 2013), ecología (Thorsten Wiegand, 2014), ciencias económicas (Bowsher, 2007), epidemiología (Diggle, Guan, Hart, Paize and Stanton, 2010), ingeniería de materiales (Ohser and Mücklich, 2000), medicina (Andersen and Hahn, 2016; Diggle et al., 1991; Hahn, 2012), seismología (Tranbarger and Schoenberg, 2010) entre otras Illian et al. (2008). Existe además una extensa bibliografía que cubre un amplio espectro tanto teórico (Chiu et al., 2013; Daley and Vere-Jones, 2003, 2007; Møller and Waagepetersen, 2004), como práctico. El aspecto práctico va desde la resolución de problemas que provienen de la realidad y de bases de datos cada vez más complejas y exigentes (Diggle, 2013), hasta la implementación de las diversas metodologías en software sencillo, libre y asequible (Baddeley et al., 2015; Gabriel et al., 2013).

Uno de los supuestos habituales en el trabajo con procesos puntuales es que se dispone de un único patrón observado en la realidad, es decir un único conjunto de puntos localizado en una región de observación. La distribución de este patrón es matemáticamente compleja y esto ha llevado a que se propongan en la literatura algunos *summary statistics* que pueden mostrar o describir la naturaleza de dicha distribución. Normalmente estos estadísticos son funciones o números reales que se basan en las distancias entre los puntos del patrón o en los conteos. Por

ejemplo, la intensidad de primer-orden describe la densidad de los puntos de un patrón puntual. Las famosas características de segundo orden (Chiu et al., 2013; Illian et al., 2008), son habitualmente utilizadas por los científicos puesto que estas describen la atracción o repulsión que pueda existir entre puntos del patrón. En general estas funciones o descriptores pueden ser estimados de forma no-paramétrica y algunos de los más conocidos y utilizados en la literatura son la *K*-función de Ripley (Ripley, 1977) y su respectiva versión normalizada, la *L*-función (Besag and Diggle, 1977); la *J*-función (van Lieshout and Baddeley, 1996) y la pair-correlation function.

Algunas bases de datos de patrones puntuales contienen más de un patrón observado, en este caso responder las preguntas que suelen traer consigo las bases de datos resulta complejo y pocos autores han trabajado en el tema. Baddeley et al. (1987, 1993) presentó un enfoque utilizando *ratio-regressions* en patrones puntuales tridimensionales replicados utilizando la *K*-función tridimensional. Diggle et al. (1991) presentó un enfoque de análisis de la varianza en una aplicación de neuroanatomía clínica a través de la comparación de *K*-funciones estimadas e inferencia no-paramétrica a través de un test de Monte Carlo. Este método fue posteriormente mejorado por Diggle et al. (2000) y por Hahn (2012) quien desarrolló un test de permutación estudentizado para la comparación de varios grupos de patrones puntuales espaciales. Algunos autores (Bagchi and Illian, 2015; Landau and Everall, 2008; Landau et al., 2004; Myllymäki et al., 2014) han incluido, por ejemplo, predictores categóricos y continuos en modelos lineales de efectos mixtos o marcas discretas para la explicación de la variación en la estructura espacial de patrones replicados.

El objetivo fundamental de esta tesis es introducir un nuevo conjunto de test para comparar *K*-funciones estimadas (o en general cualquier descriptor funcional adecuado). Se dispone de dos líneas paralelas para el desarrollo de los nuevos test.

En primer lugar, se extiende el test de permutación estudentizado propuesto por Hahn (2012) al caso espacio-temporal. En vista de que el estudio de los procesos puntuales espacio-temporales no ha sido tan profundo en la literatura como el estudio de los procesos puntuales espaciales, se hace un resumen completo, a modo de texto de consulta, de las técnicas disponibles y los enfoques de análisis en el contexto espacio-temporal y se aplican algunas de estas técnicas a tres bases de datos de diferentes campos científicos (González et al., 2016). El nuevo test de permutación espacio-temporal muestra un rendimiento adecuado en términos de

nivel empírico y de potencia. Este test está motivado por una base de datos compleja: el registro de tornados en estados unidos en un periodo de 36 años. Es por esto que además del test en si mismo, se han desarrollado algunas herramientas adicionales como un estimador no-separable de la intensidad espacio-temporal de primer orden, que permite un análisis mucho más realista del fenómeno a través del nuevo test.

