Spatial Statistics 18 (2016) 505–544

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Spatio-temporal point process statistics: A review☆

Jonatan A. González [a,*], Francisco J. Rodríguez-Cortés [a], Ottmar Cronie [b], Jorge Mateu [a]

[a] Department of Mathematics, University Jaume I, E-12071, Castellón, Spain

[b] Department of Mathematics and Mathematical Statistics, Umeå University, Umeå, Sweden

ARTICLE INFO

Article history:
Received 26 February 2016
Accepted 27 October 2016
Available online 3 November 2016

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

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.

☆ Work partially funded by Grant MTM2013-43917-P from the Spanish Ministry of Science and Education, and Grant P1-1B2015-40 from University Jaume I.

* Corresponding author.
E-mail addresses: jmonsalv@uji.es (J.A. González), cortesf@uji.es (F.J. Rodríguez-Cortés), ottmar.cronie@umu.se (O. Cronie), mateu@uji.es (J. Mateu).

http://dx.doi.org/10.1016/j.spasta.2016.10.002
2211-6753/

Contents

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) and 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. Diggle et al., 1995; Daley and Vere-Jones, 2003; Schoenberg et al., 2006; Vere-Jones, 2009; Greenspan, 2013). 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 (Møller and Sørensen, 1994; van Lieshout, 2000; Berthelsen and Møller, 2002; Møller and Waagepetersen, 2004; Daley and Vere-Jones, 2008).
   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 $T_1, \ldots, T_k \in T$ 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, Møller and Díaz-Avalos, 2010; Clements et al., 2011; Karpman et al., 2013; Pereira et al., 2013; Altieri et al., 2015 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) give 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; Karr, 1991; Snyder and Miller, 1991; Daley and Vere-Jones, 2003; Vere-Jones, 2009; Diggle, 2013 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. Ogata, 1988; Ogata, 1998, Daley and Vere-Jones, 2003; Ogata and Zhuang, 2006), which have become somewhat the main tool for the analysis of earthquakes (Marsan and Lengliné, 2008; Adelfio and Ogata, 2010; Marsan and Lengliné, 2010; Mohler et al., 2011; Adelfio and Chiodi, 2015); 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 (Musmeci and Vere-Jones, 1992; Zhuang et al., 2002; Peng et al., 2005; Adelfio and Chiodi, 2010; van Lieshout and Stein, 2012). Also, regarding mechanistic models in general, spatio-temporal separability is a key issue (see Ogata, 1988; Schoenberg, 2003; Schoenberg, 2004; Assunção and Maia, 2007; Chang and Schoenberg, 2011; Díaz-Avalos et al., 2014). 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 collection 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; cf. 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 (2015). 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 2 with a description of our datasets and in Section 3 we present a set-up of spatio-temporal point processes. Section 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 5, and mechanistic models in Section 6. The paper ends with an overall discussion and depicts some existing and ongoing research lines.

2. Datasets

Below we describe the three datasets 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.

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.

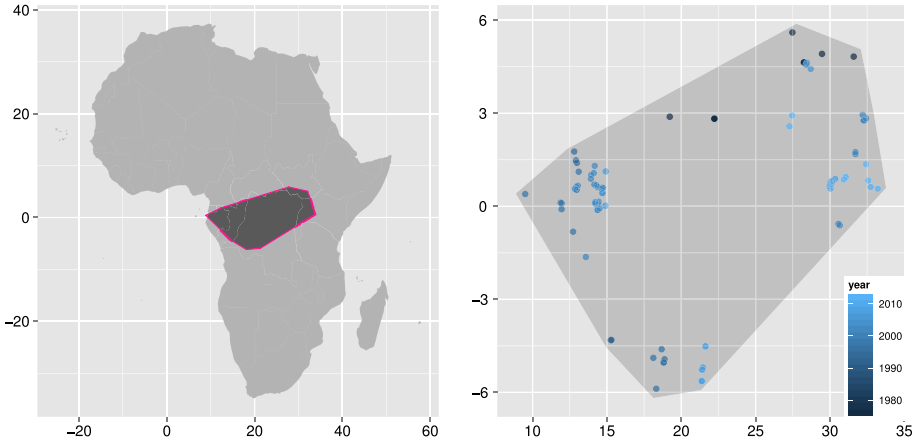

Fig. 1. Locations of 96 past Ebola outbreaks occurred in a large region of Africa during the years 1976–2012, the region of interest is formed by the smallest polygon containing all locations (left panel), here time is treated as a quantitative mark; dark dots correspond to the oldest events and light dots correspond to the most recent outbreaks (right panel).

Three outliers were removed in order to make easier the interpretation of results. The data locations are depicted in Fig. 1.

2.2. *Euphausia glacialis*

Euphausia glacialis is a type of Antarctic krill of the family *Euphausiidae* and a member of the species *Euphausia superba* (WoRMS, 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 10 000–30 000 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 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 57 006 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 Fig. 2.

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 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 Fig. 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

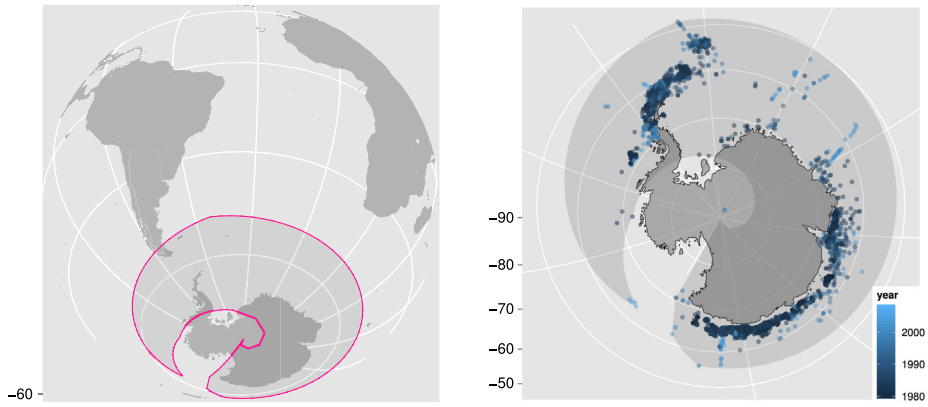

Fig. 2. Left panel: sampling area of swarms of Euphausia glacialis in the Antarctic marine environment. Right panel: orthogonal projection of locations corresponding to a sample of 7263 swarms enclosed by the smallest polygon containing all the events taken between 1980 and 2008; the darker points correspond to older records.

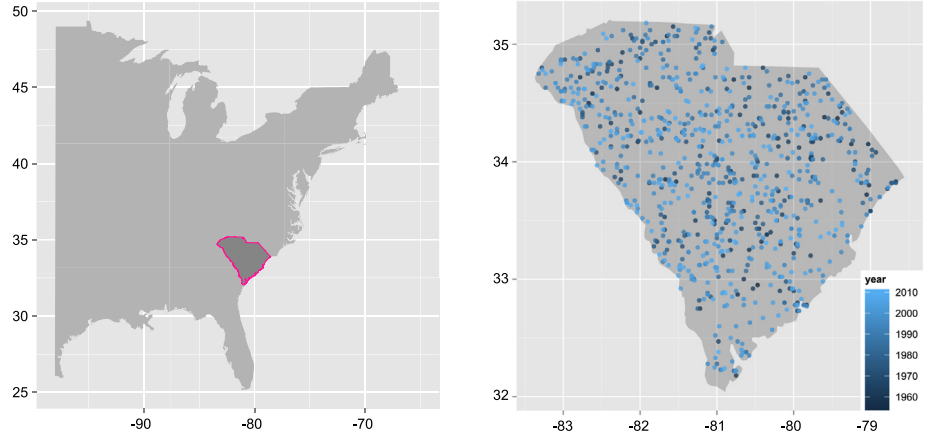

Fig. 3. Left panel: South Carolina state within US Right panel: starting locations of tornado occurrences reports over South Carolina between 1953–2012; the darker points correspond to older records.

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 of a purely spatial point pattern/process with locations $\mathbf{u}_i$ (Daley and Vere-Jones, 2008; 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

$$B[(\mathbf{u}, v), r, t] = B[\mathbf{u}, r] \times [v - t, v + t] = \{(\mathbf{a}, b) \in W \times T : \|\mathbf{u} - \mathbf{a}\| \leq r, |v - b| \leq t\}, \quad (1)$$

where $B[\mathbf{u}, r] = \{\mathbf{a} \in W : \|\mathbf{u} - \mathbf{a}\| \leq r\}$ 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 (Møller and

Ghorbani, 2012; Cronie and van Lieshout, 2015), 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 $\mathbf{r}X = \{(\mathbf{ru}, v) : (\mathbf{u}, v) \in X\}$ 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

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

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 $\mathbf{r}X_{\text{space}}$.

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.

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, 2008, p. 268), which states that, given a spatio-temporal point process X, for any non-negative function h on $(\mathbb{R}^2 \times \mathbb{R})^k$,

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

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

$$\begin{aligned}\mathbb{P}(N(\mathrm{d}\xi_1) = 1, \ldots, N(\mathrm{d}\xi_k) = 1) &= \mathbb{P}(X \cap \mathrm{d}\xi_1 \neq \emptyset, \ldots, X \cap \mathrm{d}\xi_k \neq \emptyset) \\ &= \lambda^{(k)}(\xi_1, \ldots, \xi_k) \prod_{i=1}^{k} \mathrm{d}\xi_i,\end{aligned}$$

for infinitesimal spatio-temporal regions $\mathrm{d}\xi_1, \ldots, \mathrm{d}\xi_k \subseteq W \times T$ with $\mathrm{d}\xi_i = \mathrm{d}\mathbf{u}_i \times \mathrm{d}v_i$ and size $|\mathrm{d}\xi_i| = \mathrm{d}\mathbf{u}_i\mathrm{d}v_i$, $i = 1, \ldots, k$. 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 $W \times T$.

