CENTRE FOR **STOCHASTIC GEOMETRY**
AND ADVANCED **BIOIMAGING**

Jean-François Coeurjolly, Jesper Møller and Rasmus Waagepetersen

Conditioning in spatial point processes

No. 14, December 2015

Conditioning in spatial point processes

Jean-François Coeurjolly[1], Jesper Møller[2] and Rasmus Waagepetersen[2]

[1]Laboratory Jean Kuntzmann, Statistics Department, Grenoble Alpes University, France, email: Jean-Francois.Coeurjolly@upmf-grenoble.fr.
[2]Department of Mathematical Sciences, Aalborg University, Denmark, email: jm@math.aau.dk, rw@math.aau.dk

Abstract

This tutorial provides an introduction to conditioning in spatial point processes or so-called Palm distributions. Initially, in the context of finite point processes, we give an explicit definition of Palm distributions in terms of their density functions. Then we review Palm distributions in the general case. Finally we discuss some examples of specific models and applications.

Keywords: Cox process; Gibbs process; joint intensities; log Gaussian Cox process; Palm likelihood; reduced Palm distribution; shot noise Cox processes; summary statistics.

1 Introduction

A spatial point process $\mathbf{X}$ is briefly speaking a random subset of the d-dimensional Euclidean space $\mathbb{R}^d$, where $d = 2, 3$ are the cases of most practical importance. When studying spatial point process models and making statistical inference, the conditional distribution of $\mathbf{X}$ given a realization of $\mathbf{X}$ on some specified region or given the locations of one or more points in $\mathbf{X}$ plays an important role, see e.g. Møller and Waagepetersen (2004) and Chiu *et al.* (2013). In this paper we focus on the latter type of conditional distributions which are formally defined in terms of so-called Palm distributions, first introduced by Palm (1943) for stationary point processes on the real line. Rigorous definitions and generalizations of Palm distributions to $\mathbb{R}^d$ and more abstract spaces have mainly been developed in probability theory, see Jagers (1973) for references and an historical account. Palm distributions are, at least among many applied statisticians and among most students, considered one of the more difficult topics in the field of spatial point processes. This is partly due to the general definition of Palm distributions which relies on measure theoretical results, see e.g. Møller and Waagepetersen (2004) and Daley and Vere-Jones (2008). The account of conditional distributions for point processes in Last (1990) is mainly intended for probabilists and is not easily accessible due to an abstract setting and extensive use of measure theory.

This tutorial provides an introduction to Palm distributions for spatial point processes. Our setting and background material on point processes are given in Section 2. Section 3, in the context of finite point processes, provides an explicit definition of Palm distributions in terms of their density functions while Section 4 reviews Palm distributions in the general case. Section 5 discusses examples of Palm distributions for specific models and Section 6 considers applications of Palm distributions in the statistical literature.

2 Prerequisites

2.1 Setting and notation

We view a point process as a random locally finite subset $\mathbf{X}$ of a Borel set $S \subseteq \mathbb{R}^d$; for measure theoretical details, see e.g. Møller and Waagepetersen (2004) or Daley and Vere-Jones (2003). Denoting $\mathbf{X}_B = \mathbf{X} \cap B$ the restriction of $\mathbf{X}$ to a set $B \subseteq S$, and $N(B)$ the number of points in $\mathbf{X}_B$, local finiteness of $\mathbf{X}$ means that $N(B) < \infty$ almost surely (a.s.) whenever B is bounded. We denote by $\mathcal{B}_0$ the family of all bounded Borel subsets of S and by $\mathcal{N}$ the state space consisting of the locally finite subsets (or point configurations) of S. Section 3 considers the case where S is bounded and hence $\mathcal{N}$ is all finite subsets of S, while Section 4 deals with the general case where S is arbitrary, i.e., including the case $S = \mathbb{R}^d$.

2.2 Poisson process

The Poisson process is of its own interest and also used for constructing other point processes as demonstrated in Section 2.3 and Section 5.

Suppose $\rho : S \mapsto [0, \infty)$ is a locally integrable function, that is, $\alpha(B) := \int_B \rho(x)\,\mathrm{d}x < \infty$ whenever $B \in \mathcal{B}_0$. Then $\mathbf{X}$ is a *Poisson process* with intensity function ρ if for any $B \in \mathcal{B}_0$, $N(B)$ is Poisson distributed with mean $\alpha(B)$, and conditional on $N(B) = n$, the n points are independent and identically distributed, with a density proportional to ρ (if $\alpha(B) = 0$, then $N(B) = 0$). In fact, this definition is equivalent to that for any $B \in \mathcal{B}_0$ and any non-negative measurable function h on $\{\mathbf{x} \cap B \mid \mathbf{x} \in \mathcal{N}\}$, letting $|B|$ denote the Lebesgue measure of B,

$$\mathrm{E}h(\mathbf{X}_B) = \sum_{n=0}^{\infty} \frac{\exp(-|B|)}{n!} \int_B \cdots \int_B h(\{x_1, \ldots, x_n\})\rho(x_1)\cdots\rho(x_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n\,, \tag{2.1}$$

where for $n = 0$ the term is $\exp(-|B|)h(\emptyset)$, where $\emptyset$ is the empty point configuration.

Note that the definition of a Poisson process only requires the existence of the intensity measure α, since a point of the process restricted to $B \in \mathcal{B}_0$ has probability distribution $\alpha(\cdot \cap B)/\alpha(B)$ provided $\alpha(B) > 0$. We shall use this extension of the definition in Section 5.3.2.