En segundo lugar, se considera un diseño ANOVA two-way donde las observaciones son patrones puntuales espaciales y donde además, se dispone de réplicas por cada una de las combinaciones de los tratamientos. Esta metodología viene motivada por un experimento de ingeniería de materiales en donde se han medido las localizaciones de las burbujas resultantes en una celda de flotación en un tiempo fijo y se han variado los niveles de dos factores implicados en el experimento. Se desarrolla un esquema de análisis de la varianza para experimentos factoriales, es decir, desarrollamos estadísticos eficientes para testear, de forma no-paramétrica, la influencia de los factores y el posible efecto de interacción. Para esto se proponen varias posibilidades para el procedimiento inferencia a través de permutaciones aleatorias y se prueba que son útiles en la práctica.

La tesis está organizada como sigue. En primer lugar se presenta la información introductoria (Capítulos 1 y 2). La particularidad de los capítulos subsiguientes es que son auto-contenidos, puesto que corresponden a artículos ya publicados (en el caso del Capítulo 3), o en proceso de publicación (Capítulos 4 y 5). Por lo tanto se pueden abordar de forma independiente.

En el Capítulo 2, se presenta un breve resumen de la teoría de los procesos puntuales espaciales. Se proporcionan los conceptos básicos a nivel abstracto, se introducen algunas medidas típicas importantes. Se definen teóricamente algunos modelos clásicos de procesos puntuales. Se presentan los summary-statistics espaciales y finalmente se enseñan algunos estimadores no paramétricos de dichos summary-statistics.

El Capítulo 3 se comienza con una breve descripción de las bases de datos que se analizan. Posteriormente se presenta una introducción a los procesos puntuales espacio-temporales. Luego se presentan algunas características (descriptores) de primer y segundo orden para procesos puntuales espacio-temporales. Se presenta un resumen de los modelos empíricos y mecanísticos y finalmente se concluye con una discusión general y algunas futuras ideas de investigación.

En el Capítulo 4 se considera primero una descripción de los datos de los tornados en los U.S. A continuación se procede a estudiar la función de intensidad

de primer orden espacio-temporal y a analizar la hipótesis de separabilidad en dos temporadas de tornados, fría y cálida. Se considera una partición espacial de las regiones del mapa con el fin de simplificar el análisis al considerarse los elementos de la partición como posibles réplicas de un proceso puntual subyacente. Se comprueba el rendimiento del test de permutación espacio-temporal y se aplica a la base de datos. Finalmente, se discuten algunas preguntas y problemas. Este capítulo motiva interesantes líneas de investigación. Por ejemplo, el desarrollo de test adecuados que no requieran la hipótesis de intercambiabilidad Hahn (2012). Adicionalmente, la selección del ancho de banda en el contexto espacio-temporal y la posible inclusión de dimensiones adicionales en los patrones observados como covariables o marcas.

El Capítulo 5 comienza con una descripción del procedimiento de ingeniería mediante el que se obtienen los datos de las burbujas en la celda de flotación. Posteriormente se hace una descripción completa de la base de datos. Se presenta un pequeño apartado donde se describen las herramientas matemáticas utilizadas. Se da paso entonces a una motivación del análisis ANOVA de dos factores partiendo desde el caso más simple para llegar al diseño ANOVA two-way balanceado. Se proponen los estadísticos necesarios para el análisis y se presenta un escenario de simulación para comprobar que los estadísticos funcionan adecuadamente. Finalmente se aplican los estadísticos a la base de datos. Este capítulo también motiva algunas líneas de investigación, tales como la extensión a diseños más complejos como diseños anidados y diseños con efectos fijos y mixtos a través de la utilización de estadísticos de tipo Fisher. El uso de descriptores funcionales más complejos (que incluyan marcas por ejemplo) en el diseño ANOVA también constituye un campo abierto de investigación. Finalmente, la definición de los estadísticos del diseño ANOVA pero utilizando distancias funcionales complejas entre patrones Mateu et al. (2015) o entre funciones, podría ser un desafiante campo de investigación.