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, 2005; Gabriel and Diggle, 2009), given by

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

4.2.1. First-order spatio-temporal separability

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

$$\lambda(\mathbf{u}, v) = \lambda_1(\mathbf{u})\lambda_2(v), \tag{2}$$

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 and Diggle, 2009; Møller and Ghorbani, 2012; Gabriel, 2014). 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

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

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.

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

$$\hat{\lambda}\left(\mathbf{u}, v\right) = \frac{1}{n}\left(\hat{\lambda}_{\text{space}}\left(\mathbf{u}\right)\hat{\lambda}_{\text{time}}\left(v\right)\right). \tag{3}$$

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,

$$\hat{\lambda}_{\text{space}}\left(\mathbf{u}\right) = \sum_{i=1}^{n} \frac{\kappa_{\epsilon}\left(\mathbf{u} - \mathbf{u}_i\right)}{c_{W\epsilon}\left(\mathbf{u}_i\right)}, \quad \mathbf{u} \in W. \tag{4}$$

Here

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

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

$$c_{W\epsilon}\left(\mathbf{u}_i\right) = \int_{W} \kappa_{\epsilon}\left(\mathbf{u} - \mathbf{u}_i\right) \mathrm{d}\mathbf{u}$$

is an edge-correction factor included in the estimation to guarantee that $\int_W \hat{\lambda}_{\text{space}}\left(\mathbf{u}\right) \mathrm{d}\mathbf{u} = n$ (Diggle, 1985; van Lieshout, 2011; Ghorbani, 2013). Similarly, we may also estimate $\lambda_{\text{time}}\left(v\right)$ 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) for the estimation of $\lambda\left(\mathbf{u}, v\right)$.

In the literature one finds that it is common to model both the spatial and the temporal intensity components by using either kernel estimators or parametric methods, which in the latter case gives rise to a semi-parametric estimation approach for $\lambda\left(\mathbf{u}, v\right)$. For example, Gabriel and Diggle (2009) estimate the temporal 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).

4.3.1. 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 estimation of the spatial and temporal intensities of the Ebola outbreaks described in Section 2.1 by means of kernel smoothing (Diggle, 2013 and references therein). A two-dimensional kernel intensity estimator, with bandwidth $\epsilon = (2.64, 0.45)$ 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 Gaussian 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). Fig. 4 illustrates the estimates of the spatial (left)

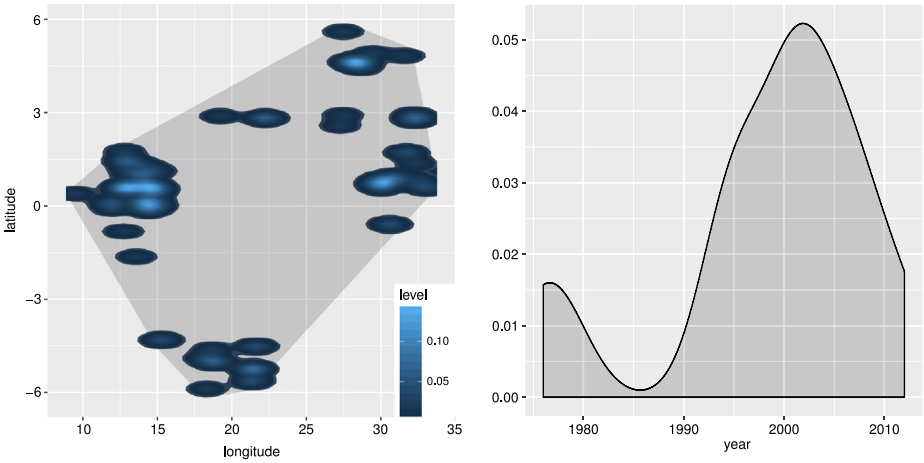

Fig. 4. Spatial (left) and temporal (right) kernel-based estimates of the normalised intensity functions for the Ebola outbreaks data.

and temporal (right) densities, understood as the intensities normalised by the total number of cases. We can see in Fig. 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 Fig. 4 (right) we note high rates of infection around the year 2000 indicating a clear degree of temporal inhomogeneity.

For the estimation of the intensity functions of the Euphausia glacialis dataset described in Section 2.2, we again use a two-dimensional kernel with an axis-aligned bivariate Gaussian kernel for the spatial intensity and a bandwidth of $\epsilon = (13.18, 0.43)$ 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 Fig. 5. In Fig. 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 Fig. 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 2.3, in which we set a spatial bandwidth of $\epsilon = (0.24, 0.16)$ degrees and a temporal bandwidth of $\delta = 3.70$ years, are depicted in Fig. 6. In this case we note that there is small variation throughout the entire spatial 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.

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* $\lambda^*(\mathbf{u}, v|\mathcal{H}_v)$ 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* $\mathcal{H}_v$, $v \in T$, consisting of the set of locations and times of all events of the process that occur prior to time

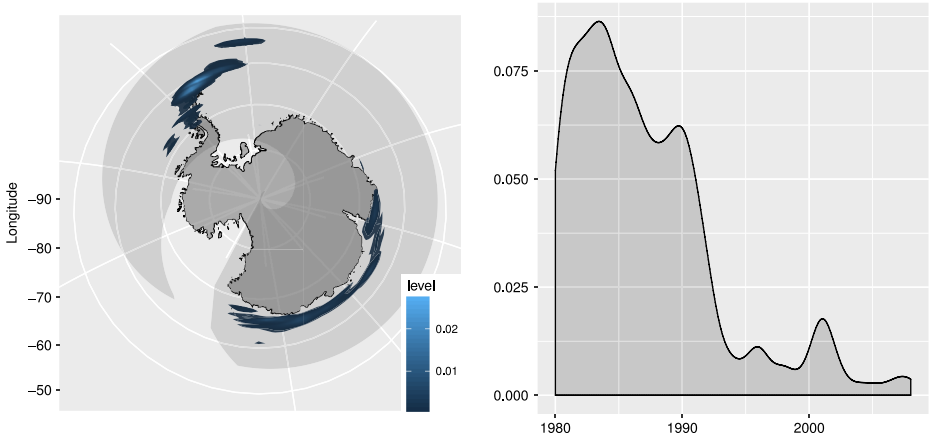

Fig. 5. Spatial (left) and temporal (right) kernel-based estimation of the normalised intensity functions for the Euphausia glacialis data.

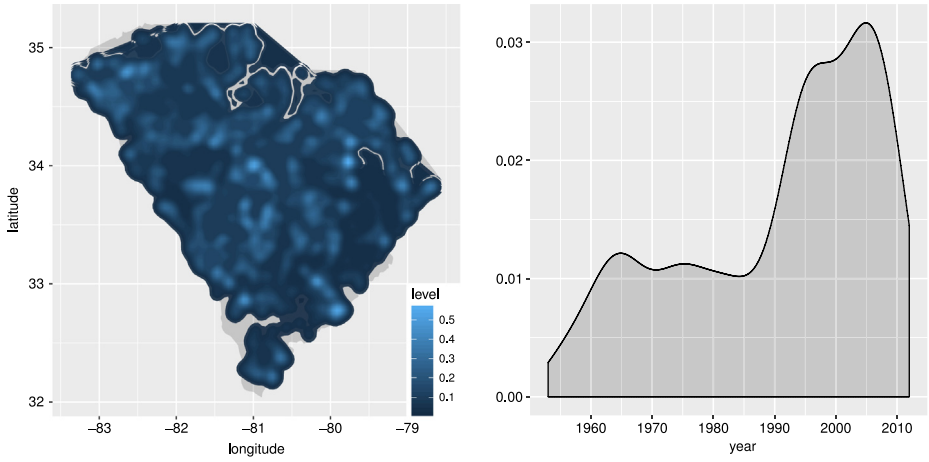

Fig. 6. Spatial (left) and temporal (right) kernel-based estimation of normalised intensity functions for the tornado in South-Carolina data.

v. In other words, $\mathcal{H}_v$ 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 Møller et al. (2016), we have that

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

We are assuming that the underlying process X is *orderly* in the sense of Diggle et al. (2010a,b), 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 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

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

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

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,

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

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 $\mathrm{d}\mathbf{u} \times \mathrm{d}v$.

4.6. Edge-correction

Edge-correction methods/factors have been widely studied in the spatial case (see e.g. Ripley, 1988; Stoyan and Stoyan, 1994; Baddeley et al., 2000 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, 2005), 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.

4.6.1. 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 (Diggle, 2005).

4.6.2. 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

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

4.6.3. Modified border method

As an extension of the method proposed by Baddeley and Turner (2000) and 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 $|W \times T|$.

4.6.4. 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 $v_i - v_j$.