2.3 Finite point processes specified by a density

Assume S is bounded, let $\mathbf{Z}$ be a unit rate Poisson process on S, and assume the distribution of $\mathbf{X}$ is absolutely continuous with respect to the distribution of $\mathbf{Z}$ (in short with respect to $\mathbf{Z}$) with density f. Thus, for any non-negative measurable function h on $\mathcal{N}$,

$$\mathrm{E}h(\mathbf{X}) = \mathrm{E}\{f(\mathbf{Z})h(\mathbf{Z})\}. \tag{2.2}$$

Moreover, by (2.1),

$$\mathrm{E}h(\mathbf{X}) = \sum_{n=0}^{\infty} \frac{\exp(-|S|)}{n!} \int_S \cdots \int_S h(\{x_1,\ldots,x_n\}) f(\{x_1,\ldots,x_n\})\,\mathrm{d}x_1 \cdots \mathrm{d}x_n. \tag{2.3}$$

This motivates considering probability statements in terms of $\exp(-|S|)f(\cdot)$. For example, with $h(\mathbf{x}) = 1(\mathbf{x} = \emptyset)$, where $1(\cdot)$ denotes indicator function, we obtain that $P(\mathbf{X} = \emptyset)$ is $\exp(-|S|)f(\emptyset)$. Further, for $n \geq 1$,

$$\exp(-|S|)f(\{x_1,\ldots,x_n\})\ \mathrm{d}x_1 \cdots \mathrm{d}x_n$$

is the probability that $\mathbf{X}$ consists of precisely n points with one point in each of n infinitesimally small disjoint sets $B_1,\ldots,B_n$ around $x_1,\ldots,x_n$ with volumes $\mathrm{d}x_1,\ldots,\mathrm{d}x_n$, respectively. Loosely speaking this event is '$\mathbf{X} = \{x_1,\ldots,x_n\}$'.

Suppose we have observed $\mathbf{X}_B = \mathbf{x}_B$ and we wish to predict the remaining point process $\mathbf{X}_{S\setminus B}$. Then it is natural to consider the conditional distribution of $\mathbf{X}_{S\setminus B}$ given $\mathbf{X}_B = \mathbf{x}_B$. By definition of a Poisson process, $\mathbf{Z}_B$ and $\mathbf{Z}_{S\setminus B}$ ($\mathbf{Z} = \mathbf{Z}_B \cup \mathbf{Z}_{S\setminus B}$) are each independent unit rate Poisson processes on respectively B and $S \setminus B$. Thus, in analogy with conditional densities for multivariate data, this conditional distribution can be specified in terms of the conditional density

$$f_{S\setminus B}(\mathbf{x}_{S\setminus B}|\mathbf{x}_B) = \frac{f(\mathbf{x}_B \cup \mathbf{x}_{S\setminus B})}{f_B(\mathbf{x}_B)}$$

with respect to $\mathbf{Z}_{S\setminus B}$ and where

$$f_B(\mathbf{x}_B) = \mathrm{E}f(\mathbf{Z}_{S\setminus B} \cup \mathbf{x}_B)$$

is the marginal density of $\mathbf{X}_B$ with respect to $\mathbf{Z}_B$. Thus the conditional distribution given a realization of $\mathbf{X}$ on some prespecified region B is conceptually quite straightforward. Conditioning on that some prespeficied points belong to $\mathbf{X}$ is more intricate but an explicit account of this is provided in the next section where it is still assumed that $\mathbf{X}$ is specified in terms of a density.

3 Palm distributions in the finite case

For understanding the definition of a Palm distribution, it is useful to assume first that S is bounded and that $\mathbf{X}$ has a density as introduced in Section 2.3 with respect to a unit rate Poisson process $\mathbf{Z}$. We make this assumption in the present section, while the general case will be treated in Section 4.

3.1 Conditional intensity and joint intensities

Suppose f is *hereditary*, i.e., for any pairwise distinct $x_0, x_1, \ldots, x_n \in S$, we have $f(\{x_1, \ldots, x_n\}) > 0$ whenever $f(\{x_0, x_1, \ldots, x_n\}) > 0$. We can then define the so-called nth order *Papangelou conditional intensity* by

$$\lambda^{(n)}(x_1, \ldots, x_n, \mathbf{x}) = f(\mathbf{x} \cup \{x_1, \ldots, x_n\})/f(\mathbf{x}) \tag{3.1}$$

for pairwise distinct $x_1, \ldots, x_n \in S$ and $\mathbf{x} \in \mathcal{N} \setminus \{x_1, \ldots, x_n\}$, setting $0/0 = 0$. By the previous interpretation of f, $\lambda^{(n)}(x_1, \ldots, x_n, \mathbf{x})\,\mathrm{d}x_1 \cdots \mathrm{d}x_n$ can be considered as the conditional probability of observing one point in each of the aforementioned infinitesimally small sets B_i, conditional on that $\mathbf{X}$ outside $\bigcup_{i=1}^n B_i$ agrees with $\mathbf{x}$.

For any $n = 1, 2, \ldots$, we define for pairwise distinct $x_1, \ldots, x_n \in S$ the nth order *joint intensity function* $\rho^{(n)}$ by

$$\rho^{(n)}(x_1, \ldots, x_n) = \mathrm{E} f(\mathbf{Z} \cup \{x_1, \ldots, x_n\}) \tag{3.2}$$

provided the right hand side exists. Particularly, $\rho = \rho^{(1)}$ is the usual *intensity* function. If f is hereditary, then $\rho^{(n)}(x_1, \ldots, x_n) = \mathrm{E}\lambda^{(n)}(x_1, \ldots, x_n, \mathbf{X})$ and by the interpretation of $\lambda^{(n)}$ it follows that $\rho^{(n)}(x_1, \ldots, x_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n$ can be viewed as the probability that $\mathbf{X}$ has a point in each of n infinitesimally small sets around $x_1, \ldots, x_n$ with volumes $\mathrm{d}x_1, \ldots, \mathrm{d}x_n$, respectively. Loosely speaking, this event is '$x_1, \ldots, x_n \in \mathbf{X}$'.

Combining (2.2) and (3.2) with either (2.3) or the extended Slivnyak-Mecke formula for the Poisson process given later in (5.1), it is straightforwardly seen that

$$\begin{aligned} &\mathrm{E} \sum_{x_1, \ldots, x_n \in \mathbf{X}}^{\neq} h(x_1, \ldots, x_n) \\ &\quad = \int_S \cdots \int_S h(x_1, \ldots, x_n) \rho^{(n)}(x_1, \ldots, x_n)\,\mathrm{d}x_1 \ldots \mathrm{d}x_n \end{aligned} \tag{3.3}$$

for any non-negative measurable function h on S^n, where $\neq$ over the summation sign means that $x_1, \ldots, x_n$ are pairwise distinct. Denoting $N = N(S)$ the number of points in $\mathbf{X}$, the left hand side in (3.3) with $h = 1$ is seen to be the factorial moment $\mathrm{E}\{N(N-1) \cdots (N-n+1)\}$.

3.2 Definition of Palm distributions in the finite case

Now, suppose $x_1, \ldots, x_n \in S$ are pairwise distinct and $\rho^{(n)}(x_1, \ldots, x_n) > 0$. Then we define the *reduced Palm distribution* of $\mathbf{X}$ given points at $x_1, \ldots, x_n$ as the distribution $\mathrm{P}^!_{x_1, \ldots, x_n}$ for the point process $\mathbf{X}^!_{x_1, \ldots, x_n}$ with density

$$f_{x_1, \ldots, x_n}(\mathbf{x}) = \frac{f(\mathbf{x} \cup \{x_1, \ldots, x_n\})}{\rho^{(n)}(x_1, \ldots, x_n)}, \qquad \mathbf{x} \in \mathcal{N},\ \mathbf{x} \cap \{x_1, \ldots, x_n\} = \emptyset, \tag{3.4}$$

with respect to $\mathbf{Z}$. If $x_1, \ldots, x_n \in S$ are not pairwise distinct or $\rho^{(n)}(x_1, \ldots, x_n)$ is zero, the choice of $\mathbf{X}^!_{x_1, \ldots, x_n}$ and its distribution $\mathrm{P}^!_{x_1, \ldots, x_n}$ is not of any importance for the results in this paper. Furthermore, the (non-reduced) *Palm distribution* of $\mathbf{X}$ given points at $x_1, \ldots, x_n$ is simply the distribution of the union $\mathbf{X}^!_{x_1, \ldots, x_n} \cup \{x_1, \ldots, x_n\}$.

3.3 Remarks

Note that by the previous infinitesimal interpretations of f and $\rho^{(n)}$, we can view $\exp(-|S|)f_{x_1,\ldots,x_n}(\mathbf{x})$ as the 'joint probability' that $\mathbf{X}$ equals the union $\mathbf{x}\cup\{x_1,\ldots,x_n\}$ divided by the 'probability' that $x_1,\ldots,x_n \in \mathbf{X}$. Thus $\mathrm{P}^!_{x_1,\ldots,x_n}$ has an interpretation as the conditional distribution of $\mathbf{X}\setminus\{x_1,\ldots,x_n\}$ given that $x_1,\ldots,x_n \in \mathbf{X}$. Conversely, by (3.4) with $\mathbf{x}=\emptyset$ and the remark just below (2.3),

$$\exp(-|S|)f(\{x_1,\ldots,x_n\}) = \rho^{(n)}(x_1,\ldots,x_n)\mathrm{P}\left(\mathbf{X}^!_{\{x_1,\ldots,x_n\}}=\emptyset\right) \tag{3.5}$$

provides a factorization into the 'probability' of observing $\{x_1,\ldots,x_n\}$ times the conditional probability of not observing further points.

We obtain immediately from (3.2) and (3.4) that for any pairwise distinct $x_1,\ldots,$ $x_n \in S$ and $m=1,2,\ldots,$ $\mathbf{X}^!_{x_1,\ldots,x_n}$ has mth order joint intensity function

$$\rho^{(m)}_{x_1,\ldots,x_n}(u_1,\ldots,u_m) = \begin{cases} \frac{\rho^{(m+n)}(u_1,\ldots,u_m,x_1,\ldots,x_n)}{\rho^{(n)}(x_1,\ldots,x_n)} & \text{if } \rho^{(n)}(x_1,\ldots,x_n)>0 \\ 0 & \text{otherwise} \end{cases} \tag{3.6}$$

for pairwise distinct $u_1,\ldots,u_m \in S\setminus\{x_1,\ldots,x_n\}$. Moreover, the so-called pair correlation function is for $u,v\in S$ defined as

$$g(u,v) = \rho^{(2)}(u,v)/\{\rho(u)\rho(v)\}$$

provided $\rho(u)\rho(v)>0$ (otherwise we set $g(u,v)=0$). If $\rho(u)\rho(v)>0$, then

$$g(u,v) = \rho_v(u)/\rho(u) = \rho_u(v)/\rho(v), \tag{3.7}$$

cf. (3.6). Thus, $g(u,v)>1$ ($g(u,v)<1$) means that the presence of a point at u yields an elevated (decreased) intensity at v and vice versa.

For later use, notice that

$$\begin{aligned} &\mathrm{E}\sum^{\neq}_{x_1,\ldots,x_n\in\mathbf{X}} h(x_1,\ldots,x_n,\mathbf{X}\setminus\{x_1,\ldots,x_n\}) \\ &\quad= \int_S\cdots\int_S \mathrm{E}h(x_1,\ldots,x_n,\mathbf{X}^!_{x_1,\ldots,x_n})\rho^{(n)}(x_1,\ldots,x_n)\,\mathrm{d}x_1\cdots\mathrm{d}x_n \end{aligned} \tag{3.8}$$

for any non-negative measurable function h on $S^n\times\mathcal{N}$. This is called the Campbell-Mecke formula and is straightforwardly verified using (2.3) and (3.4). Assuming f is hereditary and rewriting the expectation in the right hand side of (3.8) in terms of

$$f_{x_1,\ldots,x_n}(\mathbf{x}) = f(\mathbf{x})\lambda^{(n)}(x_1,\ldots,x_n,\mathbf{x})/\rho^{(n)}(x_1,\ldots,x_n)\,,$$

the finite point process case of the celebrated *Georgii-Nguyen-Zessin (GNZ) formula*

$$\begin{aligned} &\mathrm{E}\sum^{\neq}_{x_1,\ldots,x_n\in\mathbf{X}} h(x_1,\ldots,x_n,\mathbf{X}\setminus\{x_1,\ldots,x_n\}) \\ &\quad= \int_S\cdots\int_S \mathrm{E}h(x_1,\ldots,x_n,\mathbf{X})\lambda^{(n)}(x_1,\ldots,x_n,\mathbf{X})\,\mathrm{d}x_1\cdots\mathrm{d}x_n \end{aligned} \tag{3.9}$$

is obtained (Georgii, 1976; Nguyen and Zessin, 1979). We return to the GNZ formula in connection to Gibbs processes in Section 5.2.

4 Palm distributions in the general case

The definitions and results in Section 3 extend to the general case where S is any Borel subset of $\mathbb{R}^d$. However, if $|S| = \infty$, the unit rate Poisson process on S will be infinite and we can not in general assume that $\mathbf{X}$ is absolutely continuous with respect to the distribution of this process. Thus we do not longer have the direct definitions (3.2) and (3.4) of $\rho^{(n)}$ and $\mathbf{X}^!_{x_1,\ldots,x_n}$ in terms of density functions.