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* (Illian et al., 2008; Gabriel and Diggle, 2009; Møller and Ghorbani, 2012) is defined as

$$g(\xi_1, \xi_2) = \frac{\lambda^{(2)}(\xi_1, \xi_2)}{\lambda(\xi_1)\lambda(\xi_2)}, \quad \xi_1, \xi_2 \in W \times T. \tag{5}$$

For a spatio-temporal Poisson process (i.e. a completely random process, see Sections 5.1 and 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 (Eq. (2)), the pair correlation function is said to be *separable* (Møller and Ghorbani, 2012) if

$$g((\mathbf{u}, v), (\mathbf{s}, l)) = g_1(\mathbf{u}, \mathbf{s})g_2(v, l), \tag{6}$$

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

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

4.7.1. Estimation of the pair correlation function

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

Given $X = \{(\mathbf{u}_i, v_i)\}_{i=1}^n$, the spatio-temporal pair correlation function defined in (5) can be estimated by

$$\hat{g}(r,t) = \frac{1}{4\pi r} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{\kappa_{1\epsilon}(\|\mathbf{u}_i - \mathbf{u}_j\| - r)\kappa_{2\delta}(|v_i - v_j| - t)}{\hat{\lambda}(\mathbf{u}_i, v_i)\,\hat{\lambda}(\mathbf{u}_j, v_j)\, w_{ij}}, \quad r > \epsilon, t > \delta, \tag{7}$$

where $\kappa_{1\epsilon}$ 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 et al. (2013), Gabriel (2014) and Rodríguez-Cortés et al. (2014). As in Section 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.

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

$$K_{\text{inhom}}(r,t) = \int_{\mathbb{R}^2} \int_{\mathbb{R}} \mathbf{1}\{(\mathbf{u}, v) \in B_{rt}\}\, \bar{g}(\mathbf{u}, v)\mathrm{d}\mathbf{u}\mathrm{d}v, \tag{8}$$

for $r > 0$ and $t > 0$. Note that in (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),

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

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 5.1 and 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

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

where

$$g_{\text{space}}(\mathbf{u}, \mathbf{s}) = g_{\text{space}}(\mathbf{u} - \mathbf{s}) = \int_T \int_T p_2(v) p_2(l) g(\mathbf{u} - \mathbf{s}, v - l)\mathrm{d}v\mathrm{d}l,$$

$$g_{\text{time}}(v, l) = g_{\text{time}}(v - l) = \int_W \int_W p_1(\mathbf{u}) p_1(\mathbf{s}) g(\mathbf{u} - \mathbf{s}, v - l)\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{s},$$

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, 2008); $\mathbb{P}^{!(\mathbf{u},v)}(X \in \cdot)$ 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 $K(r, t) = K_{\text{inhom}}(r, t)$, 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.

$$g(r, t) = \frac{1}{4\pi r} \frac{\partial^2 K(r, t)}{\partial r \partial t}, \quad r > 0,\ t > 0. \tag{9}$$

4.8.1. 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

$$\hat{K}(r, t) = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{\mathbf{1}\left[\left\|\mathbf{u}_i - \mathbf{u}_j\right\| \leq r\right] \mathbf{1}\left[\left|v_i - v_j\right| \leq t\right]}{\hat{\lambda}\left(\mathbf{u}_i, v_i\right) \hat{\lambda}\left(\mathbf{u}_j, v_j\right) w_{ij}}. \tag{10}$$

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

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

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

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.

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.

$$J_{\text{inhom}}(r,t) = \frac{1-G_{\text{inhom}}(r,t)}{1-F_{\text{inhom}}(r,t)}, \tag{11}$$

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

$$J_{\text{inhom}}(r,t) - 1 = \underline{\lambda}(2\pi r^2 t - K_{\text{inhom}}(r,t)) + \beta, \tag{12}$$

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.

4.10.1. 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, $\Psi_X = X\cap(W_{\ominus r}\times T_{\ominus t})$, $\Psi_L = L\cap(W_{\ominus r}\times T_{\ominus t})$,

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

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

$$\frac{1}{|\Psi_X|}\sum_{(\mathbf{s},l)\in\Psi_X}\prod_{(\mathbf{u},v)\in\Omega_-^{(\mathbf{s},l)}}\left(1-\frac{\hat{\underline{\lambda}}}{\hat{\lambda}(\mathbf{u},v)}\right), \tag{13}$$

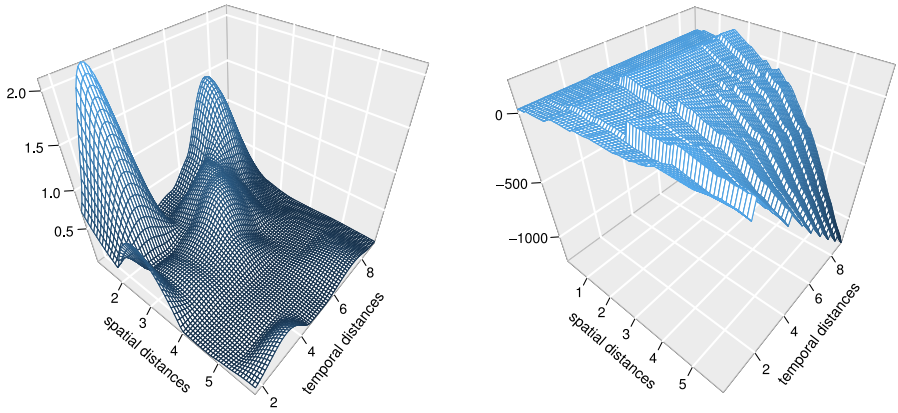

Fig. 7. Spatio-temporal summary statistics $\hat{g}(r,t)$ (left) and $\hat{K}(r,t) - 2\pi r^2 t$ (right) for the Ebola outbreaks data.

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

$$\frac{1}{|\Psi_L|} \sum_{(\mathbf{s},l)\in\Psi_L} \prod_{(\mathbf{u},v)\in\Omega_+^{(\mathbf{s},l)}} \left(1 - \frac{\hat{\underline{\lambda}}}{\hat{\lambda}(\mathbf{u},v)}\right), \tag{14}$$

where $\hat{\underline{\lambda}} = \min_{(\mathbf{u},v)} \hat{\lambda}(\mathbf{u},v)$. The ratio of the estimators given in (13) and (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).

4.11. Estimation of summary statistics for the considered examples

We keep here our assumption of first-order separability, that is, Eq. (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. (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 4.3.1). 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).

Fig. 7 illustrates the estimates of g (left) and K (right) for the Ebola outbreaks pattern described in Section 2.1. The K-function is illustrated by subtracting the theoretical surface under an inhomogeneous Poisson process with the same intensity function (see Sections 5.1 and 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 $\epsilon = 0.99^\circ$ 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 (Section 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 2.2 are displayed in Fig. 8. We use a spatial Epanechnikov kernel and a temporal biweight kernel for the pair correlation estimation, where the bandwidths are $\epsilon = 0.01^\circ$ and $\delta = 1.2$

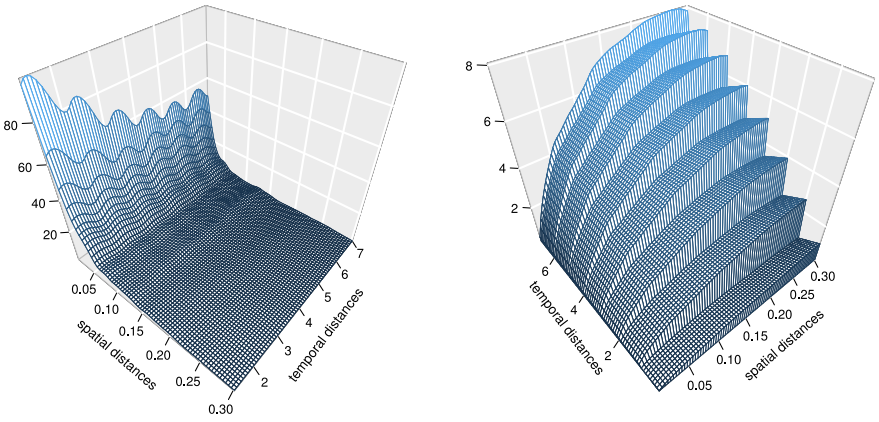

Fig. 8. $\hat{g}(r,t)$ (left) and $\hat{K}(r,t) - 2\pi r^2 t$ (right) summary statistics for the Euphausia glacialis data.

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 Fig. 2. However, we perform this analysis as a first descriptive approach to understand the nature of the pattern interactions. In Section 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 2.3 are depicted in Fig. 9. We use a spatial Epanechnikov kernel and a temporal biweight kernel with bandwidths $\epsilon = 0.12°$ 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 Fig. 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 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 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).

Fig. 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 5.1 and 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.

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 $r > 0, 0 \leq \phi \leq \pi$,

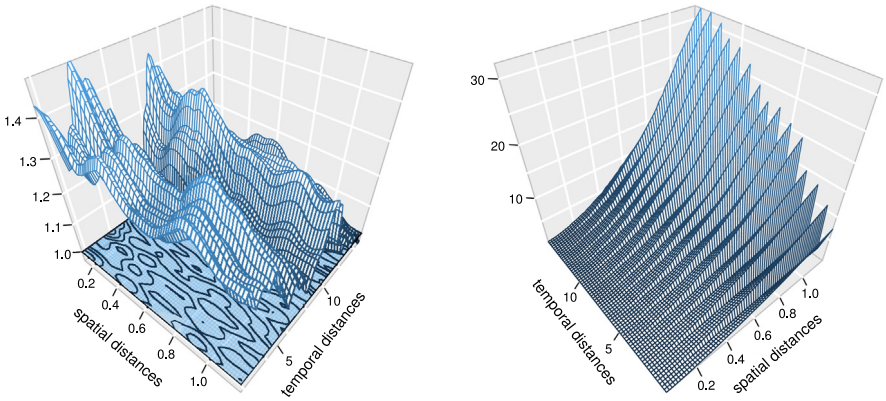

Fig. 9. $\hat{g}(r,t)$ (left) and $\hat{K}(r,t) - 2\pi r^2 t$ (right) summary statistics for tornadoes in South-Carolina data. The plane $g = 1$ is shown with the contours.