4.1 Definition of Palm distributions in the general case

In fact (3.3) is usually taken as the definition of the nth order joint intensity for $\mathbf{X}$, provided there exists such a non-negative measurable function $\rho^{(n)}$. Technically speaking, viewing the left hand side in (3.3) as an integral $\int h \, \mathrm{d}\alpha^{(n)}$, where $\alpha^{(n)}$ is called the nth order factorial moment measure on S^n, $\rho^{(n)}$ is assumed to be a density for $\alpha^{(n)}$ with respect to Lebesgue measure on S^n. Here, $\alpha^{(n)}$ is required to be a locally finite measure, i.e.,

$$\int_B \cdots \int_B \rho^{(n)}(x_1,\ldots,x_n) \, \mathrm{d}x_1 \cdots \mathrm{d}x_n < \infty \qquad \text{for all } B \in \mathcal{B}_0. \tag{4.1}$$

Thereby the Campbell-Mecke formula (3.8) can be used as the definition of the reduced Palm distributions, where their existence follows by measure theoretical arguments, see e.g. Møller and Waagepetersen (2004). This definition thus extends in a mathematically sound manner the previous definition of Palm distributions to point processes in general but seems intuitively less appealing.

4.2 Remarks

In the general setting, $\rho^{(n)}(x_1,\ldots,x_n)$ and $\mathrm{P}^!_{x_1,\ldots,x_n}$ are then clearly only determined up to a Lebesgue nullset of S^n. For simplicity and since usually there are natural choices of $\rho^{(n)}(x_1,\ldots,x_n)$ and $\mathrm{P}^!_{x_1,\ldots,x_n}$, such nullsets are often ignored. Further, like in the finite case, $\rho^{(n)}(x_1,\ldots,x_n)$ and $\mathrm{P}^!_{x_1,\ldots,x_n}$ are invariant under permutations of the points $x_1,\ldots,x_n$, and

$$\left(\mathbf{X}^!_{x_1,\ldots,x_m}\right)^!_{x_{m+1},\ldots,x_n} = \mathbf{X}^!_{x_1,\ldots,x_n} \tag{4.2}$$

whenever $0 < m < n$ and $x_1,\ldots,x_n$ are pairwise distinct.

Suppose that $\mathbf{X}$ is *stationary*, i.e., its distribution is invariant under translations in $\mathbb{R}^d$ and so $S = \mathbb{R}^d$ (unless $\mathbf{X} = \emptyset$ which is not a case of our interest). This is a specially tractable case, which makes an alternative description of Palm distributions possible. Let ρ denote the constant intensity of $\mathbf{X}$ and let o denote the origin in $\mathbb{R}^d$. First, we define

$$\mathrm{P}^!_o(F) = \frac{1}{\rho|B|} \mathrm{E} \sum_{x \in \mathbf{X}_B} 1(\mathbf{X} \setminus \{x\} - x \in F) \tag{4.3}$$

for any $B \in \mathcal{B}_0$ with $|B| > 0$, where by stationarity of $\mathbf{X}$ the right hand side does not depend on the choice of B. Second, we define

$$\mathrm{P}^!_x(F) = \mathrm{P}^!_o(F - x) \tag{4.4}$$

for any $x \in \mathbb{R}^d$. One can then check that the $\mathrm{P}^!_x$, $x \in \mathbb{R}^d$, defined in this way satisfy (3.8) so that (4.4) indeed defines a Palm distribution, see Appendix C.2 in Møller and Waagepetersen (2004) for details. Note that (4.4) is equivalent to that $\mathbf{X}^!_x - x$ and $\mathbf{X}^!_o$ are identically distributed. The reduced Palm distribution $\mathrm{P}^!_o$ is often interpreted as the 'conditional distribution for the further points in $\mathbf{X}$ given a typical point of $\mathbf{X}$'.

5 Examples of Palm distributions

For some classes of point processes, explicit characterizations of the Palm distributions are possible. Below we consider Poisson processes, Gibbs processes, and log Gaussian Cox processes (LGCPs) which share the property that their Palm distributions of any order are again respectively Poisson, Gibbs, and LGCPs. We also consider shot-noise Cox processes, where one point Palm distributions are not shot-noise Cox processes but have simple characterizations as cluster processes.

5.1 Poisson processes

In the finite case, by (2.1), a Poisson process $\mathbf{X}$ with intensity function ρ has density $f(\mathbf{x}) \propto \prod_{u \in \mathbf{x}} \rho(u)$, and so by (3.4), $\mathbf{X}^!_{x_1,\dots,x_n}$ is distributed as $\mathbf{X}$. In the general case, we appeal to the extended Slivnyak-Mecke formula which for a Poisson process $\mathbf{X}$ with intensity function ρ states that

$$\begin{aligned} \mathrm{E} \sum_{x_1,\dots,x_n \in \mathbf{X}}^{\neq} & h(x_1,\dots,x_n,\mathbf{X} \setminus \{x_1,\dots,x_n\}) \\ &= \int_S \cdots \int_S \mathrm{E} h(x_1,\dots,x_n,\mathbf{X}) \rho(x_1) \cdots \rho(x_n) \, \mathrm{d}x_1 \cdots \mathrm{d}x_n \end{aligned} \tag{5.1}$$

for any non-negative measurable function h on $S^n \times \mathcal{N}$, see Theorem 3.3 in Møller and Waagepetersen (2004) and the references therein. This implies that

$$\rho^{(n)}(x_1,\dots,x_n) = \rho(x_1) \cdots \rho(x_n)$$

and $\mathbf{X}^!_{x_1,\dots,x_n}$ is just distributed as $\mathbf{X}$. In fact, the property that $\mathbf{X}^!_x \sim \mathbf{X}$ for all $x \in S$ is characterizing the Poisson process, see e.g. Proposition 5 in Jagers (1973). Further, it makes it possible to calculate various useful functional summaries, see e.g. Møller and Waagepetersen (2004), and constructions such as stationary Poisson-Voronoi tessellations become manageable, see Møller (1989, 1994).

5.2 Gibbs processes

Gibbs processes play an important role in statistical physics and spatial statistics, see Møller and Waagepetersen (2004) and the references therein. Below, for ease of presentation, we consider first a finite Gibbs process.

A finite Gibbs process on a bounded set $S \subset \mathbb{R}^d$ is usually specified in terms of its density or equivalently in terms of the Papangelou conditional intensity, where the density is of the form

$$f(\mathbf{x}) = \exp\Big\{-\sum_{\mathbf{y}\subseteq\mathbf{x}} \Phi(\mathbf{y})\Big\}$$

for a so-called potential function Φ on $\mathcal{N}$. It follows that the nth order Palm distribution of a Gibbs process with respect to $x_1,\dots,x_n$ is itself a Gibbs process with potential function $\Phi_{x_1,\dots,x_n}(\mathbf{y}) = \Phi(\{x_1,\dots,x_n\}\cup\mathbf{y})$. Moreover, for pairwise distinct $u_1,\dots,u_m,x_1,\dots,x_n \in S$ and $\mathbf{x}\in\mathcal{N}\setminus\{u_1,\dots,u_m,x_1,\dots,x_n\}$, the mth order Papangelou conditional intensity of $\mathbf{X}^!_{x_1,\dots,x_n}$ is simply

$$\lambda^{!(m)}_{x_1,\dots,x_n}(u_1,\dots,u_m,\mathbf{x}) = \lambda^{(m)}(u_1,\dots,u_m,\mathbf{x}\cup\{x_1,\dots,x_n\}).$$

For instance, a first order inhomogeneous pairwise interaction Gibbs point process has first order potential $\Phi(\{u\}) = \Phi_1(u)$, second order potential $\Phi(\{u,v\}) = \Phi_2(v-u)$, and $\Phi(\mathbf{y}) = 0$ whenever the cardinality of $\mathbf{y}$ is larger than two; see Møller and Waagepetersen (2004) for conditions on the functions Φ_1 and Φ_2 ensuring that the model is well-defined. The Strauss model (Strauss, 1975; Kelly and Ripley, 1976) is a particular case with $\Phi_1(u) = \theta_1 \in \mathbb{R}$ and $\Phi_2(u-v) = \theta_2 1(\|u-v\| \le R)$, for $\theta_2 \ge 0$ and $0 < R < \infty$. The Palm process $\mathbf{X}^!_{x_1,\dots,x_n}$ becomes again an inhomogeneous pairwise interaction Gibbs process with inhomogeneous first order potential $\Phi_{x_1,\dots,x_n}(\{u\}) = \Phi_1(u)+\sum_{i=1}^n \Phi_2(u-x_i)$ and second order potential identical to that of $\mathbf{X}$.

In the general case, a Gibbs process can be defined (Nguyen and Zessin, 1979) in terms of the GNZ formula (3.9) briefly discussed at the end of Section 3: $\mathbf{X}$ is a Gibbs point process with Papangelou conditional intensity λ if λ is a non-negative measurable function on $S\times\mathcal{N}$ such that

$$\mathrm{E}\sum_{x\in\mathbf{X}} h(x,\mathbf{X}\setminus\{x\}) = \mathrm{E}\int_S \lambda(x,\mathbf{X})h(x,\mathbf{X})\,\mathrm{d}x \tag{5.2}$$

for any non-negative measurable function h on $S\times\mathcal{N}$. For conditions ensuring that (5.2) holds, we refer to Ruelle (1969) and Georgii (1988).

By the extensions of (3.3) and (3.8) to the general case, (5.2) implies $\rho(x) = \mathrm{E}\lambda(x,\mathbf{X})$. Unfortunately, in general it is not feasible to express $\rho(x) = \mathrm{E}\lambda(x,\mathbf{X})$ on closed form, though approximations exist (Baddeley and Nair, 2012). Also, for Gibbs processes, the pair correlation function $g(u,v)$ can be below or above 1 depending on u and v (see e.g. pages 240-241 in Illian *et al.*, 2008), and so from (3.7), $\rho_v(u)$ may be smaller or larger than $\rho(u)$, depending on u and v. Moreover, for pairwise distinct $x_1,\dots,x_n\in S$, $\mathrm{P}^!_{x_1,\dots,x_n}$ is absolutely continuous with respect to the distribution of $\mathbf{X}$, with density $\lambda^{(n)}(x_1,\dots,x_n,\cdot)$, where

$$\begin{aligned}\lambda^{(n)}(x_1,\dots,x_n,\mathbf{x}) = {} & \lambda(x_1,\mathbf{x})\lambda(x_2,\mathbf{x}\cup\{x_1\}) \\ & \cdots\lambda(x_n,\mathbf{x}\cup\{x_1,\dots,x_{n-1}\})\end{aligned}$$

for $x_1,\dots,x_n\in S$ and $\mathbf{x}\in\mathcal{N}$. This follows from (3.9) and (5.2) and is in accordance with (3.1) and (4.2).

5.3 Cox processes

Let $\mathbf{\Lambda} = \{\Lambda(x)\}_{x\in S}$ be a non-negative random field such that $\mathbf{\Lambda}$ is locally integrable a.s., that is, for any $B \in \mathcal{B}_0$, the integral $\int_B \Lambda(x)\,\mathrm{d}x$ exists and is finite a.s. Suppose $\mathbf{X}$ is a Cox process with random intensity function $\mathbf{\Lambda}$, i.e., conditional on $\mathbf{\Lambda}$, $\mathbf{X}$ is a Poisson process with intensity function $\mathbf{\Lambda}$. Apart from very simple models of $\mathbf{\Lambda}$ such as all $\Lambda(x)$ being equal to the same random variable following e.g. a gamma distribution, the density of $\mathbf{X}$ restricted to a set $B \in \mathcal{B}_0$ is intractable. However, if $\mathbf{\Lambda}$ has moments of any order $n = 1, 2, \ldots$, then by conditioning on Λ and using (3.8) and (5.1), we immediately get the following: For any pairwise distinct $x_1, \ldots, x_n \in S$ and any non-negative measurable function h on $S^n \times \mathcal{N}$, the product densities are

$$\rho^{(n)}(x_1, \ldots, x_n) = \mathrm{E}\Big\{\prod_{i=1}^{n} \Lambda(x_i)\Big\} \tag{5.3}$$

and the reduced Palm distributions satisfy

$$\begin{aligned} &\mathrm{E}\left\{h\left(x_1, \ldots, x_n, \mathbf{X}^!_{x_1,\ldots,x_n}\right)\right\} \rho^{(n)}(x_1, \ldots, x_n) \\ &= \mathrm{E}\Big\{h(x_1, \ldots, x_n, \mathbf{X}) \prod_{i=1}^{n} \Lambda(x_i)\Big\}. \end{aligned} \tag{5.4}$$

In the sequel, we consider distributions of $\mathbf{\Lambda}$, where (5.3)-(5.4) become useful.