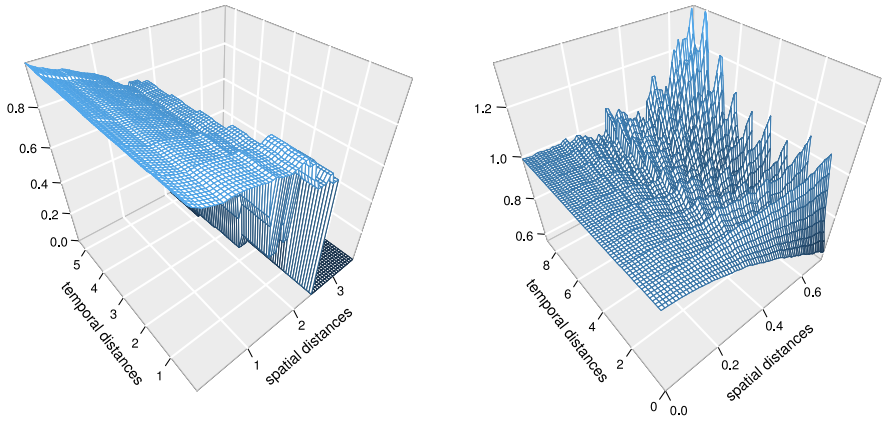

Fig. 10. $\hat{J}(r,t)$ function estimated for the Ebola outbreaks (left) data and tornado data (right).

$t > 0$ 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 4.3) and considers an edge-correction according to the nature of the underlying process. The estimator is given by

$$\hat{K}_\varphi(r,t,\phi) = \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} \frac{\mathbf{1}\{\|\mathbf{u}_i - \mathbf{u}_j\| \le r, |\theta(\mathbf{u}_i,\mathbf{u}_j)| \le \phi, |v_i - v_j| \le t\}}{\hat{\lambda}(\mathbf{u}_i, v_i)\,\hat{\lambda}(\mathbf{u}_j, v_j)\, w_{ij}}, \tag{15}$$

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

$$\vartheta_{(r_1,t_1),(r_2,t_2)}(\phi) = \frac{\int_0^{\phi}\int_{t_1}^{t_2}\int_{r_1}^{r_2} \mathrm{d}K_{\varphi}(r,t,\psi)}{\int_0^{\pi}\int_{t_1}^{t_2}\int_{r_1}^{r_2} \mathrm{d}K_{\varphi}(r,t,\psi)}, \quad r_2 > r_1 \geq 0,\ t_2 > t_1 \geq 0. \tag{16}$$

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 (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),

$$\vartheta^*_{(r_1,t_1),(r_2,t_2)}(\phi) = \vartheta_{(r_1,t_1),(r_2,t_2)}(\phi+\alpha) - \vartheta_{(r_1,t_1),(r_2,t_2)}(\phi-\alpha), \quad 0 \leq \phi \leq \pi, \tag{17}$$

where $0 < \alpha < \phi$ is a fixed prescribed angle interval, which provides the direction in which anisotropic effects are tested. Using (15), we obtain estimators of (16) and (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. Ohser and Stoyan, 1981; Baddeley et al., 2015) depicted in Fig. 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 Fig. 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 4.3.1.

5. Spatio-temporal empirical models

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 $A \times B$.

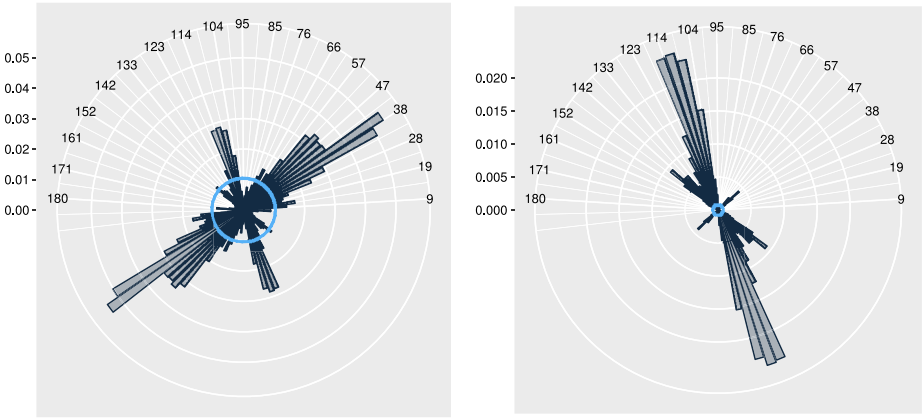

Fig. 11. Anticlockwise rose histogram of the short and middle orientation distribution $\hat{\vartheta}_{(0,1),(0.1,5)}$ (left) and of the middle range orientation distribution $\hat{\vartheta}_{(0.1,1),(0.3,5)}$(right), for the sample of Euphausia glacialis displayed in the right-side of Fig. 2, and described in Section 2.2. The blue circle corresponds to the median of each range. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

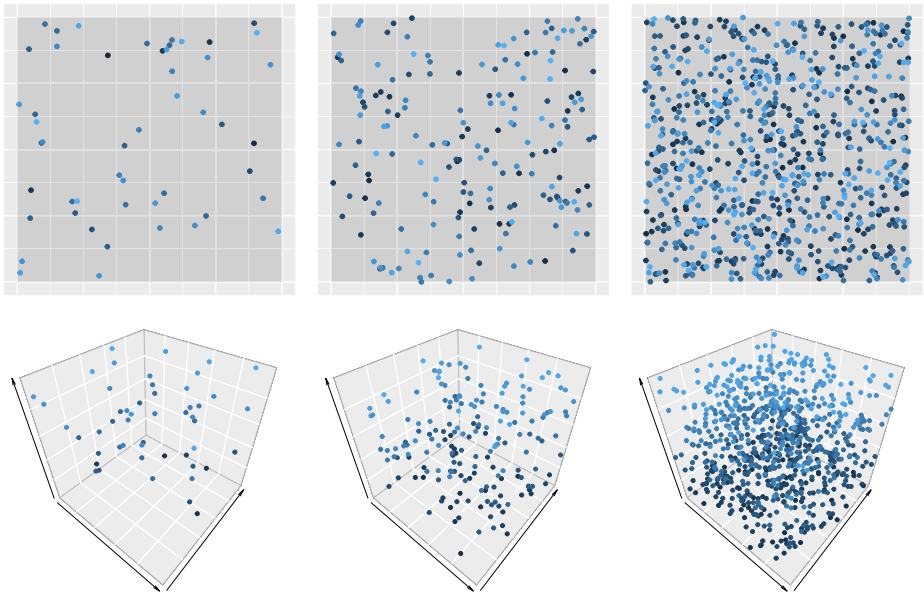

Fig. 12. Simulated realisations of a spatio-temporal homogeneous Poisson point process with $\lambda = 50$, $\lambda = 150$ and $\lambda = 800$ in left, central and right panels respectively in $W \times T = [0, 1]^2 \times [0, 1]$, the darker points correspond to older events.

It follows that all product densities exist and are given by

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

Fig. 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.

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
$$\lambda(x, y, v) = a \sin\left(\pi\sqrt{\left(\tfrac{3\pi}{2}x - 2\right)^2 + \left(\tfrac{3\pi}{2}y - 2\right)^2} - 1\right) + 2, \tag{18}$$
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 Fig. 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.

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

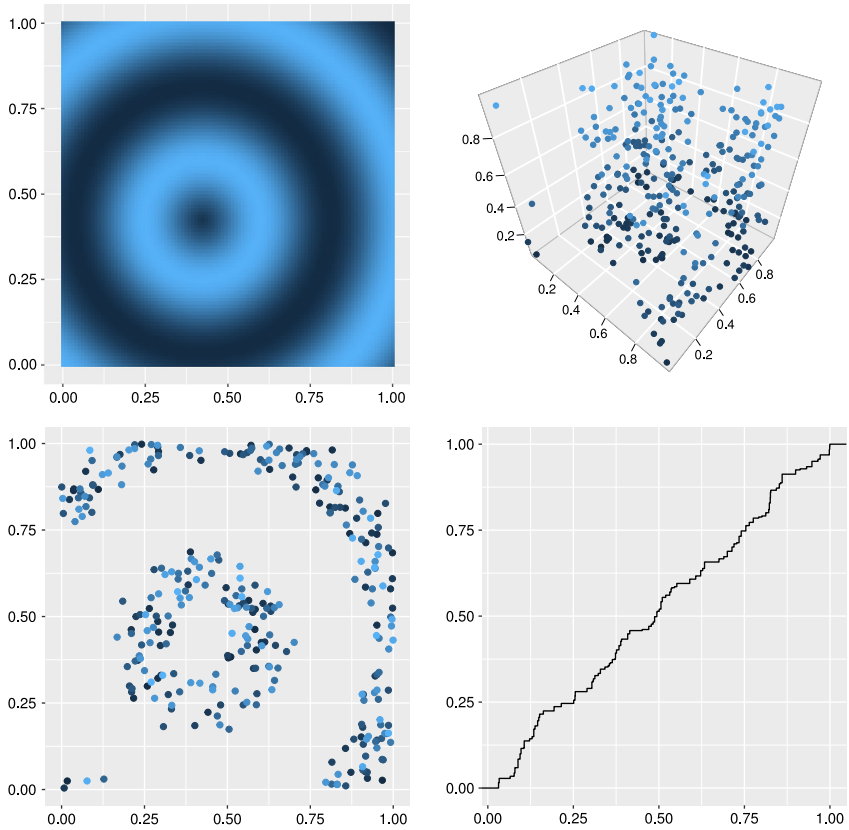

Fig. 13. Simulated realisation of a spatio-temporal inhomogeneous Poisson point process in $W \times T = [0, 1]^2 \times [0, 1]$, with λ given by (18). On the upper left-side panel the spatial intensity function is displayed, the points of the realisation on the right-side panel represented in a three-dimensional 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.