5.3.1 Log Gaussian Cox processes

Let $\Lambda(x) = \exp\{Y(x)\}$, where $\mathbf{Y} = \{Y(x)\}_{x\in S}$ is a Gaussian process with mean function μ and covariance function c so that $\mathbf{\Lambda}$ is locally integrable a.s. (simple conditions ensuring this are given in Møller *et al.*, 1998). Then $\mathbf{X}$ is a log Gaussian Cox process (LGCP) as introduced by Coles and Jones (1991) in astronomy and independently by Møller *et al.* (1998) in statistics. By Møller *et al.* (1998, Theorem 1), for pairwise distinct $x_1, \ldots, x_n \in S$,

$$\rho^{(n)}(x_1, \ldots, x_n) = \Big\{\prod_{i=1}^{n} \rho(x_i)\Big\}\Big\{\prod_{1\le i<j\le n} g(x_i, x_j)\Big\}, \tag{5.5}$$

where $\rho(x) = \exp\{\mu(x) + c(x, x)/2\}$ is the intensity function and the pair correlation function (3.7) is $g(u, v) = \exp\{c(u, v)\}$. The intensity of $\mathbf{X}^!_{x_1,\ldots,x_n}$ takes the form

$$\rho_{x_1,\ldots,x_n}(u) = \rho(u) \prod_{i=1}^{n} g(u, x_i) \tag{5.6}$$

so in the common case where c is positive, the intensity of $\mathbf{X}^!_{x_1,\ldots,x_n}$ is larger than that of $\mathbf{X}$.

In Coeurjolly *et al.* (2015) it is verified that for pairwise distinct $x_1, \ldots, x_n \in S$, $\mathbf{X}^!_{x_1,\ldots,x_n}$ is an LGCP with underlying Gaussian process $\{Y(x) + \sum_{i=1}^{n} c(x, x_i)\}_{x\in S}$. Note that this Gaussian process also has covariance function c but its mean function is $\mu_{x_1,\ldots,x_n}(x) = \mu(x) + \sum_{i=1}^{n} c(x, x_i)$. Coeurjolly *et al.* (2015) discuss how this result can be exploited for functional summaries. Moreover, if the covariance function c is non-negative, $\mathbf{X}$ is distributed as an independent thinning of $\mathbf{X}^!_{x_1,\ldots,x_n}$ with inclusion probabilities $p(x) = \exp\{-\sum_{i=1}^{n} c(x, x_i)\}$.

5.3.2 Shot noise Cox processes

For a shot noise Cox process (Møller, 2003),

$$\Lambda(x) = \sum_j \gamma_j k(c_j, x),$$

where $k(c_j, \cdot)$ is a kernel (i.e., a density function for a continuous d-dimensional random variable) and the (c_j, γ_j) are the points of a Poisson process $\mathbf{\Phi}$ on $\mathbb{R}^d \times (0, \infty)$ with intensity measure α so that $\mathbf{\Lambda}$ becomes locally integrable a.s. It can be viewed as a cluster process $\mathbf{X} = \bigcup_j \mathbf{Y}_j$, where conditional on $\mathbf{\Phi}$, the cluster $\mathbf{Y}_j$ is a Poisson process with intensity function $\gamma_j k(c_j, \cdot)$ and the clusters are independent.

The intensity function is

$$\rho(x) = \int \gamma k(c, x)\, \mathrm{d}\alpha(c, \gamma),$$

provided the integral is finite for all $x \in S$. Making this assumption, it can be verified (Proposition 2 in Møller, 2003) that for $x \in S$ with $\rho(x) > 0$, $\mathbf{X}_x^!$ is a Cox process with random intensity function $\Lambda(\cdot) + \Lambda_x(\cdot)$, where $\Lambda_x(\cdot) = \gamma_x k(c_x, \cdot)$, and where (c_x, γ_x) is a random variable independent of $\mathbf{\Phi}$ and defined on $S \times (0, \infty)$ such that for any Borel set $B \subseteq S \times (0, \infty)$,

$$\mathrm{P}\left\{(c_x, \gamma_x) \in B\right\} = \frac{\int_B \gamma k(c, x)\, \mathrm{d}\alpha(c, \gamma)}{\rho(x)}.$$

In other words, $\mathbf{X}_x^!$ is distributed as $\mathbf{X} \cup \mathbf{Y}_x$, where $\mathbf{Y}_x$ is independent of $\mathbf{X}$ and conditional on (c_x, γ_x), the 'extra cluster' $\mathbf{Y}_x$ is a finite Poisson process with intensity function $\gamma_x k(c_x, \cdot)$. Thus, as for an LGCP with positive covariance function, $P_x^!$ has a higher intensity than $\mathbf{X}_x^!$.

For instance, if $\mathrm{d}\alpha(c, \gamma) = \mathrm{d}c\, \mathrm{d}\chi(\gamma)$, where χ is a locally finite measure on $(0, \infty)$, then $\rho(x) = \kappa f(x)$, where it is assumed that $\kappa = \int \gamma\, \mathrm{d}\chi(\gamma) < \infty$ and $f(x) = \int k(c, x)\, \mathrm{d}c < \infty$, and furthermore, for $\rho(x) > 0$, c_x and γ_x are independent, c_x follows the density $k(\cdot, x)/f(x)$, and $\mathrm{P}(\gamma_x \in A) = \kappa^{-1} \int_A \gamma\, \mathrm{d}\chi(\gamma)$. The special case of a Neyman-Scott process (Neyman and Scott, 1958) occurs when $S = \mathbb{R}^d$, χ is concentrated at a given value $\gamma > 0$, $\chi(\{\gamma\}) < \infty$, and $k(c, \cdot) = k_o(\cdot - c)$, where k_o is a density function. Then $\mathbf{X}$ is stationary, $\rho = \kappa = \gamma \chi(\{\gamma\})$, c_x has density $k_o(x - \cdot)$, and conditional on c_x, $\mathbf{Y}_x$ is a finite Poisson process with intensity function $\gamma k_o(\cdot - c_x)$. Examples include a (modified) Thomas process, where k_o is a zero-mean normal density, and a Matérn cluster process, where k_o is a uniform density on a ball centered at the origin. For $n > 1$, the nth order reduced Palm distributions become more complicated.

In a Neyman-Scott process, the number of points in the clusters are independent and identically Poisson distributed. A stationary Poisson cluster process is obtained by replacing the Poisson distribution by any discrete distribution on the non-negative integers. Finally, we notice that the Palm distribution for stationary Poisson cluster processes and more generally infinitely divisible point processes can also be derived, see Chiu *et al.* (2013) and the references therein.

6 Examples of applications

In this section we review a number of applications of Palm distributions in spatial statistics.

6.1 Functional summary statistics

Below we briefly consider two popular functional summary statistics, which are used for exploratory purposes as well as model fitting and model assessment.

First, suppose $\mathbf{X}$ is stationary, with intensity $\rho > 0$. The nearest-neighbour distribution function G is defined by $G(t) = \mathrm{P}_o^!\{\mathbf{X} \cap b(o,t) \neq \emptyset\}$, where $b(o,t)$ is the ball centered at o and of radius $t > 0$. Thus $G(t)$ is interpreted as the probability of having a point within distance t from a typical point. Moreover, Ripley's K-function (Ripley, 1976) times ρ is defined by $\rho K(t) = \mathrm{E} \sum_{v \in \mathbf{X}_o^!} 1(\|v\| \leq t)$, that is, the expected number of further points within distance t of a typical point.

Second, if the pair correlation function $g(u,v) = g_0(u-v)$ only depends on $v-u$ (see (3.7)), the definition of the K-function can be extended: The inhomogeneous K-function (Baddeley *et al.*, 2000) is defined by

$$K(t) = \int_{\|v\| \leq t} g_0(v) \, \mathrm{d}v.$$

By (3.7), it follows that

$$K(t) = \mathrm{E} \sum_{v \in \mathbf{X}_u^!} \frac{1(\|v-u\| \leq t)}{\rho(v)}$$

for any $u \in S$ with $\rho(u) > 0$. If for $\|v-u\| \leq t$, $\rho(v)$ is close to $\rho(u)$, we obtain $\rho(u)K(t) \approx \mathrm{E} \sum_{v \in \mathbf{X}_u^!} 1(\|v-u\| \leq t)$. This is a 'local' version of the interpretation of $K(t)$ in the stationary case.

Nonparametric estimation of K and G is based on empirical versions obtained from (4.3). For some parametric Poisson and Cox process models, K or G are expressible on closed form and may be compared with corresponding nonparametric estimates when finding parameter estimates or assessing a fitted model. See Møller and Waagepetersen (2007) and the references therein.

6.2 Prediction given partial observation of a point process

Suppose S is bounded and we observe a point process $\mathbf{Y}$ contained in a finite point process $\mathbf{X}$ specified by some density f with respect to the unit rate Poisson process $\mathbf{Z}$. If $B \subset S$ with $|B| > 0$ and $\mathbf{Y} = \mathbf{X}_B$, then prediction of $\mathbf{X}_{S \setminus B}$ given $\mathbf{Y} = \mathbf{y}$ can be based on the conditional density $f_{S \setminus B}(\cdot|\mathbf{y})$ introduced in Section 2.3. On the other hand, if we just know that $\mathbf{y} \subseteq \mathbf{X}$, then it could be tempting to try to predict $\mathbf{X} \setminus \mathbf{y}$ using $\mathbf{X}_{\mathbf{y}}^!$. This would in general be incorrect. For instance, for an LGCP with positive covariance function, the intensity of $\mathbf{X}_{\mathbf{y}}^!$ can be much larger than the one of $\mathbf{X}$, cf. (5.6). Thus on average $\mathbf{X}_{\mathbf{y}}^! \cup \mathbf{y}$ would contain more points than $\mathbf{X}$. The issue here is that the reduced Palm distribution is concerned with the conditional distribution of $\mathbf{X}$ conditional on that *prespecified* points fall in $\mathbf{X}$. Hence the sampling mechanism that leads from $\mathbf{X}$ to $\mathbf{Y}$ must be taken into account. For instance, if the distribution of $\mathbf{Y}$ conditional on $\mathbf{X} = \mathbf{x}$ is specified by a probability density function $p(\cdot|\mathbf{x})$ (on the set of all subsets of $\mathbf{x}$), then by Proposition 1 in Baddeley *et al.* (2000), the marginal density of $\mathbf{Y}$ with respect to $\mathbf{Z}$ is

$$g(\mathbf{y}) = \rho^{(n)}(\mathbf{y}) \exp(|S|) \mathrm{E}\left\{p(\mathbf{y}|\mathbf{X}^!_{\mathbf{y}} \cup \mathbf{y})\right\},$$

where $n = n(\mathbf{y})$ is the cardinality of $\mathbf{y}$. Thus the conditional distribution of $\mathbf{X} \setminus \mathbf{y}$ given $\mathbf{Y} = \mathbf{y}$ has density

$$f(\mathbf{x}|\mathbf{y}) = p(\mathbf{y}|\mathbf{x} \cup \mathbf{y}) f(\mathbf{x} \cup \mathbf{y}) \exp(|S|)/g(\mathbf{y})$$

with respect to $\mathbf{Z}$.

6.3 Matérn-thinned Cox processes

Some applications of spatial point processes require models that combine clustering at a large scale with regularity at a local scale (Lavancier and Møller, 2015). Andersen and Hahn (2015) study a class of so-called Matérn thinned Cox processes where (clustered) Cox processes are subjected to dependent Matérn type II thinning (Matérn, 1986) that introduces regularity in the resulting point processes. The intensity function and second-order joint intensity of the Matérn-thinned Cox process is expressed in terms of univariate and bivariate inclusion probabilities which in turn are expressed in terms of one- and two-point Palm probabilities for an independently marked version of the underlying Cox process. In case of an underlying shot-noise Cox process, explicit expressions for the univariate inclusion probabilities are obtained using the simple characterization of one-point Palm distributions described in Section 5.3.2.

6.4 Palm likelihood

Minimum contrast estimators based on the K-function or the pair correlation function or composite likelihood methods are standard methods to fit parametric models (see e.g. Jolivet, 1991; Guan, 2006; Møller and Waagepetersen, 2007; Waagepetersen and Guan, 2009; Biscio and Lavancier, 2015). Tanaka *et al.* (2008) proposed an approach based on Palm intensities to fit parametric stationary models, which is briefly presented below.