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

$$\log \lambda(\mathbf{u}, v) = \sum_{j=1}^{p} \beta_j z_j(\mathbf{u}, v), \tag{19}$$

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

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 $\lambda_p(\mathbf{u}, v)$.
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

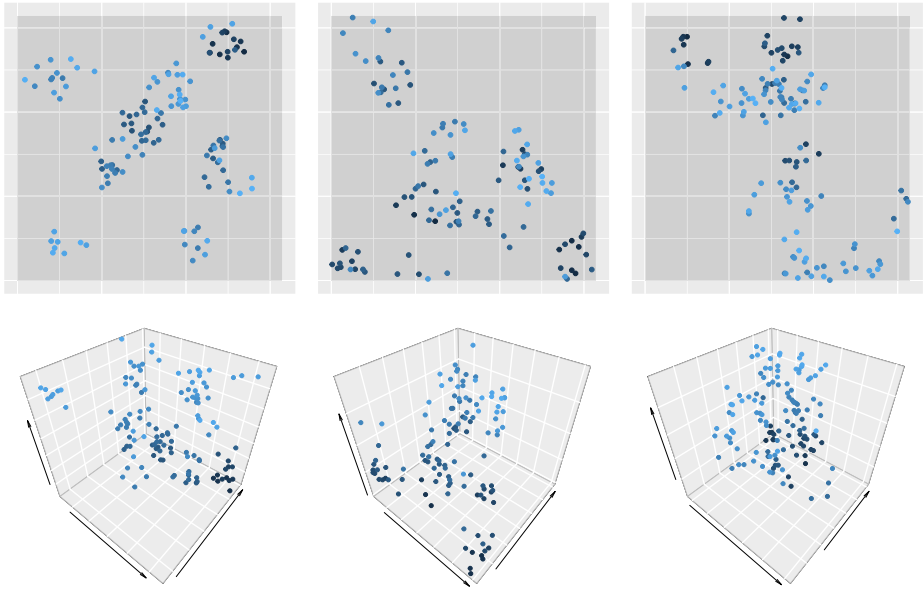

Fig. 14. Three examples of spatio-temporal Neyman–Scott cluster patterns over the rectangular region $[0, \pi]^2$ and along the unit temporal interval [0, 1], darker points correspond to older occurrences, and $m_c = 8$. The distributions of offspring are: normal (left patterns), exponential (middle patterns), and uniform (right patterns).

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 Fig. 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.

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).

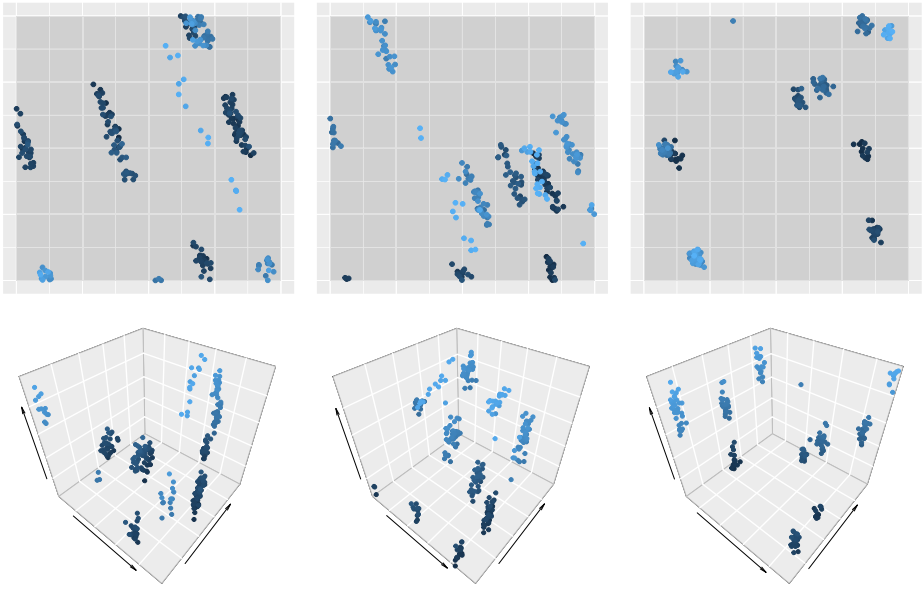

Fig. 15. Realisations of geometric anisotropic Poisson cluster processes in $W \times T = [0, 1]^2 \times [0, 1]$. The three cases correspond to $\zeta = 0.05, 0.30, 0.80$ and left, central and right patterns, respectively. Dark dots correspond to older events in time.

Fig. 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 $\zeta \in \{0.05, 0.30, 0.80\}$.

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

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

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

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

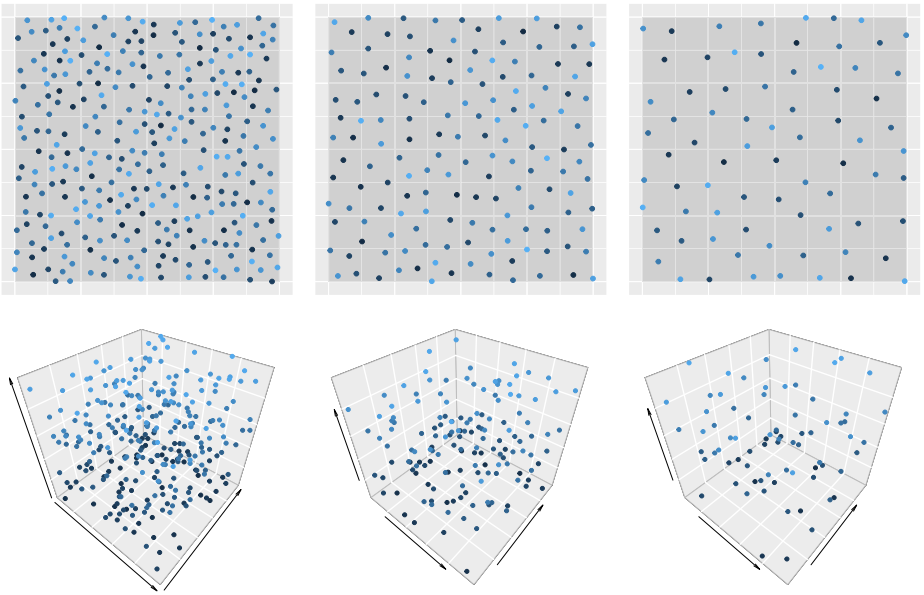

Fig. 16. Simulated realisations of simple sequential spatio-temporal inhibition point processes with $(\delta_{\mathbf{u}}, \delta_v, n) = (0.04, 0.001, 300)$ (left), $(\delta_{\mathbf{u}}, \delta_v, n) = (0.07, 0.005, 150)$ (centre) and $(\delta_{\mathbf{u}}, \delta_v, n) = (0.10, 0.001, 75)$ (right) in $W \times T = [0, 1]^2 \times [0, 1]$. The darker points correspond to older locations.

To obtain a larger class of inhibition processes, it is possible to extend condition (ii) of the above algorithm definition by introducing functions $p_{\mathbf{u}}(\mathbf{s})$ 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. Fig. 16 shows examples of three different single realisations of simple sequential inhibition point processes.

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

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

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.

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; 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

$$\Lambda(\mathbf{u}, v) = \lambda(\mathbf{u}, v)S(\mathbf{u}, v), \tag{20}$$

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 (20), we can say that $\Lambda(\mathbf{u}, v)$ is *first-order separable* if (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 Møller and Díaz-Avalos (2010) and Diggle (2013), 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 (20), is given by

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

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 (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

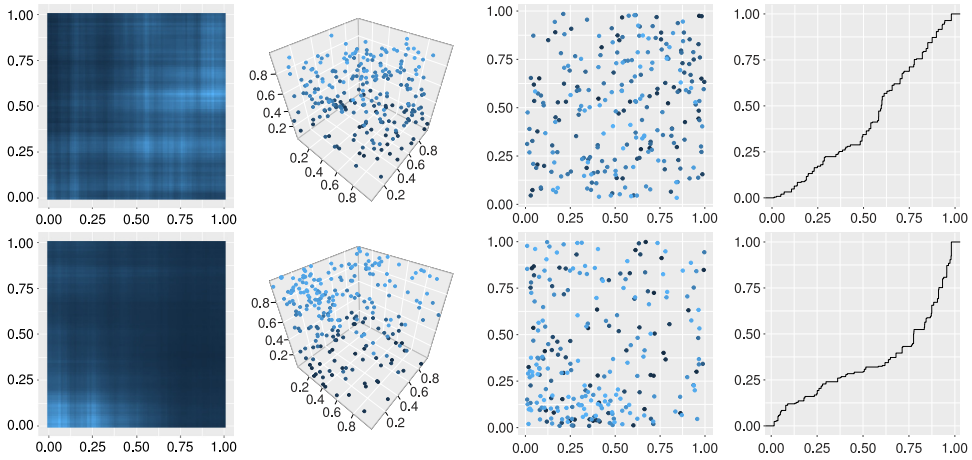

Fig. 17. Spatial intensities (left), realisations (two central panels), and cumulative distributions of times (right) of log-Gaussian Cox processes. Non-separable (top) and the separable (bottom) cases.