Given a parametric model $g(u, v) = g_0(v-u; \theta)$ for the pair correlation function of $\mathbf{X}$ and a location $u \in S$, the intensity function of $\mathbf{X}^!_u$ is $\rho_u(v; \theta) = \rho g_0(v-u; \theta)$ where ρ is the constant intensity of $\mathbf{X}$ assumed here to be known. Following Schoenberg (2005), the so-called log composite likelihood score

$$\sum_{v \in \mathbf{X}^!_u \cap b(u,R)} \frac{\mathrm{d}}{\mathrm{d}\theta} \log \rho_u(v; \theta) - \int_{b(u,R)} \rho_u(v; \theta) \, \mathrm{d}v$$

forms an unbiased estimating function for θ, where $R > 0$ is a user-specified tuning parameter. Usually $\mathbf{X}^!_u$ is not known. However, suppose that $\mathbf{X}$ is observed on $W \in$

$\mathcal{B}_0$ and in order to introduce a border correction let $W \ominus R = \{u \in W \mid b(u,R) \subseteq W\}$. Then, by (4.3),

$$\sum_{\substack{u \in \mathbf{X} \cap W \ominus R, \\ v \in \mathbf{X} \cap b(u,R)}}^{\neq} \frac{\mathrm{d}}{\mathrm{d}\theta} \log \rho_u(v;\theta) - N(W \ominus R) \int_{b(o,R)} \rho_o(v;\theta)\,\mathrm{d}v \tag{6.1}$$

is an unbiased estimate of the above composite likelihood score times $\rho|W \ominus R|$. Tanaka *et al.* (2008) coined the antiderivative of (6.1) the Palm likelihood. Asymptotic properties of Palm likelihood parameter estimates are studied by Prokešová and Jensen (2013) who also proposed the border correction applied in (6.1).

Acknowledgments

J. Møller and R. Waagepetersen are supported by the Danish Council for Independent Research | Natural Sciences, grant 12-124675, "Mathematical and Statistical Analysis of Spatial Data", and by the "Centre for Stochastic Geometry and Advanced Bioimaging", funded by grant 8721 from the Villum Foundation. J.-F. Coeurjolly is supported by ANR-11-LABX-0025 PERSYVAL-Lab (2011, project OculoNimbus).

References

Andersen, I. & Hahn, U. (2015). Matérn thinned Cox point processes. *Spatial Statistics* `http://dx.doi.org/10.1016/j.spatsta.2015.10.005`.

Baddeley, A. & Nair, G. (2012). Fast approximation of the intensity of Gibbs point processes. *Electronic Journal of Statistics* **6**, 1155–1169.

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

Biscio, C. & Lavancier, F. (2015). Contrast estimation for parametric stationary determinantal point processes. Submitted for publication. Available at arXiv:1510.04222.

Chiu, S., Stoyan, D., Kendall, W. S. & Mecke, J. (2013). *Stochastic Geometry and Its Applications*. Wiley, Chichester, 3rd edition.

Coeurjolly, J.-F., Møller, J. & Waagepetersen, R. (2015). Palm distributions for log Gaussian Cox processes. Submitted for publication. Available at arXiv:1506.04576.

Coles, P. & Jones, B. (1991). A lognormal model for the cosmological mass distribution. *Monthly Notices of the Royal Astronomical Society* **248**, 1–13.

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

Daley, D. & Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes, Volume II: General Theory and Structure*. Springer, New York, 2nd edition.

Georgii, H.-O. (1976). Canonical and grand canonical Gibbs states for continuum systems. *Communications in Mathematical Physics* **48**, 31–51.

Georgii, H.-O. (1988). *Gibbs Measures and Phase Transition*. Walter de Gruyter, Berlin.

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

Illian, J., Penttinen, A., Stoyan, H. & Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. Statistics in Practice, Wiley, Chichester.

Jagers, P. (1973). On Palm probabilities. *Probability Theory and Related Fields* **26**, 17–32.

Jolivet, E. (1991). Moment estimation for stationary point processes in $\mathbf{R}^d$. In: *Spatial Statistics and Imaging (Brunswick, ME, 1988)*, volume 20 of *IMS Lecture Notes Monograph Series*, Institute Mathematical Statistics, Hayward, CA, 138–149.

Kelly, F. & Ripley, B. (1976). A note on Strauss' model for clustering. *Biometrika* **63**, 357–360.

Last, G. (1990). Some remarks on conditional distributions for point processes. *Stochastic Processes and Their Applications* **34**, 121–135.

Lavancier, F. & Møller, J. (2015). Modelling aggregation on the large scale and regularity on the small scale in spatial point pattern datasets. *Research Report 7*, Centre for Stochastic Geometry and Advanced Bioimaging, to appear in Scandinavian Journal of Statistics.

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

Møller, J. (1989). Random tessellations in $\mathbf{R}^d$. *Advances in Applied Probability* **21**, 37–73.

Møller, J. (1994). *Lectures on Random Voronoi Tessellations*. Lecture Notes in Statistics 87, Springer-Verlag, New York.

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

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

Møller, J. & Waagepetersen, R. P. (2007). Modern spatial point process modelling and inference (with discussion). *Scandinavian Journal of Statistics* **34**, 643–711.

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

Neyman, J. & Scott, E. L. (1958). Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **20**, 1–43.

Nguyen, X. & Zessin, H. (1979). Integral and differential characterizations of Gibbs processes. *Mathematische Nachrichten* **88**, 105–115.

Palm, C. (1943). Intensitätsschwankungen im Fernsprechverkehr. *Ericssons Techniks* **44**, 1–189.

Prokešová, M. & Jensen, E. (2013). Asymptotic Palm likelihood theory for stationary point processes. *Annals of the Institute of Statistical Mathematics* **65**(2), 387–412.

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

Ruelle, D. (1969). *Statistical Mechanics: Rigorous Results*. W.A. Benjamin, Reading, Massachusetts.

Schoenberg, F. (2005). Consistent parametric estimation of the intensity of a spatial-temporal point process. *Journal of Statistical Planning and Inference* **128**, 79–93.

Strauss, D. (1975). A model for clustering. *Biometrika* **63**, 467–475.

Tanaka, U., Ogata, Y. & Stoyan, D. (2008). Parameter estimation for Neyman-Scott processes. *Biometrical Journal* **50**, 43–57.

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