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 Fig. 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 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 $X = \{(\mathbf{u}_i, v_i) \in W \times T : i = 1, \ldots, n\}$ is given by (Diggle, 2013; Diggle et al., 2013)

$$\ell(\theta; X) = \mathbb{E}_{\Lambda|\theta}(\ell^*(\Lambda; X)), \tag{21}$$

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 (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

$$[\mathcal{T}|Z] = \int_{\Theta} [\mathcal{T}|Z;\theta][\theta|Z]\mathrm{d}\theta, \tag{22}$$

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 (22) incorporates parameter uncertainty in a way that is both natural and elegant.

5.9. Spatio-temporal stationary Poisson cluster and shot-noise Cox processes

These cluster processes are built as follows. The spatial distribution of the offspring 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 5.7. Here, the residual process $S(\cdot)$ in Eq. (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

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

where $\phi_{\sigma^2}(\|\mathbf{u}\|)$ is the density of a zero-mean bivariate isotropic normal distribution with variance σ^2 and $\mathcal{E}_{\alpha}$ 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 $\mathcal{L}$ 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

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

where κ is a kernel function (weight) and $\mathcal{L}$ is a non-negative Lévy basis (see Hellmund et al., 2008; Beneš et al., 2015), 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.

6. Spatio-temporal mechanistic models

6.1. Poisson processes

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

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

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

We recall from Section 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)) is directly followed (see Choi and Hall, 1999).

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.

6.2.1. 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 $m = 0, 1, \ldots$ with the following branching structure. We first have $(\mathbf{u}_i, v_i)$, which is said to be of generation 0. Given the $0, \ldots, m$ 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.

6.2.2. 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, Marsan and Lengliné, 2008, 2010; Adelfio and Ogata, 2010; Mohler et al., 2011; Adelfio and Chiodi, 2015 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 (Musmeci and Vere-Jones, 1992; Zhuang et al., 2002; Peng et al., 2005; Adelfio and Chiodi, 2010; van Lieshout and Stein, 2012).

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 $\lambda^*(\mathbf{u}, v, M|\mathcal{H}_v) = j(M)\lambda^*(\mathbf{u}, v|\mathcal{H}_v)$, where

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

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.

6.3. Likelihood inference

We assume that the process is orderly (see Diggle et al., 2010a,b 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

$$L(\lambda^*) = \sum_{i=1}^{n} \log \lambda^*\left(\mathbf{u}_i, v_i|\mathcal{H}_{v_i}\right) - \int_T \int_W \lambda^*(\mathbf{u}, v|\mathcal{H}_v)\, \mathrm{d}\mathbf{u}\mathrm{d}v. \tag{23}$$

Hence, likelihood-based inference is quite straightforward for any model for which the conditional intensity is specified. It is only required that $\lambda^*(\mathbf{u}, v|\mathcal{H}_v)$ 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 $\lambda^*(\mathbf{u}, v|\mathcal{H}_v)$, and this may be difficult, even impossible. Secondly, the integrand of the integral term on the right-hand side of Eq. (23) is often a complicated function with many local modes, so the accurate evaluation of the integral becomes very complicated.

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. (23) called *partial likelihood*.

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

$$p_i = \frac{\lambda^*\left(\mathbf{u}_i, v_i|\mathcal{H}_{v_i}\right)}{\int_W \lambda^*\left(\mathbf{x}, v_i|\mathcal{H}_{v_i}\right) d\mathbf{x}}, \tag{24}$$

and the partial log-likelihood is given by

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

Usually, the integral in Eq. (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 $j = 1, \ldots, N$ for some $N \geq n$, $L_p(\lambda^*)$ becomes (see Møller and Sørensen, 1994; Diggle, 2006)

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

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

6.5. Separability of conditional intensities

Separability in the context of mechanistic models assumes, similarly to the first-order intensity case (Eq. (2)), a multiplicative form for the conditional intensity function

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

where $\lambda_1^*(\cdot|\mathcal{H}_t)$ and $\lambda_2^*(\cdot|\mathcal{H}_t)$ 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. Schoenberg (2004), Assunção and Maia (2007), Chang and Schoenberg (2011) 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 $\hat{\lambda}^*_{\mathrm{NS}}(\mathbf{u}, v|\mathcal{H}_t)$, and its separable counterpart estimator $\hat{\lambda}^*_{\mathrm{S}}(\mathbf{u}, v|\mathcal{H}_t)$ 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

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

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

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

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.

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 $j = 1, 2, \ldots, m$). The quantile score, γ_L, is defined as the fraction of simulated likelihoods that are less than the observed catalogue likelihood

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

If γ_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 γ_N is defined in a similar way, and the model is rejected if γ_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,

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

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

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

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

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

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) established, the distribution of raw and Pearson residuals tends 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.

8. Discussion and ongoing research

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. Ho and Stoyan, 2008; Diggle et al., 2010c).

For completeness, we do not want to finish this discussion without mentioning some ongoing research that is performed in parallel to the writing of this review paper. In this last section of the

paper we want to both indicate incoming papers in the field and indicate current trends that will help researchers in identifying new interesting areas for future developments.

A recent unpublished paper by Ghorbani et al. (unpublished) presents an extension of the non-parametric edge-corrected Ohser-type kernel estimator of the spatio-temporal second-order product density function. They derive some properties and perform a simulation study to compare the approximation of the variance with Monte-Carlo outputs and they apply the resulting estimator to analyse the spatio-temporal distribution of the invasive meningococcal disease in the Rhineland Regional Council in Germany.

Since little research has been conducted for replicated point patterns, even though they are becoming quite common, we find some analysis of variance in the spatial case (see e.g. Diggle et al., 1991; Baddeley et al., 1993; Hahn, 2012). There is also forthcoming research related to replicated point patterns and comparisons in the spatio-temporal context; González et al. (unpublished) provide some techniques for the analysis of spatio-temporal complex datasets in the presence of approximate replication. Classical inference for spatio-temporal point patterns is difficult so they propose a bootstrap procedure and a permutation test for attaching p-values to observed group differences between the pattern descriptor functions, in a method which has analogies to classical ANOVA. The validity of hypothesis testing procedures is demonstrated through simulation experiments under the null and several different alternative hypotheses. This approach is used in the analysis of tornado locations in the US along some temporal intervals and irregular pixels.

Stoyan et al. (unpublished) consider completely stationary spatio-temporal point processes transformed into two stationary marked point processes by taking the times or locations as marks. For these marked point processes, *mark variograms* are defined. These variograms give important information about the spatio-temporal behaviour of the original process, and the authors provide theoretical forms for such mark variograms, for some particular spatio-temporal point process models.

Finally, following the initial work on determinantal point process introduced by Lavancier et al. (2015), Ghorbani (unpublished) extends this methodology to the spatio-temporal scenario.

Acknowledgements

The authors are grateful to all professionals who have provided information and help. In particular they are grateful to the reviewers and the Associate Editor who have provided substantial amounts of constructive comments and feedback on earlier versions of the paper.

References

Adelfio, G., Chiodi, M., 2010. Diagnostics for nonparametric estimation in space–time seismic processes. J. Environ. Stat. 1 (2), 1–13.
Adelfio, G., Chiodi, M., 2015. Alternated estimation in semi-parametric space–time branching-type point processes with application to seismic catalogs. Stoch. Environ. Res. Risk Assess. 29 (2), 443–450.
Adelfio, G., Ogata, Y., 2010. Hybrid kernel estimates of space–time earthquake occurrence rates using the epidemic-type aftershock sequence model. Ann. Inst. Statist. Math. 62 (1), 127–143.
Adelfio, G., Schoenberg, F.P., 2009. Point process diagnostics based on weighted second-order statistics and their asymptotic properties. Ann. Inst. Statist. Math. 61 (4), 929–948.
Altieri, L., Scott, E., Cocchi, D., Illian, J., 2015. A changepoint analysis of spatio-temporal point processes. Spat. Stat. 14 (Part B), 197–207.
Assunção, R., Maia, A., 2007. A note on testing separability in spatial–temporal marked point processes. Biometrics 63 (1), 290–294.
Baddeley, A., Møller, J., Waagepetersen, R., 2000. Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. Stat. Neerl. 54, 329–350.
Baddeley, A., Moyeed, R., Howard, C., Boyde, A., 1993. Analysis of a three-dimensional point pattern with replication. J. Roy. Statist. Soc. Ser. C 42 (4), 641–668.
Baddeley, A., Rubak, E., Turner, R., 2015. Spatial Point Patterns: Methodology and Applications with R. In: Chapman & Hall Interdisciplinary Statistics Series, CRC Press, Boca Raton, Florida.
Baddeley, A., Turner, R., 2000. Practical maximum pseudolikelihood for spatial point patterns. Aust. N. Z. J. Stat. 42 (3), 283–322.
Baddeley, A., Turner, R., Møller, J., Hazelton, M., 2005. Residual analysis for spatial point processes. J. R. Stat. Soc. Ser. B Stat. Methodol. 67 (5), 617–666.
Banerjee, S., Carlin, B.P., Gelfand, A.E., 2014. Hierarchical Modeling and Analysis for Spatial Data, second ed. In: Chapman & Hall Monographs on Statistics & Applied Probability, CRC Press, Boca Raton, Florida.
Beneš, V., Prokešová, M., Helisová, K.S., Zikmundová, M., 2015. Space-Time Models in Stochastic Geometry. Springer International Publishing, pp. 205–232.

Berman, M., Diggle, P.J., 1989. Estimating weighted integrals of the second-order intensity of a spatial point process. J. R. Stat. Soc. Ser. B Stat. Methodol. 51 (1), 81–92.
Berthelsen, K., Møller, J., 2002. Spatial jump processes and perfect simulation. In: Mecke, K.R., Stoyan, D. (Eds.), Morphology of Condensed Matter: Physics and Geometry of Spatially Complex Systems. Spriger-Verlag, Berlin, pp. 391–417.
Brix, A., Diggle, P.J., 2001. Spatiotemporal prediction for log-Gaussian Cox processes. J. R. Stat. Soc. Ser. B Stat. Methodol. 63 (4), 823–841.
Brown, P., Karesen, K., Roberts, G., Tonellato, S., 2000. Blur-generated non-separable space–time models. J. R. Stat. Soc. Ser. B Stat. Methodol. 62 (4), 847–860.
Chakraborty, A., Gelfand, A.E., Wilson, A.M., Latimer, A.M., Silander, J.A., 2011. Point pattern modelling for degraded presence-only data over large regions. J. Roy. Statist. Soc. Ser. C 60 (5), 757–776.
Chang, C.-H., Schoenberg, F., 2011. Testing separability in marked multidimensional point processes with covariates. Ann. Inst. Statist. Math. 63 (6), 1103–1122.
Chiu, S.N., Stoyan, D., Kendall, W.S., Mecke, J., 2013. Stochastic Geometry and its Applications, third ed. In: Wiley Series in Probability and Statistics, John Wiley & Sons, Chichester.
Choi, E., Hall, P., 1999. Nonparametric approach to analysis of space–time data on earthquake occurrences. J. Comput. Graph. Statist. 8, 733–748.
Clements, R.A., Schoenberg, F., Schorlemmer, D., 2011. Residual analysis methods for space–time point processes with applications to earthquake forecast models in California. Ann. Appl. Stat. 5 (4), 2549–2571.
Comas, C., Rodriguez-Cortes, F.J., Mateu, J., 2015. Second-order analysis of anisotropic spatiotemporal point process data. Stat. Neerl. 69 (1), 49–66.
Cox, D.R., 1955. Some statistical methods connected with series of events. J. R. Stat. Soc. Ser. B Stat. Methodol. 17 (2), 129–164.
Cox, D.R., 1975. Partial likelihood. Biometrika 62 (2), 269–276.
Cox, D.R., Isham, V., 1980. Point Processes. Chapman & Hall. CRC Press, London.
Cressie, N., Collins, B., 2001. Analysis of spatial point patterns using bundles of product density lisa functions. J. Agric. Biol. Environ. Stat. 6 (1), 118–135.
Cressie, N., Huang, H.-C., 1999. Classes of nonseparable, spatio-temporal stationary covariance functions. J. Amer. Statist. Assoc. 94 (448), 1330–1340.
Cressie, N., Wikle, C.K., 2011. Statistics for Spatio-Temporal Data. In: Wiley Series in Probability and Statistics, John Wiley & Sons, New Jersey.
Cronie, O., Särkkä, A., 2011. Some edge correction methods for marked spatio-temporal point process models. Comput. Statist. Data Anal. 55 (7), 2209–2220.
Cronie, O., van Lieshout, M., 2015. A J-function for inhomogeneous spatio-temporal point processes. Scand. J. Statist. 42 (2), 562–579.
Daley, D., Vere-Jones, D., 2003. An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods, second ed. Springer-Verlag, New York.
Daley, D., Vere-Jones, D., 2008. An Introduction to the Theory of Point Processes. Volume II: General Theory and Structure, second ed. Springer-Verlag, New York.
Deza, M.M., Deza, E., 2009. Encyclopedia of Distances. Springer-Verlag, Berlin.
Díaz-Avalos, C., Juan, P., Mateu, J., 2014. Significance tests for covariate-dependent trends in inhomogeneous spatio-temporal point processes. Stoch. Environ. Res. Risk Assess. 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. Appl. Stat. 34, 138–147.
Diggle, P.J., 2005. Spatio-Temporal Point Processes: Methods and Applications. Johns Hopkins University, Dept. of Biostatistics, pp. 1–43.
Diggle, P.J., 2006. Spatio-temporal point processes, partial likelihood, foot and mouth disease. Stat. Methods Med. Res. 15 (4), 325–336.
Diggle, P.J., 2013. Statistical Analysis of Spatial and Spatio-Temporal Point Patterns, third ed. In: Chapman & Hall Monographs on Statistics & Applied Probability, CRC Press, Boca Raton, Florida.
Diggle, P.J., Chetwynd, A., Häggkvist, R., Morris, S., 1995. Second-order analysis of space–time clustering. Stat. Methods Med. Res. 4, 124–136.
Diggle, P.J., Guan, Y., Hart, A., Paize, F., Stanton, M., 2010a. Estimating individual-level risk in spatial epidemiology using spatially aggregated information on the population at risk. J. Amer. Statist. Assoc. 105 (492), 1394–1402.
Diggle, P.J., Kaimi, I., Abellana, R., 2010b. Partial-likelihood analysis of spatio-temporal point-process data. Biometrics 66 (2), 347–354.
Diggle, P.J., Lange, N., Beneš, F.M., 1991. Analysis of variance for replicated spatial point patterns in clinical neuroanatomy. J. Amer. Statist. Assoc. 86 (415), 618–625.
Diggle, P.J., Menezes, R., Su, T.-l., 2010c. Geostatistical inference under preferential sampling. J. Roy. Statist. Soc. Ser. C 59 (2), 191–232.
Diggle, P.J., Moraga, P., Rowlingson, B., Taylor, B., 2013. Spatial and spatio-temporal log-Gaussian Cox processes: Extending the geostatistical paradigm. Statist. Sci. 28, 542–563.
Diggle, P.J., Rowlingson, B., Su, T.-l., 2005. Point process methodology for on-line spatio-temporal disease surveillance. Environmetrics 16 (5), 423–434.
Finkenstadt, B., Held, L., Isham, V., 2007. Statistical Methods for Spatio-Temporal Systems. In: 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. Methodol. Comput. Appl. Probab. 16 (2), 411–431.
Gabriel, E., Diggle, P.J., 2009. Second-order analysis of inhomogeneous spatio-temporal point process data. Stat. Neerl. 63 (1), 43–51.
Gabriel, E., Rowlingson, B., Diggle, P.J., 2013. stpp: An R package for plotting, simulating and analyzing spatio-temporal point patterns. J. Stat. Softw. 53 (2), 1–29. 4.
Gelfand, A.E., Diggle, P.J., Guttorp, P., Fuentes, M. (Eds.), 2010. Handbook of Spatial Statistics. In: 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: Barndorff-Nielsen, O.E., Kendall, W.S., van Lieshout, M.N.M. (Eds.), Stochastic Geometry (Toulouse, 1996). In: Chapman & Hall Monographs on Statistics & Applied Probability, vol. 80. CRC Press, Boca Raton, Florida, pp. 79–140.
Ghorbani, M., 2013. Testing the weak stationarity of a spatio-temporal point process. Stoch. Environ. Res. Risk Assess. 27 (2), 517–524.
Ghorbani, M., 2016. Spatio-temporal determinantal point processes (unpublished).
Ghorbani, M., Mateu, J., Rodríguez-Cortés, F.J., 2016. Statistical properties of the spatio-temporal product density function estimator, with a special view to Poisson processes (unpublished).
Gneiting, T., 2002. Nonseparable, stationary covariance functions for space–time data. J. Amer. Statist. Assoc. 97 (458), 590–600.
Gneiting, T., Guttorp, P., 2010. Continuous parameter spatio-temporal processes. In: Gelfand, A.E., Diggle, P.J., Fuentes, M., Guttorp, P. (Eds.), Handbook of Spatial Statistics. In: Chapman & Hall Handbooks of Modern Statistical Methods, CRC Press, Boca Raton, Florida, pp. 427–436.
González, J., Hahn, U., Mateu, J., 2016. Analysis of spatio-temporal point patterns with replication (unpublished).
Greenspan, B.M., 2013. Survey of some recent advances in spatial–temporal point processes (Ph.D. thesis), University of California, Los Angeles.
Hahn, U., 2012. A studentized permutation test for the comparison of spatial point patterns. J. Amer. Statist. Assoc. 107, 754–764.
Hahn, U., Vedel Jensen, E.B., 2015. Hidden second-order stationary spatial point processes. Scand. J. Statist. 43 (2), 455–475.
Hamner, W., Hamner, P., Strand, S., Gilmer, R., 1983. Behavior of antarctic krill, euphausia superba: Chemoreception, feeding, schooling, and molting. Science 220 (4595), 433–435.
Hellmund, G., Prokešová, M., Jensen, E.B.V., 2008. Lévy-based Cox point processes. Adv. Appl. Probab. 40 (3), 603–629.
Ho, L.P., Stoyan, D., 2008. Modelling marked point patterns by intensity-marked Cox processes. Statist. Probab. Lett. 78 (10), 1194–1199.
Illian, J., Penttinen, A., Stoyan, H., Stoyan, D., 2008. Statistical Analysis and Modelling of Spatial Point Patterns. John Wiley & Sons, Chichester.
Jafari-Mamaghani, M., Andersson, M., Krieger, P., 2010. Spatial point pattern analysis of neurons using Ripley's K- function in 3D. Front. Neuroinf. 4 (9), 1–10.
Karpman, D., Ferreira, M.A., 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 ed. Marcel Dekker, New York.
Lavancier, F., Møller, J., Rubak, E., 2015. Determinantal point process models and statistical inference. J. R. Stat. Soc. Ser. B Stat. Methodol. 77 (4), 853–877.
Lund, J., 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. J. Statist. Plann. Inference (116), 489–501.
Ma, C., 2008. Recent developments on the construction of spatio-temporal covariance models. Stoch. Environ. Res. Risk Assess. 22 (Supplement 1), S39S47.
Marsan, D., Lengliné, O., 2008. Extending earthquakes' reach through cascading. Science 319 (5866), 1076–1079.
Marsan, D., Lengliné, O., 2010. A new estimation of the decay of aftershock density with distance to the mainshock. J. Geophys. Res. Solid Earth 115 (B9), b09302.
Mohler, G.O., Short, M.B., Brantingham, P.J., Schoenberg, F.P., Tita, G.E., 2011. Self-exciting point process modeling of crime. J. Amer. Statist. Assoc. 106 (493), 100–108.
Møller, J., 2003. Shot noise cox processes. Adv. Appl. Probab. 35 (3), 614–640.
Møller, J., Díaz-Avalos, C., 2010. Structured spatio-temporal shot-noise Cox point process models, with a view to modelling forest fires. Scand. J. Statist. 37 (1), 2–25.
Møller, J., Ghorbani, M., 2012. Aspects of second-order analysis of structured inhomogeneous spatio-temporal point processes. Stat. Neerl. 66 (4), 472–491.
Møller, J., Ghorbani, M., 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., Rasmussen, J.G., 2006. Approximate simulation of hawkes processes. Methodol. Comput. Appl. Probab. 8 (1), 53–64.
Møller, J., Safavimanesh, F., Rasmussen, J., 2015. The cylindrical K-function and Poisson line cluster point processes. ArXiv e-prints.
Møller, J., Sørensen, M., 1994. Statistical analysis of a spatial birth-and-death process model with a view to modelling linear dune fields. Scand. J. Statist. 21, 1–19.
Møller, J., Syversveen, A.R., Waagepetersen, R.P., 1998. Log Gaussian Cox processes. Scand. J. Statist. 25 (3), 451–482.
Møller, J., Toftaker, H., 2014. Geometric anisotropic spatial point pattern analysis and Cox processes. Scand. J. Statist. 41, 414–435.
Møller, J., Waagepetersen, R.P., 2004. Statistical Inference and Simulation for Spatial Point Processes. In: Chapman & Hall Monographs on Statistics & Applied Probability, CRC Press, Boca Raton, Florida.
Musmeci, F., Vere-Jones, D., 1992. A space–time clustering model for historical earthquakes. Ann. Inst. Statist. Math. 44 (1), 1–11.
Mylne, A., Brady, O.J., Huang, Z., Pigott, D.M., Golding, N., Kraemer, M.U., Hay, S.I., 2014. A comprehensive database of the geographic spread of past human ebola outbreaks. Sci. Data. URL http://www.nature.com/sdata.
Ogata, Y., 1988. Statistical models for earthquake occurrences and residual analysis for point processes. J. Amer. Statist. Assoc. 83 (401), 9–27.
Ogata, Y., 1998. Space–time point-process models for earthquake occurrences. Ann. Inst. Statist. Math. 50 (2), 379–402.
Ogata, Y., Zhuang, J., 2006. Space–time ETAS models and an improved extension. Tectonophysics 413 (1), 13–23.
Ohser, J., Stoyan, D., 1981. On the second-order and orientation analysis of planar stationary point processes. Biom. J. 23, 523–533.
Peng, R.D., Schoenberg, F.P., Woods, J.A., 2005. A spacetime conditional intensity model for evaluating a wildfire Hazard index. J. Amer. Statist. Assoc. 100 (469), 26–35.
Pereira, P., Turkman, K.F., Turkman, M.A.A., Sá, A., 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., Kie, J.G., 2004. Modeling animal movements using stochastic differential equations. Environmetrics 15 (7), 643–657.

Prokešová, M., Dvořák, J., 2013. Statistics for inhomogeneous space–time shot-noise Cox processes. Methodol. Comput. Appl. Probab. 16 (2), 433–449.
Rathbun, S.L., Cressie, N., 1994. A space–time survival point process for a longleaf pine forest in southern Georgia. J. Amer. Statist. Assoc. 89 (428), 1164–1174.
Ripley, B.D., 1977. Modelling spatial patterns (with discussion). J. R. Stat. Soc. Ser. B Stat. Methodol. 39 (2), 172–212.
Ripley, B.D., 1988. Statistical Inference for Spatial Processes. Cambridge University Press, Cambridge.
Rodrigues, A., Diggle, P.J., 2010. A class of convolution-based models for spatio-temporal processes with non-separable covariance structure. Scand. J. Statist. 37, 553–567.
Rodrigues, A., Diggle, P.J., 2012. Bayesian estimation and prediction for inhomogeneous spatiotemporal log-Gaussian Cox processes using low-rank models, with application to criminal surveillance. J. Amer. Statist. Assoc. 107 (497), 93–101.
Rodríguez-Cortés, F.J., Ghorbani, M., Mateu, J., Stoyan, D., 2014. On the expected value and variance for an estimator of the spatio-temporal product density function. In: Aalborg University Research Report Series (0), Department of Mathematical Sciences, pp. 1–19.
Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc. Ser. B Stat. Methodol. 71 (2), 319–392.
Schoenberg, F.P., 2003. Multidimensional residual analysis of point process models for earthquake occurrences. J. Amer. Statist. Assoc. 98 (464).
Schoenberg, F.P., 2004. Testing separability in spatial–temporal marked point processes. Biometrics 60 (2), 471–481.
Schoenberg, F.P., Brillinger, D.R., Guttorp, P., 2006. Point processes, spatial–temporal. In: El-Shaarawi, A.H., Piegorsch, W.W. (Eds.), Encyclopedia of Environmetrics. John Wiley & Sons, Chichester, pp. 1573–1577 (Chapter 3).
Schoenberg, F.P., Cochran, J.J., Cox, L.A., Keskinocak, P., Kharoufeh, J.P., Smith, J.C., 2010. Introduction to point processes. In: Wiley Encyclopedia of Operations Research and Management Science. John Wiley & Sons, pp. 2438–2811.
Schorlemmer, D., Gerstenberger, M., Wiemer, S., Jackson, D., Rhoades, D., 2007. Earthquake likelihood model testing. Seismol. Res. Lett. 78 (1), 17–29.
Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. In: Chapman & Hall Monographs on Statistics & Applied Probability, CRC Press, Boca Raton, Florida.
Snyder, D.L., Miller, M.I., 1991. Random Point Processes in Time and Space, second ed. Springer-Verlag, New York.
Stoyan, D., Rodíguez-Cortés, F.J., Mateu, J., Gille, W., 2016. Mark variograms for spatio-temporal point processes (unpublished).
Stoyan, D., Stoyan, H., 1994. Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics. John Wiley & Sons, Chichester.
Tamayo-Uria, I., Mateu, J., Diggle, P.J., 2014. Modelling of the spatio-temporal distribution of rat sightings in an urban environment. Spat. Stat. 9, 192–206.
Taylor, B., Diggle, P.J., 2014. INLA or MCMC? A tutorial and comparative evaluation for spatial prediction in log-Gaussian Cox processes. J. Stat. Comput. Simul. 84 (10), 2266–2284.
Van de Putte, A., Youdjou, N., Danis, B., 2015. The Antarctic biodiversity information facility. URL http://www.biodiversity.aq.
van Lieshout, M.N.M., 2000. Markov Point Processes and their Applications. Imperial College Press, London, London.
van Lieshout, M.N.M., 2011. On estimation of the intensity function of a point process. Methodol. Comput. Appl. Probab. 14 (3), 567–578.
van Lieshout, M.N.M., Baddeley, A.J., 1996. A nonparametric measure of spatial interaction in point patterns. Stat. Neerl. 50 (3), 344–361.
van Lieshout, M.N.M., Stein, A., 2012. Earthquake modelling at the country level using aggregated spatio-temporal point processes. Math. Geosci. 44 (3), 309–326.
Veen, A., 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. Environ. Ecol. Stat. 16, 173–195.
WoRMS, E.B., 2015. World register of marine species (worms). http://www.marinespecies.org.
Zhuang, J., 2006. Second-order residual analysis of spatiotemporal point processes and applications in model evaluation. J. R. Stat. Soc. Ser. B Stat. Methodol. 68 (4), 635–653.
Zhuang, J., Ogata, Y., Vere-Jones, D., 2002. Stochastic declustering of space–time earthquake occurrences. J. Amer. Statist. Assoc. 97 (458), 369–380.